

Achim Bühl · Peter Zöfel

# Erweiterte Datenanalyse mit SPSS

Statistik und Data Mining

Achim Bühl · Peter Zöfel

Erweiterte Datenanalyse mit SPSS

Achim Bühl · Peter Zöfel

# Erweiterte Datenanalyse mit SPSS

*Statistik und Data Mining*

Springer Fachmedien Wiesbaden GmbH

Bibliografische Information Der Deutschen Bibliothek  
Die Deutsche Bibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie;  
detaillierte bibliografische Daten sind im Internet über <<http://dnb.ddb.de>> abrufbar.

1. Auflage Dezember 2002

Alle Rechte vorbehalten

© Springer Fachmedien Wiesbaden 2002

Ursprünglich erschienen bei Westdeutscher Verlag GmbH, Wiesbaden 2002.

Lektorat: Frank Engelhardt

[www.westdeutscher-verlag.de](http://www.westdeutscher-verlag.de)



Das Werk einschließlich aller seiner Teile ist urheberrechtlich geschützt. Jede Verwertung außerhalb der engen Grenzen des Urheberrechtsgesetzes ist ohne Zustimmung des Verlags unzulässig und strafbar. Das gilt insbesondere für Vervielfältigungen, Übersetzungen, Mikroverfilmungen und die Einspeicherung und Verarbeitung in elektronischen Systemen.

Die Wiedergabe von Gebrauchsnamen, Handelsnamen, Warenbezeichnungen usw. in diesem Werk berechtigt auch ohne besondere Kennzeichnung nicht zu der Annahme, dass solche Namen im Sinne der Warenzeichen- und Markenschutz-Gesetzgebung als frei zu betrachten wären und daher von jedermann benutzt werden dürften.

Umschlaggestaltung: Horst Dieter Bürkle, Darmstadt

Titelbild: Fotostudie Erika Althaus, Marburg

Gedruckt auf säurefreiem und chlorfrei gebleichtem Papier

ISBN 978-3-531-13821-3      ISBN 978-3-322-89603-2 (eBook)

DOI 10.1007/978-3-322-89603-2

<b>Vorwort .....</b>	<b>9</b>
<b>Überblick .....</b>	<b>11</b>
<b>1 Entscheidungsbaum-Analyse .....</b>	<b>13</b>
1.1 Installation des Moduls AnswerTree .....	14
1.2 Einführendes Beispiel aus dem Marketing-Bereich .....	17
1.2.1 Erstellen einer Analysedatei.....	20
1.2.2 Erzeugung und Interpretation eines Baumdiagramms .....	24
1.2.3 Gewinnübersicht.....	27
1.2.4 Verwendung von Scores.....	30
1.2.5 Speichern eines Projekts.....	32
1.3 Vertiefendes Beispiel aus der Politikwissenschaft .....	33
1.3.1 Erstellen einer Analysedatei.....	36
1.3.2 Erzeugung und Interpretation eines Baumdiagramms .....	37
1.3.3 Ausdrucken eines Baumdiagramms.....	38
1.3.4 Darstellungsweisen des Baumfensters .....	39
1.3.5 Risikoubersicht.....	41
1.3.6 Gewinnübersicht.....	42
1.4 Der CHAID-Algorithmus als Aufbaumethode .....	44
1.4.1 Erstellen einer Analysedatei.....	45
1.4.2 Erzeugung und Interpretation eines Baumdiagramms .....	50
1.4.3 Anzeigen von Grafiken .....	55
1.4.4 Gewinnübersicht.....	57
1.5 Der Exhaustive-CHAID-Algorithmus als Aufbaumethode.....	60
1.5.1 Erstellen einer Analysedatei.....	60
1.5.2 Erzeugung und Interpretation eines Baumdiagramms .....	63
1.5.3 Gewinnübersicht.....	65
1.5.4 Gestaltung der Baumstruktur .....	67
1.6 Der C&RT-Algorithmus als Aufbaumethode.....	69
1.6.1 Erstellen einer Analysedatei.....	69
1.6.2 Erzeugung und Interpretation eines Baumdiagramms .....	72
1.6.3 Gewinnübersicht.....	74
1.7 Der Quest-Algorithmus als Aufbaumethode.....	76
1.7.1 Erstellen einer Analysedatei.....	76
1.7.2 Erzeugung und Interpretation eines Baumdiagramms .....	80
1.7.3 Gewinnübersicht.....	81

