DEMOGRAPHIC ANALYSIS OF STUDENT EVALUATIONS

A Thesis
Submitted to the Graduate Faculty
of the
North Dakota State University
of Agriculture and Applied Science

By

Lucas James Huebner

In Partial Fulfillment of the Requirements
for the Degree of
MASTER OF SCIENCE

Major Department:
Statistics

May 2015

Fargo, North Dakota

North Dakota State University
Graduate School

Title

Demographic Analysis of Student Evaluations

By

Lucas James Huebner

The Supervisory Committee certifies that this ***disquisition*** complies with North Dakota State University's regulations and meets the accepted standards for the degree of

MASTER OF SCIENCE

SUPERVISORY COMMITTEE:

Dr. Rhonda Magel
Chair

Dr. William Martin

Dr. Yarong Yang

Curt Doetkott

Approved:

5/13/2015
Date

Dr. Rhonda Magel
Department Chair

ABSTRACT

Data was collected from North Dakota State University's student rating of instructor's forms during the fall of 2013 and the spring of 2014. This thesis investigates differences between male and female instructor's ratings, as well as attempts to describe outcomes using other demographics. T-tests were performed comparing the means of class averages for male and female instructors for each question on the student evaluation. There was not a difference for the mean class averages between male and female instructors when the whole university was considered and when only looking at the College of Science and Math. The analysis conducted also shows that male students tend to rate male instructors higher and female students tend to rate female instructors higher.

ACKNOWLEDGEMENTS

I would like to thank my advisor, Dr. Rhonda Magel, for the time and effort she put in during process of writing and reviewing this thesis, as well getting me through this program. I appreciate all of her advice and motivation.

A special thanks goes out to Curt Doetkott, for his exemplary patience while I struggled to modify his SAS code to create raw data sets. Also, for his assistance to help me learn more about how to analyze data through SAS which has helped me grow and will be instrumental to my future.

Lastly, I'd like to thank all members of my committee, Dr. Rhonda Magel, Dr. Yarong Yang, Dr. William Martin, and Curt Doetkott for their time and input that they put into my thesis.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Figure Page

LIST OF APPENDIX TABLES

LIST OF APPENDIX FIGURES

1. INTRODUCTION

Measuring the effectiveness of a professor's ability to communicate and teach a class is a challenging task. The most commonly used way of evaluating teaching at North Dakota State University is by using student ratings of instruction. One benefit of the student rating of instruction form is the ease of data collection. Towards the end of each semester, every professor is required to allow students to complete the student ratings of instruction form about the class. The students are not required to do this, but almost all do. Another benefit of using the student rating of instruction form is how much data gets collected over a short period of time. After just a couple of semesters, there will be tens of thousands of observations. This data has become more and more important too as personnel decisions are starting to be more dependent upon them.

Student response data for the academic year from 2013-2014 was collected, including 2092 classes and 18,371 student responses. The student rating of instruction form consists of 16 questions about a professor's ability to teach a course. Each question has a minimum rating of 1 and a maximum rating of 5. Prior to 2013, the form consisted of six questions. These questions are the first six questions on the form given in Appendix D. It was felt by some faculty on this campus that these six questions were not "valid" questions for use in evaluating student rating of instruction. Some faculty felt the questions might actually have gender bias. As a result, a new set of 10 questions was proposed. Since the original set of six questions had been used for almost 20 years, many of the faculty on campus were reluctant to just change over to the new set of questions. A compromise was reached and it was decided that the new set of questions would be added to the old set and all 16 questions would be put on the form for students to answer.

A faculty member would have the choice of using student responses to the first 6 questions or student responses to the last 10 questions in their review for tenure and promotion. Demographic questions for students were also added to the evaluation form. These questions consisted of the gender of the student, the expected grade they will receive at the end of the course, what year they are in school, and whether the course is required or an elective.

There are four main areas of research pertaining to this student rating of instruction form (SROI) that we would like to address. The first thing that we would like to investigate is the relationship of the questions to each other. In particular, we would like to know if any two of the set of new 10 questions are highly correlated to each other. If high correlation exists, the new set of 10 questions could be reduced. We would also like to investigate the relationship between the old set of questions and the new set of questions.

In the next phase of research we will assess how the average student response of a class is related to the demographics within their class. The class demographic variables considered will be the percent of students in the class that are required to take the course for their major of study, percent of males, percent of freshman and sophomores, percent of students that expect to receive either an A or a B in the course, and finally the gender of the instructor.

The third area of research is to investigate whether the class average responses for each question are different for male and female instructors. Question 12, which asks the student to rate the instructors availability will be of particular interest. Bennett (1982) surveyed undergraduate students and asked them how much personal attention they

received from their professors. The students reported getting more personal attention from their female professors, yet rated them lower as far as availability outside of class. We will investigate whether the mean class response differs between male and female instructors for this question at NDSU.

The last phase of our research is investigate whether or not there is a bias between a student's gender and the professor's gender in student responses. Bachen, McLoughlin, and Garcia (1999) found that there wasn't a difference in overall ratings, but when asked to respond to a qualitative question about their instructors, the male students were particularly harsh on female instructors. One half of the male students said negative things about their female instructors. Sprague and Massoni (2005) also found that students were more likely to say their favorite teacher ever was the same gender as them. We will look at each question individually, 1 through 16, and look only at classes with at least 5 males and 5 females in them who actually responded to the gender question. If there is no bias, the proportion of classes taught by a female instructor in which the average male student response is higher should be 50 percent. We will compare that proportion with the proportion of classes taught by male instructors in which the male student response is higher. If a bias does exist, we will investigate how the bias in the first set of six questions compares to the bias in the new set of 10 questions.

2. LITERATURE REVIEW

A variety of different kinds of research has been done regarding student ratings of instructors. Many of these studies take a deeper look into the gender expectations between male and female instructors. This is important because the emphasis of the use of student evaluations has increased at universities. Many universities look at student evaluations to make promotion decisions. A nationwide survey was done at liberal arts colleges that showed in 1978 just 54.8% of administrators used student evaluations. This widely increased as the same poll was done in 1998 and 88.1% of universities used some form of student ratings of teaching (Seldin 1999).

Sprague and Massoni (2005) note that there are specific gender expectations for a college professor. A male who is more "masculine" and a female who is more "feminine" will receive higher scores. However, it seems that for a female professor to live up to her "feminine" expectations it is more difficult than for a man to be "masculine". This is largely due in part to the amount of time needed for a female to give students individual attention and can be hampered by a large class size. A male is not typically expected to be as readily available. This difference has been shown in many universities from a survey question regarding instructor availability.

According to Read and Raghunandan (2001), student rating of instructors are not a good measurement of a students learning or the instructor's performance. Instead, the instructor's gender, age, race, and national origin play a larger influence in their evaluations. Along with specific demographics of the instructor, Read and Raghunandan (2001) also suggest the students happiness at the end of the semester (i.e. going to receive an "A" rather than a "C") will give more of a reflection on evaluations than how the

instructor actually performed. More interest has been placed in student evaluations because historically universities have placed a larger emphasis on the researching part of a professor's job when it comes to promotion. Now many universities are starting to stress the importance of effective teaching, so a properly designed student evaluation can serve as an invaluable instrument for assessing a professor's true performance in the classroom.

Arbuckle and Williams (2003) did a study trying to see if there really is a gender perception. They ran a study in a psychology class where students were divided into 4 groups and listened to a 35-minute presentation. The presentation was recorded by a 45 year old woman and students were asked to guess her gender based on the voice. The response was close to 50-50 whether or not the voice was male or female. So of the 4 groups, one was told the voice was a male under the age of 35, another a male over 55, and the other two groups were told it was a female under 35 and a female over 55. The results showed that the young male was rated in the highest in every category. Especially in the categories of enthusiasm, interest, confidence, and voice tone.

There are, however, some sources that argue that gender is not indeed a factor at all. According Cashin's research, (Cashin, 1999) there is not a gender bias, and if anything women are rated higher than men. Feldman's research agrees that women are rated higher than men, but the difference was very trivial (Feldman 1992, 1993). Aleamoni (1999) also reported that in a majority of studies that there is no difference between the genders of the instructors' ratings.

Christine Bachen, Moira McLoughlin, and Sara Garcia (1999) did a survey and discovered that on a quantitative scale gender did not necessarily make a difference.

However, when asked to answer how they felt about their instructor in qualitative terms, female instructors were criticized for not being as challenging and as professional as their male counterparts. Some universities (notably the University of Mississippi and the University of Colorado) are starting to make up for this disparity by throwing out any evaluations that mention anything to do with gender and are unrelated to teaching (Laube, Massoni, Sprague, Ferber, 2007).

3. METHODOLOGY

3.1. Correlations

The beginning of our research starts by looking at the questions 7-16. These 10 questions are the new set. We will calculate the Pearson's correlation coefficient between each question, to see if any two questions are correlated. If a question is highly correlated with another question, it makes that part of the survey redundant because you are getting similar information.

3.2. New questions vs. old questions

Next we will look at how the old questions correspond with the new questions. Pearson correlation coefficients for questions 1 through 6 against questions 7 through 16 are calculated. Also, for each of the first 6 questions, we will perform a stepwise regression to see if any of the class responses to the first 6 questions can be explained by a combination of the class responses to the last 10 questions. There will be 6 regressions in all. The dependent variable for the first regression will be the average class response for question 1. Ten independent variables will be considered for entry into the model. These variables will be the average class responses for questions 7 through 16. For the other 5 regressions, the independent variables will be the same, and the dependent variables will be the average class response for questions 2, 3, 4, 5, and 6, respectively.

Stepwise regression will be used to determine which of the independent variables to put in a model. The first step will start with no independent variables in the model. Then SAS will check each independent variable and choose the lowest p-value of all of the independent variables and put that variable in the model. SAS will then check the model to be sure that variable has a p-value of less than .15. Once that has occurred SAS

will then assess the rest of the variables and take the independent variable with the next lowest p-value and check to make sure it's p-value is less than 0.15 as well. This process continues until all significant variables are acquired and the model is complete.

The next step of the stepwise regression is to examine our model. The first thing we'll do is to look at the global or overall F-test. The F-test looks at the entire model put together and tells us how good of a job our model is doing at explaining the dependent variable. We will look at the p-value to see if it is significant. If it is significant, we'll want to look at the R squared value and see what percentage of variation in values of the dependent variable our model is explaining. We can also interpret our beta coefficients. An example of this would be if our dependent variable was the average class response to question 1, and the independent variable was the average class response to question 7. With the corresponding beta parameter estimate equal to 0.5, then if there is an increase in the average response to question 7 by 1, the model would predict an increase in the average class response to question 1 by 0.5.

