

SAS mit der Maus ?!

Rainer Muche
Friederike Rohlmann
Abteilung Biometrie und Med. Dokumentation,
Universität Ulm
Schwabstraße 13
89070 Ulm
rainer.muche@medizin.uni-ulm.de

Zusammenfassung

Das Statistiksoftwarepaket SAS ist in der klinischen Forschung weltweit sehr verbreitet. Die Möglichkeiten sind in Bezug auf Datenmanagement und Auswertungen sehr umfangreich. Zur Programmierung steht eine komplexe Syntax-Sprache zur Verfügung. Für Erstanwender stellt das Erlernen dieser Syntax-Sprache allerdings eine relativ hohe Hürde dar. Eine Alternative sind in heutiger Zeit maus- und menügesteuerte Oberflächen. Von SAS gibt es seit einiger Zeit ebenfalls solche Oberflächen, die jedoch nicht routinemäßig genutzt werden. Es werden in dieser Arbeit der Leistungsumfang und die Vor- und Nachteile des SAS-Analyst und des SAS-Enterprise Guide unter dem Gesichtspunkt des Einsatzes zur Einführung und Unterstützung in SAS genauer vorgestellt. Als Fazit kann man festhalten: Für einen ersten Einsatz von Statistiksoftware sind beide Produkte gut nutzbar. Sie helfen beim Einstieg in die Welt der Statistiksoftware, da die Einarbeitung das Erlernen einer komplexen Syntax-Sprache nicht voraussetzt. Aber: auf lange Sicht ersetzen diese Produkte den Einsatz der Syntax-Sprache im professionellen Umfeld der klinischen Forschung nicht, können aber in der Lernphase den Einstieg in SAS erleichtern.

Schlüsselworte: SAS-Analyst, SAS Enterprise Guide, SAS Learning Edition

2 Einleitung

Das Statistiksoftwarepaket SAS ist in der klinischen Forschung weltweit sehr verbreitet. Dies gilt sowohl für den universitären Einsatz z.B. an Universitätskliniken als auch in der pharmazeutischen Industrie. Die Möglichkeiten dieses Programmpaketes sind in Bezug auf Datenmanagement und Auswertungsmöglichkeiten sehr umfangreich. Zur Programmierung steht eine komplexe Syntax-Sprache zur Verfügung. Für Erstanwender stellt das Erlernen dieser Syntax-Sprache allerdings eine relativ hohe

Hürde dar; eine umfangreiche Einarbeitungszeit muss deshalb eingeplant werden. Aus diesen Gründen haben viele Ausbildungsstellen SAS in ihren Lehrplan aufgenommen.

Eine Alternative sind in heutiger Zeit maus- und menügesteuerte Statistiksoftwarepakete, deren Oberfläche an den Quasi-Standard der Microsoft-Windowsprodukte angepasst ist. Von SAS gibt es seit einiger Zeit ebenfalls solche Oberflächen, die jedoch nicht routinemäßig genutzt werden. Ziel dieser Untersuchung war es, diese Produkte bekannt zu machen und ihre Eigenschaften bzgl. des Einsatzes im Alltag zu untersuchen.

Es werden der Leistungsumfang und die Vor- und Nachteile der folgenden Programme, unter dem Gesichtspunkt des Einsatzes zur Einführung und Unterstützung in SAS genauer vorgestellt.

- SAS-Analyst (Version 8 und 9) [1]
- SAS Enterprise Guide 2.1 (in SAS Learning Edition 2.0) [2,6]

2 Kriterien der Bewertung

Die Bewertung der Programme für einen ersten Einsatz bzw. als maus- und menügesteuerte Alternative zur Syntax-Ebene erfolgt anhand der nachfolgend genannten Kriterien. Im Mittelpunkt stehen technische Details, Dateneinlese- und Datenmanagementfunktionen sowie die Auswahl an statistischen Methoden.

