

Änderungen in der mausgesteuerten SAS-Analyst – Oberfläche von Version 6.12 zu Version 8

Rainer Muche, Andreas Habel, Friederike Rohlmann

Abteilung Biometrie und Medizinische Dokumentation, Universität Ulm
89075 Ulm
Tel. 0731 / 50-26903
E-Mail: rainer.muche@medizin.uni-ulm.de

Abstract

Beim SAS-Analyst handelt es sich um ein Zusatzpaket zum Software-Programm SAS. Es wurde ab der Version 6.12 eingeführt und ermöglicht durch eine maus- und menügesteuerte Oberfläche einen leichteren Zugriff auf Verfahren der statistischen Auswertung. Zwischen den Analyst-Versionen von SAS 6.12 und 8 gibt es einige wesentliche Änderungen und Erweiterungen.

Die wichtigsten Änderungen betreffen das Einlesen von Daten und das Datenmanagement. Wichtige Erweiterungen betreffen Ausgabe, statistische Methoden und Grafikmöglichkeiten. Insbesondere sind nun Informationsmöglichkeiten zum Analyst-Programm vorhanden, die in der alten Version noch völlig fehlten (Handbuch, Online-Hilfe etc.).

1. Einleitung

Im vergangenen Jahr haben wir auf der KSFE von ersten Erfahrungen mit SAS-Kursen auf der Analyst-Oberfläche berichtet [1]. Zum damaligen Zeitpunkt war noch die SAS-Version 6.12 aktuell. Die Erarbeitung eines kursbegleitenden Skripts [2] war sehr aufwendig und hat gezeigt, dass die Analyst-Version 6.12 noch viele zum Teil erhebliche Mängel aufweist. Einige Zeit später wurden unsere Rechner auf die SAS-Version 8 umgestellt.

Bei der Umsetzung des Kurses, insbesondere der Kursmaterialien haben wir einige wesentliche Unterschiede zwischen den beiden Analyst-Versionen ausgemacht. Insgesamt hat die SAS-Analyst-Oberfläche der Version 8 von vielen Änderungen profitiert, die auf die Verbesserungen des zugrundeliegenden SAS-Systems zurückzuführen sind. Aber es wurden auch etliche Änderungen und Erweiterungen vorgenommen, die nur der Verbesserung des Analyst dienen.

2. Kurzvorstellung der SAS-Analyst-Oberfläche

In der Version 8 des SAS-Systems gehört die Analyst-Oberfläche zum Standard-Lieferumfang. Sie steht unter **SOLUTIONS** → **ANALYSIS** → **ANALYST** zur Verfügung.

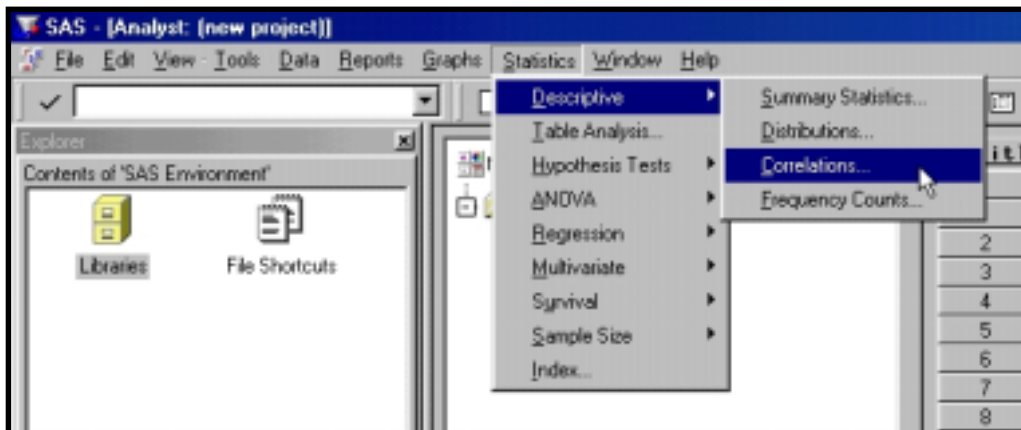
Einige Spezifika des SAS-Analyst sind:

