

# Datenmanagement und Datenanalyse mit dem SAS-System

Vom Einsteiger zum Profi

Von  
Dipl.-Psych.  
Christian FG Schendera

R.Oldenbourg Verlag München Wien

# Inhaltsverzeichnis

<b>Vorwort und Danksagung.....</b>	<b>V</b>
<b>Übersicht und Aufbau des Buches.....</b>	<b>IX</b>
<b>Inhaltsverzeichnis.....</b>	<b>XIII</b>
<b>I. Das SAS System.....</b>	<b>1</b>
<b>Übersicht.....</b>	<b>3</b>
<b>1. Die Anfänge von SAS und der Mensch im Mittelpunkt: Dr. Jim Goodnight.....</b>	<b>5</b>
<b>2. Das System hinter dem SAS System.....</b>	<b>16</b>
2.1. Anwenderorientierung.....	16
2.2. Das Konzept der MultiVendor Architecture.....	17
2.3. SAS als Programmierungsumgebung.....	19
2.4. Input: Eingabe, Einlesen, und Zugriff auf Daten unterschiedlichster Formate.....	21
2.5. Datenhaltungen und Data Warehouses.....	24
2.6. Ausblick: Dynamische Evolution des SAS Systems.....	25
<b>3. Viel mehr als nur Statistik: Das SAS System.....</b>	<b>27</b>
3.1. Das SAS System.....	27
3.2. Das Intelligence Layer Modell.....	28
3.2.1. Ebene: Enterprise Intelligence.....	29
3.2.2. Ebene: Suppliers, Organization, Customers.....	30
3.2.3. Ebene: Intelligence Architecture.....	36
3.3. Weitere SAS Produkte und Anwendungen.....	43
3.4. Übersichten zur SAS Software: Wirtschaftsbereiche, Lösungen, Produkte und der benutzerfreundliche Einstieg in die Datenanalyse.....	45
3.5. Erweiterungen, Änderungen und Verbesserungen in Version 8: Eine Auswahl.....	50
<b>4. SAS Produkte für die grafische und statistische Analyse von Daten.....</b>	<b>54</b>
4.1. SAS/TNSIGHT Software - Interaktive Datenanalyse.....	54
4.2. SAS/LAB - Ursache-Wirkungsmodelle mit Interpretation.....	56
4.3. Analyst Application - Interaktive Visualisierungen und statistische Analysen.....	57
4.4. SAS Research Analyst für professionelle Statistik im Labor.....	60

4.5.	SAS/ETS Software für Zeitreihen und Mehrgleichungsmodelle SAS Time Series Forecasting System für Trendprognosen.....	62
4.5.1.	SAS/ETS Software für Zeitreihen und Mehrgleichungsmodelle.....	62
4.5.2.	SAS Time Series Forecasting System für Trendprognosen.....	64
4.6.	SAS/GIS Software - Geografisches Informations- und Analysesystem.....	65
4.7.	SAS/SPECTARVIEW - Analyse und Visualisierung sehr großer Datenmengen.....	67
4.8.	SAS Market Research Application für Marktforschung.....	69
4.9.	Enterprise Guide - Analyse und Reporting Anwendung für Windows.....	70
4.10.	Management größter Projekte - PROJMAN Application.....	72
4.11.	JMP - Interaktive Datenanalyse und Versuchsplanung.....	73
4.12.	SAS/ASSIST - Der Power-Zugang zum SAS System.....	76
4.13.	Weitere SAS Produkte.....	77
<b>5.</b>	<b>SAS Prozeduren für die grafische und statistische Analyse von Daten.....</b>	<b>78</b>
5.1.	Basisverfahren.....	78
5.2.	Grafische Analyse (SAS/GRAPH Software).....	79
5.3.	Statistische Analyse (SAS/STAT Software).....	81
5.3.1.	Grundverfahren.....	82
5.3.2.	Professionelle Statistik.....	83
5.4.	Berichte, Grafiken und Präsentation.....	92
5.4.1.	Tabellen und Listen (PROC TABULATE, PROC REPORT, DataStep).....	93
5.4.2.	Grafiken (SAS/GRAPH).....	94
5.4.3.	Der SAS Enterprise Reporter.....	94
<b>II.</b>	<b>Der Einstieg SAS per Maussteuerung.....</b>	<b>97</b>
	<b>Übersicht.....</b>	<b>99</b>
<b>6.</b>	<b>Die Analyst Application.....</b>	<b>103</b>
6.1.	Einführung.....	103
6.1.	Start.....	105
6.2.	Daten-Management.....	107
6.2.1.	Modifikationen auf Variablen-Ebene.....	107
6.2.2.	Modifikationen auf Datensatz-Ebene.....	109
6.2.3.	Übungen.....	110
6.3.	Grafiken und Berichte.....	110
6.3.1.	Grafiken.....	110

