



# KSFE 2012

16. Konferenz der SAS® Anwender  
in Forschung und Entwicklung

8.–9. März 2012 an der Technischen Universität Dresden





Liebe SAS Interessierte,

das Competence Center for Business Intelligence der Fakultät Wirtschaftswissenschaften richtet am 8. und 9. März 2012 die 16. Konferenz der SAS Anwender in Forschung und Entwicklung (KSFE) an der Technischen Universität Dresden aus.

Ein Highlight gleich zu Beginn der Konferenz ist der Eröffnungsvortrag von John Sall, Mitbegründer und Executive Vice President von SAS sowie Entwickler der JMP Software und Leiter des Unternehmensbereichs JMP, der interaktive und visuelle Software zur Datenanalyse kreiert.

Der spannende Eröffnungsvortrag trägt den Titel:

*Speeding Up Exploration*  
*Discover more, and find out in time for decisions to matter more*

Danach bieten wir den Tagungsteilnehmern an den beiden folgenden Tagen ein weiterhin interessantes Programm: Anwender berichten in ansprechenden Vorträgen von ihren Erfahrungen im Umgang mit der SAS® Software und geben ihr Fachwissen an andere Teilnehmer weiter. Neue Anwender können in einführenden Übersichtsvorträgen und Tutorien die ersten Schritte beim Umgang mit SAS erlernen.

Am Vortag der Konferenz gibt es erstmalig neben den etablierten Tutorien die Möglichkeit, als Besucher der KSFE zu besonderen Konditionen an einer speziell für die KSFE2012 angebotenen SAS Zertifizierung zum SAS® Certified Base Programmer for SAS®9 oder SAS® Certified Advanced Programmer for SAS®9 teilzunehmen. Wir freuen uns als Ausrichter sehr, dass dieses Angebot auf eine so große Resonanz gestoßen ist.

Diskussionmöglichkeiten für die Teilnehmer gibt es im Rahmen einer Posterausstellung, in den Konferenzpausen, bei den Ständen der zahlreichen Firmen oder auch während der Abendveranstaltung. Diese findet an Bord der „MS August der Starke“, einem stolzen Vertreter der Sächsischen Dampfschiffahrt, statt. Auf der Fahrt von Dresden nach Pillnitz können Sie die historische Kulisse der Dresdner Altstadt bewundern. Ein ausgiebiges Buffet wird für das leibliche, rege Diskussionen mit Kollegen Ihres Fachs für das geistige Wohl sorgen.

Wie in den Jahren zuvor werden alle Vorträge wieder vom Publikum bewertet und die drei Besten mit dem „KSFE Best Paper Award“ ausgezeichnet.

Im Namen aller Organisatoren und gemeinsam mit dem Verein KSFE e.V. freuen wir uns auf einen regen Austausch!

Ihr Chairman



Andreas Hilbert

### Organisationskomitee

**Ralf Minkenberg**

KSFE e.V. und Boehringer Ingelheim

**Dr. Carina Ortseifen**

KSFE e.V. und Universitätsrechenzentrum Heidelberg

**Karin Pees**

KSFE e.V. und SAS Deutschland

### Lokale Organisation

**Prof. Dr. Andreas Hilbert**

Technische Universität Dresden

**Alexander E. Müller**

Technische Universität Dresden

**Andreas Schieber**

Technische Universität Dresden

**Tobias Weiß**

Technische Universität Dresden

### Kontakt:

Competence Center for Business Intelligence  
Technische Universität Dresden, Fakultät Wirtschaftswissenschaften

Prof. Dr. Andreas Hilbert

Münchner Platz 3

D - 01187 Dresden

Telefon: +49 (0) 351 463 32268

Telefax: +49 (0) 351 463 32736

E-Mail: [ksfe2012@tu-dresden.de](mailto:ksfe2012@tu-dresden.de)

Web: [ksfe2012.wiwi.tu-dresden.de](http://ksfe2012.wiwi.tu-dresden.de)

### Öffnungszeiten: Tagungsbüro

Mittwoch, 07.03.12 12:00 - 18:00 Uhr

Donnerstag, 08.03.12 08:00 - 18:00 Uhr

Freitag, 09.03.12 08:00 - 14:00 Uhr

Zu diesen Zeiten ist am Donnerstag und Freitag auch die Garderobe gegenüber dem Raum HSZ 02 besetzt.

**Programm**

**Tutorien im Vorfeld der Konferenz | Mittwoch, 07. März 2012**

<b>SAS Zertifizierung</b> Hörsaalzentrum / Fakultät Wirtschaftswissenschaften		<b>Tutorium I</b> Hörsaalzentrum HSZ	<b>Tutorium II</b> Hörsaalzentrum HSZ
14:00 – 18:00	Zertifizierung zum <b>SAS Certified Base Programmer for SAS 9</b> oder <b>SAS Certified Advanced Programmer for SAS 9</b>	<b>Statistische Grafiken mit ODS Graphics: aussagekräftig und ansprechend</b>  Dorothee Hildebrand, <i>SAS Deutschland</i>	<b>Einführung in die SAS Makrosprache</b>  Grischa Pfister, <i>iCASUS GmbH</i>

**Vorträge | Donnerstag, 08. März 2012**

<b>Eröffnung</b> Hörsaalzentrum HSZ 02  Moderation: A. Hilbert, <i>Technische Universität Dresden</i>			
10:00 – 11:00	<b>Begrüßung der Konferenzteilnehmer und Eröffnung der 16. KSFE</b>  <b>Eröffnungsvortrag „Speeding Up Exploration“</b>  John Sall, <i>Co-Founder und Executive Vice President SAS</i>		
		<b>Ausbildung und Lehre</b> Hörsaalzentrum HSZ 03  Leitung: A. Hilbert, <i>TU Dresden</i>	<b>Statistik I</b> Hörsaalzentrum HSZ 04  Leitung: R. Muche, <i>Universität Ulm</i>
11:30 – 12:00		<b>SAS OnDemand for Academics – erste Erfahrungsberichte</b>  Katja Deutsch, <i>Universität Koblenz- Landau</i> Stefan Marquardt, <i>Hochschule Fulda</i>	<b>Hierarchische Bayes'sche Netze: Eine Methode, um Dokumentensammlungen in den hierarchischen Strukturen zu klassifizieren – so, wie ein Mensch die Texte schreibt</b>  Ulrich Reincke, <i>SAS Deutschland</i>
12:00 – 12:30		<b>SAS Enterprise Miner in der Lehre – von der Einführung zur Zertifizierung</b>  Andreas Schieber, <i>Technische Universität Dresden</i>	<b>Bestimmung der optimalen Fallzahl in Dosis-Findungs- Studien mit aktiver Kontrolle und linearen sowie nicht- linearen Dosis-Wirkungs- Profilen</b>  Hans-Joachim Helms, Tim Friede, Norbert Bender, <i>Universität Göttingen</i>
12:30 – 13:00		<b>SAS Know-how in der Organisation entwickeln und pflegen – Ressourcen für die Ausbildung und für das Wissensmanagement</b>  Andreas Mangold, <i>HMS Analytical Software</i>	<b>Randomisierungsverfahren mit SAS</b>  B.P. Jäger, C. Malsch, K.-E. Biebler, <i>Ernst-Moritz-Arndt- Universität Greifswald</i>  P.E. Rudolph, <i>Universität Dummerstorf</i>

# Programm

13:00 – 14:00	<b>Mittagspause</b>		
	<b>Text und Web Mining I</b> Hörsaalzentrum HSZ 02  Leitung: N. Kleekamp, <i>Universität Jena</i>	<b>Programmierung</b> Hörsaalzentrum HSZ 03  Leitung: H. Ramroth, <i>Institute of Public Health, Heidelberg</i>	<b>Schnittstellen</b> Hörsaalzentrum HSZ 04  Leitung: C. Ortseifen, <i>Universität Heidelberg</i>
14:00 – 14:30	<b>Einsatz des neuen Text Miner 5.1 zur Nachrichtenklassifikation von Online Medien</b>  Ulrich Reincke, <i>SAS Deutschland</i>	<b>Tipps zur SAS Programmierung in der PIAFStat-Umgebung</b>  Andrea Zenk, Volker Michel, <i>Landesforschungsanstalt für Landwirtschaft und Fischerei Mecklenburg-Vorpommern</i>	<b>Webservices und SAS – PROC SOAP, PROC HTTP und der ganze REST</b>  Martin Haffner, Andreas Mangold, <i>HMS Analytical Software GmbH</i>
14:30 – 15:00	<b>Social Media Analysis</b>  Kai Heinrich, <i>Technische Universität Dresden</i>	<b>Erstellen flexibler Makros am Beispiel einer SQL-Routine</b>  Andreas Deckert, <i>Institute of Public Health, Heidelberg</i>	<b>Das Beste aus zwei Welten – Aufruf von R-Funktionen mit PROC IML</b>  Dr. Peter Beyer, <i>HMS Analytical Software GmbH</i>
15:00 – 15:30	<b>Posterausstellung</b>		
15:30 – 16:00	<b>Kaffeepause</b>		
		<b>SAS/Base</b> Hörsaalzentrum HSZ 03  Leitung: P. Würzburger	<b>Statistik II</b> Hörsaalzentrum HSZ 04  Leitung: R.-H. Boedeker, <i>Universität Gießen</i>
16:00 – 16:30		<b>SAS Backstage</b>  Biljana Gigic, <i>Nationales Centrum für Tumorerkrankungen, Heidelberg</i> Andreas Deckert, <i>Institute of Public Health, Heidelberg</i>	<b>Der Umgang mit fehlenden Werten in Verlaufsstudien im Falle von Drop-outs</b>  Benjamin Mayer, <i>Universität Ulm</i>
16:30 – 17:00		<b>Tipps, Tricks und Stolperfallen zum Umwandeln von Variablen</b>  Sabine Erbslöh, <i>Accovion GmbH</i>	
17:00 – 17:30		<b>Proc Transpose oder Do-it-yourself</b>  Dr. Heribert Ramroth, <i>Institute of Public Health, Heidelberg</i>	<b>Modellierung rechts-zentrierter Zähldaten – ein SAS Makro unter Nutzung von PROC NL MIXED</b>  Hannes-Friedrich Ulbrich, <i>Bayer Pharma AG</i>
18:30 – 22:00	<b>Abendveranstaltung an Bord der „MS August der Starke“ der Sächsischen Dampfschiffahrt</b> Das Schiff legt pünktlich um 19:00 Uhr ab.		

Vorträge | Freitag, 09. März 2012

<b>Neuere Anwendungen</b> Hörsaalzentrum HSZ 02  Leitung: J.Petzold, SAS <i>Deutschland</i>		<b>JMP</b> Hörsaalzentrum HSZ 03  Leitung: H.-P. Altenburg, <i>Siemens Healthcare Diagnostics GmbH</i>	<b>Text und Web Mining II</b> Hörsaalzentrum HSZ 04  Leitung: S. Mangold, <i>HMS Analytical Software GmbH</i>
09:00 – 09:30	<b>Die neue Style Engine in ODS – Cascading Style Sheets mit SAS 9.3</b>  Gregor Hermann, SAS <i>Deutschland</i>	<b>JMP10: Sneak Preview für Entwicklung und Modellierung</b>  Bernd Heinen, <i>SAS US JMP</i>	<b>Erweiterungen der Analysemöglichkeiten im neuen Enterprise Miner 7.1 sowie im Model Manager</b>  Martin Schütz, <i>SAS Deutschland</i>
09:30 – 10:00	<b>SAS in der Praxis – Tipps und Tricks für das SAS Web Report Studio</b>  Andreas Bachert, <i>HMS Analytical Software GmbH</i>	<b>Statistische Versuchsplanung mit JMP – von der Klassik zur Moderne</b>  David Meintrup, <i>Hochschule für Angewandte Wissenschaften, Ingolstadt</i>	<b>Einstieg in Text Analytics für SAS Enterprise Guide Anwender: Von der Datenquelle zum Bericht – und darüber hinaus</b>  Johannes Lang, <i>HMS Analytical Software GmbH</i>
10:00 – 10:30	<b>Große Datenmengen knacken mit SAS High Performance Analytics</b>  Martin Schütz, SAS <i>Deutschland</i>		
10:30 – 11:00	<b>Kaffeepause</b>		
<b>Reporting</b> Hörsaalzentrum HSZ 02  Leitung: F. Mader, <i>Boehringer Ingelheim Pharma GmbH</i>		<b>Statistik III</b> Hörsaalzentrum HSZ 03  Leitung: R. Minkenber, <i>Boehringer Ingelheim Pharma GmbH</i>	<b>Freie Themen</b> Hörsaalzentrum HSZ 04  Leitung: P. Krüger, <i>Ostdeutsche Versicherung AG</i>
11:00 – 11:30	<b>Implementation of ICH E3 based SAS reporting for a clinical trial – issues and solutions</b>  Arne Ring, Ruth L. Coleman, <i>Diabetes Trials Unit, OCDEM, Oxford University</i>	<b>Ein Makro zur Berechnung von Diskriminanz- und Reklassifizierungsstatistiken für eine Verbesserung eines Prädiktionsmodells bei Anwendung der Cox Regression</b>  Kristin Mühlenbruch, Wolfgang Bernigau, Matthias Schulze, <i>Deutsches Institut für Ernährungsforschung, Potsdam- Rehbrücke</i> Lars Berglund, <i>Uppsala Clinical Research Center</i>	<b>Verwaltung eines Metadatenrepositories im DI 3.4 mittels einer externen Exceldatei</b>  Dolic Dubravko, <i>dsquare.de für Santander</i>
11:30 – 12:00	<b>Performante Erzeugung von Berichten in (vor)formatierten Exceldateien</b>  Stephan Frenzel, Sarah Baumert, Dr. Robert Koob, <i>Kybeidos GmbH</i> Ralf Neumüller, <i>beyondbytes</i>	<b>Ein modulares Statistikmakro-Paket für direkte und indirekte Meta-Analysen</b>  Dr. Hendrik Schmidt, Dr. Clemens Tilke-Schöneck, <i>Boehringer Ingelheim Pharma GmbH</i>	<b>Methoden der Risikobewertung</b>  Jörg Schmidtke, <i>BioMath GmbH</i>

