

# Inhalt

**Vorwort zur zweiten Auflage — VII**

**Vorwort zur ersten Auflage — VIII**

**Abbildungsverzeichnis — XIX**

**Tabellenverzeichnis — XXI**

**Verzeichnis der R-Codes — XXIII**

**Verzeichnis der R-Grafiken — XXXI**

## **1 Einführung — 1**

- 1.1 Über dieses Buch — 1
- 1.2 Durchführung eines empirischen Projekts — 1
  - 1.2.1 Vorbereitung und Datenerhebung — 1
  - 1.2.2 Explorative Datenanalyse — 1
  - 1.2.3 Modellierung und Ergebnisdarstellung — 1
- 1.3 Grundlagen statistischer Modellierung — 2
- 1.4 Inferenz — 4
- 1.5 Reversion to the Mean — 5
- 1.6 Hinweise zu R — 5
- 1.7 Literatur — 7

## **2 Lineare Regression — 8**

- 2.1 Grundlagen der linearen Regression — 8
- 2.2 Verwendete R-Pakete — 11
- 2.3 Einführungsbeispiel — 11
- 2.4 Interpretation und Inferenz der Koeffizienten — 16
  - 2.4.1 Interpretation der Koeffizienten – metrisch — 16
  - 2.4.2 Interpretation der Koeffizienten – kategorial — 20
  - 2.4.3 Inferenz der Koeffizienten — 22
- 2.5 Erweiterung der Modellgleichung — 24
  - 2.5.1 Interaktion — 24
  - 2.5.2 Verschachtelte Modelle — 30
  - 2.5.3 Formelsyntax — 32
- 2.6 Globale Modellgüte und Inferenz — 33
  - 2.6.1 Globale Modellgüte — 33
  - 2.6.2 Inferenz für das Modell als Ganzes — 37

2.7	Anwendungsvoraussetzungen und Regressionsdiagnostik —	38
2.7.1	Variablentransformation —	39
2.7.2	Robustheit und einflussreiche Beobachtungen —	44
2.7.3	Nichtlinearer Zusammenhang und Fehlspezifikation —	51
2.7.4	Erwartungswert der Residuen —	61
2.7.5	Endogenität der unabhängigen Variablen —	63
2.7.6	Heteroskedastizität —	66
2.7.7	Autokorrelation —	74
2.7.8	Multikollinearität —	83
2.7.9	Normalverteilung der Residuen —	87
2.7.10	Verschiedene diagnostische Grafiken —	91
2.7.11	Stationarität der Variablen —	92
2.8	Modellselektion und Wichtigkeit der Variablen —	93
2.8.1	Modellselektion —	93
2.8.2	Erklärung —	94
2.8.3	Modellselektion mit der <code>step()</code> -Funktion —	94
2.8.4	Wichtigkeit der Variablen —	97
2.8.5	Vorhersage —	98
2.9	Literatur —	100
2.9.1	Weiterführende Literatur —	100
2.9.2	Anwendungsbeispiele —	100
<b>3</b>	<b>Panelregression —</b>	<b>102</b>
3.1	Grundlagen der Panelregression —	102
3.2	Verwendete R-Pakete —	103
3.3	Vorbereitung der Panelregression —	104
3.4	Ergänzungen für die explorative Datenanalyse —	106
3.5	Gepooltes Modell —	107
3.5.1	Durchführung der gepoolten Regression —	108
3.5.2	Interpretation des gepoolten Modells —	109
3.6	Fixed-Effects-Modell —	110
3.6.1	Anwendung des Fixed-Effects-Modells —	112
3.6.2	Überprüfung auf fixe Effekte —	114
3.6.3	Vor- und Nachteile des Fixed-Effects-Modells —	114
3.7	Random-Effects-Modell —	115
3.7.1	Anwendung des Random-Effects-Modells —	116
3.7.2	Überprüfung auf zufällige Effekte —	118
3.7.3	Überprüfung der Voraussetzungen für das Random-Effects-Modell —	119
3.7.4	Vor- und Nachteile des Random-Effects-Modells —	120
3.8	Vergleich und Auswahl der Modelle —	121
3.9	Regressionsdiagnostik —	123
3.9.1	Heteroskedastizität —	123

