

BIOST/STAT 571: Final Project (Overview)

Final Project

- **Objective:** Develop a “new” method or approach with the aim of publishing in a statistical or applied journal
- **Well Regarded Journals** (not comprehensive list!!)
 - Statistical Methods: Biometrika, JASA T&M, JRSS-B, Biometrics
 - Applied Methods: JASA A&CS, Biometrics (Practice), Biostatistics, Statistics in Medicine, JRSS-C (Applied Statistics)
 - Computational: CSDA, Journal of Statistical Computation and Simulation, JCGS
 - Informatics: Bioinformatics, Genetic Epidemiology, PLoS Journals, Neuroimage
 - Applied Journals: Journal of Clinical Oncology, American Journal of Human Genetics, Nature Journals

What is a “New” Method?

- De novo frameworks?
- Adaption of prior frameworks
 - Translation to new context
 - Extensions of existing frameworks
 - Bells and whistles and Cute tricks
- A “new” method does not truly need to be “new”
 - Very little in statistics is truly new

What Goes into Development of a New Method

1. The problem that is being solved
2. Methods and statistical framework for solving problem
3. Justification and Evaluation of the method
 - Why are you solving this problem?
 - Why do you need a “new” or different approach?
 - Can you study the theoretical properties of your approach?
 - Computational considerations and algorithms?
 - Under what circumstances will your approach work? When will it fail?
 - What happens with real data?
4. Iterate through 1-3

How to Start Developing a “New” Method: Identifying a Problem

No universal approaches, but some options include the following:

- Motivation from data
 - Is there some characteristic of the data that the “usual” methods cannot handle?
 - Is there are question arising from the data that nobody has answered before?
 - Are there standard questions (from other data sets) for which methods do not exist?
- Motivation from previous methods
 - Under what situations do existing methods fail?
 - Are there situations that an existing approach cannot handle? Can we do better?
 - Can we apply/translate an existing method to a new context?
 - I found a cool trick. Can I try incorporating it into an existing method?

Building Your Method

- Focus on specific aspects of the problem that you want to address
- Depends on what you're trying to do:
 - Better model
 - Better algorithms
 - Better theorys
 - Etc.

Justification and Evaluation a Method

- Why someone should care (most important part)
 - Sometimes, honest lies
- Theory and properties (depends on the journal)
 - Asymptotics usually
 - Finite sample theory rarely
 - Computation
- Simulations
- Data Applications
- Generally: No method universally wins, just want to show that yours **can** win and issue guidance

Paper Structure

- No definitive structure for papers: depends on context and the journal
- A Rough typical structure: (Not necessarily section headings!)
 - Introduction
 - Methods and theory
 - Results
 - Discussion
- Good idea: follow structure of relevant papers
- Bad Idea: follow structure of relevant papers
- Main idea: how would you explain and justify your approach to others?

Introduction

- By far the most important part of the paper
- What background material is necessary?
- What is the problem that you are solving?
- Why is the problem important?
- What related work has already been done?
- What is the approach that you are taking?

Methods

- (Sometimes) prior related work
- Proposed approaches and models
- Algorithms for implementation
- Theory:
 - Justification for your approach
 - Theoretical comparisons of your method to existing approaches
- This may take multiple sections in a paper

Results

- Empirical evaluation of your method (should back up what you say in the intro)
 - Comparisons with existing approaches and relevant metrics for evaluation
 - Comparisons of different options of your approach (e.g. should one use CV vs. AIC)
 - Sensitivity analysis
- Simulation scenarios (sometimes in methods) and simulation results
- Real data applications:
 - Show that your method works on real data
 - What insights does your method provide that are new or unusual?
 - For your final project:
 - Apply your method to real data
 - Do not need to give significant insights (or even work well, e.g. poor type I error control due to sample size)
 - Need to explain what you're seeing in terms of behavior of your method

Discussion/Conclusions

- What did you do in this paper and what did you show?
- Recap of when your method wins and when it loses.
- What are the options that go into your method? Recommendations for which to use?
- What are things that you would have liked to investigate further but are outside of the scope? Future research?
- What are things that others are likely to pick on? Pre-empt their comments.

Details of your Final Project

- **Groups:**

- 3-5 students each (recommend 4 students)
- *Pick your own*
- Finalized groups by weekend: communicated to me and the TAs

- **Grading:**

- The paper will be evaluated on the basis of originality/creativity, scholarship (including appropriate literature citations), clarity, organization and relevance to class goals.
- Not all group members may receive same score: you will be asked for relative contributions (HW5)
- Creativity and thoroughness of evaluation count

- **Due:** 5pm on Thursday, March 17, 2022

- E-mailed to: Instructor, TA's and ALL group members

Deliverable: The Paper

- Ideally: something publishable or that is close to publishable as a methods paper
- Length:
 - No restriction as long as it is complete as a paper
 - Expect about +/- 10-20 pages double spaced (not including any figures)
- Format:
 - No set format, but probably good idea to follow usual structure
- Note:
 - HW: can you do this?
 - Project: what can you do?

Grading

- Relevance to class goals
 - Must be (somewhat) related to correlated, clustered, longitudinal data
- Originality/Creativity
 - What is “new”: the problem? The solution? Something else?
- Scholarship
 - Justification of your approach
 - Literature review
 - Simulations
 - Evaluation of strengths (and weaknesses) of the approach, particularly relative to other ideas
- Clarity and organization
 - Can you clearly convey the problem and the solution?
 - Does your argument make sense?
 - Does the paper make logical sense or have you just copied someone else’s template?

Further details

- Most methods papers take a long time to develop. Not everything needs to be fully complete.
 - Real data: depending on your method, it need not be applied to real data due to the limitations on time
 - I do not expect your paper to be directly publishable upon submission – but the closer it is to something complete, the better the score.
- The most important thing will be the importance of the problem (which you argue for) and the core method (including rough implementation)