## Programm

12:00 – 12:30	<b>ODS Report Writing Interface – Neue Möglichkeiten der Berichtserstattung</b>  Sebastian Reimann, <i>viadee Unternehmensberatung GmbH</i>	<b>Berechnung des relativen Überlebens auf Basis von Krebsregisterdaten mit SAS</b>  Dr. Olaf Schoffer, <i>Universitäts KrebsCentrum, Universitätsklinikum Carl Gustav Carus an der TU Dresden</i>	<b>Datenvisualisierung mit SAS Visual Analytics Explorer</b>  Gregor Hermann, <i>SAS Deutschland</i>
<b>Abschluss</b> Hörsaalzentrum HSZ 02  Moderation: R. Minkenber, <i>Vorstand KSFE e.V.</i>			
12:45 – 13:15	<b>Verabschiedung</b> durch Andreas Hilbert, <i>TU Dresden</i> und Ralf Minkenber, <i>Vorstand KSFE e.V.</i> <b>Preisvergabe des Best Paper Award der 16. KSFE 2012</b> durch Michael Wetzel, <i>Systematika GmbH</i> <b>Ausblick auf die 17. KSFE 2013</b>		
13:15 – 14:00	<b>Mittagspause und Veranstaltungsende</b>		

## Begleitende Posterausstellung

Neben den Tutorien und Vorträgen findet während der Konferenz eine Posterausstellung mit folgenden Beiträgen statt:

### **Laufzeitoptimierung bei der Verknüpfung großer Datenmengen. Ein Vergleich zwischen MERGE im DATA STEP und JOIN in PROC SQL**

*Cerstin Erler, Institut für Arbeitsmarkt- und Berufsforschung der Bundesagentur für Arbeit (IAB), Nürnberg*

### **Ein SAS Macro zur Auswertung von SAE-Daten**

*Thomas Bruckner, Andreas Deckert, Universität Heidelberg*

### **SAS Macro zur Kombination von möglichen in Zusammenhang stehenden Faktoren bezüglich einer Zielgröße beschrieben anhand eines aktuellen Beispieldatensatzes**

*Martina Malzer, Annette Pfahlberg, Wolfgang Uter, Olaf Gefeller, alle Universität Erlangen-Nürnberg; Janice Hegewald, TU Dresden*

### **Medizinische Statistik mit ??? in SAS<sup>®</sup> 9.3 – ein Plädoyer für eine Windows-nahe Oberfläche**

*Rainer Muche, Benjamin Mayer, Universität Ulm*

### **DATA Step Component Interface – neue Objekte im DATA Step**

*Sebastian Reimann, viadee Unternehmensberatung GmbH*

### **Nie wieder Post sortieren – vollautomatisierte Verarbeitung von Kundenpost mittels Text-Mining-Verfahren**

*Thomas Hemmler, Horst Meyer, Sergej Steinberg, Bauer Systems KG*

### **Anwendung eines SAS/STAT-Modells ohne SAS/STAT in einem CRM-Projekt**

*Tobias Otte, Timm Euler, viadee Unternehmensberatung GmbH*

Die Autoren bieten interessierten Konferenzteilnehmern am Donnerstag, den 08. März 2012, von 15:00 bis 15:30 Uhr direkt bei ihren Postern Gelegenheit zur Diskussion.

## **SAS Zertifizierung exklusiv für KSFE Teilnehmer**

Erstmalig können die KSFE Teilnehmer die Gelegenheit nutzen und ihre soliden SAS Kenntnisse schwarz auf weiß dokumentieren lassen. Am Mittwoch von 14:00 - 18:00 Uhr bieten wir für angemeldete Teilnehmer parallel zu den Tutorien zwei Zertifizierungsmöglichkeiten:

### **SAS® Certified Base Programmer for SAS®9**

Die Zertifizierung zum „SAS Certified Base Programmer for SAS®9“ belegt Ihre grundlegenden Kenntnisse des Datenmanagements mit der SAS Software. Das Bestehen dieses Tests ist Voraussetzung für die Teilnahme an der Zertifizierung „SAS Certified Advanced Programmer for SAS®9“.

### **SAS® Certified Advanced Programmer for SAS®9**

Die Zertifizierung zum „SAS Certified Advanced Programmer for SAS®9“ überprüft fortgeschrittene Kenntnisse in der SAS Programmierung. Der Test erfordert hohe Professionalität im Programmieren mit SAS.

Alternativ ist die Teilnahme an den Zertifizierungstests zu BI, DI, Platform Administration oder Predictive Modeling möglich. Weitere Infos zur Zertifizierung allgemein unter: [www.sas.de/cert](http://www.sas.de/cert)

Die Teilnehmer treffen sich im Hörsaalzentrum und begeben sich dann – gemeinsam mit einem der lokalen Organisatoren – zu den Computerräumen der Fakultät Wirtschaftswissenschaften.

## **Tutorium I: Statistische Grafiken mit ODS Graphics: aussagekräftig und ansprechend**

### **Dorothee Hildebrand**

*SAS Deutschland*

Gute effektive Grafiken sind in der geschäftlichen und wissenschaftlichen Analyse heute selbstverständlich, aber auch bei komplexen Fragestellungen wichtiger denn je. Sie liefern Datenmuster und –unterschiede, stellen statistische Auswertungen dar und werfen mit dem Visualisieren der Ergebnisse neue Fragestellungen auf.

Die Graph Templates in ODS sind eine Erweiterung des Output Delivery Systems in SAS, mit denen in statistischen Prozeduren Ergebnisse direkt sowohl in Tabellenform als auch in Graphiken gleichermaßen erstellt werden. Die Ausgabe kann an unterschiedliche Zielorte wie HTML, pdf oder in andere Formate gestellt werden, wobei eine konsistente Erscheinung und ein einheitlicher Stil für Tabellen und Graphiken erreicht wird.

Da die SAS Software Graph Templates standardmäßig zur Verfügung stellt, müssen wir eigentlich nichts über deren Programmierung wissen, um diese zu nutzen. Wir können aber mit etwas Hintergrundwissen Graphiken anpassen und individuelle Graphiken selbst erstellen. Auf Grundlage der Graph Template Language werden dem Benutzer verschiedene Komponenten mit unterschiedlichen Möglichkeiten und unterschiedlicher Komplexität bereitgestellt, um individuell angepasste Grafiken zu erzeugen:

Graph Templates als Standard und die Möglichkeit, editierbare Graphiken zu erstellen, Prozeduren der Statistical Graphics Familie, ODS Graphics Designer und die Graph Template Language. Auf diesem Tutorium sollen die verschiedenen Komponenten eingeführt und mit Beispielen aus dem geschäftlichen Leben sowie wissenschaftlichen und klinischen Studien vorgestellt werden.

Außerdem wollen wir Ihnen das eine oder anderen neue Highlight, das SAS9.3 in die Graphiken mitbringt, nicht vorenthalten.

### **Tutorium II: Einführung in die SAS Makrosprache**

**Grischa Pfister**  
*iCASUS GmbH*

Die SAS Makrosprache, vor Jahren fast schon für tot erklärt, feiert fröhlich Urständ. Sie ist heute die Basis von Stored Processes und damit ein zentraler Bestandteil der SAS 9 Platform. Grund genug, erneut eine Einführung anzubieten.

Die Makrosprache erweitert SAS/BASE um eine andere Ebene der Programmierung. Steht auf der unteren Ebene die analytische Perspektive im Vordergrund (wie müssen Daten aufbereitet und ausgewertet werden?), geht es auf der zweiten Ebene darum, aus einzelnen Programmen Bausteine zu entwickeln, die in verschiedenen Kontexten und Kombinationen wiederverwendet werden können.

Mit Hilfe der Makrosprache ist es unter anderem möglich:

- SAS Systeminformationen abzufragen
- Beschriftungen zu flexibilisieren
- Die Ausführung von Daten- oder Prozedurschritten an Bedingungen zu knüpfen
- Programmteile in Schleifen zu fassen

In dem Tutorium werden die Grundlagen der Makrosprache vorgestellt und anhand von Beispielen demonstriert. Es richtet sich an SAS-Anwender, die bereits über Erfahrungen mit SAS/BASE verfügen.

## **Kurzfassungen Vorträge**

*Donnerstag, 8. März 2012, 10:00-11:00, Hörsaalzentrum HSZ 02*

### **Begrüßung der Konferenzteilnehmer, Eröffnung der 16. KSFE und Eröffnungsvortrag:**

#### **Speeding Up Exploration**

**John Sall**

*Co-Founder und Executive Vice President SAS*

*Donnerstag, 8. März 2012, 14:00-14:30, Hörsaalzentrum HSZ 02*

### **Einsatz des neuen Text Miner 5.1 zur Nachrichtenklassifikation von Online Medien**

**Ulrich Reincke**

*SAS Deutschland*

*ulrich.reincke@ger.sas.com*

Im Vortrag werden die wichtigsten Neuerungen des SAS Text Miner 5.1 vorgestellt und in einer Fallstudie zur Nachrichtenklassifikation genutzt. In der Fallstudie wird zunächst die Textdatenaufbereitung über die SAS Internet Crawler Technologie in Kombination mit Google vorgestellt, um eine Basistabelle für die Text Mining Analysen aufzubauen. Dann werden anhand der Prozessschritte in der Text Mining Analyse zur Nachrichtenklassifikation die neuen Analyseknotten des neuen SAS Text Miners 5.1 vorgestellt und eingesetzt. Hierzu gehören die Analyseknotten:

- Text Parsing
- Text Filter
- Text Topic
- Text Cluster
- Segment Profiler

*Donnerstag, 8. März 2012, 14:30-15:00, Hörsaalzentrum HSZ 02*

### **Social Media Analysis in Twitter**

**Kai Heinrich**

*Technische Universität Dresden*

*kai.heinrich@tu-dresden.de*

Die Analyse von Kundenmeinungen über Produkte und Technologien ist ein absoluter Schlüsselfaktor für Unternehmen. Schnelle Reaktionen auf Datenanalysen, welche aus Echtzeitmeinungsbildern gewonnen werden, können einem Unternehmen einen erheblichen Wettbewerbsvorteil verschaffen. Im Rahmen der evolutionären Entwicklung von Web 2.0 Anwendungen können die steigenden Nutzerzahlen solcher Social Media Anwendungen und die darin generierten Meinungsbilder nicht ignoriert werden. Besonders Web und Microblogs erfreuen sich großer Beliebtheit unter den Nutzern. Microblogs, wie z.B. Twitter ermöglichen die Kommunikation über kurze Statustexte im Rahmen eines sozialen Netzwerks.

Im Zuge des Vortrages sollen folgende Fragen geklärt werden:

- Eignet sich Twitter zur Analyse von Kundenmeinungen

- Welche Preprocessing-Maßnahmen sind notwendig um die Twitter Daten auswerten zu können
- Wie können die Social Media Komponenten von SAS bei der Analyse unterstützen

*Donnerstag, 8. März 2012, 11:30-12:00, Hörsaalzentrum HSZ 03*

### **SAS OnDemand for Academics - Erste Erfahrungsberichte**

**Katja Deutsch<sup>1</sup>, Steffen Marquardt<sup>2</sup>**

<sup>1</sup>*Universität Koblenz-Lindau*

<sup>2</sup>*Hochschule Fulda*

*katja.deutsch@altran.com, s.marquardt@accantec.de*

An der Uni Koblenz-Lindau und der Hochschule Fulda wurden im abgelaufenen Sommer Semester 2011 der Enterprise Miner bzw. der Enterprise Guide im Rahmen des "SAS OnDemand for Academics" Programms eingesetzt. Im Vortrag werden die Grundlagen und Prozesse des Programms erläutert sowie die Erfahrungen aus Koblenz und Fulda dargestellt. Dabei wird sowohl auf die Hochschul/Dozenten- als auch Studentensicht eingegangen. Ziel des Vortrags ist es, interessierten Hochschulen umfassende Informationen über das Programm aus Anwendersicht zu geben.

*Donnerstag, 8. März 2012, 12:00-12:30, Hörsaalzentrum HSZ 03*

### **SAS Enterprise Miner in der Lehre – von der Einführung zur Zertifizierung**

**Andreas Schieber**

*Technische Universität Dresden*

*andreas.schieber@tu-dresden.de*

An der Professur für Wirtschaftsinformatik, insb. Informationssysteme im Dienstleistungsbereich, fand im Wintersemester 2010/2011 erstmalig eine Zertifizierung für die Software SAS Enterprise Miner 6.1 statt. Bereits im Sommersemester 2010 sind die Studenten im Rahmen der Vorlesung „Business Intelligence, insb. Data Mining“ sowie der begleitenden Rechnerübung „Data Mining“ mit der Software in Berührung bekommen. Prof. Hilbert vermittelte den Studenten während der Vorlesung theoretische Grundlagen zu Vorgehen und Analysemöglichkeiten im Rahmen von Data-Mining-Projekten. In der Rechnerübung unter Leitung von Herrn Schieber wendeten die Studenten ihr erlangtes Wissen mit Hilfe des SAS Enterprise Miners praktisch an. Die Kenntnisse wurden am Ende des Semesters sowohl durch eine Klausur über den theoretischen Vorlesungsteil, als auch durch eine Prüfung zur Rechnerübung nachgewiesen. Diejenigen Studenten, die die Rechnerprüfung mit mindestens guter Note bestanden hatten, wurden am Anfang des Wintersemesters von der Professur zu dem Seminar „Advanced Business Analytics with SAS Enterprise Miner 6.1“ eingeladen, an dessen Ende die Zertifizierung stand. In Vorbereitung auf die Zertifizierung wurde den Studenten der offizielle SAS-Schulungsordner bereitgestellt. Nach dem selbständigen Erarbeiten der Ordnerinhalte und -aufgaben wurden an zwei Blockterminen im PC-Pool der Fakultät besondere Teilaspekte aus dem Themenbereich vertieft und ausstehende Fragen geklärt.

