

Hans-Jürgen Andreß

GLIM

Verallgemeinerte lineare Modelle

Programmiersprachen

Einführung in die Programmiersprache Ada, von M. Nagl

Einführung in die Programmiersprache BASIC,
von W.-D. Schwill und R. Weibezahn

BASIC in der medizinischen Statistik, von H. Ackermann

Einführung in die Programmiersprache COBOL,
von W.-M. Kähler

Einführung in die Programmiersprache FORTRAN 77,
von G. Lamprecht

FORTRAN-Trainer, von H. Kohler

Pascal-Trainer, von H. Kohler

Multivariate Statistik in den Natur- und Verhaltenswissenschaften
von C.-M. Haf und T. Cheaib

Einführung in die Programmiersprache Pascal,
von K.-H. Becker und G. Lamprecht

PEARL, Process and Experiment Automation Realtime
Language, von W. Werum und H. Windauer

Einführung in die Programmiersprache SIMULA,
von G. Lamprecht

Einführung in das Datenanalysesystem SPSS,
von W.-M. Kähler

GLIM: Verallgemeinerte lineare Modelle
von H.-J. Andreß

Vieweg

Hans-Jürgen Andreß

GLIM

Verallgemeinerte lineare Modelle

Mit 7 Abbildungen und 30 Tabellen



Friedr. Vieweg & Sohn Braunschweig/Wiesbaden

CIP-Kurztitelaufnahme der Deutschen Bibliothek

Andress, Hans-Jürgen:

GLIM: verallgemeinerte lineare Modelle /

Hans-Jürgen Andress. — Braunschweig; Wiesbaden:

Vieweg, 1986.

ISBN-13: 978-3-528-04354-4

Das in diesem Buch enthaltene Programm-Material ist mit keiner Art verbunden. Der Autor übernimmt infolgedessen keine \ folgende oder sonstige Haftung übernehmen, die auf irgendeine Ar Materials oder Teilen davon entsteht.

1986

Alle Rechte vorbehalten

© Friedr. Vieweg & Sohn Verlagsgesellschaft mbH, Braunschweig 1986

Die Vervielfältigung und Übertragung einzelner Textabschnitte, Zeichnungen oder Bilder, auch für Zwecke der Unterrichtsgestaltung, gestattet das Urheberrecht nur, wenn sie mit dem Verlag vorher vereinbart wurden. Im Einzelfall muß über die Zahlung einer Gebühr für die Nutzung fremden geistigen Eigentums entschieden werden. Das gilt für die Vervielfältigung durch alle Verfahren einschließlich Speicherung und jede Übertragung auf Papier, Transparente, Filme, Bänder, Platten und andere Medien. Dieser Vermerk umfaßt nicht die in den §§ 53 und 54 URG ausdrücklich erwähnten Ausnahmen.

Umschlaggestaltung: Peter Lenz, Wiesbaden

ISBN-13: 978-3-528-04354-4 e-ISBN-13: 978-3-322-85654-8

DOI: 10.1007/978-3-322-85654-8

Vorwort

Multivariate Analyseverfahren haben in den letzten Jahren einen gewaltigen Aufschwung in den Sozialwissenschaften erfahren. Die Vielfalt der Methoden ist für den Forschungspraktiker kaum noch zu überblicken. Mit diesem Buch wird der Versuch unternommen, ein allgemeines Modell statistischer Datenanalyse vorzustellen, das mehrere multivariate Analyseverfahren als Spezialfall enthält. Dazu zählen die klassischen Anwendungen der Regressions-, Varianz- und Kovarianzanalyse, aber auch verschiedene Methoden zur Analyse von kreuztabellierten Daten (log-lineare und logistische Modelle, GSK-Ansatz).

Dieses verallgemeinerte lineare Modell wurde von NELDER und WEDDERBURN (1972) vorgeschlagen. Für praktische Auswertungen mit diesem Ansatz steht das Programmpaket GLIM zur Verfügung, das von der Numerical Algorithms Group vertrieben wird. Bisher sind sowohl der statistische Ansatz als auch das Programm GLIM in der Forschungspraxis nur sehr zögernd aufgenommen worden. Das liegt sicherlich zum großen Teil daran, daß der Grad der statistischen Formalisierung relativ hoch und der Umfang der Programmdokumentation relativ gering ist. Vor allem fehlt es an konkreten Beispielen, wie man mit einem Datensatz z. B. eine Regressionsanalyse oder ein log-lineares Modell berechnet.

