A FINAL PROJECT DESIGN: THREE PHASES FOR SUCCESS

Carrie E. A. Grant
Flagler College, Mathematics and Technology Department
74 King Street, Saint Augustine, FL 32084
grantc@flagler.edu

Abstract

In this paper, I will explain the format of a final project in an introductory statistics course that incorporates three phases that are due throughout the semester. This paper will outline the details of each phase of the project as well as data collection, resources, and the reports. Furthermore, an explanation of why this format was adopted and the results of this project design will be discussed.

1. Introduction

In 2005, the American Statistical Association published the Guidelines of Assessment and Instruction in Statistical Education (GAISE) [1]. In 2016, a revised GAISE report [2] was published. In both reports, the recommendations remained the same. The recommendations are (1) teach statistical thinking, (2) focus on conceptual understanding, (3) integrate real data with a context and a purpose, (4) foster active learning, (5) use technology to explore concepts and analyze data, and (6) use assessments for improvement and evaluation of student work. All of these recommendations support the inclusion of a semester long comprehensive final project in an introductory statistics course. Therefore, final projects have become a common assessment in many introductory statistics courses.

One challenge of a final comprehensive project is that it is often due at the end of the semester. Students and instructors alike are filled with other due dates and expectations at the end of the semester. Thus, dividing the project into three distinct phases that correspond to the three units of study in a traditional statistics course eases the burden of one large assessment to create and grade at the end of the semester.

The three phases of the project design explained in this paper focus on (1) descriptive statistics, (2) association between variables, and (3) inferential statistics. The three phases use the same sample data but each phase analyzes the data statistically different. Phase one uses descriptive statistics to describe the sample of data gathered through graphical displays and summary statistics. In phase two, the sample data is divided into two smaller samples and three comparisons are made between these samples; one using numerical data and two using categorical data. Phase three requires students to use inferential statistics to make conclusions regarding the population based on the sample data. Students run two hypothesis tests and create two corresponding confidence intervals, one using one sample and one comparing two samples.

This project design outlined in this paper has been used by a team of instructors for traditional face-to-face classes as well as in the online environment. Thus, careful design of instruction, implementation, and grading has been considered to make this project an effective assessment that also provides students with a deeper understanding of the statistical process.

2. Background

MAT 223 Statistics is a required general education course at Flagler College, a small private teaching college of 2500 students in St. Augustine, Florida. Approximately 350 students are enrolled in MAT 223 Statistics each academic semester. This course is taught by a team of six instructors and is completely coordinated. Serving as the course coordinator, I design a common syllabus with the input from the other instructors on the team each semester. This syllabus serves to coordinate all sections of the course. All instructors on the team use the same assessments (homework, quizzes, exams, and final project) and use the same course policies.

Prior to fall semester 2018, a final project was designed and used by all instructors, but was not divided into different phases. The project design incorporated a project proposal, a project template, and a rough draft. These resources were intended to help the students be successful in the final draft of the project and to ease the final grading of the project. Even with these measures in place, many of the students procrastinated and waited to begin the final draft of the project until the day before it was due. Since the project was designed to be comprehensive, the project was long. So, not only was it a challenge to create by the students, it was a challenge to grade by the instructors. Many instructors would grade the rough draft, if submitted, so in essence, many instructors were grading the final project twice.

During the summer term 2018, I designed the phased project outlined in this paper as a way to ease the end of the semester burden the previous design created. This design has been successfully integrated into MAT 223 Statistics by the team of instructors at Flagler College teaching statistics during the fall semester 2018 and spring semester 2019.

3. The Resources

Flagler College uses Canvas for its learning management system and all instructors are required to use Canvas for attendance and grades. The statistics team at Flagler College uses course integration to link MyStatLab^{TM} content into their Canvas courses. They also use modules on Canvas to deliver content and information.

One of the modules on Canvas available to all students enrolled in MAT 223 Statistics is the Final Project Information Module. I created this module for the instructors on the statistics team to copy into their Canvas courses. This module contains multiple resources for the students to guide them to success on the three phases of the project.

In the Final Project Information Module, students will find an overview of each phase of the project, a document containing a list of all the surveys created to gather the sample

data, a grading rubric for each phase of the project, and multiple instructional videos to guide students to success for each phase of the project.

For Phase One, there are five instructional videos. These videos cover how to rename columns in a StatCrunch™ spreadsheet and how to save the sample data, how to create graphical displays of categorical and numerical data, how to calculate summary statistics needed for phase one, and how to create a StatCrunch™ report. The report construction video is extremely important to the students as it teaches them how to embed StatCrunch™ results and data into the report.

For Phase Two, there are also five instructional videos. These videos cover how to define the two samples, how to create stacked boxplots and five-number summaries, how to create split bar graphs, how to create contingency tables and, again, how to create a StatCrunch™ report. The report construction video for this phase focuses on the necessary components of phase two and how to embed the results correctly.

For Phase Three, there are four instructional videos. These videos cover how to correctly define the two hypothesis tests, how to run a one-proportion hypothesis test and create a one-proportion confidence interval, how to run a two-proportion hypothesis test and create a two-proportion confidence interval, and again how to create a StatCrunch™ report.

