
Jürgen Hedderich · Lothar Sachs

Angewandte Statistik

Methodensammlung mit R

16., überarbeitete und erweiterte Auflage

Inhaltsverzeichnis

Inhaltsverzeichnis	ix
Verzeichnis der Übersichten	xxii
Abbildungsverzeichnis	xxv
Tabellenverzeichnis	xxx
1 Einführung	1
1.1 Definition und Aufgaben der Statistik	1
1.2 Wissenschaftliche Arbeitstechnik	4
1.2.1 Daten und Modelle	4
1.2.2 Kreisprozesse	5
1.2.3 Modelle in der Statistik	7
1.3 Statistik und wissenschaftliche Methode	9
1.3.1 Wiederholbare Erfahrungen	9
1.3.2 Inferenz: Deduktion und Induktion	10
1.3.3 Beobachtungsstudien	11
1.3.3.1 Hinweise zu einer Befragung	12
1.3.3.2 Hinweise zu einer Erhebung	13
1.3.4 Systematische Fehler (Bias)	14
1.3.5 Fehlende Angaben	15
1.3.6 Deskriptive Statistik	16
1.3.6.1 Dimensionalität	17
1.3.6.2 Data Editing	17
1.3.7 Explorativer Ansatz	18
1.3.8 Konfirmativer Ansatz	19
1.3.9 Merkmale, Grundgesamtheit, Stichprobe	20
1.3.9.1 Offene und geschlossene Grundgesamtheiten	21
1.3.10 Stichproben	21
1.3.11 Zufallsstichproben	22
1.4 Datenanalyse	23
1.4.1 Merkmalsarten	23

1.4.2	Von Beobachtungen zu Daten - Skalierung	24
1.4.2.1	Zulässige Skalentransformationen	26
1.4.3	Datenstruktur, Datengewinnung und Datenerfassung	27
2	Grundlagen aus der Mathematik	30
2.1	Logische und relationale Operatoren	30
2.2	Mengen	31
2.2.1	Begriffsbildung	31
2.2.2	Mengenoperationen	32
2.3	(Grund-) Rechenarten	34
2.3.1	Summen und Produkte	34
2.3.1.1	Das Summenzeichen	34
2.3.1.2	Spezielle Summen	37
2.3.1.3	Multiplikation und Division; Fakultät	38
2.3.2	Potenzen und Wurzeln	39
2.3.3	Logarithmen	41
2.3.4	Rundungen	42
2.3.5	Rechnen mit fehlerbehafteten Zahlen	44
2.4	Einführung in die Matrixalgebra	45
2.4.1	Definition und Schreibweise	45
2.4.2	Matrixoperationen	46
2.4.2.1	Matrixaddition und -subtraktion	46
2.4.2.2	Matrixmultiplikation	47
2.4.3	Determinanten	49
2.4.4	Die Inverse Matrix	50
2.4.5	Lineare Abhängigkeit, Rang einer Matrix	50
2.4.6	Lineare Gleichungssysteme	51
2.4.7	Eigenwerte und Eigenvektoren	52
2.5	Funktionen	53
2.5.1	Lineare Funktionen	53
2.5.2	Nichtlineare Funktionen	54
2.5.2.1	Polynomfunktionen	54
2.5.3	Periodische Funktionen	55
2.5.4	Exponentialfunktion und logarithmische Funktion	56
2.5.4.1	Wachstumsfunktionen	56
2.5.5	Fläche unter einer Funktion: Integral	58
2.6	Kombinatorik	59
2.6.1	Permutationen	59
2.6.2	Binomialkoeffizient	61
2.6.2.1	Rechnen mit dem Binomialkoeffizienten - weitere Formeln	62
2.6.3	Kombinationen	63
2.6.4	Zerlegung einer Menge	64
2.6.5	Bootstrap-Stichproben	65
2.6.6	Das Pascalsche Dreieck	65
2.6.7	Der Multinomialkoeffizient	66
3	Deskriptive Statistik	68
3.1	Häufigkeiten	69
3.1.1	Absolute und relative Häufigkeiten	69
3.1.2	Sinnvolle Quotienten: Verhältniszahlen	71
3.1.3	Prozentwerte	73
3.1.4	Balken- und Kreisdiagramme	74

