Rejoinder

Ronald A. Thisted

I am grateful to Professor Huber and Dr. Chambers for their thoughtful and helpful comments on computing environments. They afford me the opportunity to elaborate on some obscurities and to reassess some of the issues in the light of passing time.

As Huber notes, the passage of time makes nearly every writing on computation obsolete by the time it reaches the printed page. The first draft of the manuscript was completed in the spring of 1984. What was *avant garde* 2 years ago may be "old-fashioned" today. (Indeed one of my examples [Lotus 1-2-3] has now been supplanted by more powerful software.) Even by the standards of 1984, however, interactive statistical programs such as GLIM or Minitab were hardly at the forefront of computing. The choice to speak in terms of familiar and widely available software as the vehicle for statistical *computations* was conscious and deliberate. Every person doing data analysis is doing so within *some* environment, consisting of the sum total of tools needed to get the job done. Like it or not, the environment with which most working statisticians are familiar is a decidedly old-fashioned one, in which paper and pencil continue to play a prominent role, and the workhorse is an interactive statistical program running on a time-shared computer. A recently completed study of computing resources available in major Ph.D.-granting Departments of Statistics (Eddy et al., 1986) indicates that many, perhaps most, departments in which statistical research is done have little more than access to standard computer packages via terminals. Within that context, it is still possible to improve the computing environment for data analysis dramatically and at relatively little cost. Stand-alone workstations appear to be the wave of the future, but interactive programs continue to be the staple of the present for most readers of this journal. It is important, therefore, to think of statistical strategies in terms of the tools available to and used by the majority of data analysts, and to ask whether those tools can be integrated in more effective ways.

Time has worked another change. Neither of the two statistical systems mentioned prominently by the discussants was generally available when the paper was written. *S* had been available to research institutions on an informal basis, but it could not be purchased as a commercial product before summer of 1984. ISP should be available to the public in early 1986. Each reflects considerable thought and effort to make it easier to integrate nonstatistical subsidiary tasks with the data analytic computations themselves. They represent one approach—a decidedly successful one—to addressing some of the issues raised in the paper. This approach evolves from the concept of a statistical package. On this view, the same program that does the statistics should do (or make available) the supporting work, too.

An alternative approach to that of the discussants is based on a rather different metaphor, that of a workbench which makes it easier to use existing tools. Interactive packages are widely available and are here to stay. For many, they will be the primary resource for statistical analysis for the foreseeable future. (In computing, this means approximately 2 years.) Many hundreds of person-years have been invested, not only by their designers in creating quality software, but also by their users in learning how to employ them effectively in data analysis. I am reluctant to relegate them to the scrap heap; it seems preferable to build on such an investment rather than to discard it. On this view, the approach is on *integrating existing resources.*

An example will illustrate the difference in approaches. It is now possible to implement several of the ideas of the paper using an Apple Macintosh computer with a modem and three off-the-shelf programs. One program is a terminal emulator which makes it possible to use the Macintosh either as a text terminal or—in a separate window—as a graphics terminal; in addition, this emulation program makes it easy to keep an automatic transcript of an entire terminal session. The second program is a word processor which allows four different windows to be active at once, each in effect, is a separate notebook. The third program makes it possible to switch from terminal to word processor virtually instantaneously, and to copy portions of text and graphics from one to the other. Thus, for instance, one can immediately implement the three-ring binder notion in conjunction with *any* interactive statistical package running on *any* time-sharing computer to which one has access. I routinely employ this single computing environment to use Minitab on a DECSystem-20 running TOPS-20, GLIM on a SUN running UNIX, and SAS on an IBM 3081 running MTS. The components added to the environment are simply tools for integration; they are ignorant of the underlying tools which do statistical computations. A more detailed account can be found in Thisted (1986).

Chambers discusses two main points, and I find

Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve, and extend access to *Statistical Science.* www.jstor.org

them succinct expressions almost entirely congruent with my own views. However, I draw rather different conclusions from these principles which stem in part from a different understanding of the term "computing environment." For Chambers, it appears to be a set of computing hardware, together with software which makes good use of that hardware. On this view a software "environment" differs from a statistical package in that the former a) is cognizant of the underlying hardware, b) takes on many of the tasks usually associated with operating systems, and c) integrates operating system tasks with statistical procedures. Such an environment can be built, of course, only by an expert. I prefer to think of a computing environment as something one tailors for oneself from available hardware and software so that, on this view, the environment consists of the collection of tools at hand and the ways in which they are integrated with one another.

With this distinction in mind it is easier to appreciate Chambers' conclusion that the high end personal computers are not adequate to support a complete computing environment. But such a system *may* provide all of the additional integrating tools needed to alter radically an existing computing environment based on time-shared computers.

Chambers' second point is reinforced by Huber, that the major gains we as statisticians have seen have been byproducts of the work of nonstatisticians, and that this is likely not to change. While in general this is true, Chambers and Huber both are overly modest; each has made important contributions both to the ways in which we think about data analysis and to the computing systems within which we carry out these analyses.

Professor Huber's closing remarks are most salient. It is virtually impossible to predict what hardware and software resources will be available 3 years hence. This fact makes it all the more important for statisticians to focus on their strategies for data analysis, and to consider ways in which new computer tools can make old ones more effective.

ADDITIONAL REFERENCES

EDDY, W. F., HUBER, P. J., MCCLURE, D. E., MOORE, D. S., STUETZLE, W. and THISTED, R. A. (1986). *Computers in Statistical Research*, report of a workshop on the uses of computers in statistical research, sponsored by the Office of Naval Research and the Institute of Mathematical Statistics.

THISTED, R. A. (1986). Tools for data analysis management. In *Computer Science and Statistics: Eighteenth Symposium on the Interface*, to appear.