TEACHING STATISTICS USING A REAL TIME ONLINE DATABASE CREATED BY STUDENTS

Carl Lee and Felix Famoye
Central Michigan University, United States
carl.lee@cmich.edu

The advancement of computer technology creates unlimited opportunities for teaching and learning statistical concepts. A significant impact is the paradigm shift from a passive teaching-centered to an active learning-centered environment. Although one should not make a paradigm shift solely for the sake of technology, there is no doubt that technology will play a crucial role in this transformation. Research has suggested that meaningful learning takes place when students are actively involved in constructing knowledge themselves through their own experiences and active participation. This article proposes an active learning environment for introductory statistics courses using an online real-time database created by students. The experience of implementing the active learning activities using the real-time online database will be shared. Some strengths and weaknesses will be discussed.

INTRODUCTION

The modern society is becoming more and more quantitative. Data come from everywhere. Decision making can no longer be based on intuition in the modern global business. Decision making based on data is inevitable. Being able to conduct a proper statistical investigation and apply statistical thinking to analyze business problems is becoming a norm in industry. For many years, educators have been looking for and developing innovative approach to integrate more 'real world' data and lead students to experience the process of statistical investigation in introductory statistics (e.g., Moore, 1997). Among these innovative strategies, hands-on activity seems to be one of the most promising strategies that allow students to experience the process of statistical investigation. The issue of technology seems to be less a problem due to the fact that advanced technology is becoming easier to use and young students live in the modern technology era. They have less fear for new technology.

The purpose of this article is to introduce a real-time online database for hosting data collected by students. The paper will address 'what is,' 'how to,' and to share our experience on the use of real-time hands-on activities for teaching statistical concepts in an introductory statistics course.

WHAT IS THE REAL-TIME ONLINE DATABASE?

- *Do you use hands-on activities in your class?*
- *Would you like to share your hands-on activities?*
- *Would you be interested in using data collected by students from different classes at different institutions?*
- *Would you be interested in sharing your students' data with others?*
- *Does it take more class time than you would like to spend on hands-on activities?*
- *Do you have to enter the hands-on activity data yourself after the class period?*

If your answer to any of the above questions is "YES," then the Real-Time Online Database approach should be beneficial to your class. Figure 1 lists the real-time online hands-on activities that are currently available on the web site http://stat.cst.cmich.edu/statact/.

Each activity is a hands-on activity. The instructor facilitates the activity and guides students through the process of a statistical investigation that involves: Defining the problem, measuring the variables, collecting the data, cleansing and manipulating the data, analyzing, interpreting and reporting the results. The importance of training students to go through the entire process of statistical investigation has been documented in many papers. For example, Hoerl and Snee (2002) and Kettering (1995) talked about the use of statistical thinking and the importance of using real world projects in statistics courses.

Activity	Title
Distance	How far are you away from home?
Hand_Size	Is hand size a good predictor of height?
Exercise	How does one minute of exercise affect your pulse rate?
Raisins	How many raisins are in a 1/2 oz box?
College_Life	Are you satisfied with your university?
Left_Right_Hand	Are you left handed or right handed?
Sampling	Random sampling vs. subjective sampling. Which is better?
Vote	Which party will you vote for?
Draw_Line	Can you draw a straight line without a ruler?
Random Selection	Which numbers are popular?

Figure 1: Real-time Online Hands-on Activities

The process of conducting a real-time online activity is similar to conducting hands-on activities proposed by statistics educators and practitioners. The difference is the use of online database to capture the data entered by each individual student during the class period. The following is a typical process of facilitating a real-time online activity.

1. Choose an activity and present the problem scenario to students.
2. Guide the discussion on the possible variables of interest and how to measure and collect the data with students.
3. Guide students to enter their own data into the online database.
4. Guide students to extract data (either all, or subset) from the database, and to use a designated software (any statistical software available in your institution) to input the data.
5. Lead students to analyze the data and discuss various statistical concepts based on the results.
6. Each activity can either be conducted in class, assigned as homework or proposed as a group project.

Traditionally, a hands-on activity requires students to collect their own data and have either the instructor or each individual student enter the data. This process either takes place during or after the class. It is often difficult to go through the data collection and analysis without interruption.

With the online database, each individual student only needs to enter his/her own data. The sample size increases over time. The extraction of the data is done immediately after the data entry. The discussion about data measurement, outliers, and sampling errors using the 'real' data generated from students themselves, which are usually not available in textbooks, can be facilitated in class without interruption. This process mimics the entire process of solving a real world problem in a short time period in class. The online database is available for any instructor at different institutions. The only requirement is an access to Internet.

