Statistical Understanding of Undergraduate Students

Sophia I. Sunseri

Submitted to the Psychology Department, Allegheny College

in partial fulfillment of the requirements for the degree of

BACHELOR OF SCIENCE

April 5th, 2023

Statistical Understanding of Undergraduate Students

A Senior Comprehensive Project

Sophia I. Sunseri

Allegheny College

April 5th, 2023

This work is mine unless otherwise cited.

Pledge: __Sophia I Sunseri__

Approved:

_____

Dr. Christopher Normile

_____

Dr. Lydia Eckstein

Acknowledgements

I would like to thank my parents for always supporting me no matter what choices I was facing, and whatever decisions I came to. Thank you to my friends, especially Meg, Dag, and Sul, for sticking by me and helping me whenever I needed it, whether it was by participating in my study, or making me laugh and giving me a break from the stress of senior year. Thank you to my professors, especially Dr. B. and Professor Normile, for helping me through these past four years, and helping me plan out what my future holds and what steps I need to take to get myself there. Everyone who has helped me with this experience holds a very special place in my heart.

Table of Contents

Abstract

Statistical cognition is the processes, representations, and activities involved in acquiring and using statistical knowledge by looking at: (1) how people acquire/use statistical knowledge, (2) how they *should* think about statistical concepts, as well as (3) closing the gap between the previous two ideas (Cumming et al., 2008). This paper looked at the connections between statistical cognition, prior knowledge, statistical self-efficacy, major discipline, and gender. Participants were provided with a statistical self-efficacy questionnaire designed to measure their belief in their ability to complete statistical tasks. Participants then completed a statistical cognition survey designed to measure how much knowledge the students have about key statistical concepts. Results revealed no significant difference in statistical cognition across disciplines.Furthermore, their prior knowledge in statistics did not significantly affect statistical cognition scores. However, the gender of the participant did have a statistically significant effect on self-efficacy, in that women had lower self-efficacy scores than men. This information tells us that statistical education across disciplines may be fairly consistent, prior experience does not limit someone's knowledge, and women are suffering from the stereotype threat which leads them to believe they cannot excel in math as well as men can even if their scores reveal otherwise.

Statistical Understanding of Undergraduate Students

It should be no surprise that statistics is typically considered a more daunting subject. It is something that many people struggle with, but few want to actually discuss. Statistical illiteracy is a problem that is not contained to a single population, or education level; it is seen equally through the highly- and less-educated (Gigerenzer, 2008). It is also believed to be a necessity to be statistically literate, although little urgency is seen to actually change the way statistics are taught (Cerrito, 1999). Even topics that may seem elementary, or simple, in nature—such as averages—are found to have misconceptions associated with them (Batanero, 2002).

Statistical cognition is the processes, representations, and activities involved in acquiring and using statistical knowledge (Cumming et al., 2008). This field looks at: (1) how people acquire/use statistical knowledge, (2) how they *should* think about statistical concepts, as well as (3) closing the gap between the previous two ideas (Cumming et al., 2008). It is believed that once it is admitted that statistical literacy/cognition is an important area of understanding, it is then necessary to discuss relevant skills, ideas, and concepts which are involved with statistics (Batanero, 2002).

In this paper, I will also discuss topics which could affect statistical cognition: self-efficacy and stereotype threat. The goal of this paper is to delve into how these three topics intertwine, and how they may be affected by gender, discipline, and prior knowledge.

Stereotype Threat

Stereotype threat is "the risk of confirming a negative stereotypic expectation about one's group" (Keller & Dauenheimer, 2012, p. 371). For example, a common stereotype is that women are bad drivers. The topic of stereotypes and stereotype threat is a common focus in research, looking at many different stereotypes. Some examples of this are: as people age, they become

more dependent on others (Coudin & Alexopoulos, 2010); women are not as good at sports as men (Hively & El-Alayli, 2014); and immigrants have low verbal abilities (Froehlich et al., 2016).

In Coudin and Alexopoulos (2010), two experiments were conducted. In the first, they wanted to study the effects of aging stereotypes on loneliness and risk-taking behaviors. They tested this by presenting participants with a speech to read aloud which differed across three conditions (positive, negative, and neutral), then having them complete a self-recorded questionnaire to assess their dependability on others. This study supported their hypothesis which stated that negative stereotype exposure would lead to an increase in self-reported loneliness of participants. In the second, Coudin and Alexopoulos (2010) looked at how stereotypes would influence dependency behaviors of participants. Rather than reading a speech, as was done in the first experiment, participants were asked to listen to a recorded speech, then complete the same self-reporting questionnaire. The results from this experiment showed that participants exposed to negative stereotypes felt a health deterioration and less extraversion than their positively affected counterparts.

The goal of Hively & El-Alayli (2014) was to compare female and male athlete's performance (1) when a stereotype was present and (2) when the threat of a stereotype was removed. This was done by assessing participants through an athletic activity—tennis or basketball—and having participants read a card which would either promote or mitigate the stereotype threat prior to said activity. When promoting the stereotype, the card expressed that it was likely for gender differences to be found. When mitigating the threat, the card expressed that no differences were expected across genders. They found that when concentration was needed in the tasks, and the stereotype was present, women performed worse than men.

Froehlich et al. (2016) did their research on the theory of intelligence (TOI), group membership, and the stereotype of verbal abilities in immigrants. They hypothesized that TOI and performance would positively relate to each other after stereotype exposure. Participants completed a verbal assessment and a self-reported TOI questionnaire. Froehlich et al. (2016) found that when the negative stereotype had been made clear, there was a greater chance of participants believing their intelligence was fixed; however, they specified that the influence of TOI on performance was uncorrelated with stereotype endorsement.

There was a general theme across these three studies that showed that when a negative stereotype was explicitly stated, performance was reduced.

With that in mind, a common stereotype aimed at women, is that women are not as good as men when math performance is observed (Tomasetto et al., 2011). However, this belief has expanded to include that minorities are expected to underperform on mathematical tasks (Steele & Aronson, 1995). It has also been found that anything which reminds women and minorities of this stereotype can diminish their performance, and even their motivation. These findings have then brought some suggestions as to how to lessen the threat. The main suggestion, proposed by Johns et al. (2005), was to educate the affected groups about what stereotype threat is, similar to how introductory psychology textbooks discuss it. The results even showed that the women who learned about stereotype threat, did not suffer from impaired math scores during the assessment. Johns et al. (2005) attributes this to a source of external arousal which was presented through the education.

Stereotype threat theory argues that situational aspects are most important, meaning that when an individual feels that they could be judged or treated differently due to a stereotype, they are more likely to be affected by it (Spencer et al., 1999). This theory suggests that lessening

negative stereotypes in evaluation moments could lead to an increase in performance by stigmatized social groups. This can be done by explicitly stating whether or not gender differences have been found, or are expected to be found; in terms of reducing the stereotype, it would be stated that there is no expectation/support of finding a gender difference (Spencer et al, 1999; Keller & Dauenheimer, 2012). In addition, Keller & Dauenheimer (2012) found that minimizing stereotype threat can increase mathematical performance in women, even in an area which is typically male dominated—such as problem solving. In order to do this, Keller & Dauenheimer (2012) took 74 students across two schools and had them complete a short math test, followed by a questionnaire to assess the participants' anxiety and emotional reactions. To manipulate stereotype threat, the test had an introduction which stated that the test had previously shown no gender differences between male and female scores *or* that the test did show gender differences between male and female scores. By stating in the introduction that gender differences were not expected to be found, Keller & Dauenheimer (2012) reduced stereotype threat, and found that female participants had higher math scores than the group which did not have a reduction in stereotype threat—the group which read that gender differences would be found.

