

Inhalt

Vorwort	15
1 Einführung	17
1.1 Statistik und Sozialwissenschaften	17
1.2 Warum Statistik?	19
1.3 Statistische Modelle	19
1.4 Aufbau des Buchs	21
A Grundlagen	25
2 Vom Erhebungsinstrument zur Datenmatrix	25
2.1 Operationalisierung und Messung	25
2.2 Skalenniveau	28
2.3 Zulässige Transformationen	31
2.4 Die Datenmatrix: Untersuchungseinheiten, Variablen und Ausprägungen	36
2.5 Zusammenfassung	39
3 Univariate Häufigkeitsverteilungen	41
3.1 Häufigkeitstabellen	43
3.1.1 Häufigkeitstabellen für ungruppierte Daten	44
3.1.2 Häufigkeitstabellen für gruppierte Daten	48
3.2 Empirische Verteilungsfunktion und Quantile	50
3.2.1 Bestimmung von Quantilwerten bei ungruppierten Daten	52
3.2.2 Bestimmung von Quantilwerten bei gruppierten Daten	54
3.3 Graphische Darstellung von Verteilungen	56
3.3.1 Visualisierung von Häufigkeitsverteilungen metrischer Variablen	56
3.3.2 Visualisierung bei normalskalierten Variablen	61
3.4 Zusammenfassung	63
4 Kennwerte univariater Verteilungen	67
4.1 Typische Werte oder Lagemaße	67
4.1.1 Modus	68
4.1.2 Median	68
4.1.3 Arithmetisches Mittel	72

4.1.4	Geometrisches Mittel	77
4.1.5	Welches Lagemaß soll eingesetzt werden?	79
4.2	Streuungsmaße	83
4.2.1	Spannweite, Quartilabstand und durchschnittliche Abweichung	84
4.2.2	Variation, Varianz, Standardabweichung und Variationskoeffizient	88
4.2.3	Devianz: ein Streuungsmaß für nominalskalierte Variablen	96
4.3	Weitere Verteilungskenngrößen: Schiefe und Steilheit	101
4.3.1	Schiefe von Verteilungen	101
4.3.2	Steilheit einer Verteilung	103
4.4	Zusammenfassung	103
B	Population und Stichprobe	107
5	Wahrscheinlichkeitstheorie: die Modellierung unsicherer Ereignisse	107
5.1	Zufallsexperimente und Wahrscheinlichkeiten	109
5.1.1	Zufallsexperiment und Ereignisraum	109
5.1.2	Disjunkte und komplementäre Ereignisse	111
5.1.3	Wahrscheinlichkeiten und die Axiome der Wahrscheinlichkeitstheorie	115
5.2	Wahrscheinlichkeiten komplexer Ereignisse	117
5.2.1	Wahrscheinlichkeit der Vereinigung von Ereignissen: das Additionstheorem	118
5.2.2	Bedingte Wahrscheinlichkeiten	120
5.2.3	Wahrscheinlichkeit der Überschneidung von Ereignissen: das Multiplikationstheorem	122
5.2.4	Statistische Unabhängigkeit	126
5.3	Statistisches Modell und Realität	128
5.3.1	Wahrscheinlichkeiten, relative Häufigkeiten und das Gesetz der großen Zahl	128
5.3.2	Wahrscheinlichkeiten als Grenzwerte relativer Häufigkeiten	133
5.4	Stichprobenziehung als Zufallsexperiment	135
5.4.1	Einfache Zufallsauswahl ohne Zurücklegen	137
5.4.2	Einfache Zufallsauswahl mit Zurücklegen	144
5.5	Zufallsvariablen	146
5.5.1	Zufallsvariable und Wahrscheinlichkeitsverteilungen	146
5.5.2	Erwartungswert und Varianz von Zufallsvariablen	150