6.3.2.	Berichte.....	114
6.3.3.	Übungen.....	115
6.4.	Statistik.....	115
6.4.1.	Deskriptive Statistik.....	115
6.4.2.	Hypothesentests.....	116
6.4.3.	Auswertung von Tabellen.....	119
6.4.4.	Varianzanalysen.....	121
6.4.5.	Stichproben- und Power-Berechnungen.....	124
6.4.6.	Nicht vorgestellte Möglichkeiten von SAS/ANALYST.....	126
6.4.7.	Übungen.....	127
6.5.	Einstellungen.....	127
<b>SAS/INSIGHT.....</b>		<b>129</b>
7.1.	Einführung, Übersicht und Steuerung.....	129
7.1.1.	Einführung.....	129
7.1.2.	Übersicht (Techniken).....	130
7.1.3.	Steuerung.....	132
7.2.	Daten-Management.....	134
7.2.1.	Interaktive Dateneingabe.....	135
7.2.2.	Öffnen eines Datasets + Daten-Management.....	135
7.2.3.	Übungen.....	136
7.3.	Datenexploration.....	137
7.3.1.	1 Dimensional (Y).....	137
7.3.2.	2 Dimensional (YX).....	139
7.3.3.	3 Dimensional (ZYG).....	141
7.3.4.	Übungen.....	142
7.4.	Multivariate Analysen: Analyse komplexer Variablenbeziehungen.....	143
7.4.1.	Korrelation.....	143
7.4.2.	ANOVA.....	143
7.4.3.	Regression.....	145
7.4.4.	Hauptkomponentenanalyse.....	152
7.4.5.	Übungen.....	154
<b>SAS/ASSIST Software.....</b>		<b>155</b>
8.0.	Einführung.....	155
8.1.	Überblick zu SAS/ASSIST.....	155
8.2.	Start von SAS/ASSIST.....	157
8.3.	Das WorkPlace Menü und seine Buttons.....	157
8.3.1.	<i>Data Management</i> .....	158
8.3.2.	<i>Report Writing</i> (Berichterstellung).....	159
8.3.3.	<i>Graphics</i> .....	161
8.3.4.	<i>Data Analysis</i> .....	164
8.3.5.	<i>Planning Tools</i> (Planwerkzeuge).....	168

8.3.6.	<i>EIS</i> .....	169
8.3.7	Remote Connect.....	169
8.3.8.	<i>Results</i> .....	169
8.3.9.	<i>Setup</i> .....	170
8.3.10.	<i>Index</i> .....	170
8.3.11.	<i>Exit</i> .....	170
8.4.	Übungen.....	170
<b>9.</b>	<b>Enterprise Guide</b> .....	<b>174</b>
9.1.	Leistungsmerkmale.....	174
9.2.	Orientierung im Enterprise Guide.....	180
9.2.1.	Starten und Beenden des Enterprise Guide.....	180
9.2.2.	Erster Überblick zu den Fenstern und Listen.....	180
9.3.	Das Auswahlfenster.....	184
9.3.1.	Öffnen einer bestehenden Datei.....	184
9.3.2.	Anlegen einer neuen Datei.....	185
9.4.	Die Menüleisten.....	185
9.4.1.	„Datei“.....	185
9.4.2.	„Bearbeiten“.....	186
9.4.3.	„Ansicht“.....	188
9.4.4.	„Einfügen“.....	188
9.4.5.	„Format“.....	189
9.4.6.	„Extras“.....	191
9.4.7.	„Daten“.....	195
9.4.8.	„Analyse“.....	200
9.4.9.	„Grafik“.....	204
9.4.10.	„Code“.....	206
9.4.11.	„Fenster“.....	207
9.4.12.	„Hilfe“.....	207
9.5.	Übungen.....	208
<b>III.</b>	<b>Syntaxsteuerung</b> .....	<b>211</b>
	<b>Übersicht</b> .....	<b>213</b>
<b>10.</b>	<b>Daten-Management</b> .....	<b>216</b>
10.0.	Einführung.....	216
10.1.	Was ist Daten-Management? Braucht man das? Wieso Syntax?.....	217
10.1.1.	Wieso Syntax?.....	217
10.1.2.	Was ist Daten-Management? Braucht man das?.....	218
10.2.	Die Programmierumgebung.....	220
10.2.1.	Die Arbeitsfenster - Übersicht.....	220
10.2.2.	(Enhanced) Program Editor.....	222
10.2.3.	Log.....	223
10.2.4.	Output.....	225

