

Lothar Sachs  
Jürgen Hedderich

# Angewandte Statistik

Methodensammlung mit R

Zwölfte, vollständig neu bearbeitete Auflage  
mit 142 Abbildungen  
und 180 Tabellen

---

# Inhaltsverzeichnis

<b>1</b>	<b>Einführung</b>	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	7
1.3.1	Wiederholbare Erfahrungen	7
1.3.2	Deskriptive Statistik	8
1.3.3	Explorativer Ansatz	9
1.3.4	Konfirmativer Ansatz	10
1.3.5	Merkmale, Grundgesamtheit, Stichprobe	11
1.3.6	Stichproben	12
1.3.7	Zufallsstichproben	13
1.4	Datenanalyse	14
1.4.1	Klassierung von Merkmalen	14
1.4.2	Skalierung von Variablen	15
1.4.3	Daten	17
<b>2</b>	<b>Grundlagen aus der Mathematik</b>	20
2.1	Logische und relationale Operatoren	20
2.2	Mengen	21
2.2.1	Begriffsbildung	21
2.2.2	Mengenoperationen	22
2.3	(Grund-) Rechenarten	23
2.3.1	Summen und Produkte	24
2.3.2	Potenzen und Wurzeln	29
2.3.3	Logarithmen	30
2.3.4	Rundungen	32
2.3.5	Rechnen mit fehlerbehafteten Zahlen	33
2.4	Einführung in die Matrixalgebra	34
2.4.1	Definition und Schreibweise	34
2.4.2	Matrixoperationen	35
2.4.3	Determinanten	39
2.4.4	Die Inverse Matrix	39
2.4.5	Lineare Abhängigkeit, Rang einer Matrix	40
2.4.6	Lineare Gleichungssysteme	41

2.4.7	Eigenwerte und Eigenvektoren	41
2.5	Funktionen	42
2.5.1	Lineare Funktionen	43
2.5.2	Nichtlineare Funktionen	44
2.5.3	Periodische Funktionen	45
2.5.4	Exponentialfunktion und logarithmische Funktion	46
2.5.5	Flächen unter einer Funktion - Integrale	46
2.6	Kombinatorik	47
2.6.1	Permutationen	47
2.6.2	Kombinationen - der Binomialkoeffizient	49
2.6.3	Kombinationen mit Wiederholungen und mit Berücksichtigung der Anordnung	50
2.6.4	Zerlegung einer Menge	52
2.6.5	Das Pascalsche Dreieck	52
2.6.6	Der Multinomialkoeffizient	54
3	<b>Deskriptive Statistik</b>	55
3.1	Häufigkeiten	56
3.1.1	Absolute und relative Häufigkeiten	56
3.1.2	Sinnvolle Quotienten: Verhältniszahlen	57
3.1.3	Prozentwerte	59
3.1.4	Torten- und Balkendiagramme	59
3.1.5	Tabellen	60
3.1.6	Bedingte Häufigkeiten	62
3.2	Beschreibung von Ordinaldaten	62
3.2.1	Medianwert und andere Quartile	63
3.2.2	Quantile	64
3.2.3	Streuung ordinal skalierten Daten	65
3.2.4	Punktdiagramm und Box-Plot	66
3.2.5	Korrelationskoeffizient nach Kendall	67
3.3	Beschreibung von metrischen Daten	68
3.3.1	Arithmetischer Mittelwert	68
3.3.2	Standardabweichung, Varianz	70
3.3.3	Variationskoeffizient	72
3.3.4	Der $(\bar{x} \pm s)$ -Bereich	73
3.3.5	Klassierte Messwerte; Berechnung des Mittelwertes und der Standardabweichung	74
3.3.6	Das gewogene arithmetische Mittel, die gewogene Varianz und das gewichtete arithmetische Mittel	74
3.3.7	Geometrischer Mittelwert	76
3.3.8	Harmonischer Mittelwert	78
3.4	Häufigkeitsverteilung	80
3.4.1	Histogramm	80
3.4.2	Stamm-Blatt Darstellung	83
3.5	Konzentration; Gini Index	83
3.6	Maßzahlen für den Zusammenhang metrischer Daten	85
3.6.1	Punktwolken	85
3.6.2	Die empirische Kovarianz	85
3.6.3	Der empirische Korrelationskoeffizient	87
3.6.4	Der Rangkorrelationskoeffizient	88
3.6.5	Typisierung korrelativer Zusammenhänge	90
3.6.6	Die lineare Regression	91

