

Berichte aus der Statistik

**Christine Becker,  
Heinz Redlich (Hrsg.)**

**Data Mining und Statistik  
in Hochschule und Wirtschaft**

Proceedings der 7. Konferenz der SAS®-Anwender  
in Forschung und Entwicklung (KSFE)



Shaker Verlag  
Aachen 2003

**Bibliografische Information der Deutschen Bibliothek**

Die Deutsche Bibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.ddb.de> abrufbar.

Copyright Shaker Verlag 2003

Alle Rechte, auch das des auszugsweisen Nachdruckes, der auszugsweisen oder vollständigen Wiedergabe, der Speicherung in Datenverarbeitungsanlagen und der Übersetzung, vorbehalten.

Printed in Germany.

ISBN 3-8322-2111-5

ISSN 1619-0963

Shaker Verlag GmbH • Postfach 101818 • 52018 Aachen  
Telefon: 02407 / 95 96 - 0 • Telefax: 02407 / 95 96 - 9  
Internet: [www.shaker.de](http://www.shaker.de) • eMail: [info@shaker.de](mailto:info@shaker.de)

## Vorwort

Der vorliegende Proceedingsband enthält die ausgearbeiteten Beiträge zur 7. Konferenz der SAS Anwender in Forschung und Entwicklung (KSFE), die vom 20. bis 21. Februar 2003 von der Zentralen Einrichtung für Informationsverarbeitung und Kommunikation (ZEIK) und dem Fachbereich Mathematische Statistik der Universität Potsdam ausgerichtet wurde.

Mit dem Themenschwerpunkt *Data Mining & Statistik in Hochschule und Wirtschaft* informierte diese Konferenz, die mit über 320 Teilnehmern aus Deutschland, der Schweiz und Polen sehr gut besucht war, über das aktuelle Forschungsgeschehen an deutschen Hochschulen und die Weiterentwicklungen im Bereich Datenmanagement und Datenanalyse. Die SAS Software wird in über 100 Hochschulen in ganz Deutschland eingesetzt. Die nunmehr zum 7. Mal durchgeführte Konferenz versteht sich als Forum für den interdisziplinären Austausch zwischen Hochschulen und Unternehmen. Dieses Konzept ging auch in diesem Jahr für Veranstalter und Besucher voll auf – etwa die Hälfte aller Teilnehmer kamen in diesem Jahr aus der freien Wirtschaft. Die Besucher der 7. KSFE konnten sich über die neuesten Entwicklungen in den Bereichen Data Mining, Statistik und Datenmanagement informieren. SAS stellte u.a. den neuen SAS Enterprise Miner mit seiner Text-Mining-Technologie vor.

Wir möchten den folgenden Firmen danken, die durch ihr Sponsoring die Konferenz maßgeblich unterstützten: dem Hauptsponsor SAS Institute GmbH (Heidelberg), der Condat Informationssysteme AG (Consulting Partner im SAS Alliance Program) sowie der Firma Systematika GmbH (Heidelberg), die auch den Preis für den besten Beitrag stiftete. Diesen Preis erhielt in diesem Jahr Herr Andreas Christmann (Universität Dortmund) für sein Tutorium *Generalisierte lineare Modelle mit SAS 8e*.

Unser besonderer Dank geht an das KSFE-Organisationskomitee für die sehr gute Zusammenarbeit bei der Vorbereitung und Durchführung der Tagung; ihre Erfahrungen aus den sechs früheren KSFE-Konferenzen erleichterte uns die Arbeit wesentlich.

Danken möchten wir auch den Mitarbeiterinnen und Mitarbeitern der Zentralen Einrichtung für Informationsverarbeitung und Kommunikation sowie dem Fachbereich Mathematische Statistik der Universität Potsdam für ihren Einsatz vor, während und nach der Konferenz. Der Erfolg dieser Tagung wurde durch ihr Engagement erst möglich.

Informationen über die nächste KSFE finden sie unter <http://www.ksfe.de>.