Now that we have interpreted our model, we need to check the assumptions. A regression model has four assumptions: random errors are independent, error terms are approximately normal, constant variance in the dependent variable for each setting of the independent variables, and error terms should have a mean of zero. In this case, each class is different and one should be able to assume that class responses are independent of one another. To check the normality assumption of the error terms, we use a quantile-quantile plot to check the distribution of the residuals. We can also conduct a Shapiro-Wilk test, however, the p-value of the Shapiro-Wilk can be sensitive with very large sample sizes so looking at the quantile-quantile plot is more effective. To check the

constant variance assumption, we plot the residuals against the predicted values. If the bandwidth is about the same, we can assume the variance is approximately constant. The residuals should also be above and below zero with an average of approximately zero.

3.3. Instructor gender

In the next part of our research we will compare class responses between male and female instructors. Using a t-test, we will test whether the means of the average class responses to a given question are the same for male and female instructors, or if they are significantly different. Sixteen t-tests will be conducted: one for each question.

3.4. Demographics

We will now conduct several regressions to determine if there is a relationship between the class response to each question and certain demographics. For the first regression model, the dependent variable is the average class response to question 1. The independent demographic variables considered are the percent of students in the class where the class is a required course for their major, the proportion of males in the class, the proportion of students that are freshman or sophomores, the proportion of students expecting to receive either an A or a B in the course, and the gender of the instructor as an indicator variable. The same step-by-step process will be used to analyze this linear regression model as was used before. Each of the beta values will be investigated to see how the variable affects evaluation scores, negatively or positively. This process will be repeated using each question as the dependent variable, so there will be 16 models in total.

3.5. Student-instructor gender interaction

In the last part of the research, we want to see if there is a gendered interaction between the gender of the instructor and the gender of the student. Only classes with at least 5 male students and 5 female students responding to the gender question on the SROI form will be considered. A Z-test to test for differences in proportions will be performed. Based on student responses to each question on the SROI form, proportion 1 is the proportion of classes taught by female instructors in which the average male student response was higher than the average female student response. For each question, proportion 2 is the proportion of classes taught by male instructors in which the average male student response was higher than the average female student response. The null hypothesis is that the two proportions are equal and the alternative hypothesis is that they are different. This test will be conducted for each of the 16 questions.

4. RESULTS

4.1. Correlations

Pearson's correlation coefficients were found for each combination of questions 7 through 16. The correlation matrix is shown in Table 4.1.

Table 4.1. Correlation matrix for new set of questions

	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	Q15	Q16
Q7	1	.6945	.6898	.5532	.5312	.5786	.5921	.7017	.6414	.7420
Q8	.6945	1	.7506	.6100	.5367	.5527	.5821	.6601	.6468	.6156
Q9	.6898	.7506	1	.5934	.5367	.5364	.5774	.6631	.6075	.6184
Q10	.5532	.6100	.5934	1	.5461	..4933	.5461	.5749	.5127	.5159
Q11	.5312	.5367	.5367	.5461	1	.4590	.4713	.5389	.4925	.5175
Q12	.5786	.5527	.5364	.4933	.4590	1	.6595	.6663	.5750	.5442
Q13	.5921	.5821	.5774	.5461	.4713	.6595	1	.7443	.5798	.5555
Q14	.7017	.6610	.6631	.5749	.5389	.6663	.7443	1	.6778	.6502
Q15	.6414	.6468	.6075	.5127	.4943	.5750	.5798	.6778	1	.6475
Q16	.7420	.6156	.6184	.5159	.5175	.5442	.5555	.6502	.6475	1

Every single one of these relationships is also statistically significant. That does not come as a surprise as the sample size is over 18,000 for each observation causing the test to be sensitive to any relationship. All of the relationships are positive, meaning as one question on the survey gets rated higher, another question will also get rated higher. This makes sense because usually a "good" instructor will be strong in many areas and will receive a higher overall score. Most of the questions have a strong positive correlation, with just a few having a very strong relationship. Questions 7 and 16, 7 and

14, 13 and 14, and 8 and 9 have relatively strong relationships with correlation coefficients above .7. Question 7 and 16 have a correlation of .7420, meaning responses to question 7 explain 55% of the variation in responses to question 16.

Looking at how the old questions (1 through 6) correlate with the new questions (7 through 16), another correlation matrix was created in Table 4.2.

Table 4.2. Correlations between old questions and new questions

	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	Q15	Q16
Q1	.7543	.6856	.7177	.5505	.5424	.5488	.5815	.6963	.6525	.6445
Q2	.7504	.6690	.7001	.5251	.4841	.5515	.5724	.6927	.6193	.6293
Q3	.7137	.6687	.7013	.5284	.4824	.5393	.5707	.6672	.5891	.6080
Q4	.6693	.6404	.6612	.5322	.5401	.5112	.5379	.6384	.6156	.6141
Q5	.5603	.5550	.5611	.6483	.5274	.4633	.5084	.5679	.4761	.5101
Q6	.5650	.5502	.5706	.4887	.6872	.4516	.5449	.5509	.4949	.5249

Again, all relationships are positive correlations. The only questions with correlations of .7 or higher are questions 1 and 7, 1 and 9, 2 and 7, 2 and 9, 3 and 7, and 3 and 9. Question 7 and question 9 have correlations of greater than .7 with 3 of the old questions.

4.2. Regression on old questions

Six ordinary least squares stepwise regressions were conducted with the response variables being the average class response to questions 1, 2, 3, 4, 5, and 6, respectively. There are 10 independent variables considered for entry in each model, the average class

responses to questions 7 through 16. The model results for question 1 are given in Tables 4.3-4.5.

Table 4.3. Regression results for question 1

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	627.8152235	104.6358706	1263.96	<.0001
Error	2086	172.6879301	0.0827842		
Corrected Total	2092	800.5031536			

Table 4.4. Coefficient results for question 1

R-Square	Coeff Var	Root MSE	average1 Mean
0.784276	6.733355	0.287723	4.273093

Table 4.5. Parameter results for question 1

Source	DF	Type III SS	Mean Square	F Value	Pr > F
average7	1	28.30149501	28.30149501	341.87	<.0001
average9	1	20.31199734	20.31199734	245.36	<.0001
average14	1	11.68379635	11.68379635	141.14	<.0001
average15	1	2.60588316	2.60588316	31.48	<.0001
average16	1	1.14373756	1.14373756	13.82	0.0002
average11	1	0.48145968	0.48145968	5.82	0.0160

Table 4.3 shows the overall model to be significant with a p-value of less than .0001. The R squared value is 0.784276, meaning that approximately 78.43% of the variation in class responses to question 1 is described by class responses to questions 7, 9, 11, 14, 15, and 16. The assumptions are checked by evaluating the residual plots found in Figure 4.1.

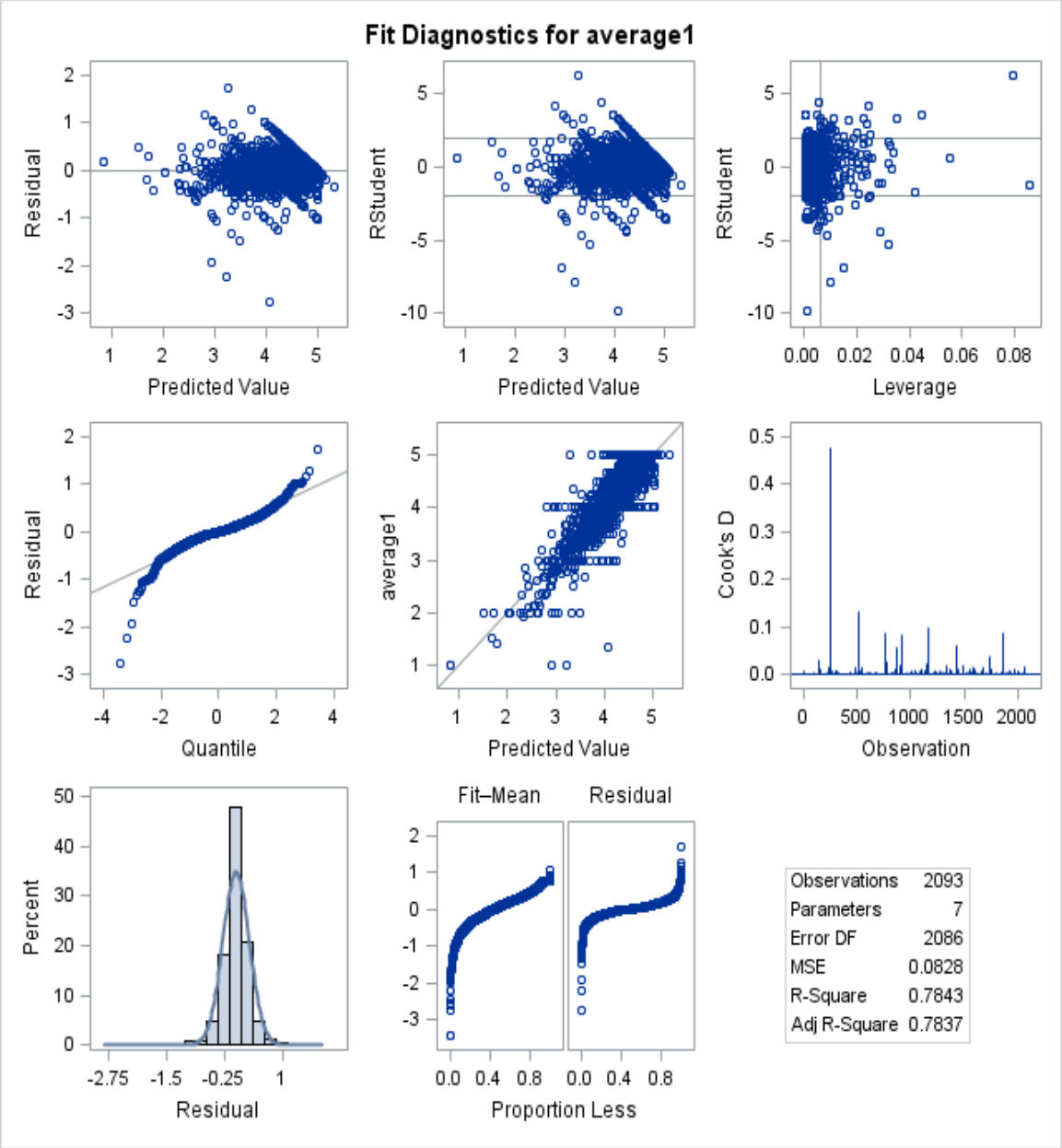


Figure 4.1. Residual plots for model 1

From the graph in row 1 column 1, we can see that the residuals are evenly distributed above and below 0, and are randomly distributed with no real pattern. In the graph in row 2 column 1, we can see the distribution is approximately normal, with a little discrepancy on the tails, but least squares regression is very robust when it comes to the normality assumption. The shape of the distribution is also seen in the histogram in

row 3 column 1. The histogram has a bell shape, with the majority of the data in the center. Finally the linear relationship is shown in row 2 column 2. It is plainly seen that there is a positive linear correlation between the class average for question 1 and the predicted values. We do not want to interpret any of the beta parameters because there will be problems with multicollinearity. The independent variables are all too closely related that the beta values will not have regular interpretations.

