

THE STATISTICAL CONSULTANT



Section on Statistical Consulting
Karen Copeland, Editor; Christopher Holloman, Assistant Editor

American Statistical Association
Fall 2007; Volume 24, No. 3

IN THIS ISSUE

- Undergraduate Consulting as an Introductory Course - Presented at JSM 2007
- JSM2007 Roundtable Summary: “Statistical Consulting in Medical Devices: Opportunities and Challenges”
- Q & A: Question from a Section Member, Answered by a Section Member
- Resources for Improving Your E-mail Use
- Comments from the Chair
- Notes from the Editor

Undergraduate Consulting as an Introductory Course - Presented at JSM 2007

Jim Bentley, University of Redlands, Redlands CA *

Classes in statistical consulting have long been a part of graduate programs in statistics. Undergraduate opportunities within these schools exist, but are less common and often reserved for junior and senior statistics majors. The consulting class offered by the University of Redlands

*jim.bentley@redlands.edu

Executive Committee of the ASA Section on Statistical Consulting

Chair: Stuart Gansky, sgansky@itsa.ucsf.edu

Chair-Elect: Brenda L. Gaydos, gaydos_brenda@lilly.com

Past Chair: Philip M. Dixon, pdixon@iastate.edu

Program Chair (for 2007 JSM): Phillip Chapman, phil@stat.colostate.edu

Program Chair-Elect (for 2008 JSM): James Grady, jgrady@utmb.edu

Secretary / Treasurer: Larry D. Haugh, haugh@cems.uvm.edu

Publications Officer: Paul Stewart, paul_stewart@unc.edu

Webmaster: Chuck Kincaid, ckincaid@comsys.com

Newsletter Editor: Karen Copeland, karen@boulderstats.com

Assistant Newsletter Editor: Christopher Holloman, holloman@stat.osu.edu

At Large Members:

Anamaria S. Kazanis, akazanis@umich.edu

N. Shirlene Pearson, pearson@smu.edu

Karen Pieper, karen.pieper@duke.edu

Council of Sections Representatives:

Edward D. Rothman (Senior), erothman@umich.edu

Susan McGorray (Junior), spmcg@biostat.ufl.edu

Department of Mathematics was designed to be accessible to undergraduate students across all disciplines, and to encourage these students to continue in their studies of statistics.

Initial need for a consulting class was recognized by Rick Cornez. While reviewing folders as a member of the U of R Faculty Review Committee, he noticed that many faculty members in many disciplines were working on projects that might benefit from some statistical or mathematical support. Together with Steve Morics and Michael Bloxham, Rick created a consulting class (Math 301) through an NSF Project Intermath grant (NSF-DUE-9555414).

Rick, Steve, and Mike envisioned a class where groups of students from different majors provided statistical and mathematical support to researchers on campus. The course was modeled along the lines of an Environmental Studies program "studio" course where teams of students with different strengths and abilities worked on real-world problems of environmental design. Math 301 was advertised within departments that had their own introductory statistics courses, with a plea that faculty encourage students who had shown great interest or exceptional ability in these courses to sign up for the consulting course. Initially offered in the fall semester of 1999, Math 301 has been offered at least once a year since. Recently it has become a spring semester course - this allows for fall groveling for projects and preparation.

The University of Redlands is a small, liberal arts university. Class sizes for Math 301 have ranged between 6 and 18 students. When team-taught, having 18 students was acceptable, but the number of projects that had to be overseen suggested that about 9 students (three projects) was better for a single faculty member.

By numbering, Math 301 is an upper-level course which is typically populated by juniors and seniors. However, the only prerequisite for the course is an introductory statistics class from any of the departments on campus (mathematics, psychology, biology, and economics) or AP credit. By

the time the students take 301 most have forgotten what they learned in their previous statistics course. Those students coming through the mathematical probability and statistics sequence often have not had an introductory course. Many of the students report never having been exposed to anything more than computing means and medians, and the occasional t or chi-square test.

Teams of three students are formed based upon their interests in the project topics and their backgrounds. Those with strong skills in programming, math majors, etc. are spread amongst the projects. In no case is a student allowed to work alone. The structure of the teams is intended to help the students recognize the importance of being able to work with others during a consultancy.

The class meets twice a week for 80 minutes. The first three weeks to one month are spent reviewing topics the students should have seen in their AP level courses. The topics are often selected based upon the instructor's prediction of the techniques that the upcoming projects will require. Full explanations of the statistical techniques are rarely given, thus leaving plenty of room for the students to research the topics they now know exist.

As an example, a faculty member may be submitting data they collected on students' self-reporting of their aptitudes in certain areas of application. Since the data is ordinal, the class will be encouraged to look at the differences between qualitative and quantitative data and which statistics and plots are correct for each type of data. Mini-projects that stress the selection and application of techniques that are appropriate for the given, "canned" data are provided. These allow the instructor the opportunity to correct misconceptions that may exist due to previous course work (*e.g.*, what is bias, variability, a p-value, etc.).

