SPSS 16.0 Schneller Einstieg

Weitere Informationen zu SPSS®-Software-Produkten finden Sie auf unserer Website unter der Adresse *http://www.spss.com* oder wenden Sie sich an

SPSS Inc.
233 South Wacker Drive, 11th Floor
Chicago, IL 60606-6412, USA
Tel.: (312) 651-3000
Fax: (312) 651-3668

SPSS ist eine eingetragene Marke, und weitere Produktnamen sind Marken der SPSS Inc. für Computerprogramme von SPSS Inc. Die Herstellung oder Verbreitung von Materialien, die diese Programme beschreiben, ist ohne die schriftliche Erlaubnis des Eigentümers der Marke und der Lizenzrechte der Software und der Copyrights der veröffentlichten Materialien verboten.

Die SOFTWARE und die Dokumentation werden mit BESCHRÄNKTEN RECHTEN zur Verfügung gestellt. Verwendung, Vervielfältigung und Veröffentlichung durch die Regierung unterliegen den Beschränkungen in Unterabschnitt (c)(1)(ii) von The Rights in Technical Data and Computer Software unter 52.227-7013. Vertragspartner/Hersteller ist SPSS Inc., 233 South Wacker Drive, 11th Floor, Chicago, IL 60606-6412.
Patentnr. 7.023.453

Allgemeiner Hinweis: Andere in diesem Dokument verwendete Produktnamen werden nur zu Identifikationszwecken genannt und können Marken der entsprechenden Unternehmen sein.

Windows ist eine eingetragene Marke der Microsoft Corporation.

Apple, Mac und das Mac-Logo sind Marken von Apple Computer, Inc., die in den USA und in anderen Ländern eingetragen sind.

Dieses Produkt verwendet WinWrap Basic, Copyright 1993–2007, Polar Engineering and Consulting, *http://www.winwrap.com*.

SPSS 16.0 Schneller Einstieg
Copyright © 2007 SPSS Inc.
Alle Rechte vorbehalten.
Gedruckt in Irland.

Ohne schriftliche Erlaubnis der SPSS GmbH Software darf kein Teil dieses Handbuchs für irgendwelche Zwecke oder in irgendeiner Form mit irgendwelchen Mitteln, elektronisch oder mechanisch, mittels Fotokopie, durch Aufzeichnung oder durch andere Informationsspeicherungssysteme reproduziert werden.

ISBN-13: 978-1-56827-825-4
ISBN-10: 1-56827-825-X

1 2 3 4 5 6 7 8 9 0 10 09 08 07

Vorwort

Das Handbuch *SPSS 16.0 Schneller Einstieg* enthält Lerneinheiten, in denen Sie die verschiedenen Komponenten von SPSS kennen lernen. Dieses Handbuch ist für alle Betriebssystemversionen der Software gleichermaßen geeignet: Windows, Macintosh und Linux. Sie können die Lerneinheiten nacheinander durcharbeiten oder gezielt die Themen auswählen, zu denen Sie weitere Informationen wünschen. Sie können das vorliegende Handbuch zusammen mit dem Online-Lernprogramm von SPSS Base 16.0 einsetzen. Sie müssen das Online-Lernprogramm jedoch nicht verwenden; die hier enthaltenen Lerneinheiten lassen sich auch separat nutzen.

SPSS 16.0

SPSS 16.0 ist ein umfassendes System zum Analysieren von Daten. Mit SPSS können Sie Daten aus nahezu allen Dateitypen entnehmen und aus ihnen Berichte in Tabellenform, Diagramme sowie grafische Darstellungen von Verteilungen und Trends, deskriptive Statistiken und komplexe statistische Analysen erstellen.

SPSS macht die statistische Analyse auch unerfahrenen Benutzern zugänglich und erleichtert erfahrenen Benutzern die Arbeit. Aufgrund der im Programm vorhandenen Menüs und Dialogfelder können selbst komplexe Analysen durchgeführt werden, ohne dass Befehlssyntax eingegeben werden muss. Der Daten-Editor läßt sich wie eine unkomplizierte und effiziente Tabellenkalkulation zur Eingabe von Daten und zum Durchsuchen der Arbeitsdatei verwenden.

Ressourcen im Internet

Die Website von SPSS (*http://www.spss.com*) enthält Antworten auf häufig gefragte Fragen zur Installation von SPSS und der Arbeit mit dem Programm, Datendateien und andere nützliche Informationen.

Außerdem steht das SPSS-Diskussionsforum im USENET (keine offizielle USENET-Gruppe von SPSS) jedem an SPSS-Produkten Interessierten offen. Die USENET-Adresse lautet *comp.soft-sys.stat.spss*. In diesem Forum werden Computer-, Statistik- und Bedienungsfragen diskutiert, die sich auf SPSS-Software beziehen.

Sie können auch eine E-Mail-Liste abonnieren, die über ein Gateway an die USENET-Gruppe weitergeleitet wird. Wenn Sie diese Liste abonnieren möchten, senden Sie eine E-Mail-Nachricht an *listserv@listserv.uga.edu*. Die E-Mail-Nachricht muss den folgenden Text enthalten: subscribe SPSSX-L Vorname Nachname. Sie können dann Nachrichten an die Liste senden, indem Sie Ihre E-Mail wie folgt adressieren: *listserv@listserv.uga.edu*.

Weitere Veröffentlichungen

Weitere Informationen über die Funktionen von SPSS Base 16.0 und die Arbeit mit diesem Programm finden Sie im *Benutzerhandbuch zu SPSS Base 16.0*, das auch Informationen über Standarddiagramme enthält. Beispiele für statistische Prozeduren in SPSS Base 16.0 finden Sie im Hilfesystem, das mit der Software installiert wird. Algorithmen für die statistischen Prozeduren sind auf der Produkt-CD-ROM verfügbar.

Den Menüs und Dialogfeldern von SPSS liegt eine Befehlssprache zugrunde. Auf einige fortgeschrittene Funktionen des System kann nur mithilfe der Befehlssyntax zugegriffen werden. (Diese Funktionen sind in der Studentenversion nicht verfügbar.) Die vollständige Befehlssyntax wird unter *SPSS 16.0 Command Syntax Reference* beschrieben. Dieses Dokument ist im PDF-Format im Menü "Hilfe" verfügbar.

Weitere Exemplare von Handbüchern für SPSS-Produkte können direkt bei SPSS Inc. bestellt werden. Wenden Sie sich bei telefonischen Bestellungen in den USA und Kanada unter 800-543-2185 direkt an SPSS Inc. Wenden Sie sich bei telefonischen Bestellungen außerhalb von Nordamerika an Ihr regionales SPSS-Büro. Die entsprechenden Kontaktinformationen finden Sie auf der SPSS-Website unter *http://www.spss.com/worldwide*.

Das Handbuch *SPSS Statistical Procedures Companion* von Marija Norušis wurde von Prentice Hall veröffentlicht. Es enthält einen Überblick über die Prozeduren in SPSS Base sowie zur logistischen Regression und zu allgemeinen linearen Modellen. Das Handbuch *SPSS Advanced Statistical Procedures Companion* wurde ebenfalls von Prentice Hall veröffentlicht. Es enthält einen Überblick über die Prozeduren in den Modulen "Advanced" und "Regression".

SPSS -Optionen

Die folgenden Optionen sind als Erweiterungsmodule der Vollversion (nicht der Studentenversion) von SPSS Base erhältlich:

SPSS Regression Models™ bietet Verfahren zur Datenanalyse, die über herkömmliche lineare Statistikmodelle hinausgehen. Es beinhaltet Prozeduren für Probit-Analyse, logistische Regression, Gewichtungsschätzungen, zweistufige Kleinste-Quadrate-Regression und allgemeine nichtlineare Regression.

SPSS Advanced Models™ umfasst vor allem Verfahren, die in der fortgeschrittenen experimentellen und biomedizinischen Forschung Anwendung finden. Dies beinhaltet beispielsweise Prozeduren für allgemeine lineare Modelle (GLM), lineare gemischte Modelle, Varianzkomponentenanalyse, loglineare Analysen, ordinale Regression, versicherungsstatistische Sterbetafeln, die Überlebensanalyse nach Kaplan-Meyer sowie die grundlegende und erweiterte Cox-Regression.

SPSS Tables™ dient dem Erstellen einer großen Auswahl von Tabellenberichten in Präsentationsqualität. Mit dieser Option können beispielsweise komplexe Stub- und Banner-Tabellen erstellt und Daten von Mehrfachantworten angezeigt werden.

SPSS Trends™ bietet Funktionen zum Ausführen umfangreicher Prognosen sowie Zeitreihenanalysen mit Modellen für mehrfache Kurvenanpassung, mit Glättungsmodellen und Methoden zum Schätzen autoregressiver Funktionen.

SPSS Categories® bietet Funktionen zum Ausführen und Optimieren von Skalierungsprozeduren, u. a. Korrespondenzanalysen.

SPSS Conjoint™ bietet eine realistische Methode zum Messen, wie sich einzelne Produktmerkmale auf die Präferenzen von Konsumenten und Bürgern auswirken. Mit SPSS Conjoint können Sie einfach messen, welche Auswirkungen es hat, wenn einzelne Produktmerkmale im Kontext einer Gruppe von Produktmerkmalen gegeneinander abgewägt werden, genau wie Konsumenten dies bei Kaufentscheidungen tun.

SPSS Exact Tests™ berechnet exakte *P*-Werte für statistische Tests bei einer kleinen Anzahl oder sehr ungleichmäßig verteilten Stichproben, bei denen herkömmliche Tests nur ungenaue Ergebnisse liefern. Diese Option ist nur unter Windows-Betriebssystemen verfügbar.

SPSS Missing Value Analysis™ dient zum Beschreiben von Mustern bei fehlenden Daten, zum Schätzen von Mittelwerten und anderen statistischen Größen sowie zum Ersetzen von Werten für fehlende Beobachtungen.

SPSS Complex Samples™ ermöglicht Experten auf den Gebieten Umfragen, Marktforschung, Gesundheitswesen und Öffentliche Meinung sowie Sozialwissenschaftlern, die das Verfahren der Stichprobenumfrage verwenden, ihre Stichprobenpläne mit komplexen Stichproben in die Datenanalyse zu integrieren.

SPSS Classification Tree™ dient zur Erstellung eines baumbasierten Klassifizierungsmodells. Die Fälle werden in Gruppen klassifiziert oder es werden Werte für eine abhängige Variable (Zielvariable) auf der Grundlage der Werte von unabhängigen Variablen (Einflussvariablen) vorhergesagt. Die Prozedur umfasst Validierungswerkzeuge für die explorative und die bestätigende Klassifikationsanalyse.

Mit **SPSS Data Preparation™** erhalten Sie rasch eine visuelle Ansicht Ihrer Daten. Damit verfügen Sie über die Möglichkeit, Validierungsregeln anzuwenden, mit denen Sie ungültige Datenwerte identifizieren können. Sie können Regeln erstellen, mit denen Werte außerhalb des Bereichs, fehlende Werte oder leere Werte gekennzeichnet werden. Sie können außerdem Variablen speichern, mit denen individuelle Regelverletzungen sowie die Gesamtanzahl von Regelverletzungen je Fall aufgezeichnet werden. Im Lieferumfang des Programms befindet sich ein Satz von vordefinierten Regeln, die Sie kopieren und bearbeiten können.

SPSS Neural Networks™ dient zur Unterstützung von Geschäftsentscheidungen durch die Vorhersage der Nachfrage für ein Produkt als Funktion von Preis und anderen Variablen oder durch die Kategorisierung der Kunden auf der Grundlage ihrer Kaufgewohnheiten und demografischen Merkmale. Neuronale Netzwerke sind Modellierungswerkzeuge für nichtlineare Daten. Sie können zur Modellierung komplexer Beziehungen zwischen Eingaben und Ausgaben und zum Auffinden von Mustern in Daten verwendet werden.

Amos™ (**a**nalysis of **mo**ment **s**tructures) verwendet Modellierung von Strukturgleichungen, um konzeptuelle Modelle zu bestätigen und zu erklären, die auf Einstellungen, Wahrnehmungen und anderen Faktoren beruhen, die bestimmten Verhaltensweisen zugrunde liegen.

Zur Produktgruppe von SPSS gehören außerdem Anwendungen für Dateneingabe, Textanalyse, Klassifikation, neuronale Netzwerke und Dienstleistungen für Unternehmen im Bereich der Prognose.

Ausbildungsseminare

SPSS Inc. bietet öffentliche und unternehmensinterne Seminare für SPSS an. Alle Seminare beinhalten auch praktische Übungen. SPSS-Seminare werden regelmäßig in größeren Städten in den USA und Europa angeboten. Wenn Sie weitere Informationen zu diesen Schulungen wünschen, wenden Sie sich an Ihr regionales SPSS-Büro, das Sie auf der SPSS-Website unter *http://www.spss.com/worldwide* finden.

Technischer Support

Kunden von SPSS mit Wartungsvertrag können den Technischen Support von SPSS in Anspruch nehmen. (Wenn Sie die Studentenversion erworben haben, lesen Sie bitte den entsprechenden Abschnitt über den Technischen Support für die Studentenversion. Für weitere Informationen siehe Technischer Support für Studenten auf S. vii.) Kunden können sich an den Technischen Support wenden, wenn sie Hilfe bei der Arbeit mit SPSS-Produkten oder bei der Installation in einer der unterstützten Hardware-Umgebungen benötigen. Informationen über den Technischen Support finden Sie auf der Website von SPSS unter *http://www.spss.com* oder wenden Sie sich an Ihr regionales SPSS-Büro, das Sie auf der SPSS-Website unter *http://www.spss.com/worldwide* finden. Bei einem Anruf werden Sie nach Ihrem Namen, dem Namen Ihrer Organisation und Ihrer Seriennummer gefragt.

Kundenmeinungen

Ihre Meinung ist uns wichtig. Teilen Sie uns bitte Ihre Erfahrungen mit SPSS-Produkten mit. Insbesondere haben wir Interesse an neuen, interessanten Anwendungsgebieten von SPSS. Senden Sie uns eine E-Mail an *suggest@spss.com* oder schreiben Sie an: SPSS Inc., Attn: Director of Product Planning, 233 South Wacker Drive, 11th Floor, Chicago IL 60606-6412.

SPSS 16.0 für Windows – Studentenversion

SPSS 16.0 für Windows – Studentenversion ist eine eingeschränkte, aber dennoch leistungsstarke Version von SPSS Base 16.0.

Leistungsspektrum

Die Studentenversion enthält alle wichtigen Werkzeuge für die Datenanalyse, die in der Vollversion von SPSS Base enthalten sind. Dazu gehören:

- Dateneditor in Tabellenkalkulationsform zur Eingabe, Bearbeitung und Anzeige von Datendateien.
- Statistische Prozeduren, unter anderem *T*-Tests, Varianzanalyse und Kreuztabellen.
- Interaktive Grafiken, mit denen Sie Diagrammelemente und Variablen dynamisch ändern bzw. hinzufügen können. Die Änderungen werden sofort nach ihrer Festlegung angezeigt.
- Hochauflösende Standardgrafiken für eine breite Palette von Diagrammen und Tabellen zu Analyse- und Präsentationszwecken.

Einschränkungen

Die Studentenversion ist für den Unterricht gedacht und darf nur von Schülern/Studenten und Lehrpersonal zu Lehrzwecken verwendet werden. Die Studentenversion enthält nicht alle Funktionen von SPSS Base 16.0. Für SPSS 16.0 für Windows – Studentenversion gelten folgende Einschränkungen:

- Datendateien dürfen nicht mehr als 50 Variablen enthalten.
- Datendateien können nicht mehr als 1.500 Fälle enthalten. SPSS-Erweiterungsmodule (z. B. Regression Models oder Advanced Models) können nicht mit der Studentenversion verwendet werden.
- Die SPSS-Befehlssyntax ist für den Benutzer nicht verfügbar. Das bedeutet, dass Analysen nicht wiederholt werden können, indem eine Reihe von Befehlen in einer Syntax- bzw. einer "Job"-Datei gespeichert wird, wie dies in der Vollversion von SPSS möglich ist.
- Skripterstellung und Automatisierung stehen nicht zur Verfügung. Sie können also keine Skripts zur Automatisierung häufig wiederholter Aufgaben verwenden, wie in der Vollversion von SPSS der Fall.

Technischer Support für Studenten

Studenten sollten Technischen Support von Ihrem Lehrpersonal erhalten oder von Support-Mitarbeitern vor Ort, die ihnen von Ihrem Lehrpersonal genannt werden. Technischer Support von SPSS für die Studentenversion von SPSS 16.0 ist *nur für Lehrpersonal verfügbar, das das System zu Unterrichtszwecken einsetzt*.

Bevor Sie Ihren Lehrer um Unterstützung bitten, notieren Sie sich bitte die untenstehenden Informationen. Ohne diese Informationen kann Ihr Lehrer Ihnen nicht helfen:

- Den Typ des verwendeten Computers sowie den verfügbaren Arbeitsspeicher (RAM) und den freien Festplattenspeicher.
- Das Betriebssystem des Computers.
- Eine anschauliche Beschreibung der Vorgänge und Ihrer Tätigkeit beim Auftreten des Problems. Versuchen Sie nach Möglichkeit das Problem mit einer der im Lieferumfang des Programms enthaltenen Dateien mit Beispieldaten zu reproduzieren.
- Den genauen Wortlaut aller Fehler- oder Warnmeldungen, die auf dem Bildschirm angezeigt wurden.
- Die von Ihnen selbst unternommenen Versuche, das Problem zu beheben.

Technischer Support für Lehrpersonal

Lehrpersonal, das die Studentenversion für Unterrichtszwecke verwendet, kann sich an den Technischen Support von SPSS wenden. In den Vereinigten Staaten und in Kanada können Sie den Technischen Support von SPSS telefonisch unter 312651-3410 oder per E-Mail unter *support@spss.com* erreichen. Geben Sie dabei Ihren Namen, Ihre Position und den Namen Ihrer Bildungseinrichtung an.

Lehrpersonal außerhalb der Vereinigten Staaten und von Kanada wendet sich an ihr regionales SPSS-Büro, das auf der SPSS-Website unter *http://www.spss.com/worldwide* aufgeführt ist.

Inhalt

1 Einführung 1

Beispieldateien . . . 1
Öffnen einer Datendatei . . . 1
Durchführen einer Analyse . . . 3
Anzeigen von Ergebnissen . . . 7
Erstellen von Diagrammen . . . 8

2 Einlesen von Daten 11

Grundlegende Struktur einer Datendatei im SPSS-Format . . . 11
Einlesen einer Datendatei im SPSS-Format . . . 11
Einlesen von Daten aus Tabellenkalkulationsdateien . . . 12
Einlesen von Daten aus einer Datenbank . . . 14
Einlesen von Daten aus einer Textdatei . . . 19
Speichern von Daten . . . 26

3 Verwenden des Daten-Editors 28

Eingeben numerischer Daten . . . 28
Eingeben von String-Daten . . . 31
Definieren von Daten . . . 33
- Hinzufügen von Variablenlabels . . . 33
- Ändern des Datentyps und Formats von Variablen . . . 34
- Hinzufügen von Wertelabels für numerische Variablen . . . 34
- Hinzufügen von Wertelabels für String-Variablen . . . 36
- Verwenden von Wertelabels bei der Dateneingabe . . . 37
- Umgang mit fehlenden Daten . . . 38
- Fehlende Werte für eine numerische Variable . . . 39
- Fehlende Werte für eine String-Variable . . . 40
- Kopieren und Einfügen von Variablenattributen . . . 41
- Definieren von Variableneigenschaften für kategoriale Variablen . . . 45

4 Arbeiten mit mehreren Datenquellen 51

Grundsätzlicher Umgang mit mehreren Datenquellen 51
Arbeiten mit mehreren Daten-Sets in der Befehlssyntax 52
Kopieren und Einfügen von Informationen zwischen Daten-Sets 53
Umbenennen von Daten-Sets 53
Unterdrücken der Anzeige mehrerer Daten-Sets. 54

5 Untersuchen von Auswertungsstatistiken für einzelne Variablen 55

Messniveau 55
Auswertungsmaße für kategoriale Daten 56
Diagramme für kategoriale Daten 57
Auswertungsmaße für metrische Variablen 59
Histogramme für metrische Variablen 61

6 Erstellen und Bearbeiten von Diagrammen 63

Grundlagen der Diagrammerstellung 63
Verwenden der Galerie für die Diagrammerstellung 64
Definieren von Variablen und Statistiken 65
Hinzufügen von Text. 68
Erstellen des Diagramms 69
Grundlagen der Diagrammbearbeitung 69
Auswählen von Diagrammelementen. 70
Verwendung des Fensters “Eigenschaften” 71
Ändern der Balkenfarben. 71
Formatieren der Zahlen in Teilstrichbeschriftungen 73
Bearbeiten von Text in einem Diagramm 75
Anzeigen von Datenbeschriftungen 76
Verwendung von Vorlagen 77
Optionen zum Definieren von Diagrammen. 82

7 Arbeiten mit Ausgaben 86

Arbeiten mit dem Viewer 86

Verwenden des Pivot-Tabellen-Editors 88
Aufrufen von Definitionen für statistische Begriffe in der Ausgabe 88
Pivot-Tabellen 89
Erstellen und Anzeigen von Schichten 91
Bearbeiten von Tabellen 92
Ausblenden von Zeilen und Spalten 93
Ändern der Anzeigeformate für Daten 94
Tabellenvorlagen 96
Verwenden von vordefinierten Formaten 96
Anpassen von Tabellenvorlagen 97
Ändern der Standard-Tabellenformate 99
Ändern der anfänglichen Einstellungen für die Anzeige 100
Anzeige von Variablen- und Wertelabels 102
Verwenden von Ergebnissen in anderen Anwendungen 103
Einfügen von Ergebnissen als Tabellen in Word 104
Einfügen von Ergebnissen als Text 105
Exportieren der Ergebnisse in Microsoft Word-, PowerPoint- und Excel-Dateien 107
Exportieren von Ergebnissen als PDF 115
Exportieren von Ergebnissen als HTML 118

8 Arbeiten mit Syntax 120

Übernehmen von Befehlssyntax 120
Bearbeiten von Befehlssyntax 122
Öffnen und Ausführen einer Syntaxdatei 123

9 Ändern von Datenwerten 124

Erstellen einer kategorialen Variablen aus einer metrischen Variablen 124
Berechnen von neuen Variablen 130
Verwenden von Funktionen in Ausdrücken 132
Verwendung bedingter Ausdrücke 134
Arbeiten mit Datumsangaben und Uhrzeiten 135
Berechnen des Zeitabstands zwischen zwei Datumsangaben 136
Hinzufügen einer Dauer zu einem Datum 139

10 Sortieren und Auswählen von Daten 141

Sortieren von Daten 141
Verarbeitung von aufgeteilten Dateien 142
Sortieren von Fällen für die Verarbeitung von aufgeteilten Dateien. 144
Aktivieren und Deaktivieren der Verarbeitung von aufgeteilten Dateien 144
Auswählen von Teilmengen von Fällen 144
Auswählen von Teilmengen von Fällen anhand eines bedingten Ausdrucks 145
Auswählen einer Zufallsstichprobe aus den Fällen. 146
Auswählen eines Zeit- oder Fallbereichs 147
Behandlung von nicht ausgewählten Fällen 148
Status der Fallauswahl 148

11 Weitere statistische Prozeduren 150

Zusammenfassen von Daten. 150
Explorative Datenanalyse 150
Weitere Informationen zum Zusammenfassen von Daten 151
Vergleichen von Mittelwerten. 152
Mittelwerte 152
T-Test bei gepaarten Stichproben 154
Weitere Informationen zum Vergleichen von Mittelwerten 155
ANOVA-Modelle 156
Univariate Varianzanalyse 156
Korrelation zwischen Variablen 157
Bivariate Korrelationen 157
Partielle Korrelationen. 157
Regressionsanalyse 158
Lineare Regression 158
Nichtparametrische Tests. 160
Chi-Quadrat. 160

Anhang

A Beispieldateien *163*

Index *175*

Kapitel

1

Einführung

Dieses Handbuch enthält eine Reihe von Lernprogrammen, die Sie befähigen sollen, nützliche Analysen an Ihren Daten durchzuführen. Sie können die Lerneinheiten nacheinander durcharbeiten oder gezielt die Themen auswählen, zu denen Sie weitere Informationen wünschen.

Dieses Kapitel enthält eine Einführung in die Grundfunktionen und eine Vorführung einer typischen Sitzung. Sie werden eine zuvor definierte Datendatei im SPSS-Format abrufen und anschließend eine einfache statistische Zusammenfassung und ein Diagramm erzeugen.

Viele der in diesem Kapitel angesprochenen Themen werden in den folgenden Kapiteln dann ausführlicher erläutert. An dieser Stelle hoffen wir, Ihnen die Grundlagen für ein tieferes Verständnis späterer Lernprogramme vermitteln zu können.

Beispieldateien

Für die meisten hier vorgestellten Beispiele wird die Datendatei *demo.sav* verwendet. Bei dieser Datendatei handelt es sich um eine fiktive Befragung von mehreren tausend Personen, die grundlegende demografische Daten und Verbraucherinformationen enthält.

Die zusammen mit dem Produkt installierten Beispieldateien finden Sie im Unterverzeichnis *Samples* des Installationsverzeichnisses.

Öffnen einer Datendatei

So öffnen Sie eine Datendatei:

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Datei
  Öffnen
    Daten...

Wahlweise können Sie auch auf der Symbolleiste auf die Schaltfläche “Datei öffnen” klicken.

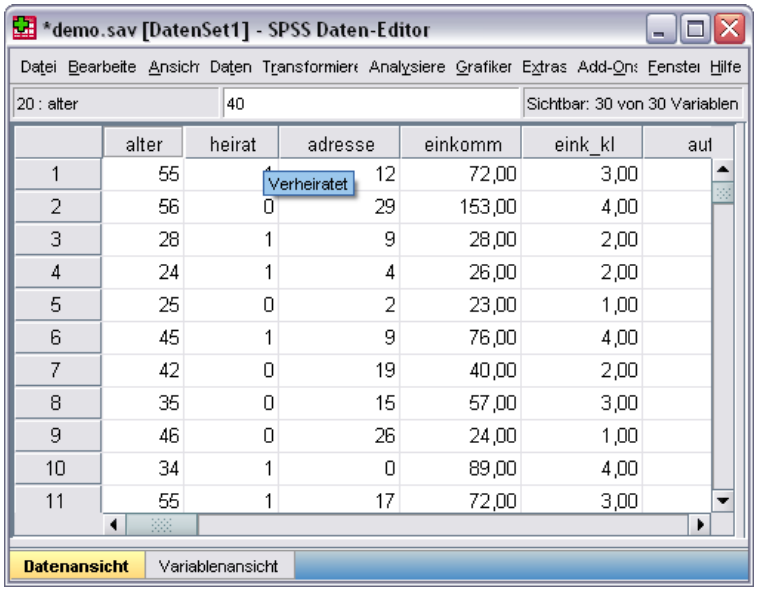
Abbildung 1-1
Schaltfläche “Datei öffnen” auf der Symbolleiste

Ein Dialogfeld zum Öffnen von Dateien wird angezeigt.

In der Standardeinstellung werden Datendateien im SPSS-Format (Erweiterung *.sav*) angezeigt.

In diesem Beispiel wird die Datendatei *demo.sav* verwendet.

Abbildung 1-2
Datei "demo.sav" im Daten-Editor.



	alter	heirat	adresse	einkomm	eink_kl	aut
1	55	1	12	72,00	3,00	
2	56	0	29	153,00	4,00	
3	28	1	9	28,00	2,00	
4	24	1	4	26,00	2,00	
5	25	0	2	23,00	1,00	
6	45	1	9	76,00	4,00	
7	42	0	19	40,00	2,00	
8	35	0	15	57,00	3,00	
9	46	0	26	24,00	1,00	
10	34	1	0	89,00	4,00	
11	55	1	17	72,00	3,00	

Die Datendatei wird im Daten-Editor angezeigt. Wenn Sie im Daten-Editor mit dem Mauszeiger auf einen Variablennamen (die Spaltenüberschrift) zeigen, wird ein ausführliches Variablenlabel angezeigt (sofern für diese Variable ein Label definiert wurde).

In der Standardeinstellung werden die tatsächlichen Datenwerte angezeigt. So zeigen Sie Labels an:

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

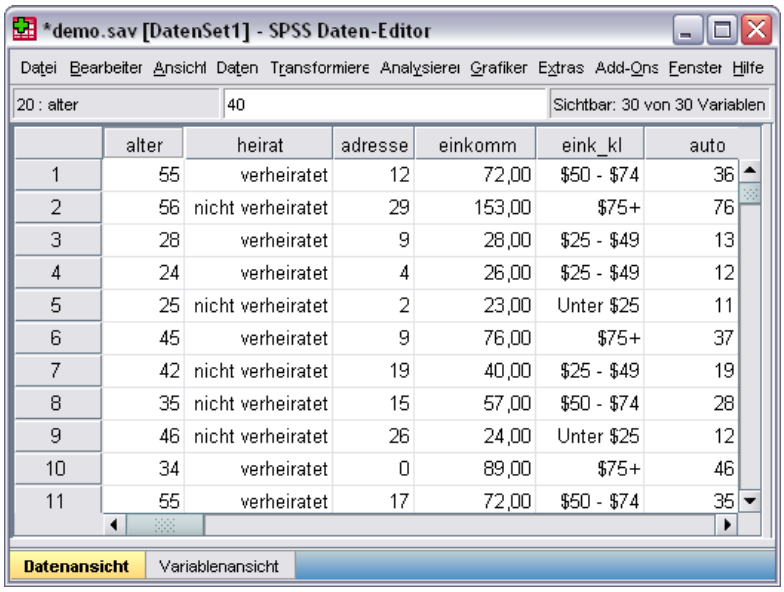
Ansicht
  Wertelabels

Sie können auch die Schaltfläche "Wertelabels" auf der Symbolleiste verwenden.

Abbildung 1-3
Schaltfläche "Wertelabels"

Es werden jetzt aussagekräftige Wertelabels angezeigt, um die Interpretation der Antworten zu erleichtern.

Abbildung 1-4
Wertelabels im Daten-Editor



	alter	heirat	adresse	einkomm	eink_kl	auto
1	55	verheiratet	12	72,00	$50 - $74	36
2	56	nicht verheiratet	29	153,00	$75+	76
3	28	verheiratet	9	28,00	$25 - $49	13
4	24	verheiratet	4	26,00	$25 - $49	12
5	25	nicht verheiratet	2	23,00	Unter $25	11
6	45	verheiratet	9	76,00	$75+	37
7	42	nicht verheiratet	19	40,00	$25 - $49	19
8	35	nicht verheiratet	15	57,00	$50 - $74	28
9	46	nicht verheiratet	26	24,00	Unter $25	12
10	34	verheiratet	0	89,00	$75+	46
11	55	verheiratet	17	72,00	$50 - $74	35

Durchführen einer Analyse

Das Menü “Analysieren” besteht aus einer Liste von allgemeinen Kategorien für Berichte und statistische Analysen.

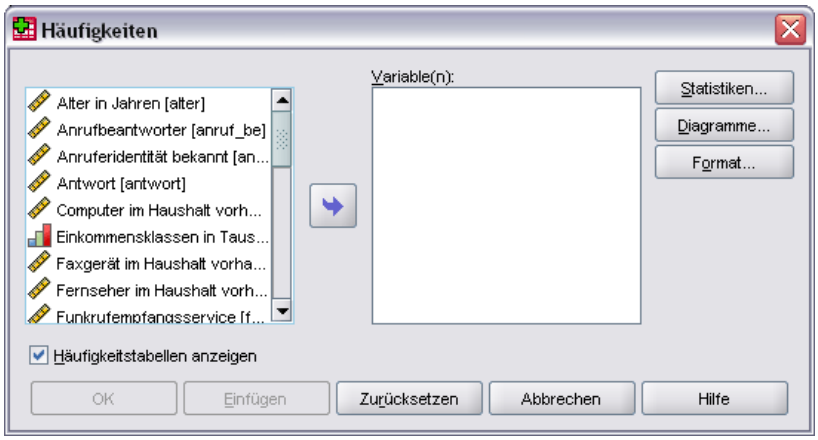
Zunächst soll eine einfache Häufigkeitstabelle erstellt werden.

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Analysieren
  Deskriptive Statistiken
    Häufigkeiten...

Das Dialogfeld “Häufigkeiten” wird angezeigt.

Abbildung 1-5
Dialogfeld “Häufigkeiten”

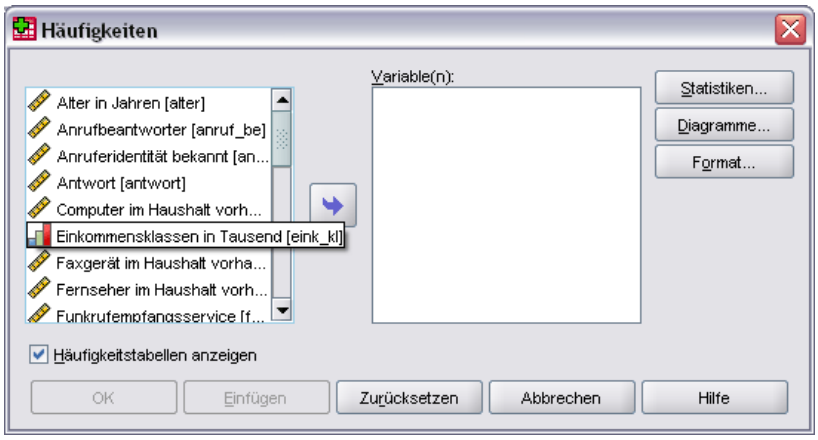


Ein Symbol neben jeder Variablen stellt Informationen zum Datentyp und Messniveau bereit.

Messniveau	**Datentyp**			
	Numerisch	String	Datum	Zeit
Metrisch		entfällt		
Ordinal				
Nominal				

▶ Klicken Sie auf die Variable *Einkommensklassen in Tausend [eink_kl]*.

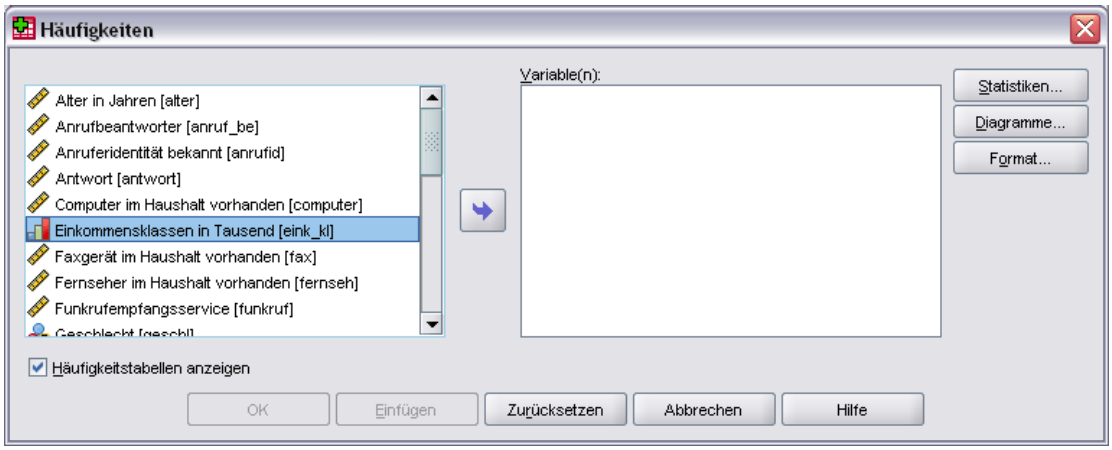
Abbildung 1-6
Variablenlabels und Variablennamen im Dialogfeld “Häufigkeiten”



Wenn das Variablenlabel und/oder der Variablenname in der Liste nicht vollständig angezeigt wird, können Sie das vollständige Label bzw. den vollständigen Namen anzeigen, indem Sie mit dem Cursor darauf zeigen. Der Variablenname *inccat* wird in eckigen Klammern nach dem beschreibenden Variablenlabel angezeigt. Das Variablenlabel lautet *Einkommensklassen in Tausend*. Wenn kein Variablenlabel vorhanden ist, wird im Listenfeld nur der Variablenname angezeigt.

Sie können die Größe von Dialogfeldern ebenso ändern wie die von Fenstern, einfach indem Sie auf die äußeren Ränder oder Ecken klicken und daran ziehen. Wenn Sie beispielsweise das Dialogfeld verbreitern, werden auch die Variablenlisten breiter.

Abbildung 1-7
Dialogfeld nach der Größenänderung



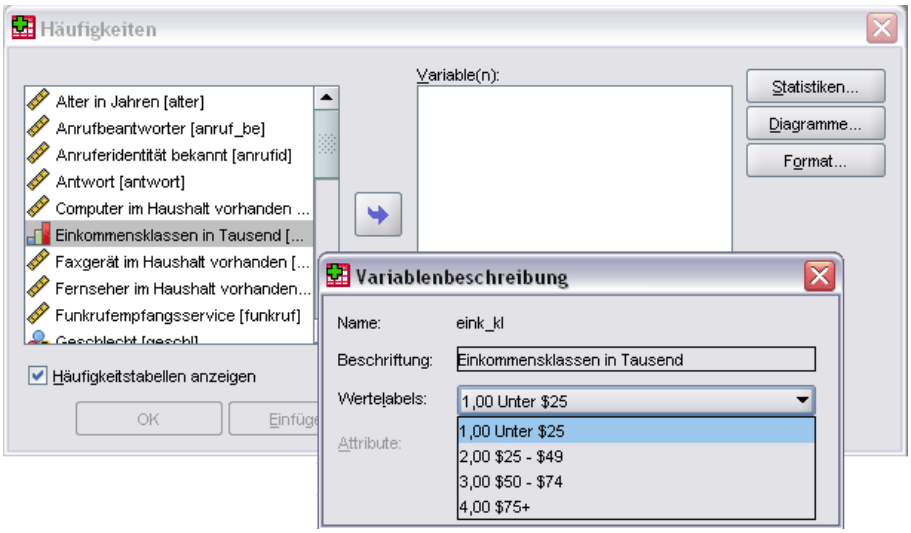
Wählen Sie im Dialogfeld die Variablen, die in die Analyse einbezogen werden sollen, aus der Liste der Quellvariablen auf der linken Seite aus und verschieben sie in die Liste “Variable(n)” auf der rechten Seite. Die Schaltfläche OK (mit der die Analyse ausgeführt wird) wird erst aktiviert, wenn mindestens eine Variable in die Liste “Variable(n)” übernommen wurde.

Weitere Informationen erhalten Sie, indem Sie mit der rechten Maustaste auf eine beliebige Variable in der Liste klicken.

- Klicken Sie mit der rechten Maustaste auf *Einkommensklassen in Tausend [eink_kl]* und wählen Sie die Option Variablenbeschreibung aus.

- Klicken Sie in der Dropdown-Liste “Wertelabels” auf den Abwärtspfeil.

Abbildung 1-8
Für Einkommensklassen definierte Labels

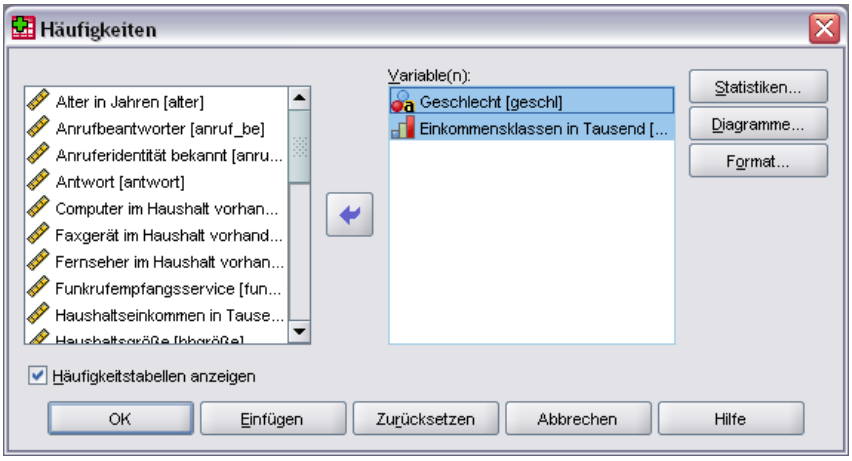


Alle für die Variable definierten Wertelabels werden angezeigt.

▶ Klicken Sie in der Quellvariablenliste auf *Geschlecht [geschl]* und ziehen Sie die Variable in die Zielliste “Variable(n)”.

▶ Klicken Sie in der Quellliste auf *Einkommensklassen in Tausend [eink_kl]* und ziehen Sie die Variable in die Zielliste.

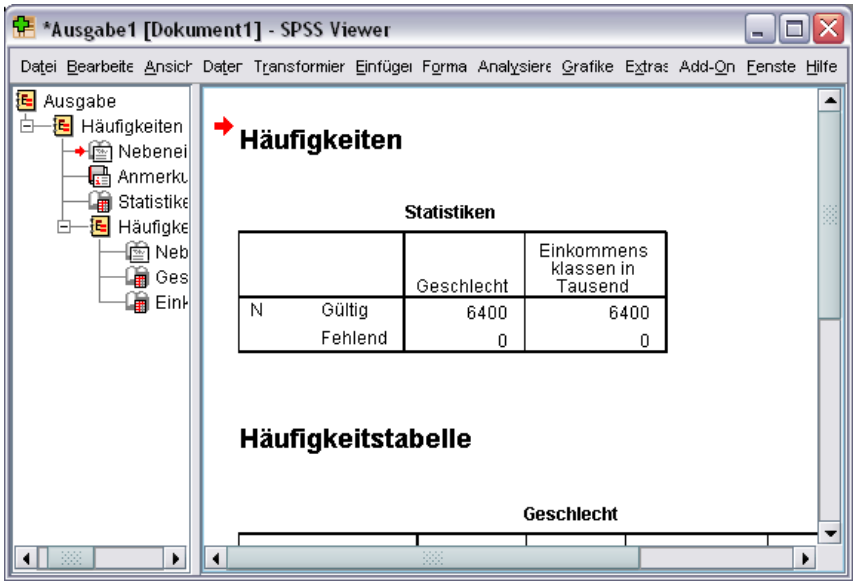
Abbildung 1-9
Für die Analyse ausgewählte Variablen



▶ Klicken Sie auf OK, um die Prozedur auszuführen.

Anzeigen von Ergebnissen

Abbildung 1-10
Viewer-Fenster

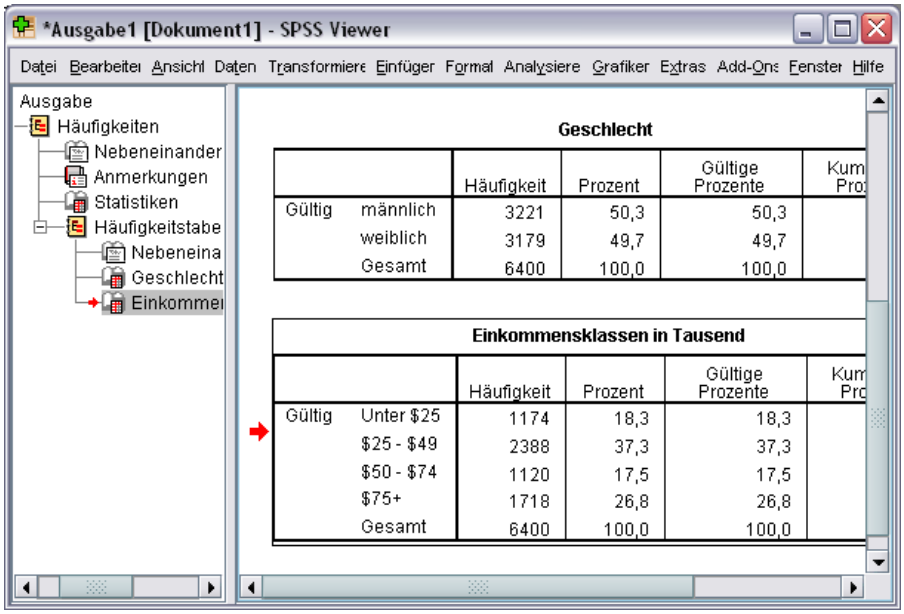


Die Ergebnisse werden im Fenster “Viewer” angezeigt.

Sie können zu jedem Element im Viewer gelangen, indem Sie es einfach im Gliederungsfenster auswählen.

▶ Klicken Sie auf Einkommensklassen in Tausend [eink_kl].

Abbildung 1-11
Häufigkeitstabelle für Einkommensklassen



Die Häufigkeitstabelle für die Einkommensklassen wird angezeigt. Diese Häufigkeitstabelle zeigt die Anzahl und den Prozentsatz der Personen in den einzelnen Einkommensklassen an.

Erstellen von Diagrammen

Zwar können auch mit einigen statistischen Prozeduren Diagramme erstellt werden, Sie können zum Erstellen von Diagrammen jedoch auch das Menü “Grafiken” verwenden.

So können Sie beispielsweise eine Grafik erstellen, in der die Beziehung zwischen dem Besitz von schnurlosen Telefonen und Palm Pilots dargestellt wird.

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

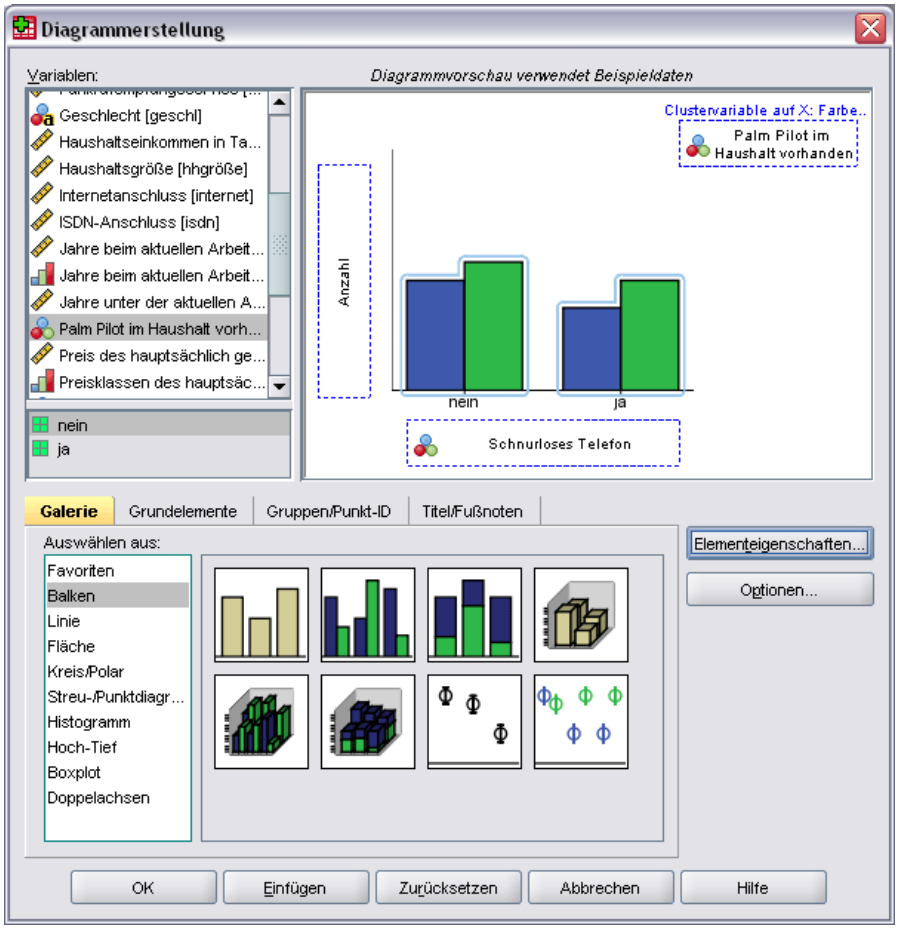
Grafiken
  Diagrammerstellung...

▶ Klicken Sie auf die Registerkarte Galerie, falls diese nicht bereits ausgewählt ist.

▶ Klicken Sie auf Balken, falls dies nicht bereits ausgewählt ist.

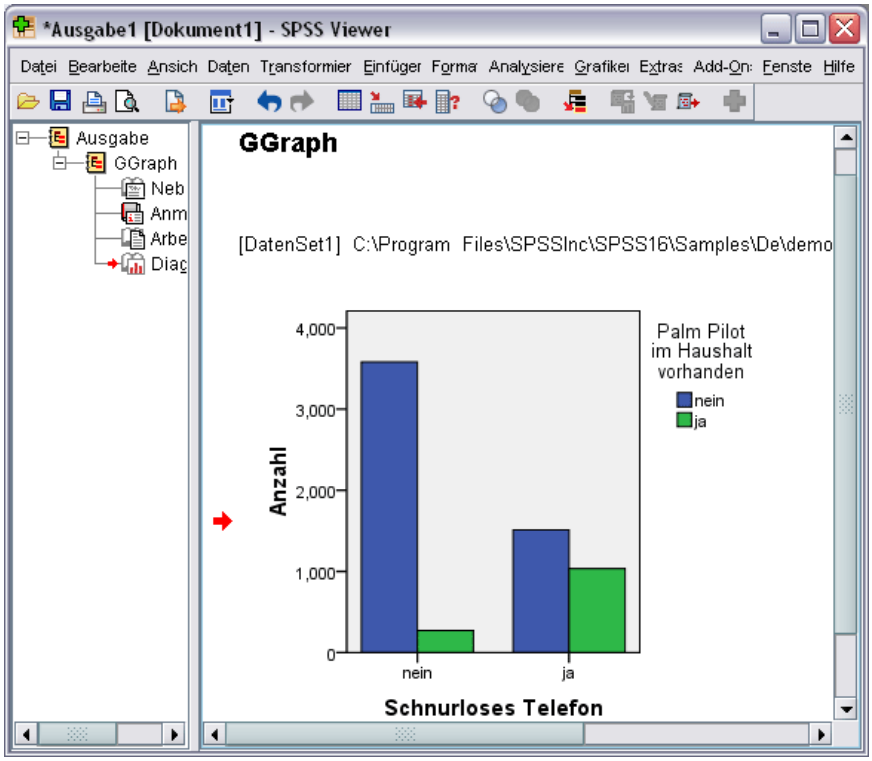
▶ Ziehen Sie das Symbol “Gruppiertes Balkendiagramm” auf die Zeichenfläche. Dies ist der große Bereich oberhalb der Galerie.

Abbildung 1-12
Dialogfeld "Diagrammerstellung"



- Führen Sie einen Bildlauf in der Liste "Variablen" durch, klicken Sie mit der rechten Maustaste auf *Schnurloses Telefon [kabellos]* und wählen Sie als Messniveau Nominal aus.
- Ziehen Sie die Variable *Schnurloses Telefon [kabellos]* auf die *x*-Achse.
- Klicken Sie mit der rechten Maustaste auf *Palm Pilot im Haushalt vorhanden [palm]* und wählen Sie als Messniveau Nominal aus.
- Ziehen Sie die Variable *Palm Pilot im Haushalt vorhanden [palm]* auf die Cluster-Ablagezone, die sich in der rechten oberen Ecke der Zeichenfläche befindet.
- Klicken Sie auf OK, um das Diagramm zu erstellen.

Abbildung 1-13
Balkendiagramm im Viewer-Fenster



Das Balkendiagramm wird im Viewer angezeigt. Das Diagramm zeigt, dass Benutzer von schnurlosen Telefonen weitaus häufiger Palm Pilots besitzen als Personen ohne schnurloses Telefon.

Diagramme und Tabellen können bearbeitet werden, indem Sie im Inhaltsfenster des Viewer-Fensters darauf doppelklicken. Außerdem können die Ergebnisse kopiert und in andere Anwendungen eingefügt werden. Diese Themen werden später behandelt.

Kapitel

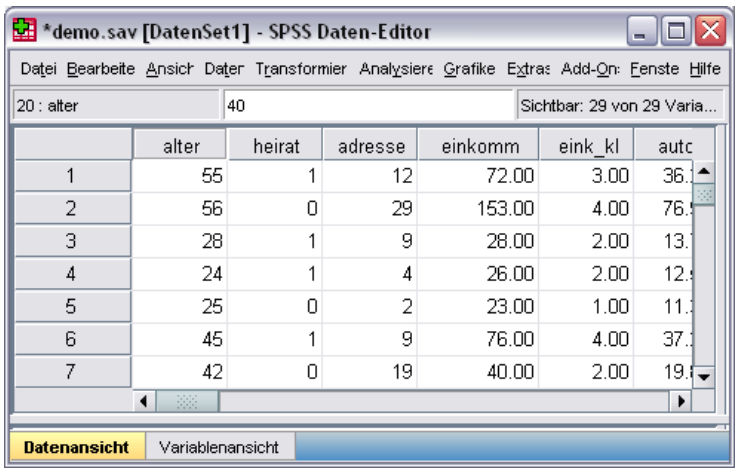
2

Einlesen von Daten

Daten können entweder direkt eingegeben oder aus einer Reihe unterschiedlicher Quellen importiert werden. In diesem Kapitel wird das Einlesen von Daten erläutert, die in Datendateien im SPSS-Format, Tabellenkalkulationsanwendungen, wie Microsoft Excel, Datenbankanwendungen, wie Microsoft Access, und Textdateien gespeichert sind.

Grundlegende Struktur einer Datendatei im SPSS-Format

Abbildung 2-1
Daten-Editor



Datendateien im SPSS-Format sind in Fälle (Zeilen) und Variablen (Spalten) untergliedert. In dieser Datendatei stellen die Fälle die einzelnen Befragten einer Umfrage dar. Die Variablen stellen die Antworten auf die einzelnen Fragen dar, die in der Umfrage gestellt wurden.

Einlesen einer Datendatei im SPSS-Format

Datendateien im SPSS-Format mit der Dateinamenerweiterung *.sav* enthalten die gespeicherten Daten. So öffnen Sie demo.sav, eine Beispieldatei, die bei der Installation des Programms installiert wurde:

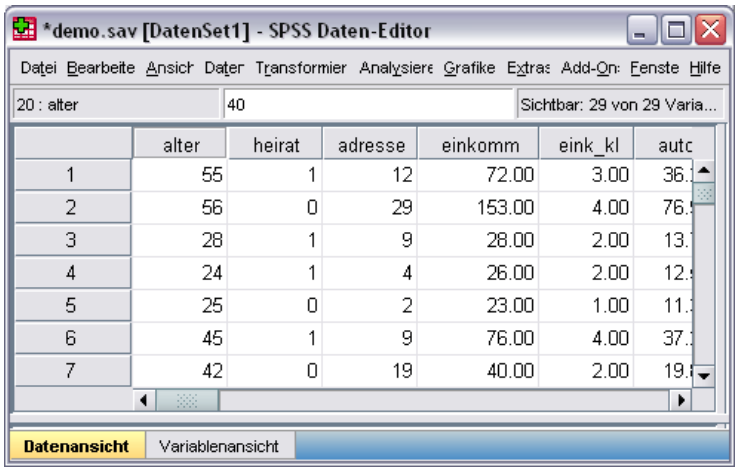
▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Datei
  Öffnen
    Daten...

▶ Wechseln Sie zu der Datei *demo.sav* und öffnen Sie sie. Für weitere Informationen siehe Beispieldateien in Anhang A auf S. 163.

Die Daten werden nun im Daten-Editor angezeigt.

Abbildung 2-2
Geöffnete Datendatei



	alter	heirat	adresse	einkomm	eink_kl	autc
1	55	1	12	72.00	3.00	36.
2	56	0	29	153.00	4.00	76.
3	28	1	9	28.00	2.00	13.
4	24	1	4	26.00	2.00	12.
5	25	0	2	23.00	1.00	11.
6	45	1	9	76.00	4.00	37.
7	42	0	19	40.00	2.00	19.

Einlesen von Daten aus Tabellenkalkulationsdateien

Statt alle Daten direkt in den Daten-Editor einzugeben, können Sie Daten aus Anwendungen wie Microsoft Excel einlesen. Sie können auch die Spaltenüberschriften als Variablennamen einlesen.

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

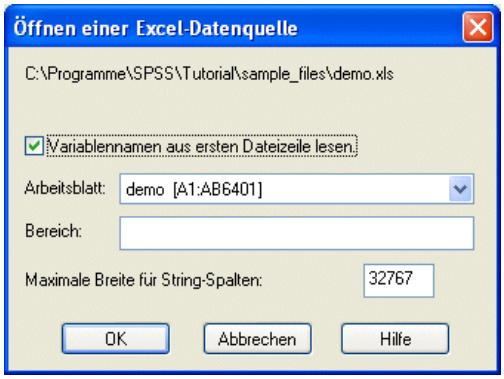
Datei
  Öffnen
    Daten...

▶ Wählen Sie Excel (*.xls) als anzuzeigenden Dateityp aus.

▶ Öffnen Sie die Datei *demo.xls*. Für weitere Informationen siehe Beispieldateien in Anhang A auf S. 163.

Das Dialogfeld “Öffnen einer Excel-Datenquelle” wird angezeigt. Hier können Sie angeben, ob das Tabellenblatt Variablennamen enthält und welche Zellen importiert werden sollen. In Excel 95 und späteren Versionen können Sie auch angeben, welche Arbeitsblätter importiert werden sollen.