3.1.5	Tabellen	74
3.1.5.1	Rechteckdiagramm und Mosaikplot	76
3.1.6	Bedingte Häufigkeiten	76
3.1.6.1	Simpson's Paradox (Vergleich von Anteilswerten)	77
3.1.6.2	Assoziationsmaße für Kontingenztafeln	77
3.2	Beschreibung von Ordinaldaten	79
3.2.1	Medianwert und andere Quantile	80
3.2.2	Klassierte Beobachtungen: Berechnung von Quantilen	82
3.2.3	Streuung ordinal skaliertes Daten	83
3.2.4	Punktplot und Box-Plot	83
3.2.5	Korrelationskoeffizient nach Kendall	85
3.2.6	Partielle Rangkorrelation	86
3.3	Beschreibung von metrischen Daten	89
3.3.1	Arithmetischer Mittelwert	89
3.3.2	Standardabweichung, Varianz	92
3.3.3	Ergänzung und Kombination von Mittelwerten und Varianzen	93
3.3.4	Variationskoeffizient	94
3.3.5	Der Streubereich um den Mittelwert	95
3.3.6	Mittelwert und Standardabweichung aus klassierten Messwerten	95
3.3.7	Das gewogene arithmetische Mittel und die gewogene Varianz	96
3.3.8	Geometrischer Mittelwert	98
3.3.9	Harmonischer Mittelwert	102
3.4	Fehlerrechnung	105
3.4.1	Fehler in Messungen; Präzision und Richtigkeit	105
3.4.2	Standardfehler von Mehrfachbestimmungen	106
3.4.3	Fehlerfortpflanzung	107
3.4.4	Präzision von Messungen	108
3.4.5	Reliabilitätsanalyse	111
3.5	Häufigkeitsverteilung	114
3.5.1	Histogramm	114
3.5.2	Stamm-Blatt Darstellung	116
3.5.3	Pareto-Diagramm	117
3.6	Konzentrationsmaße: Herfindahl-Index und Gini-Koeffizient	119
3.7	Maßzahlen für den Zusammenhang metrischer Daten	121
3.7.1	Typisierung korrelativer Zusammenhänge	121
3.7.2	Punktwolken	124
3.7.3	Empirische Kovarianz	125
3.7.4	Empirischer Korrelationskoeffizient	125
3.7.5	Autokorrelation	127
3.7.6	Rangkorrelationskoeffizient nach Spearman	129
3.7.6.1	Spearman'sche Rangkorrelation bei Bindungen	130
3.7.7	Partieller Korrelationskoeffizient	130
3.7.7.1	Multipler Korrelationskoeffizient	132
3.7.8	Lineare Regression	132
3.7.9	Spezielle Schätzungen der Regressionsgeraden	136
3.7.9.1	Das Bartlett-Verfahren	136
3.7.9.2	Das Kerrich-Verfahren	137
3.7.9.3	Orthogonale kleinste Quadrate	138
3.7.10	Robuste lineare Regression	138
3.7.11	Nichtlineare Regression	141
3.7.11.1	Einige linearisierende Transformationen	148

3.8	Nichtparametrische Regression	150
3.8.1	Regressogramm, gleitende Mittelwerte und Kernschätzer	151
3.8.2	Kubische Spline-Interpolation	153
4	Wahrscheinlichkeiten	155
4.1	Zufallsexperiment, Ereignis	156
4.2	Begriff der Wahrscheinlichkeit	159
4.2.1	Definition nach Laplace	159
4.2.1.1	Relative Häufigkeit und Wahrscheinlichkeit	160
4.2.1.2	Chancen (Odds)	161
4.2.2	Axiome nach Kolmogoroff	161
4.2.3	Additionssatz	163
4.2.3.1	Einschluss- und Ausschlussformel (Siebformel)	164
4.2.4	Paradoxon der ersten Ziffer (Benford's Law)	167
4.3	Bedingte Wahrscheinlichkeiten, stochastische Unabhängigkeit	168
4.3.1	Bedingte Wahrscheinlichkeit	168
4.3.1.1	Multiplikationssatz	168
4.3.1.2	Risikomaße [wird durch Abschnitt 7.7.3 ergänzt]	170
4.3.1.3	Wahrscheinlichkeiten aus einer Sterbetafel	171
4.3.1.4	Baumdiagramm und Pfadregeln	172
4.3.2	Stochastische Unabhängigkeit	173
4.3.2.1	Stochastische Unabhängigkeit für drei und mehr Ereignisse	175
4.3.2.2	Unvereinbarkeit und stochastische Unabhängigkeit	176
4.3.2.3	Ungleichungen nach Bonferroni	176
4.3.2.4	Bedingte Wahrscheinlichkeiten und Korrelation	177
4.3.2.5	Dreizehn Beispiele zur stochastischen Unabhängigkeit	177
4.4	Bayessches Theorem	181
4.4.1	Bayessches Theorem und Pfadregel	183
4.4.2	Acht Beispiele zum Bayesschen Theorem	183
4.5	Der diagnostische Test	186
4.5.1	ROC - Analyse	191
4.5.2	Der Likelihoodquotient	192
4.5.3	Entscheidungsanalyse nach A.J. Vickers	195
4.6	Maßzahlen in der Epidemiologie	196
4.6.1	Prävalenz und Inzidenz	197
4.6.2	Der Impfeffekt	198
4.6.3	Standardisierte Raten am Beispiel der Mortalität	199
5	Zufallsvariablen, Verteilungen	202
5.1	Die Zufallsvariable	202
5.1.1	Wahrscheinlichkeitsfunktion, Wahrscheinlichkeitsdichte und Verteilungsfunktion	203
5.1.1.1	Rechenregeln zur Verteilungsfunktion	207
5.1.1.2	Empirische Verteilungsfunktion	208
5.2	Maßzahlen zur Kennzeichnung einer Verteilung	208
5.2.1	Erwartungswert	209
5.2.2	Varianz	210
5.2.2.1	Ungleichungen mit Wahrscheinlichkeiten	212
5.2.3	Momente: Schiefe und Exzess	215
5.2.3.1	Berechnung der empirischen Momente	216
5.2.3.2	Potenzmomente	217
5.2.3.3	Quantilsmaße zu Schiefe und Exzess	221