A study done including children from Spain and Chile found that in elementary math knowledge, there was no significant difference in gender between the boys and girls (Pina et al., 2021). However, this could be attributed to a cultural difference, as another study done in the United States found that children began identifying with stereotype threat as early as six-years-old. In this study, Bien et al., (2017) conducted two experiments. The first included three tasks: (1) children listened to a story which described a "really, really smart" individual, then they were asked to choose who they believed the story to be about (out of an unfamiliar

lineup of two men and two women), (2) children were presented with pairs of adults of the same- or different-gender adults, and had to choose which one they believed be "really, really smart," and (3) children completed puzzles which depicted objects (for example, a hammer) or attributes (for example, smart) and then had to decide if the pictures best represented unfamiliar men or women. In the second study, the same tasks were completed, but instead of only comparing male and female adults, male and female children were also compared. At age five, the girls saw no difference between themselves and their male counterparts, but the following year, the girls believed that the boys were smarter than they were (Bian et al., 2017).

Self-Efficacy

Self-efficacy is the belief which students hold about their own academic capabilities (Usher & Pajares, 2009). For example, a student who is high in self-efficacy believes that they have the ability to do well and succeed in school. A person's self-efficacy can determine what types of activities they may participate in, how much effort they are willing to put into said activities, and how long they "persevere in the face of adversity" (Tipton & Worthington, 1984). Not only can self-efficacy limit the effort and activities a person participates in, it can go so far as to limit a person in terms of occupation and their self-determination (Lightsey, 1999). In addition, a student's self-efficacy is a large predictor of their motivation, and their actual achievement (Kaufmann et al., 2022).

In terms of the development of self-efficacy, Bandura (1997; as cited by Usher & Pajares, 2009) considered four themes which could be at play: mastery experience, vicarious experience, social persuasions, and emotional and physiological states. Mastery experience was defined in Bandura's paper as "the interpreted result of one's previous attainments" (Bandura, 1997; as cited by Usher & Pajares, 2009, p. 89). So, if a child completed an academic assignment well and with

ease, then found out that their peers had struggled with the same assignment and did not do as well, that mastery experience was believed to increase their self-efficacy. In addition, there is vicarious experience. For this, instead of an individual focusing on their own experience, they look to others'. In other words, people would compare themselves to one another, which would then affect their self-efficacy. For example, imagine a runner who consistently runs with their friend who has better endurance than them. That runner may start comparing their own performance/endurance with their friend, and begin to feel poorly about their own ability, which could then diminish their performance.

Bandura (1997; as cited by Usher & Pajares, 2009) continued to describe social persuasions, which are created through human interactions. Students often look to others to judge their own abilities—this could be their peers, adults in their lives, or even their teachers—before they determine what their own academic abilities are (Usher & Pajares, 2009). If a child were to see their classmates willingly participating in class while they did not, or found that their teacher was using terms/concepts which they did not understand, their self-efficacy was believed to be negatively affected as a result of social persuasions. The final influence which Bandura (1997; as cited by Usher & Pajares, 2009) identified was emotional and physiological states. These states could be feelings of anxiety or apprehension, feelings of satisfaction and excitement, or even muscle tension/relaxation. Any of these feelings can play into how a person feels about their own abilities in a particular area.

Mathematical Self-Efficacy

Math self-efficacy is a subset of self-efficacy: the belief which students hold about their mathematical capabilities. Grigg et al. (2018) used a sample of 400 students in Eastern Australia to test math self-efficacy, interest, intentions, and achievements. Math self-efficacy, interest, and

intentions were evaluated using a self-report questionnaire, white achievement was determined by using the students' math grade for that academic year. They found that positive math self-efficacy positively affected mathematical achievement in regards to class and standardized grades.

Math anxiety—feelings of stress, tension, and/or anxiety related to math performance—(Richardson & Suinn, 1972), has been found to be related to math self-efficacy (Samuel & Warner, 2019). To determine the connection between math anxiety and self-efficacy, Samuel & Warner (2019), used a combination of mindfulness practices and growth mindsets in the classroom to ease math anxieties. The mindfulness and mindset practices included deep breathing and positive affirmations. Students were instructed to repeat the deep breathing techniques whenever they felt anxious during class, and urged teachers to follow the same practice. In the end, Samuel & Warner (2019) found that those who were in the intervention group, the group which used mindfulness and mindset practices, had lowered math anxieties and increased math self-efficacy.

Current Study

The current study will be focusing on the connection between types of discipline (Humanities, Social Sciences, Natural Sciences), statistical cognition, and statistical self-efficacy. Previous research focuses on mathematical self-efficacy, rather than statistical self-efficacy. Other research also looks at undergraduate students as a whole, whereas this study will look at the separate disciplines in comparison to each other. It is hypothesized that there will be a difference in statistical cognition across disciplines. This is because of research which looked at disciplines like business, social sciences, and lab sciences, and found that statistics courses across disciplines are not consistent nor fully applicable for what different majors utilize

(Schwab-McCoy, 2019). It is also hypothesized that students who have previously taken a statistics course will have better statistical cognition when compared to those who have not previously taken a statistics course; previous research has found that students who already have statistical/mathematical knowledge tend to have better success in statistics courses (Schutz et al., 1998). Lastly, it is hypothesized that there will be a difference in statistical self-efficacy across genders, due to previous research which found that self-efficacy across genders start to blossom at a young age (Bian et al., 2017).

Method

Participants

Participants were 40 undergraduate students who were at least 18 years of age. Two students were unable to participate due to not reaching the age limit, which brought the usable sample size down to 38. Participants were recruited through SONA, which is an online recruitment site where students are able to sign up to participate in psychology research. For psychology student participants, they were compensated with one SONA credit to use towards a psychology class they were enrolled in at the time–if that course required SONA credits. Students who were not enrolled in any psychology classes were not able to be compensated with any type of credit. Those students were recruited through convenience sampling by the use of flyers (see Appendix A) to advertise the study. The study itself was conducted through Qualtrics, with a separate Qualtrics Form for an optional raffle to win a gift card (see Appendix B).

Materials

Statistics Self-Efficacy Scale

The first section within the survey was an adapted version of a self-efficacy scale (May, 2009; see Appendix C). This was a self-reporting questionnaire, where participants rated their

confidence regarding statistics and their performance related to it. This scale consisted of 22 items. An example of an item in this section is "*I get nervous when I have to use statistics outside of school."* This section was scored on a 7-point Likert scale (1 = Never, 7 = Always). In order to get a composition of self-efficacy, an average was taken of the 22 items for each participant. Items 1, 2, 4, 7, 10, 11, 13, 18, and 19 were reverse-coded.

