DEVELOPING PROJECTS BASED ON STUDENTS' DATA IN INTRODUCTORY STATISTICS

Aklilu Zeleke
Michigan State University, United States
Carl Lee and John Daniels
Central Michigan University, United States
zeleke@msu.edu

Over the last decade, the use of real world projects in introductory statistics courses has increased in popularity. Real world projects provide students with an opportunity to learn the entire process of a statistical investigation. Such projects fit the principles of active learning well. However, due to the time and effort required by both the instructor and students, it is difficult to sustain the project activity for a long time period. Hence, the final project reports are often disappointing. Through an NSF funded project, we have constructed a real-time online database. Students collect their own data and enter it into the database. Various activities are now available at http://stat.cst.cmich.edu/statact/. We assign group projects using the data collected by the students themselves. This paper shares how the process of statistical investigation is implemented into the project by using the students' own data.

INTRODUCTION

The use of projects in an introductory statistics has gained a great deal of attention in the recent statistics reform movement. Many statistics educators consider the use of projects as an authentic assessment tool for assessing student's understanding of the 'big picture' of statistical concepts (GAISE, 2005).

The use of projects in introductory statistics has been reported in many literatures. For instance, Fillebrown (1994) reported her experience of using projects in an elementary statistics course for non-science majors. Binnie (2002) discusses how he uses projects to encourage statistical thinking. Vaughan (2003) talked about his experience of teaching statistical concepts with student specific datasets. The unique feature in Vaughan's approach is to have students analyze their heating bills. Short and Pigeon (1998) talked about the planning and pilot at the stage of data collection for projects. Love (2000) discussed assessment of student projects by having students judging other students' projects. Garfield, Chance and delMas (2005) provide a through guideline for conducting projects and present a variety of sample projects in their WEBARTIST NSF project website (https://ore.gen.umn.edu/artist//index.html). The emphasis on statistical thinking in industry is another important reason for using projects in statistics courses (Kettering, 1995; Hoerl and Snee, 2002).

PROJECTS USING REAL-TIME ONLINE ACTIVITY DATABASE

Many types of projects have been proposed in the literature. They range from very structured small projects with cleaned data provided by the instructor to very unstructured large scale research projects that require a team of students to design their own study and collect their own data. The data sources for most projects are either provided by the instructors, by asking students to find real world data from existing sources such as internet, media and so on, or by asking students to design their own research studies that are related to students' discipline or are of interest to students themselves. In this article, we present yet another type of projects used in our classes in the previous year. The data source for our projects is from a real-time online activity database. Students conduct the activities, collect their own data, and enter their data into the online database. Data are collected from different classes in different institutions and are added into the database. Currently, we have developed ten activities located at http://stat.cst.cmich.edu/statact/. These activities have been used in classes, as homework, and as projects. This article reports on how we design our projects for our classes and shares our experiences of using the real-time data as student projects in three different introductory statistics courses.

BACKGROUND

The class taught by Zeleke was at Alma College (AC), a private liberal arts college in Michigan. The majority of students plan to go for advance degrees in medical or law schools. Most students are at the first or second year level. The classes taught by Daniels and Lee are at Central Michigan University (CMU), a public research-intensive regional university also in Michigan. The majority of students in their classes are from the College of Business Administration. Most students are at the second or third year level. The courses are taught using the real time hands-on activities whenever possible. The activity that was used to design the student project is the question**:** "Does one minute of exercise increase pulse rates dramatically?' The following is a typical process for conducting a hands-on activity:

(1) The instructor introduces the activity, facilitates the discussion on how to measure and collect data, and leads students to collect their own data. For the 'One minute exercise' activity, students are asked to discuss how to measure pulse rates and the potential measurement errors. The instructor turns on the music provided on the data entry site and leads students to actually do one minute of aerobic exercise. The pulse rates are taken before and right after the exercise.

