

Inhaltsverzeichnis

1. Einleitung	1
2. Beziehungen zwischen zwei Variablen	5
2.1 Einleitung—Beispiele	5
2.2 Darstellung der Verteilung zweidimensionaler Merkmale	7
2.2.1 Kontingenztafeln bei diskreten Merkmalen	8
2.2.2 Grafische Darstellung bei diskreten Merkmalen	12
2.2.3 Maßzahlen zur Beschreibung der Verteilung bei stetigen und gemischt stetig-diskreten Merkmalen	13
2.2.4 Grafische Darstellung der Verteilung stetiger bzw. gemischt stetig-diskreter Merkmale	15
2.3 Maßzahlen für den Zusammenhang zweier nominaler Merkmale	18
2.3.1 Pearsons χ^2 -Statistik	19
2.3.2 Der Odds-Ratio	23
2.4 Rangkorrelationskoeffizient von Spearman	28
2.5 Zusammenhang zwischen zwei stetigen Merkmalen	32
3. Deskriptive univariate lineare Regression	41
3.1 Einleitung	41
3.2 Plots und Hypothesen	44
3.3 Prinzip der kleinsten Quadrate	45
3.3.1 Bestimmung der Schätzungen	47
3.3.2 Herleitung der Kleinst-Quadrate-Schätzungen	47
3.3.3 Eigenschaften der Regressionsgeraden	50

3.4	Güte der Anpassung	53
3.4.1	Varianzanalyse	53
3.4.2	Korrelation	56
3.5	Residualanalyse	62
3.6	Lineare Transformation der Originaldaten	62
3.7	Multiple lineare Regression und nichtlineare Regression	64
3.8	Polynomiale Regression	66
3.9	Lineare Regression mit kategorialen Regressoren	70
3.10	Spezielle nichtlineare Modelle	74
3.10.1	Wachstumskurven	74
3.10.2	Zeit als Regressor	76
3.11	Zeitreihen	78
3.11.1	Einleitung	78
3.11.2	Kurvendiagramme	78
3.11.3	Zerlegung von Zeitreihen	79
3.11.4	Fehlende Werte, äquidistante Zeitpunkte	80
3.11.5	Gleitende Durchschnitte	81
3.11.6	Saisonale Komponente, konstante Saisonfigur	82
3.11.7	Modell für den linearen Trend	87
4.	Das klassische multiple lineare Regressionsmodell.....	89
4.1	Deskriptive multiple lineare Regression	89
4.2	Prinzip der kleinsten Quadrate	90
4.3	Geometrische Eigenschaften der Kleinste-Quadrat-Schätzung	95
4.4	Beste lineare erwartungstreue Schätzung	102
4.4.1	Lineare Schätzer	102
4.4.2	Mean-Square-Error	103
4.4.3	Beste lineare erwartungstreue Schätzung	105
4.4.4	Schätzung von σ^2	111
4.5	Multikollinearität	112

4.5.1	Extreme Multikollinearität und Schätzbarkeit	112
4.5.2	Schwache Multikollinearität	114
4.5.3	Identifikation und Quantifizierung von Multikollinearität	118
4.6	Ökonometrische Gleichungen vom Regressionstyp	125
4.6.1	Stochastische Regressoren	125
4.6.2	Instrumental-Variablen Schätzer (IVS)	125
4.6.3	Scheinbar unverbundene Regressionen	126
4.7	Klassische Normalregression	128
4.8	Prüfen von linearen Hypothesen	131
4.9	Varianzanalyse und Güte der Anpassung	138
4.9.1	Univariate Regression	138
4.9.2	Univariate Regression mit einer Dummyvariablen	143
4.9.3	Multiple Regression	145
4.9.4	Ein komplexes Beispiel	149
4.9.5	Grafische Darstellung	153
4.10	Tests auf Parameterkonstanz	155
4.10.1	Der Prognosetest von Chow	155
4.10.2	Der Test von Hansen	160
4.10.3	Tests mit rekursiver Schätzung	164
4.10.4	Tests mit Prognosefehlern	164
4.10.5	CUSUM und CUSUMSQ-Tests	165
4.10.6	Tests auf Strukturwechsel	167
4.11	Die kanonische Form	176
4.12	Methoden zur Überwindung von Multikollinearität	177
4.12.1	Hauptkomponenten-Regression	177
4.12.2	Ridge-Schätzung	178
4.12.3	Shrinkage-Schätzer	183
4.13	Minimax-Schätzung	184
4.13.1	Ungleichungsrestriktionen	184