1.8	Vergleichende Betrachtung der Algorithmen.....	83
<b>2</b>	<b>Korrespondenzanalyse .....</b>	<b>85</b>
2.1	Einfache Korrespondenzanalyse.....	86
2.1.1	Das Prinzip der einfachen Korrespondenzanalyse.....	87
2.1.2	Beispiel einer Produktpositionierung.....	97
2.1.3	Das Seriationsproblem in der Archäologie.....	102
2.2	Multiple Korrespondenzanalyse mit Nominalvariablen.....	105
2.2.1	Einführendes Beispiel aus dem Marketing-Bereich.....	105
2.2.2	Zweites Beispiel: Der Untergang der Titanic.....	113
2.3	Multiple Korrespondenzanalyse mit beliebigen Variablen.....	116
2.3.1	Erstes Beispiel: Alle Variablen numerisch.....	117
2.3.2	Zweites Beispiel: Numerische und nominalskalierte Variablen.....	121
2.4	Kanonische Korrespondenzanalyse.....	127
<b>3</b>	<b>Kategoriale Regression.....</b>	<b>141</b>
3.1	Prinzip der Kategorienquantifikationen.....	141
3.2	Zweites Beispiel: Lernziele.....	148
3.3	Drittes Beispiel: Der Untergang der Titanic.....	153
<b>4</b>	<b>Multidimensionale Skalierung.....</b>	<b>157</b>
4.1	Symmetrische Matrix von Ähnlichkeiten.....	158
4.2	Asymmetrische Matrix von Ähnlichkeiten.....	162
4.3	Ähnlichkeiten aus Daten erstellen.....	163
4.4	Multidimensionale Skalierung und Faktorenanalyse.....	169
<b>5</b>	<b>Conjoint-Analyse .....</b>	<b>173</b>
5.1	Zielsetzung.....	173
5.2	Vorstellung eines Beispiels.....	175
5.3	Erstellung eines orthogonalen Designs.....	175
5.4	Die Ausgabe des orthogonalen Designs.....	180
5.5	Die Prozedur CONJOINT.....	184
<b>6</b>	<b>Pfadanalyse.....</b>	<b>193</b>
6.1	Installation von Amos.....	194
6.2	Einführendes Beispiel: Multiple lineare Regression.....	197
6.3	Zweites Beispiel: Faktorenanalyse.....	207
6.4	Übungsbeispiel.....	209

---

<b>7</b>	<b>Exakte Testmethoden.....</b>	<b>211</b>
7.1	Exakte p-Werte.....	213
7.2	Monte-Carlo-Methode.....	215
7.3	Integration in das Basis-Modul.....	217
7.4	Nichtparametrische exakte Tests.....	222
7.4.1	Vergleich von zwei unabhängigen Stichproben.....	223
7.4.2	Vergleich von zwei abhängigen Stichproben.....	224
7.4.3	Vergleich von mehr als zwei unabhängigen Stichproben.....	226
7.4.4	Vergleich von mehr als zwei abhängigen Stichproben.....	227
7.4.5	Ein-Stichproben-Tests.....	228
7.4.6	Schnelle Berechnung.....	231
7.5	Statistiken für Kreuztabellen.....	231
<b>8</b>	<b>Zeitreihenanalyse.....</b>	<b>239</b>
8.1	Datumsvariablen.....	245
8.2	Grafische Darstellung von Zeitreihen.....	247
8.3	Behandlung von fehlenden Werten.....	251
8.4	Erstellung neuer Zeitreihen.....	252
8.5	Autokorrelationen.....	255
8.6	Kreuzkorrelationen.....	260
8.7	Exponentielles Glätten.....	266
8.8	Autoregression.....	272
8.9	ARIMA-Modelle.....	282
8.9.1	Identifikation des Modells.....	283
8.9.2	Durchrechnung des ARIMA-Modells.....	287
8.10	Saisonale ARIMA-Modelle.....	291
8.10.1	Identifikation des Modells.....	291
8.10.2	Durchrechnung des ARIMA-Modells.....	295
8.11	Saisonale Zerlegung.....	296
8.11.1	Additives Modell.....	297
8.11.2	Multiplikatives Modell.....	300
8.12	Spektralanalyse.....	301
<b>9</b>	<b>Erstellung präsentationsreifer Tabellen.....</b>	<b>309</b>
9.1	Vorstellung eines Auswertungsbeispiels.....	309
9.2	Einfache Tabellen.....	315
9.2.1	Mehrere Zeilenvariablen.....	316
9.2.2	Hinzufügen einer zweiten Dimension (Spaltenvariablen).....	317

