

FORMULATING STATISTICAL QUESTIONS AND IMPLEMENTING STATISTICS PROJECTS IN AN INTRODUCTORY APPLIED STATISTICS COURSE

Katherine Taylor Halvorsen

Department of Mathematics and Statistics, Smith College, United States of America
khalvors@smith.edu

Students taking introductory statistics in the Mathematics and Statistics Department at Smith College, a liberal arts college for women in Northampton, Massachusetts, conduct independent research projects as part of their coursework. Instructions for the project, given to students with the syllabus, include a list of deliverables: research proposal, peer reviews, data, progress reports, draft analyses, and due dates for each. Students, working in groups of two to four, write a research proposal that includes their research question and an outline of their data collection and analysis methods. Before students begin data collection, research proposals must be approved by the instructor(s) and, when human or animal subjects are used, by the Smith College Institutional Review Board. Students carry out the research during the term and present their work in a poster session, oral presentation, or a research paper at the end of term.

INTRODUCTION

Since 1989, the Smith College course, Introduction to Probability and Statistics, has required students to formulate a research question on a topic of interest to them, write a formal proposal describing the work they will do to answer the question, and following approval of their proposal, carry out the research. The motivation for incorporating these projects into the course is the firm conviction that students learn statistics by doing statistics and that doing statistics means formulating a statistical question, designing a plan for collecting relevant data, using appropriate statistical methods for analyzing the data, and presenting results in a public setting such as a poster, an oral presentation, or a paper. Projects such as these offer instructors additional opportunities to teach students to develop statistical questions out of broader research questions, to design data collection and data analysis plans, and to write and speak clearly about statistical issues.

The importance of student projects is supported by the American Statistical Association's 2005 document, *Guidelines for Assessment and Instruction in Statistics Education*. Their report recommends that instruction in statistics focus on statistical problem solving, using real data, active learning, and appropriate use of technology. The Guidelines define statistical problem solving as a four-step process: formulating questions that can be addressed through the collection and analysis of data, collecting and organizing appropriate data, analyzing the data, and interpreting results. We incorporate these steps into our course projects.

The course in which we conduct these projects, Introduction to Probability and Statistics, has a prerequisite of one semester of calculus or discrete mathematics. This prerequisite acts primarily as a filter to ensure a certain level of mathematical maturity in the students. Most of the students who enroll are mathematics or science majors in their first or second year at Smith. Typically, the course enrolls 40 to 50 students each semester. The course meets weekly for three seventy-minute lecture/discussion periods, and students meet, in groups of 15 to 20, for one eighty-minute laboratory period each week. The labs are used for computer exercises and data collection and analysis activities. Some of the project work takes place in the labs.

This paper discusses each of the steps instructors and students take in the student-conducted research project. It describes how the process works from the student's, and from the instructor's, points of view. Copies of the instructions and schedule students receive at the beginning of term are available on request from the author.

INITIAL INSTRUCTIONS

On the first day of class students receive, in addition to the syllabus, a five-page handout describing the class project. We ask them to select a research question that interests them and to find one or two partners from the class who are willing to conduct the research with them. Projects can incorporate original experiments or observational studies, including surveys, or they can reanalyze existing data that they find on the internet or obtain from faculty, or other sources.

Because real-world research is typically conducted in research teams, we believe students need to develop the skills necessary for working in teams. Thus, we strongly discourage students from working alone, and we limit group sizes to 2 to 4 so that each student will have a large part of the responsibility for the final product. We spend some class time talking about developing a professional working relationship with their research partners, and the importance of setting and meeting agreed upon deadlines. We believe this time is well spent; we have had few dysfunctional working groups in a 20-year span of observing student-driven class projects.

The initial project handout also contains a list of the intermediate steps students will need to follow in carrying out their project, and it provides explanations of each step. These steps include handing in each of the following deliverables: the initial outline, a proposal, peer reviews of other students' proposals, an interim progress report, their data in an electronic spreadsheet, draft analyses, final report drafts, the final poster, peer reviews of other posters, and a reflective paper on their experience conducting the research. Each step in the process has a due date, specified in the initial instructions. Students can study examples of project outlines and a sample project proposal on the course website. We also maintain a list of titles of previous student projects that is available to students when they are planning their own research.

We state some general rules for the project at the outset. The work, both the analysis and the write-up, must be the original work of the research team. Students may consult other sources for information about the non-statistical, substantive issues in their projects, but they must credit these sources in their poster or paper. They are told to consult the instructors about statistical questions, and they are told explicitly and frequently reminded that, "*Projects that involve using human subjects (interviewing, measuring, etc.) or live animals must be reviewed by the Smith College Institutional Review Board (IRB), and you need to allow time for this in planning your project.* Institutional review ensures that the subjects will be treated properly, their privacy protected, and their consent obtained. If you need to use human subjects, we will show you how to prepare your proposal for institutional review."

