

Providing Data Analysis Support at a Major Research University using the SAS® System

Joy Munk Smith, Department of Statistics, NCSU

Sandra Biggs Donaghy, Department of Statistics, NCSU

INTRODUCTION

North Carolina State University (NCSU) is a large research university with researchers and students in many departments who require data analysis assistance. As statistical computing consultants in the Department of Statistics at NCSU, we provide statistical consulting services to campus members. Our service includes individualized consulting to researchers and graduate students as well as teaching SAS short courses.

In recent years, staffing levels have been cut and incredible technological changes have taken place. There is now a large diversity of computing platforms and data entry methods. Our clients are becoming more computer literate and discovering new features they want to explore. As a result, their need for assistance is growing. These developments require us to search for more efficient ways to meet our clients' needs.

In this paper we will describe the statistical consultation we provide and the resources we use to assist our clients. While an increasing number of researchers and students are able to perform a portion of their data analysis, communication between the statistical computing consultants and researchers greatly improves the overall quality of the reported results. Ensuring the quality and accuracy of data and data analysis is one of our most significant contributions.

DEPARTMENTAL HISTORY

At fifty-four years of age, the Department of Statistics at NCSU is one of the oldest statistics departments in the United States. The Department's formation was motivated by the

need for efficient experimental design and valid analysis of research data in agriculture and life sciences. The department's mission, as advanced by our first department head, Gertrude Cox, is to develop, teach and apply the science of statistics. Since its inception the Department has remained closer to the field of applied statistics than perhaps any other department in the United States. Supported by the North Carolina Agricultural Research Service, the faculty and staff of the Department provide statistical consultation to the faculty and graduate students in the College of Agriculture and Life Sciences. Additional funding from the Institute of Statistics allows the Department to extend these efforts, as much as resources allow, to the entire campus.

JOB RESPONSIBILITIES

Consulting and Quality Assurance

Our consulting services are available to all faculty, staff, and graduate students affiliated with NCSU, although the majority of our clients are in the College of Agriculture and Life Sciences. Most of these clients are graduate students or research scientists.

The predominate software package used for data analysis at NCSU is the SAS System. The development of the SAS software began at NCSU in the sixties to facilitate analysis of research data and the software has continued to be used since that time. The expanded use of the SAS System is due to the quality and depth of the SAS System's data management and analysis tools, the exceptional user support the Institute provides, and the price discounts the Institute provides to degree granting institutions. The products we use most frequently include Base SAS®, SAS/STAT®, SAS/IML®,

SAS/FSP®, SAS/GRAPH® and
SAS/ACCESS®.

Clients at NCSU use several releases of SAS including version 5.18 through version 6.08 on the IBM® MVS mainframe, version 6.09 on UNIX workstations, and version 6.04 through version 6.11 on personal computers. Although SAS software is largely system independent, some version and hardware differences and incompatibilities remain.

One of our primary goals is to ensure the high quality of data analysis performed and reported by students and researchers affiliated with NCSU. We also attempt to provide our clients with a "User Friendly Interface" to the SAS System.

Our clients' knowledge of statistics and computing varies widely, so our approach to consultation is adjusted for their specific background. In our consultation we make a conscious effort to ask the appropriate questions to help solve our clients' problems. A number of clients are very well versed in math, statistics and computing, and only require answers to specific problems and questions. Some clients, however, lack the appropriate training to plan, perform and understand the needed analysis. These clients require a significant amount of assistance. They have to work especially hard in order to understand their analysis. Other clients are computer phobic and dread using computers or even trying to perform an analysis. These students require special handling. In addition there are some clients who want their data analysis done for them. However, our departmental policy requires students in degree programs to perform their data analysis themselves. We provide assistance and answer questions, but we do not write all their code, shoulder the responsibilities for data quality or write reports.

Due to our close relationship with the College of Agriculture and Life Sciences, we provide a high level of service to a number of the College's research scientists. We maintain databases from year to year, perform complicated analysis, generate graphs and tables, and create data entry screens for some of these clients. Staffing levels prohibit us from

providing this level of service to the entire university population.

Training - SAS short courses

We are able to improve our clients' knowledge of the SAS System through short courses. This training reduces the need for time consuming one-to-one explanations. Many of our clients find that having a basic understanding of SAS software helps them acquire a job when they graduate. We take training very seriously and are always looking for new ways to enhance our teaching effectiveness.

The availability of computer-equipped classrooms allows us to provide hands-on training which is more effective than lectures. We teach SAS courses at the beginning of each semester and again in the summer. The short course is approximately thirteen hours of instruction. These classes are taught in a lab equipped with 35 UNIX workstations. Each workstation is normally shared by two students. Classes are usually full even though the short courses are not required and students receive no credit. In the past we wrote all of our own training material; however, we have found that SAS trainer kits provide an excellent source of material. We integrate some material from the trainer kits with our own examples and site specific information.