(2) The instructor directs the students to the online data collection site for data entry. For the 'One minute exercise activity,' the data collected from students are Gender, Weight, Height, Before-exercise pulse rate and After-exercise pulse rate. The BMI (body mass index) is automatically computed and also recorded.

(3) Data can then be downloaded either from the entire database, the very own class or a random sample from the database. The data can then be used for various purposes to engage and enhance student learning either within class or outside of the class.

THE DESIGN OF THE PROJECT

After the data collection ends, students are asked to submit two written reports. Their reports should provide answers to questions like "Is there a dramatic change in pulse rate after one minute of exercise?", "Is there a difference between male and female pulse rate changes after one minute of exercise?" and "What is the relationship between the pulse rate change and an individual's Body Mass Index(BMI)?" To help them write the report the following case scenario was established. "*These days, rising health insurance costs are big concerns for many employers. As a result many companies encourage employees to follow a healthy life style as one measure of cutting health related costs. In our days, heart related diseases are among the deadliest and most expensive to treat. It is believed that regular exercise is one way of reducing the risk of having a heart related disease. The goal of this project is to find out if exercise and risk factors for heart disease have some relation."*

In the first (midterm) report students use graphical and descriptive techniques to analyze their data and find answers to the above questions. In their final report they must use inferential statistics.

THE ASSESSMENT OF THE PROJECT

Student Groupings

We observe that in general, most students are excited about working on projects. Alma College (AC) students worked in groups of three. There was no major objection to working in groups for these students. One possible explanation might be that Alma students are residential college students. In addition, most of these students know each other from taking many classes together and have a similar background (99% are Whites, and the majority come from rural Michigan). At Central Michigan University (CMU) it was recommended to form a team of two to three students. About 30% of these students, however, chose to work independently on the projects. The students working alone scored slightly higher on their overall grade (both paper and course), but the results were not statistically significant. Some reasons given against working in groups are time management, lack of familiarity with other students, and more confidence in working alone.

A Brief Summary of Project Assessment

The papers in each class were graded by the course instructor. All three instructors used the same rubric. Students were given this rubric when the project was assigned. Each project report was worth of about 10% of the final grade. This is a significant amount to get students paying attention to the project. The students' projects were graded on five categories. Each category was graded out of five points. Table 1 gives a summary of the assessment for each class. The first number is the mean (μ) and the second number is the standard deviation (σ) of responses.

Table 1

	Carl Lee's Class (18 Groups)	John Daniels' Class(16 Groups)	Aklilu Zeleke's Class (10 Groups)
Purpose	$\mu = 4.444, \sigma = 0.984$	$\mu = 4.375, \sigma = 0.806$	$\mu = 4.1, \sigma = 1.1$
Method	$\mu = 3.222, \sigma = 1.003$	$\mu = 4.313, \sigma = 1.015$	$\mu = 3.7, \sigma = 0.949$
Results	$\mu = 3.444, \sigma = 1.149$	$\mu = 3.938, \sigma = 0.772$	$\mu = 3.1, \sigma = 0.876$
Discussion	$\mu = 4.111, \sigma = 0.963$	$\mu = 4.000, \sigma = 0.895$	$\mu = 3.8, \sigma = 0.789$
Overall	$\mu = 3.333, \sigma = 0.840$	$\mu = 4.188, \sigma = 0.359$	$\mu = 3.675, \sigma = 0.667$

Students from both institutions have done well on category 1(Purpose). The lowest scores were in category 3 (Results) for Daniels' and Zeleke's classes while 2 (Method) was the lowest for Lee's class. Highest standard deviation for CMU is in category 2 while for AC it is category 1. Finally, for both institutions, we observe that grading categories had no statistically significant correlation between them. A student's success in one section, particularly in the Purpose and Methods sections, did not guarantee success in another section. Again, this reemphasizes the point that many students are adept at memorizing and regurgitating information but are not particularly effective when independent thinking is required. Since each paper was only graded by the instructor who taught the corresponding section, there exists grading differences among the three instructors. The grading shown in these tables reflects the instructor's differences. However, for all the three instructors, the lowest scores are found in the Results or Method category.