WHY REAL-TIME ONLINE ACTIVITIES?

Technology has become essential in the teaching and practice of statistics. Recent studies show that an activity-based learning strategy enhances students' conceptual understanding of statistics and mathematics (Smith, 1998; Brush, 1997), provides students with opportunities to work as a group, and sharpens their communication skills (Magel, 1998). It is also found to be especially beneficial to the less-prepared students in mathematics (Giraud, 1997). In addition, Schroeder (1993) pointed out that most students have learning styles that are different from those of the instructor. An active learning environment can actively engage students with different learning styles (Garfield, 1993).

Applying real world data, creating and facilitating active learning environment in an introductory statistics course have received a great deal of attention during the recent decade of statistics education reform. The most recent effort to promote the reform of statistics education is the GAISE Report (2005), which recommends: (1) Emphasize statistical literacy and develop

statistical thinking; (2) Use real data; (3) Stress conceptual understanding rather than mere knowledge of procedures; (4) Foster active learning in the classroom; (5) Use technology for developing conceptual understanding and analyzing data; (6) Use assessments to improve and evaluate student learning. The emphasis of statistical thinking in industry is another important reason for using real world data and projects in statistics courses (Kettering, 1995; Hoerl and Snee, 2002).

The use of the student's own data is also recommended by the GAISE report. Many instructors have reported their experience of using this type of data and indicated that its source is closely related to the student's own interest and that it is easier to engage students in the process of learning. For instance, Romero *et al*. (1995) used a personal survey filled in by the students on the first day of class. Gnanadesikan *et al*. (1997) developed a collection of hands-on activities to engage students in active learning. Gnanadesikan *et al*. (1997) advocated that "such activities promote the teaching of statistics more as an experimental science and less as a traditional course in mathematics. An activity-based approach enhances learning by improving the students' attention, motivation, and understanding." Many hands-on activities can be found in the *Activity-based Statistics* book (Scheaffer *et al*., 2004). *Workshop Statistics* by Rossman and Chance (2003) is another useful resource for an activity-based introductory statistics. One can find many activities from the *Journal of Statistics Education* at http://www.amstat.org/jse/.

HOW TO CONDUCT A REAL-TIME ONLINE ACTIVITY

There are ten real-time online hands-on activities currently available. The information for each activity includes 'Data Entry,' 'Data Download,' 'Variable Description,' and 'Online Assessment.' Each activity is developed based on a problem scenario, which is designed to accomplish a set of learning objectives. Different activities may have overlapping learning objectives for the purpose of reinforcement. Figure 2 summarizes the process of facilitating an activity. There are two ways to participate:

1. Have students conduct the real-time online hands-on activities. This requires a registration by the instructor to request an 'ACTIVITY CODE' to be used in the 'Data Entry.' The data entered will be stored in the online database.
2. Use the available data in the real-time online database. This does not need any registration. An instructor can guide students to the 'Download' page to download the data generated by students in other classes.

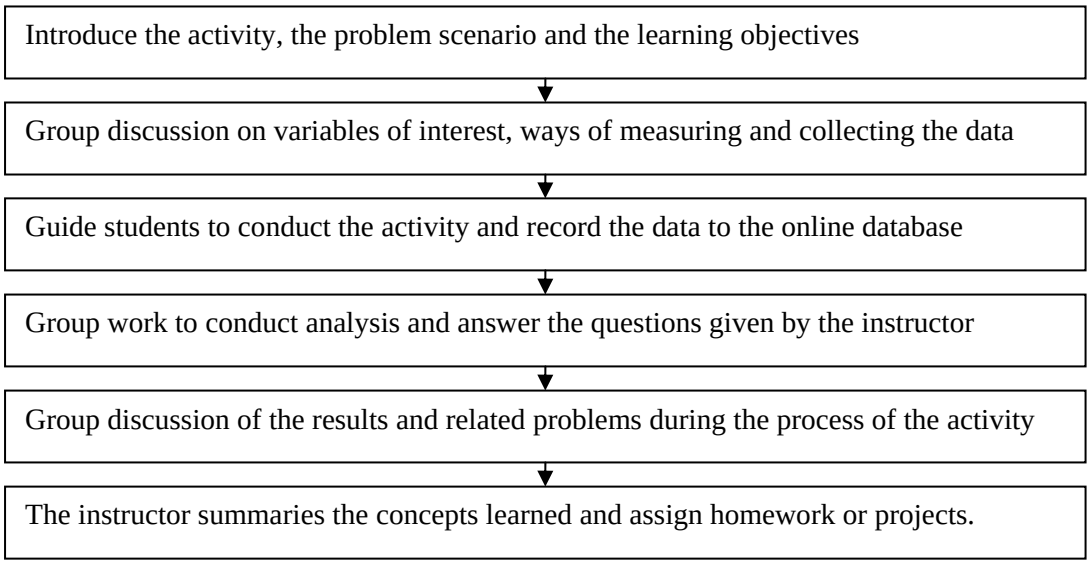

Figure 2: Guidelines for facilitating a real-time online activity

In the previous year, we have taught two different introductory statistics courses. One is mainly for business school students and one is for science and technology students. In addition to using the real-time hands-on activities, we also use a textbook for each course. Different

