

Jürgen Hedderich • Lothar Sachs

Angewandte Statistik

Methodensammlung mit R

Vierzehnte, überarbeitete und ergänzte Auflage

Inhaltsverzeichnis

Verzeichnis der Übersichten	xxi
--	-----

Abbildungsverzeichnis	xxiii
------------------------------------	-------

Tabellenverzeichnis	xxvii
----------------------------------	-------

1 Einführung	1
1.1 Definition und Aufgaben der Statistik	1
1.2 Wissenschaftliche Arbeitstechnik	3
1.2.1 Daten und Modelle	3
1.2.2 Kreisprozesse	4
1.2.3 Modelle in der Statistik	6
1.3 Statistik und wissenschaftliche Methode	8
1.3.1 Wiederholbare Erfahrungen	8
1.3.2 Deskriptive Statistik	10
1.3.2.1 Dimensionalität	11
1.3.2.2 Data Editing	11
1.3.3 Explorativer Ansatz	12
1.3.4 Konfirmativer Ansatz	12
1.3.5 Merkmale, Grundgesamtheit, Stichprobe	14
1.3.5.1 Offene und geschlossene Grundgesamtheiten	14
1.3.6 Stichproben	15
1.3.7 Zufallsstichproben	15
1.4 Datenanalyse	16
1.4.1 Merkmalsarten	16
1.4.2 Variablen-Übersicht	17
1.4.2.1 Zulässige Skalentransformationen	19
1.4.3 Datenstruktur, Datengewinnung und Datenerfassung	19
2 Grundlagen aus der Mathematik	23
2.1 Logische und relationale Operatoren	23
2.2 Mengen	24
2.2.1 Begriffsbildung	24
2.2.2 Mengenoperationen	25

2.3	(Grund-) Rechenarten	27
2.3.1	Summen und Produkte	27
2.3.1.1	Das Summenzeichen	27
2.3.1.2	Spezielle Summen	30
2.3.1.3	Multiplikation und Division; Fakultät	31
2.3.2	Potenzen und Wurzeln	32
2.3.3	Logarithmen	33
2.3.4	Rundungen	35
2.3.5	Rechnen mit fehlerbehafteten Zahlen	36
2.4	Einführung in die Matrixalgebra	37
2.4.1	Definition und Schreibweise	38
2.4.2	Matrixoperationen	39
2.4.2.1	Matrixaddition und -subtraktion	39
2.4.2.2	Matrixmultiplikation	39
2.4.3	Determinanten	42
2.4.4	Die Inverse Matrix	42
2.4.5	Lineare Abhängigkeit, Rang einer Matrix	43
2.4.6	Lineare Gleichungssysteme	44
2.4.7	Eigenwerte und Eigenvektoren	44
2.5	Funktionen	46
2.5.1	Lineare Funktionen	46
2.5.2	Nichtlineare Funktionen	47
2.5.2.1	Polynomfunktionen	47
2.5.3	Periodische Funktionen	48
2.5.4	Exponentialfunktion und logarithmische Funktion	49
2.5.4.1	Wachstumsfunktionen	49
2.5.5	Fläche unter einer Funktion: Integral	51
2.6	Kombinatorik	52
2.6.1	Permutationen	52
2.6.2	Kombinationen - der Binomialkoeffizient	53
2.6.2.1	Rechnen mit dem Binomialkoeffizienten - weitere Formeln	54
2.6.3	Kombinationen mit Wiederholungen und mit Berücksichtigung der Anordnung	55
2.6.4	Zerlegung einer Menge	56
2.6.5	Das Pascalsche Dreieck	57
2.6.6	Der Multinomialkoeffizient	58
3	Deskriptive Statistik	60
3.1	Häufigkeiten	61
3.1.1	Absolute und relative Häufigkeiten	61
3.1.2	Sinnvolle Quotienten: Verhältniszahlen	63
3.1.3	Prozentwerte	64
3.1.4	Torten- und Balkendiagramme	65
3.1.5	Tabellen	66
3.1.5.1	Rechteckdiagramm und Mosaikplot	67
3.1.6	Bedingte Häufigkeiten	67
3.1.6.1	Simpson's Paradox (Vergleich von Anteilswerten)	68
3.2	Beschreibung von Ordinaldaten	69
3.2.1	Medianwert und andere Quartile	70
3.2.2	Quantile	71
3.2.3	Streuung ordinal skalierten Daten	71
3.2.4	Punktdiagramm und Box-Plot	72

