Instructor Guide

Discovery Projects in Introductory Statistics

Sherry L. Hix, Ph.D.
Dianna J. Spence, Ph.D.

Development of this resource supported by NSF grant award DUE-1021584

Table of Contents

Overview

This is an Instructor Guide to be used in an introductory statistics course using projects to facilitate student learning. The projects used in these materials include a linear regression project and a group comparison project. The purpose of the Instructor Guide is to detail the structure and design of the projects. Implementing projects in order to accomplish student learning requires a different perspective than accomplishing student learning through more traditional methods. This guide should provide insight about that perspective and detail how to successfully implement projects for the purpose of student learning.

The Instructor Guide describes the five phases of the projects with specific information pertaining to each project explained separately. Suggestions for timelines are also presented. The appendices contain many examples that may be shared with students at the instructor's discretion.

This guide is not a course outline, but an outline about using projects to help students learn. The projects should be embedded in an introductory course where instruction is ongoing about the statistical inferences and analyses. The projects do not replace instruction, and the guide is not the course curriculum.

Rationale for Projects

According to the National Council of Teachers of Mathematics [NCTM] (2000), assessment should "guide and enhance [students'] learning" (p. 22). A carefully planned and orchestrated project with proper scaffolding is one form of assessment in mathematics that may be used to deliver such learning. Projects provide students an opportunity to perform work in authentic, contextualized settings, mimicking the work of professionals (The Buck Institute for Education and Boise State University; Donovan, Bransford, & Pellegrino, 1999). This type of student-led inquiry may lead to the development of mathematical proficiency, especially the often difficult to acquire productive disposition about mathematics (Kilpatrick, Swafford, & Findell, 2001), and contribute to a development of the necessary habits of mind particular to good problem-solving (Schoenfeld, 1987).

The projects used in this guide are issued from a knowledge-centered environment that requires learning and doing with understanding (Donovan, Bransford, & Pellegrino, 1999). The design of these projects enables students to "organize and consolidate thinking through communication" (NCTM, 2000, p. 348). Landrum and Smith (2007) suggest statistics education in particular should include projects that require data collection and analysis. Hogg (1997) claims students develop an appreciation for statistical concepts when they participate in their own statistical inquiry, and that this appreciation leads to a deeper understanding of statistical thinking and practice (also Roseth, Garfield, & Ben-Zvi, 2008). Synthesizing research findings in statistical education, the Guidelines for Assessment and Instruction in Statistics Education [GAISE] (2005) recommend using real data in alternative settings like student-led projects in order to develop statistical thinking in students.

In a typical introductory statistics course, students learn about test statistics and statistical inferences, often working through many textbook examples in order to practice procedures for each. Projects allow students to conduct tests with actual data, collected by the students themselves. The tests are then analyzed and interpreted in a research report. The students choose a research question, variables, populations, samples, and appropriate tests. In this setting, students are required to become active learners.

In introductory-level college mathematics courses, such as the ones making use of the projects in these materials, many students report taking the course because it is a requirement. Unfortunately, students are often indifferent to the content of the course and sometimes even openly hostile. After using projects similar to the ones described in these materials, many instructors find that students who have typically struggled in the abstractness of previous mathematics courses become very excited about the grounded work of the projects. The students should develop a stronger sense of statistical thinking as they struggle to make decisions about their own project.

Instructors report many learning opportunities that arise from the implementation of the projects in these materials. For instance, initially students often are unable to differentiate between statistical tests for different sets of data. In one such case, students who need to collect data from a sample in order to make inferences using a *t*-test sometimes choose data that require a chi-square test instead. Similarly, students may collect data and try to use a two-sample independent *t*-test when the two samples are actually matched pairs. These misconceptions present the opportunity to foster contextual learning and correct misunderstanding. These reported examples have occurred as a result of students collecting and analyzing their own data and making choices about how to infer about a larger population, not from reading a textbook or completing a problem set. In other words, students were in a situation that required the development of statistical thinking and reasoning, even if the students were not focused necessarily on learning a stated course objective. The opportunity to learn, with careful scaffolding strategically placed, is presented, and students, possibly unwittingly, often seize it.

