Statistical Science
1986, Vol. 1, No. 4, 419–453

Computers in Statistical Research

Workshop on the Use of Computers in Statistical Research/William F. Eddy, Chairman

Abstract. During the years since 1982, research in a small number of leading statistics departments in the United States has undergone a dramatic change as these departments have acquired significant computational resources of their own. Primary support for the acquisition of this equipment has come from the instrumentation programs of the National Science Foundation and the Department of Defense. The purpose of this study is to provide a timely assessment of the current state of computing resources in statistics departments and a projection of future needs. During 1985 the Workshop conducted a mail survey of leading research-doctorate statistics departments and held extensive discussions during two two-day meetings. We have arrived at a series of recommendations directed, variously, at statistics departments, university administrations, professional organizations, and research sponsors.

Key words and phrases: Communication networks, computational statistics, computer resources, departmental infrastructure, hardware configurations, theory and methods research.

1. INTRODUCTION

July 1, 1982 was a red-letter day for basic statistical research. On that day the first funds were awarded by the Division of Mathematical Sciences of the National Science Foundation (NSF) under its program Scientific Computing Research Equipment in the Mathematical Sciences (SCREMS). Shortly thereafter the first funds were awarded by the Department of Defense (DoD) under its University Research Instrumentation Program (DURIP). These funds provided some academic statistics departments an opportunity to acquire their own computer equipment. Prior to this date, statistical research had been treated both by the Federal funding agencies and by universities themselves as a pencil and paper activity.

Statistical practice has long made use of computers for the analysis of data. Recently, computers have allowed the acquisition and analysis of much larger data sets than in the past. Correspondingly, there has been a need for larger and more complex statistical models. Simultaneously, computer technology has allowed the development of new statistical methods. For example, it is now possible to use computer-based graphics as a method of processing and displaying information. Such widely used statistical methods as log linear models and generalized linear models would not have been developed without computers. More recent analytical methods such as projection pursuit and bootstrapping could not exist without powerful computers.

In short, computers are an essential part of statistical research. It is therefore necessary that statistical research groups acquire adequate computer hardware to facilitate this research. The situation is no different than in the "big" sciences, such as physics and chemistry, which have owned and operated their own specialized computational facilities for a number of years.

Simultaneously, computers have generated some enormous, as yet uncharted, research opportunities for statisticians. In particular, there are clear opportunities in image processing, automated control, and signal detection. Other areas more specifically related

---

This article is the report of a Workshop on the present and future needs for computer equipment and operating expenses for computing facilities to support statistical research. The members of the Workshop were William F. Eddy (Chairman), Associate Professor, Department of Statistics, Carnegie-Mellon University, Pittsburgh, Pennsylvania 15213; Peter J. Huber, Professor, Department of Statistics, Harvard University, Cambridge, Massachusetts 02115; Donald E. McClure, Professor and Chairman, Division of Applied Mathematics, Brown University, Providence, Rhode Island 02912; David S. Moore, Professor, Department of Statistics, Purdue University, West Lafayette, Indiana 47907; Werner Stuetzle, Associate Professor, Department of Statistics, University of Washington, Seattle, Washington 98195; and Ronald A. Thisted, Associate Professor, Department of Statistics, University of Chicago, Chicago, Illinois 60637.

to computing include the development of tools for scientific computing, new computer architectures, and artificial intelligence. Most of these topics provide new opportunities for collaboration which, itself, could have great benefit. Statistics has the opportunity for cross-fertilization with other sciences and computers form the common ground.

In our review of the current state of computing resources in statistics departments we found a surprising diversity. There is great variation in the quality and quantity of actual facilities, in the perceived research needs for facilities, and in the sophistication and knowledge about possible facilities. There are departments whose computer facilities are state of the art, even for computer science departments, and there are departments that have no facilities of their own. There are departments which express great need for new equipment and some which are still digesting what they have acquired. There are departments which have "inside" knowledge about next generation hardware and there are departments which seem unaware of the basic value of having their own facilities.

How much support for computational facilities in statistics is enough? We know of no method for deducing the "answer" from obvious first principles. On the other hand we know what dollars will buy. Useful microcomputers can be purchased for $5,000 to $10,000; workstations for $20,000 to $40,000; and minicomputers for $100,000 and up. Thus, 10 faculty researchers in one department can be supported with a capital expenditure in the range of $150,000 to $300,000. The equipment can be expected to have a useful life of no more than 5 years. Skilled programmers can be hired for no less than $35,000 per 12-month year. Hardware maintenance costs about 1% of equipment list price per month. Software maintenance costs are more varied but 10–20% of hardware maintenance is not uncommon. Thus, a crude but useful summary figure is that *$10,000 per researcher per year on a continuing basis will provide a department substantial computational resources with adequate operating support.*

2. SURVEY OF CURRENT RESOURCES AND FUTURE NEEDS

A portion of our assessment of current and planned resources for research computing in statistics is based on responses to a survey mailed in March 1985 to all departments in the United States granting degrees in statistics. The survey instrument contained a series of questions designed to gather detailed information on computer equipment owned by a department and another series designed to gather detailed information on equipment planned for future acquisition. Additional sections of the document gathered information on support personnel (including faculty effort), network accessibility, costs of acquisition and operation, and adequacy of existing facilities.

The core target population for the survey consisted of major Ph.D.-granting statistics departments. This population, defined primarily from the National Research Council study *An Assessment of Research-Doctorate Programs in the United States*, is listed in Table 2.1. Two stages of follow-up were carried out to increase response from this core population. All data in this section refer to the 30 responding departments in this group.

In addition, responses to the initial mailing were received from three other groups. Thirteen biostatistics departments, 20 mathematical sciences departments, and 13 miscellaneous others (management, behavioral sciences, small statistics groups, and Canadian statistics departments) responded. The initial response rate was low for these populations, and no follow-up was attempted. Moreover, these responses represent less well defined populations than the independent research-doctorate statistics departments. For these reasons, data from these additional re-

TABLE 2.1
Computing resources sample population

Respondents (n = 30)	Nonrespondents (n = 12)
University of California, Berkeley	Columbia University
Carnegie-Mellon University	University of California, Davis
Colorado State University	Florida State University
University of Connecticut	George Washington University
University of Chicago	University of Iowa
University of Florida	Kansas State University
University of Georgia	North Carolina State University
Harvard University	Pennsylvania State University
University of Illinois	University of California, Riverside
Iowa State University	Southern Methodist University
University of Kentucky	State University of New York, Buffalo
University of Michigan	Temple University
Michigan State University	
University of Minnesota	
University of Missouri–Columbia	
University of North Carolina	
Ohio State University	
Oklahoma State University	
Oregon State University	
University of Pennsylvania	
Purdue University	
University of Rochester	
Rutgers University	
Stanford University	
Texas A&M University	
VPI and State University	
University of Washington	
University of Wisconsin	
University of Wyoming	
Yale University	

sponses are not used in this section. Summary comments concerning these groups are made in Appendix I, however.

2.1 Present Resources

The responding departments vary widely in both scope and type of computer hardware owned. Eleven of the 30 have multiuser systems (9 VAX, 1 PDP 11/73, 1 Pyramid). Another 5 have workstations that provide some similar capabilities. The remainder apparently rely on central systems and/or microcomputers. With regard to graphics capability, 16 departments have workstations or graphics terminals capable of interactive graphics. Another 5 have microcomputers (Macintosh, IBM PC/XT) with some graphics capability, and 2 more rely on suitable plotters or printers for graphics. The remaining departments have only unsophisticated microcomputers (Apple IIe, IBM PC). Software is presumably even more varied. Table 2.2 summarizes departmental equipment.

2.2 Hardware Ordered or Planned

Fifteen departments plan to acquire additional microcomputers. (One has ordered 27 DEC PRO 380's.) Six departments plan graphics terminals, 7 will add workstations, and 4 will install multiuser systems. Of the 14 departments not now possessing multiuser systems/workstations, 4 plan to acquire them and others have or will acquire powerful microcomputers such as the PC/XT or PC/AT. In the near future, only 7 departments will have nothing more powerful than an IBM PC dedicated to research use.

2.3 Personnel

Only 4 departments employ undergraduates assigned to computing, but 17 assign such duties to graduate assistants (5 less than one full-time equivalent (FTE), 9 one or two FTE, 3 more than two FTE). Eighteen have no nonstudent staff, 3 less than one staff FTE, and 8 have one or two FTE. Seventeen allocate no faculty time to support of computing, and only 3 departments assign as much as 25% of a faculty member's effort (usually spread among several individuals). Despite the lack of allocation of faculty time, we believe there is considerable application of faculty time to support of computing activities. The biostatistics responses (see Appendix I) offer a notable contrast in staff support and general level of organization of computing activities.