Statistical Cognition Questionnaire

The survey consisted of 22 items, which were focused on the knowledge of statistics (see Appendix D). Some of these statistical concepts were elementary, while others may be considered more advanced. These items were developed from material typically presented in introductory statistics courses. An example of the type of survey item which was seen is "*An outlier is an extreme score in a data set.*" All scale items were measured on a dichotomous "*True"* or "*False"* scale. If an item was answered correctly, participants received one point, and if an item was answered incorrectly, the participant received no points. Once the questionnaire was completed, the points were added together to get a total, final score. This final score was used during analysis.

Demographics

Few demographic questions were asked, but there was a short section at the end of the survey in order to collect additional information about participants (see Appendix E). All information collected was confidential, and did not have any identifying information which could link a person to data. The main demographics asked were age; gender; if statistics courses have been taken, and if so, in what field; and intended major(s). In order to avoid confusion about what majors fell into what discipline, participants were asked for their specific major(s), and the discipline was coded after data collection was completed.

Procedure

Participants completed this study at home, online. For participants who used SONA, there was a link on SONA which took them to the survey on Qualtrics. If participants did not use SONA, and were recruited through the flyer, a QR code was present on said flyers, which directed them to the survey online. They were provided with the informed consent form (see Appendix F). If consent was granted, they were then forwarded to the survey itself.

First, participants completed the statistical self-efficacy questionnaire. This questionnaire was presented across three pages, so as not to overwhelm participants with too many items at once. Participants then moved on to the statistical cognition questionnaire. They clicked through a series of statements about statistics. Each item of the survey was presented individually. To discourage participants from simply clicking a response, and to promote comprehension of each item, the responses were randomized for each item. The end of the survey was the demographics section, where all demographic questions were presented on a single page. No questions/items were required, and participants had the option to skip any and all questions/items which they did not wish to answer, for whatever reason. After participants finished their demographics, they were debriefed (see Appendix G), and thanked for their participation.

Results

Preliminary Analyses

Missing Data

In the self-efficacy portion of the survey, there were three missing data points across three participants. Median imputation was used to fill in the missing scores. Because this was done on a 7-point Likert scale, the median self-efficacy score of four was used for those values to appropriately run the analyses.

Three participants failed to answer several questions,so casewise deletion was used because there was not enough information to test for any hypothesis. All subsequent analyses were run using 35 participants, unless otherwise noted.

Statistical Cognition Against Guessing

Due to the statistical cognition section being a *True* or *False* format, it would be fairly simple to guess and get an objectively high score. So, a one-sample t-test was run against a sample mean of 0.50 to check for the likelihood of guessing.

This test showed that students' ($M = 0.69$, $SD = 0.09$) answers on the statistical cognition survey were significantly different from the sample ($\mu = 0.50$), $t(34) = 12.69$, $p < .001$, Cohen's $d = 2.15$ (see Figure 1). This is a large effect size, and also means that the sample population mean is 2.15 standard deviations above the theoretical mean of 0.50. This means that it is statistically unlikely that the results for statistical cognition were obtained through guessing.

Main Analyses

Statistical Cognition Across Disciplines

To look at how statistical cognition was affected by major discipline, an ANOVA was run using the scored results with discipline (Humanities, Social Science, Natural Science, and an interdisciplinary group to accommodate for double majors/interdisciplinary programs). Originally, the goal was to have a total of six discipline groups to look at the Humanities, Social Sciences, Natural Sciences, and combinations of these. However, due to small samples (see Figure 2), several groups needed to be combined. So, all participants who recorded a double major were placed into the interdisciplinary group for analysis, and the interdisciplinary combinations were all combined.

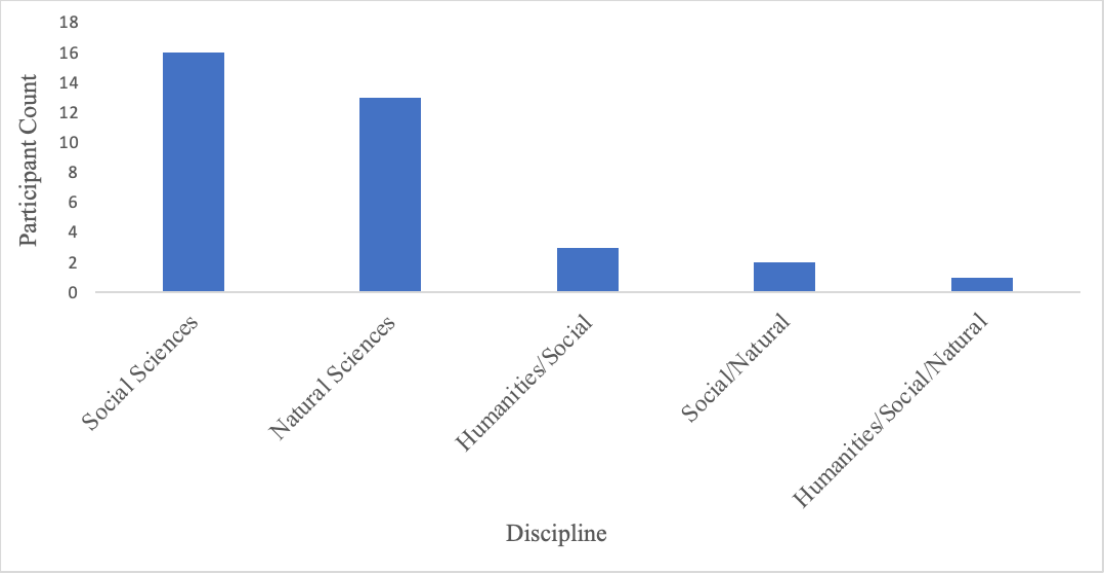


Figure 2. Total number of participants in each discipline, taken from raw data.

After the discipline combinations were regrouped, Figure 3 shows the new sample spread.

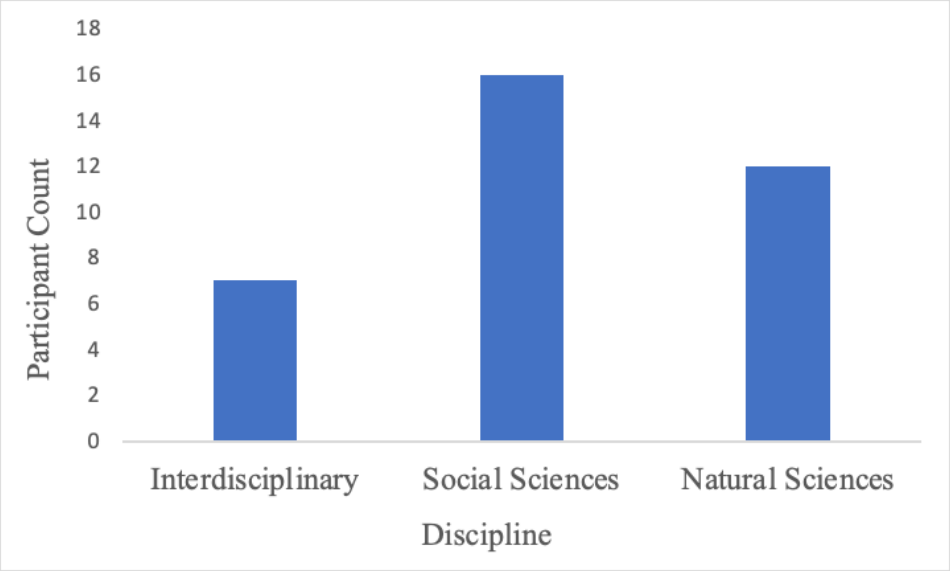


Figure 3. Total number of participants in each discipline, after regrouping.