PROJECT OUTLINE

At the end of the second week of class students must hand in the project outline. Each research group hands in one copy of the outline, and all students in the group get credit for the assignment if this step is completed on time. The outline consists of a one-page form on which they very briefly describe the proposed project. We provide a blank form on our class website. On the form, students specify who will be included in their working group and which one of them will be their primary liaison with the class instructors. They give the topic of their project, the type of study (experimental or observational), type of data (original or secondary), the type of units (human, non-human vertebrates, or other), the research question, the explanatory variables, the response variable, and a description of the comparison they will make in their research (e.g., group means) or the association they will examine. Students must collect a minimum of two variables on each unit, and we discourage them from collecting more than 5 or 6. For example, proposed surveys must fit on half a page. We use this step in the research process to encourage students to find partners and a suitable project early in the semester, and the outline serves the students by providing early feedback on their thinking about statistical questions.

PROJECT PROPOSAL

The research proposal consists of a paper, three to five pages in length that describes the research question and the methods students plan to use to both collect and analyze their data. It provides a blueprint for subsequent work on the project. The fundamental goal of the proposal is to convince the reviewers that the question is worthy of the research effort and that the students have a plan of action that will result in successful achievement of these objectives. Students hand in three copies of the proposal, one for the instructor's review and two for peer reviews.

We require the proposal to follow a fixed format. This makes review and evaluation of a large volume of proposals by each of several reviewers work more smoothly. Proposals must be typed, double-spaced, on 8.5 x 11 inch, white paper, and include a cover sheet with the names and contact information for the authors, the course number and name, the date, and the course instructors' names. The cover sheet also must contain a statement: "This proposal (requires/does

not require) institutional review.” We prioritize projects that require IRB review because that review process takes longer than reviews of other projects.

The first section of the proposal provides an introduction to the students’ research topic that, if well written, may serve as the introduction to their final report or poster. Students describe what they already know about their topic and explain why the topic is of interest to them. They describe their research question, and they restate their research question as a statistical question that can be answered through the collection and analysis of data.

Subsequent sections of the proposal describe the specific methods students will use to collect, organize and analyze their data. In each section, students answer questions listed in the project instructions about their proposed research. We ask students to describe their methods clearly enough so that someone else could collect the data for them without further instruction.

Students describe their experimental or observational units, the population from which the units are drawn, and the proposed number of units they will use. They address the issue of sampling bias by discussing what possible biases could occur and how their plans will minimize the bias. If the units are drawn at random from a population or if the units will be assigned at random to treatment conditions, students describe the randomization process they will use. They tell what variables they will measure on each unit, give operational definitions for each variable, describe how each variable will be measured, and explain why each variable is needed to answer the research question. They identify each variable as an explanatory variable or as a response variable. Students also need to explain their planned research procedures starting with the collection of the units and describe each step in the application of the treatments (if experimental), the collection of the data, and the processing of the data. If the research will use a survey, then a copy of the survey form must accompany the proposal, and if the data is collected in an experiment, students must provide a copy of the form they will use to record the data as it is collected.

If human subjects will be involved in the research, the researchers must not deceive the potential subjects nor put them at risk in any way. Students, whose research involves human subjects, must include in their proposal a draft of a letter that will be used to invite subjects to participate in the research. Potential subjects must be told that their participation is voluntary and that they may stop their participation at any time, before or after the study begins. Students also draft a consent form that subjects will be asked to sign as proof of their agreement to participate. The consent form must clearly explain the study, state that their participation is voluntary, and assure participants of the confidentiality of their responses. The proposal must show how the researchers will ensure that confidentiality through de-identification of the data. Students who want to collect data off campus, in a store or in front of a business establishment, must provide a draft of a letter to the business owner asking permission to collect data at that location. The Smith Institutional Review Board provides explicit guidelines on the contents of the consent form, and students must meet the IRB standard.

Students write their research proposal within the first three weeks of the course. During those weeks the class has reviewed graphical techniques for displaying data and numerical methods for summarizing data including both univariate and bivariate tools, such as parallel boxplots, scatterplots, correlation, and simple linear regression, material that most of them have seen in high school mathematics or science classes. They have also completed a unit on study design and data collection methods, but they have had no exposure to formal inference procedures. Thus, we cannot ask them to fully describe their statistical analysis plans in their proposal. Instead, we ask students to describe which graphs and tables they expect to use to display their results and to anticipate the relationships they expect to find between the response and the explanatory variables. As students learn inferential methods for comparing groups later in the semester, we ask them to incorporate these into their project data analysis.

Students include in the appendix to the proposal any necessary supplements such as the data collection form, a cover letter to the IRB, the letter inviting volunteers to participate, the informed-consent statement, and any requests to third parties involved in the research. Each research group submits one proposal for the group, and they hand in three copies that will be used in the review process.

PEER REVIEW OF PROJECT PROPOSALS

The course instructor, the lab instructor, and the students get involved in proposal review. We ask each student to read and assess someone else's project proposal. We emphasize that these reviews are meant to be helpful to the recipients, and we want students to find positive aspects of the proposal to compliment as well as issues that need clarification. We provide guidelines for the written review. We emphasize that comments should be helpful, not hurtful, and that vague positive feedback is not as helpful as tactful constructive comments for improving the proposal.

