

WOLF-MICHAEL KÄHLER

STATISTISCHE DATENANALYSE MIT
SPSS/PC+

Eine Einführung in Grundlagen und Anwendung

Dritte, verbesserte und erweiterte Auflage



Inhaltsverzeichnis

1	Vorbereitungen zur Datenanalyse	1
1.1	Beispiel einer empirischen Untersuchung	1
1.2	Einsatz von SPSS/PC+	2
1.3	Kodierung und Strukturierung der Daten	5
1.3.1	Kodeplan	5
1.3.2	Kodierung von fehlenden Werten	6
1.3.3	Datenmatrix	6
1.3.4	Daten-Datei	7
1.3.5	Erfassungsvorschrift	8
1.4	Datenerfassung mit dem REVIEW-Editor	9
1.4.1	Beginn des Dialogs mit dem SPSS-System	9
1.4.2	Aktivierung des REVIEW-Editors	10
1.4.3	Dateneingabe im Review-Schirm	11
1.4.4	Editier-Hilfen	12
1.4.5	Datensicherung und Dialogende	15
1.5	Meßniveau der Merkmale	17
2	Einsatz des SPSS-Systems	18
2.1	Ein SPSS-Programm zur Häufigkeitsauszählung	18
2.1.1	Die SPSS-Kommandos	18
2.1.2	SPSS-file und Variable	19
2.1.3	Häufigkeitsauszählung	21

2.2	Das Arbeiten im Submit-Modus	22
2.2.1	Arbeiten ohne SPSS-Menü-System	23
2.2.2	Arbeiten mit dem SPSS-Menü-System	30
2.2.3	Die Dateien SPSS.LOG, SPSS.LIS, SCRATCH.PAD und SPSSPROF.INI	36
2.3	Aufbau eines SPSS-Programms	38
2.3.1	Programm-Struktur	38
2.3.2	Struktur eines SPSS-Kommandos	40
2.3.3	Struktur der Spezifikationen und Trennzeichen	41
3	Vereinbarung, Beschreibung und Veränderung des SPSS- files	43
3.1	Beschreibung der Dateneingabe (DATA LIST, BEGIN DATA, END DATA)	43
3.1.1	Syntax des DATA LIST-Kommandos	43
3.1.2	Variablennamen	44
3.1.3	Eingabe ganzzahliger Werte	45
3.1.4	Eingabe von Leerzeichen	46
3.1.5	Variablenliste	47
3.1.6	Inklusive Variablenlisten	47
3.1.7	Eingabe nicht ganzzahliger Werte	48
3.1.8	Mehrere Datensätze pro Case	49
3.1.9	Alphanumerische Variable und alphanumerische Werte	50
3.1.10	Dateneingabe ohne Daten-Datei (BEGIN DATA, END DATA)	50
3.1.11	Kompakte Datenhaltung (SET/COMPRESS)	51
3.2	Variablen- und Werteetiketten	52
3.2.1	Etikettierung von Variablen (VARIABLE LABELS)	52
3.2.2	Etikettierung von Werten (VALUE LABELS)	53
3.3	Vereinbarung von missing values (MISSING VALUE)	55
3.3.1	Syntax des MISSING VALUE-Kommandos	55
3.3.2	Der system-missing value	56

3.4	Veränderung und Ergänzung des SPSS-files (COMPUTE, RECODE)	57
3.4.1	Beispiel für eine Werteänderung	57
3.4.2	Das Kommando COMPUTE	57
3.4.3	Die Kommandos RECODE und AUTORECODE . . .	58
3.4.3.1	Rekodierungsvorschrift	58
3.4.3.2	Die Schlüsselwörter ELSE, THRU, LOWEST und HIGHEST	59
3.4.3.3	Automatische Umwandlung mit AUTORECODE . . .	60
3.5	Überprüfung der Eingabedaten	61
3.5.1	Eingabefehler	61
3.5.2	Überprüfung von Werten	61
3.5.3	Temporäre Datenauswahl (PROCESS IF)	62
3.5.4	Datenausgabe in die Listing-Datei (LIST)	63
3.6	Inhalt des SPSS-files (DISPLAY, MODIFY VARS)	64
4	Beschreibung von Merkmalen	66
4.1	Häufigkeitsverteilungen und Statistiken (FREQUENCIES, DESCRIPTIVES)	66
4.1.1	Ausgabe von Häufigkeitsverteilungen (FREQUENCIES)	66
4.1.2	Steuerung der Ausgabe (FORMAT, MISSING)	68
4.1.3	Ausgabe von Histogrammen (HISTOGRAM)	71
4.1.4	Ausgabe von Balkendiagrammen (BARChart)	72
4.1.5	Berechnung von Statistiken	73
4.1.5.1	Die Subkommandos PERCENTILES und NTILES . . .	73
4.1.5.2	Das Subkommando STATISTICS	75
4.1.6	Berechnung von Statistiken für kontinuierliche Merk- male (DESCRIPTIVES)	81

