Introduction

Litigation poses a series of interesting problems whose solutions are simplified through the use of computers and efficient, easy-to-use data management systems. The opportunities for using computers in support of litigation have been increasingly recognized by members of the legal profession (see Computer/Law Journal, Vol. 1, No. 4, 1979, "Computer-Related Evidence Law"). With the decreasing cost of microcomputers, the general increase in computer power for a given dollar, and the increasing complexity of litigation, more and more attorneys are looking toward computers to provide support.

Given the large amount of information needed in the preparation of a case, what is really needed is a means of storing, organizing, rapidly retrieving, summarizing, and analyzing information. Thus the primary need in litigation support is for an easy-to-use, efficient data management system. Further, in some forms of litigation at least part of the evidence is based on statistical information. The role of the consultant is often twofold. Not only must the consultant implement the data management system, but also often perform the analyses, generate the exhibits and testify concerning the results. What the consultant needs are the tools to accomplish these tasks.

The SAS system provides a unique and useful set of data management and statistical analysis tools, in one package. There appears to be a misconception among many individuals that the SAS system is for statistical analysis and nothing else. A large number of users fail to realize that the real power of the SAS system is in the area of data management. It is common knowledge among individuals who make a living analyzing "real world" data that the actual analysis of the data may consume only a small part of the resources and manpower allocated to a project. It is the preparation of the data prior to the analysis that consumes the largest portion of the resources and usually generates the most frustration.

The application to be described here centers around the use of the SAS system on an IBM 3081-Model D in the support of Employment Litigation. While the particular application to be described centers around the use of the SAS system on a large mainframe, with its increasing availability on smaller computer systems, depending on the size of the database, similar applications could be implemented on smaller systems. It should also be noted that microcomputers do form an integral part of the particular application to be described here.

The SAS System has been found to be particularly useful at three of the major stages in the litigation process: 1) Data preparation/database maintenance, including data entry, auditing, cleanup, and validation; 2) The generation of procedural documents and materials, including Requests for Production, Admission and Findings of Fact; and 3) The performing of statistical analyses and the generation of exhibits.

Data Preparation

Aside for the legal issues of a case, perhaps the biggest problem facing an attorney is dealing with the massive amount of information that must be analyzed. This is particularly the situation with employment litigation where a significant proportion of the evidence to be presented is statistically based. This is not to say however, that the need to generate and present statistically based evidence is the only factor contributing to the massive amount of data associated with complex litigation. If data is viewed simply as pertinent information, then the potential amount of data associated with a case is rather large indeed. In line with this definition of data, the term "data analysis" will be used here in a broad sense and not limited to statistical analysis alone. Data analysis as it is to be used here would include data preparation, maintenance, synthesis, manipulation, analysis, and summarization applied to a variety of types of data both quantitative and qualitative.

As most practitioners of the art of data analysis are aware, a considerable amount of time, effort, and resources are expended simply getting the data ready to analyze. In fact, only a relatively small part of the resources are consumed actually performing the analyses.

The data preparation stage is of critical importance in litigation support. Not only must the data that is to be used meet the stringent requirements of the courts with respect to accuracy (usually an error rate of at or less than 2%), but the data must be easily and readily accessible. Data preparation in not, however, always a straightforward task. It is simply wishful thinking to believe that there are standard accepted "cookbook" approaches to the problem. As in most situations, the selection of applicable and useful methods is influenced by a variety of factors, including the general characteristics of the data, the resources available, the time available, and the particular questions that have been raised. While no "cookbook" approach exists that is simultaneously specific and general enough to fit all situations, there are a series of general guidelines that deserve mention.

It must be remembered that the primary goal is the conversion of the data to computer databases that can be easily searched and analyzed. The specific approaches that must be taken are dictated by the characteristics of the data including its original format.
Data Sources and Media:

What are the sources of the data? In employment litigation, these include any and all personnel and employment records. The list of possible data sources is simply too large to list here, however, the data sources that exist may be classified into two broad areas, (1) information (both quantitative and qualitative) which will be analyzed and summarized for trial presentation and (2) information that will be used to keep track of the process and aid in the procedural requirements of the process. For one list of possible sources included under (1) above, see "The Plaintiff's Discovery Checklist", Equal Employment Practice Guide, Federal Bar Association, May, 1980, Vol. II. Examples of the types of data included under (2) will be discussed in more detail later and include such things as on-line storage of Depositions, generation of Requests for Admission and Production, etc.

