

---

# **Mathematisch-statistische Datenanalyse**

von Henning Läuter und Richard Pincus

Mit 70 Abbildungen und 9 Tabellen



---

Akademie-Verlag Berlin

1989

---

# Inhalt

Einführung . . . . .	13
<b>1. Grundprinzipien der statistischen Datenanalyse . . . . .</b>	<b>18</b>
<b>1.1. Vorläufige Datenanalyse . . . . .</b>	<b>18</b>
<b>1.1.1. Eindimensionale Darstellung von Daten . . . . .</b>	<b>19</b>
<b>1.1.2. Gruppierung von mehrdimensionalen Daten . . . . .</b>	<b>28</b>
<b>1.1.2.1. Charakterisierung von Clusterverfahren . . . . .</b>	<b>28</b>
<b>1.1.2.2. Minimaldistanzverfahren . . . . .</b>	<b>29</b>
<b>1.1.2.3. Berechnung optimaler Cluster . . . . .</b>	<b>30</b>
<b>1.1.3. Darstellungen mit orthogonalen Funktionen . . . . .</b>	<b>32</b>
<b>1.1.3.1. Stückweise lineare Funktionen . . . . .</b>	<b>33</b>
<b>1.1.3.2. Trigonometrisches Funktionensystem . . . . .</b>	<b>35</b>
<b>1.1.4. Transformation der Daten . . . . .</b>	<b>37</b>
<b>1.2. Statistische Modelle . . . . .</b>	<b>38</b>
<b>1.3. Likelihoodfunktionen . . . . .</b>	<b>40</b>
<b>1.4. Suffiziente Statistiken . . . . .</b>	<b>42</b>
<b>1.5. Vollständige Statistiken . . . . .</b>	<b>43</b>
<b>1.6. Schlußweisen in statistischen Modellen . . . . .</b>	<b>44</b>
<b>1.6.1. Statistische Entscheidungsprobleme . . . . .</b>	<b>44</b>
<b>1.6.1.1. Testen von Hypothesen . . . . .</b>	<b>45</b>
<b>1.6.1.2. Mehrentscheidungsverfahren . . . . .</b>	<b>48</b>
<b>1.6.1.3. Multiple Tests . . . . .</b>	<b>48</b>
<b>1.6.1.4. Punktschätzungen . . . . .</b>	<b>49</b>
<b>1.6.1.5. Konfidenzintervalle . . . . .</b>	<b>49</b>
<b>1.6.2. Entscheidungsprinzipien . . . . .</b>	<b>50</b>
<b>1.6.2.1. Maximum-Likelihood-Prinzip . . . . .</b>	<b>51</b>
<b>1.6.2.2. Kleinste-Quadrat-Methode . . . . .</b>	<b>53</b>
<b>1.6.2.3. Bayessches Prinzip . . . . .</b>	<b>55</b>
<b>1.6.3. Optimale statistische Entscheidungsfunktionen . . . . .</b>	<b>56</b>
<b>1.7. Resampling-Methoden . . . . .</b>	<b>59</b>
<b>2. Varianzanalyse . . . . .</b>	<b>61</b>
<b>2.1. Mittelwertvergleiche . . . . .</b>	<b>61</b>
<b>2.1.1. Das lineare Regressionsmodell . . . . .</b>	<b>63</b>
<b>2.1.1.1. Die Kleinste-Quadrat-Schätzung . . . . .</b>	<b>64</b>
<b>2.1.1.2. Zentrierte Daten . . . . .</b>	<b>64</b>
<b>2.1.1.3. Schätzbare Funktionen . . . . .</b>	<b>66</b>
<b>2.1.1.4. Lineare Restriktionen . . . . .</b>	<b>67</b>
<b>Versuchspläne von vollem Rang (69), Reparametrisierungen (69)</b>	
<b>2.1.1.5. Schätzung der Varianz . . . . .</b>	<b>70</b>
<b>2.1.1.6. Hypothesentests . . . . .</b>	<b>70</b>
<b>Testbare Teilhypothesen (72)</b>	