3.2.5	Korrelationskoeffizient nach Kendall	73
3.2.6	Partielle Rangkorrelation	75
3.3	Beschreibung von metrischen Daten	78
3.3.1	Arithmetischer Mittelwert	78
3.3.2	Standardabweichung, Varianz	80
3.3.2.1	Arithmetische Mittelwerte und Standardabweichungen vergleichbar gemacht	82
3.3.3	Variationskoeffizient	82
3.3.4	Der $(x \pm s)$ -Bereich	83
3.3.5	Klassierte Messwerte; Berechnung des Mittelwertes und der Standardabweichung	83
3.3.6	Das gewogene arithmetische Mittel, die gewogene Varianz und das gewichtete arithmetische Mittel	84
3.3.7	Geometrischer Mittelwert	85
3.3.8	Harmonischer Mittelwert	89
3.4	Fehlerrechnung	92
3.4.1	Fehler in Messungen	92
3.4.2	Standardfehler von Mehrfachbestimmungen	93
3.4.3	Fehlerfortpflanzung	94
3.4.4	Präzision von Messungen	95
3.4.5	Reliabilitätsanalyse	98
3.5	Häufigkeitsverteilung	100
3.5.1	Histogramm	101
3.5.2	Stamm-Blatt Darstellung	103
3.5.3	Pareto-Diagramm	104
3.6	Konzentrationsmaße: Herfindahl-Index und Gini-Index	105
3.7	Maßzahlen für den Zusammenhang metrischer Daten	107
3.7.1	Punktwolken	108
3.7.2	Empirische Kovarianz	109
3.7.3	Empirischer Korrelationskoeffizient	109
3.7.4	Rangkorrelationskoeffizient	110
3.7.4.1	Spearman'sche Rangkorrelation bei Bindungen	112
3.7.5	Typisierung korrelativer Zusammenhänge	112
3.7.6	Partieller Korrelationskoeffizient	114
3.7.6.1	Multipler Korrelationskoeffizient	115
3.7.7	Lineare Regression	116
3.7.8	Spezielle Schätzungen der Regressionsgeraden	119
3.7.8.1	Das Bartlett-Verfahren	119
3.7.8.2	Das Kerrich-Verfahren	120
3.7.8.3	Orthogonale kleinste Quadrate	121
3.7.9	Robuste lineare Regression	122
3.8	Nichtlineare Regression	124
3.8.1	Einige linearisierende Transformationen	131
4	Wahrscheinlichkeiten	134
4.1	Zufallsexperiment, Ereignis	135
4.2	Begriff der Wahrscheinlichkeit	138
4.2.1	Definition nach Laplace	138
4.2.1.1	Relative Häufigkeit und Wahrscheinlichkeit	139
4.2.1.2	Chancen (Odds)	139
4.2.2	Axiome nach Kolmogoroff	140
4.2.3	Additionssatz	141

4.2.3.1	Einschluss- und Ausschlussformel (Siebformel)	142
4.2.4	Paradoxon der ersten Ziffer (Benford's Law)	145
4.3	Bedingte Wahrscheinlichkeiten, stochastische Unabhängigkeit	146
4.3.1	Bedingte Wahrscheinlichkeit	146
4.3.1.1	Multiplikationssatz	146
4.3.1.2	Risikomaße [wird durch Abschnitt 7.6.3.1 ergänzt]	148
4.3.1.3	Wahrscheinlichkeiten aus einer Sterbetafel	149
4.3.1.4	Baumdiagramm und Pfadregeln	149
4.3.2	Stochastische Unabhängigkeit	150
4.3.2.1	Stochastische Unabhängigkeit für drei und mehr Ereignisse	152
4.3.2.2	Unvereinbarkeit und stochastische Unabhängigkeit	153
4.3.2.3	Ungleichungen nach Bonferroni	154
4.3.2.4	Bedingte Wahrscheinlichkeiten und Korrelation	154
4.3.2.5	Dreizehn Beispiele zur stochastischen Unabhängigkeit	155
4.4	Bayessches Theorem	158
4.4.1	Bayessches Theorem und Pfadregel	160
4.4.2	Acht Beispiele zum Bayesschen Theorem	161
4.5	Der diagnostische Test	163
4.5.1	ROC - Analyse	167
4.5.2	Der Likelihoodquotient	169
4.5.3	Entscheidungsanalyse nach A.J. Vickers	171
4.6	Maßzahlen in der Epidemiologie	172
4.6.1	Prävalenz und Inzidenz	173
4.6.2	Standardisierungen	175
5	Zufallsvariablen, Verteilungen	177
5.1	Die Zufallsvariable	177
5.1.1	Wahrscheinlichkeitsfunktion, Wahrscheinlichkeitsdichte und Verteilungsfunktion	178
5.1.1.1	Rechenregeln zur Verteilungsfunktion	182
5.1.1.2	Empirische Verteilungsfunktion	183
5.2	Maßzahlen zur Kennzeichnung einer Verteilung	183
5.2.1	Erwartungswert	184
5.2.2	Varianz	186
5.2.2.1	Ungleichungen mit Wahrscheinlichkeiten	188
5.2.3	Momente: Schiefe und Exzess	190
5.2.3.1	Berechnung der empirischen Momente	191
5.2.3.2	Potenzmomente	192
5.2.3.3	Quantilmaße zu Schiefe und Exzess	195
5.3	Diskrete Verteilungen	197
5.3.1	Das Urnenmodell	197
5.3.2	Gleichverteilung	198
5.3.3	Binomialverteilung	200
5.3.3.1	Bernoulli-Versuch	200
5.3.3.2	Binomial-Verteilung	201
5.3.3.3	Approximation der Binomialverteilung durch die Standardnormalverteilung	207
5.3.3.4	Approximation der Binomialverteilung durch die Poisson-Verteilung	210
5.3.4	Multinomialverteilung (Polynomialverteilung)	211
5.3.5	Poisson-Verteilung	213
5.3.5.1	Der Dispersionsindex	219