9.2.3	Hinzufügen einer dritten Dimension (Tabellenvariablen) .....	320
9.2.4	Verschachtelte und gestapelte Ausgaben.....	321
9.2.5	Prozentangaben .....	327
9.2.6	Gesamtwerte.....	333
9.2.7	Mittelwerte und weitere zusammenfassende Statistiken .....	337
9.2.8	Möglichkeiten der Formatierung .....	340
9.3	Allgemeine Tabellen .....	342
9.3.1	Gestapelte und verschachtelte Variablen.....	343
9.3.2	Zellenstatistiken .....	344
9.3.3	Gesamtwerte.....	346
9.4	Auswertung von Mehrfachantworten.....	346
9.4.1	Dichotome Methode.....	346
9.4.2	Kategoriale Methode.....	349
9.4.3	Kreuztabellen mit Setvariablen .....	350
9.4.4	Prozentangaben bei Setvariablen .....	351
9.5	Häufigkeitstabellen .....	354
9.5.1	Beispiele für Häufigkeitstabellen.....	354
9.5.2	Prozentangaben und Gesamtwerte .....	355
9.5.3	Verwendung von Untergruppen.....	356
<b>10</b>	<b>Berichte und Gruppenwechsel.....</b>	<b>361</b>
10.1	Zeilenweise Berichte.....	361
10.1.1	Erstellen eines einfachen Berichts .....	361
10.1.2	Zweistufiger Gruppenwechsel .....	366
10.1.3	Dreistufiger Gruppenwechsel .....	367
10.1.4	Layout eines Berichts .....	371
10.2	Spaltenweise Berichte .....	376
10.3	Komprimierte Berichtsangabe .....	382
10.4	Übungsaufgaben.....	384
	<b>Inhalt der CD .....</b>	<b>385</b>
	<b>Literaturverzeichnis.....</b>	<b>387</b>
	<b>Stichwortverzeichnis.....</b>	<b>391</b>



---

# Vorwort

Liebe Leserin, lieber Leser,

SPSS ist mit seinen Modulen Base, Advanced Models und Regression Models das weltweit verbreitetste Programmsystem zur statistischen Datenanalyse. Zu diesen klassischen Modulen von SPSS gibt es daher auch umfangreiche deutschsprachige Literatur.

Im Laufe der Zeit sind weitere Module hinzugekommen, wie Categories zur Analyse von Kategorialdaten, Conjoint zur Conjoint-Analyse, Trends zur Zeitreihenanalyse, Exact Tests zur Durchführung von Kreuztabellen-Analysen und nichtparametrischen Tests bei kleinen Fallzahlen oder Tables zur Erstellung präsentationsreifer Tabellen. Ferner gibt es eigenständige Programme wie AnswerTree zur Entscheidungsbaum-Analyse oder Amos zur Pfadanalyse, die von SPSS vertrieben werden und SPSS-Datendateien verarbeiten können.

Zu all diesen Programmen gibt es fast keine deutschsprachige Literatur, was deren Verbreitung nicht eben förderlich sein dürfte. Die Handbücher von SPSS beschreiben zwar ausführlich die Bedienung der Programme, als dikaktische Einführung sind sie aber weniger geeignet.

So soll dieses Buch dazu dienen, das Interesse an diesen Programmen und deren Verständnis durch passende Beispiele zu wecken. Die Datenbeispiele sind auf einer CD beigefügt, damit sie von den Leserinnen und Lesern selbst nachvollzogen und gegebenenfalls modifiziert werden können.

Die Gliederung des Buches erfolgt nicht nach den einzelnen Modulen bzw. Programmen, sondern anhand der betreffenden statistischen Verfahren. Eine entsprechende Zuordnung wird in einem einführenden Überblick gegeben.

Die Verfasser danken der SPSS GmbH Software in München für die wie immer freundliche Unterstützung. Den Leserinnen und Lesern wünschen wir, dass sie die für ihre Zwecke in Frage kommenden Methoden finden und nutzbringend anwenden können. Für den Fall, dass Sie Anregungen oder Kritik äußern möchten oder Nachfragen haben, geben wir Ihnen unsere Email-Adressen bekannt.

*Marburg und Heidelberg  
im Oktober 2002*

*PD Dr. Achim Bühl  
Institut für Soziologie der Universität Heidelberg  
achim.buehl@urz.uni-heidelberg.de*

*Peter Zöfel  
Hochschulrechenzentrum der Universität Marburg  
zoefel@mailier.uni-marburg.de*

---

# Überblick

SPSS gilt als das weltweit verbreitetste Programmsystem zur statistischen Datenanalyse. Der Grundbestandteil ist dabei das Basismodul, das alle Möglichkeiten der Datenerfassung und Datenmodifikation bietet und das die meisten zur Verfügung stehenden Verfahren enthält.

Zur klassischen Grundausstattung von SPSS gehören ferner die beiden Module Advanced Models und Regression Models. Zusammen mit dem Base-Modul umfassen diese Module das, was an statistischen Analysemethoden bei den früheren Großrechner-Versionen zur Verfügung stand.

