

Statistics 103

Probability and Statistical Inference

Instructions for Data Analysis Project

You've learned lots about doing statistical analyses. It's time to work without a net....

Due Dates

Project Proposal due date: **February 21 (or any time before Spring Break).**

Completed project due date: **April 19**, presented at poster sessions in lab sections.

General Description

For the data analysis project, you address some questions that interest you with the statistical methodology we learn in Statistics 103. You choose the question; you decide how to collect data; you do the analyses. The questions can address almost any topic (although I have veto power), including topics in economics, psychology, sociology, natural science, medicine, public policy, sports, law, etc.

The project requires you to synthesize all the material from the course. Hence, it's one of the best ways to solidify your understanding of statistical methods. Plus, you get answers to issues that pique your intellectual curiosity.

You should work in groups of two to three people on the project. Larger or smaller groups must be granted special permission from the instructor. You can work with people in different lab sections than yours.

Your project will be presented in a poster session during the last week of lab sections. In a poster session, each group makes visual materials that explain the project. Then, people wander around looking at the posters and talking to the presenters, thereby learning about the various projects. Poster sessions are extremely common at professional conferences in many disciplines, including statistics. In our poster session, some members of each group are stationed at the poster to answer questions, while the others wander around to examine the projects. The poster-sitters and wanderers switch off after the wanderers have examined all the posters.

There is no formal write-up of your project, i.e., no term paper is written. Each person must present or be part of a presentation of their group's project. The poster is handed in and graded. Your presentations factor in to the grade. You also will anonymously evaluate each other's contributions to the overall project.

You should get started on the project as early as possible, particularly in thinking about procuring data and collecting background information. Keep in mind that by the end of lectures, you will have learned many statistical techniques, such as hypothesis testing, confidence intervals, and regression. These techniques will help you address your question of interest.

Some ideas for projects

The most important aspects of any statistical analysis are stating questions and collecting data. Hence, to get the full experience of running your own study, the project requires you to analyze data that you collect. It is not permissible to use data sets that have been put together by others. You are permitted to collect data off of the web; however, you must be the one who decides on the analyses and puts the data set together.

Good projects begin with very clear and well-defined hypotheses. You should think of questions that interest you first, then worry about how to collect and analyze data to address those questions. Generally, vague topics lead to uninteresting projects. For example, surveying Duke undergraduates to see which sex studies more doesn't yield a whole lot of interesting conclusions. On the other hand, it would be interesting to hypothesize why men or women study more, and then figure out how to collect and analyze data to test your hypotheses.

Below is a list of some successful project topics that have been done by past statistics students. This isn't a list that you have to pick from; in fact, you'll get a higher grade if you come up with something else. Instead, consider the list a tool for generating ideas.

1. Are men more likely than women to help someone who has dropped his or her books? Does the sex of the book dropper matter?
2. Does having the pictures on puzzle pieces shorten the time to complete the puzzle relative to not having the pictures?
3. Does eating popcorn affect people's enjoyment of movies?
4. Does drinking caffeine affect students' performance on tests?
5. Does wearing shoes affect the height of a vertical jump?
6. Does the quality of Duke students' relationship with their freshman roommate affect the quality of their overall experience at Duke?
7. Does the Chronicle fairly represent all students' voices at Duke?
8. Does birth order affect academic success at Duke?
9. Do actors' races affect which television programs Duke students are willing to watch?
10. What is more important to Duke students when choosing a major: interest in the subject, career aspirations, family influence, or ability in the subject?
11. Do FOCUS students at Duke eat, sleep, and go to parties with different frequencies than non-FOCUS students?
12. Are people like the descriptions of their horoscope sign?
13. Are people rational when playing prisoner's dilemma games?
14. Is team payroll related to winning percentages in professional sports?
15. Can we predict the order of the NFL draft based on characteristics of the players?
16. Do the results of federal elections have an effect on stock prices?
17. Is there a correlation between female empowerment and AIDS prevalence in nations across the world?
18. Do certain subpopulations get mammograms more frequently than others?
19. Are members of certain subpopulations (e.g., racial, ethnic, or educational backgrounds) more likely to receive the death penalty?
20. Are policies that reduce governmental debt also associated with reduction in quality of life?

It is important to be thoughtful about, and provide an adequate description of, the methods and design of the study. Report on the possible biases associated with your data collection. You also need to be realistic in planning your research design: can you carry out what you have planned within a reasonable time period and investment of your own energy? The quality of the final product is what counts, not just the amount of perspiration that went into it! Finally, you should make use of the concepts and methods learned in this course, and not just general knowledge, in planning and completing this type of project.