Implementing Discovery Projects

Instructor and Student Roles

These discovery projects are not ones in which the mathematics is hidden or concealed. Rather, the projects are opportunities for students to engage with statistics in order to understand statistics better. Instructors are not to abdicate their role; rather, they are to instruct. But after instruction, the students must perform and produce.

The students create a product that should *demonstrate their level of statistical understanding*. For this reason, achieving statistical significance is not the primary focus of the project. The process and explanation of the research provide insight into the student's understanding.

Instruction in the form of guidance should be delivered before, during, and after the project. Students may not benefit in any form from complex projects if the proper instructor supports and scaffolds are not in place. Students often experience difficulties, which include, but are not limited to, "initiating inquiry, directing investigations, managing time…" (Thomas, 2000, p.34). If students are preoccupied with administrative details associated with meeting project requirements, they may miss the opportunity to gain the intended understanding of statistical inquiry. Therefore, instructors should carefully plan, articulate, and implement instruction in order to assist students in learning as they are fulfilling project requirements.

Project Phases

Although there is a great deal of flexibility in how instructors choose to implement these projects, this section provides a guide for planning and implementing the various phases of these discovery projects. Most of these guidelines apply to both the linear regression project and the group comparison project. Additional details specific to each type of project will be addressed as needed.

These projects require five phases: project planning, data collection, data analysis, project report, project presentation. Some phases require students to cycle through the phase a second (or third) time. At any phase that the instructor decides to "check" the project work, whether a grade is assigned or not, additional cycles may be needed for students. If, for example, the instructor decides to check the data after they have been collected, the instructor should offer suggestions for improvement as needed and allow students to collect more (or different) data as necessary.

Project Planning

Forming Student Teams (Optional)

Projects such as these are excellent opportunities to have students work in teams, although teamwork is not a requirement. The structure and success of teams within a classroom really depends on the individual students and on classroom dynamics. Teams can vary in size but should probably range from 2 to 4 members. Additional requirements based on number of team members can be beneficial. For example, a team of 3 students can be required to have a larger sample size than a team of 2 students. Instructors may also give students the option to work in teams or to work alone.

The instructor may guide student team selection in several ways. Teams may be formed from students with the same major or similar majors. In this setting, students may be required to focus on research questions relevant to their particular fields of study. Another option is to have teams formed from students with similar available work time for the project to guarantee adequate and feasible group meetings. The instructor may also group students according to competencies required of the project. For example, a team may be formed in which one student possess strong computer skills, another is proficient with the calculator, a third student has research experience, and a fourth is completely new to all of these. Theoretically, students on a team such as this can learn from the expertise available within the group. The instructor may also choose to allow students to select teams.

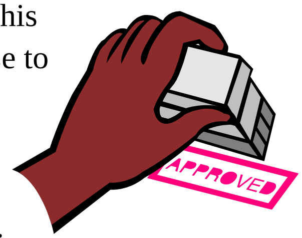
Selecting Variables and Identifying a Research Question

For many students, the freedom afforded in discovery projects can be paralyzing. In order to combat this situation, it is often beneficial to have students consider data and how they are collected. This process may help students to determine variables of interest and ultimately define a research question.

As stated in the data collection section, students may use three different techniques for gathering data: measuring and recording, administering surveys, or accessing reputable Internet sites. Instructors can show students several examples of each of these techniques, followed with actual summaries from previous student projects. This provides an opportunity for students to contemplate variables of interests, as well as different research questions. Refer to Appendix A for a collection of reputable Internet data sites; to Appendix B for a variety of survey-based variables and constructs; to Appendix C for a collection of validated surveys; and to Appendix D for sample student projects.

Documenting and Approving the Plan

Although these projects are not intended to be teacher-led or pre-packaged, certain natural checkpoints exist for the instructor to provide scaffolding and guidance. After students have planned the project, they should submit a well-defined plan to the instructor. This may be as simple as a one-page form or as comprehensive as a formal research proposal. The instructor should review the plan for appropriateness of variables, data collection technique, and data analysis method. After this review, the students will begin collecting data or refine the plan if instructor approval was not obtained. Students should not proceed from this phase until the instructor has granted approval. The instructor may choose to require students to submit evidence of the approval with the final project; this evidence may be in the form of a signed or initialed plan document. This requirement also allows the instructor to hold the student accountable for completing the project that was described in the proposal.



