Abstract

WEBER, VICTORIA LYNN. Defining the Relationship Between Learning Statistical Computing and a Student's Ability to Think and Reason Statistically. (Under the direction of Dr. Hollylynne Lee).

The purpose of this study was to begin examining the relationship that exists between a student's use of statistical computing tools and the ways in which students think statistically. To explore this relationship, a qualitative research study was designed to collect multiple instances of statistics problem solving with statistical computing tools in both written form and through a one-on-one problem solving session with the researcher. Fourteen students in a second semester statistics course completed a total of 91 written assignments across a semester. Additionally, a subset of 5 volunteer students also completed task-based interviews. These assessments required the students to use statistical computing software and their core knowledge of statistics to give complete results. A qualitative analysis of the students' work was conducted and an initial definition of the relationship between statistical computing and thinking was formed. The results of the study are organized in three manuscript Chapters.

The purpose of the first manuscript is to show how written work can be used to assess statistical thinking. The paper discusses some popular forms of written assessment used by instructors of statistics and other quantitative fields. An argumentation structure, similar to a model defined by Toulmin (1958) is defined. More detail is then given on three of the writing assignments utilized in the course, their intended assessment purpose and the types of answers teachers may expect to see if they were to give these assessments to their students.

The purpose of the second manuscript is to begin defining the relationship between statistical computing and statistical thinking. Through a qualitative study of the written assessments, five relational patterns between statistical computing and statistical thinking are

identified and illustrated with examples from students work. Implications for teaching and research due to these patterns follow.

The purpose of the third manuscript is to provide an in-depth analysis of how statistical computing is used by students to solve statistical problems. Since written work does not always give the best picture of the problem solving process, task-based interviews from 5 students were conducted and analyzed to provide a more in depth view of how students use statistical computing tools to solve statistical problems. Mimicking a problem solving analysis structure originally created by Lee and Hollebrands, the ways in which students use statistical computing while solving problems are described, and ways in which instructors may be able to help students solve problems using statistical computing software are given.

Results from the study indicate that there is a relationship between statistical thinking and statistical computing. When students are able to use statistical computing and possess good problem solving habits, they tend to produce logical arguments that show the strength of their statistical thinking sophistication. However, when students do not possess the capability to utilize the statistical computing technologies, statistical problem solving becomes very difficult or nearly impossible for some students.

Defining the Relationship Between Learning Statistical Computing and a Student's Ability to Think and Reason Statistically

by
Victoria Lynn Weber

A dissertation submitted to the Graduate Faculty of
North Carolina State University
In partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Mathematics Education

Raleigh, North Carolina
2017

APPROVED BY:

Dr. Karen Keene	Dr. Herle McGowan
Dr. Hollylynne Lee Chair of Advisory Committee	Dr. Justin Post

Biography

Victoria Weber was born in Cleveland, Ohio on February 6, 1984. She is the daughter of Harold and Lynn Weber and sister to Matthew Weber. After graduating from Berea High School in 2002, she attended Ohio University where she received a bachelor's in actuarial science (2006) and a masters of mathematics with an emphasis in education (2007). After graduation she moved south to teach mathematics at Surry Community College in Dobson, North Carolina for three years. During her time at Surry she developed her love of teaching and her love of statistics, which is what prompted her to continue her education.

In 2010 she moved to Raleigh, North Carolina to earn a master's in statistics from North Carolina State University. After graduation she was offered a position at Meredith College, an all-women's college, in Raleigh, NC. She has been teaching statistics and mathematics there for more than five years.

While teaching at Meredith, she decided to go back to school to earn her PhD in mathematics and statistics education. In the spring of 2014 she started in the mathematics education program at North Carolina State University under the direction of Dr. Hollylynne Lee.

After graduation, Victoria plans to start a new position teaching statistics at the University of Notre Dame in South Bend, Indiana.

Acknowledgements

First and foremost, I would like to thank my advisor, Hollylynne. Without her "many smiles" I probably never would have come this far. She spent many hours discussing the path this document should take and making edits to drafts. She has been very patient with me and has been an amazing mentor. I cannot imagine having completed my PhD with anyone else at the helm.

I would also like to thank my committee members, Karen, Herle and Justin for always being so flexible. Trying to make the schedules of 5 very busy academic professionals mesh is no easy task. Even finding an hour or two in a month where we were all free sometimes seemed impossible. But somehow we made it work.

Next, I would like to thank my family members. To my loving fiancé Roger, for being so very patient with me as I finished my dissertation and always offering me help when you knew the stress was getting me. To my parents, Harold and Lynn, who always told me I needed to get an education before getting married, you probably never thought I would take it this seriously. To my brother, Matthew, for teaching me how to write a term paper several years ago. And to my grandparents, Jack and Julia, for granting your generous financial support that helped to make my dreams come true.

I also need to send some love to my friends that helped me throughout all of this. To each of my statistics grad student friends, I send you thanks for encouraging me as I earned my master's degree. In particular I want to thank Sam for being the best roommate ever and never letting me get down on myself for not succeeding. To my mathematics education friends, thank you for great conversations and never letting me check Facebook during writing group. And to Kristen, who helped me to learn ggplot to create some awesome graphs for this report.

Finally, I would like to give thanks to God. One morning, in the fall of 2016 after a particularly hard day at school, I had thought of quitting. Despite all my hard work, I just didn't feel like I could do it anymore. I prayed to God that He would show me a sign to guide me in that decision. I went to church, I listened to the sermon and I just didn't feel like I had gotten the answer to my question. But on the way out of the parking lot, on that beautiful fall day, I fell in line behind a vehicle whose license plate said "NVRQUIT". God answered that prayer, and I am very thankful that He did.

Table of Contents

List of Tables

List of Figures

Chapter 1 : Introduction

Statistics Technology: Where we are

With the creation of products such as Fathom (Finzer, 2000), TinkerPlots (Konold, 2007) and StatCrunch (West, Wu, & Heydt, 2004), the integration of technology into statistics classes has become widespread. These "learning tools" (Moore, 1997), are useful for quick calculations, but more importantly, their dynamic nature and ease of use make it so that students can spend more time exploring, visualizing and interacting with data to help them learn statistical concepts (Moore, 1997). This has recently become a high priority in statistics classrooms trying to adhere to the Guidelines for Assessment and Instruction in Statistics Education (GAISE) college standards (American Statistical Association, 2005; American Statistical Association, 2016).

Several researchers have found evidence that using these learning tools are helpful in developing students' statistical thinking and reasoning skills, especially in introductory statistics courses. Several studies (delMas, Garfield, & Zieffler, 2014; Garfield, delMas, & Zieffler, 2012; Watson, 2014) have looked at students' ability to understand inference when using a program like TinkerPlots to conduct a simulation. For example, Garfield, delMas and Zieffler (2012) found that students in an introductory statistics course using TinkerPlots for randomization testing "developed good statistical thinking about statistical inference" (p. 896). Chance et al. (2007) also claim that these technologies are good for automating calculations, emphasizing data exploration and visualizing abstract concepts, which are all elements that could improve student learning.

Students majoring in STEM fields or minoring in statistics may be required to take statistics courses *beyond* the introductory level. For example, Brieger and Hardin (2012) discuss the need for more sophisticated statistical knowledge in students studying in the health sciences to improve on the quality of analysis currently being conducted in clinical trials. Additionally, Bisgaard (1991) argues that engineering students need a course in statistics that focuses on probability and experimental design, with engineering specific examples presented.

One of the overarching goals of these courses is to prepare students who need to conduct more advanced statistical analysis in their field, as opposed to learning statistics mainly to improve their statistical literacy. Since the goals of the course are different from regular introductory statistics courses, the technology that is used in them may also need to be different. Programs that are often used in professional settings, which Moore has dubbed technology that “does statistics” (Moore, 1997) may be more helpful in preparing students for their future careers. These tools, such as R and SAS, tend to be more powerful and more flexible for large scale statistical analysis, though less user friendly than the tools used for “learning statistics”.

Especially with the introduction of the field of Data Science – a field that combines computer science and statistics - one such course that has become more prevalent in statistics curricula is statistical computing (Baumer, 2015; Chance, et al., 2014; Hardin, et al., 2015). This is a course that encompasses the breadth of topics that merge statistics and computer science (Monahan, 2004). Users must be comfortable using a wide array of software tools - and perhaps hardware – while understanding programming and logic to fully complete the spectrum of statistical tasks they need to perform.

While there has been a push to implement such courses in undergraduate curricula (Gentle, 2004; Hardin, et al., 2015; Nolan & Temple Lang, 2010) the research on how students learn using programs that are designed for doing statistics has not kept up with research on how students learn using programs that are designed for learning statistics. Specifically, little research has been done to document what affordances and constraints statistical computing technologies might have on a student’s ability to think and reason statistically.

Statistics Technology: Where we are going

The expectation for statistics majors and minors to learn computing tools in their curriculum has recently grown (Chance, et al., 2014; Horton, Baumer, & Wickham, 2014; Nolan & Temple Lang, 2010). Within statistics courses, students are learning data cleaning, graphical displays and multiple techniques to aid them in their analyses (Hardin, et al., 2015). It is not enough to expect students to teach themselves the majority of the technology while the instructor teaches only the statistical concepts. If students are forced to learn the technology mainly on their own, they will focus on getting the answers to the problems they have been assigned and often pick up bad programming habits due to the lack of instruction and direction (Nolan & Temple Lang, 2010). Conversely, focusing too much on how one uses the technology to conduct statistical tests may teach the students that it is not important to explore the data before conducting formal inference procedures. Thus, it is important for teachers to find the appropriate balance between teaching the statistics and teaching the technology that is used to conduct the analysis.

Despite the cautions of the ASA about including programs like R and SAS in introductory statistics courses (American Statistical Association, 2016), some students can successfully learn these higher level languages while still grasping the concepts traditionally taught in introductory statistics courses. Horton and his colleagues were able to successfully introduce the statistical language R into their introductory statistics courses. Using RStudio and the MOSAIC platform, their students were able to perform higher level statistical processes with a more difficult technology earlier in their statistical careers. They do note however, that in teaching the statistical computing ideas, there should be more examples and activities, but less content taught overall (Horton, Baumer, & Wickham, 2014; Kaplan, Horton, & Pruim, 2015).

Additionally, teaching statistical computing topics can been seen as a chicken and the egg dilemma. Using technology to explore data and analyze data is one of the six recommendations of the GAISE college standards for introductory statistics courses (American Statistical Association, 2005; American Statistical Association, 2016). However, the same reports also caution about introducing higher level technologies - those normally seen in statistical computing courses - too early. While these tools are powerful, they may confuse students and may be more of a hindrance than help. This raises the question of when and how statistical computing should be taught when students need to learn the technology. An answer to these questions may be found if we can determine if and how a students' use of statistical computing technology is related to how students think and reason statistically.

Thus, this research aim is to develop a better understanding of the ways in which statistical computing affects a student's ability to think and reason statistically, and how a student's ability to think and reason statistically may affect their use of statistical computing to solve statistical problems.

Research Questions

Through this research, the aim is to determine if there is a relationship between a student's statistical thinking and statistical computing abilities and being to develop a description of the relationship that exists between statistical thinking and statistical computing. The main research questions can be summarized as:

- What are the ways in which statistical thinking and statistical computing may be related to each other, including the ways in which the two items help and hinder students' learning?

- In what ways are statistical computing techniques used by students in a problem solving setting and how do differences in statistical computing sophistication affect students' abilities to solve statistical problems with statistical computing technologies?

Organization of the Document

In Chapter 2, the reader will find a literature review where statistical thinking and computing are defined and a framework for identification is made. In Chapter 3, a description of the study design and methodology is given. Chapters 4 through 7 contain manuscripts that are to be submitted to reputable journals in statistics education. Specifically, Chapter 4 discusses the benefits of using written work to assess students' statistical thinking and gives a structure to assist students in writing statistical arguments. Chapter 5 begins to define the relationship observed between statistical thinking and statistical computing in students' written work. Chapter 6 shows how students use statistical computing in problem solving. A discussion and conclusion of all work is given in Chapter 7.

Chapter 2 : Literature Review and Theoretical Framework

Being able to think and reason statistically are two important outcomes of any statistics class. The GAISE college report identifies "emphasizing statistical literacy and developing statistical thinking" (American Statistical Association, 2005; American Statistical Association, 2016) as its first of six recommendations for college statistics courses. Many statistics teachers would probably agree that one of the main goals of their course has been accomplished when students are able to demonstrate higher levels of statistical thinking and reasoning.

While we can agree that we want our students to think and reason statistically, this may look different depending on who you ask. Several statistics education researchers have attempted to define statistical thinking and reasoning, leading to a mass of definitions, yet little consensus on the proper meanings.

Statistical Reasoning

While thinking and reasoning normally go hand-in-hand, these two terms do not mean the same thing (Garfield & Ben-Zvi, 2008). Reasoning is a specific form of thinking that is aimed at reaching justifiable conclusions. This can include logical reasoning, scientific reasoning or moral reasoning (Barrouillet & Gauffroy, 2013).

In a somewhat broad sense, statistical reasoning is "the way people reason for statistical ideas and make sense of statistical information" (Ben-Zvi & Garfield, 2004, p. 42). This could include connecting statistical concepts to justify patterns observed in the data, combining ideas of data and chance, understanding and being able to explain statistical processes, accurately interpreting results of statistical studies (Garfield & Ben-Zvi, 2008) making predictions from the data (Lovett, 2001) and justifying the use of given statistical tests (delMas R., 2004). In short, two key ideas for statistical reasoning are that students can explain *why* and explain *how* different statistical concepts work (Garfield & Ben-Zvi, 2008).

The current research on statistical reasoning focuses on either developing or assessing student's reasoning ability for given statistical topics. For example, some of these topics include students' reasoning about data (Chinn & Brewer, 2001; Lee H. S., et al., 2014) students reasoning about samples or sampling distributions (Ainley, Gould, & Pratt, 2015; Garfield, delMas, & Zieffler, 2012) students' reasoning about distributions and variation (Reading & Reid, 2006) and students' reasoning about inference (Zieffler, Garfield, delMas, & Reading, 2008).

A common theme of the above cited research is that the authors use statistical technology meant for teaching statistics, to either help develop or assess their students' statistical reasoning. As will be described in more detail below, technology can be very helpful to aid in student learning. It helps to automate procedures which could reduce cognitive load (Baglin, 2013; Biehler, Ben-Zvi, Bakker, & Makar, 2012) so the student can focus on recognizing patterns that emerge and making decisions about statistical tests instead of spending their mental energy on creating graphs and summaries. The technology can open the door for students to think about data and chance in new ways, for example, by helping them to present the data in different ways (Lee, et al., 2014) or allowing them to explore new methods for conducting inference (delMas, Garfield, & Chance, 1999). Part of this research study will be to determine if tools meant for doing statistics provide the same affordances to students.

Statistical Thinking

Again, while there is no consensus on what statistical thinking and reasoning are, this author believes that statistical reasoning is a subset of statistical thinking. As defined above, reasoning is a type of thinking (Barrouillet & Gauffroy, 2013), however, statistical thinking goes beyond the depth of knowledge that is required for statistical reasoning. In essence, a person who can reason statistically understands which formulas to use, how to use those formulas, and why they should use it. A statistical thinker could do all this and also potentially find new directions for the analysis or find new meanings within the given data set.

In general, thinking can be defined as using one's knowledge and ability to coordinate that knowledge to answer questions, make judgement calls and plan procedures (Barrouillet & Gauffroy, 2013). This means that thinking can take on several forms, such as problem solving, decision making, judgment and planning. Hence someone who exhibits statistical thinking might be more specifically defined as one who uses data to solve problems, make decisions and judgements or plan an analysis.

Similar to definitions of statistical reasoning, Ben-Zvi and Garfield define statistical thinking as "an understanding of why and how statistical investigations are conducted and the big ideas that underlie statistical investigations" (Ben-Zvi & Garfield, 2004, p. 7). Chance adds to this that a statistical thinker can apply concepts they have learned in a course to other situations and are likely to use their knowledge to question or investigate issues they encounter (Chance, 2002).

In more detail, Wild and Pfannkuch (1999) describe a framework to be used as a tool in organizing the different dimensions of statistical thinking. One of the dimensions of the framework defines five types of thinking that are fundamental to statistical thinking. These will be used to identify statistical thinking and reasoning in this research and are described in more detail below.

Recognizing the need for data. The first fundamental type of thinking is that, in doing analyses, one needs to recognize the need for data. One way that this could apply is to have students include a statement about their data in justifying the results that they have found. In their framework to assess student's statistical reasoning, Maker and Rubin (2009) state that students should make statistical inference statements as “Probabilistic – Generalization – From data”. The authors don’t consider a given conclusion complete unless the student includes a statement about data that was used.

Another way that this could apply is that students recognize that data they are using was properly collected, by including some element of randomization (Ben-Zvi & Garfield, 2004). Students need to be skeptical about how data was collected (Chance, 2002). Thinking about how the data was obtained and ensuring no bias exists in the data, is important to ensure that proper conclusions can be made. Evidence based on personal experience can also lead to poor decisions (American Statistical Association, 2016). Students need to be aware that anecdotal evidence is not enough to make proper statistical inference (Moore, 1990). Statistical decisions should be based on purposefully collected data and not on speculative observations.

Transnumeration. The second type of statistical thinking (Wild & Pfannkuch, 1999) is transnumeration. Transnumeration is a student’s ability to recognize that different presentations of data might tell a different story. Chance (2002) describes habits of mind of statistical thinkers, one of which states that statistical thinkers are constantly thinking about the variables that are involved in the study and they are curious of ways they can represent and visualize patterns in that data.

This means that exploratory data analysis (EDA) is extremely important for statistical thinkers. Ben-Zvi believes that EDA is not just a set of techniques that are taught to students, but it is a tool for the students to use, to help them explore and make sense of the data. Proper exploratory data analysis goes beyond a surface understanding of the data. This process is meant to give a statistician a preliminary understanding of what they might find if they dig deeper into the data and it can help to provide direction to the analysis (Ben-Zvi, 2004). However, this process isn't always easy, as finding and describing patterns in the data can take a high level of abstraction (delMas R. , 2004).

Lee et al. (2014) attempted to describe elements of transnumeration that pre-service teachers exhibit when analyzing data. They found that this may include making different graphical displays for the same set of data - such as comparing the output for both a dot plot and a box plot, using different bin widths in a histogram to tell a different story, or using multiple graphs in a dynamic software to demonstrate linking between variables. They also found that this can extend beyond graphical presentations of data, such as creating statistical summaries or using the original data table to find individuals who might not be conforming to the expectations the rest of the data set adheres to.

Consideration of variation. The third type of statistical thinking (Wild & Pfannkuch, 1999) is knowing that variability is an inherent part of a statistical analysis. Good statistical thinkers understand that all data collection involves variability (American Statistical Association, 2016). Whether this variability is because of individual differences, or if it is because of repeated measures on the same individual, students need to recognize that variability exists in all data and that it is natural (Moore, 1990). As an example, Pfannkuch and Wild (2004) discuss the importance of identifying variability in data and making attempts to reduce or even eliminate it through designing better experiments.

Being able to quantify and use the variability of the data is also an essential part of being a good statistical thinker. According to the GAISE report "it is essential to work on the development of skills that will allow students to think critically about … the quantification and explanation of variability" (American Statistical Association, 2016, p. 13). Without this understanding of variability, sampling distributions and subsequently, inferential statistics will be more difficult to understand (Chance, delMas, & Garfield, 2004). If a student cannot easily think or reason about the variability of the data, then they will likely struggle to understand why specific models work and why those models are appropriate for inference.

Reasoning with statistical models. The fourth element of statistical thinking (Wild & Pfannkuch, 1999) is knowing when and why it is appropriate to use a statistical model and the consequences of using an incorrect model. According to McCullagh (2002), a statistical model is a set of probability distributions. When conducting inference, traditional analysis is dependent on defining the probability distribution that best models the data, when parameter values are assumed true. For example, in a hypothesis test for a proportion, provided that certain conditions are met, a normal distribution with mean π and variance $\frac{\pi*(1-\pi)}{n}$ will approximately model the possible sample proportions that may occur. However, students need to recognize that, among other things, this model will only be a good fit for the data if a sample of size *n* is taken from a binomial population with probability of success, π.

As George Box famously quoted, “essentially, all models are wrong, but some are useful.” (Box & Draper, 1987). As a statistical thinker, it is up to the student to determine which model should be used and whether the model they have chosen is appropriate (Chance, 2002; Garfield & Ben-Zvi, 2008). To demonstrate, one might say for the example above that if the sample size is large enough, a normal distribution is an appropriate model for the sampling distribution of the sample proportion.

In addition to this, statistical thinkers must have a deep understanding of theories underlying the statistical models they choose to use and understand any limitations the chosen model might have for analysis (Garfield & Ben-Zvi, 2008). It is not enough for the students to just know that the sample size must be large - for example, $n*\pi \geq 10$ and $n*(1-\pi) \geq 10$ - in order to conduct inference. They should also understand the reasons why those conditions are in place and the consequences of not meeting those conditions. For our example, if the sample size is not large enough, then the central limit theorem doesn't apply. The true sampling distribution that models the data will be skewed, implying that a normal model would not be an appropriate fit for the data. The consequences would be that any inference that is conducted using the normal model is likely incorrect.

Integrating the statistical and contextual. The last type of thinking described by Wild and Pfannkuch (1999) is being able to synthesize context clues from the problem with one's own statistical knowledge. The first reason that this is important is because by knowing the context of the problem, students can then better plan the data collection and analysis (Garfield & Ben-Zvi, 2008). For example, when students are aware of the context of the problem, they may be better aware of sources of variability that exist naturally in the data, and may be able to take steps to reduce that variability in the analysis (Moore, 1990). It might also be the case that through contextual knowledge the student knows the distribution of the population will be skewed, implying that a large enough sample needs to be taken to ensure that the central limit theorem will apply, yielding an approximately normal sampling distribution of the statistics.

The second reason that it is important for students to be able to synthesize context clues from the problem with their own statistical knowledge, is that it helps them to make better conclusions from the data (Wild & Pfannkuch, 1999). One of the habits of mind that a statistical thinker should have is that they are continuously reminding themselves of the context of the data so that they can give their final conclusions in the context of the problem, avoiding the meaningless use of statistical jargon (Chance, 2002).

Table 2.1 gives a summary of the information presented on statistical thinking and reasoning. Thinking and reasoning were defined separately above to give the reader a sense of the types of outcomes that will be identified in student work throughout the research. However, a distinction between the two will not be made when identifying them in the research artifacts.

Table 2.1 Framework for identifying statistical thinking

Aspect of Statistical Thinking	Affordance for the student
Recognizing the need for data	• gives a statement about the data in the interpretation of the results • is able to identify if proper data collection techniques are used • recognizes that anecdotal evidence is not enough to conduct statistical inference
Transnumeration	• is able to think about the variables in the study and is curious of ways to represent them • uses exploratory data analysis tools for exploration, and not just prescribed methods • uses multiple representations to interpret and make sense of the data
Considerations of variation	• is conscientious that all data is variable • is able to redesign an experiment to reduce or eliminate variability • is able to quantify and explain variability in a way that helps them understand statistical inference
Reasoning with statistical models	• knows which type of model to use for specific types of data • knows when a specific model is appropriate to use for a given data set • understands theories underlying statistical models
Integrating statistical and contextual	• is able to plan data collection and analysis based on contextual knowledge of the data • is able to make more reasonable conclusions using the context of the data. • avoids unnecessary statistical jargon in their interpretations of the data

Statistical Computing

Data science and statistical computing tools are no doubt becoming more common in the statistics classroom (Horton, Baumer, & Wickham, 2014; Baumer, Cetinkaya-Rundel, Bray, Loi, & Horton, 2014). While not everyone agrees on what precisely constitutes statistical computing - for example Hardin, et al. (2015) describe 7 different courses dubbed statistical computing - most agree that our students need to learn how to integrate computer science into the entire statistical analysis process.

Some common themes occur in the literature of topics that could be covered in an undergraduate level statistical computing course. These topics include graphical displays, exploratory data analysis, data cleaning, non-parametric statistics including bootstrapping, simulations, Markov chain Monte Carlo (MCMC) , the use of more than one language - such as teaching students R, but supplementing with a secondary language such as SQL – and teaching students presentation methods like LaTeX in R Markdown or HTML for web design (Dierker, Kaparakis, Rose, Selya, & Beveridge, 2012; Gentleman, 2004; Hardin, et al., 2015; Monahan, 2004; Nolan & Temple Lang, 2010).

Suppose a statistician has been tasked with conducting a statistical analysis. If her only job was to put the data into her favorite statistical programming package, run a few tests to find results, then report those back to her employer, she really wouldn't be doing statistical computing. The use of the technology in this case is what Pea (1987) dubs an amplifier. The technology is being used to simplify tasks, helping the statistician to use their time more efficiently. According to Nolan and Temple Lang (2010), statistical computing is

> about using the computer to make sense of data. It is inherently interactive, with the next step determined dynamically by the content we see in the previous steps. It is the tightly coupled combination of programming, visualization and statistical reasoning (p. 5).

Suppose instead our statistician did the following. First, she went to a website and scraped the data she needed, using data mining tools. Then upon inspection of her results, she realized the data was not in a format suitable for analysis. She then decided to use her database skills – like

SQL - to clean and sort the data for relevant information. Next she explored the data through graphical displays to find patterns or points of concern which helped her determine the type of statistical test she was going to use in her analysis, and if conditions for inference for that test were met. Next, she used the technology to crunch the numbers and find an answer for her problem. During the analysis, she saw that data she collected behaved differently over time, so she decided to create an interactive presentation of the results on a webpage that helped to demonstrate this. We can see that this process is much different than the first one given above. This statistician is engaging in statistical computing.

Every one of these steps is not necessary to be considered as doing statistical computing. However, statistical computing is not just about using a computer to get the answer to a statistical problem. Statistical computing is the integration of technology into all steps of a statistical analysis, from data collection to presentation. As we saw above, this could include data mining, data cleaning, exploratory data analysis, statistical analysis and presentation of results.

Affordances of statistical computing. The main advantage that students probably recognize about taking a statistical computing course is that the use of statistical computing skills has become more prevalent in the work force. In their paper, Hardin et al. (2015) speak of several students who had written to their professors after graduating and telling them the usefulness of the statistical computing course they had taken. This is no wonder, since students graduating at the Bachelor's and Master’s level spend much of their time in the workforce using skills like “retrieving, filtering and cleaning data and doing initial exploratory data analysis” (Nolan & Temple Lang, 2010, p. 4)

Phillip, Schuler-Brown and Way (2013) note that while it is becoming more and more acceptable to discuss the work preparation aspects of any program, thinking about schooling in such a way limits other important goals that students might have when obtaining a higher education, such as intellectual, social, civic, ethical and aesthetic reasons.

In particular, students can benefit intellectually from taking a statistical computing course. While, statistics and computer science are different fields, they both require many of the same skills for a student to be successful, such as logic and abstraction (Wing, 2008; delMas R. , 2004) and both help their students to develop these skills. Some researchers have found (Kafura, Bart, & Chowdhury, 2015; Yuen & Robbins, 2014) that when students not majoring in computer science or statistics learned computing with a statistical context, both their computational and statistical thinking grew and supported the other. In particular "programming assists students in understanding how data is organized, structured, and presented so that they can accurately analyze it in both numeric and graphical form" (Yuen & Robbins, 2014, p. 14) while the "quantitative concepts were also developing their programming skills".

One major reason that researchers have found to support students learning statistical computing is that (if done properly) it automates computational procedures and helps to reduce the cognitive load on the student (Baglin, 2013). "Attention to key statistical concepts seems to be easier if the cognitive load required for computation and graph drawing is minimized by software" (Biehler, Ben-Zvi, Bakker, & Makar, 2012, p. 680). Students can more easily explore trends in the data (exploratory data analysis) when they aren't worried about the nuances of how a graph is drawn. Or they can spend more time interpreting and understanding why the results of a statistical study make sense since they haven't spent all their cognitive capacity trying to manually compute the results of that analysis.

The second major benefit of studying statistical computing is increased computational thinking skills. As defined by Wing (2008), computational thinking is a mixture of mathematical, scientific and engineering thinking that is not constrained by the bounds of reality. It allows one to think critically and abstractly about real world problems, while allowing for the use of the almost infinite power of the computer to solve problems that might not otherwise have solutions strategies in the real world. A person that can think computationally has problem solving skills that allow them to formulate a solution strategy which is then conveyed to a computer, telling it how to implement the given strategy (Wolfram, 2016).

Another advantage to using technology in a statistics classroom is that it offers alternate techniques to solving problems than have been unavailable in the past. For example, using simulation-based methods for developing distributions and solving statistical inference problems has recently become a popular teaching technique (delMas, Garfield, & Chance, 1999; Garfield, delMas, & Zieffler, 2012; Holcomb, Chance, Tietjen, & Cobb, 2010; Lane-Getaz, 2010; Tintle, VanderStoep, Holmes, Quisenberry, & Swanson, 2011). A major advantage to this type of analysis is that students don't have to have an in depth understanding of the properties of distributions, such as the normal distribution (Pfannkuch, et al., 2011). Instead, students conceptualize the idea of *what would my statistics look like if we could repeat this process several times?*; a concept that is foundational for several topics in statistics, including sampling distributions, confidence intervals and hypothesis testing. Instead of spending time teaching students how to read a normal distribution table, they are using software to generate p-values and explore the idea of "extreme" statistics.

One example of this is the CATALST program (Garfield, delMas, & Zieffler, 2012). The goal of the program was to make sure that students learning statistics didn't see it as a "cookbook" of procedures to be followed. By the end of the course, the researchers wanted their students to know how to "cook", or in other words, know the essentials of statistics, without being someone who is simply following a statistical recipe that their instructor gave them. The authors found positive results for their program and noted that students' reasoning about statistics was on par with students who had taken a traditional introductory statistics course.

The final benefit of utilizing statistical computing is learning multiple programming languages or programming tools. Ideally, this would help students to perform the different tasks of statistical computing that they are expected to do in order to conduct a complete statistical analysis, such as data collection, data cleaning, analysis and presentation. In addition to the practical applications of learning multiple languages and tools, students may also develop decision making skills so that they know which of these tools is best for solving different aspects of their statistical problem (Nolan & Temple Lang, 2010). Brown and Kass (2009) claim that success in statistics requires that students have adaptable strategies for solving problems but that our students have a tendency to "attack problems using blunt instruments and naïve attitudes" (p. 105). By teaching statistical computing to our students, we hopefully give them sharp tools and the knowledge of when to use them.

Learning multiple programming languages could also help students develop pattern recognition skills. Nolan and Temple Lang (2010) argue that the main goal in teaching programming is for students to be able to extract concepts from a specific problem and transfer concepts to the specifics of other languages and environments. When programming languages are taught together, students can see the patterns that emerge between them and get a better understanding of programming as a general concept instead of the idiosyncrasies of individual languages. Even when students only learn one language, developing code based on examples can be a helpful tool for learning syntax.

In summary, learning statistical computing provides students with the opportunities to increase their knowledge of the tools and become more proficient statisticians in the real world. Students can use their knowledge of statistical computing tools to efficiently create graphs and summaries of their data. In doing this, they have more cognitive capacity to spend on discerning patterns they observe or understanding concepts they are trying to grasp. Additionally, learning statistical computing may impact one's computational thinking which could in turn improve the student's problem solving skills, their logic and reasoning skills and their sophistication to think critically and abstractly. Finally, the technology could be used to implement new or multiple statistical analyses, allowing the student different perspectives on statistical concepts.

Table 2.2 synthesizes the information described in this section. Based on the research that has been done, these are believed to be the primary elements of the affordances that statistical computing has to offer students, in terms of intellectual benefits.

Table 2.2 Framework for identifying the benefits of statistical computing

Aspect of Statistical Computing	**Affordance for the student**
Automate computational procedures	• creates multiple graphs or summaries to make sense of the data • uses the technology to conduct the analysis, then makes proper inferences
Increase computational thinking	• creates a solution strategy and can communicate it to the computer • is able to think critically or abstractly about a concept that the technology is demonstrating
Offer new methods to explore concepts	• does not follow a prescribed set of procedures to answer a question • comes up with their own methods and reasons for using the technology in analysis
Aid in pattern recognition and decision making	• recognizes patterns, either in the way things are coded or in how the results behave • uses the output from the technology to influence their decision of where to go next in the analysis

It is believed that there is a symbiotic relationship between a student's ability to think and reason statistically and their use of statistical computing benefits (Figure 2.1). This paper will explore this relationship in more detail and describe some potential implications of relationships that have been observed.

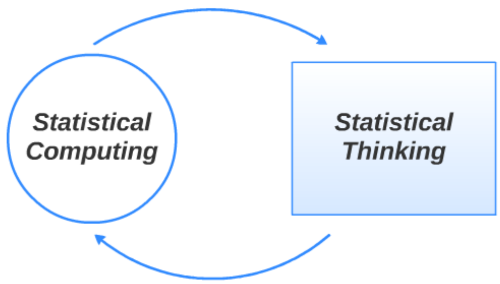


Figure 2.1 Proposed relationship between statistical thinking and statistical computing

Assessing Statistical Thinking and Statistical Computing

Now that a framework has been given to identify statistical thinking and statistical computing, a medium for students to present these constructs is needed. Structured written assessments may be useful in helping students demonstrate their statistical thinking and statistical computing.