4.2	Beschreibung von Merkmalen durch einen Report (REPORT)	84
4.2.1	Beispiel	84
4.2.2	Das Kommando REPORT	87
4.2.3	Ausgabe von Statistiken (SUMMARY)	88
4.2.3.1	Einfache Statistiken	88
4.2.3.2	Zusammengesetzte Statistiken	91
4.2.3.3	Gestaltung der Ausgabe von Statistiken	94
4.2.4	Vereinbarung der Spalten-Variablen (VARIABLES)	96
4.2.5	Vereinbarung einer Break-Variablen (BREAK)	99
4.2.6	Gestaltung der Reportausgabe (FORMAT)	103
4.2.7	Textausgabe in Kopf- und Fußzeilenbereiche (TITLE, FOOTNOTE)	106
4.2.8	Ausgabe von Variablenwerten (LIST, SUMSPACE)	109
4.2.9	Verrechnung von missing values (MISSING)	110
4.2.10	Report-Struktur bei mehreren Break-Variablen	111
4.2.11	Ausgabe von Strings (STRING)	115
4.3	Sortierung des SPSS-files (SORT CASES)	117
4.4	Vereinfachte Reportausgabe für intervallskalierte Merkmale (MEANS)	120
4.5	Untersuchung von Merkmalen (EXAMINE)	122
4.5.1	Beschreibung von Verteilungen durch Histogramme	122
4.5.2	Beschreibung von Verteilungen durch "Stem-and-leaf"-Plots	126
4.5.3	Boxplots	127
4.5.4	Gruppenvergleiche	130
4.5.5	Überprüfung auf Normalverteilung	134
4.5.6	Schätzung der zentralen Tendenz	136
4.5.7	Berechnung von Perzentilwerten	137
4.5.8	Zusammenfassung	138
5	Beschreibung der Beziehung von Merkmalen	140
5.1	Analyse von Kontingenz-Tabellen (CROSSTABS)	140
5.1.1	Die gemeinsame Häufigkeitsverteilung zweier Merkmale	140
5.1.2	Ausgabe von Kontingenz-Tabellen	142
5.1.3	Steuerung der Tabellenausgabe	144

5.1.4	Statistischer Zusammenhang zwischen nominalskalierten Merkmalen	148
5.1.4.1	Chi-Quadrat	150
5.1.4.2	Phi-Koeffizient	151
5.1.4.3	Cramer's V	152
5.1.4.4	Kontingenzkoeffizient C	152
5.1.4.5	Der Likelihood-Quotienten-Chi-Quadrat-Wert	153
5.1.4.6	Das PRE-Maß Lambda	153
5.1.4.7	Der Tau-Koeffizient von Goodman und Kruskal	156
5.1.4.8	Cohen's Kappa	156
5.1.4.9	Das relative Risiko	157
5.1.5	Statistischer Zusammenhang zwischen ordinalskalierten Merkmalen	159
5.1.5.1	Konkordante und diskordante Paare	160
5.1.5.2	Positive und negative Beziehungen	161
5.1.5.3	Der Gamma-Koeffizient	161
5.1.5.4	Der Koeffizient Somers' d	163
5.1.5.5	Kendall's Tau-B und Tau-C	164
5.1.6	Statistischer Zusammenhang zwischen intervallskalierten Merkmalen	165
5.1.6.1	Korrelationskoeffizient r	165
5.1.6.2	Spearman's Rho	167
5.1.6.3	Der Koeffizient Eta-Quadrat	168
5.1.6.4	Berechnung von r und Eta	169
5.1.7	Inferenzstatistisches Schließen	170
5.1.8	CROSSTABS im Ganzzahl-Modus	175

