

## **SAS Software in der Lehre des Fachbereichs Statistik der Universität Dortmund - Ein Erfahrungsbericht –**

Andreas Christmann

Fachbereich Statistik

Universität Dortmund

Dortmund

christmann@statistik.uni-dortmund.de

Magdalena Thöne

Fachbereich Statistik

Universität Dortmund

Dortmund

thoene@statistik.uni-dortmund.de

Guido Knapp

Fachbereich Statistik

Universität Dortmund

Dortmund

knapp@statistik.uni-dortmund.de

Joachim Hartung

Fachbereich Statistik

Universität Dortmund

Dortmund

hartung@statistik.uni-dortmund.de

### **Zusammenfassung**

Am Fachbereich Statistik der Universität Dortmund können Studierende den Diplom-Studiengang Statistik oder den Bachelor-/Masterstudiengang Datenanalyse und Datenmanagement / Datenwissenschaft absolvieren. Die Ausbildung mit der SAS Software ist im Pflicht- bzw. Wahlpflichtbereich beider Studiengänge integriert. Anwendung findet SAS zunächst in den Pflichtveranstaltungen "Fallstudien", in denen unterschiedliche Datensätze ausgewertet und entsprechende Berichte verfasst werden.

Als erster Einstieg wird regelmäßig ein Einführungskurs in SAS als sechstägiger Intensivkurs (Blockveranstaltung) durchgeführt.

Die Vorlesung "Datenanalyse mit SAS" ist eine einsemestrige Veranstaltung mit sechs Semesterwochenstunden. Der DATA Step wird ausführlich behandelt, anschließend wird der Einsatz ausgewählter SAS Prozeduren dargestellt. Eine interdisziplinäre Veranstaltung der Fachbereiche Statistik und Informatik zum Thema "Wissensentdeckung in Datenbanken / Data Mining" wird ebenfalls regelmäßig angeboten.

Der SAS Enterprise Miner wurde bisher in zwei Lehrveranstaltungen intensiv eingesetzt. In "Fallstudien II" wurden Klassifikations- und Prognoseverfahren, die der SAS Enterprise Miner zur Verfügung stellt, in einer großen epidemiologischen Studie verwendet. Zusätzlich wurden in einem Seminar die Grundlagen der Klassifikations- und Prognoseverfahren erarbeitet und deren adäquate Umsetzung im SAS Enterprise Miner anhand von Fallbeispielen illustriert.

**Schlüsselworte:** SAS in der Lehre, Datenanalyse mit SAS, Enterprise Miner.

# **1 SAS in Lehre und Studium am Fachbereich Statistik**

## **1.1 Das Studium am Fachbereich Statistik**

Das Studienangebot des Fachbereichs Statistik der Universität Dortmund besteht derzeit in der Hauptsache aus den zwei Studienrichtungen Statistik und Datenanalyse:

- Statistik – Diplom-Studiengang (9 Semester), sowie
- Datenanalyse und Datenmanagement – Bachelor-Studiengang, B.Sc. (6 Semester),
- Datenwissenschaft – Master-Studiengang, M.Sc. (4 Semester)

Daneben findet die Ausbildung im Bereich Statistik für diverse andere Studiengänge der Universität statt (Nebenfach, Service-Veranstaltungen etc.). Am Fachbereich Statistik sind momentan rund 400 Studierende in den drei Studiengängen (Diplom, Bachelor, Master) eingeschrieben.

