ABSTRACT

BEEMAN, JENNIFER LEIGH SLOAN. Assessing Understanding of Sampling Distributions and Differences in Learning Amongst Different Learning Styles. (Under the direction of Dr. Roger Woodard and Dr. David Dickey.)

Research has found that students successfully complete an introductory course in statistics without fully comprehending the underlying theory or being able to exhibit statistical reasoning. This is particularly true for the understanding about the sampling distribution of the mean, a crucial concept for statistical inference.

This study investigates two different types of students, sensor and intuitor, based on a preferred style of learning, in how they answer questions about sampling distributions and how they demonstrate reasoning and logic regarding these answers. Ten sections of introductory statistics are taught at North Carolina State University over three semesters and data is collected from eight final exam questions. These final exam questions are analyzed using both quantitative and qualitative analysis techniques. Differences between the groups are summarized, if they exist. This study also explores foundational questions and their relationship to understanding the sampling distribution of the mean.

The analysis of data showed differences between the sensor and intuitor on three of the eight questions, but the level of significance is not very high. No evidence is found to support that if a student understands a series of questions thought to be foundational to understanding of sampling distributions then they should exhibit statistical reasoning. Although there was no evidence to show that there is a largely observable difference between the sensor and intuitor on all eight activities, there is evidence to support common

misconceptions that students reason about a sampling distribution of means as though it were a simple random sample. Further research is needed to investigate the lexical ambiguity amongst key terms and definitions and how it affects the performance in the classroom as well as how students fail to make the connection between a simple random sample and a sampling distribution of the mean.

Assessing Understanding of Sampling Distributions and Differences in Learning Amongst Different Learning Styles

by

Jennifer Leigh Sloan Beeman

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

Statistics

Raleigh, North Carolina

2013

APPROVED BY:

Dr. Herle McGowan	Dr. Hollylynne Stohl Lee
Dr. David Dickey Vice-Chair of Advisory Committee	Dr. Roger Woodard Chair of Advisory Committee

DEDICATION

Dedicated to my daddy, I will always be his little girl. To all my children, this is for you.

BIOGRAPHY

29 August 1980……………Born – Greenville, SC

2002……………………...…B.S. Mathematics, Charleston Southern University.

2004……………..………… M.S. Statistics, North Carolina State University.

2002 – 2005……………….. Teaching Assistant, Department of Statistics,

North Carolina State University

2005-2006…………………. Student Employee, Data Mining,

Glaxo Smith Kline Pharmaceuticals.

2005 – 2006…………………Research Assistant,

Statistical and Applied Mathematical Sciences Institute

2005-present……………….. Advanced Analytics Software Tester,

Forecasting and Time Series, SAS Institute.

Publications

Leonard, Michael, Jennifer (Sloan) Lee, Taiyeong Lee, Bruce Elsheimer. "An Introduction to Similarity Analysis using SAS®"

http://www.forecasters.org/submissions08/MichaelLeonardISF2008.pdf

Fields of Study

Major Field: Statistics

Area of Emphasis: Statistics Education

ACKNOWLEDGEMENTS

First and foremost I want to thank my advisor Dr. Roger Woodard for his patience and encouragement as I traveled along this journey.

I want to thank my committee members, Dr. Stohl Lee and Dr. McGowan for providing me with invaluable feedback and support throughout this process. I also want to thank Dr. Dickey for his willingness to jump into this project and provide his invaluable insight and wisdom.

I am grateful for two of my colleagues at SAS Institute, Armistead Sapp and Robert Allison, for providing insight, motivation and guidance. I don't know anyone else who would choose to proof read a dissertation for fun in order to help a friend. And I thank all of my managers at SAS; Bruce Elsheimer, Jim Metcalf, and Bengt Pederson, for keeping me gainfully employed, and especially Dr. Jim Goodnight, for encouraging me to finish.

Finally, I want to thank my family for being so supportive of me over the last 10 years; most importantly, my husband, Rick Beeman who joined me in this journey only during the latter part, and pushed me to finish, and my beautiful daughter Sloan Violet, for sacrificing countless nights of cuddle time in order to let me finish writing. I also want to thank my uncles, Dr. Lee Sloan, and Dr. Jerry McDowell, my aunt Dolores McDowell, and especially, my parents, David and Judy Sloan; for being so supportive throughout the past decade. I could not have done this without any of you and for that I am very grateful.

I write this in memory of Dr. Kay Clayton whose last words to me were, "Do NOT give up on your PhD!!! One never knows what the future holds, so get that degree!"

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1: Introduction

Perhaps the most important topic for introductory statistics is the concept of sampling distributions. It is a lynch pin for most of the topics in introductory statistics, such as confidence intervals and hypothesis tests and it is often viewed as the key to statistical inference. I explored the topic of sampling distributions and how students are learning interrelated concepts in an introductory statistics course at North Carolina State University.

Rationale of the Study

The idea of the sampling distribution of the mean and its associated characteristics such as center, shape and variability are core concepts presented in an introductory statistics course. It is obvious, from both personal observation and the literature in statistics education, students struggle with understanding the concept of a sampling distribution of the mean.

The American Statistical Association (ASA) released a report presenting guidelines for the assessment and instruction in statistics education at the college level in February of 2005, called Guidelines for Assessment and Instruction in Statistics Education **(**GAISE**).** They reported that at two year colleges, enrollment in statistics courses grew to 74% the size of calculus enrollments by 2000 (Aliaga, Cobb, Cuff, Garfield, Gould, Lock, & Witmer, 2005). George Cobb notes, as cited by Aliaga,et. al., "Any introductory course should take as its main goal helping students to learn the basic elements of statistical thinking" (Aliaga, et al., 2005, p. 8). The goal for today's students tends to focus less on learning a set of tools and procedures and more on the conceptual understanding of statistical literacy and thinking (Aliaga, et al., 2005). In the guidelines, the goals for students in an introductory course includes the understanding of basic ideas of statistical inference, which involves "the concept of a sampling distribution and how it applies to making a statistical inference based on samples of data (including the idea of standard error)" (Aliaga, et al., 2005, p. 12).

Even though the topic of sampling distributions is so important, it is also one of the most difficult for many students. What causes these difficulties for students? Can we improve our methods of teaching to make this topic easier for students to understand? In this research I explore a student's learning style as one possible factor that might influence understanding of sampling distributions.

A learning style can be defined as a student's preferred style of learning. Specific to my research, I looked at a dimension of learning called, Sensing/Intuition. Briefly, a sensing learner (sensor) prefers real world examples, and to memorize facts and formula and plug in numbers. An intuitive learner (intuitor) deals better with abstract ideas and dislikes memorization and "plug in chug" type calculations (Felder & Solomon, 1999). I have chosen this dimension of learning styles based on the characteristic of concrete versus abstract thinking which is described in greater detail, in later sections.

The purpose of this dissertation is to explore the possibility of a learning style affect upon understanding, and to attempt to find differences between sensors and intuitors. I identified the misconceptions that students have in understanding the concept of sampling distributions and explored the possibility that students with different styles of learning will approach the topic differently. I explored the different styles of learning and the ways in which these students, possessing different preferences for learning, process the same information.

I believe that there is a difference between specific groups of students and that their learning styles affect the way they may understand the statistical concept of sampling distribution and relevant properties. I attempted to determine if there were any differences between groups of students and showed differences, if they exist, in the way in which the student exhibits an understanding or lack of understanding on the subject matter.

Context of the Study

This study took place at North Carolina State University. On average, NCSU enrolls nearly 25,000 undergraduate students with over 90% of them from North Carolina. Over 50% of the undergraduate students enrolled at NCSU are non-science and math majors (NC State University, 2012). By the fall 2009 semester, the student to faculty ratio was listed as 17:1 (NC State University, 2012).

The university offers an introduction to statistics course at the 300 level through the statistics department. There are no prerequisites for this course and this is the only introductory statistics course offered to non-math and science majors.

> "The syllabus covers introductory topics such as, relationships between two variables using graphical techniques, simple linear regression and correlation methods, producing data using experiment design and sampling. Elementary probability and the basic notions of statistical inference including confidence interval estimation and tests of hypothesis. One and two sample t-tests, one-way analysis of variance, and inference for count data and regression" (NC State University, 2012).

Problem Statement

Despite the findings in previous research, the fact remains: There is no definitive solution to correcting student misconceptions. We, as educators lack specific methods for teaching the topic of sampling distributions in order to maximize learning and understanding. My aim is to help facilitate the development of a method of teaching that improves the students' conceptualization of sampling distribution while reducing the recurrent misconceptions by establishing differences between types of students and assessing levels of reasoning.

This research has two parts. The main goal is to determine if there are any observable differences between sensors and intuitors in the way they understand sampling distributions. I plan to assess this in two parts. In order to look at student differences, I used quantitative and qualitative data analysis techniques on data collected from questions presented on a final exam.

The major question addressed in my research is: Do sensors learn (understand) sampling distribution differently than intuitors? If no differences exist, how then do we teach the concept of sampling distributions to harvest a deeper understanding and support the comprehension of the underlying relationships between many interrelated concepts?

Research Question 1.

Are there any differences between sensors and intuitors in their responses to sampling distributions activities?

Hypothesis. I expected that overall; intuitors would be more likely than sensors to understand the concept of the sampling distribution. I believe that sensors and intuitors understand sampling distributions differently, and that intuitors are more likely to comprehend more abstract concepts than the sensor. I believe that sensors are more likely than intuitors to fail to apply a formula or definition in practice, but that the sensor will be better at fact memorization and recalling a formula. My belief is that a sensor be stronger than an intuitor at solving an equation involving “plug and chug” style, but that the intuitor will be stronger than the sensor at solving complex problems and applying the information given in a word problem to solving the answer involving the understanding of the sampling distribution of means.

Research Question 2.

Are there any differences between sensors and intuitors in how they use logic and reasoning in qualitative responses to sampling distribution questions?

Hypothesis. I believe that a sensor will be more likely to provide a good rationalization but fail to apply it to answering a question, meaning, the sensor might be able to give a definition or formula, but will be unable to apply it in context. I believe that an intuitor will be more likely to apply definitions in context, but may be less likely to give definitions verbatim, and might be more likely to give justifications in their own terms.

Evaluating these answers is quite complex and will involve the evaluation of the answer to a question about the sampling distribution using mixed methods; both quantitative and qualitative data analysis. The quantitative data were categorized as either correct or incorrect, and then within each group, I looked at how the student supports their answer using qualitative analysis. I explored the use of three qualitative analysis methods: grounded theory, text mining, and contextual analysis. I found that Contextual Analysis using SAS was a strong alternative to grounded theory and much faster as it involves computer logic, but some of the precision of grounded theory was lost in the automated categorization.

I expected that overall; intuitors would be more likely than sensors to understand the concept of the sampling distribution. Since the majority of students are sensors, I evaluated whether sensors comprehended these difficult and abstract statistical concepts, since they make up the vast majority of the classroom population (Myers-Briggs Foundation. (n.d.)).

My hypothesis that sensors and intuitors understand sampling distributions differently, and intuitors are more likely to comprehend more abstract concepts than sensors is based on the premise that intuitors are stronger at thinking abstractly. I believe that sensors will be more likely than intuitors to memorize a definition but then fail to apply this definition from a practical perspective.

First I explored the research question: Are there any differences in how sensors and intuitors respond to a question regarding the sampling distribution of the mean. Can I find any observable differences between how a sensor versus an intuitor makes a multiple choice answer selection on various questions concerning the sampling distribution of means?

Next I explored the research question: Are there any differences between sensors and intuitors in how they used logic and reasoning in response to questions regarding the sampling distribution of means. Can I find any observable differences between how a sensor and intuitor provide reasoning in giving rationalization for answers regarding the sampling distribution.

I explored the following questions as well: Do students select the correct answer but fail to support with sound reasoning, or do they support an incorrect answer with reasoning that leads us to believe they "understand" the information or definitions, but fail to apply them successfully. My research lead to determining the reasoning of the student and their ability to apply this reasoning to questions about the sampling distribution. When I finished the analysis, I described the differences between sensors and intuitors in how they both understand and apply their reasoning to sampling distributions.

Since the majority of students taking the course are sensors, it was important to determine if sensors understand the difficult and abstract statistical concepts.

Definitions:

There are various terms that will be used throughout this paper, and thus it is important to define each term.

Learning Styles. A learning style can be defined as a person's preferred method of instruction or study that is most effective for them to best learn (Pashler, McDaniel, Rohrer, & Bjork, 2008, p. 106). There are various methods used to determine one's learning styles. In the following chapter, I discuss, in detail, a few of the popular methods and then explain why I chose the method I did for my research. **Conceptual Understanding.** Conceptual understanding can be defined as the understanding of concepts and ideas required for the process of conceptual development. It is also part of being able to determine the connections (i.e. relationships) between different statistical concepts and techniques necessary for success in problem solving (statistically) (Gal & Garfield, 1997, p. 8-9).

Statistical Reasoning. "Statistical Reasoning may be defined as the way people reason with statistical ideas and make sense of statistical information." (Garfield, 2002, p. 1) This involves interpreting information from data and graphics as well as statistical summaries, and making inference or judgment about such data. Garfield claims that the underlying part of statistical reasoning is conceptual understanding of important ideas such as spread, center, shape, randomness, and sampling. Research by Garfield also concludes that students who do well in a statistical course do not often do well when their statistical reasoning is measured.

Statistical Literacy. Statistical literacy can be defined as understanding the basic language and tools of statistics, knowing what specific terms mean, understanding symbols, and recognizing as well as being able to interpret different representations of data (Rumsey, 2002).

What to expect in the next chapters

Chapter 2 provides a literature review of research for statistics education, learning styles, and statistical reasoning that are pertinent to the issue of assessing, understanding and developing reasoning specifically in relation to sampling distributions. Chapter 3 discusses the quantitative methods I used to both collect and analyze the data from an "Introduction to Statistics" course at North Carolina State University in order to answer the first research question posed in this chapter. Chapter 4 gives a detailed quantitative data analysis of all relevant data from the final exam questions. Chapter 5 discusses the qualitative methods I used to analyze open response questions and Chapter 6 gives a detailed data analysis of the responses using three qualitative analysis methods. Finally, Chapter 7 discusses any data limitations I encountered with this study, the conclusions I was able to draw from my data, and implications for further research on a student's understanding of sampling distributions in relation to learning styles.

Chapter 2: Literature Review

Statistics Education

In 1972 Robert Hogg credits Walter Federer with stating, “One note of warning: if the statistical profession doesn’t do something about the teaching of statistics, other groups will” (p. 8). In the same year, Robert Hogg used this quote, along with other advice, to challenge statisticians, especially those in the teaching profession, to submit manuscripts containing innovative programs, which are presently not being used elsewhere. Hogg’s goal was to establish a series in *The American Statistician* called Statistical Education. Hogg points to alternative methods of statistics education, involving the idea that students learn by doing, as the best way to learn statistics. Most of these methods involve the collection and analysis of real data, done by the students, which, in Hogg's opinion, leads to better learning. Hogg comments in 1972, “Certainly, I have never heard of a person learning how to swim by watching the swimming instructor and I believe that a similar statement can be made about statistics” (p 8). Richard Freund states, “(we must develop) a student who can recognize a problem and the solution appropriate to it rather than one who has a technique for which he must find a problem to fit” (Hogg, 1972, p. 10).

In *The American Statistician*, Roger Carlson (1978) wrote, “The reason (few students come away with the kind of understanding that we think they need) for this, is a failure to confront the pedagogical problems of statistics. Until efforts are put into pedagogy, overall results of understanding will remain about the same” (p. 140). Pedagogy is the art or profession of teaching (Merriam Webster Incorporated, 2013).

In the early to mid-1980s several papers were written presenting methods of teaching statistics. They ranged from teaching basic level statistics to more complex inference based methods and theories. All of these papers adopt a common underlying theme; improve the way material is presented so that students have a better overall understanding and grasp of the key statistical ideas and concepts.

Hubert Blalock, Jr. (1987) wrote a paper titled “Some General Goals in Teaching Statistics.” His first point was that students often approach a course in statistics with much fear and anxiety, which often does not disappear with the first class. He suggests that by encouraging the students to think differently about statistics, we can help them overcome, if not avoid, those moments during a test in which their minds go blank. Blalock suggests that this can be done by teaching the students to think a problem through in a persistent manner, rather than having them memorize formulas and techniques. Certainly students need to learn key facts and ideas, but memorization is often what causes the students to go blank in the first place. If they had an alternative solution, for times when they do draw a blank then they would still be able to solve the problem.

Garfield and Ahlgren (1988) published a paper giving implications for further research in statistics education by listing common misconceptions in statistical reasoning, and how to address these issues. In describing the misconceptions involving descriptive statistics such as randomness, sample, distribution and average, the authors state “there is evidence that conceptual difficulties abound for these topics, too,” despite the minimal amount of attention that had been given to these simple yet powerful ideas. (p.53). Pollatsek, Lima and Well (1981) argue that often the students view something simple, such as the mean, as a computational act rather than conceptual, which actually may inhibit the student from learning with a more sufficient understanding involving relationships between and among statistical concepts and ideas. “Students’ knowledge often seems limited to computational formulas, and many simple problems require more general, relational, knowledge of concepts. (It is) justified that thorough understanding of the material can wait until the second course (or later)” (Pollatsek, et. al., 1981, p. 202). Justification that I believe is neither sound, nor appropriate. Since statistical reasoning and inference is the cornerstone for a more complex understanding of statistics, I believe that failure to teach crucial statistical concepts that facilitate deeper understanding and reasoning early may result in more complex confusion. The encompassing idea behind all of this research emerges into the idea of conceptual understanding or knowledge which is defined in the previous chapter.

Conceptual knowledge is just one of the many components in statistics education. Most introductory statistics teachers are aware that the terminology itself can be quite cumbersome; it is common for the student to become quickly lost amongst the morass of terminology (Singleton, 1989). Singleton (1989) suggested that we teach the basic concepts, techniques, and principles of sampling through an extended concrete example. The exercise progressively builds upon students’ understanding; beginning with the concepts of random selection and sample, and then continuing to simple random sampling. Singleton (1989) encourages hands on experience in drawing samples and calculating sample estimates.

Despite building on concepts and using hands on experiences, students still fail to comprehend many statistical concepts and ideas and often misunderstand key components to deeper understanding. These misconceptions filter through to advanced topics and cause a complete disconnect in learning.

In 1999, delMas, Garfield and Chance discussed classroom research in statistics, termed active research, presenting a four step model, with the main goal of gaining insight into problems, their definitions, and their sources. They developed a model for classroom research in order to address their concern that: while new methods are being introduced into the classroom and shared among teachers, no tangible data or evidence are collected to support their claims. (delMas, Garfield, & Chance, 1999) Many teachers have incorporated various simulation activities to try to enable the students' understanding of difficult concepts such as sampling distribution, but found that after presenting these new methods, students still demonstrate a lack of statistical understanding as well as possess "troubling misconceptions about sampling distributions" (delMas, Garfield, & Chance, 1999, p. 2).

The main concern of Chance delMas and Garfield is that "many students who pass a statistics course do not develop the deep understanding needed to integrate these concepts (ideas of sample, population, sampling, and sampling variability) and apply them in their reasoning, ... we see the concept of sampling distribution as crucial to the understanding of statistical inference" (1999, p. 3).

The idea of a sampling distribution is a key component to statistical inference but often comes with much difficulty in understanding and carries with it a great amount of misconception. According to Chance, delMas, & Garfield; "Students tend to look for rules and pattern but rarely understand the underlying relationships causing the patterns they see" (Chance, delMas, & Garfield, 2004, p. 311). For example, students do not necessarily understand why the sampling distribution of means becomes narrower and more normal as the sample size increases. Therefore, when asked to make predictions about plausible distributions of samples for a given size of varying population shapes, the students resort to rules that are often incorrectly memorized, or inconsistently applied, and tend not to think through the generation process for these distributions. As a result, students confuse the difference between the distributions of several sample means with the distribution of one sample of data.

Furthermore, the Chance, delMas, & Garfield also observe that having students use technology along with making predictions under differing conditions through simulation, and then testing the accuracy of their predictions against the simulated results does appear "to improve the impact of the technology on their reasoning" (2004, p. 312). Research shows that "students are more apt to correct their misconceptions and to construct more lasting connections with their existing knowledge framework" (Chance, delMas, & Garfield, 2004, p. 312). Technology alone is not enough to cultivate a deeper understanding of key statistical concepts and abstractions. Their research concludes that, allowing a student's direct interaction with software, as well as the ability to test their formulated ideas may have some effect upon the student's ability to reason and process, but is not likely to cause a significant change in the student's deeper levels of understanding. Good software coupled with concise directions pointing students towards important features does not ensure that the student will understand and while contradictory, is necessary for conceptual change; it has not been proven to be sufficient (delMas, Garfield & Chance, 1999, p. 12).

The problem, in part, of developing a thorough comprehension of sampling distributions, can be attributed to a student's lack of understanding of associated concepts, such as standard deviation, distribution, as well as shape and spread. Evidence also suggests that although students follow rules to correctly predict the performance of sample means, they are still unable to correctly describe the process that creates a sampling distribution, suggesting that a student still cannot fully integrate evidence about sampling distributions, in spite of their ability to predict behavior correctly.

Many authors take note at the exceeding number of statistical misconceptions, referenced in the next section. There are very few studies that tell us how to avoid further encouragement of these misconceptions within our students. One major reason for the lack of literature is that prior to 2000, few people based their research on how students think, omitting a major psychological mechanism in learning. Prior to that time, there had been many contributions to the teaching of statistics, but a very limited amount of data collected through research, to support these contributed theories and ideas.

In spite of this, over the last ten years, there has been an increase in the number of publications geared at the topic of statistical reasoning, both in how to develop statistical reasoning in the student as well as how to assess it. The following sections will discuss recent literature on statistical reasoning and student misconceptions. A section on how students prefer to learn will follow and then I related statistical reasoning with how students prefer to learn, and how I believe the connection between student learning preferences and statistical reasoning will help guide us in better teaching and assessing of statistical understanding in introductory students.

Statistical Reasoning

The underlying theme of statistical reasoning has proven to be a hot topic for statistics education research, with the goal of assessing student understanding about sampling distributions and other underlying concepts. Statistical thinking is applying statistical thoughts to new problems and then recognizing questions of your own (Rumsey, 2002). Garfield and Chance (2000) define statistical reasoning as “the way people reason with statistical ideas and make sense of statistical information,” which encompasses drawing conclusions based on data sets, statistical synopses of data, or representations of data. Richard Nisbett (1993) summarized reasoning and concluded that it can be improved by learning statistical rules. Contradictory research has indicated that merely understanding statistical rules does not necessarily imply statistical reasoning (Garfield, 2002, p. 3). Statistical reasoning cannot be upheld without statistical literacy, or correct statistical thinking. Statistical reasoning implies the understanding of important concepts such as: location and variation, distribution, sampling and randomness, as well as association, and aims at interpreting statistical results and making inferences (Tempelaar, Gijselaers, & van der Loeff, 2006, p. 2).

In 2000, Garfield and Chance summarize the current practices being used to assess students in statistics education and suggest alternative assessment methods (Garfield & Chance, 2000, p. 100). With a main goal of developing statistical literacy and reasoning, among other more basic learning goals, the authors suggest an approach to student evaluation, in contrast to the more traditional methods of assigning grades via quizzes, homework, and other task based assignments such as hands on projects, portfolios of work, concept maps as well as news critiques and minute papers.

"The Statistical Reasoning Assessment (SRA) was developed and validated as part of the ChancePlus Project funded by the NSF," (attributed to the work of both Garfield (1998) and Konold (1989)), (Garfield and Chance, 2000, p.114). The SRA was the only existing tool, at that time, which was used in assessing students' ability to understand statistical concepts and apply statistical reasoning (Garfield & Chance, 2000, p. 114). The SRA is a 20 question multiple choice test, published in 1998 that became widely available after the Garfield 2003 publication in the Statistics Education Research Journal: "Assessing Statistical Reasoning" (Tempelaar, 2004, p. 1).

The SRA lists 8 correct and 8 incorrect reasoning scales. The correct reasoning scales are listed below. The measured misconceptions are listed in the following section, titled, Student Misconceptions. The following is taken from Table 1 in Garfield and Chance (2000, p. 118).

Table 1. Correct Reasoning Scales

1.	"Correctly interprets probabilities."
2.	"Understands how to select an appropriate average."
3.	"Correctly computes probability"
4.	"Understands independence"
5.	"Understands sampling variability"
6.	"Distinguishes between correlation and causation"
7.	"Correctly interprets two-way tables"
8.	"Understands the importance of large samples"

Garfield (2003) suggests that "traditional forms of assessment such as multiple choice are too narrow to provide sufficient information about student learning" and concludes that despite the growing emphasis on the development of statistical reasoning in students, it is still a challenging task to assess statistical reasoning and this area needs more attention in the research literature. She also adds that "there is still ample room for more studies that develop new assessments of statistical reasoning" (p. 22, 31).

Building upon what Garfield (2003) suggested regarding needing more complex methods of assessing statistical reasoning, Lipson (2003) found that since many statistics students are not mathematically trained, the abstract ideas in statistics tend to be proven rather than solved, meaning that students are looking for formulas and equations to answer questions in a statistics course. Lipson (2003) concluded that the understanding of the sampling distribution is related to a development of conceptual knowledge in regards to statistical inference, and confirms that “in order to develop conceptual understanding of the procedures of statistical inference, the empirical representation of the sampling distribution is an important component of the student’s schema for sampling distribution” (p. 282-83).

Students exhibiting learning approaches based on higher efforts (such as greater efforts at test preparation or longer amount of time studying) tend to have lower statistical reasoning scores and a greater number of misconceptions than students that have different approaches at learning (Tempelaar, Gijselaers, & van der Loeff, 2006). They did not elaborate what these different learning approaches might be, but they do hypothesize that: students that exhibit a deeper learning approach are ideal because they tend to focus on underlying meanings of the main principles, ideas and applications. Tempelaar, Gijselaers, & van der Loeff, (2006) hypothesized that students exhibiting a surface learning approach to statistics are less ideal since their focus is mainly on root learning and memorization, with no real attempts at gaining a deeper understanding, and further hypothesized that students with the deeper understanding approach will have a positive effect upon the student’s level of statistical reasoning upon entering the statistics course, in opposition to the suggested negative relationship between surface understanding and statistical reasoning. The authors administered an inventory of learning styles by Vermunt and Vermetten (2004) and conclude that correlations between types of learning related to deeper understanding do exist, but they are quite small, albeit significant. This avenue has not been pursued any further and the Vermunt and Vermetten learning style index is included in my Learning Styles section to follow.

Most teachers instruct on concepts and procedures with the hope that statistical reasoning develops. This approach is often blind to the overall process of providing proper instruction to ensure a solid foundation for statistical literacy, inference and reasoning, and tends to leave the process of developing understanding and reasoning to the student rather than as part of the task of the teacher.

Several suggestions are made for administering an assessment of statistical reasoning: Presenting case studies and real life examples, drawing concept maps, providing critiques of statistical ideas or issues in the media, writing minute papers on what has been learned; presenting enhanced multiple choice items that map answers with appropriate explanations, and presenting the statistical reasoning assessment (Garfield & Chance, 2002). There has been no clear consensus on how to help students develop statistical reasoning, an ongoing goal of researchers and educators alike.

Measuring a student's statistical reasoning is something we are now equipped to ascertain due to the summary by Garfield and Chance on the various assessment methods that can be used to determine a student's reasoning; apart from traditional assignment of grades, a task more directly linked to procedural reasoning. One tool, in particular, is the Statistical Reasoning Assessment Tool (SRA), which entails 20 multiple choice questions that offer reasoning along with the response, with both correct and incorrect ideas (Garfield, 2003). Students taking this assessment are encouraged to select the option which best represents their own line of thinking. Common misconceptions taken from previous research are used in developing this tool, and it has been noted that despite being able to correctly determine the outcome of a problem, students still exhibit a lack of understanding in the processes underlying the solution. In developing the tool, researchers first established that typical evaluations of conceptual understanding are inadequate, since students receiving good grades still uphold misconceptions and misunderstand main ideas (Garfield & Chance, 2000). An example of this would be the misconception that the sampling distribution of means will have the same shape as the parent population. The implementation of simulation software into the classroom addresses the misconceptions by illustrating a visual representation of sampling, but no evidence has been presented to support its overall effectiveness at circumventing this misconception.

Research by Saldanha and Thompson (2002a) suggested that the underlying concepts of sampling distributions are abstract in nature, and a higher level of reasoning is needed to capture all of its properties. They define this abstract reasoning as relating to all parts of a problem and how they interact and support one another (Saldanha and Thompson, 2002a). In sampling distributions, many students possess a more simple kind of reasoning rather than abstract reasoning. A more simple type of reasoning would only involve viewing the parts that make up the whole, in terms of definitions rather than abstractions. A simple way to reason about the sampling process is that the samples drawn are parts that make up the population. Each individual sample is just a part of the population with no relationships between samples made (Saldanha & Thompson, 2002a).

Abstract reasoning about sampling distributions, encompasses the parts as simultaneously making up the whole, such that the means of the repeated samples are simultaneously approximating the mean of the population. Likewise, the samples will simultaneously vary, but will fall within boundaries (Saldanha & Thompson, 2002a).

Thompson concludes that generally, poor performing students have been found to view sampling as a more simple process, a building block from one step to the next, whereas very few students view sampling distributions as an abstract process (Saldanha & Thompson, 2002a). For sampling distributions, abstract reasoning involves repeatedly sampling from a population, recording a statistic, and tracking the accumulation of statistics as they distribute along a range of possibilities. Poor students view sampling in a simple context. Their sense of variability does not extend to the idea of distribution and their ability to coordinate various levels of activity such as repeated sampling is weak (Saldanha & Thompson, 2002a).

The authors believe that, if we can teach students to think abstractly regarding the sampling distribution (of the mean), then we will be able to better teach the subject, thus enabling a deeper understanding in statistical inference (Saldanha & Thompson, 2002b).

A "mature and well-articulated image of distribution supports quantifying the expectation of a particular kind of sampling outcome and thus quantifying one's confidence in a sampling outcomes' representativeness" (Saldanha & Thompson, 2002a, p. 267). This entails contrasting the individual sample result against an aggregate of similar sample results in order to compare the one against many. This image of simultaneity is central to being able to think in abstract terms. Saldanha and Thompson (2002a) conclude that very few students are able to think abstractly and speculate that the complex activities of sampling distribution tend to overshadow the more simplistic ideas of sampling variability that are presented early in their experiment.

Inferential statistics is founded on reasoning about the characteristic of a population on the basis of information obtained in a sample. Effective reasoning requires understanding that, while individual samples in a set of samples may vary tremendously from a population, means of all possible samples of a given size will follow a predictable pattern as described by the Central Limit Theorem (Heid, M., Perkinson, D., Peters, S., & Fratto, C., 2005). "To understand sampling distributions, one needs to build on understandings of distribution (including center, spread and shape), population, samples, and the act of sampling" (Heid, et al., 2005, p. 1). In this understanding students must be able to distinguish among sample, sampling distribution, distributions of means, and sampling distribution of sample means; four concepts that sound very similar but exhibit different behavior and properties (Heid, et al., 2005).

One of the difficulties of developing statistical reasoning concerning sampling distributions comes with what Heid, Perkinson, Peters and Fratto (2005) call making and managing distinctions in regards to sampling distributions. The authors reaffirm that "the concept of sampling distribution is one of the most complex topics encountered in an introductory statistics course" (p. 1). The similarities of these terms primarily exist in the nomenclature. For example, the root words amongst these terms are sample and distribution, listed in nearly every word, and the verbiage itself confuses the definitions to the student who is often hearing them for the first time in a formal setting. To further confound the issue, the definition of a sampling distribution, a distribution of means and sampling distribution of means contains the other terms.

The authors also noted the need to account for the fact that while some properties hold true under any conditions, some need to be restricted, for example, to a subset of distributions of sample means (not the set of all possible samples) (Heid, et al., 2005).

The distinctions referred to by the authors include distinctions in language, in sampling, in probability, and in measures of spread and center. The distinctions in language are “critical because similar language is used to describe different essential concepts: sample, sample distribution, distribution of sample means, and sampling distribution of sample means. Understanding of sampling distribution involves a nesting of concepts and navigating this web of concepts requires making distinctions about the level of nesting about which one is thinking” (Heid, et al., 2005, p. 2). They continue to explain that the understanding of the measures of a population’s center and spread is needed to reason about possible values of the center and spread for samples drawn from the population. “Understanding about possible values for the measures of center and spread for a sample leads to the need to understand the” abstract nature of sample statistics, and understanding this abstract nature is central to understanding about the set of means of all possible samples (of a given size) (p.2). Therefore, to be able to reason about sampling distribution, students must have a sense of how likely a sample mean is rather than whether or not the occurrence of the specific mean is possible (Heid, et al., 2005).

Further, the students also need to understand that there are properties of the sampling distribution that are absolutely true, such as the mean of the population is equal to the mean of the sampling distribution. Applying this to sample size, students must also be able to distinguish the effect sample size has on the sampling distribution characteristics as well as the characteristics of the individual sample (Heid, et al., 2005).

In conclusion of the study, the authors find that it is imperative that precise language is used to distinguish between statistical terms with precision as well as make distinctions between the closely related concepts within the complexity of sampling distributions. They state that the everyday meanings of statistical terms are less precise than the statistical meaning, which compounds the issues students have with the difficulty in reasoning with statistical terms. “The use of informal language impacts one’s ability to make distinctions that would be available through the use of precise mathematical language,” (Heid, et al., 2005, p. 5). Being able to develop statistical reasoning is further complicated by the difficulty in making distinctions among interrelated terms and ideas.

Makar and Confrey (2005) reiterate that the movement of statistics education has been driven away from rote calculations and operations and towards the development of students’ thinking about what makes sense. The authors discuss and evaluate the use of standard and nonstandard statistical language when discussing the notion of variation and explain that while standard language is ideal, it is important to recognize that the concepts being described are far from simplistic, and that students “often articulate meaning in statistics through the use of less than conventional terminology” (Makar & Confrey, 2005, p. 50). By attending to nonstandard statistical language, the authors believe that you can gain insight into student thinking and communicate, to students, that statistics can be understood even without proper statistical talk, but do not imply that you should teach nonstandard statistical language. By acknowledging only the conventional language, opportunity to gain insight into students reasoning and thought process could be lost. It can also be noted that when students use their own language, they are making sense of statistics in their own terms (Makar & Confrey, 2005). Using standard language to explain or define a statistical process may involve more memorization and less actual understanding of the concepts described.

Reading and Reid (2006) conducted a study on reasoning about distributions from the perspective of variation. In the comprehensive sense, they view variation as the explanation or amount of the observable characteristic variability. They expect the students to include in their basic descriptions about distributions, concepts of shape, center and spread as well as outliers, or deviations from the pattern (Reading & Reid, 2006). The goal of their study was "to develop and describe a hierarchy of reasoning about distribution" using SOLO Taxonomy as a suitable framework. (Reading & Reid, 2006, p. 47). In general, SOLO (Structure of the Observed Learning Outcome) Taxonomy is a model consisting "of five modes of functioning, with levels of achievement identifiable within each of these modes" (Reading & Reid, 2006, p. 48). John B. Biggs and Kevin F. Collis (1982) are attributed with developing this model in order to evaluate learning. The SOLO taxonomy suggests: Prestructural, Unistructural, Multistructural, and Relational Understanding. In relation to the elements of distribution and inference, I give a summary of their table:

Table 2. SOLO Taxonomy

SOLO LEVEL	Explanation of Reasoning
Prestructural	No reference to key elements of distribution. Recognizes the concept but does not use it to make inferential statements.
Unistructural	Focuses on one key element (center, shape, spread) Indicate a correct understanding of the concept of distribution with one inferential statement.
Multistructural	More than one key element of distribution is focused upon. Shows correct understanding of the distribution concept using multiple inferential statements.
Relational	Develops relational links between various key elements of distribution Links concepts together to exhibit a strong understanding of the concept of distribution.

Students are divided into four categories regarding variation: no consideration, weak consideration, developing consideration and strong consideration (of variation). When asking students about the sampling distribution and relating to variation, the students with weaker association compared the distribution of sample means to the parent population and had expectations for it to be similar in shape and variation. Other students expected the sampling distribution to have greater variation than the population. Students with some developing consideration of variation considered the variation in relation to shape or sample size (Reading & Reid, 2006).

The study concludes that most students with weak consideration of variation, had responses that focused on some measure of the range of data and the location (mean), while others incorporated terminology to suggest they considered more than just the center and spread, using both standard (outliers, skewed) and non-standard (clumped, compact) language. “Responses attempting to link some key elements and/or use them for comparative purposes did so incorrectly," and they attribute these types of response to more than half of the students (Reading & Reid, 2006, p. 56).

A third of the students exhibited a developing consideration of variation, and often made a link between center and spread, as well as explained variation within and between distributions. Few of the students were able to apply their understandings to the complex idea of sampling distribution of the mean (Reading & Reid, 2006). Students with a "developed consideration of variation were able to draw these key elements together to reason better about distribution, through the language they used and the links they made." Responses indicated that a student's ability to understand and articulate variation may be indicative of the ability to reason about distributions, and that their understanding was refined about center, density, spread, skewness and outliers suggesting that how students reason about variation can be partially solved through the development of their consideration of variation, and includes the use of non-standard language. Finally, the authors conclude that students who do not have the correct concepts or language necessary to describe variation are consequently unable to reason about distribution in context. In conclusion, "tasks designed to assess outcomes in core themes can also be used to identify indicators of reasoning about distribution." (Reading & Reid, 2006, p. 63).

Garfield and Ben-Zvi (2007), regarding teaching students about the ideas of distribution, center, and variability, state that they should first be introduced informally in intuitive ways such as data exploration and simple experiment design, and then introduced to the ideas of sampling distribution in a more formal context, in relation to the Central Limit Theorem. They do not mention informal language, just an informal approach to the ideas, such that rather than first introducing rules and definitions, they introduce the intuitive ideas and visual demonstrations. The authors add that the concepts of center, shape and spread should be combined and introduced together rather than be presented as separate isolated topics in order to guide students to a better understanding.

More recent studies have heeded this call for the development of statistical thinking, and examine "the existence and development of reasoning about important statistical ideas" through both formal and informal assessment methods. Most of these studies focusing on the assessment of students' statistical reasoning reveal that despite formal instruction, students still fail to exhibit correct reasoning about important statistical ideas and concepts (Zieffler, Garfield, Alt, Dupuis, Holleque, & Chang, 2008, p. 5). Both quantitative and qualitative approaches have been used to assess statistical reasoning. Researchers have found puzzling results of a low to zero correlation between course performance and an aggregation of reasoning skills, and that "students with strong effort-based learning approaches tended to have lower correct reasoning scores, and higher misconception scores, than students with different learning approaches" (Zieffler, Garfield, Alt, et al., 2008, p. 5).

Qualitative methods to assess statistical understanding such as metaphor, SOLO taxonomy, and an APOS theory (Action, Process, Object, Schema) are introduced. Groth and Bergner (2005) examined the use of metaphors to assess students understanding of statistical ideas by asking them "to write down their own metaphor for the concept of a sample" and discuss its strengths and weaknesses. They then used SOLO Taxonomy to categorize the responses, as cited by Zieffler, Garfield, Alt, et al. (2008), "building on the framework developed by Watson and Moritz (2000), Groth and Bergner (2005) concluded that despite the successful completion of the introductory statistics class, many students possess a conception of sample seemingly limited to a collection of data or group of objects." (p. 6).

Zieffler, Garfield, Alt, Dupuis, Holleque, and Chang, (2008) suggest that it might be "more useful to have several different assessments, each evaluating reasoning about a particular statistical concept (such as sampling). The use of qualitative methods should be further explored and may offer ways to develop quantitative instruments based on questions used in the qualitative studies" (p. 7). The authors also suggest examining "students' response patterns across assessment items to help identify inconsistencies in their statistical reasoning" and further, that final exam scores are not useful metrics for evaluating statistical reasoning or knowledge (Zieffler, Garfield, Alt, et al., 2008, p. 15).

Kaplan, Fisher and Rogness (2009) suggest that the lexical ambiguity in statistics relates to the difficulty of developing statistical reasoning. They found that the words average and spread have the most ambiguity to them since students typically associate average with middle, and sometimes spread can refer to an even coat, such as buttering toast, and that instructors should address these problematic associations early on in the class. The authors suggest letting the student first use the word spread (a synonym for variability), rather than introducing it. They also think that the word random, which means unlikely in everyday language, such as referring to bumping into an old friend as being "so random" (p. 15). Since something random in statistics is quite structured where we can know the likelihood of the outcome, it does not take on the same meaning as the definition of the word taken from common language.

In 2012, Kaplan, Rogness and Fisher (2012) argue against the use of the word spread, and for the correction of increasing the use of the word variability to replace this term, summarizing that spread can take on other meanings such as dispersion, students have a prior knowledge of the word that are often related to misconceptions, and they found that students did not "demonstrate strong statistical meanings of the word spread at the end of a one-semester statistics course" (p. 57). Lexical ambiguity plays a huge role in the difficulties encompassing the understanding of sampling distributions. Common words like spread, mean, average, median, sample, and distribution are confounded with statistical meanings, causing students greater difficulty in comprehension. Students need to be able to eliminate prior understanding of certain words from everyday language and learn the definitions in the statistical sense, which is not as easy to do as described. I believe that students face confusion in an introduction statistics class early on, due to this ambiguity, which creates a greater hurdle for them to overcome in order to keep up in class, and build their knowledge and understanding on the interrelated concepts. This is a large stumbling block for statistics educators.

Mohd Nor and Idris (2010) used the SOLO Taxonomy to assess levels of inferential reasoning when students compare box plots.

Because I plan to administer questions that deal with interrelated concepts and ideas that link to sampling distribution, the Solo taxonomy will prove useful in providing analysis of research results.

Student Misconceptions

In statistics education, there is a big disconnect between the introductory concepts and more abstract ideas, such as sampling distributions and variability. The same major misconceptions surface every semester, despite efforts to correct and/or teach around them. There has been a "substantial increase in the concern for professional research on and guidance of the statistical education process" (Callaert, 1998, p. 1209). Daily exposure of statistical information leads to further illiteracy and commonplace misconceptions that overwhelm the general public and either lead astray future statistics students, or overwrite/compete with prior instruction lent to students in the past. Data are often displayed in graphs in various types of media, but people often do not pay attention to the labels on the axes of often hastily produced graphics that are not well thought out. Careful thought and consideration of data presentation are necessary in order to be clear and honest as well as accurate (Callaert, 1998).

It has been suggested that students enter the statistics classroom with preconceived theories and intuitions that are often contradictory to convention (Konold, 1995). Prior theories are often difficult to adjust, and students can possess opposing beliefs about a particular situation. Students enter the classroom with prior knowledge, either learned, or an unlearned, intuitive knowledge. Sometimes this knowledge corresponds to correct reasoning skills and can be recognized as a true prior knowledge, but more often, the unlearned knowledge is incorrect and belongs to a body of statistical misconceptions (Tempelaar, 2004). Even students who can make correct probability computations have a tendency to fall back on their faulty reasoning when asked to make judgments or inferences about the uncertain event outside of exam based tasks. Tempelaar refers to this as the “last learned, first forgotten principle” and justifies the need for special attention to these misconceptions be brought out in class (Tempelaar, 2004, p. 4).

The SRA, as discussed previously, provides eight scales of student misconceptions, as taken from Garfield and Chance (2000, p. 118).

Table 3. Eight Scales of Student Misconceptions

1.	"Misconceptions involving averages (common number, median vs. mean, outliers)"
2.	"Outcome Orientation (single events versus a series of events)"
3.	"Good Samples have to represent a high percentage of the population (in order to be good)"
4.	"Law of Small Numbers (small samples best represent the population so are preferred)"
5.	"Representativeness misconception (likelihood of a sample is estimated on how closely it resembles the population)"
6.	"Correlation implies causation"
7.	"Equiprobability Bias (Events of unequal chance are viewed as equally likely)"
8.	"Groups can only be compared if they are the same size"

Most specifically, common misconceptions in understanding sampling distributions taken from Chance, delMas and Garfield (2004, p. 302), are as follows:

Table 4. Common Sampling Distributions Misconceptions

• "Believe sampling distribution should look like the population (for sample size $n > 1$)."
• "Think sampling distribution should look more like the population as the sample size increases (generalizes expectations for a single sample of observed values to a sampling distribution)."
• "Predict that sampling distributions for small and large sample sizes have the same variability."
• "Believe sampling distributions for large samples have more variability."
• "Do not understand that a sampling distribution is a distribution of sample statistics."
• "Confuse one sample (real data) with all possible samples (in distribution) or potential samples."
• "Pay attention to the wrong things, for example, heights of histogram bars."
• "Think the mean of a positive skewed distribution will be greater than the mean of the sampling distribution for samples taken from this population."

These misconceptions remain each semester, despite efforts by educators to adjust teaching methods to correct these maladies, but without much improvement. Research has shown that when students are able to observe their misconceptions and correct them appropriately, they are more likely to construct lasting connections with their existing knowledge framework (Chance, delMas, & Garfield, 2004).

Learning Styles

In order to dig deeper into how a student is understanding statistics, it is important to assess how a student thinks and reasons. Perhaps by delving into the minds of the students as well as the process underlying the way they think, learn, and evaluate their understanding, I can better ascertain student limitations. One way to delve into the mind of the student is by assessing the students preferred style of learning, which will help make sense of how a student prefers to learn.

Learning styles have been defined in various ways. In *Experiential Learning,* David A. Kolb stated, “Learning is the process whereby knowledge is created through the transformation of experience" (Kolb, 1984, p. 38).

Richard Felder defines learning styles as "characteristic strengths and preferences in the ways they (students) take in and process information" (Felder, 1996, p. 18). In an earlier article, Felder and Silverman state, "a learning style model classifies students according to where they fit on a number of scales pertaining to the ways they receive and process information" (Felder & Silverman, 1988, p. 674).

Learning Styles are tendencies and preferences that students have for taking in processing information in a particular way, as well as responses to various instructional environments. Preferences are not reliable indicators of strength and weaknesses however, stronger preferences are more likely to have strengths associated within the dimension. Learning styles have become an effective tool not only in assessing a student’s preferred way to learn, but also in establishing methods to meet these unique learning needs, and to improve our current teaching methods and tools.

Significant evidence has been found to support the use of learning styles in education. For example, Felder (1993) notes that mismatches between a student and instructor (between the learning preference and teaching style) can lead to lower grades, less interest in the subject, and in extreme mismatch cases, could even push the student to change their major. This problem could be minimized if the teachers' styles are modified in order to accommodate all learning styles at least some of the time. Felder and Henriques (1995) conclude that the quality of student learning and attitudes are unfortunate effects of mismatches between student and teaching learning styles in a language class. (p. 21).

Felder, Felder and Dietz (1998) conducted a longitudinal study in which they implemented an experimental track into the Chemistry curriculum for chemistry majors. One track of CHE majors were taught in the traditional format, and the other track of students were taught in an experimental manner. The experimental track used methods to cater to all learning style preferences. The authors concluded that while they cannot predict that the experimental method (adapting to teaching using all the LS methods), will yield the same results, they can state with confidence that making changes will ultimately bring about improved performance and the only limitation of the experiment is the potential of the instructor and students' ability to learn (Felder, Felder, & Dietz, 1998, p. 479). Graduation rates (aka student retention) were also greater for the experimental track (Felder, Felder, & Dietz, 1998, p. 469).

Learning styles have been used in several areas of education research to try to establish a link between the ways students understand material and the way teachers are presenting the material. Richard Felder and Linda Silverman state, "How much a student learns in a class is governed in part by that student's native ability and prior preparation, but also by the compatibility of his or her learning style and the instructor's teaching style" (Felder & Silverman, 1988, p. 674). According to Felder and Brent (2005), several authors conclude that greater learning occurs when learning styles are met rather than mismatched, but does not fail to mention that the "point of identifying learning styles is not to label individual students and tailor instruction to fit their preferences" (p. 62). When a learning style is mismatched, the student may become uncomfortable to effectively learn, whereas a complete match might fail to challenge a student to develop critical learning skills in a less than preferred style of learning. Felder and Brent (2005) conclude that the optimal teaching style has balance so that students are both comfortable and challenged at any given time, encouraging growth and development as a student and believe that the most important aspect learning styles is the application by instructors in designing a balanced teaching approach (p. 62).

Several instruments exist to evaluate learning styles. Perhaps one of the earliest instruments that could point to a learning style was the Myers Briggs Type Indicator (MBTI), created in 1942. Although originally developed as a personality evaluation, some of the dimensions of the MBTI allude to the learning preferences of an individual. As interest in learning styles increased, other instruments became available to determine the preferred learning styles such as, Kolb's Experiential Learning Style, Fleming's V.A.R.K., Felder's Index of Learning Styles (ILS), and many others.

Learning Style Instruments

Many Learning Style Instruments have been developed to try to ascertain a student's style of learning and categorize the types of styles they are measuring. There are many different types of learning styles instruments readily available to use in the classroom. Some instruments have been validated and some have not, but all have a unique way of categorizing the way in which a student learns. I explain some of the styles that I chose to investigate before choosing one to use in our research.

Felder and Spurlin (2005) suggest looking only at moderate to strong preferences based on implications from studies conducted (Ryerson University, San Jose State University, Arizona State University and Brazilian Universities) because mild preferences may exhibit a shift back and forth and thus when carrying out research on learning styles differences in behaviors, you want to observe the students who are less prone to waver. (p. 105)

MBTI. Learning Styles are used to evaluate a person's preferred way of learning. The Myers Briggs Type Indicator (MBTI) developed by Katherine Briggs and her daughter Isabel Briggs Myers during World War II, inspired by theories and ideals in Carl Jung's work, Psychological Types.

The MBTI consists of 93 questions. The subject completing the MBTI has a choice of two possible answers for each question, or he or she may opt to leave the question unanswered. Based on the answers to these questions, the subject is put into one of two groups within each of four dimensions. These dimensions can be referred to as "attitudes", "perception", "decision making" and "specific attitudes of the functions." The "attitudes" define how a person orients and receives their energy, and can be categorized as either Introvert or Extrovert (Myers-Briggs Foundation, n.d.).

The "perceptions" define how a person perceives information and the category is either Sensing or Intuition. "Decision making" defines how a person makes their decisions, and can be categorized as either Thinking or Feeling. The "specific attitudes of the functions" defines how a person acts and orients with the outside world, and the category is either Judging or Perceiving. For example, a person classified as sensing has a preference for practical realism, whereas a student with a preference for intuition provides a view of all possibilities. Essentially, the sensing person likes real world applications and does not tend to contemplate all possibilities, but rather, just a small subset, whereas the intuitor will be able to think in a more abstract manner, thus enabling them to consider many possibilities or outcomes.

A large number of studies have collected data on the MBTI. A summary of these categorizations is given in Table 5. This table gives an estimated relative frequency of each MBTI personality type in the United States Population and gives information within each dimension. The largest difference in population percentage estimates is found between the sensor and the intuitor.

Table 5. MBTI statistics of Population Averages

Dimension	Category	Estimated Range	Estimated Average
Attitudes	Extrovert	45-53%	49.3%
	Introvert	47-55%	50.7%
Perception	Sensor	66-74%	73.3%
	Intuitor	26-34%	26.7%
Decision Making	Thinking	40-50%	40.2%
	Feeling	50-60%	59.8%
Attitude Function	Judgment	54-60%	54.1%
	Perceiving	40-46%	45.9%

Note. Adapted from Myers-Briggs Foundation. (n.d.).

The MBTI suffers under much scrutiny from several authors. It has been said to fail to be consistent by several researchers. For example Dr. David Pittenger states that the MBTI cannot be statistically supported. According to Pittenger, many are not convinced that the test can produce any significant conclusions. Any psychological test must answer basic issues concerning foundation, namely statistical structure, reliability and validity (Pittenger, 1993a, p. 48). First addressed is statistical structure.

> "The standard error of measure for each of the four dimensions is fairly large. Unfortunately, the MBTI method of scoring obscures this important distinction. It classifies people into a rigid dichotomy. Thus, two people could have raw scores that are close to one another but that define different classifications. This occurs because there are cutoff points that divide the dimensions. When the score is above the cutoff, one classification is given; if the score is below the cutoff, the opposite classification is given." (Pittenger, 1993a, p. 50).

Since the MBTI uses an "absolute classification scheme for people, it is possible for people with relatively similar scores to (become) labeled with much different personalities" (Pittenger, 1993a, p. 50).

The second statistical issue that must be addressed is reliability. Reliability refers to the consistency of the test. The test is unreliable if it is administered to the same group of people several times and different results are achieved each time. If the test is in fact highly reliable then we would expect the same result with each measure of the same subject. An unreliable test presents the question of whether the change in score is due to a change in the subject or to an inconsistent test process. A primary statistical method for evaluating the reliability of an instrument is to administer the instrument at two distinct times to the same person, to evaluate or what is known as the "test-retest" reliability.

Since the MBTI personality type has been said to be a "constant characteristic," we expect the instrument to give similar results for an individual. Short test-retest intervals such as five weeks have shown that as many as 50% are classified into a different type, as shown by several studies (Pittenger, 1993a, p. 49). Pittenger states that "it is hard to reconcile a test that allows individuals to make radical shifts in their type" (Pittenger, 1993b, p. 472).

Changes such as these leave us, the researchers, with questions: How is it possible for a change in personality to occur over a short period of time? Does it maintain, or does it return? What if it always changes and never rests?

Thirdly, we must look at the validity of the MBTI. By using factor analysis, Pittenger (1993a) assessed whether or not the four dimensions of the MBTI really do exist. A factor analysis is used to analyze correlations among the questions in a test. Results should show that there are four “factors of questions.” Each question within a factor should correlate with the other questions in the factor. And each question within the factor will likely be related to the MBTI dimension that it measures. A researcher expects each factor to be independent of the other factors, and would like to see most of the differences among an individual to be accounted for by the factor (Pittenger, 1993a, p. 50).

In a study by Sipps, Alexander, and Freidt, (1985) six different factors were found in the MBTI along with a high measurement error (p. 791). Specifically, 83 percent of the differences among the students could not be accounted for by the MBTI questionnaire, which means that the factors in the statistical analysis of the MBTI are inconsistent with the MBTI theory (Pittenger, 1993a, p. 50). In summary, the MBTI cannot be supported with statistical analysis.

KOLB Learning Style Index. A second instrument of evaluating learning styles is the Kolb learning style model, (Kolb, 1984) called the Learning style inventory (LSI). This instrument establishes combinations of perceiving and processing information that are used to classify a learner into one of four types based upon first how they take information in, concrete versus reflective; and second, how they manage this information, active versus reflective. The four types are titled, type I (concrete, reflective), type II (abstract, reflective), type III (abstract, active) and type IV (concrete, active). A type I would be a reflective learner, a type II would be a theoretical learner, a type III would be a pragmatic learner, and a type IV be an active learner (Kolb & Kolb, 2005).

A reflective learner will learn best by thinking about what information they just received and performing a concrete activity such as taking part in a discussion or writing down notes. A type II, theoretical learner, will best benefit by listening to a lecture and thinking quietly about it. A type III, pragmatic learner, will best benefit by applying their newly learned skills to an activity. A type IV, active learner, will learn best from hands on activities. Often active learners solve problems intuitively and adapt to specific immediate circumstances (Kolb & Kolb, 2005).

Kolb's LSI has been validated for internal validity and reliability. "The comparison of the between item and within item correlation suggest support for internal construct validity" (Kayes, 2005, p. 256). Kayes (2005) provides evidence to support the Internal Validity and Reliability of the LSI Version 3 in 1999, published in 2005, and concludes that the LSI-3 does have internal reliability and consistency but calls for further research (p. 255).

Metallidou and Platsidou (2008) note that the "LSI_1985 is one of the most widely distributed instruments used in higher education" (p.115). They confirm the internal consistency reliability but found the construct validity to be problematic and suggest the instrument to be improved for this. (Metallidou and Platsidou, 2008, p. 117)

VARK. A third instrument that is used to evaluate learning styles is called VARK (Fleming, 1995). It is a questionnaire that measures students' preferences based upon the way a student wants to take in or give out information. Each letter in VARK represents a preference towards how a student desires to take in information. The V represents Visual, meaning that the learner prefers to take in or present information through visual aids such as graphs, charts, diagrams, and symbols as ways of representing the information. The A represents Aural, or hearing. A student that has an aural preference learns best from lectures, tutorials and discussions with others, as they prefer "heard" information. The R represents a

Read/Write preference. A student with this learning style has a preference for information that can be read. They prefer to read and write as a method for taking in information. The K represents a kinesthetic preference, which is a preference to those students who most want to learn from the use of all of their senses: sight, sound, touch, smell, and taste. This preference is multi-modal, and is related to the use of experience and practice, through simulation or real life experiences. This learning style evaluation is performed by presenting the learner with 16 questions, and from the responses, indicates which area (V, A, R, or K) the student has a stronger preference in the giving and taking in of information. After completion of the 16 questions, the student is given a number for each of the four dimensions, with the total in each of the four dimensions summing to 16. Whichever dimension contains the highest number indicates the students learning preference. The VARK questionnaire is free to use and available to take on the web (Fleming, n.d.a.).

The VARK learning styles questionnaire has not been validated statistically, and has only been validated through the evaluation of the person taking the test. After the questionnaire has been completed, the diagnosis is presented, and the student is then asked if they feel like this category assigned to them: Is a match, does not match, or do not know. The questionnaire has yet to be statistically validated, and it has been said that due to its structure, it is unlikely that it will ever be validated. Presently the website's Frequently Asked Questions Section answers the question: "When will there be statistical data on VARK's reliability and validity." The response is:

> "The VARK questionnaire is difficult to use with current statistical methods of validation because of its structure. VARK replicates how decisions are made using

many preferences and the multiple answers for each question make statistical analysis very difficult. There is some work being done now (April 2007)" (Fleming, n.d.b).

Index of Learning Styles. The final learning styles instrument that I address in this paper, is the Index of Learning Styles (ILS), developed by Richard Felder in 1991 (Felder & Brent, 2005). This instrument uses a 44 question survey to evaluate a student on each of four learning dimensions: sensing or intuitive, visual or verbal, active or reflective, and global or sequential.

Felder and Soloman (1999) define the categories within each dimension as follows. Sensing learners prefer to solve problems using well established methods, like learning facts, and dislike surprises and complications. They tend to be good at fact memorization and hands on activities and are patient with details. Intuitive learners like innovation, prefer to discover relationships and possibilities, and dislike repetition as well as "plug and chug" courses which involve routine calculations and a great deal of memorization. They tend to be faster workers than their careful sensor counterpart, and are more innovative than sensors that are more likely to resent being tested on material not clearly covered in class. Intuitors are often more comfortable with abstraction and mathematical formulations as well as grasping new concepts. Sensors dislike courses that have no obvious real world connection as they tend to be more practical (Felder & Solomon, 1999).

Visual learners learn best through what they see such as graphs, diagrams, charts, demonstrations, whereas Verbal learners learn best through words, such as oral and written explanations. Active learners prefer working in groups, best understand information by doing something active with it whereas Reflective learners prefer to work alone, think quietly, and internally process new information. Sequential learners prefer to work in steps, and often build understanding from one step to the next, following logical steps in finding a solution. Global learners prefer to learn in large steps, often solving complex problems understanding the big picture first (Felder & Solomon, 1999).

Rather than evaluating a student and telling them their Learning Style is absolute, a weakness of the MBTI, Felder states that the learning style is merely the students preferred way of learning, not an either/or type of categorization. Some students are more strongly weighted in one direction or the other, whereas others may only indicate a slight preference towards one category over the other (Felder & Brent, 2005)

This instrument has been widely evaluated for validity and consistency. Several analyses of the ILS have been published indicating that the ILS is adequate and sufficient. “Reliability and validity data justified a claim that the ILS is a suitable instrument for assessing learning styles” and is “best used to allow individuals to compare the strengths of their relative learning preferences” (Felder & Spurlin, 2005, p. 110).

Dr. Felder recently informed me that “The MBTI was my introduction to the whole area of personality types, learning styles through the book "People Types and Tiger Stripes", and my wife, it's a long story.” He also stated that “It includes N/S, which is taken directly from the MBTI, and A/R, which comes directly from Kolb's model but is at least a first cousin to E/I on the MBTI.” (Felder, personal communication, May 2008).

Felder found that sensors and intuitors are very different in the way they prefer to learn and tend to pursue fields of study that cater to these preferences. For example, intuitors are more apt to do well in the area of math and science. According to Felder, “most science courses focus on abstract concepts, theories and formulas putting sensors at a distinct disadvantage" (Felder, 1993). Math and science classes are often taught by intuitive learners, and thus are taught to the level of understanding of the intuitor. This often discourages a sensor from pursuing a degree in the particular field of study, not just because of the difficulty of the course, but due to the lack of ability to reach and relate to their level of understanding. Sensors often have to translate symbols (words and algebraic variables) "into concrete mental images in order to understand them" (Felder, 1993). The process can be lengthy often leaving sensors without enough time on tests, resulting in sensors scoring lower grades than intuitors in lecture courses, which, in effect, weeds them out of the field, "even though they are as likely as intuitors to succeed in scientific careers" (Felder, 1993). By teaching a subject that is inherently an intuitive idea; in a way that is appealing to sensors will bridge the gap, and actually enable the students, to reach the second tier.

