DEVELOPING AND ASSESSING WRITTEN COMMUNICATION SKILLS IN LARGE INTRODUCTORY STATISTICS CLASSES

Elizabeth Johnson and David Holmes,
Department of Statistics, George Mason University, Fairfax, VA, USA
ejohns40@gmu.edu

A learning outcome for most introductory undergraduate statistics classes is for students to be able to clearly when class sizes are large or when graduate students are required to grade these assignments. This communicate statistical results from the analysis of authentic data. Achieving consistency in grading written assignments and providing appropriate feedback to students, however, is difficult especially paper will describe a method for developing assignments that requires students to use statistical software to analyze a large data set, designing a scoring rubric to grade students' written communication on these assignments and a training protocol to enable graduate students to consistently apply the scoring rubric to these assignments.

INTRODUCTION

The use of authentic data to motivate students and the use of statistical software to analyze this data has become an essential component to most post-secondary introductory statistics courses. The American Statistical Association (ASA) Curriculum Guidelines for Undergraduate Programs in Statistical Science (2016) recognizes the increased importance of data science, the use of real applications and the ability to communicate statistical results as vital elements in the development of a statistician. The ASA guidelines note that strong communication skills complement technical knowledge and that students should be given frequent opportunities to refine their communication skills and that this should be tied directly to instruction in technical skills. The Guidelines for Assessment and Instruction in Statistics Education (GAISE) College Report (2016) also documented the need for post-secondary students to be able to integrate real data with a context and purpose and the need for appropriate assessments to improve and evaluate student learning. The GAISE report lists nine goals of an introductory statistics class and the majority of these goals require assessments of the students' statistical understanding through their written communication. For example, "students should be able to interpret and draw conclusion from standard output from statistical software" would require students to write a conclusion to some type of analysis. These goals, however, produce difficulties for faculty as regards achieving consistency in grading written assignments and providing appropriate feedback to students. These difficulties are exaggerated when the class sizes are large and require graduate teaching assistants (GTA's) to grade these assignments. In this paper we will describe a method developed for assessing undergraduate students' written communication skills on the statistical analysis of real data in large enrollment courses.

METHOD

This method of assessing students' written responses was developed for an introductory statistics course at a higher education institution (HEI) in the U.S. This course has multiple sections with course enrollment averaging 1,000 students per semester. The course is not part of the Statistics major but is a component of the university's general education quantitative reasoning curriculum requirement. In this course, students are required to complete four data analysis assignments (DA's) per semester. These assignments require students to use a common data set and to provide written responses to questions involving both descriptive and inferential statistics using a statistical software package.

These four DA's were designed to cover the course learning objectives in this non-calculus based course. The first DA covered techniques for data visualization, organizing categorical data, and numerical summaries of center and variation. The second DA covered exploring associations between quantitative variables, using simulation to compare empirical and theoretical probabilities, and modeling random events with both Normal and Binomial distributions. The third DA covered use of an applet to investigate a sampling distribution of a proportion, factors that influence confidence intervals for population proportions, and constructing and interpreting hypothesis tests

and confidence intervals for a proportion. The fourth DA covered constructing and interpreting hypothesis tests and confidence intervals for one sample and two sample (paired and unpaired) means. All of these assignments required the students to use a common data set: the *Census at School* data. This data set, which is available through the ASA, was used as it contains the results of an international survey of students in grades 4-12 and provides numerous categorical and quantitative responses. For these assignments a random sample of 500 responses from high school students in the United States was provided to the students for use in completing the DA's throughout the semester. The DA's required the students to include elements of good technical writing in their final submissions e.g., using complete and coherent sentences, titling and labeling graphical displays, using correct comparative language and providing units when appropriate.

Below is an example of a DA question from the third assignment of the semester. This item required the students to select a sample from the larger data set, use that sample to construct and interpret a confidence interval and finally compare their confidence interval to the proportion from the population.

Example: DA#3 Confidence Interval for a Proportion

The Census at School survey item shown below asked high school students the following question.

30. Which of these methods do you most often use to communicate with your friends? Select one.

☐ In person   ☐ Text messaging   ☐ Internet chat or instant messaging

☐ Telephone (landline) ☐ E-mail   ☐ Other

Fifty ☐ Cell phone   ☐ Myspace, Facebook, other social networking sites, or blog   ethod to communicate with their friends. We will use this value as the population proportion (parameter) representing the percent of all high school students who use "Text messaging" as their preferred method of communicating with their friends. Select a simple random sample of 40 students from this population.

3a. Find the proportion of students *in your sample of size 40* who use "Text messaging" as their preferred method of communicating with their friends. Copy this table into your document and state the proportion of students *in your sample of size 40* who use "Text messaging" as their preferred method of communicating with their friends rounded to two decimal places.

3b. Use your sample data to construct a confidence interval. Since you selected a random sample from the population and the population is at least ten times larger than our sample of size 40, we must only check that the final condition required to apply the CLT (the sample is large enough) has been met. Perform this necessary check, show all calculations and state if all conditions have been met.

3c. Use *your sample of 40 students* to construct a 95% confidence interval for the proportion of all students who use "Text messaging" as their preferred method of communicating with their friends.

3d. Use your result from part 3c to state the lower and upper limits of the confidence interval as a percentage rounded to one decimal place in the following format (lower limit %, upper limit %).

3e. Interpret this confidence interval in a complete sentence in the context of the question.

