CLINFO USER'S GUIDE: RELEASE THREE

PREPARED FOR THE NATIONAL INSTITUTES OF HEALTH

G. F. GRONER,
N. A. PALLEY, M. D. HOPWOOD,
W. L. SIBLEY, B. FISHMAN

R-1543-3-NIH
SEPTEMBER 1977
This research was supported by the National Institutes of Health under Contract No. NO1-RR-5-2111. Views or conclusions expressed herein should not be interpreted as representing the official opinion or policy of the National Institutes of Health.
CLINFO USER’S GUIDE:
RELEASE THREE

PREPARED FOR THE NATIONAL INSTITUTES OF HEALTH

G. F. GRONER,
N. A. PALLEY, M. D. HOPWOOD,
W. L. SIBLEY, B. FISHMAN

R-1543-3-NIH
SEPTEMBER 1977

Rand
SANTA MONICA, CA 90406
PREFACE

This manual provides most of the information needed to enter, retrieve, display, and analyze clinical research data with the CLINFO prototype system. It replaces Release 2 of the CLINFO User’s Guide (which should be discarded) and includes descriptions of many new facilities and options. It also describes Release 3, the final version of the prototype.

The prototype described in this report has been developed as part of the CLINFO project, a scientific inquiry sponsored by the Division of Research Resources (DRR) of the National Institutes of Health (NIH). The goals of the project are to identify and characterize the information-analytic tasks and the information flows in clinical research, and to develop methods for facilitating these tasks and flows.

As implemented, the prototype provides data management and analysis capability for clinical researchers. In addition, it collects data that are used to evaluate its effectiveness in clinical research centers, where it is being implemented on an experimental basis. This release reflects many changes and improvements suggested by users of the system in earlier evaluations. It incorporates suggestions from these continuing evaluations to provide a set of functional specifications for a data processing tool that can be used throughout the clinical research community.

The CLINFO project is being conducted by clinical investigators at the Baylor College of Medicine, the University of Washington, and Vanderbilt University; by information scientists at The Rand Corporation; and by staff members of the DRR. A clinical investigator at the University of Oklahoma participated previously. The project has thus far

- Broadly characterized the information-related activities of the clinical research process.
- Identified research data management and analysis as major problems in clinical research.
- Carried out an extensive survey among clinical investigators which indicated that these problems are widespread.
- Examined a number of existing systems aimed at alleviating these problems.
- Prepared an initial functional description of a prototype data management and analysis system.
- Implemented this prototype in three clinical research centers, using standard minicomputer hardware, an operating system supplied by the manufacturer, and specialized computer programs written in the BASIC language.
- Evaluated the effectiveness of the prototype and developed functional specifications and diffusion plans for a widely available CLINFO system.

Project results to date are described in a number of articles and in the following Rand Corporation reports:


CLINFO prototypes are currently operating experimentally at three clinical research sites that have established suitable physical and scientific environments in cooperation with the CLINFO project investigators. The first prototype has been operating at the Baylor College of Medicine in Houston, Texas, since January 1976. The CLINFO project director there is Dr. Howard K. Thompson, Jr., and the system manager is John C. Mabry. The second prototype has been at the University of Washington School of Medicine in Seattle, Washington, since September 1976. T. Graham Christopher, M.D., is the CLINFO project director, and Ronald O. Ling is the system manager. The third prototype has been operating at the Vanderbilt University School of Medicine, Nashville, Tennessee, since February 1977. William W. Lacy, M.D., is the CLINFO project director, and Joel L. O'Connor is the system manager.

This report is primarily intended for those who will be actively employing the CLINFO prototype in their clinical research activities, but it should also be of interest to members of the NIH staff, to information scientists concerned with the information processing needs of clinical investigators, and to all clinical investigators interested in data management aids.

As CLINFO is an experiment in its own right, users are encouraged to be innovative and to attempt operations and manipulations which may not be explicitly described in the CLINFO documentation. We hope that the results of these innovative pursuits, both successes and failures, will be reported to us; such information provides insights into the needs of the system and can be of considerable value to the CLINFO project investigators.
ACKNOWLEDGMENTS

Although the authors accept responsibility for the contents of this report, many others deserve credit for their contributions to the CLINFO Prototype Data Management and Analysis System. Our former colleagues Steven S. Zucker and William H. Josephs participated in the system design and implementation. Marie Hall and JoAnn Lockett have previously been responsible for maintaining the computer hardware upon which we develop, test, and use the CLINFO software; Ms. Lockett has made other contributions as well. Dina K. Dickerson updated portions of the guide to help create this version.

Our co-contractors, Howard K. Thompson, Jr., of the Baylor College of Medicine (the first CLINFO site), T. Graham Christopher, of the University of Washington (the second CLINFO site), and Arthur W. Nunnery, of the University of Oklahoma, deserve special credit. They have collaborated with us since the beginning of the CLINFO project in July 1972, educating us about clinical research, participating in developing the project’s research plan, ensuring that the design of the prototype system meets the needs of a substantial number of clinical investigators, and evaluating the prototype itself. John C. Mabry, of the Baylor College of Medicine, Ronald O. Ling, of the University of Washington, and William W. Lacy and Joel L. O’Connor, of Vanderbilt University, have also influenced the system design. The users of the prototype systems in Houston, Seattle and Nashville have been most helpful in uncovering deficiencies and problems and in keeping our attention focused on the practical needs of the clinical investigator.

The CLINFO project and system could not be successful without continued guidance, encouragement, and support from several staff members of the National Institutes of Health and from the group of NIH advisors for the CLINFO project. We especially want to thank our project officer, William R. Baker, Jr., whose foresight and persistence have made our endeavors both exciting and possible.

Our colleague Thomas L. Lincoln deserves credit for having educated all of us at Rand over the past several years about the need for such a system and for using the prototype, as we develop it, extensively and in many stressful and novel ways.

Finally, the authors acknowledge the assistance of Jacqueline Bowens, the project secretary, who typed successive drafts of this report, using the WYLBUR text editor. Ms. Bowens has also helped to operate the CLINFO computer hardware at Rand.
CONTENTS

PREFACE .......................................................... iii

ACKNOWLEDGMENTS ............................................ v

Section

I. INTRODUCTION ................................................ 1
   What CLINFO Does ............................................. 1
   CLINFO Terminology .......................................... 3
   Overview of CLINFO Operation ............................. 5
   Use of this Manual ............................................ 14

II. USING THE SYSTEM ........................................... 17
    The Essentials ................................................. 17
    CLINFO Activities ............................................ 24
    Data Organization and Files ............................... 38
    The CLINFO User’s Terminal ............................... 40

III. EXAMPLE OF A CLINFO SESSION .......................... 43

IV. HELPFUL HINTS .............................................. 59

V. REFERENCE MANUAL .......................................... 69
   Analyze ......................................................... 71
      Option 0: Select a Worksheet ............................ 72
      Option 1: Descriptive Statistics ....................... 72
      Option 2: T-Test ............................................. 76
      Option 3: Chi-Square Test ................................. 81
      Option 4: Linear Regression ............................. 85
      Option 5: ANOVA ............................................. 91
      Option 6: Cross Tab ......................................... 94
      Option 7: Scatter Plot ..................................... 99
      Option 8: Histogram ....................................... 106
      Option 9: Frequency Distribution ...................... 111
      Option 10: Normality Test ................................ 115
      Option 11: Non-Parametric Tests ......................... 118
      Option 12: Life Table Analysis .......................... 123
      Option 13: Radioimmunoassay ............................. 130
   Calculate ...................................................... 142
      Option 0: Select a Worksheet ............................ 146
      Option 1: Arithmetic Expression ........................ 147
      Option 2: Logical Expressions (AND) .................. 151
      Option 3: Logical Expressions (OR) .................... 154
      Option 4: Special Calculations .......................... 155
      Option 5: Time Difference ................................ 158
      Option 6: Statistical Calculations ..................... 161

vii
<table>
<thead>
<tr>
<th>Command</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Describe</td>
<td>165</td>
</tr>
<tr>
<td>Option 1: Edit the Schema</td>
<td>168</td>
</tr>
<tr>
<td>Option 2: Print a Formatted Schema</td>
<td>178</td>
</tr>
<tr>
<td>Option 3: Create/Delete Study Files</td>
<td>180</td>
</tr>
<tr>
<td>Enter</td>
<td>182</td>
</tr>
<tr>
<td>Option 0: Add New Patient Abbreviations into the Study</td>
<td>183</td>
</tr>
<tr>
<td>Option 1: Enter Data into the Update File</td>
<td>183</td>
</tr>
<tr>
<td>Option 2: Copy and Screen Data from a Worksheet into the Update File</td>
<td>189</td>
</tr>
<tr>
<td>Option 3: Review Data in the Update File</td>
<td>195</td>
</tr>
<tr>
<td>Examine</td>
<td>199</td>
</tr>
<tr>
<td>Files</td>
<td>203</td>
</tr>
<tr>
<td>Option 1: List Worksheets</td>
<td>204</td>
</tr>
<tr>
<td>Option 2: List Subsets</td>
<td>206</td>
</tr>
<tr>
<td>Option 3: List Response Files</td>
<td>206</td>
</tr>
<tr>
<td>Option 4: List Utility Files</td>
<td>206</td>
</tr>
<tr>
<td>Goodbye</td>
<td>208</td>
</tr>
<tr>
<td>Message</td>
<td>209</td>
</tr>
<tr>
<td>Response Files</td>
<td>210</td>
</tr>
<tr>
<td>Option 1: Execute a Response File</td>
<td>211</td>
</tr>
<tr>
<td>Option 2: Create a Response File</td>
<td>212</td>
</tr>
<tr>
<td>Option 3: Edit a Response File</td>
<td>214</td>
</tr>
<tr>
<td>Option 4: Discard a Response File</td>
<td>219</td>
</tr>
<tr>
<td>Option 5: Copy a Response File</td>
<td>219</td>
</tr>
<tr>
<td>Option 6: Append a Response File to Another</td>
<td>220</td>
</tr>
<tr>
<td>Retrieve</td>
<td>221</td>
</tr>
<tr>
<td>Option 0: Select a Worksheet</td>
<td>222</td>
</tr>
<tr>
<td>Option 1: Event Display</td>
<td>222</td>
</tr>
<tr>
<td>Option 2: Study Data File to Worksheet</td>
<td>223</td>
</tr>
<tr>
<td>Option 3: Study Data File to Terminal/Communication File</td>
<td>241</td>
</tr>
<tr>
<td>Option 4: A Worksheet to a Worksheet</td>
<td>249</td>
</tr>
<tr>
<td>Option 5: Worksheet to Communication File</td>
<td>254</td>
</tr>
<tr>
<td>Option 6: Communication File to Worksheet</td>
<td>257</td>
</tr>
<tr>
<td>Subset</td>
<td>259</td>
</tr>
<tr>
<td>Option 1: A List of Patient Abbreviations</td>
<td>260</td>
</tr>
<tr>
<td>Option 2: Patients Belonging to All of a List of Sets</td>
<td>261</td>
</tr>
<tr>
<td>Option 3: Patients Belonging to at Least One of a List of Sets</td>
<td>262</td>
</tr>
<tr>
<td>Option 4: Patients Whose SDF Data Satisfy All Conditions in a List</td>
<td>263</td>
</tr>
<tr>
<td>Option 5: Patients Whose SDF Data Satisfy Any of the Conditions in a List</td>
<td>266</td>
</tr>
<tr>
<td>Option 6: Patients Whose Worksheet Data Satisfy All Conditions in a List</td>
<td>267</td>
</tr>
<tr>
<td>Option 7: Patients Whose Worksheet Data Satisfy at Least One Condition in a List</td>
<td>270</td>
</tr>
<tr>
<td>Option 8: List Patient Abbreviations from a Set</td>
<td>270</td>
</tr>
<tr>
<td>Option 9: List All of Your Subset Names</td>
<td>271</td>
</tr>
<tr>
<td>Option 10: Discard a Particular Set</td>
<td>272</td>
</tr>
</tbody>
</table>
Worksheet ................................................................. 273
Option 0: Select a Worksheet ........................................... 274
Option 1: Display this Worksheet ...................................... 275
Option 2: Label Worksheet Rows ..................................... 279
Option 3: Label Worksheet Columns ................................. 280
Option 4: Enter Data by Rows ......................................... 280
Option 5: Enter Data by Columns ...................................... 282
Option 6: Enter Data by Cells .......................................... 282
Option 7: Discard this Worksheet ..................................... 283
Option 8: Sort this Worksheet .......................................... 283
Option 9: List Your Worksheets ........................................ 286
Option 10: Print this Worksheet ........................................ 287
Option 11: Edit this Worksheet ......................................... 288

GLOSSARY ................................................................. 295

CLINFO MENUS ......................................................... inside back cover
I. INTRODUCTION

The CLINFO Prototype Data Management and Analysis System is a computing system for the storage, retrieval, and analysis of data (called clinical study data) pertaining to medical research on human subjects. It also provides a means for evaluating the efficacy of such systems in the environment of General Clinical Research Centers (GCRCs) funded by the GCRC branch of the Division of Research Resources, National Institutes of Health (NIH).

GCRCs are small inpatient/outpatient research units, generally with fewer than 20 beds and usually located within a university-affiliated hospital, that serve as institutional research resources for improving the quality of care that physicians offer. They provide patient care, special prescribed dietary services, biological sample collection, and specialized laboratory facilities. Approximately 80 GCRCs are funded by NIH. The CLINFO experiment is designed to determine the requirements for a GCRC data processing tool and its computational/statistical support staff.

The CLINFO prototype is based on the premise that clinical research data are a precious commodity representing considerable intellectual, organizational, and financial investment. They are derived from tests and measurements made on human subjects, and they hold the promise of new knowledge in the treatment or prevention of human disease. The unique set of data collected during the study of a particular group of patients, under a single protocol, is the focus of the CLINFO prototype.

The data for a number of studies may be physically stored within the CLINFO system, and several investigators may use the system simultaneously. However, each investigator should view the system as his personal data storage, manipulation, and analysis tool. The design of each study will probably be unique in terms of the number and identity of variables, the frequency of data collection, the number of subjects, and the treatment and control conditions under which the data are collected; nevertheless, there are characteristics and dimensions common to all collections of clinical research data.

WHAT CLINFO DOES

The CLINFO prototype was designed to support clinical research data processing activities by providing an interactive computer system that

- Allows the investigator to describe the contents of a computer-based study data file.
- Provides for data entry into the study data file and several forms of data screening and encoding.
- Protects the investigator's data from loss due to computer malfunction.
- Allows the investigator to extract from the list of all patients in the study those patients who have particular characteristics.
- Allows for the creation of worksheets of data from a variety of sources, in particular, from the investigator's computer-based study data file.
- Allows the investigator to perform any of a number of mathematical or statistical analyses upon these worksheets and produce simple reports, plots, and graphs from them.
- Allows for the transformation of CLINFO data files into a form compatible with a simple programming language, BASIC.
- Provides for collection of statistics to aid in performing the CLINFO experiments and managing the CLINFO prototype as a clinical research center resource.
- Provides an operating environment for the prototype system which allows the inclusion of additional system features that are not part of the original design.

The CLINFO capabilities are provided by a set of computer programs that have been developed and continuously refined at The Rand Corporation, under the sponsorship of the NIH Division of Research Resources. The prototype is implemented on a Data General minicomputer which, with all associated hardware except user terminals, costs about $100,000. Up to eight terminals may be connected to the minicomputer; however, it is expected that only four of them will be in use simultaneously.

The minicomputer is physically installed in or near the clinical research center, and its operation is overseen by a resident system manager responsible for day-to-day operation, and by a CLINFO project director with overall responsibility for the local CLINFO experiment. Designed for the active investigator-users of a GCRC, CLINFO provides facilities for about 30 current studies resident in the system.

To use the prototype, the user logs onto a free-standing, remote terminal with a password that has been assigned to him and then proceeds to describe his study, enter data, retrieve previously entered data, create patient subsets, or analyze and display data. The password is required both to maintain the privacy of the research data and to protect it against malicious or inadvertent tampering. Passwords are assigned by the system manager.

The system is highly integrated and allows the user to perform a variety of functions without the need to write programs or to reformat data when passing them from one function to another. The CLINFO user does not have to remember or formulate complex commands in order to use the system. Rather, CLINFO asks him about the next step to take while suggesting possible responses, and he responds by typing the name of a variable or worksheet or a short word such as yes or no. Whenever a choice among functions is called for, the prototype displays on a CRT (cathode ray tube) screen the current options available and requests a response. By typing in one-word commands beginning with "I", however, experienced users may move from function to function without the need to pass through intermediate menus of choices. On command, a small printer attached to the terminal will copy the contents of the CRT screen onto paper.

The design philosophy of the CLINFO prototype requires that it be very "forgiving"; thus, inexperienced users may learn to use the system by logging on and trying the options presented, without being concerned that they will harm the system or their data in any way.
The resident CLINFO system manager at each site is available to provide assistance to CLINFO users. The system manager holds formal training sessions for new users and instructs, assists, and advises both new and experienced users at their request. There are also certain system functions that only the system manager can carry out. These include, for example, maintaining backup copies of all user data, retrieving copies of files that may have been deleted inadvertently, preparing files of data for analysis on other systems, and assigning passwords to new users. In addition, the system manager oversees the maintenance of the CLINFO system, writes special-purpose application programs for users, and responds to users' problems; he also communicates with the system designers about user requests, actual or suspected errors in the system, and system utilization. In addition, each CLINFO site is visited periodically by a Rand Corporation CLINFO system representative, who learns about users' needs and problems first-hand, informs users about system changes that are planned or in progress, and assists users as necessary.

CLINFO TERMINOLOGY

Patients, Items, and Panels

The unit of analysis is usually the patient (or case or subject). The investigator assigns a unique patient abbreviation (e.g., initials or hospital number) to each subject. A number of different observations, measurements, and tests are made on each subject. We call these measurements, tests, and classifications items (e.g., systolic blood pressure, height, sex), and we refer to the specific quantities or qualities associated with the items as values (e.g., 140 (torr), 63 (inches), male).

The values for certain sets of items are usually collected at the same time (for example, height, weight, and other physical data or a number of measurements made on a single laboratory sample). Such collections of item values are termed panels. In the CLINFO system the panels and items are named by the user.

The data for a single study may include a number of instances of each panel, where each instance corresponds to a different data-collection time. For example, if blood were drawn every 15 minutes after a drug is given, each resultant set of blood chemistry values would correspond to an instance of the blood panel, and the data collection times for a particular research subject would be the times that his blood is drawn. At the time of data entry, each instance of a panel is uniquely identified by its context, i.e., by the panel name, patient abbreviation, and time of data collection.

Time and Events

The protocol defining the clinical research study usually specifies the times at which measurements are to be taken relative to some experimental procedure, but the data are actually collected at real clock times which are often different for different patients. For the purposes of accuracy and accountability, it is useful (but not necessary) for the data to be identified by the (clock) time of collection; but for review and analysis, it is necessary to align the data in time relative to a fixed
occurrence across all of the patients, or relative to the occurrence of a measurement whose time cannot be predicted.

The CLINFO system allows the user to specify events that may be used during retrieval to indicate a relative zero time, around which instances of panels (or items) may be arranged. The system always recognizes two events: begin, which is the time of the first data collection, and end, the time of the most recent instance of any panel. Other events may be specified to be the time of the first, last, maximum, or minimum value of any item, or the time of occurrence of a specific value of an item. Thus, whereas an item may have different values at different times, an event has only a single value which is a time; and whereas individual item values in the database do not change, the time of an event may change as data are entered and the item associated with the event takes on new values.

Numbers, Codes, and Text

The error-free recording of study data is an obvious requirement for meaningful analysis but is not easily assured. The CLINFO system approaches this problem by providing user-specified error-checking procedures as part of the data-entry activity.

The data recorded by the CLINFO system are of two types, numeric and textual. Most of the data stored by CLINFO will be of the numeric type, which includes real numbers, times, dates, and codes. Short words or character strings may be used to represent predefined values of items, e.g., male and female might be used as values of the item sex, and lung, breast, liver, colon, etc., might be used as values of the item primary (i.e., primary site). Entered character strings are translated into numeric codes for storage, and the numeric codes may be decoded into character strings for display.

The textual type of data provides for the recording of comments and observations. Textual data may be retrieved and displayed, but the system does not provide for analysis of them (e.g., it cannot search for a specific word or phrase), nor does any error checking take place. (The CLINFO hardware does support a text editor called Clem, not described here, which is useful for such applications.)

The Study-Specific Schema

The CLINFO prototype (1) “understands” the association among items in a panel, (2) recognizes the names of items, (3) screens values upon entry for conformity with user-specified ranges or codes, and (4) flags events, all in a manner specific to an individual study. A good deal of information must be supplied to the system, and this is done by a procedure called describe, which assists the user in constructing a data dictionary that describes the items to be included in his study. This data dictionary, called a schema, includes the names of panels, items, and events, the data types of the items, the kinds of screening to be performed, and the equivalence between the numeric codes and character strings.

CLINFO Activities

Describe is one of the major groups of functions in CLINFO called activities. Other activities provide for entering, retrieving, and analyzing data, creating pa-
tient subsets, and manipulating rectangular arrays of data. An individual function (e.g., listing the patients in a specified subset, or performing a chi-square test) is called a subactivity or option. A new CLINFO user may perform a function by selecting an activity from a menu (i.e., a list displayed on the CRT), selecting a subactivity, and then responding to a series of prompts. An experienced user may go directly to the subactivity by typing the activity name and the subactivity number or by using keys labeled with abbreviated activity names.

Subactivities are generally designed to perform commonly used, discrete functions. For example, Frequency Distribution, analyze, option 9, computes the numbers of times that distinct numbers or codes occur. Some more complex or less common functions require that several subactivities be used in combination. For example, a cumulative distribution is plotted by computing frequencies with analyze, option 9, obtaining the cumulative distribution by applying calculate; option 4 (Cumulative Sum) to the results, then plotting the distribution with analyze, option 7 (Scatter Plot). Such sequences are similar to those performed manually and are easily devised once the user is familiar with the system’s capabilities. Other functions, particularly those involving data retrieval, that do not correspond to typical manual operations are more difficult to devise, and help from the system manager may be required. Helpful hints listed below indicate how to perform some of these more complex functions. If a particular sequence of steps comes to be used repeatedly, this sequence can be stored in a CLINFO response file (see below) to simplify the processing.

**Patient Set**

The user may find it useful to identify different groupings of the patients in the study and to examine the data for these groups separately. These groups are called patient sets or subsets and they are established by using the subset activity. Since sets are merely named, ordered lists of patient abbreviations, a patient may belong to any number of sets. When retrieving data, the user has the option of specifying the set named all (which contains all of the patients on a study) or any other set. Subsetting can also be used in the inverse manner by finding those patients who are (or are not) members of specified sets. A set refers to a list of patients, not to their data.

**OVERVIEW OF CLINFO OPERATION**

An ideal data analysis system would perform as shown in Fig. I-1. That is, data are entered into the system in some predefined order; they are then rearranged according to the requirements of the proposed analyses, and those analyses are carried out, producing the “results.”

In the real world of clinical research, however, the data become available a little at a time, and often in unpredictable order. Furthermore, the organization of the data may be quite complex, with some items collected only once, other items collected repeatedly, and other items (or sets of items) missing for some patients. But most statistical analyses require that the data upon which they operate be in neat, two-dimensional arrays, and such arrays are also a natural form for human obser-
Fig. I-1—An ideal data-analysis system.

vation of data. In order to accommodate these conflicting needs (i.e., ease of data entry, manipulation, and analysis), the CLINFO system provides two complementary facilities for storing experimental data: the Study Data File (SDF) and work-sheets. Data stored in the SDF are organized as shown in Fig. I-2. The SDF may be thought of as having three dimensions: patient, time, and item. The value of an item (or the group of values for items in the same panel) is represented by a point (or a line of points) in the three-dimensional space of the SDF. Because different patients may have different numbers of repetitions of some panels, the space may be very sparsely filled even when all the collected data are entered.

Worksheets are two-dimensional arrays which may be created (and destroyed) at the option of the user and which permit easy manipulation and analysis of data. Data may be entered directly into worksheets by a variety of methods, discussed below. When data are entered in these ways, the rows and columns of the worksheets may stand for whatever the user wishes. However, when the worksheets contain data retrieved from the SDF, they may be of only one of three major types, diagrammed in Fig. I-3. They are termed "patient," "item," or "time" worksheets, and each represents a different slice through the SDF. This relationship is indicated in Fig. I-3 by showing the "third" dimension as a dotted line. Since the slices are two-dimensional, they can contain values for only one "point" on the excluded dimension. Thus, in the patient worksheet, the rows can represent different items, and the columns, different times (or instances). Every value in that worksheet belongs to the same patient, as illustrated in Fig. I-4. A single patient worksheet might contain all of the data stored for that patient.

Similarly, the other two worksheet types result from the retrieval of (1) many values of a single item (for a group of patients at a number of times) or (2) many
Fig. I-2—Organization of data stored in the SDF. Each data value is associated with a particular patient and a particular data collection time for a schema item. Multiple data values may be stored for a single item, each with a different data collection time.

Fig. I-3—Types of worksheets retrieved from the SDF. Each of the three types represents a slice through the SDF. A patient worksheet contains values for multiple items at multiple times for a single patient.

values collected at a particular time (or relative time) for a group of patients and a number of different items.

The primary relationships among the SDF, the worksheets, the original data, and the "results" are diagrammed in Fig. I-5. The arrows represent CLINFO activities; that is, user-directed operations that transform the data as they pass through the CLINFO system. The four activities illustrated in this simplified diagram are:

1. Enter (data entry into the SDF).
2. Retrieve (retrieval of portions of the SDF into worksheets).
3. Analyze (the generation of statistical results or other reports, graphs, or charts).
4. Worksheet (the creation and direct entry of data into worksheets).
Fig. I-4—A patient worksheet. Each cell is the value of a particular (or relative) time.

<table>
<thead>
<tr>
<th>ITEM 1</th>
<th>ITEM 1 AT TIME 1</th>
<th>TIME 1</th>
<th>TIME 2</th>
<th>TIME 3...TIME t</th>
</tr>
</thead>
<tbody>
<tr>
<td>ITEM 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ITEM 3</td>
<td></td>
<td></td>
<td>VALUE OF ITEM 3 AT TIME 3</td>
<td></td>
</tr>
<tr>
<td>. . .</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ITEM n</td>
<td></td>
<td></td>
<td>VALUE OF ITEM n AT TIME 2</td>
<td></td>
</tr>
</tbody>
</table>

Fig. I-5—Schematic relationships among major CLINFO activities (arrows), data storage (ovals), and data input/output. Primary storage is in the SDF; all analyses are done on worksheets.
These activities may be used in several different sequences, including the following three:

1. Enter data into the SDF (enter); retrieve data from the SDF into work-
sheets (retrieve); then analyze or plot worksheet data (analyze).

2. Create, name, and label worksheets (worksheet); enter data into work-
sheets (worksheet); then analyze or plot the data (analyze).

3. Create worksheets and enter data into them (worksheet); analyze the data
and store resultant derived data in new worksheets (analyze); discard the
original worksheets (worksheet); enter the derived data into the SDF (en-
ter); retrieve data from the SDF into worksheets that have different struc-
tures and contents than the original ones (retrieve); then analyze or sort
the data in the new worksheets (analyze).

Each of the major activities has within it various options. For example, the
worksheet activity provides for the naming and titling of new worksheets, the
labeling of their rows and columns, the entry of data into the worksheet (by row,
column, or cell), sorting of data within worksheets, and rearranging rows and
columns within existing worksheets.

As described above, the SDF is a highly structured file into which user data are
placed, and from which they are retrieved. The particular structure of an investiga-
tor’s SDF depends on his particular study—the number of cases, the number of
different items (variables), the time relationship among the items, the number of
replications of groups of items, etc. This information (and more discussed below)
is contained formally or informally in the investigator’s experimental protocol. It
must be communicated to the CLINFO system so that the SDF can be constructed
to represent the data structure of the study, so that names may be associated with
items, and so that values of items may be screened before they are entered into the
SDF.

This requires the introduction of several more concepts, including two more
activities and two more files. Figure I-6 illustrates these new concepts and presents
a more complete functional picture of the CLINFO system.

One of the new files in Fig. I-6 is the schema, which contains a full description
of the SDF; the investigator uses the schema to make CLINFO specific to his
problem. All data added to the SDF are first checked against the schema to see that
they conform to the investigator’s specifications. The arrows representing the enter
and examine activities are thus shown as passing through the schema. The activity
that permits the user to construct the schema is called describe. This activity is used
only at the start of a study; once constructed, the schema is “locked” to protect it
against accidental changes and to protect the “meaning” of each stored item value
from inadvertent changes. The Update File provides a buffer between various data
entry procedures and the SDF. It holds data temporarily until they can be merged
into the SDF. Thus, data entry is not slowed down by the rather complicated
process of inserting new data into the proper locations in the SDF. Furthermore,
the data in the Update File may be reviewed (using option 3 of the enter activity)
and changes may be made before merging. The examine activity permits the user
to view the contents of the SDF directly, to modify (or delete) the values of individual
items (such changes are automatically flagged for data-auditing purposes), and
to delete entire panels.
Fig. I-6—A more complete functional representation of the CLINFO system. The schema, derived from the experimental protocol, mediates all communication with the SDF. Data enter the SDF via the Update File.

Although there are many variations, the activities illustrated in Fig. I-6 are normally used in the following sequence:

- Describe and create the schema (describe).
- Enter patients into the study (enter).
- After trial data entry (enter) and retrieval (retrieve), modify the schema (describe) until it is satisfactory.
- Enter data into the Update File (it cannot be entered directly into the SDF) and review it (enter).
- Ask the system manager to merge (i.e., reorganize and store) the Update File into the SDF, then discard the Update File (merge).
- Examine and possibly modify data in the SDF (examine).
- Retrieve data from the SDF (possibly for a subset of patients—see the following discussion) into worksheets (retrieve).
- Create additional worksheets (worksheets), enter original data (enter) and the results of analyses (analyze) into worksheets, and enter worksheet data into the Update File (enter).
- Enlarge retrieved worksheets or reorganize data within them (worksheet), sort data in worksheets based on values of variables (worksheet), move data from one worksheet to another (retrieve—not illustrated), and perform calculations on worksheet data and store the results directly in the same worksheets (calculate—not illustrated).
- Analyze or plot worksheet data and possibly store the results in new worksheets (analyze).
Another important activity, subset, is illustrated in Fig. I-7. It comprises subactivities which allow the creation of sets of patients according to logical and arithmetic criteria specified for data in the SDF (options 4 and 5) or for data in worksheets (options 6 and 7). A set may also be created by naming its members explicitly (option 1) or by performing logical operations on existing sets (options 2 and 3). Once sets are created, retrieval from the SDF may be restricted to data belonging to members of a particular patient set. This is represented in Fig. I-7 by an arrow showing retrieval of data from the SDF to worksheets, going "through" the Subset File.

Several other activities are also shown in Fig. I-7. The calculate activity and a subactivity of retrieve are both represented by the same arrow. Calculate permits arithmetic, logical, and special operations to be performed on worksheet rows or columns, and stores the results in the same worksheet. Arrows going back and forth between worksheets and results represent the capability of some analyze subactivities to store their results in worksheets for purposes of plotting or additional analysis. Retrieve, option 4, permits the user to copy portions of worksheets (or entire worksheets) into other (or the same) worksheets. This is illustrated in Fig. I-8, where a new worksheet, C, has been built which contains sections of worksheets A and B.

Another group of activities which make the system easier and more convenient to use is shown in Fig. I-9. These activities allow the display of data from the SDF, lists of patient sets or worksheets (in the files activity), lists of members of a particular set (in the subset activity), and portions of worksheets or entire work-

![Fig. I-7—Relationships of the subset activity to other major CLINFO activities. Patient sets may be created by listing patient abbreviations, specifying conditions on data in the SDF, or specifying conditions on data in worksheets. Retrievals may then be limited to patients in specified sets.](image-url)
Fig. I-8—Copying of data from worksheets into other worksheets or different locations of the same worksheet using retrieve, option 4. Segments of worksheets, defined by row and column bounds, may be copied into other worksheets or copied into different locations in the same worksheet. In this example, values for variables α through μ for two groups of cases (A-M and CC-MM) are copied into a new worksheet C.

Fig. I-9—CLINFO options that permit display of data and other important study information. For example, files, option 2, presents a list of patient sets, while subset, option 8, lists the members of a particular set.
sheets (in the worksheet activity). If hard copy is desired, the contents of the CRT screen may be printed out on the small printer attached to the terminal; for more extensive output (e.g., an entire large worksheet), it is more convenient to use the system printer (using worksheet, option 10).

Figure I-10 introduces the concept of the Communication File. This file and the activities that operate upon it allow the exchange of data between CLINFO (the region within the dark frame), BASIC programs that run on the same computer as CLINFO, and external systems. For example, this exchange may be made via magnetic tape written by another computer or paper tape from an instrument such as a scintillation counter (arrow 1); or data values produced by using a BASIC program may be stored in a Communication File (arrow 2); or data in the SDF or a worksheet may be copied into a Communication File for transmittal to a BASIC program or external system (retrieve, options 3 and 5); or properly formatted data in a Communication File may be stored in a CLINFO worksheet (retrieve, option 6).

This capability allows the user to write (or have written) BASIC programs to carry out special analyses not presently provided by the CLINFO system. The results of such analyses may be displayed directly, may be passed back into CLINFO worksheets, or may be passed back (through a worksheet and the Update File) into the SDF.

A final activity, shown in Fig. I-11, is called response files. It enables a user to carry out a sequence of CLINFO activities repeatedly without reentering the required series of responses. In this figure, the user creates a Response File (using response, option 2), then carries out a sequence of activities (A, B, etc.) by entering

---

Fig. I-10—Using the Communication File, data may be moved back and forth from the CLINFO system to BASIC programs or, via computer tape, to other machines. Once data have been entered into worksheets, either directly or from the Communication File, they may be merged with the SDF.
Fig. I-11—The creation of a Response File in the CLINFO system. Response Files are created by carrying out the desired sequence of activities; the user's responses are automatically recorded.

a series of responses (a₁, a₂, etc.) to the normal system prompts. The responses are recorded in a Response File which is named and saved.

Figure I-12 represents the editing of the Response File (using response, option 3). In this example, the response actually recorded, bₙ, is replaced by a question mark which indicates that the user will supply a response each time the sequence is executed.

In Fig. I-13, the user executes the Response File (using response file, option 1). The system reads and acts upon the sequence of responses, a₁, a₂, ..., aₙ, b₁, from the Response File until a question mark is encountered. It then waits for the user to respond. After acting upon the user's response, the system continues reading responses from the file, b₂, n₁, n₂, etc., and acting upon them.

USE OF THIS MANUAL

We recommend that the new user begin by reading this Introduction. It presents a broad picture of the purposes, functions, and capabilities of the CLINFO prototype and introduces the terminology used throughout the guide.
Fig. I-12—Editing the Response File. The user's responses may be edited, for example, to replace a specific response with a ?, which allows the user to enter a different response each time the Response File is run.

Fig. I-13—Execution of the Response File. The original sequence of activities is run with all prompts displayed and with most responses taken from the Response File; a question mark in the Response File causes it to wait for a user response from the terminal.
Section II contains more detail about each CLINFO activity and about the organization of data in the CLINFO system and gives a detailed description of the user's terminal. Short, practical instructions for logging on and moving about within the CLINFO system are given on pp. 19-23.

Section III leads the user through an example of the use of the prototype with realistic clinical research data.

Section IV contains suggestions and ideas derived from the use of the system by its designers and its initial users. It provides guidance about how to effectively accomplish tasks and use CLINFO functions.

The Reference Manual, Sec. V, contains detailed descriptions of each system feature and function. It is arranged alphabetically so that particular functions, commands, and features may be accessed as required in the day-to-day use of the system. In addition, all of the CLINFO menus are printed on the inside back cover.
II. USING THE SYSTEM

THE ESSENTIALS

This section presents the essential information required to use the CLINFO prototype system.

Using the Terminal*

Turning It On. The Ann Arbor (AA) terminal and the attached printer are sketched in Fig. II-1. The power switch (a slide switch) is mounted on the side of the terminal. Place the switch in the forward position to turn the terminal on; the sound of a fan will indicate that the power is on. Press the Data Terminals and Communications (DTC) printer POWER switch; it should light up. Also press the LINE/LOCAL switch if the line portion is not lighted, and press the top of the DTC RESET switch.

Getting the Attention of the System. The ESC (escape) key, at the upper left-hand corner of the keyboard, is used to get the attention of the system both to log on (see below) and to interrupt ongoing operations. It is only active while you hold down the yellow key marked CLINFO, on the third row down, at the left-hand edge of the keyboard.

Where You Are. The small “dash” on the screen, called the cursor, is usually positioned following the last system prompt and moves one space as each character of your response is typed in.

Responding to System Prompts. After typing in your response (and seeing that it is correct on the screen), you must press the RETURN key (the large key at the right edge of the alphanumeric keyboard). Many prompts may be responded to simply with a RETURN.

Commands. A CLINFO activity is initiated by pressing the key labeled with that activity name (abbreviated, in yellow) and simultaneously pressing the yellow CLINFO key, then typing an option number (if an option is desired), then hitting the RETURN key.

Correcting Errors. After typing a response, and before hitting the RETURN key, you may correct your response by backing up over it (thereby erasing) character by character, using the BS (backspace) key directly above the RETURN key.

Precautions. It is important to keep in mind the following features of the system and terminal:

- The system distinguishes between upper and lower case. If the TTY LOCK (lower left) key is lighted, press it to deactivate it; when on, it acts as an upper-case lock for the alphabetic characters. The SHIFT LOCK key (just

* For more details, see pp. 40-42.
Fig. II—The Ann Arbor terminal and the attached DFC printer.
above the SHIFT key at the lower left of the keyboard) locks into place when pressed; to release it, hit the SHIFT key.
- Holding down keys (including RETURN) for more than a second causes them to repeat their function.

Logging On

Each time you use the CLINFO system, you must "log on," i.e., you must use the password issued to you by the system manager to identify the study with which you wish to work. If more than one person is to have access to the same study (e.g., the principal investigator, a fellow, and a technician), each will be issued his own password.

To log on:

1. Make sure that the terminal is on (see "Using the Terminal," above).
2. Hold down the yellow CLINFO key and hit the ESC key—the system should respond, "PASSWORD:"
3. Enter your password, e.g., 1234MYSTUDY, and hit the RETURN key. The password will not be displayed as you enter it, so be careful to enter it correctly. Common errors are:
   - Using the wrong case (e.g., the SHIFT LOCK or TTY LOCK may be on).
   - Confusing the letter "o" with the numeral zero.
   - Confusing the letter "t" with the numeral one.
   - Inserting a blank between the numbers and the letters, e.g., 1234 MYSTUDY.
4. The system should respond by clearing the screen, perhaps displaying a "message of the day," indicating the current worksheet and the size of any existing Update File, and prompting for the name of the first activity to be used. If this occurs, proceed as described in "Communication Between Activities," below. If, instead, the system responds

   UNKNOWN PASSWORD
   CLINFO RELEASE 3.0 AT YOUR SERVICE

make sure you are using the correct password and repeat from step 2 (i.e., hit the CLINFO and ESC keys again).

If the system responds

   ID IN USE
   CLINFO RELEASE 3.0 AT YOUR SERVICE

someone else is currently logged on to your study. If no one else is authorized to access your study, check with the system manager; if other people are authorized to use it, wait until they are finished and try again.

If the system does not respond to the CLINFO/ESC keys with "PASSWORD," check with the system manager—he may be doing backup or some other system function, or the terminal may not be properly connected to the computer.
If you were not the first to use the terminal after turning it on, it might have a message of the form

**CLINFO RELEASE 3.0 AT YOUR SERVICE**

on the screen; proceed as above.

---

**CLINFO Activities**

**Summary of Activity Functions.** The CLINFO activities and their functions are listed below. The activity names, prefixed with `!`, are the commands that call up the activities themselves (see "Communication Between Activities," below). The activities are described more fully later in this section; for complete details, see the Reference Manual (Sec. V).

The options within each major activity are listed in the menus on the inside back cover and are briefly described later in a table entitled "Summary of Activity Options." They are also listed in the Table of Contents and are detailed, in alphabetical order, in the Reference Manual.

<table>
<thead>
<tr>
<th>Activity</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>analyze</td>
<td>Plot and perform statistical analyses upon data in worksheets; store results in other worksheets.</td>
</tr>
<tr>
<td>calculate</td>
<td>Specify and carry out arbitrary arithmetic, logical, and other operations upon data in rows or columns of worksheets; store results in the same worksheet.</td>
</tr>
<tr>
<td>describe</td>
<td>Construct, edit, and print the schema and create and delete study files.</td>
</tr>
<tr>
<td>enter</td>
<td>Enter patients into a study; enter data into the Update File.</td>
</tr>
<tr>
<td>examine</td>
<td>Examine and edit SDF data panel by panel.</td>
</tr>
<tr>
<td>files</td>
<td>Display ordered lists of worksheets, subsets, Response Files, and utility files.</td>
</tr>
<tr>
<td>goodbye</td>
<td>Terminate the current CLINFO session.</td>
</tr>
<tr>
<td>message</td>
<td>Record messages for the system manager.</td>
</tr>
<tr>
<td>response files</td>
<td>Create, edit, and execute lists of CLINFO commands for frequently used sequences of activities.</td>
</tr>
<tr>
<td>retrieve</td>
<td>Extract data from the SDF, from worksheets, and from a Communication File; place them into worksheets or Communication Files, or display them.</td>
</tr>
<tr>
<td>subset</td>
<td>Create sets of patient abbreviations by listing, from other sets, or by specifying conditions upon data in the SDF or in worksheets.</td>
</tr>
<tr>
<td>worksheet</td>
<td>Create, label, enter data into, modify, and discard worksheets.</td>
</tr>
</tbody>
</table>
**Communication Between Activities.** In order to move from one activity to another (e.g., from *worksheet* to *analyze*) or from one option to another within an activity, (e.g., from *Display this Worksheet* (option 1) to *Label Worksheet Rows* (option 4)), the user generally types commands that start with an exclamation mark (!). The commands for these moves are listed below. Commands may be typed in full or may be entered by using the yellow CLINFO key in conjunction with one of the labeled activity keys. Commands may be entered whenever the system is waiting for a response.

<table>
<thead>
<tr>
<th>To go from:</th>
<th>To:</th>
<th>Type:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Any activity</td>
<td>Main menu of activity choices</td>
<td>!_done</td>
</tr>
<tr>
<td>Any activity</td>
<td>Menu of options for activity α*</td>
<td>!_a (e.g., _lana)</td>
</tr>
<tr>
<td>Activity α</td>
<td>Option number n of activity β</td>
<td>!_β,n (e.g., _lana,7)</td>
</tr>
<tr>
<td>Anywhere in activity α</td>
<td>Option number n of activity α</td>
<td>!_n (e.g., !_7)</td>
</tr>
<tr>
<td>Option number n of activity α</td>
<td>The same option of the same activity</td>
<td>!_ or CLINFO/ESC</td>
</tr>
<tr>
<td>Menu of options for activity α</td>
<td>Option number n of activity α</td>
<td>_n (e.g., _7)</td>
</tr>
<tr>
<td>Any activity</td>
<td>Terminate the session</td>
<td>!_goodbye</td>
</tr>
</tbody>
</table>

* α and β represent CLINFO activity names, which may be abbreviated to their first three letters.

It is important to note that:

- Some activities will prompt for obligatory information before they will let you move to another option or activity.
- Any activity name may be abbreviated to its first three letters, e.g., _lana_-alyze may be entered as _lana_.
- Pressing CLINFO/ESC will interrupt the current activity and return you to the beginning of that activity; you need not wait for a prompt, as with !_.
- A complete command may always be used in place of an abbreviated command. For example, to go from anywhere in one activity to option _n of the same activity, !_a,_n and !_n are equivalent, and to go from option _n of one activity to the start of the same option of the same activity, !_a,_n and !_n and !_ are all equivalent.
Simultaneously pressing the yellow CLINFO key and a key with a yellow label (e.g., \texttt{ana}) is equivalent to typing the label. Thus, simultaneously pressing the CLINFO key and the \texttt{ana} key, then typing 7 produces the same results as typing \texttt{ana,7}.

**How to Find Essential Information About Your Study**

In your use of the CLINFO system, you will frequently need to refer to lists of patients, variables, worksheets, and other information peculiar to your study. The activities that will produce the various lists are shown below. Keeping paper copies of your lists will prove very helpful, since you will not need to leave your current activity in order to recall what patients have been entered, what worksheets have been created, how item names are spelled, and so forth.

<table>
<thead>
<tr>
<th>To List</th>
<th>Enter</th>
</tr>
</thead>
<tbody>
<tr>
<td>Activity menu</td>
<td>\texttt{!done}</td>
</tr>
<tr>
<td>Events</td>
<td>\texttt{!describe, 1 or !describe, 2} *</td>
</tr>
<tr>
<td>Items</td>
<td>\texttt{!describe, 1 or !describe, 2} *</td>
</tr>
<tr>
<td>Panels</td>
<td>\texttt{!describe, 1 or !describe, 2} *</td>
</tr>
<tr>
<td>Patient abbreviations</td>
<td>\texttt{!subset, 8 (with set “all”)}</td>
</tr>
<tr>
<td>Response Files</td>
<td>\texttt{!files, 3}</td>
</tr>
<tr>
<td>Subsets</td>
<td>\texttt{!subset, 9 or !files, 2}</td>
</tr>
<tr>
<td>Worksheets</td>
<td>\texttt{!work, 9 or !files, 1} **</td>
</tr>
</tbody>
</table>

**Helpful Hints**

Below are some suggestions about how to use CLINFO activities to accomplish frequently performed tasks. Additional suggestions appear in Sec. IV, Helpful Hints.

- Create a new worksheet. Enter \texttt{!work,0} and the name of the new worksheet, its title (i.e., a short description), and size.

- Move data within a worksheet so that it is lined up for analysis. Enter \texttt{!ret,4} and name the worksheet as both the source and destination, then proceed.

- Leave a message for the system manager to merge your data, lock your schema, or provide other assistance. Enter \texttt{!mes}, leave a message, and then proceed. Also, a request to merge can be issued by responding yes if prompted about this when logging off.

* Describe, option 2, produces a nicely formatted listing of the schema on the system printer; option 1 produces a more compact listing either on the system printer or on the terminal.

** Worksheet, option 3, generates a list of complete worksheet descriptions on the terminal; files, option 1, produces a much more compact listing with less detail but permits sorting by name, size, etc.
• Make a Response File readable. While creating it, use the complete  
  lact-name,n notation instead of the shorter in notation. When editing  
  it, follow a question mark with a  
  space and an explanatory comment.

• Add, delete, or move entire rows  
  or columns within a worksheet. Enter lwork,11 and proceed.

• Find out if your Update File data  
  have been merged. When signing on or off or after  
  entering lent, see if an Update File  
  exists.

• Print the entire current work-  
  sheet. Enter lwork,10 and proceed.

• Print an entire Response File. Enter lres,3, followed by the name  
  of the Response File and print.

• Make a worksheet with standard  
  labels. Execute a Response File that creates  
  a worksheet and labels it.

• Copy a worksheet. Use retrieve,5 to store the  
  worksheet in a Communication File, then use retrieve,6 to retrieve  
  the same Communication File and  
  store it in a new worksheet.

• Create a patient set that lists  
  patients in the same order in  
  which they appear in a sorted  
  worksheet. Use subset,7 to form a new set  
  that contains all the listed  
  patients.

---

**System Manager Functions**

The system manager

• Instructs and assists users.
• Assigns passwords for new studies.
• “Locks” and “unlocks” schema.
• Merges the Update File into the SDF.
• Writes special BASIC programs.
• Transfers externally created machine-readable data to CLINFO.
• Sees to the maintenance of the CLINFO computer.
• Documents and reports to the system implementers about system error  
  messages.
• Copies all user data to tape each day (i.e., does “backup”).
• Installs new CLINFO software and informs users of changes.
Size Limitations

Limitations on various CLINFO parameters are listed below:

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Maximum Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patients per study</td>
<td>392</td>
</tr>
<tr>
<td>Variables (items)</td>
<td>392</td>
</tr>
<tr>
<td>Panels</td>
<td>56</td>
</tr>
<tr>
<td>Events</td>
<td>28</td>
</tr>
<tr>
<td>Items/panel</td>
<td>30</td>
</tr>
<tr>
<td>Repetitions of panels</td>
<td>No practical limit</td>
</tr>
<tr>
<td>Worksheet size</td>
<td>From 1 row by 799 columns (or vice versa) to 47 rows by 47 columns (see p. 273)</td>
</tr>
<tr>
<td>Text item length</td>
<td>70 characters</td>
</tr>
<tr>
<td>Patient identifiers</td>
<td></td>
</tr>
<tr>
<td>Worksheet names</td>
<td></td>
</tr>
<tr>
<td>Item names</td>
<td></td>
</tr>
<tr>
<td>Panel names</td>
<td>8 characters</td>
</tr>
<tr>
<td>Subset names</td>
<td></td>
</tr>
<tr>
<td>Character-string item values</td>
<td></td>
</tr>
</tbody>
</table>

System Errors

Occasionally the CLINFO system will make an error. Because, ideally, this should never occur, and because it is important for us to eliminate all such "bugs," we would appreciate it if you would inform your system manager about system errors and the circumstances under which they occur. When a detectable system error occurs, a message will appear on the top line of your CRT screen, and you will be prevented from proceeding normally. The message describes the error and says TYPE quit TO QUIT. At this point, push the PRINT key to copy the screen contents on the printer, call the system manager, explain what happened, and give him the hard copy of the screen contents. Typing quit will log you off and will allow you to initiate a new session. We will fix the error as soon as possible.

CLINFO ACTIVITIES

The CLINFO system is organized into several more or less independent activities. The major processing activities are analyze, calculate, describe, enter, examine, response files, retrieve, subset, and worksheet; the files, goodbye, and message activities perform simple auxiliary tasks. To provide easy access to specific functions, the major activities usually comprise a collection of options (e.g., analyze options include ANOVA and T-Test). Each of the major activities is described briefly below.

ANALYZE

Analyze is the set of data analysis subactivities that allow the investigator to
perform mathematical and statistical analyses upon, and produce reports, plots, and graphs from, the worksheets. The analytical functions, such as t-test, chi-square, etc., are accessed via a selection process. The alternatives are displayed and the user chooses the specific type of analysis that he requires.

The analysis procedures assume that the data upon which they are to operate are contained in user-specified rows or columns of a worksheet. Certain procedures that produce output in tabular form will place that output into a worksheet if requested. For example, the output worksheet of the Linear Regression option includes the residuals and predicted values which may be plotted using Scatter Plot.

The data analysis subactivities are called Analysis of Variance, Chi-Square Test, Cross Tabs, Descriptive Statistics, Frequency Distribution, Histogram, Life Table Analysis, Linear Regression, Non-Parametric Paired Tests, Normality Test, Radioimmunoassay, Scatter Plot, and T-Test. (Data sorting, an analysis-like function, is included in the worksheet activity as option 8, Sort This Worksheet; calculate, a separate activity, performs user-defined analyses.)

In all the data analysis subactivities, missing values are identified as such and do not enter into calculations.

**CALCULATE**

*Calculate* allows the user to specify and compute functions of data in worksheet rows and columns; results are stored in specified rows or columns of the same worksheet. The calculate options give the user a modest amount of programming capability. They are useful when the user wants to compute a function that is not provided by the analyze activity or when he wants the results stored in the same worksheet as the data.

Specifically, *calculate* provides for

- Defining and evaluating arbitrary arithmetic expressions (e.g., row 9 = row 2 + 5 \times \log (row 8)) that refer to rows or columns of a worksheet.
- Defining and evaluating arbitrary logical expressions (involving ANDs or ORs) that refer to rows or columns of a worksheet.
- Special calculations including replacing missing values with a specified constant and vice versa, cumulative sum, and cumulative product.
- Computing time differences, in specified units, between dates and/or times stored in worksheet rows or columns.
- Statistical calculations including counts, sums, means, standard deviations, and standard errors of data in worksheet rows or columns.

It is often useful to add rows or columns to a worksheet (by using worksheet, option 11) before using *calculate*.

**DESCRIBE**

*Describe* allows the investigator to specify, in a standardized manner, the characteristics and the structure of data that he wishes to collect. This specification is called the study schema. The schema furnishes the flexibility required for entry and retrieval of data:
• It provides a means of associating data items into groups; for example, vital signs may be a group of items (corresponding to a CLINFO panel) whose values are measured at essentially the same time by the same person.
• It provides a dictionary which enables the investigator to describe, name, and specify the range or list of acceptable values of data items to be entered into the SDF.
• It allows the user to specify acceptable character-string (i.e., short word) responses and the numeric codes associated with them.
• It provides a means for describing to the system the relationship between events and the values of their associated items. An event is an optional signal within the CLINFO system that provides a point of reference to which other item values may be related in time. For example, an event may represent the administration of a drug, the first occurrence of a measured item, a specific value, or the maximum or minimum of the values entered for an item.
• It allows the user to personally create study files (viz., schema, SDF, and list of patients) for the purpose of testing his schema design by trial data entry and retrieval. If the user finds his schema design unsatisfactory and he has entered only a small amount of data, he may delete the patients and data entered for them, then revise the schema. If a substantial amount of data has been entered, only the system manager can revise the schema; he can make certain kinds of changes while still preserving the user's data.

The optimal design of a schema is critical to both the execution of the protocol which it represents and the effective use of the CLINFO prototype. The system manager will assist inexperienced users in the initial design and entry of a schema and will provide examples of existing well-designed schemata.

**ENTER**

*Enter* is the data entry activity which allows the user to

- Add patients to the study.
- Enter data into the Update File (which is then merged into the SDF).
- Review the Update File.
- Transfer data from a worksheet to the Update File.

As data values are entered, *enter* performs three functions:

- It relates the values to the item descriptions in the schema and does the indicated screening for acceptability and encoding.
- When the sort/merge procedure (see below) is used, *enter* automatically reevaluates the times associated with events.
- It automatically associates with the data values additional identifiers, i.e., date and time of data collection, date and time of data entry and the identity of the data enterer.

One function of the schema is to prevent erroneous data values from entering the SDF. However, what seems to be an error may sometimes be an exceptional
value. CLINFO will question a value that is out of the previously specified range; if the data enteror insists, the system will accept such a value where practicable but will flag it as exceptional and will record the enteror's initials. This inclusion of additional data identifiers and flags for exceptional values provides the investigator with an avenue for auditing his data.

In order to make both interactive data entry and retrieval flexible and efficient, and to implement the CLINFO prototype system on a small, economical computer, it is necessary to process and reorganize the data after they have been entered. This processing is performed by the CLINFO system manager, using the sort/merge procedure, either on request or at a specified time of day. Although data may be reviewed immediately after they are entered, they may not be modified in, or retrieved from, the SDF until after this procedure is executed. The user who wishes to analyze new data immediately may enter the data into a worksheet (using the worksheet activity) and analyze them in that form.

**EXAMINE**

Examining permits the user to display, modify, and delete data values stored in the SDF. Data are displayed by specifying the patient, panel name, and data collection time (i.e., the context). Successive instances of panels may be displayed without respecifying the context. Items whose values are changed are flagged. Examine also provides for determining how many patients and panel instances have been entered and how full the SDF is.

**RESPONSE FILES**

If a sequence of responses is entered frequently with little variation, it is sometimes convenient to save these in a Response File, edit the file to allow for the variations, and later execute the file as appropriate. When executing a Response File, the CLINFO prototype behaves as if the stored responses were entered by the user. The user may watch as execution progresses and enter only those responses that generally change each time the Response File is executed.

In order to maintain generality and flexibility, the data entry portion of the enter activity is quite passive, i.e., the user must indicate when information about a new patient or panel are to be entered. If the data enteror wishes, however, a Response File can be set up that prompts for new patients, panels, and data items at the right points, and the enteror needs only to enter the data values.

The response file activity provides for

- Creating a Response File by example, i.e., by responding to the normal prompts and saving the responses.
- Editing a Response File, e.g., to indicate where the response is to be provided by the user instead of the file or to indicate that a sequence of prompts is to be repeated.
- Executing a Response File.
- Discarding a Response File.
- Making a copy of a Response File to create a variation of it.
- Appending one Response File to another.
RETRIEVE

Retrieve retrieves previously entered data from various sources (the SDF, worksheets, and Communication Files) for purposes of review, display, and analysis; it also stores those values into worksheets and Communication Files or displays them on a terminal.

Retrieve allows the investigator to describe which data he wants to retrieve from the SDF and how he wants them organized, by

- Specifying the patient(s) or patient set of interest.
- Specifying the item(s) or panel(s) for which values are to be retrieved.
- Specifying the time unit that is used for alignment (minute, hour, day, week, month, year, or a multiple of one of these units).
- Specifying the time interval, relative to an event or absolute date and time, that is to be used for data selection.
- Specifying logical conditions to restrict the values retrieved.

The investigator can use an event to single out a particular occurrence of a data value so that its date and time serve as a reference point for the data retrieval process. He may also restrict retrieved data values to those that satisfy specified conditions; in this way, he can discover which patients have data values that satisfy certain conditions and at the same time retrieve their data.

The time associated with an event is recomputed whenever the SDF is updated by the sort/merge procedure. Thus, if the time of an event is defined to be the data collection time of the last value of a variable, it always refers to the most recently measured value.

When retrieve is used to store data into or retrieve data from a Communication File, the user may specify that the Communication File belongs to a study different from the one he is currently accessing. This provides a convenient mechanism for moving data from one study to another.

SUBSET

Subset groups patients into sets, according to specified criteria. When the criteria are specified and the system creates a set by selecting the patients whose characteristics match the criteria, the user names the set and enters comments that describe it. A patient set is a list of patients; it does not include any data. Data for the patients in the set are retrieved by using the retrieve activity and referencing the set name. Also, patients may be selected from a set to create a new set. All patients on the study belong to the set called all.

A set of patients may be created from

- A list of patient abbreviations.
- Those belonging to all of a list of sets.
- Those belonging to at least one of a list of sets.
- Those in a named set whose data in the SDF satisfy all conditions in a list or at least one condition in a list.
- Those named in a worksheet and whose worksheet data satisfy all conditions in a list or at least one condition in a list.
A condition is satisfied when a specified item has a numerical value that lies within a specified range or has a character-string value that matches a specified string. When a set is created by examining data in the SDF, a single condition is satisfied if any value within a specified time interval meets the condition. Depending on the user's specifications, conditions must be satisfied either simultaneously or at any time within the same time interval. As with the retrieve activity, the time interval may be specified as the time between one time-marker (i.e., event or absolute date and time) and another, a list of time units relative to a time-marker, or a range of time units relative to a time-marker.

**WORKSHEET**

*Worksheet* is the activity used for creating, listing, discarding, labeling, displaying, printing, entering data into, editing, and sorting data in worksheets.

The Select subactivity is used to name, specify the size of, and create a worksheet. It is also used to select the current worksheet that is to be further processed by worksheet, analyze, calculate, or retrieve functions. (For convenience, the analyze, calculate, and retrieve activities also have a Select a Worksheet subactivity.)

Two labeling functions are used to label worksheet rows and columns with arbitrary 8-character strings. Labels may also be changed or deleted.

Worksheet data may be entered (or deleted) by row, by column, or by cell. When entering data by row, the user specifies the range of rows and columns for which he wants to enter data; then for each row he is prompted for values to be entered. Cell data entry is convenient for entering a few data values or for changing values; the user simply specifies a row, a column, and a value. Values may be entered as numbers, dates, or character strings. Dates may be entered in a format such as 5/22/75 or 7/4/1976.

The Display subactivity displays a specified 20-row by 6-column portion of the current worksheet on the terminal screen. Different portions are displayed by specifying different starting (i.e., upper left) rows and columns. Values representing dates are presented in a format such as 5/22/1975. Missing values are displayed as three dots (...). If so requested, the Display subactivity decodes encoded values and displays them as character strings.

The entire worksheet may be printed on the system printer by using the Print subactivity. This is particularly useful for larger worksheets.

The Sort subactivity sorts all the rows (or columns) of worksheet data by a single column (or row). The sort may be performed in either ascending or descending order.

The Edit subactivity is used to rearrange the contents of an existing worksheet; to move, copy, or delete rows (or columns); and to insert additional rows (or columns). It may also be used to change a worksheet's name and title.

**Summary of Activity Options**

The options within each major activity are briefly described in the following tables. The examine, goodbye, and message activities are excluded as they do not have options.
<table>
<thead>
<tr>
<th>Activity Option</th>
<th>Name and Function</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ANALYZE</strong></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td><strong>Select a Worksheet:</strong> Provides for specifying which existing worksheet is to be the current one, i.e., which contains the data to be analyzed; also provides for creating a new worksheet.</td>
</tr>
<tr>
<td>1</td>
<td><strong>Descriptive Statistics:</strong> Computes counts, means, medians, ranges, standard deviations, and other statistics for data in specified portions of selected rows or columns of the current worksheet; optionally stores results in a new worksheet.</td>
</tr>
<tr>
<td>2</td>
<td><strong>T-Test:</strong> Tests hypotheses concerning the means of one or two populations by use of the Student t-statistic.</td>
</tr>
<tr>
<td>3</td>
<td><strong>Chi-Square Test:</strong> Tests the hypothesis that two categorizing variables are independent, i.e., that there is no relationship between the distribution of values of one and the values of the other. The test is based on data representing frequencies of cases categorized on two variables.</td>
</tr>
<tr>
<td>4</td>
<td><strong>Linear Regression:</strong> Estimates the linear relationship between one (dependent) variable (for which data are stored in a portion of a worksheet row or column) and up to three (independent) variables; provides for transforming the variables prior to the fit:</td>
</tr>
<tr>
<td></td>
<td>$y = a + bx$</td>
</tr>
<tr>
<td></td>
<td>$sqrt(y) = a + bx$</td>
</tr>
<tr>
<td></td>
<td>$y = a + b/x$</td>
</tr>
<tr>
<td></td>
<td>2nd degree polynomial</td>
</tr>
<tr>
<td></td>
<td>$ln(y) = a + b ln(x)$</td>
</tr>
<tr>
<td></td>
<td>3rd degree polynomial</td>
</tr>
<tr>
<td></td>
<td>$ln(y) = a + bx$</td>
</tr>
<tr>
<td></td>
<td>Original and predicated values and residuals may be stored in a worksheet and may be plotted using analyze, option 7.</td>
</tr>
<tr>
<td>5</td>
<td><strong>ANOVA—Analysis of Variance:</strong> One-way analysis of variance tests the hypothesis that from two to ten groups of samples are derived from the same population, i.e., have the same mean. Group numbers and values in the output worksheet may be plotted (with analyze, option 7) to create a scatter-plot type of histogram.</td>
</tr>
<tr>
<td>6</td>
<td><strong>Cross Tab:</strong> Provides frequency counts of cases categorized on two variables, calculates the total frequency for each category, and calculates the means of each variable as categorized by the other. The user specifies the categories by entering cut-points. The output worksheet may be used as input to analyze, option 3, Chi-Square Test.</td>
</tr>
<tr>
<td>7</td>
<td><strong>Scatter Plot:</strong> Plots bar charts, scatter plots, and connected line drawings; uses data for one dependent variable and up to five independent variables stored in portions of worksheet rows or columns.</td>
</tr>
</tbody>
</table>
8 **Histogram**: Creates and displays a horizontal frequency histogram from data in a portion of a worksheet row or column; computes counts, frequencies, and cumulative frequencies for continuous data (with user-specified cut-points) and for categorical data.

9 **Frequency Distribution**: Counts the number of occurrences of each distinct value in a portion of a specified worksheet row or column; displays the values, sorted in ascending value, and their associated frequencies. The output worksheet may be used to plot vertical histograms (with `analyze`, option 7) or to compute cumulative frequency distribution (with `calculate`, option 4).

10 **Normality Test**: Uses the Kolmogorov-Smirnov statistic to test the hypothesis that a specified sample was drawn from a normally distributed population; produces an automatically scaled horizontal frequency histogram for the sample data (in a user-selected worksheet row or column).

11 **Non-parametric Tests**: Includes non-parametric tests (i.e., tests that do not assume normality) on variables which represent pairs of measurements made on the same cases:

- Spearman Rank Correlation Test—tests the hypothesis that the two variables are independent.
- Sign Test—tests the hypothesis that the median of the difference between pairs of measurements is zero. Useful when both observations of a pair are made under similar conditions but different pairs are observed under different conditions.
- Wilcoxon Signed Ranks Test—tests the hypothesis that the median of a group of observations is equal to some specified value or that the median of the differences between two groups is equal to zero.

Includes the non-parametric non-paired test:

- Wilcoxon Two-Sample Rank Sum Test—tests the hypothesis that the two samples are drawn from the same population.

12 **Life Table Analysis**: The current worksheet contains survival data (time on study and whether or not the endpoint of interest has occurred) for groups or individuals in one or two samples; computes for each time on study (1) the probability of survival to that time, (2) the expected lifetime at that time, and (3) the probability of arriving at the endpoint within the time interval; stores the results in a worksheet for possible plotting (with `analyze`, option 7) or other manipulation; compares the two life tables (if there are two samples) with a Wilcoxon test.
ANALYZE (continued)

13 **Radioimmunoassay:** Estimates the concentration of ligand (i.e., the substance whose concentration is being determined by a laboratory radioimmunoassay) for a number of unknowns. Suboptions provide for the following:
   1. Builds a tube layout and saves it in a worksheet.
   2. Edits an existing tube layout.
   3. Using an existing tube layout, fits a standard curve to the data, plots the fitted curve, estimates the unknowns, and saves the results in a worksheet.
   4. Prints an existing tube layout.

**CALCULATE**

0 **Select a Worksheet:** Provides for specifying which existing worksheet is to be the current one, i.e., which contains the data to be processed; also provides for creating a new worksheet.

1 **Arithmetic Expression:** Provides for defining and evaluating arbitrary arithmetic expressions (e.g., Row3 = (Row4 + 5.6) × Row 7/15) that refer to data in worksheet rows or columns; stores the results in specified rows or columns of the same worksheet.

2 **Logical Expressions (AND):** Provides for defining and evaluating arbitrary logical expressions involving the AND operator (e.g., Set Row7 = 1 if Row3 is greater than 5.6 AND Row4 is less than 13.6) that refer to data in worksheet rows or columns; stores the results in specified rows or columns of the same worksheet.

3 **Logical Expressions (OR):** Provides for defining and evaluating arbitrary logical expressions involving the OR operator (e.g., Set Row7 = 1 if Row3 is greater than 5.6 OR Row4 is less than 13.6) that refer to data in worksheet rows or columns; stores the results in specified rows or columns of the same worksheet.

4 **Special Calculations:** Suboptions provide for the following:
   1. Replaces all missing values in a (selected portion of a) row or column with a specified constant.
   2. Replaces each occurrence of a selected constant in a (specified portion of a) row or column with a missing value.
   3. For each cell in a (specified portion of a) row (or column), computes the cumulative sum of row (or column) cell values up to that one; stores the results in corresponding cells of a specified row (or column). That is, for each k, computes \( \text{SUM}(x_i) \) where \( i=j \) to \( k \).
   4. For each cell in a (specified portion of a) row (or column), computes the cumulative product of row (or column) cell values up to that one; stores the results in corresponding cells of a specified row (or column). That is, for each k, computes \( \text{PRODUCT}(x_i) \) where \( i=j \) to \( k \).
CALCULATE (continued)

5. **Time Difference**: Calculates differences between two dates, two times, or two date-time pairs in rows or columns of a worksheet. Results, expressed in time units ranging from minutes to years, are stored in a row or column of the same worksheet.

6. **Statistical Calculations**: Calculates statistics for data in specified rows (or columns) of a worksheet and stores the result in the corresponding rows (or columns) of a specified column (or row) of the same worksheet. For example, if means are computed for columns 5 through 10, the results might be stored in columns 5 through 10 of row 20. The available statistics are:
   1. Counts of non-missing values
   2. Sums
   3. Means
   4. Standard Deviations
   5. Standard Errors

**DESCRIBE**

1. **Edit the Schema**: Provides for constructing, editing, and reviewing the schema for a particular study.

2. **Print a Formatted Schema**: Produces a specially formatted schema listing on the system printer. This listing is easier to read than those produced using option 1.

3. **Create/Delete Study Files**: Creates study files (SDF and patient file) for trial data entry and retrieval after the user constructs a schema; after the user enters a modest amount of data, provides for deleting the study files in order to enable the user to revise the schema.

**ENTER**

0. **Add New Patient Abbreviations into the Study**: Provides for entering the patient abbreviations that will be used throughout the study. A patient can be referenced only after his abbreviation has been entered in this manner.

1. **Enter Data into the Update File**: In response to prompts, the user types patient abbreviations, data collection dates and times, and item values for direct entry into the Update File. Values are screened for validity as they are typed. The system manager must run `merge` to store the values in the SDF.

2. **Copy and Screen Data from a Worksheet into the Update File**: Provides for specifying the data in an existing worksheet that are to be copied into the Update File. Values are screened for validity as the system stores them in the Update File. The system manager must run `merge` to store the values in the SDF.
3 **Review Data in the Update File:** Provides for reviewing data values prior to running *merge*. Values may be reviewed by patient or in the entered sequence. Review may start with the data entered first (and go forward) or with the most recently entered data (and go backward). Entire panels may be deleted.

**FILES**

1 **List Worksheets:** Lists worksheet names and other characteristics ordered by
   - 1. Filename, i.e., worksheet name
   - 2. Size
   - 3. Date and time last modified
   - 4. Date of last access
   - 5. Unordered

2 **List Subsets:** Lists Subsets in the orders shown above.

3 **List Response Files:** Lists Response Files in the orders shown above.

4 **List Utility Files:** Lists utility files, i.e., Communication Files, Patient File, Schema File, SDF, and Update File, in the orders shown above.

**RESPONSE FILES**

1 **Execute a Response File:** The system reads responses from the specified file, executes some as if entered by the user, and (where a question mark appears in the file) prompts the user for others.

2 **Create a Response File:** After entering this subactivity, the user proceeds to use the system in a normal manner. The system records his responses and stores them in a Response File.

3 **Edit a Response File:** Provides for modifying a Response File created with option 2 to add, delete, or change responses, to list the file, to indicate cycling (i.e., looping) within the Response File, or to change some responses to question mark (thereby indicating that the user should be prompted for them during execution).

4 **Discard a Response File:** Discards a Response File that is no longer useful.

5 **Copy a Response File:** Copies a Response File and names the copy. The copy can then be edited to create a variation of the original.

6 **Append a Response File:** Appends one Response File to another to create a longer one. This avoids the necessity of creating a large Response File (with option 2) in a single step.
**RETRIEVE**

0 **Select a Worksheet**: Provides for specifying which existing worksheet is to be the current one, i.e., which is to be used in options 4 and 5; also provides for creating a new worksheet.

1 **Event Display**: Displays information about events for individual patients. This information is the name, type, date, and time associated with the event and the name and value of the triggering item.

2 **Study Date File to Worksheet**: Provides for retrieving numeric data about a specific patient, item, or time from the SDF and for storing them in a new worksheet. A time interval (relative to an event or an absolute date and time) and a time unit (which determines the resolution of the retrieved data) may be specified. Retrieved values may be restricted to those that satisfy specified conditions.

3 **Study Data File to Terminal/Communication File**: Numeric and textual data for specified patients, panels, and times are displayed and/or are stored in a Communication File, a panel at a time. All values are identified by patient, panel, time, and item name. The Communication File may be processed by a BASIC program, e.g., to produce a report or to transfer the data to another computer.

4 **A Worksheet to a Worksheet**: Copies a portion of a worksheet and stores the copied data and labels in specified locations of either the same worksheet or a different worksheet. This is useful for properly positioning data to be analyzed and for building a summary worksheet from several others.

5 **Worksheet to Communication File**: Copies the contents of a worksheet into a Communication File and associates that file with the same study, a different study, or an account in which BASIC programs can be written. Generally, a Communication File is processed by a specially written BASIC program; it may be transferred to another computer via magnetic tape. This option may be used together with option 6 to copy entire worksheets or to move worksheets from one study to another.

6 **Communication File to Worksheet**: Copies the contents of a properly formatted Communication File (in the same study, a different study, or an account in which BASIC programs can be written) into a new worksheet. The Communication File may be created by option 5 or by a BASIC program. It may contain data transferred from another computer.
SUBSET

1. **A List of Patient Abbreviations**: Provides for creating a patient set by naming its constituent patients; useful when specific inclusion criteria cannot be stated.

2. **Patients Belonging to All of a List of Sets**: Creates the set of patients who are members of all of one list of selected existing sets and who are not members of any of another list.

3. **Patients Belonging to at Least One of a List of Sets**: Creates a set of patients who are members of any one of a list of selected existing sets or who are not members of at least one set in another list.

4. **Patients Whose SDF Data Satisfy All Conditions in a List**: A time interval is specified and a set of conditions is placed on item values; SDF data within the time interval are examined and a patient becomes a member of the new set if these data satisfy all the conditions.

5. **Patients Whose SDF Data Satisfy Any of the Conditions in a List**: A time interval is specified and a set of conditions is placed on item values; SDF data within the time interval are examined and a patient becomes a member of the new set if any of the conditions is satisfied by the data.

6. **Patients Whose Worksheet Data Satisfy All Conditions in a List**: Creates a set of patients whose data in a selected worksheet satisfy all the stated conditions.

7. **Patients Whose Worksheet Data Satisfy at Least One Condition in a List**: Creates a set of patients whose data in a selected worksheet satisfy any of the stated conditions.

8. **List Patient Abbreviations from a Set**: Lists all the members of a selected set; all patients on the study are listed if the set called all is specified.

9. **List All of Your Set Names**: Lists the name, creation date, number of members, and description of every patient set associated with the study.

10. **Discard a Particular Set**: Discards a specified set that is no longer useful.

WORKSHEET

0. **Select a Worksheet**: Creates a new worksheet having the user-specified name, title (i.e., descriptive commentary), number of rows, and number of columns; also provides for specifying the current worksheet, i.e., the one that is to be subsequently processed.
WORKSHEET (continued)

1. **Display This Worksheet**: Displays up to 20 rows and 6 columns of the current worksheet, starting with a specified row and column. Numeric codes for categorical items described in the schema may be displayed as character strings.

2. **Label Worksheet Rows**: Provides for labeling rows with words up to 8 characters long. Existing labels may be changed. Labels may be blanked out by entering three dots (\ldots\).

3. **Label Worksheet Columns**: Provides for labeling columns in the same manner as labeling rows with option 2.

4. **Enter Data by Rows**: Provides for entering data, a row at a time, into a specified range of worksheet rows. There is no distinction between entering new values and replacing existing ones. Character strings may be entered for categorical items described in this schema. Dates are entered in the form 10/27/72. Missing values are entered (or existing values are deleted) by typing three dots (\ldots).

5. **Enter Data by Columns**: Provides for entering columns of data into a worksheet. This option is otherwise similar to option 4.

6. **Enter Data by Cells**: Provides for entering data into specified cells of the current worksheet; useful when only a few values are to be entered or changed. This option is otherwise similar to option 4.

7. **Discard This Worksheet**: Discards a worksheet that is no longer useful.

8. **Sort This Worksheet**: Reorders worksheet rows by sorting the values in a column or reorders columns by sorting the values in a row. This is a very useful function because it can be used to group patients and their data according to values of one or more variables. *Analyze* and *calculate* subactivities can then be applied to the resultant individual groups of data.

9. **List Your Worksheets**: Lists the names, titles, created and last modified dates, and sizes of all worksheets.

10. **Print This Worksheet**: Prints the entire current worksheet on the system line printer.

11. **Edit This Worksheet**: Provides for changing the current worksheet's name and title and for adding, deleting, moving, and copying rows and columns within the current worksheet. This function is particularly useful for enlarging the worksheet so that more data can be added.
DATA ORGANIZATION AND FILES

Data stored by the CLINFO system are organized to facilitate the management and analysis of the clinical research study. The study is a single experiment, based on an approved protocol and carried out by one or more investigators. An investigator may have more than one study active on the CLINFO system, but each is completely separate and distinct. Even if the same subject (i.e., patient or normal volunteer) happens to be on two different studies, data relating to him must be entered separately for each study, and access to each study is obtained by a different study name. Thus any single interactive session, from log on to log off, is carried out in the context of a single study. Communication Files may be used to transfer data from one CLINFO study to another.

The CLINFO system stores data in two types of data file. Each study has a single, carefully protected, highly structured collection of data called the Study Data File (SDF). Data are entered into the SDF under control of the schema, and data to be examined or analyzed are retrieved from the SDF. CLINFO users may also create, modify, and discard files called worksheets, as required. Data may be entered directly into a worksheet or may be retrieved from the SDF or another worksheet. Data must be stored in a worksheet in order to be analyzed.

Study Data File

The SDF of a study is immediately accessible as long as the study is active; it is also copied onto a permanent medium each day to ensure against loss. The SDF may be thought of as having three dimensions: patient, item (or panel), and time of data collection. Measurements for any item or set of items may be repeated many times or only a few times, in which case the SDF may be large but sparse. Storage of data in an SDF permits great flexibility in data entry and retrieval.

Data are entered into the SDF by means of the enter activity, which utilizes the schema to prompt for item values and to check the data against user-specified ranges of values or codes and which stores the time of data collection along with each entry. Data are retrieved from the SDF by means of the retrieve activity.

