GIVING AN INTERNET COURSE ON STATISTICS TO ENVIRONMENTAL SCIENTISTS IN EIGHT COUNTRIES

Bryan F. J. Manly
University of Sao Paulo, Brazil; Western EcoSystems Technology, United States
bryanmanly@lycos.com

In this paper I describe my experiences teaching a four week internet course on environmental statistics over the internet in 2004. The course was run by an organization called Statistics.com, which provides many statistics courses this way on a regular basis. The course was based on my book on environmental statistics, which the participants had to have. I discuss the preparation of the materials, the workload for participants, coursework, and the assessment process. It seems inevitable that online courses will be more common in the future, provided by private companies, universities, and other organizations. There were some minor problems with providing the course described here, but the format is basically sound for situations where a group of professionals want to learn about a particular topic, but do not have the time to take a regular course.

INTRODUCTION

Statistics.com has been in the business of providing internet courses on statistical topics for several years. They approached me in 2004 with the idea of including a course on environmental statistics in their offerings because of the book that I wrote on that topic some years ago (Manly, 2001). They also knew that I regularly give short courses on this and other statistical topics. I agreed to give the course because I thought that it would not require too much preparation and I was interested to try the experience of teaching statistics in this different way.

PREPARATION OF COURSE MATERIALS

There is apparently a standard format for all internet courses provided by Statistics.com. Each course is four weeks long, with one lesson each week. There is an outline of the course which in one or two pages describes the overall aims of the course, and the topics covered in each of the four lessons.

My description of the overall aims was as follows:

> This course is intended to introduce environmental scientists and managers to the statistical methods that will be useful for them in their work. Many of these methods would be covered in a standard course on statistics, but some of the topics that are covered here would not be included in such a course. For example, the statistics of environmental impact assessment would be most unlikely to be mentioned in a standard statistics course. The course is based on the book *Statistics for Environmental Science and Management* by Bryan Manly (published in 2001 by Chapman and Hall/CRC, Boca Raton, Florida). The mathematical background required for the course has been kept to the minimum level possible. Only an understanding of basic algebra and arithmetic is needed. Some initial knowledge of statistical methods is assumed, such as would be obtained from one introductory course covering the ideas of random variation, statistical distributions, tests of significance, and confidence intervals. However, if the course was taken some time ago then there is a supplementary refresher component with the present course.

Clearly this course was not aimed at statistics specialists. Rather, like the course text, it was aimed mainly at scientists and managers who need to use statistical methods in their work, but have a limited background is mathematics and statistics.

I wrote the course text because my experience has been that statistics is a very important component of the discussions and arguments that surround many controversial environmental issues, and sometimes when this is not the case, then it should be. Also, the type of statistics involved is not usually covered in traditional applied statistics courses. My aim with the course

was to cover the areas that I thought scientists and managers should be aware of, but also raise what I think are important aspects of how statistics should be used.

The four lessons were:

(1) *Background and Sampling*, covering examples illustrating the role of statistics in environmental science, followed by the discussion of environmental sampling strategies (random sampling, stratified sampling, etc.) and the analysis of the data obtained.

(2) *Environmental Data Analysis*, covering models for data (standard statistical distributions, linear regression, analysis of variance, and generalized linear models), and statistical inference under different conditions (observational and experimental studies, quasi-experiments, design-based and model-based inference, tests of significance and confidence intervals, computer-intensive methods, avoiding pseudoreplication, multiple testing procedures, meta analysis, and Bayesian methods).

(3) *Monitoring and Impact Assessment*, covering the use of control charts and CUSUM charts with monitoring data, testing for a change in distribution using before and after samples, designs and analyses for impact assessment (BACI designs, impact-control designs, before-after designs, impact gradient designs, possible inferences, assessing site reclamation, the problems with usual tests of significance, and testing for bioequivalence), and time series analysis (components of time series, serial correlation, test for randomness, detection of change points and trend, and more complicated time series models).

(4) *Spatial and Censored Data, and Monte Carlo Risk Assessment*, covering types of spatial data, analysis of count data, randomness in where points are located, the Mantel randomization test for spatial correlation, the geostatistical methods (variograms and kriging), correlation between variables in space, censored data (estimation from a single sample, estimation of quantiles, comparing the means of two or more samples, and regression), and the Monte Carlo approach to probabilistic risk assessment.

For each lesson the participants were given a brief description of why the material is important, and then sections of the course text to read. Because of the large number of topics covered some of these were classified as supplementary, so that the participants only needed to cover these if they had time, or could look at them after the course was ended if necessary. Each lesson also had a set of exercises for the student to complete and submit by the start of the following week.

