

Ludwig Fahrmeir
Thomas Kneib
Stefan Lang

Regression

Modelle, Methoden
und Anwendungen

Zweite Auflage

fyA Springer

Inhaltsverzeichnis

1	Einführung	1
1.1	Anwendungsbeispiele	4
1.2	Erste Schritte	11
1.2.1	Beschreibung der Verteilung der Variablen	11
1.2.2	Grafische Zusammenhangsanalyse	13
	Stetige erklärende Variablen	13
	Kategoriale erklärende Variablen	16
2	Regressionsmodelle	19
2.1	Einführung	19
2.2	Lineare Regressionsmodelle	20
2.2.1	Das einfache lineare Regressionsmodell	20
2.2.2	Das multiple lineare Regressionsmodell	24
2.3	Regression bei binären Zielvariablen: Das Logit-Modell	30
2.4	Gemischte Modelle	35
2.5	Einfache nichtparametrische Regression	40
2.6	Additive Regression	44
2.7	Generalisierte additive Regression	47
2.8	Geoadditive Regression	49
2.9	Modelle im Überblick	55
2.9.1	Lineare Modelle (LM, Kapitel 3)	55
2.9.2	Logit-Modell (Kapitel 4)	56
2.9.3	Poisson-Regression (Kapitel 4)	56
2.9.4	Generalisierte lineare Modelle (GLM, Kapitel 4, 5)	56
2.9.5	Lineare gemischte Modelle (LMM, Kapitel 6)	56
2.9.6	Additive Modelle und Erweiterungen (AM, Kapitel 7, 8)	57
2.9.7	Generalisierte additive (gemischte) Modelle (GAMM, Kapitel 8)	58
2.9.8	Strukturiert-additive Regression (STAR, Kapitel 8)	58
3	Lineare Regressionsmodelle	59
3.1	Das klassische lineare Modell	59
3.1.1	Modelldefinition	59
3.1.2	Modellparameter, Schätzungen und Residuen	63
3.1.3	Diskussion der Modellannahmen	64
	Linearität des Einflusses der Kovariablen	64
	Homoskedastische Varianz der Störgrößen	64

Inhaltsverzeichnis

	Unkorreliertheit der Störgrößen	66
	Additivität der Störgrößen	70
3.1.4	Modellierung des Einflusses der Kovariablen	72
	Metrische Kovariablen	72
	Kategoriale Kovariablen	80
	Interaktionen zwischen Kovariablen	83
3.2	Parameterschätzungen	90
3.2.1	Schätzung der Regressionskoeffizienten	90
	Die Methode der kleinsten Quadrate	90
	Maximum-Likelihood-Schätzung	92
	Geschätzte Werte und Residuen	93
3.2.2	Schätzung der Varianz der Störgrößen	94
	Maximum-Likelihood-Schätzung	94
	Restringierte Maximum-Likelihood-Schätzung	94
3.2.3	Eigenschaften der Schätzungen	95
	Geometrische Eigenschaften des KQ-Schätzers	95
	Streuungszerlegung und Bestimmtheitsmaß	98
	Statistische Eigenschaften ohne spezielle Verteilungsannahmen	101
	Statistische Eigenschaften bei Normalverteilungsannahme	103
	Asymptotische Eigenschaften des KQ-Schätzers	105
	Statistische Eigenschaften der Residuen	107
	Standardisierte und studentisierte Residuen	108
3.3	Hypothesentests und Konfidenzintervalle	111
3.3.1	F-Test	113
	Zusammenhang mit dem Wald-Test	115
	F-Test für einige spezielle Testprobleme	115
	Asymptotische Eigenschaften des F-Tests	119
3.3.2	Konfidenzbereiche und Prognoseintervalle	119
	Konfidenzintervalle und Ellipsoide für die Regressionskoeffizienten	119
	Prognoseintervalle	121
3.4	Das allgemeine lineare Regressionsmodell	124
3.4.1	Modelldefinition	124
3.4.2	Gewichtete Methode der kleinsten Quadrate	125
	Gruppierte Daten	127
3.4.3	Heteroskedastische Fehler	128
	Diagnose heteroskedastischer Fehler	129
	Maßnahmen bei Heteroskedastizität	132
3.4.4	Autokorrelierte Fehler	136
	Autokorrelation erster Ordnung	? 137
	Diagnose autokorrelierter Störungen	139
	Maßnahmen bei Autokorrelation erster Ordnung	142

