

Inhaltsverzeichnis

1 Einführung in die Statistik	1
1.1 Eingrenzungen des Begriffs „Statistik“	1
1.1.1 Komprimierende Kennwerte	1
1.1.2 Staatswissenschaftliche Disziplin	2
1.1.3 Wahrscheinlichkeitstheoretisch fundierte Datenwissenschaft	2
1.2 Grundzüge statistischer Methoden	4
1.2.1 System und Zufall	4
1.2.2 Irrtums- und Sicherheitswahrscheinlichkeiten	6
1.2.3 Deskription und Induktion	8
1.2.4 Empirischer und theoretischer Kalkül	10
1.3 Teilkreiche, Spezialgebiete und Grundlagen	11
1.3.1 Datengewinnung und Aufbereitung	11
1.3.2 Spezielle Analysemethoden und Instrumente	14
1.3.3 Wissenschaftsspezifische Ausrichtungen	17
1.3.4 Mathematische und technische Grundlagen	18
Teil 1: Deskriptive Statistik – Empirischer Kalkül	19
2 Einführung in die deskriptive Statistik	20
2.1 Daten, Datensätze und Variablen	20
2.2 Grundgesamtheit, Merkmalstypen und Skalierungsarten	22
2.3 Empirische Verteilungen und Zusammenhänge	24
3 Gesamtbeschreibung empirischer Verteilungen	27
3.1 Tabellarische Darstellungsmöglichkeiten	27
3.1.1 Elementare Begriffe und Notation	27
3.1.2 Häufigkeitstabellen bei Urlisten	29
3.1.3 Häufigkeitstabellen bei klassierten Daten	31
3.2 Grafische Darstellungsmöglichkeiten	32
3.2.1 Kreis-, Säulen- und Balkendiagramme	32
3.2.2 Stamm-Blatt-Diagramme	34
3.2.3 Histogramme (Häufigkeitsdichten)	35
3.2.4 Boxplots	39

3.3	Empirische Verteilungsfunktion (EVF).....	43
3.3.1	Häufigkeitsfunktion und EVF bei URListen	43
3.3.2	Häufigkeitsdichtefunktion und EVF bei klassierten Daten	45
4	Spezifizierende Beschreibung empirischer Verteilungen	49
4.1	Spezifika empirischer Verteilungen	49
4.2	Lagekennwerte	52
4.2.1	Arithmetisches Mittel	52
4.2.2	Median.....	56
4.2.3	Modalwert	59
4.2.4	Fechner'sche Lageregeln	59
4.3	Spezielle Lagekennwerte	61
4.3.1	Arithmetisches Mittel bei gruppierten Daten	61
4.3.2	Quantile	62
4.3.3	Geometrisches Mittel	65
4.4	Streuungskennwerte	69
4.4.1	Spannweite.....	69
4.4.2	Mittlere absolute Abweichungen	70
4.4.3	Median absoluter Abweichungen	72
4.4.4	Varianz, Standardabweichung und Schwankungsintervalle	73
4.5	Spezielle Streuungskennwerte	76
4.5.1	Varianz bei gruppierten Daten	76
4.5.2	Quantilsabstände	80
4.5.3	Variationskoeffizient	81
4.6	Standardisierung mittels Lage und Streuung	83
4.7	Messung von Schiefe	86
4.8	Darstellung und Messung von Konzentration	88
4.8.1	Lorenz-Kurve	88
4.8.2	Gini-Koeffizient	91
4.9	Spezifische Eigenschaften empirischer Kennwerte.....	94
4.9.1	Minimumeigenschaft des arithmetischen Mittels	94
4.9.2	Minimumeigenschaft des Medians	96
4.9.3	Transformationseigenschaften	98
4.9.4	Robustheit	103
5	Beschreibung und Analyse empirischer Zusammenhänge	105
5.1	Zusammenhänge zwischen kategorialen Merkmalen.....	105
5.1.1	Kontingenztabellen	105
	Verteilungen in Kontingenztabellen	105
	Empirische Abhängigkeit und Unabhängigkeit.....	110
	Konzeptionelle Vorüberlegungen für ein Zusammenhangsmaß	114