4.13.2	Das Minimaxprinzip	187
5.	Modelle der Varianzanalyse	195
5.1	Varianzanalyse als spezielles lineares Modell	195
5.2	Einfaktorielle Varianzanalyse	196
5.2.1	Darstellung als restriktives Modell	197
5.2.2	Zerlegung der Fehlerquadratsumme	199
5.2.3	Schätzung von σ^2 durch MQ_{Residual}	203
5.3	Vergleich von einzelnen Mittelwerten	205
5.3.1	Lineare Kontraste	205
5.3.2	Kontraste in den totalen (summierten) Responsewerten im balanzierten Fall	209
5.4	Multiple Vergleiche	215
5.4.1	Einleitung	215
5.4.2	Experimentweise Vergleiche	216
5.4.3	Vergleichsbezogene Prozeduren	217
5.5	Rangvarianzanalyse im vollständig randomisierten Versuchsplan	221
5.5.1	Kruskal–Wallis–Test	221
5.5.2	Multiple Vergleiche	225
5.6	Zwei- und Mehrfaktorielle Varianzanalyse	227
5.7	Zweifaktorielle Experimente mit Wechselwirkung (Modell mit festen Effekten)	231
5.8	Zweifaktorielles Experiment in Effektkodierung	237
5.9	2^k -faktorielles Experiment	245
5.9.1	Spezialfall: 2^2 -Experiment	246
5.9.2	Das 2^3 -Experiment	248
6.	Exakte und stochastische lineare Restriktionen	255
6.1	Verwendung von Zusatzinformation	255
6.2	Die restriktive KQ-Schätzung	256
6.3	Schrittweise Einbeziehung von exakten linearen Restriktionen	259

6.4	Verzernte lineare Restriktionen und MSE-Vergleich mit der KQS	265
6.5	MSE-Matrix-Vergleiche zwischen zwei verzernten Schätzern ..	268
6.6	MSE-Matrix-Vergleich zwischen zwei linearen verzernten Schätzern	275
6.7	MSE-Vergleich zweier (verzerter) restriktiver Schätzer	277
6.8	Stochastische lineare Restriktionen	285
6.8.1	Mixed Schätzer	285
6.8.2	Annahmen zur Kovarianzmatrix V	287
6.8.3	Verzernte stochastische Restriktionen	290
6.9	Abgeschwächte lineare Restriktionen	294
6.9.1	Schwache r -Erwartungstreue	294
6.9.2	Optimale schwach r -erwartungstreue Schätzer	295
6.9.3	Optimale Ersetzung von β	299
6.9.4	RLSE als Ersatz für den mixed Schätzer	301
7.	Das verallgemeinerte lineare Regressionsmodell	303
7.1	Einleitung	303
7.2	Optimale lineare Schätzungen von β	303
7.3	Aitken-Schätzung	311
7.4	Fehlspezifikation der Kovarianzmatrix	313
7.5	Heteroskedastie und Autoregression	315
8.	Vorhersage von y im verallgemeinerten Regressionsmodell	323
8.1	Das Vorhersagemodell	323
8.2	Optimale inhomogene Vorhersage	325
8.3	Optimale homogene Vorhersagen	327
8.4	MSE-Matrix-Vergleiche zwischen optimalen und klassischen Vorhersagen	330
8.4.1	Vergleich klassische – optimale Vorhersage nach der y_* -Superiorität	333