10.2.5.	Results.....	226
10.2.6.	Befehle.....	227
10.3	Daten-Management Einführung und Grundoperationen.....	228
10.3.1.	Grundoperationen auf Programm-Ebene.....	229
10.3.2.	Basisprogrammierungen für das Daten-Management auf Datensatz-Ebene.....	230
10.3.3.	Basisprogrammierungen für das Daten-Management auf der Variablen-Ebene.....	239
10.3.4.	Übungen.....	243
10.4.	Professionelles Daten-Management I Variablenebene.....	244
10.4.1.	Möglichkeiten des Rekodierens von Variablenwerten.....	245
10.4.2.	Möglichkeiten des Berechnens neuer Werten und Variablen....	255
10.4.3.	Zaubern mit Zeichen.....	266
10.5.	Professionelles Datenmanagement II Datensatzebene.....	269
10.5.1.	Aufteilen von Datensätzen.....	269
10.5.2.	Zusammenfügen von Datensätzen (u.a. SET, MERGE, UPDATE).....	273
10.5.3.	Umstrukturieren von Datensätzen.....	277
10.5.4.	Übungen.....	284
10.6.	Datenmanagement III SQL Programmierung.....	285
10.6.1.	Anlegen und Löschen eines Datensatzes.....	286
10.6.2.	Arbeit mit Spalten (Variablen).....	287
10.6.3.	Arbeit mit Zeilen (Fällen, Beobachtungen, Subjekten).....	288
10.6.4.	Erste Arbeit mit Zeilen und Spalten: Subsets und Subqueries...	290
10.6.5.	Übungen.....	291
11.	<b>Deskriptive Statistik.....</b>	<b>292</b>
11.0.	Einführung.....	292
11.1.	Was ist deskriptive Statistik? Sinn und Grenzen.....	293
11.2.	Statistische Grundkonzepte Grundbegriffe und Grundannahmen.....	294
11.2.1.	Theorie und Praxis einer Messung Grundgesamtheit und Stichprobe.....	295
11.2.3.	Variablen Definitionen und Funktionen.....	304
11.2.4.	Übungen.....	315
11.3.	Deskriptive Statistiken (Lage-, Streuungs- und Formmaße). . .	317
11.3.1.	Lagemaße.....	320
11.3.2.	Streuungsmaße.....	331
11.3.3.	Formmaße.....	342
11.4.	Berechnung von Lage-, Streu- und Formmaßen mit SAS.....	344
11.4.1.	SAS Prozeduren für die deskriptive Statistik: Eine Übersicht...	344
11.4.2.	Univariate deskriptive Statistiken: PROC UNIVARIATE.....	346
11.4.3.	Häufigkeitstabellen: PROC FREQ.....	351
11.4.4.	Gruppierte Mittelwerte und mehr: PROC MEANS.....	354

