

Inhaltsverzeichnis

1	Einleitung	13
1.1	Was ist und was soll Datenanalyse?	13
1.2	Exploratorische und konfirmatorische Datenanalyse	22
1.3	Grundprinzipien der exploratorischen Datenanalyse	24
1.4	Aufbau des Buches	26
1.5	Benutzung von PC-ISP	27
1.5.1	Installation PC-ISP/DGS Personal	27
1.5.2	On-line Hilfe	29
1.5.3	Syntax, Workspace und I/O	29
2	Exploratorische Datenanalyse	33
2.1	Urliste, empirische Verteilungsfunktion und Histogramm	33
2.2	„Aktive“ Histogramme und Stamm-und-Blatt-Darstellungen	41
2.3	Kennzahlen einer eindimensionalen stetigen Variablen	44
2.3.1	Maßzahlen der Lage	45
2.3.2	Spezielle Ordnungsstatistiken	48
2.3.3	Modalwert	53
2.3.4	Streuungsmaße	54
2.3.5	Verteilungsformen und Formmaße	59
2.3.6	Ausreißer	66
2.3.7	Box-Plots	68
2.4	Streudiagramme	70
3	Vergleich der Verteilungsform zweier Datenmengen	75
3.1	Stamm-und-Blatt-Darstellung, Histogramm und Box-Plot	75
3.2	Empirische Q-Q-Plots	80
3.3	Theoretische Q-Q-Plots	84
3.4	P-P-Plots	85
4	Allgemeines univariates lineares Modell	87

5	Varianzanalyse und additive Modelle	91
5.1	Einfaktorielle Varianzanalyse	91
5.1.1	Einfache Varianzanalyse für Modell I	93
5.1.2	Einfache Varianzanalyse für Modell II	95
5.2	Zweifaktorielle Varianzanalyse	99
5.2.1	Vollständige Kreuzklassifikation	101
5.2.2	Hierarchische Klassifikation für Modell II	105
5.3	Kovarianzanalyse	107
5.3.1	Zweifache Klassifikation mit einer Kovariablen und einfacher Klassenbesetzung	109
5.3.2	Zweifache Klassifikation mit einer Kovariablen und mehrfacher Klassenbesetzung	111
5.4	Verteilungsfreie Varianzanalysen	112
5.4.1	Abhängige Stichproben — Friedmann-Test	112
5.4.2	Unabhängige Stichproben — Kruskal-Wallis-Test	113
5.5	Medianglättung	114
5.6	Medianglättung von Drei-Wege-Tafeln	117
6	Regression und Korrelation	119
6.1	Einfache lineare Regression	119
6.2	Regressionsdiagnostik	128
6.3	Residuenanalyse	131
6.4	Alternative Methoden für kontinuierliche Einflußgrößen	134
6.4.1	Resistente Geraden	134
6.4.2	Robuste Glättungskurven	136
6.5	Multiple lineare Regression	140
6.5.1	Modell I der multiplen linearen Regression	142
6.5.2	Modell II der multiplen linearen Regression	144
6.5.3	Modelle mit Fehlern in den Variablen	144
6.5.4	Modelle mit stochastischen Parametern	144
6.6	Modellvergleich und Merkmalsselektion	145
6.7	Ridge-Regression	150
6.8	Korrelationsanalyse	155
6.8.1	Multiple und partielle Korrelationskoeffizienten	156
6.8.2	Nichtparametrische Korrelationskoeffizienten	157

