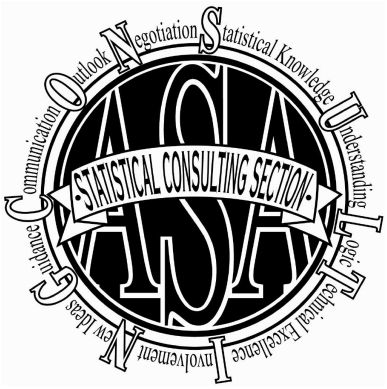
THE STATISTICAL CONSULTANT



Section on Statistical Consulting
Tzu-Cheg Kao, Co-Editor; Karen Copeland, Co-Editor

American Statistical Association
Summer 2005; Volume 22, No. 2

IN THIS ISSUE

Newsletter to go Electronic

The Executive Committee of our Section is working to move the newsletter to an electronic format. The intent is to eliminate many of the costs associated with the newsletter and funnel these funds into other section programs such as the round table conference calls. An electronic format will also reduce space restrictions on articles so we can add more content to the newsletter.

To make sure you receive the electronic newsletter (or, more likely, a link to the newsletter) it is **important** that we have a current email address for you. To verify that ASA has your correct email address, follow the "Members Only" link at `www.amstat.org` and log in to your account. (You will need your Member ID to log in.) Once there, verify the email address listed in your address record.

Feedback, comments, questions, concerns, article ideas, author ideas, etc. can be sent to Karen Copeland at `karen@boulderstats.com`.

Executive Committee of the ASA Section on Statistical Consulting

Chair: Susan J. Devlin, (732) 424-2498, sdevlin@theartemisgroup.com
Chair-Elect: Philip M. Dixon, (515) 294-2142, pdixon@iastate.edu
Past Chair: Jane F. Pendergast, (319) 384-5028 jane-pendergast@uiowa.edu
Program Chair (for 2005 JSM): Susan P. McGorray, (352) 294-0019, spmcg@biostat.ufl.edu
Program Chair-Elect (for 2006 JSM): Todd G. Nick, (513) 636-0462, todd.nick@cchmc.org
Secretary / Treasurer: Elaine I. Allen, (781) 239-6413, allenie@babson.edu
Publications Officer: Christina M. Gullion, (503) 335-6356, christina.gullion@kpchr.org
Webmaster: Chuck Kincaid, (269) 344-4100, ckincaid@comsys.com
Newsletter Co-Editor: Tzu-Cheg Kao, (301) 295-9756, tkao@usuhs.mil
Newsletter Co-Editor: Karen Copeland, (720) 564-0661, karen@boulderstats.com
Executive Committee: Anamaria S. Kazanis, (734)769-7100 x 6206, akazanis@umich.edu
Timothy A. Max, (503) 808-2060, tmax@fs.fed.us
Council of Sections Representatives: Keith E. Muller, (919) 966-7272, keith_muller@unc.edu
Brenda L. Gaydos, (317) 277-1982, blg@lilly.com
Council of Sections Governing Board Vice Chair: Russell V. Lenth, (319) 335-0814, russell-lenth@uiowa.edu

Announcement of Travel Awards Competition for 2006

The Section on Statistical Consulting invites proposals for Topic (Special) Contributed Paper Sessions for the Joint Statistical Meetings to be held August 6-10, 2006, in Seattle, Washington.

Proposals submitted will compete for up to three travel awards. Each award consists of $500 and a registration fee waiver for the organizer. Each organizer will be responsible for lining up speakers and for coordinating submission of abstracts in the fall of 2005.

A proposal should consist of a one-page description of the session, including a description of the session theme, the types of papers or discussions to be included, and the intended audience. A list of possible speakers is helpful.

The proposal submission deadline is September 1, 2005. Feel free to contact the program chair, Todd Nick, for additional information or to discuss ideas.

Send your proposal to:

Todd G. Nick, Ph.D.
Cincinnati Children's Hospital Medical Center
Center for Epidemiology and Biostatistics
3333 Burnet Ave, MLC #5041, Cincinnati, OH 45229
Phone: (513) 636-0462, Fax: (513) 636-1254, Email: todd.nick@cchmc.org

JSM 2005 Section on Statistical Consulting Events

A special thank you goes out to Sue McGorray for her work as the program chair for JSM 2005. Our section's line-up includes invited sessions, roundtable lunches, and three topic-contributed sessions arranged by travel award winners. These travel award winners will receive $500 and a registration fee waiver for their efforts.

