

Gliederung

Abbildungsverzeichnis.....	X
Tabellenverzeichnis	XX
Abkürzungsverzeichnis.....	XXV
0. Einführung	1
1. Die Ausreißerproblematik.....	4
1.1. Ausreißerobjekte in multivariaten Datensätzen.....	5
1.2. Die Vielfalt des multivariaten Ausreißer-Begriffs	6
1.3. Spezialproblem multiple Ausreißer, Masking und Swamping.....	8
2. Multiple Ausreißeranalyse in der Rohdatenmatrix	12
2.1. Das schrittweise Selektionsverfahren nach SIMONOFF	12
2.1.1. Grundgerüst und Probleme bei der Rückwärtsselektion.....	12
2.1.2. Festlegung des kritischen Wertes zur Identifikation der potentiellen Ausreißerkandidaten.....	14
2.1.3. Spezielle Anwendung des Algorithmus für multivariate Daten	15
2.1.3.1. Eigenschaften des Single-Linkage-Algorithmus für die Ausreißeranalyse	15
2.1.3.2. Die Mahalanobis-Distanz als Teststatistik.....	16
2.1.4. Ablauf des Verfahrens.....	17
2.1.5. Abschließende Bemerkungen.....	18
2.2. Die sequentielle Anwendung der Wilks-Ausreißer-Streuungskoeffizienten.....	21
2.2.1. Die Grundlage des Verfahrens: Die Teststatistik nach WILKS	21
2.2.2. Die Vorgehensweise des sequentiellen Verfahrens	23
2.2.3. Überlegungen zur Festlegung der kritischen Werte	25
2.3. Multiple Ausreißeridentifikation nach HADI.....	28
2.3.1. Die Wahl robuster Startschätzer.....	28

2.3.2. Darstellung des Algorithmus zur Identifikation von multivariaten Ausreißern.....	30
2.3.3. Anmerkungen zum Testverfahren.....	32
2.4. Modifikation der Identifikationsmethode nach HADI	34
2.4.1. Darstellung der Modifikationen	35
2.4.2. Die Vorteilhaftigkeit der Modifikation im Vergleich zur Ursprungsmethode	37
2.5. Eine weitere Modifikation für die multiple Ausreißeranalyse	37
2.5.1. Der MVE-Schätzer.....	38
2.5.2. Integration der MVE-Schätzer in den Algorithmus des Testverfahrens.....	39
2.5.3. Eigenschaften des Verfahrens mit modifizierter Ausgangsbasis.....	39
3. Graphische Verfahren für das Multiple-Ausreißer-Problem	41
3.1. Die Methodik nach BACON-SHONE/FUNG.....	41
3.1.1. Kurzdarstellung der Teststatistik.....	41
3.1.2. Theoretischer Hintergrund des graphischen Verfahrens	42
3.1.2.1. Verteilungstheoretische Probleme der Teststatistik.....	42
3.1.2.2. Approximationsmöglichkeiten für die Nullverteilung der Quantile	44
3.1.3. Abschließende Hinweise zum Verfahren.....	46
3.2. Das Stalaktit-Diagramm nach ATKINSON/MULIRA	47
3.2.1. Grundlagen und Vorgehensweise der Vorwärts-Identifikationsmethode	47
3.2.2. Die Konstruktion des Stalaktit-Diagramms	50
3.2.3. Modifikationen für eine effiziente Ausreißeranalyse.....	51
3.2.3.1. Normalisierung des Stalaktit-Diagramms.....	51
3.2.3.2. Wiederholte Festlegung der Basisstichprobe - Resampling	54
3.2.4. Abschließende Anmerkungen zum Verfahren	55
4. Ausgewählte Verfahren für Ausreißer-Spezialuntersuchungen.....	57
4.1. Ausreißeranalyse in bezug auf Korrelationsstrukturen zwischen den Merkmalsvariablen.....	57