Practical Advice: It is often easier to collect accurate experimental data than accurate survey data. Nonresponse tends to be less of an issue with projects based on experiments than with those based on surveys. I strongly encourage you to consider experiments as opposed to surveys. For those who want to do surveys, consider using students in dorms or certain courses as target populations. Make every effort to get a random sample, and try to keep track of the characteristics of nonrespondents. You will have nonresponse; your project won't be penalized for nonresponse as long as you document it and hypothesize how it might affect your results.

Project proposal

Your group should **HAND IN ONE PROJECT PROPOSAL** (with all group members' names and section leaders on it) by the proposal due date given above. The proposal is a page or so describing what you plan to do. Be as specific as possible, describing what question you want to investigate and generally how you plan to obtain data. The instructor and TAs will return the proposals to you with comments. The more detailed your proposal, the better feedback you get! Your proposal should address the following questions:

- What is the topic of your project?
- What are the main issues or problems you plan to address?
- What are your plans for obtaining background information (if needed) about your project?
- Describe the data that you plan on using or collecting, including the variables measured. You don't have to give a detailed version of your data collection design; you will hand in detailed design plans on the design due date given above.
- What questions and/or concerns do you have about your project?

The project proposal is not graded. It exists primarily for you to get feedback on your project idea.

Project grading guidelines

You will be graded by your TAs or instructor. Graders will be looking for the following characteristics:

1. **Consistency:** Did you answer your question of interest?
2. **Clarity:** Is it easy for your reader to understand what you did and the arguments you made?
3. **Relevancy:** Did you use statistical techniques wisely to address your question?
4. **Interest:** Did you tackle a challenging, interesting question (good), or did you just collect descriptive statistics (bad)?

Some suggestions for scoring high on these criteria, and suggestions you should keep in mind whenever you write anything, are the following:

1. Know your audience. In this case, you should design the poster for an audience of Statistics 101 students. You may want to have your classmates examine the poster for clarity.
2. State your question up front, and use statistics to help answer it. The statistics should not drive the question; the question should drive the statistics.
3. Don't just collect data and publish it, rather have a specific question in mind. Otherwise, you wind up being hard-pressed to come up with something challenging and interesting.
4. Most importantly, talk to your instructor and TAs for advice. You can ask them, for example, about your planned methods of analysis and see what they think.
5. Be selective with computer output to help clarity.

If you are using techniques we learned in class, you do not have to re-explain the techniques. That hurts clarity. If you are using techniques that we did not cover in class, you should definitely explain the

techniques. That is clarity!

Guidelines for making an effective poster

An effective poster communicates your project in a clear and concise fashion. The poster should address the following six points:

1. **Statement of the problem:** Describe the questions you address and any key issues surrounding the questions.
2. **Data collection:** Explain how you collect data. Include any questions you asked. Also, include response rates.
3. **Analyses:** Describe the analyses you did. Be ready to explain why you believe these methods are justified.
4. **Results:** Present relevant descriptive statistics (e.g., number of men and women surveyed, if that is important). Include tables or graphs that support your analyses (be judicious here--too many tables and graphs hurts the clarity of your message).
5. **Conclusions:** Answer your question of interest.
6. **Discussion:** What implications do your results have for the population you sampled from? What could be done to improve the study if it was done again? What types of biases might exist?

A useful method of poster presentation is to tape several sheets of paper describing your project to a large poster board (or just to the wall if you don't have poster board). You should strive to make the poster clear. Avoid unnecessary clutter, and don't put too much information on any one page. Think about what you would want to see on other people's posters as guidance for creating your poster.

Procedures for when group members are not contributing their fair share

Each group should spread the work among members so that everyone shares in the project. If some group members do not contribute their assigned workload, or are unwilling to take on work, your group may petition to have such group members dropped from the group. The process of this petition proceeds as follows:

- 1) Send an e-mail to the instructor explaining how the group members have not contributed adequately. **ALL MEMBERS OF THE GROUP MUST BE SENT THIS E-MAIL.** This is to ensure that everything is done openly.
- 2) The instructor will arrange a meeting with the group. Subjects of a petition who fail to attend the arranged meeting will be dropped from the group.
- 3) At this meeting, the instructor will make a decision on the petition.

These petitions can be made until April 1. After this date, groups will not be split up. Students who have been dropped from groups must find another group or get special permission to work alone from the instructor. After one of these meetings, any group member who does not contribute after promising to do so will be dropped from the group.