5.3.5.2	Approximation der Poissonverteilung durch die Standardnormalverteilung	221
5.3.6	Negative Binomialverteilung	222
5.3.7	Geometrische Verteilung	226
5.3.8	Hypergeometrische Verteilung	227
5.3.8.1	Approximationen der Hypergeometrischen Verteilung	230
5.4	Stetige Verteilungen	231
5.4.1	Gleichverteilung	231
5.4.2	Standard-Beta-Verteilung	232
5.4.3	Normalverteilung	236
5.4.3.1	Zentrale Schwankungsintervalle der Standardnormalverteilung	242
5.4.3.2	Familie der Normalverteilungen	243
5.4.3.3	Hinweise und Beispiele zur Normalverteilung	244
5.4.3.4	Zentraler Grenzwertsatz	248
5.4.4	Lognormalverteilung	249
5.4.4.1	Schätzung der Maßzahlen einer Lognormalverteilung	251
5.4.4.2	Empirische Maßzahlen einer Lognormalverteilung	253
5.4.5	Exponentialverteilung	255
5.4.6	Weibull-Verteilung	257
5.4.7	Gamma-Verteilung	258
5.5	Testverteilungen	261
5.5.1	Student-Verteilung (t)	261
5.5.1.1	Nichtzentrale t -Verteilung	264
5.5.2	Chiquadrat-Verteilung (χ^2)	265
5.5.3	Fisher-Verteilung (F)	268
5.5.4	Interpolation von Tabellenwerten	273
5.5.4.1	Interpolieren von Zwischenwerten der F -Verteilung	274
5.5.5	Verteilungen wichtiger Stichprobenfunktionen aus normalverteilten Grundgesamtheiten	276
5.6	Verteilung zweidimensionaler Zufallsvariablen	278
5.6.1	Modellbildung	278
5.6.1.1	Einführendes Beispiel	278
5.6.1.2	Verteilungsfunktion	279
5.6.2	Randverteilungen und Unabhängigkeit	280
5.6.2.1	Bedingte Verteilung und Unabhängigkeit	282
5.6.2.2	Satz von Bayes für Zufallsvariablen	283
5.6.3	Korrelationskoeffizient	284
5.6.4	Zweidimensionale Normalverteilung	285
6	Schätzen	287
6.1	Stichproben-Erhebung	287
6.2	Zufallsstichproben und Zufallszahlen	288
6.2.1	Spezielle Stichprobenverfahren	291
6.2.1.1	Schlussziffernauswahl	292
6.2.1.2	Geburtstagsverfahren	293
6.3	Das Schätzen von Parametern	293
6.3.1	Vorbemerkungen	293
6.3.1.1	Der Schluss von der Stichprobe auf die Grundgesamtheit	293
6.3.1.2	Punktschätzung und Intervallschätzung	294
6.3.1.3	Schätzwert und Schätzfunktion	294
6.3.2	Wünschenswerte Eigenschaften von Schätzfunktionen	295
6.3.2.1	S^2 ist eine unverzerrte Schätzfunktion, nicht aber S	296