4.1.1. Der RV-Koeffizient nach ESCOUFIER	57
4.1.2. Die Einflußfunktion als Untersuchungsinstrument.....	61
4.1.3. Die Identifikation von Korrelationsausreißern.....	63
4.1.4. Abschließende Hinweise	65
4.2. Die Hauptkomponentenanalyse als Identifikationsinstrument für spezielle Ausreißertypen.....	66
4.2.1. Kurzdarstellung des theoretischen Hintergrundes der Hauptkomponentenanalyse	66
4.2.2. Die Anwendung der Hauptkomponentenanalyse zur Ausreißeridentifikation	67
4.2.3. Interessante Projektionen mehrdimensionaler Daten in der verallgemeinerten Hauptkomponentenanalyse	69
4.2.4. Modifikationen für eine effiziente Ausreißeranalyse.....	71
4.2.4.1. Standardisierung der Hauptkomponenten.....	71
4.2.4.2. Rotation der Hauptkomponenten.....	72
4.2.4.2.1. Die grundlegenden Ansatzpunkte.....	72
4.2.4.2.2. Rotation der gleichmäßigen Komponenten	73
4.2.4.2.3. Abschließende Anmerkungen.....	74
4.3. Das Identifikationsverfahren nach BUTTLER zur Lokalisation der Ausreißerkomponenten des Beobachtungsvektors.....	75
4.3.1. Darstellung der relevanten Statistiken	75
4.3.2. Zusatzinformationen und die Konstruktion des Identifikationstableaus	76
4.3.3. Abschließende Anmerkungen.....	78
5. Praktische Datenanalysen	79
5.1. Identifikation von potentiellen Ausreißerkandidaten der Luftverschmutzungsdaten.....	80
5.2. Identifikation von potentiellen Ausreißerkandidaten der Bilanzkennzahlen.....	102
5.3. Identifikation von potentiellen Ausreißerkandidaten der Kostendaten	119
6. Kritische Würdigung der Testverfahren	138
7. Einflußreiche Beobachtungen in ausgewählten multivariaten Verfahren.....	145

7.1. Einflußreiche Beobachtungen.....	145
7.2. Die Störungstheorie für Eigenwert-Eigenvektor-Probleme	148
7.3. Einflußmessung in der Hauptkomponentenanalyse	150
7.3.1. Grundsätzliche Aspekte bei der Hauptkomponenten-Einflußmessung	151
7.3.2. Die Identifikation von einflußreichen Beobachtungsvektoren über Einflußkurven der Eigenwerte und Eigenvektoren	155
7.3.2.1. Die Herleitung der relevanten Einflußkurven.....	155
7.3.2.2. Stichproben-Versionen	156
7.3.2.3. Abschließende Anmerkungen.....	161
7.3.3. Einflußanalyse des interessierenden Hauptkomponenten- Unterraums	162
7.3.3.1. Ein möglicher Konstruktionsansatz für Einflußfunktionen zur Untersuchung eines interessierenden Unterraumes.....	162
7.3.3.2. Die Stichproben-Versionen der vorgestellten Einflußfunktionen	165
7.3.3.3. Die Identifikation von einflußreichen Beobachtungen.....	167
7.3.3.4. Abschließende Bemerkungen	168
7.3.4. Zusätzliche Berücksichtigung der quadratischen Störungsexpansionen im Einflußfunktionsansatz	169
7.3.4.1. Die quadratischen Expansionen.....	169
7.3.4.2. Anwendung der quadratischen Terme in der Hauptkomponenten-Einflußanalyse	171
7.3.5. Weitere Stabilitätsindikatoren für ausgewählte Hauptkomponenten nach BENASSENI.....	172
7.3.5.1. Überlegungen zur Herleitung eines geeigneten Similaritätsmaßes.....	172
7.3.5.2. Mögliche Ansatzpunkte	173
7.3.5.3. Die Definition geeigneter Similaritätsmaße	177
7.3.5.4. Die Sensitivität der Koeffizienten in der Einflußidentifikation.....	178

