Technology, Projects, and Elementary Statistics

T. W. Judson
Department of Mathematics and Computer Science
University of Portland
judson@up.edu

Abstract

Students can use technology to complete projects that unify the ideas and concepts taught in an elementary statistics course. Courses in elementary statistics enjoy a wide clientele, including students from the natural sciences, education, the humanities, and the social sciences. In addition to those students fulfilling requirements for their major field of study, many students elect to study statistics as a means of satisfying university core curriculum requirements for mathematics. Students who are majoring in disciplines such as biology or the social sciences often take more advanced courses that build on what they have encountered in a first course. On the other hand, elementary statistics is often the only university level mathematics course that many students will take. It is often the case that students in an elementary statistics course have a weak background in mathematics with little or no knowledge of calculus. Software such as Minitab or Excel can be employed as a pedagogical tool to aid the weaker student yet allow the better prepared student the opportunity to undertake more advanced projects.

1 A Course in Elementary Statistics

Students studying elementary statistics come from a variety of disciplines, including biology, education, the humanities, and the social sciences. Some students enroll in a statistics course to fulfill a requirement for their major field of study, and those who are majoring in disciplines such as biology and the social sciences often take more advanced courses that build on what they have learned. Other students enroll in elementary statistics in order to satisfy a university mathematics requirement. Elementary statistics is often the only university-level mathematics course that this latter group of students will experience during their university career. For these reasons elementary statistics is as important as the university calculus classes that we teach.

The Mathematical Association of America makes recommendations for university mathematics curriculum in the United States through the Committee on the Undergraduate Program in Mathematics (CUPM). The CUPM suggests that a non-calculus based statistics course include descriptive statistics, a brief exposure to probability, sampling techniques and design of statistical experiments, regression, statistical inference, and hypothesis testing. When teaching random sampling and experimental design, students should gain enough experience selecting random samples and setting up experiments to make the concept of a sampling distribution easy to grasp. Students also need experience organizing and describing data with table and graphs. They should learn about univariate descriptive statistics and understand the meaning of mean, median, mode, variance and standard deviation, and percentiles. Students should learn about bivariate descriptive statistics such as correlation and fitting lines by least squares. It is important for students to understand confidence intervals, one- and two-sample hypothesis testing, and the use of P-values to assess evidence against null hypothesis. Chi-square tests and inference in simple linear regression should also be introduced.

Probability is a necessary topic in any statistics course; however, due to the limited amount of time in a one-term statistics course, there is a danger of including too much probability. It is necessary to keep reminding oneself what topics from probability will be needed to learn elementary statistics. Probability topics covered in a one-term course should include general probability, random variables, some standard distributions, and the Central Limit Theorem [1].

2 Using Technology to Teach Statistics

The feasibility of completing large statistical projects was severely limited before the days of inexpensive computing. Studies involving large data sets were rare, and analysis was usually limited to relatively simple models. Projects involving data sets consisting of more than 1000 observations were very unusual. The situation has changed dramatically in the last few years. Now, even hand-held calculators can deal with large amounts of data, and serious studies can be done with no more than the aid of a desktop computer. In addition to doing lengthy calculations, technology makes it easy to do accurate graphs and plots [2].

A variety of technologies can be used to teach elementary statistics; however, any choice of technology must satisfy certain criteria. First, the technology must be easy to learn for both student and teacher, yet it must be powerful enough to handle reasonably large data sets and able to do a reasonable number of statistical calculations such as computing sample means and standard deviations, correlations, and regression. The capacity to perform a wide variety of statistical tests is essential. One must be able to create basic graphs such as box plots, scatter plots, histograms, and residual plots. Finally, it should be easy to enter, store, and maintain data. Although we use Minitab to teach elementary statistics at the University of Portland, other statistical software such as JMP, SPSS, or SYSTAT work well. Statistical software is periodically reviewed in the *American Statistician*. One can also use spreadsheets such as Excel or calculators such as the Texas Instruments TI–83; however, care must be taken if one chooses to use a spreadsheet, since spreadsheets were not designed to do statistics. For example, Excel is not able to handle missing values like Minitab. Hand-held technology is often more convenient, but data is easy to store, transfer, and print, if one uses computer software such as Minitab. In any case, technology must be reasonably priced and easy to maintain [2, 3, 4, 5].