- **Installation, Rechnerausstattung:** Untersucht wird die Einfachheit der Installation. Es interessiert, ob die Fragen, die das System während der Installation an den Benutzer stellt, allgemeinverständlich sind. Angegeben wird, welche Ausstattung der Rechner benötigt.
- **Dateneinlesen:** Untersucht wird, welche Dateiformate eingelesen werden können. Als Quasi-Standard sollten MS-Excel-Dateien lesbar sein. Einschränkungen im Leistungsumfang werden dokumentiert.
- **Datenmanagement:** Folgende Managementfunktionen sollten mindestens vorhanden sein: Zusammenführen von Dateien über Schlüsselvariable, Funktionen zur Definition neuer Variablen, Klassierung und Codierung von Variablen, Datenfilter
- **Statistische Methoden:** Neben den Standardmethoden (Deskriptive Statistik, Korrelations-, Regressionsrechnung) werden speziell Grafikmöglichkeiten und

einige multivariate Verfahren betrachtet. An Grafikmöglichkeiten sollten u.a. Boxplots, Mittelwert- Standardabweichungsabbildungen, Histogramme verfügbar sein. Bei den multivariaten Verfahren werden verallgemeinerte lineare Regressionsmodelle (Lineare Modelle, ANOVA, Logistische Regression, Cox-Regression) untersucht. Möglichkeiten für Faktor-, Cluster- und / oder Diskriminanzanalysen werden auf Vorhandensein überprüft.

- **Manual / Online-Hilfe / Literatur:** Ein gut lesbares Handbuch und / oder Online-Hilfe sollte zur Verfügung stehen, darüber hinaus möglichst weitere Informationsmöglichkeiten über Sekundärliteratur / Internet.

3 SAS-Analyst (Version 8 und 9)

<http://support.sas.com/rnd/app/da/analyst.html> (ca. 150 € / Jahr im akadem. Bereich)

SAS-Analyst ist eine maus- und menügesteuerte Oberfläche, eingebettet in das Statistikpaket SAS. Sie ist für Neueinsteiger interessant, da die Oberfläche ähnlich der Windows-Oberfläche gestaltet wurde. SAS-Analyst kann allerdings nicht separat laufen, sondern erfordert die Installation von SAS. Die Oberfläche wird bei Installation von SAS automatisch mitinstalliert und kann über **SOLUTIONS → ANALYSIS → ANALYST** aufgerufen werden. Die Installation mehrerer SAS-Module (Base, STAT, GRAPH, IML, ACCESS, ASSIST) wird benötigt, um den vollen Leistungsumfang dieser Oberfläche nutzen zu können.

SAS-Analyst hat eine schnell zu erlernende Oberfläche, die Funktionalität zum Dateneinlesen und Datenmanagement sind komplett eingebettet. Deskriptive Statistik, Ein- und Zweistichproben tests sowie die wichtigsten Regressionsmodelle sind vorhanden und nutzbar. Von den multivariaten Verfahren ist nur die Hauptkomponentenmethode verfügbar. Die Grafikmöglichkeiten sind nicht immer ausreichend, der mitgelieferte Grafikeditor hat einen geringen Leistungsumfang. Es gibt englische Online-Hilfe, Internet-Unterstützung [1] und deutsche Sekundärliteratur [3].

Der für die Auswertung vom System genutzte SAS-Code kann aufgerufen und abgespeichert werden, so dass das System auch zur Code-Generierung genutzt werden kann. Insgesamt ist die SAS-Analyst-Oberfläche gut einsetzbar mit Schwächen in der Präsentation, die Auswertungen sind sehr differenziert und präzise.

SAS-Analyst wird auch in der SAS-Version 9 mitgeliefert, allerdings nicht mehr weitergepflegt, da nach Firmenpolitik der SAS Enterprise Guide (s.u.) als menügesteuerte Oberfläche etabliert werden soll.

Installation, Rechnerausstattung

Installation problemlos über SAS-Installation (automatische Mitinstallation). Die Rechnerausstattung muss für SAS 8.2 bzw. 9 geeignet sein, sinnvoll: 512 MB RAM, Prozessor > 1 GHz, Festplatte > 10 GB.

Oberfläche / Nutzung

Die Oberfläche ist an MS-Office angepasst, intuitiv erfassbar und leicht zu erlernen.