6.3.3	Gesetz der großen Zahlen	297
6.3.4	Der mittlere quadratische Fehler	297
6.4	Schätzverfahren für Maßzahlen einer Verteilung	298
6.4.1	Momentenmethode	298
6.4.2	Schätzung nach der größten Erwartung (MLE)	299
6.4.2.1	ML-Schätzer zur Binomialverteilung	301
6.4.2.2	ML-Schätzer zur Negativen Binomialverteilung	302
6.4.2.3	ML-Schätzer zur Poisson-Verteilung	304
6.4.2.4	ML-Schätzer zur Normalverteilung	304
6.4.3	Schätzung nach dem kleinsten Fehler (OLS)	305
6.5	Konfidenzintervalle	306
6.6	Konfidenzintervall für π - Anteilswert aus einer dichotomen Grundgesamtheit	309
6.6.1	Approximation durch die Normalverteilung	310
6.6.2	Nullergebnisse bzw. Vollergebnisse ($\hat{p} = 0$ bzw. $\hat{p} = 1$)	315
6.6.3	Schnellschätzung der Vertrauensgrenzen anhand einer beobachteten relativen Häufigkeit nach Clopper und Pearson	317
6.6.4	Konfidenzintervall für die Differenz zweier Anteile $\pi_1 - \pi_2$	317
6.6.5	Konfidenzintervall für das Verhältnis zweier Anteile π_1/π_2	320
6.6.6	Konfidenzintervalle für standardisierte Raten	322
6.6.7	Schätzung des Mindestumfangs einer Stichprobe bei ausgezählten Werten	323
6.7	Konfidenzintervalle für μ bei Normalverteilung	325
6.7.1	Vertrauenswahrscheinlichkeit und Irrtumswahrscheinlichkeit	325
6.7.2	Konfidenzintervall für den Erwartungswert μ	326
6.7.3	Mindestzahl von Beobachtungen zur Schätzung eines Mittelwertes	329
6.7.4	Konfidenzintervall für die Differenz $\mu_1 - \mu_2$	329
6.7.5	Konfidenzintervall für den Erwartungswert μ_d aus Paardifferenzen	331
6.7.6	Konfidenzintervall für das Verhältnis μ_1/μ_2	332
6.7.7	Konfidenzintervall für Verhältniszahlen	332
6.7.8	Konfidenzintervall für den Erwartungswert einer Lognormalverteilung	333
6.8	Konfidenzintervall für die mittlere absolute Abweichung	334
6.9	Konfidenzintervall für den Median	336
6.9.1	Verteilungsunabhängige Konfidenzintervalle für beliebige Quantile	340
6.10	Konfidenzintervalle nach dem Bootstrap-Verfahren	341
6.11	Konfidenzintervall für σ^2 bzw. für σ	343
6.11.1	Konfidenzintervall für den Quotienten zweier Varianzen σ_1^2/σ_2^2	345
6.11.2	Mindestzahl von Beobachtungen zur Schätzung einer Standardabweichung	345
6.12	Konfidenzintervall für den Erwartungswert λ einer Poisson-Verteilung	346
6.13	Weibull-Verteilung	350
6.13.1	Bestimmung der Parameter	350
6.13.2	Konfidenzintervall für die Weibull-Gerade	351
6.14	Konfidenzintervalle für die Parameter einer linearen Regression	352
6.14.1	Die Schätzung einiger Standardabweichungen	352
6.14.2	Konfidenzintervalle für den Regressionskoeffizienten, für den Achsenabschnitt und für die Restvarianz	357
6.14.3	Konfidenzintervalle und Prädiktionsintervalle für die Regressionsgerade	357
6.14.4	Inverse Prädiktion aus einer linearen Regression	361
6.15	Konfidenzintervall für den Korrelationskoeffizienten ρ	362
6.16	Übereinstimmung und Präzision von Messwerten	364
6.16.1	Übereinstimmung von Messwertreihen nach Bland-Altman	364
6.16.2	Vergleich der Präzision zweier Messinstrumente oder zweier Messmethoden	366