In the nine remaining weeks the students take full responsibility for the projects. They meet with the data donors to determine what is required. These meetings may be in person, by phone, or by email, but are always at the donor's convenience. Before the students do any work with the data donor, they sign a legally binding confidentiality contract. In industry and outside of the University this is necessary. For colleagues and student donors the contract is a courtesy. The student consultants then devise and implement a plan for answering the donor's question.

The weeks are often split into a day of project discussions and a day of working on the projects with my help immediately available. During the discussion days, weekly reports are handed out and the teams critique each other. This gives the students in other groups the opportunity to view statistical techniques that are not necessarily appropriate for their particular problem. The individuals in the team being critiqued usually receive positive criticism although when a team is seen to be slacking off the questions sometimes become quite pointed. Using feedback generated by their weekly reports and presentations, and occasional conversations with the data donors, the teams revise their ideas.

At the end of the semester students give the data donor a final written report derived from the weekly reports. When possible, the donor is invited onto campus for a presentation covering the main points in the report. The written report is essentially the statistical analysis section of the donor's paper or talk along with the students' interpretation of the data (to allow the donor to discern if there was any misunderstanding of the data itself) and the team's recommendations for additional work if needed.

The course is graded C/NC with an instructor's written evaluation of the student's contribution and knowledge that is added to the student's file. By the end of the 13 weeks these are usually easy to write.

Past data donors have included faculty members, administrators, and graduate students. These donors tended to be respondents to an annual plea at faculty meetings for help in finding projects. Those who used Math 301's consulting services once often used the services again or suggested that others make use of the class.

Somewhat less common were projects from off campus. These consultancies require that the instructor have contacts in local government, industry, and national epidemiological studies. While the topics of these projects are sometimes of greater interest to the students - their non-academic, real-world applications seemed to make them appealing - they are typically more difficult to obtain.

Early in the course the students are introduced to the five steps at which a statistician should contribute to a study:

1. Question determination.
2. Study design.
3. Data collection and maintenance.
4. Analysis.
5. Communication of results.

These form the basis for the reports and evaluations previously mentioned.

Depending upon the project that the students work on, they may be responsible for all, or a few, of the steps. In all cases they must communicate results in a technically correct, but user friendly form. This interdisciplinary interaction makes a consulting class particularly useful within a liberal arts setting.

Below are listed a few of the projects that the students were involved with over the last few years. The statistical techniques that were used have been included to try to give an indication of the level of knowledge that the students attain. Because much of the data comes from colleagues who were encouraged to get in touch with the instructor at the start of their projects, many of the projects allowed the students to get involved at the question determination stage.

Wes Bernardini is an archaeologist in the university's Department of Sociology and Anthropology. His work on Hopi Indian history led him to a problem with pottery sherd misclassification as bichrome (white and black) or polychrome (red, white, and black) during surface surveys (picking up pieces while walking across fields). The students used logistic regression and Bayesian models to help correct the over and undercounts that were taking place. They also suggested different methods for classifying sherds based upon attributes other than color.

As part of the university's Western Association of Schools and Colleges (WASC) accreditation process, Tony Mueller, Director of Community Service Learning, requested that students look at the effectiveness of the Community Service Activity Class (CSAC). While the analysis ultimately required nothing more exotic than contingency tables and trellis graphs, the students had to work hard to produce a document that accurately portrayed the information in the data to a number of non-statistical administrators.

The advisor of a graduate student in environmental studies suggested that she have the Math 301 students help her determine how to efficiently collect information on the predation of Mojave Desert tortoises by ravens. The data consisted of regurgitated pellet contents and nest location information. The students were able to help her design a reasonable collection scheme and analyze the collected remains using Poisson regression and spatial techniques (variograms, etc.).

The local Chamber of Commerce was seeking ways to attract students to downtown businesses. In describing the data that the Chamber had collected, students used a number of graphs and tables to show multivariate associations. This project was followed by a request from the Chamber for the creation and analysis of a survey to determine which types of businesses the residents of Redlands wanted in town. The second project showed the students the importance of their previous work

and required that they address the question determination, design, and data collection phases that they had missed in the original project.

Eliant is the largest consumer research company in the country that caters to the building industry, conducting over 300,000 homebuyer surveys annually. The company provides builders with information and strategies to improve the homebuyer's experience thereby increasing homebuyer satisfaction and building-industry rankings. Eliant's three homeowner satisfaction surveys generate, almost exclusively, ordinal data. After helping to restructure the surveys, the students looked at the association between likelihood of referral and satisfaction with various home ownership experiences using CART, neural networks, and ordinal logistic regression.