Using a computer to do statistical calculations on a data set is only one application of technology in an elementary statistics course. Presenting even a small data set to students on the blackboard is an inconvenient and time-consuming process. Too often students are busy copying the complete data set into their notes and miss the point of the example. This process is made easier if data can be presented electronically and students can access the data sets on disk, over a local area network, or over the Internet. Ideally, the data sets need to be prepared in advance and be available to students both before and after lectures.

3 Projects

Many students in an elementary statistics course are less than enthusiastic. They view mathematics courses as theoretical, unrelated to real life, and uninteresting. Even good students sometimes study statistics by memorizing algorithms and facts that they quickly forget after the final examination. Assigning projects can help address these problems. The template approach to problem solving is less prevalent by the very nature of projects, and students can apply what they have learned in an elementary statistics course to a problem of their own creation. Many types of problems are suitable for student investigations. Past projects have covered a very wide range of topics.

- Do students living in dormitories get more sleep than those students living at home or in apartments do?
- Which students use the library the most?
- Do consumers prefer generic brands of breakfast cereal to name brands?
- Is there a relationship between students' grade point averages and alcohol use?
- Do first-year players in the National Basketball Association have a better free throw average at home than at away games?
- Is there a relationship between the number of Olympic medals won by a country during the summer Olympics and the country's population or per capita income?

Projects present challenges to both the students and the instructor, but technology has removed the burden of tedious computation. It is important to spend one or two class periods early in the semester showing students how to use software, but generally they learn quickly. Usually, a short in-class demonstration is all that is needed to introduce a new feature of the software after the initial learning period has passed.

Team projects offer some advantages over individual projects. Teams can accomplish larger projects, students learn to work in groups, and the burden on the instructor of managing and grading projects is reduced. However, team projects require careful management by the instructor. It is essential to set intermediate deadlines and monitor the progress of each group carefully; otherwise, many teams will begin their projects a few days before the due date. I organize students into teams of three or four individuals, and each group is responsible for posing their own problem. Teams are directed to choose a question that they find interesting and that can be solved by the collection and analysis of data. If left to their own devices, some teams will begin the project by collecting data, not by posing a problem. In such cases, teams will often collect unneeded and irrelevant data and fail to obtain the data that they really want. Students should pose the problem as precisely as possible from the outset. I generally allow the teams considerable freedom, but each team is required to submit a project proposal explaining the project, the population that they want to sample, how they will collect the data, and what type of model that they will use for data analysis.

I encourage projects where teams collect their own data, since this will help students better understand the ideas of sampling theory. However, a few groups choose to use data that has been collected by someone else for a different purpose. Good sources of data for projects of this type are reports published by various government agencies as well as the Internet. I encourage groups to consult other sources for information, but they are required to credit such sources in their final report.

I have students submit the following items at deadlines that are appropriately spaced throughout the term. Teams cannot move forward with the next phase of their project until they have my approval.

- *A list of group members.* I let students choose their own team members, and I assign those students who are not part of any group either to an existing team or form new teams. Surprisingly, this strategy has worked well. If one team member does not do his or her share of the work, it is usually not a problem since two students can reasonably complete most projects.

- *The preliminary report.* Teams submit their research question and a description of the methods that they will use to collect and analyze data. The preliminary report gives me an opportunity to review each project before the team begins collecting data. I direct students to measure at least two variables but caution them to measure no more than three. At this point in the course, they do not know enough statistics to set the sample size to obtain a specified margin of error, but I tell them to measure at least 30–50 subjects, even more if it is reasonably possible. I also request that teams provide enough detail about their sampling methods so that I could select their sample for them. If student groups require resources from a third party, I have them include a copy of their written request to that party. I do not allow students to send a request unless it is first approved by me. I meet with each group before I give final approval to the project.

- *A copy of the data set.* Students hand in their data as a printout from a Minitab file.

