

STAT 301 Statistics I Winter 2011

Project Assignment

This project provides another way to demonstrate that you have learned important concepts and skills related to collecting and analyzing data to address interesting research questions. You may work with one partner on this project assignment, submitting one report with both names, provided that both of you contribute substantially to all aspects of the project.

For this project, you are to investigate a topic of your own choosing that involves comparing two groups. You must use either random assignment or random sampling in your data collection.

The steps involved are to:

1. Formulate a research question that can be addressed by conducting such a study;
2. Develop a plan for collecting data to investigate the research question;
3. Collect data according to that plan;
4. Conduct a descriptive analysis of the data with graphical and numerical summaries;
5. Perform an inferential analysis of the data with simulation and probabilistic methods;
6. Write a detailed report that describes all aspects of your investigation and summarizes its conclusions.

Guidelines and advice:

- Your study must involve comparing two groups.
- The response variable can be either categorical or quantitative. (We will learn how to analyze data with a quantitative response variable in the last third of the course.)
- You must have at least 15 experimental/observational units in each group.
 - More is better, because it gives your test more power (a better chance of detecting a difference between the groups if there really is one).
 - Your experimental/observational units do not have to be human beings!
- Random assignment is usually much easier to implement than random sampling.
 - If you conduct an observational study, it must involve *genuine* random sampling.
 - You may not use *gender* as the explanatory variable.
- You may also collect data on a third variable (such as gender) if you'd like.
- Your report must be word-processed, with computer output integrated into the report as appropriate.
- Make sure that it's feasible to collect the data over the next few weeks.
- Select a topic and research question that are genuinely interesting to you.

Timeline:

- Submit a statement of your topic, research question, and detailed data collection plan (to me via email) by Wed March 2. Be sure to describe how your data collection plan makes use of randomness. Also clearly identify your observational units and your explanatory and response variables. Wait for feedback from me before you collect data.
- Collect your data, enter into computer, submit data file to me via email by Fri March 11.
- Submit final report by 4pm on Wed March 16.

Your report will be graded on:

- Creativity of the research question
- Quality of the data collection plan and implementation
- Correctness of the analyses
- Thoroughness of the analyses
- Validity of the conclusions drawn
- Quality of the written presentation

To stimulate your creative energy in thinking about possible topics, some examples of topics and research questions that have been investigated by previous students include:

- Is balsa wood more elastic after it has been immersed in water?
- Do students perform better on a memory task while listening to classical music than pop music?
- Do students express more dissatisfaction with their roommate depending on how the question is worded?
- Are people more likely to let you borrow their cell phone to make a call depending on how you are dressed?
- Do students do worse at pinning the tail on a donkey if they are spun around first?
- Does asking a stranger to watch your laptop increase the chance that they will stop a would-be thief?
- Does time to answer the door depend on whether a doorbell was rung or the door was knocked upon?
- Is a customer more likely to respond to survey questions from a store employee than from a non-employee?

You should also feel free to extend some of the research questions from studies that we have investigated in class.

Project report:

Some general advice:

- Give your report an appropriate name, preferably a catchy name that gives an indication of what your project was about.
- Make sure that it is word-processed and well-formatted.
- Revise and proof-read draft reports multiple times to make sure that it reads well.
- Provide labels for all tables and graphs. Then refer to these labels in your text. For example, you might write: "Figure 1 shows a dotplot of the results, which reveals that ..."
- Organize your report into sections.

Advice about what to include in each section follows:

1. Introduction

- Describe your research question.
- Explain what motivated you to investigate this question.
- Describe any expectations you started with, concerning what you might learn about your research question.

- Indicate what your explanatory and response variables are, and classify them as categorical or quantitative.

2. *Data Collection*

- Provide lots of detail about how you actually collected your data, enough detail that the reader will understand exactly what your data collection entailed.
- Specify whether your study is observational or experimental.
- Indicate what your project's observational/experimental units are.
- Make clear how randomness played a role in your data collection plan, either random sampling or random assignment or both.
- Give much detail about how you selected the observational/experimental units.
- For whichever type of randomness you used (random sampling or random assignment), give much detail about how you actually implemented that randomness.
- Describe how you measured/recorded the values of your variables.
 - If you used a survey, provide a copy of the survey and mention how you decided on how to word your questions.
 - If you measured something, describe how you collected the measurements.
- Describe any efforts you made to control for other variables that could be related to what you are studying.

3. *Descriptive Analysis*

- Present and describe what you learn from graphical displays of your data
 - Segmented bar graph for categorical response
 - Dotplots and/or boxplots and/or histograms for quantitative response
- Present and describe what you learn from numerical summaries of your data
 - Conditional proportions and/or relative risk and/or odds ratio (categorical)
 - Mean and/or SD and/or five-number summary (quantitative)

4. *Inferential Analysis*

- Conduct an appropriate significance test of your research question
 - Randomization-based test and/or normal/*t*-based test
- For a randomization-based test
 - State the appropriate null and alternative hypotheses, being sure to define any symbols that you introduce.
 - Describe the assumption behind the simulation analysis.
 - Present a graph of the simulation results and describe what it reveals.
 - Report an approximate p-value.
 - Summarize and explain your conclusion.
- For a normal/*t*-based test
 - State the appropriate null and alternative hypotheses, being sure to define any symbols that you introduce.
 - Check the technical conditions of the procedure and comment on whether they appear to be satisfied.
 - Calculate the test statistic and p-value, and present these with an appropriate graph.
 - Summarize and explain your conclusion.

- Produce a confidence interval for a relevant parameter
 - Also interpret what the interval reveals, being sure to describe the parameter clearly.

5. *Conclusion*

- Summarize what your analysis has revealed about your research question.
 - Did you find a statistically significant difference between the groups?
 - Can you legitimately draw a cause-effect conclusion?
 - To what population is it reasonable to generalize your findings?
- Describe any limitations of your study and your conclusions.
- Comment on ways in which your investigation went as you expected and also on ways that surprised you.
- Mention follow-up questions that you might like to investigate if you had more time and resources.