The data that is needed and used in litigation support may be originally stored on a variety of media. The general nature of the media or format of the data when it is delivered to the consultant will usually vary depending on whether the consultant is working for the Plaintiff or the Defendant. If the consultant is employed by the Plaintiff, then it is quite likely the information will be supplied in the form of paper documents which must be converted to a computer readable format. If, however, the consultant is employed by the Defendant, it is quite likely that the bulk of the employee records may already exist in a computer readable format or even as part of some database management system. In general, however, the information that is provided will be in the form of paper documents, computer diskettes, or computer tapes. Each of these different media requires a slightly different strategy for their conversion.

Building the Databases

One of the key needs in litigation support is for a system that allows for the storage, integration, retrieval and analysis of a variety of different types of data. What is needed is some type of centralized data management system. This is not to say that the goal should be the building of one gigantic database. Such an approach is not only costly, but will prove rather cumbersome. Rather, what is needed is a library of databases or data sets that can easily be cross-referenced and merged. The key to such a data management system is integration. The databases must be integrated, not only to allow for generating the analyses, but also to allow for cross-checking the data conversion procedures and to determine the validity and accuracy of the data.

In litigation support the time that may be available for the retrieval of information and the generation of specific analyses is seldom under the control of the consultant. The timely production of information is of key importance. This can not be easily be done if the data is scattered across a variety of incompatible databases or data sets.

Why SAS?: There are several advantages to using the SAS system for building and maintaining such a centralized information management system, aside from the general advantages of ease of use, flexibility and availability.

The primary advantages to using the SAS system are its flexibility and the fact that most of the functions that are needed can be performed within the system, including data preparation, report generation, statistical analysis, and the generation of exhibits. Even if the SAS system does not meet a specific need, its data manipulation capabilities allow the user to easily write a custom output file that can be used with other systems.

One of the specific advantages of the SAS system is the MERGE statement. Data from a variety of different sources or documents can be stored as separate SAS data sets. If, however, there is at least one common variable, hopefully unique, shared by one or more of the data sets then these data sets can be easily merged to allow for the direct comparison and analysis of data across data sets. For example, it may be that each of the data sets contains a unique employee identifier such as Social Security Number, or Employee Number, or a unique identifier may be constructed based on an Employee Number and a Crew or Plant Number.

The second major advantage of the SAS system is simply its data manipulation capabilities. Both numeric and text variables can be easily manipulated. With such data manipulation capabilities, it is also quite easy to generate custom analyses files. For example, the questions that have been raised may concern movement to and from different job titles. The information that exists on the data sets may consist of a payroll record, one for each pay period, containing the job title the individual held during that pay period. The question, however, is not when an individual held a job title, but rather, how long and what job titles were held before and after. Using the SAS system, an analysis file which contains a job title with the first and last dates held and the previous and future job titles can be built. The amount of time that is needed to build such an analysis file may vary depending on the exact nature of the variables that are being considered and the complexity of the questions that have been raised. However, it will be suggested that in the absence of a data management system with the capabilities of the SAS system, that the construction of such an analysis file can present a formidable task.

The third advantage is SAS/FSP, particularly FSEDIT. FSEDIT provides not only
a means of data entry, but perhaps more useful, a means of editing and cleaning up data that has been entered using other means. The author has had no problem using FSEDIT to edit SAS data sets with as many as 100,000, 100 to 150 byte records.

The primary disadvantages to using the SAS system have centered around its past availability only on mainframes or mini's. Access to a mainframe is not always easy to come by. Hopefully with Version 6 this will change. A second disadvantage concerning data entry, revolves around the fact that only one individual can easily access a given SAS data set at one time. This problem, however, can be circumvented by entering data into multiple data sets and then concatenating the data sets.

Converting Data to SAS data sets:

Given that the SAS system is to be used, then the task of converting the data to SAS data sets still exists. The nature of the approach to be taken depends on the media on which the original data exists. For example, there needs to be a different strategy employed if the original information exists on paper documents, microfilm, computer diskettes, computer tape or existing computer databases. Even if the needed information exists on a current computer database system, it will be strongly recommended that the information that is needed and only the information that is needed be extracted and used to build a separate database which can be made up of one or more data sets (Yancey, 1982).