Potsdam, im Jahre 2003

Christine Becker

Heinz Redlich



# Inhalt

<i>H.-P. Altenburg</i> <b>Über die Aufdeckung von Ernährungsmustern in epidemiologischen Studien</b>	1
<i>K. Bruderermann</i> <b>Strategie der statistischen Versuchsmethodik. Ein anschaulicher Experimentalvortrag mit Flugobjekten.</b>	23
<i>A. Christmann</i> <b>Generalisierte lineare Modelle mit SAS 8e</b>	
<i>T. G. Grobe</i> <b>Extraktion von Informationen aus binären Dateien: Ein SAS-Makro zur Erfassung von strukturierten Meta-Informationen aus Bilddateien aktueller Digitalkameras</b>	37
<i>T. G. Grobe</i> <b>Aufarbeitung von überlappenden zu eindeutig abgegrenzten Zeitintervallen unter Beibehaltung der in Primärdaten enthaltenen Informationen</b>	45
<i>M. Hensel</i> <b>Dynamische Programmcode-Erzeugung mit CALL EXECUTE - Möglichkeiten, Grenzen, Alternativen</b>	49
<i>W. Herff, C. Ortseifen, G. Pfister, H. Stürzl</i> <b>Tipps &amp; Tricks - für einen leichteren Umgang mit der Base SAS Software</b>	57
<i>Y. Hoffmann</i> <b>Regionalisierung von Business-Graphiken für die Kartographie mit SAS/GRAPH</b>	97
<i>H.-P. Höschel</i> <b>Auswertung der SAS-Stellenanzeigen in der FAZ von 1994-2002</b>	111
<i>W. Janiesch</i> <b>Von Statistik zu Analytical Intelligence</b>	117
<i>B. Kalina, K. Klenke, O. Gefeller, A. Pfahlberg</i> <b>Automatische Modellselektion bei der loglinearen Modellierung in SAS</b>	125

<i>M. Kawohl, D. Spruck</i> <b>SASGraphs2PDF - Erstellung einer PDF Datei aus SAS Graphiken mit Hilfe von ODS PDF</b>	133
<i>J. Lauter</i> <b>Hochdimensionale Statistik – Mathematik und Computer als Werkzeuge</b>	139
<i>R. Muche, C. Ring, I. Assfalg</i> <b>Ein SAS-Makro fur die Bootstrap-Validierung logistischer Prognosemodelle</b>	141
<i>C. Pares-Salvador, C. Tuckmantel</i> <b>bioWORKSTM – Losungsbeispiele zur Automatisierung statistischer Verfahren in der Pharmaindustrie</b>	151
<i>S. Pohl, S. Steinberg</i> <b>Heuristische Verfahren zur Aggregation addierbarer Zeitreihen</b>	157
<i>T. Rudiger</i> <b>Response-emails via SAS</b>	167
<i>P. E. Rudolph, A. Tuchscherer, M. Tuchscherer, B. Jager</i> <b>Bootstrap mit SAS</b>	175
<i>V. Schultze-Pawlitschko, M. Kersting</i> <b>Clusteranalyse mit SAS fur unterschiedlich lange Verlaufskurven mit verschiedenen Startpunkten</b>	195
<i>M. Seyerle</i> <b>Customer Lifetime Value</b>	205
<i>M. Spiliopoulou, A. Schulz, K. Winkler</i> <b>Text Mining an der Borse: Einfluss von Ad-hoc-Mitteilungen auf die Kursentwicklung</b>	215
<i>C. Tuckmantel</i> <b>Qualitatssicherung industrieller Prozesse mit dem SAS Enterprise Miner</b>	229
<i>S. Witte, K. Jensen</i> <b>Ein Konfidenzband fur ROC-Kurven mit SAS</b>	239

*M. Wodny, B. Jäger, K.-E. Biebler*

249

**Ausgleichende natürliche kubische Splines und die Schätzung des  
Glättungsparameters in SAS**