Daneben gibt es weitere Zusatzmodule bzw. eigenständige Programme zu bestimmten Anwendungen, die ebenfalls von SPSS vertrieben werden. Die meistbenutzten von ihnen werden in diesem Buch beschrieben. Es sind dies AnswerTree, Categories, Conjoint, Amos, Exact Tests, Trends und Tables.

Diese Programme sind in einzelnen Kapiteln dargestellt, wobei als Kapitelüberschrift nicht der Name des Programms, sondern die Bezeichnung der jeweiligen statistischen Analyse verwendet wurde. Categories mit seinen Methoden zur Korrespondenzanalyse, der kategorialen Regression und multidimensionalen Skalierung wurde dabei auf drei Kapitel aufgeteilt. Die Zuordnung der Programme zu den Kapiteln ist der folgenden Übersicht zu entnehmen.

Kapitel	Modul bzw. Programm	Prozedur
1	AnswerTree 3.0	
2.1	Categories	CORRESPONDENCE
2.2	Categories	HOMALS
2.3	Categories	CATPCA
2.4	Categories	OVERALS
3	Categories	CATREG
4	Categories	PROXSCAL
5	Conjoint	
6	Amos 4.0	
7	Exact Tests	
8	Trends	
9	Tables	
10	Base	

AnswerTree ist ein Programm zur Entscheidungsbaum-Analyse und bietet vier verschiedene Methoden, um eine Population anhand von Vorhersagevariablen in verschiedene Teilpopulationen (Segmente) einzuteilen, so dass eine abhängige Variable

in den verschiedenen Segmenten signifikant unterschiedliche Ausprägungen hat. Es ist das Nachfolgeprogramm von CHAID (CHisquared Automatic Interaction Detector); dieser Algorithmus ist nunmehr eine der vier Methoden von AnswerTree. AnswerTree gilt als typisches Data-Mining-Programm.

Mit dem Schlagwort Data Mining wird in jüngster Zeit das Entdecken von Mustern und Zusammenhängen in großen Datenbeständen mit dem Ziel entsprechender Vorhersagen verstanden. Als weitere Methoden des Data Mining gelten u. a. Clusteranalyse, Faktorenanalyse, Regressionsanalyse, Diskriminanzanalyse und neuronale Netze.

Das Modul Categories enthält sechs verschiedene Verfahren zur Analyse kategorialer Daten. Im vorliegenden Buch wurde versucht, eine gewisse Ordnung in diese Verfahren zu bringen, wobei vier Verfahren der Korrespondenzanalyse zugeordnet wurden. Eine Regressionsmethode, bei der in einfacher Weise auch mehrfach abgestufte Nominalvariablen aufgenommen werden können, ist die kategoriale Regression. Schließlich enthält Categories mit der Prozedur PROXSCAL ein Verfahren zur multidimensionalen Skalierung, welches der Prozedur ALSICAL im Base-Modul sehr ähnlich ist.

Die Conjoint-Analyse wird im Marktforschungsbereich eingesetzt und untersucht Eigenschaften von Produkten auf ihre Attraktivität. Dazu müssen befragte Personen ausgewählte Eigenschaftskombinationen anhand ihrer persönlichen Präferenz in eine Rangfolge bringen, woraus sich so genannte Teilnutzenwerte für die einzelnen Kategorien jeder Eigenschaft berechnen lassen.

Amos (Analysis of moment structures) bietet Verfahren zur Lösung linearer Strukturgleichungsmodelle an. Ziel des Programms ist es, komplexe theoretische Zusammenhänge zwischen verschiedenen Variablen in Form einer Pfad- bzw. Kausalanalyse zu prüfen und durch passende Koeffizienten zu beschreiben. Dabei muss der Anwender auf grafischem Weg ein theoretisches Modell in Form eines Pfaddiagramms vorgeben, in das neben direkt beobachteten Variablen auch so genannte latente Variablen einbezogen werden können. Amos wird von SPSS als Nachfolgeprogramm von LISREL vertrieben.

Das Modul Exact Tests dient der Berechnung der exakten Irrtumswahrscheinlichkeit (p-Wert) beim Vorliegen kleiner Fallzahlen beim Chi-Quadrat-Test und den nichtparametrischen Tests. Da diese Tests teilweise sehr rechenintensiv sind, wird als Alternative eine Monte-Carlo-Methode angeboten.

Das Modul Trends enthält verschiedene Verfahren zur Zeitreihenanalyse wie ARIMA-Modelle, exponentielles Glätten, saisonale Zerlegung und Spektralanalyse.