A stepwise regression was then performed with the class average of question 2 used as the dependent variable. The model results for question 2 are given in the Tables 4.6-4.8.

Table 4.6. Regression results for question 2

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	9	609.9280629	67.7697848	745.78	<.0001
Error	2080	189.0124338	0.0908714		
Corrected Total	2089	798.9404967			

Table 4.7. Coefficient results for question 2

R-Square	Coeff Var	Root MSE	average2 Mean
0.763421	6.949990	0.301449	4.337399

Table 4.8. Parameter results for question 2

Source	DF	Type III SS	Mean Square	F Value	Pr > F
average7	1	30.90296442	30.90296442	340.07	<.0001
average8	1	0.46031292	0.46031292	5.07	0.0245
average9	1	12.14224759	12.14224759	133.62	<.0001
average11	1	0.91293786	0.91293786	10.05	0.0015
average12	1	0.35479432	0.35479432	3.90	0.0483
average13	1	0.75975579	0.75975579	8.36	0.0039
average14	1	9.03072284	9.03072284	99.38	<.0001
average15	1	3.09219197	3.09219197	34.03	<.0001
average16	1	0.69726769	0.69726769	7.67	0.0057

Again, this model is significant for predicting the class average of question 2. Nine of the independent variables are significant. The only variable to get dropped from the stepwise regression was the class average for question 10. The R squared value was .7634, so our model describes 76.34% of the variation of class average responses to question 2. Residual plots were again conducted and it appears that the model assumptions were met.

Next, a stepwise regression model for the class average of question 3 is found and the results are given in Tables 4.9-4.11.

Table 4.9. Regression results for question 3

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	654.4771989	109.0795331	864.14	<.0001
Error	2085	263.1881576	0.1262293		
Corrected Total	2091	917.6653564			

Table 4.10. Coefficient results for question 3

R-Square	Coeff Var	Root MSE	average3 Mean
0.713198	8.327812	0.355288	4.266279

Table 4.11. Parameter results for question 3

Source	DF	Type III SS	Mean Square	F Value	Pr > F
average7	1	21.13088379	21.13088379	167.40	<.0001
average8	1	0.75381640	0.75381640	5.97	0.0146
average9	1	24.49786992	24.49786992	194.07	<.0001
average13	1	0.48795612	0.48795612	3.87	0.0494
average14	1	2.41117372	2.41117372	19.10	<.0001
average15	1	4.25695979	4.25695979	33.72	<.0001

The stepwise model for question 3 contains 6 independent variables, the class averages of question 7, 8, 9, 13, 14, and 15. The overall model was significant with an R squared of .7132. Residual plots indicate the assumptions appear to be valid.

A stepwise regression model was found with a dependent variable of the class average of question 4. Results are given in Tables 4.12-4.14.

Table 4.12. Regression results for question 4

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	9	507.9904219	56.4433802	539.49	<.0001
Error	2081	217.7210750	0.1046233		
Corrected Total	2090	725.7114969			

Table 4.13. Coefficient results for question 4

R-Square	Coeff Var	Root MSE	average4 Mean
0.699989	7.686012	0.323455	4.208362

Table 4.14. Parameter results for question 4

Source	DF	Type III SS	Mean Square	F Value	Pr > F
average7	1	2.52340764	2.52340764	24.12	<.0001
average8	1	5.55088875	5.55088875	53.06	<.0001
average9	1	6.54602903	6.54602903	62.57	<.0001
average10	1	0.31655098	0.31655098	3.03	0.0821
average11	1	4.40229521	4.40229521	42.08	<.0001
average13	1	0.35998289	0.35998289	3.44	0.0637
average14	1	3.52788865	3.52788865	33.72	<.0001
average15	1	5.30976318	5.30976318	50.75	<.0001
average16	1	1.34070668	1.34070668	12.81	0.0004

There are 9 independent variables in the model for question 4. The only variable that was not included was the class average of question 12. The overall model had an R squared value of .6999. Residual plots indicate the assumptions appear to be valid.

Another stepwise regression was conducted with the class average of question 5 as the dependent variable. Results are given in Tables 4.15-4.17.

Table 4.15. Regression results for question 5

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	426.9673657	71.1612276	565.20	<.0001
Error	2085	262.5111234	0.1259046		
Corrected Total	2091	689.4784891			

Table 4.16. Coefficient results for question 5

R-Square	Coeff Var	Root MSE	average5 Mean
0.619261	8.169075	0.354830	4.343581

Table 4.17. Parameter results for question 5

Source	DF	Type III SS	Mean Square	F Value	Pr > F
average7	1	3.88461604	3.88461604	30.85	<.0001
average10	1	58.26012300	58.26012300	462.73	<.0001
average11	1	9.38714287	9.38714287	74.56	<.0001
average12	1	0.62066981	0.62066981	4.93	0.0265
average14	1	12.53030455	12.53030455	99.52	<.0001
average15	1	0.32192225	0.32192225	2.56	0.1100

Six independent variables were selected for the question 5 model. The independent variables were the class averages for question 7, 10, 11, 12, 14 and 15. The overall model has an R squared of .6192. The assumptions were reasonably met.

The last stepwise regression was performed trying to predict the class average for question 6. Results are given in Tables 4.18-4.20.

Table 4.18. Regression results for question 6

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	7	438.2116267	62.6016610	607.88	<.0001
Error	2084	214.6187298	0.1029840		
Corrected Total	2091	652.8303565			

Table 4.19. Coefficient results for question 6

R-Square	Coeff Var	Root MSE	average6 Mean
0.671249	7.649557	0.320911	4.195161

Table 4.20. Parameter results for question 6

Source	DF	Type III SS	Mean Square	F Value	Pr > F
average7	1	2.32829248	2.32829248	22.61	<.0001
average8	1	2.39154283	2.39154283	23.22	<.0001
average9	1	3.11373803	3.11373803	30.24	<.0001
average10	1	0.43162296	0.43162296	4.19	0.0408
average11	1	79.51635938	79.51635938	772.12	<.0001
average13	1	1.68276998	1.68276998	16.34	<.0001
average14	1	6.29833899	6.29833899	61.16	<.0001

Seven independent variables were selected for the question 6 model. The R squared value is .6712 and the model is significant. Residual plots indicate the assumptions appear to be valid.

4.3. Comparing averages for male/female instructors

T-tests were conducted to determine if there is a difference between the average class responses to each question in the survey for male and female instructors. First, all colleges at the university were used. The results are shown in Table 4.21. A more detailed list of tables for each of the 16 questions can be found in Appendix A.

Table 4.21. Instructor gender comparison for all colleges

Question	Female Instructor Sample Mean	Male Instructor Sample Mean	Difference (Female - Male)	p-value
1	4.2703	4.2758	-0.00552	0.8404
2	4.3301	4.3420	-0.0119	0.6628
3	4.2936	4.2472	0.0464	0.1108
4	4.1927	4.2207	-0.0280	0.2832
5	4.3222	4.3617	-0.0395	0.1242
6	4.2371	4.1667	0.0704	0.0041 **
7	4.3359	4.3250	0.0109	0.6593
8	4.2859	4.2671	0.0188	0.4520
9	4.2319	4.2166	0.0153	0.5723
10	4.3324	4.3441	-0.0117	0.6359
11	4.2201	4.1705	0.0496	0.0327 **
12	4.3247	4.3149	0.00979	0.6848
13	4.2960	4.2896	0.00644	0.8016
14	4.2817	4.2610	0.0207	0.4131
15	4.3334	4.2909	0.0425	0.0548 *
16	4.2866	4.2998	-0.0133	0.5819

** Significant at 0.05

* Significant at 0.1

For question 1, males had a higher sample mean by a narrow margin of 0.00552. The variances were deemed equal and the p-value is 0.8404. The decision is to fail to

reject the null hypothesis of equal means, there is not enough evidence to conclude that these means are different.

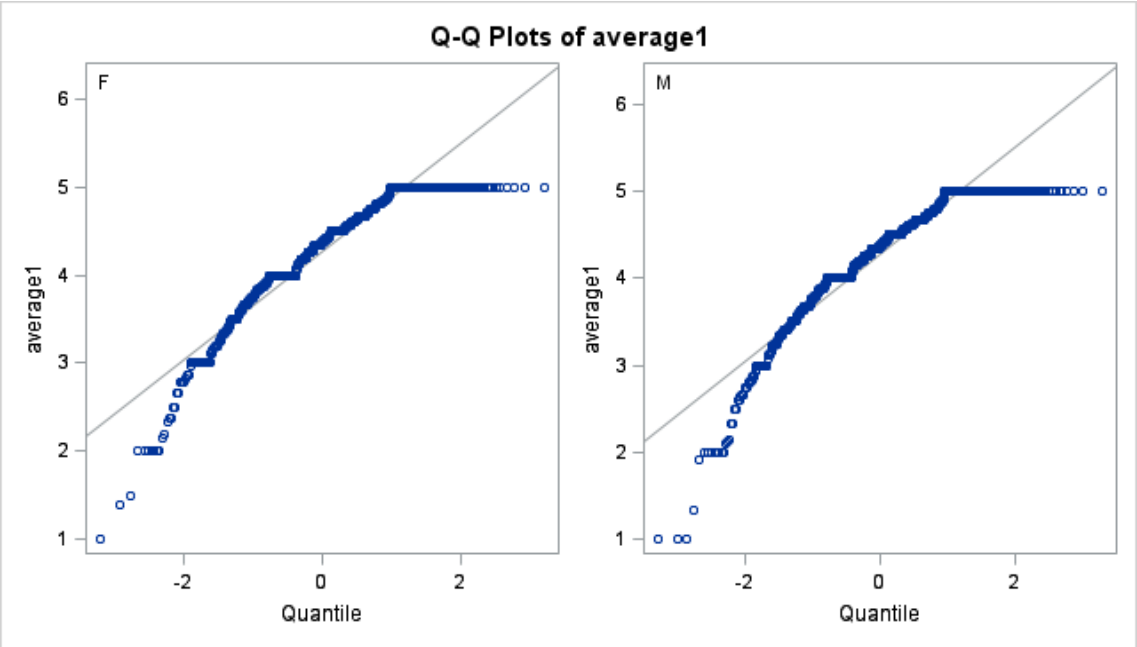


Figure 4.2. Q-Q plots for question 1

Looking at the Q-Q plot, the data is discrete on a Likert scale from 1 to 5, which is why there at the top right the plot flattens out because the data has a maximum of 5. The Q-Q plots are okay, and along with using the central limit theorem, because we have over 887 classes taught by female instructors and 1199 classes taught by male instructors, the assumption of normality is reasonably met. Female instructors had a significantly higher average class response on questions 6, 11, and marginally significantly higher on question 15.