## **1.2 Einsatz statistischer Programmpakete (insbes. SAS) im Studium: z.B. Fallstudien**

Die Studierenden erlernen im Laufe ihres Studiums verschiedene statistische Programm-Pakete, wobei die SAS Software neben R bzw. S-Plus zum "Standard" gehört. Die SAS Software wird von den Studierenden vor allem eingesetzt in den Praktika (Fallstudien I und II), in denen die Studierenden in verschiedenen Projekten lernen, mit realen Datensätzen umzugehen, diese unter verschiedenen Aspekten auszuwerten und einen Studienbericht zu erstellen. In der Veranstaltung "Fallstudien I" steht die Vermittlung und der Einsatz eines breiten Methodenrepertoires im Vordergrund: in etwa sieben kleineren Projekten werden Fragestellungen aus verschiedenen Bereichen bearbeitet; zu jedem dieser Projekte gehört die Analyse des zugehörigen Datensatzes – jeweils unter bestimmten Gesichtspunkten – sowie die Erstellung eines Berichtes. Das eingesetzte Methodenspektrum geht von der deskriptiven Analyse über Zwei- und Mehr-Gruppen-Vergleiche zu (unter anderem) Kontingenztafelanalysen, Korrelations- und Regressionsanalysen inklusive Variablenselektionsverfahren und logistischer Regression.

Während in der Lehrveranstaltung "Fallstudien I" die Methoden eher vorgegeben werden, sind die Aufgaben in "Fallstudien II" wesentlich komplexer, die Projekte entsprechend größer und länger. In der Regel beschäftigen sich die Studierenden hier während eines Semesters mit zwei Projekten, die wiederum mit einem Bericht und mündlichen Präsentationen abgeschlossen werden. In dieser Veranstaltung wird von Seiten der Lehrenden eher die Fragestellung vorgegeben, die Lösungswege sind dann

von den Studierendenteams gemeinsam zu entwickeln (Methodenauswahl, Adaption an das vorliegende Problem und die vorhandenen Daten etc.).

In den Veranstaltungen "Fallstudien" müssen die Studierenden in der Lage sein, statistische Programmpakete einzusetzen bzw. sie üben sich im Verlaufe dieser Veranstaltungen im Einsatz der Programme. Zwar ist es den Studierenden freigestellt, welche Programme sie verwenden, jedoch setzt ein großer Teil der Studierenden hier die SAS Software ein, da sie die gesamte Breite des verwendeten Methodenspektrums gut unterstützt und da die Studierenden auf diese Weise einen guten Einblick in spätere berufliche Tätigkeitsfelder erhalten und dies unter "Realbedingungen" ausprobieren möchten.

Neben dem Einsatz der SAS Software in den Fallstudien und in anderen Veranstaltungen (z.B. für die Bearbeitung von Übungsaufgaben und kleineren Projekten) spielt die SAS Software auch eine Rolle als Gegenstand der Ausbildung. Unser Erfahrungsbericht spiegelt im Wesentlichen diesen Aspekt des Einsatzes von SAS in der Lehre wieder, da insbesondere in den Fallstudien die Studierenden frei sind im Einsatz ihrer Mittel (insbesondere beim Software-Einsatz), vieles – und so auch die Verwendung der Software - steht hier unter dem Motto "learning by doing". Allerdings erhalten die Studierenden hier in allen Aspekten auch Unterstützung und Beratung durch die Lehrenden. Eine formalisierte Ausbildung in statistischer Software findet jedoch in diesen Veranstaltungen nicht statt; hierzu werden im Wesentlichen die im folgenden beschriebenen Veranstaltungen angeboten.

## **2 Ausbildung mit der SAS Software am Fachbereich Statistik**

### **2.1 Einführungskurs in SAS**

Als Einstieg in die SAS Software wird im allgemeinen ein "SAS-Einführungskurs" gewählt, der als Kompaktkurs vor Beginn des Semesters angeboten wird. Dieser Intensivkurs dauert sechs Tage; neben einer vorlesungsartigen Einführung in die verschiedenen Elemente von SAS mit vielen praktischen Vorführungen steht jeweils der Nachmittag für Übungen zur Verfügung, in denen die Studierenden anhand kleinerer Übungsaufgaben die vorgestellten Schritte und Prozeduren selbst einsetzen und ausprobieren sollen.

