

Statistical Research: Some Advice for Beginners

Hamada and Sitter

Jennifer Feder

Johns Hopkins Bloomberg School of Public Health

4 September 2009

Quick overview of select resources

Searching for articles

- **JSTOR** - Journal Storage
- **Google Scholar**
- **PubMed** - life sciences journals, biomedical articles

Access to articles and books

- On campus: connect to internet through Hopkins secure network
- Off campus: connect through **myJH**

Keeping track of everything

- **Papers software** (\$25, for Macs)
- **Zotero** (Firefox extension)
- Other tools?

So many great journals

Why I love acronyms

- Statistical & Biostatistical Journals

- JASA - Journal of the American Statistical Association
- JRSS (A, B, & C) series - Journal of the Royal Statistical Society
- Biometrics - Journal of the International Biometric Society
- Biometrika - Oxford University Press
- Biostatistics
- Annals of (Probability, Applied Probability, Statistics, Applied Statistics) - Institute of Mathematical Statistics
- ⋮

- Public Health & Medical Journals

- JAMA - Journal of the American Medical Association
- Epidemiology
- AJE - American Journal of Epidemiology
- ⋮

- ⋮

Research?!?!???

- “When do I start?”
- “How do I start?”
- “How do I find out what has already been done?”
- “How do I make progress?”
- “How do I finish?”
- “What else can I do?”

“Start now”

- Class projects could turn into research
- When choosing a project, combine it with previous knowledge and topics of interest
- Treat a course project as a research project
- When taking classes, read related material from a variety of sources and look at the papers referenced in class or in the books

How do I start?

- Work on something that interests you
- Work in a new application area not previously considered by statisticians → any advance you make a contribution
- Collaborate with a subject matter expert
- Ask professors about what they're working on
- Ask for advice from professors about a direction to go in for a problem you're interested in

Take advantage of emails from professors looking for research assistants or students to work on projects, big and small

Once you've stated a problem

Two dominant opinions

- Begin with a thorough literature review
 - Results in learning the field and techniques in use & avoids the chance of “rediscovering the wheel”
 - Time consuming and tends to lead the researcher to view the problem the same way it's been viewed previously
- Attack the potential problem first and look to the literature later
 - Avoids too much influence by those who have already been working on related problems
 - Danger of wasting effort re-solving a problem

A compromise:

- Read a few papers (a review paper, if available, can be helpful)
- Explore the problem in new ways

How do I find out what has already been done?

Identifying relevant work

- Library search engines, electronic journals
- Straight key word search in Google (or Bing, others)
- Look at references of recent papers to identify previous key papers
- Citation indices → identify other recent papers that cite these previous key papers
- Don't forget about similar problems in different statistical fields.
 - e.g. reliability & survival analysis
- Related problems in different disciplines: economics, computer science, etc.
- Recent developments may be reported in conference programs & proceedings, technical reports

How do I find out what has already been done?

Digesting the Literature

- At first: look for discussion on the ramifications of theorems rather than wading through details of proofs
- Read abstracts, introductions, and conclusions
- Questions to consider (but not get overwhelmed by) when reading
 - Are data available and how are they modeled?
 - What assumptions are being made about the models?
 - What are the issues? Are there new issues?
 - Is it an analysis or design problem?
 - What methods are employed (e.g. nonparametric or parametric, frequentist or Bayesian)?
 - How can the problem be extended?

How do I make progress?

Attacking your problem

- Use simple examples
- Consistently spend time working on the problem (e.g. an hour every day)
- Explain the problem to others
- Use the computer!
 - Try examples to explore whether conjectures are true or not
 - Begin with something simple and then extend slowly toward more complex
 - Try situations near the edge of your assumptions
- Computer is no substitute for thought and understanding
 - Work through some small examples

How do I make progress?

The Moment of Discovery

- There is no substitute for hard work
- Examples of long stretches of discouraging attempts with no results, only to solve the problem in an hour

“One needs to live with the problem, have it percolate, think about it over and over until it is always there in the back of your mind”

Beginning to Finish

- You now have the basic plot of the story and you need to tell the story well
- “Play the devil’s advocate” by
 - Criticize and challenge your own results
 - Understand the advantages and disadvantages
 - Identify competitors & compare through application to real examples, simulation study, and/or theoretical comparison
- Anticipate what others may ask
- Articulate your ideas to someone else
- Write down your story
- Remain aware of new research problems that may develop from your work

How do I finish?

Writing it Down

- Write as you go along → helps to focus and organize your thoughts
- As you have results, quickly write them up, disregarding eloquence
- Draft the main sections before writing the intro and discussions sections
- Questions to ask yourself
 - What are the main results you wish to communicate?
 - How can you help the reader to understand and enjoy what you have to say?
- Capture the reader's interest early on without getting bogged down in notation and technical details
- Be brief and concise, accessible to the reader
- No substitute for writing and rewriting

Talking About It

- Present your project to gauge the audience's response and obtain constructive criticism
- Dedicate each slide to a single idea
- Text should be easily readable
- Use graphs rather than tables
- A slide shouldn't be filled with mathematics or proofs
- Consider time constraints
 - e.g. sketch of the novelties of your problem or key aspects of the simulation

What else can I do?

- Become a project or research assistant
- Be a summer intern in industry, business, or government
- Attend conferences
- Attend seminars
- Attend Journal Club!
- “Anything and everything that can help you have fun exploring the unknown”

- How did you come up with a research question?
- How did you start your research process?
 - Did you favor beginning with a thorough literature review, jumping right in, or something else?
- How do you go about conducting a literature review?
- What are some useful strategies for reading a journal article? Questions to ask yourself while reading?
- Have you experienced the “Moment of Discovery?” What happened?
- Advice about writing up your work? Presenting it?