5.6	Bedeutung von Kennwerteverteilungen bei der Generalisierung von Stichprobenergebnissen	152
5.7	Zusammenfassung	154
6	Hypergeometrische Verteilung und Binomialverteilung: Kennwerteverteilungen von Häufigkeiten und Anteilen	159
6.1	Wahrscheinlichkeitsverteilung der Häufigkeit eines binären Merkmals bei einfacher Zufallsauswahl ohne Zurücklegen: die hypergeometrische Verteilung	162
6.2	Wahrscheinlichkeitsverteilung der Häufigkeit eines binären Merkmals bei einfacher Zufallsauswahl mit Zurücklegen: die Binomialverteilung	167
6.3	Wahrscheinlichkeitsverteilung des Anteils eines binären Merkmals	169
6.4	Beziehungen zwischen hypergeometrischer Verteilung und Binomialverteilung	175
6.5	Einfache Zufallsauswahlen und reale Stichprobenpläne	177
6.6	Zusammenfassung	179
7	Die Normalverteilung und verwandte Wahrscheinlichkeitsverteilungen: Kennwerteverteilungen von Mittelwerten und Varianzen	182
7.1	Wahrscheinlichkeitsdichten kontinuierlicher Zufallsvariablen	182
7.2	Normalverteilungen	189
7.2.1	Die Gauß'sche Normalverteilung	189
7.2.2	Standardnormalverteilung	191
7.3	Asymptotische Normalverteilungen nach dem zentralen Grenzwertsatz	195
7.3.1	Der zentrale Grenzwertsatz	196
7.3.2	Erwartungswerte und Varianzen von Stichprobenmittelwerten	198
7.3.3	Anwendung des zentralen Grenzwertsatzes auf Mittelwerte	201
7.3.4	Anwendung des zentralen Grenzwertsatzes auf Häufigkeiten und Anteile	204
7.4	Verwandte der Normalverteilung: χ^2 -, T- und F-Verteilung	209
7.4.1	Kennwerteverteilung der Variation bei normalverteilten Merkmalen	209
7.4.2	Kennwerteverteilung von Mittelwerten bei normalverteilten Merkmalen	215

7.4.3	Kennwerteverteilungen von Varianzverhältnissen	219
7.5	Zusammenfassung	222
C	Schätzen und Testen	225
8	Schätzen von Populationswerten	225
8.1	Zufallsauswahlen und repräsentative Stichproben	225
8.2	Stichprobenkennwerte als Schätzer für Populationsparameter und die Eigenschaften von Schätzern	227
8.2.1	Erwartungstreue	228
8.2.2	Effizienz	229
8.2.3	Konsistenz	230
8.3	Punktschätzung	232
8.3.1	Schätzung von Populationsanteilen	232
8.3.2	Schätzung von Populationsmittelwerten	234
8.3.3	Schätzung von Varianzen	235
8.4	Intervallschätzung	237
8.4.1	Logik von Konfidenzintervallen	238
8.4.2	Interpretation von Konfidenzintervallen	242
8.4.3	Berechnung von Konfidenzintervallen	244
8.4.4	Berechnung der benötigten Fallzahl	249
8.5	Zusammenfassung	250
9	Prinzipien der statistischen Hypothesenprüfung	253
9.1	Die Logik der statistischen Hypothesenprüfung	253
9.1.1	Kennwerteverteilung bei Gültigkeit der Null- und der Alternativhypothese	253
9.1.2	Fehlermöglichkeiten bei der statistischen Hypothesenprüfung	256
9.1.3	Inhaltliche Relevanz der Null- und der Alternativhypothese	261
9.2	Die Vorgehensweise bei der Prüfung statistischer Hypothesen	264
9.3	Beziehung zwischen statistischen Hypothesentests und der Berechnung von Konfidenzintervallen	272
9.4	Zusammenfassung	275
10	Anwendung der statistischen Hypothesenprüfung	277
10.1	Prüfung von Hypothesen über einen Populationswert: Anteile und Mittelwerte	277
10.1.1	Hypothesen über einen Populationsanteil	277
10.1.2	Hypothesen über einen Populationsmittelwert	280

10.2	Hypothesen über zwei Populationsmittelwerte	283
10.2.1	Hypothesen über die Differenz von Mittelwerten bei unabhängigen Stichproben	285
10.2.2	Hypothesen über die Differenz von Mittelwerten bei abhängigen Stichproben	291
10.3	Hypothesen über zwei Populationsanteile	296
10.3.1	Hypothesen über die Differenz von Anteilen bei unabhängigen Stichproben	296
10.3.2	Hypothesen über Anteilsdifferenzen bei abhängigen Stichproben	300
10.4	Zusammenfassung	304
D Zusammenhang zwischen zwei Variablen		307
11	Bivariater Zusammenhang in der Vierfeldertabelle	307
11.1	Die bivariate Häufigkeitsverteilung in einer Kreuztabelle	307
11.2	Bedingte und unbedingte Häufigkeitsverteilungen	312
11.3	Zusammenhangsmaße in der Vierfeldertabelle	317
11.3.1	Asymmetrischer Zusammenhang: die Prozent- satzdifferenz	317
11.3.2	Konfidenzintervalle und Hypothesenprüfung bei Prozentsatzdifferenzen	323
11.4	Prüfung der Unabhängigkeit in der Vierfeldertabelle	329
11.4.1	Ermittlung der bei Unabhängigkeit erwarteten Häufigkeiten	331
11.4.2	Pearsons Teststatistik Chiquadrat χ^2	332
11.5	Symmetrische Zusammenhangsmaße in der Vierfeldertabelle	335
11.6	Zusammenfassung	339
12	Bivariater Zusammenhang in der Mehrfeldertabelle	343
12.1	Bivariate Häufigkeitsverteilungen in Mehrfeldertabellen	343
12.2	Prüfung der Unabhängigkeit in Mehrfeldertabellen	346
12.3	Exkurs: Pearsons Chiquadrattest als Anpassungstest	351
12.4	Zusammenhangsmaße für Mehrfeldertabellen bei nominalskalierten Variablen	354
12.4.1	Symmetrische Zusammenhangsmaße für Mehr- feldertabellen: nominalskalierte Variablen	355
12.4.2	Asymmetrische Zusammenhangsmaße für Mehr- feldertabellen: nominalskalierte Variablen	357