Abbildung 2-3
Dialogfeld "Öffnen einer Excel-Datenquelle"



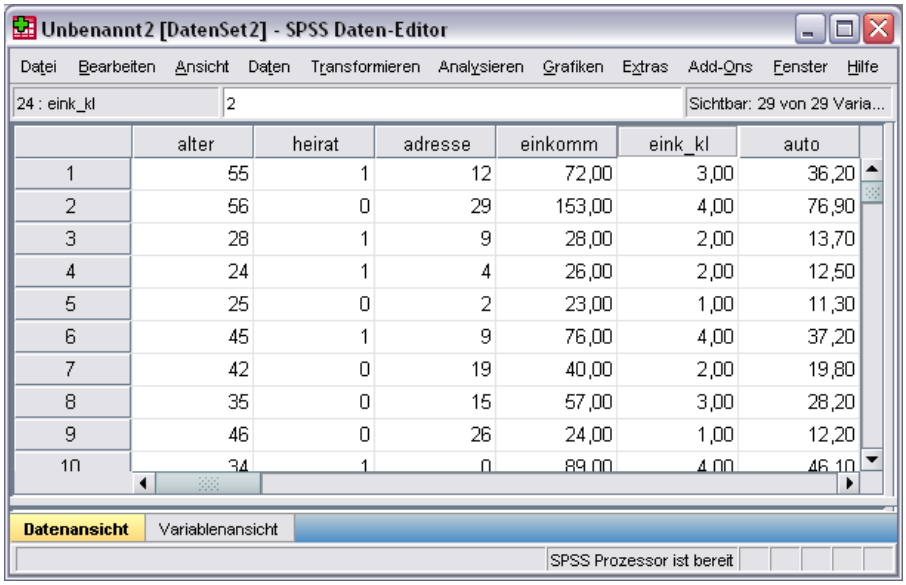
▶ Stellen Sie sicher, dass Variablennamen aus ersten Dateizeile lesen aktiviert ist. Bei dieser Option werden Spaltenüberschriften als Variablennamen eingelesen.

Wenn die Spaltenüberschriften nicht den Regeln für Variablennamen von SPSS entsprechen, werden sie in gültige Variablennamen umgewandelt und die ursprünglichen Spaltenüberschriften werden als Variablenlabels gespeichert. Wenn Sie lediglich einen Teil des Tabellenblatts importieren möchten, geben Sie im Textfeld "Bereich" den zu importierenden Zellenbereich an.

▶ Klicken Sie auf OK, um die Excel-Datei einzulesen.

Die Daten werden jetzt im Daten-Editor angezeigt, wobei die Spaltenüberschriften als Variablennamen verwendet werden. Da die Variablennamen keine Leerzeichen enthalten dürfen, wurden die Leerzeichen aus den ursprünglichen Spaltenüberschriften entfernt. So würde *fam statu* in der Excel-Datei beispielsweise zur Variable *heirat* in SPSS. Die ursprüngliche Überschrift der Spalte wird als Variablenlabel übernommen.

Abbildung 2-4
Importierte Excel-Daten



	alter	heirat	adresse	einkomm	eink_kl	auto
1	55	1	12	72,00	3,00	36,20
2	56	0	29	153,00	4,00	76,90
3	28	1	9	28,00	2,00	13,70
4	24	1	4	26,00	2,00	12,50
5	25	0	2	23,00	1,00	11,30
6	45	1	9	76,00	4,00	37,20
7	42	0	19	40,00	2,00	19,80
8	35	0	15	57,00	3,00	28,20
9	46	0	26	24,00	1,00	12,20
10	34	1	0	89,00	4,00	46,10

Einlesen von Daten aus einer Datenbank

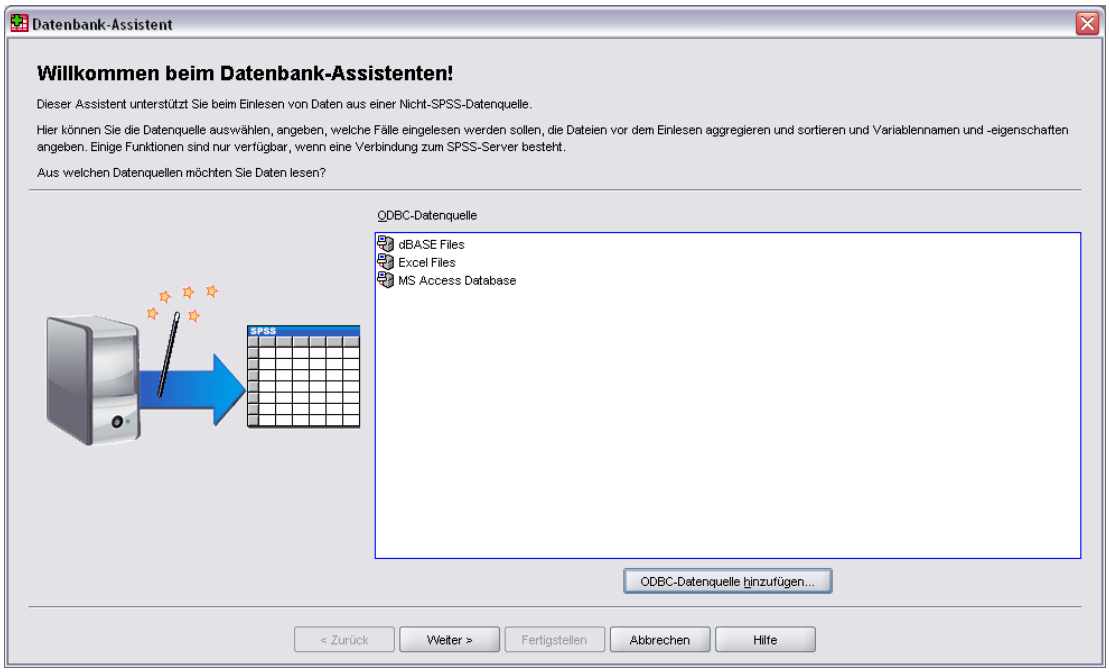
Daten aus Datenbankquellen können mithilfe des Datenbank-Assistenten problemlos importiert werden. Jede Datenbank, bei der ODBC-Treiber (Open Database Connectivity) verwendet werden, kann nach der Installation der Treiber direkt eingelesen werden. Auf der Installations-CD befinden sich ODBC-Treiber für viele Datenbankformate. Weitere Treiber können von Drittanbietern bezogen werden. In diesem Beispiel wird Microsoft Access, eine der gängigsten Datenbankanwendungen, behandelt.

Anmerkung: Dieses Beispiel bezieht sich ausschließlich auf Microsoft Windows. Außerdem ist ein ODBC-Treiber für Access erforderlich. Die Schritte sind auf den anderen Plattformen ähnlich. Es kann jedoch ein Drittanbieter-ODBC-Treiber für Access erforderlich sein.

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Datei
  Datenbank öffnen
    Neue Abfrage...

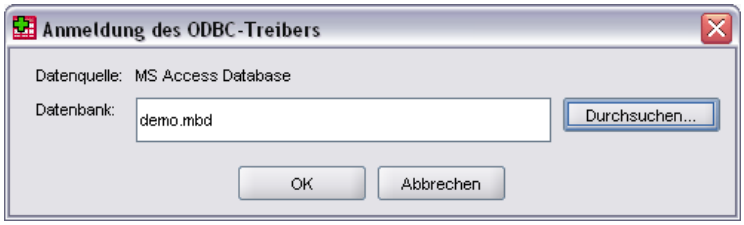
Abbildung 2-5
Begrüßungsdialogfeld des Datenbank-Assistenten



▶ Wählen Sie in der Liste der Datenquellen die Option MS Access-Datenbank aus und klicken Sie dann auf Weiter.

Anmerkung: Abhängig von Ihrer Installation werden auf der linken Seite des Assistenten möglicherweise auch OLE DB-Datenquellen angezeigt (nur bei Windows-Betriebssystemen). In diesem Beispiel wird jedoch die Liste der ODBC-Datenquellen verwendet, die auf der rechten Seite angezeigt wird.

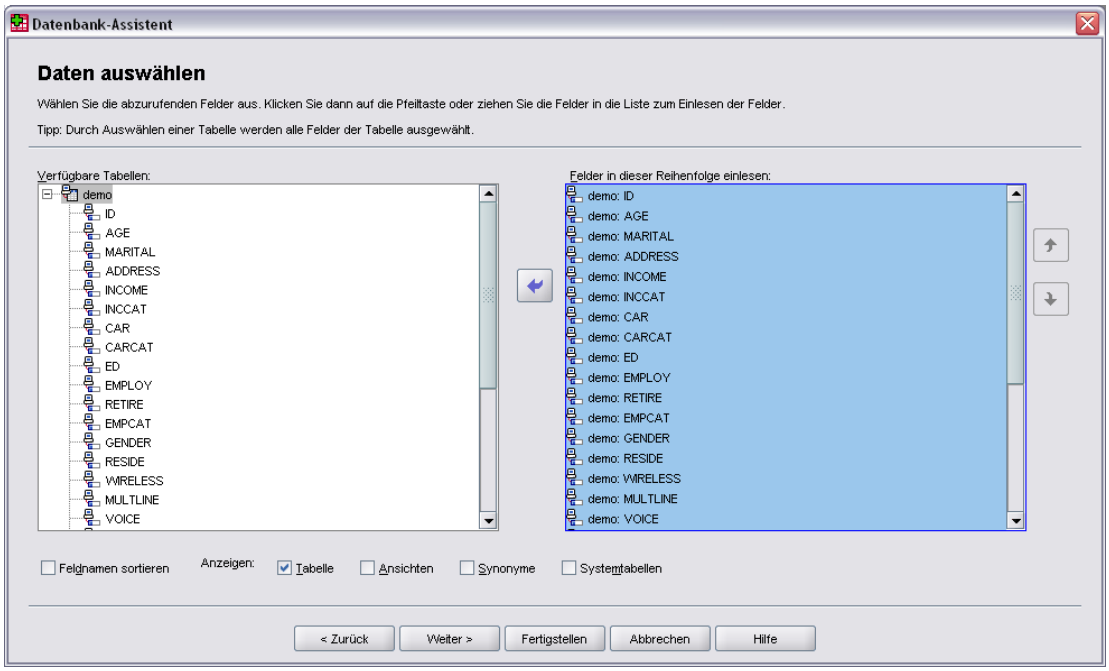
Abbildung 2-6
Dialogfeld "Anmeldung des ODBC-Treibers"



- Klicken Sie auf Durchsuchen, um die zu öffnende Access-Datenbank zu suchen.
- Öffnen Sie die Datei *demo.mdb*. Für weitere Informationen siehe Beispieldateien in Anhang A auf S. 163.
- Klicken Sie im Anmeldungs-Dialogfeld auf OK.

Im folgenden Schritt können Sie die zu importierenden Tabellen und Variablen angeben.

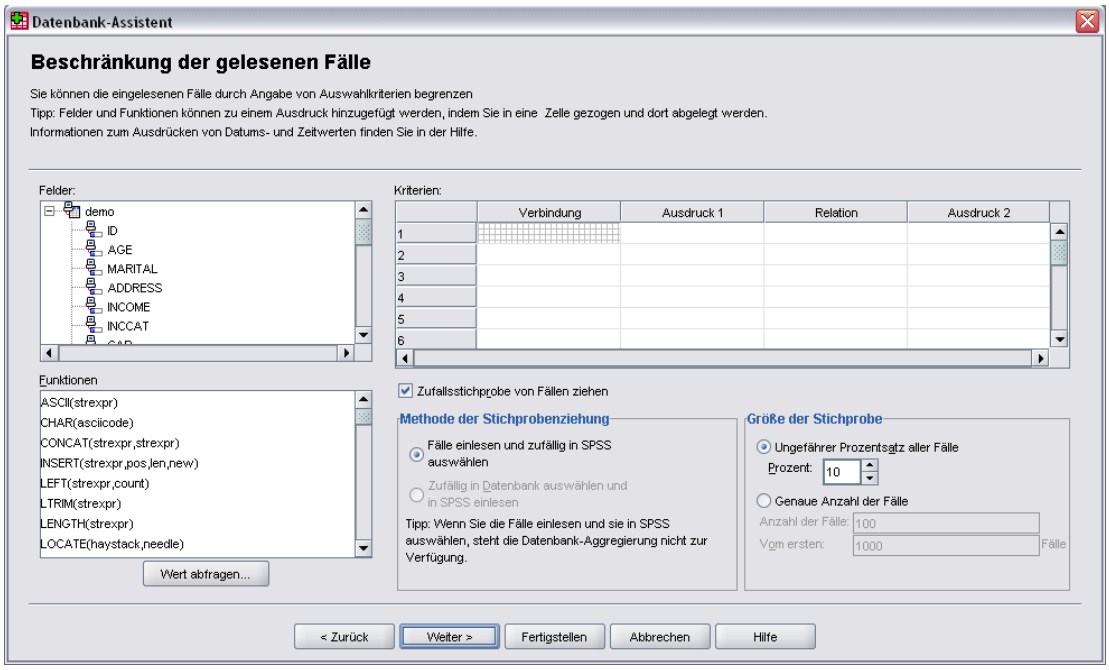
Abbildung 2-7
Schritt "Daten auswählen"



- Ziehen Sie die gesamte Demo-Tabelle in die Liste "Felder in dieser Reihenfolge einlesen".
- Klicken Sie auf Weiter.

Im folgenden Schritt wählen Sie aus, welche Datensätze (Fälle) importiert werden sollen.

Abbildung 2-8
Schritt "Beschränkung der gelesenen Fälle"

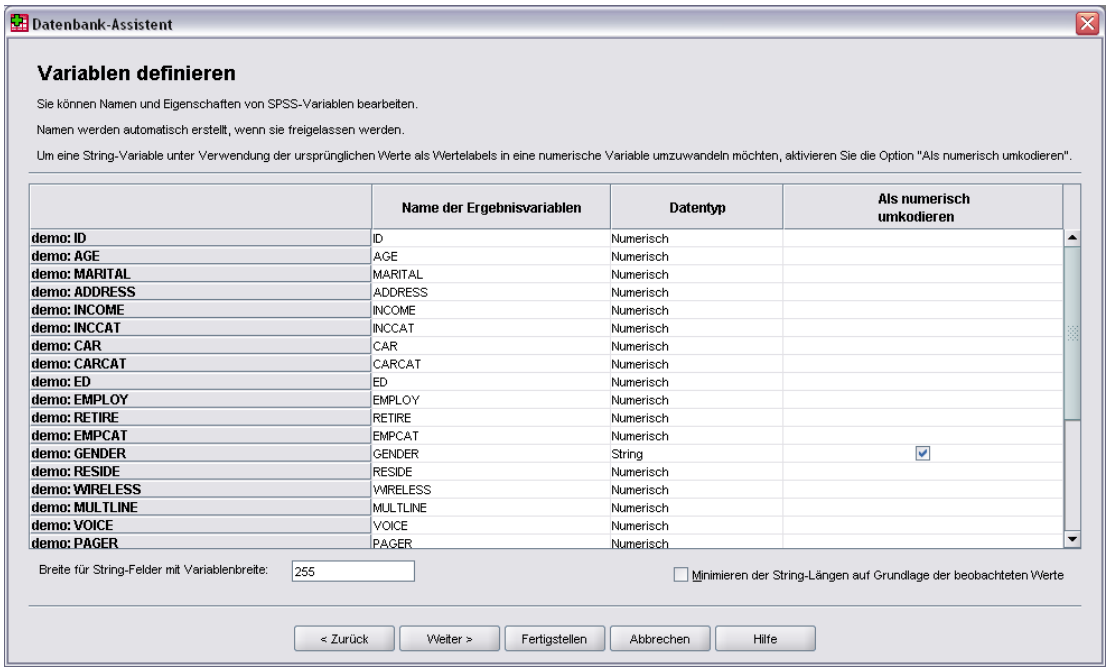


Wenn Sie nicht alle Fälle importieren möchten, können Sie eine Untergruppe von Fällen (beispielsweise Männer, die älter sind als 30) oder auch eine Zufallsstichprobe von Fällen aus der Datenquelle importieren. Bei großen Datenquellen soll die Anzahl der Fälle möglicherweise auf eine kleine, repräsentative Auswahl begrenzt werden, um die Verarbeitungszeit zu verkürzen.

▶ Klicken Sie zum Fortfahren auf Weiter.

Feldnamen werden für die Erstellung von Variablennamen verwendet. Falls erforderlich, werden die Namen in gültige Variablennamen konvertiert. Die ursprünglichen Feldnamen werden als Variablenlabels übernommen. Die Variablennamen können vor dem Importieren der Datenbank geändert werden.

Abbildung 2-9
Schritt "Variablen definieren"

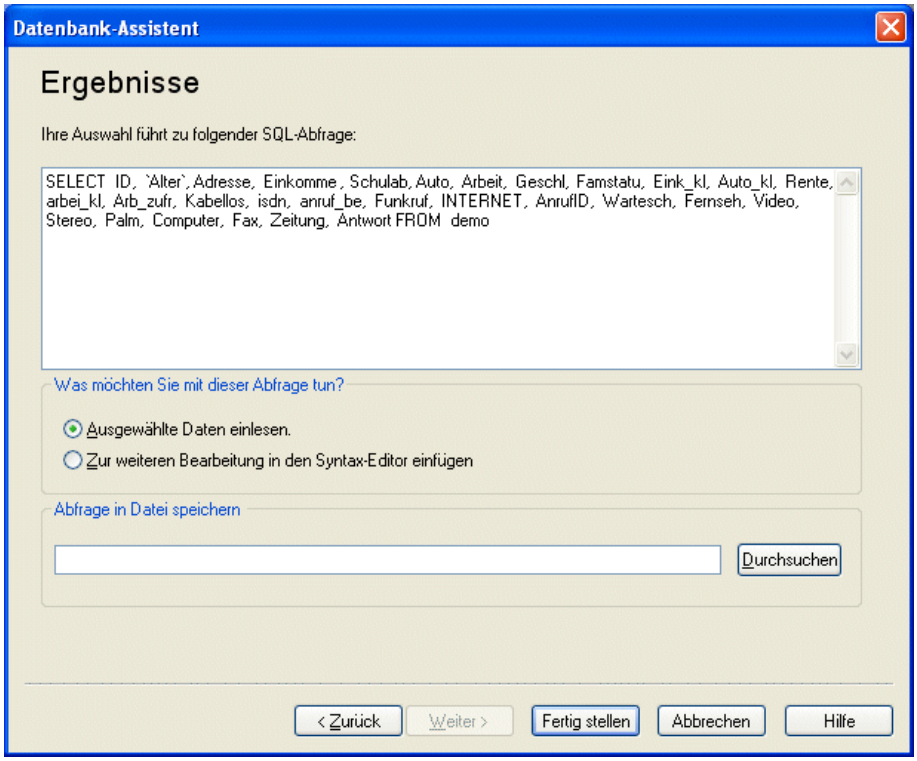


- Klicken Sie im Feld "Geschlecht" auf die Zelle Als numerisch umkodieren. Durch diese Option werden String-Variablen in ganzzahlige Variablen umgewandelt und der ursprüngliche Wert wird als Wertelabel für die neue Variable beibehalten.

- Klicken Sie zum Fortfahren auf Weiter.

Die aus Ihren Angaben im Datenbank-Assistenten erstellte SQL-Anweisung wird im Schritt “Ergebnisse” angezeigt. Diese Anweisung kann sofort ausgeführt oder für die spätere Verwendung in einer Datei gespeichert werden.

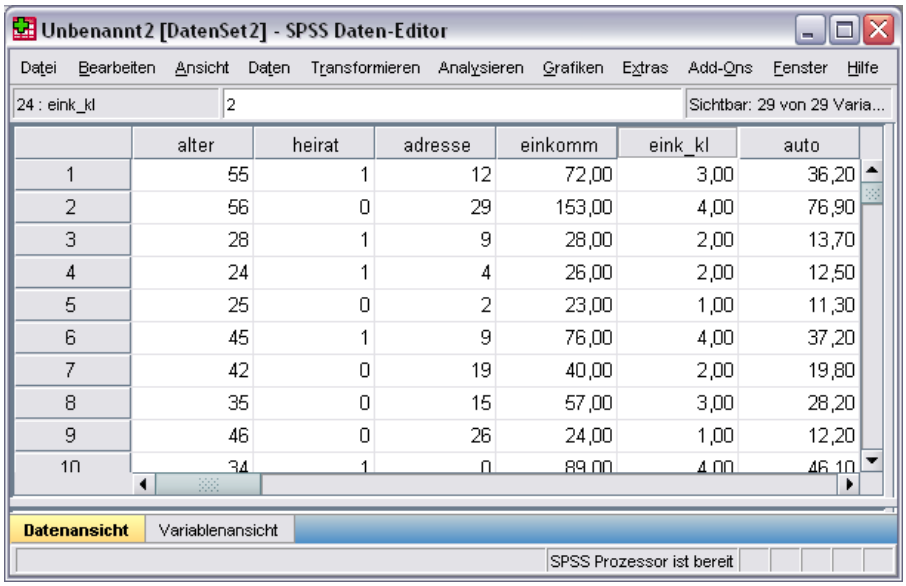
Abbildung 2-10
Schritt “Ergebnisse”



► Klicken Sie auf Fertig stellen, um die Daten zu importieren.

Alle in der Access-Datenbank zum Importieren ausgewählten Daten stehen jetzt im Daten-Editor zur Verfügung.

Abbildung 2-11
Aus einer Access-Datenbank importierte Daten



	alter	heirat	adresse	einkomm	eink_kl	auto
1	55	1	12	72,00	3,00	36,20
2	56	0	29	153,00	4,00	76,90
3	28	1	9	28,00	2,00	13,70
4	24	1	4	26,00	2,00	12,50
5	25	0	2	23,00	1,00	11,30
6	45	1	9	76,00	4,00	37,20
7	42	0	19	40,00	2,00	19,80
8	35	0	15	57,00	3,00	28,20
9	46	0	26	24,00	1,00	12,20
10	34	1	0	89,00	4,00	46,10

Einlesen von Daten aus einer Textdatei

Textdateien stellen eine weitere gängige Datenquelle dar. Bei vielen Tabellenkalkulationsprogrammen und Datenbanken kann der Inhalt in einer Reihe von Textdateiformaten gespeichert werden. Der Begriff Komma- oder Tabulator-getrennte Datei bezieht sich auf das Trennzeichen (Kommata oder Tabulatoren), das die einzelnen Variablen in einer Datenzeile voneinander trennt. In diesem Beispiel sind die Daten Tabulator-getrennt.

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

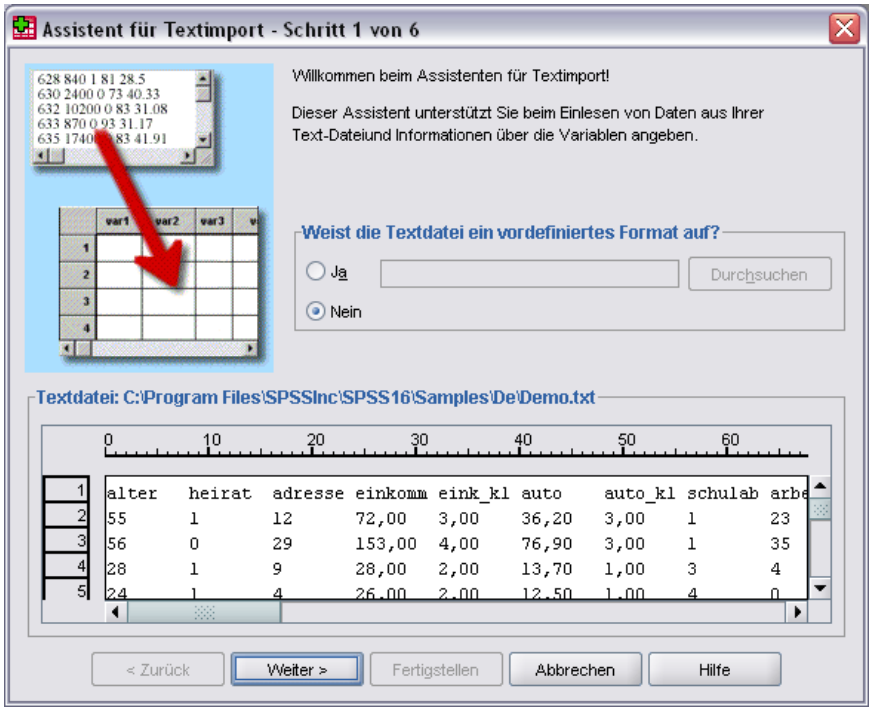
Datei
  Textdaten einlesen...

▶ Wählen Sie Text (*.txt) als anzuzeigenden Dateityp aus.

▶ Öffnen Sie die Datei *demo.txt*. Für weitere Informationen siehe Beispieldateien in Anhang A auf S. 163.

Der Assistent für Textimport unterstützt Sie bei der Definition, wie die angegebene Textdatei interpretiert werden soll.

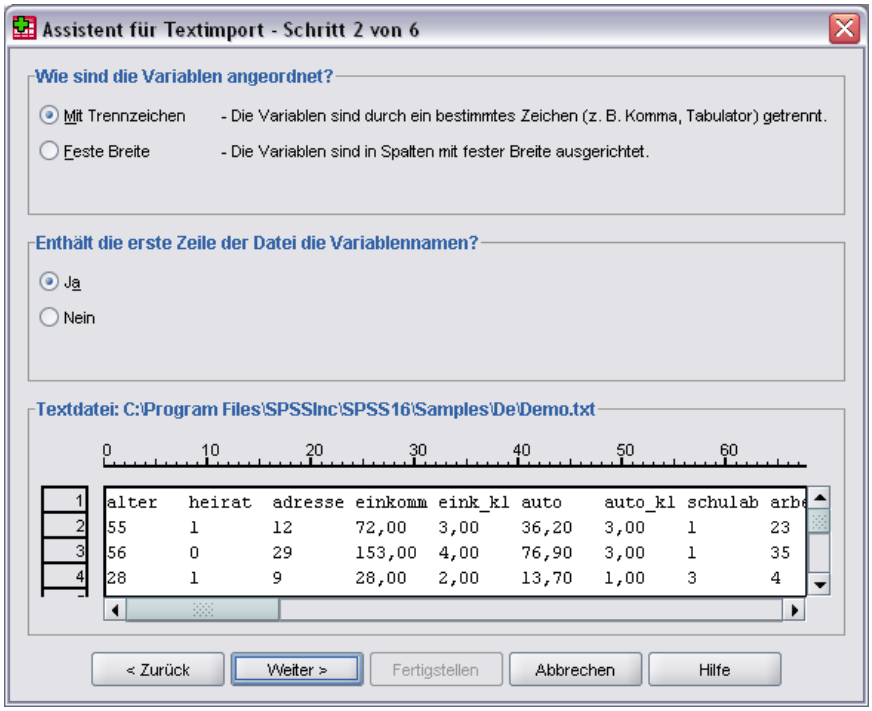
Abbildung 2-12
Assistent für Textimport: Schritt 1 von 6



- In Schritt 1 können Sie ein vordefiniertes Format auswählen oder im Assistenten ein neues Format erstellen. Wählen Sie Nein aus, um anzugeben, dass ein neues Format erstellt werden soll.
- Klicken Sie zum Fortfahren auf Weiter.

Wie bereits erwähnt, werden in dieser Datei Tabulatoren als Trennzeichen verwendet. Außerdem werden die Variablennamen in der obersten Zeile dieser Datei definiert.

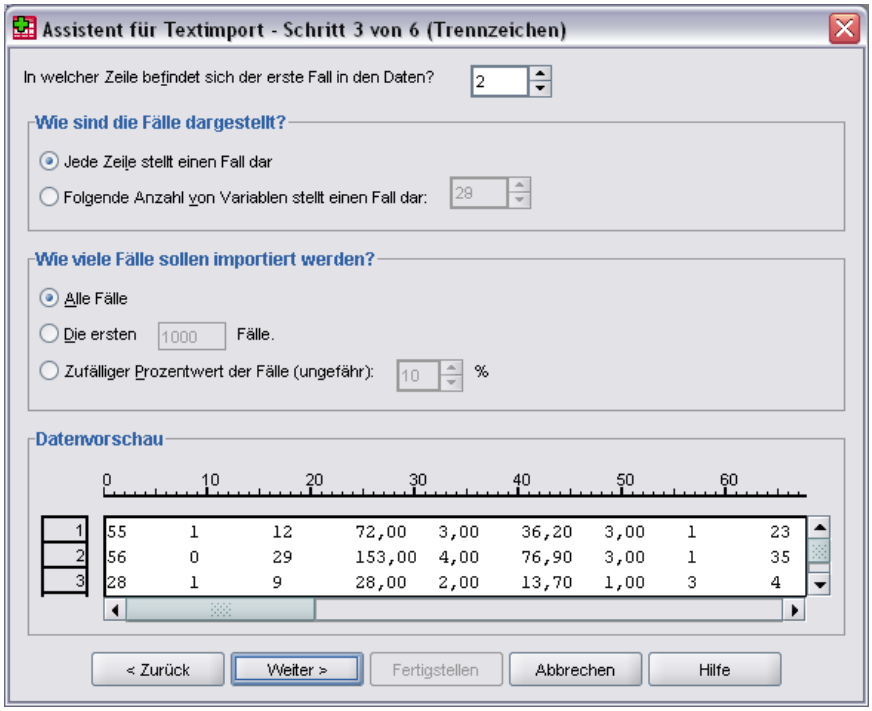
Abbildung 2-13
Assistent für Textimport: Schritt 2 von 6



- Wählen Sie Mit Trennzeichen aus, um festzulegen, dass die Daten durch Trennzeichen voneinander getrennt werden.
- Wählen Sie Ja aus, um festzulegen, dass die Variablennamen aus der ersten Zeile der Datei gelesen werden sollen.
- Klicken Sie zum Fortfahren auf Weiter.

▶ Geben Sie im oberen Abschnitt des nächsten Dialogfelds 2 ein, um festzulegen, dass sich die erste Datenzeile auf der zweiten Zeile der Textdatei befindet.

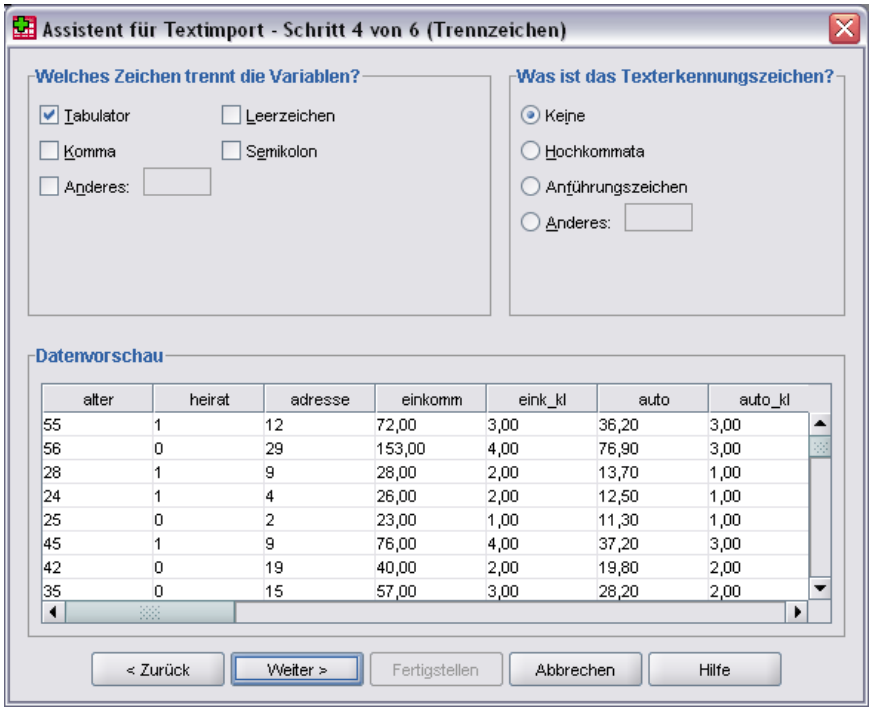
Abbildung 2-14
Assistent für Textimport: Schritt 3 von 6



▶ Behalten Sie die Standardwerte dieses Dialogfelds bei und klicken Sie dann zum Fortfahren auf Weiter.

In der Datenvorschau in Schritt 4 können Sie sich vergewissern, dass die Daten ordnungsgemäß eingelesen werden.

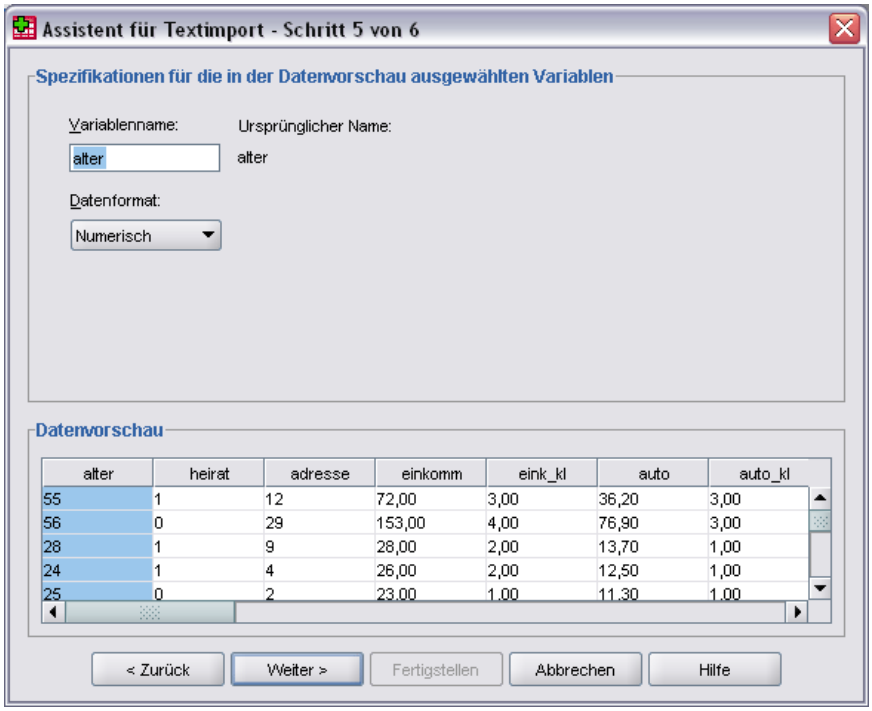
Abbildung 2-15
Assistent für Textimport: Schritt 4 von 6



- ▶ Wählen Sie Tabulator aus und deaktivieren Sie die anderen Optionen.
- ▶ Klicken Sie zum Fortfahren auf Weiter.

Da die Variablennamen aufgrund der Formatierungsbeschränkungen möglicherweise abgeschnitten wurden, können Sie unpassende Namen in diesem Dialogfeld ändern.

Abbildung 2-16
Assistent für Textimport: Schritt 5 von 6



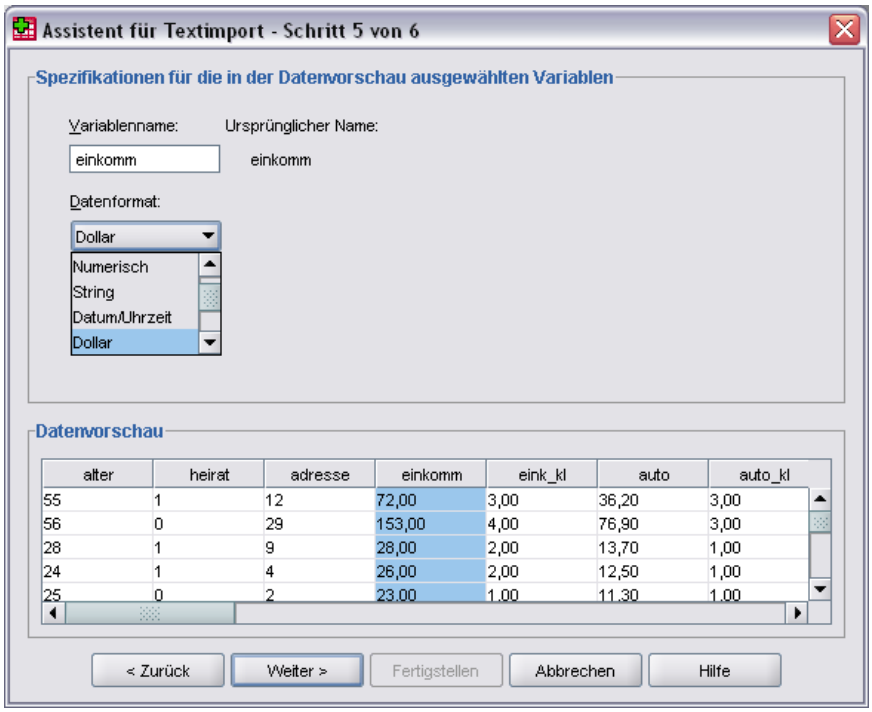
An dieser Stelle können auch Datentypen definiert werden. So kann beispielsweise davon ausgegangen werden, dass die Einkommensvariable einen bestimmten Wert in Dollar enthält.

So ändern Sie den Datentyp:

▶ Wählen Sie unter “Datenvorschau” die zu ändernde Variable aus, in diesem Fall *einkomm*.

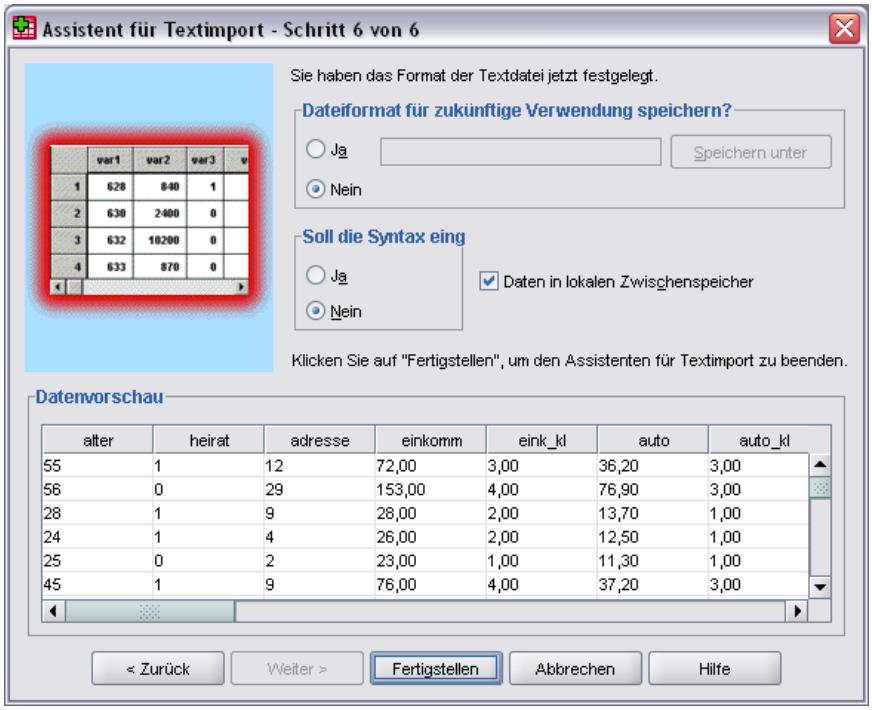
▶ Wählen Sie in der Liste “Datenformat” den Eintrag Dollar aus.

Abbildung 2-17
Ändern des Datentyps



▶ Klicken Sie zum Fortfahren auf Weiter.

Abbildung 2-18
Assistent für Textimport: Schritt 6 von 6



▶ Behalten Sie die Standardeinstellungen in diesem Dialogfeld bei und klicken Sie auf Fertig stellen, um die Daten zu importieren.

Speichern von Daten

Wenn Sie eine Datendatei im SPSS-Format speichern möchten, muss der Daten-Editor das aktive Fenster sein.

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Datei
  Speichern

▶ Wechseln Sie zum gewünschten Verzeichnis.

▶ Geben Sie einen Namen für die Datei ein.

Mithilfe der Schaltfläche “Variablen” können Sie auswählen, welche Variablen im Daten-Editor in der Datendatei im SPSS-Format gespeichert werden sollen. In der Standardeinstellung werden alle Variablen im Daten-Editor gespeichert.

▶ Klicken Sie auf Speichern.

In der Titelleiste des Daten-Editors wird nun der angegebene Dateiname angezeigt. Hiermit wird bestätigt, dass die Datei erfolgreich als Datendatei im SPSS-Format gespeichert wurde. Die Datei enthält sowohl Informationen zu den Variablen (Namen, Datentypen und, sofern vorhanden, Labels und Codes für fehlende Werte) sowie alle Datenwerte.

Kapitel

3

Verwenden des Daten-Editors

Der Daten-Editor zeigt den Inhalt der aktiven Datendatei an. Die Informationen im Daten-Editor bestehen aus Variablen und Fällen.

- In der Datenansicht stellen Spalten Variablen und Zeilen Fälle (Beobachtungen) dar.
- In der Variablenansicht entspricht eine Zeile einer Variablen und eine Spalte einem Attribut, das dieser Variablen zugeordnet ist.

Variablen dienen der Darstellung der verschiedenen Typen von Daten, die Sie zusammengestellt haben. Häufig wird dies mit einer Umfrage verglichen. Die Antwort auf jede Frage einer Umfrage entspricht einer Variablen. Es gibt viele verschiedene Variablentypen, z. B. Zahlen, Strings, Währungen und Datumsangaben.

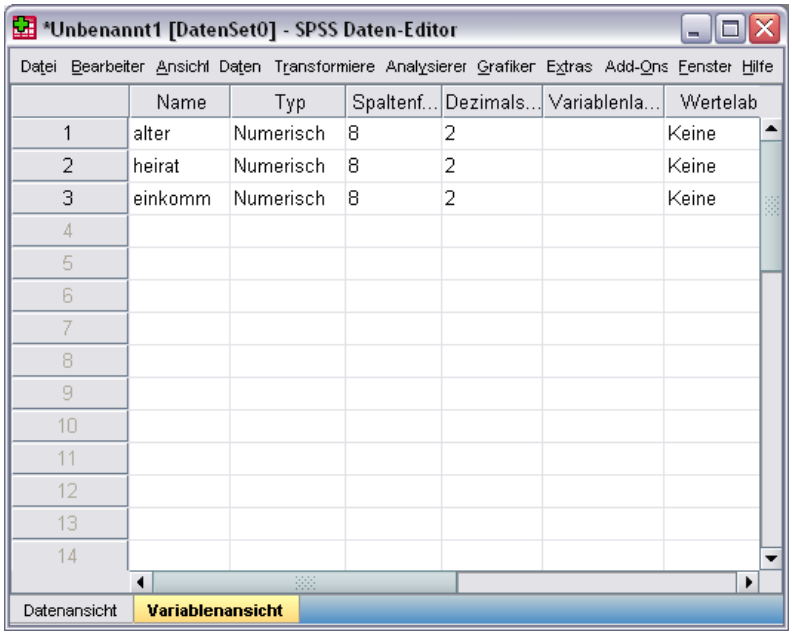
Eingeben numerischer Daten

Daten können im Daten-Editor eingegeben werden. Dies kann bei kleinen Datendateien und bei kleineren Änderungen in größeren Datendateien nützlich sein.

▶ Klicken Sie am unteren Rand des Fensters “Daten-Editor” auf die Registerkarte Variablenansicht.

Sie müssen die Variablen definieren, die verwendet werden sollen. In diesem Fall werden nur drei Variablen benötigt. *alter*, *heirat* und *einkomme*.

Abbildung 3-1
Variablennamen in der Variablenansicht



▶ Geben Sie in der ersten Zeile der ersten Spalte alter ein.

▶ Geben Sie in der zweiten Zeile heirat ein.

▶ Geben Sie in der dritten Zeile einkomme ein.

Neue Variablen erhalten automatisch den Datentyp “Numerisch”.

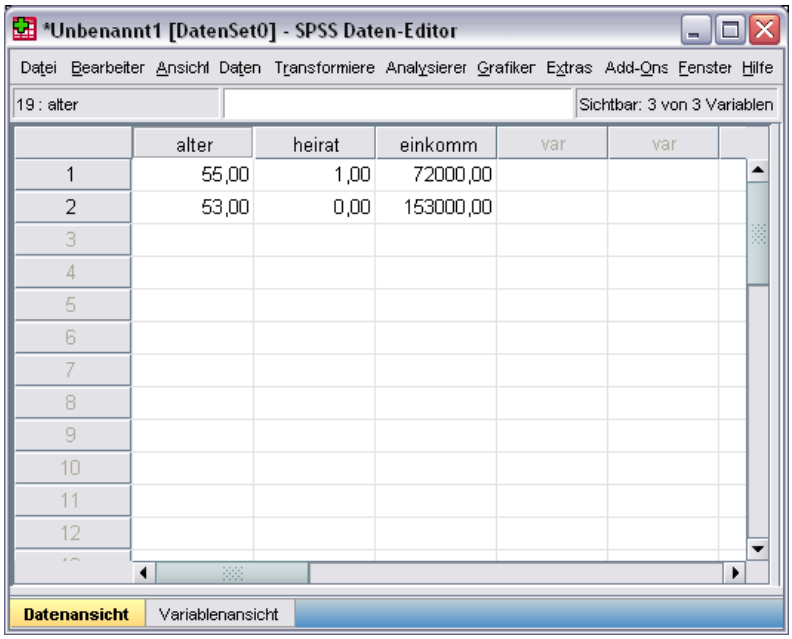
Wenn Sie keine Variablennamen eingeben, werden automatisch eindeutige Namen generiert. Diese Namen sind allerdings nicht aussagekräftig und es wird empfohlen, sie nicht für große Datendateien zu verwenden.

▶ Klicken Sie auf die Registerkarte Datenansicht, um die Eingabe von Daten fortzusetzen.

Die in der Variablenansicht eingegebenen Namen sind nun die Überschriften der ersten drei Spalten der Datenansicht.

Beginnen Sie mit der Dateneingabe in der ersten Zeile der ersten Spalte.

Abbildung 3-2
In der Datenansicht eingegebene Werte



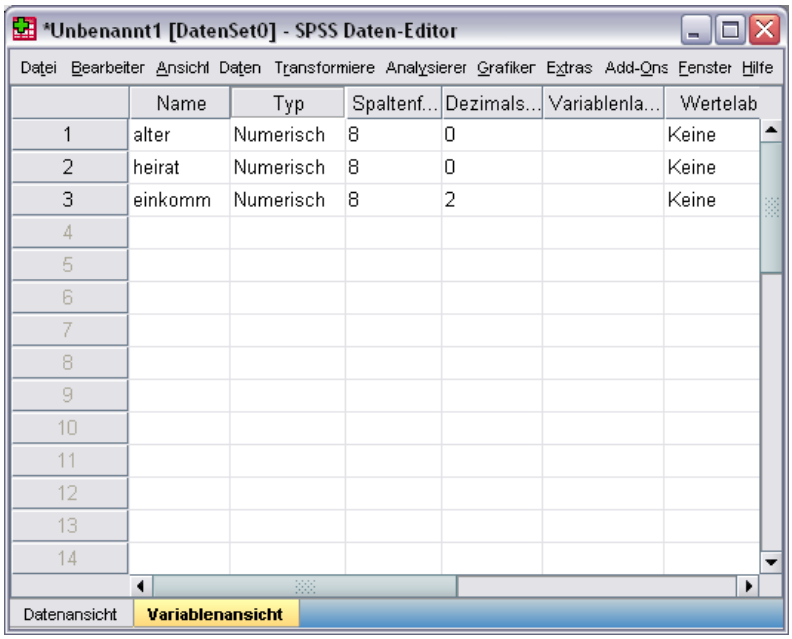
- Geben Sie in der Spalte *alter* den Wert 55 ein.
- Geben Sie in der Spalte *heirat* den Wert 1 ein.
- Geben Sie in der Spalte *einkomme* den Wert 72000 ein.
- Positionieren Sie den Cursor in der zweiten Zeile der ersten Spalte und geben Sie die Daten für die nächste Person ein.
- Geben Sie in der Spalte *alter* den Wert 53 ein.
- Geben Sie in der Spalte *heirat* den Wert 0 ein.
- Geben Sie in der Spalte *einkomme* den Wert 153000 ein.

Im Moment werden die Werte in den Spalten *alter* und *heirat* mit Dezimalpunkten dargestellt, obwohl die Werte als ganze Zahlen gemeint sind. So blenden Sie die Dezimalpunkte bei diesen Variablen aus:

- Klicken Sie am unteren Rand des Fensters “Daten-Editor” auf die Registerkarte Variablenansicht.
- Geben Sie der Spalte *Dezimalstellen* der Zeile *alter* den Wert 0 ein, um die Dezimalstellen auszublenden.

▶ Geben Sie der Spalte *Dezimalstellen* der Zeile *heirat* den Wert 0 ein, um die Dezimalstellen auszublenden.

Abbildung 3-3
Aktualisierte Eigenschaft "Dezimalstellen" für "alter" und "heirat"



	Name	Typ	Spaltenf...	Dezimals...	Variablenla...	Wertelab
1	alter	Numerisch	8	0		Keine
2	heirat	Numerisch	8	0		Keine
3	einkomm	Numerisch	8	2		Keine
4						
5						
6						
7						
8						
9						
10						
11						
12						
13						
14						

Eingeben von String-Daten

Im Daten-Editor können auch nichtnumerische Daten wie Textstrings eingegeben werden.

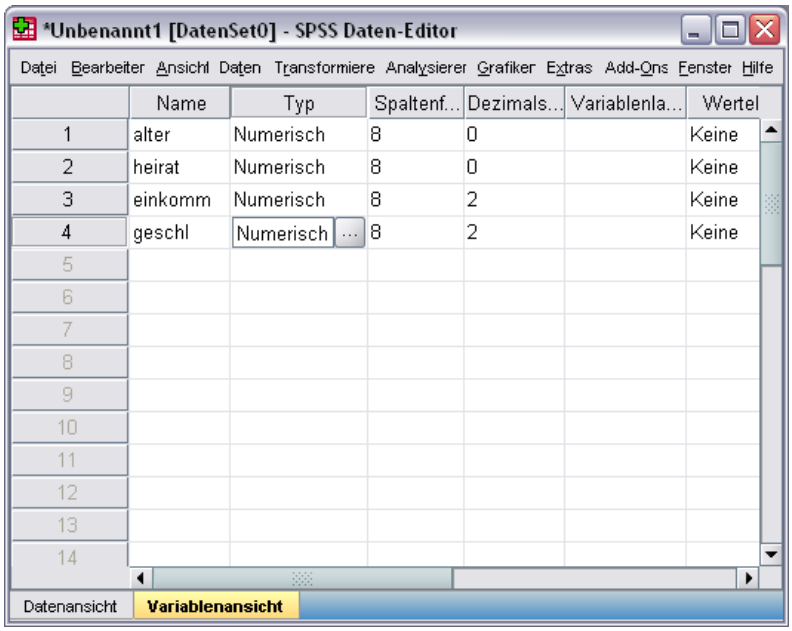
▶ Klicken Sie am unteren Rand des Fensters "Daten-Editor" auf die Registerkarte Variablenansicht.

▶ Geben Sie in der ersten Zelle der ersten Leerzeile als Variablennamen geschl ein.

▶ Klicken Sie auf die Zelle *Typ* neben Ihrer Eingabe.

▶ Klicken Sie auf die Schaltfläche rechts neben der Zelle *Typ*, um das Dialogfeld “Variablentyp definieren” zu öffnen.

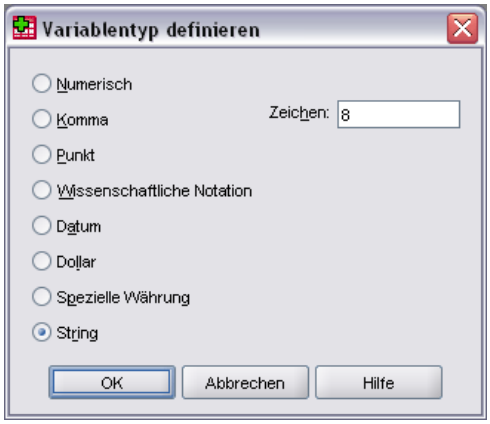
Abbildung 3-4
In der Zelle “Typ” für “geschl” angezeigte Schaltfläche



▶ Wählen Sie String aus, um den Variablentyp anzugeben.

▶ Klicken Sie auf OK, um die Auswahl zu speichern und zum Daten-Editor zurückzukehren.

Abbildung 3-5
Dialogfeld “Variablentyp definieren”



Definieren von Daten

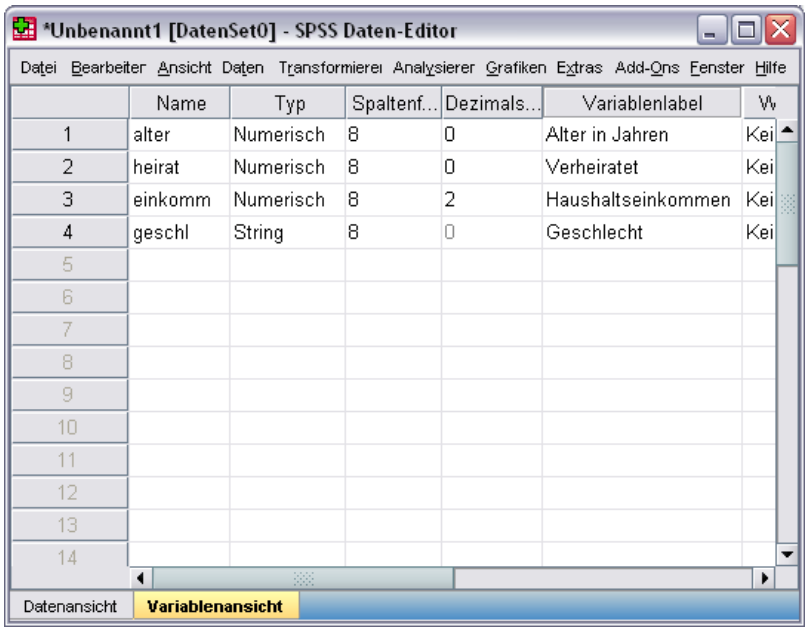
Neben Datentypen können Sie auch aussagekräftige Variablenlabels und Wertelabels für Variablennamen und Datenwerte definieren. Diese aussagekräftigen Labels werden in statistischen Berichten und Diagrammen verwendet.

Hinzufügen von Variablenlabels

Labels sollen Beschreibungen von Variablen darstellen. Diese Beschreibungen sind oft längere Versionen von Variablennamen. Labels können bis zu 255 Byte umfassen. Diese Labels werden in der Ausgabe verwendet, um die Variablen zu identifizieren.

- ▶ Klicken Sie am unteren Rand des Fensters “Daten-Editor” auf die Registerkarte Variablenansicht.
- ▶ Geben Sie Alter in Jahren in der Spalte *Variablenlabel* der Zeile *alter* ein.
- ▶ Geben Sie Verheiratet in der Spalte *Variablenlabel* der Zeile *heirat* ein.
- ▶ Geben Sie Haushaltseinkommen in der Spalte *Variablenlabel* der Zeile *einkomme* ein.
- ▶ Geben Sie Geschlecht in der Spalte *Variablenlabel* der Zeile *geschl* ein.

Abbildung 3-6
In der Variablenansicht eingegebene Variablenlabels



	Name	Typ	Spaltenf...	Dezimals...	Variablenlabel	W
1	alter	Numerisch	8	0	Alter in Jahren	Kei
2	heirat	Numerisch	8	0	Verheiratet	Kei
3	einkomm	Numerisch	8	2	Haushaltseinkommen	Kei
4	geschl	String	8	0	Geschlecht	Kei
5						
6						
7						
8						
9						
10						
11						
12						
13						
14						

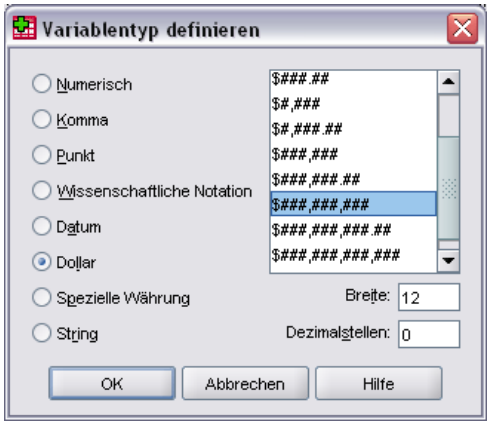
Ändern des Datentyps und Formats von Variablen

In der Spalte *Typ* wird der aktuelle Datentyp für jede Variable angezeigt. Meist werden die Datentypen “Numerisch” und “String” verwendet, aber es werden viele weitere Formate unterstützt. In der aktuellen Datendatei ist die Variable *einkomm* als numerischer Typ definiert.

▶ Klicken Sie auf die Zelle *Typ* in der Zeile *einkomme* und klicken Sie dann auf die Schaltfläche auf der rechten Seite der Zelle, um das Dialogfeld “Variablentyp definieren” zu öffnen.

▶ Wählen Sie Dollar aus.

Abbildung 3-7
Dialogfeld “Variablentyp definieren”



Die Formatierungsoptionen für den derzeit ausgewählten Datentyp werden angezeigt.

▶ Wählen Sie für das Währungsformat in diesem Beispiel $###,###,### aus.

▶ Klicken Sie auf OK, um die Änderungen zu speichern.

Hinzufügen von Wertelabels für numerische Variablen

Mit Wertelabels können Variablenwerte einem String-Label zugeordnet werden. In diesem Beispiel sind für die Variable *heirat* zwei Werte zulässig. 0 bedeutet, dass die Person ledig ist, und 1 bedeutet, dass sie verheiratet ist.

▶ Klicken Sie auf die Zelle *Wertelabels* in der Zeile *heirat* und klicken Sie dann auf die Schaltfläche auf der rechten Seite der Zelle, um das Dialogfeld “Wertelabels definieren” zu öffnen.

Wert ist der eigentliche numerische Wert.

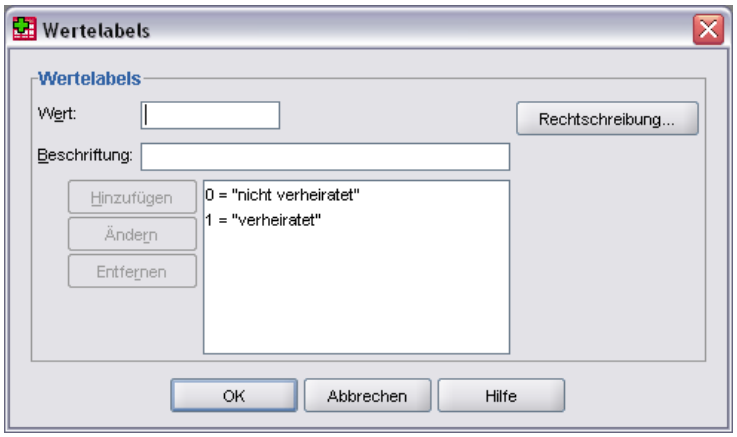
Wertelabel ist das String-Label, das dem angegebenen numerischen Wert zugeordnet wird.

