Statistical Science
1991, Vol. 6, No. 3, 217–218

In This Issue

The federal statistical system collects enormous amounts of microdata at substantial cost, for a variety of administrative purposes. It is reasonable that these data also be available for research and business purposes. Many researchers seek to use these data, aided by an ever-developing array of statistical techniques, and with ever-increasing computer power. George Duncan and Robert Pearson, in our first article, address one risk of making such data available, namely, protecting the confidentiality of the individuals and institutions surveyed. When multiple data sources are available, there is increased risk that users will combine files and discover the identity of various subjects. This issue will become even more important in the future, and so we welcome this timely contribution and discussion of the difficulties that government and other institutions face in protecting confidentiality while providing useful data.

Statisticians know that much can be learned about a population without knowing the specifics of any individual, as, for example, when individual measurements are made with random error. Duncan and Pearson suggest ways to introduce such randomness into the data deliberately, and the use of other related techniques for protecting individual identities. Also considered are the legal and the technical issues encountered when restricting or controlling access to data. Ultimately, the self-imposed restraints made by researchers, and their willingness to accept limitations, will be crucial if such data sets are to continue to be available. Three discussants, two in the federal government, add to these deliberations by suggesting additional techniques for masking data. They also consider ethics, legalities and other difficulties of maintaining confidentiality. Janet Norwood, the current Commissioner of Labor Statistics, and former President of the American Statistical Association, provides the perspective of one data steward. She notes the difficulty for statistical agencies of applying confidentiality standards uniformly and, as a consequence, the tendencies of the agencies to be overly protective of data.

Christopher Chatfield, from the University of Bath, has written a variety of texts and papers in recent years on how to do useful statistics. Those familiar with his recent Chapman and Hall books, *Introduction to Multivariate Analysis* (1980), *Statistics for Technology* (3rd edition, 1989), *Problem Solving: A Statistician's Guide* (1988) and *The Analysis of Time Series: An Introduction* (4th edition, 1989), know this. Chatfield always is concerned not only with the introduction of formal statistical methods, but also with the surrounding issues related to how data are collected, issues of computation, consultation, report writing, and the other concerns that confront everyone involved in practical data analysis. In his article here, "Avoiding Statistical Pitfalls," Chatfield provides a range of real-life familiar examples that would frustrate the naive statistician who expects real data to be as well behaved as those of the classroom. Our readers will find the variety of Chatfield's examples fascinating, ranging from the commonplace problem of identifying outliers, to the esoteric, such as explaining the 1986 Challenger space shuttle catastrophe. The group of discussants makes it clear that these statistical issues, addressed by Chatfield, and enhanced by the discussants, are fascinating to encounter. The challenges they present actually provide a lot of the fun of doing statistics. You will find much humor, as well as seriousness, in the comments of Ned Glick, in the other commentaries and in the author's rejoinder.

Peter Jagers, a Swedish statistician, was asked to give one of the 1990 Special Invited Papers by the IMS. His topic, "The Growth and Stabilization of Populations," is treated here. Jagers starts by reminding us that mathematical population theory is not the study of human populations, but rather it is a more general idea that applies also to biology, to particle physics and to a range of other populations. As such, the subject is quite mathematical, in that ideas are abstracted from a variety of applications. Jagers reviews applications, examples and the connections with renewal theory, Poisson theory, genetics and mutations. The discussants, who praise Jager's elegant approach, expand on this topic by raising new questions about long-run behavior, heterogeneity and the geographical structure of populations.

While the formula for the variance of a binomial random variable is widely known to be $E(X - np)^2 = npq$, the expected absolute deviation, $E|X - np|$, would seem to be hopelessly complicated. Yet in 1730 Abraham De Moivre gave a simple formula for this, which is virtually unknown today. Persi Diaconis, who was interviewed in the August 1986 *Statistical Science*, and Sandy Zabell, who wrote about R. A. Fisher and inverse probability in the August 1989 issue, review De Moivre's work and its implications. Included as a postscript is the fascinating history of Sir Alexander Cuming, a scoundrel in many ways, but also an inspirer of both Stirling and De Moivre. After presenting De Moivre's work, its applications and its descendants, the authors extend these results to Pearson families, using orthogonal polynomials for such fami-