SAS® Applications at the Frankfurt University Hospital

Harald Pitz, Frankfurt University Hospital
Marcus Frenz, Frankfurt University Hospital
Hans-Peter Howaldt, Frankfurt University Hospital

1. Introduction

Since 1988 the SAS System is available at the Frankfurt University Hospital. For a test period of 2-3 years 10 licences of the PC-Version were purchased. A main reason for this decision was the announcement of procedures for survival analysis, especially PROC LIFETEST including the Kaplan-Meier survival function.

The SAS System is used in the departments of biochemistry, cardiology, internal medicine, maxillo-facial surgery, medical education, neurology, in a tumor registry, in the central laboratory and in a research group running HIV-studies.

To estimate the further demands of SAS licences a questionnaire was developed and distributed to all SAS users. The most important questions and results will be described first. Four of the largest applications will be presented.

2. Questionnaire

The SAS System is distributed by the department of medical informatics. To get an overview of the departments and their use of SAS a questionnaire was developed. Beside common questions about the research fields, hardware and software other than SAS detailed questions were asked about the frequency and duration of using the SAS System and the single modules, manuals, advantages and disadvantages, operating systems, training, and alternatives.

The most interesting results can be summarized as follows:

The main areas where the SAS System is used are adminstrative statistics (e.g. OP-statistics) and research statistics (e.g. survival analysis). Most of the users are physicians with more or less experience with computers. Only two computer scientists are working with SAS.

SAS Modules

Presently the hospital has licensed Base SAS®, SAS/GRAPH®, SAS/STAT®, SAS/IML®, SAS/AF® and SAS/FSP® software.

SAS/STAT is used in all departments. The data is entered to the SAS System very differently. Some departments use SAS/FSP and PROC FSEDIT, others use DBASE or the medical record system BAIK [Giere88] which provides special medical data types not included in the SAS System. A special interface has been developed to transfer data from BAIK to SAS [Pitz91].

Producing graphics by means of SAS/GRAPH is hardly practiced in the hospital because the customization of graphics is a very troublesome work with this module. Furthermore the results cannot be compared to those obtained by popular PC-graphic software such as Freelance Plus or Harvard Graphics. SAS/GRAPH is only used at those departments where curves consisting of data computed by SAS procedures have to be plotted. For example PROC GPLOT is used to plot Kaplan-Meier survival curves. A plot macro has been developed which is able to plot survival curves and censored variables in one graph.

A user interface based on SAS/AF is used by 2 institutions.

Hardware and OS

The departments mainly use 80386 or 80386SX computers. Due to the PC-DOS restriction of 640 KB memory some departments have asked for an OS-2 or UNIX-Version of SAS.

Advantages and disadvantages

The wide capabilities of the SAS System were most often mentioned. Other aspects were the possibility of automation, processing of large datasets and the modular design of the system.

The disadvantage of the large initial period to familiarize with the SAS System was pointed out by all users. This is an extremely important factor in a hospital environment where most of the physicians are non-EDP-experienced. Handling the large set of manuals is quite troublesome for occasional users. The help, menu and cmenu facilities are mainly useful to those who are familiar with the general concepts of the system. SAS/ASSIST® facilitates the use of the SAS System but does not provide the complete functionality. Many popular PC-based database systems are easier to handle than the SAS System. Other disadvantages mentioned are problems due to the memory limitation of PC-DOS, the complicated syntax and the difficult generation and customization of graphics.

3. SAS Applications in the Hospital

Several departments and institutions performing studies use the SAS System for analysis of the data. Four examples give an overview about the largest applications.

3.1 The DÖSAK tumor registry

Overview

In 1989 the German-Austrian-Swiss cooperative group on tumors of the maxillo-facial surgery (DÖSAK) established a multicentric tumor registry for head and neck cancer. 78 departments of maxillo-facial surgery in the three participating countries fill in standardized documentation forms and send them to the tumor registry in Frankfurt. During the first 21 months more than 2,800 patients could be collected.

The registry's software environment mainly consists of the medical record system BAIK, a BAIK-SAS interface and a SAS application for analysis.

BAIK

The medical record system BAIK is used for data entry, generation of free-text reports and thesaurus based free-text search. A large set of plausibility-programs has been developed to check logical consistency of the incoming data. BAIK produces a free-text report for every patient containing the most important information. This allows the physicians to check the correctness and completeness of the documentation. An online-query system allows to perform powerful free-text search with the support of a thesaurus containing various relations. We primarily use this thesaurus to detect words with different spelling (e.g. tumor - tumour) or synonyms (e.g. tumor - malignancy).

