SUCCESSFUL STATISTICAL CONSULTING: SOME ISSUES IN THE EDUCATION OF STATISTICIANS AND PROFESSIONAL PRACTITIONERS

Douglas C. Montgomery, Georgia Institute of Technology

ABSTRACT

The effective use of statistics is an important part of the problem-solving process in industrial and engineering research. This paper identifies those aspects of the training and education of statistical consultants, that contribute to their success in the industrial environment. A number of issues regarding the professional practice of statistics are also discussed.

1. INTRODUCTION

The purpose of this paper is to discuss several aspects of the training of statisticians, the role of the statistician as consultant, and some principles of professional practice that lead to successful statistical consulting. The focus is on the industrial, academic, and the government systems environments. I exclude any discussion of the life or social sciences because of limited experience in those fields. However, I believe that the principles that apply in the industrial and engineering environment are highly portable. We will discuss the type of academic background and training that is likely to produce a successful statistical consultant, and present some suggested guidelines for how the statistician can most effectively operate in the consultant's role.

Many opportunities for statistical consulting exist. I believe that these opportunities are expanding, in part because of the recent emphasis on quality and reliability in manufacturing, and because of the increasing general awareness of the usefulness of statistics in business decision-making. There is a significant shortage of individuals with appropriate training and experience to be effective internal or external consultants in this field.

An effective consulting statistician is a scientist who takes a leadership role in the identification of problems, drawing management's attention to these problems, convincing them of the need for their solution, developing appropriate alternative approaches to the solution of these problems, analyzing the available alternatives, and identifying and implementing the solution. Consequently, consulting statisticians have interests that range across the spectrum of business, science, engineering, problem solving, statistics, and computing. The strongest personal characteristic of such an individual is an interest on working on practical problems. The statistical consultant must also have the ability to see the big picture, because of the great diversity of problems and environments that he will encounter. Finally, a consulting statistician must have good oral and written communication skills, because of the emphasis on defining problems, and communicating with both technical and nontechnical personnel concerning the statistical methods used in the solution.

2. ISSUES IN TRAINING AND EDUCATION

Consulting statisticians are usually generalists. They are often presented with problems from many different subject-matter fields and employ a number of different statistical techniques for analyzing them. Consequently, during their professional practice they become exposed to a wide range of statistical tools. However, it is not unusual for them to be highly expert in one or two areas, such as regression analysis and the design of experiments. During their formal academic training, it is impossible for statisticians to be exposed to all of the statistical techniques that they will find useful during their professional careers. However, it is possible to identify a number of techniques which are frequently used by consulting statisticians, and which should form the core of their formal academic training programs. These techniques are as follows:

- Regression analysis
- Basic statistical analysis of data
- Graphical analysis and presentation of data
- Design of experiments
- Analysis of variance
- Process control techniques
- Acceptance sampling methods
- Time series analysis
- Categorical data analysis

These techniques find extensive application in the industrial, engineering, and business environment, and are listed in approximate order of frequency of use. Naturally, this order reflects my own professional experience, and I am sure that other consulting statisticians would not agree entirely with the items on my list.

In addition to the core subjects listed above, there are a number of other techniques that are used on a less frequent basis, but which should be in the consulting statistician's repertoire:

- Nonparametric statistical methods
- Simulation and Monte Carlo techniques
- Nonlinear estimation
- Survey sampling and survey design
- Reliability and life testing
- Probability/stochastic modeling

The area of application will greatly influence the entries on this list and the order of importance. For example, in the electronics industry reliability and life testing are used very extensively. Furthermore, many of the subjects on this list are not static; for instance, time
series analysis will be used much more extensively in the future. The increasing availability of good computer software and the growing general awareness of the powerful nature of this technique will greatly expand its use in the next few years.

In addition to the statistics subjects mentioned above, it is also essential that the statistical consultant be familiar with at least one compiler level programming language and have some experience with most of the popular statistical packages, such as SAS, BMDP, SPSS, MINITAB, and OMNITAB. The computational aspect of statistics is extremely important, and the consulting statistician must have an adequate appreciation of this aspect of his field. It is quite likely that he will have to do his own computer programming, at least occasionally, and either utilize commercial software or work with a professional programming staff in most applications.

