Statistical Practice

36-726 Spring, 2007

Course Policies and Syllabus

Instructor	: Brian Junker	**Class Location**	: 231B Baker Hall
Office	: 132E Baker Hall	**Meeting Times**	: MWF 1:30–2:20
E-mail	: brian@stat.cmu.edu	**Office Hours**	: By Appointment
Webpage	: http://www.stat.cmu.edu/~brian/726		
Phone	: x8874		

Overview

Most statisticians in industry, government, and academia spend a fair bit of their time providing advice and analyzing data for colleagues whose specialties are in some other field. Often, the contribution of the statistician in a scientific collaboration is to sharpen the focus of an investigation by identifying what may, and what may not, be learned from a particular set of data. It is part of the appeal of our discipline that it is application-oriented, yet in most coursework there is inadequate time to take the applications seriously and discover what they are about. The main purpose of this course is to give students some experience in digging more deeply into scientific problems that involve statistical thinking. It is a course intended to aid you in making the transition from being a student of statistics to being a statistics professional.

Whether collaborating with a researcher who has a problem that could motivate development of new methodologies, or helping someone analyze and interpret their data using conventional techniques, communication skills are essential for success. Through practice and discussion, the class will help you to improve your ability to interact with and understand a client. A practicing statistician must be able to summarize results and advice in a clear and useful manner, both verbally and in writing. Work on projects in this course should help you improve these skills as well. Often, a lasting benefit to the investigator will be the education he or she receives from the statistician. This is another aspect of the process that improves with experience, and which you should be conscious of in your work here.

Textbooks, Resources, and Statistical Methods

There are no assigned textbooks for this course. Those of you who can do your project work in R will appreciate having a current edition of Venables & Ripley's *Modern Applied Statistics with S* handy. Web resources for R and other statistical packages will also be useful.

One good general rule about statistical analysis is K.I.S.S. (keep it simple statistics), since results from simpler methods—when they are appropriate—are always easier to interpret and explain. Often you can "get" everything from well-made graphs or other data summaries, and sophisticated modeling and inference just confirms what you already see.

Nevertheless many of you may need to learn new methods or adapt old ones for your data analyses. There will be no formal lectures on methods in the class. In class or in individual group meetings, I can suggest ideas and resources for you to pursue on your own.

Objectives

This course aims to help you develop your skills so that you will be able to:

- Determine the nature of an investigator's question, and a good way to formulate it statistically;
- Focus listening and speaking on important statistical issues; and
- Write about statistics for people who will use what you have written.

In addition, we will read and discuss various ethical, practical, and interpersonal aspects of consulting so that you can develop your own sense of values as a practicing statistician.

Projects

By having project presentations at the end of Fall Semester in 36-701, we hope that projects and teams will get underway quickly near the beginning of sprong semester, giving you the maximum time to interact with your clients and their data, and produce the most sensible analysis you can. The project teams that we set up last fall are as follows:

Project A: Astrostatistics
Investigator: Andrew Connolly, Dept of Physics & Astronomy, U. Pitt; & Google
Team: Han Liu, Joey Richard

Project B: Biochemical Markers of Neuron Growth
Investigator: Alison Barth, Biological Sciences, Carnegie Mellon
Team: Drew Crossett, Steven Wasik

Project C: Neonatal Ferret Visual System
Investigator: Bill Eddy, Statistics, Carnegie Mellon
Team: Gabrielle Fijas, Gary Klein

Project D: Analysis and Design of Alumni Surveys
Investigator: Judy Cole, Associate Vice President for Alumni Relations, Carnegie Mellon
Team: Yi Jiang, Zhanwu Liu

Project E: Project Listen in Ghana
Investigator: Jack Mostow, Robotics, Carnegie Mellon
Team: Elise Olson, Nathaniel Anozie, April Galyardt

Project F: Identifying Neurons by Firing Features
Investigator: Marc Sommer, Dept. of Neuroscience, U. Pitt
Team: Bellone Gaia, Shuhei Okomura

Format, Projects, Grading

Your team of two or three students will work on the on problems brought in by our clients last fall. Each team will work with one client on one project for the term; at the end of the term each team will present its project/results to the class and produce a written research report for the investigator (and the instructors), summarizing the work that they've done.

Here are some major events in the course. Exact dates will be announced in class.