▶ Geben Sie 0 im Feld “Wert” ein.

▶ Geben Sie nicht verheiratet im Feld “Label” ein.

▶ Klicken Sie auf Hinzufügen, um dieses Label der Liste hinzuzufügen.

Abbildung 3-8
Dialogfeld "Wertelabels definieren"



▶ Geben Sie im Feld "Wert" den Wert 1 und anschließend im Feld "Label" den Wert Verheiratet ein.

▶ Klicken Sie auf Hinzufügen und anschließend auf OK, um die Änderungen zu speichern und zum Daten-Editor zurückzukehren.

Diese Labels können auch in der Datenansicht angezeigt werden, wodurch die Daten u. U. besser lesbar werden.

▶ Klicken Sie am unteren Rand des Fensters "Daten-Editor" auf die Registerkarte Datenansicht.

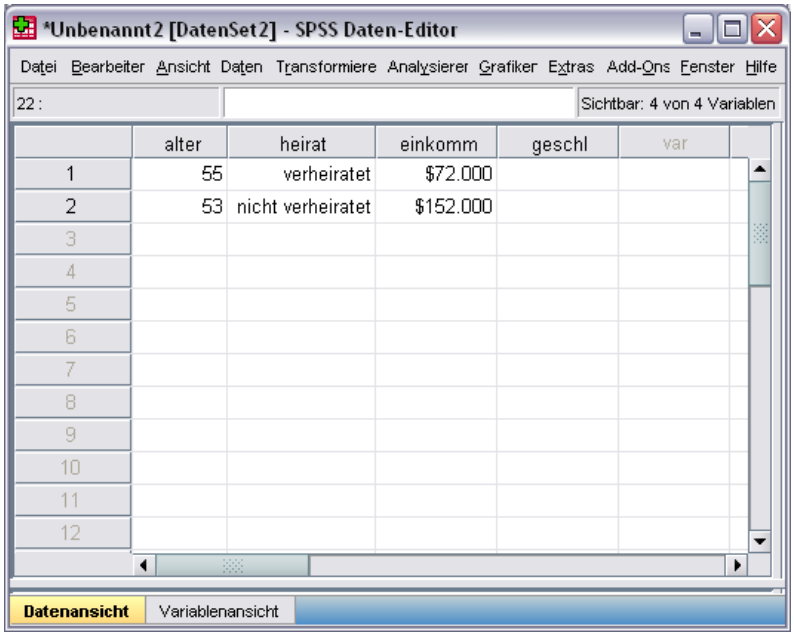
▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Ansicht
  Wertelabels

Die Labels werden nun in einer Liste angezeigt, wenn Sie im Daten-Editor Werte eingeben. Dies hat den Vorteil, dass zulässige Antworten vorgeschlagen werden und die Antwort aussagekräftiger ist.

Wenn das Menüelement “Wertelabels” bereits aktiviert wurde (neben dem Menüelement wird ein Häkchen angezeigt) und Sie erneut Wertelabels auswählen, wird die Anzeige von Wertelabels *deaktiviert*.

Abbildung 3-9
Wertelabels in der Datenansicht



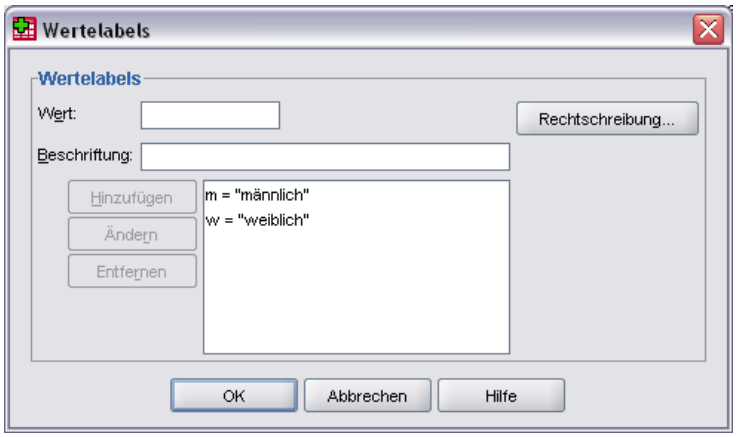
Hinzufügen von Wertelabels für String-Variablen

Auch bei String-Variablen können Wertelabels notwendig sein. Angenommen, in den Daten werden die Buchstaben *M* und *F* zur Bezeichnung des Geschlechts einer Person verwendet. Dann kann durch Wertelabels angegebenen werden, dass *MMännlich* und *FWeiblich* bedeutet.

- Klicken Sie am unteren Rand des Fensters “Daten-Editor” auf die Registerkarte Variablenansicht.
- Klicken Sie auf die Zelle *Wertelabels* in der Zeile *geschl* und klicken Sie dann auf die Schaltfläche auf der rechten Seite der Zelle, um das Dialogfeld “Wertelabels definieren” zu öffnen.
- Geben Sie im Feld “Wert” den Wert W und anschließend im Feld “Label” den Wert Weiblich ein.

▶ Klicken Sie auf Hinzufügen, um dieses Label der Datendatei hinzuzufügen.

Abbildung 3-10
Dialogfeld "Wertelabels definieren"



▶ Geben Sie im Feld "Wert" den Wert M und anschließend im Feld "Label" den Wert Männlich ein.

▶ Klicken Sie auf Hinzufügen und anschließend auf OK, um die Änderungen zu speichern und zum Daten-Editor zurückzukehren.

Achten Sie darauf, dass die Schreibweise einheitlich ist, da bei String-Werten zwischen Groß- und Kleinschreibung unterschieden wird. Es wird zwischen einem kleinen *m* und einem großen *M* unterschieden.

Verwenden von Wertelabels bei der Dateneingabe

Wertelabels können zur Dateneingabe verwendet werden.

▶ Klicken Sie am unteren Rand des Fensters "Daten-Editor" auf die Registerkarte Datenansicht.

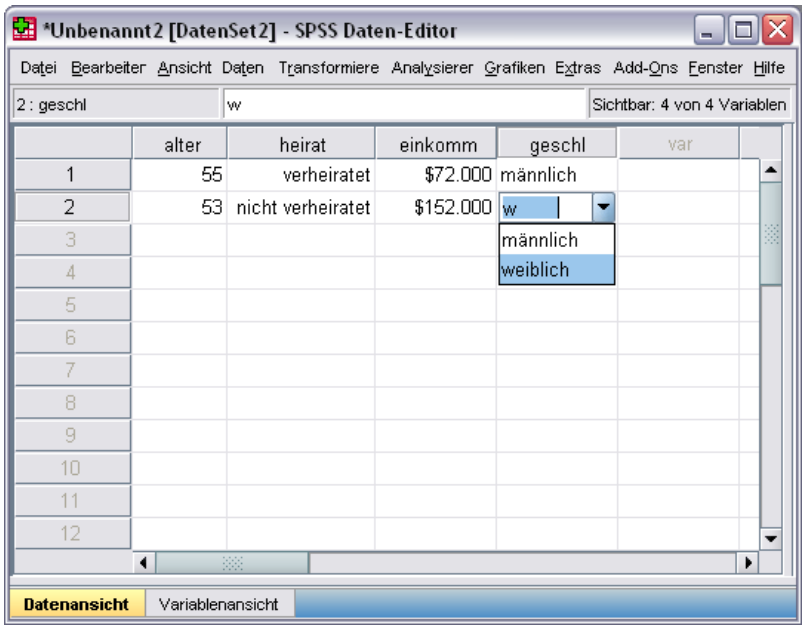
▶ Wählen Sie in der ersten Zeile die Zelle für *geschl* aus.

▶ Klicken Sie auf die Schaltfläche rechts neben der Zelle und wählen Sie aus der Dropdown-Liste den Wert Männlich aus.

▶ Wählen Sie in der zweiten Zeile die Zelle für *geschl* aus.

▶ Klicken Sie auf die Schaltfläche rechts neben der Zelle und wählen Sie aus der Dropdown-Liste den Wert Weiblich aus.

Abbildung 3-11
Verwendung von Wertelabels zur Auswahl von Werten



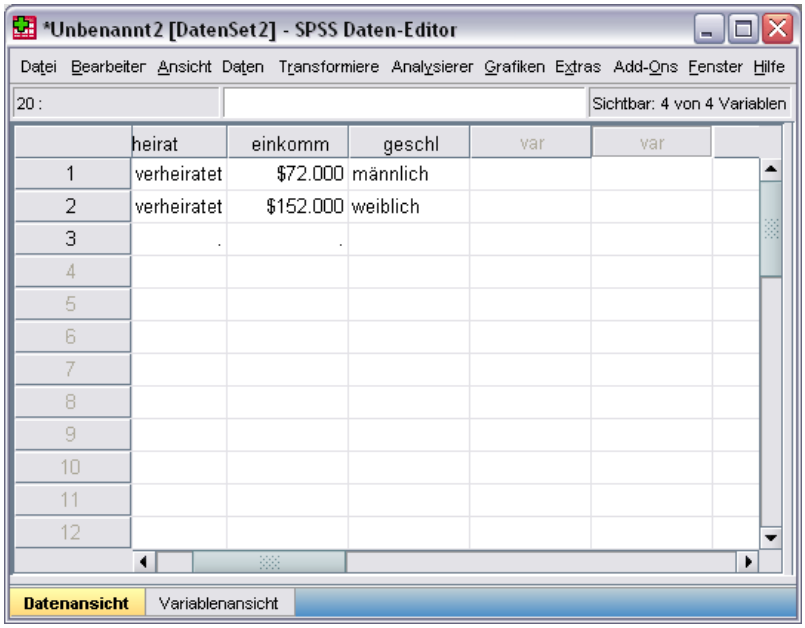
Es werden nur definierte Werte aufgelistet. Dadurch wird sichergestellt, dass die Daten im erwarteten Format eingegeben werden.

Umgang mit fehlenden Daten

Fehlende oder ungültige Werte stellen ein häufig auftretendes Problem dar, das nicht vernachlässigt werden darf. Die Teilnehmer an einer Umfrage antworten vielleicht absichtlich nicht auf eine Frage oder sie wissen keine Antwort oder die Form der Antwort ist unerwartet. Wenn Sie keine Maßnahmen zum Filtern oder Identifizieren dieser Daten treffen, ist das Ergebnis der Analyse möglicherweise ungenau.

Bei numerischen Daten werden leere Datenfelder oder Felder mit ungültiger Eingabe in systemdefiniert fehlende Felder konvertiert. Dies wird durch einen einzelnen Punkt gekennzeichnet.

Abbildung 3-12
Als Punkte angezeigte fehlende Werte



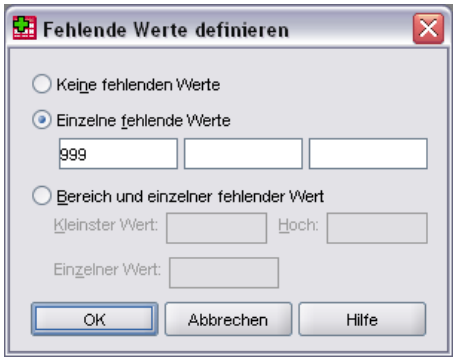
Für die Analyse kann es wichtig sein zu wissen, warum ein Wert fehlt. Es könnte beispielsweise wichtig sein, zwischen Personen zu unterscheiden, die eine Frage nicht beantworten wollten, und solchen, die nicht geantwortet haben, weil die Frage sie nicht betraf.

Fehlende Werte für eine numerische Variable

- Klicken Sie am unteren Rand des Fensters “Daten-Editor” auf die Registerkarte Variablenansicht.
- Klicken Sie auf die Zelle *Fehlende Werte* in der Zeile *alter* und klicken Sie dann auf die Schaltfläche auf der rechten Seite der Zelle, um das Dialogfeld “Fehlende Werte definieren” zu öffnen.

  In diesem Dialogfeld können Sie bis zu drei verschiedene fehlende Werte oder einen Wertebereich und zusätzlich einen einzelnen Wert angeben.

Abbildung 3-13
Dialogfeld “Fehlende Werte definieren”

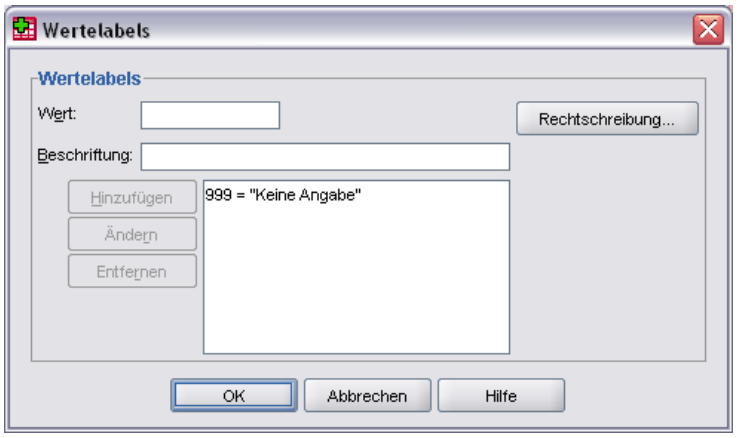


- Wählen Sie Einzelne fehlende Werte aus.
- Geben Sie im ersten Textfeld 999 ein und lassen Sie die beiden Textfelder anderen leer.
- Klicken Sie auf OK, um die Änderungen zu speichern und zum Daten-Editor zurückzukehren.

Nachdem nun der Wert für fehlende Daten definiert wurde, kann diesem Wert ein Label zugeordnet werden.

- Klicken Sie auf die Zelle *Wertelabels* in der Zeile *alter* und klicken Sie dann auf die Schaltfläche auf der rechten Seite der Zelle, um das Dialogfeld "Wertelabels definieren" zu öffnen.
- Geben Sie 999 im Feld "Wert" ein.
- Geben Sie Keine Angabe im Feld "Label" ein.

Abbildung 3-14
Dialogfeld "Wertelabels definieren"



- Klicken Sie auf Hinzufügen, um dieses Label der Datendatei hinzuzufügen.
- Klicken Sie auf OK, um die Änderungen zu speichern und zum Daten-Editor zurückzukehren.

Fehlende Werte für eine String-Variable

Fehlende Werte für String-Variablen werden ähnlich wie solche für numerische Werte behandelt. Im Gegensatz zu numerischen Variablen werden Leerfelder bei String-Variablen jedoch nicht als systemdefiniert fehlend betrachtet. Sie werden statt dessen als Leerstring interpretiert.

- Klicken Sie am unteren Rand des Fensters "Daten-Editor" auf die Registerkarte Variablenansicht.
- Klicken Sie auf die Zelle *Fehlende Werte* in der Zeile *geschl* und klicken Sie dann auf die Schaltfläche auf der rechten Seite der Zelle, um das Dialogfeld "Fehlende Werte definieren" zu öffnen.
- Wählen Sie Einzelne fehlende Werte aus.
- Geben Sie KA im ersten Textfeld ein.

Bei fehlenden Werten für String-Variablen wird zwischen Groß- und Kleinschreibung unterschieden. Der Wert *na* wird daher nicht als fehlender Wert behandelt.

▶ Klicken Sie auf OK, um die Änderungen zu speichern und zum Daten-Editor zurückzukehren.

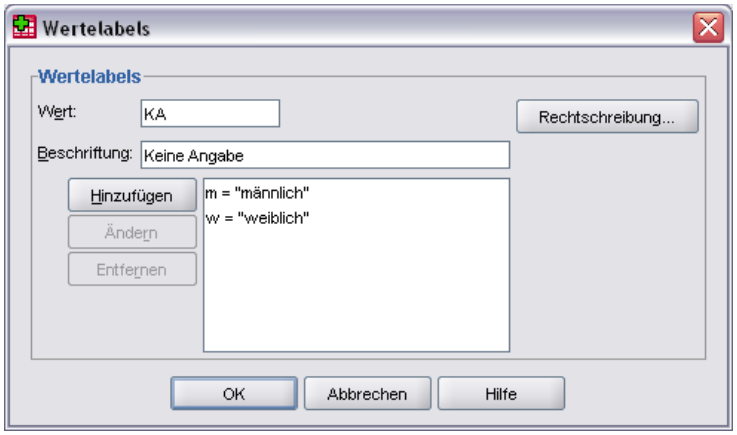
Jetzt können Sie ein Label für den fehlenden Wert definieren.

▶ Klicken Sie auf die Zelle *Wertelabels* in der Zeile *geschl* und klicken Sie dann auf die Schaltfläche auf der rechten Seite der Zelle, um das Dialogfeld “Wertelabels definieren” zu öffnen.

▶ Geben Sie KA im Feld “Wert” ein.

▶ Geben Sie Keine Angabe im Feld “Label” ein.

Abbildung 3-15
Dialogfeld “Wertelabels definieren”



▶ Klicken Sie auf Hinzufügen, um dieses Label zu Ihrem Projekt hinzuzufügen.

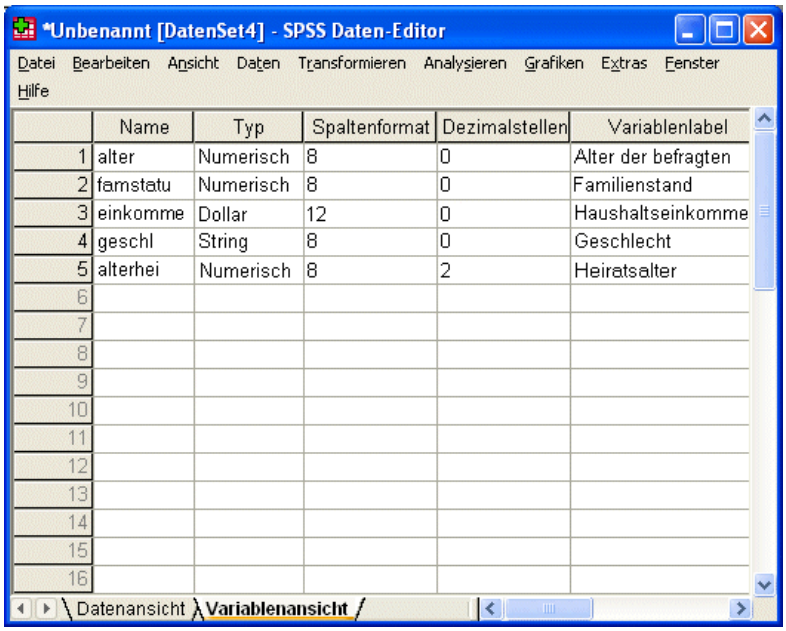
▶ Klicken Sie auf OK, um die Änderungen zu speichern und zum Daten-Editor zurückzukehren.

Kopieren und Einfügen von Variablenattributen

Nachdem Sie Variablenattribute für Variablen definiert haben, können Sie diese Attribute kopieren und anderen Variablen zuweisen.

▶ Geben Sie in der Variablenansicht in der ersten Zelle der ersten leeren Zeile alterhei ein.

Abbildung 3-16
Variable "alterhei" in der Variablenansicht



	Name	Typ	Spaltenformat	Dezimalstellen	Variablenlabel
1	alter	Numerisch	8	0	Alter der befragten
2	famstatu	Numerisch	8	0	Familienstand
3	einkomme	Dollar	12	0	Haushaltseinkomme
4	geschl	String	8	0	Geschlecht
5	alterhei	Numerisch	8	2	Heiratsalter
6					
7					
8					
9					
10					
11					
12					
13					
14					
15					
16					

▶ Geben Sie in der Spalte *Variablenlabel* die Benennung Heiratsalter ein.

▶ Klicken Sie in der Zeile *alter* auf die Zelle *Wertelabels*.

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Bearbeiten
  Kopieren

▶ Klicken Sie in der Zeile *alterhei* auf die Zelle *Wertelabels*.

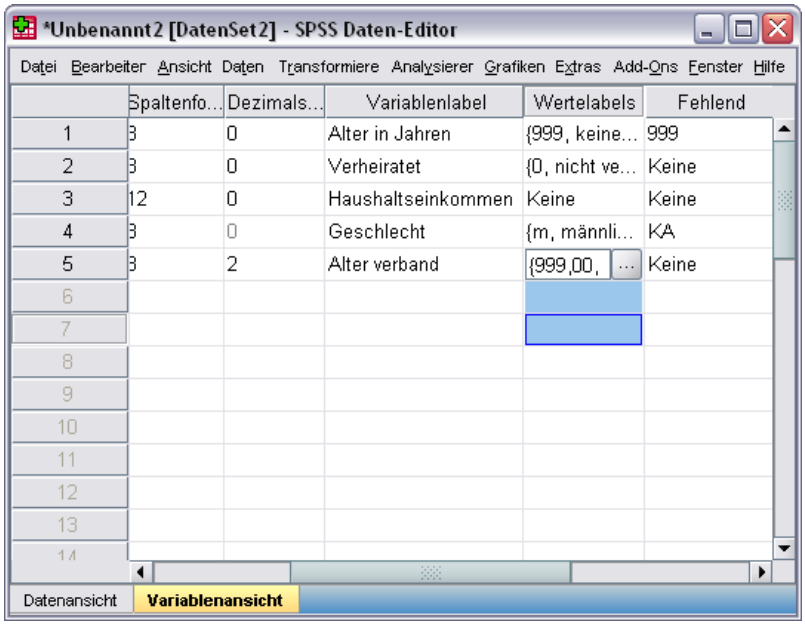
▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Bearbeiten
  Einfügen

Die für die Variable *alter* definierten Werte werden nun der Variablen *alterhei* zugewiesen.

Um die Attribute mehreren Variablen gleichzeitig zuzuweisen, wählen Sie als Ziel mehrere Zellen gleichzeitig aus. (Klicken Sie dazu mit der Maus und ziehen Sie den Mauszeiger bei gedrückter Maustaste die Spalte hinunter.)

Abbildung 3-17
Mehrere Zellen ausgewählt



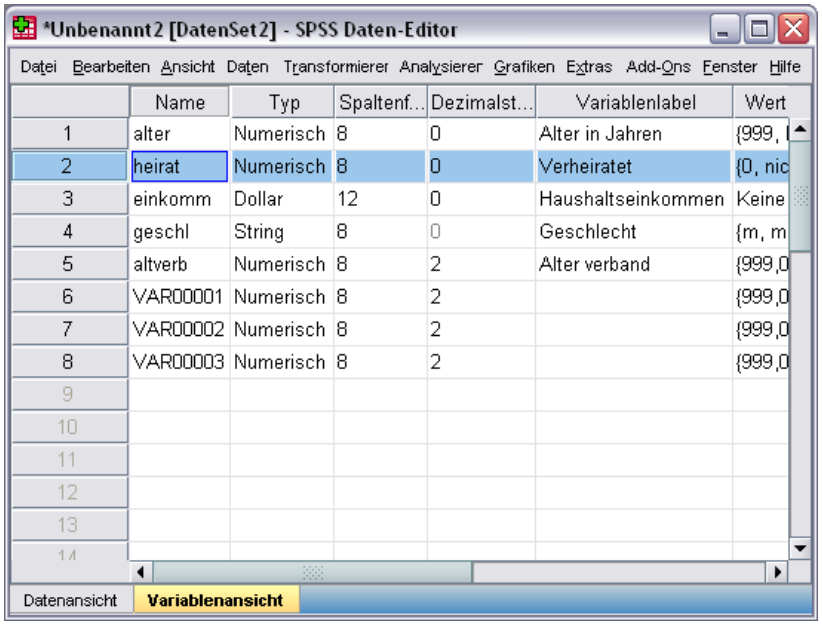
Wenn Sie das Attribut einfügen, wird es allen ausgewählten Zellen zugewiesen.

Beim Einfügen der Werte in leere Zeilen werden automatisch neue Variablen generiert.

So kopieren Sie alle Attribute aus einer Variablen in eine andere Variable:

▶ Klicken Sie in der Zeile *heirat* auf die Zeilennummer.

Abbildung 3-18
Ausgewählte Zeile



	Name	Typ	Spaltenf...	Dezimalst...	Variablenlabel	Wert
1	alter	Numerisch	8	0	Alter in Jahren	{999,
2	heirat	Numerisch	8	0	Verheiratet	{0, nic
3	einkomm	Dollar	12	0	Haushaltseinkommen	Keine
4	geschl	String	8	0	Geschlecht	{m, m
5	altverb	Numerisch	8	2	Alter verband	{999,0
6	VAR00001	Numerisch	8	2		{999,0
7	VAR00002	Numerisch	8	2		{999,0
8	VAR00003	Numerisch	8	2		{999,0
9						
10						
11						
12						
13						
14						

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Bearbeiten
  Kopieren

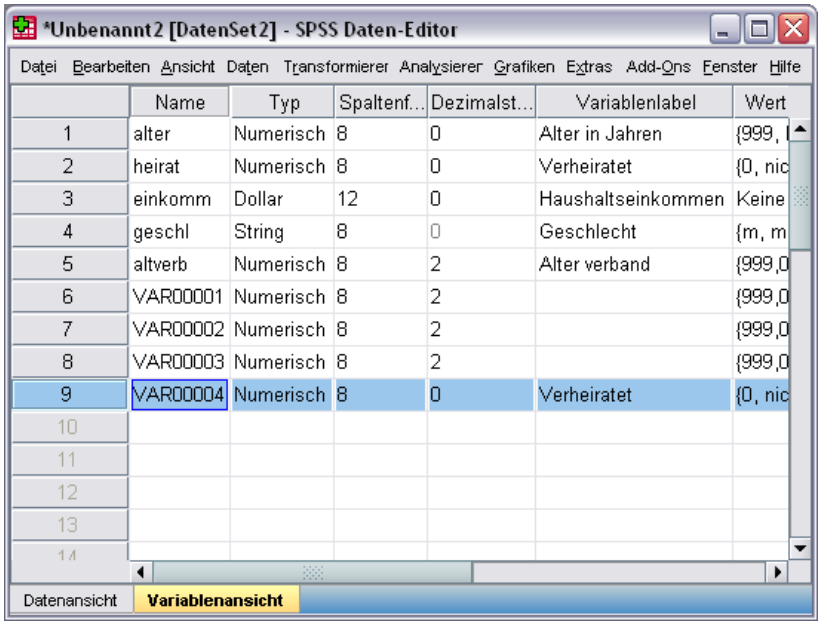
▶ Klicken Sie auf die Zeilennummer der ersten leeren Zeile.

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Bearbeiten
  Einfügen

Alle Attribute der Variablen *heirat* werden der neuen Variablen zugewiesen.

Abbildung 3-19
Alle Werte wurden in die Zeile eingefügt.



	Name	Typ	Spaltenf...	Dezimalst...	Variablenlabel	Wert
1	alter	Numerisch	8	0	Alter in Jahren	{999,
2	heirat	Numerisch	8	0	Verheiratet	{0, nic
3	einkomm	Dollar	12	0	Haushaltseinkommen	Keine
4	geschl	String	8	0	Geschlecht	{m, m
5	altverb	Numerisch	8	2	Alter verband	{999,0
6	VAR00001	Numerisch	8	2		{999,0
7	VAR00002	Numerisch	8	2		{999,0
8	VAR00003	Numerisch	8	2		{999,0
9	VAR00004	Numerisch	8	0	Verheiratet	{0, nic
10						
11						
12						
13						
14						

Definieren von Variableneigenschaften für kategoriale Variablen

Bei kategorialen (nominalen, ordinalen) Daten können Sie mithilfe von “Variableneigenschaften definieren” Wertelabels und andere Variableneigenschaften definieren. Die Prozedur “Variableneigenschaften definieren” ermöglicht Folgendes:

- Durchsuchen der tatsächlichen Datenwerte und Auflisten aller eindeutigen Datenwerte für die ausgewählten Variablen.
- Ermitteln der Werte ohne Label und Bereitstellen einer Funktion zur automatischen Beschriftung.
- Kopieren definierter Wertelabels aus einer anderen Variablen in die ausgewählte Variable oder von der ausgewählten Variablen in andere Variablen.

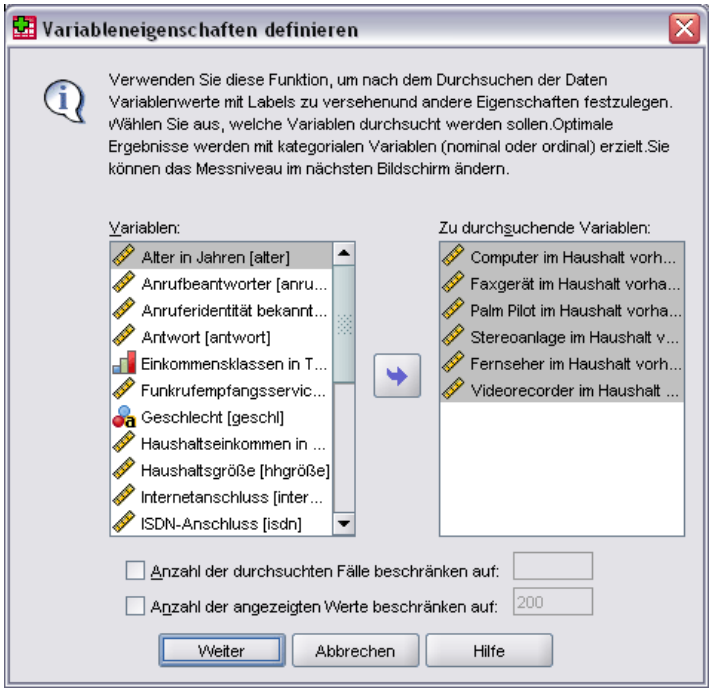
In diesem Beispiel wird die Datendatei *demo.sav* verwendet. Für diese Datendatei liegen bereits definierte Wertelabels vor. Daher soll ein Wert eingegeben werden, für den noch kein Wertelabel definiert ist:

▶ Klicken Sie in der Datenansicht des Daten-Editors auf die erste Datenzelle für die Variable *computer* (Sie müssen möglicherweise einen Bildlauf nach rechts durchführen) und geben Sie anschließend den Wert 99 ein.

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Daten
  Variableneigenschaften definieren...

Abbildung 3-20
Anfangsdialogfeld für "Variableneigenschaften definieren"



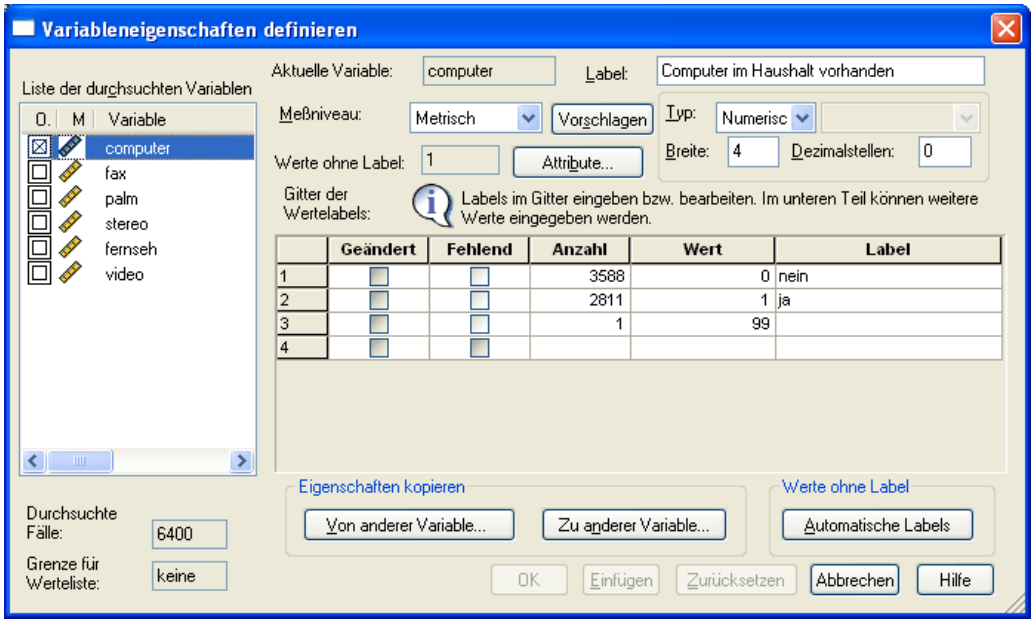
Im ersten Dialogfeld für "Variableneigenschaften definieren" wählen Sie die nominalen oder ordinalen Variablen aus, für die Sie Wertelabels und/oder andere Eigenschaften definieren möchten.

▶ Verschieben Sie die Elemente *Computer im Haushalt vorhanden [computer]* bis *Videorekorder im Haushalt vorhanden [video]* mit Ziehen und Ablegen in die Liste der zu durchsuchenden Variablen.

Sie haben möglicherweise bemerkt, dass die Symbole für das Messniveau bei allen ausgewählten Variablen angeben, dass es sich um metrische Variablen und nicht um kategoriale Variablen handelt. Bei allen ausgewählten Variablen in diesem Beispiel handelt es sich in Wirklichkeit um kategoriale Variablen, bei denen die numerischen Werte "0" und "1" für *Nein* bzw. *Ja* stehen – und eine der Variableneigenschaften, die mithilfe von "Variableneigenschaften definieren" geändert wird, ist das Messniveau.

▶ Klicken Sie auf Weiter.

Abbildung 3-21
Hauptdialogfeld "Variableneigenschaften definieren"



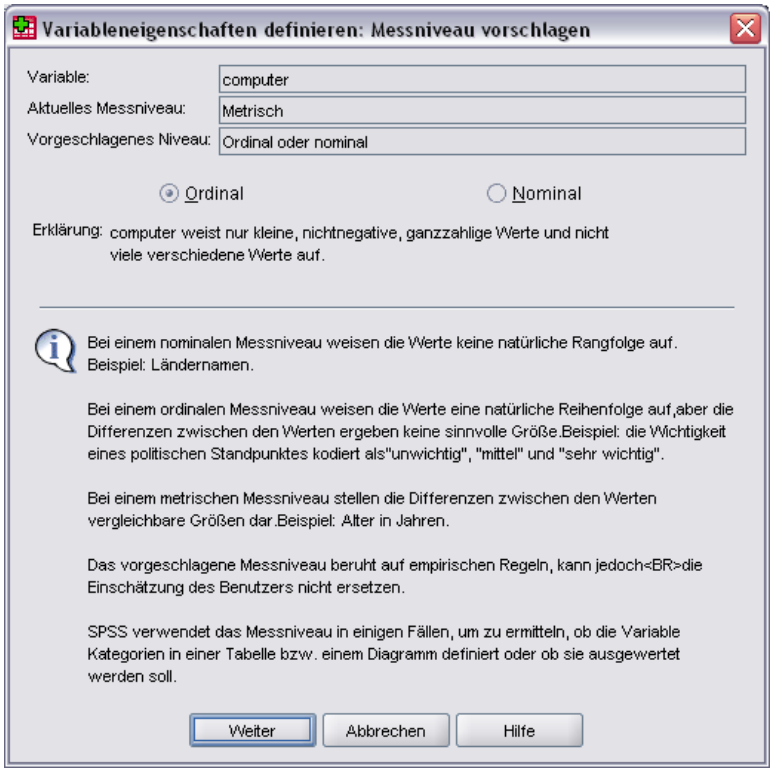
▶ Wählen Sie in der Liste der durchsuchten Variablen den Eintrag *computer* aus.

Das derzeitige Messniveau für die ausgewählte Variable ist metrisch. Sie können das Messniveau ändern, indem Sie ein Niveau aus der Dropdown-Liste auswählen. Alternativ können Sie ein Messniveau von "Variableneigenschaften definieren" vorschlagen lassen.

▶ Klicken Sie auf Vorschlagen.

Das Dialogfeld "Messniveau vorschlagen" wird geöffnet.

Abbildung 3-22
Dialogfeld "Messniveau vorschlagen"



Da die Variable nicht sehr viele verschiedene Werte aufweist und alle durchsuchten Fälle ganzzahlige Werte enthalten, ist das geeignete Messniveau vermutlich ordinal oder nominal.

▶ Wählen Sie Ordinal aus und klicken Sie anschließend auf Weiter.

Das Messniveau für die ausgewählte Variable ist nun ordinal.

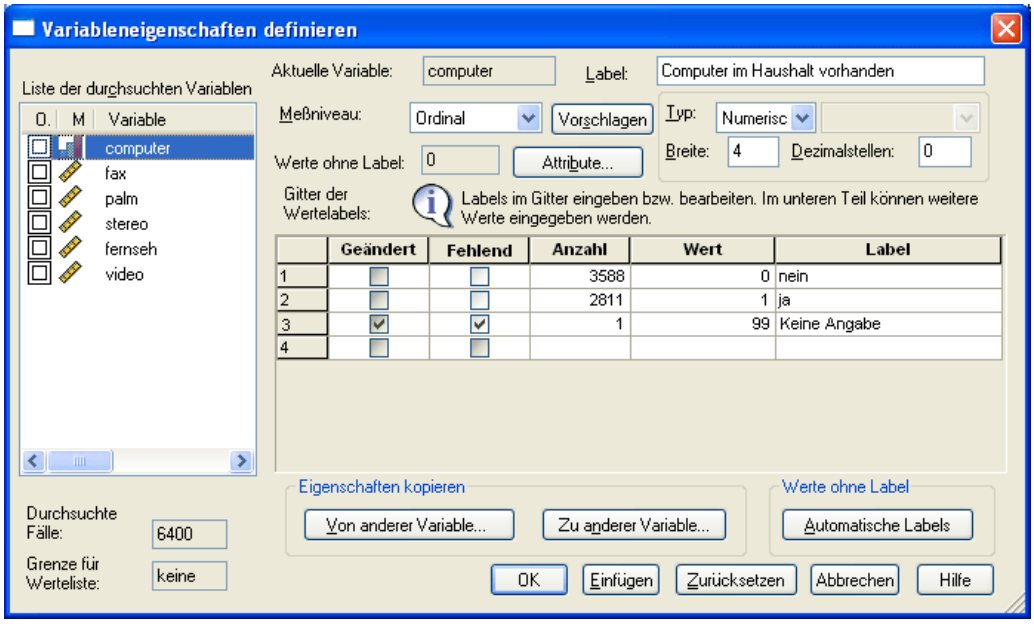
Im Gitter der Wertelabels werden alle eindeutigen Datenwerte für die ausgewählte Variable, alle definierten Wertelabels für diese Werte und die Anzahl der Vorkommnisse jedes Werts in den durchsuchten Fällen angezeigt.

Der in der Datenansicht eingegebene Wert, 99, wird im Gitter angezeigt. Die Anzahl beträgt nur 1, da nur der Wert für einen Fall geändert wurde, und die Spalte *Label* ist leer, da noch kein Wertelabel für 99 definiert wurde. Ein X in der ersten Spalte in der Liste der durchsuchten Variablen gibt außerdem an, dass die ausgewählte Variable mindestens einen beobachteten Wert ohne definiertes Datenlabel aufweist.

▶ Geben Sie in der Spalte *Label* für den Wert "99" Keine Angabe ein.

▶ Aktivieren Sie das Kontrollkästchen in der Spalte *Fehlende Werte* für den Wert 99, um den Wert 99 als **benutzerdefiniert fehlend** zu kennzeichnen.
Datenwerte, die als benutzerdefiniert fehlende Werte angegeben sind, werden zur Sonderbehandlung gekennzeichnet und von den meisten Berechnungen ausgeschlossen.

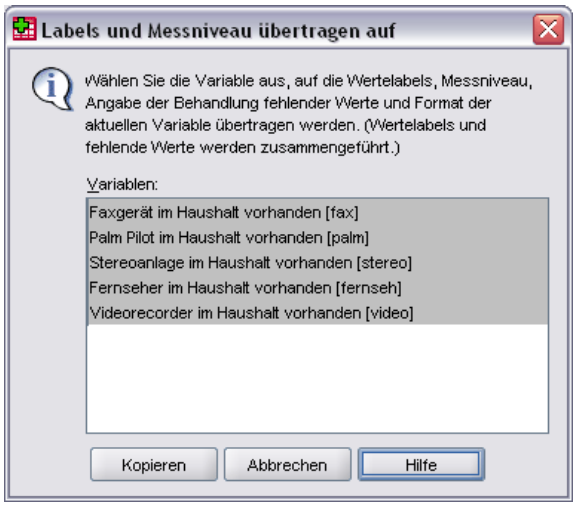
Abbildung 3-23
Neue Variableneigenschaften für "computer" definiert.



Bevor Sie die Bearbeitung der Variableneigenschaften für *computer* abschließen, sollten Sie den anderen Variablen in der Liste dasselbe Messniveau, dieselben Wertelabels und dieselben Definitionen für fehlende Werte zuweisen.

▶ Klicken Sie im Gruppenfeld "Eigenschaften kopieren" auf Zu anderer Variable.

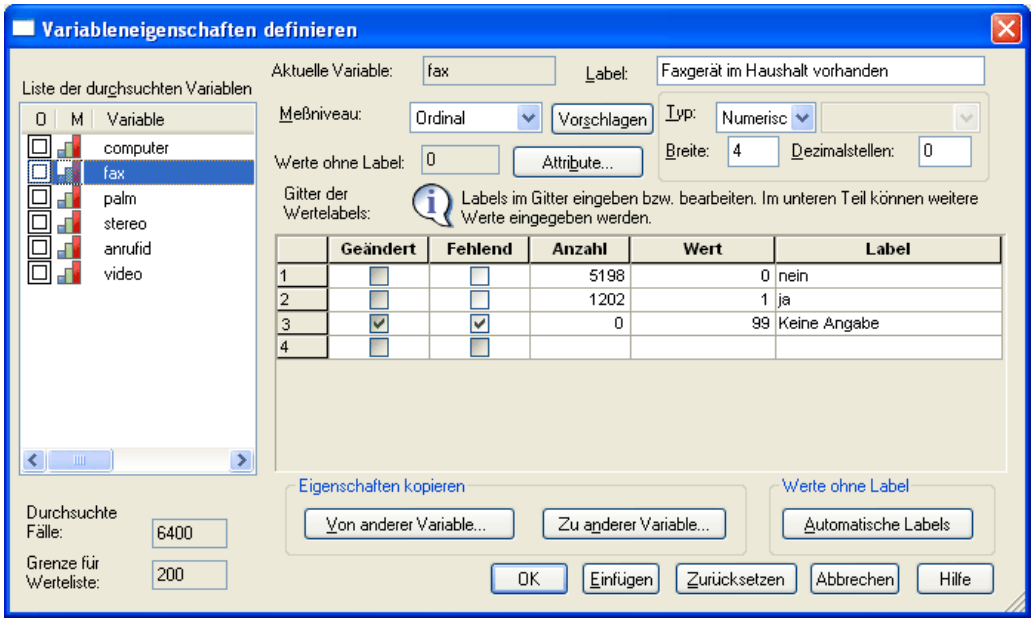
Abbildung 3-24
Dialogfeld "Labels und Messniveau übertragen"



▶ Wählen Sie im Dialogfeld "Labels und Messniveau übertragen" alle Variablen in der Liste aus und klicken Sie anschließend auf Kopieren.

Wenn Sie jetzt in der Liste der durchsuchten Variablen im Hauptdialogfeld von “Variableneigenschaften definieren” eine andere Variable auswählen, stellen Sie fest, dass alle Variablen ordinal sind, wobei der Wert “99” als benutzerdefinierter fehlender Wert mit dem Wertelabel *Keine Angabe* definiert ist.

Abbildung 3-25
Neue definierte Variableneigenschaften für “fax”.



▶ Klicken Sie auf OK, um alle Variableneigenschaften zu speichern, die Sie definiert haben.

Kapitel

4

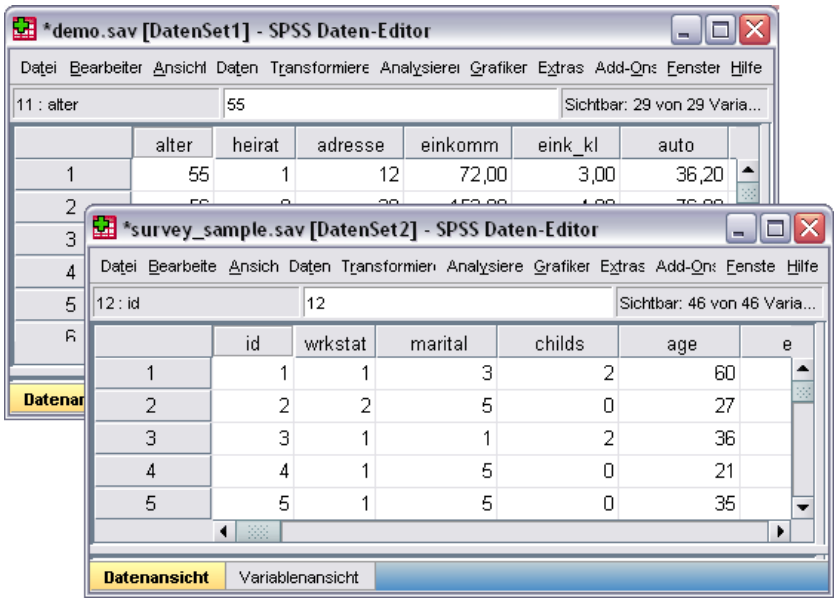
Arbeiten mit mehreren Datenquellen

Ab Version 14.0 können mehrere Datenquellen gleichzeitig geöffnet sein. Dies vereinfacht Folgendes:

- Wechseln zwischen Datenquellen
- Vergleichen der verschiedenen Datenquellen
- Kopieren und Einfügen von Daten zwischen Datenquellen
- Erstellen von Teilmengen der Fälle und/oder Variablen für die Analyse
- Zusammenführen von verschiedenen Datenquellen mit unterschiedlichen Datenformaten (beispielsweise Tabellenkalkulationsblätter, Datenbanken, Textdaten), ohne dass zuerst jede Datenquelle gespeichert werden muss.

Grundsätzlicher Umgang mit mehreren Datenquellen

Abbildung 4-1
Zwei gleichzeitig geöffnete Datenquellen

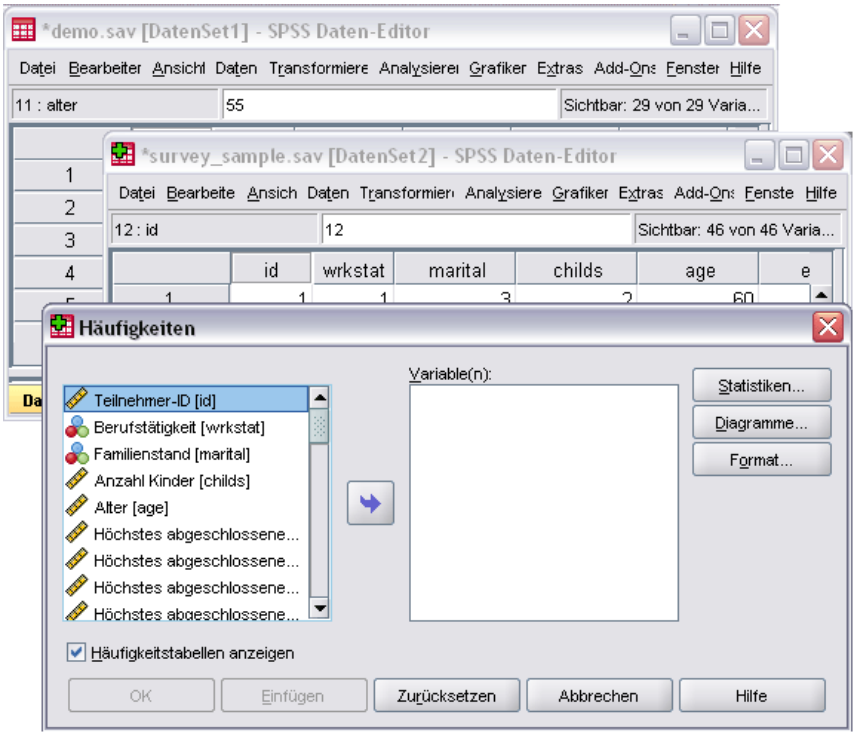


Standardmäßig wird jede Datenquelle, die Sie öffnen, in einem neuen Fenster des Daten-Editors angezeigt.

- Jede zuvor geöffnete Datenquelle bleibt geöffnet und für die weitere Verwendung verfügbar.

- Wenn Sie eine Datenquelle öffnen, wird diese automatisch zur **Arbeitsdatei**.
- Sie können die Arbeitsdatei wechseln, indem Sie auf eine beliebige Stelle im Fenster "Daten-Editor" der gewünschten Datenquelle klicken oder das Fenster "Daten-Editor" für diese Datenquelle aus dem Menü "Fenster" auswählen.
- Nur die Variablen in der Arbeitsdatei sind für die Analyse verfügbar.

Abbildung 4-2
Variablenliste mit Variablen aus der Arbeitsdatei



- Sie können die Arbeitsdatei nicht wechseln, wenn ein Dialogfeld geöffnet ist, mit dem auf die Daten zugegriffen wird (einschließlich aller Dialogfelder, in denen Variablenlisten angezeigt werden).
- Während einer Sitzung muss mindestens ein Fenster des Daten-Editors geöffnet sein. Wenn Sie das letzte geöffnete Fenster des Daten-Editors schließen, wird SPSS automatisch beendet. Dabei werden Sie aufgefordert, die Änderungen zu speichern.

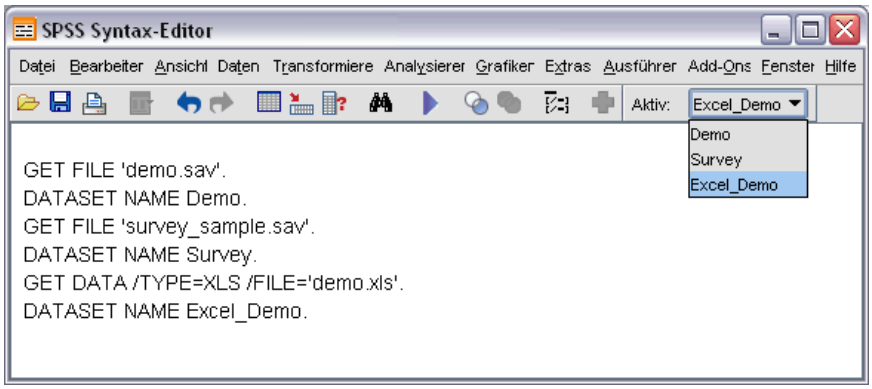
Arbeiten mit mehreren Daten-Sets in der Befehlssyntax

Wenn Sie die Befehlssyntax verwenden, um Datenquellen zu öffnen (z. B. `GET FILE` oder `GET DATA`), muss mithilfe von `DATASET NAME` jedes Daten-Set eindeutig benannt werden, damit mehr als eine Datenquelle gleichzeitig geöffnet werden kann.

Bei der Arbeit mit der Befehlssyntax wird der Name des aktiven Daten-Sets (Arbeitsdatei) in der Symbolleiste des Syntaxfensters angezeigt. Mit den folgenden Aktionen kann die Arbeitsdatei geändert werden:

- Über den Befehl `DATASET ACTIVATE`.
- Klicken Sie an eine beliebige Stelle im Fenster "Daten-Editor" eines Daten-Sets.
- Auswählen eines Daten-Set-Namens in der Symbolleiste im Syntaxfenster.

Abbildung 4-3
Offene Daten-Sets, die in der Symbolleiste des Syntaxfensters angezeigt werden.



Kopieren und Einfügen von Informationen zwischen Daten-Sets

Sowohl Daten als auch die Attribute zur Variablendefinition werden grundsätzlich auf die gleiche Art aus einem Daten-Set in ein anderes Daten-Set kopieren, in der Sie Informationen innerhalb einer einzelnen Datendatei kopieren und einfügen.

- Beim Kopieren und Einfügen von ausgewählten Datenzellen in der Datenansicht werden nur die Datenwerte eingefügt, nicht die Attribute zur Variablendefinition.
- Beim Kopieren und Einfügen einer vollständigen Variablen in der Datenansicht durch Auswählen der Variablen im Spaltenkopf werden alle Daten und alle Attribute zur Variablendefinition für diese Variable eingefügt.
- Beim Kopieren und Einfügen von Attributen zur Variablendefinition oder vollständigen Variablen in der Variablenansicht werden die ausgewählten Attribute (oder die vollständige Variablendefinition) eingefügt, nicht jedoch die Datenwerte.

Umbenennen von Daten-Sets

Wenn Sie eine Datenquelle über die Menüs und Dialogfelder öffnen, wird jeder Datenquelle automatisch der Name *DatenSetn* zugewiesen, wobei *n* eine fortlaufende Ganzzahl ist. Wenn Sie eine Datenquelle mit der Befehlssyntax öffnen, wird dem Daten-Set kein Name zugewiesen,

sofern Sie nicht ausdrücklich mit `DATASET NAME` einen Namen angeben. So vergeben Sie aussagekräftigere Namen für die Daten-Sets:

▶ Wählen Sie die folgenden Optionen für das Daten-Set, dessen Namen Sie ändern möchten, aus den Menüs im Fenster des Daten-Editors aus:

Datei
  Daten-Set umbenennen...

▶ Geben Sie einen neuen Namen für das Daten-Set ein, der den Regeln für Variablennamen entspricht.

Unterdrücken der Anzeige mehrerer Daten-Sets

Wenn Sie bevorzugen, dass jeweils nur ein einziges Daten-Set zur Verfügung steht und Sie die Funktion für mehrere Daten-Sets deaktivieren möchten, gehen Sie wie folgt vor:

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Bearbeiten
  Optionen...

▶ Klicken Sie auf die Registerkarte Allgemein.

Aktivieren Sie Jeweils nur ein Daten-Set öffnen.

Kapitel

5

Untersuchen von Auswertungsstatistiken für einzelne Variablen

In diesem Kapitel werden einfache Auswertungsmaße und der Einfluß des Messniveaus einer Variablen auf den zu verwendenden Statistiktyp behandelt. Folgende Datendatei wird verwendet: *demo.sav*. Für weitere Informationen siehe Beispieldateien in Anhang A auf S. 163.

Messniveau

Je nach Messniveau sind für verschiedene Datentypen unterschiedliche Auswertungsmaße geeignet:

Kategorial. Daten mit einer begrenzten Anzahl von eindeutigen Werten bzw. Kategorien (beispielsweise Geschlecht oder Familienstand). Auch als **qualitative Daten** bezeichnet. Bei kategorialen Variablen kann es sich um String-Variablen (alphanumerisch) oder um numerische Variablen handeln, bei denen numerische Codes zum Darstellen der Kategorien verwendet werden (beispielsweise 0 = *Ledig* und 1 = *Verheiratet*). Es gibt zwei grundlegende Arten von kategorialen Daten:

- **Nominal.** Kategoriale Daten, bei denen die Kategorien keine natürliche Reihenfolge aufweisen. So ist beispielsweise die Berufskategorie *Vertrieb* weder höher noch niedriger als die Kategorie *Marketing* oder *Forschung*.
- **Ordinal.** Kategoriale Daten, bei denen die Kategorien eine sinnvolle Reihenfolge aufweisen, aber keine Distanz zwischen den Kategorien bestimmt werden kann. So liegt bei den Werten *Hoch*, *Mittel* und *Niedrig* zwar beispielsweise eine Reihenfolge vor, eine Distanz zwischen den Werten kann jedoch nicht berechnet werden.

Metrisch. Daten, die auf einer Intervall- oder Verhältnisskala gemessen werden und bei denen die Datenwerte sowohl die Reihenfolge der Werte als auch die Distanz zwischen den Werten festlegen. So ist beispielsweise ein Gehalt von $ 72.195 höher als ein Gehalt von $ 52.398 und die Distanz zwischen den Werten beträgt $ 19.797. Auch als **quantitative Daten** oder **stetige Daten** bezeichnet.

Auswertungsmaße für kategoriale Daten

Bei kategorialen Daten ist das typischste Auswertungsmaß die Zahl bzw. der Prozentsatz an Fällen in den einzelnen Kategorien. Der **Modalwert** ist die Kategorie mit der größten Anzahl von Fällen. Wenn eine große Anzahl von Kategorien vorliegt, kann bei ordinalen Daten auch der **Median** (der Wert, ober- und unterhalb dessen die Hälfte aller Fälle angesiedelt sind) ein hilfreiches Auswertungsmaß sein.

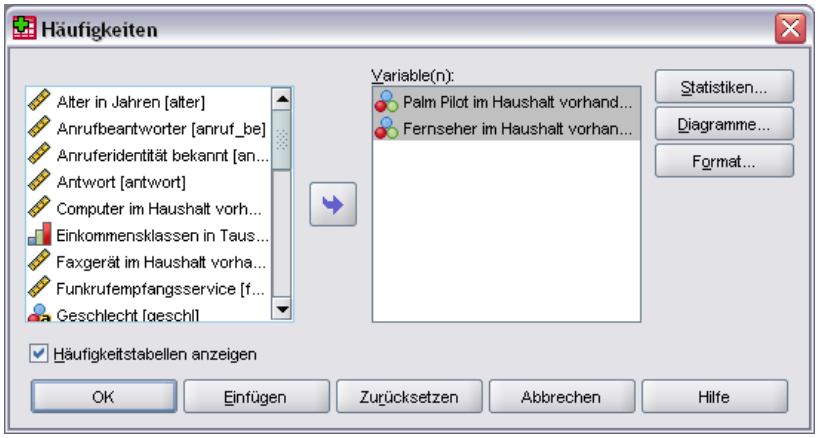
Mithilfe der Prozedur "Häufigkeiten" werden Häufigkeitstabellen erstellt, die für jeden bei einer Variablen beobachteten Wert sowohl die Zahl als auch den Prozentwert von Fällen anzeigen.

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Analysieren
  Deskriptive Statistiken
    Häufigkeiten...

