

Inhaltsverzeichnis

Vorwort	V
Verzeichnis der Beispiele	XIII
1 Einführung	1
1.1 Problemstellung und einführende Beispiele	1
1.1.1 Merkmalstypen	1
1.1.2 Zusammenhangs- und Regressionsanalyse	3
1.1.3 Einige Beispiele	4
1.1.4 Übersicht	8
1.2 Binomialverteilung, Multinomialverteilung und einfache Schätzer	9
1.2.1 Binomialverteilung und Chancen	9
1.2.2 Multinomialverteilung	13
1.3 Elementare Konzepte der metrischen Regression	15
1.3.1 Das Modell	15
1.3.2 Schätzung und Kennwerte	17
1.3.3 Die Kodierung qualitativer Einflußgrößen	18
2 Logistische Regression und Logit-Modell für binäre abhängige Größe	29
2.1 Logit-Modell für eine metrische Einflußgröße	32
2.1.1 Modelldarstellungen und Parameter	32
2.1.2 Elastizität des Logit-Modells	41
2.2 Modelle für linear spezifizierte Einflußgrößen	42
2.2.1 Logit-Modell mit linearem Einflußterm	42
2.2.2 Nichtmonotone Logitmodellierung	44
2.3 Logit-Modell bei kategorialen Einflußgrößen	45
2.3.1 Dichotome Einflußgröße im Logit-Modell	45
2.3.2 Mehrkategoriale Einflußgröße	51
2.4 Das lineare Logit-Modell ohne Interaktionen	58
2.4.1 Parameter und Interpretation	58
2.4.2 Abhängigkeit der Parameter von den einbezogenen Kova- riablen	64
2.5 Logit-Modell und Alternativen	65

3	Schätzung, Modellanpassung und Einflußgrößenanalyse	69
3.1	Parameterschätzung für Regressionsmodelle	72
3.1.1	Maximum-Likelihood-Schätzung	74
3.1.2	Gewichtete Kleinste-Quadrate-Schätzung	77
3.1.3	Minimum Logit-Schätzer	78
3.1.4	Schätzverfahren mit Penalisierung	79
3.1.5	Retrospektive Datenerhebung	80
3.2	Anpassungsgüte von Modellen	81
3.2.1	Anpassungstests	83
3.2.2	Akaike- und Schwarz-Kriterium	88
3.3	Residualanalyse	90
3.4	Überprüfung der Relevanz von Einflußgrößen: die lineare Hypothese	94
3.5	Devianz-Analyse	97
3.6	Erklärungswert von Modellen	103
3.6.1	Effektivitätsmaße vom Typ R^2	103
3.6.2	Prognosemaße	108
3.6.3	Rangkorrelationsmaße	111
3.7	Ergänzungen und Literatur	113
4	Alternative Modellierung von Response und Einflußgrößen	115
4.1	Konzeptioneller Hintergrund binärer Regressionsmodelle	116
4.1.1	Binäre Regressionsmodelle als Schwellenwertmodelle	117
4.1.2	Binäre Wahlmodelle als Modelle der Nutzenmaximierung	119
4.2	Modelltypen	121
4.2.1	Probit-Modell	122
4.2.2	Lineares Modell	123
4.2.3	Extremwertverteilungsmodelle	124
4.2.4	Komplementäres Log-Modell oder Exponentialmodell	125
4.2.5	Responsefunktion und Eindeutigkeit der Modelle	128
4.3	Die Modellierung von Interaktionswirkungen	132
4.3.1	Interaktionsmodell für gemischte Variablen	133
4.3.2	Mehrfaktorielle Modellierung mit Interaktion	139
4.4	Abweichung von der Binomialverteilung: Überdispersion	147
4.4.1	Korrelierte binäre Beobachtungen	149
4.4.2	Variabilität durch latente Variablen: verteilungsunspezifische Modellierung	149
4.4.3	Variabilität durch latente Variable: das Beta-Binomial-Modell	151
4.4.4	Generalisierte Schätzgleichungen und Quasi-Likelihood	152
4.4.5	Vernachlässigte Einflußgrößen und zufällige Effekte	154
4.5	Ergänzende Bemerkungen	156