The process involved in determining exactly what information should be included in the database is far too complex and variable to be discussed here. In general, however, if the information that is to be provided for the Plaintiff then any and all information that could be of possible use and can be obtained is included. If the information support is being provided for the defendant, only the information that is deemed to be most directly relevant to the specific questions that have been raised should be included.

Conversion of Paper Documents:

Once the documents have been obtained, numbered, the information that is to be included in the data sets identified, and the basic structure of the data sets defined, then the problem becomes one of building the data sets and then verifying that the information contained in the data sets accurately reflects the information contained on the documents.

Several different approaches have been used by the author, from the use of commercial keypunch services, to the entry of the data as sequential files on micro's, mini's and mainframes. Two approaches, however, have proved to work the best. The first of these approaches involves the direct entry of the data into a SAS data set using SAS FSEDIT. The second involves the use of microcomputers and RBASE 5000 for data entry and then uploading the data to the mainframe. There are several advantages and disadvantages to both of these approaches.

Data Entry Using SAS FSEDIT:

One of the important aspects of any data entry situation is accuracy. Systems which have audit capabilities that can be invoked at the data entry stage are particularly useful. SAS FSEDIT provides some of these capabilities. The package also allows the construction of custom data entry screens that can greatly aid in the process. The technical expertise of the individuals doing the data entry does not need to be as high or their experience as great as for individuals entering data into sequential files. Further, the data is entered directly into the data set.

There are, however, some disadvantages including following:

The audit capabilities of SAS FSEDIT are not as extensive as are sometimes needed. For example, while minimum and maximum values can be declared and checked, there is no capacity for table lookup. Additionally, only one user can enter data into a given SAS data set at one time. This last point can, however, be overcome by creating parallel data sets for data entry purposes, which are later concatenated with a simple SET f1 f2 f3;.

One last disadvantage, which will hopefully disappear with Version 6, has been that using SAS FSEDIT requires access to a mainframe or mini.

Data Entry with Microcomputers and Upload to Mainframe:

It may be that direct access to a mainframe and SAS FSEDIT is a problem, while access to microcomputers may not be. Some of the newer microcomputer database management systems such as RBASE 5000 are extremely useful and powerful. Such a system can be used to build custom data entry screens with extensive audits. For example, with RBASE 5000, not only can minimum and maximum values be set and checked, but also the value that is entered can be compared against a table of values. These custom data entry screens with built in audits can be easily constructed and implemented. Once the data have been entered, a limited amount of auditing, checking and data cleanup can also be performed using the microcomputer.

By using microcomputers, the data entry process can occur at a variety of sites, simultaneously on multiple machines, with the final product being shipped to the consultant for upload to the mainframe.

There are, however, several disadvantages to using microcomputers in this capacity. The primary ones concern speed and storage capacity. These two concerns are really interrelated and are particularly a problem if the micro system that is being used has no hard disk or similar storage media. The author has found that one very workable system includes a microcomputer hooked to an IOMEGA Bernoulli box with twin 10 meg. removable cartridges.
Not only does such a system solve the speed and mass storage problems, but also the data can be shipped back and forth via express mail in 10 meg. chunks. Further, backup becomes a trivial matter and since the cartridges are removable, they can be easily locked away to insure data security. Another disadvantage to the use of microcomputers is that the data must usually be written out in ASCII format before it is transferred to the mainframe. To convert the data to a SAS data set, an INPUT statement must be used, and columns counted, etc. It is hoped that Version 6 of SAS and the accompanying microcomputer versions will simplify this process. Also some means of uploading the data must be found. The package used by the author is KERMIT (Columbia University, 1984) which has extensive error checking features.

Conversion of Computer Diskettes:

If the data is provided on computer diskettes, it is quite unlikely that it will either be in a disk or storage format that is easily readable with the equipment available to the consultant. There are, however, occasionally a few pleasant exceptions.

If the diskette cannot be read using equipment available to the consultant then the first step is to have the diskettes converted to a format that can be read. This may involve conversion to another diskette format or the conversion to tape. If the information is converted to tape then the problems associated with the conversion of such data to SAS data sets will be discussed in more detail later.

