Proceedings of The National Conference
On Undergraduate Research (NCUR) 2014
University of Kentucky, Lexington, KY
April 3-5, 2014

A Study of Undergraduate Understanding of Sampling Variability and Regression

Lauren M. Magee
Department of Mathematics
Loyola Marymount University at
1 LMU Drive
Los Angeles, California 90045 USA

Faculty Advisor: Dr. Anna E. Bargagliotti

Abstract

Project-SET is an NSF funded project aimed at developing teacher level materials to better facilitate student learning in statistics. The project is focused on two statistics topics; sampling variability and linear regression. After an extensive literature review, four instructional videos aimed at introducing sampling variability and regression were created, two videos per topic of interest. To test the teaching capabilities of these videos, a pre- and post-test was administered to two college level statistics courses at Loyola Marymount University before and after they viewed the videos. Participants provided definitions, solved problems, and rated their confidence in answering questions pertaining to each topic. Results indicate that the videos allowed students to gain a significant amount of confidence in answering questions regarding both topics, however their ability to correctly solve questions on these topics did not improve. Upon further analysis of our data, student misconceptions about sampling variability and linear regression were identified to help direct modifications in future iterations of the instructional videos.

Keywords: Sampling Variability, Linear Regression, Statistics Education

1. Introduction

The study of statistics is growing in the United States. For example, the number of students taking AP statistics has increased from 7,500 in 1997 to 153,859 in 2012 (College Board, 2012). At the undergraduate level, many students take introductory statistics nationwide as part of their major or to fulfill a general education requirement. As data becomes more readily available and data-driven decision-making is touted as a gold standard in all facets of society, it is increasingly important for undergraduate students to be adequately prepared in statistics.

Two fundamental topics in introductory statistics are the study of sampling variability and the study of regression. At the undergraduate level, the study of sampling variability and the study of regression are integral parts of the statistics curriculum. As sampling variability and sampling distributions are the gateways to formal inference, a great deal of time is typically spent in undergraduate statistics courses studying the topic (see popular textbooks by Agresti & Franklin, Moore, McCabe, and Craig, or Mendenhall, Beaver & Beaver). Single variable and multivariate regression analysis is also emphasized in college courses.

While these topics are important and included in undergraduate courses, several papers discuss issues students have with sampling variability and regression. In relation to the topic of regression, analyzing scatter plots and fitting a line of best fit have proved to be very challenging for secondary students. These difficulties might stem from students inability to conceptualize data as a whole. They are instead prone to interpreting data as a series of individual cases which by themselves can not suggest a pattern. This case-oriented view inhibits students from identifying a correlation between two variables unless there is a strong linear relationship (Estepa & Batanero, 1996). Additionally, the use of a line of best fit in modeling data is a concept subjected to multiple misconceptions. Some misconceptions that have been indentified in the placing of the line of best fit are as follows: the line should go through the maximum number of points possible, the line must go through the origin, there must be the same number of points above and below the line, the line must pass through the left-most and right-most points or the highest and lowest points on the scatter plot, or the line must be placed visually close to a majority or cluster of the points (Sorto, White, & Lesser, 2011).

There are also quite a few misconceptions relating to the topic of sampling variability as well. Teachers at all levels maintain that variation among data is natural and an essential component of statistics (Utts, 2003). However, students have repeatedly, over multiple studies, shown difficulty identifying, accepting, and dealing with sample variation (delMas et al., 2007; Konold et al., 1997). These challenges are reinforced in a study done by Shaughessy et al. (2004) that surveyed 272 middle- and high- school students and found only 8 of their participants acknowledged any variation among data when it was present. Another study administered a pre-test and post-test to 746 undergraduate students and found that only 42.8% on the pre-test and 50% on the post-test understand unexpected patterns in sampling variability. Even fewer students, 20%, recognized the importance of estimating the sampling error when constructing an informal inference about the sample mean (delMas et al., 2007). A key component in improving students understanding of sampling variability is providing them with opportunities to use their own reasoning and methods to deal with variation (McClain, McGatha, & Hodge, 2000).

To address the teaching, learning, and application of sampling variability and regression, the NSF funded Project-SET (www.project-set.com) developed materials to enhance the ability of teachers to foster student statistical learning. The undergraduate project discussed in this paper was part of this larger project. The goal of the undergraduate study was to develop a set of instructional videos to teach sampling variability and regression and then determine whether instructional videos developed as part of a Research Experience for Undergraduates (REU) were effective in helping students increase their knowledge on the topics.