It is known that humanities and social science courses are rated higher than science and math classes. Because of this, we decided to compare the average class response for classes taught by female instructors with the average class responses taught

by male instructors for each question based on classes in the College of Science and Mathematics. The results were similar to when we included all colleges. The male instructors had significantly higher ratings on question 5. The female instructors had significantly higher ratings on questions 3, 7, and marginally significantly higher ratings on question 2. The results are given in Table 4.22. For a more detailed list of tables for each of the 16 questions in the College of Science and Mathematics, please refer to Appendix B.

Table 4.22. Instructor gender comparison for College of Science and Math

Question	Female Instructor Sample Mean	Male Instructor Sample Mean	Difference (Female–Male)	p-value
1	4.2393	4.1506	0.0887	0.1004
2	4.2855	4.1888	0.0967	0.0824 *
3	4.2515	4.1098	0.1417	0.0166 **
4	4.0381	4.0758	-0.0377	0.4725
5	4.1894	4.3100	-0.1206	0.0317 **
6	4.0063	3.9361	0.0702	0.1805
7	4.3116	4.2158	0.0958	0.0486 **
8	4.2013	4.1366	0.0647	0.1874
9	4.1465	4.0847	0.0618	0.2363
10	4.2926	4.3426	-0.0500	0.2587
11	4.0409	4.0269	0.0141	0.7816
12	4.2335	4.2077	0.0259	0.5860
13	4.2718	4.2398	0.0320	0.4847
14	4.2060	4.1444	0.0615	0.2059
15	4.1940	4.1558	0.0382	0.3757
16	4.2277	4.1930	0.0347	0.4319

** Significant at 0.05

* Significant at 0.1

4.4. Regression using demographics

Sixteen linear regressions were performed to determine if there is any relationship between the class average responses to each of the questions the class's demographics. The class average for a question was the response variable. The results for question 1 are shown in Tables 4.23-4.25, the full list of results can be found in Appendix C.

Table 4.23. Demographic regression results for question 1

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	9.28154	1.85631	10.21	<.0001
Error	140	25.45154	0.1818		
Corrected Total	145	34.73308			

Table 4.24. Demographic parameter results for question 1

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	**1**	2.82253	0.33885	8.33	<.0001
propReq	**1**	-0.14268	0.15156	-0.94	0.3481
percMales	**1**	0.28837	0.23881	1.21	0.2293
propFresh	**1**	-0.14175	0.19144	-0.74	0.4603
propAB	**1**	1.64354	0.26885	6.11	<.0001
instructorGender	**1**	0.02499	0.07514	0.33	0.7400

Table 4.25. Demographic coefficient results for question 1

Root MSE	0.42638	**R-Square**	0.2672
Dependent Mean	4.18935	**Adj R-Sq**	0.2411
Coeff Var	10.1776		

All of our models were statistically significant. The R squared values in general were not very high, they ranged from around 0.2 to 0.3 for the majority. Question 11 had

the highest R squared with a value of 0.5883. Since about 59% of the variation in class average responses for question 11 can be explained by the class demographics alone that do not involve the quality of instruction, perhaps eliminating this question should be considered. Questions 5 and 6 also had relatively high R squared values of .3935 and .4904, respectively. This indicates that approximately 40% and 50%, respectively of the variation in student responses is explained by the collected class demographics and not due to differences in instruction ability. The variable that is significant in all the regressions was the proportion of students expecting to receive an A or B for the class (p-value <.0001). The rating of the instructor for each question increased as the proportion increased. The percentage of males in the class was significant for question 7 (p-value = .0354), marginally significant for question 8 (p-value = .0994), significant for question 12 (p-value = .0462), significant for question 15 (p-value = .0046), and significant for question 16 (p-value = .0279). In all the questions for which the percentage of males was significant, it was associated with a positive coefficient indicating that for these questions as the percentage of males in the class increased, the rating of the instructor by the class increased. If the percentage of males in the class increased by ten percent, this generally increased the instructor's rating for these questions by .04. The proportion of students taking the class because it was required for their major was significant for question 6 (p-value = .031). This variable was associated with a negative coefficient. If the percentage of students taking the class because it was required increased by ten percent, the instructor rating for this question decreased by approximately .03. No other demographic variable was significant with all the demographic variables in the model.

4.5. Proportions

The proportion of classes taught by female instructors in which the average male student response is higher than the average female student response (proportion 1) is compared to the proportion of classes taught by male instructors in which the average male student response is higher than the average female student response (proportion 2) for each question. This is done for classes in which it is indicated that at least five male students and five female students responded to questions about the instruction for that class. The sample sizes were 112 classes taught by female instructors and 162 classes taught by male instructors. The results for the samples are given in Table 4.26.

Table 4.26. Proportion of classes in which male student response is higher

Question	Proportion 1	Proportion 2	Test Statistic	P-value
1	0.4375	0.5432	-1.73	0.08364 *
2	0.4821	0.5432	-1.25	0.21130
3	0.5089	0.5926	-1.37	0.17068
4	0.3750	0.5432	-2.79	0.00528 **
5	0.4286	0.5617	-2.19	0.02852 **
6	0.4821	0.6049	-2.02	0.04338 **
7	0.4123	0.5185	-1.75	0.08012 *
8	0.3947	0.5000	-1.74	0.08186 *
9	0.4474	0.5802	-2.19	0.02852 **
10	0.4561	0.5370	-1.27	0.20408
11	0.4298	0.5864	-2.59	0.00960 **
12	0.4123	0.5000	-1.45	0.14706
13	0.4649	0.4877	-0.37	0.44130
14	0.4825	0.5864	-1.71	0.08726 *
15	0.4561	0.5123	-0.87	0.38430
16	0.4298	0.5370	-1.77	0.07672

** Significant at 0.05

* Significant at 0.10

Sample proportion 1 is lower than sample proportion two for every single question. It is also significantly lower on questions 4, 5, 6, 9, and 11 and marginally significantly lower on questions 1, 7, 8, and 14. Theoretically, if the gender of the student and the gender of the instructor do not matter proportion 1 and proportion 2 should be .5. Sample proportion 1 is only greater than .5 one time on question 3 and that is at a narrow margin with a sample proportion of .5089. Sample proportion 2 is less than .5 just once and equal to .5 twice. From these results, one can conclude male students tend to rate male instructors higher and female students tend to rate female instructors higher.

5. CONCLUSION AND DISCUSSIONS

Our first research focus was investigating the relationship of the new set of questions to each other. This is done to see if the number of questions in the new set can be reduced. A correlation matrix was formed for the new questions. Questions 7 and 16 have a correlation coefficient of .7420. The correlation coefficient between question 7 and question 14 is also above .7. We would recommend dropping question 7 from the SROI form. Question 8 and Question 9 also have a high correlation of .7506. Neither of these questions have a high correlation with any other question, but the nature of both of the questions is quite similar, so it is also recommended that one of these questions be dropped. Questions 13 and 14 also have a high correlation of .7443. It seems as though these two questions could be combined together given that they both ask about feedback.

The relationships between the old questions and the new questions were investigated using regression models. The R squared values for the models were all around .7 to .8, meaning that about 70% to 80% of the variation in class responses to each of the old questions can be explained by some combination of class responses to the new questions. A correlation matrix was also calculated between the new set of questions and the old set of questions. Questions 7 and 9 had correlations of over .7 with 3 of the old questions. No old question had a correlation of more than .75 with one of the new questions.

In the second phase of our research we investigated how class demographics were related to class average responses for each question. Regression models were developed with class average responses to each question as the dependent variables. It was found that 59% of the variation in class responses for question 11 could be explained by the

class demographics. It is recommended that this question be dropped since less than 50% of the variation of class responses could be explained by differences in "instruction" ability. The demographic variables also described 32%, 39%, and 49% of the variation in class responses to questions 4, 5, and 6 respectively. Considering the amount of variation described by demographics for questions 4, 5, and 6 we recommend using the new set of questions. At least two of the six old questions are evaluating something besides the quality of the instruction, and with question 11 dropped, the new set of questions will be more suitable and less biased. For all other questions, the class demographics explained only 20-30% of the variation in class average responses. It is hoped that most of the variation in class average responses to these questions is because of differences in "instruction" ability. The percentage of students in the class expecting an A or B was the one demographic variable significant for all questions.

We next investigated how the mean of the class responses to each question for female instructors compared to that of male instructors. When classes for the entire university were considered, the sample mean response for classes taught by female instructors was not that much different than the sample mean response for male instructors. When the means for male and female instructors were compared in the College of Science and Mathematics, the sample means were also not much different. There were only a few questions in which the means between genders were significant. For two questions, the means were significantly higher for female instructors and for one question, the mean was significantly higher for male instructors. This agrees with our literature review that there is not a significant quantitative difference between instructor gender.

In this last phase of our research, it was found that the sample proportion of classes taught by female instructors in which the average male student response was higher was lower than the sample proportion of classes taught by male instructors in which the average male student response was higher being significantly lower for five questions and marginally significantly lower for four other questions. Again, this indicates male students are generally rating male instructors higher and female students are generally rating female instructors higher.

