Summer 2004

A Reference Model for Providing Statistical Consulting Services in an Academic Library Setting

Kristi Anne Thompson
University of Windsor

Daniel M. Edelstein
University of Windsor

A Reference Model for Providing Statistical Consulting Services in an Academic Library Setting

by Kristi Thompson[1] and Daniel M. Edelstein

Introduction

Princeton University Library, through its Data and Statistical Services (DSS) unit, goes further than many libraries in providing consulting on statistical methods as well as software support for data library users. Princeton requires all third and fourth year undergraduates to do independent original research papers and theses in their disciplines of concentration. These requirements create a clientele of students who need to conduct relatively sophisticated statistical analysis, but who may or may not possess the necessary skills required to do so. This paper describes the model we employ at Princeton DSS to help these patrons, drawing on our experience as data consultants to discuss how it works in practice.[2]

A Call To Integrate Statistical Literacy Into Data Library Service

Data librarians go to great lengths to make data files available to their patrons and to help their patrons locate data sources. But helping people find the data file that best suits their needs is not enough. Providing access is more than making an item available. It means making a resource useful. Data files have a layer of complexity that can make them more challenging to use than other information sources. While most college students can extract and use the information contained in a book, many do not have the statistical or technical skills required to effectively extract and use the information contained in a data file. In other words, a high proportion of our patrons are not statistically literate. Giving a data file to a patron who does not possess the tools and skills needed to analyze it is about as useful as giving a book to someone who cannot read.

Much discussion of the problem of statistical illiteracy among students revolves around the need to properly integrate statistical literacy into the academic curriculum. Schield, for example, claims that "statistical educators should develop a college-level statistical literacy course for students in majors that do not require a math or statistics course." (Schield 2004)[3] This would be a welcome achievement. While we are waiting, the data library has a niche to fill in helping make it possible for students to use statistical analysis as a research methodology. And we believe that the library has a powerful model for dealing with the immediate needs of students who need to analyze data, that is, the reference model.

An introductory information literacy course, no matter how well taught, is not a substitute for the direct assistance of a qualified and experienced reference librarian. Teaching patrons about information is important, but the essential service of a librarian is to help meet the need for information at the point when it occurs. Similarly, we believe that while teaching the concepts of statistical literacy is critical, the primary purpose of the Library's consulting facility is to help our patrons use the data resources made available by the library. Teaching statistical theory is the role of the faculty. Our role is to complement them by helping our patrons overcome the practical knowledge barriers that arise when conducting a data analysis project. The DSS consultants provide assistance with the software and the statistical techniques necessary to make use of data files. Our service also makes it easier for faculty to integrate data analysis into their curricula. In her article "Understanding Barriers to the Use of Numeric Data in Learning and Teaching," Rice concluded, "Universities should develop IT strategies that include data services and support for staff and students, and integration of empirical datasets into learning technologies." (Rice 2001)[4] We support this conclusion, but believe that expanded library service and support for statistical theory is as important as IT service and integrated learning technology.

Putting It Into Practice

The service model we have evolved is in many ways similar to traditional library reference service, but we have adapted it to meet the unique challenges of statistical consulting. Our primary role is to help students use the data resources made available by the library. Much like in a traditional academic library reference transaction, we are trying to help our patrons find an answer to a particular research question. Our aim is not to teach statistics in itself, but to provide users with the practical knowledge needed to carry out their research. As a result, we have evolved a very practical, problem-oriented and intuitive approach. Teaching statistical theory is the professors' role. As one student remarked, the DSS consultants helped explain things "in a way in which even (his) statistically-

challenged mind could understand" (Data and Statistical Services Client Database, 2003).[5]

The Consulting Interview

Each consulting transaction begins with an informal reference interview in which we try to gently extract the information we need from our students before we can start working with them:

- What is the research question?
- What level of knowledge does the student have?
- Where are the data and what form are they in?
- What are the conventions and standards of the student's area of study?

What is the research question?

Finding out the actual research question – not just the question the student thinks that he needs help with – is important. Sometimes a student will come in and say that she wants to perform a particular type of analysis using some data set. Occasionally questioning about what she really wants to find out will reveal that the approach she wants to take is incorrect or not suitable for her data. Sometimes a student will have picked up a statistical term from classmates or will vaguely remember something from a lecture that he thinks is what he needs to do. Others come in with only a very vague idea of what they want to find out, or with a question that needs to be reformulated into something that can actually be answered with the data available. And occasionally, we have oddities such as the person who wanted to explain the gender of a judge as an outcome of what law school he or she had gone to. As one student noted, we help with "logic" as well as with "seeing the limitations of (her) study" (Data and Statistical Services Client Database, 2004).

