

Vorwort und Danksagung

Dieses Buch entstand nicht in luftleerem Raum, sondern inmitten von Menschen. Bei ihnen möchte ich mich zuallererst bedanken.

Für die tatkräftige Unterstützung beim Abfassen dieses Buches bedanke ich mich bei Oliver Hiller (Dipl.Ing. (FH) für den Studiengang Biotechnologie, Berlin) und Peter Bonata (Stud.cand., Kaiserslautern).

Herrn Prof. Rainer Schlittgen (Universität Hamburg) als Herausgeber der Reihe, Lehr- und Handbücher der Statistik' und Herrn Weigert vom Oldenbourg Verlag danke ich ebenfalls für die jeweils großzügige Unterstützung.

Den Herren Professoren Klaufuß und Lamers (jeweils Pädagogische Hochschule Heidelberg), Dr. Stich (Institut für Seenforschung, Langenargen), Dr.med. Karl-Heinrich Adzersen (Heidelberger Brustkrebssstudie), Dipl.Biol. Michael Scheurig (Brutvogel-Monitoring Baden-Württemberg), und Herrn Felix Schulz (Projekt „Die Entwicklung der Delinquenz von Kindern, Jugendlichen und Heranwachsenden in Deutschland“) danke ich jeweils für die Erlaubnis, Daten ihrer Projekte für dieses Buch verwenden zu dürfen.

Den Leserinnen und Lesern möchte ich danken, dass sie diesem Buch ihre wertvolle Aufmerksamkeit schenken.

Den Dank an meine Freundinnen und Freunde möchte ich gar nicht erst in Worte zu fassen versuchen. Ohne euch gäbe es dieses Buch nicht, und ich wäre nichts ohne euch.

Dieses Buch entstand aus der jahrelangen Arbeit mit dem SAS System, und der dabei erlangten Überzeugung, dass SAS eines der besten und leistungsfähigsten der derzeit erhältlichen Systeme zur Analyse und Präsentation von Daten ist.

Ein Anliegen dieses Buches war auch, diverse Vorurteile SAS gegenüber auszuräumen. SAS sei benutzerunfreundlich. SAS sei kompliziert, heißt es. Es gebe kein „richtiges“ Handbuch zu SAS. Mit SAS könne man keine vorzeigbaren Grafiken erstellen. Und vieles andere mehr.

Das Gegenteil ist richtig. SAS ermöglicht es oft erst, überhaupt mit kompliziertesten Datenstrukturen und -analysen umgehen zu können. Viele andere Programme sind weit weniger flexibel, oder rechnen sogar so falsch, dass Tester für die Arbeit mit ihnen aufgrund ihres schlechten Abschneidens in Benchmarktests schlichtweg abraten. Die Dokumentation des SAS Systems ist hervorragend, und z.T. mehrfach mit Preisen ausgezeichnet worden. Natürlich ist die Dokumentation auf Englisch, und dennoch haben Qualität und Umfang ihren vergleichsweise günstigen Preis.

SAS ist darüber hinaus auch nicht gleich SAS. SAS ist zwar ein System, aber nicht nur ein *einzelnes* Programm. Das SAS System bietet mehrere Anwendungen, und damit viele verschiedene Zugänge an, die mit animierten Hilfe-Assistenten (Enterprise Guide), Ergebnisinterpretationen (SAS/LAB), und auch völlig ohne Programmierkenntnisse (SAS/INSIGHT, JMP) auch dem statistiktraumatisierten Anfänger einen unkomplizierten Einstieg in Welt der Daten und Algorithmen zu ermöglichen. Die Leistungsvielfalt und Mächtigkeit des SAS Systems kann und sollte nur im Gesamtzusammenhang beurteilt werden.

Dieses Buch möchte auch Perspektiven für Data Analysten, Projekte und Unternehmen aufzeigen, wenn sie sich für das SAS System entscheiden. Die Entscheidung für das SAS System ist auch eine Frage des Offenseinwollens in die Zukunft, eine Entscheidung für die Ausbaufähigkeit und Kontinuität der eigenen Arbeit und Kompetenzen, auf die enormen Einsatz- und Ausbaumöglichkeiten der statistischen Module im Gesamtkonzept des SAS Systems zurückgreifen zu wollen.

Die Komplexität der Analyse von Daten sollte ebenfalls angemessen eingeschätzt werden. Vor den Gefahren methodischer Halbbildung kann nicht oft genug gewarnt werden (siehe Kromrey, 1999). Die Dunkelziffer suboptimaler, unvollständiger, oder eindeutig falsch publizierter Ergebnisse ist nicht zu unterschätzen. Viele Untersuchungen, die die Qualität der publizierten Studien untersuchen, bestätigen diese vorsichtigen kritischen Einschätzungen. Autoren wie z.B. Nuovo, Melnikow & Chang (2002), die kritisch Versuchsplanung und Design in medizinischen Veröffentlichungen untersuchten, sprechen gar von einem Großteil.

