Part Review Projects

Projects are longer exercises that require gathering information or producing data and that emphasize writing a short essay to describe your work. Many are suitable for teams of students.

PART I PROJECTS

Project I.1. Design your own sample survey. Choose an issue of current interest to students at your school. Prepare a short (no more than five questions) questionnaire to determine opinions on this issue. Choose a sample of about 25 students, administer your questionnaire, and write a brief description of your findings. Also write a short discussion of your experiences in designing and carrying out the survey.

(Although 25 students are too few for you to be statistically confident of your results, this project centers on the practical work of a survey. You must first identify a population; if it is not possible to reach a wider student population, use students enrolled in this course. How did you choose your sample? Did the subjects find your questions clear? Did you write the questions so that it was easy to tabulate the responses? At the end, did you wish you had asked different questions?)

Project I.2. Measuring. Is the right hand generally stronger than the left in right-handed people? You can crudely measure hand strength by placing a bathroom scale on a shelf with the end protruding, then squeezing the scale between the thumb and the four fingers. The reading of the scale shows the force exerted. Using several right-handed subjects, try to determine whether this method of measuring hand strength is reliable. Write an account of your findings. For example, did you find that subjects used different grips, so that careful instructions were needed to get a consistent way of measuring? Prepare written instructions for subjects.

Project I.3. Experimenting. After you or other members of your team have refined the measurement of hand strength in the previous project, carry out the matched pairs experiment of Project 2 with at least 10 subjects. Write a report that describes the randomization, gives the data, reports the differences in strength (right hand minus left hand), and says whether your small experiment seems to show that the right hand is stronger, on the average.

Project I.4. Describe a medical study. Go to the website of the *Journal of the American Medical Association* (http://jama.ama-assn.org). Unlike the *New England Journal of Medicine* (http://content.nejm.org), *JAMA* makes the full text of some articles freely available online. Select an article from the current issue or from a past issue in the last five years that describes a study whose topic interests you. Write a newspaper article that summarizes the design and the findings of the study. (Be sure to include statistical aspects, such as observational study versus experiment and any randomization used. News accounts often neglect these facts.)

Project I.5. Do you drive an SUV? Are staff at your college more or less likely than students to drive an SUV? Design and carry out a study to find out, and write a report describing your design and your findings. You must first be clear about what "SUV" means so that each car is clearly an SUV or not. Then you must locate a suitable sample of cars—perhaps from a student parking area and a staff parking area. If the areas are large, you will want to sample the cars rather than look at all of them. Consider using a systematic sample (Exercise 4.27, page 82).

Project I.6. Data ethics. Locate a news discussion of an ethical issue that concerns statistical studies. Write your own brief summary of the debate and any conclusions you feel you can reach.

Here is an example of one way to approach this project. Testing new drugs on human subjects continues to be an ongoing concern in medical studies. How does one balance the need for knowledge with protecting subjects from possible harm? Searching the archives at the website of the *New York Times* (www.nytimes.com) for "experiments and ethics" (to use the *New York Times* search engine you should enter "+experiments +ethics"), one finds many articles, including a promising one in the issue of May 7, 2015. To read the article, you may be able to link to it directly online, but if not, you will have to either pay a fee or go to the library. You could also try searching the Web with Google. We entered "drugs and human guinea pigs" and found some possible leads.

Project I.7. Measuring income. What is the "income" of a household? Household income may determine eligibility for government programs that assist "low-income" people. Income statistics also have political effects. Political conservatives often argue that government data overstate how many people are poor because the data include only money income, leaving out the value of food stamps and subsidized housing. Political liberals reply that the government should measure money income so it can see how many people need help.

You are on the staff of a member of Congress who is considering new welfare legislation. Write an exact definition of "income" for the purpose of determining which households are eligible for welfare. A short essay will be needed. Will you include nonmoney income such as the value of food stamps or subsidized housing? Will you allow deductions for the cost of child care needed to permit the parent to work? What about assets that are worth a lot but do not produce income, such as a house?

PART II PROJECTS

Project II.1. Statistical graphics in the press. Graphs both good and bad fill the news media. Some publications, such as *USA Today*, make particularly heavy use of graphs to present data. Collect several graphs (at least five) from newspapers and magazines (not from advertisements). Include some graphs that, in your opinion, represent good style and some that represent poor style or are misleading. Use your collection as examples in a brief essay about the clarity, accuracy, and attractiveness of graphs in the press.