Diesem Mangel soll mit diesem Buch abgeholfen werden, obwohl auch hier einschränkend bemerkt werden muß, daß der Anteil der statistischen Vorarbeiten relativ hoch ist, bis man tatsächlich eine bestimmte Datenanalyse durchführen kann. Dieses Buch ist daher keine Programmbeschreibung mit Beispielen, sondern in erster Linie eine statistische Einführung in die Theorie und Praxis verallgemeinerter linearer Modelle. Es richtet sich an die Personen, die schon gewisse Erfahrungen mit multivariaten Analyseverfahren haben. Es kann aber auch ergänzend oder im Anschluß an eine Einführung in die klassische Regressionsanalyse gelesen werden. Das verwendete Beispiel betrachtet Statusveränderungen in einer Stichprobe von Berufsanfängern und untersucht die Frage, wie diese Veränderungen mit der Qualifikation einer Person und Arbeitsmarktmerkmalen variieren.

Die Idee zu diesem Buch entstand auf einem NONMET/GLIM-Workshop, den das Zentrum für Umfragen, Methoden und Analysen e.V. (Mannheim) in der Zeit vom 16. bis 20.11.1981 veranstaltete. Der Initiator und der Hauptreferent dieses Kurses, die Professoren M. Küchler und G. Arminger, haben daher maßgeblichen Anteil an der Entstehung dieses Buches. Mein besonderer Dank gilt vor allem G. Arminger, durch dessen Arbeiten auf dem Gebiet verallgemeinerter linearer Modelle dieses Werk erst möglich wurde. Ich möchte diese Gelegenheit auch nutzen, Gisela Diekmeier, Christian Kerst und Wilfried Staemmler ganz herzlich zu danken, die mit bei der technischen Erstellung des Manuskriptes geholfen haben. Alle verbliebenen Fehler liegen natürlich in der alleinigen Verantwortung des Autors.

Bielefeld, im Juli 1985

Hans-Jürgen Andreß

Verzeichnis der Abbildungen und Tabellen

Abbildungen

1.1	Verlauf einer Datenanalyse	14
2.1	Logit-Transformation	30
2.2	Graphische Darstellung einer Kreuzklassifikation mit zentrierten Effekten (Beispiel 2.2)	55
2.3	Graphische Darstellung einer Kreuzklassifikation mit auf die erste Kategorie bezogenen Effekten (Beispiel 2.2)	57
2.4	Graphische Darstellung der Zusammenhänge zwischen konditionalen und Interaktionseffekten (Beispiel 2.4)	64
3.1	Die Suche nach einem passenden Modell	96

Tabellen

1.1	Multivariate Analyseverfahren	2
1.2	Variablen des Anwendungsbeispiels	8
1.3	Auszug aus der aggregierten Datenbasis	11
1.4	Art des Tätigkeitswechsels nach Typ der Qualifikation und Wirtschaftszweig	12
1.5	Anwendungen multivariater Analyseverfahren in diesem Buch	15
2.1	Spezielle Verteilung der exponentiellen Familie	23
2.2	Lage- und Streuungsparameter in Verteilungen der exponentiellen Familie	26
2.3	Verbindungsfunktionen	29
2.4	Transformationen des abhängigen Merkmals	31
4.1	Regressionsanalyse	122
4.2	Regressionsanalyse mit Dummy-Variablen (Varianzanalyse)	126
4.3	Einfluß des Wirtschaftszweiges und der Qualifikation auf die Veränderung des sozio-ökonomischen Status	128
4.4	Kovarianzanalyse	131
4.5	Einfluß der Qualifikation und des Beschäftigungswachstums auf die Veränderung des sozio-ökonomischen Status	132
4.6	Berufliche Abstiege nach Art der Qualifikation	141
4.7	Gewichtete Regression mit aggregierten Daten	151
4.8	Einfluß der Qualifikation und des Beschäftigungswachstums auf den Anteil beruflicher Abstiege	151
4.9	Einfluß der Qualifikation und des Beschäftigungswachstums auf den Anteil beruflicher Abstiege	153
4.10	Logistische Regression mit aggregierten Daten	158

4.11	Einfluß des Wirtschaftszweiges und der Qualifikation auf das Abstiegsrisiko	159
4.12	Einfluß des Wirtschaftszweiges und der Qualifikation auf die Wahrscheinlichkeit eines beruflichen Abstieges	162
4.13	Häufigkeiten verschiedener Qualifikationstypen und Berufswechsel	167
4.14	Äquivalente logistische und log-lineare Modelle	172
4.15	Einfluß der Qualifikation auf verschiedene Arten beruflicher Wechsel	175
4.16	Einfluß der Qualifikation auf die Rangordnung von Statusänderungen	178
4.17	Logistische Regressionsmodelle für continuation ratios	180
4.18	Berufliche Abstiege nach Art der Qualifikation (Individualdaten)	182
4.19	Häufigkeiten verschiedener Qualifikationstypen und Berufswechsel (Individualdaten)	184
4.20	Logistische Regression mit Individualdaten	186
5.1	Elemente der GLIM-Sprache	194

