

Inhaltsverzeichnis

Einleitung

19

Zu diesem Buch	19
Konventionen in diesem Buch	20
Was Sie nicht lesen müssen	21
Törichte Annahmen über den Leser	21
Wie dieses Buch aufgebaut ist	21
Teil I: Datenanalyse und Grundlagen der Modellbildung	22
Teil II: Vorhersagen unter Verwendung der Regression treffen	22
Teil III: Vergleich vieler Mittelwerte mit der ANOVA	22
Teil IV: Starke Verbindungen mit Chi-Quadrat-Tests aufbauen	22
Teil V: Rebellen ohne Verteilung: Nichtparametrische Statistik	22
Teil VI: Der Top-Ten-Teil	23
Die Symbole in diesem Buch	23
Wie es weitergeht	24

Teil I

Datenanalyse und Modellbildung – Grundlagen

25

Kapitel 1

Mehr als nur Zahlenverarbeitung: Datenanalyse als Kunst und Wissenschaft

27

Datenanalyse: Nicht mehr nur für Statistiker	27
Die gute alte Zeit	28
Der Nachteil der heutigen Statistik-Software	29
Regel Nr. 1: Informieren Sie sich VOR der Verarbeitung!	29
Nichts ist ewig (nicht einmal eine Gerade)	30
Datenschnüffeln ist nicht cool!	31
(Daten-)Fischen verboten!	32
Das große Ganze: Ein Überblick über weiterführende Statistik	36
Populationsparameter	37
Stichprobenkenngroße	37
Vertrauensintervall	38
Hypothesentest	38
Varianzanalyse (ANOVA, Analysis of Variance)	39
Multiple Vergleiche	40
Interaktionseffekte	41
Korrelation	43
Lineare Regression	44
Chi-Quadrat-Tests	45
Nichtparametrische Statistik	47

Kapitel 2	
Orientierung innerhalb der statistischen Techniken	49
Qualitative und quantitative Variablen in der statistischen Analyse	49
Statistiken für qualitative Variablen	51
Anteile vergleichen	51
Einen Anteil abschätzen	51
Nach Beziehungen zwischen qualitativen Variablen suchen	52
Modelle für Vorhersagen erstellen	53
Statistik für quantitative Variablen	55
Vergleiche anstellen	55
Verbindungen erkennen	55
Vorhersagen treffen	57
Verzerrung vermeiden	58
Betrachtung der Verzerrung durch die Brille der Statistik	58
Der Varianzstreichfall: der Kampf zwischen $n - 1$ und n	59
Höchste Genauigkeit erzielen	61
Genauigkeit aus statistischer Perspektive verstehen	61
Genauigkeit mit dem Fehlerspielraum messen	61
Schlussfolgerungen treffen und Grenzen erkennen	65
Kapitel 3	
Vertrauen aufbauen und Modelle testen	67
Parameter anhand von Vertrauensintervallen schätzen	68
Die Grundlagen: Die allgemeine Form eines Vertrauensintervalls	68
Das Vertrauensintervall für einen Populationsmittelwert finden	69
Was ändert den Fehlerspielraum?	70
Ein Vertrauensintervall interpretieren	73
Modelle aufstellen und testen	74
Was sind H_0 und H_a wirklich?	75
Ihre Hinweise in einer Teststatistik zusammenfassen	75
Die Stärke des Beweises mit einem p -Wert bestimmen	76
Typ-I- und Typ-II-Fehler ausschließen	77
Die Güte eines Hypothesentests	79
Teil II	
Mit Hilfe der Regression Vorhersagen treffen	85
Kapitel 4	
Einfache lineare Regression verstehen	87
Mit Streudiagrammen und Korrelationen Beziehungen untersuchen	87
Mit Hilfe von Streudiagrammen Beziehungen untersuchen	89
Informationen mit Hilfe des Korrelationskoeffizienten zuordnen	90