Worksheets

Data may be extracted from the SDF into CLINFO worksheets in order to carry out analyses, review data, or produce reports, graphs, and tables. Data in worksheets may be copied from them and stored in the SDF. The combination of extracting data from the SDF into worksheets and copying worksheet data into the SDF provides for minimizing the number of worksheets that must be saved. It also provides for restructuring worksheets by entering worksheet data into the SDF, then retrieving these data from the SDF into new worksheets. For example, data from worksheets that apply to individual patients can be moved (through the SDF) into a worksheet that stores time-varying values of an item for a group of patients.

Worksheets are rectangular, tabular representations of study data, which are useful for the review of data and are required for data analysis. As worksheets are created by the user, they are given names by which they may be recalled, long titles which describe them, dimensions (numbers of rows and columns), and, optionally, row and column labels.
Although the data stored in worksheets are numeric, a decoding facility, which works in conjunction with the schema, permits the display of character-string values rather than the stored numeric codes.

Data may be retrieved from the SDF into a worksheet, or they may be entered in five other ways:

- By entering data directly from the terminal keyboard.
- By copying portions of one worksheet into portions of another.
- By moving and copying rows or columns within a worksheet.
- By storing results computed by analyze or calculate activities.
- By retrieving data that have been processed by a BASIC program and stored in a Communication File.

Worksheets are automatically saved by the system when they are created or modified, but they are not protected in the same way as the SDF; i.e., worksheets may be easily modified or deleted by the user. All of the data analysis programs expect the data to be in rows or columns of an existing worksheet. If a worksheet happens to be arranged so that each row represents a series of values of different variables for a single patient, then an analysis would probably be performed on individual columns because each would contain values of a single variable. In any case, the user informs the system whether rows or columns are to be analyzed and then designates, by number, the row or column to be analyzed.

In addition to the SDF and a collection of worksheets, a number of other files are used by CLINFO both for internal housekeeping and for direct use.

**Schema File**

A Schema File contains descriptions of the data to be collected for a particular study. This file is initially entered by the user and is referenced by the system for all data entry, retrieval, and subsetting functions that involve the SDF, as well as for decoding numeric values for display.

**Update File**

This file acts as a temporary, protective storage area for data to be entered into the SDF. This temporary file allows the updating of the SDF, which may be time-consuming, to be carried out when it will not interfere with normal access to the system.

**Patient File**

The Patient File is a list of patients, in order of their entry into the study, containing the patient abbreviation, i.e., the identifier by which the patient is known to the system, and information about the location of the patient's data in the SDF.

**Subset File**

The Subset File contains lists of abbreviations of patients who belong to the patient sets that the user has defined (i.e., groups of patients who have particular
characteristics or values of measured variables in common, where the variables and their values are specified by the user; this file also contains information that permits the system to locate the SDF data referring to patients in the sets.

Communication Files

These files act as buffers between the CLINFO prototype and non-CLINFO programs. They provide a link to the BASIC programming language while protecting the CLINFO processor and files from BASIC procedures. Communication Files permit data values to be transmitted from the SDF or from worksheets to BASIC programs; they can also be used to transmit data from non-CLINFO processes (e.g., data generated by BASIC programs) to CLINFO worksheets and from one study to another. In addition, a Communication File can receive data from and transmit data to a system device (e.g., a magnetic tape) that can be taken to another computer for processing.

Response Files

These files are executable lists of responses to CLINFO prompts. When a Response File is executed, the prototype behaves as if the responses were entered by the user. This is a convenient vehicle for storing and later executing a sequence of responses that is repeated often, perhaps with variations. The user creates Response Files when they are needed and discards them when they are no longer needed.

THE CLINFO USER'S TERMINAL

The user interacts with the CLINFO prototype through a CRT terminal with an attached printer. The terminal can display up to 40 lines of 80 characters each on its TV-like screen. Information and commands are transmitted to the system by use of a typewriter-like keyboard, which is described in greater detail below. The small, attached DTC printer is used for copying the contents of the screen onto paper.

The Ann Arbor Terminal

Switch Settings. The power switch is located on the right-hand side of the AA terminal and is "on" in the forward position. The two switches on the rear panel of the terminal labeled S1 and S3 should be in the right-hand position when viewed from the rear, corresponding to line and full duplex, respectively. Switch S1 may be switched to LOCAL to modify the screen's contents before copying them onto the printer, but it must be switched back to LINE to use CLINFO. Other switches on the rear panel are inoperative. The system manager may adjust an internal switch to set the terminal's operating speed to 1200 baud or 4800 baud (i.e., to 120 or 480 characters per second).

The Keyboard. Throughout this report, the keys on the terminal are referred to by enclosing the appropriate symbol in braces; e.g., [a] refers to the lower-case "a" key.
In addition to the standard typewriter-like keys and controls, the AA keyboard has some keys and functions that are specific to the CLINFO prototype and some that are not used in this application:

- The keys labeled in yellow with abbreviated activity names are used in conjunction with the yellow CLINFO key to initiate the activities.
- The control characters above most of the alphabetic keys (such as DC4 over the letter T) are disregarded by the CLINFO system.
- The 10 numeric keys and the | on the right-hand side of the keyboard may be used instead of the upper row of numeric keys. They are not affected by the SHIFT key.
- The key used for exponentiation in BASIC is \(^{\wedge}\). This is displayed as a "hat" on the screen.
- The four keys containing arrows, and the HOME, TAB, and LINE FEED keys are disabled.
- The TTY LOCK key (which lights when activated) has the same function as the SHIFT LOCK, but only for alphabetic keys. All other keys transmit their lower-case character, e.g., ^| and ^|, when TTY LOCK is on.
- The yellow CLINFO key is used to enable the escape function; that is, to log onto the CLINFO system or to terminate an ongoing activity, the operator holds down the CLINFO key while pressing the ESC key. This arrangement prevents the operator from accidentally hitting [ESC] and aborting an ongoing process. When the CLINFO key and a key labeled in yellow with an abbreviated activity name are simultaneously pressed, this is equivalent to typing the label.
- The PRINT key above the 10-key numeric pad is used to command the attached printer to copy the contents of the CRT screen. Printing is also activated by simultaneously pressing the CLINFO key and the key labeled PRINT in yellow.
- The BRK key halts the print function and returns control to the user.
- The RETURN key is used to complete all user responses; depressing the RETURN key informs the system that the response is to be read and acted upon. In many of the system dialogues, the choice of responses includes simply striking RETURN, usually to indicate the most common or expected choice.
- The BS key backspaces and deletes successive characters from the screen. Unintentional or incorrect responses may be corrected by backspacing and thereby deleting characters with the BS key, then retyping the correct response, prior to pressing the RETURN key.

**Repeat Function.** The alphabetic, numeric, and BS, RETURN, and ESC keys produce repeated characters if held down more than about a second. This can cause unexpected behavior, because a series of RETURN, or other, characters may be stored by the computer if the keys are held down; the system will then respond to the stored characters as if they had just been entered by the user. If the system does not respond promptly to a RETURN, be patient; a second RETURN will not speed the response and will probably produce an undesirable effect on the following prompt when the system does respond. Instead, use the space bar or BS key to see if the system is alive.
Getting Ahead of the System. As mentioned above, the characters issued by the keyboard are stored and acted upon in sequence by the system. This feature can be used to your advantage, since familiar sequences of commands and responses may be entered more quickly than the system can act on them; thus you do not have to wait for a familiar prompt but may enter your response and wait for the system to catch up. On the other hand, this can be dangerous because the system may behave in an unexpected manner if an inappropriate or incorrect response has been entered. If you enter more characters than the system can handle, it will "beep" to indicate that it has missed a character; wait until the next prompt to see where you are. As mentioned above, sequences of commands may also be stored in a Response File.

The DTC Printer

A DTC printer, resembling a typewriter without a keyboard, is attached to the terminal. The information displayed on the CRT screen is typed on this printer when the PRINT key on the terminal is pressed. Thus, hard copies of worksheets, analytic results, plots, and lists of files are immediately available. Also, connected lines (which cannot be displayed on the CRT) can be printed on the printer by the Scatter Plot function. While the printer is in operation, the keyboard is disabled; however, pressing the BRK key will halt the printer and reactivate the keyboard.

The DTC printer is turned on by pressing its leftmost switch, i.e., the one labeled POWER. The LINE/LINE LOCAL switch should be in LINE mode for normal use, and the PITCH rocker switch should be set at 10. The three switches labeled CR (carriage return), LF (line feed), and FF (form feed) will operate only with the LINE/LOCAL switch in LOCAL.

After paper has been loaded into the printer and aligned, the horizontal perforation should be brought about an inch above the top of the platen, then the RESET switch should be pressed. This will establish the top-of-page position. To advance the paper to a new page, before pressing the PRINT key on the terminal, press the LINE/LOCAL key on the printer, then FF, then LINE/LOCAL again to return the printer to line mode. Occasionally, particularly when the printer is first turned on or after the BRK key has been pressed to stop printing, the printer will fail to operate properly. This is usually corrected by pressing the top of the printer's RESET switch. If that does not work, turn the terminal power off and on and try again. If that does not solve the problem, inform the system manager.
### PANEL 1: History

<table>
<thead>
<tr>
<th>#</th>
<th>Name</th>
<th>Description</th>
<th>Units</th>
<th>Type</th>
<th>Range/ # Codes/TextLabel</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Case No.</td>
<td>Case No.</td>
<td>%</td>
<td>num</td>
<td>1.20</td>
</tr>
<tr>
<td>2</td>
<td>Age</td>
<td>Age</td>
<td>yr</td>
<td>num</td>
<td>18.35</td>
</tr>
<tr>
<td>3</td>
<td>Sex</td>
<td>Sex</td>
<td>#</td>
<td>char</td>
<td>code (2)</td>
</tr>
<tr>
<td>4</td>
<td>%IwIw</td>
<td>% Ideal weight</td>
<td>%</td>
<td>num</td>
<td>90.105</td>
</tr>
<tr>
<td>5</td>
<td>Dur Diab</td>
<td>Duration of diabetes</td>
<td>yr</td>
<td>num</td>
<td>5.25</td>
</tr>
<tr>
<td>6</td>
<td>Ins Req</td>
<td>Daily insulin required</td>
<td>U</td>
<td>num</td>
<td>25.60</td>
</tr>
</tbody>
</table>

### PANEL 2: Control Data

<table>
<thead>
<tr>
<th>#</th>
<th>Name</th>
<th>Description</th>
<th>Units</th>
<th>Type</th>
<th>Range/ # Codes/TextLabel</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>Insulin</td>
<td>Control Insulin</td>
<td>U</td>
<td>num</td>
<td>25.60</td>
</tr>
<tr>
<td>8</td>
<td>Glucose</td>
<td>Control Glucose</td>
<td>mg/100 ml</td>
<td>num</td>
<td>25.325</td>
</tr>
<tr>
<td>9</td>
<td>Glucose</td>
<td>Control Glucose</td>
<td>mg/100 ml</td>
<td>num</td>
<td>0.200</td>
</tr>
<tr>
<td>10</td>
<td>Alanine</td>
<td>Control Alanine</td>
<td>mM</td>
<td>num</td>
<td>1.8</td>
</tr>
<tr>
<td>11</td>
<td>Lactate</td>
<td>Control Lactate</td>
<td>mM</td>
<td>num</td>
<td>1.2</td>
</tr>
</tbody>
</table>

### PANEL 3: Treatment Data

<table>
<thead>
<tr>
<th>#</th>
<th>Name</th>
<th>Description</th>
<th>Units</th>
<th>Type</th>
<th>Range/ # Codes/TextLabel</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>Insulin</td>
<td>Treatment Insulin</td>
<td>U</td>
<td>num</td>
<td>25.60</td>
</tr>
<tr>
<td>13</td>
<td>Glucose</td>
<td>Treatment Glucose</td>
<td>mg/100 ml</td>
<td>num</td>
<td>25.325</td>
</tr>
<tr>
<td>14</td>
<td>Glucose</td>
<td>Treatment Glucose</td>
<td>mg/100 ml</td>
<td>num</td>
<td>0.200</td>
</tr>
<tr>
<td>15</td>
<td>Alanine</td>
<td>Treatment Alanine</td>
<td>mM</td>
<td>num</td>
<td>1.8</td>
</tr>
<tr>
<td>16</td>
<td>Lactate</td>
<td>Treatment Lactate</td>
<td>mM</td>
<td>num</td>
<td>1.2</td>
</tr>
</tbody>
</table>

### EVENTS

<table>
<thead>
<tr>
<th>#</th>
<th>Name</th>
<th>Description</th>
<th>Item</th>
<th>Modifier</th>
<th>Trigger</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Insulin</td>
<td>Ctrl insulin stopped</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Insulin</td>
<td>Treatment insulin stopped</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Fig. III-2— The investigator's schema as listed by CLINFO after compilation. The panel, item, and event numbers have been assigned by the schema compiler.
Fig. III-3—The dialog between the system and the user as new data are entered for the patient identified as case 1. After establishing the context, the user chooses the History panel and enters the 6 items that it contains. At the next prompt of ?, the user changes the context and is prompted to verify the entire context. He maintains the same patient and date but changes the time to 615, i.e., to 6:15 A.M. The prompt at the bottom reminds the user of the acceptable responses to ?.
III. EXAMPLE OF A CLINFO SESSION

This section presents a simple but realistic example of the use of the CLINFO prototype. In this example, we reference the CLINFO activities employed at each step in the procedure but do not describe them in detail. Generally, CLINFO displays information on the CRT screen in upper- and lower-case or all upper-case letters; prompts (i.e., requests for information from the user) are in upper-case letters, followed by a colon (:). The user responds with lower-case letters.

In this example, the role of glucagon in the pathogenesis of diabetic ketoacidosis in man will be evaluated by investigating the effect of suppression of glucagon secretion by somatostatin. We shall examine changes in plasma glucose, glucagon, alanine, and lactate concentrations after acute withdrawal of insulin from 37 patients with juvenile-type diabetes. The example is derived from a study reported by Gerich et al. in "Prevention of Human Diabetic Ketoacidosis by Somatostatin—Evidence for an Essential Role of Glucagon," New England Journal of Medicine, Vol. 292, No. 19, May 8, 1975, pp. 985-989.

We assume that the investigator has defined the protocol for the study. It involves about 40 patients, with a small amount of identifying, clinical, and historical information recorded for each. In addition, each patient is subjected to a treatment situation and a control situation in which a series of measurements are made on hourly blood samples following acute withdrawal from insulin. The treatment series is measured while somatostatin is infused at a rate of 500 micrograms/hour; the control series is measured while a 0.9 percent saline solution is infused.

The investigator's data collection form might look like Fig. III-1. Each patient would have two such forms: one with history data and the laboratory data for the control situation, and the other with only the laboratory data from the treatment samples.

The investigator, who already has a personal password, obtains an identifier for this new study from the system manager and authorizes other members of his study team—fellows, technicians, and/or secretary—to access it. He then enters a schema into the system, using the describe activity. The schema is shown in Fig. III-2, as it is printed by the system printer. Names have been chosen for the panels and items, and limits and acceptable values have been specified for each item. It is assumed that each patient has several instances of panels called control and treatment because each panel instance corresponds to the measurements on a single blood sample; the instances are distinguished by their blood sample collection times. The notation is such that glucose refers to glucose in the control situation and glucoset refers to glucose in the treatment situation. The schema has been "compiled" and the system has assigned panel and item numbers.

The schema is next "locked," with the assistance of the system manager, and at this point, data entry can begin.

After the patient abbreviations (i.e., case 1 through case 37) have been entered for the 37 subjects, using option 0 of the enter activity, option 1 of enter is used to enter History panel data for each patient, as shown in Fig. III-3. The question marks and the items preceding colons are displayed by the system; the specifications or data following question marks or system-generated colons are entered by the user.
### Somatostatin Study

**Patient Name:**

**Case No.:**

**Date:**

**Age:**

**Sex:** M F

**Percent Ideal Weight:**

**Duration of Diabetes:** (yrs.)

**Insulin Requirements:** (U)

**Lab Data**

<table>
<thead>
<tr>
<th>Date</th>
<th>Control</th>
<th>Treatment</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Sample Time</th>
<th>Insulin Given</th>
<th>Glucose</th>
<th>Glucagon</th>
<th>Alanine</th>
<th>Lactate</th>
</tr>
</thead>
</table>

Fig. III-1—The investigator’s data collection form.

If the experimental data are available, they are entered by specifying the correct context (i.e., patient, date, and time) and the appropriate panel. Data entry continues as shown in Fig. III-4. The data, from samples collected at approximately 6 a.m. and 7 a.m. on May 4, 1975, are entered into the panel called control for the patient called case 1.

After entering all the data for case 1 and some of the data for case 2, the data enterer realizes that some of the interaction is unnecessary. That is, because data are already organized by patient in a lab book with history data followed by control and treatment data, and because both the control and treatment experiments result in a fixed number of blood samples that are all drawn on the same day, the flexibility that is normally provided by CLINFO’s prompting is not required. Prompting can be bypassed by storing responses (such as “History,” “date:,” “control,” and “” in a Response File that controls data entry. The Response File is produced by using option 2 of response files and then going through the steps of entering data for a patient (but hitting the RETURN key to enter “ ” instead of entering actual data values) to provide CLINFO with an example. The Response File that has been created by example is then modified somewhat, using option 3 of response files, and is executed by using option 1.
Fig. III-4— A continuation of the dialog as the data for the items in the control panel are entered. An out-of-range value (3.22) is caught by the system and reentered correctly. The context is then changed by time:700, and another set of data is entered. An error is corrected by entering the item name and the correct value in brackets; the system then reprompts for the next item.

Data entry goes more quickly when the Response File is used because the data enterer does not need to enter the panel name or hit the RETURN key to verify the current patient, date, or time. After some or all of the data have been entered, the user, when logging off, indicates to the system manager that he would like his data merged. The system manager then runs the sort/merge procedure to store the entered data in the SDF.

Using option 2 of the retrieve activity, the user creates a worksheet called CLINICAL, shown in Fig. III-5, in which the rows represent patients and the columns represent variables. To define the retrieval, the investigator indicates the following:

- He wants to create a patients-by-items worksheet.
- The data are not time-dependent (because, in this case, he wishes to retrieve only data in the History panel).
Fig. III-5—The first 20 rows of the CLINICAL worksheet. This worksheet contains the data entered into the History panel of each patient. Note that the column labels are the names of the items in the History panel, the row labels are patient identifiers, and the “decode” option was used to display the characters m and f in column 2 in place of the stored codes 0 and 1.

- Each row of the worksheet is to correspond to a member of a patient set.
- The particular set of interest is the one called all.
- Each column is to correspond to an item.
- The items of interest are age, sex, %idl wt, dur diab, and ins req.

The column labels of CLINICAL were automatically copied from the schema by the retrieve activity. Values of the item called sex are displayed as the characters m and f. The worksheet may be checked against the original data, and corrections entered by using the examine activity.

At this point, the user might further check the accuracy and reasonableness of his data by obtaining descriptive statistics, using option 1 of analyze (Fig. III-6), a frequency histogram of continuous data using option 8 of analyze (Fig. III-7), and perhaps a scatter plot of variables that he would expect to be consistently related, such as age and duration of diabetes (Fig. III-8). This latter method is excellent for detecting outliers which may represent erroneous data.

If a questionable data point appears, the user sorts the worksheet on the variable in question (a good technique for large worksheets). He may then determine which case is associated with the largest (or smallest) value of the variable, and perhaps check back to the original data.
COL DATA FROM WORKSHEET CLINICAL CREATED 12/17/75 MODIFIED 12/17/75

TITLE Clinical Characteristics of Diabetic Patients

<table>
<thead>
<tr>
<th>COL#</th>
<th>Label</th>
<th>N</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Range</th>
<th>Mean</th>
<th>Median</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>age</td>
<td>37</td>
<td>16</td>
<td>35</td>
<td>17</td>
<td>26.646</td>
<td>27</td>
</tr>
<tr>
<td>2</td>
<td>sex</td>
<td>37</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>.54054</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>%d1 wt</td>
<td>37</td>
<td>90</td>
<td>105</td>
<td>15</td>
<td>97.1243</td>
<td>98</td>
</tr>
<tr>
<td>4</td>
<td>%d2 diab</td>
<td>37</td>
<td>5</td>
<td>25</td>
<td>20</td>
<td>12.4865</td>
<td>12</td>
</tr>
<tr>
<td>5</td>
<td>ins req</td>
<td>37</td>
<td>25</td>
<td>60</td>
<td>35</td>
<td>39.5946</td>
<td>40</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>COL#</th>
<th>Std Dev</th>
<th>Sum</th>
<th>Sum Squares</th>
<th>Variance</th>
<th>Std Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4.82026</td>
<td>986</td>
<td>27112</td>
<td>23.2349</td>
<td>.79245</td>
</tr>
<tr>
<td>2</td>
<td>.50523</td>
<td>20</td>
<td>20</td>
<td>.25526</td>
<td>.63806</td>
</tr>
<tr>
<td>3</td>
<td>3.88872</td>
<td>3601</td>
<td>3.509875+05</td>
<td>14.5064</td>
<td>.62615</td>
</tr>
<tr>
<td>4</td>
<td>5.41222</td>
<td>462</td>
<td>6826</td>
<td>29.3679</td>
<td>.89991</td>
</tr>
<tr>
<td>5</td>
<td>9.2698</td>
<td>1465</td>
<td>61099</td>
<td>85.9149</td>
<td>1.52382</td>
</tr>
</tbody>
</table>

Fig. III-6—The output of option 1 of analyze, displaying descriptive statistics for the five variables in the worksheet called CLINICAL. Observing the maximum and minimum verifies the expected limits for each variable.

<table>
<thead>
<tr>
<th>Cut Pt.</th>
<th>Freq</th>
<th>%</th>
<th>Cum %</th>
<th>10</th>
<th>20</th>
<th>30</th>
<th>40</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>3</td>
<td>8.1</td>
<td>100.0</td>
<td>***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>27</td>
<td>71.9</td>
<td>99.9</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>1</td>
<td>2.7</td>
<td>100.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Max = 35 Min = 18 Mean = 26.649 S. D. = 4.8203

Fig. III-7—Histogram, option 8 of analyze. Knowledge of the sample distribution might avoid data analysis problems.

Successive use of option 2 of the retrieve activity also creates a series of patient "flowsheets" (i.e., one worksheet for each patient) such as that shown in Fig. III-9. To specify the retrieval of such a worksheet, the investigator states that:

- He wants to create a times-by-item worksheet for a patient.
- The worksheet is for a particular patient (in this example, for case 1).
- The rows correspond to different times.
- The appropriate time unit is hour.
- The times are specified as a list (of hours) relative to an event.
- The relative hour numbers are 0, 1, 2, 3, and 10.
- The event that establishes the base time is instopc.
- Columns correspond to dates, times, and items in panels.
- The panel of interest is the one called control.
Fig. III-8— A scatter plot of age vs. duration of diabetes created using option 7 of analyze. Here the user would expect no points to lie below the diagonal, since age must be greater than duration of diabetes in each case.

WORKSHEET CASE1C
TITLE Flowsheet for case1, control data
CREATED 12/17/75  MODIFIED 12/17/75
# OF ROWS 9  
# OF COLS 7

ROWS/COLS 1 2 3 4 5 6

LABELS DATE TIME insulin glucose glucgonc alaninc
1 HR -2 5/4/1975 615 35 103.500 89.300 .322
2 HR -1 5/4/1975 700 35 97.800 90.200 .324
3 HR 0 5/4/1975 810 35 96 90.500 .318
4 HR 1 5/4/1975 917 --- 150.190 108.300 .298
6 HR 4 5/4/1975 1205 --- 257.800 127.300 .307
7 HR 6 5/4/1975 1400 --- 264.600 131.400 .335
8 HR 8 5/4/1975 1612 --- 283.700 143.800 .378

(NOTE: ... indicates missing values)

Fig. III-9— The first six columns of worksheet CASE1C, created by retrieving control panel data for case 1, by hour, and aligning the times by use of the event inscope. A similar worksheet is prepared for each patient. The rows in the different worksheets then represent measurements taken at comparable relative times.
In the first resulting worksheet (or "flowsheet"), which the investigator titles CASE1C, the row labels denote times relative to the specified event, the first two columns list the dates and times that insulin was given and samples were collected for the patient, and the remaining five columns (the last of which does not appear on the display) itemize the data values. Note that HR 0 corresponds to the last value of insulin (insulin-control case); this is because the event instopic, as defined in the schema (see Fig. III-2), occurs at the same time as the last value of instopic. When the investigator similarly retrieves data for the other patients, he observes that the dates and times of sample collection are different for each one. However, because each worksheet has values for the same relative time stored in the same row (e.g., in every patient worksheet, row 5 corresponds to the second hour after insulin withdrawal), data values for all the patients can be easily time-aligned. Although the patient worksheets are convenient for examining and verifying collected data by patient, the investigator realizes that there is a simpler way to organize his data for purposes of analysis.

Using the retrieve activity again, the investigator constructs a worksheet for each variable, such as that shown for glucose in Fig. III-10. Each row in the

<table>
<thead>
<tr>
<th>WORKSHEET GLUCOSEC</th>
</tr>
</thead>
<tbody>
<tr>
<td>TITLE Control Glucose by Patient and by Hour</td>
</tr>
<tr>
<td>CREATED 12/17/75 MODIFIED 12/17/75</td>
</tr>
<tr>
<td>OF ROWS 37 OF COLS 9</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ROWS/COLS</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>LABELS</td>
<td>HR -2</td>
<td>HR -1</td>
<td>HR 0</td>
<td>HR 1</td>
<td>HR 2</td>
<td>HR 4</td>
</tr>
<tr>
<td>1 case 1</td>
<td>103.500</td>
<td>97.800</td>
<td>96</td>
<td>156.100</td>
<td>202.300</td>
<td>257.800</td>
</tr>
<tr>
<td>2 case 2</td>
<td>96.800</td>
<td>...</td>
<td>83.800</td>
<td>...</td>
<td>185.400</td>
<td>144.200</td>
</tr>
<tr>
<td>3 case 3</td>
<td>101.300</td>
<td>95.400</td>
<td>92.800</td>
<td>147.900</td>
<td>199.200</td>
<td>254.600</td>
</tr>
<tr>
<td>4 case 4</td>
<td>95.400</td>
<td>84.300</td>
<td>87.400</td>
<td>...</td>
<td>191.300</td>
<td>249.500</td>
</tr>
<tr>
<td>5 case 5</td>
<td>111.700</td>
<td>109.800</td>
<td>108.100</td>
<td>157.700</td>
<td>211.600</td>
<td>265.700</td>
</tr>
<tr>
<td>6 case 6</td>
<td>106.500</td>
<td>94.800</td>
<td>96.200</td>
<td>146.400</td>
<td>202.300</td>
<td>254.800</td>
</tr>
<tr>
<td>7 case 7</td>
<td>90.800</td>
<td>87.500</td>
<td>83.800</td>
<td>148.200</td>
<td>185.600</td>
<td>147.200</td>
</tr>
<tr>
<td>8 case 8</td>
<td>114.800</td>
<td>118.300</td>
<td>105.700</td>
<td>163.400</td>
<td>214.800</td>
<td>268.400</td>
</tr>
<tr>
<td>9 case 9</td>
<td>103.300</td>
<td>100.700</td>
<td>104.300</td>
<td>153.800</td>
<td>205.400</td>
<td>260.500</td>
</tr>
<tr>
<td>10 case 10</td>
<td>101.700</td>
<td>97.400</td>
<td>92.000</td>
<td>147.900</td>
<td>198.600</td>
<td>254.200</td>
</tr>
<tr>
<td>11 case 11</td>
<td>103.500</td>
<td>95.800</td>
<td>96</td>
<td>152</td>
<td>202.300</td>
<td>257.800</td>
</tr>
<tr>
<td>12 case 12</td>
<td>95.800</td>
<td>84.500</td>
<td>83.800</td>
<td>152</td>
<td>181.800</td>
<td>144.200</td>
</tr>
<tr>
<td>13 case 13</td>
<td>111.300</td>
<td>109.800</td>
<td>110.700</td>
<td>158.700</td>
<td>215.600</td>
<td>265.700</td>
</tr>
<tr>
<td>14 case 14</td>
<td>104.600</td>
<td>110.300</td>
<td>108.100</td>
<td>163.400</td>
<td>214.800</td>
<td>268.400</td>
</tr>
<tr>
<td>15 case 15</td>
<td>111.300</td>
<td>95.800</td>
<td>92.800</td>
<td>147.900</td>
<td>198.600</td>
<td>254.200</td>
</tr>
<tr>
<td>16 case 16</td>
<td>103.500</td>
<td>97.400</td>
<td>96.400</td>
<td>150.100</td>
<td>202.300</td>
<td>250.800</td>
</tr>
<tr>
<td>17 case 17</td>
<td>98.400</td>
<td>...</td>
<td>87.600</td>
<td>140.400</td>
<td>195.300</td>
<td>249.500</td>
</tr>
<tr>
<td>18 case 18</td>
<td>106.300</td>
<td>105.700</td>
<td>101.300</td>
<td>153.800</td>
<td>201.400</td>
<td>267.500</td>
</tr>
<tr>
<td>19 case 19</td>
<td>90.800</td>
<td>84.500</td>
<td>83.400</td>
<td>152.400</td>
<td>185.600</td>
<td>144.200</td>
</tr>
<tr>
<td>20 case 20</td>
<td>101.300</td>
<td>98.400</td>
<td>92.000</td>
<td>147.900</td>
<td>198.600</td>
<td>254.200</td>
</tr>
</tbody>
</table>

Fig. III-10—A portion of the GLUCOSEC worksheet. The first 20 rows contain all of the hourly glucose (control glucose) values for each of the first 20 patients. Data for the other 17 patients are in the worksheet but are not shown. Also, columns 7, 8, and 9, which contain data for HR6, HR8, and HR10, are not shown. A similar worksheet is constructed for each of the experimental variables, using the retrieve activity. Note that the data are now aligned by (relative) time and that missing data are easily spotted.
worksheet GLUCOSE contains data for a particular patient, and each column contains glucose data for a time relative to the event insilopic. The data for all patients are now aligned by time, and instances of missing data may be easily observed (and the data entered if available).

Statistical computations may now be performed on the worksheets containing the time-aligned data. In our example, the investigator wishes to plot the mean value of each variable (across patients) against time, so he creates a worksheet that contains the mean values. Starting with a worksheet like GLUCOSE, he first uses Edit This Worksheet (option 11 of worksheet) to enlarge it by one row (row 38). Next, using Statistical Calculations (option 6 of calculate), he computes the means of all columns and stores the results in row 38. Sums, standard deviations, and standard errors can be similarly computed and stored, whereas other statistics such as median and sum of squares can be computed using Descriptive Statistics (option 1 of analyze) but cannot be stored directly in the same worksheet.

Next, option 4 of retrieve is used to copy row 38, containing the mean values of glucose, to row 1 of a new worksheet called MEANS (see Fig. III-11). This

retrieved, 4 WORKSHEET TO WORKSHEET COPY
THE CURRENT SOURCE WORKSHEET IS:
WORKSHEET GLUCOSE
TITLE Control Glucose by Patient and by Hour
CREATED 12/17/75 MODIFIED 12/17/75
# OF ROWS 38
# OF COLS 9
TO SELECT ANOTHER SOURCE WORKSHEET, TYPE ITS NAME. ELSE [RETURN]:
THERE IS NO CURRENT DESTINATION WORKSHEET, PLEASE SELECT ONE: MEANS
THE CURRENT DESTINATION WORKSHEET IS:
WORKSHEET MEANS
TITLE Means of Control and Treatment Data by Hour and Variable
CREATED 12/17/75 MODIFIED 12/17/75
# OF ROWS 8
# OF COLS 13
TYPE c TO COPY COLUMNS OR r TO COPY ROWS:
r
SOURCE ROW RANGE..a,b...: 38,38
SOURCE COL RANGE..a,b...: 1,9
TYPE yes TO TRANSPOSE ROWS INTO COLS, ELSE [RETURN]:
ROWS WILL BE COPIED INTO ROWS
COLS WILL BE COPIED INTO COLS
DESTINATION STARTING ROW: 1
DESTINATION STARTING COL: 1
TYPE yes TO COPY ROW LABELS, ELSE [RETURN]:
TYPE yes TO COPY COLUMN LABELS, ELSE [RETURN]: yes

Fig. III-11—The dialog for using retrieve, option 4, to copy a row (38) from the GLUCOSE worksheet, which contains means of glucose for each hour, into the first row of the MEANS worksheet. Row 38 from each of seven successive source worksheets is copied into an appropriate row (2 through 8) of the same destination worksheet.
process is repeated for each of the variables (that is, row 38 computed in each of
the worksheets retrieved for the individual variables is copied into successive rows
of the MEANS worksheet).

Response Files can be used to automate this repetitive procedure. Starting with
Create a Response File (option 2 of response files), the user goes through the above
steps—retrieving a worksheet for a variable, enlarging the worksheet by one row,
computing and storing the means of all columns, and copying the new row into the
MEANS worksheet. Next, using Edit a Response File (option 3 of response files),
he changes each of the stored responses that changes from variable to variable
(namely the item name, the name of the retrieved worksheet, and the destination
row in MEANS) to "?" to indicate which responses are to be entered from the
terminal (instead of being read from the Response File). Figure III-12 shows the

RESPONSE FILE: fig10  THIS FILE DOES NOT CYCLE  

<table>
<thead>
<tr>
<th>LINE #</th>
<th>RESPONSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.00</td>
<td>!ret,2</td>
</tr>
<tr>
<td>2.00</td>
<td>2</td>
</tr>
<tr>
<td>3.00</td>
<td>glucosec</td>
</tr>
<tr>
<td>4.00</td>
<td>2</td>
</tr>
<tr>
<td>5.00</td>
<td>all</td>
</tr>
<tr>
<td>6.00</td>
<td>2</td>
</tr>
<tr>
<td>7.00</td>
<td>3</td>
</tr>
<tr>
<td>8.00</td>
<td>-2</td>
</tr>
<tr>
<td>9.00</td>
<td>10</td>
</tr>
<tr>
<td>10.00</td>
<td>instopc</td>
</tr>
<tr>
<td>11.00</td>
<td>yes</td>
</tr>
<tr>
<td>12.00</td>
<td>4</td>
</tr>
<tr>
<td>13.00</td>
<td>glucosec</td>
</tr>
<tr>
<td>14.00</td>
<td>Control Glucose by Patient and by Hour</td>
</tr>
<tr>
<td>15.00</td>
<td>yes</td>
</tr>
<tr>
<td>16.00</td>
<td></td>
</tr>
<tr>
<td>17.00</td>
<td></td>
</tr>
<tr>
<td>18.00</td>
<td></td>
</tr>
<tr>
<td>19.00</td>
<td>!wer,11</td>
</tr>
<tr>
<td>20.00</td>
<td>r</td>
</tr>
<tr>
<td>21.00</td>
<td>add 38</td>
</tr>
<tr>
<td>22.00</td>
<td>!calc,6</td>
</tr>
<tr>
<td>23.00</td>
<td>r</td>
</tr>
<tr>
<td>24.00</td>
<td>3</td>
</tr>
<tr>
<td>25.00</td>
<td></td>
</tr>
<tr>
<td>26.00</td>
<td>#38</td>
</tr>
<tr>
<td>27.00</td>
<td>means</td>
</tr>
<tr>
<td>28.00</td>
<td></td>
</tr>
<tr>
<td>29.00</td>
<td></td>
</tr>
<tr>
<td>30.00</td>
<td>!ret,4</td>
</tr>
<tr>
<td>31.00</td>
<td></td>
</tr>
<tr>
<td>32.00</td>
<td>means</td>
</tr>
<tr>
<td>33.00</td>
<td>r</td>
</tr>
<tr>
<td>34.00</td>
<td>38,38</td>
</tr>
<tr>
<td>35.00</td>
<td>1,9</td>
</tr>
<tr>
<td>36.00</td>
<td></td>
</tr>
<tr>
<td>37.00</td>
<td>1</td>
</tr>
<tr>
<td>38.00</td>
<td>1</td>
</tr>
<tr>
<td>39.00</td>
<td></td>
</tr>
<tr>
<td>40.00</td>
<td>yes</td>
</tr>
<tr>
<td>41.00</td>
<td>!res,3</td>
</tr>
</tbody>
</table>

Fig. III-12—A Response File used in retrieving values of particular vari-
ables, computing the means across patients, and storing the results in the
MEANS worksheet. To apply to more than one variable, the changes indi-
cated in the text must be made.
Response File for our example. Lines 31 through 40 contain the responses in Fig. III-11. To generalize, Response File lines 3, 14, 15, 27, and 37 should be changed to "?"; line 41 should be deleted, and the Response File should cycle back to line 1 to repeat the procedure for the next variable.

The first 6 columns of the MEANS worksheet are shown in Fig. III-13. Scatter plots of variable versus variable and variable versus time may be constructed with this worksheet.

To examine the time variation of mean variable values, the investigator uses Scatter Plot (option 7 of analyze) to create four plots, one of which is shown in Fig. III-14. Row 1 (control values of glucose) and row 5 (treatment values of glucose) of the MEANS worksheet are plotted against the column labels (time in hours from insulin withdrawal).

**WORKSHEET MEANS**
**TITLE** Means of Control and Treatment Data by Hour and Variable
**CREATED** 12/17/75  **MODIFIED** 12/17/75
**# OF ROWS** 8  **# OF COLS** 13

<table>
<thead>
<tr>
<th>ROWS/COLS</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>LABELS</td>
<td>HR -2</td>
<td>HR -1</td>
<td>HR 0</td>
<td>HR 1</td>
<td>HR 2</td>
<td>HR 4</td>
</tr>
<tr>
<td>glucose</td>
<td>103.400</td>
<td>99.750</td>
<td>97.186</td>
<td>152.363</td>
<td>201.457</td>
<td>242.957</td>
</tr>
<tr>
<td>glucone</td>
<td>89.104</td>
<td>90.235</td>
<td>98.613</td>
<td>108.425</td>
<td>119.680</td>
<td>127.324</td>
</tr>
<tr>
<td>alanine</td>
<td>.322</td>
<td>.325</td>
<td>.720</td>
<td>.298</td>
<td>.294</td>
<td>.316</td>
</tr>
<tr>
<td>lactate</td>
<td>1.418</td>
<td>1.362</td>
<td>1.503</td>
<td>1.455</td>
<td>1.186</td>
<td>1.271</td>
</tr>
<tr>
<td>glucose</td>
<td>117.327</td>
<td>107.800</td>
<td>106.905</td>
<td>75.245</td>
<td>69.400</td>
<td>77.620</td>
</tr>
<tr>
<td>glucone</td>
<td>98.704</td>
<td>109.753</td>
<td>118.105</td>
<td>68.300</td>
<td>45.810</td>
<td>46.302</td>
</tr>
<tr>
<td>alanine</td>
<td>.371</td>
<td>.361</td>
<td>.337</td>
<td>...</td>
<td>.254</td>
<td>.406</td>
</tr>
<tr>
<td>lactate</td>
<td>1.316</td>
<td>1.434</td>
<td>1.369</td>
<td>...</td>
<td>1.268</td>
<td>1.184</td>
</tr>
</tbody>
</table>

*(NOTE: ... indicates missing values)*

Fig. III-13—A portion of the destination worksheet, MEANS, showing the first 6 columns after data have been copied into it row by row. This worksheet is now used as input to the Scatter Plot subactivity to create plots of each variable versus time, using the column labels as values of the time variable.

To make his plots easier to interpret, the investigator indicates that he would like the printer attached to his CRT terminal to draw lines between the data points; the terminal itself cannot do this. Fig. III-15 was produced in this way by a printer. Apparently the somatostatin treatment changes plasma glucose level.

The investigator is also interested in the relationships between variables in both the treatment and control experiments following withdrawal of insulin. To explore these, he uses the Scatter Plot subactivity to plot one variable in columns 4 through 13 of MEANS against another. Figure III-16 illustrates the relation between mean plasma glucose and alanine levels after withdrawal of insulin in the control experiment; Fig. III-17 shows a similar relation in the treatment experiment. These figures demonstrate that at any given alanine level there was greater hyperglycemia when glucagon secretion was unrestrained by somatostatin.

Finally, to arrive at estimated mathematical descriptions of the relations illustrated in Figs. III-16 and III-17, the investigator uses Linear Regression (option 4
PLOT FROM WORKSHEET: MEANS
TITLE: Plasma Glucose after Acute Withdrawal of Insulin

Fig. III-14—The scatter plot using column labels of the MEANS worksheet to define the x-axis and rows 1 (glucosec) and 5 (glucoset) as the y-axis variables.

PLOT FROM WORKSHEET: MEANS
TITLE: Plasma Glucose after Acute Withdrawal of Insulin

Fig. III-15—The scatter plot (similar to Fig. III-14, but with dotted lines drawn by the printer).
Fig. III-16—The scatter plot using row 3 (alanine) as the x-axis variable and row 1 (glucose) as the y-axis variable.

Fig. III-17—The scatter plot using row 7 (alanine) as the x-axis variable and row 5 (glucose) as the y-axis variable.
of analyze). The results for the regression of mean glucose (from the MEANS worksheet) upon mean alanine in the treatment experiment are shown in Fig. III-18. The estimated slope and intercept are 408 and -84, respectively; the coefficient of correlation is 0.97, with a significance level of 0.005. For the control experiment, the estimated slope and intercept are 803 and -35, respectively; the coefficient of correlation is 0.79, with a significance level of 0.06. Finally, the regression line computed in Fig. III-18 is plotted in Fig. III-19.

The CLINFO prototype has several important features that are not illustrated by this example. In addition to entering data as shown in Figs. III-3 and III-4, a user may enter data directly into a worksheet and may transfer (using option 2 of enter and sort/merge) worksheet data (that have been entered or computed) into the SDF. Data in an Update File (where they are stored until the system manager runs sort/merge) may be reviewed using option 3 of enter; data in the SDF (where they are stored permanently) may be examined and changed using the examine activity. Sets of patients with common characteristics (e.g., all females older than 25, or all patients in a particular set who have glucosec greater than 200 between 1 and 4 hours after insulin is stopped) may be created by using one of the subset options. Finally, there are a number of statistical analyses available (as analyze options) that are not illustrated here, and special analyses and reports may be produced by writing computer programs in the BASIC language and transferring data in Communication Files (using options 3, 5, and 6 of retrieve).

---

**Fig. III-18**—The results of using the Linear Regression subactivity to compute the regression of columns 4 to 13 of row 5 (glucoset) in the MEANS worksheet upon row 7 (alaninet).
Fig. III-19—The regression line, together with the data plotted in Fig. III-17.
IV. HELPFUL HINTS

This section contains suggestions made by the designers and initial users of the prototype system. The topics covered are arranged in the order in which they are discussed in Sec. V. The suggestions relate to general topics such as data preparation, as well as to specific CLINFO activities.

We invite users of the CLINFO prototype system to contribute their own suggestions for easier and more efficient use of the system. Suggestions may be submitted to local system managers or sent directly to

The Rand Corporation
1700 Main Street
Santa Monica, Ca. 90406
Att: G. F. Groner

GENERAL

- The CLINFO prototype distinguishes between upper and lower case in most instances. It is wise to avoid unnecessary use of upper-case characters (thereby minimizing use of the SHIFT key) and to avoid panel and item names that differ only in capitalization.

- If you inadvertently depress the SHIFT LOCK or the TTY LOCK key, you may not be able to log on, enter data, find patient identifiers, etc., because upper-case characters are being entered where lower-case must be used. Note: The SHIFT LOCK key does not affect the numeric pad on the right-hand side of the terminal keyboard.

- The system responds to common errors (such as the entering of letters for numbers, or references to nonexistent columns of a worksheet) with messages that allow you to reenter the correct response.

- When unanticipated errors and temporary hardware failures occur, the system issues numbered error messages. Inform the system manager and explain the circumstances when these occur so that they may be corrected.

- A recent article (H. Wohl, “The Cusum Plot: Its Utility in the Analysis of Clinical Data,” New England Journal of Medicine, Vol. 296, No. 18, May 5, 1977, pp. 1044-1045), describes the utility of using cusum plots to analyze trends. Such plots are easily created with CLINFO by using calculate, option 1, to subtract the reference value, calculate, option 4, to compute the cumulative sum, and analyze, option 7, to plot the resulting cusum.

ANALYZE

- The Select a Worksheet option of the analyze activity allows you to change the current worksheet without leaving the activity.
A paper copy of the list of row (column) labels, provided by several of the analyze options, is very useful for reference when working with large worksheets. Hard copies provide handy documentation away from the terminal.

Before performing any statistical tests that assume a normally distributed population, you should use the Normality Test option (analyze, option 10) on your sample data. It will perform a Kolmogorov-Smirnov test of normality, and a warning message will be displayed if the sample is unlikely to have been drawn from a normally distributed population.

To restrict analyses to groups of data in a worksheet (e.g., to those for patients having particular data values), use worksheet, option 8, to sort on data that can be used to delimit the groups; use worksheet, option 1, to determine the bounding rows or columns; then indicate the bounds when using an analyze subactivity.

Sometimes data that lie within a single column must be moved into two or more columns in order to be analyzed. To do this, use worksheet, option 11, to add another column, then use retrieve, option 4, to move the data within the worksheet.

The Descriptive Statistics Option

- Hard copies should be run whenever you run the Descriptive Statistics option, so that the results can be used to "audit" your worksheet data. The output is dated and gives the number of non-blank entries, max, min, etc.; out-of-range values are easily noted, as are changes in means which indicate changes to data in the analyzed rows or columns.

- You may "get ahead" of the system in selecting row or column numbers in the Descriptive Statistics option; remember to input an extra carriage return (i.e., hit the RETURN key) after the last row (or column) number if you have entered fewer than 14 numbers.

- The row numbers of the worksheet created from the output of Descriptive Statistics will not necessarily correspond to the row (or column) numbers of the input worksheet, but the original numbers are retained as data in column 1 of the output worksheet.

- Create an output worksheet only if you will need it for further computation, and delete it after you have used it successfully. Otherwise, you may accumulate a large number of unneeded worksheets.

- Descriptive Statistics takes a fairly long time for large (i.e., >200) sets of data if the median is to be calculated; don't hit the RETURN or ESC key if there is a delay between lines of output.

The T-Test Option

- Don't depend on the t-test if the associated normality test raises any doubts about its applicability to your data. Consult a statistician, who may be able to transform your data and then appropriately subject it to a t-test.
• When using hypotheses 4 through 7, both groups may be chosen from the same row or column (if there are at least 6 cases). Simply enter the same row or column number in response to the prompts; you will be further prompted for the applicable ranges.

• When using subjects as their own controls, always use the paired t-test options (8 or 9); but note that a case will be excluded if either of the pair of data is missing. Note that the correlation coefficient is calculated when using these hypotheses.

• In performing one-tailed t-tests using hypothesis 5, 7, or 9, the sample whose mean is expected to be larger (if the experiment yields positive results) should be designated as $A$.

• When using the T-Test option, choose the hypothesis that assumes whatever is already known about the samples; in other words, take into account your previous knowledge about equality of variances and the probable sign of the difference between the sample means.

The Chi-Square Option

• The Chi-Square option (analyze, option 3) should be used only when the current worksheet contains frequency data; if the contents are measurements, the results will be meaningless. The Cross-Tab option (analyze, option 6) will produce as output a frequency table that may then be analyzed with the Chi-Square option.

• If your worksheet is larger than $2 \times 2$, you may select a subarray for analysis.

• If the expected frequencies (shown below the actual cell frequencies) are too small, it may be possible to collapse some categories, using the calculate activity. The analysis can then be rerun.

• Results of the chi-square test are reported as the probability of obtaining a chi-square value greater than that computed from the data by chance, assuming that the variables used to categorize the data are independent.

• Two-by-two tables (with $n<75$) are analyzed by means of Fisher's exact test, which is not sensitive to the presence of small cell values.

The Cross Tab Option

• If the data to be analyzed are integers, then entering cut-points such as 1.5, 2.5, etc., will clarify the range of the categories, since no data can take on those values.

• To calculate the mean and standard deviation of one variable (A) stratified by another variable (B), specify A as the x-axis with no cut-points and B as the y-axis with cut-points to stratify the data.
The Scatter Plot Option

- The horizontal resolution of the Scatter Plot option (as displayed on the terminal) is twice the vertical resolution; this should be taken into account when assigning variables to the x- and y-axes.

- If you wish to compare two separately produced plots, you should override the automatic scaling feature and assign your own maxima and minima so that the scales will not vary with the particular data being plotted.

- A scatter plot with automatic scaling provides a good screening procedure, since outliers will usually have a noticeable effect on the plot. They may then be excluded by narrowing the range of the axes.

- When preparing worksheets from which time plots of your data will be produced, use the special time labels rather than including another variable in the worksheet to represent time (see analyze, option 7). The time labels save a row (or column) of the worksheet and are more self-identifying. The system will remind you of their presence when you use the worksheet as input to the Scatter Plot option.

- A “vertical histogram” may be created by first using Frequency Distribution (analyze, option 9) to compute frequencies and to store them in a worksheet. This option can then be used to plot the frequencies in the form of a bar chart.

- A “scatter-plot histogram” may be created by first using ANOVA, (analyze, option 5) to store group numbers and data values in two columns of a worksheet if data values are already stored in separate worksheet columns or rows for each group. Alternatively, group numbers and data values can be directly entered into a worksheet. This option is then used to plot data values versus group number as a scatter plot.

- A cumulative frequency distribution may be created by computing frequencies (with analyze, option 9) and storing the results in a worksheet; computing cumulative sums (with calculate, option 4); then plotting the results with this option.

The Histogram Option

- Vertical and scatter-plot-type histograms may be created as described above, under the Scatter Plot option.

The Frequency Distribution Option

- This option may be used in creating vertical histograms and cumulative frequency distributions as described above, under the Scatter Plot option.

CALCULATE

- Use worksheet, option 11, to enlarge the worksheet prior to using one of the calculate options that stores calculated values in a row or column.
• A convenient way to copy data from a row (column) range in one column (row) to the same range in a second is to use calculate, option 1, to set values in the second column (row) equal to the values in the first while specifying the row (column) range that applies.

• Be careful when using calculate, option 1, to evaluate an expression that contains a reference to an element (indicated by \# r,c) that is in the result row or column. The value of the element gets changed as a result of the calculation, then the new value of the element is used in calculating further results. For example, if the data values in row 7 are

\[
\begin{array}{cccc}
1 & 2 & 3 & 4 \\
5 & & & \\
\end{array}
\]

the evaluation of \#7 = \#7/\#7,3 results in

\[
\begin{array}{cccc}
1/3 & = .333 & 2/3 & = .667 \\
3/3 & = 1 & 4/1 & = 4 \\
5/1 & = 5 & & \\
\end{array}
\]

because the calculation changes the value of element \#7,3 from 3 to 1.

• Enter a date in the form mmd19yy to use calculate, option 4, to replace missing values with the value of a date. That is, enter June 10, 1977 as 6101977 and enter December 5, 1977 as 12051977.

**DESCRIBE**

• When preparing a schema, look at examples prepared by other investigators; get their advice and that of your system manager.

• Prepare drafts of the schema and make corrections on paper before entering it into the system. Once entered, it can be edited, but only a line at a time (that is, a line must be deleted and reentered to make a change).

• Remember that names of panels and items may have no more than 8 characters and that upper and lower case are distinguished.

• Use easy-to-remember item and panel names. You will need to remember them (or look them up) frequently as you enter and retrieve data.

• When using patient abbreviations, use whatever is in common use by those at your institution who will be entering data into the CLINFO system. If you use a hospital I.D. number, you must enter that number each time you enter data for that patient. Although patient names are convenient identifiers, they and Social Security numbers are inappropriate because of privacy considerations (and the potential risk of legal action). Initials may be used, but spelling and capitalization must be consistent; for example, RHJ and rhj are read as two different character strings by the system.

• Include a “method” item in each panel of laboratory results where the analytic method may be modified during the course of your study or where you know that more than one method will be used.
- Keep the character strings for the coded values short and consistent, since they are the form in which data will be entered and displayed. If you use y and n for yes and no in one item, don't switch order or change to YES and NO or Y and N in another item.

- In coded lists, always include at least one "unknown" and/or "other" category, and try to use the same character strings in each item in which they appear. You might want a "not yet available" response as well for some items.

- Don't include in one panel items that are not measured at the same time or whose results are not available at the same time. In the first case, you would not be able to time-stamp the items separately, and in the second, you would need to go back to an existing panel and complete the entry of the previously unavailable items.

- When preparing data collection forms, use the same names and codes that you have used in the schema.

- The formatted schema obtained by using option 2 of describe may be used as a data collection form. Also distribute copies to all those who will be working on your protocol to encourage uniformity in the reporting of item names and codes.

- Use the range-checking feature with all continuous numerical data—and use reasonable ranges. Ranges that are too narrow are annoying to the data enterer; those that are too wide won't catch errors.

- The data type of an item must match the data type of the panel in which it is included; a panel defined as "numeric" can contain items of any data type except free-form text.

- Item names, panel names, and character-string values of coded items should be meaningful to everyone involved with the study described by the schema. Because they will be typed frequently during data entry, they should also be short.

- Because the schema controls data entry, it should be designed to make data entry efficient. This can best be accomplished by describing a schema, practicing data entry with it, and using option 3 to revise the schema until data entry goes smoothly.

**ENTER**

- Data-entry forms should be arranged in the same order as the panels and items in the schema to permit efficient data entry. Response Files should be used to facilitate the entry of data (in any order) and to minimize typing.

- It is most efficient to enter data collected at the same date and time for a series of subjects, since only one context value (i.e., patient abbreviation)
To determine whether data are missing, retrieve data for all the planned collection times. Retrieve by panel or retrieve by item but specify that date and time are to be retrieved along with data values.

To find all the data for a patient regardless of the times of data collection, select the "event begin to event end" time interval and make the selection choice "all."

To retrieve data for only those patients who have particular characteristics, first use the subset activity to create an appropriate patient set.

After sorting a worksheet, use subset, option 7, to create a subset that lists patients in the same order in which they appear in the worksheet. Refer to this subset in retrieve, option 2, to retrieve patient data in the same order as in the sorted worksheet.

To make a worksheet just created by retrieve, option 2, the current worksheet, indicate that you want to display it, then either display it or transfer elsewhere after entering the worksheet display module.

In the Reference Manual text for retrieve, option 2, there is a table called TYPICAL WORKSHEET FORMATS which contains examples of retrieved worksheets and the retrieval specifications used in obtaining them. If one of the illustrative worksheets is similar to one you want to retrieve, the corresponding example retrieval specifications can be a useful guide. The text also includes a table that summarizes specification choices.

### The Retrieve Data from a Worksheet to a Worksheet Option

- Draw a rough "row-column" diagram of the manipulations that you intend to carry out. If they are complex, do them one step at a time and display the destination worksheet between steps to verify the outcome.

- Before attempting complex manipulations of worksheets, it is wise to run hard copies of (portions of) the source and destination worksheets and to indicate on the destination worksheet the source of each row or column that is to be copied into it.

- Use the facility for displaying the current worksheet each time you move data to make sure that what you intended has been carried out.

- Remember that at the end of this option, the source worksheet is still the current worksheet. If you wish to use the destination worksheet in the activities that follow, make it the current worksheet by responding affirmatively to the appropriate prompt or by going to option 0 of retrieve, analyze, calculate, or worksheet.

### The Worksheet to and from Communication File Options

- The Reference Manual text for retrieve, option 5, contains an example of a BASIC program that processes a worksheet-type Communication File.
would then need to be changed. Alternatively, if data are to be entered in
a predictable order, a Response File should be created to minimize typing.

- Remember to add new patients (using enter, option 0) before trying to
  enter data for them.

- Items not entered in a panel will be recorded as “missing”; when re-
  trieved, such values will be displayed as ... (3 periods) in worksheets.

- The date and time of earliest data collection (not data entry) define the
time associated with the event called begin. This event may be referenced
when retrieving data from the SDF.

GOODBYE

- Typing !goodbye directly logs off the user. If you have an Update File you
  may request that it be merged by responding yes to the prompt that
  appears after you type !goodbye. Other messages may be left for the sys-
  tem manager by typing !message prior to logging off.

- Dial-up users must redial after being logged off.

RESPONSE FILES

- Comments may be included in response files to remind you about what
  they do and how. Wherever a question mark appears, it may be followed
  by a space and then a comment. Thus, the following is an example of a
  response file comment:

  ? AT THE PROMPT ENTER THE WORKSHEET NAME

  Comments are displayed only when using response files, option 3.

RETRIEVE

The Event Display Option

- This option always displays the dates and times of first and most recent
data collection for any specified patient. This is useful for reviewing the
data that have been collected and entered and for setting up retrievals.

The Retrieve Data from Study Data File to Worksheet Option

- To ensure that actual data collection times lie within retrieval time units,
  choose time units centered around scheduled data collection times. For
  example, if samples are to be collected on the hour, choose hour-long units
  that start on the half-hour.
To write your own program, start with this example (which shows how to do the file handling) and modify the data processing portion to suit your needs.

**SUBSET**

- To obtain a list of all the patients in the study, use option 8 and enter "all" as the set name.
- If you have a large number of patients or a large amount of data and plan to create several subsets based on values of a few items, it is faster to create the subsets by retrieving relevant data into a worksheet and then using `subset`, option 6 or option 7, than it is to create them based directly on SDF data by using `subset`, option 4 or option 5.

**WORKSHEET**

- Hard copies of blank worksheets with rows and columns labeled make good data collection forms.
- Since 20 rows and only 6 columns may be displayed on the terminal at a time, make rows the larger dimension when creating a worksheet. You can always transpose rows and columns if necessary, using `retrieve`, option 4.
- Entering data into worksheets is much easier if the rows and columns are labeled. The prompts in `worksheet`, options 4, 5, and 6 then clearly identify the required data.
- Entries in a worksheet (and also row or column labels) may be deleted by responding ... (3 periods) to a data-value (or label) prompt.
- Worksheets are modified as data or labels are entered or the worksheet is sorted or edited, and are saved when certain subactivities (options 2 to 6, 8, or 11) are concluded. They need not be explicitly saved.
- Display the worksheet after entering data (using option 1) to make sure the data were properly entered.
- Get hard copies of complex worksheets from time to time, using option 10, as the worksheets are being constructed.
- Upper and lower case are not distinguished in assigning worksheet names.
- Entering `work,1` (with its RETURN) then hitting the RETURN key two more times will display the first 20 rows and 6 columns of the current worksheet; you don’t have to wait for the successive prompts.
- If you have just displayed the current worksheet, the command sequence
starting-row-number, [RETURN] starting-column-number, [RETURN] will display the next segment of the worksheet.

- If you wish to use the decode facility, remember to use labels identical (in spelling and case) to those specified in the schema.

- You cannot accidentally discard the current worksheet by entering only \texttt{throw} (or only \texttt{!7} while in the \texttt{worksheet} activity), since you must verify your intent to discard by typing \texttt{yes}.

- Get hard copies of the list of your worksheets; it’s faster than listing them repeatedly.

- Discard those worksheets that are no longer needed, and get new listings occasionally.

- Note any change in the \texttt{modified} date in your worksheets. It indicates that a change may have been made.

- To sort dates stored in a column or row:
  - Use \texttt{calculate}, option 1, to create a new variable
    \texttt{year=date-100*INT(date/100)},
    where date refers to the column or row of dates.
  - Sort on date.
  - Sort on year.

- A worksheet may be copied by using options 5 and 6 of \texttt{retrieve} in combination. A Response File that can be used for copying a worksheet regardless of its size is

\begin{verbatim}
1. !ret,5
2. ?
3.
4. 0
5. password
6. yes
7. !ret,6
8. 0
9. password
10. ?
11. ?
\end{verbatim}

where \texttt{password} is your log-on password, Communication File 0 already exists and is always used for this purpose, and the first ? (in line 2) corresponds to the name of the worksheet to be copied, the second to the name of the copy, and the third to the title of the copy.
V. REFERENCE MANUAL

ANALYZE .............................................................................. 71
CALCULATE ........................................................................ 142
DESCRIBE ......................................................................... 165
ENTER ............................................................................... 182
EXAMINE .......................................................................... 199
FILES ................................................................................. 203
GOODBYE ......................................................................... 208
MESSAGE .......................................................................... 209
RESPONSE FILES .............................................................. 210
RETRIEVE ......................................................................... 221
SUBSET ............................................................................... 259
WORKSHEET ...................................................................... 273
ANALYZE

Analyze allows the investigator to perform a variety of mathematical and statistical analyses upon data contained in worksheets and to produce other worksheets, reports, graphs, and plots. In general, each major analysis or display function is a separate subactivity which is called up from a menu of functions presented by analyze. Except for option 13 (which requires two worksheets), all of these functions operate on numeric (including encoded) data contained in a single worksheet. The particular information required to perform each analysis is requested interactively by the system after the appropriate analysis is chosen.

The input to an analyze subactivity may comprise the data contained in one or more user-specified columns or rows of the current worksheet. With two exceptions, analyze options permit the selection of the row range if columns are chosen (and vice versa) so that analyses may be limited to a portion of a column or row. Analyses that involve two or more variables, e.g., linear regression and paired t-tests, require the data values for each variable to be stored in a separate column or row. If the data values for two variables are in a single column, worksheet, option 11, can be used to add a new column to the worksheet, and then retrieve, option 4, can be used to copy the values of the second variable from their original column to the new column in the same worksheet.

The output is displayed on the user terminal in a format appropriate to the analysis. Portions of the output of some subactivities may be saved as another worksheet. The format of such an output worksheet is known by the system and is dependent on the type of analysis. The output worksheet may be used in the same manner as any other worksheet, and successive operations may be performed upon it. Further, data contained in an output worksheet may be combined with data from other worksheets, as described under retrieve, option 4. In addition, the data may be made available (via the Communication File) to any BASIC program.

Dialog

Analyze prompts with

```
THE CURRENT WORKSHEET IS TERNUSE
TITLE Machine Utilization - Summary by Ports Consecuted
CREATED 9/16/77    MODIFIED 9/16/77
# OF ROWS 5
# OF COLS  2

ANALYZE OPTIONS

TYPE TO

9..SELECT A WORKSHEET
1..DESCRIPTIVE STATISTICS
2..T-TEST
3..CHI-SQUARE TEST
4..LINEAR REGRESSION
5..ANOVA
6..CROSS TABS
7..SCATTER PLOT
8..HISTOGRAM
9..FREQUENCY DISTRIBUTION
10..NORMALITY TEST
11..NON-PARAMETRIC TESTS
12..LIFE TABLE ANALYSIS
13..RADIOIMMUNOASSAY

ANALYZE OPTION:
```
The user responds with a number from 0 to 13, then hits the RETURN key, and the corresponding procedure is invoked.

OPTION 0: SELECT A WORKSHEET

This subactivity performs the same function as worksheet, option 0 (it creates a new worksheet or changes the current worksheet), and the dialog is identical. It is included here to permit the user to select a worksheet upon entering the analyze activity or to change the current worksheet without leaving analyze. Also, the system takes the user directly to this subactivity if he attempts to invoke some other analyze subactivity (e.g., Descriptive Statistics) when there is no current worksheet. In that case, it automatically takes him to the initially specified subactivity after he selects a worksheet.

Dialog

The dialog proceeds:

WORKSHEET NAME PLEASE: able

If no worksheet named ABLE had been saved previously, the system would respond:

WORKSHEET ABLE DOESN'T EXIST
TYPE yes TO CREATE A WORKSHEET, ELSE [RETURN];

If worksheet ABLE exists, its name and description are written at the top of the screen as soon as it is selected.

If the user had previously attempted to go to another analyze subactivity when there was no current worksheet, he would next be taken there instead of to the list of analyze options.

Note that in the case of worksheet names, unlike item or panel names, the system does not distinguish between upper and lower case, so that ABLE, Able, and able all refer to the same worksheet.

OPTION 1: DESCRIPTIVE STATISTICS

This procedure will calculate 11 basic statistics on the numeric data specified to be in any row(s) or column(s) of a worksheet. These statistics are:

  N       Number of non-missing values in the row/col
  Min     Minimum of the non-missing values
  Max     Maximum of the non-missing values in the row/col
  Range   Max-Min
  Mean    Arithmetic mean of the non-missing values
  Median  The midmost value of the order statistic
  Std Dev Standard deviation: $\sqrt{Var}$
  Sum     Sum of the non-missing values in the row/col
  Sum Squares Sum of squares; $\Sigma x^2$
Var  Variance; \[
\frac{\sum (X_i - \bar{X})^2}{(N - 1)}
\]
Std Error  Standard error of the mean; \(\text{Std Dev} / \sqrt{N}\)

The user may display up to 14 sets of descriptive statistics on the screen at one time and may then list the results and/or save them as a worksheet and proceed with additional row or column statistics.

**Dialog**

If option 1 is selected, the current worksheet is identified, and if it is to be analyzed, the system prompts for row or column:

**TYPE** r **IF DATA ARE IN ROWS**, c **IF IN COLUMNS**: c

The system then offers the option of displaying the column labels, which the user has requested by responding yes:

```
analyze,1 DESCRIPTIVE STATISTICS
WORKSHEET CARDDATA
TITLE Cardiology Study Data
CREATED 1/31/76 MODIFIED 1/31/76
# OF ROWS 28
# OF COLS 20

TYPE yes TO LIST LABELS, ELSE HIT [RETURN] : yes

1 study#
2 entrydat
3 run#
4 age
5 conrttn
6 isoenzym
7 ekgeval
8 hdays
9 v.fib
10 probmi
11 prevmi
12 angina
13 hyperbp
14 cht
15 blank
16 surviv
17 mode
18 sex
19 echpain
20 sob
```

After displaying the labels (if requested), the system prompts:

**TYPE** yes TO EXCLUDE MEDIAN COMPUTATION; ELSE [RETURN] :

The median computation can be time-consuming for large numbers of cases and may not be required.

Next, the user selects the data to be analyzed:

**YOU MAY MAKE UP TO 14 SELECTIONS...**
**FOR EACH COL TO BE ANALYZED, ENTER ITS NUMBER FOLLOWED BY [RETURN]**
STOP BY ENTERING [RETURN] IN RESPONSE TO REQUEST FOR ANOTHER NUMBER...

COL NUMBER:

After choosing up to 14 column (or row) numbers, each followed by hitting the RETURN key, the user hits the RETURN key once more.

The next prompt allows the selection of a range of rows (i.e., cases) to be included in the computation (had rows been selected for analysis, a range of columns could be specified):

THE RANGE OF ROWS FOR THIS COMPUTATION is 1 THROUGH 28;

IF OK, HIT [RETURN]; ELSE ENTER A STARTING ROW NUMBER:

If all of the rows are to be included (as in this example), the user hits the RETURN key and the system presents the statistics in two groups as follows:

```
analyze,1 DESCRIPTIVE STATISTICS 1/31/1976, at 21:29

COL DATA FROM WORKSHEET NAMED: CARDDATA
TITLE Cardiology Study Data
CREATED 1/31/76 MODIFIED 1/31/76

<table>
<thead>
<tr>
<th>COL#</th>
<th>Label</th>
<th>N</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Range</th>
<th>Mean</th>
<th>Median</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>age</td>
<td>28</td>
<td>17</td>
<td>80</td>
<td>63</td>
<td>60.1786</td>
<td>59</td>
</tr>
<tr>
<td>8</td>
<td>hdays</td>
<td>28</td>
<td>8</td>
<td>49</td>
<td>41</td>
<td>22.3571</td>
<td>21.5000</td>
</tr>
<tr>
<td>16</td>
<td>surviv</td>
<td>28</td>
<td>1</td>
<td>9</td>
<td>8</td>
<td>2.03571</td>
<td>1.5000</td>
</tr>
<tr>
<td>18</td>
<td>sex</td>
<td>28</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1.25000</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>COL#</th>
<th>Label</th>
<th>Std Dev</th>
<th>Sum</th>
<th>Sum Squares</th>
<th>Variance</th>
<th>Std Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>age</td>
<td>12.9617</td>
<td>1685</td>
<td>1.05937E+05</td>
<td>168.006</td>
<td>2.44953</td>
</tr>
<tr>
<td>8</td>
<td>hdays</td>
<td>9.7343</td>
<td>626</td>
<td>16554</td>
<td>94.7573</td>
<td>1.33962</td>
</tr>
<tr>
<td>16</td>
<td>surviv</td>
<td>2.04545</td>
<td>57</td>
<td>229</td>
<td>4.18386</td>
<td>.38655</td>
</tr>
<tr>
<td>18</td>
<td>sex</td>
<td>.44096</td>
<td>35</td>
<td>49</td>
<td>.19444</td>
<td>.08333</td>
</tr>
</tbody>
</table>
```

TYPE yes TO SAVE RESULTS IN A WORKSHEET, ELSE HIT [RETURN]:

Missing values are automatically excluded from the computation; the column labeled N shows the number of cases having non-missing values for each specified item. Statistics such as the standard deviation, variance, and standard error are not computed for samples smaller than 3.

At this point the user may hit the RETURN key and proceed to another activity or type yes to save the results in a worksheet. If the results are to be saved, the dialog continues:

TYPE yes TO MAKE THIS NEW WORKSHEET THE CURRENT WORKSHEET: yes

The user would type yes if he wanted to display the new worksheet or perform further analyses upon it as soon as it is created; he would hit the RETURN key to continue with further analysis of CARDDATA (the current worksheet). The dialog would then continue:
WORKSHEET NAME PLEASE: des1

If the named worksheet already exists, the system would prompt:

    des1 ALREADY EXISTS; TO REPLACE IT TYPE yes ELSE [RETURN]

and the user could replace the worksheet or provide another name and continue. If the name were not the name of an existing worksheet, the system would verify that the worksheet had been created and would then continue with the prompts:

    des1 IS THE NAME OF THE WORKSHEET (IT MAY HAVE BEEN TRUNCATED)

HIT [RETURN] FOR ANALYZE OPTIONS:

A worksheet name may consist of up to eight alphanumeric characters; the name will be truncated at the first blank. Thus, the user-entered name, MYDATA 1 (with a blank between A and 1) will be truncated to MYDATA.

A portion of the output worksheet is shown below:

1 / 31 / 76, at 21:38

WORKSHEET DES1
TITLE Descriptive statistics results from COLs of WS named CARDDATA
CREATED 1/31/76 MODIFIED 1/31/76
# OF ROWS 4
# OF COLS 12

<table>
<thead>
<tr>
<th>ROWS/COLS</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>LABELS</td>
<td>COL #</td>
<td>Number</td>
<td>Minimum</td>
<td>Maximum</td>
<td>Range</td>
<td>Mean</td>
</tr>
<tr>
<td>1 age</td>
<td>4</td>
<td>28</td>
<td>17</td>
<td>80</td>
<td>63</td>
<td>60.179</td>
</tr>
<tr>
<td>2 hdays</td>
<td>8</td>
<td>28</td>
<td>8</td>
<td>49</td>
<td>41</td>
<td>22.357</td>
</tr>
<tr>
<td>3 surviv</td>
<td>16</td>
<td>28</td>
<td>1</td>
<td>9</td>
<td>8</td>
<td>2.036</td>
</tr>
<tr>
<td>4 sex</td>
<td>18</td>
<td>28</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1.250</td>
</tr>
</tbody>
</table>

(NOTE: ... indicates missing values)

Note that

- The title is created by the system and is always of the form: “Descriptive statistics results from (rows/cols) of WS named (worksheet name).”
- The 12 column labels (6 of which are shown) are the names of the statistics produced by the Descriptive Statistics subactivity, with the addition of the label for col 1, which is either ROW # or COL # depending on the source of the input data.
- The row labels are copied from the rows or columns of the current worksheet.

If the range of rows to be analyzed was to be restricted (e.g., to fewer than 28 rows, as in this example), the dialog from that point might be:

THE RANGE OF ROWS FOR THIS COMPUTATION IS 1 THROUGH 28; IF OK, HIT [RETURN]; ELSE ENTER A STARTING ROW NUMBER: 3
ENTER AN ENDING ROW NUMBER BETWEEN 3 AND 28: 20
The subsequent dialog would be the same; the results would reflect the smaller n, i.e., 18, and a note would be included in the heading of the results:

<<<ROW RANGE IS 3 THROUGH 20>>>

OPTION 2: T-TEST

This subactivity tests hypotheses concerning the means of one or two populations, by use of the Student t-statistic.

A variety of hypotheses may be tested by this method. The choice depends on prior knowledge concerning the data and the expectations of the investigator. The investigator should select an hypothesis that he expects to disprove; the t-test indicates if the hypothesis can be rejected. Thus, if he expects that the population mean of variable A is greater than K he selects the null hypothesis $\text{Mean } A \leq K$ and the test indicates if this can be rejected in favor of the alternative hypothesis $\text{Mean } A > K$. The test does not prove the alternative hypothesis; it only indicates that the data are consistent with the rejection of the null hypothesis.

This subactivity assumes that the data are in one or two rows or columns of a single worksheet. If data for two variables (A and B) are to be compared by a non-paired test, data for both variables may be within the same column (or row); the user indicates the row (or column) range for each. If two groups of data to be compared by a paired test are in one column, a new column can be added (using worksheet, option 11), the data for the second variable can be moved into it (using retrieve, option 4), and the analysis can be restricted to the appropriate rows.

The information required for carrying out the analysis is elicited by the system. This information includes

- Row(s) or column(s) that contain the data.
- The range of columns or rows to be included in the analysis.
- The number of the null hypothesis to be tested.
- A constant K (for some of the hypotheses).

The proper application of the t-test assumes that the samples were drawn from normally distributed populations. Prior to carrying out the t-test, the samples are automatically tested for normality by means of the Kolmogorov-Smirnov (K-S) test. If that test rejects the hypothesis of normality, it may be necessary to transform the population data (e.g., by log, square root, etc., using calculate) to make them appear normal before performing the t-test. The results of the K-S test are reported, and those variables suspected of being non-normal are noted in the display of the results. If normality cannot be assumed, the non-parametric tests performed by analyze, option 11, can be used.

Null Hypotheses

The choices within the t-test option are stated in the form of null hypotheses; they may be rejected in favor of the appropriate alternative hypotheses.

In all of the null hypotheses noted below, it is assumed that the population variances (Var) are unknown, and that the population data are normally distributed. If the latter is not true, and if N is small, the results may be misleading.
If the hypothesis refers to two groups, the first group chosen becomes group A, and the second, group B.

1. Mean(A) = K

The population (universal) mean is equal to a specified number; the alternatives are that Mean (A) > K or < K.

2. Mean(A) ≤ K (Note: This is displayed as Mean (A) ≤ = K.)

The population mean is not greater than a specified number; the alternative is that Mean (A) > K.

3. Mean(A) ≥ K

The population mean is not less than a specified number; the alternative is Mean (A) < K.

4. Mean(A) = Mean(B), Vars =

Two populations have the same mean, assuming population variances are equal (but unknown); the alternatives are Mean (A) > or < Mean (B).

5. Mean(A) ≤ Mean(B), Vars =

The mean of one population is less than or equal to the mean of a second population, assuming population variances are equal (but unknown); the alternative is Mean (A) > Mean (B).

6. Mean(A) = Mean(B), Vars Neq

Two populations have the same mean, assuming variances are not equal; the alternatives are Mean (A) > or < Mean (B).

7. Mean(A) ≤ Mean(B), Vars Neq

The mean of population A is less than the mean of population B, assuming the population variances are unequal; the alternative is Mean (A) > Mean (B).

8. Paired; Mean(A - B) = 0

The mean of differences in paired values is equal to zero; the alternative is Mean (A - B) > or < 0.

9. Paired; Mean(A - B) ≤ K

The mean of differences in paired values is less than or equal to a specified number; the alternative is Mean (A - B) > K.

For any of the above hypotheses the user should have in mind an alpha-value, or level of significance (e.g., .01, .05, or .10), below which he would reject the null hypothesis. If prior knowledge permits the use of a "one-tailed" test, i.e., option 2, 3, 5, 7, or 9, that is preferable, since it has the effect of increasing the size of the "rejection region" against the null hypothesis.
The procedure reports the exact percentile of the computed t-value in the statement following the t-statistic for p values between .3 and .001; outside of this range, it reports values as "greater than .3" or "less than .001." In addition, for hypotheses 4 through 7 involving two populations, an F-value is also presented so that the user can judge the assumption of equality (or inequality) of the variances.