Keys (1994) found that scaffolding written reports for students helped them to focus their thinking on pertinent information. A popular example of structuring a response comes from Toulmin's (1958) model for argumentation, originally used to aid in creating verbal arguments for ill-structured problems (Karbach, 1987). Inglis, Mejia-Ramos and Simpson (2007) discuss several studies in which college mathematics students were expected to use Toulmin's model, or a modification of it, in order to create well defined mathematical arguments.

Toulmin's model for argumentation directs users to create persuasive arguments by:

- Stating a *claim*;

- Identifying the *data* that supports the claim; and
- Backing the claim with *warrants*

If needed, the argument can also be enhanced by:

- Providing *backing* for the warrants;
- Identifying *qualifiers* to support the strength of the relationships; and
- Providing a *rebuttal* indicating circumstances where the argument may not be true (Toulmin, 1958).

To assist introductory statistics students in writing complete arguments and help them to better convey their reasoning, Woodard (2016) suggested teaching students a similar four element model, dubbed the four sentence structure. The four sentence structure directs students to:

- Answer the question;
- State relevant facts from the problem;
- State the implications that the facts imply; and
- Explain how the facts lead to the conclusion.

In relation to Toulmin's model, answering the question is similar to stating the claim and stating relevant facts is similar to identifying data to support the claim. Stating the implications that the facts imply doesn't translate directly to one of the items in Toulmin's model. Instead this element combines both supporting claims with warrants and providing backing for the warrants. Finally, depending on how a students has answered the question, explaining how the facts lead to the conclusion could be similar to providing more backing for the claim, restating the original claim or it may not translate to any of the elements in Toulmin's argumentation model (1958).

Problem Solving Methods

Another part of this research examines how students utilize statistical computing technologies to solve statistical problems. To do this, a framework for understanding problem solving will be necessary. Several authors have proposed methods for problem solving structures in mathematics, and how these methods can be used to help conduct research on students' ability to solve problems.

As one example, Schoenfeld (1985) described the knowledge and behavior necessary to adequately characterize students' mathematical problem-solving performance. This included knowledge of four items: resources, heuristics, control, and belief systems. Resources include knowing the facts, procedures, skills and algorithms that students possesses and bring to the table when they start solving the problem. Schoenfeld discusses that when we know, as researchers, what our students know about a problem, it helps us to better understand the strategies they use to solve the problem.

The second item defined by Schoenfeld (1985) are heuristics, or rules that help students to solve problems. One example of a heuristic is Polya's (1957) four step process for problem solving. These steps include the student understanding the problem, devising a plan, carrying out the plan and looking back at what they have done to check their solution. Knowing a strategy such as this one can help students to make progress on a problem, even if they don't possess all of the necessary resources to solve the problem.

Third, Schoenfeld (1985) discusses the need to understand the student's control of the problem. Control is the way a student chooses to use the resources and heuristics that they have to solve the problem. Each time a student makes a major decision about how to solve the problem or the resources they should use, they are showing their control of the problem.

Finally, according to Schoenfeld, a student's beliefs system plays a role in how students solve problems. When there is context associated with a problem, students beliefs about that context may influence how they perceive the problem should be solved. This can happen even without the student being aware of it.

Lee and Hollebrands (2006) expanded on the work of Schoenfeld (1985) and defined a framework that helps classify students' problem solving behaviors in the context of a technological task. The phases in this structure include analysis, planning, implementation, assessment, verification and organization. Definitions of these items can be found in Figure 2.2.

Analysis	Students attempt to understand the problem, its constraints, and consider useful perspectives for the problem.
Planning	Students work towards devising a sequence of actions or formulate a strategy to pursue a solution.
Implementation	Students carry out a pre-planned strategy, which may or may not be explicitly stated.
Assessment	Students take explicit actions to monitor their progress towards a solution.
Verification	Students believe they have a solution and either evaluate whether or not it is correct, or engage in activities to explain or justify their solution.
Organization	Students take actions to arrange their environment (either with paper or in the technology tool) in a way that they perceive will be beneficial in working towards a solution.

Figure 2.2 Phases of problem solving

Lee and Hollebrands (2006) originally used this structure to analyze how students solve a proportional reasoning problem using technology. The researchers had students working in pairs to solve a problem about placing a set number of fish to three different ponds, with constraints on the proportion of each type of fish that went into the different ponds. The researchers used the framework to classify the ways in which students solved the problem.

Even though the focus of this work is not proportional reasoning, students will use technology to help them solve a statistics problem. This framework has been tested and has been shown to work when analyzing how students use the computer to solve a problem and thus it seems reasonable to assume that this structure should aid in assessing the problem solving techniques students use when solving a statistical problem with technology.

Chapter 3 : Methodology

Study design

To determine how statistical thinking and statistical computing impact each other, a small-scale, two-phase qualitative research study was conducted. According to Groth (2010), if done appropriately, qualitative research can be scientific, rigorous, generalizable and objective, and can "provide a means for conducting scientific studies that contribute to our understanding of statistics teaching and learning". The first phase of the study involved collecting artifacts from students while they were taking a second course in statistics. The second phase of the study involved conducting task-based interviews with students who had completed the second course in statistics.

Quantitative methods, specifically traditional regression analyses, were not appropriate for this research. Because of the small sample of students used in the study, and potential correlation within and between subjects, conducting an effective quantitative analysis would have been difficult. More importantly though, this research is meant to be exploratory and descriptive in nature. This is why the researcher has chosen to conduct a qualitative analysis.

Participants

Participants for this study came from a small private women's college in the southeastern United States. During the fall semester of 2016, fifteen students began a second course in statistics with the primary investigator. Fourteen of these students finished the course and volunteered to participate in this study.

Statistics 2 is a continuation of introductory statistics. It includes topics of one and 2 sample inference, two way tables, simple and multiple linear regression and one and two factor ANOVA. Students were also taught introductory statistical computing practices in R while taking this course.

While the only prerequisite course for statistics two was introductory statistics, the statistical and computational backgrounds of students did vary. Some students had only taken the introductory statistics course, while some had taken a course in probability theory or statistical computing. Six students had experience with programming before taking the course, through a statistical computing course offered the previous semester. In addition, two participants were majoring in computer science.

Data Sources

According to Creswell, multiple forms of data should be collected and themes across the different types of data should be used to perform a qualitative analysis properly (Creswell, 2013). Because of this, data for this research was collected from four primary sources. These sources are teacher field notes, a demographic survey, writing assignments and a task-based interview. More information about these data sources is described below.

Data for this study was collected during two semesters, fall 2016 and early spring 2017. As part of the requirements of the statistics 2 course, students were responsible for completing several writing assignments. Additionally, students who completed the course and were not taking any courses with the primary researcher in the spring of 2017 were asked to participate in a task-based interview in January of 2017. Five students volunteered for these interviews, 2 of whom had taken the statistical computing course previously and 3 that had not.

Field Notes. Field notes were kept so that there was a record of conversations, where students discussed with their classmates or with the instructor, ideas that demonstrate statistical computing or statistical thinking. For example, in-class group work was one of the modes of instruction in the course. To find appropriate solutions for many of the group-work problems, students needed to use the computer and were encouraged to share their thinking with their group mates. As the instructor walked around the classroom assisting groups, she took notes about the student’s discussions. In addition, when students came to the instructor’s office for additional help, notes were taken after they left that may have been pertinent to the research.

Demographic Survey. Students in the second course in statistics completed a survey to determine their statistical and computational background. The survey can be found in Appendix C. The primary purpose of the survey is to determine differences in student's backgrounds in both statistics and computer science.

Writing Assignments. One of the primary source of data for this research were writing assignments. Throughout the course, students completed a total of ten individual writing assignments utilizing the four sentence structure created by Woodard (2016). In this structure, which will be described in more detail in Chapter 4, students were directed to write about solutions to statistical problems by:

- Answering the question;
- Stating relevant facts from the problem;
- Stating the implications that the facts imply; and
- Explaining how the facts led to the conclusion.

Seven of the ten assignments were used for the purposes of this research. These assignments gave students the opportunity to demonstrate their statistical computing, while encouraging them to explain their statistical thinking and reasoning. The assignments not used for research were necessary for assessment in the course but did not give the students a potential medium for demonstrating statistical thinking and statistical computing. In total, 91 documents were collected from students. All writing assignments utilized in the course can be found in Appendix A.

The general strategy of the writing assignment was to have students make a decision about a scenario using data that was provided for them. To assist students in giving reasonable responses, Chance (1997) believes that students need to be given enough direction to complete the assignments. As such, the instructor worked through the first writing assignment with students and gave them guidance as to how to write the second assignment. Students wrote their responses individually starting with the third assignment.

Most of the assignments necessitated the use of the computer to give a reasonable solution. This was done to assess if students recognized the need for data when providing responses to the prompt, one of the five elements of statistical thinking (Wild & Pfannkuch, 1999). Jolliffe (2007) believes that pairing written work with computational output provides a reasonable tool for assessment. While not all students will obtain the same results students could obtain the same output from the computer. When students are expected to interpret the output, we add another layer of differentiation that can be used to assess differences in student understanding.

While the structure of the assignments stayed the same, the purpose of the assignments varied over the semester which provided the opportunity to assess different types of statistical thinking while maintaining a certain consistency for the students. For example, some assignments called for students recognizing the need to use the data provided as opposed to their intuition to create an appropriate model for prediction (American Statistical Association, 2016; American Statistical Association, 2005; Wild & Pfannkuch, 1999). Other assignments asked student to determine if a proposed method was valid when, unbeknownst to the students, the method utilized a common misconception in statistics (Rossman & Chance, 2004). Other assignments included creating a simulation to answer the question (Tintle, et al., 2015) or critiquing the work of another individual to determine if that individual had correct thinking (Rossman & Chance, 2004).

Task-based Interviews. The fourth source of data that was used were task-based interviews. Task-based interviews are structured or semi-structured interviews where students are expected to complete a task and explain their thinking while doing so (Koichu & Harel, 2007). These types of interviews are useful tools for researchers because they allow us to see how students solve problems and they can help researchers form conjectures about students' content knowledge and the ways in which they think about and solve problems (Goldin, 1997; Koichu & Harel, 2007).

Before conducting interviews with students from the fall 2016 statistics 2 course, a pilot interview was conducted with a student who had completed statistics 2 prior to taking statistical computing in the spring of 2016. Information collected from this interview was used to modify the interview protocol. Conducting a pilot interview is also helpful in training the researcher to conduct the interviews appropriately for the primary analysis (Nguyen, McFadden, Tangen, & Beutel, 2013).

In the spring of 2017, students who had taken the statistics 2 course and who were not taking any other courses with the primary investigator were asked to participate in the task-based interviews. 5 students volunteered to take part in the interviews. These interviews lasted for approximately 45 minutes to an hour. Students completed these interviews individually. The interview was set up in three phases. Each of these phases allowed students to demonstrate their statistical computing and statistical reasoning ability in different ways.

The interview protocol can be viewed in Appendix B. The intended purpose of phase 1 of the interview was to have students use the computer to see if they understood the conditions and reasoning for doing a one-way ANOVA test. Several of the questions that were asked in this section were intended to allow students to elaborate more on the knowledge of why we do certain procedures in a statistical test, which can be a good measure of a student's statistical thinking and reasoning (Wild & Pfannkuch, 1999). The students were also asked to use the computer to "do a full statistical analysis" meaning that they were encouraged to use the computer for exploratory data analysis, statistical analysis and interpretation instead of just using the computer to conduct the hypothesis test. Or in other words, they were being encouraged to do statistical computing.

In phase 2 of the interview, students were given a question that was similar to the format of the writing assignments they had been given in class. As they were formulating their argument, students were given the chance to discuss their thoughts about the problem, as students may have difficulty putting their reasoning into words on paper, but may be able to verbalize it (Park, Park, Lee, & Lee, 2016). Students were not required to write down their final solution, but they were asked to verbalize all four parts of the four sentence structure after they had the chance to talk through the problem on their own.

During this phase of the interview students were given computer output, which they needed to utilize in order to answer the question properly. This question focused more on assessing the student's statistical thinking, as opposed to their statistical computing sophistication. However use of the computer in the first phase of the interview may have made this process easier for the student, as the two problems were related.

The third phase of the interview included questions about student's thoughts in the sequencing of the second course in statistics and the statistical computing course. Specifically, students who had taken computing prior to the second course in statistics were asked if they felt like having the extra experienced helped them to do well in statistics 2, and students who had not taken computing were asked if they thought it would have helped them to do better in the course. While the goal of this research is to focus on ways that statistical computing techniques can help or hinder a student's ability to reason statistically, research has shown that student beliefs can also play a role in how well they perform in a course (Gal, Ginsburg, & Schau, 1997; Gal, 2004).

Analysis

As described above, the goal of this research was to make sense of the phenomena that is observed in student work in order to refine the proposed connections that were made in the literature review between statistical computing and statistical thinking. The primary artifacts that were analyzed to answer the research questions were the writing assignments and the task-based interviews. The demographic survey and field notes that were collected were used as supplemental information.

Writing Assignments. Students in the second course in statistics were expected to complete ten writing assignments as part of the coursework. Seven of these writing assignments were used for the purpose of this research. Students electronically submitted their .docx document to be graded. A copy of each ungraded document was saved for the purpose of the analysis.

Before the analysis could be done, all documents needed to be formatted identically. A colleague went through and assigned a code number to each student. These were put in place of the student’s name to blind the researcher when analyzing students’ work. A color coding scheme based on the four sentence structure was applied to each document to aid in analysis. This included red text for answering the question and stating the conclusion, blue text for stating the facts and orange text for stating the implications that the facts implied An example of this color coding can be seen in Figure 3.1.

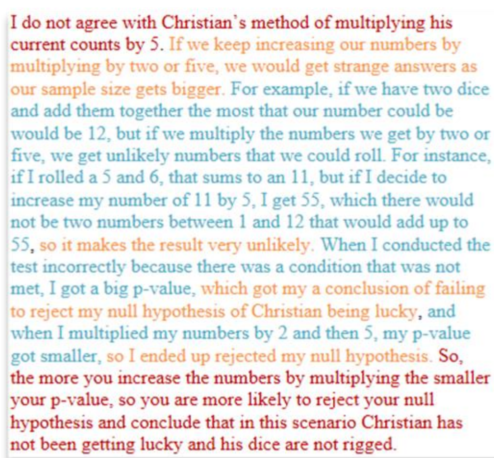

I do not agree with Christian’s method of multiplying his current counts by 5. If we keep increasing our numbers by multiplying by two or five, we would get strange answers as our sample size gets bigger. For example, if we have two dice and add them together the most that our number could be would be 12, but if we multiply the numbers we get by two or five, we get unlikely numbers that we could roll. For instance, if I rolled a 5 and 6, that sums to an 11, but if I decide to increase my number of 11 by 5, I get 55, which there would not be two numbers between 1 and 12 that would add up to 55, so it makes the result very unlikely. When I conducted the test incorrectly because there was a condition that was not met, I got a big p-value, which got my a conclusion of failing to reject my null hypothesis of Christian being lucky, and when I multiplied my numbers by 2 and then 5, my p-value got smaller, so I ended up rejected my null hypothesis. So, the more you increase the numbers by multiplying the smaller your p-value, so you are more likely to reject your null hypothesis and conclude that in this scenario Christian has not been getting lucky and his dice are not rigged.

Figure 3.1 Example color coding scheme to aid in writing assignment analysis

Once the 91 documents of students' work had been coded for instances of statistical thinking and computing as defined in the frameworks in Tables 2.1 and 2.2 respectively, the documents were then organized using document diagrams similar to Lee and Hollebrands (2006). In their work, which was influenced by Schoenfeld (1985), the authors segmented and classified student's work into categories to help visualize the type of goal students had, and whether the technology was supporting or not supporting their goal directly. An example of the document diagrams utilized in this work can be found in Figure 3.2.

A more in depth discussion of the preparation and analysis process, as well as the patterns identified in the writing assignments will be discussed in Chapter 5.

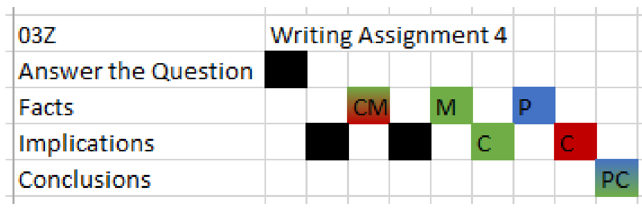


Figure 3.2 Example of a document diagram for the four sentence structure and use of statistical thinking and computing in a student's document

Task-based Interviews. Once all interviews were completed, the video of the interview and the screencast of the student's work on the computer was compiled into picture-in-picture videos. An example of this can be seen in Figure 3.3.

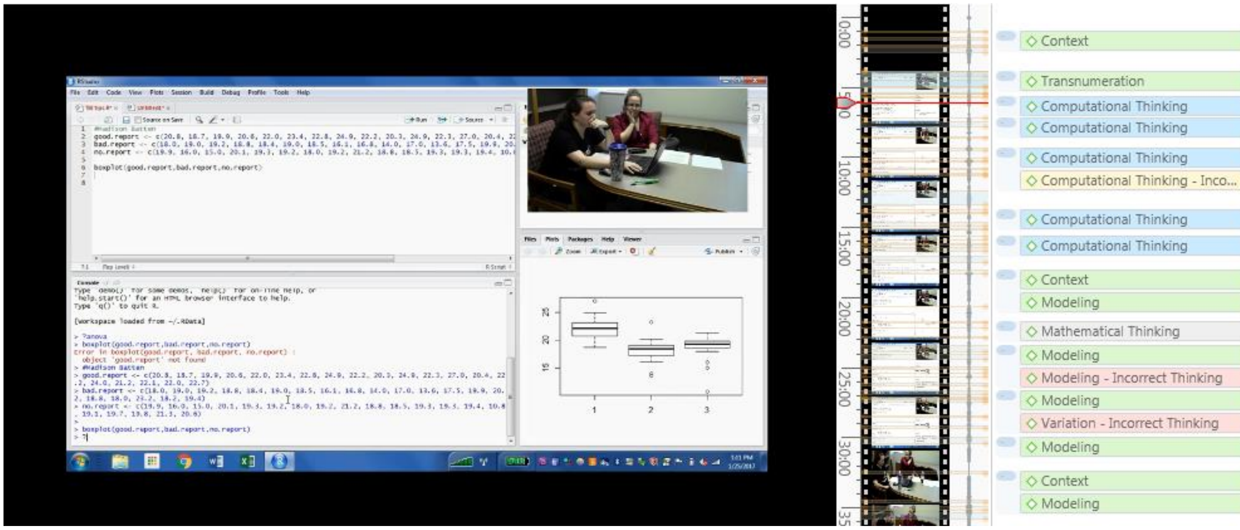


Figure 3.3 Picture-in-picture task-based interview example with statistical thinking and statistical computing codes

Each task-based interview was examined at least three times for the purposes of this analysis. In the first pass through the interviews, the videos were coded for instances of statistical thinking and computing as described in the frameworks in Tables 2.1 and 2.2. Each video was inspected a second time for accuracy of these codes, with a sample of students' work viewed and discussed among both researchers until agreement was met. In the third iteration of the videos, students' work was categorized into the problem solving phases defined in Figure 2.2. Graphics were made to assist in finding patterns. A more in depth description of how the task-based interviews were analyzed can be found in Chapter 6.

Ethical Considerations

This research has been approved by the Internal Review Boards (IRB) of both North Carolina State University and Meredith College (Appendix D and E). Consent was obtained from all participants using the Informed Consent Form found in Appendix F. Since the researcher was the instructor of the course, students were notified that their grades were in no way impacted (positively or negatively) by participation in the study. All interviews were conducted after the student had completed all coursework for their respective courses (either statistical computing or the second course in statistics). This was done so as to not unfairly affect student learning (i.e. through extra examples that non-participants didn't get to see) or grades of any students involved (or not involved) in the study. Pseudonyms are used in place of student names in all forms of dissemination of this work. It is believed that there was minimal risk for this study.

Chapter 4 : Writing Assignments to Assess Statistical Thinking

Journal

This manuscript was written to be submitted to JSE – The Journal of Statistics Education. JSE is published online by the American Statistical Association and can be accessed freely by anyone. The mission of JSE is to "disseminate knowledge for the improvement of statistics education at all levels", hence any statistics educator may find value in the materials found in the journal. JSE accepts many types of papers provided that they are relevant to their mission. One type specifically mentioned is a manuscript that demonstrates an innovative assessment. No word or page limitations are cited on the submission website.

Abstract

One of the main goals of statistics is to use data to provide evidence in support of an argument. As part of a second course in statistics, students were expected to answer prompts that required them to 1. Take a stance on an argument, 2. Defend their position with facts given in the prompt and 3. Discern the implications that those facts implied. This model, which was originally created to assess students in introductory statistics and has been adapted for the second course in statistics, takes a unique approach towards assessing students understanding of statistical concepts through writing. This paper will discuss some popular forms of writing assessment currently in use, to demonstrate the differences between the methods. I will then go into more detail on a few of the writing assignment prompts that were used in the course, their intended assessment purpose and common answers that students gave to these assignments.

Introduction

A common problem in education is determining how we can assess certain learning outcomes for our program. How many times have we sat in curriculum development meetings wordsmithing the perfect learning outcome only to have a colleague reiterate, “it sounds great, but how do we assess it?!”

Besides the need for record keeping in our institutions, assessment of our students is very important. According to Garfield (1994), the primary purpose of assessment is to improve student learning. Additionally assessments provide information to students about how well they have learned a topic and help them to determine their overall strengths and weaknesses. Assessments can also help to inform the instructor on how well students have learned the material and provide information about which topics may need to be taught more in depth or differently (Chance, 1997; Shibli, 1992).

A potential learning outcome for a statistics course could be that we want our students to think and reason statistically. However, creating appropriate assessment to evaluate a student’s ability to think and reason statistically could be difficult. Fortunately a framework for identifying statistical thinking was created by Wild and Pfannkuch (1999). Their perspective of statistical thinking includes: recognizing the need for data, transnumeration, considerations of variation, reasoning with statistical models and integrating the statistical and contextual. While being able to identify statistical thinking is important, we need to create appropriate mediums for our students to demonstrate ability to think statistically.

The purpose of this paper is define why and how written work in statistics can be beneficial to our students, to instructors and to researchers, as well as to review the ways in which instructors are currently utilizing this type of assessment. Additionally, a structure to assist students with their writing is proposed, some sample writing assignments that utilize this structure will be given, and examples of students demonstrating statistical thinking in the assignments will be discussed.

Literature Review

Using written work in statistics courses to assess students is not a new concept. In the early 1990s several authors expanded on the discussion of using written work to assess students in quantitative courses (Radke-Sharpe, 1991; Garfield J. B., 1994; Shibli, 1992). For example, Radke-Sharpe (1991) discussed the pros and cons of using written assessment and Garfield (1994) further encouraged its use by including it as a type of assessment to use in statistics courses. Writing is a tool that has been found to help students conceptualize topics, to aid instructors in assessing student understanding, and to inform researchers as to some of the ways in which students learn concepts in statistics, as well as other fields. Below, we will discuss some of the benefits of using written work and ways that written work is currently used in statistics and other quantitative courses. Additionally a new writing structure that may be helpful in assessing our students' statistical thinking is proposed.

Why use writing assignments? The Principles and Standards for School Mathematics (National Council of Teachers of Mathematics, 2000) and the Guidelines for Assessment and Instruction in Statistics Education (GAISE) (American Statistical Association, 2016) both recommend using written work to assess students. There are several benefits to our students when they are asked to use written assignments to expand on their thinking. For example, Johnson (1983) believes that "good writing is part of the problem solving process." The writing process involves thinking and reflection (Burns, 2004). It forces students to slow down (Van Dyke, Malloy, & Stallings, 2015) and organize their thoughts so that they can express them clearly. It has the added benefit of helping students to improve writing while encouraging creativity (Radke-Sharpe, 1991).

All of which may in turn help students internalize and conceptualize what they are learning (Gopen & Smith, 1990; Radke-Sharpe, 1991).

For example, written work may help our students to better understand the material. Writing removes the temptation to just memorize formulas (Shibli, 1992) and it helps students focus on the concepts in statistics (Goenner & Snaith, 2003). This could help students to understand the reasoning behind statistical models, one of the elements of statistical thinking identified by Wild and Pfannkuch (1999). Additionally, for students that are not mathematically inclined, writing may give them an outlet to demonstrate that they understand the material, even if they struggle to do so though computational procedures, which could be a confidence booster for students who are taking statistics courses as a requirement (Jolliffe, 2007).

In addition to helping our students gain a deeper understanding of the concepts, writing also helps them to become better communicators of content and conceptual knowledge (Johnson K. G., 2016). With practice, the goal of these types of assessments is for our students to come to understand the language of statistics (Chance, 1997) and use it to demonstrate what they know about statistical topics. It gives students an opportunity to "take a complex idea and explain it with clarity" (Cline, 2008, p. 401) so that they may communicate their understanding of statistics to a general audience (Goenner & Snaith, 2003; Johnson K. G., 2016). We are providing students with a type of assessment that affords them a different way to demonstrate statistical thinking in creative ways (Radke-Sharpe, 1991).

Writing assignments already in use. Educators at all levels and in several quantitative disciplines have already discovered the benefits of using students' written work for assessment and research, and have implemented their own assignments in their classrooms. Some examples include minute papers, entry and exit slips and memos for shorter written assessments to lab reports and course projects for longer written assessments.

For some, the creation of particular assessments was largely restricted by the amount of time it would take to grade and give informative feedback for larger classes. One solution utilized by Stromberg and Ramanathan (1996) was to use entry and exist slips. These are responses that students give to a question that is asked either at the beginning or the end of class. Students responses are restricted to the size of a note card. Because of the limited space to write, students are ensured to give shorter answers to the questions. Similarly, having his business and economic students write memos, Goenner (2003) ensured that students wrote coherent but concise reports. This method also had the added benefit that students were practicing with a medium they would potentially use in their careers. Even though these assessments are expeditious to complete and to grade, they can also be powerful in helping students learn. Drabick, Weisberg, Paul and Bubier (2007) found that minute papers utilized in a large psychology course helped to improve both conceptual and factual understanding.

Other types of written assessment may take more time to grade, but may provide instructors with a better understanding of their students' thinking. For example Stromberg and Ramanathan (1996) had their students write a journal that supplemented their notes. These journals provided feedback to the instructors as to how well the students understood the material and it helped students to refine their writing to be more clear and concise.

Burns (2004) had her students write about mathematical ideas so she could identify misconceptions students had about concepts. In one instance, after a lesson on fractions, a student noted that a sixteenth was the smallest possible fraction. He came to this conclusion because in the activity they had done in class, a sixteenth was the smallest unit of fraction they had observed. She was able to use this information to continue the lesson the next day and show students that fractions could get smaller and smaller if someone took the time to continue cutting pieces in half.

Course-long writing assessments have been found to help students to connect concepts. Holcomb and Ruffer (2000) created six projects for their students to complete over the course of the semester. Each of the projects covered different material, but the context was the same, providing consistency and helping students develop a thorough understanding of the data with which they were working. Chance (2002) also structured her assignments to build on each other. When she would give feedback to her students, it was meant to be used in subsequent assignments.

Not all assessments need to be semester long commitments to help students make connections. Concept mapping has also been found to help students organize their thinking and make connections between concepts. Bolte (1999) used concept mapping in conjunction with interpretive essays in her calculus class and found that students were able to make better connections between concepts discussed in the course.

In addition to being useful tools for students and instructors, writing assignments can also be beneficial for researchers. For example, Lee et al. (2014) used students' written documents to understand how pre-service teachers used transnumerative actions when solving problems. While the authors stated that written work doesn't always give an accurate picture of the entire problem solving process, they did find that using written work to assess students' statistical problem solving across multiple institutions "made things more manageable".

Nahrgang and Petersen (1986) used the written work completed by their students, paired with Blooms taxonomy to begin to understand how students perceived the action of factoring. This gave the researchers a way to compare and categorize the level of understanding their students had on the topic. Traditional problem solving methods may not have provided the researchers with similar information.

Not just having students write in statistics courses, but structuring the assignments in a specific way, may also aid researchers. Biehler (2007) had students complete written project reports where they were instructed to include an introduction, analysis and summary/conclusion sections. Through these reports Biehler was able to learn more about students understanding about distributions of random variables.

The Four Sentence Structure. Writing may be difficult for students, especially when they are expected to write about a topic that they do not fully understand (Stromberg & Ramanathan, 1996) or if they have not had much practice with technical writing. While a statistics instructor can provide students with the information they need to know to solve problems, teaching them technical writing to present strong arguments may be outside the scope of the course.

Keys (1994) found that scaffolding written reports for students helped them to focus their thinking on pertinent information. A popular example of structuring a response comes from Toulmin's (1958) model for argumentation, originally used to aid in creating verbal arguments for ill-structured problems (Karbach, 1987). Inglis, Mejia-Ramos and Simpson (2007) discuss several studies in which college mathematics students were expected to use Toulmin's model, or a modification of it, in order to create well defined mathematical arguments.

Toulmin's model for argumentation directs users to create persuasive arguments by:

- Stating a *claim*;
- Identifying the *data* that supports the claim; and
- Backing the claim with *warrants*

If needed, the argument can also be enhanced by:

- Providing *backing* for the warrants;

- Identifying *qualifiers* to support the strength of the relationships; and
- Providing a *rebuttal* indicating circumstances where the argument may not be true (Toulmin, 1958).

To assist introductory statistics students in writing complete arguments and help them to better convey their reasoning, Woodard (2016) suggested teaching students a similar four element model, dubbed the four sentence structure. Connections between the two models will be discussed below. The four sentence structure directs students to:

- Answer the question;
- State relevant facts from the problem;
- State the implications that the facts imply; and
- Explain how the facts lead to the conclusion.

While many arguments may need more than four sentences to fully answer the question, these steps give students a structure to ensure they discuss all major parts of an argument in a potentially concise manner. If students can limit themselves to four sentences and still provide convincing arguments, this would significantly reduce the amount of time it would take to grade papers for a larger class. The next sections unpack each of the four parts.

Answer the question. Stromberg and Ramanathan (1996) claim five reasons why students write poorly. One of these is that students do not read the instructions well enough and consequently do not address the main question asked. How often do we come across a response where the student has argued both sides of the question and never actually manage to tell us which side they believe to be true? It's almost as if they feel if they can hit all of the major points on both sides that they might get some credit for the problem.

Similar to Toulmin's first step of stating a claim (1958), the first step in this structure is for students to write a complete sentence that answers the question given in the prompt. This gives the students a topic sentence to follow and structure their argument on. By forcing students to answer the question in the first line of their response, we know, as readers, the direction the argument should go and we can assess whether the rest of the argument supports the decision the student has stated. According to Cline (2008, p. 401) "good mathematical writing should present an argument so that the reader can easily follow the logic of each step as we work towards the conclusion." This begins with knowing where the argument is going.

The following toy example will be used to demonstrate the different pieces of the four sentence structure:

> Max is interested in the average wing span of a particular type of insect. He randomly selected 35 of these insects and measured their wingspans. When conducting the analysis, he notices that the data appears to be right skewed and he's unsure as to if he is allowed to use the normal distribution to model the data. Is the normal distribution a good model for helping Max estimate the average wing span of the insect? Explain using the four sentence structure.

First we need to answer the question using a complete sentence: Max can use the normal distribution in his analysis.

State relevant facts from the problem. Similar to Toulmin's (1958) second element of argumentation, identifying data to support the claim, the next step for students to follow in the four sentence structure is to find and state the facts from the problem that led them to the option they have chosen. Too often, students try to argue their opinion on a topic instead of researching facts that can support their claim. As a result, students give vague or inaccurate statements (Van Dyke, Malloy, & Stallings, 2015). Stromberg and Ramanathan (1996) believe that students do this because they are not used to writing technical arguments. Every argument of substance should be backed by facts that strongly support the argument one is trying to make.

In addition to the quality of the argument, Grice argues that this also relates to the “quantity of information” suggesting that those presenting an argument should “make your contribution as informative as required” and “do not make your contribution more informative than is required.” (Grice, 1975, p. 45). In addition to providing the facts that support their claim, students should only provide the *relevant* facts. This means that students should not give more information than is required to solve a problem or argue a point. Grice (1975) believes that in doing so writers may be wasting the reader’s time, the information could be confusing or, most importantly, it may raise side issues not intended to be discussed in the argument.

Too often students use a “shotgun” approach in which they write down everything they can think of for a problem in an attempt to hit the mark with at least one of their points. Hayden believes that answers should be short and direct. When grading work for his student he states that “ambiguity or vagueness is taken as a sign of uncertainty and costs points” (Hayden, 1997, p. 204). This is why when stating the facts of the problem, students are told to only give the relevant facts for the problem.

Continuing with our example, the relevant facts that are given to us are that Max took a random sample and it was of size 35. Students may also cite that the distribution of the data was right skewed. Depending on how they focus their argument, this may or may not be important information.