Academic programs in statistics are most effective at the M.S. and Ph.D. levels. Statistics is a field best studied by an individual who is building a strong base in engineering, the physical sciences, or mathematics. Engineering is an ideal preparatory experience for the consulting statistician. It stresses problem solving, modeling and the use of analytical methods in problem solution. Most students in graduate level statistics programs come from a mathematics background. While these students are much more skilled in the theoretical aspects of statistics than in engineering and physical science backgrounds, they often lack the problem formulation and modeling skills that are the strength of engineers.

The consulting statistician's formal academic training should contain a balance between theory and methods. The theory courses are necessary because of the many non-standard situations that the consultant encounters. In many cases, he will have to modify an existing procedure, develop new methodology, or understand the assumptions and limitations of proposed methods in order to solve the problem at hand. This level of sophistication can only be achieved through a firm understanding of statistical theory. The consulting statistician should view his training in mathematical statistics as a means to an end.

It is also essential that statistical consultants receive experience in working directly with clients and developing effective problem identification and formulation skills during their formal academic training. This hands-on experience with real problems is an essential ingredient of the consultant statistician's training program. Some of this experience will be obtained as a result of courses in design of experiments, regression analysis, and time series analysis. Thesis research which is directed toward problem solving and which involves data analysis and modeling can also be an effective learning tool. However, a statistical consulting laboratory or intern program is the most effective mechanism for imparting this type of training to the statistical consultant.

Consulting laboratories and intern programs must be supervised by experienced faculty. These faculty must have an appreciation for professional practice, and an understanding of the environment in which the statistical consultant will operate. It is essential that they have reasonable consulting backgrounds themselves because the statistical consultant-in-training will learn much of the practical aspects of his craft by observation. Unfortunately, there are not many faculty who are comfortable in this role. In addition to the consulting laboratory or intern program is the scientific literature. Some of this experience will be obtained as a result of courses dealing with "analysis and interpretation of data", which are based totally on real problems and data sets from the scientific literature. Few would be a second best alternative approach.

In addition to the skills that are directly relevant to statistical analysis, the consultant also needs to develop effective oral and written communication skills. The inability to communicate effectively, particularly with nontechnically trained individuals, is a major deficiency of many potentially suburban statistical consultants. Communication skills can be developed through consulting and intern programs, course projects, and teaching.

Other business skills, such as knowledge of finance, the ability to delegate work, and public relations skills, are also extremely useful. While much of this background can be acquired after the consulting statistician's formal academic training is complete, it is strongly advised that any student with an inclination toward professional practice begin acquiring these skills during his formal educational process.

3. SOME COMMENTS ON PROFESSIONAL PRACTICE

I now turn to a discussion of several factors that impact statistical consulting in the general industrial and business environment. First, it is important to remember that the consulting statistician is never working on his own problem, but rather on other individuals' problems. Thus the role of the consultant is to assist other professionals in doing their jobs more effectively. Furthermore, because statistical techniques are useful in more fields of application than any other technical discipline, the breadth of the statistician's background is necessarily too shallow for him to be an expert in the subject matter field in which he is consultant. Sometimes this creates enormous difficulties, in that the client views the statistician as incapable of making a substantive contribution to this field. On the other hand, the consultant statistician is well trained in methodology which is highly portable. He views problems as linear models, say, and whether the linear model is imbedded in a background of electronics or pharmacology is immaterial.

The consulting statistician should make every effort to become totally involved in the
problem situation. This implies that he must meet with the clients regularly and, if necessary, assume the initiative to meet and communicate with all principals involved in the consulting situation. A consulting statistician must also be as much of the subject matter field and the terminology of that field as possible. This has the twin benefit of helping the consultant to understand the real problem and to be able to communicate to the other participants in the study.