- *A draft of the final report.* I read the drafts to make suggestions and be certain that the teams are not encountering any last minute difficulties.

- *The final report.* I give very specific instructions on the format of the final report. Projects are not only graded on their results but on their exposition. Every final report must contain a title page signed by all project team members, an introduction, and sections explaining methods, results, a discussion of the results, and a conclusion. Reports should also include references and appendices containing figures and tables refereed to in the text.

- *An in-class presentation.* The last week of the term is devoted to teams presenting their projects to the class. Teams have fifteen minutes to do their presentations including a question and answer period. I invite other faculty members to attend the presentations.

One of the greatest difficulties of having students do projects is the protection of human subjects. If teams plan to use a survey in their project, they must provide a copy of the actual questionnaire that they plan to use. The survey form must include the names of the students conducting the survey as well as the purpose of the study. In most cases students only need to say that they are doing a project for elementary statistics. Project groups must take adequate steps to protect human subjects and cannot deceive participants or put them at risk in any way. They must provide a statement that participation is voluntary and that any responses are confidential and anonymous. If the information that is requested is sensitive, the team must provide an explanation as to the steps that they are taking to assure confidentiality and anonymity. Although most undergraduate projects are not subject to the university-wide human subjects process at the University of Portland, there are other universities that require such projects to have the approval of the faculty committee that oversees protection of human subjects in research.

4 Sample Student Projects

One student project investigated whether or not students at the University of Portland had a preference for name brand breakfast cereals over generic brands. Their investigation was motivated by numerous cereal commercials claiming that there was no difference in taste between name brand and generic brand breakfast cereals. The team obtained a representative sample from 30 individuals who ate breakfast in the university dining room. As students lined up for breakfast, the team asked every fifth individual in line to participate in the experiment. Participants sampled dry cereal and were not allowed to add sugar or any other additive. In a blind taste test half of the participants sampled the generic brand cereal first and half sampled the name brand cereal first. On the data table the team posted a statement detailing the purpose of the study and the rights of the participants, and each participant was required to sign their name beside the data acknowledging their compliance with the experiment.

Out of 30 people sampled, 14 preferred the generic brand of breakfast cereal ($\hat{p} = 0.47$). The team formulated the following hypothesis test:

$$\begin{array}{rcl} H_0: \; p & = & 0.5 \\ H_a: \; p & < & 0.5, \end{array}$$

where p is the proportion of people preferring the generic breakfast cereal and H_0 was the null hypothesis that there was no preference. They calculated a P-value of 0.37 and accepted the null hypothesis. Further analysis demonstrated that it was not statistically significant whether the participants sample the generic brand first or the name brand. The group did notice that the 90 percent confidence interval for p was large, $(0.32, 0.62)$, and wrote that they would increase the sample size if there was an opportunity for further study. Although it was unnecessary to use the full power of Minitab to perform the statistical analysis for their project, the team used Minitab to create graphs for their final report [6].

Another project examined how students use the university library facilities. More specifically, the team was interested to find out if the ratio of the number students from each school or college in the university (the College of Arts and Sciences, the School of Nursing, the School of Business, the School of Education, and the School of Engineering) using the library was the same as the ratio of the students from each college enrolled at the university. Similarly, the team wanted to determine if the ratio of students from each college class (freshman, sophomore, junior, and senior) using the library was the same as the ratio of students from each class enrolled at the university. In order to obtain a representative sample, every fifth person entering the library was interviewed and sampling was carried out over the course of one week at various times of the day and evening. In total, 110 people were interviewed.

Usually, students in an elementary statistics course learn only one method of analysis for more than two variables, the chi-squared goodness-of-fit test. The team used Minitab to calculate the chi-squared statistic for each population. They accepted the null hypothesis that the ratio of the students from each college class using the library is the same as the ratio of the students from each college class enrolled at the university ($P = 0.242$). However, the null hypothesis that the ratio of students form each college using the library is the same as the ratio of students from each college enrolled at the university was rejected ($P = 0.016$) [7].