- 3.9.2 Autokorrelation — 125
- 3.9.3 Querschnittskorrelation — 127
- 3.9.4 Multikollinearität — 129
- 3.9.5 Ausreißer und einflussreiche Beobachtungen — 130
- 3.10 Modellselektion — 132
- 3.11 Literatur — 134
- 3.11.1 Weiterführende Literatur — 134
- 3.11.2 Anwendungsbeispiele — 134

## **4 Logistische Regression — 136**

- 4.1 Grundlagen der logistischen Regression — 136
  - 4.1.1 Transformation — 137
  - 4.1.2 Modellgleichung — 138
  - 4.1.3 Linkfunktion — 139
  - 4.1.4 Maximum-Likelihood-Schätzung — 140
- 4.2 Verwendete R-Pakete — 141
- 4.3 Interpretation der Koeffizienten — 142
  - 4.3.1 Direkte Interpretation — 142
  - 4.3.2 Interpretation über das Chancenverhältnis — 143
  - 4.3.3 Interpretation über den marginalen Effekt — 149
- 4.4 Globale Modellgüte — 153
  - 4.4.1 Devianz — 153
  - 4.4.2 Likelihood-Quotienten-Test — 155
  - 4.4.3 Pseudo- $R^2$  — 155
- 4.5 Inferenz der Koeffizienten — 156
  - 4.5.1 Wald-Test — 156
  - 4.5.2 Likelihood-Quotienten-Test — 157
  - 4.5.3 Inferenz der marginalen Effekte — 159
- 4.6 Interaktion — 161
- 4.7 Regressionsdiagnostik — 165
  - 4.7.1 Stichprobenumfang — 166
  - 4.7.2 Einflussreiche Beobachtungen — 166
  - 4.7.3 Nichtlinearer Zusammenhang — 172
  - 4.7.4 Multikollinearität — 174
  - 4.7.5 Robuste Varianz-Kovarianz-Schätzer — 175
- 4.8 Modellselektion und Wichtigkeit der Variablen — 176
  - 4.8.1 Modellselektion — 176
  - 4.8.2 Wichtigkeit der Variablen — 178
- 4.9 Klassifikationseigenschaften — 178
  - 4.9.1 Konfusionsmatrix — 179
  - 4.9.2 ROC-Kurve und AUC-Wert — 182
  - 4.9.3 Optimaler Cutpoint — 185

4.9.4	Korrelation —	187
4.9.5	Liftwerte und Liftkurve —	188
4.10	Literatur —	191
4.10.1	Weiterführende Literatur —	191
4.10.2	Anwendungsbeispiele —	191
<b>5</b>	<b>Logistische Panelregression —</b>	<b>192</b>
5.1	Verwendete R-Pakete —	192
5.2	Anwendungsbeispiel —	193
5.3	Fixed-Effects-Modell —	195
5.3.1	Conditional-Maximum-Likelihood-Schätzung —	196
5.3.2	Nachteile des Conditional-Likelihood-Ansatzes —	197
5.3.3	Pseudo-Demeaning —	198
5.3.4	Durchführung einer Fixed-Effects-Modellierung —	198
5.3.5	Überprüfung auf fixe Effekte —	201
5.3.6	Nachteile des Fixed-Effects-Modell —	202
5.4	Random-Effects-Modell —	203
5.4.1	Durchführung einer Random-Effects-Modellierung —	204
5.4.2	Überprüfung auf zufällige Effekte —	207
5.4.3	Überprüfung der Voraussetzungen für ein Random-Effects-Modell —	208
5.5	Interpretation der Koeffizienten —	210
5.5.1	Direkte Interpretation und Chancenverhältnis —	210
5.5.2	Marginale Effekte —	210
5.6	Regressionsdiagnostik —	214
5.6.1	Multikollinearität —	214
5.6.2	Einflussreiche Beobachtungen —	215
5.7	Modellselektion —	217
5.8	Hinweise zur Auswahl der Modellklasse —	220
5.9	Vorhersage —	221
5.10	Klassifikationseigenschaften —	222
5.10.1	Konfusionsmatrix —	223
5.10.2	ROC-Kurve und AUC-Wert —	224
5.11	Literatur —	226
5.11.1	Weiterführende Literatur —	226
5.11.2	Anwendungsbeispiele —	227
<b>6</b>	<b>DAGs und kausale Modellierung —</b>	<b>228</b>
6.1	Einführung in die kausale Modellierung —	228
6.2	Verwendete R-Pakete —	229
6.3	Elemente der DAGs —	230
6.3.1	Pfeil —	230