While stating the facts in the problem may not directly demonstrate that a student is thinking statistically, it is important that the student identify the appropriate facts to support the claim they are making. These facts are the foundation of the argument and, provided the student follows the structure, they should lead the student to the implications that the facts imply. As we will see in the next section, the implications phase is where students are, in theory, the most likely to demonstrate their statistical thinking.

***State the implications that the facts imply*.** The third step in the four sentence structure is for students to state the implications that the facts imply. This element does not translate directly to one of the items in Toulmin's model. Instead, this element combines both supporting claims with warrants and providing backing for the warrants (Toulmin, 1958).

In this step, students are required to make the connection between the facts that have been presented to them and the statistical knowledge they should have learned. A third reason why students may write poorly is that their writing follows a method in which they supply a large number of facts, but it is up to the reader to formulate the coherent argument from those facts (Stromberg & Ramanathan, 1996). Gopen and Smith (1990) found that student's early written work in a calculus course often gave a list of facts without any connections to concepts. This may suggest that students are not used to writing implications and need some extra help with this step. It might also be that the author naturally assumes that the reader knows what the facts imply, and believe that it is not necessary to make those steps evident to the reader. This is problematic because the facts of a problem are not enough to prove an argument, especially if the reader does not know what those facts imply. Even if the reader does know what the facts imply, a good argument will still have the author stating why those facts are important to avoid ambiguity in implication (Grice, 1975).

Continuing with our example, a qualified statistician would know that if a person stated: 'a random sample of 35 was taken' then this sample would be large enough to suppose that we could use the normal distribution to model the data. The reason for this, is because, in our minds, we implied the information that a random sample size of 35 means the central limit theorem applies.

However, a non-statistician might not know this. Students may have been taught in class that as long as the sample size is at least 30 or that the sample size is large, then the central limit theorem applies, and if the CLT applies, then the sampling distribution for the average will be approximately normal. However, through experience, we know that many students forget why the sampling distribution is normal and just remember that the normal distribution can be used. This means that by making students state the implications of the facts, we are forcing them to remember more than just how something works, but also why it works, a key factor in demonstrating statistical thinking (Garfield & Ben-Zvi, 2008).

If the student were using the scenario given in our toy example to solve a traditional word problem, we may likely give full credit for the problem as long as the student realizes the normal distribution should be used. If they don't write something like 'by the CLT' on their paper, we wouldn't have any demonstration from the student that they understand why they can use the normal distribution. We might often skip this step in assessing our students, perhaps because in our minds, we can make the leap between the two items very easily and we "knew what the student meant". With written work, the student has the ability to craft a statement to show exactly what they mean. By including this implication step, we are holding our students accountable for implicitly stating the train of thought that lead them to the conclusion, which is what helps us to assess their ability to think statistically.

Explain how the facts lead to the conclusion. This leads us to the final step, which is where students are expected to explain how the facts lead to the conclusion. Depending on how a student has answered the question, this could be similar to providing more backing for the claim, restating the original claim, or it may not translate to any of the elements in Toulmin's argumentation model (1958). In this element of the four sentence structure, students are expected to come full circle in their explanation and again reiterate the decision they have made for the problem.

Going back to our example, we left off with an ideal answer having the student state that a sample size of at least 30 means that the central limit theorem applies. This does not answer the question of whether or not one can use the normal distribution to solve the problem. The student would then have to state how these things all come together.

The following would demonstrate a full solution to the toy problem given above.

Max can use the normal distribution to model the average wingspan of the insects. We are told that he took a random sample of 35 insects. This means that we have a large enough sample that the central limit theorem should apply. When the central limit theorem applies, the sampling distribution of the average wingspan of the insects will be approximately normal and a normal distribution can be used to model the statistic.

Note that the above solution contains the statistical thinking constructs of *reasoning with statistical models* and *integrating the statistical and contextual*. Not all types of statistical thinking will be present in every assignment, and trying to include all of them at once may lead to an overbearing prompt. Thus, multiple assignments may need to be assigned in a semester, each having a different outcome of statistical thinking that will be assessed.

Methods

To show the applicability of some of the sample writing assignments, a qualitative research study was conducted. In this study, students completed written assignments that utilized the four sentence structure, which were then coded for instances of statistical thinking. This section describes the participants as well as the development of the writing assignment used in the study.

Participants. Participants for this study came from a small private women's college in the southeastern United States. During the fall semester of 2016, fifteen students began a second course in statistics with the primary investigator. Fourteen of these students finished the course and volunteered to participate in this study.

Writing Assignments. Throughout the course, students completed a total of ten individual writing assignments utilizing the four sentence structure described above. The general strategy of the writing assignment was to have students make a decision about a scenario using data that was provided for them. To assist students in giving reasonable responses, Chance (1997) believes that students need to be given enough direction to complete the assignments. As such, the instructor worked through the first writing assignment with students and gave them guidance as to how to write the second assignment. Students wrote their responses individually starting with the third assignment.

Most of the assignments necessitated the use of the computer to give a reasonable solution. One reason this was done, was to assess if students recognized the need for data when providing responses to the prompt, one of the five elements of statistical thinking (Wild & Pfannkuch, 1999). Jolliffe (2007) believes that pairing written work with computational output provides a reasonable tool for assessment. While not all students will obtain the same results students could obtain the same output from the computer. When students are expected to interpret the output, we add another layer of differentiation that can be used to assess differences in student understanding.

While the structure of the assignments stayed the same, the purpose of the assignments varied over the semester which provided the opportunity to assess different types of statistical thinking while maintaining a certain consistency for the students. For example, some assignments called for students recognizing the need to use the data provided as opposed to their intuition to create an appropriate model for prediction (American Statistical Association, 2016; American Statistical Association, 2005; Wild & Pfannkuch, 1999) an example of which can be seen in assignment 7 below. Other assignments asked student to determine if a proposed method was valid when, unbeknownst to the students, the method utilized a common misconception in statistics (Rossman & Chance, 2004), an example of which can be seen in assignment 3. Other assignments included creating a simulation to answer the question (Tintle, et al., 2015) or critiquing the work of another individual to determine if that individual had correct thinking (Rossman & Chance, 2004) an example of which can be seen in assignment 4 below.

Each assignment was submitted electronically. Rubrics were developed based on the expected solution strategy to answer the question. They were then used to grade the assignments and feedback was given. In one case, the rubric was amended after seeing solutions from students. Example rubrics and an explanation of the rubric amendment is described more below. Finally, graded assignments were sent back to students electronically.

Writing Assignment Examples

This section will demonstrate two example writing assignments used in the course. The purpose of each assignment and the statistical thinking outcomes that were intended will be discussed. Additionally information will be provided as to when the assignment was given in the course, an intended solution strategy, possible student solutions and misconceptions, and a potential rubric for grading. The remaining writing assignments from the course and any applicable datasets can be found in Appendix A.

Assignment 7 – Weird Statistics: Regressing a Prom Date. *Weird Statistics: Regressing a Prom Date* (Figure 4.1) necessitated students using regression output in non-traditional ways to make decisions about the data. The primary purpose of this assignment was to assess students' use of *context* to solve statistical problems and their recognition of a *need for data* over intuition.

Gary and Wyatt are trying to get dates to prom. They aren't very picky about who they take, but they really don't want to be rejected when they do find someone to ask. They have decided to create a statistical model based on survey data from students within their school to see if they can come up with some criteria for choosing a girl to ask. The data can be found here.

In this survey, the respondent was asked to report their height and also provide an "ideal height" for what they would like in a mate. Gary created a regression model where the survey respondent's height is used as a predictor variable and their ideal mate height was the response variable. The results showed that there was a significant negative linear relationship. He excitedly called Wyatt at 2 in the morning and exclaims "I'm going to ask out a girl who is taller than me, because clearly my model shows that girls ideally prefer guys who are shorter!" Wyatt hurriedly tries to discourage Gary from asking out a taller girl when his mother comes in and makes him get off the phone.

Based on the data, should Gary ask out a girl that is taller than him or shorter than him?

Figure 4.1 Weird Statistics: Regressing a Prom Date

This assignment was given to students at the end of a section discussing multiple linear regression (MLR). While MLR may not have been necessary to solve the problem, this put students in the mindset to think about multiple variables to model the response variable.

If the student was to graph the full data set, they would see the negative relationship that Gary spoke about with Wyatt. Using the *context* of the problem, students might recognize that Gary should only be looking at half of the data set: women describing their preference for their ideal mate. After separating the dataset and redoing the regression analysis, students should find that the regression line is now positive. This data shows us an example of Simpson's Paradox: a trend that occurs when two groups are viewed together disappears or changes when the groups are visualized separately.

Finding that the slope of the regression line is now positive is not the key to solving this problem. Since the slope of our regression line is positive, this tells us that shorter women ideally prefer shorter mates and that taller women ideally prefer taller mates. What Gary could do at this point is input his height in for the response variable and determine the explanatory variable value that would maximize his chances of getting a date, something not traditionally done in a statistical analysis.

However, since we do not know how tall Gary actually is, the student would need to look at the regression line (Figure 4.3) and recognize that at each point on the regression line, the value of the independent variable is less than the value of the dependent variable. This is what informs us that on average, females desire a mate that is taller than them.

An ideal response to this prompt and a rubric for grading (Figure 4.2) follow.

> Gary should ask out a young lady that is shorter than him. When graphing the data and regression line of females only, we get a graph like the one visualized in Figure 4.3. From this graph we can see that on average (at the regression line) the height of the individual is lower than the ideal height of their mate and it can be visualized that most of the data falls above the line y=x. This means that many women prefer mates that are taller than them. Hence, Gary has the best chance of not being rejected (based on height alone) if he were to find a woman that was shorter than him.

Description	Points
All four items of the four sentence structure are present.	4
The context of the data is used to determine that only females should be analyzed.	2
It is decided that Gary should ask out a girl that is shorter than him.	2
A logical reason beyond "the slope is positive" is given to argue that Gary should ask out a girl that is shorter than him.	2

Figure 4.2 Rubric for Weird Statistics: Regressing a Prom Date

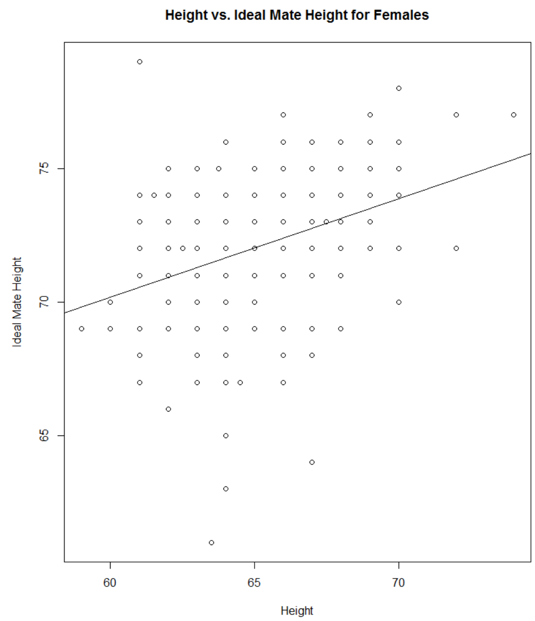


Figure 4.3 Scatterplot and regression line of the female data for Weird Statistics: Regressing a Prom Date

Most of the students in the course were able to identify that the full data set included both males and females stating their preferential mate height and that the responses for the males needed to be removed from the analysis. For example, part of Samantha's response to the prompt was:

> Gary should not ask out a girl that is taller than him… Typically, males are larger than females and would prefer a shorter mate, while females would prefer a larger mate. Likely more males were sampled than females, which would lead Gary to believe that he should ask out a girl that is taller than him… In order to determine who he should ask out based on their preference of mate height, he would need to sample only females so that he would be able to determine what their ideal mate height is so he can see if he fits it…

This solution earned Samantha a score of 7. Even though her conclusion is not given here, Samantha did include all four pieces of the four sentence structure in her full response. However, she gave irrelevant facts for the argument. When she stated that “likely more males were sampled than females” it is obvious that she didn’t use the data she was given to inform her statement as she could determine exactly how many males and females were in the data set, and it is information that is not relevant for the argument she needs to make. This means she earned 3 points for using the structure, 2 points for using the context to realize that only females should be used in the analysis and 2 points for recognizing that Gary needed to ask out a woman that was shorter than him.

While Samantha was not able to make the leap, most students that reached this point in the analysis then stated something to the effect of ‘if the slope of the regression line is positive, then that means that women prefer men that are taller than them’. While the way the prompt was intentionally worded may be leading students to this response, they may also be coming to this conclusion based on their own intuition as opposed to the *using the data* that was given. Provided that students utilized the four sentence structure appropriately, the maximum that these students could have earned on the assignment was an 8.

One way to help students to avoid this notion and to perhaps look deeper into the data that they have analyzed is to mention that the slope of the regression lines for males is also positive. By the student’s interpretation, that means that men want someone that is taller than them as well. The context may then potentially lead them to intuit that their interpretation is not entirely correct.

As it turns out, the student’s conclusions were correct, females do desire taller mates. However, they came to this conclusion based on inaccurate implications of the data and we may not have been able to fully see this if they had not used the four sentence structure to relay their responses.

Assignment 4 – Chi-square Conditions Conundrum. The goal of *Chi-square Conditions Conundrum* (Figure 4.4) was to have students think about the viability of a proposed method.. The statistical thinking outcome to be assessed in *Chi-Square Conditions Conundrum* was students' understanding of *variability* in the data.

This assignment was given to students after a section on methods for analysis for categorical data. This included using the Chi-square distribution for goodness of fit tests and tests for independence.

You are playing a game with two of your friends, Hans and Christian, both of whom are very competitive. To play this game, you toss two dice, add together the values shown and move that many spaces on the board. After 36 turns, Hans has tossed the following values. Note that this data can be found on blackboard in the file called game cheating.csv.

Turn	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
Roll	6	7	8	8	9	9	2	6	3	8	9	11
Turn	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**
Roll	4	7	9	8	6	10	10	7	4	6	12	6
Turn	**25**	**26**	**27**	**28**	**29**	**30**	**31**	**32**	**33**	**34**	**35**	**36**
Roll	9	8	6	7	3	3	2	5	7	9	6	5

At this point Christian feels like Hans has been getting very lucky on where he lands on the board and wants to see if his dice may be rigged. He decides that he is going to do a Chi-square goodness of fit on the data, but notices that one of the conditions for inference isn't met: the count for each cell must be at least 5.

Since the game will be over before there is enough data collected to meet this condition, Christian decides to multiply the counts he currently has by 5 and use that information to conduct the test. Do you agree with Christian's method? Conduct the analysis yourself and explain using the 4 sentence structure.

Figure 4.4 Chi-square Conditions Conundrum

The proposed method has the students (and Christian) multiplying the counts for the data by 5 to ensure that the counts for each cell are at least 5, one of the conditions for inference for a Chi-squared goodness of fit test. If the student agrees that Christian may proceed with the proposed method, they are conceding to the fact that data is not variable. A trained statistician would know that the data collected in this game would converge to a theoretical probability distribution over time. By multiplying the counts of the data given above by 5, we would be preventing the data the opportunity to converge. Consequently, using this method would just amplify the results currently found in the data. Graphing the data before and after transformation would confirm this, as can be seen in Figure 4.5.

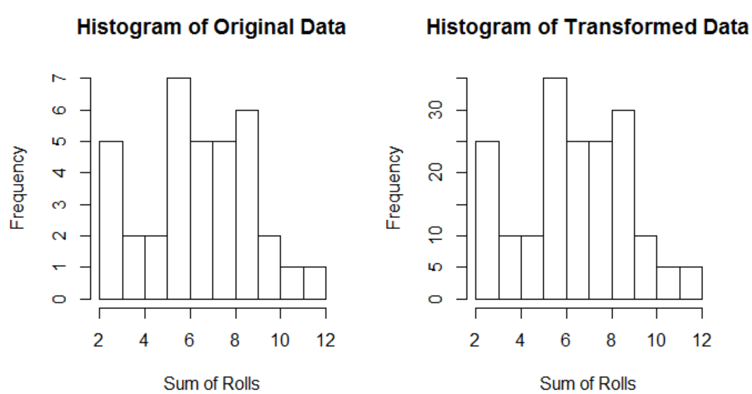


Figure 4.5 Histograms of original and transformed data for Chi-square Conditions Conundrum

An ideal response to this prompt and a rubric for grading (Figure 4.6) follow.

> I do not agree with Christian's proposed method. We are told that Christian plans to multiply the counts he currently has by 5 in order to meet the condition for inference that cell counts are at least 5. We can see from the graphs in Figure 4.5 that this method provides us with data with the exact same distribution as the original data set, although amplified by a factor of 5. Yet, the laws of statistics tell us that this distribution should

converge to some theoretical distribution, which we are not allowing it to do with this method. Since this method prevents our "new" data from being variable compared to the original data set, the given method should not be used.

Description	Points
All four items of the four sentence structure are present.	4
A statement that the method should not be used is given	1
A statement about the lack of variability in the data under the new method, is made.	3
A statement is given or implied about the importance of variability when conducting inference.	2

Figure 4.6 Rubric for Chi-square Conditions Conundrum

Most of the students in the class were able to answer this question correctly. However, a few interesting incorrect solutions were identified. One of the things that was noted was that students believe that we should do anything we can to obtain significant results. Emily and Samantha both stated that multiplying the counts by 5 was a good method, not only because it helped us meet the conditions for inference, but because the results become significant. Even though the student's arguments were completely incorrect, both utilized all four elements of the four sentence structure and earned a score of 4 on the assignment. This writing assignment provided a good opportunity to remind students that while we desire significant results, we cannot manipulate the data to obtain those results.

There was also some confusion as to what was being multiplied by 5. Isabella thought that the sums on the dice were being multiplied by 5, not the counts, and demonstrated this when she stated: "if we have two dice and add them together, the most that our number could be would be 12, but if we multiply the numbers we get by two or five, we get unlikely numbers that we could roll". This was helpful in identifying that Isabella did not understand what was meant when told the conditions for inference. Her solution contained all four elements of the structure and she did make a statement that the method shouldn't be used, earning Isabella 5 points on this assignment.

Discussion and Implications

The use of writing assignments in statistics courses can be beneficial to students, instructors and researchers. Using the four sentence structure as a model provided students with a consistent base to write their arguments. The structure helped students to include essential pieces of an argument that they may have forgotten to include otherwise, and reminded them to avoid using information that was not relevant to the argument they were trying to make (Keys, 1994). These assignments gave students the chance to prove they understood the concepts and that they were not just repeating steps in a formula (Shibli, 1992; Goenner & Snaith, 2003).

Writing assignments were found to be a useful tool for the instructor. In general, as an instructor, I felt like I was more informed on my student's progress and I was able to more easily identify the areas where students were struggling to understand the material (Stromberg & Ramanathan, 1996; Burns, 2004), as was seen above in Isabella's example for *Chi-square Conditions Conundrum*. This allowed me to help students on an individual basis, or if I saw that multiple students were struggling with the same concept, as was seen in *Weird Statistics: Regressing a Prom Date,* we could revisit a topic as a class.

In addition to being able to see topics in which students struggled, the writing assignments also allowed students to be creative in their responses (Radke-Sharpe, 1991). This helped me, as an instructor, to see ways in which students thought statistically, even if it wasn't with the intended type of statistical thinking.

While it wasn't discussed above, in assignment 3 – *Monkeying Around with Models,* the intended strategy for students to solve the problem, was to check the conditions for inference and reason that a two sample t-test would not be appropriate for the given data. However, several students in the course instead reasoned that the original model – a paired differences test – would possess less variability and would be the better model to use. Since students were thinking statistically about the problem and came up with a reasonable solution, a modified rubric was created to accommodate these responses. Instructors that use writing assignments should be prepared to deviate from their initial rubric at times.

By incorporating writing assignments that utilized the four sentence structure, one can ask questions of different forms and assess all types of statistical thinking. Student responses tended to be of a moderate length which made these assignments good for identifying what the students understand without having to spend an exorbitant amount of time grading and giving meaningful feedback (Stromberg & Ramanathan, 1996).

In addition to being a useful tool for assessing statistical thinking, it also appears that the structure may have helped improve student's statistical thinking, their writing skill or both, as the scores on the writing assignments increased on average from the beginning of the semester until the end. In Figures 4.11 and 4.12 we can see graphs of the average score on writing assignments 3 through 10 and individual scores with a line of best fit for the same assignments, respectively. Recall that the first two assignments of the semester were highly structured by the instructor, and scores on these writing assignments may not be indicative of student's ability. Despite some trouble with assignments 6 and 7 we can see a positive trend in student's scores throughout the semester, with several students doing very well on the last three assignments of the semester. However, a formal regression analysis of the data would not be appropriate since student's scores between assignments are likely correlated.

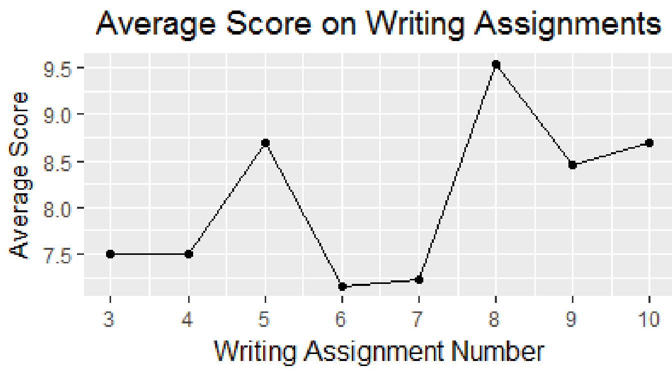


Figure 4.7 Average score on writing assignments 3 through 10

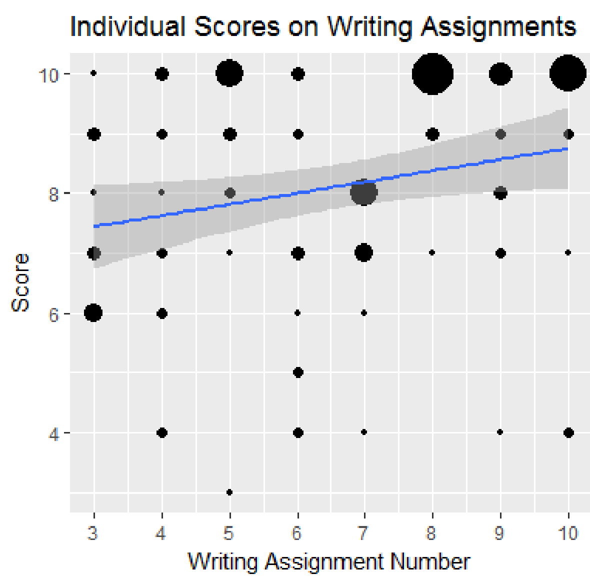


Figure 4.8 Bubble plot of scores for individuals on writing assignments 3 through 10

It is clear from Figure 4.11 that the scores on assignment 7 – *Weird Statistics: Regressing a Prom Date* were low. As was discussed above, on this assignment none of the students were able to give a full solution because they jumped to the conclusion that a positive slope meant that Gary should ask out a shorter prom date. This may have been due to the way the question was worded, or it may be that students were influenced by the context of the data to choose the solution that stated women wanted taller dates. As we can see from Figure 4.12, most students received a score of 7 or 8 on this assignment. As it turns out, by this point in the semester, most students were utilizing the four sentence structure and writing strong arguments. The only reason the average score for this assignment was low was because of students losing points for giving an incorrect reason as to why Gary should ask out a woman that was shorter than himself.

Similarly in assignment 6 – *Professors are out(liers) to get ya*, students implemented inferential procedures before checking the data. In the dataset for this assignment, there was an outlier that was influencing the significance of the regression output, which four of the fourteen students did not see, subsequently earning them scores of 4 or 5 on this assignment and bringing down the class average for this assessment. Despite teaching students to conduct an exploratory data analysis before doing any statistical test, both to understand the story behind the data and to check conditions for inference, it is obvious that some students still did not grasp the importance of utilizing descriptive statistics in an analysis. This assignment helped demonstrate that this was not an isolated incidence and more emphasis should be placed on these items when teaching the course in the future.

While not shown here, utilizing a structured written assessment can be beneficial to those that are conducting statistics education research (Biehler R. , 2007). Specifically, having students utilize the four sentence structure for these assignments aided the process of organizing student's work, which made analyzing 91 written documents more manageable (Lee, et al., 2014). This allowed me to see patterns in the ways in which students think and reason statistically. This is demonstrated in other work to be produced from this research study.

While these assignments may have been difficult for some students, as can be seen by the fact that at the end of the semester, some students were still only earning 4s and 5s on the assignments, other students really seemed to grow in their ability to write about and to think statistically. These assignments also helped me as an instructor to understand the topics that students comprehended, and the topics that they struggled with, at a deeper level than what might have been seen without the structure in place. Finally, this structure was helpful to me as a researcher, as I was able to see in what aspects of writing an argument students used statistical thinking.

Chapter 5 : How Learning Statistical Computing May Impact a Student’s Ability to Think and Reason Statistically

Journal

This manuscript was written to be a report of original empirical research for the Statistics Education Research Journal (SERJ). The purpose of SERJ is to “advance research-based knowledge that can help to improve the teaching, learning and understanding of statistics or probability at all educational levels.” This manuscript does this by providing instructors with information on the relationship between statistical computing and statistical thinking, which they could use when trying to determine how statistical computing technology can be used in their courses. SERJ has a 10,000 word limit.

Abstract

As data sets have become larger and demand for data scientists to analyze that data has become greater, using the computer in the classroom to teach appropriate statistical techniques has grown more popular. As such, there is a greater need for future statisticians to learn statistical computing tools. Students are learning to use the computer to conduct statistical analysis earlier in their educations. As with any technology, this may affect how a student learns the concepts in their field. Through a qualitative study of written work completed by students in a second course in statistics, this empirical research study begins to define the relationship that exists between a student’s ability to think statistically while utilizing statistical computing benefits.

Introduction

With the creation of products such as Fathom (Finzer, 2000), TinkerPlots (Konold, 2007) and StatCrunch (West, Wu, & Heydt, 2004), the integration of technology into statistics classes has become widespread. These "learning tools" (Moore, 1997), are useful for quick calculations, but more importantly, their dynamic nature and ease of use make it so that students can spend more time exploring, visualizing and interacting with data to help them learn the statistical concepts (Moore, 1997). This has recently become a high priority in statistics classrooms trying to adhere to the GAISE standards (American Statistical Association, 2005; American Statistical Association, 2016).

Several researchers have been able to find evidence that using these learning tools are helpful in developing students' statistical thinking and reasoning skills, especially in introductory statistics courses. For example, several studies (delMas, Garfield, & Zieffler, 2014; Garfield, delMas, & Zieffler, 2012; Watson, 2014) have looked at students' ability to understand inference when using a program like TinkerPlots to conduct a simulation. Chance et al. (2007) also claim that these technologies are good for automating calculations, emphasizing data exploration and visualizing abstract concepts, which are all elements that could improve student learning.

Students majoring in STEM fields or minoring in statistics may be required to take statistics courses *beyond* the introductory level. For example, Brieger and Hardin (2012) discuss the need for more sophisticated statistical knowledge in students studying in the health sciences to improve on the quality of analysis currently being conducted in clinical trials. Additionally, Bisgaard (1991) argues that engineering students need a course in statistics that focuses on probability and experimental design, with engineering specific examples presented.

One of the overarching goals of these courses is to prepare students who need to conduct more advanced statistical analysis in their field, as opposed to learning statistics mainly to improve their statistical literacy. Since the goals of the course are different from regular introductory statistics courses, the technology that is used in them may also need to be different. Programs that are often used in professional settings, which Moore has dubbed technology that "does statistics" (Moore, 1997) may be more helpful in preparing students for their future careers. These tools, such as R and SAS, tend to be more powerful and more flexible for large scale statistical analysis, though less user friendly than the tools used for "learning statistics".

Especially with the introduction of the field of Data Science – a field that combines computer science and statistics - one such course that has become more prevalent in statistics curricula is statistical computing (Baumer, 2015; Chance, et al., 2014; Hardin, et al., 2015). This is a course that encompasses the breadth of topics that merge statistics and computer science (Monahan, 2004). Users must be comfortable using a wide array of software tools - and perhaps hardware – while understanding programming and logic to fully complete the spectrum of statistical tasks they need to perform.

While there has been a push in the research to implement such courses in undergraduate curricula (Gentle, 2004; Hardin, et al., 2015; Nolan & Temple Lang, 2010) the research on how students learn using programs that are designed for doing statistics has not kept up with the research on how students learn using programs that are designed for learning statistics. Specifically, little research has been done to document what affordances and constraints statistical computing technologies might have on a student's ability to think and reason statistically.

One way that we may gain a better understanding about students' reasoning is through student's written work. Several researchers (Park, Park, Lee, & Lee, 2016; Parke, 2008; Peck, 2005; Weldon, 2007; Lee, et al., 2014), have found benefits for both learning and assessing statistical reasoning by having their students communicate their understanding of a problem by writing responses to questions. For example Parke claims that written work provides students with an environment that helps "students develop deeper understandings of concepts and increase their level of comfort in talking about statistics" (Parke, 2008).

Keys (1994) found that scaffolding written reports for students helped them to focus their thinking on pertinent information. A popular example of structuring a response comes from Toulmin's (1958) model for argumentation, originally used to aid in creating verbal arguments for ill-structured problems (Karbach, 1987). Inglis, Mejia-Ramos and Simpson (2007) discuss several studies in which college mathematics students were expected to use Toulmin's model, or a modification of it, in order to create well defined mathematical arguments.

Toulmin's model for argumentation directs users to create persuasive arguments by:

- Stating a *claim*;
- Identifying the *data* that supports the claim; and
- Backing the claim with *warrants*

If needed, the argument can also be enhanced by:

- Providing *backing* for the warrants;
- Identifying *qualifiers* to support the strength of the relationships; and
- Providing a *rebuttal* indicating circumstances where the argument may not be true (Toulmin, 1958).

To assist introductory statistics students in writing complete arguments and help them to better convey their reasoning, Woodard (2016) suggested teaching students a similar four element model, dubbed the four sentence structure. The four sentence structure directs students to:

- Answer the question;
- State relevant facts from the problem;
- State the implications that the facts imply; and
- Explain how the facts lead to the conclusion.

In relation to Toulmin's model, answering the question is similar to stating the claim and stating relevant facts is similar to identifying data to support the claim. Stating the implications that the facts imply doesn't translate directly to one of the items in Toulmin's model. Instead this element combines both supporting claims with warrants and providing backing for the warrants. Finally, depending on how a students has answered the question, explaining how the facts lead to the conclusion could be similar to providing more backing for the claim, restating the original claim or it may not translate to any of the elements in Toulmin's argumentation model (1958).

What This Means for Statistics Educators. The expectation for statistics majors and minors to learn computing tools in their curriculum has recently grown (Chance, et al., 2014; Horton, Baumer, & Wickham, 2014; Nolan & Temple Lang, 2010). Within statistics courses, students are learning data cleaning, graphical displays and multiple techniques to aid them in their analyses (Hardin, et al., 2015). It is not enough to expect students to teach themselves the majority of the technology while the instructor teaches only the statistical concepts. If students are forced to learn the technology on their own, they will focus on getting the answers to the problems they have been assigned and often pick up bad programming habits due to the lack of instruction and direction (Nolan & Temple Lang, 2010). Conversely, focusing too much on how one uses the technology to conduct statistical tests may teach the students that it is not important to explore the data before conducting formal inference procedures. Thus, it is important for teachers to find the appropriate balance between teaching the statistics and teaching the technology that is used to conduct the analysis.

Despite the cautions of the ASA about including programs like R and SAS in introductory statistics courses (American Statistical Association, 2016), some students are able to successfully learn these higher level languages while still grasping the concepts traditionally taught in introductory statistics courses. Horton and his colleagues were able to successfully introduce the statistical language R into their introductory statistics courses. Using RStudio and the MOSAIC platform, their students were able to perform higher level statistical processes with a more difficult technology earlier in their statistical careers. They do note however, that in teaching the statistical computing ideas, there should be more examples and activities, but less content taught overall (Horton, Baumer, & Wickham, 2014; Kaplan, Horton, & Pruim, 2015).

Additionally, teaching statistical computing topics can been seen as a chicken and the egg dilemma. Using technology to explore data and analyze data is one of the six recommendations of the GAISE standards for introductory statistics courses (American Statistical Association, 2005; American Statistical Association, 2016). However, the same reports also caution about introducing higher level technologies - those normally seen in statistical computing courses - too early. While these tools are powerful, they may confuse students and may be more of a hindrance than help. So which should be taught first, statistical concepts or the statistical computing technology; and how do we account for the negative effects of our choice? An answer to these questions may be found if we can determine if and how a students' use of statistical computing technology is related to how students think and reason statistically.

Thus, our research aim is to develop a better understanding of the ways in which statistical computing affects a student's ability to think and reason statistically, and how a student's ability to think and reason statistically may affect their use of statistical computing to solve statistical problems.

Guiding Framework

To determine how statistical computing and thinking work with and against each other, a framework was developed to identify statistical thinking and computing in students' work.