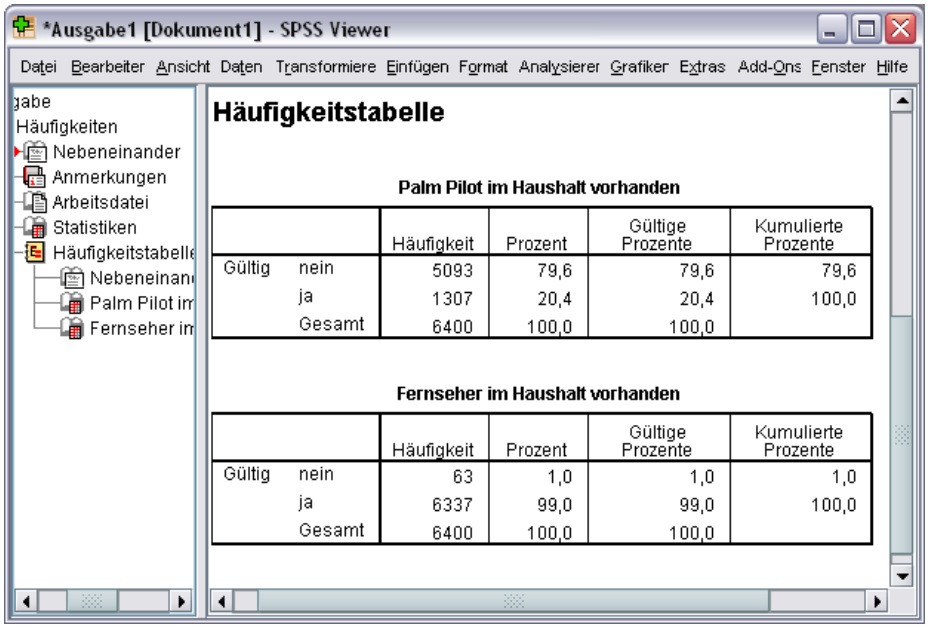
▶ Wählen Sie *Palm Pilot im Haushalt vorhanden [palm]* und *Fernseher im Haushalt vorhanden [fernseh]* aus und verschieben Sie diese in die Liste "Variable(n)".

Abbildung 5-1
Für die Analyse ausgewählte kategoriale Variablen



▶ Klicken Sie auf OK, um die Prozedur auszuführen.

Abbildung 5-2
Häufigkeitstabellen



Die Häufigkeitstabellen werden im Fenster “Viewer” angezeigt. Die Häufigkeitstabellen besagen, dass nur 20,4% der Personen Palm Pilots, jedoch fast alle (99,0%) einen Fernseher besitzen. Dies mag zwar an sich keine besonders interessante Erkenntnis sein, es könnte jedoch durchaus interessant sein, mehr über die kleine Gruppe von Personen zu erfahren, die keinen Fernseher besitzt.

Diagramme für kategoriale Daten

Die in einer Häufigkeitstabelle enthaltenen Informationen können mithilfe eines Balken- oder Kreisdiagramms grafisch dargestellt werden.

▶ Öffnen Sie das Dialogfeld “Häufigkeiten” erneut. (Die beiden Variablen sollten noch ausgewählt sein.)

Sie können auf der Symbolleiste auf die Schaltfläche “Zuletzt verwendete Dialogfelder” klicken, um schnell zu den zuletzt verwendeten Prozeduren zurückzukehren.

Abbildung 5-3
Schaltfläche “Zuletzt verwendete Dialogfelder”

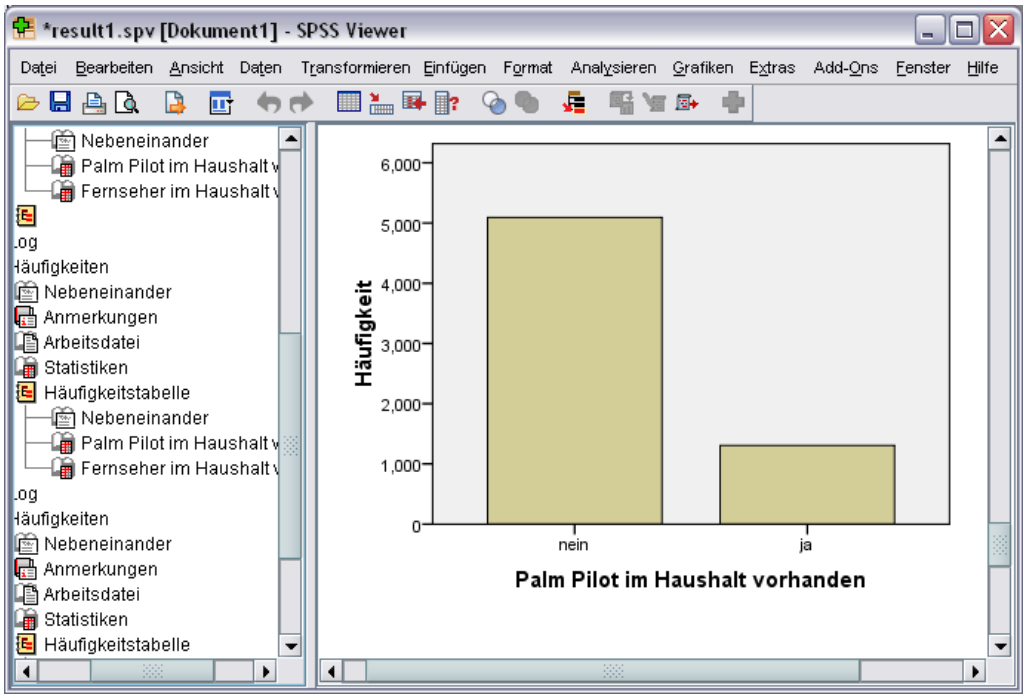
▶ Klicken Sie auf Diagramme.

▶ Wählen Sie die Option Balkendiagramme und klicken Sie anschließend auf Weiter.

Abbildung 5-4
Dialogfeld "Häufigkeiten: Diagramme"

▶ Klicken Sie im Hauptdialogfeld auf OK, um die Prozedur auszuführen.

Abbildung 5-5
Balkendiagramm



Zusätzlich zu den Häufigkeitstabellen werden dieselben Informationen jetzt in Form von Balkendiagrammen dargestellt, wodurch nun leicht zu erkennen ist, dass die meisten Personen keinen Palm Pilot, jedoch fast alle einen Fernseher besitzen.

Auswertungsmaße für metrische Variablen

Für metrische Variablen gibt es viele Auswertungsmaße, wie z. B:

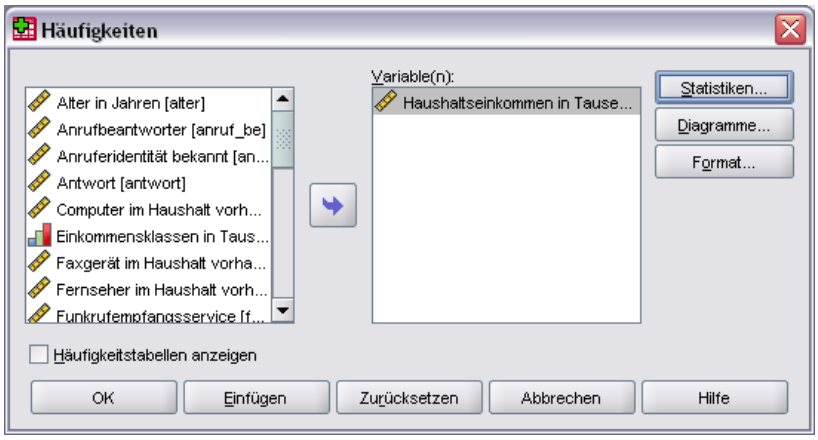
- **Maße der zentralen Tendenz.** Die am häufigsten verwendeten Maße der zentralen Tendenz sind der **Mittelwert** (arithmetisches Mittel) und der **Median** (Wert, ober- und unterhalb dessen die Hälfte aller Fälle angesiedelt sind).
- **Streuungsmaße.** Zu den Statistiken, welche die Streubreite oder die Variation in den Daten messen, gehören die Standardabweichung, das Minimum und das Maximum.

▶ Öffnen Sie das Dialogfeld “Häufigkeiten” erneut.

▶ Klicken Sie auf Zurücksetzen, um frühere Einstellungen zu löschen.

▶ Wählen Sie den Eintrag *Haushaltseinkommen in Tausend [einkomm]* aus und verschieben Sie ihn in die Liste “Variable(n)”.

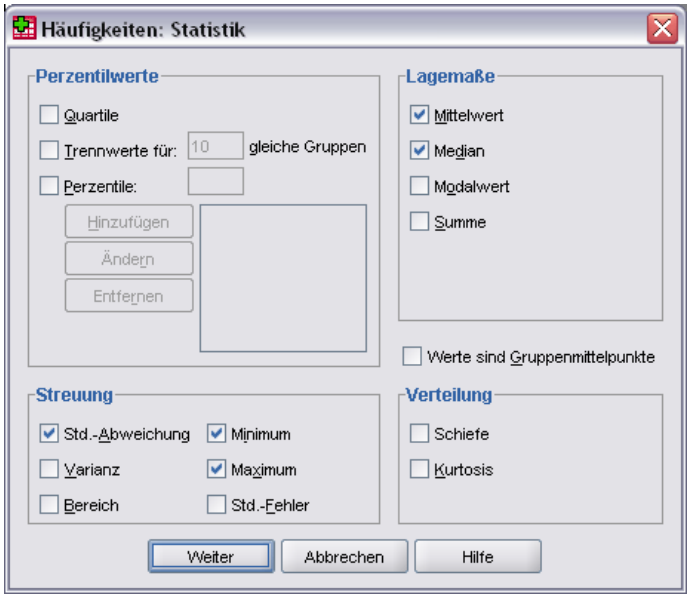
Abbildung 5-6
Für die Analyse ausgewählte metrische Variablen



▶ Klicken Sie auf Statistik.

▶ Wählen Sie die Optionen Mittelwert, Median, Std.-Abweichung, Minimum und Maximum.

Abbildung 5-7
Dialogfeld "Häufigkeiten: Statistik"



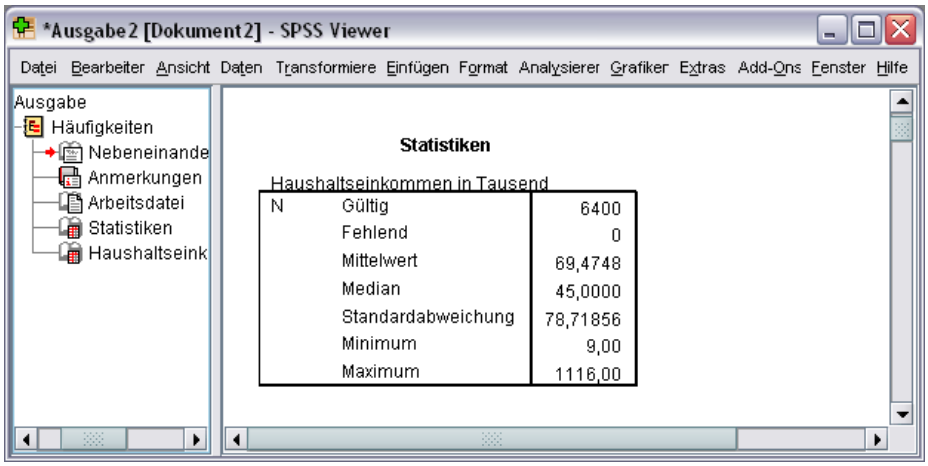
▶ Klicken Sie auf Weiter.

▶ Deaktivieren Sie (durch Entfernen der Auswahlmarkierung) Häufigkeitstabellen anzeigen im Hauptdialogfeld. (Häufigkeitstabellen sind bei metrischen Variablen in der Regel nicht besonders nützlich, da fast genauso viele unterschiedliche Werte wie Fälle in der Datendatei vorliegen können.)

▶ Klicken Sie auf OK, um die Prozedur auszuführen.

Die Tabelle Statistiken der Prozedur Häufigkeiten wird im Fenster "Viewer" angezeigt.

Abbildung 5-8
Tabelle Statistiken der Prozedur Häufigkeiten



In diesem Beispiel liegt eine große Differenz zwischen dem Mittelwert und dem Median vor. Der Mittelwert übersteigt den Mediam um fast 25.000. Dies bedeutet, dass die Werte nicht normalverteilt sind. Anhand eines Histogramms kann die Verteilung anschaulich überprüft werden.

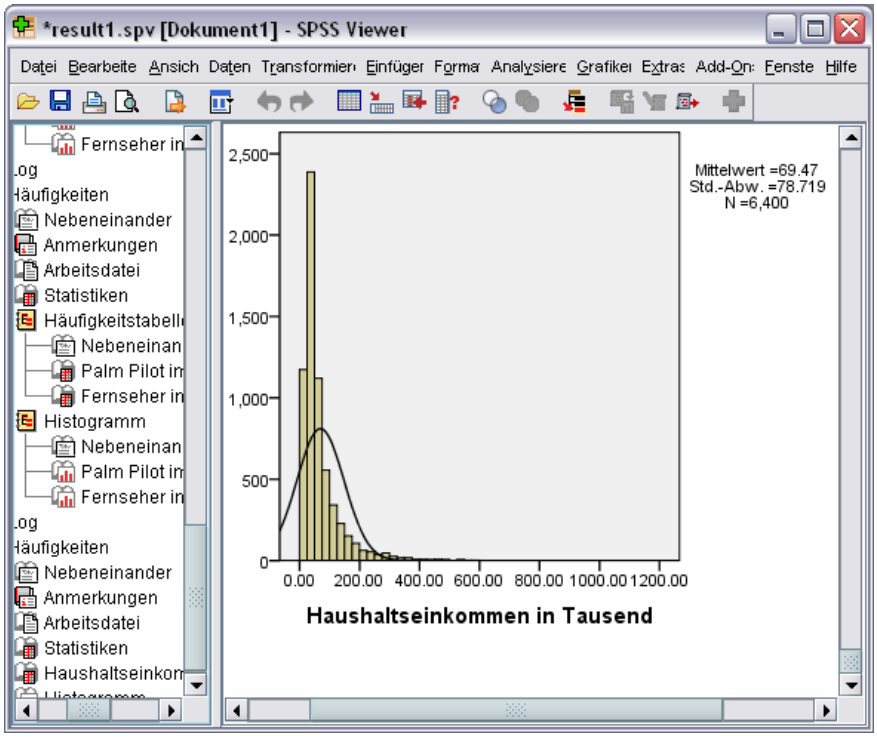
Histogramme für metrische Variablen

- ▶ Öffnen Sie das Dialogfeld “Häufigkeiten” erneut.
- ▶ Klicken Sie auf Diagramme.
- ▶ Wählen Sie die Option Histogramm und klicken Sie auf Mit Normalverteilungskurve.

Abbildung 5-9
Dialogfeld “Häufigkeiten: Diagramme”

- ▶ Klicken Sie auf Weiter und danach im Hauptdialogfeld auf OK, um die Prozedur auszuführen.

Abbildung 5-10
Histogramm



Die Mehrheit der Fälle ballt sich am unteren Ende der Skala, wobei die meisten unter 100.000 liegen. Ein paar Fälle liegen jedoch auch im Bereich um 500.000 und darüber (zu wenige, als dass sie ohne Änderung des Histogramms sichtbar wären). Die hohen Werte bei einigen wenigen Fällen haben große Auswirkungen auf den Mittelwert, aber nur eine geringe oder gar keine Auswirkung auf den Median. Daher ist der Median in diesem Beispiel ein besserer Indikator für die zentrale Tendenz.

Kapitel

6

Erstellen und Bearbeiten von Diagrammen

Sie können eine Vielzahl von Diagrammtypen erstellen und bearbeiten. In diesem Kapitel widmen wir uns der Erstellung und Bearbeitung von Balkendiagrammen. Die behandelten Prinzipien können auf alle Diagrammtypen übertragen werden.

Grundlagen der Diagrammerstellung

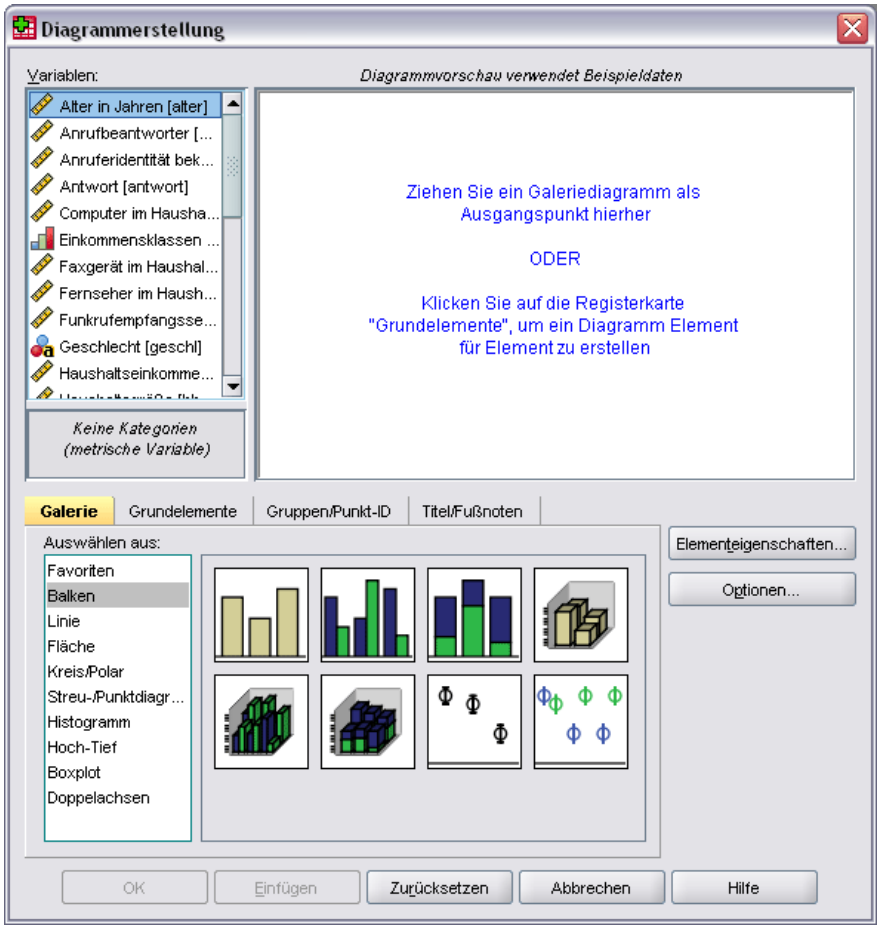
Um die Grundlagen der Diagrammerstellung zu demonstrieren, erstellen wir ein Balkendiagramm des Durchschnittseinkommens für die verschiedenen Stufen der Zufriedenheit mit der Arbeit. In diesem Beispiel wird die Datendatei *demo.sav* verwendet. Für weitere Informationen siehe Beispieldateien in Anhang A auf S. 163.

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Grafiken
  Diagrammerstellung...

Das Dialogfeld “Diagrammerstellung” ist ein interaktives Fenster, in dem Sie beim Erstellen in der Vorschau verfolgen können, wie das Diagramm aussieht.

Abbildung 6-1
Dialogfeld “Diagrammerstellung”



Verwenden der Galerie für die Diagrammerstellung

▶ Klicken Sie auf die Registerkarte Galerie, falls diese nicht bereits ausgewählt ist.

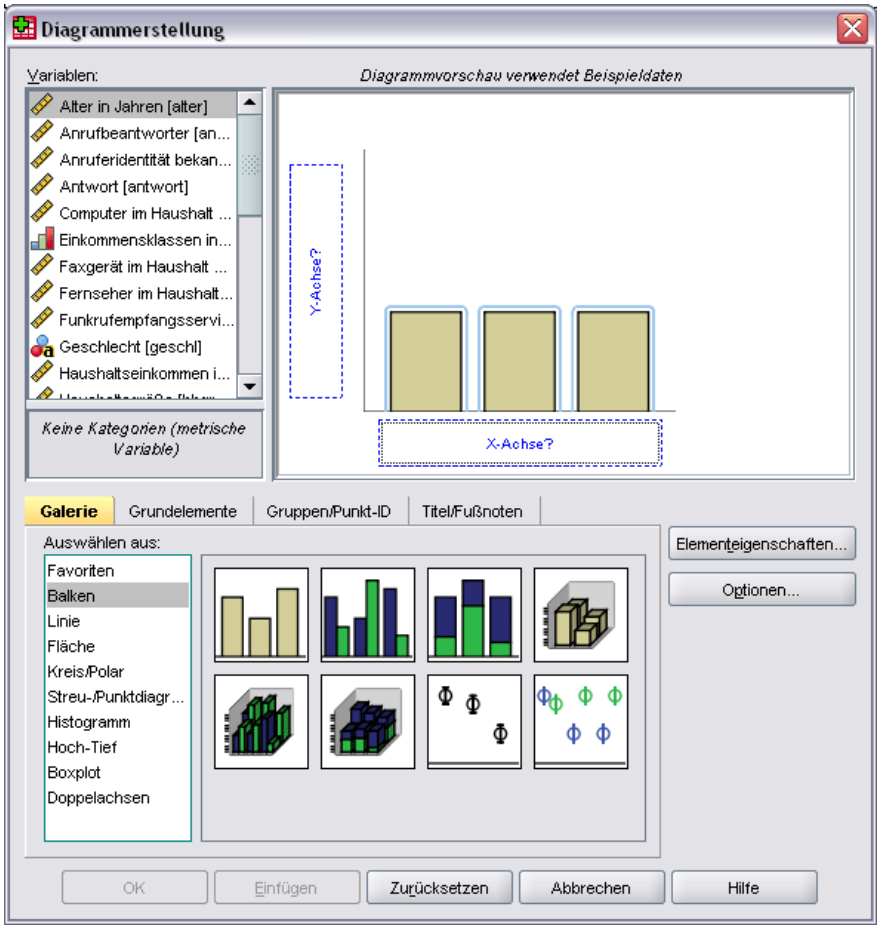
Die Galerie enthält zahlreiche verschiedene vordefinierte Diagramme, die nach Diagrammtypen angeordnet sind. Auf der Registerkarte “Grundelemente” stehen Ihnen auch Grundelemente wie Achsen und Grafikelemente zur Verfügung, mit denen Sie eigene Diagramme von Grund auf erstellen können. Die Galerie erleichtert Ihnen diese Arbeit jedoch.

▶ Klicken Sie auf Balken, falls dies nicht bereits ausgewählt ist.

Im Dialogfeld werden Symbole für die Balkendiagramme angezeigt, die in der Galerie verfügbar sind. Anhand der Bilder können Sie den jeweiligen Diagrammtyp erkennen. Wenn Sie weitere Informationen benötigen, können Sie mit dem Mauszeiger auf ein Symbol zeigen. Ihnen wird dann eine QuickInfo mit einer Beschreibung des Diagramms angezeigt.

▶ Ziehen Sie das Symbol für das einfache Balkendiagramm auf die Zeichenfläche. Dies ist der große Bereich oberhalb der Galerie. Im Dialogfeld “Diagrammerstellung” wird eine Vorschau des Diagramms auf der Zeichenfläche angezeigt. Beachten Sie, dass das Diagramm dabei nicht auf der Grundlage Ihrer Daten gezeichnet wird. Hierfür werden nur Beispieldaten verwendet.

Abbildung 6-2
Balkendiagramm auf der Zeichenfläche für die Diagrammerstellung



Definieren von Variablen und Statistiken

Das auf der Zeichenfläche angezeigte Diagramm ist noch nicht vollständig, da es weder Variablen noch Statistiken enthält, die die Größe der Balken steuern und angeben, welche Variablenkategorie den einzelnen Balken entspricht. Sie können kein Diagramm ohne Variablen und Statistiken anlegen. Sie fügen Variablen hinzu, indem Sie diese aus der Variablenliste ziehen, die sich links von der Zeichenfläche befindet.

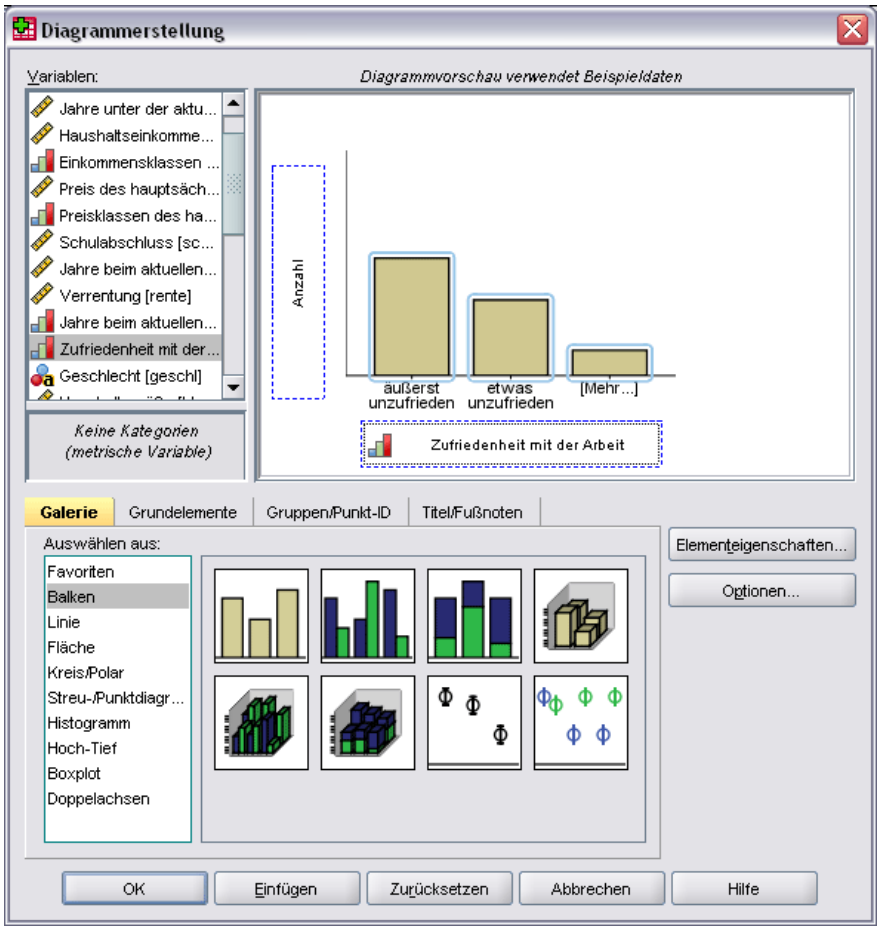
Für die Diagrammerstellung ist das Messniveau einer Variablen wichtig. Sie möchten auf der *x*-Achse die Variable *Zufriedenheit mit der Arbeit* verwenden. Das Symbol (es ähnelt einem Lineal) neben der Variablen gibt jedoch an, dass das Messniveau als metrisch definiert ist. Sie müssen für ein fehlerfreies Diagramm daher ein kategoriales Messniveau verwenden. Es ist nun

nicht nötig, wieder in die Variablenansicht zu wechseln und das Messniveau zu ändern. Sie können das Messniveau vorübergehen im Dialogfeld “Diagrammerstellung” ändern.

▶ Klicken Sie in der Liste “Variablen” mit der rechten Maustaste auf *Zufriedenheit mit der Arbeit* und wählen Sie Ordinal aus. “Ordinal” eignet sich als Messniveau, da die Kategorien in *Zufriedenheit mit der Arbeit* nach dem Zufriedenheitsgrad bewertet werden können. Beachten Sie, dass sich das Symbol nach der Änderung des Messniveaus ändert.

▶ Ziehen Sie *Zufriedenheit mit der Arbeit* nun aus der Liste “Variablen” auf die Ablagezone für die *x*-Achse.

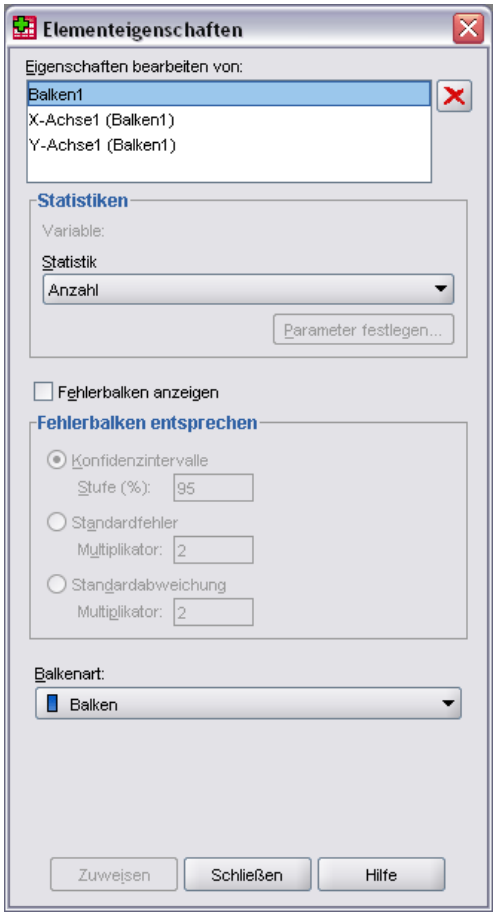
Abbildung 6-3
“Zufriedenheit mit der Arbeit” in der Ablagezone für die x-Achse



Die Ablagezone für die *y*-Achse ist standardmäßig auf die Statistik *Anzahl* festgelegt. Sie können jedoch problemlos eine andere Statistik wie Prozentsatz oder Mittelwert festlegen. Im vorliegenden Beispiel wird keine dieser Statistiken verwendet, dennoch wird das Verfahren für den Fall erläutert, dass Sie diese Statistik zu einem späteren Zeitpunkt ändern möchten.

▶ Klicken Sie auf Elementeigenschaften, um das Fenster “Elementeigenschaften” zu öffnen.

Abbildung 6-4
Fenster "Elementeigenschaften"



Im Fenster "Elementeigenschaften" können Sie die Eigenschaften der verschiedenen Diagrammelemente ändern Dabei handelt es sich um die Grafikelemente, z. B. die Balken im Balkendiagramm, und die Axen des Diagramms. Wählen Sie in der Liste "Eigenschaften bearbeiten von" ein Element aus, um die zugehörigen Eigenschaften zu ändern. Beachten Sie auch das rote *X* rechts neben der Liste. Mit dieser Schaltfläche kann ein Grafikelement aus der Zeichenfläche gelöscht werden. Hier ist Bar1 ausgewählt, daher beziehen sich die angezeigten Eigenschaften auf Grafikelemente, in diesem Fall das Grafikelement für den Balken.

In der Dropdown-Liste "Statistik" werden die verfügbaren spezifischen Statistiken angezeigt. Diese stehen in der Regel für alle Diagrammtypen zur Verfügung. Für einige Statistiken muss die Ablagezone der *y*-Achse allerdings eine Variable enthalten.

▶ Wechseln Sie in das Dialogfeld "Diagrammerstellung" zurück und ziehen Sie *Haushaltseinkommen in Tausend* aus der Liste "Variablen" auf die Ablagezone für die *y*-Achse. Da die *y*-Achse eine metrische und die *x*-Achse eine kategoriale Variable enthält (ordinal ist ein Typ des kategorialen Messniveaus), ist die Ablagezone der *y*-Achse in der Standardeinstellung auf die Statistik *Mittelwert* festgelegt. Dies sind die benötigten Variablen und Statistiken, Sie müssen die Elementeigenschaften daher nicht ändern.

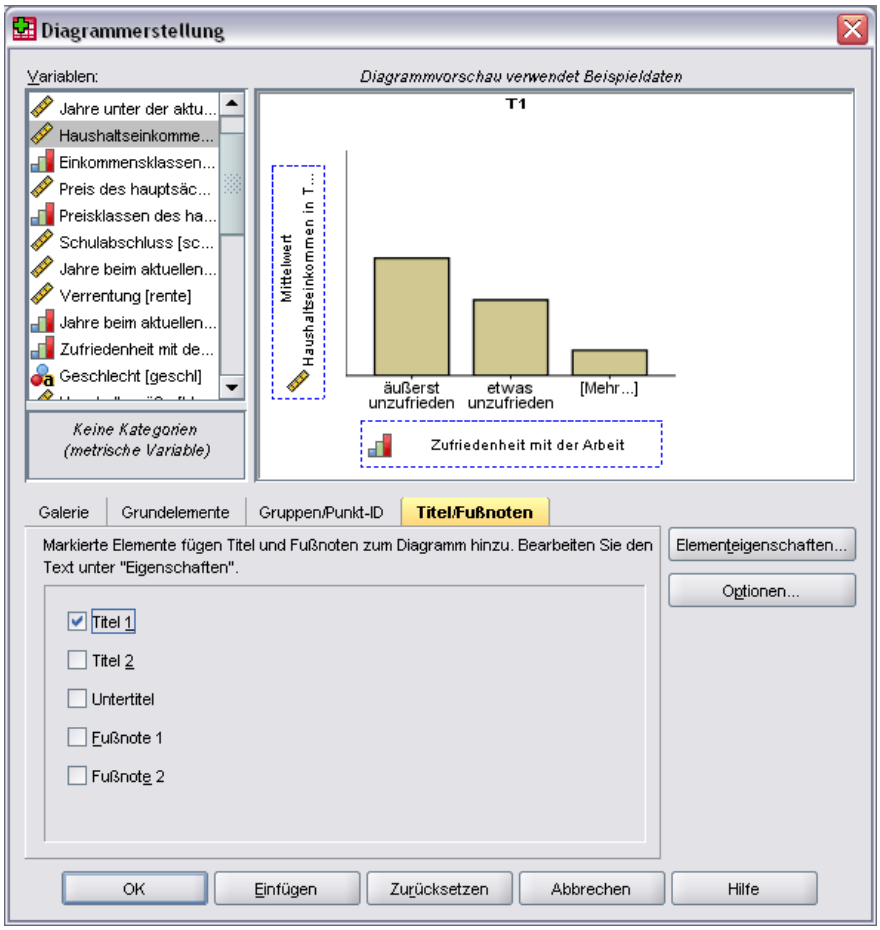
Hinzufügen von Text

Sie können Titel und Fußnoten zum Diagramm hinzufügen.

▶ Klicken Sie auf die Registerkarte Titel/Fußnoten.

▶ Wählen Sie Titel 1.

Abbildung 6-5
"Titel 1" auf der Zeichenfläche



Der Titel wird mit der Beschriftung T1 auf der Zeichenfläche angezeigt.

▶ Wählen Sie im Fenster "Elementeigenschaften" in der Liste "Eigenschaften bearbeiten von" den Eintrag Titel 1 aus.

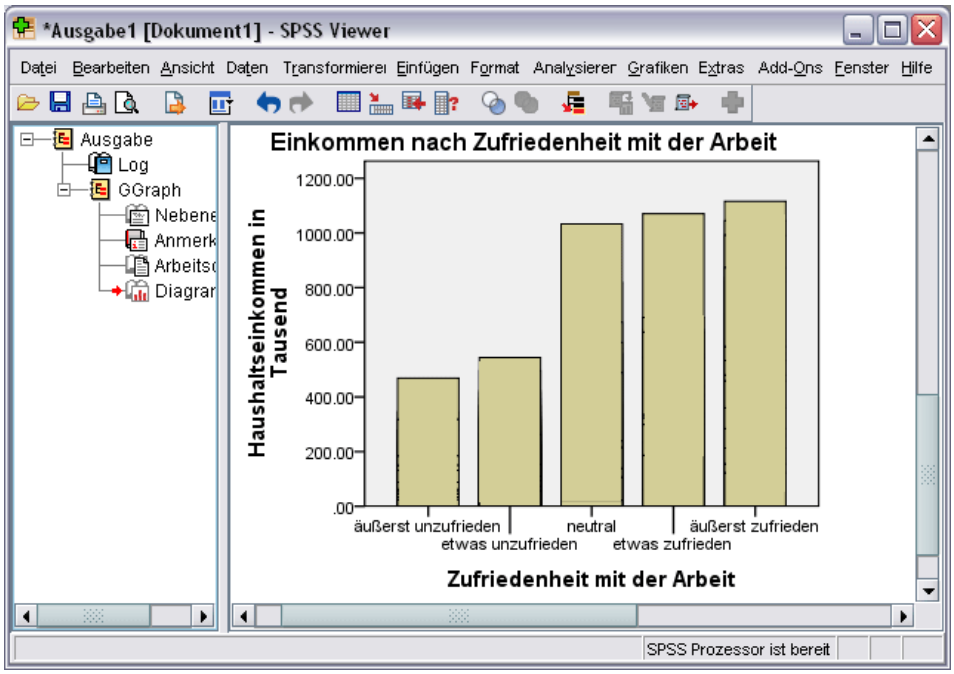
▶ Geben Sie im Textfeld "Inhalt" Einkommen nach Zufriedenheit mit der Arbeit ein. Dieser Text wird dann als Titel angezeigt.

▶ Klicken Sie auf Zuweisen, um den Text zu speichern. Der Text wird im Dialogfeld "Diagrammerstellung" zwar nicht angezeigt, ist jedoch später beim Generieren des Diagramms sichtbar.

Erstellen des Diagramms

▶ Klicken Sie auf OK, um das Balkendiagramm zu erstellen.

Abbildung 6-6
Balkendiagramm



Das Balkendiagramm zeigt, dass Befragte, die zufriedener mit ihrer Arbeit sind, in der Regel über eine höheres Haushaltseinkommen verfügen.

Grundlagen der Diagrammbearbeitung

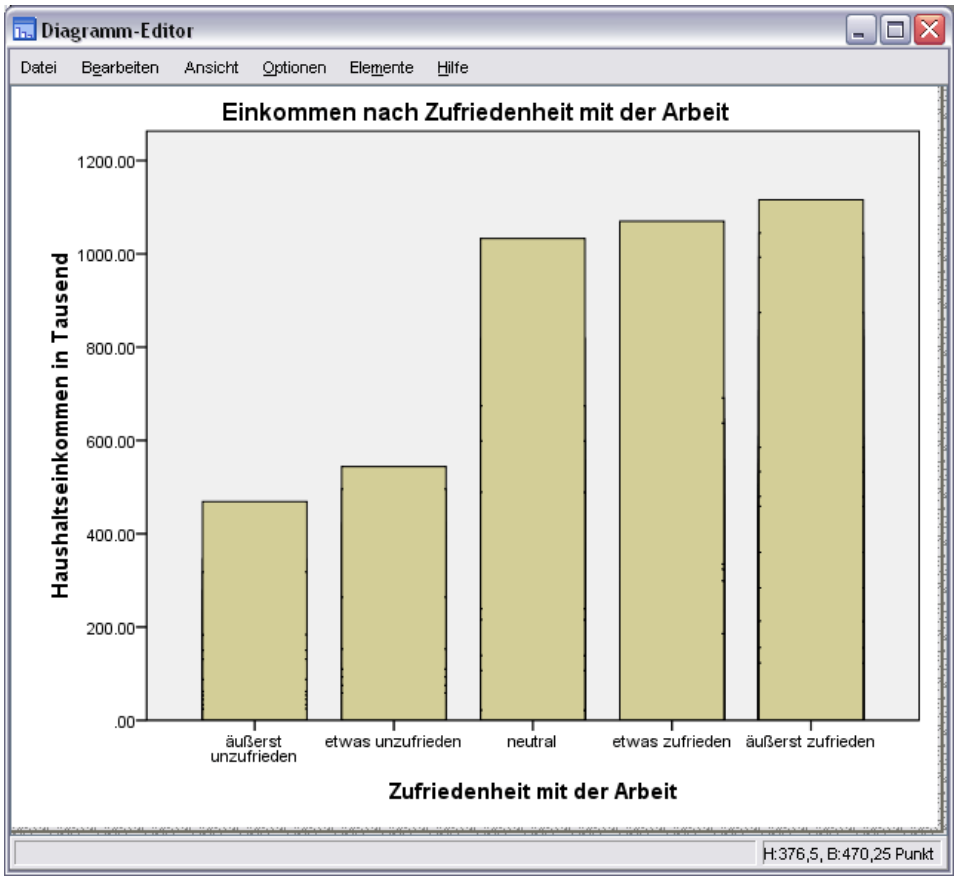
Diagramme können auf verschiedene Art bearbeitet werden. Für das soeben erstellte Beispiel-Balkendiagramm werden folgende Vorgänge durchgeführt:

- Ändern der Farben
- Formatieren der Zahlen in Teilstrichbeschriftungen
- Bearbeiten von Text
- Anzeigen von Datenwertelabels
- Verwenden von Diagrammvorlagen

Um das Diagramm zu bearbeiten, öffnen Sie es im Diagrammeditor.

▶ Doppelklicken Sie auf das Balkendiagramm, um es im Diagramm-Editor zu öffnen.

Abbildung 6-7
Balkendiagramm im Diagramm-Editor



Auswählen von Diagrammelementen

Um ein Diagrammelement zu bearbeiten, müssen Sie es zunächst auswählen.

▶ Klicken Sie auf einen der Balken. Die Rechtecke um die Balken zeigen an, dass sie ausgewählt sind.

Es gibt allgemeine Regeln für die Auswahl von Elementen in einfachen Diagrammen:

- Wenn keine Grafikelemente ausgewählt sind, können Sie durch Klicken auf ein beliebiges Grafikelement alle Grafikelemente auswählen.
- Wenn alle Grafikelemente ausgewählt sind, wird durch Klicken auf ein Grafikelement die Auswahl aller anderen Grafikelemente aufgehoben. Nur das Element, auf das geklickt wurde, bleibt ausgewählt. Sie können ein anderes Grafikelement auswählen, indem Sie darauf klicken. Um mehrere Grafikelemente auszuwählen, klicken Sie bei gedrückt gehaltener STRG-Taste auf die einzelnen Elemente.

▶ Drücken Sie ESC, um die Auswahl aller Elemente aufzuheben.

▶ Klicken Sie auf einen beliebigen Balken, um wieder alle Balken auszuwählen.

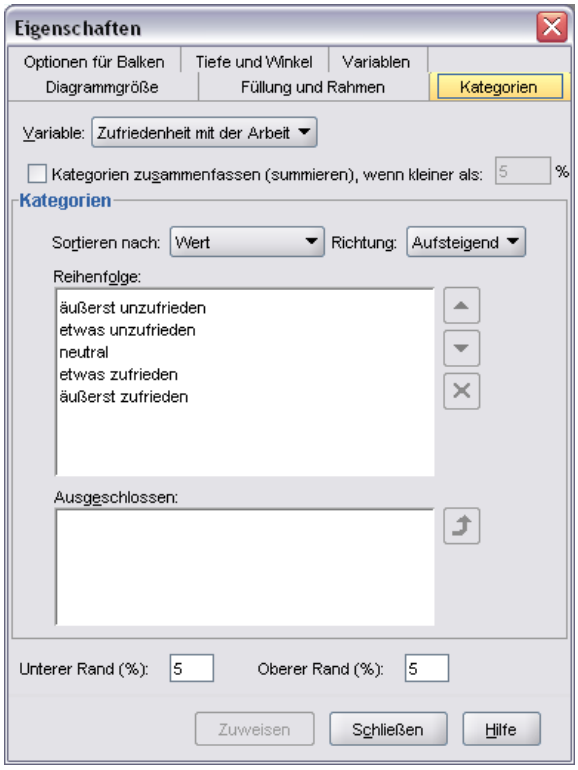
Verwendung des Fensters "Eigenschaften"

▶ Wählen Sie die folgenden Menübefehle des Diagramm-Editors aus:

Bearbeiten
  Eigenschaften

Hiermit wird das Fenster "Eigenschaften" geöffnet, das die Registerkarten für die ausgewählten Balken anzeigt. Diese Registerkarten ändern sich in Abhängigkeit von dem im Diagramm-Editor ausgewählten Diagrammelement. Wenn Sie beispielsweise anstelle der Balken einen Textrahmen ausgewählt hätten, würden im Fenster "Eigenschaften" andere Registerkarten angezeigt werden. Mithilfe dieser Registerkarten können Sie den größten Teil der Diagrammbearbeitung erledigen.

Abbildung 6-8
Fenster "Eigenschaften"

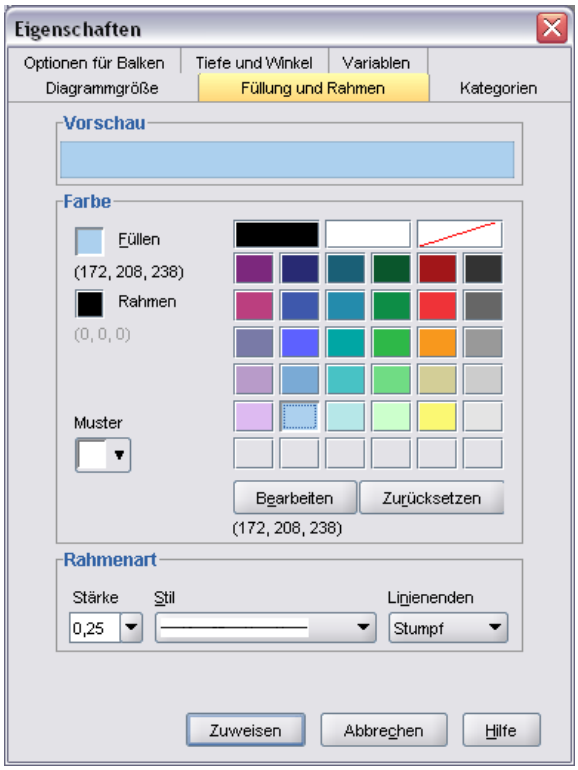


Ändern der Balkenfarben

Zunächst ändern Sie die Farbe der Balken. Auf der Registerkarte "Füllung und Rahmen" geben Sie die Farbattribute von Grafikelementen (außer Linien und Markierungen) an.

- Klicken Sie auf die Registerkarte Füllung und Rahmen.
- Klicken Sie auf das Muster neben “Füllung”, um anzuzeigen, dass Sie die Füllfarbe der Balken ändern möchten. Die Zahlen unter dem Muster geben die Rot-, Grün- und Blaueinstellungen für die aktuelle Farbe an.
- Klicken Sie auf die hellblaue Farbe, das zweite Feld von links in der zweiten Zeile von unten.

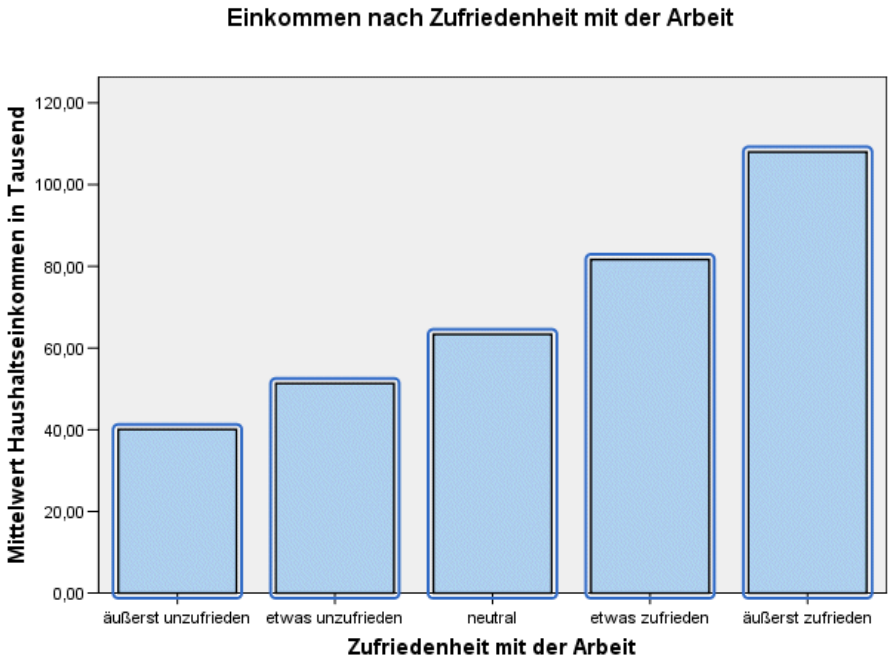
Abbildung 6-9
Registerkarte “Füllung und Rahmen”



- Klicken Sie auf Zuweisen.

Die Balken im Diagramm sind nun hellblau.

Abbildung 6-10
Bearbeitetes Balkendiagramm mit blauen Balken



Formatieren der Zahlen in Teilstrichbeschriftungen

Beachten Sie, dass die Zahlen auf der *y*-Achse auf Tausenderwerte skaliert sind. Um das Diagramm ansprechender und leichter interpretierbar zu gestalten, ändern wir das Zahlenformat in den Teilstrichbeschriftungen und bearbeiten dann den Achsentitel entsprechend.

- Wählen Sie die Teilstrichbeschriftungen für die *y*-Achse aus, indem Sie auf eine der Beschriftungen klicken.
- Um das Fenster “Eigenschaften” erneut zu öffnen (sofern es zuvor geschlossen wurde), wählen Sie in den Menüs folgende Optionsfolge:

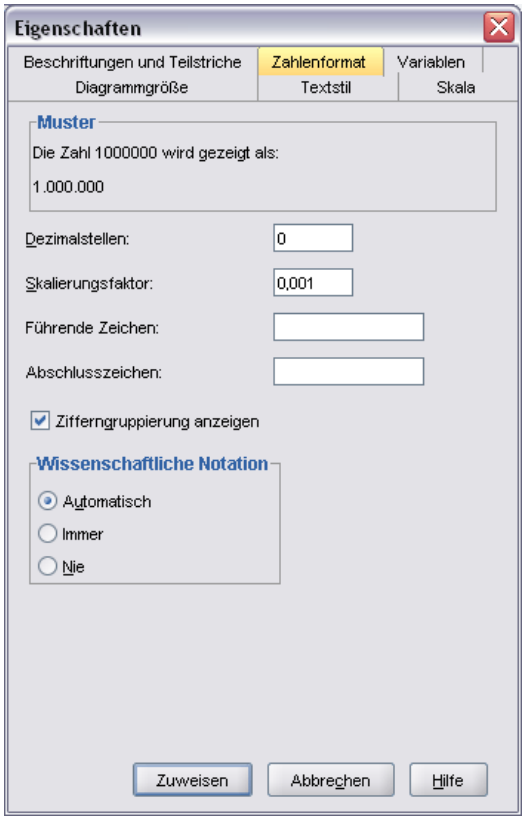
  Bearbeiten
  Eigenschaften

  Anmerkung: Von jetzt an gehen wir davon aus, dass das Fenster “Eigenschaften” geöffnet ist. Falls Sie das Fenster “Eigenschaften” geschlossen haben, können Sie es wie oben beschrieben wieder öffnen. Alternativ können Sie das Fenster mit der Tastenkombination STRG+T öffnen.
- Klicken Sie auf die Registerkarte Zahlenformat.
- Wenn in den Teilstrichbeschriftungen keine Dezimalstellen angezeigt werden sollen, geben Sie in das Textfeld “Dezimalstellen” den Wert 0 ein.

- Geben Sie im Textfeld “Skalierungsfaktor” den Wert 0,001 ein. Der Skalierungsfaktor ist die Zahl, durch die der Diagramm-Editor die anzuzeigende Zahl dividiert. Da 0,001 eine Bruchzahl ist, *erhöht* eine Division durch diesen Wert die Zahlen in den Teilstrichbeschriftungen um den Faktor 1.000. Die Zahlen werden daher nicht mehr in Tausenderwerten angeben, sondern sind nun unskaliert.

- Wählen Sie die Option Zifferngruppierung anzeigen. Bei der Zifferngruppierung wird ein Zeichen (durch die Ländereinstellung Ihres Computers bestimmt) verwendet, um jede Tausenderstelle in der Zahl zu markieren.

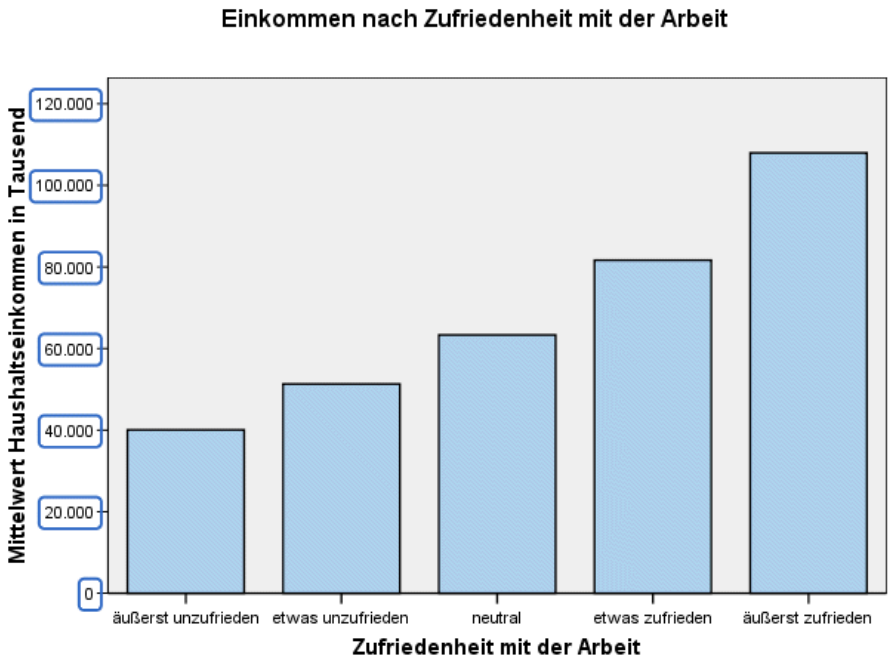
Abbildung 6-11
Registerkarte “Zahlenformat”



- Klicken Sie auf Zuweisen.

Die Teilstrichbeschriftungen zeigen nun das neue Zahlenformat an: Es gibt keine Dezimalstellen, die Zahlen sind nicht mehr skaliert, und die Tausenderstellen sind durch ein Zeichen gekennzeichnet.

Abbildung 6-12
Bearbeitetes Balkendiagramm mit neuem Zahlenformat



Bearbeiten von Text in einem Diagramm

Da Sie nun das Zahlenformat der Teilstrichbeschriftungen geändert haben, ist der Achsentitel nicht mehr korrekt. Daher werden Sie als nächstes den Achsentitel entsprechend dem neuen Zahlenformat ändern.

Anmerkung: Sie brauchen das Fenster “Eigenschaften” nicht zu öffnen, um Text zu bearbeiten. Sie können Texte direkt im Diagramm bearbeiten.

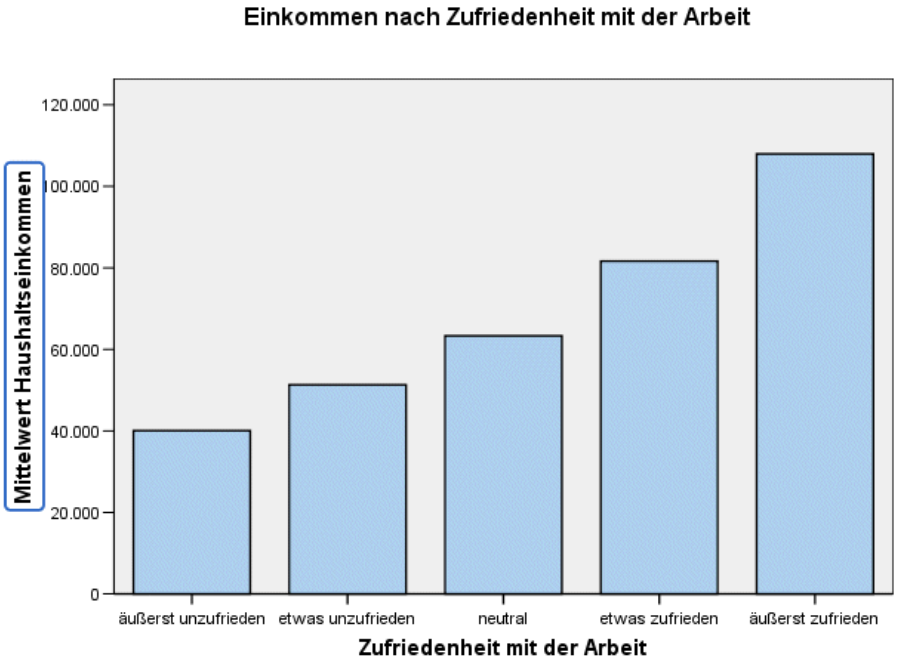
- Klicken Sie einmal auf den Titel der *y*-Achse, um ihn auszuwählen.
- Klicken Sie erneut auf den Achsentitel, um den Bearbeitungsmodus zu starten. Im Bearbeitungsmodus positioniert der Diagramm-Editor alle gedrehten Textelemente horizontal. Außerdem wird ein blinkender roter Balkencursor angezeigt (im Beispiel nicht zu sehen).

▶ Löschen Sie folgenden Text:

in Tausend

▶ Drücken Sie die Eingabetaste, um den Bearbeitungsmodus zu verlassen und den Achsentitel zu übernehmen. Der Achsentitel beschreibt nun den Inhalt der Teilstrichbeschriftungen korrekt.

Abbildung 6-13
Balkendiagramm mit bearbeitetem Titel der y-Achse



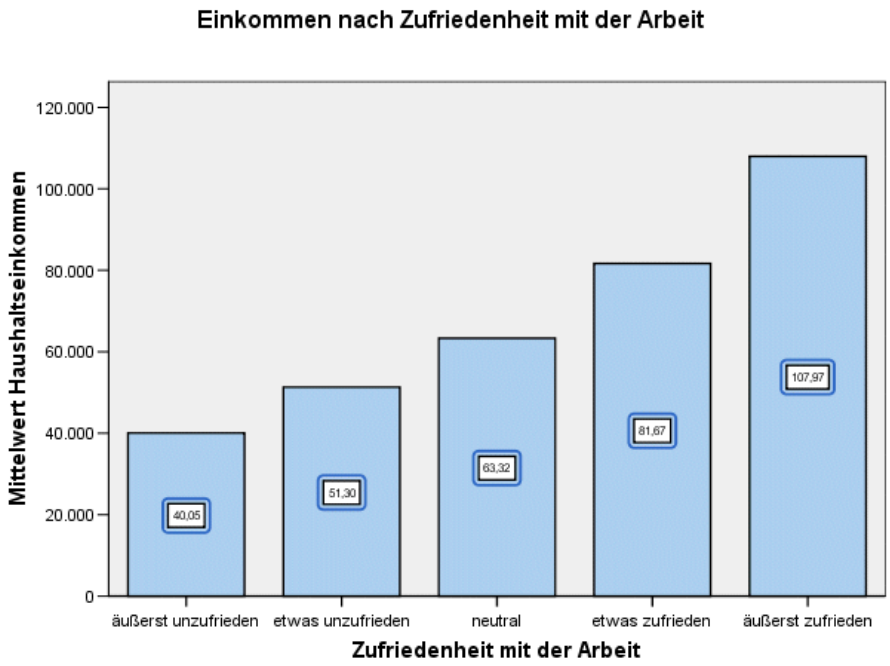
Anzeigen von Datenbeschriftungen

Eine weitere häufige Aufgabe besteht in der Anzeige der exakten Werte, die den Grafikelementen zugeordnet sind (in diesem Beispiel den Balken). Diese Werte werden in Datenbeschriftungen angezeigt.