5.3	Diskrete Verteilungen	222
5.3.1	Das Urnenmodell	222
5.3.2	Gleichverteilung	224
5.3.3	Binomialverteilung	225
5.3.3.1	Bernoulli-Versuch	225
5.3.3.2	Binomial-Verteilung	226
5.3.3.3	Approximation der Binomialverteilung durch die Standardnormalverteilung	233
5.3.3.4	Approximation der Binomialverteilung durch die Poisson-Verteilung	235
5.3.4	Multinomialverteilung (Polynomialverteilung)	236
5.3.5	Poisson-Verteilung	238
5.3.5.1	Der Dispersionsindex	244
5.3.5.2	Approximation der Poissonverteilung durch die Standardnormalverteilung	246
5.3.6	Negative Binomialverteilung	246
5.3.7	Geometrische Verteilung	251
5.3.8	Hypergeometrische Verteilung	252
5.3.8.1	Approximationen der Hypergeometrischen Verteilung	256
5.3.9	Negative Hypergeometrische Verteilung	256
5.4	Stetige Verteilungen	258
5.4.1	Gleichverteilung	258
5.4.2	Standard-Beta-Verteilung	259
5.4.3	Normalverteilung	264
5.4.3.1	Zentrale Schwankungsintervalle der Standardnormalverteilung	269
5.4.3.2	Hinweise und Beispiele zur Normalverteilung	271
5.4.3.3	Zentraler Grenzwertsatz	276
5.4.4	Halbnormalverteilung	277
5.4.5	Gestutzte Normalverteilung	280
5.4.6	Lognormalverteilung	282
5.4.6.1	Schätzung der Maßzahlen einer Lognormalverteilung	285
5.4.6.2	Empirische Maßzahlen einer Lognormalverteilung	286
5.4.7	Exponentialverteilung	288
5.4.8	Weibull-Verteilung	290
5.4.9	Extremwertverteilung Typ I (Gumbel-Verteilung)	292
5.4.10	Gamma-Verteilung	295
5.5	Testverteilungen	298
5.5.1	Student-Verteilung (t-Verteilung)	298
5.5.1.1	Nichtzentrale t-Verteilung	302
5.5.2	Chiquadrat-Verteilung	303
5.5.2.1	Nichtzentrale Chiquadrat-Verteilung	306
5.5.3	Fisher-Verteilung (F-Verteilung)	307
5.5.4	Interpolation von Tabellenwerten und P-Werten	309
5.5.4.1	Interpolieren von Zwischenwerten der F -Verteilung	313
5.6	Verteilung zweidimensionaler Zufallsvariablen	316
5.6.1	Modellbildung	316
5.6.1.1	Einführendes Beispiel	316
5.6.1.2	Verteilungsfunktion	316
5.6.2	Randverteilungen und Unabhängigkeit	318
5.6.2.1	Bedingte Verteilung und Unabhängigkeit	320
5.6.2.2	Satz von Bayes für Zufallsvariablen	321