Schließlich werden zwei erweiterte Möglichkeiten der Tabellenerstellung beschrieben. Das Modul Tables dient zur Erstellung präsentationsreifer Tabellen. Verglichen mit den Layouts der gängigen Häufigkeits- und Kreuztabellen stehen zusätzliche Tabellenformate zur Verfügung. Ferner wird die Erstellung von Berichten (Reports) beschrieben. Diese Möglichkeit mit dem wesentlichen Element des Gruppenwechsels wird bereits im Basismodul angeboten.

# Entscheidungsbaum-Analyse

Das Modul AnswerTree teilt eine Population anhand von Vorhersagevariablen (Prediktoren) in verschiedene Teilpopulationen (Segmente) ein, so dass eine abhängige Variable in den verschiedenen Segmenten signifikant unterschiedliche Ausprägungen hat. Die einzelnen Segmente können dann bezüglich des Grades dieser Ausprägungen in eine Reihenfolge gebracht werden. Das wichtigste Hilfsmittel hierzu ist die Darstellung in so genannten Baumdiagrammen. Anhand des Entscheidungsbaumes ist auch erkennbar, welche Vorhersagevariablen einen starken oder auch gar keinen Einfluss auf die abhängige Variable besitzen.

Im Bereich der Markt- und Meinungsforschung wird AnswerTree zum Durchführen von Klassifizierungen und Segmentierungen u. a. in folgenden Bereichen eingesetzt:

- ▶ *Werbeaktionen*: Festgestellt werden soll, ob bestimmte soziale Gruppen etwa bei Postwurfsendungen eine höhere Rücklaufquote aufweisen als andere.
- ▶ *Wahlforschung*: Bestimmt werden kann, welche Merkmale, wie z. B. Alter, Geschlecht, Beruf, Religionszugehörigkeit, sich besonders auf die Wahlentscheidung auswirken.
- ▶ *Produktforschung*: Erforscht wird, welche Variablen, wie z. B. Preis, Verpackung, Merkmale des Kunden, Lage des Ortes, sich verkaufsfördernd auswirken.
- ▶ *Bonitätsprüfung*: Getestet wird auf der Basis gegebener Informationen wie Alter, Einkommen, Besitz, Schulden, ob eine einzelne Person als kreditwürdig gilt.

Der Aufbau des so genannten Entscheidungsbaumes geschieht bei AnswerTree automatisch, so dass ein Teil der Datenanalyse dem Benutzer von der Maschine abgenommen wird. Zur Verfügung stehen vier verschiedene Algorithmen, die wir im Folgenden noch erläutern werden.

Entscheidungsbaum-Analysen werden im Allgemeinen zur Identifikation von Objekten (Personen oder Sachen) eingesetzt, die mit hoher Wahrscheinlichkeit einer bestimmten Klasse angehören (*Segmentierung*), zur Gliederung von Fällen in verschiedene Kategorien, wie z. B. Probanden mit hohem, mittlerem und niedrigem Risiko (*Schichtung*) sowie zur Vorhersage von Zuordnungen bzw. perspektivischen Ereignissen anhand von Vorhersageattributen (*Prognostik*).

Die Leistungsstärken des Moduls AnswerTree sollen im Folgenden anhand einiger Beispiele verdeutlicht werden, wobei nach einer kurzen Beschreibung der Installation des Moduls anhand einer typischen Anwendung aus dem Marketing-Bereich in das Programm eingeführt wird, gefolgt von einem vertiefenden Beispiel aus der Politikwissenschaft. Die vier verschiedenen Algorithmen, die AnswerTree für Seg-

mentierungsanalysen zur Verfügung stellt, werden sodann näher dargestellt und abschließend einer vergleichenden Betrachtung unterzogen.

## 1.1 Installation des Moduls AnswerTree

Wir gehen bei der folgenden Beschreibung davon aus, dass die für das Modul AnswerTree erforderlichen Systemvoraussetzungen erfüllt sind.

- Legen Sie die Installations-CD-ROM von AnswerTree in das CD-ROM-Laufwerk.
- Starten Sie die Installationsroutine durch Aufruf der Datei setup.exe im Ordner AnswerTree.

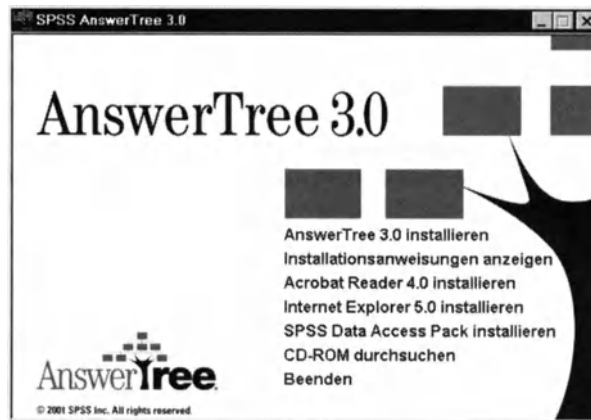


Bild 1.1: Startbildschirm der Installationsroutine

- Klicken Sie auf die Option *AnswerTree installieren*.

