UNDERGRADUATES CAN CONDUCT ORIGINAL RESEARCH AND LEARN DATA ANALYSIS IN THE PROCESS ®

Richard Baldy
California State University
USA

Undergraduates at California State University, Chico's College of Agriculture do experimental research including data analysis. Desire to see if treatments differ, motivates students to learn inferential statistics. Students analyze data with ANOVA programs written for Microsoft Excel. These programs return an analysis with one or at most a few mouse clicks; they handle missing data – a problem in real world experiments. In statewide science competitions, our students routinely place first or second.

INTRODUCTION

Complacency among science teachers satisfied with the state of American undergraduate science education might change to concern, if they were to consider the parallels between science education and music education. Imagine a music education program in which nearly all graduates could appreciate music, but could not create, nor perform music. Such is the state of science education; our science programs graduate students who can appreciate science, but cannot create knowledge by doing (performing) science. Walker (2001) polled 180 B.S. degree holders entering the California State University, San Jose science teaching program to learn how many had formulated a hypothesis, developed a protocol, carried out the experiment, and analyzed its data. Of the 180, two had direct research experience. Yet, all these science-teachers-in-training will become our grammar school, high school, and in many cases our community college science instructors without the benefit of ever actually doing science. This state of affairs is akin to having teachers teach music without ever picking up an instrument.

A principal reason for our failure to teach science majors to do science maybe the belief that applied inferential statistics is too difficult for undergraduates. I will argue in this paper that this belief is outdated. Computers and easy-to-use data analysis programs make data analysis the easy part of experimentation. Push students to do experiments and you will create the best environment to teach inferential statistics. Students are primed to learn statistics, because they are curious to determine if their treatments are significant or not.

In the College of Agriculture, California State University, Chico we have developed two courses, which deal with problems that can be approached through scientific experimentation. The courses are still evolving, but already we are sure that our curriculum is teaching students how to do science. Before describing the courses, I will mention three assessments that encourage us to continue with this curriculum that includes year-long, two-course sequence in experimentation.

1) Deans from other colleges, after hearing the student research reports, have remarked that some students' research rivals that of master degree candidates. Members of the agricultural industry in attendance have been excited and surprised by the quality of the research and the presentations.
2) Each year the university offers 10 creativity awards for undergraduates from all disciplines. Students in the College of Agriculture make up about two percent of the campus population. Our students routinely receive 20-40 percent of the awards which they use to support their research.
3) The California State University system has 23 colleges with programs in natural science and four with programs in agriculture. Students from the natural science departments and the agriculture departments are lumped together for the statewide annual undergraduate research competition from; our students always bring home first and/or second place awards.

THREE FACTORS THAT PROMOTE SUCCESS

I attribute our success to three factors: 1) We compel students to begin their experiment immediately, 2) we set up situations where second semester students teach first semester students, and 3) we provide easy to use data analysis software. Our success is not attributable to teaching a select group of students. All of our students take the experimentation course. Most are not academically gifted. In our state, the universities of choice for the academically gifted include the University of California and private universities such as Stanford. The California State University system in which I teach accepts the next lower tier of students.

Require Students to Start Their Experiments Early

Almost all experimentation class members are completing their final university year. During the first semester the class meets for eight hours per week – two 4 hour lab/discussions. Some students come to the class with a research topic in mind, for example, research to conduct on their family's or employer's property. Students without a research idea spend the course's first four hours at a "science fair." Faculty, university farm staff, local agricultural extension agents, and students who have completed their research describe ongoing research projects and suggest research ideas. As Bausell (1994) observed, students will not learn to do research until they begin to do research. Therefore, I give students just one week to pick a research project, state the experiment's null hypothesis or hypotheses, settle on the treatments, and decide what will be the response variable, and how it will be measured. I prod students with weekly tests, roll calls, and the certainly of a poor grade, if they fall behind. These academic threats provide enough "activation energy" for students to start their research.