Seit einiger Zeit wird dieser Kurs jeweils vor Beginn des Wintersemesters gehalten. Aufgrund steigender Nachfrage wird der Kurs nunmehr auch zwischen dem Wintersemester und dem Sommersemester angeboten; die Teilnehmerzahlen liegen i.A.

zwischen 20 und 50 Studierenden. Die meisten Teilnehmenden dieser Kurse sind Studierende des Fachbereichs Statistik, in der Regel am Ende des Grundstudiums oder zu Beginn des Hauptstudiums; der Kurs ist jedoch auch offen für Studierende anderer Fachbereiche.

Inhalte des Kurses sind Grundlagen der SAS Software.

- Es wird das Einlesen verschiedenster Daten behandelt und besonders dem DATA-Step wird viel Platz eingeräumt.
- Der Datenzugang wird ausführlich behandelt (Einleseformate), anschließend wird auf einfache Datenschnitte eingegangen (Variablen erzeugen, Gruppieren von Daten, if-then-else-Abfragen), erste Prozeduren zur Datenverwaltung (sortieren, drucken, formatieren, aber auch das Zusammenfügen verschiedener Dateien) sowie zur Zusammenfassung von Daten (Proc Means, Freq, Tabulate, Plot) werden erläutert.
- Auch auf verschiedene Output-Möglichkeiten wird eingegangen (z.B. ODS sowie diverse Ausgabeformate und Druckoptionen); ebenso werden interaktive Module kurz präsentiert (SAS Insight).
- Grundlegende statistische Prozeduren werden kurz vorgestellt, sie sind jedoch nicht Hauptgegenstand des Kurses (Proc Univariate, Means, Freq, Corr, Reg, ANOVA); auch die Grafik-Prozeduren werden kurz angerissen (Proc Gchart, Gplot).
- Um den Studierenden weitere Möglichkeiten von SAS zu verdeutlichen, wird das Makro-Konzept kurz vorgestellt. Besonders auf Wunsch der Studierenden wird inzwischen auch in die Matrix-Sprache IML eingeführt.

Auf diese Weise wird versucht, in einer kompakten intensiven Veranstaltung den Studierenden einen ersten Einblick und einen Überblick über die SAS Software zu geben, so dass sie, nachdem sie die Grundstrukturen kennen gelernt haben, bei späteren Veranstaltungen (wie z.B. bei den Fallstudien) in der Lage sind, die SAS Software einzusetzen und sich weiter mit den Möglichkeiten der Software vertraut zu machen.

Um das Selbststudium weiter zu unterstützen und das Verfolgen der im Kompaktkurs in relativ kurzer Zeit vorgestellten Möglichkeiten, Prozeduren und Befehle von SAS zu erleichtern, wurde für den Einführungskurs ein Skript erstellt, das auch später noch zum Nachschlagen verwendet werden kann. Viele Beispielprogramme helfen, sich weiter mit der Software auseinanderzusetzen; daneben wird auch die SAS-Hilfe vorgestellt sowie einige Literatur angegeben, die den Studierenden beim weiteren Arbeiten mit SAS hilfreich sein kann (vgl. Literaturverzeichnis).

Eine regelmäßige Evaluation am Ende des Kurses zeigt, dass die Studierenden mit diesem Angebot recht zufrieden sind; andererseits gibt es hier auch immer wieder

Anregungen für Verbesserungen und Veränderungen für nachfolgende Kurse, die sich aber bisher (bis auf die erwähnte Aufnahme des Themas IML in den Kurs) auf eher kleinere Punkte bezogen haben. Insbesondere der letzte, sehr große Kurs hat wieder deutlich gezeigt, dass es bei einer Software-Einführung wichtig ist, während der Übungen ausreichend Hilfestellung geben zu können (Betreuungsverhältnis). Die vorliegenden Gruppengrößen im Zusammenhang mit den vorhandenen Rechnerpools lassen es leider nicht zu, die Einführung in SAS direkt am PC anzubieten; insgesamt hat sich aber auch das Vorgehen mit Vorlesung und Übungen bewährt.