11.4.5.	Statistiken in Tabellenform: PROC TABULATE.....	357
11.4.6.	Übungen.....	365
<b>12.</b>	<b>Ein- und mehrdimensionale Tabellenanalysen mit PROC FREQ....</b>	<b>368</b>
12.0.	Übersicht.....	368
12.1.	Einführung in die Analyse von Kreuztabellen.....	368
12.1.1.	Grundlagen.....	370
12.1.2.	Übersicht: Tabellenstrukturen und Tests.....	377
12.2.	Tabellenstrukturen und Tests.....	381
12.2.1.	Tests für 1 dimensionale Tabellen.....	382
12.2.2.	Tests für 2 dimensionale Tabellen.....	382
12.2.3.	Tests für mehrdimensionale Tabellen.....	394
12.3.	Skalenniveaus und Maße für die Stärke des Zusammenhangs ...	397
12.3.1.	Maße für Intervallniveau.....	400
12.3.2.	Maße für Ordinalniveau.....	400
12.3.3.	Maße für Nominalniveau.....	402
12.4.	Beispiele für die Berechnung von Tabellenstatistiken.....	404
12.5.	Voraussetzungen zur Durchführung einer Tabellenanalyse.....	412
12.6.	Syntax der Prozedur PROC FREQ.....	413
12.6.1.	Statistik.....	413
12.6.2.	Einstellung weiterer Parameter und Optionen.....	416
<b>13.</b>	<b>Einführung in die Inferenzstatistik.....</b>	<b>420</b>
13.1.	Was ist Inferenzstatistik? Sinn und Grenzen.....	420
13.1.1.	Inferenzstatistik als ein System von Modellen und Regeln.....	421
13.1.2.	Die Messtheorie als Brücke zu anderen Wissenschaften.....	422
13.1.3.	Die Bedeutung von Skalenniveaus.....	425
13.1.4.	Sinn und Grenzen der Inferenzstatistik.....	427
13.2.	Einführung in die Theorie des Hypothesentests.....	428
13.2.1.	Hypothesenarten, Sicherheiten, Fehlerarten und Fragerichtungen.....	429
13.2.2.	Beispiel: Ein Käfer im Labyrinth.....	433
13.2.3.	Einfluss des Stichprobenumfangs auf $\alpha$ - und $\beta$ -Fehler.....	435
13.2.4.	Ermittlung des Stichprobenumfangs.....	436
13.3.	Verteilungen als Modelle und zur Prüfung.....	439
13.3.1.	Verteilungen als Modelle für Daten.....	439
13.3.2.	Prüfverteilungen.....	449
13.4.	Einführung in die Praxis des Hypothesentests.....	452
13.4.1.	Einführung: Punkt- und Intervallschätzung.....	452
13.4.2.	Hypothesentests für Punktschätzer.....	453
13.4.3.	Intervallschätzung.....	482
13.5.	Übersicht der dargestellten inferenzstatistischen Verfahren.....	485

<b>14.</b>	<b>Korrelatton und Regression CORR REG.....</b>	<b>489</b>
14.1.	Einführung in die Korrelationsanalyse mit PROC CORR.....	489
14.1.1.	Korrelationskoeffizient r nach Pearson: Maß für linearen Zusammenhang.....	491
14.1.2.	Rangkorrelationskoeffizient R nach Spearman: Maß für monotonen Zusammenhang.....	497
14.1.3.	Die Syntax der Prozedur CORR.....	502
14.2.	Einführung in die Regressionsanalyse mit PROC REG.....	503
14.2.1.	Der Grundgedanke der einfachen linearen Regression.....	504
14.2.2.	Berechnung einer linearen Regression mit PROC REG.....	508
14.2.3.	Erste Regressionsgrafiken mit PROC REG.....	511
14.2.4.	Voraussetzungen für die Berechnung einer einfachen linearen Regression.....	514
14.2.5.	Die Syntax der Prozedur REG.....	516
<b>15.</b>	<b>Varianzanalysen ANOVA GLMNPARIWAY NESTED.....</b>	<b>519</b>
15.1.	Einführung in die Varianzanalyse.....	519
15.2.	Die Berechnung von Varianzanalysen in SAS.....	520
15.3.	PROC ANOVA für balancierte Designs.....	521
15.3.1.	Syntax.....	521
15.3.2.	Beispiel-Analyse: Einfaktorielle-univariate ANOVA.....	522
15.4.	Prozedur GLM für Allgemeine Lineare Modelle.....	526
15.4.1.	Syntax.....	526
15.4.2.	Beispielanalyse: Zweifaktorielle ANOVA.....	527
15.5.	PROC NPARIWAY für nichtparametrische (Rang) Varianzanalysen.....	530
15.5.1.	Syntax.....	530
15.5.2.	Beispiel-Analyse: Nichtparametrische Rangvarianzanalyse....	531
15.6.	Prozedur NESTED für vollständig hierarchisch genestete Designs.....	534
15.7.	Hotelling's P: Ein „t-Test“ für mehrere abhängige Variablen..	537
15.8.	Übungen.....	539
<b>16.</b>	<b>Clusteranalyse CLUSTER FASTCLUS.....</b>	<b>540</b>
16.1.	Einführung in die Clusteranalyse.....	540
16.1.1.	Partitionierende Verfahren.....	541
16.1.2.	Hierarchische Verfahren.....	542
16.2.	Exkurs: Das %DISTNEW Makro.....	544
16.3.	Die Berechnung einer Clusteranalyse in SAS.....	546
16.3.1.	Die partitionierende Clusteranalyse.....	546
16.3.2.	Die hierarchische Clusteranalyse.....	552
16.4.	Anmerkung zu den Voraussetzungen für die Berechnung einer Clusteranalyse.....	570