**Dialog**

```plaintext
analyze, 2 T-TEST

WORKSHEET CARDDATA
TITLE Cardiology Study Data
CREATED 12/9/76 MODIFIED 12/9/76
# OF ROWS 28
# OF COLS 2

HYPOTHESIS OPTIONS

<table>
<thead>
<tr>
<th>NULL HYPOTHESIS</th>
<th>ALTERNATIVE HYPOTHESIS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Mean(A) = K</td>
<td>Mean(A) &gt; or &lt; K</td>
</tr>
<tr>
<td>2. Mean(A) &lt;= K</td>
<td>Mean(A) &gt; K</td>
</tr>
<tr>
<td>3. Mean(A) &gt;= K</td>
<td>Mean(A) &lt; K</td>
</tr>
<tr>
<td>4. Mean(A) = Mean(B), Vars = VarA</td>
<td>Mean(A) &gt; or &lt; Mean(B)</td>
</tr>
<tr>
<td>5. Mean(A) = Mean(B), Vars = VarA</td>
<td>Mean(A) &gt; Mean(B)</td>
</tr>
<tr>
<td>6. Mean(A) = Mean(B), Vars NEQ</td>
<td>Mean(A) &gt; or &lt; Mean(B)</td>
</tr>
<tr>
<td>7. Mean(A) = Mean(B), Vars NEQ</td>
<td>Mean(A) &gt; Mean(B)</td>
</tr>
<tr>
<td>8. PAIRED / Mean(A-B) = 0</td>
<td>Mean(A-B) &gt; or &lt; 0</td>
</tr>
<tr>
<td>9. PAIRED / Mean(A-B) &lt;= K</td>
<td>Mean(A-B) &gt; K</td>
</tr>
</tbody>
</table>

ENTER HYPOTHESIS NUMBER:
```

The screen is blanked, the chosen hypothesis is displayed, and the user is requested to indicate whether the data are in rows or columns of the current worksheet. He then has the option of listing the column (or row) labels:

**TYPE yes TO LIST COL NAMES, ELSE HIT [RETURN];**

Depending on the hypothesis chosen, the system requests the column (or row) numbers of the data to be designated group A and group B and the value of the constant K. The user in this example has decided to test the (two-tailed) hypothesis that the mean age of the population from which this sample was drawn is equal to 55:

```plaintext
===> HYPOTHESIS: 1. Mean(A) = K

SELECT THE COL(S) TO BE TESTED BY ENTERING THE NUMBER OF EACH FOLLOWED BY [RETURN]...

GROUP A- NUMBER: 4

ENTER VALUE OF THE CONSTANT K: 55.
```

The user is then prompted for the row range:

**THE RANGE OF ROWS FOR THIS COMPUTATION IS 1 THROUGH 28; IF OK, HIT [RETURN]; ELSE ENTER A STARTING ROW NUMBER:**
After the user hits the RETURN key, thus including the entire row range, the results are displayed:

```
COL DATA / WS: SCHAPPER
TITLE Schaffer Cardiology Study
CREATED 5/15/75 MODIFIED 10/8/75

NULL HYPOTHESIS: Mean(age) = 55

COL# Label    N    Mean    Std Dev    Variance    Std Error
        =========    ======    ======    =======    =======
        4    age     28  68.1786  12.9617  168.006    2.44953

The t-statistic = 2.114 with 27.0 degrees of freedom.
***** T-Test Interpretation ***************

The (two-tailed) probability of a t-statistic at least as large as that obtained is equal to 0.044, (that is, p=0.044) The null hypothesis (of equality) MAY be rejected.
With one chance in 23 of being wrong, you may assume that the population mean of Group A is LARGER than the constant K.

The null hypothesis is restated using group labels and the value of the constant, and the t-statistic and its associated probability are computed. In the example above, the assumption of age equal to 55 could be rejected at any significance level greater than 0.044, e.g., at p = 0.05.

Had a variable (e.g., sex) been chosen which was not normally distributed, the following displays would have been presented:

NORMALLITY TEST
TESTING COL 18

NORMALLITY TEST RESULTS FOR COL 18, (sex)

Kolmogorov-Smirnov D(n) = .286; The critical value ( alpha=.20 ) is .190

***** CAUTION: THERE IS LESS THAN ONE CHANCE IN FIVE THAT THIS SAMPLE
***** WAS DRAWN FROM A NORMALLY DISTRIBUTED POPULATION.
***** CONSULT A STATISTICIAN BEFORE PERFORMING TESTS WHICH ASSUME NORMALITY.

TYPE yes FOR MORE DETAILS; ELSE HIT (RETURN):

analyze,2 T-TEST 3/15/1976, at 14:00

COL DATA FROM WORKSHEET: SCHAPPER
TITLE Schaffer Cardiology Study
CREATED 5/15/75 MODIFIED 10/8/75

NULL HYPOTHESIS: Mean(sex) = 3

COL# Label    N    Mean    Std Dev    Variance    Std Error
        =========    ======    ======    =======    =======
        18    sex     28  1.25000  .44096    .19444    .08333 ***

*** THERE IS ONLY ONE CHANCE IN FIVE THAT THE POPULATION FROM WHICH
THIS SAMPLE WAS CHOSEN IS NORMALLY DISTRIBUTED; THE T-TEST MAY NOT
BE VALID WITHOUT TRANSFORMING THIS VARIABLE.

The t-statistic = -21.000 with 27.0 degrees of freedom.

A two-tailed test with significance level greater than or equal to 0.000
WOULd reject the null hypothesis in favor of the alternative of a
negative difference.
Tests involving paired data (i.e., hypotheses 8 and 9) assume that the pairs of values for each case are in corresponding locations in the rows or columns selected. If either variable is missing for a case, that case will be excluded from the computation.

In the example below using hypothesis 9, the worksheet is shown, and then the t-test result. In this case the data represent the initial weights of six monkeys and their weights after a 24-hour fast. A one-tailed test is appropriate, since it can be assumed that there was no weight gain; the only issue is whether or not there was a weight loss, and thus the null hypothesis is that of "no weight loss," not "no change."

Although there are six cases, one of them is excluded because of a missing data value.

```
WORKSHEET FEMOBEESE
TITLE Weight comparison for female obese monkeys
CREATED 10/14/77 MODIFIED 10/14/77
# OF ROWS 6  # OF COLS 2

ROWS/COLS 1 2
LABELS wt day1 wt day2
1 57240 15.600 15.200
2 67044 7.400 7.100
3 67278 10.900 10.100
4 78186 9.600 9.100
5 70439 12.300 ...
6 65096 16.100 9.700
```

CUL DATA / WS: FEMOBEESE
TITLE WEIGHT COMPARISON FOR FEMALE OBSESE MONKEYS
CREATED 10/14/77 MODIFIED 10/14/77

NULL HYPOTHESIS: Paired; Mean(wt day1 - wt day2) = 0

<table>
<thead>
<tr>
<th>COL#</th>
<th>Label</th>
<th>N</th>
<th>Mean</th>
<th>Std Dev</th>
<th>Variance</th>
<th>Std Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>wt day1</td>
<td>5</td>
<td>10.720</td>
<td>3.02111</td>
<td>9.1271</td>
<td>1.35108</td>
</tr>
<tr>
<td>2</td>
<td>wt day2</td>
<td>5</td>
<td>10.2500</td>
<td>2.98999</td>
<td>8.94002</td>
<td>1.33716</td>
</tr>
<tr>
<td>0</td>
<td>Diff</td>
<td>5</td>
<td>.47000</td>
<td>.20494</td>
<td>.04200</td>
<td>.09165</td>
</tr>
</tbody>
</table>

The t-statistic = 5.128 with 4.0 degrees of freedom.
The product-moment correlation (r) = 1.00
The 95% confidence interval for rho is between 0.96 and 1.00

***** T-Test Interpretation ***************

The (two-tailed) probability of a t-statistic at least as large as
that obtained is equal to 0.007, (that is, p=0.007)
The null hypothesis (of equality) MAY be rejected.
With one chance in 146 of being wrong, you may assume that
the population mean of the paired differences (A-B) is LARGER than zero.

The results show that the null hypothesis (of no weight loss) could be rejected at the 0.01 level. Note that the correlation coefficient and the 95 percent confidence interval are computed, since the two variables are assumed to have been measured on the same subjects.
If these same data had not been paired but were derived from two different groups (for example, one group fed normally and another on a restricted diet), they would properly be analyzed using hypothesis 5. The results are shown below:

**COL DATA / WS: FEMORESE**
**TITLE WEIGHT COMPARISON FOR FEMALE OBESE MONKEYS**
**CREATED 10/14/77 MODIFIED 10/14/77**

**NULL HYPOTHESIS:** Mean(wt day1) \(\leq\) Mean(wt day2), Variances equal

<table>
<thead>
<tr>
<th>COL#</th>
<th>Label</th>
<th>N</th>
<th>Mean</th>
<th>Std Dev</th>
<th>Variance</th>
<th>Std Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>wt day1</td>
<td>6</td>
<td>10.9833</td>
<td>2.77809</td>
<td>7.71777</td>
<td>1.13415</td>
</tr>
<tr>
<td>2</td>
<td>wt day2</td>
<td>5</td>
<td>10.2500</td>
<td>2.98999</td>
<td>8.94002</td>
<td>1.33716</td>
</tr>
</tbody>
</table>

The t-statistic = 8.421 with 9.0 degrees of freedom.

***** Test of Equality of Variances **************

The F-ratio = 1.15837 with degrees of freedom = 4, 5.

The probability of obtaining an F-value greater than or equal to 1.16, is greater than 0.3. Equal variances MAY be assumed.

***** T-Test Interpretation **************

The (one-tailed) probability of a t-statistic at least as large as that obtained is greater than 0.30.

The null hypothesis MAY NOT be rejected.

Note that there are now six cases shown for the first variable, and that the null hypothesis cannot be rejected even at the .1 level. Moral: Use the correct hypothesis.

**OPTION 3: CHI-SQUARE TEST**

The Chi-Square subactivity tests the hypothesis that two categorizing variables are independent, i.e., that there is no relationship between the distribution of values of one and the values of another. The test is based on data representing frequencies of cases categorized on two variables.

The data are contained in a worksheet specified by the user. A maximum of 11 rows and 11 columns of frequency data may be analyzed. If the worksheet is larger than 11 \(\times\) 11, the user must select a subarray for analysis; if the worksheet dimensions are between 3 \(\times\) 3 and 11 \(\times\) 11, he may select a subarray. Prior to using this option, the user may enter frequency data directly into a worksheet. Or, given a worksheet containing values for at least two variables for a number of cases, he may generate a worksheet containing two-variable frequency data by use of the Cross Tab subactivity, analyze, option 6.

The output for tables larger than 2 \(\times\) 2 includes the chi-square statistic, the degrees of freedom, the probability of obtaining a chi-square greater than or equal to that computed, the row and column distributions and percentages, and the expected frequencies for each cell. If there are fewer than 7 row categories, expected frequencies are displayed automatically; otherwise, they may be displayed optionally.
If the frequency table is $2 \times 2$, it will be analyzed according to the following rules ($N$ = total table frequency):

- $N < 40$: Fisher's exact test only
- $40 < N < 75$: Fisher's and chi-square
- $N > 75$: Chi-square only

For any size table, if more than 20 percent of the expected cell frequencies are less than 5 or if a single expected frequency is less than 1, the chi-square test will not be performed.

The exact test computes the two-tailed probability of obtaining a cell size distribution at least as extreme as that in the present table, given the existing row and column frequencies, and under the assumption of independence of the row and column categorizing variables. For tables with large frequencies, the computation will take a noticeable amount of time.

**Dialog**

When the Chi-Square subactivity is selected, if the current worksheet is $2 \times 2$ (or $4 \times 4$ if produced by the Cross Tab subactivity), it is analyzed under the assumption that it contains frequency data. If it is larger than $2 \times 2$, the user is prompted to select a subarray as follows:

**TYPE yes TO SELECT A SUB-ARRAY FOR ANALYSIS; ELSE [RETURN].**

The current worksheet and the results of performing the chi-square test upon that entire worksheet are shown below. The data represent cases categorized simultaneously by eye color (rows) and by hair color (columns).

```
1 / 31 /76, at 23:29
WORKSHEET EYESHAIR
TITLE Frequency table of eye color vs. hair color
CREATED 1/31/76   MODIFIED 1/31/76
# OF ROWS 3
# OF COLS 3

ROWS/COLS  1  2  3
LABELS Blond Brunette Black
1 Brown    4  7  9
2 Blue     8  3  1
3 Green    4  4  0

(NOTE: ... indicates missing values)
```
The results indicate that the likelihood of obtaining a chi-square value as large as that calculated, under the assumption of the independence of the categorizing variable, is 0.019. That indicates that the hypothesis of independence can be rejected at any confidence level greater than or equal to 0.019.

Had the user chosen to analyze only a portion of the worksheet, the dialog might proceed:
The resulting analysis would be as shown below. Note that since there were fewer than 40 cases and the portion of the worksheet analyzed was a $2 \times 2$ table, only Fisher's exact test was performed.

```
WORKSHEET: EYESHAIR
CREATED 6/7/76 MODIFIED 6/7/76
TITLE: Frequency table of eye color v. hair color
3 ROWS x 3 COLS
For a $2 \times 2$ table with $n < 4^n$, only Fisher's Exact Test will be performed.

2 3 TTL ROWS

1- 7 9 16 8p.0
8.0 8.0 8.0

2- 3 1 4 20.0
2.0 2.0 20.0

TTL 10 10 20
COLS 50.0 50.0 100.0

**** Fisher's Exact Test Results ****
Given the marginal totals, the probability of obtaining the above cell size distribution (or cell size distributions less likely) by chance is 0.582, if the two factors were independent.

Hit [RETURN] for ANALYZE options...
```

In the next example, a $2 \times 2$ table containing 41 cases is analyzed. Since there are more than 40 and fewer than 76 cases, both the chi-square and Fisher's exact test are performed.

```
6 / 7 /76, at 11:27
WORKSHEET HAIR/EYES
TITLE Eye color v. hair color
CREATED 3/11/76 MODIFIED 3/11/76
# OF ROWS 2
# OF COLS 2

ROWS/COLS 1 2
LABELS Blond NotBlond
1 Brown 4 16
2 NotBrown 12 9

(NOTE: ... indicates missing values)
```
**OPTION 4: LINEAR REGRESSION**

The Linear Regression option performs a least-squares linear regression of one row (or column) of a worksheet upon up to three other rows (or columns). It includes several suboptions that approximate curve-fitting by fitting a line to data for transformed independent and dependent variables. The transformations are:

\[
y = a + b \cdot \frac{1}{x}
\]

\[
\ln(y) = a + b \cdot \ln(x)
\]

\[
\sqrt{y} = a + b \cdot x
\]

2nd degree polynomial  
3rd degree polynomial

Using logarithm to the base e  
Using square root

If data values are not already sorted in this format, retrieve, option 4, can be used to copy them from one position in a worksheet to another or from one worksheet to another. It may be necessary to first use worksheet, option 11, to add new rows or columns. The reported results include the regression equation, 95 percent confidence limits for the slope and intercept values, 95 percent confidence limits for the sample r and r-squared, and predicted dependent variable values and 95 percent confidence limits for user-supplied independent variable values. Results may be stored in an output worksheet. Except for the polynomials, the transformed (linearized) original and predicted and residual values, as well as the regression line, may be plotted with the Scatter Plot subactivity, (analyze, option 7). For regressions with a single independent variable, the user may force the regression line through the origin and, if so, he may weight the points as \(1/x^2\).
Dialog

When the Linear Regression option is selected, the system identifies the current worksheet and prompts the user to select rows or columns for analysis. The row or column labels may be displayed. The user then chooses one of the linear regression suboptions; in this example he chooses suboption 1. Next he selects the rows (or columns) to be the dependent and independent variables; the initial dialog is shown below. In this example, only a single independent variable was selected by hitting the RETURN key in response to the prompt for the second independent variable; a later example will deal with multiple linear regression.

```
analyze,4 Linear Regression

WORKSHEET HTWT
TITLE Height v. Weight for patients on experimental diet
CREATED 3/11/76 MODIFIED 5/5/76
100 Rows x 2 Cols

TYPE r for ROW DATA, TYPE c for COLUMN DATA: c

LINEAR REGRESSION OPTIONS

- TYPE FOR
  1...MULTIPLE LINEAR
  2...Y=A+B/X
  3...Ln(Y)=A+B*Ln(X)
  4...Ln(Y)=A+B*X
  5...Sqrt(Y)=A+B*X
  6...2nd DEGREE POLYNOMIAL
  7...3rd DEGREE POLYNOMIAL

OPTION: 1
```

```
analyze,4 Linear Regression

WORKSHEET HTWT
TITLE Height v. Weight for patients on experimental diet
CREATED 3/11/76 MODIFIED 5/5/76
100 Rows x 2 Cols

TYPE yes TO LIST LABELS, ELSE HIT [RETURN]: yes

1 Height
2 Weight

ENTER THE COL NUMBER OF THE DEPENDENT VARIABLE; 2

ENTER THE COL NUMBER OF THE FIRST INDEPENDENT VARIABLE; 1

ENTER THE COL NUMBER OF THE SECOND INDEPENDENT VARIABLE; ELSE [RETURN];
```

The user is then given the opportunity to select a range of rows:

```
THE RANGE OF ROWS FOR THIS COMPUTATION IS 1 THROUGH 100;
IF OK, HIT [RETURN]; ELSE ENTER A STARTING ROW NUMBER:
```

In this case, the user has chosen to analyze the entire worksheet.

The screen is then blanked and a flashing message indicates that the data are being extracted from the worksheet. If there are cases (i.e., rows if columns were chosen) with missing values in either variable, that information is reported on the
screen. Next, the user is asked if a forced fit through the origin is desired (if a single independent variable was chosen); if so, he is then asked if the y-values are to be weighted as $1/x^2$. If neither option is desired, the regression is carried out and the results are displayed as shown below.

```
analyze, 4  LINEAR REGRESSION  12/4/1976, AT 21:17
COL DATA FROM WORKSHEET HTWT
TITLE Height v. Weight for patients on experimental diet
CREATED 3/11/76  MODIFIED 5/5/76

Regression of COL 1 (Height) upon COL 2 (Weight); Sample size = 96
Regression Equation (Y=A+B*X): Height = 52.497 + .07851 * Weight

Standard Error of Estimate (Sy.x): 3.7673 ( = sqrt(1334.07/94) )

<table>
<thead>
<tr>
<th>Std. Error</th>
<th>95% Conf. Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept (A): 38449</td>
<td>51.731</td>
</tr>
<tr>
<td>Slope (B): .91627</td>
<td>.04629</td>
</tr>
</tbody>
</table>

Mean Std. Dev. (Population Est.)
-------------------------------------
Height 66.268 4.1884
Weight 175.31 23.827

Coefficient of Correlation:

<table>
<thead>
<tr>
<th>R</th>
<th>R-SQ</th>
<th>F-RATIO</th>
<th>Sig. Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>.447</td>
<td>.199</td>
<td>23.424</td>
<td>&lt; .001</td>
</tr>
</tbody>
</table>

95% Confidence interval for R:

R-Sq: .271
R-Sq: .073

ENTER A VALUE FOR Weight TO ESTIMATE Height; ELSE HIT [RETURN]:

In the Standard Error of Estimate line of text the sum of squares (1334.07) and the degrees of freedom (94) used to compute the standard error (Sy.x=3.7673) are reported, so that comparisons may be made with forced regressions to test for improved fit.

If the selected rows (or columns) were labeled, these labels are used in the statement of the regression equation. The user may respond to the prompt at the bottom of the display with a value for the x-variable; the system will then compute the estimated and 95 percent confidence limits for the corresponding y-value. Any number of such values may be calculated as illustrated below.

For Height = 62:
Estimated Weight = 155.11; 95% Confidence Interval: 116.66 193.56

For Height = 72:
Estimated Weight = 200.59; 95% Confidence Interval: 161.79 239.40

For Height = 58:
Estimated Weight = 136.91; 95% Confidence Interval: 97.313 176.51

ENTER A VALUE FOR Height TO ESTIMATE Weight; ELSE HIT [RETURN]:

If the user responds to the last prompt with the RETURN key, he is asked if he wishes to construct a worksheet from the results of the linear regression. To do so,
he must supply a name; the title is automatically constructed by the system, using the variable names, if available. A portion of the worksheet that was created in carrying out the example is shown below.

3 / 12 /76, at 11:29

**WORKSHEET REGWT**

**TITLE Regression of COL 2. (Weight 1), upon COL 1. (Height 1), FROM WS HTWT**

**CREATED 3/13/76**

**MODIFIED 3/12/76**

<table>
<thead>
<tr>
<th>ROWS/COLS</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>LABELS</td>
<td>Height</td>
<td>Weight</td>
<td>Y-Pred</td>
<td>Residual</td>
</tr>
<tr>
<td>1</td>
<td>72</td>
<td>175</td>
<td>200.595</td>
<td>-25.595</td>
</tr>
<tr>
<td>2</td>
<td>72</td>
<td>215</td>
<td>200.595</td>
<td>14.405</td>
</tr>
<tr>
<td>3</td>
<td>71</td>
<td>194</td>
<td>196.846</td>
<td>-2.846</td>
</tr>
<tr>
<td>4</td>
<td>71</td>
<td>164</td>
<td>196.846</td>
<td>-12.846</td>
</tr>
<tr>
<td>5</td>
<td>71</td>
<td>172</td>
<td>196.846</td>
<td>-24.846</td>
</tr>
<tr>
<td>6</td>
<td>71</td>
<td>281</td>
<td>196.046</td>
<td>-9.954</td>
</tr>
<tr>
<td>7</td>
<td>71</td>
<td>198</td>
<td>196.046</td>
<td>1.954</td>
</tr>
<tr>
<td>8</td>
<td>78</td>
<td>210</td>
<td>191.497</td>
<td>18.503</td>
</tr>
<tr>
<td>9</td>
<td>78</td>
<td>200</td>
<td>191.497</td>
<td>0.503</td>
</tr>
<tr>
<td>10</td>
<td>78</td>
<td>214</td>
<td>191.497</td>
<td>22.503</td>
</tr>
<tr>
<td>11</td>
<td>76</td>
<td>178</td>
<td>191.497</td>
<td>-13.497</td>
</tr>
<tr>
<td>12</td>
<td>78</td>
<td>165</td>
<td>191.497</td>
<td>-26.497</td>
</tr>
<tr>
<td>13</td>
<td>78</td>
<td>167</td>
<td>191.497</td>
<td>-6.497</td>
</tr>
<tr>
<td>14</td>
<td>69</td>
<td>215</td>
<td>186.948</td>
<td>28.052</td>
</tr>
<tr>
<td>15</td>
<td>69</td>
<td>167</td>
<td>186.948</td>
<td>-24.052</td>
</tr>
<tr>
<td>16</td>
<td>69</td>
<td>221</td>
<td>186.948</td>
<td>34.052</td>
</tr>
<tr>
<td>17</td>
<td>69</td>
<td>162</td>
<td>186.948</td>
<td>-24.948</td>
</tr>
<tr>
<td>18</td>
<td>69</td>
<td>286</td>
<td>186.948</td>
<td>19.052</td>
</tr>
<tr>
<td>19</td>
<td>69</td>
<td>217</td>
<td>186.948</td>
<td>10.052</td>
</tr>
<tr>
<td>20</td>
<td>69</td>
<td>217</td>
<td>186.949</td>
<td>10.052</td>
</tr>
</tbody>
</table>

**NOTE:** ... indicates missing values

**TYPE yes TO CONTINUE WITH worksheet ELSE**

**TYPE THE STARTING ROW 1**

The columns are always in the order shown: x, y, Y-Pred, and Residual. The original variable labels are used if available. If the data are transformed, the transformations are indicated, e.g., as Ln(X) and Ln(Y), in the column labels. If the cases had been labeled, these would appear as row labels. The user may make this worksheet the current worksheet and perform a scatter plot or other analysis on the results. A scatter plot of the residual values versus the predicted weight (Y-Pred) is shown below.

**PLOT FROM WORKSHEET: REGWT ON 3/12/76 AT 11:33**

**TITLE: Plot of residuals (col 4) v. predicted weight (col 3) from regression.**
The following example shows the part of the dialog involved in producing a multiple linear regression from the data in the worksheet below:

<table>
<thead>
<tr>
<th>LABELS</th>
<th>air flow</th>
<th>temp</th>
<th>concentr</th>
<th>stk loss</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>88</td>
<td>27</td>
<td>89</td>
<td>42</td>
</tr>
<tr>
<td>2</td>
<td>88</td>
<td>27</td>
<td>88</td>
<td>37</td>
</tr>
<tr>
<td>3</td>
<td>75</td>
<td>25</td>
<td>90</td>
<td>37</td>
</tr>
<tr>
<td>4</td>
<td>62</td>
<td>24</td>
<td>87</td>
<td>28</td>
</tr>
<tr>
<td>5</td>
<td>62</td>
<td>22</td>
<td>87</td>
<td>18</td>
</tr>
<tr>
<td>6</td>
<td>62</td>
<td>23</td>
<td>87</td>
<td>18</td>
</tr>
<tr>
<td>7</td>
<td>62</td>
<td>24</td>
<td>93</td>
<td>19</td>
</tr>
<tr>
<td>8</td>
<td>62</td>
<td>24</td>
<td>93</td>
<td>26</td>
</tr>
<tr>
<td>9</td>
<td>58</td>
<td>23</td>
<td>87</td>
<td>15</td>
</tr>
<tr>
<td>10</td>
<td>58</td>
<td>18</td>
<td>88</td>
<td>14</td>
</tr>
<tr>
<td>11</td>
<td>58</td>
<td>18</td>
<td>89</td>
<td>14</td>
</tr>
<tr>
<td>12</td>
<td>58</td>
<td>17</td>
<td>88</td>
<td>13</td>
</tr>
<tr>
<td>13</td>
<td>58</td>
<td>18</td>
<td>88</td>
<td>13</td>
</tr>
<tr>
<td>14</td>
<td>58</td>
<td>19</td>
<td>93</td>
<td>12</td>
</tr>
<tr>
<td>15</td>
<td>58</td>
<td>18</td>
<td>89</td>
<td>8</td>
</tr>
<tr>
<td>16</td>
<td>58</td>
<td>18</td>
<td>86</td>
<td>7</td>
</tr>
<tr>
<td>17</td>
<td>58</td>
<td>19</td>
<td>86</td>
<td>8</td>
</tr>
<tr>
<td>18</td>
<td>58</td>
<td>19</td>
<td>72</td>
<td>8</td>
</tr>
<tr>
<td>19</td>
<td>58</td>
<td>19</td>
<td>72</td>
<td>8</td>
</tr>
<tr>
<td>20</td>
<td>58</td>
<td>20</td>
<td>82</td>
<td>15</td>
</tr>
</tbody>
</table>

(Note: ... indicates missing values)

Type yes to continue with worksheet else type the starting row #:

In this example, the user wishes to perform a regression with column 4 as the dependent variable and the first 3 columns as the independent variables.

Enter the col number of the dependent variable; 4
Enter the col number of the first independent variable; 1
Enter the col number of the second independent variable; else [return]; 2
Enter the col number of the third independent variable; else [return]; 3

The results are shown below. The F-ratio tests the hypothesis that $R$ equals zero. This may be rejected at an alpha level less than 0.001. The coefficients and their standard errors are tabulated beneath the restatement of the regression equation. Substituting the values into the equation yields

$$\text{stk loss} = -39.9 + .716 \text{ (air flow)} + 1.3 \text{ (temp)} -.152 \text{ (concentr)}.$$

The standard error of the estimate (2.9) reflects the residual variation in the dependent variable. The means and standard deviations of all of the variables are displayed next. Finally, the partial and simple correlation coefficients are displayed in tabular form, along with F-ratios and significance levels.
ANALYZE, 4

ROW DATA FROM WORKSHEET BROWNLEE
TITLE Example from Brownlee p 454.
CREATED 12/29/76 MODIFIED 12/29/76

N = 21, R = .956, R-SQ = .914, F-RATIO = 59.89, SIG LEVEL = .000

THE EQUATION: stk loss = A + B2(air flow) + B3(temp) + B4(concetr)

COEFFICIENTS OF REGRESSION AND STANDARD ERRORS

A
-39.913

B2 .71567

B3 1.2952

B4 -.1572W

(COEFF)

.7978P

.13487

.36894

.15633

(STD ERR)

STANDARD ERROR OF ESTIMATE FOR SAMPLE = 2.9183

STANDARD DEVIATIONS (POPULATION ESTIMATES) AND MEANS

stk loss air flow temp concentr

17.524 60.429 21.095 86.286 (MEAN)

10.172 9.1684 3.1606 5.3587 (STD DEV)

COEFFICIENTS OF PARTIAL CORRELATION

VARIABLES R F-RATIO SIG LPV

R12.34 .953 28.157 .000

R13.24 .649 12.305 .003

R14.23 .922 51.240 .000

R1.34 .978 30.220 .000

R1.2 .920 164.191 .000

R1.3 .875 62.368 .000

R1.4 .400 3.615 .073

R2.3 .782 29.888 .000

R2.4 .560 6.340 .021

R3.4 .391 3.428 .088

(LOWER-ORDER)

TYPE yes IF YOU WISH TO ESTIMATE stk loss; ELSE [RETURN]:

If estimates of the dependent variable are desired, the user responds yes to the last prompt and enters appropriate values for the three independent variables in response to system prompts.

ENTER VALUE FOR air flow: 58
ENTER VALUE FOR temp: 120
ENTER VALUE FOR concentr: 82
ESTIMATED stk loss IS: 15.428
CONFIDENCE LIMITS FOR ESTIMATED VALUE AT 95% ARE 8.41456 AND 22.9245

TYPE yes IF YOU WISH ANOTHER ESTIMATE OF stk loss; yes
ENTER VALUE FOR air flow: 58
ENTER VALUE FOR temp: 125
ENTER VALUE FOR concentr: 98
ESTIMATED stk loss IS: 14.553
CONFIDENCE LIMITS FOR ESTIMATED VALUE AT 95% ARE 7.54768 AND 21.5576

TYPE yes IF YOU WISH ANOTHER ESTIMATE OF stk loss:

TYPE yes TO SAVE VARS, PREDICTED, RESID. AS A WS; ELSE [RETURN]:
The system responds with an estimated value (15.02) and the 95 percent confidence limits (i.e., the mean plus and minus approximately two standard errors). Additional estimates may be made by responding yes to the last prompt. If the RETURN key is hit, the user is prompted to save the results in a worksheet. Such a worksheet, shown below, may be used to produce residual plots against the independent variables or to locate cases with extreme residuals (e.g., case 4).

**WORKSHEET 80**
**TITLE** REGRESSION OF (stk loss) ON (air flow), (temp), (concentr)/ WS BROWNLEE
**CREATED** 12/29/76 **MODIFIED** 12/29/76
# OF ROWS 21 # OF COLS 6

<table>
<thead>
<tr>
<th>ROWS/COLS</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>LABELS</td>
<td>air flow</td>
<td>temp</td>
<td>concentr</td>
<td>stk loss</td>
<td>y-pred</td>
<td>residual</td>
</tr>
<tr>
<td>1</td>
<td>89</td>
<td>27</td>
<td>99</td>
<td>42</td>
<td>38.765</td>
<td>3.235</td>
</tr>
<tr>
<td>2</td>
<td>80</td>
<td>27</td>
<td>88</td>
<td>37</td>
<td>38.917</td>
<td>-1.917</td>
</tr>
<tr>
<td>3</td>
<td>75</td>
<td>25</td>
<td>90</td>
<td>37</td>
<td>32.444</td>
<td>4.556</td>
</tr>
<tr>
<td>4</td>
<td>62</td>
<td>74</td>
<td>87</td>
<td>28</td>
<td>22.382</td>
<td>5.698</td>
</tr>
<tr>
<td>5</td>
<td>62</td>
<td>22</td>
<td>87</td>
<td>18</td>
<td>19.712</td>
<td>-1.712</td>
</tr>
<tr>
<td>6</td>
<td>62</td>
<td>23</td>
<td>87</td>
<td>18</td>
<td>21.007</td>
<td>-3.007</td>
</tr>
<tr>
<td>7</td>
<td>62</td>
<td>24</td>
<td>93</td>
<td>19</td>
<td>21.389</td>
<td>-2.389</td>
</tr>
<tr>
<td>8</td>
<td>62</td>
<td>24</td>
<td>93</td>
<td>28</td>
<td>21.389</td>
<td>-3.389</td>
</tr>
<tr>
<td>9</td>
<td>58</td>
<td>23</td>
<td>87</td>
<td>15</td>
<td>18.144</td>
<td>-1.144</td>
</tr>
<tr>
<td>10</td>
<td>58</td>
<td>18</td>
<td>88</td>
<td>14</td>
<td>12.734</td>
<td>1.266</td>
</tr>
<tr>
<td>11</td>
<td>58</td>
<td>18</td>
<td>89</td>
<td>14</td>
<td>11.364</td>
<td>2.636</td>
</tr>
<tr>
<td>12</td>
<td>58</td>
<td>17</td>
<td>88</td>
<td>13</td>
<td>10.221</td>
<td>2.779</td>
</tr>
<tr>
<td>13</td>
<td>58</td>
<td>18</td>
<td>82</td>
<td>11</td>
<td>12.429</td>
<td>-1.429</td>
</tr>
<tr>
<td>14</td>
<td>58</td>
<td>19</td>
<td>93</td>
<td>12</td>
<td>12.950</td>
<td>-0.950</td>
</tr>
<tr>
<td>15</td>
<td>58</td>
<td>18</td>
<td>89</td>
<td>8</td>
<td>6.538</td>
<td>2.362</td>
</tr>
<tr>
<td>16</td>
<td>58</td>
<td>18</td>
<td>86</td>
<td>7</td>
<td>6.095</td>
<td>0.905</td>
</tr>
<tr>
<td>17</td>
<td>58</td>
<td>19</td>
<td>77</td>
<td>8</td>
<td>9.521</td>
<td>-1.521</td>
</tr>
<tr>
<td>18</td>
<td>58</td>
<td>19</td>
<td>79</td>
<td>8</td>
<td>8.456</td>
<td>-1.456</td>
</tr>
<tr>
<td>19</td>
<td>56</td>
<td>20</td>
<td>98</td>
<td>9</td>
<td>9.599</td>
<td>-0.599</td>
</tr>
<tr>
<td>20</td>
<td>56</td>
<td>70</td>
<td>82</td>
<td>15</td>
<td>13.588</td>
<td>1.412</td>
</tr>
</tbody>
</table>

**OPTION 5: ANOVA**

The ANOVA (one-way analysis of variance) activity provides for testing the hypothesis that from two to ten groups of samples were derived from the same population, i.e., have the same mean. The number of cases in the groups need not be equal. For hypothesis testing concerning the means of one or two groups, the t-test (analyze, option 2) is appropriate. However, the use of the t-test for multiple comparisons changes the level of significance of the overall test (e.g., to compare 5 means would require 10 t-tests), so that the overall significance level might be much higher than that of a single test. This increases the probability of rejecting the null hypothesis. The analysis of variance estimates the variability of all of the cases combined and compares that with the pooled variabilities derived from each group.

The F-statistic ratio reflects the variability of cases in the combined sample relative to the variability within the groups. Based on the number of cases and the number of groups, the option computes the probability of obtaining an F as large as or larger than that found under the hypothesis of equal means. If that probability is less than the desired level of significance, the hypothesis of equal means may be rejected (with a chance of error equal to the computed probability).
The ANOVA option assumes that the data for the separate groups reside in different rows or columns of a worksheet. If this is not the case, retrieve, option 4, may be used to move data within a worksheet or from one worksheet to another. It may be necessary to first add rows or columns to a worksheet by using worksheet, option 11.

Group numbers and values may be stored in an output worksheet. When the Scatter Plot subactivity, analyze, option 7, is used to plot column 2 of this worksheet against column 1, the result is a scatter-plot type of histogram with one vertical array of plotted points for each group.

**Dialog**

The data to be analyzed are in a worksheet named PRESS, shown below:

```
WORKSHEET PRESS
TITLE Systolic Pressures for Control Group and Two Treatment Groups
CREATED 11/10/77  MODIFIED 11/10/77
# OF ROWS 10
# OF COLS 3

ROWS/COLS

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>L</td>
<td>CON</td>
<td>DRUG A</td>
</tr>
<tr>
<td>1</td>
<td>154</td>
<td>140</td>
</tr>
<tr>
<td>2</td>
<td>181</td>
<td>129</td>
</tr>
<tr>
<td>3</td>
<td>160</td>
<td>138</td>
</tr>
<tr>
<td>4</td>
<td>149</td>
<td>161</td>
</tr>
<tr>
<td>5</td>
<td>137</td>
<td>128</td>
</tr>
<tr>
<td>6</td>
<td>163</td>
<td>143</td>
</tr>
<tr>
<td>7</td>
<td>170</td>
<td>151</td>
</tr>
<tr>
<td>8</td>
<td>152</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>170</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
```

After indicating that the groups are in columns in the current worksheet, the user may list the labels; he then indicates the number of groups and the column number of each. This interaction is shown below.

```
TYPE YES TO LIST LABELS; ELSE HIT RETURN: YES

1 CONTROL
2 DRUG A
3 DRUG B

ENTER THE NUMBER OF GROUPS: (BETWEEN 2 AND 10): 3

ENTER A COL NUMBER FOR EACH GROUP, FOLLOWED BY [RETURN]:

THE COL NUMBER OF GROUP 1 IS: 1
SELECTED COLS: 1,

THE COL NUMBER OF GROUP 2 IS: 2
SELECTED COLS: 1, 2,

THE COL NUMBER OF GROUP 3 IS: 3
SELECTED COLS: 1, 2, 3,
```

The analysis is then performed, and the results are displayed in the standard ANOVA format, as shown below.
ONE WAY ANOVA 11/10/77, at 14:8

Col DATA FROM WS NAMED: PRESS
TITLE Systolic Pressures for Control Group and Two Treatment Groups
CREATED 11/10/77 MODIFIED 11/10/77

<table>
<thead>
<tr>
<th>Group</th>
<th>Col#</th>
<th>Label</th>
<th>N</th>
<th>Mean</th>
<th>Std Dev</th>
<th>Std Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>Control</td>
<td>9</td>
<td>160.444</td>
<td>14.2482</td>
<td>4.74673</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>Drug A</td>
<td>7</td>
<td>141.429</td>
<td>11.7315</td>
<td>4.43407</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>Drug B</td>
<td>8</td>
<td>139.750</td>
<td>14.1598</td>
<td>4.00625</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Total</td>
<td>24</td>
<td>148.000</td>
<td>16.2751</td>
<td>3.32214</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sum of Sq.</th>
<th>Mean Square</th>
<th>Computed F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between</td>
<td>2</td>
<td>220.1856</td>
<td>110.093</td>
<td>6.10835</td>
</tr>
<tr>
<td>Within</td>
<td>21</td>
<td>183.402</td>
<td>8.7382</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>23</td>
<td>6692.58</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The probability of obtaining an F-value greater than or equal to 6.11,
with 2 numerator degrees of freedom and 21 denominator degrees
of freedom, is 0.008
Thus the hypothesis of equal means MAY BE REJECTED at a level
of significance GREATER THAN OR EQUAL TO 0.008

TYPE yes TO PLACE RESULTS IN A WORKSHEET; ELSE (RETURN):

In this case, the “between” mean square (i.e., the sum of the squares of the deviations of the group mean and the category means divided by the number of groups less 1) is 110.093; the “within” mean square (the sum of the squared deviation of each case from its respective category mean divided by the number of cases less the number of groups) is 8.7382. Their ratio is the computed F, approximately 6.1. The probability of obtaining 6.1 or greater, given the number of cases and groups, is 0.008 (under the hypothesis of equal means). Thus the hypothesis may be rejected at a significance level of .05 or .01, but not at .005. This result suggests that the treatments had some effect on the systolic pressure of the experimental groups.

A portion of the output worksheet generated by this option and a scatter-plot type of histogram produced using analyze, option 7, are shown below.
OPTION 6: CROSS TAB

The Cross Tab option provides frequency counts of cases categorized on two variables. It also calculates the total frequency for each category of the two variables, and the means of each variable as categorized by the other variable. The output worksheet may be used as input to the Chi-Square Test, analyze, option 3.

The two variables are termed the x-axis and y-axis variables and are chosen from the rows or columns of a worksheet. As usual, the computation may be restricted to a range of rows (or columns). The user specifies the categories by entering up to 10 cut-points for each variable. The cut-points are treated as upper inclusive bounds for the categories. Thus, specifying the three cut-points 1, 7, and 9.5 establishes four categories:

<table>
<thead>
<tr>
<th>Categories</th>
<th>Bounds</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>&gt;9.5</td>
</tr>
<tr>
<td>3</td>
<td>&gt;7, ≤9.5</td>
</tr>
<tr>
<td>2</td>
<td>&gt;1, ≤7</td>
</tr>
<tr>
<td>1</td>
<td>&lt;1</td>
</tr>
</tbody>
</table>
The x-axis variable cut-points define the columns of the cross tab display, and the y-axis cut-points define the rows. The column means are those of the y-axis variable (for the cases as categorized by the x-axis variable), and the row means are those of the x-axis variable (for the cases as categorized by the y-axis variable). An additional feature calculates and displays the standard deviation (as well as the mean) of the x-axis variable for cases grouped by categories defined for the y-axis variable whenever no cut-points are defined for the x-axis variable.

Each time cut-points are defined, they are saved and may be reused without being reentered if, for example, a series of cross tab tables are to be generated wherein one of the variables remains the same.

Dialog

Upon selecting the Cross Tab option, the user is informed of the current worksheet, asked to select rows or columns for the analysis, and given the opportunity to display the row or column labels, and is then asked to select an x-axis variable and a y-axis variable, as shown below.

analyze, 6 CROSS-TABS

WORKSHEET ERDATA
TITLE Emergency room data
CREATED 1/26/76   MODIFIED 1/26/76
# OF ROWS 250
# OF COLS 3

TYPE yes TO LIST LABELS, ELSE HIT [RETURN]: yes
1 Age
2 Trt time
3

ENTER THE NUMBER OF THE 'X-AXIS' VARIABLE: 1

In this example, the user chooses Age as the x-axis variable and Trt time (time in minutes until discharge in an emergency room situation) as the y-axis variable. He is then prompted to indicate the range of rows to be included.

The screen clears and a message is displayed indicating that data are being extracted from the worksheet; the user is then given the opportunity to enter cut-points or use the previously entered cut-points for the two variables. In the example below, the user has entered new cut-points for Age, the x-axis variable, and has used the stored cut-points for Trt time, the y-axis variable. The cut-points must be entered in ascending order; hitting the RETURN key in response to a prompt for a cut-point terminates the prompting sequence, as does entering the 10th cut-point for either variable.
the 'x-axis' variable is col 1, age

the following 0 cut-points were used previously:

    type yes to use; else hit (return) to enter new cut-points:

enter up to 10 age cut points; hit (return) after each one -
(hit (return) again to terminate prompting)

    age    cut point 1 : 5
    age    cut point 2 : 15
    age    cut point 3 : 25
    age    cut point 4 : 55
    age    cut point 5 :

the 'y-axis' variable is col 2, trt time

the following 4 cut-points were used previously: 30 60 90 120

    type yes to use; else hit (return) to enter new cut-points: yes

the screen clears and a flashing message is displayed indicating that the system
is checking for missing data, then tallying data. if cases (i.e., worksheet rows) are
found that have missing data in one or both of the variables, a message is displayed
similar to that shown below.

tallying data

*******************************************************************************
*** note: there are missing data in the input to this analysis.
*** both the counts and the means may be affected.
***
*** 6 pairs are excluded because both values are undefined;
***
*** 1 pair is excluded because the 'x-axis' value is undefined;
***
*** 3 pairs are excluded because the 'y-axis' value is undefined;
***

the screen again clears, and the cross tab display is presented as shown below.
The circled letters on the example indicate the following:

A = the label (if present) of the worksheet column containing the x-axis variable.
B = the label of the column containing the y-axis variable.
C = the four Age cut-points, which define five Age categories.
D = the four Trt time cut-points, which define five categories.
E = the number of cases having Age greater than 25 and less than or equal to 55 and Trt time greater than 120.
F = the total number of cases with Trt time greater than 120 (and non-missing values for Age).
G = the total number of cases with Age greater than 25 and less than or equal to 55 (and non-missing values for Trt time).
H = the mean age for all cases with Trt time greater than 90 and less than or equal to 120.
I = the mean Trt time for all cases with Age greater than 55.
J = the total number of cases included in the analysis (cases with missing values in either variable are excluded from the analysis).
The user is then given the opportunity to save the results in a worksheet. If he does not wish to save the results, he may hit the RETURN key and, by way of the next prompt, return to the cut-point specification prompting sequence for the same two variables. To save the results, the user must provide a name; the title is supplied by the system. The worksheet created from the example is shown below.

A comparison with the cross tab display will clarify the arrangement of the data. Note that the cut-points are made into labels of the worksheet (along with the variable names) and that the frequency data are contained in the upper-left 6 row \times 6 column subarray within the worksheet. If this worksheet is used as input to the Chi-Square subactivity, that subactivity will recognize the origin of the worksheet and perform a chi-square test on the subarray of frequencies, rather than on the entire worksheet; the two additional rows and columns are disregarded, so it is not necessary to select a subarray explicitly.

If the user wishes to compute the mean and standard deviation of a variable for cases categorized by another variable (e.g., the mean and standard deviation of age for groups having Trt time \( \leq 30 \text{ min} \), \( >30 \text{ min} \), etc.), he may do so by specifying the first variable (in this case, Age) as the x-axis variable, and specifying no cut-points (by responding to the prompt for the first cut-point with the RETURN key). This interaction, using the same worksheet, is shown below.
Note that the previously stored cut-points are not used for Age but are used for Trt time. Following the missing-value display, the results are presented as shown below.

<table>
<thead>
<tr>
<th>Age</th>
<th>Mean</th>
<th>S.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>74</td>
<td>74</td>
<td>39.1</td>
</tr>
<tr>
<td>46</td>
<td>46</td>
<td>40.8</td>
</tr>
<tr>
<td>48</td>
<td>48</td>
<td>48.6</td>
</tr>
<tr>
<td>48</td>
<td>48</td>
<td>32.1</td>
</tr>
<tr>
<td>24</td>
<td>24</td>
<td>45</td>
</tr>
<tr>
<td>240</td>
<td>240</td>
<td></td>
</tr>
</tbody>
</table>

Mean of 185.00

Trt time

The user is next prompted to save the results if he wishes to.

OPTION 7: SCATTER PLOT

The Scatter Plot subactivity permits the user to display on the terminal both scatter plots of up to five variables plotted against any other variable and bar charts defined by two variables. He may copy anything on the screen onto the DTC printer, and he may also have the printer draw line segments between selected sets of displayed data points. Furthermore, if the current worksheet was created by the Linear Regression subactivity, a single response will cause the regression line to be printed together with the original data.

The data values are contained in rows or columns of a worksheet. As in other analyze options, the user specifies which rows or columns correspond to variables of interest, and he may then indicate that only those values within a particular column or row range should be used.

The plot has a resolution of 60 points horizontally and 30 points vertically. Up to five variables may be plotted simultaneously against another variable; the same x- and y-axis scales apply to all the plots. Individual data values are displayed as plotting symbols; the symbols o, x, +, * and # are used to distinguish the plots, which may be overlaid. Multiple values occurring at the same x-y location are indicated by the digits 2 through 9; the symbol @ indicates that more than 9 cases occur at the same location. This subactivity also provides for producing a line drawing on the DTC printer.
The x- and y-axes are automatically scaled to include the minimum and maximum values of the x and y variables while preserving "rational" intervals (e.g., a power of 10 times .1, .125, .5, .75). The user may specify an approximate range for each axis and may exclude points which, as outliers, might cause undesirable scaling of the plot. If points are excluded, that fact is noted on the display. Because the system uses ranges that result in rational intervals and easily interpreted plots, the ranges used may be somewhat wider than those specified by the user.

If row or column labels have been entered in the proper format, they may be used as variables for plotting. The formats are:

SEC snnn
MIN snnn
HR snnn
DAY snnn
WK snnn
MON snnn
YR snnn
# snnn
INT snnn

where snnn stands for a plus or minus sign followed by up to 3 digits or an unsigned number of up to 4 digits, e.g., snnn could be 3, 47, 123, 1084 or −289. The entire label must be 8 characters long, with the time unit abbreviation and blanks in the first 4 positions and snnn in the last 4 positions. The time unit must be the same in all the labels that are interpreted as plottable variables. When the retrieve activity retrieves data from the SDF and builds a worksheet with different rows (or columns) representing different times or instances, it labels the rows (or columns) according to these rules.

Either upper- or lower-case letters may be used arbitrarily in these special labels; MIN, min, Min, etc., are all equivalent.

As everywhere else in the system, the information displayed on the screen may be copied onto the attached DTC printer by pushing the PRINT key. In addition, here the plot can be copied onto the printer by responding to a prompt, and then all the points for some or all of the variables may be connected with lines (in the order in which the data values appear in the worksheet). Furthermore, when data points are printed because of the user's positive response to the printing prompt, they are positioned on the paper with the accuracy of the attached printer whether or not lines are drawn between them. Worksheets generated by analyze, option 4, (Linear Regression) are recognized by this option; the regression line superimposed on the two-variable scatter plot may be obtained on the DTC printer with a single response (see the example below).

The Scatter Plot subactivity is accessed by hitting the 7 key in response to the prompt following the display of the analyze menu or by entering lana/alyze,7 in response to any system prompt. After this, the name and description of the current worksheet are displayed and the dialog proceeds.

The worksheet to be used in the following example is shown below; note that the column labels are in one of the formats that allows the use of the numeric part of the label as a variable.
Dialog

Option 7 first prompts for rows or columns. Assuming that the variables of interest are in rows of the worksheet, it prompts for optional display of the row labels.

The system then prompts for the numbers of the rows to be used as the x- and y-axis variables of the scatter plot. In this example, the user has chosen row 7, alaninet, as the x-axis variable. The system displays the label of the chosen row and then prompts for the y-axis variables as shown below. In this case, up to five y-axis variables may be chosen.
The system then gives the user an opportunity to limit the data values plotted to those within a specified range of columns. In this example, he hits the RETURN key to indicate that he wants to plot data in all columns of the two rows he has already chosen. Next, the system displays the minimum and maximum values found in the worksheet for the two selected variables.

1 OF THE 13 PLOTTING PAIRS CONTAIN AN UNDEFINED VALUE
THEY DO NOT ENTER INTO THE COMPUTATION OF THE MAXIMA AND MINIMA SHOWN BELOW
THEY WILL NOT APPEAR ON THE SCATTER PLOT

<table>
<thead>
<tr>
<th>VARIABLE</th>
<th>MINIMUM</th>
<th>MAXIMUM</th>
</tr>
</thead>
<tbody>
<tr>
<td>x: ROW 7 (alaninet)</td>
<td>.337</td>
<td>.632</td>
</tr>
<tr>
<td>y: ROW 5 (glucoset)</td>
<td>69.400</td>
<td>175.825</td>
</tr>
<tr>
<td>y: ACCUMULATED</td>
<td>69.400</td>
<td>175.825</td>
</tr>
</tbody>
</table>

TYPE yes TO USE OTHER MINIMA AND MAXIMA FOR PLOT SCALING, ELSE [RETURN]:

At this point the user may allow the system to scale the plot so as to include the entire range of both variables or, as is shown in the example, he may assign new maxima and minima. This feature allows the exclusion of outliers which might cause the majority of the data points to be crowded into a small portion of the display. It also permits the generation of several plots that have the same x and y scales, regardless of the range of the data, thus permitting easier comparison of plots made at different times.

In this example, the user has made the x-axis range .1 to .7 instead of 0.337 to .632, and the y-axis range 0 to 300 instead of 69.4 to 175.8. If the user responds with the RETURN key to any of the prompts, the current value would be left unchanged.

The system then displays the rescaled plot. After the user types a plot title, the plot appears as shown below.

The user is then prompted about drawing lines on the printer, and if he does not respond positively, he is taken back to the beginning of Scatter Plot.

PLOT FROM WORKSHEET: MEANS ON 3/2/77 AT 17:46
TITLE: Relation between Mean Plasma Glucose and Alanine Levels—Treatment
The following dialog illustrates (1) the use of the numerical portions of properly formatted labels (e.g., -2 in the label HR -2) as if they were values of a variable with time units (hours in this example), (2) the use of more than one y-axis variable, and (3) the use of the printer to create a line drawing. As before, the row labels may be displayed, followed by dialog, which results in the following display:

YOU MAY SELECT AT MOST 5 Y VARIABLES, HIT [RETURN] TO TERMINATE THE LIST
TO SELECT A ROW, TYPE ITS NUMBER
(0 WILL SELECT TIME IN HRS, FROM COLUMN LABELS, AS A VARIABLE)

Y-AXIS VARIABLE:

X:HR          Y:glucosec glucoset

The user has responded with 0 (zero) to the system request for the x-axis variable, and with 1 and 5 to successive requests for y-axis variables. The interaction is terminated by hitting the RETURN key.

Note that the system recognized the format of the column labels of worksheet MEANS as representing hours. Otherwise the message (0 WILL SELECT: . . .) would not be printed, and the system would not accept 0 (zero) as a response to the prompt for a row number. Here again, the user wants to plot all the data in the selected rows, so he hits the RETURN key (answering affirmatively) when asked if all 13 columns should be used. Otherwise, he would enter a starting column number and an ending column number (of the data to be plotted), possibly after listing the column labels.

When the maxima and minima are displayed, the user leaves the x (HR) minimum and maximum as -2 and 18 (by hitting the RETURN key when prompted for values) but changes the Y ACCUMULATED minimum and maximum (i.e., the overall y-variable range considering both glucosec and glucoset) to 0 and 300, respectively.

After the plot appears on the screen, the user types a plot title.

To begin creating a plot with connected points, the user types yes in response to

TYPE yes TO PRINT THIS WITH LINES, ELSE HIT [RETURN]:

He is then asked to type the plotting symbol for each set of points that he wants lines drawn between or hit the RETURN key to have lines drawn between all sets of points. Data points that do not have lines drawn between them are copied (within the accuracy of the attached printer) as a scatter plot. In this example, hitting the RETURN key or responding o results in drawing lines through both sets of points, as shown below. The response o would draw lines only through the glucosec data points and would copy the unconnected glucoset data points. The response x would draw lines through only the glucoset data points.

The system first rolls the paper to a new page, deletes the data points from the screen, then behaves as if the user had pushed the PRINT key, i.e., it locks the keyboard and copies the screen contents. Next, the system clears the screen and displays Greek and other unusual characters. Finally, it prints line segments between points, as shown below in the plot.
Note in the plot that the system has used the x-axis range —4 to 20 hours, even though the minimum and maximum values of time are —2 and 18 hours, respectively. This is because the system always uses six intervals and —4 to 24 divides into six exact 4-hour intervals, whereas the range —2 to 18 does not.

The following example illustrates the capability of the system to plot a regression line and to produce residual plots when the current worksheet is generated by *analyze*, option 4 (Linear Regression).

Upon entering the Scatter Plot option with such a worksheet, the user is prompted:

**THIS IS A REGRESSION WORKSHEET, TYPE yes TO PLOT THE LINE:**

If the user hits the RETURN key, the dialog proceeds as with any input worksheet; typing *yes* results in the following sequence:

- The dependent variable (vertical axis) is plotted against the independent variable (horizontal axis); the ends of the regression line are indicated by the symbol x.
- The system prompts for TITLE.
- The paper is advanced in the DTC printer, and the scatter plot is copied.
• The screen is blanked, the message

PLEASE WAIT WHILE THE SCREEN IS USED IN PREPARING YOUR CURVE

followed by Greek and other symbols is displayed.

• The regression line is plotted over the scatter plot.

Such a printed regression line is shown below:

The system assumes that the worksheet created by Linear Regression has not been modified. If it has been, the regression line plot may be incorrect. In such a case, the user should answer the regression line prompt by hitting the RETURN key and then create the appropriate plot step by step.

Finally, this subactivity can be used to create a bar chart. If the user selects only one y-variable, he is asked if he wishes a bar chart display. If he responds yes, he is asked to enter the character width of the bars. This width, which can be up to five characters, is generally selected to be the number of character positions between two displayed data points. In the following chart, the bar width is 5 characters:
OPTION 8: HISTOGRAM

The Histogram subactivity creates a horizontal frequency histogram from data contained in a specified row or column of the current worksheet. It also computes counts, frequencies, and cumulative frequencies for both continuous and categorical data.

The category frequencies are represented by up to 40 asterisks, and the number of cases represented by a single asterisk is automatically scaled to conform to that limit; an appropriate frequency scale is displayed at the top of the histogram. Several options provide for different treatment of categorical (i.e., coded) and continuous data. Users who prefer a "vertical histogram" with less auxiliary information displayed may use analyze, option 9, to compute a frequency distribution, save the results in a worksheet, then use Scatter Plot (analyze, option 7) to display a bar chart.

Continuous Data

When treating continuous data, the user has the option of entering up to 15 cut-points or allowing the system to compute cut-points automatically. The cut-points represent upper inclusive bounds for each interval. For example, if the cut-points are chosen to be 10, 20, and 30, there will be four intervals defined as follows:
Category 4 >30
Category 3 >20 and ≤30
Category 2 >10 and ≤20
Category 1 ≤10

The automatic cut-point feature computes the mean, range, and variance of the data and generates a number of categories appropriate to the size of the sample with "nice" cut-points, i.e., powers of 10 times 1, 1.5, 2, 3, etc. The automatic cut-point algorithm assumes that the data are reasonably normally distributed; consequently, highly skewed or other extreme distributions may require manual entry of cut-points for the most informative and useful display of the data. The automatic cut-point option also computes the Kolmogorov-Smirnov (K-S) statistic under the assumption of normality and reports the result.

Categorical Data

If the user indicates that the data are categorical, the system assumes that all legitimate data values will be integers between zero and 29, and constructs a frequency histogram showing the number of cases having each integer value within the range specified. Categories with zero frequencies may be excluded at the user's option. All non-integer data values and all integer values outside the specified range are grouped under the category "other." If the row or column is labeled with the name of a schema item, the appropriate character strings will be substituted for the category codes.

Dialog

When analyze, option 8, is chosen, the system identifies the current worksheet, asks the user to indicate whether the set of data to be used resides in a row or column of the current worksheet, and then may display the row or column labels. In the example below, the user has selected column 6, HT (height), for the creation of the histogram.

```
analyze,8 HISTOGRAM
WORKSHEET BLOOD
TITLE Blood pressure data
CREATED 12/9/75 MODIFIED 12/9/75
100 ROWS X 6 COLS

TYPE yes TO LIST LABELS, ELSE HIT [RETURN]: yes
1 S#
2 DOE
3 AGE
4 R/S
5 EDHT
6 HT

ENTER THE COL NUMBER OF THE VARIABLE: 6
```
The user is given the opportunity to restrict the range of rows (in this case), after which the screen is cleared and the message EXTRACTING DATA is displayed until it is replaced by the histogram suboption menu shown below.

---

PLOT FROM WORKSHEET: PRESSANA

analyze,8 HISTOGRAM 3/14/1976, AT 14:53
COL 6 (HT ) FROM WORKSHEET: BLOOD CREATED 12/9/75 MODIFIED 12/9/75
TITLE Blood pressure data

THE VARIABLE IS COL 6, HT

| TYPE: | 1 to enter CUT-POINTS for CONTINUOUS data; |
|       | 2 for AUTOMATIC CUT-POINTS with NORMALITY TEST; |
|       | 3 to enter CODE RANGE for CATEGORICAL data; |
|       | 4 for ALL CODES (0-29) excluding empty categories; |

ENTER 1, 2, 3, or 4: 1

---

Suboption 1 (enter cut-points for continuous data)

Typing 1 initiates a series of prompts for up to 15 cut-points. If fewer than 15 cut-points are entered, as in the example below, hitting the RETURN key in response to the next cut-point prompt terminates the prompting sequence.

ENTER UP TO 15 CUT-POINTS, EACH FOLLOWED BY [RETURN],
(Hitting [RETURN] only, terminates the prompting sequence...)

HT  CUT POINT 1 : 50
HT  CUT POINT 2 : 55
HT  CUT POINT 3 : 60
HT  CUT POINT 4 : 65
HT  CUT POINT 5 : 70
HT  CUT POINT 6 : 75
HT  CUT POINT 7 :

The screen is cleared and the message TALLYING DATA flashes until it is replaced by the histogram display shown below.
The first line of the display indicates the column (or row) number of the named worksheet from which the data were extracted. If the row or column has been labeled, the label appears in parentheses, e.g., (HT ). In the example above, the highest interval contains two values greater than 75, and the lowest interval would contain values less than or equal to 50. Since the interval (>65, ≤70) contains more than 40 cases, the scaling has been adjusted so that each asterisk represents two cases. The note at the bottom of the histogram reminds the user that the total does not include undefined values. Note that if an asterisk represents 8 (>1) cases, rounding may cause the last asterisk in a row to stand for fewer than 8 cases, and a single case in a category may not generate an asterisk. Typing yes in response to the final prompt causes the system to present the Histogram submenu again while retaining the same variable.

Suboption 2 (automatic cut-points with normality test)

The example below shows the result of entering 2 in response to the Histogram submenu prompt. The same data are used as in the above example.
Although the same data were used, the distribution appears quite different because of
the narrower intervals selected by the automatic cut-point algorithm. In addition to
the information provided by suboption 1, this option lists the maximum, minimum,
mean, and standard deviation of the data and also tests the hypothesis that
this sample of data was drawn from a normal population by use of the K-S test
carried out at an alpha of 0.2. As indicated, the K-S statistic did not exceed the
critical value, and thus the assumption of population normality cannot be rejected.
Note that this is not "proof" that the population is normal, but only lack of evidence
(at the 80 percent confidence level) that it is not normal.

Suboption 3 (enter code range for categorical data)

The variable selected for this example is an encoded item KDHT (known duration of hypertension) having the following values and meanings defined in the schema:

<table>
<thead>
<tr>
<th>Code</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>no history</td>
</tr>
<tr>
<td>1</td>
<td>&lt;6 mos.</td>
</tr>
<tr>
<td>2</td>
<td>6-12 mos.</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>20+ years</td>
</tr>
</tbody>
</table>

Selecting the variable and choosing suboption 3 leads to the following prompting sequence and responses:

ENTER THE RANGE OF CATEGORY CODES (0-29) ...

MINIMUM: 0
MAXIMUM: 8

TYPE yes to EXCLUDE EMPTY CATEGORIES; ELSE HIT [RETURN]:

Hitting the RETURN key (and thus including any empty categories) produces the following histogram:

COL 5 (KDHT ) / WORKSHEET BLOOD 3/14/1976, AT 15:14

<table>
<thead>
<tr>
<th>Category</th>
<th>Freq</th>
<th>%</th>
<th>Cum %</th>
<th>10</th>
<th>20</th>
<th>30</th>
<th>40</th>
</tr>
</thead>
<tbody>
<tr>
<td>E</td>
<td>1</td>
<td>0.0</td>
<td>100.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20+ yrs</td>
<td>2</td>
<td>2.0</td>
<td>100.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>K 15-19 yr</td>
<td></td>
<td>5.0</td>
<td>98.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D 10-14 yr</td>
<td></td>
<td>9.0</td>
<td>93.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H 5-9 yrs</td>
<td>11</td>
<td>11.0</td>
<td>64.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T 1-4 yrs</td>
<td>19</td>
<td>19.0</td>
<td>73.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6-12 mos</td>
<td>8</td>
<td>8.0</td>
<td>54.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;6 mos</td>
<td>8</td>
<td>8.0</td>
<td>46.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>no hist</td>
<td>38</td>
<td>38.0</td>
<td>100.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Total: 100 100.0

TYPE yes TO ENTER NEW CUT-POINTS; ELSE HIT [RETURN]:
Note that the left-hand column is now labeled Category rather than Cut-pt (cut-point), category 8 is included even though empty, and codes for which character strings are in the schema have been replaced by those strings.

Another example using the same data illustrates the effect of choosing different limits:

**ENTER THE RANGE OF CATEGORY CODES ( 0-29 ) . . .**

**MINIMUM: 2**

**MAXIMUM: 5**

**TYPE yes TO EXCLUDE EMPTY CATEGORIES; ELSE HIT [RETURN]**

<table>
<thead>
<tr>
<th>Category</th>
<th>Freq</th>
<th>%</th>
<th>Cum %</th>
<th>20</th>
<th>40</th>
<th>60</th>
<th>80</th>
</tr>
</thead>
<tbody>
<tr>
<td>K (Other)-</td>
<td>53</td>
<td>53.0</td>
<td>100.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D 10-14 yr-</td>
<td>9</td>
<td>9.0</td>
<td>47.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H 5-9 yrs -</td>
<td>11</td>
<td>11.0</td>
<td>38.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T 1-4 yrs -</td>
<td>19</td>
<td>19.0</td>
<td>27.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6-12 mos-</td>
<td>8</td>
<td>8.0</td>
<td>8.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Total:</strong></td>
<td>100</td>
<td>100.0</td>
<td>100.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**TYPE yes TO ENTER NEW CUT-POINTS; ELSE HIT [RETURN]:**

Note that the cases previously included in categories corresponding to codes 0, 1, 6, and 7 are now included in the Other category.

**Suboption 4 (all codes 0-29), excluding empty categories**

This choice involves no further prompting and has the same effect as specifying a range of 0 to 29 and excluding empty categories in suboption 3.

**OPTION 9: FREQUENCY DISTRIBUTION**

This option counts the number of occurrences of each distinct value in a specified row or column of the current worksheet and displays the values, sorted in ascending order, and their associated frequencies on the screen, 150 pairs at a time. It is useful for data editing and screening purposes. If the variable has been labeled in the worksheet, that label will be displayed over the "value" columns. The user may save the results as a 2 column × n row worksheet for further processing and display (e.g., vertical histograms and cumulative frequency distributions may be generated by use of the Scatter Plot subactivity and the calculate activity). The output worksheet is limited to 599 rows, i.e., 599 distinct values.

**Dialog**

After selecting rows or columns for analysis and optionally viewing the labels, the user selects a variable (Age) and hits the RETURN key, as shown below.
He is then given the option of restricting the range of rows to be analyzed:

**THE RANGE OF ROWS FOR THIS COMPUTATION IS 1 THROUGH 100; IF OK, HIT [RETURN]; ELSE ENTER A STARTING ROW NUMBER.**

Hitting the RETURN key selects the entire range.

Each value of Age and its associated frequency is then displayed, and the number of missing values is reported.

<table>
<thead>
<tr>
<th>Age</th>
<th>FREQ</th>
<th>Age</th>
<th>FREQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>1</td>
<td>46</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>48</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>51</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>52</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>55</td>
<td>2</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>56</td>
<td>2</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>60</td>
<td>1</td>
</tr>
<tr>
<td>12</td>
<td>2</td>
<td>62</td>
<td>1</td>
</tr>
<tr>
<td>14</td>
<td>1</td>
<td>65</td>
<td>1</td>
</tr>
<tr>
<td>18</td>
<td>1</td>
<td>66</td>
<td>1</td>
</tr>
<tr>
<td>19</td>
<td>1</td>
<td>68</td>
<td>1</td>
</tr>
<tr>
<td>20</td>
<td>2</td>
<td>69</td>
<td>2</td>
</tr>
<tr>
<td>21</td>
<td>1</td>
<td>70</td>
<td>2</td>
</tr>
<tr>
<td>22</td>
<td>2</td>
<td>71</td>
<td>1</td>
</tr>
<tr>
<td>24</td>
<td>1</td>
<td>74</td>
<td>3</td>
</tr>
<tr>
<td>25</td>
<td>1</td>
<td>75</td>
<td>1</td>
</tr>
<tr>
<td>26</td>
<td>3</td>
<td>78</td>
<td>2</td>
</tr>
<tr>
<td>27</td>
<td>2</td>
<td>79</td>
<td>1</td>
</tr>
<tr>
<td>28</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>36</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>37</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>41</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>43</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>45</td>
<td>3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**TYPE yes TO PLACE RESULTS IN A WORKSHEET; ELSE [RETURN]:**
Had there been more than 150 value-frequency pairs, the following message would be displayed:

THERE ARE MORE FREQUENCIES TO BE DISPLAYED;
HIT [RETURN] TO CONTINUE:

Hitting the RETURN key would then display up to 150 more value-frequency pairs. When all have been displayed, the user may save the results in a worksheet by responding yes to the worksheet prompt and naming the new worksheet. Otherwise, he may continue to select other variables to be analyzed.

The first 20 rows of the resultant worksheet are shown below. The title is automatically generated and the row labels are of the form "# nn" which permits them to be used as a variable in Scatter Plot.

```
6 / 17 / 76, at 11:59

WORKSHEET AGEDIST
TITLE Frequency Distribution of Col # 1 (Age), from WS AGETIME
CREATED 6/17/76 MODIFIED 6/17/76
# OF ROWS 48
# OF COLS 2

ROWS/COLS 1 2
LABELS Age FREQUENCY
1 # 1 2 1
2 # 2 3 1
3 # 3 4 2
4 # 4 5 1
5 # 5 7 1
6 # 6 8 1
7 # 7 10 1
8 # 8 12 2
9 # 9 14 1
10 # 10 16 1
11 # 11 19 1
12 # 12 20 2
13 # 13 21 1
14 # 14 22 2
15 # 15 24 1
16 # 16 25 1
17 # 17 26 3
18 # 18 27 2
19 # 19 28 2
20 # 20 29 3
```

(NOTE: ... indicates missing values)

TYPE yes TO CONTINUE WITH worksheet ELSE
TYPE THE STARTING ROW #:

Applying calculate, option 4 (Within-Column Cumulative Sum), to column 2 of the resultant worksheet, followed by Scatter Plot (analyze, option 7), produces the following cumulative distribution plot:
The following example illustrates the construction of a vertical bar chart from the Frequency Distribution output worksheet. The original worksheet contained diagnostic codes for 289 cases. The first 20 rows of the output worksheet are shown below.

WORKSHEET CODED
TITLE: Frequency Distribution of Co1 # 6{ DiagCode }, from WS ERDATA
CREATED 6/1/76 MODIFIED 6/1/76
# OF ROWS 23 OF COLS 2

ROWS/COLS 1 2
LABELS  DiagCode  FREQUENCY
1 # 1  1  95
2 # 2  4  3
3 # 3  5  1
4 # 4  6  2
5 # 5  7  6
6 # 6  9  4
7 # 7 14  2
8 # 8 15  1
9 # 9 25  1
10 # 10 26 10
11 # 11 27  1
12 # 12 2240 1
13 # 13 2570 22
14 # 14 2573 10
15 # 15 3569 13
16 # 16 4120 1
17 # 17 4150 28
18 # 18 5164 1
19 # 19 6390 6
20 # 20 7240 25
Note that the range of the codes is such that a plot using DiagCode as the x-axis variable would cause some of the low-valued codes to appear as points in the same column (because of the 60-point limit to the horizontal resolution on the screen). If the row labels are used as the x-axis variable, however, the screen resolution is sufficient to resolve every point (from 1 to 23). The resulting bar chart is shown below.

**OPTION 10: NORMALITY TEST**

The Normality Test option tests the hypothesis that a specified sample was drawn from a normally distributed population. It also produces an automatically scaled horizontal frequency histogram of data in a single user-selected worksheet row or column. The intervals are computed on the basis of the number of cases in the sample and the standard deviation. The cut-points (which are upper inclusive bounds) are then adjusted to be “nice” values (i.e., 1, 1.25, 1.50, 2, 3, etc., times a power of 10). The category frequencies are represented by up to 40 asterisks. The number of cases represented by a single asterisk is scaled to conform to that limit.

The option then computes the Kolmogorov-Smirnov (K-S) statistic, which is the absolute value of the difference in proportions of the cumulative distribution in the sample categories and in corresponding categories of a normal distribution with mean and standard deviation equal to that of the sample. This statistic is then compared to critical K-S values. If the statistic exceeds a critical value, the hypothesis of normality of the population from which the sample was drawn may be
rejected at the reported probability; the system reports the largest critical value (corresponding to probabilities of .2, .1, .05, and .01) exceeded by the statistic. If no critical value is exceeded by the statistic, that indicates only that there is not sufficient evidence to reject the hypothesis of population normality at the 80 percent confidence level; it is not proof that the population is normally distributed.

The choice of an alpha of .2 provides a large (conservative) rejection region, and thus will occasionally reject the hypothesis of normality even when the population has a normal distribution (namely, one time in five, in the long run). If a critical value is exceeded, care should be taken in employing any statistical procedure (such as the t-test, linear regression, or analysis of variance) which assumes reasonable normality. Often a transformation may be made on the sample data (using the calculate activity) which will result in a more nearly normal distribution. The transformed data may then be analyzed using parametric tests. Alternatively, the tests performed by analyze, option 11, can be used because they do not require the assumption that the population is normally distributed. 

When in doubt, see a statistician.

Dialog

After the user selects this option, the current worksheet is displayed and the user is prompted to select rows or columns of the worksheet for analysis. He may optionally display the row (or column) labels and then select a single row or column for analysis, as shown below.

```
analyze,10 NORMALITY TEST

WORKSHEET HTVWT
TITLE Height v. Weight for patients on experimental diet
CREATED 3/11/76 MODIFIED 5/5/76
100 ROWS x 2 COLS

TYPE yes TO LIST LABELS, ELSE HIT [RETURN]: yes
1 Height
2 Weight

ENTER THE COLUMN NUMBER OF THE VARIABLE: 2
```

The screen is cleared and a blinking message indicates that the analysis is proceeding. A histogram is then presented, as shown below, beneath which is a statement of the results of the K.S test.
analyze.10  NORMALITY TEST  
COL 2 (Weight) /WORKSHEET HTVWT  12/16/1976, AT 17:01

<table>
<thead>
<tr>
<th>Cut Pt.</th>
<th>Freq</th>
<th>%</th>
<th>Cum %</th>
<th>10</th>
<th>20</th>
<th>30</th>
<th>40</th>
</tr>
</thead>
<tbody>
<tr>
<td>220</td>
<td>6</td>
<td>6.3</td>
<td>97.9</td>
<td>100.0 **</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>230</td>
<td>11</td>
<td>11.5</td>
<td>91.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>250</td>
<td>10</td>
<td>18.4</td>
<td>80.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>170</td>
<td>8</td>
<td>8.3</td>
<td>69.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>160</td>
<td>10</td>
<td>10.4</td>
<td>61.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>150</td>
<td>23</td>
<td>24.0</td>
<td>51.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>140</td>
<td>13</td>
<td>13.5</td>
<td>27.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>130</td>
<td>8</td>
<td>8.3</td>
<td>13.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>120</td>
<td>5</td>
<td>5.2</td>
<td>5.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Total: 96 100.0  NOTE: 4 undefined values were excluded.

Maximum = 232  Minimum = 124  Mean = 175.31  S. D. = 23.827

Kolmogorov-Smirnov D(96) = .094; The critical value (alpha=.20, N=96) = .109

**** This test, done at an 80% confidence level, cannot reject the
**** assumption that the sample was drawn from a normally distributed
**** population.

TYPE yes FOR MORE DETAILS; ELSE HIT [RETURN]: yes

In this example, the hypothesis of normality cannot be rejected. If the user wishes to see more details of the test, he may respond yes to the last prompt, and the following display will be presented:

analyze.10  NORMALITY TEST  12/16/1976, AT 17:03

COLUMN A = ACTUAL CUMULATIVE DISTRIBUTION
COLUMN B = CUMULATIVE DISTRIBUTION OF N( 175.31 , 23.827 )
COLUMN C = ABS(A-B)/96

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>7</td>
<td>.0268</td>
</tr>
<tr>
<td>13</td>
<td>14</td>
<td>.0164</td>
</tr>
<tr>
<td>26</td>
<td>25</td>
<td>.0164</td>
</tr>
<tr>
<td>49</td>
<td>40</td>
<td>.0938 &lt;=</td>
</tr>
<tr>
<td>59</td>
<td>55</td>
<td>.0417</td>
</tr>
<tr>
<td>67</td>
<td>70</td>
<td>.0313</td>
</tr>
<tr>
<td>77</td>
<td>82</td>
<td>.0521</td>
</tr>
<tr>
<td>88</td>
<td>99</td>
<td>.0184</td>
</tr>
<tr>
<td>94</td>
<td>93</td>
<td>.0184</td>
</tr>
<tr>
<td>96</td>
<td>96</td>
<td>.0000</td>
</tr>
</tbody>
</table>

The K-S D(n) = .094, the largest absolute difference in proportions.
The critical value of the K-S statistic, for alpha=.20 and n= 96 is .109.

HIT [RETURN] TO CONTINUE WITH NORMALITY TEST:
The three columns represent the cumulative number of cases in each interval in
the sample, the number in an equivalent normal distribution (i.e., one with the
same n, mean, and standard deviation), and the (absolute) difference in proportions
in each interval. The maximum difference (indicated by the arrow) is the statistic
that is compared to the K-S critical value.

The following histogram illustrates a sample which can be assumed to be non-
normal (at the 80 percent confidence level).

Had the statistic been large enough to exceed the critical value corresponding to
an alpha of .1, the message would report:

... LESS THAN ONE CHANCE IN 10 ...

Again, it is important to test samples for normality before applying parametric
statistical tests; consult a statistician before proceeding if the sample does not
appear to be normal.

OPTION II: NON-PARAMETRIC TESTS

This subactivity performs four non-parametric tests (i.e., tests that do not as-
sume normal distributions) upon the data in any two rows or columns of a work-
sheet. Three of these tests assume that the variables represent pairs of measure-
ments made on the same cases; the fourth does not make such an assumption. If
data are not already in the appropriate format, they may be moved (from the same
or another worksheet) by using *retrieve*, option 4, possibly after adding rows or
columns with *worksheet*, option 11.

The *paired-value* tests include the following:

- The Spearman Rank Correlation Test

  This tests the hypothesis that two variables are independent. The results
  include the *r*-statistic, as well as the probability of attaining an *r* as large
  as or larger than that computed, under the assumption that the variables
  are unrelated in the population. Appropriate adjustment is made for ties
  in ranks. This test does not assume normality or equal interval data, only
  that the values represent ordered rankings of some pair of population
  attributes.

- Sign Test

  The sign test tests the hypothesis that the *median* of the difference be-
  tween pairs of measurements is zero. It reports the number of non-zero
  differences, the smaller number of like-signed differences (the *r*-statistic),
  and the two-tailed probability of obtaining an *r* at least as small. This test
  is most useful when each of the two observations of a given pair was made
  under similar conditions but the different pairs were observed under differ-
  ent conditions. The *t*-test is preferable if all of the distributional assump-
  tions can be met.

- Wilcoxon Signed Ranks Test

  This extension of the sign test may be used to test the hypothesis that the
  *median* of a group of observations is equal to some specified value or that
  the median of the differences between two groups is equal to zero. The
  results include the number of non-zero differences between pairs, the
  algebraic sum of the signed ranks, the smaller sum of like signed ranks
  (the *r*-statistic), and the probability of obtaining a *t* at least that small.

