Betting your Grade: Using Introductory Statistics to Teach Data-Driven Decision Making

Mark Archibald
Grove City College

Abstract

This paper describes a technique for enhancing student motivation and providing meaningful context for introductory statistics. Additionally, students learn how to make – and the importance of – data-based decisions. Sophomore students in Grove City College's experimental methods lab course were unable to perform or understand basic statistical data analyses such as confidence intervals, regression and tests of hypotheses. Basic instruction using Matlab's statistical toolbox allowed some students to perform the analyses by rote, but the understanding was poor, and many students continued to struggle. In 2014, the structure of the course was changed, reducing the number of experiments and increasing the time devoted to data analysis by spending two weeks on each lab. During the first week, students conduct the experiment and record data, followed by an introduction to the statistics principles required to answer the lab questions. The following week, after a more detailed explanation, students are required to demonstrate an understanding of the statistical method by making predictions based on their own measured data. For example, in an experiment involving linear regression, students use their data from the previous week and a regression model to predict the outcome of three new trials for which the instructor provided values of the independent variable. The students then experimentally measure the outcomes. Their grade for the demonstration depend on the width of their prediction intervals and the fraction of trials within their predicted intervals. This provides very high motivation for learning the statistical method, and also set an example for how engineers make decisions based on data. Application of statistics to their end-of-semester student-design lab assignment show that students better understand statistical methods and their applications. Other metrics, such as quiz scores and course evaluations, did not show clear improvement, possibly due to increased rigor on the quizzes.

Keywords

First year students, statistics, engineering education, experimental methods course

Introduction

This paper describes a technique for introducing first or second year students to statistics in a context that enhances motivation and adds meaningful context to the learning environment. Additionally, students learn how to make data-based decisions and the importance of data-driven decisions. Sophomore students at Grove City College take an experimental methods lab course. Due to a quirk in the curriculum, most students have not taken a statistics course and have little or no experience with even basic statistical methods. As a result, they were unable to perform or understand even simple data analyses such as confidence intervals, regression and tests of hypotheses. Basic instruction using Matlab's statistical toolbox allowed some students to perform the analyses by rote, but the understanding was poor, and many students continued to struggle. In 2014, the structure of the course was changed in order to address the problem. The number of experiments was reduced and two weeks were spent on each experiment. During the first week, students conducted the experiment and recorded data. During the last hour of the lab, the instructor presented questions that could not be answered without a statistical analysis. Students were then introduced to the statistics principles required to answer the questions. The following week, the instructor spent the initial portion of lab explaining in more detail the relevant statistical methods. Students were then required to demonstrate an understanding of the principles by making predictions based on their measured data. For example, in an experiment involving linear regression, students used their previous week data and a regression model to predict the outcome of three new trials for which the instructor provided values of the independent variable. The students then experimentally measured the outcome. Their grade for the demo depended on the width of their prediction intervals and the number of trials within their predicted intervals. The initial goas were to motivate students and give meaningful context for the statistical methods taught. It soon became clear that the method also help students make rational, data-based decisions.

The importance of student motivation is widely recognized. Korolkiewicz and Chiera describe an approach to teaching statistics to first year students in non-math majors.[1] These students often did not understand the application of math or statistics to their discipline, suffered math anxiety, or were poorly motivated. It was found that a shift in emphasis away from statistics and mathematics and toward business and real-life situations (perceived a more relevant) led to improved outcomes. Siepermann used a card game to teach fundamental statistical principles to business science majors.[2] Students were motivated to lean the statistical principles when they realized it helped them win in a game. The approach outlined in this paper – using student demonstrations – can be thought of as a game. Siepermann found that serious games can be effective for teaching unpopular subjects if the topic is inherently tied to game itself. This suggests that the student demonstrations will be most effective if the statistical method involved is critical for solving the task at hand. Within the context of teaching experimental methods, both Bridges[3] and Anthony[4] emphasized the importance experimental planning, critical thinking, and analytical competence.

Methods/Implementation

Either three or four student demonstration experiences are provided during the first half of the semester. The statistical principles covered include descriptive statistics and prediction, regression analysis, hypothesis testing, and (only in some years) ANOVA. The specific experiments used for each lab have varied somewhat since the initial implementation of student demonstrations in 2014 in order to improve the effectiveness of the lab. Although many different experiments could be used to teach each concept, there are certain characteristics that improve the effectiveness of the experiment for teaching the statistical principles involved.