▶ Wählen Sie die folgenden Menübefehle des Diagramm-Editors aus:

Elemente
  Datenbeschriftungen einblenden

Abbildung 6-14
Balkendiagramm mit Datenbeschriftungen



Jeder Balken im Diagramm zeigt nun das genaue mittlere Haushaltseinkommen an. Beachten Sie, dass alle Einheiten in Tausendern angegeben werden. Sie könnten also die Registerkarte "Zahlenformat" erneut verwenden, um den Skalierungsfaktor zu ändern.

Verwendung von Vorlagen

Wenn Sie eine Reihe von routinemäßigen Änderungen an Ihren Diagrammen vornehmen, können Sie den für das Erstellen und Bearbeiten der Diagramme erforderlichen Zeitaufwand durch eine Diagrammvorlage reduzieren. In einer Diagrammvorlage werden die Attribute eines bestimmten Diagramms gespeichert. Sie können die Vorlage dann beim Erstellen bzw. Bearbeiten eines Diagramms anwenden.

Wir speichern das aktuelle Diagramm als Vorlage und weisen diese dann bei der Erstellung eines neuen Diagramms zu.

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Datei
  Diagrammvorlage speichern...

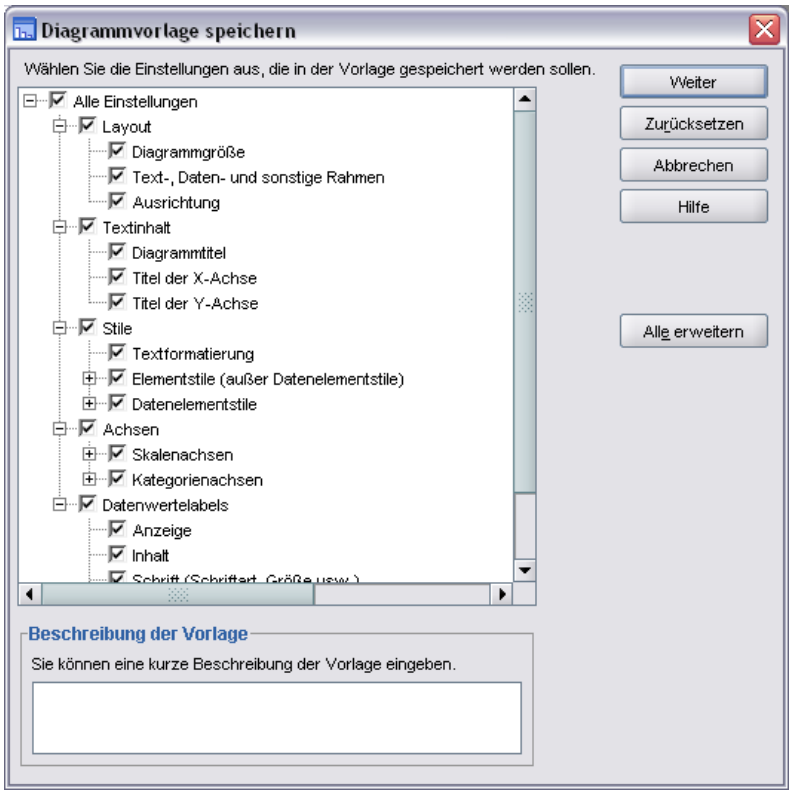
Im Dialogfeld "Diagrammvorlage speichern" können Sie angeben, welche Diagrammattribute in die Vorlage aufgenommen werden sollen.

Wenn Sie eines der Elemente in der Baumansicht erweitern, sehen Sie, welche speziellen Attribute mit dem Diagramm gespeichert werden können. Wenn Sie beispielsweise den Bereich Skalenachsen des Baums erweitern, sehen Sie alle Attribute der Datenwertelabel, die die Vorlage beinhaltet. Sie können ein beliebiges Attribut für die Aufnahme in die Vorlage auswählen.

▶ Wählen Sie Alle Einstellungen aus, um alle verfügbaren Diagrammattribute in die Vorlage aufzunehmen.

Sie können außerdem eine kurze Beschreibung der Vorlage eingeben. Diese Beschreibung wird sichtbar, wenn die Vorlage zugewiesen wird.

Abbildung 6-15
Dialogfeld "Diagrammvorlage speichern"



▶ Klicken Sie auf Weiter.

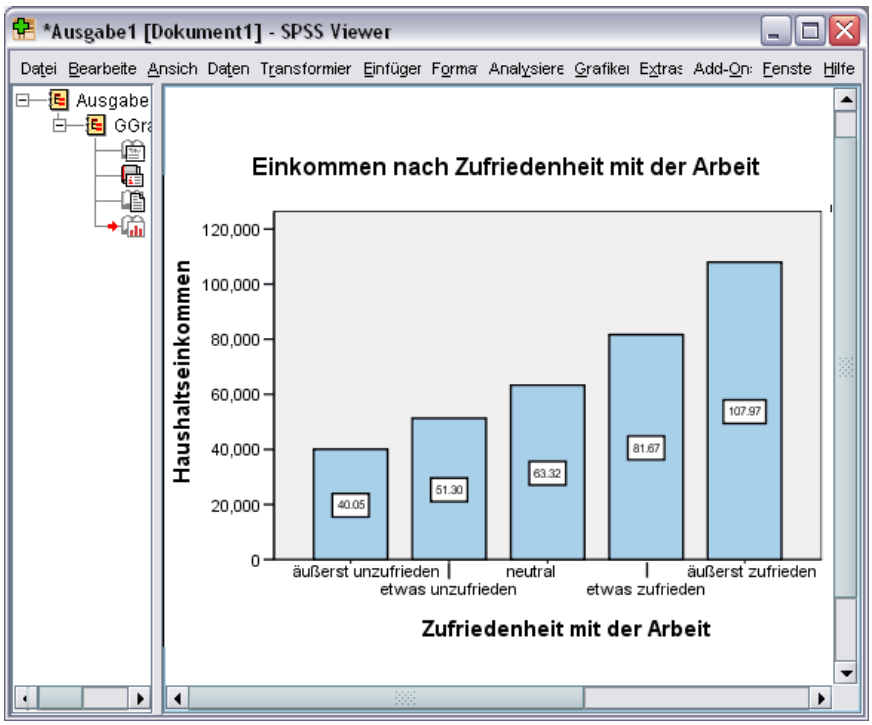
▶ Geben Sie im Dialogfeld "Vorlage speichern" einen Ort und einen Dateinamen für die Vorlage ein.

▶ Klicken Sie abschließend auf Speichern.

Sie können die Vorlage beim Erstellen eines Diagramms oder im Diagramm-Editor zuweisen. Im folgenden Beispiel weisen wir sie beim Erstellen eines Diagramms zu.

▶ Schließen Sie den Diagramm-Editor. Das aktualisierte Diagramm wird im Viewer angezeigt.

Abbildung 6-16
Aktualisiertes Balkendiagramm im Viewer.



▶ Wählen Sie die folgenden Befehle aus den Menüs des Viewers aus:

Grafiken
  Diagrammerstellung...

Das Dialogfeld “Diagrammerstellung” kann sich die Variablen “merken”, die Sie beim Erstellen des ursprünglichen Diagramms eingegeben haben. Hier soll jedoch ein etwas anderes Diagramm erstellt werden, damit Sie sehen können, wie Diagramme durch das Zuweisen einer Vorlage formatiert werden.

▶ Entfernen Sie *Zufriedenheit mit der Arbeit* von der *x*-Achse, indem Sie die Variable aus der Ablagezone zurück in die Liste “Variablen” ziehen. Sie können auch auf die Ablagezone klicken und anschließend die Entf-Taste drücken.

▶ Klicken Sie in der Liste “Variablen” mit der rechten Maustaste auf *Schulabschluss* und wählen Sie Ordinal aus.

▶ Ziehen Sie *Schulabschluss* nun aus der Liste “Variablen” auf die Ablagezone für die *x*-Achse.

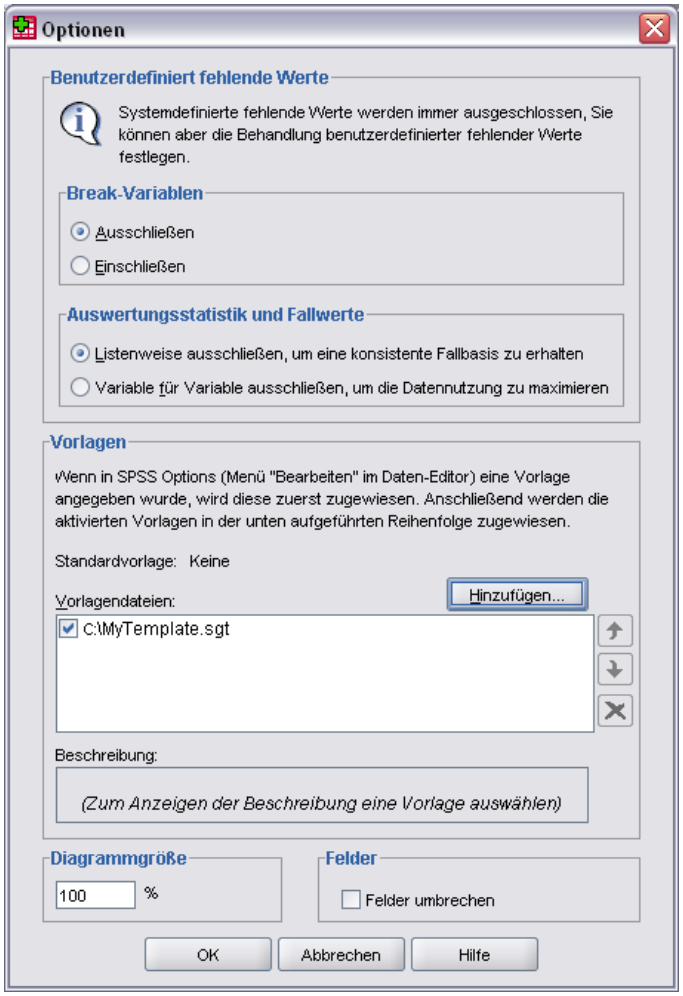
Der Titel stimmt nicht mehr, daher wird er nun gelöscht.

▶ Heben Sie auf der Registerkarte “Titel/Fußnoten” die Auswahl von Titel 1 auf.

Als nächstes wird die Vorlage angegeben, die auf das neue Diagramm angewendet wird.

- Klicken Sie auf Optionen.
- Klicken Sie im Dialogfeld “Optionen” im Gruppenfeld “Vorlagen” auf Hinzufügen.
- Suchen Sie im Dialogfeld “Vorlagendateien suchen” die Vorlagendatei, die Sie zuvor im Dialogfeld “Diagrammvorlage speichern” gespeichert haben.
- Wählen sie diese Datei aus und klicken Sie auf Öffnen.

Abbildung 6-17
Dialogfeld “Optionen” mit Vorlage

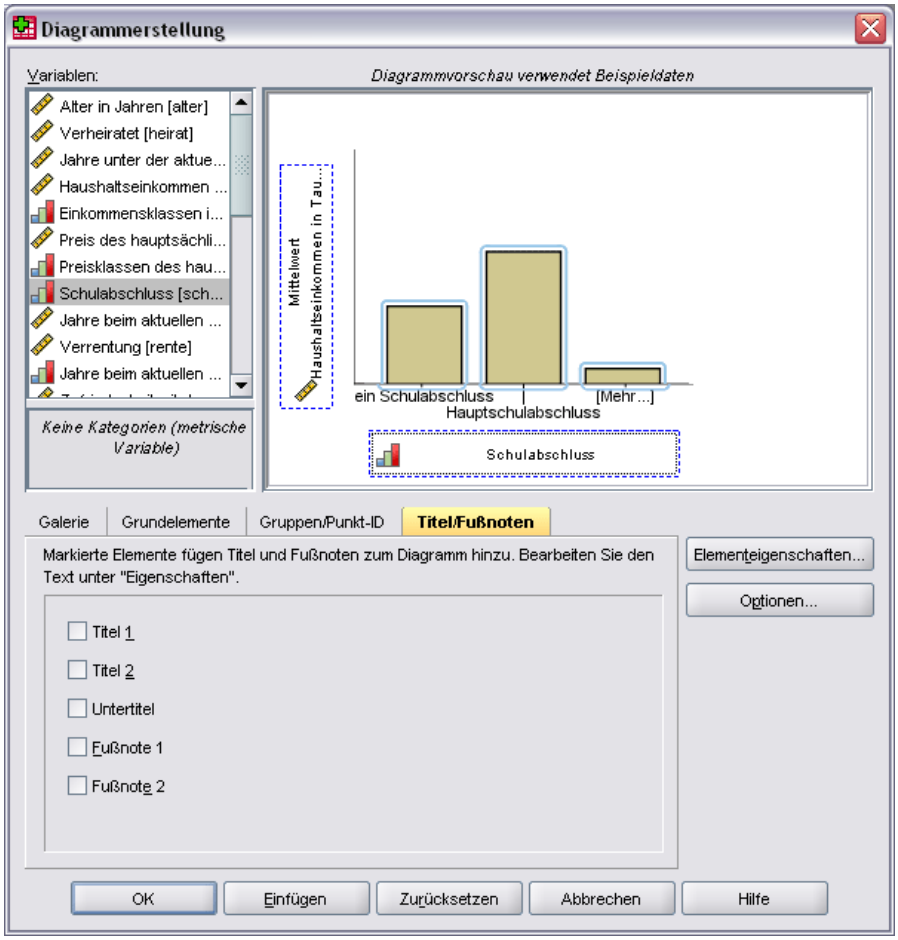


Im Dialogfeld “Optionen” wird der Dateipfad der ausgewählten Vorlage angezeigt.

(In unserem Beispiel lautet dieser Pfad: *C:\MyTemplate.sgt*.)

- Klicken Sie auf OK, um das Dialogfeld “Optionen” zu schließen.

Abbildung 6-18
Dialogfeld "Diagrammerstellung" mit belegten Ablagezonen

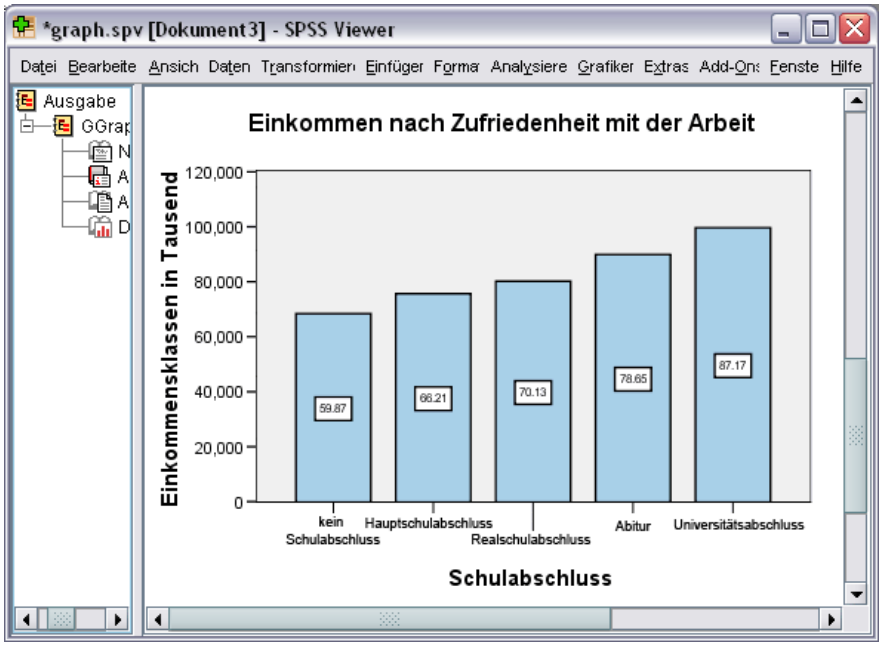


▶ Klicken Sie auf im Dialogfeld "Diagrammerstellung" auf OK, um das Diagramm zu erstellen und die Vorlage zuzuweisen.

Die Formatierung im neuen Diagramm stimmt mit der Formatierung in dem zuvor erstellten und bearbeiteten Diagramm überein. Die Variablen auf der *x*-Achse unterscheiden sich zwar, ansonsten ähneln sich die Diagramme jedoch. Beachten Sie, dass der Titel des vorausgegangenen Diagramms sich immer noch in der Vorlage befindet, obwohl Sie ihn in der "Diagrammerstellung" gelöscht hatten.

Wenn Sie nach dem Erstellen eines Diagramms Vorlagen zuweisen möchten, können Sie dies im Diagramm-Editor tun. (Wählen Sie hierfür im Menü “Datei” die Option Diagrammvorlage zuweisen.)

Abbildung 6-19
Aktualisiertes Balkendiagramm im Viewer.



Optionen zum Definieren von Diagrammen

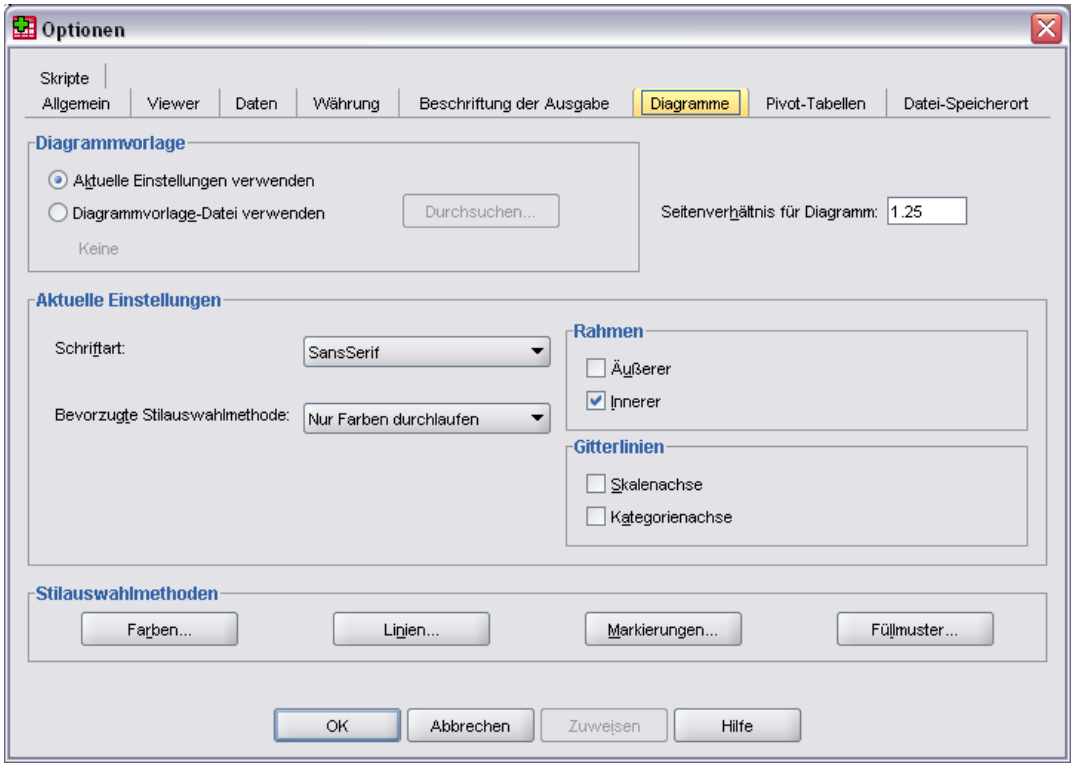
Sie können Diagramme nicht nur mithilfe von Vorlagen formatieren, sondern auch mit den Optionen verschiedene Aspekte der Diagrammerstellung steuern.

▶ Wählen Sie die folgenden Befehle im Menü des Daten-Editors oder des Viewers aus:

Bearbeiten
  Optionen...

Das Dialogfeld “Optionen” enthält zahlreiche Konfigurationseinstellungen. Klicken Sie auf die Registerkarte Diagramme, um die verfügbaren Optionen anzuzeigen.

Abbildung 6-20
Registerkarte “Diagramme” im Dialogfeld “Optionen”



Die Optionen legen fest, wie ein Diagramm erstellt wird. Für jedes neue Diagramm können Sie folgende Aspekte angeben:

- Ob die aktuellen Einstellungen oder eine Vorlage verwendet werden soll.
- Das Verhältnis von Breite zu Höhe (Seitenverhältnis).
- Wenn Sie keine Vorlage verwenden: die für die Formatierung zu verwendenden Einstellungen.
- Die Stilauswahlmethoden für Grafikelemente.

Stilauswahlmethoden ermöglichen die Angabe des Stils von Grafikelementen in neuen Diagrammen. In diesem Beispiel betrachten wir die Details für die Stilauswahlmethode für die Farbe.

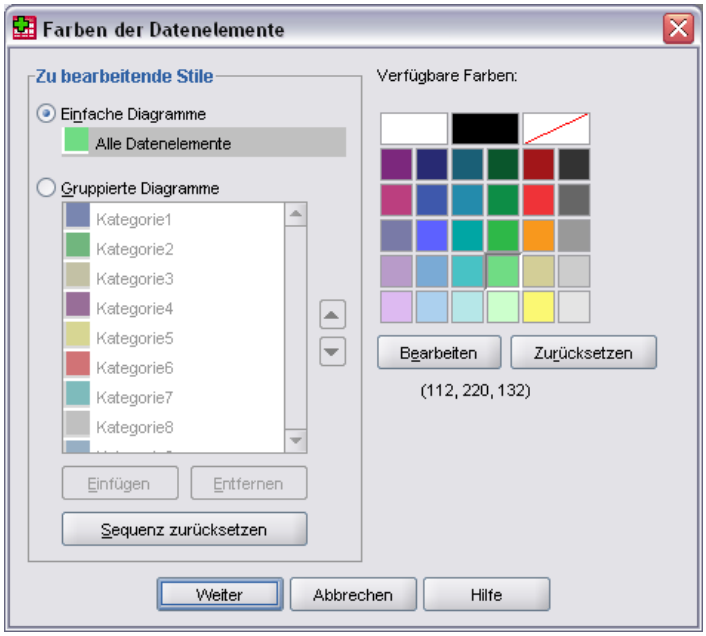
▶ Klicken Sie auf Farben, um das Dialogfeld “Farben der Datenelemente” zu öffnen.

Bei einfachen Diagrammen verwendet der Diagramm-Editor einen einzigen, von Ihnen angegebenen Stil. Bei gruppierten Diagrammen verwendet der Diagramm-Editor eine Menge von Stilen, die er für jede Gruppe (Kategorie) im Diagramm durchgeht.

▶ Wählen Sie die Option Einfache Diagramme.

▶ Wählen Sie die hellgrüne Farbe aus, das dritte Feld von rechts in der zweiten Zeile von unten.

Abbildung 6-21
Dialogfeld "Farben der Datenelemente"



- Klicken Sie auf Weiter.
- Klicken Sie im Dialogfeld "Optionen" auf OK, um die Änderungen an der Stilauswahlmethode für die Farbe zu speichern.

  Die Grafikelemente in allen neuen einfachen Diagrammen sind nun hellgrün.

- Wählen Sie die folgenden Befehle im Menü des Daten-Editors oder des Viewers aus:

  Grafiken
  Diagrammerstellung...

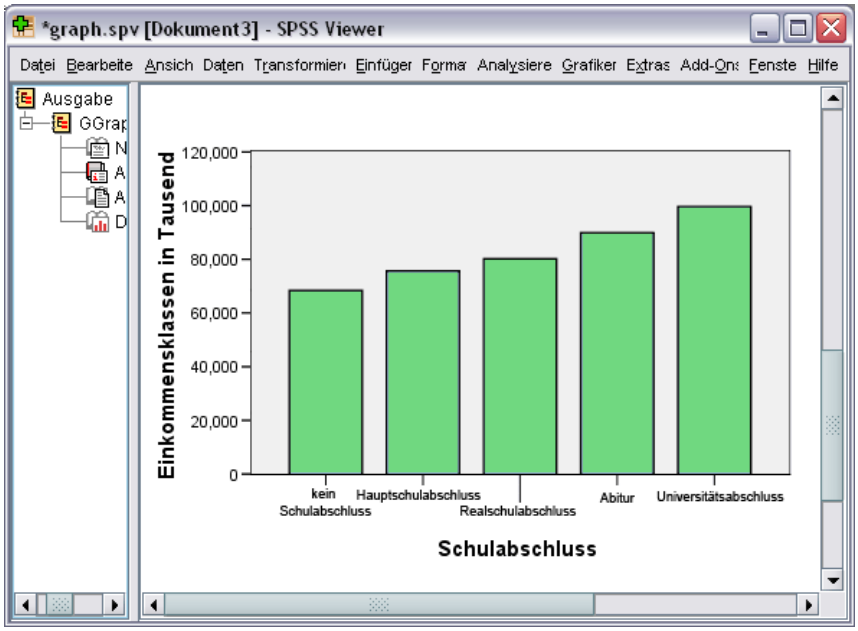
  In der "Diagrammerstellung" wird das zuletzt erstellte Diagramm angezeigt. Wie Sie wissen, wurde diesem Diagramm eine Vorlage zugewiesen. Diese Vorlage soll nicht mehr verwendet werden.

- Klicken Sie auf Optionen.
- Deaktivieren Sie die zuvor hinzugefügte Vorlage. Sie könnten die Vorlage auch löschen, indem Sie auf das rote *X* klicken. Wenn Sie die Vorlage jedoch nur deaktivieren, statt sie zu löschen, steht sie Ihnen später noch zur Verfügung.
- Klicken Sie auf OK, um das Diagramm zu erstellen.

  Die Balken im neuen Diagramm sind hellgrün. Dieses Diagramm unterscheidet sich vom vorausgegangenen Diagramm auch in anderen Aspekten. Es gibt keinen Titel, die Achsenbeschriftungen weisen Tausender-Einheiten auf, und es sind keine Datenbeschriftungen

vorhanden. Die Ursache für diese Unterschiede ist, dass die Vorlage nicht auf das Diagramm angewendet wurde.

Abbildung 6-22
Aktualisiertes Balkendiagramm im Viewer.



Kapitel

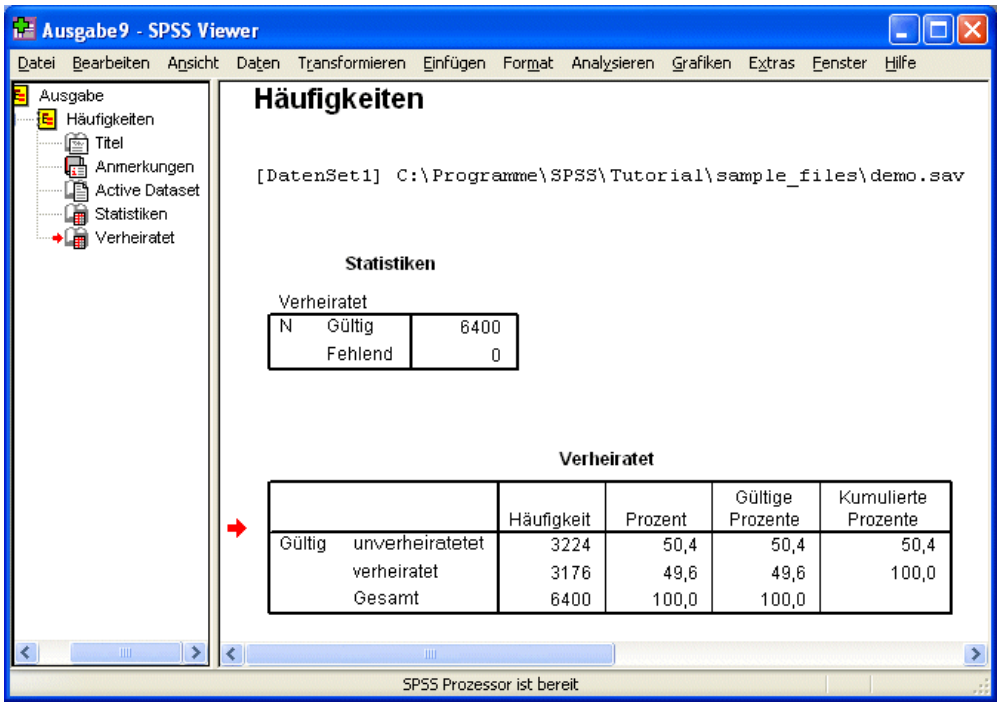
7

Arbeiten mit Ausgaben

Die Ergebnisse aus der Ausführung statistischer Prozeduren werden im Viewer angezeigt. Bei der Ausgabe kann es sich um Statistiktabellen, Diagramme, Grafiken oder Text handeln, je nachdem, welche Optionen beim Ausführen der Prozedur ausgewählt wurden. In diesem Abschnitt werden die Dateien *viewertut.spv* und *demo.sav* verwendet. Für weitere Informationen siehe Beispieldateien in Anhang A auf S. 163.

Arbeiten mit dem Viewer

Abbildung 7-1
Viewer



Das Viewer-Fenster ist in zwei Fensterbereiche unterteilt. Das **Gliederungsfenster** enthält eine Gliederung aller im Viewer enthaltenen Daten. Das **Inhaltsfenster** enthält Statistiktabellen, Diagramme und Textausgaben.

Mithilfe der Bildlaufleisten können Sie vertikal und horizontal durch den Inhalt des Fensters navigieren. Wenn Sie im Gliederungsfenster auf ein Element klicken, wird es im Inhaltsfenster anzeigt.

Wenn die Tabelle im Viewer nicht vollständig angezeigt werden kann oder die Gliederungsansicht zu schmal ist, können Sie das Fenster einfach vergrößern.

▶ Klicken Sie auf den rechten Rahmen des Gliederungsfensters und ziehen Sie diesen auf die gewünschte Breite.

Ein Symbol in Form eines offenen Buches im Gliederungsfenster weist darauf hin, dass das betreffende Element derzeit im Viewer angezeigt wird, obwohl es sich möglicherweise im gegenwärtig nicht sichtbaren Bereich des Inhaltsfensters befindet.

▶ Wenn Sie eine Tabelle oder ein Diagramm ausblenden möchten, können Sie im Gliederungsfenster auf das entsprechende Buchsymbol doppelklicken.

Das Symbol mit dem offenen Buch ändert sich zu einem geschlossenen Buch, woran Sie erkennen, dass die zugeordneten Daten ausgeblendet sind.

▶ Wenn die ausgeblendete Ausgabe wieder angezeigt werden soll, doppelklicken Sie auf das Symbol mit dem geschlossenen Buch.

Außerdem können Sie die gesamte Ausgabe aus einer bestimmten statistischen Prozedur oder die gesamte Ausgabe im Viewer ausblenden.

▶ Klicken Sie links neben der Prozedur, deren Ergebnisse ausgeblendet werden sollen, auf die Schaltfläche mit dem Minuszeichen (−) oder klicken Sie auf die Schaltfläche neben dem obersten Element der Gliederung, um die gesamte Ausgabe auszublenden.

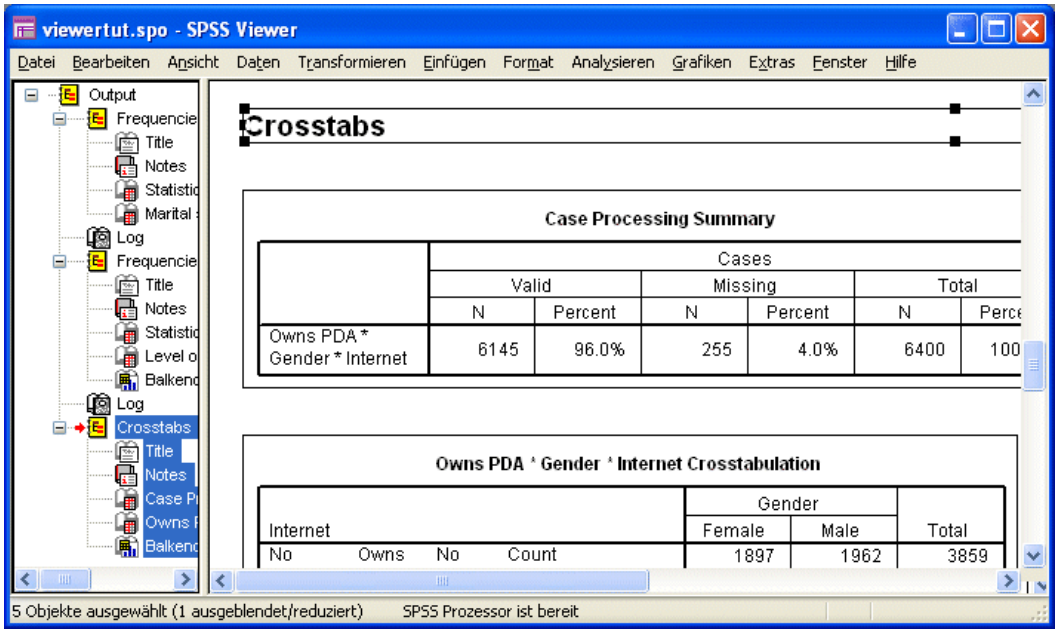
Die Gliederung wird reduziert, um optisch darauf hinzuweisen, dass die Ergebnisse nicht mehr angezeigt werden.

Sie können die Anzeigereihenfolge für die Ausgabe auch ändern.

▶ Klicken Sie im Gliederungsbereich auf die Elemente, die Sie verschieben möchten.

▶ Ziehen Sie die ausgewählten Elemente an die neue Position in der Gliederung und lassen Sie die Maustaste los.

Abbildung 7-2
Neu geordnete Ausgaben im Viewer



Sie können Ausgabeelemente auch verschieben, indem Sie auf diese klicken und sie in das Inhaltsfenster ziehen.

Verwenden des Pivot-Tabellen-Editors

Die Ergebnisse der meisten statistischen Prozeduren werden in **Pivot-Tabellen** angezeigt.

Aufrufen von Definitionen für statistische Begriffe in der Ausgabe

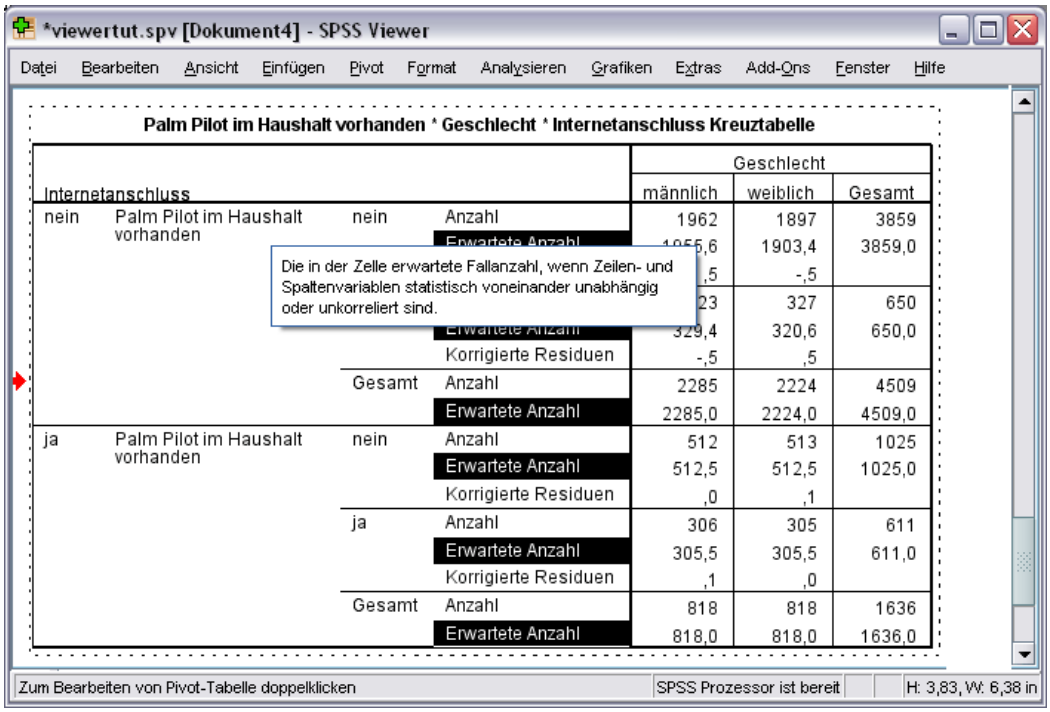
In der Ausgabe werden zahlreiche statistische Begriffe angezeigt. Auf die Definitionen dieser Begriffe können Sie direkt im Viewer zugreifen.

▶ Doppelklicken Sie auf die Tabelle *Palm Pilot im Haushalt vorhanden * Geschlecht * Internetanschluss Kreuztabelle*.

▶ Klicken Sie mit der rechten Maustaste auf *Erwartete Anzahl* und wählen Sie im Kontextmenü Direkthilfe aus.

Die Definition wird in einem Popup-Fenster angezeigt.

Abbildung 7-3
Popup-Definition



Internetanschluss				Geschlecht männlich	weiblich	Gesamt
nein	Palm Pilot im Haushalt vorhanden	nein	Anzahl	1962	1897	3859
			Erwartete Anzahl	1955,6	1903,4	3859,0
			Korrigierte Residuen	,5	-,5	
			Anzahl	323	327	650
			Erwartete Anzahl	329,4	320,6	650,0
			Korrigierte Residuen	-,5	,5	
		Gesamt	Anzahl	2285	2224	4509
			Erwartete Anzahl	2285,0	2224,0	4509,0
ja	Palm Pilot im Haushalt vorhanden	nein	Anzahl	512	513	1025
			Erwartete Anzahl	512,5	512,5	1025,0
			Korrigierte Residuen	,0	,1	
		ja	Anzahl	306	305	611
			Erwartete Anzahl	305,5	305,5	611,0
			Korrigierte Residuen	,1	,0	
		Gesamt	Anzahl	818	818	1636
			Erwartete Anzahl	818,0	818,0	1636,0

Pivot-Tabellen

Möglicherweise werden die Daten in den erstellten Standardtabellen unübersichtlich angezeigt. In Pivot-Tabellen können Sie Zeilen und Spalten transponieren (vertauschen), die Reihenfolge der Daten in der Tabelle ändern und die Tabelle auf vielfältige Art und Weise bearbeiten. Durch Transponieren von Zeilen und Spalten wird aus einer kurzen, breiten Tabelle eine lange, schmale Tabelle. Das Ändern des Tabellen-Layouts wirkt sich nicht auf die Ergebnisse aus. Es bietet nur eine Möglichkeit, Informationen auf andere bzw. angemessenere Weise darzustellen.

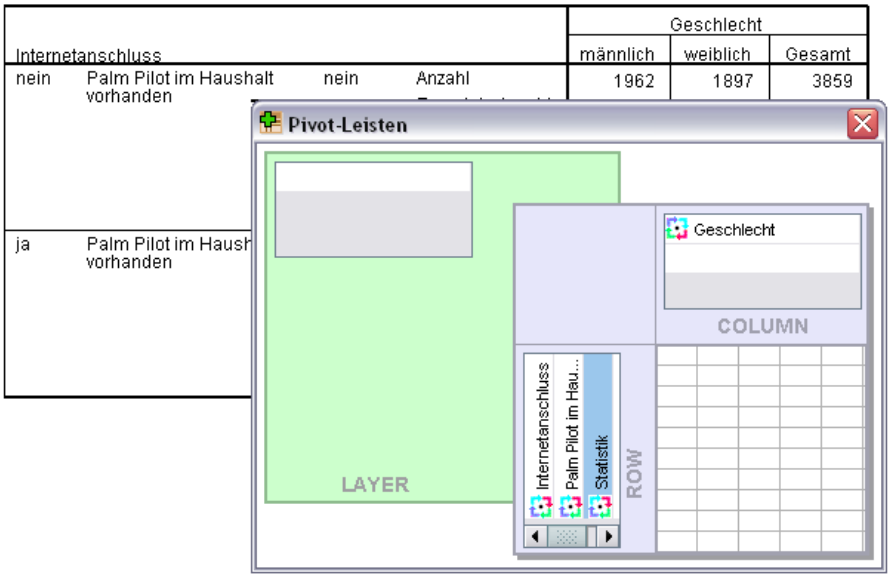
▶ Falls sie noch nicht aktiviert ist, doppelklicken Sie auf die Tabelle *Palm Pilot im Haushalt vorhanden * Geschlecht * Internetanschluss Kreuztabelle*. um sie zu aktivieren.

▶ Wenn das Fenster “Pivot-Leisten” ausgeblendet ist, wählen Sie die folgenden Menübefehle aus:

Pivot
  Pivot-Leisten

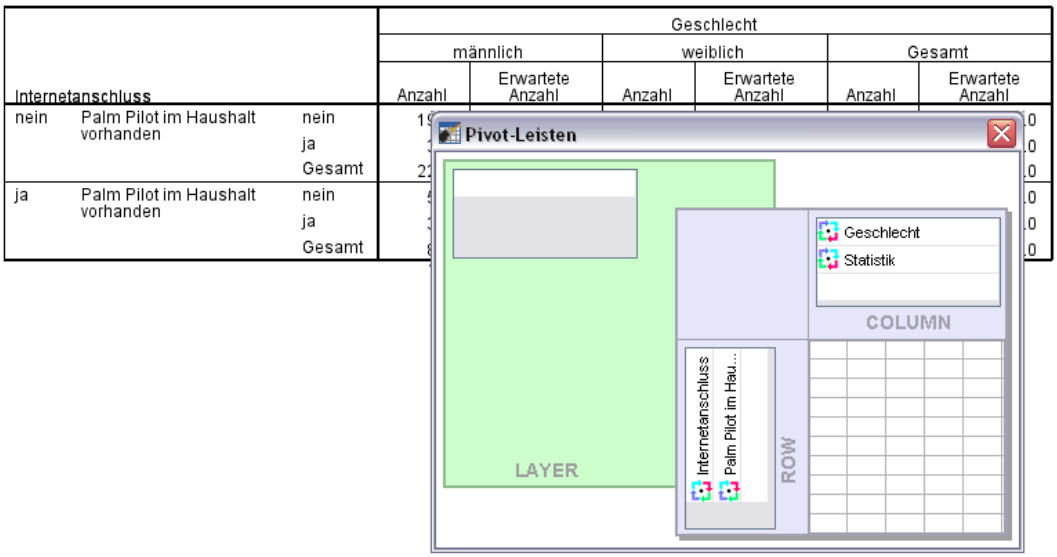
Mithilfe von Pivot-Leisten können Daten zwischen Spalten, Daten und Schichten verschoben werden.

Abbildung 7-4
Pivot-Leisten



▶ Ziehen Sie das Element *Statistik* aus der Zeilendimension in die Spaltendimension, unterhalb von *Geschlecht*. Die Tabelle wird unmittelbar neu konfiguriert und die Änderungen werden angezeigt.

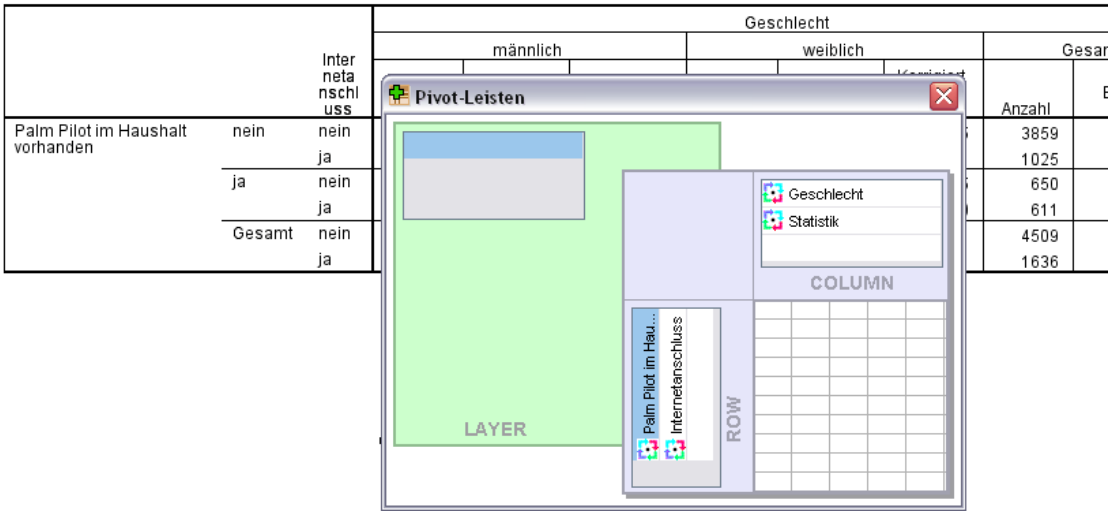
Abbildung 7-5
Verschieben von Zeilen in Spalten



Die Reihenfolge der Elemente auf der Pivot-Leiste entspricht der Reihenfolge der Elemente in der Tabelle.

▶ Ziehen Sie das Element *Palm Pilot im Haushalt vorhanden* vor das Element *Internetanschluss* in der Zeilendimension und legen Sie es dort ab, um die Reihenfolge dieser beiden Zeilen zu vertauschen.

Abbildung 7-6
Vertauschen von Variablen in der Zeilendimension

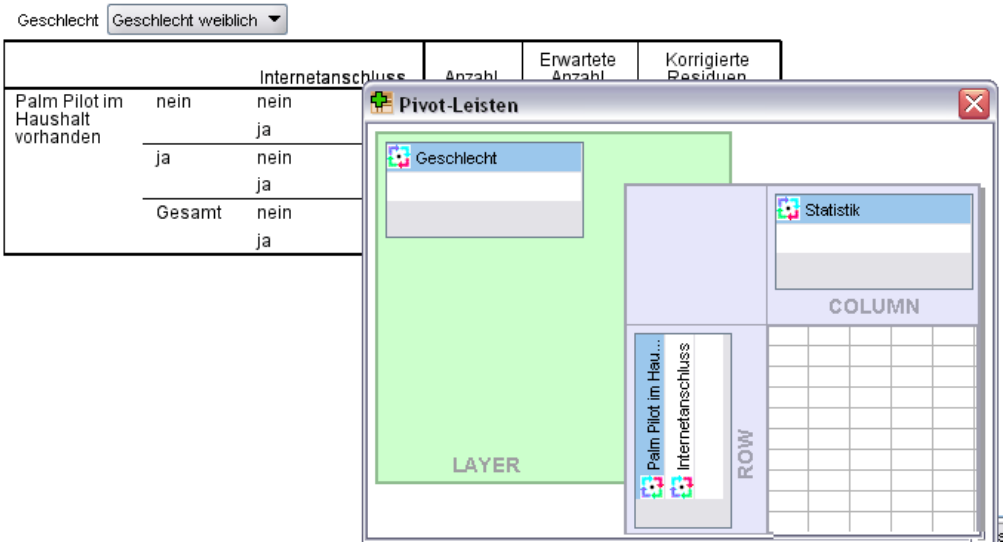


Erstellen und Anzeigen von Schichten

Schichten können den Umgang mit großen Tabellen mit verschachtelten Datenkategorien erleichtern. Durch das Erstellen von Schichten wird die Tabelle übersichtlicher und damit leichter verständlich.

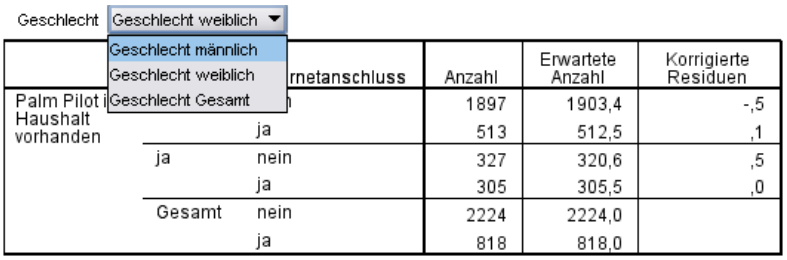
▶ Ziehen Sie das Element *Geschlecht* aus der Spaltendimension in die Schichtendimension.

Abbildung 7-7
Pivot-Symbol "Geschlecht" in der Schichtendimension



Um eine andere Schicht anzuzeigen, wählen Sie eine Kategorie aus der Dropdown-Liste in der Tabelle aus.

Abbildung 7-8
Auswählen einer Schicht



Bearbeiten von Tabellen

Wenn Sie keine eigene Tabellenvorlage erstellt haben, werden Pivot-Tabellen mit einer Standardformatierung erstellt. Die Formatierung des gesamten Texts in einer Tabelle kann geändert werden. Sie können unter anderem folgende Formatierungen ändern: Name der Schriftart, Schriftgröße, Schriftschnitt (fett oder kursiv) und Farbe.

▶ Doppelklicken Sie auf die Tabelle *Schulabschluss*.

▶ Wenn die Formatierungs-Symbolleiste ausgeblendet ist, wählen Sie die folgenden Menübefehle aus:

Ansicht
  Symbolleiste

▶ Klicken Sie auf den Titel *Schulabschluss*.

▶ Wählen Sie in der Liste der Schriftgrößen 12 aus.

▶ Um die Farbe des Titeltexts zu ändern, klicken Sie auf das Werkzeug für die Textfarbe und wählen Sie eine neue Farbe aus.

Abbildung 7-9
Neu formatierter Titeltext in der Pivot-Tabelle

Schulabschluss

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	kein Schulabschluss	1390	21,7	21,7	21,7
	Hauptschulabschluss	1936	30,2	30,2	52,0
	Realschulabschluss	1360	21,2	21,2	73,2
	Abitur	1355	21,2	21,2	94,4
	Universitätsabschluss	359	5,6	5,6	100,0
	Gesamt	6400	100,0	100,0	

Sie können auch den Inhalt der Tabellen und Beschriftungen bearbeiten. Sie können beispielsweise den Titel dieser Tabelle ändern.

▶ Doppelklicken Sie auf den Titel.

▶ Geben Sie als neue Beschriftung Ausbildungsniveau ein.

Anmerkung: Wenn die Werte in einer Tabelle geändert werden, werden Gesamtergebnisse und andere Statistiken nicht neu berechnet.

Ausblenden von Zeilen und Spalten

Einige der in der Tabelle dargestellten Daten sind möglicherweise unwichtig oder machen die Tabelle unnötig kompliziert. Sie können ganze Zeilen oder Spalten ausblenden, ohne dass Daten verlorengehen.

▶ Falls sie noch nicht aktiviert ist, doppelklicken Sie auf die Tabelle *Ausbildungsniveau*. um sie zu aktivieren.

▶ Klicken Sie auf die Spaltenbeschriftung *Gültige Prozente*, um diese Spalte auszuwählen.

▶ Wählen Sie im Menü “Bearbeiten” bzw. im Kontextmenü folgende Optionen:

Auswählen
  Datenzellen und Beschriftung

▶ Wählen Sie im Menü “Ansicht” die Option Ausblenden bzw. im Kontextmenü die Option Kategorie ausblenden.

Die Spalte wird nun ausgeblendet, jedoch nicht gelöscht.

Abbildung 7-10
Die Spalte "Gültige Prozente" wurde in der Tabelle ausgeblendet.

Schulabschluss

		Häufigkeit	Prozent	Kumulierte Prozente
Gültig	kein Schulabschluss	1390	21,7	21,7
	Hauptschulabschluss	1936	30,2	52,0
	Realschulabschluss	1360	21,2	73,2
	Abitur	1355	21,2	94,4
	Universitätsabschluss	359	5,6	100,0
	Gesamt	6400	100,0	

So blenden Sie die Spalte wieder ein:

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Ansicht
  Alles einblenden

Zeilen können genau wie Spalten aus- und eingeblendet werden.

Ändern der Anzeigeformate für Daten

Sie können das Anzeigeformat für Daten in Pivot-Tabellen ganz einfach ändern.

▶ Falls sie noch nicht aktiviert ist, doppelklicken Sie auf die Tabelle *Ausbildungsniveau*. um sie zu aktivieren.

▶ Klicken Sie auf die Spaltenbeschriftung *Prozent*, um diese Spalte auszuwählen.

▶ Wählen Sie im Menü "Bearbeiten" bzw. im Kontextmenü folgende Optionen:

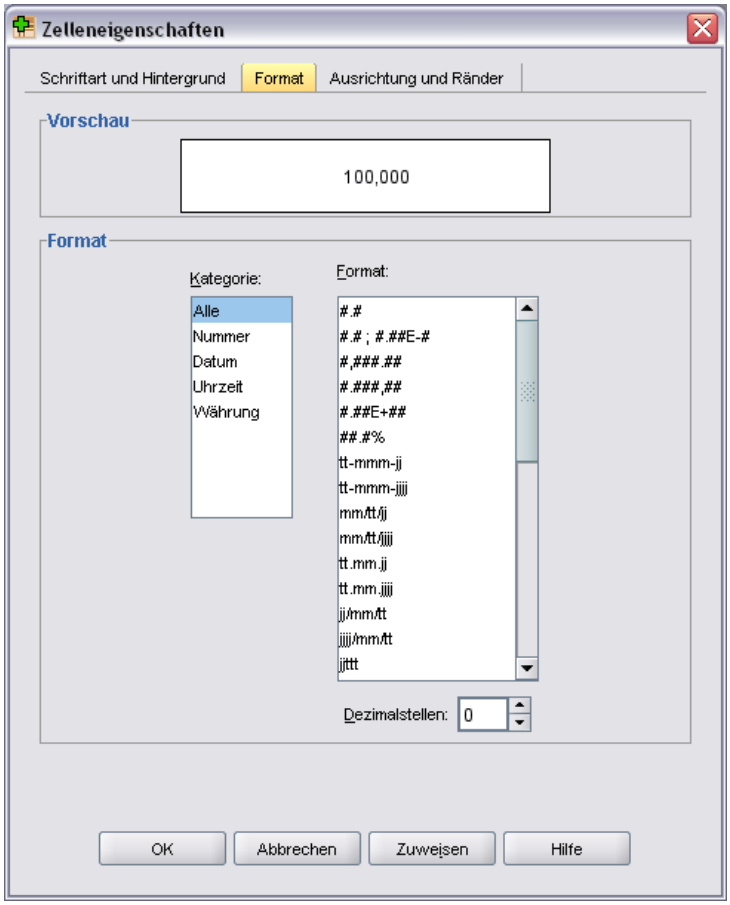
Auswählen
  Datenzellen

▶ Wählen Sie im Menü "Format" bzw. im Kontextmenü die Option Zelleneigenschaften.

▶ Klicken Sie auf die Registerkarte Formatwert.

▶ Geben Sie im Feld “Dezimalstellen” den Wert 0 ein, damit alle Dezimalpunkte in dieser Spalte ausgeblendet werden.

Abbildung 7-11
“Zelleneigenschaften”, Registerkarte “Format”



In diesem Dialogfeld können auch Datentyp und Format geändert werden.

▶ Wählen Sie in der Liste “Kategorie” einfach den gewünschten Typ und anschließend in der Liste “Format” das Format für diesen Typ aus.

▶ Klicken Sie auf OK oder Zuweisen, um Ihre Änderungen zu übernehmen.

Abbildung 7-12
Spalte “Prozent” mit ausgeblendeten Dezimalstellen

Schulabschluss

		Häufigkeit	Prozent	Kumulierte Prozente
Gültig	kein Schulabschluss	1390	22	21,7
	Hauptschulabschluss	1936	30	52,0
	Realschulabschluss	1360	21	73,2
	Abitur	1355	21	94,4
	Universitätsabschluss	359	6	100,0
	Gesamt	6400	100	

Die Dezimalstellen in der Spalte *Prozent* sind jetzt ausgeblendet.

Tabellenvorlagen

Das Erscheinungsbild der Tabellen ist der entscheidende Faktor für die Präsentation übersichtlicher, prägnanter und aussagekräftiger Ergebnisse. Wenn eine Tabelle unübersichtlich ist, sind die darin enthaltenen Informationen möglicherweise schwer verständlich.

Verwenden von vordefinierten Formaten

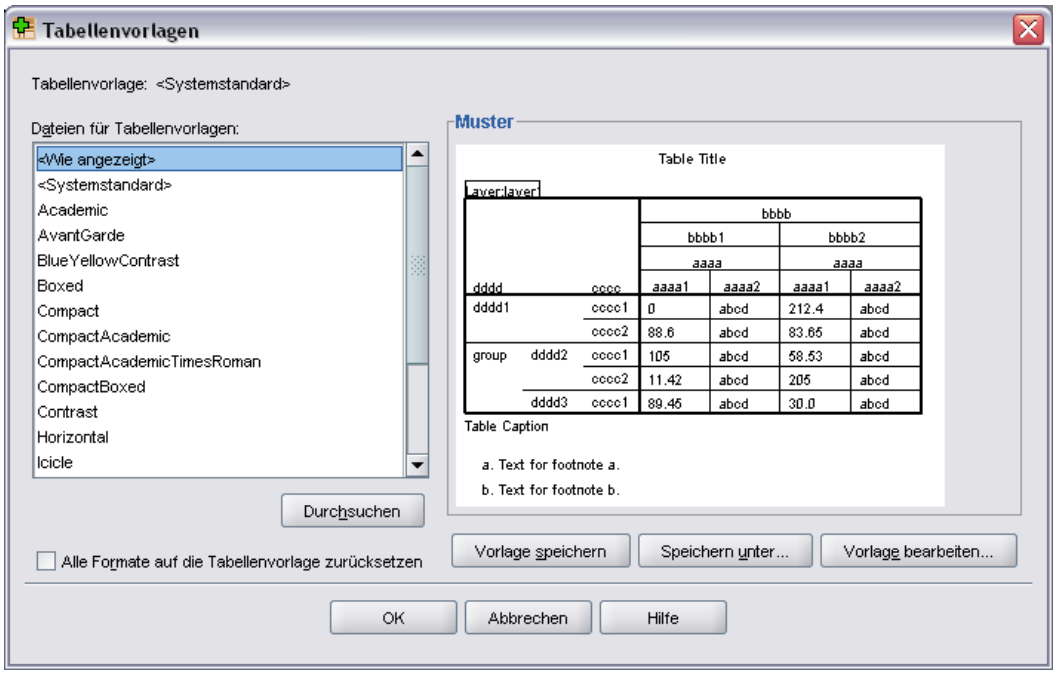
▶ Doppelklicken Sie auf die Tabelle *Verheiratet*.