Following are some questions that may be of use as the instructor reviews the plan:

- Are the correct types of variables identified (e.g., quantitative variables for linear regression and t-tests)?
- Is the research question asked in reference to the population, not the sample?
- Does the research question accurately reflect the variables that have been identified?
- Is the data collected from a sample of the population and not the entire population?
- If a significance test is required, is the correct test chosen?
- For t-tests: is the correct kind of test identified (e.g., matched pairs or 2 independent samples)?
- If data will be collected from a website, is it reputable?
- If data are collected from a survey, is the survey a proper one for the research question and data needed?
- If the project requires participation from human subjects, have the students completed appropriate permission forms?

Appendix T1 provides examples of plan forms and assignment sheets describing a project proposal.

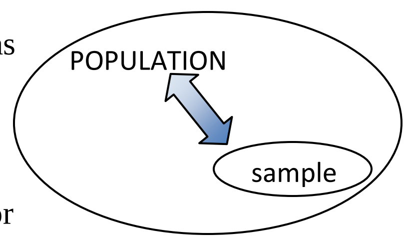
If the instructor allows students to conduct research with human subjects, then instructors should determine the requirements of the school's Institutional Review Board (IRB) or similar governing body. Students may be required to complete additional forms based on the policies of the IRB. Appendix T2 shows an example of a permission form sometimes used for student research.

Data Collection

For the purpose of these projects, students may collect data in three ways.

- Students may record information or take measurements to produce data. Examples of recording information may include recording consumer prices at particular stores or nutritional information from food package labels. Examples of measurements might include timing how long a participant can hold his or her breath or weighing items such as textbooks or newborn puppies.
- Students may administer surveys to participants. Appendix B and Appendix C contain surveys, survey items, and constructs for students to use as a resource. Students also may choose to use surveys from other sources, or they may create their own. In any case, the survey to be used must be submitted with the project plan to the instructor for approval. Surveys are easily flawed or may not align with the construct to be investigated. The students must consider many characteristics of good surveys, for example, neutrality, clarity, and precision of questions. They must also consider how the survey will be "scored." Depending on the student decisions about surveys, this part of the planning phase may require more time and instructor guidance.
- Students may access Internet sites to find existing data. Appendix A contains examples of links to data in four categories: government and community data, restaurant nutrition data, sports data, and retail/consumer (general) data. This resource may serve first as a brainstorming tool to help the students begin to think about the different kinds of research questions they can investigate. It also helps the students to consider the tremendous amount of data available to them.

The instructor should prepare students to use good sampling techniques in the projects. While it is important to note that these projects will very often use convenience sampling, students can think about different strategies to make these samples as good as possible given the constraints under which they work. As in any research, the research report should address what steps were taken to avoid bias in the sample and whether the student believes the sample is representative of the population.



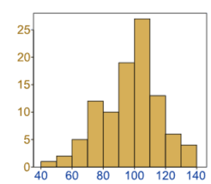
Additionally, the expanse of data available to students can serve as an obstacle. Often students choose a research question and do not appropriately distinguish between the sample and the population it is intended to represent. For example, a student may state that the sample is all Major League Baseball pitchers for the current year. Mistakes such as these should be evident in the project plan and the instructor should assist the students in refining the plan if necessary. If the student in the example above intends the population to be MLB pitchers in the current year, no statistical inference is needed since the student is going to use data from all individuals in the population. This type of confusion regarding samples and populations can lead students to overlook the need to sample, which undermines a focal point of the project. There are many ways to address this problem. The student could use a specific sampling strategy to choose a certain number of MLB pitchers from the current population of pitchers. The student could also choose a certain number of MLB pitchers from multiple years, which would extend the population to be MLB pitchers in general and not just the current year. Alternatively, the instructor could require the student to select a different population for which data are not available for all individuals.

