

Einführung in die Statistik

Analyse und Modellierung von Daten

Von
Univ.-Prof. Dr. Rainer Schlittgen

9., durchgesehene Auflage

R. Oldenbourg Verlag München Wien

Inhaltsverzeichnis

1	Statistische Daten	1
1.1	Erhebungen	1
	Problemstellung	1
	Begriffsbildungen	1
1.2	Statistische Variablen	4
	Begriffsbildungen	4
	Klassifikationen	5
	Transformationen	7
1.3	Die Datenmatrix	9
1.4	Aufgaben	11
2	Darstellung univariater Datensätze	13
2.1	Häufigkeitsverteilungen diskreter Variablen	13
	Problemstellung	13
	Häufigkeiten	14
	Graphische Darstellungen	15
2.2	Häufigkeitsverteilungen stetiger Variablen	17
	Problemstellung	17
	Geordneter Datensatz	18
	Stemleaf-Diagramm	19
	Häufigkeitstabelle	22
	Das Histogramm	23
2.3	Die empirische Verteilungsfunktion	26
	Einführendes Beispiel	26
	Empirische Verteilungsfunktion	26
	Klassierte Daten	28
2.4	Quantile	31
	Problemstellung	31
	Bestimmung der Quantile	31
	Spezielle Quantile	34
	Quantildiagramme	34
	5-Zahlen-Zusammenfassung	37
	Box-Plots	38
2.5	Aufgaben	40
3	Darstellung multivariater Datensätze	44
3.1	Diskrete Variablen	44
	Einführendes Beispiel	44
	Bivariate Kontingenztabellen	45
	Bedingte relative Häufigkeiten	46
	Mehrdimensionale Kontingenztabellen	47

3.2	Stetige Variablen.	48
	Streudiagramm.	48
	Konvexe Hüllen.	49
	Histogramm.	50
	Mehrdimensionale Daten.	52
3.3	Aufgaben.	54
4	Grundlagen der Wahrscheinlichkeitsrechnung	57
4.1	Zufallsexperimente und Ereignisse.	57
	Zufallsexperimente	57
	Ereignisse.	59
4.2	Wahrscheinlichkeiten.	64
	Chancen von Ereignissen.	64
	Häufigkeiten von Ereignissen.	65
	Statistische Wahrscheinlichkeiten.	66
	Axiome von Kolmogorov.	68
	Das Gleichmöglichkeitsmodell.	71
	Zufallsauswahlen aus endlichen Grundgesamtheiten.	72
4.3	Bedingte Wahrscheinlichkeit und Unabhängigkeit	77
	Bedingte Wahrscheinlichkeit.	77
	Unabhängigkeit	80
4.4	Aufgaben.	82
5	Zufallsvariablen	87
5.1	Zufallsvariablen.	87
	Definition.	88
	Verteilungsfunktion.	89
5.2	Diskrete Zufallsvariablen.	91
	Wahrscheinlichkeitsfunktion.	91
	Bestimmung eines Verteilungsmodells.	92
	Verteilungsfunktion.	95
5.3	Stetige Zufallsvariablen.	96
	Dichtefunktion.	96
	Verteilungsfunktion.	99
	Bestimmung eines Verteilungsmodells.	100
5.4	Theoretische Quantile.	102
	Definition.	102
	Quantildiagramme.	103
5.5	Mehrdimensionale Zufallsvariablen.	108
	Grundlagen.	108
	Unabhängigkeit	111
5.6	Aufgaben.	112
6	Lageparameter	116
6.1	Empirische Lageparameter.	116
	Problemstellung.	116
	Erfassung des Niveaus.	116
	Der Median.	117

	Das arithmetische Mittel	118
	Berechnung des arithmetischen Mittels	119
	Eigenschaften des arithmetischen Mittels	121
	Weitere Lageparameter	124
6.2	Theoretische Lageparameter	127
	Der Erwartungswert	127
	Eigenschaften des Erwartungswertes	129
	Der Median	131
	Der Modus	132
6.3	Aufgaben	132
7	Streuungsparameter	136
7.1	Empirische Streuungsparameter	136
	Problemstellung	136
	Durchschnittliche Abweichung und Standardabweichung	136
	Varianz	139
	Weitere Streuungsparameter	142
7.2	Die theoretische Varianz	144
	Definition der Varianz	144
	Eigenschaften	146
	Die Tschebyschev-Ungleichung	148
7.3	Aufgaben	151
8	Schiefeparameter und Datentransformation	154
8.1	Schiefeparameter	154
	Einführendes Beispiel	154
	Schiefe theoretischer Verteilungen	154
	Theoretische Parameter	157
	Schiefe empirische Verteilungen	158
8.2	Datentransformation	162
	Problemstellung	162
	Nichtlineare Transformation	162
	Auswahl einer Transformation zur Symmetrisierung	164
	Stabilisierung der Streuung	168
8.3	Aufgaben	171
9	Parameter multivariater Verteilungen	172
9.1	Empirische Verteilungen	172
	Problemstellung	172
	Die Kovarianz	173
	Der Korrelationskoeffizient von Bravais-Pearson	176
	Der Rangkorrelationskoeffizient von Spearman	179
	Der Phi-Koeffizient	181
9.2	Theoretische Verteilungen	183
9.3	Aufgaben	185