Summary of Learning Styles. There is a lot of information and research available to tell us about distinctive learning styles. Understanding that differences between individuals in how they prefer to learn do exist and learning what these differences are, can help us to teach these individuals in unique ways that will reach all the learning styles rather than just a few. Learning style differences emerge in thought patterns, overall classroom feelings, and ultimately how a student best learns material presented to them. If using a learning style instrument, it is best to use one that has been statistically validated and has been proved reliable.

Sampling Distributions

The integration of two seemingly contrasting ideals underlies the understanding of sampling distributions and inference. The contrasting ideas suggested are: sampling representativeness, the idea that the sample will have the similar properties to the parent population; and sampling variability, the idea that different samples will differ. Research shows that students exhibit one line of thought or the other, and find it extremely difficult to inter relate the two concepts to form a coherent understanding (Rubin, Bruce, & Tenney 1991, p. 314).

In our study we aim to compare the way two different groups of students understand statistics. Specifically we want to compare the understanding of the sampling distribution between students classified as strong sensors with those classified as strong intuitors, according to Felder's ILS.

The concept of sampling distributions is a relatively abstract concept. Of the four dimensions of the Felder instrument, I expect the differences between sensors and intuitors to be the most pronounced because intuitors are often expected to readily take in an abstract idea whereas sensors often need a more concrete description to understand the same concept. The differences between a sensor and intuitor is defined, but not very well understood, especially in practice.

According to various tests, problem solving with well-known methods, fact memorization, and making real world connections are traits known to be exhibited by sensors, a category within a dimension of the Index of Learning Styles. A sensor would wish to solve a problem using well known methods, which would fall into the same line of thinking as what delMas and Chance and Garfield described as learning rules and patterns (Chance, delMas, & Garfield, 2004, p. 311).

Key Connections to Project

The concept of sampling distributions is crucial to understanding more complex concepts and ideas in statistics and "elements of statistical reasoning have become requisite for a wide range of fields of study" (Garfield & Ahlgren, 1988). Unfortunately, it is a difficult topic for students to learn and is a concept in which students often become frustrated and even turned off from statistics. In, *Reasoning About Sampling Distributions,* Chance, delMas & Garfield (2004) list four concepts that are fundamental for understanding sampling distributions. The following are what the authors consider as "prerequisite knowledge to learning about sampling distributions" (p. 300):

- The idea of variability
- The idea of a distribution
- The normal distribution
- The idea of sampling

Understanding sampling distributions requires a student to incorporate and apply concepts from various parts of the statistics course as well as reason about theoretical behavior of many samples.

> This is a "distinct, intangible thought process for most students. The Central Limit Theorem provides a theoretical model of the behavior of sampling distributions, but students often have difficulty mapping this model to applied contexts. As a result, students fail to develop a deep understanding of the concept of sampling distributions

and therefore often develop only a mechanical knowledge of statistical inference" (Chance, delMas, & Garfield, 2004, p. 311).

Learning Style Connection. Understanding the sampling distribution and its properties seems to be a major theme in the difficulties encountered in statistical education. Students often struggle with the concept of sample and sample means, as well as the distribution of a sample or a distribution of sample means. It is important to find a way to better educate students in this area. Since the concept of the sampling distribution is more abstract in nature as it involves possibilities and relationships, it is not easily grasped by all students, especially if the student prefers to learn with material based on real world examples and concrete concepts. I believe that there is a major difference between a sensor and intuitor in understanding these key statistical ideas. If I can discover differences between these two categories, it could play a key role in helping students to better understand this abstract concept.

Statistical Reasoning and Learning Styles. In relating learning styles to the types of reasoning mentioned in the previous sections, I made the following observations. A sensor is more likely to think simply and in terms of definitions and formulas because they are fact oriented and prefer to memorize the process in regards to the sampling distribution. For example, in regards to the shape of a sampling distribution, a sensor might view each random sample as equally likely to occur and thus might believe that each sample mean is equally likely. This leads to the belief that the likelihood of each mean is not taken into consideration, implying a student might not conceive that multiple samples give the same mean, they may reason in simple terms rather than abstractly. They might observe each sample mean, but might not view the likelihood of the sample mean to estimate the population mean, and therefore may likely argue that it is necessary to see the actual data in order to predict shape or distribution of the sampling distribution of means. This similar line of thinking for a sensor could also be used to support the location of the center for the distribution of means. The sensor might be more likely to state that they will need all the data in order to predict the center of the sampling distribution of means for a given population rather than think in abstract terms and might reason that they could make an estimate of the mean of the sampling distribution based on the population mean.

An intuitor might be more likely than a sensor to think abstractly as they prefer to think in terms of possibilities and relationships. Abstract reasoning entails the process of thinking of concepts and ideas at the same time and how they relate to one another. For example, the idea that a random sample will exhibit similar shape as the population from which it is drawn, and the mean of that sample will be close to the mean of the population. A person that can reason abstractly should be able to think about multiple random samples simultaneously as well as each sample having a distinct mean that will fall within a boundary around the population mean.

In regards to the shape of the distribution, abstract reasoning will lead to the realization that the majority of sample means will fall in the "center" or around the mean of the population since the probability of the occurrence is greater and the tail values will have less likelihood of occurring and thus not as many means will occur at those points. This example of abstract reasoning differs from simple reasoning in that simple reasoning consists of applying a set of rules to a problem in determining the outcome. It is contrived of repetition and fact memorization, and usually consists of rules generated from routine and well defined practices. A sensor is more likely to exhibit simple reasoning based on this definition whereas an intuitor is more likely to exhibit abstract reasoning.

Chance, delMas and Garfield (2004), propose six components of statistical reasoning behavior: fluency, rules, consistency, integration, equilibrium, and confidence (p. 309). The dimension "rules" means that the student identifies and uses a formal rule to make predictions and explanations. The process of using rules to explain an outcome or process, or come to a certain conclusion is a trait that we would expect to see most likely surface in the sensor.

The dimension of "integration" entails "the extent to which ideas, concepts and procedures are connected" (Chance, delMas, & Garfield, 2004, p. 309). This is the core of abstract reasoning. The idea of integration involves the reliance of statistical relations among variables or information to calculate likelihoods of certain outcomes, an idea that is core to the characteristic of intuitive thinking.

The dimensions of fluency denote "how well the student understands and uses appropriate terminology, concepts, and procedures" (Chance, delMas, & Garfield, 2004, p. 309). Since a sensor is more likely to memorize facts, definitions, and formulas, I would expect to see a sensor to be more fluent in providing statistical reasoning, but perhaps not as strong in the area of applying the idea in practice.

The dimension of consistency means "the presence or absence of contradictory statements" (Chance, delMas, & Garfield, 2004, p. 309). Students will be asked multiple interrelated questions, and to provide justification about them. I am not sure if any student will exhibit consistency across questions within a related concept, based purely on a learning preference, but do believe that a student, who exhibits statistical reasoning, will show consistency. This student that portrays consistency will likely think abstractly, since these types of thinking and reasoning often align with solid statistical reasoning.

A sensor may be more fluent in terminology since they prefer to memorize facts and information, but may be inconsistent in the use of the concepts. Consistency really depends on the understanding of the student, and perhaps an intuitor may be more consistent if they actually understand the relationships between processes, but this needs to be determined on a per student basis and then validated. For example, do they consistently state that the distribution of sample means will be less variable than the population, or does this statement (of position) vary with posing the question in different ways? Do they relate information about variability to the fact that a larger sample size will yield a sampling distribution with less variability? This fact is often confusing for the student and I describe why I think that it is confusing. The population consists of all the data and has higher variability than a sampling distribution of means, which inevitably has fewer observations. So here we are stating that a distribution (of means) with fewer observations than the population, has a smaller variance, but if we take a larger sample and plot the means, this distribution of means, with more observations in the sample, but the same number of observations in the sampling distribution, will have a smaller variance than the previously mentioned one. The dimensions of equilibrium and confidence are not ones that can be easily tied to specific reasoning process or learning style.

The fact that the population has different qualities than the sampling distribution is one that is not easy to grasp. The intuitor might be more likely to comprehend this concept because it deals with relationships between several intertwining ideals. A sensor may not be able to coordinate these relationships, but might likely memorize certain key facts and ideas that will lead them to draw similar conclusions.

This can be a tricky process for the sensor because while they are memorizing that a larger sample size will lead to a smaller variance, they will most likely confuse this with comparing the population to the sampling distribution. Since the population has more observations than the sampling distribution, a student may argue that the population will have a smaller variance, completely disregarding the fact that the population consists of all values, and the sampling distribution of the mean consists of all possible means, which are bounded. Many students have made the mistake of stating that a sampling distribution of a larger sample size will vary more because there are more values chosen. Likelihood of outcomes is not considered which further supports the idea that students are thinking in simple terms rather than abstractly.

Theoretical Framework. The main theoretical framework of this study is the idea of statistical reasoning in regards to the sampling distribution of means, and how I would expect a sensing learner versus an intuitive learner to organize their thought process regarding four key components upon which sampling distributions build, namely: variability, distribution, normal distribution, and sampling. Table 6 expounds upon this topic. Using Chance, delMas and Garfield's (2004) terms of "Prerequisite knowledge to learning about sampling distribution", I define what I have termed, a "key component", use the authors' description of the component, and then explain how I believe a sensor or intuitor will likely fit into this description based on their preferred method of learning (p. 300).

Table 6. Conceptual Framework

Learning Style		Anticipated Reasoning of a Sensor	Anticipated Reasoning of an Intuitor
Key Component		Sensors are likely to be fluent in terminology as they prefer fact memorization, but are also likely to be inconsistent in using the concepts correctly or in application.	Intuitors prefer abstract Ideas, relationships and possibilities. Intuitors are likely to apply a concept to a problem, even if they have not seen the exact problem before; an intuitor is more likely to be able to make the connection. An intuitor may be less likely to memorize facts word for word and less likely to memorize formulas.
Variability	"What is a variable? What does it mean to say observations vary? Students need an understanding of the spread of a distribution in contrast to common misconceptions of smoothness or variety." (Chance, delMas, & Garfield, 2004, p 300)	A sensor is likely to view variability incorrectly. I expect a sensor is likely to believe that a variable is something based on observation that has any given value. It is likely that a sensor will believe that varying observations means they are changing within an individual group and not always the same. Since variance has an inverse relationship with sample size, but at the same time, the population has the largest variance whereas a sample has a smaller variance, which is counterintuitive since as sample size increase, variance decreases. This is likely to confuse the sensor that could have memorized the relationship but might not connect this in practice.	An intuitor is likely to view variability properly, being able to think abstractly, intuitor will likely expect the sampling distribution of means to have a smaller variance with increasing sample size. Intuitors will likely believe that a variable is a measure of some quantity. I believe the intuitor will likely be able to conceptualize the idea of variability and understand that the more observations in a sample, the smaller the variance will be.

Table 6. Continued

Distribution	"Students should be able to read and interpret graphical displays of quantitative data and describe the overall pattern of variation. This includes being able to describe different distributions of data, characterizing the shape, center and spread; and being able to compare different distributions on these characteristics. " (Chance, delMas, & Garfield, 2004, p 300)	A sensor is likely to believe that the distribution of a sample will be the same as a distribution of the population, BUT might possible believe that it will exhibit normal properties if they memorize definitions about normality and fail to apply them in the proper situation. The overarching concept of a symmetric/normal distribution may confuse the facts with application. It is likely that a sensor will believe that you need to observe the data in order to predict the shape or center of a distribution.	An intuitor is likely to be able to infer that the shape of a random sample will be close to the shape of the population with a mean similar to the population mean based on the trait that an intuitor thinks in terms of possibilities and relationships. It is likely that an intuitor will be able to draw inferences regarding the shape and center of a distribution based on the information given, and will not need any additional information.
Normal Distribution	"This includes properties of the normal distribution and how a normal distribution may look different due to changes in variability and center." (Chance, delMas, & Garfield, 2004, p 300)	A sensor is likely to believe that all things are normal due to ability to memorize definitions about normality, but I also suspect that they might be more likely to believe you need to actually see the evidence (data) before believing that this normalcy exists	An intuitor is likely to believe that the sampling distribution of means behaves normally. Likely to be able to think abstractly and view the distribution of a mean in terms of its relationship with the distribution.

Table 6. Continued

Sampling	"This includes random samples and how they are representative of the population. Students should be comfortable distinguishing between a sample statistic and a population parameter and should also be able to consider how sample statistics vary from sample to sample but follow a predictable pattern." (Chance, delMas, & Garfield, 2004, p 300)	Since a sensor tends to memorize facts and formulas, I think a sensor is likely to think that a sample is going to have a normal distribution because of the overarching theme of normality in relation to sampling distributions. I do believe that a sensor is likely to believe that a sampling distribution of means will follow a normal shape, but I do not think that a sensor will likely be able to consider how these statistics vary from sample to sample but follow a predictable pattern.	An intuitor is likely to believe that a sample is a subset of a population and thus will have the same properties and shape as the parent population. Likewise, I believe the intuitor will likely think that a sample mean follows a predictable pattern based on the mean of the population. I do not think that an intuitor will be likely to use the correct verbiage due to their lack of concern with details, but will likely be able to infer the correct relationships and patterns.

Summary

This chapter was intended as a review of the literature relevant to two main topics: Current literature and research in statistics education and learning styles in education research. Some of the early studies in Statistics Education attempted to identify and correct student misconceptions, but later efforts moved towards understanding how students think and process information.

Chapter 3 will describe the methods used in my research, provide an analysis of the different types of data collected is presented in Chapter 4, and finally a review of the conclusions, limitations and implications of this research will be given in Chapter 5.

Chapter 3: Quantitative Methodology

Introduction

For the quantitative portion of the study, I administered final exam questions to all Introduction to Statistics, students. The data analysis began the ILS survey. I plotted the four ILS dimensions (Sensing/Intuition, Visual/Verbal, Sequential/Global, and Active/Reflective) against one another to identify if there were any observable correlations between the Sensing/Intuitor and the other three dimensions: (Active/Reflective, Global/Sequential, Visual/Verbal). This is to ensure the dimensions are independent, though it has been noted that there is a correlation between Sensing-Intuitive and Sequential-global preferences (Felder, 2005, p. 108). The ILS survey is used to select a population of interest. Only students that sign a consent form are eligible to participate in my study. The learning style of the student is unknown to the researcher, until the data are gathered and the analysis is performed. Although students are selected to participate in an interview based upon learning styles, the interview is conducted without that information. Data collection and analysis are performed without knowledge of the learning style until it becomes necessary for comparison of the two groups.

I first administered the consent form and the ILS survey to ensure I had a representative sample of data, and second, to be able to subset the data for answering the research questions.

Initially, I only administered a few final exam questions, but as the data were analyzed in the earlier semesters, I began to question the logic behind the way students were selecting answers for multiple choice questions. I then adjusted the questions as well as added more in an attempt to capture the underlying thought process of the student. These adjustments are made in the Spring and Fall semesters in 2008.

The nature of Research Question 1 (Are there any differences between sensors and intuitors in their responses to sampling distributions activities?) is intended to highlight the differences between sensors and intuitors in how they answered questions involving the sampling distribution of the mean and evaluated the answer as either correct or incorrect.

Overview of Methods Used

I analyzed the results of 5 semesters of Final Exam Questions for ST311 from Fall 2006 to Fall 2008. There are approximately 10 sections taught each semester, by various teachers/graduate students. Each instructor followed the same format/course outline and used the same materials to teach the class; such as hands on classroom activities, applet demonstrations, and other methods aimed at meeting all learning preferences.

The first semester (Fall 2006) was used as a pilot study. A small sample of students participated in interviews for the spring and fall semesters of 2007. In the spring of 2008, I no longer conducted interviews and I added a required free response portion to the foundational knowledge questions presented on the final exam. This was in order to determine if the student could explain a reason for making a multiple choice selection.

Participant Selection

Each section of a ST311 course started with approximately 65 students. In each semester NCSU has approximately 10 sections of ST311. Students were informed of the study on the second day of class, and were asked to go online and view the consent form. (See Appendix A).

Students all had to fill out the ILS questionnaire online as a first homework assignment. The results were input into a spreadsheet, and students having an extreme preference towards Sensing or Intuition qualified for the study, based on the merit that I will only get useful information about the differences between sensors and intuitors if I observe the subjects at the extremes (Felder & Spurlin, 2005, p. 105). Since the ILS ranges from -11 to 11, (odd values only), I chose to filter the extreme values as the last three observable data points at either end of the spectrum, meaning that only students with a score of 7, 9 or 11 qualified as eligible for this study (likewise a score of -7, -9 or -11).

Data Analysis

In order to analyze how a student reasons about sampling distributions and answer the research questions, I performed quantitative analysis, to determine if there was a difference in how the students answered questions about sampling distributions based on a correct or incorrect response. The analysis methods are described in this chapter.

Quantitative Methods

Final Exam Questions. I administered 8 final exam questions to try to answer my Research Questions. In each of these analyses, I performed a simple Fisher's Exact Test. The test is useful because I have such a small sample size and am examining the significance of the relationship between the two categories, namely sensors versus intuitors.

Analysis Tool: Fisher's Exact Test. A Fisher's Exact Test is a statistical test used to determine if there are nonrandom associations between two categorical variables, in my example, these variables are sensing students and intuitive students. This test does not make any assumptions regarding the distribution of the data. I performed a Fisher's exact test on a

2x2 table as well as (when possible) an mx2 table where "m" represents the number of possible answer choices. The hypothesis H_o is that the intuitor and the sensor are equally likely to answer questions in the same way, versus the alternative hypothesis, H_a, that there is an observable difference between how a sensor and an intuitor thinks about, or chooses an answer regarding sampling distributions. I used PROC FREQ in SAS® to perform these tests and evaluate them at the 0.05 level, though I may note when a test is significant at the 0.10 level. A significant result, one having a p-value less than 0.05 will indicate that we can safely REJECT the null hypothesis, H_o, in favor of supporting the alternative hypothesis, H_a, namely, that there is significant evidence to believe that sensors and intuitors behave differently.

Sampling Distribution Question taken from CAOS. The CAOS test (Comprehensive Assessment of Outcomes in a First Statistics Course), was designed to provide an instrument that would assess students' statistical reasoning after any first course in statistics. Since a main emphasis of Introductory Statistics is that the students understand the concept of Sampling distributions and its properties, a good way to assess the understanding of the basic concept of the sampling distribution of a mean versus a sample, is to present relevant questions, taken from CAOS, since it has proved to be a valid measure of the importance of learning outcomes in statistic (ARTIST, 2006).

From CAOS, a pretest/posttest was conducted to the understanding of how to select an appropriate sampling distribution for a particular population and sample size, asks students to select a graph among three histograms that represent a sampling distribution of sample means for a given sample size. The test was designed and validated to measure the most important learning outcomes for students in a non-math, first statistics course. There are a noticeable number of students who selected the correct response on the pretest but answered incorrectly on the posttest.

Data analysis begins with collection of the final exam multiple choice answers to answer the first research question.

Research Question 1: Are there any differences between sensors and intuitors in their responses to sampling distributions activities?

The first two exam question used to assess students' understanding and comprehension of sampling distributions will be a histogram of a parent population. The student will be asked the following questions which will precede a graph with four solutions. These questions are taken from the CAOS test described above.

1. Which of the histograms represents a single random sample of 500 values from this population?
2. Which of the histograms represents a distribution of 500 sample means from random samples of size 100 from this population?

By administering and evaluating this question, I compared the two categories within the Sensing/Intuitive dimension and identified if any differences exist. Likewise, I evaluated whether or not the student falls into the common misconception that a sampling distribution of means will mirror the parent population. The same four histograms are presented for both questions. One of the selections is blatantly incorrect for either question. This is to establish if the student is merely guessing at an answer.

In the first two semesters, questions presented to the students give a right skewed distribution as in Figure 1. The first question for my analysis presents the students with a right skewed distribution. In the latter groups of semesters, the students are given either a version with a bimodal distribution or a version with a right skewed distribution. The answer choices for every question present a distribution of 500 sample means from random samples of size 100, a distribution of a single random sample of 500 values taken from the population, as well as two “distractor” choices. By this I mean, a histogram of values that is not the correct answer for either question, but intended to distract the student from the correct answer if careful thought is not exercised.

The first question presented consists of two parts. The student is given a population distribution and asked to answer questions, first about the distribution of a sample and then about the distribution of sample means.

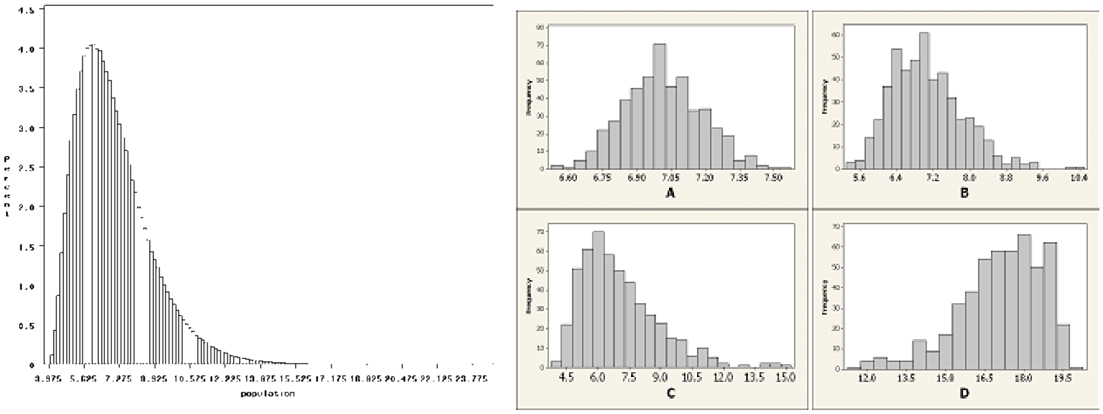


***Figure 1*. Final Exam Fall 2006 - Spring 2007**. The graph at the left is a distribution for a population with mean 7.0 and standard deviation 1.7.

After careful consideration and a discussion with Bob delMas at USCOTS 2007, I decided to put all of the multiple choices on the same numeric scale. In the last three semesters of data collection, fall 2007- fall 2008, the students received one of two versions of the test.

The answer choices for this version of the final exam question follow this theme: Choice A is a distribution of 500 sample means from simple random samples of size 100. Choice B is the distribution of 500 sample means from a simple random sample of size 10. Choice C is a distribution of a simple random sample of size 500. Choice D is a distribution of 500 sample means from a sample of size 100, with a center that is greater than the population center, as in Figure 2.

In the first two semesters, the distractor histogram was a distribution of 500 sample means from a smaller sample size, and a left skewed distribution. The left skewed distribution was clearly random, and only students “guessing” or not paying careful attention will make this as a selection. In the last three semesters, the distractor histogram was again, a distribution of 500 sample means taken from a small sample size (so it appears normal, but with a large variance), and a distribution of 500 sample means taken from a sample of size 100, but centered at a different mean. This was done to assess whether or not the student is paying attention to the details, or if they are just looking for a normal distribution and ignoring all pertinent information

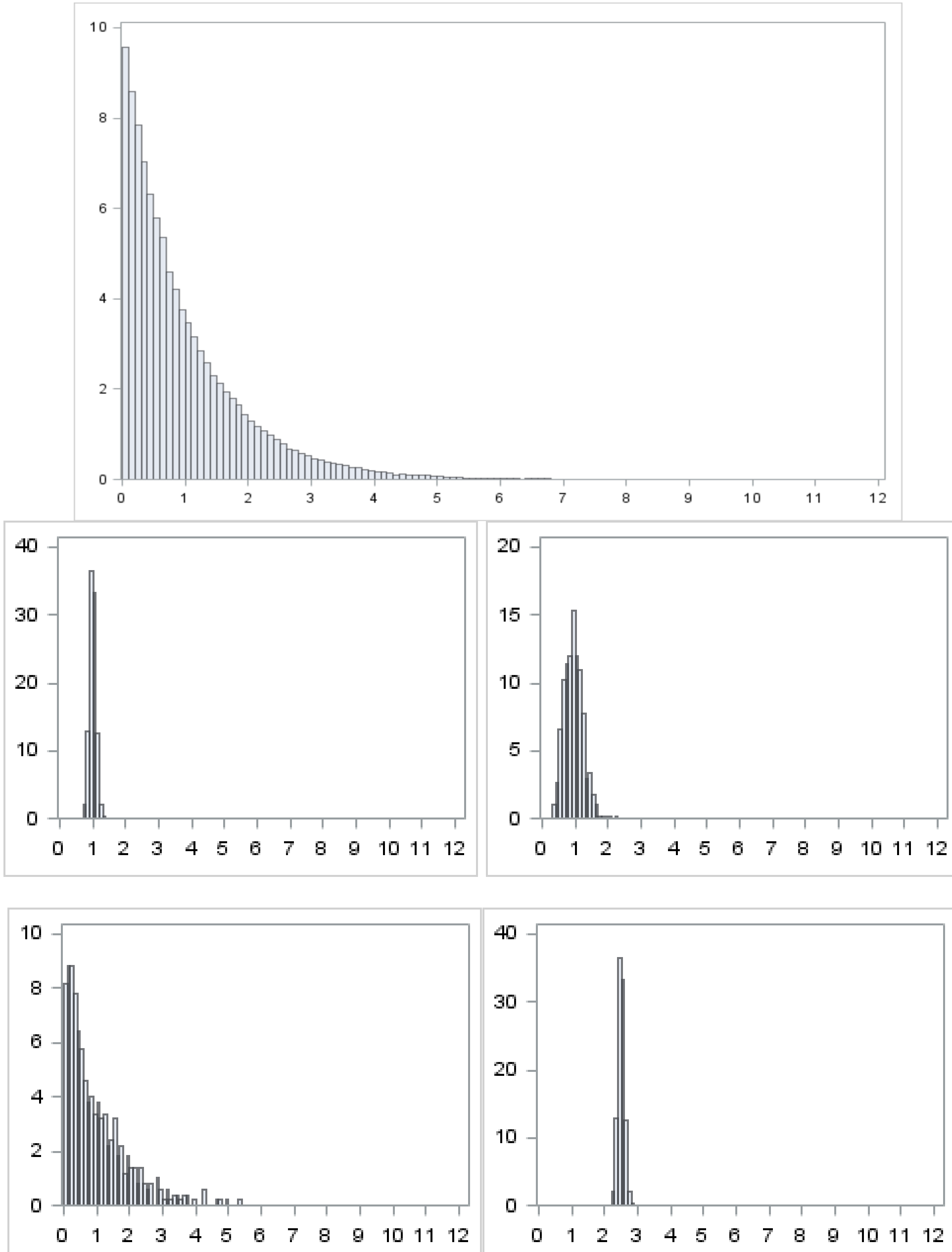


Figure 2. Final Exam Fall 2007 - Fall 2008, Version A. The graph at the top is a distribution for a population with mean 1.0 and standard deviation 1.0.

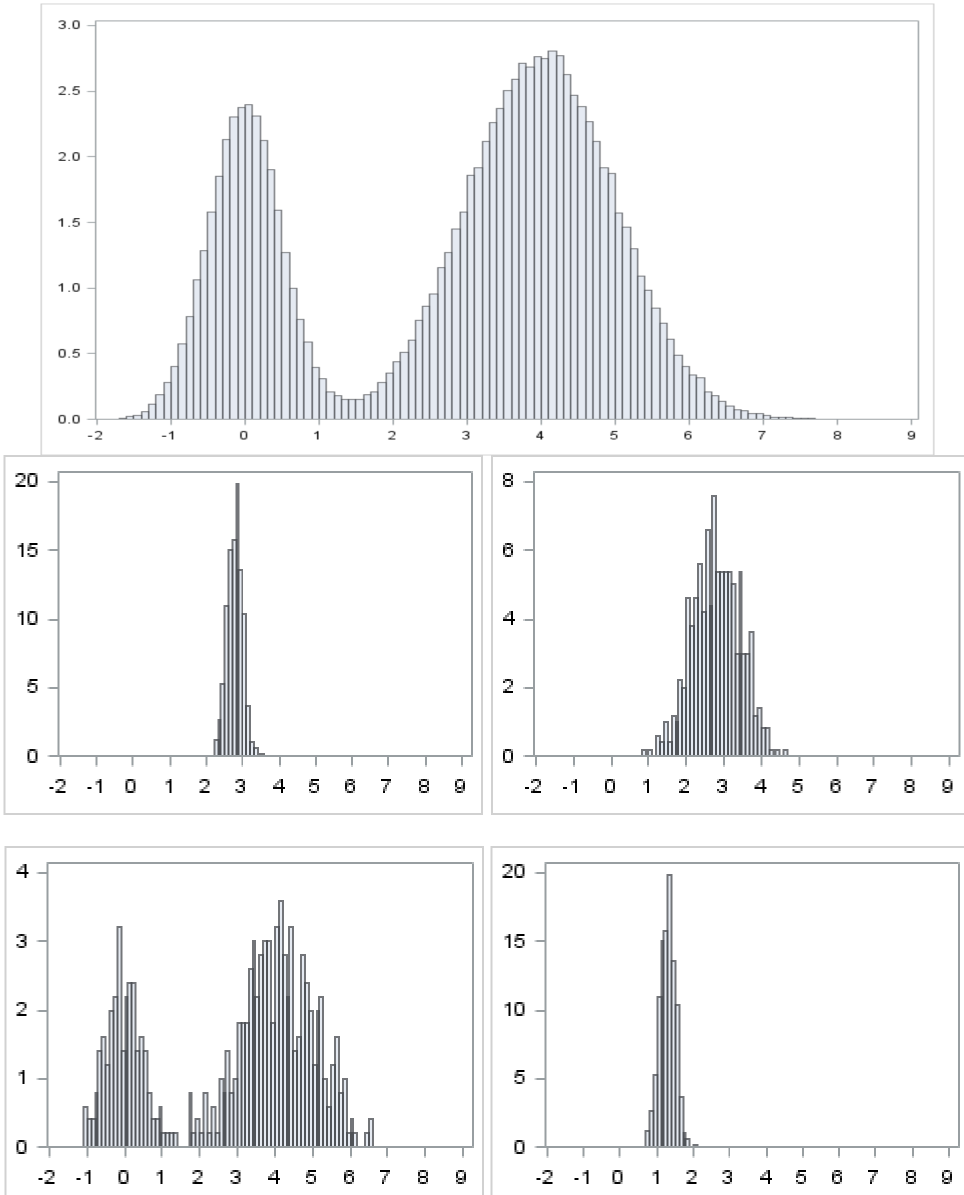


Figure 3. Final Exam Fall 2007 - Fall 2008, Version B. Five histograms are presented above. The graph at the top is a distribution for a population with mean 2.8 and standard deviation 2.

The answer choices for both versions follow the same theme. Choice A is the distribution of 500 sample means from a simple random sample of size 100. Choice B is the distribution of 500 sample means from a simple random sample of size 10. Choice C is a distribution of a simple random sample of size 500. Choice D is a distribution of 500 sample means from a sample of size 100, with a center that is greater than the population center.

Written Multiple Choice Questions. Data analysis then continued with collection of the final exam multiple choice answers to answer the first research question**.** Research Question 1: Are there any differences between sensors and intuitors in their responses to sampling distributions activities?

Students are given one of two test versions. The main difference in the test version is the ordering of the test questions, to discourage the student from cheating. Two of the test questions are presented in two formats, a word problem or a more straight forward question, set up to determine if the way a question is presented affects the student's ability to answer correctly.

I performed a Fisher's Exact test on the multiple choice answers, first on a 2x2 table of the correct answer versus incorrect answer between sensors and intuitors. The Fisher's Exact Test is useful for small samples in comparing behavior between two groups.

I performed this test on six final exam questions to determine an answer to the first research question: Are there any differences between sensors and intuitors in their responses to sampling distributions activities?

The students are presented six word problems, with multiple choice answers. The p-value from a Fisher's Exact Test will indicate whether or not the values are significantly different. I also performed an exact test on the version to make sure that there were no significant differences between Version A and Version B, as this would bias the results of my study.

Evaluating a Z-statistic: Probability and Range. Another facet that I plan to examine is the difference between sensors and intuitors in the way in which they solve a word problem involving a normal distribution in evaluating a z-statistic and p-value. The question is presented in two different ways and randomly distributed to the students on either version of the exam. Both versions of the question are asking the student to calculate the probability that the observed value lies between an upper and lower value. The first version of the question will involve a story about a wheat field and a random sample of bushels gathered. The second version of the question is straightforward and presents just the facts. Both questions provided the student with the population mean and standard deviation.

> A farmer in western Kansas knows that the amount of wheat on each acre of his farm is normally distributed with a mean of 43 bushels with a standard deviation of 4 bushels. If we select an acre at random from his farm, what is the probability that it yields between 40 and 50 bushels?

The second way this problem was presented was more complex in nature, and thus was initially thought to be more difficult for the sensor:

We have a random variable X that has a normal distribution. This normal distribution has a mean of 67 and a standard deviation of 8. What is the probability that the variable X is between 50 and 80?

Upon further examination of the question, I realized that the first question may be more complex in nature, and thus more difficult for the sensor to answer. The second form of the question is straightforward and possibly easier for a student to plug in values, a defining quality of a sensor. If sensors are represented as noted, then they might tend to recall the formula and simply plug in the numbers. An intuitor may be equally likely to answer the second question correctly, but due to the nature of the intuitor to read into underlying meanings as well as realize possibilities and relationships, I believe that an intuitor would perform better on question one due to the complex nature.

The next question presented to the student, asks them to calculate a z-statistic and p-value. The question is presented in two possible ways, but there is no distinguishable difference between the two questions except that one asks about cow's weight and the other asks about wheat bushel yield.

A dairy farmer knows that the weights of his mature milk cows follows a normal distribution with a mean weight of 1350 pounds and standard deviation 40 pounds. The farmer is going to select 20 cows at random to be used in a special program. What is the probability that the average weight of these 20 cows will be over 1370 pounds?

The second way this was worded is as follows:

> A farmer in western Kansas knows that the amount of wheat on each acre of his farm is normally distributed with a mean of 38 bushels and a standard deviation of 6 bushels. If we select 20 acres at random, what is the probability that the average yield is over 42 bushels?

I performed analysis on the question to see if the students exhibit any differences in their ability to solve the problem. My hypothesis is that the intuitor is more likely to do better due to the reasons previously given, namely the likely ability of an intuitor to connect ideas in practice and use inferential skills.

Foundational Knowledge: Likelihood, Center, Spread, and Shape. The remaining four questions cover the aspects deemed necessary as building blocks to understanding sampling distributions. The questions from this problem aim at understanding how students perceive the center, shape, and variability of the sampling distribution in comparison with the parent population. The student is encouraged to explain their choice with a written explanation that is examined in the following section.

Likelihood. First, a question of likelihood is presented. This question requires an understanding that a smaller sample mean is less likely to be representative of the population parameter.

> According to a recent report the average income of people in Adair County, Missouri is $31,023. Which of the following is more likely? Pick one and explain.
>
> a. We take a random sample of 10 people from this county and find that the average is over $50,000.
> b. We take a random sample of 200 people from this county and find that the average is over $50,000.
> c. We do not have enough information to answer this question.

The final portion of the assessment presents three questions based on this problem:

> A large multinational company knows that the average age of their employees is 34 years. They also know that the standard deviation of the ages of these employees is 8 years. A manager from human resources is going to randomly select a sample of 100 employees. Consider the sampling distribution of the mean of a sample of size 100 from this population. From the information given, which of the following statements are true about this sampling distribution?

Center

Which of the following is true? Pick one and explain.

a. The sampling distribution will be centered at 34.

b. We do not know where the sampling distribution will be centered.

Spread

Which of the following is true? Pick one and explain.

a. The sampling distribution will have a higher standard deviation than the population.

b. The sampling distribution will have the same standard deviation as the population.

c. The sampling distribution will have a smaller standard deviation than the population.

Shape

Which of the following is true? Pick one and explain.

a. We know that the shape of the sampling distribution of the mean will be right skewed.

b. We know that the shape of the sampling distribution of the mean will be approximately symmetric.

c. We cannot tell what the shape sampling distribution of the mean will look like.

I believe that these eight questions provided adequate support to answer the first research question: Are there any differences between sensors and intuitors in their responses to sampling distributions activities?

Ethical considerations

In order to comply with standards of the IRB, I had all students who wished to participate, read and sign the CONSENT document, Appendix A. Students were allowed to opt out by checking, "I do not consent". These students were not included in any of our data collection, including the ILS survey, focus group interviews, and final exam questions.

Timeline

Fall 2006 - data collection

Spring 2007 - data collection

Fall 2007 - data collection

Spring 2008 - data collection

Fall 2008 - data collection

Table 7. Timeline of data collection, analysis, and participant count

Semester	Data Collection	Analysis	Participants (#)
Fall 2006	Exam Multiple Choice Questions	Fisher's Exact test on MC questions	103
Spring 2007	Exam Multiple Choice Questions	Fisher's Exact test on MC questions	126
Fall 2007	Exam Multiple Choice Questions	Fisher's Exact test on MC questions	129
Spring 2008	Exam Multiple Choice Questions	Fisher's Exact test on MC questions	180
Fall 2008	Exam Multiple Choice Questions	Fisher's Exact test on MC questions	144
January 2009 - June 2009	NONE	Data Analysis and writing a first draft	ALL

Summary

In this chapter, I described the quantitative research methodology I employed in my research. Five semesters of introductory statistics classes were taught between 2006 and 2008 at NCSU, using the same syllabus and teaching materials and objectives. I collected ILS survey data and multiple final exam responses.

Chapter 4: Quantitative Results

Overall, 81% of the students filled out the consent form, and over 50% of the students agreed to participate in the study. Only students that consented to participate in the study and had an extreme ILS score in the Sensing/Intuitive dimension were used in the study. As stated in the previous chapter, I defined the extreme values as the last three observable data points at either end of the spectrum, meaning that only students with a score of 7, 9 or 11 qualified as eligible for this study (likewise a score of -7, -9 or -11). From this subset of students, only the students that completed the course and took the final exam are included in Table 8.

As presented in the table below, 682 students that consented to participate also qualified for the study based on their ILS score. All of the other consenting students did not qualify. This table shows the number of students consenting to participate in the study, and the shaded bars in Figure 4 display eligible students, based on the ILS score.

Table 8. Participant Rate

Semester	Roster Size	Submitted	Consented	Eligible participant	Extreme Sensors	Extreme Intuitors
Fall 2006	679	540 (79.5%)	335 (49.3%)	103	92	11
Spring 2007	682	653 (95.7%)	492 (72.1%)	180	150	30
Fall 2007	643	624 (97.0%)	444 (69.1%)	126	101	25
Spring 2008	655	477 (72.8%)	351 (53.5%)	144	108	36
Fall 2008	722	464 (64.3%)	349 (48.3%)	129	110	19
TOTALS	**3381**	**2758 (81.6%)**	**1971 (58.3%)**	**682 (20%)**	**561**	**121**

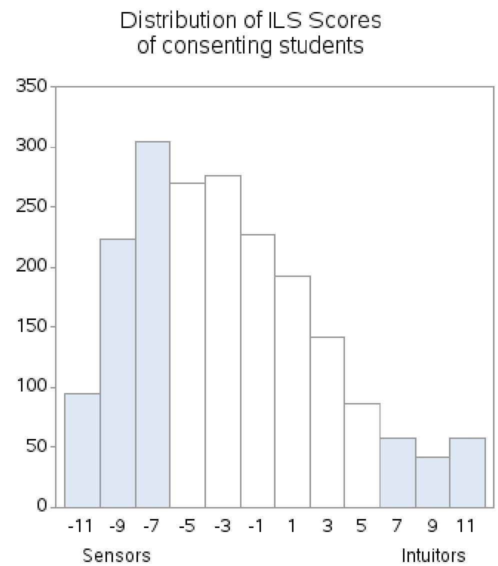


Figure 4. Distribution of ILS Scores of Consenting Students.

My first semester analysis was exploratory and aimed at getting an understanding of the problem and discovering ways to observe and assess understanding.

The distribution of sensor and intuitor groups yield 73.75% sensor versus 26.25% intuitor, compare to the MBTI demographic results of 73.3% sensors and 26.7% intuitors (Myers-Briggs Foundation, n.d.).

Research Question 1.

Are there any differences between sensors and intuitors in their responses to sampling distributions activities?

Hypothesis. I expect that overall; intuitors will be more likely than sensors to understand the concept of the sampling distribution. I believe that sensors and intuitors understand sampling distributions differently, and that intuitors are more likely to comprehend more abstract concepts than the sensor. I believe that sensors are more likely than intuitors to fail to apply a formula or definition in practice, but that the sensor will be better at fact memorization and recalling a formula. My belief is that a sensor be stronger than an intuitor at solving an equation involving “plug and chug” style, but that the intuitor will be stronger than the sensor at solving complex problems and applying the information given in a word problem to solving the answer involving the understanding of the sampling distribution of means.

The following questions, which are presented in the Introductory Statistics final exam, are taken from the CAOS exam, detailed in Chapter 3, since it has been validated as a good measure of the students understanding.

To gain a better insight of the differences between the sensing student and the intuitive student in the way they understand sampling distributions, I performed analysis on several questions from the final exam. I recorded results from five semesters of students that took the final exam.

The variation between Fall 2006/Spring 2007 and Fall 2007/Spring 2008/Fall 2008 lends to two groups of results. The following results intend to list the exam question, address the nature of the question along with the intended purpose in assessment, and finally list the results collected over a number of semesters.

Students are presented a histogram of a population distribution and are first asked to select from four sampling distribution histograms the one that best represents a histogram of a single random sample of 500 values from the population. The second part of this question asked the student to identify the histogram which best represents a distribution of 500 sample means from random samples of size 100 from the population.

In this question, there are several points to analyze. First, does the student think that the sampling distribution of means should look like the parent population? It is important in writing this question to present a parent population that did not have a normal distribution so that the selection made by the student would not be confounded by this fact. Students are either presented a skewed distribution, a uniform distribution, or a bimodal distribution. The solutions presented to the students are three normal distributions having different variance and center, and one simple random sample, taken from the parent population, to represent the solution for question one.

Fall 2006 – Spring 2007. The first part of the analysis comes from the Fall 2006 and Spring 2007 semesters. This exam question as shown in Figure 1, gives answer selections that are not on the same scale. For the questions regarding distribution of sample and sample means, the answer choices are as follows.

Fall 2006 and Spring 2007 answer choices: (See Figure 1)

a. Histogram of 500 means taken from a SRS of size 100 from the population.

b. Histogram of 500 means taken from a SRS of size 4 from the population

c. Histogram of 500 SRS taken from the population

d. Histogram of a left skewed distribution (completely random)

Graphic Distribution on a Simple Random Sample.

Which of the histograms represents a simple random sample of 500 values from this population? See Figure 1.

For the distribution of a sample of 500, over 50% of sensors and intuitors correctly chose C. I had 283 students participate in the study in the first two semesters of data analysis, 41 intuitors and 242 sensors. The results of the distribution of answer selections are presented in Table 9.

Table 9. Fall 2006 - Spring 2007: Graphic Question on a SRS of 500 observations

ILS Category / ANSWER	Intuitors	Sensors
A	9 (21.95%)	65 (26.86%)
B	9 (21.95%)	41 (16.94%)
C*	21 (51.22%)	123 (50.83%)
D	2 (4.88%)	13 (5.37%)
TOTAL	41	242
		p-value =1

Note: * indicates the correct answer. Percentages indicate the proportion of students within the ILS category choosing that answer out of all within the category.

The p-value is calculated by comparing the total number of correct intuitors versus the total number of correct sensors, where all incorrect answers are evenly weighted. A fisher's test for exactness is used due to the small number of observations. The p-value of 1 indicates that this test is not significant; hence we do not reject H_o which states that sensors and intuitors are equally likely to make the correct answer choice.

It is important to observe and mention that the proportion of students selecting a normal distribution (either selection A or B) is nearly the same as the proportion of students making the correct choice. I believe that this is due to the same reason as explained by the students in the pilot study: The student recalls the underlying theme of normality, and thus lends to the logic that a normal distribution would be the best choice.

Graphic Distribution on a Sampling Distribution of Means.

> Which of the histograms represents a distribution of 500 sample means from random samples of size 100 from this population?

Next I examined the number of sensors and intuitors that chose a normal distribution as the selection for the sampling distribution of means. Most of the students chose the correct normal distribution as indicated in Table 10.

Table 10. Fall 2006 - Spring 2007: Graphic Question on distribution of 500 means

ILS Category / ANSWER SELECTION	Intuitors	Sensors
A*	21 (51.22%)	99 (40.91%)
B	3 (7.32%)	57 (23.55%)
C	13 (31.71%)	66 (27.27%)
D	4 (9.76%)	20 (8.26%)
TOTAL	41	242
		p-value=0.2348

Note. * shows the correct answer. Percentages indicate the proportion of students within the ILS category choosing that answer out of all within the category

The p-value is calculated by comparing the total number of correct intuitors versus the total number of correct intuitors, where all incorrect answers are evenly weighted. Overall, I observe no significant difference between the sensors and intuitors in making the correct selection, or in choosing either of the normal distributions. It is interesting to look at the number of students that chose B for their choice. Can we infer from this that the sensors are more likely to believe that while the distribution of sample means is normal, they do not have a good understanding that a large sample size such as n=100 is adequate to have a smaller varying distribution?

Due to this vast difference in proportions between sensing and intuitive students on selecting answer “B”, I performed a Fisher’s Exact Test on the likelihood that a sensor is more likely than the intuitor to choose “B” which indicates a distribution of sample means that is normal but has a larger variance and slight skew than the other answer choices. This is with a p-value of 0.02, meaning that we reject H_o that sensors are just as likely as intuitors to select answer “B” in favor of the alternative that they are likely to be different. I do not believe that the scale in this test was fair (due to previous comments regarding this) and hence, will not use this result in conclusions.

Comparison/Interaction of Simple Random Sample and Sampling Distributions. Next I considered the responses of both questions, where the Simple Random Sample (SAMPLE) answer is represented on the row and the Sampling Distribution of Means (MEANS) answer is represented on the column.

Table 11. Fall 2006 - Spring 2007: Interaction of SAMPLE and MEANS

SAMPLE \ MEANS	A*		B		C		D	
	Intuitor	Sensor	Intuitor	Sensor	Intuitor	Sensor	Intuitor	Sensor
A	1 (2.44%)	7 (2.89%)	1 (2.44%)	16 (6.61%)	6 (14.63%)	34 (14.10%)	1 (2.44%)	8 (3.31%)
B	4 (9.76%)	20 (8.26%)	0	1 (0.41%)	3 (7.32%)	16 (6.61%)	2 (4.88%)	4 (1.65%)
C*	15* (36.59)	67* (27.69)	2 (4.88%)	33 (13.64)	3 (7.32%)	15 (6.20%)	1 (2.44%)	8 (3.31%)
D	1 (2.44%)	5 (2.07%)	0	7 (2.89%)	1 (2.44%)	1 (0.41%)	0	0

Note: * indicates the correct answer. Percentages indicate the proportion of students within the ILS category choosing that answer out of all within the category.

It is interesting to note that 67 sensors (27.69%) and 15 intuitors (36.59%) answered both questions correctly. When I performed a Fisher's Exact Test I observe a p-value of 0.2661 which is not significant thus I do not reject H_o. If I ignore how the student interpreted the spread in the question regarding the means and look at the Fisher's exact Test for sensors versus intuitors, we have 100 sensors and 17 intuitors with a partially correct answer, again, ignoring spread, the p-value is 1.0. This leads me to believe that intuitors are not any more likely than sensors to take spread into account. I analyzed this data in similar ways in data for upcoming semesters, adjusting the graphs so that they are all on the same scale.

Fall 2007 – Fall 2008. Following the Spring 2007 semester, the final exam was slightly altered putting all answer choices on the same scale. Two separate versions were presented to the students for the Fall 2007, Spring 2008 and Fall 2008 semesters. The first version of the question displays a right skewed distribution and the second version of the question displayed a bimodal distribution as displayed in Figure 2 and Figure 3. The answer selections for both versions were:

a. Histogram of 500 means from a SRS of size 100 taken from the population

b. Histogram of 500 means from a SRS of size 10 taken from the population

c. Histogram of 500 samples from a SRS taken from the population

d. Histogram of 500 means from a SRS of size 100 taken from a population with a larger mean than the population presented in the problem.

First the overall results are displayed, and then the results are given by version to test for a difference between version A and version B.

Graphic Distribution on a Simple Random Sample.

Which of the histograms represents a simple random sample of 500 values from this population? See Figures 2 and 3.

For the distribution of a sample of 500, nearly 60% of both sensors and intuitors correctly chose C. I had 399 students participate in the study in the final three semesters of data analysis, 80 strong intuitors and 319 strong sensors. The results of the distribution of answer selections for the question on simple random samples are presented in Table 12.

Table 12. Fall 2007 - Fall 2008: Graphic Question on SRS of 500 observations

SAMPLE	Intuitors	Sensors
A	11 (13.75%)	43 (13.52%)
B	15 (18.75%)	69 (21.70%)
C*	49 (61.25%)	187 (58.62%)
D	5 (6.25%)	20 (6.29%)
TOTAL	80	319
		p-value: 0.7037

Note: * indicates the correct answer. Percentages indicate the proportion of students within the ILS category choosing that answer out of total number of students within the category.

The observed p-value of 0.7037 indicates that we have no reason to reject H_o suggesting that there is not a significant difference between how an intuitor and a sensor answer this question regarding the distribution of a simple random sample. All other counts seem to be relatively proportional within the answer selection. The Fisher's Exact test on the 4x2 table yields a p-value of 0.9559.

Next I examined the counts within each test version. Most of the students chose the correct normal distribution as indicated in Table 13 and Table 14, however it appears that sensors were less likely to do well on Version A and intuitors were less likely to do as well on Version B. Despite the varying proportions, this is not a statistically significant number, and thus I am still led to believe that sensors are just as likely as intuitors to make the same answer selection regarding this question.

Table 13. Fall 2007 - Fall 2008: Graphic Question on SRS of 500, Version A

SAMPLE	Intuitors	Sensors
A	5 (11.90%)	22 (13.75%)
B	6 (14.29%)	35 (21.88%)
C*	30 (71.43%)	97 (60.63%)
D	1 (2.38%)	6 (3.75%)
		p-value: 0.2144

Note: * indicates the correct answer. Percentages indicate the proportion of students within the ILS category choosing that answer out of total number of students within the category.

Table 14. Fall 2007 - Fall 2008: Graphic Question on SRS of 500, Version B

SAMPLE	Intuitors	Sensors
A	6 (15.79%)	21 (13.29%)
B	9 (23.68%)	34 (21.52%)
C*	19 (50.00%)	89 (56.33%)
D	4 (10.53%)	14 (8.86%)
		p-value: 0.5863

Note: * indicates the correct answer. Percentages indicate the proportion of students within the ILS category choosing that answer out of total number of students within the category.

Graphic Distribution on a Sampling Distribution of Means.

Which of the histograms represents a distribution of 500 sample means from random samples of size 100 from this population? See Figures 2 and 3.

I examined the number of sensors and. intuitors that chose a normal distribution as the selection for the sampling distribution of means. Most of the students did not choose the correct normal distribution as indicated in Table 15, but instead chose answer B, selecting the sampling distribution with a larger variance. The p-value for the test that intuitors and sensors are equally likely to select the correct answer is 0.1120 indicating that intuitors might be slightly more likely to select the correct answer, but this is not significant. The Fisher's Exact test on the 4x2 table yields a p-value of 0.1630, also not significant, leading me to the conclusion that sensors and intuitors are equally likely to make the same answer choices and there is no observable difference between these two categories, at least in respect to this question as well as the question regarding the distribution of asimple random sample.

Table 15. Fall 2007 - Fall 2008: Graphic Question on Distribution of Means

MEANS	Intuitors	Sensors
A*	26 (32.50%)	74 (26.27%)
B	33 (41.25%)	123 (38.68%)
C	15 (18.75%)	76 (23.90%)
D	6 (7.50%)	45 (14.15%)
		p-value: 0.1120

Note: * indicates the correct answer. Percentages indicate the proportion of students within the ILS category choosing that answer out of total number of students within the category.

Next I examined the counts within each test version as presented in Table 16 and Table 17, however it appears that sensors were less likely to do well on Version A and intuitors were less likely to do as well on Version B. Despite the varying proportions, this is not a statistically significant number, and thus I am still led to believe that sensors are just as likely as intuitors to make the same answer selection regarding this question.

Table 16. Fall 2007 - Fall 2008: Version A, Means

MEANS	Intuitors	Sensors
A*	12 (28.57%)	43 (26.88%)
B	17 (40.48%)	43 (26.88%)
C	10 (23.81%)	48 (30.00%)
D	3 (7.14%)	26 (16.25%)
		p-value: 0.8468

Note: * indicates the correct answer. Percentages indicate the proportion of students within the ILS category choosing that answer out of total number of students within the category.

Table 17. Fall 2007 - Fall 2008: Version B, Means

MEANS	Intuitors	Sensors
A*	14 (36.84%)	31 (19.62%)
B	16 (42.11%)	80 (50.63%)
C	5 (13.16%)	28 (17.72%)
D	3 (7.89%)	19 (12.03%)
		p-value: 0.0314

Note: * indicates the correct answer. Percentages indicate the proportion of students within the ILS category choosing that answer out of total number of students within the category.

This p-value of 0.03 appears to be significant, but after reviewing the other data, this is most likely due to a type I error since there are no significant differences detected in the other analyses. The Fisher's Exact Test on the 4x2 table yields a p-value of 0.1956 leading me to believe that the significance observed was due to error.

Comparison/Interaction of Simple Random Sample and Sampling Distributions. Next we take a look at the responses of both questions, where the Simple Random Sample (SAMPLE) answer is represented on the row and the Sampling Distribution of Means (MEANS) answer is represented on the column. I simplified the table to only Correct and Incorrect answers and calculated a p-value using Fisher's Exact Test on the likelihood that sensors and intuitors get both questions correct.

Table 18. Fall 2007 - Fall 2008: Interaction of SAMPLE and MEANS

Sample \ Means	CORRECT		INCORRECT	
	Intuitor	Sensor	Intuitor	Sensor
CORRECT	18 (22.5%)	50 (15.67%)	31 (38.75%)	137 (42.95%)
INCORRECT	8 (10.00%)	25 (7.84%)	23 (28.75%)	107 (33.54%)
				p-value=0.1823

Note: Percentages indicate the proportion of students within the ILS category choosing that answer out of total number of students within the category.

In Table 18, I looked at the combination of answer selection for both the question regarding the simple random sample of size 500 and the question regarding the sampling distribution of 500 means. After calculating the Fisher's Exact Test P-value for the number of sensors versus intuitors to answer both correct in the Fall07-Fall08 semesters, the p-value is 0.1823 which indicates that there is no difference in how sensors and intuitors answer both of these questions regarding sampling distributions.

The p-values are assessing the number of correct answers versus number of incorrect answers between the categories of sensor and intuitor, and indicates using a Fisher's Exact test, as to whether or not there is a significant difference in how they make a selection regarding distribution (either of means or of a simple random sample). There seems to be no significant difference between sensors and intuitors in the way they answer questions about the sampling distribution, except in version B, in the question regarding sample means. Intuitors seem more likely to choose the correct selection for the distribution of sample means, but we cannot be sure this is not due to a Type I error.

Foundation Questions. The remaining four questions cover the aspects deemed necessary as building blocks to understanding sampling distributions. The questions from this problem aim at understanding how students perceive the center, shape, and variability of the sampling distribution in comparison with the parent population

Likelihood

According to a recent report the average income of people in Adair County, Missouri is $31,023. Which of the following is more likely? Pick one and explain. a. We take a random sample of 10 people from this county and find that the average is over $50,000. b. We take a random sample of 200 people from this county and find that the average is over $50,000. c. We do not have enough information to answer this question.

Table 19. Fall 2007 - Fall 2008: Likelihood

Likelihood	Intuitors	Sensors
A*	50 (63.29%)	165 (51.72%)
B	8 (10.13%)	55 (17.24%)
C	21 (26.58%)	99 (31.03%)
		p-value: 0.1027

Note: * indicates the correct answer. Percentages indicate the proportion of students within the ILS category choosing that answer out of total number of students within the category.

The p-value tests the hypothesis, Are sensors and intuitors equally likely to make the correct answer selection. The p-value of 0.1027 indicates that it is nearly significant at the 0.10 level, such that I reject H_o in favor of intuitors being more likely than sensors to answer this question correctly. I believe that this is approaching significance in favor of intuitors being more likely than sensors to answer this question correctly.

I am interested to see if the data I collected on the response justification supports this finding in the following qualitative section; however the p-value on the 3x2 contingency table is 0.1443, leading me to fail to reject the null hypothesis.

The remainder of the assessment presents three questions based on this problem

A large multinational company knows that the average age of their employees is 34 years. They also know that the standard deviation of the ages of these employees is 8 years. A manager from human resources is going to randomly select a sample of 100 employees. Consider the sampling distribution of the mean of a sample of size 100 from this population. From the information given, which of the following statements are true about this sampling distribution?

Center

Which of the following is true? Pick one and explain.

a. The sampling distribution will be centered at 34.

b. We do not know where the sampling distribution will be centered.

Table 20. Fall 2007 - Fall 2008: Center

Center	Intuitors	Sensors
A*	59 (73.75%)	219 (68.65%)
B	21 (26.25%)	100 (31.35%)
		p-value: 0.4163

Note: * indicates the correct answer. Percentages indicate the proportion of students within the ILS category choosing that answer out of total number of students within the category.

Table 20 shows the proportion of intuitors and sensors that chose the correct answer for the question regarding the center of the sampling distribution of means. The p-value is 0.4163 which is not significant, thus we do not reject H_o.

Spread

Which of the following is true? Pick one and explain.

a. The sampling distribution will have a higher standard deviation than the population.

b. The sampling distribution will have the same standard deviation as the population.

c. The sampling distribution will have a smaller standard deviation than the population.

Table 21. Fall 2007 - Fall 2008: Spread

SPREAD	Intuitors	Sensors
A	27 (33.75%)	89 (27.90%)
B	21 (26.25%)	78 (24.45%)
C*	32 (40.00%)	152 (47.65%)
		p-value = 0. 2591

Note: * indicates the correct answer. Percentages indicate the proportion of students within the ILS category choosing that answer out of total number of students within the category.

Table 21 shows the proportion of intuitors and sensors that chose the correct answer for the question regarding the standard deviation of the sampling distribution of means versus the population standard deviation. The p-value is 0.2591 which is not significant, thus we do not reject H_o. Furthermore the p-value for the 3x2 table is 0.4140.

Shape

Which of the following is true? Pick one and explain.

a. We know that the shape of the sampling distribution of the mean will be right skewed.

b. We know that the shape of the sampling distribution of the mean will be approximately symmetric.

c. We cannot tell what the shape sampling distribution of the mean will look like.

Table 22. Fall 2007 - Fall 2008: Shape

Shape	Intuitors	Sensors
A	33 (41.25%)	124 (38.87%)
B*	30 (37.50%)	105 (32.92%)
C	17 (21.25%)	90 (28.21%)
		p-value: 0.4318

Note: * indicates the correct answer. Percentages indicate the proportion of students within the ILS category choosing that answer out of total number of students within the category.

The p-value for H_o is 0.4318 leading me to reject the alternate hypothesis that sensors and intuitors are different in how they answer a question regarding a sampling distribution. The p-value for the 3x2 table is 0.4444.

Evaluating a Z-Statistic: Probability and Range. The questions to follow are on the last two semesters' final exams. First I presented a question regarding the probability that X is in a range, and then I presented a question asking the student to calculate the probability that X is greater than a given value. Since sensors are better at memorizing facts and formulas, I hypothesize that overall, a sensor will perform better on these questions. There are two different formats to this question and I do believe that a sensor might have more trouble with one version over the other, since one is a word problem whereas the other is a straightforward problem. I suspect this only because a sensor prefers to plug in numbers, and the second version of this problem makes that easier to do.

First, I asked the student to calculate the probability that a value X is between a range of two numbers, given a normal distribution. Version I asked the question in a word problem and version II asked the question with straightforward numbers. The first version of the question will involve a story about a wheat field and a random sample of bushels gathered. The second version of the question will be straightforward and present just the facts. Both questions provided the student with the population mean and standard deviation.

> A farmer in western Kansas knows that the amount of wheat on each acre of his farm is normally distributed with a mean of 43 bushels with a standard deviation of 4 bushels. If we select an acre at random from his farm, what is the probability that it yields between 40 and 50 bushels?

The second way this problem was worded is as follows:

> We have a random variable X that has a normal distribution. This normal distribution has a mean of 67 and a standard deviation of 8. What is the probability that the variable X is between 50 and 80?

My hypothesis is that an intuitor would be more likely to perform better on Version I one due to the complex nature, whereas a sensor would be more likely to perform better on Version II due to the straightforward presentation of the numbers, void of confusion with the word problem information. The results are presented in Tables 23, 24 and 25.

Table 23. Spring 2008 - Fall 2008: Likely in Range, Version I

Version I	Intuitor	Sensor
Correct	22 (73.337%)	81 (77.14%)
Incorrect	8 (26.67%)	24 (22.86%)
		p-value 0.6356

Note: Percentages indicate the proportion of students within the ILS category choosing that answer out of total number of students within the category.

This p-value of 0.6356 indicates that there is not a significant difference between the sensor and intuitor in calculating this result, leading me to believe that a sensor is no more likely than an intuitor to solve this version of the problem, and thus I would not reject the null hypothesis in this instance.