In light of the importance of these topics, it is important to uncover whether undergraduate students have a deep understanding of these topics at different points in their studies. The research questions to guide the study are: (1) Do the instructional videos increase student ability to answer typical sampling variability and regression questions? (2) Do the instructional videos increase student confidence in their ability to solve questions in both topics? (3) Do the instructional videos increase the accuracy of students' topic definitions? The goal of this paper is to report the results from the study.

2. Methods

During the summer of 2013, as part of an REU, the author developed four videos—two for sampling variability and two for regression, outlining the appropriate steps to solve a problem concerning either topic. These videos are a compilation of PowerPoint slides accompanied by audio that explains the processes further. The four videos can be viewed at the following links: http://vimeo.com/63795125, http://vimeo.com/63795126, http://vimeo.com/63777934, http://vimeo.com/63795266. To test if these video explanations were significant in improving student learning, a pre and post-test was developed using fluidsurveys.com for undergraduate students to complete and offer feedback.

Prior to completing the tests, demographic questions were asked of participants.

1. *Name*
2. *Major*
3. *Expected Year of Graduation*
4. *Have you taken a Statistics course prior to your current enrollment? (yes/no)*
5. *Did you cover sampling variability in these courses?(yes/no)*
6. *Did you cover regression in these courses?(yes/no)*

Then the students were asked to give a definition of their personal understanding of sampling variability and regression and rate their confidence in their ability to solve problems related to such topics.

7. *Are you familiar with the concept of sampling variability? (yes/no)*
8. *Are you familiar with the concept of regression? (yes/no)*
9. *Please explain your personal understanding of sampling variability?*
10. *Please explain your personal understanding of regression?*

11. *How comfortable are you with questions involving sampling variability?(Very comfortable/Moderately comfortable/Neutral/Moderately Uncomfortable/Very Uncomfortable)*
12. *How comfortable are you with questions involving regression?(Very comfortable/Moderately comfortable/Neutral/Moderately Uncomfortable/Very Uncomfortable)*

The demographic questions were followed by the pre-test, consisting of two multifaceted problems that covered different aspects of the topics. The pre-test problems were adapted from tasks for sampling variability and regression available on the Illustrative Mathematics Project (IMP) website (illustrativemath.org). The following questions were edited to fit the online survey format. The first question gave participants a table of the number of U.S. households each year between 1998 and 2004. A scatterplot was also shown and participants were asked if the graph was linear. After establishing linearity, participants were asked to find the line of best fit and then interpret the slope and y-intercept in the terms of the problem.

Table 1. The table below shows the number of households in the U.S. in the years 1998-2004 [data source: www.census.gov].

Year	1998	1999	2000	2001	2002	2003	2004
Households (in thousands)	97,107	98,990	99,627	101,018	102,528	103,874	104,705

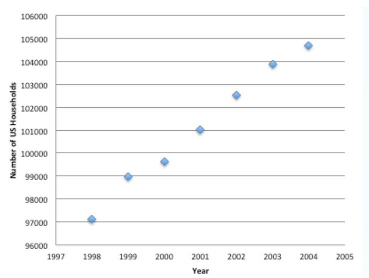


Figure 1.

1. *Does the graphical representation of these data look linear? Explain.*
2. *Determine which of the following options you think represents the line of best fit. (Option A, B, C, D, or Other)*

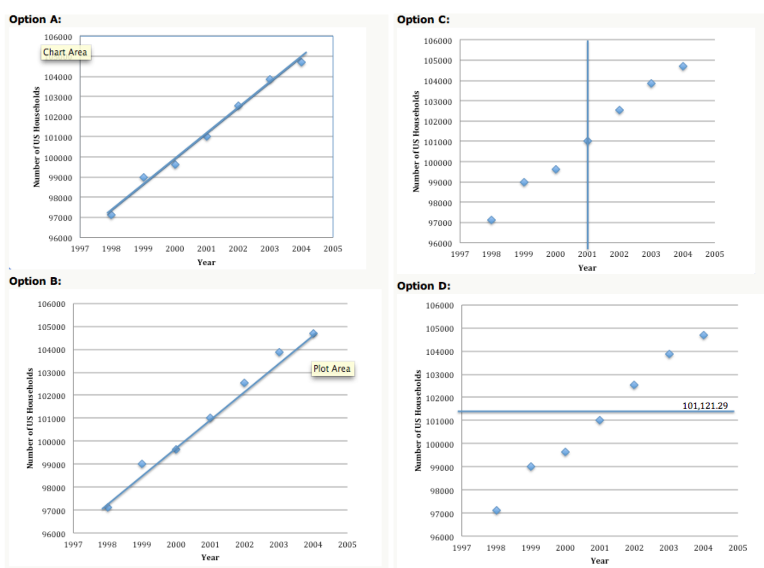


Figure 2.