- maus- und menügesteuerte Oberfläche
- integrierte Dateneingabe- und Datenmanagementmöglichkeiten
- auf das SAS-System aufgesetzt, nicht unabhängig lauffähig
- es gibt einige Statistikkösungen, die im "normalen" SAS-System nicht vorhanden sind (z. B. Fallzahlberechnungen, Poweranalysen)

In Abbildung 1 ist die Startseite des Analyst dargestellt. Am oberen Fensterrand befindet sich die in Anlehnung an MS-Windows aufgebaute Menüleiste, in der Pull-Down-Menüs mit der Maus angewählt werden können. Im Teilfenster am linken Rand werden die Auswertungsergebnisse als Symbole in der Form eines Explorers oder Auswertungsbaumes angeordnet und können während der Sitzung wieder aufgerufen, gespeichert und ausgedruckt werden. Zu jeder Auswertung gibt es im Explorer auch den Eintrag **Code**. Er enthält den jeweiligen SAS-Syntaxcode, der in diesem Ordner durchgeführten

Auswertung zugrunde liegt. Auch dieser Code ist speicherbar und kann zur üblichen SAS-Programmierung verwendet werden. Im rechten Teil des Analyst-Fensters gibt es ein Tabellenblatt, in das Daten eingegeben oder eingelesen werden können. Die manuelle Eingabe von Daten ist nicht sehr glücklich gelöst und hat viele Fehler. Dazu gibt es einige Anmerkungen im Abschnitt "Daten einlesen"

Abbildung 1: Die SAS-Analyst – Oberfläche



Die Auswertungen und sonstige Analyst-Funktionen werden durch Pull-Down-Menüs ausgewählt (Abbildung 2). Es öffnen sich Dialogfenster, in denen die erforderlichen Angaben zu den Auswertungen ausgewählt oder eingegeben werden können. Weitere Optionen, Einstellungen und Anweisungen lassen sich über Dialogfenster der nächste Ebene vornehmen und sind über entsprechende Schaltflächen erreichbar (Abbildung 3).