Table 24. Spring 2008 - Fall 2008: Likely in Range, Version II

Version II	Intuitor	Sensor
Correct	13 (54.17%)	88 (77.88%)
Incorrect	11 (45.83%)	25 (22.12%)
		p-value 0.0224

Note: Percentages indicate the proportion of students within the ILS category choosing that answer out of total number of students within the category.

This small p-value of 0.0224 indicates that there is a significant difference between the sensor and intuitor in calculating this result, leading me to believe that sensor are more likely than intuitors to solve this problem correctly. I would reject the null hypothesis in favor of the alternative that sensors are more likely than intuitors to correctly answer this question.

It is interesting to see that the sensing student actually performed better on the straightforward version than the intuitor, but in the question regarding the word problem, the two groups performed about the same within the version, however the sensor performed consistently the same between versions, with 77% of the students. This observation supports the claim that states a sensor is more likely to prefer a problem involving the use of a formula and "plug and chug" whereas the intuitor dislikes these types of problems.

To view the overall results between the two groups, to gauge whether this significance still held, I combined both versions of the question and observed significant results with a p-value of 0.0779, as shown in Table 25. This provides a small amount of evidence that sensors are more likely to correctly solve a word problem using a (memorized) equation than an intuitor. I could state that this has some significance at the 0.10 level, but I would attribute this significance to the second version of the test where the sensors performed better than the intuitor.

Table 25. Spring 2008 - Fall 2008: Likely in Range, Combined Versions

Likely in Range, Combined	Intuitor	Sensor
Correct	35 (47.06%)	169 (66.37%)
Incorrect	19 (52.94%)	49 (33.63%)
		p-value 0.0779

Note: Percentages indicate the proportion of students within the ILS category choosing that answer out of total number of students within the category.

The final question presented to the students, asks them to calculate a z-statistic and p-value. The question is presented in two possible ways, but there is no distinguishable difference between the two questions aside that one asks about cow's weight and the other asks about wheat bushel yield.

> A dairy farmer knows that the weights of his mature milk cows follows a normal distribution with a mean weight of 1350 pounds and standard deviation 40 pounds. The farmer is going to select 20 cows at random to be used in a special program. What is the probability that the average weight of these 20 cows will be over 1370 pounds?

The second way this was worded is as follows:

> A farmer in western Kansas knows that the amount of wheat on each acre of his farm is normally distributed with a mean of 38 bushels and a standard deviation of 6 bushels. If we select 20 acres at random, what is the probability that the average yield is over 42 bushels?

Since these versions are not different in how the question was presented, I did not analyze the results within each version, but rather, combined the results.

Table 26. Spring 2008 - Fall 2008: Likely Mean, Combined Versions

Combined Versions	Intuitor	Sensor
Correct	22 (40.74%)	126 (57.80%)
Incorrect	32 (59.26%)	92 (42.20%)
		p-value 0.0321

Note: Percentages indicate the proportion of students within the ILS category choosing that answer out of total number of students within the category.

The Fisher's Exact Test yields a p-value of 0.0321 and is significant at the 0.05 level, indicating that the sensor is more likely than an intuitor to correctly solve this problem, supporting my hypothesis.

Quantitative Analysis Summary. The aim of the quantitative data analysis was to answer the question: Are there any differences between sensors and intuitors in their responses to sampling distributions activities?

My hypothesis was that overall; intuitors will be more likely than sensors to understand the concept of the sampling distribution. I believed that sensors would be better at fact memorization and recalling a formula. My belief was that a sensor would be stronger than an intuitor at solving an equation involving "plug and chug" style, but that the intuitor would be stronger than the sensor at solving complex problems and applying the information given in a word problem, to solving the answer involving the understanding of the sampling distribution of means.

From my presented analysis of the multiple questions above, regarding sampling distributions, I failed to reject the null hypothesis, implying that there is no significant difference between a sensor and intuitor in how they answer question involving sampling distributions in regard to the graphical questions on sampling distributions. I also failed to reject the null hypothesis for questions involving the center, shape and spread of the sampling distribution. There is not sufficient evidence to support the rejection of the null hypothesis.

For the question involving the likelihood of an outcome, there is a small amount of evidence ($p=0.1027$) that the intuitor is more likely than the sensor to correctly answer this question, but it is not conclusive as this is only approaching significance.

Finally, sensors performed better than intuitors on the question involving solving the formula except for one question, where there was no observable difference. This leads me to reject the null hypothesis in two out of three questions, in favor of my alternative that the sensing student would perform better than the intuitor on a question involving plug and chug type calculations. Even though the Version I test was not significant, sensors did still perform slightly better than intuitors.

The only foundational question in which the sensor outperformed the intuitor was the one dealing with the standard deviation of the sampling distribution, possibly because a sensor was relying upon memorization of a formula in answering this question, however it was not statistically significant. Otherwise, the intuitors seemed to outperform the sensors, though statistical tests of significance show that intuitors and sensors are equally likely to answer the questions the same way, with the exception of the significance in the likelihood test already mentioned. This lack of significance only suggests that there is not sufficient evidence to reject.

Analysis Summary

In this chapter I performed quantitative analysis on final exam data. I compared sensors and intuitors in how they answered questions about a sampling distribution in an attempt to answer the question, are there any differences in how sensors and intuitors respond to questions on sampling distributions. I observed significant differences between sensors and intuitors regarding the questions involving the memorization and application of a formula, “Evaluating a Z-statistic”. This was in favor of the sensor being more likely to answer the question correctly.

This conclusion is in agreement with the observations I hypothesized about sensors and intuitors, since a sensor is more likely to memorize a formula as well as more likely to enjoy plugging in numbers whereas an intuitor is less likely to want to perform these rote activities, it makes sense that the sensor performed better on this question.

Secondly, in this Quantitative analysis, I observed nearly significant differences between a sensor and intuitor, in favor of the intuitors, involving the question on whether a small sample or a larger sample is more likely to have a mean greater than the observed population mean. I believe that this is due to the abstract nature of the problem as it involves the conceptualization of the mean of a sample of two different sample sizes and estimating the likelihood of the sample mean. Sensors were less likely to view this abstractly and were more likely to believe that the larger sample would obtain a larger mean. I believe sensors looked at this in a simplistic way of reasoning rather than using abstract reasoning. Intuitors were more likely to understand that a smaller sample is more likely to be skewed and influenced by outliers and that a larger sample was more representative, reasoning abstractly.

I believe that I observed this difference because sensors were able to memorize the definition and idea, but were unable to correctly apply this idea whereas the intuitors were able to correctly apply the concept.

Table 27. Results Analysis Summary A

Research Questions	Data	Framework	Analysis
Are there any differences between sensors and intuitors in their responses to sampling distributions activities?	Multiple choice data from final exam: Graphic based question regarding a distribution of a sample and the distribution of sample means Questions regarding center, shape, spread, and likelihood of a sampling distribution.	Expect to see intuitors more likely to think abstractly and thus more likely to answer questions involving sampling distributions correctly since intuitors are more likely to think abstractly and connect ideas in practice whereas a sensor is more likely to memorize a formula and plug in value to an equation. I expect a sensor to perform better when there is simply a formula to crank out an answer.	Fisher's exact Test

Table 28. Results Analysis Summary B

Semester	Data Collection	Analysis	Participants (#)
Fall 2006	Exam Multiple Choice	Fisher's Exact test on MC questions	103
Spring 2007	Exam Multiple Choice	Fisher's Exact test on MC questions	126
Fall 2007	Exam Multiple Choice	Fisher's Exact test on MC questions	129
Spring 2008	Exam Multiple Choice Short Answer Questions	Fisher's Exact test on MC and short answer questions	180
Fall 2008	Exam Multiple Choice Short Answer Questions	Fisher's Exact test on MC and short answer questions	144

Table 29. Summary of Questions Examined in the Quantitative Analysis and Results

Data Collected	Semester	Significant Results
Graph: Simple Random Sample	Fall 06 - Spring 07	No
Graph: Simple Random Sample	Fall 07 - Fall 08	No
Graph: Sampling Distribution of Means	Fall 06 - Spring 07	No
Graph: Sampling Distribution of Means	Fall 07 - Fall 08	No
Foundation: Likelihood	Fall 07 - Fall 08	*No (p= 0.1027)
Foundation: Center	Fall 07 - Fall 08	No
Foundation: Spread	Fall 07 - Fall 08	No
Foundation: Shape	Fall 07 - Fall 08	No
Z-statistic – Range: Version 1	Spring 08 - Fall 08	No
Range: Version 2	Spring 08 - Fall 08	Yes (p= 0.02)
Z-statistic - Pr(X > xbar): Combined	Spring 08 - Fall 08	Yes (p= 0.03)

*approaching significance at a 0.10 level

Chapter 5: Qualitative Methodology

In order to gather more in depth data about the student and their thought process, I involved mixed methods for my analysis. The nature of Research Question 2 (Are there any differences between sensors and intuitors in how they use logic and reasoning in qualitative responses to sampling distribution questions?) is intended to provide this in depth analysis, highlighting the differences between how sensors and intuitors justify responses for questions involving the sampling distribution of the means.

Due to the underlying ideas that a student must understand in order to pull the concepts of sampling distributions together (Chance, delMas, & Garfield, 2004, p. 300), I did not believe that assessing a student's ability to answer questions regarding a sampling distribution using multiple choice format would be enough to definitively determine if the student truly understands the material. Prior research has already shown that there is a zero to low correlation between course performance and statistical reasoning, such that statistical reasoning is unrelated to students' performance (Tempelaar, Gijselaers, & van der Loeff, 2006).

A larger problem developed, since each statistical concept is indistinguishably linked to other similar concepts, these associations between terms meant that my research had to encompass the assessment of a pre-requisite or foundational knowledge deemed necessary for statistical understanding regarding sampling distributions (Chance, delMas, & Garfield 2004, p 300). I found that the best way to reach this foundational knowledge was through free response questions, allowing the student to justify their reasoning for a multiple choice selection, using their own words, hence the need for qualitative analysis to accompany the quantitative analysis.

I used the responses to these questions to answer the second research question: Are there any differences between sensors and intuitors in how they use logic and reasoning in qualitative responses to sampling distribution questions? I looked at how the student supported their answer. Do students select the correct answer but fail to support it, or do they support an incorrect answer with reasoning that leads us to believe they "know" the information or definition but fail to apply in context?

Overview of Methods Used

I analyzed the results of 3 semesters of Final Exam Questions for ST311 from Fall 2007 to Fall 2008. There are approximately 10 sections taught each semester, by various teachers/graduate students. Each instructor followed the same format/course outline and used the same materials to teach the class; such as hands on classroom activities, applet demonstrations, and other methods aimed at meeting all learning preferences. Participant selection is the same as described in the Quantitative Methodology section.

Interview Selection

I conducted interviews in the first two semesters. At most, three sensors and three intuitors were randomly selected at one time. I then emailed each student and asked them if they wanted to participate in the interview, and offered the incentive of a $10 gift card to the NCSU Campus Bookstore. (See Appendix D).

After a student agreed, we would set up a meeting time to conduct the interview. If a student did not wish to participate, I randomly selected a different student using PROC SURVEYSELEC, and invited them to participate in the interview.

Focus Groups and Interviews

In the Fall 2006 semester, I conducted my first focus groups and a total of six students from the entire population of eligible students participated. In the first group, three sensors participated. Only one student in the group could recall the applet from class. The other two students had never seen the applet before. Unfortunately, after the interview, I discovered that they belonged to the only section out of 10 that did not have technology available in the classroom to allow the instructor to demonstrate the exercise. The second focus group consisted of only one sensor, and the third focus group consisted of two intuitors. Focus group questions were adjusted as the interview progressed to better facilitate gaps in student understanding. I had anticipated observing a deeper understanding out of the group than what the students exhibited.

When students were asked to describe the third graph of the applet (a display of the distribution of sample means) the sensors all commented on the shape, saying that it looks symmetric. When asked for another way to state that, one sensor said “Normal”. In my second focus group, I only had one sensor attend. The dialogue with the student and the facilitator is given in Appendix B.

From these focus groups, I gathered evidence to support the common misconception that if you take the distribution of the sample of means from any distribution (with a significant size) you will get a distribution similar to the parent distribution. The intuitors recalled the applet prior to seeing it, but exhibited no distinct differences from the sensors. The differences between the second sensor (who seemed to grasp the concept fairly well) and the other sensors and intuitors may have been due to intelligence and understanding rather than specific to the student's learning style.

Interviews

In my exploratory study, I conducted a focus group with several students from the same category of the S/N dimension, and detailed in Appendix B. I felt that isolating each student, one at a time, would lend to gathering better information about the thought process and understanding of the individual. The same applet is presented in each semester, but the method and delivery of questions are improved upon based on information gathered from the previous semester.

I conducted interviews to see if I could better understand the students' thought process as they answered graphical questions about a sampling distribution of means, and then interacted with a web-applet. For each interview, the student(s) is presented a question regarding a sampling distribution of means from a given population. The student is then shown a web-applet demonstration of individual sampling leading to a sampling distribution of means, and then asked to answer questions. After the presentation, the student is allowed to use the applet to answer the next two questions, and then the remaining two questions are presented without any help from the online applet.

In the 2006 fall semester Pilot Study, the interview format was that of a Focus Group, where multiple students from the same learning style category participated at once.

I had six students participate all-together in three separate focus groups. The first session consisted of three sensors, the second was one sensor, and the third consisted of two intuitors Conducting a focus group proved to be ineffective since student interaction biased how a student recalled information. I was also not able to observe how each student observed a demonstration, since they all were able to speak at will. This style was not used in subsequent semesters.

For the spring 2007 and fall 2007 semesters, I conducted one-on-one task based interviews. In the spring of 2007, five students participated in the interview: three sensors and two intuitors. In the fall of 2007, I handed the student a pre-assessment questionnaire prior to asking questions about the web applet. This was based on what Chance et al. (2004) consider as foundational knowledge required to understand sampling distributions. Three sensors and five intuitors participated in these one-on-one interviews for the fall 2007 semester.

The main purpose of the interview was to see if I could collect detailed information about how a student views a sampling distribution of means. Secondary to that, I wanted to see if the applet interaction immediately affected the way the student thought through the problem, and if the student was able to learn or retain the information after guiding the interactive tool to answer a few questions. Finally, I wanted to test if the foundational knowledge was actually a precursor to the understanding of sampling distributions, or if misconceptions in any of the four areas foreshadowed their ability, or lack thereof, to answer questions regarding a sampling distribution of means.

Although I had to take care to select from both the sensors and intuitors, I did not have any knowledge of which student belonged to which group prior to the interview. I was able to make my own speculations during the interview, but I did not have this information prior to the interviews for the spring and fall semesters of 2007. Details of the Task Based Interviews are listed in Appendix P.

Data Analysis

In order to analyze how a student uses logic and reasoning about sampling distribution questions, I performed a qualitative analysis on short answer responses from four questions given on the final exam. This qualitative analysis was to determine if the student was able to explain why they chose an answer. Three analysis methods are used and are described in this chapter.

Qualitative Methods

The initial qualitative data analysis came from focus groups and task based interviews. Students were asked questions and videotaped performing activities. I then scored the answers and gave an assessment of the student's rationale based on answers. This was more of a pre-emptive analysis to answer the second research question:

Research Question 2. Are there any differences between sensors and intuitors in how they use logic and reasoning in qualitative responses to sampling distribution questions?

Focus groups evolved in the 1950s from scientists that began looking for an alternative way to conduct job interviews in the late 1930s. Typically, a job interview consisted of a one on one interview with a list of predetermined questions and close ended response choices. A new approach called nondirective interviewing shifts the focus directly to the “reality of the interviewee” with open ended questions that allow unbounded responses with no direction as to what the interview is looking for in a response, eliminating the bias (Krueger & Casey, 2000, p. 3). The “intent of the focus group is to promote self-disclosure among participants” (Krueger & Casey, 2000, p. 4). There are five characteristics of a focus group. They contain people (1) possessing a specific characteristic (2); they provide qualitative data (3) through a focused discussion (4) helping in understanding their viewpoints on the topic of interest (5) (Kruger & Casey, 2000, p. 6).

Goldin suggests a framework consisting of four exploratory stages for each main question. The four stages are:

> "Posing the question (free problem solving), with sufficient time for the (student) to respond and only nondirective follow-up questions, such as, 'Can you tell me more about that?', Minimal heuristic suggestions, if the response is not spontaneous, such as, 'Can you show me using some of these materials?', The guided use of heuristic suggestions, again only when the requested description or anticipated behavior does not occur spontaneously, such as, 'Do you see a pattern in the cards?', and Exploratory, metacognitive questions, such as, 'Do you think you could explain how you thought about the problem?'” (Goldin, 2000, p. 523).

Task based interviews can be used to gain insight into the students' thought process and reasoning. “Structured, task based interviews for the study of mathematical behavior involve minimally a subject (problem solver), and an interviewer (the clinician)”, and some tasks such as activities, problems or questions (Goldin, 2000, p. 519). The goal of these interviews is to analyze "verbal and nonverbal behavior or interactions" in order "to make inferences about mathematical thinking, learning, and/or problem solving of the subjects" (Goldin, 2000, p. 519). As a qualitative assessment tool, the task based interview also has great potential for evaluating whether we are succeeding" (Goldin, 2000, p. 524). Strength of tasked based interview methods is the potential for extension from research to assessment (Goldin, 2000, p. 525).

> "In comparison with conventional, paper and pencil test-based methods, task-based interviews make it possible to focus research attention more directly on the subjects' processes of addressing mathematical tasks, rather than just on the patterns of correct and incorrect answers in the results they produce," making it possible to delve "into a variety of important topics more deeply than is possible by other experimental means" (Goldin, 2000, p. 520).

Interview Based Methods. One way to assess how students think is by conducting focus groups and task based interviews. Focus groups are an example of an approach called nondirective interviewing which shifts the focus directly to the reality of the student with open ended questions that allow unbounded responses and no direction as to what the interview is looking for in a response, eliminating the bias. Task based interviews allow the student to perform tasks while the interviewer observes.

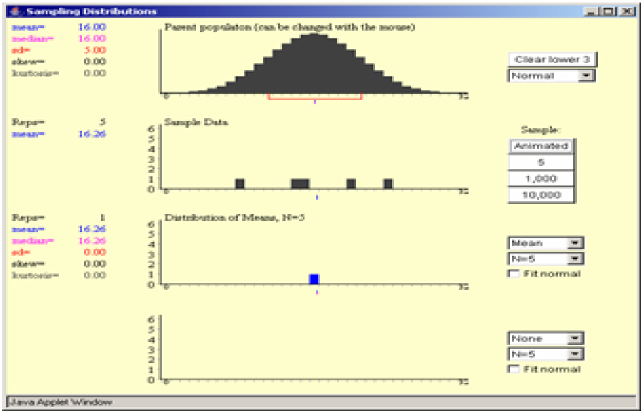
Pilot Study. In the exploratory semester, Fall 2006, I conducted a focus group to assess the students' perception of a sampling distribution applet presented to them in class, for the sensors and for the intuitors. I randomly selected 15 students from the group with strong sensor preferences and 10 students from the group with strong intuitor preferences. I deemed a student to have a strong preference if their score was greater than ±7 for the designated Learning Style. I would have also selected 15 intuitors, but only 13 students consented to the study and met all requirements. To draw the random sample of students, I used PROC SURVEYSELECT from SAS

Of the first group of sensors I had three students agree to participate. In the focus group I asked questions regarding a web applet that the students had previously seen in their classroom lectures (Rice Virtual Lab in Statistics, n.d.).

I first asked participants if they recall the applet from class. The applet is shown again during the focus group since most of the students could not recall much information about it. Comments concerning the applet as well as the statistics course in general were recorded. I also asked questions about their attitudes and opinions of their introductory statistics class and how they feel about the way the material is presented. The invitation and full questions presented are in the appendices. Below is a snapshot of the applet that is presented to each student.

- Think back to the applet that was demonstrated during your class on Wednesday. Does anyone recall what that applet was or what it demonstrated?
- Does anyone recall what this applet looked like?

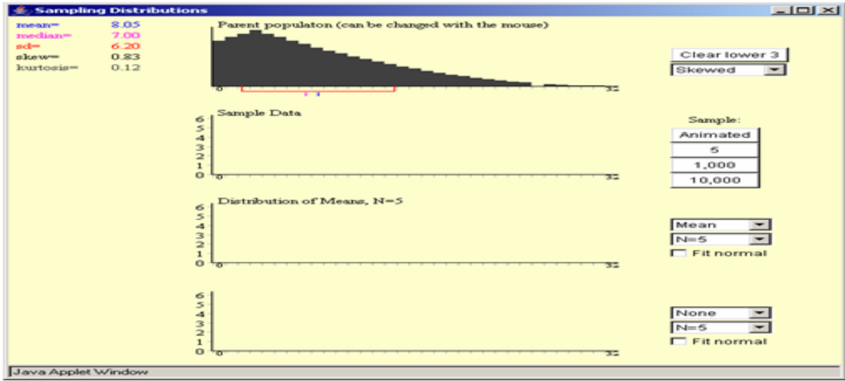
To remind us all of the applet that was being used I will bring up a screen shot on the computer projector.



- In looking at this applet can anyone tell me what the upper graphic represents?
- What does the second graphic represent? (Indicate graph with mouse.)
- And what about the third graphic?
- Does anyone remember what happens when we run the animation multiple times?

Figure 5. Applet and questions asked during interview Part 1.

- If we change the Parent Population to have this type of shape, how would things be different?



- Overall what do you feel was the point of this applet?

Have you covered Confidence intervals in your class yet?"
"How do confidence intervals relate to the ideas we see in this applet?"
"Now some general questions about the course:"
"How do you feel the course is going in general?"
"What things about the course do you feel help you learn the material the most?"
"What things would you change if you could?"
"Is there anything else you would like to tell us?"

Figure 6. Applet and questions asked during interview Part 2.

I used this evidence to form my interview. Since the student interaction in the focus group biased the responses, I moved to task-based interviews, and blinded myself to which students were sensors and which were intuitors.

Task Based Interviews. After conducting a preliminary semester of research, I chose to administer task based interviews rather than focus groups, since I noticed that the student interaction in the focus group biased the responses of other participants within the same group. In the fall semester of 2007, I added to the interview format by performing a pre-assessment prior to the task based interview (Appendix E)**.**

I asked the student to answer several questions covering fundamental concepts based on what Chance et al. (2004), deem as the foundation to the understanding of sampling distributions. These questions aim to quantify the students' understanding of sample, distribution, normal distribution, and population; the four areas essential to statistical inference in (but not limited to) sampling distribution theory.

Students were first presented with a worksheet with pre-assessment questions (see Appendix E). I neither observe the answers nor discuss them with the student. The purpose of the questionnaire is to later observe if the student exhibits the required foundation that Chance et al. attribute to learning about sampling distributions. (Chance et al., 2004, p. 300).

Following a pre-assessment questionnaire (see Appendix E) I asked the student questions about sampling distributions by presenting a histogram of a parent population followed by six smaller histograms. I ask the student to select which histogram best represents a distribution of 500 sample means of sample size 100. Following this question, I have the student participate in hands on computer activities involving a web applet on sampling distributions which I first demonstrate, and then present Question 2. Questions 3 and 4 were to be used along with the applet (the student was allowed/encouraged to use the applet to simulate the scenario that they were presented). Then the final question was presented without any further aid.

The hope in conducting exercises involving the applet is that it will help the students to grasp the ideas and fully connect the concepts together, enabling them to observe their misconceptions and adjust for them through manipulating the sampling distribution of means interactively.

Following this portion of the interview, students are then asked about their statistics course, what they feel works, doesn't work, and general ideas and thoughts regarding the course.

Qualitative Analysis Tools: Grounded Theory, Text Mining and Content Analysis. The goal of the qualitative analysis is to answer the second research question:

Research Question 2. Are there any differences between sensors and intuitors in how they use logic and reasoning in qualitative responses to sampling distribution questions?

All three methods of qualitative analysis are used to analyze 399 responses from each of 4 final exam questions, or 1596 observations. It is my goal to compare these three qualitative methods to one another to see if the same conclusions are met.

Qualitative data analysis techniques are useful to examine the thought process behind the students' answers regarding sampling distribution properties and the relationship to the population, by presenting four questions to the student to assess understanding of center, shape, spread, and likelihood. This data analysis will be used to directly answer the second research question: Are there any differences between sensors and intuitors in how they use logic and reasoning in qualitative responses to sampling distribution questions? I plan to use Grounded Theory, Text Mining, and Content Categorization; the latter two using SAS Software, to perform analysis of the free responses. There are 399 responses given for four questions, yielding nearly 1600 strings of text necessary for analysis.

According to Groth (2010), the two decades prior to the 21st century could "be called the golden age of qualitative research in mathematics education," adding that "the value of qualitative research in statistics education should not be underestimated" as it contributes to the understanding of statistics teaching and learning. (p. 7). Qualitative methods allow educators to discover students' ideas of statistics and discover how they arrive at (sometimes false) beliefs, and even though statistics is essentially a quantitative discipline, it contains a core of qualitative elements that are necessary such as categories and definitions (Petocz & Newberry, 2010, p. 125).

Grounded theory attempts to capture general analysis principles and describes investigative strategies which apply to any social analysis, independent to the classes of data (Atkinson & Delamont, 2006, p. 833). Grounded Theory research not only describes the data, but intends to discover or generate a theory. The key idea is that data from participants are used to "ground" or generate the developing theory. In grounded theory, the views of a large group of participants shape a general explanation of a process, interaction or action, as a qualitative research design (Creswell, 2007, p. 63).

In, *A comparative Study on Feature Selection in Text Categorization*, Yang and Pedersen (1997) describe statistical learning of text categorization as "the problem of automatically assigning predefined categories to free text documents" (Yang & Pedersen, 1997, p. 412). The authors continue to define several methods of categorization, such as document frequency thresholding (DF), information gain (IG), mutual information (MI) Chi Squared statistic (CHI) and term strength (TS).

> "Document Frequency - number of documents in which a term occurs. Simplest technique for vocabulary reduction. Usually considered an Ad-hoc approach b/c of a widely received assumption in information retrieval. That is, low DF terms are assumed to be relatively informative and therefore should not be removed aggressively.
>
> Information Gain - frequently employed as a term-goodness criterion in the field of machine learning. It measures the number of bits of information obtained for category prediction by knowing the presence or absence of a term in a document. (Essentially it is a summation of the product of conditional (probability and log probability).
>
> Mutual Information - criterion commonly used in statistical language modeling of word associations and related applications" (Yang & Pedersen, 1997, p. 413).

Yu, Jannasch-Pennell & DiGangi, (2011) state that Text mining has emerged as a research method and is now considered by researchers to be a viable qualitative research method (p. 730). Like grounded theory, it discourages preconceptions and encourages open-mindedness. Text mining is similar to content analysis. "The criteria of good text mining adhere to those in qualitative research in terms of reliability and validity" (Yu, Jannasch –

Pennell, & DiGangi, 2011, p. 730). Text mining is “defined as process of extracting useful information from document collections through the identification and exploration of interesting patterns” (Yu et al., 2011, p. 731). Grounded theory was developed as a means to conduct data exploration with an open mind, such that the categories can emerge from the data, although it is not possible for a researcher to remain unbiased or to be completely uncontaminated by any preconceived notions (Yu et al., 2011). Often a researcher will “force” the data from certain unconscious prejudices, by imposing implied structures on the phenomenon under study and fitting the data into existing framework. Text mining remedies this by initially using automated algorithms, which restrain the researcher from making premature decisions in the research process (Yu et al., 2011, p. 732).

Similar to grounded theory where the initial categories that are extracted from the data are compared against new data and new categories might emerge, or old ones could collapse and be revised; text mining algorithms are designed to learn from the data by revising the categories as well. Notwithstanding, text mining is viewed as a supplement to human judgment, not a replacement (Kostoff, Morse, & Oncu, 2007, p. 171).

A text miner makes qualitative judgment to set key words, and then decides on constraints. Human judgment is then used to set constraints and keywords and pass the output to pattern recognition. This is an iterative process that continues until an optimal solution is obtained. (Yu et al., 2011, p. 732).

Natural Language processing (NLP) in text mining “can uncover patterns and provide predictive information based on a more sophisticated understanding of language” (Yu et al., 2011, p. 735). NLP is a subfield of computer linguistics and artificial intelligence, using algorithms that focus on automatic analytics of human language. Well written software can address the different parts of speech a word can take on, in an effort to understand the data as though it were read by a human coder (Yu et al., 2011, p. 735).

In text mining, the most crucial form of reliability is replicability - meaning you can achieve the same result from different coders under different circumstances at different points in time. "Text mining is confined to textural analysis whereas the scope of Content Analysis expands to audio and video" (Yu et al., 2011, p. 739). With a well-trained researcher and a well written algorithm, text mining can maintain a high degree of consistency. Text mining allows categories to emerge from the data.

In order to review the responses for over 300 students on four final exam questions, I plan to use SAS Enterprise Text Miner and Content Categorization, two separate packages that might prove useful in analyzing these types of free response data

> "SAS Text Miner discovers information buried in document collections. By automatically reading text and delivering algorithms for rigorous, advanced analyses, the solution makes it possible to grasp future trends and act on new opportunities more precisely and with less risk. It includes advanced linguistic capabilities within the core data mining solution of SAS® Enterprise Miner™ so you can easily extend text mining insights into structured data mining and predictive analysis" (SAS Institute, Inc., 2013).

Grounded Theory. For the grounded theory analysis, I carefully read each of the responses, and then summarized them into groups. This took several iterations of data parsing. First I grouped the responses by writing them into one of many groups, and then consolidated the groups into 10 broad topics to encompass all of the responses. The groups were then summarized as one of ten categories. Common misconceptions, as noted in the previous section, were well represented in the data.

The advantage of using grounded theory in qualitative data analysis is that I was able to ascertain “correct reasoning” from the data. I could interpret what the student was trying to say despite misspelling and the misuse of words (sometimes the use of words was not precise but I was able to infer the concept based on the answer selection along with the verbiage.) A big disadvantage of using grounded theory in analyzing these responses is that it is quite time consuming, and not ideal for more than 500 observations. I had 399 observations for four questions, giving me nearly 1600 questions to categorize. This process took several weeks to complete. I did this one question at a time.

For each question, I recorded the student’s response and placed it into a general category. The categories emerge through the data based on what the student is talking about. I then took the generalized categories and grouped them into 10 broad categories to encompass the underlying meaning of the response, and allowed only 10 categories to emerge. Each question contained a “need more information” category, which I feel is significant in that students believe that, despite having all the information, if they just knew more about the population total or sample taken, they would have the answer, being unable to draw inference from the information given, despite having all the necessary information and tools needed to draw specific conclusions. It was important to carefully read each response, and sometimes interpret to the best of my ability what the student was trying to convey.

Text Mining. After using a grounded theory approach, I sought after more sophisticated computing methods to test whether or not I could use advanced computer language to create categories from the student's responses, with the goal in mind to answer the research question; Are there any differences between sensors and intuitors in how they use logic and reasoning in qualitative responses to sampling distribution questions?

The goal was to see if the software would produce similar categories to what I created by hand, since SAS Text miner uses a modified grounded theory approach. I am not an expert in the use of SAS Text Miner, and thus a more sophisticated use might be applied given a greater level of experience. Text Mining "mines" the data and creates "Topic" categories, using keywords taken from the document. I use the words category and topic interchangeably.

For this analysis, I input the student's responses, one question at a time, and let the software process the topics. The software provided categories consisting of keywords, which were weighted within each category. Each category is given a "cutoff" score such that if an observation falls below this value, it will not fall into the topic. This cutoff and the keyword weights are used to determine if an observation contains enough weight to fall into the specified category. Each observation is given a raw score based on the keywords and overall weight. This raw score is measured against the cutoff value to determine if the observation belongs within the category. If the raw score of the observation is less than the cutoff value, the observation is not categorized into the topic. Likewise, some words are categorized into more than one Topic. When this occurs, I mathematically assigned the observation to the topic holding the highest difference between raw score and cutoff score.

For example, suppose topic 1 has cutoff 1.5 and topic 2 has cutoff 0.5. If observation A has a raw score of 1.6 for topic 1 and 1.2 for topic 2, the observation would be assigned to topic 2 because it has a stronger score for the category. Here is the math:

$$(1.6-1.5) = \text{Difference } 1 = 0.10 < 0.7 = \text{Difference } 2 = (1.2 - 0.5)$$

I performed this calculation using a SAS data step for every observation, on each of the 10 raw scores assigned. Some observations were not assigned to any categories and were placed into a category appropriately called “NO CATEGORY”.

After doing an initial pass of the data for Center and Spread, I took the data and created categories for each group (one for the sensor and one for the intuitor) to see if there were any major differences. This made sense since the sensors and intuitors might have used different terminology, but creating the Topics together confounded this information. By separating the groups (sensor and intuitor) away from each other, I was able to determine if different topics emerged as well. These are the results that I have displayed below for each of the four questions.

Content Analysis. Content Analysis also called Contextual Analysis is a rule based qualitative analysis method in which the user defines rules and the observations are assigned into categories based on the rules. This research method is a well-known method to analyze documents, allowing the researcher to filter text into categories in which different words and phrases share the same meanings. Content analysis is useful for making “replicable and valid inferences from data to their context, with the purpose of providing knowledge, new insights, a representation of facts and a practical guide to action” (Elo & Kyngäs, 2008, p. 108). Content analysis is a very simple technique that does not offer detailed statistical analysis, but for the purposes of this research, the categorizing of responses will prove beneficial to comparing the groups.

Common strategies of content analysis include examination of the data and making notes, condensing the data in order to answer the research question, organizing the condensed data in order to identify themes or patterns, and then verify, test and confirm the patterns and themes. Grounded theory, content analysis and text mining share common goals as it is data-driven research allowing categories to emerge from the data, where the latter two utilize computer algorithms. (Yu et al., 739)

In my work, categories are created based on rules. Rules are created, first based on information found in the literature regarding misconceptions as well as what I expect to see from a correct answer. Variations of misconceptions and correct answers emerge from the grounded theory as well as iteratively processing the data with the initial rules in order to capture as much data as possible. Text that contains certain words or phrases indicates that the document falls into a specific category. Observations are then categorized based on the rules and a final comparison between the groups is conducted.

For this method of qualitative analysis, I used SAS Content Categorization, soon to be marketed as Contextual Analytics in SAS Text Analytics solutions. "SAS Content Categorization enables you to define a taxonomy of categories and concepts that develops and identifies metadata about your information." (SAS Institute, Inc., 2012). It allows you to write rules for categories that are used to classify data into groups. These groups will then be useful in comparison of intuitors and sensors. The references in the following section will use the terminology Content Categorization which is synonymous with Content Analysis.

Final Exam Questions. The questions are "multiple choice" format and I ask the student to provide justification for each selection to help me to evaluate the thought process for the student making the selection. This should reduce the noise of random answer selection. I also plan to compare the answers across a subject to check for consistency in understanding.

I believe it is likely to observe correct answers in which the student fails to support the answer choice logically. This requirement for answer justification should lend to a more accurate and precise analysis of the students reasoning and thought process. Since it has been shown that students are able to perform well in a statistics course without exhibiting statistical reasoning, I aim to verify this within the four questions I have asked. I hope to be able to decisively declare whether or not the student exhibits a sound understanding of the concepts being presented.

Since there are multiple aspects underlying sampling distribution theory; the questions to follow are additional measures of the students' understanding rather than specific questions regarding the sampling distribution of means. Concepts of shape, center, spread, as well as likelihood and variability among sample sizes are addressed in the questions presented.

These final exam questions delve further into the understanding of the student and assess their proficiency in these four areas. I plan to use this analysis to answer the second research question: Are there any differences between sensors and intuitors in how they use logic and reasoning in qualitative responses to sampling distribution questions?

My goal is to measure any relationship between the way the student selects a distribution of sample means and the way they perceive the variability or shape of the population and sampling distribution, individually as well as in relation from one to the other. Students are asked the following questions and then asked to justify the reason for the selection they made. This is a significant addition to our research because it enables us to evaluate correct reasoning along with the correct answer selection, reducing the possibility of random selection.

Foundational Knowledge: Likelihood, Center, Spread, and Shape. The first portion of this assessment presents three questions based on this problem:

Likelihood

According to a recent report the average income of people in Adair County, Missouri is $31,023. Which of the following is more likely? Pick one and explain.

a. We take a random sample of 10 people from this county and find that the average is over $50,000.
b. We take a random sample of 200 people from this county and find that the average is over $50,000.
c. We do not have enough information to answer this question.

A large multinational company knows that the average age of their employees is 34 years. They also know that the standard deviation of the ages of these employees is 8 years. A manager from human resources is going to randomly select a sample of 100 employees. Consider the sampling distribution of the mean of a sample of size 100 from this population. From the information given, which of the following statements are true about this sampling distribution?

The questions from this problem aim at understanding how students perceive the center, shape, and variability of the sampling distribution in comparison with the parent population. The student is encouraged to explain their choices and is penalized if they do not.

Center

Which of the following is true? Pick one and explain.

a. The sampling distribution will be centered at 34.
b. We do not know where the sampling distribution will be centered.

Spread

Which of the following is true? Pick one and explain.

a. The sampling distribution will have a higher standard deviation than the population.
b. The sampling distribution will have the same standard deviation as the population.
c. The sampling distribution will have a smaller standard deviation than the population.

Shape

Which of the following is true? Pick one and explain.

a. We know that the shape of the sampling distribution of the mean will be right skewed.
b. We know that the shape of the sampling distribution of the mean will be approximately symmetric.
c. We cannot tell what the shape sampling distribution of the mean will look like

Additionally a question of likelihood is presented. This question requires an understanding that a smaller sample mean is less likely to be representative of the population parameter.

Ethical considerations

In order to comply with standards of the IRB, I had all students who wished to participate, read and sign the CONSENT document, Appendix A. Students were allowed to opt out by checking, “I do not consent”. These students were not included in any of our data collection, including the ILS survey, focus group interviews, and final exam questions.

Timeline

Fall 2006 – PILOT STUDY/ Focus Groups

Spring 2007 - PILOT STUDY/Task Based Interviews

Fall 2007 - Data Collection

Spring 2008 - Data Collection

Fall 2008 - Data Collection

Table 30. Timeline of data collection, analysis, and participant count

Semester	Data Collection	Analysis	Participants (#)
Fall 2006	Focus Group Pilot	Rudimentary analysis of Focus Group recording and notes	6
Spring 2007	Task Based Interviews	Analysis of Task Based Interviews	13
Fall 2007	Free Response Questions	Grounded Theory, Content Categorization and Text Mining on Free Response Questions	129
Spring 2008	Task Based Interviews Free Response Questions	Grounded Theory, Content Categorization and Text Mining on Free Response Questions	180
Fall 2008	Task Based Interviews Free Response Questions	Grounded Theory, Content Categorization and Text Mining on Free Response Questions	144
January 2009 - June 2009	NONE	Data Analysis and writing a first draft	ALL

Summary

In this chapter, I described the qualitative research methodology I employed in my research. Three semesters of introductory statistics classes were taught between 2007 and 2008 at NCSU, using the same syllabus and teaching materials and objectives. I collected ILS survey data and written responses to statistical reasoning questions that accompanied final exam questions. Interviews were conducted with 13 students from the semesters in 2007. This chapter itemized these different forms of data and explained the rationale behind collecting each form of data.

Chapter 6: Qualitative Results

The second research question and hypotheses are:

Research Question 2.

Are there any differences between sensors and intuitors in how they use logic and reasoning in qualitative responses to sampling distribution questions?

Hypothesis. I believe that a sensor will be more likely to provide a good rationalization but fail to apply it to answering a question, meaning, the sensor might be able to give a definition or formula, but will be unable to apply it in context. I believe that an intuitor will be more likely to apply definitions in context, but may be less likely to give definitions verbatim, and might be more likely to give justifications in their own terms.

Evaluating these answers is quite complex and will involve the evaluation of the answer to a question about the sampling distribution using mixed methods; both quantitative and qualitative data analysis. The quantitative was categorized as either correct or incorrect, and then within each group, I looked at how the student supported their answer using qualitative analysis.

I performed qualitative analysis on the four questions that were written to assess foundational knowledge of sampling distributions and the responses given by students. I looked at this information and compared it to how the student is answering the question about the distribution of sample means. Did the student select the graph consistently with how they are solving the question about the location of the distribution of mean center, shape, and spread? Or is there a lack of consistency in this approach? Likewise, if the student answers all four questions correctly, I would expect them to exhibit strong knowledge in sampling distribution theory and thus be able to correctly solve the graphic questions.

Data Analysis. I administered multiple choice questions with a free response section, requiring the students to justify their multiple choice selection for each of four questions. Each question is related to foundational ideas of sampling distributions, namely: center, shape, spread, and likelihood as presented on pages 95-96 .

Final exams were collected, and the students who fit into our criterion as previously discussed were selected. Responses were transcribed along with answers to the other questions of interest. This data was carefully entered into a database for qualitative analysis. Final exams administered in the Fall of 2007, Spring 2008, and Fall 2008 semesters included these four free response questions analyzed in this section.

In this section, I detailed the qualitative analysis results by first describing the grounded theory approach to categorizing and analyzing the data. Second I discuss a text mining approach using SAS® Enterprise Miner 12.1 and the text mining application. Third, I discuss a categorization approach using SAS® Content Categorization Studio, and finally, I compared the methods, presenting both the strengths and weaknesses of each and gave details on the results in a side by side comparison, when possible.

Grounded Theory.

In my first approach to the qualitative data analysis, I used grounded theory and allowed the categories to emerge from the responses. The following sections detail the categories that emerged for each question, Likelihood, Center, Spread, and Shape; as well as the statistical representation of the data for both sensors and intuitors. I finally compared the categories of sensor versus intuitor for statistical significance, and decide based on the results whether or not one dimension was more likely to use correct reasoning than the other.

Likelihood.

First I detail the following question on the likelihood of an outcome.

> According to a recent report the average income of people in Adair County, Missouri is $31,023. Which of the following is more likely? Pick one and explain.
>
> a. We take a random sample of 10 people from this county and find that the average is over $50,000.
>
> b. We take a random sample of 200 people from this county and find that the average is over $50,000.
>
> c. We do not have enough information to answer this question.

Overall Results. The following 10 categories emerged from the data of responses from the students, listed in Table 31. The categories preceded with a double asterisk indicate a correct justification to the answer. A correct answer and correct justification is indicated by a single asterisk and listed in parentheses if it does not correspond with the count for the response of the entire group.

Table 31. Grounded Theory, Likelihood: Categories

Smaller samples are less representative of the population, more affected by outliers and thus more likely to have a mean greater than the population mean.	**Intuitor**	**Sensor**	**Total**
1: With a larger sample, variability increases so the spread is greater	2 (2.5%)	7 (2.19%)	9 (2.26%)
2: The mean increases in a larger sample	0	4 (1.25%)	3 (0.7%)
**3: We need a larger sample to represent the entire population	3 (*0) (3.75%)	22 (*2) (6.58%)	24 (*2) (6.02%)
**4: A small sample is not representative of the population	*9 (11.25%)	*30 (9.40%)	*40 (10.03%)
5: Larger samples make the outcome more likely	3 (3.75%)	24 (7.52%)	27 (6.77%)
**6: Using the CLT, as n gets larger, the sampling distribution approaches a normal distribution with mean: μ and std: σ	*1 (1.25%)	7 (*6) (2.19%)	9 (*7) (2.26%)
**7: Larger values of n put you closer to the actual mean, so the smaller n is more likely to yield this result	*19 (23.75%)	*60 (18.81%)	*79 (19.79%)
**8: A large sample has less variability and a small sample has more variability, and larger margin of error	*5 (6.25%)	20 (*18) (6.27%)	25 (*23) (6.27%)
9: Need more information in order to solve this	27 (33.75%)	115 (36.05%)	142 (35.59%)
**10: Small samples are skewed and thus not representative of μ. Outliers will affect the sample mean more in a smaller sample	*9 (11.25%)	25 (*24) (7.84%)	34 (*33) (8.52%)
NO RESPONSE	2 (2.5%)	5 (1.57%)	7 (1.75%)
TOTAL	80	319	399

Note: * Correct Answer AND Correct Reasoning. ** Correct Reasoning. Correct Reasoning responses are highlighted. The final column represents percentage of the entire population. Numbers in Parentheses with an asterisk represent the count of correct reasoning AND correct answer selection. If the response is ** as a correct reason, but the * count is not in parentheses, then all students in this instance have the correct answer and correct reasoning. The number inside the parentheses below the counts represents the percentage of the corresponding population.

Example: Response - Three intuitors gave this reason, and none of them had the correct answer.

This number is 3.75% of all intuitors

21 sensors gave this reason and 2 of them also had a correct answer selection.

This number is 6.58% of all sensors.

Table 31 details the distribution of correct reasoning and responses for intuitors and sensors. More intuitors than sensors answered this question correctly and gave a correct reasoning for the answer selection. Table 32 illustrates the distribution of responses by intuitors and sensors and highlights the proportions with both a correct answer and a correct justification.

Table 32. Grounded Theory, Likelihood: Answer and Reasoning

Likelihood	Intuitors	Sensors
A*	50 (63.29%)	165 (51.72%)
correct reasoning **	44 (55.00%)	140 (43.88%)
B	8 (10.13%)	55 (17.24%)
C	21 (26.58%)	99 (31.03%)
		p-value**= 0.0801

Note: *Correct Answer. ** Correct answer AND correct Reasoning. Percentages indicate the proportion of students within the ILS category choosing that answer out of total number of students within the category.

The distribution of responses between sensors and intuitors is similar as indicated by the percentages below the counts in Table 32. I also calculated p-values using Fisher's Exact Test on two separate items. First I calculated the probability that an intuitor was to make a correct answer selection regardless of correct reasoning more often than a sensor, as done in the qualitative section. The p-value for this comparison is 0.1027.

Next, I performed a Fisher's exact test on the correct answer / correct reasoning combination between sensors and intuitors. The p-value for this is also 0.0801. The latter p-value is significant at the 0.10 level, indicating we should reject H_o, in favor of intuitors being more likely than sensors to correctly reason and answer about this question involving likelihood.

I believe that this evidence is adequate to support that intuitors and sensors do reason about this problem differently, and in the least, intuitors are more likely than sensors to correctly answer this problem and justify that answer, even if they are only slightly more likely to do so.

Correct Reasoning. Correct reasoning does not always indicate that the student selected the correct answer. Many students responded similar to "We need a larger sample to represent the entire population", but chose "b" "We take a random sample of 200 people from this county and find that the average is over $50,000," as their (incorrect) answer. Thus students' open responses indicated that they felt they needed a larger sample size in order to represent the population and that without a larger sample, they would not achieve μ. It was apparent that while the rationalization was absolutely correct, the student completely disregarded the fact that the population mean was only $31,023 and the question was asking how likely it would be to achieve a mean MUCH GREATER than the true mean. Two sensors using this justification actually chose the correct answer. Sensors tend to be good at fact memorization and are patient with details. Despite these characteristics, only two were able to make this connection.

Table 33. Grounded Theory, Likelihood: Correct Reasoning Categories, Subset

Response 3: We need a larger sample to represent the entire population	LEARNING STYLE
• N=10, no enough to prove something, when working with a normal distribution, we need at least 30 from n=10 we cannot draw a conclusion • Bigger pop size the better the chances of the approximate average being right • We would need to take a sample of 200 b/c the other option was too small a sample • * Larger sample more likely to represent the county • * The more you sample the closer the average gets to the average of the population	Sensor Sensor Intuitor Intuitor Intuitor
Response 4: A small sample is not representative of the population	
• Smaller sample has more likely to have greater mean • Smaller sample is more likely to misrepresent the population • * Smaller sample more likely to have higher mean than population	Intuitor Intuitor Sensor
Response 6: Using the CLT, as n gets larger, the sampling distribution approaches a normal distribution with mean: μ and std: σ	
• As the sample size gets larger the mean becomes closer to the actual value and to a normal distribution • Larger sample size means more likely to reflect mean of the whole population • * As n increases, so does z and as z increases, probability/ likeliness of having income 50k of n people	Intuitor Sensor Sensor
Response 7: Larger values of n put you closer to the actual mean, so the smaller n is more likely to yield this result	
• Smaller group more likely • * In order for the average to be below 50K, there must only be a few people with an income over 50K. If the average of 200 people was over 50K then the report would have been wrong • * Less sampled, making it more likely that sample mean would be a lot greater than the real mean	Sensor Sensor Intuitor
Response 8: A large sample has less variability and a small sample has more variability, and larger margin of error	
• Different results would occur due to a sample size that is too small • Smaller sample have greater STANDARD DEVIATION • Margin of error is reduced with a larger sample size, b/c of the equation as well as df for the t-value. It’s more likely to get extreme data with smaller sample sizes • * With smaller sample, the deviation is more likely to be great enough to allow that possibility	Intuitor Sensor Intuitor Sensor
Response 10: Small samples are skewed and thus not representative of μ. Outliers will affect the sample mean more in a smaller sample	
• Less in sample, more chance to be skewed • More likely to get with small sample because chance already low. This random sample could be more different b/c the number of persons picked is so small it could be skewed • * Smaller sample and could be a rich area	Sensor Sensor Intuitor

Note: A correct answer (a) coupled with the correct reasoning is preceded by an asterisk (*).

An additional hypothesis that I tested: Are sensors more likely to use correct reasoning but choose an incorrect answer? This is due to the sensors nature of fact memorization, but inability to think abstractly, meaning, the sensor is less likely than an intuitor to apply a definition to answer selection. In order to test this, I only created a 2x2 table of Incorrect Response and Correct Reasoning versus the other three combinations namely: Incorrect Response and Incorrect Reasoning, Correct Response and Correct Reasoning, and Correct Response and Incorrect Reasoning, called “Other Combinations”.

Table 34. Grounded Theory, Likelihood: Correct Reasoning and Incorrect Response

Reasoning and Response	Intuitors	Sensors
Correct Reason with Incorrect Answer	3 (3.75%)	24 (7.52%)
Other Combinations	77 (96.25%)	295 (92.48%)
		p-value= 0.3207

From this resulting p-value of 0.3207, we do not have enough evidence to reject the null hypothesis in favor of sensors and intuitors being equally likely to provide a correct response and make an incorrect answer about the likelihood of an outcome.

Misconceptions. Misconceptions from literature that emerged in the data, come from Garfield and Chance (2000) p. 118.

Law of Small Numbers (small samples best represent the population so are preferred) This misconception is represented with response 1, 2, and 5. Students state that a large sample has a larger mean, a larger variance, and are more likely to achieve an outcome that seems improbable based on the information about the population. Five (6.25%) intuitors fell into the misconception categories, whereas 34 (10.67%) sensors exhibited these misconceptions.

I generated another 2x2 table for a Fisher's exact test to test if there is a significant difference between sensors and intuitors in exhibiting misconceptions in their reasoning.

Table 35. Grounded Theory, Likelihood: Misconceptions.

Misconception Reasoning	Intuitors	Sensors
Misconception	5 (6.25%)	34 (10.66%)
Other	75 (93.75%)	285 (79.34%)
		p-value= 0.2954

This test indicates that we would fail to reject the null hypothesis such that there is not enough evidence to suggest that sensors are more likely than intuitors to exhibit misconceptions in their reasoning about the likelihood of an outcome.

Table 36. Grounded Theory, Likelihood: Misconception Categories, Subset

Response 1: With a larger sample, variability increases so the spread is greater	Learning Style
• Income n=200 most likely vary more than n=100, so more likely average over 50K • With fewer people sampled, the SAMPLING DISTRIBUTION will be smaller b/c the sample itself is still small, as more people are surveyed, the variance increases and the average will be closer to the true average of the county • More sampled more likely to get higher average since spread will be greater as n increases • If we take this option we have a greater chance at choosing a width spread of income levels and thus having a greater probability of proving this correct	Sensor Sensor Intuitor Intuitor
Response 2: The mean increases in a larger sample	
• As your sample size increase, the mean will also increase. N=10 not reliable • If you have a larger sample it will help the average to go higher • Would need more samples to be so big so n=200	Sensor Sensor Sensor
Response 5: Larger samples make the outcome more likely	
• We have a larger sample to fulfill the expectations that the average income is high: the sample size allows us to cover more data • Bigger sample size more accurately reports average income • The more taken for the sample the better chance there will be for the average to be over 50K • The bigger the sample size you have the closer you can get to true proportion	Intuitor Sensor Sensor Intuitor

More Information Needed. The only remaining response was that “more information was required in order to make a conclusion”, of which 27 (33.75%) intuitors and 115 (36.05%) sensors used as a reason for their answer

Table 37. Grounded Theory, Likelihood: More Information Needed, Subset

Response 9: Need more information in order to solve this	Learning Style
• Only know the average not the range. Need more info • We don't' know population total so need more info • We don't' know anything about Adair county except that one #, 31K. We need more information • Because we don't know the standard deviation of this data • A sample size or even pop. Is unknown thus it is uncertain if the average is 31K	Intuitor Sensor Intuitor Sensor Intuitor

Some of the students stated they needed to know the range of the population, while others wanted to know the standard deviation. Still others want to know how many are in the population of interest. The fact that they had the population mean and were asked which sample was least likely to estimate the population mean escaped them. Overall, the vast majority of the students stated that they needed to know more information in order to solve this problem, and this was not limited to either intuitor or sensor, but to the population as a whole, 33.75% of intuitors and 36.05% of sensors stated they needed more information in order to solve this problem.

If this trend continues throughout the remaining discussion, it would be wise to make note that a new "misconception" might be emerging, such that students believe they need more information in order to answer a problem rather than draw inference from the details given.

Center.

A large multinational company knows that the average age of their employees is 34 years. They also know that the standard deviation of the ages of these employees is 8 years. A manager from human resources is going to randomly select a sample of 100 employees. Consider the sampling distribution of the mean of a sample of size 100 from this population. From the information given, which of the following statements are true about this sampling distribution?

Which of the following is true? Pick one and explain.

a. The sampling distribution will be centered at 34.
b. We do not know where the sampling distribution will be centered.

Overall Results. The following 10 categories emerged from the data of responses from the students, listed in Table 38. The categories preceded with a double asterisk indicate a correct justification to the answer. A correct answer and correct justification is indicated by a single asterisk and listed in parentheses if it does not correspond with the count for the response of the entire group.

The following 10 categories emerged from the data of responses from the students regarding the location of the center of the sampling distribution. The majority of both intuitors and sensors stated that the center is at the mean or average. The distribution of the remaining response categories is relatively the same between the groups, as shown in Table 38 below.

Table 38. Grounded Theory, Center: Categories

The center of the sampling distribution will be at the mean of the population	Intuitor	Sensor	Total
**1: Center at the population mean / μ	*21 (26.25%)	96 (*95) (30.09%)	117 (*116) (29.32%)
**2: Larger sample is a representative of the entire population so we know the mean	*7 (8.75%)	30 (*29) (9.40%)	37 (*36) (9.27%)
3: The center is at μ because it is the median	0 (0.00%)	2 (0.63%)	2 (0.50%)
**4: The sampling distribution is NORMAL and therefore it is centered at μ since n>30. Central Limit Theorem	*14 (17.50%)	*39 (12.23%)	*53 (13.28%)
5: This is a small sample so the center could differ.	3 (3.75%)	9 (2.82%)	12 (3.01%)
6: Outliers will make it skewed, so we don't know	5 (4.55%)	17 (5.33%)	22 (5.51%)
7: We need to know the median to know the center	2 (2.50%)	11 (3.45%)	13 (3.26%)
8: We do not have enough information. The sample average is unknown. We need the population distribution. This is random and we don't know if it is normal.	8 (10.0%)	35 (10.97%)	43 (10.78%)
9: The sample mean is not always equal to μ.	7 (8.75%)	27 (8.46%)	34 (8.52%)
10: This is random and could vary with each sample size. We aren't sure how close to the population. It might change	3 (3.75%)	26 (8.15%)	29 (7.27%)
No Response	10 (12.50%)	27 (8.46%)	37 (9.27%)

Note: * Correct Answer AND Correct Reasoning. ** Correct Reasoning. Correct Reasoning responses are highlighted. *The final column represents proportions of the entire population.* Numbers in Parentheses with an asterisk represent the count of correct reasoning and correct answer selection. If the response is ** as a correct reason, but the * count is not in parentheses, then all students in this instance have the correct answer and correct reasoning. The number inside parentheses below the counts represent percentages of the corresponding population

Example:

Response 1. 21 intuitors gave this reason, and all of them had the correct answer.
This number is 26.25% of all intuitors
96 sensors gave this reason and 95 of them had the correct answer.
This number is 30.09% of all sensors.

Table 38 details the distribution of correct reasoning and responses for intuitors and sensors. More intuitors than sensors answered this question correctly and gave a correct reasoning for the answer selection. Table 39 illustrates the distribution of responses by intuitors and sensors and highlights the proportions with both a correct answer and a correct justification for the answer. There is no significant difference between intuitors and sensors, and the p-value is 0.9006.

Table 39. Grounded Theory, Center: Answer and Reasoning

Center	Intuitors	Sensors
A*	59 (73.75%)	219 (68.65%)
correct reasoning**	42 (52.50%**)	163 (51.10%**)
B	21 (26.25%)	100 (31.35%)
		p-values**: 0.9006

Note: *Correct Answer. ** Correct answer AND correct Reasoning. Percentages indicate the proportion of students within the ILS category choosing that answer out of total number of students within the category.

In my opinion, this was the easiest question from the exam used in my analysis with around 70% of both sensors and intuitors choosing the correct answer. Only 51% of the sensors used correct reasoning to justify their answer versus 50% of the intuitors, which clearly did not have any significance using a Fisher's Exact Test with a p-value of 0.9.

Correct Reasoning. Table 40 lists the correct reasoning responses. The first mentions the average or the mean and states that this is where the center is located. This is a correct reason, though lacks a deep thought or line of reasoning. The second response mentions that a larger sample is representative of the entire population and therefore we know the mean. I had to take into consideration that this question is talking about the sampling distribution being of a large size and not just a simple random sample, but the truth is that I cannot be certain that the student isn't confusing sample with sampling distribution, which is a prime example of the lexical ambiguity and morass of terminology getting confounded in the student's mind. I don't believe this to be a very strong correct reasoning, but it is (somewhat correct). Finally the fourth correct response mentions the Central Limit theorem or uses it to define and explain the justification for the problem. I would consider this the best response, and believe that the student has a good understanding of sampling distributions.

Table 40. Grounded Theory, Center: Correct Reasoning Categories, Subset

Response 1: Center at the population mean / μ	Learning Style
• it is centered around the average • centered at 34 because it is the average • because it is the average • This is the mean where the sampling distribution is centered	Sensor Sensor Intuitor Intuitor
Response 2: Larger sample is a representative of the entire population so we know the mean	
• Many samples from the population the mean will center around the true mean • The Sampling Distribution is representative of the population • chose appropriate sample size thus they will be equal • the center will be 34 because that is the average age of the employees so the Sampling distribution should be an accurate representation with 8+/- from the SAMPLING DISTRIBUTION of the mean	Sensor Sensor Intuitor Intuitor
Response 4: The sampling distribution is NORMAL and therefore it is centered at μ since n>30. Central Limit Theorem	
• because statistics have a predictable pattern and inferences can be made on just a few datasets or information • SAMPLING DISTRIBUTION will follow normal distribution patterns • The sampling distribution will be normally centered around the actual population parameter, • sample size > 30 so normal properties apply and Sampling Distribution is centered at 34,	Sensor Sensor Intuitor Intuitor

The distribution of responses between sensors and intuitors seems pretty similar as indicated by the percentages below the counts in Table 38. No big differences can be observed between how a sensor and how an intuitor justified an answer as the distribution of categories between groups is fairly even. It does seem that more sensors than intuitors are looking for more information about how close the sampling distribution is to the population (response 10) with 8% sensors versus only 3% intuitors having this response, but this is not significantly different.

An additional hypothesis that I tested: Are sensors more likely to use correct reasoning but choose an incorrect answer? This is due to the sensors nature of fact memorization, but inability to think abstractly, meaning, the sensor is less likely than an intuitor to apply a definition to answer selection. In order to test this, I only created a 2x2 table of Incorrect Response and Correct Reasoning versus the other three combinations namely: Incorrect Response and Incorrect Reasoning, Correct Response and Correct Reasoning, and Correct Response and Incorrect Reasoning, called “Other Combinations”.

Table 41. Grounded Theory, Center: Correct Reasoning and Incorrect Response.

Reasoning and Response	Intuitors	Sensors
Correct Reason with Incorrect Answer	0	2 (0.63%)
Other Combinations	80 (100%)	317 (99.37%)
		p-value= 1

From this resulting p-value of 1, we do not have enough evidence to reject the null hypothesis in favor of sensors and intuitors being equally likely to provide a correct response and make an incorrect answer.

Misconceptions. Misconceptions involving averages (common number, median vs. mean, outliers) (Garfield & Chance, 2000, p. 118).

Table 42. Grounded Theory, Center: Misconceptions of a Sample: Subset

Response 3: The center is at μ because it is the median	Learning Style
• Center is at 34 because it is the median • Because it's the median	Sensor Sensor
Response 7: We need to know the median to know the center	
• Not enough info. Need median • Population not normal so mean ne median so not certain of center • Average is just 34, not the median • Do not know center of distribution because we do not know median of the sample	Sensor Sensor Intuitor Intuitor

Only two students gave this response and both responses stated the center is at 34 because it is the median. We could argue that this is correct, since the sampling distribution is normal, the mean and the median will be the same, but this only applies to the sampling distribution of means. The median of the population might fall elsewhere.

Response 7 falls along the same thought regarding the median, stating that, we need to know the median to know the center. This indicates that they are not sure what the median is, but once they know, they will know where the center lies, disregarding the fact that they know the population mean, and the sampling distribution of means follows a predictable pattern.