6.16.3	Intraklassen-Korrelation: Interrater-Reliabilität	367
6.17	Toleranzgrenzen	369
6.17.1	Verteilungsunabhängige Toleranzgrenzen	371
6.18	Voraussageintervalle (Prädiktionsintervalle)	372
6.18.1	Voraussageintervalle für den Mittelwert von m künftigen Beobachtungen	374
6.18.2	Voraussageintervalle für alle künftigen Beobachtungen	374
6.18.3	Voraussageintervalle für die Standardabweichung von m künftigen Beobachtungen	376
6.19	Bayes-Schätzung	376
6.19.1	A-priori Verteilungen (Prior)	379
6.19.2	Parameterschätzung nach Bayes	381
7	Hypothesentest	385
7.1	Der statistische Test	385
7.1.1	Entscheidungsprinzipien	386
7.1.2	Statistische Hypothesen und Testentscheidungen	387
7.1.3	Wie oft wird eine wahre Nullhypothese abgelehnt?	390
7.1.4	Statistischer Test - Schritt für Schritt	391
7.1.4.1	Der Likelihood-Quotient; das Neyman-Pearson Lemma	394
7.1.5	Powerfunktion und Operationscharakteristik	395
7.1.6	Die Operationscharakteristik	400
7.1.6.1	Die OC-Kurve in der Qualitätskontrolle	400
7.1.7	Die Formulierung von Hypothesen	403
7.1.7.1	Alternativhypothese als Gegensatz zur Nullhypothese	403
7.1.7.2	Durch Daten angeregte Hypothesen	403
7.1.8	Der P-Wert nach R.A. Fisher	404
7.1.9	Äquivalenztests	405
7.1.10	Verteilungsunabhängige Verfahren	406
7.2	Tests der Verteilung (Anpassungstests, goodness of fit)	408
7.2.1	Der Quotient R/s	409
7.2.2	Überprüfung des 3. und 4. Momentes	409
7.2.3	Das Wahrscheinlichkeitsnetz, QQ-Plot	413
7.2.4	Box-Cox-Transformation	416
7.2.5	Der Chiquadrat-Anpassungstest	417
7.2.6	Kolmogoroff-Smirnoff-Anpassungstest	421
7.2.6.1	Anpassung an eine Poisson-Verteilung	424
7.2.7	Shapiro-Wilk Test	426
7.2.8	Anderson-Darling Test	426
7.2.9	Ausreißerproblem	427
7.2.9.1	Grubbs-Test für Ausreißer	429
7.2.9.2	Q-Test nach Dixon für $3 \leq n \leq 8$	430
7.2.9.3	Standardisierte Extremabweichungen	430
7.3	Einstichprobenverfahren	431
7.3.1	Hypothesen zu Wahrscheinlichkeiten	432
7.3.1.1	Binomialtest	432
7.3.1.2	Binomialtest - Approximation durch die Normalverteilung	434
7.3.1.3	Binomialtest - Fallzahlab-schätzung	435
7.3.1.4	Binomialtest: Likelihood-Quotienten-Test	437
7.3.2	Hypothesen zu Erwartungswerten, die sich auf einen empirischen Mittelwert beziehen	438
7.3.2.1	Einstichproben-t-Test	438
7.3.2.2	Fallzahlab-schätzung zum Einstichproben-t-Test	441