What level of knowledge does the student have?

The level of knowledge of the student is sometimes readily apparent, particularly in those cases when the answer is "none." On other occasions we need to ask some gentle probing questions: are you familiar with this type of analysis? Have you worked with data files before? Before we can proceed, we need to know how closely the student's level of knowledge matches the level at which the analysis needs to be done. It is necessary to establish this at the start so that we do not inadvertently confuse or discourage students by giving them explanations that they do not have the background to understand.

Where is the data and what form is it in?

Usually, students come in with the data or knowing how to get it easily. Occasionally our questioning will reveal that a student is using the wrong dataset, or needs to merge in additional data. Sometimes students have come in wanting to perform a regression on a dataset that was only available as a set of summary statistics. Very frequently the data needs to be extracted, reshaped, combined with other data files, converted to another format, or recoded before anything else can be done.

What are the conventions and standards of the student's area of study?

The discipline that the student is working in will often affect what statistical advice we will give her. For example, we have found that biology students are often required to do nonparametric tests in situations where social scientists would not be. The area of study also can affect which results are reported and what language is used. In addition, the level that the analysis is done at, how rigorous or sophisticated it needs to be, can depend on the student's discipline, the year she is in, the scale of her project, and the standards of the department.

To demonstrate more clearly how the process works in practice, consider the following example of a typical encounter. A student comes in and asks the consultant to show him how to "get means in SPSS." Rather than immediately providing the answer, the consultant first asks the student about his dataset and research problem, and realizes that in fact he wants to do a t-test for the difference between means of some dependent variable grouped by some subgroup variable, such as gender. And given the type of research project the student is working on and the departmental standards for that type of work, the student needs to control for several other variables, so the consultant explains these issues and helps him to run and interpret multiple regression instead.

Different Approaches For Different Students

Our patrons exhibit a wide range of both technical and statistical ability and experience, and we need to take this into account when deciding how to approach each individual problem. We have informally grouped our students into three basic types to help us explain the range of different approaches we need to use.

The least sophisticated group consists of students with absolutely no knowledge of data or statistics whatsoever, who have somehow found themselves needing to conduct an analysis. Sometimes we encounter students who have collected or come across some data that they want to use, but do not know what to do with it. One example was a philosophy student we worked with last year who had conducted a survey of other students to get their reactions to various moral dilemmas. With students like this we often need to start by getting their data into a useable computer form. Then we work with them to find out what they want to know, and teach them in an intuitive way the statistical procedures they need to find that out. These students often have nowhere else to turn because statistics

simply is not taught in their department.

This group also includes students (and, occasionally, faculty members) who encountered a question in their research, or came across an enticing entry in the library catalog, and came to the library hoping to get a book or nicely formatted table they could read the information from. Instead, they were given a data file and told that the information was in there somewhere, if they could only decode it. We often start our discussion with these patrons by saying something like "A data file is sort of like an Excel spreadsheet…"

The second group contains students who have some background in the mathematical and theoretical basis of statistics, but have trouble applying it. University statistics courses are often taught in a theoretical and abstract way that leaves students under-equipped to deal with the practicalities of analyzing real-life data critically. Some may just need help with the software, while others may understand the mathematics of a regression but have no sense of what variables need to be included, or how to code categorical variables. And many seem to simply have trouble understanding how the theoretical concepts they have memorized can be used to make sense of actual data. An exchange that occurred while helping a student with his third year undergraduate research captures this well. Consultant: "Your explanatory variable is significant." Student: "Great!" Pause. "But what does it mean?"

The most capable group includes students with a solid level of statistical knowledge who are trying to do something ambitious that requires specialized programming skills or otherwise advanced knowledge. We also occasionally encounter students who need to find a valid statistical test that will let them find out something unusual, or need to learn the correct model to use in some peculiar circumstance. Answering this last type of question involves pure library research skills. We keep assorted statistical reference material on hand and have favorite trusted web sites to look at, and we also maintain contacts with graduate students and faculty who can help us with particularly difficult questions.

With all of these students our work often resembles data counseling as much as data consulting. We work with our patrons to explore the data and their research, and give them the opportunity to discuss what they are doing with someone knowledgeable who is not involved in evaluating them. We ask questions to help them clarify in their own minds what they are doing, and encourage them to explore possibilities. Our focus is on helping students to do the best work they are capable of themselves, and simply listening to them is often the most effective way to accomplish this.

We want to make sure that students understand what they are doing – not the details of the statistical theory behind it so much as an intuitive understanding of the concept behind the actions they are performing and the results they are getting. For example, if a student needs to do a probit analysis, we will not go into detail about the distribution and how it was derived, but we will give a couple of examples to explain why linear regression breaks down in the case of a binary dependent variable and may do a sketch of the probit curve to help her grasp why it works better.