Only two (2.5%) of the intuitors used either of these misconceptions, whereas 11 (4.1%) of the sensors did. A difference that is not statistically significant.

More misconceptions emerged from the data. Taken from Chance, delMas and Garfield (2004, p. 302).

- Do not understand that a sampling distribution is a distribution of sample statistics.

- Confuse one sample (real data) with all possible samples (in distribution) or potential samples.

The listed misconceptions above are a reoccurring theme I often see in the answers that students give. For example, a student discussing the size of the sample in Response 5, lead me to believe they are thinking of the sampling distribution of means as a simple random sample and not a distribution of all possible means. Response 6 involves outliers and skewness, which alludes to the fact that this might be a single sample rather than a collection of samples.

Table 43. Grounded Theory, Center: Misconceptions of a sample: Subset

Response 5: This is a small sample so the center could differ	Learning Style
• center could be very different because again, we will be using only a small sample of the larger population • Only 100 people in sample so mean could vary • Sample size is small and without large sample size, do not know how closer the distribution is to population distribution • small sample so might not be centered at mu w/ n=100	Sensor Sensor Intuitor Intuitor
Response 6: Outliers will make it skewed, so we don't know	
• Large sample but do not know if normal because no info about outliers, may be skewed • because it may be skewed either to the left or right • The employees chosen could exclude some of the outliers, which would affect the mean • do not know if data are normally distributed so data can be heavily skewed in a certain direction, making the mean change but not the median	Sensor Sensor Intuitor Intuitor
Response 9: The sample mean is not always equal to μ.	
• cannot predict it will be exactly 34 • the mean may be 34 but we do not know the sample mean is 34 • we do not know the exact mean location, but we can generally conclude that it will be around 34, but it's not guaranteed • 34 is mean of the population. The sample mean could be a different value	Sensor Sensor Intuitor Intuitor
Response 10: This is random and could vary with each sample size. We aren't sure how close to the population. It might change.	
• 100 random employees so we do not know where the SAMPLING DISTRIBUTION of the mean will be centered • do not know where Sampling distribution centered because only one sample. For Sampling Distribution you need distributed values in same sample and same population • do not know because it is random and they might randomly select all older • Because you could randomly select many younger workers and the mean of the sample will be much lower than that of the parameter	Sensor Sensor Intuitor Intuitor

The student is conceptualizing the center of the sampling distribution as a single (random) sample, rather than an accumulation of sample means taking from all possible samples of size 100. This is a fact that is often overlooked in a basic statistics course, but is quite pivotal in the understanding of sampling distributions. This misconception held 18 (22.5%) intuitors and 79 (24.76%) of sensors within the categories.

Table 44. Grounded Theory, Center: Misconceptions.

Misconception Reasoning	Intuitors	Sensors
Misconception	20 (25%)	92 (28.84%)
Other	60 (75%)	227 (71.16%)
		p-value= 0.5783

I generated another 2x2 table (above) to conduct a Fisher's exact test to determine if there is a significant difference between sensors and intuitors in exhibiting misconceptions in their reasoning. This test indicates that we would fail to reject the null hypothesis (p=0.5783) such that there is not enough evidence to suggest that sensors are more likely than intuitors to exhibit misconceptions in their reasoning about the likelihood of an outcome.

More Information Needed. Finally, Response 8 states more information about the population distribution or the specific sample taken is needed. Table 45 details a few of the responses within this category. Only 10% of both the intuitors and sensors fell into this category.

Table 45. Grounded Theory, Center: More Information Needed, Subset

Response 8: We do not have enough information. The sample average is unknown. We need the population distribution. This is random and we don't know if it is normal.	Learning Style
• The mean may vary from population mean because do not know data and is a simple random sample • No significant information to determine where the Sampling Distribution will be. We need xbar-mu_o/ sigma/root(n) all in order to solve for Sampling Distribution • do not now until the sample is complete • do not know center because do not have xbar	Sensor Sensor Intuitor Intuitor

Despite Response 7 mentioning the median, we could also group it into this category of "More information needed" since it states we need the median to calculate the center, similarly, response 10 which states the uncertainty, could also be grouped with this information. Adding these two responses gives us 16.25% of intuitors, 22.57% of sensors, and 21.31% of the total population.

Overall, the largest response rates fell inside the correct reasoning. As stated previously, the question regarding the center is likely the easiest question in our analysis, since the word center is synonymous with average, middle, or mean.

Spread.

A large multinational company knows that the average age of their employees is 34 years. They also know that the standard deviation of the ages of these employees is 8 years. A manager from human resources is going to randomly select a sample of 100 employees. Consider the sampling distribution of the mean of a sample of size 100 from this population. From the information given, which of the following statements are true about this sampling distribution?

Which of the following is true? Pick one and explain.

a. The sampling distribution will have a higher standard deviation than the population.

b. The sampling distribution will have the same standard deviation as the population.

c. The sampling distribution will have a smaller standard deviation than the population.

Overall Results. The following 10 categories emerged from the data of responses from the students, listed in Table 46. The categories preceded with a double asterisk indicate a correct justification to the answer. A correct answer and correct justification is indicated by a single asterisk and listed in parentheses if it does not correspond with the count for the response of the entire group.

The following 10 categories emerged from the data of responses from the students regarding the spread of the sampling distribution with respect to the parent population. The distribution of responses between intuitors and sensors is about the same with the exception of two categories. More sensors than intuitors stated the Central Limit Theorem or the formula to calculate the standard deviation, but not at any level of significance, however the p-value from the Fisher's Exact test is 0.1358. I would suspect that with more collection of data, I might be able to observe a significant difference, with sensors being more likely than intuitors to give a formula or recite the central limit theorem. This is expected, since a characteristic of a sensor is that they prefer to memorize facts and formulas and while the results are not statistically significant, I do believe they are meaningful.

Finally, more intuitors than sensors failed to provide a response, with a p-value of 0.002. The test is evaluating whether or not this is a random occurrence or if intuitors are more likely than sensors to not leave a response, with respect to this question. This p-value indicates that, at least for this question, it is statistically significant that more intuitors failed to provide a justification for their answer. Intuitors are less likely to pay attention to small details and perhaps missed the part that stated, Explain.

Table 46. Grounded Theory, Spread: Categories

The sampling distribution will have a smaller STD than the population because it is a distribution of sample means that represent the population mean	**Intuitor**	**Sensor**	**Total**
1: Since the sample size is smaller than the population size, n<N and s<σ. A larger sample size gives a larger STD	10 (12.50%)	53 (16.61%)	63 (15.79%)
2: Because the mean is larger in the sampling distribution, s > σ. The numbers will be further away from the mean.	1 (1.25%)	4 (1.25%)	5 (1.25%)
**3: The sampling distribution has a smaller spread therefore a smaller STD than the population. The sampling distribution is symmetric and centered around μ.	*5 (6.25%)	*22 (6.90%)	*27 (6.77%)
4: Small sample has a larger STD therefore since n < N, s > σ. Larger n has a smaller std so σ is small. (correct logic wrong application)	18 (22.50%)	79 (24.76%)	97 (24.31%)
**5: Formula is σ/√(n%), Central Limit Theorem	*6 (7.50%)	*46 (14.42%)	*52 (13.03%)
6: Outliers can affect the population more and therefore the population has a larger STD therefore s<σ.	2 (2.50%)	5 (1.57%)	7 (1.75%)
** 7: Sampling distribution (is of the mean), is repeated samples (not individual observations) and therefore has a smaller STD.	*2 (2.50%)	*9 (2.82%)	*11 (2.76%)
8: Outliers have less of an effect upon a larger N such as the population therefore s > σ.	0	3 (0.94%)	3 (0.75%)
9: Sampling distribution has a wider range of numbers	2 (2.50%)	4 (1.25%)	6 (1.5%)
10: STD of the samples should match the population. It won't change just because the number changes. Sample size has no effect on STD.	18 (22.5%)	70 (21.94%)	88 (22.06%)
NO RESPONSE	16 (20.0%)	24 (7.52%)	40 (10.03%)

Note: * Correct Answer AND Correct Reasoning. ** Correct Reasoning. Correct Reasoning responses are highlighted. *The final column represents proportions of the entire population.* Numbers in Parentheses with an asterisk represent the count of correct reasoning and correct answer selection. If the response is ** as a correct reason, but the * count is not in parentheses, then all students in this instance have the correct answer and correct reasoning. The number inside parentheses below the counts represent percentages of the corresponding population

Example:

Response 3. 5 intuitors gave this reason, and all of them had the correct answer.
This number is 6.25% of all intuitors
22 sensors gave this reason and all of them had the correct answer.
This number is 6.9% of all sensors.

Table 46 details the distribution of correct reasoning and responses for intuitors and sensors. More sensors answered this question correctly and gave a correct reasoning for the answer selection. Table 47 illustrates the distribution of responses by intuitors and sensors and highlights the proportions with both a correct answer and a correct justification for the answer. There is no significant difference between intuitors and sensors, and the p-value is 0.1381.

Table 47. Grounded Theory, Spread: Answer and Reasoning

SPREAD	Intuitors	Sensors
A	27 (33.75%)	89 (27.90%)
B	21 (26.25%)	78 (24.45%)
C*	32 (40.00%)	152 (47.65%)
Correct Reasoning**	13 (16.25 %)	77 (24.14%)
		*p-value**: 0.1381*

Note: *Correct Answer. ** Correct answer AND correct Reasoning. Percentages indicate the proportion of students within the ILS category choosing that answer out of total number of students within the category.

I am not surprised that while over 40% of both sensors and intuitors chose the correct answer for this problem, only a small proportion of them were able to correctly justify the answer, though it does seem that sensors are more likely than intuitors to justify the answer selection correctly, although this was not statistically significant.

Correct Reasoning. There are three different correct reasoning categories. The first states that the spread is smaller. The second gives the formula for the sampling distribution standard deviation or recites the Central Limit theorem (or both) and the final response, which I believe gives the strongest support, actually conceptualizes this problem as repeated sample means rather than individual observations. I would expect the students that answered along the lines of response 7 to exhibit solid understanding on sampling distributions. It is not surprising that only 11 students, 2%, used this line of thinking to justify the problem.

Table 48. Grounded Theory, Spread: Correct Reasoning Categories, Subset

Response	Learning Style
Response 3: The sampling distribution has a smaller spread therefore a smaller STD than the population. The sampling distribution is symmetric and centered around μ.	Learning Style
• Whenever a sampling distribution is done, the standard deviation decreases	Sensor
• the standard deviation will be smaller because it is a sampling distribution of the mean	Sensor
• Again, the graph will look different, smaller thinner perhaps, but normal	Intuitor
• the sampling distribution is centered around the mean but is tighter due to being averages of groups rather than individuals so the standard deviation is smaller	Intuitor
Response 5 Formula is $\sigma/\sqrt{(n)}$, Central Limit Theorem	
• The bigger size sample, sigma/sqrt(n), the smaller the standard deviation	Sensor
• because as the sample size gets larger the mean becomes closer to the actual value and to a normal distribution	Sensor
• because it increases your denominator of the standard deviation: sigma/ sqrt(n)	Intuitor
• formula < sigma for all values of n > 1 (not sure if words or formula)	Intuitor
Response 7 Sampling distribution (is of the mean), is repeated samples (not individual observations) and therefore has a smaller STD.	
• More sample statistics with more people, increase in n, decrease in STANDARD DEVIATION	Sensor
• STANDARD DEVIATION is smaller because multiple samples	Sensor
• sampling distribution is the distribution of the mean so values are closer together	Intuitor
• Higher sample implies smaller STANDARD DEVIATION, so sample has smaller standard deviation than pop	Intuitor

I also tested the hypothesis: Are sensors more likely to use correct reasoning but choose an incorrect answer? This is due to the sensors nature of fact memorization, but inability to think abstractly, meaning, the sensor is less likely than an intuitor to apply a definition to answer selection. In order to test this, I only created a 2x2 table of Incorrect Response and Correct Reasoning versus the other three combinations namely: Incorrect Response and Incorrect Reasoning, Correct Response and Correct Reasoning, and Correct Response and Incorrect Reasoning, called “Other Combinations”.

Table 49. Grounded Theory, Spread: Correct Reasoning and Incorrect Response.

Reasoning and Response	Intuitors	Sensors
Correct Reason with Incorrect Answer	0	0
Other Combinations	80 (100%)	319 (100%)
		p-value= **

Since these are both zero we cannot calculate the p-value, but it is not significant since the totals are the same.

Misconceptions. A common misconception is that "the variability will increase with sample size because more data means more spread" (Chance, delMas, & Garfield, 2004, p. 302). Response 1 captures this sentiment, with students stating that since the sample size is smaller than the population size, the sampling distribution will have a smaller standard deviation than the population. The answer here is correct, the sampling distribution will have a smaller variation than the population, but not because the population has a larger sample size. 63 of the students used this line of reasoning and got the correct answer, 10 (12.5%) intuitors and 53 (16.61%) sensors.

Table 50. Grounded Theory, Spread: Misconceptions About Variance, Subset

Response 1: Since the sample size is smaller than the population size, $n<N$ and $s<\sigma$. A larger sample size gives a larger STD	Learning Styles
• get people close to same ages because smaller random sample	Sensor
• The STANDARD DEVIATION will be smaller because the random employees may all be around the same age	Sensor
• because less observations in sample than in pop	Intuitor
• larger pop increase standard deviation	Intuitor

Response 2 is a small category, and while I wonder if I could have placed the five observations into a pre-existing category, I felt like it had enough uniqueness to establish itself. The students in this category mention the mean of the population and how the mean or data of the sampling distribution are further away from mu, and therefore the standard deviation must be larger in the sampling distribution. I cannot pretend that I fully understand what the student is trying to express here in regards to this explanation, but I think it falls along the lines of this misconception, “Think the mean of a positive skewed distribution will be greater than the mean of the sampling distribution for samples taken from this population” (Chance, delMas, & Garfield, 2004, p. 302). Only one intuitor (1.25%) and four sensors (1.25%) used this response.

Table 51. Grounded Theory, Spread: Misconception about the Mean, Subset

Response 2: Because the mean is larger in the sampling distribution, s > σ. The numbers will be further away from the mean.	Learning Style
• if the mean is larger the sampling distribution will also be larger	Sensor
• smaller standard deviation than population because the mean of the population is greater than the mean of the SAMPLING Distribution	Sensor
• Mean for population likely to be larger than sample which will make our standard deviation larger with more subjects to count for which makes distance from mean greater	Sensor
• The sampling distributions standard deviation will be higher than the population's standard deviation because the numbers in the sampling distribution will be further away from the mean.	Intuitor

The remaining misconceptions fall into this category taken from, Chance, delMas and Garfield (2004, p. 302),

- Do not understand that a sampling distribution is a distribution of sample statistics.
- Confuse one sample (real data) with all possible samples (in distribution) or potential samples.

Response 4 is another common misconception that has not been completely captured in the literature, but can be easily summarized by applying the misconception of the student's conceptualization of a sampling distribution as a simple random sample to the definition of the standard deviation of the sampling distribution. Namely; the students believed that since the sample size is smaller than the population size, the standard deviation of the sampling distribution will be larger. This shows an understanding that the sampling distribution will decrease with an increase in sample size, however, the student fails to make the connection that this is not a sample but rather a collection of samples. This is the simple type of thinking that I have discussed in prior chapters. Students know one concept but think about it simply rather than connecting the ideas together in an abstract sense. If the size is smaller, then the standard deviation must be larger, forgoing the fact that we are not comparing “apples to apples”. This was the largest category of reasoning with 97 students (24.31%), of which 18 (22.50%) were intuitors and the remaining 79 (24.76%) were sensors.

Table 52. Grounded Theory, Spread: Misconceptions, Subset

Response 4: Small sample has a larger STD therefore since $n < N$, $s > \sigma$. Larger n has a smaller std so σ is small. (correct logic wrong application)	Learning Style
• less people in sampling distribution so higher standard deviation	Sensor
• SAMPLING Distribution will be higher standard deviation than population because sample size is smaller than population and sqrt(n) of standard deviation is being divided by a smaller number when makes the whole deviation larger	Sensor
• because as number in study decrease, the standard deviation increases because there is more variability	Intuitor
• the smaller the population the larger the standard deviation will be due to increase in variability	Intuitor

Response 6 and 8 are very similar in that the student mentions the presence of outliers and how they affect the standard deviation. From Garfield and Chance (2000, p. 118), Misconceptions involving averages (common number, median vs. mean, outliers). The difference is that in Response 6, the student believes that the data will be more spread out in the population whereas in Response 8, the student believes that outliers are more likely to cause the sampling distribution to be more varied. Only a small portion of students fell into either of these categories with 2.5% of the total population.

Table 53. Grounded Theory, Spread: Misconception involving Outliers, Subset

Response 6: Outliers can affect the population more and therefore the population has a larger STD thus s<σ	Learning Style
• May be extreme outliers in age in the population that would affect it • Sample is less likely to be extreme than a single data point • Since the max and min ages are likely not to be selected randomly, the standard deviation should decrease • likelihood of sampling extremes is lower so standard deviation of sample will be smaller	Sensor Sensor Intuitor Intuitor
Response 8: Outliers have less of an effect upon a larger N such as the population therefore s>σ	
• larger standard deviation will be present because with a right skewed dist. mean, a larger standard deviation will be present • higher chance of outliers in a random sample • because outliers have less of an effect on large groups	Sensor Sensor Sensor

Response 9 is was a pretty vague generalization, but the data did not fall into any other category with ease. Only six students used this rationale of spread or wider range with 2 intuitors (2.5%) and 4 sensors (1.25%). A subset of responses, taken as a simple random sample, is presented in Table 54.

Table 54. Grounded Theory, Spread: Miscellaneous Category, Subset

Response 9: Sampling distribution has a wider range of numbers	Learning Style
• SAMPLING DISTRIBUTION will have larger STANDARD DEVIATION then there will be more variation in ages used in samples than in population • because there's a wider range of numbers in the sampling distribution of the mean • sample of the population has a larger range of data to cover to support the overall population • The sample will be more spread out than the population distribution	Sensor Sensor Intuitor Intuitor

A new misconception that emerged from these observations was the second largest category in the data, namely, that the sampling distribution will have the same variability as the population. While this is similar to what Chance et al. (2004) describe "Predict that sampling distributions for small and large sample sizes have the same variability" (p. 302), it is not exactly the same since our students are comparing the sampling distribution to the population, not to a sampling distribution with a different sample size. All of the students that fell into this category chose answer selection "b" which states that there is no difference; hence the logic was in line with the answer selection. 22% of the population chose this as their answer.

Table 55 Grounded Theory, Spread: Misconceptions on Standard Deviation, Subset

Response 10: STD of the samples should match the population. It won't change just because the number changes. Sample size has no effect on STD.	Learning Style
• same because only the mean and median are affected • since same mean, same STANDARD DEVIATION • SAMPLING DISTRIBUTION is representative of population • standard deviation should not change just because the mean changes	Sensor Sensor Intuitor Intuitor

Though we did not observe any significant differences between sensors and intuitors, we did gain some useful insights into student misconceptions. The biggest misconception to play out in our data is that students perceive a sampling distribution of means as a simple random sample rather than a collection of sample means.

I generated a 2x2 table for a Fisher's exact test to test if there is a significant difference between how sensors and intuitors exhibit misconceptions in their reasoning.

Table 56. Grounded Theory, Spread: Misconceptions

Misconception Reasoning	Intuitors	Sensors
Misconception	51 (63.75%)	218 (68.34%)
Other	29 (36.25%)	101 (31.66%)
		p-value= 0.4273

This test indicates that we would fail to reject the null hypothesis such that there is not enough evidence to suggest that sensors are more likely than intuitors to exhibit misconceptions in their reasoning about the likelihood of an outcome.

More Information Needed. Surprisingly I did not discover any students in this category, and thus one did not emerge for the question on the standard deviation of the sampling distribution in relation to the parent population.

Shape.

> A large multinational company knows that the average age of their employees is 34 years. They also know that the standard deviation of the ages of these employees is 8 years. A manager from human resources is going to randomly select a sample of 100 employees. Consider the sampling distribution of the mean of a sample of size 100 from this population. From the information given, which of the following statements are true about this sampling distribution?
>
> Which of the following is true? Pick one and explain.
>
> a. We know that the shape will be right skewed.
>
> b. We know that the shape will be approximately symmetric.
>
> c. We cannot tell what the shape will look like.

Overall Results. The following 10 categories emerged from the data of responses from the students, listed in Table 57. The categories preceded with a double asterisk indicate a correct justification to the answer. A correct answer and correct justification is indicated by a single asterisk and listed in parentheses if it does not correspond with the count for the response of the entire group.

Table 57. Grounded Theory, Shape: Categories

We know the shape of a sampling distribution is going to be normal with mean μ and std, σ/√(n)	**Intuitor d**	**Sensor**	**TOTAL**
**1: Normal, bell-shaped, because n>30 and using the CLT. Xbar is normal(μ,(σ/√(n))	*8 (10.00)	44 (*43) (13.8%)	52 (*51) (13.03%)
2: Should have the shape of the population since the sample size is large	6 (7.50%)	27 (8.46%)	33 (8.27%)
** 3: Sampling distribution of the mean is always normal around mean. It is symmetric.	*5 (6.25%)	*17 (5.33%)	22 (5.51%)
4: We can predict the shape because we know μ and σ	3 (3.75%)	11 (3.45%)	14 (3.51%)
5: It is normal because it is a random sample and a large sample size	9 (11.25%)	21 (6.58%)	30 (7.52%)
6: We don't have enough information, we need the data points, we need to perform a test, we need to know if it is normal. We need the mean.	11 (13.75%)	57 (17.87%)	68 (17.04%)
7: We need the median to know the shape	0 (0.00%)	15 (4.70%)	15 (3.76%)
8: We need the shape of the parent population to know. It is not normal	20 (25.00%)	75 (23.51%)	95 (23.81%)
9: It may have outliers which affect the shape. It could be skewed. Shape is flatter for the mean	5 (6.25%)	21 (6.58%)	26 (6.52%)
10: We don't know the standard deviation.	1 (1.25%)	1 (0.31%)	2 (0.5%)
NO RESPONSE	12 (15.00%)	30 (9.40%)	42 (10.53%)

Note: * Correct Answer AND Correct Reasoning. ** Correct Reasoning. Correct Reasoning responses are highlighted. The final column represents proportions of the entire population. Numbers in Parentheses with an asterisk represent the count of correct reasoning AND correct answer selection. If the response is ** as a correct reason, but the * count is not in parentheses, then all students in this instance have the correct answer and correct reasoning. The number inside parentheses below the counts represent percentages of the corresponding population

Example:

Response 1. 8 intuitors gave this reason, and none of them had the correct answer.

This number is 10% of all intuitors

44 sensors gave this reason and 43 of them also had a correct answer selection.

This number is 13.8% of all sensors.

Table 57 details the distribution of correct reasoning and responses for intuitors and sensors. More intuitors than sensors answered this question correctly and gave a correct reasoning for the answer selection. Table 58 illustrates the distribution of responses by intuitors and sensors and highlights the proportions with both a correct answer and a correct justification for the answer. There is no significant difference between intuitors and sensors in giving a correct response for a correct answer with a p-value of 0.7464.

Table 58. Grounded Theory, Shape: Answer and Reasoning

Shape	Intuitors	Sensors
A	33 (41.25%)	124 (38.87%)
B*	30 (37.50%)	105 (32.92%)
Correct Reasoning**	13 (16.25%**)	60 (18.81%**)
C	17 (21.25%)	90 (28.21%)
		p-value**: 0.7464

Note: *Correct Answer. ** Correct answer AND correct Reasoning. Percentages indicate the proportion of students within the ILS category choosing that answer out of total number of students within the category.

Correct Reasoning. There are only two correct reasoning categories for this question. The first uses the Central Limit Theorem and the second uses the rationale that the sampling distribution is always normal and symmetric. The first response seems more detailed and sophisticated, but if the sensor can memorize a definition, I could argue that the second correct response takes more generalization and less rehearsal. Of the intuitors 16.25% used this rationale whereas, 18.81% of the sensors did. More sensors used the Central Limit Theorem definition and formula than intuitors, but the difference is not significant.

Table 59. Grounded Theory, Shape: Correct Reasoning Categories, Subset

Response	Learning Style
Response 1: The shape will be normal, or bell shaped. This is because n>30 and we can use the central limit theorem. We know that $\overline{x}$ is distributed normal (μ,($\sigma/\sqrt{(n)}$)).	Learning Style
• SAMPLING DISTRIBUTION is symmetric because it is a predictable pattern among samples of same size and same population, • Shape of SAMPLING Distribution bell shape because approximately normal because sample size > 30 and CLT, • if sufficiently large or normal the sampling distribution will be normal symmetric distribution around the mean of the population, though more tightly centered. • always know what shape will look like	Sensor Sensor Intuitor Intuitor
Response 3: The sampling distribution of the mean is always normal around the mean and symmetric	
• since we're taking a sample the curve will be more symmetric because it is of means • shape of SAMPLING DISTRIBUTION of the mean will be approximately symmetric because calculating means • SAMPLING DISTRIBUTION follow normal distribution, • cause we are dealing with sample MEANS and not individual values, the distribution would be normal-symmetric	Sensor Sensor Intuitor Intuitor

Additionally I tested the hypothesis: Are sensors more likely to use correct reasoning but choose an incorrect answer? This is due to the sensors nature of fact memorization, but inability to think abstractly, meaning, the sensor is less likely than an intuitor to apply a definition to answer selection. In order to test this, I only created a 2x2 table of Incorrect Response and Correct Reasoning versus the other three combinations namely: Incorrect Response and Incorrect Reasoning, Correct Response and Correct Reasoning, and Correct Response and Incorrect Reasoning, called "Other Combinations".

Table 60. Grounded Theory, Shape: Correct Reasoning and Incorrect Response.

Reasoning and Response	Intuitors	Sensors
Correct Reason with Incorrect Answer	0	1 (0.3%)
Other Combinations	80 (100%)	318 (99.7%)
		p-value= 1

From this resulting p-value of 1, we do not reject the null hypothesis in favor of sensors and intuitors being equally likely to provide a correct response and make an incorrect answer.

Response 4 shows a lower level of correct reasoning, but not very strong reasoning. The students support the answer by stating that by knowing μ and σ, they also know the shape of the sampling distribution. Of the 14 students to use this reasoning, only one gave an incorrect answer. For the remaining 13 students to use this reasoning and answer correctly, I believe they used some correct logic in their answer, just not enough for me to group these students with those who strongly justified their answer.

Table 61. Grounded Theory, Shape: Partially Correct Response Categories, Subset

Response 4: We can predict the shape because we know μ and σ	Learning Style
• We can make predictions on what the sampling distribution will look like based on our knowledge of the average age and STANDARD DEVIATION of the original set of data, we will not know exactly what the distribution	Sensor
• Basic shape will be the same because we know average age and standard deviation	Intuitor
• Know shape because we know mean and standard deviation	Intuitor

The one student who answered incorrectly with reason 4 stated "Population and sampling distribution will look same because same standard deviation and mean". The sentiment that they will know the shape because they know the average and standard deviation is echoed throughout the remaining 11 responses, but does not indicate to me that the student understood that the shape of the sampling distribution of means follows a predictable pattern of normality, but rather based it on data observations.

Misconceptions. More misconceptions emerged from the data. Taken from Chance, delMas and Garfield (2004, p. 302),

- Do not understand that a sampling distribution is a distribution of sample statistics.
- Confuse one sample (real data) with all possible samples (in distribution) or potential samples.

Response 5 almost seems also along the right lines, but is not quite right. I believe that the student is alluding to a simple random sample rather than a distribution of sample means.

Table 62. Grounded Theory, Shape: Misconception of Sample, Subset

Response 5: It is normal because it is a random sample taken of a large sample size	Learning Style
• Normal because the sample size is 100 people	Sensor
• because sample is random the shape would be normally distributed,	Sensor
• the total of all sampling distributions will be skewed but not the single sampling distribution	Intuitor
• SAMPLING Distribution should be somewhat normal because sample size is large	Intuitor

These answers sound like correct reasoning at first pass, but with careful consideration, the student is comparing the sampling distribution to a random sample. The student has the right idea about a larger sample size producing a normal distribution (with repeated samples of the mean), but the student fails to address this as a collection of sample means, and not a single random sample.

A small, 6.52% of the population mentioned the presence of outliers, and that it would make the distribution skewed.

Table 63. Grounded Theory, Shape: Misconception of Sample, Subset

Response 9: It may have outliers which affect the shape. It could be skewed. Shape is flatter for the mean	Learning Style
• Right skew because outliers	Sensor
• Do not know because the outliers are unknown	Sensor
• Skewed right because of outliers	Intuitor
• Do not know of outliers so cannot predict shape	Intuitor

These students fall into the misconception that this is a sample and not a distribution of sample statistics, confusing one sample with all possible samples. There are no significant differences between sensors and intuitors in this topic.

Common misconceptions regarding the shape of a sampling distribution prevail in our data. One of the most common misconceptions is when a student believes that the shape of the sampling distribution should reflect the shape of the population.

> "Believe sampling distribution should look like the population (for sample size $n > 1$)" and "think sampling distribution should look more like the population as the sample size increases (generalizes expectations for a single sample of observed values to a sampling distribution)" (Chance, delMas, & Garfield, 2004, p. 302).

Table 64. Grounded Theory, Shape: Misconception of Parent, Subset

Response 2: Should have the shape of the population since the sample size is large	Learning Style
• Since n=100 graph should resemble the population shape • Since large sample, shape relative same as population • Sample shape should be like parent if it is normal, then sample should be normal because it used same data • Shape of Sampling Distribution will be normal if population is normal	Sensor Sensor Intuitor Intuitor
Response 8: We need the shape of the parent population to know. It is not normal	
• Expect our sample to have the same right skewed shape as our population • If population is right skewed then so is the sampling distribution • If population is skewed so will the distribution. • Data skewed, mean is larger than median, distribution is skewed right	Sensor Sensor Intuitor Intuitor

This widely recognized misconception is thoroughly supported by the data in this project. Since both responses mention the parent population, I combined them together. 32% of the population talked about the parent population with 32.5% intuitors and 31.97% sensors. There is clearly no significant difference in sensors and intuitors believing that the shape of the sampling distribution will represent the parent population.

I generated a 2x2 table for a Fisher's exact test to test if there is a significant difference between how sensors and intuitors exhibit misconceptions in their reasoning.

Table 65. Grounded Theory, Shape: Misconceptions.

Misconception Reasoning	Intuitors	Sensors
Misconception	40 (50%)	144 (45.14%)
Other	40 (50%)	175 (54.86%)
		p-value= 0.4539

This test indicates that we would fail to reject the null hypothesis such that there is not enough evidence to suggest that sensors are more likely than intuitors to exhibit misconceptions in their reasoning about the likelihood of an outcome.

More Information Needed. The new misconception that has emerged from most of the data is that the student needs more information in order to solve this problem.

Table 66. Grounded Theory, Shape: Not Enough Information, Subset

Response 6: We don't have enough information, we need the data points. We need to perform a test; we need to know if it is normal. We need the mean.	Learning Style
• cannot tell yet because we haven't surveyed anyone yet to get the answer, • since we do not know population size, we do not know if the sample size of 100 is relatively small or large, and thus do not know if it is relatively large enough to mimic the shape of the population • need more info to determine shape. • we do not have enough info	Sensor Sensor Intuitor Intuitor
Response 7: We need the median to know the shape	
• Need the median could be outliers • all have is average age and standard deviation, need median, q1, q3, and outliers to determine shape.	Sensor Sensor
Response 10: We don't know the standard deviation	
• We do not know the average or mean of the Sampling Distribution so we do not know the shape. • it depends on the sample's variability	Sensor Intuitor

All three of these categories (Response 6, 7 and 10) fall into a larger scope where the student believes more information is needed about the sampled data or the population in order to be able to answer a question about the sampling distribution. Further, some of these responses lead me to believe that that student is confusing a sampling distribution of means

Response rates between intuitors and sensors are nearly proportional for all of the categories that emerged in the topic of Shape of the sampling distribution. While 46% of the population answered this question correctly, only 22% were able to support their answers with rationale thinking, leaving a large amount of room for improvement in student understanding.

Foundational Questions Combined. Finally, I combined the four foundational questions and the grounded theory responses in order to determine if the students who used good reasoning for one question, kept this for the remaining questions. This would exhibit sound understanding of sampling distributions if the student was able to consistently answer a question about sampling distributions correctly. Only 13 (3.26%) of the students answered all four questions correctly and provided correct reasoning, 3 (3.75%) of which were intuitors and the remaining 10 (3.13%) sensors. Table 67 may seem a bit confusing but is a 4 dimensional display of the four questions and whether the student used correct reasoning or not across all four. The largest proportion of students answered all but the likelihood question incorrectly. Perhaps this is indicative of the lack of connection between sample and sampling distribution, or rather, that students understand sample, but are unable to apply their knowledge of a single sample to a collection of samples. One of the biggest reasons that the students exhibited incorrect reasoning was because they likened a sampling distribution to a single, simple random sample. If we were to replace the word “simple random sample” with sampling distribution, or collection of sample means, and continue to use their logic related to this, the student would have had the correct logic. Meaning, the student said everything correctly, except for the fact they called it a simple random sample rather than a collection of sample means. Perhaps then, the issue is not in understanding the definition and properties of a sampling distribution, but rather, realizing it is a collection of means from simple random samples and the overall confusion in terminology between sample and sampling distribution of means.

I also created a table for both intuitors and sensors (see Appendix L). The majority of sensors and intuitors both fell into the same category of only the likelihood question correct. The weight of the data fell into the bottom part of the chart, meaning that the most combinations were of incorrect responses rather than a large combination of correct ones.

Only 16.29% of the students answered at least three questions correctly, with 12.5 % intuitors and 17.25% sensors.

Table 67. Grounded Theory: Interaction of Spread, Shape, Center, and Likelihood

All Students				SHAPE				
				CORRECT		INCORRECT		
				LIKELIHOOD		LIKELIHOOD		
				CORRECT	INCORRECT	CORRECT	INCORRECT	
SPREAD	CORRECT	Center	Correct	13 (3.26%)	9 (2.26%)	17 (4.26%)	22 (5.51%)	61
			Incorrect	6 (1.50%)	4 (1.00%)	20 (5.01%)	14 (3.51%)	44
	INCORRECT	Center	Correct	20 (5.01%)	10 (2.51%)	50 (12.53%)	66 (16.54%)	146
			Incorrect	6 (1.50%)	7 (1.75%)	78 (19.55%)	57 (14.29%)	148
				45	30	165	159	399

Note: Numbers in parentheses represent the percentage for the collection of data.

Next I performed analysis on the students who answered all four questions correctly, to determine if they also correctly answered the two graphical questions regarding sampling distributions, or the other two questions. Only two of the students who answered all four foundational questions correctly also answered both graphic questions correctly as well as the other two remaining questions, one was a sensor and the other an intuitor.

It is hard to perform additional analysis on these 13 students since the sample is so small. All 13 of them correctly answered the questions involving the z-statistic. It is hard to determine whether or not these foundational questions actually indicate that the student understood sampling distributions since the majority of students did not exhibit statistical reasoning about the foundational questions, based on my interpretation of the results. The majority of the students were only able to answer no more than one question from the graph of the sample and means correctly and correctly reason about no more than one question from the foundational questions. From this, it is hard to determine whether or not the foundational questions really did determine understanding of sampling distributions since the students failed to show understanding for both foundational question and the sampling distribution of the mean.

Grounded Theory Summary. Overall, I attempted to use grounded theory to answer this question; Are there any differences between sensors and intuitors in how they use logic and reasoning in qualitative responses to sampling distribution questions? Although we did not observe any significant results, I could argue that the small amount of significance in the likelihood question, with a p-value of 0.0772 and an even smaller significance regarding the question on Spread with p-value of 0.1358 which indicates that there might be some differences between sensors and intuitors, just not at a 0.05 significance level and the latter not at the 0.10 level either.

Table 68. Summary of Grounded Theory

Data Collected	Semester	Analysis performed	Significant Results
Foundation: Likelihood	Fall 07 – Fall 08	Correct Reasoning & Correct Response	Partial (at the 0.10 level)
		Correct Reasoning & Incorrect Response	No
		Misconceptions	No
Foundation: Center	Fall 07 – Fall 08	Correct Reasoning & Correct Response	No
		Correct Reasoning & Incorrect Response	No
		Misconceptions	No
Foundation: Spread	Fall 07 - Fall 08	Correct Reasoning & Correct Response	No
		Correct Reasoning & Incorrect Response	No
		Misconceptions	No
Foundation: Shape	Fall 07 – Fall 08	Correct Reasoning & Correct Response	No
		Correct Reasoning & Incorrect Response	No
		Misconceptions	No

The significance in the question on likelihood indicates that an intuitor might be more likely than a sensor to answer this question and rationalize it properly. This question involves the conceptualization of two sample sizes and their means, asking the student to decide which one is more likely to have a mean much greater than the population mean. It is not surprising that the intuitor is more likely to answer this correctly given the fact that intuitors are more abstract thinkers.

Conversely, the question on Spread has a small amount of significance in the opposite direction. Since sensors are likely to memorize formulas and definitions, they are more likely to have memorized the definition for the standard deviation of the sampling distribution, which is what the data also illustrate. Their ability to answer this question correctly and justify the answer might be related to their tendency to memorize a formula and plug in the numbers and then directly apply this to the question. More evidence would be required in order to make a definitive statement about these groups.

Text mining using SAS.

I used advanced computer language to create categories from the student's responses. The goal was to see if the software would produce similar categories to what I created using grounded theory, since SAS Text miner uses a modified grounded theory approach. I am not an expert in the use of SAS Text Miner, and thus a more sophisticated use might be applied given a greater level of experience. Text Mining "mines" the data and creates "Topic" categories, using keywords taken from the document. I use the words category and topic interchangeably. I attempt this for all four questions and present the results below.

Likelihood.

According to a recent report the average income of people in Adair County, Missouri is $31,023. Which of the following is more likely? Pick one and explain.

a. We take a random sample of 10 people from this county and find that the average is over $50,000.

b. We take a random sample of 200 people from this county and find that the average is over $50,000.

c. We do not have enough information to answer this question.

Table 69. Text Mining, Likelihood: Topics, Intuitor

TOPIC Likelihood : INTUITOR	Response Choice			Total
	a	b	c	
1 - Small, Small Sample Size, Sample, Population, likely	13	1	0	14 (17.5%)
2 - Closer, Population, Mean, sample size	5	2	2	9 (11.3%)
3 - Random, Random Sample, Skew, predict	5	1	1	7 (8.75%)
4 - Average, value, Predict	2	1	4	7 (8.75%)
5 - Large, Large sample size	5	0	0	5 (6.25%)
6 - Great, Sample, Mean	5	2	0	7 (8.75%)
7 - High, Little, value	4	1	1	6 (7.5%)
8 - Sample Size, Result, Entire	7	0	1	8 (10.0%)
9 - Data, Mean, Little	2	0	2	4 (5.0%)
10 - Standard Deviation, Outlier, know	1	0	6	7 (8.75%)
NO TOPIC	1	0	4	5 (6.25%)

***Note:* *Percentages are based on the total number of intuitors in the population (80). One observation is missing, thus the total count here is only 79.**

Table 70. Text Mining, Likelihood: Topics, Sensor

TOPIC Likelihood : Sensor	Response Choice			Total
	a	b	c	
1- Likely, small sample size, small, sample, large, likely, find, high,	36	1	0	37 (11.6%)
2- average, value, find, likely	11	10	19	34 (10.6%)
3- high, chance, value, average, great, little, pick, mu, good, small sample size, small, high chance	24	3	1	28 (8.8%)
4 - large, sample size, large, increase, chance, good, sample, variability	14	16	1	31 (9.7%)
5 - random, random sample, sample, large, large sample size	2	7	12	21 (6.6%)
6 - mean, closer, increase, actual, population	23	2	3	28 (8.8%)
7 - predict, basis, likely, standard deviation, know, size, population	5	2	20	27 (8.5%)
8 - population, average, representative, size, entire population, entire, large, good, represent, big, sample, closer	21	10	6	37 (11.6%)
9 - outlier, data, small sample size, sample size, skew, small, great, mean, effect	25	3	0	28 (8.8%)
10 - Information, know, population, total	2	0	29	31 (9.7%)
NO TOPIC	0	1	5	6 (1.9%)
*Percentages are based on the total number of intuitors in the population (319)				

Tables 69 and 70 display the two groups (intuitor and sensor) for the Likelihood question presented above. There is some indication that a certain grouping may allude to good reasoning with the correct answer, but there is not any way to know for sure without checking all the answers manually, as done with grounded theory in the previous section. Printing a subset of responses confirmed this, and is shown in its entirety in Appendix M. Furthermore, I do believe that the grouping of terms can shed some light on what the students are talking about, but only in how they use phrases and terminology, not how they reason about them.

Topic categories are similar but not identical. Before delving further into any of this type of analysis, I looked at the breakdown of the categories to see what sort of information I could achieve. The topics for the sensors contain more keywords than the topics for the intuitors, which is likely due to the fact that there is 4 times more data for the sensor than the intuitor. None of the topics stand out as having the most number of observations, and further, you cannot determine whether or not the student understood the problem by looking at the keywords assigned to the topic paired with a correct or incorrect answer.

The information as presented in Tables 69 and 70 does not shed any light on the reasoning behind the students thought process, only the keywords and short phrases being used. Additionally, there are some observations that were not categorized, and thus it would be still be necessary to categorize them by hand. Text mining might be useful to gather buzz words, but is not going to help answer my second research question.

Center.

A large multinational company knows that the average age of their employees is 34 years. They also know that the standard deviation of the ages of these employees is 8 years. A manager from human resources is going to randomly select a sample of 100 employees. Consider the sampling distribution of the mean of a sample of size 100 from this population. From the information given, which of the following statements are true about this sampling distribution?

Which of the following is true? Pick one and explain.

a. The sampling distribution will be centered at 34.

b. We do not know where the sampling distribution will be centered.

Table 71. Text Mining, Center: Topics, Intuitor

TOPIC Center - INTUITOR	Response Choice		Total
	a	b	
1- Sample Size, distribution, normal, large	7	2	9 (11.25%)
2-large enough sample, enough, large, sample	3	1	4 (5.0%)
3- sample distribution, center, normal	6	4	10 (12.5%)
4- average, know, sampling distribution	5	3	8 (10.0%)
5- population, sample, distribution, mean	3	6	9 (11.25%)
6- data, normally, distribute, know, median	1	2	3 (3.75%)
7- mu, sampling distribution, center	8	2	10 (12.5%)
8- center, do not know, normal distribution, average	4	3	7 (8.75%)
9- likely, close to, standard deviation, mean, sample size	1	2	3 (3.75%)
10 - select, randomly, sampling distribution, random, parameter	4	2	6 (7.5%)
NO TOPIC	0	1	1 (1.25%)
NO RESPONSE	5	4	9 (11.25%)

Note: *Percentages are based on the total number of intuitors in the population (80)

Table 72. Text Mining, Center: Topics, Sensor

TOPIC Center - SENSOR	Response Choice		Total
	a	b	
1 - Sampling distribution, mean, center, always, population	38	14	52 (16.3%)
2 - large, large sample size, sample size	26	5	31 (9.72%)
3 - average (population, center, median, vary)	27	11	38 (11.91%)
4 - normal, distribution, population, large, sample	17	5	22 (6.90%)
5 - random, random sample, sample, population, outlier, do not know	5	8	13 (4.08%)
6 - mu, xbar, central limit theorem, normal, sampling distribution, large, approximately	25	3	28 (8.78%)
7 - population, sample, different, statistic, standard deviation	14	24	38 (11.91%)
8 -know, normal, standard deviation	7	11	18 (5.64%)
9 - median, skew, outlier	2	20	22 (6.90%)
10 - do not know, data, select, distribution, randomly	3	19	22 (6.90%)
NO TOPIC	4	10	14 (4.39%)
NO RESPONSE	11	11	22 (6.90%)

Note: *Percentages are based on the total number of intuitors in the population (319)

Tables 71 and 72 display the two groups (intuitors and sensor) for the Center question presented above. The response rate between answer choice A and B seem to be distributed randomly. The pattern within each topic doesn't allude to any heavy weighting one direction or the other to indicate a topic might best capture one of the choices with the exception of a few of the responses mentioning outliers, skewness, or "do not know" (Topics 9 and 10). Topics mentioning normality and the large sample size tend to be related more to answer choice A, but it is not absolute. Interestingly enough, the intuitors did not seem to mention the median, outliers or skewness, at least in context of keywords in this text mining analysis.

Despite these interesting observations, there is not any way to know for sure what rationale the student used without checking all the answers manually, as done with grounded theory above. A subset of responses confirmed this, and is shown in its entirety (see Appendix M).

Spread.

A large multinational company knows that the average age of their employees is 34 years. They also know that the standard deviation of the ages of these employees is 8 years. A manager from human resources is going to randomly select a sample of 100 employees. Consider the sampling distribution of the mean of a sample of size 100 from this population. From the information given, which of the following statements are true about this sampling distribution?

Which of the following is true? Pick one and explain.

a. The sampling distribution will have a higher standard deviation than the population.

b. The sampling distribution will have the same standard deviation as the population.

c. The sampling distribution will have a smaller standard deviation than the population.

Table 73. Text Mining, Spread: Topics, Intuitor

TOPIC Spread : INTUITOR	Response Choice			Total
	a	b	c	
sample, distribution, large, overall, population	4	3	1	8 (10.0%)
small, increase, standard deviation, people, population	4	1	6	11 (13.75%)
variability, more, increase, large, sample size	4	0	6	10 (12.5%)
correct formula, less, sigma, great, small	2	0	6	8 (10.0%)
mean, sampling distribution, distribution, standard deviation, small	1	4	3	8 (10.0%)
sample size, overall, small, data, distribution	3	1	2	6 (7.5%)
number, data, great, population, large	3	1	0	4 (5.0%)
representative, population, large sample size, large, sampling distribution	0	6	1	7 (8.75%)
NO CATEGORY	0	2	2	4 (5.0%)
NO RESPONSE	5	3	4	12 (15.0%)

Note: *Percentages are based on the total number of intuitors in the population (80)

Table 74. Text Mining, Spread: Topics, Sensor

TOPIC Spread : SENSOR	Response Choice			]Total
	a	b	c	
Large, Small, Sample, N, Population, Sample Size, small sample, size, entire	20	2	14	36 (11.29%)
Mean, Sampling distribution, population, same	6	14	22	42 (13.17%)
same, size, sample, stay, population, standard deviation	0	32	0	32 (10.03%)
decrease, increase, sample size, size	14	1	4	19 (5.96%)
more, people, spread, sample, little, variability	19	0	18	37 (11.60%)
correct formula, less, correct, formula, sigma	2	0	31	33 (10.3%)
distribution, sampling, lower, small, sample	8	2	14	24 (7.5%)
great, number, sample, employee, age, population, mean, small	12	7	21	40 (12.5%)
n, sqrt, deviation, standard, population, small, high, people	8	4	13	25 (7.84%)
sample size, same, large sample size, large, small sample size, small, data, less, stay	11	20	16	47 (14.73%)
NO CATEGORY	1	2	1	4 (1.29%)
NO RESPONSE	1	8	11	20 (6.27%)

Note: *Percentages are based on the total number of intuitors in the population (319)

Prior to conducting this analysis, rather than enter in the correct formula, I exchanged it for the phrase “correct formula”. This was to eliminate the keywords that might occur with individual portions of the formula. 10% of sensors and intuitors were assigned to the category topic containing the keyword “correct formula”.

Similar to the question regarding likelihood, the data seem to be distributed randomly between answers A and C. Answer B has more weight with the Topic mentioning the word “same” since it states that the standard deviations will be the same. Otherwise, the topics seem to contain what I would consider common keywords in regards to the size of a sampling distribution. There doesn't appear to be any major difference between the topics of the sensor versus the intuitor.

Shape.

> A large multinational company knows that the average age of their employees is 34 years. They also know that the standard deviation of the ages of these employees is 8 years. A manager from human resources is going to randomly select a sample of 100 employees. Consider the sampling distribution of the mean of a sample of size 100 from this population. From the information given, which of the following statements are true about this sampling distribution?
>
> Which of the following is true? Pick one and explain.
>
> a. We know that the shape will be right skewed.
> b. We know that the shape will be approximately symmetric.
> c. We cannot tell what the shape will look like.

Table 75. Text Mining, Shape: Topics, Intuitor

TOPIC Shape : INTUITOR	Response Choice			Total
	a	b	c	
1 - right, skew	9	0	0	9 (11.25%)
2 - normal, approximately	2	3	2	7 (8.75%)
3 - sample size, large, large sample size, resemble	1	6	0	7 (8.75%)
4 - know, standard deviation, mean	0	3	3	6 (7.5%)
5 - normal distribution, mean, sample distribution	0	8	0	8 (10.0%)
6 - symmetric, shape	0	4	0	4 (5.0%)
7 - population, resemble, skew	8	0	0	8 (10.0%)
8 - sample, shape	4	0	3	7 (8.75%)
9 - do not know, data, distribution, median	1	1	3	5 (6.25%)
10 - information, enough, sampling	0	1	4	5 (6.25%)
NO TOPIC	1	1	1	3 (3.75%)
NO RESPONSE	7	3	0	10 (12.5%)

Note: *Percentages are based on the total number of intuitors in the population (80)

Table 76. Text Mining, Shape: Topics, Sensor

TOPIC SHAPE - SENSOR	Response Choice			Total
	a	b	c	
1 - Right, skew, mean, population	37	2	4	43 (13.48%)
2- large, central limit theorem, normal, sample size, approximately, distribution	4	22	1	27 (8.46%)
3- median, know, outlier, mean, data, determine	7	2	28	37 (11.60%)
4- normal bell, bell, shape, normal, curve, sampling distribution, shape	6	26	4	36 (11.29%)
5- random, population, distribution, population distribution, sample, select	18	2	8	28 (8.78%)
6- average, look, standard deviation, sampling distribution, know	9	9	7	25 (7.84%)
7- sample, large, symmetric, enough, large sample	6	13	7	26 (8.15%)
8 - symmetric, distribution, mean, always, sample	4	15	2	21 (6.58%)
9- shape, sample distribution, population, mean, sampling	16	6	4	26 (8.15%)
10- sample, enough, data, different, distribution	5	3	10	18 (5.64%)
NO TOPIC	0	0	9	9 (2.82%)
NO RESPONSE	12	5	6	23 (7.21%)

Note: *Percentages are based on the total number of intuitors in the population (319)

The first thing that stands out to me is that the Central Limit Theorem is not a keyword for the Topics of the intuitor, nor is the phrase bell shaped mentioned anywhere in the intuitor column of topics. Once again, intuitors also do not mention outliers. A majority of the students mentioning the word skew selected answer “a”, which makes sense since the answer contains the word. It would be interesting to see what students who did not select “a” but mentioned “skew” were talking about, which only occurred with the sensors. A quick view of the data where a student mentions “skew” but does not make answer selection “a” is given in Tables 77 and 78.

Table 77. Text Mining, Shape: Response Subset, Sensor

Text Mining: mention Skew but choose answer other than "a"
• Do not know which values are picked, so could be skewed, or symmetric • Cannot tell shape because pop shape is right skewed • The numbers could be off or skewed depending on outliers of data • We do not have enough information on how old the different employees are or how they may be skewed • Do not know what the data look like, could be normal or skewed • Because Sampling Distribution may end up looking different than the population, we'd expect it to look normal but it could be skewed

A quick print of the data where "skew" was mentioned but the answer selection was not "a" revealed that there are 4 intuitors that also fell into this category. I suppose that they were double categorized in text mining and thus assigned to a category with a greater weight, hence not the one mentioning skew, which is appropriate since I would hope that an observation that mentions skew but does not choose answer "a" to hold greater weight in another Topic Category.

Table 78. Text Mining, Shape: Response Subset, Intuitor

Text Mining: mention Skew but choose answer other than "a". Intuitor.
• It is normal if population is normal but not necessarily skewed just because population is • Could be skewed we do not know • We do not know that the distribution is normal w/o first doing calculations to standardize it. It is possible the data from the sample could be skewed. • The total of all sampling distributions will be skewed but not the single sampling distribution

Almost all of these responses listed mention an uncertainty about the shape.

Text Mining Summary. Although we can see the use of words and phrases together to create a category, I am not able to tell much more information about the topic, such as; does it represent any sort of correct or incorrect reasoning? The answer to this question is simply, no, it does not. While text mining does give common groupings of words, terms and word phrases, it fails to provide any insight into what a student is thinking regarding their answer to the question and justification.

In comparison to grounded theory, as conducted in the previous section, text mining might be faster, but it is not a good candidate for further analysis other than providing a list of terms that are being used together with highest frequency. Phrases need to be observed together, in context, in order to make sense in light of answering the research question: Are there any differences between sensors and intuitors in how they use logic and reasoning in qualitative responses to sampling distribution questions?

The information as presented in the tables above does not shed any light on the reasoning behind the students thought process, only the keywords and short phrases being used by the students. While some words are more strongly used by one group over another, it does not give any bearing as to how the student is thinking about the problem. Additionally, there are some observations that were not categorized, and thus it would be still be necessary to categorize them by hand. Text mining might be useful to gather buzz words, useful for other analysis methods, but is not going to help answer my second research question.

Content Analysis.

The final method of qualitative analysis that I employed is Content Analysis using SAS Content Categorization. From this point forward I refer to the content analysis as Content Categorization. "SAS Content Categorization enables you to define a taxonomy of categories and concepts that develops and identifies metadata about your information." (SAS Institute, Inc., 2012). It allows you to write rules for categories that will be used to classify data into groups. A rule can be written to establish a concept you are looking for within a category.

For each of the four questions presented, I wrote rules to categorize a response as Correct, Incorrect, Needs More Information, and No Response. Some of the questions have a fifth category that will be discussed in the analysis, but these four mentioned are uniform for all of the questions. Each rule is listed in Appendix N. SAS Content Categorization assigns the responses to a rule. Multiple rule assignments are avoided by creating a rule and telling it that it cannot contain information that would assign it into another rule.

Here is a quick example that may not contain the entire rule as listed in the Appendix N.

Correct:

(OR, (SENT, "center", "mean"),

(AND, (OR, "average", "34", "Mean", "mu"), (NOT, "NOT"))

This rule states that if either of these items occur, the text shall be categorized as "Correct": The sentence contains the words "center" and "mean" OR, one of the words "mean" "average" or "34" and not the word "not".

The next text details a sample of an incorrect rule:

Incorrect:

(OR,

(SENT, "NOT", (OR, "mean", "mu", "34"),

(SENT, "outliers"),

(AND, "different", (OR, "average", "center", "mean"),

This rule states that if the text contains either of the three items, "not" and one of the following words, "mean", "mu" or "34", the sentence contains the word "outliers" or, the text contains the word "different" and one of the following words "average", "center" and "mean".

I used some of the categories that emerged in my grounded theory analysis to help establish the rules and then allowed the data to help write a few more. I also used common misconceptions to the rules for incorrect reasoning. After a few iterations of testing the rule on the categories I updated the rules to attempt to encompass those that were left blank.

At first some of the responses were assigned to multiple categories, but I re-wrote the rules by adding the following logic, If A + B then not A. conversely, if B+A then B could still be true. For example if a sentence contained correct logic AND incorrect logic, then the category assigned would be Incorrect. I also had to add in logic that stated, if A and NOT B then A, such as, if correct and not incorrect, then correct. Any of the phrases in the incorrect logic would discount a correct answer.

Here is an example of a sentence that would follow this rationale.

"I would expect it to be centered at the mean, but I do not know for sure because there could be outliers" This student has expressed the sentiment that the center should be at the mean, and had they stopped writing, then they would have shown correct reasoning. However the student did not stop writing and decided that they could not know for sure due to the presence of outliers.

Likewise, if the student mentions needing more information, it doesn't matter what they state about the rest of the problem, it will be categorized as needing more information to solve.

I detail the results for each of the four questions as presented in the previous sections. I randomly selected 10 observations to make sure that the text was in agreement with the category. I attempted to answer this question in this analysis: Are there any differences between sensors and intuitors in how they use logic and reasoning in qualitative responses to sampling distribution question?

Finally, I compared Content Categorization analysis to the Grounded theory analysis.

Likelihood.

According to a recent report the average income of people in Adair County, Missouri is $31,023. Which of the following is more likely? Pick one and explain.

a. We take a random sample of 10 people from this county and find that the average is over $50,000.

b. We take a random sample of 200 people from this county and find that the average is over $50,000.

c. We do not have enough information to answer this question.

I created four categories for this question: Correct, Incorrect, More Information and No Response.

Overall Results. I created four categories using Content Categorization: Correct, Incorrect, More Information and No Response. A fifth category, "uncategorized" was created by default and contained any responses that did not fall into one of the four categories. The results for both intuitors and sensors are listed in Table 79.

Table 79. Content Categorization, Likelihood.

RESPONSE	Intuitor	Sensor	Total
Correct	33 (41.25%)	104 (32.60%)	137 (34.34%)
Incorrect	5 (6.25%)	19 (5.96%)	24 (6.02%)
More Information	15 (18.75%)	66 (20.69%)	81 (20.30%)
No Response	2 (2.5%)	5 (1.57%)	7 (1.75%)
Uncategorized	25 (31.25%)	125 (39.18%)	150 (37.59%)
TOTAL	80	319	399

Note: Numbers in Parentheses are percentages based on column totals

One of the issues with this question was the complexity of the responses. One way that I could have worked around this, perhaps, would have been to perform the analysis for each multiple choice response, rather than lumping all three responses together, however, this is not the approach I took to this problem and thus will only be listed in the next chapter.

The 2x2 Fisher's Exact Test (Correct Reasoning versus Not Correct) yields a p-value of 0.1498 which is not significant. We must also take into account that over 37% of the data is uncategorized, and thus this is not likely an accurate measure. The 2x5 test yields a p-value of 0.5103 indicating it is not significant in favor of the table probabilities being equal.

While the counts in this table are very similar between groups, there is a larger number of intuitors categorized into the "Correct" category. This was the largest difference between categories assigned when comparing sensors to intuitors. I must point out that the categories are not correctly assigned for all of the responses as noted below, and over 37% of the observations are not categorized, so I do not put any weight on this comparison since so much of the data is categorized incorrectly or not at all.

Table 80 is a simple 2x2 table of correct responses and correct answer (a) and a Fisher's Exact Test Performed on the data shows, with a p-value of 0.2295, that there is no significant difference between sensors and intuitors in providing both a correct reason and a correct answer in the question involving the likelihood of an outcome.

Table 80. Content Categorization, Likelihood: Correct Reasoning and Correct Answer

LIKELY MEAN	Intuitor	Sensor
Correct	30 (37.50%)	97 (30.41%)
Incorrect	50 (62.5%)	222 (69.59%)
Total	80	319
		p-value 0.2295

Note: Numbers in Parentheses are percentages based on column totals.

An additional hypothesis that I tested: Are sensors more likely to use correct reasoning but choose an incorrect answer? This is due to the sensors nature of fact memorization, but inability to think abstractly, meaning, the sensor is less likely than an intuitor to apply a definition to answer selection. In order to test this, I only created a 2x2 table of Incorrect Response and Correct Reasoning versus the other three combinations namely: Incorrect Response and Incorrect Reasoning, Correct Response and Correct Reasoning, and Correct Response and Incorrect Reasoning, called “Other Combinations”.

Table 81. Content Categorization, Likelihood: Correct Reason and Incorrect Answer

Reasoning and Response	Intuitors	Sensors
Correct Reason with Incorrect Answer	3 (3.75%)	7 (2.19%)
Other Combinations	77 (96.25%)	312 (97.81%)
		p-value=0.4266

From this resulting p-value of 0.4266, we do not reject the null hypothesis which states sensors and intuitors are equally likely to provide a correct response and make an incorrect answer.

The answers involving likelihood are complex due to the nature of the question. The student is being asked to predict the likelihood of a mean for two sample sizes and discuss. The student often mentions both samples in the response as well as the standard deviation and uses words like smaller, larger, more variable, less variable, all within the same sentence. In my opinion, this makes it harder for sophisticated machines to categorize, because the words can go into either of the categories. Table 82 illustrates a subset of responses and the dimension of the student's learning style. The subset of response were randomly selected using PROC SURVEYSELCT and simple random sampling, sample size = 4.

Subset of Responses.

Table 82. Content Categorization, Likelihood: Subset

CATEGORY	RESPONSE	Learning Style
CORRECT		
	More likely to pick in small sample	Intuitor
	Smaller sample size more susceptible to variability and more likely to stray from the true average	Intuitor
	Very unlikely for n=200, average be much greater than mu, however if take small sample size n=10 it very well may be that our sample average would be higher because didn't take a representative sample	Sensor
	margin of error and standard deviation error are greater with a smaller sample size so there is more variance in the salary and a higher chance that the salary will show up higher than it actually is.	Sensor
INCORRECT		
	more likely to have less conflicting outliers in large sample so more likely to be > 50K	Intuitor
	Smaller sample more likely, large is more like population mean	Intuitor
	because larger the sample the more likely the average income will be over 50K	Sensor
	to conduct a study need a bigger sample than 10 or 200 people	Sensor
MORE INFO		
	not enough info to know how to skew the data	Intuitor
	we do not have enough info	Intuitor
	only know the average and not the range. Need more info	Sensor
	do not know the number of people who are accounted for. Need more info	Sensor
UNCATEGORIZED		
	We have no reason to believe that the average of 31K is not accurate	Sensor
	this would be simple random sample and enough sample size	Sensor
	sample size needs to be proportional to the area being studied, 10 people would not be a sufficient sample of the population	Intuitor
	we have no idea how many people live in the county	Intuitor

Several of the responses labeled "incorrect" were actually correct responses. This question had the most number of uncategorized responses as well as the most incorrectly categorized. I believe this is because the students often used the same terms to describe both instances. This is a situation where writing rules for each answer selection would be useful since often the student describes the reason but does not repeat the answer they chose, thus context clues are useful, such as knowing the response chosen at the time the rules are written. This could reduce some of the confusion between answers.