2.4 Other Resources

Sixteen departments mentioned no networks extending off campus. This is no doubt a reflection of lack of use as much as lack of availability. The networks most often listed as available were BITnet (11 departments), ARPAnet (7), USEnet (5), and TELEnet (3). Nine departments have some access to supercomputers and 2 more plan such access.

2.5 Costs

The survey requested information on both departmental budget cost and external funding in 1984–1985 for several categories of computing-related expenses. Unlike the biostatistics groups (see Appendix I), many statistics groups could not provide detailed cost information. The variety of financial arrangements and accounting systems make the resulting data hard to

TABLE 2.2
Present departmental equipment

Microcomputers											
Number owned	0	1	2	3	4	5	6	7	9	16	17
Departments	7	7	4	2	1	3	1	1	2	1	1
Graphics terminals											
Number owned	0	1	2	3	4	5	7				
Departments	16	6	3	1	2	1	1				
Workstations											
Number owned	0	1	2	3	4						
Departments	21	2	3	3	1						
Other terminals											
Number owned	0	1–4	5–8		9–12	13–16		17–20	20+	NR[a]	
Departments	5	7	5		3	1		4	4	1	
Plotters											
Number owned	0	1	2	3	6						
Departments	18	7	3	1	1						
Printers											
Number owned	0	1	2	3	4	6	8	9	16	17	
Departments	6	10	5	2	2	1	1	1	1	1	

[a] NR, no response.

TABLE 2.3
Operating costs (in thousands of dollars)

Departmental costs:	0,	3,	15,	79,	343
External funding:	0,	0,	6,	25,	174

TABLE 2.4
Departmental priorities

	Rank								Mean
	1	2	3	4	5	6	7	NR[a]	
Hardware acquisition	13	2	4	4	3	1	0	3	2.44
Software acquisition	3	14	5	4	0	1	1	2	2.68
Support personnel	5	8	4	0	1	2	1	9	2.71
Operation, maintenance	5	3	8	1	7	1	2	3	3.48
Faculty release time	1	1	4	5	7	1	3	8	4.41
Student research	0	1	2	11	2	2	4	8	4.64
Network access	2	0	2	0	1	9	5	11	5.37

[a] NR, no response.

summarize. Six departments did not respond, and 6 claimed no direct cost to the department even though all 6 claim to own some hardware and several claim to own substantial amounts. Four respondents mentioned fee systems in which most operating costs are paid by grants and contracts. For the record, the 5-number summaries (minimum, lower quartile, median, upper quartile, maximum) for the respondents are given in Table 2.3.

2.6 Needs and Priorities

Survey respondents were asked to rank seven categories as priorities for additional funding. The distribution of ranks is given in Table 2.4. Here "NR" indicates no response, and often implies a low priority since several respondents ranked only their top 3 or 4 categories.

Priorities vary with the current status of department facilities. For example, 10 of the 14 departments without a multiuser system or workstation ranked "hardware acquisition" as their first priority. Only 3 of the 16 departments with such equipment ranked hardware acquisition first, but 9 ranked "support personnel" either first or second. Support for software acquisition is quite consistent among both groups of departments, while those with such equipment give "operation and maintenance" higher priority than do those without. The departments with systems or workstations ranked "operation and maintenance" either quite high (11 of 16 placed it in the top 3) or very low (the other 5 ranked it 5 or lower or did not rank it). Clearly individual circumstances vary quite a bit. The remaining options ("network access," "faculty release time," and "student research support") were generally accorded lower priority. The rankings of support personnel versus faculty release time probably reflect a consensus that technical support should be provided by technical staff rather than by faculty—who often do this job at present without release time.

3. COMPUTATION AND STATISTICAL RESEARCH

Academic statistics departments have three major missions: teaching, research, and consulting. Computers are fundamentally changing the way all of these missions are accomplished. While it is not always possible to separate these activities (e.g., research is one aspect of graduate education), the focus of this report is on the research uses of computers. We have identified four categories of such use. Beyond the uses of computers for research in statistics per se, there are strong interdisciplinary connections that bind statistics to the use of computation in other basic sciences. Section 3.5 below discusses some of the effects of the interdisciplinary connections on how computers and computational methods are used in the sciences.

3.1 Infrastructure

The influence of computers on those statistics departments which have acquired their own is not only directly on the research itself, but also on the support of research activities. One major aspect of this influence is communication. Computer mail provides an opportunity for rapid, asynchronous, written communication between researchers. The easy transfer of data, programs, and manuscripts has a profound positive impact on the speed and nature of collaborative research. These exchanges take place within the confines of one department, across disciplines within a university, and, when access to national networks is available, between universities. It is the opinion of this panel that the low ranking given network access

by respondents to our survey merely reflects the fact that one cannot appreciate the important role of electronic communication before one has become accustomed to the use of it.

Another major aspect of the influence of computers is in document production. The main outputs from statistical research and data analysis are written reports. Computers can provide the software tools for text editing and document formatting. Document formatting software together with an adequate output device provides the capability to produce technical reports with complex equations of near typeset quality. The more powerful systems allow for integration of text, formulas, data, programs, and graphics in a single document.

3.2 Data Analysis

The practice of data analysis has been dramatically altered by access to interactive computing. Statistics department computer facilities can provide a direct benefit to other disciplines through this interaction. Analysis can be performed step by step, with immediate feedback in the form of numerical results and, more importantly, graphs produced on the screen of a computer terminal. Newer statistical languages combine the possibility of performing complex operations, such as regression and cluster analysis, on large data sets with the flexibility familiar from the pencil and paper analysis of small data sets. The data analyst is no longer forced to squeeze the problem at hand into a form and framework that is treatable by one of the standard mainframe statistical packages. Improvisation and experimentation are greatly facilitated. The ability to look at data quickly, easily, and in a large number of ways can give a whole new quality to a consulting program and suggest new problems for statistical research.

3.3 Research in Theory and Methods

Many of the recent advances in statistical methodology are unthinkable without the availability of computers. These methods, such as robust estimation, the bootstrap, log linear models, Bayesian methods, nonparametric regression, the Cox model, etc. require a computer for their application. All nonlinear estimation techniques are impractical without computers. Also, computer experimentation is essential for the evaluation of the performance of the various methods and the understanding of their limitations. This reflects the transition of statistics from a purely mathematical to an increasingly experimental discipline. Computers are a tool for the majority of statistical researchers; the increased availability of computers has provided a new impetus to methodological research.

3.4 Research in Computational Statistics

Computers cannot only be used as a tool for statistical research—use of computers in statistics is also becoming a research topic in its own right. Some of these areas of research are:

- Use of high interaction graphics for both input to and output from statistical analysis. Fast, high resolution displays not only allow for the faster production of more pictures with higher quality; they also lead to invention and implementation of qualitatively new methods of looking at data.
- Computing environments for data analysis. Modern statistical languages offer great flexibility and freedom for improvisation. This gives rise to a whole new area of research problems. For example, it becomes crucial to be able to keep and organize records of the intermediate results of an analysis, to go back to previous states, and to repeat selected parts of the analysis in slightly modified form.
- Artificial intelligence and statistics. Some feedback occurs in both directions between artificial intelligence (AI) and statistics. The potential exists for far more interaction. There is ongoing research, for example, in the use of AI methods for the creation of *expert systems* to guide the statistically unsophisticated user through the steps of a statistical analysis. In the other direction, there is considerable activity in the use of frameworks built on statistical theory for modeling AI systems. Bayesian models, in particular, provide a conceptual basis for integrating prior knowledge with incomplete or noise-corrupted observations in order to determine (statistically) optimal decisions. Additionally, statisticians are well qualified to assist AI researchers in the evaluation of the resulting systems.
- Statistical methods for novel computer architectures. Several research groups are exploring the use of new computer architectures for computationally intensive statistical applications. The special-purpose architectures include pipelined and multiple processor systems ranging from common high speed array processors to experimental development systems with many processors and reconfigurable communications.

Research in these latter areas is at the forefront of the interface between statistics, computer engineering, and computer science. The phenomenal developments in computational power are becoming a driving force themselves for research in statistics, as they are in the sciences in general. The availability of resources such as supercomputers, designed for vector operations, has made it possible to do research in statistical methods that was unthinkable on sequential, scalar machines. For instance, algorithm development and

experimentation with maximum-likelihood methods for positron emission tomography reconstructions was facilitated by the availability of a supercomputer; the availability today of general purpose supercomputers and array processors with peak speeds in excess of 100 million floating point operations per second is, for the first time, making true three-dimensional reconstructions conceivable. Research in statistical methods per se is greatly enhanced by the easy access of the research community to supercomputers. For example, the use of bootstrapping for evaluation of performance of nonlinear parameter estimators, simulation experiments with combinatorial testing procedures, and numerical quadrature for multivariate integrals in moderate numbers of dimensions are feasible with today's fastest computers.