In the future, we would like to add a few one day classes in specific areas, such as SAS/INSIGHT®, SAS/IML®, SAS/GRAPH®, SAS/TUTOR® and SAS/ASSIST®. The new Statistics Application, a graphic interface for statistical analysis under development by SAS Institute, looks promising. We will follow its development and possibly incorporate it into our classes.

One of the major difficulties we encounter in our teaching is dealing with the wide range of students' computing and statistical skill levels. It is hard to pace a class to suit everyone. We keep the pace fairly fast and encourage students who are having trouble to practice the class examples. We set aside practice time after each class, and we are available to answer their questions. We include practice exercises with every class handout.

We maintain a library of SAS documentation that is available for check out by class participants. This library includes introductory SAS manuals and books by SAS users.

Training - Other Resources

The Department of Statistics teaches a semester course in statistical computing that makes extensive use of the SAS System. In addition, a supervised consulting course is required for all statistics graduate students. This course is designed to train future consulting statisticians by exposing them to applied statistical problems.

The NCSU computing center staff also teach SAS classes. Their classes are less statistical than our short course and are held at different times during the semester. These additional course offerings help satisfy the growing demand for SAS training.

NCSU licenses the SAS/TUTOR product and purchases many of the SAS training videos. We store data sets from the instructor kits on-line and allow students to check out the corresponding SAS Course Notes. We inform our clients of the Institute's Public Training Classes, a great value especially with the 50% discount offered to individuals affiliated with degree granting institutions.

CHALLENGES AND REWARDS

Providing statistical computing support is both challenging and rewarding. We provide support by appointment, telephone and e-mail. We face an endless stream of clients with an interesting array of research data to analyze. We work with researchers from such areas as air and water quality, psychology, sociology, genetics, crop breeding, variety testing, milk production, educational testing, soils and marketing. While some problems are difficult, we view them as a challenge and learning opportunity.

Most of our clients are appreciative of our efforts, the guidance we provide and the time we save them. A short consulting session to answer some basic questions can save them hours or even days. Most students are under great pressure to complete their research and graduate, and we often are able to reduce their stress by providing assistance.

Our department supports our professional development by providing time and partial funding to attend SAS Users Group meetings locally, regionally and nationally. These meetings provide an opportunity for us to share experiences with other users and meet with the SAS Institute staff who announce new software features and Institute plans for future releases. Attending the meetings allows us to expand our knowledge of available tools and renews our enthusiasm.

Documentation has always been a dilemma. We use the SAS subscription service to ensure a supply of current manuals, but it is impossible to read even a small percentage of the documentation. Explaining to the novice user where to look for information in SAS manuals is a struggle as well. We have not found any good solutions to these problems but are able to help them locate the necessary documentation. Perhaps a CD-ROM version of the documentation with a good search engine will help.

We provide some in-house documentation and plan to take advantage of the Institute's SelecText program. In the past, all of our in-house documentation has been on paper, but we are currently working to make it available electronically.

Being familiar with current information pertinent to statistical computing is central to our effectiveness. This is very challenging due to the rapid development of the SAS System, the proliferation of SAS books, the growth of the Internet, and the rapid evolution of computing. Many of the resources we use are listed in the footnote section.

While the SAS System is largely platform independent, there still remain technical difficulties in supporting SAS software in a distributed computing environment. Transporting data sets and printing are the source of many of these problems.

STATISTICAL CONSULTING

In statistical consulting we provide guidance regarding appropriate data management, data

analysis and statistical interpretation. We often provide our clients with first-level statistical support, and in many cases solve their problems without assistance from the statistics faculty.

The statistics department faculty provides in-depth statistical consulting. Each researcher or student decides when to request assistance from the statistics faculty. Our clients benefit from consulting with departmental faculty, and we encourage them to take advantage of this service.

Students and researchers with a weak background in statistics may not understand the directions they receive from faculty members during consulting sessions. We can usually explain the directions because we work with the faculty regularly and are familiar with how they approach data analysis. However, clients who are untrained in statistics are at a disadvantage. Students who are unprepared become very frustrated and confused. Our department tries to require clients to have at least two graduate level statistics courses.

We encourage graduate students to communicate regularly with their committee members and share the progress of their data analysis. Possible differences of opinion regarding the best data analysis methods are best resolved very early in the process.

We stress to our clients the need for data checking. Checking the data carefully before running analyses can prevent the need to repeat the analysis multiple times. We remind our clients that the absence of error messages is no guarantee that their analysis is correct. Only correct and careful analysis can ensure accurate results. Although this seems obvious, caution is often discarded, data is not carefully checked, and errors result. It is amazing how easily and how often this happens. We help our clients uncover problems that could cause delays and embarrassment.

CONSULTANT'S TOOLBOX

We search high and low for solutions to our statistical computing problems. The reference

and footnote sections of this paper list many of the resources we use.

Our large library of SAS manuals is essential. We look forward to online availability of documentation with the appropriate search tools. In addition to the reference manuals, we use the Master Index to SAS System Documentation, and the Master Index to SUGI Proceedings to search for specific topics. The SUGI and SESUG Proceedings are a valuable resource for information. Articles in "Observations®: The Technical Journal for SAS Software Users" are packed with detailed information. While all the articles do not apply to our work, the ones that do provide a thorough coverage of the topic.