Comparison with Grounded Theory. In my grounded theory results, I had four categories, Correct, Incorrect, Needs More Information, and No Response. These align with the Content categorization categories that I created and Table 83 displays the comparison of the two groups. To compare the two analysis methods, I created a variable called GT_likelihood and assigned it to the value "Correct" if the student's grounded theory response fell into one of the six categories that were described as correct responses in the grounded theory section; likewise an incorrect response was assigned to the value of "Incorrect". Responses requiring more information fell into the "More Information" category, and finally those failing to provide a response fell into the "No Response" category.

For comparison purposes, I used PROC COMPARE in SAS and compared the variables GT_likelihood and CC_likelihood, where GT_likelihood is one of four categories from grounded theory analysis and CC_likelihood is one of four categories form content categorization analysis. I omitted the analysis where the CC_likelihood variable was missing since 150 observations were uncategorized. There were 45 variables that did not compare equal. I spent my time analyzing each response to see which method categorized the values correctly.

Here is one example where content categorization removed the bias that knowing the response would bring in. The following student chose answer "c", "Not enough information". Here is the student's justification: "A larger sample more likely to have true mean, smaller could either be bigger or smaller than the mean." While the student exhibits understanding that a smaller sample will be less likely to hit the true mean, they fail to apply this knowledge in context, even though the answer is in their reply. Essentially the student states that the smaller sample is more likely to have a value greater than or smaller than the true mean. The question asks the student which sample size is more likely to have a mean greater than the population mean, explicitly by stating, greater than 50K. I am not surprised that this student was a sensor since they failed to make the connection despite knowing the facts. Content Categorization labeled this as “Correct” whereas Grounded theory labeled it as “More Information” which was consistent with the answer selection but not the student.

Conversely, this incorrect response “we have a larger sample to fulfill the expectations that the average income is high: the sample size allows us to cover more data” was categorized as “More Information” by content categorization, but as “incorrect” by grounded theory. The correct category would be incorrect since the student states they believe that a larger sample will have a larger mean. I believe content categorization caught this because of the words “more data” which is commonly used when the student expected more information will help them solve the problem.

Here is another example that content categorization correctly categorized: “More likely because it is larger and could be skewed. Need standard deviation to determine”. The student chose answer “b” stating they think the larger sample size is more likely to achieve a mean greater than 50K, but the response states they believe they need the standard deviation to determine. I categorized this as “incorrect” in my grounded theory since the student stated they expected the mean to be greater with the larger sample and chose answer “b”. I suppose either of these could be viewed as correct categorizations.

Here is a response that is categorized incorrectly by both methods “x-mu/ (8/(sqrt(n))) = (31023-34) /(8/sqrt(n)). I am still not sure how to categorize this. Content Categorization labeled it as “more information” and grounded theory incorrectly labeled it as “Correct”. The number 34 is mu from a previous problem and 8 is standard deviation from the other problem. As expected, this student is a sensor (prefers formulas and plug and chug) and is looking to plug in numbers to solve this problem. This is obviously an error on my part in the grounded theory.

In my breakdown of these 45 unequal categorizations, 21 were correctly categorized by Content Categorization and 21 were correctly categorized by Grounded Theory, one could have been categorized as either because the student used both incorrect reasoning and stated they needed more information. The remaining 2 were incorrectly categorized by both methods.

Table 83 illustrates the comparison of the Content Categorization and the Grounded theory. We would expect that if the methods are similar, the majority of the data would lie on the diagonal axis. I do not feel like Content Categorization did as good a job on this question as it did on the subsequent ones.

Table 83. Content Categorization vs. Grounded Theory: Likelihood.

Grounded Theory Category / Content Categorization Category	CORRECT	INCORRECT	MORE INFORMATION	NO RESPONSE	TOTAL
CORRECT	118 (29.75%)	3 (0.75%)	16 (4.01%)	0	137 (34.34%)
INCORRECT	16 (4.01%)	7 (1.75%)	1 (0.25%)	0	24 (6.02%)
MORE INFORMATION	6 (1.5%)	3 (0.75%)	72 (18.05%)	0	81 (20.30%)
NO RESPONSE	0	0	0	7 (1.75%)	7 (1.75%)
Uncategorized	70 (17.54%)	27 (6.77%)	53 (13.28%)	0	150 (37.59%)
	210	40	142	7	399

Note: *Amounts in parentheses are population percentages

With the exception of the Incorrect\Correct alignment, the bulk of the data do exist on the diagonal of this table. We still cannot account for the uncategorized questions and it seems that when there was a mismatch between content categorization and grounded theory, 50% of the time either one was right. This goes to show that it takes a lot of careful observation in the analysis, regardless of the method you use.

In the end, I was unable to determine if there were any differences between sensors and intuitors in how they use logic and reasoning in qualitative responses to sampling distribution questions because the results I achieved using Content Categorization were not ideal.

Center.

A large multinational company knows that the average age of their employees is 34 years. They also know that the standard deviation of the ages of these employees is 8 years. A manager from human resources is going to randomly select a sample of 100 employees. Consider the sampling distribution of the mean of a sample of size 100 from this population. From the information given, which of the following statements are true about this sampling distribution?

Which of the following is true? Pick one and explain.

a. The sampling distribution will be centered at 34.

Overall Results. I created four categories using Content Categorization: Correct, Incorrect, More Information and No Response. A fifth category, "uncategorized" was created by default and contained any responses that did not fall into one of the four categories. The results for both intuitors and sensors are listed in Table 84.

Table 84. Content Categorization, Center

RESPONSE	Intuitor	Sensor	Total
Correct	37 (46.25%)	149 (46.71%)	186 (46.62%)
Incorrect	13 (16.25%)	79 (24.76%)	92 (23.06%)
More Information	12 (15.00%)	55 (17.24%)	67 (16.79%)
No Response	10 (12.50%)	23 (7.21%)	33 (8.27%)
Uncategorized	8 (10.00%)	13 (4.08%)	21 (5.26%)
TOTAL	80	319	399

Note: Numbers in Parentheses are percentages based on column totals

Most of the numbers in Table 84 are about the same proportion, except that more sensors than intuitors gave an incorrect response. Since our question of interest is to see if there are any differences in how intuitors and sensors use logic and reasoning in qualitative responses to sampling distribution questions, I performed a Fisher's exact test on the 2x5 table using PROC FREQ in SAS. The p-value is 0.0767 indicating there is no significance difference between sensors and intuitors in how they were categorized, at the 0.05 level meaning the probabilities are equally likely between groups.

Table 85 is a simple 2x2 table of correct responses and correct answer (a) and a Fisher's Exact Test Performed on the data shows, with a p-value of 1, that there is no significant difference between sensors and intuitors in providing both a correct reason and a correct answer for this question involving the center of the sampling distribution of means.

Table 85. Content Categorization, Center: Correct Reasoning and Correct Answer

CENTER	Intuitor	Sensor
Correct	36 (45.00%)	144 (45.14%)
Incorrect	44 (55.00%)	175 (54.86%)
Total	80	319
		p-value 1.0000

Note: Numbers in Parentheses are percentages based on column totals

An additional hypothesis that I tested: Are sensors more likely to use correct reasoning but choose an incorrect answer? This is due to the sensors nature of fact memorization, but inability to think abstractly, meaning, the sensor is less likely than an intuitor to apply a definition to answer selection. In order to test this, I only created a 2x2 table of Incorrect Response and Correct Reasoning versus the other three combinations namely: Incorrect Response and Incorrect Reasoning, Correct Response and Correct Reasoning, and Correct Response and Incorrect Reasoning, called “Other Combinations”.

Table 86. Content Categorization, Center: Correct Reason and Incorrect Answer

Reasoning and Response	Intuitors	Sensors
Correct Reason with Incorrect Answer	1 (1.25%)	5 (1.57%)
Other Combinations	79 (98.75%)	314 (98.43%)
		p-value=1.000

From this resulting p-value of 1, we do not reject the null hypothesis which states sensors and intuitors are equally likely to provide a correct response and make an incorrect answer.

Subset of Responses. This question was fairly easy to categorize and unlike the previous question, only 5.26% of the observations did not receive a category. Table 87 displays a subset of responses and the learning style. I used PROC SURVEYSELECT in SAS and took a simple random sample of size 5 using the Learning Style as Strata, meaning that I took a sample of size 5 for both sensors and intuitors. I listed four of the responses below, two for each dimension of the Sensing/Intuitor category.

Table 87. Content Categorization, Center: Response Subset

CATEGORY	RESPONSE	Learning Style
CORRECT		
	because Sampling Distribution has mean mu and standard deviation sigma/root(n)	Intuitor
	Average=mean	Intuitor
	centered at mu, the Sampling Distribution allows us to examine the variability in sample statistics, including how they differ	Sensor
	Center at average	Sensor
INCORRECT		
	The employees chosen could exclude some of the outliers, which would affect the mean.	Intuitor
	Average age is 34 and simple random sample of 100 follows normal distribution then average of the sample should be the same	Intuitor
	xbar has a mean mu not mu_0	Sensor
	cannot know center just by mean alone, need median as well	Sensor
MORE INFO		
	we do not know whether or not the data are normally distributed	Intuitor
	do not know center because do not have xbar	Intuitor
	not enough information	Sensor
	Do not know Sampling distribution because do not' have an X or Y (z-score) to prove answer, also sigma is at 8, so it cannot be centered at 34, 34 is the mean. Ok so we have n large and if n large it's normal but don't know.	Sensor
Uncategorized		
	because we cannot figure that out with the information we're given	Intuitor
	the sample size is large thus the sample distribution would be centered.	Intuitor
	because the sample could consist of the older proportion of workers or vice versa	Sensor
	multinational so the sample will select people from any area and thus not be concise	Sensor

From what I observed in this random sample, there are no incorrectly categorized values, with the exception of those not assigned to a category. The reason these are not categorized is due to the kind of language used, or the poorly phrased responses. The students do not use the words like spread or outliers or do not know, so these were not easily categorized alone with the other observations, but with only 21 observations uncategorized, it would be easy to go back and categorize them manually. This result is listed in Appendix O.

In the incorrect category, the response “Average age is 34 and simple random sample of 100 follows normal distribution then average of the sample should be the same” appears to be incorrectly categorized. The student correctly selected answer “a” and told me that the center should be at the average, but the student also stated that this was a simple random sample of size 100 and thus normal, when actually it is a sampling distribution, which is a collection of sample means from samples of size 100. So while this student sounds like they know what they are talking about, the fact that the student stated this is a simple random sample leads me to believe that they have confused sample with sampling distribution. Perhaps the student used incorrect terminology and actually meant “sampling distribution” rather than “sample”. There is no way that we can tell this information from what we are given here. This is a common misconception amongst students in introductory statistics that we continue to observe in our data.

Comparison with Grounded Theory. For the grounded theory results, I have four categories, Correct, Incorrect, Needs More Information, and No Response. These align with the Content categorization categories. Table 90 displays the comparison of the two groups.

To compare the two analysis methods, I created a variable called GT_center and assigned it to the value “Correct” if the student’s grounded theory response fell into one of the three categories that were described as correct responses in the grounded theory section, likewise an incorrect response was assigned to the value of “Incorrect”. Responses stating the need for more information fell into the “More Information” category, and finally those failing to provide a response fell into the “No Response” category. For comparison purposes, I used

PROC COMPARE in SAS and compared the variables GT_center and CC_center, where CC_center is the content categorization center category (SAS Institute, Inc., 2012).

There are 58 differences between the grounded theory analysis and the content categorization, not including the uncategorized responses. Since "MORE INFORMATION" and "INCORRECT" can be viewed as similar categories, both describing a type of incorrect response, I limited further comparison only to observations where one was categorized as correct and the other as incorrect. I felt like this was more useful than comparing two very similar categories. Leaving out the more information and incorrect differences, there are 22 observations that compare different, four of which are marked No Response by only one method.

There were 37 observations marked "NO RESPONSE" in grounded theory and only 33 in content categorization. I printed the observations where either of the CC_center or GT_center variables was categorized as "no response" but not both. I expected to see only two observations in the output, but as listed in Table 88, there were four observations where only one of the methods found this to be NO RESPONSE. The actual responses, as written by the student are listed in Table 88, followed by the categorization by both Grounded Theory and then Content Categorization. I then gave my interpretation of what I actually expect the categorization to be.

Table 88. Content Categorization vs. Grounded Theory: Center, No Response

RESPONSE	GT_center	CC_center	ACTUAL
centered at the mean because it is a sample mean	NO RESPONSE	Correct	Correct
since a & b are the same I choose c	NO RESPONSE		?
xbar has a mean mu not mu_0	NO RESPONSE	Incorrect	Incorrect
All of the information leads to this	NO RESPONSE		?

The first observation listed in Table 88 must have been a careless evaluation on my part in the grounded theory. The student gave a correct response. Content Categorization captured this. The second response was due to a bad final exam (where the answer choices for both a and b were both "The mean is centered at 34" so by default the student opted for the one unique choice, however I have no way to categorize this though I believe that this was a good use of logic and process of elimination. The third response listed is incorrect as captured by content categorization and the last response is partially correct, the n is large enough to get a normal distribution based on the central limit theorem, but the student does not go into enough detail for me to think this is sound reasoning.

The remaining 18 observations that compare differently, observe values in which one analysis method categorizes the response as correct and the other categorizes it as incorrect. Here is a confusing response: "[the] sampling statistic should capture the true mean and large enough sample size does not influence mean but outlier does". At quick glance, this student has a correct answer. They understand that this is a sampling statistic, but then the student mentions the presence of an outlier. The student gave the correct response, but it is confusing as to why they mention that the mean is affected by outliers. The use of the word outlier caused Content Categorization to categorize as an incorrect response whereas in grounded theory, I categorized it as Correct. I am still not sure which the better categorization method is. The following is also confusing to categorize “It does not change”. Since the question refers to the center of the sampling distribution, I assume the student means that the mean does not change, but I am sure this information is not enough evidence to mark as correct reasoning, nor do I think it is incorrect, but will categorize it as incorrect, since I have to choose a category, and a lack of compelling evidence showing understanding leads me to believe this is a lack of understanding.

Of the remaining 18 differences, 10 of them were correctly categorized by Content Categorization method. The student talks about a simple random sample rather than a sampling distribution or the distribution of a collection of statistics. For example, consider either of the following: “a random sample represents the population,” or “because the average of the population is 34, a random sample should also be close to this average.” Both are inherently true, but the student misses the big picture by viewing this as a simple random sample rather than a collection of sample means, and thus content categorization accounts for this. Table 89 displays the 18 differences and the “re-categorization”.

Table 89. Content Categorization vs. Grounded Theory: Center unmatched categories

Learning Style	Response	CC_center	GT_center	Additional Analysis
Intuitor	Average age is 34 and simple random sample of 100 follows normal distribution then average of the sample should be the same	Incorrect	Correct	Incorrect
Intuitor	Because the sample size is larger and random. Xbar is normal when n is large	Incorrect	Correct	Correct
Sensor	The mean of Sampling Distribution is close to true mean but no way to know if it is exactly 34	Correct	Incorrect	Incorrect
Sensor	Population center is 34, sample can be somewhere else	Correct	Incorrect	Incorrect
Sensor	Sampling Distribution can be treated as normal but we do not' know the center	correct	Incorrect	Incorrect
Sensor	since sample is large and random it can be considered a normal distribution which centers the mean around 34	Incorrect	Correct	Incorrect
Sensor	the SAMPLING DISTRIBUTION of the mean will be centered at 34 is true because the mean is always the center for sampling distribution along with sigma and whether or not it is normal	Incorrect	Correct	Correct
Sensor	It does not' change	Incorrect	Correct	Incorrect
Sensor	Although it does not state it is normal, they have actual mean of all employees (large sample size) so distribution should be centered around average	Incorrect	Correct	Correct
Sensor	Normal distribution so it should have a location and the center does not change	Incorrect	Correct	Correct
Sensor	sampling statistic should capture the true mean and large enough sample size does not influence mean but outlier does	Incorrect	Correct	Correct
Sensor	Large random sample is a good representative of the actual population and we'd expect to see it centered around 34 as well	Incorrect	Correct	incorrect
Sensor	it is a SIMPLE RANDOM SAMPLE so the mean is in the center bc histogram is symmetric	Incorrect	Correct	incorrect
Sensor	large random sample so it's the same	Incorrect	Correct	incorrect
Sensor	Because the average of the population is 34, a random sample should also be close to this average	Incorrect	Correct	incorrect
Sensor	SIMPLE RANDOM SAMPLE should be representative of population so Sampling Distribution should be centered at 34	Incorrect	Correct	incorrect
Sensor	Random sample of this size will represent the average age	Incorrect	Correct	incorrect
Sensor	Random sample rep. population	Incorrect	Correct	incorrect

The remaining 8 responses were correctly categorized using grounded theory, including the two mentioned as confusing and hard to categorize above. Fortunately, only 4.5% of the values did not match the grounded theory analysis, and of those, only 2% were incorrectly categorized based on additional follow-up analysis as indicated in Table 89.

Table 90 shows the overall comparison of the two analysis methods. It is desirable for the values to be large on the diagonal since this indicates that the methods are in agreement.

Table 90. Content Categorization vs. Grounded Theory: Center

Grounded Theory Category / Content Categorization Category	CORRECT	INCORRECT	MORE INFORMATION	NO RESPONSE	TOTAL
CORRECT	181 (45.36%)	3 (0.75%)	1 (0.25%)	1 (0.25%)	186 (46.62%)
INCORRECT	15 (3.76%)	71 (17.79%)	5 (1.25%)	1 (0.25%)	92 (23.06%)
MORE INFORMATION	0	32 (8.02%)	35 (8.77%)	0	67 (16.79%)
NO RESPONSE	0	0	0	33 (8.27%)	33 (8.27%)
Uncategorized	11 (2.76%)	6 (1.50%)	2 (0.50%)	2 (0.50%)	21 (5.26%)
	207	112	43	37	399

Note: *Amounts in parentheses are population percentages

Because of the association between “More Information” and “Incorrect”, I combined the groups to show a comparison of only three categories.

Table 91. Content Categorization vs. Grounded Theory: Center Simplified

Content Categorization Category \ Grounded Theory Category	CORRECT	INCORRECT	NO RESPONSE	TOTAL
CORRECT	181 (45.36%)	4 (1.00%)	1 (0.25%)	186 (46.62%)
INCORRECT	15 (3.76%)	143 (35.84%)	1 (0.25%)	159 (39.85%)
NO RESPONSE	0	0	33 (8.27%)	33 (8.27%)
Uncategorized	11 (2.76%)	8 (2.01%)	2 (0.50%)	21 (5.26%)
	207	155	37	399

Note: *Amounts in parentheses are population percentages

In this simplified table, the data along the diagonal show a majority of the data, with the exception of the 5.26% of data that is uncategorized and only a small percent of the observations are not in agreement with the grounded theory analysis. This gives promising hope for the use of SAS Content Categorization in future projects. It takes careful writing and review of the rules and categories over multiple iterations to make sure it has been done correctly, but can be done with much more speed than grounded theory and can then be applied to the same question in future classes. This question does not present any information regarding any differences between how sensors and intuitors using logic and reasoning in answering questions regarding sampling distributions.

Spread.

> A large multinational company knows that the average age of their employees is 34 years. They also know that the standard deviation of the ages of these employees is 8 years. A manager from human resources is going to randomly select a sample of 100 employees. Consider the sampling distribution of the mean of a sample of size 100 from this population. From the information given, which of the following statements are true about this sampling distribution?
>
> Which of the following is true? Pick one and explain.
>
> a The sampling distribution will have a higher standard deviation than the population.
>
> b The sampling distribution will have the same standard deviation as the population.
>
> c The sampling distribution will have a smaller standard deviation than the population.

Overall Results. First I created four categories using Content Categorization: Correct, Incorrect, Incorrect- SAME, and No Response. A fifth category, "uncategorized" was created by default and contained any responses that did not fall into one of the four categories. The results for both intuitors and sensors are listed in Table 92. I did not create a more information category for this question because it was not listed as any of the answer choices and did not appear in the grounded theory analysis.

Table 92. Content Categorization, Spread

RESPONSE	Intuitor	Sensor	Total
Correct	24 (33.33%)	120 (37.62%)	144 (36.09%)
Incorrect	16 (20.00%)	65 (20.38%)	81 (20.30%)
Incorrect - SAME	17 (21.25%)	75 (23.51%)	92 (23.06%)
No Response	15 (18.75%)	29 (9.09%)	44 (11.03%)
Uncategorized	8 (10.00%)	30 (9.40%)	38 (9.52%)
	80	319	399

Note: Numbers in Parentheses are percentages based on column totals

The category, Incorrect – SAME indicates that the student felt the sampling distribution would have the same standard deviation as the population. The Fisher's Exact Test on the 2x6 table yields a p-value of 0.1822, which is not significant. The 2x2 table of correct versus other classification give a p-value of 0.2416 and thus is also not significant. From Table 92 and the p-values from the analysis, I concluded that there is no difference in how sensors and intuitors reason about the standard deviation of the sampling distribution in comparison with the parent population.

Table 93 is a simple 2x2 table of correct responses and correct answer (”c”) and a Fisher's Exact Test Performed on the data shows, with a p-value of 0.0295, indicates that there is a significant difference between sensors and intuitors in providing both a correct reason and a correct answer for this question involving the standard deviation of the sampling distribution of means, in favor of the sensors being more likely to provide a correct answer.

Table 93. Content Categorization, Spread: Correct Reasoning and Correct Answer

SPREAD	Intuitor	Sensor
Correct	12 (15.00%)	85 (26.65%)
Incorrect	68 (85.00%)	234 (73.35%)
Total	80	319
		p-value 0.0295

Note: Numbers in Parentheses are percentages based on column totals

An additional hypothesis that I tested: Are sensors more likely to use correct reasoning but choose an incorrect answer? This is due to the sensors nature of fact memorization, but inability to think abstractly, meaning, the sensor is less likely than an intuitor to apply a definition to answer selection. In order to test this, I only created a 2x2 table of Incorrect Response and Correct Reasoning versus the other three combinations namely: Incorrect Response and Incorrect Reasoning, Correct Response and Correct Reasoning, and Correct Response and Incorrect Reasoning, called “Other Combinations”.

Table 94. Content Categorization, Spread: Correct Reason and Incorrect Answer

Reasoning and Response	Intuitors	Sensors
Correct Reason with Incorrect Answer	12 (15%)	35 (10.97%)
Other Combinations	68 (85%)	284 (89.03%)
		p-value=0.3333

From this resulting p-value of 0.3333, we do not reject the null hypothesis which states sensors and intuitors are equally likely to provide a correct response and make an incorrect answer.

Subset of Results. Table 95 displays a subset of responses and the learning style. I used PROC SURVEYSELECT in SAS and took a simple random sample of size 5 using the Learning Style as Strata, meaning that I took a sample of size 5 for both sensors and intuitors. I listed four of the responses below, two for each dimension of the Sensing/Intuitor category.

Table 95. Content Categorization, Spread: Subset

CATEGORY	RESPONSE		Learning Style
CORRECT			
	Sigma/sqrt(n) < sigma	c	Intuitor
	SAMPLING DISTRIBUTION is representative of population	b	Intuitor
	sample smaller than a population, so the standard deviation would be more dramatic and more noticeable	a	Sensor
	formula so smaller standard deviation for sample	C	Sensor
INCORRECT			
	with fewer samples (n) the variance will be higher	a	Intuitor
	because the sample size is smaller than the population, the standard deviation will be bigger.	a	Intuitor
	smaller standard deviation than population because the mean of the population is greater than the mean of the SAMPLING Distribution	c	Sensor
	the standard deviation will be larger because the sample size is smaller	a	Sensor
SAME			
	standard deviation should not change just because the mean changes	b	Intuitor
	the standard deviation should not change from the sampling distribution	b	Intuitor
	(guessing) it will be the same as the other because it's from the same data	b	Sensor
	The standard deviation should reflect the same standard deviation as the population	b	Sensor
Uncategorized			
	the standard deviation should still reflect the average even when the sample is 100	b	Intuitor
	with sampling distribution you are able to quantify the variability from the parameter	c	Intuitor
	because it has more subjects in it. Population should vary more	c	Sensor
			Sensor

Not all of these observations are correctly categorized as we can see in a few of those listed above. For example, the student categorized as correct stated that "the sampling distribution is representative of the population", which is an incorrect statement. Likewise the following response "sample smaller than a population, so the standard deviation would be more dramatic and more noticeable" is also incorrectly categorized. I think the software keyed in on the word "smaller" and did not take into account the context, such that the word small refers to the size of the sample, and not standard deviation. This is something that could be adjusted in the rules and taken into account. Similar to the likelihood question, this type of question might be better evaluated for each of the three multiple choice selections so that the context of the response takes into consideration the answer the student is trying to support.

Comparison with Grounded Theory. For the grounded theory results, I have four categories, Correct, Incorrect, Incorrect- Same, and No Response. These align with the Content categorization categories. Table 96 displays the comparison of the two groups.

To compare the two analysis methods, I created a variable called GT_spread and assigned it to the value “Correct” if the student’s grounded theory response fell into one of the three categories that were described as correct responses in the grounded theory section, likewise an incorrect response was assigned to the value of “Incorrect”, except when the incorrect response stated that the two standard deviations should be the same. Responses failing to provide a response fell into the “No Response” category. For comparison purposes, I used PROC COMPARE in SAS and compared the variables GT_spread and CC_spread, where CC_spread is the content categorization spread category.

Table 96. Content Categorization vs. Grounded Theory: Spread

Grounded Theory Category / Content Categorization Category	CORRECT	INCORRECT	INCORRECT-SAME	NO RESPONSE	TOTAL
CORRECT	71 (17.79%)	62 (15.54%)	11 (2.76%)	0	144 (36.09%)
INCORRECT	11 (2.76%)	65 (16.29%)	4 (1.00%)	1 (0.25%)	81 (20.30%)
INCORRECT-SAME	7 (1.75%)	15 (3.76%)	68 (17.04%)	2 (0.50%)	92 (23.06%)
NO RESPONSE	2 (0.50%)	3 (0.75%)	3 (0.75%)	36 (9.02%)	44 (11.03%)
Uncategorized	14 (3.51%)	19 (4.76%)	4 (1.00%)	1 (0.25%)	38 (9.52%)
	210	40	142	7	399

Note: *Amounts in parentheses are population percentages

This biggest misalignment in this comparison is the observations that are classified as Correct by content Categorization and Incorrect in grounded theory with nearly half of the observations being misclassified. One of the analysis methods has to be correct, so I went through them line by line to see which was actually correct, and found that of the 62 classified as Correct by Content Categorization and Incorrect by Grounded theory, 14 were actually correct and the remaining 48 were incorrect. Likewise, of the 11 categorized as Incorrect by Content Categorization, only 4 were incorrect and the remaining 7 were actually correct.

These properly assigned categories might make the significance of Table 93 less significant since I have a collection of incorrectly assigned categories. Overall, I suspect that grounded theory analysis did a better job assigning the categories, though with a little practice, I believe that content categorization could perform as good as, if not better than grounded theory. Due to this poor categorization, I suspect there is not a significant difference in Table 93, but rather, significance due to error. We have yet to observe a significant difference in how sensors and intuitors reason about sampling distributions.

Shape.

A large multinational company knows that the average age of their employees is 34 years. They also know that the standard deviation of the ages of these employees is 8 years. A manager from human resources is going to randomly select a sample of 100 employees. Consider the sampling distribution of the mean of a sample of size 100 from this population. From the information given, which of the following statements are true about this sampling distribution?

Which of the following is true? Pick one and explain.

a. We know that the shape will be right skewed.
b. We know that the shape will be approximately symmetric.
c. We cannot tell what the shape will look like.

I created five categories using Content Categorization: Correct, Incorrect, Parent, Need More Information, and No Response. A sixth category, "uncategorized" was created by default and contained any responses that did not fall into one of the four categories. The category, "parent" indicates that the student incorrectly believed the shape to resemble the parent population. The results for both intuitors and sensors are listed in Table 96.

Table 97. Content Categorization, Shape

RESPONSE	Intuitor	Sensor	Total
Correct	16 (20.00%)	79 (24.76%)	95 (23.81%)
Incorrect	10 (12.50%)	45 (14.11%)	55 (13.78%)
Parent	25 (31.25%)	88 (27.59%)	113 (28.32%)
More Information	9 (11.25%)	65 (7.52%)	74 (18.55%)
No Response	11 (13.75%)	24 (7.52%)	35 (8.77%)
Uncategorized	9 (11.25%)	18 (5.64%)	27 (6.77%)
	80	319	399

Note: Numbers in Parentheses are percentages based on column totals.

Overall Results. Comparison between groups is fairly proportional between the groups. The largest difference is between those failing to provide a response. I performed a Fisher's Exact Test on the entire 2X6 table and the p-value is 0.1559. The 2x2 table of Correct versus Not-Correct has a p-value of 0.4630. There is not any significance in the way sensors and intuitors answer this question. Only a small proportion of the data were left uncategorized.

Table 97 is a simple 2x2 table of correct responses and correct answer ("b") and a Fisher's Exact Test Performed on the data shows, with a p-value of 0.3735, indicates that there is not a significant difference between sensors and intuitors in providing both a correct reason and a correct answer for this question involving the shape of the sampling distribution of means.

Table 98. Content Categorization, Shape: Correct Reasoning and Correct Answer

SPREAD	Intuitor	Sensor
Correct	15 (18.75%)	77 (24.14%)
Incorrect	65 (81.25%)	242 (75.86%)
Total	80	319
		p-value 0.3735

Note: Numbers in Parentheses are percentages based on column totals

An additional hypothesis that I tested: Are sensors more likely to use correct reasoning but choose an incorrect answer? This is due to the sensors nature of fact memorization, but inability to think abstractly, meaning, the sensor is less likely than an intuitor to apply a definition to answer selection. In order to test this, I only created a 2x2 table of Incorrect Response and Correct Reasoning versus the other three combinations namely: Incorrect Response and Incorrect Reasoning, Correct Response and Correct Reasoning, and Correct Response and Incorrect Reasoning, called “Other Combinations”.

Table 99. Content Categorization, Shape: Correct Reason and Incorrect Answer

Reasoning and Response	Intuitors	Sensors
Correct Reason with Incorrect Answer	1 (1.25%)	2 (0.63%)
Other Combinations	79 (98.75%)	317 (99.37%)
		p-value=0.4899

From this resulting p-value of 0.4899, we do not reject the null hypothesis which states sensors and intuitors are equally likely to provide a correct response and make an incorrect answer.

Subset of Responses. Table 100 displays a subset of responses and the learning style. I used PROC SURVEYSELECT in SAS and took a simple random sample of size 5 using the Learning Style as Strata, meaning that I took a sample of size 5 for both sensors and intuitors. I listed four of the responses below, two for each dimension of the Sensing/Intuitor category.

Table 100. Content Categorization, Shape: Subset

CATEGORY	RESPONSE	Answer Selection	Learning Style
CORRECT			
	SAMPLING Distribution should be somewhat normal b/c sample size is large	b	Intuitor
	It has a normal distribution	b	Intuitor
	Central Limit Theorem	b	Sensor
	the sample mean should be approximately symmetric b/c should be normal	b	Sensor
INCORRECT			
	Right skew b/c outliers	a	Intuitor
	people under 18 not likely included but older would, thus skewing it to the right	a	Intuitor
	b/c the standard deviation of the employees is 8 and the average is 34, compared to age the it will skew right	a	Sensor
	skew right if mu > median and skew left if mu < median	a	Sensor
PARENT			
	the SAMPLING DISTRIBUTION will resemble the overall population given the large sample size and the mean will be skewed if the population is skewed	a	Intuitor
	sample shape should be like parent if it is normal, then sample should be normal b/c it used same data	a	Intuitor
	the sampling distribution estimates the population distribution	a	Sensor
	Represents the population	a	Sensor
MORE INFO			
	we don't know which 100 people we may choose and the ages may average higher or lower than 34	c	Intuitor
	we don't have enough info	c	Intuitor
	we must know more information of the sample.	c	Sensor
	no because we have yet to calculate the results	c	Sensor
Uncategorized		c	
	depends on sample's variability	c	Intuitor
	always know what shape will look like	b	Intuitor
	The point of the sampling distribution is to center it about the average	b	Sensor
	ages and the mean are related	a	Sensor

Comparison with Grounded Theory. For the grounded theory results, I have five categories, Correct, Incorrect, Incorrect-Parent, More Information and No Response. These align with the Content categorization categories. Table 101 displays the comparison of the two groups.

To compare the two analysis methods, I created a variable called GT_shape and assigned it to the value “Correct” if the student’s grounded theory response fell into one of the three categories that were described as correct responses in the grounded theory section, likewise an incorrect response was assigned to the value of “Incorrect”, except when the incorrect response stated that the shape should resemble the parent population. Responses stating more information was required fell into the “More Information” category, and those failing to provide a response fell into the “No Response” category. For comparison purposes, I used PROC COMPARE in SAS and compared the variables GT_ shape and CC_ shape, where CC_ shape is the content categorization spread category.

Table 101. Content Categorization vs. Grounded Theory: Shape

Grounded Theory Category / Content Categorization Category	CORRECT	INCORRECT	PARENT	MORE INFORMATION	NO RESPONSE	TOTAL
CORRECT	60 (15.04%)	28 (7.02%)	1 (0.25%)	3 (0.75%)	3 (0.75%)	95 (23.81%)
INCORRECT	4 (1.00%)	14 (3.51%)	28 (7.02%)	8 (2.01%)	1 (0.25%)	55 (13.78%)
PARENT	8 (2.01%)	7 (1.75%)	89 (22.31%)	6 (1.50%)	3 (0.75%)	113 (28.32%)
MORE INFORMATION	0	13 (3.26%)	3 (0.75%)	58 (14.54%)	0	74 (18.55%)
NO RESPONSE	0	0	0	0	35 (8.77%)	35 (8.77%)
Uncategorized	3 (0.75%)	7 (1.75%)	7 (1.75%)	10 (2.51%)	0	27 (6.77%)
						399

Note: Amounts in parentheses are population percentages.

There are 28 incorrect responses that were marked correct by Content Categorization. The only other major difference along the diagonal of the table was between incorrect responses and incorrect-parent response. These could be combined since both responses do indicate an incorrect response. Table 102 combines Parent and incorrect since they can be grouped together as an incorrect answer.

Table 102. Content Categorization vs. Grounded Theory: Shape Simplified

Grounded Theory Category / Content Categorization Category	CORRECT	INCORRECT	MORE INFORMATION	NO RESPONSE	TOTAL
CORRECT	60 (15.04%)	29 (7.02%)	3 (0.75%)	3 (0.75%)	95 (23.81%)
INCORRECT	12 (3.00%)	138 (34.59%)	14 (3.51%)	4 (1.00%)	168 (42.11%)
MORE INFORMATION	0	16 (3.26%)	58 (14.54%)	0	74 (18.55%)
NO RESPONSE	0	0	0	35 (8.77%)	35 (8.77%)
Uncategorized	3 (0.75%)	14 (1.75%)	10 (2.51%)	0	27 (6.77%)
					399

Note: Amounts in parentheses are population percentages.

The major differences between grounded theory and content categorization are now in the 29 marked correct by content categorization and incorrect by grounded theory. Similarly there are 4 responses marked incorrect by content categorization that are marked correct by grounded theory. Of the 29 marked correct by content categorization and incorrect by grounded theory, 9 are actually incorrect, the remaining 19 are correct. Likewise only one of the four marked incorrect is actually correct “Because we are dealing with means and not individual values, the distribution would be normal-symmetric”.

Despite the incorrect categorizations, content categorization does a pretty good job with this question regarding the shape of the sampling distribution. At the end of this analysis, I am able to answer that sensors and intuitors do not reason differently about sampling distributions, based on the data collected.

Content Analysis Summary. Most of the 399 observations were categorized for all four questions, with the exception of the question on likelihood; the categories were properly assigned, with a small margin of error. I would be able to quickly categorize those left uncategorized by hand since there was only a small proportion uncategorized except for the question regarding likelihood which might have been better categorized if I created rules based upon which answer the student selected.

In general, Content Categorization is useful if you have a good idea of how a student is going to answer a question. It does take some trial and error and some "tweaking", but I was able to write solid rules for most of the questions and feel that with a little more practice and experience using the software, I would be able to better categorize the questions involving spread and likelihood.

The reason I believe that those two questions were more difficult to categorize is because of the use of the words smaller and larger regarding sample size and standard deviation. The terminology within the question often refers to the answer selection and thus writing rules specific to how the student answered the question might prove beneficial for these two questions. The major issue I had with taking this approach and why I did not separate the rules for how the student answered the question is because I expected to see a student provide the right rationalization but an incorrect answer and conversely, incorrect rationalization and a correct answer. Had I separated the responses by how the student answered the question and then performed analysis, I might have done a better job at categorizing, but my goal was to assess the reasoning separate from how the student answered the question.

Overall, Content Categorization did a great job categorizing the responses and will save a lot of time if using this for exam questions en masse. Grounded theory can provide the foundation for writing the rules, but after a few iterations, the rules and categorizes can be generalized for future semesters on the same question, however lexical ambiguity and imprecise language can confuse the software and cause the user to have to create a longer list of synonyms. I found that content categorization was very useful and the results obtained were nearly the same as from the grounded theory, as shown in the analysis above.

Table 103. Content Categorization Summary

Data Collected	Semester	Analysis performed	Significant Results
Foundation: Likelihood	Fall 07 – Fall 08	Correct Reasoning & Correct Response	No
		Correct Reasoning & Incorrect Response	No
Foundation: Center	Fall 07 – Fall 08	Correct Reasoning & Correct Response	No
		Correct Reasoning & Incorrect Response	No
Foundation: Spread	Fall 07 - Fall 08	Correct Reasoning & Correct Response	*Yes (p=0.0295)
		Correct Reasoning & Incorrect Response	No
Foundation: Shape	Fall 07 – Fall 08	Correct Reasoning & Correct Response	No
		Correct Reasoning & Incorrect Response	No

*likely due to poor categorization of correct responses

Summary of Analysis Methods

The purpose of the qualitative analysis was to answer the question: Are there any differences between sensors and intuitors in how they use logic and reasoning in qualitative responses to sampling distribution questions?

All three analysis methods had strengths and weaknesses which are listed in Table 104 in a comparison of three qualitative analysis methods.

Table 104. Analysis Methods Summary

ANALYSIS METHOD	STRENGTHS	WEAKNESSES	Significance?
Grounded Theory	By hand so easier to interpret meaning.	Takes a long amount of time to do. Might assign wrong category - User error.	Likelihood question is significant at the 0.10 level in favor of Intuitors being more likely to provide a correct justification with a correct answer.
Text Mining	Good for grabbing key words and phrases being used most often by the students. Useful to create a synonym list for other analysis methods.	Not useful in determining understanding and reasoning.	N/A
Content Categorization	Quick and easy to use once rules are established. Fast. Relatively efficient. Eliminates bias caused by knowing how the student answered the question.	Learning curve in writing rules and using software. Might incorrectly assign wrong category – rule error. Terminology and slang are not easily detected and synonyms and words with double meanings have to be carefully coded and evaluated.	No significant results.

Grounded theory was probably the best method as it allowed the researcher to interpret the text in context to what the student is describing as well as lends to human interpretation of words and phrases that might not be commonly used in answering questions in an introductory statistics course.

While grounded theory has major strengths, it also has weaknesses as it is very time consuming and is not always a plausible analysis method when you have a 10 sections of 160 students each per semester, especially if you wish to categorize four questions for each student, yielding 6400 observations. Another weakness of grounded theory is that human error comes into play as the research is a fallible subject, despite best attempts to remain pure and unbiased. The researcher has to use their best judgment in determining if the answer is correct and can make mistakes. When pressed for time, this is not the quickest method and the less time you have, the more likely you are to make mistakes. For this research, time was not a factor, but human error still played a role. Grounded theory is useful in determining if the student used correct reasoning to justify an answer and did so effectively.

Text mining can be useful in determining words and phrases that students are using to answer the specific questions. It is quick and painless to use and parse over the large amounts of data. Text mining can be useful to generate lists of phrases and terms being used to describe a problem, but does not lend to understanding how a student reasons about an answer as the topics generated in text mining are simply words and do not indicate anything regarding understanding.

Finally, Content Categorization proved to be very effective and fast, though not without problems of its own. The problems regarding center and shape were easier to evaluate than the ones involving the likelihood and standard deviation. It is my belief that this is because the latter problems involved the use of the same phrases to describe two different scenarios, and without context of the answer choice, it would be easy to misclassify. I do believe that with some training, it will be possible to expand this method to all four questions and not just the two simple ones. I felt that over all, I did achieve some success using this method for analysis. Grounded theory will always win out, but we live in a fast paced world and it is not always efficient or probable. I believe that SAS Content

Categorization is on par with my grounded theory analysis, and perhaps even did better in some places, and thus would be considered an equal to grounded theory, just with added caution and careful checking. Once established, rules can be used for future exams on the same questions and will definitely save a lot of time.

Qualitative Analysis Summary

In this chapter I performed qualitative analysis on final exam data. I compared sensors and intuitors in how they answered questions about a sampling distribution in an attempt to answer the questions, are there any differences in how sensors and intuitors use reason and logic with their answers on questions about sampling distributions.

Secondly, in the Qualitative analysis methods, I observed significant differences between a sensor and intuitor, in favor of the intuitors, involving the question on whether a small sample or a larger sample is more likely to have a mean greater than the observed population mean. I believe that this is due to the abstract nature of the problem as it involves the conceptualization of the mean of a sample of two different sample sizes and estimating the likelihood of the sample mean. Sensors were less likely to view this abstractly and were more likely to believe that the larger sample would obtain a larger mean. I believe sensors looked at this in a simplistic way of reasoning rather than using abstract reasoning. Intuitors were more likely to understand that a smaller sample is more likely to be skewed and influenced by outliers and that a larger sample was more representative, reasoning abstractly. Both sensors and intuitors stated that they needed a larger sample to represent the entire population, but sensors applied this information incorrectly by stating that they were more likely to observe a mean greater than the population mean from taking a larger sample size. I believe that we observed this difference because sensors were able to memorize the definition and idea, but were unable to correctly apply this idea whereas the intuitors were able to correctly apply the concept.

The biggest conclusion from this qualitative data analysis is that a large number of students from both groups confuse a simple random sample with a distribution of sample means and apply the definition and properties of one to the other as though they are interchangeable. This seems to be the biggest underlying cause of all misconceptions. Students who thought the standard deviation should be larger since a sampling distribution has fewer observations than the population is thinking about the sampling distribution of means in terms of a simple random sample. This same statement can be made about the students who believe the sampling distribution will look like the parent population or that the center is unknown until we know what the data look like. Similarly, answers stating the need of more data lead me to believe that the student has conceptualized in terms of a simple random sample, which is greatly affected by the observations of which it is composed, rather than a sampling distribution of means, which exhibits properties of normality.

Table 105. Results Analysis Summary A

Research Questions	Data	Framework	Analysis
Are there any differences between sensors and intuitors in how they use logic and reasoning in qualitative responses to sampling distribution questions?	Free response data from final exam on questions regarding center, shape, spread and likelihood of a sampling distribution of means.	I expect that an intuitor is more likely than a sensor to rationalize about why they chose a certain answer and think in terms of a collection of sample means rather than think of a sampling distribution as a simple random sample. I expect a sensor to give a formula or recite a definition but not apply it successfully. This is because sensors are more likely than intuitors to memorize a formula and plug in numbers.	Grounded Theory and Content Categorization Analysis using Fisher's exact test to compare groups.

Table 106. Results Analysis Summary B

Semester	Data Collection	Analysis	Participants (#)
Fall 2006	Focus Group Pilot	Rudimentary analysis of Focus Group recording and notes	6
Spring 2007	Task Based Interviews	Analysis of Task Based Interviews	13
Fall 2007	Free Response Questions	Grounded Theory, Content Categorization and Text Mining on Free Response Questions	129
Spring 2008	Free Response Questions	Grounded Theory, Content Categorization and Text Mining on Free Response Questions	180
Fall 2008	Free Response Questions	Grounded Theory, Content Categorization and Text Mining on Free Response Questions	144

Table 107. Summary of Questions Examined in the Qualitative Analysis and Results

Data Collected	Semester	Significant Results
Foundation: Likelihood	Fall 07 - Fall 08	*Yes (p= 0.0801)
Foundation: Center	Fall 07 - Fall 08	No
Foundation: Spread	Fall 07 - Fall 08	No
Foundation: Shape	Fall 07 - Fall 08	No

* Significance found in Grounded Theory at the 0.10 level.

I collected a large amount of data about two different types of students, sensors and intuitors. Students' explanations about their answer selection revealed a lack of conceptual understanding of sampling distributions, which was not surprising. I was, however, surprised to see that students who exhibited strong understanding on one question failed to carry this understanding to the next question. Only 3% of the population showed correct understanding for all four of the foundational questions.

Essentially sensors and intuitors were just as likely to answer a question correctly as they were to provide justification for the answer, with the exception of the questions mentioned above. Students seemed to fail to exhibit a sound understanding of sampling distributions and even those providing a correct answer in one question failed to do so for all of the questions leading me to believe that overall, the students do not grasp the concept of the sampling distribution of means. What follows in the next chapter is a discussion of my conclusions, the limitations of my study and suggestions for further research.

Chapter 7: Conclusions, Limitations and Implications

Overview of Conclusions

Over the course of this research, I collected a substantial amount of information and data from an introductory statistics course, over a period of five semesters. In this paper I explored the answer to two questions. First: Are there any differences between sensors and intuitors in their responses to sampling distributions activities? Second: Are there any differences between sensors and intuitors in how they use logic and reasoning in qualitative responses to sampling distribution questions? If I were to sum up my conclusions from this research in an abridged manner, here is what I would say:

1) My first hypothesis was that intuitors are more likely than sensors to answer complex questions about sampling distributions. There is not enough evidence to suggest a difference between how sensors and intuitors answer questions about sampling distributions. I did find a small amount approaching significance on the question involving the likelihood of an outcome in favor of intuitors being more likely than sensors to correctly answer the question. If this is true, then the effort to teach students about the likelihood of outcomes from a given population, with respect to sampling distribution of the mean of various sizes should be improved.
2) My second hypothesis was that sensors are more likely than intuitors to solve a formula correctly. I rejected the null hypothesis that there is no difference between sensors and intuitors in favor of my alternative. I found adequate evidence to suggest that sensors are more likely than intuitors to solve a problem involving a formula or equation. Since the majority of our students are sensors, I

think this is something that is adequately taught in class, though intuitors should be encouraged to do more work on the rote memorization of the formulas since this is something an intuitor prefers not to do.

3) The third hypothesis was explored in the qualitative analysis suggesting that intuitors are more likely than sensors to give a correct answer and correct reason. I was only able to verify this hypothesis on the question involving the likelihood of an outcome. I rejected the null hypothesis that there are no differences between the groups in favor of our alternative hypothesis. For the remaining questions involving center, shape and spread; there was not enough evidence to suggest that intuitors outperformed sensors, thus I failed to reject the hypothesis.
4) My final hypothesis was that sensors are more likely than intuitors to give a correct reason but make an incorrect answer selection. There was not enough evidence to support this hypothesis in any of the four questions, and in fact, most of the students who exhibited correct reasoning also selected the correct answer.

Overall, I was unable to detect any large significant differences between sensors and intuitors in how they answered questions and reasoned about sampling distributions. I did find significant results in how the students answered a question involving the use of a formula in favor of the sensor being more likely to correctly solve the problem than the intuitor. It should be noted that I did find nearly significant results in both the qualitative and quantitative analysis regarding the likelihood of an outcome in favor of the intuitor.

Table 108. All Analysis Methods and Results

Data Collected	Semester	Significant Quantitative	Significant Qualitative
Graph: Simple Random Sample	Fall 06 - Spring 07	No	N/A
Graph: Simple Random Sample	Fall 07 - Fall 08	No	N/A
Graph: Sampling Distribution of Means	Fall 06 - Spring 07	No	N/A
Graph: Sampling Distribution of Means	Fall 07 - Fall 08	No	N/A
Foundation: Likelihood	Fall 07 - Fall 08	*No (0.10 level)	**Yes (Grounded Theory Analysis)
Foundation: Center	Fall 07 - Fall 08	No	No
Foundation: Spread	Fall 07 - Fall 08	No	No
Foundation: Shape	Fall 07 - Fall 08	No	No
Z-statistic – Range (Version 1)	Spring 08 – Fall 08	No	N/A
Range (Version 2)	Spring 08 – Fall 08	Yes	N/A
Z-statistic - Pr(X > xbar)	Spring 08 – Fall 08	Yes	N/A

*Approaching significance at the 0.10 level
** Significant at the 0.10 level

Limitations of the Study

There are several issues and concerns that I feel need to be discussed, that might be affecting this research.

Instrument Effect. The most obvious to me, would be the wording of the questions involving the sampling distribution of means.

The following question reads:

> A large multinational company knows that the average age of their employees is 34 years. They also know that the standard deviation of the ages of these employees is 8 years. A manager from human resources is going to randomly select a sample of 100 employees. Consider the sampling distribution of the mean of a sample of size 100

from this population. From the information given, which of the following statements are true about this sampling distribution?

It might be confusing the student to mention that we are going to randomly select one sample of size 100 and then ask them to consider the sampling distribution of the mean to answer the questions that follow. Many of the students, answered the three questions related to this text, in terms of a sample, rather than discussing the answer in terms regarding the sampling distribution of the mean. I am unable to differentiate between the shortcoming of the instrument and the lack of understanding by the student from the evidence obtained in this study.

I would be interested to see these series of questions administered in future final exams with the sentence “A manager from human resources is going to randomly select a sample of 100 employees” removed. This is to determine if the number of students talking about a simple random sample rather than a sampling distribution of means is reduced.

For the question involving the likelihood of a sample mean from two different sample sizes, some of the students were talking about the larger sample size being more likely, ignoring the fact that the population mean is given. In this question, I did not call it the “population mean” but simply stated that the “average income of the people” in the county was $31,023. Perhaps this is something that needs to be spelled out more clearly by calling it a population average, though I am not sure that this would have made a difference. Another interesting thing to point out regarding this question is that the third answer choice was: “We do not have enough information to answer this question,” which was surprisingly the most popular choice. It would be interesting to see what would happen if we removed this answer as an option, or changed it to: “Both outcomes are equally likely to observe a mean of $50,000”. Research has recently been published to state that when students are presented this option, in a question of likelihood, students are most likely to state that all outcomes are equally likely (Watson and Callingham, 2013).

The overall limitation of the instrument is that I was working within the context of a final exam rather than administering a free standing series of questions. Perhaps better information regarding the student’s understanding could have been obtained had I been able to present the same questions in multiple formats, in an effort to reduce the noise of the instrument itself. That way I could focus the instrument purely on the understanding of sampling distributions and its properties rather than having them nested inside a comprehensive exam.

With respect to the likelihood question, if I were able to ask this question in more than one way, I could vary the answer selections from: need more information, both outcomes are equally likely, or remove a third option all together and force the student to make a prediction. Does our presentation of answer choices affect the outcome?

The Instructor Effect. Another obvious factor to consider would be the teacher effect. ST311 is an introductory statistics course, designed by senior faculty, but taught by graduate students from the NCSU Statistics department. At times, this is the first course ever taught by the graduate student. The bulk of the data was collected over the course of three semesters and over approximately 11 sections, giving us 28 distinct courses being taught by 20 different teachers. Because the teaching assistants (TAs) are instructed to follow a rigid teaching plan that is carefully laid out, I did not perform any analysis to account for this effect.

The Researcher Effect. I must take into account personal bias when evaluating the data. While I did keep the student's learning style hidden when categorizing the response for each of the four problems for the 399 students, I did not hide the answer to the question as I found it useful in determining whether or not the student used correct reasoning. Often the student's response was vague enough that only by knowing which answer they selected was I able to properly categorize it. Knowing a student had a correct answer might have biased the categorization towards a correct response, giving them credit and reading between the lines, more or less interpreting on my own, what the student really meant to say, which might not have been unbiased. I believe that the Content Categorization eliminated this bias, and a merging of the two methods would be best in going forward for these types of analysis.

The Learning Style Effect. An additional question I must ask is whether or not I can generalize about the entire student population regarding the behavior and characteristics I observed through limiting my study to the learning style extremes. The study was restricted to the extreme preferences of a sensor and an intuitor in order to try to capture the true properties exhibited by these learning styles. Research has indicated that students not having extreme preferences, tend to be more adaptable to the other style of learning more so than the extreme counterpart. Having said this, I need to consider whether or not I can generalize my study and findings to the entire population.

Discussion

The understanding of sampling distributions remains a major problem for statistics educators despite recent advances of introducing technology into the classroom. Sampling distributions are the lynch pin for most of the topics in introductory statistics, such as confidence intervals and hypothesis tests, and it is often viewed as the key to statistical inference. Statistical reasoning is a difficult concept to assess, and is often misaligned with classroom performance as discussed in the literature.

It is difficult to create an assessment item to capture students' understanding of sampling distributions, which is why I wrote four questions involving free response, and included questions involving the interpretation of a graph as well as word problems involving the evaluation of a z-statistic. I was fortunate to find some questions from CAOS involving sampling distributions to model a few of my final exam questions. The foundational questions were written to assess whether the student possessed foundational knowledge deemed essential to the understanding of sampling distributions. The main issue with this is that the majority of the students, while selecting correct answers, did not provide a correct reason and thus the majority of the students failed to exhibit understanding, at least in written word, regarding sampling distributions, making it difficult to make connections between the foundation and the sampling distribution graphs.

What is clear from my analysis is that students are able to provide correct answers but unable to justify the reason for doing so. It is also clear that students are misusing terms and confusing the definition of sampling distribution with the concept of a sample. It is uncertain as to whether the students really do understand but use the wrong terminology, or do they confuse a sample for a sampling distribution? It seems to me that they are using the definition of sampling distribution of the mean and calling it a sample, and thus perhaps are just confused in the terminology, but I am unable to make a definitive conclusion based on these results.

From these findings, it seems that the ST311 course is being taught at a level that reaches both sensors and intuitors the same, and thus should not be catered towards a specific learning style, but instead, continued to be taught towards all preferences for learning. There is an exception on the questions involving formulas and likelihood.

For the questions involving a formula, intuitors could be encouraged towards better memorization. For the question on the likelihood of a specific outcome, we could increase the efforts of the instruction to improve this understanding, since significant or nearly significant results are found in both my qualitative and quantitative analysis. In regards to the likelihood question, it is interesting that we see a difference between sensors and intuitors in favor of the intuitor. I believe that the confusion between sample and sampling distribution is something that is lurking here, and would like to see this teased out in future research.

The major lexical ambiguity that I discovered involves the confusion of sample and sampling distribution. Many students talked about a sample in terms of a sampling distribution and a sampling distribution in terms of a sample, regardless of learning style. These students stated that they needed to see which values were selected in order to know the outcome (of the center, shape, or spread), in regards to the sampling distribution of the mean. The students talked about the likelihood of the sample mean being greater than the population mean in terms of more values (larger n) making an outcome more likely. I'm not certain whether this pertained to the confusion of sample and sampling distribution, or the blatant disregard of the given information about the population. What is interesting is that the attribute of a sensor is that they are more likely to pay attention to details, but if this were true, then the detail about the population mean would have been something they might have been more likely than an intuitor to observe, but was not the case in this problem.

While I was unable to fully support my hypothesis regarding the differences between sensors and intuitors, I believe we did find some useful evidence in how students are reasoning, or failing to reason, about sampling distributions. I also found it interesting that the students were looking for more information any time the option was presented to them rather than drawing their own conclusions. The student seemed to expect to see the problem solved for them, but when they were not presented with the option to state more data was needed, the student often found a solution to the problem, albeit incorrect at times.

It was disappointing that the blatant differences I expected to see between the intuitor and sensor did not occur. I do not believe that with the collection of more data, the differences will suddenly exist, thus my encouragement would be to stop looking at learning styles as an answer, but perhaps, put the focus on lexical ambiguity and focus more teaching around the major confusion between sample and sampling distribution.

Future research should focus on developing an instrument better aimed at gaining student understanding of sampling distributions. The ARTIST instrument is a valued and useful tool that we could consider implementing in our future tests, but there is little written regarding the validity and reliability of the tool.

Implications for Future Research

I think it would be useful to continue to explore the use of SAS Content Categorization, in future research in statistics education for qualitative data analysis. SAS Content Categorization is now packaged as SAS Contextual Analysis. I think with a little practice it will prove to be a useful tool in conducting statistical analysis on qualitative data. Once rules are established for a problem, it can be used for semesters to follow. I believe that SAS Contextual Analysis, is nearly equivalent to grounded theory analysis and much faster and easier to use. Caution should be used when writing rules by which to evaluate your data, but with practice and care, I believe it will, in the least, offer a solid companion to grounded theory analysis in data categorization and analysis and reduce the amount of “by hand” work.

I also think it would be wise to continue to explore this problem, perhaps not between sensors and intuitors, but maybe within the population of students as a whole. It would be interesting to explore why intuitors grasp the concept of the likelihood of a sample mean better than the sensor, and whether or not this has any bearing on the confusion between sample and sampling distribution. Teachers should use great care in discussing likelihood of outcomes and emphasize that all outcomes do not carry the same likelihood of occurring. I think this idea may be more abstract than the properties of normality in the sampling distribution of the mean, and thus exercises regarding the likelihood of outcomes being unequally likely should be covered.

It would be prudent to write questions comparing a sample to a sampling distribution in such a way that the students are forced to make a discernment regarding the two separate concepts, and at least force the students to discuss the concepts side by side, similar to the graphic questions presented, but perhaps, also in word form.

Finally, it is my belief after conducting this research, that lexical ambiguity and the confusion between two terms such as sample and sampling distribution is the most obvious avenue to explore. Students exhibited understanding of both concepts but failed to disassociate the separate properties, or think about them collectively in the right way. Teachers should use extra care when covering these definitions and emphasize the properties of each, first starting with a sample and then expounding to a sampling distribution of means. I think it would be useful for the teachers to continue to stress the differences between sample and sampling distribution of means throughout the course as well as educate the students about the common confusion in the terms. I think with knowledge comes power, and being aware of the common confusions between the terms is a good step in the direction of being able to overcome the obstacle.

Summary

I hope that my research has encouraged future researchers to conduct a mixed method analysis in a better attempt to understand the students and thought process driving the students' behavior in a statistics course. My research did not achieve the lofty goals I had set forth, but I do believe that a non-result is still a result in its own right. For instance, I did discover a few differences between sensors and intuitors, but the differences that I observed were not as extreme or obvious as I had hoped. I believe the introduction of Contextual Analysis in this framework brings a great contribution to this type of analysis and should be explored in future use cases. I found that mixed method analysis was crucial to exploring the questions I set forth, and I hope this motivates future teachers and researchers alike to continue to research how students understand sampling distributions in an effort to gain better insight into the problems students have with the introductory course. I believe that with better understanding will derive more efficient instruction, which, in turn, will result in students exiting the course with a deeper and more profound knowledge of the subject matter.

References

Aliaga, M., Cobb, G., Cuff, C., Garfield, J., Gould, R., Lock, R., & Witmer, J. (2005). GAISE college report. In *American Statistical Association*. Retrieved from www.amstat.org/education/gaise

Assessment Resource Tools for Improving Statistical Thinking (ARTIST) (2006). The CAOS Test. Retrieved from https://apps3.cehd.umn.edu/artist/caos.html.

Atkinson, P. & Delamont, S. (2006). Analytic Perspectives. *Handbook of Qualitative Research*, Ch. 32, 821-840.

Biggs, J. B. & Collis, K. F. (1982). Evaluating the Quality of Learning: The SOLO Taxonomy. New York: Academic Press.

Blalock Jr., H. M. (1987). Some General Goals in Teaching Statistics. *Teaching Sociology, 15*(2), 164-172. Teaching research methods and statistics.

Callaert, H. (1998). Stumbling blocks on the road towards statistical literacy. Limburgs Universitair Centrum.

Carlson, R. (1978). Review of Special Issue on Statistical Education. *The American Statistician*, *32*(4).

Chance, B., delMas, R., & Garfield, J. (2004). Reasoning about sampling distributions. In D. Ben-Zvi & J. Garfield (Eds.), *The Challenge of Developing Statistical Literacy Reasoning and Thinking*, 295-32). Springer. doi: 10.1007/1-4020-2278-6_13

Creswell, J. W. (2007). Qualitative Inquiry and Research Design: *Choosing among Five approaches Second edition.*

delMas, R., Garfield, J., Chance, B. (1999). A Model of Classroom Research In Action: Developing Simulation Activities to Improve students' Statistical Reasoning. *Journal of Statistics Education 7*(3). 1-19.

Elo, S., & Kyngäs, H. (2008). The qualitative content analysis process. *Journal of advanced nursing*, *62*(1), 107-115

Everson, M., Zieffler, A., & Garfield, J. (2008). Implementing New Reform Guidelines in Teaching Introductory College Statistics Courses. *Teaching Statistics, 30*(3). 66-70.