The BAIK-SAS interface

A BAIK-SAS interface has been implemented to export data entered in BAIK to the SAS System. The interface allows individual selection of patients and data-items. An ASCII-file containing all necessary statements including length, label, format and informat statements to build up a SAS dataset is automatically produced.

The SAS Application

A SAS application has been developed to perform various analyses and statistics. Based on frequently use of macros a large set of modules have been implemented and allow automated analysis of either the complete database, of a specific institution or individually for all institutions. The capability of automation is very important to handle the large amount of data. The application automatically generates titles including the selected set of data and the number of observations. Graphical output can be produced for the screen, a laser printer, a graphics catalog or as CGM metafile.

The registry uses the same methods for survival analysis as described in 3.2.

The user-interface is actually based on %WINDOW and %DISPLAY statements and allows easy input of the necessary options. An interface using SAS/AF software is under development.

A more detailed description of the application can be found in [Pitz91].

3.2 Department of maxillo-facial surgery

Tasks

Since 1983 the head- and neck tumors treated at the department of maxillo-facial surgery are documented by means of the BAIK system. The evaluation of patient's outcome is done by survival analysis [Howaldt90]. A comprehensive data collection for each patient has been built up in order to identify those parameters which influence survival. They are commonly called prognostically relevant factors.

The knowledge of prognostic factors is essential for performing controlled clinical trials which is the only way to confirm progress in cancer treatment. Furthermore these factors help physicians to optimize the therapy for each patient. Some prognostic factors can be collected before any cancer-specific treatment, e.g. tumorsize, depth of infiltration, fixation of lymphnodes etc. Other factors are available after surgery.

Beside clinical findings the usefulness of other therapy modalities e.g. chemotherapy is to be explored.

The prognosis of cancer patients is most adequately measured by the duration of survival beginning with the first diagnosis of the tumor. The intention is to explain the observed survival with potential prognostic factors.

Parametric and semiparametric survival analysis

An appropriate tool to evaluate prognostic factors is the SAS/STAT procedure LIFETEST which implements fully parametric survival analysis. Parametric survival analysis needs information about the underlying distribution of failure times. Usually considered distributions are exponential, gamma, Weibull, Gompertz-Makeham, compound exponential, log normal and log logistic. PROC LIFEREG implements four of them. An appropriate distributional form has to be assessed before starting analysis. The semiparametric COX-model which is implemented in the SUGI Supplemental Library is an alternative method of survival analysis, if the distribution form is unknown and no reasonable assessment can be achieved. Unfortunately the SUGI-Library is not available

on the PC-Version but the COX-model can be implemented in the Matrix-Programming-Language SAS/IML with a reasonable amount of work. We have choosen the fully parametric survival analysis with the log-normal distributional form for our analyses. Comparison with the Cox model showed high similarities of the corresponding covariate effects. This can also be expected from theoretical results [Galler90].

Exploring prognostic factors with PROC LIFEREG

First the data are exported from BAIK to the SAS System. With the help of the SAS datastep and the facilities of the SAS macros the raw data is processed and datasets with the whole collective and some subgroups are created. For each sample we build up a prognostic model by means of PROC LIFEREG. An experienced physician selects a set of potential prognostically relevant factors. Dichotomous variables can be directly used as covariates (e.g. fixation of lymphnodes (yes/no)). Categorial and ordinal-scaled variables with n different values have to be mapped to n-1 dummy-variables (e.g. location of the tumor). Cardinal scaled variables such as tumorsize can be used as covariates either directly (if a linear influence can be suggested), transformed (so that the transformed variable has a linear influence) or have to be mapped to dummies, each representing an interval of the response-scale. The partitioning of the response-scale should be based on a histogram of the variable.

Now we can test our prognosis model. To reduce the model we can eliminate step by step the most insignificant covariate, until the log-likelihood of the model becomes significantly smaller (more negative). Such models may still contain insignificant single effects but keeping such variables turned out to be of important prognostic value. The result of the stepwise elimination is a parsimonious model of prognostic relevant factors. The prognostic influence of the covariate can easily be computed using the output of PROC LIFEREG.

Departmental Information system

Since 1988 the department has established an information system [Pitz89]. One of the main features is a module for the documentation of the surgical operations including an automated book of operations. The data is used to produce a yearly operation room statistic for the department and each surgeon with SAS.