1. The experimental method data collection should be fairly simple so students can concentrate on what the data means rather than how to collect the data.

2. The results should clearly require the statistical method used to obtain the desired outcome of the student demonstration.

3. The experiment should be interesting or provide a novel insight in order to motivate students.

Examples are provided for labs that have been used, along with qualitative assessments of how well each met the criteria above and student reactions. In all cases, the lectures emphasized calculation and interpretation rather than deep theoretical understanding of the statistical principles involved. Each student group was required to submit their data at the end of the first week. The instructor reviewed and compiled the data, which was used to establish the rubric for the second week demonstration. Experiment 1 will be treated in some depth to illustrate the method used. Other experiments will be briefly summarized.

Experiment 1, Tensile Strength Of Copy Paper: The objective of this lab was to teach students basic descriptive statistical methods, including familiarization with the Normal distribution and probabilities. In a one-hour lecture, students were introduced – at a very basic level – to sample and population mean and standard deviation, the Normal distribution, Normal distribution plots, and probabilities associated with the normal distribution. The first week, students were given narrow strips of 20lb copy paper, which they glued into loops. The loops were suspended from a round bar with a yoke hung from the bottom. Weights were incrementally added to the yoke until the strip of paper fractured. Up to 25 samples were taken. Students were encouraged to record the relative humidity in the lab. The following week, Students were given an opportunity to repeat the experiment. (The lab is not climate controlled, so humidity varies, and the paper strength varies with humidity. This provided a nice opportunity to discuss extraneous variables that can affect an experiment.) When they were confident in their data, each group was asked to predict the failure load for three subsequent trials. Specifically, students provided the range of loads in which failure would occur and the probability of failure in the predicted range based on their own data (under the Assumption the data was normally distributed.) This provided a dilemma for the students: the rubric favored a narrow range

The grade for the lab consisted of two parts, a lab memo and the student demonstration. The memo (weighted more heavily than the demo) consisted of an abbreviated lab report using a template provided by the instructor. The student demonstration grade was based on the student group's ability to use statistics to predict failure. The rubric, shown in Equation (1.1), is based on the range r of the predicted interval (range = upper limit – lower limit) and how many of the three trials failed within the predicted interval (n). The rubric provides a small bonus for narrow prediction ranges, but a significant penalty for either very large ranges or for any trials that fall outside the predicted range. If the student group does not correctly calculate the probabilities corresponding to their chosen range, a penalty is assessed.

$$\begin{aligned} & r = (UL - LL) \\ & \text{if}(r <= 1000) \qquad s = \frac{(120 - .02r)n}{3} \\ & \text{if}(r > 1000) \qquad s = \frac{(350 - .25r)n}{3} \end{aligned} \tag{1.1}$$

If all three trials are within the range, the resulting score is shown in Figure 1. Note that for this example, if all three loops fail within the predicted range, score of 100% is obtained if the range is exactly 1000 grams. Smaller ranges give bonus, but larger ranges incur penalties. This places the students in a dilemma: choosing a smaller interval increases the likelihood that a trial will be out of range, but could potentially lead to a bonus. This dilemma provides a strong motivation to understand how to calculate probabilities based on experimental data.