Abbildung 2: Aufruf der Anwendungen durch Pull-Down-Menüs

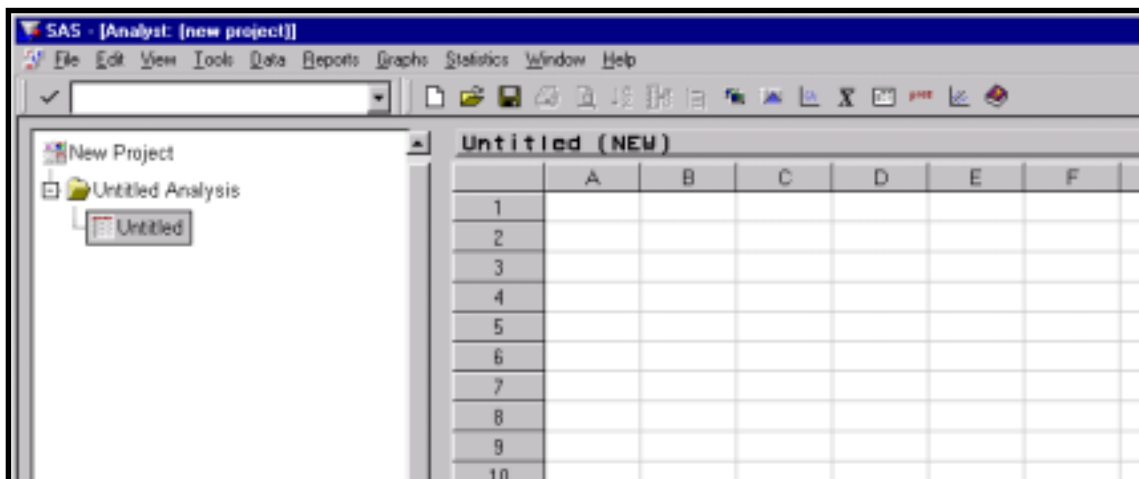
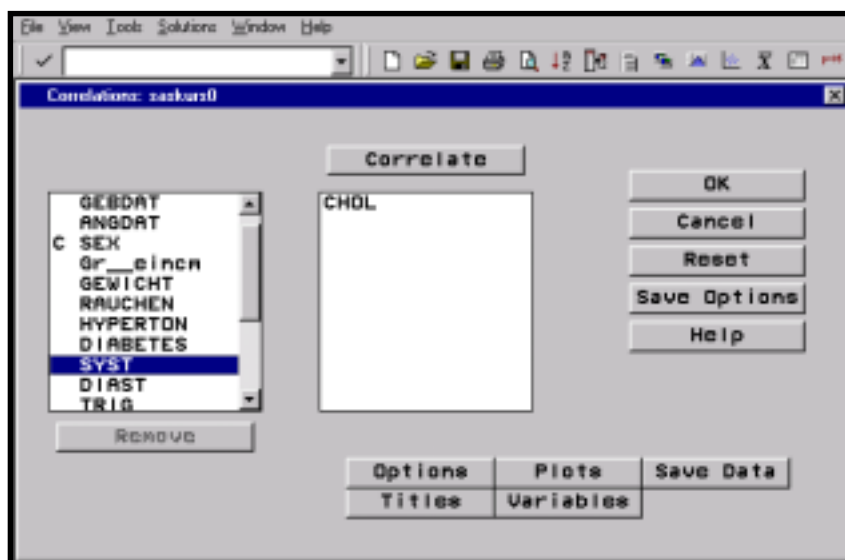


Abbildung 3: Aufruf der Auswertungen durch Eingabe und Auswahl im Dialogfenster

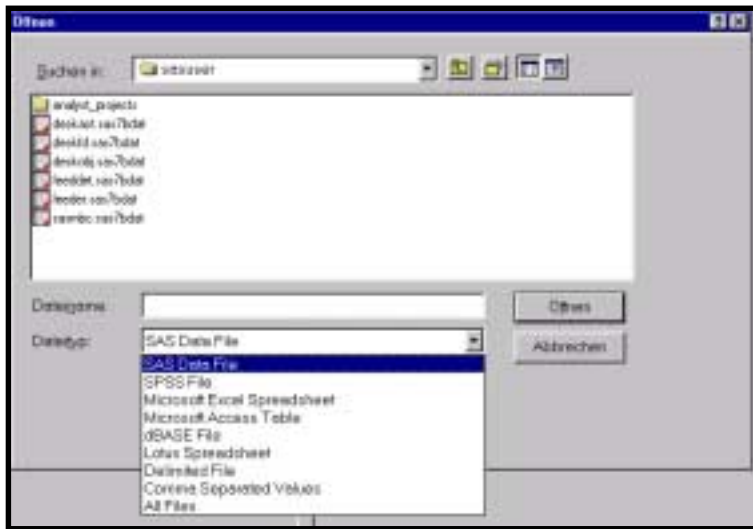
3. Veränderungen zwischen Version 6.12 und Version 8

Im Folgenden werden die wesentlichen Änderungen und Erweiterungen zwischen den beiden Versionen erläutert. Gegliedert sind die Gegenüberstellungen in die Bereiche "Daten einlesen", "Datenmanagement", "Daten- und Ergebnisausgabe", "Statistik-" und "Grafikmöglichkeiten".