Die Professur für Wirtschaftsinformatik, insb. Informationssysteme im Dienstleistungsbereich, unter Leitung von Prof. Hilbert, kooperiert seit vielen Jahren mit dem Softwarehersteller SAS und ermöglichte so als eine der ersten Universitäten in Deutschland eine entsprechende Zertifizierung für die Studenten. Aufgrund des Erfolgs der Zertifizierung – auch im Vergleich mit anderen Universitäten – wird sie von der Professur zukünftig regelmäßig angeboten.

Donnerstag, 8. März 2012, 12:30-13:00, Hörsaalzentrum HSZ 03

## **SAS-Know-how in der Organisation entwickeln und pflegen - Ressourcen für die Ausbildung und für das Wissensmanagement**

**Andreas Mangold**

*HMS Analytical Software GmbH Heidelberg  
andreas.mangold@analytical-software.de*

So mächtig die SAS-Software auch ist - ohne Anwender und Programmierer mit entsprechender Ausbildung und Erfahrung gibt sie uns keine Ergebnisse. Dazu kommt, dass die Software ständig weiterentwickelt wird, ebenso die Schnittstellen zu anderen Produkten. Technische Fähigkeiten sind in der Regel nicht ausreichend. Kenntnisse von Methoden (z.B. Statistik oder Software Engineering) und des jeweiligen fachlichen Einsatzgebietes sind ebenfalls erforderlich. Daher stellt sich für Organisationen, die SAS (und nebenher meist noch eine Reihe weiterer Softwareprodukte) im betriebswirtschaftlichen Bereich oder in Forschung und Entwicklung einsetzen, die Frage, wie das erforderliche Wissen systematisch entwickelt und gepflegt werden kann. Folgende Themen werden in diesem Vortrag angesprochen:

- Dimensionen des Wissens und Könnens: Technisch, anwendungsbezogen, methodisch und sozial
- Ausbildungsmöglichkeiten für SAS-Technologie in Deutschland an Hochschulen, bei SAS Institute und bei anderen Unternehmen
- Know-how-Ressourcen für SAS aus dem Internet für die tägliche Arbeit: Online-Hilfe, Blogs und Communities
- Systematisches Wissensmanagement in der Organisation: Unser Ansatz besteht aus Wissenszielen und Wissensprofilen für die Mitarbeiter, Wissenszirkel für die interne Schulung, Schulungen und Zertifizierungen bei externen Anbietern sowie einem unterstützenden System auf Basis von Semantic Mediawiki

Donnerstag, 8. März 2012, 14:00-14:30, Hörsaalzentrum HSZ 03

## **Tipps zur SAS-Programmierung in der PIAFStat-Umgebung**

**Andrea Zenk, Volker Michel**

*Landesforschungsanstalt für Landwirtschaft und Fischerei Mecklenburg-Vorpommern, Gülzow  
a.zenk@lfa.mvnet.de, v.michel@lfa.mvnet.de*

PIAFStat als ein Modul des PIAF-Systems [2] verwaltet in einer Verfahrensbibliothek SASCode für die Datenanalyse und biostatistische Auswertung von landwirtschaftlichen Feldversuchen und Versuchsserien [1]. Zur Erstellung von PIAFStat-Verfahren sind umfassende Kenntnisse sowohl im PIAFStat-Programm als auch in der SAS-Welt notwendig. In diesem Beitrag wird über Erfahrungen zur SAS-Programmierung für PIAFStat berichtet. Es wird erläutert, wie ein ausgewogener Grad von Automatismus und Flexibilität im Verfahren erreicht werden kann. Daneben werden Tipps zu SAS direkt gegeben. So werden unter anderem folgende Fragen beantwortet:

- Wie kann ein PIAFStat-Verfahren in SAS direkt programmiert und getestet werden?
- Wie können Ergebnisse gebündelt abgespeichert werden?
- Wie ist auf einfachste Art das Überschreiben von Ergebnis-Dateien zu verhindern?
- Wie können Systeminformationen einbezogen werden?
- Wie kann das Zusammenspiel von SAS z.B. mit einer ACCESS-Datenbank gestaltet werden?

Zur Veranschaulichung dieser und weiterer Tipps wird ein aktuelles PIAFStat-Verfahren [3] genutzt.

### Literatur:

- [1] Michel, V., Schmidtke, J. und Zenk, A. (2008): Die Bibliothek biometrischer Verfahren in PIAFStat. 2008, [http://www.landwirtschaftmv.de/cms2/LFA\\_prod/LFA/content/de/Fachinformationen/Sorten/Beitraege/?&seite=2&artikel=2301](http://www.landwirtschaftmv.de/cms2/LFA_prod/LFA/content/de/Fachinformationen/Sorten/Beitraege/?&seite=2&artikel=2301)
- [2] Michel, V., Zenk, A., (2007): Konzipierung, Entwicklung und Einführung von PIAF. Mitteilungen LFA 37, 2007, 59-60.
- [3] Zenk, A. (2011): PIAFStat-Verfahren Demo 2011. Computer Program

Donnerstag, 8. März 2012, 14:30-15:00, Hörsaalzentrum HSZ 03

## **Erstellen flexibler Makros am Beispiel einer SQL-Routine**

### **Andreas Deckert**

*Institute of Public Health, Heidelberg*

*a.deckert@uni-heidelberg.de*

Auf der letzten KSFE wurde vom Autor eine verbesserte SQL-basierte Methode zum Matching von Fällen und Kontrollen in epidemiologischen Fall-Kontroll-Studien vorgestellt. Diese Methode wurde mittlerweile in ein Makro übersetzt. Kernstück dieses Makros ist eine SQL-Routine, die situations-flexibel programmiert wurde. Je nach Anzahl und Art der angegebenen Variablen generiert der SAS-Interpreter dabei die jeweils notwendige Anzahl von Anweisungen im SELECT- und im WHERE-Statement während der Laufzeit selbst. Der Vortrag erläutert kurz den Hintergrund des Matchings und die Funktionsweise der SQL-Anweisung. Danach wird ausgehend von einem konkret definierten Anwendungsfall Schritt für Schritt ein Makro entwickelt, das mehrere verschiedene Anwendungsmöglichkeiten benutzerfreundlich bearbeiten kann.

### Literatur:

- [1] Kawabata H et.al. Using SAS to Match Cases for Case Control Studies. SUGI 29, Paper 173- 29
- [2] Deckert A. 1:N Matching von Fällen und Kontrollen: Propensity Score vs. PROC SQL. KSFE 2011 - Proceedings der 11. Konferenz der SAS-Anwender in Forschung und Entwicklung.

Donnerstag, 8. März 2012, 16:00-16:30, Hörsaalzentrum HSZ 03

## **SAS Backstage**

### **Biljana Gigic<sup>1</sup>, Andreas Deckert<sup>2</sup>**

<sup>1</sup>*Nationales Centrum für Tumorerkrankungen, Heidelberg*

<sup>2</sup>*Institute of Public Health, Heidelberg*

*biljana.gigic@nct-heidelberg.de, a.deckert@uni-heidelberg.de*

In diesem Vortrag sollen die Hintergrundprozesse in SAS während einfacher Input- und Merge-Anweisungen beleuchtet werden. Das Verständnis dieser Prozesse erleichtert den richtigen Umgang beim Einlesen von Rohdaten und Zusammenführen von Tabellen. Anhand von Beispielen soll konkret die Funktion der Buffer-Verwaltung und des Program Data Vector (PDV) beschrieben werden.

Erläuterung: Ein SAS DATA Step wird grundsätzlich in zwei aufeinanderfolgenden Phasen exekutiert, die Kompilierung und Ausführung. Neben einer Syntaxprüfung wird während der Kompilierungsphase der Input Buffer, der PDV und die Dateibeschreibung generiert. In der Ausführungsphase wird die Output-Datei erzeugt. Dabei verwaltet der Input Buffer den zu verarbeitenden Datensatz. Die einzelnen Beobachtungen werden im PDV gespeichert und nach jedem Durchgang des DATA Steps in die Output-Datei transferiert.

Literatur:

[1] "SAS 9.2 Language Reference: Concepts, Second Edition", SAS Publishing, 2010 <http://www.urz.uni-heidelberg.de/statistik/kurse/online/kurs6.html> (letzter Aufruf 29.09.2011)

Donnerstag, 8. März 2012, 16:30-17:00, Hörsaalzentrum HSZ 03

**Tipps, Tricks und Stolperfallen zum Umwandeln von Variablen****Sabine Erbslöh**

Accovion GmbH, Abteilung Statistical Programming, Eschborn  
[sabine.erbsloeh@accovion.com](mailto:sabine.erbsloeh@accovion.com)

Bei der Arbeit mit Daten aus klinischen Studien müssen die Rohdaten für eine Auswertung erst entsprechend aufbereitet werden. Dazu sind oft Umwandlungen nötig, z.B. von numerisch in alphanumerisch oder von mehreren einzelnen Variablen in eine Datumsvariable. Bei der Weiterbearbeitung erlebt man dann manche Überraschung, z.B. wenn man nicht daran gedacht hat, die Länge einer alphanumerischen Variablen vorher zu definieren, wenn die Rohdaten nicht immer so aussehen wie erwartet, oder wenn es durch Standardeinstellungen beim Umwandeln zu unerwarteten Leerzeichen oder gar zum Abschneiden von Zeichen kommt.

Oft gibt es auch ganz spezielle Vorgaben, die für einen Datensatz erfüllt sein sollen. Dann ist es hilfreich zu wissen, wie man beispielsweise alle Formate oder Label entfernt oder Variablen initialisiert, ohne sich Gedanken machen zu müssen, ob sie numerisch oder alphanumerisch sind. Auch die Reihenfolge der Variablen in einem Datensatz kann wichtig sein, sei es mit führenden Identifizierungsmerkmalen oder in alphabetischer Reihenfolge. In diesem Beitrag werden Beispiele zur Umwandlung von Variablen unter Verwendung von Funktionen, Formaten und Prozeduren vorgestellt. Dabei wird auch speziell auf Stolperfallen eingegangen.

Donnerstag, 8. März 2012, 17:00-17:30, Hörsaalzentrum HSZ 03

**Proc Transpose oder Do-it-yourself****Dr Heribert Ramroth**

Institut für Public Health, Heidelberg  
[heribert.ramroth@uni-heidelberg.de](mailto:heribert.ramroth@uni-heidelberg.de)

Bei der Datenauswertung taucht regelmäßig die Situation auf, Daten in eine andere Struktur überführen zu müssen, als diese aktuell vorhanden sind. Diese Situation tritt z.B. bei Anwendung eines Makros mit vorgegebener Datenstruktur auf, oder aber wenn die Ergebnisse in einer ansprechenden Form aufbereitet werden sollen. Eine nützliche Prozedur ist hier PROC TRANSPOSE.

Folgende Fragen stellen sich für das Datenmanagement:

- bietet PROC TRANSPOSE eine schnell zu realisierende Lösung an;
- kann es sinnvoller sein, die Daten selbst in mehreren, möglicherweise übersichtlicheren , Datenschriften umzuordnen,
- oder c) gibt es sinnvolle Alternativen wie PROC SQL (vielleicht sogar PROC IML)?

Der Beitrag bietet anhand von Beispielen eine Übersicht über die Möglichkeiten von PROC TRANSPOSE, zeigt aber auch Vor- und Nachteile möglicher Alternativen auf.

Literatur:

[1] SAS Help

Donnerstag, 8. März 2012, 11:30-12:00, Hörsaalzentrum HSZ 04

### **Hierarchische Bayes'sche Netze: Eine Methode, um Dokumentensammlungen in den hierarchischen Strukturen zu klassifizieren - so, wie ein Mensch die Texte schreibt.**

**Ulrich Reincke**

*SAS Deutschland*

*ulrich.reincke@ger.sas.com*

In diesem Vortrag wird mit der Prozedur TMBelief ein neues Text Mining Verfahren vorgestellt, um den Prozess der semantischen Erschließung und Klassifikation zu automatisieren und so einen besseren Zugang zu den im Text enthaltenen Informationen zu ermöglichen. Bei den Bayes'schen Belief Netzen handelt es sich um ein Verfahren der überwachten statistischen Relationserkennung. Dabei kann auf die Grundstruktur der Texte und ihre Bausteine in der Analyse zugegriffen werden. Die Analyse kann also die Strukturen nutzen die ein Verfasser beim Schreiben aufbaut, d.h. Worte, Sätze, Absätze, Abschnitte, Kapitel, Klassifikationsthemen. Eine vom Nutzer ausgewählte Teilmenge der gefundenen hoch gewichteten Polaritätsworte und ihre Relationen zu den Zielvariablen kann anschließend zum Aufbau eines Klassifikationsschemas oder einer Ontologie genutzt werden um die Inhalte der Sammlung zu repräsentieren. Die vielseitigen Möglichkeiten der Prozedur TMBelief werden in einer konkreten Fallstudie anschaulich vorgeführt. Ihr Anwendungsfeld geht weit über die Analyse von Texten hinaus.

Donnerstag, 8. März 2012, 12:00-12:30, Hörsaalzentrum HSZ 04

### **Bestimmung der optimalen Fallzahl in Dosis-Findungs-Studien mit aktiver Kontrolle und linearen sowie nichtlinearen Dosis-Wirkungs-Profilen.**

**Hans-Joachim Helms, Tim Friede**

*Universität Göttingen, Abteilung Medizinische Statistik*

*hans-joachim.helms@med.uni-goettingen.de, tim.friede@med.uni-goettingen.de*

In einer Dosis-Findungs-Studie mit aktiver Kontrolle wird ein neuartiges Medikament in verschiedenen Dosis-Stufen mit einem bereits bekannt Medikament (der aktiven Kontrolle) verglichen. Ziel dieser Studien ist das Finden der Dosis-Konzentration  $d$ , welche zum gleichen Wirkungsgrad führt, wie die aktive Kontrolle. Dafür muss der Schnittpunkt der linearen oder nichtlinearen Dosis-Wirkungskurve mit dem Wirkungsgrad der aktiven Kontrolle berechnet werden.