3.5 Interdisciplinary Cross-currents

It is widely recognized that computers and computational methods of inquiry are having a phenomenal impact on the way that scientific research is done. This has been documented in *Report of the Panel on Large-scale Computing in Science and Engineering* (1982) (the Lax report), in *Renewing U.S. Mathematics, Critical Resource for the Future* (1984) (the David report), in *Future Directions in Computational Mathematics, Algorithms, and Scientific Software* (1985) (the Rheinboldt report), and elsewhere. The uses within statistics research identified above are indicators of this impact of computers on statistics in its own right.

There are simultaneous parallel effects and developments in all of the basic sciences. The commonality of the developments in each of the sciences presents new opportunities for significant interdisciplinary cooperation. In particular, the special needs and ideas to which research statisticians are the first to respond have the potential of making substantial contributions to other sciences. Further, the common features of computational problems encountered in different areas of science have the potential to foster new interdisciplinary research to which statisticians can make pivotal contributions.

For example, graphical data-analytical methods developed by statisticians—which can in fact trace some of their motivation and origins to interdisciplinary projects involving statisticians and physicists—provide extremely powerful computational tools to all scientists concerned with description, analysis, and interpretation of high-dimensional data sets. In addition, statisticians working with high interaction computer-graphical analysis have identified the need for better computing environments, computing environments that include graphical programming languages, facilities for input and output of graphical information, and operating systems geared to graphical interaction. When this need is fulfilled it will have a striking benefit for many other scientists. For instance, a computational fluid dynamicist at a workstation, communicating with a remote supercomputer, will benefit greatly from the use of graphical methods to analyze multidimensional data on the multiple variables produced by computational models for turbulence.

Further, there are significant contributions by statisticians to basic science and technology, beyond providing tools and methods such as high-interaction graphical data analysis to other computational scientists. For example, in cognitive sciences some so-called connectionist models of machine learning and machine intelligence (realizable by novel parallel computer architectures) draw very heavily on principles of Bayesian inference to define how they make decisions, formulate hypotheses, or incorporate new information with prior information. Similarly, image reconstruction methods used in medical tomographic systems and in nonmedical applications for nondestructive testing have been successfully built on nonparametric statistical formulations and fundamental statistical principles (maximum likelihood and Bayes optimality). Such areas of scientific inquiry and technological development are driven by advances in computers and they rely or draw strongly on basic statistical ideas for their foundations.

4. CREATING THE NECESSARY CAPABILITIES

The survey results indicate a very different perception of needs between those departments which have moderate or substantial computing resources and those which have only minimal access to computing. The latter group emphasizes the need for support of hardware acquisition, while the former stresses the need for support of technical staff and hardware maintenance. This difference reflects a change in focus that departments undergo as efforts shift from obtaining a basic resource toward making it useful and usable on an on-going basis. The specific needs of statistics departments depend in large measure on two factors: (i) the current level of computational maturity in the department, and (ii) the level of research efforts in computational statistics per se. Since research in computational statistics is confined almost completely to those departments in which computing already plays an important role, these two factors can very nearly be thought of as a single dimension. We turn now to an assessment of the needs of departments which fall at various points on this scale. We consider three points on the continuum: the department just emerging from the Dark Ages of the late 1960s, the depart-

ment ready to take additional steps beyond initial hardware acquisition, and the research group engaged in statistical computing research.

At one end of the spectrum, there are departments whose major computing is limited to a large mainframe computer on campus, often operating in batch mode, with limited or no facilities for interactive computing. This setting represents the scientific computing environment of two decades ago, and is inadequate for most of the computing required for statistical research today. Such departments may have obtained one or more personal computers of power comparable to the IBM PC or PC/XT; access to even such limited local computing power often makes the need for expanded resources evident. Approximately 40% of the respondents to our survey fall into this category. Even though acquisition of hardware may well be the top priority for departments which now have little or no access to computation, even the short term needs of such departments cannot be met solely by obtaining suitable computer equipment. The remainder of this section outlines the computing resources that are fundamental to modern statistical research and practice.

4.1 Hardware Requirements

There are a number of different hardware configurations which can be used effectively to meet the needs of statistical science. The strategies which are possible depend on several factors, including the level of funds that can be brought to bear on the problem, the size of the department, and the extent to which some needs can be met adequately through existing campus facilities. Although some headway can be made with less than $10,000, such a funding level nearly dictates purchase of one or two personal computers. Such an expenditure, while helpful, can only be considered a short term step. Stand-alone microcomputers such as the IBM PC/XT do not currently have sufficient power to do manuscript processing at the departmental level, for instance; power roughly equivalent to a VAX 11/750 or a SUN workstation seems to be the minimum requirement today.

The main choice seems to be between one or more workstations and a larger multiuser minicomputer. As used here, a "workstation" is a stand-alone computer, generally with high-resolution graphics, often with software to support multiple windows, and always with substantial computing power. A typical workstation configuration would include a CPU with virtual memory capability, two megabytes (MB) or more of random access memory, 40–80 MB of disk storage, floppy disk or tape storage, high resolution graphics display, telecommunications, and graphics software. Workstations are generally intended by their manufacturers to be single user machines, although often a single user workstation can effectively support a larger user community, of whom perhaps two to four may be using the machine at one time. Workstations cost between $20,000 and $40,000 each (although the price is expected to drop to less than half that in the next 2 years). Multiuser super-minicomputers cost around $100,000. Several workstations can be linked together so as to share a single large disk and to communicate at high speeds via a network; such a network makes it possible to expand resources by adding more workstations.

The hardware needed to make either a network of workstations or a multiuser system effective would include (in addition to the CPU) a laser printer, adequate disk storage, and a number of terminals for remote access to the system. An output device of near typeset quality is necessary, both for preparation of text and for printing hard copies of graphics displays. Laser printers with resolution of 300 dots per inch are adequate in this respect and are becoming common and relatively inexpensive. Terminals which do not reside in the same room as the computer will require additional hardware to connect to the computer via telephone or direct cabling.

4.2 Software Requirements

An unrecognized need among those who lack hardware is that software is necessary and often expensive. Even with substantial academic discounts, software packages useful in statistical research can assume a major role in the on-going budget of a departmental facility. Many of these products require an annual license fee; some are licensed on a per CPU basis.

Operating Systems. No general purpose computer will run without an operating system. Most hardware comes bundled together with an operating system. However, maintenance and upgrades to the operating system can represent a substantial unanticipated cost.

Interactive Data Analysis. The day-to-day practice of data analysis, which is the source of much innovation in statistical methodology and theory, depends increasingly on easy, routine access to interactive software for statistical data analysis. Examples of such programs include S, ISP, GLIM, and MINITAB. One or more of these programs should be available on the departmental computer.

Programming Languages. The cost of programming languages such as C, Fortran, and Pascal may not be included in the price of a computing system. Such languages are indispensable for statistical research.

Mathematical Software. Program libraries such as NAG and IMSL that provide tested, high-quality subroutines for performing standard statistical and mathematical computations should be available.

Electronic Mail. Although software for electronic

mail is included with some systems (such as most UNIX systems), in some it is not. The availability of mail software, and its compatibility with use on national networks, should be ascertained at an early stage. A mail system with store-and-forward capabilities is highly desirable.

Standard Statistical Packages. Access to large standard statistical packages is essential; fortunately, this is the area in which departments are probably best served today. Continuing access to standard packages must be assured. At a minimum, this requires adequate funding to make access to these packages on the available mainframes feasible on a routine basis. Departmental computing facilities are generally ill-suited to providing such access. Large software packages generally entail large annual license fees, and require substantial efforts on the part of support staff to maintain, update, and document. Because of their size, they also absorb a considerable fraction of the available computing resources (CPU and I/O) on computers of the size typically used in departmental facilities, thus degrading system performance for other activities. A department contemplating a facility of its own should *not* consider it a replacement for a large central mainframe, particularly with respect to standard packages.

Technical Manuscript Preparation. Software for text formatting will allow preparation of mathematical equations of arbitrary complexity. Such software is often referred to as software for mathematical typesetting. Examples of such programs are TEX, eqn/troff, and SCRIBE. It is imperative that the printer and the typesetting software be mutually compatible in order to avoid having capabilities in principle that are not realizable in practice.

Other Software. Examples include special purpose language processors (LISP, etc.), special purpose graphics software (DISSPLA, etc.), symbolic algebra packages (MACSYMA, REDUCE, etc.), multiple precision arithmetic packages, and so forth.