5.1.2	Zusammenhangsmaße für Kontingenztabellen	116
	Chi-Quadrat-Koeffizient	116
	Mittlere quadratische Kontingenz	119
	Kontingenzkoeffizient nach Pearson	120
	Transformationseigenschaften der Zusammenhangsmaße	122
5.1.3	Grafische Analysemöglichkeiten	124
	Gestapelte und gruppierte Säulendiagramme	124
	Segmentierte Säulen- und Balkendiagramme	125
	Assoziationsplots	126
	Mosaikplots	127
	Spineplots	128
5.2	Zusammenhänge zwischen metrischen Merkmalen	129
5.2.1	Grafische Analysemöglichkeiten	129
	Streudiagramme	129
	Streudiagramm-Matrizen	131
	Hexagonalplots	132
5.2.2	Zusammenhangsmaße für metrische Merkmale	132
	Empirische Kovarianz	132
	Empirischer Korrelationskoeffizient nach Pearson	137
	Empirische Kovarianz- und Korrelationsmatrizen	145
	Korrelationskoeffizient nach Spearman	146
5.2.3	Einfache lineare Regression	150
	Zweck und allgemeine Vorgehensweise	150
	KQ-Methode (L_2 -Regression)	153
	LAD-Methode (L_1 -Regression)	158
	Kritische Punkte und Alternativen	163
5.3	Ergänzende und vertiefende Themen	166
5.3.1	Zusammenhänge in anderen Skalierungsfällen	166
	Grafische Analysemöglichkeiten	166
	Zusammenhangsmaße	169
	Spezielle Regressionsmodelle	171
5.3.2	Grafische Darstellung höherdimensionaler Daten	174
	Höherdimensionale Mosaikplots	174
	3D-Punktwolken	176
	Biplots	176
5.3.3	Wichtige Aspekte bei der Analyse empirischer Zusammenhänge ..	183
	Empirische Zusammenhänge und Kausalität	183
	Spezielle Formen von Kausalität	186
	Systematische Verzerrungen	188
	Zufällige Schwankungen	198

Teil 2: Wahrscheinlichkeitsrechnung – Theoretischer Kalkül	203
6 Einführung in die Wahrscheinlichkeitsrechnung	204
6.1 Wichtige Grundbegriffe und Regeln	204
6.1.1 Interpretation von Zufall und Wahrscheinlichkeiten	204
6.1.2 Elementare Mengenlehre	209
6.1.3 Axiomatischer Wahrscheinlichkeitsbegriff und Regeln	216
6.2 Rechnen mit abhängigen und unabhängigen Ereignissen	226
6.2.1 Bedingte Wahrscheinlichkeiten und Rechenregeln	226
6.2.2 Stochastische Unabhängigkeit von Ereignissen	233
6.2.3 Kalkül nach der Formel von Bayes	241
7 Theoretische Verteilungen und Abhängigkeiten	245
7.1 Zufallsvariablen und deren Wahrscheinlichkeitsverteilungen	245
7.1.1 Eindimensionale Zufallsvariablen	245
Konzept und Ausblick	245
Diskrete Zufallsvariablen und Additionskalkül	247
Stetige Zufallsvariablen und Integrationskalkül	250
Theoretische Verteilungsfunktion	255
Wahrscheinlichkeitsverteilung	259
7.1.2 Mehrdimensionale Zufallsvariablen	260
Konzept und Ausblick	260
Zweidimensionale diskrete Zufallsvariablen	262
Zweidimensionale stetige Zufallsvariablen	268
Stochastische Abhängigkeit und Unabhängigkeit	274
Höherdimensionale Wahrscheinlichkeitsverteilungen	278
7.1.3 Verteilung von Funktionen von Zufallsvariablen	287
Funktionen einer Zufallsvariable	287
Funktionen aus mehreren Zufallsvariablen	293
7.2 Theoretische Kennwerte	296
7.2.1 Kennwerte in Bezug auf Lage und Streuung	297
Erwartungswert	297
Theoretische Varianz und Standardabweichung	312
Theoretische Quantile und theoretischer Median	316
7.2.2 Kennwerte in Bezug auf Abhängigkeiten	319
Bedingte Erwartungswerte und Varianzen	319
Theoretische Kovarianz und Korrelation	327
Theoretische Regressionskoeffizienten	333
7.2.3 Spezifische Eigenschaften theoretischer Kennwerte	334
Minimumeigenschaften von Lagekennwerten	334
Wichtige Transformationseigenschaften	335
Endliche und nicht endliche theoretische Momente	338
7.3 Spezielle eindimensionale Verteilungen	340
7.3.1 Spezielle diskrete Verteilungen	341
Elementare Kombinatorik	341

Einpunktverteilung	344
Bernoulli-Verteilung	344
Binomialverteilung	346
Poisson-Verteilung	350
7.3.2 Spezielle stetige Verteilungen	354
Stetige Gleichverteilung	354
Exponentialverteilung	356
Normalverteilung	361
7.4 Verteilung stochastischer Summen und Mittelwerte	369
7.4.1 Exakte Aussagen	370
Erwartungswerte und Varianzen	370
Verteilungen unter bestimmten Ausgangsverteilungen	373
7.4.2 Asymptotische und approximative Aussagen	379
Gesetz der großen Zahlen (GGZ)	379
Zentraler Grenzwertsatz (ZGWS)	384
8 Ergänzungen und Verallgemeinerungen	394
8.1 Weitere eindimensionale Verteilungen	394
8.1.1 Weitere diskrete Verteilungen	394
8.1.2 Weitere stetige Verteilungen	396
8.1.3 Sonstige eindimensionale Verteilungen	400
8.2 Ergänzungen zu asymptotischen Aussagen	402
8.2.1 Alternative GGZ- und ZGWS-Varianten	403
8.2.2 Stochastische Konvergenz und Verteilungskonvergenz	407
8.3 Einige multivariate Verallgemeinerungen	411
8.3.1 Multivariate Abhängigkeiten	411
Abhängigkeit und Unabhängigkeit von Zufallsvektoren	411
Bedingte Abhängigkeit und Unabhängigkeit	414
8.3.2 Kenngrößen multivariater Abhängigkeiten	415
Theoretische Kovarianz- und Korrelationsmatrix	415
Bedingte Kennwerte für $n > 2$	416
8.3.3 Sonstige Verallgemeinerungen	419
Teil 3: Induktive Statistik – Verbindung von Empirie und Theorie	427
9 Einführung in die induktive Statistik	428
9.1 Modellierung und Handhabung von Schätz- und Testproblemen	428
9.2 Was versteht man unter einer Stichprobe?	432
9.3 Formale Unterscheidungsebenen	442