Do you have an idea for a session for 2006? Consider entering the 2006 travel award competition for a chance to offset some of your travel expenses!

In addition to the technical program events listed below, the section business meeting and mixer will be held on Monday, August 8, from 5:30 to 7:00 p.m. Come for food, section business, door prizes, and networking with other section members.

Evaluation of Statistical Consulting Programs in an Academic Setting
Sunday, August 7, 4:00 p.m.
Topic-contributed papers (travel award winner)
Organizer and Chair: Manuela Huso

Practical Guidance for Collaborators and Their Clients
Monday, August 8, 8:30 a.m.
Invited papers
Organizer and Chair: Harold Dyck

Working with Clients and Researchers: Communication is the Key
Monday, August 8, 12:30 p.m.
Roundtable luncheon (fee event)
Organizer: Todd G. Nick
Presenter: Thomas Loughin

Recent Work in Power and Sample Size
Tuesday, August 9, 10:30 a.m.
Invited papers
Organizer and Chair: John Castelloe

Study Design for Genetic Epidemiology
Wednesday, August 10, 10:30 a.m.
Topic-contributed papers (travel award winner)
Organizer and Chair: Stephen Lake

Delivering Client-Focused Results
Wednesday, August 10, 12:30 p.m.
Roundtable luncheon (fee event)
Organizer: Todd G. Nick
Presenter: Karl Heiner

Career Success: Statisticians Collaborating and Thinking Beyond Statistics
Thursday, August 11, 8:30 a.m.
Invited (joint sponsored)
Organizer and Chair: Janet Myhre

Statistical Consulting: From Client Acquisition to Project Reporting
Thursday, August 11, 8:30 a.m.
Topic-contributed panel (travel award winner)
Organizer: Karen Copeland
Chair: Sue McGorray

Budgeting: Data Management and Statistical Analyses

Anamaria S. Kazanis, MA, VA Ann Arbor Healthcare System*

A statistical analysis is only as good as the data being analyzed. And data are only as good as the process used to generate them.

When budgeting for a statistical analyst to participate in a research project, funds need to be assigned for three stages of a research project: the proposal, the project, and subsequent publications. The following article is adapted from various project proposals, and it illustrates many of the line items needed in a budget for the data collection and statistical analysis portions of a project.

The three stages are each assigned to a section of the article with details on the needs for that stage outlined within the section. These particular projects were submitted to and funded by the National Institutes of Health (NIH) from 2000 through 2004.

Proposal

The statistical analyst should be engaged for a minimum of 10 to 30 hours to provide input on focus groups, power analysis/sample size considerations, research design and methods, and pilot or preliminary studies.

The budget line items for this stage of the project could include data collection, data coding and documentation, data entry, data management, statistical analyses, and necessary software and hardware.

Focus groups of experts on the field and populations similar to the study's target need to be convened to provide input on the appropriateness of the theoretical framework and measures to be applied.

A preliminary power analysis is performed to determine the minimum sample size necessary to detect significant effects on the dependent measures. With this information, the investigators and analyst jointly produce the research design, methods and budget sections of the study proposal.

The statistical analyst should look for threats to validity and verify that the planned procedures and size of the study will be adequate to address its goals. She can also suggest ways to maximize the efficient use of the available resources and contribute to the decision-making process.

Project

Data Collection

Computers and peripheral hardware, paper and pencil, or a combination of both media could be used for data collection. It might be necessary to engage the services of not only a statistical analyst but also a data manager and, possibly, a computer programmer to assist in the data collection stage.

During the questionnaire-design process, the statistical analyst will assign the formatting of variables in a manner consistent with his role in future statistical analyses and the data manager will be in charge of compliance and documentation.

The electronic format can be labor intensive at the onset of the project. This format requires the participation of a programmer engaged for 20 to 30 hours of programming per instrument/measure.

The statistical analyst should be available for consultation with the programmer throughout the process and should allow about four to six hours for each instrument/measure to be programmed.

*VA Ann Arbor Healthcare System, Health Science Research & Development, 2215 Fuller Road, Ann Arbor, MI 48105, akazanis@umich.edu

The traditional pencil-and-paper format requires the engagement of a substantial number of research assistants to code and enter the data. The integrity of the resulting data is only as good as the coding book, the coders, and the data entry personnel.