Die Aufgabe des in SAS/IML erstellten Programmes ist das Finden einer optimalen Fallzahl, so dass das Konfidenzintervall der gesuchten Dosis-Konzentration  $d$  eine vorgegebene Breite in  $\gamma$ -% der Fälle nicht überschreitet.

Zuerst soll die Problemstellung erläutert und anschließend die bisherigen Ergebnisse vorgestellt werden. Abschließend wird ein Ausblick auf die weitere Entwicklung und bestehende Probleme gegeben.

#### Literatur:

- [1] Bretz, F. et al, 2008. Practical considerations for optimal designs in clinical dose finding studies. *Statistics in Medicine*
- [2] Fisch, R. und Stehlau, G., 2008. A Simplified Approach to Calibration Confidence Sets. *The American Statistician*

Donnerstag, 8. März 2012, 12:30-13:00, Hörsaalzentrum HSZ 04

## Randomisierungsverfahren mit SAS

Jäger, B.P.<sup>1</sup>, Malsch, C.<sup>1</sup>, Rudolph, P.E.<sup>2</sup>, Biebler, K.-E.<sup>1</sup>

<sup>1</sup>Ernst-Moritz-Arndt-Universität Greifswald, Institut für Biometrie und Med. Informatik

<sup>2</sup>Leibniz-Institut für Nutztierbiologie (FBN) Dummerstorf, FB Genetik und Biometrie

[bjaeager@biometrie.uni-greifswald.de](mailto:bjaeager@biometrie.uni-greifswald.de), [malsch@biometrie.uni-greifswald.de](mailto:malsch@biometrie.uni-greifswald.de), [pe.rudolph@fbn-dummerstorf.de](mailto:pe.rudolph@fbn-dummerstorf.de),

[biebler@biometrie.uni-greifswald.de](mailto:biebler@biometrie.uni-greifswald.de)

Kontrollierte klinische Studien verwenden Randomisierungsverfahren, deren Aufgabe es ist, die vorhandenen Probanden auf z. B. eine Kontroll- und eine Behandlungsgruppe der Studie zu verteilen, ohne dass subjektive Einflüsse des Untersuchers wirksam werden. Ein wichtiges Nebenkriterium der Randomisierung ist die Balanciertheit, d. h., die Umfänge der Studiengruppen sollen gleich groß sein. Der Grund hierfür ist, dass Statistische Verfahren zum Vergleich von  $m$  Gruppen mit den Umfängen  $n_1, n_2, \dots, n_m$  in der Regel für den balancierten Fall die größte Power besitzen. Die gewöhnliche Randomisierung, für die (bei zwei Gruppen) das Werfen einer Münze Modell steht, führt nur ausnahmsweise zu Balanciertheit. Deshalb sind zahlreiche Randomisierungsverfahren entwickelt worden, um wenigstens näherungsweise Balanciertheit zu erreichen.

Für die Randomisierungsverfahren nach Wei (1978), Efron (1971) und Atkinson (1982), die Replacement-Randomisierungen von Pocock (1983) und von Abel (1987), die Blockrandomisierung und das Big Stick Design von Soares und Wu (1983), das Square Root Design und das Two Coin Design werden die Algorithmen kurz erläutert. Weiterhin wird ein SAS-Makro vorgestellt, mit dem Randomisierungen nach allen vorgestellten Verfahren und mit allen denkbaren Parametern realisierbar sind.

Mit diesem Makro wurde das erwartete Balanceverhalten mit einfachen Simulationsexperimenten geprüft. Untersuchungen zur Bewertung des Einflusses des Randomisierungsverfahrens auf die jeweils verwendeten statistischen Verfahren sind ebenfalls möglich. Verallgemeinerungen auf mehr als zwei Therapiearme und zusätzliche Prognosefaktoren können leicht angepasst werden.

### Literatur:

- [1] Abel, U. (1987): Modified replacement randomization. *Statistics in Medicine*, 6127-135
- [2] Atkinson, A.C. (1982): Optimum biased coin designs for sequential clinical trials with prognostic factors. *Biometrika* 69.
- [3] Efron, B. (1971): Forcing a sequential experiment to be balanced. *Biometrika* 58, 403-417 Kundt, G. (2002): Randomisierungsverfahren für kontrollierte klinische Studien, Shaker Verlag Aachen.
- [4] Nordle, O., Brantmark, B. (1977): A self-adjusting randomization plan for allocation of patients into two treatment groups. *Clinical Pharmacology and Therapeutics* 22, 825- 830.
- [5] Pocock, S. J. (1983): *Clinical trials: A practical Approach*. John Wiley & Sons.
- [6] SAS Institute Inc., (2004). *SAS/STAT 9.1 User Guide*. Cary, NC: SAS Institute Inc.
- [7] Soares, J. F.; Wu, C. F. J. (1983): Some restricted randomization rules in sequential designs. *Commun. Statist. Theor. Meth.* 17, 2017-2034.
- [8] Wei, L. J. (1978): The adaptive biased coin design for sequential experiments. *The Annals of Statistics* 6, 92-100.

Donnerstag, 8. März 2012, 14:00-14:30, Hörsaalzentrum HSZ 04

## Webservices und SAS - PROC SOAP, PROC HTTP und der ganze REST

Martin Haffner, Andreas Mangold

HMS Analytical Software GmbH, Heidelberg

[martin.haffner@analytical-software.de](mailto:martin.haffner@analytical-software.de), [andreas.mangold@analytical-software.de](mailto:andreas.mangold@analytical-software.de)

Keine Softwareanwendung ist so universell, dass sie für alle Aufgaben einsetzbar ist. Das Zusammenspiel vieler Softwaresysteme ist in der Wirtschaft vorherrschend und auch in der Forschung und Entwicklung häufig

nötig. Webservices bilden eine universelle Schnittstelle zwischen den Systemen, sozusagen eine Lingua franca, eine "gemeinsame Sprache" für den Aufruf von Funktionen zwischen verschiedenen Softwareanwendungen.

Für SAS-Anwender haben Webservices zwei Anwendungsgebiete: das Konsumieren und das Bereitstellen von Webservices. Zum einen können Webservices, die von anderen Anwendungen oder öffentlich im Internet bereitgestellt wurden, durch ein SAS-Programm konsumiert (aufgerufen) werden. Zum anderen kann man jedes SAS-Programm als Webservice bereitstellen und damit anderen Anwendungen den Aufruf ermöglichen. Dieser Vortrag konzentriert sich auf die Funktionen der SAS-Software, die das Konsumieren von Webservices ermöglichen, und zwar anhand eines Beispiels aus der Geographie.

Folgendes kommt zur Sprache:

- Kurze Einführung in die Bedeutung und die technischen Grundlagen von Webservices
- Konsumieren von SOAP-basierten Webservices mit PROC SOAP, mit den Data-Step-Funktionen SOAPxxx und mit der XML Libname Engine (XMLTYPE=WSDL)
- Konsumieren von REST-basierten Webservices mit PROC HTTP
- Zugriff auf die XML-Strukturen, die von Webservices benötigt und geliefert werden

Die Programmbeispiele werden veröffentlicht und können anschließend von jedem nachvollzogen werden. Zum Schluss wird noch kurz auf die SAS BI Webservices eingegangen, die es ermöglichen, Stored Processes als Webservices bereitzustellen.

### Literatur:

- [1] World Wide Web Consortium: Web Services Architecture
- [2] SAS Institute Inc.: SAS(R) 9.3 Functions and CALL Routines: Reference— SOAPWEB Function
- [3] SAS Institute Inc.: Base SAS(R) 9.3 Procedures Guide— SOAP Procedure
- [4] SAS Institute Inc.: Base SAS(R) 9.3 Procedures Guide— HTTP Procedure
- [5] SAS Institute Inc.: SAS(R) 9.2 XML LIBNAME Engine: User's Guide, Second Edition— LIBNAME Statement Syntax
- [6] SAS Institute Inc.: SAS 9.3 BI Web Services Developer's Guide
- [7] Andreas Adlichhammer: Vortrag auf der KSFE 2011 in Heidelberg— XML mit SAS leicht gemacht

*Donnerstag, 8. März 2012, 14:30-15:00, Hörsaalzentrum HSZ 04*

## **Das Beste aus zwei Welten - Aufruf von R-Funktionen mit PROC IML**

### **Dr. Peter Beyer**

*HMS Analytical Software GmbH, Heidelberg*

*peter.beyer@analytical-software.de*

Mit der Prozedur PROC IML stellt SAS eine leistungsfähige Schnittstelle zur Open Source Software R bereit. Dadurch ist eine einfache Nutzung von speziellen R-Funktionen, Graphiken und frei verfügbaren Zusatzpaketen aus SAS heraus möglich. Eben das Beste aus zwei Welten.

Die Zuhörer bekommen die Schnittstelle PROC IML zu R vorgestellt und erhalten Antworten auf die Fragen:

- Welche Installationsvoraussetzungen gibt es?
- Wie führt man R-Code aus?
- Auf welche Art übermittelt man SAS-Datasets nach R und Ergebnisse wieder zurück?
- Kann man mit R erzeugte Graphiken in SAS Reports einbetten?
- Wie werden Fehlermeldungen von R in SAS abfangen?
- Welche Einschränkungen sind mit der Schnittstelle verbunden?

Einfache Codebeispiele zu den einzelnen Fragestellungen vertiefen die Nutzbarkeit und die Anwendungsmöglichkeit dieser Schnittstelle.

Literatur:

- [1] SAS Institute Inc.: SAS/IML(R) 9.3 User's Guide - Calling Functions in the R Language

*Donnerstag, 8. März 2012, 16:00-17:00, Hörsaalzentrum HSZ 04*

## **Der Umgang mit fehlenden Werten in Verlaufsstudien im Falle von Drop-outs**

### **Benjamin Mayer**

*Institut für Epidemiologie und Medizinische Biometrie, Universität Ulm  
benjamin.mayer@uni-ulm.de*

Fehlende Werte sind nahezu allgegenwärtig in der Forschung [1]. Eines der Hauptprobleme fehlender Daten stellt die mögliche Verzerrung der interessierenden Parameter dar, die aus den Daten geschätzt werden sollen. Dies kann dann zu einer Verfälschung und Einschränkung der Aussagekraft von Untersuchungsergebnissen führen. Es gibt verschiedene Ansätze dem Problem fehlender Werte zu begegnen [1,4]. Dies wurde in dem Übersichtsvortrag [6] von Frau Steindorf auf der 15. KSFE in Heidelberg sehr schön dargelegt. Insbesondere im Rahmen von Longitudinalstudien ist die vollständige Erfassung aller relevanten Daten komplex und kann daher kaum realisiert werden. Es bedarf dabei zusätzlich den Umgang mit so genannten Drop-outs, die einen Spezialfall fehlender Werte darstellen. In diesem Vortrag sollen verschiedene Ansätze von Ersetzungsmethoden speziell für die Situation von Drop-outs in Verlaufsstudien vorgestellt werden [2,3]. Es wird diskutiert, welche Voraussetzungen für die korrekte Anwendung der einzelnen Methoden erfüllt sein müssen und auf welchen statistischen Eigenschaften sie basieren. Deren Performance soll anhand geeigneter SAS-Programme (PROC MI [5] bzw. spezielle Makros) jeweils veranschaulicht werden.

Literatur:

- [1] Allison PD (2001) Missing Data. Sage University Papers Series on Quantitative Applications in the Social Sciences 136. Thousand Oaks, CA: Sage
- [2] Mayer B (2011) Fehlende Werte in klinischen Verlaufsstudien— der Umgang mit Studienabbrechern. Dissertation an der Universität Ulm
- [3] Newman DA (2003) Longitudinal Modelling with Randomly and Systematically Missing Data: A Simulation of Ad Hoc, Maximum Likelihood and Multiple Imputation Techniques. Organizational Research Methods 6: 328-362
- [4] Rubin DB (1987) Multiple Imputation for Nonresponse in Surveys. John Wiley & Sons, New York
- [5] SAS Online-Dokumentation: PROC MI. URL: [http://support.sas.com/documentation/cdl/en/statug/63033/HTML/default/viewer.htm#mi\\_toc.htm](http://support.sas.com/documentation/cdl/en/statug/63033/HTML/default/viewer.htm#mi_toc.htm)
- [6] Steindorf K, Kuß O (2011) Multiple Imputation— der State-of-the-Art-Umgang mit fehlenden Werten. SAS-Wiki ([http://de.saswiki.org/images/c/c7/15.\\_KSFE\\_2011\\_-\\_Steindorf\\_-\\_Multiple\\_Imputation.pdf](http://de.saswiki.org/images/c/c7/15._KSFE_2011_-_Steindorf_-_Multiple_Imputation.pdf))

*Donnerstag, 8. März 2012, 17:00-17:30, Hörsaalzentrum HSZ 04*

## **Modellierung rechts-zensierter Zähldaten - ein SAS-Makro unter Nutzung von PROC NLMIXED**

### **Hannes-Friedrich Ulbrich**

*Bayer Pharma AG, Berlin  
hannesfriedrich.ulbrich@bayer.com*

In einem Tiermodell zum Metastasenbefall der Lunge werden die einzelnen Metastasen gezählt. Bei starkem Metastasenbefall ist davon auszugehen, dass sich wegen Zusammenwachsens die exakte Metastasenanzahl nicht ermitteln lässt, die Zählraten sind rechts-zensiert.

Chou und Steenhard (2009) beschreiben im SAS Global Forum ein von ihnen entwickeltes Makro mit den Worten "... flexible in allowing one to estimate a variety of count regression models including: zero-inflated, hurdle, censored, truncated, finite mixture, semi-parametric, squared polynomial expansion, and generalized heterogeneous". Leider ist das Makro nicht veröffentlicht.

Mit der Einführung von PROC COUNTREG (SAS/ETS 9.22) sind Poisson- und Negativbinomial-Regressionen und ihre 'zero-inflated' Pendanten ohne Zusatzmodul modellierbar, Regression mit zensierten Zählraten hingegen erfordern weiterhin eine separate Lösung.