5	Multinomiale Modelle für ungeordnete Kategorien	159
5.1	Modellbildung bei mehrkategorialer abhängiger Variable	160
5.2	Das multinomiale Logit-Modell	162
5.3	Das multinomiale Modell mit kategorienspezifischen Charakteristiken	174
5.4	Das multinomiale Logit-Modell als verallgemeinertes lineares Modell	176
5.5	Einfache Verzweigungsmodelle	178
5.6	Modellierung als Wahlmodelle der Nutzenmaximierung	180
5.6.1	Probabilistische Wahlmodelle	180
5.6.2	Paarvergleichsmodelle	185
5.6.3	Unabhängige Störgrößen in probabilistischen Wahlmodellen	187
5.6.4	Modell der Elimination von Aspekten	192
5.6.5	Verzweigungsmodelle und das genestete Logit-Modell . . .	194
5.7	Schätzen und Testen für multinomiale Modelle	197
5.7.1	Maximum-Likelihood Schätzung	198
5.7.2	Anpassungstests und Residuen	199
5.7.3	Einflußgrößenanalyse	202
5.8	Ergänzungen und weitere Literatur	202
6	Regression mit ordinaler abhängiger Variable	205
6.1	Das Schwellenwert- oder kumulative Modell	209
6.1.1	Ableitung des Modells	209
6.1.2	Spezielle kumulative Modelle	214
6.1.3	Verallgemeinerte Schwellenwertmodelle: kategorienspezi- fische Parameter	219
6.2	Das sequentielle Modell	221
6.2.1	Ableitung des Modells	222
6.2.2	Spezielle sequentielle Modelle	223
6.2.3	Verallgemeinerte sequentielle Modelle	225
6.2.4	Schätzung sequentieller Modelle als binäre Modelle	226
6.3	Schätzen und Testen für ordinale Modelle	231
6.3.1	Kumulative Modelle	232
6.3.2	Sequentielle Modelle	234
6.4	Ordinale Modelle versus klassisches lineares Regressionsmodell .	235
6.5	Kumulatives versus sequentielles Modell	239
6.6	Bemerkungen und weitere Literatur	241
7	Zählraten und die Analyse von Kontingenztafeln: das loglineare Modell	243
7.1	Die Poisson-Verteilung	245
7.2	Poisson-Regression	249
7.2.1	Das Grund-Modell der Poisson-Regression	249
7.2.2	Maximum-Likelihood-Schätzung	249

7.2.3	Poisson-Regression versus metrischer Regression	251
7.2.4	Poisson-Regression mit zusätzlichem Parameter	255
7.2.5	Einflußgrößen und Modellanpassung	257
7.3	Poisson-Regression mit Dispersion	259
7.3.1	Unbeobachtete Heterogenität: das konjugierte Gamma-Poisson-Modell	259
7.3.2	Unbeobachtete Heterogenität: zufällige Effekte	260
7.3.3	Dispersionsmodellierung	261
7.4	Analyse von Kontingenztafeln	263
7.4.1	Typen der Kontingenztafel-Analyse	263
7.4.2	Das zweidimensionale loglineare Modell	268
7.4.3	Drei- und höherdimensionale Modelle	270
7.4.4	Loglineare und Logit-Modelle	280
7.5	Inferenz in loglinearen Modellen	283
7.6	Ergänzende Bemerkungen	284
8	Nonparametrische Regression I: Glättungsverfahren	287
8.1	Lokale Regression für binäre abhängige Variable	289
8.1.1	Grundkonzept lokaler Anpassung	289
8.1.2	Kerngewichte und Glättungssteuerung	291
8.1.3	Ausgleich zwischen Varianz und Verzerrung	294
8.1.4	Polynomiale Erweiterung	297
8.1.5	Erweiterung auf mehrere Kovariablen	299
8.2	Ansätze mit Penalisierung	300
8.2.1	Penalisierung für metrischen Response	300
8.2.2	Penalisierung für kategorialen Response	303
8.3	Semiparametrische Erweiterung: Das partiell lineare Modell	304
8.4	Generalisierte additive Modelle	304
8.5	Modellierung effektmodifizierender Variablen	305
8.6	Schätzalgorithmen	310
8.6.1	Iterative Schätzung der penalisierten Likelihood	310
8.6.2	Backfitting-Algorithmus für additive Modelle	311
8.6.3	Backfitting-Algorithmus für generalisierte additive Modelle	313
8.7	Weitere Literatur	314
9	Nonparametrische Regression II: Klassifikations- und Regressionsbäume	317
9.1	Grundkonzept	317
9.2	Zugelassene Verzweigungen	319
9.3	Verzweigungskriterium	321
9.3.1	Tests auf Homogenität als Verzweigungskriterium	321