Datenimport / Datenmanagement

- Excel-Dateien werden mit Variablennamen und Formaten (Datum, num., alphanum.) übernommen
- Management komplett vorhanden, mit geringem Aufwand erlernbar

Statistische Methoden

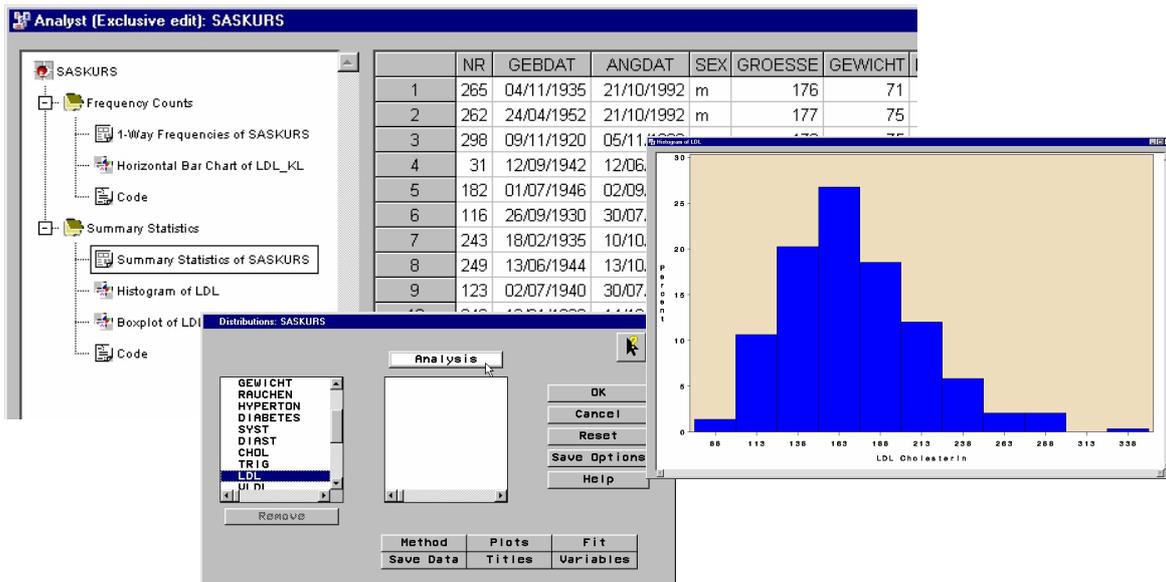
- Basismethoden für deskriptive Statistik komplett vorhanden
- Grafikarten umfangreich (u.a. Box-Plots), aber wenig Lay-out-Möglichkeiten, mäßige Präsentabilität
- sehr umfangreiches Methodenspektrum
- statistische 2- und Mehrstichproben tests vorhanden
- Survival-Analyse und Regressionsanalysen vollständig vorhanden
- Weitere Verfahren: u.a. Power-Analyse, Fallzahlplanung
- Faktoren-, Cluster- und Diskriminanzanalyse nicht vorhanden

Manual / Online-Hilfe / Literatur

- umfangreiche und inhaltlich tiefgehende Online-Hilfe (SAS System Help, u.a. mit stat. Formeln), weitere Hilfe auch in der SAS OnlineDOC
- Sekundärliteratur: Muche, R; Habel; A; Rohlmann, F: Medizinische Statistik mit SAS-Analyst. Springer Verlag, Heidelberg (2000)

Fazit zur Handhabung des Programms

Die Einarbeitung ist leicht. Standardauswertungen sind bei Windows-Kenntnissen schnell durchzuführen. Umfangreiches Methodenspektrum – zusätzliche Möglichkeiten, wenn entsprechende SAS-Module vorhanden sind (keine zusätzlichen Kosten).