Project II.2. Roll your own regression. Choose two quantitative variables that you think have a roughly straight-line relationship. Gather data on these variables and do a statistical analysis: make a scatterplot, find the correlation, find the regression line (use a statistical calculator or software), and draw the line on your plot. Then write a report on your work. Some examples of suitable pairs of variables are the following:

(a) The height and arm span of a group of people.

(b) The height and walking stride length of a group of people.

(c) The price per ounce and bottle size in ounces for several brands of shampoo and several bottle sizes for each brand.

Project II.3. High school dropouts. Write a factual report on high school dropouts in the United States. The following are examples of questions you might address: Which states have the highest percentages of adults who did not finish high school? How do the earnings and employment rates of dropouts compare with those of other adults? Is the percentage who fail to finish high school higher among blacks and Hispanics than among whites?

The Census Bureau website will supply you with data. Go to

www.census.gov/hhes/socdemo/education/.

Project II.4. Association is not causation. Write a snappy, attention-getting article on the theme that "association is not causation." Use pointed but not-too-serious examples like those in Example 6 (page 353) and Exercise 15.30 (page 361) of Chapter 15, or this one: there is an association between long hair and height (because women tend to have longer hair than men but also tend to be shorter), but cutting a person's hair will not make him or her taller. Be clear, but don't be technical. Imagine that you are writing for high school students.

Project II.5. Military spending. Here are data on U.S. spending for national defense for the fiscal years between 1940 and 2015 from www.whitehouse.gov/omb/budget/. See the pdf file available by clicking on *Historical Tables.* Look in Section 3 of this pdf file. The units are billions of dollars (this is serious money).

Year:	1940	1945	1950	1955	1960	1965	1970	1975
Military spending:	1.7	83.0	13.7	42.7	48.1	50.6	81.7	86.5
Year:	1980	1985	1990	1995	2000	2005	2010	2015
Military spending:	134.0	252.7	299.3	272.1	294.5	495.3	693.5	597.5

Write an essay that describes the changes in military spending in real terms during this period from just before World War II until a decade after the end of the cold war. Do the necessary calculations and write a brief description that ties military spending to the major military events of this period: World War II (1941–1945), the Korean War (1950–1953), the Vietnam War (roughly 1964–1975), the end of the cold war after the fall of the Berlin Wall in 1989, and the U.S. war with Iraq (beginning in March 2003). You may want to look at years not included in the table to help you as you write your essay.

Project II.6. Your pulse rate. What is your "resting pulse rate"? Of course, even if you measure your pulse rate while resting, it may vary from day to day and with the time of day. Measure your resting pulse rate at least six times each day (spaced throughout the day) for at least four days. Write a discussion that includes a description of how you made your measurements and an analysis of your data. Based on the data, what would you say when someone asks you what your resting pulse rate is? (If several students do this project, you can discuss variation in pulse rate among a group of individuals as well.)

Project II.7. The dates of coins. Coins are stamped with the year in which they were minted. Collect data from at least 50 coins of each denomination: pennies, nickels, dimes, and quarters. Write a description of the distribution of dates on coins now in circulation, including graphs and numerical descriptions. Are there differences among the denominations? Did you find any outliers?

PART III PROJECTS

Project III.1. A bit of history. On page 409, we said, "The systematic study of randomness ... began when seventeenth-century French gamblers asked French mathematicians for help in figuring out the 'fair value' of bets on games of chance." Pierre de Fermat and Blaise Pascal were two of the mathematicians who responded. Both are interesting characters. Choose one of these

men. Write a brief essay giving his dates, some anecdotes you find noteworthy from his life, and at least one example of a probability problem he studied. (A Web search on the name will produce abundant information. Remember to use your own words in writing your essay.)

Project III.2. Reacting to risks. On page 418, we quoted a writer as saying, "Few of us would leave a baby sleeping alone in a house while we drove off on a 10-minute errand, even though car-crash risks are much greater than home risks." Take it as a fact that the probability that the baby will be injured in the car is very much higher than the probability of any harm occurring at home in the same time period. Would you leave the baby alone? Explain your reasons in a short essay. If you would not leave the baby alone, be sure to explain why you choose to ignore the probabilities.

Project III.3. First digits. Here is a remarkable fact: the first digits of the numbers in long tables are usually *not* equally likely to have any of the 10 possible values 0, 1, 2, 3, 4, 5, 6, 7, 8, and 9. The digit 1 tends to occur with probability roughly 0.3, the digit 2 with probability about 0.17, and so on. You can find more information about this fact, called "Benford's law," on the Web or in two articles by Theodore P. Hill, "The Difficulty of Faking Data," *Chance*, 12, No. 3 (1999), pp. 27–31; and "The First Digit Phenomenon," *American Scientist,* 86 (1998), pp. 358–363. You don't have to read these articles for this project.