5.6.3	Korrelationskoeffizient	322
5.6.4	Zweidimensionale Normalverteilung	323
6	Schätzen	325
6.1	Stichproben-Erhebung	325
6.2	Zufallsstichproben und Zufallszahlen	326
6.2.1	Spezielle Stichprobenverfahren	331
6.3	Das Schätzen von Parametern	334
6.3.1	Vorbemerkungen	335
6.3.1.1	Der Schluss von der Stichprobe auf die Grundgesamtheit	335
6.3.1.2	Punktschätzung und Intervallschätzung	335
6.3.1.3	Schätzwert und Schätzfunktion	336
6.3.2	Wünschenswerte Eigenschaften von Schätzfunktionen	336
6.3.2.1	Unverzerrte Schätzfunktion für die Varianz	337
6.3.3	Gesetz der großen Zahlen	338
6.3.4	Der mittlere quadratische Fehler	338
6.4	Schätzverfahren für Maßzahlen einer Verteilung	340
6.4.1	Momentenmethode	340
6.4.2	Schätzung nach der größten Erwartung (MLE)	341
6.4.2.1	ML-Schätzer zur Binomialverteilung	343
6.4.2.2	ML-Schätzer zur Negativen Binomialverteilung	343
6.4.2.3	ML-Schätzer zur Poisson-Verteilung	345
6.4.2.4	ML-Schätzer zur Normalverteilung	346
6.4.2.5	ML-Schätzer zur gestutzten Normalverteilung	347
6.4.3	Schätzung nach dem kleinsten Fehler (OLS)	348
6.5	Intervallschätzung - Konfidenzintervalle	349
6.6	Konfidenzintervall für Anteilswerte	352
6.6.1	Approximation durch die Normalverteilung	356
6.6.2	Nullergebnisse bzw. Vollergebnisse	358
6.6.3	Schnellschätzung der Vertrauensgrenzen anhand einer beobachteten relativen Häufigkeit nach Clopper und Pearson	359
6.6.4	Konfidenzintervall für die Differenz zweier Anteile	360
6.6.5	Konfidenzintervall für das Verhältnis zweier Anteile	362
6.6.6	Mindestumfang einer Stichprobe zur Schätzung eines Anteils	364
6.6.7	Simultane Konfidenzintervalle für multinomiale Anteile	367
6.7	Konfidenzintervalle für den Erwartungswert einer Poisson-Verteilung	368
6.7.1	Zentrale („exakte“) Konfidenzintervalle für den Erwartungswert	369
6.7.2	Nicht-zentrale (kürzeste) Konfidenzintervalle nach Crow und Gardner	370
6.7.3	Konfidenzintervall für das Verhältnis zweier Raten (Poisson-Verteilung)	372
6.7.4	Konfidenzintervalle für standardisierte Raten	373
6.8	Konfidenzintervalle für den Erwartungswert bei Normalverteilung	376
6.8.1	Vertrauenswahrscheinlichkeit und Irrtumswahrscheinlichkeit	376
6.8.2	Konfidenzintervall für den Erwartungswert einer Normalverteilung	378
6.8.3	Konfidenzintervall für die Differenz zweier Erwartungswerte	381
6.8.4	Konfidenzintervall für den Erwartungswert aus Paardifferenzen	382
6.8.5	Konfidenzintervall für das Verhältnis zweier Erwartungswerte	383
6.8.6	Konfidenzintervall für Verhältniszahlen	384
6.8.7	Konfidenzintervall für den Erwartungswert einer Lognormalverteilung	385
6.9	Konfidenzintervall für die mittlere absolute Abweichung	386
6.10	Konfidenzintervall für den Median	387
6.10.1	Konfidenzintervall für die Differenz und den Quotienten von Medianen	391
6.10.2	Verteilungsunabhängige Konfidenzintervalle für beliebige Quantile	393

6.11	Konfidenzintervalle nach dem Bootstrap-Verfahren	394
6.12	Konfidenzintervall für die Varianz bzw. die Standardabweichung	397
6.12.1	Konfidenzintervall für den Variationskoeffizienten	400
6.12.2	Konfidenzintervall für den Quotienten zweier Varianzen	400
6.13	Weibull-Verteilung	401
6.13.1	Bestimmung der Parameter	401
6.13.2	Konfidenzintervall für die Weibull-Gerade	403
6.14	Konfidenzintervalle für die Parameter einer linearen Regression	404
6.14.1	Die Schätzung einiger Standardabweichungen	404
6.14.2	Konfidenzintervalle für den Regressionskoeffizienten, für den Achsenabschnitt und für die Restvarianz	408
6.14.3	Konfidenzintervalle und Prädiktionsintervalle für die Regressionsgerade ..	410
6.14.4	Inverse Prädiktion aus einer linearen Regression	413
6.15	Konfidenzintervall für den Korrelationskoeffizienten nach Pearson	414
6.16	Übereinstimmung und Präzision von Messwerten	416
6.16.1	Übereinstimmung von Messwertreihen nach Bland-Altman	417
6.16.2	Regressionsverfahren zur Übereinstimmung zweier Messwertreihen	419
6.16.3	Vergleich der Präzision und der Genauigkeit zweier Messwertreihen	422
6.16.4	Der Konkordanz-Korrelationskoeffizient	423
6.16.5	Intraklassen-Korrelation: Interrater-Reliabilität	425
6.17	Toleranzgrenzen	428
6.17.1	Verteilungsunabhängige Toleranzgrenzen	430
6.18	Voraussageintervalle (Prädiktionsintervalle)	431
6.18.1	Voraussageintervalle für den Mittelwert aus künftigen Beobachtungen	433
6.18.2	Voraussageintervalle für alle künftigen Beobachtungen	433
6.18.3	Voraussageintervalle für die Standardabweichung aus künftigen Beobachtungen	435
6.19	Bayes-Schätzung	435
6.19.1	A-priori Verteilungen (Prior)	439
6.19.2	Parameterschätzung nach Bayes	441
7	Hypothesentest	445
7.1	Der statistische Test	445
7.1.1	Entscheidungsprinzipien	446
7.1.2	Statistische Hypothesen und Testentscheidungen	447
7.1.3	Die Formulierung von Hypothesen	451
7.1.3.1	Alternativhypothese als Gegensatz zur Nullhypothese	451
7.1.3.2	Durch Daten angeregte Hypothesen	451
7.1.4	Der <i>P</i> -Wert nach R.A. Fisher	452
7.1.5	Wie oft wird eine wahre Nullhypothese abgelehnt?	454
7.1.6	Statistischer Test - Schritt für Schritt	455
7.1.6.1	Der Likelihood-Quotient; das Neyman-Pearson Lemma	458
7.1.7	Powerfunktion und Operationscharakteristik	459
7.1.8	Die Operationscharakteristik	464
7.1.8.1	Die OC-Kurve in der Qualitätskontrolle	464
7.1.9	Äquivalenztests	467
7.1.10	Verteilungsunabhängige Verfahren	467
7.2	Tests der Verteilung (Anpassungstests, goodness-of-fit)	470
7.2.1	Der Quotient R/s	471
7.2.2	Überprüfung des 3. und 4. Momentes	473
7.2.3	Der Quantile-Quantile Plot	475
7.2.4	Box-Cox-Transformation	477