4 SAS Enterprise Guide (in SAS Learning Edition)

<http://support.sas.com/rnd/le/index.html> (125 \$ für 3 J. für SAS-Learning Edition)

Der SAS Enterprise Guide ist eine neu entwickelte mausgesteuerte Oberfläche, keine Weiterentwicklung von SAS-Analyst. Sie muss beim Einsatz in der SAS-Umgebung als Extra-Modul zusätzlich angeschafft werden. Entwickelt wurde diese Oberfläche als Thin-Client-Anwendung, d.h. dass das SAS-System und die Daten auf einem beliebigen Version-8/9-SAS-Server installiert sein können. Somit ist diese Anwendung universeller einsetzbar als die SAS-Analyst-Oberfläche, die auf PC-Installationen angewiesen ist.

Ein wichtiger Unterschied zwischen den beiden Oberflächen ist die Umsetzung des Datenmanagements. Sie erfolgt im SAS-Enterprise Guide fast vollständig mit PROC SQL, einer sehr mächtigen SAS-Prozedur für das Datenmanagement. Allerdings ist das Erlernen der Funktionalität unserer Meinung nach nicht unbedingt einfacher geworden, wenn man das Prinzip von SQL nicht kennt. Die beiden von SAS zu beziehenden Einführungsbroschüren helfen hier nicht immer weiter [4,5].

Der Leistungsumfang der zur Zeit gültigen Version 2.1 des Enterprise Guide entspricht im Wesentlichen der von SAS-Analyst, einige multivariate Verfahren wie Cluster-, Diskriminanz- und Faktorenanalyse sind dazugekommen. In dieser Version gibt es drei verschiedene Grafikarten, die den Entwicklungsstand des Grafikpaketes von SAS widerspiegeln.

Seit etwa einem Jahr wird von SAS ein vom Leistungsumfang reduziertes SAS-Paket mit integriertem SAS Enterprise Guide für geringe Kosten (125 \$) als sogenannte SAS Learning Edition [6] angeboten. Zu beziehen ist diese 3 Jahre gültige Edition z.B. über SAS Publishing.

Installation, Rechnerausstattung

Bestellung Online; Microsoft Version 95, 98, NT 4.0, 2000, Me und XP; 390 MB Festplattenbedarf, 64 MB RAM Minimum; Installation der SAS Learning Edition problemlos, kann aber nicht gleichzeitig auf einem Rechner mit schon vorhandener SAS-Installation installiert werden.

Datenimport / Datenmanagement

- Daten aus Excel, Access und im ASCII-Code importierbar
- Datenmanagementfunktionen komplett realisiert
- Datenausgabe in HTML-, RTF-, PDF-Format möglich

Statistische Methoden

- Basismethoden für deskriptive Statistik komplett vorhanden
- Erzeugung von Grafiken ohne Nachbearbeitungsmöglichkeiten; unkonventionelle Boxplots
- statistische 2- und Mehrstichprobentests vorhanden
- Survival- und Regressionsanalyse komplett vorhanden
- Faktoren-, Cluster- und Diskriminanzanalyse vorhanden

Manual / Online-Hilfe / Literatur

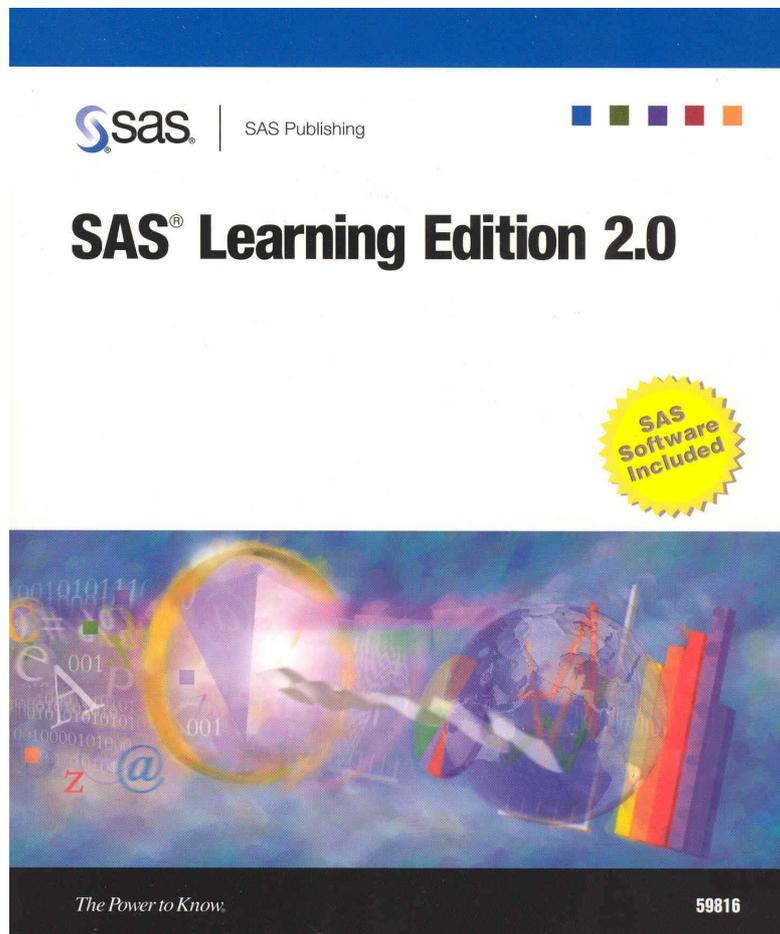
englisches Manual im Lieferumfang; deutschsprachige Kurzfassung auf Anfrage; englischsprachige Online-Hilfe, z. T. aber nicht oder wenig verständlich