Felder, R. M. (1993). Reaching the Second Tier: Learning and Teaching Styles in College Science Education. *Journal of College Science Teaching, 23*(5), 286-290. Retrieved from http://www4.ncsu.edu/unity/lockers/users/f/felder/public/Papers/Secondtier.html

Felder, R. (1996). Matters of Style. *ASEE Prism, 6*(4), 18-23. Retrieved from http://www4.ncsu.edu/unity/lockers/users/f/felder/public/Papers/LS-Prism.htm

Felder, R., & Brent, R. (2005). Understanding Student Differences. *Journal of Engineering Education, 94*(1), 57-72.

Felder, R., Felder, G., & Dietz, E. (1998). A Longitudinal Study of Engineering Student Performance and Retention. V. Comparisons with Traditionally-Taught Students. *Journal of Engineering Education, 87*(4), 469-480.

Felder, R., &. Henriques, E. (1995). Learning and Teaching Styles in Foreign and Second Language Education. *Foreign Language Annals, 28*(1), 21-31.

Felder, R. M., & Silverman, L. K. (1988). Learning and Teaching Styles in Engineering Education. *Engr. Education, 78*(7), 674-681.

Felder, R., & Soloman, B. (1999). Learning Styles and Strategies. Retrieved from http://www4.ncsu.edu/unity/lockers/users/f/felder/public/ILSdir/styles.htm

Felder, R., & Spurlin, J. (2005). Applications, Reliability and Validity of the Index of Learning Styles. *International Journal of Engineering Education, (21)*1, 103-112.

Fleming, N. (n.d.a). The VARK Questionnaire. Retrieved from http://www.vark-learn.com/english/page.asp?p=questionnaire

Fleming, N. (n.d.b). Frequently Asked Questions. Retrieved from http://www.vark-learn.com/english/page.asp?p=faq

Fleming, N. D. (1995). *I'm different; not dumb. Modes of presentation (VARK) in the tertiary classroom*, in Zelmer, A., (ed.) Research and Development in Higher Education, Proceedings of the 1995 Annual Conference of the Higher Education and Research Development Society of Australasia (HERDSA), *HERDSA*, *18*, 308-313.

Gal, I. & Garfield, J. (1997). Curricular Goals and assessment challenges in statistics Education. *The Assessment Challenge in Statistics Education*, 1-13.

Garfield, J. (1998). The Statistical Reasoning Assessment: Development and Validation of a Research Tool. *Proceedings of the Fifth International Conference on Teaching Statistics*, ed. L. Pereira-Mendoza, Voorburg, The Netherlands: International Statistical Institute, 781-786.

Garfield, J. (2002). The Challenge of Developing Statistical Reasoning. *Journal of Statistics Education, 10*(3).

Garfield, J. (2003). Assessing Statistical Reasoning. *Statistics Education Research Journal, 2*(1). P 22-38.

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.

Garfield, J., & Chance, B. (2000). Assessment in Statistics Education: Issues and Challenges. *Mathematical Thinking and Learning, 2*(1&2), 99-125.

Garfield, J., delMas, R. (2010). A web site that provides resources for assessing students' statistical literacy, reasoning and thinking. *Teaching Statistics, 32*(1), 2-7.

Goldin, G. A. (2000). A Scientific Perspective of Structured, Task-Based Interviews in Mathematics Education Research. In Kelly, A., and Lesh, R. A. (eds.), *Handbook of Research Design in Mathematics and Science Education,* Lawrence Erlbaum Associates, Mahwah, NJ, 517-545.

Groth, R. (2010). Situating Qualitative modes of inquiry within the discipline of Statistics Education Research. *Statistics Education Research Journal*. *9*(2), 7-21.

Groth, R., & Bergner, J. (2005), Pre-Service Elementary School Teachers' Metaphors for the Concept of Statistical Sample, *Statistics Education Research Journal*, *4*(2), 27-42. Retrieved from http://www.stat.auckland.ac.nz/~iase/serj/SERJ4(2)_groth_bergner.pdf

Heid, M., Perkinson, D., Peters, S., & Fratto, C. (2005). Making and managing distinctions: The case of sampling distributions. In G. M. Lloyd, M. Wilson, J. L. M. Wilkins, & S. L. Behm (Eds.), Proceedings of the 27th Annual Meeting, International Group for Psychology of Mathematics Education – North American Chapter [CD]. Blacksburg, VA: Virginia Polytechnic Institute and State University.

Hogg, R. V. (1972). On Statistical Education. *The American Statistician, 26*(3), 8-11. Retrieved from http://www.jstor.org/stable/2682858

Kaplan, J., Fisher, D., & Rogness, N. (2009). Lexical Ambiguity in Statistics: What do students know about the words association, average, confidence, random and spread? *Journal Of Statistics Education. 17*(3), 1-19.

Kaplan, J., Rogness, N., & Fisher, D. (2012). Lexical Ambiguity: making a case against Spread. *Teaching Statistic, 34*(2), 56-60.

Kayes, D. C. (2005). Internal Validity and Reliability of Kolb's Learning Style Inventory Version 3 (1999). *Journal of Business and Psychology, 20*(2), 249-257. doi: 10.1007/s10869-005-8262-4

Kolb, D. A. (1984). Experiential Learning: Experience as the Source of Learning and Development. Englewood Cliffs, NJ: Prentice Hall.

Kolb, D. A., & Kolb, A. Y. (2005). The Learning Style Inventory Version 3.1. Boston, MA: Hay Resources Direct.

Konold, C. (1989). Informal Conceptions of Probability. *Cognition and Instruction, 6*(1), 59-98.

Konold, C. (1995). Issues in Assessing Conceptual Understanding in Probability and Statistics. *Journal of Statistics Education, 3*(1).

Kostoff, R., Morse, S., & Oncu, S. (2007). The Seminal Literature of Anthrax Research. *Critical Reviews in Microbiology, 33*(3), 171-181.

Krueger, R. A., & Casey, M. A. (2000). Focus groups: A practical guide for applied research. Sage Publications.

Lipson, K. (2003). The role of the sampling distribution in understanding statistical inference. *Mathematics Education Research Journal, 15*(3), 270-287.

Makar, K., & Confrey, J. (2005). “Variation-Talk”: Articulating Meaning in Statistics. *Statistics Education Research Journal*, *4*(1), 27-54.

Mathews, D., and Clark, J. (2003), "Successful Students’ Conceptions of Mean, Standard Deviation and the Central Limit Theorem," unpublished manuscript. Retrieved from http://www1.hollins.edu/faculty/clarkjm/papers.htm

McKenzie Jr. J., & Goldman, R. (2006). Questions to Assess the Understanding of Statistical Concepts. ICOTS-7, 2006.

Merriam Webster Incorporated (2013). Pedagogy. Retrieved from http://www.merriam-webster.com/dictionary/pedagogy

Metallidou, P., & Platsidou, M. (2008). Kolb’s Learning Style Inventory-1985: Validity Issues and Relations with Metacognitive Knowledge about Problem-Solving Strategies. *Learning and Individual Differences*, *18*(1), 114-119.

Mohd Nor, N., & Idris, N. (2010). Assessing Students’ Informal Inferential Reasoning using SOLO Taxonomy based Framework. *Procedia Social and Behavioral Sciences, 2,* 4805-4809.

Myers-Briggs Foundation (n.d.). How frequent is my type? Retrieved from http://www.myersbriggs.org/my-mbti-personality-type/my-mbti-results/how-frequent-is-my-type.asp

NC State University (2012). University Planning and Analysis. Retrieved from http://upa.ncsu.edu/univ/fact/student-statistics

Nisbett, R. (1993). Rules for Reasoning. Mahwah, NJ: Lawrence Erlbaum.

Pashler, H., McDaniel, M., Rohrer, D., & Bjork, R. (2008). Learning Styles: Concepts and Evidence. *Psychological Science in the Public Interest*, *9*(3), 105-119.

Petocz, A., & Newberry, G. (2010). On conceptual Analysis as the primary qualitative approach to statistics education research in psychology. *Statistics Education Research Journal*, *9*(2), 123-145.

Pittenger, D. J. (1993a). Measuring the MBTI… and coming up short. *Journal of Career Planning and Employment*, *54*, 48-52.

Pollatsek, A., Lima, S., & Well, A. (1981). Concept or Computation: Students' understanding of the mean. *Educational Studies in Mathematics*, *12*(2), 191-204.

Reading, C. & Reid, J. (2006). An Emerging Hierarchy of Reasoning About Distribution: From a Variation Perspective. *Statistics Education Research Journal*, *5*(2), 46-68.

Rice Virtual Lab in Statistics (n.d.). Sampling Distribution. Retrieved from http://onlinestatbook.com/stat_sim/sampling_dist/index.html

Rubin, A., Bruce, B., & Tenney, Y. (1991). Learning about Sampling: Trouble at the Core of Statistics. In D. Vere-Jones (Ed.), *Proceedings of the Third International Conference on Teaching Statistics*, *1*, 314-319). Voorburg, The Netherlands: International Statistical Institute.

Rumsey, D. J. (2002). Statistical Literacy as a Goal for Introductory Statistics Courses. *Journal of Statistics Education*, *10*(3).

Saldanha, L., & Thompson, P. (2002a). Conceptions of sample and their relationships to statistical inference. *Educational Studies in Mathematics, 51*(3), 257–270. doi: org/10.1023/A:1023692604014

Saldanha, L. & Thompson, P. (2002b). Students' scheme-based understanding of sampling distributions and its relationship to statistical inference. In D. Mewborn (Ed.), *Proceedings of the Twenty-fourth Annual Meeting of the International Group for the Psychology of Mathematics Education*. Athens, GA.

SAS Institute Inc. (2012). SAS Content Categorization Studio 12.1: User's Guide. Cary, NC: SAS Institute Inc.

SAS Institute Inc. (2013). SAS Text Miner 5.1: User's Guide. Cary, NC. SAS Institute Inc. Retrieved from http://www.sas.com/text-analytics/text-miner/index.html

Singleton Jr., R. (1989). On Teaching Sampling: A Classroom Demonstration of Concepts, Principles, and Techniques. *Teaching Sociology*, *17*(3), 351-355.

Sipps, G., Alexander, R., & Freidt, L. (1985). Item analysis of the Myers- Briggs Type Indicator. *Educational and Psychological Measurement, 45,* 789-796.

Tempelaar, D. (2004). Statistical Reasoning Assessment: an Analysis of the SRA Instrument. October 22, 2004, Presented at the ARTIST Roundtable conference on Assessment in Statistics held at Lawrence University, Aug 1-4. 2004.

Tempelaar, D. T., Gijselaers, W. H., & van der Loeff, S. S. (2006). Puzzles in Statistical Reasoning. *Journal of Statistics Education, 14*(1).

Vermunt, J. D. and Vermetten, Y. J. (2004), “Patterns in Student Learning: Relationships Between Learning Strategies, Conceptions of Learning, and Learning Orientations,” *Educational Psychology Review*, *16*(4), 359-384.

Watson, Jane and Callingham, Rosemary (2013). Likelihood and Sample Size: The understandings of students and their teachers. *The Journal of Mathematical Behavior, 32*(3), 660-672.

Watson, J. M., & Moritz, J. B. (2000). Developing Concepts of Sampling. *Journal for Research in Mathematics Education, 31*(1), 44-70.

Yang, Y., & Pedersen, J. (1997). "A Comparative Study on Feature Selection in Text Categorization.” *Proceedings of the Fourteenth International Conference on Machine Learning* (ICML'97).

Yu, C., Jannasch-Pennell, A. & DiGangi, S. (2011). Compatibility between Text Mining and Qualitative Research in the Perspectives of Grounded theory, Content Analysis and Reliability. *The Qualitative Report*, *16*(3), 730-44.

Zieffler, A., Garfield, J., Alt, S., Dupuis, D., Holleque, K., & Chang, B. (2008). What does Research Suggest About the Teaching and Learning of Introductory Statistics at the College Level? A Review of the Literature. *Journal Of Statistics Education*, *16*(2), 1-26.

Zieffler, A., Garfield, J., delMas, R., & Reading, C. (2008). A Framework to support research on informal inferential reasoning” *Statistics Education Research Journal*, *7*(2), 40-58.

APPENDICES

Appendix A - Informed Consent Form

ST311 Students

Redesign and Improvement of ST311

Dr. Roger Woodard, Principal Investigator

( ) I am 18 years of age or older

( ) I am less than 18 years of age

Statistics 311 is using new teaching methods to improve the teaching (and learning) of statistics. The purpose of this project is to use new teaching methods with improved lesson plans and computer based technology to better illustrate statistical ideas and concepts. You will contribute to this research by completing all regular class assignments and projects, and allowing researchers to have access to your work on these materials. In addition, you will complete the following:

1. Index of Learning Style Questionnaire.

2. Survey of personal attitudes and feelings towards statistics.

3. Exams that will assess your understanding of key statistical concepts and ideas covered in the course, including a comprehensive final exam.

In addition to the aforementioned items, several students may be asked to participate in a focus group, either via email, or in a face to face format. This will allow researchers to gain a deeper understanding of your reasoning and thought process while completing such tasks.

There will be no risk associated with your participation in the research study, Participation is voluntary. Your grade in the course will not be affected by your decision to participate in the study and the person(s) involved in grading assignments and exams will not have information regarding your participation (or therein lack of) or results of any voluntary assignments and surveys. The knowledge we gain from your experiences will add to the knowledge base of effective teaching methods in this course, especially with regard to how technology based methods affect overall understanding. The information derived from the class activities and assignments will be kept strictly confidential, with your name removed from the work. It will be stored securely in a locked file and will be made available only to the researchers unless you specifically give permission in writing to do so. No reference will be made to you name either in oral or written reports or through transcripts that could link you individually to the study.

You are free to withdraw from the study at any time; however you will still participate in all of the activities that are class requirements. If you have any questions at any time, you may contact Dr. Pam Arroway at 515-1955. Her address is 202 Patterson Hall, NC State University. If you feel as though you have not been treated according to the descriptions in this form, or your rights as a participant in research have been violated during the course of this project, you may contact Dr. Matthew Zingraff, Chairperson of the NCSU Human subjects Committee, Box 8101, NCSU Campus.

CONSENT

I have read and understood the above information. I have received a copy of this form. I agree to participate in this study.

Participants' Signature________________________________Date______________

Appendix B - Focus Group Fall 2006

Group1: (Three sensors participated)

Transcript of sensors Focus Group:

Over the past two weeks you guys have been looking at the sampling distribution

Do you recall an applet you looked at in class?

Participant1 and Participant2 : No. We did not do this in our class.

Participant3: I'm not sure.

Well I guess I will just have to show you. This is the Applet, Does this look familiar?

Participant3:Yes

Can everyone see this ok?

So has anyone else ever seen this before?

P2: I don't think we have. I've never seen that.

Since you have already seen this, can you tell us what this top graphic represented?

P3:The normal distribution

And this represents the parent population, consisting of all the data.

Now when I animate this, can you tell me what happens.

P3: like a child?

This is a representation of 5 observations being randomly drawn from the parent population.

What do you think the third graphic represents

P1: the mean

Now if I repeat this exercise, another mean will drop.

What do you notice about the distribution of the mean after taking 10,000 samples

P3: It's symmetric.

What is another way to term this?

P3: Normal

So what if I show you this distribution? (showing a skew) Have you guys looked at a distribution like this in your class?

P3: yes

What does this graph represent?

P1: Right skewed

So what do you think is going to be different about this distribution (animation) compared to the normal distribution?

P1: the mean will be skewed (others in agreement)

You think that, but let’s watch what happens. Let’s draw a sample of 25 observations.

Here is the mean. Let’s repeat this a few times.

so what do you notice about this distribution (pointing to the 3rd graph)

This is a sample taken from this right? (pointing to the 2nd graph)

the mean drops down, correct?

all: yes

So what do you notice about the distribution of the means of samples drawn from this parent population?

It is ...

p3: Normal

So when we run this animation 5 times, what do you notice?

P3: not normal

So what kind of conclusion can you draw about taking repeated samples and looking at the means?

P3: You have to take a lot of them to get anything..

anything?

P2 P3: a normal distribution

So what can you tell me about this left skewed distribution?

What would you expect to see after taking 5 samples?

The distribution of the mean would be expected to be....

P1: More varied than if you took a lot of samples

How many samples do you think it will take until I start to see a normal distribution?

P3: uh, 30

Where did you get that number from?

P2: we learned that in class maybe.

P3: It was just a guess, I don't remember

Have you covered normal distributions in class?

What have you covered most recently in your class?

P1: Probability

What about confidence intervals?

Do you think you can make any relation b/w confidence intervals and this graph?

Would you be able to see a confidence interval being drawn from this graph?

The population has a range between 0 and 32

P1: we have that chart in our book.

What is one value for any sample taken that we know cannot be true, or cannot occur from any random sample taken?

Is 5 a probable value that can come from this population?

P2: No probably not

P3: yeah (agreeing)

Ok, the range of the population is from zero to 32.

So what do you think? Can a sample of five contain the value 5 in it from this parent distribution?

It could

p3: yeah

So, how do you feel about your stats course in general?

P2: confusing

What are you confused about?

P2: The material. I guess we move too fast in class. I'm like overwhelmed

What about you

P3: it's not my favorite subject. I’m confused

How?

P1: I don't think it's necessarily hard, but the way our teacher presents the data is way too confusing.

So it's not the material, but how it the material is conveyed?

P1: He teaches it like a text book. Like a chemistry text book. They don't explain anything, they just give the facts and the numbers and send you on your way. My roommate has stats too, and his notes are really organized and his examples are just laid out so nicely. He comes home and teaches me.

ME: How do you feel about your teacher?

P3: I feel like she's taught it before, but she doesn't know how to present it (so that I understand it)

Are all your homework’s just webassign? Or do you have other kinds of homework’s

P3: we have webassign assignments and some in class assignments, that if we don’t' finish we can take home to complete

Do you feel like webassign makes it harder for you? Or do you have any feelings one way or the other?

Do you like webassign?

p3: no

p1: I like webassign because my roommate has the same stuff so he teaches me. The questions we get are random. But if the questions on the test are not on webassign, then I will not know how to do it.

So how does your teacher present the material in the class?

P2: He usually just writes stuff down and then everyone will be really confused and then he'll have to go and explain it.

What about your section?

P3: She normally tries to explain it

What do you think helps you learn things better? Do you like to see things acted out.

P2: Examples

Seeing this distribution is different from hearing it

P3: yeah

P1: um hum

Do you feel like you are getting a wide variety of examples in your classroom to cover all different situations and circumstances, or

P1: he gives examples but gives examples that he just makes up, so that they don't fit certain parts of the test.

Like p<= 15 or something like that. He just made it up. So then he's, like forget those steps. And then he went onto the next part of it

P2: I don’t know what he was doing, he'll give an example and then he'll mess up and be like. Wait.

so preparedness is lacking

P1P2P3: yeah

What do you think will help you out with this class?

P1: Well my chemistry class has given us a practice test last week, and this week we are going over the solutions.

I would like that. It would be helpful

P2:p3: yeah

So what things would you change of your courses if you could

p3: nothing. I'm ok.

p2: just how things are taught.

What so how would you want it taught? In what way?

p2: Just more examples and explain like what exactly he is doing

instead of just throwing it out there and kind of having us figure it out and that kind of thing.

Ok. What about you?

p1: he's the type of professor that writes everything out and goes through examples. But I think for him it would be better if he went through part of it, gave an example and then built from there instead of really taking it apart

I guess.

So you want the information required for the problem and then look at a real problem and then go into the information with a more detailed description of what it means?

P1-p3: yeah

Anything else you want to add?

p1: I don't think that stat is that hard, I think it is just the way that the information is being presented.

I can figure out the webassign problems myself most of the time.

P1: I think that is a good idea that all the stats classes are pretty much on the same page. I mean it helps that my roommate's class is covering the same material at the same time.

Anything else?

Silence

Thank you for coming!

Group2:

Focus group 2: (Two Intuitors participated)

What do you recall from the applet you saw in your class?

P1: demonstrating the higher the population the closer to the estimated probability it will be.

Is that ? Am I thinking about the right graph?

Ok, can you guys see this graph?

Looking at this graph, what does this top graph represent?

P1: the normal distribution?

P2: the population right?

p2: And that is one standard deviation? The red line?

What about this second graph. When I hit animate, what does this represent?

P1: the sample taken

Anything else?

P2: 5 of them?

What about the third graph?

P1: the distribution of the mean.

Ok so when I animate this 10000 times, what do you notice?

P1: the distribution is normal

Let's look at a different distribution. Do you recall the coin exercise you did in class?

Where you looked at the dates?

P1P2: yes we did

What did you remember about the distribution of the dates on the coin?

What is like the first graph or this second one?

P2: second one

How do you think that things would differ if you were to draw a sample from this distribution compared to drawing a sample from a normal distribution?

P2: it's going to be skewed as well? The sample?

Ok. So if I draw a sample from this distribution, what do you expect to see in these two lower graphs?

P2: Both distributions (of the sample drawn and the mean obtained) will be skewed as well

Ok. So let's see what happens when we run this animation

If I keep taking samples, you told me you expect the distribution of the samples to be high over here and lower on this side.

P2: um hum.

How many samples do you think I should take to get a familiar distribution from these means?

P2: is it one minus, it's like 15 is in there somewhere. I don't know

P1: Yeah we talked about like 30 or 40 is a large enough sample.

P2: doesn't that depend on what you are talking about though? Like what kind of graph?

Ok, so look at this graph, I have taken a large sample and plotted the means.

What do you notice about the distribution of these means from the skewed distribution?

P1: It's normal

So one thing you can learn from this exercise is that the distribution of the means of a sample drawn from a population will be normal.

Have you covered confidence intervals in your class?

Both: yes

Is there something you can state about confidence intervals by looking at this graph?

P2: you might be more sure that they will be on the side of the ... well I'm not too sure

Now I have some general questions about your class.

How do you feel your course is going in general?

P2: um, well for me I don't know like this is something I am taking as sort of a requirement and I feel that a lot of people in st311 are doing the same thing as I'm doing.

So I get the impression that it is not people's most important class, and it might be kind of like a chore to go as well

I do feel like there is a lot of material to cover and it's covered in a short period of time.

It's going well, but it's kind of like a struggle to get really into taking the class I guess. And I feel like that is how many feel. And sometimes our instructor is a little hard to understand. His handwriting is not very good and he understands things but as far as explaining things and putting things into perspective of real life that doesn't really come through sometimes and I think a lot of us are looking for that

P1: I feel the same way. Except I feel like my teacher is often times not prepared for class. She'll like go back over examples and change the numbers in the examples. She'll put an example on the board, and realize the numbers are a little difficult and she'll change it to something else. I pay attention in class, but it's really hard to follow what is going on (when she is always going back and changing things she just said or wrote down)

I feel like it is a lot of material that is very different. Like I'm having a hard time, originally the course was pretty easy.

Now I'm having a more difficult time understanding it, especially because I don't retain information from class in what she is teaching us.

I'm just really confused in class. And the webassign has only 3 submissions, but even with the book I'm struggling with doing the equations right and I'm just kind of lost and confused.

P2: I just write everything down in class and going home and when I'm doing my HW or webassign or studying, going back and trying to make the connections, but a lot of times it's just me writing down as much as I can off the board in hope that I can understand it later.

I've taken a stats course in stats, but it wasn't foreign to me. I can totally see how someone who has never heard these terms before would be really confused.

I just think they (the instructors) lack the skills to explain it properly and teaching skills altogether. I think they understand it but don't know how to explain it and put it into terms that would help me to see the big picture.

ME: When you get an example in class, how does that work?

P1: I have no idea what is going on in class right now

P2: we have a lot of in class assignments but we do have in class assignments and we work together, but they are not as hard as something on a test or webassign. But it's hard to do without notes. Sometimes he goes through them on his on, and people will ask him questions, but it seems hard to follow.

Have you guys used your books and tried to follow along to help supplement what you are doing in class?

P1P2: yes

(Note to self: none of the sensors said they were actually reading the book and both of the intuitors said they did use the book and try to do the examples)

P2: one of the things he has done recently is give us specific problems from the book to work on, but sometimes there isn't an answer so one thing hard is um, I guess I just have trouble doing things and I’m not sure if they are right or not but I’ve finished them, it's almost like I wish I had verification that I was doing this right.

Or maybe in the next class, at the beginning, if he would just work through the problems step by step so I could see if I had it right, or see where I made the mistake.

Have you guys tried the office hours? You can see any teacher at any time allotment. And maybe they will have a better fresher insight to help you understand in a way perhaps your instructor doesn't teach?

- I then go to show them that there are ST311 TA's in the office M-F 10-5 (or something like that)

Is there anything about this class that you like or that is helping you?

P1: when we do class activities I feel like I learn the most, like when we did the coin one. We actually had to calculate the numbers so we had to sit there and figure out what we were doing whereas in class we had to like try to figure out what is going on. I'm like along for the ride, and I’m just writing down what she's writing down and trying to understand while she talks us through it, but when we're given class activities I have to like do it on my own, and I get to discuss it with other people and we can figure it out how to work through it.

Do you feel like your peers are helping you to understand?

P1: yeah, together we're better, but we're all lost

So what would you change if you could?

P1: work in groups, more class activities, and more submissions on the webassign. Maybe 5 would be awesome.

P2: I think there is a lot of material to cover in a short period of time, so basically everyone goes home and works on this on their own, so if we had more study guides, more problems to pull from. Or a database of problems or a practice test to work on. I may get my webassign, but I am still unsure on stuff and still need more practice

P2: we need someone who has a different angle on things. A different point is sometimes all you need.

So the same information given to you in 5 different ways.

p2: because different people learn different ways

Right!

P1: And I’m sure it is something that will help us in the future, I mean one day we can relate back even when we're watching the news and the polls and stuff. I'm sure I’ll actually use this in the future.

Group 3: (One Intuitor)

A few weeks ago you looked at an applet in your class covering the sampling distribution.

Do you recall like from a website?

p6: She doesn't post anything like that; she always puts everything on the board.

She never does any kind of exercise?

p6: Well what do you mean by applet?

Ok have a look at this... (Turning the computer around). Have you seen anything like this in your class?

p6: Well that was pretty much like our whole test from last time the bell shaped curve and standardized tables and stuff

Ok so if I were to ask you what this top graph represents, you would say?

(12 seconds pass)

p6: I don't see a name besides Parent Population.

Well what kind of distribution would you say this was?

p6: oh a normal distribution

The parent population consists of all the data.

p6: Right

So the second graph is blank, but if I click on this button right here

It says draw a sample, and I'm going to draw a sample of five.

So can you describe what you saw happen on the second graph when I hit the animate button?

p6: it appears that 5 random parents were chosen out of the initial population sample

Ok, these are actually called 5 random samples

What do you see occur on the third graph?

p6: It appears to be the mean of the sample

So what do you think will happen if I do this five times? This means I would draw five samples of size five.

What do you think you will see happen down here? (Pointing to the 3rd graph)

p6: You would probably get a more precise estimate of the mean?

Ok, well let's see... (Take 5 random samples) and repeat a few times

What do you notice about this distribution here?

p6: it is starting to form the same distribution curve as the normal distribution.

Is this mean the same as this mean here? (Pointing to the mean of the 3rd graph and comparing it to the first graph)

p6: It looks that way. (It is

So what kind of conclusion can we draw about taking a large number of samples from the population and looking at the mean of those samples

p6: You will get roughly the same size, uh same mean, depending on the size, if you take a larger sample you will get a more, uh, get more equivalent mean.

Anything else?

Ok. So what we have noticed is that the distribution of the means of the samples is also distributed normally.

p6: uh hum.

What would you say about this distribution?

p6: that it is skewed to the right?

What do you think might be different about this distribution and the one we looked at previously p6: there are more outliers, definitely to the right, so it's not a normal distribution or a bell shaped curve

What do you think will happen when I click on animate this time?

p6: it would be more likely to get samples from the left of the population, those numbers are more probable

What do you think might happen to the distribution of the mean when we draw from this population?

p6: you might get a sort of normal distribution that is more left than the previous one

Ok, so you think that it will still be normal?

p6: probably

So I drew a sample, and it is heavily weighted towards the left side (like he said)

And the sample mean is a little to the left of the parent mean

The parent population has a mean of?

p6: 8.08

And the sample population has a mean of?

p6: 6.64

Ok, so what do you think would happen if we took 10,000 samples of size 25?

p6: it will likely get the same distribution curve as before

The distribution curve as we had with the normal?

p6: no it’s skewed to the right

Ok. Well let’s take another sample.

What do you notice?

Ok, so what do you think is the overall point of this applet? (Did you do this in your class?)

p6:No.

To show what happens in comparison to an originally normal distribution when you take a sample size and a skewed distribution and take a sample size.

And what would you say about drawing a large sample versus a small sample?

p6: Ok it definitely gives you a better answer

By answer you mean?

p6: better distribution, like a better idea of what like actually happens

Ok so looking at this graph, if I take a sample size of 25, the distribution looks like

p6: uh, just a couple of bars

Ok, but if I took 1000?

p6: then you would get a normal distribution, roughly

And if you had to guess a certain number of samples you might have to take to get this normal distribution, how many would you guess?

p6: probably 500 or so

Ok that is right, but I think you could take as few as 30 to get close to normal, but 500 is a good safe answer.

p6: yes it is

Have you covered confidence intervals in your class?

p6: yes

Is there any way you would relate confidence intervals to this?

p6: well you can say that you are I guess we can go to the 68, 95, what is it, 97 we talked about

And with that picture right there can be 68% confident the mean is in what appears to be the STD of 5.

Did you do the M&M exercise in class?

Can you relate that to this?

p6: there were a certain amount of blue m&m's we had to count. I think it said something about we were something % confident that there’s got to be 3 to 9 blue m&ms in a bag of m&m's.

3 to 9?

p6: yeah, like between 3 and 9 blues in a bag

Oh I see that is your confidence interval: (3, 9)

So in your class everyone got a bag of m&m's (or per group)

p6: yeah

So each bag of m&m's would represent

p6: a population

Well, the m&m factory I suppose is the "population"

p6: a sample.

So did you have to count the m&m's?

p6: yeah the total number

Ok and then you had to take the total number of?

p6: blues out of the total number of the sample

Ok so you were finding proportions?

p6: yes

Ok and did the proportions you found match up to what you expected?

p6: I think she might have mentioned it after we figured out the proportions

Like she told us the answer it was supposed to be

Ok and did you compare your sample to the samples of the other people in the class?

p6: yup

Ok so this is sort of like when I take 5 samples here and that is sort of the same thing and this is like the mean

p6: uh huh (yes)

And what you could do is make a histogram of the outcome (proportion) of each sample (each group’s bag of m&m's)

p6: we didn't do histograms for this exercise.

But maybe you did for the coin exercise?

p6: yup.

And the coin exercise probably looked like this (showing now a left tailed skewed distribution)

p6: yeah it did

So, overall did you feel like this gave you a good example or demonstration of the sampling distribution?

p6: yeah

How would you compare this to the way your teacher explained it in words. Drawing a sample, taking the mean

Do you think this is the same?

p6: it was a pretty much the same, but a little bit easier to see a comparison when you can like do it this quickly

Then have it like the next day in class when you just do a part of it

So you prefer watching something than just listening to words

p6: yeah, usually I zone out when people talk all the time

Ok, so let me ask you some general questions about your class

p6: ok

How do you feel your statistics class is going in general?

P6: pretty good

Any things you would say specifically that you like about the class?

p6: one thing I do like about my teacher is the way that she goes over the stuff. Even though she talks, she also writes the it all down on the board

I guess I'm more of a visual learner, so like when she gives an example, I like how she gives the example it helps me with the webassign

Which do you prefer between hands on example from your teacher like an M&M activity, and when your teacher works out an example?

p6: most definitely the hands on

Ok anything you want to see

p6: what I would really like to have is like to have an example and then do it myself so I can see if I need to change something

So that doing it myself really help reinforce what the example she was giving.

Does she do that?

p6: for the in class assignment

What would you change if you could?

p6: the only problems I have are like it's redundant. Like she goes over things multiple times.

In the same way?

p6: often different ways, but I pick up on math relatively fast, so it's a little hard to keep focused on that

So you find that this class comes pretty easy to you

p6: I do have an A in here I think.

I have a little bit of troubles sometimes, like I don't' get it right away sometimes (webassign)

But usually on the second or third try.

So when you get a problem wrong on webassign, how do you go about trying to figure out what you did wrong with those problems?

p6: that would be when I actually open up my book and look at the stuff and compare what she says and figure it out that way

So other than that is there anything else you want to add?

p6: it gets a little difficult with the construction going on, but there is nothing really else.

Sorry, unfortunately I cannot help you with that.

Appendix C – Focus Group Procedures for ST311

Subject selection:

To avoid self-selection bias subjects will be selected using a stratified random sample. This sample will be selected from the list of all ST311 students who have completed the Index of Learning Styles. The sample will be stratified based on the learning styles of these students.

Setting:

Between 5 and 10 subjects from each group will be brought together in a conference room. They will all sit around a conference table. Each seat will have bottled water, a note pad and pen. At one end of the table will be a projector screen. The moderator will have a laptop computer and a projector to display materials.

Prototype script:

Introduction:

I want to welcome all of you to our focus group. My name is _________ and I will be leading this focus group. By participating in this focus group you are helping us improve our statistics course. We are trying to determine what types of materials work best for different types of learners. To do this we would like to talk about some of the materials you have seen in class and in your homework assignments. For those of you who have never been involved in a focus group before, I as the moderator will ask you a series of questions about the materials you have been using in class. Based on your comments we will ask further questions and perhaps ask you to comment on the ideas of the other members of the group.

We expect this session to last between 20 minutes and one hour.

Before we begin I would like to remind you that your participation in this focus group is voluntary. Your grade in ST311 will not be penalized in anyway if you choose not to participate. Also, I am not involved in the grading of ST311. We are recording these sessions but those involved with grading your course will not hear those recordings. The transcripts of these recording will not indicate your names or any identifying information, but instead you will be listed as participant 1, participant 2, etc. I would like to remind you that at any time you can choose stop participating in this group.

Are there any questions about the procedures?

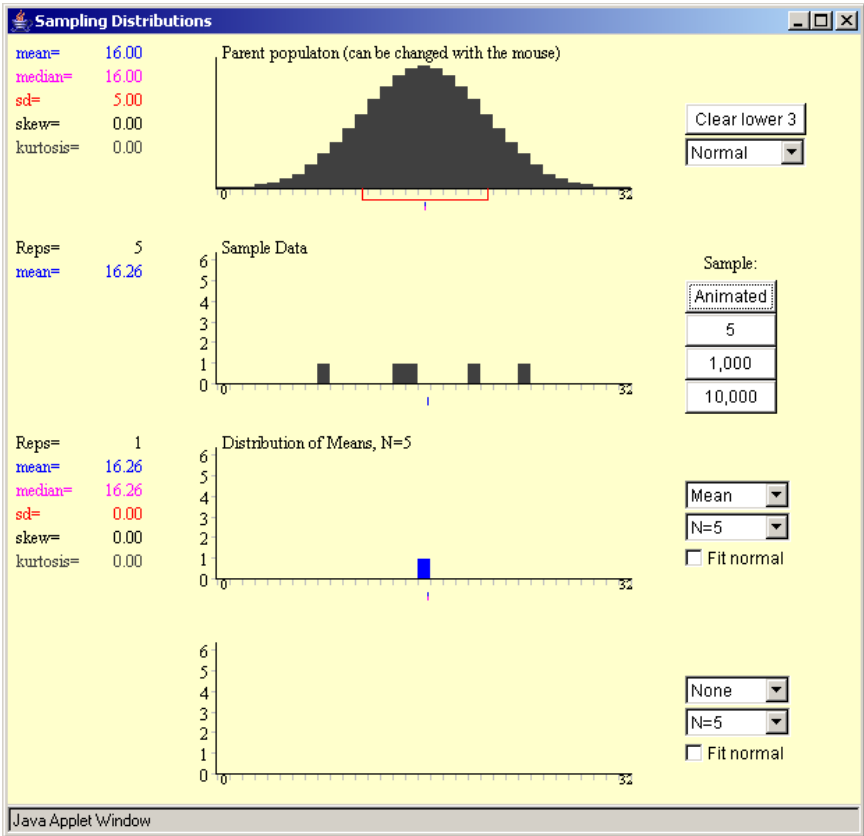
(Pause 10 to 15 seconds and try to make eye contact with each participant to insure that they all are in agreement and that they have no questions.)

Ok, let's get started

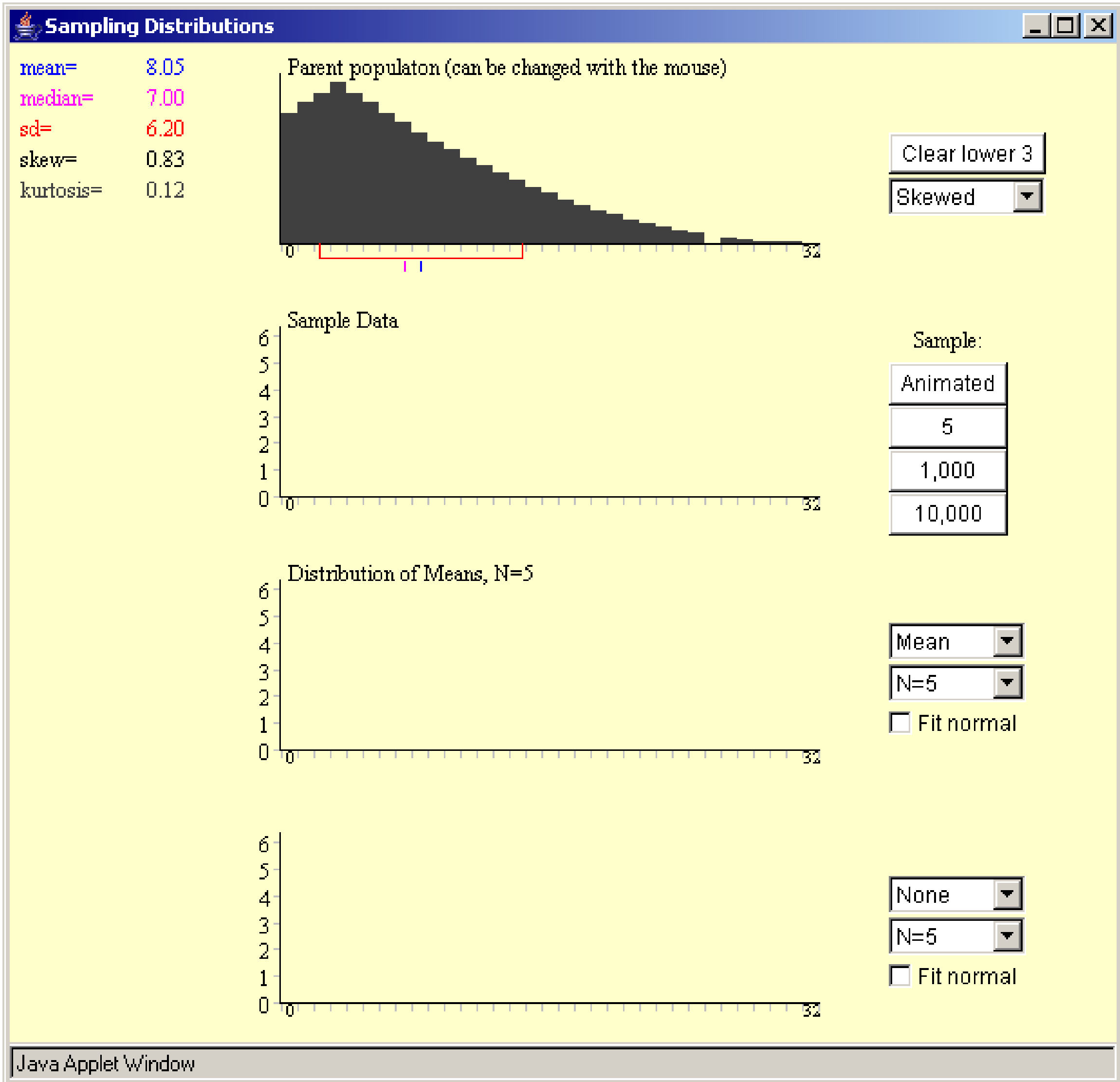
(These questions will vary based on the specific material that is of interest.)

- Think back to the applet that was demonstrated during your class on Wednesday. Does anyone recall what that applet was or what it demonstrated?
- Does anyone recall what this applet looked like?

To remind us all of the applet that was being used I will bring up a screen shot on the computer projector.



- In looking at this applet can anyone tell me what the upper graphic represents?
- What does the second graphic represent? (Indicate graph with mouse.)
- And what about the third graphic?
- Does anyone remember what happens when we run the animation multiple times?
- If we change the Parent Population to have this type of shape, how would things be different?



- Overall what do you feel was the point of this applet?

“Have you covered Confidence intervals in your class yet?"

"How do confidence intervals relate to the ideas we see in this applet?"

"Now some general questions about the course:"

“How do you feel the course is going in general?"

"What things about the course do you feel help you learn the material the most?"

"What things would you change if you could?"

"Is there anything else you would like to tell us?"

Appendix D – Invitation

Dear {Insert Name Here}:

You may remember completing a survey of Learning Styles at the beginning of the semester. We are conducting a study to try to improve our statistics course. Due to your unique Learning Style, we would like to get your opinions on the course and on the topics you are learning by having you participate in an interview. The interview will be completely informal, and nothing you do or say will affect your grade in your class. We will simply ask you some questions about the material and have a discussion the questions and what meaning you find in them. The interview should take between 20 and 45 minutes.

We value your thoughts and opinions, and this will give you a chance to shape the way future students are taught. As a further incentive we will give a $10 bookstore gift card to all students who participate in this interview, which will be handed out at the end of the session.

I would like to schedule a time that best fits your schedule where we could meet and conduct this interview. Could you let me know times during the next few weeks (after spring break) when we could schedule this interview? If you cannot attend, or simply wish to not participate, please email me and let me know so that I can select other students in your place.

Thank you for your time and interest in helping to improve Statistics Education and this course.

Kind Regards,

Jennifer Sloan
jsloan@stat.ncsu.edu
NCSU Department of Statistics

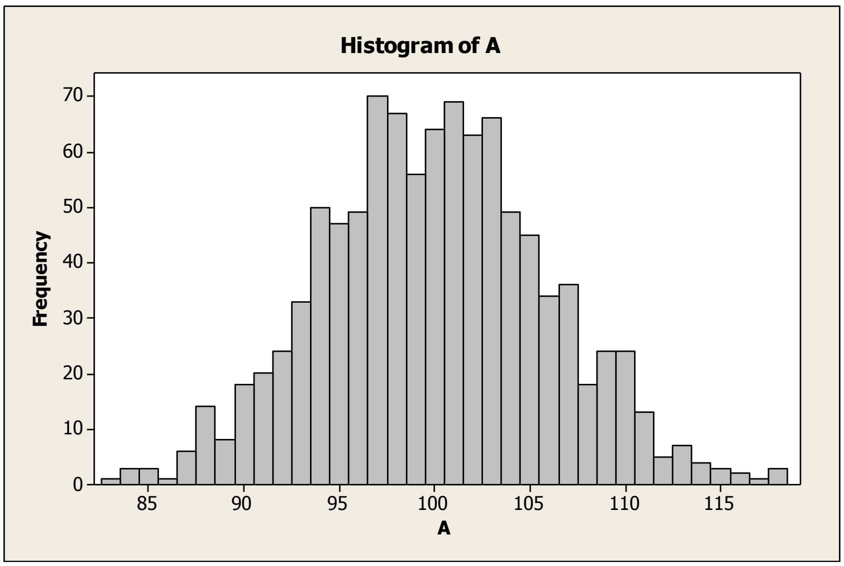
Appendix E – Pre-assessment

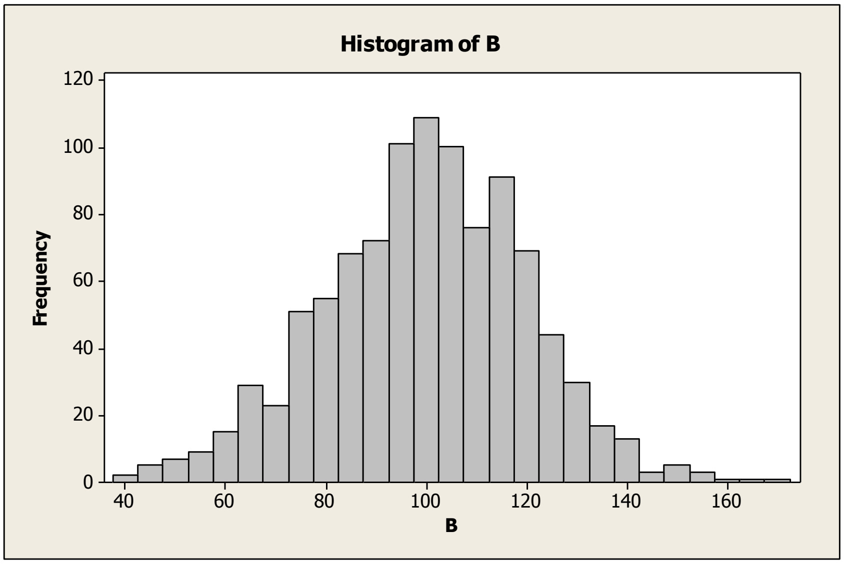
What is a variable?

What does it mean to say that observations vary: In words?

Now draw a picture to represent several observations that vary:

- Compare the two plots below. What can you say about their variability? Which is larger? Which is smaller? Or are they the same? Explain your choice.



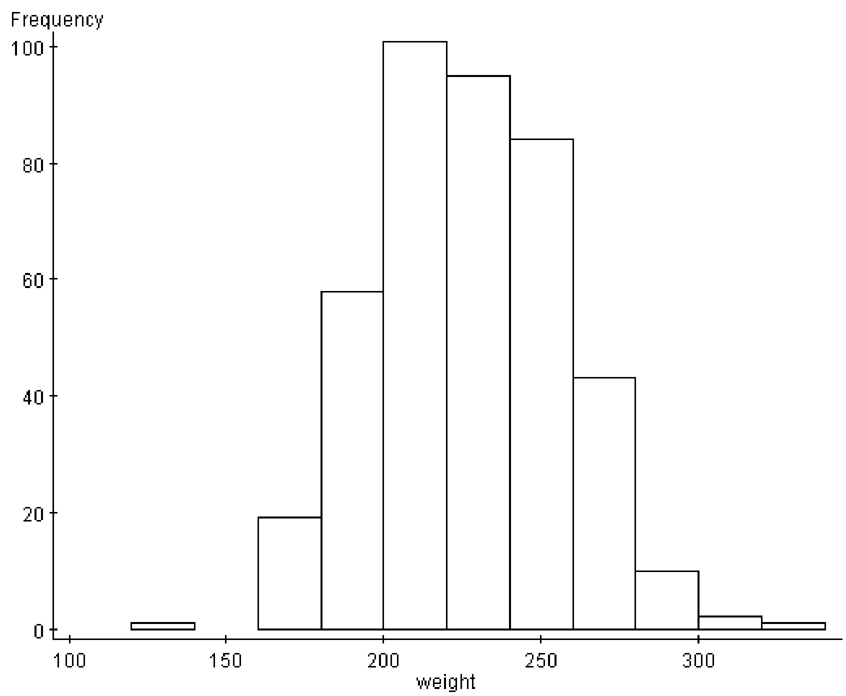


Do you feel the histogram above would represent a normal distribution?

What do I mean when I say; I have a distribution of 100 random variables?

What does a population distribution consist of?

The histogram below presents the distribution of the weights of players from the NBA.



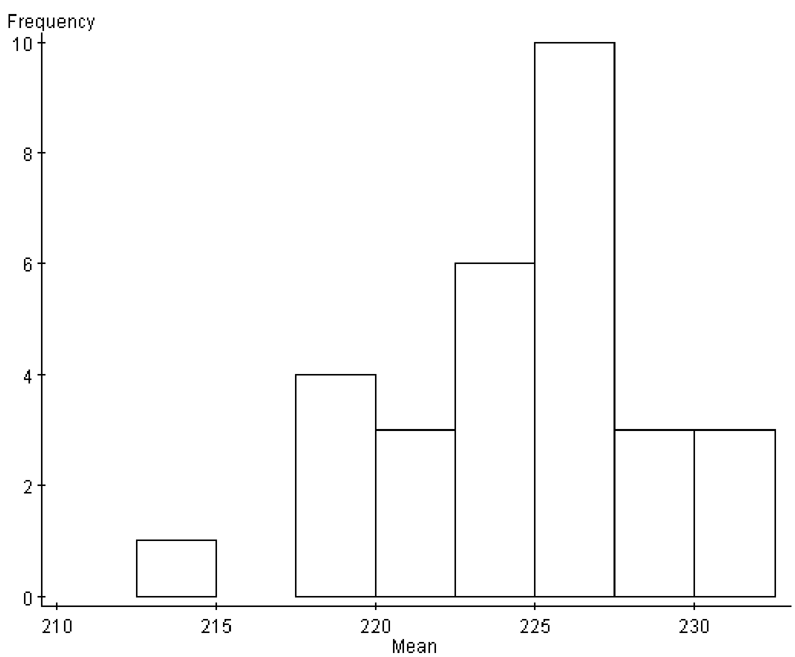
Kendric Perkins of the Boston Celtics weighs 280 pounds. Where would he be on this graph?

Based on this histogram what can you tell me about the weight of players from the NBA?

Approximately how many players in the NBA weigh less than 150 pounds?

Approximately how many NBA players weigh 300 pounds or more?

The graph below presents the *average* weights of each team (about 16 players each) in the NBA.



Kendric Perkins of the Boston Celtics weighs 280 pounds. Could you locate him on this graph? Explain why or why not.

If the parent population has a standard deviation of 12, what would the standard deviation of the second graph?

What does it mean to draw a random sample from a population?

At a large private college in upstate New York a professor asked the 50 students in his class about the number of hours of they are taking. In his class the average was 17.6 hours. In checking with the college registrar the professor found that for all students at that school the average number of hours taken was 14.3 hours.

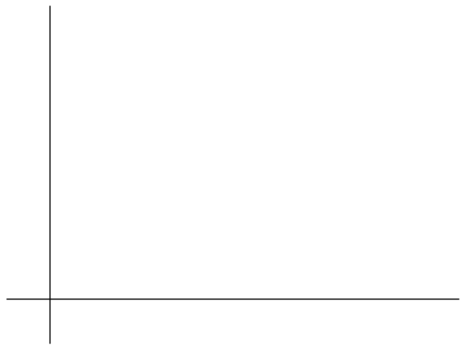
In the example above is 17.6 a sample statistic or a population parameter?

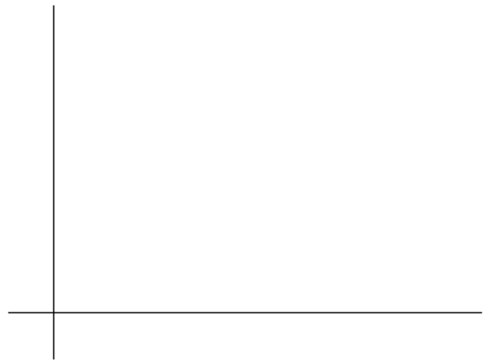
Would 14.3 be μ or $\bar{x}$. Explain your answer.

What is a NORMAL Distribution?

Please draw a NORMAL distribution on the Axes below, and label the axes:

Now draw the NORMAL distribution, but this time with an increase in variance as well as one with a smaller variance.





If the mean increased by 5, show how this would change from your original graph: What can you tell me about the area underneath the curve? What does it represent?

Appendix F – Responses

Likelihood question:

Since students were asked to give a reason for their answer, these reasons were analyzed and the number of students with the correct answer and correct reasoning are displayed below

Smaller samples are less representative of the population, more affected by outliers and thus more likely to have a mean greater than the population mean.	**Intuitor**	**Sensor**
1: With a larger sample, variability increases so the spread is greater	2 (1.82)	7 (1.49)
2: The mean increases in a larger sample	0 (0)	3 (0.64)
3: We need a larger sample to represent the entire population	3 (2.73)	21 (4.48)
**4: A small sample is not representative of the population	10 (9.09)	30 (6.40)
5: Larger samples make the outcome more likely	3 (2.73)	24 (5.12)
**6: Using the CLT, as n gets larger, the sampling distribution approaches a normal distribution with mean: μ and std: σ	1 (0.91)	8 (1.71)
**7: Larger values of n put you closer to the actual mean, so the smaller n is more likely to yield this result	12 (10.91)	32 (6.82)
**8: A large sample has less variability and a small sample has more variability, and larger margin of error	7 (6.36)	25 (5.33)
9: Need more information in order to solve this	22 (20.00)	97 (20.68)
**10: Small samples are skewed and thus not representative of μ. Outliers will affect the sample mean more in a smaller sample	20 (18.18)	67 (14.29)
NO RESPONSE:	0	

CENTER QUESTION

The center of the sampling distribution will be at the mean of the population	**Intuitor**	**Sensor**	**Total**
**1: Center at the population mean / μ	21 (26.25)	96 (30.09)	117 (29.32)
**2: Larger sample is a representative of the entire population so we know the mean	7 (8.75)	30 (9.40)	37 (9.27)
3: The center is at μ because it is the median	0 (0.00)	2 (0.63)	2 (0.50)
**4: The sampling distribution is NORMAL and therefore it is centered at μ since n>30. Central Limit Theorem	14 (17.50)	39 (12.23)	53 (13.28)
5: This is a small sample so the center could differ.	3 (3.75)	9 (2.82)	12 (3.01)
6: Outliers will make it skewed, so we don't know	5 (4.55)	17 (5.33)	22 (5.51)
7: We need to know the median to know the center	2 (2.50)	11 (3.45)	13 (3.26)
8: We do not have enough information. The sample average is unknown. We need the population distribution. This is random and we don't know if it is normal.	8 (10.0)	35 (10.97)	43 (10.78)
9: The sample mean is not always equal to μ.	7 (8.75)	27 (8.46)	34 (8.52)
10: This is random and could vary with each sample size. We aren't sure how close to the population. It might change	3 (3.75)	26 (8.15)	29 (7.27)
No Response	10 (12.50)	27 (8.46)	37 (9.27)

SHAPE QUESTION

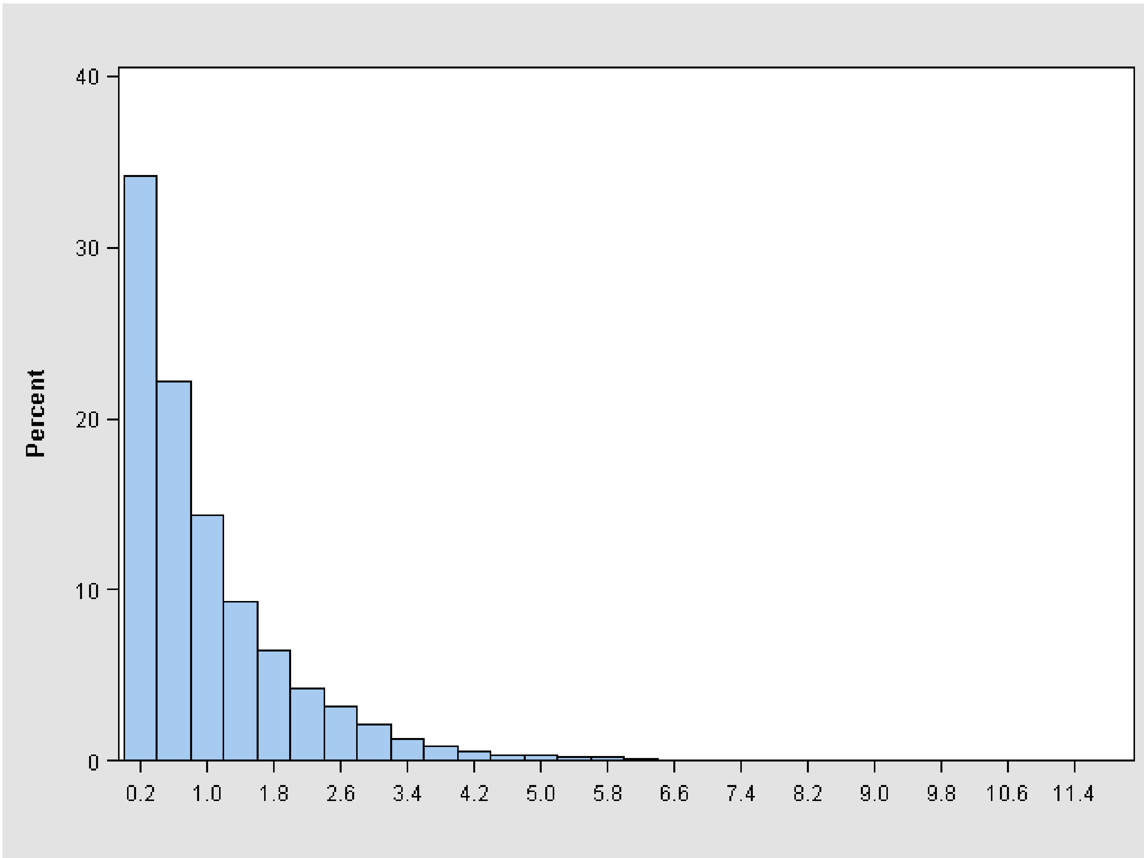
We know the shape of a sampling distribution is going to be normal with mean μ and std, σ/√(n)	**Intuitor**	**Sensor**
**1: Normal, bell-shaped, because n>30 and using the CLT. Xbar is normal(μ,(σ/√(n)));	8 (10.00)	44 (13.8)
2: Should have the shape of the population since the sample size is large	6 (7.50)	27 (8.46)
** 3: Sampling distribution of the mean is always normal around mean. It is symmetric.	5 (6.25)	17 (5.33)
4: We can predict the shape because we know μ and σ	3 (3.75)	11 (3.45)
**5: It is normal because it is a random sample and a large sample size	9 (11.25)	21 (6.58)
6: We don't have enough information, we need the data points, we need to perform a test, we need to know if it is normal. We need the mean. Random.	11 (13.75)	57 (17.87)
7: We need the median to know the shape	0 (0.00)	15 (4.70)
8: Right skewed. We need the shape of the parent population to know. It is not normal	20 (25.00)	75 (23.51)
9: It may have outliers which affect the shape. It could be skewed. Shape is flatter for the mean	5 (6.25)	21 (6.58)
10: We don't know the standard deviation.	1 (1.25)	1 (0.31)
NO RESPONSE	12 (15.00)	30 (9.40)

Spread

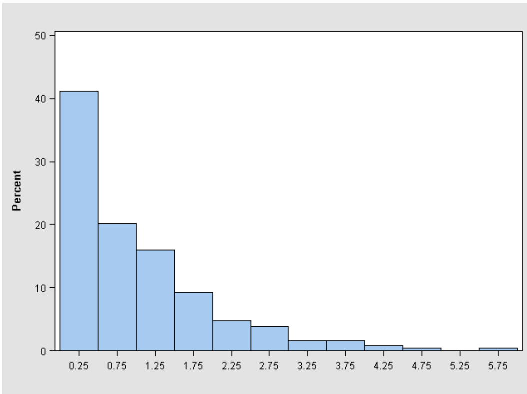
The sampling distribution will have a smaller STD than the population because it is a distribution of sample means that represent the population mean	**Intuitor**	**Sensor**
1: Since the sample size is smaller than the population size, $n<N$ therefore $s<\sigma$. Inversely, a larger sample size gives a larger STD	8 (7.27)	49 (10.45)
2: The mean of the population is greater than the sample mean therefore the STD is larger in the population.	2 (1.82)	4 (0.85)
**3: The sampling distribution has a smaller spread therefore a smaller STD than the population. The sampling distribution is symmetric and centered around μ.	7 (6.36)	28 (5.97)
4: Small sample has a larger STD therefore since $n < N$, $s > \sigma$. INVERSELY: a larger n has a smaller std so σ is small.	15 (13.64)	65 (13.86)
**5: Formula is $\sigma/\sqrt{(n)}$, also CLT	6 (5.45)	44 (9.38)
6: Outliers can affect the population more and therefore the population has a larger STD therefore $s<\sigma$.	1 (0.91)	10 (2.13)
7: Sampling distribution (is of the mean), is repeated samples (not individual observations) and therefore has a smaller STD.	2 (1.82)	6 (1.28)
8: Outliers have less of an effect upon a larger N such as the population therefore $s > \sigma$.	2 (1.82)	7 (1.49)
9: Sampling distribution has a wider range of numbers	3 (2.73)	12 (2.56)
10: STD of the samples should match the population. It won't change just because the number changes. Sample size has no effect on the STD.	18 (16.36)	70 (14.93)
No Response	46 (41.82)	174 (37.10)

Appendix G – Task Based Interview Format

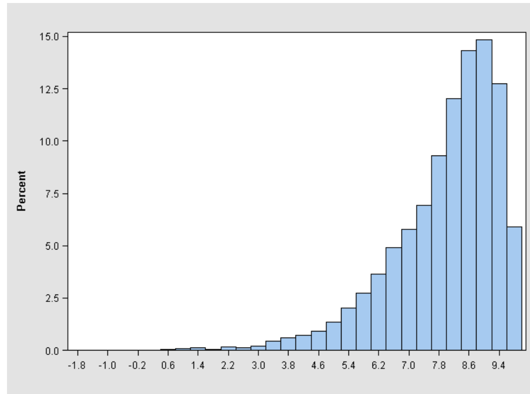
Question 1:



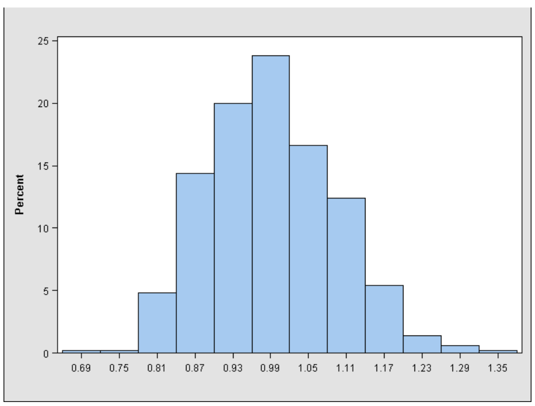
Which of the histograms to the right represents a distribution of 500 sample means from random samples of size 100 from this population?