3.1 Änderungen im Bereich "Daten einlesen"

Einlesen externer Daten

Der Standard für das Einlesen externer Daten in der Version 6.12 war der Import-Wizzard. Dieser führte den Nutzer durch eigene Dialogfenster und ist deshalb für ungeübte Nutzer etwas gewöhnungsbedürftig. Der Import-Wizzard wurde für das SAS-System in der Version 8 beibehalten. Allerdings steht er jetzt nicht mehr in der Analyst-Oberfläche zur Verfügung. Externe Daten-Files können - nach MS-Windows-Standard - über **FILE** → **OPEN** geöffnet werden. Im entsprechenden Dialogfenster ist ein Pull-Down-Menü eingerichtet, über das der passende Dateityp ausgewählt werden kann (Abbildung 4). Mit diesem Vorgehen ist man nun kompatibel zu sonstigen MS-Windows-Programmen. Gegenüber dem Import-Wizzard der Version 6.12 sind die einzulesenden Dateitypen um SPSS-Files und MS-Access-Tabellen erweitert worden. Auch das Einlesen von MS-Excel-Dateien an sich hat sich verbessert. Das Einlesen von Variablen versagt nun nicht mehr, wenn in Variablennamen der MS-Excel Tabelle Umlaute oder Sonderzeichen verwendet wurden. Allerdings werden diese nun in SAS bzw. im Analyst durch Unterstriche ersetzt. Leider gilt dies nur für MS-Excel-Dateien ab der Version 5.

Abbildung 4: Daten einlesen über FILE → OPEN

Manuelle Dateneingabe

Die Dateneingabe über das Tabellenblatt war in der Version 6.12 nicht zu empfehlen [2]. So sind z.B. direkte Werte-Eingaben auf 8 Zeichen beschränkt, die Formatierung der Variablen geschieht automatisch nach der ersten Werteingabe ohne Korrekturmöglichkeit und die Nutzung zur Dateneingabe wird dadurch erschwert, dass nach Eingabe durch die Enter-Taste der Cursor nicht automatisch ins nächste Feld springt. Diese negative Einschätzung zur Nutzung des Tabellenblattes hat sich auch in der Version 8 nicht wesentlich geändert. Es gibt Verbesserungen, die darauf zurückzuführen sind, dass sich das SAS-System insgesamt verändert hat, z.B. was die Länge der Variablennamen angeht. Ausserdem gibt es nun die Möglichkeit, Variableneigenschaften nachträglich zu verändern. Allerdings hat sich dabei wohl ein Programmierfehler eingeschlichen, denn egal, welches Format der Variablen zugeordnet wird, lassen sich nicht mehr als 16 Zeichen in die Tabellenfelder eingeben. Damit ist das Tabellenblatt als Dateneingabe-Tool nicht mehr allgemein geeignet.

Desweiteren fällt als Kuriosität auf, dass der Variablenname im Tabellenblatt bis zu 12 Zeichen lang sein kann (warum eigentlich diese Beschränkung?), wenn vorher keine Werte in die entsprechende Spalte eingetragen wurden. Gibt man zuerst Daten ein und vergibt den Variablennamen im nachhinein, kann er nur noch maximal 8 Zeichen lang sein.

Datumseingaben können nicht direkt in das Tabellenblatt vorgenommen werden. Sie werden als alphanumerische Variable interpretiert. Zur Eingabe muss man erst eine numerische Variable neu erzeugen und dieser ein Datumsformat zuweisen.

Als Konsequenz empfehlen wir weiterhin, umfangreichere und komplexere Datenerfassung nicht auf der Analyst-Oberfläche vorzunehmen, sondern dazu auf andere Software (z.B. MS-Excel) zurückzugreifen.

3.2 Änderungen im Bereich "Datenmanagement"

In den Datenmanagement-Möglichkeiten gibt es wesentliche Verbesserungen. Diese waren in der Version 6.12 sehr stiefmütterlich behandelt worden. So lassen sich z.B. die Variableneigenschaften über **DATA → Column Properties** auch nach oder während der Dateneingabe verändern.