Although these samples do not violate the assumption for equality of variances (Levene's $p = .582$), it should be noted that the sample sizes do vary a decent bit, so these results should be considered with caution.

When looking at the Q-Q plot for this spread to test for the assumption of normality, the residual data points appear to deviate from the theoretical slope, violating the assumption of normality. To adjust for this, a nonparametric equivalent was run: the Kruskal-Wallis H Test. The results from this mirrored the results of the ANOVA of not being significant, so the ANOVA results are instead being reported.

The ANOVA showed that participants who are majoring in an interdisciplinary field ($M = 15.23$, $SD = 1.38$), those who are majoring in a social science ($M = 15.38$, $SD = 1.75$), and those majoring in a natural science ($M = 15.00$, $SD = 2.59$) did not have a statistically significant difference in their statistical cognition scores, $F(2, 32) = 0.12$, $p = .886$, $\eta^2 = .01$ (see Figure 4). The effect of η^2 tells us that 1% of variance within statistical cognition scores can be explained by the type of discipline which participants were in.

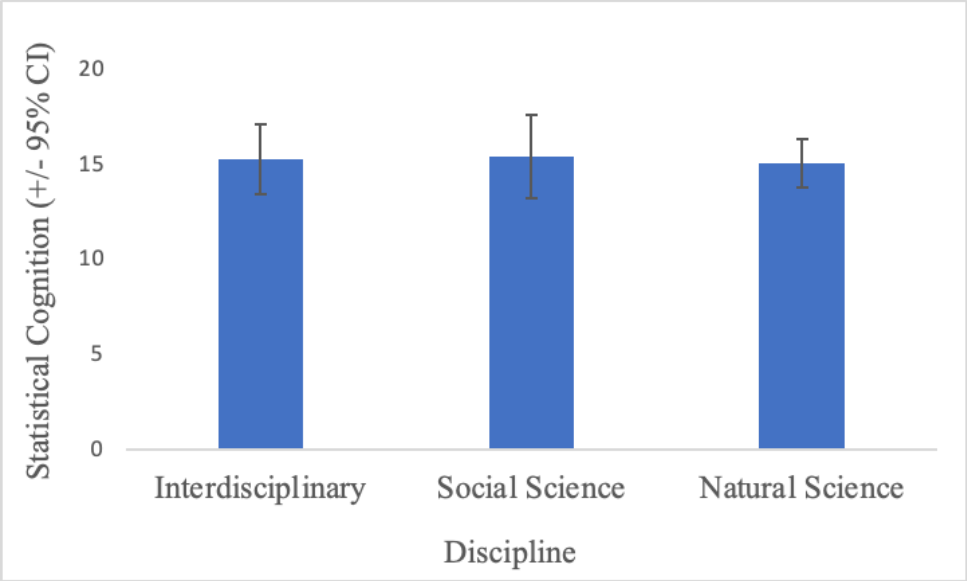


Figure 4. Statistical cognition means within disciplines. Higher scores show a higher number of correctly answered statistical cognition statements. Error bars represent 95% confidence intervals.

Statistical Cognition with Previous Experience

With focus on whether or not participants had taken a statistics course before and their statistical cognition score, an independent samples t-test was run. The independent variable used was whether or not a statistics course was previously taken.

There was a violation of the assumption of equal variances due to extremely uneven group sizes between those who have taken a statistics course before ($N = 29$) and those who have not ($N = 6$), with Levene's $p < .001$, as well as a violation of the assumption of normality using the Shapiro-Wilk's test for those who have taken statistics ($p < .001$) and for those who have not taken statistics ($p = .031$). So, instead, a nonparametric equivalent test was run; in this case, the Mann-Whitney U test. The results from the Mann-Whitney U test mirrored those from the t-test—that there is not a significant difference—so the results from the t-test have been reported.

This analysis revealed that students who have previously taken a statistics course before ($M = 15.35$, $SD = 1.50$) did not have a significantly different statistical cognition score from

those who have not previously taken a statistics course ($M = 14.67$, $SD = 3.67$), $t(33) = 0.76$, $p = .526$, Cohen's $d = 0.34$ (see Figure 5). With this Cohen's d value, we know that there is a relatively small effect size of a participant's previous experience on their final statistical cognition score. With this information, we also know that the mean of those who have taken statistics is .34 standard deviations above the mean of those who have not previously taken statistics.

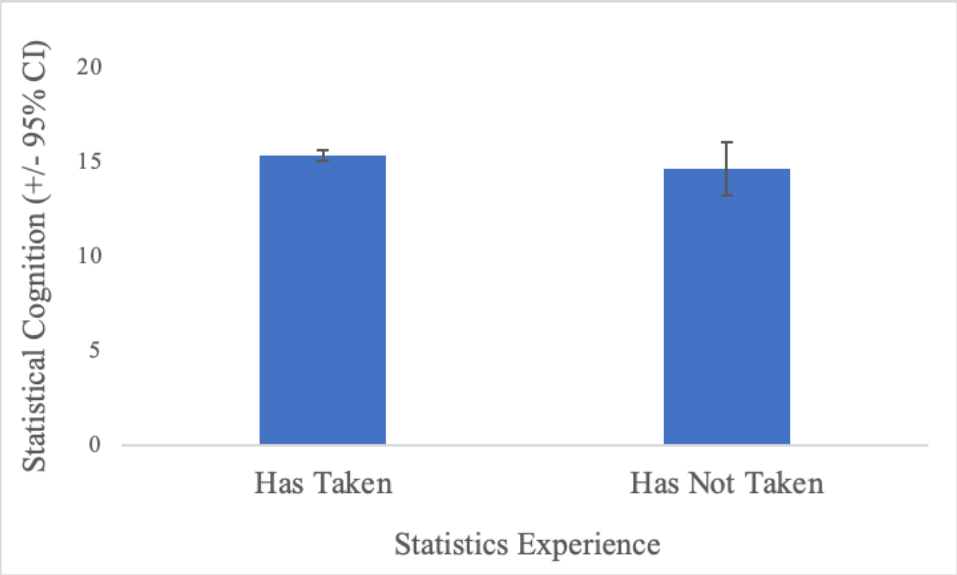


Figure 5. Statistical cognition means split by experience levels. Higher scores show a higher number of correctly answered statistical cognition statements. Error bars represent 95% confidence intervals.

Gender and Statistical Self-Efficacy

To test whether or not there was a significant difference in statistical self-efficacy across genders, an ANOVA was the original plan. However, due to small sample sizes (see Figure 6), those who self-described gender or identified as nonbinary were excluded from the analysis due to the total sample size they would have contributed ($n = 2$) for fear of violating the assumption of equal variances. This changed the sample size to 33.