6.3.2	Chain —	230
6.3.3	Fork —	232
6.3.4	Inverted Fork —	234
6.4	Kausale Modellierung —	236
6.4.1	Chain —	236
6.4.2	Fork —	237
6.4.3	Inverted Fork —	237
6.4.4	Alle Variablen im Datensatz —	238
6.4.5	Allgemeine Vorgehensweise —	242
6.5	Bias —	245
6.5.1	Unbeobachtete Variablen —	245
6.5.2	Collider- und Sample-Selection-Bias —	246
6.6	Zusammenfassung —	249
6.7	Literatur —	249
6.7.1	Weiterführende Literatur —	249
6.7.2	Anwendungsbeispiele —	250
<b>7</b>	<b>Instrumentvariablen —</b>	<b>251</b>
7.1	Grundlagen Instrumentvariablen —	251
7.2	Auswahl der Instrumentvariablen —	253
7.2.1	Anforderungen an Instrumentvariablen —	253
7.2.2	Beispiele —	253
7.2.3	Ausblick —	255
7.3	Zweistufige lineare Regression —	255
7.4	Allgemeines Modell —	257
7.5	Verwendete R-Pakete —	257
7.6	Anwendungsbeispiel —	258
7.6.1	Auswahl der Instrumentvariablen —	258
7.6.2	Lineare Regression als Vergleichsmodell —	260
7.6.3	Manuelle Durchführung der zweistufigen linearen Regression —	261
7.6.4	Instrumentvariablenregression —	261
7.6.5	Bericht —	264
7.7	Überprüfung der Validität der Instrumente —	264
7.7.1	Endogenität der Variablen von Interesse —	264
7.7.2	Relevanz der Instrumente —	265
7.7.3	Exogenität der Instrumente —	266
7.7.4	Durchführung der Tests —	266
7.8	Übertragung auf die Panelregression —	268
7.8.1	Instrumentvariablenregression —	270
7.8.2	Alternative mit $p_{lm}$ —	271
7.9	Überprüfung der Validität der Instrumente – Update —	272
7.9.1	Endogenität der Variablen von Interesse —	273

7.9.2	Relevanz der Instrumente —	273
7.9.3	Exogenität der Instrumente —	274
7.10	Binäre abhängige Variable —	276
7.11	Literatur —	276
7.11.1	Weiterführende Literatur —	276
7.11.2	Anwendungsbeispiele —	277
<b>8</b>	<b>Ereignisstudie —</b>	<b>278</b>
8.1	Einführung —	278
8.2	Renditen —	279
8.2.1	Normale Rendite —	279
8.2.2	Ereignis- und Schätzfenster —	283
8.2.3	Abnormale Renditen —	284
8.3	Testverfahren —	284
8.3.1	t-Test —	285
8.3.2	Cross-Sectional-Dependence-Test (CSD-Test) —	285
8.3.3	Cross-Sectional-Independence-Test (CSI-Test) —	286
8.3.4	Corrado-Rang-Test —	286
8.4	Verwendete R-Pakete —	287
8.5	Anwendungsbeispiel —	288
8.5.1	Ereignisstudie mit mehreren Unternehmen —	288
8.5.2	Ereignisstudie mit einem Unternehmen —	294
8.6	Ausblick —	296
8.7	Literatur —	296
8.7.1	Weiterführende Literatur —	296
8.7.2	Anwendungsbeispiele —	297
<b>9</b>	<b>Klassifikation und Regression mit Bäumen und Random Forest —</b>	<b>298</b>
9.1	Grundlagen baumbasierter Verfahren —	298
9.2	Verwendete R-Pakete —	299
9.3	Klassifikations- und Regressionsbäume —	300
9.3.1	Klassifikationsbäume —	300
9.3.2	Wichtige Parameter der Funktion <code>rpart()</code> —	307
9.3.3	Beschneiden der Bäume —	308
9.3.4	Regressionsbäume —	311
9.3.5	Vor- und Nachteile baumbasierter Verfahren —	318
9.4	Random Forest —	319
9.4.1	Bootstrapping —	319
9.4.2	Bootstrapping bei Random Forest —	320
9.4.3	Random Forest für die Klassifikation —	320
9.4.4	Random Forest für die Regression —	325
9.4.5	Vorhersage —	327