5.2	Beschreibung der Beziehung von intervallskalierten Merkmalen	176
5.2.1	Graphische Beschreibung (PLOT)	176
5.2.1.1	Streudiagramm	176
5.2.1.2	Regressionsgerade	177
5.2.1.3	Gestaltung des Layouts	178
5.2.1.4	Verarbeitung von missing values	181
5.2.1.5	Kontrollvariable	181
5.2.1.6	Überlagerung von Streudiagrammen	182
5.2.1.7	Aufbau des PLOT-Kommandos	182
5.2.2	Korrelationskoeffizient von Bravais-Pearson (CORRELATION)	184
5.2.3	Vergleich von Mittelwerten (MEANS)	187
5.2.3.1	Das STATISTICS-Subkommando	187
5.2.3.2	Varianzanalyse-Tafel	188
5.2.3.3	Linearitäts-Test	189
5.2.4	Mittelwertvergleich für zwei Gruppen (T-TEST)	191
5.2.4.1	Der T-Test	191
5.2.4.2	T-Test für abhängige Stichproben	193
5.2.4.3	Das Kommando T-TEST	195
6	Veränderung des SPSS-files	196
6.1	Berechnung von arithmetischen Ausdrücken (COMPUTE)	196
6.2	Rekodierung von Variablenwerten (RECODE)	199
6.3	Bedingte Zuweisung (IF)	201
6.4	Auszählung von Werten (COUNT)	203
6.5	Gewichtung von Cases (WEIGHT)	204

6.6	Datenauswahl	206
6.6.1	Gezielte Auswahl von Cases (PROCESS IF, SELECT IF)	206
6.6.2	Überprüfung der Satzfolge	208
6.6.3	Auswahl der ersten Cases (N)	209
6.6.4	Zufällige Auswahl von Cases (SAMPLE, SET/SEED)	210
7	Protokollausgaben des SPSS-Systems	211
7.1	Ausgabe von Kommandos und Analyseergebnissen (SET, SHOW)	211
7.2	Ausgabe von Seitenüberschriften (TITLE, SUBTITLE)	213
7.3	Kommentierung von SPSS-Kommandos (*)	214
8	Datenausgabe	215
8.1	Ausgabe von Variablenwerten(LIST, WRITE)	215
8.2	Bestimmung der Ergebnis-Datei (SET/RESULTS)	216
8.3	Ausgabeformate (FORMATS)	217
8.4	Datenausgabe bei den Auswertungsverfahren	217
9	Das Arbeiten mit SPSS-files und Datenaustausch	219
9.1	Sicherung des SPSS-files (SAVE, SYSFILE INFO)	219
9.2	Wiederherstellung des SPSS-files (GET)	222
9.3	Zusammenfassung von SPSS-files (JOIN)	224
9.3.1	Zusammenführung paralleler SPSS-files	224
9.3.2	Zusammenführung von nicht-parallelen SPSS-files	225
9.3.3	Aneinanderreihung von gleichstrukturierten SPSS-files	228
9.3.4	Mischen von gleichstrukturierten SPSS-files	230
9.4	Transponieren des SPSS-files (FLIP)	231

9.5	Datenaustausch mit Fremdsystemen	233
9.5.1	Erstellung einer portierbaren Sicherungs-Datei (EXPORT)	233
9.5.2	Umwandlung von portierbaren Sicherungs-Dateien in SPSS-files (IMPORT)	234
9.5.3	Datenaustausch mit dem Datenbanksystem dBASE und Tabellenkalkulationsprogrammen (TRANSLATE)	235
10	Speicherung von Rangwerten und Statistiken	238
10.1	Speicherung von Rangwerten (RANK)	238
10.1.1	Rangwerte und Bindungen	238
10.1.2	Berechnung von Spearman's Rho	241
10.1.3	Transformation der Rangwerte	241
10.1.4	Überprüfung auf Normalverteilung	244
10.2	Speicherung von Statistiken (AGGREGATE)	244
10.2.1	Beispiel	244
10.2.2	Indikator-Variable	247
10.2.3	Statistiken	248
10.2.4	Missing values	250
10.2.5	Syntax des AGGREGATE-Kommandos	250
11	Varianzanalyse	251
11.1	Einfaktorielle Varianzanalyse (ONEWAY)	251
11.1.1	Voraussetzungen und Nullhypothese	251
11.1.2	Varianzanalyse-Tafel	252
11.1.3	Überprüfung der Test-Voraussetzungen	253
11.1.4	Vergleiche einzelner Faktorstufen	254
11.1.5	"A priori"-Vergleiche	257
11.1.6	Trend-Tests	259
11.1.7	Syntax des ONEWAY-Kommandos	260
11.1.8	Eingabe von Statistiken in Matrixform	262