Percent
0 10 20 30 40 50
0.25 0.75 1.25 1.75 2.25 2.75 3.25 3.75 4.25 4.75 5.25 5.75

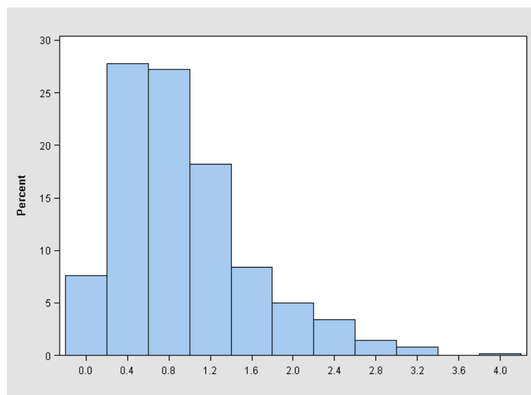
A


Percent
0.0 2.5 5.0 7.5 10.0 12.5 15.0
-1.8 -1.0 -0.2 0.6 1.4 2.2 3.0 3.8 4.6 5.4 6.2 7.0 7.8 8.6 9.4

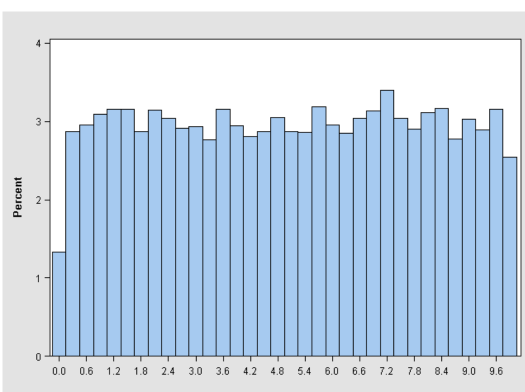
B


Percent
0 5 10 15 20 25
0.69 0.75 0.81 0.87 0.93 0.99 1.05 1.11 1.17 1.23 1.29 1.35

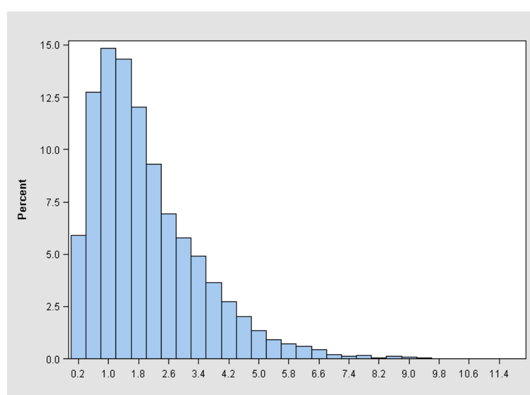
C


Percent
0 5 10 15 20 25 30
0.0 0.4 0.8 1.2 1.6 2.0 2.4 2.8 3.2 3.6 4.0

D

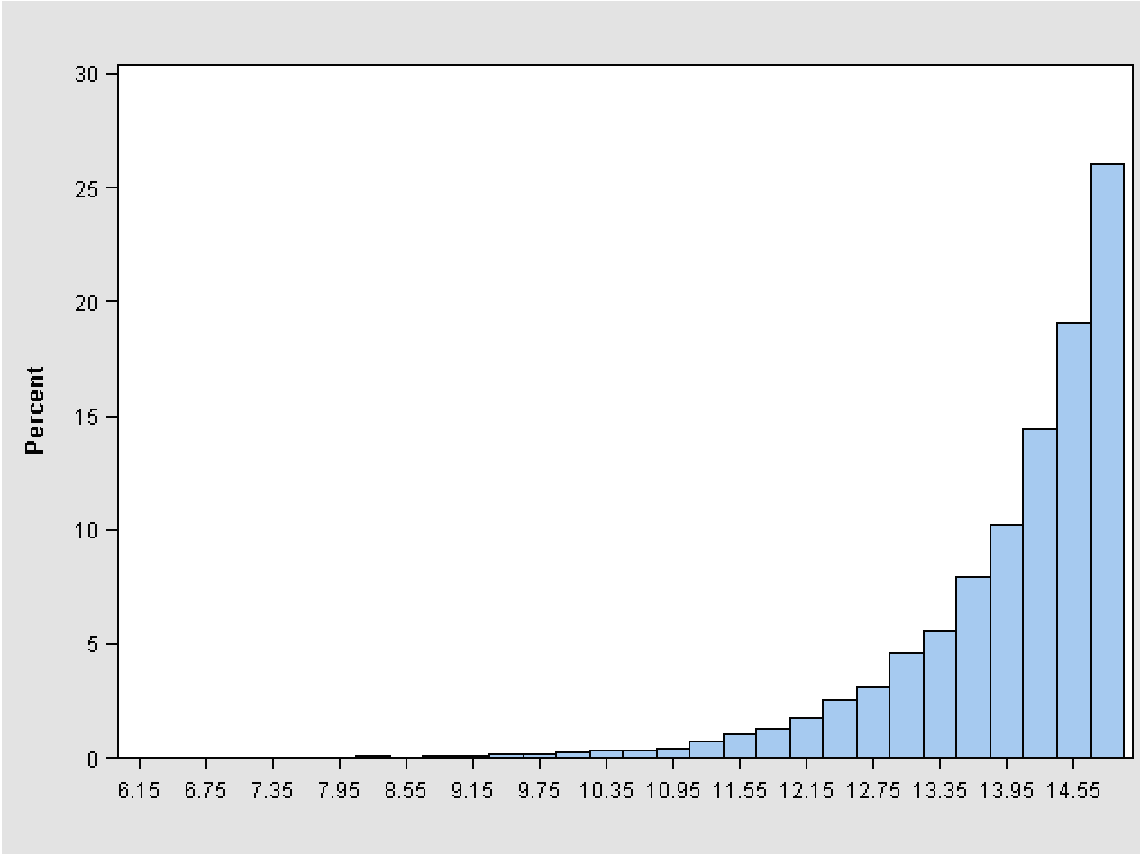

Percent
0 1 2 3 4
0.0 0.6 1.2 1.8 2.4 3.0 3.6 4.2 4.8 5.4 6.0 6.6 7.2 7.8 8.4 9.0 9.6

E

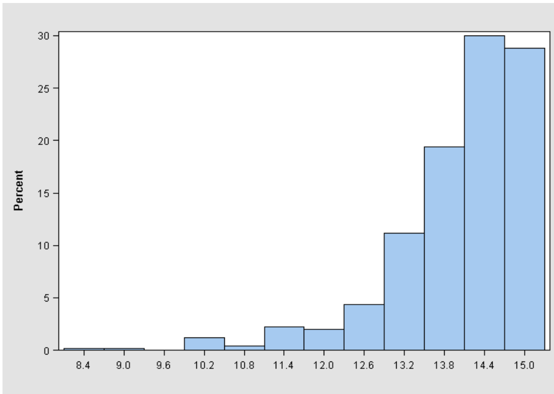

Percent
0.0 2.5 5.0 7.5 10.0 12.5 15.0
0.2 1.0 1.8 2.6 3.4 4.2 5.0 5.8 6.6 7.4 8.2 9.0 9.8 10.6 11.4

F

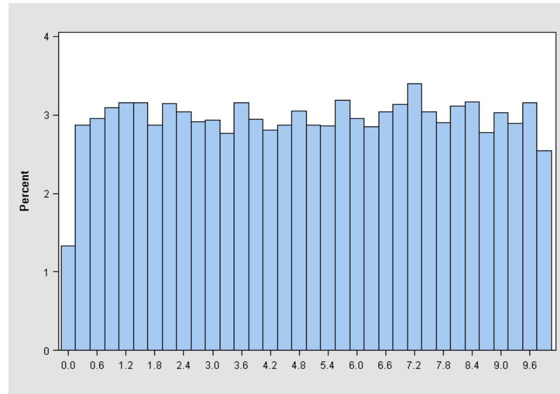
Question 2:



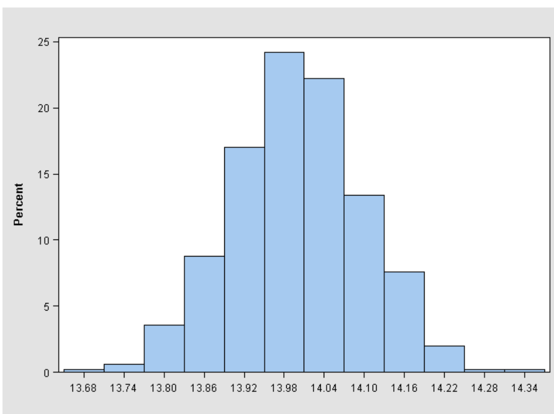
Which of the histograms to the right represents a distribution of 500 sample means from random samples of size 100 from this population?


Percent
0
5
10
15
20
25
30
8.4 9.0 9.6 10.2 10.8 11.4 12.0 12.6 13.2 13.8 14.4 15.0

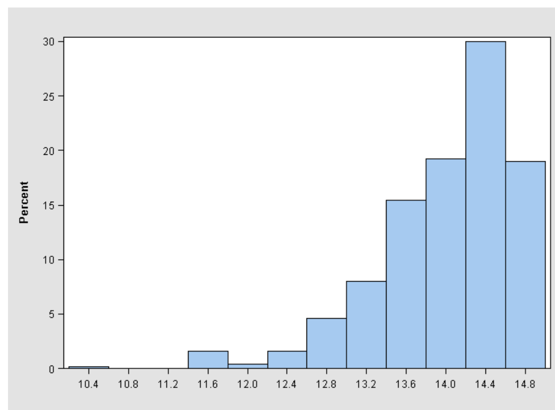
A


Percent
0
1
2
3
4
0.0 0.6 1.2 1.8 2.4 3.0 3.6 4.2 4.8 5.4 6.0 6.6 7.2 7.8 8.4 9.0 9.6

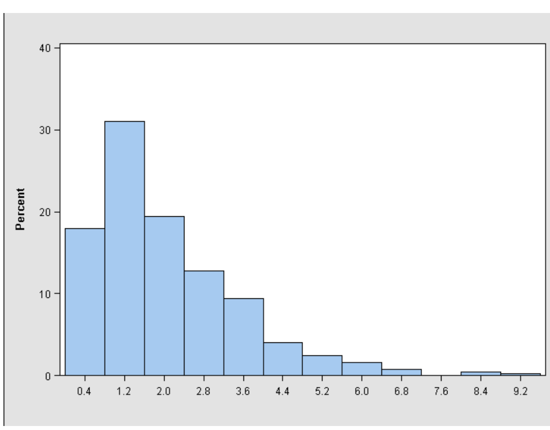
B


Percent
0
5
10
15
20
25
13.68 13.74 13.80 13.86 13.92 13.98 14.04 14.10 14.16 14.22 14.28 14.34

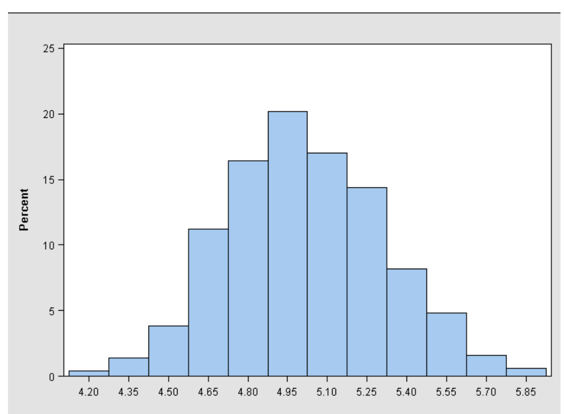
C


Percent
0
5
10
15
20
25
30
10.4 10.8 11.2 11.6 12.0 12.4 12.8 13.2 13.6 14.0 14.4 14.8

D

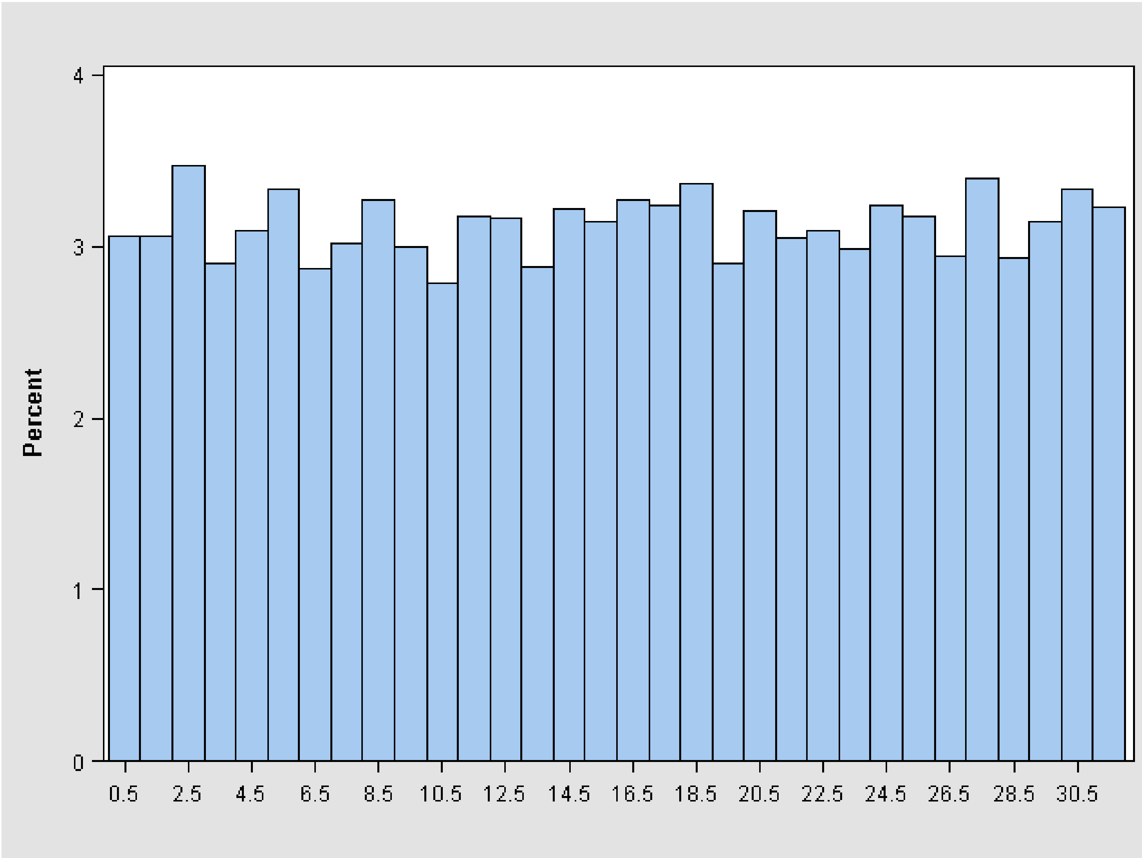

Percent
0
10
20
30
40
0.4 1.2 2.0 2.8 3.6 4.4 5.2 6.0 6.8 7.6 8.4 9.2

E

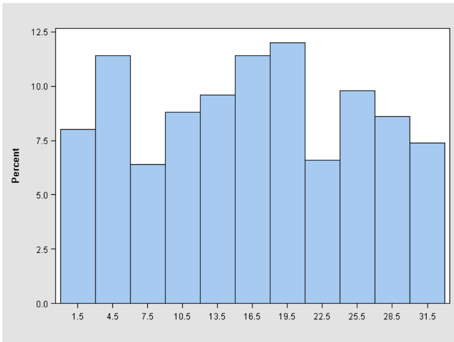

Percent
0
5
10
15
20
25
4.20 4.35 4.50 4.65 4.80 4.95 5.10 5.25 5.40 5.55 5.70 5.85

F

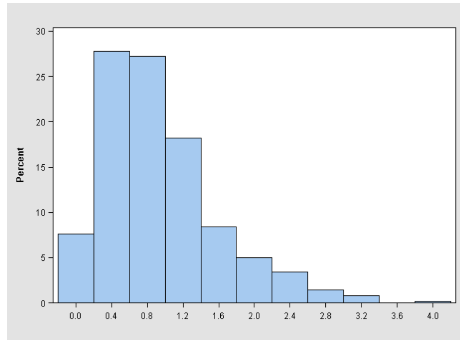
Question 3:



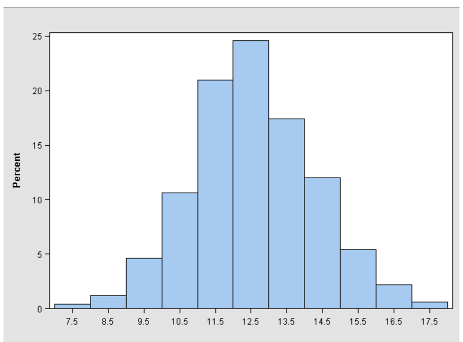
Which of the histograms to the right represents a distribution of 1000 sample means from random samples of size 25 from this population? You may use the applet previously demonstrated to assist with this problem.


Percent
12.5
10.0
7.5
5.0
2.5
0.0
1.5 4.5 7.5 10.5 13.5 16.5 19.5 22.5 25.5 28.5 31.5

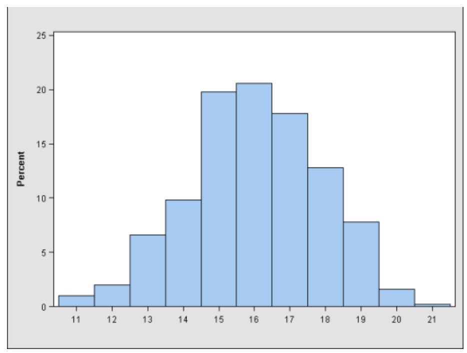
A


Percent
30
25
20
15
10
5
0
0.0 0.4 0.8 1.2 1.6 2.0 2.4 2.8 3.2 3.6 4.0

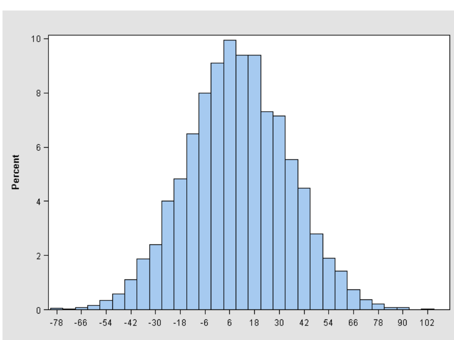
B


Percent
25
20
15
10
5
0
7.5 8.5 9.5 10.5 11.5 12.5 13.5 14.5 15.5 16.5 17.5

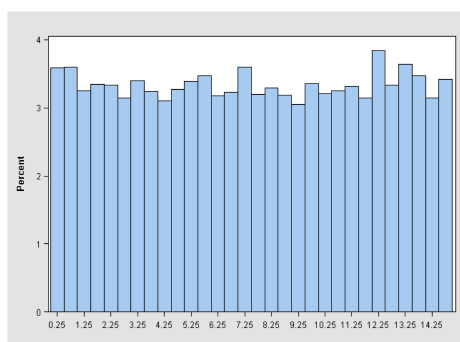
C


Percent
25
20
15
10
5
0
11 12 13 14 15 16 17 18 19 20 21

D

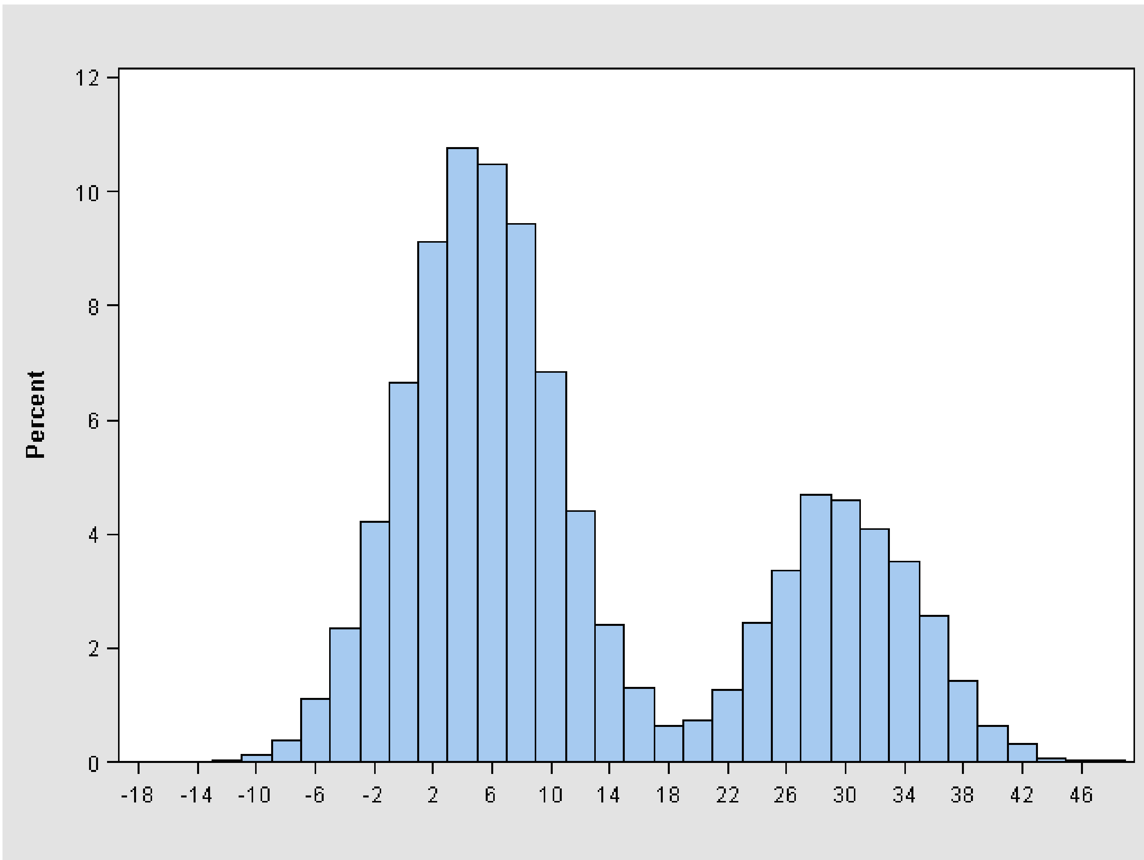

Percent
10
8
6
4
2
0
-78 -66 -54 -42 -30 -18 -6 6 18 30 42 54 66 78 90 102

E

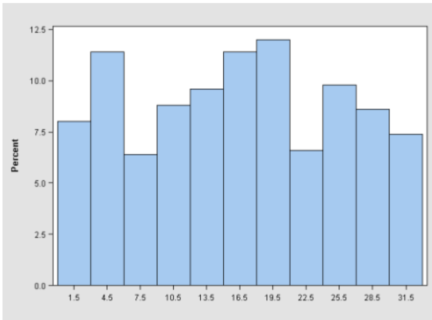

Percent
4
3
2
1
0
0.25 1.25 2.25 3.25 4.25 5.25 6.25 7.25 8.25 9.25 10.25 11.25 12.25 13.25 14.25

F

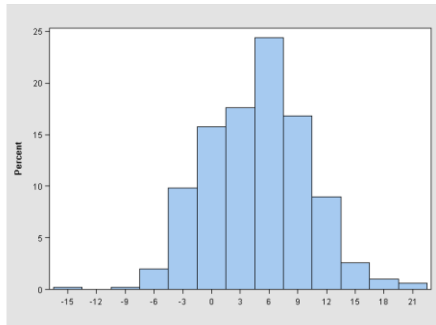
Question 4:



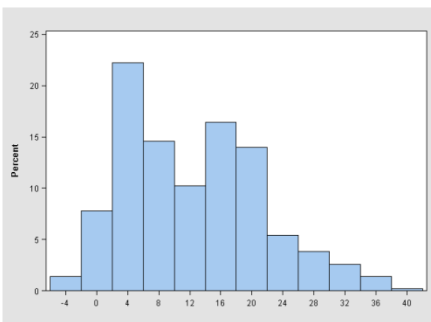
Which of the histograms to the right represents a distribution of 500 sample means from random samples of size 50 from this population? You may again use the applet to assist you with this problem.


Percent
0.0
2.5
5.0
7.5
10.0
12.5
1.5 4.5 7.5 10.5 13.5 16.5 19.5 22.5 25.5 28.5 31.5

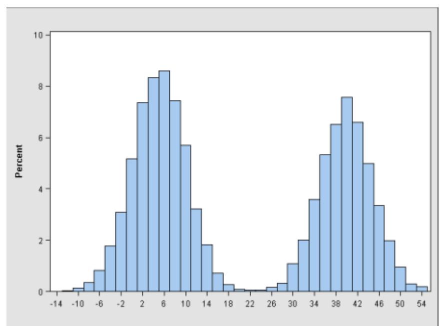
A


Percent
0
5
10
15
20
25
-15 -12 -9 -6 -3 0 3 6 9 12 15 18 21

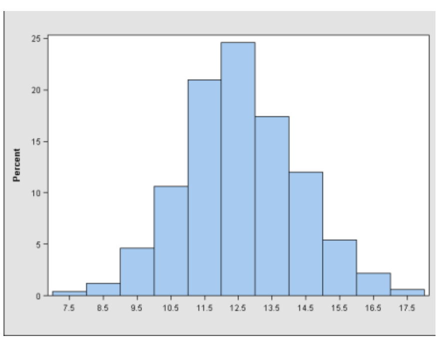
B


Percent
0
5
10
15
20
25
-4 0 4 8 12 16 20 24 28 32 36 40

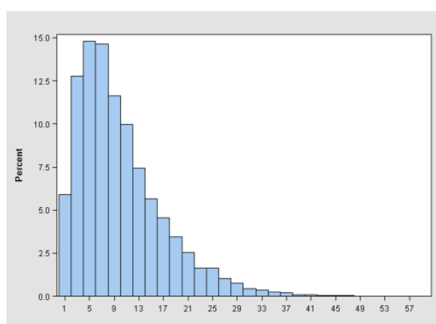
C


Percent
0
2
4
6
8
10
-14 -10 -6 -2 2 6 10 14 18 22 26 30 34 38 42 46 50 54

D

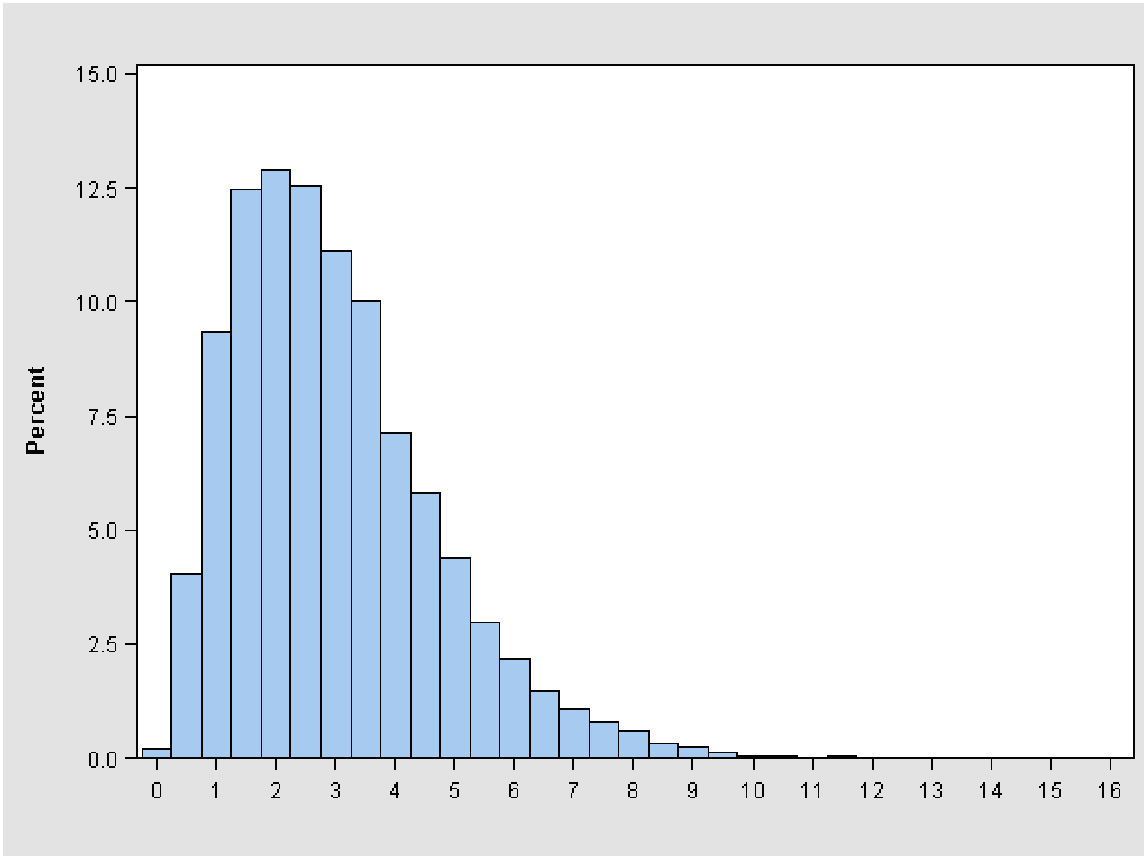

Percent
0
5
10
15
20
25
7.5 8.5 9.5 10.5 11.5 12.5 13.5 14.5 15.5 16.5 17.5

E

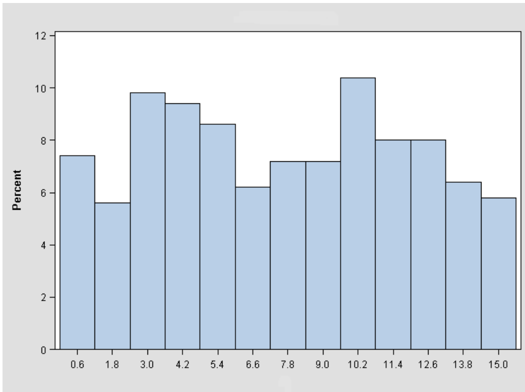

Percent
0.0
2.5
5.0
7.5
10.0
12.5
15.0
1 5 9 13 17 21 25 29 33 37 41 45 49 53 57

F

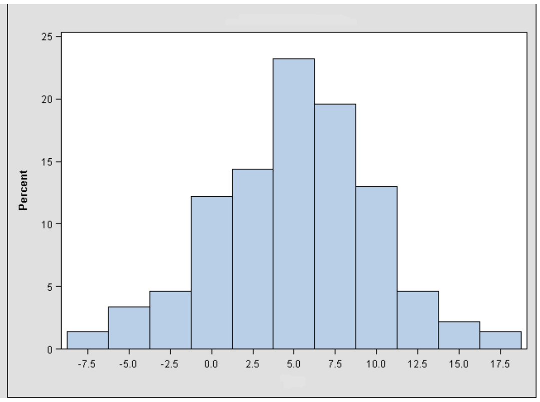
Question 5:



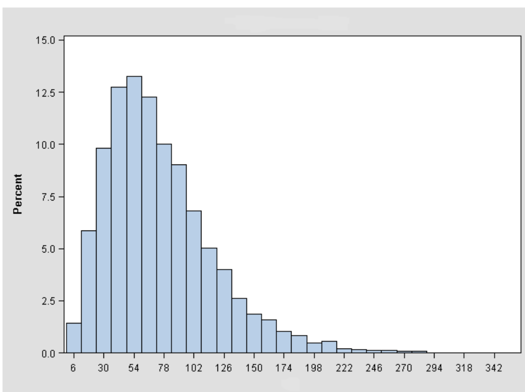
Which of the histograms to the right represents a distribution of 500 sample means from random samples of size 100 from this population?


Percent
0 2 4 6 8 10 12
0.6 1.8 3.0 4.2 5.4 6.6 7.8 9.0 10.2 11.4 12.6 13.8 15.0

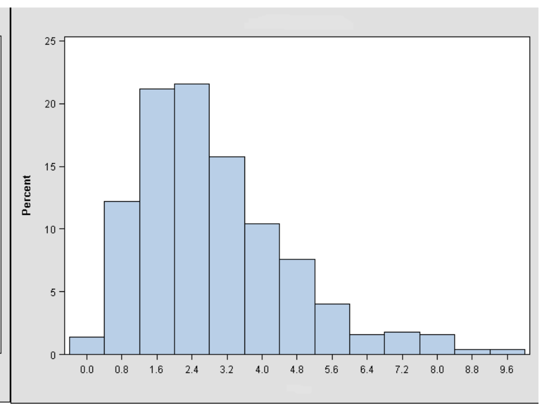
A


Percent
0 5 10 15 20 25
-7.5 -5.0 -2.5 0.0 2.5 5.0 7.5 10.0 12.5 15.0 17.5

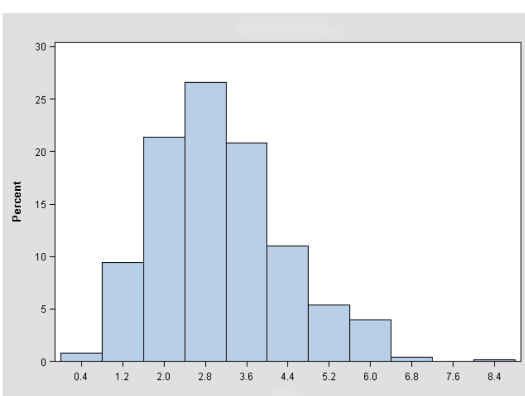
B


Percent
0.0 2.5 5.0 7.5 10.0 12.5 15.0
6 30 54 78 102 126 150 174 198 222 246 270 294 318 342

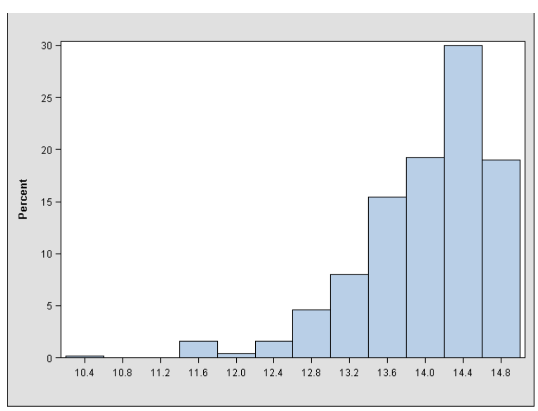
C


Percent
0 5 10 15 20 25
0.0 0.8 1.6 2.4 3.2 4.0 4.8 5.6 6.4 7.2 8.0 8.8 9.6

D


Percent
0 5 10 15 20 25 30
0.4 1.2 2.0 2.8 3.6 4.4 5.2 6.0 6.8 7.6 8.4

E


Percent
0 5 10 15 20 25 30
10.4 10.8 11.2 11.6 12.0 12.4 12.8 13.2 13.6 14.0 14.4 14.8

F

Appendix H – Task Based Interview Fall 2007

Interview 1: Intuitor =9

Question 1:

For sample size of n=2, she chose B, which is the correct answer. Her reasoning is that the shape was similar.

For sample size of n=100, she chose C because, it was a random selection.

Applet Demo:

Participant understood that the means would fall in different locations. With an increased sample size, she believed that the means would fall in a general area.

Question 2:

For sample size of n=2, she chose F. It was her choice because she recalled from the demonstration that a smaller sample had more variability, and F looked like the parent. F was a sample of size 1. She was stuck between A and F. A would have been the correct choice.

For sample size n=100, she chose B. She chose B because it was more "focused" and this is a word she used previously to describe a smaller variance.

Question 3:

Participant used the applet a few times, but not enough to show the distribution. She then decided to do it 1000 times to show what the distribution would look like. For sample of size n=2, she chose F. This was her choice, she said, because it has less variability than B (the sample of size one) because it was still wide and she knew that is what she wanted to choose.

I had to inform her that one of them had a sample of size one so that she could determine from more information.

For n=100 she chose D. This, she said, was because it was less variable than the other graphs. This was the correct choice as well.

Question 4:

Participant used the applet to draw a uniform distribution, she took a sample of size 2 and repeated 1000 times.

Using this information she chose A. She chose A because it showed high variability, but still had a slightly normal shape to it.

For a sample of size 100, she chose E. This was because it was the most focused and thus had the least amount of variability. She does remark that it could be any of the normal distributions, but that E is the least variable and thus that lead to her choice.

Question 5:

Participant used the applet to draw the bimodal distribution. She then took a sample of size 2 and repeated 1000 times. From this she chose D since it was starting to show normality, but not quite.

For sample of size 100, she chose A because it had the least amount of variability.

Question 6:

Participant chose C for the sample of size 2. This was because it was a little less variable than A (sample size one) but still had a large amount of variability

For sample size 100, she chose F because it was less variable and had the right shape.

COMMENTS:

Dislikes in class activities b/c she feels under pressure and does not feel like she's actually learning something.

Interview 2: Intuitor=9

Question 1:

For sample size 4 he Chose A. Did so because he thought it would look like the parent. Chose F for n=100 because when you take a larger sample size, it will more closely approximate these numbers that are estimates of the population.

Applet Demo:

He understood that the variance would decrease as the sample size increased. He also knew the values would tend to fall around the mean of the parent. He did not use the applet to answer the other questions because he felt he didn't need it. He was almost right for every question.

Question 2:

For sample size 4, he chose F. He approached the problem correctly and realized that he would need a sample standard deviation of 0.5 and he felt that F was the best choice for that since it too had a standard deviation of about 0.5.

For sample size 100, he chose D because he wanted something with a smaller spread, since he thought the sample std would be 1/10 which was correct thinking. He chose D over A because he felt that A wasn't quite what he wanted. He also ignored A since he used it in his answer for question 1.

Question 3:

For sample size 4 he chose F. He was stuck between F and B. He didn't want to pick F because he had already chosen F for the previous problem. I encouraged him that the solutions are in no particular order and he shouldn't let that persuade him and his decision. He didn't choose B because it was too varied. So he concluded the solution must be F.

For sample size 100 he chose E. The reason he chose E was because he felt like it has a standard deviation of 1/10. He also thought that E looked and felt more normal than D.

Question 4:

For sample size 4 he chose A. This was the correct answer. He ruled out D because it was more random. He ruled out F because he wanted a mean of 3/20. He found the center and then calculated the STD 3 places out to see if most of the data was captured.

For sample size 100 he chose E. He used the same line of thinking and was looking for the choice with a std of 0.03.

Question 5:

For sample size 4 he chose D. Again this was the correct solution. He did not use the applet to aid in any of his thinking for questions 3-5. He felt the normal pattern would begin to emerge, but not absolutely normal.

For sample size 100 he chose A. This was because it had a standard deviation of 0.2, so he wanted all the data to be captured between the mean and 0.6 away from the mean which is 3*STD.

Question 6:

For sample size 4 he chose C. This was because he wanted most of the distribution to be within .75 of the mean and exhibit a normal shape.

For sample size 100 he chose F. The correct answer was actually E, but the difference b/w E and F was that E was for a sample of size 100 and F was for a sample of size 30.

Interview 3: Intuitor=7

Question 1:

She chose C for sample size 4, and for sample size 100 she chose E. She was just guessing these solutions because she had no clue.

Applet Demo:

Student watched and expected to see the means (of repeated sampling) fall around the same area. The participant noted that the shape of many repeated means was more symmetrical than the parent population. When asked if she thought a larger sample size would have this same shape, she felt it would change. She also felt that it would have a higher variability.

I then demonstrated the difference between a sample of size 25 versus a sample of size 5.

After the applet demonstration, she was able to say that with a smaller sample size you get a larger variance, and vice versa.

She wanted to go back to number one and change her answers so I allowed it.

This time, after seeing the applet demonstration, she chose B for the sample of size 4 which was the correct answer, and she chose F for the sample of size 100, which again was the correct answer.

She did, however, state that the mean of the sample would have a different mean than that of the parent. I am not sure that she “gets it” at this point, and displayed a weakness in sampling, sample size, and variability.

Question 2:

For a sample of size 4 she was stuck between B and C. She chose B over C because she said “it was not as distributed”. By this she meant varied. For sample of size 100 she chose F.

This was because “it looked more symmetrical and its mean was about 19.” She did not consider any of the other symmetric distributions which were also centered at 19. This is a good example of Intuitors being faster workers making careless mistakes. Having said this, she did select the correct distribution, but this one had a sample of size 15, not 100.

Question 3:

For sample of size 4 she chose B. Since the parent was a normal distribution, it was probably more difficult to select a solution without using the other information given in the question, namely the Standard Deviation.

She chose B because “it resembles that graph”. She also considered F, but did not choose it.

F was the correct solution

For sample of size 100 she chose D because it was skinnier than the sample size of 25. She also considered E, but stuck with d because “it looks a lot more narrow than sample of size 25.

Question 4:

For sample of size 4, she chose A. This was because “by looking at the graph (applet) it was more normalish”.

For sample of size 100 she chose between E and C. I asked her what would make her choose C over E. She had no answer. Then I asked her what would make her choose E over C. She said “well it has a smaller standard deviation.”

Question 5:

For sample of size 4 she chose D. This was “because I expected it to somewhat resemble this graph (pointing to the bimodal parent) but I wouldn’t expect it to resemble it like that, (pointing to selection B).

For sample size 100, she chose A “because it was normal and the smallest.” (meaning least varied)

While she used the applet for 3 and 4, she did not use it for question 5.

Probably because she noticed the trend and was able to determine the solution without the use of the computer aide.

Question 6:

For sample of size 4 she chose C. This was “because it looks sort of like the parent but it’s more normalish.”

For sample of size 100 she chose E. This was “because the mean was 2.5 and it is probably what the graph would look like.”

General questions:

Dislikes that the class is fill in the blank notes. The teacher will hand out notes with blanks and it makes the class pretty boring because you are just looking to find the right word.

The examples are also fill in the blank.

She likes the group activities for the candy.

Interview 4: Intuitor=11

Question 1:

For sample size 4 he originally pointed to B as the solution for sample of size 4 and to A for sample of size 100. He stated that "the more samples you take you are going to get closer to this (pointing at the parent population). So this one (B) is not quite there. Actually I'm going to change my answer. I mean the more samples you take of a higher sample size, the closer you are going to get to this in the end, (again pointing at the Parent population). I think. No it's going to be centered around the mean. If you take all the means from all the possible samples and averaged them, you would get the actual mean of that population. So the more you take, the closer to the mean you are going to get. Ok I'm just going to choose F for the first one and D for the sample of size 100. I can sort of justify it, never mind. I'm just guessing at this point. I'm a bit confused. "

Applet Demo:

He notes that the larger sample size has lower variability and recognizes that the sampling distribution of a sample size greater than one will begin to exhibit a normal shape.

Question 2:

For sample size 4 he chose B. He chose it because "it was normal, but had the most amount of variability. It would not be C because it does not have a normal distribution and F looks like it would be somewhere between 4 and 100 because the frequency is higher and the variability is less." He chose A for sample size 100 because it had "the right mean and smallest variance."

Question 3:

Initially (without the applet) he chose B for sample of size four and D for sample of size 100. I then asked him if he knew of a way to derive the standard deviation of a sample of size 4, if he was given the standard deviation of the parent population.

He then recalled the formula, and calculated the standard deviation of the sample.

“It’s going to be ½ of it, which is why the variability decreases, the standard deviation decreases with a larger sample size.”

I then asked him if his selection had a standard deviation of ½ .

He then tries to use the applet to check the standard deviation. He draws a sample of size 5 and finds that the standard deviation is approximately ½ of the population standard deviation. He notes that his sample of size 25 has a standard deviation of about 1/5 of the parent population.

“So, as far as judging the standard deviations of these (referring to the 6 selections) This one looks like it has a standard deviation of ½ (pointing to selection F).”

He chose B for sample size 100 because he said “the standard deviation would be about 0.3 and about all the data fall here so this one would be the correct choice.”

Question 4:

He used the applet to draw samples of size 2 and 25, and noted that the shape of the distribution of means was still normal, despite that he had a uniform parent population.

“In that case, based on shape, I would probably say that A was sample size 4.” He then calculated the standard deviation of the sample for size four, and confirmed that his selection had the right std. He then noted that for sample size 100, he would have a standard deviation of 0.03, and stated that “I’m just generally using the outlier as the third standard deviation away from the mean, about 97% of the observations will fall within that value.”

Question 5:

“I’ve never seen this graph before (remarking about the bimodal distribution) but I assume that it is going to be normal since most of the answers are normal. So the next thing I’m going to do is find out the standard deviation. So for size 4 the standard deviation is going to be 1, and the mean is 2.8. So this is close (pointing to D) to it, but it’s not really normal, but I’ll go with this one because it’s probably the best answer. And the 10 one (referring to a sample of size 100) is going to have a standard deviation of 1/5, so I’m probably going to go with A, because it looks like it’s closer to a 1/5 standard deviation.

He did not opt to use the applet with this problem.

Question 6:

“since the standard deviation of the parent is 0.5, the std for sample size 4 will be 0.25, so everything is pretty much going to fall w/in 0.75 of the mean, so, I’ll go with C, because it looks to be the closest. And for 100, it is going to have a std of 0.05, which E looks like it is closest to.”

Interview 5: Intuitor=7

Question 1:

For sample size 4 she chose B. This is “because the mean should be around one and I just feel like for this sample size it should be this one.”

For sample size 100 she chose F. This is “because it should be less varied.”

Applet Demo:

She recognized that with an increasing sample size variability will decrease.

She then stated that with sample of size 4 the variance would divide in half.

Question 2:

For sample of size 4, she initially chose B, but then remarked that she was stuck between C and B. She liked B because “it’s varied, but C is more varied and has more observations (or means) that way (referring to the tail of the left side).” Then she chose C”

For sample size 100, she chose between D and E. Initially she chose E, but then looked at the mean and realized that the graph was not centered about 19, so she chose D because it was centered at 19.

Question 3:

She used the applet to solve this problem.

She selected sample of size 5 and 25 and repeated about 15 times.

She chose B because “it’s centered around ‘here’ (pointing to the mean), and I think that and you add quartiles, and then look at the parent quartiles, and it just makes sense.”

For sample of size 100 she chose C. "Because the quartiles still have those outliers, its centered around 0 and it still should be a little varied. But I don't think it would be as varied as F or as small as D, so I chose C."

Question 4:

Using the applet she took a sample of size 5 and size 25. She took 1000 samples, and then compared the mean of the applet to the question at hand.

I had to inform her that the applet had a different mean than her question, so that she should not use those values (from the applet) as a guide to her solution.

For sample of size 4 she chose A. This was because "a sample of size 2 (pointing to the applet) is like really spread out, but it's kind of in the middle, and more spread out compared to all the other ones."

For sample of size 100 she chose E because she was looking for a really skinny one. "it's really tall and skinny"

Question 5:

For sample size four she initially chose C "because it is varied but not too varied." When I asked her again about her selection for sample size four she stated, "it actually might be D, because I was using the applet. Originally I thought it was too varied, but then I was looking for C and sample size 25, being my size four, but then I realized it was D."

For sample size 100, she chose A, because it had the smallest amount of variability.

Question 6:

For sample size 4 she chose C. “I chose it because it is not exactly going to look like that graph (pointing to the parent), but it will still be varied, and so this would be my only choice. The others are not varied enough.”

For sample of size 100 she chose B. This was because it was not as varied but still had some variability.

Interview 6: Sensor= -7

Question 1:

For sample size n=4 she chose A. This was because it looked like the parent but was not quite like the parent because with size four you wouldn’t be able to achieve it exactly. For sample of size 100, she chose A because it looks like the parent and with a large sample size she thinks you will achieve a sampling distribution with shape just like the parent.

Applet Demonstration:

Student remarks that (the distribution of means for) sample size of 5 is wider than a sample of size 25, whereas 25 is more centered. And she also remarked that a sample of size 16 is still more spread out that a sample of size 25, but not as wide as a sample of size 5. (based on observing the applet demonstration)

Question 2:

Contrary to the selections the student chose for question 1, this time she was looking for distributions that were more normal, and narrow. Not like the parent, as she chose for the first question. This was due to her observations from the applet demonstration.

Student chose F for a sample of size 4.

This was because "Because I did it sort of like the last one, where it is wider if you have less samples and they get steeper if you have more samples."

Student chooses E for a sample of size 100.

This is because "it got narrower".

She did not pay attention to any of the other graphs for the choices, and completely ignored the center or standard error.

The selection E was actually a sample of size 30, shifted over so that it did not even capture the mean of 19.

Was she in a hurry? Did she just not think that the information in the question was relevant?

D was the exact same shape and spread as E, only it was centered around the mean of 19.

Question 3:

Student was able to use the applet for this question.

After reading the question, she asked if the mean of standard deviation would be the same on the applet.

I let her know that they would not be the same since the scale was different, and also pointed out that she could not simulate a sample of size 100, but would need to draw conclusions, and perhaps make inferences, based on what she observes.

By using the applet, she simulated 1000 iterations of sample size 5 and sample size 25. She then used the images she created on the applet to select her answers.

She first chose a solution for sample of size 100. She chose F initially and then chose B for a sample of size 4.

She said she chose F for sample of size 100 because, “Like you said, this is not sample of size 100, (pointing to the distribution of sample size 25 on the applet) so I feel like it (the solution) would be a bit wider.”

-(me)- so you think that sample of size 100 would be more variable than sample of size 25?

“yes”.

-(me)- So the solution you chose is for size 100, not for size 25. Do you still think that based on what I’m still telling you that F would be the correct answer?

“oh no!” (she confused herself with the applet, thinking that the applet distribution was of size 100 and her solution she needed was for size 25)

“so I feel it would be closer to C because it is narrower”

She stated that she chose B for a sample of size 4 because “it looks like this one” pointing to the applet sampling distribution of size 5.

B was actually too varied and the correct solution should have been F for a sample of size 4.

C was a sample of size 15, so the solution should have been D, but this time she had the center and spread correct.

Question 4:

Again, the student was able to use the applet.

She changed the distribution to a Uniform, and then drew 5 samples of sizes 5 and 25. She then immediately clicked on 1000 samples and made her selections.

For a sample of size 4 she chose A.

“Because I feel the base is getting wider, and D kind of confused me.”

-so what are the differences between D and A?

“the shape of it”

-And you are looking for ?

“more of a distribution curve and this one (D) is more scattered”

(by this I think she means normal curve)

-and you immediately chose E for the sample of size 100

“Well this one is narrower than the sample of size 25, and I think this one will be narrow as well.”

Question 5:

Again, she used the applet and hand drew a bimodal distribution.

She then clicked on 1000 samples.

She chose D for the sample of size 4.

“Because when I looked at (sample size) 2, there still seemed to be a distinction between the two. It looked like it was spread out, but it didn’t show a bimodal”

-now what about a sample of size 100.

“I don’t really see a difference between E and F”

-so try to look at the graph closely and see if you can tell me what is different.

(the student looks at the two and then looks at the applet distribution of n=25.)

After looking for a few seconds she chose F and I asked her why?

“Cause I felt like, (pause) now I’m looking at C. (scratches out here solution) Now I’m going to go with C. Because if you have this bimodal (pointing at the parent population) you will pick up some within this range (the left mode), and I feel like the tip is going to be closer to this side (pointing to the left size of the peak in C) and I feel like if you look at 3.3 on this graph (F) then it's not enough, where C picks it up."

-So what is the mean of the parent population?

"is it a negative?"

-so I point to the question and show her

"oh, 2.8"

I then point out the mean of the applet parent that she created and asked her to look at the means of the sample of size 2 and 25.

She read them off the applet and then I ask

-Are the means of the samples close to the mean of the parent?

"right, now these 2 look alike, (pointing to C and F), It's between C and F."

(note: C is a distribution of size 15, F is of size 30,)

-so I then prompt, If you are looking for a sample of size 100, what are the differences between C and F?

"the narrowness, the width. I guess when we have more samples it would be wider"

-(HMMM?) You think it will be wider with an increase in samples? Does variability increase or decrease with sample size?

"oh it gets narrower"

-(I'm still not sure she really is understanding this. She keeps reverting back to stating that a larger sample size yields a wider sampling distribution."

She finally chose F because it is narrower. The actual solution was A

Question 6:

For sample size n=4, she chose C.

"Once again I looked at the mean (based on my prompt from the previous question). And then I looked at the size, and it's going to be wider with a small sample"

For sample of size 100 she chose B.

"Well I had to choose between B, D and E."

-so what's the difference between those 3?

"Well I knew it wasn't going to be D because it is too wide, and it needs to be narrow."

-So why not E?

"I don't know, because it didn't look like it could be a choice. So I just chose B."

Additional Comments:

How do you feel about the class?

"Not very well. I'm really good at math, and actually trying to become a math tutor, but the professor confuses not only himself but the rest of the students"

Doesn't demand control of the classroom.

And he cannot answer questions presented to him in the class. He often answers with Yeah, not giving us direct answers.

What do you like about the class?

Examples are good. I can understand them better when he shows them

I hate that the exam is multiple choice. I usually get the numerical answers right but the multiple choice ones wrong.

Hands on activities are not that great. She does not take much from it since she feels like the other students are just as lost as her. She prefers him to walk her through the examples.

Interview 7: Sensor = -9
Question 1:

For sample size 4 she chose A because

“When you choose four at a time, it’s still going to be a lot like this (pointing to the parent population)”

For a sample of size 100 she chose A again.

-so you think that no matter what sample size you take, when you take 500 repeated samples you are going to get a distribution that looks like this parent?

“yeah”

Applet Demo:

-do you think when I continue to draw samples (of size 5) the means are going to fall into a general area or spread all across?

“I think with more data they will fall into a general area because there is more data over there (pointing to the left side of a right skewed graph)”

I then show 500 samples and ask her to find the mean of the parent population.

“8.0”

And what is the mean of the sampling distribution?”

“7.9”

-So you notice that they are very close.

“right”

-Do you think if I take a distribution of a larger sample size this plot would be the same or different?

“I think the bell curve will be different”

-Do you think it will be wider or skinner?

“skinnier, but I’m not really sure”

I then do a few more demos and she gets that the sampling distribution will have a smaller spread with a larger sample size

Question 2:

For sample of size 4 she chose C and for sample of size 100 she chose D.

“ Because with four you will have a really big spread because of the data, but with 100 you are going to get… (she pauses, cancels out her answer of D and changes it to B) the same ones over and over again, (referring to B.)”

Question 3:

With applet:

She takes samples of size 2 and of size 5, 1000 times.

She chose F for sample of size 4

“because compared to B, it won’t have that high of a variance, but will still have more compared to the other solutions”

For sample of size 100 she simulates samples of size 20 and 25 1000 times.

She was choosing between C, D and E.

She chose E because she felt like it was in between C and D and she’s looking for something kind of normal and felt it was the best fit.

Question 4:

With applet: She drew samples of size 2 and 5. And then chose A as her solution.

Because it looked like the best choice for what she saw on the applet based on the shape.

For sample of size 100 she simulated 16 and 25, 1000 times.

She chose E "because as it goes from 16 to 25 in your sample it continues to go up and get skinnier. So if you have 100, it would have to be like my choice"

Question 5:

She drew the bimodal and then simulated sample of size 2 and 5, 1000 times.

She chose D

"because it's kind of the in between for those and it will have a pretty high variance. And you will get a little more of a normal"

She then drew size 16 and 25, 1000 times and chose A.

This was because it was the skinniest one

Question 6:

For sample of size 4 she chose A. She still did not grasp that the sample would look

"Because I was looking for more of a spread

And she chose E for sample of size 100 because it was the skinniest AND "and because of the mean being 2.5"

Additional comments:

I then simulated the problem and showed her the sample of size 2 and 5 to lead her to choose C as a solution for the sample of size 4. I reinforced that anytime you have a sample size larger than 1, you will have a wide spread, but will not look like the parent b/c it will start to exhibit properties of normality.

She did state that she felt the applet was useful in explaining these concepts and that she wished her teacher would have shown her this applet.

She liked the hands on activities.

Needs more examples. Only getting one, and would prefer to see more than one problem worked.

Homework questions are worded differently than presented in class, and that is confusing to the student b/c she prefers to see things as presented to her before.

To solve a problem she pulls out all the numbers and tries to figure out how to calculate the

Appendix I – Task Based Interview Fall 2007 pre Assessment

Intuitor PRE-Assessment

What is a variable:

1- A number in a problem to help us solve the problem
2- Some quantity or quality that is measured for some notable change
3- Something/a number you want to measure/calculate
4- A value that varies depending on the sample taken
5- A letter representing an unknown number

What does it mean to say that observations Vary?

1- changes
2- Variation could be due to different relative frameworks by which something is viewed or variation occurs due to some stimulus or treatment
3- The observation/conclusion may not always be the same number/answer
4- When taking different samples, the observation you're looking for (dependent variable) may change depending on your sample. It may also mean that two people looking at the same data may draw different conclusions.
5- There are many different possible outcomes for the sample

WITH PICTURES?

1- drew a grid with some x's all over

2- happy face to smiling face, 2 rectangular prisms facing different ways

3- Drew histogram with age (x-axis) and grade received on y-axis all were three different heights.

4- Histogram of years (x-axis) and Number of hamburgers eaten in billions (y-axis), and # eaten between years vary

5- 3 normal curves with different areas shaded. The varying part is X, the number you got.

Which has more variability?

1- B.

2- B has more b/c it has a larger range and yes it represents a normal distribution (roughly speaking)

3- B, since the range of the variables is greater. And, yes it is representative of a normal distribution based on its shape.

4- B. Because it has a wider spread. And yes it represents a normal based on its shape

5- B. Yes it's shape is bell shaped curve, and it is fairly symmetrical on both sides.

What does a population distribution consist of?

1- Was not sure what it consists of.

2- With regard to populations, distributions concern the actual "count" of a specific parameter. Based on sampling, we aim to use statistics to estimate actual distributions of some characteristic, (the parameter of interest)

3- They can represent facts or opinions. This participant sort of confused this with a sample, stating that not all possible outcomes will occur b/c people will be more likely to say one thing over another.

4- Mean and standard deviation.

5- Not sure

Id an observation on a POPULATION histogram:

1- yes.

2- Yes

3- Yes

4- Yes

5- yes

Using histogram, tell about weight of players in NBA

1- most between 170 and 300

2- mostly weigh b/w 175 and 260 lbs.

3- Most of them weigh between 200 and 250 lbs. All of them fall between 100 and 400

4- Majority fall between 150 and 300. But varies from about 120 to about 350.

5- It varies. Most data lie between 130 and 300

How many <150?

1- 1 player

2- Three

3- One player

4- 1 or 2

5- 1

How many > 300?

1- 4 players

2- Five

3- Three

4- 3 or 4

5- 3

ID a single observation on a HISTOGRAM OF (team) AVERAGES:

1- no(but not sure b/c video not on point)

2- no, this is an ecological fallacy. Based on these numbers, only assumptions about teams (the aggregate unit) can be made

3- No because the graph represents the average weights of each team and thus the graph is not specific enough to locate one individual

4- No because his weight is not on the graph and because teams could consist of very heavy of light individuals. (not quite getting it here)

5- No, I would need other players weights to get a mean for the team

Can you find STD of Sampling Distribution if pop has STD of 12?

1 – this was not on this person’s assessment

2 – from graphical inference estimate standard deviation of graph 2 is about 5

3- Thought that the sampling distribution would have the same standard deviation as the parent population

4 - standard deviation = 3. He knew the formula and derived the solution from it.

5- 3. She recalled the formula.

Sample Stat vs. Pop Parameter

1- Both correct

2- both correct

3- Both correct

4- Both correct

5- Both correct

Normal Distribution:

1- drew it correct. Drew one with higher variance as wide and "mushed". The smaller variance graph was tall and skinny and more "focused". Drew the shifted mean distribution over 5 positions. She displayed an understanding that areas with taller peaks are more probable to occur.

2- Normal distribution is symmetric, characterized completely by mean and std dev. Drew a normal correct and labeled mu, sigma, and 2*sigma and frequency on y axis. Larger variance has wide smaller has tall skinny. Mu+5 was the shifted one. The area is still the same. Probability density function is all points of probability like along the x-axis and the area of a true normal distribution is one. (in agreement with the laws of probability) and given areas represent a probability that a value will fall in a given range.

3- A normal distribution is symmetrical. Did not label the axes, but did locate the mean. Drew a tall skinny for smaller variance and fatter one for larger variance. Knew that it would shift over and remain the same shape if the mean increased by 5.

4- Normal distribution is bell shaped or symmetric. He drew a standardized normal curve and labeled the confidence limits up to 3*sigma, He drew a tall skinny for a smaller variance and a fat wide for a larger variance. He was able to draw a shifted normal curve for the mean increased by 5.

5- A normal distribution can be represented by a bell shaped curve or a large random sample. Drew the curve with mean of zero and y-axis is the frequency. The small variance was taller and a bit skinnier and the larger variance was more spread out. She was able to draw the shifted mean normal correctly.

Sensor PRE-Assessment:

What is a variable:

Something that changes

Something you are testing

An item that can or cannot be changed based on the observation. It can affect outcome.

