Statistik mit SAS

Von Prof. Dr. rer. nat. Julius Dufner
Priv.-Doz. rer. nat. Uwe Jensen
Dr. rer. nat. Erich Schumacher

Universität Hohenheim

Mit zahlreichen Abbildungen,
Beispielen und Übungsaufgaben

Springer Fachmedien Wiesbaden GmbH

Prof. Dr. rer. nat. Julius D. fner

Geboren 1941 in Freiburg . Br. Von 1960 b s 1967 Studium der Mathematik und Physik an der Universität Freiberg. Ab 1967 Assistententätigkeit am Mathematischen Institut der Universität Freiburg Promotion 1971. Von 1972 bis 1974 Assistententätigkeit an der Pädagogischen Hochschule Freiburg, zweites Staatsexamen. Dozent an der FH Darmstadt von 1974 bis 1976, ab 1976 an der Berufspädagogischen Hochschule Esslingen, 1979 Professor. Von 1988 an Professor an der Universität Hohenheim.

Privatdozent Dr. rer. nat. Uwe Jensen

Geboren 1950 in Bremen. Von 1971 bis 1976 Studium der Mathematik, Physik und Betriebswirtschaft an der Technischen Universität Braunschweig, Diplom 1976. 1979 Promotion und 1987 Habilitation im Fach Mathematik an der Universität Stuttgart-Hohenheim. 1976/77 Industrietätigkeit in Frankfurt. Von 1977 bis 1980 Wiss. Angestellter, seit 1980 Akademischer Rat/Oberrat am Institut für Angewandte Mathematik und Statistik der Universität Hohenheim.

Dr. rer. nat. Erich Schumacher

Geboren 1945 in Bonladen. Studium der Mathematik an der Universität Stuttgart, Diplom 1969. Von 1970 bis 1974 Wiss. Assistent, seit 1975 Wiss. Angestellter am Institut für Angewandte Mathematik und Statistik der Universität Hohenheim. 1979 Promotion in Hohenheim.

Die Deutsche Bibliothek – CIP-Einheitsaufnahme

Dufner, Julius:
Statistik mit SAS : mit Beispielen und Übungsaufgaben /
von Julius Dufner ; Uwe Jensen ; Erich Schumacher. –
Stuttgart : Teubner, 1992
(Teubner-Studienbücher : Mathematik)
ISBN 978-3-519-02088-2 ISBN 978-3-322-94766-6 (eBook)
DOI 10.1007/978-3-322-94766-6
NE: Jensen, Uwe:; Schumacher, Erich:

Das Werk einschließlich aller seiner Teile ist urheberrechtlich geschützt. Jede Verwertung außerhalb der engen Grenzen des Urheberrechtsgesetzes ist ohne Zustimmung des Verlages unzulässig und strafbar. Das gilt besonders für Vervielfältigungen, Übersetzungen, Mikroverfilmungen und die Einspeicherung und Verarbeitung in elektronischen Systemen.

© Springer Fachmedien Wiesbaden 1992
Ursprünglich erschienen bei B. G. Teubner Stuttgart 1992
Gesamtherstellung: Druckhaus Beltz, Hemsbach/Bergstraße
Umschlagentwurf: P.P.K,S-Konzepte T. Koch, Ostfildern/Stuttgart

Vorwort

Aufgrund einer in den letzten Jahren sprunghaft gewachsenen Verfügbarkeit über Rechnerkapazitäten, insbesondere im Bereich der Personal Computer (PC), lassen sich heute auch umfangreiche und aufwendige statistische Datenanalysén innerhalb kürzester Zeit ausführen. Die zunehmende Bedeutung der Statistik in nahezu allen Wissenschaftsdisziplinen geht nicht zuletzt zurück auf diese gewachsenen Möglichkeiten, eine statistische Datenanalyse praktisch durchzuführen. Dafür ist ein geeignetes Statistik-Softwarepaket erforderlich. SAS (Statistical Analysis System) zählt zu den am weitesten verbreiteten und leistungsfähigsten Software-Systemen dieser Art.

Das Buch richtet sich an den Anwender statistischer Verfahren. Damit ist einerseits der Nichtmathematiker gemeint, der durch Beobachtungen oder aufgrund von Experimenten Daten gesammelt hat und diese für eine geordnete Darstellung aufbereiten möchte und Schlußfolgerungen aus den gewonnenen Daten ziehen will. Dazu werden Verfahren der beschreibenden und der beurteilenden Statistik herangezogen. Diese Verfahren sollen dann mit Hilfe einer leistungsfähigen Statistik-Software auf einem Rechner umgesetzt werden.