The students are given more than adequate instruction on how to correctly complete the three phases of the project. Most of the instruction is completed outside of class, but since the instructional videos can be replayed more than once, the students embrace these virtual instructions.

Typically, instructors explain to student groups that they should meet out of class with three laptops. Simultaneously, one laptop should be used to play the video instructions, one laptop should be used to have the sample phase of the report open (available on the StatCrunch™ group page for each section of the course), and one laptop should be used to construct the report.

4. Data Collection – The Survey and Samples

At the beginning of the semester, twenty surveys were created. Each survey followed the same format; the first two questions were gender identification and age followed by two more questions that required a numerical response and six more questions that resulted in a categorical response. The first questions typically were demographic in nature while the last questions focused on student opinion. Each survey was based on a common theme so the questions were related to one another. For example, Flagler College uses a plus and minus grading scale, so one survey focused on this policy. A document containing all the surveys is made available to the students in the Final Project Information Module on Canvas.

A StatCrunch™ survey was created for each of the twenty surveys and then four homework assignments on MyStatLab™ were made each containing five surveys. The homework was assigned to all students enrolled in MAT 223 Statistics. Thus, over 300 students completed each survey each semester. For the project, students are divided into groups of

two or three and were assigned a unique sample of 150 student responses. Multiple samples of size 150 from each survey were created using the sampling feature on StatCrunch™.

5. StatCrunch Groups

For each section of MAT 223 Statistics, the instructors created a StatCrunch™ group. The purpose of the group page was to share data and reports. The unique samples of size 150 were made available to the student groups on these group pages in the third week of classes. Each phase of the project was submitted as a StatCrunch™ report shared with the StatCrunch™ group. The reports contained the graphs, tables, summary statistics, test and confidence interval results, and data embedded into the report for ease of grading.

One student in each of the student groups was designated the leader of the group. This student would be the student who would own the data and the reports. Some groups chose a different leader for each phase of the project while others used the same leader for each phase. Prior to constructing the first phase of the project, the student groups were required to clean up the samples by renaming the columns to reflect the sample data and then saving the data to their StatCrunch™ account. The groups were also instructed to create their results and save these results on their StatCrunch™ account for future use as they worked on each phase of the project.

6. Phase One

For the first phase of the project, students are required to use descriptive statistics to describe the ten variables in their sample of 150 responses. The report consists of an introduction, a section on data collection, a list of the survey questions, and then an appropriate graph, summary statistics, and short description of the distribution of data gathered for each variable.

For the three categorical variables, the students are required to create either a bar graph or a pie chart with a title and the percent within each category displayed on the graph. A short description followed which must include a statement of the mode of the distribution.

For the seven numerical variables, the students are required to create either a histogram or a dotplot with the frequency displayed on the vertical axis and then calculate the following summary statistics: the sample size, the mean, the median, the mode, the standard deviation, the range, and the interquartile range. A description of the distribution must include the shape of the distribution, the best measure of center and spread, a discussion of any possible outliers, and a clear understanding of what the data reveals.

The StatCrunch™ report for this phase of the project includes the graphs and summary statistics embedded in the correct location in the report and the updated sample data with renamed columns.

7. Phase Two

For the second phase of the project, the students limit the focus of their report from all ten variables to four variables, three categorical variables and one numerical variable. One of the categorical variables is used to divide the sample of 150 students into two smaller samples. The report consists of an introduction, three comparisons, and a conclusion. The

three comparisons will be made between the two samples defined in the introduction. One comparison is with a numerical variable and the other two comparisons involve categorical variables.

Students can use any of the categorical variables to define their two samples, but the samples must be close in size. One common idea is to compare responses between males and females although there are many other options that students can choose to use. For example, for the sample results regarding the plus and minus grading scale at Flagler College, students could compare those students who do support the addition of an A+ with those students who do not support the addition of an A+ as long as these two groups are similar in size. The two samples defined in this phase will be used again in the third phase of the project.

In the introduction, students define their two samples, create a bar graph with a title representing the two samples with the sample size identified on the graph, and give a brief introduction to the three comparisons that will be included in the report.

For the first comparison, students compare one of the numerical variables between their two samples. The numerical variable selected cannot be the variable age but rather one of the other two numerical variables. Students are required to create horizontal stacked botplots with a title that identifies outliers, compute the five-number summary for each sample, and then write a short paragraph comparing the results between the two samples.

For the second comparison, students compare one of the categorical variables between the two defined samples. Students create a split bar graph with a title that has the two samples on the x-axis and the columns representing the variable. The percent within each sample is identified in the split bar at the top of each bar. A short discussion of the comparison follows that compares the percent within each sample.

For the third comparison, students compare a different categorical variable between the two defined samples. Students create a contingency table with the rows representing the two samples and the columns representing the variable. Students define one of the responses to the categorical variable as a success and then compute the probability of success for the entire sample of 150 students and the conditional probability of success for the two smaller samples. Students write a short paragraph to compare the calculated probabilities and determine if a relationship exists between the two samples in regards to the variable.