12.5	Zusammenhangsmaße für Mehrfeldertabellen bei ordinalskalierten Variablen	366
12.5.1	Logik der Paarvergleiche	368
12.5.2	Symmetrische Zusammenhangsmaße für Mehr- feldertabellen: ordinalskalierte Variablen	373
12.5.3	Asymmetrisches Zusammenhangsmaß für Mehr- feldertabellen: ordinalskalierte Variablen	375
12.5.4	Beziehungen zwischen den Maßen für ordinal- skalierte Variablen	376
12.6	Zusammenfassung	378
13	Bivariater Zusammenhang bei metrischen Variablen: Grundlagen von Regression und Korrelation	382
13.1	Das Regressionsmodell in der Stichprobe	388
13.1.1	Die Regressionsgleichung in der Stichprobe	388
13.1.2	Methode der kleinsten Quadrate	389
13.1.3	Berechnung der Regressionskoeffizienten in der Stichprobe	389
13.1.4	Vorhersagewerte: die Regressionsgerade in der Stichprobe	391
13.2	Interpretation der Regressionsgeraden	394
13.2.1	Extrapolation für nicht beobachtete Werte	394
13.2.2	Die Bedeutung der Regressionskoeffizienten	394
13.2.3	Prognose und Kausalerklärung	397
13.2.4	Regressionsresiduen	398
13.3	Kovarianz und Korrelation	398
13.3.1	Kovarianz	398
13.3.2	Korrelation	401
13.4	Stichprobeneigenschaften von Vorhersagewerten und Residuen	407
13.4.1	Eigenschaften von Vorhersagewerten	407
13.4.2	Eigenschaften der Residuen	408
13.4.3	Varianzzerlegung und Determinations- koeffizient	410
13.4.4	Die PRE-Logik in der linearen Regression	414
13.4.5	Standardisierte Regressionskoeffizienten	417
13.5	Zusammenfassung	419
14	Bivariater Zusammenhang bei metrischen Variablen: Schätzen und Testen von Populationsparametern im linearen Regressions- modell	424

14.1	Generalisierung der aus Stichprobendaten berechneten Regressionsgeraden	425
14.1.1	Voraussetzungen für die Generalisierung der Regressionsgeraden	425
14.1.2	Standardfehler der Regressionskoeffizienten	428
14.1.3	Konfidenzintervalle für die Regressionskoeffizienten α und β	431
14.1.4	Prüfung von Hypothesen über die Regressionskoeffizienten α und β	433
14.1.5	Standardfehler und Konfidenzintervall für den Vorhersagewert von Y	435
14.1.6	Prüfung von Hypothesen über Kovarianzen, Korrelationen und Determinationskoeffizienten	441
14.2	Kontrolle der Modellannahmen des linearen Regressionsmodells	443
14.2.1	Residuenanalyse zur Beurteilung von Linearität und Homoskedastizität	445
14.2.2	Prüfung der Normalverteilungsannahme	448
14.2.3	Ausreißerwerte und einflussreiche Fälle	449
14.3	T-Test von Mittelwertdifferenzen in unabhängigen Stichproben	451
14.4	Zusammenfassung	459
E	Drittvariablenkontrolle	463
15	Drittvariablenkontrolle in Kreuztabellen	463
15.1	Aufbau einer trivariaten Kreuztabelle	463
15.2	Beziehungsmuster bei Drittvariablenkontrolle	471
15.2.1	Additiver Effekt bei unkorrelierten erklärenden Variablen	472
15.2.2	Scheinkausalität	475
15.2.3	Drittvariable Z als intervenierende Variable	476
15.2.4	Konfundierung	480
15.2.5	Interaktion	483
15.3	Korrelation und Kausalität	487
15.4	Chiquadrattests in der trivariaten Tabellenanalyse	489
15.5	Zusammenfassung	500
16	Drittvariablenkontrolle in der linearen Regression: Trivariate Regression	503
16.1	Trivariates Regressionsmodell in der Stichprobe	506

