

Die Reihe

Biometrie und Medizinische Informatik Greifswalder Seminarberichte

wird herausgegeben von:

Universitätsprofessor Dr. rer. nat. habil. Karl-Ernst Biebler
und Dr. rer. nat. Bernd Jäger

Institut für Biometrie und Medizinische Informatik der
Ernst-Moritz-Arndt-Universität Greifswald
Rathenaustraße 48
17487 Greifswald

Anliegen der Greifswalder Seminarberichte Biometrie und Medizinische Informatik ist es, die Zusammenarbeit im Berührungsfeld von Medizin, Biowissenschaften, Mathematik und Informatik zu fördern. In diesem Sinne steht die Mitarbeit an den Publikationen allen denen offen, die sich einer solchen Vermittlung zwischen den Wissenschaften verpflichtet fühlen, und die dazu einen Beitrag leisten wollen.

Dieses Buch ist sorgfältig hergestellt worden. Die Autoren und Herausgeber sowie der Verlag garantieren jedoch nicht, dass es fehlerfrei ist.

Die Leser werden darauf hingewiesen, auf versehentliche Fehler zu achten. Insbesondere sollen hier mitgeteilte Rechnerprogramme vor Verwendung gründlich getestet werden.

Biometrie und medizinische Informatik
Greifswalder Seminarberichte

**Michael Wodny, Karl-Ernst Biebler, Bernd Jäger,
Paul Eberhard Rudolph, Gunter Teumer**

Statistikpraktikum mit SAS®

Shaker Verlag
Aachen 2010

Bibliografische Information der Deutschen Nationalbibliothek

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

Copyright Shaker Verlag 2010

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

Printed in Germany.

ISBN 978-3-8322-8785-6

ISSN 1439-5320

Shaker Verlag GmbH • Postfach 101818 • 52018 Aachen

Telefon: 02407 / 95 96 - 0 • Telefax: 02407 / 95 96 - 9

Internet: www.shaker.de • E-Mail: info@shaker.de

Statistikpraktikum mit SAS[®]

Michael Wodny
Karl-Ernst Biebler
Bernd Jäger
Paul Eberhard Rudolph
Gunter Teumer

Über die Autoren

Univ.-Prof. Dr. rer. nat. habil. Karl-Ernst Biebler ist Direktor des Instituts für Biometrie und Medizinische Informatik an der Medizinischen Fakultät der Ernst-Moritz-Arndt-Universität Greifswald.

Dr. rer. nat. Bernd Jäger und Dr. rer. med. Michael Wodny sind wissenschaftliche Mitarbeiter am gleichnamigen Institut.

Dr. rer. nat. Paul Eberhard Rudolph arbeitet als wissenschaftlicher Mitarbeiter am Leibniz-Institut für Nutztierbiologie, Dummerstorf.

Dr. rer. nat. habil. Gunter Teumer ist freiberuflicher Privatdozent und wissenschaftlicher Mitarbeiter am Institut für Mathematik und Informatik der Mathematisch-Naturwissenschaftlichen Fakultät der Ernst-Moritz-Arndt-Universität Greifswald.

Vorwort

Die Anwendung statistischer Verfahren gehört zu den selbstverständlichen Arbeitsaufgaben im Forschungsbetrieb vieler Disziplinen. In wirtschaftswissenschaftlichen, technischen und naturwissenschaftlichen Studiengängen sowie auch im Medizinstudium wird dazu Methodenwissen vermittelt.

Die praktische Umsetzung solcher Verfahren stellt sich jedoch oft als neue Anforderung heraus. Nur in Ausnahmefällen kann ein Taschenrechner genügen. Für Datenerfassung, Datenbehandlung und die Auswertungen sind vielfältige Softwaresysteme kommerziell verfügbar. Entsprechend dem Zweck wird man Software auswählen und sich deren Möglichkeiten erschließen.

In dem vorliegenden Buch werden Beispiele für die Nutzung des Statistik-Softwaresystems SAS[®] gegeben. Dieses System ist im industriellen Bereich weit verbreitet und stellt für die wissenschaftlichen Datenauswertungen in der Medizin gewissermaßen den internationalen Standard dar. Es bietet alle erforderlichen Möglichkeiten der Datenbearbeitung, stellt eine enorme Vielfalt an Verfahren zur Datenauswertung bereit und gestattet über Schnittstellen zu anderen Programmiersprachen individuelle Problemlösungen.

Offenbar kann man ein solches komplexes System nicht umfassend kennen. Deswegen brauchen selbst Einsteiger jedoch keinesfalls auf die angebotenen Möglichkeiten zu verzichten.

Im vorliegenden Buch werden typische Aufgabenstellungen wie Datenarbeit, Erstellung von Grafiken, Berechnung statistischer Testgrößen, Anpassung von Modellfunktionen an Daten sowie explorative Datenanalysen an Beispielen behandelt. Auf den statistischen bzw. mathematischen Hintergrund wird dabei nicht genauer eingegangen, die Methodenkenntnis ist meist vorausgesetzt. Bei Bedarf studiere man weitere Literatur (z.B. Rasch u. A. (1996), Sachs und Hedderich (2009), Hartung u. A. (1984), Biebler und Jäger (2008)). Die von uns verwendete Terminologie vermeidet allzu strengen Fachbezug.