10	Diskrete Verteilungsmodelle	188
10.1	Die Binomialverteilung	189
	Einführendes Beispiel	189
	Herleitung der Verteilung	190
	Eigenschaften	193
	Anpassung an empirische Verteilungen	194
	Die hypergeometrische Verteilung	194
	Die Multinomialverteilung	196
10.2	Die Poisson-Verteilung	197
	Problemstellung	197
	Ableitung und Eigenschaften	198
	Anpassung an empirische Verteilungen	200
10.3	Die geometrische Verteilung	201
	Einführendes Beispiel	201
	Ableitung und Eigenschaften	202
10.4	Die negative Binomialverteilung	204
	Ableitung und Eigenschaften	204
	Anpassung an empirische Verteilungen	205
10.5	Zur Auswahl eines diskreten Verteilungsmodells	208
	Einführendes Beispiel	208
	Das Auswahlendiagramm	208
10.6	Aufgaben	212
11	Stetige Verteilungsmodelle	215
11.1	Stetige Gleichverteilung	215
11.2	Die Pareto-Verteilung	216
	Definition	216
	Anpassung an empirische Verteilungen	217
11.3	Die Exponentialverteilung	219
	Ableitung und Eigenschaften	219
	Anpassung an empirische Verteilungen	222
11.4	Die Gamma-Verteilung	225
	Ableitung und Eigenschaften	225
11.5	Aufgaben	227
12	Die Normalverteilung	229
12.1	Grundlagen	229
	Problemstellung	229
	Definition und Bedeutung der Parameter	230
	Eigenschaften	231
	Anpassung an empirische Verteilungen	235
12.2	Approximation von Verteilungen	238
	Problemstellung	238
	Der zentrale Grenzwertsatz	239
	Approximation der Binomialverteilung	240
	Approximation der Poisson-Verteilung	243

12.3	Die logarithmische Normalverteilung	244
	Problemstellung	244
	Definition und Eigenschaften	245
	Anpassung an empirische Verteilungen	245
12.4	Die bivariate Normalverteilung	248
	Problemstellung	248
	Ableitung und Eigenschaften	249
	Anpassung an empirische Verteilungen	252
12.5	Aufgaben	254
13	Stichprobenfunktionen	258
13.1	Grundlagen	258
	Einführendes Beispiel	258
	Definition	259
13.2	Spezielle Stichprobenfunktionen	261
	Das arithmetische Mittel	261
	Die relative Häufigkeit	263
	Die empirische Verteilungsfunktion	265
	Monte-Carlo-Simulation	266
13.3	Aufgaben	267
14	Schätzen von Parametern	269
14.1	Schätzfunktionen und ihre Eigenschaften	269
	Problemstellung	269
	Schätzfunktionen	271
	Mittlerer quadratischer Fehler	271
	Erwartungstreue	275
	Konsistenz	277
14.2	Anwendungsorientierte Aspekte	278
	Problemstellung	278
	Robustheit	279
	Sensitivitätskurve und Bruchpunkt	280
	Datenvalidierung mit der Sensitivitätskurve	282
	Standardfehler	285
14.3	Schätzmethoden	288
	Problemstellung	288
	Momentenmethode	288
	Maximum-Likelihood-Methode	289
	Bestimmung des Maximums von $L(\hat{\theta})$	293
	Mehrere Parameter	293
14.4	Übersicht	295
14.5	Aufgaben	296
15	Konfidenzintervalle	299
15.1	Konfidenzintervall für den Median	299
	Problemstellung	299
	Ableitung des Konfidenzintervalles	300
	Das Konfidenzintervall	301