The two-group test test is

- Wilcoxon Two-Sample Rank Sum Test

  This tests the hypothesis that two samples are drawn from the same
  population.

Dialog

The data to be analyzed are shown in the worksheet below. They represent
pairs of blood chemistry measurements made on a series of 15 cases after the
administration of two different drugs.
This option prompts with a menu and the user selects 1, paired-value tests. These tests are appropriate when pairs of measurements are made under the same conditions.

```
<table>
<thead>
<tr>
<th>TYPE TO SELECT:</th>
</tr>
</thead>
<tbody>
<tr>
<td>1......PAIRED-VALUE TESTS</td>
</tr>
<tr>
<td>Sign Test</td>
</tr>
<tr>
<td>Spearman Rank Correlation</td>
</tr>
<tr>
<td>Wilcoxon Signed Rank Test</td>
</tr>
<tr>
<td>2......TWO-GROUP TEST</td>
</tr>
<tr>
<td>Wilcoxon 2-group Rank Sum</td>
</tr>
</tbody>
</table>
```

After inquiring whether rows or columns are to be analyzed and displaying the labels, if the user requests them, the system prompts for the number of the first row or column to be tested. In this example, column 1 is chosen; the system then prompts:

TO COMPARISON THE MEDIAN OF COLUMN 1 WITH A CONSTANT, ENTER A NUMBER: 
ELSE [RETURN]

If the median of a single group of observations is to be compared with a constant, that constant is entered here (see the example below). Otherwise, the user hits the RETURN key and the system prompts:
ENTER THE COLUMN NUMBER OF THE SECOND VARIABLE:

Following the entry of a legitimate column number (i.e., 2), the user is given the opportunity of restricting the range of the computation. The system then indicates that the analysis is proceeding by a series of flashing messages—"SORTING COL1", "RANKING", etc. The results are then displayed as shown in the example below.

THE VARIABLES ARE COLUMN 1, Drug A AND COLUMN 2, Drug E

** SIGN TEST RESULTS **

Number of non-zero differences = 15
Smaller number of like-signed differences = 5 (the \( t \)-statistic) 

The (TWO-TAILED) probability that \( t \) is less than or equal to 5, under the assumption that the median of the differences is zero, is 0.216.

** SPEARMAN RANK CORRELATION **

SPEARMAN \( R = 0.961; N = 15 \)
(Medians) COl 1: 29, COL 2: 32.4, Difference: -3.39999

The probability of obtaining an \( R \) greater than or equal to 0.961, under the hypothesis of independence, is less than 0.01.

** WILCOXON SIGNED RANKS TEST **

Number of non-zero differences between pairs = 15
Algebraic sum of signed ranks = -61
Smaller sum of like-signed ranks = 29.5 (the \( T \)-statistic)

The (TWO-TAILED) probability that \( T \) is less than or equal to 29.5, assuming that the median difference between the two groups is zero, and that rank-sums are normally distributed, is 0.085.

A message is displayed if any of the rows (or columns) have been excluded because one or both values were missing for the variables. The sign test results do not give sufficient evidence that the medians are different in the two groups (\( p = 0.216 \)). The Spearman Rank Correlation results include the medians of the two groups and the difference between the medians, as well as the computed \( R \) (0.961). The \( p \) value is less than 0.001, indicating a large positive correlation between the groups. The Wilcoxon test, also of zero difference between medians, yields a \( p \) value of 0.085. Thus, the null hypothesis could be rejected at the 0.1 level.

If the hypothesis to be tested is that the median value of one of the groups is equal to a specific value, that value (27.5 in the example below) is entered in response to the prompt:

TO COMPARE THE MEDIAN OF COLUMN 1 WITH A CONSTANT,
ENTER A NUMBER;
ELSE [RETURN]: 27.5

In this case, the Spearman correlation will not be performed, since only one group is involved (even though this is identified as a "paired-value" test). The results of the other two tests, shown below, both indicate that the actual median of column
1, (29), is different from 2.5. Again, the Wilcoxon test yields a lower p value. Note that the tests are two-tailed; the one-tailed p value would be half of that reported and would be appropriate in this case.

THE VARIABLES ARE COLUMN 1, Drug A AND THE HYPOTHETICAL MEDIAN, 27.5.

** SIGN TEST RESULTS **

Number of non-zero differences = 15
Smaller number of like-signed differences = 3 (the r-statistic)

The (TWO-TAILED) probability that r is less than or equal to 3, under the assumption that the median of the differences is zero, is 0.035.

MEDIAN OF COL 1: 29
HYPOTHETICAL MEDIAN: 27.5

** WILCOXON SIGNED RANKS TEST **

Number of non-zero differences between pairs = 15
Algebraic sum of signed ranks = 98
Smaller sum of like-signed ranks = 11 (the T-statistic)

The (TWO-TAILED) probability that T is less than or equal to 11, assuming that the median difference between the two groups is zero, and that rank-sums are normally distributed, is 0.006.

If measurements were not made under the same conditions, a two-group, non-paired test could be applied. In this case, only one test, the Wilcoxon two-sample rank sum test is available. To use it the user selects suboption 2 when presented with the menu, then proceeds as before. The results for the DRUGTEST data are shown below. They are consistent with the previous results (computed under the paired assumption) in that they indicate that the assumption that both samples were drawn from the same population cannot be rejected at the .05 level of significance.

analyze.11 WILCOXON NON-PAIRED RANK SUM TEST

WORKSHEET DRUGTEST
TITLE Comparison of response to two experimental drugs
CREATED 12/38/76 MODIFIED 12/38/76
15 ROWS BY 2 COLUMNS

GROUP 1 (COL 1), Drug A N= 15
GROUP 2 (COL 2), N= 15

The sum of the ranks of Group 1 is equal to 223.0;

>>> There were 2 sets of ties <<<
Normal approximation used; z = 0.39

If you assume that both samples were drawn from the same population, then

* The (two-tailed) probability of achieving a sum of ranks
  less than or equal to 223.0, OR
  greater than or equal to 242.0 is 0.693;

>>> NOTE: To reject the assumption at the .05 level (two-tailed)
  the sum of ranks for Group 1 must be;
  less than or equal to 185, OR
  greater than or equal to 280;

HIT [RETURN] TO CONTINUE WITH ANALYZE 11....
OPTION 12: LIFE TABLE ANALYSIS

The Life Table Analysis option estimates the proportion of patients from a population whose "life expectancy" exceeds some known time. The Kaplan-Meier product-limit technique is used to make the estimate. This is a maximum likelihood non-parametric estimation procedure which utilizes all the available information in a sample containing right-censored data. By life expectancy is meant the time to the occurrence of some event of interest (e.g., death, onset of symptoms, or recovery from disease). A right-censored observation is one for which the time of occurrence of the event of interest is unknown but for which we have the information that the patient "survived" at least up to some known time, i.e., it is not known when or if the event occurred, but it is known that it did not occur before a time which can be specified. Whereas this discussion and the system refer to "death," this actually means death or any other end event that is being investigated.

The system computes the cumulative probability of surviving for a single life table or a pair of them. If a pair of tables is used, the system computes Gehan's Wilcoxon test for equality of the tables and reports whether or not the data indicate that they differ.

Because of the non-parametric nature of these estimates, the cumulative probability of survival may be estimated only at points for which there is an observation of a "death." No appropriate method of interpolation between points exists.

The input for this option may be in one of two general forms: (1) grouped data, and (2) individual data. For grouped data, each entry contains the ending time of an interval, the number of "deaths," and the number of right-censored observations during that interval. For example, row 3 of worksheet GEHAN2 is for the interval 9 to 14 weeks, during which there were 2 deaths and 2 patients lost to the study.

---


For individual data, observations for each individual patient must include either on-study time and off-study time (i.e., the time the patient goes on study and the time of the event of interest or loss from study) or elapsed time. They must also include an off-study indicator, i.e., an indication of the meaning of the observations (e.g., died from the disease under study, died accidentally, lost to follow-up, or last visit). Worksheet GEHAN contains the individual data from which worksheet GEHAN2 was constructed. Column 1 contains the number of weeks that the patient was on the study, and column 2 contains the indication of the type of observation (1 = death, 0 = death did not occur while on study).

<table>
<thead>
<tr>
<th>ROWS/COLS</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>LABELS</td>
<td>time wks</td>
<td>remission</td>
</tr>
<tr>
<td>1</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>7</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>10</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>13</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>14</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>22</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>23</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>11</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>12</td>
<td>10</td>
<td>1</td>
</tr>
<tr>
<td>13</td>
<td>11</td>
<td>1</td>
</tr>
<tr>
<td>14</td>
<td>17</td>
<td>1</td>
</tr>
<tr>
<td>15</td>
<td>19</td>
<td>1</td>
</tr>
<tr>
<td>16</td>
<td>20</td>
<td>1</td>
</tr>
<tr>
<td>17</td>
<td>25</td>
<td>1</td>
</tr>
<tr>
<td>18</td>
<td>32</td>
<td>1</td>
</tr>
<tr>
<td>19</td>
<td>34</td>
<td>1</td>
</tr>
<tr>
<td>20</td>
<td>35</td>
<td>1</td>
</tr>
<tr>
<td>21</td>
<td>32</td>
<td>1</td>
</tr>
<tr>
<td>22</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>23</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>24</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>25</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>26</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>27</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>28</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>29</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>30</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>31</td>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td>32</td>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td>33</td>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td>34</td>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td>35</td>
<td>11</td>
<td>0</td>
</tr>
<tr>
<td>36</td>
<td>11</td>
<td>0</td>
</tr>
<tr>
<td>37</td>
<td>12</td>
<td>0</td>
</tr>
<tr>
<td>38</td>
<td>15</td>
<td>0</td>
</tr>
<tr>
<td>39</td>
<td>17</td>
<td>0</td>
</tr>
<tr>
<td>40</td>
<td>22</td>
<td>0</td>
</tr>
<tr>
<td>41</td>
<td>23</td>
<td>0</td>
</tr>
<tr>
<td>42</td>
<td>12</td>
<td>0</td>
</tr>
<tr>
<td>43</td>
<td></td>
<td></td>
</tr>
<tr>
<td>44</td>
<td></td>
<td></td>
</tr>
<tr>
<td>45</td>
<td></td>
<td></td>
</tr>
<tr>
<td>46</td>
<td></td>
<td></td>
</tr>
<tr>
<td>47</td>
<td></td>
<td></td>
</tr>
<tr>
<td>48</td>
<td></td>
<td></td>
</tr>
<tr>
<td>49</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
When two life tables are to be computed from one worksheet, their data must be in consecutive rows (columns) and their data elements must lie in the same columns (rows).

Dialog

After the user selects the Life Table Analysis option, the current worksheet is identified and the user is prompted to select rows or columns of the worksheet for analysis. In this example, the user types c to indicate that different data values for an individual or group are in different worksheet columns, and that different rows represent different individuals or groups. If he had responded r, then column and row would be interchanged in the following discussion.

The system inquires if the data are in the form of individual observations or are grouped. In this example the user hits the RETURN key, indicating the use of individual data. The system next prompts for the starting row to be used in the computation of the life table. Since two separate life tables are to be computed (the first in rows 1 through 21, and the second in 22 through 50), the user responds by entering 1. Next the user is prompted for the ending row number. After he responds 21, he is asked if a second life table is to be computed. In this example, the user responds yes. The system then requests the starting row number for the second life table. Since the default range is acceptable, the user responds by hitting the RETURN key.

```
analyze,12 PRODUCT-LIMIT LIFE TABLE ON 8/23/1977 AT 11:12
WORKSHEET GEHAN 50 ROW X 2 COLS
TITLE data from Gehan's paper
CREATED 3/4/77 MODIFIED 3/4/77

<< INDIVIDUAL DATA >>

THE RANGE OF ROWS FOR THIS COMPUTATION IS 1 THROUGH 50;
IF OK, HIT [RETURN]; ELSE ENTER A STARTING ROW NUMBER: 1

ENTER AN ENDING ROW NUMBER BETWEEN 2 AND 50: 21

TYPE yes TO SELECT RANGE FOR A SECOND LIFE TABLE, ELSE [RETURN]: yes
THE RANGE OF ROWS FOR THIS COMPUTATION IS 22 THROUGH 50;
IF OK, HIT [RETURN]; ELSE ENTER A STARTING ROW NUMBER:
```

The system next allows the user to have the column labels displayed, and then prompts for the format of the time-on-study data.
In this example, the user specifies time option 1, since the worksheet contains the number of weeks on study for each observation. Option 2 is used when each observation contains the dates (and/or times of day) that the patient entered and left the study.

The system next prompts for the column containing time on study (i.e., the elapsed time between on and off study) and the column containing the off-study indicator.

This is followed by a request for the first code indicating a loss due to "death" (i.e., the end event of interest). If the column label for the off-study indicator corresponds to an item in the schema and if that column is said to contain encoded data, the system will expect the response to be a character string; if not, the response must be an integer between 0 and 9. The user may specify up to five different indicator codes for a loss due to "death" and up to five indicators for a censored observation. He responds to the prompt for the next code with the RETURN key to indicate that he has specified all the codes.

In response to a prompt, the user enters the units in which time on study is measured.

The system next proceeds with the calculation and displays the results. After an initial pass through the data, the system displays a summary describing the observations in each life table.
The system then pauses while the user decides if these numbers are sensible (i.e., if the off-study indicator codes were properly specified). If everything looks all right, the user hits the RETURN key and the system proceeds with the computation of the generalized Wilcoxon test for equality of the life tables.

Under the null hypothesis that the two life tables are drawn from the same populations, the Wilcoxon test statistic (W) is approximately normally distributed with mean zero. The null hypothesis may be rejected when the z value \(W/\text{STD}(W)\) is large in absolute value. The appropriate critical value of z depends on whether a one- or a two-tailed test is to be used and on the level of significance that the investigator deems reasonable. If the alternative hypothesis is that the second sample is drawn from a population with longer (shorter) expected life than the first, the appropriate test to use is a one-tailed test, rejecting the null hypothesis for large negative (positive) values of z. If the alternative hypothesis is simply that the two populations are different, the appropriate test is two-tailed. In this example, the test statistic W has a value of 271 and an approximate standard deviation of 73.77, giving a z of 3.67. This leads to the rejection of the null hypothesis that the two samples have the same time to "failure" (i.e., to the end event of interest) in favor of the alternative hypothesis that the first sample comes from a population with a greater expected time to failure. The significance level is .001.

The system next displays the maximum time on study for the first sample and prompts the user to enter a new, alternative value or to hit the RETURN key. This value is used as the maximum obtainable life expectancy (with probability zero of living longer) in computing the conditional life expectancy. The effect of this is illustrated below. In this example, the user chooses to use 35 weeks, the time on study for the longest censored observation.

\[ W = 271 \quad \text{VAR}(W) = 5441.56 \quad \text{STD}(W) = 73.766 \quad z = 3.674 \]

A one-tailed test with a significance level greater than or equal to .001 would reject the hypothesis that time to failure is greater in the 1st sample.

The maximum time in the first sample is 35 wks.

To reset, enter new value else [return]:

```
analyze,12 PRODUCT-LIMIT LIFE TABLE ON 9/6/1977 AT 14:46
WORKSHEET GEHAN 50 ROW X 2 COLS << ROWS 1 to 21 and 22 to 50 >>
TITLE data from Gehan's paper CREATED 3/4/77 MODIFIED 3/4/77
<< INDIVIDUAL DATA >>

GENERALIZED WILCOXON TEST FOR COMPARING TWO LIFE TABLES

W=271 \quad \text{VAR}(W)=5441.56 \quad \text{STD}(W)=73.766 \quad z=3.674

A one-tailed test with a significance level greater than or equal to .001 would reject the hypothesis that time to failure is greater in the 1st sample.

The maximum time in the first sample is 35 wks.

To reset, enter new value else [return]:
```
The results for the first sample are then displayed on the screen. There is one row for each different time at which there is a "death." Column 2 (Elp time units) contains the elapsed times on study; column 3 (Psurviv), the cumulative probabilities of surviving this long; and column 5 (RISK), the risk, i.e., the probability of "dying" before this time given that the patient was "alive" at the end of the previous time period. Column 4 (E(t)) gives the life expectancy conditional on an individual having survived up to this point. The computation of life expectancy is somewhat arbitrary if the last (i.e., the longest lived) observation in the sample is not a death, since the maximum life expectancy is not known and must be arbitrarily set.

<table>
<thead>
<tr>
<th>Elp WKS</th>
<th>Psurviv</th>
<th>E(t)</th>
<th>RISK</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.857</td>
<td>23.287</td>
<td>.143</td>
</tr>
<tr>
<td>2</td>
<td>.807</td>
<td>26.169</td>
<td>.059</td>
</tr>
<tr>
<td>3</td>
<td>.753</td>
<td>27.367</td>
<td>.067</td>
</tr>
<tr>
<td>4</td>
<td>.698</td>
<td>28.687</td>
<td>.083</td>
</tr>
<tr>
<td>5</td>
<td>.627</td>
<td>30.026</td>
<td>.091</td>
</tr>
<tr>
<td>6</td>
<td>.538</td>
<td>31.429</td>
<td>.143</td>
</tr>
<tr>
<td>7</td>
<td>.440</td>
<td>33.000</td>
<td>.167</td>
</tr>
</tbody>
</table>

The maximum time in the second sample is 23 WKS

To reset, enter new value else [return]:

The effect of maximum life expectancy is illustrated in the following table, which is computed from the same data as the preceding one but which uses a maximum life expectancy of 50 rather than 35 weeks.

<table>
<thead>
<tr>
<th>Elp WKS</th>
<th>Psurviv</th>
<th>E(t)</th>
<th>RISK</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.857</td>
<td>30.010</td>
<td>.143</td>
</tr>
<tr>
<td>2</td>
<td>.807</td>
<td>34.012</td>
<td>.059</td>
</tr>
<tr>
<td>3</td>
<td>.753</td>
<td>35.700</td>
<td>.067</td>
</tr>
<tr>
<td>4</td>
<td>.698</td>
<td>37.536</td>
<td>.083</td>
</tr>
<tr>
<td>5</td>
<td>.627</td>
<td>39.766</td>
<td>.091</td>
</tr>
<tr>
<td>6</td>
<td>.538</td>
<td>42.143</td>
<td>.143</td>
</tr>
<tr>
<td>7</td>
<td>.440</td>
<td>45.300</td>
<td>.167</td>
</tr>
</tbody>
</table>

The maximum time in the second sample is 23 WKS

To reset, enter new value else [return]:

The maximum time in the first sample is 35 WKS

To reset, enter new value else [return]: 50
The system repeats the last two steps for the second sample and then allows the user to save the results in a worksheet. The resulting worksheet is in a form that allows for the convenient plotting (with `analyze`, option 7) of two survival (or life expectancy or risk) curves.

**WORKSHEET SURVIVAL**

<table>
<thead>
<tr>
<th>ROWS/COLS</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>LABELS</td>
<td>1p WKS</td>
<td>Psurvival</td>
<td>E(t)1</td>
<td>RISK1</td>
<td>Psurvival2</td>
<td>E(t)2</td>
</tr>
<tr>
<td>1</td>
<td>6</td>
<td>.657</td>
<td>23.287</td>
<td>.143</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>2</td>
<td>7</td>
<td>.807</td>
<td>26.169</td>
<td>.059</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>3</td>
<td>10</td>
<td>.753</td>
<td>27.367</td>
<td>.067</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>4</td>
<td>13</td>
<td>.690</td>
<td>28.007</td>
<td>.083</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>5</td>
<td>16</td>
<td>.627</td>
<td>30.026</td>
<td>.091</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>6</td>
<td>22</td>
<td>.538</td>
<td>31.429</td>
<td>.143</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>7</td>
<td>23</td>
<td>.448</td>
<td>33</td>
<td>.167</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

| ROWS/COLS | 7 |
|LABELS    | RISK2 |
| 1         | ... |
| 2         | ... |
| 3         | ... |
| 4         | ... |
| 5         | ... |
| 6         | ... |
| 7         | ... |
| 8         | .895  |
| 9         | .805  |
| 10        | .859  |
| 11        | .125  |
| 12        | .143  |
| 13        | .333  |
| 14        | .250  |
| 15        | .333  |
| 16        | .250  |
| 17        | .333  |
| 18        | .500  |
| 19        | 1     |
OPTION 13: RADIOIMMUNOASSAY

The RIA option provides a means of estimating the concentration of ligand (i.e., the substance whose concentration is being determined by a laboratory radioimmunoassay) from count data for a number of unknowns. In order to accomplish this, several operations are necessary. First, a tube layout protocol must be prepared and stored as a worksheet. Next, a standard curve is calibrated to the data for this assay. Finally, by means of interpolation, the standard curve is used to estimate the concentration in each unknown.

In order to minimize the labor in running the analysis, this option is divided into four suboptions. Three of these (suboptions 1, 2, and 4) are related to establishing, modifying, and examining tube layouts. Suboption 3 utilizes an existing tube layout to calibrate a standard curve and, if this is found to be satisfactory, to estimate the concentration for the unknown sample.

In performing the analysis, the RIA option follows the general approach advocated by David Rodbard. This procedure is based on estimating a linear transformation of the standard curve which relates a response variable to the concentration of unlabeled ligand in the standard tubes. The values of the response variable are first estimated from the raw count data and are then transformed so that an approximately linear response function can be estimated.

The most widely used response variable is the ratio of counts for ligand that is bound (B) to the counts for ligand that is bound in the absence of unlabeled antigen (B0). Rodbard and others have shown on theoretical and empirical grounds that

$$\logit(B/B_0) = b_0 + b_1 \cdot \ln(\text{concentration})$$

is approximately linear and that under ideal conditions $b_1$ is exactly 1. The logit transformation is defined by

$$\logit(x) = \ln(x/(1-x)) \quad x < 1$$

undefined \quad x \geq 1

Though there are some undesirable statistical properties associated with using this relationship to estimate the concentration in the unknowns, computational tractability dictates against the use of more elaborate statistical techniques, and experience indicates that this procedure is generally quite satisfactory.

The procedure used in CLINFO differs slightly from that used by Rodbard. In estimating $B_i/B_0$ from the data, Rodbard uses separate estimates for each data point and then uses all points in fitting the standard curve. In CLINFO, we compute a single estimate of $B_i/B_0$ for each concentration level, using the information that replicates are multiple measurements of the same quantity, and then use the single estimates for each level in fitting the standard curve. We employ this procedure because it reduces the bias introduced by a non-linear transformation. To see this, consider the situation pictured below:

For the three points at the left, the logit of the mean is less than the mean of the logit, while for the points at the right, the logit of the mean is greater than the mean of the logit. The nature of the logit transformation is such that, in effect, more weight is given to the highest-count points for low-dose replicates and to the lowest-count points for high-dose replicates. This tends to bias downward (make more negative) the estimated slope of the linearized standard curve. From this, it would appear that the slope produced by Rodbard's procedure would always be more negative than CLINFO's, but this is not necessarily so, because under some circumstances his procedure requires that some points be excluded from the analysis.
Points for which the value of B/B0 is greater than 1 cannot be included in the regression because the logit of x is undefined for x≥1. It is not uncommon, for some observations at very low concentrations of ligand, to exceed the mean number of counts for the zero-concentration standard, but it is rare for the mean number of counts in any of the standard tubes to exceed this. Thus it is possible, using Rodbard’s procedure, to systematically exclude the high-count tubes from the analysis, thus biasing the estimate of the slope of the standard curve toward less negative values.

There are four alternative response-variable/standard-curve combinations to choose from as suboptions:

1. F/B, i.e., free/bound.
2. B/T, i.e., bound/total counts.
3. (T–C)/T–C0), where T is total counts, C is counts in the supernatant (i.e., in the free portion), and C0 is C at zero concentration.
4. B/B0, i.e., bound/bound at zero concentration.

Response variables 3 and 4, which utilize the logit transformation of B/B0 or equivalent, are the most commonly used. Response variable 3 is used when the free portion (the supernatant) is counted, and response variable 4 is used when the bound portion (the precipitate) is counted.

Response variable 1 is the form recommended by R. P. Ekins.* He has shown that the bound-to-free ratio (B/F) is a linear function of the concentration if logit(B/B0) is linear in ln(concentration) (i.e., if logit(B/B0) = b0 + b1+ln(concentration)) and the slope coefficient (b1) is equal to −1. This response variable involves no non-linear transformation. Thus, it should be considered for making the concentration estimates if response variable 3 or 4 yields a slope estimate that is very near −1. On theoretical grounds, we expect the slope to be near −1 in most cases, so response variable 1 should often be a good choice. When response variable 1 is used with data for which the supernatant is counted, the resulting standard curve will look somewhat peculiar, but the estimates produced will be satisfactory.

Dialog

When the RIA option is selected, CLINFO identifies the current worksheet and prompts the user to select the suboption to be performed.

Suboption 1 (build tube layout)

In this example, suboption 1 is selected first in order to set up the tube layout protocol. The system prompts for the maximum number of tubes which will be employed and the units for the concentration of ligand. The responses for this example are 90 and nanogram, respectively.

A table containing a running tally of tube assignments is displayed and the user is prompted for the location (tube number) of the beginning of the first set of tubes for non-specific bonding (NSB) and the number of replicates in this set. In this example, tube 45 and 4 replicates are indicated.

<table>
<thead>
<tr>
<th>TUBE TYPE</th>
<th>TUBES USED</th>
<th># OF SETS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Background</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Standard</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Unknown</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Blank</td>
<td>90</td>
<td>0</td>
</tr>
</tbody>
</table>

-----BACKGROUND (NSB)-----

<table>
<thead>
<tr>
<th>Group</th>
<th>Start # of tubes</th>
<th>Group</th>
<th>Start # of tubes</th>
<th>Group</th>
<th>Start # of tubes</th>
<th>Group</th>
<th>Start # of tubes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

ENTER STARTING TUBE NUMBER FOR 1ST BACKGROUND (NSB) GROUP: 45

ENTER THE NUMBER OF REPLICATES IN THIS GROUP, IF 3 [RETURN]: 4

CLINFO then prompts for the location of the next set of NSB tubes. The RETURN key is hit, indicating that only a single set is being used. The system next prompts for the location and number of replicates for the total count tubes. Had there been more than one set of NSB tubes, it would also have prompted for the number of NSB groups associated with the total count.

ENTER STARTING TUBE NUMBER FOR 'TOTAL COUNT': 1

ENTER THE NUMBER OF REPLICATES IN THIS GROUP, IF 4 [RETURN]:

NEXT TUBE 449

This is followed by a request for the starting location and number of replicates for the standard tubes. Here, the responses are 5 and RETURN (indicating 4 replicates). CLINFO then requests the concentration for each standard (it is assumed that the standards occupy successive locations and that all standards have the same number of replicates). To indicate that the concentrations for all the standards have been entered, the user hits the RETURN key in response to the request for the concentration of the next standard. The system requires that a zero concentration standard be specified and will not build a tube layout without one.
ANALYZE, 13

DOSE CONC. UNITS: nanogram

=== TOTAL COUNT TUBE ===
Start # of Bkgrnd
tube tubes group
1 4 1

=== BACKGROUND (NSB) ===
| Group Start # of | Group Start # of | Group Start # of | Group Start # of |
| # tube tubes | # tube tubes | # tube tubes | # tube tubes |
| 1 45 4 |

=== STANDARD TUBES ===
| # Start # of Conc. Bkgrnd | # Start # of Conc. Bkgrnd |
| tube tubes group | tube tubes group |

ENTER STARTING TUBE NUMBER FOR STANDARDS: 5
ENTER THE NUMBER OF REPLICATES PER STANDARD, IF 4 (RETURN):

ENTER CONCENTRATION FOR STANDARD # 1: 0
NEXT TUBE #5

ENTER CONCENTRATION FOR STANDARD # 2: 500
NEXT TUBE #9

ENTER CONCENTRATION FOR STANDARD # 10, ELSE (RETURN): 1
NEXT TUBE #45

After the last standard is entered, the starting location of the first set of unknowns and the number of unknowns in the set are requested. The user responds 49 and 6, respectively. The system next requests the number of dilutions per patient for these unknowns. In the example, the user responds with the RETURN key, since he is not using multiple dilutions for each patient. (CLINFO allows up to 5 dilutions per unknown; a dilution of 2:1 would be entered as 2.) All unknowns in a set are assumed to have the same number of replicates and the same dilutions. The system next prompts for the number of replicates per unknown and the dilution factor to be used in adjusting the estimated concentration. It then prompts for the starting location of the next set of unknowns. This process is repeated until the user responds with the RETURN key to the prompt for the starting location of the next set of unknowns. In this example, two sets of unknowns are established, one starting at tube 49 with 6 unknowns consisting of 4 replicates each and the other starting at tube 73 with 6 unknowns of 2 replicates.
DOSE CONC. UNITS: nonogram

== TOTAL COUNT TUBE ==
Start # of Bkgrnd tube tubes group
1 4 1

== BACKGROUND (NSB) ==
<table>
<thead>
<tr>
<th>Group Start # of tubes</th>
<th>group</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>45</td>
</tr>
</tbody>
</table>

== STANDARD TUBES ==
<table>
<thead>
<tr>
<th># Start # of Conc. Bkgrnd tube tubes group</th>
<th># Start # of Conc. Bkgrnd tube tubes group</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 5 4 8.0 1</td>
<td>2 9 4 500.0 1</td>
</tr>
<tr>
<td>3 13 4 250.0 1</td>
<td>4 17 4 1000.0 1</td>
</tr>
<tr>
<td>5 21 4 50.0 1</td>
<td>6 25 4 25.0 1</td>
</tr>
<tr>
<td>7 29 4 10.0 1</td>
<td>8 33 4 5.0 1</td>
</tr>
<tr>
<td>9 37 4 2.5 1</td>
<td>10 41 4 1.0 1</td>
</tr>
</tbody>
</table>

== UNKNOWNS ==
| # thru # Start Tubes Dil. Bkgrnd tube /set group |
|-----------------------------------------------|-----------------------------------------------|

ENTER STARTING TUBE NUMBER FOR 1st SET OF UNKNOWNS: 49
ENTER THE NUMBER OF UNKNOWNS IN THIS SET: 6

ENTER THE NUMBER OF DILUTIONS PER PATIENT; IF 1 [RETURN]:
ENTER NUMBER OF REPLICATES PER UNKNOWN: 4
ENTER THE BACKGROUND (NSB) GROUP #: IF OTHER THEN 1, ELSE [RETURN]:

ENTER DILUTION FACTOR FOR THESE UNKNOWNS; IF 1 [RETURN]:
ENTER STARTING TUBE NUMBER FOR 2nd SET OF UNKNOWNS ELSE [RETURN]: 73
ENTER THE NUMBER OF UNKNOWNS IN THIS SET: 6
ENTER THE NUMBER OF DILUTIONS PER PATIENT; IF 1 [RETURN]:
ENTER NUMBER OF REPLICATES PER UNKNOWN: 2
**ANALYZE, 13**

**RADIOIMMUNOASSAY**

**DOSE CONC. UNITS: nanogram**

---**TOTAL COUNT' TUBE ===**

<table>
<thead>
<tr>
<th>Start # of Bkgnd</th>
<th>tube tubes group</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4</td>
</tr>
</tbody>
</table>

---**BACKGROUND (NSB)**---

<table>
<thead>
<tr>
<th>Group</th>
<th>Start # of</th>
<th>Group Start # of</th>
<th>Group Start # of</th>
<th>Group Start # of</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>tube tubes</td>
<td>tube tubes group</td>
<td>tube tubes group</td>
<td>tube tubes group</td>
</tr>
<tr>
<td>1</td>
<td>45</td>
<td>4</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---**STANDARD TUBES ===**

<table>
<thead>
<tr>
<th>#</th>
<th>Start # of Conc. Bkgnd</th>
<th>#</th>
<th>Start # of Conc. Bkgnd</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>tube tubes group</td>
<td></td>
<td>tube tubes group</td>
</tr>
<tr>
<td>1</td>
<td>5 4 0.0 1</td>
<td>2</td>
<td>9 4 500.0 1</td>
</tr>
<tr>
<td>3</td>
<td>13 4 200.0 1</td>
<td>4</td>
<td>17 4 100.0 1</td>
</tr>
<tr>
<td>5</td>
<td>21 4 50.0 1</td>
<td>6</td>
<td>25 4 25.0 1</td>
</tr>
<tr>
<td>7</td>
<td>29 4 10.0 1</td>
<td>8</td>
<td>33 4 5.0 1</td>
</tr>
<tr>
<td>9</td>
<td>37 4 2.5 1</td>
<td>10</td>
<td>41 4 1.0 1</td>
</tr>
</tbody>
</table>

---**UNKNOWNs ===**

<table>
<thead>
<tr>
<th># thru</th>
<th>Start Tubes Dil. Bkgnd</th>
<th># thru</th>
<th>Start Tubes Dil. Bkgnd</th>
</tr>
</thead>
<tbody>
<tr>
<td>tube</td>
<td>/set</td>
<td>tube</td>
<td>/set</td>
</tr>
<tr>
<td>group</td>
<td></td>
<td>group</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>6 49 4 1.0 1</td>
<td>7</td>
<td>12 73 2 1.0 1</td>
</tr>
</tbody>
</table>

*TYPE yes TO SAVE THIS TUBE LAYOUT IN A WORKSHEET; ELSE [RETURN]; yes*

**NEXT TUBE #85**

After the last set of unknowns is set up, CLINFO asks the user whether this tube layout is to be saved in a worksheet. A reply of yes is required if this tube layout is to be saved.

---**Worksheets**---

**ANALYZE, 13**

**RADIOIMMUNOASSAY**

**WORKSHEET NAME** 24 ROW X 2 COLS

**TUBE LAYOUT Concentration units: nanogram**

**CREATED 8/9/77 MODIFIED 8/9/77**

**| Tubes | NSB | Stnd Conc/ group Ukn Dilu |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>1</td>
<td>4 1 0.0</td>
</tr>
<tr>
<td>Bkrd #1</td>
<td>45</td>
<td>48 1 0.0</td>
</tr>
<tr>
<td>Stnd #1</td>
<td>5</td>
<td>5 1 0.0</td>
</tr>
<tr>
<td>Stnd #2</td>
<td>9</td>
<td>12 1 500.0</td>
</tr>
<tr>
<td>Stnd #3</td>
<td>13</td>
<td>16 1 250.0</td>
</tr>
<tr>
<td>Stnd #4</td>
<td>17</td>
<td>20 1 100.0</td>
</tr>
<tr>
<td>Stnd #5</td>
<td>21</td>
<td>24 1 50.0</td>
</tr>
<tr>
<td>Stnd #6</td>
<td>25</td>
<td>28 1 25.0</td>
</tr>
<tr>
<td>Stnd #7</td>
<td>29</td>
<td>32 1 10.0</td>
</tr>
<tr>
<td>Stnd #8</td>
<td>33</td>
<td>36 1 5.0</td>
</tr>
<tr>
<td>Stnd #9</td>
<td>37</td>
<td>40 1 2.5</td>
</tr>
<tr>
<td>Stnd #10</td>
<td>41</td>
<td>44 1 1.0</td>
</tr>
<tr>
<td>Ukn #1</td>
<td>49</td>
<td>52 1 1.0</td>
</tr>
<tr>
<td>Ukn #2</td>
<td>53</td>
<td>56 1 1.0</td>
</tr>
<tr>
<td>Ukn #3</td>
<td>57</td>
<td>60 1 1.0</td>
</tr>
<tr>
<td>Ukn #4</td>
<td>61</td>
<td>64 1 1.0</td>
</tr>
<tr>
<td>Ukn #5</td>
<td>65</td>
<td>68 1 1.0</td>
</tr>
<tr>
<td>Ukn #6</td>
<td>69</td>
<td>72 1 1.0</td>
</tr>
<tr>
<td>Ukn #7</td>
<td>73</td>
<td>74 1 1.0</td>
</tr>
<tr>
<td>Ukn #8</td>
<td>75</td>
<td>76 1 1.0</td>
</tr>
<tr>
<td>Ukn #9</td>
<td>77</td>
<td>78 1 1.0</td>
</tr>
<tr>
<td>Ukn #10</td>
<td>79</td>
<td>80 1 1.0</td>
</tr>
<tr>
<td>Ukn #11</td>
<td>81</td>
<td>82 1 1.0</td>
</tr>
<tr>
<td>Ukn #12</td>
<td>83</td>
<td>84 1 1.0</td>
</tr>
</tbody>
</table>
Suboption 3 (run RIA analysis)

Suboption 3 is used to perform an analysis of RIA data stored in a worksheet.

### Worksheet Labels

<table>
<thead>
<tr>
<th>HOUSE/COILS</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>12957</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>12667</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>15851</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>6</td>
<td>12553</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>6187</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>6227</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>7</td>
<td>6344</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>6272</td>
<td>1</td>
</tr>
<tr>
<td>9</td>
<td>500</td>
<td>1450</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>100</td>
<td>1450</td>
<td>1</td>
</tr>
<tr>
<td>11</td>
<td>11</td>
<td>1571</td>
<td>1</td>
</tr>
<tr>
<td>12</td>
<td>12</td>
<td>1571</td>
<td>1</td>
</tr>
<tr>
<td>13</td>
<td>13</td>
<td>1574</td>
<td>1</td>
</tr>
<tr>
<td>14</td>
<td>14</td>
<td>1574</td>
<td>1</td>
</tr>
<tr>
<td>15</td>
<td>15</td>
<td>1549</td>
<td>1</td>
</tr>
<tr>
<td>16</td>
<td>16</td>
<td>1549</td>
<td>1</td>
</tr>
<tr>
<td>17</td>
<td>17</td>
<td>2059</td>
<td>1</td>
</tr>
<tr>
<td>18</td>
<td>18</td>
<td>2059</td>
<td>1</td>
</tr>
<tr>
<td>19</td>
<td>19</td>
<td>2114</td>
<td>1</td>
</tr>
<tr>
<td>20</td>
<td>20</td>
<td>2114</td>
<td>1</td>
</tr>
<tr>
<td>21</td>
<td>21</td>
<td>3916</td>
<td>1</td>
</tr>
<tr>
<td>22</td>
<td>22</td>
<td>3916</td>
<td>1</td>
</tr>
<tr>
<td>23</td>
<td>23</td>
<td>4128</td>
<td>1</td>
</tr>
<tr>
<td>24</td>
<td>24</td>
<td>4128</td>
<td>1</td>
</tr>
<tr>
<td>25</td>
<td>25</td>
<td>3048</td>
<td>1</td>
</tr>
<tr>
<td>26</td>
<td>26</td>
<td>3048</td>
<td>1</td>
</tr>
<tr>
<td>27</td>
<td>27</td>
<td>3614</td>
<td>1</td>
</tr>
<tr>
<td>28</td>
<td>28</td>
<td>3614</td>
<td>1</td>
</tr>
<tr>
<td>29</td>
<td>29</td>
<td>4909</td>
<td>1</td>
</tr>
<tr>
<td>30</td>
<td>30</td>
<td>4909</td>
<td>1</td>
</tr>
<tr>
<td>31</td>
<td>31</td>
<td>5726</td>
<td>2</td>
</tr>
<tr>
<td>32</td>
<td>32</td>
<td>5726</td>
<td>2</td>
</tr>
<tr>
<td>33</td>
<td>33</td>
<td>5760</td>
<td>3</td>
</tr>
<tr>
<td>34</td>
<td>34</td>
<td>5760</td>
<td>3</td>
</tr>
<tr>
<td>35</td>
<td>35</td>
<td>6077</td>
<td>3</td>
</tr>
<tr>
<td>36</td>
<td>36</td>
<td>6077</td>
<td>3</td>
</tr>
<tr>
<td>37</td>
<td>37</td>
<td>6129</td>
<td>3</td>
</tr>
<tr>
<td>38</td>
<td>38</td>
<td>6129</td>
<td>3</td>
</tr>
<tr>
<td>39</td>
<td>39</td>
<td>6129</td>
<td>3</td>
</tr>
<tr>
<td>40</td>
<td>40</td>
<td>6129</td>
<td>3</td>
</tr>
<tr>
<td>41</td>
<td>41</td>
<td>6200</td>
<td>3</td>
</tr>
<tr>
<td>42</td>
<td>42</td>
<td>6200</td>
<td>3</td>
</tr>
<tr>
<td>43</td>
<td>43</td>
<td>6200</td>
<td>3</td>
</tr>
<tr>
<td>44</td>
<td>44</td>
<td>6355</td>
<td>3</td>
</tr>
<tr>
<td>45</td>
<td>45</td>
<td>6355</td>
<td>3</td>
</tr>
<tr>
<td>46</td>
<td>46</td>
<td>6355</td>
<td>3</td>
</tr>
<tr>
<td>47</td>
<td>47</td>
<td>6355</td>
<td>3</td>
</tr>
<tr>
<td>48</td>
<td>48</td>
<td>6355</td>
<td>3</td>
</tr>
<tr>
<td>49</td>
<td>49</td>
<td>6355</td>
<td>3</td>
</tr>
<tr>
<td>50</td>
<td>50</td>
<td>6355</td>
<td>3</td>
</tr>
</tbody>
</table>

### Worksheet Labels

<table>
<thead>
<tr>
<th>HOUSE/COILS</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>51</td>
<td>u1</td>
<td>51</td>
</tr>
<tr>
<td>2</td>
<td>52</td>
<td>u2</td>
<td>52</td>
</tr>
<tr>
<td>3</td>
<td>53</td>
<td>u3</td>
<td>53</td>
</tr>
<tr>
<td>4</td>
<td>54</td>
<td>u4</td>
<td>54</td>
</tr>
<tr>
<td>5</td>
<td>55</td>
<td>u5</td>
<td>55</td>
</tr>
<tr>
<td>6</td>
<td>56</td>
<td>u6</td>
<td>56</td>
</tr>
<tr>
<td>7</td>
<td>57</td>
<td>u7</td>
<td>57</td>
</tr>
<tr>
<td>8</td>
<td>58</td>
<td>u8</td>
<td>58</td>
</tr>
<tr>
<td>9</td>
<td>59</td>
<td>u9</td>
<td>59</td>
</tr>
<tr>
<td>10</td>
<td>60</td>
<td>u10</td>
<td>60</td>
</tr>
<tr>
<td>11</td>
<td>61</td>
<td>u11</td>
<td>61</td>
</tr>
<tr>
<td>12</td>
<td>62</td>
<td>u12</td>
<td>62</td>
</tr>
<tr>
<td>13</td>
<td>63</td>
<td>u13</td>
<td>63</td>
</tr>
<tr>
<td>14</td>
<td>64</td>
<td>u14</td>
<td>64</td>
</tr>
<tr>
<td>15</td>
<td>65</td>
<td>u15</td>
<td>65</td>
</tr>
<tr>
<td>16</td>
<td>66</td>
<td>u16</td>
<td>66</td>
</tr>
<tr>
<td>17</td>
<td>67</td>
<td>u17</td>
<td>67</td>
</tr>
<tr>
<td>18</td>
<td>68</td>
<td>u18</td>
<td>68</td>
</tr>
<tr>
<td>19</td>
<td>69</td>
<td>u19</td>
<td>69</td>
</tr>
<tr>
<td>20</td>
<td>70</td>
<td>u20</td>
<td>70</td>
</tr>
<tr>
<td>21</td>
<td>71</td>
<td>u21</td>
<td>71</td>
</tr>
<tr>
<td>22</td>
<td>72</td>
<td>u22</td>
<td>72</td>
</tr>
<tr>
<td>23</td>
<td>73</td>
<td>u23</td>
<td>73</td>
</tr>
<tr>
<td>24</td>
<td>74</td>
<td>u24</td>
<td>74</td>
</tr>
<tr>
<td>25</td>
<td>75</td>
<td>u25</td>
<td>75</td>
</tr>
<tr>
<td>26</td>
<td>76</td>
<td>u26</td>
<td>76</td>
</tr>
<tr>
<td>27</td>
<td>77</td>
<td>u27</td>
<td>77</td>
</tr>
<tr>
<td>28</td>
<td>78</td>
<td>u28</td>
<td>78</td>
</tr>
<tr>
<td>29</td>
<td>79</td>
<td>u29</td>
<td>79</td>
</tr>
<tr>
<td>30</td>
<td>80</td>
<td>u30</td>
<td>80</td>
</tr>
<tr>
<td>31</td>
<td>81</td>
<td>u31</td>
<td>81</td>
</tr>
<tr>
<td>32</td>
<td>82</td>
<td>u32</td>
<td>82</td>
</tr>
<tr>
<td>33</td>
<td>83</td>
<td>u33</td>
<td>83</td>
</tr>
<tr>
<td>34</td>
<td>84</td>
<td>u34</td>
<td>84</td>
</tr>
<tr>
<td>35</td>
<td>85</td>
<td>u35</td>
<td>85</td>
</tr>
<tr>
<td>36</td>
<td>86</td>
<td>u36</td>
<td>86</td>
</tr>
<tr>
<td>37</td>
<td>87</td>
<td>u37</td>
<td>87</td>
</tr>
<tr>
<td>38</td>
<td>88</td>
<td>u38</td>
<td>88</td>
</tr>
<tr>
<td>39</td>
<td>89</td>
<td>u39</td>
<td>89</td>
</tr>
<tr>
<td>40</td>
<td>90</td>
<td>u40</td>
<td>90</td>
</tr>
</tbody>
</table>

**CLINFO** first prompts for the name of the tube layout worksheet. In this example, the current worksheet contains the tube layout, so the RETURN key is hit in response. The system next prompts for the name of the worksheet containing the data for the analysis. The RIA option assumes that data are in columns and requires a column of tube numbers, a column of row counts, and, if the counting times are not the same for all tubes, a column of counting times. The column labels may be displayed. The user specifies the number of the column that contains the tube numbers and either the one that contains the counts per minute (actually the counts per unit time) or those that contain counts and counting times.
CLINFO next prompts for the form of the response variable to be used in this analysis. In this example, response variable 4, BOUND/B0, is selected because in this laboratory analysis the precipitate was counted.

The system then calibrates the standard curve and displays some summary information about the standard. It is important to check these numbers before proceeding further. When response variable 3 or 4 is used, the value of the slope should be close to \(-1.0\). Large deviations from this value indicate that some of the samples in this run are probably contaminated. The standard error (Std. Err.) provides a measure of the variation of the standard points about the computed standard curve. This number should not vary greatly from analysis to analysis for the same substance. A value much greater than normal standard error is also an indication of possible problems with the analysis. After displaying these results, the system prompts for a RETURN key to wait for the user to scrutinize, and possibly print, the results.
CLINFO next plots the fitted standard curve and the actual standard points on the terminal. If any of the standard points has an undefined value of the response variable, it is plotted as having a value of 12 so that it is removed from the rest of the curve. By responding yes and then x to the next two prompts, the user obtains a plot of the standard curve on the attached printer. At this point, it is important to examine the standard curve for evidence of significant non-linearity or bad outlying points. In this example, the standard curve is deemed to be adequate, so the user requests results for the unknowns by responding yes to the next prompt.
The system finally displays the results. These may be stored in a worksheet by responding yes to the appropriate prompt. The results for each unknown give the mean value of the response variable, the logit of the mean (for all response variables but free/bound), and the estimated concentration based on the mean. If the mean value of the response variable is outside the range used to estimate the standard curve, or there is a large amount of variance within the replicate for an unknown, a warning message is printed. In this example, the logit of 9.2 is outside the range of the standards. If B/B0 is greater than 1 for an unknown, no estimate can be made.

<table>
<thead>
<tr>
<th>TUBES</th>
<th>B/B0</th>
<th>LOGIT</th>
<th>CONC.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unk #1</td>
<td>49</td>
<td>52</td>
<td>.64</td>
</tr>
<tr>
<td>Unk #2</td>
<td>53</td>
<td>56</td>
<td>.97</td>
</tr>
<tr>
<td>Unk #3</td>
<td>57</td>
<td>60</td>
<td>4.2</td>
</tr>
<tr>
<td>Unk #4</td>
<td>61</td>
<td>64</td>
<td>5.9</td>
</tr>
<tr>
<td>Unk #5</td>
<td>65</td>
<td>68</td>
<td>2.6</td>
</tr>
<tr>
<td>Unk #6</td>
<td>69</td>
<td>72</td>
<td>1.1</td>
</tr>
<tr>
<td>Unk #7</td>
<td>73</td>
<td>74</td>
<td>4.4</td>
</tr>
<tr>
<td>Unk #8</td>
<td>75</td>
<td>76</td>
<td>4.4</td>
</tr>
<tr>
<td>Unk #9</td>
<td>77</td>
<td>78</td>
<td>4.5</td>
</tr>
<tr>
<td>Unk #10</td>
<td>79</td>
<td>80</td>
<td>4.1</td>
</tr>
<tr>
<td>Unk #11</td>
<td>81</td>
<td>82</td>
<td>7.3</td>
</tr>
<tr>
<td>Unk #12</td>
<td>83</td>
<td>84</td>
<td>.65</td>
</tr>
</tbody>
</table>

TYPE yes TO SAVE RESULTS IN A WORKSHEET; ELSE (RETURN):

* EXCESSIVE AMOUNT OF VARIATION AMONG REPLICATES FOR THIS UNKNOWN ----

** ---------- VALUE OUT OF RANGE OF STANDARD CURVE ----------

The following three figures illustrate the results obtained when response variable 1 (free/bound) is used. The estimated concentrations are similar to, but different from, those obtained using response variable 4.
**ANALYZE, 13**

**RADIOIMMUNOASSAY**

**WORKSHEET** RPSHLAY 24 ROW X 2 COLS

**TITLE** TUBE LAYOUT  Concentration units: nanograms

**CREATED** 8/9/77  **MODIFIED** 8/9/77

---

**CONCENTRATIONS**

<table>
<thead>
<tr>
<th>TUBES</th>
<th>B/F</th>
<th>CONC.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unk #1</td>
<td>49</td>
<td>52</td>
</tr>
<tr>
<td>Unk #2</td>
<td>53</td>
<td>56</td>
</tr>
<tr>
<td>Unk #3</td>
<td>57</td>
<td>68</td>
</tr>
<tr>
<td>Unk #4</td>
<td>61</td>
<td>64</td>
</tr>
<tr>
<td>Unk #5</td>
<td>65</td>
<td>68</td>
</tr>
<tr>
<td>Unk #6</td>
<td>69</td>
<td>72</td>
</tr>
<tr>
<td>Unk #7</td>
<td>73</td>
<td>74</td>
</tr>
<tr>
<td>Unk #8</td>
<td>75</td>
<td>76</td>
</tr>
<tr>
<td>Unk #9</td>
<td>77</td>
<td>78</td>
</tr>
<tr>
<td>Unk #10</td>
<td>79</td>
<td>80</td>
</tr>
<tr>
<td>Unk #11</td>
<td>81</td>
<td>82</td>
</tr>
<tr>
<td>Unk #12</td>
<td>83</td>
<td>84</td>
</tr>
</tbody>
</table>

**Pooled Variance for B/F = 0.374415**

---

*Excesive amount of variation among replicates for this unknown ****

** Value out of range of standard curve ========
CALCULATE

The calculate activity gives the user the flexibility of performing arbitrary arithmetic, logical, time difference, and statistical computations on data stored in designated rows or columns of a worksheet. Like a desk calculator that operates on worksheets, it may be used to perform calculations that are not provided for by available CLINFO options.

The input to a calculate subactivity may be the data contained in one or more rows or columns of the current worksheet; the output is recorded in a designated row or column of the same worksheet. When calculate subactivities are first invoked, the user indicates whether he wants to perform computations on data in rows or in columns of the worksheet. The choice of rows generally means that data in one or more rows will be operated upon, column by column, and the results will be stored in another row. Thus, “add two rows” usually means add the data value in column 1 of one row to the value in column 1 of the other row and store the sum in a third row, add the data values in the second columns of the two rows and store the sum in the second column of the third row, etc. When calculations are performed on columns, data within the specified columns are similarly processed row by row.

If the user chooses arithmetic expression (calculate, option 1), he intends to define the subsequent calculations by an expression that may include row (or column) numbers, real numbers, arithmetic operators, and system-defined functions. For example, he may wish to accomplish the following calculations:

Row 3 = (Row 4 + 5.6) • Row 7

If he chooses AND or OR logical expressions, he intends to write a combination of arithmetic and logical expressions that define a set of conditions. With AND expressions, the result to be stored in the designated row (or column) is 1 if all the conditions are satisfied; otherwise it is 0. With OR, the result is 1 if any of the conditions is satisfied; otherwise it is 0. As an example,

Set Row 7 = 1 if Row 3 is greater than 5.6

Special calculations refer to such operations as replacing missing values in a row (or column) with a constant and computing the cumulative sum of values within a row (or column). Time-difference calculations can be performed for either dates, times, or dates and times, and the units of the results can be specified as anything from minutes to years. Means, standard errors, and standard deviations as well as counts and sums can be calculated using the Statistical Calculations option. The entire range of choices is as follows:

<table>
<thead>
<tr>
<th>CALCULATE OPTIONS</th>
</tr>
</thead>
<tbody>
<tr>
<td>TYPE TO</td>
</tr>
<tr>
<td>0. SELECT A WORKSHEET</td>
</tr>
<tr>
<td>1. ARITHMETIC EXPRESSION</td>
</tr>
<tr>
<td>2. LOGICAL EXPRESSIONS (AND)</td>
</tr>
<tr>
<td>3. LOGICAL EXPRESSIONS (OR)</td>
</tr>
<tr>
<td>4. SPECIAL CALCULATIONS</td>
</tr>
<tr>
<td>5. TIME DIFFERENCE</td>
</tr>
<tr>
<td>6. STATISTICAL CALCULATIONS</td>
</tr>
</tbody>
</table>
The arithmetic operators are:

<table>
<thead>
<tr>
<th>Operator</th>
<th>Meaning</th>
<th>Terminal Shift and Key</th>
</tr>
</thead>
<tbody>
<tr>
<td>^</td>
<td>Exponentiation</td>
<td>Lower-case RS</td>
</tr>
<tr>
<td>*</td>
<td>Multiplication</td>
<td>Upper-case *;</td>
</tr>
<tr>
<td>/</td>
<td>Division</td>
<td>Lower-case ?/</td>
</tr>
<tr>
<td>+</td>
<td>Addition</td>
<td>Upper-case +;</td>
</tr>
<tr>
<td>-</td>
<td>Subtraction</td>
<td>Lower-case -=</td>
</tr>
</tbody>
</table>

An arithmetic expression (expr) is evaluated in the following order, proceeding from left to right:

1. Any expr within parentheses is evaluated before any expr that is not in parentheses. When parenthetical exprs are nested, the innermost expr is always evaluated first.
2. Exponentiation is performed.
3. Multiplication and division are performed, with equal priority.
4. Addition and subtraction are performed, with equal priority.
5. When two operators (e.g., * and /) are of equal precedence, evaluation proceeds from left to right.

One consequence of this sequence is that the expression A/B*C is equivalent to (A/B)*C, not to A/(B*C), i.e., A is divided by B, and the quotient is multiplied by C.

Relational operators are used to compare two exprs in a logical expression. These expressions are combined with ANDs and ORs. A logical expression is of the form

expr1 relational operator expr2

The relational operators are:

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Meaning</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>=</td>
<td>Equal</td>
<td>A=B</td>
</tr>
<tr>
<td>&lt;</td>
<td>Less than</td>
<td>A &lt; B</td>
</tr>
<tr>
<td>&lt;=</td>
<td>Less than or equal</td>
<td>A &lt;= B</td>
</tr>
<tr>
<td>&gt;</td>
<td>Greater than</td>
<td>A &gt; B</td>
</tr>
<tr>
<td>&gt;=</td>
<td>Greater than or equal</td>
<td>A &gt;= B</td>
</tr>
<tr>
<td>&lt;&gt;</td>
<td>Not equal</td>
<td>A &lt;&gt; B</td>
</tr>
</tbody>
</table>

An example of a logical expression is

col 5 >= 6 • col 7

A function performs a system-defined calculation upon the value of an arithmetic expression, e.g., functions can compute the square root or the logarithm of a single number that results from evaluating an arithmetic expression. When applied to a row (or column) of values, the calculation is performed on each constituent value. A function may be used as an expression or may be included as part of an expression. It has a three-character mnemonic name and is followed by an arithmetic expression in parentheses, upon which it performs its calculation. The following functions are available:
Function               Value Produced

rnd(expr)             Random number between 0 and 1
sgn(expr)             The algebraic sign of expr (i.e., +1 or -1)
int(expr)             The integer value of expr
abs(expr)             The absolute value of expr
sqr(expr)             The square root of expr (expr≥0)
exp(expr)             e raised to the power expr (−178≤ expr≤175)
log(expr)             The natural logarithm of expr (expr>0)
fnl(expr)             The logarithm to the base 10 of expr (expr>0)
sin(expr)             The sine of expr (expr in radians)
cos(expr)             The cosine of expr (expr in radians)
tan(expr)             The tangent of expr (expr in radians)
atan(expr)            The arctangent of expr (result in radians)

An example of a function used as part of an arithmetic expression is

col 5 = log(3.14*col 7)

As in most other CLINFO activities, rows and columns are referred to by number. To aid the user in determining which numbers to use, the system will display row or column labels at appropriate places in the interaction.

Because expressions can be written that refer both to real numbers and to row or column numbers, a special symbol must be used to distinguish them. The symbol # is used in this activity when referring to a row or column number in an expression (in options 1, 2, 3 and 5), e.g., #7 refers to row 7 if performing row calculations and to column 7 if performing column calculations. It is not possible to refer to both rows and columns in the same arithmetic or logical expression. In the expression 3*#3, the first 3 refers to the real number and the second 3 refers to row or column 3. Since this notation is unusual, the system reminds the user at key points. To avoid confusion, # may be optionally used whenever referring to a row or column number, e.g., # may be used when specifying a bounding row or column number even though CLINFO does not require this. Furthermore, the notation #r,c may be used in calculate options 1, 2, and 3 to refer to the contents of a worksheet cell, e.g., #3,7 is used to refer to the contents of row 3, column 7 in an arithmetic or logical expression.

Because CLINFO worksheets may contain missing values (displayed as three periods, . . .), and because calculations involving missing values are not generally well defined, this activity treats missing values as special cases. The system follows two general rules:

1. If a value must be known in order to compute a result and that value is missing, the result is undefined and therefore its value is missing as well.
2. Whenever a row or column calculation encounters a missing value, the system informs the user.

One consequence of the first rule is that the value of any arithmetic expression is missing if any constituent value is missing. If row calculations are being performed, this rule is applied on a column-by-column basis; if column calculations are performed, it is applied on a row-by-row basis.
The application of the first rule to logical expressions is a little more complicated. If a result is true (i.e., has the value 1) when A is true OR B is true (e.g., if \(\text{glucose}>200\) OR \(\text{insulin}<50\)), and B has a missing value, the system examines the value of A. If A is true, the result is true (because it does not depend on B); if A is false or missing, the result is missing (because it cannot be determined without knowing the value of B). If a result is true when A is true AND B is true and the value of B is missing, the system again examines the value of A. If A is false, the result is false (because it would be false regardless of the value of B); if A is true or missing, the result is missing (because it depends on the value of B). This logic is not quite correct, because there are cases where if the system were "smart" enough, it could determine the result (e.g., when the the result is true if \(A=B\) OR \(A<>B\)). The second rule above is a means of warning the user when such a situation is possible. The application of the first rule to logical expressions is summarized in the following table:

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>A OR B</th>
<th>A AND B</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td>T</td>
<td>F</td>
<td>T</td>
<td>F</td>
</tr>
<tr>
<td>T</td>
<td>...</td>
<td>T</td>
<td>...</td>
</tr>
<tr>
<td>F</td>
<td>T</td>
<td>T</td>
<td>F</td>
</tr>
<tr>
<td>F</td>
<td>F</td>
<td>F</td>
<td>F</td>
</tr>
<tr>
<td>F</td>
<td>...</td>
<td>...</td>
<td>F</td>
</tr>
<tr>
<td>...</td>
<td>T</td>
<td>T</td>
<td>...</td>
</tr>
<tr>
<td>...</td>
<td>F</td>
<td>F</td>
<td>F</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

A and B are logical expressions that may be true (T), false (F), or missing (...). The value of A OR B or A AND B also may be T, F or ... The value stored in a worksheet is actually the number 1 for true, 0 for false, or missing.

The subactivities Statistical Calculations and Special Calculations allow the user to handle missing values in other ways. For example, the former counts the non-missing values in a column or row, and the latter converts all missing values in a column or row to a specified real number.

When performing row calculations, the user may not wish to use all the columns in the rows of interest, or he may not wish to store the results in an entire row. To handle this situation, the system allows him to limit the calculations to a specified column range. A similar limitation may be placed on rows when performing column calculations.

The results of a row calculation are stored in a row of the current worksheet; the results of a column calculation are stored in a column. Results will be stored in whatever row or column is specified—including one involved in the calculation, one that is not involved but contains data, or one that contains only missing values. It is important, therefore, to be careful when specifying where the results are to be stored.

When the user designates the row or column in which results are to be stored, he is prompted for a label. He may either enter an eight-character label or hit the RETURN key to leave the existing label unchanged (and possibly missing).
Once a calculation is completely specified, the system performs it. Usually, the calculation successfully goes to completion and the user is so informed. In some cases, e.g., when there is an attempted division by zero or an attempted square root of a negative number, the calculation is aborted. In such cases, the result row or column is left unchanged; not even the successfully computed values are stored.

Upon completion of a calculation, the user may indicate, by responding to a prompt, that he wants to display the current worksheet (and examine the results). The effect is similar to typing !work, 1 (to go to worksheet, option 1), with the important exception that after looking at the worksheet, the user may respond yes to the prompt TYPE YES TO CONTINUE WITH calculate, n and return directly to the current subactivity, i.e., to option n of calculate.

**Dialog**

*Calculate* prompts with

```
<table>
<thead>
<tr>
<th>TYPE TO</th>
</tr>
</thead>
<tbody>
<tr>
<td>8. SELECT A WORKSHEET</td>
</tr>
<tr>
<td>1. ARITHMETIC EXPRESSION</td>
</tr>
<tr>
<td>2. LOGICAL EXPRESSIONS (AND)</td>
</tr>
<tr>
<td>3. LOGICAL EXPRESSIONS (OR)</td>
</tr>
<tr>
<td>4. SPECIAL CALCULATIONS</td>
</tr>
<tr>
<td>5. TIME DIFFERENCE</td>
</tr>
<tr>
<td>6. STATISTICAL CALCULATIONS</td>
</tr>
</tbody>
</table>
```

The user responds with a number from 0 to 6, and the corresponding option is invoked.

**OPTION 0: SELECT A WORKSHEET**

This subactivity performs the same function as *worksheet*, option 0, and the dialog is identical. It is included in *calculate* for the convenience of the user who wants to select a worksheet upon entering *calculate* or who wants to change the current worksheet without leaving this activity. Also, the system takes the user directly to Select a Worksheet if he attempts to invoke any other *calculate* subactivity (e.g., Special Calculations) when there is no current worksheet. In that case, he is automatically returned to the initially specified subactivity after he selects a worksheet.

Several uses of *calculate* are illustrated in the following. All the examples except that for the Time Difference and Statistical Calculations subactivities apply to the worksheet:
OPTION 1: ARITHMETIC EXPRESSION

This option provides for defining and evaluating arbitrary arithmetic expressions (e.g., Row 3 = Row7+(5.6+Row4)) that refer to data in worksheet cells and either rows or columns. It stores results in specified rows or columns of the same worksheet. Appropriate arithmetic operators and functions and the treatment of missing values are described above under Calculate. After the user indicates that calculations are to involve rows or columns, the symbol # must be used in expressions, as #n, to designate a particular row or column number. This symbol may be optionally used when specifying result and bounding rows and columns. The notation #r,c is used in an expression to refer to the data in row r, column c of the current worksheet.

We shall begin with several arithmetic calculations on rows of data. The user enters local,1 to invoke this subactivity and types r to indicate that the calculations apply to rows.

First, we shall assume that the user wants to add the column values in row 1 to those in row 2 and store the results in corresponding columns of row 4; this is called adding rows 1 and 2. After selecting row calculations he displays the row labels. He enters 4 as the number of the result row, names the result row label 1+2 so that it clearly identifies the contents of that row, and then types the arithmetic expression #1+#2 to define the calculation. Because imbedded blanks are of no consequence, the expressions #1+#2 and #1+#2 are equivalent.
The system next informs the user of its interpretation of his specification. ROW #4 designates the result row, 1+2 in parentheses is to become the label for that row, and #1 + #2 is the expression to be evaluated. When given an opportunity to restrict the range of columns involved in the calculation, the user hits the RETURN key to indicate that he does not care to do so.

ROW #4 (1+2 )=1+2

This calculation will be performed for columns 1 to 6
Type yes to select another range of columns, else hit [RETURN]:

Finally, the system repeats the calculation specification (including the column range), indicates that the calculation is in progress, indicates that it is complete, warns that missing values were encountered, and then lets the user display his worksheet directly (i.e., without explicitly going to the worksheet activity). Each calculation specification involves all the steps just described, but the following examples do not illustrate them in detail.

The worksheet now appears as shown below:

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>LABELS</td>
<td>col 1</td>
<td>col 2</td>
<td>col 3</td>
<td>col 4</td>
<td>col 5</td>
<td>col 6</td>
</tr>
<tr>
<td>1 row #1</td>
<td>2</td>
<td>4</td>
<td>6</td>
<td>8</td>
<td></td>
<td>12</td>
</tr>
<tr>
<td>2 row #2</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>5</td>
<td></td>
<td>8</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>4 #1+2</td>
<td>2</td>
<td></td>
<td></td>
<td>11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note that the column-by-column sum of the first two rows is now stored in the first four columns of row 4. The values in columns 5 and 6 of row 4 are missing; the sums that would be stored in these columns cannot be computed because row 1 has a missing value in column 5 and row 2 has a missing value in column 6.

To continue with arithmetic expressions for rows, the user responds yes to the prompt at the bottom of the worksheet display. If the user wishes to stay in this subactivity but perform calculations on columns, he should input /I instead of yes in order to have an opportunity to choose between row and column calculations. Again, these steps are repeated each time, although they are not illustrated in the following examples.

In the next example, the user explores the effect of restricting the data that are processed during row calculations to those in particular columns. He first defines the result row, the result row label, and the arithmetic expression.
ARITHMETIC EXPRESSION INVOLVING ROWS
RESULT ROW: 5
LABEL FOR ROW ##5(1-2): 1-2
ALL ROW NUMBERS MUST BE PRECEDED BY A # SIGN
TYPE THE EXPRESSION: #1-#2

This time, he responds yes to select a column range other than 1 to 6 and responds yes again when asked if he wants to see the column labels. He then defines the range of columns in which data are to be processed by entering the numbers of the starting and ending columns of the range.

ROW ##5(1-2) =#1-#2

<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>col</td>
<td>col</td>
<td>col</td>
<td>col</td>
<td>col</td>
<td>col</td>
<td>col</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td></td>
</tr>
</tbody>
</table>

STARTING COLUMN: 2
ENDING COLUMN: 4

When the system displays the complete calculation specification, the listed starting and ending column numbers are 2 and 4 (the numbers he has just entered), rather than 1 and 12 as in the previous example.

When the results are displayed, row 5 contains missing values in columns 5 and 6 for the same reason explained above. However, in this case, the result in row 5, column 1, is also missing because column 1 was excluded from the calculation. In general, when column data are not processed, the value that was previously stored in the corresponding column of the result row is left unchanged. If the value 13.6 had been stored in row 5, column 1, prior to the calculation, then 13.6 would remain there. In this case, the value previously stored in that cell was missing so it remains missing.

ROWS/COLS

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>LABELS</td>
<td>col #1</td>
<td>col #2</td>
<td>col #3</td>
<td>col #4</td>
<td>col #5</td>
<td>col #6</td>
</tr>
<tr>
<td>1 row #1</td>
<td>2</td>
<td>4</td>
<td>6</td>
<td>8</td>
<td>***</td>
<td>12</td>
</tr>
<tr>
<td>2 row #2</td>
<td>*</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>***</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>5</td>
<td>8</td>
<td>11</td>
<td>***</td>
<td>***</td>
</tr>
<tr>
<td>4 1-2</td>
<td>***</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>***</td>
<td>***</td>
</tr>
<tr>
<td>5 1-2</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
</tr>
</tbody>
</table>

Next, a calculation is defined which would set values in row 6 to the ratio of those in rows 1 and 2. Because row 2 contains a zero in column 1 and division by zero is not allowed, the calculation terminates. No results are stored in row 6 even though there are no other zero values in row 2. To compute those ratios that can be computed, the user could exclude column 1 from the calculation as described above. Another alternative would be to set zeros to missing values, as described below under Special Calculations.

To evaluate a more complex expression involving other arithmetic operators and both row numbers and real numbers, the user defines
ROW #7(expr ) = #1* #2 ∧ 3 – #4/2

where #2 ∧ 3 means raise values in row 2 to the third power. Note that the expression includes row 4, one of the computed rows, in addition to the two original rows.

When the user defines

ROW #8(log ) = log( #2)

the computation is terminated and

ARGUMENT RANGE FOR FUNCTION EXCEEDED.

is displayed. This is because the natural logarithm of zero (in row 2, column 1) cannot be evaluated.

After performing the calculations defined by

ROW #8(log ) = log( #1/2)

and by

ROW #9(int ) = #2 + 2*(int( #8) + 1)

where int( #8) means take the integer part of values in row 8, the worksheet has the contents

<table>
<thead>
<tr>
<th>ROWS/COLS</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>LABELS</td>
<td>col 1</td>
<td>col 2</td>
<td>col 3</td>
<td>col 4</td>
<td>col 5</td>
<td>col 6</td>
</tr>
<tr>
<td>1 row #1</td>
<td>2</td>
<td>4</td>
<td>6</td>
<td>8</td>
<td>...</td>
<td>12</td>
</tr>
<tr>
<td>2 row #2</td>
<td>8</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>...</td>
</tr>
<tr>
<td>3 expr</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>8</td>
<td>11</td>
<td>...</td>
</tr>
<tr>
<td>4 1+2</td>
<td>2</td>
<td>5</td>
<td>8</td>
<td>11</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>5 1-2</td>
<td>...</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>7 expr</td>
<td>1</td>
<td>1.580</td>
<td>44</td>
<td>218.590</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>8 log</td>
<td>2.26208-08</td>
<td>.693</td>
<td>1.099</td>
<td>1.386</td>
<td>...</td>
<td>1.792</td>
</tr>
<tr>
<td>9 int</td>
<td>2</td>
<td>3</td>
<td>6</td>
<td>7</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>10</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

Note that whereas log(1)=0, the computed value in row 8, column 1 represents a small computational error.

To explore the use of column calculations, we use retrieve, option 4, to copy the values and labels in rows 1 through 10 and in columns 1 through 3 of worksheet ROWCOMP into a new worksheet called COLCOMP. The user then returns to calculate, option 1, and indicates that he wants to perform column calculations. By interacting with the system just as he would when specifying row calculations, he specifies

COLUMN #5(1+2...) = #1+ #2

and

COLUMN #6(1-2...) = #1- #2
When asked by the system if he wants to restrict the column calculations to certain rows, the user declines to do so; as a result, all ten rows copied into the worksheet take part in the calculations. After the calculations, the display of COL-COMP looks like:

**WORKSHEET COL-COMP**

**TITLE**: Illustrate the use of column calculations

*CREATED 11/11/77*  
*MODIFIED 11/11/77*

*OF ROWS*: 20  
*OF COLS*: 12

<table>
<thead>
<tr>
<th>ROWS/COLS</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>LABELS</td>
<td>col #1</td>
<td>col #2</td>
<td>col #3</td>
<td>1+2</td>
<td>1-2</td>
<td></td>
</tr>
<tr>
<td>1 row #1</td>
<td>2</td>
<td>4</td>
<td>6</td>
<td>...</td>
<td>6</td>
<td>-2</td>
</tr>
<tr>
<td>2 row #2</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>...</td>
<td>1</td>
<td>-1</td>
</tr>
<tr>
<td>3</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>4 1+2</td>
<td>7</td>
<td>5</td>
<td>8</td>
<td>...</td>
<td>7</td>
<td>-3</td>
</tr>
<tr>
<td>5 1-2</td>
<td>...</td>
<td>3</td>
<td>4</td>
<td>...</td>
<td>3</td>
<td>...</td>
</tr>
<tr>
<td>6</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>7 expr</td>
<td>1.508</td>
<td>44</td>
<td>...</td>
<td>...</td>
<td>1.508</td>
<td>-2.508</td>
</tr>
<tr>
<td>8 log</td>
<td>2.26208e-08</td>
<td>.693</td>
<td>1.099</td>
<td>...</td>
<td>.693</td>
<td>-.693</td>
</tr>
<tr>
<td>9 int</td>
<td>3</td>
<td>3</td>
<td>6</td>
<td>...</td>
<td>5</td>
<td>-1</td>
</tr>
<tr>
<td>10</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>11</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>12</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>13</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>14</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>15</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>16</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>17</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>18</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>19</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>20</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

(Note: ... indicates missing values)

**TYPE**: yes to continue with calculate, 1 ELSE  
**TYPE**: the starting row #:

---

**OPTION 2: LOGICAL EXPRESSIONS (AND)**

This option provides for defining and evaluating arbitrary logical expressions involving the AND operator. It stores results in specified rows or columns of the current worksheet. Logical expressions are used to determine when all or any of a set of conditions is satisfied. Each condition is defined by an expression such as #3<5 or #2>1/2. If all conditions must be satisfied, an AND expression is formed (i.e., a logical expression like #3<5 AND #2>1/2); if any condition must be satisfied, an OR expression is formed (i.e., a logical expression like #3<5 OR #2>1/2). If a set of expressions is satisfied (i.e., if all (for AND) or any (for OR) of the constituent expressions is true), the result is the number 1; otherwise, the result is zero. Missing values are treated as explained above under Calculate.

The specification of calculations involving logical expressions is similar to that of calculations involving arithmetic expressions. One major difference is that whereas only one arithmetic expression is used in specifying a calculation, several logical expressions may be combined, as just described. Another difference is that logical expressions include arithmetic expressions whose values are compared using relational operators (listed above under Calculate), but the reverse is not true.
The specification of row (or column) calculations based on logical expressions involves the following steps:

1. Invoke calculate, option 2, for AND; option 3, for OR.
2. Choose row (or column) calculations.
3. Display the row (column) labels (this step is optional).
4. Enter the result row (column) number and label.
5. Type a series of logical AND or OR expressions (depending on which option was chosen in step 1); terminate the series by hitting the RETURN key.
6. Restrict the calculations to a range of columns (rows) after displaying the column (row) labels (both of these steps are optional).
7. The system lists the complete specification, then carries out the calculations. The calculation may be terminated if a constituent arithmetic expression cannot be evaluated. If missing values are encountered, the system so indicates.
8. Display the results stored in the current worksheet (optional).

ROWCOMP is the current worksheet in all the following examples.
After the user completes the first four steps listed above (with 10 as the result row number) and types the logical expression \#2>=1, the terminal screen looks like:

calculate.2 LOGICAL EXPRESSIONS (AND)
WORKSHEET ROWCOMP
TITLE Illustrate the use of row/column calculations
CREATED 11/11/77 MODIFIED 11/11/77
# OF ROWS 20
# OF COLS 12

LOGICAL EXPRESSIONS (AND) INVOLVING ROWS
TYPE yes TO SEE ROW LABELS, ELSE HIT (RETURN): yes
1 row #1
 2 row #2
 3
 4 1+2
 5 1-2
 6
 7 expr
 8 log
 9 int

RESULT ROW: 10
LABEL FOR ROW #10( ): AND.
ALL ROW NUMBERS MUST BE PRECEDED BY A # SIGN
TYPE A SERIES OF LOGICAL EXPRESSIONS TO BE ANDED TOGETHER.
STOP BY HITTING (RETURN) IN RESPONSE TO A PROMPT FOR ANOTHER EXPRESSION.
LOGICAL EXPRESSION 1: \#2>=1

After the second logical expression \#1<8 is entered, the lower portion looks like:

ROW #10(AND )=1 IF \#2>=1 AND \#1<8
ROW #10(AND )=0 OTHERWISE
ALL ROW NUMBERS MUST BE PRECEDED BY A # SIGN
TYPE A SERIES OF LOGICAL EXPRESSIONS TO BE ANDED TOGETHER.
STOP BY HITTING (RETURN) IN RESPONSE TO A PROMPT FOR ANOTHER EXPRESSION.
LOGICAL EXPRESSION 3:
and just after the calculation is completed, the following is displayed:

ROW #10 (AND 1) = 1 IF #2 > 1 AND #1 < 8
ROW #10 (AND 1) = 0 OTHERWISE
STARTING COLUMN: #1
ENDING COLUMN: #6
CARRYING OUT CALCULATION
CALCULATION COMPLETED
WARNING: THE CALCULATIONS PERFORMED INVOLVED MISSING VALUES.
TYPE yes TO DISPLAY THE RESULTS, ELSE HIT [RETURN]: yes

The results are

ROWS/COLS 1 2 3 4 5 6

<table>
<thead>
<tr>
<th>LABELS</th>
<th>col #1</th>
<th>col #2</th>
<th>col #3</th>
<th>col #4</th>
<th>col #5</th>
<th>col #6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 row</td>
<td>2</td>
<td>4</td>
<td>6</td>
<td>8</td>
<td>...</td>
<td>12</td>
</tr>
<tr>
<td>2 row</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>...</td>
</tr>
<tr>
<td>3</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>4 1+2</td>
<td>2</td>
<td>5</td>
<td>8</td>
<td>11</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>5 1-2</td>
<td>...</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>6</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>7 exp</td>
<td>...</td>
<td>1.500</td>
<td>44</td>
<td>210.500</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>8 log</td>
<td>2.2628E-08</td>
<td>0.693</td>
<td>1.099</td>
<td>1.386</td>
<td>...</td>
<td>1.792</td>
</tr>
<tr>
<td>9 int</td>
<td>2</td>
<td>3</td>
<td>6</td>
<td>7</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>10 AND</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>...</td>
<td>0</td>
</tr>
<tr>
<td>11</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

The result in column 5 is missing, because even though #2 > 1 is true (because the value in row 2, i.e., the number 4, is greater than or equal to 1), it cannot be determined if the corresponding value in row 1 is less than 8. The result in column 6 is 0 (i.e., false) even though the corresponding value in row 2 is missing, because the value in row 1 is greater than 8 (and #1 < 8 is false); whenever one of the conditions is not satisfied, it is not possible for the overall result to be anything but false.

