

Berichte aus der Statistik

**Andreas Christmann,  
Claus Weihs (Hrsg.)**

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

Proceedings der 6. Konferenz der SAS<sup>®</sup>-Anwender  
in Forschung und Entwicklung (KSFE)

Shaker Verlag  
Aachen 2003

Die Deutsche Bibliothek - CIP-Einheitsaufnahme

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

Proceedings der 6. Konferenz der SAS®-Anwender in Forschung  
und Entwicklung (KSFE) / Andreas Christmann, Claus Weihs (Hrsg.).

Aachen : Shaker, 2003

(Berichte aus der Statistik)

ISBN3-8322-1244-2

Copyright Shaker Verlag 2003

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

Printed in Germany.

ISBN 3-8322-1244-2

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

Dieser Proceedingsband enthält die ausgearbeiteten Beiträge zur 6. Konferenz der SAS Anwender in Forschung und Entwicklung (KSFE), die vom 28. Februar bis 1. März 2002 vom Hochschulrechenzentrum und dem Fachbereich Statistik der Universität Dortmund ausgerichtet wurde.

Mit dem Themenschwerpunkt *Data Mining* informierte diese Konferenz, die mit fast 400 Teilnehmern einen neuen Besucherrekord verbuchte und damit zu den größten Konferenzen zu diesem Thema in Deutschland zählte, über das aktuelle Forschungsgeschehen an deutschen Hochschulen, die Weiterentwicklung dieser Technologie und gleichzeitig über den Bedarf von hochqualifizierten Absolventen im Bereich Datenmanagement und Datenanalyse. Die SAS Software wird in rund 100 Hochschulen in ganz Deutschland eingesetzt. Die jährlich stattfindende Konferenz versteht sich als Forum für den interdisziplinären Austausch zwischen Hochschulen und Unternehmen – ein Konzept, das für Veranstalter und Besucher voll aufging: mehr als die Hälfte aller Teilnehmer kamen in diesem Jahr aus der freien Wirtschaft. Die Teilnehmer der 6. KSFE konnten sich über die neuesten Entwicklungen in den Bereichen Data Mining, Statistik und Datenmanagement informieren. SAS stellte neben dem SAS Enterprise Miner erstmals seine neue Text-Mining-Technologie vor.

Danken möchten wir den folgenden Firmen, die durch ihr Sponsoring diese Konferenz maßgeblich unterstützten: dem Hauptsponsor SAS (Heidelberg) sowie Bayer AG Integrated Analyses (Wuppertal), Info Ware GmbH (Heidelberg) und Schering AG (Berlin). Dem Sonderforschungsbereich 475 „Komplexitätsreduktion in multivariaten Datenstrukturen“ an der Universität Dortmund danken wir für die finanzielle Unterstützung der eingeladenen Vortragenden.

Den von der Firma Systematika GmbH in Heidelberg gestiftete Preis für den besten Beitrag erhielten Frau Dr. Carina Ortseifen (Universität Heidelberg), Herr Grischa Pfister (Info Ware, Heidelberg) und Herr Heinrich Stürzl (Bade Bering, Marburg) für ihren gemeinsamen Vortrag zum Thema Tipps & Tricks rund um die SAS-Software.

Unser besonderer Dank geht an Frau Petra Schlager und Herrn Dr. Manfred Thibud vom Hochschulrechenzentrum der Universität Dortmund, deren Unterstützung in allen Phasen der Planung und Durchführung der Konferenz sowie der Erstellung des Proceedingsbands und der zugehörigen CD wesentlich zum Gelingen der Konferenz beigetragen hat.

Wir möchten dem KSFE-Organisationskomitee danken für die gute Zusammenarbeit bei der Vorbereitung und Durchführung der Tagung; ihre Erfahrung aus früheren KSFE-Konferenzen erleichterte uns die Arbeit erheblich.

Danken möchten wir auch dem Leiter des Hochschulrechenzentrums, Herrn Günter Schwichtenberg, für sein Verständnis und seine Unterstützung. Frau Anja Lorenz und Herrn Björn Schubert gilt unser Dank für ihre professionelle technische Unterstützung bei der Erstellung des Proceedingsbands. Nicht zuletzt danken wir allen beteiligten Mitarbeitern und Mitarbeiterinnen des Hochschulrechenzentrums, des Fachbereichs Statistik und des SFB 475 für Ihren Einsatz vor, während und nach der Konferenz. Durch Ihr Engagement wurde der Erfolg dieser Tagung erst möglich.

Informationen über die nächste KSFE finden Sie unter [www.ksfe.de](http://www.ksfe.de).