Der Leser wird durchaus unterschiedliche Zugriffsmöglichkeiten auf SAS[®] vorfinden. Das mag die eigene weiterführende Erschließung des Systems erleichtern. Wir haben die Erfahrung gemacht, dass Doktoranden bereits nach einem einführenden Kurs von 6 Stunden weitgehend auf sich gestellt erfolgreich mit SAS[®] arbeiten konnten.

Mit diesem Buch möchten die Autoren einen Beitrag leisten zur Unterstützung individueller wissenschaftlicher Arbeit.

Michael Wodny
Karl-Ernst Biebler
Bernd Jäger
Paul Eberhard Rudolph
Gunter Teumer

Greifswald und Dummerstorf, Januar 2010

Inhaltsverzeichnis

Einleitung.....	9
1 Dateiarbeit.....	11
1. 1 Zu den Fenstern im SAS® und im Enterprise Guide	11
1. 2 Anlegen einer SAS® -Datendatei.....	13
1. 3 Variablen und Datensätze für eine Teildatei auswählen.....	18
1. 3. 1 Auswahl einzelner Variablen.....	19
1. 3. 2 Auswahl einzelner Datensätze	20
1. 3. 3 Auswahl von Variablen und Datensätzen.....	21
1. 4 Dateien zusammenfügen	22
1. 4. 1 Dateien hintereinander schreiben	22
1. 4. 2 Dateien nebeneinander schreiben	24
1. 4. 3 Dateien mit Hilfe eines Schlüsselfeldes verbinden	25
1. 5 Erzeugen neuer Variablen aus vorhandenen	26
1. 6 Fremddateien eingeben.....	29
1. 7 Daten editieren.....	31
1. 8 Makroprogrammierung in SAS®	34
2 Grafiken erstellen mit dem SAS® Enterprise Guide	39
2. 1 Erstellen eines einfachen Balkendiagramms	39
2. 2 Veränderungen des einfachen Balkendiagramms.....	40
2. 3 Programmcode aus SAS® Enterprise Guide verändern	42
2. 4 Weitere Grafikmöglichkeiten im SAS® Enterprise Guide.....	45
2. 5 Beispiele	46
3 Der Signifikanztest	54
3. 1 Der t-Test für unverbundene Stichproben	57
3. 2 Der t-Test für verbundene Stichproben	60
3. 3 Die χ^2 -Tests	62
3. 3. 1 Der Homogenitätstest	63
3. 3. 2 Der Unabhängigkeitstest.....	68
3. 3. 3 Der χ^2 -Test bei gegebener Tafel	68
3. 3. 4 Der χ^2 -Anpassungstest.....	70
3. 4 Der U-Test.....	73
3. 5. Der Vorzeichentest und der Wilcoxon-Rangsummentest.....	76
3. 6 Der Symmetrietest nach Bowker.....	79
3. 7 Der Test von Kolmogorov und Smirnov	84
3. 8 Der Median-Test.....	87
3. 9 Der unbedingte Barnard-Test	89
3. 10 Der Friedman-Test.....	94
3. 11 Der Trendtest von Cox und Stuart.....	98

3. 12	Der Trendtest nach Jonckheere	103
4	Regression, Korrelation, Kurvenanpassung.....	110
4. 1	Wahrscheinlichkeitstheoretische Grundlagen	110
4. 1. 1	Die Polynomialverteilungen	115
4. 1. 2	Die zweidimensionalen Normalverteilungen	118
4. 2	Regression	121
4. 2. 1	Allgemeiner Regressionsansatz	121
4. 2. 2	Spezielle Regressionsansätze	122
4. 3	Korrelation.....	123
4. 4	Statistik.....	124
4. 5	SAS® – Programme	125
4. 5. 1	Der Pearson'sche Korrelationskoeffizient.....	125
4. 5. 2	Lineare Regression mit Konfidenzintervall.....	129
4. 5. 3	Der Korrelationskoeffizient nach Spearman	132
4. 6	Kurvenanpassung	135
4. 6. 1	Polynomialer Ansatz	135
4. 6. 2	Allgemeiner linearer Ansatz	139
5	Diskriminanzanalyse.....	149
5. 1	Nächste-Nachbarn-Diskriminanzanalyse	151
5. 2	Wahrscheinlichkeiten von Fehlklassifikationen	154
5. 3	Schätzung von a-posteriori-Wahrscheinlichkeiten	155
5. 4	SAS®-Prozeduren für Diskriminanzanalysen	155
5. 5	Die SAS®-Prozedur DISCRIM.....	157
5. 6	Die Nächste-Nachbarn-Diskriminanzanalyse.....	161
6	Überlebenszeitanalyse	169
6. 1	Wahrscheinlichkeitstheoretische Modellierung	169
6. 2	Sterbetafeln.....	173
6. 3	Datenauswertung	174
6. 3. 1	Überlebenszeitanalyse nach der Sterbetafelmethode.....	174
6. 3. 2	Überlebenszeitanalyse nach Kaplan und Meier.....	179
	Literatur	189
	Stichwortverzeichnis.....	191