Ein einfaches lineares Regressionsmodell erstellen	91
Die beste Gerade für die Modellierung Ihrer Daten	91
Der y -Schnittpunkt der Regressionsgeraden	92
Die Steigung der Regressionsgeraden	93
Schätzungen anhand der Regressionsgeraden	94
Prüfen, ob das Modell passt (Daten, nicht Kleider!)	94
Definition der Bedingungen	94
Die Residuen finden und untersuchen	96
Mit r^2 messen, ob das Modell geeignet ist	99
Ausreißer	100
Korrekte Schlüsse ziehen	101
Vermeiden Sie, Ursache- und Wirkungsdiskussionen	102
Extrapolation: No-No!	102
Die Grenzen eines einfachen linearen Regressionsmodells kennen	103

Kapitel 5**Wenn zwei Variablen besser sind als eine: Multiple Regression****105**

Das multiple Regressionsmodell	105
Die Verwendungszwecke der multiplen Regression erkennen	105
Die allgemeine Form des multiplen Regressionsmodells	106
Die Analyseschritte	106
Alle x und y betrachten	107
Daten sammeln	108
Mögliche Beziehungen erkennen	109
Streudiagramme erstellen	109
Korrelationen: Untersuchung der Verbindung	110
Auf Multikollinearität prüfen	113
Das am besten angepasste Modell finden	114
Die Koeffizienten des multiplen Regressionsmodells bestimmen	114
Die Koeffizienten interpretieren	115
Koeffizienten testen	116
Vorhersage von y anhand der x -Variablen	118
Prüfen, wie gut das Modell angepasst ist	119
Die Voraussetzungen	119
Ein Plan für die Prüfung der Voraussetzungen	119

Kapitel 6**Ein Schritt vor und zwei zurück: Auswahl des Regressionsmodells****125**

Abstoßdistanzen schätzen – Der ultimative Kick	125
Variablen-Brainstorming und -Datensammlung	126
Streudiagramme und Korrelationen untersuchen	127
Die Vorwärtsselektion	130
Variablen hinzufügen – eine nach der anderen	131
Wie gut passt das Modell?	132
Die Vorwärtsselektion für die Abstoßdistanzen	133

Und jetzt das Ganze von hinten: Die Rückwärtsselektion	136
Variablen nacheinander eliminieren	136
Passt das Modell?	138
Variablen ausschließen, um die Abstoßdistanz zu bestimmen	138
Das Verfahren der besten Teilmengen	140
Alle Modelle erzeugen und dann das Beste davon auswählen	140
Die besten Teilmengen für das Beispiel der Abstoßdistanz	141
Vergleich der Modellauswahlverfahren	143
Warum erzeugen nicht alle Verfahren dieselben Ergebnisse?	143
Welche Vor- und Nachteile haben die einzelnen Verfahrensweisen?	144
Kapitel 7	
Mit Daten in die Kurve gehen: Nichtlineare Regression	147
Am Anfang war das Streudiagramm	147
Polynome für Kurven nutzen	149
Zurück zu den Polynomen	149
Suche nach dem besten Polynomialmodell	151
Die Prüfung mit einem Polynom zweiten Grades bestehen	152
Bewertung der Anpassung eines Polynomialmodells	155
Vorhersagen treffen	158
Nach oben? Nach unten? Exponentiell!	159
Zurück zu den exponentiellen Modellen	160
Die Suche nach dem besten exponentiellen Modell	160
Geheimnisse in exponentieller Geschwindigkeit verbreiten	162
Kapitel 8	
Ja, Nein, Vielleicht: Vorhersagen mit logistischer Regression	167
Aufstellung des logistischen Regressionsmodells	167
Definition eines logistischen Regressionsmodells	167
Mit einer S-Kurve Wahrscheinlichkeiten abschätzen	168
Interpretation der Koeffizienten des logistischen Regressionsmodells	168
Abschätzen der Wahrscheinlichkeit, dass ein Film ein Kassenschlager wird – mit der logistischen Regression	169
Allgemeine Schritte für die logistische Regression	170
Die Analyse in Minitab	171
Die Koeffizienten finden und das Modell erstellen	172
Schätzung von p	173
Die Anpassung des Modells überprüfen	174