textbooks are used in these courses. Some instructors use more textbook materials, and some use mainly the real-time hands-on activities for instruction. Whenever the activity is used for instruction, the instructors followed the above guideline.

The activity 'How well can your hand size predict your height?' is used to demonstrate how a hands-on activity is conducted in a classroom setting. This activity is used for introducing correlation and simple linear regression. The subjects for group discussion include:

1. How do you measure hand size? This usually leads to discussions related to measuring devices and measurement errors. The instructor emphasizes the importance of finding a measurement that is repeatable with less variation and robust to the measuring device. The online survey captures 'Hand Length,' the distance from the wrist to the top of the middle finger; and 'Hand Width,' the distance between the top of the thumb and the little finger when the hand is completely stretched out. The 'Hand Length' can be measured repeatedly with very small measurement errors, while the 'Hand Width' may have large variation depending on how much you stretch out the hand. Students learn the importance of reducing measurement errors from this simple activity.
2. Which one, the 'Hand Length' or 'Hand Width' will have larger variation? Students are asked to think about the consequence of measuring system on the actual measurement. Students are asked to discuss and present their practical guesses with their reasons, which will be justified at the analysis stage. This discussion gives students some flavour of setting 'hypothesis' without introducing the formal terminology.
3. Which one, the 'Hand Length' or 'Hand Width' is a better predictor of height? Again, students are asked to state their 'practical guesses,' which will be justified by the data.

In the analysis stage, students are asked to use graphical techniques as well as numerical summaries to answer a list of questions. The same activity is used later for estimation, hypothesis testing and inference about the regression line. We have developed a set of assessment questions based on each activity. These questions will be available online.

HOW THE REAL-TIME ACTIVITIES ARE USED FOR GROUP PROJECTS

Projects are perhaps the best assessment tool that can effectively take into account the six recommendations by the GAISE report (2005). Many types of projects have been proposed in the literature. The types of projects 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. With the real-time database, it is a natural fit to assign group projects using the student generated data. An example of the project is described in the following.

> Problem Scenario: 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. Based on the data collected in class about pulse rate and BMI, your group should study and write a report that addresses the following issues:
>
> a) Is there a dramatic change in pulse rate after one minute of exercises?
> b) Is there a difference between male and female pulse rate changes after one minute of exercise?
> c) What is the relationship between the pulse rate change and an individual's BMI?
>
> Project Layout: Your project report should have four main sections. Any reference quoted must be given in the Reference section.
>
> *Section 1. Purpose of the Study:* In this section, you should describe the following items: (a) Search the Internet, for example the American Heart Association for information on the effect of exercise on health related issues. Some factors that you should look into are pulse rate change, body mass index (BMI), blood pressure and cholesterol levels. (b) Describe the purpose of the study. (c) Describe your hypothesis on the effect of one minute exercise.

Section 2. Method: In this section, you should (a) discuss the sampling method used in collecting your pulse rate data and some potential drawbacks of this sampling method, (b) describe the data analysis techniques you apply.

Section 3. Results: In this section, you should present the evidence from your analysis to address each of the issues given in the problem scenario. Tables and graphs should be created as part of your report, not simply attach the computer output. Appropriate interpretations of the results based on the context are very critical. It is important to describe your findings in a way that people who do not know statistics will understand your results.

Section 4. Discussions: What did you learn from this project? How can the techniques used to analyze this data be used in other cases? Give another example!

Section 5. References

Students are required to submit two reports. The first report (the midterm project) addresses issues raised in the problem scenario using only descriptive and graphical techniques. The final report will include inferential statistics. A rubric is also provided to students so that they are aware of how the project will be assessed.

DISCUSSION

The real-time online activity database that hosts data generated by students is designed to streamline the process of conducting in-class hands-on activities. Instructors no longer need to enter the data themselves. This saves class time to engage students in the process of statistical investigation by having students to work on their own 'real data.' Data generated from students themselves tend to draw their attention and motivate their interest more than a dataset that is disconnected with their everyday life. In addition, the fact that the data continue to grow provides unlimited use of the same activity. The data will not be the same from semester to semester.