3f. Recall that the population proportion was found to be 58%. Does your confidence interval support this value? Explain your answer in a sentence.

ASSESSMENT

As there are four questions per assignment and four assignments per student and 1,000 students per semester, we had approximately 16,000 questions to grade. These DA's are graded by Graduate Teaching Assistants (GTAs). Our GTAs are students who are enrolled in our Statistics Master's and Doctoral programs but who do not necessarily have any prior teaching experience. Our GTAs' main responsibilities include working in the statistics tutoring center and grading. The GTAs come from diverse backgrounds and, for many, English is not their first language. In short, we had many assignments for untrained graders.

A detailed scoring rubric was therefore developed for each question on the four DA's. The goal of each scoring rubric was to ensure consistency in the grading process for these assignments while providing feedback to the student on their work. The scoring rubric was designed to assess the overall elements of good technical writing as well as presentation of data and written communication of results.

Faculty members conducted grading training sessions with the GTAs after each DA assignment and its scoring rubric had been developed but before the assignment was released to the classes. During these sessions the faulty members stressed the importance of consistency when grading. By reviewing the scoring rubrics there was time to revise any questions or scoring guidelines based on feedback from the GTAs before these assignments were released. The GTAs found these training sessions to be particularly helpful in their understanding of what the faculty were specifically looking for in a student response to a question.

To ensure that the GTAs were following the rubric, faculty members randomly selected student papers and back-graded the GTAs' scores. If there was a discrepancy between the GTA and faculty scores, the faculty member worked with the GTA to determine and clarify the source of the score disagreement before the scores were released to the students.

Below is the scoring rubric that was developed for the question shown previously. As this question required students to select their own random sample, the rubric had to contain grading contingencies for occasions when the students did not follow the instructions and used the entire population of responses and also if the students selected a sample where the conditions for conducting inference were not met. The GTAs also had to be trained to recognize that student responses would vary.

Example: Scoring Rubric DA#3 Question 3

Question 3: Confidence Intervals for One Sample Proportion (Total 25 points)
Each student must select their own random sample. If the student uses the entire data set and not a sample, they do not receive these 6 points. Then grade the remainder of their response using the entire data set where n=253 and p=0.5816 for this variable. (6 points)
Table showing % of total for all four methods of communication (1 point)   Selection of 'text messaging' and rounded value (1 point).
n$\hat{p}$ > 10 correctly checked (1 point); n(1-$\hat{p}$) > 10 correctly checked (1 point) and states if final condition has been met (1 point)
Correct variable selection (1 point); Correct "Text messaging" for success entered (1 point); Correct confidence level of 0.95 (1 point); Output (1 point).
State lower and upper limits (1 point) of interval as a percent (1 point) rounded to one decimal place (1 point)
Answer should contain: "95% confident" (1 point); Interval "Captures p" or "contains p" (1 point); Interval values from part 3(d) (1 point); Sentence in context stating the parameter of interest (2 points)
Correctly states (yes/no) whether the confidence interval supports the population proportion (1 point) by determining whether or not their interval contains the true population proportion value of 58% (1 point).
*Answers will vary according to their selected sample. In a few cases the condition in part 3b may not be met and the conclusion may also differ.

RESULTS

Initial feedback on the use of the scoring rubrics is that they have resulted in an increase in the grading consistency across the multiple sections of this course and they have decreased student confusion over their grades. One method for measuring these outcomes was by observing the decrease in the number of emails faculty received that semester from students questioning the scoring of their assignments. In previous semesters there was a high volume of student emails and office visits in which students asked to have their grades reviewed due to either inconsistency in

the scoring or they were uncertain as to why they lost points. An additional measure of the effectiveness of these assignments and the scoring system was obtained by reviewing the end-of-course student evaluations. As this is not a required course for the vast majority of the students, they therefore tend not to comment at all on these evaluation forms. Any written response, therefore, is considered valuable information. Some examples of student responses included:

"*The data analysis assignments really helped with understanding the material.*"
"*The DA's helped a lot.*"
"*The data analysis really helped me learn.*"
"*DAs were very helpful*"
"*I thought that the DA's helped to solidify my understanding of the material.*"

CONCLUSIONS

The use of scoring rubrics on the DA's has allowed us to assess and provide timely feedback on a student's written communication of a statistical analysis using real data in courses with very large enrollments. In addition, these scoring rubrics have allowed us to score questions that utilized applets and simulation studies in which student responses will vary. A side benefit of the development of these activities was that we created reusable assignments. Since we selected a random sample of responses from the Census at School data set we can reuse the questions and their scoring rubrics in future semesters by simply selecting another random sample from the same data set. Finally, the use of scoring rubrics had an unintended benefit of helping our graduate students. The training sessions helped the GTAs to understand what the faculty expected from the students in terms of written assignments which aided them with their work in the statistics tutoring center.

REFERENCES

American Statistical Association (2016), *Curriculum Guidelines for Undergraduate Programs in Statistical Science*, http://www.amstat.org/education/pdfs/guidelines2014-11-15.pdf

GAISE College Report ASA Revision Committee (2016). *Guidelines for Assessment and Instruction in Statistics Education College Report.* Retrieved from http://www.amstat.org/education/gaise.

American Statistical Association (2016), *Census at School* [Data set]. Retrieved from http://ww2.amstat.org/CensusAtSchool/