Inhaltsverzeichnis

1 Einleitung	1
1.1 Multivariate Analyseverfahren – Ein Überblick	2
1.2 Verallgemeinerte lineare Modelle und ihre Realisierung in dem Programmpaket GLIM	5
1.3 Ein Anwendungsbeispiel	7
1.3.1 Fragestellung und betrachtete Merkmale	7
1.3.2 Individualdaten, aggregierte Daten, Kreuztabellen	10
1.4 Gliederung dieses Buches und notwendige Vorkenntnisse	14
2 Spezifikation verallgemeinerter linearer Modelle: Theoretische Grundlagen	18
2.1 Struktur der verallgemeinerten linearen Modelle	18
2.2 Verteilungsannahmen bei verallgemeinerten linearen Modellen	21
2.3 Abhängigkeitsbeziehungen in verallgemeinerten linearen Modellen	27
2.4 Erklärende Variablen in verallgemeinerten linearen Modellen	31
2.4.1 Regressionsanalyse	36
2.4.2 Varianzanalyse (Regressionsanalyse mit Dummy-Variablen)	40
2.4.2.1 Reparametrisierung und Effektkodierung	42
2.4.2.2 Spezielle Designmatrizen: Interaktionen, konditionale und geordnete Effekte	51
2.4.2.3 Varianzanalyse und Kreuztabellenanalyse: Das Problem fehlender Beobachtungen	66
2.4.3 Kovarianzanalyse	69
2.4.4 Konstruktion der Matrix der unabhängigen Variablen innerhalb des Programmpakets GLIM	71
2.5 Spezifikation verallgemeinerter linearer Modelle – Zusammenfassung an Hand des Beispiels	75
3 Überprüfung verallgemeinerter linearer Modelle: Theoretische Grundlagen	80
3.1 Schätzung der Regressionsparameter	81
3.2 Konfidenzintervalle der Regressionsparameter und Hypothesentests	88
3.3 Modellanpassung	94
3.3.1 Exkurs: Berechnung der Devianz für zwei Submodelle des GLIM- Ansatzes	100
3.3.2 Modellanpassungstests	103
3.3.3 Schätzung des Skalenparameters und Signifikanztests für Zwei- Parameter-Verteilungen	106
3.4 Eigenschaften der Testverfahren	111

3.5 Multiple und partielle Bestimmtheitsmaße für verallgemeinere lineare Modelle	112
3.6 Analyse der Residuen	114
4 Anwendung verallgemeinerter linearer Modelle: Eine empirische Analyse beruflicher Wechsel	118
4.1 Metrische Zielvariablen: Statusänderungen bei Berufswechseln	119
4.1.1 Regressionsanalyse	119
4.1.2 Varianzanalyse	125
4.1.3 Kovarianzanalyse	132
4.2 Dichotome Zielvariablen: Abstiegsrisiko in verschiedenen Subgruppen	135
4.2.1 Probleme der Analyse dichotomer Zielvariablen	136
4.2.2 Analyse dichotomer Zielvariablen mit gewichteter Regression	143
4.2.2.1 Grundprinzipien des GSK-Ansatzes (Minimum-Chi-Quadrat-Methode)	143
4.2.2.2 Ein empirisches Beispiel: Ein lineares Modell des Abstiegsrisikos	149
4.2.3 Analyse dichotomer Zielvariablen mit Maximum-Likelihood-Schätzungen	154
4.2.3.1 Theoretische Grundlagen	155
4.2.3.2 Ein empirisches Beispiel: Ein logistisches Modell des Abstiegsrisikos	157
4.2.4 Unvollständige Kreuztabellen	161
4.3 Polytome Zielvariablen: Verschiedene Arten von Tätigkeitswechseln	165
4.3.1 Analyse polytomer Zielvariablen mit gewichteter Regression und Maximum-Likelihood-Schätzungen	173
4.3.2 Ein empirisches Beispiel: Ein multivariates logistisches Modell verschiedener Tätigkeitswechsel	173
4.3.3 Ordinale Zielvariablen	177
4.4 Diskrete Zielvariablen mit Individualdaten	180
4.4.1 Kreuztabellen und Individualdaten	181
4.4.2 Ein empirisches Beispiel: Ein logistisches Modell des individuellen Abstiegsrisikos	185
4.5 Fortgeschrittene Anwendungen verallgemeinerter linearer Modelle	188
5 GLIM als Programmiersprache	190
5.1 Sprachdefinition	191
5.2 Dateneingabe und Druckausgabe	198
5.3 Datenmanipulation	201
5.3.1 Berechnung von Variablen	202
5.3.2 Umkodierung von Variablen	204
5.3.3 Aggregation von Daten und Generierung von Kreuztabellen	206
5.3.4 Auswahl von Fällen	209
5.4 Macros	210

Statistischer Anhang	215
Datenanhang	231
Programmanhang	232
GLIM: Kommandoübersicht	236
Glossar	241
Literaturverzeichnis	244
Autorenregister	247
Stichwortregister	248