9.3.2	Weitere Verzweigungskriterien	328
9.4	Größe eines Baumes	330
9.5	Ergänzungen	332
10	Kategoriale Prognose und Diskriminanzanalyse	337
10.1	Bayes-Zuordnung als diskriminanzanalytisches Verfahren	340
10.1.1	Grundkonzept	340
10.1.2	Bayes-Zuordnung und Fehlerraten	342
10.1.3	Fehlklassifikationswahrscheinlichkeiten	343
10.1.4	Bayes-Regel und Diskriminanzfunktionen	345
10.1.5	Logit-Modell und normalverteilte Merkmale	348
10.1.6	Logit-Modell und binäre Merkmale	349
10.1.7	Grenzen der Bayes Zuordnung: Maximum-Likelihood-Regel	350
10.1.8	Kostenoptimale Bayes-Zuordnung	353
10.2	Geschätzte Zuordnungsregeln	356
10.2.1	Stichproben und geschätzte Zuordnungsregeln	356
10.2.2	Prognosefehler – Direkte Prognose der Klassenzugehörigkeit	358
10.2.3	Prognosefehler – alternative Schadensfunktionen	360
11	Elemente der Schätz- und Testtheorie	369
11.1	Intervallschätzung für Wahrscheinlichkeiten, Chancen und Logits .	369
11.2	Schätzung für binäre und mehrkategoriale Regressionsmodelle . . .	374
11.2.1	Prinzip der Maximum-Likelihood-Schätzung	374
11.2.2	Maximum-Likelihood-Schätzung – Binäre Modelle	376
11.2.3	Maximum-Likelihood-Schätzung – mehrkategoriale Modelle	381
11.2.4	Kleinste-Quadrate-Schätzung – binäre Modelle	384
11.2.5	Kleinste-Quadrate-Schätzung – mehrkategoriale Modelle . .	388
11.3	Schätzung im Rahmen des generalisierten linearen Modells	390
11.3.1	Das generalisierte lineare Modell	390
11.3.2	Maximum Likelihood-Schätzung	392
11.3.3	Anpassungstest	394
11.4	Schätzung loglinearer Modelle für Kontingenztafeln	395
11.5	Überdispersion und Quasi-Likelihood	397
Anhang A:	Verteilungen	399
A.1	Binomialverteilung	399
A.2	Multinomialverteilung	399
A.3	Poisson-Verteilung	400
A.4	Negative Binomialverteilung	401
A.5	Multivariate Normalverteilung	401
A.6	Logistische Verteilung	402

A.7	Exponentialverteilung	403
A.8	Extremwertverteilungen	403
A.9	Dirichlet- und Beta-Verteilung	404
A.10	Gamma-Verteilung	406
Anhang B:	Einige statistische Werkzeuge	409
B.1	Lineare Algebra	409
B.2	Formel von Taylor	411
B.3	Delta Methode	412
Anhang C:	Verwendete Daten	413
C.1	Staubbelastung	413
C.2	Dauer der Arbeitslosigkeit	413
C.3	Pkw-Besitz	414
C.4	Klinische Schmerzstudie	414
C.5	Parteipräferenzen	415
C.6	Paarvergleichsdaten	415
C.7	Bewertung von Fachbereichen	415
C.8	Neugegründete Unternehmen	415
C.9	Konkurse in Berlin	417
C.10	Habilitationen in Deutschland	417
C.11	Umsatzstärkste Unternehmen in der BRD	417
C.12	Konjunktur-Surveys	417
C.13	Herpes Encephalitiden	418
Anhang D:	Software	419
D.1	SAS®-Programmpaket	419
D.2	BMDP	420
D.3	GENSTAT	420
D.4	EGRET	421
D.5	GLIM4	421
D.6	GLAMOUR	421
D.7	SPSS/PC+	422
D.8	LIMDEP	422
D.9	XPLORE	422
D.10	S-PLUS	422
Literaturverzeichnis		424
Autorenindex		439
Sachindex		443