

## FINAL PROJECT

---

### The goal

The goal of the final project is for you to complete a data analysis from beginning (scientific background and motivation) to end (conclusions and scientific implications), which includes most importantly presenting your results to a *non-statistical* audience. You will need to find a data set; no two people may work on the same data set. I encourage you to consider a data set of your own from your own field of research or RA work. I have several data sets available on the class web site, but they are not necessarily very interesting; look elsewhere first.

The methods used or explored must be those covered in this course or an extension of those covered in this course.

**Each student must discuss with me on or before 6 November her/his proposed project.** Send a project proposal to me by email (just a few sentences of description of the data and the scientific questions to be answered) and I will let you know whether or not it is acceptable.

### The structure of the written report

Your written report should be structured like a short report to be submitted to a journal in your discipline or a medical/public health discipline of interest to you. (Two exceptions: I do not expect a full literature review in your Background section. I do not expect you to include an abstract.) A strongly suggested outline is:

**Background/Objectives.** What are the objectives of or motivation for the work? What background science is needed for the audience to understand the work? Are there specific hypotheses to be addressed?

**Methods.** What are the variables of interest and how were the data collected? What was the sampled population? What technique(s) did you use to analyze the data: exploratory/descriptive, modeling, and diagnostics? BRIEFLY describe these; do NOT describe every model you considered!

**Results.** What do the data look like? What are the main results of your analyses? How do you interpret the results? How do the results relate to your objectives?

**Discussion.** Summarize the main points of your talk. Discuss any scientific or statistical problems with the data, the analyses, or the interpretation. Mention any future research directions.

You should strenuously refrain from technical statistical discussions. **The focus should be on the science, not the statistics.** An excellent report clearly and concisely communicates the

results and their implications to the reader. A clearly presented simple analysis is preferable to a poorly justified or explained complex analysis.

If you are not experienced in writing reports like this, spend a little time browsing the American Journal of Public Health, especially papers there that are slightly more statistically sophisticated (i.e., papers that used statistics more complex than two-sample t-tests and contingency table  $\chi^2$ -tests).

### Preparing the written report

You have SIX WEEKS to plan, carry out analyses for, and prepare this report. **You should spend a considerable amount of time on preparing the report AFTER your analyses are done.** Reports must be typed using double spacing, at least 1 inch margins, and 11pt or bigger font size. **They may not be longer than six pages of text.** You may include 2-3 tables and/or figures, and any necessary reference list, in addition to the six pages.

### Grading

The project is worth in total 50 points: **35 points for the initial submission** (due Thur 29 Nov), and **15 points for the final revised submission** (due Mon 17 Dec). Yes, you read that correctly; I will read and grade your submission, hand it back, and expect you to revise it and resubmit it.

The grading algorithm for the first submission will follow these guidelines:

- **Technical Details** (10 points) Is the analysis problem or research focus of the report formulated clearly using appropriate terminology? Are the appropriate statistical tools used properly? Are the assumptions checked when possible? Are the research or analysis conclusions justified?
- **Thoroughness** (10 points) Is there enough background science given so that a scientist not experienced in your field can understand the general ideas of your study, and the implications of your study for your population of interest? Was anything obvious overlooked, for example, a too-simple procedure performed in place of a more appropriate procedure, an assumption not checked, conclusions not stated for each objective? Is the work of others appropriately cited, and are references for those cited materials included? Are analysis interpretations and implications described in plain English (no statistical jargon)?
- **Clarity** (8 points) Is the report easy to understand? Does it follow a logical structure, both within and across sections? Is it concise and clearly written? Is the grammar correct?
- **Innovation** (7 points): Were the data or was the science particularly challenging or complex? Was any extra information provided which added to the quality of the report? Was the report exceptionally well done, smoothly flowing and with an easily understandable presentation of information? These are points to be explicitly earned; the default score for this category will be 0!

The grading algorithm for the second submission will follow these guidelines:

- **Background/Objectives** (2 points)
- **Methods** (2 points)
- **Results** (4 points)
- **Discussion** (2 points)
- **Overall organization, clarity, and flow** (5 points)

For each section, I will be looking to see whether the relevant information for each section (as shown on p. 1) is included. For overall clarity and flow, I will be looking to reward papers which are well organized within and across sections, convey details clearly and succinctly, and flow well from beginning to end.

**DO NOT CONSULT WITH ANY FELLOW STUDENT ABOUT THIS PROJECT AT ANY TIME.** All questions should be addressed to me or to Yanni or to Jia only. You may consult any reference books and journals in addition to the textbook and your course notes.

### Due dates

Tue 6 Nov: Project idea due

Thur 29 Nov: Project due in class

Mon 17 Dec: Project revision due by 5:00 pm

With such a large class, I would prefer that *you* do the printing of your report; therefore please turn in a paper copy, rather than a file by email.