Wesentlich verbessert haben sich die Werkzeuge zur Umkodierung oder Klassierung von Variablen. Zum Beispiel musste man zur Klassierung einer Variablen in Version 6.12 einen relativ komplizierten Algorithmus in ein Fenster eintippen (s. Abbildung 5). Das bedeutete hohe Fehleranfälligkeit, insbesondere für ungeübte Nutzer. Jetzt gibt es für diese Aufgabe ein selbst erklärendes, einfach zu bedienendes Dialogfenster (Abbildung 5).

Abbildung 5: Klassierung einer Variable in V 6.12 und V 8

V6: Algorithmus zur Klassierung

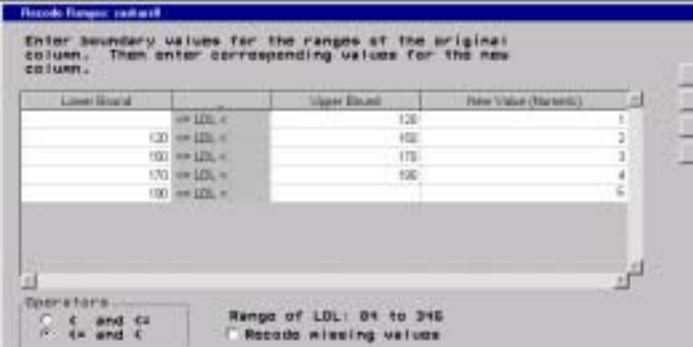
(LDL<120)*1+

((LDL>=120) AND (LDL<150))*2+

((LDL>=150) AND (LDL<170))*3+

((LDL>=170) AND (LDL<190))*4+

(LDL>=190)*5



In der Version 6.12 fehlte (unverständlicherweise) die Möglichkeit zum Zusammenführen von Dateien (Set und Merge). Beide Tools sind jetzt implementiert worden. Auch konnten Daten nur in aufsteigender Richtung sortiert werden. Version 8 ermöglicht nun auch die absteigende Sortierung.

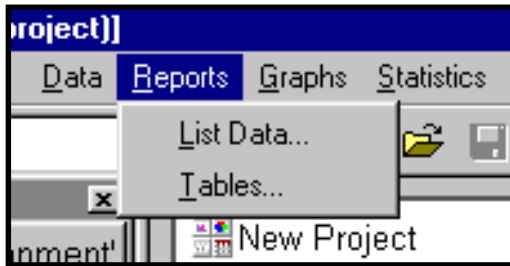
Das Abspeichern und Drucken von Outputs und Grafiken wurde dem MS-Windows-Standard angepasst und kann jetzt ohne Probleme mit **FILE → SAVE** bzw. **FILE → PRINT** vorgenommen werden.

Bei der Definition der Überschriften (Titles) konnte man in der alten Version anwählen, ob Datum und/oder Seitenzahl automatisch mit ausgegeben werden sollen. In der neuen Version hat man nun auch die Möglichkeit, sich die Datenfilterfunktion mitausgeben zu lassen, sofern eine solche zur Einschränkung des Datensatzes für eine Auswertung zur Anwendung kam. Diese Dokumentationsmöglichkeit ist sehr nützlich für die spätere Interpretierbarkeit des Outputs.