3.5	Bayesianische lineare Modelle	147
3.5.1	Priori-Verteilungen	147
3.5.2	Vollständig bedingte Dichten und MCMC-Inferenz	149
3.5.3	Posteriori-Verteilung	152
3.6	Modellwahl und Variablenselektion	152
3.6.1	Auswirkungen Bias, Varianz und Prognosegüte	155
	Auswirkungen auf Bias und Varianz des KQ-Schätzers $\hat{\beta}$	156
	Auswirkung der Modellspezifikation auf die Prognosegüte	157
3.6.2	Modellwahlkriterien	159
	Das korrigierte Bestimmtheitsmaß	160
	Mallows' C_p	161
	Informationskriterium nach Akaike AIC	161
	Kreuzvalidierung	161
	Bayesianisches Informationskriterium BIC	162
3.6.3	Praktische Verwendung der Modellwahlkriterien	163
3.6.4	Modelldiagnose	168
	Überprüfen der Modellannahmen	168
	Kollinearitätsanalyse	170
	Ausreißer- und Einflussanalyse	173
	Alternative Modellierungsansätze nach Modelldiagnose	179
3.7	Bemerkungen und Ergänzungen	180
3.7.1	Literaturhinweise	180
3.7.2	Beweise	181
4	Generalisierte lineare Modelle	189
4.1	Binäre Regression	189
4.1.1	Binäre Regressionsmodelle	189
	Logit-Modell	190
	Probit-Modell	191
	Komplementäres log-log-Modell	191
	Binäre Modelle als Schwellenwertmodelle latenter linearer Modelle	193
	Parameterinterpretation	194
	Gruppierte Daten	195
	Überdispersion (Overdispersion)	197
4.1.2	Maximum-Likelihood-Schätzung	198
	Vergleich mit ML- bzw. KQ-Schätzung im linearen Modell	201
	Iterative numerische Berechnung des ML-Schätzers	202
	Asymptotische Eigenschaften des ML-Schätzers	203
4.1.3	Testen linearer Hypothesen	204
4.1.4	Kriterien zur Modellanpassung und Modellwahl	205
4.2	Regression für Zähldaten	210

- 4.2.1 Modelle für Zähldaten 210
 - Log-lineares Poisson-Modell 210
 - Lineares Poisson-Modell 210
 - Überdispersion 210
- 4.2.2 Schätzen und Testen: Likelihood-Inferenz 212
 - Maximum-Likelihood-Schätzung 212
 - Testen linearer Hypothesen 213
 - Kriterien zur Modellanpassung und Modell wähl. 213
 - Schätzung des Überdispersions-Parameters 213
- 4.3 Modelle für positive stetige Zielvariablen 215
 - Gamma-Regression 217
 - Inverse Gauß-Verteilung 217
- 4.4 Generalisierte Lineare Modelle 217
 - 4.4.1 Allgemeine Modelldefinition 217
 - 4.4.2 Likelihood-Inferenz 220
 - Asymptotische Eigenschaften des ML-Schätzers 223
 - Schätzung des Skalierungs- oder Überdispersionsparameters. 224
 - Testen linearer Hypothesen 224
 - Kriterien zur Modellanpassung und Modellwahl. 225
- 4.5 Quasi-Likelihood-Modelle 226
- 4.6 Bayesianische generalisierte lineare Modelle 228
- 4.7 Bemerkungen und Ergänzungen 233