3. *Use technology (i.e. a graphing calculator), to find the equation of the line of best fit.*
4. *Looking at the equation you entered above, interpret the slope and y-intercept of the equation in the context of the problem.*

The second question, dealing with the topic of sampling variability, concerns a Valentine's Day contest that asks what proportion of red and white marbles are present in a jar. The proportions are given from 100 samples that were taken. Participants are then asked to draw conclusions about the marbles jar from the sampling distribution provided.

A hotel holds a Valentine's Day contest where guest are invited to estimate the percentage of red marbles in a huge clear jar containing both red marbles and white marbles. There are 11,000 total marbles in the jar: 3696 are red, 7304 are white. The actual percentage of red marbles in the entire jar (33.6% = 3696/11000) is known to some members of the hotel staff.
Any guest who makes an estimate that is within 9 percentage points of red marbles in the jar wins a prize, so any estimate from 24.6% to 42.6% will be considered a winner. To help with the estimating, a guest is allowed to take a random sample of 16 marbles from the jar in order to come up with an estimate. (Note: when this occurs, the marbles are then returned to the jar after counting.)
One of the hotel employees who does not know that the true percentage of red marbles in the jar is 33.6% is asked to record the results of the first 100 random samples. A table and dotplot of the results appear below.

Table 2.

Percentage of red marbles in the sample of size 16	Number of times the percentage was obtained
12.50%	4
18.75%	8
25.00%	15
31.25%	22
37.50%	20
43.75%	12
50.00%	12
56.25%	4
62.50%	2
68.75%	1
Total:	100

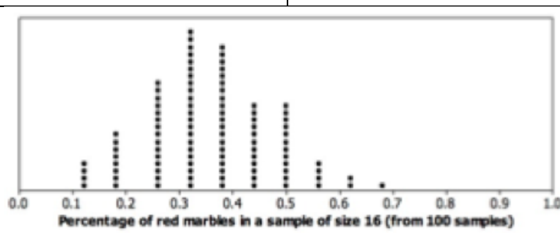


Figure 2.

For example, 15 of the random samples had exactly 25.00% red marbles, only 2 of the random samples had exactly 62.50% red marbles, and so on.

1. *Assuming that each of the 100 guest who took a random sample used their random sample's red marble percentage to estimate the whole jar's red marble percentage. Based on the table above, how many of these guests would be "winners"?*
2. *How many of the 100 guest obtained a sample that was more than half red marbles?*
3. *Should we be concerned that none of the samples had a red marble percentage of exactly 33.6% even though that value is the true red marble percentage for the whole jar? Explain briefly why a guest can't obtain a sample red marble percentage of 33.6% for a random sample of size 16.*

4. *Recall that the hotel employee who made the table and dotplot above didn't know that the real percentage of red marbles in the entire jar was 33.6%. If another person thought that half of the marbles in the jar were red, briefly explain how the hotel employee could use the doplot and table results to challenge this person's claim. Specifically, what aspects of the table and dotplot would encourage the employee to challenge the claim?*

After the completion of the pre-test problems, the students were asked again to define the terms and rate their confidence with their problem solving skills. The point of this mid-test survey was to gauge if students had altered their responses due to their exposure to questions regarding both topics. Students may have gain understanding on what is expected from them when answering questions centered around these topics, so their definitions and confidence in dealing with them may have altered.

At this point, the instructional video footage, as described earlier, was shown to the students. The sampling variability video lasted 10 minutes and the regression video lasted 12 with both part 1 and part 2 of the videos combined. After watching the videos, participants were again asked to define the terms and rate their confidence with them. In addition, they were also asked questions about the success of our videos and whether they aided in their understanding. To summarize, participants were asked to define the terms and rate their confindence at three points in time: at the start, in the middle after viewing problems, and at the end after viewing the videos.