In another example, the user enters the four logical expressions shown below:

LOGICAL EXPRESSIONS (AND) INVOLVING ROWS
ROW 11 (AND4expr) = 1 IF int(#4/#6) > 2 AND #8 > 5 AND #9 < 5 AND #10 > 3
ROW 11 (AND4expr) = # OTHERWISE
ALL ROW NUMBERS MUST BE PRECEDED BY A # SIGN
TYPE A SERIES OF LOGICAL EXPRESSIONS TO BE ANDED TOGETHER.
STOP BY HITTING [RETURN] IN RESPONSE TO A PROMPT FOR ANOTHER EXPRESSION.
LOGICAL EXPRESSION 5:

The results are

ROWS/COLS 1 2 3 4 5 6

<table>
<thead>
<tr>
<th>LABELS</th>
<th>col #1</th>
<th>col #2</th>
<th>col #3</th>
<th>col #4</th>
<th>col #5</th>
<th>col #6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 row</td>
<td>2</td>
<td>4</td>
<td>6</td>
<td>8</td>
<td>...</td>
<td>12</td>
</tr>
<tr>
<td>2 row</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>...</td>
</tr>
<tr>
<td>3</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>4 1+2</td>
<td>2</td>
<td>5</td>
<td>8</td>
<td>11</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>5 1-2</td>
<td>...</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>6</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>7 exp</td>
<td>...</td>
<td>1.500</td>
<td>44</td>
<td>210.500</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>8 log</td>
<td>2.2628E-08</td>
<td>0.693</td>
<td>1.099</td>
<td>1.386</td>
<td>...</td>
<td>1.792</td>
</tr>
<tr>
<td>9 int</td>
<td>2</td>
<td>3</td>
<td>6</td>
<td>7</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>10 AND</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>...</td>
<td>0</td>
</tr>
<tr>
<td>11 AND4expr</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>...</td>
<td>0</td>
</tr>
<tr>
<td>12</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>
OPTION 3: LOGICAL EXPRESSIONS (OR)

This option provides for defining and evaluating arbitrary logical expressions involving the OR operator. It stores results in specified rows or columns of the current worksheet. The eight steps listed above for logical calculations apply to option 3 as well as option 2. Relational operators and the treatment of missing values are described above under Calculate. In the first example of logical expressions requiring OR, two expressions (\#2>=1 and \#1<8) are entered. The screen then displays the completed ORed statements:

LOGICAL EXPRESSIONS (OR) INVOLVING ROWS
ROW \#12(OR \#1) =1 IF \#2>=1 OR \#1<8
ROW \#12(OR \#1) =0 OTHERWISE
ALL ROW NUMBERS MUST BE PRECEDED BY A # SIGN.
TYPE A SERIES OF LOGICAL EXPRESSIONS TO BE ORed TOGETHER.
STOP BY HITTING [RETURN] IN RESPONSE TO A PROMPT FOR ANOTHER EXPRESSION.

LOGICAL EXPRESSION 3:

At this point, the user may continue to enter statements requiring OR, or he may hit the RETURN key to terminate the procedure. Note that these are the same two expressions entered in the first AND example, but here they use OR instead of AND.

In a second example, the specification is

ROW \#13(OR expr) =1 IF \#11=\#12 OR \#9<5
ROW \#13(OR expr) =0 OTHERWISE
STARTING COLUMN: \#1
ENDING COLUMN: \#6
CARRYING OUT CALCULATION
CALCULATION COMPLETED
WARNING: THE CALCULATIONS PERFORMED INVOLVED MISSING VALUES.
TYPE yes TO DISPLAY THE RESULTS, ELSE HIT [RETURN]:

The results are

<table>
<thead>
<tr>
<th>ROWS/COLS</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>LABELS</td>
<td>col #1</td>
<td>col #2</td>
<td>col #3</td>
<td>col #4</td>
<td>col #5</td>
<td>col #6</td>
</tr>
<tr>
<td>1 row #1</td>
<td>2</td>
<td>4</td>
<td>6</td>
<td>8</td>
<td>...</td>
<td>12</td>
</tr>
<tr>
<td>2 row #2</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>...</td>
</tr>
<tr>
<td>3</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>4 1+2</td>
<td>2</td>
<td>5</td>
<td>8</td>
<td>11</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>5 1-2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>6</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>7 expr</td>
<td>-1</td>
<td>1,500</td>
<td>44</td>
<td>216,500</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>8 log</td>
<td>2.2628E-08</td>
<td>0.693</td>
<td>1.099</td>
<td>1.386</td>
<td>...</td>
<td>1.792</td>
</tr>
<tr>
<td>9 int</td>
<td>2</td>
<td>3</td>
<td>6</td>
<td>7</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>10 AND</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>...</td>
<td>0</td>
</tr>
<tr>
<td>11 AND4expr</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>...</td>
</tr>
<tr>
<td>12 OR expr</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>13 OR expr</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>14</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

The results in row 12 depend on values in rows 1 and 2, some of which are missing. In column 5, the value in row 2 is greater than or equal to 1, so the result is 1 (i.e., true) regardless of the value in row 1. In column 6, the expression involving row
1 (#1<8) is false, so the result is undefined because it is either true or false depending on the (presently undefined) value in row 2. The results in columns 5 and 6 of row 13 are missing because neither constituent expression can be evaluated there.

**OPTION 4: SPECIAL CALCULATIONS**

To specify a special calculation, the user selects a single type of calculation from a menu, instead of typing an expression. The special calculation suboptions that apply to rows are listed below; an identical set of options applies to columns.

**SPECIAL CALCULATIONS OPTIONS**

<table>
<thead>
<tr>
<th>TYPE TO CALCULATE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. ROW WITH MISSING VALUES REPLACED BY A CONSTANT</td>
</tr>
<tr>
<td>2. ROW WITH SELECTED CONSTANT REPLACED BY MISSING VALUE</td>
</tr>
<tr>
<td>3. WITHIN ROW CUMULATIVE SUM</td>
</tr>
<tr>
<td>4. WITHIN ROW CUMULATIVE PRODUCT</td>
</tr>
</tbody>
</table>

When row calculations are performed, the suboptions apply to only a single row. As usual, a range of columns may also be specified. Whereas most row calculations combine values in corresponding columns, two types of special calculation (3 and 4) compute the cumulative sum or product of values within a row. The same two special calculations may use particular constants (0 or 1) in place of missing values.

To avoid making the following descriptions unduly complicated, we will state them only in terms of row calculations. However, if "column" is substituted for "row" and vice versa, they will apply to column calculations.

**Row with Missing Values Replaced by a Constant.** This special calculation replaces missing values with a specified constant. It is useful when missing values must be replaced by a number such as 0 or 1 to complete a calculation, or when, for purposes of display but not calculation, the user wants to replace them with a number that is shown as a character-string code such as unknown. The user specifies a source row and a result row (which might be the same row). The values in the source row are copied into the result row, with missing values replaced by the specified constant.

The user enters the result row number and label, the source row number, and the constant:

```
REPLACE MISSING VALUES WITH A CONSTANT IN A ROW
RESULT ROW: 1
LABEL FOR ROW #1(row #1): ... to 0
SOURCE ROW: 1
CONSTANT TO REPLACE MISSING VALUES: 0
```

The result is
Row with Selected Constant Replaced by Missing Values. This special calculation replaces occurrences of a specified constant with missing values. It is useful when missing values in a row have been replaced with a constant, and the user now wishes to convert the constant back to indicate missing values. It is also useful when logical expressions are used to create subsets of patients or item values. Logical expressions are used to store 1s in a row when a set of conditions are satisfied and to store 0s when the conditions are not satisfied. This special calculation can be used to convert the 0s to indicators of missing values. Then the data values that satisfy the conditions can be copied into a result row by multiplying their row by the row of 1s and missing values. The result row then contains only the desired values; other columns in the result row show missing values.

The user specifies a source row and a result row (which can be the same row). The values in the source row are copied into the result row with the specified constant replaced by a missing value indicator wherever it appears. The major portion of the dialog is

```
REPLACE CONSTANT WITH MISSING VALUE IN A ROW
RESULT ROW: 1
LABEL FOR ROW 1(... to 0): 0 to ...

SOURCE ROW: 1
CONSTANT TO BE REPLACED WITH MISSING VALUE: 0
```

The effect of this is shown at the end of the next subsection.

Within-Row Cumulative Sum. This special calculation is useful in computations that involve the sum of values i through j for several js. It computes the sum of values within a source row. The value stored in the nth column of the result row is equal to the sum of the values in the starting column through the nth column of the source row, i.e., it is the cumulative sum of source row values up to the corresponding column. Unless otherwise specified, the starting column is column 1 and the ending column is the last one that has a label or a non-missing value.

If a missing value is encountered in the source row, the cumulative sums stored in the corresponding column and in all subsequent columns are undefined, because a sum is undefined if one of its component values is missing. To avoid this possibility, the user may choose to use zeros in place of missing values and thereby have the system calculate the cumulative sum of non-missing values.

When using this special calculation, the user enters the result row number and label, enters the source row number, and responds to a prompt about the treatment of missing values. Although it is not illustrated here, the user may also change the starting and ending column numbers.
The cumulative sums appear in row 3. If zero were not used in place of missing values, the values in column 5 and subsequent columns of row 3 would be missing.

<table>
<thead>
<tr>
<th>ROWS/COLS</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>LABELS</td>
<td>col #1</td>
<td>col #2</td>
<td>col #3</td>
<td>col #4</td>
<td>col #5</td>
<td>col #6</td>
</tr>
<tr>
<td>1 8 to ...</td>
<td>2</td>
<td>4</td>
<td>6</td>
<td>8</td>
<td>...</td>
<td>12</td>
</tr>
<tr>
<td>2 row #2</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>32</td>
</tr>
<tr>
<td>3 cumsum 1</td>
<td>2</td>
<td>6</td>
<td>12</td>
<td>20</td>
<td>20</td>
<td>32</td>
</tr>
<tr>
<td>4 1/2</td>
<td>2</td>
<td>5</td>
<td>8</td>
<td>11</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>5 1/2</td>
<td>...</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>6</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

**Within-Row Cumulative Product.** This special calculation is useful in computations such as those of geometric means, which involve the product of values $i$ through $j$ for several $j$s. It is like the cumulative sum except that the cumulative product of values within a source row is calculated for each column between (and including) the starting and ending columns. The value stored in a column of the result row is the product of values in the source row from the starting column through that column.

Because a product is undefined if any of the multipliers is undefined, the cumulative products stored in a column of a result row and in all subsequent columns are undefined once a missing value is encountered in the source row. To avoid this, the user may choose to have the system use the number 1 in place of any missing value (since a zero would result in all succeeding values being zero). In that case, the system calculates the cumulative product of non-missing values.

As with other calculations, the user enters the result and source row numbers and the result row label. He also responds to a prompt about the treatment of missing values. He may also specify the starting and ending column numbers; otherwise, the starting column is column 1 and the ending column is the last one that is labeled or contains non-missing values.

The cumulative products are stored in row 6. If the number 1 were not used in place of missing values, all the values in row 6 would be missing because the first multiplier (in column 1 of row 5) is a missing value. One consequence of this is that the
1 in row 6, column 1, is an anomaly, rather than a correct result. However, the values in subsequent columns of row 6 are correct.

<table>
<thead>
<tr>
<th>ROWS/COLS</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>LABELS</td>
<td>col #1</td>
<td>col #2</td>
<td>col #3</td>
<td>col #4</td>
<td>col #5</td>
<td>col #6</td>
</tr>
<tr>
<td>1 0 to</td>
<td>2</td>
<td>4</td>
<td>6</td>
<td>8</td>
<td>...</td>
<td>12</td>
</tr>
<tr>
<td>2 row #2</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>...</td>
</tr>
<tr>
<td>3 cussum 1</td>
<td>2</td>
<td>6</td>
<td>12</td>
<td>20</td>
<td>20</td>
<td>32</td>
</tr>
<tr>
<td>4 1+2</td>
<td>2</td>
<td>5</td>
<td>8</td>
<td>11</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>5 1-2</td>
<td>...</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>6 cumprod5</td>
<td>1</td>
<td>3</td>
<td>12</td>
<td>60</td>
<td>60</td>
<td>60</td>
</tr>
</tbody>
</table>

**OPTION 5: TIME DIFFERENCE**

This subactivity calculates differences between values that represent time. The differences between two dates, between two times, or between two dates and times may be specified. The values may be stored in worksheet rows or columns or may be entered as constants. The results can be expressed in time units ranging from minutes to 365-day years. The notation #n is used to refer to row or column n when specifying the location of date or time data.

In the following examples, differences will be taken between rows; of course, the equivalent operations can be performed between columns. The worksheet, TIMEDIFF, will illustrate the three functions of calculate, option 5.

**WORKSHEET TIMEDIFF**

**TITLE** Illustrate Time Difference Calculations

**CREATED** 3/3/77  **MODIFIED** 3/3/77

| # OF ROWS | 10 |
| # OF COLS | 6  |

<table>
<thead>
<tr>
<th>ROWS/COLS</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 time 1</td>
<td>1</td>
<td>83</td>
<td>1215</td>
<td>1645</td>
<td>1630</td>
<td>2350</td>
</tr>
<tr>
<td>4 time 2</td>
<td>715</td>
<td>1420</td>
<td>...</td>
<td>1215</td>
<td>1630</td>
<td>215</td>
</tr>
<tr>
<td>5</td>
<td>...</td>
<td>4</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>6</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>7</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>8</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>9</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>10</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

The top two rows are dates, and rows 3 and 4 are times. These examples also indicate that dates must be of the form mm/dd/yy (or mm/dd/yyyy) and that times must be expressed in 24-hour time with no colons or abbreviations (i.e., 1300 instead of 1:00 P.M.). Either 2400 or 0 (with the date for the next day) may be used to represent midnight. Once row (or column) calculations have been chosen, the following suboptions are available:
In the first example, the user has chosen to specify dates and times, a result row (5) was labeled result1, and the row numbers of the dates and times were entered in response to prompts, with earlier times first. Note that a constant date or time can be entered instead of the number of a particular row on which operations are to be performed.

Once the operation is defined, the option of selecting another range of columns is presented. The time units can then be selected. In this example, the results will be expressed in years:

The results of these three examples will be shown below. The next example illustrates the use of difference between dates. Row 6 is labeled result2, and row 1 will be subtracted from row 2. In this example, days are selected as the result units.
ROW 6(result1 )=DATE1 TO DATE2
DATE1=#1 DATE2=#2
THIS CALCULATION WILL BE PERFORMED FOR COLUMNS 1 TO 6
TYPE yes TO SELECT ANOTHER RANGE OF COLUMNS, ELSE HIT [RETURN]:

SELECT UNITS FOR RESULTS

<table>
<thead>
<tr>
<th>TYPE TO EXPRESS RESULTS IN</th>
</tr>
</thead>
<tbody>
<tr>
<td>1..MINUTES</td>
</tr>
<tr>
<td>2..HOURS</td>
</tr>
<tr>
<td>3..DAYS</td>
</tr>
<tr>
<td>4..WEEKS</td>
</tr>
<tr>
<td>5..30-DAY INTERVALS</td>
</tr>
<tr>
<td>6..365-DAY INTERVALS</td>
</tr>
</tbody>
</table>

UNITS CHOICE: 3

In the final example, times only are to be expressed. Row 7 is labeled result3, and row 3 is subtracted from row 4. Results will be expressed in minutes.

ROW 7(result3 )=TIME1 TO TIME2
TIME1=#3 TIME2=#4
THIS CALCULATION WILL BE PERFORMED FOR COLUMNS 1 TO 6
TYPE yes TO SELECT ANOTHER RANGE OF COLUMNS, ELSE HIT [RETURN]:

SELECT UNITS FOR RESULTS

<table>
<thead>
<tr>
<th>TYPE TO EXPRESS RESULTS IN</th>
</tr>
</thead>
<tbody>
<tr>
<td>1..MINUTES</td>
</tr>
<tr>
<td>2..HOURS</td>
</tr>
<tr>
<td>3..DAYS</td>
</tr>
<tr>
<td>4..WEEKS</td>
</tr>
<tr>
<td>5..30-DAY INTERVALS</td>
</tr>
<tr>
<td>6..365-DAY INTERVALS</td>
</tr>
</tbody>
</table>

UNITS CHOICE: 1

The completed worksheet is shown below. Row 5 shows the difference in hours between dates and times (date 1, time 1 subtracted from date 2, time 2); row 6 shows the difference in days between date 1 and date 2; row 7 shows the difference in minutes between time 1 and time 2.

ROWS/COLS  1   2   3   4   5   6
LABELS

| 3 time 1 | 1 | 830 | 1215 | 1645 | 1830 | 2350 |
| 4 time 2 | 715 | 1420 | ... | 1215 | 1805 | 215 |
| 5 result1 | 280.138 | 77.655 | ... | ... | 31.250 | .118 |
| 6 result1 | 730.50 | 28344 | 19826 | ... | 11486 | 44 |
| 7 result3 | 434 | 350 | ... | -270 | 455 | -1295 |
| 8 | ... | ... | ... | ... | ... | ... |
| 9 | ... | ... | ... | ... | ... | ... |
| 10 | ... | ... | ... | ... | ... | ... |
OPTION 6: STATISTICAL CALCULATIONS

Below are the five suboptions available for performing statistical calculations on rows. A statistic is computed for the data in each specified column and the result is stored in the result row. Equivalent operations are available for columns. When this subactivity is used, row n may be designated by typing either #n or n.

<table>
<thead>
<tr>
<th>STATISTICAL CALCULATIONS INVOLVING ROWS</th>
</tr>
</thead>
<tbody>
<tr>
<td>STATISTICAL CALCULATIONS OPTIONS</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>1..ROW OF COUNTS OF NON-MISSING COLUMN VALUES</td>
</tr>
<tr>
<td>2..ROW OF SUMS OF COLUMN VALUES</td>
</tr>
<tr>
<td>3..ROW OF MEANS OF COLUMN VALUES</td>
</tr>
<tr>
<td>4..ROW OF STANDARD DEVIATIONS OF COLUMN VALUES</td>
</tr>
<tr>
<td>5..ROW OF STANDARD ERRORS OF COLUMN VALUES</td>
</tr>
</tbody>
</table>

OPTION: 3

The data used in the following examples and all of the results are shown at the end of this section.

**Row of Counts of Non-Missing Column Values.** The first step in using this type of statistical calculation is to specify and label the result row. Then, to define the calculation, the user either specifies a range of rows to be used or indicates that all rows (i.e., those that have labels or non-missing values) should be used. For each column, the system counts the number of non-missing values within the specified range of rows and stores the count in the corresponding column of the result row. Although this type of calculation is used when the values of interest are stored in columns, it is considered a row calculation because it counts across rows and because the results are stored in a row.

The user specifies:

```
ROW #14 OF COUNTS OF NON-MISSING COLUMN VALUES
TYPE yes TO SEE ROW LABELS, ELSE HIT [RETURN]:
RESULT ROW: 14
LABEL FOR ROW #14( )=COUNT THIS CALCULATION WILL BE PERFORMED FOR ROWS 1 TO 13
TYPE yes TO SELECT ANOTHER RANGE OF ROWS, ELSE HIT [RETURN]: yes
STARTING ROW: 1
ENDING ROW: 7
```

and the system responds:

```
ROW #14(COUNT )=COUNT OF NON-MISSING COLUMN VALUES CONTAINED IN ROW #1 . . . ROW #7
STARTING COLUMN: #1
ENDING COLUMN: #10
CALCULATION COMPLETED
TYPE yes TO DISPLAY THE RESULTS, ELSE HIT [RETURN]:
```

Results will be shown for all statistical calculations at the end of this subsection.
Row of Sums of Column Values. Here again the user specifies a range of rows or indicates that all the rows that have labels or non-missing values should be used. For each column the system computes the sum of the values stored in the specified range of rows and stores the sum in the corresponding column of the result row. For example, if the result row is 15 and the starting and ending rows are 3 and 8, this specification is equivalent to

ROW #15 = #3 + #4 + #5 + #6 + #7 + #8

It is, therefore, shorthand for a calculation that could be specified using an arithmetic expression. An important difference is that here the value 0 may be used whenever a missing value is encountered. This is important, because the user would not always want the sum to be missing whenever a constituent value was missing. It is important to specify the row range carefully to exclude, for example, the row in which counts are stored.

To compute the sum of the values counted above, the user specifies:

ROW OF SUMS OF COLUMN VALUES
TYPE yes TO SEE ROW LABELS, ELSE HIT [RETURN]:
RESULT ROW: 15
LABEL FOR ROW #15:
THIS CALCULATION WILL BE PERFORMED FOR ROWS 1 TO 15
TYPE yes TO SELECT ANOTHER RANGE OF ROWS, ELSE HIT [RETURN]: yes
STARTING ROW: 3
ENDING ROW: 8

In some cases, it is meaningful to use 0 in place of non-missing values in order to compute the sum of the non-missing values. In other cases, this would not be appropriate, and the sum should be undefined if missing values are involved. The system responds:

ROW #15(SUM 3to8)=ROW #3 + . . . + ROW #8

STARTING COLUMN: #1
ENDING COLUMN: #10
TYPE yes TO USE THE CONSTANT 0 FOR MISSING VALUES, ELSE HIT [RETURN]:
(A RESULT COMPUTED USING A MISSING VALUE IS MISSING UNLESS THE CONSTANT IS USED.): yes

Rows of Means, Standard Deviations, or Standard Errors of Non-Missing Values. All of these calculations are performed in the same manner by the user. A result row is specified and labeled. The range of rows upon which the calculations are to be performed can be limited. In this example, calculations will be performed on rows 1 through 13.

Means, standard deviations, and standard errors are computed, ignoring all missing values. The following examples illustrate the procedure for calculating means and storing them in a row which is already labeled.
The following two examples illustrate the use of option 6, Statistical Calculations, for standard deviations and standard errors. Row 10 is already labeled S.D., and row 11 is already labeled S.E.

The counts, sums, means, standard deviations, and standard errors are now stored in the worksheet, as shown below:
## CALCULATE, 6

4 / 1 /77, at 14:56

**WORKSHEET GLUCOSE**

**TITLE Control Glucose Values with Statistics**

**CREATED 3/3/77**  **MODIFIED 4/1/77**

<table>
<thead>
<tr>
<th>ROWS/COLS</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>LABELS</strong></td>
<td>HR -2</td>
<td>HR -1</td>
<td>HR 0</td>
<td>HR 1</td>
<td>HR 2</td>
<td>HR</td>
</tr>
<tr>
<td>1 case 1</td>
<td>103.500</td>
<td>97.800</td>
<td>96</td>
<td>150.100</td>
<td>202.300</td>
<td>...</td>
</tr>
<tr>
<td>2 case 2</td>
<td>90.000</td>
<td>84.500</td>
<td>83.800</td>
<td>...</td>
<td>185.000</td>
<td>...</td>
</tr>
<tr>
<td>3 case 3</td>
<td>101.300</td>
<td>95.400</td>
<td>92.800</td>
<td>147.900</td>
<td>198.800</td>
<td>200</td>
</tr>
<tr>
<td>4 case 4</td>
<td>103.400</td>
<td>99.750</td>
<td>97.186</td>
<td>152.383</td>
<td>201.457</td>
<td>200</td>
</tr>
<tr>
<td>5 case 5</td>
<td>103.400</td>
<td>99.750</td>
<td>97.186</td>
<td>152.383</td>
<td>201.457</td>
<td>...</td>
</tr>
<tr>
<td>6 case 6</td>
<td>103.400</td>
<td>99.750</td>
<td>97.186</td>
<td>152.383</td>
<td>201.457</td>
<td>...</td>
</tr>
<tr>
<td>7 case 7</td>
<td>100.700</td>
<td>...</td>
<td>153.800</td>
<td>...</td>
<td>205.400</td>
<td>...</td>
</tr>
<tr>
<td>8</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>9 means</td>
<td>100.928</td>
<td>96.158</td>
<td>102.565</td>
<td>152.030</td>
<td>199.524</td>
<td>200</td>
</tr>
<tr>
<td>10 S.D.</td>
<td>4.615</td>
<td>5.965</td>
<td>23.183</td>
<td>3.531</td>
<td>6.358</td>
<td>...</td>
</tr>
<tr>
<td>11 S.E.</td>
<td>1.744</td>
<td>2.435</td>
<td>8.732</td>
<td>1.579</td>
<td>2.403</td>
<td>...</td>
</tr>
<tr>
<td>12</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>13</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>14 COUNT</td>
<td>7</td>
<td>6</td>
<td>7</td>
<td>5</td>
<td>7</td>
<td>2</td>
</tr>
<tr>
<td>15 SUM 8to8</td>
<td>512.200</td>
<td>394.650</td>
<td>538.157</td>
<td>610.049</td>
<td>1008.57</td>
<td>400</td>
</tr>
</tbody>
</table>
Option 1 of describe is a specialized text editor and compiler that allows the user to construct, edit, and review the schema for a particular study (see Sec. II for a discussion of the schema). Option 2 provides for printing the schema in an expanded format on the system printer. A compact listing of the schema, containing the names and descriptions of all panels, items, and events, may be generated by entering `ides,1` and then entering the command `print`. Option 3 allows the user to create and delete study files (and lock and unlock the schema) for the purpose of testing his schema design.

**Schema Elements**

The schema comprises a number of lines of text, each of which describes a panel, item, or event. A line is constructed by entering its line number and then the appropriate descriptive elements, all in response to a series of system prompts in option 1.

Recall that items are organized into panels, i.e., data values considered to be measured at the same time (e.g., vital signs or the results of a set of measurements on a single blood sample). A panel is defined by a descriptive panel line, followed by a number of lines that define its constituent items. Each item, i.e., variable, to be collected in a study is defined in turn by a line that names it and that specifies its attributes.

Events act as time markers and permit alignment of data values (by use of the retrieve activity) relative to some significant occurrence specified by the user. Retrieve always recognizes the events `begin` (associated with the earliest time and date of data collection for a patient) and `end` (associated with the last, or most recent, time and date of data collection for a patient). Other events may be defined by the user and each takes on the time of a specific occurrence of an item, namely, the time that the first or last value was measured, the time that the first or last specific value was measured, or the time that the maximum or minimum value occurred.

**Schema Design**

Much information relevant to schema design is contained in Secs. I, II, and III and in the Reference Manual descriptions of data entry (`enter`) and data retrieval (`retrieve`). The data entry and especially the retrieval operations must be kept in mind in the schema design phase. Some general guidelines are suggested below:

- Prepare an initial draft of the schema and make paper copies as schema construction proceeds. Make corrections on paper before entering them into the system. Once entered, a schema can be edited, but only a line at a time; that is, a line must be deleted and reentered to make a change.
- Compile the schema frequently during construction; this will locate errors while they can be easily corrected.
The data type of an item must match the data type of the panel in which it is included; a panel defined as "numeric" can contain items of any data type except free-form text.

Use easy-to-remember item and panel names. You will need to remember them (or look them up) frequently as you enter and retrieve data.

Include a "method" item in each panel of laboratory results where the analytic method is likely to be modified during the course of your study or where you know that more than one method will be used.

Keep the character strings for the coded values short, consistent, and mnemonic (remember, upper and lower case are distinguished), since they are the form in which the data will be entered and displayed. If you use y and n for yes and no in one item, don't switch order or change to YES and NO or Y and N in another item.

In coded lists, don't forget to include at least one "unknown" and/or "other" category, and try to use the same character strings in each item in which they appear. You might want a "not yet available" response as well for some items.

Don't include in a panel items that are not measured at the same time or whose results are not available at the same time. In the first case, you would not be able to time-stamp the items separately, and in the second, you would need to go back to an existing panel and complete the entry of the previously unavailable items.

Keep the number of items in a panel below about 15; more items make retrieval by panel awkward.

Distribute copies of the compiled schema listing, obtained using describe, option 2, to all those who will be working on your protocol and encourage uniformity in the reporting of item names and codes.

When preparing data collection forms, use the same names and codes that you have used in the schema. For some purposes, you may be able to use the formatted schema itself as a data collection form.

Use the range-checking feature with all continuous numerical data—and use reasonable ranges. Too narrow a range will be annoying to the data enterer; too wide a range won't catch errors.

Because the schema controls data entry, it should be designed to make data entry efficient; e.g., items should be listed in the same order as they appear on data collection forms or lab slips. This can best be accomplished by describing a schema, practicing data entry with it, and revising the schema until data entry goes smoothly. Response Files (see the discussion of response files) should be used to simplify data entry and to change the data entry order from the order of items in the schema.

Schema Construction

Schema construction and modification are critical procedures that affect the entire course of the study; they should be done in consultation with the system manager. Generally, several iterations will be made in the preparation of the schema. To prevent users from inadvertently changing a schema (and thereby changing the meaning of previously stored data) after study data have been collect-
ed, it must be "locked" prior to the entry of the first patient and must be unlocked for modification.

It is strongly recommended that the user perform trial data entry and retrieval prior to finalizing the schema and entering extensive amounts of data. Describe, option 3, allows the user to lock the schema and then unlock it after entering a modest amount of data. Trying out a completed or partially completed schema by entering a small but representative amount of data and then trying typical retrievals is sure to uncover problems which may be corrected with little wasted effort. Although the user may lock the final version of his schema, it is recommended that this be done in collaboration with the system manager. Limited changes to the schema may be made after extensive data entry, but only by the system manager.

Items are associated with panels by adding item lines following the panel line. Events may be inserted anywhere in a schema; they are associated with an item by entering that item's name as a parameter. When a schema is being constructed, there may be up to three versions of it in the system (see the flow diagram included with the description of option 1 that illustrates the effects of responses to the command prompt);

1. The saved text version, which the user may only list (by using the list or print command in option 1 or by using option 2).
2. A copy of the saved text version, which the user may list and edit and then save, thus replacing the saved version.
3. The internal "compiled" version, which the system generates by operating on the saved text version when requested to do so by the user.

At any stage in the construction of the schema, the saved text version may be compiled. The compilation associates panels and items, assigns internal panel and item numbers, checks item and panel names for duplicates, and checks item attributes for consistency with the associated panel. Errors are reported to the user and must be corrected before compilation is completed. Following compilation, but preceding the entry of study data, panel and item lines may be added to or deleted from any position in the text version of the schema, changes may be made to any line (by deleting and reentering the line), and the schema then may be recompiled.

The following table specifies the parameters of the schema:

<table>
<thead>
<tr>
<th></th>
<th>Minimum Number</th>
<th>Maximum Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Panels</td>
<td>1</td>
<td>56</td>
</tr>
<tr>
<td>Items</td>
<td>1</td>
<td>392</td>
</tr>
<tr>
<td>Codes/item</td>
<td>1</td>
<td>30</td>
</tr>
<tr>
<td>Items/panel</td>
<td>1</td>
<td>30</td>
</tr>
<tr>
<td>Events</td>
<td>0</td>
<td>28</td>
</tr>
<tr>
<td>Lines of text</td>
<td>2</td>
<td>500</td>
</tr>
</tbody>
</table>

Dialog

When the user enters the describe activity, the following menu is displayed:
To create, edit, or list the current schema, the user enters 1 and proceeds as shown below. Option 2 will produce an expanded version of the schema on the system printer which is easy to read but may be quite lengthy. Option 3 provides for locking and unlocking the schema by creating and deleting study files.

**OPTION 1: EDIT THE SCHEMA**

The following is an example of the text version of a partially completed schema which will be referenced in the following paragraphs. This schema was constructed by responding to a series of prompts, illustrated below.

### SCHEMA for hypertens on 1/13/78 at 15:31

<table>
<thead>
<tr>
<th>PANEL</th>
<th>ITEM</th>
<th>FIELD</th>
<th>TYPE</th>
<th>DESCRIPTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>10.0</td>
<td>1</td>
<td>name</td>
<td>text</td>
<td>Patient name</td>
</tr>
<tr>
<td>20.0</td>
<td>1</td>
<td>phone</td>
<td>text</td>
<td>Patient phone</td>
</tr>
<tr>
<td>30.0</td>
<td>2</td>
<td>address</td>
<td>text</td>
<td>Patient address</td>
</tr>
<tr>
<td>40.0</td>
<td>3</td>
<td>name</td>
<td>text</td>
<td>Patient name</td>
</tr>
<tr>
<td>50.0</td>
<td>2</td>
<td>Phys</td>
<td>numeric</td>
<td>Physical data</td>
</tr>
<tr>
<td>60.0</td>
<td>5</td>
<td>age</td>
<td>char.code(3)</td>
<td>Age at last birthday</td>
</tr>
<tr>
<td>70.0</td>
<td>6</td>
<td>date</td>
<td>date</td>
<td>Admission date</td>
</tr>
<tr>
<td>80.0</td>
<td>7</td>
<td>weight</td>
<td>num.range</td>
<td>Weight at admission</td>
</tr>
<tr>
<td>90.0</td>
<td>8</td>
<td>control</td>
<td>char.code(3)</td>
<td>Control group code</td>
</tr>
<tr>
<td>100.0</td>
<td>9</td>
<td>vital</td>
<td>vital.sign</td>
<td>Vital signs</td>
</tr>
<tr>
<td>110.0</td>
<td>10</td>
<td>systolic.pressure</td>
<td>torr</td>
<td>Systolic pressure</td>
</tr>
<tr>
<td>120.0</td>
<td>11</td>
<td>diastolic.pressure</td>
<td>torr</td>
<td>Diastolic pressure</td>
</tr>
<tr>
<td>130.0</td>
<td>12</td>
<td>heart.rate</td>
<td>心率</td>
<td>Heart rate</td>
</tr>
<tr>
<td>140.0</td>
<td>13</td>
<td>core.temp</td>
<td>temporal</td>
<td>Core temperature oral</td>
</tr>
<tr>
<td>150.0</td>
<td>14</td>
<td>urine</td>
<td>urine.measure</td>
<td>24 hour urine measure</td>
</tr>
<tr>
<td>160.0</td>
<td>15</td>
<td>potassium</td>
<td>serum level</td>
<td>Potassium serum level</td>
</tr>
<tr>
<td>170.0</td>
<td>16</td>
<td>blood.salt</td>
<td>serum level</td>
<td>Blood salt serum level</td>
</tr>
<tr>
<td>180.0</td>
<td>17</td>
<td>glucose</td>
<td>urine glucose</td>
<td>Glucose urine glucose</td>
</tr>
</tbody>
</table>

### OPTIONS

- **1. Edit the Schema**
- **2. Print a Formatted Schema**
- **3. Create/Delete Study Files**
A listing of this same schema in its formatted version is included in the description of option 2. Each line in the text version of the schema contains two common elements:

- A line number (of the form xxx.x). This is the number by which the line is referenced during schema construction and editing.
- A line type (PANEL, ITEM, EVENT, or “blank item”). Blank items, e.g., line 4.0, serve as place holders and will be explained more fully below.

PANEL, ITEM, and EVENT lines contain, in addition, a series of phrases separated by commas. The meaning of these phrases is determined by the line type and the position of the phrase in the line. The last phrase belonging to an EVENT, ITEM, or PANEL is followed by a semicolon. Lines with non-integer numbers, such as 6.1 and 10.1 above, are termed continuation lines and are provided by describe when there is not sufficient space on a single line for the required phrases. They cannot be referenced directly by the user, but referencing the continued line references its continuation also.

When the schema is listed again after compilation, the panel, item, and event numbers appear (e.g., the 1 in line 2.0 and the 6 in line 14.0). If new items or panels are inserted prior to compilation, all the succeeding items or panels will be renumbered. The compilation process assigns panel numbers to consecutive panels and item numbers to consecutive items (across panels). These numbers are useful for keeping track of the number of panels and items during schema construction and for reference after the schema is completed.

Levels of Prompts

The schema is constructed and modified by responding to a series of system prompts. There is a hierarchical relationship among the prompts, with four main levels. (Not all levels will be invoked in every series of prompts. The details of the responses will be explained later.)

1. COMMAND: Responses permit construction, manipulation, and listing of the schema.
2. ENTRY TYPE: Responses determine whether a panel, item, or event is to be described.
3. DATA TYPE: Responses specify whether data are numeric, textual, etc.
4. SCREEN TYPE: Responses specify the type and the details of data screening procedures.

The figure below shows all of the prompts that are encountered in describe, option 1, and all of the fixed-choice responses. The prompts are shown in capital letters followed by a colon (e.g., COMMAND:, DESCRIPTION:), while the fixed responses are shown in lower case, branching from the prompt (e.g., delete #, max, numeric). Most prompts require a response; those that can be answered with the RETURN key are followed by an asterisk (*). Responses to some prompts, like DESCRIPTION:, are up to the user but are limited to a maximum number of characters. The end of any sequence returns the user to the COMMAND: prompt. This is indicated by an arrow pointing to (C). Any prompt may be answered with !; this will also return to the COMMAND: prompt but will discard responses so far entered in the current sequence.
System prompts and user responses in DESCRIBE

**COMMAND:**
- load
- list #
- delete #
- add
- renumber
- print
- compile
- save

Responses to COMMAND:
- REPLACE COPY WITH ORIGINAL: (yes, no)

**ENTRY TYPE:**
- ABBREV: (8 char)
- EVENT ABBREV: (8 char)
- ITEM ABBREV: (8 char)
- DESCRIPTION: (20 char)
- DESCRIPTION: (20 char)
- DESCRIPTION: (20 char)
- PANEL TYPE:
- CONTROLLING ITEM ABBREV: (8 char)
- OCCURRENCE:
  - numeric
  - textual
  - first
  - last
  - max
  - min

TRIGGERING VALUE: *

**DATA TYPE:**
- time
- date
- numeric

Responses to DATA TYPE:
- SCREEN TYPE: *
  - range
    - [RETURN]
  - check
    - [RETURN]
- SCREEN TYPE: *
  - char
  - text
  - [RETURN]

LOWER BOUND: (n)
UPPER BOUND: (n)
HOW MANY: (1-30)
CHECK 0: (n)
CHECK n-1: (n)

HOW MANY CODES: (1-30)
CODE 0: (8 char)
CODE n-1: (8 char)

TEXT LENGTH: (1-70)

CONFIDENTIAL

Note: Entering ' ' in response to any prompt returns to the COMMAND: prompt.

*Response other than [RETURN] not required.
At the first level, the system prompts with `COMMAND:`; followed by a list of allowable responses, described below. Depending on the response, the system will either perform the required function or will issue a series of second-level prompts as described in the next section.

Responses to the `COMMAND:` Prompt

- **load.** This command discards the existing working copy of the schema and makes a new one from the saved text version. It is used when the user wants to abandon the existing working copy because it contains errors.
- **save.** This command replaces the saved text version of the schema with the working copy and leaves the copy intact.
- **list #** (where `#` is an integer). This command lists up to 25 lines of the working copy of the schema, starting with a specified line, "#". If a line number is not specified, the first 25 lines of the schema are listed. The command `list last` will cause the last 25 lines of the schema to be displayed.
- **delete #** (where `#` is an integer). This command deletes the schema line whose number corresponds to "#" from the working copy of the schema, as well as continuation lines numbered #.1, etc. Note: No change occurs in the display of the lines until another `list` command is issued.
- **add #** (where `#` is an integer). This command adds a line numbered "#" to the working copy, where "#" is an integer not already used. The system then prompts for `ENTRY TYPE:` (e.g., panel, item, or event) and issues second-level prompts for the details appropriate to that type as described above under Schema Elements. Note: No change occurs in the display of the lines until another `list` command is issued. If the phrases for a line do not fit within 64 characters, the system provides additional lines numbered #.1, #.2, etc.
- **renumber.** If additional lines are to be inserted between two lines that have no integer between them (e.g., lines 7 and 8), the command `renumber` is used to renumber all lines of the working copy with even numbers prior to inserting the new line. It has no effect on the display. To see the renumbered lines, the user must issue another `list` command.
- **print.** This command causes the entire current working copy of the schema to be printed on the system printer in compact format. The schema may be printed in expanded format using option 2. The current screen contents may be printed on the attached DTC printer at any time by pressing the PRINT key.
- **compile.** This command causes the saved text version of the schema, not the working copy, to be processed into the compiled schema in a form suitable for use by other CLINFO activities (e.g., `enter`). The saved text version is modified to include panel, item, and event numbers. Before compiling, the system asks if the working copy is to become the saved text version (e.g., if changes have been made in the working copy since the last save command).
- **lactivity-name.** When the user has finished his editing activity and has had the opportunity to save the working copy of the schema, he can go on to another CLINFO activity by this standard means.
The figure below illustrates the relationships among the three versions of the schema and the effects of the various responses to the COMMAND: prompt:

**Responses to Panel, Item, and Event Prompts**

When the user adds a new line to the schema, he is prompted for ENTRY TYPE, as described above. Depending on the response, an appropriate sequence of prompts will obtain the information necessary to specify a panel, item, or event. An explanation of the prompts follows. The permissible responses (shown in parentheses) may be user-defined character strings, numbers, or one of a set of fixed responses. The latter are shown in italics.

**Panel Prompts**

- Panel Abbreviation (at most 8 characters): An abbreviated name (e.g., hist in line 2.0 above), unique within a schema.
- Panel Description (at most 20 characters): A textual description of the panel (e.g., physical data in line 18.0).
- Panel Type (*numeric* or *text*). The type of panel, determined by whether it comprises all numeric items or all textual items (i.e., free text). Items belonging to a numeric panel have values that are stored as numbers. They may be numeric or encoded data, times, or dates.

**Item Prompts**

Note: If the item is a "blank item" the following phrases are not prompted for:
• Item Abbreviation (at most 8 characters): An abbreviation to be used for reference and labeling (e.g., AGE in line 10.0 above), unique within a schema.

• Item Description (at most 20 characters): A textual description of the item (e.g., Height in line 22.0).

• Item Units (at most 10 characters): A description of the units of measurement for the item. These units (e.g., years in line 10.0) are purely descriptive and do not enter into any computations. However, the units should be chosen so that data values lie in the range \(-9999.99\) to \(+9999.99\) to avoid later displaying data in scientific notation. Note that the text items in the above schema contain an asterisk in their unit positions; the asterisk indicates "not applicable."

• Item Data Type (\textit{num}, \textit{char}, \textit{time}, \textit{date}, or \textit{text}:) An indicator that specifies the type of data to be entered for the data item; each is explained below.

\textit{num} (numeric): The datum is zero or any positive or negative real number between \(5.4E-79\) and \(7.2E75\) in magnitude.

\textit{char} (encoded character string): The datum is one of a list of up to 30 character strings (e.g., male, female; yes, no, unknown, etc.) which will be coded and stored as an integer from 0 to 29 (see Item Screening Procedures below).

\textit{time}: The datum is a 24-hour time entered in the form hhmm (e.g., 1315).

\textit{date}: The datum is a date entered as mm/dd/yy (e.g., 3/17/75) or as mm/dd/yyyy (e.g., 9/5/1893).

\textit{text} (free text): The datum is up to 70 characters of text, including blanks (e.g., Previous history of high plasma bananase).

• Item Screening Type (\textit{none}, \textit{range}, \textit{check}, or \textit{code}): An indicator that specifies the type of screening procedure to be performed on the data item values during data entry.

• Item Screening Procedures: The specification of the details of the screening procedure to be carried out during data entry (with the \textit{enter} activity) for the indicated screening type. These details will be prompted for, depending upon the item screening type chosen for the items.

The following table indicates the permissible screening procedures that may be applied to the five item data types.
<table>
<thead>
<tr>
<th>Item Data Type</th>
<th>Screening Procedure Type</th>
<th>Range</th>
<th>Check</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>num</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>char</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>time</td>
<td>x</td>
<td></td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>date</td>
<td>x</td>
<td></td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>text</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: The only screening performed on text Item Data Types is for specified string length.

The information that will be prompted for and the resultant screening procedures for each item screening type are shown in the Dialog section below. If the user hits the RETURN key in response to the request for a screening procedure, only type checking will be performed.

- Item Confidentiality (yes or no): An indicator formerly used to prevent the value of a sufficiently sensitive item, such as the patient name, from appearing in the version of the instrumentation data that was sent to the computer science contractor as part of the CLINFO experiment. The indicator still functions, but instrumentation data are not being collected.

**Event Prompts**

- Event Name (at most 8 characters): A short name used to reference the event.
- Event Description (at most 20 characters): A short description of the event.
- Controlling Item Abbreviation (at most 8 characters): The name of the data item on whose value the time assigned to a given event depends.
- Occurrence (first, last, max, or min): Since many values may be recorded for an item, one value of the item must be picked to represent the event. This is done by specifying one of the modifiers first, last, max, min. For example, we may refer to the maximum DBP of the study and name the event MDP, or to the first DBP and name the event DBP1. If first or last are chosen, an additional specification is optional (see below).
- Triggering Value (a numeric value or a character string): A value that may be specified by the user if he wishes the event to take on the time of a particular value (other than the max or min) of the controlling item. The value must be a number (e.g., 125) or one of the user-defined character strings (e.g., yes). Times or dates may not be used as triggering values for events.

**Dialog**

Schema creation and editing are done on a copy of the saved text version of the schema (even if the "copy" is blank, as it is initially). This allows any editing session to be aborted without affecting the permanent schema. It has the disadvantage of
requiring the user to replace the permanent text version of the schema with the copy at the end of a successful editing session, by use of the save command (otherwise the copy is lost when the user goes on to some other activity). The system protects the user by asking if the working copy is to be saved before it will proceed to another activity.

If the schema has been locked, the user is so informed upon choosing describe, option 1, and may use only the list or print commands. Otherwise, he may proceed to construct or modify the schema by adding a line number in response to COMMAND: (by typing add #), which results in a series of prompts starting with

ENTRY TYPE; panel, event, item, blank item or [RETURN]:

The prompts resulting from each possible response are illustrated below. (See Responses to Panel, Item, and Event Prompts above for more details on the meanings of each option.)

Incorrect responses are rejected and the prompt is reissued. The prompting sequence may be broken by responding ! to any prompt. The system then discards all information about the particular line and prompts again with COMMAND: ...

Entering a Panel. The following dialog illustrates the entry of information necessary to define a panel. Each line comprises the system prompt (up to the last colon) and the user’s response.

COMMAND:load,save,print,renumber,compile,list #,delete #,add #:add 240
ENTRY TYPE; panel, event, item, blank item, OR [RETURN]: panel
ABBREV: lab——
DESCRIPTION: Laboratory data——
PANEL TYPE; numeric OR text: num

COMMAND:load,save,print,renumber,compile,list #,delete #,add #:list 240

(Screen clears)

.

.

.

240.0 PANEL   lab   ,Laboratory data   ,numeric;

The user has

- Chosen to enter a new line, 240.
- Designated it as a panel.
- Named the panel lab.
- Described the panel as Laboratory data.
- Specified the panel as numeric (the first three letters are sufficient).
- Listed the line thus created.

Note that when all appropriate second-level responses to the ENTRY TYPE prompt have been made, describe issues the COMMAND: prompt again.
Entering an Item. The following dialog illustrates the entry of information necessary to define an item:

COMMAND: load, save, print, renumber, compile, list #, delete #, add #: add 250
ENTRY TYPE: panel, event, item, blank item, OR [RETURN]: item
ITEM ABBREV: UNA——
DESCRIPTION: Urine Na——
UNITS: meq/l
DATA TYPE: time, date, num, char, OR text: num
SCREEN TYPE: range, check, OR [RETURN]: range
LOWER Bound: 5
UPPER Bound: 300
CONFIDENTIAL: yes, ELSE [RETURN]:

COMMAND: load, save, print, renumber, compile, list #, delete #, add #: list 240

(Screen clears)

240.0 PANEL  lab, laboratory data, numeric;
250.0 ITEM  UNA, Urine Na, meq/l, num, range, (5, 300),
250.1  no;

COMMAND: load, save, print, renumber, compile, list #, delete #, add #:

The user has
- Chosen to enter a new line, 250.
- Designated it as an item.
- Entered the item abbreviation, UNA, and given it a description.
- Specified the units.
- Specified that the item is numeric.
- Declared the screening type to be range and entered the appropriate limits.
- Indicated (by not typing yes) that the item is not to be flagged as confidential.
- Listed lines 240 and following.

Note that line 250.1 has been added by the system to accommodate the portion of the item line that would exceed 80 characters.

The example below illustrates the entry of a schema line for a coded item.

COMMAND: load, save, print, renumber, compile, list #, delete #, add #: add 247
ENTRY TYPE: panel, event, item, blank item, OR [RETURN]: item
ITEM ABBREV: PROT——
DESCRIPTION: Protenuria——
As in the previous example, the user has entered a line number, designated it as an item, given it the abbreviation PROT, described it as Protenuria, hit the RETURN key in response to the UNITS: prompt, then specified the item to be of type char.

The system then assumes that the data screening type is code and proceeds to prompt the user appropriately. The user

- Enters the number of codes (4).
- Enters the character strings to be represented by each of the codes zero through 3.
- Declares, by pressing the RETURN key, that the item is not confidential.

Listing line 240 clears the screen and displays the schema with the newly created line:

240.0 PANEL    lab ,Laboratory data ,numeric;
247.0 ITEM     PROT ,Protenuria * ,char,code (4)
247.1          (prot neg,1+ ,2+ ,3+ ),no;
250.0 ITEM     UNA ,Urine Na ,meq/l ,num ,range, (5,300),
250.1          no ;

Entering a Blank Item.

COMMAND:load,save,print,renumber,compile,list #,delete #,add #: add 245 ENTRY TYPE; panel,event,item,blank item, OR [RETURN]: blank item

Upon listing, the schema now appears as shown below. Note that no details need to be specified in order to create a blank item, and that the blank item has been inserted between lines 240, the panel line, and 247, the first item. At a later time, after data collection has begun, this blank item may be replaced with a real item without affecting the organization of the schema.

240.0 PANEL    lab ,Laboratory data ,numeric;
245.0 ITEM     ***** ,.................................,num ,none,* ,no ;
247.0 ITEM     PROT ,Protenuria * ,char,code (4),
247.1          (prot neg,1+ ,2+ ,3+ ),no ;
250.0 ITEM     UNA ,Urine Na ,meq/l ,num ,range,(5,300),
250.1          no ;
Entering an Event

COMMAND: load, save, print, renumber, compile, list #, delete #, add #:
add 270
ENTRY TYPE: panel, event, item, blank item, OR [RETURN]: event
EVENT ABBREV: NEGPROT
DESCRIPTION: Negative Protenuria
CONTROLLING ITEM ABBREV: PROT:
MODIFIER: first, last, last, OR min: last
ENTER A PARTICULAR TRIGGERING VALUE, OR HIT
[RETURN] FOR ANY VALUE: prot neg

COMMAND: load, save, print, renumber, compile, list #, delete #, add #:

Had max or min been chosen as the modifier, the

ENTER A PARTICULAR TRIGGERING VALUE . . .:

prompt would not have been displayed.

The user has

- Created a new line, 270.
- Declared it to be an event.
- Given it a name, NEGPROT, and a description.
- Specified the item, PROT, which controls the event.
- Specified that the event is to be the last time that the item PROT takes
  on the value prot neg.

As data are collected, additional instances of PROT which have the value prot neg
but have later data collection times may be added to the SDF. As a result, the time
associated with the event NEGPROT may change.

OPTION 2: PRINT A FORMATTED SCHEMA

This option produces a specially formatted listing on the system printer of the
saved text version of the schema. The listing is useful for reference purposes and
may be used as a data collection form.

It differs from the version available under option 1 in several ways:
- Each new page starts with a panel heading.
- The various fields are aligned and have headings.
- There are no line numbers.

The listing below is of the same schema used as the example in option 1.
**Schema hypotens 1/13/1978 1532**

### PANEL 1 demos

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
<th>Units</th>
<th>ITyrel Range</th>
<th># Codes</th>
<th>Text Len</th>
<th>ICl</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 name</td>
<td>patient name</td>
<td>text</td>
<td>(10)</td>
<td>C</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 address</td>
<td>patient address</td>
<td>text</td>
<td>(20)</td>
<td>C</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 phone</td>
<td>patient home phone</td>
<td>text</td>
<td>(8)</td>
<td>C</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### PANEL 2 Phys

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
<th>Units</th>
<th>ITyrel Range</th>
<th># Codes</th>
<th>Text Len</th>
<th>ICl</th>
</tr>
</thead>
<tbody>
<tr>
<td>4 adm date</td>
<td>date admitted to stu</td>
<td>date</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 sex</td>
<td>sex</td>
<td>char</td>
<td>code (3)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 age</td>
<td>age at last brithd</td>
<td>years</td>
<td>num 25.75</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7 height</td>
<td>height at admission</td>
<td>cm</td>
<td>num 150.250</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8 weight</td>
<td>weight at adm</td>
<td>kg</td>
<td>num 40.150</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9 control</td>
<td>control group code</td>
<td>char</td>
<td>code (3)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### PANEL 3 vitals

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
<th>Units</th>
<th>ITyrel Range</th>
<th># Codes</th>
<th>Text Len</th>
<th>ICl</th>
</tr>
</thead>
<tbody>
<tr>
<td>10 sys</td>
<td>systolic pres cuff</td>
<td>torr</td>
<td>num 90.200</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11 dia</td>
<td>diastolic bp cuff</td>
<td>torr</td>
<td>num 50.150</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12 hrt</td>
<td>heart rate</td>
<td>/min</td>
<td>num 60.160</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13 temp</td>
<td>core temp</td>
<td>oral</td>
<td>deg o</td>
<td>num 95.107</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### PANEL 4 chem

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
<th>Units</th>
<th>ITyrel Range</th>
<th># Codes</th>
<th>Text Len</th>
<th>ICl</th>
</tr>
</thead>
<tbody>
<tr>
<td>14 sodium</td>
<td>plasma sodium level</td>
<td>meq/l</td>
<td>num 110.200</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15 potass</td>
<td>serum k level</td>
<td>meq/l</td>
<td>num 3.5.5.8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16 sgot</td>
<td>blood sgot level</td>
<td>meq/l</td>
<td>num 0.25</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### PANEL 5 urine

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
<th>Units</th>
<th>ITyrel Range</th>
<th># Codes</th>
<th>Text Len</th>
<th>ICl</th>
</tr>
</thead>
<tbody>
<tr>
<td>17 urinor</td>
<td>24 hour urine output</td>
<td>ml</td>
<td>num 0.600</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18 glc</td>
<td>urine glucose</td>
<td>g</td>
<td>num 0.100</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### PANEL 6 lab

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
<th>Units</th>
<th>ITyrel Range</th>
<th># Codes</th>
<th>Text Len</th>
<th>ICl</th>
</tr>
</thead>
<tbody>
<tr>
<td>19 *****001</td>
<td>***************</td>
<td>num</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20 PROT</td>
<td>Protenuria</td>
<td>char</td>
<td>code (4)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21 UNA</td>
<td>Urine Na</td>
<td>meq/l</td>
<td>num 5.300</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Schema hypotens 1/13/1978 1532**

### EVENTS

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
<th>Item</th>
<th>Modifier</th>
<th>Trigger</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 NEGPROT</td>
<td>Negative Protenuria</td>
<td>PROT</td>
<td>last</td>
<td>prot neg</td>
</tr>
</tbody>
</table>
OPTION 3: CREATE/DELETE STUDY FILES

This option permits the user to personally create and delete his study files (i.e., SDF and Patient File) and, as a consequence, to lock and unlock his schema. To create his study files, the user must create and compile his schema, then specify the number of patients and the number of panel instances per patient he expects. The system creates the study files, locks the schema, and allows the user to enter and retrieve data on a trial basis. If the user finds that his schema is unsatisfactory and he has entered only a small amount of data (no more than 10 patients and no more than 100 panel instances), he may use this option to unlock the schema by deleting all of the patients and their data. The schema may then be modified using describe, option 1. However, if the user has entered a substantial amount of data before deciding to change the schema, he cannot personally delete his files; only the system manager can affect the study files, because he has the ability to preserve valuable data that otherwise would be lost.

When the user is beyond the trial stage and is specifying the number of patients and panel instances he actually expects in his study, he should be realistic and not make these numbers larger than necessary. If he reserves too much disk space for his study by making these numbers too large, he wastes space that could otherwise be used by his colleagues and himself.

Dialog

Creating Study Files. The user selects describe, option 3, when he is prepared to test his schema. If the schema has not been compiled, the system responds with

HIT RETURN TO CONTINUE WITH describe:
THE STUDY DATA FILE CANNOT BE BUILT UNTIL THERE IS A SCHEMA.
USE l describe TO ENTER AND COMPILE A SCHEMA.

Describe, option 1, allows the user to compile his schema. After successful schema compilation, he may return to option 3 to lock his schema. The system prompts for the number of patients on the study and the number of instances (i.e., repetitions) of each panel expected for the average patient. Small numbers of patients (10 or fewer) and panel instances should be specified for trial data entry and retrieval; the numbers can be made larger later. The dialog proceeds as follows:

HOW MANY PATIENTS DO YOU EXPECT ON THIS STUDY? 5
THE NUMBER OF PATIENTS IS 5
HOW MANY INSTANCES PER PATIENT OF PANEL 1 (demog ) DO YOU EXPECT? 1
HOW MANY INSTANCES PER PATIENT OF PANEL 6 (lab ) DO YOU EXPECT? 8

After the user provides the number of instances per patient expected for all of the panels, the system prompts with

TYPE yes TO ACTUALLY CREATE THE STUDY FILES, ELSE TYPE no:

If the user finds it necessary to redefine his study parameters (i.e., the patient and instance numbers), he types no and repeats the process of creating his study files.
If he is satisfied with the parameters as defined, he types yes, and CLINFO presents information regarding the file sizes which have been allocated for the study.

FILE SIZES IN CHARACTERS (SCHEMA FILES LOCKED AND STUDY DATA FILES BUILT)

<table>
<thead>
<tr>
<th>File Type</th>
<th>Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>STUDY DATA FILE</td>
<td>48128</td>
</tr>
<tr>
<td>PATIENT FILE</td>
<td>512</td>
</tr>
<tr>
<td>SCHEMA FILES</td>
<td>46080</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td><strong>94720</strong></td>
</tr>
</tbody>
</table>

HIT {RETURN} TO CONTINUE WITH describe:

After hitting the RETURN key, the user is notified that his schema files are locked and may be altered only by the system manager. The user may now proceed with data entry and, subsequently, data retrieval functions.

**Deleting Study Files.** Option 3 is used to delete trial study files and to unlock the schema before modifying it or before creating the actual study files. The system assumes that the user intends to delete his study files if he enters this option when study files already exist. If the user has entered more than 10 patients or more than 100 panel instances, he is prevented from personally deleting his files and is so notified by a message like

YOU HAVE 5 PATIENTS AND 141 PANELS IN YOUR STUDY DATA FILES
THE SYSTEM MANAGER MUST DELETE THESE FILES. HIT {RETURN} TO CONTINUE:

Hitting the RETURN key takes the user to the menu of describe options.

If the user has entered data for no more than 10 patients and no more than 100 panel instances, he is prompted as follows:

YOU HAVE 5 PATIENTS AND 80 PANELS IN YOUR STUDY DATA FILES
TYPE yes TO DELETE THEM, ELSE HIT {RETURN} TO CONTINUE WITH describe: yes

**WARNING** YOU WILL DELETE ALL OF YOUR PATIENTS AND ALL OF THEIR DATA
TYPE delete IF THAT IS WHAT YOU INTEND, ELSE HIT {RETURN}: delete
THE SCHEMA IS UNLOCKED, TYPE yes TO MODIFY IT, ELSE HIT {RETURN}: yes

Typing delete above deletes the SDF, the Patient File, and the compiled schema and also unlocks the schema for modification. Typing yes in response to the last prompt is equivalent to typing !describe, 1, i.e., it takes the user to option 1. Hitting the RETURN key in response to the last prompt takes the user back to the start of option 3 with no study files or compiled schema; he must at least compile the schema (with describe, option 1) before creating new study files.
Enter provides for entering data either directly or from a worksheet into the Update File (from which they are merged into the SDF), adding new patient abbreviations to a study, and reviewing data already entered into the Update File.

The SDF is organized to facilitate efficient data retrieval because the clinical investigator’s primary interaction with the system is in activities involving retrieval and analysis. This organization is not necessarily optimal for data entry, however, and as a result, updating of the SDF during data entry could consume considerable amounts of time and resources. Therefore, an Update File is created as data are entered; the data in this file are then merged with the SDF by a separate sort/merge procedure that is run by the system manager when the computing load permits. Newly entered data cannot be corrected in or retrieved from the SDF until this has been accomplished. (However, new data can always be entered directly into a worksheet for immediate retrieval and analysis.)

Values for data items must be identified by patient abbreviation, date of data collection, and time of data collection. These three identifiers are referred to as the data context. Data entry with enter, option 1, is usually convenient when data are to be entered in an arbitrary order, e.g., when they are entered from lab slips, or when repeated values for different patients, items, or times are not already in tabular form. Option 2 is useful when data values are, or can easily be, stored in worksheets, or when derived values are to be entered into the SDF from worksheets.

Dialog

The user types ENTER to use this activity, and the system prompts:

THE CURRENT UPDATE FILE CONTAINS 3228 CHARACTERS

ENTER OPTIONS

<table>
<thead>
<tr>
<th>TYPE TO</th>
</tr>
</thead>
<tbody>
<tr>
<td>9..ADD NEW PATIENT ABBREVIATIONS INTO THE STUDY</td>
</tr>
<tr>
<td>1..ENTER DATA INTO THE UPDATE FILE</td>
</tr>
<tr>
<td>2..COPY AND SCREEN DATA FROM A WORKSHEET INTO THE UPDATE FILE</td>
</tr>
<tr>
<td>3..REVIEW DATA IN THE UPDATE FILE</td>
</tr>
</tbody>
</table>

ENTER OPTION:

The first line is shown only if there is an Update File that has not yet been merged into the SDF. If the schema has not been locked (see describe), the following message will be displayed when the user tries to call for the enter activity:

NO PATIENT FILE HAS BEEN ALLOCATED.
CONTACT THE SYSTEM MANAGER

The schema must be locked before patients or their data may be entered into a study.
OPTION 0: ADD NEW PATIENT ABBREVIATIONS INTO THE STUDY

This option provides for entering the patient abbreviations that will be used throughout the study. A patient can be referenced only after his 8-character patient abbreviation has been added to the Patient File through the use of this option. The patient abbreviation is used to establish the context each time data are entered. Thus, it should be the one by which the subject is generally identified for protocol data collection.

The dialog proceeds:

PATIENT ABBREV: Case1
PATIENT Case1 WILL BE ADDED TO THE PROTOCOL

PATIENT ABBREV: lent, I

TYPE yes TO DELETE ALL PATIENTS ADDED DURING THIS ACTIVITY, ELSE [RETURN]:

The user has added a patient with the abbreviated name Case1 to the list of patients and is then prompted for another. He indicates that there are no more patients to be added and that he would like to enter data by typing lent, I. If he types yes in response to the last prompt, all the patient abbreviations just entered (since enter, option 0, was last invoked) will be discarded.

To protect privacy (and to prevent possible legal action), patient names and Social Security numbers should not be used as patient abbreviations.

OPTION 1: ENTER DATA INTO THE UPDATE FILE

In response to prompts, the user types patient abbreviations, data collection dates and times and item values for direct entry into the Update File. Values are screened for validity as they are typed. The system manager must run merge to store the values in the Study Data File.

The prompts and responses for this option occur at several levels. In general terms, these are:

- Responses to the ENTER . . . ? prompt.
- Responses to context prompts.
- Responses to item-name prompts.

After the user selects option 1, the context is established and the ENTER . . . ? prompt is displayed. The entire prompt (showing the response options) is displayed only with the first ?, or if an error is made; otherwise, only the question mark is displayed. The user is prompted with ?

1. Following a context change.
2. Following the entry (in response to an item-name prompt) of a value for the last item in a panel.
3. Following the entry of a value for a single item using the item-name: item-value command.
4. Following certain errors.
The _enter_ commands and several short prompting sequences are shown below. A more detailed illustration of data entry is given in Sec. III. If a great deal of data are to be entered in a fixed sequence, a Response File should be constructed (see _response files_). A brief example is shown at the end of the discussion of this option.

**Establishing or Changing the Context**

The system establishes the initial context by prompting for the user's initials and then for the first patient abbreviation and the date and time of data collection. It is important to note that the date and time are those which identify subsequent values in the SDF and should correspond to the data or sample collection, _not_ to CLINFO data entry. This context then applies to all subsequent data entered until the user changes the context. The following example shows both the initial prompts and illustrative user responses (following the colons):

```
PLEASE TYPE YOUR INITIALS: abc

ADDING NEW DATA FOR

PATIENT ABBREV: Wiley K
DATE OF SAMPLE: 12/13/74
TIME OF SAMPLE: 1335
ENTER panel-name OR item-name:item-value OR context OR
patient:patient-abbrev OR date:date-value OR time:time-value

? 
```

Responses to ? that result in changing the context are _context, patient:patient-abbreviation, date:date-value, and time:time-value_, where the first word (shown here in italics) must be entered as shown and the value after the colon is user-specified.

**patient:patient-abbreviation**

The patient-abbreviation command (e.g., _patient: Wiley K_) in response to ? changes the current context. It must be used prior to entering data for a different patient. Any data entered following this command will be identified by the new context. The abbreviation following _patient_: cannot exceed eight characters, including spaces and punctuation (if any). The abbreviation is checked against the list of patient abbreviations that have previously been entered using _enter_, option 0.

If there is a match with the current list of patient abbreviations, the system responds with further prompts, giving the user the opportunity to change the current context (indicated by the values in parentheses) by entering a different patient abbreviation, date, or time or to leave a value unchanged by pressing the RETURN key in response to any of the prompts.

```
patient: Wiley K

ADDING NEW DATA FOR

PATIENT ABBREV (RETURN) FOR Wiley K:

DATE OF SAMPLE (RETURN) FOR 12/13/1974:

TIME OF SAMPLE (RETURN) FOR 1335:
```
If there is no match with the list of patient abbreviations, the system issues the message:

Wiley K HAS NOT BEEN ENTERED AS A PATIENT.

and then reprompts with ?.

date: date-value

The *date* command (e.g., *date: 12/14/74*) in response to ? also changes the current context. It or one of the other context-changing commands must be used prior to entering data for a different date of sample collection. The system checks the date for correct form and permissible month/day/year combinations. The year may be a two-digit number (e.g., 74) referring to a year in the twentieth century or a four-digit number (e.g., 1974 or 1887). If the date as entered is legitimate, the next prompt is again the sequence:

**ADDING NEW DATA FOR**

PATIENT ABBREV ([RETURN] FOR Wiley K):

DATE OF SAMPLE ([RETURN] FOR 12/14/1974):

TIME OF SAMPLE ([RETURN] FOR 1335):

with the patient, time, and new date repeated for verification. If the date value entered is not legitimate, the system will give an error message and reprompt with ?.

time: time-value

The *time* command (e.g., *time: 1300*) in response to ? also changes the current context. It or one of the other context-changing commands must be used prior to entering data for a different time of sample collection. The system checks the entry for correct format (i.e., hhmm expressed in 24-hour clock time) and for legitimate range (i.e., 0000-2400). The number 0 refers to midnight of the previous date; 2400 refers to midnight of the current date. If the time as entered is legitimate, the next prompt is the verification sequence:

**ADDING NEW DATA FOR**

PATIENT ABBREV ([RETURN] FOR Wiley K):

DATE OF SAMPLE ([RETURN] FOR 12/13/1974):

TIME OF SAMPLE ([RETURN] FOR 1300):

If the time entry is not legitimate, the system issues an error message followed by the prompt ?.

Thus, for any change in the current context, the system prompts for possible changes in each of the elements of the context. In each case, pressing the RETURN key maintains the current context for data entered subsequently.
context

The context command may be used to display or change the current context (i.e., patient, date, and time). The system responds with the following prompts, which indicate the current context in parentheses; the user may change the context by entering a different patient abbreviation, date, or time, or he may leave a value unchanged by pressing the RETURN key in response to any prompt.

To display the context and change the time from, say, 1335 to 1630, leaving the patient and date of collection unchanged, the user would respond as follows:

? context

ADDITION NEW DATA FOR

PATIENT ABBREV ([RETURN] FOR Case 13): __________

DATE OF SAMPLE ([RETURN] FOR 12/13/1974): ________

TIME OF SAMPLE ([RETURN] FOR 1335): 1630

Entering Data by Panels and Making Corrections

Panel-abbreviation

Typing the name of a panel (e.g., vitals) in response to ? causes the system to prompt for all the items in the named panel; the user does not need to name the individual items. He may respond to any item prompt with an item name and value in brackets, like [item-name:item-value], in order to correct a previous entry, or he may enter a panel name in brackets, like [panel-name], to initiate item prompts for that panel. For example:

? vitals

temp:102

resp:[RETURN]

dbp:50

50 IS OUT OF RANGE

IF OK, TYPE YOUR INITIALS, ELSE HIT [RETURN]:abc

sbp:[temp:102.3]

sbp:120

pulse:[blood]

glucose:
Prompting will continue for each item in the blood panel; after the last item in this panel, the system will again prompt with a question mark.

The above example shows five different kinds of response:

1. 102 and 120, which are legitimate values for the items.
2. 50, which is out of range according to the schema but is verified as being acceptable by the user.
3. [RETURN], which skips this item and goes on to the next; the value of resp for this panel instance is left undefined.
4. [temp:102.3], a parenthetical remark (in the format [item-name: item-value]) that corrects a previous entry in this, or another, panel. Such remarks may be used with any item name. In the example above, although the item sbp is prompted for, the value of temp is corrected instead; then sbp is prompted for again.
5. [blood], which has the format [panel name]; this response terminates the prompting for this panel and initiates prompting for items (starting with glucose) in the blood panel.

Not illustrated is the permitted reply [ ], which terminates the prompting for this panel and causes the system to prompt with ?. The reply [context] may also be used to leave this panel and cause the system to prompt for patient, date, and time.

When an out-of-range entry is made, the system redisplay the entry. If the out-of-range value was entered intentionally, the system proceeds to the next item after requesting verification; if the entry was erroneous, the system reprompts for the item value. If a value of the wrong type or an inappropriate character string is entered, the system redispalyes the entry with a message and immediately reprompts for the item value.

**Entering a Value for an Item**

**item-name:item-value**

The item-name:item-value response (e.g., resp:14) to (?) produces an Update File entry for this item identified with the current context. Until the context is changed, an incorrect entry may be changed by simply reentering it (e.g., resp:15) in response to ? (square brackets are necessary only if the current prompt is other than ?). After the context has been changed, the appropriate context must be reestablished (via the patient, date, time, or context commands) and the item value reentered. If an inappropriate or out-of-range value is entered, the system responds as described above.

**Determining the Context**

The context command causes the system to display the current patient, date, and time and to prompt for changes as shown above.

**Terminating the Data Entry Session**

The activity-name command (e.g., hwork or henter.3) terminates the data entry session and proceeds to the named activity. Prior to leaving this subactivity, the system prompts
TYPE yes TO DELETE ALL DATA ENTERED DURING THIS ACTIVITY, ELSE [RETURN]:

If the user types yes, all the data entered since the prompt PLEASE TYPE YOUR INITIALS; ADDING NEW DATA FOR ... will be discarded. If he responds by hitting the RETURN key, none of the data will be discarded.

Storing Data in the Study Data File

As data are entered, they are stored in the Update File, not in the SDF. They cannot be changed with the examine activity or retrieved until the system manager moves them to the SDF with the sort/merge procedure. This is usually done on request or at a prearranged time. However, data in the Update File can be examined and can be discarded, using enter, option 3. New data can be analyzed immediately if entered into a worksheet.

Using Response Files for Data Entry

If the schema is at all complex, data entry may be made much more efficient by use of the Response File facility, which permits the user to specify the sequence of prompts for individual items, for entire panels, and for context changes. (For complete details, see the discussion of response files; the examples below are only outlines.) For example, if the schema has three panels, history, bp, and urine, a Response File of the following form might be created:

```
1.0ilent,1
2.0de(data enterer's initials)
3.0?(prompt for patient)
4.0?(prompt for date)
5.0?(prompt for time)
6.0hist(response to ENTER ...)
7.0?(prompt for 1st hist item)
8.0?(prompt for 2nd hist item)
9.0[bp](change to bp panel before completing list)
10.0?(prompt for first 2 items)
11.0?then skip to 4th)
12.0(no value to be entered for 3rd item)
13.0?
14.0time?(prompt for a new time)
15.0(verify the patient, date and time)
16.0
17.0
18.0bpt(prompt for 2nd instance of bp panel)

23.0gluc?(prompt for the value gluc)
```

(cycle to line 3 to prompt for another patient, etc.)
This Response File, when executed, prompts for the patient, date, and time; permits the user to enter two items from the hist panel; changes panels (keeping the same context); prompts for the first, second, and fourth item in bp; and lets the user enter a new time, add another bp panel, add the value of gluc in the urine panel, and then change the context for a new patient and repeat the cycle. The command in line 23.0 is originally entered with a value (e.g., gluc:24) and is then edited, using response, option 3, to the form shown. A series of such commands permits the user to enter items in any order, independent of their panels.

During the execution of a Response File, the user may enter parenthetical remarks (e.g., syspress:120) to correct a previously entered value or may enter a value for an item not included in the Response File.

OPTION 2: COPY AND SCREEN DATA FROM A WORKSHEET INTO THE UPDATE FILE

This option of enter permits the user to copy data into the Update File from a worksheet. (Data may be copied in the opposite direction—into a worksheet from the SDF—by using retrieve, option 2.) The data are screened in the same way as data entered by use of option 1. The Update File may then be merged into the SDF.

There are several circumstances in which the user may wish to move data from a worksheet into the SDF. For example:

- It may be easier to enter data for portions of studies, or for entire studies, into worksheets than directly into the Update File, using enter, option 1. This would be the case in a retrospective study, or any study for which all data were available concurrently. As another example, when data are coded, values may be entered into worksheets (using option 4 or 5 of worksheet) as either numbers or codes, whereas with option 1 of enter, they must be entered as codes (which usually requires more typing than does the entering of numbers).
- Data to be stored in the SDF may be the results of computations to be performed on “raw data” which are not themselves to be stored in the SDF.
- Some data to be stored in the SDF may need to be derived from data already there (which must be retrieved into worksheets for processing). The derived data, now in the worksheet, may then be placed in the SDF without manual entry.
- Data may be available on paper tape or magnetic tape, from which they may be moved into worksheets, using option 6 of retrieve. From there, they may be further processed or moved to the SDF.

As with option 1, the context (i.e., the patient abbreviation, date, and time) must be established for each item value (or set of item values in a panel) before values can be entered into the Update File.

The worksheet must be of the general patient form (where rows and columns may be interchanged):
or of the general *time* form:

\[
\begin{array}{cccc}
\text{Patient 1} & \text{Patient 2} & \cdots & \text{Patient } p \\
\text{Item 1} & \cdot & \cdot & \cdot \\
\text{Item 2} & \cdot & \cdot & \cdot \\
\vdots & \vdots & \ddots & \ddots \\
\text{Item } i & \cdot & \cdot & \cdot \\
\end{array}
\]

Dates and times may be specified in a number of ways:

- By direct entry of an actual data collection date and time which apply to the entire worksheet (for a *time* worksheet).
- By indicating that the system is to use properly formatted labels as times relative to an event or to an entered absolute date and time (for a *patient* worksheet).
- By indicating which rows (or columns) contain data collection dates (and times) for sets of different item values in columns (or rows).

In this last case, the worksheet might be of the form:

\[
\begin{array}{cccc}
1 & 2 & \cdots & p \\
\text{Patient 1} & \cdot & \cdot & \cdot \\
\text{Patient 2} & \cdot & \cdot & \cdot \\
\vdots & \vdots & \ddots & \ddots \\
\text{Patient } p & \cdot & \cdot & \cdot \\
\end{array}
\]

Here the first two rows (labeled DATE and TIME) are assumed to contain absolute dates and times which establish the data collection times for items 1 through \(i\) for each patient. Note that worksheets of this form, or of either of the two preceding forms, may be created using *retrieve*, option 2.

The items to which the values belong are established by labeling the rows or columns with schema-defined item names.
Finally, the patient context is established either by associating data in rows (or columns) with row (or column) labels that are patient abbreviations or by directly entering a patient abbreviation as the context for the data in a range of rows (or columns). For example, the user might create an item-by-time worksheet (i.e., a patient-type worksheet) in which each group of items or times applies to a different patient. Even though such a worksheet cannot be retrieved with retrieve, option 2, its data can be copied into the Update File by entering a patient abbreviation for each row or column range that applies to a different patient.

**Dialog**

**Example 1: Absolute Dates and Times in Columns (or Rows)**

The worksheet used in this first example is shown below. It is a patient-type worksheet with columns of dates and times (1 and 2). The dates and times and the first three item columns (3 through 5) could have been retrieved from the SDF or entered directly into the worksheet. The last column has been computed from columns 4 and 5 by use of the calculate activity. The user now wishes to move the calculated data into the Update File (and then into the SDF). The label of column 6 is an item name in the schema.