The final part of the second phase of the report is a conclusion that summarizes the results of the comparisons.

The StatCrunch™ report for this phase of the project includes the graphs, contingency table, and summary statistics embedded in the correct location in the report and the updated sample data with renamed columns.

8. Phase Three

For the third phase of the project, students run two hypothesis tests and construct two corresponding confidence intervals. One of the two categorical variables discussed in the second phase of the report is the focus of the third phase of the project. Typically, students will choose the variable from the contingency table, but either categorical variable is appropriate.

This report consists of an introduction, a one-proportion hypothesis test, a one-proportion confidence interval, a two-proportion hypothesis test, a two-proportion confidence interval and then a conclusion. First a test of majority is run for the entire population and a corresponding confidence interval is created. The two-proportion hypothesis test and the corresponding two-proportion confidence interval uses the two samples defined in the second phase of the project and the same categorical variable used in the one-proportion hypothesis test and confidence interval.

In the introduction, students are required to again define the two samples from phase two of the project, include the bar graph with title and sample sizes above the bars for the two samples, and then introduce the hypothesis tests and confidence intervals they plan to run and create in this phase of the project.

For the one-proportion hypothesis test, the population of interest is all 2500 students at Flagler College and the sample is the 150 students represented by the sample data. Students are required to give a brief introduction explaining the test of majority they plan to run, a pie chart representing the variable of interest that includes a title and the percent within each category, the four steps of the hypothesis test (hypothesize, prepare, compute, and interpret), and a final conclusion regarding the original claim.

The corresponding one-proportion confidence interval estimates the population proportion that was tested in the first hypothesis test. Students are required to include an introduction explaining what they plan to estimate identifying the appropriate level of confidence, and the three steps of creating a confidence interval (prepare, compute, and interpret). The interpretation needs to include the level of confidence and a statement of the population parameter being estimated.

The two-proportion hypothesis test is a test of difference between the two population proportions. The two populations are based on the samples defined in phase two. For example, if the samples defined in phase two were males and females, then the two populations are the male students at Flagler College and the female students at Flagler College. If, as in the example with the plus and minus grading scale, the two samples are those students who support the addition of an A+ and those students who do not support the addition of an A+, the two populations would be all students at Flagler College who support the addition of an A+ grade and all students at Flagler College who do not support an A+ grade.

For the two-proportion hypothesis test, students must include a brief introduction explaining the hypothesis test, a contingency table with the samples identified on the rows and the variable from the first hypothesis test as the columns, the four steps of a hypothesis

test (hypothesize, prepare, compute, and interpret), and a final conclusion regarding the original claim.

The corresponding two-proportion confidence interval estimates the difference in the population proportions that were tested in the second hypothesis test. For this section of the report, students must include an introduction explaining what they plan to estimate identifying the appropriate level of confidence, and the three steps of creating a confidence interval (prepare, compute, and interpret). The interpretation needs to include the level of confidence and a statement of the difference being estimated. Furthermore, an explanation of the importance of zero being contained in the confidence interval, if appropriate.

The StatCrunch™ report for this phase of the project includes the graphs, contingency table, results to the hypothesis tests and results to the confidence intervals embedded in the correct location in the report and the updated sample data with renamed columns.

9. Conclusions

Overall, the new phased project design with the enhanced resources available to the students has been a great success, although there are some limitations.

Instructors admit that the new design reduces grading of both the rough draft and final project, limits student procrastination, eliminates grading a major report at the end of the semester, and spreads the grading throughout the semester.

According to student comments on course evaluations, the resources available to them completely support their success on the project and make the project extremely easy to complete as long as they follow the detailed directions.

As with any group project, there are challenges that are difficult to overcome. One challenge that I hope may be changed with updates is that multiple StatCrunch™ accounts cannot create one comprehensive report, so one student becomes responsible to complete each phase. Other challenges that are at the core of any group project in an undergraduate course include students dropping the course and students not working together. These are challenges I am sure I cannot overcome, but hope to minimize.

Furthermore, I have found that with all the resources available to the students, students may not use them. A student commented that "You can lead a horse to water, but you cannot make them drink!" This saying is extremely true of the students enrolled in MAT 223 Statistics. No matter how many resources that are created for them to be successful in any aspect of the course, I cannot make them open and use these resources.

References

[1] Aliaga, M., Cobb, G., Cuff, C., Garfield, J., Gould, R., Lock, R., Moore, T., Rossman, A., Stephenson, B., Utts, J., Velleman, P., Witmer, J., *Guidelines for Assessment and Instruction in Statistics Education (GAISE) College Report*, American Statistical Association, 2005.

[2] Carver, R., Everson, M., Gabrosek, J., Horton, N., Lock, R., Mocko, M., Rossman, A., Rowell, G. H., Velleman, P., Witmer, J., and Wood, B., *Guidelines for Assessment and Instruction in Statistics Education (GAISE) College Report*, American Statistical Association, 2016.