What does it mean to say that observations Vary?

Results or outcomes can be different

Observations can vary dependent on a person’s view of the situation or data, or they can vary based on different data that is found

People have different views and/or opinions on something that is seen.

WITH PICTURES?

Many distributions were drawn (bimodal, right skew, left skew, normal)

Scatterplot(no correlation) and Scatterplot(correlation)

Normal distribution, and stated that they vary with time. Or that time is the varying factor.

Which has more variability: All chose B b/c of the spread

What does a population distribution consist of?

See video

Consists of differing pieces of data from a population

Amounts of people

Id an observation on a POPULATION histogram:

All three correctly did this

Using histogram, tell about weight of players in NBA

Avg is 225

Majority weight over 160 lbs.

Avg is around 230-240

How many <150? 1 (all three)

How many > 300? 3 (all three)

ID a single observation on a HISTOGRAM OF (team) AVERAGES:

See video (?)

These increments are pretty small, so if he is 280. She then re-reads the problem and states, oh, you won't be able to find his average weight, but maybe his team's average weight.

No (confused) because the average of the team was around 240 which are not represented on this graph (despite repeated attempts to show her that the histogram is of average per team, not person)

Can you find STD of sampling distribution if population has STD of 12?

Yes, (uses formula) = 3. is this on her own? (see video)

Less than 12. (so understands it will be less) I then give her the formula and she calculates.

Had help with formula. Had no idea otherwise.

Sample Stat vs. Pop Parameter

All were correct in this

Normal Distribution:

BELL CURVE. Count (y) and age (x). marked the mean

Drew smaller variance normal from 0-4. larger one 0 -1000

Shifted to the right by 5.

Bar Graph (histogram). Count (y) height (x)

Drew smaller variance skinny. larger one is fat. Needed help with this one. She was not sure how to represent the idea of variance graphically, so I would say she lacked a solid understanding of this concept. Stated that it would Shift to the right by 5.

The student stated it was a Bell curve. She then tried to recall a formula, but could not do so. She drew the normal distribution correctly. Only labeled x at center. Drew one fat and one skinny.

The student drew the distribution with a shifted mean taller to represent the increase of 5.

Appendix J – Task Based Interview Fall 2007 pre Assessment Results

Sensor PRE-Assessment:

What is a variable:

1. Something that Changes
2. Something you are testing
3. An item that can or cannot be changed based on the observation. It can affect outcome.

What does it mean to say that observations Vary?

1. Results or outcomes can be different
2. Observations can vary dependent on a person's view of the situation or data, or they can vary based on different data that is found.
3. People have different views and/or opinions on something that is seen.

WITH PICTURES?

1. Many distributions were drawn (bimodal, right, left skew, normal)
2. Scatterplot(no correlation) and Scatterplot(correlation)
3. Normal distribution, and stated that they vary with time. Or that time is the varying factor.

Which has more variability: All chose B b/c of the spread

What does a population distribution consist of?

See video

1. Consists of differing pieces of data from a population
2. Amounts of people

Identify an observation on a POPULATION histogram:

All three correctly did this

Using histogram, tell about weight of players in NBA

1. Avg is 225
2. Majority weight over 160 lbs.
3. Avg is around 230-240

How many <150? 1 (all three)

How many > 300? 3 (all three)

Identify a single observation on a HISTOGRAM OF (team) AVERAGES:

1. **Not sure**
2. These increments are pretty small, so if he is 280. She then re-reads the problem and states, oh, you won't be able to find his average weight, but maybe his team's average weight.
3. No (confused) because the average of the team was around 240 which is not represented on this graph (despite repeated attempts to show her that the histogram is of average per team, not person)

Can you find STD of sampling distribution if pop has STD of 12?

1. Yes, (uses formula) = 3.
2. Less than 12. (understands it will be less) I then give her the formula and she calculates.

3. Had help with formula. Had no idea otherwise.

Sample Stat vs. Pop Parameter

All were correct in this

Normal Distribution:

1. BELL CURVE. Count (y) and age (x). marked the mean

   Drew smaller variance normal from 0-4. larger one 0 -1000

   Shifted to the right by 5.

2. Bar Graph (histogram). Count (y) height (x)

   Drew smaller variance skinny. larger one is fat. Needed help with this one. She was not sure how to represent the idea of variance graphically, so I would say she lacked a solid understanding of this concept. Stated that it would Shift to the right by 5.

3. The student stated it was a Bell curve. She then tried to recall a formula, but could not do so. She drew the normal distribution correctly. Only labeled x at center. Drew one fat and one skinny. The student drew the distribution with a shifted mean taller to represent the increase of 5.

Appendix K – Interviews

Based on the results from the preliminary round of focus groups, we decided that it was better to switch to task based interviews to eliminate the interaction between students since we are unable to assume that all sensors or intuitors understand the material at the same level based on just the learning style.

We conducted one on one interviews with both sensors and intuitors. The format of the interview was as follows:

A histogram was presented to the student and six choices were given. The student was asked to select which one best represented a distribution of 500 sample means of sample size 100. The actual handout is given in Appendix G. Following this question, the applet on sampling distributions was demonstrated and then the Question 2 is presented. Questions 3 and 4 were to be used along with the applet (the student was allowed/encouraged to use the applet to simulate the scenario that they were presented). Then the final question was presented without any further aid. Finally a few follow up questions were asked specific to attitudes and opinions about their statistics class.

Three sensors and two intuitors participated in these interviews for the spring 2007 semester.

In answering the first question, without any aid, recall, or assistance, two of the three sensors selected the correct solution to this problem as well as one of the two intuitors. One sensor stated that she chose "C" because everything in class is normal, so speculated that the solution would be normal. The theme of normality clearly stood out in this students mind, but did not carry with it any context or deeper meaning, as she was unable to elaborate any further. She will be referred to as Sensor_One. The other sensor (Sensor_Two) that chose the correct solution stated that she had no basis and it was a complete guess. The sensor that selected an incorrect choice (Sensor_Three), selected the distribution that looked most like the parent population because that is what she expected from a distribution of sample means.

The intuitor that selected the correct response stated that “I don’t really understand quite what is going on, and I seem to remember something about if you get a large enough sample size and then it is normal, and that (pointing to his selection) is normal.” This shows that he has a slight understanding of the central limit theorem, but from this information, I was unable to gather how deep his understanding was. I will refer to him as Intuitor_One. The other intuitor (Intuitor_Two) chose the distribution that looked most like the parent and stated that his selection was a guess.

All five students watched the applet demonstration and then were asked to describe what they saw. Intuitor_One states that the more sample mean you take; it is going to the real population average. Intuitor_Two states that the means would keep stacking up and that all the means will fall into the same area and form a symmetric shape.

Sensor_One expects the means to change because we are taking a different sample each time, but not much because it will bulk in the middle. Sensor_Two expects the pattern (of the means dropping down on top of one another to continue), and Sensor_Three notes all the means are grouped together and form a normal distribution.

Overall, we believe that the students are getting an idea of “shape” with sampling distributions, but are not paying attention to other key details such as center or spread.

Progressing to the remainder of the interview, Intuitor_One was able to correctly answer each question. He worked quickly through the problems and did not use the applet to solve the problems, but instead, used it to verify the solutions he chose. He paid attention to both the shape and the center of the answer choices. Several solutions were normal distributions and could have easily been selected, but one was scaled outside the range of the population mean, and thus was an infeasible solution.

Intuitor_Two was able to take the applet and correctly solve the remaining problems. He chose the distributions with symmetric shapes and used the numbers to verify his selection. When asked about the class, Intuitor_Two stated that the wording was very confusing and he prefers to work through problems rather than attend a lecture where all the information is presented on the chalk board.

For question 2, Sensor_One chose a left skewed distribution to match the parent population. She thought it was a good representation of the entire population. For question 3, she chose the correct answer and paid attention to the numbers at the bottom of the graph. For question 4, she chose a highly varied normal distribution, and for the final question, she chose a right skewed distribution that was tending towards normality because she felt like the bulk on the left side best captured the solution, but wanted the tail to be represented as well. It is clear to me that, even with the applet demonstration; this student was not making the proper connections. When asked how she felt about the course, she stated that she prefers to see a similar problem worked out so that she can just plug in the numbers. She also stated that she has trouble relating symbols to formulas, for example, she asked, What does xbar mean, and how do I get it?

Sensor_Two chose a solution with the correct shape, but did not notice that the center was incorrect. She does the same thing for question three, she chooses the normally distributed graph, but ignores the detail. She correctly answers the remaining two questions and based her solution on their shape. She comments that each question has two or three solutions that are normally distributed which is making it difficult for her to decide. When asked about the class she states that she prefers to plug and chug and hands on activities.

Sensor_Three chose the correct answer for problem two basing it upon the symmetric shape. For question three, she chose a normal distribution, but ignored the number detail, and thus chose the solution that was centered on a lower mean. For the fourth problem, she chose between the two normal distributions, but ultimately chose the incorrect solution, again not taking heed to the detail. She did the same thing for the final question, selecting the normal distribution, but ignoring the mean given at the bottom of the graph. When asked about the course, she states that she likes the examples, but dislikes hands on activities (waste of time) and dislikes being quizzed on new material.

It was surprising that the two intuitors paid more attention to the detail of the graphs than the sensors. Even though sensors are more detail oriented, perhaps the thought process leading them to look at the numbers is more intuitive. Since the sensor was not told to look at the details of the graph, perhaps they did not think that it was an important feature, whereas the intuitor instinctively knew that this information was just as significant as the shape of the graph to the solution.

In the fall semester of 2007, we added to the interview format by performing a pre-assessment prior to the task based interview (see Appendix E).

Before the students were asked to answer questions about sampling distributions, we first gave them a handout to complete, asking questions to try to see what their background knowledge was regarding sampling, variability, normal distribution, and distribution. We were able to interview five intuitors and three sensors for this semester.

The first question asked the student to state what a variable was. The sensors stated that it was something that changes, something you are testing, and an item that can or cannot be changed based on the observation, but affects outcome. The intuitors stated that a variable was a number, a quantity, a measure, a value that will vary and an unknown number. These responses between the groups differ in that the intuitors are thinking of a variable in terms of a number or quantity, whereas the sensor is thinking of it in more general and vague terms, an item that may or may not change.

When asked what it meant to say that observations vary, the sensors stated that it meant a point of view or the data can change, or the outcomes will be different. The intuitors stated that the observations will change, not always be the same, there are different possible outcomes, changes due to something (as in a dependent variable), and that the framework by which something is viewed could cause variation (not really answering the question, but more-so stating why variation occurs). Here, the sensors are talking about in more specific terms, and the intuitors are being more general.

All students were able to select which normal distribution had more variability and attributed their response to reasoning that it had a wider spread or range. None of the students were really able to state that a population distribution consisted of everything in the population, or all possibilities, or anything of the sort. One sensor stated that it consisted of amounts of people, so counts. The other sensor stated that it consisted of differing pieces of data from a population. The third did not answer this question. Two intuitors were not sure how to answer this question, one stated that a population distribution consisted of a mean and a standard deviation, and the fourth stated that a population distribution represents facts or opinions, which I sort of believe she was confusing this with a sample because she also stated that not all possible outcomes would occur because people are more likely to say one thing over another. The fifth intuitor wrote that "with regards to populations, distributions concern the actual 'count' of a specific parameter. Based on sampling, we aim to use statistics to estimate actual distributions of some characteristic (the parameter of interest)".

All students could identify an observation on a population histogram. They were then asked if they could identify the same observation on the distribution of average team weights and to explain why or why not. All five intuitors stated that they could not identify the value, where four of them stated they needed the team average, and that this single point could not be represented by averages. The fifth intuitor attempted to find the observation, and upon realizing the scale did not capture it, stated that his weight was not on the graph. (The weight is 280, and the averages ranged from 210 to 240).

All of the sensors immediately tried to find the point 280 on the graph, and ignored the fact that the histogram was of the team averages. Two stated that they could not find it because it was not on the graph, the other sensor stated that the average of the team was 240 which were not represented on the graph. This sensor took the first histogram (of all the players) and used that average in their reasoning.

This exercise made it clear that sensors and intuitors are thinking differently about sampling distributions. The intuitors did not have a difficult time thinking about the histogram of averages and realizing that a single observation could not be found. The sensors, however, had complications thinking about a single observation as part of a whole, the whole being the team, and the single observation being the individual player. They had a difficult time reasoning why a single observation could not be depicted on a chart of averages, and even with some explanation, still appeared to be confused.

The students were then asked if they could calculate the standard deviation of the sampling distribution if the standard deviation of the population was 12. (There are 16 players in each team). One of the sensors immediately used the formula and was able to calculate the value. The other two sensors had no idea, but knew that there was a formula to get to it. One of these sensors stated that it would be smaller than the population standard deviation, but had no recollection of what the formula was.

Two of the intuitors knew the formula and immediately calculated the standard deviation. One of the intuitors did not have this question because it was added after the first interview. One intuitor thought that it would be the same as the parent population, and the other eyeballed the sampling distribution and stated it would be five, based purely on the graph.

The next question presented in this pre-assessment asked the students to identify (from a word problem) if a value given was a sample statistic or a population parameter, and then state if the sample statistic given was mu or xbar. Both groups were able to identify these terms from the problem.

Finally the students were asked to draw a normal distribution, and then below it, draw one with a large variance and one with a smaller variance, and then draw what the graph would look like if the mean increased by 5. All of the intuitors drew a tall and narrow normal curve to represent small variance and a fat and wide normal curve to represent larger variance. They also all correctly drew the shifted curve (to represent an increase in mean). Two of the sensors were able to do the same thing, but the third sensor drew a taller curve to represent the mean increase by five.

Overall, the sensors and intuitors had some similar conceptions of these ideas, but several of the thoughts differed. An idea of sampling distribution of means was not as easily grasped by the sensors.

The overall understanding of these key concepts was above average, but not exceptional. There are definitely gaps in the thought process and understanding of components leading up to sampling distributions, and the assessment of the sampling distribution questions may lead to our further understanding of these differences.

Appendix L – Four Foundational Questions Intuitor and Sensor

Intuitors, four foundational questions				***SHAPE***				
				CORRECT		INCORRECT		
				LIKELIHOOD		***LIKELIHOOD***		
				CORRECT	INCORRECT	CORRECT	INCORRECT	
SPREAD	CORRECT	***Center***	Correct	3	1	1	5	10
			Incorrect	1	0	3	2	6
	INCORRECT	***Center***	Correct	4	3	15	10	32
			Incorrect	1	0	19	12	32
				9	4	38	29	80

Sensors, four foundational questions				SHAPE				
				CORRECT		INCORRECT		
				LIKELIHOOD		***LIKELIHOOD***		
				CORRECT	INCORRECT	CORRECT	INCORRECT	
SPREAD	CORRECT	***Center***	Correct	10	8	16	17	51
			Incorrect	5	4	17	12	38
	INCORRECT	***Center***	Correct	16	7	35	56	114
			Incorrect	5	7	59	45	116
								319

Appendix M – Text Mining Per Group

Additionally a question of likelihood is presented. This question requires an understanding that a smaller sample mean is less likely to be representative of the population parameter.

2. According to a recent report the average income of people in Adair County, Missouri is $31,023. Which of the following is more likely? Pick one and explain.
    a. We take a random sample of 10 people from this county and find that the average is over $50,000.
    b. We take a random sample of 200 people from this county and find that the average is over $50,000.
    c. We do not have enough information to answer this question.

TOPIC Likelihood : INTUITOR	a	b	c
Small, Small Sample Size, Sample, Population, likely	13	1	0
1 More likely to pick in small sample 2 Smaller sample increases the chance of wealthy people 3 sample of 10 people is more likely because it represents a smaller number of the sample.			
Closer, Population, Mean, sample size	5	2	2
1 it depends if the distribution is skewed or normal. If its normal, then it is more likely since the more the people there are in the sample the closer to the actual mean it will be. If it is skewed it might be larger. 2 sample means are usually centered about the population mean 3 option a seems more likely due to the small sample size. As N increases, should get closer to true population			
Random, Random Sample, Skew, predict	5	1	1
1 More likely skewed 2 this random sample could be more different because the number of persons picked is so small it could be skewed 3 we have a basis to predict and a random sample of 10 people is way too small			
Average, value, Predict	2	1	4
1 we have no prior knowledge of the sample except an average so to predict whether 10 or 200 would have average over 50K is impossible to guess 2 50K is a great deal over 31K. We aren't able to see which one will be more likely to predict the average because we have no information on the max and min income and I feel that would greatly help. 3 The sample will probably be centered around the same average as the			

population			
large, Large sample size	5	0	0
1 Having a larger sample ensures selecting some people from the lower quartile. 2 larger the sample the less likely you are to get an extreme sample since larger sample chance events are more likely to even out. 3 margin of error is reduced with a larger sample size, because of the equation as well as df for the t-value. It is more likely to get extreme data with smaller sample sizes.			
Great, Sample, Mean	5	2	0
1 less sampled, making it more likely that sample mea would be a lot greater than the real mean 2 Smaller sample more likely, large is more like population mean 3 more sampled more likely to get higher average since spread will be greater as n increases.			
High, Little, value	4	1	1
1 no details as to why either should be higher 2 higher sample better representative 3 10 people sample is not very accurate and could have people with higher income			
Sample Size, Result, Entire	7	0	1
1 different results would occur due to a sample size that is too small 2 The parameter is for the entire town in which we do not' know how many people live. We could take either sample size and could still get such a result 3 Amount of error increases when sample size is smaller so the probability of getting a different mean score goes up with decreased sample size			
Data, Mean, Little	2	0	2

1 not enough info to know how to skew the data 2 because we aren't given any more data than just one mean. 3 small sample will decrease the ability for the data to be normal like the population and this is small			
Standard Deviation, Outlier, know	1	0	6
1 Do not know the standard deviation or the outliers 2 Standard deviation will change with n, so no idea 3 No STANDARD DEVIATION			
NO RESPONSE -			
NO TOPIC	1	0	4
- we have no idea how many people live in the county - we do not have enough info - we are not told what proportion of people make over 50K - because this is unrelated to the report - no reply			

TOPIC Likelihood : Sensor	a	b	c
Likely, small sample size, small, sample, large, likely, find, high,	36	1	0
- A sample size of 10 is not enough to cover entire pop of interest. It is likely the average could be over 50K with such small sample size - Sample size is small so more likely - with smaller sample, the deviation is more likely to be great enough to allow that possibility			
average, value, find, likely	11	10	19
- because the average income is 31K there's no way to tell who is likely to have an income of over 50K - Bigger sample size more accurately reports average income - if we did a SIMPLE RANDOM SAMPLE then we have no idea who makes over 50K for their average income. To be normally distributed, there has to be more than 30 subjects anyways.			
high, chance, value, average, great, little, pick, mu, good, small sample size, small, high chance	24	3	1
- Less in sample, more chance to be skewed - The likeliness of picking 10 people who happen to have high incomes for this county is higher than the chance of picking n200 people with high incomes. - chance of 10 people with average over 50K is higher than chance of 200 having an average over 50K			
large, sample size, large, increase, chance, good, sample, variability	14	16	1
- variability will increase with larger sample size - the larger the sample size, the lower the variability in the sampling distribution			

- larger sample gives less variability			
random, random sample, sample, large, large sample size	2	7	12
- if the average is mu, we cannot possibly fairly say that we will have an average of 50K in our random sample. - The random sample is over 30 which gives a better prediction. Also the report states all people - Sample size does not affect likelihood of events if they are random			
mean, closer, increase, actual, population	23	2	3
- as n increases the distribution becomes more normal - the high income is a lot greater than the mean which probably means these 10 people are the wealthier of the people in the county - With 10 people, sample is small enough it could still miss the normal distribution but as the sample gets larger, it should go closer to average			
predict, basis, likely, standard deviation, know, size, population	5	2	20
- we have no idea which 10 people or which 200 people we'll choose, so we do not have a basis for predicting which is more likely to have an average over 50K - smaller sample have greater STANDARD DEVIATION - less subjects less likely to be like the population			
population, average, representative, size, entire population, entire, large, good, represent, big, sample, closer	21	10	6
- Small size less rep of population - Sample of 10 people is not large enough to represent entire population and one person with a high income will change the average much more with 10 people than with 200. - it is a very small sample causing it to be less representative of the			

population			
outlier, data, small sample size, sample size, skew, small, great, mean, effect	25	3	0
- because sample size is much smaller so outliers would have a substantial offset on the data - those 10 people could easily be outliers based on what part of town they live - more people sampled the less outliers will affect the average. If sample 10 and 2 have very large income, overall mean goes up			
Information, know, population, total	2	0	29
- We do not' have enough information - only know the average and not the range. Need more info - we do not know population total so need more info			
NO RESPONSE	2	0	3
NO TOPIC	0	1	5
1 There is not enough information 2 Random so no way to tell 3 Not enough information 4 n=10, no enough to prove something, when working with a normal distribution, we need at least n=30 from n=10 we cannot draw a conclusion 5 Not enough information 6 Not enough evidence			

A large multinational company knows that the average age of their employees is 34 years. They also know that the standard deviation of the ages of these employees is 8 years. A manager from human resources is going to randomly select a sample of 100 employees. Consider the sampling distribution of the mean of a sample of size 100 from this population. From the information given, which of the following statements are true about this sampling distribution?

The questions from this problem aim at understanding how students perceive the center, shape, and variability of the sampling distribution in comparison with the parent population. The student is encouraged to explain their choices and is penalized if they do not.

3. Which of the following is true? Pick one and explain.
    a. The sampling distribution will be centered at 34.
    b. We do not know where the sampling distribution will be centered.
4. Which of the following is true? Pick one and explain.
    a. The sampling distribution will have a higher standard deviation than the population.
    b. The sampling distribution will have the same standard deviation as the population.
    c. The sampling distribution will have a smaller standard deviation than the population.
5. Which of the following is true? Pick one and explain.

a. We know that the shape of the sampling distribution of the mean will be right skewed.

b. We know that the shape of the sampling distribution of the mean will be approximately symmetric.

c. We cannot tell what the shape sampling distribution of the mean will look like.

TOPIC Center - INTUITOR	a	b
1- Sample Size, distribution, normal, large	7	2
- because the sample size is large the distribution will be approximately normal - chose appropriate sample size thus they will be equal - because the sample size is larger and random. Xbar is normal when n is large		
2-large enough sample, enough, large, sample	3	1
- because it is a large enough sample - you have a large enough sample the average should be 34 - 100 is a large enough sample to be considered normal		
3- sample distribution, center, normal	6	4
- because sampling distribution is centered at the mean - in the sample there could be an outlier that throws off the mean, which could end up in the sampling distribution not being centered at 34 - sampling distribution is centered at 34 because normal		
4- average, know, sampling distribution	5	3
- know average age and n is large so centered at 34 - Average=mean		

- because it’s the average age		
5- population, sample, distribution, mean	3	6
- it’s not the entire population, just a sample so mean can change - Sampling distribution will be centered around 34 - 34 is mean of the population. The sample mean could be a different value		
6- data, normally, distribute, know, median	1	2
- we do not know whether or not the data are normally distributed - do not know if data are normally distributed so data can be heavily skewed in a certain direction, making the mean change but not the median - This is due to the large sample and randomization, causing the sample to be normally distributed		
7- mu, sampling distribution, center	8	2
- SAMPLING DISTRIBUTION have same mean as pop - since mu=34 it should be centered there - Sampling Distribution always settled at mean		
8- center, do not know, normal distribution, average	4	3
- 34 is mean which put it in center - N>30 so normal distribution - do not know if it’s a normal distribution so cannot say if 34 will be the center		
9- likely, close to, standard deviation, mean, sample size	1	2
- do not know because it is random and they might randomly select all older - Because you could randomly select many younger workers and the mean of the sample will be much lower than that of the parameter - With such a low sampling distribution you can assume that out of 100 people randomly selected their average age would be 34		
10-select, randomly, sampling distribution, random, parameter	4	2

- mean is 34 and standard deviation is low - mean is likely to be around 34 since large n - since the sample is only 100 w/ standard deviation of 8, the mean could be close to 34 but not likely exact		
NO RESPONSE	5	4
NO TOPIC	0	1
because we cannot figure that out with the information we're given		

TOPIC Center - SENSOR	a	b
1 - Sampling distribution, mean, center, always, population	38	14
- the SAMPLING DISTRIBUTION of the mean will be centered at 34 is true because the mean is always the center for sampling distribution along with sigma and whether or not it is normal - if the mean is 34 then the sampling distribution will be centered around this - Sampling Distribution unknown because we have yet to calculate the results from the Sampling Distribution of 100 people		
2 - large, large sample size, sample size	26	5
- large sample is likely to have mean centered at true mean - Large samples - because large sample size, we can conclude the SAMPLING DISTRIBUTION will be centered at 34		
3 - average (population, center, median, vary)	27	11
- Just because the company average age employee is 34 does not suggest that will be the average age of the employees in the sample of 100. - Center around average - Centered around the average		
4 - normal, distribution, population, large, sample	17	5
- normal distribution if population distribution (100) is normal - the mean is 34 therefore the distribution is centered there and the standard deviation are from there also - Normal distribution so it should have a location and the center does not change		
5 - random, random sample, sample, population, outlier, do not know	5	8
- sample is completely random so no way of knowing - since sample is large and random it can be considered a normal distribution which		

centers the mean around 34 - 100 random employees so we do not know where the SAMPLING DISTRIBUTION of the mean will be centered		
6 - mu, xbar, central limit theorem, normal, sampling distribution, large, approximately	25	3
- center at 34 because mu=34 and ages will fall within standard deviation of 8 - xbar is approximately normal when n>30 - according to CLT for large SAMPLING DISTRIBUTION, center will be at mu		
7 - population, sample, different, statistic, standard deviation	14	24
- Mean for sample is same as population - The mean of large sample properties reflects that of the population being studied or looked at. - the sampling distribution should reflect the overall population the statistic is about		
8 -know, normal, standard deviation	7	11
- it is known that it is 34 for everyone in the company - Do not know how many employees at company so cannot determine the standard error - we have not worked it out to know		
9 - median, skew, outlier	2	20
- Should be skewed so not represent entire population - With such a large number it could possibly be skewed or bimodal - Outlier will affect the mean		
10 - do not know, data, select, distribution, randomly	3	19
- The mean may vary from population mean because do not know data and is a simple random sample		

- SAMPLING DISTRIBUTION contains all the data sample sets that could be used but we do not know actual center - because the data of the sample has not been collected yet. The only numbers (data) available is for the entire company not, the sample		
NO RESPONSE	11	11
NO TOPIC	4	10
1 Because not enough info (and because are two same and it says pick one.) 2 because a simple random sample is going to be conducted 3 we do not have enough facts 4 It does not' change 5 cannot predict it will be exactly 34 6 Not enough info 7 not enough information 8 ? 9 because 34 is the age of most people 10 Until we work it out 11 Not enough info 12 all of the information leads to this 13 because we do not have enough info 14 since a & b are the same I choose c		

TOPIC Spread : INTUITOR	a	b	c
sample, distribution, large, overall, population	4	3	1
- the average age is 34 so with a sample of 100 the average age and sample distribution will be represented so the standard deviation should be about the same - less people in sample than population the standard deviation will be larger - sample of the population has a larger range of data to cover to support the overall population			
small, increase, standard deviation, people, population	4	1	6
1 Again, the graph will look different, smaller thinner perhaps, but normal 2 the denominator will be smaller and the STANDARD DEVIATION will increase for the sampling distribution 3 sampling distribution has smaller spread than population so smaller standard deviation			
+variability, more, increase, large, sample size	4	0	6
1 larger pop increase standard deviation 2 with sampling distribution you are able to quantify the variability from the parameter 3 the smaller the population the larger the sampling distribution will be due to increase in variability			
correct formula, less, sigma, great, small	2	0	6
1 correct formula therefore smaller 2 correct formula less than sigma 3 correct formula less than sigma			
mean, sampling distribution, distribution, standard deviation, small	1	4	3
- because the sampling distribution reflects the standard deviation of the population - standard deviation should not change just because the mean changes			

- sampling distribution is the distribution of the mean so values are closer together			
sample size, overall, small, data, distribution	3	1	2
- even with sample size numbers decreased to sampling distribution will stay the same - because the sample size is smaller than the population, the standard deviation will be bigger. - sample size small and sqrt(n) smaller so standard deviation will be divided less			
number, data, great, population, large	3	1	0
1 fewer numbers in SAMPLING DISTRIBUTION and the possibility of a greater range of data. 2 Standard deviation of mean is greater than population (incorrect formula) 3 sample size has statistically sound number of subjects, the data will be similar to the pop			
representative, population, large sample size, large, sampling distribution	0	6	1
- true because not fully representative of entire company - SAMPLING DISTRIBUTION is representative of population - Again, if it is an accurate representation of the population it should have the same standard deviation of the population.			
NO RESPONSE	5	3	4
NO CATEGORY	0	2	2
rumor has it			

DO not know size didn’t change so same because it is a normal distribution and CLT applies			

TOPIC Spread : SENSOR	a	b	c
Large, Small, Sample, N, Population, Sample Size, small sample, size, entire	20	2	14
- the 100 randomly selected will have a larger variance than the entire population make the samples STANDARD DEVIATION larger -since n=100 population would have a larger amount and smaller standard deviation - smaller sample, larger standard deviation			
Mean, Sampling distribution, population, same	6	14	22
- because SAMPLING DISTRIBUTION of means and they'll be clumped together more in a symmetric shape. - mean is average population could have outliers - because the sampling distribution is a much smaller group than the population creating a smaller standard deviation			
same, size, sample, stay, population, standard deviation	0	32	0
- because they have the same sample size - The standard deviation will also remain the same unless it is multiplied/divided by some number - the numbers are distributed the same			
decrease, increase, sample size, size	14	1	4
- because the sample got smaller and the standard deviation will increase - spread will increase because there'll be less data points and outliers will affect the standard deviation a lot - Spread will decrease			
more, people, spread, sample, little, variability	19	0	18
- Less people therefore standard deviation higher - Smaller sample size so more variability			

- if you have a smaller sample size there is going to be more spread to the data and more of a chance that the data will be further from the mean			
correct formula, less, correct, formula, sigma	2	0	31
- the standard deviation of sample will be less than the population because we're using correct formula - standard deviation=correct formula so standard deviation depends on sample size, so sample standard deviation less than population standard deviation. - since fewer in sample the standard deviation will be larger according to correct formula = s			
distribution, sampling, lower, small, sample	8	2	14
- SAMPLING Distribution is smaller so there will be less deviation - The SAMPLING Distribution is based off of averages of sample statistics making the standard deviation smaller - Standard deviation will be smaller because less people they're sampling			
great, number, sample, employee, age, population, mean, small	12	7	21
- Smaller standard deviation because sample and not whole population - average age of population is probably greater than 34 therefore the standard deviation would be higher so the company would be smaller - population is a greater number			
n, sqrt, deviation, standard, population, small, high, people	8	4	13
- SAMPLING Distribution is smaller so there will be less deviation - since n=100 population would have a larger amount and smaller standard deviation - Less people in the population therefore smaller standard deviation (n less than N so standard deviation smaller)			
sample size, same, large sample size, large, small sample size, small, data, less,	11	20	16

stay			
- because they have the same sample size - Smaller sample size so more variability - if you have a smaller sample size there is going to be more spread to the data and more of a chance that the data will be further from the mean			
NO RESPONSE	1	8	11
NO CATEGORY	1	2	1
1 Because it has to 2 because there are less subjects 3 we know this because it is a normal distribution 4 A few outliers might be randomly chosen			

TOPIC Shape : INTUITOR	a	b	c
1 - right, skew	9	0	0
- people under 18 not likely included but older would, thus skewing it to the right. - if it is skewed right we already know that there will be more people 34 and younger and less older workers, skewed right indicates less in upper quartile - if the population of employee ages are right skewed then the sampling distribution will also be right skewed because the sample is coming from a population that is right skewed			
2 - normal, approximately	2	3	2
- shape is approximately normal if the population parameter is approximately normal. - shape of Sampling Distribution will be normal if population is normal - because it is not normal we cannot tell because it does not follow the Central Limit Theorem anymore			
3 - sample size, large, large sample size, resemble	1	6	0
- because of the law of large numbers - the SAMPLING DISTRIBUTION will resemble the overall population given the large sample size and the mean will be skewed if the population is skewed - Because CLT we know that the SAMPLING DISTRIBUTION will be approximately normal w/ large sample size			
4 - know, standard deviation, mean	0	3	3
- Do not know because the outliers are unknown - always know what shape will look like - know what shape of Sampling Distribution will look like because we know mean, size and standard deviation			
5 - normal distribution, mean, sample distribution	0	8	0

- SAMPLING DISTRIBUTION follow a normal distribution around the mean - SAMPLING DISTRIBUTION follow normal distribution - It has a normal distribution			
6 - symmetric, shape	0	4	0
- SAMPLING DISTRIBUTION will follow a symmetric shape because they are randomly chosen - shape will be smaller spread with taller bell curve - With xbar and sigma we can tell the shape should be a bell curve			
7 - population, resemble, skew	8	0	0
- the skewness of the sampling distribution should accurately represent the population - Will be right skewed like the population - if population is skewed so will the distribution			
8 - sample, shape	4	0	3
- Samples take on relative shape of the parent population - depends on sample's variability - Shape of sample should reflect the whole			
9 - do not know, data, distribution, median	1	1	3
- we do not know which 100 people we may choose and the ages may average higher or lower than 34 - could be skewed we do not know - we do not know that the distribution is normal w/o first doing calculations to standardize it. it is possible the data from the sample could be skewed			
10 - information, enough, sampling	0	1	4
- we do not have enough info - need more info to determine shape			

- the total of all sampling distributions will be skewed but not the single sampling distribution			
NO RESPONSE	7	3	0
NO TOPIC	1	1	1
1 Inference about statistics 2 The older employees are much less likely to be chosen 3 Because again you could select majority of all younger workers			

TOPIC SHAPE -SENSOR	a	b	c
1 - Right, skew, mean, population	37	2	4
- if population is right skewed then the SAMPLING DISTRIBUTION is also right skewed because the SAMPLING DISTRIBUTION is from the population - Population skew right therefore SAMPLING DISTRIBUTION of mean also right skew - If employee ages are right skewed then the graph will be too			
2- large, central limit theorem, normal, sample size, approximately, distribution	4	22	1
-Normal because the sample size is 100 people -Because of the CLT, xbar is approximately normal when the sample size is large -Central Limit Theorem, will be normal			
3- median, know, outlier, mean, data, determine	7	2	28
-Do not know of outliers so cannot predict shape - because we're not given all the data points, just the mean, so we cannot predict outliers or other things - reason we know this is that the company could have a lot of outliers on that end causing the mean to be 34			
4- normal bell, bell, shape, normal, curve, sampling distribution, shape	6	26	4
- because of CLT, with n=100 we'll have a normal distribution - can only assume the shape of Sampling Distribution but because of CLT can assume its bell shaped - normal curve			
5- random, population, distribution, population distribution, sample, select	18	2	8
- With a random sample of this size it should follow the trend of the whole company.			

- cannot tell because randomly selecting employees, we won't know the ages of the people being picked - The mean will be the same as the parent population since it was randomly selected			
6- average, look, standard deviation, sampling distribution, know	9	9	7
- because we do not know what the average age is in the sample - SAMPLING DISTRIBUTION of average ages of all employees and the averages will be symmetrically distributed - because Sampling Distribution may end up looking different than the population, we'd expect it to look normal but it could be skewed			
7- sample, large, symmetric, enough, large sample	6	13	7
- the sample is large enough to resemble the sampling distribution of the parent population to some extent - if we have a large enough sample it will be normally distributed - Sampling Distribution allows us to predict the variability of a sample statistic, which tells us about its shape			
8 - symmetric, distribution, mean, always, sample	4	15	2
- do not know which values are picked, so could be skewed, or symmetric - Distribution of the means should be fairly symmetric - we know this because it is an SIMPLE RANDOM SAMPLE which has a symmetric distribution			
9- shape, sample distribution, population, mean, sampling	16	6	4
- Shape of population - since n-large, the Sampling Distribution will be similar to population shape - Sampling Distribution of mean follows same shape as population			
10- sample, enough, data, different, distribution	5	3	10

- Chances are the sample will mimic the population - not enough info - we do not have enough information on how old the different employees are or how they may be skewed			
NO RESPONSE	12	5	6
NO TOPIC	0	0	9
1 you cannot tell because there may be younger or older working at the company 2 no because we have yet to calculate the results 3 not enough information 4 Needs to be a set range of ages 5 cannot tell until we do the sampling 6 we need to work it out 7 we're not told how many employee ages were taken 8 cannot tell yet because we haven't surveyed anyone yet to get the answer 9 We do not have a z-variable. We have an x and no y			

Appendix N – Content Categorization Rules

The following are the same for all four questions:

More Information:

```
(OR,

(SENT, "DO NOT", "KNOW"),

(SENT, "MORE", "INFORMATION"),

(SENT, "NEED", "DATA"),

(SENT, "not enough")

)
```

NO RESPONSE:

```
(OR, "NO RESPONSE", "BLANK")
```

Likelihood

Correct:

```
(AND,

(OR,

(DIST_5, "small", "exceed"),

(SENT, (DIST_4, (OR,"small","smaller"), "MORE", (OR, "VARY", "variable",
"outlier", "outliers"))),

(SENT, (DIST_4, (OR,"small","smaller"), "NOT", "REPRESENT")),

(AND, (OR,"small","smaller"), "sample", (NOT, "large")),

(SENT, (OR, "CLT", "Central Limit Theorem")),
```

```
(SENT, "less subjects", "less likely"),

(SENT, "LARGE", "actual"),

(SENT, "LARGE", "LESS VARAIBLE"),

(SENT, (DIST_4, "small", (OR,"skew","skewed","skewness","skews"))),

(SENT, "SMALL", (OR, (OR,"outlier","outliers"), "EXTREME"), "Affect"),

(SENT, "more likely", "Small"),

(DIST_7, (OR, "small sample", "less", "small"), (OR,"skew","skewed")),

(SENT, (DIST_5, "large sample", "normal")),

(SENT, "Less", "sample", "skewed"),

(SENT, "larger", "closer", "to mu"),

(AND, "small", (NOT, (OR, "large", "200")))

),

(NOT ,

(OR, "NO RESPONSE", "BLANK" )

))
```

INCORRECT:

```
(AND,

(OR,

(SENT, (DIST_4, "LARGE", (OR, "INCREASE", "INCREASES", "INCREASING",
"MORE"),
       (OR, "VARY", "outlier", "variable", "outliers") ) ),

(SENT, "MEAN", "INCREASE@" ),

(DIST_3, "NEED", "LARGE"),

(DIST_6, "LARGE", "SAMPLE", "LIKELY"),

(SENT, "likely", (OR, "large", "larger")),

(DIST_8, "NEED", "200"),

(SENT, "variability", "increase", "larger sample")
```

```
),

(NOT,

(OR,

(OR, "NO RESPONSE", "BLANK" ),

(DIST_5, "small", "exceed"),

(SENT, (DIST_4, (OR,"small","smaller"), "MORE", (OR, "VARY", "variable",
"outlier", "outliers"))),

(SENT, (DIST_4, (OR,"small","smaller"), "NOT", "REPRESENT")),

(AND, (OR,"small","smaller"), "sample", (NOT, "large")),

(SENT, (OR, "CLT", "Central Limit Theorem")),

(SENT, "less subjects", "less likely"),

(SENT, "LARGE", "actual"),

(SENT, "LARGE", "LESS VARAIBLE"),

(SENT, (DIST_4, "small", (OR,"skew","skewed","skewness","skews"))),

(SENT, "SMALL", (OR, (OR,"outlier","outliers"), "EXTREME"), "Affect"),

(SENT, "more likely", "Small"),

(DIST_7, (OR, "small sample", "less", "small"), (OR,"skew","skewed")),

(SENT, (DIST_5, "large sample", "normal")),

(SENT, "Less", "sample", "skewed"),

(SENT, "larger", "closer", "to mu"),

(AND, "small", (NOT, (OR, "large", "200")))

)))
```

Center

CORRECT:

```
(AND,

(OR,

(SENT, ( AND,
(OR,"Center","Centers","center","centered","centering","centers"),
       (OR, "mean", "Average"), (NOT, (OR, "cannot", "NOT")))),

(SENT, (AND, "normal", (NOT, (OR, "Do NOT KNOW", "NOT", "cannot")))),

(SENT, (AND, "MEAN", (NOT, (OR, "NOT", "cannot", "different",
"outlier")))),

(SENT, (AND,(OR,"large","larger", "bigger", "big"),
       (OR,"sampling","sample","sampled","samples"),
       (OR,"represent","represented","representing","represents"))),

(SENT, (OR, "CLT", "Central Limit Theorem")),

(SENT, (AND, "NORMAL", "MU", (NOT, (OR, "NOT", "CANNOT")))),

(SENT, (AND, (OR, "centered", "center"), (OR, "mu", "34", "mean", "xbar"),
(NOT, (OR, "NOT", "CANNOT")))),

(SENT, (AND, "average", "34", (NOT, (OR, "NOT", "CANNOT")))),

(AND, "34", (NOT, (OR, "NOT", "CANNOT")))

),

(NOT,
(OR,

(AND, (OR,"Center","Centers","center","centered","centering","centers",
"mean", "average"),
      (OR, "NOT", "cannot")),

(SENT, "MEDIAN"),

(AND, (OR,"small","smaller"),
(OR,"sampling","sample","sampled","samples")),

(SENT, (OR,"outlier","outliers", "SKEW", "SKEWED")),

(AND, "NOT", "KNOW"),

(AND, "NOT", "MU"),

(AND, "random", (OR,"varied","varies","vary","varying", "variation")),
```

```
(AND, "RANDOM SAMPLE"),

(AND, (SENT, "random", "sample", (OR, "cannot", "NOT"))),

(AND, "random", "sample", (NOT, "sampling distribution")),

(OR, "changing","change","changed","changes"),

(AND, (DIST_4, (OR, "NO", "DO NOT"), (OR, "FACTS", "KNOW",
"Information"))),

(AND, "different", (OR, "average", "center", "mean")),

(AND, "not", "Enough", (OR,"info","information")),

(SENT, "NOT", (OR,"represent","represented","representing","represents")),

(SENT, "Data", "not", "collected"),

(AND, "need", "more"))
))
```

INCORRECT:

```
(OR,

(AND, (OR,"Center","Centers","center","centered","centering","centers",
"mean", "average"),
      (OR, "NOT", "cannot")),

(SENT, "MEDIAN"),

(AND, (OR,"small","smaller"),
(OR,"sampling","sample","sampled","samples")),

(SENT, (OR,"outlier","outliers", "SKEW", "SKEWED")),

(AND, "NOT", "KNOW"),

(AND, "NOT", "MU"),

(AND, "random", (OR,"varied","varies","vary","varying", "variation")),

(AND, "RANDOM SAMPLE"),

(AND, (SENT, "random", "sample", (OR, "cannot", "NOT"))),

(AND, "random", "sample", (NOT, "sampling distribution")),

(OR, "changing","change","changed","changes"),

(AND, "different", (OR, "average", "center", "mean")),
```

```
)


SPREAD

CORRECT:

(AND,

(OR,

(SENT, (DIST_4, "larger", (OR, "size", "n", "number")), (DIST_4,
"smaller", (OR, "spread", "variance"))),

(OR, "Central limit theorem", "CLT"),

(SENT, "CORRECT formula"),

(SENT, "formula"),

(AND, (DIST_3, "repeated", "samples"), (DIST_3, "smaller", "std")),

(SENT, "Sampling distribution", "always", "Small"),

(AND, (OR, "small", "smaller"), (OR, "standard deviation", "std"), (NOT,
"larger")),


(AND, (OR, "smaller", "decrease"), (NOT, "larger"))

),

(NOT,


 (OR,

(SENT, (DIST_6, "smaller", (OR, "size", "n", "number"), "population"),
       (DIST_4, "population", (OR, "spread", "variance", "variation",
"variability", "range")) ),

(SENT, "INCORRECT formula"),

(AND, (OR, "bigger", "larger", "higher", "greater"),  (OR, "standard
deviation", "std")),

(AND, (DIST_4, (OR, "large", "larger"), "N"),
      (DIST_4, (OR, "smaller", "small"), (OR, " std", "Standard
Deviation") )
  ),
```

```
(AND, (DIST_4, (OR, "smaller", "small"), "sample"),
      (DIST_4, (OR, "larger", "large", "bigger"), (OR, "std", "Standard
Deviation"))
  ),

(SENT, "Sampling distribution", "always", (OR, "Small", "smaller")),

(AND, "Wider", "Sampling Distribution"),

(OR, "do not vary", "does not vary"),

(AND, (OR, "NO", "NOT"), "change"),

(SENT, (DIST_2, (OR, "sampling", "sample"), (OR,"distribution", "dist")),
       (DIST_3, (OR, "larger", "large"), (OR, "std", "standard
deviation"))
      ),

(AND, "same", (NOT, (OR, "larger", "large", "small", "smaller"))),

(SENT, (OR, "not", "no"), "change")


)))
```

INCORRECT_SPREAD:

```
(AND,

(OR,

(SENT, (DIST_6, "smaller", (OR, "size", "n", "number"), "population"),
       (DIST_4, "population", (OR, "spread", "variance", "variation",
"variability", "range")) ),

(SENT, "INCORRECT formula"),

(AND, (OR, "bigger", "larger", "higher", "greater"),  (OR, "standard
deviation", "std")),

(AND, (DIST_4, (OR, "large", "larger"), "N"),
      (DIST_4, (OR, "smaller", "small"), (OR, " std", "Standard
Deviation") )
  ),

(AND, (DIST_4, (OR, "smaller", "small"), "sample"),
```

```
      (DIST_4, (OR, "larger", "large", "bigger"), (OR, "std", "Standard
Deviation"))
  ),

(SENT, "Sampling distribution", "always", (OR, "Small", "smaller")),

(AND, "Wider", "Sampling Distribution"),


(SENT, (DIST_2, (OR, "sampling", "sample"), (OR,"distribution", "dist")),
       (DIST_3, (OR, "larger", "large"), (OR, "std", "standard
deviation"))
      ),

(AND, "same", (NOT, (OR, "larger", "large", "small", "smaller")))

),

(NOT,

(OR,

(OR, "do not vary", "does not vary"),

(AND, (OR, "NO", "NOT"), "change"),

(SENT, "same"),

(SENT, (OR, "not", "no"), "change")

)))
```

SAME_SPREAD:

```
(OR,

(OR, "do not vary", "does not vary"),

(AND, (OR, "NO", "NOT"), "change"),

(AND, "same", (NOT, (OR, "ages", "age"))),

(SENT, (OR, "not", "no"), "change")

)
```

SHAPE:

Correct_shape:

```
(AND,

(OR,
(AND, (OR, "BELL", "Bell-shaped", "symmetric", "symmetrically",
"symmetric/normal","normal","normally"),
      (NOT, (OR, "NOT", "cannot"))),
(OR, "Central limit theorem", "CLT"),
(SENT, "Sampling distribution", "always", (OR, "NORMAL", "SYMMETRIC")),
(SENT, "LAW OF", "LARGE NUMBERS"),
(AND, (OR, "PREDICTABLE", "predict"), (OR, "PATTERN", "statistic"))

),

(NOT,


(OR,
(SENT, "PARENT"),
(SENT, "POPULATION"),
(AND, "NOT", "normal"),
(SENT, (OR,"skew","skewed","skewing","skews")),
(SENT, "MEDIAN"),
(AND, (OR, "random", "randomly"), (OR, "different", "differ")),
(AND, (OR, (OR,"outlier","outliers"),
(OR,"skew","skewed","skewing","skews"))),

(SENT, "WHOLE"),
(AND, (OR, "like", "need", "same", "resemble"), (OR, "PARENT",
"POPULATION")),
(DIST_5, (AND, (OR, "POP", "POPULATION", "actual distribution"), (NOT,
"NOT"))),
(AND, "shape", "SIMILAR", (NOT, (OR, "NORMAL", "SYMMETRIC"))),


(DIST_4, "NEED",  "MORE", "INFORMATION"),
(SENT, "NEED"),
(SENT, "DO NOT", "KNOW"),
(SENT, "MORE", "INFORMATION"),
(SENT, "NEED", "DATA"),
(SENT, "NOT", "ENOUGH",  (OR, "INFO", "information")),
(AND, "CANNOT", (OR, "TELL", "PREDICT")),
(SENT, "NOT", (OR,"exact","exactly")),
```

```
(SENT, "NOT", "Certain")

)))
```

INCORRECT_SHAPE

```
(AND,

(OR,

(AND, "NOT", "normal"),
(SENT, (OR,"skew","skewed","skewing","skews")),

(SENT, "MEDIAN"),

(AND, (OR, "random", "randomly"), (OR, "different", "differ")),

(AND, (OR, (OR,"outlier","outliers"),
(OR,"skew","skewed","skewing","skews")))

  ),

(NOT,

(OR,
(SENT, "PARENT"),
(SENT, "Population"),
(SENT, "WHOLE"),
(AND, (OR, "like", "need", "same"), (OR, "PARENT", "POPULATION")),
(DIST_5, (AND, (OR, "POP", "POPULATION", "actual distribution"), (NOT,
"NOT"))),
(AND, "shape", "SIMILAR", (NOT, (OR, "NORMAL", "SYMMETRIC"))),
(SENT, "SYMMETRIC"),

(DIST_4, "NEED",  "MORE", "INFORMATION"),
(SENT, "NEED"),
(SENT, "DO NOT", "KNOW"),
(SENT, "MORE", "INFORMATION"),
(SENT, "NEED", "DATA"),
(SENT, "NOT", "ENOUGH",  (OR, "INFO", "information")),
(AND, "CANNOT", (OR, "TELL", "PREDICT")),
(SENT, "NOT", (OR,"exact","exactly")),
(SENT, "NOT", "Certain")


)))
```

PARENT_SHAPE:

```
(AND,

(OR,

(SENT, "WHOLE"),

(AND, (OR, "like", "need", "same", "resemble"), (OR, "PARENT",
"POPULATION")),

(DIST_5, (AND, (OR, "POP", "POPULATION", "actual distribution"), (NOT,
"NOT"))),

(AND, "shape", "SIMILAR", (NOT, (OR, "NORMAL", "SYMMETRIC")))

),

(NOT,

   (OR,

(DIST_4, "NEED",  "MORE", "INFORMATION"),
(SENT, "NEED"),
(SENT, "DO NOT", "KNOW"),
(SENT, "MORE", "INFORMATION"),
(SENT, "NEED", "DATA"),
(SENT, "NOT", "ENOUGH",  (OR, "INFO", "information")),
(AND, "CANNOT", (OR, "TELL", "PREDICT")),
(SENT, "NOT", (OR,"exact","exactly")),
(SENT, "NOT", "Certain")

    )
        )
            )
```

Appendix O – Content Categorization, Center Uncategorized Values

RESPONSE	Learning Style	NEW CATEGORY
chose appropriate sample size thus they will be equal	Intuitor	Correct
Because you could randomly select many younger workers and the mean of the sample will be much lower than that of the parameter	Intuitor	Incorrect
because we cannot figure that out with the information we're given	Intuitor	Incorrect
the sample size is large thus the sample distribution would be centered.	Intuitor	Incorrect
because it is a large enough sample	Intuitor	Incorrect
the distribution may have a mean of 34 but it could also be less than that	Intuitor	Incorrect
The sampling distribution will be the same as the population	Intuitor	Correct
although it says it should be centered around 34, you might have gotten a bunch of people younger or older	Intuitor	Incorrect
since STANDARD DEVIATION is so high that indicates there'll be a broad spectrum of ages	Sensor	Incorrect
Large samples	Sensor	Correct
because these 100 employees since chosen randomly may be slightly different from total population of employees but close	Sensor	Incorrect
if parent and sample of 100 will have same shape	Sensor	Incorrect
because the sample could consist of the older proportion of workers or vice versa	Sensor	Incorrect
multinational so the sample will select people from any area and thus not be concise	Sensor	Incorrect

that is the parameter of the data	Sensor	Correct
Sample size is large enough and randomly chose	Sensor	Incorrect
because statistics have a predictable pattern and inferences can be made on just a few datasets or information	Sensor	Correct
More likely bell shaped around mean	Sensor	Correct
the n is large	Sensor	Incorrect
all of the information leads to this	Sensor	Incorrect
since a & b are the same I choose c	Sensor	Incorrect

Appendix P - Task Based Interviews

Based on the results from the preliminary round of focus groups, I decided that it was better to switch to task based interviews to eliminate the interaction between students since we are unable to assume that all sensors or intuitors understand the material at the same level based on just the learning style.

I conducted one on one interviews with both sensors and intuitors. The format of the interview is detailed in the previous chapter.

Three sensors and two intuitors participated in these interviews for the Spring 2007 semester. In answering the first question, (see Appendix G) without any aid, recall, or assistance, two of the three sensors selected the correct solution to this problem as well as one of the two intuitors. One sensor stated that she chose “C” because everything in class is normal, so speculated that the solution would be normal. The theme of normality clearly stood out in this student’s mind, but did not carry with it any context or deeper meaning, as she was unable to elaborate any further. She will be referred to as Sensor_One. The other sensor (Sensor_Two) that chose the correct solution stated that she had no basis and it was a complete guess. The sensor that selected an incorrect choice (Sensor_Three), selected the distribution that looked most like the parent population because that is what she expected from a distribution of sample means.

The intuitor that selected the correct response stated that “I don’t really understand quite what is going on, and I seem to remember something about if you get a large enough sample size and then it is normal, and that (pointing to his selection) is normal.” This shows that he has a slight understanding of the Central Limit Theorem, but from this information, I was unable to gather how deep his understanding was. I will refer to him as Intuitor_One. The other intuitor (Intuitor_Two) chose the distribution that looked most like the parent and stated that his selection was a guess.

All five students watched the applet demonstration and then were asked to describe what they saw. Intuitor_One states that the more sample means you take; it is going to the real population average. Intuitor_Two states that the means would keep stacking up and that all the means will fall into the same area and form a symmetric shape.

Sensor_One expects the means to change because we are taking a different sample each time, but not much because it will bulk in the middle. Sensor_Two expects the pattern (of the means dropping down on top of one another) to continue, and Sensor_Three notes all the means are grouped together and form a normal distribution.

Overall, I believe that the students are getting an idea of “shape” with sampling distributions, but are not paying attention to other key details such as center or spread.

Progressing to the remainder of the interview, Intuitor_One was able to correctly answer each question. He worked quickly through the problems and did not use the applet to solve the problems, but instead, used it to verify the solutions he chose. He paid attention to both the shape and the center of the answer choices. Several solutions were normal distributions and could have easily been selected, but one was scaled outside the range of the population mean, and thus was an infeasible solution.

Intuitor_Two was able to take the applet and correctly solve the remaining problems. He chose the distributions with symmetric shapes and used the numbers to verify his selection. When asked about the class, Intuitor_Two stated that the wording was very confusing and he prefers to work through problems rather than attend a lecture where all the information is presented on the chalk board.

For question 2, Sensor_One chose a left skewed distribution to match the parent population. She thought it was a good representation of the entire population. For question 3, she chose the correct answer and paid attention to the numbers at the bottom of the graph. For question 4, she chose a highly varied normal distribution, and for the final question, she chose a right skewed distribution that was tending towards normality because she felt like the bulk on the left side best captured the solution, but wanted the tail to be represented as well. Even with the applet demonstration, this student was not making the proper connections. When asked how she felt about the course, she stated that she prefers to see a similar problem worked out so that she can just plug in the numbers. She also stated that she has trouble relating symbols to formulas, for example, she asked, What does xbar mean, and how do I get it?

Sensor_Two chose a solution with the correct shape, but did not notice that the center was incorrect. She does the same thing for question three, she chooses the normally distributed graph, but ignores the detail. She correctly answers the remaining two questions and based her solution on their shape. She comments that each question has two or three solutions that are normally distributed which is making it difficult for her to decide. When asked about the class she states that she prefers to plug and chug and hands on activities.

Sensor_Three chose the correct answer for problem two basing it upon the symmetric shape. For question three, she chose a normal distribution, but ignored the number detail, and thus chose the solution that was centered on a lower mean. For the fourth problem, she chose between the two normal distributions, but ultimately chose the incorrect solution, again not taking heed to the detail. She did the same thing for the final question, selecting the normal distribution, but ignoring the mean given at the bottom of the graph. When asked about the course, she states that she likes the examples, but dislikes hands on activities (waste of time) and dislikes being quizzed on new material.

It was surprising that the two intuitors paid more attention to the detail of the graphs than the sensors. Even though sensors are more detail oriented, perhaps the thought process leading them to look at the numbers is more intuitive. Since the sensor was not told to look at the details of the graph, perhaps they did not think that it was an important feature, whereas the intuitor instinctively knew that this information was just as significant as the shape of the graph to the solution.

In the fall 2007 semester, 3 sensors and 5 intuitors were interviewed. Before the students were asked to answer questions about sampling distributions (Appendix G), I first gave them a handout to complete (Appendix E), asking questions to try to see what their background knowledge was regarding sampling, variability, normal distribution, and distribution. The full detail of the results to the first handout is presented in Appendices I and J.

The first question asked the student to state what a variable was. The sensors stated that it was something that changes, something you are testing, and an item that can or cannot be changed based on the observation, but affects outcome. The intuitors stated that a variable was a number, a quantity, a measure, a value that will vary and an unknown number. These responses between the groups differ in that the intuitors are thinking of a variable in terms of a number or quantity, whereas the sensor is thinking of it in more general and vague terms, an item that may or may not change.

When asked what it meant to say that observations vary, the sensors stated that it meant a point of view or the data can change, or the outcomes will be different. The intuitors stated that the observations will change, not always be the same, there are different possible outcomes, changes due to something (as in a dependent variable), and that the framework by which something is viewed could cause variation (not really answering the question, but more-so stating why variation occurs). Here, the sensors are talking about in more specific terms, and the intuitors are being more general.

All students were able to select which normal distribution had more variability and attributed their response to reasoning that it had a wider spread or range. None of the students were really able to state that a population distribution consisted of everything in the population, or all possibilities, or anything of the sort. One sensor stated that it consisted of amounts of people, so counts. The other sensor stated that it consisted of differing pieces of data from a population. The third did not answer this question. Two intuitors were not sure how to answer this question, one stated that a population distribution consisted of a mean and a standard deviation, and the fourth stated that a population distribution represents facts or opinions, which I sort of believe she was confusing this with a sample because she also stated that not all possible outcomes would occur because people are more likely to say one thing over another. The fifth intuitor wrote that "with regards to populations, distributions concern the actual 'count' of a specific parameter. Based on sampling, we aim to use statistics to estimate actual distributions of some characteristic (the parameter of interest)".

All students could identify an observation on a population histogram. They were then asked if they could identify the same observation on the distribution of average team weights and to explain why or why not. All five intuitors stated that they could not identify the value, where four of them stated they needed the team average, and that this single point could not be represented by averages. The fifth intuitor attempted to find the observation, and upon realizing the scale did not capture it, stated that his weight was not on the graph. (The weight is 280, and the averages ranged from 210 to 240).

All of the sensors immediately tried to find the point 280 on the graph, and ignored the fact that the histogram was of the team averages. Two stated that they could not find it because it was not on the graph; the other sensor stated that the average of the team was 240 which were not represented on the graph. This sensor took the first histogram (of all the players) and used that average in their reasoning.

This exercise made it clear that sensors and intuitors are thinking differently about sampling distributions. The intuitors did not have a difficult time thinking about the histogram of averages and realizing that a single observation could not be found. The sensors, however, had complications thinking about a single observation as part of a whole, the whole being the team, and the single observation being the individual player. They had a difficult time reasoning why a single observation could not be depicted on a chart of averages, and even with some explanation, still appeared to be confused.

The students were then asked if they could calculate the standard deviation of the sampling distribution if the standard deviation of the population was 12. (There are 16 players in each team). One of the sensors immediately used the formula and was able to calculate the value. The other two sensors had no idea, but knew that there was a formula to get to it. One of these sensors stated that it would be smaller than the population standard deviation, but had no recollection of what the formula was.

Two of the intuitors knew the formula and immediately calculated the standard deviation. One of the intuitors did not have this question because it was added after the first interview. One intuitor thought that it would be the same as the parent population, and the other eyeballed the sampling distribution and stated it would be five, based purely on the graph.

The next question presented in this pre-assessment asked the students to identify (from a word problem) if a value given was a sample statistic or a population parameter, and then state if the sample statistic given was mu or xbar. Both groups were able to identify these terms from the problem.

Finally the students were asked to draw a normal distribution, and then below it, draw one with a large variance and one with a smaller variance, and then draw what the graph would look like if the mean increased by 5. All of the intuitors drew a tall and narrow normal curve to represent small variance and a fat and wide normal curve to represent larger variance. They also all correctly drew the shifted curve (to represent an increase in mean). Two of the sensors were able to do the same thing, but the third sensor drew a taller curve to represent the mean increase by five.

Overall, the sensors and intuitors had some similar conceptions of these ideas, but several of the thoughts differed. An idea of sampling distribution of means was not as easily grasped by the sensors.

The overall understanding of these key concepts was above average, but not exceptional. There are definitely gaps in the thought process and understanding of components leading up to sampling distributions, and the assessment of the sampling distribution questions may lead to our further understanding of these differences. Despite observing differences between sensors and intuitors, I was unable to confidently answer either of the research questions involving the student's ability to answer or reason about sampling distributions, and thus discontinued conducting interviews in subsequent semesters. However I did find the interviews to play an integral role in developing the final exam questions.