If the size of the media is the same as that used by the equipment that the consultant has access to, it may be that one of the utility routines such as UNIFORM that allow microcomputers to read diskettes written on different machines may be used. For example, by using UNIFORM, IBM PC's or compatibles can be used to read diskettes written on Osborne or Xerox machines, to mention a few.

Let us assume that the data has been provided in a diskette format that can be easily read by the consultant or converted to such a format. If the data resides on the diskette is written as ASCII, then it is simply a matter of uploading the data to the mainframe and using an INPUT statement. It may be that the data is written as ASCII, but in some strange configuration that will need to be converted. Depending on the complexity of the conversion process and the availability of existing software to do the conversion, a decision will need to be made as to where the conversion will be attempted, on the micro or on the mainframe.

If a custom routine will need to be written, depending on the complexity of the routine, it will be suggested that the data manipulation and handling characteristics of the SAS system can be used to their best advantage. However, the decision as to where the conversion will occur also depends on the amount of data that is involved, cost, and available resources.

Conversion of Computer Tapes:

It may be that the data has been provided on computer tape. If the tape was written on a compatible machine and in a compatible format, then there are fewer problems. However, it is more likely that a considerable amount of translation will have to be done. Assuming that the information can be converted, it still may be that the data may be in a structure that must be translated. For example, the data may have been written using COBOL and include COBOL "OCCURS" types of data structures. There are several approaches that can be taken in this situation. The first would be to use a program written in the same language to write the data out an ASCII. It should be noted, however, that the SAS system can be used to read a variety of tape formats, including those written using COBOL. The primary goal would, however, be to get the data in a format that can be read by the SAS system.

Data Auditing Strategies:

In litigation support, the accuracy of the data is of prime importance. The term accuracy as it is and will be used here refers to how well the data on the computer data sets reflects the information contained on the documents or other sources from which it was derived. Accuracy as it is used here is not intended to refer to how well the information contained on the data sets reflects reality.

Auditing at Time of Data Entry: If the data must be entered from paper documents, then ideally, the majority of the auditing should be applied during the data entry process, since it will be at this point that the documents will be before the person doing the data entry and most easily checked. Audit runs performed after the data has been entered reveal potential problems which require that the original document be retrieved and checked.

The two data entry methods which have been discussed, using SAS FSEDIT and microcomputer entry with upload to a mainframe, both have advantages and disadvantages.

With the current version of SAS FSEDIT (pre-Version 5), for example, minimum and maximum values can be set and used to flag out of range values at the time of data entry. However, it is usually the case that more complex audit strategies need to be employed. For example, in the case of payroll information, it may be that hours worked, pay rate and gross pay are all entered. This information can be used to flag occurrences where the data is not equal to gross pay. The original document can then be immediately checked to see if there was a data entry error. If such an error has occurred, then it can be corrected on the spot. If it is
in fact the case that the values contained on the document do not add up, then a note can be added using a note field that can be used to identify such records in the future. Further, the microcomputer database package that is being used, RBASE 5000, incorporates table lookup capabilities. If a value for a variable does not exist in a table, the entry will be flagged to be checked.

In general, the auditing capabilities of the microcomputer routines have been found to be considerably more extensive than those of SAS FSEDIT and it is for this reason that when complex auditing at the data entry stage needs to be done, the use of the microcomputer routines may be preferred.

Post Data Entry Auditing: The advantages of the microcomputer based software at the data entry stage, begin to disappear if the primary auditing is to be done after the data has been entered or if the data is converted from another computer readable format. While a limited amount of auditing can be done, such as checking for duplicate records, missing records, records out of sequence, and even some of the more complex auditing tasks such as the one mentioned with respect to hours worked, pay rate and gross pay, the time involved in accomplishing such tasks on a microcomputer can perhaps be better spent. This is particularly the case once the SAS data sets have been built. The data manipulation and programming capabilities of the SAS system can greatly simplify the auditing process. This is particularly the case if data from two or more different data sets need to be compared or if the data sets contain a significant number of records, say 1,000 or more.