Je nach Daten und Analyse sind u.U. Kenntnisse vieler Disziplinen, Ansätze und Theorien nötig, u.a. der Objekt- und Wissenschaftstheorie, Mess- bzw. Testtheorie, Statistik, und der Informatik. Alle Rechenverfahren setzen z.B. statistische Voraussetzungen voraus, die die Daten erfüllen müssen, um überhaupt ausgewertet werden zu können. Verantwortlichen in den einzelnen Phasen des Forschungsprozesses kommt die Aufgabe zu, die Sicherstellung der statistischen Voraussetzungen der jeweiligen statistischen Verfahren zu gewährleisten. Kein Statistikprogramm, auch nicht SAS, kann diese Voraussetzungen überprüfen; dies können nur die jeweiligen Anwender. Eine sorgfältige Planung vor der Erhebung erspart viele Schwierigkeiten bei der Auswertung und Interpretation. Versäumnisse Dritter bei der Erhebung bzw. Kodierung der Daten kann ein Data Analyst weder verantworten, und nur in den seltensten Fällen „retten“. Forschung ist Einstellungssache. Gegenüber Qualität, Präzision und Verantwortung. Ergebnisse aus Wissenschaft und Forschung fließen in die Alltagsrealität zurück und bilden dort die Grundlage für zukünftiges Handeln und Entscheiden. Dieser immer mögliche Anwendungsrückbezug sollte für jeden Data Analysten Anlass genug sein, so kompetent methodisches Reflexionsbewusstsein anzuwenden, dass für die Qualität der Forschung in jedem Fall auch die wissenschaftliche und soziale Verantwortung übernommen werden kann. Nur professionell angewendete Statistikprogramme ge-

währleisten, dass Wissenschaft und Forschung ernst genommen werden. SAS stellt dazu das hochleistungsfähige SAS System für u.a. die Analyse und Präsentation von Daten zur Verfügung. Die Verantwortung für seine eigenen Fähigkeiten kann nur der jeweilige Anwender übernehmen.

Dieses Buch versucht, die Bürde des Einstiegs in die Welt des Managements und der Analyse von Daten ein bisschen leichter zu machen, und die Möglichkeiten des SAS Systems ansatzweise anzudeuten. Neben der gezielt erarbeiteten Didaktik vom mausgesteuerten Einstieg bis hin zur komplexen Syntaxprogrammierung ist hervorzuheben, dass zu fast allen Verfahren die wichtigsten konzeptionellen bzw. statistischen Voraussetzungen zusammengestellt wurden, die es idealerweise *vor* der Erhebung und Analyse von Daten zu beachten gilt. Die wichtigste Literatur ist im Literaturverzeichnis zusammengestellt. Ein in weit fortgeschrittener Vorbereitung befindliches SAS Buch für Profis (Schendera & Hiller, o.J.) wird u.a. viele andere statistische Verfahren (u.a. Varianzanalyse für gemischte Effekte, Mehrfachregression, Nichtlineare Regression) behandeln, und an zahlreichen Beispielen die SAS Makroprogrammierung, wie auch weitere SAS Anwendungen (u.a. das Output Delivery System) und Produkte u.a. auch aus SAS Version 9 vorstellen.

Dieses Buch wird zeigen können, warum Ihre Entscheidung für das SAS System eine gute Entscheidung war.

Ich habe mich bemüht, das Buch so aktuell wie möglich zu halten. Wie die Leserinnen und Leser selbst nachvollziehen werden können, unterliegt das SAS System einer derart dynamischen Weiterentwicklung, dass nicht ausgeschlossen werden kann, dass einige Informationen bei Drucklegung nicht mehr optimal aktuell sind. Weil sich das SAS System in Bezug auf Produkte, Lösungen und Qualität kontinuierlich weiterentwickelt, kann diese Einführung keinen Anspruch auf Aktualität und Vollständigkeit haben. Diese Zusammenstellung leistet allenfalls eine erste Übersicht und Orientierung im Bereich der statistischen und grafischen Datenanalyse und -präsentation mit der SAS Software (Stand: 06/2003), und hofft zumindest einen annähernden Einblick in das Leistungspotential und die offene Dynamik des SAS Systems vermitteln zu können. Sollten Leser feststellen, dass das Buch nicht mehr der Aktualität von SAS entspricht, so ist die Verantwortung beim Verfasser, und sicher nicht bei SAS zu suchen. Aus dem notwendigen Rückgriff auf die umfangreichen technischen Informationen über die SAS Software kann keinesfalls geschlossen werden, dass dieses Buch in irgendeiner Hinsicht von SAS autorisiert sei.

Übersicht und Aufbau des Buches

Dieses Buch ist in drei große Einheiten untergliedert: SAS als System, SAS per Maussteuerung (Einstieg), und SAS per Syntaxsteuerung.