SPSS-Setup bereitet den so genannten „InstallShield-Assistenten“ vor, der Sie durch den Installationsvorgang begleitet.

- Klicken Sie auf den Schalter *Weiter*. Auf dem Bildschirm erscheinen die SPSS-Lizenzvereinbarungen für AnswerTree.
- Bestätigen Sie die aufgeführten Bedingungen mit „Ja“.
- Sie werden nunmehr aufgefordert, die erforderlichen Benutzerinformationen einzugeben (Benutzername, Firmenname, Seriennummer). Bestätigen Sie Ihre Eingaben mit *Weiter*.
- Sie erhalten die Möglichkeit, das Zielverzeichnis zu bestimmen, in das AnswerTree installiert werden soll. Die Voreinstellung bezüglich des Zielverzeichnisses (C:\Programme\AnswerTree) sollten Sie mit *Weiter* bestätigen.

Möchten Sie AnswerTree jedoch in einem anderen Verzeichnis installieren, so klicken Sie auf den Schalter *Durchsuchen*. Es öffnet sich die Dialogbox *Ordner auswählen*. Sie können hier das gewünschte Verzeichnis festlegen.

- Haben Sie die Voreinstellungen mit *Weiter* bestätigt, so werden Sie nach der gewünschten Installationsart gefragt. In der Regel dürfte dies die Einzelplatzinstallation sein. Bestätigen Sie daher die Voreinstellung mit *Weiter*.
- Haben Sie AnswerTree mit einem Lizenzcode erhalten, so werden Sie nunmehr aufgefordert, diesen einzutragen. Geben Sie die Codenummer für die Produktlizenz ein. Achten Sie darauf, dass die jeweiligen Ziffernblöcke des Codes genau von einem Leerzeichen getrennt eingegeben werden müssen.
- Bestätigen Sie die Eintragungen mit *Weiter*, und starten Sie den Kopiervorgang.

Sie erhalten zum Schluss eine Bestätigung, dass die Installation von AnswerTree auf Ihrem Computer abgeschlossen ist.

- Klicken Sie auf den Schalter *Fertigstellen*.
- Sie sehen wieder den Startbildschirm der Installationsroutine (siehe Bild 1.1). Klicken Sie hier auf den Schalter *Beenden*.

Wir gehen davon aus, dass Sie AnswerTree in Zukunft häufiger benutzen und es griffbereit halten wollen. Wir schlagen Ihnen daher vor, eine Verknüpfung zu erstellen.

- Klicken Sie mit der rechten Maustaste auf einen freien Bereich des Desktops von Windows. Es öffnet sich das Kontextmenü.
- Im Kontextmenü zeigen Sie auf den Befehl *Neu*.
- Zeigen Sie dann auf die Option *Verknüpfung*. Auf dem Desktop von Windows sehen nun diese Menüs.

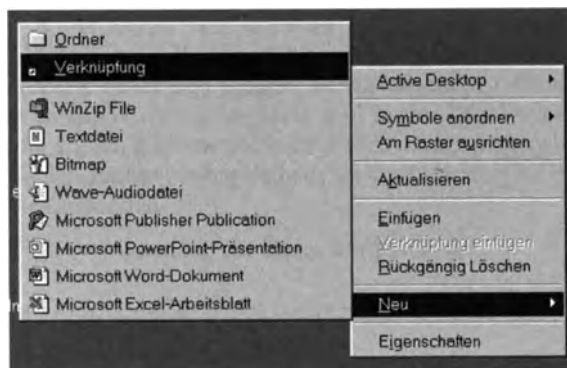


Bild 1.2: Verknüpfung erstellen

- Nachdem Sie mit der linken Maustaste auf *Verknüpfung* geklickt haben, öffnet sich die Dialogbox *Verknüpfung erstellen*.

- Tragen Sie in die Dialogbox *Verknüpfung erstellen* den Pfad und den Namen der ausführenden Datei ein; in der Regel dürfte dies die Pfadbezeichnung „C:\Programme\AnswerTree\atree.exe“ sein, oder nutzen Sie den Schalter *Durchsuchen*, wenn Sie den Pfad nicht genau wissen. Sie gelangen mit Hilfe dieser Schaltfläche in die Verzeichnisstruktur des Rechners und können dort die Datei atree.exe suchen.
- Bestätigen Sie mit *Weiter*.