An analysis of data provided by CRAB (Cancer Research and Biostatistics, Seattle) dealt with finding genetic markers that are indicative of the effectiveness of specific treatment regimes for multiple myeloma patients. CRAB's data provided students with an opportunity to use microarray analysis.

Students in Math 301 looked at a subset of the variables from the Boston Area Health Study, which was selected by the principal epidemiologist. They were asked to model behavioral characteristics that the investigators had not had time to evaluate. One, non-math major member of the team used logistic regression to generate a report that was of near master's level quality. He also was able to convey the findings to the other team members. His interest in the project led him to consider biostatistics as a career path.

Math 301 is not only fun for the students. What makes this course interesting for the instructor is that it uses undergraduates with varied backgrounds and fresh ideas on projects that appeal to them. Math 301 does not create consulting statisticians. Instead it makes better consumers of statistics. Having been on the consultant side of the consultation, the students are better prepared to communicate with "real" statisticians when working on their own projects. Working through various parts of the consulting process, the students learn that calling a statistician early is a good idea.

The students are not the only ones who benefit from the course. The data donor gets a group of intelligent and motivated students and a Ph.D. statistician for free. In essence, the class acts as a consulting unit for the company without the company having to do any hiring.

The university has benefited from its new ties with industry and local government. On campus in both the sciences and the humanities, previously unknown techniques for dealing with statistical problems are trickling down into the classroom. Albeit on a smaller scale, the same benefits that are afforded by a graduate level statistical consulting course can be provided by an undergraduate consulting course.

JSM2007 Roundtable Summary: "Statistical Consulting in Medical Devices: Opportunities and Challenges"

Jianxiong Chu, Mathematical Statistician (Biomedical), General and Surgical Devices Branch, Division of Biostatistics, Office of Surveillance and Biometrics, Center for Devices and Radiological Health (CDRH), Food and Drug Administration *

*jianxiong.chu@fda.hhs.gov

We had a full table of 8 participants representing government (n=1), industry (n=3), academia (n=2), and independent consulting (n=2):

- Jianxiong Chu (Lead Discussant), CDRH, FDA
- Terri Henderson, Johnson & Johnson Vision Care, Inc.
- Mark Grant, BlueCross and BlueShield Association
- Françoise Iweins, Abbott Laboratories
- Madhu Mazumdar, Cornell University
- Michael Longnecker, Texas A & M University
- Andy Mugglin, University of Minnesota
- Alicia Toledano, Biostatistics Consulting, LLC

As a kick-off, the lead discussant briefly explained the legal definition of medical devices and how devices differ from drugs in terms of primary mode of action (*i.e.*, not through chemical action and not dependent on being metabolized). Regulatory pathways for medical device approval or clearance (IDE, PMA, HDE and 510K) were also briefly described.

We next discussed the abundant opportunities statisticians have in the development of new medical devices which are being accelerated by recent scientific advances in bioengineering, biomaterials, genomics and nanotechnology. Since a majority of over 20,000 device companies in the world are small with less than 100 employees and do not have stable in-house statisticians, we agreed that there will be an increasing demand for outside statistical consulting service in the medical device industry. Jianxiong also mentioned some opportunities within government agencies. For example, there is a constant need for statisticians to serve on the 21 medical devices advisory panel committees.

A brief discussion followed on some of the common statistical challenges in device trials (*e.g.*, small sample size, center effect, open-label, non-randomized control and historical control). Independent consultants Alicia and Andy expressed the need for more guidelines from FDA regarding device trials (in addition to the recent Draft Guidance for the Use of Bayesian Statistics in Medical Device Clinical Trials). They also pointed out that networking with statisticians with extensive experiences in the real world is particularly important for a relatively new statistician who just starts to get into the consulting business. We agreed that, to promote more professional interactions among statisticians working in the device area, there is an urgent need to hold an annual statistical conference exclusively for medical devices. We are looking forward to participating in the AdvaMed Statistics Conference for Devices and Diagnostics sometime in mid-April of 2008 (AdvaMed: Advanced Medical Technology Association).

In summary, we had a very pleasant and enjoyable luncheon discussion and hope that more and more statisticians will be interested in working in the device area so that together we can make a great contribution to promoting high-quality health care through devices.

Q & A: Question from a Section Member, Answered by a Section Member

Question: I am looking for professional and general liability insurance to cover my consulting activities. Any suggestion of a source or carrier?

Disclaimer: Insurance is a business decision that should be researched and decided upon based on each consultant's particular needs. The following is some general information that was provided to the questioner.

Answer:

1. If you are only consulting in the medical/pharmaceutical field then investigate RAPS (Regulatory Affairs Professional Society) or ACRP (Association of Clinical Research Professionals) for coverage options for their members.
2. If you consult outside of the medical field then an association to investigate is NASE (National Association of Self Employed). In addition, check rates from standard insurance carriers such as State Farm.