## **2.2 SAS Enterprise Miner**

Der SAS Enterprise Miner in der Version 4.0 bzw. Version 4.1 stand dem Lehrstuhl von Prof. Hartung für den Einsatz in der Lehre sechs Semester lang zur Verfügung, und zwar vom Sommersemester 2001 bis zum Wintersemester 2003 / 2004. Der Einsatz dieser Software in Lehrveranstaltungen hing im Wesentlichen von den Lehraufgaben der Mitarbeiter des Lehrstuhls in den einzelnen Semestern ab.

Um jedoch einer größtmöglichen Anzahl von interessierten Studierenden einen Einblick in die Benutzung und Mächtigkeit der Software zu geben, wurden außerhalb der Vorlesungszeiträume regelmäßig Einführungsveranstaltungen als Kompaktkurse angeboten. Dieses Angebot wurde in der Regel von Studierenden im Hauptstudium des Diplomstudiengangs Statistik und in der letzten Zeit auch von Studierenden im Master-Studiengang Datenwissenschaften wahr genommen.

In diesen Kompaktkursen wurde die grundlegende SEMMA (Selecting, Exploring, Modifying, Modeling, Assessing) Philosophie des Data Mining mit dem SAS Enterprise Miner vorgestellt. Es wurde eine Einführung bzw. ein Einblick in die wichtigsten Knoten gegeben. Diese Knoten sind: Input Data Source, Insight, Distribution Explorer, Multiplot, Data Partition, Filter Outliers, Replacement, Transform Variables, Regression, Tree, Neural Network, Assessment, Clustering, Association.

In Lehrveranstaltungen des Fachbereichs Statistik ist der SAS Enterprise Miner in zwei Hauptstudiumsveranstaltungen des Diplomstudiengangs Statistik intensiv eingesetzt worden. Dabei handelte es sich um die Veranstaltung "Fallstudien II" und ein Haupt-Seminar.

In der Veranstaltung "Fallstudien II", in der Studierende im Hauptstudium des Diplomstudiengangs Statistik in der Regel Probleme aus der Praxis lösen sollen, stand ein Datensatz einer epidemiologischen Studie zu Oberbauchbeschwerden zur Verfügung. In dieser Studie wurden das Management und der Verlauf der funktionellen Dyspepsie in Arztpraxen dokumentiert. Bei der funktionellen Dyspepsie liegen im Gegensatz zur organischen Dyspepsie (z. B. Ulkusleiden) keine Läsionen vor. In der

hausärztlichen Praxis sind etwa 80% der Dyspepsie-Fälle funktioneller Art. Der Leidensdruck der Patienten mit funktioneller Dyspepsie ist größer als bei Patienten mit organischer Dyspepsie, da erstere häufiger ihre täglichen Aktivitäten z.B. durch Unwohlsein unterbrechen müssen.

Der Datensatz umfasste 2736 Patienten. Für jeden Patienten lag eine Beurteilung der Symptome sowie eine Einschätzung der Lebensqualität vor. Darüber hinaus wurden prognostische Faktoren wie Alter, Geschlecht, Lebensgewohnheiten etc. erhoben. Von den Ärzten wurden die Patienten einer der drei folgenden initialen Therapieformen zugewiesen: Probatorische medikamentöse Therapie, Endoskopie mit anschließender medikamentöser Therapie oder Wait-and-see Strategie.

Die Patienten wurden über ein Jahr beobachtet und dann als Responder/Non-Responder bzgl. der Reduktion der Beschwerden und als Responder/Non-Responder bzgl. der Verbesserung der Lebensqualität klassifiziert.

In der Lehrveranstaltung "Fallstudien II" sollten die Studierenden geeignete Prognosemodelle für die Responder ermitteln. Dazu sollten verschiedene Prognosemodelle mit den Enterprise Miner Knoten Logistische Regression, Entscheidungsbaum und Neuronales Netz erstellt werden. Ziel war es, sinnvolle prognostische Faktoren zu ermitteln. Aber es sollten auch Sensitivitätsanalysen durchgeführt werden z.B. hinsichtlich der Ersetzungsmethoden für fehlende Werte im Replacement-Knoten oder hinsichtlich des Selektionskriteriums in der schrittweisen Logistischen Regression.