For example, one problem that is often encountered is establishing the uniqueness of the individuals contained in the data sets. It may be that a question has been raised as to whether or not an individual or set of individuals were given the opportunity to work the number of hours specified in a contract. To determine this the number of hours that were worked by each unique individual must be determined. Perhaps a number of unique or nearly unique identifiers exist, including name, social security number, and/or employee number. Names usually can not be used alone due to creative spellings and, in some instances, two or more individuals having the same name. The best identifier is usually Social Security Number, however, there is always the possibility of data entry or reporting errors. With the SAS system, it is a rather trivial task to generate name within Social Security Number and Social Security Number within in name lists that can be used to check the data. Similar lists can also be produced with microcomputer based software, given enough time and storage.

For the more complex audit procedures which require the merging of information for two or more data sets, the capabilities of the SAS system can rarely be met by the microcomputer based software. There are also certain auditing strategies that require checking the sequence in which particular events occurred. If each record represents an occurrence on a given date, then such audit procedures are relatively straightforward using the SAS system, but present an almost overwhelming task for most microcomputer based software.

Data Cleanup:

If the data has been entered from paper documents, then the audit runs which are performed will usually reveal at least a few data entry errors. The question then becomes one of determining the best way to correct these errors. If the data entry and the auditing has been done with microcomputers then the cleanup will perhaps best be done on the microcomputers before the data is uploaded.

If the data has already been converted to SAS data sets, then FSEDIT provides an ideal means of cleaning up the data. When the final cleanup is being done, it will be strongly recommended that one person be in charge of this process with respect to each data set. Using such an approach, the one person at time limitations of the SAS system with respect to accessing the data sets is no longer a problem. A distinction must be made at this point between data cleanup and data standardization. If the changes that are to be made are to make the computer version of the data more closely match the original documents, then the use of SAS FSEDIT with the direct updating of data items is still recommended. However, there are some situations that require the existence of an audit trail. For example, it may be that there are some obvious problems with the data such as two different Social Security Numbers that differ by only one digit for the same individual. To insure that an individual is uniquely identified, the Social Security Numbers will need to be standardized. To leave an audit trail, a separate SAS program can be written, saved and used to document the changes. This insures that the process that was undertaken can be easily reconstructed.

Data Validation:

Separate from the auditing strategies which are designed to check how closely the information contained on the computer data sets match the information contained on the original documents, is the process of data validation. Data validation involves determining how consistent the information contained on the data sets is both within each data set and across data sets. For example the validation process might include checking the range of values for a given variable. Given a weekly payroll, a notation that an individual was paid 96 hours in a given week should indicate that a closer look should be taken at that particular record. More likely the validation process will require cross checking data contained on
two or more data sets. For example, one data set may contain records that purport to indicate that an individual was promoted to a higher paying job on a given date or that an individual was hired on a given date. A cross-check of the payroll records, however, may reveal that the individual started receiving pay at the higher rate 6 weeks prior to the time that the promotion was noted or that the individual who was hired never appeared on the payroll.

The implementation of the strategies needed to validate the data is usually considerably more complex than the implementation of the data auditing strategies. While some data validation can be performed using microcomputer based software, it will be strongly recommended that the validation process be delayed until the data has been converted to SAS data sets. The data manipulation and management capabilities found in the SAS system can greatly simplify the validation process.

The primary advantage of the SAS system can be seen when data from a variety of data sets needs to be cross-checked. The match/merge capabilities of the SAS system greatly simplifies this process. For example, given personnel records and payroll records keyed by Social Security Numbers, it is a rather easy to merge these data sets and cross-check the information contained on each.

Procedural Functions

The data management capabilities of the SAS system can be quite useful with respect to some of the procedural aspects of the litigation process. To better understand some of the procedural functions, it would perhaps be best to spend a little time reviewing some of the concepts involved. The first of these concepts is the notion of discovery.

One of the critical stages in the litigation process is discovery. At this stage the court will require that the defendant provide data or information to the plaintiff to allow the plaintiff to try and establish a "prima facie" case. The courts will often require that the defendant provide copies of personnel documents and records to the plaintiff. The plaintiff will file Interrogatories and/or Requests for Production asking for answers to a series of questions or requesting that specific documents be produced.