Die erste Einheit stellt das SAS System vor, und zwar nicht nur als simples Analyseprogramm für die Datenverarbeitung und Statistik, sondern mit seinem gesamten Leistungspotential als Unternehmen und System. Aber selbst diese sicher eindrucksvolle Präsentation wird dem vollen Potenzial des SAS System nicht gerecht werden. Der interessierte Leser wird an diesem ersten Überblick erkennen können, welche Entwicklungsmöglichkeiten das SAS System bieten kann.

Die zweite Einheit stellt mehrere Module, Oberflächen und Anwendungen für den unkomplizierten mausgesteuerten Einstieg in SAS vor, u.a. SAS/ANALYST, SAS/ASSIST, ENTERPRISE GUIDE, und SAS/INSIGHT. Mit einfacher Maussteuerung können relativ unkompliziert bereits komplexe Berechnungen bzw. Datenpräsentationen erzeugt werden. Diese Einheit stellt vorwiegend die Benutzerführung dieser Module, Oberflächen und Anwendungen vor. Deskriptive Parameter und inferenzstatistische Verfahren werden in der Einheit zur Syntaxsteuerung erläutert. Nicht nur, um eine gewisse Redundanz zu vermeiden, sondern weil zu erwarten ist, dass der Leser mit der Zeit selbst auf die Vorteile der Syntaxsteuerung umsteigen wird, und im Laufe der Zeit die Vorteile dieser Darstellung nutzen wird.

Übersicht 1: Aufbau des Buches

Einheit	Stichwort	Kapitel
Das SAS System	Dr. Jim Goodnight	1
	Das System hinter dem SAS System	2
	Das SAS System	3
	SAS Produkte für die grafische und statistische Analyse von Daten	4
	SAS Prozeduren für die grafische und statistische Analyse von Daten	5
Der Einstieg: SAS per Maus	Analyst Application	6
	SAS/INSIGHT	7
	SAS/ASSIST	8
	Enterprise Guide	9
SAS per Syntax: SAS Base Language, SAS/QC, SAS/STAT, SAS/GRAPH	Daten-Management	10
	Deskriptive Statistik u.a. UNIVARIATE, MEANS, TABULATE	11
	Tabellenanalysen mit PROC FREQ	12
	Einführung in die Inferenzstatistik u.a. TTEST, NPARIWAY	13
	Korrelation und Regression CORR, REG	14
	Varianzanalysen u.a. ANOVA, GLM, NESTED, NPARIWAY	15
	Clusteranalyse CLUSTER, FASTCLUS	16
	Kovarianzanalyse GLM	17
	Faktorenanalyse FACTOR	18
	Pfadanalyse CALIS	19
	Überlebenszeitanalyse LIFETEST	20
	Hauptkomponentenanalyse PRINCOMP	21
	Prognose FORECAST	22
	Grafiken mit SAS u.a. SAS/GRAPH	23
Ausblick und mehr	„Mercury“ SAS Version 9	24
	Verzeichnis der Übersichten	25
	Literatur	26
	Hinweise zu den Daten	27
	Kontakt zu SAS	28
	Ihre Meinung zu diesem Buch	29
	Stichwortverzeichnis	30
	Autor	31

Die dritte Einheit behandelt die Syntaxsteuerung von SAS, und teilt sich in vier Untereinheiten auf. Der erste Teil behandelt das Daten-Management, der zweite Teil die deskriptive Statistik, der dritte Teil ausgesuchte Verfahren der Inferenzstatistik, und der vierte Teil stellt die Programmierung von Grafiken mit dem SAS System vor. Die Kapitel zur Statistik enthalten jeweils Angaben über die wich-

tigsten Voraussetzung der Verfahren, Rechenbeispiele, erläuterter Code und Output. Das umfangreiche Kapitel zu SAS Grafiken beginnt zunächst bei der Vorstellung der gebräuchlichsten Grafiktypen, und steigert anschließend die Programmierung von SAS Grafiken von ganz einfachen Beispielen bis hin zu komplexen Grafiken mit Annotate und SAS Makros. Die Syntax wird ausführlich erläutert. Abschließend werden spezielle Themen wie z.B. Import/Export von SAS Grafiken und das SAS Output Delivery System angesprochen. Die letzte Einheit stellt u.a. einen Ausblick auf die SAS Version 9 („Mercury“), und einen Überblick über die Übersichten dieses Buches, die verwendete Literatur, und die Kontaktmöglichkeiten zu SAS bzw. zum Autor bereit.

Dieses Buch verwendet häufig englische Begriffe, z.B. bei Statistiken und bei der Bedienung des SAS Systems. Dies geschah in voller Absicht. Die Online und Druckdokumentation und damit auch die Terminologie von SAS ist fast ausnahmslos englisch, ebenfalls wie die meisten Menüoberflächen (Ausnahme: Enterprise Guide). Die Überlegung war einfach die: Wenn dieses Buch auch die Nomenklatur des SAS System vermittelt, dann können sich Leserinnen und Leser leichter in Dokumentation und Bedienung orientieren.