The first part of this framework, which identifies statistical thinking in student's work, was developed based on the ideas of Wild and Pfannkuch (1999) where they discuss five elements that define statistical thinking. These include recognizing the need for data, transnumeration, considerations of variation, reasoning with statistical models and integrating the statistical and contextual. With additions and adaptations from supplemental articles (Ben-Zvi, 2004; Ben-Zvi & Garfield, 2004; American Statistical Association, 2016; Chance, 2002; Chance, delMas, & Garfield, 2004; delMas, 2004; Garfield & Ben-Zvi, 2008; Lee, et al., 2014; Makar & Rubin, 2009; McCullagh, 2002; Moore, 1990; Pfannkuch & Wild, 2004), the first part of the framework in Table 5.1 describes how statistical thinking will be identified in student work.

Unlike statistical thinking, research on the benefits of statistical computing has not been well defined. Current research focuses heavily on the topics that should be taught in a statistical computing course (Dierker, Kaparakis, Rose, Selya, & Beveridge, 2012; Gentleman, 2004; Hardin, et al., 2015; Monahan, 2004; Nolan & Temple Lang, 2010), as opposed to the benefits of being able to use statistical computing in one's work. While individual benefits of statistical computing have been discussed in other articles (Ainley, Gould, & Pratt, 2015; Baglin, 2013; Biehler, Ben-Zvi, Bakker, & Makar, 2012; Brown & Kass, 2009; Garfield, delMas, & Zieffler, 2012; Nolan & Temple Lang, 2010; Pfannkuch, et al., 2011; Wing, 2008; Wolfram, 2016), a succinct collection of these benefits was not found. As such, a new framework was created to identify statistical computing in students' work. The second part of the framework defined in Table 5.1 describes how the benefits of statistical computing were identified in student work.

Table 5.1 Framework for identifying statistical thinking

Aspect of Statistical Thinking	**Affordance for the student**
Recognizing the need for data	• gives a statement about the data in the interpretation of the results • is able to identify if proper data collection techniques are used • recognizes that anecdotal evidence is not enough to conduct statistical inference
Transnumeration	• is able to think about the variables in the study and is curious of ways to represent them • uses exploratory data analysis tools for exploration, and not just prescribed methods • uses multiple representations to interpret and make sense of the data
Considerations of variation	• is conscientious that all data is variable • is able to redesign an experiment to reduce or eliminate variability • is able to quantify and explain variability in a way that helps them understand statistical inference
Reasoning with statistical models	• knows which type of model to use for specific types of data • knows when a specific model is appropriate to use for a given data set • understands theories underlying statistical models
Integrating statistical and contextual	• is able to plan data collection and analysis based on contextual knowledge of the data • is able to make more reasonable conclusions using the context of the data. • avoids unnecessary statistical jargon in their interpretations of the data

Table 5.2. Framework for identifying statistical computing

Aspect of Statistical Computing	**Affordance for the student**
Automate computational procedures	• creates multiple graphs or summaries to make sense of the data • uses the technology to conduct the analysis, then makes proper inferences
Increase computational thinking	• creates a solution strategy and can communicate it to the computer • is able to think critically or abstractly about a concept that the technology is demonstrating
Offer new methods to explore concepts	• does not follow a prescribed set of procedures to answer a question • comes up with their own methods and reasons for using the technology in analysis
Aid in pattern recognition and decision making	• recognizes patterns, either in the way things are coded or in how the results behave • uses the output from the technology to influence their decision of where to go next in the analysis

It is believed that there is a symbiotic relationship between a student's ability to think and reason statistically and their use of statistical computing benefits (Figure 5.1). This paper will explore this relationship in more detail and describe some potential implications of relationships that have been observed.

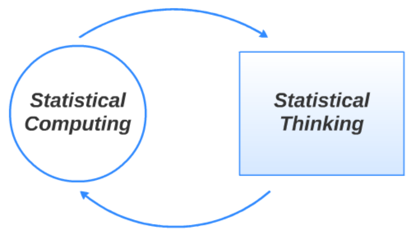


Figure 5.1 Relationship between statistical thinking and statistical computing

Methods

To determine how statistical thinking and statistical computing impact each other, a small-scale, qualitative research study was conducted. According to Groth (2010), if done appropriately, qualitative research can be scientific, rigorous, generalizable, and objective and can "provide a means for conducting scientific studies that contribute to our understanding of statistics teaching and learning".

Participants. Participants for this study came from a small private women's college in the southeastern United States. During the fall semester of 2016, fifteen students began a second course in statistics with the primary investigator. Fourteen of these students finished the course and volunteered to participate in this study.

Statistics 2 is a continuation of introductory statistics. It includes topics of one and two sample inference, two way tables, simple and multiple linear regression and one and two factor ANOVA. Students were also taught introductory statistical computing practices in R while taking this course.

While the only prerequisite course for statistics 2 was introductory statistics, the statistical and computational backgrounds of students did vary. Some students had only taken the introductory statistics course, while some had taken a course in probability theory or statistical computing. Six students had experience with programming before taking the course, through a statistical computing course offered the previous semester. In addition, one participant was majoring in computer science.

While there was variation in the student's backgrounds, the goal of this research was to look at the work of students holistically, not to look at their individual differences. As such, specific differences between students who had statistical computing experience and those who did not will not be discussed in this work.

Data Sources. Data for this project was collected during the fall semester of 2016. This data took the form of field notes and writing assignments.

Field Notes. Field notes were kept so that there was a record of conversations, where students discussed with their classmates or with the instructor, ideas that demonstrate statistical computing or statistical thinking. For example, in-class group work was one of the modes of instruction in the course. To find appropriate solutions for many of the group-work problems, students needed to use the computer and were encouraged to share their thinking with their group mates. As the instructor walked around the classroom assisting groups, she took notes about the student's discussions. In addition, when students came to the instructor's office for additional help, notes were taken after they left that may have been pertinent to the research.

Writing Assignments. The primary source of data for this research were writing assignments. Throughout the course, students completed a total of eleven individual writing assignments utilizing the four sentence structure described above (Woodard, 2016). Eight of the eleven assignments were used for the purposes of this research. These assignments gave students the opportunity to demonstrate their statistical computing, while encouraging them to explain their statistical thinking and reasoning. The three assignments not being used for research were necessary for assessment in the course but did not give the students a potential medium for demonstrating statistical thinking and statistical computing. In total 91 documents were collected from students. The writing assignments used for analysis, along with all assignments used in the course can be found in the Appendix A.

The general strategy of the writing assignment was to have students make a decision about a scenario using data that was provided for them, however, the opportunities for statistical thinking and computing within the assignments did vary. For example, some assignments called for students recognizing the need to use the data provided as opposed to their intuition to create an appropriate model for prediction (American Statistical Association, 2005; American Statistical Association, 2016; Wild & Pfannkuch, 1999). Other assignments asked student to determine if a proposed method was valid when, unbeknownst to the students, the method utilized a common misconception in statistics (Rossman & Chance, 2004). Other assignments included creating a simulation to answer the question (Tintle, et al., 2015) and critiquing the work of another individual to determine if that individual had correct thinking (Rossman & Chance, 2004).

In order to make an appropriate decision and justify that decision, students were expected to use the computer to support their argument. Using technology to help create an argument was a necessary but not a sufficient condition to assess whether the student was seen as doing statistical computing. It should be noted that since about half of the students in the course had not taken statistical computing prior, the level of required computer use on each of the writing assignments, especially towards the beginning of the course, needed to be somewhat limited. To ensure that every student in the course could complete the writing assignments, yet still collect data that could be used for this research, these assignments were purposefully designed to allow students to be engaged in the problem no matter their level of understanding of computing. Students were encouraged to use the technology for more than just computational purposes, however this was not required for the student to get credit in the course for the writing assignments.

Overview of Analysis

Recall the purpose of this research was to determine how statistical computing and thinking influence each other. The first step to doing this was to blind the documents. Each document was formatted identically and student's names were replaced with reference codes. After analysis, these reference codes were replaced with pseudonyms.

Next, the 91 documents were partitioned into logical groupings that showed different cycles of thoughts the students were having while writing. This was done in similarity to Lee et al. (2014). In their article, they partitioned student work using the four phases of a statistical investigation, similarly defined by Wild and Pfannkuch (1999): choose focus, represent and explore data, analyze, and interpret data and make decisions. Since students in this study were asked to write their responses using the four sentence structure, this was used to partition the documents. An example of how a document was color-coded to indicate the partitioning can be seen on the left of Figure 5.2. Only three colors were used to partition the assignments. The first and fourth pieces of the four sentence structure essentially asks students to decide which option they support and were both color-coded in red. The second part of the structure asked students to state the *Facts* in the problem. These statements were given blue text. Finally students were to explain the *Implications* that the facts imply. These statements were given orange text.

Once the assignments were segmented into their color-coded partitions, instances of statistical thinking and computing were identified in the students' documents, using the categories in the framework in Table 5.1. Whole partitions of the document were used when assigning the statistical thinking and statistical computing codes defined in the framework, to demonstrate the co-occurrence of these constructs. (Note, the term *code* is used here in the qualitative analysis sense, not the programming sense). This included instances where the student had correct and incorrect work for both statistical thinking and statistical computing. Thus, it became necessary to identify in the coding process whether a students' work was correct or incorrect. An example of the color-coding scheme used to do this can be seen on the right of Figure 5.2. Green codes were assigned to correct statistical thinking, red codes were assigned to incorrect statistical thinking, blue codes were assigned to correct statistical computing, and yellow codes (not shown in Figure 5.2) were assigned to incorrect statistical computing. For example, as shown on the right side of Figure 5.2, the student was correctly using context and modeling as part of their statistical thinking (green) and correctly engaging in pattern recognition using statistical computing (blue), but also incorrectly used modeling and context during parts of their response (red).

Finally if one partition in the document appeared to link directly to another partition in the document, these connections were noted. This can also be seen on the right of Figure 5.2. The reader will note the use of the term "causes" in these connections was used to identify a potential relationship between two codes and may better be understood as "led to".

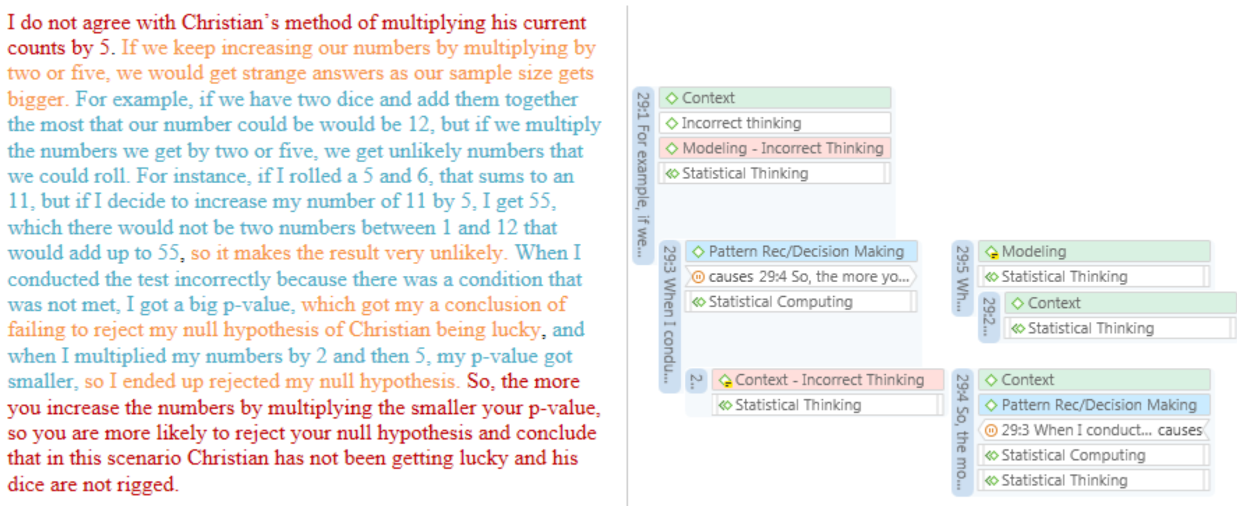


Figure 5.2 Example document from writing assignment 4. Shows partitioning, color-coding and connections made in the document

Once the 91 documents of students' work had been coded for instances of statistical thinking and computing and the connections between them, the documents were then organized using document diagrams similar to Lee and Hollebrands (2006). In their work, which was influenced by Schoenfeld (1985), the authors segmented and classified student's work into 6 categories to help visualize the type of goal students had, and whether the technology was supporting or not supporting their goal directly. For the purposes of this research, the student's work was segmented into four sections corresponding to the elements of the four sentence structure. This was done to diagram the structure of student's arguments, including if and when the student was using statistical computing or thinking to support their argument, and which types of statistical computing or thinking they were using. Colors and lettered codes were assigned based on the process described above. As an example of the document diagram structure for the document shown in Figure 5.2, see Figure 5.3.

In Figure 5.3 a black tile means that the student gave a statement in that segment of text, but it did not demonstrate any statistical thinking or computing. The other colors that are shown on the document diagram match the colors that were used in the qualitative analysis software (green/red for correct/incorrect statistical thinking and blue/yellow for correct/incorrect statistical computing). When two colors and codes are shown together, it means that they were identified in the document for the same segment of text. For example, the first *Fact* in the document in Figure 5.3 was identified as both green and red, as well as a C and an M. Since green and C were first, this means the student correctly integrated the contextual and statistical into their argument. The red and M are represented second, meaning the student reasoned incorrectly about statistical models.

The letters C, M, V, D and T were used with green and red blocks to identify the statistical thinking codes of context, modeling, variation, need for data and transnumeration, respectively. The letters A, C, N and P were used with blue and yellow blocks to identify the statistical computing codes of automation of computational procedures, computational thinking, new methods and pattern recognition/decision making, respectively. Once all 91 documents were represented with these document diagrams, they were printed so they could be arranged side-by-side to identify themes across all student's work.

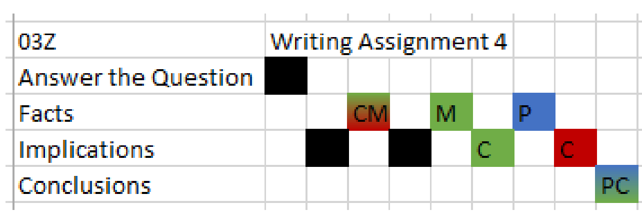


Figure 5.3 Example of a document diagram for the four sentence structure and use of statistical thinking and computing in a student's document

The following steps describe the process that was completed to identify themes:

- Separated out the 23 document diagrams that had no blue or yellow blocks (no statistical computing). These were then subcategorized into two groups
    - Documents from assignment 3

        - Other writing assignments
- Noted that most students had not used any statistical thinking or statistical computing in the A*nswer the Question* partition and subsequently sorted remaining students by their first *Facts* partition that included statistical thinking, statistical computing or both. These subcategories included:
    - Correct statistical computing only
    - Correct statistical thinking only
    - Correct statistical computing with correct statistical thinking
    - Correct statistical computing with incorrect statistical thinking
    - Incorrect statistical computing with incorrect statistical thinking
- Identified the 30 document diagrams that had only black, green and blue (correct statistical thinking and computing). No subthemes were identified in these documents.
- Tried to identify patterns within each of the sub bullets in step 2 and noticed that statistical computing was not used in the *Implications* of any of the document diagrams.
- Identified the 10 documents that had incorrect statistical computing. These were subcategorized into
    - Having correct programming but an incorrect strategy
    - Having a correct strategy but incorrect programming
    - Having both incorrect programming and an incorrect strategy
- Collapsed the remaining 28 documents into one group that consisted of arguments with correct statistical computing with correct and incorrect statistical thinking. These were subcategorized by the type of incorrect statistical thinking identified. This included all five types of statistical thinking, as well as a misunderstandings of basic statistical facts.

Results

The detailed qualitative analysis process resulted in five themes about how students' illustrated statistical thinking and statistical computing in their written work. These themes will be used to organize the presentation of results:

1. Students' written work does not always provide evidence of statistical computing.
2. Students' arguments use statistical computing to provide *Facts*, not *Implications*.

3. Students' written arguments may demonstrate correct statistical thinking and statistical computing.
4. Students sometimes apply incorrect methods or have incorrect computational thinking in their evidence for arguments.
5. Students' correct application of statistical computing does not always lead to correct statistical thinking due to pitfalls in student's work.

Each theme is explained, and then examples from students' work are used to illustrate the finding.

Student Work may not contain any statistical computing. The first theme that was noted in the student document diagrams was that some arguments did not appear to utilize statistical computing. Ten of the twenty-three documents that fell into this theme came from the third writing assignment given to students. This also happened to be the first writing assignment students would complete that required the use of data provided to them.

In this assignment, students were to determine if using a two sample t-test could be used instead of a paired t-test, for a data set that had only 36 subjects and response variables that did not come from normally distributed populations. Of the ten students who did not use computing to answer this question, nine gave incorrect responses to the prompt. Several of these students reasoned statistically about why a paired t-test was better than a two sample t-test, but did not use the computer to support implicitly why the two-sample t-test should not be used.

For twelve of the remaining thirteen documents it was obvious from the students' writing that they had utilized the computer to come to a solution, but using the definitions in the framework led to categorizing the responses as transnumeration, part of their statistical thinking activities. For example, in assignment 5, students were to use data about Hondas and Toyotas to find the best model for predicting the price of the cars. In Isabella's argument, she provided several summary statistics for different models to make a decision on the best model for predicting price. However, in her response she gave no indication that the computer was used to find these values, so they were coded only as transnumeration. While they could have been found by hand, it could probably be inferred that the student used the software to find these values.

Statistical computing provides facts, not implications. The second theme identified is different from the other four in that it spans all of the documents, as opposed to classifying the documents into mutually exclusive groups. Specifically it was noted that statistical computing was not identified within the *Implication* segment in any of the 91 arguments. It was identified in the *Conclusion* of two arguments, but every other instance it was found to represent the *Facts* in the argument. We can see an example of this in the document diagram shown in Figure 5.3. Thus, it appears that students' rely on their statistical computing when presenting facts that support their argument, and do not explicitly refer to statistical computing activities when stating final implications from their work.

Sometimes student work had correct statistical thinking and statistical computing only. Statistical thinking and computing are highly integrated when a student uses the computer to analyze data and make decisions. Thirty of the ninety-one documents, spanning all of the writing assignments, were found to have only correct thinking and computing assessed in their work, with nineteen of these documents coming from students who had taken statistical computing. It did not matter if students used the computer first, gave a statistical thinking justification first, or a mixture of both. These arguments tended to utilize several thinking constructs integrated with statistical computing. Below is an example of this type of solution.

In assignment 4 students were to determine if, based on the dice rolls in a game, an individual named Hans was cheating. The intended strategy for solving this problem would be to see if the dice rolls conformed to the theoretical distribution for the sum of two dice, using a goodness of fit test. However, the data set they had been given did not meet the conditions for inference for a Chi-square goodness of fit test. Specifically, the observed cell counts were not at least 5. Students were asked if multiplying the observed cell counts by 5 to satisfy the condition would be an appropriate solution to this problem.

In her argument against this proposed strategy, Natalie stated that the more times the dice are rolled, the more normal the data should become (meaning the sums that are observed in the data set). To demonstrate that this was not happening, she used the software to create a histogram of the cell counts before and after they were multiplied by 5. She noted that the summations from the original data had the same ratio as the data after being multiplied by 5, hence there was no variation as expected. She used the lack of variation in the data to help her decided against the validity of using the proposed method. She noted that if the results from the proposed method were to be used that we could state that Hans was cheating, however the results are not reliable.

This example demonstrates a complex integration of several statistical thinking constructs utilized with automation of computational procedures. First, Natalie stated that the summations when graphed in a histogram should become more normal. While the shape of the histogram does become more symmetric and somewhat bell shaped, the student is misusing the term normal here to categorize what the distribution should look like. In actuality, the data should converge to the theoretical probability distribution, which is what the student's wording was taken as. While her terminology is incorrect, this student is demonstrating reasoning with statistical models.

The student then used the computer to multiply the counts in the data by 5 and create a new histogram to visualize the data. Because her intent was to speed up the process of graphing the proposed method's distribution and make sense of the data, this was seen as both automated procedure and transnumeration occurring in tandem.

From the output, Natalie then noted that there was no variation in the distributions the two histograms were portraying. If more data had been collected in the experiment, we would not expect the distributions to be exactly the same. The one with more data values should start to converge on the theoretical distribution. This demonstrates the student's understanding for variation in data.

Finally when reporting the results of the study, if the new method was to be used, the conclusion we should come to was that Hans was cheating. She implied that she didn't actually believe this to be true since the results of the analysis were not reliable. Hence, she stated her conclusion in the context of the problem without complicating her solution using statistical jargon.

In this argument, Natalie used the computer to present her with the *Facts* necessary to answer the question correctly, yet the computer alone was not enough to give a reasonable solution. She needed to integrate her statistical knowledge to produce a logical argument. In particular, automation of computational procedures was seen here being sparked by the statistical modeling knowledge of the student, it occurred with transnumeration and it led to considerations of variation and integrating the statistical and contextual.

While all arguments found in this theme may not have been this detailed, they all shared the commonality that statistical computing and thinking were used in tandem to help form logical arguments. Automation of computational procedures was by far the most commonly used statistical computing code, followed by computational thinking. While pattern recognition and decision making, and new methods were also seen, they were not nearly as common.

Connections between thinking and computing varied depending on the student and the argument. A summary of these findings include that

- A need for data and reasoning with statistical models were the *precursors to* the students using statistical computing.

- Statistical computing techniques were found to *occur with* each of the statistical thinking constructs except for need for data, though not normally with more than one at a time. Table 5.2 shows the number of co-occurrences observed of correct statistical thinking and statistical computing within each partition, across all 91 documents. This highlights that automation of computational procedures often occurred with statistical thinking codes, with reasoning for statistical models being the most prominent co-identified code.
- Statistical computing generally *led to Implications* involving the statistical thinking codes of reasoning with statistical models, integrating the statistical and contextual and considerations of variation.

A visualization of this is depicted in Figure 5.4.

Table 5.3 Number of co-occurrences between statistical thinking and statistical computing codes in ninety-one documents

	Automation	Computational Thinking	Pattern Recognition/ Decision Making	New Methods
Modeling	7	0	1	0
Context	3	2	1	0
Transnumeration	4	1	0	2
Variation	2	3	1	1

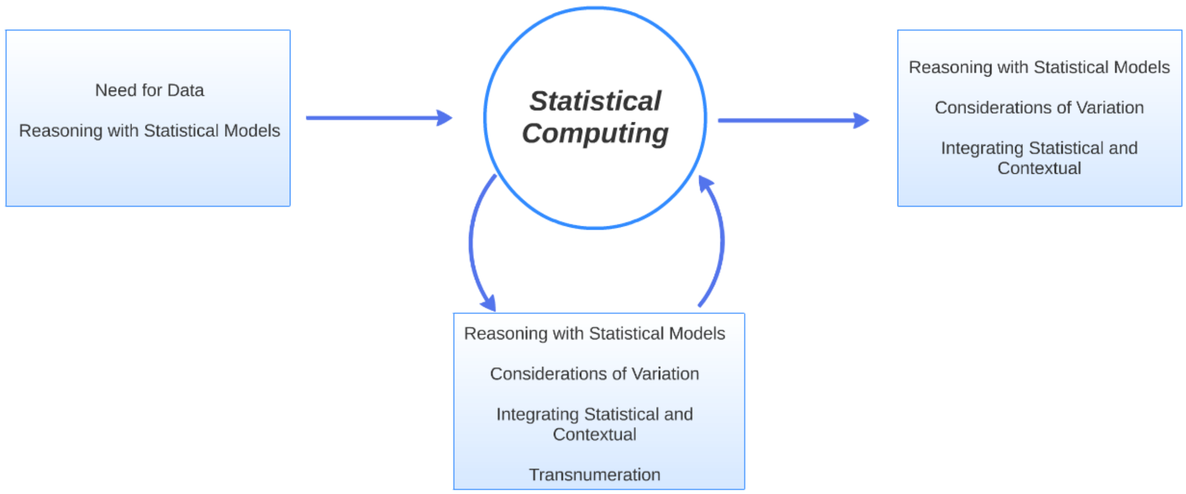


Figure 5.4 Relationship between statistical thinking and statistical computing when correct thinking and computing are utilized

Sometimes student work had incorrect statistical computing. While good statistical computing and thinking led to good arguments, when students have poor computing skills, their arguments are much weaker and tend to be inaccurate. Poor statistical computing was only observed in ten student documents, with eight of these documents coming from students who did not take the statistical computing course. Specifically in these assignments, the statistical computing errors that were observed were identified as errors in computational thinking and new methods. These ten documents were categorized by the following groupings:

1. Students were able to program their work but were unable to properly explain the strategy that they used. (4 documents)
2. The students had a good strategy for completing the problem, but had incorrect programming. (4 documents)
3. The student had a poor strategy and did not know how to program the problem. (2 documents)

Correct programming, incorrect strategy. Student work in this category had programming that was correct but stated strategies that were incorrect. The three documents that fell into this category had mostly correct arguments with the exception of the strategy the student thought they were conducting on the computer. An example student document is described below.

Recall that in assignment 4, students needed to determine if multiplying the data counts in a data set would provide accurate inference for assessing if one of the players had cheated. As an optional bonus follow up assignment, students were asked to program a simulation, to determine the likelihood of the data that was originally observed in the problem. In an ideal solution strategy, students would sum two randomly selected values between one and six, 36 times to simulate 36 rolls of the dice. They would use that data to create Chi-squared values, comparing the simulated data to the expected theoretical distribution of rolling two fair six-sided dice. Finally, they should compare the chi-square value they had found from the original data to the chi-squared values they had produced through simulation and determine the likelihood of the original chi-squared value.

One student, Cora, found correct results for the simulation and was able to give a reasonable argument to the prompt. However in describing the strategy that she used in her simulation to find her results she stated:

> I produced a simulation that produces 36 sample chi squared values, the same number that Hans rolled. Then I did 10,000 iterations of these values.

Notice that Cora believes she is creating 36 chi-squared values, instead of 36 tosses of the dice and one corresponding chi-squared value. In actuality, when reviewing her program, it was found that she created 36 sums, to which she created a single chi-squared value, repeating the process 10,000 times. This means that her program was correct, yet she was unable to describe what the program was actually doing.

Correct strategy, incorrect programming. The second type of problem that was observed when a document was coded for incorrect statistical computing was when a student had a good strategy for solving a problem but was unable to program the problem as anticipated. This produced two results in the student documents. In the first case, the student was lucky enough to get a wrong solution that was similar to the right solution, and therefore came up with a reasonable argument to the prompt. The other scenario was if the student came up with an incorrect and imprecise solution, which resulted in an incorrect argument.

As an example, in assignment 6, students were given a data set that included the height of several individuals (in inches) and their GPAs. The data set included an outlier for an individual that had incorrectly reported their height as 6.5 inches. To give an appropriate response to this prompt, students would have had to explore the data to see that there was an outlier, remove it from the data set (or clean it if they assumed the person to be 6.5 feet tall) then rerun the regression analysis to see if the relationship between the two variables was strong enough to suggest that height influences one's GPA.

Samantha had developed a proper strategy to see if the proposed relationship existed; however, she had some difficulty in programming the proposed strategy. She first graphed the data for height and GPA, then added the regression line to the plot. However, in creating her line of "best fit", she produced a regression line that did not accurately fit the data. Reproducing these results has been difficult, so an accurate description of what she actually did cannot be given. Although based on her response to the prompt, she believed that the line she created was correct, indicating that she had an error in her program that she did not recognize. Instead of Samantha trying to determine computationally why the regression line did not fit the data, she tried to explain statistically why she obtained her graphical results (Figure 5.5).

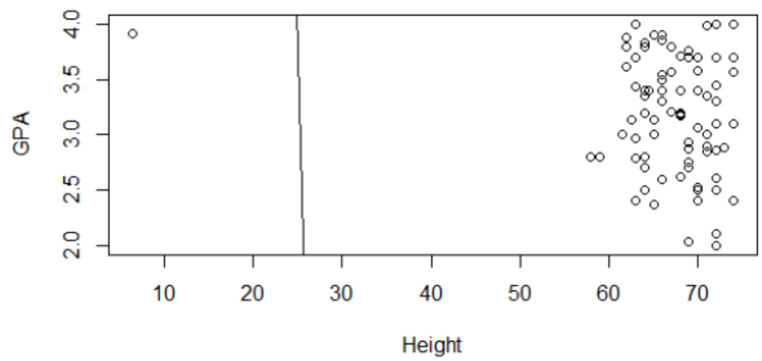


Figure 5.5 Demonstration of a correct strategy but incorrect programming

Her justification of how the outlier influenced the regression line was that it

> drags the regression line to the left quite a bit. An outlier this extremely (sic) can affect a number of values including the slope and intercept, as well as the p-value. Although it can affect the p-value, in this situation, excluding the outlier would strengthen the p-value and would still suggest that there is a relationship between height and GPA

While there is some truth to her justification, such as outliers can affect the slope, intercept and p-value, these are not the reasons why her regression line is much different from the expected line. Removing the outlier also would not strengthen the p-value in this particular case.

Incorrect programming, incorrect strategy. The third type of error that was observed, was that students were not able to propose an accurate strategy to solve the problem, and therefore also had incorrect programming. This was observed in two documents. Since both of these students tried something new, not only was their work categorized as computational thinking, but also applying new methods. The students in this group attempted to use the internet as a source to help them understand how to program their work. Since they had no initial strategy to fall back on, they trusted the procedure they found and tried to justify (statistically) the results they observed in their computer output.

Moira did this in assignment 10 when she was asked to use a simulation to determine the effects of the Bonferroni correction on the type I error rate. Specifically, students were to determine if the correction gave a type I error rate exactly at, or below the significance level. In the software (R), Moira used the following procedure in her simulation: p.adjust(0.05, "bonferroni", n=1000). In trying to justify the results that the software gave her, she stated:

> we adjust our p-value to the "Bonferroni" method, for 1000 iterations. The output tells us that p (our overall type I error rate) is now 1 or very close to one. This means that if our initial p value was 0.5 to begin with, our rate of making a type 1 error is very close to one or 100%.

In actuality, what this procedure has done is adjust the given significance level using the Bonferroni correction with 1000 comparisons. It finds the minimum value of 1 and the number of comparisons multiplied by the original significance level, α. This is why she found a value of 1 when she ran her program. However, since she did not understand what the program was supposed to be doing, she came up with a reason for getting the results that she did that do not match what the statistical theory states she should get.

The above examples illustrate how incorrect statistical computing can impact a student's ability to think statistically. Only incorrect computational thinking and new methods were observed in student's work. These statistical computing constructs were not *proceeded by* any statistical thinking constructs, hence a lack of thinking about the problem statistically before using the computer may have led to the students' incorrect statistical computing.

Incorrect statistical computing only *co-occurred with* incorrect transnumeration, consideration of variation and reasoning with statistical models, with modeling being the most prominent co-occurring statistical thinking code. Hence if a student was trying to think and reason about the problem statistically, their flawed thinking may have been a contributor to the student's incorrect statistical computing.

Incorrect statistical computing *led to* both correct and incorrect statistical thinking. As observed above with Cora, despite having stating an incorrect programming strategy, she still obtained the correct output and created a logical argument based on that output. Additionally despite Samantha having incorrect output and obtaining a final solution that was incorrect, the student had a good strategy to solve the problem and was able to think statistically about the output she had obtained. When a student's incorrect statistical computing did lead to incorrect statistical thinking, it was observed from those that were missing both the ability to program and form a strategy for the problem.

The difference between correct and incorrect statistical thinking as a product of incorrect computing, seems to depend on the level of error performed by the student. If they had a small error in their work (either a programming error, or a strategy error, but not both) they tended to produce correct statistical thinking. If the student had larger errors in their work, such as having inaccuracies with both programming and strategy, then they were unable to think about the problem statistically. This suggests that having the sophistication to even partially think computationally helps improve students' solutions and may mean that they have stronger statistical thinking skills. The diagram in Figure 5.6 summarizes these results.