Patron Management: Outreach And Intake

Almost all of our consulting is done on a walk-in basis in our computer lab during open consulting hours. We prefer not to make appointments, as we have found that scheduling an extended block of time to work with a single patron is generally an inefficient and unfair way of dividing our time. Working together in our computer lab, we can serve eight or more students at a time, answering questions as they arise while encouraging our patrons to work independently as much as they are able. Naturally, some students do require larger blocks of time, particularly when they are beginning a project. We hire graduate students as assistants and train them to deal with routine questions so that during these busy periods we can focus on the students who need more involved assistance.

Close to half of our students come from the department of Economics, with Politics, Sociology, Public Policy and Psychology supplying most of the remainder. However, over the last year or so we have also assisted patrons from History, Computer Science, Geology, Philosophy, English, Bioethics, Engineering, Religion and Evolutionary Biology.

Our outreach efforts include giving presentations to groups of majors in the social science departments, either by meeting with the students in groups or by attending sessions of departmental workshops, together with either a subject librarian or the data librarian. We also meet with new graduate students and encourage them to send us their students as well as use our services themselves. Many of our patrons are also referred to us by the data, economics, and other social science librarians. As word of our service spreads around campus, we also get many customers through word of mouth. Our web site has also become an increasingly important promotional tool.

Working With Faculty

Most of our consultations are with undergraduates doing independent research projects under the guidance of a faculty member or graduate student who advises them. We frequently need to decide what is appropriate to teach or advise students in the area of statistical methodology, given that their advisor is also supposed to offer help in this area, and certainly will be evaluating the choices that the student makes. When it comes to methodology questions, we will give a definitive answer if one exists. However, frequently the right answer to a statistical question is a matter of

opinion. When a question of this type arises, we may offer suggestions, but we also discuss possible alternative options, and will strongly encourage students to seek help from a professor. We generally try to get a sense of how much support a student is getting from other sources. This can influence how much direct methodological advice we will give. In cases where an advisor is working closely with the student we will defer to or, where necessary, reinforce the professor's advice. In cases where a student is getting less help, we will do our best to make up the lack.

In a few cases we have found ourselves dealing with students whose advisors were giving them advice that was unambiguously wrong. Often this is a result of miscommunication, and we are able to find the source of the misunderstanding and resolve the problem. In other cases the problem is an actual lack of knowledge, as in the case of the philosophy student whose professor did not understand the concept of statistical significance. Advisors who have found themselves dealing with an area outside their realm of knowledge are often relieved when they learn that the data consulting facility is available to assist their students. If handled with care and tact, incidents such as this can improve our relationships with the various departments with whom that we find ourselves working.

Conclusion

Once, we asked a student what the next stage of his research was and why he wanted to perform a particularly complex procedure. He said, "I don't know what the point of this is. I'll just finish this step and then my advisor will tell me what to do next." This attitude, and the teaching style which fosters it, is antithetical to our approach.

Our goal is to make it possible for non-statistically literate students to both conduct and understand analyses using sophisticated statistical techniques. We make it possible for our students to do their work themselves, and encourage learning by doing. We teach students enough to get them started, then have them dive into actual statistical analysis as quickly as possible. Most of them start to catch on quickly, and then proceed with their work with increasing confidence. From that point, we act as a resource they can consult at the inevitable bumps in the road. Frequently there are quick questions, and sometimes they need to pull over for more extended consultation. The flexibility of this approach allows our students to quickly take control of their projects, through acquiring and using the appropriate level of statistical knowledge. Often they are encouraged to learn more statistics and move on to more ambitious projects. But, even if not, they leave having accomplished something worthwhile.

Notes

[1] Contact: Kristi A. Thompson, Princeton University Library. Phone: +1 609-258-6053 http://dss.princeton.edu/ Email: kristit@princeton.edu.

[2] Parts of this paper were presented at the IASSIST conference, 2004, Madison, Wisconsin by Daniel M. Edelstein and Kristi Thompson.

[3] Schield, Milo (2004), "Statistical Literacy Curriculum Design." 2004 IASE Roundtable, Lund Sweden. Available at: http://www.augsburg.edu/ppages/~schield/MiloPapers/2004SchieldIASE.pdf

[4] Rice, Robin (2001), "Understanding Barriers to the Use of Numeric Data in Learning and Teaching." *IASSIST Quarterly*. Vol. 25, No. 1. Pg. 5-9.

[5] Data and Statistical Services Client Database (2001-2004) A database used to track usage of the Data and Statistical Services computer lab, in which computer lab users enter details of their lab usage, including optional comments, Princeton University.