

Moderne nichtparametrische Verfahren der Risikoanalyse

Helge Toutenburg

Moderne nicht- parametrische Verfahren der Risikoanalyse

Eine anwendungsorientierte Einführung für Mediziner,
Soziologen und Statistiker

Mit 34 Abbildungen



Springer-Verlag Berlin Heidelberg GmbH

Professor Dr. Dr. Helge Toutenburg
Universität München
Institut für Statistik und
Wissenschaftstheorie
Akademiestraße 1
D-8000 München 40

ISBN 978-3-7908-0592-5 ISBN 978-3-642-58246-2 (eBook)
DOI 10.1007/978-3-642-58246-2

Dieses Werk ist urheberrechtlich geschützt. Die dadurch begründeten Rechte, insbesondere die der Übersetzung, des Nachdruckes, des Vortrags, der Entnahme von Abbildungen und Tabellen, der Funk- sendungen, der Mikroverfilmung oder der Vervielfältigung auf anderen Wegen und der Speicherung in Datenverarbeitungsanlagen, bleiben, auch bei nur auszugsweiser Verwertung, vorbehalten. Eine Vervielfältigung dieses Werkes oder von Teilen dieses Werkes ist auch im Einzelfall nur in den Grenzen der gesetzlichen Bestimmungen des Urheberrechtsgesetzes der Bundesrepublik Deutschland vom 9. September 1965 in der Fassung vom 24. Juni 1985 zulässig. Sie ist grundsätzlich vergütungspflichtig. Zuwiderhandlungen unterliegen den Strafbestimmungen des Urheberrechtsgesetzes.

© Springer-Verlag Berlin Heidelberg 1992

Ursprünglich erschienen bei Physica-Verlag Heidelberg 1992

Die Wiedergabe von Gebrauchsnamen, Handelsnamen, Warenbezeichnungen usw. in diesem Werk berechtigt auch ohne besondere Kennzeichnung nicht zu der Annahme, daß solche Namen im Sinne der Warenzeichen- und Markenschutz-Gesetzgebung als frei zu betrachten wären und daher von jedermann benutzt werden dürften.

7100/7130-543210 - Gedruckt auf säurefreiem Papier

Vorwort

Das vorliegende Buch entstand auf Anregung von Anwendern und Studenten und stellt eine Erweiterung und Vertiefung meines Vorlesungsskripts „Medizinische Statistik“ für den Studiengang Diplom-Statistik im Sommersemester 1991 an der Universität München dar.

Ein wesentliches Ziel ist die Information über Modelle, Studientypen und spezifische Verfahren der nichtparametrischen Statistik, die sich insbesondere mit Raten, Adjustierung von Raten, nichtzufälligem Datenverlust (Zensierung) und der Modellierung von zeitabhängigen Verläufen befassen.

Semiparametrische und parametrische Ansätze zur Einbeziehung von prognostischen Faktoren stellen grundlegende Erweiterungen der Gedanken der klassischen Regression dar. Mit der komplexen Modellanalyse von Kontingenztafeln über Wilk's G^2 , Logit- und Logistische Regression, Residualanalyse bis hin zur Risikomodellierung bei Lebensdauerdaten soll ein theoretisch fundierter, mit zahlreichen Beispielen unersetzter Überblick über moderne Verfahren dieses Zweiges der Statistik gegeben werden.

In dankenswerter Weise hat Herr Oberarzt Dr. Winfried Walther (Zahnärztliche Akademie Karlsruhe) aktuelle Datensätze über Risikofaktoren und Verlaufsdaten von Konuskronen zur Verfügung gestellt. Herr Christian Heumann hat zahlreiche Algorithmen programmiert sowie Beispiele gerechnet und kontrolliert. Herr Harald Huber hat mit großer Sorgfalt dieses Buchmanuskript geschrieben.

Die jetzt vorliegende Fassung erfüllt sicherlich nicht alle Wünsche, was Vollständigkeit und stärkere Behandlung von Beispielen aus der Soziologie oder anderen Bereichen wie etwa der Volkswirtschaft angeht. Dies liegt aber zum großen Teil darin begründet, daß die behandelten Methoden dort erst begrenzt eingesetzt werden und somit solide Datensätze kaum zugänglich sind.

Angesichts des Fehlens eines Buches mit vergleichbarem Inhalt auf dem deutschen Markt habe ich mich trotzdem entschlossen, das Buch in der jetzt vorliegenden Fassung zu publizieren — also unter momentanem Verzicht auf die Einbindung weiterer hochaktueller Stoffgebiete wie etwa Datenverlust durch Noncompliance und die notwendige Adjustierung durch Follow-up-Studien, Bindungseffekte durch Mehrfachbeobachtungen an einer Person oder Adjustierung von unvollständigen Kontingenztafeln in Abhängigkeit vom Fehlend-Mechanismus.

An alle Leser und Nutzer des Buches wende ich mich mit der Bitte, mich über Fehler oder andere Unzulänglichkeiten zu informieren.