In den Seminaren im Hauptstudium des Diplomstudiengangs Statistik soll der Student /die Studentin selbständig wissenschaftliche Literatur verstehen, verarbeiten und darstellen lernen. Sie haben also sowohl eine wissenschaftliche als auch eine didaktische Funktion, welche beide insbesondere im Hinblick auf die Diplomarbeit von Bedeutung sind. Mit dem Enterprise Miner wurde in dem Seminar Theorie und Praxis verknüpft. Die Seminarteilnehmer stellten anhand wissenschaftlicher Literatur die Methoden der Statistik und des Data Mining vor und präsentierten die Umsetzung mit dem Enterprise Miner anhand kleiner Fallbeispiele zur Kreditwürdigkeitsprüfung, Warenkorbanalyse oder Marktsegmentierung. Die Beiträge des Seminars lauteten:

- Data Mining und der SAS Enterprise Miner™ 4.0 – Eine Einführung und Übersicht
- Explorative Datenanalyse im Enterprise Miner
- Warenkorbanalyse
- Ausreißeridentifikation und Ersetzungsverfahren für fehlende Werte
- Marktsegmentierung und Clusteranalyse
- Selbstorganisierende Karten und Kohonen-Netze
- Logistische Regression
- Entscheidungsbäume

- Künstliche Neuronale Netze
- Der Vergleich der Prognosegüte verschiedener Modelle
- Group Processing- und Ensemble Knoten

In Abhängigkeit vom inhaltlichen Umfang der einzelnen Themen und der Mächtigkeit der Enterprise Miner Knoten wurden die Beiträge von bis zu drei Studierenden in einer oder in zwei Seminarsitzungen präsentiert. In diesem Seminar bestand die Möglichkeit, verfeinerte Prognosemethoden wie Boosting oder Bagging im Ensemble Knoten oder Self-Organizing Maps und Kohonen-Netze zu besprechen, die bis dahin nicht zur "Standard"-Ausbildung eines Diplom-Statistikers / einer Diplom-Statistikerin gehörten.

### **2.3 Datenanalyse mit SAS**

Die Studierenden des Bachelor-Master-Studienganges müssen eine Lehrveranstaltung "Datenanalyse mit SAS" (pro Woche 4 Stunden Vorlesung und 2 Stunden praktische Übung im Rechnerpool) absolvieren. Diese Veranstaltung, die jährlich im Sommersemester angeboten wird, gehört für die Studierenden im Diplom-Studiengang Statistik zum Wahlpflichtbereich "Spezialgebiete der Statistik". Die Teilnehmerzahl liegt in der Regel zwischen 30 und 60 Personen. Während der Einführungskurs im Wesentlichen die Grundlagen der Software vorstellt, steht hier die statistische Datenanalyse im Vordergrund. Es wird vorausgesetzt, dass die Studierenden ein Basiswissen über SAS bereits erworben haben. Die Veranstaltung dient der Vertiefung der SAS-Kenntnisse. Wesentliche Bestandteile dieser Lehrveranstaltung in Bezug zur SAS Software sind:

- DATA-Step
- Einfache Prozeduren (PRINT, COMPARE CONTENTS, CORR, MEANS, UNIVARIATE, FREQ, etc.)
- SQL, sowie Vor- und Nachteile von SQL und "klassischer" SAS-Programmierung
- Import und Export von Daten sowie ODS
- Grafiken (SAS/GRAPH und SAS/INSIGHT)
- Regression, Varianzanalyse (REG, GLM)
- Generalisierte lineare Modelle (GENMOD, LOGISTIC, PROBIT, CATMOD)
- Nichtparametrische Verfahren (NPAR1WAY, FREQ)
- Multivariate Verfahren (PRINCOMP, FACTOR, CLUSTER)
- Zeitreihenanalyse (SAS/ETS)
- Makros
- IML

Während der Vorlesung liegt das Augenmerk auf der statistischen Theorie, grundlegender SAS-Syntax und einfachen Beispielen zur Illustration. In der Übung werden die in der Vorlesung erlernten SAS-Befehle eingesetzt und es erfolgt ein selbständiges Erarbeiten weiterer Sprachelemente und Optionen.