3.6.7	Spezielle Schätzungen der Regressionsgeraden	93
3.6.8	Robuste lineare Regression	97
3.7	Nichtlineare Regression	99
3.7.1	Einige linearisierende Transformationen	104
<b>4</b>	<b>Wahrscheinlichkeiten</b>	<b>108</b>
4.1	Zufallsexperiment, Ereignis	109
4.2	Begriff der Wahrscheinlichkeit	111
4.2.1	Definition nach Laplace	111
4.2.2	Axiome nach Kolmogoroff	113
4.3	Bedingte Wahrscheinlichkeiten, stochastische Unabhängigkeit	116
4.3.1	Bedingte Wahrscheinlichkeit	116
4.3.2	Stochastische Unabhängigkeit	120
4.4	Bayessches Theorem	128
4.4.1	Bayessches Theorem und Pfadregel	129
4.4.2	Acht Beispiele zum Bayesschen Theorem	130
4.5	Der diagnostische Test	132
4.5.1	ROC - Analyse	136
4.5.2	Der Likelihoodquotient	137
4.6	Maßzahlen in der Epidemiologie	139
4.6.1	Prävalenz und Inzidenz	139
4.6.2	Standardisierungen	141
<b>5</b>	<b>Zufallsvariablen, Verteilungen</b>	<b>144</b>
5.1	Die Zufallsvariable	144
5.1.1	Wahrscheinlichkeitsfunktion, Wahrscheinlichkeitsdichte und Verteilungsfunktion	145
5.2	Maßzahlen zur Kennzeichnung der Verteilung	150
5.2.1	Erwartungswert	151
5.2.2	Varianz	152
5.2.3	Momente: Schiefe und Exzess	154
5.3	Diskrete Verteilungen	161
5.3.1	Das Urnenmodell	161
5.3.2	Gleichverteilung	163
5.3.3	Binomialverteilung	164
5.3.4	Poisson-Verteilung	173
5.3.5	Negative Binomial-Verteilung	181
5.3.6	Hypergeometrische Verteilung	186
5.4	Stetige Verteilungen	190
5.4.1	Gleichverteilung	190
5.4.2	Normalverteilung	191
5.4.3	Lognormalverteilung	204
5.4.4	Exponentialverteilung	207
5.4.5	Weibull-Verteilung	209
5.5	Testverteilungen	210
5.5.1	Student-Verteilung ( $t$ )	211
5.5.2	Chiquadrat-Verteilung ( $\chi^2$ )	215
5.5.3	Fisher-Verteilung ( $F$ )	216
5.5.4	Verteilungen wichtiger Stichprobenfunktionen aus normalverteilten Grundgesamtheiten	222
5.6	Verteilung zweidimensionaler Zufallsvariablen	224
5.6.1	Modellbildung	224

5.6.2	Randverteilungen und Unabhängigkeit	226
5.6.3	Korrelationskoeffizient	230
5.6.4	Zweidimensionale Normalverteilung	231
5.6.5	Multinomialverteilung (Polynomialverteilung)	233
<b>6</b>	<b>Schätzen</b>	<b>235</b>
6.1	Zufallsstichproben und Zufallszahlen	235
6.1.1	Spezielle Stichprobenverfahren	238
6.2	Das Schätzen von Parametern	239
6.2.1	Vorbemerkungen	240
6.2.2	Wünschenswerte Eigenschaften von Schätzfunktionen	241
6.2.3	Gesetz der großen Zahlen	243
6.2.4	Der mittlere quadratische Fehler	243
6.3	Schätzverfahren für Maßzahlen einer Verteilung	245
6.3.1	Momentenmethode	245
6.3.2	Schätzung nach der größten Erwartung (MLE)	246
6.3.3	Kleinster Fehler (OLS)	251
6.4	Konfidenzintervalle	252
6.5	Konfidenzintervall für einen Anteilswert aus einer dichotomen Grundgesamtheit ( $\pi$ )	254
6.5.1	Approximation durch die Normalverteilung	256
6.5.2	Sonderfälle mit $\hat{p} = 0$ bzw. $\hat{p} = 1$	258
6.5.3	Schnellschätzung der Vertrauensgrenzen anhand einer beobachteten relativen Häufigkeit nach Clopper und Pearson	259
6.5.4	Angenähertes 95%-Konfidenzintervall für $\pi_1 - \pi_2$ ( $n_1$ und $n_2$ groß)	261
6.5.5	Schätzung des Mindestumfangs einer Stichprobe bei ausgezählten Werten	262
6.6	Konfidenzintervalle für $\mu$ bei Normalverteilung	263
6.6.1	Vertrauenswahrscheinlichkeit und Irrtumswahrscheinlichkeit	263
6.6.2	Konfidenzintervall für den Erwartungswert $\mu$	265
6.6.3	Konfidenzintervall für die Differenz $\mu_1 - \mu_2$	267
6.6.4	Das Konfidenzintervall für den Erwartungswert $\mu_d$ der Paardifferenzen	269
6.6.5	Konfidenzintervall für das Verhältnis $\mu_1/\mu_2$	269
6.6.6	Mindestzahl von Beobachtungen zur Schätzung eines Mittelwertes	271
6.7	Konfidenzintervall für die mittlere absolute Abweichung	271
6.8	Konfidenzintervall für den Median	273
6.8.1	Angenäherte verteilungsunabhängige Konfidenzintervalle für beliebige Quantile	274
6.9	Konfidenzintervalle nach dem Bootstrap-Verfahren	275
6.10	Konfidenzintervall für $\sigma^2$ bzw. $\sigma$	278
6.10.1	Konfidenzintervall für den Variationskoeffizienten $\gamma$	279
6.10.2	Konfidenzintervall für den Quotienten zweier Varianzen $\sigma_1^2/\sigma_2^2$	279
6.10.3	Mindestzahl von Beobachtungen zur Schätzung einer Standardabweichung	280
6.11	Konfidenzintervall für den Erwartungswert $\lambda$ einer Poisson-Verteilung	280
6.12	Weibull-Verteilung	284
6.12.1	Bestimmung der Parameter	284
6.12.2	Das Konfidenzintervall für die Weibull-Gerade	285
6.13	Konfidenzintervalle für die Parameter einer linearen Regression	286
6.13.1	Die Schätzung einiger Standardabweichungen	286
6.13.2	Konfidenzintervalle für den Regressionskoeffizienten, für den Achsenabschnitt und für die Restvarianz	291
6.13.3	Konfidenzintervalle und Prädiktionsintervalle für die Regressionsgerade	292
6.13.4	Inverse Prädiktion aus einer linearen Regression	296