7.3.2.3	Einstichprobentest auf Äquivalenz	442
7.3.3	Einstichproben-Median-Test	444
7.3.4	Vergleich einer empirischen Varianz mit ihrem Parameter	445
7.3.5	Asymptotischer Test für den Variationskoeffizienten	446
7.3.6	Prüfung der Zufallsmäßigkeit einer Folge von Alternativdaten oder von Messwerten	447
7.3.6.1	Die sukzessive Differenzenstreuung	447
7.3.6.2	Der Iterationstest für die Prüfung, ob eine Folge von Alternativdaten oder von Messwerten zufallsgemäß ist	449
7.3.6.3	Phasenhäufigkeitstest von Wallis und Moore	450
7.3.6.4	Der Vorzeichen-Trendtest von Cox und Stuart	452
7.3.6.5	Variabilität der zentralen Tendenz	453
7.3.7	Prüfung der Erwartungswerte von Poisson-Verteilungen	453
7.4	Zweistichprobenverfahren	455
7.4.1	Vergleich zweier Varianzen (F-Test)	455
7.4.1.1	Varianzvergleich bei kleinem bis mittlerem Stichprobenumfang	455
7.4.1.2	Varianzvergleich bei mittlerem bis großem Stichprobenumfang	457
7.4.1.3	Varianzvergleich bei großem bis sehr großem Stichprobenumfang ($n_1, n_2 \gtrsim 100$)	458
7.4.1.4	Minimale Stichprobenumfänge für den F-Test	458
7.4.1.5	Vergleich der Streuung zweier kleiner Stichproben nach Pillai und Buenaventura	458
7.4.2	Rangdispersionstest von Siegel und Tukey	459
7.4.3	Ansari-Bradley-Test	463
7.4.4	t-Test für unabhängige Stichproben	465
7.4.4.1	Unbekannte aber gleiche Varianzen	465
7.4.4.2	t-Test bei unbekanntem Varianzen, die möglicherweise ungleich sind	470
7.4.4.3	Fallzahlabschätzung für den t-Test für zwei unabhängige Stichproben	473
7.4.5	t-Test für Paardifferenzen	476
7.4.5.1	Gepaarte Beobachtungen	476
7.4.5.2	Absolute oder prozentuale Veränderungen	478
7.4.5.3	t-Test für paarweise angeordnete Messwerte	479
7.4.5.4	Prüfung der Gleichheit zweier Varianzen paariger Stichproben	481
7.4.6	Wilcoxon-Rangsummentest für zwei unabhängige Stichproben (U-Test)	481
7.4.6.1	Der U-Test bei Rangaufteilung	488
7.4.6.2	Effektstärke im Vergleich unabhängiger Stichproben	491
7.4.6.3	Fallzahlabschätzung für den U-Test	494
7.4.7	Wilcoxon-Paardifferenzentest	494
7.4.7.1	Konfidenzintervall für den Median $\hat{\mu}_d$ der Paardifferenzen	498
7.4.7.2	Der Maximum-Test für Paardifferenzen	499
7.4.7.3	Der Vorzeichentest von Dixon und Mood	499
7.4.7.4	Stichprobenumfang für den Vorzeichentest / Wilcoxon-Test für Paardifferenzen	501
7.4.8	Vergleich zweier unabhängiger Stichproben nach Kolmogoroff und Smirnov	502
7.4.9	Cramér-von Mises Test	505
7.4.10	Einige weitere verteilungsunabhängige Verfahren für den Vergleich unabhängiger Stichproben	506

7.4.10.1	Der Zwei-Stichproben Dispersionstest: Count Five ($n_x = n_y > 25$)	507
7.4.10.2	Rosenbaumsche Schnelltests	508
7.4.10.3	Permutationstest, Randomisierungstest	508
7.4.10.4	Der Vergleich zweier unabhängiger Stichproben: Schnelltest nach Tukey	512
7.4.10.5	Der Median-Test	514
7.4.11	Zweistichprobentest auf Äquivalenz	515
7.4.11.1	Test auf Bioäquivalenz	516
7.5	Mehrstichprobenverfahren, varianzanalytische Methoden	519
7.5.1	Prüfung der Gleichheit mehrerer Varianzen	520
7.5.1.1	Prüfung der Gleichheit mehrerer Varianzen nach Hartley	520
7.5.1.2	Prüfung der Gleichheit mehrerer Varianzen nach Cochran	521
7.5.1.3	Prüfung der Gleichheit mehrerer Varianzen nach Bartlett	522
7.5.1.4	Robuster Test auf Varianzheterogenität nach Levene in der Brown-Forsythe-Version	524
7.5.2	Einfache Varianzanalyse	525
7.5.2.1	Stichprobenumfänge für den varianzanalytischen Mittelwertvergleich für maximal sechs Stichprobengruppen	529
7.5.3	Mehrfacher Hypothesentest	529
7.5.3.1	Multiple Testproblem	529
7.5.3.2	Adjustierung von P-Werten	532
7.5.4	Multiple paarweise Vergleiche und weitere Mittelwertverfahren	536
7.5.4.1	Beurteilung linearer Kontraste nach Scheffé	536
7.5.4.2	Bildung homogener Gruppen von Mittelwerten anhand des nach Hayter modifizierten LSD-Tests, eines Lücken-Tests für geordnete μ_i	539
7.5.4.3	Bereichsschätzung für μ_{max} und Auswahl der „besten“ \bar{x}_i	540
7.5.4.4	Multiple Vergleiche nach Tukey-Kramer	541
7.5.4.5	Multiple Vergleiche nach Games-Howell	544
7.5.4.6	Multiple Vergleiche nach Dunnett	544
7.5.4.7	Multipler Vergleich von Mittelwerten mit dem Gesamtmittel (Maximum-Modulus-Ansatz)	547
7.5.5	H-Test von Kruskal und Wallis	549
7.5.5.1	Multiple paarweise Vergleiche mittlerer Ränge	551
7.5.5.2	H-Test mit Stichproben-Untergruppen	556
7.5.5.3	H-Test-Variante: Vergleich eines Standards mit mehreren Behandlungen	557
7.5.5.4	Trendtest nach Jonckheere: Vergleich mehrerer geordneter Verteilungsfunktionen oder Mediane	558
7.5.6	Varianzanalyse für Messwiederholungen (Blockvarianzanalyse)	561
7.5.7	Friedman-Test	562
7.5.7.1	Multiple paarweise Vergleiche nach Wilcoxon und Wilcox	567
7.5.7.2	Page-Test für geordnete Alternativen, die auf Friedman- Rangsummen basieren	570
7.5.8	Zweifache Varianzanalyse	572
7.5.9	Analysc von wiederholten Messungen	578
7.5.9.1	Typisierung wiederholter Messungen	578
7.5.9.2	ANOVA für wiederholte Messungen (gemischte Modelle)	581
7.5.10	Prinzipien der Versuchsplanung	583
7.6	Die Analyse von Häufigkeiten	590