The data manager will supervise all aspects of generation, verification and storage of the data, requiring his services each time for four to six hours per instrument/measure. There is an inverse exponential relationship between the number and severity of errors and the amount of expert statistical verification of the data.

The mixed-media format could require the engagement of a statistical analyst, a data manager, computer programmers and research assistants to program, code, enter, and verify the data. A great number of studies use this medium since it provides the greatest flexibility.

Data Coding

Codebooks need to be written with the participation of a statistical analyst in the selection of the appropriate codes and formatting of the data, with special attention to the cases of missing, unknown, uncodeable, "too large," or conflicting data.

The data manager is responsible for writing the coding guidelines as well as continuous updating of any and all changes to the codebooks. The information contained in the codebook is programmed in the instrument when electronic data collection is used.

If the pencil-and-paper format is chosen, each survey will need to be manually coded. It is important to retain the services of a statistical analyst as consultant throughout the coding process, allotting about four to six hours for each instrument/measure.

Double coding is the preferred method of professionals of the field. This process consists of two people coding the data. The reduction on the number and complexity of possible errors, with their exponential ramifications in time, effort, and cost, totally justify this method.

Data entry is needed for data recorded by the pencil-and-paper and mixed-media formats. The electronic format for data collection transfers the data in digital form to an environment compatible with established statistical software packages, thus eliminating the need for data entry.

Once the data have been entered, the data manager/analyst will need at least three to five weeks to familiarize himself with the characteristics of the data and run univariate statistics.

Every single error from data entry compounds itself during data management to represent over 100 hours of work searching for the source and nature of the error then correcting and documenting.

Professional data entry requires a formatting program written for each version of the questionnaire. Changes in questions will affect the formatting program and require payment for the development of a new program.

Each questionnaire should be keyed twice: once in the entry mode of the program and then a second time in a verify mode. During the verify pass, for each field displayed, if the two entries do not match, the program alerts the verify operator who then reconciles the data. This process allows for near 100 percent accuracy in the final output file.

A preferred format for data is ASCII fixed format with no delimiters. This type of data file determines the separation of the raw data from all the data management environments.

An informal survey of various data entry companies produced the following cost estimates, in 2004 dollars:

a. Program Setup: \$70 to \$75 per hour

b. Data Entry:

	0.4¢–0.6¢ per char. keyed
+	0.4¢–0.6¢ per char. verified
or	0.6¢–0.7¢ per char. keyed & verified

An example follows: $N = 500$ subjects, with pretest, post-test, and 3-, 6-, and 12-month follow-up (FU) questionnaires. Assume that all follow-up questionnaires are similar in layout and design. Some time and cost estimates are presented in the following table.

	PROGRAM			DATA ENTRY (per subject)			
Instrument	**Characters**	**Time**	**Set Up**	**Keyed**		**Verified**	
Pretest	1,000	7 hr	\$490–\$525	\$5		\$5	
Post-test	600	5 hr	\$350–\$375	\$3		\$3	
3-mo FU	600	5 hr	\$350–\$375	\$3		\$3	
6-mo FU	600	1 hr	\$70–\$75	\$3		\$3	
12-mo FU	600	1 hr	\$70–\$75	\$3		\$3	
Subtotal	3,400	19 hr	\$1,330–\$1,425	\$17	+	\$17	= \$34
TOTAL	**\$18,330–\$18,425 for $N = 500$**						

Data Management

The data manager and the statistical analyst could be the same person depending on the size, length of time, and complexity of the project. Detailed documentation should be kept, starting with data entry and ending with "final" data files created by cleaning data and, if appropriate, merging data from various instruments or different data collection points (Antonakos, Miller and Caruso, 2002).

The data manager should be hired from the beginning of the project to allow for familiarization with the data and facilitate the resolution of issues appearing at later dates. The data manager and the statistical analyst should be part of the decision-making and supervision processes at all stages of data coding, entry, checking, cleaning, merging, storage, or other manipulation.

The data manager will perform the actual operations related to the maintenance of the data files as well as supervise and implement confidentiality of subject information, data integrity, and security (Berman, Ganksy, Guillion, Loughin and Sanchez, 2003).

It is important to avoid the misguided use of project funds contracting less-expensive or less-skilled workers to perform data management tasks. Even though the proposed statistical framework would ensure the quality of the research design and methods, the data set may be plagued with mistakes due to lack of skill in data handling and processing. This results in unrealistic time frames assigned to the analysis processes (Antonakos and Kazanis, 2003).