4.3 Communications Requirements

There are many computer networks which can provide access to statisticians nationwide; these include BITnet, CSnet, ARPAnet, USEnet, and MAILnet. The particular choice will be determined by the existing availability of computer networks on campus, the mailing software available (as noted above), and constraints imposed by the operating system in use. Connection to such a network requires that the departmental computer be able to communicate with the outside world using either a telephone link or a local area network (LAN) link. Each national network may require additional specialized hardware to connect the computer to the communications channel.

Departmental computing facilities must also have some capability for communicating with other computers in the department or on campus. A modem attached at one end to a serial port on the computer, and at the other end to a telephone line provides an inexpensive, but relatively slow, link to the outside world. Depending on departmental and university resources available, faster links involving dedicated telephone lines or LANs can be used instead.

It is most desirable to be able to transmit data (and programs) from the departmental facility to the mainframes on which standard packages reside, and to be able to transfer data and text between the departmental computer and personal microcomputers. In addition to the communications capabilities just discussed, file-transfer software is also needed to accomplish this. Standard error-detecting protocols exist for file-transfer between computers; *Kermit* is one such protocol which is implemented in public domain software, which is widely used, and which is available for a variety of different computers.

4.4 Support Staff

Of overriding importance for the continued operation of a departmental facility is the need for *technical support staff.* Although this is a universal need, it is one which must be emphasized when a department is considering its own computational facility. Effective use of computing resources requires individuals with expertise and the ability to share their expertise. In addition, it is necessary for someone to perform such tasks as identification of and negotiation with hardware vendors, arranging for maintenance contracts on hardware, scheduling preventive maintenance, arranging for hardware repair (when necessary), obtaining and installing software, installing updates to existing software, maintaining adequate inventory of computer supplies, diagnosing problems, and the like. To the extent that these tasks are either delegated to or simply taken on by a faculty member, it amounts to a net *reduction* of effort being spent on research. Functions better served by staff than by faculty include the following areas.

Operations. No computing facility runs without individuals operating and maintaining equipment. Our survey indicates that at present substantial amounts of faculty time are being spent on activities solely devoted to operations. This is largely due to lack of support for technical staff both from university administrators and from funding agencies. Among departments having adequate computing facilities, *support for technical staff constitutes a major unfilled need which is currently being met by diverting faculty efforts from research to technical support.*

Programming. Coupled with the need for opera-

tions staff, research productivity can be greatly enhanced with on-site programming expertise. This comes in two flavors: systems programming which supports and simplifies use of the facilities and applications programming which supports specific research projects. A technical staff member can play either or both roles more effectively than can either graduate students or faculty. In smaller departments, the operations and programming personnel may overlap substantially.

Planning and Training. Planning is essential for departments, their administrations, and funding agencies. Areas in which departments need planning and support include: establishing a suitable *cost-recovery* mechanism, *acquiring and maintaining software, training* current and future department members, monitoring usage of computing resources and anticipating future needs, and monitoring and dealing effectively with the amount of *faculty and student time* devoted to support of the departmental facility.

4.5 Physical Plant and Maintenance

Every computing facility at every level of power and sophistication has certain costs which are essential to meet if the facility is to usefully support research activities. These are so important, yet so frequently neglected in planning and support, that we list some of the more important ones here.

Maintenance costs for hardware generally amount to 1% of the *list* price of computing equipment per month. This item, generally overlooked by start-up departments, is a continuing headache for departments with existing facilities. Our survey indicates that *adequate provision for maintenance and repair is a major need.*

Space for computing equipment, including sufficient working areas for terminal clusters, graphics laboratories, reference works, maintenance facilities, storage, and staff.

Site preparation is often both expensive and forgotten in planning. This item includes such things as air conditioning and humidity control, providing adequate electrical circuits and power supply, furniture, telephone installation and equipment charges.

Cabling is a necessary expense if terminals are to be provided in every office, or even at a site separated from the physical location of the departmental computer. This includes the cost of cable trays both between and within computing rooms. Because the costs of cabling are primarily labor costs, it is wise to install enough cable for the remotely foreseeable future rather than solely for current needs. On the other hand, because communications technology will surely change it is also desirable to plan for the possibility of installing additional cables in the future.

4.6 Discussion

There is a strong need for departments obtaining facilities of their own to do so with a view toward the ongoing costs in dollars and in people, so as to minimize these costs at the outset. The single most practical way of doing so is by installing a hardware and software environment that is *standard.* By this, we mean that there is great value in having a system which duplicates one that already exists elsewhere in the university. A statistics department with few existing computing resources should generally not be the trendsetter on campus in terms of new hardware or operating systems. The more nearly "off-the-shelf" a configuration is, the less likely it is that the department will have to absorb the high costs associated with learning about a new system. This principle applies at both a higher and a lower level. To the extent that a department's facilities are compatible with those in other statistics departments in the country, it will be relatively easy to make use of the experience of those other departments and to obtain software developed (or debugged) elsewhere. To the extent that the computers within the department (especially microcomputers) are similar to one another, the department can benefit from synergies and will not have to develop means of communication between different computers and different operating systems. *Standardization is more important to long range effectiveness than either price or performance.* Appendix II contains descriptions of the facilities available in several departments; these descriptions are intended to provide departments taking their first steps an indication of what some other departments have done about standardization.

Once an adequate hardware configuration is installed, the primary costs are three-fold: space, maintenance, and personnel. Computing, like laboratory work in other sciences, requires space for hardware and for people. This need will have to be met by the university, and because university administrations are not generally accustomed to thinking of statistical research as being either experimental or requiring research equipment, meeting this need may be extremely difficult. Computing hardware needs maintenance, repair, and replacement. A plan for providing this support is an essential part of a coordinated plan for departmental computing. Providing the funds for continuing operation may require the implementation of an accounting system to allow cost recovery from grants and contracts for the user of the facilities. It appears from a comparison of the statistics and biostatistics responses to the survey that many statistics groups have an inadequate system of recording the costs of computing. Computing costs are often lost in such categories as "supplies and expenses." Awareness

of actual costs incurred is essential for planning and for seeking additional funding from internal or external sources. A very brief description of some existing accounting systems is given in Appendix III.

5. EXPANDING THE CAPABILITIES

The department that has an adequate computing facility generally finds that computing has become an integral departmental activity. The continuing needs of such a department are rather different from those outlined in Section 4 and are focussed less on the CPU and the "main box" and more on peripherals which make computing more effective for those whose research depends in large measure on computation. Assuming that a basic departmental facility consisting of a multiuser system exists, it is useful to divide the additional needs into two categories, those which directly extend the resources necessary for research in statistical theory and methods, and those which represent additional capabilities needed for research in computational statistics.

5.1 Statistical Theory and Methods Research

The hardware requirements will generally involve adding performance-enhancing peripherals to the departmental computer. These include such things as more memory, more communications ports, floating-point accelerators, graphics output devices, and additional disk capacity. Rarely will it involve a move to a larger CPU.

Large scale noninteractive computation that arises in statistical research requires more computing power than a single departmental resource can provide cost-effectively. Modest needs of this sort can be met through running at the lowest available rates on campus mainframes. For large scale problems, *access to supercomputers*, such as that made available to statistical researchers through recent Federal initiatives, is essential. Researchers with large scale needs for interactive computing will generally require dedicated computing equipment for such work; this may involve a move to workstations.

Quite apart from the number of CPU cycles needed for statistical computation, productive research will increasingly require those capabilities universally associated with workstations. Needs here include improved programming environments based upon multiple windows, dynamic debugging, multi-window editors, and virtual memory. Better graphics is a major need; by this we mean high-resolution raster graphics devices, typically with oversized displays so that several windows (at least three) can be simultaneously displayed, and adequate hard copy facilities. Color graphics capabilities are needed for some research programs as well.

The software needs will largely involve special-purpose software tools such as symbolic algebra systems (MACSYMA, REDUCE, SMP, MAPLE), languages or systems used by only one or two individuals within the department (LISP, SCHEME, GKS), and database systems.

A major requirement for making supercomputer access more useful is to establish high-speed communications links for transmitting programs to and for receiving output from supercomputers.

The greatest need for enhancing the utility of a departmental facility is to make it easier to access. Having a terminal for *every* faculty member, and a large pool of terminals for graduate students, is a major need for the computationally mature department. This serves to integrate computing more fully into department activities. Adequate access to departmental resources has a large and positive effect on graduate education. Increasing access, however, means increased needs for space, maintenance, and support personnel.

5.2 Computational Statistics Research

The previous sections largely concern departments with research programs in statistical theory and methods, and the computing resources discussed have been those necessary to support those research efforts. There is a small but growing number of researchers whose work is focussed on computational statistics per se. These efforts include, but are not limited to, such activities as designing computing environments for data analysis, developing algorithms for statistical computations on new computer architectures, applying methods of artificial intelligence to statistical data analysis, employing statistical methods to problems in artificial intelligence, and constructing systems for high interaction graphics.