- Class meetings for the first week of classes will be for organization and orientation. We will talk a bit about the consulting process, and also set up regular meeting times for the teams.
- During the second week of class, each team will give a short (10-15 min) presentation of their project, based on (a) the investigator's presentation last fall; (b) your first meeting(s) with the investigator; and (c) any initial work or thought you've had on the project.
- Student teams will arrange one or more meetings with the investigator, as needed, and begin working on the problem.
- Approximately once every 4 weeks, one member of each team will give the entire class a review and update on each project. *Each team member will be expected to do this at least once; the same person may not give two monthly reports in a row.*
- A written interim report on your project will be due to me on March 9, 2007, **by email** [I will be in Washington DC on that day]. I will read these reports and give you feedback on March 19, 2007. I may also ask the team to organize interim team presentations to the class.
- Toward the end of the semester, teams will make final presentations to the entire class. All of you will be expected to write a summary of each team's oral presentation.
- Final written reports are due to the instructors and the investigators no later than the last day of classes, May 4, 2007. We will discuss how these final reports might be organized as the course progresses.

As you can see, most of the class periods will be taken up by presentations by outside investigators or student teams. You are expected to participate actively in class and to interact both in class and outside class with the investigators. In addition to working on projects you will also be doing various reading, writing, and presentation assignments to hone your skills as a consumer and producer of statistical research.

Readings

When we are not having presentations by investigators or students, we will be discussing various aspects of statistical practice. Some representative readings follow. I will assign readings from this list or elsewhere, for you to present and discuss in class.

Communication & Writing:

Alley, M. (1986). *The Craft of Scientific Writing, 3rd ed.* New York: Springer-Verlag. There are many examples at http://www.writing.eng.vt.edu and http://www.me.vt.edu/people/faculty/alley.html.

Hoadley, A. Bruce and Kettenring, J.R. (1990) Communications between statisticians and engineers/physical scientists. (Com: pp.249–274). *Technometrics*, **32**, 243–247.

Boen, James R. and Zahn, Douglas A. (1982) *The Human Side of Statistical Consulting.* Wadsworth, Belmont CA. *out of print*

Ehrenberg, A.S.C. (1981) The problem of numeracy. *The American Statistician*, **35**, 67–71.

——— (1982) Writing technical papers or reports. *The American Statistician*, **36**, 326–329.

Gopen, George D. and Swan, Judith A. (1990) The science of scientific writing. *American Scientist*, **78**, 550–558.

Scientific Thinking:

Chatfield, Christopher (1991) Avoiding statistical pitfalls (with discussion). *Statistical Science*, **6**, 240–268.

Bolles, Robert C. (1988) Why you should avoid statistics. *Biological Psychiatry*, **23**, 79–85.

Platt, John R. (1964) Strong Inference. *Science*, **146**,

Cohen, Jacob (1990) Things I have learned (so far). *American Psychologist*, **45**, 1304–1312.

Ethics:

Committee on Professional Ethics, American Statistical Association (1999). *Ethical guidelines for statistical practice.* http://www.amstat.org/profession/index.cfm?fuseaction=ethicalstatistics Additional resources available at http://www.tcnj.edu/˜asaethic

Ellenberg, Jonas H. (1983) Ethical guidelines for statistical practice: a historical perspective. *The American Statistician*, **37**, 1–4.

Finney, David J. (1991) Ethical aspects of statistical practice. *Biometrika*, **47**, 331–339.

Consulting and Statistical Practice:

Hunter, William G. (1981) The practice of statistics: The real world is an idea whose time has come. *The American Statistician*, **35**, 72–76.

Joiner, Brian L. (1975, 1979) Some advice to young statisticians; Consultant's check list. Unpublished papers.

Kirk, Roger E. (1991) Statistical consulting in a university: Dealing with people and other challenges. *The American Statistician*, **45**, 28–34.

Marquardt, Donald W. (1979) Statistical consulting in industry. *The American Statistician*, **33**, 102–107.

Tweedie, R. (1998). Consulting: Real problems, real interactions, real outcomes, *Statistical Science, 13,* 1–29.

Zahn, D. A. and Isenberg, D. J. (1983). Nonstatistical aspects of statistical consulting, *The American Statistician, 37,* 297–302.

The Office of Equal Opportunity Services provides support services for both physically disabled and learning disabled students. For individualized academic adjustment based on a documented disability, contact Equal Opportunity Services at eos@andrew.cmu.edu or (412) 268-2012.