The selection of statistical software for use on microcomputers requires a thorough definition of the goal, plus knowledge of the capabilities of the user, the hardware and the software. These factors form a complex, interacting system which is limited by its weakest component. This paper suggests criteria to be used in the selection of software and hardware for statistical applications. Additional sources of information and several examples of statistical software are cited.

Introduction

The recent proliferation of microcomputers and applications software has put computational power in the hands of general practitioners and specialists of many fields. This paper examines the requirements of engineers and scientists who wish to analyze and interpret their data using statistical programs on microcomputers. The perspective is that of someone who has had limited exposure to statistics or computer science, but wishes to employ the procedures of both fields to solve scientific or technical problems.

The system to be examined has four components: the goal, the user, the hardware and the software. The first component, the goal or desired product, will not be realized until the other three components have been acquired, integrated and put into service. The user or responsible manager must have a clear idea of the objectives and the type of product desired in order to select the three resource components. The user must be cognizant of his or her skill levels pertaining to computers and statistics. The type of computer and its operating system will limit the type of programs, the speed of computation, the precision of the results and the type of hardcopy product available. Software is available from several sources with a variety of capabilities, features, hardware requirements and assumptions regarding the skill levels of users. A satisfactory balance must be achieved in this four component system.

* References to specific hardware and software types and product names are for illustrative purposes only and do not constitute endorsement or recommendation. Any emphasis placed on specific systems reflects the author's experience and/or perception of current market conditions.
The Four Components

The Goals

The first step is to define the problem. What are the source, format and extent of the data to be analyzed? What analyses are to be performed? What type of report is expected?

Data may consist of small sets of observations taken, for example, from a series of laboratory experiments. Such data sets may be entered into a program from the keyboard with relative ease and do not require extensive computer memory. Hence, demands on the hardware and software are not severe and file formats are not a major consideration. Extensive data bases may be obtained from on-line instrumentation, data base programs or other sources on a variety of machines. The amount of memory in the computer and the capability of the software to handle large files and import files written in various formats become important for large data sets.

A wide variety of statistical procedures have been programmed in various statistical packages. Typically, one begins by employing exploratory data analysis procedures to gain insight into the nature and structure of the data and to search for possible relationships: appropriate procedures are graphical as well as numerical. Formal statistical analyses include the testing of various statistical hypotheses, such as a comparison of samples means, the estimation of model parameters and testing of the adequacy of models. The design of experiments, the structure of the resulting data set and the appropriate statistical methods are closely related. The number and type of dependent and independent variables in the data set will determine the statistical methods to be used. For example, specific methods are applicable if the data set is a time series.

Reporting may consist of any type of output, from numbers displayed on the computer console to elaborate printed reports complete with multi-colour graphics. The type of reporting should be appropriate considering the end use of the results and the audience for whom the report is intended. The user's requirements for reporting will influence not only the hardware to be acquired but also the capabilities of the software, including its device drivers, editing capabilities, graphics capabilities and output format options.

The User

The user should be familiar with the computer to be employed, including both the hardware and the operating system. Knowledge of the ways in which the various components communicate with each other is necessary in order to configure the basic system and the applications packages to be used on the system. Methods of allocating and accessing memory space are also important, particularly considering the large size of many statistics packages.

The user should also be familiar with statistical procedures used in the software package. One potential disadvantage of statistical software is that it facilitates misuse