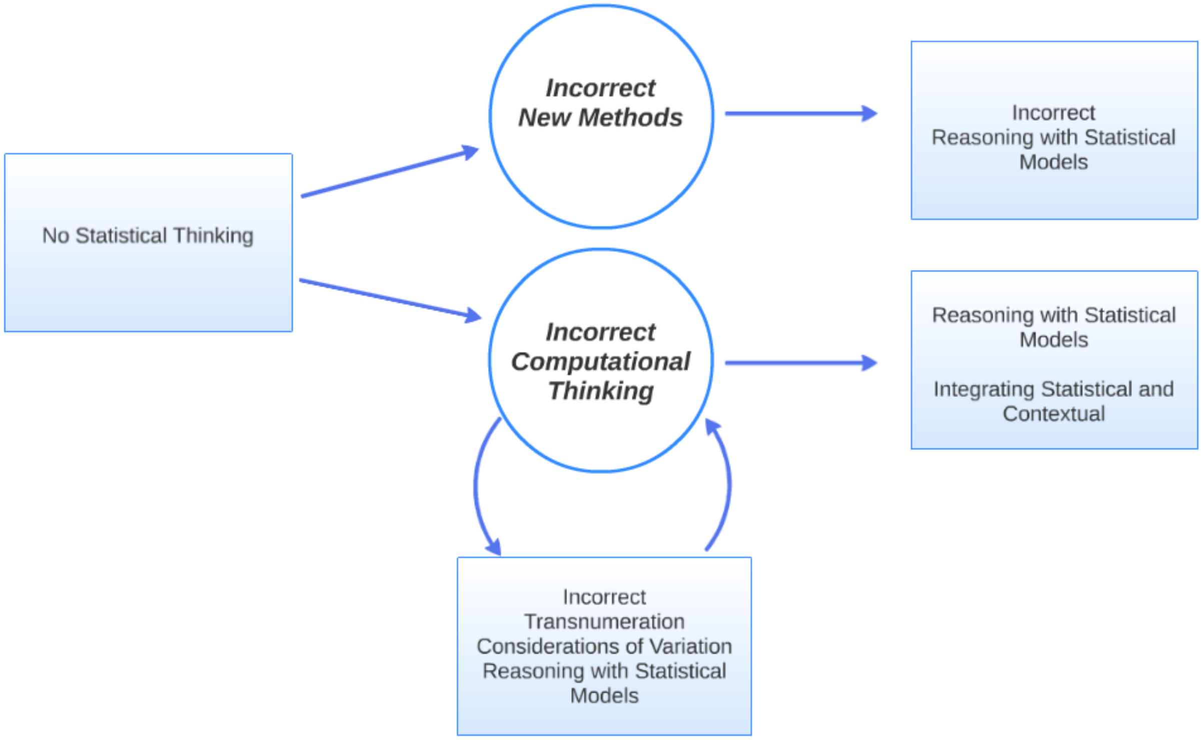


Figure 5.6 Relationship between statistical thinking and incorrect statistical computing

Sometimes student work had correct statistical computing with some or all incorrect statistical thinking. The final theme that was identified from the students' documents was that some students are able to use the computer correctly in all aspects of their work, but are lacking in their statistical thinking abilities and therefore come to incorrect conclusions. Incorrect statistical thinking with correct statistical computing comprised 28 of the 91 documents. What these documents demonstrated was that

1. Students may rely heavily on the capabilities of the computer to get a solution and are quick to trust the results that they obtain.
2. Students may use the computer to find solutions that already confirm suspicions they have about the data.
3. Students who have a poor statistical knowledge base may draw improper *Facts*, and thus have incorrect *Implications*, from the computer output.

The following examples from student work demonstrates how correct statistical computing paired with incorrect statistical thinking can produce inaccurate results.

Analysis without checking the data. Sometimes students would use the computer to find a solution to an inferential statistics problem, without first exploring the data that was given to them.

For example, in assignment 6, Chloe described an accurate solution strategy to test for a significant relationship between height and GPA, had all conditions for inference been met. However, she obtained her solution without first graphing the data to see what patterns emerged. Recall that in the height versus GPA assignment, there was an extreme outlier that actually makes the regression output appear to have a significant negative relationship. Yet, when the outlier is removed, there is no longer a significant relationship between the variables. The use of technology to analyze the data without first exploring the data led Chloe to an inaccurate solution.

Another example of this is when students do not take the time to consider the variability of data once the computer has given them a solution. Recall that in assignment 10, students were to create a simulation that helped them determine if the Bonferroni correction gave them a p-value at or below the significance level. Most students were able to define a simulation that allowed them to determine a p-value, however five students that completed the assignment stated their results in such a way that demonstrates they did not consider the effects of running the simulation multiple times. For example, Allison accurately programmed her problem and gave a good written response to the prompt, except for her statement:

> out of the 10,000 iterations I found that 4.84% were significant. This tells us that the Bonferroni adjustment is giving us a less than 5% type I error.

What Allison failed to realize is that the simulation she conducted will not always give a value of 4.84% and it may not even always give a value less than 5%. The computer has the capability of running her simulation fairly quickly, and it would have taken her a minute to run her program a few more times to come up with a reasonable range of values for the type I error.

Following directly from Allison's example, a third way students heedlessly solved problems is that even though they have the computer power at their fingertips, they may not always use it. An example of this comes from the work of Isabella in assignment 8. In this assignment, students were given several explanatory variables and were asked to determine which of them would contribute to the best model in predicting the price of a car. Despite having the choice of 11 explanatory variables to use in her model, Isabella decided to only try combinations of 3 of the variables. The computer could have done the work of reducing down the size of the data if the student just would have put the variables in her original model. Instead, Isabella believed that she had found the "best" model for the data, when in actuality, her classmates were finding models with much higher adjusted r-squared values and much lower AICs.

Context, not data driven decisions. A second problem that students have is that they may let the context of the data drive their decisions, even when they have used the computer to find a solution. For example, in assignment 7, students were asked if a person's height influences their ideal mate height. Specifically, they were trying to help an individual named Gary determine if he should ask out a girl that was shorter or taller than him. Initial regression of the data indicated a negative relationship between the variables (because of Simpson's paradox when using both males and females in the data set). While many students used the context of the data to realize that they needed to separate the data by gender, they were not able to use the output they obtained to give a logical reason why Gary should ask out a girl that is shorter than him.

When conducting the regression analysis on the data that contained only females looking for male companions, Cora found that there was a significant positive relationship between height and ideal mate height. The correct interpretation of this positive relationship is that shorter women prefer shorter men and taller women prefer taller men. However Cora, as well as several of her classmates, stated that a positive slope meant that a woman would want a man that is taller than them. While one could argue that this alone might mean that Cora was hastily finishing the problem and was not actually using context to inform her solution, a conversation with her proves otherwise. When writing the program for the assignment, Cora had come to get help in separating data. During that conversation, she had stated that she knew that she wanted a man that was taller than her, which she felt was common among her classmates, so the negative slope that Gary had initially found didn't make sense. This tells us that she believes the solution should be that Gary should ask out a female that is shorter than him, and a positive slope is the key to proving it.

Poor knowledge of basic statistical facts. A lack of understanding of basic statistical facts led several students to incorrect decisions in their documents. Three problems that were observed were that students had a poor understanding of statistical hypotheses and tended to switch the null and alternate hypotheses; students would read the numbers from the output incorrectly, and not check their work when the number did not make sense; and students would muddle the interpretations of selection criteria for model selection.

As an example of a student having a poor understanding of statistical hypotheses we can look to the work of Lilly. In trying to determine if there was a significant relationship between height and GPA (assignment 6), Lilly was able to use the computer to remove the outlier and conduct a proper analysis with the given data. However, after finding these results, which included a p-value that was greater than 10%, she gave an inaccurate solution to the problem when she stated:

> I knew that meant that we accept the null hypothesis to believe that we have sufficient evidence to show that there is a very, very weak relationship between the two.

Despite being able to use the computer to produce accurate results, this student found the wrong conclusion because she 1) defined her hypotheses reversed from what they should be and 2) accepted the null hypotheses.

As an example of a student reporting incorrect numbers from output, and not checking the results when they did not make statistical sense, we can look to the work of Natalie. In a model selection problem, Natalie read the scientific notation output incorrectly and stated that the p-value was almost 2. In a previous part of the assignment she had properly read the scientific notation which indicates that this was not an error in mathematical understanding, but statistical understanding. She made no mention that this p-value was impossible, or that she attempted to check her work when she obtained this p-value. Instead, since the value was so large, she decided not to reject her null hypothesis, when she actually should have.

Finally, students had trouble with remembering how each selection criteria worked in model selection. For example, in assignment 8, after creating several potential models to predict the price of a car, Olivia stated that the best model from the set of potential models was the one that has the *highest* AIC. Another example is from Genevieve who in assignment 5 stated that the model with the *lowest* R-squared value is the better model. Despite using appropriate programming and obtaining the correct output for these problems, each of these students got the wrong answer to their prompt due to a weak understanding of basic statistical facts.

Even though all of these examples demonstrate a student who has correct statistical computing and incorrect statistical thinking within the same argument, correct statistical computing does not necessarily lead to bad statistical thinking. As was demonstrated above, there were several examples of students' work where the student was able to use the computer appropriately and draw reasonable statistical conclusions. The difference in the students who could make appropriate conclusions versus the students in this section are the outside influences affecting the students. We can see that the pitfalls of implementing inferential procedures before checking the data seeing the computer as the authority, using intuition instead of the data and poor knowledge of basic statistical facts may actually be what led the students to their incorrect statistical thinking, despite their correct statistical computing. This has been summarized in Figure 5.7.

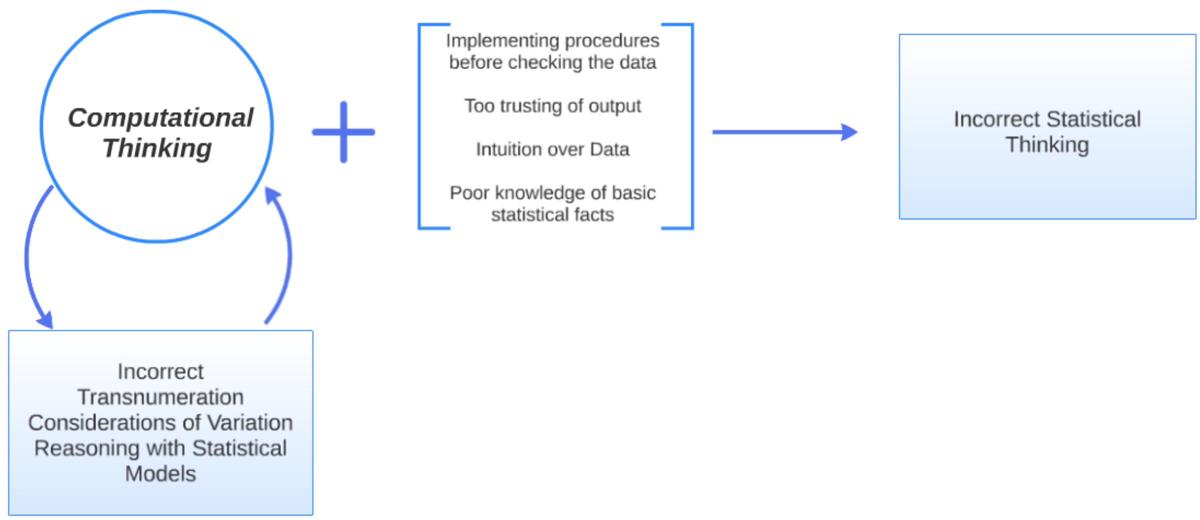


Figure 5.7 Relationship between statistical thinking and statistical computing when pitfalls are present

Implications for Teaching and Research

In well-defined statistical arguments, statistical thinking and computing are highly integrated. Statistical thinking constructs such as the need for data or reasoning with statistical models helped students identify the need for statistical computing tools. Statistical computing and thinking were then used in tandem to provide *Facts* for the argument, which led to *Implications* and *Conclusions* that again involved statistical thinking.

However, even if students were able to initially think statistically and properly utilize statistical computing, this may result in incorrect statistical thinking if the student possesses some pitfalls in their work, such as implementing inferential procedures before checking the data or being too trusting of the computer output.

When students did not first think statistically about the problem, they were more likely to exhibit incorrect statistical computing. Despite this, some students were able to reason statistically about the results they obtained, but still came to incorrect solutions to the prompt. Similarly, some students who did not demonstrate any statistical computing could still think statistically about a problem, but again did not come to correct solutions.

While there is still much work to be done in defining the relationship between statistical thinking and statistical computing, it is clear from the examples given above that there is a relationship between the two. Below are some potential implications for teaching and research based on the results of this study.

Implications for teaching. Based on this research five implications for teaching statistics with a computational element were identified. The first implication is that if you desire specific items from students, make sure to give students specific instructions to include them (Peck, 2005). While Chance et al. believe it is important to give detailed instructions so students don't get lost in the nuances of the technology (Chance, Ben-Zvi, Garfield, & Medina, 2007), it may also be evident from this research that explicit instructions may be needed just to get students to utilize the technology, as was seen with a majority of students in assignment 3.

The second implication for teaching is that student work should be assessed on the process, not on the final solution (American Statistical Association, 2016). As was discussed above, students who have poor computational thinking may either have a poor strategy for solving a statistical problem, program the problem incorrectly, or both. While students could obtain the correct solution for a problem even if they had poor computational thinking, this was likely due to luck. Instructors who are integrating technology in statistical assignments should take care to assess student work, not on the correct answer, but on the correct process and explanation, to ensure the student really knows what they are doing. While sometimes it is very obvious that students don't understand, such as was seen with Samantha, and her incorrect line of best fit, other times it is not and may be overlooked, such as was seen with Cora when she stated that Gary should ask out a female that was shorter than him based on an incorrect interpretation of her scatterplot.

One way to ensure that students understand the process, is by making sure that they don't just blindly follow example procedures. This could entail having students do problems that are programmed completely different from examples seen in class, or maybe more constructively, having students give an explanation of the program they have written. Holcomb et al. did this when they asked students who had worked through a simulation activity to answer questions about what they had done and how the simulation related to the original problem (Holcomb, Chance, Rossman, & Cobb, 2010). By having students explain their intended strategy, this may help instructors to ensure that the student can not only program, but that they understand what they are programming and they can communicate it to others. If they aren't able to explain the procedure, they probably won't be able to reproduce it later, or use it in a modified way when needed in future applications.

The third implication for teaching is that students need to recognize that statistical computing is necessary but not sufficient to create complete arguments with data. As was discussed above, the computer was found to be useful in helping students identify *Facts* that they then use to draw *Implications* from the data. When working with large data sets it is fairly necessary to use the computer to produce output and create appropriate arguments in answering statistical questions (Nolan & Temple Lang, 2010). Without the aid of the computer, the students wouldn't have the information they need in order to inform these *Implications*. Assuming *Facts* and *Implications* are both necessary to developing logical arguments, statistical computing is necessary to create complete statistical solutions.

However, the computer alone does not answer statistical questions. As stated in the GAISE College Report, "It will not be helpful for students to know about the tools and procedures that can be used to analyze data if students don't first understand the underlying concepts" (American Statistical Association, 2016). The computer can output *Facts* but students need to pair this output with their statistical knowledge to make decisions and present logical arguments. If students do not possess the appropriate statistical knowledge, the computer output will not be helpful in solving problems. As we saw with Olivia and Genevieve when they were confusing model selection criteria, both obtained correct output, but having poor statistical knowledge led them both to incorrect solutions. Being able to utilize statistical computing is not enough. Students need to have a strong statistical thinking base to use the computer output to make these arguments and the intuition to understand that the output from the computer may not always be accurate, either due to statistical or computational errors.

The fourth implication for teaching is that students should be encouraged to check their computing procedure both by themselves and with the aid of others (Chance, 1997; Nosek, 1998). While it is good for students to explore the internet and to learn to troubleshoot problems on their own, teachers may want to encourage their students (especially those that are new to programming) to either check with the instructor to see if they are programming correctly, or to discuss the procedure with a classmate to see if what they are doing makes sense. If Moira had checked her work with a partner, she may have come up with a better strategy for her simulation of the Bonferroni correction. While this didn't appear to be a significant problem with this set of students, it may be in other classes if the teacher is inaccessible or if the student has procrastinated on doing the assignment and doesn't have the chance to ask others for help.

The final implication for teaching found in this research is that students may need help in identifying pitfalls in their work, such as blindly trusting the computer output (Chance, 2002) or using one's intuition over the data (Pfannkuch, Wild, & Regan, 2014), and that they may need support to overcome these pitfalls. As was seen above, these pitfalls paired with good statistical computing, can lead students to incorrect statistical arguments. This means that it is not enough for a student to utilize the technology, students need to be able to overcome the pitfalls they may succumb to, to prevent them from making mistakes like the ones we saw above.

Implications for research. This small-scale qualitative analysis was meant to bring forward some of the relationships that may exist when students are using statistical thinking and statistical computing to answer questions with data. Due to the small sample size and observational nature of the data, it is hard to say which if any of the relationships found herein are causal or merely correlational. Future studies should include more students, if possible, and data should be collected in an experimental or quasi-experimental fashion to confirm or deny any of the relationships defined in this research.

Additionally, as identified by Lee et al. (2014) the writing assignments utilized for analysis only give a snapshot in time of the way that students worked through the problem to answer the questions. While students were asked to discuss their thought process in detail when creating these documents, what they say they did likely left out several key details from their problem solving process. This was demonstrated by the lack of perceived statistical computing in several assignments, despite having calculated solutions. This may be due to having difficulty writing about the statistical process as Park et al. (2016) found.

Understanding the entire process students took to answer these questions could change the way that some of the documents were coded, and consequently the relationships identified from this research. Future research could involve analyzing videotaped task-based interviews, or video and screencasts of students' working in class, to see the entire process students follow when answering these types of questions (Lee & Hollebrands, 2006).

Finally, researchers utilizing the framework defined in this report may need to infer the use of the computer in certain situations as students may not explicitly state its usage in written work (Gal, 2004). As was discussed above, the overlap between transnumeration and automation of computational thinking can be difficult to distinguish in written work. Despite students not specifically stating the use of the computer in their work, the researcher could infer that the computer was used to produce the transnumerative thinking. In these cases, statistical computing benefits were utilized, but without the deduction of the researcher, these items may be miscoded.

Statistical thinking and statistical computing are highly intertwined concepts that are important in helping students develop strong statistical arguments with data. If students are to succeed in using large data sets to make decisions, they will need to have both strong statistical thinking and statistical computing skills. We can offer these skills to our students, but there is still much work to be done on how we can best serve our students in this manner.

Chapter 6 : How Students Use Statistical Computing in Problem Solving

Journal

This article was written to contribute to the theory of learning and assessing students' work when they use statistical computing tools, for Technology Innovations in Statistics Education (TISE). According to the TISE *about* section, the purpose of the journal is to "report on studies of the use of technology to improve statistics learning at all levels". Specifically they accept research papers for empirical studies that "contribute to a theory of learning or a design of a technology or address the effectiveness of a particular technological tool." No page limit is given for articles submitted to this journal. The purpose of this paper is to discuss how students use statistical computing tools when problem solving, which could be helpful to instructors when trying to design lessons or create homework problems that utilize statistical computing tools.

Abstract

As data science becomes more popular, more statistics programs are training their students to better understand and utilize the tools that their future courses and their careers will likely require. Since these tools are being used in the classrooms, in lieu of more user friendly statistical software, student's conceptual understanding of statistics might not be as refined as those students who have had opportunities to learn statistics with programs such as TinkerPlots or StatCrunch. Using task-based interviews of students who completed a second course in statistics, the ways in which students utilize statistical computing tools, specifically R, while going through problem solving phases is analyzed and patterns that emerged are discussed.

Introduction

Utilizing computing tools has become very important in the field of statistics and in courses where statistics are taught. The GAISE report suggests incorporating computing tools in the classroom to explore concepts and analyze data (American Statistical Association, 2005; American Statistical Association, 2016). The reason for doing this is that it assists students in expediting computations so they have more time and cognitive capacity to spend on developing an understanding of the material. Some technologies may be helpful in taking the focus off of the procedural aspects of statistical analysis and help to put more focus on the conceptual understanding of topics (Watson, 2014; Fitzallen, 2013).

Several researchers (delMas, Garfield, & Zieffler, 2014; Ben-Zvi, 2006; Lee, et al., 2014) have shown that teaching students statistics with "learning tools" (Moore, New pedagogy and new content: The case of statistics, 1997), positively impacts students' ability to think and reason statistically which could be a stepping stone towards solving statistical problems. These menu driven tools (Chance, Ben-Zvi, Garfield, & Medina, 2007) tend to be more user friendly by integrating a graphical user interface. Additionally learning tools put more importance on having dynamic and interactive media to aid students learning (Biehler R. , Sofware tools and mathematics education: The case of statistics., 1993).

For example, delMas, Garfield and Zieffler (2014) found that using TinkerPlots in the CATALYST "Randomize, Repeat, Reject" method helped students to think and reason statistically about variation and the unusualness of data. Also using TinkerPlots, Ben-Zvi (2006) found that students were able to scaffold the idea of inference through simulated samples. This led students to be able to "discuss their thoughts and actions, explain their inferences and argue about data-based claims" (Ben-Zvi, Scaffolding students' informal inference and argumentation, 2006, p. 5) all things that may arguably help to assess, if not improve, problem solving skills.

Even though these "learning tools" have been found to be helpful in teaching students statistics, not all introductory statistics instructors have opted to use them. One alternative option that is being utilized are tools designed for "doing statistics" (Moore, New pedagogy and new content: The case of statistics, 1997). These tools tend to be less user friendly and require command driven communication as opposed to a menu driven method, to do statistical analysis (Biehler R. , Sofware tools and mathematics education: The case of statistics., 1993). The purpose of these tools is to do the intense analysis that is required of statisticians. While knowing how to use these tools will be important for majors and minors in statistics, and others that may need them for school or research, since they are not as user friendly, there is a learning curve for our students before they can fully utilize them in the classroom.

The expectation for statistics majors and minors to learn computing tools in their curriculum is growing (Chance, et al., 2014; Horton, Baumer, & Wickham, 2014; Nolan & Temple Lang, 2010). Within statistics courses, students are learning data cleaning, graphical displays and multiple techniques to aid them in their analyses (Hardin, et al., 2015). Yet, we cannot expect students to teach themselves the majority of the technology while the instructor teaches only the statistical concepts. If students are forced to learn the technology on their own, they will focus on getting the answers to the problems they have been assigned and often pick up bad programming habits due to the lack of instruction and direction (Nolan & Temple Lang, 2010). Conversely, focusing too much on how one uses the technology to conduct statistical tests may teach the students that it is not important to explore the data before conducting formal inference procedures.

Thus, it is important for teachers to find the appropriate balance between teaching the statistics and teaching the technology that is used to conduct the analysis.

Despite the cautions of the ASA about including programs like R and SAS in introductory statistics courses (American Statistical Association, 2016), some students are able to successfully learn these higher level languages while still grasping the concepts traditionally taught in introductory statistics courses. Horton and his colleagues were able to successfully introduce the statistical language R into their introductory statistics courses. Using RStudio and the MOSAIC platform, their students were able to perform higher level statistical processes with a more difficult technology earlier in their statistical careers. They do note however, that in teaching the statistical computing ideas, there should be more examples and activities, but less content taught overall (Horton, Baumer, & Wickham, 2014; Kaplan, Horton, & Pruim, 2015).

Horton and his colleagues are not alone. Others have started to use tools meant for "doing statistics" in place of "learning tools" in their classrooms, with the intention of helping students prepare for upper level courses and their future careers. (Horton, Baumer, & Wickham, 2014; Baumer, Cetinkaya-Rundel, Bray, Loi, & Horton, 2014). While these authors were able to show that students can utilize the tools and show a greater appreciation for them after using them for a semester, there could be some drawbacks to using these types of technology. While students may be more prepared technologically to do statistical analysis, little has been done to understand the ramifications of choosing to use command based statistical computing tools. In particular, the goal of the research in this paper is to understand the ways in which students utilize statistical computing tool when solving statistical problems and how such use affects a student's ability to solve statistical problems.

Framework

In order to examine students' problem solving with statistical computing tools, it is useful to first describe a working definition of statistical computing, identify affordances of statistical computing in students' work, and be able to classify problem solving phases. These aspects will work together as a guiding framework for this research.

What is Statistical Computing? Data science and statistical computing tools are no doubt becoming more common in the statistics classroom (Horton, Baumer, & Wickham, 2014; Baumer, Cetinkaya-Rundel, Bray, Loi, & Horton, 2014). While not everyone agrees on what precisely constitutes statistical computing - for example Hardin, et al. (2015) describe 7 different courses dubbed statistical computing - most agree that our students need to learn how to integrate computer science into the entire statistical analysis process.

Some common themes occur in the literature of topics that could be covered in an undergraduate level statistical computing course. These topics include graphical displays, exploratory data analysis, data cleaning, non-parametric statistics including bootstrapping, simulations, MCMC, the use of more than one language - such as teaching students R, but supplementing with a secondary language such as SQL – and teaching students presentation methods like LaTeX in R Markdown or HTML for web design (Dierker, Kaparakis, Rose, Selya, & Beveridge, 2012; Gentleman, 2004; Hardin, et al., 2015; Monahan, 2004; Nolan & Temple Lang, 2010).

Suppose a statistician has been tasked with conducting a statistical analysis. If her only job was to put the data into her favorite statistical programming package, run a few tests to find results, then report those back to her employer, she really wouldn't be doing statistical computing. The use of the technology in this case is what Pea (1987) dubs an amplifier. The technology is being used to simplify tasks, helping the statistician to use their time more efficiently. According to Nolan and Temple Lang (2010), statistical computing is

> about using the computer to make sense of data. It is inherently interactive, with the next step determined dynamically by the content we see in the previous steps. It is the tightly coupled combination of programming, visualization and statistical reasoning (p. 5).

Suppose instead our statistician did the following. First, she went to a website and scraped the data she needed, using data mining tools. Then upon inspection of her results, she realized the data was not in a format suitable for analysis. She then decided to use her database tool knowledge – like SQL - to clean and sort the data for relevant information. Next she explored the data through graphical displays to find patterns or points of concern which helped her determine the type of statistical test she was going to use in her analysis, and if conditions for inference for that test were met. Next, she used the technology to crunch the numbers and find an answer for her problem. During the analysis, she saw that data she collected behaved differently over time, so she decided to create an interactive presentation of the results on a webpage that helped to demonstrate this. We can see that this process is much different than the first one given above. This statistician is engaging in statistical computing.

Every one of these steps is not necessary to be considered as doing statistical computing. However, statistical computing is not just about using a computer to get the answer to a statistical problem. Statistical computing is the integration of technology into all steps of a statistical analysis, from data collection to presentation. As we saw above, this could include data mining, data cleaning, exploratory data analysis, statistical analysis and presentation of results.

Benefits of Statistical Computing. The main advantage that students probably recognize about taking a statistical computing course is that the use of statistical computing skills has become more prevalent in the work force. In their paper, Hardin et al. (2015) speak of several students who had written to their professors after graduating and telling them the usefulness of the statistical computing course they had taken. This is no wonder, since students graduating at the Bachelor's and Master's level spend much of their time in the workforce using skills like "retrieving, filtering and cleaning data and doing initial exploratory data analysis" (Nolan & Temple Lang, 2010, p. 4).

Phillip, Schuler-Brown and Way (2013) note that while it is becoming more and more acceptable to discuss the work preparation aspects of any program, thinking about schooling in such a way limits other important goals that students might have when obtaining a higher education, such as intellectual, social, civic, ethical and aesthetic reasons.

In particular, students can benefit intellectually from taking a statistical computing course. While, statistics and computer science are different fields, they both require many of the same skills for a student to be successful, such as logic and abstraction (Wing, 2008; delMas R. , 2004) and both help their students to develop these skills. Some researchers have found (Kafura, Bart, & Chowdhury, 2015; Yuen & Robbins, 2014) that when students not majoring in computer science or statistics learned computing with a statistical context, both their computational and statistical thinking grew and supported the other. In particular "programming assists students in understanding how data is organized, structured, and presented so that they can accurately analyze it in both numeric and graphical form" (Yuen & Robbins, 2014, p. 14) while the "quantitative concepts were also developing their programming skills".

One major reason that researchers have found to support students learning statistical computing is that (if done properly) it automates computational procedures and helps to reduce the cognitive load on the student (Baglin, 2013). “Attention to key statistical concepts seems to be easier if the cognitive load required for computation and graph drawing is minimized by software” (Biehler, Ben-Zvi, Bakker, & Makar, 2012, p. 680). Students can more easily explore trends in the data (exploratory data analysis) when they aren’t worried about the nuances of how a graph is drawn. Or they can spend more time interpreting and understanding why the results of a statistical study make sense since they haven’t spent all their cognitive capacity trying to manually compute the results of that analysis.

The second major benefit of studying statistical computing is increased computational thinking skills. As defined by Wing (2008), computational thinking is a mixture of mathematical, scientific and engineering thinking that is not constrained by the bounds of reality. It allows one to think critically and abstractly about real world problems, while allowing for the use of the almost infinite power of the computer to solve problems that might not otherwise have solutions strategies in the real world. A person that can think computationally has problem solving skills that allow them to formulate a solution strategy which is then conveyed to a computer, telling it how to implement the given strategy (Wolfram, 2016).

Another advantage to using technology in a statistics classroom is that it offers alternate techniques to solving problems than have been unavailable in the past. For example, using simulation-based methods for developing distributions and solving statistical inference problems has recently become a popular teaching technique (delMas, Garfield, & Chance, 1999; Garfield, delMas, & Zieffler, 2012; Holcomb, Chance, Tietjen, & Cobb, 2010; Lane-Getaz, 2010; Tintle, VanderStoep, Holmes, Quisenberry, & Swanson, 2011). A major advantage to this type of analysis is that students don't have to have an in depth understanding of the properties of distributions, such as the normal distribution (Pfannkuch, et al., 2011). Instead, students conceptualize the idea of *what would my statistics look like if we could repeat this process several times?*; a concept that is foundational for several topics in statistics, including sampling distributions, confidence intervals and hypothesis testing. Instead of spending time teaching students how to read a normal distribution table, they are using software to generate p-values and explore the idea of "extreme" statistics.

One example of this is the CATALST program (Garfield, delMas, & Zieffler, 2012). Their goal was to make sure that students learning statistics didn't see it as a "cookbook" of procedures to be followed. By the end of the course, they wanted their students to know how to "cook", or in other words, know the essentials of statistics, without being someone who is simply following a statistical recipe that their instructor gave them. The authors found positive results for their program and noted that students' reasoning about statistics was on par with students who had taken a traditional introductory statistics course.

The final benefit of utilizing statistical computing is learning multiple programming languages or programming tools. Ideally, this would help students to perform the different tasks of statistical computing that they are expected to do in order to conduct a complete statistical analysis, such as data collection, data cleaning, analysis and presentation. In addition to the practical applications of learning multiple languages and tools, students may also develop decision making skills so that they know which of these tools is best for solving different aspects of their statistical problem (Nolan & Temple Lang, 2010). Brown and Kass (2009) claim that success in statistics requires that students have adaptable strategies for solving problems but that our students have a tendency to "attack problems using blunt instruments and naïve attitudes" (p. 105). By teaching statistical computing to our students, we hopefully give them sharp tools and the knowledge of when to use them.

Learning multiple programming languages could also help students develop pattern recognition skills. Nolan and Temple Lang (2010) argue that the main goal in teaching programming is for students to be able to extract concepts from a specific problem and transfer concepts to the specifics of other languages and environments. When programming languages are taught together, students can see the patterns that emerge between them and get a better understanding of programming as a general concept instead of the idiosyncrasies of individual languages. Even when students only learn one language, developing code based on examples can be a helpful tool for learning syntax.

In summary, learning statistical computing provides students with the opportunities to increase their knowledge of the tools and become more proficient statisticians in the real world. Students can use their knowledge of statistical computing tools to efficiently create graphs and summaries of their data. In doing this, they have more cognitive capacity to spend on discerning patterns they observe or understanding concepts they are trying to grasp. Additionally, learning statistical computing may impact one's computational thinking which could in turn improve the student's problem solving skills, their logic and reasoning skills and their sophistication to think critically and abstractly. Finally, the technology could be used to implement new or multiple statistical analyses, allowing the student different perspectives on statistical concepts.

Table 6.1 synthesizes the information described in this section. Based on the research that has been done, these are believed to be the primary elements of the affordances that statistical computing has to offer students, in terms of intellectual benefits.

Table 6.1 Affordances for different aspects of statistical computing

Aspect of Statistical Computing	**Affordance for the student**
Automate computational procedures	• creates multiple graphs or summaries to make sense of the data • uses the technology to conduct the analysis, then makes proper inferences
Increase computational thinking	• creates a solution strategy and can communicate it to the computer • is able to think critically or abstractly about a concept that the technology is demonstrating
Offer new methods to explore concepts	• does not follow a prescribed set of procedures to answer a question • comes up with their own methods and reasons for using the technology in analysis
Aid in pattern recognition and decision making	• recognizes patterns, either in the way things are coded or in how the results behave • uses the output from the technology to influence their decision of where to go next in the analysis

Assessing Problem Solving. In addition to understanding how students use technology to solve statistical problems, another purpose of this research is to see how students use these tools in different phases of problem solving. To reach this goal, it will be useful to identify different phases of problems solving. Several authors have proposed methods for problem solving structures, but for the purposes of this study, we began with the structure created by Lee and Hollebrands (2006) based on the work of Schoenfeld (1985). The phases in this structure include analysis, planning, implementation, assessment, verification and organization. Definitions of these items can be found in Figure 6.1.

Analysis	Students attempt to understand the problem, its constraints, and consider useful perspectives for the problem.
Planning	Students work towards devising a sequence of actions or formulate a strategy to pursue a solution.
Implementation	Students carry out a pre-planned strategy, which may or may not be explicitly stated.
Assessment	Students take explicit actions to monitor their progress towards a solution.
Verification	Students believe they have a solution and either evaluate whether or not it is correct, or engage in activities to explain or justify their solution.
Organization	Students take actions to arrange their environment (either with paper or in the technology tool) in a way that they perceive will be beneficial in working towards a solution.