7.6.1	Vergleich zweier relativer Häufigkeiten	590
7.6.2	Analyse von Vierfeldertafeln	592
7.6.2.1	Fallzahl und Power zum Vierfeldertest	597
7.6.2.2	Minimales n für den Vierfeldertest	598
7.6.2.3	Vorsicht vor Trugschlüssen beim Vierfeldertest	599
7.6.3	Spezielle Risiko- und Effektmaße	600
7.6.3.1	Odds Ratio und relatives Risiko	600
7.6.3.2	Konfidenzintervalle für das Relative Risiko ψ und für die Odds Ratio ω	605
7.6.3.3	Stichprobenumfänge für das Relative Risiko ψ und für die Odds Ratio ω	606
7.6.3.4	Der expositionsbedingte Anteil Erkrankter: Population Attributable Risk	608
7.6.3.5	Number Needed to Treat (NNT)	610
7.6.4	Exakter Test nach R.A. Fisher	611
7.6.5	Äquivalenz zweier Binomialwahrscheinlichkeiten	614
7.6.6	Der von McNemar modifizierte Vorzeichentest	617
7.6.7	Test nach Mantel-Haenszel	622
7.6.7.1	Breslow-Day-Test	624
7.6.7.2	Die Kombination von Vierfeldertafeln	626
7.6.8	Der $k \cdot 2$ -Felder- χ^2 -Test nach Brandt und Snedecor	628
7.6.8.1	Homogenitätstest nach Ryan (Lückentest)	635
7.6.9	Cochran-Armitage Test auf linearen Trend	637
7.6.10	Vergleich mehrerer Anteile mit einem vorgegebenen Anteilswert (Standard)	640
7.6.11	Die Analyse von Kontingenztafeln des Typs $r \cdot c$	642
7.6.11.1	Kontingenzkoeffizient - Stärke des Zusammenhangs	651
7.6.11.2	Lokalisation der stochastischen Abhängigkeit nach Hommel	652
7.6.11.3	Simultane Paarvergleiche nach Royen	654
7.6.11.4	Prüfung auf Trend: Der Anteil der linearen Regression an der Gesamtvariation	657
7.6.12	Bowker-Test auf Symmetrie in quadratischen Mehrfeldertafeln	660
7.6.13	Marginalhomogenitätstest nach Lehman	661
7.6.14	Q-Test nach Cochran	663
7.6.14.1	Simultane Konfidenzintervalle für paarweise Differenzen von Erfolgsanteilen	664
7.6.15	Cohen's Kappa-Koeffizient κ	666
7.6.15.1	Das gewichtete κ (Kappa)	669
7.6.15.2	Das κ für mehrfache Beurteilungen (Multi-Rater)	670
7.6.16	Kendalls Konkordanzkoeffizient W	672
7.7	Hypothesentests zur Korrelation und Regression	674
7.7.1	Prüfung des Vorhandenseins einer Korrelation	675
7.7.2	\hat{z} -Transformation nach R.A. Fisher	679
7.7.2.1	Weitere Anwendungen der \hat{z} -Transformation	680
7.7.2.2	Der Vergleich mehrerer Korrelationskoeffizienten	682
7.7.3	Prüfung der Linearität einer Regression	683
7.7.4	Hypothesentests zu den Parametern einer Regression	684
7.7.5	Prüfung des Rang-Korrelationskoeffizienten ρ_S (Rho-S)	688
7.7.6	Prüfung des Rang-Korrelationskoeffizienten τ (Tau)	689