Such research programs have highly diverse needs for computing equipment. In this arena the equipment needs are driven by specific research projects. What these projects have in common is that they are generally on the cutting edge of computer science research, and consequently require hardware that is not available in large quantities. Support of these activities may well involve substantial hardware expense.

Whereas a major component of expense required to support research in statistical theory and research is directed toward *obtaining* specialized software, a corresponding major aspect of research in computational statistics directly involves *creating* specialized software. Additional technical staff, particularly programming staff, is required to make effective progress in these areas.

6. RECOMMENDATIONS

The panel's recommendations are directed separately to statistics departments, to university administrations, to the professional organizations of university-based research statisticians, and to the principal sponsors of research in statistics. The goal of the recommendations is to provide direction for the departments for coherent development of the computing resources that are needed for research in statistics and to identify priorities for the funding agencies, foundations, industries, universities, and professional societies who support this research.

6.1 Recommendations to Statistics Departments

The individual department necessarily assumes the responsibility for developing its own resources. To carry out this responsibility, we recommend:

- That the department formulate a systematic development plan for building its computing resources, including careful internal records of computing costs by category;
- That the department seek advice from other departments which have already developed their resources to the level planned;
- That the development plan place the highest priority on *standardization* for computing hardware, software, and intersystem communications;
- That the development plan ensure access to facilities for *all* of the department's faculty and graduate students;
- That potential sources of funding (Federal, and, possibly, state or local, government agencies, industrial sponsors, private foundations) for implementation of the development plan be identified and approached.

The formulation of a coherent and comprehensive development plan is critical for attaining the goal of an easy to use and fully integrated departmental computing facility. There are many components to even the simplest computing system. A development plan can ensure that the system that is acquired will fulfill the department's computing requirements, that it is comprised of compatible components, and that it can be expanded as department needs grow and change. While it is very difficult to project, say, 3–5 years ahead, it is useful as part of the planning activity to think about amortization and replacement of the hardware.

Departments that now have no computing resources are well advised to start with a so-called "turnkey" system in which all considerations about adequacy and compatibility among hardware components (user interfaces—terminals or microcomputers, processors, storage media, output devices) and among software components (operating systems, development tools, language compilers and interpreters, program packages) are addressed by the vendor. The alternative to the turnkey approach is to invest a large amount of faculty and/or staff time and energy to do fact-finding on the individual elements of the computing system and to resolve the adequacy and compatibility issues. *Standardization is more important to long range effectiveness than either price or performance.* As a department gains experience with computing systems, the do it yourself approach becomes less burdensome and gives the department greater flexibility in tailoring a system to meet its unique needs.

A department should always seek advice from other departments who have already developed their resources to the level planned. In Appendix II of this report, the resources of several statistics departments are described in some detail. Departments such as these should be contacted for firsthand information about alternative development strategies and existing systems. Advice from experienced users is particularly helpful for tempering the exaggerated claims of hardware and software vendors.

The computer industry is continually evolving toward higher levels of standardization. The extent to which a system adheres to hardware and software standards will determine how easy it is to expand the system, to transport the user's software from one system to another, and to establish network communications between systems. It is much easier to use a system which recognizes industry standards than it is to use a system that is in a world of its own.

The standards are usually not universal, and hence some choices must be made. For instance, among personal computer operating systems both MS-DOS and CP/M are widely used and implementable on a variety of different machines. At a different level, both TCP/IP and DECnet are common networking protocols. Where choices between "standards" such as these must be made, a department should consider how its system will be integrated with other computing resources at the same institution and how the department wants its facility to mesh with systems in use by other research statistics departments. Again, advice from experienced users is helpful.

Departmental computing resources should be as accessible as possible to all faculty and graduate students. There should be no obstacles to the facility's use in terms of where equipment is physically located or in terms of accounting mechanisms which control who is authorized to use the equipment. A great advantage of autonomous departmental computing systems over central mainframe facilities is that such barriers to access can be removed.

Two of the major sources of funding for acquisition of departmental computing systems have been the Special Projects Program in the Division of

Mathematical Sciences at the National Science Foundation and the Department of Defense University Research Instrumentation Program administered through the Office of Naval Research, the United States Army Research Office, and the Air Force Office of Scientific Research. The latter program has usually been tied to already existing basic research contracts from DoD agencies. We recommend that departments which have ties to other mission-oriented Federal agencies (e.g., USDA, DOE, NIH) seek computer equipment funding in support of on-going contracts. Also, computer manufacturers will often give free or heavily discounted equipment in consideration of cooperative research and development projects with universities. Cooperative arrangements with vendors should, of course, be entered cautiously.

6.2 Recommendations to University Administrations

University administrations are unaccustomed to thinking of statistical research as a laboratory science. As a consequence only the most enlightened administrations can be expected to assist a department in the acquisition and operation of its own computing facilities. For a department that is unaccustomed to the acquisition of the large research grants required for a departmental computing facility the assistance of the central administration is essential.

Departmental facilities cannot be expected to satisfy all computational needs for statistics research. For example, these needs will typically include access to standard statistical packages which really require mainframe computer resources and thus are more appropriately supported at a central computer center. Statistics department needs are likely to extend beyond the individual university and include resources such as network links with other universities and use of remote super computers. All of these resources can be made directly and immediately accessible at a departmental facility through various types of intercomputer communications. The communications can range from simple modem/telephone connections to sophisticated networking that combines a department's local area network with a campus internetwork and with an interuniversity network such as BITnet. The department and its institution together should ensure that effective data communications are available.

To assist statistics departments in the development of their own computational resources we recommend to university administrations:

- That statistics departments be provided assistance in (i) the formulation of a development plan and (ii) the identification of and approach to potential sources of funding to implement the plan;
- That central computer center facilities at the institution be integrated with statistics department facilities through a campus communications network or through traditional telecommunications;
- That central computer facilities provide (i) continued access to standard software packages for statistical analysis and (ii) access to one or more national computer networks.

6.3 Recommendations to Professional Societies

One of the greatest needs of a department that is formulating and carrying out a development plan is easy access to reliable information about computing systems. Effective communication mechanisms and good sources of information are equally important for the departments that have the computing resources and need to train and keep their members abreast of the continual evolution of uses of computers for statistical research. The professional societies, in particular the American Statistical Association and the Institute for Mathematical Statistics, are designed for disseminating information among their members. In order to provide for the transfer of information about computers and their use for statistics research, we recommend:

- That the principal professional societies sponsor workshops, in conjunction with their regular meetings, directed to members of academic research-doctorate statistics departments and focused on uses of computers of varied types for research in statistics;
- That the professional societies promote the use of a common network and standard network protocols for communications among research-doctorate statistics departments;
- That the professional societies facilitate electronic communications between their individual members by including network mailing addresses on widely accessible networks (e.g., BITnet, ARPAnet, CSnet) in their membership directories.

Our recommendation for a workshop on uses of computers for statistics research is motivated by the need to provide instruction to new and potential users of departmental computing facilities. There can be a great deal of inertia to overcome in order to get reseachers to use new computational techniques in their work. The inertia is natural since it takes a lot of one's time to learn what the capabilities of a system are and how to use those capabilities to one's advantage. A workshop would provide the opportunity for live demonstrations of how systems ranging in sophistication and complexity from personal computers, through multiuser time-shared systems and workstations, to special purpose hardware for high interaction graphics

and large scale computing are actually used as research tools.

The professional societies have traditionally fostered communications among their members. In the interest of initiating and improving computer communications between all statistics departments, the professional organizations should promote standardization and commonality of interuniversity networking. CSnet, linking departments of Computer Science, is a prime example of an effective communication network linking departments with common interests.

6.4 Recommendations to Research Sponsors

The costs of equipping a department with a minimal configuration of hardware and software are substantial when measured on the same modest scale as the kinds of budgetary line items to which statistics departments are accustomed. Special funding programs of Federal agencies and cooperative arrangements with computer manufacturers have been vitally important in enabling recent expansions of computing facilities by leading research departments. The sustained support of Federal funding agencies and industry will be needed both to build a broader base of resources for computing in statistics departments and to help with the ongoing costs of maintaining the required research facilities. To provide the necessary basis of support, the panel recommends:

- That equipment funding programs such as Scientific Computing Research Equipment in the Mathematical Sciences sponsored by the Special Projects Program, Division of Mathematical Sciences at the National Science Foundation and the University Research Instrumentation Program of the Department of Defense basic research offices be maintained;
- That the National Science Foundation support a research initiative in Interactive Computing Environments for Scientific Research to bring together the expertise of statisticians, computer scientists, mathematicians, and researchers from the physical sciences, life sciences, and quantitative social sciences who use computational methods to advance their science;
- That funding policies of the Statistics Programs within the Federal agencies recognize the need for research support at the project level when considering requests of individual researchers for equipment, research staff, and other on-going costs associated with operation of a computing facility;
- That the equipment funding programs (SCREMS, DURIP) continue to address the needs for special purpose computers and graphics devices of the research groups at the leading edge of statistical/computational theory and methods;
- That computer hardware and software manufacturers continue to expand their sponsorship of research through cooperative arrangements with research-doctorate statistics departments.