Ein für rechtszensierte Zählraten entwickeltes SAS Makro wird vorgestellt und anhand eines Beispieldatensatzes werden die Bedeutung der Modellierung unter Anerkennung der Zensierungen erläutert. Mögliche Makro-Erweiterungen sowie potentiell weitere Anwendungsfälle werden diskutiert.

### Literatur:

- [1] Cameron AC, Trivedi PK. Regression Analysis of Count Data. Cambridge, ...: Cambridge University Press, 1998
- [2] Chou N-T, Steenhard D. A Flexible Count Data Regression Model Using SAS PROC NLMIXED.
- [3] SAS Global Forum 2009, paper 250
- [4] Hilbe JM. Negative Binomial Regression. Cambridge, ...: Cambridge University Press, 2007
- [5] Jung BC, Jhun M, Song SH. Testing for overdispersion in a censored Poisson regression model. *Statistics* 40: 533–543, 2006

*Freitag, 9. März 2012, 09:00-09:30, Hörsaalzentrum HSZ 02*

### **Die neue Style Engine in ODS – Cascading Style Sheets mit SAS 9.3**

**Gregor Hermann**

*SAS Deutschland*

*gregor.herrmann@ger.sas.com*

Mit SAS 9.3 können die Formatierungsmöglichkeiten mit Cascading Style Sheets erstmals nahezu vollständig genutzt werden. Grundlage dafür ist die völlig überarbeitete Style-Engine in ODS. Der Vortrag beschreibt die Änderungen und demonstriert deren Anwendung an praktischen Beispielen.

*Freitag, 9. März 2012, 09:30-10:00, Hörsaalzentrum HSZ 02*

### **SAS in der Praxis - Tipps und Tricks für SAS Web Report Studio**

**Andreas Bachert**

*HMS Analytical Software GmbH, Heidelberg*

*andreas.bachert@analytical-software.de*

SAS Web Report Studio, eine Benutzeroberfläche für die Berichterstellung im Webbrowser, bietet Anwendern ohne SAS-Kenntnisse vielfältige Möglichkeiten, Auswertungen in tabellarischer oder grafischer Form zu erstellen, zu speichern und zu publizieren. Web Report Studio verwendet sogenannte Information Maps, um auf die Daten zuzugreifen.

Dieser Vortrag erklärt kurz die Bedeutung und die wichtigsten Funktionen von SAS Web Report Studio und gibt dann Tipps für den praktischen Einsatz, z.B. für folgende Fragen:

- Vorgehensweise bei Änderungen an der zugrundeliegenden Information Map
- Was ist zu beachten, wenn eine Information Map umbenannt wurde bzw. an einen anderen Ort verschoben wurde?
- Performance von dynamischen Prompts steigern
- Berichte miteinander verlinken
- Unterschiedliche Berichtsabschnitte mit den gleichen Filterbedingungen versorgen
- Tipps zum Seitenlayout und zum Kopieren von Teilen von Reports
- Zellen in Abhängigkeit des Wertes einfärben
- Einsatz von Web Report Studio in verschiedenen Webbrowsern

Literatur:

[1] SAS Institute Inc.: SAS Products— SAS Web Report Studio

*Freitag, 9. März 2012, 10:00-10:30, Hörsaalzentrum HSZ 02*

## **Große Datenmengen knacken mit SAS High Performance Analytics**

**Martin Schütz**

SAS Deutschland

*martin.schuetz@ger.sas.com*

Höchste Performance für Datenabfragen wird schon seit Langem durch Daten-Partitionierung und Abfrage-Parallelisierung und seit Kurzem durch In-Memory-Speicherung adressiert. Über Datenabfragen hinaus eine Lösung für Höchste Performance echter Analytischer Verfahren wie Regressionen, Neuronale Netze, etc. bieten die entsprechenden Datenbanksysteme in der Regel nicht.

Mit SAS High Performance Analytics können größte Datenmengen mit komplexen Analytischen Verfahren in kürzester Zeit verarbeitet bzw. „geknackt“ werden.

*Freitag, 9. März 2012, 11:00-11:30, Hörsaalzentrum HSZ 02*

## **Implementation of ICH E3 based SAS reporting for a clinical trial – issues and solutions**

**Arne Ring, Ruth L. Coleman**

Diabetes Trials Unit, OCDEM, Oxford University, Oxford, OX3 7LJ, UK

*arne.ring@dtu.ox.ac.uk*

Introduction:

The Diabetes Trials Unit (DTU) conducts trials in diabetes and cardiovascular diseases ranging from proof of concept and efficacy studies to randomised multicentre and international controlled trials. The Statistics and Modelling Group (SMG) within the DTU provides support to these trials by offering advice and guidance in their design and conduct stages and by performing analyses, which are often based on SAS programming.

Objective:

These SMG activities have been carried out on an ongoing basis for many years, and have recently been expanded to support an ICH E3 conform clinical trials report within a few weeks after locking the data base. The trial was an exploratory Phase I trial in crossover design, that evaluated endpoints derived from repeated

measurements. Issues that arose during the implementation of the SAP and the output generation in SAS are described.

### Methods:

The SAP specified some MIXED effects models along with tabulated descriptions of group results as primary and secondary analyses. In addition, listings of derived endpoints were to be created. At the start of programming, all tables were categorised into different display types, aiming for transferability of program code across different parts of the analysis. In addition, derived data sets were created for the efficacy analysis to consolidate the structure of the data.

It was decided to use PROC REPORT for all outputs, despite the statisticians had limited experience with this procedure. To learn about PROC REPORT facilities, various papers from the SUGI websites (SAS Users Group International – now: SAS Global Users Group) that explained efficient ways of data listing and tabulation were used.

Due the time constraints during programming, some features of tabulation with PROC REPORT were not used. For example, some results were derived using PROC MEANS or PROC FREQ, and the outcome was only listed using PROC REPORT. This was done similarly for result presentation of the PROC MIXED. Hereby, the format of the output could be defined flexibly flexibly, but without the need of sophisticated methods of output generation (e.g. ODS TEMPLATES).

All displays were generated using ODS LISTING into TXT files. The major advantage of this approach was that all results (and TXT files for sections titles) could be automatically linked together to one file and converted into a PDF file, in which the analysis was presented using a monospace font.

This linking and conversion was supported by a (DOS-) batch file which also allowed efficient generation of the whole analysis output into a single output file. One part of the analyses was performed using R (as a dedicated package was available), and the resulting PDF output was linked into the clinical trial report PDF file.

One dedicated issue in SAS was the difficult usage of the PAGENO options, as this option does not allow outputting the page number after the creation of the output. Therefore, page numbering was finally performed directly within the PDF.

The program validation was planned using a risk based evaluation of the trial objectives and the delivery of results. The level of validation was specified similar to [1]. All of the validation had to be completed before distribution of the results.

### Results and conclusion

Overall, the time taken from the start of programming to the final readout of the study results was less than 9 weeks (with data base lock in-between). The main analysis comprised 130 pages of output based on about 25 display templates. Due to the limitation of time all outputs were based on stand-alone SAS programs without macro functionality.

The process of programming was finally reviewed to establish output macros (similar to [2]) that could be applied to create displays across trials more efficiently.

### Literatur:

- [1] C Meurer – 11. KSFE 2007. Validierung von SAS-Programmen für die Auswertung und Dokumentation klinischer Daten - Prozess, Umfang und Dokumentation der Validierung
- [2] E Arnold – 14. KSFE 2010. Eine Vorlage – zahlreiche Outputs: Entwicklung einer effizienten Strategie für die Studienauswertung.

Freitag, 9. März 2012, 11:30-12:00, Hörsaalzentrum HSZ 02

## Performante Erzeugung von Berichten in (vor)formatierten Excel-Dateien

**Stephan Frenzel<sup>1</sup>, Ralf Neumüller<sup>2</sup>, Sarah Baumert<sup>1</sup>, Dr. Robert Koob<sup>1</sup>**

<sup>1</sup>Kybeidos GmbH, Heidelberg

<sup>2</sup>Beyondbytes, Ober-Ramstadt

*stephan.frenzel@kybeidos.de; rn@beyondbytes.de; sarah.baumert@kybeidos.de; robert.koob@kybeidos.de*

Trotz hervorragender Möglichkeiten zur Datenpräsentation mit SAS werden Berichte von Fachabteilungen zumeist in Form einer formatierten Excel-Datei gewünscht, welche die Richtlinien des Corporate Designs im Unternehmen erfüllt und gegebenenfalls Excel-Formeln, -Diagramme und VBA-Makros enthält.

Microsoft Technologien eröffnen neue Möglichkeiten, direkt aus einer SAS/Base-Prozedur formatierte Excel-Dateien zu befüllen. Mittels Microsofts Data Access Objects und ActiveX werden SQL-Statements ausgeführt, die SAS-Daten in Excel-Arbeitsmappen übertragen. Der Zugriff erfolgt hierbei ausschließlich auf der Basis von Desktop-/Client-Funktionen. Durch diese Methode können 100% SAS-Funktionalität mit 100% Excel-Funktionalität kombiniert werden. Es gibt keine Einschränkungen - jeder denkbare Excel-Report auch mit mehreren Arbeitsblättern inklusive Diagrammen und Makros wird möglich.

Die Implementierung dieser Funktionalität in eine SAS/Base-Prozedur besticht daher durch folgende Vorzüge:

- Hohe Performance (Massenbericht-Erzeugung durch Batch-Verarbeitung)
- Weitreichende Funktionalität (SAS-Datenimport in formatierte Excel-Arbeitsmappen, die Diagramme, Formeln und VBA-Makros enthalten)
- Einfache Bedienung (native SAS-Prozedur)
- Schnelle Implementierung (SAS-Base und SQL Kenntnisse ausreichend)
- Lauffähig auch auf kleinen Client-Rechnersystemen

Freitag, 9. März 2012, 12:00-12:30, Hörsaalzentrum HSZ 02

## ODS Report Writing Interface - Neue Möglichkeiten der Berichtserstellung

**Sebastian Reimann**

*viadee Unternehmensberatung GmbH*

*sebastian.reimann@viadee.d*

Die Erstellung von pixel-genauen Auswertungen mit SAS Bordmitteln stellte bislang immer hohe Herausforderungen an den ambitionierten SAS Entwickler. Häufig lief dies auf Umsetzung mittels DATA \_NULL\_ Step und vielen PUT-Statements hinaus, um so die Ausgabe wunschgemäß zu erstellen.

Mit dem ODS Report Writing Interface stellt SAS über das Component Object Interface eine Methode zur objektgesteuerten Reporterstellung bereit. Hierbei können neben den ODS Features wie Farben, Bilder, proportionale Schriftarten und Unicode-Zeichen auch Positionierungsfunktionen zur pixel-genauen Platzierung von Objekten genutzt werden. Im Rahmen des Vortrags soll das Interesse geweckt werden, die neuen DATA Step Funktionen im produktiven Umfeld zu nutzen.

### Literatur:

[1] SAS: The Power To Show— Paper 313-2009

Freitag, 9. März 2012, 09:00-09:30, Hörsaalzentrum HSZ 03

### **JMP10: Sneak Preview für Entwicklung und Modellierung**

**Bernd Heinen**

*SAS Institute*

*bernd.heinen@jmp.com*

Anfang März 2012 wird die neue Version JMP10 zur Verfügung stehen. Mit dem "Application Builder" steht eine neue Entwicklungsumgebung zur Verfügung, die sowohl für die schnelle Entwicklung und Verallgemeinerung von erstellten Reports als auch für Programmierung komplexer Anwendung neue Möglichkeiten bietet. Gleichzeitig erleichtert die neue graphische Oberfläche die Gestaltung von Benutzerdialogen durch die Platzierung grafischer Elemente.

Die Methoden, die zur Berechnung statistischer Modelle genutzt werden, sind um Partial Least Squares erweitert worden. Weiterhin stehen zusätzliche nichtlineare Funktionen anwendungsbereit zur Verfügung.

Die genannten Funktionen werden an Hand von Beispielen demonstriert.

Freitag, 9. März 2012, 09:30-10:30, Hörsaalzentrum HSZ 03

### **Statistische Versuchsplanung mit JMP - von der Klassik zur Moderne**

**Meintrup, David**

*Hochschule für Angewandte Wissenschaften, Fakultät Maschinenbau, Ingolstadt*

*david.meintrup@haw-ingolstadt.de*

In der Statistischen Versuchsplanung (DoE - Design of Experiments) wird mit statistischen Methoden eine Serie von Einzelexperimenten aufgestellt, die den Versuchsraum optimal erfassen und eine effiziente Analyse der Versuchsergebnisse ermöglichen. So kann man für beliebige Anwendungsgebiete, in denen experimentell gearbeitet wird, Produkte und Prozesse verbessern, die Robustheit erhöhen und Kosten sparen.

In diesem Vortrag werden wir mit der Statistik-Software JMP, einer SAS-Lösung, deren Stärken in der Interaktivität und Visualisierung liegen, anhand realer Datenbeispiele Versuchspläne erstellen, analysieren und optimieren. Wir setzen dabei weder Vorkenntnisse im Umgang mit JMP noch in der Statistischen Versuchsplanung voraus.

Wir beginnen mit einem Screening-Design, das zur Auswahl der wichtigen Faktoren in einer komplexen, mehrfaktoriellen Versuchssituation dient. Anschließend zeigen wir mit Hilfe eines Wirkungsflächendesigns, wie mehrere Zielgrößen gleichzeitig optimiert werden können.

Die moderne Versuchsplanung geht mit Hilfe von optimalen Plänen über diese klassischen Ansätze hinaus. In JMP steht zur Erstellung von optimalen Versuchsplänen die sehr leistungsfähige und flexible Plattform "Custom Design" zur Verfügung. In unserem letzten Beispiel werden wir zeigen, wie diese Plattform verwendet werden kann, um Versuchspläne zu erstellen, die exakt auf die Bedürfnisse und Nebenbedingungen des jeweiligen Experiments zugeschnitten sind.

#### Literatur:

- [1] Goos, Jones: Optimal Design of Experiments, Wiley, 2011. Montgomery: Design and Analysis of Experiments, Wiley, 2008.