7.2.5	Der Chiquadrat-Anpassungstest	479
7.2.6	Kolmogoroff-Smirnoff Anpassungstest	483
7.2.6.1	Anpassung an eine Poisson-Verteilung	486
7.2.7	Shapiro-Wilk Test	487
7.2.8	Anderson-Darling Test	488
7.2.9	Ausreißerproblem	489
7.2.9.1	Grubbs-Test für Ausreißer	491
7.2.9.2	Q-Test nach Dixon für kleine Stichproben	492
7.2.9.3	Standardisierte Extremabweichungen	493
7.3	Einstichprobenverfahren	494
7.3.1	Hypothesen zu Wahrscheinlichkeiten	494
7.3.1.1	Binomialtest	494
7.3.1.2	Binomialtest - Approximation durch die Normalverteilung	496
7.3.1.3	Binomialtest - Fallzahlabstschätzung	498
7.3.1.4	Binomialtest: Likelihood-Quotienten-Test	500
7.3.2	Hypothesen zu Erwartungswerten, die sich auf einen empirischen Mittelwert beziehen	501
7.3.2.1	Einstichproben- <i>t</i> -Test	501
7.3.2.2	Fallzahlabstschätzung zum Einstichproben- <i>t</i> -Test	504
7.3.2.3	Einstichprobentest auf Äquivalenz	505
7.3.3	Einstichproben-Median-Test	507
7.3.4	Vergleich einer empirischen Varianz mit ihrem Parameter	509
7.3.5	Asymptotischer Test für den Variationskoeffizienten	509
7.3.6	Prüfung der Zufällsmäßigkeit einer Folge von Alternativdaten oder von Messwerten	510
7.3.6.1	Die sukzessive Differenzenstreuung	510
7.3.6.2	Der Iterationstest für die Prüfung, ob eine Folge von Alternativdaten oder von Messwerten zufallsgemäß ist	512
7.3.6.3	Phasenhäufigkeitstest von Wallis und Moore	514
7.3.6.4	Der Vorzeichen-Trendtest von Cox und Stuart	515
7.3.6.5	Variabilität der zentralen Tendenz	516
7.3.7	Prüfung der Erwartungswerte von Poisson-Verteilungen	517
7.4	Zweistichprobenverfahren	520
7.4.1	Vergleich zweier Varianzen (F-Test)	520
7.4.1.1	Varianzvergleich bei kleinem bis mittlerem Stichprobenumfang ..	520
7.4.1.2	Varianzvergleich bei mittlerem bis großem Stichprobenumfang ..	522
7.4.1.3	Varianzvergleich bei großem bis sehr großem Stichprobenumfang	523
7.4.1.4	Stichprobenumfang und Power für den F-Test	523
7.4.1.5	Vergleich der Streuung zweier kleiner Stichproben nach Pillai und Buenaventura	524
7.4.1.6	Vergleich zweier Variationskoeffizienten	524
7.4.2	Rangdispersionstest von Siegel und Tukey	526
7.4.3	Ansari-Bradley Test	530
7.4.4	<i>t</i> -Test für unabhängige Stichproben	533
7.4.4.1	Unbekannte aber gleiche Varianzen	533
7.4.4.2	<i>t</i> -Test bei unbekanntem Varianzen, die möglicherweise ungleich sind	538
7.4.4.3	Fallzahlabstschätzung für den <i>t</i> -Test für zwei unabhängige Stichproben	541
7.4.4.4	Bootstrap: <i>t</i> -Test Variante	545
7.4.4.5	Multivariater <i>t</i> -Test: Hotelling's T^2	546