Computation, especially large scale computation, is recognized as having a significant impact on methods of inquiry in basic scientific and engineering research. The increased use of computing has in turn increased the importance of developing methods for interpretation of large volumes of data, including high-dimensional data, and for succinct presentation of the analyses. The increased use of computational methods has also accentuated the need for interactive computing environments where it is easy for the user to interact with graphical displays. Ideally such interaction will include two-way communication with a display, permitting the user to input instructions to the system in graphical languages—above and beyond the use of the graphical display as an output device. Statisticians have been leaders in the development of some of these techniques, as described in Section 3.5.

We believe that statisticians can make significant contributions to satisfying the needs for computing environments that are conducive to interpretation in graphical terms and to succinct presentation of analyses for the large volumes of high-dimensional data produced from computationally intensive methods of scientific research. An interdisciplinary initiative to develop better interactive computing environments will benefit all of the sciences that are making advances using computational methods of inquiry.

Additionally, statisticians are making contributions to interdisciplinary science that extend beyond methodology. Specific examples in cognitive science and image processing are discussed in Section 3.5. An interdisciplinary initiative will provide both statisticians and others an opportunity for further cross-fertilization of ideas.

Cooperative arrangements with hardware and software manufacturers have been critically important to leading statistics research departments during their recent expansion of computing facilities. Sometimes these agreements have provided direct benefit to the manufacturers: occasionally, in terms of a marketable product; often, in terms of an idea that needs further development to become marketable. Most often there is no direct benefit to the manufacturer other than in tax benefits. The indirect benefits are well known; the need for continuing industrial support of basic scientific research in the United States is obvious.

APPENDIX I. SURVEY RESPONSES OUTSIDE THE CORE POPULATION

I.1 Biostatistics

The most commonly mentioned research use of computing for these 13 respondents was analysis of

data from clinical trials or epidemiological studies. The responses reflect a higher degree of administrative organization than in the core population of statistics respondents. *Every* biostatistics group has nonstudent staff, 10 have faculty time assigned to computing support, and most were able to provide a detailed budget breakdown by categories. Several groups provide computing services on a large scale. For example, the Johns Hopkins University Department of Biostatistics operates an Academic Data Center for the university. There is therefore great variation among the responses in such items as equipment inventory and budget. Eight of the respondents operate multiuser systems, and the 5-number summary of departmental budget costs is (in thousands of dollars): 2, 32, 76, 117, and 245. The National Institutes of Health were mentioned as supporting agencies by 10 of the 13 respondents. None mentioned NSF or DoD agencies, except those sharing equipment with statistics departments at the same institution.

The biostatistics community is well ahead of statistics departments in providing staff support and in careful accounting of the costs of maintaining computing facilities. It may be behind in communications (only 4 have access to national networks, 3 of these to BITnet) and in interactive graphics (only 4 have graphics terminals or workstations). Expressed needs also differ. Biostatistics groups do want graphics (4 mentioned this), but also better database management systems (3 mentions) and more CPU power (3 mentions). One mentioned special hardware for studies in biomedical signal analysis and image processing. The ranked priorities responses are given in Table I.1.

I.2 Mathematical Sciences

These 20 respondents represent a varied population. They award degrees in statistics, and so are located at institutions without separate departments of statistics. Statisticians are generally a small fraction of the faculty. Both research uses of computing and computing facilities are (with several notable exceptions) less developed in these departments than in the statistics and biostatistics departments. Only 2 operate multiuser systems and 5 have workstations. Eight have graphics terminals, and 4 mentioned graphics as an unfilled need. Only 1 has nonstudent staff assigned to computing, while 5 have faculty who devote at least 5% of their time to support of computing. Four have access to national networks (3 to BITnet). Three have access to super computer facilities, and 2 more will soon have such access.

Some respondents in this group felt no need for computing beyond the resources of university central systems and a few microcomputers. Others expressed strong needs that are reflected in the fact that 11 of the 20 ranked hardware acquisition as their first priority. As might be expected in the circumstances, few respondents cited external sources of funding for

TABLE I.1
Departmental priorities: Biostatistics

	Rank								Mean
	1	2	3	4	5	6	7	NR[a]	
Hardware acquisition	6	2	3	1	1	0	0	0	2.15
Software acquisition	4	5	1	1	1	0	0	1	2.17
Support personnel	1	1	5	0	2	1	0	3	3.40
Student research	0	3	0	6	0	0	3	1	4.25
Faculty release time	2	0	2	0	2	2	3	2	4.64
Network access	0	2	1	0	1	5	2	2	5.09
Operation, maintenance	0	0	1	2	3	2	2	3	5.20

[a] NR, no response.

TABLE I.2
Departmental priorities: Mathematical Sciences

	Rank								Mean
	1	2	3	4	5	6	7	NR[a]	
Hardware acquisition	11	3	0	1	0	1	0	4	1.69
Software acquisition	1	7	4	0	1	1	0	6	2.71
Operation, maintenance	2	1	4	3	1	1	1	7	3.54
Faculty release time	0	2	4	3	2	1	0	8	3.67
Support personnel	1	0	2	4	2	3	1	7	4.46
Network access	1	1	2	0	3	1	5	7	5.00
Student research	0	1	0	2	2	3	4	8	5.50

[a] NR, no response.

research computing. Three mention NSF, 1 ONR, and 2 credit other agencies. The ranked priorities responses are given in Table I.2.

I.3 Other Departments

This group contains small United States statistics departments, Canadian statistics departments, and statistics groups in management and behavioral sciences. The 13 responses were for the most part incomplete (every category in the priorities ranking item had a majority of nonresponse). These groups rely almost entirely on central facilities and microcomputers (11 of the 13 have microcomputers). The only clear pattern is need for both hardware and software (6 ranked each of hardware and software acquisition as priority 1 or 2). In view of the diverse population and low response, further analysis is not appropriate.

APPENDIX II. SOME EXAMPLE CONFIGURATIONS

II.1 Department of Statistics: University of Chicago

Population Served. Approximately 12 faculty members and 15 graduate students.

Hardware Configuration. Three Sun 2/120 workstations connected via Ethernet. One of the workstations ("galton") is configured to be a file server, managing 260 MB of disk storage for itself and the other two stations ("karl" and "egon"). Each machine has 2 MB of random access memory.

Karl and egon are standard Sun workstations. Neither has disk of its own; rather each relies on file service via galton for disk storage. The transmission speed of the Ethernet is sufficiently rapid that there is no noticeable delay associated with disk access. In addition, egon has been supplied with a medium resolution color monitor (and associated controlling hardware), as well as a floating-point accelerator.

Galton does not have graphics capability and plays much the same departmental role that a standard minicomputer such as a VAX 11/750 would. Attached to galton are 16 serial ports for terminals and telecommunications, and a ¼-inch streaming tape drive for disk backup. In addition, galton drives an Imagen 8/300 laser printer and has three dial-up telephone lines attached to three of the ports. One of the ports is used for a 4800-baud dedicated line to an IBM PC/XT, which is also used for data entry and file transfer. Archives of programs and manuscripts are generally kept on floppy disks created by transfering files from galton to the PC/XT. The department owns approximately six display terminals (Wyse-50 and Wyse-75), a Tektronix 4013 graphics terminal with hard copy unit, and several hard copy terminals. The department also owns two Macintosh computers which are used both as terminals and graphics devices (with hard copy) for galton. Several display terminals reside in faculty offices, where they communicate with galton at 4800 baud using the University's digital telecommunications network. This requires a device in each faculty office whose cost is approximately $1200; this device includes simultaneous voice and data transmission, and can also be used in place of a modem for data communications outside the University.

Software. The network runs Berkeley 4.2bsd UNIX, as supplied by Sun Microsystems. This distribution includes eqn/troff typesetting software. The printer software is the UNIX software supplied by Imagen with essentially no modifications. Both Kermit and macput/macget file transfer programs have been installed; each is available in the public domain. Linpack and Eispack, public domain mathematical software libraries for numerical linear algebra and eigen analysis, respectively, have also been installed. A version of TEX is available but it is not yet fully operational due to the unavailability of 300 dot per inch fonts. Versions of GLIM and PLOT-10 (the latter is a library of Fortran-callable graphics subroutines available from Tektronix) are available. Within the next 6 months installation of a version of LISP (either Portable Standard Lisp from Utah, or Franz Lisp from Franz, Inc.), the S language, and EMACS (a powerful text editor) is planned. Additionally, a symbolic algebra package such as REDUCE or MACSYMA may be obtained.

