

Some Research Skills

STAT8801 Statistical Consulting

School of Statistics
University of Minnesota

March 10, 2010

A consulting statistician must be jack of all trades and master of some.

You've learned a lot about statistics from your courses, but there are many, many other methods out there.

Methodology Talks

Instead of case studies, for the last five weeks, you will each give a presentation of eight minutes on a randomly assigned method you may not have heard of.

Your presentation should

- Describe what the method is for.
- Give an example of a potential or preferably real application.
- Give a few references where others could go to find out more, including at least one book and/or journal article.
- Is software available for this method? If so, tell us about it.
- Include a **one-page** handout that summarizes your talk and provides references.

Sources

- On-line:
 - ▶ Journals
 - ▶ Encyclopedia
 - ▶ Software documentation
 - ▶ Books
 - ▶ Unpublished work
- Off-line:
 - ▶ Books
 - ▶ Journals
 - ▶ Encyclopedia
 - ▶ Software documentation
 - ▶ Unpublished work

Journal articles

Pros:

- Comprehensive
- Up to date
- Complete
- Usually the primary source

Cons:

- Can be difficult to find
- Often difficult to understand
- Idea/notation drift

Books

Pros:

- Usually less technical than journal articles
- Often better written/easier to understand than journal articles
- Generally not on-line

Cons:

- Often only cover “big” subjects
- Often poorly indexed

Encyclopedia

Here I'm mostly talking about The Encyclopedia of Statistical Sciences.

Pros:

- Wide variety of topics
- Well written, understandable articles
- Easy to understand

Cons:

- Very expensive
- Not widely available
- Doesn't cover everything

On-line

A google search on well-chosen key words can turn up dozens of relevant and helpful links. But ...

Here be Dragons!

Web items are not authoritative. Any jerk with a computer can put stuff up.

Know your source!

Tally Ho!

Let's go hunting for something, say **discrete choice models** and **multinomial logit models**.

The easiest thing is just to google and see what we find.

<http://www.google.com>

Jackpot!

OK, we may have a winner. The very first link is to an entire book on discrete choice models by a professor at UC Berkeley. We also find lecture notes from NYU and similar items. These are more authoritative than you generally find.

More helpful links

Wikipedia www.wikipedia.org

Google scholar scholar.google.com; from off-campus, use
<http://www.lib.umn.edu/slog.phtml?url=http://scholar.google.com>

Current Index to Statistics <http://query.statindex.org> or from the
"Other links" link on School of Stats homepage, or from
off-campus
<http://www.lib.umn.edu/slog.phtml?url=http://query.statindex.org>

Getting papers

How about Parks' 1980 paper on the multinomial logit? It sounds pretty early.

Parks, Richard W. (1980) "On the estimation of multinomial logit models from relative frequency data" Journal of Econometrics, 13, 293-303

How do we find it?

Making a COPY of an article \neq READING an article.

The library homepage provides access to many resources, including e-journals, indexes, data sources and many others.

To find a journal article...

- 1 Go to www.lib.umn.edu.
- 2 Click on e-journals.
- 3 Search for econometrics.
- 4 Choose Journal of Econometrics.
- 5 Fill in year and page number.

Eventually, you'll get to the article (you may need id and password).

Web of Science

Accessible from the library's indexes page. You can now search for the paper of interest. You can

- Find papers that cited it.
- Find papers that cited them.
- Find the papers they cited.

Then you start looking these up online as well.

Indexes

We also like to trace : who did this paper cite, and who cited this paper.

Citation index Can be used to trace ideas forwards and backwards via references. Go to www.lib.umn.edu and click on indexes, then "Cited reference indexes", then click on Science Citation Index.

Google scholar scholar.google.com also presents citations

Lexis-Nexis academic general contemporary search of news sources

Software documentation

Caveat emptor, let the buyer beware.

SAS I [Gary] love to hate SAS, but they have good, extensive documentation. Go to support.sas.com. Search for "multinomial logit" or "discrete choice".

R Documentation for R is more diverse. Good starting points are rseek.org, a search engine that searches only known R sites, and www.r-project.org, with links to manuals, FAQs, the R Newsletter, the R Wiki, and books.

Specialized programs Some areas use specialized programs; e.g., **Mark**, welcome.warnercnr.colostate.edu/~gwhite/mark/mark.htm for mark-recapture studies of animal populations, and **Winsteps**, www.winsteps.com in educational testing. Finding out what they do can be a challenge.

Hits on amazon.com on 3/10/10:

R statistics	5,582
SAS	3,852
SPSS	2,223
Stata	1,108
Minitab	1,016
JMP	403
Lisrel	181

- 1 www.fedstats.gov provides a central gateway to many federal statistics (that is, numbers)
- 2 Other governmental agencies (such as states) often make data available on line.
- 3 Private organizations also provide data; e.g. www.jewishdatabank.org.