Freitag, 9. März 2012, 11:00-11:30, Hörsaalzentrum HSZ 03

## Ein Makro zur Berechnung von Diskriminanz- und Reklassifizierungsstatistiken für die Verbesserung eines Prädiktionsmodells bei Anwendung der Cox Regression.

Kristin Mühlenbruch<sup>1</sup>, Wolfgang Bernigau<sup>1</sup>, Lars Berglund<sup>2</sup>, Matthias Schulze<sup>1</sup>

<sup>1</sup>Deutsches Institut für Ernährungsforschung Potsdam-Rehbrücke, Abteilungen Molekulare Epidemiologie und Epidemiologie, Nuthetal

<sup>2</sup>Uppsala Clinical Research Center, Uppsala

kristin.muehlenbruch@dife.de, bernigau@dife.de, lars.berglund@ucr.uu.se, mschulze@dife.de

Die epidemiologische Forschung setzt sich mit der Identifizierung von Determinanten für bestimmte Erkrankungen, vor allem für sogenannte Zivilisationskrankheiten, auseinander. In diesem Zusammenhang hat sich ein Bereich entwickelt, der sich mit der Definition von Risiko-Prädiktionsmodellen auseinandersetzt. Risikoprädiktionsmodelle gibt es für verschiedene Erkrankungen, hier ist vor allem der Framingham Risk Score zu nennen, der für kardiovaskuläre Erkrankungen entwickelt wurde [1], und der Deutsche Diabetes Risiko Test [2]. Die Entwicklung solcher Prädiktionsmodelle ist aber oftmals nur der erste Schritt. Das Ziel ist das beste Vorhersagemodell zu finden und das initiale Prädiktionsmodell durch Hinzunahme von neuen Markern zu verbessern. Diese Verbesserung ist messbar. Zum einen dient die ROC (Receiver Operator Characteristic) Kurve und AUC (Area Under the Curve) zur Berechnung der diskriminativen Fähigkeiten des Prädiktionsmodells. Ein Vergleich zweier korrelierter ROC-Kurven wird mit Hilfe eines p-Werts für die Differenz ausgedrückt [3]. Diese von DeLong et al. (1988) entwickelte Methode ist bereits in die SAS Prozedur Proc Logistic integriert. Eine weitere Möglichkeit, die Verbesserung eines Modells zu berechnen, bieten sogenannte Reklassifizierungstatistiken. Pencina et al. (2008) schlagen hier vor allem Net Reclassification Improvement (NRI) und Integrated Discrimination Improvement (IDI) vor [4]. Zur Berechnung dieser Statistiken für den Vergleich zweier Prädiktionsmodelle wurde ein SAS Makro zur Verfügung gestellt, welches unter Verwendung berechneter Risiken diese Statistiken sowie Reklassifizierungstabellen ausgibt [5]. Bisher wurden die Risiken (Erkrankungswahrscheinlichkeiten) mit Proc Logistic berechnet und ROC-AUC Werte sowie ein Vergleich zweier ROC-Kurven war nicht in dieses Makro %nriidi integriert.

Das Ziel der Weiterentwicklung dieses Makros zu %reclassification\_phreg war, zum einen weiterhin Proc Logistic für die Berechnung und den Vergleich von zwei ROC-Kurven zu ermöglichen und die Berechnung der Reklassifizierungstatistiken auf der Basis von Survivaldaten zu übertragen. Hierfür wurde eine Cox Regression verwendet und die Risiken, die für die Berechnung des NRI und IDI benötigt werden, mit Proc Phreg berechnet. Das Makro soll nun die Möglichkeit bieten, die genannten Berechnungen auf unterschiedliche prospektive Studiendaten anzuwenden.

### Literatur:

- [1] Wilson, P.W., et al., Prediction of coronary heart disease using risk factor categories. *Circulation*, 1998. 97(18): p. 1837-47.
- [2] Schulze, M.B., et al., An accurate risk score based on anthropometric, dietary, and lifestyle factors to predict the development of type 2 diabetes. *Diabetes Care*, 2007. 30(3): p. 510-5.
- [3] DeLong, E.R., D.M. DeLong, and D.L. Clarke-Pearson, Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*, 1988. 44(3): p. 837-45.
- [4] Pencina, M.J., et al., Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Stat Med*, 2008. 27(2): p. 157-72; discussion 207-12.
- [5] Sundstrom, J., et al., Useful tests of usefulness of new risk factors: tools for assessing reclassification and discrimination. *Scand J Public Health*, 2011. 39(4): p. 439-41.

Freitag, 9. März 2012, 11:30-12:00, Hörsaalzentrum HSZ 03

### **Ein modulares Statistikmakro-Paket für direkte & indirekte Meta-Analysen**

**Dr. Hendrik Schmidt, Dr. Clemens Tilke**

*Boehringer Ingelheim Pharma GmbH & Co. KG, Global Biometrics & Clinical Applications, Ingelheim am Rhein  
hendrik.schmidt@boehringer-ingelheim.com, clemens.tilke@web.de*

Meta-Analysen klinischer Studien haben in den vergangenen Jahren stark an Bedeutung gewonnen. Die Ausbreitung und Anwendung meta-analytischer Methoden wurde und wird hierbei durch die Verfügbarkeit anwenderfreundlicher Software-Komplettlösungen und Makros in verschiedenen Programmiersprachen unterstützt.

Wir präsentieren eine modulare Designstruktur für ein SAS-Makropaket meta-analytischer Methoden inklusive Datenein- und ausgabe. Implementiert sind direkte Methoden wie die Inverse Varianzmethode, die Mantel-Haenszel-Methode und Peto's Ansatz sowie Methoden für indirekte Vergleiche, jeweils für unterschiedliche Effektmaße (z.B. risk ratio, odds ratio), sowie verschiedene Korrekturmöglichkeiten für Nullereignisse bei binären Daten. Graphische Darstellungsmöglichkeiten beinhalten u.a. den klassischen Forest-Plot.

Wegen der vielfältigen Analysemethoden und der sich gerade im Detail unterscheidenden graphischen Darstellungen erlaubt der modulare Designansatz eine klare und benutzerfreundliche Handhabung der Makros sowie deren übersichtliche Pflege. Darüber hinaus wird die Einbindung neuer oder erweiterter Meta-Analysemethoden stark erleichtert.

Freitag, 9. März 2012, 12:00-12:30, Hörsaalzentrum HSZ 03

### **Berechnung des relativen Überlebens auf Basis von Krebsregisterdaten mit SAS**

**Dr. Olaf Schoffer**

*Tumorepidemiologie im Universitäts KrebsCentrum Dresden Universitätsklinikum Carl Gustav Carus an der TU Dresden & Medizinische Fakultät Carl Gustav Carus der TU Dresden  
olaf.schoffer@uniklinikum-dresden.de*

Die Berechnung von Überlebendauerverteilungen ist eine essentielle Anforderung der epidemiologischen Forschung für Erkrankungen mit erhöhtem Sterberisiko. Neben der spezifischen Erkrankung beeinflusst die Altersstruktur der betrachteten Patientengruppe die Überlebendauerverteilung erheblich. Zur besseren Vergleichbarkeit wird der Einfluss der Altersstruktur insbesondere für Analysen auf Basis von Krebsregisterdaten häufig durch Verwendung relativer Überlebensraten eliminiert.

In der vorliegenden Präsentation wird die Methodik zur Überlebenszeitanalyse (Aktuarialschätzung, Kaplan-Meier-Verfahren, Breslow-Schätzer) sowie zur Berechnung relativer Überlebensraten (natürliche Mortalität, erwartetes Überleben, Ederer II) kurz vorgestellt.

Anschließend wird eine neukonzeptionierte Umsetzung in SAS beschrieben und dabei auf folgende Aspekte eingegangen:

- Basis: RSURV von R. Koch
- Verwendete Methodik: Breslow, Ederer II
- PROC LIFETEST (Neuerungen seit SAS 9.2)
- ODS (...GRAPHICS, ...OUTPUT, ...PDF/RTF)
- SQL (Bestimmung des kartesischen Produktes)
- Fortgeschrittene DATA-Step-Funktionalität (ARRAY, MERGE mit LAST.) Makro-Funktionalität

Diese Umsetzung wird zudem mit anderen bestehenden Lösungen hinsichtlich der Rahmendaten (Plattform, Methoden, Bedienung, Flexibilität) und der erzeugten Ergebnisse (absolute und relative Überlebenskurven) verglichen.

Literatur:

- [1] Allison, P. (2010): Survival Analysis Using SAS - A Practical Guide (2nd ed.), SAS Publishing, Cary, NC.16. KSFE 2012 Dresden 2/2
- [2] Dickman, P. (2004): Estimating and modelling relative survival using SAS, manual: [http://www.pauldickman.com/rsmode/sas\\_colon/](http://www.pauldickman.com/rsmode/sas_colon/).
- [3] Koch, R.: Analyse von Überlebensdauerverteilungen unter Berücksichtigung der Altersstruktur der Patienten in Klinischen Krebsregistern, in: Jahresbericht 2001 des Tumorzentrums Dresden e.V., 67-74.
- [4] SAS Institute Inc. (2010): SAS/STAT 9.22 User's Guide, SAS Publishing, Cary, NC.

*Freitag, 9. März 2012, 09:00-09:30, Hörsaalzentrum HSZ 04*

## **Erweiterungen der Analysemöglichkeiten im neuen Enterprise Miner 7.1 sowie im Model Manager**

**Martin Schütz**

SAS Deutschland

[martin.schuetz@ger.sas.com](mailto:martin.schuetz@ger.sas.com)

Im Vortrag werden die Highlights des neuen SAS Enterprise Miner 7.1 sowie die Integration in einer industrialisierten Data Mining Model Factory in Verbindung mit dem Model Manager vorgestellt. Zu den im Vortrag neuen Analysemethoden, die im Vortrag abgehandelt werden gehören:

- Survival Data Mining Analysis
- Rate Making for Insurance
- Time Series Data Mining (experimentell)
- Support Vector Machines (experimentell)
- Variablenselektionsmethode LARS für binäre Zielvariablen

*Freitag, 9. März 2012, 09:30-10:00, Hörsaalzentrum HSZ 04*

## **Einstieg in Text Analytics für SAS Enterprise Guide Anwender: Von der Datenquelle zum Bericht - und darüber hinaus...**

**Johannes Lang**

HMS Analytical Software GmbH, Heidelberg

[johannes.lang@analytical-software.de](mailto:johannes.lang@analytical-software.de)

Immer mehr Firmen erkennen das Potential ihrer unstrukturierten Daten in Textform, welche meist nicht in bestehende analytische Umgebungen (z.B. Data Warehouse) integriert sind.

Mit Hilfe von Text-Analytics-Verfahren ist es möglich, wertvolle Informationen aus Kunden-E-Mails, Call-Center-Notizen oder Vertragsdokumenten automatisch aufzubereiten und in eine Analyseplattform zu integrieren - auf Wunsch mit Berücksichtigung von Informationen aus dem Web.

Allein - wie fängt man damit an?

Um den Einstieg für Interessierte zu erleichtern, wird anhand des intuitiven Anwenderwerkzeugs SAS Enterprise Guide gezeigt, wie man unstrukturierte Daten einliest und vorverarbeitet, so dass anschließend einfache tabellarische und grafische Auswertungen möglich sind. Es wird auch gezeigt, wie fehlende Funktionalität in Form von externen Java-Bibliotheken eingebunden kann.

Behandelte Schritte:

- Definition der Datenquelle(n)
- Vereinzelung der Textdaten (Tokenisierung)
- Einbindung von Stop- und Startwortlisten zur Datenreduktion
- Wortstammreduktion (Stemming)
- Erstellung eines Keywords-in-Context Berichts
- Erzeugung eines Top-50-Keywords Diagramms
- Export der Ergebnisse

Auf spezialisierte Tools (SAS Text Miner, SAS Enterprise Categorization Studio) wird ebenfalls hingewiesen, die dem geeigneten Anwender weitere Möglichkeiten zur Datenanalyse bieten.

*Freitag, 9. März 2012, 11:00-11:30, Hörsaalzentrum HSZ 04*

### **Verwaltung eines Metadatenrepositories im DI 3.4 mittels einer externen Exceldatei**

#### **Dubravko Dolic**

*dsquare.de für Santander (Mönchengladbach)*

*dubravko.dolic@dsquare.de*

Im Unternehmen wird das DI Studio in der Version 3.4 eingesetzt, um ETL Strecken zu modellieren, während ansonsten noch viel SAS Base Code eingesetzt wird. Zur Implementation eines Datenmodells wurde das Modell in einem Metadatenrepository angelegt. Auf Grund der hausinternen Prozesse unterliegt das Datenmodell teilweise Änderungen. Diese müssen über eine Excel Datei verwaltet werden. Damit die Änderungen aus der Excel Tabelle im Metadatenrepository nachgefahren werden können, wird in SAS Base eine Verwaltung der Metadatenobjekte realisiert. Die Verwaltung entnimmt die Metadateninformationen aus der Excel Datei und ändert oder löscht die Informationen im Metadatenrepository.

Dieses Vorgehen ist ungewöhnlich und sicherlich in der Standardstruktur nicht so vorgesehen. Trotzdem kommen immer wieder Situationen vor, in denen direkt auf die Metadatenobjekte zugegriffen werden muss. Daher zeigt dieser Ansatz, wie auch der SAS Base Programmierer mit Metadaten umgehen kann.

Gliederung:

1. Kurzer Überblick Metadatenverwaltung im DI Studio/Metadatenserver
2. Problemaufriss: Datenmodell in Excel
3. Überblick über die Methoden zur Verwaltung von Metdadatenobjekten
4. Codebeispiele

Freitag, 9. März 2012, 11:30-12:00, Hörsaalzentrum HSZ 04

## Methoden der Risikobewertung

**Jörg Schmidtke**

*BioMath GmbH, Rostock*

*joerg.schmidtke@biomath.de*

In diesem Beitrag wird zunächst der Begriff des Risikos hinsichtlich der praktischen und der daraus resultierenden statistischen Fragestellung abgegrenzt. Im Weiteren werden die folgenden beiden Themenkomplexe behandelt:

### Theorie

Die statistische Grundlage für die Risikobewertung bilden Äquivalenz- und Signifikanztests in einem speziellen Versuchsdesign. Die Versuchsanlage, das mathematische Modell und die Auswertungsmethode solcher Experimente wird an einem praktischen Beispiel vorgestellt.