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Format
  Tabellenvorlagen...

Im Dialogfeld "Tabellenvorlagen" wird eine Reihe vordefinierter Stile angezeigt. Wenn Sie einen Stil in der Liste markieren, wird im Vorschaufenster rechts daneben eine Vorschau angezeigt.

Abbildung 7-13
Dialogfeld "Tabellenvorlagen"



Sie können den Stil unverändert übernehmen oder diesen nach Ihren Bedürfnissen anpassen.

▶ Wenn Sie einen vorhandenen Stil verwenden möchten, wählen Sie ihn aus und klicken Sie auf OK.

Anpassen von Tabellenvorlagen

Sie können ein Format individuell anpassen. Fast alle Aspekte einer Tabelle von der Vordergrundfarbe bis hin zur Rahmenart können angepasst werden.

▶ Doppelklicken Sie auf die Tabelle *Verheiratet*.

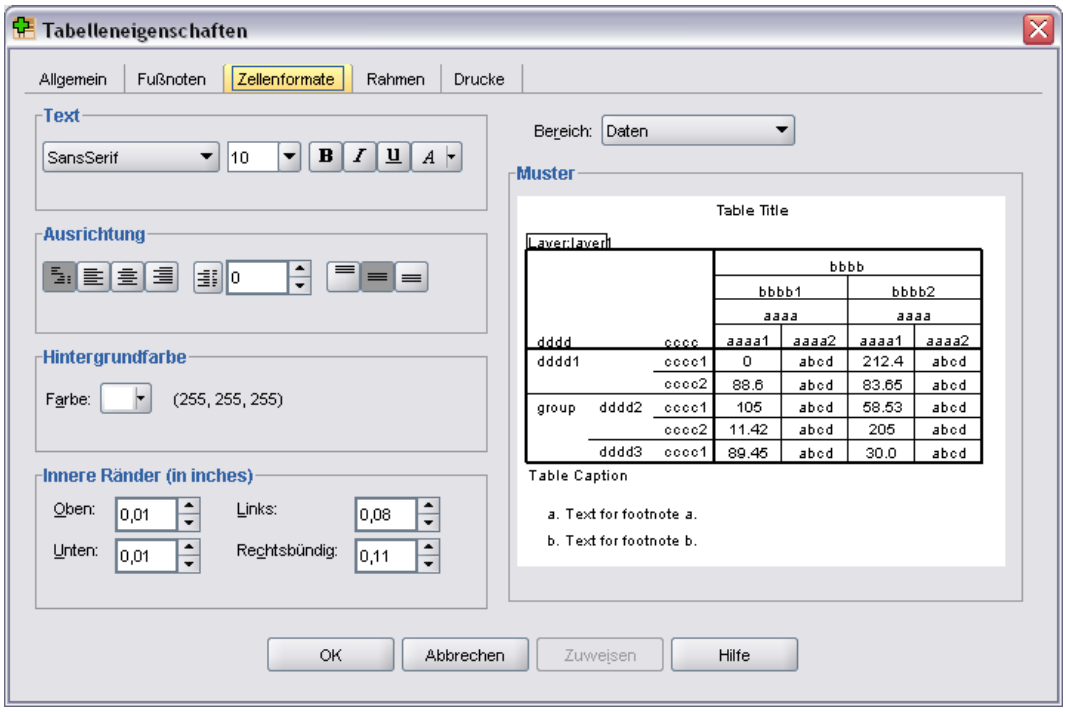
▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Format
  Tabellenvorlagen...

▶ Wählen Sie den Stil aus, der dem gewünschten Format am ehesten entspricht, und klicken Sie auf Tabellenvorlage bearbeiten.

▶ Wenn Sie auf die Registerkarte Zellenformate klicken, werden die Formatierungsoptionen angezeigt.

Abbildung 7-14
Dialogfeld "Tabelleneigenschaften"



Zu den Formatierungsoptionen gehören: Name der Schriftart, Schriftgröße, Schriftschnitt und Farbe. Weitere Optionen sind Ausrichtung, Farben für Text und Hintergrund und Randgrößen.

Im Vorschaufenster rechts können Sie erkennen, wie sich Änderungen der Formatierung auf die Tabelle auswirken. Jeder Bereich der Tabelle kann auf andere Weise formatiert werden. Der Titel soll sich von den angezeigten Daten vermutlich unterscheiden. Um einen Tabellenbereich für die Bearbeitung auszuwählen, können Sie entweder einen Bereich anhand des Namens in der Dropdown-Liste "Bereich" auswählen oder im Vorschaufenster auf den zu ändernden Bereich klicken.

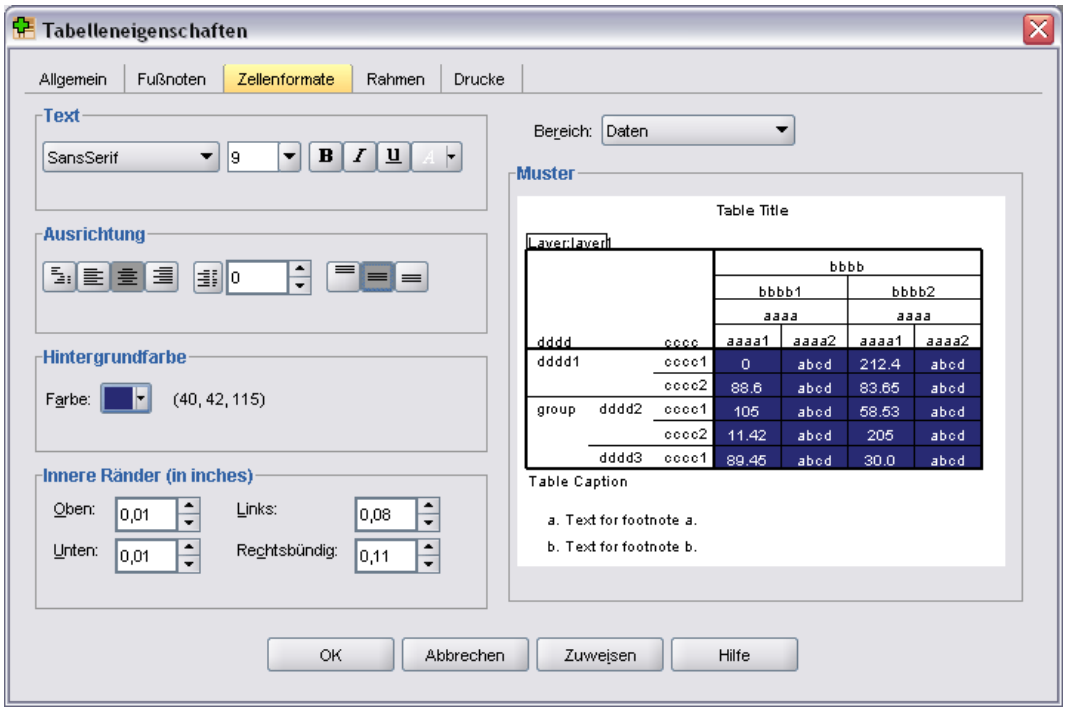
▶ Wählen Sie in der Dropdown-Liste “Bereich” die Option Daten aus.

▶ Wählen Sie aus der Dropdown-Palette “Hintergrund” einen neue Farbe aus.

▶ Wählen Sie anschließend eine neue Textfarbe aus.

Der neue Stil wird im Vorschaufenster angezeigt.

Abbildung 7-15
So ändern Sie Zellenformate von Tabellen:



▶ Klicken Sie auf OK, um zum Dialogfeld “Tabellenvorlagen” zurückzukehren.

Sie können den neuen Stil speichern, damit Sie ihn für neue Tabellen wiederverwenden können.

▶ Klicken Sie auf Speichern unter.

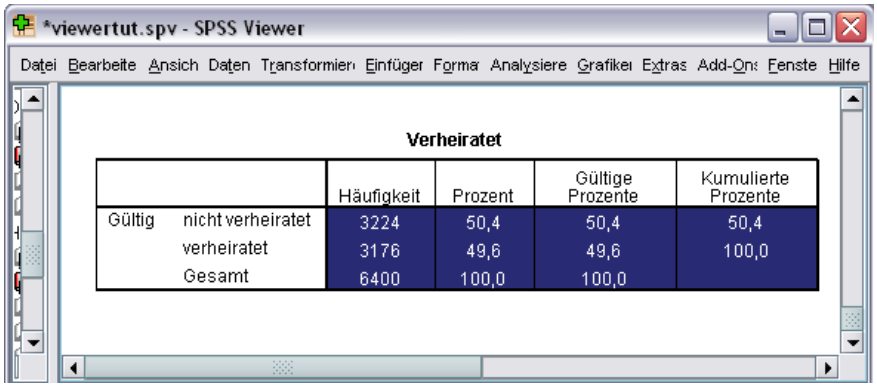
▶ Wechseln Sie zum gewünschten Zielverzeichnis und geben Sie im Textfeld “Dateiname” einen Namen für den neuen Stil ein.

▶ Klicken Sie auf Speichern.

▶ Klicken Sie auf OK, um die Änderungen zu übernehmen und zum Viewer zurückzukehren.

Die Tabelle enthält jetzt die angegebenen benutzerdefinierten Formatierungen.

Abbildung 7-16
Benutzerdefinierte Tabellenvorlage



		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	nicht verheiratet	3224	50,4	50,4	50,4
	verheiratet	3176	49,6	49,6	100,0
	Gesamt	6400	100,0	100,0	

Ändern der Standard-Tabellenformate

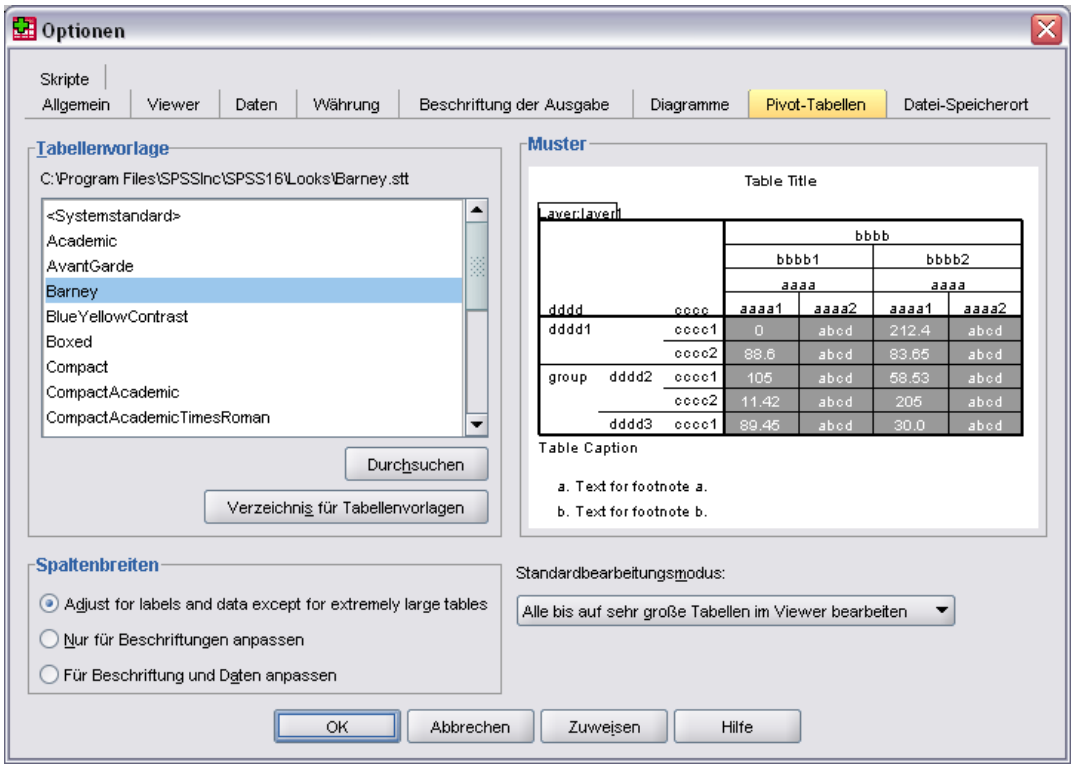
Auch wenn Sie die Möglichkeit haben, das Format einer Tabelle nach ihrer Erstellung zu ändern, kann es effizienter sein, das Standard-Tabellenformat zu drucken, sodass Sie nicht bei jeder Erstellung einer Tabelle das Format ändern müssen.

Wenn Sie die Standard-Tabellenvorlage für Pivot-Tabellen ändern möchten, wählen Sie die folgenden Menübefehle aus:

Bearbeiten
  Optionen...

▶ Klicken Sie im Dialogfeld “Optionen” auf die Registerkarte Pivot-Tabellen.

Abbildung 7-17
Dialogfeld “Optionen”



▶ Wählen Sie für die Tabellenvorlage den Stil aus, den Sie für alle neuen Tabellen verwenden möchten.

Im Vorschaufenster rechts können Sie eine Vorschau der verschiedenen Tabellenvorlagen anzeigen.

▶ Klicken Sie auf OK, um die Einstellungen zu speichern und das Dialogfeld zu schließen.

Alle Tabellen, die nach der Änderung der Standard-Tabellenvorlage erstellt werden, entsprechen automatisch den neuen Formatierungsvorschriften.

Ändern der anfänglichen Einstellungen für die Anzeige

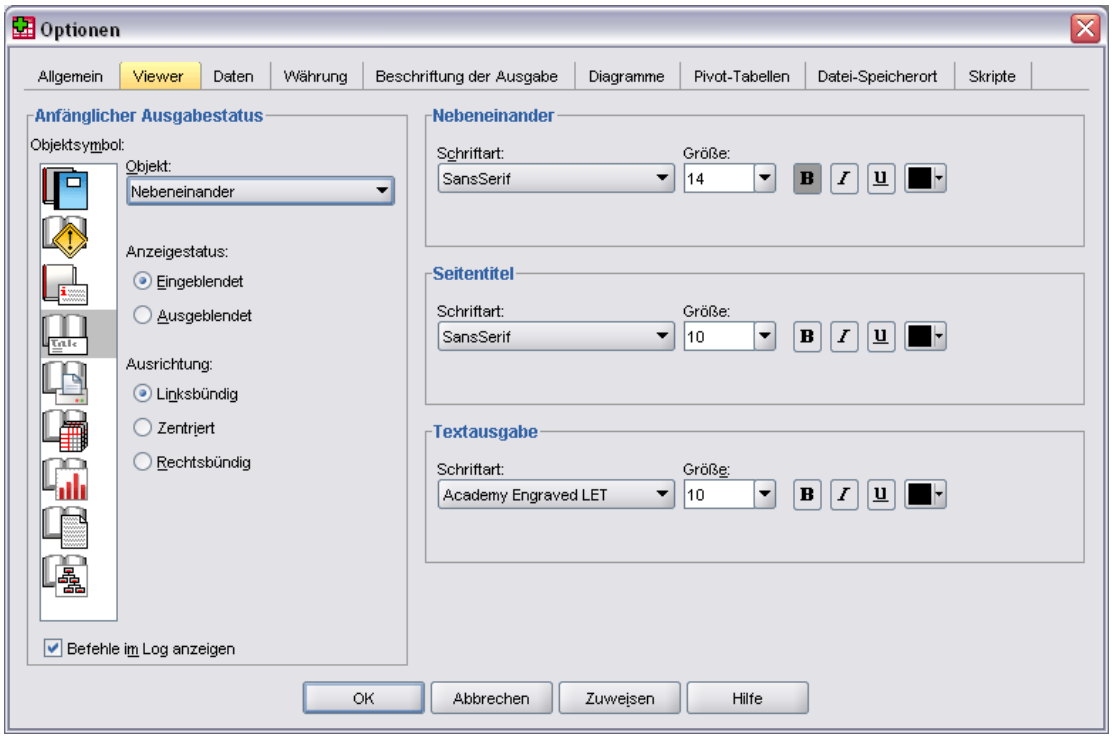
Zu den ursprünglichen Anzeigeeinstellungen gehören die Ausrichtung der Objekte im Viewer, die Breite des Viewer-Fensters und die Angabe, ob Objekte standardmäßig ein- oder ausgeblendet werden. So ändern Sie diese Einstellungen:

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Bearbeiten
  Optionen...

▶ Klicken Sie auf die Registerkarte Viewer.

Abbildung 7-18
Viewer-Optionen



Die Einstellungen werden für jedes Objekt einzeln festgelegt. Sie können beispielsweise die Anzeige von Diagrammen anpassen, ohne damit die Form zu ändern, in der die Tabellen angezeigt werden. Wählen Sie einfach das Objekt aus, das Sie anpassen möchten und nehmen Sie die gewünschten Änderungen vor.

▶ Klicken Sie auf das Symbol Titel, um dessen Einstellungen anzeigen zu lassen.

▶ Klicken Sie auf Zentriert, wenn alle Titel im Viewer horizontal zentriert angezeigt werden sollen.

Elemente wie Protokolle und Warnungen, die die Ausgabe häufig unübersichtlich machen, können ebenfalls ausgeblendet werden. Wenn Sie auf ein Symbol doppelklicken, werden die Anzeigeeigenschaften des Objekts automatisch geändert.

▶ Wenn Sie auf das Symbol Warnungen doppelklicken, werden Warnungen in der Ausgabe ausgeblendet.

▶ Klicken Sie auf OK, um die Änderungen zu speichern und das Dialogfeld zu schließen.

Anzeige von Variablen- und Wertelabels

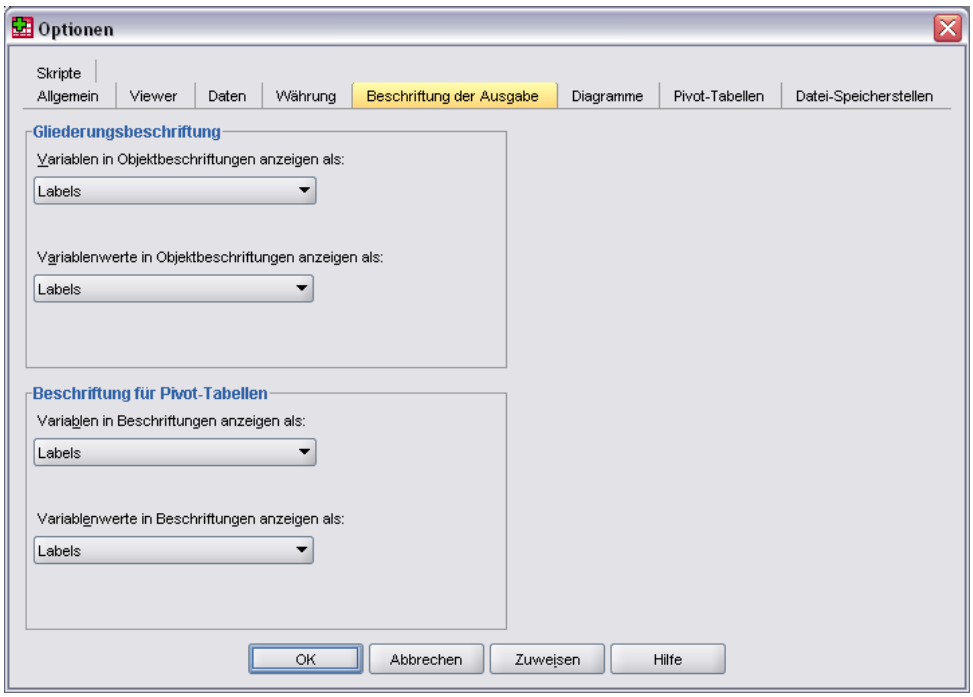
Das Anzeigen der Variablen- und Wertelabels ist meist wirkungsvoller als das Anzeigen des Variablennamens und des eigentlichen Datenwerts. In manchen Fällen kann es jedoch günstig sein, sowohl die Namen als auch die Labels anzuzeigen.

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Bearbeiten
  Optionen...

▶ Klicken Sie auf die Registerkarte Beschriftung der Ausgabe.

Abbildung 7-19
Optionen: Beschriftung der Ausgabe

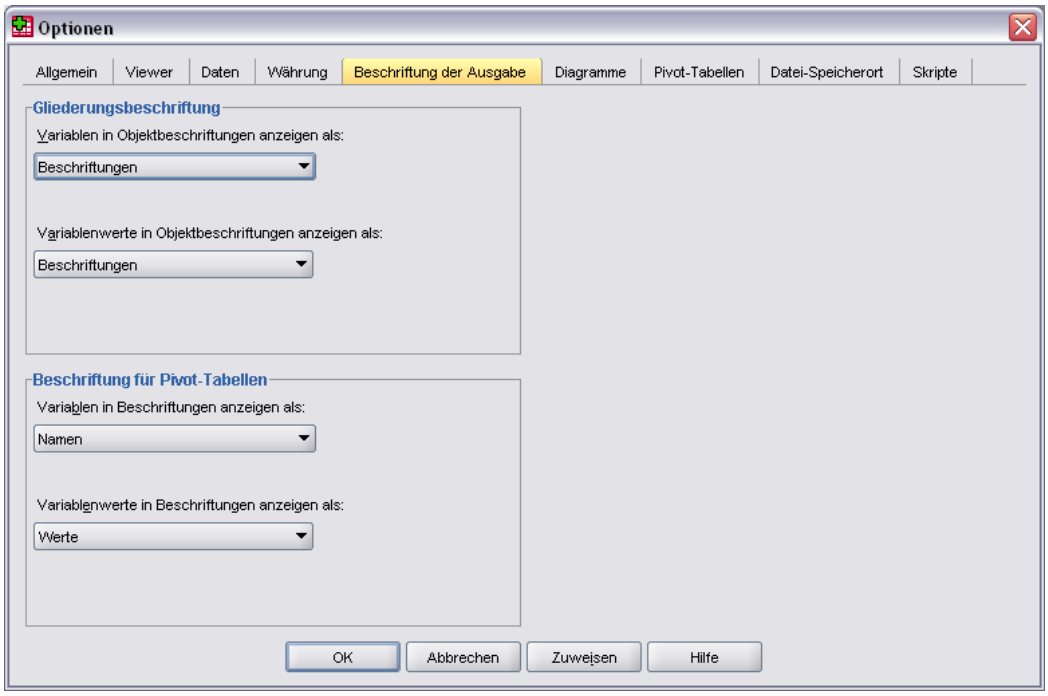


Sie können unterschiedliche Einstellungen für das Gliederungs- und das Inhaltsfenster angeben. So können Sie beispielsweise im Gliederungsfenster Labels und im Inhaltsfenster Variablennamen und Datenwerte anzeigen lassen:

▶ Wählen Sie im Gruppenfeld “Beschriftung für Pivot-Tabellen” aus der Dropdown-Liste “Variablen in Beschriftungen anzeigen als” die Option Namen aus, um Variablennamen anstelle von Labels anzuzeigen.

▶ Wählen Sie anschließend aus der Dropdown-Liste “Variablenwerte in Beschriftungen anzeigen als:” die Option **Werte** aus, um Datenwerte anstelle von Labels anzuzeigen.

Abbildung 7-20
Einstellungen für die Beschriftung der Pivot-Tabellen



In den weiteren in dieser Sitzung erstellten Tabellen werden diese Änderungen berücksichtigt.

Abbildung 7-21
Anzeige von Variablennamen und Werten

Verheiratet

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	0	3224	50,4	50,4	50,4
	1	3176	49,6	49,6	100,0
	Gesamt	6400	100,0	100,0	

Verwenden von Ergebnissen in anderen Anwendungen

Die Ergebnisse können in vielen Anwendungen verwendet werden. Sie könnten zum Beispiel eine Tabelle oder ein Diagramm in eine Präsentation oder einen Bericht einfügen.

Die folgenden Beispiele beziehen sich auf Microsoft Word, sind aber möglicherweise auch auf andere Textverarbeitungsprogramme anwendbar.

Einfügen von Ergebnissen als Tabellen in Word

Sie können Pivot-Tabellen in Word als echte Word-Tabellen einfügen. Alle Tabellenattribute wie Schriftgrößen und Farben werden beibehalten. Da die Tabelle im Word-Tabellenformat eingefügt wird, können Sie sie in Word genau wie jede andere Tabelle bearbeiten.

▶ Klicken Sie im Viewer auf die Tabelle *Verheiratet*.

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Bearbeiten
  Kopieren

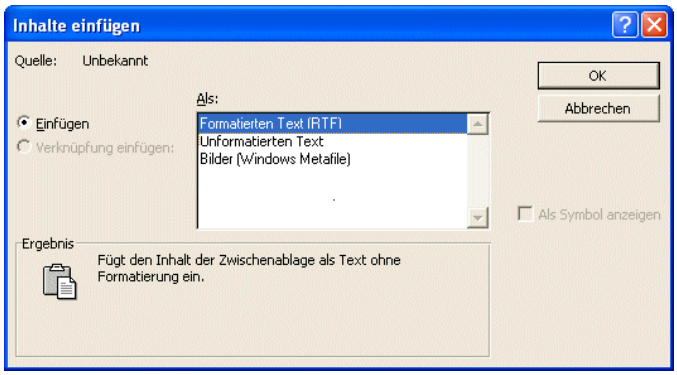
▶ Öffnen Sie das Textverarbeitungsprogramm.

▶ Wählen Sie die folgenden Befehle aus den Menüs der Textverarbeitung aus:

Bearbeiten
  Inhalte einfügen...

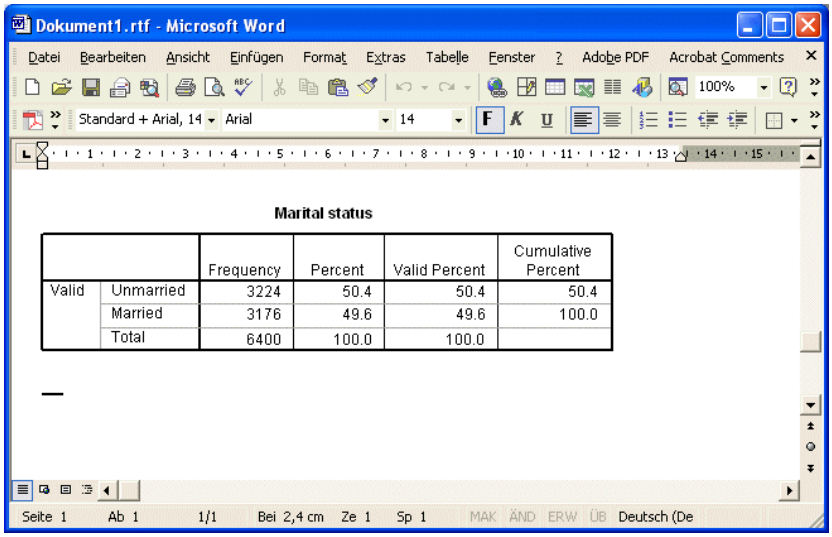
▶ Wählen Sie im Dialogfeld "Inhalte einfügen" die Option Formatierten Text (RTF) aus.

Abbildung 7-22
Dialogfeld "Inhalte einfügen"



▶ Klicken Sie auf OK, um die Ergebnisse in das aktuelle Dokument einzufügen.

Abbildung 7-23
In Word angezeigte Pivot-Tabelle



Die Tabelle wird nun im Dokument angezeigt. Sie können das Format anpassen, die Daten bearbeiten und die Größe der Tabelle nach Ihren Wünschen ändern.

Einfügen von Ergebnissen als Text

Pivot-Tabellen können in andere Anwendungen als einfacher Text kopiert werden. Bei diesem Verfahren werden zwar keine Formatierungen übernommen, Sie können die Tabellendaten jedoch nach dem Einfügen in die Zielanwendung bearbeiten.

▶ Klicken Sie im Viewer auf die Tabelle *Verheiratet*.

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Bearbeiten
  Kopieren

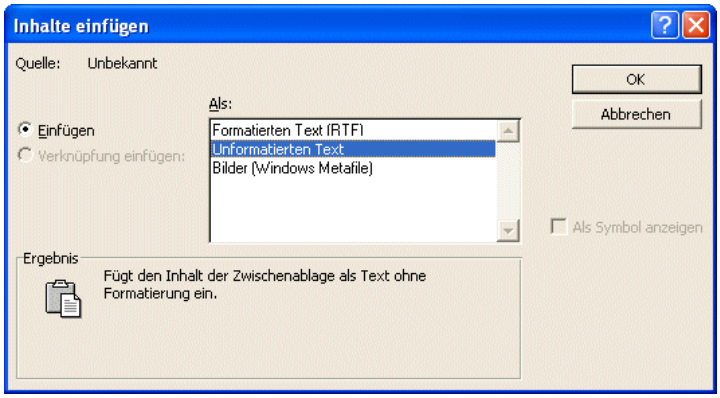
▶ Öffnen Sie das Textverarbeitungsprogramm.

▶ Wählen Sie die folgenden Befehle aus den Menüs der Textverarbeitung aus:

Bearbeiten
  Inhalte einfügen...

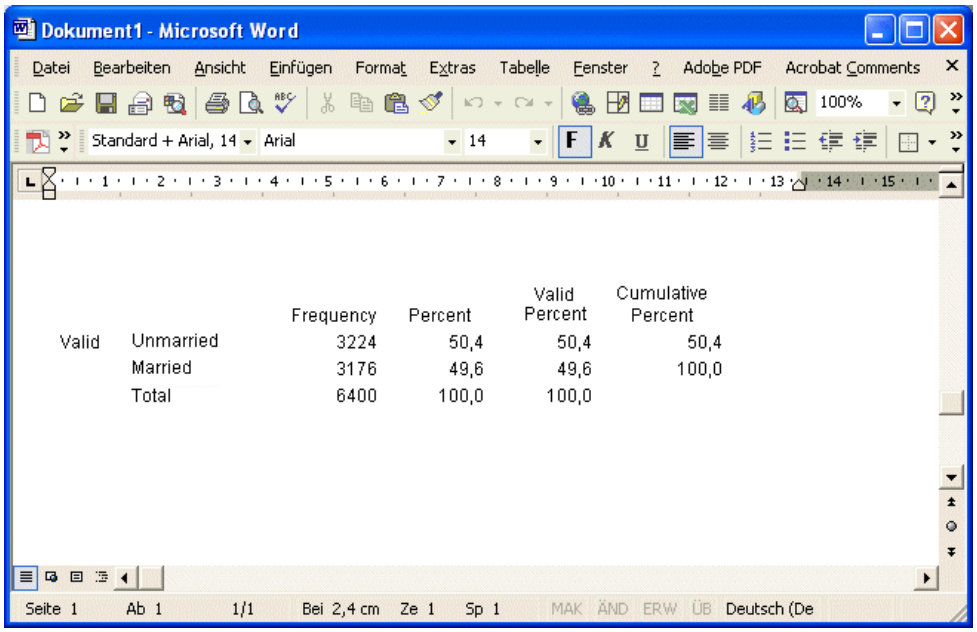
▶ Wählen Sie im Dialogfeld “Inhalte Einfügen” die Option Unformatierter Text aus.

Abbildung 7-24
Dialogfeld “Inhalte einfügen”



▶ Klicken Sie auf OK, um die Ergebnisse in das aktuelle Dokument einzufügen.

Abbildung 7-25
In Word angezeigte Pivot-Tabelle



		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Unmarried	3224	50,4	50,4	50,4
	Married	3176	49,6	49,6	100,0
	Total	6400	100,0	100,0	

Die Spalten der Tabelle sind durch Tabulatoren getrennt. Sie können die Spaltenbreiten ändern, indem Sie die Tabstops in der Textverarbeitung entsprechend korrigieren.

Exportieren der Ergebnisse in Microsoft Word-, PowerPoint- und Excel-Dateien

Sie können Ergebnisse in eine Microsoft Word-, PowerPoint- oder Excel-Datei exportieren. Sie können ausgewählte Elemente oder auch alle Elemente im Viewer exportieren. In diesem Abschnitt werden die Dateien *msouttut.spv* und *demo.sav* verwendet. Für weitere Informationen siehe Beispieldateien in Anhang A auf S. 163.

Anmerkung: Der Export nach PowerPoint ist nur unter Windows-Betriebssystemen und nicht in der Studentenversion verfügbar.

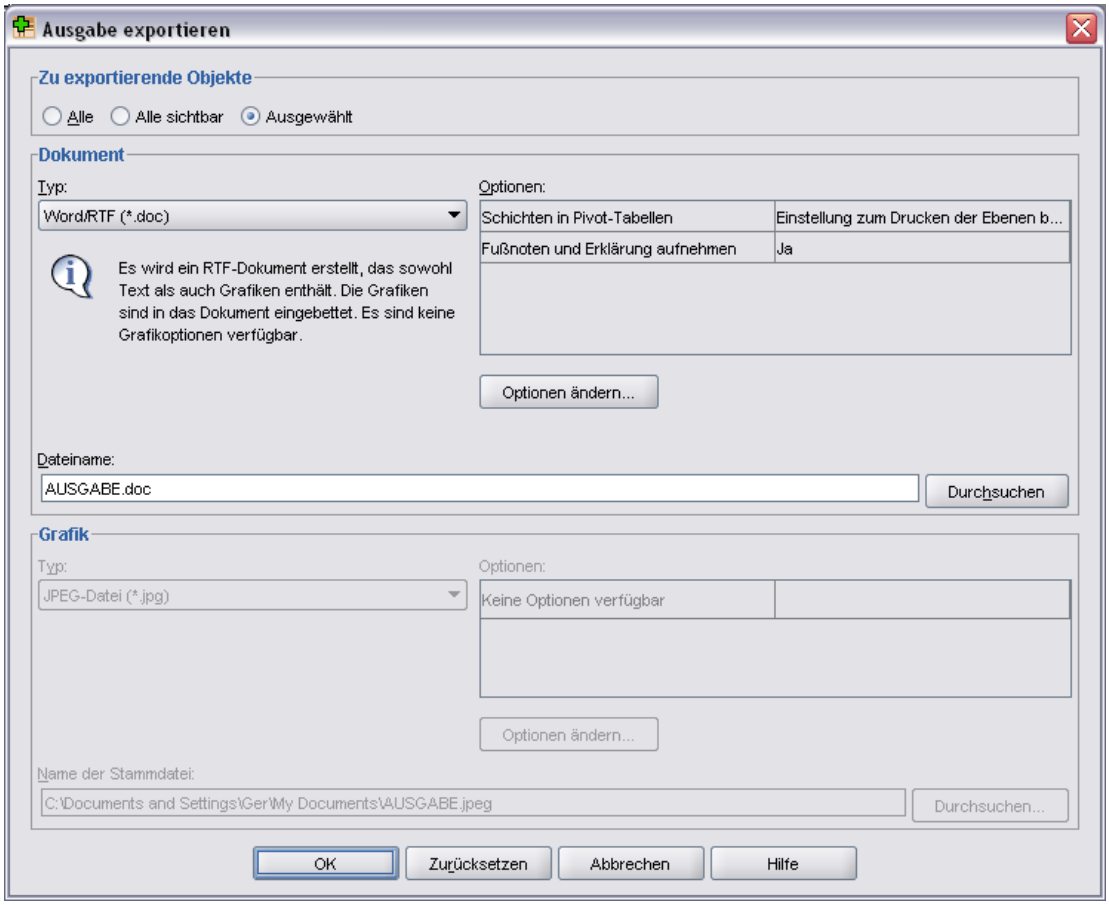
Im Gliederungsfenster des Viewers können Sie spezielle Elemente für den Export auswählen. Sie müssen keine speziellen Elemente auswählen.

▶ Wählen Sie die folgenden Befehle aus den Menüs des Viewers aus:

Datei
  Exportieren...

Anstatt alle Objekte im Viewer zu exportieren, können Sie auch nur sichtbare Objekte (geöffnete Bücher im Gliederungsfenster) exportieren oder nur die Objekte, die Sie im Gliederungsfenster ausgewählt haben. Wenn Sie im Gliederungsfenster keine Elemente ausgewählt haben, steht Ihnen die Option zum Export ausgewählter Objekte nicht zur Verfügung.

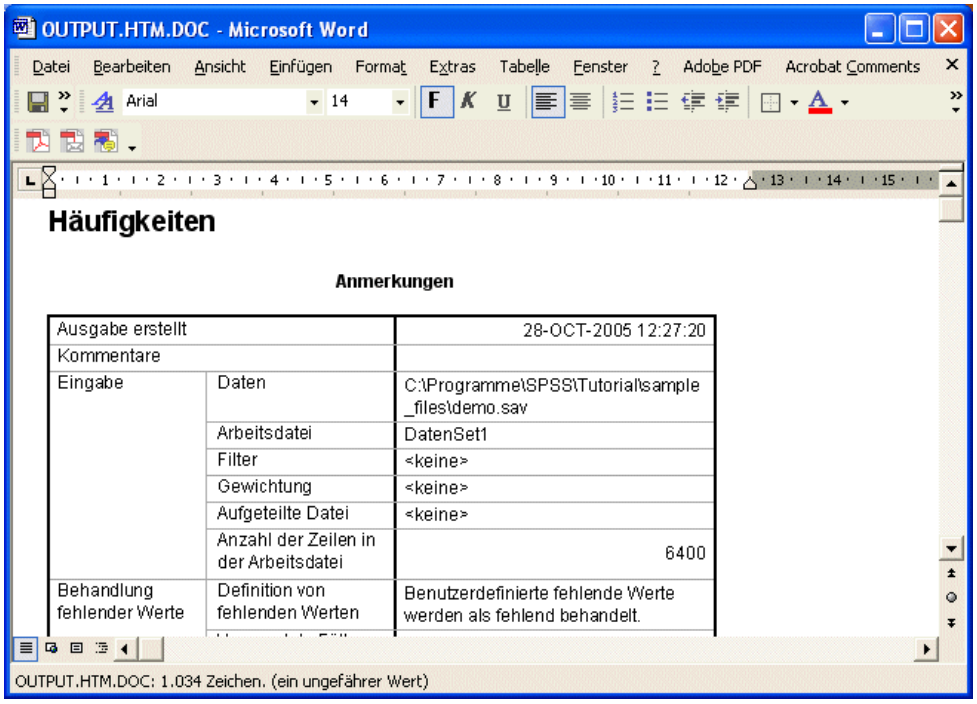
Abbildung 7-26
Dialogfeld "Ausgabe exportieren"



- Wählen Sie in der Gruppe “Zu exportierende Objekte” die Option Alle aus.
- Wählen Sie in der Dropdown-Liste “Typ” die Option Word/RTF-Datei (*.doc) aus.
- Klicken Sie auf OK, um die Word-Datei zu erstellen.

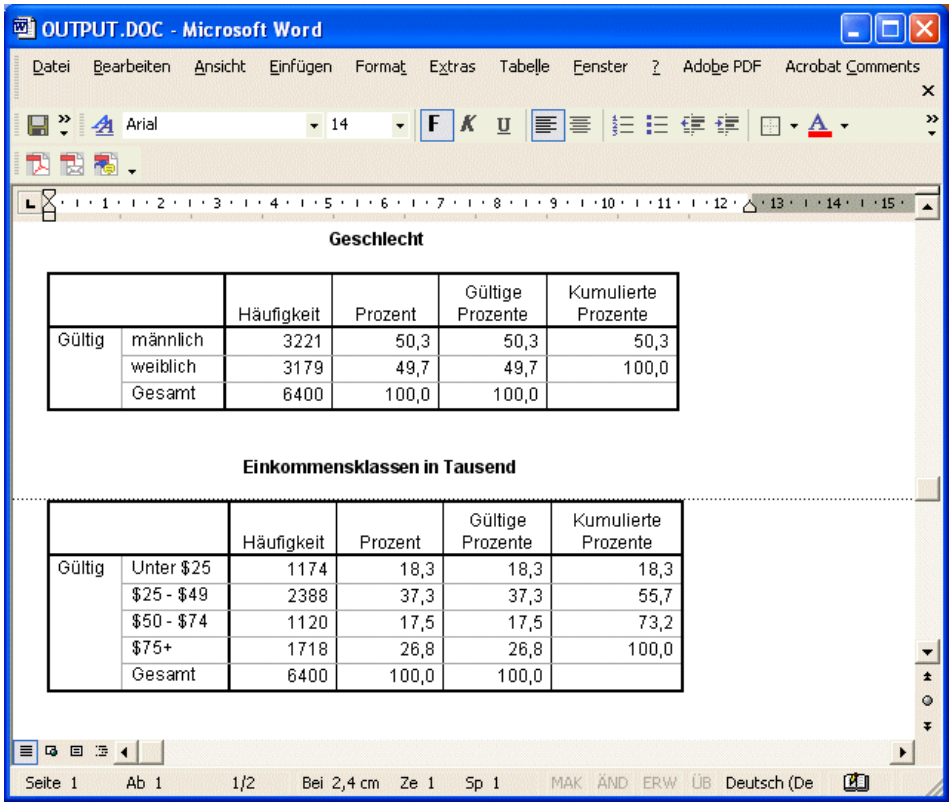
Wenn Sie die resultierende Datei in Word öffnen, können Sie sehen, wie die Ergebnisse exportiert wurden. Anmerkungen, die keine sichtbaren Objekte sind, werden in Word angezeigt, da Sie ausgewählt haben, dass alle Objekte exportiert werden sollen.

Abbildung 7-27
Output.doc in Word



Aus Pivot-Tabellen werden Word-Tabellen. Dabei bleiben alle Formatierungen der ursprünglichen Pivot-Tabelle erhalten, einschließlich Schriftarten, Farben, Rahmen usw.

Abbildung 7-28
Pivot-Tabellen in Word



Geschlecht

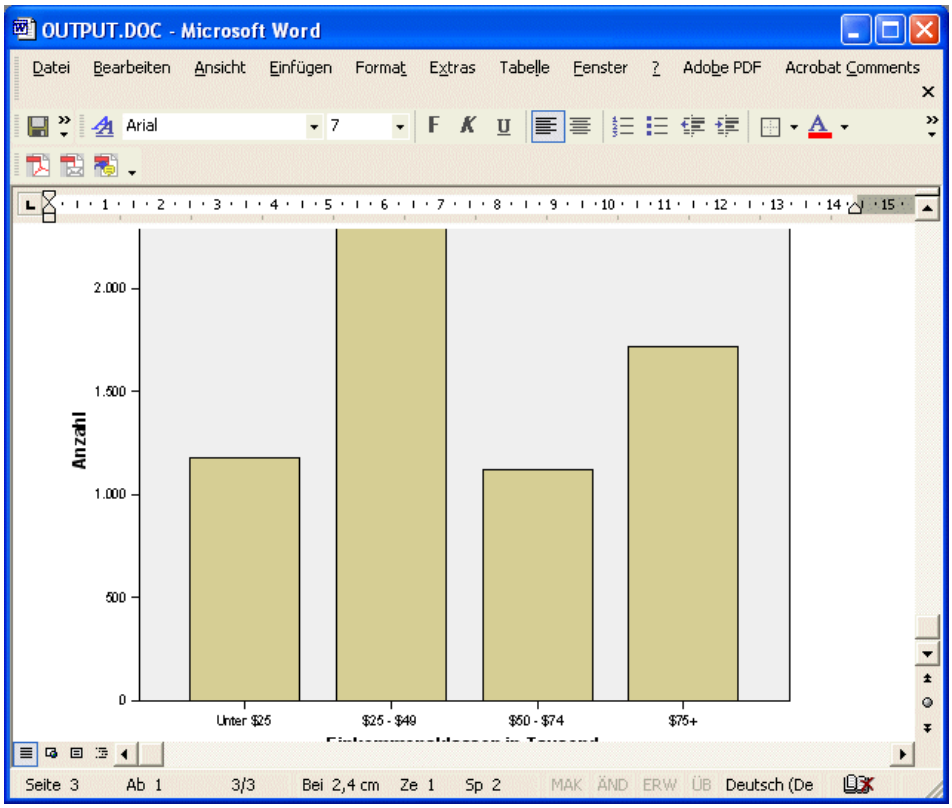
		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	männlich	3221	50,3	50,3	50,3
	weiblich	3179	49,7	49,7	100,0
	Gesamt	6400	100,0	100,0	

Einkommensklassen in Tausend

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	Unter $25	1174	18,3	18,3	18,3
	$25 - $49	2388	37,3	37,3	55,7
	$50 - $74	1120	17,5	17,5	73,2
	$75+	1718	26,8	26,8	100,0
	Gesamt	6400	100,0	100,0	

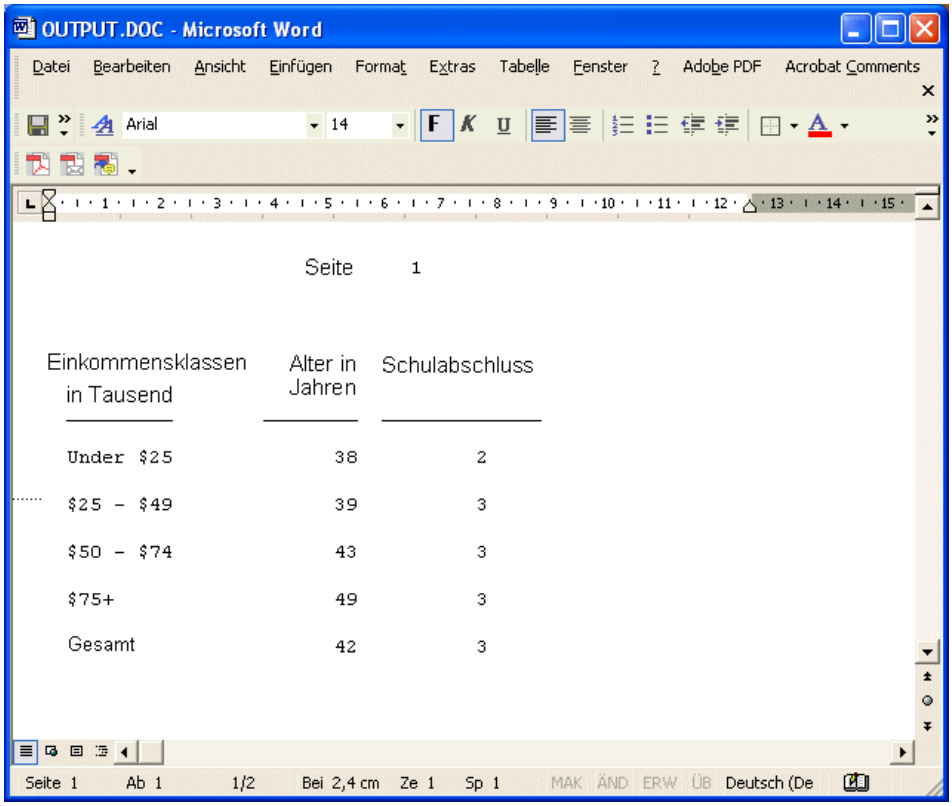
Diagramme werden als Grafiken in das Word-Dokument aufgenommen.

Abbildung 7-29
Diagramme in Word



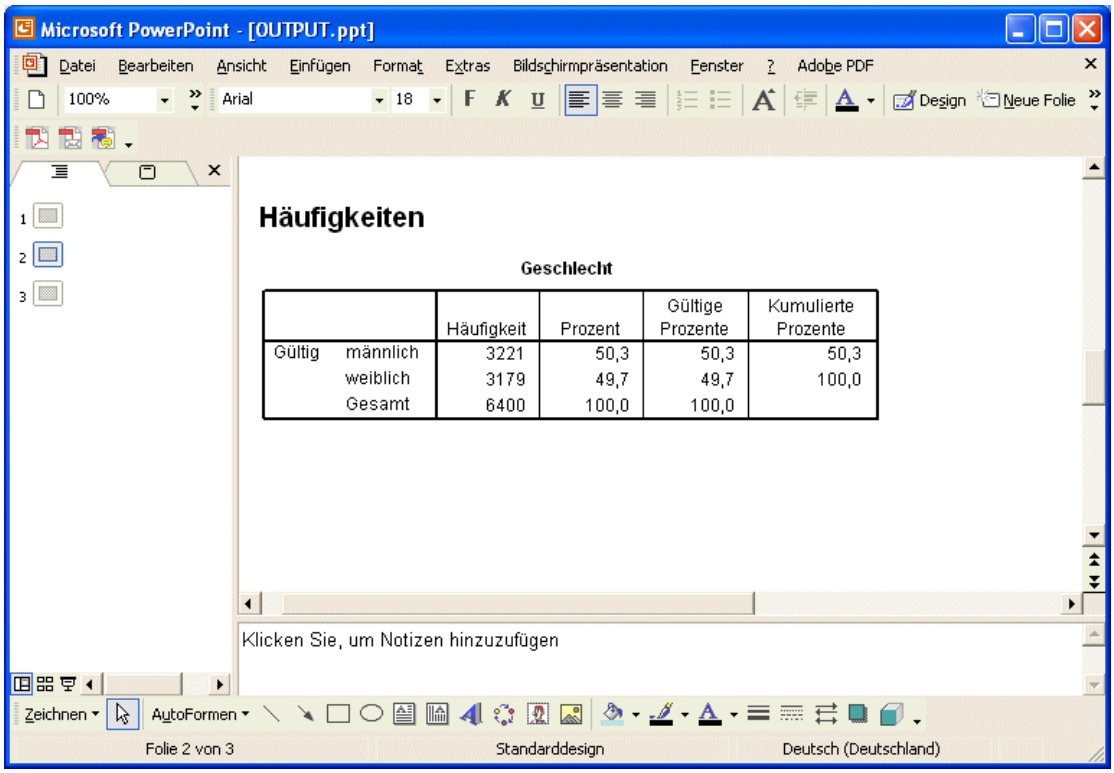
Die Textausgabe wird in derselben Schriftart angezeigt, die auch für das Textobjekt im Viewer verwendet wird. Zur richtigen Ausrichtung sollte für Textausgaben ein Zeichensatz mit fester Zeichenbreite (festem Abstand) verwendet werden.

Abbildung 7-30
Textausgabe in Word



Beim Exportieren in eine PowerPoint-Datei werden die einzelnen exportierten Elemente jeweils auf eine separate Folie plaziert. Aus Pivot-Tabellen werden Word-Tabellen in PowerPoint. Dabei bleiben alle Formatierungen der ursprünglichen Pivot-Tabelle erhalten, einschließlich Schriftarten, Farben, Rahmen usw.

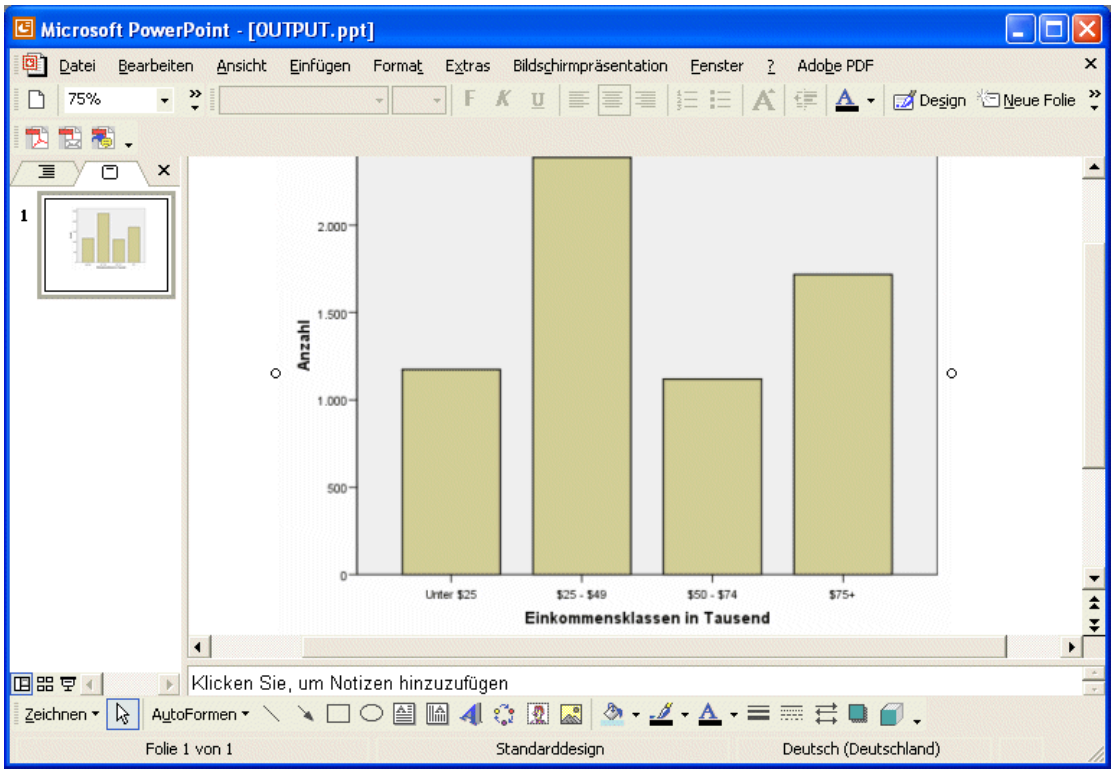
Abbildung 7-31
Pivot-Tabellen in PowerPoint



		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	männlich	3221	50,3	50,3	50,3
	weiblich	3179	49,7	49,7	100,0
	Gesamt	6400	100,0	100,0	

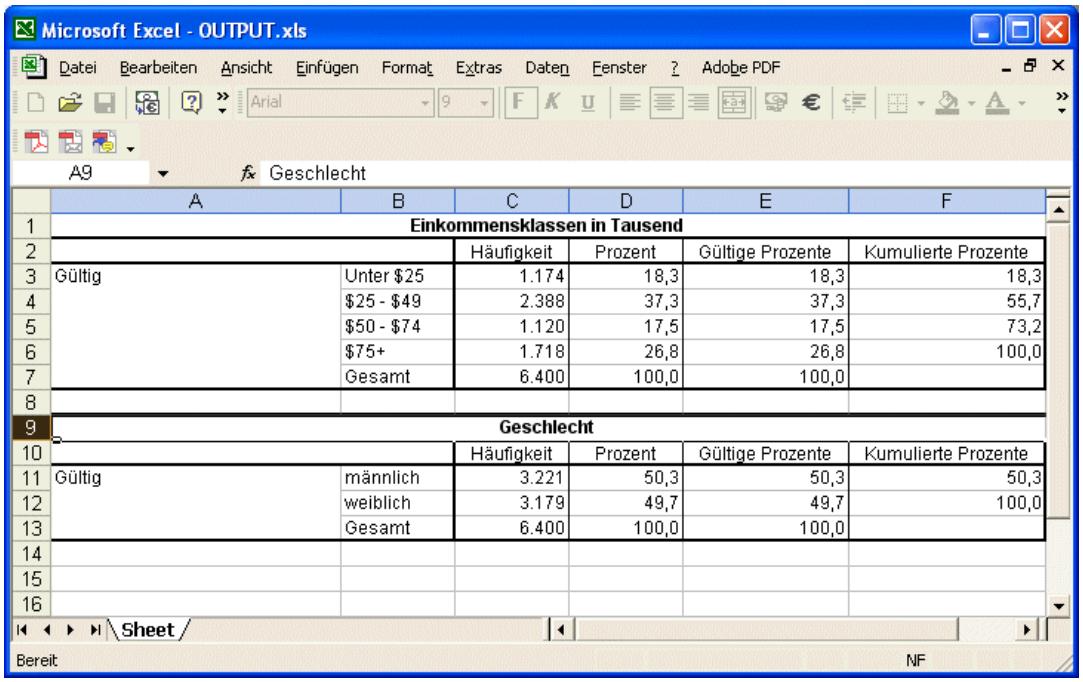
Die für den Export in PowerPoint ausgewählten Diagramme werden in die PowerPoint-Datei eingebettet.

Abbildung 7-32
Diagramme in PowerPoint



Beim Export in eine Excel-Datei werden die Ergebnisse in anderer Form exportiert.

Abbildung 7-33
Output.xls in Excel

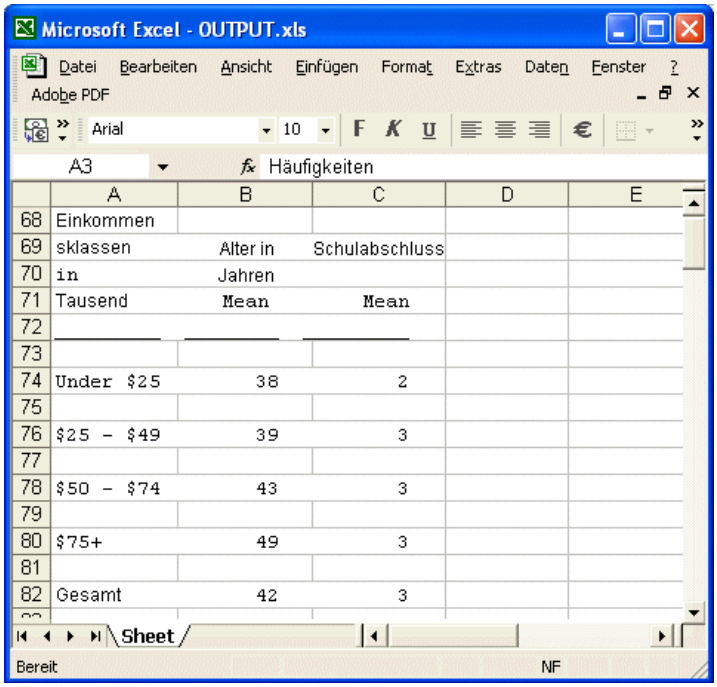


	A	B	C	D	E	F
1	Einkommensklassen in Tausend					
2			Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
3	Gültig	Unter $25	1.174	18,3	18,3	18,3
4		$25 - $49	2.388	37,3	37,3	55,7
5		$50 - $74	1.120	17,5	17,5	73,2
6		$75+	1.718	26,8	26,8	100,0
7		Gesamt	6.400	100,0	100,0	
8						
9	Geschlecht					
10			Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
11	Gültig	männlich	3.221	50,3	50,3	50,3
12		weiblich	3.179	49,7	49,7	100,0
13		Gesamt	6.400	100,0	100,0	
14						
15						
16						

Die Zeilen, Spalten und Zellen von Pivot-Tabellen werden zu Excel-Zeilen, -Spalten und -Zellen.