Figure 6.1 Phases of problem solving

Methods

To determine how statistical computing is used to aid in problem solving, a small-scale qualitative research study was conducted. According to Groth (2010), if done appropriately, qualitative research can be scientific, rigorous, generalizable, and objective and can "provide a means for conducting scientific studies that contribute to our understanding of statistics teaching and learning". Qualitative methods are appropriate for answering the research question because the goal is to do an in-depth examination of relationships that exist between statistical computing and problem solving. For example this question can be answered by determining when and what types of statistical computing are utilized in each of the problem solving phases, as well as how students use statistical computing in those phases.

Participants. Participants for this study came from a small private women's college in the southeastern United States who took a second course in statistics during the fall semester of 2016. Statistics 2 is a continuation of introductory statistics. It includes topics of one and two sample inference, two way tables, simple and multiple linear regression and one and two factor ANOVA. Students were also taught introductory statistical computing practices in RStudio while taking this course. Fourteen students completed the course with the primary investigator.

In total, 5 students volunteered to participate. Despite being a volunteer and convenience sample the variation among the students made for a rich collective case. The following describes the statistical and computational backgrounds of these students beyond the statistics 2 course. They are ranked by what the researcher believes is the most experienced to the least experienced as far as total knowledge of computing and statistics. These rankings are based on information gained from the interview and a demographic survey as well as observations made by the researcher while the students were in class. Each of these students was taught by the researcher in the statistics 2 courses and three students, Allison, Cora and Samantha, were pupils in at least one other course.

- Allison: Statistical Computing using R, tutored Intro Statistics, participated in DataFest.
  (Ranked 1st in statistical knowledge and 1st in computational knowledge)
- Cora: Statistical Computing using R, intro statistics grader.
  (Ranked 2nd in statistical knowledge and 3rd in computational knowledge)
- Genevieve: Psychology Statistics course which included SPSS.
  (Ranked 3rd in statistical knowledge and 4th in computational knowledge)
- Moira: Computer science major with experience in SQL.
  (Ranked 5th in statistical knowledge and 2nd in computational knowledge)
- Samantha: Introduction to R through an independent study in Intro Statistics.
  (Ranked 4th in statistical knowledge and 5th in computational knowledge)

Interviews. In the spring of 2017, after completion of Statistics 2, five students completed a task-based interview with the primary investigator. Task-based interviews are structured or semi-structured interviews where students are expected to complete a task and explain their thinking while doing so (Koichu & Harel, 2007). These types of interviews are useful tools for researchers because they allow us to see how students solve problems and they can help researchers form conjectures about students' content knowledge and the ways in which they think about and solve problems (Goldin, 1997; Koichu & Harel, 2007).

The interviews began with informing students that their actions would be recorded both by video and by screen capture. They were then instructed to try and solve the problems to the best of their ability, using the software that was provided for them, and to talk about what they were thinking as they solved the problem. Students were also told that they could use any resources that they would have had they been working on this problem at home.

While there were three segments conducted in the interview, only the first segment was used in this analysis. In this segment, students were given a prompt describing an experiment that was conducted to study the effects on tipping behavior when a waitress wrote different statements about future weather patterns on the bill. The full prompt can be found in Appendix B. The students were prompted to "conduct a full statistical analysis" using the data that was provided to them. Occasionally during the interview, if students got off track or were completely stuck, the interviewer would try to help them get back on track or give leading questions to assist them in working towards a solution. The interviewer also intervened when the student had not discussed their thinking for long periods of time.

Since students were taught to use RStudio during their Statistics 2 course, they were expected to use that software to complete the task. A laptop was present at the interview which had RStudio loaded and was screen recording the work they did during the interview. Students were provided with the following information in an R script.

```
good.report <- c(20.8, 18.7, 19.9, 20.6, 22.0, 23.4, 22.8, 24.9, 22.2, 20.3, 24.9, 22.3, 27.0, 20.4, 22.2,
24.0, 21.2, 22.1, 22.0, 22.7)
bad.report <- c(18.0, 19.0, 19.2, 18.8, 18.4, 19.0, 18.5, 16.1, 16.8, 14.0, 17.0, 13.6, 17.5, 19.9, 20.2,
18.8, 18.0, 23.2, 18.2, 19.4)
no.report <- c(19.9, 16.0, 15.0, 20.1, 19.3, 19.2, 18.0, 19.2, 21.2, 18.8, 18.5, 19.3, 19.3, 19.4, 10.8,
19.1, 19.7, 19.8, 21.3, 20.6)
```

The following is a list of anticipated steps students would need to complete to procure a full analysis:

- Identify that the problem requires a one-way ANOVA analysis.
- Graph the data to see if there is a potential relationship between the weather report and the percentage of the total bill given as a tip.
- Check the conditions for inference for ANOVA to see if the statistical analysis was appropriate. These conditions include:
    1. Subjects must be randomly assigned to groups and that the groups are independent.
    2. The quantitative value within each group must be normally distributed or the sample within each group must be sufficiently large for the central limit theorem to apply.
    3. The variances for each group should be equal.
    4. There are no outliers that influence the results unduly.
- Format the data so that it can be used in R. This included concatenating the given quantitative variable and creating a categorical variable that describes the type of weather report given.
- Use R to conduct the one-way ANOVA procedure.
- Interpret the results that the computer had given in the context of the problem.

Even though each student was told to conduct "a complete analysis", they were not specifically told to do descriptive statistics prior to doing inferential statistics. This was purposeful, to see if students bother to utilize graphical procedures and check conditions for inference prior to analysis. As mentioned above, using the computer to determine the p-value and interpreting results does not constitute statistical computing. It is incorporating the computer into several aspects of the analysis that constitutes statistical computing (Nolan & Temple Lang, 2010).

Additionally, as was noted above, the data was given in the particular format to add a complication to the usual statistical inference procedures that students would have conducted in class. Normally, the students would be given data where the quantitative variable was given in one single vector of values, with a second vector comprising the explanatory variable. While students never needed to do this new procedure in class, they were given tools, such as the concatenation [`c( )`] function and the replicate [`rep( )`] function, that could assist with such a process if they realized it needed to be done.

This was done to see if students could apply functions learned in other domains to problems they were facing in the present (Chance, 1997). A help file was created prior to the interviews that formatted the data in a way the students would be able to use it for the analysis. This file was only given to students to assist them at this point in the interview if this perturbation was outside their "zone of potential construction" (Steffe & Olive, 2002) and it looked like no more forward progress would be made.

Analysis. Each task-based interview was examined at least three times for the purposes of this analysis. In the first pass through the interviews, the videos were coded for instances of statistical computing as described in the framework in Table 6.1. Each video was inspected a second time for accuracy of these codes, with a sample of students' work viewed and discussed among both researchers until agreement was met. In the third iteration of the videos, students' work was categorized into the problem solving stages defined above.

In the final pass through the interviews, it became evident that students were using the technology for *planning* in ways not entirely characterized by the problem solving framework (Figure 6.1). Because of this, another category, *research*, was added to the problem solving phases. Since students were asked to use a computer to aid in solving a statistical problem, they had access to the internet and each of them used it to research potential programming code to solve the problem. Initially, these instances were identified as planning. However, it was later given its own category to distinguish it from a plan the students created on their own versus a plan they created in conjunction with doing research on the internet. This addition can be found in Figure 6.2.

Analysis	Students attempt to understand the problem, its constraints, and consider useful perspectives for the problem.
Planning	Students work towards devising a sequence of actions or formulate a strategy to pursue a solution.
Research	Students use the resources available to them to aid in forming a strategy to pursue a solution.
Implementation	Students carry out a pre-planned strategy, which may or may not be explicitly stated.
Assessment	Students take explicit actions to monitor their progress towards a solution.
Verification	Students believe they have a solution and either evaluate whether or not it is correct, or engage in activities to explain or justify their solution.
Organization	Students take actions to arrange their environment (either with paper or in the technology tool) in a way that they perceive will be beneficial in working towards a solution.

Figure 6.2 Revised phases of problem solving

Finally, graphics were made from the information gained in the interviews to help make sense of the large amount of data collected which aided in answering the research question.

The graphic in Figure 6.3 gives information about the order in which students used the problem solving phases, how long students spent in each of the problem solving phases and how long students spent in the interview solving the problem.

Figure 6.3 Sequence of problem solving phases in the task-based interviews

Table 6.2 gives information on the number of times each of the statistical computing codes were identified in each of the problem solving phases as an aggregate count across the five students.

Table 6.2 Number of statistical computing actions used in each of the problem solving phases

Key: Automation (A), Computational Thinking (CT), Incorrect Computational Thinking (ICT), Pattern Recognition and Decision Making (PRDM), Incorrect Pattern Recognition and Decision Making (IPRDM), New Methods (NM), Incorrect New Methods (INM)

	A	CT	ICT	PRDM	IPRDM	NM	INM
Analysis	0	7	9	3	0	0	0
Assessment	0	9	7	1	0	0	0
Implementation	2	12	15	1	1	3	0
Organization	0	0	0	0	0	0	0
Planning	0	13	10	3	0	0	1
Research	0	6	1	1	0	3	0
Verification	1	2	0	1	0	0	0

The Ribbon chart in Figure 6.4 gives information on the types of statistical computing that were utilized in each of the problem solving phases for individual students, and the proportion of time during their interview that they spent on the problem solving phases.

The reader should note that the grey sections define the proportion of time for interviewer intervention. The order of the phases in the graph is alphabetical, as that is the default of the graphing software. This does not indicate the order in which students used the problem solving phases in the interviews. Any statistical computing benefit that occurred during these problem solving phases is overlaid. Note that each of the statistical computing codes was not observed every time the student was in a particular problem solving phase, but that it showed up at some point in that phase during the course of the interview. The codes listed in the graphic below are the same identified in the key in Table 6.2.

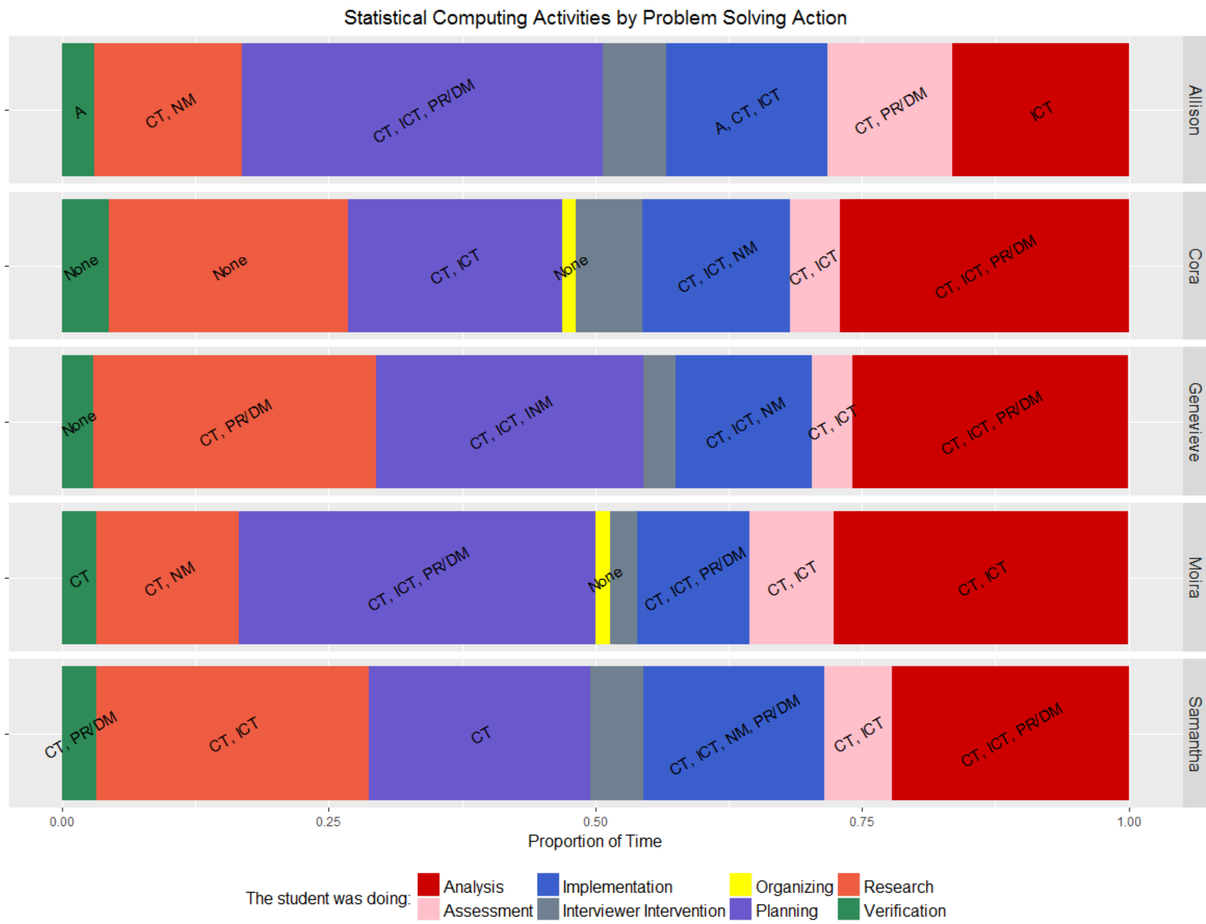


Figure 6.4 Proportion of time each student spent on the different problem solving phases

Results

In this section we will discuss findings from the interviews as they pertain to how students use technology to solve problems and differences in the ways students solve problems with technology.

Engagement in Problem Solving Phases. It is clear from Figure 6.3 that students completing the interview spent varying amounts of time both completing the task and in different phases of problem solving. One thing that each of the students have in common though is the first two phases of the problem solving process were the same. In particular each student first spent time *analyzing* the problem and then *planning*. This should make sense given the structure of the interview. Students first needed to read the prompt and then formulate a statistical strategy in order to solve the problem. With a small exception from Samantha and Moira, both needing to reread the prompt before deciding on an appropriate statistical analysis tool, each student then received more instructions from the interviewer on what to do next. This is why there is a grey block for “interviewer intervention” in the problem solving diagrams for each student. This is where students were told to conduct a complete statistical analysis.

After this, the paths of the students diverged. Allison, the student who was the most experienced both statistically and computationally, immediately started researching how to do a one-way ANOVA using R. She did this first by using the help menu in RStudio and then by searching the internet. Cora, probably the next most statistically experienced student, first organized her screen and then started planning what it was she wanted to do in order to solve the problem. The other three students needed to go back to the prompt first and again make sense of what it was they were trying to find (analyzing).

Even though students took different paths, each student found themselves at some point struggling with how to write a piece of the program necessary to solve the problem. This led each student to a sequence which included *research, planning, implementation* and *assessment*. Essentially, students would use the internet to research example code, they would plan what they thought they should program to mimic the results online for their problem, they would then implement that plan and check the results that that they obtained to see if they made sense.

We can see from Figure 6.3 that Moira and Cora, (ranked 2 and 3 in computational knowledge, respectively) more than the other three students, moved through this process of four phases quickly and often, using what they saw as a learning tool to determine what they needed to code to solve the problem. While sometimes one of the steps was omitted, they tended to use this strategy a lot. Allison (ranked 1st in both computational and statistical knowledge) utilized this method just once. It was apparent from her behavior in the video that she had a general idea of what she needed to do, and with very little *research* she was able to do it. Cora and Moira were less knowledgeable of the statistics that needed to be done but still had the computer knowledge to fall back on and were able to solve the problem through a series of trial and error.

Genevieve and Samantha struggled with this process a little bit more and as we can see from Figure 6.3, had large time periods of *research*. It was clear in the interviews that they did not know what to look for online and spent a lot of time trying to make sense of what they did find. In a completely opposite dynamic from Moira and Cora, it was as if they were afraid to try something on the computer and learn from it. This could be due to a lack of experience in using the computer, or it may be that their problem solving strategy in a general sense was to ask for help when they were stuck, as opposed to figuring it out themselves. Both students asked me a lot of questions and Genevieve in particular, was visibly and verbally distraught that I would not answer her questions.

While this information gives us a bit of insight into how the students were solving problems, we still have yet to see how they used statistical computing in those phases.

Use of Statistical Computing Tools in Problem Solving. In this section, we will take a closer look at the information portrayed in Table 6.2 and Figure 6.4, above. Observances of the four aspects of statistical computing are described in the sections that follow.

Automation of Computational Procedures. Recall that automation of computational procedures included students creating graphs or summaries to make sense of the data, using the technology to conduct the analysis and make proper inference, and utilizing the technology to maintain a reduced cognitive load. The ultimate goal with this benefit of statistical computing is to have students be proficient enough at programming the software that they can do it without much thought, and they do not lose sight of the goal of the problem. The technology is helping the student solve the problem, not standing in their way of trying to do what they wish to accomplish. We can see from Table 6.2 and Figure 6.4 that Automation was only seen in the *implementation* and *verification* problem solving phases and it was only utilized by Allison.

Allison utilized automation for computational procedures twice in the *implementation* phase. In one example, she was trying to use the technology to create a graph so that she could determine if conducting a one-way ANOVA procedure was necessary. She discussed her plan to visualize the data in side by side boxplots to determine if it was reasonable to attempt a one-way ANOVA procedure.

> *Allison*: Do I need to look at the histograms to see?
> *Interviewer*: If that's what you feel like you need to do, then you can go ahead and do that.
> *Allison*: Well, or if you looked at boxplots or the five number summaries of each one and you compared. You know what I mean? If you compared each thing, just so, first off you can see if there is a really obvious difference between them. Because if there isn't, usually, I don't know. I just like to graph.

After creating the graph, Allison was easily convinced that there appeared to be a difference between the boxplots and that conducting the analysis would be appropriate.

> *Allison*: So looking at this, you can definitely tell that there is a difference in the average tip given between those (moves the curser between the first and second group) and so that, you know, I think it's worth looking at.

This is a great example of using automation of computational procedures because Allison had no trouble creating the graphs, and once they were prepared she was able to interpret them and use the information she had found to inform her next step in the analysis.

As we can see from Figure 6.4, no other student was able to benefit from automating computational procedures. In fact, all five students struggled with formatting the data properly to be able to use the technology to conduct inference, and spent a good deal of time just trying to get the data formatted appropriately. This led to students forgetting the goal of the problem (not maintaining reduced cognitive load) and having to stop what they were doing to remind themselves of their goal several times. We can see this in Figure 6.3, with the multitude of times students had to go back to the *analysis* phase of problem solving.

For example, it took Samantha almost 28 minutes from the start of the problem until she found the output for the analysis. By the time she had gotten to that point, she had forgotten that she was just trying to see *if* there was a difference in the groups as opposed to which group had the largest tip percentages.

> *Samantha*: My p-value is less than .01. So that would mean that is significant, in which case I need to do one of those three methods, I think.
> *Interviewer*: Okay. So does the question ask you to figure out, uh, which one is bigger or if one is bigger?
> *Samantha*: No. I don't think so. Wait. (reads the prompt again) Okay, yeah, the data supports that there are differences because I have a p-value that is less than .01.

Note that while Samantha had been able to tell me that there are differences, she did not give any context to the problem when stating her results. Paired with the fact that she couldn't remember in general what it was she was trying to find from the problem, it's probably not a coincidence that she did not give specific details about the context of the problem when interpreting her results.

Computational Thinking. Recall from the theoretical framework that computational thinking entails three characteristics. The first characteristic is that the students can create a computational solution strategy, as opposed to a statistical solution strategy, for the problem. Second, students who are thinking computationally can communicate their solution strategy to the computer. The third aspect of computational thinking is being able to think critically or abstractly about a concept that the technology is demonstrating.

It should be no surprise to the readers that computational thinking was often seen in the *planning* and *implementing* phase of problem solving. Being able to create a solution strategy is essentially the student devising a plan for what they want to program and communicating that solution strategy is equivalent to them implementing their plan. These two items were not always observed together, though. Several times, students would give a computational solution strategy, but then were never able to implement it.

For example, when Moira reached the point in the interview where she discovered the data was not formatted properly, she knew that she needed to join the data together into one vector. She even presented a plan for how to do this using SQL.

> *Moira*: I remember this happening. I do. Um, what was it? I'm getting it confused with my computer science stuff. I mean like um, we do a JOINTABLE kind of thing.

In this *planning* phase, she had a computational strategy, but she could not remember how to implement it in R. Even though she didn't implement her proposed plan, she was still thinking computationally.

While not a common occurrence, the third element of computational thinking, thinking critically and abstractly while utilizing the computer, was also observed. In a discussion with Allison about concatenating the data in a *planning* phase, she was attempting to visualize what the data would look like after she ran the concatenation. Even though her thinking wasn't completely correct, she was still thinking abstractly about the data manipulation before she ran the program.

> *Allison*: If we did c parentheses all of these, then it would be all of them together. That's probably not real.
> *Interviewer*: So what would that look like? You've taken linear algebra. What would that look like as a vector?
> *Allison*: Well currently we have three vectors, one for each thing. Oh, so it would just be a 3 by however many tip matrix… But I just want it to be one column.

Based on her dialog, we can tell that Allison is visualizing a matrix that consists of 3 rows by 20 columns. While this doesn't fit her plan of creating one long vector, it eventually helps her towards the desired data format to conduct the analysis.

While computational thinking was seen in each of the problem solving phases except *organization* (which didn't include any statistical computing constructs), an interesting pattern occurred in which the students were utilizing computational thinking. As was discussed above, when students were stuck and didn't know how to program the software, they fell into a pattern of *research-planning-implementation-assessment*. This essentially entailed the student using the help menu or online examples to assist them in determining what they needed to do next. Students would find some code, then try to mimic it with the problem they were given and then check their work.

This is one of the reasons why incorrect computational thinking was so prevalent in the students' work. For example, students were marked as having incorrect computational thinking in the *research* phase if they did not search for example program that would help them to solve the problem. Or the student was marked as having incorrect computational thinking in the *planning* phase if they had found an example program they thought would work, but misinterpreted what they thought the code actually did. Students were marked as having incorrect computational thinking in the *implementation* phase if they were unable to communicate their planned code to the computer. Finally, students were marked as having incorrect computational thinking in the *assessment* phase if they incorrectly interpreted what an error code meant.

We can see from Table 6.2 that students had most of their incorrect computational thinking in the *planning* and *implementation* phases. However, even though students struggled to determine what a program was doing or how to tell the computer to perform certain actions, some students were able to use the information they gained from their incorrect work to move forward in the problem solving process. This means that a student having incorrect computational thinking may not be a bad thing, as long as they were learning from it. It was a matter of the student's persistence level as to how far this strategy helped them.

Pattern Recognition and Decision Making. Recall that pattern recognition included students being able to recognize patterns either in the way things were programmed or in how the results behave. Decision making was defined as students using the output from the technology to influence their decision of where to go next in the analysis. While identified in other problem solving phases, pattern recognition and decision making was seen most often in *planning* and *analysis*.

As an example of pattern recognition in student work, we can look back to Moira's conversation about trying to format the response variable properly. When she was creating a strategy for how to format the data in a *planning* phase, she fell back to her SQL knowledge to describe how she might format the data appropriately for the analysis.

One point of interest about this situation was the little amount of time and effort it took Moira to formulate her plan. Even though she was strategizing in the wrong programming language, she did not struggle to think of a way to format the data in a way that she could use to help solve the problem. She thought about how to use the tools in her toolbox, as opposed to the "blunt instruments" (Brown & Kass, 2009) in front of her, to help answer the question. If she had been allowed to use SQL to solve the problem, or if she had known that SQL commands could be used with certain R packages, she may have been able to format the data properly without any programming hints.

Another way in which pattern recognition and decision making was identified was in the *research* phase, when students were trying determine how something was programmed in an example they found online, then adapted it to fit the structure needed for the problem they were trying to solve. While it wasn't common, it is of note to give the distinction between how pattern recognition and decision making were found in *research* versus *planning*. For example, when Genevieve was trying to determine how to join her quantitative variables together, she found an example online of the concatenate function and accurately compared the variables on the website to the ones she had in her problem.

Genevieve: Okay so like let's say n and s, n would be good message, or whatever and the s would be like the bad one, so here they're just like combining good and the bad.

This particular example was coded under the *research* phase because Genevieve was comparing to an example that she had in front of her. Moira's example above and another by Allison not given here were both coded under the *planning* phases because they were referring back to examples in memory to make these comparisons.

Pattern recognition and decision making was also seen when students were trying to make sense of their problem in the *analysis* phase. For example, after a long struggle with getting the data into the appropriate format, Cora could not get the regression analysis to run. This was because she had inadvertently switched her explanatory and response variables in her program and R cannot create a linear model with a categorical response variable, with the commands she had given the software. After doing some research online, she found an example that she was able to compare to. She couldn't see where she was going wrong with her program, so the following conversation ensued between her and the interviewer:

Interviewer: So the model in the example that you are looking at online, which variable is the explanatory variable and which is the response variable?
Cora: So they're saying <pause> weight is the response variable. <mutters to herself: Does that help me?> Oh! So I did it backwards.

We can see from this example that Cora was able to compare her problem to another solution and make a decision about the way her variables should be used in her analysis. Hence pattern recognition and decision making helped her with her *analysis* phase. Similar conversations occurred with Samantha and Genevieve in their interviews.

New Methods. According to the theoretical framework defined above, new methods to explore concepts involved students not following a prescribed set of procedures to answer a question or coming up with their own methods and reasons for using the technology in analysis. While six instances of new methods were noted in the students' work, specifically in the *implementation* and *research* phases of problem solving, not enough information was found to be able to make generalizations about the use of this statistical computing construct in the problem solving phases. However, an example that comes from Samantha, gives good warning to students who are trying to find new ways to solve problems through research on the internet.

After Samantha had determined that she needed to do a one-way ANOVA procedure, which included first creating the linear model, she found a website that demonstrated how to create a linear model in R. Unfortunately not all websites are created equal. The particular website that she had found gave her the computer code $lm(x{\sim}y)$ (Figure 6.5). From her statistics background, she recognized that x represents the explanatory variable and y represents the response variable. She was able to then match that to the explanatory and response variables that she had defined previously from her problem.

Samantha used this computer code without recognizing the computational error. In this case, having some good statistical knowledge but poor statistical computing sophistication, left the student without the ability to check her solution. Fortunately, due to the nature of the data (Group being categorical and Tips being quantitative) R gave her an error when she later ran her program. Through a guess and check procedure to fix the error that she didn't understand, she switched the variables around and came to the appropriate solution.

R Linear Model Function

`lm()` is a linear model function, such like linear regression analysis.

```
lm(formula, data, subset, weights, ...)
```
`formula`: model description, such as x ~ y
`data`: optional, variables in the model
`subset`: optional, a subset vector of observations to be used in the fitting process
`weights`: optional, a vector of weights to be used in the fitting process

Let's create two vectors, and then fit a linear model:

```
>x <- c(rep(1:20))
>y <- x * 2
>f <- lm(x ~ y)
>f
Call:
lm(formula = x ~ y)
```

Figure 6.5 Website with misleading information that was utilized by Samantha

Discussion

In this section, we discuss what the results mean in terms of the research questions and some implications for teaching that these results may have.

How do students use the technology to solve problems? It is clear from the results of this study that students use the affordance of statistical computing differently across the different phases of problem solving. Three main characteristics of the ways in which students use the technology to solve problems were identified. These are that students use the technology to automate procedures, to produce output to help guide future programming and to do research. These characteristics are given in a decreasing order of preference of attainment. That is, we would most like our students to be able to use the technology to automate procedures since it frees up cognitive capacity to be able to understand other parts of the problem (Chance, Ben-Zvi, Garfield, & Medina, 2007) instead of spending lots of time researching, not making visual forward progress towards a solution.

Experienced students can use technology to automate procedures. While really only one student, Allison, demonstrated that she could use the statistical computing technology to automate procedures, we do have evidence that other students may also be able to do this. It would be ideal if all of our students could automate computational procedures, so that the technology is helping, not hindering their problem solving (Chance, Ben-Zvi, Garfield, & Medina, 2007). However, even Allison was only able to automate her procedures three times during the course of the interview. It appears that while students can get to this point, utilizing automation of computational procedures can be elusive to those that lack experience and understanding in both statistics and computing.

This is important to note, because students that are using statistical computing technologies to solve problems may be struggling to remember the programming necessary to complete the problem, as was seen in several of the examples given above. If they are lacking in either statistical or computational knowledge, the student may spend several minutes just trying to communicate basic statistical pieces to the computer.

As we saw with Cora for example, the order of the response and explanatory variables in the function were keeping her from obtaining a solution. Being able to communicate this to the computer had little to do with a student's statistical knowledge, but it was frustrating for her as she struggled to get the appropriate output from the computer. Remember too, that Cora had taken a course in statistical computing and she still had some difficulty with some of the basic commands in R. If she struggled with programming these commands, students with little to no experience will likely have much more trouble than she did.

Students use computing output to decide what to do next. Despite the struggle that these students face in recalling the syntax of the command-based languages, it is encouraging to know that students who are not yet able to automate procedures still have a method to be able to solve problems using technology. As was discussed above, each of the students in at least one point during their interview utilized a problem solving strategy of *research-planning-implementing-assessment*. We saw that students were using the output from this process, whether correct or incorrect, to help them revise their programming strategy and try something new to work towards a solution. Cora and Moira in particular used this method iteratively several times in their interview, which helped them both solve the problem. Because they were persistent in this method, it did not matter that they were not yet able to automate the computational procedures in R. This strategy helped them come to a solution.

Encouraging students to use this sort of process at least gives them a strategy to solve problems when nothing else seems to be working. Getting to this point still requires having more computational knowledge than some students will have. If the student doesn't have any clue as to what to research, they will struggle to find a program that they can mimic in the remaining problem solving phases. Making sure students have a strong programming basis is therefore important.

Students use the computer to research solution strategies. Finally, when students are unable to automate procedures and they struggle to formulate a problem solving plan, and the students are not able to ask other people for assistance, it appears that students will research (sometimes reluctantly) what needs to be done to solve the problem. As we saw with Samantha and Genevieve, students may spend a lot of time researching when they aren't sure how else to solve the problem. This means that the students are spending less time in the *planning*, *implementation* and *assessment* phases, and not making any visible forward progress towards a solution.

Perhaps with more experience and faith in themselves to complete problems, students will begin *researching* less and start *planning* and *implementing* more. Ideally, students like Samantha and Genevieve could work their way up to continuously using the four phase process described above that both Cora and Moira used.

Implications for Teaching. Statistical computing technologies can be useful for students who are planning future careers in statistics, or who may need the technologies in future classes, hence the appeal for introducing the software in introductory statistics classes (Horton, Baumer, & Wickham, 2014; Baumer, Cetinkaya-Rundel, Bray, Loi, & Horton, 2014). However, starting students on these technologies too soon may be more of a hindrance to their problem solving ability than it is a help.

As should be expected, it appears that students who struggle the least in using statistical computing technologies may be those that are the most comfortable with technology in general. Allison, who was familiar with statistics and computing, Cora, who had taken statistical computing and Moira who was familiar with computer programming, all found solutions to the problem without too much prompting from the interviewer. While we can’t always assure that students will come into our classes with strong computing backgrounds, if we are going to use statistical computing software, there are some implications from this research that can help us to help our students who may have weaker computational backgrounds.

Ideally, we would like our students to “move up the ladder” in their computational sophistication and problem solving strategies. Students who are able to automate procedures, like Allison, are at the top or moving towards the top of the ladder. They know what they need to program and how to program it, meaning that the technology is working for them, not against them. There is always more that the student can learn, so until the student is an expert, they will never really reach the top of the ladder.

The next rung down on the ladder is when a student has a problem solving strategy but cannot automate the computational procedures for the problem, like Cora and Moira. Even though neither of these students knew exactly what needed to be programmed, they each had a problem solving strategy to help them find the code they needed. Helping these students to refine their programming sophistication may help them move towards the next rung of the ladder. One may suggest to students that are at this point that they try programming without their notes and without looking at other people's code. In doing so, students may start to remember code and in the future automation of procedures will be more likely.

The step below on our metaphorical ladder is when the student does not possess the sophistication to program a solution for the problem and they do not have a problem strategy to assist them in obtaining appropriate code. We saw that with Samantha and Genevieve, they spent a lot of time researching how to do the problem and they needed a lot of assistance from the interviewer to be able to find a solution to the problem. Suggesting a problem solving strategy to these students, like the one that Moira and Cora utilized, could help students when they get stuck in the future. However Nolan and Temple Lang (2010) warn that for many students it can be difficult for students to go from basic programing skills to embracing a problem solving strategy. It takes time and experience for them to get to the point that Cora and Moira were at.

The interviews also demonstrated that when a student becomes stuck while solving a problem, they may easily get frustrated (Nolan & Temple Lang, 2010) and ask for help before really trying anything on their own. This was evident in the interviews when Genevieve became verbally agitated that I would not answer her questions. At one point when she was struggling, I asked her what her normal strategy would be, when unable to solve a problem. She told me that she would ask the teacher for help. However, this seems to have hindered her ability to program in the long run. We need to encourage students like this to struggle through problems on their own before seeking help to build their persistence in problem solving. While we are a resource to our students, we should not become a crutch. Suggesting to students a problem solving strategy of *research-planning-implementing* and *assessment* might be helpful for them to being programming on their own and move up on the ladder.