Andererseits richtet sich dieses Buch auch an den Mathematik-Studenten (Dozenten) mit Interesse an der angewandten Stochastik, der die in den Statistikvorlesungen vermittelten Verfahren mit Hilfe eines Computers realisieren möchte. Auch der erfahrene Statistiker kann, so hoffen wir, an der einen oder anderen Stelle Nutzen aus diesem Buch ziehen.

Vorausgesetzt wird in jedem Fall ein Grundkurs in Statistik oder mathematischer Stochastik, wie er eigentlich in allen natur- und sozialwissenschaftlichen Disziplinen im Grundstudium angeboten wird. Die benötigten Begriffe und Resultate werden zwar alle noch einmal zusammengestellt und knapp erläutert, nicht jedoch in der Form, wie es für ein Lehrbuch der Statistik angebracht wäre.

In der Bereitstellung und Verfügbarkeit einer großen Vielfalt von statistischen Verfahren durch Statistik-Software-Systeme, die in immer kürzeren Abständen um neue Module mit immer komplexeren Methoden bereichert wird, liegt auch eine gewisse Gefahr. Das Wissen des Anwenders um die Hintergründe dieser Verfahren hält oft nicht

Schritt mit dieser rasanten Entwicklung. Ein solches Hintergrundwissen erscheint unserer Meinung nach, zumindest zu einem gewissen Grad, auch für den Anwender erforderlich zu sein, damit er das seinem Problem angemessene Modell auswählen kann, die Modellvoraussetzungen versteht, aus den Resultaten der Rechnung die richtigen Schlüsse ziehen kann und nicht zu Fehlinterpretationen geführt wird.

Im vorliegenden Buch werden deswegen zu allen statistischen Verfahren die Modelle erläutert und die Voraussetzungen zur Anwendung des jeweiligen Verfahrens genannt. Dann wird, zumeist anhand eines Beispiels, die Durchführung mit Hilfe von SAS beschrieben durch Angabe des Programm-Textes und dessen Erläuterung. Ein solches Programm führt zu graphischen Darstellungen und/oder zu Ausgabedateien, die im Text kurz Output genannt werden. Daher schließt sich an die Durchführung mit Hilfe von SAS eine ausführliche Erläuterung und Interpretation des Output an.

Der Leser soll dadurch in die Lage versetzt werden, sein statistisches Problem mit Hilfe von SAS zu lösen, weitgehend ohne auf die für den Anfänger abschreckend umfangreichen SAS-Handbücher zurückgreifen zu müssen. Das Buch geht insbesondere auf die Anwendung von SAS auf dem PC ein. Hierzu sind Grundkenntnisse des Betriebssystems DOS von Vorteil. Obwohl sich das Buch auf die PC-Version von SAS bezieht, ist es mit wenigen Einschränkungen auch für den Benutzer der Großrechner-Version geeignet.

Das vorliegende Buch ist weder ein Lehrbuch der Statistik noch eine systematische Einführung in SAS. Schwerpunkt der Darstellungen sind die Konzepte der Statistik, SAS dient als Werkzeug zur Realisierung dieser Konzepte. Daher ist das Buch auch gegliedert nach methodischen Gesichtspunkten der Statistik. SAS wird nur soweit vorgestellt, wie es zur Umsetzung der einzelnen statistischen Methoden notwendig ist. Deshalb kann dieses Buch auch kein Ersatz für die äußerst umfangreichen SAS-Handbücher sein, die immer dann herangezogen werden sollten, wenn man zusätzliche Möglichkeiten ausschöpfen möchte. Auf einige dieser zusätzlichen Möglichkeiten wird im Text durch Verweise auf die entsprechenden SAS-Dokumentationen hingewiesen.