One thing the literature told us was female instructors were consistently rated lower on their availability and that female instructors needed to put more time in to receive better ratings. Our research tended to disagree with this. Question 12 on the SROI form asks if the instructor was available to assist students outside of class. On question 12, female instructors received higher ratings in our research than male instructors.

We did consider the first six questions (old set) in terms of gender bias with the last ten questions (new set) on the SROI form. The sample mean class average response for female instructors and male instructors was not much different. The means for classes in the College of Science and Mathematics showed little differences in class means between male and female instructors. In the case of the proportions of classes taught by female instructors in which the male student response was larger, those were significantly lower for three of the first six questions (old) and significantly or marginally significantly lower in six of the last ten questions (new). We suggest using the new set of questions after dropping questions 7 and 11, combining questions 8 and 9, and combining questions 13 and 14. Over 70% of the variations in class responses to each of the

questions 1 through 3 are explained by class responses to the new questions 7-16. A large amount of the variations in class responses for questions 4-6 are explained by class demographics and not quality of instruction. The same is true for question 11 in the new set. Class responses to question 7 are highly correlated with class responses to other questions in the new set. Class responses to questions 8 and 9 are highly correlated and these two questions could be concluded into one question. Class responses to 13 and 14 are also highly correlated and these two questions could be combined into one question. With these suggestions the form could be reduced to 6 questions.

REFERENCES

Aleamoni, L. (1999). Student Rating Myths Versus Research Facts from 1924 to 1998. *Journal of Personnel Evaluation in Education* 13(2):153-66.

Arbuckle, J., and Williams, B.D. (2003). Students' perceptions of expressiveness: age and gender effects on teacher evaluations. *Sex Roles*, 49, 507-516.

Bachen, C., McLoughlin, M., and Garcia, S. (1999). Assessing the Role of Gender in College Students' Evaluations of Faculty. *Communication Education* 48:193-210

Bennett, S. K. (1982). Student perceptions of and expectations for male and female instructors: Evidence relating to the question of gender bias in teaching evaluation. *Journal of Educational Psychology*, 74, 170–179.

Cashin, W. (1999). Student Ratings of Teaching: Uses and Misuses. In Changing Practices in Evaluating Teaching; A Practical Guide to Improved Faculty Performance and Promotion/Tenure Decisions, ed. Peter Seldin, 25-44. Boston: Anker Publishing.

Feldman, K. (1993). College Students' Views of Male and Female College Teachers: Part II-Evidence from Students' Evaluations of their Classroom Teachers. *Research in Higher Education* 34:151-211.

_____. (1992). College Students' Views of Male and Female College Teachers: Part I-Evidence from the Social Laboratory and Experiments. *Research in Higher Education* 33:317-73.

Laube, H., Massoni, K., Sprague, J., and Ferber, A.L. (2007). The impact of gender on the evaluation of teaching: what we know and what we can do. *NWSA Journal*, 19, 87-104.

Read, W., Rama, D.V., and Raghunandan K. (2001). The relationship between student evaluations of teaching and faculty evaluations. *Journal of Education for Business*, 76, 189-192.

Seldin, P. (1999). Current practices-good and bad-nationally. In P. Seldin (Ed.), Changing practices in evaluating teaching: A practical guide to improved faculty performance and pro-motion/tenure decisions (pp. 1–24). Boston, MA: Anker

Sprague, J, Massoni, K. (2005). Student Evaluations and Gendered Expectations: What We Can't Count Can Hurt Us. *Sex Roles* 53, 779-93.

APPENDIX A. MEAN GENDER RESULTS FOR QUESTIONS 1-16 ALL COLLEGES

Table A1. Mean gender results question 1

sex	N	Mean	Std Dev	Std Err	Minimum	Maximum
F	887	4.2703	0.6180	0.0208	1.0000	5.0000
M	1199	4.2758	0.6191	0.0179	1.0000	5.0000
Diff (1-2)		-0.00552	0.6186	0.0274		

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	2084	-0.20	0.8405
Satterthwaite	Unequal	1910.9	-0.20	0.8404

Table A2. Mean gender results question 2

sex	N	Mean	Std Dev	Std Err	Minimum	Maximum
F	887	4.3301	0.6187	0.0208	1.0000	5.0000
M	1199	4.3420	0.6188	0.0179	1.0000	5.0000
Diff (1-2)		-0.0119	0.6188	0.0274		

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	2084	-0.44	0.6628
Satterthwaite	Unequal	1909.3	-0.44	0.6628

Table A3. Mean gender results question 3

sex	N	Mean	Std Dev	Std Err	Minimum	Maximum
F	887	4.2936	0.6469	0.0217	1.2500	5.0000
M	1199	4.2472	0.6710	0.0194	1.0000	5.0000
Diff (1-2)		0.0464	0.6609	0.0293		

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	2084	1.59	0.1127
Satterthwaite	Unequal	1946	1.60	0.1108

Table A4. Mean gender results question 4

sex	N	Mean	Std Dev	Std Err	Minimum	Maximum
F	887	4.1927	0.5889	0.0198	1.6000	5.0000
M	1199	4.2207	0.5878	0.0170	1.0000	5.0000
Diff (1-2)		-0.0280	0.5882	0.0261		

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	2084	-1.07	0.2831
Satterthwaite	Unequal	1907	-1.07	0.2832

Table A5. Mean gender results question 5

sex	N	Mean	Std Dev	Std Err	Minimum	Maximum
F	887	4.3222	0.5981	0.0201	1.0000	5.0000
M	1199	4.3617	0.5553	0.0160	1.0000	5.0000
Diff (1-2)		-0.0395	0.5739	0.0254		

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	2084	-1.56	0.1200
Satterthwaite	Unequal	1826.9	-1.54	0.1242

Table A6. Mean gender results question 6

sex	N	Mean	Std Dev	Std Err	Minimum	Maximum
F	887	4.2371	0.5477	0.0184	1.0000	5.0000
M	1199	4.1667	0.5614	0.0162	1.0000	5.0000
Diff (1-2)		0.0704	0.5556	0.0246		

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	2084	2.86	0.0043
Satterthwaite	Unequal	1934.4	2.87	0.0041

Table A7. Mean gender results question 7

sex	N	Mean	Std Dev	Std Err	Minimum	Maximum
F	887	4.3359	0.5629	0.0189	2.0000	5.0000
M	1199	4.3250	0.5547	0.0160	1.0000	5.0000
Diff (1-2)		0.0109	0.5582	0.0247		

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	2084	0.44	0.6586
Satterthwaite	Unequal	1893.5	0.44	0.6593

Table A8. Mean gender results question 8

sex	N	Mean	Std Dev	Std Err	Minimum	Maximum
F	887	4.2859	0.5621	0.0189	1.0000	5.0000
M	1199	4.2671	0.5698	0.0165	1.0000	5.0000
Diff (1-2)		0.0188	0.5665	0.0251		

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	2084	0.75	0.4529
Satterthwaite	Unequal	1923.1	0.75	0.4520

Table A9. Mean gender results question 9

sex	N	Mean	Std Dev	Std Err	Minimum	Maximum
F	887	4.2319	0.6026	0.0202	1.2500	5.0000
M	1199	4.2166	0.6196	0.0179	1.0000	5.0000
Diff (1-2)		0.0153	0.6124	0.0271		

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	2084	0.56	0.5739
Satterthwaite	Unequal	1937.3	0.56	0.5723

Table A10. Mean gender results question 10

sex	N	Mean	Std Dev	Std Err	Minimum	Maximum
F	887	4.3324	0.5664	0.0190	1.0000	5.0000
M	1199	4.3441	0.5415	0.0156	1.0000	5.0000
Diff (1-2)		-0.0117	0.5522	0.0245		

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	2084	-0.48	0.6336
Satterthwaite	Unequal	1860.1	-0.47	0.6359

Table A11. Mean gender results question 11

sex	N	Mean	Std Dev	Std Err	Minimum	Maximum
F	887	4.2201	0.5119	0.0172	1.0000	5.0000
M	1199	4.1705	0.5399	0.0156	1.0000	5.0000
Diff (1-2)		0.0496	0.5281	0.0234		

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	2084	2.12	0.0340
Satterthwaite	Unequal	1961.7	2.14	0.0327

Table A12. Mean gender results question 12

sex	N	Mean	Std Dev	Std Err	Minimum	Maximum
F	887	4.3247	0.5343	0.0179	1.7500	5.0000
M	1199	4.3149	0.5582	0.0161	1.0000	5.0000
Diff (1-2)		0.00979	0.5482	0.0243		

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	2084	0.40	0.6867
Satterthwaite	Unequal	1952.8	0.41	0.6848

Table A13. Mean gender results question 13

sex	N	Mean	Std Dev	Std Err	Minimum	Maximum
F	887	4.2960	0.5839	0.0196	1.2500	5.0000
M	1199	4.2896	0.5715	0.0165	1.0000	5.0000
Diff (1-2)		0.00644	0.5768	0.0255		

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	2084	0.25	0.8009
Satterthwaite	Unequal	1886.2	0.25	0.8016

Table A14. Mean gender results question 14

sex	N	Mean	Std Dev	Std Err	Minimum	Maximum
F	887	4.2817	0.5706	0.0192	1.6000	5.0000
M	1199	4.2610	0.5718	0.0165	1.0000	5.0000
Diff (1-2)		0.0207	0.5713	0.0253		

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	2084	0.82	0.4132
Satterthwaite	Unequal	1911.4	0.82	0.4131

Table A15. Mean gender results question 15

sex	N	Mean	Std Dev	Std Err	Minimum	Maximum
F	887	4.3334	0.4834	0.0162	2.0000	5.0000
M	1199	4.2909	0.5212	0.0151	1.0000	5.0000
Diff (1-2)		0.0425	0.5054	0.0224		

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	2084	1.90	0.0576
Satterthwaite	Unequal	1981.4	1.92	0.0548

Table A16. Mean gender results question 16

sex	N	Mean	Std Dev	Std Err	Minimum	Maximum
F	887	4.2866	0.5496	0.0185	1.0000	5.0000
M	1199	4.2998	0.5350	0.0155	1.0000	5.0000
Diff (1-2)		-0.0133	0.5413	0.0240		

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	2084	-0.55	0.5804
Satterthwaite	Unequal	1880.2	-0.55	0.5819

