

A Class Project in Survey Sampling

Andrew Gelman* and Deborah Nolan†

July 1, 2001

Courses in quantitative methods typically require students to analyze previously collected data. There is great value in this exercise especially when they analyze data collected to address real problems from the world outside the textbook. Yet, in our experience it can be difficult for students to bring the analytic skills learned in the classroom to an independent project.

More recently, activity-based learning has gained popularity. When students actually collect data, they get a feel for the practical struggles and small decisions needed in real data gathering, and many key concepts in survey sampling can be well illustrated through these endeavors. Unfortunately, however, these data collection activities are often limited to taking measurements on fellow classmates, where the data have little intrinsic value and the data collection techniques do not represent sound statistical practice. Yet, sending students off to collect meaningful data on their own can be a very tall order; we have found that the results are usually disappointing because the task is so difficult.

As an alternative, we have developed a whole-class project that we think melds the best of these two approaches to learning. The entire class works together on one project to design a survey and collect and analyze data that address a real question. We have successfully used this project in our upper level statistics courses, but it is also appropriate for courses in quantitative methods in the social and life sciences. We have been very pleased with the outcomes, and we have found that the students take pride in the accomplishments of the class.

*Department of Statistics, Columbia University, New York

†Department of Statistics, University of California, Berkeley

Organization

This is a major endeavor that takes the better part of a semester to complete, although the effort required from each student is limited to a few hours over one or two weeks. To prepare, we choose the topic of study before the semester starts and collect background materials (newspaper clippings, research papers, and reports) to frame the problem, contact authorities on the topic for advice, and arrange guest lectures for our class. We break the project into smaller manageable tasks: questionnaire design, sampling plan development, data collection, analysis, and report writing. Groups of students work on different tasks, which the instructor oversees.

We introduce the project early in the semester, and set deadlines for completion of the main pieces. If the class is large, more than one group works on the same task. For example, two groups may design questionnaires, the advantage being that we merge the best parts of each into one questionnaire. One large group of about six students takes charge of data collection and validation, but everyone in the class helps collect data.

The project is an ongoing theme throughout the term. We schedule groups to report to the rest of the class on their part of the project, and we use these reports to manage the project, to solicit input from all of the students on each aspect of the project, and as a means for teaching survey sampling. Each group writes a summary of their contribution, which appears in the final report as an appendix.

Survey Topics

We choose our survey topics according to current activities and political issues on campus, and the group that we study is usually some subset of the student population. In other fields, the topic of the survey can be tied to the course material, provided the relevant population is available for study. We present three of our class projects here to give a flavor for the variety of issues that can be addressed. The questionnaires can be found on the Web at www.stat.berkeley.edu/~nolan/surveys.

Video games

The goal of this survey was to inform a committee of statistics faculty who were designing new interactive computer labs for statistics classes about students preferences and attitudes toward video and computer games. The target population was students who would be using the computer labs in statistics courses, the sampled population was all undergraduates enrolled in Stat 21, Fall 1994, at the university, and the sampling frame was the list of students who took the midterm exam in the week before the survey was conducted. A simple random sample of 95 of the 310 students were selected for the sample. To reduce nonresponse (92 of the 95 responded), several steps were taken:

1. Each section of the class (a section has 30 students and meets twice a week with a teaching assistant) was visited on the day that the exams were returned to the students. Before attending section, the teaching assistants had been informed of the project, and the instructor for the course asked that they accommodate the data collectors. The surveys were handed out and collected in section.
2. The data collectors returned to the same section later in the week to reach any students who had not come to the earlier section meeting.
3. Finally, the names of the students who had not attended either of the two section meetings were written on the blackboard in the main lecture hall. The instructor made an announcement at the beginning of class for the listed students to raise their hands, and questionnaires were distributed to them and collected in class.

Math majors

The goal of this survey was to study the differences between male and female majors in regards to their course work, self confidence, and future plans in mathematics. The faculty and staff of the Mathematics Department gave input into the design of the questionnaire. For motivation, we provided the class with related findings from studies of students in mathematics and science. The target population was all math majors at

Berkeley. The sampling frame was the list of declared and registered mathematics majors at the university in January, 1995. A complete census was taken because there were only 115 majors. We obtained permission from the instructors of math courses to approach the majors at the beginning of the class to complete the survey. For the few majors who were not enrolled in a math course that semester, we found them outside of the meeting places of their other classes.

In the semester following the survey, we led a small group of undergraduate students on a project to carefully analyze the results of the survey and compare them to a qualitative study of female math majors from around the country. The findings from this project were published in Adhikari et al. (1998).

Community service

California Governor Gray Davis proposed a community service graduation requirement for all college students in order to instill in them a “sense of obligation to the future.” A campus task force was formed to prepare a response to the Governor’s proposal. To help the task force determine the current level of student participation in service activities, and the general attitude of students toward community service, we volunteered to work with the task force to design a survey on the topic.

For background material, we read current news stories such as, “Educators, Politicians Ponder Forcing Students to Volunteer,” May 18, 1999, *SF Chronicle*, and the Rand report, “Combining Service and Learning in Higher Education,” Ondaatje & Zakaras. We met with members of the task force to find out about existing avenues at the university for students to perform community service. These included campus programs such as Cal Corp., America Reads, and America Counts, off-campus activities through churches and student societies, and faculty-led service-learning courses.