Es öffnet sich die Dialogbox *Programmbezeichnung wählen*.

- Tragen Sie in die Dialogbox *Programmbezeichnung wählen* „AnswerTree“ ein.
- Bestätigen Sie mit *Fertig stellen*, um das Erstellen der Verknüpfung abzuschließen.

Die Verknüpfung ist nunmehr erstellt. Sie können AnswerTree in Zukunft über zwei verschiedene Wege starten, zum einen von Ihrem Desktop aus mit Hilfe des AnswerTree-Icons (siehe Bild 1.3).



*Bild 1.3: AnswerTree-Icon*

Die andere Möglichkeit besteht darin, AnswerTree aus dem SPSS-Programm aufzurufen, und zwar mit Hilfe der Menüwahl

*Analysieren*  
*AnswerTree*

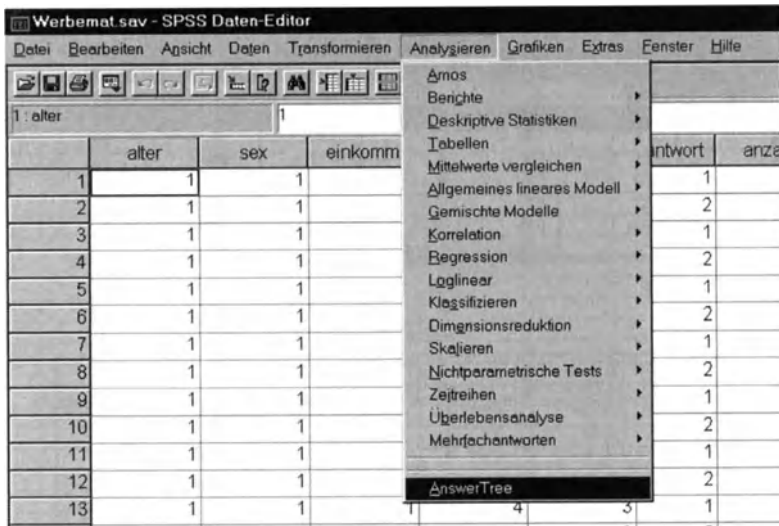


Bild 1.4: Starten von AnswerTree aus SPSS

Die Option *AnswerTree* wurde von der Installationsroutine automatisch in das Pull-Down-Menü *Analisieren* von SPSS hinzugefügt.

## 1.2 Einführendes Beispiel aus dem Marketing-Bereich

Häufige Anwendung findet AnswerTree im Marketing-Bereich, dem daher auch das erste Beispiel entnommen ist. Ein Verlag, der präzise Informationen über seinen potentiellen Kundenkreis hat, vertreibt eine populärwissenschaftliche Zeitschrift und verschickt entsprechendes Werbematerial. Besteller und Nichtbesteller der Zeitschrift werden dabei mit einigen sozialstatistischen Daten in einer Datei festgehalten. Die entsprechenden Angaben sind in der Datei *werbemat.sav* gespeichert. Die Variable *anzahl* gibt an, wie viele Fälle mit den angegebenen Kategorien der anderen Variablen jeweils vorhanden sind. Diese Variable ist demnach als Häufigkeitsvariable zu verwenden.

- Starten Sie SPSS, und laden Sie die Datei *werbemat.sav*.



Bild 1.5: Datei werbemat.sav im Daten-Editor

- Treffen Sie die Menüwahl

*Daten*

*Fälle gewichten...*

- Tragen Sie in der Dialogbox *Fälle gewichten* die Variable *anzahl* als Häufigkeitsvariable ein.
- Erstellen Sie mit Hilfe der Menüwahl

*Analysieren*

*Deskriptive Statistiken*

*Häufigkeiten...*

Häufigkeitstabellen der übrigen Variablen. Diese sind im Folgenden wiedergegeben.

**Antwort**

	Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig				
Besteller	2648	6,1	6,1	6,1
Nichtbesteller	41007	93,9	93,9	100,0
Gesamt	43655	100,0	100,0	

**Alter**

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	<= 25	7946	18,2	18,2	18,2
	26-35	10349	23,7	23,7	41,9
	36-45	4212	9,6	9,6	51,6
	46-55	6093	14,0	14,0	65,5
	56-65	7645	17,5	17,5	83,0
	> 65	7410	17,0	17,0	100,0
Gesamt		43655	100,0	100,0	

**berufliche Taetigkeit**

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	eher geistig	19314	44,2	45,2	45,2
	eher koerperlich	7268	16,6	17,0	62,2
	nicht berufstaetig	16129	36,9	37,8	100,0
	Gesamt	42711	97,8	100,0	
Fehlend	4	944	2,2		
Gesamt		43655	100,0		