- 5 Kategoriale Regressionsmodelle 235**
 - 5.1 Einführung 235
 - Multinomialverteilung 236
 - Daten 237
 - 5.2 Modelle für ungeordnete Kategorien 238
 - Nominale Modelle und latente Nutzenmodelle 241
 - 5.3 Ordinale Modelle 242
 - Das kumulative oder Schwellenwert-Modell 242
 - Das sequentielle Modell 245
 - 5.4 Schätzen und Testen: Likelihood-Inferenz 247
 - Numerische Bestimmung des ML-Schätzers. 249
 - Asymptotische Eigenschaften und Tests linearer Hypothesen. 249
 - 5.5 Bemerkungen und Ergänzungen 252

- 6 Gemischte Modelle 253**
 - 6.1 Lineare gemischte Modelle für Longitudinal- und Clusterdaten 254
 - 6.2 Das allgemeine lineare gemischte Modell 259
 - 6.3 Likelihood-Inferenz für LMM 261
 - 6.3.1 Schätzung bei bekannter Kovarianzstruktur. 261

6.3.2	Schätzung der Kovarianzstruktur	263
6.3.3	Schätzung fixer und zufälliger Effekte	264
6.3.4	Hypothesentests	266
6.4	Likelihood-Inferenz für Longitudinal- und Clusterdaten-Modelle	268
6.5	Bayesianische gemischte lineare Modelle	271
	Posteriori-Verteilung bei bekannter Kovarianzstruktur	273
	Empirische Bayes-Schätzung	273
	Volle Bayes-Schätzung	274
6.6	Generalisierte lineare gemischte Modelle	278
	6.6.1 Definition und Eigenschaften von GLMM	278
	GLMM für Longitudinal- und Clusterdaten	279
	GLMM in allgemeiner Form	279
	Kategoriale gemischte Regressionsmodelle	282
6.7	Likelihood- und Bayes-Inferenz in GLMM	284
	6.7.1 Penalisierte Likelihood- und empirische Bayes-Schätzung	284
	6.7.2 Volle Bayes-Inferenz mit MCMC	287
6.8	Bemerkungen und Ergänzungen	289
 Nichtparametrische Regression		 291
7.1	Univariate Glättung	292
	7.1.1 Polynom-Splines	293
	Polynom-Splines und trunkeerte Potenzen	296
	Einfluss der Knoten auf die Schätzung	301
	B-Splines	303
	7.1.2 Penalisierte Splines (P-Splines)	306
	P-Splines basierend auf der TP-Basis	307
	P-Splines basierend auf B-Splines	309
	Penalisierte KQ-Schätzung	311
	Bayesianische P-Splines	316
	7.1.3 Allgemeine Penalisierungsansätze	320
	7.1.4 Glättungssplines	323
	7.1.5 Random Walks	326
	7.1.6 Kriging	327
	Klassisches Kriging	327
	Kriging als Glättungsverfahren für Zeitreihen	330
	Kriging als Glättungsverfahren der nichtparametrische Regression	331
	7.1.7 Lokale Glättungsverfahren	333
	Nächste-Nachbarn-Schätzer	333
	Lokal polynomiale Regression und Nadaraya-Watson-Schätzer	335
	Loess	339
	7.1.8 Allgemeine Streudiagramm-Glätter	340

	Lineare Glättungsverfahren	340
	Konfidenzintervalle und -bänder	342
	Äquivalente Freiheitsgrade (effektive Parameterzahl)	345
	Schätzung der Fehlervarianz	347
	Bias-Varianz-Trade Off	348
7.1.9	Wahl des Glättungsparameters	350
	Glättungsparameterwahl basierend auf Optimalitätskriterien	350
	Repräsentation von Penalisierungsansätzen als gemischte Modelle	353
	Bayesianische Glättungsparameterwahl basierend auf MCMC	357
7.1.10	Adaptive Verfahren	359
	Multivariate adaptive Regressions-Splines (MARS)	359
	Regressionsbäume	361
	Bayesianische adaptive Verfahren I: Model Averaging	364
	Bayesianische adaptive Verfahren II: Reversible Jump MCMC	366
7.2	Bivariate Glättung und räumliche Effekte	368
7.2.1	Tensorprodukt-P-Splines	371
	Tensorprodukt-Basen	371
	2D-Penalierungsansätze	375
7.2.2	Radiale Basisfunktionen	379
7.2.3	Kriging	381
	Klassische Geostatistik	382
	Kriging als Basisfunktionenansatz	384
	Schätzung von Kriging-Modellen	385
7.2.4	Markov-Zufallsfelder	387
	Nachbarschaften und penalisiertes KQ-Kriterium	387
	Bayesianische Modellformulierung	389
	Räumlich autoregressive Prozesse	393
7.2.5	Fazit	393
7.2.6	Lokale und adaptive Glättungsverfahren	394
7.3	Höherdimensionale Glättung	395
7.4	Bemerkungen und Ergänzungen	397
	Strukturiert-additive Regression	399
8.1	Additive Modelle	399
8.2	Geoadditve Regression	404
8.3	Modelle mit Interaktionen	407
	8.3.1 Modelle mit variierenden Koeffizienten	408
	8.3.2 Interaktion zwischen zwei metrischen Kovariablen	410
8.4	Strukturiert-additive Regression	413
8.5	Inferenz	419
	8.5.1 Penalisierte KQ- bzw- Likelihood-Schätzung	420