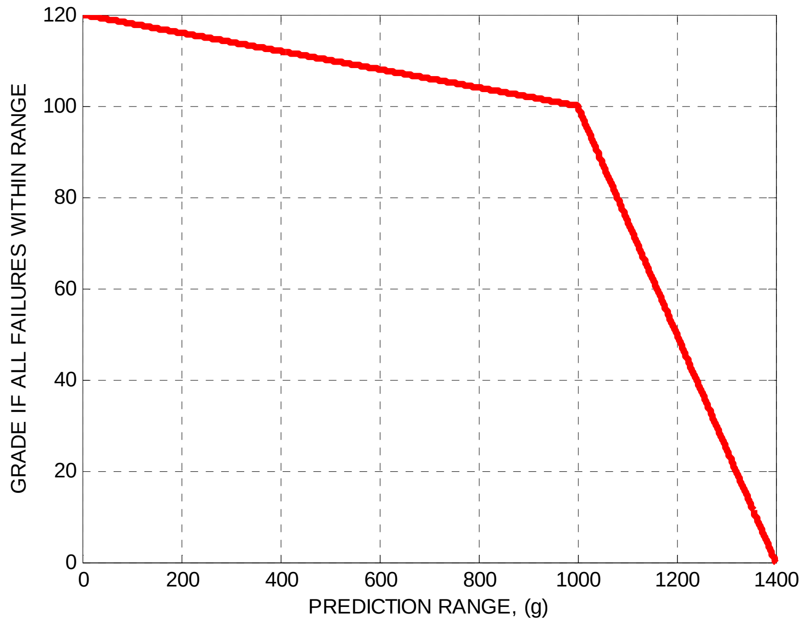

Figure 1 Scoring rubric for Lab 1 student demonstration

Most students are strongly motivated by the demonstration, and do well. Some students are stressed by the concept that their grade has an element of chance involved. These students are assured that they have control over that probability, and advised that much of engineering requires trade-off decisions. Failures outside the predicted range do occur for some groups each year. Usually, these are due either to calculation errors or judgement errors (choosing too small an interval in the hopes of getting a bonus.) Rarely, a student group will not understand the rubric, and as a result improperly choose a very inappropriate range above the maximum limit. Since this results in an automatic score of zero, these groups have been permitted to repeat the demonstration.

Two shortcomings in lab planning are worth noting as precautions for what not to do. In the first year this demonstration was required, the instructor used a small tensile tester to load the strips for the student demonstrations. It soon became clear that the tensile tester loaded the paper strips more smoothly than was done by the students during the initial data collection, resulting in a statistically significant difference in breaking load between the two loading methods. Subsequently, students were instructed to conduct their demonstration on the same equipment and using the same methods as were used in taking the original data. The second problem appeared in the second year, when many student groups found their predictions to be very poor, with many out-of-range results. Investigation showed that the humidity in the lab was significantly higher on week two, during the demonstrations, than week one, when the data was collected. The lab was changed to allow students to take a second set of data at the beginning of lab on week two. They are encouraged to compare data from the two weeks, and make a rational choice as to which data to use for their demonstration prediction – first week, second week, or both weeks. The final data set used must be submitted to ensure the instructor and student group share the same data set.

This experiment meets all three of the criteria for an effective lab. The data collection is easy (although it can be tedious at times), the prediction would be difficult to do accurately without using basic statistics, and students find it interesting. Most students are quite surprised at the strength of copy paper.

Experiment 2, Pendulum-Launched Projectile: The objective of this lab was to teach students to use linear regression to define the relationship between two measured variables, and to use the resulting model to make predictions. Students were given a pendulum assembly – effectively a steel hammer hung from a pivot – and a steel ball. The pendulum was released from a variable height h above the table and struck the steel ball resting on a platform. The horizontal distance the ball traveled until it struck the table was measured and recorded as variable d. After multiple trials with different pendulum release heights, regression analysis was used to model the distance as a function of release height. Students used Matlab with the Statistical Toolbox to fit the model. As with all statistical methods in this course, the emphasis was on correct implementation and interpretation rather than theory. This was especially true with the prediction intervals, which were provided by the Matlab functions and treated as a "black box".

During the first week, students conducted the experiment and collected the data. A lecture on regression analysis was also provided. In the second week, students conducted their demonstration. Since there was no clear theoretical model order for the height/distance relationship, students had to select and justify an appropriate order. They were then given three randomly generated release heights and asked to provide prediction intervals for each. As with the first lab, students were free to choose a range corresponding to any desired probability, and the grading rubric was similar. This approach clearly does not teach regression analysis at the algorithmic level, but it does teach students to correctly use and interpret regression results.

This lab has been used for three years with very little modification. Data collection is easy and regression is required to generate prediction intervals. Students – at least initially – seem to enjoy knocking balls across the lab tables. The demonstration appears to give students more motivation for learning how to conduct and interpret a regression analysis.