### Praxis

Alle Schritte der Risikobewertung werden an Hand von SAS-Code detailliert erläutert. Dabei wird stets der Bezug zum Auswertungsmodell hergestellt. Kernstück des SAS-Programms ist die Prozedur MIXED.

Dieses Vorgehen der Risikobewertung wird unter anderem im EU-Zulassungsverfahren von gentechnisch veränderten Pflanzen eingesetzt, welches als praktisches Beispiel in diesem Beitrag verwendet wird [1].

### Literatur:

[1] EFSA 2009. Scientific Opinion on Statistical considerations for the safety evaluation of GMOs. European Food Safety Authority, Parma.

Freitag, 9. März 2012, 12:00-12:30, Hörsaalzentrum HSZ 04

## Datenvisualisierung mit SAS Visual Analytics Explorer

**Gregor Hermann**

*SAS Deutschland*

SAS erweitert die BI Reporting Suite um eine intuitive, Web-basierte Benutzeroberfläche, die es Fachanwendern ermöglichen soll, bei der Analyse ihrer Daten ein Höchstmaß an Freiheitsgraden zu nutzen und neue Erkenntnisse zu gewinnen. Die Anwendung unterstützt zahlreiche unterschiedliche Datenquellen und bietet verschiedenste visuelle Optionen zur grafischen „Exploration“ der Daten, ergänzt durch analytische Funktionen von SAS.

Erleben sie den neuen SAS Visual Analytics Explorer live.

## Kurzfassungen Poster

Donnerstag, 8. März 2012, 15:00-15:30, Hörsaalzentrum

### **Laufzeitoptimierung bei der Verknüpfung großer Datenmengen. Ein Vergleich zwischen MERGE im DATA STEP und JOIN in PROC SQL**

**Cerstin Erler**

Institut für Arbeitsmarkt- und Berufsforschung der Bundesagentur für Arbeit (IAB), Nürnberg

cerstin.erler@iab.de

Mit einem zugrunde liegenden sehr großen Datensatz werden verschiedene Varianten des Zusammenspiels zweier Datensätze durchgeführt.

1. MERGE im DATA STEP mit und ohne Verwendung von Indizes
2. JOIN unter PROC SQL mit und ohne Verwendung von Indizes

Dieses Vorgehen wird für Datensätze unterschiedlicher Größe (Stichproben des Ausgangsmaterials) wiederholt und ist deshalb als MACRO geschrieben. Bei jedem MACRO-Lauf wird die Laufzeit gemessen und dokumentiert. Auf dem Poster wird anhand leicht lesbarer Diagramme das Ergebnis dargestellt. Ziel ist es, für das Datenmanagement mit großen und sehr großen Datensätzen aussagekräftige Handlungsempfehlungen abzuleiten, mit welchem Vorgehen (Prozeduren etc.) die Laufzeit optimiert werden kann und wie die Inputdaten (und Indizes) organisiert sein sollten, um diese schnellstmöglich verarbeiten zu können.

Daten: Integrierte Erwerbsbiographien mit 338 GB Daten und 72 GB Index (1.752.424.787 Zeilen und 39 Spalten). Es handelt sich um einen Datensatz, der Forschungsfragen der Volkswirtschaft und Sozialwissenschaft beantworten kann.

Software: SAS Enterprise Guide 4.2 unter Windows Server 2003 mit 4 Prozessoren und 16 GB RAM sowie SAS Base 9.2 unter Unix mit 4 Prozessoren und 32 GB RAM

#### Literatur:

- [1] SAS Online Documentation SAS Base 9.2

## **Ein SAS Macro zur Auswertung von SAE-Daten**

**Thomas Bruckner, Andreas Deckert**

Institut für Medizinische Biometrie und Informatik Universität Heidelberg

bruckner@imbi.uni-heidelberg.de, deckert@imbi.uni-heidelberg.de

Die Auswertung von Daten zur Sicherheit einer medizinischen Intervention beinhalten statistische Methoden, die in bei der Auswertung der Effizienz einer Therapie oder Intervention nicht in Betracht gezogen werden. Diese Methoden beinhalten spezielle Schätzer für Raten (z.B. "exposure adjusted incidence rates") und auch graphische Methoden zur Darstellung der Entwicklung der unerwünschten Ereignisse im Studienverlauf (mean cumulative function - MCF). Das vorgestellte SAS-Macro stellt diese Methoden zur Verfügung. Am Beispiel unerwünschter Ereignisse einer chirurgischen Studie werden diese statistischen Mittel dargestellt.

#### Literatur:

- [1] Siddiqui O: Statistical Methods to Analyse Adverse Events Data of Randomized Clinical Trials. J Biopharm Stat; 19889-899, 2009
- [2] Diener MK, et al: Efficacy of stapler versus hand-sewn closure after distal pancreatectomy (DISPACT): a randomised, controlled multicentre trial. Lancet; 377(9776):1514-22, 2011

## **SAS Macro zur Kombination von möglichen in Zusammenhang stehenden Faktoren bezüglich einer Zielgröße beschrieben anhand eines aktuellen Beispieldatensatzes**

**Martina Malzer<sup>1</sup>, Annette Pfahlberg<sup>1</sup>, Wolfgang Uter<sup>1</sup>, Janice Hegewald<sup>2</sup>, Olaf Gefeller<sup>1</sup>**

<sup>1</sup>Institut für Medizininformatik, Biometrie und Epidemiologie, Universität Erlangen-Nürnberg;

<sup>2</sup>Institut und Poliklinik für Arbeits- und Sozialmedizin, Technische Universität Dresden

martina.malzer@imbe.med.uni-erlangen.de

Die UV-Lichtempfindlichkeit der Haut steht direkt im Zusammenhang mit der Augenfarbe, Haarfarbe, den Sommersprossen, der Anzahl der Muttermale an beiden Armen und der Hauthelligkeit (in unserem Beispiel gemessen mittels Reflektometer). Die Empfindlichkeit der Haut kann in verschiedene Hauttypen eingeteilt werden (Fitzpatrick- Hauttyp). Allerdings ist der Hauttyp in einem selbst auszufüllenden Fragebogen schwer ermittelbar. Aus diesem Grund soll versucht werden, den Lichthauttyp auf der Basis anderer, reliabel messbarer konstitutioneller Merkmale zu bestimmen.

Es wurde ein automatisiertes SAS Macro verwendet, welches obige Merkmale mittels p-Wert basierendem CART geeignet miteinander kombiniert. Für die Realisierung einer p-Wert-basierten CART-Analyse innerhalb des automatisierten SAS Macros steht in SAS die FREQ-Prozedur zur Verfügung, durch diese die p-Werte errechnet werden. In einem nächsten Schritt werden die p-Werte der durchgeführten Tests miteinander verglichen, um so den kleinsten p-Wert für einen Split (Aufteilung) erhalten zu können. Daraus resultiert EIN binärer Klassifikationsbaum, in dem die oben genannten Merkmale bestmöglich miteinander kombiniert dargestellt werden, um die Zuordnung in einen Hauttyp erleichtern zu können.

Das Verfahren wird anhand eines Beispieldatensatzes über eine Befragung zum Thema Sonnenschutz / Verhalten in der Sonne bei jungen Erwachsenen illustriert.

### Literatur:

- [1] Uter W., Pfahlberg A., Kalina B., Kölmel K. F., Gefeller O., "Inter-relation between variables determining constitutional UV sensitivity in Caucasian children". *Photodermatol Photoimmunol Photomed*, 20: 9-13, Blackwell Munksgaard, 2004.
- [2] Fitzpatrick T. B., "The validity and practicality of sun-reactive skin types I through VI". *Arch Dermatol*, 124: 869-871, 1988.
- [3] Breiman L., Friedman J. H., Olshen R. A., Stone C. J., "Classification and Regression Trees". Wadsworth Inc., Monterey, California - USA, 1984.
- [4] Guggenmoos-Holzmann I., Wernecke K. D., "Medizinische Statistik". Blackwell Wissenschaftsverlag, Berlin - Wien, Oxford, Edinburgh, Boston, Melbourne, Paris, Yokohama, 1996.
- [5] SAS Institute Inc., SAS Macro Language: Reference, Version 8. SAS Institute Inc, Cary, NC - USA, 1999.

## **Medizinische Statistik mit ??? in SAS® 9.3 – ein Plädoyer für eine Windows-nahe Oberfläche**

**Rainer Muche, Benjamin Mayer**

Institut für Epidemiologie und Medizinische Biometrie, Universität Ulm

rainer.muche@uni-ulm.de

Im Studium der Humanmedizin wird an der Universität Ulm seit Jahren den Studierenden die Medizinstatistik / Biometrie auch anhand einer Statistiksoftware nahegebracht. Ziel ist dabei, nicht nur die Konzepte der Medizinstatistik zu lehren sondern auch die Umsetzung in Forschungsprojekten wie der Dissertation zu ermöglichen. Da über eine Landeslizenz SAS auf den PC-Pools der Universität und als Studierendenversion kostengünstig zur Verfügung steht war es naheliegend, eine auf SAS basierende maus- und menügesteuerte Oberfläche zu nutzen. Diese ist seit der Version 6.12 mit der SAS-Analyst-Oberfläche [5] angeboten und von uns entsprechend aufbereitet worden [1]. Diese Applikation ist von SAS zur Version 8 noch einmal gründlich überarbeitet worden, danach nicht mehr. Eine "revolutionäre Entwicklung" [4] sieht sicher anders aus. Dennoch

war uns wichtig, dass die Oberfläche der üblichen Maus - und Menüsteuerung der Windows-Applikationen angenähert ist, da so die Studierenden ihre entsprechenden Kenntnisse (MS-Excel, MS-Word) einsetzen können und sich "nur" um die spezifischen Einstellungen kümmern müssen. Außerdem haben wir in Fortbildungen den SAS- Analyst als Syntaxgenerator nutzen können, da alle Auswertungen als Syntaxskript ohne zusätzlichen Aufwand angesehen und abgespeichert werden können.

In SAS Version 9.2 kam dann beim Aufruf des SAS-Analyst folgende Meldung: "The Analyst application will no longer be available after SAS Release 9.2." Dementsprechend kann man jetzt in SAS 9.3 die Applikation nicht mehr finden. Allerdings fehlt in der Dokumentation zu SAS 9.3 (z.B. in "What's new in SAS/STAT 9.3") jeglicher Hinweis auf diesen Umstand und entsprechende Empfehlungen für Alternativen.

Seit einigen Jahren gibt es mit dem Enterprise Guide eine alternative Oberfläche in SAS. Diese grafische Oberfläche ist aber leider für unsere Studierenden nicht intuitiv einfach zu nutzen, da das Konzept sich doch sehr von dem der Maus- und Menüsteuerung der Office-Anwendungen unter Windows unterscheidet. Außerdem basiert die Programmierung der Syntaxsprache im Hintergrund hauptsächlich auf PROC SQL, welches für Medizinstudierende eine zusätzliche Hürde darstellt. Ein Einsatz der Oberfläche zur Einführung in SAS ist daher für uns im klinischen Umfeld nicht sinnvoll. In einem Nebenfach für Informatikstudierende mit Kenntnissen im Bereich Datenbanken wäre ein Einsatz diskutabel. Ein angedachter Ersatz des SAS-Analyst durch den Enterprise Guide haben wir deshalb nicht weiterverfolgt [2] sondern überlegen, auf die kostenfreie Oberfläche RExcel auszuweichen [3], bis SAS wieder eine adäquate Oberfläche anbietet.

### Literatur:

- [1] Muche R, Habel A, Rohlmann F (2000) Medizinische Statistik mit SAS-Analyst. Springer Verlag, Heidelberg
- [2] Muche R, Plank G, Babik T, Molz E, Orlogi S (2009) Medizinische Statistik mit dem SAS Enterprise Guide 4.1. Unveröffentlichtes Manuskript
- [3] Muche R, Lanzinger S, Rau M (2011) Medizinische Statistik mit R und Excel. Springer Verlag, Heidelberg
- [4] Ortseifen C (1998) Der SAS-Analyst— eine revolutionäre Entwicklung? <http://www.urz.uni-heidelberg.de/statistik/kurse/sastreff/bn-98-04-03/index.htm>
- [5] SAS Institute Inc (2003) The SAS-Analyst application, 2nd Edition. Sas Institute Inc., Cary NC [http://support.sas.com/documentation/onlinedoc/91pdf/sasdoc\\_91/stat\\_analystapp\\_7761.pdf](http://support.sas.com/documentation/onlinedoc/91pdf/sasdoc_91/stat_analystapp_7761.pdf)

## **DATA Step Component Interface – neue Objekte im DATA Step**

### **Sebastian Reimann**

viadee IT-Unternehmensberatung, Münster  
sebastian.reimann@viadee.de

Über das Component Interface erhält der klassische DATA Step neue Funktionalität. In diesem Vortrag wird gezeigt, wie das Hash-Objekt und der Hash-Iterator genutzt werden können, um effiziente Lookup-Methoden bei großen Tabellen zu ermöglichen. Weiterhin wird die Integration vorhandener Java-Businesslogik in SAS Programme gezeigt. Hierzu werden die Möglichkeiten des Javas Objekt im DATA Step näher analysiert.

### Literatur:

- [1] SAS 9.3 Component Objects: Reference Documentation

## **Nie wieder Post sortieren – vollautomatisierte Verarbeitung von Kundenpost mittels Text-Mining-Verfahren**

**Thomas Hemmler, Horst Meyer, Dr. Sergej Steinberg**

Bauer Systems KG - Abteilung Business Intelligence, Hamburg

thomas.hemmler@bauermedia.com, horst.meyer@bauermedia.com, sergej.steinberg@bauermedia.com

In der Praxis stellt die Verarbeitung von Kundenpost einen großen Zeitaufwand dar. Kann ein Computer die Anliegen der Kunden erkennen und den gesamten Prozess zufriedenstellend bearbeiten? Anders gesagt: Lässt sich die Verarbeitung der "in freien Form" geschriebenen Kundenpost vollständig automatisieren?