Requests for Production: Part of the Discovery process involves the generation of requests for production. The original data that was obtained and is to be analyzed was most likely acquired via a request for production. Once the initial data sets have been built, it may become clear that the information is either incomplete or contradictory and additional requests for production may be needed. It will be assumed that the additional information that is needed is required to answer some specific questions that have been raised by a preliminary review of the initial data. For example, it may be that a preliminary review payroll data indicates that individuals may have worked more than 40 hours in a given week and yet were not paid overtime. If the data sets have been constructed to include the document numbers that are assigned to exhibits that are to be used in a trial, then the SAS system can be used to computer generate the requests for admission with specific references to the documents in question. Further, the general practice is to sequentially number the requests for production so that they can be uniquely identified. The report writing capabilities of the SAS system provide a means of doing this.

Requests for Admission: Once the data sets have been constructed, audited and validated, then before the information contained on the data sets can be used, it must be admitted as evidence. Given the amount of data that is involved the SAS system provides a way of computer generating these requests. The evidence may be based on a variety of sources. For example, it may be necessary to have as evidence the payroll record for each pay period for each individual under consideration. The total number of these records may be in the tens of thousands. To make sure that all of the data that is needed is placed into evidence, the required information would need to be extracted and typed out in an appropriate format. The generation of the Request for Admission gives the opposing side the opportunity to review the data and object if necessary.

Requests for Admission: Once the data sets have been constructed, audited and validated, then before the information contained on the data sets can be used, it must be admitted as evidence. Given the amount of data that is involved the SAS system provides a way of computer generating these requests. The evidence may be based on a variety of sources. For example, it may be necessary to have as evidence the payroll record for each pay period for each individual under consideration. The total number of these records may be in the tens of thousands. To make sure that all of the data that is needed is placed into evidence, the required information would need to be extracted and typed out in an appropriate format. The generation of the Request for Admission gives the opposing side the opportunity to review the data and object if necessary.

Again, the general procedure for generating the Requests for Admission would involve sequentially numbering each specific request. The SAS system can be used to computer generate the requests and simultaneously sequentially number them.

Findings of Fact: Prior to the trial, a list of the specific facts that are to be considered in the case is normally generated by each side. In employment litigation, where specific employment practices and employment histories may be in question, this list can be very large. In fact, in many cases it is not uncommon to encounter Findings of Fact that contain 10,000 or more entries. The report writing capabilities of the SAS system can be used to generate these Findings of Fact.
Miscellaneous Applications: There are a variety of additional tasks which must be performed in the litigation process, not directly related to the production of procedural documents. There is always a considerable amount of information associated with a given case that is not contained on the documents obtained via discovery. Some of this information is vital to the attorneys for the preparation of motions, briefs, appeals, etc.

For example, a considerable amount of information may be obtained from depositions or from the actual trial transcripts. Such information may consist of hundreds or even thousands of pages. The minimum that is needed is some sort of indexing system. The SAS system can be used to build such an indexing system or to store the actual contents of the documents. Such an application is a bit more creative than those that are normally encountered. To use the SAS system to build an indexing system, key words, defining key categories or legal issues are determined and then the documents are scanned by the attorneys to identify the occurrence of testimony relevant to each of these categories. A data set is then built containing the category, location, page and line numbers of this information on the original documents. A more ambitious approach involves entering the entire text of the documents and using the SAS system to perform key word or key string searches. A similar approach can be taken using microcomputer based software such as RBASE 5000.

Analyses and Exhibit Generation

Statistical analyses provide only one source for exhibits. A number of the exhibits may also consist of lists of individuals who meet certain criteria. This is particularly the case when a class action is involved. The criteria that are used are often dictated both by the court and the results of the statistical analyses that have been performed.

Statistical Analysis: There is simply not enough room here to discuss or even give an overview of the various statistical techniques that have been or might be used in litigation support. The following general points will, however, be mentioned.

Do not look for "cookbook" approaches, they do not exist.

Do not expect to receive much guidance from the courts. The "rules of thumb" that are employed by some of the courts with respect to the use and interpretation of statistical evidence are seldom based on sound statistical theory. There are some exceptions, however, see Boykin v. Georgia-Pacific Corporation, 706 F.2d 1384, 1983.