The target population was all undergraduates at the university, and the sampling frame was the Registrars list of students enrolled in the Spring '00 semester. We recommended stratifying the sample according to major and college. The questionnaire did not ask for personal characteristics because this information was available from other sources.

The class fine-tuned the questionnaire in two stages of pilot studies. Due to confidentiality, our data collection and subsequent analysis was limited to the pilot studies. We forwarded the findings to the committee as examples of how they could summarize the complete survey results.

Carrying it out

Although only a subset of the students work on any one task, we organize class discussions on all the topics. We provide students with handouts to clarify each group's job and to focus our in-class discussions. These instruction sheets are available on the Web at www.stat.berkeley.edu/~nolan/surveys. We briefly describe here how the tasks were integrated into the course, and to give an idea as to what we expect from the students.

The sampling plan

We asked the team designing the sampling plan to come to class prepared to identify: the target population, i.e. the group we want to study; the sampling frame, i.e. the list to use to reach the members of the target population, and the sampled population, that is, the group from which the sample will be taken. They were to also supply a probability method to choose individuals for the sample and the sample size.

After the in-class discussion, the group draws up a formal sampling plan and submits it for review and finalization by the instructor.

The Questionnaire

The entire class first brainstorms to come up with a list of information that we want to find out about the population, keeping in mind that some information may be sensitive or otherwise hard to collect. Then we ask smaller groups of students to bring sample questions to class for discussion. We provide these groups with a list of examples of what not to do in order to assist them in reducing bias (eg. Do you not believe that video games could possibly not be educational? Do you use illegally copied video games on

your PC? How many hours do you spend a week on average playing video games? Do you like math?)

The groups must try their questionnaires out on a few guinea pigs to uncover points of confusion and other problems and improve the questionnaire. We combine and edit their efforts to produce the final questionnaire.

Data collection

We expect each member of the class to spend two hours collecting data for the survey. The data collection team is responsible for providing explicit instructions on how to do it. They oversee the data collection process, including handling nonresponse, data entry, and cleaning. The handout has a list of the steps involved with examples, eg. to preserve anonymity in one survey, we provided each data collector a manila envelope and a list of students to contact; a student's name was crossed off this list when he or she returned the questionnaire; and completed questionnaires were turned in for data entry without any personal identification on them.

We ask the data collection team to give frequent progress reports in class in order to keep the data collection on track and help out when problems arise.

Data Analysis

The appropriate analysis of the data depends on the questions asked. At first we have the analysis team(s) supply the class with preliminary findings, including basic tabulations and graphical summaries of the responses to each of the questions. With these summaries in hand, we brainstorm with the class about what further analyses should be prepared. We emphasize that it is important to provide a clear and complete picture of those surveyed, and this includes looking at the relationships between variables through say cross-tabulations and scatter plots.

The analysis team supplies the report writing team with a summary of its findings as well as detailed numerical and graphical results.

The Write-up

We always have a target audience in mind when summarizing the findings of the survey. For example, we ask the report team to write a memo to the Statistics Department committee in charge of designing new computer laboratories, a press release to the student newspaper on the future plans of math majors, or a report for the Chancellor's office with recommendations on mandatory community service work for undergraduates.

Not all of the numbers, charts, and graphs provided by the analysis group make it into the body of the final report. The writing team selects those which succinctly and accurately describe the population of interest. To help shape the final report, the writing team presents a rough draft to the class for critique.

We ask that the final report includes the following information: the purpose of the survey, a summary of findings and conclusions, a brief description of the survey methodology, and a detailed description of the findings with supporting numerical and graphical statistics. We also ask that the report authors put the results in context with other findings in the area, discuss the problems encountered, including nonresponse and other potential sources of survey bias, and include as appendices the reports from the other groups.

Reflections

We found that our students really enjoy working on these large, purposeful projects. The whole-class format has many advantages. Students can choose the task for which they are best suited, and the value of the role they play in the process is evident to everyone in the class because the entire process is viewed collectively. In addition, the work load for each student is kept to a reasonable level.

Although the class is directly involved in the design and implementation of the survey, the instructor oversees and guides the project. As a result, the students learn by doing and they learn through the advice of the instructor. The instructor's involvement at each step helps secure a successful outcome, and it sets a good example of how to carry

out independent research projects.

These projects generate a lot of enthusiasm in the class. This enthusiasm is evident in the follow up projects that we have led. Students have worked individually with us to carry out further surveys and to write honors theses on these topics, and they have worked in groups to further analyze the results of the surveys. Some of these efforts have resulted in publications (Adhikari et al, 1997, Nolan & Speed, Chapter 2, 2000).

References

- Adhikari, A. and Nolan, D. (1997) A view of mathematics from an undergraduate perspective. In *Women in Mathematics: Scaling the Heights*, MAA Notes, No. 46, 97-104.
- Nolan, D. and Speed, T. (2000) *Stat Labs: Learning Theory through Applications*. New York: Springer-Verlag.