A third project investigated the distributions of colors in a bag of M&M candy. The team selected a random sample of 20 bags of plain M&Ms, each of the same size. The students calculated confidence intervals for proportions of each of the different colors of the candy: red, orange, yellow, green, blue, and brown. This task would have been very time-consuming without the aid of technology. Then the team compared the results of their sample to the distribution of colors claimed by the Mars Candy Company, which they found on the company's World Wide Web. Although the group also used a chi-squared goodness-of-fit test to determine that the observed distribution agreed with the expected distribution of colors, they did a great deal of pairwise comparison. That is, they compared the means of red verses blue M&Ms, and so on. With the exception of brown and orange M&Ms, the team found the sample proportions agreed with the expected proportions. This project demonstrates the difficulties encountered when a group performs an investigation with more than two or three variables [8].

5 Comments and Conclusions

Projects provide a capstone experience for a course in elementary statistics; however, some careful planning is necessary to make the experience of doing projects worthwhile. First, choose appropriate technology and spend time early in the course introducing the technology. After the initial instruction period, use technology in the classroom frequently if possible and introduce new features of the technology on a regular basis. Set up intermediate deadlines and meet with each project team at least one or two times during the course. Be careful about the types of projects that you approve and encourage projects that will best illustrate the topics that you are covering in the course, especially projects that involve linear regression, hypothesis testing, comparing sample means and proportions, and chi-squared tests. Discourage projects that have two many variables. In my experience students are usually enthusiastic about doing projects, especially team projects.

References

[1] Lynn A. Steen, editor. *Reshaping College Mathematics*. Number 13 in MAA Notes. Mathematical Association of America, Washington, D.C., 1989.

[2] Ronald A. Thisted and Paul F. Velleman. Computers and modern statistics. In *Perspectives on Contemporary Statistics*, number 21 in MAA Notes, pages 41–53. Mathematical Association of America, Washington, D.C., 1992.

[3] Deane E. Arganbright. Using spreadsheets in teaching statistics. In *Statistics for the Twenty-First Century*, number 26 in MAA Notes, pages 226–242. Mathematical Association of America, Washington, D.C., 1992.

[4] Iris Brann Fetta. Graphing calculator enhanced introductory statistics. In *Statistics for the Twenty-First Century*, number 26 in MAA Notes, pages 258–270. Mathematical Association of America, Washington, D.C., 1992.

[5] Marilyn K. Pelosi, Theresa M. Sandifer, and Jerzy J. Letkowski. *Doing Statistics with Excel for Windows^{TM} Version 5.0: An Introductory Course Supplement for Explorations in Data Analysis.* John Wiley & Sons, Inc., New York, 1996.

[6] Kathy Kenny, Joe Massoth, and Barbara Hall. Generic and name brand cereals: Is there a difference. Project report for Mth 161. Elementary Statistics, 1997.

[7] Sharon Clark, Mai Hoang, Lillian Driver, and Joseph P. Browne, C.S.C. A survey of the use of the library at the university of portland. Project report for Mth 161. Elementary Statistics, 1997.

[8] Micki Bergantzel, Kelsi Compton-Griffith, Anne Davis, and Mandi Van Dyke. Project M&M. Project report for Mth 161. Elementary Statistics, 1997.

[9] Florence Gordon and Sheldon Gordon, editors. *Statistics for the Twenty-First Century.* Number 26 in MAA Notes. Mathematical Association of America, Washington, D.C., 1992.

[10] Harry V. Roberts. Student-conducted projects in introductory statistics courses. In *Statistics for the Twenty-First Century*, number 26 in MAA Notes, pages 109–121. Mathematical Association of America, Washington, D.C., 1992.

[11] David C. Hoaglin and David S. Moore, editors. *Perspectives on Contemporary Statistics.* Number 21 in MAA Notes. Mathematical Association of America, Washington, D.C., 1992.

[12] David S. Moore. *The Basic Practice of Statistics*. W. H. Freeman and Company, New York, 1995.

[13] Lynn A. Steen, editor. *Heeding the Call for Change: Suggestions for Curricular Action*. Number 22 in MAA Notes. Mathematical Association of America, Washington, D.C., 1992.