SOME DIFFERENT MODELS FOR INTERACTING WITH RESEARCHERS AND STUDENTS IN OTHER DISCIPLINES

Bryan F.J. Manly
Western EcoSystems Technology Inc., United States of America
bmanly@west-inc.com

In this talk I discuss my experiences in terms of giving classes and workshops on statistics to researchers and students in other disciplines, working with individual researchers, writing books to introduce statistical ideas in a particular topic area to researchers and scientists, and situations where the statistician is part of a large group that is investigating some problem, and the group includes many scientists who are very knowledgeable about standard statistical methods.

INTRODUCTION

Statisticians interact with researchers and students in other disciplines in different ways. Perhaps the most common way is when the statistician teaches an introductory or advanced course on statistical methods to a class of students who are mainly not specializing in statistics. However, there are many possible alternatives to this, including the statistician collaborating with an expert in another subject area to collect and analyse a particular set of data, or to develop a new method of analysis for an unusual type of data that occurs in the subject area.

This talk is about my experiences when interacting with people from disciplines other than statistics. It is based on my time as a teacher and researcher at universities in several different countries, the presentation of many short courses and workshops at various locations around the world, writing books aimed mainly at audiences of non-statisticians, my work since 2000 as a statistical consultant for a private company in the United States and, more recently, running internet courses on environmental and ecological statistics for scientists working in this area.

WORKSHOPS AND SHORT COURSES

I taught statistics at all university levels from first year to postgraduate at the University of Salford in the UK (1967-70), the University of Papua New Guinea (1970-73), the University of Otago in New Zealand (1973-2000), and the University of Sao Paulo in Brazil (2004-05). I started giving workshops and short courses on statistics for students and researchers in other subject areas in 1988, where I think of a short course as a number of lectures on a particular topic and a workshop consisting of lectures plus practical work by the participants.

Altogether I have presented about 50 workshops in various places around the world. The last one was in 2009 on Environmental Sampling and Data Analysis and this was for workers in several government departments in the Cook Islands in the Pacific Ocean.

I see great value in workshops and short courses providing that they are presented at an appropriate level, with lectures by the instructor being combined with practical work by the participants. For example a good plan seems to be for each day to consist of two or three presentations by the instructor, followed by at least the same amount of time for the participants to work on exercises related to the presentations. Also, when the participants are working scientists it can be very useful and interesting to have some sessions where these scientists talk about specific problems that they have with the collection and analysis of data. This can even lead to future collaborative research.

One reason why workshops and short courses can be very valuable is that it often seems to be the case that scientists and researchers have to collect and analyse data as part of their work but do not have any way to get help from a statistician when the data are unusual and they are not sure how to deal with it.

Since 2004 I have also been running statistics courses over the internet for a company called Statistics.com. Again these courses are aimed mainly at scientists working in other disciplines that need to use statistical methods either often or occasionally. The first internet course that I taught was on Environmental Statistics. This has been repeated every year since 2004. Then in 2007 I started another internet course on Environmental and Ecological Sampling, that has also been repeated every year since then.

In C. Reading (Ed.), *Data and context in statistics education: Towards an evidence-based society. Proceedings of the Eighth International Conference on Teaching Statistics (ICOTS8, July, 2010), Ljubljana, Slovenia*. Voorburg, The Netherlands: International Statistical Institute. www.stat.auckland.ac.nz/~iase/publications.php

Internet courses with Statistics.com involve the participants reading specified parts of a book or course notes each week and then working on some exercises, which are marked and returned with comments. This continues for four weeks. Credit certificates are available for the successful completion of individual courses with the American Council of Education (an organization involving more than 1800 universities and colleges in the USA), and there is a Program in Advanced Statistical Study that requires the successful completion of eight courses within five years.

These internet courses are good because they can be taken anywhere in the world providing an internet connection is available. However, they are not as good as a workshop covering the same material over a period of several days where everyone is together because the interaction between the instructor and the other participants is then obviously better than it is with an internet course. It would be good if an internet course could include some video conferencing, but with participants all over the world finding a good time is not possible.

WRITING BOOKS

My books are mainly aimed at students who are not majoring in statistics, and scientists who have to use statistics but are not statisticians. These books mostly arose either from university courses that I have taught or as a result of collaborative research with biologists.

In 1968 I was an Assistant Lecturer in Mathematics at the University of Salford in the UK. By chance I was the only person in the mathematics coffee room when Mike Parr, a biologist, came in to ask for help with analysing some data from a mark-recapture study on dragonflies. I had never met a biologist before, and I had no idea what a mark-recapture study involved. However, that meeting changed my career forever.