Nach einer Einführung in SAS in den ersten beiden Kapiteln wird die beschreibende Statistik in Kapitel 3 an Hand einer Reihe von Beispielen behandelt. In Kapitel 4 werden die Grundlagen der Wahrscheinlichkeitstheorie und Statistik in knapper Form zusammengestellt und soweit

beschrieben, wie es im weiteren benötigt wird. In Kapitel 5 werden einige grundlegende Verfahren der Statistik vorgestellt. Dazu zählen neben den Ein- und Zweistichprobentests unter Normalverteilungsannahme auch Anpassungstests und die nichtparametrischen bzw. verteilungsfreien Verfahren. Die letzten beiden Kapitel 6 und 7 beinhalten eine Reihe von Verfahren der Varianzanalyse und Regressionsrechnung, die unter dem Begriff lineare Modelle zusammengefaßt werden können. Darin werden auch einige Probleme angeschnitten, die mehr den fortgeschrittenen Statistiker ansprechen, wie z.B. spezielle Randomisationsstrukturen, unbalancierte Daten bei Mehrfachklassifikation, Kovarianzanalyse. Das abschließende Literaturverzeichnis haben wir zur besseren Orientierung um einige Hinweise zu Lehrbüchern und weiterführender Literatur ergänzt.

Bei der erforderlichen Auswahl der Themen haben wir uns von dem Prinzip leiten lassen, einerseits möglichst einfache und grundlegende Verfahren der Statistik vorzustellen und andererseits einige komplexere Methoden zu behandeln, die unserer Erfahrung nach häufig in der Praxis verwandt werden. Gerade in diesem letzten Punkt stützt sich die subjektive Auswahl auf unsere mehrjährige Beratungspraxis und die Zusammenarbeit mit "Anwendern" in Hohenheim. Natürlich konnten dabei einige für die Anwendung interessante Gebiete, wie z.B. multivariate Methoden und Zeitreihenanalyse, nicht in dieses Buch aufgenommen werden.

Wir haben uns bemüht, Computer-Englisch und Abkürzungen weitgehend zu vermeiden. Allerdings erschien es uns sinnvoll, einige Wörter wie z. B. Output im Text wie ein deutsches Wort zu verwenden, da eine direkte Übersetzung, etwa Ausstoß, umständlich und sinnentstellend erscheint. Zu den übernommenen Anglizismen zählt auch, daß im gesamten Text ein Dezimalpunkt statt des im Deutschen üblichen Kommas verwendet wird. Von SAS reservierte Schlüsselwörter (DATA, PROC, UNIVARIATE,...) werden in Großbuchstaben wiedergegeben. Programmtexte und Ausgabedateien sind durch einen Rahmen hervorgehoben. Da die Ausgabedateien der einheitlichen Darstellung wegen ebenfalls in Proportionalschrift gesetzt wurden, können kleine Abweichungen in der Form gegenüber der Bildschirmausgabe auftreten. Disketten mit allen Beispiel-Programmtexten können von uns gegen eine Schutzgebühr bezogen werden.

Schließlich ist es uns eine angenehme Pflicht denen zu danken, die am Zustandekommen dieses Buches beteiligt waren. Dazu zählen eine Reihe von Studenten und uns verbundene Kollegen, die durch fortwährende Diskussionen und Anregungen direkt oder indirekt an der Gestaltung des Buches mitgewirkt haben. Unser Dank gilt Herrn Heinz Becker, der bei der Überprüfung der Programmtexte behilflich war. Ganz herzlich möchten wir uns auch bei unserer **EXP**ertin Frau Regina Schulze bedanken, die uns bei der Erstellung des Textes im Textverarbeitungssystem **EXP** unterstützt hat. Gerne erwähnen wir auch dankend die angenehme Zusammenarbeit mit Herrn Dr. Spuhler vom Teubner Verlag.

Den Benutzern dieses Buches empfehlen wir, die Beispiele auch als Übungsaufgaben anzusehen und diese durch Variieren, Umstellen und Ergänzungen zu einer eigenen kleinen Programmsammlung auszubauen. Dabei wünschen wir viel Erfolg und möglichst wenige rote Fehlermeldungen.

Stuttgart-Hohenheim, im Sommer 1992

Julius Dufner, Uwe Jensen, Erich Schumacher

Inhaltsverzeichnis

Kapitel 1 SAS für Personal Computer

Kapitel 2 Das SAS-Programmsystem

Kapitel 3 Beschreibende Statistik

Kapitel 4 Grundlagen der Wahrscheinlichkeitstheorie und Statistik

Kapitel 5 Beurteilende Statistik - Grundlegende Verfahren

Kapitel 6 Varianzanalyse