In studies where data is to be collected, it is essential that the consultant visit the data collection site. He should observe the preparation and operation of the equipment, and the procedures being used to actually collect the data. He should discuss the data collection process with the operators, as well as with the professional staff, and read any reports, journals, or books in the subject matter field which will help him to understand the data collection process. I have often said that data is the raw material of statistics, and that more research and development projects are spoiled by faulty, erroneous, and inaccurate data collection than by any other cause. Consequently, whenever it is at all possible to do so, the consulting statistician should personally monitor and supervise the data collection process to the closest extent practical. It is much easier to resolve data problems at their source than months later after the experimental apparatus has been disassembled or the data collection budget expended.

Successful statistical consulting also requires a good problem. There are several conditions that define a good problem, and the experienced statistical consultant should learn to identify these. A good problem is one that is important to the client. It is not effective to work on minor problems, even as pilot studies to test new methodology. A major contribution to an unimportant problem is still not much of a contribution. Another characteristic of a good problem is that it can be clearly formulated, and the statistical versus nonstatistical aspects of the problem clearly identified. The statistical consultant has the responsibility of making known to the client the areas of the problem in which his contribution is likely to be made, and his limitations in the subject matter field. I have often found that obtaining a clear statement of the problem, and an agreement of the part of the client as to exactly what needs to be done to solve it was the major factor in obtaining a satisfactory solution. Far too many projects have ill-defined and fuzzy objectives, and many clients, while being well versed in their own subject matter fields, have only a limited understanding of experimentation and statistical analysis.

An issue related to problem identification is avoiding type III errors. Most statisticians are familiar with type I and type II errors, and are well trained in methodology to control the risks of these errors, but are often unaware of the third type of error. As defined by Kimball (1957), a type III error is the error committed by successfully solving the wrong problem. I suppose that one could also define a type IV error as obtaining the wrong solution to the wrong problem, but let us assume technical competence on the part of the statistician.

I have no way of knowing the incidence rate of errors of the third kind, but I feel that all of us, particularly in our early years of practice, have committed them. I can think of many occasions where I have helped clients make t-tests, build regression models, or perform analyses of variance, thinking that I was giving the right answer to the right question. I hope that at least in a majority of these cases I did give the right answer to the question that was asked. Unfortunately, I have come to suspect that in many cases the question that is asked is only peripherally related to the real problem. In this way a great many type III errors are committed. Sometimes the error can be blamed on the client, who doesn't believe that the statistician has adequate knowledge in the subject matter field and that any attempt at a complete explanation would be a waste of his valuable time. In other instances, the consulting statistician is at fault for not becoming sufficiently familiar with the problem to enable him to give intelligent advice, or he is too quick to apply a standard solution to what at first appearances is a familiar problem.

I think that type III errors are sufficiently important that every potential consulting statistician needs to be exposed to them, preferably in his academic training. To some extent this is accomplished by statistical consulting laboratories and intern programs that form a part of many M.S. and Ph.D. programs in statistics. Unfortunately, however, many faculty seem to be unaware of the existence of type III errors, and because they form a part of the experience base and cannot be handled analytically, they feel that they have no place in the formal academic training program. However, I feel that this is a serious issue, and one that the consulting statistician must continually guard against.

Finally, a key ingredient to successful statistical consulting is the consultant's ability to stay abreast of developments in the field. This is extremely difficult to do, for some areas of statistics are moving very rapidly such as regression and linear models, and time series analysis. Nevertheless, what would constitute effective and appropriate statistical practice in regression five years ago would hardly be adequate today. Continuing education programs can be helpful in this regard, if carefully selected.

In summary, I believe that successful statistical consulting is a combination of the consultant's background, the training he receives in his formal education, and his ability to identify, formulate, and solve problems effectively in the industrial environment. Some of the skills required can be taught, while others cannot. As in any business, the consultant's success will depend on his possession of the correct...
mix of these attributes, and his ability to combine them in a complex decision-making process to produce satisfactory results for his clients.

REFERENCES


Contact Author

Dr. Douglas C. Montgomery
School of Industrial and Systems Engineering
Georgia Institute of Technology
Atlanta, Georgia 30332
Phone: (414)894-2306