Two last questions were administered to the students, so that their progress could be observed outside of self-report.

The second questions regarding regression, to test participants after viewing the videos, offered a graphical representation of the height a ball was dropped and its subsequent height after a bounce. Participants were again asked to explain and find the line of best-fit, as well as predict a bounce height based off of the height the ball was dropped at.

Please answer the following problems related to sampling variability and regression.
An eighth grade class was investigating how high a golf ball bounces when it is dropped from different heights. Students were given eight set heights to drop the ball from. Then they dropped a golf ball from each of those heights and measured how high the ball bounced back up. Below is the scatterplot they made of their data. The table below displays the data as well.

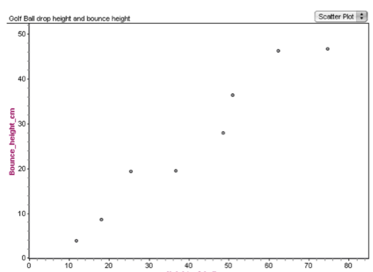


Figure 3.

Table 3.

Height of Ball	Height of Bounce
11	4
19	8
26	20
38	20
50	48
52	52
62	62
76	78

The teacher gave each student a copy of the scatterplot and asked them to find the line of best fit for the data. The teacher then asked each student to explain why they put the line where they did, defining the criterion for the line of best fit.

1. *Here is an answer of one of the students in the class:*

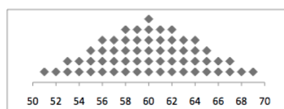
The line needs to start at (0,0) then go through the most dots. I got my line to go through two of the dots so I put it there. Does the student's reasoning always work? Why or why not? If not, give an example where it would not work.

2. *Please explain where you believe the line of best fit would be places on a graphical representation of the above data.*
3. *Find a linear function h that models the bounce height as a function of the drop height.*
4. *Find h(38) and interpret your answer in terms of the bounce height and drop height.*

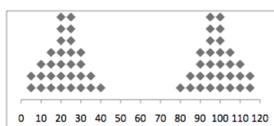
The final question, offered up the sampling distribution of two students' text message sending over a 60-day interval. Students were asked to use these distributions and their means and standard deviations to determine which student texted the most.

Sally and Jamal are students in Mr. Miller's Statistics class. One day in class they get into a discussion about text messaging. Sally is convinced that she does more text messaging than Jamal because she thinks that girls like to communicate with each other more than guys do. Jamal thinks he sends/receives as many text messages as Sally because he recently received a new smartphone for his birthday and is still enjoying being able to communicate more easily with his friends. In their statistics class, they have recently learned about summary statistics, populations, distributions, and sampling.
They decide to use what they have learned to investigate and compare the amount of text messaging they do. The teacher of the class asks Jamal and Sally to provide the class with the their texting record for the past 60 days without looking at it. Here is a dot plot representing the number of texts on each of the 60 days for Sally and Jamal. Although the class does not know this information, note that over the past 60 days, both Sally and Jamal texted an average of 60 texts per day.

Population Distribution Dot Plots



Dot Plot Showing Population Distribution of Text Messages for Sally



Dot Plot Showing Population Distribution of Text Messages for Jamal

Figure 4.

Sally and Jamal are interested in comparing their average number of texts. At this point, the teachers ask each student in the class to draw a random sample of size 10 from the 60 day data they collected. There are 30 students in the class. For each random sample drawn, each student computes the average number of texts for that sample and plots the number on a dot plot. Here are the dot plots they find:

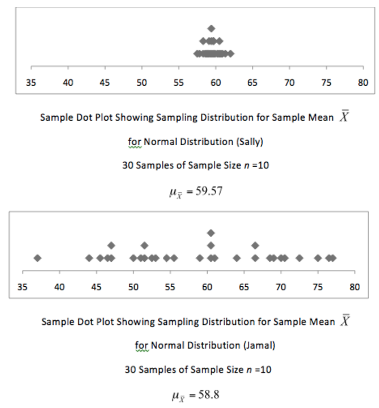


Figure 5.

Using the sample dot plots, the teachers ask the class the following questions.