<table>
<thead>
<tr>
<th>ROWS/COLS</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>LABELS</td>
<td>DATE</td>
<td>TIME</td>
<td>PROT</td>
<td>UNA</td>
<td>UCR</td>
<td>N/C*100</td>
</tr>
<tr>
<td>1</td>
<td>1/2/1977</td>
<td>905</td>
<td>prot neg</td>
<td>52</td>
<td>.200</td>
<td>2.650</td>
</tr>
<tr>
<td>3</td>
<td>1/5/1977</td>
<td>655</td>
<td>1+</td>
<td>101</td>
<td>.200</td>
<td>1.262</td>
</tr>
<tr>
<td>4</td>
<td>1/7/1977</td>
<td>915</td>
<td>prot neg</td>
<td>115</td>
<td>.200</td>
<td>1.278</td>
</tr>
<tr>
<td>5</td>
<td>1/8/1977</td>
<td>892</td>
<td>1+</td>
<td>112</td>
<td>.200</td>
<td>1.600</td>
</tr>
<tr>
<td>6</td>
<td>1/9/1977</td>
<td>833</td>
<td>2+</td>
<td>130</td>
<td>.200</td>
<td>1.300</td>
</tr>
<tr>
<td>7</td>
<td>1/10/1977</td>
<td>902</td>
<td>1+</td>
<td>90</td>
<td>.300</td>
<td>2.250</td>
</tr>
<tr>
<td>8</td>
<td>1/12/1977</td>
<td>928</td>
<td>prot neg</td>
<td>86</td>
<td>.300</td>
<td>2.867</td>
</tr>
</tbody>
</table>

The first prompt is

**TYPE c IF DATA ITEMS OCCUR IN COLUMNS, OR r IF IN ROWS: c**

Here the items are in columns, so the user responds c and is then prompted for the time specification option. Note that the entire procedure is symmetric; had the items been in rows, the words row and column would be interchanged in the following prompts:

**DATA COLLECTION TIME OPTIONS**

1. ABSOLUTE DATES AND TIMES IN COLUMNS
2. A SINGLE DATE AND TIME OF DAY OR AN EVENT
3. TIMES RELATIVE TO AN EVENT AS ROW LABELS

**TIME OPTION:1**

Option 1 is chosen because there are different times associated with each row and because the first two columns contain absolute dates and times. (Option 3 cannot
be used because there are no row labels in this case.) The system then displays the column labels (if so requested) and prompts for the location of dates and times:

COLUMN LABELS

1 DATE
2 TIME
3 PROT
4 UNA
5 UCR
6 N/C*100

COLUMN NUMBER FOR DATES ([RETURN] SELECTS 1):

COLUMN NUMBER FOR TIMES ([RETURN] SELECTS A TIME OF 1200): 2

The user might respond to the time prompt by hitting the RETURN key if there were no exact data collection times or if the exact time were not important.

To allow the user to store only selected columns (i.e., items), the system prompts:

CHOOSE A RANGE OF COLUMNS FOR ITEMS ([RETURN] SELECTS 1,6): 6, 6

Here, the user wishes to store only the computed item in column 6, so he responds 6, 6. Had all of the data been entered directly into the worksheet (instead of being retrieved from the SDF), he could now store it by specifying 3, 6.

The system then verifies that the label N/C*100 is the name of a schema item; if it is not, an error message is displayed.

Next, the user is asked if he wants to display row labels. After displaying the labels (or declining to display them), he is asked if they are patient abbreviations. This inquiry is made because at this point patients may be identified either by row labels or by entering their abbreviations. The user hits the RETURN key to indicate that row labels are not patient abbreviations.

TYPE yes IF PATIENT ABBREVS ARE IN ROW LABELS, ELSE [RETURN]:

In either case, the user is next asked to select a range of rows (i.e., different instances). In this case, he hits the RETURN key to indicate that data from all eight rows apply to the patient whose abbreviation will next be entered and that they should all be copied into the Update File.

CHOOSE A RANGE OF ROWS ([RETURN] SELECTS 1,8):

If data for more than one patient are contained in the worksheet (e.g., the first four rows belong to case 2, and the second four belong to case 3), the user would select the rows for the first patient and later select rows for the other patients.

Finally, the system prompts for the patient abbreviation appropriate to this range of rows:

THE PATIENT ABBREV. FOR THIS RANGE ([RETURN] SELECTS case1): case234

The user responds case234 and the system prompts for additional specifications:
NOTE: YOU HAVE COMPLETED A DATA TRANSFER SPECIFICATION
TYPE yes TO TRANSFER THE DATA, ELSE [RETURN] TO ADD
ANOTHER SPECIFICATION:

In this case, he responds yes and the system copies the data. If data for additional
patients were to be copied into the Update File, the user would hit the RETURN
key. Now the system copies each row of values, listing each row number as it
proceeds:

PATIENT: case123
PROCESSING ROW 1
PROCESSING ROW 2

PROCESSING ROW 8
HIT [RETURN] TO CONTINUE WITH enter,2:

Had a date or time been in error, or had a data value been out of range, an error
message would be displayed. The user may override the range check by entering
his initials, as in the example below:

PATIENT: case123
PROCESSING ROW 1
EITHER DATE OR TIME IN ROW 2 IS INCORRECT
PROCESSING ROW 2
PROCESSING ROW 3
PROCESSING ROW 4
PROCESSING ROW 5
PROCESSING ROW 6
310 IS OUT OF RANGE FOR N/C•100
TYPE YOUR INITIALS TO ACCEPT IT, ELSE HIT [RETURN]: np
PROCESSING ROW 7
PROCESSING ROW 8
HIT [RETURN] TO CONTINUE WITH enter,2:

Because the date or time is incorrect, none of the data in row 2 would be stored in
the Update File.

Example 2: A Single Date and Time of Day or an Event

The worksheet to be used is shown below:

<table>
<thead>
<tr>
<th>ROWS/COLS</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>LABELS</td>
<td>age</td>
<td>sex</td>
<td>idnl</td>
<td>wt</td>
<td>dur</td>
</tr>
<tr>
<td>1 case 1</td>
<td>27</td>
<td>0</td>
<td>100</td>
<td>0</td>
<td>35</td>
</tr>
<tr>
<td>2 case 2</td>
<td>24</td>
<td>0</td>
<td>98</td>
<td>0</td>
<td>45</td>
</tr>
<tr>
<td>3 case 3</td>
<td>32</td>
<td>0</td>
<td>98</td>
<td>0</td>
<td>45</td>
</tr>
<tr>
<td>4 case 4</td>
<td>34</td>
<td>0</td>
<td>94</td>
<td>0</td>
<td>50</td>
</tr>
<tr>
<td>5 case 5</td>
<td>21</td>
<td>0</td>
<td>96</td>
<td>0</td>
<td>30</td>
</tr>
<tr>
<td>6 case 6</td>
<td>23</td>
<td>0</td>
<td>98</td>
<td>0</td>
<td>45</td>
</tr>
<tr>
<td>7 case 7</td>
<td>28</td>
<td>0</td>
<td>99</td>
<td>0</td>
<td>45</td>
</tr>
</tbody>
</table>
It is a time worksheet, where columns represent different items and rows represent different patients.

The exact time of collection is not important for these data; by responding to later prompts, the user may store them all with the identical absolute time context, or with the same time relative to a specified event. The latter results in a different absolute time for each patient because the event has a different absolute time for each patient.

After indicating that data items occur in columns, selecting time option 2, and displaying (or declining to display) the column (item) labels, the user is prompted with:

CHOSE A RANGE OF COLUMNS FOR ITEMS ([RETURN] SELECTS 1,5):

At this point, he could exclude some of the items; here, he hits the RETURN key and selects all 5 items. He is next prompted:

EVENT NAME OR DATE AS A TIME ORIGIN ([RETURN] SELECTS begin):

Entering the name of an event will, after a few more prompts, cause the data for each patient (i.e., in each row) to be stored with a data collection time specific to that patient; if an absolute date and time are entered (as illustrated later), all of the selected patients' data will be stored with identical data collection times. Here the user responds with the RETURN key, thus selecting the event begin, which takes on the earliest data collection time entered for each patient. The next two prompts allow the time associated with each row to be offset from the named event:

THE TIME MAY BE OFFSET FROM begin BY A NUMBER OF TIME UNITS
TYPE THE UNIT (min hr day wk mon yr) OR HIT [RETURN] FOR NO
OFFSET: day

HOW MANY days:  -1

Here the user has indicated that he wishes to store the data for each case with a time one day earlier than the current earliest data collection time.

Finally, the user is prompted:

TYPE yes IF PATIENT ABBREV'S ARE IN ROW LABELS, ELSE [RETURN]: yes

and then:

CHOSE A RANGE OF ROWS ([RETURN] SELECTS 1,7):

If row labels were not patient abbreviations, the user could choose a row (or range of rows) and assign a patient abbreviation to it (or them).

If he wished to store all of the data with the identical data collection time, the user would respond to the time origin prompt as follows:

EVENT NAME OR DATE AS A TIME ORIGIN ([RETURN] SELECTS begin):
12/25/76

and would then be prompted for a time:

TIME ([RETURN] SELECTS 1200): 1000
Example 3: Times Relative to an Event or Row (or Column) Labels

The worksheet used in this example is shown below. It is a patient worksheet, with all data referring to a single patient.

<table>
<thead>
<tr>
<th>ROWS/COLS</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>LABELS</td>
<td>glucose</td>
<td>glucagon</td>
<td>alanine</td>
<td>lactate</td>
</tr>
<tr>
<td>1 HR -2</td>
<td>103.500</td>
<td>89.300</td>
<td>.322</td>
<td>1.420</td>
</tr>
<tr>
<td>2 HR -1</td>
<td>97.800</td>
<td>90.200</td>
<td>.324</td>
<td>1.300</td>
</tr>
<tr>
<td>3 HR 0</td>
<td>96</td>
<td>90.500</td>
<td>.318</td>
<td>1.500</td>
</tr>
<tr>
<td>4 HR 1</td>
<td>150.100</td>
<td>108.300</td>
<td>.298</td>
<td>1.450</td>
</tr>
<tr>
<td>5 HR 2</td>
<td>202.300</td>
<td>119.700</td>
<td>.291</td>
<td>1.180</td>
</tr>
<tr>
<td>6 HR 3</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>7 HR 4</td>
<td>257.800</td>
<td>127.300</td>
<td>.307</td>
<td>1.270</td>
</tr>
<tr>
<td>8 HR 5</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

Here the row labels are times relative to an event which the user may specify.

After the user indicates that data items occur in columns, selects time option 3, and displays (or declines to display) the column labels, the system prompts:

CHOOSE A RANGE OF COLUMNS FOR ITEMS ((RETURN) SELECTS 1,4); 1,3

Here the user has chosen to enter the first three items. After displaying (or declining to display) the row labels (i.e., the relative times), the user is prompted:

EVENT NAME OR DATE AS A TIME ORIGIN ((RETURN) SELECTS begin): instop

He has responded with the name of an event which, in this case, is the time at which insulin was stopped for this patient. Had that event not yet been defined (because no insulin data had been collected in the SDF), the user might have entered an absolute date and time as a relative zero time marker.

He is next prompted for a range of rows (i.e., relative times) of data to be entered:

CHOOSE A RANGE OF ROWS ((RETURN) SELECTS 1,8); 3,7

Here he has selected the data associated with hours 0 through 4.

Finally, the system requests the patient abbreviation, which completes the identification of the data:

THE PATIENT ABBREV. FOR THIS RANGE ((RETURN) SELECTS case 1): case 2

The system either lists the first patient in the study or remembers the last patient for whom data were stored (during this use of enter, option 2) and allows the user to select that patient by hitting the RETURN key.

OPTION 3: REVIEW DATA IN THE UPDATE FILE

This subactivity permits the user to review the contents of his Update File. When data are entered using enter, option 1 or option 2, they are stored in the Update File, where they are identified by patient abbreviation, panel, date and time of sample, enterer’s initials, and date and time of entry. They are later stored in
the SDF and deleted from the Update File when the system manager runs the sort/merge program. With enter, option 3, the user may examine the data that have been entered but not merged, find which patients are represented in the Update File, determine which entered numeric, date, or time values lie outside the range specified in the schema, and delete panels of data from the Update File. He may specify a particular patient from the list provided and may choose to review the panels in order from the first entered to the most recent, or vice versa. These choices apply to the entire session; to change options, the user must reenter the subactivity (by typing !.)

The user should be aware that a new panel is created during data entry each time the context (patient, date, or time) or panel is changed, even if the change is only a correction to previously entered values. Suppose, for example, that a patient, date, and time are specified; that data are entered for the History panel which include the item sex; and that, without changing the context, data are entered for the control panel. The user then realizes that he entered the value for sex incorrectly, so he reenters it. Although the sort/merge program handles this situation by storing in the SDF only the most recently entered values for each context, in this case the Update File contains two History panels for the same patient, date, and time, the last containing only the corrected value for sex.

The number of similar panels in the Update File depends on the exact sequence of data entry interactions. If the History panel is completed, the control panel is completed, a value for sex is entered, and then the context is changed, the Update File contains a History panel, a control panel, and another History panel that contains only a value for sex. Because the subactivity displays one panel at a time in the order in which they are stored in the Update File (or in the opposite order), the user may not come across the latest, corrected value for sex until he displays the latest History panel. If he is not aware of this, he may not realize that the value for sex has been corrected. If a second value for sex had been entered before entry of the History panel had been completed, then two values for sex would appear in that panel, but there would not be a second History panel. If a second value for sex had been entered during the entry of the control panel (e.g., by entering sexual in response to the prompt for alaninec), then the Update File would contain a History panel, a control panel that includes the first few values entered, a History panel containing sex:s, and another control panel starting with a value for alaninec. As mentioned above, only the most recently entered values for each variable are entered into the SDF during the next sort/merge process. When a panel is displayed, any value that lies outside the schema-specified range is marked OUT OF RANGE.

Dialog

In this example, the user reviews the Update File after data have been entered as illustrated in Figs. III-3 and III-4 (in Sec. III), and before they have been merged into the SDF. When the user enters the subactivity, if there is a current Update File, the system prompts:

TYPE yes TO EXAMINE THE DATA FOR A PATIENT AT A TIME, ELSE [RETURN]; yes
The user has chosen to select patients for review rather than reviewing the entire Update File in order of entry. He is then prompted:

**TYPE** yes TO REVIEW THE MOST RECENTLY ENTERED DATA FIRST, ELSE {RETURN}:

Entering yes would permit review of the Update File backwards from the last entered panel (for a particular patient). This is most useful if a large amount of data have been entered since the last merge and the user wishes to make periodic checks. It also assists the user in finding his place if the data entry session has been interrupted.

In either case, the system prompts:

**THE FOLLOWING PATIENTS HAVE DATA IN THE UPDATE FILE**

*case 1*

*case 3*

HIT {RETURN} TO CONTINUE:

Here, data for only two patients have been entered. After the user hits the RETURN key, the system prompts for a patient abbreviation from the list:

**PATIENT ABBREV. (HIT {RETURN} FOR A LIST): case 1**

The History panel and the two instances of the control panel for the patient identified as case 1 may then be displayed:

**PATIENT:** case 1 **PANEL:** History **INITIALS:** gfg

**DATE OF SAMPLE:** 5/3/1975 **DATE OF ENTRY:** 3/19/1976

**TIME OF SAMPLE:** 08:00 **TIME OF ENTRY:** 10:44

- case no.: 1
- age: 27
- sex: m
- % idol wt: 100
- dur diab: 12
- ins req: 35

**TYPE** yes TO DELETE THIS BLOCK, ELSE HIT {RETURN}:

**PATIENT:** case 1 **PANEL:** control **INITIALS:** gfg

**DATE OF SAMPLE:** 5/4/1975 **DATE OF ENTRY:** 3/19/1976

**TIME OF SAMPLE:** 06:15 **TIME OF ENTRY:** 10:47

- insulinc: 35
- glucose: 103.5
- glucose: 89.5
- alanine: .322
- lactate: 1.42

**TYPE** yes TO DELETE THIS BLOCK, ELSE HIT {RETURN}
ENTER, 3

PATIENT: case 1
DATE OF SAMPLE: 5/4/1975
TIME OF SAMPLE: 07:00
INITIALS: gfg
DATE OF ENTRY: 3/19/1976
TIME OF ENTRY: 10:48
INsulfinc: 35
Glucose: 98.7
Glucose: 97.8
Glucgone: 90.2
Alanine: 324
Lactate: 1.57

 TYPE yes TO DELETE THIS BLOCK, ELSE HIT {RETURN}:

The second control panel has glucose listed twice because first the value 98.7 was entered and then the value 97.8 was entered. During entry of the first control panel, an out-of-range value was entered for alanine, but when the system called the enterer's attention to this, he indicated that the value should not be accepted and he corrected it. Alanine does not appear twice because the out-of-range value was not stored.

After displaying the last panel and hitting the RETURN key, the user is given an opportunity to select another patient.

PATIENT ABBREV. (HIT {RETURN} FOR A LIST): case 3
PATIENT case 1 HAS NO MORE DATA IN THE UPDATE FILE

An item is listed only if a value was entered for it, i.e., if a value had not been entered for Insulfinc, then Insulfinc would not be displayed even though it is one of the items in the control panel. The item names and values are indented to leave room at the left for the words OUT OF RANGE next to an out-of-range value that the enterer verified was correct. In the following panel for case 3, Insulfinc was not entered, and the entered, verified value for alanine was outside the specified range of 0.1 to 0.8:

PATIENT: case 3
DATE OF SAMPLE: 5/7/1975
TIME OF SAMPLE: 08:12
INITIALS: gfg
DATE OF ENTRY: 3/22/1976
TIME OF ENTRY: 11:29
Glucose: 203
Glucgon: 189
OUT OF RANGE
Alanine: 3.5
Lactate: 1.7

The above examples illustrate numeric panels. Text panels are displayed similarly except that there is no space left for the words OUT OF RANGE. This is because text values are not checked for validity.

If the user wishes to review the panels in the opposite order, or to review the entire Update File without specifying patients, he can enter /and reply appropriately to the prompts.
EXAMINE

This activity permits the direct examination, editing, and deletion of data stored in the SDF. It also provides for reviewing SDF usage. The user identifies a patient, a panel, and a starting time, and the system then displays data for that patient and panel, instance by instance.

Dialog

After prompting for the data enterer's initials, the system asks if the user wishes to review his SDF usage. He responds yes if he does; he hits the RETURN key if he does not.

**TYPE yes TO REVIEW YOUR STUDY DATA FILE USAGE: yes.**

**FOR 10 PATIENTS (OF AN ESTIMATED 10), THE PANEL USAGE IS:**

<table>
<thead>
<tr>
<th>PANEL</th>
<th>(History)</th>
<th>ESTIMATED</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7</td>
<td>10/1/PATIENT</td>
</tr>
<tr>
<td>2</td>
<td>65</td>
<td>100/1/PATIENT</td>
</tr>
</tbody>
</table>

**ESTIMATED = 10 (1/PATIENT)**

**ESTIMATED = 100 (10/PATIENT)**

**ESTIMATED = 100 (10/PATIENT)**

**THE STUDY DATA FILE IS 66% FULL**

After each instance is displayed, the user is given the opportunity to delete the entire panel instance, or to edit the panel contents by changing the values of items, adding values for missing items, or changing values to missing. All changes are flagged with an appropriate code, which is displayed beside the value. The same screening is applied to changes and additions made in the SDF by use of this option as it operates in the enter activity; out-of-range values can be accepted if responses to prompts indicate that they are intentional.

When a panel instance is displayed, anything unusual about an item value is noted by a character immediately to the left of the item name:

- **M** means that the item value has been modified, i.e., it is a different value than was originally stored (via an enter subactivity) in the SDF for the same patient, date, and time.
- **E** means that, according to the schema, the value exceeds the specified range, but the data enterer indicated that the value was to be accepted anyway.
- **X** means that a modified value is out of range. The indicators E and X do not apply to char or check data types because in these cases the schema specification cannot be overridden.

Next, the system successively issues the following prompts (the prompts are in upper case, and the user's responses are in lower case):
PATIENT ABBREV.: case 1

EVENT NAME OR DATE: begin

PANEL NAME: control

The responses above would elicit the first instance of the control panel for the patient whose abbreviation is case 1. If an instance that occurred at a known date and time were desired, that date and time could be entered.

Following the control response, the following data would be displayed:

PATIENT: case 1 PANEL: control INITIALS: g/f
TIME OF SAMPLE: 06:00 TIME OF ENTRY: 12:04

insulin  . . .
glucose  103.5
glucagon  . . .
alanine  . . .
M lactate  . . .

TYPE delete, edit, patient, OR HIT |RETURN| :

The heading identifies the name, patient, date, and time of the sample, as well as the initials of the data enterer and the date and time of data entry. In this instance, four of the five items in this panel have missing values. The M preceding the item name lactatec indicates that its value has been modified.

Hitting the RETURN key then displays the next instance of the same panel:

PATIENT: case 1 PANEL: control INITIALS: g/f
TIME OF SAMPLE: 06:15 TIME OF ENTRY: 11:47

insulin  35
glucose  103.5
glucagon  89.3
alanine  .322
lactatec  1.42

TYPE delete, edit, patient, OR HIT |RETURN| :

If the user wished to delete the entire instance, he would respond delete to the prompt. This is not the same as setting all of the values to missing; in that case, the panel might still show up in certain retrievals, whereas deleting the panel removes it entirely from the SDF.

Responding to the prompt with edit causes the system to further prompt:

ITEM ABBREV. OR HIT |RETURN|: glucose

and, after the user enters glucose, prompt with the named item:

glucose: 130.5
When the user enters the value 130.5 and hits the RETURN key, the data are redisplayed with the modified value, and the original prompt is repeated:

```
insulinc 35
M glucosec 130.5
```

```


TYPE delete, edit, patient, OR HIT \RETURN\ : 

The following sequence (after a response of edit to the last prompt) replaces the value of insulinc with the missing value symbol:

ITEM ABBREV. OR HIT \RETURN\ : insulinc

insulinc: ...

That is, the user enters ... (3 periods) in response to the insulinc prompt. The data are again displayed:

```
M insulinc ...
M glucosec 130.5
```

If a value were entered which was out of the screening range for that item (as specified in the schema), the system would respond as follows:

```
ITEM ABBREV. OR HIT \RETURN\ : insulinc

insulinc: 1.5

1.5 IS OUT OF RANGE, TYPE yes TO ACCEPT IT, ELSE HIT \RETURN\ : yes

X insulinc 1.5
M glucosec 130.5
```

The X indicates that a value was changed and the new value is out of range; an E would indicate that an original value is out of range. If the item were of the type char, only those character strings associated with the item in the schema would be accepted; non-numeric values would not be accepted for numeric items.

Entering patient in response to the TYPE delete, edit, patient prompt allows the user to change the context, in a manner analogous to that of the enter activity. The user is prompted as follows:
PATIENT ABBREV. (HIT [RETURN] FOR case 1): case 2

EVENT NAME OR DATE (HIT [RETURN] FOR begin): end

PANEL NAME (HIT [RETURN] FOR control):

Here the user has indicated that he wishes to see the last instance of the panel called control for the patient called case 2.

Note that the system uses the most recent context (patient, event, and panel) values as defaults; hitting the RETURN key would specify the value in parentheses.

In response, the system displays the data shown below:

<table>
<thead>
<tr>
<th>PATIENT: case 2</th>
<th>PANEL: control</th>
<th>INITIALS: gfg</th>
</tr>
</thead>
<tbody>
<tr>
<td>DATE OF SAMPLE:</td>
<td>5/7/1975</td>
<td>DATE OF ENTRY: 11/26/1975</td>
</tr>
<tr>
<td>TIME OF SAMPLE:</td>
<td>19:00</td>
<td>TIME OF ENTRY: 12/14</td>
</tr>
</tbody>
</table>

insuline . . .
glucosec 271.6
glucogne . . .
alaninec . . .
lactatec . . .

TYPE delete, edit, patient, OR HIT [RETURN] :

Thus, the user may move about within the SDF by specifying different patients, panels, and times; and for a particular patient and panel, he may step through successive instances by hitting the RETURN key.
The *files* activity permits the user to display ordered lists and descriptions of his worksheets, subsets, Response Files, or utility files. The description of a file comprises the file’s name, its size in bytes (one byte is the storage required for one character, four bytes are required to store a number), the date and time the file was last modified, and the date it was last accessed. Utility files include all system files that are not listed separately.

The sizes of some typical worksheets are as follows:

<table>
<thead>
<tr>
<th>Dimensions (rows × cols)</th>
<th>Size (bytes)</th>
</tr>
</thead>
<tbody>
<tr>
<td>20 × 6</td>
<td>809</td>
</tr>
<tr>
<td>20 × 12</td>
<td>1341</td>
</tr>
<tr>
<td>20 × 36</td>
<td>3461</td>
</tr>
<tr>
<td>20 × 107</td>
<td>9735</td>
</tr>
<tr>
<td>47 × 47a</td>
<td>9747</td>
</tr>
</tbody>
</table>

* Largest worksheet.

The following storage space (in bytes) is required for worksheets:

\[ 4 \times n \times m + 8 \times (n + m) + 120 + \text{a small, variable amount of overhead}, \]

where \( n \) and \( m \) are the numbers of rows or columns, and 120 is the space required for the title.

The *files* activity is useful for obtaining a quick listing of worksheets or other files for reference purposes. It is also used to determine what files have been accumulated, their size, and if they are in active use. It can help the investigator and system manager to manage the disk storage space allocated to the investigator. When the investigator’s space or the entire disk becomes full, files may have to be copied onto magnetic tape or discarded. Prime candidates for discarding are old, large worksheets that have not been accessed recently. Other candidates are subsets and Response Files that have not been accessed recently.

The user types *files* (or *file*) to enter this activity and then can choose from the following options:

**FILES OPTIONS**

```
<table>
<thead>
<tr>
<th>TYPE</th>
<th>TO DISPLAY INFORMATION ABOUT</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>WORKSHEETS</td>
</tr>
<tr>
<td>2</td>
<td>SUBSETS</td>
</tr>
<tr>
<td>3</td>
<td>RESPONSE FILES</td>
</tr>
<tr>
<td>4</td>
<td>UTILITY FILES</td>
</tr>
</tbody>
</table>
```

**FILES OPTION: 1**
OPTION 1: LIST WORKSHEETS

Once the type of file (worksheets, in this subactivity) has been chosen, five ordering options are possible:

<table>
<thead>
<tr>
<th>ORDER OPTIONS:</th>
</tr>
</thead>
<tbody>
<tr>
<td>TYPE TO ORDER THESE FILES BY</td>
</tr>
<tr>
<td>1. ... FILENAME</td>
</tr>
<tr>
<td>2. ... SIZE IN CHARACTERS</td>
</tr>
<tr>
<td>3. ... DATE AND TIME LAST MODIFIED</td>
</tr>
<tr>
<td>4. ... DATE OF LAST ACCESS</td>
</tr>
<tr>
<td>5. ... NONE OF THE ABOVE (UNORDERED)</td>
</tr>
</tbody>
</table>

ORDER OPTION: 1

If there are more than 100 worksheets, the following message will appear after the files have been processed:

THE MAXIMUM OF 100 WORKSHEETS HAVE BEEN PROCESSED.
TYPE yes TO ORDER AND DISPLAY THESE, ELSE [RETURN] TO PROCESS THE NEXT 100 WORKSHEETS:

Usually this message does not appear, and the user immediately chooses either ascending or descending sort order.

TYPE yes FOR ASCENDING ORDER, HIT [RETURN] FOR DESCENDING ORDER: yes

If the user has chosen order option 5, i.e., he does not care to order the listing, CLINFO lists his worksheets immediately without asking anything further about sort order.

A maximum of 30 worksheet descriptions will be displayed on the screen. Hitting the RETURN key will display the next 30. When all worksheets have been displayed, the following message will appear:

END OF ENTRIES. HIT [RETURN] TO RESPECIFY ORDER OPTIONS:

Hitting the RETURN key again will bring the user back to the ordering options for listing worksheets.

Below is an example of a listing of worksheets by file name:
## Files, 1

### Ascending List of Worksheets Ordered by Filename

<table>
<thead>
<tr>
<th>Filename</th>
<th>Size (in bytes)</th>
<th>Date Last Modified</th>
<th>Date Last Accessed</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACC1FLR</td>
<td>361</td>
<td>06/25/76 11:23</td>
<td>06/25/76</td>
</tr>
<tr>
<td>ANI1FLR</td>
<td>1197</td>
<td>06/28/76 11:43</td>
<td>06/28/76</td>
</tr>
<tr>
<td>BAS1FLR</td>
<td>5019</td>
<td>08/25/76 13:59</td>
<td>09/03/76</td>
</tr>
<tr>
<td>CALCULAT</td>
<td>1453</td>
<td>08/25/76 11:53</td>
<td>08/25/76</td>
</tr>
<tr>
<td>CF01FLR</td>
<td>297</td>
<td>06/28/76 11:32</td>
<td>06/28/76</td>
</tr>
<tr>
<td>CH11FLR</td>
<td>297</td>
<td>06/25/76 11:12</td>
<td>06/25/76</td>
</tr>
<tr>
<td>COLCOMP</td>
<td>1339</td>
<td>08/09/76 11:11</td>
<td>08/09/76</td>
</tr>
<tr>
<td>COM1FLR</td>
<td>297</td>
<td>06/25/76 11:38</td>
<td>06/25/76</td>
</tr>
<tr>
<td>CP11FLR</td>
<td>297</td>
<td>06/25/76 11:58</td>
<td>06/25/76</td>
</tr>
<tr>
<td>DES1FLR</td>
<td>425</td>
<td>06/25/76 10:45</td>
<td>06/25/76</td>
</tr>
<tr>
<td>DUMMY</td>
<td>683</td>
<td>06/25/76 09:07</td>
<td>12/29/76</td>
</tr>
<tr>
<td>EXP1FLR</td>
<td>489</td>
<td>06/25/76 11:47</td>
<td>06/25/76</td>
</tr>
<tr>
<td>EXPENSES</td>
<td>9755</td>
<td>06/28/76 13:36</td>
<td>07/19/76</td>
</tr>
<tr>
<td>FHD11FLR</td>
<td>619</td>
<td>06/25/76 09:25</td>
<td>06/25/76</td>
</tr>
<tr>
<td>JUNK</td>
<td>2995</td>
<td>07/23/76 09:51</td>
<td>09/01/76</td>
</tr>
<tr>
<td>MHD1FLR</td>
<td>297</td>
<td>06/25/76 09:16</td>
<td>06/25/76</td>
</tr>
<tr>
<td>MTAL1FLR</td>
<td>619</td>
<td>06/25/76 09:21</td>
<td>06/25/76</td>
</tr>
<tr>
<td>OTH11FLR</td>
<td>297</td>
<td>06/28/76 11:27</td>
<td>06/26/76</td>
</tr>
<tr>
<td>OTHS1FLR</td>
<td>297</td>
<td>06/25/76 11:02</td>
<td>06/25/76</td>
</tr>
<tr>
<td>RES1FLR</td>
<td>361</td>
<td>06/25/76 11:17</td>
<td>06/25/76</td>
</tr>
<tr>
<td>RET1FLR</td>
<td>619</td>
<td>06/25/76 11:17</td>
<td>06/25/76</td>
</tr>
<tr>
<td>REV1FLR</td>
<td>297</td>
<td>06/28/76 11:29</td>
<td>06/28/76</td>
</tr>
<tr>
<td>TIMEDIFF</td>
<td>489</td>
<td>08/28/76 14:21</td>
<td>08/25/76</td>
</tr>
<tr>
<td>TRN3FLR</td>
<td>489</td>
<td>06/25/76 11:42</td>
<td>06/25/76</td>
</tr>
<tr>
<td>UND1FLR</td>
<td>297</td>
<td>06/25/76 11:47</td>
<td>06/25/76</td>
</tr>
<tr>
<td>UNDS1FLR</td>
<td>297</td>
<td>06/25/76 11:47</td>
<td>06/25/76</td>
</tr>
<tr>
<td>UTLFLR</td>
<td>361</td>
<td>06/28/76 11:07</td>
<td>06/28/76</td>
</tr>
<tr>
<td>WUL1FLR</td>
<td>297</td>
<td>06/28/76 11:15</td>
<td>06/28/76</td>
</tr>
</tbody>
</table>

Hit [RETURN] to continue listing your worksheets:

Worksheets may also be listed by size in characters (i.e., in bytes) in ascending order:

### Files, 1

### Ascending List of Worksheets Ordered by Size in Characters

<table>
<thead>
<tr>
<th>Filename</th>
<th>Size (in bytes)</th>
<th>Date Last Modified</th>
<th>Date Last Accessed</th>
</tr>
</thead>
<tbody>
<tr>
<td>CF01FLR</td>
<td>297</td>
<td>06/28/76 11:32</td>
<td>06/25/76</td>
</tr>
<tr>
<td>CH11FLR</td>
<td>297</td>
<td>06/25/76 11:12</td>
<td>06/25/76</td>
</tr>
<tr>
<td>COM1FLR</td>
<td>297</td>
<td>06/25/76 11:30</td>
<td>06/25/76</td>
</tr>
<tr>
<td>CP11FLR</td>
<td>297</td>
<td>06/25/76 11:59</td>
<td>06/25/76</td>
</tr>
<tr>
<td>OTNO1FLR</td>
<td>297</td>
<td>06/25/76 09:16</td>
<td>06/25/76</td>
</tr>
<tr>
<td>OTNS1FLR</td>
<td>297</td>
<td>06/28/76 11:27</td>
<td>06/28/76</td>
</tr>
<tr>
<td>REV1FLR</td>
<td>297</td>
<td>06/28/76 11:29</td>
<td>06/28/76</td>
</tr>
<tr>
<td>TRN1FLR</td>
<td>297</td>
<td>06/25/76 11:47</td>
<td>06/25/76</td>
</tr>
<tr>
<td>UND1FLR</td>
<td>297</td>
<td>06/25/76 11:02</td>
<td>06/25/76</td>
</tr>
<tr>
<td>UNDS1FLR</td>
<td>297</td>
<td>06/25/76 11:42</td>
<td>06/25/76</td>
</tr>
<tr>
<td>WUL1FLR</td>
<td>297</td>
<td>06/28/76 11:15</td>
<td>06/28/76</td>
</tr>
<tr>
<td>ACC1FLR</td>
<td>361</td>
<td>06/25/76 11:23</td>
<td>06/25/76</td>
</tr>
<tr>
<td>RES1FLR</td>
<td>361</td>
<td>06/25/76 11:37</td>
<td>06/25/76</td>
</tr>
<tr>
<td>UNDO1FLR</td>
<td>361</td>
<td>06/28/76 11:40</td>
<td>06/28/76</td>
</tr>
<tr>
<td>UTL1FLR</td>
<td>361</td>
<td>06/28/76 11:07</td>
<td>06/28/76</td>
</tr>
<tr>
<td>DES1FLR</td>
<td>425</td>
<td>06/25/76 11:07</td>
<td>06/25/76</td>
</tr>
<tr>
<td>ENT1FLR</td>
<td>489</td>
<td>06/28/76 14:21</td>
<td>06/25/76</td>
</tr>
<tr>
<td>TIMEDIFF</td>
<td>489</td>
<td>08/28/76 11:02</td>
<td>06/28/76</td>
</tr>
<tr>
<td>TRN3FLR</td>
<td>489</td>
<td>08/28/76 09:25</td>
<td>06/25/76</td>
</tr>
<tr>
<td>MTAL1FLR</td>
<td>619</td>
<td>06/28/76 09:21</td>
<td>06/25/76</td>
</tr>
<tr>
<td>RET1FLR</td>
<td>619</td>
<td>06/25/76 11:17</td>
<td>06/25/76</td>
</tr>
<tr>
<td>DUMMY</td>
<td>683</td>
<td>06/25/76 09:07</td>
<td>12/29/76</td>
</tr>
<tr>
<td>ANI1FLR</td>
<td>1197</td>
<td>06/28/76 11:43</td>
<td>08/09/76</td>
</tr>
<tr>
<td>CALCULAT</td>
<td>1453</td>
<td>06/25/76 11:53</td>
<td>08/25/76</td>
</tr>
<tr>
<td>COLCOMP</td>
<td>1339</td>
<td>08/09/76 11:11</td>
<td>08/09/76</td>
</tr>
<tr>
<td>BAS1FLR</td>
<td>5819</td>
<td>06/25/76 13:59</td>
<td>08/03/76</td>
</tr>
<tr>
<td>EXPENSES</td>
<td>9755</td>
<td>06/28/76 13:36</td>
<td>07/19/76</td>
</tr>
</tbody>
</table>

Hit [RETURN] to continue listing your worksheets:
And worksheet may be listed in descending order of the dates on which they were last accessed:

<table>
<thead>
<tr>
<th>FILENAME</th>
<th>SIZE</th>
<th>DATE MODIFIED</th>
<th>TIME MODIFIED</th>
<th>DATE LAST ACCESS</th>
</tr>
</thead>
<tbody>
<tr>
<td>DUMMY</td>
<td>563</td>
<td>06/25/76</td>
<td>09:07</td>
<td>12/29/76</td>
</tr>
<tr>
<td>BASIIPLR</td>
<td>5429</td>
<td>06/25/76</td>
<td>13:59</td>
<td>09/03/76</td>
</tr>
<tr>
<td>JUNK</td>
<td>2965</td>
<td>07/23/76</td>
<td>09:51</td>
<td>09/01/76</td>
</tr>
<tr>
<td>TIMEDIFF</td>
<td>489</td>
<td>06/20/76</td>
<td>14:21</td>
<td>05/25/76</td>
</tr>
<tr>
<td>CALCULAT</td>
<td>1453</td>
<td>06/25/76</td>
<td>11:51</td>
<td>06/25/76</td>
</tr>
<tr>
<td>COLCOMP</td>
<td>1239</td>
<td>08/09/76</td>
<td>11:11</td>
<td>08/09/76</td>
</tr>
<tr>
<td>EXPENSES</td>
<td>9755</td>
<td>06/28/76</td>
<td>13:36</td>
<td>07/13/76</td>
</tr>
<tr>
<td>WGR1FLR</td>
<td>297</td>
<td>06/28/76</td>
<td>11:15</td>
<td>06/28/76</td>
</tr>
<tr>
<td>UTL1FLR</td>
<td>361</td>
<td>06/28/76</td>
<td>11:07</td>
<td>06/28/76</td>
</tr>
<tr>
<td>UNG1FLR</td>
<td>361</td>
<td>06/28/76</td>
<td>11:18</td>
<td>06/28/76</td>
</tr>
<tr>
<td>TM31FLR</td>
<td>489</td>
<td>06/28/76</td>
<td>11:02</td>
<td>06/28/76</td>
</tr>
<tr>
<td>REV1FLR</td>
<td>297</td>
<td>06/28/76</td>
<td>11:29</td>
<td>06/28/76</td>
</tr>
<tr>
<td>OTMT1FLR</td>
<td>297</td>
<td>06/28/76</td>
<td>11:27</td>
<td>06/28/76</td>
</tr>
<tr>
<td>CP01FLR</td>
<td>297</td>
<td>06/28/76</td>
<td>11:32</td>
<td>06/28/76</td>
</tr>
<tr>
<td>AN11FLR</td>
<td>1197</td>
<td>06/28/76</td>
<td>11:43</td>
<td>06/28/76</td>
</tr>
<tr>
<td>UNDA1FLR</td>
<td>297</td>
<td>06/25/76</td>
<td>11:42</td>
<td>06/25/76</td>
</tr>
<tr>
<td>UNDA2FLR</td>
<td>297</td>
<td>06/25/76</td>
<td>10:26</td>
<td>06/25/76</td>
</tr>
<tr>
<td>TCH41FLR</td>
<td>297</td>
<td>06/25/76</td>
<td>11:47</td>
<td>06/25/76</td>
</tr>
<tr>
<td>RET1FLR</td>
<td>619</td>
<td>06/25/76</td>
<td>11:17</td>
<td>06/25/76</td>
</tr>
<tr>
<td>RES1FLR</td>
<td>361</td>
<td>06/25/76</td>
<td>11:37</td>
<td>06/25/76</td>
</tr>
<tr>
<td>OTGS1FLR</td>
<td>297</td>
<td>06/25/76</td>
<td>11:02</td>
<td>06/25/76</td>
</tr>
<tr>
<td>MTNL1FLR</td>
<td>619</td>
<td>06/25/76</td>
<td>09:21</td>
<td>06/25/76</td>
</tr>
<tr>
<td>MHDI1FLR</td>
<td>297</td>
<td>06/25/76</td>
<td>09:16</td>
<td>06/25/76</td>
</tr>
<tr>
<td>FHD11FLR</td>
<td>619</td>
<td>06/25/76</td>
<td>09:25</td>
<td>06/25/76</td>
</tr>
<tr>
<td>EN11FLR</td>
<td>499</td>
<td>06/25/76</td>
<td>11:07</td>
<td>06/25/76</td>
</tr>
<tr>
<td>DES1FLR</td>
<td>425</td>
<td>06/25/76</td>
<td>10:45</td>
<td>06/25/76</td>
</tr>
<tr>
<td>CPUI1FLR</td>
<td>297</td>
<td>06/25/76</td>
<td>11:50</td>
<td>06/25/76</td>
</tr>
<tr>
<td>CON1FLR</td>
<td>297</td>
<td>06/25/76</td>
<td>11:30</td>
<td>06/25/76</td>
</tr>
<tr>
<td>CHR11FLR</td>
<td>297</td>
<td>06/25/76</td>
<td>11:12</td>
<td>06/25/76</td>
</tr>
<tr>
<td>ACC1FLR</td>
<td>361</td>
<td>06/25/76</td>
<td>11:23</td>
<td>06/25/76</td>
</tr>
</tbody>
</table>

HIT (RETURN) TO CONTINUE LISTING YOUR WORKSHEETS:

The use of worksheet, option 9, can cause the dates of last access to be misleading because it accesses worksheets to obtain information about them and thereby updates the dates of last access.

OPTION 2: LIST SUBSETS

This option is basically the same as option 1 of files, except that subsets are listed. Refer to the description of files, option 1.

OPTION 3: LIST RESPONSE FILES

This option is basically the same as option 1 of files, except that Response Files are listed. Refer to the description of files, option 1.

OPTION 4: LIST UTILITY FILES

This option provides for listing information about utility files, i.e., all system files except worksheets, subsets, and Response Files. Utility files listed are
- Communication Files, called COMMF1, COMMF2, etc.
- The Patient File, called PATIENT
- Three Schema Files, called SCHEMA.AM, SCHEMA.AT, and SCHEMA.C
- The SDF, called PSDF (Permanent Study Data File)
- The Update File, called UPDATE

An example of a list of Utility Files ordered by file name is shown below.

### files.4  ASCENDING LIST OF UTILITY FILES ORDERED BY FILENAME

<table>
<thead>
<tr>
<th>FILENAME</th>
<th>SIZE IN BYTES</th>
<th>DATE LAST MODIFIED</th>
<th>DATE LAST ACCESSED</th>
</tr>
</thead>
<tbody>
<tr>
<td>COMMF1</td>
<td>4333</td>
<td>04/06/76 09:37</td>
<td>04/06/76</td>
</tr>
<tr>
<td>COMMF2</td>
<td>1883</td>
<td>03/30/76 09:18</td>
<td>03/30/76</td>
</tr>
<tr>
<td>COMMF3</td>
<td>68</td>
<td>03/19/76 09:59</td>
<td>03/19/76</td>
</tr>
<tr>
<td>COMMF5</td>
<td>204</td>
<td>04/01/76 09:17</td>
<td>04/01/76</td>
</tr>
<tr>
<td>COMMF6</td>
<td>288</td>
<td>08/17/76 10:10</td>
<td>08/17/76</td>
</tr>
<tr>
<td>COMMF7</td>
<td>3100</td>
<td>12/01/76 15:00</td>
<td>12/01/76</td>
</tr>
<tr>
<td>PATIENT</td>
<td>1024</td>
<td>12/08/76 09:46</td>
<td>12/09/76</td>
</tr>
<tr>
<td>PSDF</td>
<td>14336</td>
<td>12/28/76 09:46</td>
<td>12/29/76</td>
</tr>
<tr>
<td>SCHEMA.AM</td>
<td>4096</td>
<td>12/13/75 16:00</td>
<td>12/16/76</td>
</tr>
<tr>
<td>SCHEMA.AT</td>
<td>37988</td>
<td>12/13/75 16:00</td>
<td>12/16/76</td>
</tr>
<tr>
<td>SCHEMA.C</td>
<td>1024</td>
<td>12/13/75 16:00</td>
<td>12/16/76</td>
</tr>
<tr>
<td>UPDATE</td>
<td>88</td>
<td>12/29/76 10:02</td>
<td>12/29/76</td>
</tr>
</tbody>
</table>

END OF ENTRIES.  HIT (RETURN) TO RESPECIFY ORDER OPTIONS:

Except for listing utility files instead of worksheets, this option is basically the same as option 1 of files. For details about its function, refer to the description of files, option 1.
GOODBYE terminates the CLINFO session. If the user has an Update File it allows him to leave a request for the system manager to merge the data into the SDF. The user should use the messages activity to send other messages to the system manager prior to typing 'goodbye'. The merge request and other messages are left in a special file that the system manager reads at least once a day.

Dialog

At any system request for a response, the user may enter 'goodbye'. Unless the user has an Update File, his CLINFO session is immediately terminated. If he does have an Update File, CLINFO asks if he wants to leave a data merge request, then terminates the session.

*** THE UPDATE FILE CONTAINS 1192 CHARACTERS ***

TYPE yes TO REQUEST THE SYSTEM MANAGER TO MERGE THAT DATA WITH THE STUDY DATA
ELSE, HIT [RETURN]: yes

The following is an example of a final message which appears as the terminal is inactivated after typing 'goodbye'.

08/09/75  14:23  SIGN-OFF,  05  
08/09/75  14:23  CPU-USED,   1  
08/09/75  14:24  I/O-USED,  13  

CLINFO RELEASE 3 AT YOUR SERVICE

The first line indicates that the terminal is being signed off. The second indicates the central processor time used, in tenths of a second. The third line shows the number of input/output statements issued during the terminal session.

The last line indicates that the terminal is inactive and that the user may sign on, using the CLINFO and ESC keys.

Special note for dial-up users: When using the system remotely over a dial-up line, the CLINFO AT YOUR SERVICE message is garbled in the last several characters, and the telephone connection is broken automatically by the system at logoff. Thus it is necessary to redial in order to log on to the system again.
The *message* activity allows the user to leave requests and other messages for the system manager and to record comments about his experience with the prototype. Messages are stored in a special file that the system manager reads at least once a day. They are acted upon as necessary.

**Dialog**

In response to any prompt, the user may enter `!message` or `!mes`. The system prompts for messages and the user may type several lines, each followed by hitting the RETURN key. After entering his messages, the user may type `!done` on a new line to obtain the list of activities or he may type `!activity-name` (e.g., `!analyze` or `!ana`) or `!activity name, subactivity number` (e.g., `!ana, 1`) to go directly to an activity or a subactivity.

**TYPE YOUR MESSAGE. WHEN DONE, TYPE !activity-name OR !done ON A NEW LINE**

MESSAGE:

Tommy--I don't understand some of the results produced by the regression function. Can we get together on it Thurs afternoon?

John--Thanks for getting Rand to include thru-the-origin regression; now how about non-linear regression?

!ana, 4
The response files activity allows the user to create, edit, execute, discard, copy, and append Response Files, i.e., lists of responses to CLINFO prompts. Response Files simplify the execution of procedures that are performed repeatedly with only minor variations. A Response File is created for a particular procedure; when it is executed, a specified subactivity is initiated and the system then goes on to display prompts and information in the usual manner. However, the user responds to only those prompts that are necessary to provide for the variations; fixed responses are read from the Response File for the fixed part of the procedure. Response Files are useful for:

- Entering data in a well-defined sequence; e.g., the Response File initiates prompts for changing the patient or time and prompts for items in particular panels, and the user provides the values.
- Retrieving particular kinds of data and reorganizing them; e.g., the Response File indicates that data are to be retrieved for a patient, the user enters the patient abbreviation, then the Response File provides the time specification and item names and sorts the retrieved worksheet.
- Copying portions of a series of worksheets into a single worksheet; the user provides the source worksheet names and destination starting rows, and the Response File provides all the other information.
- Repeatedly performing row/column calculations; the user specifies the result column and arithmetic expression, but the Response File contains all of the other required specifications.

A Response File is created by example, i.e., by entering a sequence of responses (which are stored in the file) while using the CLINFO system in the normal manner. It is then edited to change, add, or delete prompts; to indicate that a response is to be provided by the user instead of the file; or to specify cycling from the end of the file back to some previous response. Edit is the only response files subactivity that requires any explicit interaction with the Response File. Figures I-11, I-12, and I-13 in Sec. I illustrate the creation, editing, and execution of Response Files.

Dialog

Response files displays the menu of subactivities and prompts for an option:

RESPONSE FILE OPTIONS

<table>
<thead>
<tr>
<th>TYPE TO</th>
</tr>
</thead>
<tbody>
<tr>
<td>1..EXECUTE A RESPONSE FILE</td>
</tr>
<tr>
<td>2..CREATE A RESPONSE FILE</td>
</tr>
<tr>
<td>3..EDIT A RESPONSE FILE</td>
</tr>
<tr>
<td>4..DISCARD A RESPONSE FILE</td>
</tr>
<tr>
<td>5..COPY A RESPONSE FILE</td>
</tr>
<tr>
<td>6..APPEND A RESPONSE FILE</td>
</tr>
</tbody>
</table>

RESPONSE FILE OPTION:

A response causes the appropriate subactivity to be invoked.
OPTION 1: EXECUTE A RESPONSE FILE

To execute a Response File, the user specifies it by name. The system then proceeds to read some responses from the file and to prompt the user for others. The user is prompted whenever a question mark appears as a response in the Response File.

The Response File remains in control until terminated by the user or by an error condition. The user terminates it by

- Typing ! without any following text.
- Typing /response.
- Simultaneously hitting the CLINFO and ESC keys.

A terminating error condition is one that occurs when a response read from the Response File is unacceptable. This would happen if the Response File were edited incorrectly or if a user response resulted in more or fewer prompts than the Response File was designed to accommodate. Unexpected prompts would occur, for example, if there were no current worksheet when one was expected, if the Response File was designed assuming one path through an activity but another was taken, or if the user transferred unexpectedly to another activity while a Response File was in control. In such cases, the Response File is not terminated immediately, but when the first unacceptable response occurs. When a Response File is terminated, the user may proceed normally but must enter all required responses.

Dialog

The following example illustrates the execution of a Response File that was created as shown below in option 2 and edited as discussed below in option 3.

The user enters the name of the Response File as

RESPONSE FILE NAME: padata

and the interaction then proceeds as shown below, where the [SYSTEM] and [USER] annotations indicate whether the system, i.e., the Response File, or the user provided the response.

PLEASE TYPE YOUR INITIALS: gfg

[SYSTEM]

ADDING NEW DATA FOR

PATIENT ABBREV: andy g__

[USER]

DATE: 7/15/75___

[USER]

TIME: 1215

[USER]

ENTER panel_name OR item_name:item_value OR context OR patient:patient_abbr OR date:date_value OR time:time_value.

? History

[SYSTEM]

case no.: 9

[USER]

age: 24

[USER]

sex: f

[USER]

bdi: wt: 87

[USER]

87 IS OUT OF RANGE.
IF OK, TYPE YOUR INITIALS, ELSE HIT [RETURN]: gfg  [USER]
dur div: 15  [USER]
ins req : 38  [USER]
? context
ADDDING NEW DATA FOR
PATIENT ABBREV(andy g ):  [SYSTEM]
DATE(7/15/1975): ____  [USER]
TIME(1215): 1530  [USER]
? control
insuline:  [USER]
glucose: 205  [USER]
glucone: 187  [USER]
alanine: .36  [USER]
lactate: 1.7  [USER]
? context
ADDDING NEW DATA FOR
PATIENT ABBREV(andy g ): case 3  [USER]
DATE(7/15/1975): ____  [USER]
TIME(1510): 915  [USER]
? History
case no.:  

The Response File provides all the responses that result in possible context changes and in prompts for item values. The user specifies the context and supplies the item values. Each response that the user enters corresponds to the response ? in the Response File.

An out-of-range value that was entered for %idwt was accepted without terminating the Response File even though it resulted in an extra prompt, because Response File handling assumes that one extra prompt will occur whenever the user makes what is treated as an error.

Except for the response to the out-of-range prompt, all of the responses from the first gfg to the last context correspond one-to-one with question marks or other responses in the Response File called ptdata. Subsequent responses result from cycling back to the prompt in the Response File that follows gfg.

OPTION 2: CREATE A RESPONSE FILE

A Response File is created by providing a name for the file and then using the system in the normal manner. The system records the user's responses and stores them in the named Response File. Because the Response File must usually be edited using option 3, it is not essential that each response correspond to what is
ultimately stored in the file. Some responses that must be entered in order to proceed with file creation will be changed to ? to indicate that they are to be entered by the user during Response File execution. It is important, however, to take the same paths that are to be taken when the Response File is executed, in order to minimize file editing.

Response File creation is terminated by

- Simultaneously hitting the CLINFO and ESC keys.
- Typing ' without any following text.
- Typing $response (or $res,3 to go directly to Edit).

In the last case, the characters $response will be included in the file and must be deleted. The activity name,option number notation can be used within a Response File in order to go from one activity to another during the Response File execution; in fact, this is recommended.

User errors, i.e., responses that result in reprompts, are not stored in the Response File by this subactivity.

**Dialog**

This example illustrates the creation of a Response File that controls data entry. Its editing is in the detailed discussion of option 3, and its execution, in the discussion of option 1.

The name of the new Response File is entered, the first activity is specified, and the dialog then proceeds as if that activity (and subsequent activities) were being used directly. In this case, Response File creation is terminated by typing $res,3.

**RESPONSE FILE NAME:** ptdata

**TYPE THE FIRST ACTIVITY (E.G. $analyze,1):** $ent,1

**TO TERMINATE THE RESPONSE FILE, TYPE ! OR $response OR HIT [CTRL][ESC]**

**PLEASE TYPE YOUR INITIALS:** gfg

**ADDING NEW DATA FOR**

**PATIENT ABBREV:** case 1

**DATE:** 3/22/76

**TIME:** 1500

**ENTER panel_name OR item_name:item_value OR context OR**

patient:patient_abbr OR date:date_value OR time:time_value.

? History

* case no.:

* age :

* sex :

* idl wt :

* dur diab:

* ins req :

? control
Here the values entered for items are unimportant because they will be replaced in the Response File by question marks to indicate that they are to be provided by the user during Response File execution. It is important, however, to provide some acceptable response to each prompt and to enter responses, such as History, which dictate the number of following prompts and which are to remain in the Response File.

This example is particularly simple because it involves only a single activity. In general, transfers to activities (except response files itself) may be included in Response files by entering appropriate responses (e.g., \texttt{\textbackslash work,1}) during file creation.

**OPTION 3: EDIT A RESPONSE FILE**

This subactivity provides for adding, changing, or deleting responses in a Response File; for listing the file on the user's terminal; for printing it on the system printer; and for specifying a cycle from the end of the Response File to a particular response. One reason for editing a Response File is to change a specific response that was entered during file creation to a question mark. Whereas most responses in the file will be interpreted by the system as if they were entered directly by the user, the question mark signifies that the system should prompt the user in the usual fashion and should use his response instead of reading a response from the file. Another reason for editing a Response File is to provide for repeated operations \textit{within} the file. A cycle may be established such that after reaching the end of the file, Response File execution continues from a specified response.

Data entry with \texttt{enter}, option 1, is facilitated by including extended question-mark-type responses in the Response File. When using \texttt{enter}, option 1, the data enterer may respond to question mark prompts by indicating what he is entering (by typing \texttt{patient, date, time} or an item name followed by \texttt{?}) and then providing a value. To minimize typing during data entry, a response such as

\begin{verbatim}
patient:?
date:?
time:?
item-name:?
\end{verbatim}

may be included in the Response File. When ooe of these is encountered during Response File execution, CLINFO displays the response up to the question mark (e.g., it displays \texttt{patient}) and expects the data enterer to enter the data value (e.g., a patient abbreviation).
After indicating which Response File is to be edited, all editing is effected by responding to the prompt

COMMANDS: print, cycle to #, list #, delete #, change #, add #, ? #

COMMAND:

The user responds by typing two or more letters of one of the listed commands (e.g., change or ch), followed (except when the command is print) by the number of a response in the file. For example, he responds ch 3 to change prompt 3. When change # or add # is entered, he is prompted for the new response.

The editing commands have the following meanings:

- list or list # (where # is a response number): Lists up to 20 responses starting with response number #; lists responses starting with the first response if # is not specified or refers to a nonexistent response number.
- add # (where # is in the form x.xxx and there is no response number #): Adds a new response with number #, where # may be before, after, or between existing responses. The user is prompted with

RESPONSE:

and enters the new response. The response ? means that the Response File user is to enter the actual response during Response File execution. The new response does not appear on the screen until the list command is executed.

- delete # (where # is a response number): Deletes response number #. The change does not appear until the list command is executed.
- change # (where # is a response number): Deletes the existing response number # and adds a new response with the same number. The user is prompted with

RESPONSE:

and enters the new response. The change is not displayed until the list command is executed.

- ? # (where # is a response number): Deletes the existing response number and replaces it with the response ?. The response ? means that the Response File user is to enter the actual response when the Response File is executed. This command is equivalent to entering change # and then responding ? to RESPONSE. The change is not displayed until the list command is executed.

- print: Prints the entire Response File on the system printer.
- cycle to # (where # is a response number): Creates a cycle from the end of the Response File to the specified response. During execution, the response specified here is used after the last one in the file is used; the file continues to execute and cycle until it is terminated. The person editing the file should be careful to ensure that the specified "cycle to" response and subsequent responses have the intended meaning when they follow the last one. During editing, the terminal screen always indicates which, if any, response is cycled to. The cycle response can be changed by entering
cycle #, where # is the new cycle response; the file can be caused not to
cycle by entering cycle n, where n is not an existing response number.

Whenever the user leaves this subactivity, the current state of the Response File
is automatically saved. The response numbers have no utility outside this subactivity,
however, so they are not maintained. Instead, when this subactivity is invoked,
each response in sequence is given an integer number. One consequence of this is
that after adding responses with fractional numbers and deleting responses, the
user may "clean up" the numbers by invoking the subactivity. Another is that the
response numbers on a printer listing may not reflect the current numbers even if
there have been no additions or deletions. The system handles cycles properly, i.e.,
the file will continue to cycle to the previously specified response even if its number
has changed.

A Response File that has been edited should always be executed to test it
before it is used operationally.

Dialog

In this example, we shall edit the Response File created in the discussion of
option 2. The edited Response File was executed in the discussion of option 1.

As in other response file subactivities, the dialog begins with the entry of the
name of the Response File:

RESPONSE FILE NAME: ptdata

Next, the list command is entered to display the contents of the Response File on
the screen. The ptdata file has only 20 responses, so they are all displayed at once.

Response, 3 EDIT A RESPONSE FILE
RESPONSE FILE: ptdata THIS FILE DOES NOT CYCLE

<table>
<thead>
<tr>
<th>LINE</th>
<th>RESPONSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.00</td>
<td>!ent,1</td>
</tr>
<tr>
<td>2.00</td>
<td>gifs</td>
</tr>
<tr>
<td>3.00</td>
<td>case 1</td>
</tr>
<tr>
<td>4.00</td>
<td>3/22/76</td>
</tr>
<tr>
<td>5.00</td>
<td>1500</td>
</tr>
<tr>
<td>6.00</td>
<td>History</td>
</tr>
<tr>
<td>7.00</td>
<td></td>
</tr>
<tr>
<td>8.00</td>
<td></td>
</tr>
<tr>
<td>9.00</td>
<td></td>
</tr>
<tr>
<td>10.00</td>
<td></td>
</tr>
<tr>
<td>11.00</td>
<td></td>
</tr>
<tr>
<td>12.00</td>
<td></td>
</tr>
<tr>
<td>13.00</td>
<td>control</td>
</tr>
<tr>
<td>14.00</td>
<td></td>
</tr>
<tr>
<td>15.00</td>
<td></td>
</tr>
<tr>
<td>16.00</td>
<td></td>
</tr>
<tr>
<td>17.00</td>
<td></td>
</tr>
<tr>
<td>18.00</td>
<td>!res,3</td>
</tr>
<tr>
<td>19.00</td>
<td></td>
</tr>
<tr>
<td>20.00</td>
<td>yes</td>
</tr>
</tbody>
</table>

COMMANDS: print, cycle to #, list #, delete #, change #, add #, ?, #
COMMAND: ? 3
To change the Response File to allow the context (patient, date, and time) to be specified and item values to be entered, existing responses numbered 3 through 5, 7 through 12, and 14 through 18 are changed to question marks. The first such change is shown above.

The last two responses in the file should not be included (they were stored when Response File execution was terminated), so they are deleted. The deletion of 'res,3 after the changes from fixed responses to ? were made and the changed file was listed is shown below:

```
response,3 EDIT A RESPONSE FILE
RESPONSE FILE: ptdata    THIS FILE DOES NOT CYCLE

LINE #    RESPONSE
  1.00    'tent,1
  2.00    gig
  3.00    ?
  4.00    ?
  5.00    ?
  6.00    History
  7.00    ?
  8.00    ?
  9.00    ?
 10.00    ?
 11.00    ?
 12.00    ?
 13.00    control
 14.00    ?
 15.00    ?
 16.00    ?
 17.00    ?
 18.00    ?
 19.00    'res,3
 20.00    yes

COMMANDS: print, cycle to $, list $, delete $, change $, add $, ? $
COMMAND: del 19.
```

Because the Response File user, i.e., the data enterer, may wish to change the data collection date or time prior to entering the control panel, additional responses are inserted between response 12 and response 13. The added responses 12.1 through 12.4 are listed below. The file now has more than 20 lines, so the last few are not displayed when the list command is entered.
The *context* response (on line 12.10) initiates prompts for patient, date, and time. The [RETURN] response, which is shown as a blank in line 12.20, means that the patient abbreviation will not be changed at this point. The next two responses are ? to allow the Response File user to change the date and time.

The Response File is completed by editing it to cause it to prompt for another patient after data are entered for the first patient. Once data entry has begun, i.e., after data have been entered for the first patient, data for successive patients are entered by responding *context* and then responding to the context (i.e., patient, date, and time) prompts. Because responses 3 through 5 are responses to the context prompts, data entry can be continued by including *context* at the end of the Response File, then cycling to response 3. The context response is, therefore, added to the file as response 19. Finally, editing is completed by entering *cycle 3* as the command. This means that after *context* is used as a response, the responses starting with line 3 will be used. After adding response 19 and entering *cycle 3* and then *list 17*, the following is displayed on the screen:
If instead of providing for the entry of the entire History panel, only values for
sex, age, and %idl wt are to be entered, responses 6 through 12 can be replaced by
sex:?
age:?
%idl wt:?

During Response File execution the data enterer is prompted with the item name
(e.g., sex:) and enters the value. Similarly, to provide for changing the data collection
date and time, but not the patient abbreviation, responses 12.10 through 12.40
can be replaced by
date:?
time:?

OPTION 4: DISCARD A RESPONSE FILE

A Response File is discarded when a user invokes this subactivity, enters the
Response File's name, and then verifies that this is his intent.

Dialog

After the user enters a Response File name, the system repeats this name in
a prompt. The user types yes to discard the named file. If he does not want to discard
the file named in the prompt, he may hit the RETURN key. Upon completion of this
subactivity, the response files menu is displayed.

RESPONSE FILE NAME: ptdata

ptdata IS THE FILE NAME (IT MAY HAVE BEEN TRUNCATED)

TYPE yes TO DISCARD RESPONSE FILE ptdata :yes

OPTION 5: COPY A RESPONSE FILE

This option allows the user to copy an existing Response File and to name the
copy. The copy can be used like any other Response File, e.g., it can be edited (with
response, option 3) to create a variation of the original. The original, or source,
Response File is left unchanged.

Dialog

The user enters the name of the source Response File and then the name of the
copy. In this example the user makes a copy of Response File ptdata and names
the copy ptdata1.
SOURCE RESPONSE FILE NAME: ptdata

COPY RESPONSE FILE ptdata AND CALL IT
RESPONSE FILE NAME: ptdata1
ptdata IS THE FILE NAME (IT MAY HAVE BEEN TRUNCATED)
ptdata1 IS THE FILE NAME (IT MAY HAVE BEEN TRUNCATED)

OPTION 6: APPEND A RESPONSE FILE TO ANOTHER

This option appends a copy of one Response File to an existing Response File to lengthen the latter. As a result, the lengthened Response File no longer remains in its original form. The use of this option avoids the necessity of creating a large Response File in a single step with response, option 2. It is also useful for performing a fixed series of repetitive tasks for which individual Response Files have already been created.

Dialog

The user is asked to supply the name of the Response File that is to be copied and appended, and the name of the Response File that is to be lengthened by the addition of the copy. The latter Response File is modified and no longer remains in its original form. The user may then edit, execute, or copy the lengthened Response File, depending on his needs.

In this example, the user indicates that he wants to append a copy of Response File analife to the end of Response File retlife.

NAME OF RESPONSE FILE TO BE APPENDED TO ANOTHER: analife
APPEND A COPY OF RESPONSE FILE analife TO RESPONSE FILE: retlife
analife IS THE FILE NAME (IT MAY HAVE BEEN TRUNCATED)
retlife IS THE FILE NAME (IT MAY HAVE BEEN TRUNCATED)
TYPE yes TO EDIT THE RESPONSE FILE retlife, ELSE HIT [RETURN]
64 LINES FROM FILE analife APPENDED TO 37 LINES OF FILE retlife

Typing yes in response to the last prompt is similar to typing .3 (to go to response, option 3) followed by the name of the lengthened Response File.

A Response File may contain no more than 500 responses. Thus, when one Response File is appended to another, the responses that are copied (from the beginning of the Response File that is being appended) may be limited to those that make the lengthened Response File 500 responses long. If Response File retlife in the above example already contained 490 responses, the user would get the message

10 LINES FROM FILE analife APPENDED TO 490 LINES OF FILE retlife

even though Response File analife contains 64 responses. Furthermore, when the user edits a Response File that has been truncated to 500 responses, he is notified that this has occurred.
The *retrieve* activity is used to copy data values from various sources, namely, from the SDF, from a worksheet, or from a Communication File; it then either stores those values in a worksheet or in a Communication File, or it displays them at a user terminal.

In order for data values to be retrieved from the SDF and organized, they must be aligned in time. Clinical data are seldom collected simultaneously (an exception being data values collected as a panel in the CLINFO sense). However, it is not always necessary to align items with great precision. For example, the fact that two different data item values are collected on the same day may be sufficient. On the other hand, coarse time intervals may include several occurrences of a data item, so it is necessary to choose among them to refer to a single value.

*Retrieve* options 2 and 3 allow the user to deal with these requirements by

- Specifying the time unit that is used for alignment, viz., minute, hour, day, week, month, year, or multiple units.
- Specifying what time interval to use for data selection, viz., the time interval between two dates (and times) or events (including the system-defined events `begin` and `end`), a list of time units following a date or event, or a range of time units following a date or event.
- Specifying which of the possible multiple instances within a time unit or interval to use, viz., the first, last, maximum, minimum, mean, or all instances.
- Specifying "retrieve all data" and thereby obtaining either
  - the *only* set of data values or
  - one panel or group of items each day from the first day to the last day.

As noted, the user can cause a time interval to start at a specified calendar date and clock time or at the time of occurrence of a particular event. Recall that an event, in concept, is a time marker placed in a patient's data file. That marker singles out a patient-specific date and time that can be used as a reference point for the data retrieval process.

**Dialog**

A number of subactivities are available within the *retrieve* activity. These subactivities are presented as a menu and the user chooses among them by typing an option number. The menu has the form:
The response causes the appropriate subactivity to be invoked. Each subactivity is described below.

Note that it is possible to copy data into the SDF from a worksheet via option 2 of the enter activity, where the entries are checked and possibly encoded.

OPTION 0: SELECT A WORKSHEET

This subactivity performs the same function as worksheet, option 0, and the dialog is identical. It is included here for the convenience of the user who wants to select a worksheet upon entering the retrieve activity or who wants to change the current worksheet without leaving retrieve. Also, the system takes the user directly to this subactivity if he attempts to invoke some other retrieve subactivity when there is no current worksheet (for example, if he attempts to copy a portion of a worksheet into another worksheet). In that case, the system takes him to his initially specified subactivity after he selects a worksheet.

OPTION 1: EVENT DISPLAY

This subactivity displays information about events for individual patients, such as the dates and times of the first and most recent data collection for a patient, when a drug is given, and the specific value of an item associated with the occurrence of an event.

The user enters the patient abbreviation of interest, and in response, the system displays the event information and prompts for the next patient. The information displayed includes the name, type, date, time, item, and item value associated with each user-defined and system-defined event.

Dialog

In this example the user enters the patient abbreviation case 1. Each event defined in the schema is listed, as are the system-defined events begin (associated with the earliest time of data sample) and end (associated with the most recent time of data sample). Displayed column headings have the following meanings:
EVENT refers to the name of an event.
DATE and TIME define the time associated with an event; i.e., the time at which the event was triggered. This is the same as the time marked by the event for retrieval purposes.
ITEM is the name of the item that triggers the event when it takes on its first (in time) value, last value, maximum (in magnitude) value, minimum value, or first or last instance of a specific value.
TYPE is FIRST, LAST, MAX, or MIN.
VALUE is the item’s value at the indicated time.

<table>
<thead>
<tr>
<th>EVENT</th>
<th>TYPE</th>
<th>DATE</th>
<th>TIME</th>
<th>ITEM</th>
<th>VALUE</th>
</tr>
</thead>
<tbody>
<tr>
<td>begin</td>
<td></td>
<td>5/ 3/1975</td>
<td>00:00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>instopc</td>
<td>LAST</td>
<td>5/ 4/1975</td>
<td>08:10</td>
<td>insulin</td>
<td>35</td>
</tr>
<tr>
<td>instopt</td>
<td>LAST</td>
<td>5/ 5/1975</td>
<td>09:00</td>
<td>insulin</td>
<td>35</td>
</tr>
<tr>
<td>end</td>
<td></td>
<td>5/ 5/1975</td>
<td>09:00</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

If no data are stored in the SDF for the item associated with an event, no time is associated with the event. This is indicated by *UNK* in the DATE and TIME columns and by ... (3 periods) in the VALUE column. The events called begin and end have a date and time associated with them as soon as any data are stored for the patient.

OPTION 2: STUDY DATA FILE TO WORKSHEET

This subactivity provides for retrieving numeric data about a specific patient, item, or time from the investigator’s primary data base (the SDF) and for storing them in a new worksheet for examination and analysis. A time interval (relative to an event or an absolute calendar date and clock time) and a time unit (which determines the resolution of the retrieved data) may be specified. Retrieved values may be restricted to those that satisfy specified conditions.

Because the SDF is the major data repository, but data must be in a worksheet in order to be analyzed, this subactivity is at the heart of the CLINFO system. It can be used to retrieve data and organize worksheets in many different ways and is thus one of the most flexible and powerful, but complex, subactivities. This subactivity must be thoroughly understood in order to take full advantage of the CLINFO system.

The numeric data to be retrieved (i.e., numbers, codes, dates, and times) are those for particular, specified patients, items (i.e., variables), and times. Each of these three data “dimensions” may be used to define: (1) the type of worksheet to be constructed, (2) the worksheet rows, and (3) the worksheet columns.

If data are retrieved that pertain to a particular patient, then worksheet rows correspond to different items and worksheet columns must correspond to different times; alternatively, if rows correspond to times, then columns must correspond to items.
• If SDF data about a particular item are retrieved, then rows (or columns) are associated with different patients and columns (or rows) are associated with different data collection times.
• Finally, if the retrieved worksheet applies to a specific time, then its rows must correspond to different patients and its columns must correspond to different items, or vice versa.

The rows and columns of the retrieved worksheet are automatically labeled by the system with the names of the appropriate patients, items, and times.

Specifying the Type of Worksheet To Be Retrieved

To specify a retrieval, the user first indicates the dimension and the specific value of that dimension that apply to all the retrieved data; the value may be a patient abbreviation, an item name, or a time specification. To specify a patient, the user enters the abbreviation of a patient for whom data have already been entered. To specify an item, the user enters the name of an item that is defined in the schema. The time specification is more complex.

The Single Time for a Patients by Items Worksheet. When the retrieved data in a worksheet are said to be measured at a single time (i.e., a patient versus item worksheet), this time is a single, relative time of data collection for all of the specified patients and items. This time is defined by selecting an instance (i.e., a data value occurrence) within a specified time interval. As an example, the user may wish to retrieve one of the following for a list of patients:

• The maximum value measured during the study for each of a list of items,
• The first value of each of a list of items that may have been measured on the 5th, 12th, or 19th day following a transplant, or
• The mean (over the time interval) values of a list of items measured between 5 hours before noon on January 23, 1975, following the injection of a drug, and 10 hours afterward.

The time is specified by first selecting a time unit, next defining a time interval, and then selecting one data value for the entire time interval. The unit (or subinterval) choice may be minute, hour, day, 7-day week, 30-day month, 365-day year, or multiple units; the unit is specified first to simplify the dialog. The user may also specify "retrieve all data" and obtain the only (or the first, in case of multiple instances) value of each item for each patient. Selection of "retrieve all data" avoids the requirement for further time specification. The multiple-unit option allows the user to define subintervals as integer multiples of time units, e.g., 3 days or 2 weeks.

The time interval may be defined as:
The time between two time markers (e.g., from begin to end),
A list of time units from a time marker (e.g., 5, 7, and 12 days from the event called transplant), or
A range of time units relative to a time marker (e.g., -5 through 10 hours from 1/23/75 at 1200).

The time markers may be the system-defined events (i.e., begin and end), user-defined events that are described in the schema, or calendar dates and clock times specified by the user.

The selection choice that results in a single data value for the entire time interval may be the maximum, minimum, first, last, or mean value. The choices maximum, minimum, and mean apply only to individual items and cannot be used when entire panels are to be retrieved.

Specifying Rows and Columns of the Retrieved Worksheet

The user specifies the worksheet rows and columns after specifying the individual patient, item, or time to which all of the retrieved values in the worksheet refer.

Patients. If the worksheet does not apply to a single patient, rows or columns (but not both) may be defined to contain data for either a list of patients or a named patient set. That is, each row (or column) will represent a single patient. The patient set may be the one called all, which includes all the patients in the study, or it might be a set created using the subset activity.

Items. Items that are to define rows or columns (but, again, not both) may be specified as

- A list of numeric items,
- All numeric items described in the schema,
- A list of numeric panels, or
- All numeric panels.

When individual items are selected, they appear in the selected order in the resultant worksheet; if all items are specified, they appear in the worksheet in the same order in which they appear in the schema. The data collection dates and times for the item values may optionally be retrieved and stored in the resultant worksheet. If they are retrieved, the worksheet will have a DATE column (or row) and a TIME column (or row) preceding a series of columns (or rows) for all the selected items in the same panel. The dates and times will be stored for each such series of items.

Similarly, panels appear in the resultant worksheet in the order in which they are selected or (if "all panels" is selected) in the same order as in the schema. When an instance of a panel is retrieved, the date and time of its collection, as well as values for all constituent items, are always stored in the resultant worksheet.

Times. The time specification for a retrieved worksheet that has its rows or columns associated with individual times is similar to that described above. Its components are
1. Time unit or subinterval—the time resolution of retrieved data.
2. Time interval—the time range of interest.
3. Selection choice—the single value retrieved in each time unit in case more than one is stored.

The user specifies the time interval over which he wants to retrieve data and divides this interval into time units, or subintervals, such that (usually) one data value lies in each subinterval. During the interaction, the time unit is specified before the time interval to simplify the dialog. The result is that a single value is retrieved for each time unit. In cases where more than one value is measured during a time unit (e.g., during the hour between 8 A.M. and 9 A.M.), the selection choice first, last, max, min, or mean is used to limit the retrieval to a single value.

Instead of providing a detailed specification, the user may select “retrieve all data” when prompted for a time unit. This is equivalent to specifying

- Time unit = day
- Time interval = event begin to event end
- Selection choice = all

It results in retrieving all data values measured each day for all days on the study. The resultant worksheet identifies the ith data value with #i as a row or column label.

**Time Units.** Here again, the possible time units are minute, hour, day, 7-day week, 30-day month, 365-day year, or integer multiples of these. The subinterval defined by a day-long time unit starts at 0001 (i.e., at one minute after midnight) on a calendar date and ends at 2400 (i.e., midnight) on that same date. An hour-long interval begins on the hour (e.g., at 8:00 A.M.) and ends at 59 minutes after the hour (e.g., at 8:59 A.M.). Longer subintervals (i.e., weeks, months, and years) begin at 0001 on the day that contains the specified event or date.

Because measurements are not made and samples are not drawn at precisely specified times, the time associated with a datum might not lie within the expected unit-long subinterval. Thus, if a measurement is scheduled for 8:30 A.M., it could be as much as 30 minutes early or 29 minutes late and still fall within the 8:00 A.M. to 8:59 A.M. subinterval, but one scheduled for 8:00 A.M. would fall outside the expected interval if it is measured even one minute early.

**Time Intervals.** The time interval may be defined as all the time units between two time markers (i.e., events or dates and times), a list of time units from a specified time marker, or a range of time units from a time marker. As before, allowable events may be begin, end, or any event described in the schema that has been assigned a value prior to retrieval. The time units specified in the list or range may be either negative or positive numbers. Negative values are convenient for specifying times prior to an event; e.g., to retrieve data for the most recent seven days, the user would specify — 6 through 0 days from the event called end.

When specifying the time interval, the user is asked if he wants to select only those time units containing data. If he responds affirmatively, the resultant worksheet contains rows (or columns) for only those units (e.g., hour 1, hour 3, hour 4) for which data values are stored in the SDF. If he declines, the resultant worksheet contains a row (or column) for every unit in the interval; e.g., even if no data were
recorded for hour 2, the resultant worksheet would contain an hour 2 row of all missing values. The affirmative choice results in a concise worksheet when data are collected irregularly over a long period of time; the negative choice results in a worksheet that makes missing values apparent.

**Selection Choices.** The selection choice is used to select a single value from several that might have been measured during a unit-long subinterval. Allowable choices are maximum, minimum, first, last, mean, and all values. If all values is selected, all values measured during the entire time interval are retrieved. This is useful for retrieving and examining all data regardless of data collection time, but it does not time-align data across patients. When all values are retrieved, the resulting worksheet has one column or row for each retrieved value, i.e., there would be columns or rows labeled #1, #2, #3, etc. At most, 199 such values may be retrieved for any patient or item. The other selection choices retrieve a single value for each time unit, so the resulting worksheet has one column or row for each unit-long subinterval in the entire time interval, e.g., it has columns or rows labeled DAY 1, DAY 2, DAY 3, etc. Labels of worksheets created using multiple units have the form INT #. All worksheet time-unit labels generated by this subactivity are recognized by Scatter Plot (analyze, option 7) for plotting purposes.

When data for several patients are retrieved and one of the first five selection choices is used (i.e., the choice is not "all values"), data for the patients are aligned by comparable relative times. For example, if the retrieved worksheet has the system-generated column label DAY 3, then the data values in that column correspond to those measured on the third day from the specified event (or absolute date and time) for each of the patients. However, when "all values" is selected, the corresponding worksheet columns or rows are labeled with instance numbers only, and although all data values in the column labeled #7 correspond to the seventh measurement of the specified item for the specified patient, they may not have comparable relative times because data may be collected at different rates for different patients.

All selection choices apply when items are listed or when all items are chosen, but the choices maximum, minimum, and mean are not applicable to the retrieval of entire panels. This is because the maximum (or minimum) values for the various items might occur in different panel instances and because mean implies the generation of a single panel that represents several.

**A Time Specification Example.** The process of time specification is illustrated in the diagram below, which represents a planned experiment where a series of samples were to be drawn from a particular patient on the hour; i.e., at 6:00 A.M., 7:00 A.M., etc. They were actually drawn (as indicated in the diagram by arrows labeled S1, S2, etc.) at 6:15 A.M., 7:00 A.M., 8:10 A.M., etc. The event of interest, i.e., the last insulin injection, was to occur at 8:00 A.M. in conjunction with the third sample, but it actually occurred at 8:10 A.M.
In this example, the investigator wants to retrieve the data for every data collection hour starting at two hours before the event, so he selects 1-hour subintervals m thru n from event X (or date, time). He then enters the relative starting and ending hours — 2 and 10, provides the event name, and indicates that only those intervals containing data are to be selected. Because the time unit is hour, the system will treat the event as if it occurred at 8:00 A.M., and the relative time "Hour 0" will be 8:00 A.M.

The investigator next chooses the data value he wants to retrieve for each hour-long subinterval. If he chooses first value, the system will retrieve the first values measured after 6:00 A.M., 7:00 A.M., 8:00 A.M., etc., and will find values for the sample labeled S1, S2, S3, S4, S5, S7, and S9. Values for samples S6 and S8 will not be retrieved. In the worksheet, the retrieved values will be labeled HR −2, HR −1, HR 0, HR 1, HR 2, HR 4, and HR 6. If the investigator chooses last value, the system will retrieve the last value measured during each hour-long subinterval and will find those for S1, S2, S3, S4, S6, S8, and S9. In this case, values for samples S5 and S7 will not be retrieved. These results are shown in the diagram below.

In the worksheet, the retrieved values will be labeled HR −2, HR −1, HR 0, HR 1, HR 2, HR 4, and HR 6, even though S6 is normally the HR 3 (11:00 A.M.) sample and S8 is normally the HR 5 (1:00 P.M.) sample. If all values is the selection choice, then values for all the samples will be retrieved; they will be labeled #1, #2, #3, etc., in the worksheet.

Knowing that samples are actually drawn a little before or a little after the hour, the investigator might choose to avoid problems by using fractional hour
numbers. For example, he might enter the hour numbers $-2.5, -1.5, -0.5, 0.5, 1.5$, etc., instead of the range $-2$ to $10$. The effects of this specification are illustrated in the following diagram.

When an event occurs at 8:10 A.M., the subintervals in this case would be those starting at 5:30 A.M., 6:30 A.M., 7:30 A.M., 8:30 A.M., 9:30 A.M., etc. In the diagram above, there is one sample in each of these subintervals; so regardless of whether the selection choice was first value, last value, or all values, all the values would be retrieved. In the worksheet, the retrieved values would be properly labeled HR $-2.5$, HR $-1.5$, HR $-0.5$, HR $1.5$, etc.

**Specifying Logical Conditions to Restrict the Values Retrieved**

When the retrieval specification (i.e., the specification of worksheet type and rows and columns) is otherwise complete, the user is given an opportunity to specify logical conditions that restrict the data values retrieved. Then the system will retrieve only data values from those panels in which these or other values satisfy the specified conditions. A condition is specified by naming an item (e.g., glucagon or sex) and an acceptable range or code (e.g., 175,200 or m), just as a condition is specified when using the subset activity (see subset, options 4 and 5, for examples).

If, for example, the user specifies that only the values associated with glucagon in the range 175 to 200 for the specified patient(s) and time(s) are to be retrieved, then all panels that contain the glucagon item (and that otherwise satisfy the retrieval specification) are examined, but only data from those in which the glucagon lies in the specified range are stored in the resultant worksheet. Thus, if glucagon is in the plasma panel and this panel is listed as one to be retrieved, only those plasma panel values corresponding to the patients and times for which glucagon is in the specified range would be retrieved. Similarly, if glucose and glucagon are in the same panel and an item-type worksheet for glucose is specified, the only glucose values stored in the resultant worksheet would be those from panels which also contain glucagon values in the specified range. Only conditions placed on an item that is in the same panel as one retrieved have an effect on the retrieval. Thus, if sex and glucose are not in the same panel, conditions placed on values of sex have no effect on which values of glucose are retrieved.
As noted earlier, this feature is used for retrieving those values that satisfy specified conditions or that are related to such values. It is also useful for creating lists of patients whose data values satisfy specified conditions and, at the same time, retrieving data for these patients. For example, a retrieval that calls for patients in worksheet rows and demographic items in columns, with the conditions $35 \leq \text{age} \leq 50$ and $\text{sex} = M$ specified, would result in a worksheet that lists the patients in this age and sex category as row labels and contains their demographic item values as elements. (There would be no rows for patients whose data do not satisfy the conditions.) The patients whose labels appear in the retrieved worksheet may formally be made members of a set, which can be referred to elsewhere by name, by using subset, option 7, and specifying a condition that is true for all the patients.

Retrieval

Because of the complexity of the retrieval specification, the user is presented with updated summaries of his specification as he proceeds. At any point, he may type ! to redefine the entire specification.

When the retrieval is completed, the system reports the dimensions of the new worksheet in which the values are to be stored, and the user names the worksheet and enters its title. He may then display the retrieved worksheet. Because a worksheet cannot contain more than about 2000 data values, it is possible to retrieve more data than can be stored in a single worksheet. When this happens, a worksheet is created which has all the specified rows but which may have fewer than the number of columns implied by the retrieval specification. A warning appears on the CRT, telling the user how many more columns are needed to contain all of the data. The user may then specify another retrieval starting with the first patient, item, or time that was not retrieved.

Example Retrieval Specifications and Resultant Worksheets

Some typical worksheet formats are shown below, with instructions for retrieving data into them. System prompts are in upper case; user responses are in lower case. Notes within sketches refer to content; they are not actual prompts and responses.
### TYPICAL WORKSHEET FORMATS

#### FORMAT OF WORKSHEET TO BE RETRIEVED

**USER SPECIFICATIONS**

All the Data Collected for Patient *case 1*:

<table>
<thead>
<tr>
<th>Instances</th>
<th>#1</th>
<th>#2</th>
<th>.</th>
<th>.</th>
<th>.</th>
<th>.</th>
</tr>
</thead>
<tbody>
<tr>
<td>DATE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TIME</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item 2</td>
<td>.</td>
<td>.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>.</td>
<td>.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All Data Values for Case 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Worksheet Type:** times vs items  
**Patient Abbrev:** case 1  
**Rows Are:** all panels  
**Unit Choice:** hour (unit does not matter when *Selection Choice* is *all*)  
**Time Interval Choice:** event begin to event end  
**Select Only Intervals With Data:** yes  
**Select Choice:** all  
**Restrict Panels:** no

#### Treatment Panel Data Collected for Patient 1534 on Each of the 7 Days Before 3/15/75:

<table>
<thead>
<tr>
<th>Treatment Panel</th>
<th>DATE</th>
<th>TIME</th>
<th>glucose</th>
<th>glucagon</th>
</tr>
</thead>
<tbody>
<tr>
<td>Day -6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Day -5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Daily Treatment Panel</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Values for Patient 1534</td>
<td></td>
</tr>
<tr>
<td>Day 0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Worksheet Type:** times vs items  
**Patient Abbrev:** 1534  
**Rows Are:** times  
**Unit Choice:** day  
**Time Interval Choice:** -6 through 0 days from 3/15/75 at 1200  
**Select Only Intervals With Data:** no  
**Select Choice:** first  
**Columns Are:** panels  
**Panels:** treatment  
**Restrict Panels:** no
For All Patients: Glucose Values Averaged Over Each 2-Hour Interval from the Time Insulin is Stopped:

<table>
<thead>
<tr>
<th>2-hour Intervals</th>
</tr>
</thead>
<tbody>
<tr>
<td>INT 0</td>
</tr>
</tbody>
</table>

- Patient 1
- Patient 2
- Glucose Values Averaged Over Each 2-hour Interval

**WORKSHEET TYPE:** times vs patients
**ITEM:** glucose
**ROWS ARE:** patient set
**SUBSET NAME:** all
**UNIT CHOICE:** multiple units
**UNIT TO BE MULTIPLIED:** hour
**NUMBER OF HOURS:** 2
**TIME INTERVAL CHOICE:** event instart to event end
**SELECT ONLY INTERVALS WITH DATA:** no
**SELECTION CHOICE:** mean value
**RESTRICT PANELS:** no

---

Same as Above but Retrieval Restricted to Glucose Values from Panels in which Glucose $\geq 200$ and $0.31 \leq$ Alanine $\leq 0.4$:

<table>
<thead>
<tr>
<th>2-hour Intervals</th>
</tr>
</thead>
<tbody>
<tr>
<td>INT 0</td>
</tr>
</tbody>
</table>

- Patient 1
- Patient 2
- Glucose Values from Selected Panels

**RESTRICT PANELS:** yes
**ANY OR ALL CONDITIONS:** all
**ITEM:** glucose
**RANGE:** 200,500
**ITEM:** alanine
**RANGE:** .31, .4
FORMAT OF WORKSHEET TO BE RETRIEVED

Demographic Data for Patients in the Set "old":

<table>
<thead>
<tr>
<th>age</th>
<th>sex</th>
<th>wt</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Patients in the Set "old":

Data Values that are Measured Only Once

WORKSHEET TYPE: patients vs items
UNIT CHOICE: retrieve all data
ROWS ARE: patient set
SUBSET NAME: old
COLUMNS ARE: items
DATE AND TIME REQUIRED: no
ITEMS: age, sex, wt
RESTRICT PANELS: no

For All Patients in the Study: Values of Glucose, Glucagon, Alanine, and Lactate
Selected from Those Panels Measured in the Fourth Hour
After Insulin is Stopped and in which Glucose ≥ 200:

<table>
<thead>
<tr>
<th>glucose</th>
<th>lactate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Patients with Glucose ≥ 200

Values Measured in Hour 4 After Stopping Insulin

WORKSHEET TYPE: patients vs items
UNIT CHOICE: hour
TIME INTERVAL CHOICE: list from event
SELECT ONLY INTERVALS WITH DATA: no
INTERVAL LIST: 1-hour interval 4
EVENT: instop
SELECTION CHOICE: first value
ROWS ARE: patient set
COLUMNS ARE: items
DATE AND TIME REQUIRED: no
ITEMS: glucose, glucagon, alanine, lactate
RESTRICT PANELS: yes
ANY OR ALL CONDITIONS: all
ITEM: glucose
RANGE: 200,500
Summary

The following table summarizes the specification choices:

<table>
<thead>
<tr>
<th>Retrieved data may refer to a Single</th>
<th>Which is Defined by:</th>
<th>And Yields:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient</td>
<td>a patient abbreviation</td>
<td>An Item x Time WS</td>
</tr>
<tr>
<td>Item</td>
<td>an item name</td>
<td>A Patient x Time WS</td>
</tr>
<tr>
<td>Time</td>
<td>a selection criterion operating over a time interval</td>
<td>A Patient x Item WS</td>
</tr>
</tbody>
</table>

ROWS or COLUMNS may represent Which are defined by specifying:

PATIENTS
- A list of patient abbreviations, or
- The name of a patient set

ITEMS (together with optional calendar dates and clock times)
- A list of items, or
- All items, or
- A list of panels, or
- All panels
- Retrieve all data, or
- A unit; minute, hour, day, week, month, year
- Over a time interval; defined by
  - Time-marker to time-marker
  - A list of unit numbers from a time-marker
  - A range of unit numbers from a time-marker
  - A selection; first, last, min, max, mean

Relative TIMES

1 If panels are used to define rows (or columns), the date and time of the panel will occupy two rows (or columns) of each panel.
2 Or integer multiples of such units.
3 Optionally only those units containing data are retrieved.
4 Time-markers may be: Events, e.g., begin (date and time of first data collection), end (time of last data collection), or events defined by the user in the schema; or dates and times.
5 Unit numbers may be any real numbers; e.g., -3 to precede a time marker, or 2.5, 3.5, etc., for nominal times on the hour.
6 Range must be less than 200 units.
7 Min, max, and mean cannot be applied to panels; first and last may give different results when applied to panels and items; all will result in instances which may not yield comparable relative times.
Dialog

To illustrate four different types of retrieval specifications, we shall describe the creation of three worksheets that appear in Sec. III: (1) CLINICAL, which contains the first values measured for a list of patients and items; (2) CASE1C, which contains data for the patient called case 1; and (3) GLUCOSE2C, which contains values of the item glucose measured for all patients at several times, and shall describe the creation of a worksheet like CASE1C but which includes only those values that satisfy specified conditions.

**Retrieval of a Worksheet for a Specific Time.** To retrieve worksheet CLINICAL, the user specifies that data are to be retrieved for "all" times while remembering that there is actually only a single data collection time. He does this by indicating that the worksheet type is patients vs items for a single time, and then selecting retrieve all data as the time unit choice. The interactions are shown below.