The range of statistical procedures available in the SAS system is sufficient to satisfy the majority of the needs in litigation support. If an particular method is not available, then it may be possible with PROC MATRIX to perform the necessary calculations. It may also be possible to write out the data for use with another analysis package. For example, prior to the release of Version 5, the SAS system did not have a good, easy to use, Log-linear routine, while SPSS-X did (LOGLINEAR, HILLOGLINEAR). PROC TOSPPS makes it a rather simple task to write a SPSS-X getfile to be analyzed with one of these two routines. Even if another package is to be used to do the majority of the analysis, it will still be suggested that the data management capabilities and data manipulation capabilities of the SAS system make it a preferred choice for storing the data.

One of the real advantages of the SAS system is that summary statistics such as means, standard deviations, counts, etc. can be written out to files for latter analysis or used for producing graphs. This is particularly useful in preparing exhibits for presentation. The general output format produced by most statistical analysis routines, including the SAS system is appropriate for most statisticians and researchers, but not for presentation in court.

Modeling: The modeling that can occur in support of litigation covers a broad area of circumstances and applications. The modeling may be based on statistical models or on employment practices or contracts, etc.

For example, consider a situation where one of the questions concerns promotion and potential differences between the length of time it takes for males and females to be promoted. One approach would involve deriving a model (usually a regression model) that can be used to predict the length of time until promotion for the males. Assuming equal opportunity and an equal functioning of employment practices and policies then the model that has been derived for the males would be applied to the females and the differences between the observed and actual times for promotion examined. A plot of the residuals can be rather revealing.

A second example of an application of modeling could involve questions as to whether or not the terms of a contract are being met. For example, a given contract may specify the policies and procedures for acquiring vacation or sick leave or for payment of overtime, etc. The specifics of the contract can be used to build a model which would calculate the expected values for an employee under the terms of the contract. These expected values would then be compared against the observed values. The use of modeling can also be extended to a variety of selection situations, such as hiring, selection for promotion, etc.

Graphics: One of the best media for summarizing and presenting the results of statistical analyses in support of litigation is graphics. SAS/GRAPH provides a variety of tools and approaches for accomplishing this task. In litigation support, the capability of...
rapidly and inexpensively generating graphics is of prime importance. It is quite likely that in complex litigation as many as 200 or 300 charts will need to be produced in a relatively short time period. Since these charts are designed to communicate the results of the analyses, the format and contents of each of the charts is of prime importance.

Given that the data is stored as SAS data sets, then the production of most of the charts are relatively straightforward.

In the past, one of the disadvantages to using SAS/GRAPH was the lack of a means of annotating the graphs. Charts which are to be presented as evidence often require more detailed annotation than the average statistical chart. The capabilities of Version 5 should, however, solve some of the shortcomings of SAS/GRAPH in this area.

Even in situations where SAS/GRAPH does not satisfy the requirements with respect to the production of graphics, this has not proved to be an insurmountable problem. The approach that has been taken has been to use the SAS system to write ASCII files containing the information that is to be graphed and download this information using KERMIT to a microcomputer. Microcomputer based graphics software such as Microsoft CHART has then been used to produce the necessary graphics.

While the exact situation may vary for each user, the author faces a situation where SAS/GRAPH can not be used to directly drive a graphics device and the only output media is a ZETA plotter or a file that can be transferred to another machine and then viewed on a graphics device. This situation greatly impedes the development process. The primary advantage of using the microcomputer based graphics routines is immediate feedback along with the added flexibility. However, in the situation where a large number of charts must be prepared, the speed and convenience of SAS/GRAPH will often compensate for the additional development time. Hopefully the availability of SAS/GRAPH on the micro will solve some of these problems.

Summary

It has not been the purpose of this paper to suggest that the SAS system provides all of the tools that are needed for litigation support, but rather to point out that the primary need in such an application is for a fast, efficient, easy to use data management system. The SAS system can provide the central core of such a system. The data management and manipulation capabilities of the SAS system provide the means of creating data files that can be used with the SAS system or a variety of other packages. One key point to remember in litigation support is that to get the job done you will have to use what is available, when it is available. It is just that if the SAS system is available, you will most likely find that the time available for sleep and other necessary bodily activities more closely resembles the normal requirements than if other systems have to be used.

**REFERENCES**