Teil III	
Viele Mittelwerte vergleichen – mit der Varianzanalyse	177
Kapitel 9	
Einfache Varianzanalyse	179
Zwei Mittelwerte mit einem t-Test vergleichen	179
Mehr Mittelwerte mit einer ANOVA vergleichen	181
Kerne spucken: Eine Situation, die nach einer ANOVA schreit	181
Die ANOVA-Schritte	182
Die Bedingungen prüfen	183
Unabhängigkeit prüfen	183
Was ist schon normal?	183
Die Streuung	185
Die Hypothesen aufstellen	185
Der F-Test	186
Die ANOVA in Minitab	186
Die Varianz in Quadratsummen zerlegen	187
Die mittleren Quadratsummen suchen	188
Die F-Statistik bestimmen	189
Schlüsse aus der ANOVA ziehen	190
Was kommt als Nächstes?	192
Die Anpassung des ANOVA-Modells überprüfen	192
Kapitel 10	
Mit multiplen Vergleichen Paare finden	195
Nach der ANOVA	195
Mit Hilfe von Fisher und Tukey unterschiedliche Mittelwerte dingfest machen	197
Mit Fisher's LSD Unterschiede erkennen	198
Mit dem Tukey-Test vergleichen	200
Kapitel 11	
Weiter mit der zweifachen ANOVA!	203
Das Modell für die zweifache ANOVA aufstellen	203
Die Behandlungen festlegen	204
Die Quadratsummen	204
Interaktionseffekte verstehen	205
Was ist eine Interaktion überhaupt?	206
Interaktion und Interaktionsdiagramme	206
Die Terme in der zweifachen ANOVA testen	209
Die zweifache ANOVA-Tafel erstellen	210
Ergebnisse interpretieren: Zahlen und Grafiken	210

Kapitel 12	
Grenzenlos: Wie Regression und ANOVA zusammenhängen	215
Regression in den Augen der Variation	215
Variabilität in den y feststellen und nach x suchen, die sie erklären	215
Die Variabilität in der Internet-Benutzung erkennen	216
Eine Erklärung für die Internet-Nutzung suchen	216
Ergebnisse durch Regression	217
Bewertung der Anpassung des Regressionsmodells	219
Regression und ANOVA: Treffen der Modelle	220
Quadratsummen vergleichen	220
Regression und die ANOVA-Tafel	223
F- und t-Statistik: Die letzte Hürde	224
Teil IV	
Starke Verbindungen mit Chi-Quadrat-Tests herstellen	227
Kapitel 13	
Mit Kreuztabellen Assoziationen bilden	229
Eine Kreuztabelle erstellen	230
Daten in einer Kreuztabelle anordnen	230
Die Zellenwerte eintragen	231
Randsummen erstellen	231
Die Wahrscheinlichkeiten zerlegen	232
Randwahrscheinlichkeiten	233
Gemeinsame Wahrscheinlichkeiten	234
Bedingte Wahrscheinlichkeiten	235
Unabhängigkeit anstreben	240
Auf Unabhängigkeit zwischen zwei Kategorien prüfen	241
Auf Unabhängigkeit zwischen zwei Variablen prüfen	242
Das Simpson-Paradoxon – keine Zauberei	243
Das Simpson-Paradoxon untersuchen	243
Die Frage nach dem Warum: Simpson-Paradoxon	246
Achten Sie immer auf das Simpson-Paradoxon!	247
Kapitel 14	
Ausreichend unabhängig für den Chi-Quadrat-Test sein	249
Ein Hypothesentest auf Unabhängigkeit	250
Daten sammeln und anordnen	251
Die Hypothesen festlegen	252
Bestimmung der erwarteten Zellenwerte	252
Die Bedingungen für den Test prüfen	255
Die Chi-Quadrat-Teststatistik berechnen	255