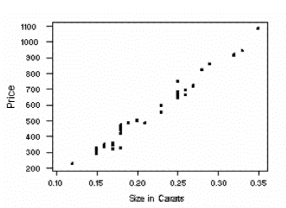
Data Analysis

Although the students cannot collect data until a project plan is approved, the actual data collection will sometimes not match the approved plan. For example, a student's project plan states that participants will be asked for an average highway driving speed. But when the student actually collects the data, the student only asked whether or not the participant has exceeded the speed limit. The data analysis phase then becomes problematic for the students because the variables do not match the analysis required for the project. In the example described above, the student did not collect data for the quantitative variable of driving speed needed for the linear regression or t-test which comprise the analysis for the project. However, if the students should make this mistake, most will quickly understand their mistake and even be able to understand how it should be corrected. This is an example of the conceptual understanding that may develop as students interact with data in ways that mirror statisticians.

Data Analysis Requirements for Linear Regression Project

First, descriptive statistics must be given for each of the two quantitative variables. A table giving the sample size, mean, standard deviation, and 5-number summary should be provided for each variable, as well as a histogram that illustrates the distribution of that variable.



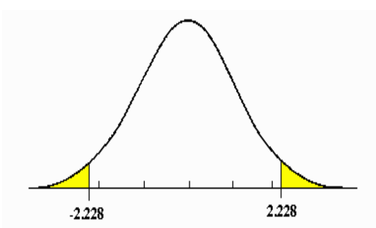


After providing descriptive statistics for each of the variables, students should explore the relationship between the two variables. Using available technology, students should generate a scatter plot with a graph of the regression line, value of r and R^2, and equation of the regression line. Each of these will be interpreted and discussed in the project report (see below).

The instructor may also give additional requirements or optional extensions. One such extension is determining the significance of the correlation coefficient r and interpreting the significance within the context of the problem, to be included in the project report. Another extension is the introduction of categorical variables. For example, if a student performs a regression analysis on the cost of wine and the rating of the wine, the student could also classify each wine in the data set as red or white and perform a regression analysis for each category separately.

Data Analysis Requirements for Group Comparison Project

The group comparison project used in this guide is oriented toward using t-tests. The instructor may choose to extend the scope of the project to include other kinds of statistical comparisons. In that case, similar requirements would hold. Students must choose the appropriate significance test and explain why the chosen test is appropriate for the research question posed. Assuming the group comparison test selected is a t-test, students must identify a matched pairs or 2 independent samples t-test and defend their choice. For either type of t-test, the student will be working with one or more quantitative variables for which descriptive statistics will be needed. For a matched pairs test, three sets of data should be given. For instance, in a before/after design, there will be one set of data each for the "before" values, the "after" values, and the differences. For a test with 2 independent samples, two sets of data should be given. For example, in a design comparing males and females, there will be one set of data for the males and one set for the females. Descriptive statistics must be given for each data set: A table giving the sample size, mean, standard deviation, and 5-number summary should be provided for each variable, as well as a histogram that illustrates the distribution of that variable.



After providing descriptive statistics, students must run the appropriate significance test to make the statistical comparisons of the collected data. Students should determine and state the null and alternative hypotheses. They should use technology or tables to find the test statistic, degrees of freedom, and the *p*-value of the significance test. Each of these will be interpreted and discussed in the project report (see below).

Several additional requirements or extensions may be provided for the comparison project. Students may be required to construct confidence intervals and interpret an appropriate range of values for the mean within that confidence interval. Students also may be given permission to conduct randomized experiments as opposed to simply conducting an observational study.

Other Data Analysis Requirements

If the student has collected data for a categorical variable in a project, the student should report descriptive statistics for that variable as well. These should include a frequency table, relative frequency table, and a bar chart and/or a pie chart.

Project Report

The project report is the culminating part of the project since it requires the student to synthesize findings for an audience. It will reveal much about the student's conceptual understanding. The instructor and student should view the report as the link between the data and the statistics generated by the data. The report explicitly makes the inferences (or not, depending on the outcome of the statistical tests) about the population as the sample allows.