**Monatliches Nettoeinkommen**

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	< 2000 DM	3573	8,2	34,0	34,0
	2000-3499 DM	4040	9,3	38,5	72,5
	3500-4999 DM	1845	4,2	17,6	90,1
	>= 5000 DM	1041	2,4	9,9	100,0
	Gesamt	10499	24,0	100,0	
Fehlend	5	33156	76,0		
Gesamt		43655	100,0		

**Schulbildung**

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	Hauptschule	21734	49,8	50,1	50,1
	Mittlere Reife	12743	29,2	29,4	79,4
	Abitur	8940	20,5	20,6	100,0
	Gesamt	43417	99,5	100,0	
Fehlend	4	238	,5		
Gesamt		43655	100,0		

**Geschlecht**

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	weiblich	16440	37,7	37,7	37,7
	männlich	27215	62,3	62,3	100,0
Gesamt		43655	100,0	100,0	

Insgesamt wurden also Werbematerialien an 43655 Personen versandt. Die Variable antwort gibt an, dass hiervon 2648 Personen die Zeitschrift abonniert haben, was

6,1 % entspricht. Von größtem Interesse für den Verlag ist es, diejenigen Personengruppen herauszufinden, bei denen die Erfolgsquote am höchsten ist.

Dies ist ein typisches Anwendungsbeispiel für AnswerTree. Dabei ist die Variable *antwort* als abhängige Variable zu betrachten, wobei anhand der Kategorien der Prediktoren (Vorhersagevariablen) die Gesamtpopulation so in verschiedene Teilpopulationen (Segmente) zu unterteilen ist, dass in diesen Segmenten die Erfolgsquote (also die prozentuale Anzahl der Besteller) signifikant unterschiedlich ist. Die Teilpopulationen mit der höchsten Erfolgsquote sollten vom Verlag bei künftigem Mailing in besonderer Weise berücksichtigt werden.

Die von AnswerTree vorgenommene Segmentierung wird in Form eines Baumdiagramms angezeigt. Zusammenfassende Ergebnisse werden in so genannten Risikoübersichten dargestellt.

### 1.2.1 Erstellen einer Analysedatei

- Um das vorgestellte Beispiel mit AnswerTree zu analysieren, verlassen Sie SPSS und klicken Sie doppelt auf das AnswerTree-Icon.
- Bestätigen Sie die Dialogbox *Login beim Server* mit *OK*.

Es öffnet sich die Start-Dialogbox von AnswerTree.

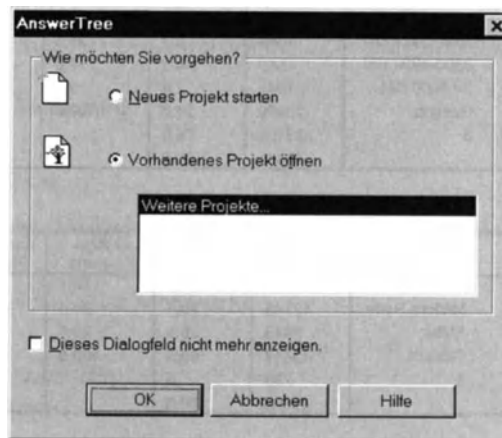


Bild 1.6: Start-Dialogbox

Sie können an dieser Stelle entscheiden, ob Sie ein neues Projekt starten oder ein vorhandenes Projekt öffnen wollen. Schließlich können Sie noch festlegen, ob das Start-Dialogfeld auch in Zukunft angezeigt werden soll. Wir empfehlen Ihnen, diese Option zu deaktivieren, da das Start-Dialogfeld redundant ist. Wurde diese Option auf Ihrem Rechner bereits deaktiviert, so sehen Sie das Start-Dialogfeld nicht. Wählen Sie in diesem Fall aus dem Menü die Option

*Datei**Neues Projekt...*

Im vorliegenden Fall ist die Start-Dialogbox noch aktiviert.

- Wählen Sie die Option *Neues Projekt starten*.

Es öffnet sich die Dialogbox *Datenquelle auswählen*.



Bild 1.7: Dialogbox *Datenquelle auswählen*

- Aktivieren Sie die Option *SPSS-Datendatei*, und bestätigen Sie mit *OK*.
- Es öffnet sich die Dialogbox *Datei öffnen*. Suchen Sie die Datei *werbemat.sav*, und bestätigen Sie anschließend durch Klicken auf den Schalter *Öffnen*.

Sie erblicken nunmehr den Baum-Assistenten, der Sie nach der Methode fragt, die Sie verwenden wollen.