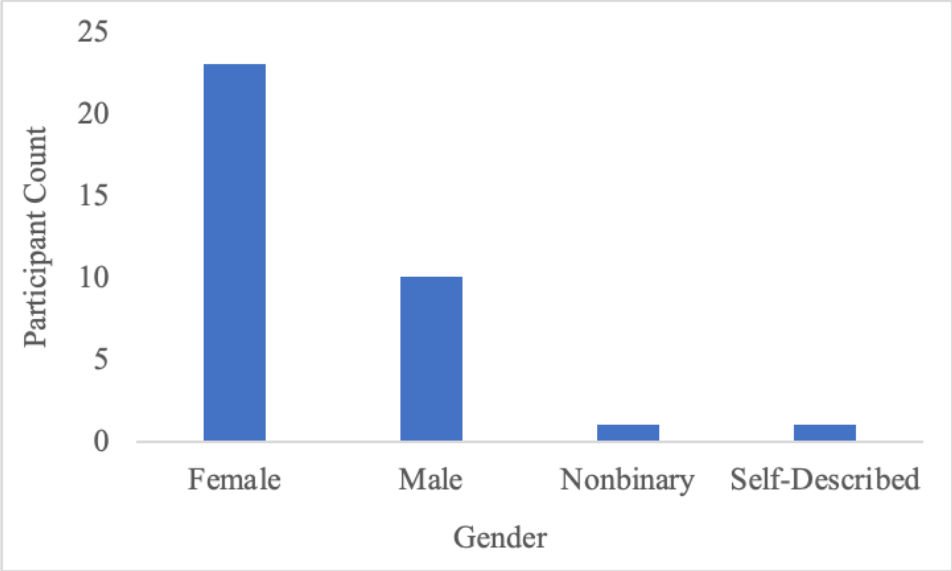


Figure 6. Total number of participants split by gender, taken from raw data. This chart shows 35 participants, two more than used in analyses.

Instead, the male and female groups were compared using an independent samples t-test. Neither group, female (p = .488) nor male (p = .323) violated the Shapiro-Wilk assumption of normality. Levene's test of equal variances was also not violated (p = .622).

The results showed that participants who identified as male (M = 4.79, SD = 1.03) had significantly higher statistical self-efficacy than those who identified as female (M = 3.60, SD = 0.90), $t(31) = -3.33$, p = .002, Cohen's d = -1.26 (see Figure 7). Our Cohen's d value tells us that there is a large effect of gender on statistical self-efficacy, more specifically, the mean of the female group's statistical cognition score is 1.26 standard deviations below the mean of the male group's statistical cognition group.

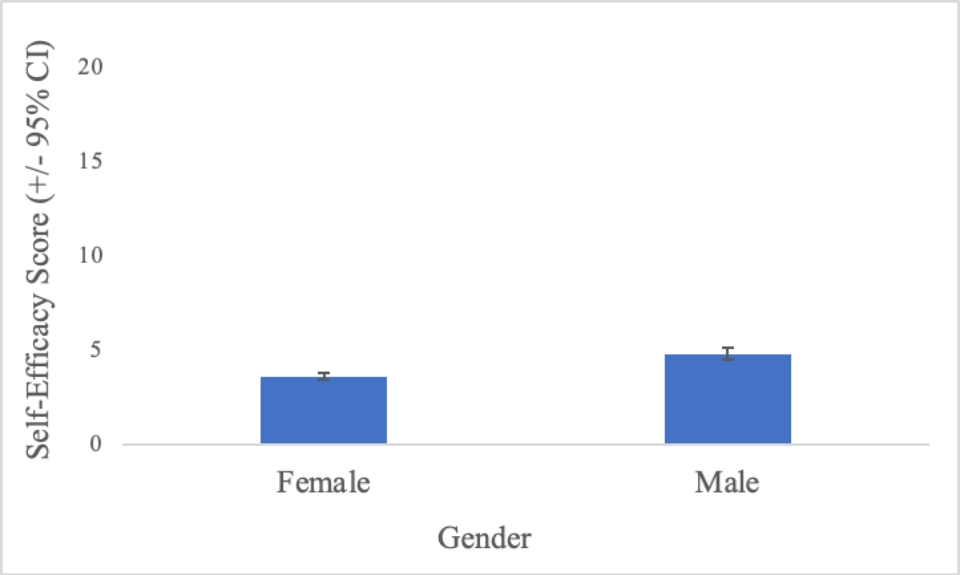


Figure 7. Self-efficacy means split by gender. Higher scores show a higher self-efficacy score. Error bars represent a 95% confidence interval.

Discussion

Overall, students who are in different discipline fields, whether that be the social sciences, natural sciences, or an interdisciplinary field, showed no significant difference when it came to statistical cognition. It was also found that participants who had previously taken a statistics course(s), whether at Allegheny College or a different institution, did not have significant differences in cognition scores. However, there was a significant difference between the statistical self-efficacy scores of males and females, with males having higher self-efficacy.

Theoretical Implications

The difference in statistical self-efficacy between genders was not a huge surprise. In fact, these results were in line with those found in previous studies, more than likely due to the common stereotype threat involved with women and mathematics (Johns et al., 2005). If women are already displaying these trends, and there is evidence that minorities are following along the same path (Johns et al., 2005), then that means fields which rely on mathematical knowledge will

end up being dominated by White men, leading to an extreme lack of diversity. The fact that women are *still* following this trend after nearly two decades, tells us that as a society, something is being done wrong. Previous research has found that if stereotype threat can be reduced in some way, the mathematical abilities of women improve (Keller & Dauenheimer, 2012). Starting young, since so many women start to find diminished self-efficacy as early as six-years-old (Bian et al., 2017), and combating stereotype threat could possibly lead to an increase in mathematical performance by women, and their representation in related fields.

Again, there was no significant effect of discipline on statistical cognition. This has two main-takeaways: (1) the education of statistics across all disciplines has a similar coverage, and (2) that knowledge is not complete, and can always get better. This is a very important takeaway, because it gives argument to support the equal knowledge of all disciplines. Previous research has shown that earth scientists often believe that those in the social sciences are less competent than those in the natural sciences (Kirby et al., 2019) and the humanities are seen as less valuable than "the sciences" in the eyes of the public (Liu et al., 2022). It is important to acknowledge these perceptions, but also understand that there is no true basis for them.

Applied Implications

This research can be applied to classroom settings at all levels/experiences, especially if further research is done. By having a better understanding of what children/teenagers know about statistics, and what some of the reasons may be for their lack of understanding, we can become better equipped to teach them and help them to succeed.

Limitations/Future Research

One obvious limitation to this study was the small sample size, with only 38 participants, before a few needed to be excluded. This left only 35 points of data to work with. Even those

analyses which did not fail the equal variance tests (gender against self-efficacy and discipline against cognition), could have had a more even and diverse spread if more participants were able to be recruited. To go along with this, more participants would also lead to an increase of power, the ability of analyses to show significant results when they actually exist. So, just because there was no significant difference found, it does not mean that there is *not* a difference.