DISCUSSION

Areas of Difficulties Students Encountered

The project described in this paper uses data about students and gathered by students. The process simulates the typical statistical investigation process used in real world projects without overwhelming students, and at the same time, provides opportunities for students to experience every aspect of a real world project. We think this is a unique feature compared to other types of projects. However, the difficulties that students encountered are similar to those based on other types of projects. Hence, the findings could also be helpful for instructors who use other types of projects for their classes. In the following we summarize our assessment of the difficulties students showed in their reports. This assessment has provided a useful insight for improving our teaching in the future.

A) Poor report organization and improper quoting of references

Although the report layout was provided to students ahead of time, some students did not follow the layout. Two problems are particularly troublesome. One is the quoting of references. Only a few students were able to properly quote references and list the source of references. The other is writing a report similar to doing homework. Computer output is copied and pasted into the report without any interpretation based on the context. The variable names used in the output are not revised or explained.

B) Misuse of terminology

This appears to be a common problem among most students. We encountered many students making statements such as:

'It has been proven that being physically active reduces your chance of heart disease.'

'The data I collected 'PROVES' that this acceleration impulse isn't very significant for the average person.'

'*Our hypothesis proves to be true.*'

'*There appears to be no correlation between the gender of the student and his/her pulse rate change.'*

C) Misinterpretation of results

The set up of null and alternative hypothesis and their interpretation seem to be difficult for many students. Various misinterpretations are identified, such as:

'p-value (.002) falls into the 'do not reject region of the hypothesis test at alpha = 5%' a wrong conclusion.'

'The null hypothesis is 'Pulse rate change after one minute exercise is dramatic.'

'The claim is 'there is no significant difference between males and females in pulse rate change after one minute exercise.'

'After looking at the (a descriptive summary), I have concluded that for the first question my hypothesis will be that the exercise will not have a dramatic effect on the heart rate.'

D) Inappropriate Application of Concepts.

i. Right Chart / Wrong Data: An example of this mistake was one student's use of Total Pulse Difference vs. Gender on a Bar Chart.

ii. Right Data / Wrong Chart: Two of the most commonly observed mistakes were placing Pulse difference vs. BMI on a Bar Chart. Another common mistake was providing Pulse rates vs. Student Number on a time series plot.

iii. Inappropriate Use of Summary Statistics

The following are examples of inappropriate use of summary statistics.

'*Our class had a range of 60 for the before pulse and a range of 76 for the after pulse rate. Therefore, the difference after one minute of exercise is 16 BPM. On average we feel that is dramatic (see graph one).'*

'The difference in the pulse of people before and after one minute of exercise is almost a fourth of the average pulse rate after exercise and is almost a third of the pulse rate before exercise. A 33% increase in pulse rate seems substantial enough.'

E) Statements without any supporting evidence.

Very often, a student will make a statement or opinion without any supporting evidence developed in the Methods and Results sections. An example of this can be seen in the following statement:

'Is there a difference between the two [genders]? Yes, and I think there will always be a difference. The two genders will never have the same amount.'

Such strong opinions using words such as "always" and "never" certainly require supporting evidence. No evidence was provided to support this claim. In fact the group from which this statement was quoted was not able to do correct hypothesis testing.

F) Results out of Context.

Students very often state a result or conclusion without relating such a statement back to the Purpose section of the paper. The instructor is left thinking "So what does this mean?" An example of such a pairing of disconnected statements is as follows:

i. Purpose Section *: 'The purpose of this study is to find out whether there was a dramatic change in pulse rate after one minute of exercise by looking at a person's body mass index, their gender, their blood pressure, and their cholesterol levels.'*

ii. Results Section*: 'Thus, there is a slight difference between men and women in their pulse rate change after one minute of exercise. When comparing BMI to pulse rate change we found that there was nothing in common between the two.'*

Areas Where Students Did Well

At the other end of the spectrum, we wish to acknowledge those areas in which the student's showed exemplary skills.