16.5.	Übungen.....	572
<b>17.</b>	<b>Kovarianzanalyse GLM.....</b>	<b>573</b>
17.1.	Einführung in die Kovarianzanalyse.....	573
17.2.	Berechnung einer Kovarianzanalyse mit der Prozedur GLM....	574
17.3.	Zwei Beispiele: pH-Werte und Phosphorkonzentrationen.....	577
17.3.1.	Zwei Syntaxbeispiele: pH-Werte (einfach), Phosphorkonzentrationen (differenzierend) ....	578
17.3.2.	Output und Erläuterung des pH-Beispiels.....	580
17.3.3.	Output und Erläuterung des Phosphor-Beispiels.....	582
17.4.	Die Syntax der Prozedur GLM.....	586
17.5.	Voraussetzungen einer Kovarianzanalyse und ihre Überprüfung.....	589
17.6.	Übungen.....	593
<b>18.</b>	<b>Faktorenanalyse FACTOR.....</b>	<b>594</b>
18.1.	Einführung in die Faktorenanalyse.....	594
18.1.1.	Grundprinzip.....	596
18.1.2.	Varianten der Faktorenanalyse.....	599
18.1.3.	Kriterien zur Bestimmung der Faktoren: Anzahl und Interpretation.....	602
18.2.	Die Berechnung einer Faktorenanalyse in SAS.....	605
18.2.1.	Syntax der Prozedur FACTOR.....	605
18.2.2.	Beispiel-Analyse mit Syntax, Output und Interpretation.....	609
18.2.3.	Allgemeine Hinweise zur Interpretation.....	620
18.3.	Voraussetzungen für das Berechnen einer Faktorenanalyse.....	621
18.4.	Übungen.....	625
<b>19.</b>	<b>Pfadanalyse CALIS.....</b>	<b>626</b>
19.1.	Einführung in die Pfadanalyse.....	626
19.2.	Die Berechnung einer Pfadanalyse in SAS.....	627
19.2.1.	RAM Model Option.....	628
19.2.2.	Die Anpassungsgüte.....	630
19.2.3.	Die Pfadkoeffizienten.....	631
19.2.4.	Voraussetzungen für das Berechnen einer Pfadanalyse mit PROC CALIS.....	631
19.3.	Syntax.....	633
19.4.	Beispiel-Berechnung.....	637
19.4.1.	Modell.....	637
19.4.2.	Programm.....	638
19.4.3.	Output.....	639
19.5.	Übungen.....	646