7.3.5.5. Darstellung der Stichproben-Versionen	179
7.3.5.6. Abschließende Hinweise	180
7.3.6. Praktische Datenanalysen.....	182
7.3.6.1. Einflußuntersuchung der Milchbestandteildaten	182
7.3.6.2. Untersuchung der Luftverschmutzungsdaten	200
7.3.6.3. Einflußuntersuchung für die Verdichtung der Bilanzkennzahlen auf 14 Faktoren.....	215
7.4. Einflußmessung in der Faktorenanalyse.....	217
7.4.1. Kurzdarstellung der Grundgleichungen und Annahmen der Faktorenanalyse	218
7.4.2. Einflußmessung in der Hauptfaktorenanalyse	219
7.4.2.1. Darstellung der Konstruktionsschritte für die Einflußfunktionen.....	220
7.4.2.2. Konzentration einer Einflußanalyse auf die Konstruktion der empirischen Einflußkurve	224
7.4.3. Sensitivitätsmessung in der Maximum-Likelihood-Faktorenanalyse.....	226
7.4.3.1. Kurzdarstellung der Maximum-Likelihood-Faktorenanalyse	226
7.4.3.2. Die theoretischen Einflußfunktionen für die Einzelrestvarianz und die Varianz der gemeinsamen Faktoren	228
7.4.3.3. Darstellung des eigentlichen Identifikationsverfahrens	230
7.4.3.4. Überprüfung der theoretischen Einflußfunktionen auf Skaleninvarianz.....	232
7.4.3.5. Weitere Einflußmaße für die Maximum-Likelihood-Faktorenanalyse	233
7.4.3.6. Abschließende Anmerkungen zu den vorgestellten Diagnosemaßen.....	235
7.4.4. Zusätzliche Berücksichtigung der quadratischen Koeffizienten für eine effiziente Einflußidentifikation.....	236
7.4.4.1. Hauptfaktorenanalyse	236

7.4.4.2. Faktorenanalyse auf Basis der Korrelationsmatrix.....	239
7.4.4.3. Anwendung der quadratischen Expansionen.....	240
7.4.4.4. Abschließende Anmerkungen.....	240
7.4.4.5. Die quadratischen Störungsexpansionen für die Maximum-Likelihood-Faktorenanalyse	242
7.4.5. Sensitivitätskoeffizienten für die Faktorladungsmatrix	244
7.4.5.1. Der RV-Koeffizient als Konstruktionsansatz	244
7.4.5.2. Zusätzliche Berücksichtigung der quadratischen Expansionen.....	246
7.4.6. Weitere Einflußmaße in der Faktorenanalyse	248
7.4.6.1. Jackknife-Distanzmaße für die Hauptfaktorenanalyse	248
7.4.6.2. Eine Einflußfunktion für den χ^2 -Anpassungstest in der Maximum-Likelihood-Faktorenanalyse	250
7.4.6.3. Einflußfunktionen für die Determinanten der Einzelrest- und Modellkovarianz.....	252
7.4.6.4. Likelihood-Distanzmessung für die Maximum-Likelihood-Faktorenanalyse	254
7.4.6.5. Abschließende Anmerkungen.....	257
7.4.7. Praktische Anwendungsbeispiele für die iterative Hauptfaktorenanalyse.....	259
7.4.7.1. Untersuchung der Milchbestandteildaten	259
7.4.7.2. Untersuchung ausgewählter Bilanzkennzahlen.....	273
7.4.8. Einflußmessung für die Maximum-Likelihood-Faktorenanalyse.....	284
7.4.8.1. Einflußmessung für eine Maximum- Likelihood-Faktorenanalyse der Aktienkursdaten.....	284
7.4.8.2. Einflußmessung für eine Maximum-Likelihood- Faktorenanalyse der Bilanzkennzahlen	297
7.5. Einflußanalyse in der Diskriminanzanalyse	307
7.5.1. Einflußanalyse für interessierende Statistiken der linearen Diskriminanzanalyse nach CAMPBELL.....	308
7.5.1.1. Theoretische Einflußfunktionen für die lineare Diskriminanzanalyse.....	308