In addition to written documentation, SAS Institute provides online information. Their staff continues to enhance the information available from their World Wide Web site¹. The Institute provides technical information in the technical notes² directory and outstanding problems and usage notes in the SAS Notes³. The programs and data sets from "Observations"⁴, graphics macros⁵ written by Michael Friendly and other information can be downloaded from their ftp site⁶.

The SAS-L news group⁷ has a knowledgeable and active membership who answer questions in a skilled and professional manner. We search the SAS-L archives⁸ for answers to specific questions. Since Netscape version 2.02 appears to have a problem accessing these archives due to the length of the file names, we use Mosaic version 2.1.1 for searching the SAS-L archives.

The SAS File Contribution Server⁹, developed in 1995, is an Internet repository for user contributed SAS programs, catalogs and SAS/Toolkit applications. Occasionally there are programs of interest to our clients at this site.

The NCSU computing center staff maintains a help desk and a help database¹⁰. Their consultants are knowledgeable, and their database is rich with information directly related to our campus computing needs. Their support personnel are the authors of several useful handouts.

We share years of accumulated sample programs with our clients. Sample programs often enable them to work more independently. Several members of our department have written documentation. This documentation may provide general introductions for students in a specific class or detailed information on specific tasks. Work continues on making these documents available from our department's World Wide Web home page¹¹.

As a last resort we call SAS Institute's Technical Support Division. Their support staff is always helpful. Permission to call the Institute is granted to a limited number of NCSU support personnel.

In addition to these SAS resources, we have found DBMS/COPY¹² software to be a user friendly package that moves data sets in and out of SAS easily and accurately.

CONCLUSION

The diverse support we provide includes answering questions about SAS installation and programming, statistical analysis, statistical interpretation, computing platforms and other software packages. There are always many problems to solve and never enough time. We must be familiar with the latest information and the best ways to find solutions to our problems. Difficult problems can be enjoyable because they provide an opportunity to learn new techniques.

The consulting and training we provide helps ensure accurate data analysis and provides statistical methods training for present and future researchers. The statistical and computing consulting service we offer is highly valued by our clients.

ACKNOWLEDGMENTS

We would like to thank Dr. Tom Gerig, Head of the Department of Statistics at NCSU, for his support. We also sincerely appreciate the assistance and guidance we receive from our departmental faculty.

REFERENCES

SAS Institute Inc., Master Index to SAS System Documentation, Version 6, Fourth Edition, Cary, NC: SAS Institute Inc., 1992

SAS Institute Inc., Master Index to SUGI Proceedings, 1995 Edition, Cary, NC: SAS Institute Inc., 1995

SAS Institute Inc., "Observations, The Technical Journal for SAS Software Users", Cary, NC: SAS Institute Inc.

FOOTNOTES

¹SAS Institute Home Page
<http://www.sas.com/>

²Directory of SAS technical notes
<http://www.sas.com/techsup/download/technote>

³Index of Version 6 SAS Notes
<http://www.sas.com/www-bin/unotes-V6-wais.pl>

⁴Directory of "Observations" information
<ftp://ftp.sas.com/techsup/download/observations>

⁵Michael Friendly's Graphics Macros
<ftp://ftp.sas.com/techsup/download/stat/statgraf.sas>

⁶SAS Institute FTP Download site
<ftp://ftp.sas.com/techsup/download/>

⁷SAS-L Newsgroup
<newsgroup://comp.soft-sys.sas>

⁸Index of SAS-L archives
<gopher://jse.stat.ncsu.edu/11/othergroups/sas/>

⁹SAS File Contribution Server
<ftp://ftp.uga.edu/pub/sas/contrib/user.doc>

¹⁰NCSU Help Database
<http://www2.ncsu.edu/unity/lockers/public/consuli/CustomerQ/cqsearch.html>

¹¹NCSU Department of Statistics Home Page
<http://www2.ncsu.edu/ncsu/pams/stat/dept.html>

¹²DBMS/COPY, Conceptual Software, Inc.,
9660 Hillcroft #510, Houston, TX 77096.
copyright 1994-1995
(713)721-4200

AUTHORS

Sandy Donaghy
Senior Statistician
Department of Statistics, NCSU
Raleigh, NC 27695-8203
sandy_donaghy@ncsu.edu

Joy Munk Smith
Applications Analyst Programming Consultant
Department of Statistics, NCSU
Raleigh, NC 27695-8203
joy_smith@ncsu.edu

TRADEMARKS

SAS, SAS/ACCESS, SAS/ASSIST, SAS/FSP,
SAS/GRAPH, SAS/IML, SAS/INSIGHT,
SAS/STAT, SAS/TUTOR and "Observations"
are registered trademarks or trademarks of SAS
Institute, Inc. in the USA and other countries.
IBM is a registered trademark of International
Business Machines Corporation. ® indicates
USA registration.

Other brand and product names are registered
trademarks or trademarks of their respective
companies.