Die Leistung wird mittels einer 2-stündigen Klausur überprüft.

## **2.4 Wissensentdeckung in Datenbanken / Data Mining**

Zum Pflichtprogramm der Studierenden des Bachelor-Master-Studienganges gehört die Veranstaltung "Wissensentdeckung in Datenbanken", die gemeinsam von Lehrenden aus den Fachbereichen Statistik und Informatik angeboten wird. Die Teilnehmerzahl liegt in der Regel zwischen 30 und 40 Personen. Pro Woche umfasst die Veranstaltung 4 Stunden Vorlesung und 2 Stunden praktische Übung im Rechnerpool. Diese Veranstaltung wird zudem von Studierenden der entsprechenden Diplom-Studiengänge Statistik und Informatik im Wahlpflichtbereich (Spezialgebiete) belegt. Zu den Themen gehören:

- Data Mining Strategie CRISP (entwickelt von DaimlerChrysler, NCR und SPSS)
- Datenbanken und Data Cube
- Stichprobenverfahren
- Clusterverfahren (PROC CLUSTER)
- Klassifikationsverfahren (PROC DISCRIM, PROC LOGISTIC)
- Generalisierte lineare Modelle (PROC GENMOD, PROC LOGISTIC)
- Generalisierte additive Modelle (PROC GAM, sowie entsprechende Funktionen von R und S-Plus)
- Schätzung von Kovarianzmatrizen
- Entscheidungsbäume
- Zeitreihen
- Support Vector Machine (SVM)
- Konvexe Risikominimierung und Robustheit

In den Übungen wird wahlweise GNU-Software, wie z.B. die an der Universität Waikato entwickelte Software Weka und das Statistik-Programm R, sowie selbstverständlich SAS eingesetzt. Derzeit werden die klassischen SAS-Module aus der Hochschullandeslizenz NRW (u.a. BASE, STAT, GRAPH, ETS, IML) eingesetzt. Geplant ist zudem der Einsatz des SAS/Enterprise Miners und eine Beschreibung der SEMMA-Strategie, die von SAS verwendet wird.

### **3 SAS im Studium am Fachbereich Statistik Einsatz und Verfügbarkeit der SAS Software unter dem Blickwinkel "Studierende"**

Während der Einführungskurs in SAS und die Vorlesungen "Datenanalyse mit SAS" sowie "Wissensentdeckung in Datenbanken (Data Mining)" regelmäßig am Fachbereich angeboten werden, kam der SAS Enterprise Miner nur im Wahlpflichtbereich zum Einsatz. Hier spielen sowohl die Nachfrage und das Interesse von Seiten der Studierenden als auch die Verfügbarkeit des SAS Enterprise Miners an sich sowie der notwendigen Lehrkapazitäten eine Rolle; der SAS Enterprise Miner wurde vom Lehrstuhl von Herrn Hartung bisher mehrfach, aber unregelmäßig, in der Lehre eingesetzt.

Die SAS Software (Version 8) steht den Studierenden in den PC-Pools des Fachbereichs zur Verfügung. Viele Studierende arbeiten aber auch auf ihren privaten Rechnern mit einer SAS-Version, die sie zu günstigen Konditionen beim Hochschulrechenzentrum (im Rahmen der Landeslizenz) erwerben können. Dies spielt insbesondere für die Phase der Abschlussarbeit (Diplomarbeit etc.) eine wichtige Rolle für die Studierenden, während zuvor für die Bearbeitung von Übungsaufgaben oder kleineren Projekten, sofern dies hauptsächlich am eigenen Rechner daheim und nicht in einem der Pools des Fachbereichs geschieht, häufiger auf frei verfügbare Software zurückgegriffen wird.