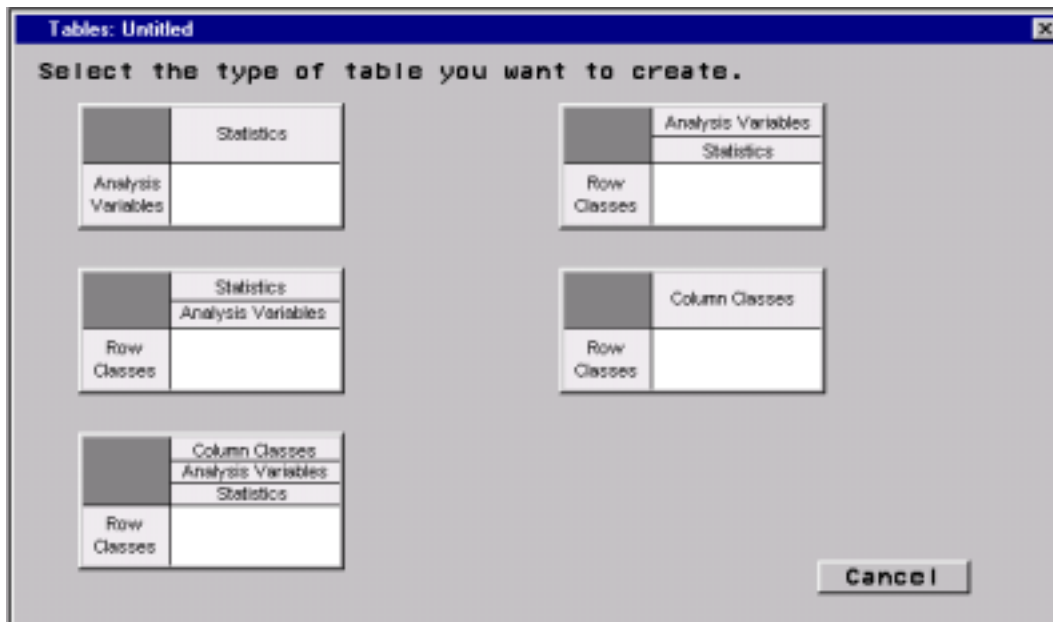
Schließlich wurden auch noch die Dokumentationsmöglichkeiten im Auswertungsbaum verbessert. Die einzelnen Zweige können neu benannt werden und so statt der Standardbenennungen sprechende Namen erhalten. Außerdem lassen sich Auswertungen zu sogenannten Projekten zusammenführen. Auch die jeweils bearbeiteten Datensätze werden nun im Auswertungsbaum dokumentiert, so dass die Auswertungen insgesamt gut nachvollziehbar werden.

3.3 Änderungen im Bereich “Daten- und Ergebnisausgabe”

In der Version 8 gibt es einen neuen Menüpunkt **REPORTS** (Abbildung 6), unter dem man ein PROC PRINT und 5 verschiedene PROC TABULATE-Tabellen (Abbildung 7) auswählen kann. Diese Möglichkeiten sind neu. So einfache, aber wichtige Tools gab es in der Version 6.12 nicht.

Abbildung 6: Menüpunkt REPORTS

Folgende 5 Tabellenversionen aus PROC TABULATE, die die wohl notwendigsten Varianten abdecken, sind implementiert (Abbildung 7).

Abbildung 7: Tabellenversionen in REPORTS → TABLES

Teilweise ist das ODS integriert, zumindest gibt es die Möglichkeit, die Ergebnisse im HTML-Format zu erhalten.

4. Änderungen im Bereich "Statistik"

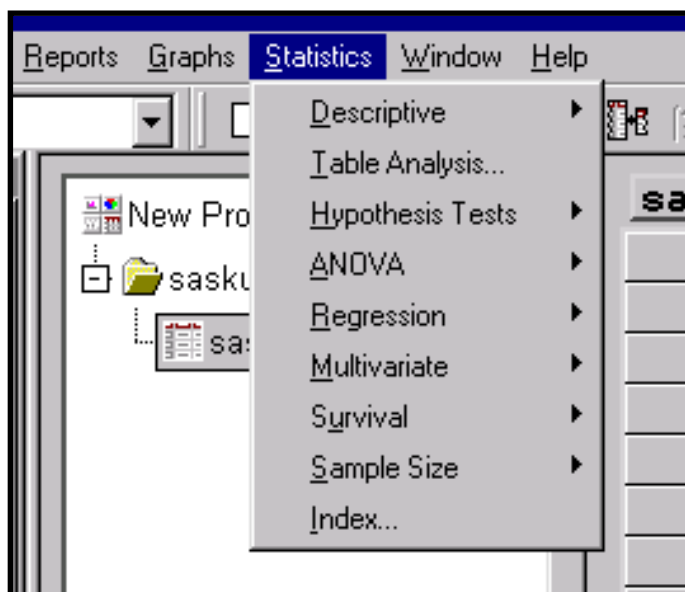
Die folgende Abbildung 8 gibt einen groben Eindruck, welche Auswertungsmöglichkeiten im SAS-Analyst der Version 8 enthalten sind. Im Vergleich zur Version 6.12 neu hinzugekommen sind die Menüs **SURVIVAL** und **MULTIVARIATE** sowie einige Menüpunkte unter **ANOVA**.