11.2	Mehrfaktorielle Varianzanalyse (ANOVA)	263
11.2.1	Beispiel	263
11.2.2	Varianzanalyse-Tafel	264
11.2.3	Zellenbesetzungen	265
11.2.4	Graphische Darstellung	266
11.2.5	Mehr als zwei Faktoren	268
11.2.6	Multiple Klassifikationsanalyse	268
11.2.7	Kovarianzanalyse	269
11.2.8	Syntax des ANOVA-Kommandos	271
11.2.9	Das Subkommando OPTIONS	271
11.2.10	Das Subkommando STATISTICS	273
12	Nichtparametrische Testverfahren (NPAR TESTS)	274
12.1	Vergleich mit einer theoretischen Verteilung	274
12.2	Vergleich zwischen empirisch ermittelten Verteilungen	276
12.2.1	Paarvergleich bei abhängigen Stichproben	277
12.2.2	Vergleich mehrerer Verteilungen bei abhängigen Stichproben	278
12.2.3	Verteilungs-Vergleich bei zwei unabhängigen Stichproben	279
12.2.4	Verteilungs-Vergleich bei mehreren unabhängigen Stichproben	281
12.3	Iterationstest für dichotomisierte Merkmale	282
12.4	Die Subkommandos OPTIONS und STATISTICS	283
13	Regressionsanalyse (REGRESSION)	284
13.1	Beschreibung der linearen Beziehung und Anpassungsgüte	284
13.2	Überprüfung der Linearitätsannahme	288
13.3	Voraussetzungen zur Durchführung von statistischen Tests	288

13.4	Identifikation von statistischen Ausreißern	293
13.5	Multikollinearität	296
13.6	Methoden der schrittweisen Regression	298
13.7	Syntax des Kommandos REGRESSION	300
14	Itemanalyse (RELIABILITY)	311
14.1	Skalenbildung	311
14.2	Vereinbarung von Skalen und Cronbach's Alpha	312
14.3	Korrelative Beschreibung von Skalen	313
14.4	Weitere Reliabilitätskoeffizienten für Skalen	317
15	Faktorenanalyse (FACTOR)	318
15.1	Die Hauptkomponentenanalyse	318
15.2	Extraktion und Festlegung der Faktorenzahl	320
15.3	Rotation zur Einfachstruktur	322
15.4	Vorabprüfung auf die Existenz gemeinsamer Faktoren	325
15.5	Weitere Verfahren zur Durchführung einer Faktorenanalyse	326
15.6	Sicherung der Faktorenwerte	328
15.7	Anzeige von Statistiken	328
15.8	Eingabe und Ausgabe von Matrizen	330
16	Clusteranalysen (CLUSTER, QUICK CLUSTER)	332
16.1	Verfahren und Ziele der Clusteranalyse	332
16.2	Ergebnisse der Clusteranalyse	334
16.3	Cluster-Kriterien zur Fusionierung	339
16.4	Sicherung und Bereitstellung von Distanz- und Ähnlichkeits- matrizen	343
16.5	Clusteranalyse für große Fallzahlen (QUICK CLUSTER)	345

17 Diskriminanzanalyse (DSCRIMINANT)	353
17.1 Zielsetzung der linearen Diskriminanzanalyse	353
17.2 Ein Beispiel	354
17.3 Kriterien zur Güte der Gruppentrennung	358
17.4 Syntax des Kommandos DSCRIMINANT	362
Anhang	370
A.1 Einführung in das Arbeiten unter MS-DOS	370
A.2 Der REVIEW-Editor (REVIEW)	374
A.3 Inhalt der Dateien SPSS.LOG und SPSS.LIS	379
A.4 Das Arbeiten im Dialog-Modus	381
A.5 Das Arbeiten im Batch-Modus	385
A.6 Formatfreie Dateneingabe (DATA LIST FREE)	387
A.7 Der Quick-Editor (QED)	389
A.8 Das SET-Kommando	398
A.9 Ausführung von MS-DOS-Kommandos (EXECUTE)	401
A.10 Datenerfassung mit DATA ENTRY II (DE)	402
Syntax der Kommandos	406
Modul-Struktur von SPSS/PC+	429
Literaturverzeichnis	430
Index	431