The reports are formal and include nine elements. Students should be aware that this requirement does not necessarily dictate nine paragraphs. For example, two elements may be discussed in one paragraph while another element may require multiple paragraphs for full discussion. The reports should include these nine elements:

- introduction;
- clearly defined population(s);
- clearly defined variable(s);
- description of data collection;
- description of study design;
- descriptive statistics for variables;
- statistical analyses as specified for each project;
- interpretation of statistical analyses; and
- discussion of conclusions.

The report is fully described in the project instructions. A rubric that clearly defines the report will assist the student in writing it as well as the instructor in scoring it. Examples of scoring rubrics for both projects are in Appendix T3. The student should include the approved project plan with the report so that the instructor can verify that the project actually aligns with the approved plan.

Project Report for Linear Regression Project

The report elements outlined above describe general reporting requirements. For the linear regression project, the following specific components should be included within these elements:

- The expected correlation and its strength should be addressed within the introduction and/or the description of the study design.
- The distinction between the explanatory and response variables needs to be addressed in the definition of variables. This entails indentifying which variable is explanatory and which is response and explaining why.
- The statistical analyses should include the following:
  - scatterplot with the regression line
  - value of r and interpretation of its meaning
  - regression line equation
  - a prediction using the regression equation
  - discussion of slope of regression line and its meaning
  - value of R^2 and interpretation of its meaning
  - significance of correlation with discussion of significance level (if applicable)

Project Report for Group Comparison Project

As with the linear regression project, the comparison project report requires the following specific components within the general report framework.

- The description of the study design should address the type of significance test used, including such details as whether the test is 1-sided or 2-sided, left-tailed or right-tailed.
- If a matched pairs design is chosen, the student should explain clearly how the pairs of values were matched and how the difference should be computed.
- The statistical analyses must include:
    - the test statistic
    - degrees of freedom
    - *p*-value of the significance test.
- The *p*-value must be properly interpreted and significance levels should be discussed accurately.
- The decision to reject or fail to reject the null hypothesis should be stated clearly in one or more sentences that interpret the results in the context of the original research question.
- The scope of inference should be discussed: if the analysis warrants an inference to a population, the specific population should be stated and the inference should be explicit. The student should also indicate whether or not a cause-and-effect relationship is warranted and explain why.

The instructor must also determine the extent to which the students will be required to discuss how thoroughly the conditions of the test were met. For example, students may need to discuss the randomness of the sample, the sample size, or the distribution of the variables.

Project Presentation

In addition to assigning a project report, the instructor may also require students to give an in-class presentation of their project design and results. These presentations should be brief (for example, the presentation may be as short as five minutes) and include visual aids (for example, PowerPoint slides) with appropriate graphs and text with correct mathematical symbols. Detailed rubrics for presentation of each project can be found in Appendix T3.

While the report requires the student to provide a robust explanation of the data, research, and conclusions, the presentation requires the student to synthesize this information into an informative and concise presentation of the key points of the research. The rubrics clearly outline necessary parts for the student and instructor for scoring purposes, and the presentation of these parts reveal the student's comprehension of the statistics.

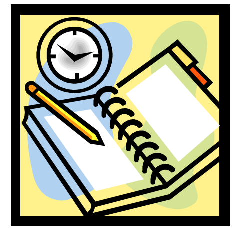
Project Timelines

It is helpful to both students and instructors to define a project timeline in advance. This timeline will identify when each phase of the project should happen and due dates for certain milestones along the way. Depending on the organization and parameters of the project, the full project cycle may take as little as 2 weeks or as much as 6 weeks. Following are suggested timelines with milestones for a 6-week project, a 4-week project, and a 2-week project.

2-week project	4-week project	6-week project
Week 1: • Identify variables & research question • Complete project plan and obtain approval • Collect data	Week 1: • Identify variables & research question • Complete project plan and obtain approval	Week 1: • Brainstorm ideas • Identify variables & research question
Week 2: • Analyze data • Complete report • Present results	Week 2: • Collect data	Week 2: • Complete formal project proposal
	Week 3: • Analyze data	Week 3: • Obtain approval • Collect data
	Week 4: • Complete report • Present results	Week 4: • Collect data
		Week 5: • Analyze data
		Week 6: • Complete report • Present results

A major difference in the timelines above is the degree of formality of the plan or proposal. The 6-week timeline specifies a project proposal, which may be a formal paper. The 4-week timeline requires the students to submit a project plan, which is often a prepared form that the student completes. The 2-week timeline requires a project plan form also, which may be less detailed than the 4-week plan form. Sample project plan forms are shown in Appendix T1.