Im Poster wird gezeigt, dass es bei der Heterogenität der eingehenden Kundenpost immer wieder Fälle gibt, die eine gesonderte Abarbeitung durch eine Fachkraft erfordern. Auch wenn sich dieser Prozess nicht vollständig automatisieren lässt, so lassen sich doch Klassen von Dokumenten herausfiltern, die "gelesen", vollautomatisch dem richtigen Kunden zugeordnet, "verstanden" und "beantwortet" werden können.

Mit in SAS-Base programmierten Algorithmen wird gezeigt, wie man Dokumente klassifizieren und Informationen aus dem Text extrahieren und anhand bekannter Daten validieren kann. Ebenfalls wird gezeigt, wie mittels SAS ein geeignetes Qualitätsmanagement betrieben werden kann, indem man Test- und Kontrollgruppen (Test- und Kontrolldokumente) bildet.

Es stellt sich auch die Frage, was sagt uns die eingehende Post über unsere Kunden? Können wir das mit dem SAS Text-Mining aus den Briefen über die Kunden gewonnene "Wissen" für die Verbesserung der Geschäftsprozesse gewinnbringend einsetzen?

### Literatur:

[1] Weiss, S.: Text mining: predictive methods for analyzing unstructured information, New York; 2005

## **Anwendung eines SAS/STAT-Modells ohne SAS/STAT in einem CRM-Projekt**

**Tobias Otte, Timm Euler**

viadee IT-Unternehmensberatung, Münster

tobias.otte@viadee.de, timm.euler@viadee.de

Für eine Analyse der Kundenloyalität mit Data Mining bei einem Finanzdienstleister stand uns SAS/STAT nur während der Modellbildung zur Verfügung, aber nicht zur Modellanwendung. Daher wurde ein logistisches Regressionsmodell mit SAS/STAT erzeugt und dann mit SAS/BASE ausgelesen und zur Vorhersage eingesetzt. Das Poster berichtet vom Projekthintergrund, der technischen Umsetzung und den Ergebnissen aus CRM-Sicht.

## Organisatorische Hinweise

### Tagungsort Dresden – Stadt der Kultur, Wissenschaft & Technologie

Dresden lebt von seinem Weltruhm als Kunst- und Kulturstadt ersten Ranges. Für Semperoper, Zwinger, Staatstheater, Philharmonie und die heiligen Hallen der Gemäldegalerien und Museen sollten Besucher auf alle Fälle Sakko und Abendkleid mit im Gepäck haben.

Daneben ist Dresden einer der modernsten High-Tech-Standorte in Europa mit besonderen Kompetenzen in der Mikroelektronik, der Informations- und Kommunikationstechnik, der Biotechnologie und der Erforschung und Herstellung neuer Werkstoffe. Auch die Umgebung Dresdens hat ihre Reize und bietet viel Entdeckenswertes: Die Sächsische Schweiz ist als Ausflugsziel zum Wandern und Klettern ideal. In Meißen lassen sich die berühmte Porzellanmanufaktur und eine gediegene Altstadt bewundern. Für die kalte Jahreszeit bieten sich im Erzgebirge ausreichend Wintersportmöglichkeiten. Im näheren Umland Dresdens und in der Stadt selbst gibt es ebenfalls viel zu entdecken. Wunderbare Barockschlösser in Moritzburg oder Pillnitz, romantische alte Dorfkerne, kleine Parks, die Geschichte gemacht haben und ihre eigenen Geschichten erzählen.

Die Konferenz selbst findet auf dem Campus Südvorstadt in den Räumen des Hörsaalzentrums der Technischen Universität Dresden statt. Das moderne Hörsaalzentrum befindet sich direkt an der B170 (Richtung Prag):



## Verpflegung während der Tagung

An allen drei Tagen stellt das lokale Organisationsteam die Versorgung mit Kaffee und Erfrischungsgetränken sicher. An mehreren Ausgabestationen im Foyer des Hörsaalzentrums werden während der dedizierten Kaffeepausen neben den Getränken auch verschiedenste sächsische Kuchenspezialitäten angeboten. In den Mittagspausen am Donnerstag und Freitag wird die Dresdner Suppenbar, DER Anbieter hochwertiger Suppen in Dresden, ein Buffet verschiedenster Suppen bereithalten. Diese werden mit diversen Zutaten und frisch gebackenem Weißbrot serviert.

Die lokalen Organisatoren möchten Sie bitten, Essen und Getränke nicht mit in die Hörsäle zu nehmen!

Die Verpflegung während der Abendveranstaltung übernimmt das Catering der Sächsischen Dampfschiffahrt. Auf dem Schiff wird ein reichhaltiges Böhmisches Buffet angeboten.

## Abendveranstaltung

Am Donnerstag, den 08. März 2012, findet für die Konferenzteilnehmer ab 18:30 Uhr eine Veranstaltung an Bord der „MS August der Starke“ der Sächsischen Dampfschiffahrt statt. Während der gut dreistündigen Fahrt wird das Schiff von Dresden aus in Richtung Pillnitz und zurück unterwegs sein. Abfahrts- und Ankunftsort ist das Terrassenufer Dresden. Vom Veranstaltungsort der Konferenz wird ab ca. 18:00 Uhr ein Bustransfer zum Terrassenufer eingerichtet – die genauen Abfahrtszeiten entnehmen Sie bitte den Konferenzinformationen vor Ort. Falls Sie noch ins Hotel möchten und selbstorganisiert zum Terrassenufer kommen, beachten Sie bitte, dass das Schiff pünktlich um 19:00 Uhr ablegt.

Bitte bringen Sie Ihren Voucher, den Sie zusammen mit Ihren Tagungsunterlagen erhalten haben, zur Abendveranstaltung mit!

## Sponsoren

Die Konferenz wird unter anderem freundlicherweise unterstützt von:

SAS Deutschland	<a href="http://www.sas.de">www.sas.de</a>
HMS Analytical Software GmbH	<a href="http://www.analytical-software.de">www.analytical-software.de</a>
iCASUS GmbH	<a href="http://www.icasus.de">www.icasus.de</a>
Systematika GmbH	<a href="http://www.systematika.com">www.systematika.com</a>
viadee Unternehmensberatung	<a href="http://www.viadee.de">www.viadee.de</a>
JMP Germany	<a href="http://www.jmp.com/de">www.jmp.com/de</a>
dsquare.de	<a href="http://www.dsquare.de">www.dsquare.de</a>
KYBEIDOS	<a href="http://www.kybeidos.de">www.kybeidos.de</a>
Condat AG	<a href="http://www.condat.de">www.condat.de</a>
Pharmanet i3	<a href="http://www.pharmanet-i3.com">www.pharmanet-i3.com</a>

## WLAN und Internet

Im Tagungsbüro stehen den Besuchern der 16. KSFE mehrere moderne Computer mit Zugang zum Internet zur Verfügung. Für diese PCs benötigen Sie keine gesonderten Zugangsdaten. Bitte beachten Sie, dass keine Drucker zur Verfügung stehen werden.

Im Bereich des Hörsaalzentrums steht flächendeckend ein drahtloses Netzwerk zur Verfügung. Besucher die bereits über einen EDUROAM-Zugang verfügen, sollten sich ohne weitere Konfiguration am EDUROAM-Netz anmelden können.

## Organisatorische Hinweise

---

Besucher ohne EDUROAM-Zugang finden in den Konferenzunterlagen und im Tagungsbüro die notwendigen Login-Daten für den Internetzugang durch ein drahtloses Konferenznetzwerk.

Bei Problemen wenden Sie sich bitte an das Tagungsbüro.

### **Bus-/Bahnfahrplan**

Eine Übersicht der Bus- und Bahnhaltstellen in der Umgebung des Hörsaalzentrums finden Sie in den Konferenzunterlagen. Ein kostenloses 2-Tages-Ticket (Donnerstag & Freitag) erhalten Sie auf Anfrage im Konferenzbüro.

Zwei abschließende Bitten von den lokalen Organisatoren:

1. In den Tagungsunterlagen haben Sie einen Fragebogen zur 16. KSFE erhalten. Bitte füllen Sie diesen am Ende der Konferenz aus und geben Sie ihn im Tagungsbüro ab.
2. Die Plastikhüllen der Namensschilder können wir auch für die nächste KSFE wieder verwenden. Geben Sie daher Ihr Namensschild bitte auch am Ende der Konferenz im Tagungsbüro ab. Danke!

Inhaltsverzeichnis

<b>Organisationskomitee</b> .....	<b>4</b>
<b>Programm</b> .....	<b>5</b>
Tutorien im Vorfeld der Konferenz   Mittwoch, 07. März 2012 .....	5
Vorträge   Donnerstag, 08. März 2012 .....	5
Vorträge   Freitag, 09. März 2012 .....	7
Begleitende Posterausstellung .....	8
SAS Zertifizierung exklusiv für KSFE Teilnehmer .....	9
Tutorium I: Statistische Grafiken mit ODS Graphics: aussagekräftig und ansprechend.....	9
Tutorium II: Einführung in die SAS Makrosprache.....	10
<b>Kurzfassungen Vorträge</b> .....	<b>11</b>
Begrüßung der Konferenzteilnehmer, Eröffnung der 16. KSFE und Eröffnungsvortrag: .....	11
Speeding Up Exploration .....	11
Einsatz des neuen Text Miner 5.1 zur Nachrichtenklassifikation von Online Medien .....	11
Social Media Analysis in Twitter .....	11
SAS OnDemand for Academics - Erste Erfahrungsberichte .....	12
SAS Enterprise Miner in der Lehre – von der Einführung zur Zertifizierung .....	12
SAS-Know-how in der Organisation entwickeln und pflegen - Ressourcen für die Ausbildung und für das Wissensmanagement .....	13
Tipps zur SAS-Programmierung in der PIAFStat-Umgebung .....	13
Erstellen flexibler Makros am Beispiel einer SQL-Routine .....	14
SAS Backstage .....	14
Tipps, Tricks und Stolperfallen zum Umwandeln von Variablen .....	15
Proc Transpose oder Do-it-yourself .....	15
Hierarchische Bayes'sche Netze: Eine Methode, um Dokumentensammlungen in den hierarchischen Strukturen zu klassifizieren - so, wie ein Mensch die Texte schreibt. ....	16
Bestimmung der optimalen Fallzahl in Dosis-Findungs-Studien mit aktiver Kontrolle und linearen sowie nichtlinearen Dosis-Wirkungs-Profilen.....	16
Randomisierungsverfahren mit SAS.....	17
Webservices und SAS - PROC SOAP, PROC HTTP und der ganze REST.....	17
Das Beste aus zwei Welten - Aufruf von R-Funktionen mit PROC IML .....	18
Der Umgang mit fehlenden Werten in Verlaufsstudien im Falle von Drop-outs .....	19
Modellierung rechts-zensierter Zähldaten - ein SAS-Makro unter Nutzung von PROC NL MIXED.....	19
Die neue Style Engine in ODS – Cascading Style Sheets mit SAS 9.3 .....	20
SAS in der Praxis - Tipps und Tricks für SAS Web Report Studio .....	20
Große Datenmengen knacken mit SAS High Performance Analytics .....	21
Implementation of ICH E3 based SAS reporting for a clinical trial – issues and solutions.....	21
Performante Erzeugung von Berichten in (vor)formatierten Excel-Dateien .....	23
ODS Report Writing Interface - Neue Möglichkeiten der Berichtserstellung .....	23
JMP10: Sneak Preview für Entwicklung und Modellierung .....	24
Statistische Versuchsplanung mit JMP - von der Klassik zur Moderne .....	24
Ein Makro zur Berechnung von Diskriminanz- und Reklassifizierungsstatistiken für die Verbesserung eines Prädiktionsmodells bei Anwendung der Cox Regression.....	25
Ein modulares Statistikmakro-Paket für direkte & indirekte Meta-Analysen .....	26
Berechnung des relativen Überlebens auf Basis von Krebsregisterdaten mit SAS .....	26
Erweiterungen der Analysemöglichkeiten im neuen Enterprise Miner 7.1 sowie im Model Manager.....	27
Einstieg in Text Analytics für SAS Enterprise Guide Anwender: Von der Datenquelle zum Bericht - und darüber hinaus.....	27
Verwaltung eines Metadatenrepositories im DI 3.4 mittels einer externen Exceldatei.....	28

## Inhaltsverzeichnis

---

Methoden der Risikobewertung.....	29
Datenvisualisierung mit SAS Visual Analytics Explorer .....	29
<b>Kurzfassungen Poster.....</b>	<b>30</b>
Laufzeitoptimierung bei der Verknüpfung großer Datenmengen. Ein Vergleich zwischen MERGE im DATA STEP und JOIN in PROC SQL .....	30
Ein SAS Macro zur Auswertung von SAE-Daten .....	30
SAS Macro zur Kombination von möglichen in Zusammenhang stehenden Faktoren bezüglich einer Zielgröße beschrieben anhand eines aktuellen Beispieldatensatzes .....	31
Medizinische Statistik mit ??? in SAS® 9.3 – ein Plädoyer für eine Windows-nahe Oberfläche .....	31
DATA Step Component Interface – neue Objekte im DATA Step .....	32
Nie wieder Post sortieren – vollautomatisierte Verarbeitung von Kundenpost mittels Text-Mining-Verfahren .....	33
Anwendung eines SAS/STAT-Modells ohne SAS/STAT in einem CRM-Projekt.....	33
<b>Organisatorische Hinweise.....</b>	<b>34</b>
Tagungsort Dresden – Stadt der Kultur, Wissenschaft & Technologie.....	34
Verpflegung während der Tagung.....	35
Abendveranstaltung .....	35
Sponsoren .....	35
WLAN und Internet.....	35
Bus-/Bahnfahrplan.....	36





# systematika

Chancen ergreifen. Mehr Wert entdecken.