Future Research. While we were able to determine some patterns in the ways that students solve problems using statistical computing, more research could be done on this problem. One of the complications in determining what was helping students to do well or not do well in the interviews was the mixed level of statistical and computational knowledge. Since students were at different levels of each, it was difficult at times to tell what was actually helping or hindering them from solving problems.

Future studies would use students with very similar statistical backgrounds, but potentially different computing backgrounds to see how the differences in computing skill affects one's sophistication in solving the statistical problems. This would be helpful to confirm or refute results that were found above.

Chapter 7 : Summary and Conclusion

The purpose of this final Chapter is to draw together the information in the three individual manuscripts and to present the overall findings to answer the research questions. Implications for teaching and research conclude the paper.

Summary

As we saw in Chapter 4, writing assignments can be useful tools for students, instructors and researchers. We saw that in the past, these assessments had been used to help students conceptualize topics, to aid instructors in assessing student understanding and to inform researchers as to some of the ways in which students learn statistics concepts.

A new method for helping students to write concise, meaningful responses to statistical questions, called the four sentence structure, was given. This method asked students to

- Answer the question;
- State relevant facts from the problem;
- State the implications that the facts imply; and
- Explain how the facts lead to the conclusion.

This method provided students in a second course in statistics with a medium to help demonstrate their statistical thinking while utilizing computing technologies to solve statistics problems. Example writing assignments, rubrics for grading and some student responses were given so that the reader had an idea of how to structure problems, responses they might expect and ways to grade, if they were to utilize similar prompts with their students.

Trends in the writing assignment grades for the students across the semester demonstrate that students' scores were, on average, increasing. This shows that using this consistent method of assessment either helped students to improve their writing and communication, their statistical thinking, or both.

In Chapter 5 we reviewed the written work of students on several assignments that involved the use of statistical computing. Themes between how students utilized statistical computing and statistical thinking together (or not) were identified. Some of these themes included

- Students' written work does not always provide evidence of statistical computing.
- Students' arguments use statistical computing to provide *Facts*, not *Implications*.
- Students' written arguments may demonstrate correct statistical thinking and statistical computing.
- Students sometimes apply incorrect methods or have incorrect computational thinking in their evidence for arguments.
- Students' correct application of statistical computing does not always lead to correct statistical thinking due to pitfalls in student's work.

In addition to these themes, the types of statistical thinking and statistical computing that student's use in arguments were also identified. It was noted that automation for computational procedures was the type of statistical computing that was most often found in students' written work. However, one of the problems that we find with writing assignments is that they do not give us a full picture of the process that students take to get to their solution (Lee, et al., 2014). Even though automation for computational procedures was the statistical computing affordance that was coded the most, it appears that when students have the opportunity to report back on what they have done to complete a problem, they may not accurately report all steps they took to reach a solution.

This research shows that students tend to write about only the process that worked to help them get their solution, not the pitfalls they faced in trying to solve the problem. This is likely due to the way the problems were structured, asking for concise answers, as opposed to detailed reports. While the four sentence structure was useful in helping to assess students and to inform the instructor of topics with which students struggled, it may have had some flaws when informing on the types of statistical computing students used. As such, to understand more about how students solve problems, we needed to see them in action.

In Chapter 6, analysis and results from task-based interviews were presented to gain a better understanding of how students use statistical computing and in which of the problem solving phases they were most likely to utilize these techniques.

One of the most prominent findings, holding contrary to what was observed in student written work, was that automation of computational procedures does not seem to be a common statistical computing tool after all. In fact, the only student who used it was Allison, who had more experience with statistics and computing than the other students who participated in the interviews. To work towards a solution, other students had to struggle with a repetitive sequence of *research*, *planning*, *implementation* and *assessment* or spend long periods of time making sense of example code that was found on the internet. Ultimately, it appears as if it is only through hindsight that students describe what they did with ease, which appears, perhaps mistakenly to a researcher, as automation of computational procedures in their written work.

The interviews gave us a better understanding of how students use statistical computing in a problem solving setting. The ways in which the technology was used seemed highly correlated with the amount of experience the student had with both statistics and computing. These methods included:

- Automating computational procedures
- Using the computer to perform a sequence of research, planning, implementing and assessment
- Spending large amounts of time researching and making sense of output found on the internet

Answering the Research Questions

At the beginning of this paper, two research questions were posed.

What are the affordances and constraints of how students use statistical computing to support their statistical thinking? The first thing of note is that despite growing up in a technology intense world, not all students naturally use the computer to solve problems. Several students did not use technology to solve the first solo writing assignment given to them (assignment 3), yet they were still able to think about the variation in the data. It is unclear as to whether these students didn't realize they needed to use the technology or if they were actively avoiding it to solve the problem. Either way, this may demonstrate that students don't inherently believe that computing is necessary to solve statistical problems.

Once students became more accustomed to using the technology to solve problems, we were then able to identify how students use the technology to solve statistical problems and how it affected a student's statistical thinking. While the writing assignments and the task-based interviews may have given some conflicting perspectives on the use of statistical computing tools, we can still see that students are capable of using statistical computing to inform their statistical thinking.

The types of statistical computing that were identified in students' written work in Chapter 5 may not be entirely accurate. We saw that automation of computational procedures was by far the type of statistical computing most often utilized by students, with computational thinking being a distant second. However, in reviewing the task-based interviews, computational thinking was found to be, by far, the most commonly used type of statistical computing and automation of computational procedures was the least. As was discussed above, this discrepancy in findings is likely due to the fact that when students write about their work, it is through hindsight that they describe what they did with ease, appearing as automation.

While there is a difference in how the computing actions should be coded, in general, the results defining the relationship between statistical computing and statistical thinking, in general, should hold.

From the writing assignments we found that when students were able to use statistical computing affordances, they were able to give logical arguments that demonstrated a great amount of statistical thinking. In Figure 7.1, we can see that statistical computing is informed by, works in conjunction with and leads to, statistical thinking. It is likely that many of the instances coded in the students' documents as automation for computational procedures would have been coded as computational thinking had we been able to observe the students in action. However since the relationship defined in Figure 7.1 merely discusses statistical computing in general, the exchange of codes would not change the relationship discussed here.

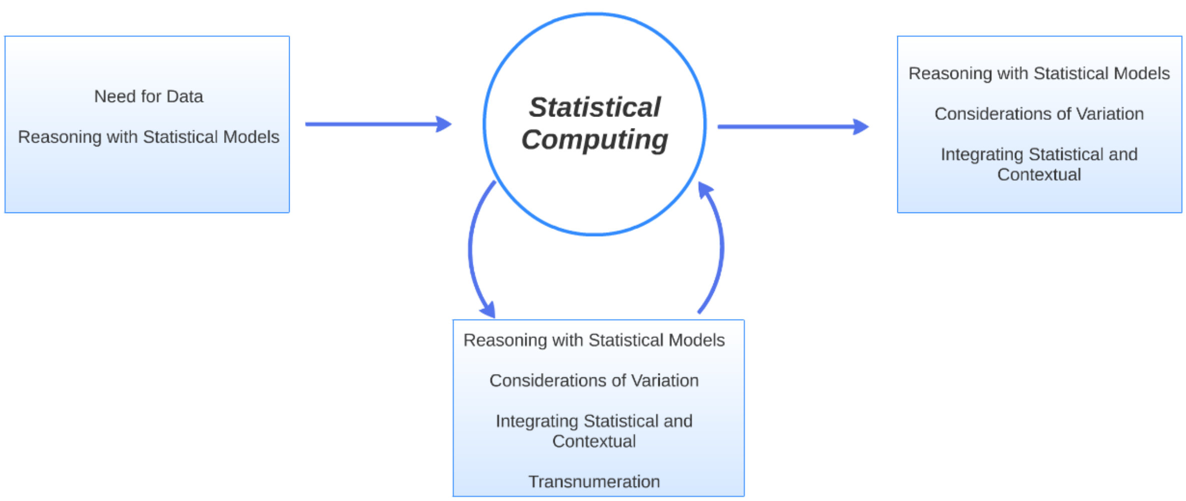


Figure 7.1 Relationship between statistical thinking and statistical computing when correct thinking and computing are utilized

Despite the results presented in Figure 7.1, this does not guarantee that a student who is good at statistical computing will produce logical results or think statistically. In fact, we can see in Figure 7.2 that several students were able to use the computer to find statistical output however they also trusted the computer too much, used their intuition of the context over the output of the data, or just had weak statistical facts, tended to produce incorrect arguments.

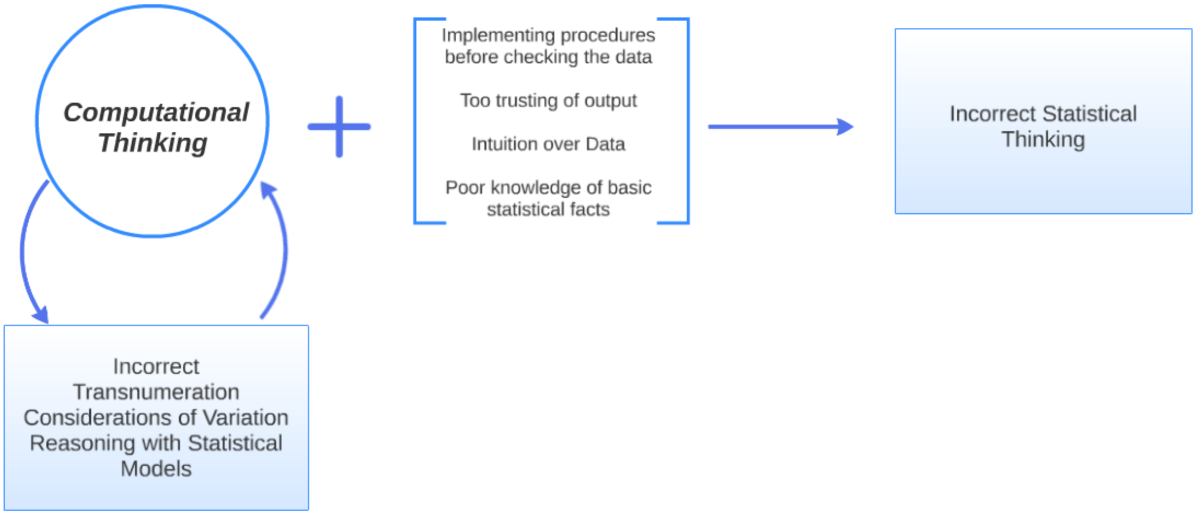


Figure 7.2 Relationship between statistical thinking and statistical computing when pitfalls are present in students' work.

This tells us that outside factors, such as implementing inferential procedures before checking the data and being too trusting of the results of the computer, can have an impact on the relationship that exists between thinking and computing. For example, we saw that in writing assignment 6, Chloe was too trusting of the output. She did not first check conditions for inference for regression and obtained incorrect results for the prompt because she found the regression output with an outlier present in the data set. In other cases, we saw that even though students were able to use the computer to produce appropriate output, they did not actually use the data to help them answer the question. Cora did this in assignment 7 when she found the regression line but did not actually use it to support her claim that Gary should ask out a woman shorter than him. These examples show us that statistical computing is only helpful to statistical thinking if students are able to overcome their poor problem solving habits.

Both the writing assignments and the task based interviews also demonstrated that not all students were able to fully utilize the affordances of the statistical computing tools to help solve problems. For example, some students were able to program their work but were unable to properly explain the strategy that they used. We saw an example of this with Cora, when she inaccurately described what was going on behind the scenes in her simulation with chi-square statistics. Another example of this was when students had a good strategy for completing the problem, but had incorrect programming, an example of which was seen in Samantha's work when she knew a regression analysis was necessary, but created the wrong regression line for the data. The third way this was observed was when students had a poor strategy and did not know how to program the problem. We saw an example of this when Moira found code online to help her with her Bonferonni adjustment simulation.

When students fell into the first two categories, which can be thought of as having a partial understanding of the statistical computing tools, they were still able to think and reason statistically. Not understanding exactly what her code was doing didn't stop Cora from being able to make a decision about whether or not the data showed that Hans was cheating. Or when Samantha produced the wrong regression line for her scatterplot, she wasn't able to give a full solution to the problem, but she was able to discuss what the ramifications of an outlier in a regression analysis would be. These tools still helped students to produce output, and make decisions with data, even if they did not answer the question prompt accurately.

The problem occurs when students fall under the third category. We saw examples of students, such as Moira using the p.adjust function in R, that did not have an appropriate strategy to program the computer, researched the internet for example code and ended up producing a completely inaccurate result after finding something that looked vaguely helpful on the internet. In this situation the students were not able to solve the problem and did not demonstrate statistical thinking. While this didn't appear to be a large problem in the writing assignments submitted for class, after viewing the task-based interviews, this appears like it might be fairly problematic for students.

Overall this research shows us that in solving statistical problems, statistical computing is highly dependent on statistical thinking and that statistical thinking is highly dependent on statistical computing. In order for our students to be successful problem solvers in statistics, they need to be strong and confident in both statistical thinking and computing.

In what ways are statistical computing techniques used by students in a problem solving setting and how do differences in statistical computing sophistication affect students' abilities to solve statistical problems with statistical computing technologies? Making generalizations about students' ability to utilize statistical computing based on computing sophistication alone is a bit difficult from the data that was collected in this research. As was noted above, the five student volunteers did have different computing sophistication, but they also had different statistical sophistication. Distinguishing between the two can be difficult, however some generalizations about what was found based on statistical and computational sophistication can be made.

We saw in Chapter 6 that the most advanced students both statistically and computationally, were able to automate their computational procedures and use the technology to answer the questions they were asked. However, this only occurred with one student in the task-based interviews, Allison. Despite her extensive background in both statistics and computing, she was only able to automate procedures three times during the first phase of her interview. This tells us that even our most experienced students may never reach a level of expertise where they feel comfortable enough to automate procedures to complete statistical tasks.

The next step down from this would be students that have a decent amount of experience with statistics, computing or both. Cora and Moira fit into this categorization of students from the task based interview group. From their interviews, it appeared that students with less than expert experience tend to struggle with the technology and it seemed to be more of a hindrance than a help in some situations. Fortunately, students that have some in depth understanding, but perhaps not an expert level, in computing or statistics (or both), seemed to be able to come up with a strategy to solve the problem, which involved iterating the problem solving phases of *research*, *planning*, *implementation* and *assessment*. While these students did not come to solutions in an expeditious manner, they were able to use the technology to help them.

Not all students are at this level of problem solving. In the task based interviews, we saw two students, Genevieve and Samantha, with some, yet limited experience in computing and statistics. When solving the problem they were given, both of these students struggled to find a strategy to solve the problem and had even less understanding of how to converse with the computer. They spent a lot of time researching the internet for example code and did not make a lot of forward progress towards a solution. Genevieve even became verbally agitated that I would not give her assistance on the problem. Hence, the technology was a major hindrance to both of them, due to lack of experience or lack of problem solving perseverance.

Implications for Teaching

A discussion on some of the implications for teaching gained from this research are given in this section. In particular, we will look at when it may be appropriate to teach statistical computing and how statistical computing impacts a student's ability to solve statistical problems.

When to teach statistical computing. There are benefits to utilizing statistical computing software in the classroom. First, as this research has shown, once students are proficient with the statistical computing technology, it can help them to solve problems and think statistically. Additionally, when these technologies are used in courses that tend to be taken earlier in the curriculum, we are offering our students training to use the computer to solve problems in preparation for upper level classes and potentially their future careers. By having the students practice with the statistical computing technology in lower level courses, this may help them to build their skill set and hopefully, get to a point where they can use the computing software without much difficulty.

However, there is a much larger learning curve to be able to use statistical computing technologies efficiently versus using graphical user interface software. Introducing the command based technology too early may have the consequence of frustrating students or keeping them from understanding key statistical concepts because they are too busy trying to learn the nuances of the technology.

Instructors should be mindful of their choice in technology when teaching statistics. If students will not need the knowledge of the computing technology in future classes, it might be better to use a tool meant for teaching statistics. Teachers should also make sure to set aside enough time in class to teach students to use the computing technology appropriately. This should be done to make sure that students aren't picking up bad computational or statistical habits once they are using the technology on their own to solve problems.

Statistical computing and problem solving. It is probably of little surprise that students with varying levels of skill in statistics and computing will perform statistical problem solving tasks differently. However, one of the main takeaways from this research is that that even students with a lot of experience may struggle to use statistical computing technologies to solve statistical problems. We saw that Allison was able to solve her task-based interview problem, but even with her experience, she struggled at times to devise a solution. Students with much less experience may become frustrated and just not do the problem.

The key to building our student's programming sophistication is to practice. However, if the less experienced students become frustrated and quit, or ask for help from others too quickly, they will likely not learn. Providing these students with a problem solving strategy utilized by our middle tier students: research-plan-implement-assess, may be the advice the lower tier students need to work through these problems. The students need to try something, as opposed to staring at potential code for long periods of time. If that code doesn't work, then try again. Learn from the computer output and try to make forward progress to a solution. Finally, don't let your students give up when they are frustrated.

Implications for Research

In this section we will discuss some of the implications for research based on observations from this work. In particular, the usefulness of the statistical computing framework and the four sentence structure will be discussed.

Statistical Computing Framework. In the literature review for this paper a framework was created to determine when students were utilizing statistical computing in their work. It was found that there are some flaws when using the framework to determine precisely which types of statistical computing are being used in written work. However, the framework seemed to do a good job of identifying and classifying the types of statistical computing in the task-based interviews.

Those wishing to use the framework for their own purposes should be mindful of the drawbacks for assessment in written work since students tend to write about their problem solving process in hindsight. This tends to make automation for computational procedures appear more prominent than it actually is. Comparing the results from the interviews to the written work also indicates that Computational Thinking is also probably more prominent in student work than they are letting on in their writing, but New Methods and Pattern Recognition/Decision Making seemed to be measured consistently in both assessments.

As such a revision to the framework should be made to account for these differences. Specifically, the affordances for automating computational procedures shouldn't focus on the product the students present, but the actions they took to get there (Nolan & Temple Lang, 2010). Being able to create the graphics and summaries to inform analysis are better classified under computational thinking, as the students needed to create a strategy and convey that strategy to the computer to create those graphics. Since we only know that the student was able to create the graphs, not that they maintained a reduced cognitive load to create them, we cannot assume that the student was able to automate the procedures when using the computer to make the graph. Finally, based on information gained from the interviews, computational thinking should also include using the technology to test and revise solution strategies. These modifications can be found in Table 7.1.

Table 7.1 Revised framework for identifying affordances of statistical computing

Aspect of Statistical Computing	**Affordance for the student**
Automate computational procedures	• uses the technology to conduct the analysis, then makes proper inferences • uses the technology to maintain a reduced cognitive load
Increase computational thinking	• creates a solution strategy and can communicate it to the computer • creates multiple graphs or summaries to make sense of the data • uses the technology to test and revise solutions strategies • is able to think critically or abstractly about a concept that the technology is demonstrating
Offer new methods to explore concepts	• does not follow a prescribed set of procedures to answer a question • comes up with their own methods and reasons for using the technology in analysis
Aid in pattern recognition and decision making	• recognizes patterns, either in the way things are coded or in how the results behave • uses the output from the technology to influence their decision of where to go next in the analysis

The Four Sentence Structure. Despite the inability of the initial statistical computing framework to accurately categorize the types of computing used in students' written work, using the four sentence structure was very helpful in assessing students' statistical thinking. This was due to the inclusion of an *implications* step, where students had to make connections they might not have otherwise stated. Researchers interested in knowing how students think statistically about particular concepts might find this assignment structure, and the methods discussed in Chapter 5 to analyze the qualitative data, to be helpful.

An additional idea to help instructors with this assessment, would be to ask students to screencast their work as they attempt a problem. Students would submit this video file along with their written work. This would not only help the instructor to see the ways in which students solve problems and the ways in which they utilize statistical computing, but it would also give researchers a way to see how students perceive their problem solving process through their writing, versus what they actually do in their video.

Future Research

As was discussed above, the original statistical computing framework created and utilized in this research may not be entirely accurate for written work. It would be interesting to see how the suggested modifications to the framework Table 7.1 affects the results found when analyzing students' written work. One way to test this may be to have students create a screen capture of their work as they solve a statistical problem, then have them write about the process using the four sentence structure. Research could then be done to compare instances of statistical computing found in students' written work to the actions they actually took, found in their screencasts. This has the additional benefit of collecting data that could be used in a comparison study to see how students portray their work in hindsight versus how they actually complete a problem. Understanding that connection could be useful to teachers and researchers trying to use written work for assessments.

Another point of research would be to see if the statistical computing framework defined in this research produces similar results with other types of statistical computing technologies. The only statistical computing technologies used in this research were R and RStudio. It might be that using other command based statistical computing technologies will produce different results. In order for large scale generalizations to be made about statistical computing, the reliability of the framework needs to be tested with other programming languages.

Finally, the research presented here focused on a small set of students, representing a convenience sample. Because of the restrictions in sampling, conducting a quantitative study would not have been appropriate. It may be of interest to reproduce the methods used in this research with a larger set of randomly selected students. Better connections between statistical thinking and statistical computing may then be defined, and potential differences between more experienced students, say those who had taken a statistical computing course prior to the research, versus less experienced students could also be made.

References

Ainley, J., Gould, R., & Pratt, D. (2015). Learning to reason from samples: Commentary from the perspectives of task design and the emergence of "big data". *Educational Studies in Mathematics, 88*(3), 405-412.

American Statistical Association. (2005). *Guidelines for assessment and instruction in statistics education.* American Statistical Association.

American Statistical Association. (2016). *Guidelines for Assessment and Instruction in Statistics Education (GAISE) College Report.* American Statistical Association.

Baglin, J. (2013). Applying a theoretical model for explaing the development of technological skills in statistics education. *Technology Innovations in Statistics Education, 7*(2).

Barrouillet, P., & Gauffroy, C. (2013). *The Development of Thinking and Reasoning.* Psychology Press.

Baumer, B. (2015). A data science course for undergraduates: Thinking with data. *The American Statistician, 69*(4), 334-342.

Baumer, B., Cetinkaya-Rundel, M., Bray, A., Loi, L., & Horton, N. J. (2014). R Markdown: Integrating a reproducible analysis tool into introductory statistics. *Technology Innovations in Statistics Education, 8*(1).

Ben-Zvi, D. (2004). Reasoning about data analysis. In D. Ben-Zvi, & J. Garfield (Eds.), *The Challenge of Developing Statistical Literacy, Reasoning and Thinking* (pp. 121 - 145). Dordrecht: Kluwer Academic Publishers.

Ben-Zvi, D. (2006). Scaffolding students' informal inference and argumentation. *Proceedings of the Seventh International Conference on Teaching Statistics*, (pp. 1 - 6).

Ben-Zvi, D., & Garfield, J. (2004). Statistical literacy, reasoning and thinking: Goals, definitions, and challenges. In D. Ben-Zvi, & J. Garfield (Eds.), *The Challenge of Developing Statistical Literacy, Reasoning and Thinking* (pp. 3 - 15). Dordrecht: Kluwer Academic Publishers.

Biehler, R. (1993). Sofware tools and mathematics education: The case of statistics. In C. Keitel, & K. Ruthven (Eds.), *Learning from Computers: Mathematics Education and Technology* (pp. 68 - 100). Berlin: Springer.

Biehler, R. (2007). Assessing students' statistical competence by means of written reports and project work. *Proceedings of the IASE Satellite Conference on Assessing Student Learning in Statistics*, (pp. 1 - 8). Guimaraes, Portugal.

Biehler, R., Ben-Zvi, D., Bakker, A., & Makar, K. (2012). Technology for enhancing statistical reasoning at the school level. In *Third International Handbook of Mathematics Education* (pp. 643-689). New York: Springer.

Bolte, L. A. (1999). Using concept maps and interpretive essays for assessment in mathematics. *School Science and Math, 99*(1), 19-30.

Box, G. E., & Draper, R. (1987). *Empirical Model-Building and Response Surfaces* (Vol. 424). New York: Wiley.

Brown, E. N., & Kass, R. E. (2009). Statistical training and curricular revision. *The American Statistician*, *63*(2), 105-110.

Burns, M. (2004). Writing in math. *Educational Leadership*, *62*(2), 30-33.

Chance, B. (1997). Experiences with authentic assessment techniques in an introductory statistics course. *Journal of Statistics Education*, *5*(3).

Chance, B. (2002). Components of statistical thinking and implications for instruction and assessment. *Journal of Statistics Education, 10*(3).

Chance, B., Ben-Zvi, D., Garfield, J., & Medina, E. (2007). The role of technology in improving student learning of statistics. *Technology Innovations in Statistics Education, 1*(1), 1-26.

Chance, B., Cohen, S., Grimshaw, S., Hardin, J., Hesterberg, T., Hoerl, R., . . . Nolan, D. (2014). *Curriculum guidelines for undergraduate programs in statistical science.* American Statistical Association.

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* (pp. 295-323). Dordrecht: Kluwer Academic Publishers.

Chinn, C. A., & Brewer, W. F. (2001). Models of data: A theory of how people evaluate data. *Cognition and Instruction, 19*(3), 323-393.

Cline, K. S. (2008). A writing-intensive statistics course. *PRIMUS, 18*(5), 399-410.

Creswell, J. W. (2013). *Qualitative Inquiry & Research Design.* Sage Publications Inc.

delMas, R. (2004). A comparison of mathematical and statistical reasoning. In D. Ben-Zvi, & J. B. Garfield (Eds.), *The Challenge of Developing Statistical Literacy, Reasoning and Thinking* (pp. 79 - 95). Dordrecht, The Netherlands: Kluwer Academic Publishers.

delMas, R. C., Garfield, J., & Chance, B. L. (1999). A model of classroom research in action: Developing simulation activities to improve student's statistical reasoning. *Journal of Statistics Education, 7*(3).

delMas, R., Garfield, J., & Zieffler, A. (2014). Using TinkerPlots to develop tertiary students' statistical thinking in a modeling-based introductory statistics class. In *Mit Werkzeugen Mathematik und Stochastik lernen - Using Tools for Learning Mathematics and Statistics* (pp. 405-420). Springer Fachmedien Wiesbaden.

Dierker, L., Kaparakis, E., Rose, J., Selya, A., & Beveridge, D. (2012). Strength in numbers: A multidisciplinary, project-based course in introductory statistics. *The Journal for Effective Teaching, 12*(2).

Drabick, D. A., Weisberg, R., Paul, L., & Bubier, J. L. (2007). Keeping it short and sweet: Brief, ungraded writing assignments facilitate learning. *Teaching of Psychology, 34*(3), 172-176.

Finzer, W. F. (2000). Design of Fathom, a dynamic statistics environment, for the teaching of mathematics. *Ninth International Congress on Mathematics Education.* Tokyo, Japan.

Fitzallen, N. E. (2013). Characterising students' interaction with TinkerPlots. *Technology Innovations in Statistics Education*, *7*(1), 1 - 19.

Gal, I. (2004). Statistical literacy: Meanings, components, responsibilities. In D. Ben-Zvi, & J. Garfield (Eds.), *The Challenge of Developing Statistical Literacy, Reasoning and Thinking* (pp. 47-78). Netherlands: Springer.

Gal, I., Ginsburg, L., & Schau, C. (1997). Monitoring attitudes and beliefs in statistics education. In I. Gal, & J. B. Garfield (Eds.), *The Assessment Challenge in Statistics Education* (pp. 37-51). International Statistical Institute.

Garfield, J. B. (1994). Beyond testing and grading: Using assessment to improve student learning. *Journal of Statistics Education, 2*(1), 1- 11.

Garfield, J., & Ben-Zvi, D. (2008). Assessment in statistics education. In *Developing Students' Statistical Reasoning* (pp. 65-89). Springer Science & Business Media.

Garfield, J., & Ben-Zvi, D. (2008). *Developing students' statistical reasoning: Connecting research and teaching practice.* Springer Science & Business Media.

Garfield, J., delMas, R., & Zieffler, A. (2012). Developing statistical modelers and thinkers in an introductory, tertiary-level statistics course. *ZDM, 44*(7), 883-898.

Gentle, J. E. (2004). Courses in statistical computing and computational statistics. *The American Statistician*, *58*(1), 2 - 5.

Gentleman, R. (2004). Some perspectives on statistical computing. *The Canadian Journal of Statistics*, *32*(3), 209-226.

Goenner, C. F., & Snaith, S. M. (2003). *Incorporating writing into an introductory business and economics statistics course: A practical approach.*

Goldin, G. A. (1997). Chapter 4: Observing mathematical problem solving through task-based interviews. *Journal for Research in Mathematics Education. Monograph*, 40 - 62; 164 - 177.

Gopen, G. D., & Smith, D. A. (1990). What's an assignment like you doing in a course like this?: Writing to learn mathematics. *College Mathematics Journal, 21*(1), 2-19.

Grice, H. P. (1975). Logic and Conversation. In P. Cole, & J. Morgan (Eds.), *Speech Acts, Syntax and Semantics 3* (pp. 41 - 58). New York: Elsevier.

Groth, R. E. (2010). Situating qualitative modes of inquiry within the discipline of statistics education research. *Statistics Education Research Journal, 9*(2), 7-21.

Grover, S., & Pea, R. (2013). Computational thinking in K-12: A review of the state of the field. *Educational Researcher, 42*(1), 38-43.

Hardin, J., Hoerl, R., Horton, N., Nolan, D., Baumer, B., Hall-Holt, O., & Ward, M. D. (2015). Data science in statistics curricula: Preparing students to "think with data". *The American Statistician, 69*(4), 343-353.

Hayden, R. (1997). Using writing to improve student learning of statistics. *Writing Across the Curriculum, 8*, 198 - 205.

Holcomb, J. P., & Ruffer, R. L. (2000). Using a term-long project sequence in introductory statistics. *The American Statistician, 54*(1), 49-53.

Holcomb, J., Chance, B., Rossman, A., & Cobb, G. (2010). Assessing student learning about statistical inference. *Proceedings of the 8th International Conference on Teaching Statistics.*

Holcomb, J., Chance, B., Tietjen, A. R., & Cobb, G. (2010). Introducing concepts of statistical inference via randomization tests. *Proceedings of the Eighth International Conference on Teaching Statistics.* Voorberg: International Statistical Institute.

Horton, N. J., Baumer, B. S., & Wickham, H. (2014). Teaching precursors to data science in introductory and secondary courses in statistics. *Proceedings of the Ninth International Conference on Teaching Statistics.* Flagstaff Arizona.

Inglis, M., Meija-Ramos, J. P., & Simpson, A. (2007). Modeling mathematical argumentation: The importance of qualification. *Educational Studies in Mathematics, 66*(1), 3 - 21.

Johnson, K. G. (2016). Incorporating writing into statistics. In *Mathematics Education* (pp. 319-334). Springer International Publishing.

Johnson, M. L. (1983). Writing in mathematics classes: A valuable tool for learning. *The Mathematics Teacher, 76*(2), 117-119.

Jolliffe, F. (2007). The changing brave new world of statistics assessment. *IASE/ISI.*

Kafura, D., Bart, A. C., & Chowdhury, B. (2015). Design and preliminary results from a computational thinking course. *2015 ACM Conference on Innovation and Technology in Computer Science Education*, (pp. 63 - 68). ACM.

Kaplan, D., Horton, N. J., & Pruim, R. (2015). *Randomization-based inference using the mosaic package.*

Karbach, J. (1987). Using Toulmin's model of argumentation. *Journal of Teaching Writing, 6*(1), 81 - 92.

Keys, C. W. (1994). The development of scientific reasoning skills in conjunction with collaborative writing assignments: An interpretive study of six ninth-grade students. *Journal of Research in Science Teaching, 31*(9), 1003-1022.

Koichu, B., & Harel, G. (2007). Triadic interaction in clinical task-based interviews with mathematics teachers. *Educational Studies in Mathematics, 65*(3), 349 - 365.

Konold, C. (2007). Designing a data analysis tool for learners. In M. C. Lovett, & P. Shah (Eds.), *Thinking with Data* (pp. 267-291).

Lane-Getaz, S. J. (2010). Linking the randomization test to reasoning about p-values and statistical significance. *Eighth International Conference on Teaching Statistics.* Ljubjana, Slovenia: Voorburg, The Netherlands.

Lee, H. S., & Hollebrands, K. F. (2006). Students' use of technological features while solving a mathematics problem. *The Journal of Mathematical Behavior, 25*(3), 252-266.

Lee, H. S., Kersaint, G., Harper, S. R., Driskell, S. O., Jones, D. L., Leatham, K. R., . . . Adu-Gyamfi, K. (2014). Teachers' use of transnumeration in solving statistical tasks with dynamic statistical software. *Statistics Education Research Journal, 13*(1), 25-52.

Lovett, M. C. (2001). A collaborative convergence on studying reasoning processes: A case study in statistics. In S. M. Carver, & D. Klahr (Eds.), *Cognition and Instruction: Twenty-five years of progress* (pp. 347-384). Mahway, New Jersey: Lawrence Erlbaum Associates, Inc.

Makar, K., & Rubin, A. (2009). A framework for thinking about informal statistical inference. *Statistics Education Research Journal, 8*(1), 82-105.