Backfitting	420
Direkte Minimierung des penalisierten KQ-Kriteriums.	421
Generalisierte STAR-Modelle.	422
Schätzung der Glättungsparameter.	422
Modellwahl und Diagnose.	423
8.5.2 Inferenz basierend auf der Repräsentation als gemischtes Modell.	423
Modellwahl und Diagnose.	425
8.5.3 Bayesianische Inferenz mit MCMC.	425
Normalverteilte Zielgrößen.	425
Latente normalverteilte Zielgrößen.	427
Nicht-normalverteilte Zielgrößen.	428
Modellwahl und Diagnose.	428
8.5.4 Software-Hinweise.	430
8.6 Fallstudie: Unterernährung in Sambia	431
8.6.1 Hinweise zur grundsätzlichen Vorgehensweise.	431
Deskriptive Analyse der Rohdaten.	431
Datenaufbereitung.	431
Grafische zweidimensionale Zusammenhangsanalyse.	432
Schätzung erster Arbeitsmodelle.	432
Modelldiagnose und Verfeinerung der Arbeitsmodelle.	432
Darstellung der Ergebnisse.	434
8.6.2 Deskriptive Analysen.	435
8.6.3 Modellierungsvarianten.	437
8.6.4 Schätzergebnisse und Modellevaluation.	438
8.7 Bemerkungen und Ergänzungen.	443
A Matrix-Algebra	445
A.1 Definition und elementare Operationen.	445
A.2 Der Rang einer Matrix.	449
A.3 Determinante und Spur einer Matrix.	451
A.4 Verallgemeinerte Inverse.	452
A.5 Eigenwerte und Eigenvektoren.	453
A.6 Quadratische Formen.	455
A.7 Differentiation von Matrixfunktionen.	457
B Wahrscheinlichkeitsrechnung und induktive Statistik	459
B.1 Einige eindimensionale Verteilungen.	459
B.2 Zufallsvektoren.	461
B.3 Die multivariate Normalverteilung.	464
B.3.1 Definition und Eigenschaften.	464
B.3.2 Die singuläre Normalverteilung.	465
B.3.3 Verteilungen quadratischer Formen.	466

B.3.4	Multivariate t-Verteilung	467
B.4	Likelihood-Inferenz	467
B.4.1	Maximum-Likelihood-Schätzung	467
B.4.2	Numerische Berechnung des ML-Schätzers	473
B.4.3	Asymptotische Eigenschaften des ML-Schätzers	475
B.4.4	Likelihood-basierte Tests für lineare Hypothesen	475
B.4.5	Modellwahl	477
B.5	Bayes-Inferenz	478
B.5.1	Grundlagen der Bayes-Inferenz	478
B.5.2	Punkt- und Intervallschätzer	480
Punktschätzer		480
Intervallschätzung		481
B.5.3	MCMC-Methoden	482
Metropolis-Hostings-Algorithmus		483
Gibbs-Sampler und Hybrid-Algorithmen		486
B.5.4	Modellwahl	488
Literaturverzeichnis		491
Index	?	497