8.4.2	Vergleich klassische – optimale Vorhersage nach der $X_*\beta$ -Superiorität	336
8.5	Vorhersagebereiche	338
9.	Sensitivitätsanalyse	343
9.1	Die Prediction-Matrix	343
9.2	Einfluss einer Beobachtung auf die Parameterschätzung	350
9.2.1	Transformation der Residuen	350
9.2.2	Algebraische Konsequenzen aus dem Wegfall einer Beobachtung	351
9.2.3	Test auf Ausreißer	353
9.3	Grafische Methoden zum Prüfen von Modellannahmen	358
9.4	Maße auf der Basis des Konfidenzellipsoids	361
10.	Modelle für kategoriale Responsevariablen	369
10.1	Generalisierte lineare Modelle	369
10.1.1	Erweiterung des Regressionsmodells	369
10.1.2	Die Struktur des Generalisierten Linearen Modells	371
10.1.3	Scorefunktion und Informationsmatrix	374
10.1.4	Maximum-Likelihood Schätzung	375
10.1.5	Testen von Hypothesen und Güte der Anpassung	379
10.1.6	Overdispersion	380
10.1.7	Quasi-Loglikelihood	382
10.2	Kontingenztafeln	384
10.2.1	Überblick	384
10.2.2	Vergleich von Anteilen	386
10.2.3	Stichproben in zweidimensionalen Kontingenztafeln ...	389
10.2.4	Likelihoodfunktion und Maximum-Likelihood Schätzungen	391
10.2.5	Testen auf Güte der Anpassung	392
10.3	GLM für Binären Response	395
10.3.1	Logitmodelle und Logistische Regression	395

10.3.2 Testen des Modells	399
10.3.3 Verteilungsfunktion als Linkfunktion	399
10.4 Logitmodelle für kategoriale Daten	400
10.5 Güte der Anpassung—Likelihood—Quotienten Test	401
10.6 Loglineare Modelle für Kategoriale Variablen	403
10.6.1 Zweidimensionale Kontingenztafeln	403
10.6.2 Dreidimensionale Kontingenztafeln	406
10.7 Der Spezialfall binärer Responsevariablen	410
10.8 Kodierung kategorialer Kovariablen	413
10.8.1 Dummy- und Effektkodierung	413
10.8.2 Kodierung von Responsemodellen	416
10.8.3 Kodierung in Modellen mit Hazardrate	417
10.9 Erweiterungen für abhängige binäre Variablen	420
10.9.1 Überblick	420
10.9.2 Modellansätze für korrelierten Response	422
10.9.3 Quasi-Likelihood-Ansatz bei korreliertem binären Response	423
10.9.4 Die GEE-Methode von Liang und Zeger	424
10.9.5 Eigenschaften der GEE Schätzung $\hat{\beta}_G$	426
10.9.6 Effizienz von GEE- und IEE-Verfahren	428
10.9.7 Die Wahl der Quasi-Korrelationsmatrix $R_i(\alpha)$	428
10.9.8 Bivariate korrelierte binäre Responsevariablen	429
10.9.9 Die GEE-Methode	430
10.9.10 Die IEE-Methode	432
10.9.11 Ein Beispiel aus der Zahnmedizin	432
10.9.12 Voller Likelihood-Ansatz für marginale Modelle	437
11. Regression bei unvollständigen Daten	439
11.1 Statistische Methoden bei fehlenden Daten	440
11.1.1 Nutzung der kompletten Fälle (complete case analysis)	440