Literatur:

- Einstieg in die Enterprise Guide TM Software. SAS Publishing, Cary NC (2001) [5]
- Getting Started with SAS Learning Edition. SAS Publishing, Cary NC (2002) [6]

Fazit zur Handhabung des Programms

Einarbeitung erschwert durch unstimmmige Literatur; dann aber einfache Bedienung des Programms; Standard-Auswertungen schnell erlernbar.



5 Ergebnisse und Diskussion/Schlussfolgerungen

Für einen ersten Einsatz von Statistiksoftware sind beide Produkte gut nutzbar. Sie helfen beim Einstieg in die Welt der Statistiksoftware, da die Einarbeitung das Erlernen einer komplexen Syntax-Sprache nicht voraussetzt. Deshalb sind diese maus- und menügesteuerten Oberflächen für kleine Auswertungen, die nicht unbedingt gut dokumentiert werden müssen, gut geeignet. Andererseits ermöglicht der bereitgestellte SAS-Code dem Anwender einen einfacheren Übergang zur SAS-Syntax, die für große Projekte und methodisch komplexere Auswertungen zu empfehlen ist. Beide Oberflächen stellen nicht den vollen Leistungsumfang des zugrundeliegenden SAS-Paketes zur Verfügung. Die durchgeführte Auswertung kann zu Dokumentationszwe-

cken auf der Syntax-Ebene gespeichert werden. Diese Dokumentation der durchgeführten Auswertungen ist nach den heutigen Maßstäben der Regelwerke zur Durchführung klinischer Studien (z.B. GCP) dringend notwendig. Die gespeicherte Auswertung ist aber nicht von den menügesteuerten Oberflächen wieder abrufbar. Dies erfordert dann die Kenntnis der Syntax-Sprache.

Unser Fazit: Auf lange Sicht ersetzen diese Produkte den Einsatz der Syntax-Sprache im professionellen Umfeld der klinischen Forschung nicht, können aber in der Lernphase den Einstieg in SAS erleichtern.

Literatur

- [1] SAS-Analyst (Version 8) <http://support.sas.com/rnd/app/da/analyst.html> (aufgerufen am 20. 12. 2003)
- [2] SAS Enterprise Guide (in SAS Learning Edition) <http://support.sas.com/le/index.html> (aufgerufen am 24.11.2004)
- [3] Muche, R; Habel, A; Rohlmann, F: Medizinische Statistik mit SAS-Analyst. Springer Verlag, Heidelberg (2000)
- [4] Getting started with Enterprise Guide 2nd Ed. SAS Publishing, No 58809 Cary NC (2002)
- [5] Der Einstieg in die Enterprise Guide Software. SAS Publishing, No 58476 Cary NC (2001)
- [6] SAS Learning Edition 2.0. SAS Publishing, No 59816 Cary NC (2004)