Experiment 3, Vibration of a SDOF Spring-Mass System: The third lab is designed to introduce students to hypothesis testing using the Student's T-test to compare the natural frequency of two similar spring-mass systems. Students use a stopwatch to measure the period of oscillation for a known mass suspended by a spring. The experiment is repeated with a different, but similar, spring, and the periods are compared. As with other labs, the first week includes a lecture on hypothesis testing, Student's T distribution, and interpretation of the resulting p-value. The student demonstration in the second week initially posed a challenge, however. Students were given a spring and mass, and told that the system was designed for a particular period of oscillation. They measured the new spring/mass period and compared it to the design value. The effectiveness of the lab was compromised due to the low variance of the resulting data –

many students correctly rejected the null hypothesis when the measured value differed only slightly from the design value. To the students, this seemed to contradict common sense, and thus the demonstration failed to provide motivation for understanding the effectiveness of a T-test. The demonstration also took longer to complete and was more difficult to evaluate than the previous demonstrations, and thus failed to satisfy the criteria for effective learning. An alternative experiment involved measuring and comparing the flexural strength of strips of different metals. This lab was also not fully effective, as the mean flexural strengths of the materials used were quite different, and the need for a test of hypothesis was unclear. For 2017, either a variant of the vibrations experiment or a new experiment will be developed.

Results

The evolution of MECE-251 over the last few years tends to make interpretations of quantitative measures such as quiz scores and student course ratings more difficult. The experiments, the quizzes, and the very nature of the course has not remained constant. However, combined with a few observations about the course changes, this data can provide some insight into the value of the demonstrations.

Students have been quizzed on similar material for several years, but the specific quizzes have changed from year to year. Prior to 2014, when the student demonstrations were initiated, average quiz scores varied significantly from year to year. In early years, there were experiments each week and students struggled to keep up with even learning the statistics principles by rote. The higher quiz scores in 2011 and 2012 are likely due to reduces expectations and less rigor. Quiz scores have been consistent since 2014, when the student demonstrations were introduced. In 2014 and 2016, the median score was well above the mean score, suggesting that most students did quite well, and a few performed poorly. While it is difficult to draw any definite conclusions from the quiz data, it is consistent with the notion that the demonstrations helped most students to do very well, but the few remaining students simply did not understand the concepts.

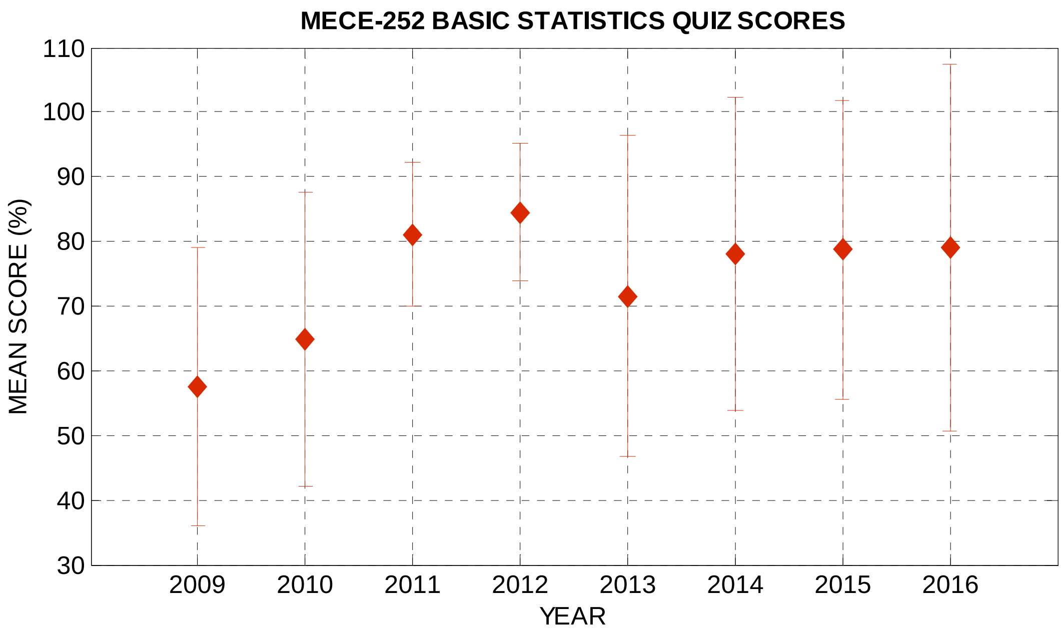

Figure 2 Basic Statistics quiz scores over time

End-of-semester student evaluations, while increasing in the last few years, also provide ambiguous information. It should be noted that the student demonstrations described in this paper do not span the entire semester and the final labs will influence the course evaluations. In early years, the scores were quite low, as low as 4.5 out of a possible 7 points in 2009. A local peak occurred in 2011 that does not have an obvious explanation. Starting in 2014, with the student demonstrations, average course evaluations have steadily climbed. Although not definitive, it is hoped that the demonstrations contributed to this trend.