Another limitation is the lack of diversity. The majority of participants identified as female, while also being in the social sciences. This is likely because of the study being published on a psychology-based recruiting website. Because of the advertisement being on a psychology-based website, and psychology being a social science, it was not a surprise that data was skewed with many participants identifying as a part of the social sciences. This also could have influenced the results since the information in the survey used was created from material presented in the psychology statistics courses.

New studies in this line of research could try to gain a more diverse group of participants, including but not limited to, race, ethnicity, gender, disciplines, age, and even the amount/type of statistics courses previously taken. As previously mentioned, the majority of participants met similar demographics. Although it is typically believed that women are not as gifted in mathematics, there has also been research to show this stereotype threat has been extending to racial minorities (Johns et al., 2005). This future research can expand as far as looking into cultural differences, since there have been differences in relation to intelligence levels and, more specifically, mathematical ability found in separate studies which were conducted in the United States and Europe/Latin America (Bian et al., 2017; Pina et al., 2021). Adult learners—anyone attending college who is above 22-years-old—have been previously found to have significantly

lower levels of math self-efficacy than their traditionally aged peers (Jameson & Fusco, 2014). It would be interesting to see whether or not this is a consistent trend almost ten years later.

Although there was no significant difference between statistical cognition scores when split by discipline, it would be interesting to do a more in depth analysis in the future to see if there was a difference between specific questions. Perhaps some disciplines are better adapted with information related to definition, while others may understand application based questions. This could be better analyzed by first validating a scale and running an exploratory factor analysis to better group items. By looking at which items correlate together, it would be easier to understand what connects them, and also, whether or not disciplines score differently depending upon the grouping.

Following a similar line of thinking, it would be interesting for future research to look into the connection between the type of statistics classes which were taken and statistical cognition scores. This could focus on whether classes in different disciplines focused on the computational/theoretical aspect of statistics, or using different programs to get results and then interpreting them. This would give a better idea about the type of education being offered in disciplines, and the differences and similarities they have. Again, to better test for this, a statistical cognition survey could be drafted, and validated, with questions which focus on the theory behind statistics, computing them, and interpreting program results. Of course, more research would need to be done prior, to determine the types of information which would be relevant to include and to form hypotheses.

Additional analyses and research could also be done to see if there is an effect of the amount of statistics courses which were taken on the participant's statistical cognition. Perhaps the amount of statistics you are exposed to—and the time you spend doing it—affects your

understanding of the topic. Previous research has found that prior achievement in math courses is a significant predictor of mathematical self-efficacy (Grigg et al., 2018). However, this research does state that this is limited to the initial level of self-efficacy, so it would be intriguing to further look into the relationship between experience and self-efficacy, specifically related to math and statistics.

References

Bandura, A. (1997). Self-efficacy: The exercise of control. *New York: Freeman.*

Batanero, C. (2002). The Roles of Models in Understanding and Improving Statistical Literacy. *International Statistical Review/Revue Internationale de Statistique, 70(1),* 37-40. https://doi.org/10.1111/j.1751-5823.2002.tb00340.x

Bian, L., Leslie, S-J., & Cimpian, A. (2017). Gender stereotypes about intellectual ability emerge early and influence children's interests. *Science, 355(6323),* 389-391. https://doi.org/10.1126/science.aah6524

Cerrito, P. B. (1999). Teaching Statistical Literacy. *College Teaching, 47(1),* 9-13. https:doi.org/10.1080/87567559909596070

Coudin, G., & Alexopoulos, T. (2010). 'Help me! I'm old!' How negative aging stereotypes create dependency among older adults. *Aging & Mental Health*, *14(5),* 516-523. https://doi.org/10.1080/13607861003713182

Cumming, G., Beyth-Marom, R., & Fidler, F. (2008). Statistical Cognition: Towards Evidence-Based Practice in Statistics and Statistics Education. *Statistics Education Research Journal*, *7*, 20-39. http://www.stat.auckland.ac.nz/serj

Diana, K. M. (2009). Mathematics self-efficacy and anxiety questionnaire [Unpublished doctoral dissertation]. University of Georgia.

Froehlich, L., Martiny, S. E., Deaux, K., Goetz, T., & Mok, S. Y. (2016). Being smart or getting smarter: Implicit theory of intelligence moderates stereotype threat and stereotype lift effects. *The British Psychological Society, 55(3),* 564-587. https://doi.org/10.1111/bjso.12144

Gigerenzer, G. (2008). Sound reasoning requires statistical understanding. *Science News*, 32. https://www.jstor.org/stable/20465736

Grigg, S., Perera, H. N., McIlveen, P., & Svetleff, Z. (2018). Relations among math self-efficacy, interest, intentions, and achievement: A social cognitive perspective. *Contemporary Educational Psychology, 53,* 73-86. https://doi.org/10.1016/j.cedpsych.2018.01.007

Hively, K., & El-Alayli, A. (2014). "You throw like a girl:" The effect of stereotype threat on women's athletic performance and gender stereotypes. *Psychology of Sport and Exercise, 15(1)*, 48-55. https://doi.org/10.1016/j.psychsport.2013.09.001

Jameson, M. M., & Fusco, B. R. (2014). Math anxiety, math self-concept, and math self-efficacy in adult learners compared to traditional undergraduate students. *Adult Education Quarterly, 64(4),* 306-322. https://doi.org/10.1177/0741713614541461

Johns, M., Schmader, T., & Martens, A. (2005). Teaching stereotype threat as a means of improving women's math performance. *Psychological Science, 16(3),* 175-179. http://www.jstor.org/stable/40064197

Kaufmann, L., Ninaus, M., Weiss, E. M., Gruber, W., & Walter, G. (2022). Self-efficacy matters: Influence of students' perceived self-efficacy on statistics anxiety. *Annals of the New York Academy of Sciences,* https://doi.org/10.1111/nyas.14797.

Keller, K., & Dauenheimer, D. (2012). Stereotype threat in the classroom: Dejection mediates the disrupting threat effect on women's math performance. *Personality and Social Psychology Bulletin, 29(3),* 371-381. https://doi.org/10.1177/0146167202250218

Kirby, C. K., Jaimes, P., Lorenz-Reaves, A. R., & Libarkin, J. C. (2019). Development of a measure to evaluate competence perceptions of natural and social science. *PubMed Central, 14(1).* https://doi.org/10.1371/journal.pone.0209311

Lightsey, R. (1999). Albert Bandura and the Exercise of Self-Efficacy. *Journal of Cognitive Psychotherapy, 13(2),* 158-166. Retrieved from https://www.proquest.com/scholarly-journals/albert-bandura-exercise-self-effiacy/docview/89070865/se-2

Liu, A., Droge, A., Kleinman, S., Thomas, L., Baciu, D. C., & Douglass, J. (2022). What everyone says: Public perceptions of the humanities in the media. *Daedalus, 151(3),* 19-39. https://www.jstor.org/stable/48681141