Resources for Improving Your E-mail Use

Karen Copeland, Boulder Statistics *

Given the events of August 15th, 2007 I know that there are many of us out there that can improve our email habits. What happened on August 15th? That was the day that a combination of an ASA server issues and "respond all" habits led to an explosion of messages to the consulting section members. I must admit that I had to laugh about the email havoc as two weeks prior I had read the book *Send: The Essential Guide to Email for Office and Home*. The main point to the book is to **think before you send**. Too often we are quick to hit the respond all button and fire off an email response. I would encourage all of you to read the *Send* book; it is a quick read and you can even skim and skip some of the later chapters (after a while you get the point and can move on). Go to <http://www.thinkbeforeyousend.com/> for more information about the book.

A second resource related to email I came across recently in a blog. It is a website and video from a site called 43folders about what they call "Inbox Zero." I listened to the video and cleaned out my inbox as I did down to zero messages. The site and video had great tips for releasing yourself from email overload. The website to visit is <http://www.43folders.com/izero/>. I will admit that over time I have not kept my inbox completely free of messages, however, the principle ideas of the Inbox Zero makes sense and has helped me to keep my inbox under control.

*karen@boulderstats.com

Comments from the Chair

Stuart Gansky, University of California, San Francisco *

- Listservs and Lessons
- JSM Invited Program
- e-Directory of Statistical Consultants

Listservs and Lessons

In August, I posted a call for JSM invited sessions to all members of the Statistical Consulting Section, and a flurry of replies and unsubscribe requests jammed mailboxes of members. This was after many of us thought steps had already been taken to avoid such problems. I apologize for any inconvenience this may have caused. Partly as a result of this small mess, ASA stepped up its implementation of a new system for email messages to members. Thus, I am happy to report this incident cannot be repeated. We will continue to post messages to the full membership judiciously about 6-8 times per year for newsletters, calls for sessions, and other timely matters.

JSM Invited Program

Despite the email snafu, Jamie Grady has done a terrific job planning the Section's invited program for the Joint Statistical Meetings in Denver (Aug 3 - 7, 2008) which has the theme "Communicating Statistics: Speaking Out and Reaching Out". In addition to our 2 allocated sessions, one of the 2 sessions our section submitted for the section competition was accepted.

- Tues 10:30-12:20 - Consulting Service Models for Academic Medical Centers: What Works, What Doesn't Work, Reaching Out and Filling the Growing Need (Shelly Hurwitz)
- Wed 8:30-10:20 - Statistical Consulting and Collaboration in Private Industries (I-Li Lu)
- Thurs 8:30-10:20 - Communicating Statistics: Speaking Out and Reaching Out (Stephan Ogenstad)

e-Directory of Statistical Consultants

We continue in the process of rolling out the Section E-Directory, a web-based database of Statistical Consultants. Please verify your profile in the MembersOnly section of www.amstat.org. Shortly we will enable the search function by content area or geographical region. This will allow clients to find consultants who fit their needs.

Please continue to contact a member of the Executive Committee with ideas for section services. As the year comes to a close, I will note that it has been a privilege to serve the section this year. I know the section is in good hands with Brenda Gaydos becoming Chair in 2008.

Sincerely
Stuart Gansky, Chair

*sgansky@itsa.ucsf.edu

Notes from the Editor

Karen Copeland, Boulder Statistics[†]

This issue of the newsletter brings me to the end of my term of service as editor. I want to thank all of you who have provided me with material for the newsletter over the past two years. My goal with the newsletter was to collect articles about consulting that would be of interest to the members; your submissions helped me in that quest. Next year Chris Holloman is stepping up to the editor position. If you are interested in getting involved with the section we are looking for an assistant editor to help Chris (contact Brenda Gaydos at gaydos.brenda@lilly.com if you are interested). Your time commitment would be a few hours three times a year.

Remember, the newsletter is for the section members at large. If you have a topic you would like to see covered or have an article you would like to share then please do so. Send articles and topic ideas to Chris at holloman@stat.osu.edu.

[†]karen@boulderstats.com

Karen Copeland, Editor

Contact information: Boulder Statistics
2456 Vine Place
Boulder, CO 80304
Phone: 720-564-0661
karen@boulderstats.com

Christopher Holloman, Assistant Editor

Contact information: The Ohio State University
Cockins Hall
1958 Neil Ave.
Columbus, OH 43210
Phone: 614-292-0738
holloman@stat.ohio-state.edu

**Remember the
Section on Statistical Consulting
website address:**

www.amstat.org/sections/cns1/

You'll find all information relating to the Section on Statistical Consulting, including our charter, officer list, section activities, past issues of *The Statistical Consultant*, minutes of past meetings, and more.