Helge Toutenburg

München, im Oktober 1991

Inhaltsverzeichnis

1	Einführung	1
1.1	Spezifikation medizinischer Daten und Datenerhebung	1
1.2	Indikation für nichtparametrisches Vorgehen	3
1.3	Motivierende Beispiele	9
2	Kontingenztafeln	19
2.1	Rangskalierung	20
2.2	Unabhängigkeit	21
2.3	Methoden zum Vergleich von Anteilen	22
2.3.1	Relatives Risiko	23
2.3.2	Odds-Ratio	24
2.3.3	Beziehung zwischen Odds-Ratio und dem relativen Risiko	27
2.3.4	Der Odds-Ratio für $I \times J$ -Tafeln	27
2.3.5	Analyse von ordinalen Zusammenhängen (Rangdaten) . .	28
2.4	Untersuchung von Zweifachklassifikationen	32
2.4.1	Die Poisson-Stichprobe	32
2.4.2	Die unabhängige multinomiale Stichprobe	33
2.4.3	Likelihood-Funktion und Maximum-Likelihood-Schätzungen	34
2.5	Tests für die Güte der Anpassung	35
2.5.1	Prüfen einer spezifizierten Multinomialverteilung (theoretische Verteilung)	36
2.5.2	Güte der Anpassung bei geschätzten erwarteten Häufigkeiten	37
2.5.3	Prüfen auf Unabhängigkeit	38
2.5.4	Likelihood-Quotienten-Test	39
2.6	Differenziertere Untersuchung von $I \times J$ -Tafeln	40
2.6.1	Modellierung des ordinalen Zusammenhangs	44
3	Modelle für binäre Responsevariablen	51
3.1	Generalisierte lineare Modelle	51
3.2	GLM für binären Response	52
3.2.1	Logit-Modelle	52
3.2.2	Loglineare Modelle	53
3.2.3	Logistische Regression	54
3.2.4	Prüfen des Modells	58

3.2.5	Verwendung von Verteilungsfunktionen als Linkfunktion	59
3.3	Logitmodelle für kategoriale Daten	60
3.4	Güte der Anpassung — Likelihood-Quotienten-Test	63
3.5	Verwendung von eingeschränkten Alternativhypothesen	65
3.5.1	Cochran-Armitage-Trend-Test	66
4	Alternative Modelle und Modelldiagnostik	69
4.1	Probitmodelle	69
4.2	Modelle mit Log-Log-Link	70
4.3	Modell-Diagnostik	71
4.3.1	Diagnostik auf der Basis der Residuen	72
4.3.2	Diagnostik in Anlehnung an das Bestimmtheitsmaß	72
4.3.3	Beispiele für die Modelldiagnostik	74
4.4	ML-Schätzung für die logistische Regression	87
4.5	Newton-Raphson-Methode	89
5	Analyse von epidemiologischen und klinischen Daten — Untersuchung des Zusammenhangs zwischen Ereignis und Exposition	93
5.1	Einleitung	93
5.2	Studientypen in der Epidemiologie	94
5.3	Prüfung auf Unabhängigkeit von Exposition und Krankheit	97
5.4	Untersuchung des Odds-Ratio für mehrere 2×2-Tafeln	101
5.5	Standardisierung und Angleichung (Adjustierung) von Raten	104
6	Schätzen von Überlebenswahrscheinlichkeiten	115
6.1	Problemstellung	115
6.2	Survivorfunktion und Hazardrate (Ein-Episoden-Fall)	121
6.3	Sterbetafel-Methode	123
6.4	Kaplan-Meier-Schätzung	130
6.5	Nichtparametrische Methoden zum Vergleich von Überlebenskurven	134
6.6	Vergleich der Methoden	135
6.7	Log-Rank-Statistik zum Vergleich von Survivorfunktionen	136
6.8	Vergleich von mehr als zwei Überlebenskurven — die Methode von Peto-Pike	139
6.9	Relation zwischen Überlebenskurven	141
6.10	Standardfehler und Konfidenzbänder für Überlebensraten und -kurven	142
6.10.1	Standardfehler der Sterbetafel-Überlebensrate	142
6.10.2	Konfidenzbereiche für die nach Kaplan-Meier geschätzte Survivorkurve	147
6.11	Einbeziehung von Kovariablen in die Überlebensanalyse	156
6.11.1	Das Proportional-Hazard-Modell von Cox	157
6.11.2	Überprüfung der Proportionalitätsannahme	158
6.11.3	Schätzung des Cox-Modells	159

6.11.4	Schätzung der Überlebensfunktion unter dem Cox-Ansatz	159
6.11.5	Einige Wahrscheinlichkeitsverteilungen für die Verweil- dauer	160
6.11.6	Modellierung der Hazardrate	162
7	Loglineare Modelle	165
7.1	Zweifache Klassifikation	165
7.2	Dreifache Klassifikation	168
7.3	Parameterschätzung im loglinearen Modell	172
7.4	Der Spezialfall des binären Response	174
7.4.1	Iterative Proportionale Anpassung (IPA)	175
7.4.2	Einbeziehung von kumulierten Verweildauern — Analyse von Raten	176
7.5	Kodierung kategorialer Einflußvariablen	189
7.5.1	Dummy- und Effektkodierung	189
7.5.2	Kodierung von Responsemodellen	192
7.5.3	Kodierung von Modellen für die Hazardrate	194
	Anhang: χ^2-Verteilung	199
	Literaturverzeichnis	200
	Sachregister	205