Jede Zeile in der Textausgabe entspricht einer Zeile in der Excel-Datei, wobei der gesamte Inhalt der Zeile in einer einzelnen Zelle enthalten ist. Diagramme werden überhaupt nicht exportiert.

Abbildung 7-34
Textausgabe in Excel



Exportieren von Ergebnissen als PDF

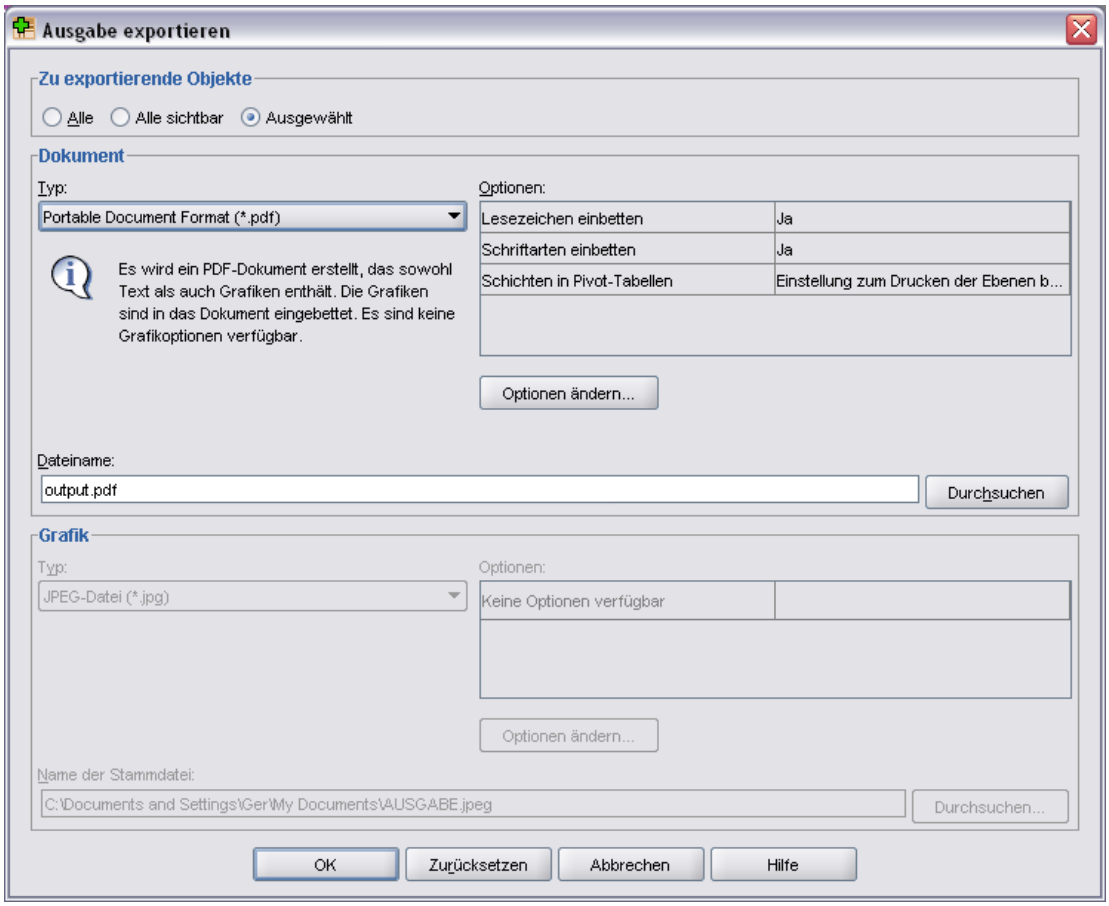
Sir können alle Elemente oder ausgewählte Elemente im Viewer in eine PDF-Datei (Portable Document Format) exportieren.

▶ Wählen Sie in dem Viewer-Fenster, das die Ergebnisse enthält, die Sie als PDF exportieren möchten, folgende Optionen aus den Menüs aus:

Datei
  Exportieren...

▶ Wählen Sie im Dialogfeld “Ausgabe exportieren” aus der Dropdown-Liste “Exportformat - Dateityp” die Option Portable Document Format aus.

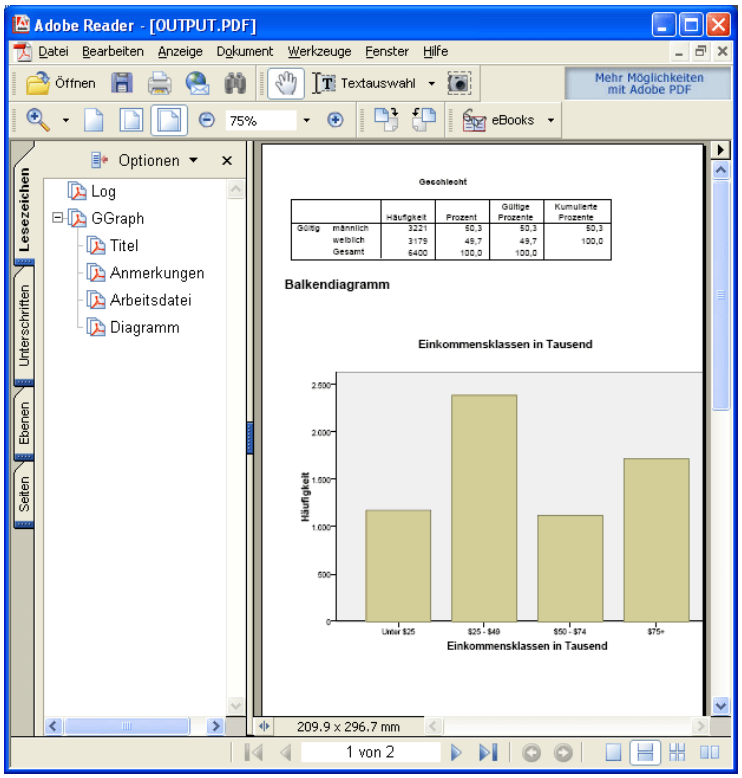
Abbildung 7-35
Dialogfeld “Ausgabe exportieren”



- Das Gliederungsfenster des Viewer-Dokuments wird in der PDF-Datei in Lesezeichen konvertiert, um die Navigation zu erleichtern.
- In PDF-Dokumenten werden Seitengröße, Ausrichtung, Ränder Inhalt und Anzeige von Kopf- und Fußzeilen sowie die Größe des gedruckten Diagramms über die Optionen für die Seiteneinrichtung (Menü “Datei”, “Seite einrichten” im Viewer-Fenster) gesteuert.
- Die Auflösung (DPI) des PDF-Dokuments ist die aktuelle Auflösungseinstellung für den Standarddrucker bzw. den aktuell ausgewählten Drucker (kann über “Seite einrichten” geändert werden). Die maximale Auflösung beträgt 1200 DPI. Wenn eine höhere Druckerauflösung eingestellt ist, wird für das PDF-Dokument eine Auflösung von 1200

DPI verwendet. *Anmerkung*: Dokumente mit höherer Auflösung können beim Drucken auf Druckern mit niedrigerer Auflösung zu schlechten Ergebnissen führen.

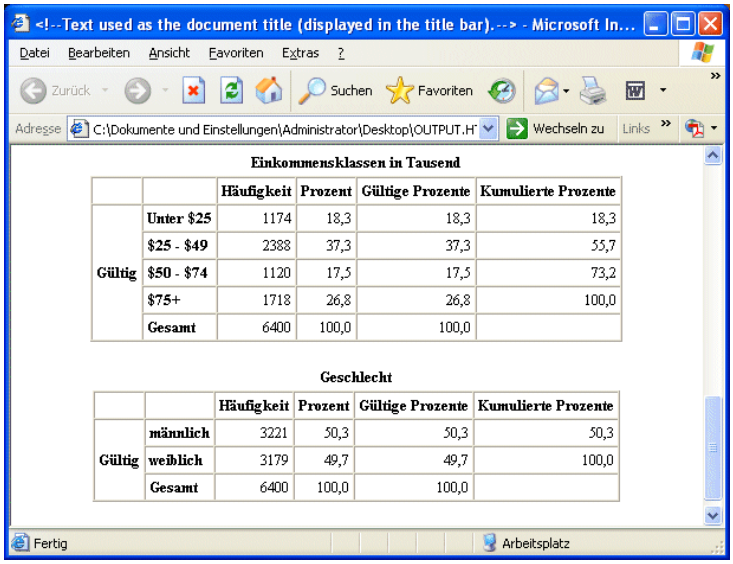
Abbildung 7-36
PDF-Datei mit Lesezeichen



Exportieren von Ergebnissen als HTML

Sie können Ergebnisse auch als HTML (Hypertext Markup Language) exportieren. Beim Speichern als HTML werden alle nichtgrafischen Ausgaben in eine einzelne HTML-Datei exportiert.

Abbildung 7-37
Output.htm im Web-Browser



Einkommensklassen in Tausend

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	Unter $25	1174	18,3	18,3	18,3
	$25 - $49	2388	37,3	37,3	55,7
	$50 - $74	1120	17,5	17,5	73,2
	$75+	1718	26,8	26,8	100,0
	Gesamt	6400	100,0	100,0	

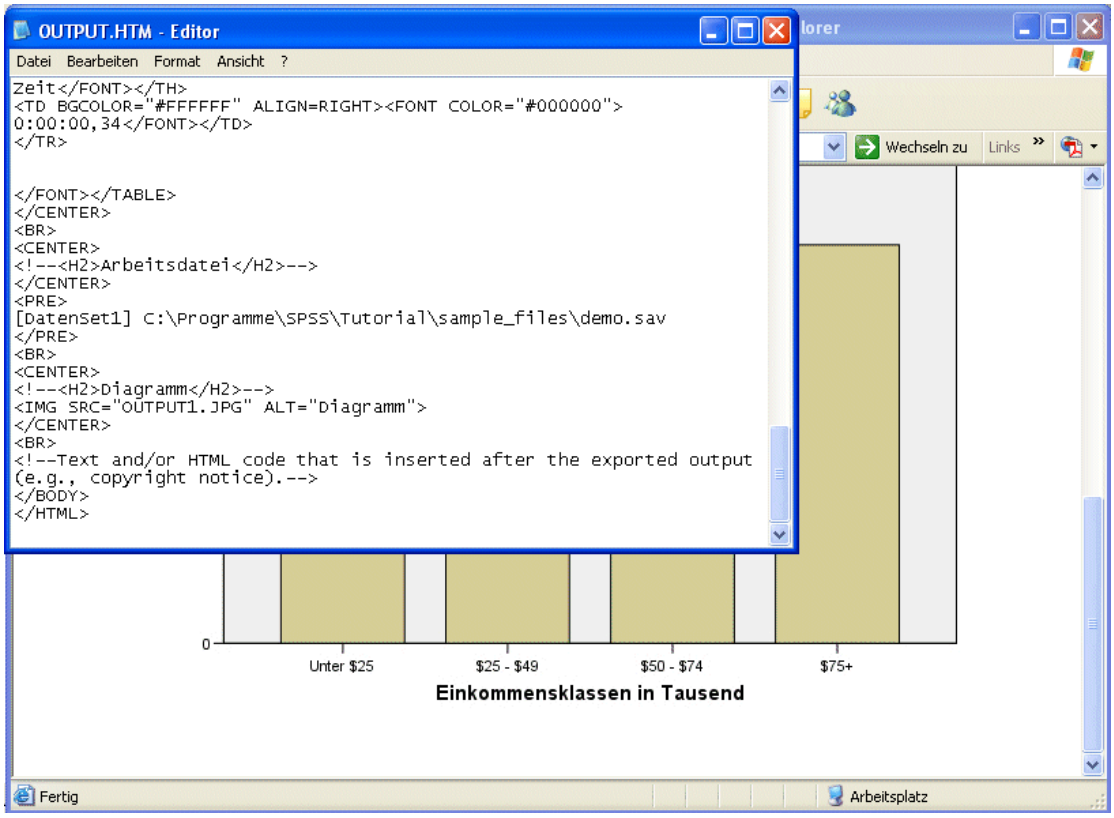
Geschlecht

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	männlich	3221	50,3	50,3	50,3
	weiblich	3179	49,7	49,7	100,0
	Gesamt	6400	100,0	100,0	

Beim Exportieren als HTML können auch Diagramme exportiert werden, jedoch nicht in eine einzelne Datei.

Jedes Diagramm wird als separate Datei in einem Format gespeichert, das Sie angeben. In die HTML-Datei werden Verweise auf diese Grafikdateien eingefügt. Es ist außerdem eine Option vorhanden, mit der alle (oder ausgewählte) Diagramme in separate Grafikdateien exportiert werden können.

Abbildung 7-38
Verweise auf Grafiken in HTML-Ausgaben



Kapitel

8

Arbeiten mit Syntax

Sie können viele übliche Aufgaben mithilfe der leistungsstarken Befehlssprache speichern und automatisieren. Sie bietet außerdem einige Funktionen, die nicht über die Menüs und Dialogfelder zur Verfügung stehen. Auf die meisten SPSS-Befehle können Sie über die Menüs und Dialogfelder zugreifen. Einige Befehle und Optionen sind aber nur in der SPSS-Befehlssprache verfügbar. Mit der Befehlssprache verfügen Sie außerdem über die Möglichkeit, Jobs in einer Syntaxdatei zu speichern. Sie können eine Analyse dann zu einem späteren Zeitpunkt wiederholen.

Eine Befehlssyntaxdatei ist einfach eine Textdatei, die SPSS-Syntaxbefehle enthält. Sie können ein Syntax-Fenster öffnen und direkt Befehle eingeben. Oftmals ist es aber einfacher, wenn Sie sich der Dialogfelder bedienen.

Für die Beispiele dieses Kapitels wird die Datendatei *demo.sav* verwendet. Für weitere Informationen siehe Beispieldateien in Anhang A auf S. 163.

Anmerkung: Die Befehlssyntax ist in der Studentenversion nicht verfügbar.

Übernehmen von Befehlssyntax

Am einfachsten kann Syntax über die Schaltfläche “Einfügen” erstellt werden, die in den meisten Dialogfeldern vorhanden ist.

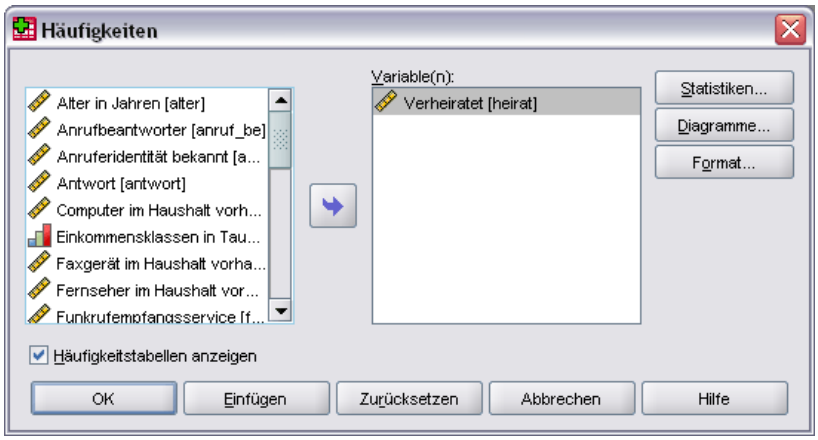
▶ Öffnen Sie *demo.sav* für dieses Beispiel.

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Analysieren
  Deskriptive Statistiken
    Häufigkeiten...

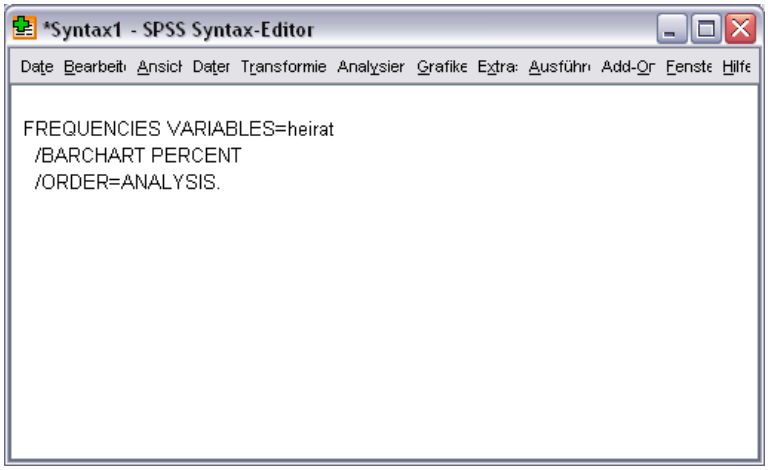
Das Dialogfeld “Häufigkeiten” wird geöffnet.

Abbildung 8-1
Dialogfeld “Häufigkeiten”



- Wählen Sie den Eintrag *Verheiratet [heirat]* aus und verschieben Sie ihn in die Liste “Variable(n)”.
- Klicken Sie auf Diagramme.
- Wählen Sie im Dialogfeld “Grafiken” die Option Balkendiagramme aus.
- Wählen Sie im Gruppenfeld “Diagrammwerte” die Option Prozente aus.
- Klicken Sie auf Weiter.
- Klicken Sie auf Einfügen, um die Syntax, die anhand der Angaben in den Dialogfeldern erstellt wurde, in den Syntax-Editor zu kopieren.

Abbildung 8-2
Syntax für Häufigkeiten



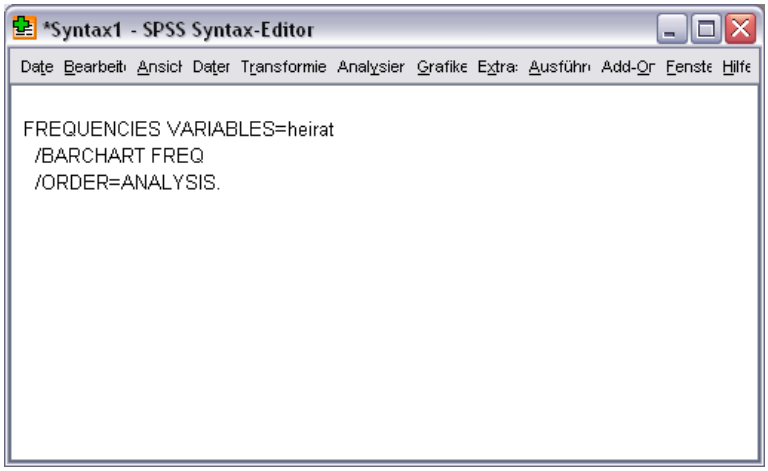
▶ Wenn Sie die gerade angezeigte Syntax ausführen möchten, wählen Sie die folgenden Menübefehle aus:

Ausführen
  Aktuellen Befehl

Bearbeiten von Befehlssyntax

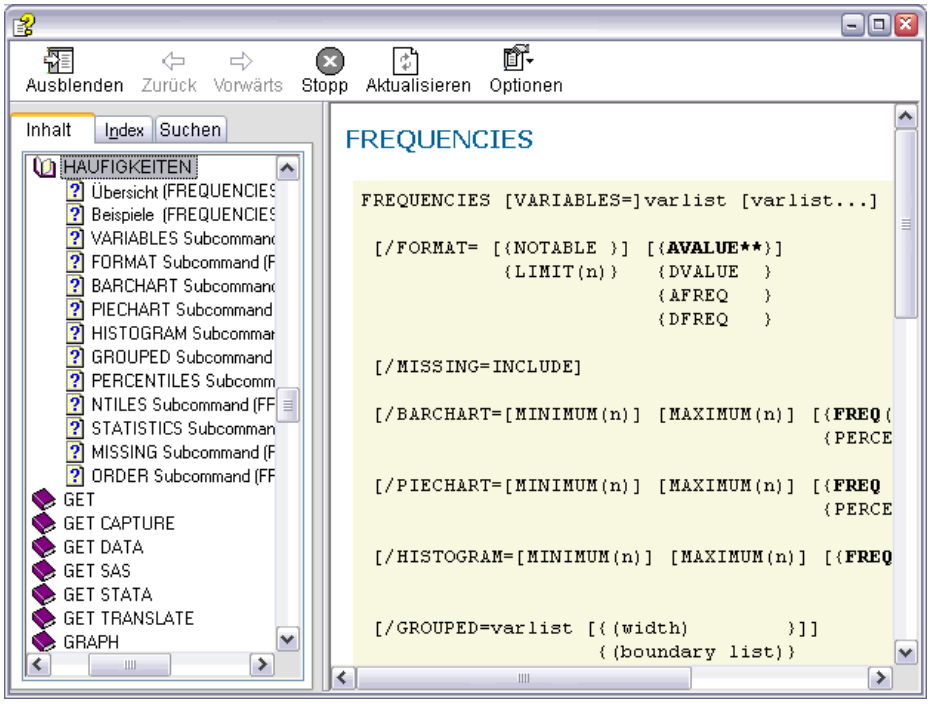
Die Befehlssyntax kann im Syntaxfenster bearbeitet werden. Sie können den Unterbefehl `/BARCHART` beispielsweise so bearbeiten, dass Häufigkeiten anstelle von Prozentwerten angezeigt werden. (Ein Unterbefehl wird durch einen Schrägstrich gekennzeichnet.)

Abbildung 8-3
Geänderte Syntax



Um alle Unterbefehle und Stichwörter zu ermitteln, die für den aktuellen Befehl verfügbar sind, drücken Sie die Taste F1. Damit gelangen Sie direkt zur Befehlssyntaxreferenz für den aktuellen Befehl.

Abbildung 8-4
Hilfe zur Syntax für den Befehl FREQUENCIES



Öffnen und Ausführen einer Syntaxdatei

▶ Zum Öffnen einer gespeicherten Syntaxdatei wählen Sie die folgenden Befehle aus den Menüs aus:

Datei
  Öffnen
    Syntax...

Ein Standarddialogfeld zum Öffnen von Dateien wird angezeigt.

▶ Wählen Sie eine Syntaxdatei aus. Wenn keine Syntaxdateien angezeigt werden, müssen Sie sicherstellen, dass Syntax (*.sps) als anzuzeigender Dateityp ausgewählt ist.

▶ Klicken Sie auf Öffnen.

▶ Verwenden Sie das Menü “Ausführen” im Syntax-Editor, um die Befehle auszuführen.

Wenn die Befehle auf eine bestimmte Datendatei angewendet werden sollen, muss die betreffende Datendatei vor der Ausführung der Befehle geöffnet werden oder Sie müssen ein Befehl zum Öffnen der Datendatei in die Syntax einbeziehen. Diesen Befehlstyp können Sie aus den Dialogfeldern übernehmen, mit denen Datendateien geöffnet werden.

Kapitel

9

Ändern von Datenwerten

Daten sind nicht immer von Anfang an optimal für Analysen oder Berichte strukturiert. Angenommen, Sie möchten beispielsweise folgende Vorgänge durchführen:

- Erstellen einer kategorialen Variablen aus einer metrischen Variablen.
- Kombinieren mehrerer Ergebniskategorien in einer einzelnen Kategorie.
- Erstellen einer neuen Variablen, welche die berechnete Differenz zwischen zwei vorhandenen Variablen darstellt.
- Berechnen des Zeitabstands zwischen zwei Daten.

In diesem Kapitel wird die Datendatei *demo.sav* verwendet. Für weitere Informationen siehe Beispieldateien in Anhang A auf S. 163.

Erstellen einer kategorialen Variablen aus einer metrischen Variablen

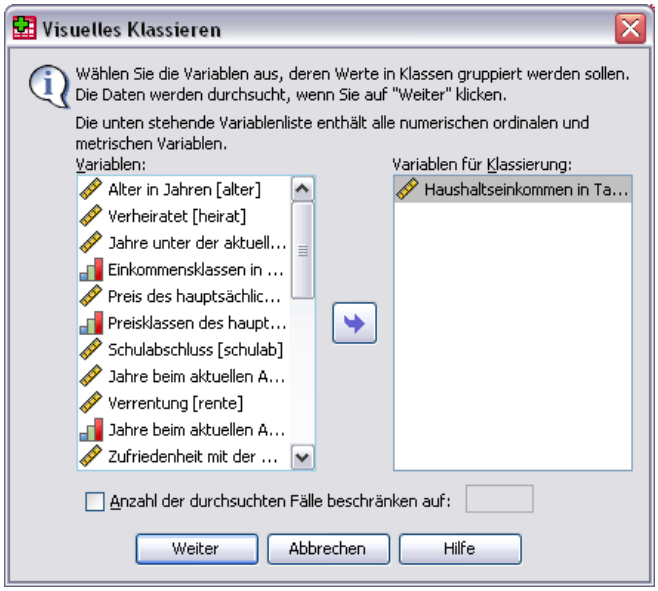
Mehrere kategoriale Variablen in der Datendatei *demo.sav* sind eigentlich von metrischen Variablen in dieser Datendatei abgeleitet. Die Variable *eink_kl* ergibt sich einfach dadurch, dass *Einkommen* in vier Kategorien unterteilt wird. Bei dieser kategorialen Variablen stehen die ganzzahligen Werte 1–4 für folgende Einkommensklassen (in Tausend): unter $25, $25–$49, $50–$74 und $75 oder höher.

So erstellen Sie die kategoriale Variable *eink_kl*:

▶ Wählen Sie die folgenden Befehle aus den Menüs im Fenster “Daten-Editor” aus:

Transformieren
  Visuelles Klassieren...

Abbildung 9-1
Visuelles Klassieren – Anfangsdialogfeld

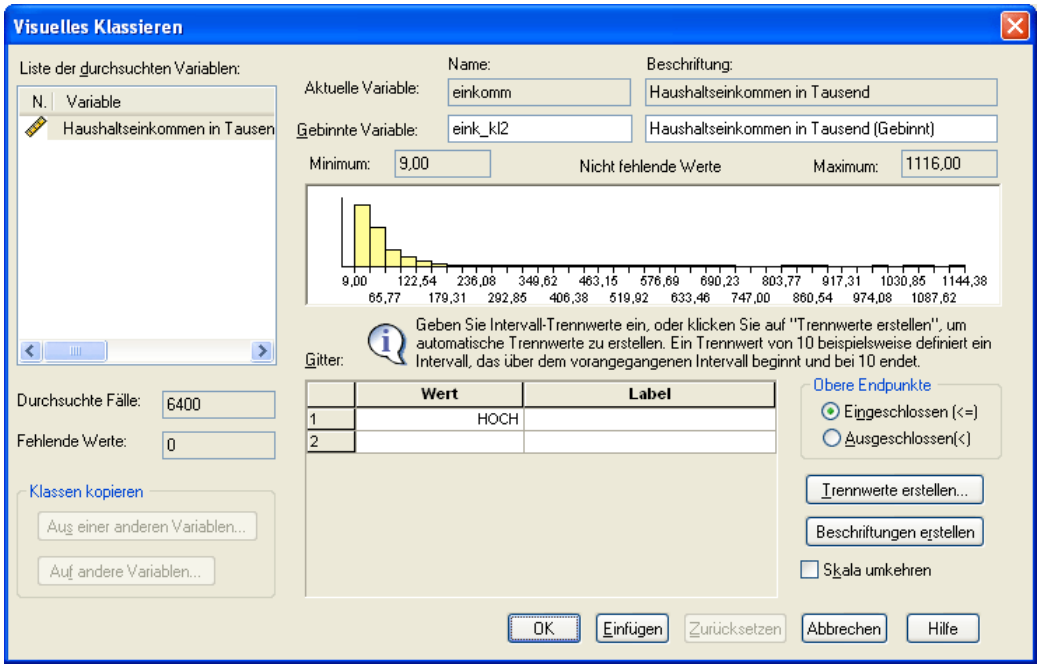


Im ersten Dialogfeld von “Visuelles Klassieren” wählen Sie die metrischen und/oder ordinalen Variablen aus, für die neue, klassierte Variablen erstellt werden sollen. **Klassieren** bedeutet, dass zwei oder mehrere nebeneinanderliegende Werte zusammengefasst und in dieselbe Kategorie eingeordnet werden.

Da die Funktion “Visuelles Klassieren” die tatsächlichen Werte in der Datendatei verwendet, um Ihnen ein sinnvolles Klassieren zu erleichtern, muss sie die Datendatei zuerst lesen. Da dies einige Zeit in Anspruch nehmen kann, wenn Ihre Datendatei eine große Anzahl von Fällen enthält, können Sie in diesem Anfangsdialogfeld auch die Anzahl der zu lesenden (durchsuchenden) Fälle begrenzen. Bei der verwendeten Beispieldatendatei ist dies nicht erforderlich. Obwohl sie mehr als 6.000 Fälle umfaßt, dauert das Durchsuchen bei dieser Anzahl von Fällen nicht besonders lang.

▶ Verschieben Sie den Eintrag *Haushaltseinkommen in Tausend (einkomm)* durch Ziehen und Ablegen aus der Liste “Variablen” in die Liste “Variablen für Klassierung” und klicken Sie anschließend auf Weiter.

Abbildung 9-2
Visuelles Klassieren – Hauptdialogfeld



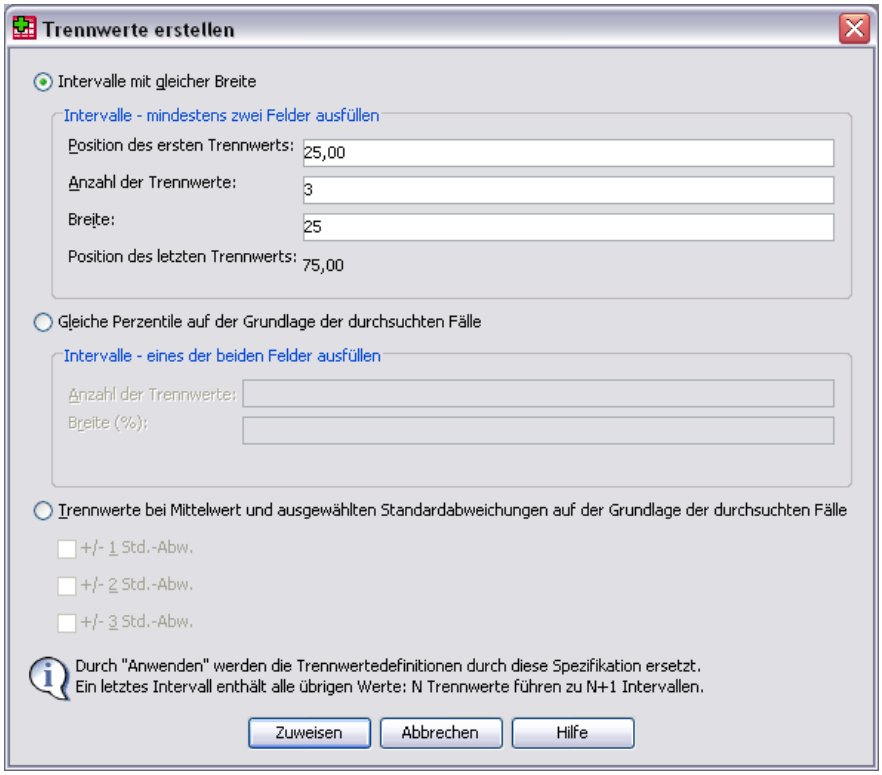
- Wählen Sie im Hauptdialogfeld von “Visuelles Klassieren” in der Liste der durchsuchten Variablen den Eintrag *Haushaltseinkommen in Tausend [einkomm]* aus.

  Ein Histogramm zeigt die Verteilung der ausgewählten Variablen an, die in diesem Fall sehr schief ist.

- Geben Sie eink_kl2 als Namen der neuen gebinnten Variablen und Einkommensklassen (in Tausend) als Variablenlabel an.

- Klicken Sie auf Trennwerte erstellen.

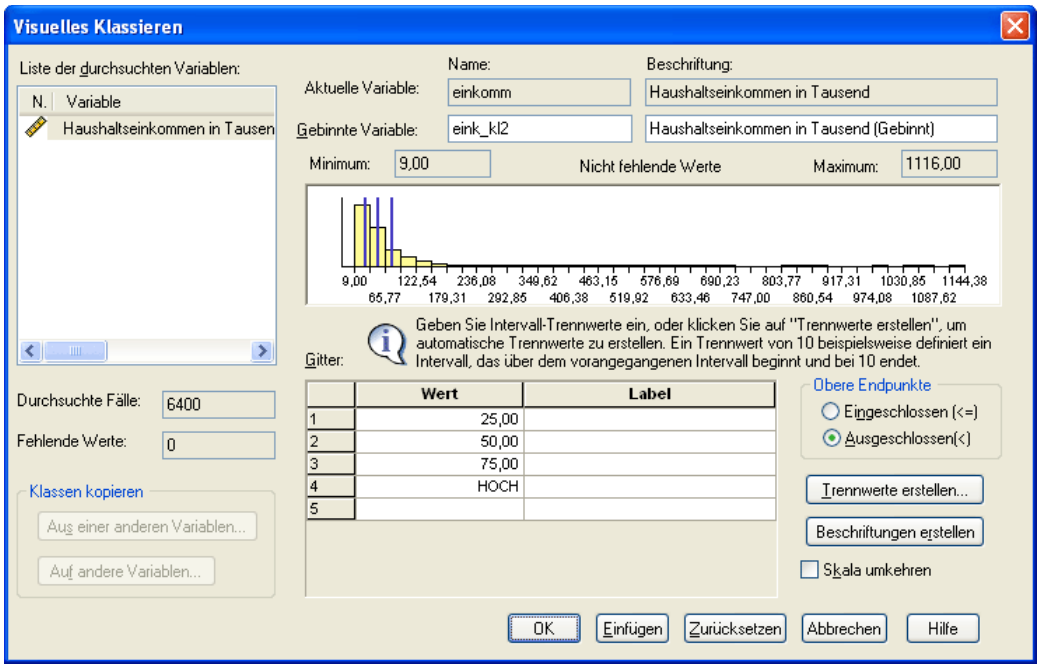
Abbildung 9-3
Visuelles Klassieren – Dialogfeld "Trennwerte erstellen"



- Wählen Sie die Option Intervalle mit gleicher Breite.
- Geben Sie 25 als ersten Trennwert, 3 als Anzahl der Trennwerte und 25 als Breite ein.

  Die Anzahl der gebinnten Kategorien ist um den Wert 1 größer als die Anzahl der Trennwerte. In diesem Beispiel weist die neue, gebinnte Variable also vier Kategorien auf, wobei die ersten drei Kategorien Bereiche von 25 (tausend) umfassen und die letzte Kategorie alle Werte über dem höchsten Trennwert 75 (tausend) enthält.
- Klicken Sie auf Zuweisen.

Abbildung 9-4
Visuelles Klassieren – Hauptdialogfeld mit festgelegten Trennwerten



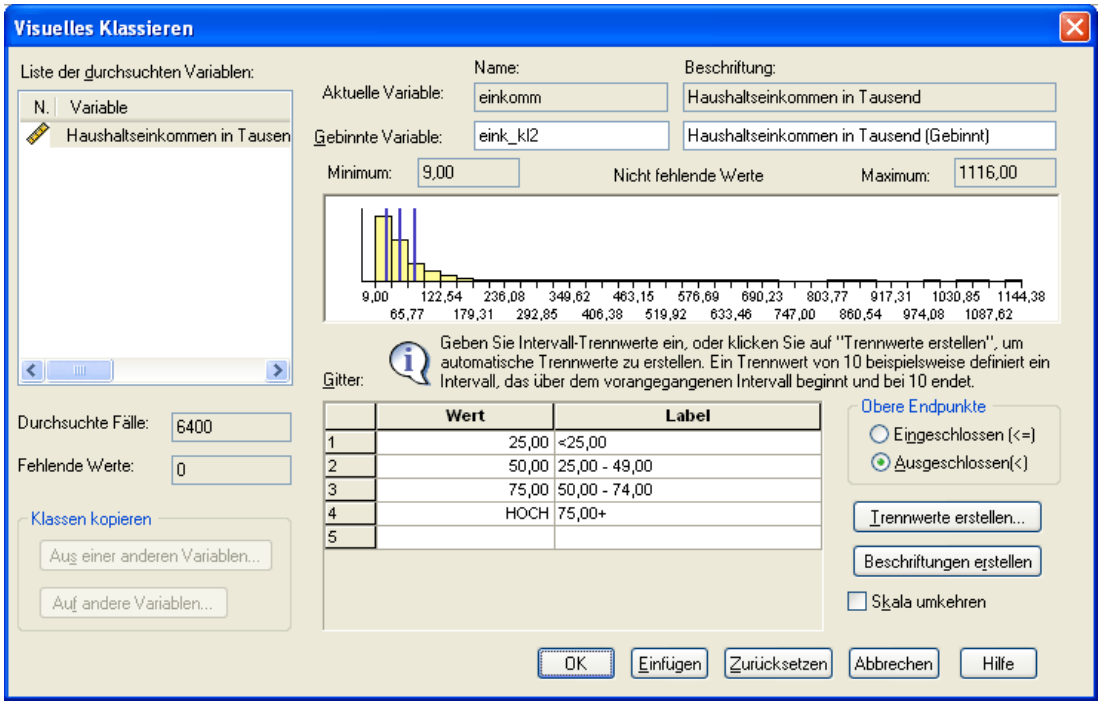
Die nun im Gitter angezeigten Werte stellen die festgelegten Trennwerte dar, die oberen Endpunkte der einzelnen Kategorien. Die Position der Trennwerte wird auch durch vertikale Linien im Histogramm angezeigt.

Standardmäßig sind diese Trennwerte in den entsprechenden Kategorien enthalten. Der erste Wert (25) bedeutet beispielsweise, dass alle Werte kleiner oder gleich 25 eingeschlossen werden. In diesem Beispiel möchten wir jedoch Kategorien erstellen, die Werte unter 25, 25–49, 50–74 und 75 oder höher enthalten.

▶ Wählen Sie in der Gruppe “Obere Endpunkte” den Eintrag Ausgeschlossen (<).

▶ Klicken Sie anschließend auf Beschriftungen erstellen.

Abbildung 9-5
Automatisch erstellte Wertelabels



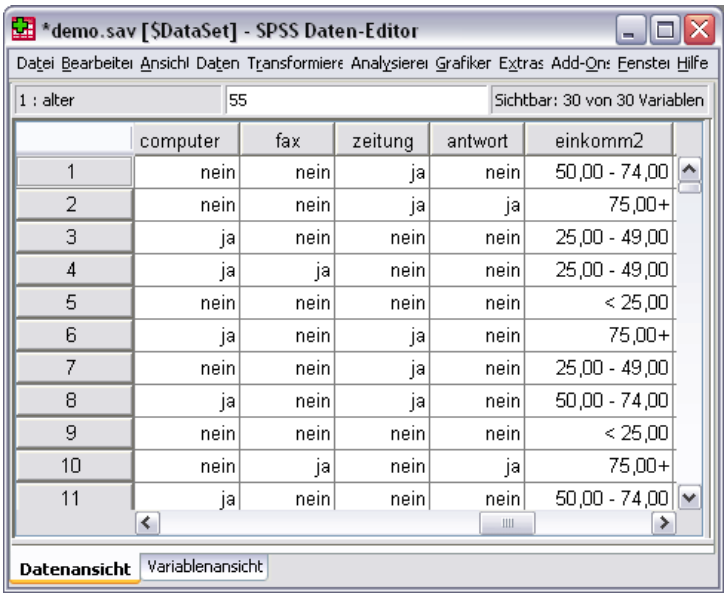
Dadurch werden automatisch beschreibende Wertelabels für die einzelnen Kategorien erstellt. Da es sich bei den Werten, die der neuen gebinnten Variablen zugewiesen werden, einfach um aufeinander folgende Ganzzahlen, beginnend bei 1, handelt, können die Wertelabels sehr hilfreich sein.

Außerdem können Sie manuell Trennwerte und Labels (Beschriftungen) im Gitter eingeben oder ändern, die Position von Trennwerten durch Ziehen und Ablegen der Trennwertlinien im Histogramm ändern und Trennwerte löschen, indem Sie die Trennwertlinien vom Histogramm weg ziehen.

▶ Klicken Sie auf OK, um die neue, gebinnte Variable zu erstellen.

Die neue Variable wird im Daten-Editor angezeigt. Da die Variable am Ende der Liste hinzugefügt wird, wird sie in der Ansicht “Daten” in der äußersten rechten Spalte und in der Ansicht “Variable” in der letzten Zeile angezeigt.

Abbildung 9-6
Neue Variable im Daten-Editor



	computer	fax	zeitung	antwort	einkomm2
1	nein	nein	ja	nein	50,00 - 74,00
2	nein	nein	ja	ja	75,00+
3	ja	nein	nein	nein	25,00 - 49,00
4	ja	ja	nein	nein	25,00 - 49,00
5	nein	nein	nein	nein	< 25,00
6	ja	nein	ja	nein	75,00+
7	nein	nein	ja	nein	25,00 - 49,00
8	ja	nein	ja	nein	50,00 - 74,00
9	nein	nein	nein	nein	< 25,00
10	nein	ja	nein	ja	75,00+
11	ja	nein	nein	nein	50,00 - 74,00

Berechnen von neuen Variablen

Sie können eine Vielfalt von mathematischen Funktionen einsetzen, um neue Variablen auf der Grundlage von komplizierten Gleichungen zu berechnen. In diesem Beispiel wird jedoch lediglich eine neue Variable berechnet, die die Differenz zwischen den Werten von zwei vorhandenen Variablen darstellt.

Die Datendatei *demo.sav* enthält eine Variable für das aktuelle Alter der Befragten und eine Variable für die Anzahl der Jahre beim aktuellen Arbeitgeber. Sie enthält dagegen keine Variable dafür, welches Alter die Befragten hatten, als sie diese Arbeitsstelle antraten. Sie können eine neue Variable erstellen, welche die berechnete Differenz zwischen dem aktuellen Alter und der Anzahl der Jahre beim aktuellen Arbeitgeber darstellt. Diese Differenz ist ungefähr das Alter, in dem die Befragten diese Arbeitsstelle antraten.

▶ Wählen Sie die folgenden Befehle aus den Menüs im Fenster “Daten-Editor” aus:

Transformieren
  Variable berechnen...

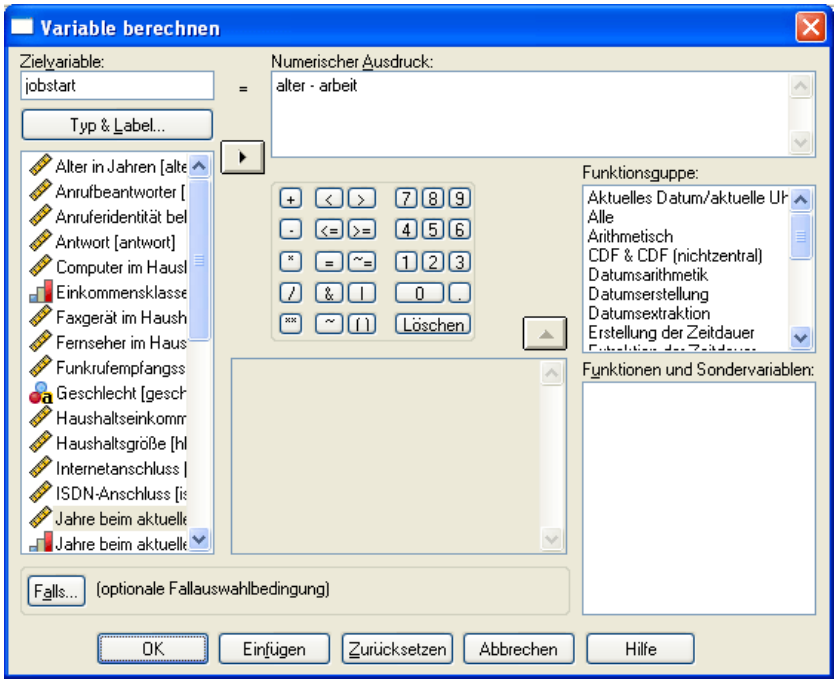
▶ Geben Sie berufanf als “Zielvariable” ein.

▶ Wählen Sie in der Quellvariablenliste *Alter in Jahren [alter]* aus und klicken Sie auf die Pfeilschaltfläche, um die Variable in das Textfeld “Numerischer Ausdruck” zu kopieren.

▶ Klicken Sie auf der Rechentastatur des Dialogfelds auf die Schaltfläche mit dem Minuszeichen (–) oder drücken Sie die MINUSTASTE auf der Tastatur.

▶ Wählen Sie *Jahre beim aktuellen Arbeitgeber [arbeit]* aus und klicken Sie auf die Pfeilschaltfläche, um die Variable in die Liste für den Ausdruck zu kopieren.

Abbildung 9-7
Dialogfeld "Variable berechnen"

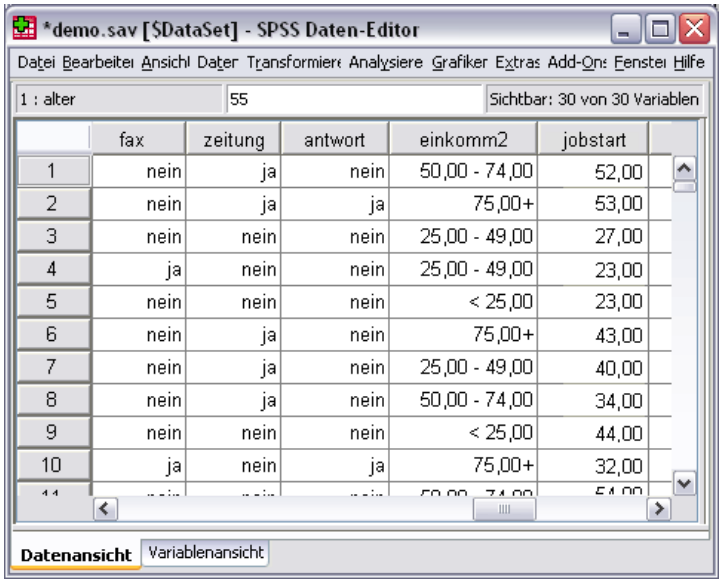


Anmerkung: Achten Sie darauf, die richtige Variable für den Arbeitsplatz auszuwählen. Es gibt auch eine umkodierte kategoriale Version der Variablen. Dies ist *nicht* die Variable, die Sie benötigen. Der numerische Ausdruck sollte *alte–arbeit* lauten und nicht *alter–arbeit_kl*.

▶ Klicken Sie auf OK, um die neue Variable zu berechnen.

Die neue Variable wird im Daten-Editor angezeigt. Da die Variable am Ende der Liste hinzugefügt wird, wird sie in der Ansicht “Daten” in der äußersten rechten Spalte und in der Ansicht “Variable” in der letzten Zeile angezeigt.

Abbildung 9-8
Neue Variable im Daten-Editor



	fax	zeitung	antwort	einkomm2	jobstart
1	nein	ja	nein	50,00 - 74,00	52,00
2	nein	ja	ja	75,00+	53,00
3	nein	nein	nein	25,00 - 49,00	27,00
4	ja	nein	nein	25,00 - 49,00	23,00
5	nein	nein	nein	< 25,00	23,00
6	nein	ja	nein	75,00+	43,00
7	nein	ja	nein	25,00 - 49,00	40,00
8	nein	ja	nein	50,00 - 74,00	34,00
9	nein	nein	nein	< 25,00	44,00
10	ja	nein	ja	75,00+	32,00

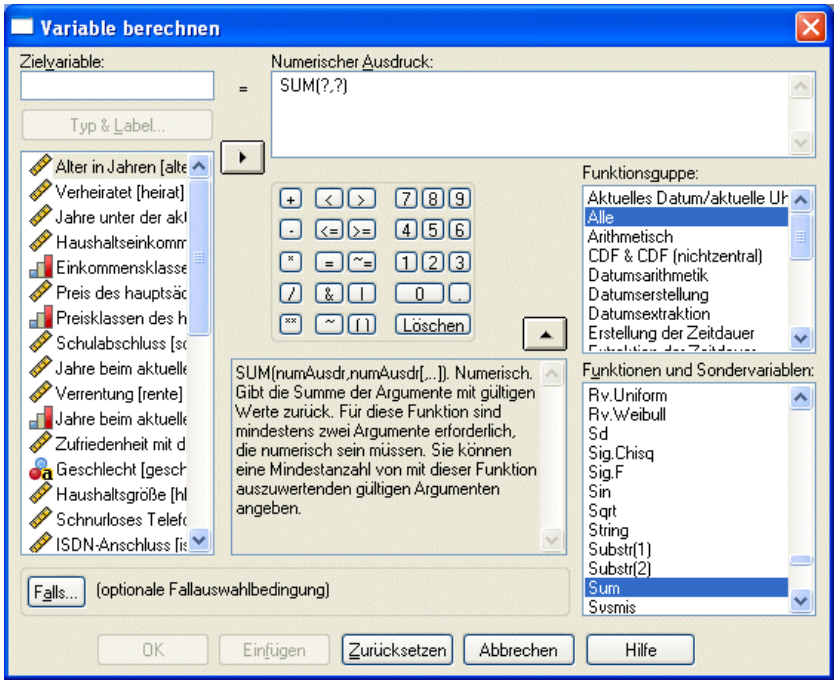
Verwenden von Funktionen in Ausdrücken

Sie können außerdem vordefinierte Funktionen in Ausdrücken verwenden. Es stehen mehr als 70 integrierte Funktionen zur Verfügung, darunter folgende:

- Arithmetische Funktionen
- Statistische Funktionen
- Verteilungsfunktionen
- Logische Funktionen
- Funktionen zur Aggregation und Extraktion von Datum und Uhrzeit
- Funktionen für fehlende Werte

- Fallübergreifende Funktionen
- String-Funktionen

Abbildung 9-9
Dialogfeld "Variable berechnen"; Funktionsgruppierung wird angezeigt.



Funktionen werden in logisch getrennte Gruppen eingeteilt, beispielsweise eine Gruppe für arithmetische Operationen und eine Gruppe zur Berechnung statistischer Metriken. Zur Erhöhung der Benutzerfreundlichkeit ist auch eine Anzahl von häufig verwendeten Systemvariablen, beispielsweise *$TIME* (aktuelles Datum und aktuelle Uhrzeit) in den entsprechenden Funktionsgruppen enthalten. Eine kurze Beschreibung der aktuell ausgewählten Funktion (in diesem Fall `SUM`) oder Systemvariablen wird in einem speziellen Bereich des Dialogfelds "Variable berechnen" angezeigt.

Einfügen einer Funktion in einen Ausdruck

So fügen Sie eine Funktion in einen Ausdruck ein:

▶ Setzen Sie den Mauszeiger an die Stelle im Ausdruck, an der die Funktion eingefügt werden soll.

▶ Wählen Sie aus der Liste "Funktionsgruppe" die geeignete Gruppe aus. Die Gruppe mit der Beschriftung Alle bietet eine Auflistung aller verfügbaren Funktionen und Systemvariablen.

▶ Doppelklicken Sie in der Liste "Funktionen und Sondervariablen" auf die Funktion. (Sie können auch die Funktion auswählen und auf den Pfeil neben der Liste "Funktionsgruppe" klicken.)

Die Funktion wird in den Ausdruck eingefügt. Wenn Sie einen Teil des Ausdrucks markieren und anschließend die Funktion einfügen, wird der markierte Teil des Ausdrucks als erstes Argument der Funktion verwendet.

Bearbeiten einer Funktion in einem Ausdruck

Die Funktion ist erst vollständig, nachdem Sie die Argumente eingegeben haben, die in der eingefügten Funktion durch Fragezeichen dargestellt werden. Die Anzahl der Fragezeichen gibt die Mindestanzahl der Argumente an, die zur Vervollständigung der Funktion erforderlich sind.

- ▶ Markieren Sie das bzw. die Fragezeichen in der eingefügten Funktion.
- ▶ Geben Sie die Argumente ein. Wenn es sich bei den Argumenten um Variablennamen handelt, können Sie sie aus der Variablenliste einfügen.

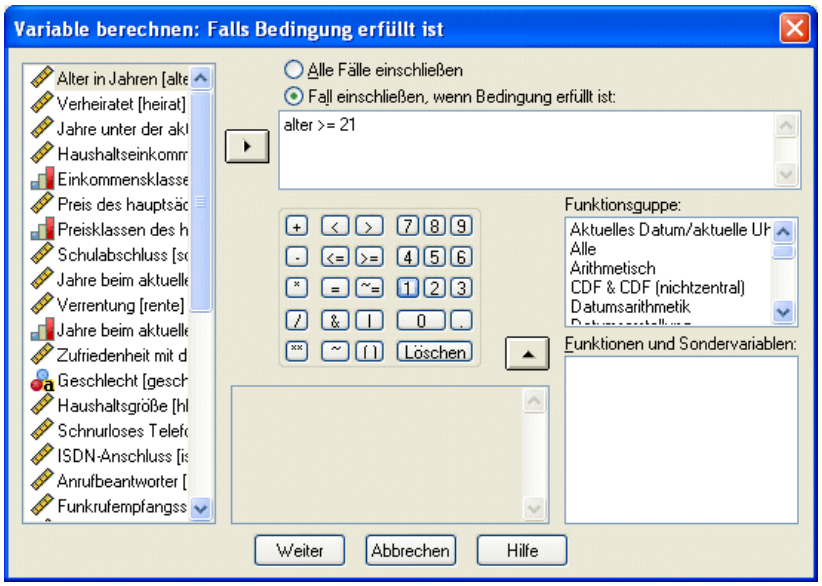
Verwendung bedingter Ausdrücke

Mit bedingten Ausdrücken, auch logische Ausdrücke genannt, können Sie Transformationen auf ausgewählte Teilmengen von Fällen anwenden. Ein bedingter Ausdruck gibt für jeden Fall den Wert “Wahr”, “Falsch” oder “Fehlend” zurück. Wenn das Ergebnis eines bedingten Ausdrucks “Wahr” lautet, wird die Transformation für den Fall durchgeführt. Wenn als Ergebnis der Wert “Falsch” oder “Fehlend” vorliegt, wird die Transformation nicht auf den Fall angewendet.

So geben Sie einen bedingten Ausdruck ein:

- ▶ Klicken Sie im Dialogfeld “Variable berechnen” auf Falls. Dadurch wird das Dialogfeld “Falls Bedingung erfüllt ist” aufgerufen.

Abbildung 9-10
Dialogfeld “Variable berechnen: Falls Bedingung erfüllt ist”



- ▶ Wählen Sie Fall einschließen, wenn Bedingung erfüllt ist aus.
- ▶ Geben Sie den bedingten Ausdruck ein.

Die meisten bedingten Ausdrücke enthalten mindestens einen relationalen Operator. Beispiel:

```
alter>=21
```

oder

```
einkomm*3<100
```

Im ersten Beispiel werden nur Fälle ausgewählt, in denen der Wert für *Alter [alter]* größer oder gleich 21 ist. Im zweiten Beispiel muss der Wert für *Haushaltseinkommen in Tausend [einkomm]* multipliziert mit 3 unter 100 liegen, damit ein Fall ausgewählt wird.

Es besteht außerdem die Möglichkeit, mehr als zwei bedingte Ausdrücke über logische Operatoren zu verbinden. Beispiel:

```
alter>=21 | schulab>=4
```

oder

```
einkomm*3<100 & schulab=5
```

Im ersten Beispiel werden Fälle ausgewählt, die die Bedingung für *Alter [alter]* oder für *Schulabschluss [schulab]* erfüllen. Im zweiten Beispiel müssen sowohl die Bedingungen für *Haushaltseinkommen in Tausend [einkomm]* als auch die Bedingungen für *Schulabschluss [schulab]* erfüllt sein, damit ein Fall ausgewählt wird.

Arbeiten mit Datumsangaben und Uhrzeiten

Mit dem Assistenten für Datum und Uhrzeit läßt sich eine Reihe von Aufgaben, die häufig mit Datumsangaben und Uhrzeiten durchgeführt werden, problemlos bewältigen. Mit diesem Assistenten können Sie folgende Aufgaben ausführen:

- Erstellen einer Datums-/Zeitvariablen aus einer String-Variablen, die ein Datum oder eine Uhrzeit enthält.
- Erstellen einer Datums-/Zeitvariablen durch Zusammenführen von Variablen, die verschiedene Teile des Datums bzw. der Uhrzeit enthalten.
- Addieren oder Subtrahieren von Werten zu bzw. von einer Datums-/Zeitvariablen (einschließlich Addition bzw. Subtraktion von zwei Datums-/Zeitvariablen).
- Extrahieren eines Teils einer Datums- oder Zeitvariablen, beispielsweise des Tags im Monat aus einer Datums-/Zeitvariablen mit dem Format mm/tt/jjjj.

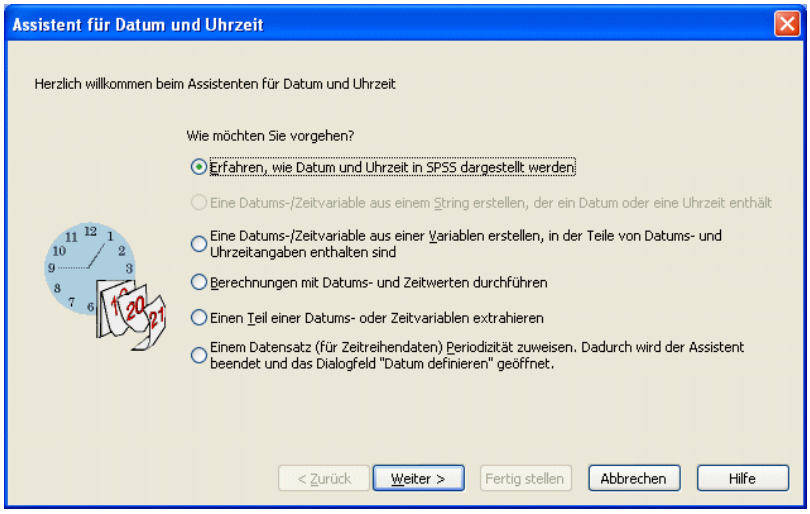
Für die Beispiele in diesem Abschnitt wird die Datendatei *upgrade.sav* verwendet. Für weitere Informationen siehe Beispieldateien in Anhang A auf S. 163.