1. *How many students in the class obtained a sample average that was more than 60 for Sally? For Jamal?*
2. *If Sally thought that she texted on average more than Jamal during the 60 day period, explain whether she could use the sample dot plots to argue her case.*

As stated above, the first research question was: Do the instructional videos increase student ability to answer typical sampling variability and regression questions? The answer to this question was sought out through the sampling variability and regression questions included in the pre- and post-test. An improvement in the proportion of participants' correct responses from the pre- to the post-test responses would be informative as to whether the videos facilitated an improvement in participant answering capabilities. The problems were multiple-choice. In an assessment of noting the improvement in proportion of correct answers, a paired samples t-test was carried out in SPSS.

The second research question was: Do the instructional videos increase student confidence in their ability to solve questions in both topics? The three checkpoints on the pre- and post-test combined serve as the basis for this analysis. Participant confidence before the pre-test, after the pre-test, and after the videos was used to measure improvements. If we noted that participant confidence was increasing, that would be indicative of an increase in their understanding, because they have obtained a greater understanding of how to handle problems of the sort.

The final question was: Do the instructional videos increase the accuracy of students' topic definitions? Originally, definitions were going to be rated based on a specific criteria, but due to the rough definition offered by participants throughout all parts of the study, more analysis was done on the content of their definitions and the misconceptions that can be extracted from such data.

3. Data & Participants

The surveys were distributed to two statistics classes at Loyola Marymount University. The first class it was distributed to was "Introduction to Probability and Statistics" (MATH 360) and is identified as an upper-division mathematics course. This course does not require any prior statistics experience, however it does have a Calculus II prerequisite. This class was composed of around 30 students, but only 12 students completed the survey and were eligible for data interpretation. Students in this course were offered extra credit based off of their participation. This contributed to the low participation seen from this group.

The second course was labeled as "Elementary Statistics" (MATH 104) and was taught by Anna Bargagliotti, the Principal Investigator of Project-SET and faculty mentor for this project. This course requires no statistics experience or pre-requisites and is typically the introductory course in statistics offered for the university. From this class, we collected 40 responses that were eligible for further data analysis. Dr. Bargagliotti made this mandatory for all of her students, and a 100% participation rate was observed.

4. Results

Starting with the background knowledge of our participants, we found that 8% of Math 360 students had taken a statistics course previously while 25% of Math 104 students had taken a prior statistics course. As Math 360 is an upper division course, we would expect them to be better prepared for the survey, however, as it appears in the proportions, the lower division Math 104 class appears to be coming into our survey with more training in statistics. Math 104 students also showed more self-reported familiarity with the topics of sampling variability and regression. These results are depicted in Table 1.

Table 4. Prior Knowledge and Experience

	Math 104	Math 360
Have you taken a Statistics course prior to your current enrollment?	25%	8%
Are you familiar with the concept of sampling variability?	65%	25%
Are you familiar with the concept of regression?	55%	25%

SV = Sampling Variability
Reg = Regression

Participants self-reported comfort level with both topics were measured three times throughout the survey. Table 2 and 3, show the proportion of participants that selected each comfort option at each check-point of the survey.

Table 5. Comfort with Sampling Variability

	Math 104			Math 360		
	Check #1	Check #2	Check #3	Check #1	Check #2	Check #3
Very Uncomfortable	17.5%	12.5%	0%	41.7%	8.3%	16.7%
Moderately Uncomfortable	22.5%	22.5%	7.5%	25%	41.7%	8.3%
Neutral	42.5%	40%	47.5%	33.3%	33.3%	50%
Moderately Comfortable	17.5%	20%	37.5%	0%	16.7%	25%
Very Comfortable	0%	5%	7.5%	0%	0%	0%

Table 6. Comfort with Regression

	Math 104			Math 360		
	Check #1	Check #2	Check #3	Check #1	Check #2	Check #3
Very Uncomfortable	22.5%	15%	5%	41.7%	8.3%	16.7%
Moderately Uncomfortable	12.5%	20%	10%	25%	41.7%	8.3%
Neutral	47.5%	35%	37.5%	33.3%	41.7%	50%
Moderately Comfortable	17.5%	22.5%	37.5%	0%	8.3%	25%
Very Comfortable	0%	7.5%	10%	0%	0%	0%

The data above provides us with evidence that our participants were becoming more comfortable with both topics throughout our survey. This is shown in the increasing proportions of Math 104 students who noted feeling "very comfortable. This pattern can be seen above in the table showing 0% of Math 104 students feeling "very Comfortable" at Check #1 and 10% feeling "Very Comfortable" at Check #3. Another indication of increased confidence is depicted in the decreasing proportion of Math 104 student feeling "Very Uncomfortable", with the proportion at Check #1 being 22.5% and the proportion at check #3 being 5%. These patterns can also be seen in the Math 360 students as well. A paired samples t-test was done on these data in SPSS. The classes were pooled together to form one data set for the t-test, because there did not appear to be any significant differences between groups. Additionally, the Math 360 sample size was very small and therefore would likely produce inaccurate statistics. The following results were observed from the t-test.