I did a lot of work with Mike Parr in those early days, and he introduced me to many of his biologist friends. This included Lawrence Cook at the University of Manchester who had a general interest on studies about natural selection of habitat and food by animals. One day he asked me to review the chapter on statistics in his latest book on natural selection. It was quite short and I asked him whether that was all that there was on this subject. He said that he thought this was the case, and I then decided that I would write a book on statistical aspects of the study of natural selection, because this would obviously be useful for the biologists working in this area.

It turned out that there was a lot more statistics in the study of natural selection than I had imagined when I talked to Lawrence Cook in about 1970. The 500 page book took a long time to write, but The Statistics of Natural Selection on Animal Populations was eventually published by Chapman and Hall in 1985.

My second book arose from a course on multivariate analysis that I set up and taught at the University of Otago. At the time this course was unusual because it was a second year course aimed at students not majoring in statistics. There was even opposition from some people in my department to the idea of teaching anything more than introductory statistics to students from other departments. However, the course went ahead because of the overwhelming support coming from the other departments. My idea with the course was to teach the methods of multivariate analysis with the minimum of mathematics, but making sure that the assumptions of different methods were properly understood.

I had printed notes for the course. One day a biologist that I did not know came to my office to ask me a question about multivariate analysis. He had some papers with him and I asked him where they came from. He said that he did not know, but they were the only explanations about multivariate methods that he had ever been able to understand. I saw that what he had was a copy of my lecture notes. At that point I decided that I needed to turn these notes into a book both to be used as a textbook for introductory courses on multivariate analysis, and for scientists to read on their own. I wanted it to be the easiest book to read that is available on the topic.

Multivariate Statistical Methods: a Primer appeared in 1986. The second edition appeared in 1994, and the third edition in 2005. With each revision reviewers have suggested adding more material, and tried to persuade me to use a particular statistical package for the examples. I have resisted that because I want this to stay being a book that is just an introduction to the most important topics, that is not very long, and that does not require the use of one particular statistical package. I have used the book as the basis for many workshops on multivariate analysis.

I have written two other books based on courses that I taught at the University of Otago. The Design and Analysis of Research Studies (Cambridge University Press, 1992) is based on my lecture notes for a course with the same name. University departments other than mathematics and statistics sometimes have courses with names like this, and this book reflects what I think those courses should contain. There are chapters on sampling designs and standard methods of analysis like regression and analysis of variance, but also discussions about more fundamental ideas like the difference between the strength of conclusions based on experimental and observational studies, the differences between true experiments and quasi-experiments, and the ethical considerations involved in some study designs.

My third book based on courses at the University of Otago is Statistics for Environmental Science and Analysis (Chapman and Hall/CRC, 1st edition 2000, 2nd edition 2009). This is based on a course that was part of a masters degree on environmental science. For this course it was necessary to assume that the students had only taken a first year statistics course several years before. Therefore some revision of basic statistical ideas was included in the course, and appears as an Appendix in the book. The book is the text for my internet course on Environmental Statistics, and I have used it as the basis for many workshops and short courses.

Three of my other books did not arise from university courses that I have given. Stage-Structured Populations: Sampling, Analysis and Simulation (Chapman and Hall, 1990) was the result of my research interest in the analysis of data on animal populations where the individuals pass through discrete recognizable development stages and there is interest in the survival rates through the different stages and the total population numbers reaching each stage. I used the material in this book for a number of workshops in the late 1980s and early 1990s.

I got interested in randomization tests and other computer-intensive methods while I was working on my book on the statistics of natural selection. These methods were just becoming feasible for biologists in the late 1980s, and I realized that a book about these methods for biologists would be extremely useful. So I wrote Randomization, Bootstrap and Monte Carlo Methods in Biology (Chapman and Hall, 1st edition 1991, 2nd edition 1997, 3rd edition 2006), and produced a computer package to do the calculations. I think that I have given more short courses and workshops based on this book than on any other subject.

Finally, there is the book Resource Selection by Animals: Statistical Design and Analysis for Field Studies (Chapman & Hall 1st Edition, 1993, 2nd Edition, 2002) written with Lyman McDonald and others. In a way this was a development of the work in the book on natural selection by animals, but the focus is a little different. It is about the food and habitat used by animals rather than the animals that survive some selection process. Resource selection studies seem to be becoming more important with time and Lyman and I and others have run numerous workshops based on the ideas in this book.