Ihre Ergebnisse in der Chi-Quadrat-Tabelle nachschlagen	258
Ihre Schlüsse	261
Zwei Tests für den Vergleich von zwei Anteilen vergleichen	263
Zurück zum Z-Test für zwei Populationsanteile	264
Chi-Quadrat-Tests und Z-Tests für eine 2×2 -Tabelle gleichsetzen	265
 Kapitel 15	
Der Chi-Quadrat-Test auf Güte der Anpassung	269
Die Teststatistik für die Güte der Anpassung bestimmen	269
Beobachtetes mit Erwartetem vergleichen	270
Die Statistik für die Güte der Anpassung berechnen	272
Interpretation der Statistik für die Güte der Anpassung mit Hilfe von Chi-Quadrat	274
Als Erstes werden die Bedingungen geprüft	275
Die Schritte für den Chi-Quadrat-Test auf Güte der Anpassung	276
 Teil V	
Rebellen ohne Verteilung	279
 Kapitel 16	
Es wird nichtparametrisch!	281
Argumente für die nichtparametrische Statistik	281
Kein Problem, wenn Bedingungen nicht erfüllt sind	281
Der Median	282
Wo also liegt der Haken?	284
Die Grundlagen der nichtparametrischen Statistik	285
Vorzeichen (Signum)	285
Rang	287
Vorzeichen-Rang	288
Rangsumme	289
 Kapitel 17	
Der Vorzeichentest und der Vorzeichen-Rangtest	291
Die Vorzeichen erkennen: Der Vorzeichentest	291
Den Median testen	293
Den Median schätzen	295
Tests für verbundene Stichproben	297
Noch einen Schritt weiter: Mit dem Vorzeichen-Rangtest	299
Eine Einschränkung des Vorzeichentests	299
Die Schritte für einen Vorzeichen-Rangtest	299
Gewichtsverlust mit Vorzeichen-Rängen	301

Kapitel 18	
Der Rangsummentest	305
Den Rangsummentest durchführen	305
Die Bedingungen prüfen	306
Der schrittweise Test	306
Die Stichprobengröße heraussetzen	308
Einen Rangsummentest durchführen: Welcher Immobilienmakler verkauft die Häuser schneller?	309
Die Bedingungen für diesen Test überprüfen	310
Die Hypothesen testen	311
Kapitel 19	
Kruskal-Wallis und Wilcox	315
Mit dem Kruskal-Wallis-Test mehr als zwei Populationen vergleichen	315
Die Bedingungen prüfen	316
Aufstellung des Testproblems	318
Den Test schrittweise durchführen	319
Die Differenzen erkennen: Der Wilcoxon-Rangsummentest	322
Paarweise Vergleiche	322
Vergleichstest stellen fest, wer anders ist	323
Die Mediane untersuchen, um festzustellen, wie sie sich unterscheiden	324
Kapitel 20	
Korrelationen mit dem Spearman'schen Rang bestimmen	327
Pearson und seine aufwändigen Bedingungen	327
Bewertungen mit Hilfe der Spearman-Rangkorrelation	329
Den Spearman-Rangkorrelationskoeffizienten bestimmen	329
Spearman in der Praxis: Talent und Leistung in eine Beziehung bringen	331
Teil VI	
Der Top-Ten-Teil	335
Kapitel 21	
Zehn Fehler in statistischen Schlüssen	337
Diese Statistiken beweisen ...	337
Es ist nicht technisch statistisch signifikant, aber ...	337
Das bedeutet, x verursacht y	338
Ich nahm an, die Daten seien normalverteilt ...	339
Ich berichte nur über »wichtige« Ergebnisse	340
Eine größere Stichprobe ist immer besser	340
Es ist nicht technisch zufällig, aber ...	341

1.000 Antworten sind 1.000 Antworten	342
Natürlich gelten diese Ergebnisse für die gesamte Population!	343
Ich habe beschlossen, das wegzulassen!	344
Kapitel 22	
Zehn Probleme aus der Praxis	347
Mittelwerte mit der einfachen ANOVA vergleichen	347
Multiple Vergleiche	348
Mit der zweifachen ANOVA zwei Faktoren betrachten	349
Vorhersage einer quantitativen Variablen mit Hilfe der Regression	350
Eine Wahrscheinlichkeit mit der logistischen Regression vorhersagen	351
Nichtlineare Regression für gekrümmte Daten	353
Mit Chi-Quadrat auf Unabhängigkeit testen	354
Spezielle Modelle mit dem Test auf Güte der Anpassung testen	355
Den Median mit dem Rangsummentest schätzen	355
Die Modellanpassung mit R^2 überprüfen	356
Appendix	
Tabellen zum Nachschlagen	359
<i>t</i> -Tabelle	359
Binomialtabelle	361
Chi-Quadrat-Tabelle	364
Rangsummentabelle	366
<i>F</i> -Tabelle	367
Stichwortverzeichnis	369