Let Advanced Students Teach Beginning Students

Because students should also be thinking about the experimental design they will use, the second class meeting introduces them to a number of experimental designs such as 1) single factor completely randomized, 2) randomized complete block, and 3) Latin-square designs. The single factor design may include a covariate. The instructors for these four hours are the second semester students. They lay out a mock experiment to demonstrate each design, explain the conditions that call for each design, and discuss the null hypothesis the experiment is designed to test. The first semester students are the experimental units and may be "trees", "cows," etc. They are assigned treatments – often a type of candy – and their responses, such as waist size, are measured to illustrate that the response variable must be numerically expressed. Because students are active participants, they tend to remember the layouts when the analysis of the designs is explained later. For the second semester students, the instructor's role forces them to review material they learned their first semester. They also convey the message "we learned this material last semester, so can you!" Second semester students serve as instructors four more times throughout the semester. In the second week of the course, they present their research and answer questions from the new students. In the fourth week, they introduce two factor experiments by setting out mock treatments arranged in completely randomized, randomized block, and split-plot designs. In week five, the advanced students teach Power Point presentation software and prepare first semester students to present their experiment's budget to a committee composed of the college dean, faculty, and university farm staff. In week eight, advanced students introduce simple and multiple regression, and analysis of count data.

Provide Easy to Use Data Analysis Software

In the previous section, I mentioned a lot of experimental designs – nearly all the designs covered in classic experimental design-data analysis texts for agricultural graduate students and professional researchers. To cover all these designs in one semester while also launching undergraduate students on their research projects is impossible, if one has to spend much time teaching a data analysis program, and if the program requires involved steps. Also, an agricultural industry advisory committee charged us with the mission of teaching students to solve problems. I interpret that charge to mean that graduates would remember how to run experiments and analyze data when they are on the job. The data analysis programs must be so easy to use that one could ignore the programs for a year and then come back to them to analyze data. In agriculture data analysis typically occurs once a year, after harvest.

I did not find a program that met the above requirements, so I wrote a suite of ANOVA/ANCOVA programs – one for each the experimental designs covered in the course – in visual basic for Microsoft Excel (Baldy, 2001). Students come to class with an acquaintance with Excel – it is usually installed on their computers. Upon graduating they will continue to use Excel for other purposes, even if they use it for data analysis only once a year.

The user selects the ANOVA/ANCOVA program for his or her experimental design. The data entry sheet has column headings appropriate for the design. For a randomized complete block design column headings are Treatment, Block, and Response. Treatment and Block columns accept number codes or descriptive words such as copper, zinc, iron for treatments and north, south, east, and west for blocks. If a datum is missing, the user enters a decimal point in place of a response value. The programs use iterative methods to compute the correct sum of squares for each term in the ANOVA table's sources of variation column when experimental designs become unbalanced due to data loss. With single factor experiments without preplanned contrasts just one mouse click yields an ANOVA table and means chart with symbols indicating if means differ significantly.

The more complicated designs, such as a split-plot design with preplanned single degree of freedom contrasts requires the user to input data and coefficients for orthogonal contrasts and to make four mouse clicks to: 1) run the analysis, 2) and 3) select means to contrast, and 4) run the contrasts. All mouse clicks are in boxes with prompts such as "click here to run." Based on the suite of data analysis programs is a suite of programs that estimate an experiment's power to declare as significant user-defined single degree of freedom orthogonal contrasts. Another program randomizes treatments for a Latin-square design. Recently I wrote a program to estimate an experiment's economic return. Excel comes with adequate simple and multiple regression capabilities, so I did not have to write regression programs. For analysis of count data students use Miller's (1997) program that tests if the proportions among categories, for example, pregnant or not pregnant, vary with treatments.

COURSE OUTCOMES

My colleague Dr. Wesley Patton guides students their second semester. That class meets four hours per week. These students mentor first semester students and complete their research including data analysis and communication of results via a scientific paper, a poster, a web posting, and an oral presentation in scientific meeting format. We videotape these presentations so students can include their presentations in their portfolios.

REFERENCES

Baldy, R.W. (2001). *Using Excel Templates for Routine Analysis of Agricultural Experiments*. Paper presented at joint ASSUME and Royal Statistical Society Statistical Computing Section Spreadsheet Meeting. Available at http://www.gla.ac.uk/External/RSS/RSScomp/spr2mtg.html

Barker, B.R. (1994). *Conducting meaningful experiments – 40 steps to becoming a scientist.* Thousand Oaks: Sage

Miller, M. (1997). *RxC - R by C: A program that performs Fisher's Exact Test on any sized contingency table through the use of the metropolis algorithm*. Available at http://bioweb.usu.edu/mpmbio/

Walker, D.B. (2001). *Inquiry as a basis for learning: A different college biology course*. Paper presented at the 167th annual meeting of the American Association for the Advancement of Science.