9.4.6	Wichtige Parameter von <code>randomForest()</code> und <code>cforest()</code> — 328
9.4.7	Wichtigkeit der Variablen — 328
9.4.8	Weitere Möglichkeiten zum Einsatz von Random Forest — 332
9.4.9	Vor- und Nachteile von Random Forest — 332
9.5	Literatur — 333
9.5.1	Weiterführende Literatur — 333
9.5.2	Anwendungsbeispiele — 333
<b>10</b>	<b>Hauptkomponentenanalyse — 334</b>
10.1	Grundlagen der Hauptkomponentenanalyse — 334
10.2	Verwendete R-Pakete — 338
10.3	Anwendungsbeispiel — 339
10.3.1	Eignung der Daten — 340
10.3.2	Anzahl der Hauptkomponenten — 342
10.3.3	Durchführung und Interpretation der Hauptkomponentenanalyse — 344
10.3.4	Zusammenfassung der Hauptkomponenten — 348
10.4	Literatur — 350
10.4.1	Weiterführende Literatur — 350
10.4.2	Anwendungsbeispiele — 351
<b>11</b>	<b>Analyse von Zeitreihen — 352</b>
11.1	Verwendete R-Pakete — 352
11.2	Grundbegriffe — 353
11.2.1	Stationarität — 354
11.2.2	Autokovarianz- und Autokorrelationsfunktion — 354
11.2.3	Stochastischer Prozess — 357
11.2.4	Differenzbildung — 358
11.2.5	Integration und Kointegration — 359
11.2.6	Stochastischer und deterministischer Trend — 362
11.2.7	Tests auf Stationarität — 363
11.3	AR- und MA-Prozesse — 367
11.3.1	AR-Prozess — 367
11.3.2	MA-Prozess — 369
11.4	AR(I)MA-Modelle — 371
11.4.1	Residuenanalyse — 371
11.4.2	Beispiel für ein ARIMA-Modell — 372
11.4.3	Verschiedene diagnostische Plots — 377
11.4.4	Automatische Schätzung der Modellparameter — 378
11.4.5	Vorhersage — 379
11.5	Rendite und Volatilität — 381
11.5.1	Rendite — 381
11.5.2	Volatilität — 381

11.5.3	Gewichtung der früheren Varianzen —	<b>382</b>
11.5.4	EWMA-Modelle —	<b>383</b>
11.6	(G)ARCH-Modelle —	<b>383</b>
11.6.1	ARCH-Modelle —	<b>384</b>
11.6.2	GARCH-Modelle —	<b>387</b>
11.6.3	Beispiel eines GARCH-Modells —	<b>389</b>
11.6.4	Ausblick —	<b>398</b>
11.7	Literatur —	<b>399</b>
11.7.1	Weiterführende Literatur —	<b>399</b>
11.7.2	Anwendungsbeispiele —	<b>399</b>

<b>Anhang —</b>	<b>401</b>
-----------------	------------

<b>Quellennachweise —</b>	<b>425</b>
---------------------------	------------

<b>Literatur —</b>	<b>427</b>
--------------------	------------

<b>Stichwortverzeichnis —</b>	<b>441</b>
-------------------------------	------------

<b>Verzeichnis der verwendeten R-Funktionen —</b>	<b>449</b>
---	------------