3.3 Office of Medical Education

Background

Since 1988 oral examinations have been introduced in addition to the nationwide written examinations. The Office of Medical Education runs a project called INTOREX to documentate and analyse the impact of these additional oral examinations on the process of studying.

Questionnaire

A questionnaire containing questions about sequence, duration, management, synchronisation and perspective of the study behaviour, preparation for the exam, orientation to learning tasks, processing of learning tasks and biographical and social data was sent to all students who enrolled themselves for the second part of the national clinical examination. To compensate local characteristics the questionnaires were also sent to a second medical school.

One questionnaire corresponds to 415 Variables. This was one of the main reasons to use the SAS System. Many database systems have a limit of 256 variables per observation which would require a splitting of the observation and results in a more difficult data-entry. Within a period of 3 years 6000 questionnaires have been returned. The data set has been created with PROC FSEDIT. The screen customization has been quite expensive due to the large set of variables and necessary subsequent changes of the dataset. Input checking while data entry has been realized by defining field attributes. As most of the variables are digits and numbers the attributes MAXIMUM, MINIMUM, REQUIRED have been used most frequently.

Data-entry and analysis

To reduce the large amount of data to be entered (6000 * 415 items) a two-phase procedure was chosen: In the first phase 1500 questionnaires were entered. Then a pre-analysis of these 1500 observations was done to find out which variables are important and which are not. The data-entry of the remaining questionnaires was then restricted to the important variables.

The main methods and procedures to analyse the questionnaires are bivariate analysis using PROC FREQ and regression analysis using PROC REG.

First results of the project can be found in [Drolshagen90].

3.4 Central Laboratory

The central laboratory uses the SAS System to process the data from various clinical studies with large datasets. Constantly there is a need to establish reference values of new blood tests (for instance indicators of myocardial infarction, tumor markers). In addition it is involved in a large study of progression of HIV-disease in the Frankfurt area.

Reference interval study

Due to the differences in races, regions and methods applied the central laboratory constantly evaluates new tests and has to establish reference values. For instance

1700 healthy employees of an industrial company were tested in 1989-1990 for over 30 laboratory parameters.

The decision when to perform additional time and cost-intensive test is very important. So a hospital population with low activities of a basic clinical chemistry test (Alkaline Phosphatase) were evaluated to prove whether further differentiation of bone and liver AP is useful. Persons with a total AP activity very well within the normal range already showed a large number of pathological bone or liver AP-values [Siede90].

For the analysis the laboratory makes intensive use of PROC UNIVARIATE (box plot), PROC TABULATE, PROC FREQ, PROC ANOVA and SAS/GRAPH procedures.

HIV-study

The laboratory is involved in a long-term observational study of 2000 people with HIV-infection. The persons are examined every 3 months. Up to now more than 4500 visits have taken place.

The intention is to monitor and find clinical chemistry markers of HIV-disease activity [Rübsamen91].

4. Discussion

Various departments have used the SAS System during the past 2-3 years. Many of them perform large projects and have developed applications to process and analyse their data.

Mainly three reasons explain the decision to use the SAS System:

1. The SAS System provides a complete set of tools for data entry, graphics and statistics. Even large projects can be passed by using only one software product. The basic concepts of the SAS systems (e.g. PROC and DATA steps, HELP function, MENU system) are present in all SAS modules. If you want to replace a SAS module by another program (e.g. some departments use BAIK or DBASE instead of PROC FSEDIT) procedures for data conversion are available.
2. The SAS System provides most of the important statistical methods and computations necessary for a medical institution. The procedures used most frequently are descriptive analysis (PROC FREQ, PROC MEANS, PROC UNIVARIATE, PROC TABULATE), regression analysis (PROC REG) and survival analysis (PROC LIFEREG, PROC LIFETEST). Unfortunately the Cox's regression model is not yet available for the PC-Version.
3. The SAS System has no remarkable limitations concerning the size of datasets. In contrast to many PC-database systems a dataset may contain more than 256 variables.
4. The SAS System allows developing an application at one time and then to run the application whenever necessary. Especially long-term observational studies require the capability of repeated analysis.