Locate at least two long tables whose entries could plausibly begin with any digit. You may choose data tables, such as populations of many cities, the number of shares traded on the New York Stock Exchange on many days, or mathematical tables such as logarithms or square roots. We hope it's clear that you can't use the table of random digits. Let's require that your examples each contain at least 300 numbers. Tally the first digits of all entries in each table. Report the distributions (in percentages) and compare them with each other, with Benford's law, and with the "equally likely" distribution.

Project III.4. Personal probability. Personal probabilities are personal, so we expect them to vary from person to person. Choose an event that most students at your school should have an opinion about, such as rain next Friday or a victory in your team's next game. Ask many students (at least 50) to tell you what probability they would assign to rain or a victory. Then analyze the data with a graph and numbers—shape, center, spread, and all that. What do your data show about personal probabilities for this future event?

Project III.5. Making decisions. Exercise 20.12 (page 471) reported the results of a study by the psychologist Amos Tversky on the effect of wording on people's decisions about chance outcomes. His subjects were college students. Repeat Tversky's study at your school. Prepare two typed cards. One says:

You are responsible for treating 600 people who have been exposed to a fatal virus. Treatment A has probability 1-in-2 of saving all 600 and probability 1-in-2 that all 600 will die. Treatment B is guaranteed to save exactly 400 of the 600 people. Which treatment will you give?

The second card says:

> *You are responsible for treating 600 people who have been exposed to a fatal virus. Treatment A has probability 1/2 of saving all 600 and probability 1/2 that all 600 will die. Treatment B will definitely lose exactly 200 of the lives. Which treatment will you give?*

Show each card to at least 25 people (25 different people for each, chosen as randomly as you can conveniently manage and chosen from people who have not studied probability). Record the choices. Tversky claims that people shown the first card tend to choose B, while those shown the

second card tend to choose A. Do your results agree with this claim? Write a brief summary of your findings: Do people use expected values in their decisions? Does the frame in which a decision is presented (the wording, for example) influence choices?

Part IV PROJECTS

Project IV.1. Reporting a medical study. Many of the major articles in medical journals concern statistically designed studies and report the results of inference, usually either P -values or 95% confidence intervals. You can find summaries of current articles on the websites of the *Journal of the American Medical Association* (jama.ama-assn.org) and the *New England Journal of Medicine* (www.nejm.org). A full copy may require paying a fee or visiting the library. Choose an article that describes a medical experiment on a topic that is understandable to those of us who lack medical training—anger and heart attacks and fiber in the diet to reduce cholesterol are two examples used in Chapters 21 and 22. Write a two-paragraph news article explaining the results of the study.

Then write a brief discussion of how you decided what to put in the news article and what to leave out. For example, if you omitted details of statistical significance or of confidence intervals, explain why. What did you say about the design of the study, and why? News writers must regularly make decisions like these.

Project IV.2. Use and abuse of inference. Few accounts of really complex statistical methods are readable without extensive training. One that is, and that is also an excellent essay on the abuse of statistical inference, is "The Real Error of Cyril Burt," a chapter in Stephen Jay Gould's *The Mismeasure of Man* (W. W. Norton, 1981). We met Cyril Burt under suspicious circumstances in Exercise 9.27 (page 201). Gould's long chapter shows that Burt and others engaged in discovering dubious patterns by using complex statistics. Read it, and write a brief explanation of why "factor analysis" failed to give a firm picture of the structure of mental ability.

Project IV.3. Roll your own statistical study. Collect your own data on two categorical variables whose relationship seems interesting. A simple example is the sex of a student and their political party preference. A more elaborate example is the year in school of a college undergraduate and his or her plans following graduation (immediate employment, further study, take some time off, ...). We won't insist on a proper SRS.

Collect your data and make a two-way table. Do an analysis that includes comparing percentages to describe the relationship between your two variables and using the chi-square statistic to assess its significance. Write a description of your study and its findings. Was your sample so small that lack of significance may not be surprising?

Project IV.4. Car colors. We have heard that more white cars are sold in the United States than any other color. What percentage of the cars driven by students at your school are white? Answer this question by collecting data and giving a confidence interval for the proportion of white cars. You might collect data by questioning a sample of students or by looking at cars in student parking areas. In your discussion, explain how you attempted to get data that are close to an SRS of student cars.