---

2.1.1.7.	Konfidenzbereiche . . . . .	73
2.1.1.8.	Optimalität der Verfahren . . . . .	75
2.1.2.	Klassifikationsmodelle . . . . .	76
2.1.2.1.	Kreuz-Klassifikation Schätzungen (77), Zerlegung der Varianz (78), Klassifikation ohne Wechselwirkung (81), <i>K</i> -faktorielle Klassifikation (81), Interpretation der Effekte (84)	77
2.1.2.2.	Hierarchische Klassifikation . . . . .	86
	Schätzungen (88), Zerlegung der Varianz (88), <i>K</i> -fache Klassifikation (88)	
2.1.2.3.	Gemischte Klassifikationstypen . . . . .	89
2.1.2.4.	Klassifikation und lineare Regressionsmodelle . . . . .	90
	Schätzbare Funktionen (91), Testbare Hypothesen (92), Reparametrisierungen (92), Konfidenzbereiche (93), Orthogonalität (94)	
2.1.2.5.	Unbalancierte Versuchspläne . . . . .	94
	Lineare Modelle (94), Reparametrisierungen und Schätzungen (95), Zerlegung der Varianz und Hypothesentests (97)	
2.1.3.	Multivariate Modelle . . . . .	100
2.1.3.1.	Mehrdimensionale Klassifikationsmodelle . . . . .	101
	Schätzung der Effekte (101), Schätzung der Kovarianz (102), Zerlegung der Varianz und Testen von Hypothesen (103), Simultane Konfidenzintervalle (105), Fehlende Beobachtungen (106)	
2.1.3.2.	Merkmalsauswahl . . . . .	108
	Merkmalsauswahl bei <i>k</i> Stichproben (110)	
2.1.4.	Multiple Mittelwertvergleiche . . . . .	111
2.1.4.1.	Der multiple t-Test . . . . .	112
2.1.4.2.	Der Scheffé-Test . . . . .	112
2.1.4.3.	Der Tukey-Test . . . . .	113
2.1.4.4.	Der Dunnett-Test . . . . .	114
2.1.4.5.	Der Newman-Keuls-Test . . . . .	115
2.1.4.6.	Bonferroni-Tests . . . . .	117
2.2.	Varianzkomponenten . . . . .	118
2.2.1.	ANOVA-Schätzungen . . . . .	119
2.2.1.1.	Zufällige Modelle . . . . .	120
	Hierarchische Klassifikation (121), Kreuzklassifikation (122)	
2.2.1.2.	Gemischte Modelle . . . . .	124
2.2.2.	MINQUE-Schätzungen . . . . .	125
	Minimum-Varianz-Schätzung (126), Berechnung der MINQUE (128), Iterative Anwendungen (130), Modifikationen von MINQUE (130)	
2.2.3.	Negative Varianzkomponenten . . . . .	131
2.2.3.1.	Negative Schätzungen . . . . .	131
2.2.3.2.	Fest oder zufällig? . . . . .	131
2.2.4.	Multivariate Varianzkomponenten . . . . .	132
2.2.4.1.	ANOVA-Schätzungen . . . . .	133
2.2.4.2.	MINQUE-Schätzungen . . . . .	133
<b>3.</b>	<b>Diskrimination und Klassifikation . . . . .</b>	<b>135</b>
3.1.	Optimale Verfahren zur Diskrimination . . . . .	137
3.1.1.	Diskriminanzverfahren mit maximaler Korrektklassifikationswahrscheinlichkeit	140
3.1.2.	Diskriminanzverfahren bei Kosten für Fehlklassifikationen . . . . .	142
3.1.3.	Diskrimination mit A-posteriori-Wahrscheinlichkeiten . . . . .	144
3.1.4.	Maximum-Likelihood-Klassifikation . . . . .	144
3.1.4.1.	Definition der Maximum-Likelihood-Regel . . . . .	144
3.1.4.2.	Autoregressive Zeitreihenmodelle . . . . .	146
3.1.5.	Substitutionsregeln . . . . .	149
3.1.5.1.	Einsetzen von Schätzungen für unbekannte Parameter . . . . .	149
3.1.5.2.	Einsetzen von Schätzungen für Dichten . . . . .	151
3.1.6.	Kombination von Diskriminationsregeln . . . . .	157