Table 7. SPSS results of a paired samples t-test

	Mean	Std. Deviation	t-value	Significance
Sampling Variability	-.86538	.74172	-8.413	3.3003E-11
Regression	-.78846	.82454	-6.896	1.9576E-9

From the data table above, we can conclude that participant comfort with both sampling variability and regression did improve over the course of our survey. With both topics having an observed p-value of less than .05, we can conclude that these results are significant.

However, because their comfort was measured through self-report, participants may have been providing us with the responses they believed we were looking for. We can see evidence of this potential bias in asking our participants about the success of our videos in teaching our participants how to solve questions involving sampling variability and regression.

Table 8. Video Success, Self-Reported "yes" Responses

	Math 104	Math 360
Did you find the videos clear and informative?	90%	83.3%
Did you learn anything new about sampling variability during the videos?	75%	83.3%
Did you learn anything new about regression during the videos?	72.5%	83.3%
Did you find the videos helpful in increasing your understanding about sampling variability?	77.5%	83.3%
Did you find the videos helpful in increasing your understanding about regression?	70%	75%

SV = Sampling Variability
Reg = Regression

The self-reported data is very encouraging, but the real signifier of a successful instructional video lies in our participants' ability to solve actual problems regarding both topics. If their abilities have increased, then we know our videos are effective, but if they have not increased, our videos may need to be modified. Table 5 and 6 depict the proportion of participants that correctly solved sampling variability and regression problems before and after the videos were shown.

Table 9. Sampling Variability Solving, Correct Responses

	Math 104		Math 360	
	Pre-Video	Post-Video	Pre-Video	Post-Video
Question #1	37.5%	15%	75%	8.3%
Question #2	57.5%	7.5%	58.3%	8.3%
Question #3	62.5%	45%	83.3%	33.3%
Question #4	70%	N/A*	83.3%	N/A*

*There was no Question #4 on sampling variability included in the post-test.

Table 10. Regression Solving, Correct Responses

	Math 104		Math 360	
	Pre-Video	Post-Video	Pre-Video	Post-Video
Question #1	95%	80%	100%	91.7%
Question #2	90%	17.5%	91.7%	16.7%
Question #3	20%	20%	41.7%	33.3%
Question #4	7.5%	12.5%	25%	25%

From the table, we observe that a lower proportion of our participants were answering the post-video questions correctly than the pre-video questions. This could be attributed to the difficulty level of the questions, as the pre-video questions paralleled the format of our videos and the post-video questions were more deviant. However, with this information, we cannot conclude that our videos were successful in increasing participant ability to solve questions in both topics of interest.

Another goal of our videos was improving how our participants defined each topic. The survey had three checkpoints asking how you would, at this stage of the survey, define the topic of sampling variability and define the topic of regression. Scoring was done for these data that accessed student definitions as either “correct” or “incorrect”. However, because of the nature of these definitions, we found it more telling to analyze the common themes among participant definitions, instead of simply providing an overall score. An overall theme among both groups definitions, was that they would pick out a small detail they noticed from the problem set and focus on it when defining the topics. This most commonly included the identification of a graph, slope-intercept form equation, and line of best fit for regression definitions. The following are actual participant definitions of regression: “the measured association of two or more variables”, “line of best fit, slope, y-intercept”, “being able to look at points where no data is given from the data we were given/found”. Sampling variability often included references to “samples and how they vary from one another”. While these terms and phrases are not incorrect, they are not very strong either, and in most cases, participants could have simply inferred the definition from the terms themselves and the contexts of the problems. Some students became frustrated with the repetitive asking of the questions and resorted to responses, such as “same as before”, which may have hindered some improvement patterns from being seen.