APPENDIX B. MEAN GENDER RESULTS FOR QUESTIONS 1-16 COLLEGE OF SCIENCE AND MATH

Table B1. Mean gender results question 1

sex	N	Mean	Std Dev	Std Err	Minimum	Maximum
F	177	4.2393	0.5589	0.0420	2.3717	5.0000
M	359	4.1506	0.6387	0.0337	1.0000	5.0000
Diff (1-2)		0.0887	0.6135	0.0563		

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	534	1.57	0.1161
Satterthwaite	Unequal	395.07	1.65	0.1004

Table B2. Mean gender results question 2

sex	N	Mean	Std Dev	Std Err	Minimum	Maximum
F	177	4.2855	0.5664	0.0426	2.4336	5.0000
M	359	4.1888	0.6756	0.0357	1.0000	5.0000
Diff (1-2)		0.0967	0.6417	0.0589		

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	534	1.64	0.1014
Satterthwaite	Unequal	410.29	1.74	0.0824

Table B3. Mean gender results question 3

sex	N	Mean	Std Dev	Std Err	Minimum	Maximum
F	177	4.2515	0.6058	0.0455	2.1947	5.0000
M	359	4.1098	0.7088	0.0374	1.0000	5.0000
Diff (1-2)		0.1417	0.6765	0.0621		

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	534	2.28	0.0230
Satterthwaite	Unequal	403.41	2.40	0.0166

Table B4. Mean gender results question 4

sex	N	Mean	Std Dev	Std Err	Minimum	Maximum
F	177	4.0381	0.5628	0.0423	2.3333	5.0000
M	359	4.0758	0.5859	0.0309	1.0000	5.0000
Diff (1-2)		-0.0377	0.5784	0.0531		

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	534	-0.71	0.4783
Satterthwaite	Unequal	363.33	-0.72	0.4725

Table B5. Mean gender results question 5

sex	N	Mean	Std Dev	Std Err	Minimum	Maximum
F	177	4.1894	0.6386	0.0480	2.0000	5.0000
M	359	4.3100	0.5427	0.0286	2.0000	5.0000
Diff (1-2)		-0.1206	0.5761	0.0529		

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	534	-2.28	0.0230
Satterthwaite	Unequal	304.65	-2.16	0.0317

Table B6. Mean gender results question 6

sex	N	Mean	Std Dev	Std Err	Minimum	Maximum
F	177	4.0063	0.5612	0.0422	2.0000	5.0000
M	359	3.9361	0.5868	0.0310	1.0000	5.0000
Diff (1-2)		0.0702	0.5785	0.0531		

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	534	1.32	0.1869
Satterthwaite	Unequal	364.76	1.34	0.1805

Table B7. Mean gender results question 7

sex	N	Mean	Std Dev	Std Err	Minimum	Maximum
F	177	4.3116	0.4841	0.0364	2.3333	5.0000
M	359	4.2158	0.6059	0.0320	1.0000	5.0000
Diff (1-2)		0.0958	0.5687	0.0522		

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	534	1.83	0.0672
Satterthwaite	Unequal	427.52	1.98	0.0486

Table B8. Mean gender results question 8

sex	N	Mean	Std Dev	Std Err	Minimum	Maximum
F	177	4.2013	0.4823	0.0362	3.0000	5.0000
M	359	4.1366	0.6250	0.0330	1.0000	5.0000
Diff (1-2)		0.0647	0.5819	0.0534		

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	534	1.21	0.2265
Satterthwaite	Unequal	439.9	1.32	0.1874

Table B9. Mean gender results question 9

sex	N	Mean	Std Dev	Std Err	Minimum	Maximum
F	177	4.1465	0.5244	0.0394	2.6991	5.0000
M	359	4.0847	0.6461	0.0341	1.0000	5.0000
Diff (1-2)		0.0618	0.6086	0.0559		

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	534	1.11	0.2694
Satterthwaite	Unequal	421.84	1.19	0.2363

Table B10. Mean gender results question 10

sex	N	Mean	Std Dev	Std Err	Minimum	Maximum
F	177	4.2926	0.4657	0.0350	2.6667	5.0000
M	359	4.3426	0.5114	0.0270	2.0000	5.0000
Diff (1-2)		-0.0500	0.4968	0.0456		

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	534	-1.10	0.2736
Satterthwaite	Unequal	381.25	-1.13	0.2587

Table B11. Mean gender results question 11

sex	N	Mean	Std Dev	Std Err	Minimum	Maximum
F	177	4.0409	0.5240	0.0394	2.6250	5.0000
M	359	4.0269	0.6053	0.0319	1.0000	5.0000
Diff (1-2)		0.0141	0.5798	0.0533		

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	534	0.26	0.7918
Satterthwaite	Unequal	398.9	0.28	0.7816

Table B12. Mean gender results question 12

sex	N	Mean	Std Dev	Std Err	Minimum	Maximum
F	177	4.2335	0.4775	0.0359	3.0000	5.0000
M	359	4.2077	0.5880	0.0310	1.0000	5.0000
Diff (1-2)		0.0259	0.5540	0.0509		

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	534	0.51	0.6115
Satterthwaite	Unequal	421.68	0.55	0.5860

Table B13. Mean gender results question 13

sex	N	Mean	Std Dev	Std Err	Minimum	Maximum
F	177	4.2718	0.4791	0.0360	2.0000	5.0000
M	359	4.2398	0.5349	0.0282	1.0000	5.0000
Diff (1-2)		0.0320	0.5171	0.0475		

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	534	0.67	0.5008
Satterthwaite	Unequal	386.95	0.70	0.4847

Table B14. Mean gender results question 14

sex	N	Mean	Std Dev	Std Err	Minimum	Maximum
F	177	4.2060	0.4996	0.0376	2.9554	5.0000
M	359	4.1444	0.5837	0.0308	1.0000	5.0000
Diff (1-2)		0.0615	0.5574	0.0512		

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	534	1.20	0.2298
Satterthwaite	Unequal	402.88	1.27	0.2059

Table B15. Mean gender results question 15

sex	N	Mean	Std Dev	Std Err	Minimum	Maximum
F	177	4.1940	0.4350	0.0327	3.0000	5.0000
M	359	4.1558	0.5303	0.0280	2.0000	5.0000
Diff (1-2)		0.0382	0.5009	0.0460		

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	534	0.83	0.4071
Satterthwaite	Unequal	418.01	0.89	0.3757

Table B16. Mean gender results question 16

sex	N	Mean	Std Dev	Std Err	Minimum	Maximum
F	177	4.2277	0.4545	0.0342	2.6667	5.0000
M	359	4.1930	0.5268	0.0278	2.0000	5.0000
Diff (1-2)		0.0347	0.5041	0.0463		

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	534	0.75	0.4545
Satterthwaite	Unequal	400.06	0.79	0.4319