---

3.1.6.1.	Logische Verknüpfung . . . . .	159
3.1.6.2.	Konvexe Kombination . . . . .	161
3.2.	Mathematische Eigenschaften von Diskriminationsregeln . . . . .	166
3.2.1.	Asymptotisch sichere Verfahren . . . . .	167
3.2.2.	Klassifikationsverfahren mit monotoner Güte . . . . .	169
3.2.3.	Unverfälschtheit von Klassifikationsverfahren . . . . .	171
3.2.4.	Minimaxeigenschaften von Klassifikationsverfahren . . . . .	171
3.3.	Schätzung der Fehlerquote für Diskriminanzverfahren . . . . .	178
3.3.1.	Schätzungen für die Fehlerquote . . . . .	178
3.3.1.1.	Resubstitutionsschätzung . . . . .	179
3.3.1.2.	Schätzung nach der U-Methode . . . . .	179
3.3.1.3.	Schätzung nach der Bootstrap-Methode . . . . .	181
3.3.1.4.	Schätzungen in parametrischen Verteilungen . . . . .	182
3.3.2.	Eigenschaften der Fehlerquotenschätzungen . . . . .	186
3.3.2.1.	Reduktion der Verzerrung der Bootstrap-Schätzung . . . . .	189
3.4.	Diskriminanzanalysen als Bestandteil von Datenanalysen . . . . .	191
3.4.1.	Existieren Unterschiede zwischen Klassen? . . . . .	192
3.4.1.1.	Normalverteilungen in den Klassen . . . . .	192
3.4.1.2.	Mischungen von Normalverteilungen in den Klassen . . . . .	195
3.4.1.3.	Paarweiser Vergleich normalverteilter Klassen . . . . .	200
3.4.2.	Klassifikation und Diskrimination . . . . .	202
3.4.2.1.	Grafische Darstellung der Diskrimination . . . . .	204
3.4.2.2.	Bewertung und Steuerung der Klassenzugehörigkeit . . . . .	207
3.4.2.3.	Auswahl wesentlicher Merkmale . . . . .	211
3.4.2.4.	Zweckmäßige Schritte bei der Diskrimination . . . . .	213
3.5.	Stellung der Diskriminanzanalyse in der Datenanalyse . . . . .	214
4.	<b>Schätzung in statistischen Modellen</b> . . . . .	215
4.1.	Parametrische und nichtparametrische Modelle . . . . .	215
4.2.	Schätzung von Parametern . . . . .	217
4.3.	Cramér-Rao-Ungleichung . . . . .	220
4.4.	Prinzipielle Verbesserungsmöglichkeit der besten erwartungstreuen Schätzung . . . . .	224
4.5.	Schätzung nichtlinearer Parameter in linearen Modellen . . . . .	225
4.5.1.	Nichtlineare Parameter in linearen Modellen . . . . .	226
4.5.2.	Schätzung nichtlinearer Parameter . . . . .	226
4.5.2.1.	Maximum-Likelihood-Schätzung von $\lambda(P)$ . . . . .	226
4.5.2.2.	Bedingte Schätzungen . . . . .	228
4.5.2.3.	Approximation der bedingten Schätzungen . . . . .	231
4.5.2.4.	Güte der bedingten Schätzungen . . . . .	232
4.5.3.	Datenangepaßte bedingte Schätzungen . . . . .	235
4.5.4.	Schätzung spezieller nichtlinearer Parameter . . . . .	236
4.5.4.1.	Schätzung höherer Momente . . . . .	237
4.5.4.2.	Schätzung von Funktionalen . . . . .	237
4.5.5.	Asymptotisch effiziente Schätzung . . . . .	238
4.6.	Schätzung linearer Parameter in eingeschränkten Modellen . . . . .	240
4.6.1.	Schätzung von Parametern in skaleninvarianten Modellen . . . . .	241
4.6.2.	Schätzung von Parametern aus beschränkten Bereichen . . . . .	245
4.6.2.1.	Eingesetzte Ridge-Schätzung . . . . .	245
4.6.2.2.	Eingeschränkte Schätzungen . . . . .	247
5.	<b>Regressionsmodelle</b> . . . . .	253
5.1.	Lineare Verlaufskurvenmodelle . . . . .	253
5.1.1.	Orthogonale Funktionen . . . . .	254
5.1.1.1.	Polynomialmodelle . . . . .	258