Networking. The three workstations comprise a single local area network based on a 10 Mb Ethernet running TCP/IP software. Within 3 months this departmental network will be attached to a campus-wide 10 Mb Ethernet, using galton as a gateway. The University of Chicago Ethernet has adopted TCP/IP as a standard. Through the campus Ethernet, the Department of Statistics will have direct access to USEnet, CSnet, and BITnet. Until the department is connected to the campus Ethernet, outside communication is achieved through a telecommunications (UUCP) link to a machine operated by the Computer Science Department, through which the campus network can be reached.

Cost. The hardware listed above, with the exception of the hard copy terminals, the Tektronix graphics equipment, and about 25% of the telecommunications equipment, has been obtained within the last year, at a total cost of approximately $135,000. Of this amount, approximately $20,000 involved site preparation costs. The three workstations and their peripherals cost approximately $80,000, after academic discounts. The laser printer cost approximately $10,000. The remaining $25,000 accounts for terminals, telecommunica-

tions equipment, the 16-port multiplexor, archive tapes, extra cabling, and microcomputers.

Technical Support. One faculty member spends approximately 25% time in providing software and hardware support. This arrangement has proven unsatisfactory. Originally conceived as being a temporary solution during the initial months of the facility's existence, it has proven difficult to find and to fund a suitable staff person.

Maintenance. The only maintenance contract is for the firmware for the laser printer. The experience of other Sun owners on campus allowed assessment of the relative merits of maintenance contracts versus return to factory on a time and materials basis; the latter decision was taken.

Adequacy. The hardware configuration described above has proven to be adequate for current needs, although this is based on only a half year of experience. More of the purely text entry tasks will be moved to off-line microcomputers in the future, giving some effective expansion of capacity.

II.2 Department of Statistics: Carnegie-Mellon University

Population Served. Approximately 15 faculty members, 6 secretarial and editorial staff, and 30 graduate students.

Hardware Configuration. One VAX 11/750 super minicomputer, 6 VAXstation II workstations, 6 VAXstation I workstations, 1 VAXstation 500 color workstation, 2 Sun 2/120 workstations, 4 IBM PC/XT personal computers, 1 Macintosh personal computer, 1 CSPI Mini-Map array processor, 5 GIGI color microcomputers, 1 QMS Lasergraphics 1200 graphics laser printer, 2 other printers, 18 terminals of various kinds, and 1 HP 7470A pen plotter.

The VAX 11/750 has 4 MB of random access memory, 912 MB of disk storage, 25 9600-baud terminal ports, 1 800/1600 bpi tape drive, 3 separate network interfaces, and a floating-point accelerator. Eight of the VAX 11/750 terminal ports are connected to the University wide Micom switching system which is connected to all faculty, staff, and graduate student offices; two ports are reserved for system support activities; three ports are connected to staff offices; two ports are connected to the graduate statistics research facility (GSRF); two ports are connected to line printers. To restrict access to the VAX 11/750 eight terminal ports are currently unused.

The VAX 11/750 is physically located in a Psychology Department machine room which was renovated with about $85,000 of Statistics Department funds. Five of the VAXstation II and four of the VAXstation I are located in faculty offices and one VAXstation II and two VAXstation I are in the GSRF. The VAXstation I is a workstation with computing power roughly equivalent to a VAX 11/730, a high resolution raster graphics display screen, a 3-button mouse, 2 MB of random access memory, 31 MB of disk storage, two 5-inch floppy disk drives, and an interface to a 10 Mb/sec Ethernet to the VAX 11/750 and the other VAXstations. The VAXstation II is equivalent in every respect to the VAXstation I except it has processing power approaching that of a VAX 11/780.

The VAXstation 500 (a relabeled Tektronix 4125) is a high resolution color workstation with an Intel 80286/80287 processor. It is located in the GSRF and has Micom access. One of the Sun workstations is located in the GSRF and one in a faculty office. These workstations are part of a joint CMU-IBM network development project and run a special version of the 4.2 bsd UNIX and a special window manager, ANDREW. The CSPI array processor is an attached processor to the VAX 11/750 and has 1.25 MB of random access memory. Its CPU is roughly twice as fast as an IBM 3083 but is somewhat more difficult to use.

The QMS Lasergraphics 1200 is a 300 pixel per inch xerographic laser-driven printer with a Motorola 68000 processor and 1 MB of random access memory. The screens of the VAXstations can be dumped directly to this printer.

Software. All of the VAXes run VAX/VMS, Versions 4.1. The standard editor is EMACS, a full screen multiwindow editor. Document production is handled by SCRIBE, a device-independent formatting system similar to TEX with output to any printer including the Xerox 9700 in the central computer center. Statistical packages include SAS, Minitab, ISP, SCA, GLIM, and RS/1. Languages include Fortran, C, Franzlisp, and Simscript II. Subroutine libraries include IMSL, LINPACK, and EISPACK. The VAXstations have GKS graphics software; the VAXstation 500 has PLOT10.

Networking. All of these machines are interconnected by a 10 Mb Ethernet using DECnet and TCP/IP software. The machines are part of CCnet, a very large local area network providing direct access to all of the machines on it including the DEC 2060s, the VAXes, and the IBM 3083 in the University Computing Center. CCnet extends beyond the CMU campus to include machines at Columbia University, Stevens Institute of Technology, Vassar College, New York University, Case-Western Reserve University, and the University of Pittsburgh. Additionally, one of the network nodes is the Westinghouse Center for Advanced Computation in the Engineering Sciences which is expected, beginning 1986, to house the CRAY 1-S supercomputer awarded in June 1985 to the CMU-Pitt-Westinghouse consortium.

Cost. The equipment listed above has been obtained with the help of grants from NSF (SCREMS) and DoD (DURIP). Total grant funds for hardware acqui-

sition over the last 4 years total about $202,000. The Carnegie-Mellon central administration has contributed approximately $57,000 for hardware acquisition. The department has contributed approximately $40,000. IBM donated about $25,000 worth of cables and the cost of the two Sun workstations. DEC has donated roughly $175,000 in discounts. Other vendors have donated an additional $25,000 in discounts. Additionally the central administration has contributed about $100,000 in site preparation expenses.

Technical Support. One faculty member spends about 25% effort on support of these facilities with the aid of several graduate students. This level of support is inadequate.

Maintenance. The general strategy has been to buy maintenance contracts initially and then as reliability is learned, contracts are cut back where they are not cost-effective. Current maintenance expenses total about $2500 per month. Annual software license fees are about $3000.

Adequacy. General experience at Carnegie-Mellon suggests that it takes about 2 years to fully assimilate (i.e., saturate) new computer hardware. Statistics Department experience is similar. They have moved from general availability of terminals in 1981, to a VAX 11/750 in 1983, to several workstations in 1985, to (they expect) general availability of workstations in 1987.

II.3 Department of Statistics: Purdue University

Population Served. Within the Statistics Department: 15 faculty members, 3 secretaries, and about 30 graduate students. Also some members of the Mathematics Department: about 5 faculty, 3 secretaries, and 5 graduate students.

Hardware Configuration. One VAX 11/780 super minicomputer with 4 MB of memory, floating-point accelerator, 512 MB of disk space, a tape drive, and 40 9600-baud terminal ports. Two of the ports are used for printers: an Imagen 8/300 laser printer and a DEC LA-100 line printer. Twenty-six ports are hardwired to faculty (17) and secretarial (4) offices and to terminal rooms (5) available for graduate students and staff. Five ports are connected to the Purdue University Computer Center (PUCC) Serial Data Switch (SDS); this allows use of the VAX from up to five non-hard-wired terminals in various places on the campus (including 3 Statistics and Math faculty offices), as well as dial up connections. Of the remaining 7 ports, 1 is used for networking, 1 for operations, and 5 are currently unused.

They have 29 nongraphics terminals: half of them are ADDS Viewpoint A-2, the others are Wyse 50, Zenith Z-29, DEC VT-100, and ADDS Regent 20. Two graphics terminals (Visual 550 with 768 × 585 resolution) are on order. They also have 6 HP7470A pen plotters which are connected between terminals and computer, using "eavesdrop" cables.

A 512 KB Macintosh microcomputer with a printer is on order; and an IBM PC/AT will soon be available on loan.

Software. The operating system is Berkeley Unix 4.2 which includes Fortran, Pascal, C, and Franz LISP languages, and the eqn/troff typesetting programs. The S and GLIM statistical packages, VAXIMA (a version of MACSYMA), and the TEX typesetting program have all been installed. Kermit is available for file transfer from microcomputers via modem connections.