McCullagh, P. (2002). What is a statistical model. *The Annals of Statistics, 30*(5), 1225-1310.

Monahan, J. (2004). Teaching statistical computing at North Carolina State University. *The American Statistician, 58*(1), 6-8.

Moore, D. S. (1990). Uncertainty. In L. A. Steen (Ed.), *On the shoulders of giants: New approachs to numeracy* (pp. 95-137). New York, New York: Andrew W. Mellon Foundation.

Moore, D. S. (1997). New pedagogy and new content: The case of statistics. *International Statistical Review*, 123-137.

Moore, D. S. (2010). *The Basic Practice of Statistics.* New York, NY: W. H. Freeman and Company.

Nahrgang, C. L., & Petersen, B. T. (1986). Using writing to learn mathematics. *The Mathematics Teacher, 79*(6), 461-465.

National Council of Teachers of Mathematics. (2000). *Principles and Standards for School Mathematics.* Reston, VA: NCTM.

Nguyen, N. T., McFadden, A., Tangen, D., & Beutel, D. (2013). Video-stimulated recall interviews in qualitative research. *Australian Association for Research in Education*, (pp. 1 - 10).

Nolan, D., & Temple Lang, D. (2010). Computing in the statistics curricula. *The American Statistician, 64*(2), 97-107.

Nosek, J. T. (1998). The case for collaborative programming. *Communications of the ACM, 41*(3), 105 - 108.

Park, M.-S., Park, M., Lee, E.-J., & Lee, K.-H. (2016). Preservice teachers' difficulties with statistical writing. In D. Ben-Zvi, & K. Makar (Eds.), *The Teaching and Learning of Statistics* (pp. 261 - 270). Springer.

Parke, C. S. (2008). Reasoning and communication in the language of statistics. *Journal of Statistics Education, 16*(1).

Pea, R. D. (1987). Cognitive technologies for mathematics education. *Cognitive science and mathematics education*, 89-122.

Peck, R. (2005). There's more to statistics than computation - Teaching students how to communicate statistical results. *IASE Satellite Conference on Statistics Education and the Communication of Statistics*, (pp. 1 - 4).

Pfannkuch, M., & Wild, C. (2004). Towards an understanding of statistical thinking. In D. Ben-Zvi, & J. Garfield (Eds.), *The Challenge of Developing Statistical Literacy, Reasoning and Thinking* (pp. 17-46). Dorcrecht: Kluwer Academic Publishers.

Pfannkuch, M., Regan, M., Wild, C., Budgett, S., Forbes, S., Harraway, J., & Parsonage, R. (2011). Inference and the introductory statistics course. *International Journal of Mathematical Education in Science and Technology, 42*(7), 903-913.

Pfannkuch, M., Wild, C., & Regan, M. (2014). Student's difficulties in practing computer-supported statistical inference: Some hypothetical generalizations from a study. In T. W. al., *Using Tools for Learning Mathematics and Statistics* (pp. 393 - 403). Springer.

Philip, T. M., Schuler-Brown, S., & Way, W. (2013). A framework for learning about big data with mobile technologies for democratic participation: Possiblities, limitations and unanticipated obstacles. *Technology, Knowledge and Learning*, *18*(3), 103-120.

Powell, A. B., Francisco, J. M., & Maher, C. A. (2003). An analytical model for studying the development of learners' mathematical ideas and reasoning using videotape data. *Journal of Mathematical Behavior*, 405-435.

Radke-Sharpe, N. (1991). Writing as a component of statistics education. *The American Statistician*, *45*(4), 292-293.

Reading, C., & Reid, J. (2006). An emerging hierarchy of reasoning about distribution: From a variation perspective. *Statistics Education Research Journal*, *5*(2), 46-68.

Rossman, A. J., & Chance, B. L. (2004). Anticipating and addressing student misconceptions. *ARTIST Conference on Assessment in Statistics*, (pp. 1-4). Lawrence University.

Shibli, A. (1992). Increasing learning with writing in quantiative and computer courses. *College Teaching*, *40*(4), 123 - 127.

Steffe, L. P., & Olive, J. (2002). Design and use of computer tools for interactive mathematical activity (TIMA). *Journal of Educational Computing Research*, *27*(1), 55-76.

Stromberg, A. J., & Ramanathan, S. (1996). Easy implementation of writing in introductory statistics courses. *The American Statistician*, *50*(2), 159-163.

Tintle, N., Chance, B., Cobb, G., Roy, S., Swanson, T., & Vanderstoep, J. (2015). Combating anti-statistical thinking using simulation-based methods throughout the undergraduate curriculum. *The American Statistician*, *69*(4), 362-370.

Tintle, N., VanderStoep, J., Holmes, V.-L., Quisenberry, B., & Swanson, T. (2011). Development and assessment of a preliminary randomization-based introductory statistics curriculum. *Journal of Statistics Education, 19*(1).

Toulmin, S. E. (1958). *The Uses of Argument.* Cambridge: Cambridge University Press.

Van Dyke, F., Malloy, E. J., & Stallings, V. (2015). Conceptual writing in college-level mathematics courses and its impact on performance and attitude. *International Journal of Mathematical Education in Science and Technology, 46*(2), 223-233.

Watson, J. (2014). TinkerPlots as an interactive tool for learning about resampling. In *Mit Werkzeugen Mathematik und Stochastik learnen - Using Tools for Learning Mathematics and Statistics* (pp. 421-436). Springer Fachmedien Wiesbaden.

Weldon, L. K. (2007). Assessment of a writing course in statistics. *IASE Satellite Conference on Assessing Student Learning in Statistics.*

West, W., Wu, Y., & Heydt, D. (2004). An introduction to StatCrunch 3.0. *Journal of Statistical Software, 9*(i05).

Wild, C. J., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review, 67*(3), 223 - 248.

Wing, J. M. (2008). Computational thinking and thinking about computing. *Philosophical Transactions of the Royal Society A, 336*(1881), 3717- 3725.

Wolfram, S. (2016, September 7). How to teach computational thinking. Retrieved from http://blog.stephenwolfram.com/2016/09/how-to-teach-computational-thinking/

Woodard, R. D. (2016, August). Emphasizing critical thinking in introductory statistics. *The Joint Statistical Meetings.* Chicago.

Yuen, T. T., & Robbins, K. A. (2014). A qualitative study of students' computational skills in a data-driven computing class. *ACM Transactions on Computing Education, 14*(4).

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: Writing Assignments

Assignment 1 – I Z what you did there

An IT staff member at a large company in Raleigh was asked to determine the average amount spent per laptop for the company last year. Unfortunately the staff member was a little disorganized and lost most of the receipts. His boss tells him that if he can come up with an estimate for the value, that would be sufficient. The IT member was able to find the receipts for 6 random purchases and found an average cost of $684.92 with a standard deviation of $17.10. He also knows that the distribution of costs on the laptops is normally distributed. He then calculated a confidence interval for the data and reported to his boss: "I am 95% confident that the true average amount we spent per laptop last year was between $671.24 and $698.60". Is the IT member's statement correct? Explain using the 4 sentence structure described below.

a. Provide a clear answer. – If the question asks you a yes or no question, state either yes or no and answer in a complete sentence. If the question asks you to choose from two options, choose one of the options and answer in a complete sentence.
b. State the RELEVANT facts that lead to your answer. – Find the information given to you in the problem that is relevant to the answer you have just given. Don't write everything you know about a topic, only what makes sense for the given problem.
c. State the additional information the facts imply. – Connect the relevant information you just stated to information you have learned that led you to the conclusion you have made.
d. Explain how the implications lead to the conclusion. – Connect any missing dots to your argument that lead you to give the answer you did.

Assignment 2 – Open is the new biased

Piper is trying to determine the proportion of teens who text and drive. She went online and found two sources that reported different results. Here is what she found.

- Source 1: A newspaper article had randomly surveyed 84 teens anonymously asking questions about different behaviors they exhibited. One of the questions asked "Do you text while driving?" The survey results stated that 76% of teens reported that they texted while driving.
- Source 2: A local high school student and member of the school newspaper posted the question "Do you text and drive" on her Facebook account. 397 of her friends responded. About 53% of them reported that they texted while driving, the results of which the student posted in the school newspaper.

Piper believes that since the second source sampled more people, the results are more reliable. Do you agree with Piper's conclusion? Explain using the four sentence structure.

Assignment 3 – Monkeying Around With Methods

The usual way to study the brain's response to sounds is to have subjects listen to both "pure tones" and a recognizable sound. Researchers were interested in monkeys' responses to pure tones versus monkey calls and if the monkeys would respond stronger to the monkey calls. The researchers fed pure tones and monkey calls directly to the brains of 36 monkeys. The response to the stimulus was measured by the firing rate of neutrons in the various areas of the brain. A paired differences test was conducted on the data. Note: the data for this problem can be found on blackboard in the file called Tone Data.csv.

Obs	Tone	Call	Obs	Tone	Call	Obs	Tone	Call	Obs	Tone	Call
1	474	500	**10**	145	42	**19**	71	134	**28**	35	103
2	256	138	**11**	141	241	**20**	68	65	**29**	31	70
3	241	485	**12**	129	194	**21**	59	182	**30**	28	192
4	226	338	**13**	113	123	**22**	59	97	**31**	26	203
5	185	194	**14**	112	182	**23**	57	318	**32**	26	135
6	174	159	**15**	102	141	**24**	56	201	**33**	21	129
7	176	341	**16**	100	118	**25**	47	279	**34**	20	193
8	168	85	**17**	74	62	**26**	46	62	**35**	20	54
9	161	303	**18**	72	112	**27**	41	84	**36**	19	66

Before data collection, one member of the research team had recommended splitting the monkeys in half using random assignment, giving the treatment group the monkey call and the control group the pure tone. The researchers then would have done a two sample t-test on the data. Solve the problem on your own first then decide: do you agree with the researcher's proposed method? Explain using the four sentence structure. Make sure to support your answer with any relevant analysis.

Assignment 4 – Chi-square Conditions Conundrum

You are playing a game with two of your friends, Hans and Christian, both of whom are very competitive. To play this game, you toss two dice, add together the values shown and move that many spaces on the board. After 36 turns, Hans has tossed the following values. Note that this data can be found on blackboard in the file called game cheating.csv.

Turn	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
Roll	6	7	8	8	9	9	2	6	3	8	9	11
Turn	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**
Roll	4	7	9	8	6	10	10	7	4	6	12	6
Turn	**25**	**26**	**27**	**28**	**29**	**30**	**31**	**32**	**33**	**34**	**35**	**36**
Roll	9	8	6	7	3	3	2	5	7	9	6	5

At this point Christian feels like Hans has been getting very lucky on where he lands on the board and wants to see if his dice may be rigged. He decides that he is going to do a Chi-square goodness of fit on the data, but notices that one of the conditions for inference isn't met: the count for each cell must be at least 5.

Since the game will be over before there is enough data collected to meet this condition, Christian decides to multiply the counts he currently has by 5 and use that information to conduct the test. Do you agree with Christian's method? Conduct the analysis yourself and explain using the 4 sentence structure.

Assignment 5 -Mini Model Selection
Johnny is a car salesman and is developing a model to help him assign prices to cars that he might want to purchase (to be resold). He has collected several variables worth of data on both Honda Accords and Toyota Camrys. This data can be found on blackboard in the file entitled car.data.csv.

He would like to have a model for predicting price based on the variable mileage. Examine the data given to you and consider different ways to build this model. Your goal is to propose at least two models that Johnny could use to predict the price of the cars. Which of your proposed models do you think is best for Johnny? Explain using the 4 sentence structure.

Assignment 6 - Professors are out(liers) to get ya!
Alice is a student at Mid-South State University (MSSU). She and her friend Alex, who are both rather tall, feel like professors may be intimidated by their height and treat them more harshly when it comes time to assign grades. Alice took an introductory statistics course and remembers her teaching harping on the fact that data beat anecdotes, so she doesn't want to jump to any conclusions based solely on the accounts of two people.

She was able to obtain some data from surveys collected from randomly selected students at the university. This data set includes information such as the student's gender, height, GPA and number of credit hours taken, among other variables. This data set can be found on blackboard as survey.csv.

Using the data from the survey, Alice examined the linear relationship between height and GPA and found evidence at the 10% level that taller students have lower GPA's. Do you agree with Alice's claim that there seems to be a negative relationship between height and GPA? Explain using the four sentence structure.

Assignment 7 – Weird Statistics: Regressing a Prom Date
Gary and Wyatt are trying to get dates to prom. They aren't very picky about who they take, but they really don't want to be rejected when they do find someone to ask. They have decided to create a statistical model based on survey data from students within their school to see if they can come up with some criteria for choosing a girl to ask. This data can be found on blackboard in the file survey2.csv.

In this survey, the respondent was asked to report their height and also provide an "ideal height" for what would like in a mate. Gary created a regression model where the survey respondent's height is used as a predictor variable and their ideal mate height was the response variable. The results showed that there was a significant negative linear relationship. He excitedly called Wyatt at 2 in the morning and exclaims "I'm going to ask out a girl who is taller than me, because clearly my model shows that girls ideally prefer guys who are shorter!" Wyatt hurriedly tries to discourage Gary from asking out a taller girl when his mother comes in and makes him get off the phone.

Based on the data, should Gary ask out a girl that is taller than him or shorter than him? Explain using the four sentence structure.

Assignment 8 - The Ultima(te) Model Selection
Johnny is back at it again. He is still trying to come up with a model to help him with purchasing cars. However, now he has learned about multiple linear regression and wants to include several variables in his model and come up with the best one to determine the price. The data for this problem can be found on blackboard in the file car data.csv.

From his successful endeavors in simple linear regression, Johnny is convinced that mileage belongs in the model and asks that you include it in future models. Using the data that is given to you, help Johnny determine the best model to predict the price of the cars. Explain why your model is the best model using the four sentence structure.

Assignment 9 - Identifying Statistical Errors
Matthew, a pediatrician, wanted to determine the average amount of calories that his 0 to 3 month old patients were getting. Specifically, he wanted to test to see if the children in his care were getting at least the recommended 400 calories per day, giving him the following hypotheses:

$$H_0: \mu = 400$$
$$H_a: \mu > 400$$

After taking a random sample of his patients, and conducting the appropriate test, he found his p-value to be 0.043. Matthew wrote a summary of his results by saying: “we reject the null hypothesis and there is sufficient evidence to conclude that the children are getting at least 400 calories per day.” When asked about the potential for a statistical error in his results he stated “No, there is no chance of a statistical error. I double checked my calculations, made sure all of my assumptions were met and I know that when my p-value is small that I reject my null hypothesis.” Is Matthew’s final statement correct? Explain using the four sentence structure.

Assignment 10 – Confirming Theoretical Through Simulation
Recall that the Bonferroni multiple comparison adjustment allows us to compare multiple pairwise comparisons but helps to maintain the correct type I error rate. That is for example, if we are comparing three populations, and we conduct pairwise t-tests on all combinations, we need to redefine our significance value to be $\alpha/3$.

Two statisticians are arguing over how well the Bonferroni correction works. One of them argues that it gives us a type I error rate that is above 5% (that is, the tests will still reject more than 5% of the time, even if the population means are all the same value). The second statistician argues that it gets us below a 5% type I error rate. Which statistician (if either) is correct? Conduct a simulation to show who is correct. Explain your results using the 4 sentence structure.

Appendix B: Interview Protocol

Introduction:
Hi _______, thank you for agreeing to participate in this task-based interview with me today. How has your semester been going? Are you taking any interesting classes? (etc.)
I really appreciate you taking the time to meet with me today. As you may know, I am currently working on collecting data so that I can write my dissertation and graduate this coming summer. The goal of the research is to see how taking statistical computing affects one's ability to think and reason statistically. Since you have taken computing and you took stats 2 prior to that, you are an ideal person to observe doing a task from a statistics 2 context. The information I collect here will be used as baseline information as I conduct interviews with students who are currently taking stats 2. To make sure I do not forget anything that you say or do, this interview is going to be video-taped and the computer screen will be screen captured.

The interview is going to take on 3 distinct parts. In the first part, I'm going to ask you to read over a prompt for a statistics 2 problem and ask you to solve the problem in any way you know or remember how. You can use the computer that's in front of you to do that. Just so you know, it will be recording your motions using a screen capture software. In the second part of the interview I'm going to give you a short writing task where you will be asked to critique a solution that someone has given to a problem. In the third part, I'm going to ask you a few questions about how you felt about taking statistics 2 and statistical computing. Do you have any questions for me, before I begin?

Phase 1: The task-based problem
To be given to the student on a separate sheet of paper:
Favorable weather has been shown to be associated with increased tipping. Will just the belief that future weather be favorable lead to higher tips? The researchers gave 60 index cards to a waitress at an Italian restaurant in New Jersey. Before delivering the bill to each customer, the waitress randomly selected a card and wrote on the bill the same message that was printed on the index card.
Twenty of the cards had the message "The weather is supposed to be really good tomorrow. I hope you enjoy the day!" Another 20 cards contained the message, "The weather is supposed to be not so good tomorrow. I hope you enjoy the day anyway!" The remaining 20 cards were blank, indicating that the waitress was not supposed to write any message.
The data for this problem can be found in R.
Do the data support the hypothesis that there are differences among the tipping percentages for the three experimental conditions?
Questions for the student:

- What are your initial thoughts on the problem? Do you think there may be a difference?
- What sort of statistical procedure do you think would be best to answer this question? Why did you choose that procedure?
- Go ahead and use the laptop now. The data has already been loaded into R, as stated in the prompt. Do a full statistical analysis to answer the question that is given in the problem.

The student may be asked questions as they are going to see what they are doing. There are certain conditions that need to be met and students might be asked to clarify what they are doing if they demonstrate that they are checking any of these conditions. Some examples that might be seen are:

- Random sample (rereading the prompt)
- Standard deviation of one group is no more than twice that of any other group (checking summary statistics)
- The data appears to be approximately normal (can be checked with boxplots or qqnorm/qqline)

Note: students may agree or disagree on whether or not the conditions for inference are met on this problem. Students should continue with the interview, even if they feel like the conditions were not met.

Using the software students should come up with an ANOVA table/p-value to answer the question.

- What were your null and alternative hypotheses for this test?
- What p-value did you find? How can you interpret the results based on this p-value?

Note: the correct answer is that at least one group is different, but the student may say that all groups are different. If they do, they will be asked to go (back) to the side by side box plots and see if all three groups actually look different.

- One of things that the ANOVA table gives us is the Sum Sq values for both report and the Residuals. What do these values mean in the context of the problem?

Note: SSReport is the numeric value of variation between the three factors. SSResidual is the numeric value of the variation of the points from their group means.

- In the ANOVA table, we divide the MSReport by the MSResiduals to get the F statistics. This value came out to be rather large (20.68). Why do we divide MSREport/MSResiduals?
- What does it mean when this quotient is large?

Note: We divide the two values because we want to see if the overall variation between groups is equivalent to the overall variation within groups. If they are the same, this ratio will be 1. If they are not the same, the resulting statistic will be large, indicating that there is at least one group that doesn't fit with the general pattern.

- You should always check your conditions for inference before doing any statistical analysis. What are the conditions for inference about ANOVA and more importantly, what could happen if any of those conditions are not met?

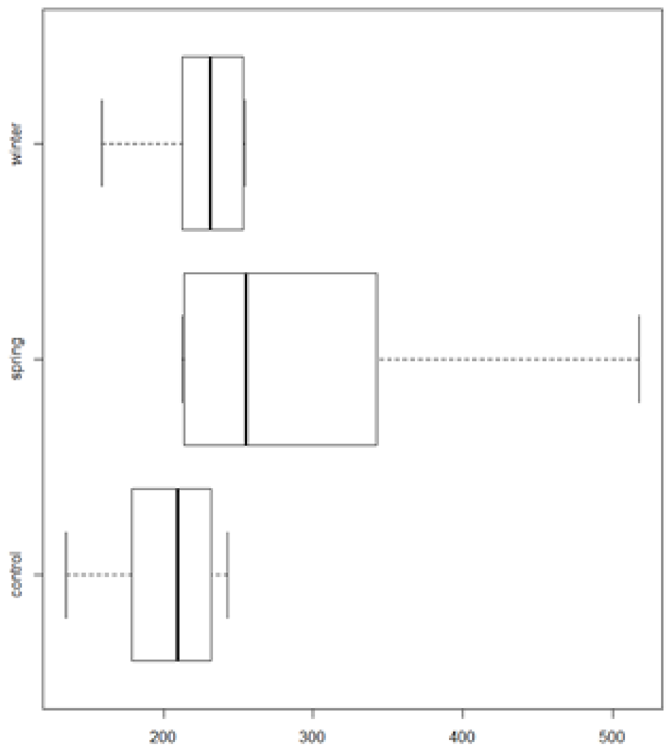
Phase 2: Critiquing a solution

Next, I'm going to give you a prompt about an ANOVA problem that a student has already answered. Your goal is to write a response stating whether or not you believe the results of the student and give reason as to why you made that choice.

Student Handout:

Data was collected for several years on the plant biomass for 18 fields exposed to different watering treatments. In each year, the fields were randomly assigned to receive added water either in the winter or the spring, or to have no added water. Anna wanted to determine if the plant biomass was significantly different under any of the three treatments. Analyzing data from the 2004 season she came up with the following side by side boxplots and ANOVA table.

```
          Df Sum Sq Mean Sq F value Pr(>F)
Treatment  2  31945   15972   2.898 0.0862 .
Residuals 15  82661    5511
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```



After reviewing the output, Anna stated the following: *“Based on the output, there is not enough evidence to show that additional watering in either the spring or the winter bares a significant difference in the plant biomass output*.”

Do you believe Anna’s results?

1. Provide a clear answer. – I state either yes or no and answer in a complete sentence.
2. State the RELEVANT facts that lead to your answer. – Find the information given to you in the problem that is relevant to the answer you have just given. Don’t write everything you know about a topic, only what makes sense for the given problem.
3. State the additional information the facts imply. – Connect the relevant information from the problem to what you already know about statistics that leads you to the conclusion you have made.
4. Explain how the implications lead to the conclusion. – Connect any missing dots to your argument that lead you to give the answer you did.

Phase 3: Questions about Stats 2 and Statistical Computing

- I'm going to ask you some questions about Computing prior to taking your Statistics 2 class.
    - Did you have any computing experience prior to taking statistics 2?
    - How comfortable did you feel with any computing going into statistics 2?
    - How comfortable did you feel with computing after you were done with statistics 2?
    - Do you think you would have done better in statistics 2 if you had taken computing first?
- Next I'm going to ask you some questions about your statistics 2 knowledge in the computing class.
    - How long had it been since you had taken statistics 2 coming into the computing class?
    - How comfortable did you feel with statistics 2 material prior to taking computing?
    - Did you feel like taking computing helped solidify any of the statistics 2 concepts you had learned prior that you had struggled with?
    - Do you think you did better in computing because you had taken statistics 2 prior?

Appendix C: Student Demographic Survey

Please answer this survey to the best of your knowledge. If you do not know the answer to one of the questions or you are unsure of an answer, please indicate this with any given responses.

1. What statistics courses have you taken in either high school or college? Please give course names if you know them. If you took a course at Meredith College, a course number will be sufficient. Please also indicate when you took these courses (semester and year).

2. What computer science courses have you taken in either high school or college? Please give course names if you know them. If you took a course at Meredith College, a course number will be sufficient. Please also indicate when you took these courses (semester and year).

3. Do you have a preferred statistical programming language or statistical software? If so, which one? Is there are reason for this preference?

4. On a scale from 1 to 10, with 1 being not at all confident and 10 being completely confident, before you starting the second course in statistics, how confident were you in your ability to use a statistical programming language? (Please circle one)

   1 2 3 4 5 6 7 8 9 10

   If you would like to offer further explanation, please do so here:

5. On a scale from 1 to 10, with one being not at all confident and 10 being completely confident, after almost a semester of using R, how confident are you in your ability to use a statistical programming language? (Please circle one)

   1 2 3 4 5 6 7 8 9 10

   If you would like to offer further explanation, please do so here:
6. What are your declared major and minor? If you have not yet declared but have decided on a major and minor, what do you plan to declare?

7. Have you participated in any competitions, consulting or research outside of your classwork that may influence your ability to do statistics or statistical computing? If so, describe the work you have done.

8. Do you have any other information about your ability to use statistics or statistical technology that you would like to share with the researcher?

Thank you for completing this survey!

Appendix D: NCSU IRB approval

IRB PROTOCOL - 9076

Title	Description	Populations	Consent	Procedures	Data Security	Risks and Benefits	Compensation	Routing and Status

<- Previous | Return to Main Menu | Next ->

Project Title

How learning statistical computing affects student's statistical thinking and reasoning.

Source of funding (if externally funded, enter PINS or RADAR number of funding proposal via 'Add New Sponsored Project Record' button below):

NCSU Faculty point of contact for this protocol:
NB: only this person has authority to submit the protocol

Lee, Hollylynne Stohl
Science, Technology, Engineering, & Mathematics Education (STEM)
Phone: 919-513-3544 Fax: 919-515-6892
COI Filed: 94188, filed 10/13/2016

Additional Personnel	
If the primary contact is not also the principal investigator, list the PI here. Make sure to enter the unity ID (ex: dapaxton) appropriately. Any other investigators should be listed in the appropriate field on the Description tab.	
Name:	Weber, Victoria Lynn (vlweber)
Email:	vlweber@ncsu.edu

Does any investigator associated with this project have a significant financial interest in, or other conflict of interest involv the sponsor of this project? (Answer No if this project is not sponsored)

No

IRB

For IRB Office use

Areas of regulatory concern:

Vulnerable populations:

- Students

Original Approval Date:

08/29/2016

Current Approval Period Begin Date:

08/29/2016

Current Approval Period End Date:

Category:

Exempt b.1, b.2

Appendix E: Meredith College IRB approval

IRB File # 752

Meredith College Institutional Review Board
Application for Approval of Research with Human Participants as Exempt from IRB Review

Project Title: How learning statistical computing affects student's statistical thinking and reasoning.

Principle Investigator (for co-investigators, submit equivalent information on a separate sheet):

VICTORIA WEBER weberviс@meredith.edu
Printed name and email

SMB 243
Local mailing address

Victoria L Weber 9/6/16
Signature and date

Faculty Supervisor (if PI is a student)

Printed name and email

Local mailing address

Signature and date

Exemption Category Checklist

Check one or more of the criteria below that make your research eligible for exemption.

✓ Research on instructional strategies, techniques, curricula, or classroom management methods conducted in established or commonly accepted educational settings.

_____ Research involving the use of educational tests (cognitive, diagnostic, aptitude, achievement) if information taken from these sources is recorded in such a manner that participants cannot be identified.

_____ Surveys or interviews in which responses are recorded in such a manner that the human participants cannot possibly be identified, either directly or through identifiers linked to the participants.

_____ Surveys or interviews in which the respondents are elected or appointed public officials or candidates for public office

_____ Observations of behavior in public situations where participants would reasonably assume the absence of privacy

_____ Collection or study of publicly available existing data, documents, records or specimens.

_____ Collection or study of existing data, documents, records or specimens if these sources are publicly available or if the information will be recorded in such a manner that an individual participant can not be identified, either directly or through identifiers linked to the participants.

1

Please prepare a concise (typically one page) description of the research to accompany this application, and include the following information:

- The goals, hypotheses or research questions guiding the investigation.
- Expected dates for starting and completing the research.
- How participants will be selected or identified.
- How many participants your plan will involve.
- Any type of compensation or reward that participants will receive.
- How you will obtain the data.
- Where and for how long the data will be stored.

IRB from NCSU has been approved.

Also please attach a copy of written materials that may assist the IRB in determining if your research is eligible for exemption from continuing review. Examples would be a cover letter, a questionnaire, a script of remarks for participants, etc.

<u>If this research project is customized by a member of the Meredith faculty to meet specific learning objectives in a course offered at Meredith College, a "blanket" exemption is possible, valid for up to two years. In lieu of the description requested above, submit a generic explanation of topics and/or methods that students will pursue, with explicit clarification of how the research assignment is relevant to the objectives. Include the course number and title, and indicate how often the course is scheduled.</u>

<u>Submit three copies to the Meredith College Institutional Review Board</u>

<u>Ledford 113</u>

Application for Exempt Review [✓ is] [____ is not] approved.

Signature of IRB Chairperson____________________

Date of decision 9/7/16, and if approved…

Approval expiration date 9/7/17

Appendix F: Informed Consent Form

NC STATE UNIVERSITY

MEREDITH

Informed Consent Form for: A Curriculum Assessment of Introductory Statistics

This informed consent form is for students enrolled in Victoria Weber's MAT 345 (Statistics II) course at Meredith College during the fall semester of 2016.

Principal Investigator: Victoria Weber
Faculty Sponsor: Hollylynne Lee
North Carolina State University/Meredith College
How Learning Statistical Computing Affects Student's Statistical Thinking and Reasoning.

This Informed Consent Form has two parts:

- **Information Sheet (to share information about the study with you)**
- **Certificate of Consent (for signatures if you choose to participate)**

You will be given a copy of the full Informed Consent Form

Part I: Information Sheet

Introduction
I am currently working on my PhD at NC State. As part of the requirements for graduation, I need to complete a research based dissertation on a topic from my field: statistics education. I am particularly interested in studying curriculum development and how taking a statistical computing course may affect your ability to think and reason about statistical topics.
You have been invited to participate in this research since you are enrolled in my MAT 345 course in the Fall of 2016. You are not guaranteed any personal benefits from being in this study. Research studies also may pose risks to those that participate. In this consent form you will find specific details about the research in which you are being asked to participate. If you do not understand something in this form it is your right to ask the researcher for clarification or more information. You do not have to decide today whether or not you wish to participate in the research. A copy of this consent form will be provided to you. If at any time you have questions about your participation, do not hesitate to contact the researcher.

Purpose of the research
Many researchers in the field of statistics have determined that learning a statistical computing language before graduation may be beneficial to those that are seeking work in the field. However, there is debate among statisticians as to *when* the best time is to teach statistical computing. My goal is to determine if teaching statistical computing to students before they have taken a second course in statistics improves their ability to think and reason statistically, creating the argument that computing should be taught earlier in a student's college career.

Type of Research Intervention
For most of the students who participate in this research, you will not need do anything more than come to class, and participate as normal. However, some students in the class may be asked if they want to participate in an video-taped interview after the semester is over. This interview will take approximately an hour to an hour and a half and will ask questions about different topics that you will have studied this semester.

Voluntary Participation
Your participation in this research is entirely voluntary. It is your choice whether to participate or not. If you choose not to participate, this will not affect your grade in this course in any way. You may change your mind later and stop participating even if you agree to participate now.

Procedures
You are being asked to participate in this research study to help make the statistics curriculum at Meredith College, and perhaps other colleges and universities, better for both you and your peers. By being part of this research project you will be asked to:

- complete this informed consent form
- complete the writing assignments which will be used for analysis.

- if selected, participate in a 1 to 1.5 hour interview

In the interview you will be asked to complete different tasks and problems relating to topics we have covered this semester. You may also be asked your opinion about this course and other statistics courses at Meredith College.

Duration
The research will take place during the course of the semester in which you are taking MAT 345 (Fall 2016). I may ask you to participate in an interview in the Spring semester of 2017.

Risks
As a student in my class, you should know that your grade will not in any way be impacted (positively or negatively) by participating in this study. The data that I collect from you will be purely informational to help better form statistics curricula.
Students that will be asked to participate in the interviews will be selected at random until a predetermined number of students agree to participate. When writing my dissertation, your names will be changed to pseudonyms to protect your identity.

Benefits
For the majority of participants in this study, there may be no direct benefit to you. However your participation will benefit students in future semesters.
For the students that participate in the interviews, you will receive a $10 gift card (place to be determined).

Confidentiality
The research being done in the community may draw attention and if you participate you may be asked questions by other people in the community. We will not be sharing information about you to anyone outside of the research team. The information that we collect from this research project will be kept private and secure.

Sharing the Results
If you wish to see the results of the study, a copy of the final version will be emailed to you upon request. Since this work is being done for my dissertation, the final results will be published for the public to read.

Right to Refuse or Withdraw
You do not have to take part in this research if you do not wish to do so, and choosing to participate will not affect your grade in this course in any way. You may stop participating in the interview at any time that you wish without your grade being affected. I will give you an opportunity at the end of the interview to review your remarks. If there is anything that you said that you would like omitted, you may ask me to do so.

Who to Contact
If you have any questions at any time about the study itself or the procedures implemented in this study, you may contact the research (Victoria Weber) at: webervic@meredith.edu.

This proposal has been reviewed and approved by the Meredith College IRB and the NCSU IRB. Both are committees whose task it is to make sure that research participants are protected from harm. If you feel you have not been treated according to the description in this form, or your rights as a participant in research have been violated during the course of this project, you may contact Deb Paxton, Regulatory Compliance Administrator at dapaxton@ncsu.edu or by phone at 919-515-4514.

Part II: Certificate of Consent

I have read and understand the above information. I have received a copy of this form. I agree to participate in this study with the understanding that I may choose not to participate or to stop participating at any time without penalty or loss of benefits to which I am otherwise entitled.

Print Name of Participant______________________________
Signature of Participant ______________________________
Date ________________________