6.13.5	Das Konfidenzintervall für den Korrelationskoeffizienten $\rho$ .....	297
6.14	Toleranzgrenzen .....	299
6.14.1	Verteilungsunabhängige Toleranzgrenzen .....	301
6.15	Übereinstimmung von Messwerten nach Bland-Altman .....	302
<b>Hypothesentest</b>	.....	<b>305</b>
7.1	Der statistische Test .....	305
7.1.1	Entscheidungsprinzipien .....	306
7.1.2	Statistische Hypothesen und Testentscheidungen .....	307
7.1.3	Statistischer Test - Schritt für Schritt .....	310
7.1.4	Powerfunktion und Operationscharakteristik .....	314
7.1.5	Die Operationscharakteristik .....	318
7.1.6	Die Formulierung von Hypothesen .....	321
7.1.7	Der P-Wert nach R.A. Fisher .....	322
7.1.8	Äquivalenztests .....	324
7.1.9	Verteilungsunabhängige Verfahren .....	325
7.2	Tests der Verteilung (goodness of fit) .....	327
7.2.1	Der Quotient $R/s$ .....	327
7.2.2	Überprüfung des 3. und 4. Momentes .....	328
7.2.3	Das Wahrscheinlichkeitsnetz, QQ-Plot .....	330
7.2.4	Der Chiquadrat-Anpassungstest .....	333
7.2.5	Kolmogoroff-Smirnoff-Anpassungstest .....	337
7.2.6	Shapiro-Wilk Test .....	341
7.2.7	Anderson-Darling Test .....	342
7.2.8	Ausreißerproblem .....	343
7.3	Einstichprobenverfahren .....	347
7.3.1	Hypothesen zu Wahrscheinlichkeiten .....	347
7.3.2	Hypothesen zu Erwartungswerten, die sich auf einen empirischen Mittelwert beziehen .....	352
7.3.3	Einstichproben-Median-Test .....	358
7.3.4	Vergleich einer empirischen Varianz mit ihrem Parameter .....	359
7.3.5	Prüfung der Zufällsmäßigkeit einer Folge von Alternativdaten oder von Messwerten .....	360
7.3.6	Prüfung der Erwartungswerte von Poisson-Verteilungen .....	366
7.4	Zweistichprobenverfahren .....	367
7.4.1	Vergleich zweier Varianzen (F-Test) .....	367
7.4.2	Rangdispersionstest von Siegel und Tukey .....	371
7.4.3	Ansari-Bradley-Test .....	375
7.4.4	t-Test für unabhängige Stichproben .....	377
7.4.5	t-Test für Paardifferenzen .....	387
7.4.6	Wilcoxon Rangsummentest für zwei unabhängige Stichproben .....	391
7.4.7	Wilcoxon-Paardifferenzentest .....	400
7.4.8	Vergleich zweier unabhängiger Stichproben nach Kolmogoroff und Smirnoff .....	405
7.4.9	Cramér-von Mises Test .....	408
7.4.10	Einige weitere verteilungsunabhängige Verfahren für den Vergleich unabhängiger Stichproben .....	410
7.4.11	Zweistichprobenentest auf Äquivalenz .....	414
7.5	Mehrstichprobenverfahren, varianzanalytische Methoden .....	418
7.5.1	Prüfung der Gleichheit mehrerer Varianzen .....	418
7.5.2	Einfache Varianzanalyse .....	423
7.5.3	Multiple Vergleiche, Multiples Testproblem .....	428