7.4.5	t-Test für Paardifferenzen	548
7.4.5.1	Gepaarte Beobachtungen	548
7.4.5.2	Absolute oder prozentuale Veränderungen	549
7.4.5.3	t-Test für paarweise angeordnete Messwerte	550
7.4.5.4	Prüfung der Gleichheit zweier Varianzen paariger Stichproben	552
7.4.6	Wilcoxon-Rangsummentest für zwei unabhängige Stichproben (U-Test)	552
7.4.6.1	Der U-Test bei Rangaufteilung	561
7.4.6.2	Effektstärke im Vergleich unabhängiger Stichproben	563
7.4.6.3	Fallzahlabeschätzung für den U-Test	566
7.4.7	Wilcoxon-Paardifferenzentest	568
7.4.7.1	Konfidenzintervall für den Median aus Paardifferenzen	572
7.4.7.2	Der Maximum-Test für Paardifferenzen	572
7.4.7.3	Der Vorzeichentest von Dixon und Mood	573
7.4.7.4	Stichprobenumfang für den Vorzeichentest / Wilcoxon-Test für Paardifferenzen	574
7.4.8	Vergleich zweier Erwartungswerte aus Poisson-Verteilungen	576
7.4.9	Vergleich zweier unabhängiger Stichproben nach Kolmogoroff/Smirnow	579
7.4.9.1	Cramér-von Mises Test	581
7.4.10	Einige weitere verteilungsunabhängige Verfahren für den Vergleich unabhängiger Stichproben	583
7.4.10.1	Der Zwei-Stichproben Dispersionstest: Count Five	583
7.4.10.2	Rosenbaumsche Schnelltests	584
7.4.10.3	Permutationstest, Randomisierungstest	585
7.4.10.4	Der Vergleich zweier unabhängiger Stichproben: Schnelltest nach Tukey	589
7.4.10.5	Der Median-Test	590
7.4.11	Zweistichprobentest auf Äquivalenz	592
7.4.11.1	Test auf Bioäquivalenz	593
7.5	Mehrfacher Hypothesentest	595
7.5.1	Multipl. Testproblem	595
7.5.2	Adjustierung von P-Werten	598
7.5.3	Kombination von P-Werten aus gleichgerichteten einseitigen Tests	603
7.6	Mehrstichprobenverfahren, varianzanalytische Methoden	605
7.6.1	Prüfung der Gleichheit mehrerer Varianzen aus normalverteilten Grundgesamtheiten	605
7.6.1.1	Prüfung der Gleichheit mehrerer Varianzen nach Hartley	605
7.6.1.2	Prüfung der Gleichheit mehrerer Varianzen nach Cochran	607
7.6.1.3	Prüfung der Gleichheit mehrerer Varianzen nach Bartlett	608
7.6.1.4	Robuster Test auf Varianzhomogenität nach Levene in der Brown-Forsythe-Version	610
7.6.2	Transformation zur Stabilisierung der Varianz	611
7.6.3	Einfache Varianzanalyse (ANOVA, analysis of variance)	613
7.6.3.1	Permutationstest zur Varianzanalyse	617
7.6.3.2	Stichprobenumfänge und Power für den varianzanalytischen Mittelwertvergleich	618
7.6.4	Multiple paarweise Vergleiche und weitere Mittelwertvergleiche	620
7.6.4.1	Multiple Vergleiche nach Tukey-Kramer	620
7.6.4.2	Multiple Vergleiche nach Games-Howell	623
7.6.4.3	Multiple Vergleiche mit einer Kontrolle nach Dunnett	625
7.6.4.4	Multiple Vergleiche: Auswahl des „Besten“ nach Hsu	627
7.6.4.5	Bereichsschätzung für μ_{max} und Auswahl der „besten“ \bar{x}_i	630