One of the main disadvantages of the SAS System is the large amount of time needed to learn the handling of the system. Most of the scientific tasks in a hospital consist of data collection and subsequent analysis including graphics and presentations. They have to be done by the physicians. Building up such an application with SAS requires at least basic EDP-knowledge and experience which is available only by some of them. For small tasks such as limited collecting of data and computing descriptive statistics the long training period needed to get acquainted to the SAS System is not worth-while, especially not for medical stuff. Under such circumstances other possibly smaller and self-explaining computer programs are preferable. For long-term studies or tasks requiring some of the various statistical procedures included in the SAS System the initial amount of time is paying.

Learning the use of the SAS System consists of at least two steps. At first you have to understand the basic concepts and features (e.g. data and procedure steps, usage of DISPLAY MANAGER). Specific guides and the tutorial for beginners are suitable for that. The problem for non-EDP-trainеed persons is to apply these concepts to the specific tasks and problems. The second step is to find the necessary procedures and parameters to perform your tasks. This is supported by manuals and several powerful on-line features such as the help function, menu, cmenu and SAS/ASSIST. Again skilled personel will profit easier from these functions than others.

Unfortunately the SAS user interface is not yet based on a standard system such as MS-WINDOWS. This could facilitate the use of the menus and windows. The editing functions of the display manager including the commands for line insertion, deletion, moving or copying have to be similar to those of a word processing system. Release 6.06 under OS/2 is an important development towards a desktop system.

5. Conclusion

At the Frankfurt University Hospital the SAS System provides a powerful tool for EDP-experienced users who plan to use it for long-term projects. Except few functions (e.g. COX's regression model) all statistical methods normally applied in medical research are implemented.

Integrating modern concepts of user interfaces (especially window systems) will lead to a higher acceptance by potential users in the medical environment.

To introduce the SAS System in the hospital a consulting service has been established to give basic instructions and information to the users and to help them finding solutions for specific tasks and problems. Two computer scientists at the department of medical informatics provided this assistance and made the successful use of the SAS system at the Frankfurt University Hospital possible.

The authors thank Dr. H. Schneider, Frankfurt University, for helpful comments.

6. References

[Giere88] Giere, W.: Treatment, teaching and research - structural requirements for medical records, the information model of BAIK. In: Proc. of the 8th Joint Conference in Medical Informatics, Tokyo 1988.

[Pitz91] Pitz, H.; Howaldt, H.P.; Frenz, M.: The Integration of SAS-Software in a Multicentric Tumor Registry. In: Proceedings of the Sixteenth Annual SUGI Conference 1991.

[Howaldt90] Howaldt, H.P.: Aufbau und Ergebnisse einer klinischen Dokumentation für Tumoren im Mund-, Kiefer- und Gesichtsbereich als Grundlage für ein überregionales Register. Habilitationsschrift. Frankfurt am Main 1990.

[Pitz89] Pitz, H.; Howaldt, H.-P.; Giere, W.: A Heterogeneous MUMPS Network as Basis for a Departmental Clinical Information System. In: Proceedings of the 6th Conference on Medical Informatics, North Holland 1989.

[Galler90] Galler, H.-P.; Pötter, U.: Zur Robustheit von Schätzern für die Dauer der Arbeitslosigkeit. In: Schneider, H.; Hujer, R. (Eds.): Herausforderungen an den Wohlfahrtstaat im strukturellen Wandel. Campus Verlag Frankfurt, New York 1990.

[Drolshagen90] Drolshagen, S.: INTOREX2 - Introducing Oral Examinations and it's impact on medical students' learning and examination preparation approach. Paper presented at the AMEE conference (Association for Medical Education in Europe). Budapest 1990.

[Siede90] Siede, W.H.; Regeniter, A.; Kunz, G.; Seiffert, U.B.: Alkaline Phosphatase Isoenzymes in Patients with Reference Range Activity. 9th Internat'l congress of clinical enzymology. Prog. Med Lab 4, 472 (1990).

[Rübsamen91] Rübsamen-Waigmann, H.; von Briesen, H.; Regeniter, A.; Goetzens, M.; Gerte, S.; Scholz, C.; Biesert, L.; Nielsen, N.C.; Andreesen, N.R.; Brede, H.P.; Büttner, C.: Indicators of HIV-disease-progression. 1. Relations to virus cultures on lymphocytes and macrophages and to a decrease of serum cholesterol. The Microbiologist 1, 23-30 (1991).

Author:

Dipl.-Inform. Harald Pitz

Department of Medical Informatics
J.W. Goethe University Hospital
Theodor-Stern-Kai 7
D-6000 Frankfurt / Main
Germany

Tel.: (49) 69 6301 6642
Fax.: (49) 69 6301 6301