15.2	Allgemeine Definition eines Konfidenzintervalles.	304
15.3	Konfidenzintervalle für den Erwartungswert	305
	Normalverteilung mit bekanntem σ^2	305
	Normalverteilung mit unbekanntem σ^2	306
	Approximative Konfidenzintervalle für p	309
15.4	Konfidenzintervalle für Streuungsparameter.	311
	Konfidenzintervall für σ^2 bei Normalverteilung	311
	Konfidenzintervalle für zwei Quartile.	313
15.5	Konfidenzintervalle für Wahrscheinlichkeiten und Anteile	315
	Problemstellung.	315
	Approximatives Konfidenzintervall für eine Wahrscheinlichkeit	316
	Approximatives Konfidenzintervall für einen Anteil.	317
	Länge der Konfidenzintervalle für p	318
15.6	Aufgaben.	320
16	Testen von Hypothesen.	323
16.1	Reine Signifikanztests.	323
	Problemstellung.	323
	Tests auf der Basis von Konfidenzintervallen.	323
	Tests auf der Basis von Prüfgrößen.	325
	Das empirische Signifikanzniveau.	327
16.2	Die Elemente der klassischen Testtheorie.	329
	Problemstellung.	329
	Formen und Hypothesen.	329
	Fehlerarten.	331
	Gütefunktion.	332
16.3	Aufgaben.	335
17	Spezielle Parametertests.	337
17.1	Einstichprobentests.	337
	Tests auf f_i bei Normalverteilung	337
	Approximative Tests auf J	338
	Test auf f_i bei symmetrischen Verteilungen.	339
	Test auf f_i bei beliebigen Verteilungen.	342
	Tests auf σ^2 bei Normalverteilung	344
	Test auf eine Wahrscheinlichkeit p	345
17.2	Parametervergleich bei unabhängigen Stichproben.	345
	Testen von Lageunterschieden.	345
	Vergleich von i_x und β_y bei Normal Verteilung: σ_x und σ_y bekannt	346
	Vergleich von i_x und n_y : Stichproben groß.	346
	Vergleich von f_x und n_y bei Normal Verteilung: σ_x und σ_y unbekannt, Stichproben klein.	348
	Vergleich von f_x und f_y	349
	Vergleich zweier Streuungen.	351
	Vergleich zweier Wahrscheinlichkeiten.	352
17.3	Übersichten über die Parametertests.	354
	Übersicht Einstichprobentests.	354

Übersicht Zweistichprobentests	355
17.4 Aufgaben	356
18 Varianzanalyse	360
18.1 Einfache Varianzanalyse	360
Einführendes Beispiel	360
Das Modell	361
Der Test bei Normalverteilung	362
Multiple Vergleiche bei Normalverteilung	366
Ungleiche Steuerungen	368
Nicht-normalverteilte Variablen	370
18.2 Zweifache Varianzanalyse	373
Einführendes Beispiel	373
Modell mit Wechselwirkungen	373
Test bei Normalverteilung	374
Eine Beobachtung pro Zelle	378
Ungleiche Varianzen	379
Nicht-normalverteilte Störungen	382
18.3 Aufgaben	384
19 Überprüfung der Anpassungsgüte von Verteilungen und der Unabhängigkeit	386
19.1 Anpassungstests	386
Problemstellung	386
Der Chi-Quadrat-Anpassungstest für vollständig spezifizierte Verteilungen	387
Der Chi-Quadrat-Anpassungstest für Verteilungen mit unbekannten Parametern	389
Das Chigramm	393
Der Kolmogorov-Smirnov-Test	394
Anpassungstest vom Korrelationstyp	401
19.2 Unabhängigkeitstests	403
Problemstellung	403
Der Chi-Quadrat-Test	404
Tests auf Korrelation	408
19.3 Aufgaben	410
20 Regressionsanalyse	414
20.1 Die empirische Regressionsgerade	415
Problemstellung	415
Kriterien für die Anpassung einer Geraden	416
Bestimmung der Regressionsgeraden	418
Das Bestimmtheitsmaß	420
20.2 Das lineare Regressionsmodell	422
Einführendes Beispiel	422
Entwicklung des Modells	423

20.3	Schätzen und Testen im linearen Regressionsmodell	425
	Problemstellung	425
	Die Kleinst-Quadrate-Schätzfunktion.	426
	Eigenschaften des KQ-Schätzers.	427
	Konfidenzintervalle.	429
	Tests.	432
20.4	Residuenanalyse	434
	Systematische Änderung des Mittels.	436
	Inhomogenität der Varianz.	438
	Ausreißer.	440
	Verletzung der Normalverteilungsannahme.	443
20.5	Linearisierung eines Zusammenhanges.	446
	Problemstellung	446
	Transformationen.	447
	Auswahl der Transformationen.	447
20.6	Aufgaben.	449
	Anhang Tabellen.	455
	A: Binomialverteilung	456
	B: Poissonverteilung	462
	C: Normalverteilung	464
	D: t-Verteilung	466
	E: Chi-Quadrat-Verteilung	468
	F: Kritische Werte des F-Tests für $\alpha = 0.05$	470
	G: Kritische Werte der Korrelationstests.	472
	H: Quantile der Wilcoxon-Vorzeichen-Rangteststatistik	474
	I: Quantile der Wilcoxon-Rangsummenstatistik	476
	Anhang: Lösungen der Aufgaben.	480
	Index	515