5. Discussion

It is important to note that we found a significant increasing in participant confidence after they had been shown the instructional videos. Participant ability to provide relevant information in defining our topics did also improve, however many had trouble compiling topic terms into a holistic definition. As addressed above, student problem solving did not improve after viewing the videos we had prepared. This again could be due to the difficulty levels of the problems as the pre and post-tests were not aligned effectively. The video instruction could have also focused too finely on one type of problem rather than offering a more holistic perspective on both topics. Additionally, by necessity, the pre and post-test was taken in one sitting, which can be very taxing on participants. Because of this, many elected to skip the solving of a few problems, especially in the post-test, so this could also lead to inaccurate results. An article written in the Journal of Applied Psychology highlights these biases (Podsakoff, MacKenzie, Lee, Odasakoff, 2003).

There were quite a few subtle differences between the classes we studied. The Math 104 course students seemed to be much more confident in their abilities from the start of the pre-test and they continued to show higher confidence levels throughout the rest of the study. Additionally, these individuals had more experience in statistics prior to their class than those in the Math 360 course, which could have added to this confidence. Additionally, the Math 360 class made the problems of regression much more algebra based than statistics, possibly illustrating their training in advanced calculus courses. However, it is important to identify the limitations to these results. The Math 360 students likely represent a biased sample, because they were only given extra-credit as an incentive instead of the Math 104 students, who were required to complete the survey. Additionally, the sample sizes of both groups were relatively small and therefore may not be representative of the population.

Current generations of students are becoming more technologically advanced and are therefore turning to online outlets, such as Youtube, to provide them with information. The videos outlined in this study were thus created to offer students an education alternative from typical literature. Moving forward, our video scripts will be edited and re-recorded for further clarity about the topics. Through this REU and research project, misconceptions were found among student definitions and solving methods, that offered incite into where the videos were falling short in their instruction. This understanding will result in more accurate videos and tests created in the future for this project.

6. Acknowledgments

Anna Keathley deserves acknowledgment for her hard work and dedication to Project-SET and the data collection process of this analysis.

7. References

1. Agresti & Franklin, Statistics: The Art and Science of Learning from Data (Boston: Pearson Education, 2013).
2. College Board (2012). Archived AP Data: Exam data from administrations since 1997. Retrieved from professionals.collegeboard.com/common/shared/html/
3. delMas, Garfield, Ooms, Chance, "Assessing students' conceptual understanding after a first course in statistics," *Statistics Education Research Journal* 6 (2007): 28-58.
4. Estepa & Batanero, "Judgments of correlation in scatter plots: Students' intuitive strategies and preconceptions," *Hiroshima Journal of Mathematics Education* 4 (1996): 25-41. http://www.ugr.es/~batanero/pages/ARTICULOS/Hiroshima1996.PDF
5. Konold, Pollatsek, Well, & Gagnon, "Students Analyzing Data: Research of Critical Barriers," *Research on the role of technology in teaching and learning statistics: Proceedings of the 1996 IASE Roundtable Conference* (1997): 152-167. https://www.srri.umass.edu/sites/srri/files/KonoldEtal1997.pdf
6. McClain, McGatha, & Hodge, "Improving data analysis through discourse," *Mathematics Teaching in the Middle School* 5 (2000): 548-553.
7. Mendenhall, Beaver & Beaver, Introduction to the Practice of Statistics 8th Edition (New York: W.H. Freeman and Company, 2014).
8. Moore, McCabe, and Craig, Probability and Statistics 12th Edition (Duxbury: Thomson Brooks/Cole, 2006).
9. Podsakoff, MacKenzie, Lee & Podsakoff, "Common method biases in behavioral research: A critical review of the loterature and recommended remedies," *Journal of Applied Psychology* 88 (2003): 879-903. doi:10.1037/0021-9010.88.5.879
10. Shaughnessy, Ciancetta & Canada, "Types of Student Reasoning on Sampling Tasks," *The Proceedings of the 27th meeting of the International Group for Psychology and Mathematics Education* 4 (2004): 177-184. http://www.emis.de/proceedings/PME28/RR/RR045_Shaughnessy.pdf
11. Sorto, White, & Lesser, Understanding Student Attempts to Find a Line of Fit," *Teaching Statistics* 33 (2011): 49-52. doi:10.1111/j.1467-9639.2010.00458.x
12. Utts, "What Educated Citizens Should Know About Statistics and Probability," *The American Statistician* 57 (2003): 74-79. doi: 10.1198/0003130031630