11.1.2	Verwendung aller verfügbaren Daten (available case analysis)	441
11.1.3	Imputation für fehlende Daten	442
11.1.4	Verfahren auf der Basis von Modellen	443
11.2	Missing-Data-Mechanismen	443
11.2.1	Indikatormatrix der fehlenden Werte	445
11.2.2	Missing Completely at Random	445
11.2.3	Missing at Random	446
11.2.4	Nichtignorierbarer Nonresponse	446
11.3	Fehlend-Muster	446
11.4	Fehlende Daten im Response	447
11.4.1	KQ-Schätzung bei vollständigem Datensatz	447
11.4.2	KQ-Schätzung nach Auffüllen fehlender Werte	448
11.4.3	Bartlett's Kovarianzanalyse	449
11.5	Fehlende Werte in der X -Matrix	450
11.5.1	Fehlende Werte und Effizienzverlust	452
11.6	Standardverfahren bei unvollständiger X -Matrix	454
11.6.1	Complete case Analyse (CC)	454
11.6.2	Available case Analyse	455
11.6.3	Maximum-Likelihood Methoden	456
11.7	Imputationsmethoden für unvollständige X -Matrizen	457
11.7.1	Zero-order Regression (ZOR)	457
11.7.2	First-order Regression (FOR)	458
11.7.3	Korrelationsmethoden für stochastisches X	460
11.7.4	Maximum-Likelihood-Schätzungen der fehlenden Werte	460
11.7.5	Gewichtete mixed Schätzung	462
11.8	Annahmen über den Fehlend-Mechanismus	467
11.9	Regressionsdiagnostik zur Identifizierung von Nicht-MCAR Prozessen	467
11.9.1	Vergleich der Mittelwerte	467

11.9.2 Vergleich der Varianz-Kovarianz-Matrizen	468
11.9.3 Diagnostische Maße aus der Sensitivitätsanalyse	468
11.9.4 Verteilung der Maße und Testprozedur	469
11.10 Behandlung von nichtignorierbarem Nichtresponse	469
11.10.1 Gemeinsame Verteilung von (Y, X) mit fehlenden Werten nur in Y	470
11.10.2 Bedingte Verteilung von Y gegeben X mit fehlenden Werten nur in Y	472
11.10.3 Bedingte Verteilung von Y gegeben X mit fehlenden Werten nur in X	473
11.11 Weitere Literatur	474
A. Matrixalgebra	475
A.1 Einführung	475
A.2 Spur einer Matrix	478
A.3 Determinanten	479
A.4 Inverse	481
A.5 Orthogonale Matrizen	482
A.6 Rang einer Matrix	483
A.7 Spalten- und Nullraum	484
A.8 Eigenwerte und Eigenvektoren	484
A.9 Zerlegung von Matrizen (Produktdarstellungen)	488
A.10 Definite Matrizen und quadratische Formen	494
A.11 Idempotente Matrizen	502
A.12 Verallgemeinerte Inverse	503
A.13 Projektoren	512
A.14 Funktionen normalverteilter Variablen	513
A.15 Differentiation von skalaren Funktionen von Matrizen	516
A.16 Stochastische Konvergenz	519

B. Tabellenanhang	521
B.1 Verteilungsfunktion $\Phi(z)$ der Standardnormalverteilung $N(0, 1)$	522
B.2 Dichtefunktion $\phi(z)$ der $N(0, 1)$ -Verteilung	524
B.3 $(1 - \alpha)$ -Quantile $c_{df;1-\alpha}$ der χ^2 -Verteilung	525
B.4 $(1 - \alpha)$ -Quantile $t_{df;1-\alpha}$ der t -Verteilung	526
B.5 $(1 - \alpha)$ -Quantile $f_{df_1, df_2; 1-\alpha}$ der F -Verteilung für $\alpha = 0.05$..	527
B.6 $(1 - \alpha/2)$ -Quantile $f_{df_1, df_2; 1-\alpha/2}$ der F -Verteilung für $\alpha = 0.05$	530
B.7 $(1 - \alpha)$ -Quantile $f_{df_1, df_2; 1-\alpha}$ der F -Verteilung für $\alpha = 0.01$..	533
B.8 $(1 - \alpha/2)$ -Quantile $f_{df_1, df_2; 1-\alpha/2}$ der F -Verteilung für $\alpha = 0.01$	536
Literaturverzeichnis	539
Sachverzeichnis	555