Pina, V., Martella, D., Chacón-Moscoso, S., Saracostti, M., & Fenollar-Cortés, J. (2021). Gender-based performance in mathematical facts and calculations in two elementary school samples from Chile and Spain: An exploratory study. *Frontiers in Psychology, 12*, 1-9. https://doi.org/10.3389/fpsyg.2021.703580

Richardson, F. C. & Suinn, R. M. (1972). The mathematics anxiety rating scale: Psychometric data. *Journal of Counseling Psychology, 19(6),* 551-554. https://doi.org/10.1037/h0033456

Samuel, T. S. & Warner, J. (2019). “I can math!”: Reducing math anxiety and increasing math self-efficacy using a mindfulness and growth mindset-based intervention in first-year students. *Community College Journal of Research and Practice, 45(3),* 205-222. https://doi.org/10.1080/10668926.2019.1666063

Schutz, P. A., Drogosz, L. M., White, V. I., & Distefano, C. (1998). Prior knowledge, attitude, and strategy use in an introduction to statistics course. *Learning and Individual Differences, 10(4),* 291-308. https://doi.org/10.1016/S1041-6080(99)80124-1

Scwab-McCoy, A. (2019). The state of statistics education research in client disciplines: Themes and trends across the university. *Journal of Statistics Education, 27(3),* 253-264. https://doi.org/10.1080/10691898.2019.1687369

Spencer, S. J., Steele, C. M., & Quinn, D. M. (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology, 35(1),* 4-28. https://doi.org/10.1006/jesp.1998.1373

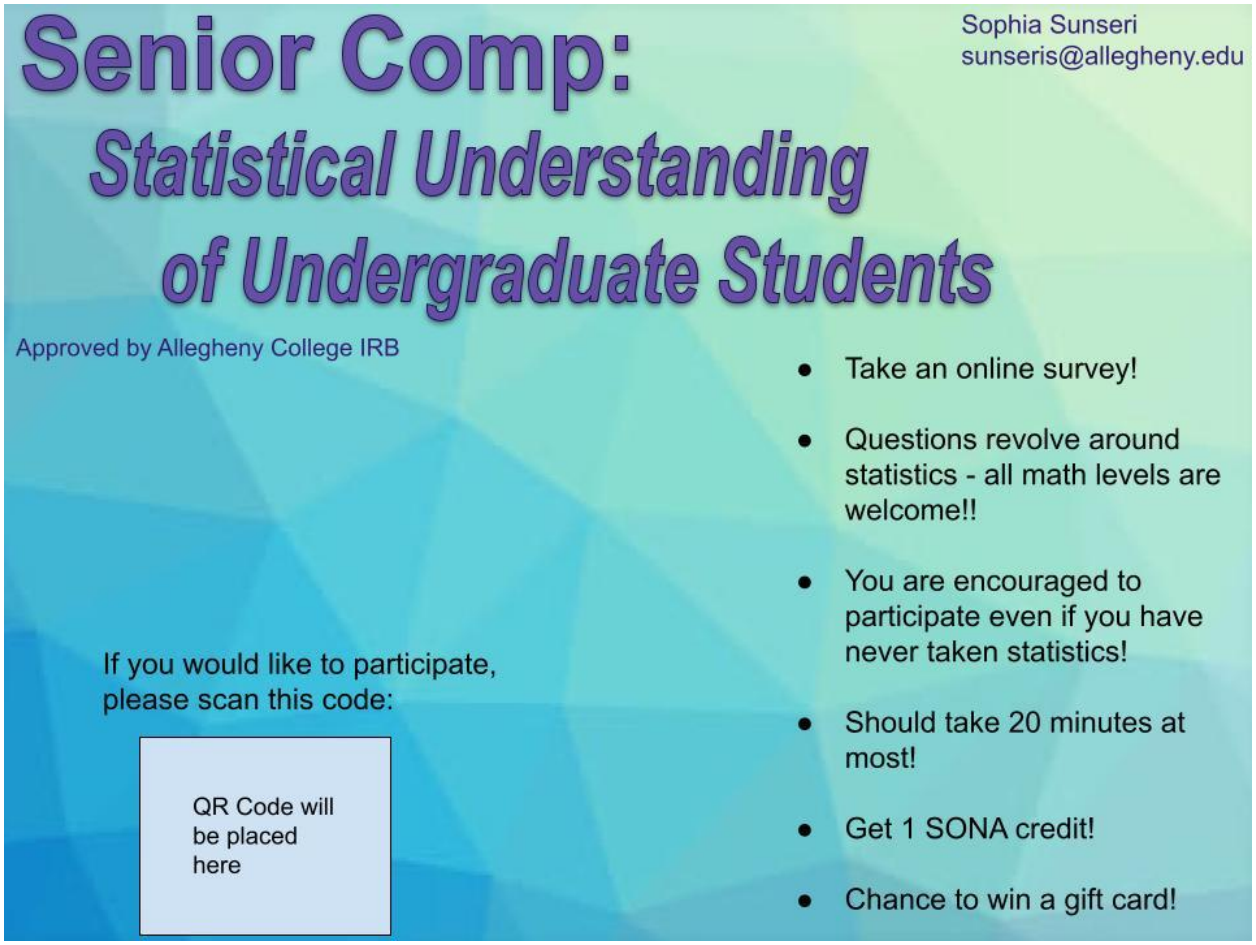
Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology, 69(5),* 797-811. https://doi.org/10.1037/0022-3514.69.5.797

Tipton, R. M., & Worthington, Jr. E. L. (1984). The Measurement of Generalized Self-Efficacy: A Study of Construct Validity. *Journal of Personality Assessment, 48(5)*, 545-548. https://doi.org/10.1207/s15327752jpa4805_14

Tomasetto, C., Alparone, F. R., & Cadinu, M. (2011). Girls' math performance under stereotype threat: The moderating role of mothers' gender stereotypes. *Developmental Psychology, 47(4),* 943-949. https://doi.org/10.1037/a0024047

Usher, E. L., & Pajares, F. (2008). Sources of self-efficacy in mathematics: A validation study. *Contemporary Educational Psychology, 34(1),* 89-101. https://doi.org/10.1016/j.cedpsych.2008.09.002

Appendix A: Recruitment Flyer



Appendix B: Raffle Form

1. Please enter your name. (Short answer)
2. Please enter your Allegheny email address. (Short answer)

Appendix C: Statistical Self-Efficacy

All items will be measured on a 7-point Likert scale:

1 = Never; 2 = Rarely; 3 = Seldom; 4 = Sometimes; 5 = Often; 6 = Usually; 7 = Always.

1. I get tense when I prepare for a statistics test.
2. I get nervous when I have to use statistics outside of school.
3. I believe I can do well on a statistics test.
4. I worry that I will not be able to get a good grade in my statistics courses.
5. I believe I can complete all of the assignments in a statistics course.
6. I believe I am the kind of person who is good at statistics.
7. I feel stressed when listening to statistics instructors in class.
8. I believe I can understand the content in a statistics course.
9. I believe I can get an "A" when I am in a statistics course.
10. I get nervous when asking questions in class.
11. Working on statistics homework is stressful for me.
12. I believe I can learn well in a statistics course.
13. I worry that I do not know enough statistics to do well in future statistics courses.
14. I feel confident when taking a statistics test.
15. I believe I am the type of person who can do statistics.
16. I feel that I will be able to do well in future statistics courses.
17. I believe I can do the statistics in a statistics course.
18. I get nervous when taking a statistics test.