7.5.4	H-Test von Kruskal und Wallis	442
7.5.5	Varianzanalyse für Messwiederholungen (Blockvarianzanalyse)	454
7.5.6	Friedman-Test	456
7.5.7	Zweifache Varianzanalyse	465
7.5.8	Prinzipien der Versuchsplanung	470
7.6	Die Analyse von Häufigkeiten	477
7.6.1	Vergleich zweier relativer Häufigkeiten	477
7.6.2	Die Analyse von Vierfeldertafeln	479
7.6.3	Odds Ratio und relatives Risiko	487
7.6.4	Exakter Fisher-Test	496
7.6.5	Der von McNemar modifizierte Vorzeichentest	497
7.6.6	Test nach Mantel-Haenszel	503
7.6.7	Der $k \cdot 2$ -Felder- $\chi^2$ -Test nach Brandt und Snedecor	507
7.6.8	Cochran-Armitage Test auf linearen Trend	516
7.6.9	Die Analyse von Zweiwegtafeln des Typs $r \cdot c$	519
7.6.10	Bowker-Test auf Symmetrie in quadratischen Mehrfeldertafeln	535
7.6.11	Cohen's Kappa-Koeffizient $\kappa$	537
7.7	Hypothesentests zur Korrelation und Regression	543
7.7.1	Prüfung des Vorhandenseins einer Korrelation	544
7.7.2	$z$ -Transformation nach R.A. Fisher	548
7.7.3	Weitere Anwendungen der $z$ -Transformation	549
7.7.4	Der Vergleich mehrerer Korrelationskoeffizienten	551
7.7.5	Prüfung der Linearität einer Regression	552
7.7.6	Prüfung der Regressionsparameter	553
7.7.7	Prüfung des Rang-Korrelationskoeffizienten $\rho_S$	557
<b>8</b>	<b>Statistische Modellbildung</b>	<b>560</b>
8.1	Einführung	560
8.2	Regressionsmodelle	562
8.2.1	Die einfache lineare Regression	562
8.2.2	Die multiple lineare Regression	566
8.2.3	Verfahren der Variablenauswahl	573
8.2.4	Nominalskalierte Einflussgrößen	576
8.3	Varianzanalyse im linearen Modell	577
8.3.1	Einfaktorielle Varianzanalyse	577
8.3.2	Zweifaktorielle Varianzanalyse	581
8.4	Logistische Regression	585
8.4.1	Hypothesentest im logistischen Regressionsmodell	589
8.4.2	Multiple logistische Regression	591
8.4.3	Interpretation der Regressionskoeffizienten (odds)	594
8.4.4	Variablenauswahl im Rahmen der Modellbildung	595
8.4.5	Residuenanalyse	597
8.5	Log-lineare Modelle	598
8.5.1	Kontingenztafeln	598
8.5.2	Log-lineares Modell am Beispiel von 2 Faktoren	602
8.5.3	Drei-dimensionale Kontingenztafeln	604
8.6	Analyse von Überlebenszeiten	609
8.6.1	Kaplan-Meier Schätzung der Überlebensfunktion	611
8.6.2	Der Logrank-Test	616
8.6.3	Parametrische Modelle für Überlebenszeiten	618
8.6.4	Das Cox-Regressionsmodell	621

<b>9</b>	<b>Einführung in R</b> .....	633
9.1	Das Konsolfenster .....	633
9.2	Objekte in R .....	636
9.3	Hilfestellung in R .....	637
9.4	Erzeugen von Daten in R mittels Funktionen .....	638
9.5	Dateneingabe: „Daten in Rahmen“ (data.frame) .....	639
9.6	Auswahl und Sortierung von Daten .....	640
9.7	Ablaufsteuerung: logische Bedingungen und Funktionen in R .....	641
9.8	Einige mathematische und statistische Funktionen .....	643
9.8.1	Formulierung von Modellgleichungen .....	645
9.9	Einfache graphische Funktionen und Hilfsmittel .....	646
<b>10</b>	<b>Übungsaufgaben zu ausgewählten Themen</b> .....	650
	<b>Lösungen der Übungsaufgaben</b> .....	657
	<b>Literaturverzeichnis</b> .....	667
	<b>Namensverzeichnis</b> .....	680
	<b>Sachverzeichnis</b> .....	684