MA 681 Fall 2016 (/people/sussman/MA_681/)

Announcement: Project Information

Oct 19, 2016

Project Proposal

Due: Monday October 31, 2016

As a first step for your project please write a project proposal. Your project proposal should be about 1/2 a page or less. I will only give a grade for completion or not. By the end of the week I will give you some feedback to give you some ideas and make some suggestions. If I think the project is very unrealistic I may ask that you try to do something else or modify it appropriately.

- If you are doing option 1 or 2 below you should answer
  - what data are you going to use,
  - what question or questions might you try to answer, and

- what types of statistical techniques do think you might use.
- If you are doing option 3, you indicate
    - which statistical technique you are going to investigate,
    - what scientific or other field are you going to investigate, and
    - one to three papers that you might think are relevant.

Project

As described in the syllabus 50% of your grade will be determined by your final project. You have 3 options for this project.

1. I will provide a selection of data sets and a description of an analysis problem. Choose one and use the methods discussed in class to explore the problem and make conclusions about the structure in the data.

2. Use data collected in your lab or publicly available to formulate an analysis problem that requires some of the statistical methods discussed in class, explore the problem using the tools developed in class and draw conclusions about the data. Please check with me to make sure the data you are going to use is appropriate.

3. Identify and survey the literature surrounding the application of a class of statistical methods to a particular field of study, and write a review article discussing the methods used and conclusions drawn in these analyses.

Directions in general

These directions apply to all three options. Below this you'll find information specific to Options 1/2 and Option 3.

Common Sections for All Papers

1. **Title**: Should be short, no colons if possible, and state the main point of the paper.
2. **Abstract**: In (ideally) 4-5 sentences explain (a) what problem you are studying, (b) why it is interesting, (c) the main approach to the problem, and (d) what are the results.
3. **Introduction**: A more lengthy (1-2 pages) explanation of the problem you are solving, why people should care, and how you are solving it. The most critical thing is never underestimate how little people know or care about what you are working on. It is your job to explain to them why they should.
4. **Specific to 1/2 or 3**: After the introduction will be parts specific to which option you chose. See below.
5. **Discussion/Conclusion**: The final section of the paper should present your main conclusions. You should also attempt to discuss what your next steps would be if you were

to continue to pursue this area. This is also a good place to discuss the advantages and drawbacks of your approach.

6. **Bibliography**: Please use a common bibliography style. The most important thing is that it is consistent and gives the relevant information.

Formatting

The paper itself should be no longer than 20 pages including the title, abstract, figures, tables, bibliography, etc. (This is the upper limit, if you get your ideas across in fewer pages that is fine.) Please use size 12 font and margins no smaller than 1 inch per side. There should be no code in the paper except perhaps pseudocode illustrating your techniques. Papers should be submitted as *pdf* files.

Figures

Your figures and tables should all be easy to understand and should each give key information related to your paper. Font size, axis labels, and a clear legend are all very important. There are lots of guides out there for making good figures so read a few and try to follow them. This is a checklist (https://github.com/neurodata/checklists/blob/master/figures.md) a friend of mine made, it has some good advice but don't feel you need to follow all of it.

Citations and Plagiarism

Please cite the important work related to your topic and any techniques and methods you used. If you replicate an analysis as part of your project that is OK, but please indicate that clearly. Plagiarism of any kind will be taken seriously and the best way to avoid this is to cite your resources carefully and to write everything in your own words. (This does not mean changing a few words in each sentence to make it different).

Revise, Revise, Revise!

Make sure to get a draft finished at least a week before the due date so that you have time to revise your paper. I also encourage you to use the resources available in your department and the university, including the BU writing center.

Code

If you like you can submit code separately from your paper. I will not read through the code in detail but if you give me directions and make it easy to run your code, I will do so. This could be a way for you to show some additional information but in general figures in your paper will be preferred. Please submit your code as either a single source file, (`.R` , `.mat` , etc) or a zip file.

Directions for Options 1 and 2

For Options 1/2, you will be analyzing some kind of data using the techniques discussed in class as well as other techniques you find on your own. The best papers will likely take tools from the literature that go beyond what we covered in class and I also encourage you to come up with your own well reasoned and statistically sound methods. On the other hand, I don't want you to only apply methods you learned about outside of class. Try to make sure you apply the techniques from class as well. A good way to do this is by combining methods from class with newer methods.

Your data analysis should try to target a question or a few questions about the data. I don't expect you to come up with new answers or new science, (though I'm guessing some of you will), but to present a well reasoned analysis that illustrates your understanding of the statistical techniques. Please note that you don't need "positive" results; as long as you performed a correct statistical analysis and correctly describe your conclusions that will be excellent.

The important points to cover are:

1. **Data**: Describe the data that you are using in enough detail so that the reader will have enough information to understand your analysis. This might include a description of how the data was collected, what the relevant variables are, as well as how the data was cleaned.
2. **Exploratory Data Analysis** Use the techniques of summary statistics, plots and tables to do an exploratory data analysis. This should further motivate the subsequent more detailed analysis and clarify the description of the data above.
3. **Methods**: Describe the methods that you will be using. State any assumptions you are making and try to justify them. (If you need to make unreasonable assumptions, that is OK, as long as you make it clear that the conclusions may be invalid and you should describe what assumptions would be more reasonable in for future work.) What models are you using? Are you using nonparametric and/or parametric techniques? How are you adapting the techniques from class for this task? This is a place to be very careful to describe the mathematics and statistics that will go into your analysis. I especially encourage you to consider comparing different methods for solving the same problem.
4. **Results**: Describe the results of your analysis, applying the methods from above. The important thing I want to see is that you clearly explain the statistical conclusions using the language of statistics. Present further figures illustrating your results. In addition to analyzing your data you may want to present some related simulations for comparisons.

Data Sets

For data sets you can use any data you like. Data from your lab will probably be best since you can use your own expertise and analyze the most interesting questions. If you don't have data from your lab you might want to look online for data in your field. Another option is to ask your

friends and colleagues if they have data they are willing to share. It is fine to use data which is associated with published research but a simple replication of published analysis will lead to poor grades.

If you find data that you think others might enjoy working on as well please post it below if you are willing to share it. Suggested data sets are found here (/people/sussman/MA_681/ProjectData.html).

Directions for Option 3

For Option 3, you will pick a statistics topic review the application of this technique to a specific field. The technique could be anything we covered in class or something outside of class. I will expect you to find a variety of papers that use the given technique in a variety of ways. While you may choose a topic like linear models, I expect you will find papers that go beyond the standard least squares analysis that we covered in class.

In your paper you should:

1. Clearly describe the main way that this technique is used in the field.
2. Describe important variations on the technique.
3. Describe the type of results that the technique has helped to demonstrate.
4. Describe any issues with how the technique has been used in the field.

You can also provide some basic simulations or real data analysis to demonstrate some of the issues desribed above.

It will be important that you be specific: **Bad**: Linear models applied to social sciences. **Good**: Bayesian linear models applied to education data.

MA 681 Fall 2016

Daniel Sussman (http://math.bu.edu/people/sussman/)
sussman@bu.edu (mailto:sussman@bu.edu)
Office Hours: MCS 236
- Tuesday 2pm-4pm
- Wednesday 11am-12pm

Sahand Farhoodi
sahand@bu.edu (mailto:sahand@bu.edu)
Office Hours: MCS B44
- Wednesday 9am-11am

This website has notes and other info for GRS MA 681: Accelerated Introduction to Statistical Methods for Quantitative Research at Boston University for Fall 2016.