7.5.1.2. Verteilungstheoretische Überlegungen.....	312
7.5.1.3. Die Stichproben-Versionen	315
7.5.1.4. Abschließende Anmerkungen.....	316
7.5.2. Ein Einflußmessungsansatz für die Mißklassifikationswahrscheinlichkeit.....	318
7.5.2.1. Elemente der Einflußmessung für die Mißklassifikationswahrscheinlichkeit	318
7.5.2.2. Einflußanalyse für jeden einzelnen Beobachtungsvektor	319
7.5.2.3. Intensivierung der Ausführungen von CAMPBELL.....	321
7.5.2.4. Die graphische Veranschaulichung der Diskriminanzanalyse-Daten.....	324
7.5.2.5. Die Identifikation von multiplen Einflußbeobachtungen	328
7.5.2.5.1. Der Gewichtungsansatz nach PREGIBON	328
7.5.2.5.2. Typen des Masking Effektes in der linearen Diskriminanzanalyse.....	332
7.5.3. Einflußanalytische Ergänzungen aus den parallelen Erkenntnissen von FUNG	334
7.5.4. Einflußmaße für die quadratische Diskriminanzanalyse.....	338
7.5.4.1. Grundzüge der quadratischen Diskriminanzanalyse	338
7.5.4.2. Die Formulierung von geeigneten Diagnosemaßen	339
7.5.4.3. Konstruktion von Q-Q-Diagrammen zur visuellen Dateninspektion	343
7.5.4.4. Abschließende Anmerkungen.....	344
7.5.5. Die Entwicklung von geeigneten Diagnosemaßen in der mehrfachen Diskriminanzanalyse	345
7.5.5.1. Kurzdarstellung der Basisgrößen in der mehrfachen Diskriminanzanalyse.....	345
7.5.5.2. Diagnosemaße für die geschätzte Gruppenzugehörigkeitswahrscheinlichkeit.....	346
7.5.5.3. Weitere Aspekte der Einflußanalyse	348

7.5.5.4. Wichtige Erkenntnisse aus Simulationsstudien und einer Fallstudie.....	350
7.5.6. Die Berücksichtigung von Fehlklassifikationskosten in der Einflußmessung.....	352
 7.5.6.1. Der Fehlklassifikationskostenansatz in der linearen Diskriminanzanalyse.....	352
 7.5.6.2. Der Fehlklassifikationskostenansatz in der quadratischen Diskriminanzanalyse.....	354
 7.5.6.3. Der Fehlklassifikationskostenansatz in der mehrfachen Diskriminanzanalyse.....	355
7.6. Praktische Datenanalysen in der Diskriminanzanalyse	357
7.6.1. Einflußmessung in der linearen Diskriminanzanalyse am Beispiel der Konkursdaten	357
7.6.2. Einflußanalyse in der quadratischen Diskriminanzanalyse für die Konkursdaten	373
7.6.3. Die Identifikation von einflußreichen Beobachtungen in der mehrfachen Diskriminanzanalyse am Beispiel der Rohöldaten	379
8. Kritische Würdigung der Einflußanalysemethoden.....	385
9. Schlußbetrachtung.....	390

Anhang

Anhang 1:

Streichungs-Mahalanobis-Distanz 392

Anhang 2:

Simulation von standardnormalverteilten Zufallsvariablen in SAS 394

Anhang 3:

Störungseffekte der Eigenwerte und Eigenvektoren:

Herleitung der Differentialkoeffizienten erster Ordnung von ϵ 395

Anhang 4:

Herleitung der Matrix-Koeffizienten nach TANAKA für
die Hauptfaktorenanalyse am Beispiel einer (3x3) Matrix 399

Anhang 5:

Überprüfung der Approximationsgenauigkeit EIC≈SIC für die Einzelrestvarianz
einer Hauptfaktorenanalyse der Milchbestandteildaten 402

Anhang 6:

Berechnung der approximativen Gamma-Quantile
nach WILK et al. (1962) für die Konkursdaten 403

Anhang 7:

IML-Makros und Programme zur Ausreißer- und Einflußanalyse 409

Anhang 8:

Datensätze 555

Literaturverzeichnis XXVII