Networking. Locally, the VAX is linked to the other VAX 11/780s on the campus: at PUCC, in the Computer Science Department, and on the Engineering Computer Network (ECN). The VAX 11/780s are currently linked to a portion of the PUCC central system (CDC 6600s) for submission and retrieval of jobs which are executed in batch mode; connections to PUCCs IBM mainframe and CYBER 205 supercomputer are expected soon. Externally, ECN provides access to USEnet, and indirect access to ARPAnet.

Cost. The equipment described above cost about $200,000, of which NSF and ONR grants provided about $150,000, and Purdue University provided the rest. Site preparation costs were minimal (under $2,000) because the VAX was located within PUCC's facility.

Technical Support. One faculty member spends about 20% effort on support and coordination, with no reduction in other duties. A graduate student spends quartertime as a consultant and applications programmer. There is a great need for a systems programmer. PUCC, which maintains the system, including tape backups, does not provide programming or software support, other than what is common to both the Statistics VAX and theirs.

Maintenance. The Statistics Department pays (with some help from the Math Department) about $2000 per month to PUCC for operation and maintenance, which includes repairs on some of the hardware. (PUCC, which maintains a number of VAX 11/780s has DEC-trained technicians and maintains an inventory of parts.) There is a separate maintenance contract on the laser printer.

Adequacy. The equipment meets current needs quite well, although they are just now beginning to feel the pinch in disk space.

II.4 Department of Computer Science and Statistics: University of Georgia

Population Served. Approximately 22 faculty members and 25 graduate students.

Hardware Configuration. One VAX 11/750 super

minicomputer, 27 PRO 380 microcomputers, 23 LA-50 dot matrix printers, 8 Hazeltine terminals, 1 LA-210 printer, 1 LXY-12 printer/plotter, and 1 Tektronix 4006 graphics terminal.

The VAX has 2 MB of memory and 456 MB disk. It has 32 terminal ports which are connected to the PRO 380s and a 800/1600 bpi tape drive.

The PRO 380s have 512 KB of memory and 10 MB of disk storage. Twenty-three of the PRO 380s are located in faculty offices, 2 are located in a computing lab, and 3 are reserved for system support activities.

Software. The VAX runs the Ultrix-32 operating system (a variant of Unix marketed by DEC); conversion to AT&T System V Unix is under consideration. The statistical package S is available. The PRO 380s run the Venix operating system (another Unix variant).

Networking. None presently. Plans are being made to connect the 4 Unix VAXes on campus.

Cost. The total cost of the equipment was about $91,000 and site preparation expenses for the VAX were another $11,000. The funds came from NSF and DoD equipment grants together with support from the Digital Equipment Corporation PACE program.

Technical Support. One full-time staff member and 3 part-time undergraduates provide all system support.

Maintenance. The VAX is maintained under a standard contract with DEC. There is currently no maintenance contract for the PRO 380s.

Adequacy. This equipment has just been installed and they have not yet learned to fully utilize it.

APPENDIX III. COST RECOVERY

Whether or not a research group accounts for use of its computer facilities and, if it does account for them, how it is done depends on a variety of conditions. These conditions include the size of the research group, the size (and hence cost) of acquisition and operation of the computer facility, the sources of the funds available to pay the costs, whether or not the group performs its own facilities management, etc. We have identified three general approaches to cost recovery and describe each one briefly below.

III.1 No Direct Cost Recovery

In very small research groups, there is no need to perform any accounting. This is the case when there is no opportunity to recover funds from external grants and contracts. For example, if the facility consists of only one principal investigator with a single Federal research grant who purchases and uses a single computer system there is obviously no need for an accounting system.

III.2 Access Charges

Under this method of cost recovery each user account is charged a fee for access to all the facilities. This method is believed to comply with OMB Circular A-21 (as amended) and involves minimal administrative overhead. The major drawback of this method is its lack of incentive for users to limit their consumption of resources. The amount of the access fee is obviously determined by the cost of operation; in the Department of Statistics at Carnegie-Mellon University, for example, this fee is currently set at $150 per month. The department operating budget pays the access charge for those accounts which are not supported by external funds.

III.3 Resource Usage Charges

This is the standard method used by large computer centers. It requires accounting software to determine resource utilization. It also requires more administration than the other methods. Rates must be determined for each resource based on the cost of supplying that resource. In actual application, the rates a department will charge are likely to be one to two orders of magnitude smaller than the rates charged by its central computer center for the same resource. This is because the cost of the resource does not (necessarily) have to be amortized over its lifetime and because the level of user support in a department facility will be significantly smaller than in a central computer center.

ACKNOWLEDGMENTS

The Workshop on the Use of Computers in Statistical Research was supported by Contract N00014-85-G-0181 from the Mathematical Sciences Division of the Office of Naval Research to the Institute of Mathematical Statistics. The Workshop was planned and organized by the Statistical Policy Committee comprising Morris H. DeGroot, Carnegie-Mellon University; Donald J. Geman, University of Massachusetts; Samuel W. Greenhouse, George Washington University; Frederick Mosteller, Harvard University; David S. Moore, Purdue University; Ingram Olkin, Stanford University (co-chairman); Ronald Pyke, University of Washington (co-chairman); and Jerome Sacks, University of Illinois.

REFERENCES

DAVID, E. E., JR. (Chairman). (1984). *Renewing U.S. Mathematics, Critical Resource for the Future.* Report of the Ad Hoc Committee on Resources for the Mathematical Sciences. National Academy Press, Washington, D. C.

LAX, P. D. (Chairman). (1982). *Report of the Panel on Large-scale Computing in Science and Engineering.* National Science Foundation, Washington, D. C.

RHEINBOLDT, W. C. (Chairman). (1985). *Future Directions in Computational Mathematics, Algorithms, and Scientific Software.* Report of the Panel on Future Directions in Computational Mathematics, Algorithms, and Scientific Software. SIAM, Philadelphia.

Comment

Jessica Utts

My discussion will be divided into two parts. The first part consists of a treatise on the responsibility which accompanies the use of computers in statistical research. I offer several recommendations to complement those in the article.

The second part is a short description of a setup which works fairly well at the University of California at Davis and was not mentioned in the report. It might be of interest to other statistics departments.

1. SCIENCE FICTION OR FUTURE FACT?

There has been a science fiction novel living in my head for the past 10 years or so. It started when I was a graduate student studying robustness and I realized that most users would think of the computational aspects of robust procedures as a black box. This story occurs 30 to 40 years in the future. There are no more statisticians. There are statistical clerks, and every university department has at least one. Research is done by collecting data and giving it to the statistical clerk, who takes it from there. The clerk feeds the data into the computer and out pops the appropriate model, estimate, or whatever, complete with the associated significance or confidence levels. These are sent to journals, along with a post hoc explanation for the results of any of the tests which turned out to be "significant." Everyone is quite happy with this arrangement. No one knows how the computer generates these answers, but everyone knows that if the computer produced them, they must be right. All sorts of interesting (and not so interesting) hypotheses are being proved this way, and when they don't agree with common sense, everyone knows that common sense must be wrong.

In the current version of the story, something finally goes wrong. I haven't worked out the details, but it is a result which contradicts common sense so much that someone (a fresh young scientist, of course) actually has the audacity to question what is happening in the computer. In order to determine what the computer should be doing, a team of scholars attempts to decipher the statistical literature. To their dismay, they find that the literature is unreadable to them. Finally, they locate a few old statisticians who have long since retired, and with their help they piece together the story. It seems that when the computer software was being developed, most statisticians didn't pay much attention. The packages which were eventually implemented were written by people who were good at selling, but who didn't really understand the concepts involved. A few statisticians tried to protest, but since they were advocating the use of their own services, no one took them seriously. After all, the journals were much more likely to publish the computerized version of the results, so why bother with the more cautious and complicated interpretations the statisticians were trying to sell?

Of course I will never write this novel, but if things continue on their present course I may very well watch it unfold from science fiction into future fact. There are even those who believe that it is already well under way. One of our graduate students told me that a recent cocktail party response to his statement that he was studying statistics was "aren't you afraid of being replaced by a computer?"

So am I against the use of computers in statistical research? Of course not. In fact, I embrace these developments. After all, the world is a complex, nonnormal, non independent and identically distributed place and complex models are much more likely to accurately describe reality. Tools like the bootstrap, high resolution graphics, and interactive data analysis programs are important and useful for applied statisticians.

What I advocate is that we as research statisticians begin to play a greater role in determining that our work is properly applied. Our techniques are simultaneously becoming more complex and more automated. They are less and less likely to be understood by nonstatisticians. I was concerned when people started using calculators which give regression coefficients without producing plots. But the potential for misuse

Jessica Utts is Associate Professor, Division of Statistics, University of California, Davis, California 95616.