Dortmund, im Januar 2003

Andreas Christmann

Claus Weihs

<i>H.-P. Altenburg</i>	1
<b>Möglichkeiten der SAS Software für die Analyse großer Datentabellen</b>	
<i>C. Berlin, M. Probach</i>	17
<b>Statistische Standard Makros für die Analyse klinischer Studiendaten</b>	
<i>H. Bleiholder, C. Janson, G. Jung</i>	23
<b>Ein System zur Erfassung, Speicherung und Auswertung landwirtschaftlicher Versuche – Einsatz von SAS in Datenbankzugriff und statistischer Analyse</b>	
<i>A. Böckenhoff, J. Hartung</i>	35
<b>Zweistufige Schätzung in der Meta-Analyse</b>	
<i>T. Bruckner, H. Dickel</i>	47
<b>Muster und Häufigkeit von Sensibilisierungen im Epikutantest: Eine SAS-Anwendung</b>	
<i>B. Jäger, M. Wodny, K.-E. Biebler, P.E. Rudolph, K. Matthies</i>	57
<b>Diskriminanzanalyse mit binären Daten</b>	
<i>W. Janiesch</i>	65
<b>Ausbildung, Forschung &amp; Anwendung: Die KSFE als Brücke zwischen Hochschule und Unternehmen</b>	
<i>S. Klein, A. Frey, U. Gauger</i>	71
<b>Veränderungsmessung mit dem Linear Logistic Test Model</b>	
<i>G. Knapp, J. Hartung</i>	79
<b>Data Mining in einer epidemiologischen Studie zu Oberbauchbeschwerden (PRESTO)</b>	
<i>O. Kuss, H. Dickel</i>	89
<b>Eine Verallgemeinerung des Cochran-Armitage-Tests auf mehrere Gruppen mit PROC LOGISTIC und PROC GENMOD</b>	
<i>W.F. Lesener</i>	97
<b>Erfassung von Internetfragebögen</b>	
<i>T. Lochmann, C. Ortseifen</i>	161
<b>Von SAS nach SPSS und umgekehrt – Umwandlung von Dateien inklusive Formaten bzw. Wertelabels</b>	

<i>E. Moll</i>	177
<b>Die biometrische Fortbildung im Geschäftsbereich des Bundesministeriums für Verbraucherschutz, Ernährung und Landwirtschaft (BMVEL)</b>	
<i>R. Muche, A. Habel, F. Rohlmann</i>	187
<b>Medizinische Statistik mit SAS-Analyst</b>	
<i>C. Ortseifen, G. Pfister, H. Stürzl</i>	197
<b>Tipps &amp; Tricks – Nützliche Lösungen zu Problemen und Fragestellungen der SAS Versionen 8, 8.1 und 8.2</b>	
<i>A. Pfahlberg, T. van der Horst, O. Gefeller</i>	219
<b>Ein SAS-Macro zur Berechnung des epidemiologischen Risikomaßes der Risk and Rate Advancement Period</b>	
<i>T. Riemer, H.H. Leuschner</i>	233
<b>Trendanalyse mit gemischten linearen Modellen SAS-Lösungen und Anwendung in der Ökologie</b>	
<i>P.E. Rudolph, P. Havemann, M. Wodny</i>	245
<b>Datentransfer zwischen SAS 8.2 und MS SQL-Server bzw. Oracle mit ODBC</b>	
<i>T. Rüdiger</i>	259
<b>Aufbau und Pflege von Kunden-Ids im Finanzdienstleistungsgewerbe</b>	
<i>V. Schultze-Pawlitshko, M. Kersting</i>	265
<b>Die Anwendung des SAS-Makros GLIMMIX bei der statistischen Auswertung von Langzeitstudien in der Ernährungsmedizin am Beispiel der DONALD-Studie</b>	
<i>K. Steuernagel</i>	273
<b>Data Mining im e-commerce am Beispiel der Deutschen Bahn AG</b>	
<i>D. Stöckler, N. Wachter, M. Selic, A. Spatz</i>	279
<b>Kundensegmentation bei der WEKA Media GmbH Ansätze zur Kundenbewertung im CRM – Möglichkeiten zur Bestimmung des Kundenpotenzials Datengestützte Validierung von Customer-Lifetime-Konzepten</b>	
<i>H. Stürzl</i>	295
<b>PDF mit SAS 8.2</b>	
<i>D. Taeger</i>	307
<b>Zugriff auf eine externe Dynamic Link Library zur Berechnung von Personenjahren und Standardisierten Mortalitäts Ratios mit SAS</b>	

<i>A. Tuchscherer, P.E. Rudolph, B. Jäger, M. Tuchscherer</i>	315
<b>Bewertung von Diskriminanzanalyseverfahren in SAS bei Nichtnormalität</b>	
<i>H.-F. Ulbrich, K. Holzapfel</i>	327
<b>eRandom</b>	
<b>das SAS®-Produkt für Randomisierungen in F&amp;E unter GCP- (Good Clinical Practice) Anforderungen</b>	
<b>Eine gemeinsame Entwicklung von SAS Deutschland und der Schering AG</b>	
<i>M. Wahl</i>	339
<b>Auswertung klinischer Studien mit SAS/PH-Clinical</b>	
<i>K. Winkler, M. Spiliopoulou</i>	347
<b>Text Mining in der Wettbewerberanalyse:</b>	
<b>Konvertierung von Textarchiven in XML-Dokumente</b>	
<i>G. Winterstein</i>	365
<b>Einsatz von Data Mining Methoden zur Prognose von studentischem Verhalten</b>	
<i>S. Witte</i>	375
<b>Macro zur Erstellung von ROC-Kurven</b>	