Statistical Analyses

The statistician needs to have a complete, detailed description of the study design and a clear exposition of the questions to be addressed.

As the project progresses, the statistical analyst will supervise all aspects related to data production and collection. This professional will examine the data for threats to validity, ranging from missing data to questionable outliers to confounders (Berman, Ganksy, Guillion, Loughin and Sanchez, 2003).

The statistical analyst should be engaged for four to six hours per week for each instrument/measure to be analyzed. Once pretest data are clean and all relevant univariate statistics are generated, the analyst will evaluate the validity of the cross-sectional aspects

of both the aims and hypotheses. The amount of time required to perform these tasks is directly proportional to the complexity of the data and hypotheses to be tested.

Allowances need to be given for regular meeting time between the members of the research team and for trouble-shooting meetings as circumstances require. Data sets collected over time might have an added level of complexity due to changing software and/or hardware versions, as well as the manner in which the new versions interact with existing versions.

Update: Software and Hardware

Software and hardware needs for the project should be included as a line item in the budget if appropriate allowances for upgrades, maintenance, and training of personnel during the course of the project should also be included in the budget.

Publications

Research projects often generate a series of articles, abstracts, presentations, and other means of presenting the results of the study as it progresses. It is only practical to include the statistical analyst as part of the team of authors writing data-based articles.

A research program is done by a research team, not just by the principal investigator. Therefore, research staff should contribute as co-authors and senior staff should be given latitude in pursuing independent contributions to ensure the success of the research program (Antonakos and Kazanis, 2003). Funds should be assigned for the time used by those members of the team co-authoring papers on the study.

Summary

It is the intention of this article to motivate sensitivity towards proper budgeting itemization in order to account for all aspects of the participation of statisticians in a grant project. It has been the experience of this author that projects tend to be too short on funds shortly after data collection concludes, to the detriment of the analysis and writing processes.

References

Antonakos, C.L. and Kazanis, A.S. (2003), "Research Process in the Health Sciences: A Focus on Methods," *Research and Theory for Nursing Practice: An International Journal* 17(3), 257-264.

Antonakos, C.L., Miller, J.M. and Caruso, C.C. (2002), "Critical Elements of Documentation for Data-Based Research," *Western Journal of Nursing Research* 24(1), 87-100.

Berman, N., Ganksy, S., Guillion, C., Loughin, T. and Sanchez, M. (2003), "When You Consult A Statistician - What To Expect," *The Statistical Consultant* 20(3).

Get Out From Behind Your Desk

Fritz Scheuren, President, American Statistical Association*

Some of you may remember the blunt style of W. Edwards Deming whose words I am paraphrasing in the title above. Now, Deming detested all "puffery" and overstatement. So his urging that statisticians get out from behind their desks was no mere turn of words.

In the 1930s when Deming began practicing, statisticians worked mainly in places like the U.S. Census Bureau, the U.S. Department of Agriculture (where he started) or Western Electric. Statistics only became a widely recognized Ph.D. granting academic discipline in the 1940s.

Most of us had well-defined roles in those days; actually, narrowly-defined might be a better way to put it. We were often human computers. That tradition continues, of course. For example, tongue in cheek, the Australian Bureau of Statistics, has a singing group called "Calculating Women." With good voices too, I might add.

But statisticians with roles that are defined by their place (desk) in a large organization are declining. More and more, we need to define ourselves, perhaps as an internal consultant, perhaps on our own. That self-defined role is a lot more challenging, risky, and, I would add, fun, too!

Of course, to create our own role we have to get better at our craft, but that craft is one of service giving, not just product creation.

"Systems thinking," one of Deming's key ideas, can seldom be done successfully alone. You need to do it with a client, seeing it through their eyes–feeling their problem with two sets of hands, theirs and yours.

So people skills, especially listening skills have to be given more emphasis in our statistical training. This has been recognized in the consulting labs that I see emerging in many academic departments of statistics.

Better still would be more workshops and mentoring experiences with the masters of this art. That is clearly a niche for the Statistical Consulting Section.

Periodically worries are expressed about our future as a profession. Well, they may be in part justified, but I am not really concerned, as long as our skills as consultants continue to be honed.