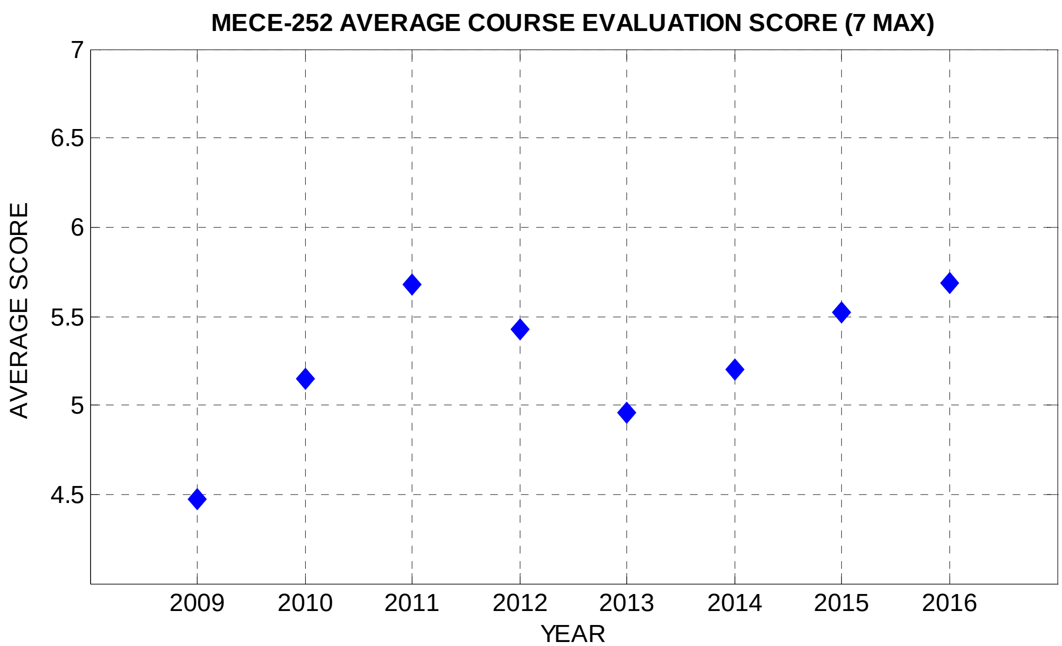

Figure 3 Average course student evaluations per year

Qualitatively, students appear to understand the basic statistical concepts more readily with the student demonstrations. Their motivation is higher, and they get two weeks to absorb the information and use it, rather than one year as in the past. Students design and implement an experiment of their own during the final three weeks of each semester. It is here that the benefits of the student demos are most clearly seen. Students are better able to select an appropriate analysis, correctly use the methods, and correctly interpret the results.

Conclusion

When students must demonstrate an understanding of basic statistical methods by using data to make predictions, they are more motivated, interested and engaged. The student demonstrations also provide an example of making engineering decisions based on data. Students are forced to make trade-off decisions, and find that statistics provides a means of maximizing their benefit. Although some students are initially threatened by the notion that their grade is at least partially determined by chance, but most become comfortable with the concepts and methods, and are able to apply them successfully to a student-designed experiment. were graded on the success of their predictions, required to make prediction

References

1 Korolkiewicz, Małgorzata Wiktoria1, Chiera, Belinda Ann, The Golden arches : An approach to teaching statistics in a first-year University service course, Springer Proceedings in Mathematics and Statistics, v 81, p 149-165, 2014.

2 Siepermann, Markus, Teaching statistics and risk management with card games, Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), v 9599, p 171-180, 2016, Games and Learning Alliance - 4th International Conference, GALA 2015.

3 Bridges, David H, Instruction in experimental methods: What Should we be teaching in laboratory courses?, 48th AIAA Aerospace Sciences Meeting Including the New Horizons Forum and Aerospace Exposition, Article number: 2010-0079, 2010.

4 Anthony, Richard J., Sondergaard, Rolf, What is needed in Experimental Methods Instruction - Perspectives from a government laboratory, 46th AIAA Aerospace Sciences Meeting and Exhibit, 2008, 46th AIAA Aerospace Sciences Meeting and Exhibit, Article number: 2008-0705; January 2008

Mark Archibald

Mark Archibald is a professor of mechanical engineering at Grove City College.