

MIT OpenCourseWare  
<http://ocw.mit.edu>

11.220 Quantitative Reasoning & Statistical Methods for Planners I  
Spring 2009

For information about citing these materials or our Terms of Use, visit: <http://ocw.mit.edu/terms>.

# Quantitative Reasoning and Statistical Methods

## Written Assignment #3: Data Analysis Exercise

Ezra Haber Glenn, AICP

Proposal Due: 4/16/09  
Final Project Due: 5/14/09

### 1 The Assignment

For this assignment you are asked to use everything you have learned in class to gather and analyze your own quantitative information, and to present your findings in a clear, thoughtful, and honest report. The actual choice of topic is entirely yours, although you must first describe it in a proposal meeting the approval of your TA.

#### 1.1 Sample Topics

To help get you started, here are some sample topics (although choosing any of these without good reason may be taken as evidence of lack of creativity on your part):

- An evaluation of changes to the Community Development Block Grant funding over the past 30 years, with special attention to the effects these changes have had on different types of cities.
- An analysis of the effects of local wetland protection laws on rates of development, using a statewide data set for New Jersey during the years 1990-2005.
- A discussion of the (hypothesized) disconnect between international funding levels for prevention activities and the damage done by different diseases.
- A exploration of voting patterns for Asian-Americans in California, to investigate whether there is evidence for more “ethnic/affinity voting” occurring in local elections (as contrasted to state/federal elections) for this population.

You might also want to look back over the readings we have had in class, and those articles/reports covered in the second written assignment, for additional inspiration.

## 2 Components

This assignment actually consists of a number of different sub-components, including a proposal, a dataset, a written report, and some supporting documentation and metadata.

### 2.1 Proposal

Before you can begin work on your final project, you *must* submit a proposal on the idea to your TA. In your proposal, please describe the following:

**Problem Statement** What questions are you seeking to answer through your project, and why do they matter? How, specifically, will *quantitative information* be used to address these particular issues?

**Hypothesis/Argument** What, if any, is the hypothesis you are hoping to test or argument you are expecting to assert? (Having a research hypothesis is highly recommended; strictly speaking, you could manage to write an interesting and insightful paper without one—and we are not looking for a full-blown scientific article or anything—but the language of hypotheses and scientific research can be helpful to structure your argument even along less rigorous lines.) Along with your research hypothesis, you might want to think about issues such as a “null hypothesis,” dependent and independent variables, confounding and lurking variables, internal and external validity, sources of bias and noise, and all the other aspects of measurement theory, scientific method, and research design we spoke of in the first half of the class.

**Data Sources** What data sources have you identified, and what issues (if any) do you anticipate in terms of acquiring them, converting formats, cleaning, correcting, modifying, joining sets, and filling gaps. What concerns might you have regarding measurement instruments, bias in sampling, missing values, and the like, and how do you intend to deal with these problems? (You do not need to say everything that could be said in this section, but we want to use this as a placeholder to remind you of the significant amount of time you may need to spend gathering the data and “making it your own.”)

Your proposal does not need to be very long—1 or 2 pages should be fine—but it should demonstrate that you have (a) identified an interesting topic that will allow you to demonstrate the full range of skills you have mastered in the class, (b) focused your thinking to a level that will be manageable given the scale of this assignment, and (c) begun to think about what information you will use and how you will use it.

Please note that timely submission of the proposal—and thoughtful consideration of the aspects of the project at this early stage—will count for up to 10% of the final grade (in addition to making your finished paper better). If

your proposal is not received by April 16, your TA will not be allowed to give you a grade for this portion.

## **2.2 Final Report**

Your final report should describe the work you did and the findings of your investigation. Although there is no set format for the final report, we recommend that you include some sort of Summary, a short Background section setting your project in context, a section on Methodology, and a section on Key Findings. Importantly, given the time and space constraints, you should not feel the need to be exhaustive or definitive—save that for your thesis. The purpose of this project is simply to get you started working with quantitative data related to your work in planning, development, design, and/or policy, and to give you a chance to show off everything you’ve learned.

### **2.2.1 A note on graphics, tables, and other ways to present data**

Depending on your topic, your skills, and your interests, some of you may be content to include graphs and tables directly into your report, while others may choose to provide them in some other format (e.g., embedded in an on-line application, plotted on a map, published as an animated flip-book, etc.). Either is fine—and we are happy to see you explore creative ways to present quantitative data—but please be mindful of the fact that we will need to read and evaluate approximately 60 reports in a very short period of time, and would prefer to not have to manage too many different pieces.