Deming was, in many ways, a role model for consultants. He wrote on it and lived most of his life as one. He never let events come to him. He sought out challenge. Best of luck in doing likewise.

EDITOR'S NOTE: Fritz has been very active for the last 11 years as a consultant on business and, more recently, human rights problems. He was formerly the Chief Mathematical Statistician at the Social Security Administration and Director of Statistics at the Internal Revenue Service.

*American Statistical Association, 1402 Ruffner Rd, Alexandria, Virginia 22302-4217, scheuren@aol.com

Comments from the Chair

Susan J. Devlin, The Artemis Group*

ASA President Fritz Scheuren's comments in this newsletter emphasize the importance of the statistical consultant to the future of statistics. This highlights an important leadership role our section could play in ASA. To meet this challenge, we need to find even more ways to support our members and the ASA.

Successful statistical consulting requires strong statistical skills and the ability to see their application to solve business and research problems. In addition, building a statistical practice has many administrative challenges, including marketing and sales, pricing, and systems support. This is especially true for small or solo practices.

Again in this issue, I am using this column to field some new ideas, request your feedback, and solicit volunteers.

Consultant Referral Website Status. Thanks for all the feedback I received on my note in the spring newsletter about the website. We received a lot of constructive input.

A planning committee has been appointed. I hope to present recommendations at the business meeting in Minneapolis. If we do not run into any legal or financial roadblocks, the winter newsletter will have an article about the plan and implementation time line.

Isolated Consultant e-Support. A network of professional friends and associates can be very instrumental to growth in statistical and business savvy. This is easiest when in a large corporation or academic institution with other statisticians around us, and corporate resources are available.

But the solo consultant isolated either geographically or organizationally is not as fortunate. This became very clear to me in 1997, when I retired from Bellcore (now Telecordia Technologies) to co-found The Artemis Group.

Having my own small group provided the freedom I sought and more hands-on time, which I love. The price was the loss of an extensive technical library and reference service, large marketing and sales department, and administrative support. The most important loss, however, was the ability to walk down the hall to a colleague to get advice.

Recently, George Cox told me about the Isolated Statistician electronic list group. These academics are often the only statistician at their institutions. I joined and was very impressed with the support this provides: quick answers to questions on textbooks, technical issues, software, teaching ideas, conferences, etc.

Is there a need for an Isolated Statistical Consultant group list? Is there such a support structure in place that we should help our members tap? Let me know your thoughts. If this is an idea to pursue, we would need a sponsor/administrator for the list. Any volunteers?

Rates Survey. Frequently, I have been asked what the going rates are for statistical consultants. Understanding what rates are reasonable and competitive can be key to the financial success of a business.

I propose that the section periodically conduct a simple electronic survey of our members concerning the rates their institutions charge for different levels of statistical support. We could then provide a range of typical rates, showing variation by experience, location, and type of consulting.

Do you think this is a valuable initiative for the section? Who is willing to help with this project?

*The Artemis Group, 33 Harvey Drive, Summit, NJ 07901-1204, sdevlin@theartemisgroup.com

Notes from the Editor

Thanks to our Section chair, Susan Devlin, for recruiting Karen Copeland to serve as the editor of *The Statistical Consultant* for 2006. Special thanks to Karen for editing a significant part of the current issue. During the transition, I have invited her to serve as the co-editor. My term as editor will expire at the end of 2005.

Again, I would like to thank Karla Genter for her continuing assistance in typesetting *The Statistical Consultant*.

As always, we would like to encourage our section members to share with us feedback, comments, questions, concerns, article ideas, author ideas, and so on that are interesting to our members. Feel free to contact Karen or me.

Submission deadlines for future issues are October 1, February 1, and May 1 for Winter, Spring and Summer issues.

Tzu-Cheg Kao, Co-Editor

Contact information: Division of Epidemiology and Biostatistics
Department of Preventive Medicine and Biometrics
Uniformed Services University of the Health Sciences
4301 Jones Bridge Road
Bethesda, MD 20814-4799
Phone: 301-295-9756
Fax: 301-295-1854
tkao@usuhs.mil

Karen Copeland, Co-Editor

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

Remember the Section on Statistical Consulting website address:

www.amstat.org/sections/cnsl/

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.

THE STATISTICAL CONSULTANT
Tzu-Cheg Kao, Co-Editor
Karen Copeland, Co-Editor
c/o American Statistical Association
1429 Duke St
Alexandria, VA 22314-3402