7.6.4.6	Multipler Vergleich von Mittelwerten mit dem Gesamtmittel: Maximum-Modulus-Ansatz	630
7.6.4.7	Beurteilung linearer Kontraste nach Scheffé	632
7.6.4.8	Bildung homogener Gruppen von Mittelwerten anhand des nach Hayter modifizierten LSD-Tests, eines Lücken-Tests für geordnete μ_i	636
7.6.5	H-Test von Kruskal und Wallis	637
7.6.5.1	Multiple paarweise Vergleiche mittlerer Ränge	640
7.6.5.2	H-Test mit Stichproben-Untergruppen	644
7.6.5.3	H-Test-Variante: Vergleich eines Standards mit mehreren Behandlungen	645
7.6.5.4	Trendtest nach Jonckheere: Vergleich mehrerer geordneter Verteilungsfunktionen, zugleich ein Trendtest für Mediane	646
7.6.6	Varianzanalyse für Messwiederholungen (Blockvarianzanalyse)	649
7.6.7	Friedman-Test	651
7.6.7.1	Multiple paarweise Vergleiche mit einer Kontrolle	654
7.6.7.2	Multiple paarweise Vergleiche nach Wilcoxon und Wilcox	656
7.6.7.3	Page-Test für geordnete Alternativen	659
7.6.7.4	Spannweitenrangtest nach Quade	661
7.6.8	Zweifache Varianzanalyse	663
7.6.9	Analyse von wiederholten Messungen	668
7.6.9.1	Typisierung wiederholter Messungen	668
7.6.9.2	ANOVA für wiederholte Messungen (gemischte Modelle)	672
7.6.10	Prinzipien der Versuchsplanung	673
7.7	Die Analyse von Häufigkeiten	680
7.7.1	Vergleich zweier relativer Häufigkeiten	680
7.7.2	Analyse von Vierfeldertafeln	684
7.7.2.1	Fallzahl und Power zum Vierfeldertest	688
7.7.2.2	Minimales n für den Vierfeldertest	690
7.7.2.3	Vorsicht vor Trugschlüssen beim Vierfeldertest	691
7.7.3	Spezielle Risiko- und Effektmaße	692
7.7.3.1	Odds Ratio und relatives Risiko	692
7.7.3.2	Konfidenzintervalle für das Relative Risiko ψ und für die Odds Ratio ω	695
7.7.3.3	Stichprobenumfänge für die Schätzung von Odds Ratio ω und Relativem Risiko ψ	697
7.7.3.4	Der expositionsbedingte Anteil Erkrankter: Population Attributable Risk	702
7.7.3.5	Number Needed to Treat (NNT)	704
7.7.4	Exakter Test nach R.A. Fisher	706
7.7.5	Äquivalenz zweier Binomialwahrscheinlichkeiten	709
7.7.6	Der von McNemar modifizierte Vorzeichenstest	712
7.7.7	Test nach Mantel-Haenszel	717
7.7.7.1	Breslow-Day-Test	719
7.7.7.2	Die Kombination von Vierfeldertafeln	722
7.7.8	Der $k \times 2$ -Felder-Chiquadrat-Test nach Brandt und Snedecor	724
7.7.8.1	Multipler Vergleich von Anteilen (Marascuilo-Prozedur)	728
7.7.8.2	Homogenitätstest nach Ryan (Lückentest)	730
7.7.8.3	Power- und Fallzahlabstimmung zum k -2-Felder-Test	731
7.7.9	Cochran-Armitage Test auf linearen Trend	732
7.7.10	Vergleich mehrerer Anteile mit einem vorgegebenen Anteilswert (Standard)	736

7.7.11	Die Analyse von Kontingenztafeln	738
7.7.11.1	Kontingenzkoeffizient - Stärke des Zusammenhangs	746
7.7.11.2	Fallzahl und Power für die Analyse von Kontingenztafeln	747
7.7.11.3	Lokalisation der stochastischen Abhängigkeit nach Hommel	748
7.7.11.4	Simultane Paarvergleiche nach Royen	750
7.7.12	Bowker-Test auf Symmetrie in quadratischen Mehrfeldertafeln	753
7.7.13	Marginalhomogenitätstest nach Lehman	754
7.7.14	Stuart-Maxwell-Test auf Homogenität der Randverteilungen	755
7.7.15	Q-Test nach Cochran	757
7.7.15.1	Simultane Konfidenzintervalle für paarweise Differenzen von Erfolgsanteilen	758
7.7.16	Cohen's Kappa-Koeffizient κ	760
7.7.16.1	Das gewichtete Kappa	763
7.7.16.2	Das Kappa für mehrfache Beurteilungen (Multi-Rater)	764
7.7.17	Kendalls Konkordanzkoeffizient W	766
7.8	Hypothesentests zur Korrelation und Regression	768
7.8.1	Hypothesentest zum Korrelationskoeffizienten (Pearson)	769
7.8.1.1	z -Transformation nach R.A. Fisher	773
7.8.1.2	Korrelation bei Mehrfachbeobachtungen	776
7.8.1.3	Fallzahl und Power zum Korrelationskoeffizienten	779
7.8.1.4	Der Vergleich mehrerer Korrelationskoeffizienten	780
7.8.2	Prüfung des Rang-Korrelationskoeffizienten nach Spearman (Rho-S)	781
7.8.3	Prüfung des Rang-Korrelationskoeffizienten nach Kendall (Tau)	783
7.8.4	Hypothesentests zu den Parametern einer Regression	786
7.8.4.1	Prüfung der Linearität einer Regression	786
7.8.4.2	Chow-Test: „Strukturbruch“ in einer linearen Regression	787
7.8.4.3	Durbin-Watson-Test: Autokorrelation in den Residuen	789
7.8.4.4	Prüfung des Regressionskoeffizienten gegen Null	792
7.8.4.5	Prüfung der Differenz zwischen einem geschätzten und einem hypothetischen Regressionskoeffizienten.	792
7.8.4.6	Prüfung der Differenz zwischen einem geschätzten und einem hypothetischen Achsenabschnitt.	793
7.8.4.7	Vergleich zweier Regressionskoeffizienten	793
7.8.4.8	Vergleich zweier Achsenabschnitte	795
8	Statistische Modellbildung	796
8.1	Einführung	796
8.2	Lineare Regressionsmodelle	800
8.2.1	Die einfache lineare Regression	800
8.2.2	Multiple lineare Regression	804
8.2.3	Überwindung von Multikollinearität in Regressionsmodellen.	809
8.2.4	Analyse der Residuen im linearen Modell	810
8.2.5	Heteroskedastizität im linearen Modell	813
8.2.6	Hypothesentest und Konfidenzintervalle zum linearen Modell	815
8.2.7	Verfahren der Variablenauswahl	816
8.2.8	Nominalskalierte Einflussgrößen	820
8.3	Varianzanalyse im linearen Modell	821
8.3.1	Einfaktorielle Varianzanalyse	821
8.3.1.1	Erwartungswert-Parametrisierung	821
8.3.1.2	Effekt-Parametrisierung: Dummy-Codierung	822
8.3.1.3	Effekt-Parametrisierung: Effekt-Codierung	823
8.3.1.4	Varianzkomponenten - ANOVA	824