## **2.3 Supporting Materials**

Beyond the actual written report and related graphs and tables, you are also responsible for submitting the following items by the deadline. Please note that the preparation and testing of these components may take significant time, especially if you forgot to document your work as you went and now need to figure out what you did, and we strongly recommend that you do not leave these parts for the end.

### **2.3.1 Data and Metadata**

Along with your final report, you must submit a clean and self-explanatory (i.e., well-documented with metadata) dataset. The easiest way to do this would be to provide the data itself (saved as a plain-text “.csv” file, a Stata “.dta” file, or an R “.Rdata” file, depending on the program you used), along with a text document containing the metadata (ideally saved as a plain-text “.txt” file as well, which is good practice for metadata and other documentation). In your metadata document, be sure to include the following:

- sources (and dates) of all data you gathered, and any modifications you made;

- the shape (how many observations? how many variables?) and format of the data;
- names, types, and units of measurements for all variables (and guidance for interpreting any codes or factors—for example, “SEX -- coded 0 for male and 1 for female”);
- an explanation of how you dealt with missing values;
- notes and references to any other documents or sources that might be helpful in interpreting this data (e.g., a description of a particular data transformation technique or measurement scale you used); and
- contact information for yourself, so someone with a question a few years from now might be able to find you to ask about it (optional).

### 2.3.2 Scripts/Do-Files

Throughout the class, we have stressed the importance of documenting the steps you undertake in quantitative analysis, both for your own benefit as well as for others who may be interested in critiquing (or building on) your work. Along with your final report, we are asking you to submit a clean and well-commented script (or in Stata-parlance, a “do-file”) capable of recreating your analysis and producing whatever tables, plots, statistical tests, models, and other analyses you conducted.

Importantly, you do *not* need to provide scripts/do-files for *all* of the early steps of your project (e.g., the data-cleaning, transformations, variable creation, exploratory analysis, etc.), but only for the final elements you used in your report. For this you can assume that we are in possession of your finished dataset and we will run the scripts on that. (In fact, we should be in possession of this data, and we will use it to check your scripts.) Typically, a script will load the data somewhere in the early steps, so don’t forget to do this.

If you are still unclear on the concept of scripts and/or do-files, I recommend the following resources:

- For Stata:
  - <http://www.cpc.unc.edu/services/computer/presentations/statatutorial/example6.html>
  - <http://www.cpc.unc.edu/services/computer/presentations/statatutorial/example18.html> (including a sample do-file)
- For R:
  - [http://cran.r-project.org/doc/contrib/Lemon-kickstart/kr\\_script.html](http://cran.r-project.org/doc/contrib/Lemon-kickstart/kr_script.html)

Before you submit all of your materials, we recommend you test out your script and data file on another computer, to make sure that it will run. (Hint: if you used any special libraries, plugins, or packages, make sure your script installs and loads them as well.) This will also help you make sure that you are using the correct file path to access your data (you should be able to assume that the data will be in the “working directory” and just use a “relative path” to it).

*Do not forget to use comments throughout your script. This will help you organize your work as you go, and will also help us figure out what you were trying to do if anything goes wrong.*

## 2.4 Boring details

The actual written report should be as long as you need it to be, but 6–8 pages is a good guide, plus graphics/tables, plus the other required components mentioned above. The proposal is due by 4:00 pm on 4/16/09. The final paper and supporting materials are due by 4:00 pm on 5/14/09.

One copy of your written report, including any printed tables and graphics, must be submitted *in hard copy* to your TA. In addition, please upload an electronic copy of your report, along with all supplementary material.

Be sure not to forget your dataset and your metadata description. Please name each of your files with a descriptive title, starting with your first initial and last name (e.g., “`eglenn_final_report.pdf`,” “`eglenn_dataset.Rdata`,” “`eglenn_metadata.txt`”).

## 3 Resources

You are responsible for whatever other outside reading you feel is necessary to address your topic and put it into a planning, policy, development, design, or social science context. Keep in mind, however, that we are not all that interested in grading this aspect of what you do—that is for you, your advisor, and your other professors to worry about. To the extent possible, we will be limiting our review to the aspects of the topic that relate to the work we have done in the class: measurement and data assembly; hypothesis testing and data analysis; organization and presentation of complex quantitative information; dealing with estimates, forecasts, and uncertainty; writing clearly and honestly about your findings; and the like. Certainly draw on outside sources to frame your investigation, but remember that the goal here is really for you to demonstrate some technical mastery of statistical methods and quantitative reasoning, not to impress us with your outside research.<sup>1</sup>

---

<sup>1</sup>The one exception here is the resourcefulness you demonstrate in identifying data that is appropriate to your topic—this sort of research, detective work, and evaluative judgement is related to the subject of the course.