Journal of Official Statistics
Vol. 5. No. 4. 1989. pp. 471–473

Letters to the Editor

Letters to the Editor will be confined to discussion of papers which have appeared in the Journal of Official Statistics and of important issues facing the statistical community.

Dear Editor:

The Chief Editor requested this letter to address the current status of environmental statistics. It is an outgrowth of a discussion we had with Editor P. Dean in Washington, D.C. in August 1989. We appreciate the opportunity to use this special issue of *Journal of Official Statistics* as a forum for our views. Our experience is mainly in government-supported environmental studies in the United States. We encourage comments from those of different perspectives.

Environmental sciences and the associated statistical methodologies are relatively young, and are still in a period of rapid advancement. Statistical progress is most obvious in (1) the growth of statistical literature in environmental science journals and increased frequency of environmental topics in statistical journals, (2) the increasing number of statistical presentations at major environmental meetings and symposia, and (3) the appearance of symposia dealing exclusively with the unique statistical problems of environmental sciences.

In any emerging discipline, such as environmental science, intuitive data analysis techniques will initially prevail. Later, established statistical methodologies will be adapted. Finally, rigorous statistical methodologies will be tailored to the specific applications within the discipline. Only a decade ago, for example, hazardous waste sites were often described by the mean and standard deviation of the observed concentrations of toxic chemicals. Site comparisons typically used Student's t-tests, often violating the assumptions of homogeneity of variance and stochastic independence. In the early 1980s, kriging techniques were adapted from geostatistics to estimate the spatial distribution of chemicals. Recently, the environmental applications of kriging have led to the development and adaptation of certain new forms. These may eventually be adopted as standard procedures in the coming decade. Many aspects of environmental statistics are in the difficult transitional phase between the simple and intuitive and the mathematically rigorous.

There is, however, much room for improvement and further growth. Three of the main challenges in the maturation of environmental statistics are (1) the complexity of measuring and interpreting environmental data, (2) communication limitations between statisticians and environmental scientists, and (3) lack of interdisciplinary awareness and education.

Most, if not all, environmental data sets are multivariate. Typically, the distributions of such data are not normal and do not lend themselves to classical analyses. Many are spatially and temporally autocorrelated. Few are reproducible. Often the number of factors affecting an environmental experiment is so great, and their form and interactions so complex, that effective modeling is difficult. Therefore, the analysis of such complex data requires close teamwork between statisticians and environmental scientists. The frequency and quality of this teamwork is often less than optimal.

Ambiguity in data quality as a result of measurement difficulties further hinders the statistical interpretation of environmental data. Often various phases of environmental studies are conducted by contractors. Some contractors are primarily motivated by profit and secondarily by product quality. Therefore, data quality and validity can be compromised. Environmental analytical chemistry is a challenging discipline. Laboratory methods can be subject to substantial imprecision, even on standard materials. Test portions are easily contaminated, leading to inaccurate or missing data. Clearly, assessing data quality is a challenge in environmental sciences. Field scientists are often confronted with uncontrollable, even unobservable, variables. The state-of-the-art of instrumentation is often insufficient to allow sophisticated measurement in the field. This limitation has recently received much attention. Tremendous improvements may soon be possible with advanced, portable instrumentation. Field portable x-ray fluorescence spectrometers and in-situ electrochemical probes capable of continuous digital logging are two examples of this new technology.

Environmental analyses are expensive and time-consuming. Often, to perform safe, financially responsible studies on potentially toxic samples, it is necessary to severely limit the numbers of analyses to be performed. As a result, the Type II error rate may be sacrificed.

Communication between statisticians and environmental scientists is sometimes ineffective. Terms such as "sample," "correlated," "parameter," and "variable" have different meanings to statisticians and environmental scientists. Often the problem is more serious than just semantics. We can recall giving a presentation, consisting of simple scatterplots and correlation coefficients of the reported concentrations of pairs of compounds, to experienced analytical chemists. Afterward, we were approached by prominent chemists who stated, "I don't understand all that statistical jargon, but it sure sounded great." Our experience indicates that communicating relatively simple statistical approaches to non-statisticians can be challenging.

Many environmental scientists are not aware of the role of a statistician in enhancing environmental data collection, evaluation, and interpretation. While some environmental scientists have experience in data analysis, many lack training in the underlying theory and properties of even simple, standard tests. We have seen many analytical chemists, uncertain how to interpret randomness in their data, deliberately delete "nonsignificant" digits until only one remains. At low reported concentrations, often entire observations are censored to prevent potential misinterpretation. This renders certain types of estimation and inference much more difficult. Conversely, statisticians are not trained in the environmental sciences. Nature is a great equalizer. Sometimes elegant and rigorous experiments are too expensive or are impractical due to inflexibility or unexpected field conditions. The statistician tends to emphasize statistical significance at some arbitrary significance level, without detailed understanding of the system or experiment. The applied scientist tends to emphasize the practical significance based on experience and knowledge of the system or experiment, without regard to probabilities. These are seldom coincident, and further effort is always necessary to reconcile these issues.

Education in experimental design is often lacking for both the statistician and the environmental scientist. Many graduate courses and textbooks in experimental design do not deal with designing an experiment before the data are collected. Rather, the student is taught to categorize an existing data set into a design and to analyze the data accordingly. We suggest a different approach. Courses and textbooks should teach the student to plan experiments that achieve experimental objectives in a practical fashion and attain optimal power for a given budget, or attain specific power at minimal expense. The graduate should be able to determine, for example, the optimal number of lakes, sampling crews, and replications in a study comparing two water sampling methods, given a fixed budget, limited time and a range of field conditions.

These are exciting times in the environmental sciences. Wise decision- and policy-making can be possible only with carefully planned experiments, high-quality data, and rigorous data analyses. Improved communication and education on the part of the statisticians and environmental scientists could facilitate these objectives.

Martin A. Stapanian and Forest C. Garner
Lockheed Engineering and Sciences Co.
1050 E. Flamingo Rd.
Las Vegas, NV 89119 USA