MIDTERM PROGRESS REPORT

Midway through the term we ask students to assess their progress and complete a progress report. We ask them to summarize, in writing, their term project and their progress on it to date. We provide a set of questions to answer in their report, and tell them that we understand that they may not be able to answer all our questions if they haven't reached that stage of their research. We particularly want to know whether anything has changed since the proposal was reviewed and approved. Have difficulties arisen and if so how have they resolved these issues? We also ask for a timetable for completion of the project. They need to list the tasks remaining to be completed, estimate how long each will take, and when each will be completed. Our goal for this one- to two-page report is to ensure that the projects are on track for successful completion, on time for the poster review on the last day of class.

OTHER INTERMEDIATE DEADLINES

Approximately one month before the end of term students hand in an electronic copy of their data and preliminary descriptive analysis of their data. We review these and return them to the students with our comments and suggestions. During the last two weeks of term, we discuss the poster format and content with students, and we use lab time for students to complete their final data analyses. We ask that posters contain each of the following items: a title, a list of the authors and their institutional addresses, an introduction, methods, results, and discussion sections. They may include references and acknowledgements where appropriate. In the week before the poster session, they hand in drafts of their posters and we give them feedback, either in person or written on the drafts. We help them work on clearly written exposition of their work.

POSTER SESSION AND FINAL PEER REVIEW

At the poster presentation session on the last day of class, students bring their completed posters, a separate abstract (for the instructor's records), and a one-page reflection paper on the research process that has just finished. We exhibit the posters in a large display hall where students and invited guests can mingle, view, and discuss the posters. Each poster receives an identification number, and the instructors randomly assign three posters to each student for peer reviewing. Each poster receives three or more reviews. The reviews focus on presentation, statistical content, and subject matter. We ask the reviewers to consider the following questions:

- *Presentation:* Is the poster attractive and well organized? Are all components present? Is it easy to follow? Is the text well written? Are the graphs and tables easy to understand?
- *Statistical content:* Did the investigators use appropriate statistical techniques? Are the techniques used correctly? Are the figures appropriate to the data?
- *Subject matter content:* Did you learn something about the subject of the poster? Is the content interesting, fun, provocative, compelling, educational, or fascinating?

The invited guests at the poster session typically include other faculty, some of whom may have provided their research lab facilities for student use or have permitted students to use data the students have collected while working on research with the faculty member during the summer or a previous semester. We also invite the local high school statistics class to attend, and we provide cookies for the occasion. We are celebrating the successful completion of a long process.

STUDENT EVALUATION OF RESEARCH PROCESS

Students complete a one-page questionnaire during the last week of the course that asks about their experiences carrying out their project. For about half the students this is the first extended group-research project they have completed in college. The others have participated in student-faculty research projects, or research projects in other courses, but for most, this is their first independent research experience. Many are surprised by the detailed planning required, and they come to recognize the importance of planning and organization in managing the details of the project. Several have commented on how important having deadlines is to completing the work on time. Most note how difficult it can be to coordinate work among team members, and yet almost all of them are able to create functional teams.

Students learn the process of research, and they also learn statistics through the project. They mention learning that data collection is often tedious and more difficult than they expected, especially those who plan surveys. They say that the project makes the material they study and practice in class much more meaningful. One student said that the research project provided her with a context for her in-class learning. They say that as they work to compose their posters and communicate their research questions, methods, and findings, the project helps them to see the importance of clarity, simplicity, and discussing their research results in the context of the original research question. Sometimes, they are even surprised by their results. One student commented, "College students do not know nearly as much about multiple sclerosis as I thought!"

CONCLUSION

The mechanics of managing 15 to 20 small groups of students doing projects may seem daunting. However, detailed planning, before the semester starts, that includes preparing the project instructions, setting the deadlines, providing resources on the course web pages, and delineating the responsibilities of the lecturer and the lab instructor all serve to make the projects possible. Projects provide students an opportunity to learn and show what they have learned outside the context of timed examinations. Strong students benefit from the experience, but the projects are particularly important for weaker students, many of whom dread taking a statistics course. Often they are relieved, sometimes pleased, to find an area within statistics that they can master and enjoy. During office hours, instructors frequently work closely with students to discuss issues that arise in the process of their research. This small-group work with an instructor helps students connect the work of the project with in-class learning, and gives instructors another opportunity to assess student effort and learning. The results from including student projects in the introductory statistics course are richly rewarding. This is particularly evident at the poster session at the end of term, when student can review what they have accomplished, see what others have produced, and receive questions and complements from their peers and guests. Their "Reflections on the Project" papers after it is all finished are priceless.

REFERENCES

- Franklin, C., Kader, G., Mewborn, D., Moreno, J., Peck, R., Perry, M., & Schaeffer, R. (2007). *Guidelines for Assessment and Instruction in Statistics Education: Pre K-12 Curriculum Framework*. Alexandria, Virginia: The American Statistical Association. Online: www.amstat.org/education/gaise.
- Halvorsen, K. T., & T. L. Moore. (2000). Section 2: Motivating, Monitoring, and Evaluating Student Projects. In T. L. Moore (Ed.), *Teaching Statistics: Resources for Undergraduate Instructors* (Notes #52, 27-32). Washington: The Mathematical Association of America.