	8.3.1.5	Klassenbildung von stetigen Einflussgrößen	825
	8.3.2	Zweifaktorielle Varianzanalyse	826
8.4		Logistische Regression	830
	8.4.1	Hypothesentest im logistischen Regressionsmodell	833
	8.4.2	Multiple logistische Regression	835
	8.4.3	Interpretation der Regressionskoeffizienten	839
	8.4.4	Variablenauswahl im Rahmen der Modellbildung	840
	8.4.5	Residuenanalyse	841
	8.4.6	Pseudo-Bestimmtheitsmaße (Pseudo- R^2)	843
8.5		Poisson-Regression und loglineare Modelle	844
	8.5.1	Poisson-Regression	844
	8.5.1.1	Dispersionsindex und Poisson-Regression	847
	8.5.2	Poisson-Regression zum relativen Risiko aus Raten	847
	8.5.3	Analyse von Kontingenztafeln	849
	8.5.4	Loglineares Modell am Beispiel von 2 Faktoren	852
	8.5.5	Dreidimensionale Kontingenztafeln	854
	8.5.5.1	Modellbildung unter verschiedenen Restriktionen	854
	8.5.5.2	Modellauswahl im loglinearen Ansatz	856
	8.5.5.3	Fünf Einschränkungen und Hinweise zum loglinearen Modell	859
8.6		Modelle zu wiederholten Messungen	860
	8.6.1	Varianzanalyse für wiederholte Messungen	860
	8.6.2	Lineare gemischte Modelle	864
	8.6.3	Analyse von Cluster-Daten	871
	8.6.4	Verallgemeinerte Schätzgleichungen	876
8.7		Analyse von Überlebenszeiten	881
	8.7.1	Kaplan-Meier Schätzung der Überlebensfunktion	884
	8.7.2	Der Logrank-Test	889
	8.7.3	Parametrische Regressionsmodelle für Überlebenszeiten	892
	8.7.3.1	Exponentielles Regressionsmodell	892
	8.7.3.2	Weibull Regressionsmodell	894
	8.7.3.3	Loglogistisches Regressionsmodell	896
	8.7.3.4	Modellwahl und Güte der Anpassung	897
	8.7.3.5	AFT-Modelle (Accelerated Failure Time)	899
	8.7.4	Das Proportional-Hazards Modell von Cox	901
	8.7.4.1	Parameter-Schätzung zum Cox-Modell	902
	8.7.4.2	Interpretation der Parameter	904
	8.7.4.3	Modellbildung; zur Auswahl geeigneter Einflussgrößen	907
	8.7.4.4	Güte der Modellanpassung, Residuenanalyse	908
9		Einführung in R	912
	9.1	Das Konsolfenster	912
	9.2	Hilfestellung in R	916
	9.3	Objekte in R	917
	9.3.1	Vektoren	917
	9.3.2	Erzeugen von Vektoren und Dateneingabe	919
	9.3.3	Faktoren in R, Klassenbildung	920
	9.3.4	Erzeugen von Matrizen und Tabellen	920
	9.3.5	Rechnen mit Matrizen in R am Beispiel der Herleitung einer Kovarianzmatrix	921
	9.3.6	Tabellenstruktur: Daten in Rahmen (data.frame)	922
	9.4	Fehlende Angaben	925
	9.5	Auswahl und Sortierung von Daten	926

9.6	Ablaufsteuerung: logische Bedingungen und Funktionen in R	928
9.7	Einige mathematische und statistische Funktionen	932
9.8	Modellbildung in R	935
9.9	Einfache graphische Funktionen und Hilfsmittel	936
Literaturverzeichnis		941
Autorenverzeichnis		971
Sachverzeichnis		979
Verzeichnis zu den Beispielen		1015
Verzeichnis zu R-Funktionen		1019