16.1.1	Berechnung der Regressionskoeffizienten in der Stichprobe	507
16.1.2	Interpretation der Regressionskoeffizienten	511
16.1.3	Partielle Regressionsgewichte und Residuenregression	515
16.1.4	Vorteile der Drittvariablenkontrolle in der linearen Regression	518
16.1.5	Anwendungsbeispiel für die trivariate Regression	520
16.2	Standardisierte Effekte und partielle Korrelation	522
16.2.1	Lineare Regression zwischen standardisierten Variablen	523
16.2.2	Berechnung der standardisierten Regressionsgewichte	525
16.2.3	Unterschiede zwischen bivariaten und partiellen Regressionsgewichten	527
16.2.4	Standardisiertes Regressionsgewicht und partielle Korrelation	529
16.3	Stichprobeneigenschaften von Vorhersagewerten und Residuen	531
16.3.1	Eigenschaften der Vorhersagewerte	531
16.3.2	Eigenschaften der Residuen	531
16.3.3	Varianzzerlegung und Determinationskoeffizient ..	532
16.4	Zusammenfassung	534
17	Von der trivariaten zur multiplen Regression	538
17.1	Schätzen und Testen im trivariaten Regressionsmodell ...	538
17.1.1	Annahmen im trivariaten Regressionsmodell	538
17.1.2	Schätzung der Residualvarianz in der Population	539
17.1.3	Standardfehler der Regressionsgewichte in der trivariaten Regression	540
17.1.4	Konfidenzintervalle für die Regressionsgewichte ..	541
17.1.5	Prüfung von Hypothesen über die Regressionsgewichte	543
17.1.6	Multikollinearität	545
17.1.7	Prüfung von Hypothesen über den Determinationskoeffizienten R^2	548
17.2	Erweiterung der trivariaten Regression zur multiplen Regression	551
17.2.1	Varianzzerlegung und F-Test	552

17.2.2 Standardisierte und unstandardisierte Regressionskoeffizienten	556
17.3 Interaktionen und nichtlineare Beziehungen im multiplen Regressionsmodell	558
17.3.1 Spezifikation von Interaktionseffekten durch Produktvariablen	559
17.3.2 Prüfung der Bedeutsamkeit zusätzlicher Prädiktoren im multiplen Regressionsmodell	562
17.3.3 Spezifikation nichtlinearer Beziehungen im linearen Regressionsmodell	564
17.3.4 Interaktionseffekte im Regressionsmodell mit nichtlinearer Beziehung	567
17.4 Nominalskalierte Prädiktoren im Regressionsmodell	571
17.4.1 Spezifikation von Indikatorvariablen	571
17.5 Zusammenfassung	578
F Ausblick	583
18 Multivariate Analysemodelle im Überblick	583
18.1 Lineare Modelle	583
18.1.1 Multivariate Regression	583
18.1.2 Hauptkomponentenanalyse	585
18.1.3 Explorative Faktorenanalyse	591
18.1.4 Konfirmatorische Faktorenanalyse	596
18.1.5 Lineare Strukturgleichungsmodelle	601
18.2 Nichtlineare Regressionsmodelle	605
18.2.1 Probitregression und logistische Regression	606
18.2.2 Log-lineare Tabellenanalyse	609
18.3 Weitere Analyseverfahren	611
18.4 Literaturhinweise	613
18.5 Zusammenfassung	614
 19 Literatur	 616
 Anhang	 621
A-1 Rechnen mit Summen, Potenzen und Logarithmen	621
A-1.1 Das Summenzeichen	621
A-1.2 Potenzen und Logarithmen	623
A-2 Mittelwerte, Standardabweichungen und Varianzen linearer Funktionen von Variablen	627
A-2.1 Lineartransformationen einer Variablen	627

A-2.2	Zusammenfassen der Verteilungen von Subgruppen	632
A-2.3	Linearkombinationen aus voneinander unabhängigen Variablen	635
A-2.4	Linearkombinationen aus voneinander abhängigen Variablen	638
A-3	Tabellen	642
	Sachregister	649