Den Studierenden ist neben dem Einsatz der Software zur Lösung der ihnen gestellten Aufgaben auch wichtig, bereits während des Studiums Erfahrungen mit statistischer Software zu machen, die sie später im Beruf ebenfalls einsetzen werden/müssen. Insbesondere bei der SAS Software ist dabei zu beobachten, dass gerade auch Studierende gegen Ende ihres Studiums, oder auch Promovierende, die Veranstaltungen mit der SAS Software besuchen und den Einsatz der Software einüben.

### **4 Zusammenfassung und Fazit**

Ein großer Teil der Studierenden am Fachbereich Statistik beschäftigt sich im Rahmen des Studiums, - sei es in bestimmten Lehrveranstaltungen, sei es im Rahmen des Selbststudiums -, auch mit der SAS Software. Daneben gehört das Erlernen von R bzw. S-Plus zum Studienprogramm. Der Einsatz von SAS erfolgt insbesondere im Bereich der Datenanalyse, z.B. im fortgeschrittenen Studium oder auch in der Abschlussarbeit.

Gerade in letzter Zeit hat sich gezeigt, dass bei den Studierenden einerseits das Interesse an der Arbeit mit SAS wieder mehr zugenommen hat, andererseits bewirken

auch die ohnehin gestiegenen Studierendenzahlen eine größere Nachfrage nach Veranstaltungen mit SAS.

Langfristig zahlt sich das Festhalten an der Programmierung mit SAS im Gegensatz zu Bedienung mit der Maus ("Klickversionen") aus; die Studierenden erkennen und lernen es zu schätzen, dass sie beim Programmieren viel mehr Einfluss auf die Verfahren haben und sie in ihrem Sinne, für ihre Daten, für ihre Probleme adaptieren können, auch wenn der Zugang zum SAS System zunächst etwas aufwendiger und schwieriger erscheint. Daher ist der SAS/Enterprise Guide kein wesentlicher Bestandteil der Lehre am Fachbereich Statistik.

Wichtigste Lerninhalte beim Einsatz der SAS Software bleiben der DATA-Step sowie die Prozeduren aus dem Modul SAS/STAT.

Erste Einsätze des SAS / Enterprise Miners in der Lehre (Fallstudien, Seminar) wurden sowohl von den Studierenden als auch den Lehrenden begrüßt, die guten Erfahrungen sprechen für die weitere Verwendung des Miners in der Lehre am Fachbereich Statistik. Ein Einsatz des SAS / Enterprise Miners in der noch neuen Veranstaltung "Wissensentdeckung in Datenbanken / Data Mining" ist wünschenswert, aber derzeit aus unterschiedlichen Gründen noch nicht realisiert.

## **Literatur**

- [1] Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., Wirth, R. (2000). CRISP-DM 1.0. Step-by-step data mining guide. (<http://www.crisp-dm.org/CRISPWP-0800.pdf>)
- [2] Delwiche, L.D., Slaughter, S.J. (1998). The Little SAS Book: A Primer. (Second Edition). Cary, NC.
- [3] Dufner, J., Jensen, U., Schumacher, E. (2004). Statistik mit SAS. (3. Auflage). Teubner, Stuttgart.
- [4] Graf, A., Ortseifen, C. (1995). Statistische und grafische Datenanalyse mit SAS. Spektrum, Heidelberg.
- [5] Krämer, W., Schoffer, O., Tschiersch, L. (2005). Datenanalyse mit SAS. Statistische Verfahren und ihre grafischen Aspekte. Springer, Berlin.
- [6] SAS Manuals / SAS Online-Hilfe, SAS Institute Inc., Cary, NC, USA.