8	Statistische Modellbildung	691
8.1	Einführung	691
8.2	Lineare Regressionsmodelle	693
8.2.1	Die einfache lineare Regression	693
8.2.2	Multiple lineare Regression	698
8.2.3	Überwindung von Multikollinearität in Regressionsmodellen	702
8.2.4	Analyse der Residuen im linearen Modell	703
8.2.5	Hypothesentest und Konfidenzintervalle zum linearen Modell	706
8.2.6	Verfahren der Variablenauswahl	708
8.2.7	Nominalskalierte Einflussgrößen	711
8.3	Varianzanalyse im linearen Modell	712
8.3.1	Einfaktorielle Varianzanalyse	712
8.3.1.1	Erwartungswert-Parametrisierung	713
8.3.1.2	Effekt-Parametrisierung: Dummy-Codierung	713
8.3.1.3	Effekt-Parametrisierung: Effekt-Codierung	714
8.3.1.4	Varianzkomponenten - ANOVA	715
8.3.2	Zweifaktorielle Varianzanalyse	716
8.4	Logistische Regression	720
8.4.1	Hypothesentest im logistischen Regressionsmodell	724
8.4.2	Multiple logistische Regression	725
8.4.3	Interpretation der Regressionskoeffizienten	729
8.4.4	Variablenauswahl im Rahmen der Modellbildung	730
8.4.5	Residuenanalyse	731
8.4.6	Pseudo-Bestimmtheitsmaße (Pseudo- R^2)	733
8.5	Poisson-Regression und loglineare Modelle	733
8.5.1	Poisson-Regression	734
8.5.1.1	Dispersionsindex und Poisson-Regression	736
8.5.2	Analyse von Kontingenztafeln	737
8.5.3	Loglineares Modell am Beispiel von 2 Faktoren	740
8.5.4	Dreidimensionale Kontingenztafeln	742
8.5.4.1	Modellbildung unter verschiedenen Restriktionen	742
8.5.4.2	Modellauswahl im loglinearen Ansatz	744
8.5.4.3	Fünf Einschränkungen und Hinweise zum loglinearen Modell	747
8.6	Modelle zu wiederholten Messungen	747
8.6.1	Varianzanalyse für wiederholte Messungen	748
8.6.2	Lineare gemischte Modelle	751
8.6.3	Analyse von Cluster-Daten	758
8.6.4	Verallgemeinerte Schätzgleichungen	763
8.7	Analyse von Überlebenszeiten	767
8.7.1	Kaplan-Meier Schätzung der Überlebensfunktion	770
8.7.2	Der Logrank-Test	775
8.7.3	Parametrische Modelle für Überlebenszeiten	777
8.7.4	Das Cox-Regressionsmodell	780
8.7.4.1	Regressionsmodelle zu Überlebenszeiten	780
8.7.4.2	Das Proportional-Hazards Modell von Cox	780
8.7.4.3	Parameter-Schätzung zum Cox-Modell	781
8.7.4.4	Interpretation der Parameter	783
8.7.4.5	Modellbildung: zur Auswahl geeigneter Einflussgrößen	785
8.7.4.6	Güte der Modellanpassung, Residualanalyse	787

9 Einführung in R	791
9.1 Das Konsolfenster	791
9.2 Hilfestellung in R	794
9.3 Objekte in R	795
9.3.1 Vektoren	795
9.3.2 Erzeugen von Vektoren und Dateneingabe	797
9.3.3 Faktoren in R, Klassenbildung	798
9.3.4 Erzeugen von Matrizen	799
9.3.5 Rechnen mit Matrizen in R am Beispiel der Herleitung einer Kovarianzmatrix	799
9.3.6 Tabellen: Daten in Rahmen (data.frame)	801
9.4 Fehlende Angaben	803
9.5 Auswahl und Sortierung von Daten	804
9.6 Ablaufsteuerung: logische Bedingungen und Funktionen in R	806
9.7 Einige mathematische und statistische Funktionen	809
9.8 Modellbildung in R	811
9.9 Einfache graphische Funktionen und Hilfsmittel	813
Literaturverzeichnis	819
Namensverzeichnis	842
Verzeichnis zur Anwendung der R-Funktionen	848
Sachverzeichnis	852