- 20. Überlebenszeitanalyse LIFETEST.....647**
  - 20.1. Einführung in die Überlebenszeitanalyse.....647
  - 20.2. Das Grundprinzip der Survivalanalyse.....648
  - 20.3. Zensierte Daten.....650
  - 20.4. Statistische Methoden zur Analyse zensierter Daten.....656
    - 20.4.1. Schätzung von S(t) mit der versicherungsmathematischen Methode.....657
    - 20.4.2. Schätzung von S(t) mit der Kaplan-Meier-Methode.....658
    - 20.4.3. Beispiele ohne und mit Zensierungen.....658
  - 20.5. Der Vergleich der Überlebenswahrscheinlichkeit mehrerer Gruppen.....662
  - 20.6. Die Berechnung einer Survivalanalyse in SAS.....663
    - 20.6.1. Syntax.....664
    - 20.6.2. Beispiel-Berechnung.....668
    - 20.6.3. Voraussetzungen für die Berechnung einer Survivalanalyse. . . 674
  - 20.7. Übungen.....676
- 21. Hauptkomponentenanalyse PRINCOMP.....677**
  - 21.1. Einführung in die Hauptkomponentenanalyse.....677
    - 21.1.1. Gemeinsamkeiten und Unterschiede zur Faktorenanalyse. ....678
    - 21.1.2. Das Prinzip der Hauptkomponentenanalyse.....679
    - 21.1.3. Bestimmung der Anzahl der Komponenten.....680
  - 21.2. Die Berechnung einer Hauptkomponentenanalyse in SAS.....680
    - 21.2.1. Syntax.....680
    - 21.2.2. Beispiel-Berechnung.....683
    - 21.2.3. Grafiken für PROC PRINCOMP.....691
    - 21.2.4. Voraussetzungen für das Berechnen einer Hauptkomponentenanalyse.....692
  - 21.3. Übungen.....694
- 22. Prognose FORECAST.....695**
  - 22.1. Einführung in die Prognose.....695
    - 22.1.1. Fragestellungen und Funktionen.....695
    - 22.1.2. Modelle der Prognose: Trendmodelle.....698
    - 22.1.3. Die Methode der exponentiellen Glättung.....700
  - 22.2. Die Berechnung einer Prognose in SAS.....703
    - 22.2.1. Syntax.....705
    - 22.2.2. Beispiel-Berechnung.....708
      - 22.2.1. Erläuterung des grafischen Outputs.....711
      - 22.2.2. Erläuterung des statistischen Outputs.....712
    - 22.2.3. Voraussetzungen für das Berechnen einer Prognose.....714
  - 22.3. Übungen.....716

<b>23.</b>	<b>Grafiken mit dem SAS System SAS/GRAPH.....</b>	<b>717</b>
23.1.	Einführung und Übersicht.....	717
23.2.	Grundformen der Grafiken.....	720
23.3.	Standardgrafiken mit der SAS/GRAPH Software.....	732
23.3.1.	Balkendiagramme hoch/vertical (2D, 3D).....	733
23.3.4.	Boxplots.....	750
23.3.5.	Histogramme.....	752
23.3.6.	P-P Plots mit PROC CAPABILITY (PPLOT-Statement).....	757
23.3.7.	Prob Plots.....	758
23.3.8.	Q-Q Plots.....	761
23.3.9.	Streu- und Liniendiagramme.....	763
23.3.10.	Ausblick: Weitere SAS Grafiken.....	770
23.3.11.	Übungen.....	781
23.4.	Zaubern mit Annotate und andere Tricks.....	782
23.4.1.	Tunen von SAS Grafiken mit Annotate.....	782
23.4.2.	Andere Möglichkeiten, andere Grafiken Spezielle Optionen, Makros, Annotate, GMAP und GREPLAY.....	795
23.5.	Export und Import von Grafiken und Tabellen (z.B. MSWORD).....	816
23.5.1.	Export von Grafiken.....	816
23.5.2.	Export von Tabellen mit ODS.....	819
<b>IV.</b>	<b>Ausblick und mehr.....</b>	<b>823</b>
<b>24.</b>	<b>„Mercury" Ein Ausblick auf SAS Version 9.....</b>	<b>825</b>
<b>25.</b>	<b>Verzeichnis der Übersichten.....</b>	<b>826</b>
<b>26.</b>	<b>Literatur.....</b>	<b>827</b>
<b>27.</b>	<b>Hinweise zu den Daten.....</b>	<b>831</b>
<b>28.</b>	<b>Kontakt zu SAS.....</b>	<b>833</b>
<b>29.</b>	<b>Ihre Meinung zu diesem Buch.....</b>	<b>834</b>
<b>30.</b>	<b>Stichwortverzeichnis.....</b>	<b>835</b>
<b>31.</b>	<b>Autor.....</b>	<b>853</b>