19. I am afraid to give an incorrect answer during my statistics class.
20. I believe I can think like a statistician.
21. I feel confident when using statistics outside of school.
22. I feel confident enough to ask questions in my statistics classes.

Note: Items 1, 2, 4, 7, 10, 11, 13, 18, and 19 will be reverse-coded.

Appendix D: Statistical Cognition Questionnaire

All items will be on a *True* or *False* scale.

1. It is best to report the mean, compared to other forms of central tendency, when data is normally distributed.
2. It is best to report the median, compared to other forms of central tendency, when data is skewed.
3. An outlier is an extreme score in a data set.
4. A data set can have only one mode.
5. A normal distribution is a bell curve.
6. A skewed graph tells us there are extreme values in the data set.
7. A normal distribution should be symmetric and be dense in the middle.
8. When the p-value is less than α, we reject the null.
9. A z-score tells us the number of standard deviations an individual is away from the sample mean.
10. A z-score is able to tell us the relationship between individual scores.
11. If the p-value is less than α, that means the result was not significant.
12. A cause of Type II error is running too many analyses.
13. Variables can correlate without actually being related to each other.
14. Correlation is the same as causation.
15. If the 95% confidence interval for a sample mean is [3,15] it means that there is a 95% chance that the population mean is between 3 and 15.
16. When there is a large sample size, the results are less likely to be statistically significant.

17. Statistical significance implies that there is real world significance.

18. Effect size focuses on how much groups overlap.

19. Power refers to the ability to detect a significant relationship when one actually exists.

20. If there is a high probability of Type II Error, then power is also high.

21. If $p < .05$, then the null hypothesis is proven wrong.

22. The null hypothesis states that there is no relationship between the variables in a study.

Appendix E: Demographic Questionnaire

1. What is your class year?

- First-Year
- Sophomore
- Junior
- Senior
- 5th year

2. What is your age? (Open-ended)
3. What is your gender?
    - Male
    - Female
    - Nonbinary
    - Prefer to self-describe: _____
    - Prefer not to say
4. What is your race?
    - White
    - Black or African American
    - American Indian or Alaska Native
    - Asian
    - Native Hawaiian or Pacific Islander
    - Other: _____
5. Are you Hispanic or Latino/a?
    - Yes, I am (Please Explain): _____

- No, I am not

6. What is/are your declared/intended major(s)? (Using square boxes)

   - Psychology

   - Biology

   - Economics

   - Mathematics

   - Global Health Studies

   - Computer Science

   - Other: _____

7. Have you taken a statistics course previously at Allegheny?

   - Yes

   - No

8. If so, in what department? (Using square boxes)

   - Psychology

   - Biology

   - Economics

   - Mathematics

   - Global Health Studies

   - Computer Science

   - Other: _____

9. Have you taken a statistics course previously NOT at Allegheny?

   - Yes

   - No

10. If so, in what department? (Using square boxes)

    - Psychology

    - Biology

    - Economics

    - Mathematics

    - Global Health Studies

    - Computer Science

    - Other: _____

Appendix F: *Informed Consent Form*

Study Title: Statistical Understanding of Undergraduate Students

Research Investigator's Name and Department: Sophia Sunseri (Psychology)

Contact information for researchers: sunseris@allegheny.edu

Introductory Statement

This study aims to examine undergraduate students' knowledge of statistics. This includes the types of materials that are traditionally taught in introductory statistics courses.

What is the purpose of this study? To get a better understanding of statistical knowledge in undergraduate students.

What will I do in this study? You will be asked to complete a short survey that should take no longer than 20 minutes. In the survey you will be asked basic demographic questions, including your major and whether or not you have taken a statistics course before. You will be asked about your knowledge and confidence regarding introductory statistics topics. You will not be asked to compute any statistics during this survey.

How long will it take me to do this? Your participation will last no longer than 20 minutes.

Are there any risks of participating in the study? Participation in this experiment will entail no more than minimal risk. However, if you are unable to continue for any reason, you will be allowed to withdraw from the experiment immediately and without penalty.

What are the benefits of participating in the study? There are no direct benefits for participating in this study. The information that you provide will help to give a better understanding about what statistical education is taught in undergraduate studies.

Will anyone know what I do or say in this study (Confidentiality)? No one will know what you say or do in this study. The researcher will not collect names. You will be asked your major, and whether you have been enrolled in previous statistics courses; however, no other identifying questions will be asked. To help ensure confidentiality, participants' responses will be combined with other participants' responses, and all data files will be stored in Qualtric's database and is only accessible to the student researcher. If the results and datasets are presented publicly, the data will be presented without any identifying features.

Will I receive any compensation for participation? Psychology students will be able to receive one SONA credit upon completion of this study. Unfortunately, there is no direct compensation for participating in this study for students not currently enrolled in a psychology course.

Is there a different way for me to receive this compensation or the benefits of this study? No.

Who can I contact for information about this study? For further information about the study contact Sophia Sunseri, (sunseris@allegheny.edu).

You are free to refuse to participate in this research project or to withdraw your consent and discontinue participation in the project at any time without penalty or loss of benefits to which you are otherwise entitled. Your participation will not affect your relationship with the institution(s) involved in this research project.

If you are not satisfied with the manner in which this study is being conducted, you may report (anonymously if you so choose) any complaints to the chair of the Allegheny College Institutional Review Board at IRB@allegheny.edu

Due to the online nature of the study, there is no spot for participants to sign their name. However, the following questions will be asked following the informed consent form:

Do you consent to participating in this study, and verify that you are at least 18 years of age?

1. Yes, I verify I am at least 18 years old and wish to participate in this study.
2. No, I do not wish to participate in this study.
3. No, I am not at least 18 years old.

Appendix G: Debrief

The purpose of this study was to see how well undergraduates at Allegheny College understand statistics. There was hope to look at differences in statistical understanding across major disciplines as well as gender.

Today, you were asked to go through the math self-efficacy, statistical scale, and demographics and answer anything you felt comfortable with.

If you have any interest in knowing the results upon completion of the research, or find that you have questions in the future, please do not hesitate to reach out to Sophia Sunseri, (sunseris@allegheny.edu). You may also contact one of the research advisors, Dr. Christopher Normile (cnormile@allegheny.edu). If you have questions/concerns regarding the process of this project, you may contact the chair of the IRB subcommittee, Dr. Monali Chowdhury (mchowdhury@allegheny.edu).

If you find this area of research interesting, consider the following articles:

Beyth-Marom, R., Fiddler, F., Cumming, G. (2008). Statistical Cognition: Towards evidence-based practice in statistics and statistics education. *Statistics Education Research Journal, 7(2),* 20-39.

Gigerenzer, G. (2008). Sound reasoning requires statistical understanding. *Science News, 174(11),* 32.

Batanero, C. (2002). The role of models in understanding and improving statistical literacy. *International Statistical Review / Revue Internationale de Statistique, 70(1),* 37-40.