Real-time databases also provide many opportunities to discuss the issues related to data measurement, data collection, sampling errors, outliers, missing data, and the types of problems we encounter in real world projects. The real-time database globalizes the locally conducted hands-on activities. Different classes from different institutions can share their locally collected data with one another.

The use of the real-time online activity for teaching statistics naturally leads to the creation of an active learning environment, since students are the central part of the activity. The data are about students and are collected by students themselves. Students are engaged in every step of a statistical investigation. The instructor in this active learning environment is a facilitator who leads and guides students through the process of a statistical investigation. Instructors can use the data for classroom activities as well as for homework or projects.

Assessment is a critical activity to reinforce student learning. A list of suggested homework problems and projects are given on the site as well. By clicking on "Activity HW/Proj," one will find a list of suggested questions for each activity. These questions or projects are not topic-oriented. They are problem-solving based questions. Therefore, one should choose the proper level of questions or projects for his/her class from the list of suggested questions.

Although there are advantages using the real-time online activities in an introductory statistics course, based on our limited evaluation from individual classes, there are a few issues that should be considered. These include: (1) The real-time online activities rely on technology, especially the Internet. On some occasions, the Internet connection may be down. This creates frustrations for instructors and students. (2) Although we did not find a problem of covering similar amount of material in our classes due to the limitation of the classroom time, one has to be selective in the choice of topics to be covered. In our experience, some times we have to reduce topics in probability and discrete distributions. (3) There are always some students who need additional help on basic computer skills. A careful plan at the beginning of the semester to speed up the familiarity of computer for these students is important. (4) The unstructured active learning

environment is more chaotic. Our experience indicates that a guided facilitation to engage students and a summary of concepts learned at the end of the activity are very important.

"Did students from the real-time database class perform better?," "Did students have more positive attitudes about statistics?" These are some questions that have to be addressed. We conducted an interview study in Spring of 2005. We are now editing the videos and beginning to analyze the interview data. We are presently collecting data to assess the content knowledge learned and students' attitudes using pre/post approach. These assessment data will be reported in the near future.

ACKNOWLEDGEMENTS

We gratefully acknowledge the support of this project by the NSF/CCLI grant #0310932 and a matching grant from Central Michigan University. We thank the session chair Dr. Juana Sanchez for many helpful comments that have helped greatly to improve the article.

REFERENCES

Brush, T. A. (1997). The effects on student achievement and attitudes when using integrated learning systems with cooperative pairs. *Educational Technology Research Development, 45*, (1), 51-64.

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

Garfield, J (1993). Teaching Statistics Using Small-Group Cooperative Learning. *Journal of Statistics Education*, 1(1), www.amstat.org/publications/jse/v5n3/garfield.html

Giraud, G. (1997). Cooperative learning and statistics instruction. *Journal of Statistics Education*, 5(3), www.amstat.org/publications/jse/v5n3/giraud.html

Gnanadesikan, M., Scheaffer, R., Watkins, A. and Witmer, J. (1997). An Activity-Based Statistics Course. *Journal of Statistics Education*, 5(2), www.amstat.org/publications/jse/v5n2/gnanadesikan.html

Hoerl, R. and Snee, R. (2002). *Statistical Thinking: Improving Business Performance*. Belmont, CA: Duxbury/Thomson Learning.

Kettering, J. (1995). What industry needs. *American Statistician*, 49(1), 2-4.

Magel, R. (1998). Using cooperative learning in a large introductory statistics class. *Journal of Statistics Education*, 6(3), www.amstat.org/publications/jse/v5n3/magel.html

Moore, D. (1997). New pedagogy and new content: the case of statistics. *International Statistical Review*, 65, 123-165.

Romero, R., Ferrer, A., Capilla, C. and Zurica, L. (1995). Teaching statistics to engineers: An innovative pedagogical experience. *Journal of Statistics Education*, 3(1), www.amstat.org/publications/jse/v5n3/romero.html.

Rossman, A. J. and Chance, B. L. (2003). *Workshop Statistics.* Emeryville. CA: Key College Publishing.

Scheaffer, R.L., Watkins, A., Witmer, J. F., Gnanadesikan, M. and Erickson, T. (2004). *Activity-Based Statistics* (2nd Edition). Emeryville, CA: Key College Publishing.

Schroeder, C. (1993). New students – New learning styles. *Change Magazine*, 25(5), 21-26.

Smith, G. (1998). Learning statistics by doing statistics. *Journal of Statistics Education*, 6(3), www.amstat.org/publications/jse/v5n3/smith.html