7 Transformationen	163
7.1 Eigenschaften von Potenztransformationen	163
7.2 Symmetrietransformationen	168
7.3 Transformationen zur stabilen Streuung in mehreren Gruppen	172
7.4 Linearitätstransformationen für X–Y–Daten	175
7.5 Transformationen in Zwei-Wege-Tafeln	178
7.6 Möglichkeiten und Grenzen von Transformationen	181
8 Kontingenztafelanalyse	185
8.1 Vier–Felder–Tafeln	185
8.2 Allgemeine Kontingenztafeln	189
8.2.1 Untersuchungsmodelle und Hypothesen	190
8.2.2 Statistische Analysen von zweidimensionalen Kontingenztafeln	192
8.3 Kodierte Tafeln	195
9 Verallgemeinerte lineare Modelle	197
9.1 Lineare Modelle	197
9.2 Verallgemeinerungsmöglichkeiten des Modells	197
9.3 Verallgemeinerte lineare Modelle	198
9.4 GLIM–Implementierung in ISP	199
9.5 Modellbildung und Beurteilung der Modellgüte	201
10 Exploratorische Verfahren der multivariaten Analyse	203
10.1 Dynamische Grafik	203
10.2 Beispiel 1: Schadstoffbelastung von Dorschen	205
10.3 Beispiel 2: Daten aus der Gesundheitsforschung	209
10.4 Beispiel 3: Belastung von Böden mit Pestiziden	213
10.5 Beispiel 4: Analysesystem für Straßendaten	215
10.6 Drei weitere Beispiele zur dynamischen Grafik	215
11 Allgemeines multivariates lineares Modell	225
11.1 Formulierung des multivariaten linearen Modells	225
11.2 Statistische Maßzahlen im multivariaten linearen Modell	226
11.3 Teststatistiken	228
11.4 Prüfen von Hypothesen	229

11.4.1	Hypothese über den Erwartungswertvektor	229
11.4.2	Ein Konfidenzbereich für den Erwartungswertvektor	231
11.4.3	Gleichheit der Erwartungswertvektoren bei abhängigen Stichproben	232
11.4.4	Hypothese der Gleichheit von Kovarianzmatrizen	234
11.4.5	Gleichheit der Erwartungswertvektoren bei unabhängigen Stichproben	235
11.5	Merkmalsselektion	237
12	Diskriminanzanalyse	241
12.1	Bayes'sches Prinzip	241
12.2	Fehlerraten	244
12.3	Merkmalsselektion	245
12.4	Diskrimination bei normalverteilten Populationen	246
12.4.1	Lineare Diskriminanzanalyse	247
12.4.2	Quadratische Diskriminanzanalyse	256
12.5	Diskrimination mit Kernfunktionen	256
12.6	Diskriminanzanalyse für Zeitreihen	259
12.6.1	Modelle im Frequenzbereich	264
12.6.2	Testen von Hypothesen im Frequenzbereich	265
12.7	Diskrimination mit kategorialen Merkmalen	270
12.7.1	Aktuarisches Modell (AKTUAR)	271
12.7.2	BIRCH-Modell (BIRWW)	271
12.7.3	BIRCH-Modell mit gleicher Wechselwirkungsstruktur (BIRGW) . .	271
12.7.4	Unabhängigkeitsmodell (UNABH)	272
12.7.5	Verfahren zur Modellwahl	273
12.8	Kopplung von Diskriminatoren	276
12.9	Hinzunahme von Verlustfunktionen zur Diskrimination	280
13	Clusteranalyse	283
13.1	Ähnlichkeiten und Distanzen	285
13.2	Algorithmen	291
13.2.1	Hierarchische Verfahren	291
13.2.2	Agglomerative Verfahren	293
13.2.3	Divisive Verfahren	303
13.2.4	K-Mittelwert-Clustering (Quick-Clustering)	304
13.3	Andere Verfahren	304
13.3.1	Faktorenanalyse	304
13.3.2	Konfigurationsfrequenzanalyse	306

14 Faktorenanalyse und Hauptkomponentenanalyse	309
14.1 Das statistische Modell der Faktorenanalyse	309
14.1.1 Prüfen der Anpassungsgüte des Faktorenmodells	312
14.1.2 Rotation der gemeinsamen Faktoren	313
14.1.3 Grafische Darstellungen	314
14.2 Hauptkomponentenanalyse	315
14.3 Hauptfaktorenanalyse	321
A ISP im Überblick	323
A.1 Befehlsreferenz	323
A.2 LET – Funktionen	328
A.3 Makro-Referenz	330
A.4 DGS-Funktionstasten	336
A.5 Dynamische Grafik	337
A.5.1 Der dgs-Befehl	337
A.5.2 Interaktionen in DGS	338
A.5.3 Übersicht des 3D-Modus	340
A.5.4 Die Scatterplot-Matrix	350
B Liste der Beispiele	353
Literaturverzeichnis	355
Sachwörterverzeichnis	369