

EXPERIENCES IN CONSULTING WITH RUSSIAN BIOLOGISTS ON APPLICATIONS OF STATISTICAL METHODS

Lyman L. McDonald

Western EcoSystems Technology, Inc., United States
lmcdonald@west-inc.com

I report on my experience in consulting with environmental biologists on applications of statistical methods while on a tour of State Nature Reserves and National Parks in Russia, 2003. The biologists that I worked with were well trained in mathematics in comparison to biologists trained in the United States. I also found some biologists with advanced graduate degrees to be well trained in analysis of time series data. However, training in many basic applied statistical methods was either absent, poor, or incorrect. This lack of knowledge of applied statistical methods greatly limits the ability of biologists to design observational studies or manipulate experiments and publish results of their work outside Russia. If there are individuals actively teaching applied statistics in Russia among the audience or among the readers of the proceedings of this conference, I apologize if I have offended anyone. The gap in knowledge of applied statistics is real, however, and I would appreciate communicating with you.

INTRODUCTION

I am President and Senior Biometrician in a private company that specializes in providing statistical consulting services to biologists conducting observational studies (e.g., Cook and Campbell, 1979) and manipulative experiments (e.g., Gotelli and Ellison, 2004) under natural and manipulated environmental conditions in the field. Most of our projects are with clients specializing in wildlife and fisheries sciences.

I have worked extensively with wildlife biologists in Alaska, USA, and Russia (formerly the Soviet Union) on projects of mutual interest. Using contacts with Russian colleagues, I arranged to travel through Siberia in 2003 from the far east to the Ural Mountains visiting biologists in a number of the State Nature Reserves (Zapovedniks), National Parks, and the Ural's Branch of the Russian Academy of Science. I volunteered my time to work with the biologists on analysis of data from their observational studies and on design and future analysis of observational studies and manipulative experiments. I report on my personal observations of the biologists' knowledge of applied statistical methods in this paper.

UNIVERSITY TRAINING IN APPLIED STATISTICAL METHODS

The biologists that I worked with in Russia were relatively well trained in basic mathematics. For example, their working knowledge of calculus and matrix algebra was better than that of biologists working in similar field jobs in the United States. However, training in many basic applied statistical methods was either absent, poor, or incorrect. I give three examples of the inadequate teaching of applied statistical methods to the Russian biologists that I worked with.

Applied Linear Statistical Models

The biologists were aware of multiple regression models, however some had been taught that they could not fit models unless all independent and dependent variable were continuous. Modern methods for fitting linear statistical models (e.g., Neter *et al.*, 1996) were largely unknown. Individuals had not had contact with common procedures such as fitting linear models with discrete predictor variables or the use of indicator variables for qualitative predictor variables.

Selection of Samples from a Finite Population

Biologists with advanced graduate degrees, whom I worked with, did not understand the concept of sampling from a population of study units in the spirit of sampling as is presented in books such as Cochran (1977) and Thompson (1992). They were unable to use a table of random

digits or a random number generator to select a simple random sample from a large study area. They were unaware of topics such as sampling with or without replacement.

Observational Studies and Manipulative Experiments

I helped analyse data from observational studies, e.g., quasi-experiments without the benefit of random assignment of “quasi-treatments” to experimental units (Cook and Campbell, 1979). The biologists found it difficult to understand why they could not draw causal relationships between the observed responses and the quasi-treatments.

Time Series Data

I found some biologists with advanced graduate degrees to be well trained in statistical analysis of time series data. I believe this to be the case for two reasons. First, time series models are inherently more “mathematical” and are likely better understood and taught by university mathematics instructors. Second, the thesis topics for granting of some of the Ph.D. degrees have involved analyses of long term monitoring data collected on biological populations, e.g., analysis of many years of morphological data collected on carcasses of a species collected by commercial or sport hunting.

PUBLICATION OF RESEARCH RESULTS IN NON-RUSSIAN JOURNALS

I was asked to review several manuscripts in English that had been submitted to journals outside Russia and had been rejected. The authors assured me that, if left in Russian and submitted to a Russian journal, the papers would be published without much effort. They could not understand why the papers were rejected. There were, of course, the usual problems of editing papers in English written by non-native speaking authors. For the papers I reviewed, however, the basic problem was that the conclusions drawn in the papers were written in a strong tone as if the results were from a manipulative experiment when in fact, these data were from observational studies. Most of the observational studies involved long term monitoring data or studies of large areas conducted under difficult conditions and are worthy of publication in my opinion if proper statistical inferences are made.

Potential confounding factors and inadequate replication with collaborative results, limited the statistical inferences and editors had no choice but to reject the papers or to ask for extensive revisions that the authors were unprepared to understand or make. The scientists had not been taught the difference between: 1) statistical inferences based on sampling and modelling in observational studies and 2) statistical inferences based on manipulative experiments.

DISCUSSION

I grant that my “sample” of Russian biologists is small and there are likely exceptions to the lack of training in applied statistical methods. The symptoms that I observed, however, point to a serious gap in the teaching of statistics at the university level in Russia and probably in many of the former states of the Soviet Union. In my interactions with Russian colleagues over a 20 year period, I met only one individual whom I would call an applied statistician. He worked for one of the federal agencies involved with estimation of the size of populations of commercially valuable wildlife populations and is now retired. Other Russian colleagues have travelled extensively in international circles of scientists and have learned much statistical theory and are putting it into practice. I am unaware of the training of mathematics and statistics instructors that could teach applied statistics to the next generation of Russian biologists, but would guess that the situation is not improving quickly. I am unaware of the extent of teaching of applied statistical methods in other disciplines of science, but my intuition is that the situation is similar.

I have sent a proposal to the U. S. Fish and Wildlife Service to teach a short course(s) on applied environmental statistics to biologists working in the State Nature Reserves and the National Parks in Russia. The Agency has sponsored short courses on Geographical Information Systems for these biologists and expressed interest in the proposal for teaching environmental statistics, but has not funded it. I know it sounds like “sour grapes” to say that, apparently the Service is willing to pay for the biologists to know how to record their data and make pretty maps, but not to run a proper statistical analyses.

As a practicing statistical consultant, I have not been active in ICOTS and may not be aware of efforts to improve the teaching of statistics in Russia and the former Soviet Union States. Assuming that the Conference has not been actively involved in Russia, I challenge ICOTS to organize a campaign to bring this gap to the attention of the international community of statisticians and to do whatever else is within its charter. The problem is real and deserves the attention of ICOTS.

If there are individuals actively teaching applied statistics in Russia among the audience or among the readers of the proceedings of this conference, I apologize if I have offended anyone and I would appreciate communicating with you.

REFERENCES

- Cochran, W. G. (1977). *Sampling Techniques* (3rd edition). New York: John Wiley and Sons.
- Cook, T.D. and Campbell, D. T. (1979). *Quasi-experimentation: Design and Analysis Issues for Field Experimentation*. Chicago: Rand McNally.
- Gotelli, N. J. and Ellison, A. M. (2004). *A Primer of Ecological Statistics*. Sunderland, Massachusetts: Sinauer.
- Neter, J., Kutner, M. H., Nachtsheim, C. J., and Wasserman, W. (1996). *Applied Linear Statistical Models* (4th edition). Boston: McGraw-Hill.
- Thompson, S. K. (1992). *Sampling*. New York: John Wiley and Sons.