

GETTING THE INFORMATION YOU NEED

- Introduction
- Resources for searching the literature
- Reading the literature
- Searching for other stuff
- Remarks

INTRODUCTION

Purpose and scope of research:

- Research is the mechanism by which knowledge is *advanced*
- Meaningful research that achieves this objective generally involves *nontrivial extension of ideas in previous work, development of new approaches and ideas, evaluation and comparison of existing and new approaches*
- Thus, to carry out meaningful research focused on a particular area, one must be *familiar* with previous developments in that area

“The literature:”

- *Journals* are the main mechanism by which research results are disseminated
- Also *books*, *monographs*, compilations of papers (*edited volumes*)
- “*Encyclopedias*”

Thus: A first step in becoming an *effective researcher* is to become effective at *searching* and *understanding* the literature

Dissertation: Your first *in-depth* research experience

- Research that constitutes a dissertation should address a previously *unaddressed* problem or issue, and the student should be an “*expert*” in the area at the conclusion.
- The first thing you will (almost always) do is become *familiar* with *past* and *current work* in the area(s) of statistical methodology relevant to your dissertation topic
- You will find that you must *continue* to consult the literature
 - To find results that may be useful in your analytical work
 - To verify that your work has not *already* been done!
- It is likely that your advisor will point you in the direction of certain “*must read*” papers (including some by the advisor)

Dissertation, continued:

- However, papers your advisor thinks you should read probably represent only a *small fraction* of those with which you *ought* to be familiar to really be an “*expert*” in the area
- As a *researcher*, it is *your responsibility* to identify and seek out literature important to your work

Other scenarios: Work on a dissertation is by no means the *only* situation where familiarity with literature is *important*

- *Academic/government lab environment* – research is a *central* activity/responsibility
- *Industry environment* – problems arise that require *specialized methods*. Are such methods available?

Searching the literature:

- You may want to do a *general search* on topics of primary importance to a particular area to become familiar with the names of the “*key players*” and the important papers
- You may want to find all the papers by the “*key players*” to trace the development of an area
- You may want to find a method appropriate for a *particular problem*
- In reading papers, you may encounter a reference to a paper by an author whose name does not sound familiar – you may want to check to see whether that author has done other, related work
- You may realize there is more than one “*buzzword*” meaning the same thing in your problem area – you may want to make sure you are not missing an entire *subset* of relevant literature

RESOURCES FOR SEARCHING THE LITERATURE

Luckily:

- There are numerous resources for *searching* the (statistical) literature and *retrieving* articles and other information
- The *web* has made it possible to do much of this without ever leaving your chair!

ISI Web of Science:

- A web-based service provided by the *Institute for Scientific Information* (ISI) to which NCSU subscribes
- ISI maintains *Citation Index* databases, which contain bibliographic information gathered from *thousands* of scholarly journals
- Web of science allows searches of these databases, and thus of papers in almost all journals in all imaginable fields of science!
- Updated *continually*
- Ample on-line help, easy to figure out
- Go to

<http://www.isiknowledge.com>

and click on ISI Web of SCIENCE

ISI Web of Science, continued:

- Searches may be done by
 - Topic
 - Author
 - Journal
 - Author address
 - Authors cited within a paper
 - Papers cited within a paper
 - Years a paper was cited
- Once an article is found, it is possible to link to other articles on the same topic
- *Easy search* – Search by topic, person, place
- *Full search* – more refined search, by year, journal title, cited references, etc.

ISI Web of Science, continued: Practical pointers

- Because it covers *numerous disciplines* (not just statistics), searches must be tailored; e.g., typing

BERGER R*

to find papers by Roger Berger will yield *zillions* of results because Berger is a common name. Better to type

BERGER RL

- Try typing BOOS DD (not a common name!)
- Similarly, inputting *too broad* a topic will bring up too many results; e.g., typing

MIXED EFFECTS

will yield papers from *all fields* in which mixed effects models are used, *not* just statistical research articles about mixed effects models

- Best bet: *Try it!*

Current Index to Statistics (CIS):

- A bibliographic database to publications in statistics and related fields, including journals, books, proceedings of conferences.
- Published in book form in the *olden days*; now available *on line*
- Main web site with description of CIS
<http://www.statindex.org/CIS/>
- *Query server* main web site
<http://query.statindex.org/>

Allows searches by

- Author or groups of authors
- Keywords
- Title of article, journal book
- During specific time ranges

Current Index to Statistics (CIS), continued:

- Try typing BERGER, ROGER
- Can type BERGER, ROGER **or** BERGER, R. **or** BERGER, R.L. to make sure you find *all* Roger L. Berger papers
- *Advantage* - restricts search to sources with statistical content
- *Disadvantage* – not as up-to-date as Web of Science
- *Thus*, do not consider CIS “*the last word*,” as articles with statistical content may also appear in journals not searched by CIS

JSTOR:

- A service to which NCSU subscribes
- A *searchable* archive of back-content of journals in numerous disciplines (including *statistics*)
- Content of numerous of the most important statistical journals, including *American Statistician*, *Annals of Statistics*, *Applied Statistics*, *Biometrics*, *Biometrika*, *Journal of the American Statistical Association*, *Journal of the Royal Statistical Society, Series B*
- Articles are available in pdf and postscript formats for downloading
- Complete issues of journals are available, often from inception up to a *moving wall* of usually five years from the present

JSTOR, continued:

- Can retrieve and *print* articles, saving a trip to the library
- Can *search* the archived journals by clicking on **SEARCH** and following the instructions
- E.g., type Roger L. Berger in author and select **Statistics journals**
- Can click on articles identified to save or print

On-line journals:

- Most journals now are published in both *print* and *electronic* forms
- May or may not be accessible to individual users
- Many statistics journals are accessible through the *NCSU libraries*; e.g., *Biometrika*, because the institution has a subscription
- Journals to which NCSU does not subscribe may not be accessible to you on-line
- Some journals require *membership* in the society that publishes the journal or an individual subscription (e.g. *Biostatistics*)
- The NCSU library offers *Interlibrary & Document Delivery Services*; see <http://www.lib.ncsu.edu/ads/ils/>

Other useful resources: For example

- *Encyclopedia of Statistics*
- *Encyclopedia of Biostatistics*
- *Encyclopedia of Environmetrics*

contain overview articles on numerous topics and are a good way to get started learning about an area and the most important references

- A *Google* search (www.google.com) on author names or key words can be surprisingly effective!

READING THE LITERATURE

How to exploit all these resources?

- It is very possible to have *too much of a good thing!*
- When tracing the development of an area, you may identify a *huge* number of seemingly relevant sources (from searches on titles, authors, *keywords*)
- *Fact of life* – all papers on a topic are not necessarily *good* or *useful*
- It can be a *difficult* task to sort through and find the “*important,*” *useful* ones
- You can't read them all in gory detail!
- To be effective, you must learn how to *sort* and *read efficiently* references identified by a search

How to sort:

- *Dissertation research* – your advisor may give you “*key*” papers to get you started; read these *first* and seek out the articles that are *cited*
- Start with the most recent sources; there is a “*mushroom effect*” as you go back in time
- *Encyclopedia* articles and *books* are sometimes a good way to start, but remember there are likely more recent sources
- Articles that are *almost always cited* are probably “*key*” to the development of the area or were among the first in the area
- A recurring set of *authors* will emerge; articles by these authors are probably “*more important*”
- A paper with an accompanying *discussion* may often be “*important*”

How to sort, continued:

- *Usually* – articles in “*main*” or “*first-tier*” journals are the most important, but this is not always true
- *Do not* print, copy, download every and any paper you identify! Instead, read the *abstract*, *introduction*, and *conclusions* or *discussion* sections. A paper might not be what you think it is.

How to read:

- Because of space considerations, journal articles often *leave out* a good deal of technical detail (“It may be shown that...”)
- Initially, you may wish to verify as many details as you can. But this can be frustrating and time consuming and may get you off the track. Read first instead for the *main ideas* and *message* to see if the paper is of interest; you can verify details later.
- You will certainly want to verify all the details in the “*seminal*” and “*key*” papers in your dissertation area
- As you learn more about an area, you will begin to see *patterns* in the kinds of results and arguments that arise
- As we will discuss, the *quality of the writing* is often a help or a hindrance
- Keep *dated*, detailed notes of any calculations you carry out or thoughts you have when reading a paper for later reference

SEARCHING FOR OTHER STUFF

Software: It would be a great help if there were *software* to carry out some specialized analysis

- Chances are, someone may have written a program to do it
- Statlib – a system for distributing statistical software, datasets, and information
- <http://lib.stat.cmu.edu/>
- S, Splus, R routines, codes of all kinds (e.g. c, c++, Mathematica) to do a variety of analyses
- Datasets
- Lots more...
- Many researchers make software *freely available* on their web sites

REMARKS

Key aspect of research: To be a successful researcher, one must keep up with the *state of the art*

- Scan the most recent issues of statistical journals
- Relevant work may be in progress or recently completed, so not published yet, but may be available on web pages of “*key players*” (Google)
- Attend statistical meetings; a good way to meet “*key players*” and discuss ideas
- E-mail