THE PARTICIPANTS

When the list of 22 participants arrived it was clear that they were nearly all professionals in the environmental area, with a few academics who wanted to know more about statistical methods in this area. Most (13) were from the USA, followed by Canada (3), Australia (2), and one each from the UK, Colombia, Switzerland, and South Africa. It transpired later that the level of statistical knowledge varied greatly from very little to extensive as far as standard statistical methods is concerned, and also in terms of the ability to do statistical calculations on a computer. This may be quite usual with this type of course. One of my PhD students from a few years ago enrolled for the R course at statistics.com. She said that by about the third week it seemed like most of the students had dropped out.

RUNNING OF THE COURSE

At the start of each of the four weeks the participants could find the week's lesson material on the internet site. They were then expected to read the specified parts of the course text. When they had questions there was a Discussion Board on the internet site where they could put the question for everyone to see. When the course started I wrote some information about myself on the Discussion Board and asked the participants to do the same, which they mostly did.

The different time zones caused some minor problems with questions because I had gone to bed when some of the participants wanted answers to questions. When I woke up and checked

the Discussion Board it was interesting to see that some participants (one or two in particular) had decided to answer the questions of other people. In some case the answers provided were somewhat wrong or misleading. Also, if I was too busy to look at the Discussion Board for a whole day, some participants seemed to get rather annoyed.

At the end of each week the participants were supposed to upload their coursework to the internet site at a location that the other participants could not access. This is usually looked at by a tutorial assistant provided by Statistics.com, but at the last moment this person was not available, so I was asked to organize this process. This caused some problems because the post-doc students that I asked to help gave marks, which they were not supposed to do, and we were not told that after a piece of coursework was commented on there was a process for making it available to the participant concerned. So for a while there was confusion because we said that the work was marked but the participants said that it had disappeared. Then some of the participants started questioning the marks. It required an e-mail to everyone from the person in charge at Statistics.com to calm everyone down.

Another problem was that some of the students wanted to include graphics and other fancy things in their coursework but the internet site had problems accepting this. Apparently, therefore, there are still some technology problems to be overcome at the website.

By the third week differences in the background of the participants were becoming obvious. Some of them were using a statistical package that they were familiar with to do the coursework. Others had decided to get a statistical package but did not have time to learn how to use it and do the course as well. Others were struggling just to do fairly simple things in Excel. For this reason I advised some of the students to leave parts of the coursework until the course was over, and then to look back at them if they wanted to be able to use the statistical method concerned. For example, doing a logistic regression for the first time was a major problem for some participants.

Many of the problems encountered were just due to my ignorance about how the system works at the web site and the standard procedure operated by Statistics.com, rather than the approach to teaching. In addition, I decided that in future I would plan the course with more allowance for the varying backgrounds of participants.

ASSESSMENT

When I started the internet course I thought that all students would be assessed based on their coursework handed in. This was not the case. As noted above, the coursework was not supposed to be marked, only commented on. Statistics.com then provides participants with a certificate showing that they have completed the course if they attempt enough of the coursework. It seems that this means that they have to be active still in the fourth week of a course.

This assessment is satisfactory for a working professional who is just doing a course for their own benefit. It is not satisfactory for a regular student who needs credit for a course towards a university degree, or a professional who wants a qualification that will be taken seriously. I have discussed this issue with representatives of one of the large publishing companies. I understand that they are considering moving into the business of producing internet courses in conjunction with one or more universities, as a new method of promoting their books. The university would then set the methods and standards for assessment, and award credit to those who reach the required standard. Providing this was done with a university with an international reputation that it is concerned to maintain, this would provide a qualification with real status.

FINAL COMMENTS

My internet course had more technical problems than most because I was not aware of some aspects of the system, and did not even realize that fact. Usually this is not so important because the tutorial assistant handles the coursework and is available for online assistance.

Judging the amount of material to be covered each week is not simple. Initially it seemed to me that I was supposed to cover my whole book, but in fact in the end I made many topics supplementary reading, because this seemed to be necessary. At any event, no one complained about there being too much material, although as noted above, I did advise some students not to attempt all of the coursework questions.

Perhaps the most difficult problem with a course like this is accommodating the varying levels of statistical and computing ability in the participants. I stated that an introductory course in statistics was a reasonable prerequisite for my course. However, anyone who has taught first year university statistics will know that students may pass the course but still not understand some basic concepts very well. Then after a few years most of the knowledge may have disappeared for someone who does not use statistical methods regularly.

Recently someone that I work with said that she was thinking of taking an internet course on Bayesian statistics. I was a little surprised because I did not think that she knew much statistics at all. She asked me if I thought the course would be suitable. I suspected not, but then it occurred to me that the obvious thing for her to do was to get a copy of the course text and see if she could understand it easily. So that is what I suggested that she did. That seems like a pretty good screening device for internet courses.

REFERENCES

Manly, B. F. J. (2001). *Statistics for Environmental Science and Management*. Boca Raton: Chapman and Hall/CRC.