Each of these proposals, regardless of the particular timeline, should include any survey a student intends to use. This is necessary and critical prior to data collection in order to prevent time-consuming student errors that may sabotage the project. The students may unwittingly violate rules of an institution's IRB, and this checkpoint facilitates compliance on the part of the student.



Another difference between the timelines is that they are dictated by the amount of time the instructor is willing to devote to the projects in class. Shorter timelines require more in-class time to implement project tasks. In the 4- or 6-week timeline, the only class time necessary is for brainstorming. This session also varies in length; it can be as simple as showing example projects in 5 – 15 minutes. The amount of time spent brainstorming is a matter of instructor preference.

The most important factor to consider in choosing a timeline is the alignment of the project phases to the timeline and structure of the course objectives. Ideally, the project phases and the in-class content should occur in tandem so that the work of the project complements the content covered in class meetings.

Independent vs. Guided Discovery

The descriptions above assume that the students will be assigned projects which require them to define their variables, research question, and data collection plan independently. In some cases, instructors want to assign a project with similar benefits to the student, but in a shorter timeframe or with more guidance.

Guided discovery projects give students a more structured framework within which to work by providing a general research question and general data collection plan, while still requiring the student to complete the project by collecting and analyzing the data and reporting the results. The student may also refine the research question within the context provided.

Appendix E contains some guided discovery project options.

References

The Buck Institute for Education & Boise State University. (n.d.). *What is project-based learning?* Retrieved March 24, 2011, from http://pbl-online.org/About/whatisPBL.htm

Donovan, M. S., Bransford, J., Pellegrino, J. (Eds.). (1999). *How people learn: Bridging research and practice*. Washington DC: National Academy Press.

Guidelines for Assessment and Instruction in Statistics Education [GAISE] (2005). *Guidelines for assessment and instruction in statistics education (GAISE) college report*. Alexandria, VA: American Statistical Association. Retrieved from http://www.amstat.org/education/gaise/GAISECollege.pdf

Hogg, R. V. (1991). Statistical education: Improvements are badly needed. *The American Statistician, 45*, 342-343.

Kilpatrick, J., Swafford, J., & Findell, B. (Eds.). (2001). *Adding it up: Helping children learn mathematics*. Washington DC: national Academy Press.

Landrum, R. E., & Smith, R. A. (2007) Creating syllabi for statistics and research methods courses. In D. S. Dunn, R. A. Smith, & B. C. Beins (Eds.), *Best practices for teaching statistics and research methods in the behavioral sciences* (pp. 45-57). Mahwah, NJ: Lawrence Erlbaum.

National Council of Teachers of Mathematics (2000). *Principles and standards for school mathematics.* Reston, VA: Author.

Roseth, C. J., Garfield, J. B., & Ben-Zvi, D. (2008). Collaboration in learning and teaching statistics. *Journal of Statistics Education, 16*(1). Retrieved from http://www.amstat.org/publications/jse/v16n1/roseth.html

Schoenfeld, A. (1987). What's all the fuss about metacognition? In A. H. Schoenfeld (Ed.), *Cognitive science and mathematics education* (pp. 189-215). Hillsdale, NJ: Erlbaum.

Thomas, J.W. (2000). *A review of project-based learning*. Retrieved July 20, 2011, from http://www.bie.org/index.php/site/RE/pbl_research/29.

Appendices

Appendix A Internet Data Sources

Appendix B Variables and Constructs

Appendix C Collection of Surveys

Appendix D Sample Student Projects

Appendix E Guided Discovery Project Assignments

Appendix T1 Project Plan Forms and Proposal Assignments

Appendix T2 Sample IRB and Permission Form

Appendix T3 Project Scoring Rubrics