10 Statistisches Schätzen	446
10.1 Punktschätzung	446
10.1.1 Schätzer und ihre Gütekriterien	446
10.1.2 Erwartungstreue	453
10.1.3 Erwartete quadratische Abweichung (MSE)	457
10.1.4 Konsistenz	459
10.1.5 Effizienz	466
10.2 Intervallschätzung	469
10.2.1 Was versteht man unter einem Konfidenzintervall?	469
10.2.2 Konfidenzintervalle für Erwartungswerte	474
10.2.3 Konfidenzintervalle für Erwartungswertdifferenzen	480
10.2.4 Weitere Konfidenzintervalle	490
10.2.5 Adäquatheit bestimmter Modellannahmen	494
10.3 Schätzmethoden	500
10.3.1 Momentenmethode	500
10.3.2 Maximum-Likelihood-Methode	503
10.3.3 Weitere Schätzmethoden	509
11 Statistisches Testen	511
11.1 Was versteht man unter einem Test?	511
11.1.1 Einführende Beispiele	511
11.1.2 Grundstruktur und Durchführung	523
11.1.3 Zusammenhang zur Intervallschätzung	526
11.2 Wichtige Aspekte beim Testen	528
11.2.1 Hypothesenwahl und Fehlerarten	529
11.2.2 Irrtumswahrscheinlichkeiten und Güte	532
11.2.3 p -Werte	543
11.2.4 Signifikanz vs. Relevanz	546
11.3 Ausgewählte Testverfahren	548
11.3.1 Tests über Erwartungswerte	548
11.3.2 Tests über Erwartungswertdifferenzen	552
11.3.3 Nichtparametrische χ^2 -Tests	556
Allgemeiner Überblick	556
χ^2 -Anpassungstest	557
χ^2 -Unabhängigkeitstest	565
11.3.4 Weitere Tests	570
Tests über Anteilswerte	570
Tests auf Unkorreliertheit und Unabhängigkeit	576
Sonstige Tests und Testmethoden	580
11.3.5 Allgemein zu beachtende Punkte	583

12 Das lineare Regressionsmodell	585
12.1 Einfaches lineares Regressionsmodell	585
12.1.1 Grundmodell und KQ-Methode	585
Statistisches Grundmodell	585
Herleitung der KQ-Schätzer	587
Eigenschaften der empirischen KQ-Regression	590
Bestimmtheitsmaß und Standardfehler der Regression	594
12.1.2 Modellannahmen und theoretische KQ-Regression	598
Ensembles von Modellannahmen	598
Theoretische KQ-Regression	605
12.1.3 Verteilungstheoretische Grundlagen	610
Verteilungen der KQ-Schätzer	610
Konsistenz und Effizienz der KQ-Schätzer	619
Schätzung der Varianzen der KQ-Schätzer	620
Verteilungen der Inferenzstatistiken	623
12.1.4 Schätzen und Testen	624
Konfidenzintervalle und Tests	624
Spezialfall: Binärer Regressor	626
Adäquatheit bestimmter Modellannahmen	629
Fallbeispiel 1: Bewässerung und Wachstum	636
Fallbeispiel 2: Klassengröße und Lernerfolg	640
12.2 Einführung in das multiple lineare Regressionsmodell	646
12.2.1 Partielle lineare KQ-Regression	646
Empirische partielle Regression	646
Theoretische partielle Regression	652
Verbindung von Empirie und Theorie	654
12.2.2 Multiple lineare KQ-Regression	656
Empirische multiple Regression	656
Theoretische multiple Regression	659
Verbindung von Empirie und Theorie	661
Statistische Modelle und Inferenz	661
12.2.3 Fallbeispiele	670
Fallbeispiel 2 fortgesetzt: Determinanten des Lernerfolgs	670
Fallbeispiel 3: Gewicht und Geschlecht	672
Fallbeispiel 4: Binäre Regressoren und ANOVA-Modelle	674
Anhang: Tabellen	678
Stichwortverzeichnis	681
Literaturverzeichnis	690