APPENDIX C. REGRESSIONS ON DEMOGRAPHIC RESULTS

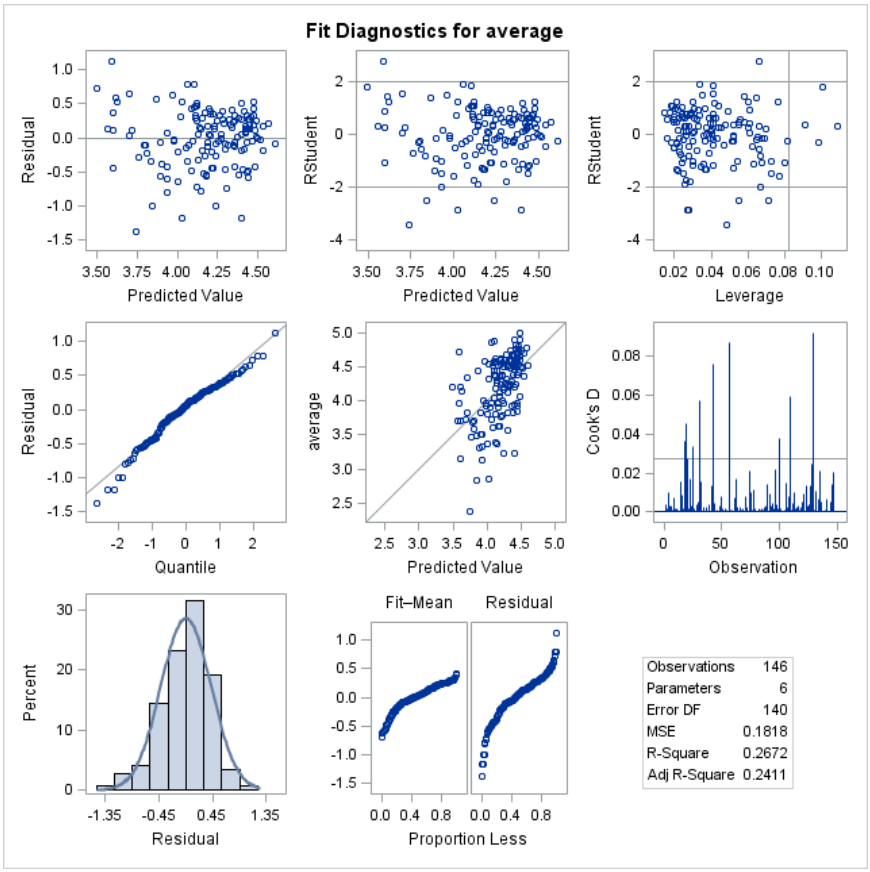


Figure C1. Residual diagnostics for question 1

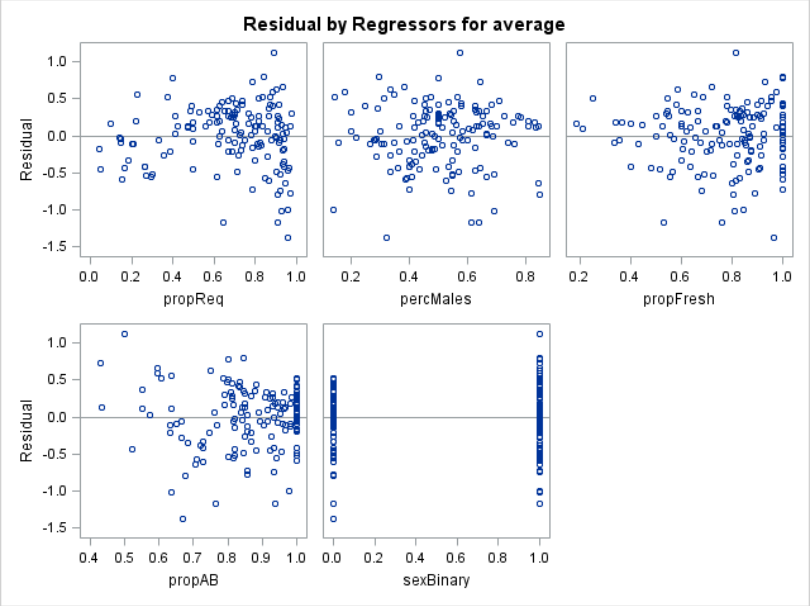


Figure C2. Residual plots for each variable for question 1

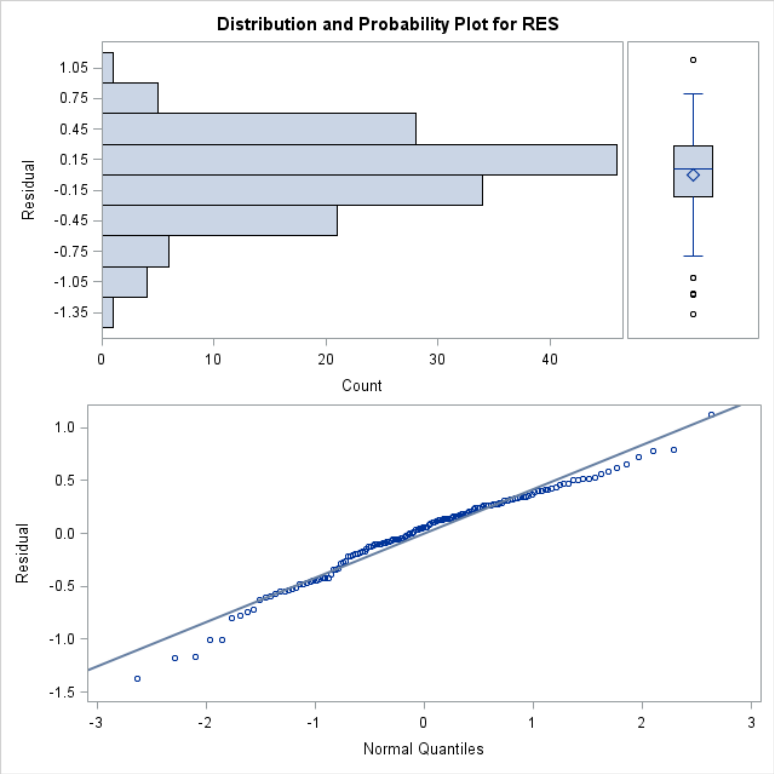


Figure C3. Histogram and q-q plot for question 1

Table C1. Demographic results for question 2

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	7.47092	1.49418	7.45	<.0001
Error	140	28.0732	0.20052		
Corrected Total	145	35.5441			

Parameter Estimates					
Variable	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**
Intercept	**1**	3.0635	0.35587	8.61	<.0001
propReq	**1**	-0.2022	0.15917	-1.27	0.2062
percMales	**1**	0.23537	0.25081	0.94	0.3496
propFresh	**1**	-0.083	0.20106	-0.41	0.6803
propAB	**1**	1.45734	0.28236	5.16	<.0001
instructorGender	**1**	0.0526	0.07891	0.67	0.5061

Root MSE	0.4478	**R-Square**	0.2102
Dependent Mean	4.26254	**Adj R-Sq**	0.182
Coeff Var	10.5054		

Table C2. Demographic results for question 3

Analysis of Variance					
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Pr > F**
Model	5	7.31912	1.46382	6.37	<.0001
Error	139	31.9341	0.22974		
Corrected Total	144	39.2533			

Parameter Estimates					
Variable	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**
Intercept	**1**	2.92889	0.38157	7.68	<.0001
propReq	**1**	-0.1236	0.17038	-0.73	0.4693
percMales	**1**	0.36982	0.26895	1.38	0.1713
propFresh	**1**	-0.0802	0.21568	-0.37	0.7105
propAB	**1**	1.43079	0.30225	4.73	<.0001
instructorGender	**1**	0.03946	0.08488	0.46	0.6427

Root MSE	0.47931	**R-Square**	0.1865
Dependent Mean	4.22168	**Adj R-Sq**	0.1572
Coeff Var	11.3536		

Table C3. Demographic results for question 4

Analysis of Variance					
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Pr > F**
Model	5	8.78193	1.75639	12.91	<.0001
Error	140	19.0463	0.13605		
Corrected Total	145	27.8283			

Parameter Estimates					
Variable	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**
Intercept	**1**	2.97174	0.29313	10.14	<.0001
propReq	**1**	-0.2058	0.13111	-1.57	0.1188
percMales	**1**	0.26466	0.20658	1.28	0.2023
propFresh	**1**	-0.1854	0.16561	-1.12	0.2649
propAB	**1**	1.52038	0.23257	6.54	<.0001
instructorGender	**1**	-0.0126	0.065	-0.19	0.8461

Root MSE	0.36884	**R-Square**	0.3156
Dependent Mean	4.12408	**Adj R-Sq**	0.2911
Coeff Var	8.94363		

Table C4. Demographic results for question 5

Analysis of Variance					
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Pr > F**
Model	5	9.46062	1.89212	18.16	<.0001
Error	140	14.5838	0.10417		
Corrected Total	145	24.0445			

Parameter Estimates					
Variable	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**
Intercept	**1**	2.80445	0.2565	10.93	<.0001
propReq	**1**	-0.1773	0.11473	-1.55	0.1245
percMales	**1**	0.21822	0.18077	1.21	0.2294
propFresh	**1**	0.03026	0.14492	0.21	0.8349
propAB	**1**	1.71669	0.20351	8.44	<.0001
instructorGender	**1**	0.0298	0.05688	0.52	0.6012

Root MSE	0.32275	**R-Square**	0.3935
Dependent Mean	4.30942	**Adj R-Sq**	0.3718
Coeff Var	7.4895		

Table C5. Demographic results for question 6

Analysis of Variance					
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Pr > F**
Model	5	16.5929	3.31857	26.94	<.0001
Error	140	17.2451	0.12318		
Corrected Total	145	33.8379			

Parameter Estimates					
Variable	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**
Intercept	**1**	2.57084	0.27892	9.22	<.0001
propReq	**1**	-0.2719	0.12476	-2.18	0.031
percMales	**1**	0.05457	0.19657	0.28	0.7817
propFresh	**1**	-0.2129	0.15758	-1.35	0.1789
propAB	**1**	2.1774	0.2213	9.84	<.0001
instructorGender	**1**	-0.0865	0.06185	-1.4	0.164

Root MSE	0.35097	**R-Square**	0.4904
Dependent Mean	4.07598	**Adj R-Sq**	0.4722
Coeff Var	8.61066		

Table C6. Demographic results for question 7

Analysis of Variance					
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Pr > F**
Model	5	6.90496	1.38099	10.65	<.0001
Error	141	18.2877	0.1297		
Corrected Total	146	25.1926			

Parameter Estimates					
Variable	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**
Intercept	**1**	3.0226	0.28522	10.6	<.0001
propReq	**1**	-0.1554	0.128	-1.21	0.2267
percMales	**1**	0.4285	0.20168	2.12	0.0354
propFresh	**1**	-0.0206	0.16085	-0.13	0.8983
propAB	**1**	1.33131	0.22644	5.88	<.0001
instructorGender	**1**	0.01946	0.06336	0.31	0.7592

Root MSE	0.36014	**R-Square**	0.2741
Dependent Mean	4.27463	**Adj R-Sq**	0.2483
Coeff Var	8.42502		

Table C7. Demographic results for question 8

Analysis of Variance					
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Pr > F**
Model	5	5.74474	1.14895	10.72	<.0001
Error	141	15.1188	0.10723		
Corrected Total	146	20.8635			

Parameter Estimates					
Variable	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**
Intercept	**1**	3.04808	0.25933	11.75	<.0001
propReq	**1**	-0.0551	0.11639	-0.47	0.6365
percMales	**1**	0.30419	0.18337	1.66	0.0994
propFresh	**1**	-0.0511	0.14625	-0.35	0.7274
propAB	**1**	1.29793	0.20589	6.3	<.0001
instructorGender	**1**	0.01017	0.05761	0.18	0.8601

Root MSE	0.32745	**R-Square**	0.2753
Dependent Mean	4.24835	**Adj R-Sq**	0.2497
Coeff Var	7.70776		

Table C8. Demographic results for question 9

Analysis of Variance					
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Pr > F**
Model	5	7.12291	1.42458	8.96	<.0001
Error	141	22.4228	0.15903		
Corrected Total	146	29.5457			

Parameter Estimates					
Variable	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**
Intercept	**1**	3.06152	0.31582	9.69	<.0001
propReq	**1**	-0.0855	0.14174	-0.6	0.5474
percMales	**1**	0.25023	0.22332	1.12	0.2644
propFresh	**1**	-0.1932	0.17811	-1.08	0.2798
propAB	**1**	1.41065	0.25074	5.63	<.0001
instructorGender	**1**	-0.0161	0.07016	-0.23	0.8194

Root MSE	0.39878	**R-Square**	0.2411
Dependent Mean	4.18761	**Adj R-Sq**	0.2142
Coeff Var	9.52289		

Table C9. Demographic results for question 10

Analysis of Variance					
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Pr > F**
Model	5	4.30065	0.86013	11.02	<.0001
Error	141	11.0027	0.07803		
Corrected Total	146	15.3033			

Parameter Estimates					
Variable	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**
Intercept	**1**	3.23286	0.22123	14.61	<.0001
propReq	**1**	0.01619	0.09929	0.16	0.8707
percMales	**1**	0.23915	0.15643	1.53	0.1286
propFresh	**1**	-0.0613	0.12476	-0.49	0.6242
propAB	**1**	1.17622	0.17564	6.7	<.0001
instructorGender	**1**	0.0673	0.04915	1.37	0.173

Root MSE	0.27934	**R-Square**	0.281
Dependent Mean	4.3665	**Adj R-Sq**	0.2555
Coeff Var	6.39744		

Table C10. Demographic results for question 11

Analysis of Variance					
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Pr > F**
Model	5	13.7695	2.7539	40.3	<.0001
Error	141	9.63571	0.06834		
Corrected Total	146	23.4052			

Parameter Estimates					
Variable	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**
Intercept	**1**	2.46348	0.20703	11.9	<.0001
propReq	**1**	-0.1418	0.09292	-1.53	0.1292
percMales	**1**	0.09657	0.14639	0.66	0.5106
propFresh	**1**	-0.1131	0.11675	-0.97	0.3342
propAB	**1**	2.10115	0.16437	12.78	<.0001
instructorGender	**1**	-0.0174	0.04599	-0.38	0.7055