---

5.1.1.2.	Tschebyscheff-Polynome . . . . .	259
5.1.1.3.	Trigonometrische Funktionen . . . . .	259
5.1.2.	Lokale und globale Modelle . . . . .	259
5.2.	Nichtlineare Regressionsmodelle . . . . .	262
5.2.1.	Schätzung der Regressionsfunktion . . . . .	264
5.2.1.1.	Maximum-Likelihood-Schätzung . . . . .	264
5.2.1.2.	Kleinste-Quadrat-Schätzung . . . . .	265
5.2.2.	Schätzung der Parameter . . . . .	266
5.2.2.1.	Bedingte Schätzungen . . . . .	267
5.2.2.2.	Datenangepaßte bedingte Schätzungen . . . . .	269
5.2.3.	Eigenschaften der Schätzungen . . . . .	269
5.2.4.	Numerische Methoden zur Berechnung von Schätzungen . . . . .	270
5.2.5.	Spezielle nichtlineare Modelle . . . . .	272
5.2.5.1.	Periodische Modelle . . . . .	272
5.2.5.2.	Versteckte Perioden in Verlaufskurven . . . . .	285
5.2.5.3.	Wachstumsmodelle . . . . .	291
<b>6.</b>	<b>Approximation für Verlaufskurven . . . . .</b>	<b>297</b>
6.1.	Glättungsfunktionen . . . . .	297
6.1.1.	Definition von Glättungsverfahren . . . . .	298
6.1.1.1.	Polynomiale Glätter . . . . .	299
6.1.1.2.	Spline-Glättungen . . . . .	301
6.1.1.3.	Nichtlineare Glättungsfunktionen . . . . .	311
6.2.	Eigenschaften von Glättungen . . . . .	311
6.2.1.	Optimale lineare Glätter . . . . .	314
6.2.2.	Güte nichtlinearer Glätter . . . . .	317
6.3.	Nichtparametrische Regression . . . . .	320
6.3.1.	Regressogramm . . . . .	321
6.3.2.	Kernschätzungen . . . . .	321
6.3.3.	Eigenschaften der Kernschätzungen . . . . .	322
<b>7.</b>	<b>Methoden der Versuchsplanung . . . . .</b>	<b>324</b>
7.1.	Charakterisierung optimaler Versuchspläne . . . . .	326
7.2.	Optimale Versuchsplanung . . . . .	328
7.2.1.	Versuchspläne für die Schätzung . . . . .	328
7.2.2.	Versuchspläne für Test- und Klassifikationsprobleme . . . . .	330
7.2.3.	Versuchspläne bei der nichtlinearen Regression . . . . .	332
7.3.	Bemerkungen zur Versuchsplanung . . . . .	333
<b>8.</b>	<b>Anpassungs- und Ausreißertests . . . . .</b>	<b>335</b>
8.1.	Anpassungstests . . . . .	335
8.1.1.	Tests vom $\chi^2$ -Typ . . . . .	335
8.1.1.1.	Anpassung an eine gegebene Verteilung . . . . .	336
8.1.1.2.	Geschätzte Parameter . . . . .	337
	Minimum-X <sup>2</sup> -Schätzungen (338), Maximum-Likelihood-Schätzungen (338), Zu-fällige Intervalle (340), Regressionsmodelle (341)	
8.1.2.	Tests vom Kolmogorow-Smirnow-Typ . . . . .	342
8.1.2.1.	Einfache Hypothesen . . . . .	342
	Multivariate Beobachtungen (343)	
8.1.2.2.	Geschätzte Parameter . . . . .	344
8.1.2.3.	Zwei Stichproben . . . . .	344
8.1.3.	Grafische Verfahren . . . . .	345
8.1.3.1.	Empirische Verteilungs-Plots . . . . .	345
8.1.3.2.	Wahrscheinlichkeits-Papier . . . . .	347

8.2.	Ausreißertests . . . . .	349
8.2.1.	Abweichungen vom Mittelwert . . . . .	349
8.2.2.	Mehrere Ausreißer . . . . .	350
8.2.2.1.	Bekannte Ausreißeranzahl . . . . .	351
8.2.2.2.	Unbekannte Anzahl von Ausreißern . . . . .	352
	Abstand benachbarter Beobachtungen (352), Wahrscheinlichkeits-Plots (353)	
8.2.3.	Multivariate Ausreißer . . . . .	353
	Multivariate studentisierte Abweichungen (353), Grafische Verfahren (355)	
<b>Anhänge</b>	. . . . .	358
A. 1.	Matrizen, verallgemeinerte Inverse . . . . .	358
A. 2.	Funktionen normalverteilter Variabler . . . . .	359
<b>Tafeln</b>	. . . . .	361
I.	Verteilungsfunktion der Normalverteilung . . . . .	361
II.	Signifikanzpunkte der $\chi^2$ -Verteilung . . . . .	362
III.	Signifikanzpunkte des zweiseitigen t-Tests . . . . .	363
IV.	Signifikanzpunkte des F-Tests . . . . .	364
V.	Signifikanzpunkte der studentisierten Variationsbreite . . . . .	366
VI.	Signifikanzpunkte des Dunnett-Tests . . . . .	368
VII.	Signifikanzpunkte des Kolmogorow-Smirnow-Anpassungstests . . . . .	370
VIII.	Signifikanzpunkte des Kolmogorow-Smirnow-Anpassungstests für die Normalverteilung mit unbekannten Mittelwerts- und Varianzparametern . . . . .	371
IX.	Signifikanzpunkte des Kolmogorow-Smirnow-Tests für die Homogenität zweier Stichproben gleichen Umfangs . . . . .	372
X.	Signifikanzpunkte für die maximale standardisierte Abweichung vom Mittelwert . . . . .	373
XI.	Signifikanzpunkte der größten Lücke in einer gleichverteilten Stichprobe . . . . .	374
<b>Literaturverzeichnis</b>	. . . . .	375
<b>Symbolverzeichnis</b>	. . . . .	387
<b>Sachverzeichnis</b>	. . . . .	389