Unter **SURVIVAL** sind die Menüs für den Aufruf von PROC LIFETEST und PROC PHREG zu finden. Für die Anwendung des Analyst in der Medizinischen Statistik war diese Erweiterung der Auswertungsmöglichkeiten dringend nötig und ist sehr zu begrüßen.

Erweiterungen in **MULTIVARIATE** geben die Möglichkeit, PROC PRINCOMP und PROC CANCECORR aufzurufen. **ANOVA** ist im Wesentlichen um die Möglichkeiten von PROC MIXED erweitert worden. Diese Auswertungsmöglichkeiten sind schon sehr speziell und nicht die typischen Methoden, die von Einsteigern bzw. unerfahrenen Nutzern gebraucht werden. Deshalb stellt sich hier die Frage, warum gerade diese Prozeduren in den Analyst

implementiert wurden. Wir hätten hier eher Standardmethoden wie PROC CLUSTER, PROC DISCRIM, PROC FACTOR, PROC CATMOD erwartet.

Abbildung 8: Auswertungsmöglichkeiten in SAS-Analyst V8



Einige Menüs sind in der Version 8 um wesentliche Optionen erweitert worden. Aus unverständlichen Gründen fehlten in der alten Version z.B. die Möglichkeiten, exakte Testergebnisse in PROC FREQ (exakter Fisher-Test) und in PROC NPAR1WAY (exakter unverbundener Wilcoxon- bzw. Kruskal-Wallis-Test) zu erhalten. Diese sind nun implementiert worden.

5. Änderungen im Bereich “Grafik”

Im Bereich Grafik wurden Elemente aus PROC GCHART (Bar Chart, Pie Chart) eingebaut, die in der Version 6.12 nicht auszuwählen waren. Den Box-Plot (PROC BOXPLOT) konnte man interessanterweise schon in Version 6.12 wählen, obwohl die zugehörige Prozedur erst für die Version 8 angekündigt wurde.

Die Voreinstellungsmöglichkeiten für Grafiken (Settings, z.B. Farbeinstellungen) sind aus dem Grafikenmenü in das Menü **TOOLS** verschoben worden. Leider gibt es keine Veränderungen bezüglich des Grafik-Editors. Wie im allgemeinen SAS-System wird der alte, schlechte Grafik-Editor immer noch eingesetzt.

Informationsmöglichkeiten / Handbücher

Allgemein wichtig bei der Nutzung von Software, also auch hier bei der Nutzung der Analyst-Oberfläche in SAS ist es, genügend gute Informations- und Hilfemöglichkeiten zu haben. In der Version 6.12 gab es bis auf das von uns selbst geschriebene Skript [2] keine Hilfen für die Analyst-Oberfläche. Dies hat sich in der neuen Version wesentlich verbessert. Folgende Informationsmöglichkeiten bietet SAS jetzt an:

- **Online-Hilfe zur Analyst-Oberfläche :** (Aufruf: HELP → SAS System HELP → Using SAS Solutions and Tools → Using the ANALYST application for statistical tasks). Darüberhinaus kann man über einen HELP-Button in den Auswertungsdialogfenstern direkte Hilfe zum aktuellen Problem erhalten.