WORKING WITH INDIVIDUAL RESEARCHERS

From talking with researchers in other subjects it seems that they often have trouble communicating with statisticians. I am not sure why exactly this is, but sometimes it seems to be because many statisticians are talking a different language and the researchers just do not understand what they are saying.

Unfortunately, I think that part of the problem is that statisticians sometimes tend to make their analyses so complicated that no one else understands what they are doing. For example, I remember when a researcher told me that they had asked a statistician for help on analysing some data related to penguin foraging times in the Antarctic. The analysis was done, and the researcher said that the complicated unbalanced analysis of variance was too difficult to understand. However, the basic question was quite simple - whether the mean foraging times were the same for males and females. Simple randomization tests that everyone could understand answered that question easily (Clarke et al., 1998).

Murtaugh (2007) argues that there is "... an abundance of needlessly complicated and confusing statistical methodology in modern ecological literature", and that when there is a choice simple analyses are better because it is "easier to follow the train of logic and verify that there are no weaknesses obscured by elaborate statistical manipulations". Murtaugh also suggests that some researchers also want a complicated analysis for their data because this might make it easier to get

their work published. However, this has not been my experience. The researchers that I have worked with have wanted an analysis that is easy to understand, so that they can explain it to other people themselves, if necessary.

WORKING WITH GROUPS OF RESEARCHERS

My statistical consulting has usually involved working with one or two researchers. However, I have in the last few years been involved with a group of scientists in the USA working for the federal government, the California state government, and water associations that are responsible for delivering water to cities and farms in California. The main concern has been the endangered delta smelt fish in the Sacramento-San Joaquin Delta in California, and ways to reverse the decline in numbers since about 2000. My work in this area in the last few years has been funded by two of the water associations, but has often involved working with scientists from the state government.

The situation with the delta smelt is very contentious. There are disagreements between the different groups involved, particularly about the adverse effects on delta smelt of pumping water from the Delta for cities and agriculture. These effects are not clear and many scientists have been trying to quantify them for years without any definite conclusions emerging.

Many of the scientists involved with the delta smelt problem are used to using standard statistical analysis methods, particularly regression, and do many analyses themselves. However, I have been asked to do some of the more complicated analyses like simulation studies to assess the power to detect the effect of some modification to the water flows in the delta on the likely location of the delta smelt. My role as a statistician in this situation has therefore been partly to do particular analyses suggested by the other scientists. However, often my role has been to review the results of analyses done by other scientists. In this respect it has been very important to act as an unbiased expert, even though my work has been funded by one of the several groups involved with the delta smelt problem.

In this type of situation I believe that it is very important for statisticians not to take sides in the sense of favouring analyses that suggest particular effects. It is very difficult for the scientists involved to be completely unbiased because it suits the organization that they work for the effects to be in a particular direction. However, there does seem to be a general understanding that the role of a statistician is just to report what the data shows in an unbiased manner.

CONCLUSIONS

The short courses and workshops that I have presented and my consulting work with researchers who are not statisticians has shown me that there is a great need for statisticians who are prepared to take the time to find out what the data analysis problems are in a particular subject area. When I first began working with biologists as a young man I was really surprised about how easy it was to find a statistical problem that needed solving with biological data, and how easy it was to get papers published in good journals by working in this area.

One example that I remember well is that I found out in 1969 that although there were several methods available for analysing mark-recapture data no one had ever compared these methods in a simulation study. So I did that study and wrote a 27 page paper that the Journal of Applied Ecology was very happy to publish. Before I joined up with the biologists I was trying to do research on sequential tests, because I got interested in this area during my undergraduate degree. I had written several papers in this area, of which some were accepted for publication and some were not. I knew that these papers were about things that would never be of much use to anyone. So I was very happy to forget all about sequential tests and work with biologists on problems that I knew were important and where the analyses that I proposed would actually be used in the future.

One big change that I have seen with statistical consulting over the last 40 years is that when I started consulting it was necessary to meet people frequently in order to work on a joint research problem. Now this is hardly ever necessary. With an internet connection a statistician can be anywhere in the world and still be working with a scientist somewhere else in the world on the design of a study or the analysis of some data. To me this is one of the things that makes statistics such a wonderful profession.

REFERENCES

Clarke, J., Manly, B. F. J., Kerry, K., Gardner, H., Franchi, E., Corsolini, S, & Focardi, S. (1998). Sex differences in Adélie penguin foraging durations. *Polar Biology, 20*, 248-58.

Murtaugh, P. A. (2007). Simplicity and complexity in ecological data analysis. *Ecology, 88*, 56-62.