Root MSE	0.26142	**R-Square**	0.5883
Dependent Mean	4.1284	**Adj R-Sq**	0.5737
Coeff Var	6.33215		

Table C11. Demographic results for question 12

Analysis of Variance					
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Pr > F**
Model	5	3.66561	0.73312	6.5	<.0001
Error	141	15.8979	0.11275		
Corrected Total	146	19.5635			

Parameter Estimates					
Variable	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**
Intercept	**1**	3.17847	0.26593	11.95	<.0001
propReq	**1**	-0.0235	0.11935	-0.2	0.8441
percMales	**1**	0.37822	0.18804	2.01	0.0462
propFresh	**1**	0.1156	0.14997	0.77	0.4421
propAB	**1**	0.92566	0.21113	4.38	<.0001
instructorGender	**1**	-0.0428	0.05908	-0.72	0.4697

Root MSE	0.33578	**R-Square**	0.1874
Dependent Mean	4.21735	**Adj R-Sq**	0.1586
Coeff Var	7.96198		

Table C12. Demographic results for question 13

Analysis of Variance					
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Pr > F**
Model	5	4.62422	0.92484	7.53	<.0001
Error	141	17.3276	0.12289		
Corrected Total	146	21.9519			

Parameter Estimates					
Variable	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**
Intercept	**1**	2.98367	0.27763	10.75	<.0001
propReq	**1**	0.02484	0.1246	0.2	0.8422
percMales	**1**	0.2387	0.19631	1.22	0.2261
propFresh	**1**	0.06559	0.15657	0.42	0.6759
propAB	**1**	1.21447	0.22042	5.51	<.0001
instructorGender	**1**	-0.0058	0.06168	-0.09	0.9257

Root MSE	0.35056	**R-Square**	0.2107
Dependent Mean	4.21458	**Adj R-Sq**	0.1827
Coeff Var	8.31774		

Table C13. Demographic results for question 14

Analysis of Variance					
Source	**DF**	**Sum of Square**	**Mean Square**	**F Value**	**Pr > F**
Model	5	6.32311	1.26462	9.67	<.0001
Error	141	18.438	0.13077		
Corrected Total	146	24.7611			

Parameter Estimates					
Variable	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**
Intercept	**1**	2.9357	0.28639	10.25	<.0001
propReq	**1**	-0.1214	0.12853	-0.94	0.3466
percMales	**1**	0.29973	0.20251	1.48	0.1411
propFresh	**1**	-0.0072	0.16151	-0.04	0.9644
propAB	**1**	1.34004	0.22737	5.89	<.0001
instructorGender	**1**	-0.0009	0.06362	-0.01	0.9884

Root MSE	0.36162	**R-Square**	0.2554
Dependent Mean	4.15258	**Adj R-Sq**	0.229
Coeff Var	8.70821		

Table C14. Demographic results for question 15

Analysis of Variance					
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Pr > F**
Model	5	3.71026	0.74205	9.1	<.0001
Error	141	11.4934	0.08151		
Corrected Total	146	15.2036			

Parameter Estimates					
Variable	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**
Intercept	**1**	3.16408	0.22611	13.99	<.0001
propReq	**1**	0.01256	0.10148	0.12	0.9016
percMales	**1**	0.4583	0.15988	2.87	0.0048
propFresh	**1**	-0.0114	0.12751	-0.09	0.9289
propAB	**1**	0.95276	0.17952	5.31	<.0001
instructorGender	**1**	0.02838	0.05023	0.57	0.5729

Root MSE	0.2855	**R-Square**	0.244
Dependent Mean	4.23173	**Adj R-Sq**	0.2172
Coeff Var	6.74676		

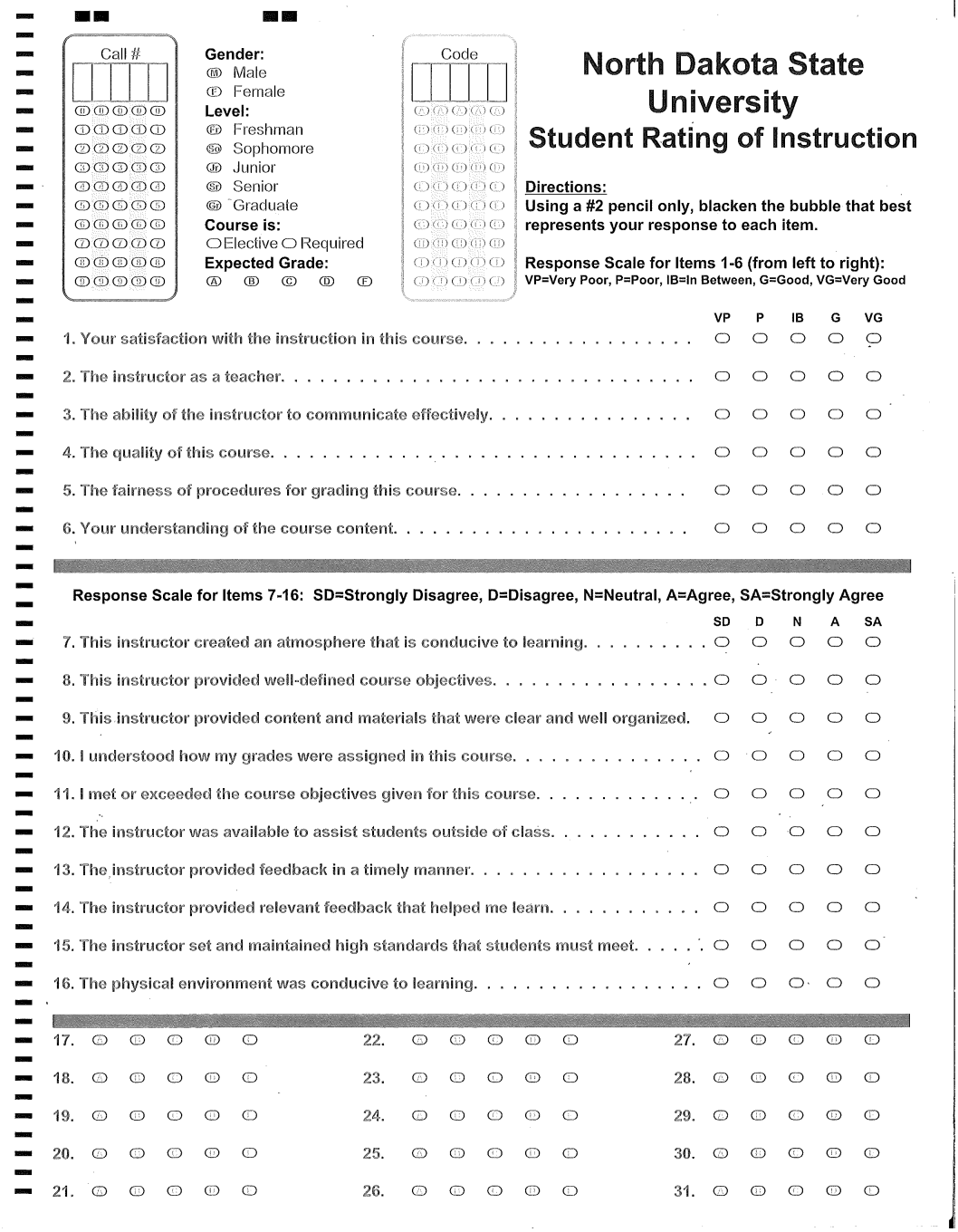
Table C15. Demographic results for question 16

Analysis of Variance					
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Pr > F**
Model	5	5.87179	1.17436	11.31	<.0001
Error	141	14.6457	0.10387		
Corrected Total	146	20.5175			

Parameter Estimates					
Variable	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**
Intercept	**1**	2.99106	0.25524	11.72	<.0001
propReq	**1**	-0.1271	0.11455	-1.11	0.2691
percMales	**1**	0.40087	0.18048	2.22	0.0279
propFresh	**1**	0.05473	0.14394	0.38	0.7043
propAB	**1**	1.25174	0.20265	6.18	<.0001
instructorGender	**1**	0.02121	0.0567	0.37	0.7089

Root MSE	0.32229	**R-Square**	0.2862
Dependent Mean	4.2387	**Adj R-Sq**	0.2609
Coeff Var	7.6035		

APPENDIX D. SROI FORM



Call #

Gender:
Ⓜ Male
Ⓕ Female
Level:
Ⓕr Freshman
Ⓢo Sophomore
Ⓙr Junior
Ⓢr Senior
Ⓖr Graduate
Course is:
○ Elective ○ Required
Expected Grade:
Ⓐ Ⓑ Ⓒ Ⓓ Ⓕ

Code

North Dakota State University
Student Rating of Instruction

Directions:
Using a #2 pencil only, blacken the bubble that best represents your response to each item.

Response Scale for Items 1-6 (from left to right):
VP=Very Poor, P=Poor, IB=In Between, G=Good, VG=Very Good

	VP	P	IB	G	VG
1. Your satisfaction with the instruction in this course.	○	○	○	○	○
2. The instructor as a teacher.	○	○	○	○	○
3. The ability of the instructor to communicate effectively.	○	○	○	○	○
4. The quality of this course.	○	○	○	○	○
5. The fairness of procedures for grading this course.	○	○	○	○	○
6. Your understanding of the course content.	○	○	○	○	○

Response Scale for Items 7-16: SD=Strongly Disagree, D=Disagree, N=Neutral, A=Agree, SA=Strongly Agree

	SD	D	N	A	SA
7. This instructor created an atmosphere that is conducive to learning.	○	○	○	○	○
8. This instructor provided well-defined course objectives.	○	○	○	○	○
9. This instructor provided content and materials that were clear and well organized.	○	○	○	○	○
10. I understood how my grades were assigned in this course.	○	○	○	○	○
11. I met or exceeded the course objectives given for this course.	○	○	○	○	○
12. The instructor was available to assist students outside of class.	○	○	○	○	○
13. The instructor provided feedback in a timely manner.	○	○	○	○	○
14. The instructor provided relevant feedback that helped me learn.	○	○	○	○	○
15. The instructor set and maintained high standards that students must meet.	○	○	○	○	○
16. The physical environment was conducive to learning.	○	○	○	○	○

17. Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ
18. Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ
19. Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ
20. Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ
21. Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ
22. Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ
23. Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ
24. Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ
25. Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ
26. Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ
27. Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ
28. Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ
29. Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ
30. Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ
31. Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