- **SAS OnlineDoc – CD** : In der dem SAS-System nun beigelegten CD, die alle SAS-Handbücher beinhaltet, gibt es jetzt auch ein entsprechendes Kapitel zur Analyst-Oberfläche. Direkt zu finden ist es unter: SASDOC/SASHTML/Analyst/Index.htm.
- **SAS-Internetseite** <http://www.sas.com/rnd/app/da/analyst.html> : Desweiteren bietet SAS auf den Internet-Seiten einige Informationen und Beschreibungen zur Analyst-Oberfläche.

Schließlich gibt es ein von uns geschriebenes Buch [3] zur Nutzung der Analyst-Oberfläche mit Anwendungen in der Medizinstatistik. Darin wird die Nutzung anhand eines medizinischen Beispiels ausführlich dargestellt, allerdings eingeschränkt auf den Lehrinhalt der Veranstaltung Biomathematik im Medizinstudium.

6. Zusammenfassung / Resümee

Zusammenfassend sehen wir folgende Vorteile in der Nutzung der Analyst-Oberfläche für Studenten bzw. SAS-Einsteiger [4] nach der Überarbeitung in der Version 8:

- Die Oberfläche ist maus- und menügesteuert, so dass das aufwendige Erlernen der SAS-Syntaxsprache wegfällt.
- Die Oberfläche ist Windows-ähnlich und erleichtert dadurch dem weniger erfahrenen PC-Nutzer den Einstieg in das Programm.
- Daten aus Standard-PC-Programmen wie MS-Excel können einfach eingelesen werden.
- Es gibt jetzt ausreichend Hilfe- und Nachschlagemöglichkeiten.
- Die Click-Befehle können als SAS-Syntax ausgegeben werden, so dass die Analyst-Oberfläche als Einstiegsmöglichkeit in die SAS-Programmierung genutzt werden kann.

Wo es Vorteile gibt, sind auch fast immer Nachteile zu anzumerken:

- Der Windows-ähnliche Aufbau der Oberfläche verführt dazu, auch das Tabellenblatt so zu verwenden, wie man es von MS-Excel gewöhnt ist. Da man in der Regel schon bei den ersten Schritten auf Fehler und Probleme stößt, ist der Anwender sofort abgeschreckt, und dieses Urteil wird auf das ganze Produkt übertragen. Daher gilt immer noch unser Empfehlung, die Datenerfassung mit anderen Programmen vorzunehmen.
- In der Analyst-Oberfläche ist nach wie vor der alte Grafikeditor des SAS-Systems integriert, der auf jeden Fall unzulänglich ist.

Insgesamt kann man feststellen, dass die Analyst-Oberfläche wesentlich besser geworden ist. Speziell die Verbesserungen im Datenmanagement führen dazu, dass man jetzt mit dem System vernünftig auswerten kann. Gerne würden wir sehen, dass alle Möglichkeiten des SAS-Systems unter einer solchen maus- und menügeführten Oberfläche nutzbar wären.

7. Literatur

- [1] R. Muche, A. Habel: SAS-Analyst – Kurse für Mediziner: erste Erfahrungen Proceedings der 4. KSFE, Giessen (2000), S. 139-144
- [2] R. Muche, A. Habel: SAS-Analyst. Einführung in die Benutzeroberfläche Analyst des SAS-Systems V6.12 mit Anwendungen in der Medizinischen Statistik. Kurs-Skript, unveröffentlicht, vom Autor zu erhalten (1999)
- [3] R. Muche, A. Habel, F. Rohlmann: Medizinische Statistik mit SAS-Analyst Springer Verlag, Heidelberg, (2000)
- [4] R. Muche, G. Büchele, A. Imhof, A. Habel: Erfahrungen mit SAS-Kursen für unterschiedliche Nutzergruppen. Proceedings der 3. KSFE, Heidelberg, (1999), S. 177-182