(A) Use of the internet to obtain meaningful background information and context.

The students appear comfortable using the internet as a research tool. A large variety of websites and companion data sources were referenced and acknowledged.

(B) Some students appeared to use intuition, instead of computational mathematics, in the analysis of quantitative data. They were able to find significant results without statistical techniques. For example, one student wrote:

> "*After looking at it [bar chart of Pulse Diff vs. BMI] I have come to the conclusion that BMI doesn't have a significant affect [sic] on how your heart rate responds to exercise. The person with the smallest BMI actually had a higher increase in heart rate than the person with the highest BMI. If we think about how we get calculate the BMI we can see that it gives a good measure of generally how large or small someone is, but does not give us any more information than that.*"

This student did not need to know or interpret the correlation coefficient between Pulse Difference and BMI. The student was able to make this statement by a visual assessment of the scatter plot without relying on any quantitative techniques.

(C) Use of Computer Software: Most students were able to use Minitab fairly well to conduct the needed analysis.

(D) Most students were able to effectively use both descriptive and graphical displays to address the objectives of the projects.

LESSONS LEARNED

In this article, we present the use of projects in our introductory statistics courses, based on students' own data about pulse rate change after one minute of exercise. The unique features of this project approach include (1) students collect their own pulse rate data and then analyze their own data. (2) Data are collected through an online database. Students from different classes in different institutions can all participate in the same activity. (3) Sample size continues to grow. The same project can be used in different classes and different semesters based on different data sets. (4) Since the data is 'real,' data are not 'clean' as those we see in textbooks. This provides opportunities for the discussion of measurement and sampling errors. The following are some lessons that we learned from this student project.

(A) A successful project requires good planning and clear communication of expectations with students.

(B) Developing projects using a real-time online database allows for smoothly conducting the same projects in different classes and different institutions.

(C) Real-time online activities can be integrated into classroom activities, lab activities, homework and projects together. There is less stress for students in terms of looking for real world data outside the class.

(D) The assessment method of projects should be communicated with students with a clearly defined rubric grading system.

(E) We find that choosing either group or individual work does not affect student performance. Allowing students to choose their own team members or to work alone seems to work well.

(F) Using student-generated data seems to increase interest and enthusiasm.

ACKNOWLEDGEMENT

The authors are grateful for the support by NSF under the NSF/CCLI grant # 0310932.

REFERENCES

Binnie, N. (2002). Using projects to encourage statistical thinking. In B. Phillips (Ed.), *Proceedings of the Sixth International Conference on Teaching of Statistics*, Cape Town. Voorburg, The Netherlands: International Statistical Institute.

Fillebrown, S. (1994). Using projects in an elementary statistics course for non-science majors. *Journal of Statistics Education*, 2(2).

GAISE College Report. (2005). Guidelines for assessment and instruction in statistics education. The American Statistical Association, http://www.amstat.org/education/gaise/.

Garfield, J., Chance, B., and delMas, R. (2005). WEBARTIST Project. https://ore.gen.umn.edu/artist//projects.html.

Hoerl, R. and Snee, R. (2002). *Statistical Thinking: Improving Business Performance.* Duxbury.
Kettering, J. (1995). What industry needs. *The American Statistician*, 49(1), 2-4.
Love. T. (2000). A different approach to project assessment. *Journal of Statistics Education*, 8(1).
Real-Time Online Activity Database. http://stat.cst.cmich.edu/statact/.
Short, T. and Pigeon, J. (1998). Protocols and pilot studies: Taking data collection projects seriously. *Journal of Statistics Education,* 6(1).
Vaughan, T. (2003). Teaching statistical concepts with student-specific datasets. *Journal of Statistics Education*, 11(1).