```plaintext
WORKSHEET TYPE CHOICE

<table>
<thead>
<tr>
<th>TYPE TO CREATE A WORKSHEET OF</th>
</tr>
</thead>
<tbody>
<tr>
<td>1...TIMES VS ITEMS (FOR A PATIENT)</td>
</tr>
<tr>
<td>2...TIMES VS PATIENTS (FOR AN ITEM)</td>
</tr>
<tr>
<td>3...PATIENTS VS ITEMS (FOR A TIME)</td>
</tr>
</tbody>
</table>

WORKSHEET TYPE: 3
```

```
retrieve,2 RETRIEVE DATA FROM STUDY FILE INTO WORKSHEETS

WORKSHEET IS : PATIENTS VS ITEMS (FOR A TIME)

A TIME INTERVAL WILL BE BROKEN DOWN INTO UNIT-LONG SUBINTERVALS, MAKE A UNIT CHOICE

<table>
<thead>
<tr>
<th>TYPE TO CHOOSE THE UNIT</th>
</tr>
</thead>
<tbody>
<tr>
<td>0...RETRIEVE ALL DATA</td>
</tr>
<tr>
<td>1...MINUTE</td>
</tr>
<tr>
<td>2...HOUR</td>
</tr>
<tr>
<td>3...DAY</td>
</tr>
<tr>
<td>4...7-DAY WEEK</td>
</tr>
<tr>
<td>5...30-DAY MONTH</td>
</tr>
<tr>
<td>6...365-DAY YEAR</td>
</tr>
<tr>
<td>7...MULTIPLE UNITS</td>
</tr>
</tbody>
</table>

UNIT CHOICE: 0
```

Worksheet rows are specified by choosing patient set as the type of specification, and naming all as the patient set of interest.
WORKSHEET IS: PATIENTS VS ITEMS (FOR A TIME)
1-day sub-int's FROM EVENT begin TO EVENT end
(FROM THE START OF THE day CONTAINING THE EVENT begin)
SELECTING THE FIRST MEASURED DURING ALL THE 1-day sub-int's

ROWS ARE
---------
<table>
<thead>
<tr>
<th>TYPE TO USE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.PATIENTS</td>
</tr>
<tr>
<td>2.PATIENT SET</td>
</tr>
<tr>
<td>3.ITEMS</td>
</tr>
<tr>
<td>4.PANELS</td>
</tr>
<tr>
<td>5.ALL ITEMS</td>
</tr>
<tr>
<td>6.ALL PANELS</td>
</tr>
</tbody>
</table>
---------

ROWS ARE: 2

Columns are specified by choosing items as the type of specification, then specifying the items called age, sex, %idl wt, dur diab, and ins req. Because the data collection dates and times are not of interest, these are not retrieved with the item values.

WORKSHEET IS: PATIENTS VS ITEMS (FOR A TIME)
1-day sub-int's FROM EVENT begin TO EVENT end
(FROM THE START OF THE day CONTAINING THE EVENT begin)
SELECTING THE FIRST MEASURED DURING ALL THE 1-day sub-int's

THE ROWS ARE: PATIENT SET (all )

COLUMNS ARE
-----------
<table>
<thead>
<tr>
<th>TYPE TO USE</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.ITEMS</td>
</tr>
<tr>
<td>4.PANELS</td>
</tr>
<tr>
<td>5.ALL ITEMS</td>
</tr>
<tr>
<td>6.ALL PANELS</td>
</tr>
</tbody>
</table>
-----------

COLUMNS ARE: 3

age
sex
%idl wt
dur diab
ins req

HIT [RETURN] TO TERMINATE THE LIST
ITEM ABBREV.: 

The complete specification which results from this interaction is displayed as:

WORKSHEET IS: PATIENTS VS ITEMS (FOR A TIME)
1-day sub-int's FROM EVENT begin TO EVENT end
(FROM THE START OF THE day CONTAINING THE EVENT begin)
SELECTING THE FIRST MEASURED DURING ALL THE 1-day sub-int's

THE ROWS ARE: PATIENT SET (all )

THE COLS ARE: ITEMS (age etc.)

TYPE yes to LOGICALLY RESTRICT RETRIEVED PANELS, ELSE HIT [RETURN]:
The user checks this specification (noting that the time specification referring to
days is equivalent to retrieve all data) and indicates that he does not want to specify
logical conditions to restrict the panels that are retrieved. Then, data are retrieved
patient-by-patient and item-by-item from the SDF. During retrieval, patient abbreviations and panel numbers appear on the screen to tell the user what data are
currently being examined. If there are patients for whom no data are retrieved,
their abbreviations are displayed at the bottom of the screen. When retrieval is
completed, the user names the resultant worksheet CLINICAL and enters a title.
This worksheet is shown below.

WORKSHEET CLINICAL
TITLE Clinical Characteristics of Diabetic Patients
CREATED 12/17/75 MODIFIED 12/17/75
# OF ROWS 7
# OF COLS 5

ROWS/COLS  1  2  3  4  5

<table>
<thead>
<tr>
<th>LABELS</th>
<th>age</th>
<th>sex</th>
<th>bmi</th>
<th>wt</th>
<th>dur</th>
<th>diag</th>
<th>ins</th>
<th>req</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 case 1</td>
<td>27</td>
<td>m</td>
<td>100</td>
<td>12</td>
<td>35</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 case 2</td>
<td>24</td>
<td>m</td>
<td>98</td>
<td>11</td>
<td>45</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 case 3</td>
<td>32</td>
<td>m</td>
<td>98</td>
<td>15</td>
<td>45</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 case 4</td>
<td>34</td>
<td>m</td>
<td>94</td>
<td>26</td>
<td>50</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 case 5</td>
<td>21</td>
<td>m</td>
<td>96</td>
<td>9</td>
<td>40</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 case 6</td>
<td>23</td>
<td>m</td>
<td>98</td>
<td>10</td>
<td>30</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7 case 7</td>
<td>28</td>
<td>m</td>
<td>99</td>
<td>15</td>
<td>45</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Retrieval of a Worksheet for a Specific Patient. To retrieve worksheet CASE1C, the user first chooses times vs items for a patient as the worksheet type
and enters the patient abbreviation case 1.

To define the rows, the user indicates that they correspond to times; chooses
hour as the time unit because his measurements are made at approximately hour
intervals; selects 1 hour subintervals m, n, etc. . . . from event X (or date, time) as
the time interval (i.e., indicates that he wants to list a series of hour numbers that
define several hour-long time units relative to an event that he will name); indicates
that he does not want the system to select the times automatically; and then enters
the hour numbers −2, −1, 0, 1, 2, 4, 6, 8, and 10 (the hours at which measurements
were to be made).

The system now prompts for the abbreviation of the event, and the user types
instopc. This means that data values in the "hour −2" row of the retrieved work-
sheet will be those measured in the clock hour that is two hours before the one in
which insulin was last given (e.g., if insulin was last given at 8:10 A.M., retrieval by
hour proceeds as if the event instopc occurred at 8:00 A.M. and the values in the
"hour −2" row would be those measured between 6:00 A.M. and 6:59 A.M.); the
values in the "hour 0" row will be those measured in the same clock hour (but not
necessarily at the same time) that insulin was last given; and so forth.

To complete the definition of the rows, the user specifies the last measured
value as the one to select if more than one value was measured during a clock hour.
In our example, only one value was measured each clock hour, so the choice is
arbitrary. If some of the planned measurements either were not made or were
made earlier or later than planned and therefore occurred in the same clock hour
as other measurements, some values would be missing on the worksheet. The
selection choice all values could be used to make certain that all values were
retrieved, regardless of when they were measured, but, as noted above, this ex-
cludes the possibility of time alignment.

To complete the specification for this retrieval, the user indicates that columns
are to correspond to panels and then indicates that the only panel to be retrieved
is the one called control. He does not wish to logically restrict the panels to be
retrieved. The complete specification which results from this interaction is dis-
played as shown below.

Worksheet is: Times vs items (For a Patient: case 1)

The rows are: Times
1-hour sub-int's -2 etc. from event instopc
(from the start of the hour containing the event instopc)
selecting the last measured each 1-hour sub-int'

The cols are: Panels (control etc.)

When retrieval is complete, the user names the resultant worksheet CASE1C,
provides a title, and displays it. Note that because entire panels were retrieved, the
date and time for each instance are included, in addition to values for the constituent items. Also, the last measured value of insulin is in the row HR 0, i.e., the event
instopc and the relative hour numbers have the intended meaning.

Worksheet CASE1C
Title: Flowsheet for case1, control data
Created 12/17/75 Modified 12/22/75
# of rows 9
# of cols 7

<table>
<thead>
<tr>
<th>Rows/cols</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Labels</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 HR -2</td>
<td>5/4/1975</td>
<td>615</td>
<td>35</td>
<td>183.500</td>
<td>89.300</td>
<td>322</td>
</tr>
<tr>
<td>2 HR -1</td>
<td>5/4/1975</td>
<td>700</td>
<td>35</td>
<td>97.000</td>
<td>98.300</td>
<td>324</td>
</tr>
<tr>
<td>3 HR 0</td>
<td>5/4/1975</td>
<td>810</td>
<td>35</td>
<td>96</td>
<td>98.500</td>
<td>318</td>
</tr>
<tr>
<td>4 HR 1</td>
<td>5/4/1975</td>
<td>917</td>
<td></td>
<td>158.100</td>
<td>86.300</td>
<td>298</td>
</tr>
<tr>
<td>5 HR 2</td>
<td>5/4/1975</td>
<td>1820</td>
<td></td>
<td>282.300</td>
<td>87.200</td>
<td>239</td>
</tr>
<tr>
<td>6 HR 3</td>
<td>5/4/1975</td>
<td>1205</td>
<td></td>
<td>257.600</td>
<td>127.300</td>
<td>307</td>
</tr>
<tr>
<td>7 HR 4</td>
<td>5/4/1975</td>
<td>1400</td>
<td></td>
<td>264.600</td>
<td>131.400</td>
<td>335</td>
</tr>
<tr>
<td>8 HR 5</td>
<td>5/4/1975</td>
<td>1612</td>
<td></td>
<td>283.700</td>
<td>143.000</td>
<td>370</td>
</tr>
</tbody>
</table>

(Note: ... indicates missing values)

Retrieval of a Worksheet for a Specific Item. This example illustrates the
retrieval of the worksheet GLUCOSEC. Here, the user first indicates that all the
retrieved data are to be values of the single item glucosec, i.e., he wants a times vs
patients worksheet for an item.

To define rows, the user indicates that each row is to apply to a different patient
in the system-defined patient set called all, i.e., data are to be retrieved for all
patients in the study. The name of a patient set that was created using the subset
activity would also be acceptable.
Columns of worksheet GLUCOSEC are defined in the same way that rows of worksheet CASE1C were defined. Because the worksheet is for a single item and rows correspond to patients, columns must represent times. As above, the user chooses hour as the time unit; selects the "list of intervals from an event" type of time interval; enters the hour numbers -2, -1, 0, etc.; indicates that these are relative to the event instopc; and chooses last value as the time selection choice. The complete specification which results from this interaction is displayed as:

WORKSHEET IS : TIMES VS PATIENTS  (FOR AN ITEM: glucosec)

THE ROWS ARE : PATIENT SET (all)

THE COLS ARE : TIMES
1-hour sub-int's -2 etc. FROM EVENT instopc
(from the start of the hour containing the event instopc)
selecting the last measured 1-hour sub-int'

All seven rows and the first six columns of the retrieved worksheet, which the user names GLUCOSEC, are shown below. The "hour -1" measurement for case 4 and the "hour 1" measurement for case 2 were either not recorded or not measured during these clock hours. The user could determine which was the case by repeating the specification such that all values instead of only the last value for each listed hour were retrieved.

WORKSHEET GLUCOSEC
TITLE Control Glucosec by Patient and by Hour
CREATED 12/17/75    MODIFIED 12/17/75
# OF ROWS 7   # OF COLS 9

<table>
<thead>
<tr>
<th>ROWS/COLS</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>LABELS</td>
<td>HR</td>
<td>-2</td>
<td>HR</td>
<td>-1</td>
<td>HR</td>
<td>0</td>
</tr>
<tr>
<td>1 case 1</td>
<td>103.500</td>
<td>97.000</td>
<td>86.000</td>
<td>78.000</td>
<td>69.000</td>
<td>59.600</td>
</tr>
<tr>
<td>2 case 2</td>
<td>90.000</td>
<td>84.500</td>
<td>71.000</td>
<td>57.000</td>
<td>45.000</td>
<td>33.000</td>
</tr>
<tr>
<td>3 case 3</td>
<td>101.300</td>
<td>95.400</td>
<td>89.200</td>
<td>83.000</td>
<td>77.000</td>
<td>71.000</td>
</tr>
<tr>
<td>4 case 4</td>
<td>75.000</td>
<td>68.500</td>
<td>61.000</td>
<td>53.000</td>
<td>45.000</td>
<td>37.000</td>
</tr>
<tr>
<td>5 case 5</td>
<td>111.700</td>
<td>109.800</td>
<td>108.100</td>
<td>106.700</td>
<td>105.000</td>
<td>103.000</td>
</tr>
<tr>
<td>6 case 6</td>
<td>114.000</td>
<td>110.300</td>
<td>108.700</td>
<td>107.400</td>
<td>106.000</td>
<td>104.500</td>
</tr>
<tr>
<td>7 case 7</td>
<td>106.500</td>
<td>104.700</td>
<td>102.100</td>
<td>100.700</td>
<td>99.300</td>
<td>98.000</td>
</tr>
</tbody>
</table>

Retrieval of a Patient Worksheet with Conditions Placed on Item Values. The final example illustrates (1) the use of a time interval specified as a range of units relative to a calendar date and clock time; and (2) the specification of logical conditions to restrict the data that are retrieved. In this worksheet, the patient is case I, the unit chosen is hours, and the time interval is 1-hour subintervals -2 through 10 from date 5/4/75 (i.e., May 4, 1975) and time 800 (i.e., 8:00 A.M.). The panel named control will have conditions placed upon it.
WORKSHEET IS : TIMES VS ITEMS  (FOR A PATIENT: case 1 )
THE ROWS ARE : TIMES

TIME INTERVAL CHOICE

<table>
<thead>
<tr>
<th>TYPE FOR A TIME INTERVAL OF</th>
</tr>
</thead>
<tbody>
<tr>
<td>1  l-hour  sub-int's FROM EVENT X (OR DATE,TIME) TO EVENT Y (OR DATE,TIME)</td>
</tr>
<tr>
<td>2  l-hour  sub-int's M, N, ETC. ... FROM EVENT X (OR DATE, TIME)</td>
</tr>
<tr>
<td>3  l-hour  sub-int's M THRU N FROM EVENT X (OR DATE,TIME)</td>
</tr>
</tbody>
</table>

1-hour  sub-int's M THRU N FROM EVENT X (OR DATE,TIME)


Only those panels in which the value of glucosec is between 100 and 200 or
glucagon falls between 100 and 120 are to be retrieved. Since this is a logical OR
statement, any rather than all is selected. If both conditions had to be simultaneou-
ly satisfied (a logical AND statement), all would have been chosen.

WORKSHEET IS : TIMES VS ITEMS  (FOR A PATIENT: case 1 )
THE ROWS ARE : TIMES
1-hour  sub-int's -2 THRU 10 FROM DATE 5/4/75

(TIME 000)
SELECTING THE LAST MEASURED EACH 1-hour  sub-int'

THE COLS ARE : PANELS  (control etc.)

TYPE yes TO LOGICALLY RESTRICT RETRIEVED PANELS, ELSE HIT (RETURN): yes

THE PANELS MAY SATISFY EITHER 'ANY' OR 'ALL' CONDITIONS IN A LIST

TYPE yes TO SELECT 'ANY', ELSE HIT (RETURN): yes

The first condition is entered:

INCLUDE EACH PANEL SATISFYING ANY OF THE APPROPRIATE FOLLOWING CONDITIONS

TO TERMINATE LIST OF CONDITIONS, HIT (RETURN)
ITEM: glucosec

IF glucosec IS IN A RANGE (E.G. 3,6) OR IS A CODE (E.G. m): 180,200

The final specification which results from this interaction is displayed as:
WORKSHEET IS : TIMES VS ITEMS  (FOR A PATIENT: case 1 )
THE ROWS ARE : TIMES
 1-hour sub-int's  -2 THRU 10 FROM DATE 5/4/75
  TIME 200
  (FROM THE START OF THE hour CONTAINING 800 , ON 5/4/75 )
  SELECTING THE LAST MEASURED EACH 1-hour sub-int'
THE COLS ARE : PANELS
  (control etc.)
  INCLUDE EACH PANEL SATISFYING ANY OF THE APPROPRIATE FOLLOWING CONDITIONS
  IF  100 <= glucosec <= 200
  OR IF  100 <= glucogonc <= 120

The retrieved worksheet is named, titled, and displayed below. Comparing the resultant worksheet CASE1LTD with CASE1, we can see that only those panels with glucosec between 100 and 200 or glucogonc between 100 and 120 were retrieved.

<table>
<thead>
<tr>
<th>ROWS/COLS</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>LABELS</td>
<td>DATE</td>
<td>TIME</td>
<td>insulin</td>
<td>glucosec</td>
<td>glucogonc</td>
<td>alamines</td>
</tr>
<tr>
<td>1 HR</td>
<td>5/4/1975</td>
<td>615</td>
<td>35</td>
<td>103.500</td>
<td>89.300</td>
<td>.322</td>
</tr>
<tr>
<td>2 HR</td>
<td>5/4/1975</td>
<td>917</td>
<td>...</td>
<td>150.100</td>
<td>108.300</td>
<td>.292</td>
</tr>
<tr>
<td>3 HR</td>
<td>5/4/1975</td>
<td>1020</td>
<td>...</td>
<td>202.300</td>
<td>119.700</td>
<td>.291</td>
</tr>
</tbody>
</table>

OPTION 3: STUDY DATA FILE TO TERMINAL/COMMUNICATION FILE

This subactivity is used to retrieve specified numeric and text data from the SDF and to display it at the user's terminal or store it in a Communication File. Once stored in a Communication File, data can be processed by a program written in the BASIC programming language. Results from the BASIC program can be displayed at the terminal, printed on the system printer, stored in a Communication File properly formatted for transferring the results into a worksheet (using retrieve, option 6), or stored on magnetic tape for transmittal to another computer system. The retrieval specification is similar to that in retrieve, option 2, but the detailed contents of panels (including date and time of sample) are displayed or stored one panel at a time. The display feature is useful when reviewing a well-defined set of numeric and text data or when checking specified data prior to storing them in a Communication File.

The retrieval specification involves specifying patients, panels, and times. Patients may be specified by entering a list of patient abbreviations or the name of a patient set. Panels are specified by entering panel names or by requesting the retrieval of all panels. As in retrieve, option 2 (where it is explained in detail), the time specification is more complex.

The time is specified by defining a time interval, dividing it into time units of a specified duration, and selecting one or all data values for each unit-long subinter-
The unit choice may be minute, hour, day, week, month, year, or multiples of these. The time interval may be defined as

- The time between two time markers (e.g., from begin to end),
- A list of time units from a time marker (e.g., 5, 7, and 12 days from the event called transplant), or
- A range of time units relative to a time marker (e.g., 5 through 10 hours from 1/23/75 at 1200.)

The time markers may be the system-defined events begin and end, user-defined events that are described in the schema, or calendar dates and clock times specified by the user. The selection choice may be first value, last value, or all values, i.e., the first, last, or all values in each unit-long subinterval in the specified time interval.

Once the retrieval is specified, the user indicates whether the data are to be displayed on his CRT or stored in a Communication File. If he chooses the former, panels of data are displayed one panel at a time. He is then given another opportunity to store the data in a Communication File.

Each study may have up to ten Communication Files, designated by one-digit numbers. Thus, a Communication File is specified by entering a digit between 0 and 9. For purposes of processing by BASIC, the name of a Communication File is COMMFILE.n, where n is the digit specified by the user. The created Communication File may be associated with the study currently being accessed or with another CLINFO study. Generally, a Communication File is transferred to a BASIC account for further processing there. The study or account to which the file is to be transferred is designated by entering the log-on password for that study or account. Thus, the transfer must be effected by, or in cooperation with, a person who has access to the destination study or account. Note that the Communication File number refers to a file in the destination study or account, and this is not necessarily the study currently being accessed.

A Communication File created with this subactivity is formatted exactly like the information displayed at the terminal, i.e., as a series of character strings grouped by patient (with patient abbreviations in alphanumeric sort order), by panel (in schema panel-number order) for each patient, and by time for each panel type. Each character string includes an identifier (e.g., DATE OF SAMPLE or an item name) and a value. The format for each panel in the file is shown below.

Each of the following lines represents a character string. The listed upper-case words and colons are included in the strings; the lower-case words refer to item names and values that depend on the panel and data.
String

PATIENT: 8-char patient abbreviation
PANEL: 8-char panel name
TYPE: NUMERIC (or TYPE: TEXT)
# OF ENTRIES: n
DATE OF SAMPLE: mm/dd/yyyy
TIME OF SAMPLE: hh:mm
1st 2-char code 8-char item name:value
2nd 2-char code 8-char item name:value
\[ \ldots \]
nth 2-char code 8-char item name:value

where the 2-char code is:

Mb for modified value (b signifies the blank character)
Eb for value exceeds range
Xb for modified value exceeds range
bh for value not changed

This code is not used (and occupies no space) for text items.

An item value is a string formatted as:

\[ \ldots (3 \text{ periods}) \text{ if missing} \]
\[ \text{a number (use the VAL function to convert to a real number)} \]
\[ \text{if a numeric item} \]
\[ 8 \text{ characters if a char item} \]
\[ 8 \text{ characters representing mmddyyyy if a date item} \]
\[ 4 \text{ characters representing hhmm if a time item} \]
\[ \text{a character string if a text item} \]

Dialog

In this example, data in two SDF panels are retrieved for all patients on the study. The data are reviewed at the terminal and stored in a Communication File, which is then processed by a BASIC program to produce a list of the names, addresses, and last visit dates of all patients who have ventricular fibrillation and who have not visited the Clinical Research Center for more than one year.

The two panels of interest are those numbered 1 and 2 in the schema. Panel 1 is a text panel called id; it comprises three text items, name, street, and city. Panel 2 is a numeric panel called data; it comprises 29 numeric items, including a char (i.e., a character-coded) item called v.fib. The item v.fib has the value yes or no, depending on whether or not ventricular fibrillation is associated with the patient's visit.

The interaction required to specify the retrieval is like that in retrieve, option 2, but here the specification order is prescribed, i.e., patients, then panels, then the time specification. In this case, the user first indicates that the patients of interest are those in the set called all (i.e., all the patients in the study) and then that the panels to be retrieved are those called id and data. To complete the specification, the user chooses day as the time unit, chooses days from event X to event Y, indicates event X is begin and event Y is end, then selects first value as the one to
be retrieved during each time unit. Because no more than one set of data is collected in a day, this time specification will result in retrieval of all the relevant data for each patient. The retrieval specification which results from this interaction is summarized below:

```
retrieve,3 RETRIEVE DATA FROM STUDY FILE TO TERMINAL/COMMUNICATION FILE
PATIENTS ARE : PATIENT SET (all)
PANELS ARE : PANELS (id etc.)
TIMES ARE : TIMES
1-day intervals FROM EVENT begin TO EVENT end
(From the start of the day CONTAINING THE EVENT begin
SELECTING THE FIRST MEASURED EACH 1-day interval
```

To review the retrieved data, the user indicates that he would like it displayed rather than stored immediately in a Communication File. The data for the first patient (who has the abbreviation I) are shown below as they appear on the screen one panel at a time. In this case, all the values in the data panel are missing except that of u,f,h. Each time the user hits the RETURN key, another panel appears. For each patient, one id panel is displayed (because name, street, and city are entered only once), followed by one or more data panels.

```
retrieve,3 STUDY DATA FILE TO TERMINAL/COMMUNICATION FILE
PATIENT: 1
PATIENT: id
PANEL: id
TYPE: TEXT
# OF ENTRIES: 3
DATE OF SAMPLE: 9/5/1974
TIME OF SAMPLE: 12:00
name : Burks Eddie
street : 4128 22nd Pl.
city : Northridge, CA 91324
hit (return) TO CONTINUE:
```
PATIENT: 1

PATIENT: data
TYPE: NUMERIC
# OF ENTRIES: 29
DATE OF SAMPLE: 9/5/1974
TIME OF SAMPLE: 12:00

entrydat:
run#
age
cont:
isoenzym:
ekgeval:
v.fib:yes
probi:
previ:
angina:
hyperbp:
chf
blank
surviv
mode
sex
chpain
rob
indigs
uncons
deathdat:
ett
hca
hsurg
lastdat
study#
gwave
******001:
******002:

HIT [RETURN] TO CONTINUE:

After all the retrieved data are displayed, the user is again asked to type the number of a Communication File or to indicate that he would like the data displayed. This time he types the number 5. Next he enters the password 2037basic to indicate that data are to be transferred to Communication File 5 in his BASIC account. To protect his privacy, the system does not display the password.

TO SEND DATA TO A COMMUNICATION FILE, TYPE ITS NUMBER (0,9):
TO SEND DATA TO YOUR TERMINAL HIT [RETURN]: 5

PASSWORD FOR THE DESIRED STUDY:

Now that the data have been retrieved and transferred, the user logs off. He then enters the password 2037basic to log on to his BASIC account, where he runs the program called address, which the system manager previously wrote for him. A partial listing of the results displayed at the terminal is shown below. The names, addresses, and last visit dates of the patients selected by the BASIC program are also listed on the system printer, as shown. (The “patients” listed below are fictional.)
*run "address
THE COMMUNICATION FILE NUMBER PLEASE (0 TO 9): 5

PATIENTS WITH V.FIB WHOSE LAST VISIT WAS MORE THAN A YEAR AGO:

Burks Eddie
4128 22nd Pl.
Northridge, CA 91324
9/ 5/1974

Machado Felipe
2830 E Walnut
Palo Alto, CA 94305
11/20/1974

Popowitz Abraham
3268 Descanso Dr.
Los Angeles, CA 90012
6/15/1974

Smith Howard E
1215 Lindhurst Dr.
San Francisco, CA 94101
3/12/1974

Encinas Joe R
4234 Degnan Ave.
Chico, CA 95926
5/17/1974

Crowley Eugene T
280 S Ardeno Pl.
Duarte, CA 91010
11/24/1973

Choi Young Ki
6462 Laurel Canyon
Los Angeles, CA 90024
12/ 3/1974

STOP AT 0620
*

PATIENTS WITH V.FIB WHOSE LAST VISIT WAS PRIOR TO 3/16/1975

Burks Eddie
4128 22nd Pl.
Northridge, CA 91324
9/ 5/1974

Machado Felipe
2830 E Walnut
Palo Alto, CA 94305
11/20/1974

Popowitz Abraham
3268 Descanso Dr.
Los Angeles, CA 90012
6/15/1974

Smith Howard E
1215 Lindhurst Dr.
San Francisco, CA 94101
3/12/1974

Encinas Joe R
4234 Degnan Ave.
Chico, CA 95926
5/17/1974

Crowley Eugene T
280 S Ardeno Pl.
Duarte, CA 91010
11/24/1973

A listing of the BASIC program that reads the Communication File created above, performs some error checking, and produces results is reproduced below. Nearly half the lines consist of commentary. Once the input and output parts of the program are working, it can be easily extended to do more elaborate processing. For
example, patient selection could be based on city or state or on items other than v.fib. Also, patients could be sorted alphabetically by last name, and their first names could be listed first.

0010 REM This example program processes data in a communication file that
0020 REM was created from PCF data. For each patient on the study there
0030 REM is an "id" panel and one or more "data" panels. The id panel is
0040 REM a text panel which contains the patient's full name, street address
0050 REM and city. The data panel includes several numeric items; of
0060 REM interest here is the coded item "v.fib" If the character-string
0070 REM value of v.fib is "yes" this means that ventricular fibrillation
0080 REM was indicated on the corresponding visit.
0090 REM
0100 REM The purpose of the program is to produce a list of the names and
0110 REM addresses of those patients who had v.fib indicated on their last
0120 REM visit and who have not visited for at least a year. It does this
0130 REM by first reading an id panel for a patient and storing his name
0140 REM and address. Next, it finds the patient's last data panel and
0150 REM records the value of v.fib in that panel and the panel's data
0160 REM collection date. If v.fib is yes and the date is at least one year
0170 REM prior to the current date, the patient's name, address and last visit
0180 REM date are stored in a file. After the entire communication file has
0190 REM been processed the file of names and addresses is listed on the
0200 REM system printer.
0210 REM
0220 REM The program assumes that the communication file contains one or
0230 REM more id panels for a patient followed by one or more data panels
0240 REM for the same patient followed by id and data panels for the next
0250 REM patient. If this is not the case an error message is printed.
0260 REM
0270 DIM N%1,131,F%103,P%143,F2%134,N%16
0280 DIM D%112,F%103,D%25,T%21
0290 DIM N%29,S%134,C%39
0300 DIM V%8(9,0)
0310 PRINT
0320 INPUT "THE COMMUNICATION FILE NUMBER PLEASE (0 TO 9): ": N%
0330 LET F%="comfile".N%
0340 OPEN FILE1.33,F%
0350 REM initialize the output file
0360 OPEN FILE2.12."output"
0370 CLOSE FILE2)
0380 PRINT
0390 PRINT "PATIENTS WITH V.FIB WHOSE LAST VISIT WAS MORE THAN A YEAR AGO:"
0400 PRINT
0410 REM Get current day of the month, month of the year=ml, year=y
0420 LET D1=SYS(1)
0430 LET M1=SYS(2)
0440 REM read the data panels
0450 REM read F%="PATIENT: abbrev and F2%="NAME: name
0460 REM read the id panel for this patient
0470 REM P3%="NAME: name where name refers to the panel being searched for
0480 LET P3%="NAME:"
0490 REA克莱FILE13.T%5,P%2
0500 IF EOF1 THEN GOTO 1360
0510 REM if this is not an id panel print an error message
0520 IF P2%<>P%3 THEN GOTO 1370
0530 REM read next panel type, entries, date and time
0540 REM and read the values of name, street, and city
0550 REA克莱FILE13.T%5,P%5,P%5,T%5,S%5,C%5
0560 IF EOF1 THEN GOTO 1360
0570 REM read the next panel
0580 REM if it is an id panel: if same patient: update
0590 REM if different patient: restart
0600 REM if not an id panel (it shouldn't be): is same rt: continue
0610 REM
0620 REA克莱FILE13.T%5,P%5,P%5
0630 IF EOF1 THEN GOTO 1360
0640 IF P2%<>P%3 THEN GOTO 0690
0650 REM Another id panel
0660 IF P%5=P%5 THEN GOTO 0550
0670 LET P1$=P4$
0680 GOTO 0950
0690 REM Not an id file
0700 IF P4$=P1$ THEN GOTO 0740
0710 PRINT "ERROR: DATA OUT OF SEQUENCE"
0720 PRINT "NEW PATIENT","P4$" "WITH","P2$" " PANEL"
0730 STOP
0740 REM If this is not a data panel print an error message
0750 LET P3$="PANEL: data"
0760 IF P2$<>P3$ THEN GOTO 1570
0770 REM Find the most recent data panel for this patient
0780 IF P3$=P1$ THEN GOTO 0720
0790 REM If the next patient is different process this panel then restart
0800 REM If the next patient is the same:
0810 REM If the next panel is a data panel: update
0820 REM If the next panel is not data: process this one then find next pt
0830 REM Read the panel type, # of entries, date and time
0840 READ FILE[1].T$.E$.D$.T$
0850 IF EOF(1) THEN GOTO 1360
0860 REM Read the entries, saving the value of v.fib
0870 LET V$="...
0880 REM E$ has the format: # OF ENTRIES:n
0890 REM C$ has the format: 8-char item name: value
0900 FOR I=1 TO VAL(E$[14:15])
0910 READ FILE[1].C$
0920 IF C$[3:10]="v.fib" THEN LET V$=V$[12:19]
0930 NEXT I
0940 REM Now read the beginnings of the next panel
0950 READ FILE[1].P4$.P2$
0960 IF EOF(1) THEN GOTO 0990
0970 IF P2$<>P3$ THEN GOTO 0990
0980 REM This is the most recent panel for this patient
0990 REM The next panel is for a different patient or isn't a data panel
1000 REM or there are no more data in the communication file
1010 REM
1020 REM If this patient didn't have v.fib on the last visit process the next
1030 REM patient if not at the end of file
1040 IF V$[1:3]<>"Yes" THEN GOTO 1320
1050 REM This patient had v.fib on his last visit
1060 REM If he hasn't been in for a year or more list his name and address
1070 REM Extract visit month, day and year from D$ which has the format
1080 REM DATE OF SAMPLE:mm/dd/yyyy
1090 LET M=VAL(D$[1:5])
1100 LET D=VAL(D$[7:9])
1110 LET Y=VAL(D$[12:14])
1120 REM Compare with present date
1130 IF Y>Y1 THEN GOTO 1220
1140 IF Y<Y1 THEN GOTO 1200
1150 REM Visit was last calendar year
1160 IF Y1+1 THEN GOTO 1220
1170 REM Visit was this calendar year
1180 IF M>Y1 THEN GOTO 1220
1190 IF M<Y1 THEN GOTO 1220
1200 REM Visit was last month, last calendar year
1210 REM Visit was this month, last calendar year
1220 REM Last visit was at least a year ago
1230 REM Print patient's name/last visit date and address at the terminal
1240 REM and store in a file for later printing.
1250 REM Name, street address, and city start in position 10 of m$. s$. c$
1260 LET C$=S$[10:29]."...
1270 LET D$[16:25]."C$
1280 PRINT C$
1290 PRINT S$[10:34]
1270 PRINT C$[10:39]
1280 PRINT "OPEN FILE[2.23, "output"
1310 CLOSE FILE(2)
1320 REM Process the next patient if not End Of File
1330 LET P1$=P4$
1340 LET P3$="PANEL: id"
1350 IF EOF(1)<>1 THEN GOTO 0520
1360 REM All communication file data have been processed.
1370 CLOSE
1380 PRINT "SELECTED PATIENT NAMES AND ADDRESSES ARE BEING LISTED"
1390 PRINT
1400 REM Print the names and addresses in output file on the system printer
1410 OPEN FILE[2.1], "$t"
OPTION 4: A WORKSHEET TO A WORKSHEET

This option copies a portion of a source worksheet and stores the copied data and labels in specified locations of a destination worksheet which may be the source worksheet or any other worksheet. Copying data from worksheet to worksheet is useful for properly positioning data to be analyzed and for building a summary worksheet from several others. Figure I-8 in Sec. I illustrates what can be accomplished.

The data are transferred by rows or columns. In addition, row and column labels may be transferred. Even if there are no data in the source worksheet, the labels may still be transferred; thus a series of empty worksheets can be set up, with identical row and/or column labels for convenience in data collection.

Dialog

The system prompts for the source worksheet, the rows and columns to be copied from the source, the destination worksheet, and the starting row and column in the destination worksheet. The worksheet to be used as the source worksheet in the following example is shown below:

```
1 / 20 /76, at 17:17
WORKSHEET CHITST
TITLE chi-square test
CREATED 1/20/76 MODIFIED 1/20/76
# OF ROWS 3
# OF COLS 3
ROWS/COLS 1 2 3
LABELS Small Medium Large
1 Red 4 7 9
2 Blue 8 3 1
3 Green 4 4 0
```

After requesting option 4 of retrieve, the dialog continues:
In this example, the user has

- Confirmed that the current worksheet is the source.
- Created the worksheet DEST to be used as the destination worksheet (alternatively, an existing worksheet could have been selected as the destination).
- Chosen to copy data by rows.
- Chosen to copy data in rows 1 through 3 and columns 1 through 3.
- Directed these rows and columns, respectively, to rows 1 through 3 and columns 1 through 3 of the destination worksheet.
• Chosen to copy both row and column labels.
• Indicated, by responding with the RETURN key to the last prompt, that he wanted to copy another row by column data matrix.

The second data copy and transfer is specified as:

```
retrieve, 4 WORKSHEET TO WORKSHEET COPY

THE CURRENT SOURCE WORKSHEET IS:

WORKSHEET CHITST
TITLE chi-square test
CREATED 1/28/76     MODIFIED 1/28/76
# OF ROWS 3
# OF COLS 3

TO SELECT ANOTHER SOURCE WORKSHEET, TYPE ITS NAME, ELSE [RETURN]:

THE CURRENT DESTINATION WORKSHEET IS:

WORKSHEET DEST
TITLE destination worksheet for chitst
CREATED 1/21/76     MODIFIED 1/21/76
# OF ROWS 6
# OF COLS 5

TO SELECT ANOTHER DESTINATION WORKSHEET, TYPE ITS NAME, ELSE [RETURN]:

TYPE c TO COPY COLUMNS OR r TO COPY ROWS: c

SOURCE COL RANGE...a,b...: 1,3
SOURCE ROW RANGE...a,b...: 1,3

TYPE yes TO TRANSPOSE COLS INTO ROWS, ELSE [RETURN]:

COLS WILL BE COPIED INTO COLS
ROWS WILL BE COPIED INTO ROWS

DESTINATION STARTING COL: 3

DESTINATION STARTING ROW: 6

TYPE yes TO COPY ROW LABELS, ELSE [RETURN]:

TYPE yes TO COPY COLUMN LABELS, ELSE [RETURN]:

WORKSHEET COPY COMPLETE

TYPE yes TO DISPLAY THE DESTINATION WORKSHEET, ELSE [RETURN]: yes
```

In this case, the system remembers the names of the existing source and destination worksheets. The user has

• Confirmed that he wants to copy data from the same source and transfer it to the same destination worksheet.
• Chosen to copy data by columns; this is similar to copying data by rows except that the user has a different mental set and column-related prompts precede row-related prompts.
• Chosen to copy data in columns 1 through 3 and rows 1 through 3 (i.e., the same data as copied above) from worksheet CHITST and transfer it to columns 3 through 6 and rows 6 through 8 of worksheet DEST.
- Chosen not to copy either row or column labels.
- Decided to display the current contents of worksheet DEST.

1/21/76, at 8:26

WORKSHEET DEST
TITLE destination worksheet for CHITST
CREATED 1/21/76 MODIFIED 1/21/76
# OF ROWS 8
# OF COLS 5

<table>
<thead>
<tr>
<th>ROWS/COLS</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>LABELS</td>
<td>Small</td>
<td>Medium</td>
<td>Large</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 Red</td>
<td>4</td>
<td>7</td>
<td>9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 Blue</td>
<td>8</td>
<td>3</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 Green</td>
<td>4</td>
<td>4</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td>4</td>
<td>7</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td>4</td>
<td>0</td>
</tr>
</tbody>
</table>

(Note: ... indicates missing values)

TYPE yes TO CONTINUE WITH retrieve, 4 ELSE TYPE THE STARTING ROW #; yes

Data and labels from worksheet CHITST have been copied and transferred into worksheet DEST as shown above. The results of the first transfer appear in the upper left-hand portion of DEST; the results of the second appear in the lower right.

When labels are copied, they may replace labels that already exist in the destination worksheet. In this example, the first transfer replaced blank labels with those shown. Because both transfers resulted in storing data in column 3, the label Large placed there by the first transfer would have been replaced by Small if column labels were copied by the second transfer. That is, if row and column labels had been copied by both transfers, the row labels would now be Red, Blue, Green, blank, blank, Red, Blue, Green, and the column labels would be Small, Medium, Small, Medium, Large.

To continue copying data, the user returns to retrieve, option 4, by responding yes to the prompt below the worksheet display. On the way, the user is asked if he would like DEST to become the current worksheet; he responds negatively because he plans to copy more data from CHITST into DEST.

The following interaction, using the same source and destination worksheets, shows the effect of transposing data so that data stored in rows of CHITST are stored in columns of DEST.
retrieve,4 WORKSHEET TO WORKSHEET COPY

THE CURRENT SOURCE WORKSHEET IS:

WORKSHEET CHITST
TITLE chi-square test
CREATED 1/20/76 MODIFIED 1/20/76
# OF ROWS 3
# OF COLS 3

TO SELECT ANOTHER SOURCE WORKSHEET, TYPE ITS NAME, ELSE [RETURN]:

THE CURRENT DESTINATION WORKSHEET IS:

WORKSHEET DEST
TITLE destination worksheet for chitst
CREATED 1/21/76 MODIFIED 1/21/76
# OF ROWS 8
# OF COLS 5

TO SELECT ANOTHER DESTINATION WORKSHEET, TYPE ITS NAME, ELSE [RETURN]:

TYPE c TO COPY COLUMNS OR r TO COPY ROWS: r

SOURCE ROW RANGE..a,b..: 1,2
SOURCE COL RANGE..a,b..: 1,2

TYPE yes TO TRANSPOSE ROWS INTO COLS, ELSE [RETURN]: yes

ROWS WILL BE COPIED INTO COLS
COLS WILL BE COPIED INTO ROWS

DESTINATION STARTING COL: 4

DESTINATION STARTING ROW: 4

TYPE yes TO COPY ROW LABELS, ELSE [RETURN]: yes

TYPE yes TO COPY COLUMN LABELS, ELSE [RETURN]: yes

WORKSHEET COPY COMPLETE

TYPE yes TO DISPLAY THE DESTINATION WORKSHEET, ELSE [RETURN]: yes

The result is then displayed:

1 / 21 /76, at 8:31

WORKSHEET DEST
TITLE destination worksheet for chitst
CREATED 1/21/76 MODIFIED 1/21/76
# OF ROWS 8
# OF COLS 5

ROWS/COLS 1 2 3 4 5

LABELS Small Medium Large Red Blue

1 Red 4 7 9 ... ... 4 5 6 7 8
2 Blue 8 3 1 ... ... 4 5 6 7 8
3 Green 4 4 0 ... ... 4 5 6 7 8
4 Small ... ... ... 4 8 5 6 7 8
5 Medium ... ... ... 7 3 6 7 8
6 ... ... ... 4 7 6 7 8
7 ... ... ... 8 3 6 7 8
8 ... ... ... 4 4 6 7 8

(NOTE: ... indicates missing values)

TYPE yes TO CONTINUE WITH retrieve,4 ELSE
TYPE THE STARTING ROW #: iana
Data and labels from rows 1 and 2 of CHITST have been transferred into columns 4 and 5 of DEST, and data and labels from columns 1 and 2 of CHITST have been transferred into rows 4 and 5 of DEST.

To analyze the data in DEST, the user types \textit{lan}a in response to the last prompt. When asked if he wants DEST to become the current worksheet, he responds yes, because he is now interested in working on DEST instead of CHITST.

\textbf{OPTION 5: WORKSHEET TO COMMUNICATION FILE}

This subactivity is used to copy the data and labels of a worksheet into a Communication File and to associate that file with the same study, a different study, or an account in which programs can be written in BASIC. Generally, a Communication File is transferred into a BASIC account where the data may be processed by an arbitrary program; the processed data are then transferred (using \textit{retrieve}, option 6) back into the original study. However, this option, together with option 6, may also be used to transfer worksheets from one CLINFO study to another. To prepare data for another computer system, a Communication File may be created and then copied onto magnetic tape. A worksheet may be copied in its entirety by transferring it into a Communication File (with option 5), then retrieving it from the Communication File (with option 6) without any intermediate processing.

The study or BASIC account that is to receive the Communication File is specified by entering the password for that study or account. Thus, the transfer must be effected by, or in cooperation with, someone who has access to the destination study or account.

Each study may have up to ten Communication Files, designated by one-digit numbers. Thus, a Communication File is specified by entering a digit from 0 to 9. For purposes of processing by BASIC, the name of a Communication File is COMM-FILE.n, where n is the digit specified by the user. The file number refers to a file in the destination study or account; it is not generally in the user's study.

A Communication File created using this subactivity contains the size, labels, and data for a single worksheet. The format is two numbers followed by two character strings followed by a matrix:

\begin{itemize}
  \item r (the number of matrix (i.e., worksheet) rows)
  \item c (the number of matrix (i.e., worksheet) columns)
  \item an \( r \times 8 \) character string (the row labels)
  \item a \( c \times 8 \) character string (the column labels)
  \item an \( r \times c \) matrix (the worksheet data)
\end{itemize}

A missing value is stored as the number $-9.99999E-07$.

\textbf{Dialog}

In this example, the worksheet called WS1 is transferred to a BASIC account, where it is processed.
The entire dialog within this subactivity is the entry of the Communication File number and the password. To maintain privacy, the system does not display the password.

```
retrieve,5 WORKSHEET TO COMMUNICATION FILE
SELECT A SOURCE WORKSHEET ([RETURN] SELECTS WS1 ):
```

```
WORKSHEET WS1
TITLE Data to be processed by a BASIC program
CREATED 11/17/77    MODIFIED 11/17/77
# OF ROWS 3         # OF COLS 5

<table>
<thead>
<tr>
<th>ROWS/COLS</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>LABELS</td>
<td>col#1</td>
<td>col#2</td>
<td>col#3</td>
<td>col#4</td>
<td>col#5</td>
</tr>
<tr>
<td>1 row#1</td>
<td>12</td>
<td>7</td>
<td>5</td>
<td>13.700</td>
<td>...</td>
</tr>
<tr>
<td>2 row#2</td>
<td>3</td>
<td>2</td>
<td>25</td>
<td>3</td>
<td>23</td>
</tr>
<tr>
<td>3 row#3</td>
<td>17</td>
<td>4.700</td>
<td>17.300</td>
<td>40.300</td>
<td>10</td>
</tr>
</tbody>
</table>
```

```
PASSWORD FOR THE STUDY RECEIVING THE COMMUNICATION FILE:
COMM FILE 3 ALREADY EXISTS, TYPE yes TO REPLACE IT, ELSE HIT [RETURN]: yes
TO CONTINUE WITH RETRIEVE, HIT [RETURN]: 1900,
```

**Processing the Communication File**

A simple BASIC program that processes a Communication File is listed below. It can be generalized to perform other processing tasks by replacing lines 360 through 440 and by possibly deleting lines 250, 280, 290, and 320.
```
0010 REM This program reads worksheet data stored in a communication file.
0020 REM prints the worksheet data, divides the first row of data by the
0030 REM second, stores the results in the first row, prints the revised
0040 REM worksheet contents, then stores the revised worksheet in the same
0050 REM communication file.
0060 REM
0070 REM The format of a worksheet-type communication file is R,C,R$,$S,A
0080 REM where:
0090 REM R = # of worksheet rows
0100 REM C = # of worksheet columns
0110 REM RX = R x 6 character long string of R row labels
0120 REM RX = C x 6 character long string of C column labels
0130 REM A = R x C matrix of worksheet data
0140 REM
0150 REM Set up constants and dimension statements
0160 REM A CLINFO missing value is represented by the number -9.99999E-07.
0170 LET U=-9.99999E-07
0180 REM Ask the user for the number of the communication file.
0190 PRINT
0200 INPUT "THE COMMUNICATION FILE NUMBER PLEASE (0 to 9): ",NS
0210 LET FS="commfile."+NS
0220 REM Open the file then read and print # of rows, # of cols, labels, data.
0230 REM OPEN FILE[1],FS
0240 REM READ FILE[1],R,C
0250 PRINT R,C
0260 REM RX=R$[R*8],CS[C*8]
0270 REM READ FILE[1],R$,CS
0280 PRINT R$,CS
0290 PRINT CS
0300 DIM A[R,C]
0310 MAT READ FILE[1],A
0320 MAT PRINT A
0330 CLOSE FILE[1]
0340 REM
0350 REM Do the worksheet processing here.
0360 REM In this example program: row#1/row#2 is stored in row#1
0370 REM In a real program error checking should be included.
0380 FOR I=1 TO C
0410 IF A[1,1]=U THEN GOTO 0430
0430 NEXT I
0440 MAT PRINT A
0450 REM store the results.
0460 OPEN FILE[1],FS
0470 WRITE FILE[1],R,C,R$,CS
0480 MAT WRITE FILE[1],A
0490 CLOSE FILE[1]
0500 END

The program's output follows. The numbers 3 and 5 are the numbers of rows
and columns in worksheet WS1. The next two lines are the row and column labels.
The following three lines are the original worksheet and matrix data values. The
last three lines represent the matrix data values after values in the first row have
been divided by corresponding values in the second row and have been restored.

<table>
<thead>
<tr>
<th>row#1</th>
<th>row#2</th>
<th>row#3</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>7</td>
<td>5</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>25</td>
</tr>
<tr>
<td>17</td>
<td>4.7</td>
<td>17.3</td>
</tr>
<tr>
<td>4</td>
<td>3.5</td>
<td>.2</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>25</td>
</tr>
<tr>
<td>17</td>
<td>4.7</td>
<td>17.3</td>
</tr>
</tbody>
</table>

END AT 0500
```
OPTION 6: COMMUNICATION FILE TO WORKSHEET

This subactivity is used to copy the contents of a properly formatted Communication File into a worksheet. The Communication File may have been created from a worksheet by the use of retrieve, option 5, or it may have been created directly by a BASIC program. The Communication File may be in the study currently being accessed, in a different study, or in a BASIC account.

The study or BASIC account that contains the Communication File is specified by entering the password for that study or account. Thus, the transfer must be effected by, or in cooperation with, someone who has access to the file.

Each study may have up to ten Communication Files, designated by numbers from 0 to 9. As mentioned above, the user generally specifies a file in another study or in a BASIC account, rather than one in the study currently being accessed. The actual name of a Communication File (for purposes of processing by BASIC) is COMMFILEn, where n is specified by the user.

A Communication File that can be copied into a worksheet by this subactivity contains the size, labels, and data for a single worksheet. The format is two numbers followed by two character strings followed by a matrix:

\[
\begin{align*}
&\text{r} \quad (\text{the number of matrix (i.e., worksheet) rows}) \\
&\text{c} \quad (\text{the number of matrix (i.e., worksheet) columns}) \\
&\text{an} \quad r \times 8 \quad \text{character string (the row labels)} \\
&\text{a} \quad c \times 8 \quad \text{character string (the column labels)} \\
&\text{an} \quad r \times c \quad \text{matrix (the worksheet data)}
\end{align*}
\]

where the number \(9.99999E-07\) represents a missing value.

When a Communication File is read, an attempt is made to ensure that the file is properly formatted, but this does not guarantee that it will be. The system notifies the user if it determines that the file is badly formatted, e.g., if r and c do not match the matrix dimensions.

Dialog

The Communication File is specified by entering the file number and password. To maintain privacy, the system does not display the password. Next, the worksheet to be created from the Communication File is named and given a title. Finally, the user has the option of displaying the worksheet.

```
retrieve,6 COMMUNICATION FILE TO WORKSHEET
THE COMMUNICATION FILE NUMBER PLEASE (0 TO 9): 3
PASSWORD FOR THE STUDY CONTAINING THE COMMUNICATION FILE:
WORKSHEET NAME PLEASE : ws2
TITLE: Data in worksheet ws1 that have been processed by a BASIC program ws2 IS THE WORKSHEET'S NAME (IT MAY HAVE BEEN TRUNCATED)
TYPE yes TO DISPLAY THE WORKSHEET, ELSE HIT [RETURN]: yes
```
In this example, the retrieved Communication File is the one created and processed in the example under retrieve, option 5. The resultant worksheet is shown below.

**WORKSHEET WS2**

**TITLE** Data in worksheet WS1 that have been processed by a BASIC program

**CREATED 11/17/77**  **MODIFIED 11/17/77**

<table>
<thead>
<tr>
<th># OF ROWS</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td># OF COLS</td>
<td>5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ROWS/COLS</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>col#1</td>
<td>col#2</td>
<td>col#3</td>
<td>col#4</td>
<td>col#5</td>
</tr>
<tr>
<td>1 row#1</td>
<td>4</td>
<td>3.500</td>
<td>5.200</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>2 row#2</td>
<td>3</td>
<td>2</td>
<td>25</td>
<td>0</td>
<td>23</td>
</tr>
<tr>
<td>3 row#3</td>
<td>17</td>
<td>4.700</td>
<td>17.300</td>
<td>48.300</td>
<td>18</td>
</tr>
</tbody>
</table>
The subset activity allows the user to specify, create, and manipulate sets of patients who have common characteristics.

Patients are placed in a set because their data in the SDF or in a worksheet satisfy user-specified criteria or because they are members of one or more existing sets. Because the selection criteria may be difficult to specify, a set may also be created simply by naming its constituent patients. Sets are created in order to determine which patients in a study satisfy certain criteria of interest or to define groups of patients for whom data are to be retrieved.

All patients in a study who are known to the system (i.e., their abbreviations have been entered using enter, option 0) belong to the set called all. A patient must be a member of this set before he can become a member of any other set.

Depending on the option used, a patient may be placed in a set if all of a list of criteria are satisfied or if any is satisfied. A criterion is satisfied if a number, code, or time falls within a specified range or if an encoded character string has a specified character-string value. Presently, it is not possible to create sets based on dates or stored textual information.

When patients are selected on the basis of their SDF data, those data are retrieved (just as they are by retrieve, option 2, which retrieves data from the SDF for storage in worksheets) and are examined. The user specifies a data collection time interval during which the criteria must be satisfied. As in retrieve, option 2, the time interval may be the time between two dates or events, a list of time units following a date or an event, or a range of time units following a date or an event. Because a worksheet is smaller and simpler in structure than the SDF, it is both simpler and faster to create sets based on worksheet data. Thus, if more than one set is to be created based on the values of a few items, it is often most efficient to retrieve the data into a worksheet and then use subset, option 6 or 7.

In addition to creating sets, the subset activity provides for listing the names and descriptions of sets, for listing the members of a patient set, and for discarding a set. A set is saved permanently until discarded. The set called all may not be discarded, but the user may create and discard other sets at will.

CLINFO treats set names as if they are comprised of all numbers and uppercase letters. Thus, set names NAME, Name, and name are equivalent. Also, like worksheet names, set names may contain up to 8 numbers and letters; they may not contain punctuation marks or imbedded blanks.

All the examples in this section apply to the data described in the example of Sec. III.

Dialog

A number of subactivities are available within the subset activity. These subactivities are presented in a menu; the user chooses among them by typing an option number. The menu has the form:
A response causes the appropriate subactivity to be invoked. Each subactivity is described below.

**OPTION 1: A LIST OF PATIENT ABBREVIATIONS**

This subactivity allows the user to create a patient set by naming its constituent patients. Each patient must be a member of the set called all, i.e., his abbreviation must have been previously entered using enter, option 0.

**Dialog**

The interaction is illustrated below. After the user names the set he is creating, he supplies the list of patient abbreviations. A patient abbreviation that is not already known to the system is rejected.

The patient list and the use of this subactivity are terminated by responding to

**PATIENT ABBREV:**

by hitting the RETURN key or entering / followed by an activity name and a subactivity number. The user is then prompted for a comment that is saved as part of the set's description.
subset,1 SET CREATION BY LIST OF PATIENT ABBREV.

NEW SET NAME PLEASE: set1
set1      IS THE SET’S NAME (IT MAY HAVE BEEN TRUNCATED)

TO TERMINATE THE LIST, HIT [RETURN]

PATIENT ABBREV.: case 1
PATIENT ABBREV.: case 2
PATIENT ABBREV.: case 35
PATIENT case 35 IS NOT IN PATIENT FILE
PATIENT ABBREV.: case 3
PATIENT ABBREV.:  

IDENTIFYING COMMENTS: First three cases

OPTION 2: PATIENTS BELONGING TO ALL OF A LIST OF SETS

This subactivity enables the user to create a set of those patients who belong to all of a list of existing sets. After naming the new set to be created, the user enters the list of existing sets that the patients must belong to. The complement of a set may be entered to indicate that the patient must not belong to that set. The system lists the patients satisfying the criteria and creates the new set.

The complement of a set is indicated by preceding a listed set name with [not] as in [not] odd. If the patients in the study have the abbreviations case 1, case 2, case 3, case 4, etc., and patients called case 1, case 3, etc., are members of the set called odd, the entering of [not] odd means that case 2, case 4, etc., are candidates for inclusion in the new set, but case 1, case 3, etc., are not.

Dialog

The user first names the set he is about to create and may then display existing set names.

subset,2 PATIENTS BELONGING TO ALL OF A LIST OF SETS

NEW SET NAME PLEASE: set2

subset,2 PATIENTS BELONGING TO ALL OF A LIST OF SETS
SET :SET2

TYPE yes TO DISPLAY YOUR SET NAMES: yes

ALL
WSSET
EVEN
LOWGLUC
SET1

In the figures here and under option 3, the list of sets that define the new set appears at the left and the list of all existing sets appears at the right. As each set (or its complement) is specified, its name is verified, and it is then added to the list of sets that define the new set.
SUBSET, 2

subset,2 PATIENTS BELONGING TO ALL OF A LIST OF SETS
SET :SET2

SELECT A PATIENT IF A MEMBER OF THE SETS

<table>
<thead>
<tr>
<th>SET1</th>
<th>ALL</th>
</tr>
</thead>
<tbody>
<tr>
<td>LOWGLUC</td>
<td>WSSET</td>
</tr>
<tr>
<td>EVEN</td>
<td>LOWGLUC</td>
</tr>
<tr>
<td>SET1</td>
<td></td>
</tr>
</tbody>
</table>

LIST OF SETS. HIT [RETURN] TO TERMINATE THE LIST
SET NAME (E.G. all OR [not]all):
even IS THE SET NAME (IT MAY HAVE BEEN TRUNCATED)

Once he has listed the constituent sets, the user may enter a comment to help identify the set:

IDENTIFYING COMMENTS: set1 AND lowgluc AND NOT even

The members of the set SET1 are case 1, case 2, and case 3; the members of lowgluc are case 1, case 2, case 3, and case 4. The cases in the set even are case 2, case 4, and case 6. The patients who belong to both of the former sets and not the latter and who become members of the new set, SET2, are therefore case 1 and case 3.

subset,2 PATIENTS BELONGING TO ALL OF A LIST OF SETS
SET :SET2

PATIENTS SATISFYING THE CRITERIA

case 1
case 3

THERE ARE 2 PATIENTS IN SET SET2 . HIT [RETURN] TO CONTINUE:

OPTION 3: PATIENTS BELONGING TO AT LEAST ONE OF A LIST OF SETS

With this subactivity, the user may create a set of those patients who belong to any one of a list of sets. After naming the set to be created, he enters the list of sets that the patients may belong to. As explained in the description of option 2, the notation [not]set-name may be used to indicate that patients who are not members of the set so designated should be included in the new set. The system lists the patients who are members of at least one of the specified sets or complements of sets and creates the new set.

Dialog

The interaction is similar to that in option 2. The user first names the set he is about to create. After displaying the list of his existing sets (or declining to do so), he enters the names of those sets whose members are to become members of the new set.
subset,3 patients belonging to at least one of a list of sets
set : set3

select a patient if a member of the sets

set1
or
lowgluc
or [not] even

list of sets. hit [return] to terminate the list
set name (e.g. all or [not]all): [not] even

lowgluc is the set name (it may have been truncated)

finally, he adds identifying comments:

subset,3 patients belonging to at least one of a list of sets
set : set3

select a patient if a member of the sets

set1
or
lowgluc
or [not] even

identifying comments: set1 or lowgluc or not even

patients case 1, case 2, and case 3 belong to the set called set1; patients case 1, case 2, case 3, and case 4 are members of the set lowgluc; and case 2, case 4, and case 6 belong to the set even (thus, case 1, case 3, case 5, and case 7 are not in even.) the above specification, therefore, results in including all of these patients except case 6 in the new set, set3.

option 4: patients whose sdf data satisfy all conditions in a list

this subactivity enables the user to specify a set of conditions that numbers, times, and character-string codes in the sdf must satisfy during a prescribed time interval; to cause the system to retrieve and examine relevant data; and to create a set that comprises those patients whose sdf data satisfy all the conditions.

the subsetting specification has three aspects:

• the patient set from which the patients are selected.
• the required values of items.
• the time interval during which items must take on their required values.

the first aspect is satisfied by naming the source patient set. the second aspect is accomplished by listing the items and inclusive value ranges of interest. when an
item has a number or time as its value, two numbers or times (within the same day) separated by a comma are entered to denote an inclusive range. For example, if the value of the item glucosec must be at least 100 and no greater than 200, glucosec is named as the item and 100,200 is entered as the range; if it must be at least 250, the range 250,999 is entered, assuming that 999 is greater than any actual value of glucosec; if the value of glucosec must be exactly 225, then this is designated by the range 225,225.

If an item is coded, i.e., if its values are predefined character strings that are encoded as integers, its required values may be specified as an inclusive range of integer codes as just described, or the required character-string values may be listed. For example, an item called color may have the character-string values red, yellow, and blue, represented by the integer codes 0, 1, and 2, respectively. If, for subsetting purposes, the values of color must be red or yellow, this may be indicated by specifying the range 0,1. If the value of color must be blue, this could be specified by designating either the range 2,2 or the specific character-string value blue. Each specified character string results in a separate condition to be satisfied when data are retrieved and examined. Because this subactivity is used when all conditions must be satisfied, it cannot be used to indicate that the value of color must be either red or yellow; subset, option 5, is suitable for that purpose.

To indicate that glucosec must be between 100 and 200 and that color must be red, the user would enter the item name glucosec and the range 100,200 and would also enter the item name color and the character string red. To select patients for whom the value of glucosec is at least 300 and the value of color is either red or yellow, the user might first use subset, option 5, to create the set of patients for whom color is either red or yellow and then use either option 4 or option 5 to select from that set those patients whose glucosec value is at least 300. Alternatively, he could use this subactivity (option 4) to enter the item name glucosec and the range 300,999 and also to enter the item name color and the integer code range 0,1.

To create a set of patients whose values of particular items are missing, the user would enter the item name, then when asked for a range or code, he would respond with . . . (i.e., 3 periods).

The third aspect of the subsetting specification requires the definition of a time interval during which the items must conform to the ranges entered by the user. The time interval may be specified as

- The time between two dates or events, e.g., between begin (the earliest data collection time) and end (the time the last data were collected), or between 12/28/76 at 1430 and 1/16/77 at 0100,
- A list of time units relative to a date or an event, e.g., during the 5th, 7th, or 19th day following transplant, or during the 3rd and 5th weeks following 11/23/76 at 1200 hours, or
- A range of time units relative to an event, e.g., between the 8th and 12th hours following insulin withdrawal, or following a specific date and time.

The time units may be minutes, hours, days, 7-day weeks, 28-day months, 365-day years, or a multiple number of any of these units; the events may be begin, end, or user-defined events in the schema. In the simplest case, the user does not care when the items take on their required values, so he can choose the time interval between begin and end, i.e., the entire study period. In the most restrictive case,
the user can require that the items take on their required values during a specific day, hour, or minute. The specification and meaning of time intervals are detailed in the discussion of retrieve, option 2.

The user may require that (1) each item may independently take on a required value any time during the same specified time interval or (2) the items must take on the desired values simultaneously, i.e., in the same instance of a panel. In the first case, if the conditions are that glucose must be in the range 100,200 and color must be red during the time interval 8 through 12 hours following insulin withdrawal, a patient will be selected if his glucose value is 175 at hour 9 but greater than 210 otherwise and his color value is red at hour 12 but yellow otherwise. That is, the patient is selected even though the conditions are satisfied at different times. The way to ensure that all items take on their required values simultaneously is to specify simultaneous occurrence. This requires that the items be in the same schema panel.

Dialog

When using this option, the user first names the set he is about to create and then names the set from which patients are to be selected. In this example, the user wants to select those patients who are male and who require high amounts of insulin (at least 45 units). He enters the new set name highins; then, because he wants to search the data for all the patients on the study, he indicates that he wants to use patients from the set called all. The search could be restricted to some other previously defined set by entering the name of that set. After the user hits the RETURN key, the terminal display screen looks like:

```
subset,4 SELECT PATIENTS SATISFYING ALL CONDITIONS
SOURCE SET: ALL
RESULT SET: HIGHINS

A TIME INTERVAL WILL BE BROKEN DOWN INTO UNIT-LONG SUBINTERVALS, MAKE A UNIT CHOICE

<table>
<thead>
<tr>
<th>TYPE TO CHOOSE THE UNIT</th>
</tr>
</thead>
<tbody>
<tr>
<td>0...RETRIEVE ALL DATA</td>
</tr>
<tr>
<td>1...MINUTE</td>
</tr>
<tr>
<td>2...HOUR</td>
</tr>
<tr>
<td>3...DAY</td>
</tr>
<tr>
<td>4...7=DAY WEEK</td>
</tr>
<tr>
<td>5...30-DAY MONTH</td>
</tr>
<tr>
<td>6...365-DAY YEAR</td>
</tr>
<tr>
<td>7...MULTIPLE UNITS</td>
</tr>
</tbody>
</table>

UNIT CHOICE: 0
```

The user is now asked to specify the time interval during which conditions must be satisfied. Because he is willing to have the conditions satisfied at any time during the study, and because he knows that values for sex and ins req are recorded only once near the beginning of the study, he chooses retrieve all data. With this type of interval specification, the time unit choice (which will be days) actually does not matter. More complex restrictions may be specified, exactly as in retrieve, option 2.
OPTION 5: PATIENTS WHOSE SDF DATA SATISFY ANY OF THE CONDITIONS IN A LIST

This subactivity enables the user to specify sets of conditions that numbers, times, and character-string codes in the SDF must satisfy during a prescribed time interval; to cause the system to retrieve and examine relevant data; and to create a set comprised of those patients whose SDF data satisfy any of the conditions.

The subsetting specification is identical to that in option 4 above, comprising

- The patient set from which the patients are selected.
- The required values of items.
- The time interval during which items must take on their required values.

These aspects are detailed under subset, option 4. Except for the fact that under option 4 patients are selected only if all of a series of conditions are satisfied, and under option 5 patients are selected if any one condition is satisfied, the two options are very similar. Therefore, we shall describe the subsetting specifications only briefly here.

The first aspect, the source patient set, is completed by naming the existing patient set from which patients are to be selected.

The second aspect requires the specification of items of interest and their required numerical ranges or character-string values. An inclusive range of numbers, times, or integer codes is specified by two such values separated by a comma. A required character-string value is specified by the string itself. For example, to select patients who either have glucosec values between 100 and 200 or have red color, the user would enter the item name glucosec with the range 100, 200 and the item name color with the character string red. To select patients whose glucosec value is at least 300 or whose color is either red or yellow, he would enter the item name glucosec with the range 300,999 (if 999 is outside the possible range), the item name color with the character string red, and the item name color (again) with the character string yellow. To select patients whose item value is missing, the user would enter . . . (i.e., 3 periods) when asked for the required range or code.

The third aspect requires the specification of a time interval, namely,

- The time between two named events or between two dates and time,
- A list of time units relative to an event (or date), or
- A range of time units relative to an event (or date).

The time unit may be a 365-day year, 30-day month, 7-day week, day, hour, minute, or a multiple number of any of these units. The event may be begin or end (the two system-defined events) or a user-defined event. The time interval specification restricts the time period over which data values are retrieved and examined. In order for a patient to be selected, at least one of the item value conditions must be satisfied at some time during this time period, or if indicated by the user, the conditions on items in the same panel must be satisfied simultaneously.

Dialog

Aside from the statement of the conditions as a series of IF . . . OR, IF . . . statements, the dialog is identical to that in option 4.
OPTION 6: PATIENTS WHOSE WORKSHEET DATA SATISFY ALL CONDITIONS IN A LIST

This subactivity enables the user to specify one or more conditions that data values stored in a worksheet must satisfy, and to create a set which comprises those patients whose data satisfy all the conditions. A condition is defined by specifying the label or number for a worksheet row or column and a range of allowable data values. Because data have already been retrieved (or otherwise entered) into a worksheet, it is not necessary to specify either a set from which patients are to be selected or a time interval. Also, because worksheet data values can be entered directly from the terminal into the worksheet or can be computed and stored in the worksheet, and because they need not correspond to items in the schema, this activity provides added flexibility. This method of subsetting is not only simple, but fast.

This option may be used to create a set with the patients in some order other than that in which they were entered using enter, option 0. For example, if patient abbreviations are row labels of a worksheet that is sorted based on values in a column and a subset of these patients is created, the resultant set will list the patients in the order in which they appear in the sorted worksheet. If this new set is used in a retrieval, data will be retrieved (using retrieve, option 2) in that same order and may be associated with the data in the sorted worksheet.

The user first enters the name of the set he wants to create. Next, he specifies the worksheet that contains the data values to be tested. Because the only information used to select the patient set is contained in the worksheet, either its row or column labels must be patient abbreviations (previously entered using enter, option 0).

After indicating whether patient abbreviations are row labels or column labels, the user specifies the conditions by entering row labels or numbers (if patient abbreviations are column labels) or column labels or numbers (if patients are in rows) and ranges of allowable values. A label need not be the name of an item defined by the schema. Normally, two numbers separated by a comma are entered (e.g., 0.13) to specify an acceptable range of values. However, if a label corresponds to a coded schema item, a character-string code may be entered to specify the condition. To indicate that a patient is to be selected if the value in the specified row or column is missing, . . . (i.e., 3 periods) is entered.

After the user has entered the conditions, verified them, and entered a comment, the system creates the set.

Dialog

In this example the worksheet called CLINIC will be used in specifying the patient set. Note that row labels correspond to patient abbreviations, that labels in columns 1 through 5 correspond to schema items, and that the label in column 6 does not correspond to a schema item because the values in that column were computed (as #6 = #1 - #4 in calculate).
Next, the user is asked to specify a series of conditions, i.e., a series of inclusive ranges of item values. After entering the item name \textit{ins req}, he is prompted for the required inclusive range and responds 45,999 (i.e., greater than or equal to 45), as shown.

\textbf{TO TERMINATE LIST OF CONDITIONS, HIT [RETURN]}

\textbf{ITEM: ins req}

\textbf{INCLUDE A PATIENT:}

\textbf{IF ins req IS IN A RANGE (E.G. 3,6) OR IS A CODE (E.G. m): 45,999}

After similarly entering the item \textit{sex} and the value \textit{m}, the user hits the RETURN key to indicate that he has finished specifying conditions. The specification has the form IF ... AND IF ... to emphasize that both conditions must be satisfied.

\textbf{INCLUDE A PATIENT:}

\textbf{IF} \quad 45 \leq \text{ins req} \leq 999

\textbf{AND IF} \quad \text{sex} = \text{m}

Finally, he is prompted:

\textbf{TYPE yes IF CONDITIONS WITHIN A PANEL MUST BE SATISFIED SIMULTANEOUSLY,}

\textbf{OR HIT [RETURN] IF THERE IS NO SUCH RESTRICTION:}

In this case, he hits the RETURN key, since each of these items occurs only once.

As the search proceeds, the system indicates the panel and patient that are being examined, and it builds a list of patients whose data satisfy all the conditions. When data for all the patients in the source (i.e., the "select from") set have been examined, the system indicates the number of selected patients and then halts. If no patients satisfy all the conditions, the system so indicates and does not create a new patient set; otherwise, it prompts for a comment which it saves with the new set name and list of selected patients.

\textbf{SOURCE SET: ALL}
\textbf{RESULT SET: HIGHINC}
\textbf{TIMES ARE} \quad 1-day \quad \text{sub-int's FROM EVENT begin TO EVENT end}
\textbf{(FROM THE START OF THE day CONTAINING THE EVENT begin)}
\textbf{SELECTING ALL} \quad \text{MEASURED DURING ALL THE 1-day sub-int's}

\textbf{INCLUDE A PATIENT:}

\textbf{IF} \quad 45 \leq \text{ins req} \leq 999

\textbf{AND IF} \quad \text{sex} = \text{m}

\textbf{IDENTIFYING COMMENTS: Insulin req} \geq 45 \text{ AND male}

case 2

case 3

case 4

case 7
After naming the set **WSI**, the user specifies the worksheet and then indicates that row labels are patient abbreviations by responding:

**WORKSHEET NAME PLEASE ([RETURN] SELECTS CURRENT): clinic**

and

**TYPE c IF PATIENTS ARE IN COLUMNS, r IF IN ROWS: r**

Next, after listing the column labels (or declining to list them), he indicates that the value of **diab age** must be less than or equal to 13.

**TO TERMINATE LIST OF CONDITIONS, HIT [RETURN]**

**COLUMN LABEL OR NUMBER: diab age**

1 age
2 sex
3 idl wt
4 dur diab
5 ins req
6 diab age

**IF dur diab IS IN A RANGE (E.G. 3,6) OR IS A CODE (E.G. m): 0,13**

After the second (and last) condition is entered, the selection specifications appear as shown.

**INCLUDE A PATIENT:**

**IF 0 <= diab age <= 13**

**AND IF 40 <= ins req <= 70**

After the user hits the RETURN key to verify the specifications and enters an identifying comment, the system selects the patients whose data satisfy the specifications and creates the set.
subset, 6 SELECT PATIENTS SATISFYING ALL CONDITIONS IN A WORKSHEET
SET :WS1

INCLUDE A PATIENT:

IF 0 <= diab age <= 13
AND IF 40 <= ins req <= 70

THERE ARE 3 PATIENTS IN SET WS1, HIT [RETURN] TO CONTINUE:

case 2
case 5
case 7

OPTION 7: PATIENTS WHOSE WORKSHEET DATA SATISFY AT LEAST ONE CONDITION IN A LIST

This subactivity enables the user to specify a set of conditions that data values stored in a worksheet must satisfy, and to create a set comprising those patients whose data satisfy at least one of the conditions. The data values examined may have been retrieved, computed, or entered from the terminal. This subactivity is substantially the same as option 6 above, except that option 6 is used when all conditions must be satisfied, while option 7 is used when any one condition must be satisfied.

After naming the set to be created, the user specifies the worksheet of interest. Because patient abbreviations must appear as either row or column labels, he next specifies which. The specifications are completed by entering labels or numbers to identify the data and by entering numbers to specify allowable ranges. The labels need not correspond to schema item names, but if a label is the name of a coded item, a character-string code may be entered. Three periods (...) may be entered to indicate that a missing value is acceptable.

OPTION 8: LIST PATIENT ABBREVIATIONS FROM A SET

This subactivity enables the user to list the patients who are members of a specified set. In particular, by specifying the set called all, he can obtain a list of all the patients who are known (to the system) to be in the study.

Dialog

When the user enters a set name, the system responds by displaying the set description and the list of its constituent patients. The user may repeat this for another set by responding to the last prompt, and (when the subactivity is reinitiated) by entering another set name. If the RETURN key is hit in response to the last prompt, the menu of subset options is displayed.
subset, 8 LIST OF PATIENT ABBREVIATIONS

EXISTING SET NAME PLEASE ([RETURN] SELECTS all):
ALL IS THE SET'S NAME (IT MAY HAVE BEEN TRUNCATED)

SET NAME: ALL, DATE CREATED: 11/26/75, 7 ENTRIES

    case 1    case 2    case 3    case 4    case 5    case 6    case 7

END OF ENTRIES, HIT [RETURN] TO CONTINUE:

subset, 8 LIST OF PATIENT ABBREVIATIONS

EXISTING SET NAME PLEASE ([RETURN] SELECTS all): lowgluc
LOWGLUC IS THE SET'S NAME (IT MAY HAVE BEEN TRUNCATED)

SET NAME: LOWGLUC, DATE CREATED: 3/8/76, 4 ENTRIES
female OR glucose<100 prior to insulin withdrawal

    case 1    case 2    case 3    case 4

END OF ENTRIES, HIT [RETURN] TO CONTINUE:

OPTION 9: LIST ALL OF YOUR SUBSET NAMES

This subactivity is used to list the names and descriptions (date created, number of constituent patients, and comments) of all the patient sets associated with the current study. Option 2 of the files activity is used to list the set names in a more compact format and in order by name, date, or size.

Dialog

The names and descriptions of the sets in the study are displayed as soon as the subactivity is invoked. If there are more set names than the screen can display, the system indicates this, and the user may hit the RETURN key to display them. The end of the directory is also indicated; when the RETURN key is hit in this case, the menu of subset options is displayed.

subset, 9 LIST OF SETS

SET NAME: ALL, DATE CREATED: 11/26/75, 7 ENTRIES

SET NAME: ODD, DATE CREATED: 1/28/76, 4 ENTRIES
odd numbered cases

SET NAME: EVEN, DATE CREATED: 1/28/76, 3 ENTRIES
even numbered cases

SET NAME: YOUNGINS, DATE CREATED: 1/28/76, 3 ENTRIES
age<30 & ins reg>40

SET NAME: HIGHGLUC, DATE CREATED: 1/28/76, 2 ENTRIES
in set youngins & 8 hr control glucose>=200

SET NAME: OLDIDOB, DATE CREATED: 1/28/76, 4 ENTRIES
age>=30 OR %id1 wt=100 OR dur diab>=15

SET NAME: LOWGLUC, DATE CREATED: 1/29/75, 4 ENTRIES
female OR glucose<100 prior to insulin withdrawal

END OF DIRECTORY, HIT [RETURN] TO CONTINUE:
OPTION 10: DISCARD A PARTICULAR SET

This subactivity is used to discard a patient set that is no longer needed. The user should use subset, option 9, periodically to review his sets and should use option 10 to discard any that are unnecessary.

The set called all cannot be discarded.

Dialog

The user names an existing set (except all), then types yes to verify that he wants to discard it. The system discards the set, informs the user of this, and then reprompts for a subset option.

If the user inadvertently types the name of the wrong set, he may hit the RETURN key instead of typing yes. This prevents the set from being discarded.

EXISTING SET NAME PLEASE: all
ALL IS THE SET'S NAME (IT MAY HAVE BEEN TRUNCATED)
ALL PATIENTS CANNOT BE DELETED

EXISTING SET NAME PLEASE: odd
odd IS THE SET'S NAME (IT MAY HAVE BEEN TRUNCATED)
TYPE yes TO DISCARD SET odd ELSE [RETURN]:

SUBSET CHOICE: 10
EXISTING SET NAME PLEASE: add
add IS THE SET'S NAME (IT MAY HAVE BEEN TRUNCATED)
TYPE yes TO DISCARD SET add ELSE [RETURN]: yes
SET add DISCARDED

SUBSET CHOICE:
A worksheet is a labeled, rectangular array of numeric data values. The worksheet activity provides for creating, labeling, entering data into, displaying, listing, sorting, printing, editing, (i.e., rearranging), and discarding worksheets. All worksheet operations performed by this and other activities are applied to the current worksheet, which is selected by using option 0.

CLINFO worksheets may be used independently of the retrieval of information from the Study Data File (SDF). That is, in addition to retrieving data from the SDF, data can be entered directly into a worksheet, and columns or rows may be labeled from the terminal. Worksheets are automatically saved and may be retrieved for analysis or display. They are created by use of option 0.

The worksheet name (by which it is recalled) is entered by the user and may be any combination of up to eight letters and numbers. The name will be truncated if it is longer than eight characters or if it contains a blank or special symbol. Note that in the case of worksheet names, unlike item or panel names, the system does not distinguish between upper and lower case, so that ABLE, ABE, and able are all equivalent.

The following table indicates the maximum number of rows (columns) for a given number of columns (rows) and the total number of data items in the resultant worksheet.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>799</td>
<td>799</td>
<td>2</td>
<td>599</td>
<td>1198</td>
<td>3</td>
<td>478</td>
<td>1434</td>
<td>4</td>
<td>398</td>
<td>1202</td>
<td>5</td>
<td>341</td>
<td>1705</td>
</tr>
<tr>
<td>6</td>
<td>298</td>
<td>1788</td>
<td>7</td>
<td>265</td>
<td>1555</td>
<td>8</td>
<td>238</td>
<td>1904</td>
<td>9</td>
<td>216</td>
<td>1944</td>
<td>10</td>
<td>190</td>
<td>1900</td>
</tr>
<tr>
<td>11</td>
<td>182</td>
<td>2002</td>
<td>12</td>
<td>169</td>
<td>2028</td>
<td>13</td>
<td>158</td>
<td>2054</td>
<td>14</td>
<td>148</td>
<td>2072</td>
<td>15</td>
<td>139</td>
<td>2005</td>
</tr>
<tr>
<td>16</td>
<td>131</td>
<td>2096</td>
<td>17</td>
<td>124</td>
<td>2108</td>
<td>18</td>
<td>118</td>
<td>2124</td>
<td>19</td>
<td>112</td>
<td>2128</td>
<td>20</td>
<td>107</td>
<td>2149</td>
</tr>
<tr>
<td>21</td>
<td>102</td>
<td>2194</td>
<td>22</td>
<td>98</td>
<td>2156</td>
<td>23</td>
<td>94</td>
<td>2166</td>
<td>24</td>
<td>90</td>
<td>2160</td>
<td>25</td>
<td>87</td>
<td>2175</td>
</tr>
<tr>
<td>26</td>
<td>83</td>
<td>2158</td>
<td>27</td>
<td>80</td>
<td>2168</td>
<td>28</td>
<td>78</td>
<td>2184</td>
<td>29</td>
<td>75</td>
<td>2175</td>
<td>30</td>
<td>73</td>
<td>2198</td>
</tr>
<tr>
<td>31</td>
<td>70</td>
<td>2170</td>
<td>32</td>
<td>68</td>
<td>2178</td>
<td>33</td>
<td>66</td>
<td>2178</td>
<td>34</td>
<td>64</td>
<td>2176</td>
<td>35</td>
<td>62</td>
<td>2170</td>
</tr>
</tbody>
</table>

Maximum Worksheet Dimensions

If I is the number of rows, then J is the maximum number of columns (or vice versa). I x J is the number of data elements in the worksheet. Certain ranges of row dimensions yield the same maximum number of columns; e.g., for from 100 to 112 rows there may be up to 19 columns.
Dialog

The system first presents the following display:

THE CURRENT WORKSHEET IS EMP
TITLE emp data
CREATED 3/29/77    MODIFIED 3/29/77
# OF ROWS 2
# OF COLS 2

WORKSHEET OPTIONS
-----------------------------------------------
TYPE FOR WORKSHEET 1
0...SELECTION
1...DISPLAY
LABELLING BY
2...ROWS
3...COLUMNS
DATA ENTRY BY
4...ROWS
5...COLUMNS
6...CELL
7...DELETION
8...SORTING
9...LISTS
10...PRINTING
11...EDITING
-----------------------------------------------

WORKSHEET OPTION:

If there were no current worksheet, the upper message would be:

THERE IS NO CURRENT WORKSHEET

and the user could select option 0.

A worksheet's modified date is changed to the current date whenever worksheet, option 2 through 6, 8, or 11, is selected. Since a worksheet is not protected from modifications, the date indicates when it was last accessed by an option that could modify it.

The worksheet that was current at the end of a session will be recalled when the user next logs on to the system.

OPTION 0: SELECT A WORKSHEET

This option permits the user to create a new worksheet having a specified name, title (i.e., descriptive commentary), number of rows, and number of columns. It also provides for selecting the current worksheet, i.e., the one that is to be subsequently processed. After this option is chosen, the dialog proceeds:

WORKSHEET NAME PLEASE: able

If no worksheet named ABLE had been created previously and the user wished to create one, he would respond yes to the next prompt:

TYPE yes TO CREATE A WORKSHEET, ELSE [RETURN]; yes
able IS THE WORKSHEET'S NAME (IT MAY HAVE BEEN TRUNCATED)
The name is echoed to verify the user's intention.

The system then prompts for a title (of up to 70 characters) and for the dimensions of the worksheet:

**TITLE:** Frequency data for male/female glucose study

**THE NUMBER (1 to 799) OF ROWS OR COLUMNS (E.G. 3 rows OR 7 cols):** 3 rows

**FOR 3 ROWS THE MAXIMUM # OF COLUMNS IS 478, # OF COLUMNS: 2**

Had the user responded 3 cols to the first worksheet-dimensions prompt, the second prompt would have been for the number of rows. Appropriate responses to the first prompt are 3 or 3 rows to specify rows first, or 2 cols to specify columns first; if neither rows nor columns is indicated, rows is assumed. The previous current worksheet display is then replaced by the name, title, and dimensions of worksheet ABLE, indicating that it has been created, and the system again prompts:

**WORKSHEET NAME PLEASE:**

This prompts the user to create a series of worksheets without reentering option 0. To label the columns of the new worksheet, for example, the user would type ABC.

Had the worksheet ABLE already existed prior to entering this option, its name and description would have appeared at the top of the screen as soon as it was selected, followed by the prompt:

**WORKSHEET OPTION:**

If the user responds with the RETURN key to

**TYPE yes TO CREATE A WORKSHEET, ELSE [RETURN]:**

then there is no current worksheet, and the system prompts:

**WORKSHEET OPTION:**

**OPTION 1: DISPLAY THIS WORKSHEET**

This option displays up to 20 rows and 6 columns of the current worksheet, starting with a specified row and column. Numeric codes for categorical items described in the schema may be displayed as character strings. Option 1 may be selected at any time by entering /I if the user is already in the *worksheet* activity or by entering /work, I if in another activity. Worksheets may also be displayed under Options 2, 4, and 6 of *retrieve* and under all *calculate* options if that is where they are created or modified.

**Dialog**

```
WORKSHEET CARDIO
TITLE Cardiology Study
CREATED 1/21/76 MODIFIED 1/21/76
# OF ROWS 28
# OF COLS 12

TYPE yes TO DECODE DATA IN THIS WORKSHEET, ELSE HIT [RETURN]:

TYPE yes TO SELECT STARTING ROW AND COLUMN, ELSE [RETURN]:
```
In this example, the name of the current worksheet is CARDIO; it was created on 1/21/76 (i.e., on January 21, 1976) and has not been modified since that date. The DECODE prompt appears if there is a compiled schema for the study; otherwise it does not. The user responded with the RETURN key to both system prompts; the effect of responding yes is explained below. The system displays up to 6 columns and up to 20 rows of the current worksheet, starting with row 1 and column 1:

1 / 21 /76, at 17:11

WORKSHEET CARDIO
TITLE Cardiology Study
CREATED 1/21/76 MODIFIED 1/21/76
\# OF ROWS 20
\# OF COLS 12

ROWS/COLS

<table>
<thead>
<tr>
<th>LABELS</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>CASE 1</td>
<td>7007</td>
<td>4/25/1970</td>
<td>51</td>
<td>0</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>CASE 2</td>
<td>7008</td>
<td>4/29/1970</td>
<td>...</td>
<td>0</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>CASE 3</td>
<td>7009</td>
<td>4/29/1970</td>
<td>17</td>
<td>0</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>CASE 4</td>
<td>7010</td>
<td>5/3/1970</td>
<td>69</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>
| CASE 5 | 7011 | 5/7/1970 | 49 | 0 | 2 | ...
| CASE 6 | 7013 | 5/17/1970 | 64 | 0 | 2 | 1 |
| CASE 7 | 7020 | 6/10/1970 | 67 | 1 | 2 | 1 |
| CASE 8 | 7024 | 7/3/1970 | 58 | 0 | 2 | 1 |
| CASE 9 | 7027 | 7/11/1970 | 67 | 0 | 1 | 1 |
| CASE 10 | 7028 | 7/26/1970 | 43 | 0 | 2 | 1 |
| CASE 11 | 7030 | 8/16/1970 | 55 | 0 | 2 | 1 |
| CASE 12 | 7032 | 8/25/1970 | 75 | 1 | 2 | 1 |
| CASE 13 | 7033 | 9/14/1970 | 70 | 0 | 2 | 1 |
| CASE 14 | 7034 | 9/15/1970 | 53 | 1 | 1 | 1 |
| CASE 15 | 7038 | 9/27/1970 | 65 | 1 | 2 | 1 |
| CASE 16 | 7043 | 10/15/1970 | 55 | 0 | 2 | 1 |
| CASE 17 | 7045 | 10/21/1970 | 58 | 0 | 1 | 1 |
| CASE 18 | 7047 | 10/29/1970 | 57 | 0 | 2 | 1 |
| CASE 19 | 7049 | 10/29/1970 | 56 | 0 | 2 | 1 |
| CASE 20 | 7056 | 12/2/1970 | 72 | 1 | 2 | 1 |

[NOTE: ... indicates missing values]

TYPE yes TO CONTINUE WITH worksheet ELSE
TYPE THE STARTING ROW #: 1
STARTING COL #: 7

Missing values are represented by ellipsis dots (i.e., ..., as in row 2, column 3). These dots are displayed in worksheet locations where no values have been entered; they should not be confused with zero, which represents an entered value. Dates are displayed with month, day, and year separated by slashes (e.g., 4/25/1970 represents April 25, 1970).

To display more of the worksheet (since this worksheet has 12 columns), the user indicates (as shown above) that he wants to redisplay the worksheet starting at row 1, column 7. Worksheet CARDIO displayed starting at row 1, column 7, appears as:
<table>
<thead>
<tr>
<th>ROWS/COLS</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>LABELS</td>
<td>prevmi</td>
<td>angina</td>
<td>hyperbp</td>
<td>surviv</td>
<td>ett</td>
<td></td>
</tr>
<tr>
<td>1 CASE 1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>6</td>
<td>...</td>
</tr>
<tr>
<td>2 CASE 2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>9</td>
<td>...</td>
</tr>
<tr>
<td>3 CASE 3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>...</td>
</tr>
<tr>
<td>4 CASE 4</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>9</td>
<td>...</td>
</tr>
<tr>
<td>5 CASE 5</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>9</td>
<td>...</td>
</tr>
<tr>
<td>6 CASE 6</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>9</td>
<td>...</td>
</tr>
<tr>
<td>7 CASE 7</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>9</td>
<td>...</td>
</tr>
<tr>
<td>8 CASE 8</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>9</td>
<td>...</td>
</tr>
<tr>
<td>9 CASE 9</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>9</td>
<td>...</td>
</tr>
<tr>
<td>10 CASE 10</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>9</td>
<td>...</td>
</tr>
<tr>
<td>11 CASE 11</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>9</td>
<td>...</td>
</tr>
<tr>
<td>12 CASE 12</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>9</td>
<td>...</td>
</tr>
<tr>
<td>13 CASE 13</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>9</td>
<td>...</td>
</tr>
<tr>
<td>14 CASE 14</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>9</td>
<td>...</td>
</tr>
<tr>
<td>15 CASE 15</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>9</td>
<td>...</td>
</tr>
<tr>
<td>16 CASE 16</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>9</td>
<td>...</td>
</tr>
<tr>
<td>17 CASE 17</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>9</td>
<td>...</td>
</tr>
<tr>
<td>18 CASE 18</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>9</td>
<td>...</td>
</tr>
<tr>
<td>19 CASE 19</td>
<td>1</td>
<td>7</td>
<td>1</td>
<td>1</td>
<td>9</td>
<td>...</td>
</tr>
<tr>
<td>20 CASE 20</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>9</td>
<td>...</td>
</tr>
</tbody>
</table>

To display the worksheet initially starting with other than the first row and column, the option 1 prompts should be answered as shown below:

```
WORKSHEET CARDIO
TITLE Cardiology Study
CREATED 1/21/76  MODIFIED 1/21/76
# OF ROWS 28
# OF COLS 12

TYPE yes TO DECODE DATA IN THIS WORKSHEET, ELSE HIT {RETURN};

TYPE yes TO SELECT STARTING ROW AND COLUMN, ELSE {RETURN}; yes

TYPE yes TO CONTINUE WITH worksheet ELSE
TYPE THE STARTING ROW #: 1
STARTING COL #: 7
```

The Decoding Facility

Worksheets contain only numbers. These may be values of numeric data or dates, times, or codes for encoded character-string data (see describe). For example, if "sex" is a coded item, its values may be entered (via enter or worksheet, options 4 through 6) as the character strings male and female, but these would be stored and displayed in the worksheet as 0 and 1 (or 1 and 0, depending on the schema). The decode facility allows the user to translate the codes into their corresponding character strings and to display those strings, in place, in worksheets.

To use the decode facility, either all of the values in the worksheet must be those of a single coded item which the user must specify, or worksheet row or column labels must correspond to coded items; in either case, data values must correspond to codes defined in the schema. In the worksheet shown above, the column labels sex, chpaim, v. fib, prevmi, angina, hyperbp, survir, and ett all correspond to coded items in the associated study schema. The decode facility requires that the schema be compiled but it need not be locked (see describe).
When initiating worksheet, option 1, the user may indicate that all or part of the decodable data in properly labeled columns or rows should be decoded when the worksheet is displayed. He does this by responding yes to the prompt:

**TYPE yes TO DECODE DATA IN THIS WORKSHEET, ELSE HIT [RETURN]:** yes

and then responding to:

**TYPE c TO DECODE COLUMNS, r FOR ROWS, OR TYPE AN ITEM NAME:**

Responding r implies that at least one of the row labels is the name of a character-coded item in the schema; responding c implies the same for column labels. Entering an item name would be appropriate if all the data in the worksheet were values of that single item (e.g., if the rows were patients and the columns were times).

As long as option 1 is not reinitiated, data will be decoded even if another portion of the worksheet is displayed.

When the current worksheet is CARDIO (shown above) and the user indicates that he wants columns decoded, with the CARDIO worksheet displayed starting at row 1, column 1, the resulting display is:

```
WORKSHEET CARDIO
TITLE Cardiology Study
CREATED 1/21/76  MODIFIED 1/21/76
# OF ROWS 28  # OF COLS 12
ROWS/COLS  1   2   3   4   5   6
LABELS    study##  entrydat  age  sex  chpsain  v.fib
1  CASE 1  7007   4/25/1978  51  male  no
2  CASE 2  7008   4/29/1978  ... male  no
3  CASE 3  7009   4/29/1978  17  male  no
4  CASE 4  7010   5/3/1978  69  male  no
5  CASE 5  7011   5/7/1978  49  male  ... no
6  CASE 6  7013   5/17/1978  64  male  no
7  CASE 7  7028   6/18/1978  67  female  no
8  CASE 2  7024   7/3/1978  58  male  no
9  CASE 9  7027   7/11/1978  67  male  no
10 CASE 10 7028   7/28/1978  43  male  no
11 CASE 11 7030   8/16/1978  55  male  no
12 CASE 12 7032   8/26/1978  75  female  no
13 CASE 13 7033   9/14/1978  70  male  no
14 CASE 14 7034   9/15/1978  53  female  no
15 CASE 15 7038   9/27/1978  65  female  no
16 CASE 16 7043  10/15/1978  55  male  no
17 CASE 17 7045  10/21/1978  58  male  no
18 CASE 18 7047  10/29/1978  57  male  no
19 CASE 19 7049  10/29/1978  56  male  no
20 CASE 20 7056  12/2/1978  72  female  no

(NOTE: ... indicates missing values)
```

**TYPE yes TO CONTINUE WITH worksheet ELSE:**

**TYPE THE STARTING ROW #: 1**

**STARTING COL #: 7**

When the user then asks to see CARDIO displayed starting at row 1, column 7, the following is displayed:
<table>
<thead>
<tr>
<th>CASE</th>
<th>prevmi</th>
<th>angina</th>
<th>hyperbp</th>
<th>surviv</th>
<th>ett</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>2</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>3</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>4</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>5</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>6</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>7</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>8</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>9</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>no</td>
</tr>
</tbody>
</table>

(Note: ... indicates missing values)

Type yes to continue with worksheet else
Type the starting row #:

This example illustrates several important features of the decode facility:

- Columns 4 through 11 have been decoded, while columns 1 through 3, which are not labeled with the names of coded items, were not affected. However, the dates in column 2 are displayed as dates regardless of whether or not the decode option is used.

- The missing value located at row 5, col 6, is not decoded and neither are other missing values.

- The blanks in columns 5 and 7 are not missing values, but rather are the character strings associated with codes in those columns (i.e., the code 2 corresponds to blank for chpain, and 3 corresponds to blank for prevmi) as specified in the schema.

- The values 3 and 9 in column 10, and 9 in column 11 do not correspond to valid item codes for surviv and ett, respectively, and so are left unchanged.

Care should be taken in using the decode facility on worksheets into which data have been entered manually, since arbitrary labels and data values can be entered without the screening provided by the enter activity. A column of survival codes, for example, with the meanings 0=survived and 1=died might mistakenly be labeled "sex", which has the codes 0=male and 1=female. If such a worksheet were then decoded, it might look reasonable, but it would be completely erroneous.

**OPTION 2: LABEL WORKSHEET ROWS**

This option provides for labeling worksheet rows with words up to 8 characters long. Existing labels may be changed. Labels may be blanked out by entering ... (i.e., three periods) for a label.
WORKSHEET 2

The user specifies the number of the first row that he would like to label. The system prompts with row numbers, the current row labels in parentheses, and eight underlined blanks (to remind the user that labels are limited to eight characters) until the last worksheet row or until the user terminates the dialog by typing \textit{activity-name}. When the user responds to a prompt by hitting the RETURN key, the corresponding row label remains unchanged from its previous value (including blank). The dialog proceeds as:

\begin{verbatim}
worksheet, 2 LABEL WORKSHEET ROWS

WORKSHEET ABLE
TITLE Frequency Data for Male/Female Glucose Study
CREATED 3/3/77 MODIFIED 3/3/77
# OF ROWS 3
# OF COLS 2

TO ENTER LABELS INTO THIS WORKSHEET TYPE THE
STARTING ROW #: 1
ROW 1 ( ) LABEL: Highgluc
ROW 2 ( ) LABEL:
ROW 3 ( ) LABEL: Lowgluc

TO ENTER MORE DATA, HIT [RETURN]:
\end{verbatim}

OPTION 3: LABEL WORKSHEET COLUMNS

This option allows the user to label worksheet columns in the same manner as rows are labeled in option 2.
An example is:

\begin{verbatim}
worksheet, 3 LABEL WORKSHEET COLUMNS

WORKSHEET ABLE
TITLE Frequency Data for Male/Female Glucose Study
CREATED 3/3/77 MODIFIED 3/3/77
# OF ROWS 3
# OF COLS 2

TO ENTER LABELS INTO THIS WORKSHEET TYPE THE
STARTING COLUMN #: 1
COLUMN 1 ( ) LABEL: Males
COLUMN 2 ( ) LABEL: Females

TO ENTER MORE DATA, HIT [RETURN]:
\end{verbatim}

OPTION 4: ENTER DATA BY ROWS

This option provides for entering data, a row at a time, into a specified range of worksheet rows and columns. There is no distinction between entering new values and replacing existing ones. Character strings may be entered for categorical items described in the schema. Dates and missing values may also be entered.

Dialog

When the user chooses \textit{worksheet}, option 4, he may first specify that he will be entering codes by responding \textit{r} or \textit{c} to the following prompt:

\textbf{TYPE} \textit{c} TO ENTER CODES FOR ITEMS IN COLUMNS, \textit{r} FOR ROWS, ELSE [RETURN]:
The user responds \textit{r} or \textit{c} to enter character strings corresponding to those specified in the schema (e.g., \textit{female} or \textit{male} as values of \textit{sex}); he responds with the RETURN key to enter numeric codes (e.g., \textit{0} or \textit{1} for \textit{sex}) for the same items. To enter character strings, labels must be names of character-coded items in the schema.

The user next specifies the range of rows and columns for which he wants to enter data. The system prompts for data \textit{column by column within rows}. The row and column labels and previously entered data, if any, are presented along with the row and column numbers. If the user does not wish to place a data value in a cell, he presses the RETURN key, and the previous value (including missing values) is retained. At any time he may enter /11 to display the current state of the worksheet, or /1 to restart this option.

\begin{verbatim}
worksheet,4 ENTER DATA BY ROWS

WORKSHEET ABLE
TITLE Frequency Data for Male/Female Glucose Study
CREATED 3/3/77 MODIFIED 3/3/77
# OF ROWS 3
# OF COLS 2

TO ENTER DATA INTO THIS WORKSHEET TYPE THE
STARTING ROW: 1
ENDING ROW: 3
STARTING COL: 1
ENDING COL: 2

ROW # 1 (Highgluc):
  COL # 1 (Males ......): 5
  COL # 2 (Females ......): 7

ROW # 2 (Lowgluc):
  COL # 1 (Males ......):
  COL # 2 (Females ......):

ROW # 3 (Lowgluc):
  COL # 1 (Males ......): 3
  COL # 2 (Females ......):

TO ENTER MORE DATA, HIT (RETURN): /11
\end{verbatim}

If the user wishes to replace an existing value with the null (i.e., missing) value, he does so by entering \ldots (three periods).

In entering dates into worksheets, the "slash" notation is used; i.e., \textit{mm/dd/yy} or \textit{mm/dd/yyyy}. If the year to be entered is not in this century, e.g., 1898, the entire year \textit{must} be entered, e.g., 9/21/1898 is entered to represent September 21, 1898. If the year is in the twentieth century, the century need not be entered; i.e., both 10/12/72 and 10/12/1972 are acceptable.

In response to the last prompt above, the user has typed /11. After selecting the starting row and column, he is presented with the display of the current worksheet:

\begin{verbatim}
WORKSHEET ABLE
TITLE Frequency Data for Male/Female Glucose Study
CREATED 3/3/77 MODIFIED 3/3/77

# OF ROWS 3
# OF COLS 2

ROWS/COLS 1 2
LABELS Males Females
1 Highgluc 5 7
2 Lowgluc ... ... 3

(NOTE: ... indicates missing values)
\end{verbatim}
Note that to complete entry of data into this worksheet, the user could choose option 4 or 5 and start at row 3, column 2, or he could choose option 6 and select only that cell.

OPTION 5: ENTER DATA BY COLUMNS

This option provides for entering columns of data into a worksheet. When the user chooses worksheet, option 5, indicates whether or not he is entering codes, and specifies the range of columns and rows for which he wants to enter data, the system prompts for data row by row within columns. The column and row labels and previously entered data, if any, are presented along with the column and row numbers. This option is similar to option 4 above.

OPTION 6: ENTER DATA BY CELLS

This option permits the user to enter data or codes cell by cell by specifying the cell row and column numbers and then the data value in response to the system prompts:

worksheet,6 ENTER DATA BY CELLS

WORKSHEET ABLE
TITLE Frequency Data for Male/Female Glucose Study
CREATED 3/3/77 MODIFIED 3/3/77
# OF ROWS 3
# OF COLS 2

TYPE c TO ENTER CODES FOR ITEMS IN COLUMNS, r FOR ROWS, ELSE [RETURN]:
TO ENTER DATA INTO THIS WORKSHEET TYPE THE
ROW #: 2
COL #: 2
ROW #: 2 { }
COL #: 2 (Females : ...): 5
ROW #: 3
COL #: 3

3 IS INCORRECT
COL #: 2
ROW #: 3 (Lowgluc )
COL #: 2 (Females : ...): 8
ROW #: 3
COL #: 1
ROW #: 3 (Lowgluc )
COL #: 1 (Males : 3): 4
ROW #: 1

Labels and previous values, if any, are presented. If the user wishes to replace an existing value with the null (i.e., missing) value, he does so by entering . . . (three periods).

The system responds, e.g., 3 IS INCORRECT, when a row or column number larger than the corresponding worksheet dimension is entered.

Responding /? to a prompt displays the current worksheet:
OPTION 7: DISCARD THIS WORKSHEET

This option provides for discarding a worksheet that is no longer useful. Selection of worksheet, option 7, prompts:

TYPE yes TO DISCARD WORKSHEET ABLE ELSE HIT [RETURN]:

where ABLE is the current worksheet. If yes is typed, the worksheet is discarded and there is no current worksheet. The user is then prompted:

WORKSHEET OPTION:

If the RETURN key is hit, the worksheet is not discarded and it remains the current worksheet.

OPTION 8: SORT THIS WORKSHEET

This subactivity allows the user to reorder worksheet rows by sorting the values in a column or to reorder columns by sorting the values in a row. Thus, if values in a column are sorted (in ascending or descending order), the contents of each row remain unchanged, but entire rows are rearranged to reflect the sort order of values in the specified column. This is a particularly useful function because it can be used to group patients and their data according to values of one or more variables. Analyze and calculate subactivities can then be applied to the resultant individual groups of data.

Values in several columns or rows may be sorted by iterating through the subactivity. The major sort is done last. For example, to group patients by diastolic blood pressure and within blood pressure by sex and within sex by age, the user first sorts the age column (or row), then sorts by sex, and finally sorts by blood pressure.

Values are sorted according to their numerical representations. Numbers are sorted in consecutive order, as one would expect. Coded values are sorted according to their numeric codes, not alphabetically according to their character-string representations.

Because dates are represented as numbers, with month and day as the most significant digits and year as the least, dates within a year are sorted correctly but
years are not. To completely sort dates stored in a row or column, use calculate, option 1, create a new variable year = date - 100 • INT(date/100), sort on date, then sort on year.

Dialog

This subactivity requires the user to indicate whether the data are to be sorted by row or by column. He may then examine the appropriate labels. After he has indicated either ascending or descending order and has selected a particular row or column, the entire worksheet is rearranged according to the data in that row or column. The user may then continue sorting or may perform another activity such as displaying the sorted worksheet.

In the following example, the CARDIO worksheet is sorted. CARDIO was illustrated in both its numeric and decoded forms above, in the discussion of worksheet, option 1, Display this Worksheet. Each worksheet row corresponds to a different patient. The user wishes to sort the worksheet such that patients are grouped by age, within age by sex, and within sex by entry date; thus he will sort by columns. The entry date sort is not necessary because the rows are already in entry date order.

After the user chooses worksheet, option 8 (e.g., by typing /work,8), and the initial display appears, he types c in response to:

TYPE r TO SORT A ROW; FOR COLUMN SORT TYPE c:

and then types yes in response to:

TYPE yes TO DISPLAY COLUMN LABELS; ELSE HIT [RETURN]:

The display now appears as shown below.

worksheet, 8 WORKSHEET SORT BY COLUMN OR ROW
WORKSHEET CARDIO
TITLE Cardiology Study
CREATED 1/21/76 MODIFIED 1/21/76
# OF ROWS 28
# OF COLS 12

TYPE yes FOR ASCENDING ORDER, [RETURN] FOR DESCENDING ORDER:

1 study
2 entrydat
3 age
4 sex
5 chpain
6 v.fib
7 prevai
8 angina
9 hyperbp
10 surviv
11 ett

The user first wants to sort on sex, such that the first group of worksheet rows is for female patients and the second group is for males. Because the sex code for female is 1 and that for male is 0, the user hits the RETURN key to indicate that
he wants values to be sorted in descending order (i.e., with 1s followed by 0s). Next, the user specifies that he wants to sort on sex by typing 4 in response to

**TYPE THE COLUMN NUMBER:**

The system indicates that it is sorting on column 4. When it is done, it displays

**TYPE yes TO SPECIFY ANOTHER SORT, ELSE HIT [RETURN]:**

In this case, the user types 11 to display the worksheet. The first 20 rows of the sorted worksheet CARDIO are shown below.

<table>
<thead>
<tr>
<th>ROWS/COLS</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>LABELS</td>
<td>study#</td>
<td>entrydat</td>
<td>age</td>
<td>sex</td>
<td>chp</td>
<td>v.fib</td>
</tr>
<tr>
<td>1 CASE 7</td>
<td>7020</td>
<td>6/10/1970</td>
<td>67</td>
<td>female</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>2 CASE 12</td>
<td>7032</td>
<td>8/28/1970</td>
<td>75</td>
<td>female</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>3 CASE 14</td>
<td>7034</td>
<td>9/15/1970</td>
<td>53</td>
<td>female</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>4 CASE 15</td>
<td>7030</td>
<td>9/27/1970</td>
<td>65</td>
<td>female</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>5 CASE 20</td>
<td>7056</td>
<td>12/2/1970</td>
<td>72</td>
<td>female</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>6 CASE 28</td>
<td>7246</td>
<td>9/17/1972</td>
<td>67</td>
<td>female</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>7 CASE 1</td>
<td>7007</td>
<td>4/25/1970</td>
<td>51</td>
<td>male</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>8 CASE 2</td>
<td>7000</td>
<td>4/29/1970</td>
<td>...</td>
<td>male</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>9 CASE 3</td>
<td>7009</td>
<td>4/29/1970</td>
<td>17</td>
<td>male</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>10 CASE 4</td>
<td>7010</td>
<td>5/3/1970</td>
<td>69</td>
<td>male</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>11 CASE 5</td>
<td>7011</td>
<td>5/7/1970</td>
<td>49</td>
<td>male</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>12 CASE 6</td>
<td>7013</td>
<td>5/17/1970</td>
<td>64</td>
<td>male</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>13 CASE 8</td>
<td>7024</td>
<td>7/3/1970</td>
<td>58</td>
<td>male</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>14 CASE 9</td>
<td>7027</td>
<td>7/11/1970</td>
<td>67</td>
<td>male</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>15 CASE 10</td>
<td>7029</td>
<td>7/28/1970</td>
<td>43</td>
<td>male</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>16 CASE 11</td>
<td>7030</td>
<td>8/16/1970</td>
<td>55</td>
<td>male</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>17 CASE 13</td>
<td>7033</td>
<td>9/14/1970</td>
<td>70</td>
<td>male</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>18 CASE 16</td>
<td>7043</td>
<td>10/15/1970</td>
<td>55</td>
<td>male</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>19 CASE 17</td>
<td>7045</td>
<td>10/21/1970</td>
<td>58</td>
<td>male</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>20 CASE 18</td>
<td>7047</td>
<td>10/29/1970</td>
<td>57</td>
<td>male</td>
<td>no</td>
<td>no</td>
</tr>
</tbody>
</table>

Noting that he accomplished what he intended thus far, the user types 8 in response to:

**TYPE THE STARTING ROW #:**

to continue sorting. He next wants to sort on age, in order to move the data for the youngest patient to row 1 and those for the oldest to the last row; to do this, he requests the system to sort on column 3 in ascending order. The first 20 rows and then rows 21 through 28 of the resultant sorted worksheet are shown below. The last row has a missing value for age, because sorting always moves rows (or columns) containing missing values to the bottom (or right side) of the worksheet.

Note that this subactivity reorganizes the current worksheet rather than moving the results into a new one. To keep a copy of the unsorted worksheet, the user must first copy it by using retrieve, option 4, or by using retrieve, options 5 and 6.
### Worksheet 8

<table>
<thead>
<tr>
<th>ROWS/COLS</th>
<th>LABELS</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 CASE 3</td>
<td>study#</td>
<td>7009</td>
<td>4/29/1970</td>
<td>17</td>
<td>male</td>
<td>no</td>
<td></td>
</tr>
<tr>
<td>2 CASE 10</td>
<td>entrydat</td>
<td>7028</td>
<td>7/26/1970</td>
<td>43</td>
<td>male</td>
<td>no</td>
<td></td>
</tr>
<tr>
<td>3 CASE 5</td>
<td>age</td>
<td>7011</td>
<td>5/7/1970</td>
<td>49</td>
<td>male</td>
<td>no</td>
<td></td>
</tr>
<tr>
<td>4 CASE 1</td>
<td>sex</td>
<td>7007</td>
<td>4/25/1970</td>
<td>51</td>
<td>male</td>
<td>no</td>
<td></td>
</tr>
<tr>
<td>5 CASE 14</td>
<td>chpaine</td>
<td>7034</td>
<td>9/15/1970</td>
<td>53</td>
<td>female</td>
<td>no</td>
<td></td>
</tr>
<tr>
<td>6 CASE 11</td>
<td>v.fib</td>
<td>7030</td>
<td>8/16/1970</td>
<td>55</td>
<td>male</td>
<td>no</td>
<td></td>
</tr>
<tr>
<td>7 CASE 16</td>
<td></td>
<td>7043</td>
<td>10/15/1970</td>
<td>55</td>
<td>male</td>
<td>no</td>
<td></td>
</tr>
<tr>
<td>8 CASE 22</td>
<td></td>
<td>7060</td>
<td>12/26/1970</td>
<td>55</td>
<td>male</td>
<td>no</td>
<td></td>
</tr>
<tr>
<td>9 CASE 19</td>
<td></td>
<td>7049</td>
<td>10/29/1970</td>
<td>56</td>
<td>male</td>
<td>no</td>
<td></td>
</tr>
<tr>
<td>10 CASE 10</td>
<td></td>
<td>7047</td>
<td>10/29/1970</td>
<td>57</td>
<td>male</td>
<td>no</td>
<td></td>
</tr>
<tr>
<td>11 CASE 8</td>
<td></td>
<td>7024</td>
<td>7/3/1970</td>
<td>58</td>
<td>male</td>
<td>no</td>
<td></td>
</tr>
<tr>
<td>12 CASE 17</td>
<td></td>
<td>7045</td>
<td>10/21/1970</td>
<td>58</td>
<td>male</td>
<td>no</td>
<td></td>
</tr>
<tr>
<td>13 CASE 25</td>
<td></td>
<td>7067</td>
<td>1/20/1971</td>
<td>58</td>
<td>male</td>
<td>no</td>
<td></td>
</tr>
<tr>
<td>14 CASE 21</td>
<td></td>
<td>7057</td>
<td>12/8/1970</td>
<td>60</td>
<td>male</td>
<td>no</td>
<td></td>
</tr>
<tr>
<td>15 CASE 6</td>
<td></td>
<td>7013</td>
<td>5/17/1970</td>
<td>64</td>
<td>male</td>
<td>no</td>
<td></td>
</tr>
<tr>
<td>16 CASE 15</td>
<td></td>
<td>7038</td>
<td>9/27/1970</td>
<td>65</td>
<td>female</td>
<td>no</td>
<td></td>
</tr>
<tr>
<td>17 CASE 23</td>
<td></td>
<td>7062</td>
<td>12/31/1970</td>
<td>65</td>
<td>male</td>
<td>no</td>
<td></td>
</tr>
<tr>
<td>18 CASE 7</td>
<td></td>
<td>7028</td>
<td>6/18/1970</td>
<td>67</td>
<td>female</td>
<td>no</td>
<td></td>
</tr>
<tr>
<td>19 CASE 28</td>
<td></td>
<td>7246</td>
<td>9/17/1972</td>
<td>67</td>
<td>female</td>
<td>no</td>
<td></td>
</tr>
<tr>
<td>20 CASE 9</td>
<td></td>
<td>7027</td>
<td>7/11/1970</td>
<td>67</td>
<td>male</td>
<td>no</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ROWS/COLS</th>
<th>LABELS</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>21 CASE 4</td>
<td>study#</td>
<td>7018</td>
<td>5/3/1970</td>
<td>69</td>
<td>male</td>
<td>no</td>
<td></td>
</tr>
<tr>
<td>22 CASE 13</td>
<td>entrydat</td>
<td>7033</td>
<td>9/14/1970</td>
<td>70</td>
<td>male</td>
<td>no</td>
<td></td>
</tr>
<tr>
<td>23 CASE 20</td>
<td>age</td>
<td>7056</td>
<td>12/2/1970</td>
<td>72</td>
<td>female</td>
<td>no</td>
<td></td>
</tr>
<tr>
<td>24 CASE 12</td>
<td>sex</td>
<td>7032</td>
<td>8/28/1970</td>
<td>75</td>
<td>female</td>
<td>no</td>
<td></td>
</tr>
<tr>
<td>25 CASE 26</td>
<td>chpaine</td>
<td>7241</td>
<td>9/10/1972</td>
<td>76</td>
<td>male</td>
<td>no</td>
<td></td>
</tr>
<tr>
<td>26 CASE 24</td>
<td>v.fib</td>
<td>7057</td>
<td>1/10/1971</td>
<td>79</td>
<td>male</td>
<td>no</td>
<td></td>
</tr>
<tr>
<td>27 CASE 27</td>
<td></td>
<td>7242</td>
<td>9/11/1972</td>
<td>80</td>
<td>male</td>
<td>no</td>
<td></td>
</tr>
<tr>
<td>28 CASE 2</td>
<td></td>
<td>7008</td>
<td>4/29/1970</td>
<td>...</td>
<td>male</td>
<td>no</td>
<td></td>
</tr>
</tbody>
</table>

If additional data are to be retrieved from the SDF and associated with data in a worksheet that has been sorted, a subset may be created (using subset, option 6) from the sorted worksheet which will have the patients in the desired order.

### Option 9: List Your Worksheets

This subactivity displays lists of worksheets and their descriptions on the CRT. More compact listings of worksheet names (without titles) may be obtained using files, option 1. The dialog for option 9 is:
### WORKSHEET LIST

<table>
<thead>
<tr>
<th>Worksheet Name</th>
<th>Created</th>
<th>Modified</th>
<th>Rows</th>
<th>Columns</th>
</tr>
</thead>
<tbody>
<tr>
<td>WORKSHEET GLUCOSEC</td>
<td>9/10/75</td>
<td>9/10/75</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>TITLE Control Glucose by Case and by Hour</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WORKSHEET CASEC</td>
<td>9/10/75</td>
<td>9/10/75</td>
<td>9</td>
<td>4</td>
</tr>
<tr>
<td>TITLE Flowsheet for Control Case 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WORKSHEET MEANURL</td>
<td>9/11/75</td>
<td>9/11/75</td>
<td>8</td>
<td>10</td>
</tr>
<tr>
<td>TITLE Means starting at Hour 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WORKSHEET CLINICAL</td>
<td>9/10/75</td>
<td>9/10/75</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>TITLE Clinical Characteristics of Diabetic Subjects</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WORKSHEET MEANS</td>
<td>9/18/75</td>
<td>2/25/76</td>
<td>8</td>
<td>13</td>
</tr>
<tr>
<td>TITLE Means of Control and Treatment Data by Hour and Variable</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WORKSHEET SCHAFFER</td>
<td>5/15/75</td>
<td>8/4/75</td>
<td>28</td>
<td>28</td>
</tr>
<tr>
<td>TITLE Schaffer Cardiology Study</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WORKSHEET SMB675</td>
<td>8/21/75</td>
<td>3/2/76</td>
<td>258</td>
<td>7</td>
</tr>
<tr>
<td>TITLE SMB675 Data for 6/75 (every 16th patient)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WORKSHEET TEST30</td>
<td>9/17/75</td>
<td>10/2/75</td>
<td>35</td>
<td>4</td>
</tr>
<tr>
<td>TITLE test 30 codes and checks</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Pressing the RETURN key continues the procedure until all worksheets have been listed.

If there are many worksheets, listing and printing all of them with this suboption can take a long time.

*Worksheets that are no longer needed should be discarded.*

Because this option accesses worksheets to obtain their titles, it changes the *date last accessed* displayed by files, option 1.

### OPTION 10: PRINT THIS WORKSHEET

This subactivity allows the user to have the entire current worksheet printed on the system line printer. For large worksheets, this is much more efficient than the use of option 1 and the attached printer. The printing of pages that would contain all missing values may be suppressed, since a worksheet's dimensions are frequently considerably larger than those required for the data.

After selecting option 10, the user can choose to decode columns and rows. He can then specify if he wishes to print all pages or only those pages containing data. The cover sheet can then be labeled. The user should enter his name or initials so that his copy can be identified if someone else removes it from the system line printer.
WORKSHEET, 10

WORKSHEET HOSPSTAY

TITLE Artery Dis, Curr Op, Post Op Days, & Hospital

# OF ROWS 391
# OF COLS 4
CREATED 3/3/77       MODIFIED 3/3/77

TYPE yes TO DECODE DATA IN THIS WORKSHEET, ELSE HIT [RETURN]: yes

TYPE c TO DECODE COLUMNS, r FOR ROWS, OR TYPE AN ITEM NAME: c

TYPE yes TO PRINT PAGES OF ALL MISSING VALUES, ELSE HIT [RETURN]:

ENTER LABEL FOR PRINTOUT COVER SHEET: gabe

When the printing is complete, the user can continue with worksheet activities or go on to some other activity:

PRINTING COMPLETE
HIT [RETURN] TO CONTINUE WITH WORKSHEET ACTIVITY:

Each page of the printed worksheet contains up to 6 columns and up to 50 rows. If there are more than 6 columns, all the rows of the first 6 columns are printed first, followed by all the rows of columns 7 through 12, etc. This makes it easier to lay out the pages to create a large, complete worksheet if one is desired.

OPTION 11: EDIT THIS WORKSHEET

Option 11 permits the user to modify existing worksheets by

1. Adding blank rows or columns.
2. Deleting rows or columns.
3. Moving a range of rows or columns.
4. Copying a range of rows or columns.
5. Changing the worksheet name.

The first four types of modification are accomplished by editing a numbered list of worksheet row and column labels; when leaving the subactivity, CLINFO modifies the worksheet itself. The editing is performed by entering the appropriate command and specifying the row or column as a decimal line number in the context of the current list of row and column labels. The changes are made in response to the commands, but the new name, title, or arrangement of rows and columns is not displayed until the list command is issued. Successive commands may be entered without leaving the subactivity. Some commands may result in fractional line numbers in the list, even though the actual row or column numbers are always consecutive integers. However, with the renumber command, the list may be renumbered at any time to assign consecutive integers to the line numbers so that they correspond to actual row and column numbers. Thus, a displayed, numbered list of row or column labels has one line for each row or column of the edited worksheet, and editing commands refer to those line numbers when specifying operations on particular rows or columns, but a line does not always have the same number as the row or column it represents.
Dialog

After the user selects the subactivity, the first prompt is:

TYPE r TO EDIT ROWS, c for COLUMNS, OR HIT [RETURN] FOR NAME OR TITLE CHANGE:

If any changes other than name or title changes are to be made to the worksheet, either r or c must be entered, indicating that all subsequent commands are to apply to rows or to columns, respectively. The message

EDIT WORKSHEET—ROWS

or

EDIT WORKSHEET—COLUMNS

is immediately displayed at the top of the screen.

The following prompt is then displayed:

COMMANDS: add #, delete #, move from #, copy from #, enlarge by #,
list #, renumber, title change, name change

COMMAND:

The command may be abbreviated to its first letter, e.g., a for add, d for delete, t for title change, etc.

The # signs refer to user-entered line numbers which may be integers or decimals greater than zero with two digits to the right of the decimal point, e.g., 2, 3.50, 121.11, etc. These line numbers appear in the list of rows or columns, and while each row or column has a line number, the number is not necessarily a row or column number.

The worksheet below is used in the following examples:

WORKSHEET EDIT DEMO
TITLE WS for demonstrating the WORKSHEET EDIT sub-option
CREATED 10/12/75 MODIFIED 10/14/75
6 ROWS x 4 COLS

ROWS/COLS | 1 | 2 | 3 | 4
---|---|---|---|---
LABELS | One | Two | Three | Four
---|---|---|---|---
1 Alpha | 1 | 2 | 3 | 4
2 Beta | 1 | 2 | 3 | 4
3 Gamma | 1 | 2 | 3 | 4
4 Delta | 1 | 2 | 3 | 4
5 Epsilon | 1 | 2 | 3 | 4
6 Zeta | 1 | 2 | 3 | 4

Examples

1. list # (or l #). The list # command should be used first to serve as a reference for the operations to follow. It should also be used periodically to determine the effects of other entered commands. If list n (or l n) is entered as a command, where n is a number in the current range of lines, up to 25 lines starting
with line n will be listed on the screen. The current range is the first to the last numbered line in the list. The list has the heading EDITING ROWS or EDITING COLUMNS. For example, if columns had been selected, list 1 would display:

**EDITING COLUMNS**

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1.00</td>
<td>One</td>
</tr>
<tr>
<td>2.00</td>
<td>Two</td>
</tr>
<tr>
<td>3.00</td>
<td>Three</td>
</tr>
<tr>
<td>4.00</td>
<td>Four</td>
</tr>
</tbody>
</table>

where 1.00, 2.00, etc., are line numbers for reference in the editing procedure, and One, Two, etc., are column labels.

If no starting line number is specified, the first 25 lines are listed.

2. **add # (or a #)**. This command is used to add a single blank row or column to the worksheet. A new row (or column) may be inserted between existing rows (or columns), may precede the first, or may follow the last. In the present example, the user could enter add 1.5, and the next list command would display

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1.00</td>
<td>One</td>
</tr>
<tr>
<td>1.50</td>
<td></td>
</tr>
<tr>
<td>2.00</td>
<td>Two</td>
</tr>
<tr>
<td>3.00</td>
<td>Three</td>
</tr>
<tr>
<td>4.00</td>
<td>Four</td>
</tr>
</tbody>
</table>

If the worksheet were then displayed, it would appear, in part, as shown below:

**6 ROWS x 5 COLUMNS**

<table>
<thead>
<tr>
<th>ROWS/COLS</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>LABELED</td>
<td>One</td>
<td>Two</td>
<td>Three</td>
<td>Four</td>
<td></td>
</tr>
<tr>
<td>1 Alpha</td>
<td>1</td>
<td>...</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>2 Beta</td>
<td>1</td>
<td>...</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>6 Zeta</td>
<td>1</td>
<td>...</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>

Note that

- A blank column (i.e., the new column 2) has been inserted between old columns 1 and 2 (still labeled One and Two).
- The columns have been renumbered and the dimensions have been changed.
- The labels remain with their original columns of data.

3. **Delete # (or d #)**. Again using our original 4-column example; if the command delete 2 is issued, the following worksheet results:
Note again that
- The columns have been renumbered.
- The dimensions have been adjusted.
- The labels have been retained.

4. **move from # (or m #)**. This command permits you to move one or more consecutive rows or columns from their original position to a new relative position in the worksheet.

When the command *move from m* is entered, where *m* is a number in the current range of lines, the system prompts, *thru #*. If *m* is entered again, only the row or column represented by line *m* in the list will be moved. If a number *n*, greater than *m*, is entered, then a range of rows or columns will be moved, i.e., those rows or columns with line numbers *m* through *n*. Finally, the system prompts to *#*. In response, the user must enter a destination number—the number of the line to which the first row or column in the range is to be moved. An existing line number may not be used as the destination line. The use of decimals permits the insertion of lines between existing lines; up to 99 lines may be inserted between two lines numbered with consecutive integers. For example, the commands

```
move 1 thru 2 to 3.5
```

applied to columns have the following effect:

<table>
<thead>
<tr>
<th>(before)</th>
<th>(after)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.00 One</td>
<td>3.00 Three</td>
</tr>
<tr>
<td>2.00 Two</td>
<td>3.50 One</td>
</tr>
<tr>
<td>3.00 Three</td>
<td>3.60 Two</td>
</tr>
<tr>
<td>4.00 Four</td>
<td>4.00 Four</td>
</tr>
</tbody>
</table>

Note that the system renumbers the "moved" lines as necessary to fit into the specified destination. The resultant worksheet would be:

<table>
<thead>
<tr>
<th>ROW/COL</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>LABELS</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 Alpha</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>2 Beta</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

The columns are renumbered as consecutive integers starting with 1.
5. **copy from # (or c #).** This command is similar to *move from #* except that the source rows or columns are left in place and are also duplicated in the list, starting at the position indicated by the response to the *to #* prompt.

For example, the command

```plaintext
copy 1 thru 2 to 3.5
```

has the following effect:

<table>
<thead>
<tr>
<th>(before)</th>
<th>(after)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.00 One</td>
<td>1.00 One</td>
</tr>
<tr>
<td>2.00 Two</td>
<td>2.00 Two</td>
</tr>
<tr>
<td>3.00 Three</td>
<td>3.00 Three</td>
</tr>
<tr>
<td>4.00 Four</td>
<td>3.60 Two</td>
</tr>
<tr>
<td></td>
<td>4.00 Four</td>
</tr>
</tbody>
</table>

6. **enlarge by # (or e #).** This command adds the specified number of blank rows or columns following the last current row or column of the worksheet. For example, the command

```plaintext
enlarge by 2
```

applied to columns has the following effect:

<table>
<thead>
<tr>
<th>(before)</th>
<th>(after)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.00 One</td>
<td>1.00 One</td>
</tr>
<tr>
<td>2.00 Two</td>
<td>2.00 Two</td>
</tr>
<tr>
<td>3.00 Three</td>
<td>3.00 Three</td>
</tr>
<tr>
<td>4.00 Four</td>
<td>4.00 Four</td>
</tr>
<tr>
<td></td>
<td>5.00</td>
</tr>
<tr>
<td></td>
<td>6.00</td>
</tr>
</tbody>
</table>

In the worksheet, columns 5 and 6 will be unlabeled and their contents will be missing values.

7. **renumber (or r).** This command causes the line numbers to be renumbered as consecutive integers. The command has no effect on the actual row or column numbers, but it makes the line numbers correspond to what the row or column numbers will be after the editing commands are processed. For example, if a series of commands have resulted in the list to the left below, a *renumber* command (followed by a *list* command) will display the list on the right:

<table>
<thead>
<tr>
<th>(before)</th>
<th>(after)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.00 One</td>
<td>1.00 One</td>
</tr>
<tr>
<td>1.50</td>
<td>2.00</td>
</tr>
<tr>
<td>2.00</td>
<td>3.00</td>
</tr>
<tr>
<td>2.10</td>
<td>4.00</td>
</tr>
<tr>
<td>2.20</td>
<td>5.00</td>
</tr>
<tr>
<td>5.00</td>
<td>6.00</td>
</tr>
<tr>
<td>7.00</td>
<td>7.00</td>
</tr>
</tbody>
</table>

The renumbered list will not appear until a *list* command is issued.
8. **title change** (or t). The command *title* permits the title of an existing worksheet to be changed. It elicits the prompt *TITLE:*, to which the user may respond with up to 70 characters of text. The title he types will replace the original worksheet title. However, the current title displayed at the top of the screen will not change until the *list* command is issued or the current worksheet heading is otherwise displayed.

9. **name change** (or n). Entering *name* causes the system to prompt *NAME:*.

The user may respond with a word of up to eight characters, which will replace the current worksheet name. If an existing worksheet name, such as *oldname*, is entered, the system responds

```
oldname ALREADY EXISTS. PLEASE CHOOSE A NEW NAME.
```

and reprompts for a name. As with title changes, the new name will appear the next time the current worksheet heading is displayed.
GLOSSARY

AA Terminal. An Ann Arbor video terminal; the standard CLINFO user's terminal.

Activity. One of the major functional groups of the CLINFO system; e.g., analyze, enter, describe. Each activity is called for by typing its name (in lower case) preceded by '!

All. The set of all patients entered into a study.

Alphanumeric. Composed of both alphabetic and numeric characters, as an alphanumeric character string.

Arithmetic expression. An algebraic expression involving real numbers, worksheet row or column numbers, functions, and arithmetic operators (+, -, *, /, and \\

Arithmetic operators. Operators used in arithmetic expressions (i.e., +, -, *, /, and \\

Assembly language. A programming language, unique to a particular computer, which corresponds to its built-in instruction repertoire.

Automatic selection. A feature in retrieve, option 2, which selects only those time subunits in which an item of interest was measured.

Backup. Noun: A copy of a file or program on tape or disk which may be used if the original is lost or destroyed. Verb: The act of making such a copy.

BASIC. A high-level programming language which runs on the CLINFO computer and can be used to perform analyses not provided by the CLINFO prototype.

Begin. A system-defined event which takes on the value of the time and date of first data collection for a patient.

Bit. The smallest quantity of information that may be stored in the computer, corresponding to one or zero.

BRK. The BREAK key on the AA terminal; it causes the attached DTC printer to stop prior to completion of its run.

BS. The backspace key on the AA terminal, used to delete typed entries prior to hitting the RETURN key.

Byte. A unit of memory large enough to store a single character.

Character. Any numeral, upper- or lower-case letter, punctuation sign, or algebraic or other symbol.

Character string. A series of characters.

Cleaned. The state of a data file after the removal of errors.

Clem (CLINFO Editing Machine). A text editor which runs on the CLINFO computer and which can be used for manipulating text and for various bookkeeping functions.

CLINFO. An NIH-sponsored study undertaken to design a system for the management and analysis of clinical research data, to be used by individual investigators. CLINFO is also the name of a yellow terminal key that is used in conjunction with other keys to move directly to an activity or subactivity.

Clinical contractor. A research physician holding a contract with the Division of Research Resources of the NIH relating to the development of the CLINFO
system; the physician in charge of the prototype system installed at his institution.

**Code.** A number which represents a meaningful, previously specified character string, and which replaces the character string for efficient computer storage and manipulation.

**Column.** A vertically arranged list in a worksheet, identified by a number, which may be labeled and which may contain missing value indicators, numbers, dates, times, or codes representing character strings.

**Communication File.** The CLINFO file onto which SDF or worksheet data may be written in order to operate upon them with BASIC programs or other programs outside of the CLINFO system; also the file onto which data may be written in order to transfer them into the system and into worksheets.

**Context.** The patient, date, and time of data collection that uniquely identify an instance of a panel within a study.

**Created date.** The date on which a worksheet was built; the created date is displayed with the worksheet.

**Cursor.** An underline that appears on the screen of the AA terminal, indicating the position of the next character to be entered.

**Date.** Dates are specially treated by CLINFO. They are entered in a format such as 11/23/77 or 11/23/1977 for November 23, 1977, or as 7/4/1776 for July 4, 1776; they are displayed as 11/23/77 or 7/4/1776; and they are stored internally as 11231977 or 7041776. The twentieth century is assumed unless otherwise specified.

**Decode.** The facility that translates a numeric code into a user-specified character string.

**DTC printer.** The hardcopy device attached to the AA terminal which creates paper copies of the information displayed on the screen of the terminal.

**Encoded** (or Coded). Having a numeric code substituted for a character string.

**End.** The system-defined event that takes on the date and time of the last entry of data for a patient.

**ESC.** The ESCAPE key on the AA terminal; pressing this key and the CLINFO key simultaneously aborts and restarts the current activity or subactivity.

**Event.** A time marker which takes on a user-defined value of time associated with a particular instance (i.e., first, last, max, min, or specific value) of an item.

**File.** A list or collection of data stored by the CLINFO prototype.

**Function.** A system-defined calculation that operates on the value of an arithmetic expression and which may be included in an expression when performing row/column calculations. A function has a three-character identifier and is followed by an expression in parentheses. The functions are abs, atn, cos, exp, fln (logarithm to the base 10), int, log (logarithm to the base e), rnd, sgn, sin, sqr (i.e., square root), and tan.

**Hardcopy device.** A printer or facsimile device that produces paper copies of information generated by the computer.

**Hardcopy terminal.** A hardcopy device with a keyboard permitting two-way user/system communication.

**Hardware.** The mechanical, electrical, and electronic components of a computer system.

**High-level** (language). A programming language employing English-like or algebra-like statements rather than numeric codes.
Histogram. A graphic presentation of the frequency of cases falling into each of a number of categories.

Instance. A particular value or set of values corresponding to a particular panel and measured at a particular time.

Interactive. A form of computer use in which the user and the system exchange information during the execution of a program.

Item. A variable or measurement for which numeric or textual values will be collected during the course of a study.

K. A symbol representing the number 1024 (2 to the 10th power), used when referring to the size of a computer memory or disk (e.g., 32K bytes of memory can store 32,768 characters).

Keyboard. An electro-mechanical device with buttons or keys representing letters, numerals, special characters, and control functions, attached to a display device such as a CRT for the purpose of entering data.

Label. The eight-character name assigned by the user or the system to a row or column of a worksheet.

Logical expression. An expression that has values of "true" (represented by the number 1) and "false" (represented by the number 0). In the CLINFO system, the user writes logical expressions involving arithmetic expressions (i.e., algebraic expressions containing real numbers, worksheet row or column numbers, functions, and arithmetic operators) and relational operators (=, <, <=, >, >=, <, >). The system uses these expressions to produce combined logical AND or OR expressions which are true if all (in the AND case) or any (in the OR case) of the constituent expressions have the “true” value.

Log-off. The process of terminating a CLINFO session.

Log-on. The process of activating the terminal and typing in your identification and that of the study you wish to work on.

Menu. A list of options, usually numbered, one of which may be chosen in responding to a prompt.

Merge. The function performed by the system manager in which he moves the data entered into the Update File (using enter) into the SDF.

Minicomputer. A small computer costing from about $10,000 to $150,000, usually slower and with smaller memory capacity than a full-size computer.

Missing Value. The value of an item or worksheet cell before an actual value has been entered or otherwise defined. A missing value is displayed as . . . . (3 dots) and is stored as -9.99999E-07. The analyze and calculate subactivities are designed to properly take missing values into consideration.

Modified date. The latest date on which worksheet options 2 through 6, 8, or 11 were selected for a current worksheet; initially the same as the created date.

Non-parametric test. A statistical test, such as those performed by analyze, option 11, that does not require the assumption that the population is normally distributed.

Numeric. A value represented by a real number, e.g., 125, 97.6, -20.

Numeric item. A schema item defined as num, date, time, or char.

Numeric panel. The name assigned in the schema to a group of numeric items.

Off-line. A term referring to procedures, such as file backup, that do not require running the CLINFO system.

Option (number). A choice within a CLINFO activity; e.g., Enter Data by Rows is option 4 in the worksheet activity.
Panel. A group of items, the values for which are collected at (or near) the same time for the same subject.

Password. A confidential code comprising a unique user identifier and a study identifier which is entered into the terminal when the user logs on.

Patient. A case or subject identified by an abbreviation and known to the CLINFO system as a member of a particular study. One of the three "dimensions" of the SDF.

PRINT. The PRINT key on the keyboard of the AA terminal which causes the attached printer to copy the contents of the screen onto paper.

Printer. A hardcopy device that produces paper copies of information generated by the computer; CLINFO uses a system printer and a DTC printer.

Prototype. An early version of a system constructed to test proposed capabilities and features.

Relational operator. One of the operators $=, <, <=, >, >=, <=$ (i.e., equal to, less than, less than or equal to, greater than, greater than or equal to, not equal to) used in logical expressions.

Response File. A sequence of CLINFO responses created, edited, and saved using the response file activity, which will be executed by the prototype as if they had been entered from the keyboard; used for carrying out complex and repetitive sequences of CLINFO operations.

RETURN. A terminal key used to indicate that the "normal" action should be taken or to signify that the user's response is complete.

Row. A horizontally arranged list in a worksheet, identified by a number, which may be labeled and which may contain missing value indicators, numbers, dates, times, or codes representing character strings.

Schema. The user-specified CLINFO data dictionary, which includes the names of panels, items, and events and which specifies the types of checking to be performed upon data when they are entered.

SDF (Study Data File). The file containing all information entered via the enter activity concerning a particular study, and from which data may be retrieved.

Simultaneous (system use). Use of the system by more than one user by means of timesharing.

Software. Computer programs, including the operating system, the CLINFO program, and user-written programs.

Special character. Characters other than numerals and upper- and lower-case letters; e.g., #, @ or +.

Study. A single experiment, based upon an approved protocol and carried out by one or more investigators. Each CLINFO session is in the context of a single particular study.

Subactivity. A specific function within an activity; e.g., T-Test, in the analyze activity. A subactivity is called into use by typing its number (chosen from the activity menu) following an activity name (e.g., analyze,2), or (if already using the same activity) by typing its number (e.g., 2).

Subsetting. Creating sets of patient identifiers according to specified characteristics which they have in common. The subset activity creates sets based on SDF data or data contained in worksheets.

System manager. An employee of the institution where the CLINFO prototype is located who is responsible for the day-to-day operation of the system and for instructing and assisting users.
System printer. The high-speed hardcopy device attached directly to the CLINFO prototype computer, which may be used to list BASIC programs, schemas, and Cem files.

System representative. An employee of The Rand Corporation who visits the CLINFO sites periodically to learn about users' needs and to inform users about system changes.

Terminal. A device having a typewriter-like keyboard permitting two-way communication between the computer and the user.

Text. One or more lines of information consisting of any series of numbers, letters, or special characters.

Time. Times are specially treated by CLINFO. They are entered and displayed as times on a 24-hour clock with the hour immediately followed by the minute. For example, 8:37 A.M. is represented as 837, and 5:03 P.M. is represented as 1703. The value 0 represents midnight of the previous date, and 2400 represents midnight of the current date.

Update File. An internal file which contains data entered via the enter activity prior to its being merged with the SDF. Data in the Update File may be reviewed (and deleted) by using enter, option 3.

Value. A quantity or quality associated with an item.

Worksheet. A rectangular array of numeric data with user-supplied name, title, and row and column labels. A worksheet may contain a maximum of approximately 2000 data items.

Yellow Key. A key on the AA keyboard used to move directly to an activity or subactivity. One is a solid-yellow key labeled CLINFO. Only keys with abbreviated activity names (lana, leal, etc.) in yellow are called yellow activity keys. Pressing the CLINFO key and a yellow activity key simultaneously is equivalent to typing the label (e.g., lana,) on the yellow activity key. Other keys with yellow labels are a PRINT key and the ESC key.