So verwenden Sie den Assistenten für Datum und Uhrzeit:

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Transformieren
  Assistent für Datum und Uhrzeit...

Abbildung 9-11
Assistent für Datum und Uhrzeit: Einführungsbildschirm



Im Einführungsbildschirm des Assistenten für Datum und Uhrzeit wird eine Reihe allgemeiner Aufgaben angezeigt. Aufgaben, die nicht auf die aktuellen Daten zutreffen, sind deaktiviert. So enthält die Datendatei *upgrade.sav* beispielsweise keine String-Variablen, sodass die Aufgabe zur Erstellung einer Datenvariablen aus einem String deaktiviert ist.

Wenn Sie sich mit Datum und Uhrzeit in SPSS nicht auskennen, können Sie die Option Erfahren, wie Datum und Uhrzeit in SPSS dargestellt werden aktivieren und auf Weiter klicken. Dadurch gelangen Sie auf einen Bildschirm, der einen kurzen Überblick über Datums-/Zeitvariablen und (über die Schaltfläche “Hilfe”) einen Link zu detaillierteren Informationen bietet.

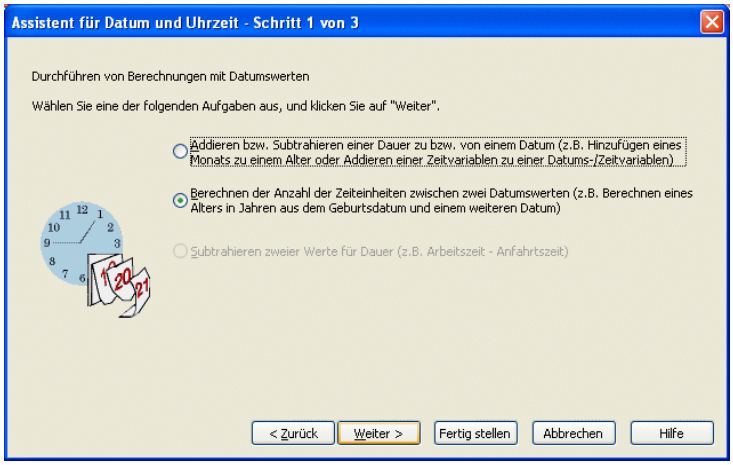
Berechnen des Zeitabstands zwischen zwei Datumsangaben

Zu den häufigsten Aufgaben, die mit Datumsangaben zu tun haben, gehört die Berechnung des Zeitabstands zwischen zwei Daten. Betrachten wir folgendes Beispiel: Ein Softwareunternehmen möchte die Verkäufe von Upgrade-Lizenzen analysieren, indem die Anzahl der Jahre ermittelt wird, die vergangen sind, seitdem die einzelnen Kunden zuletzt ein Upgrade erworben haben. Die Datendatei *upgrade.sav* enthält eine Variable für das Datum, an dem die einzelnen Kunden zuletzt ein Upgrade erwarben, und nicht die Anzahl der Jahre seit diesem Kauf. Eine neue Variable, die die Zeitspanne in Jahren zwischen dem Datum der letzten Aktualisierung und dem Datum der nächsten Produktveröffentlichung angibt, bietet ein Maß für diesen Wert.

So berechnen Sie den Zeitabstand zwischen zwei Daten:

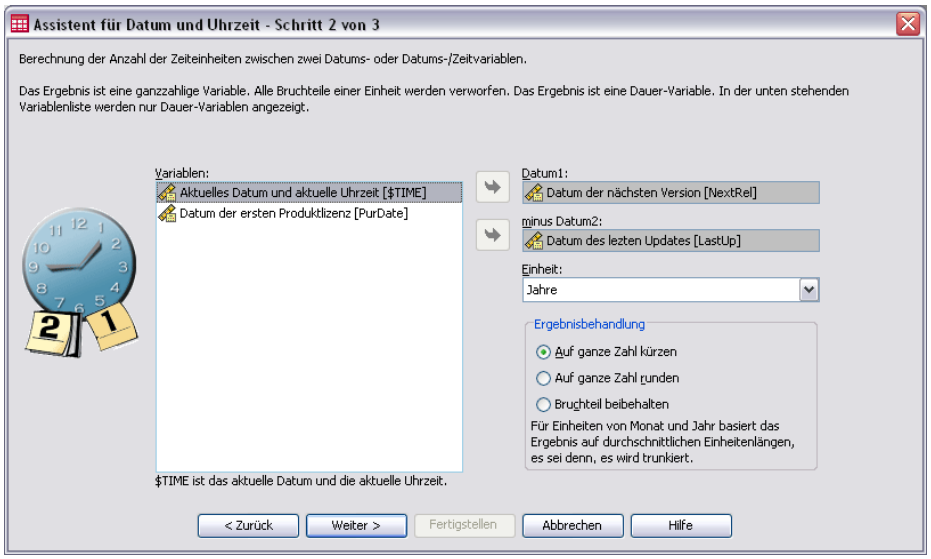
- Wählen Sie im Einführungsbildschirm des Assistenten für Datum und Uhrzeit die Option Berechnungen mit Datums- und Zeitwerten durchführen und klicken Sie anschließend auf Weiter.

Abbildung 9-12
Berechnen des Zeitabstands zwischen zwei Datumsangaben: Schritt 1



▶ Wählen Sie die Option Berechnen der Anzahl der Zeiteinheiten zwischen zwei Datumswerten und klicken Sie auf Weiter.

Abbildung 9-13
Berechnen des Zeitabstands zwischen zwei Datumsangaben: Schritt 2



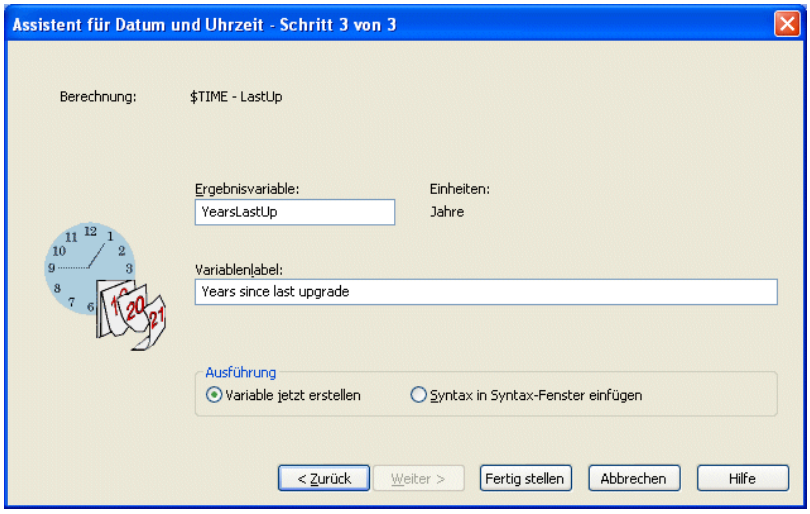
▶ Wählen Sie *Datum der nächsten Version* als “Datum1”.

▶ Wählen Sie *Datum des letzten Upgrades* als “Datum2”.

▶ Wählen Sie Jahre als Einheit und Auf ganze Zahl kürzen als Ergebnisbehandlung. (Dies sind die Standardeinstellungen.)

▶ Klicken Sie auf Weiter.

Abbildung 9-14
Berechnen des Zeitabstands zwischen zwei Datumsangaben: Schritt 3



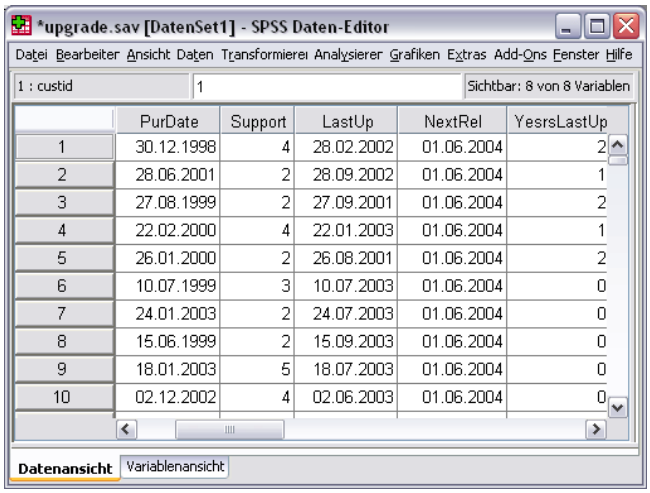
- Geben Sie *YearsLastUp* als Namen der Ergebnisvariablen ein. Ergebnisvariablen können nicht denselben Namen haben wie bestehende Variablen.

- Geben Sie *Years since last upgrade* (Jahre seit dem letzten Upgrade) als Label für die Ergebnisvariable ein. Die Variablenlabels für Ergebnisvariablen sind optional.

- Behalten Sie die Standardauswahl Variable jetzt erstellen bei und klicken Sie auf Fertig stellen, um die neue Variable zu erstellen.

  Bei der neuen Variablen *YearsLastUp*, die im Data-Editor angezeigt wird, handelt es sich um die ganzzahlige Angabe der Jahre zwischen den beiden Datumswerten. Bruchteile von Jahren wurden gekürzt.

Abbildung 9-15
Neue Variable im Daten-Editor



	PurDate	Support	LastUp	NextRel	YesrsLastUp
1	30.12.1998	4	28.02.2002	01.06.2004	2
2	28.06.2001	2	28.09.2002	01.06.2004	1
3	27.08.1999	2	27.09.2001	01.06.2004	2
4	22.02.2000	4	22.01.2003	01.06.2004	1
5	26.01.2000	2	26.08.2001	01.06.2004	2
6	10.07.1999	3	10.07.2003	01.06.2004	0
7	24.01.2003	2	24.07.2003	01.06.2004	0
8	15.06.1999	2	15.09.2003	01.06.2004	0
9	18.01.2003	5	18.07.2003	01.06.2004	0
10	02.12.2002	4	02.06.2003	01.06.2004	0

Hinzufügen einer Dauer zu einem Datum

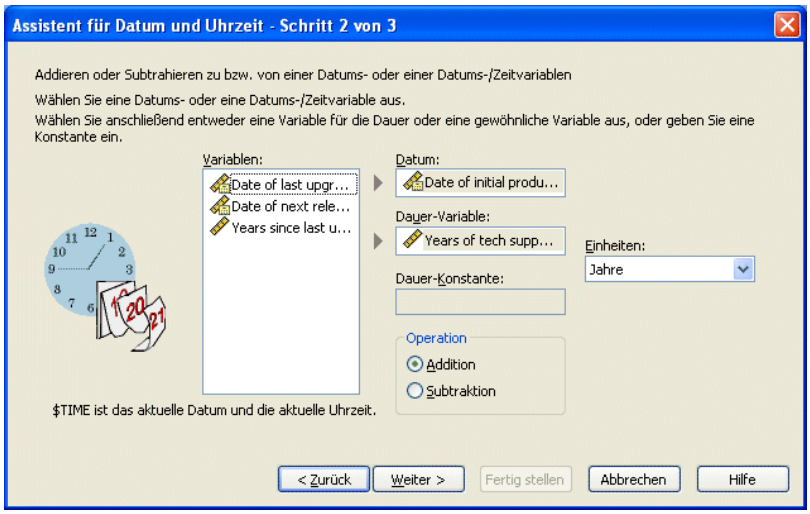
Sie können Werte für die Zeitdauer zu einem Datum addieren bzw. subtrahieren, beispielsweise 10 Tage oder 12 Monate. Setzen wir das Beispiel des Softwareunternehmens aus dem vorangegangen Abschnitt fort: Wir bestimmen das Datum, an dem der ursprüngliche Vertrag über den Technischen Support für die einzelnen Kunden ausläuft. Die Datendatei *upgrade.sav* enthält eine Variable für die Anzahl der Jahre für den vertraglich zugesicherten Support und eine Variable für das ursprüngliche Kaufdatum. Nun können wir das Enddatum des ursprünglichen Supports bestimmen, indem wir die Jahre des Supportzeitraums zum Kaufdatum hinzufügen.

So fügen Sie eine Dauer zu einem Datum hinzu:

- Wählen Sie im Einführungsbildschirm des Assistenten für Datum und Uhrzeit die Option Berechnungen mit Datums- und Zeitwerten durchführen und klicken Sie anschließend auf Weiter.

- Wählen Sie die Option Addieren bzw. Subtrahieren einer Dauer zu bzw. von einem Datum und klicken Sie auf Weiter.

Abbildung 9-16
Hinzufügen einer Dauer zu einem Datum: Schritt 2



- Wählen Sie als Datum die Option *Date of initial product license* (Datum der ursprünglichen Produktlizenz).

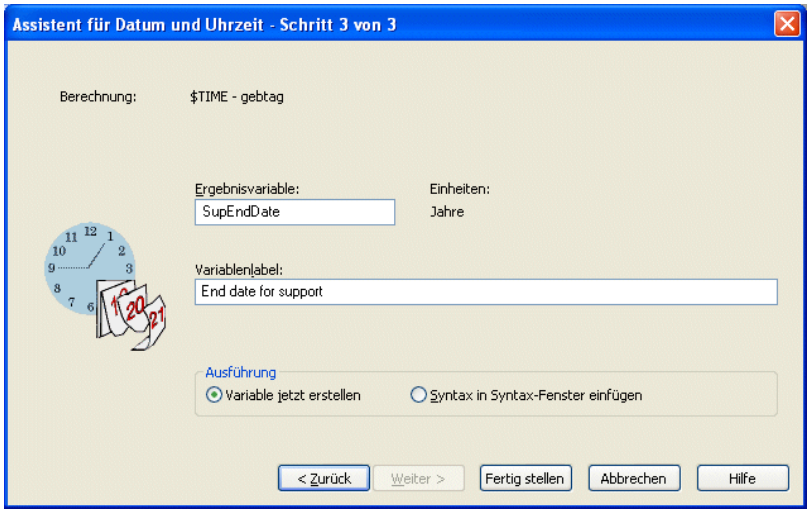
- Wählen Sie *Years of tech support* (Jahre des Technischen Supports) als Dauer-Variable aus.

  Da es sich bei *Years of tech support* (Jahre des Technischen Supports) einfach um eine numerische Variable handelt, müssen Sie angeben, welche Einheiten verwendet werden sollen, wenn diese Variable als Dauer hinzugefügt wird.

- Wählen Sie in der Dropdown-Liste für die Einheiten die Option Jahre aus.

- Klicken Sie auf Weiter.

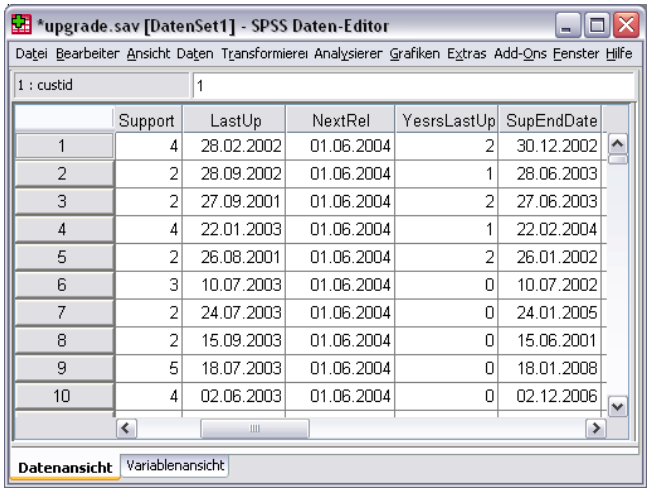
Abbildung 9-17
Hinzufügen einer Dauer zu einem Datum: Schritt 3



- Geben Sie *SupEndDate* als Namen der Ergebnisvariablen ein. Ergebnisvariablen können nicht denselben Namen haben wie bestehende Variablen.
- Geben Sie *End date for support* (Enddatum für den Support) als Label für die Ergebnisvariable ein. Die Variablenlabels für Ergebnisvariablen sind optional.
- Klicken Sie auf Fertig stellen, um die neue Variable zu erstellen.

Die neue Variable wird im Daten-Editor angezeigt.

Abbildung 9-18
Neue Variable im Daten-Editor



	Support	LastUp	NextRel	YesrsLastUp	SupEndDate
1	4	28.02.2002	01.06.2004	2	30.12.2002
2	2	28.09.2002	01.06.2004	1	28.06.2003
3	2	27.09.2001	01.06.2004	2	27.06.2003
4	4	22.01.2003	01.06.2004	1	22.02.2004
5	2	26.08.2001	01.06.2004	2	26.01.2002
6	3	10.07.2003	01.06.2004	0	10.07.2002
7	2	24.07.2003	01.06.2004	0	24.01.2005
8	2	15.09.2003	01.06.2004	0	15.06.2001
9	5	18.07.2003	01.06.2004	0	18.01.2008
10	4	02.06.2003	01.06.2004	0	02.12.2006

Kapitel
10

Sortieren und Auswählen von Daten

Datendateien liegen nicht immer genau in der Form vor, die Sie gerade benötigen. Um Daten für die Analyse vorzubereiten, stehen Ihnen mehrere Möglichkeiten zur Dateitransformation zur Verfügung, darunter folgende Funktionen:

- **Sortieren von Daten.** Sie können Fälle nach dem Wert einer oder mehrerer Variablen sortieren lassen.
- **Auswählen von Teilmengen von Fällen.** Sie können die Analyse auf eine Teilmenge von Fällen beschränken oder Analysen für verschiedene Teilmengen gleichzeitig vornehmen.

Für die Beispiele dieses Kapitels wird die Datendatei *demo.sav* verwendet. Für weitere Informationen siehe Beispieldateien in Anhang A auf S. 163.

Sortieren von Daten

Das Sortieren von Fällen, d. h. der Zeilen der Datendatei, ist für bestimmte Analysetypen hilfreich und mitunter auch erforderlich.

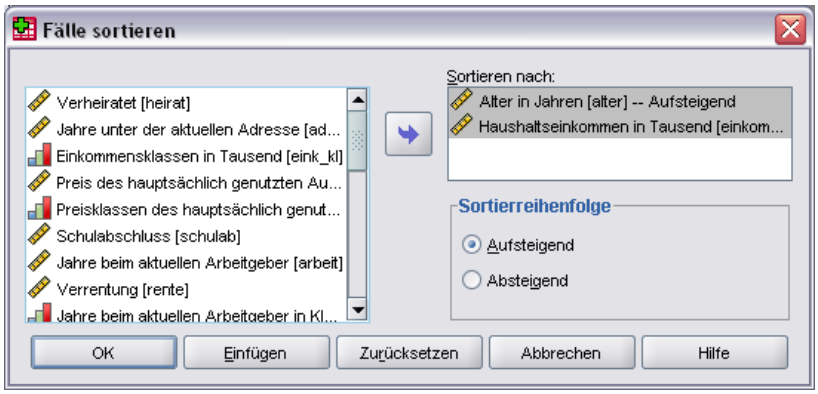
Um die Reihenfolge der Fälle in der Datendatei auf der Grundlage des Werts mindestens einer Sortiervariablen zu ändern, gehen Sie folgendermaßen vor:

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Daten
  Fälle sortieren

Das Dialogfeld “Fälle sortieren” wird angezeigt.

Abbildung 10-1
Dialogfeld “Fälle sortieren”



▶ Fügen Sie der Liste “Sortieren nach” die Variablen *Alter in Jahren [alter]* und *Haushaltseinkommen in Tausend [einkomm]* hinzu.

Bei der Auswahl mehrerer Sortiervariablen bestimmt die Reihenfolge, in der diese in der Liste “Sortieren nach” angezeigt werden, die Sortierreihenfolge der Fälle. In diesem Beispiel werden die Fälle auf der Grundlage der Liste “Sortieren nach” nach dem Wert von *Haushaltseinkommen in Tausend [einkomm]* innerhalb von Kategorien von *Alter in Jahren [alter]* sortiert. Bei String-Variablen stehen Großbuchstaben in der Sortierreihenfolge vor den entsprechenden Kleinbuchstaben. Der String-Wert *Ja* steht in der Sortierreihenfolge beispielsweise vor *ja*.

Verarbeitung von aufgeteilten Dateien

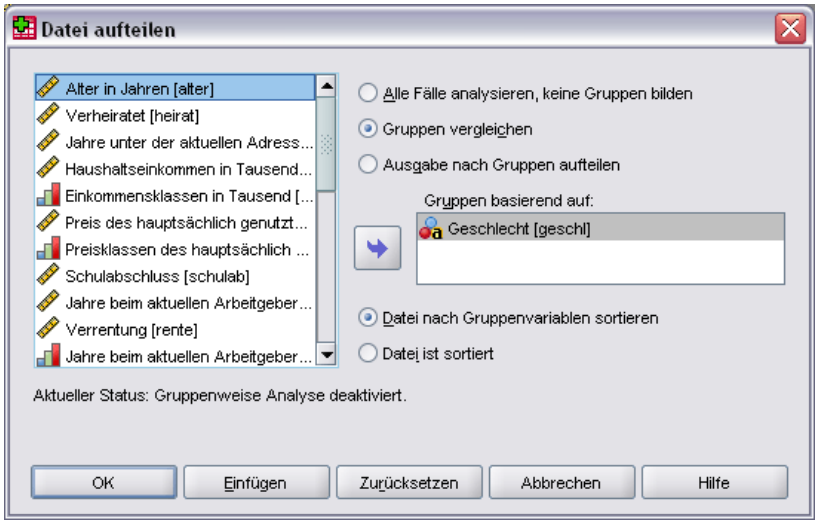
So teilen Sie die Datendatei zu Analysezwecken in einzelne Gruppen auf:

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Daten
  Datei aufteilen...

Das Dialogfeld “Datei aufteilen” wird geöffnet.

Abbildung 10-2
Dialogfeld “Datei aufteilen”



▶ Aktivieren Sie das Optionsfeld Gruppen vergleichen oder Ausgabe nach Gruppen aufteilen. (Die nach diesen Schritten aufgeführten Beispiele zeigen die Unterschiede zwischen den beiden Optionen.)

▶ Wählen Sie *Geschlecht [geschl]* aus, um die Datei für diese Variablen in getrennte Gruppen aufzuteilen.

Sie können numerische Variablen sowie kurze und lange String-Variablen als Gruppenvariablen verwenden. Für jede der durch die Gruppenvariablen definierten Untergruppen wird eine gesonderte Analyse durchgeführt. Bei der Auswahl mehrerer Gruppenvariablen bestimmt die

Reihenfolge, in der diese in der Liste “Gruppen basierend auf” angezeigt werden, die Gruppierung der Fälle.

Wenn Sie Gruppen vergleichen auswählen, werden die Ergebnisse aller Gruppen der aufgeteilten Datei in derselben Tabelle bzw. denselben Tabellen aufgeführt, wie beispielsweise in der folgenden Tabelle mit Auswertungsstatistiken gezeigt, die von der Prozedur “Häufigkeiten” erstellt wurde.

Abbildung 10-3
Ausgabe für die aufgeteilte Datei mit einer einzelnen Pivot-Tabelle

Statistik

Haushaltseinkommen in Tausend			
weiblich	N	Gültige	3179
		Fehlend	0
	Mittelwert		68,7798
	Median		44,0000
	Standardab weichung		75,73510
männlich	N	Gültige	3221
		Fehlend	0
	Mittelwert		70,1608
	Median		45,0000
	Standardab weichung		81,56216

Wenn Sie Ausgabe nach Gruppen aufteilen auswählen und die Prozedur “Häufigkeiten” durchführen, werden zwei Pivot-Tabellen angelegt: eine Tabelle für Frauen und eine Tabelle für Männer.

Abbildung 10-4
Ausgabe für die aufgeteilte Datei mit der Pivot-Tabelle für Frauen

Statistik[a]

Haushaltseinkommen in Tausend		
N	Gültige	3179
	Fehlend	0
Mittelwert		68,7798
Median		44,0000
Standardab weichung		75,73510

a. Geschlecht = weiblich

Abbildung 10-5
Ausgabe für die aufgeteilte Datei mit der Pivot-Tabelle für Männer

Statistik[a]

Haushaltseinkommen in Tausend		
N	Gültige	3221
	Fehlend	0
Mittelwert		70,1608
Median		45,0000
Standardab weichung		81,56216

a. Geschlecht = männlich

Sortieren von Fällen für die Verarbeitung von aufgeteilten Dateien

In der Prozedur “Datei aufteilen” wird eine neue Untergruppe erstellt, sobald ein anderer Wert für eine der Gruppenvariablen ermittelt wird. Es ist daher wichtig, die Fälle anhand der Werte der Gruppenvariablen zu sortieren, ehe Sie die aufgeteilte Datei verarbeiten.

In der Standardeinstellung wird die Datendatei in der Prozedur “Datei aufteilen” auf der Grundlage der Werte der Gruppenvariablen sortiert. Wenn die Datei bereits in der richtigen Reihenfolge sortiert ist, können Sie Verarbeitungszeit sparen, indem Sie Datei ist sortiert auswählen.

Aktivieren und Deaktivieren der Verarbeitung von aufgeteilten Dateien

Nachdem Sie die Verarbeitung von aufgeteilten Dateien aktiviert haben, bleibt sie für den Rest der Sitzung aktiv. Sie muss ausdrücklich deaktiviert werden.

- **Alle Fälle analysieren.** Diese Option deaktiviert die Verarbeitung von aufgeteilten Dateien.
- **Gruppen vergleichen** und **Ausgabe nach Gruppen aufteilen**. Diese Option aktiviert die Verarbeitung von aufgeteilten Dateien.

Wenn die Verarbeitung von aufgeteilten Dateien aktiviert ist, wird in der Statusleiste unten im Anwendungsfenster die Meldung Datei aufteilen an angezeigt.

Auswählen von Teilmengen von Fällen

Sie können die Analyse anhand von Kriterien, zu denen Variablen und komplexe Ausdrücke gehören, auf eine bestimmte Untergruppe beschränken. Sie können auch eine Zufallsstichprobe aus den Fällen auswählen. Die Kriterien zum Festlegen der Untergruppen können folgende Elemente enthalten:

- Variablenwerte und -bereiche
- Datums- und Zeitbereiche
- Fallnummern (Zeilennummern)
- Arithmetische Ausdrücke
- Logische Ausdrücke
- Funktionen

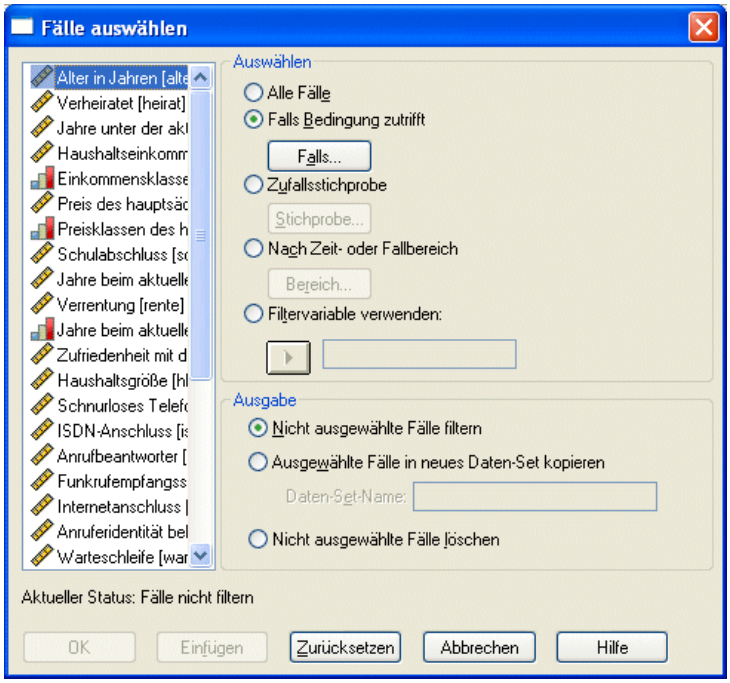
So wählen Sie eine Teilmenge der Fälle für die Analyse aus:

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Daten
  Fälle auswählen...

Dadurch wird das Dialogfeld “Fälle auswählen” aufgerufen.

Abbildung 10-6
Dialogfeld "Fälle auswählen"



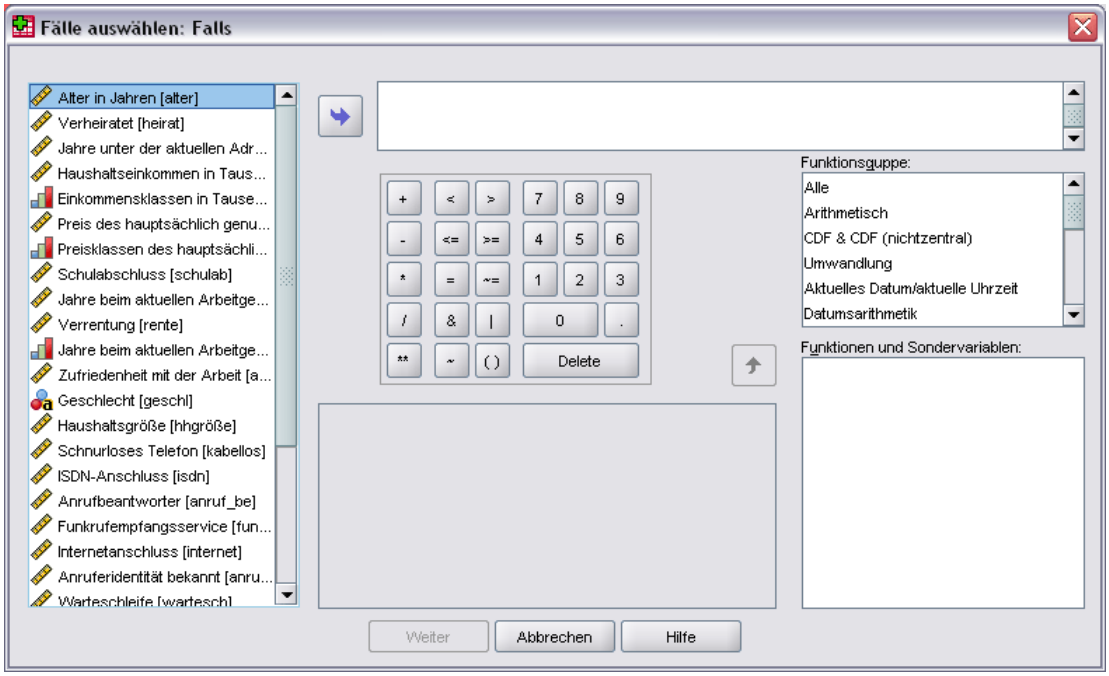
Auswählen von Teilmengen von Fällen anhand eines bedingten Ausdrucks

So wählen Sie Fälle auf der Grundlage eines bedingten Ausdrucks aus:

▶ Wählen Sie Falls Bedingung zutrifft aus und klicken Sie im Dialogfeld "Fälle auswählen" auf Falls.

Dadurch wird das Dialogfeld "Fälle auswählen: Falls" aufgerufen.

Abbildung 10-7
Dialogfeld "Fälle auswählen: Falls"



Der bedingte Ausdruck kann vorhandene Variablennamen, Konstanten, arithmetische Operatoren, logische Operatoren, relationale Operatoren und Funktionen enthalten. Sie können den Ausdruck im Textfeld genau wie Text in einem Ausgabefenster eingeben und bearbeiten. Sie können auch die Rechentastatur, die Variablenliste und die Funktionsliste verwenden, um Elemente in den Ausdruck einzufügen. Für weitere Informationen siehe Verwendung bedingter Ausdrücke in Kapitel 9 auf S. 134.

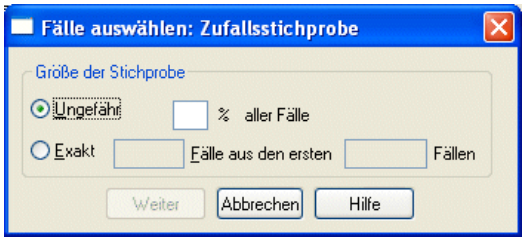
Auswählen einer Zufallsstichprobe aus den Fällen

So erhalten Sie eine Zufallsstichprobe:

- ▶ Wählen Sie im Dialogfeld "Fälle auswählen" die Option Zufallsstichprobe aus.
- ▶ Klicken Sie auf Stichprobe.

Dadurch wird das Dialogfeld "Fälle auswählen: Zufallsstichprobe" aufgerufen.

Abbildung 10-8
Dialogfeld "Fälle auswählen: Zufallsstichprobe"



Für die Stichprobengröße stehen die folgenden Optionen zur Auswahl:

- **Ungefähr.** Geben Sie einen Prozentsatz ein. SPSS erstellt eine Zufallsstichprobe, die ungefähr den angegebenen Prozentsatz aller Fälle enthält.
- **Exakt.** Geben Sie die gewünschte Anzahl der Fälle ein. Sie müssen außerdem die Anzahl der Fälle angeben, aus denen die Stichprobe gezogen werden soll. Diese zweite Zahl muss kleiner oder gleich der Gesamtanzahl der Fälle in der Datendatei sein. Wenn die angegebene Anzahl die Gesamtanzahl der Fälle in der Datendatei übersteigt, enthält die Stichprobe entsprechend weniger Fälle als die geforderte Anzahl.

Auswählen eines Zeit- oder Fallbereichs

So wählen Sie einen Fallbereich anhand von Datum, Uhrzeit oder Beobachtungs- bzw. Zeilennummern aus.

▶ Wählen Sie die Option Nach Zeit- oder Fallbereich aus und klicken Sie im Dialogfeld "Fälle auswählen" auf Bereich.

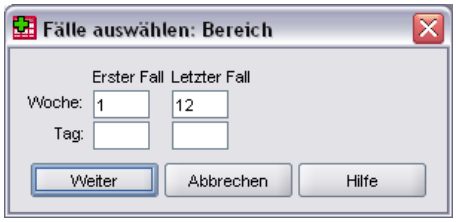
Dadurch wird das Dialogfeld "Fälle auswählen: Bereich" geöffnet, in dem Sie einen Bereich von Beobachtungs- bzw. Zeilennummern auswählen können.

Abbildung 10-9
Dialogfeld "Fälle auswählen: Bereich"

- **Erster Fall.** Geben Sie das Startdatum und/oder die Startzeit für den Bereich ein. Wenn keine Datumsvariablen definiert sind, geben Sie die erste Beobachtungsnummer (die Zeilennummer im Daten-Editor, wenn die Option "Datei aufteilen" deaktiviert ist) ein. Wenn Sie im Feld "Letzter Fall" keinen Wert angeben, werden alle Fälle vom Startdatum/von der Startzeit bis zum Ende der Zeitreihe ausgewählt.
- **Letzter Fall.** Geben Sie das Enddatum und/oder die Abschlusszeit für den Bereich ein. Wenn keine Datumsvariablen definiert sind, geben Sie die letzte Beobachtungsnummer (die Zeilennummer im Daten-Editor, wenn die Option "Datei aufteilen" deaktiviert ist) ein. Wenn Sie im Feld "Erster Fall" keinen Wert angeben, werden alle Fälle vom Beginn der Zeitreihe bis zum Enddatum/zur Abschlusszeit ausgewählt.

Für Zeitreihendaten mit definierten Datumsvariablen können Sie einen Datumsbereich und/oder Zeitbereich auf der Grundlage von definierten Datumsvariablen auswählen. Jeder Fall steht für Beobachtungen zu einem anderen Zeitpunkt und die Datei ist in chronologischer Reihenfolge sortiert.

Abbildung 10-10
Dialogfeld "Fälle auswählen: Bereich" (Zeitreihen)



So generieren Sie Datumsvariablen für Zeitreihendaten:

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Daten
  Datum definieren...

Behandlung von nicht ausgewählten Fällen

Die folgenden Optionen stehen für die Behandlung nicht ausgewählter Fälle zur Auswahl:

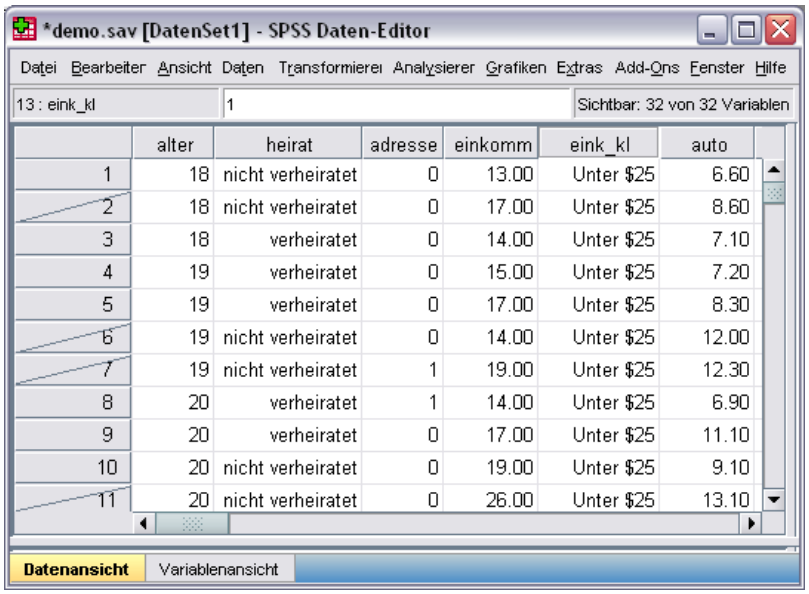
- **Nicht ausgewählte Fälle filtern.** Nicht ausgewählte Fälle werden nicht in die Analyse aufgenommen, verbleiben jedoch im Daten-Set. Sie können die nicht ausgewählten Fälle später in der Sitzung verwenden, wenn Sie die Filterfunktion deaktivieren. Wenn Sie eine Zufallsstichprobe oder Fälle anhand eines bedingten Ausdrucks auswählen, wird die Variable *filter_$* mit dem Wert 1 für ausgewählte Fälle und dem Wert 0 für nicht ausgewählte Fälle erzeugt.
- **Ausgewählte Fälle in neues Daten-Set kopieren.** Die ausgewählten Fälle werden in ein neues Daten-Set kopiert, das ursprüngliche Daten-Set bleibt unverändert. Nicht ausgewählte Fälle werden nicht in das neue Daten-Set aufgenommen. Sie verbleiben im ursprünglichen Zustand im ursprünglichen Daten-Set.
- **Nicht ausgewählte Fälle löschen.** Nicht ausgewählte Fälle werden aus dem Daten-Set gelöscht. Gelöschte Fälle können nur wiederhergestellt werden, indem Sie die Datei ohne Speichern der Änderungen schließen und sie dann erneut öffnen. Wenn Sie die Änderungen in der Datendatei speichern, werden die Fälle dauerhaft gelöscht.

Anmerkung: Wenn Sie nicht ausgewählte Fälle löschen und die Datei speichern, können die Fälle nicht wiederhergestellt werden.

Status der Fallauswahl

Wenn Sie eine Teilmenge von Fällen ausgewählt, nicht ausgewählte Fälle jedoch nicht verworfen haben, sind die nicht ausgewählten Fälle im Daten-Editor mit einer diagonalen Linie durch die Zeilennummer gekennzeichnet.

Abbildung 10-11
Status der Fallauswahl



	alter	heirat	adresse	einkomm	eink_kl	auto
1	18	nicht verheiratet	0	13.00	Unter $25	6.60
2	18	nicht verheiratet	0	17.00	Unter $25	8.60
3	18	verheiratet	0	14.00	Unter $25	7.10
4	19	verheiratet	0	15.00	Unter $25	7.20
5	19	verheiratet	0	17.00	Unter $25	8.30
6	19	nicht verheiratet	0	14.00	Unter $25	12.00
7	19	nicht verheiratet	1	19.00	Unter $25	12.30
8	20	verheiratet	1	14.00	Unter $25	6.90
9	20	verheiratet	0	17.00	Unter $25	11.10
10	20	nicht verheiratet	0	19.00	Unter $25	9.10
11	20	nicht verheiratet	0	26.00	Unter $25	13.10

Kapitel

11

Weitere statistische Prozeduren

In diesem Kapitel finden Sie kurze Beispiele zu ausgewählten statistischen Prozeduren. Die Prozeduren sind in der Reihenfolge geordnet, in der sie im Menü "Analysieren" angezeigt werden.

Die Beispiele dienen zur Veranschaulichung von exemplarischen Angaben, die zum Ausführen einer statistischen Prozedur benötigt werden. Für die Beispiele in diesem Kapitel wird die Datendatei *demo.sav* verwendet, mit Ausnahme der folgenden Beispiele:

- Für das Beispiel für den *T*-Test bei gepaarten Stichproben wird die Datendatei *dietstudy.sav* verwendet. Hierbei handelt es sich um eine hypothetische Datendatei mit den Ergebnissen einer Studie zur "Stillman-Diät". In den Beispielen dieses Kapitels müssen Sie die Prozeduren selbst ausführen, damit die Ausgabe angezeigt wird.
- Für das Beispiel zur Korrelation wird die Datei *Employee data.sav* verwendet, die historische Daten über die Mitarbeiter eines Unternehmens enthält.
- Für das Beispiel zur Prozedur "Exponentielles Glätten" wird die Datendatei *Inventor.sav* verwendet. Diese Datei enthält Lagerdaten, die über einen Zeitraum von 70 Tagen gesammelt wurden.

Informationen zu einzelnen Elementen in einem Dialogfeld erhalten Sie, wenn Sie auf Hilfe klicken. Wenn Sie eine bestimmte Statistik, beispielsweise Perzentile, suchen, verwenden Sie den Index oder die Suchfunktion im Hilfesystem. Weitere Informationen über die Interpretationsmöglichkeiten der Ergebnisse, die mit diesen Prozeduren erzielt werden können, erhalten Sie in Lehrbüchern für Statistik oder Datenanalyse.

Zusammenfassen von Daten

Über das Menü "Analysieren" und das Untermenü "Deskriptive Statistiken" stehen Ihnen Verfahren für die Zusammenfassung von Daten mit Statistiken und Diagrammen zur Verfügung.

Explorative Datenanalyse

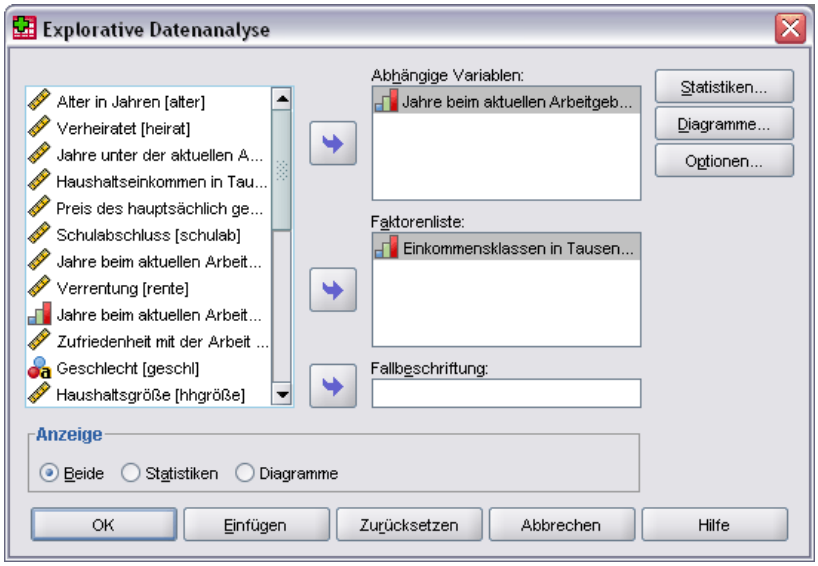
Angenommen, Sie möchten die Verteilung für "Jahre beim aktuellen Arbeitgeber" für die einzelnen Einkommensklassen näher betrachten. Mit Hilfe der Prozedur "Explorative Datenanalyse" können Sie die Verteilung für "Jahre beim aktuellen Arbeitgeber" innerhalb der Kategorien einer anderen Variablen untersuchen.

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Analysieren
  Deskriptive Statistiken
    Explorative Datenanalyse

Dadurch wird das Dialogfeld “Explorative Datenanalyse” aufgerufen.

Abbildung 11-1
Dialogfeld “Explorative Datenanalyse”



▶ Wählen Sie die Variable *Jahre beim aktuellen Arbeitgeber [arbeit]* aus, und verschieben Sie sie in die Liste “Abhängige Variablen”.

▶ Wählen Sie die Variable *Einkommensklassen in Tausend [eink_kl]* aus, und verschieben Sie sie in das Feld “Faktorenliste”.

▶ Klicken Sie auf OK, um die Prozedur “Explorative Datenanalyse” auszuführen.

In der Ausgabe werden für die Jahre beim aktuellen Arbeitgeber in den einzelnen Einkommensklassen deskriptive Statistiken und ein Stengel-Blatt-Diagramm angezeigt. Außerdem umfaßt die Anzeige im Viewer ein Boxplot (im Standardformat), in dem die Jahre beim aktuellen Arbeitgeber nach Einkommensklassen verglichen werden. Für jede Kategorie werden im Boxplot der Median und der Interquartilbereich (25. bis 75. Perzentil) sowie Ausreißer (durch O gekennzeichnet) und Extremwerte (durch * gekennzeichnet) angezeigt.

Weitere Informationen zum Zusammenfassen von Daten

Daten können auf viele verschiedene Weisen ausgewertet werden. Zum Berechnen von Medianen oder Perzentilen verwenden Sie z. B. die Prozedur “Häufigkeiten” oder “Explorative Datenanalyse”. Weitere Methoden sind:

- **Deskriptive Statistiken.** Für das Einkommen können standardisierte Werte (auch *Z*-Werte genannt) berechnet werden. Verwenden Sie die Prozedur “Deskriptive Statistiken”, und wählen Sie Standardisierte Werte als Variable speichern.

- **Kreuztabellen.** Mit dem Verfahren “Kreuztabellen” können Sie die Beziehung zwischen zwei oder mehr kategorialen Variablen anzeigen.
- **Prozedur “Zusammenfassen”.** Mit Hilfe der Prozedur “Zusammenfassen” können Sie im Ausgabefenster eine Auflistung der tatsächlichen Werte für Alter, Geschlecht und Einkommen der ersten 25 oder 50 Fälle schreiben.

▶ Wenn Sie die Prozedur “Zusammenfassen” ausführen möchten, wählen Sie die folgenden Menübefehle aus:

Analysieren
  Berichte
    Fälle zusammenfassen

Vergleichen von Mittelwerten

Im Menü “Analysieren” stehen Ihnen über das Untermenü “Mittelwerte vergleichen” verschiedene Verfahren zur Verfügung, mit denen Sie deskriptive Statistiken anzeigen lassen und testen können, ob die Differenzen zwischen zwei Mittelwerten sowohl für unabhängige als auch für gepaarte Stichproben signifikant sind. Mit der Prozedur “Einfaktorielle ANOVA” können Sie auch testen, ob signifikante Differenzen zwischen mehr als zwei unabhängigen Mittelwerten vorliegen.

Mittelwerte

In der Datei *demo.sav* sind mehrere Variablen für die Einteilung der Personen in Gruppen verfügbar. Um diese Gruppen anschließend zu vergleichen, können verschiedene Statistiken berechnet werden. So können Sie beispielsweise das durchschnittliche Haushaltseinkommen

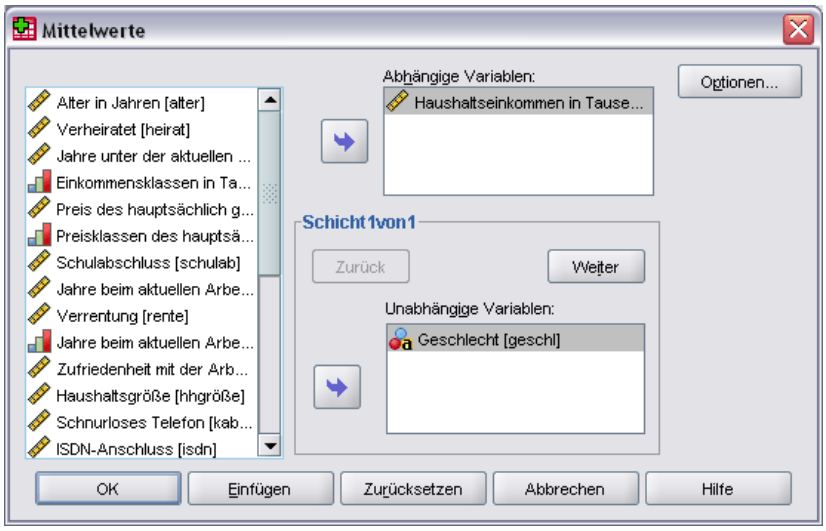
(Mittelwerte) für Männer und Frauen berechnen. Mittelwerte können mit den folgenden Schritten berechnet werden:

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Analysieren
  Mittelwerte vergleichen
    Mittelwerte

Dadurch wird das Dialogfeld "Mittelwerte" aufgerufen.

Abbildung 11-2
Dialogfeld "Mittelwerte" (Schicht 1)

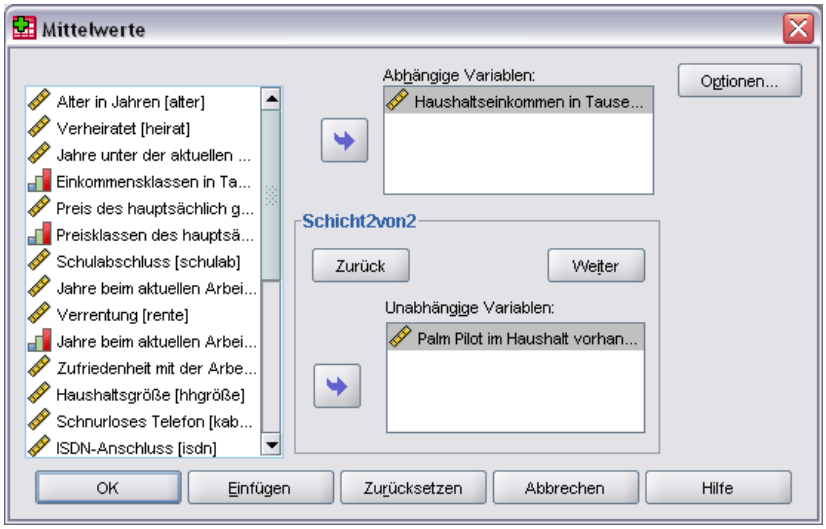


▶ Wählen Sie die Variable *Haushaltseinkommen in Tausend [einkomm]* aus, und verschieben Sie sie in die Liste "Abhängige Variablen".

▶ Wählen Sie die Variable *Geschlecht [geschl]* aus, und verschieben Sie sie in die Liste "Unabhängige Variablen" in Schicht 1.

▶ Klicken Sie auf Weiter, um eine weitere Schicht zu erstellen.

Abbildung 11-3
Dialogfeld "Mittelwerte" (Schicht 2)



- Wählen Sie die Variable *Palm Pilot im Haushalt vorhanden [palm]* aus, und verschieben Sie sie in die Liste "Unabhängige Variablen" in Schicht 2.
- Klicken Sie auf OK, um die Prozedur auszuführen.

T-Test bei gepaarten Stichproben

Stichproben werden gepaart, wenn aufgrund der spezifischen Datenstruktur zwei Beobachtungen zu demselben Einzelwert oder zwei Beobachtungen vorhanden sind, die mit zwei Einzelwerten (z. B. Zwillingen) einer anderen Variablen übereinstimmen. In der Datendatei *dietstudy.sav* wird für jede Person, die an der Studie teilnahm, das Anfangs- und das Endgewicht angegeben. Wenn die Diät erfolgreich war, ist ein signifikanter Unterschied zwischen dem Gewicht des Teilnehmers vor und nach der Studie zu erwarten.

Um einen *T*-Test für die Werte für das Anfangs- und das Endgewicht durchzuführen, führen Sie die folgenden Schritte aus:

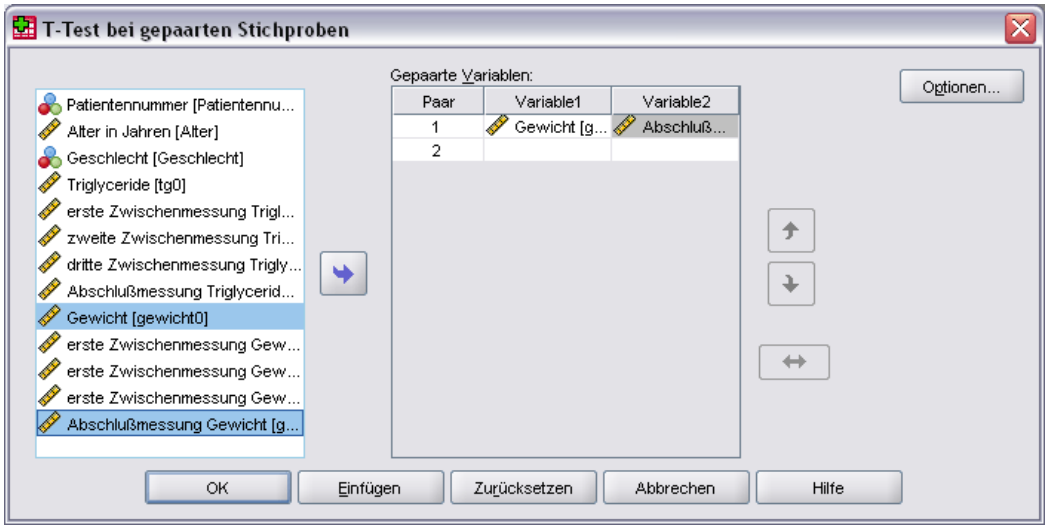
- Öffnen Sie die Datendatei *dietstudy.sav*. Für weitere Informationen siehe Beispieldateien in Anhang A auf S. 163.

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Analysieren
Mittelwerte vergleichen
T-Test bei gepaarten Stichproben

Damit wird das Dialogfeld “T-Test bei gepaarten Stichproben” geöffnet.

Abbildung 11-4
Dialogfeld “T-Test bei gepaarten Stichproben”



▶ Wählen Sie *Gewicht* und *Abschlussmessung Gewicht* als gepaarte Variablen aus.

▶ Klicken Sie auf OK, um die Prozedur auszuführen.

Die Ergebnisse zeigen, dass das Endgewicht deutlich vom Anfangsgewicht abweicht. Dies entspricht der geringen Wahrscheinlichkeit, die in der Spalte *Sig. (2-seitig)* der Tabelle “Test bei gepaarten Stichproben” angezeigt wird.

Weitere Informationen zum Vergleichen von Mittelwerten

In den folgenden Beispielen werden einige Möglichkeiten aufgeführt, mit denen Sie Mittelwerte über andere Prozeduren vergleichen können.

- **T-Test bei unabhängigen Stichproben.** Wenn Sie Mittelwerte einer Variablen in mehreren unabhängigen Gruppen mit einem *T*-Test vergleichen, sind die Stichproben unabhängig. Die in der Datei *demo.sav* aufgeführten Männer und Frauen können durch die Variable *Geschlecht [geschl]* in unabhängige Gruppen eingeteilt werden. Mit Hilfe eines *T*-Tests können Sie bestimmen, ob die Mittelwerte für das Haushaltseinkommen bei Männern und Frauen übereinstimmen.
- **T-Test bei einer Stichprobe.** Sie können testen, ob das Haushaltseinkommen von Personen mit einem Hochschulabschluss vom nationalen Durchschnitt bzw. vom Durchschnitt des Bundeslandes abweicht. Verwenden Sie Fälle auswählen im Menü “Daten”, um die Fälle auszuwählen, für die *Schulabschluss [schulab]* >= 4 gilt. Führen Sie dann die Prozedur

"T-Test bei einer Stichprobe" aus, um den Wert für *Haushaltseinkommen in Tausend [einkomm]* mit dem Testwert 75 zu vergleichen.

- **Einfaktorielle ANOVA.** Die Variable *Schulabschluss [schulab]* unterteilt die Arbeitnehmer nach ihrem Schulabschluss in fünf unabhängige Gruppen. Mit der Prozedur "Einfaktorielle ANOVA" können Sie testen, ob sich die Mittelwerte von *Haushaltseinkommen in Tausend [einkomm]* der fünf Gruppen signifikant unterscheiden.

ANOVA-Modelle

Im Menü "Analysieren" stehen Ihnen über das Untermenü "Allgemeines lineares Modell" verschiedene Verfahren zur Verfügung, mit denen Sie univariate Varianzanalysemodelle testen können. (Wenn nur ein Faktor vorhanden ist, kann auch die Prozedur "Einfaktorielle ANOVA" im Untermenü "Mittelwerte vergleichen" verwendet werden.)

Univariate Varianzanalyse

Mit der Prozedur "GLM - Univariat" können Sie eine Varianzanalyse für faktorielle Designs durchführen. Mit einem einfachen faktoriellen Design können Sie testen, ob das Haushaltseinkommen einer Person und ihre Zufriedenheit mit der Arbeit Auswirkungen auf die Jahre beim aktuellen Arbeitgeber haben.

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Analysieren
  Allgemeines lineares Modell
    Univariat...

Dadurch wird das Dialogfeld "Univariat" aufgerufen.

Abbildung 11-5
Dialogfeld "Univariat"

- Wählen Sie die Variable *Jahre beim aktuellen Arbeitgeber [arbeit]* aus, und verschieben Sie sie in die Liste “Abhängige Variable”.
- Wählen Sie die Variablen *Einkommensklassen in Tausend [eink_kl]* und *Zufriedenheit mit der Arbeit [arb_zufr]* aus, und verschieben Sie sie in die Liste “Feste Faktoren”.
- Klicken Sie auf OK, um die Prozedur auszuführen.

  In der Tabelle “Tests der Zwischensubjekteffekte” ist zu erkennen, daß die Auswirkungen von Einkommen und Arbeitszufriedenheit eindeutig signifikant sind und daß das beobachtete Signifikanzniveau der Wechselwirkung von Einkommen und Arbeitszufriedenheit 0,000 beträgt. Weitere Informationen über Interpretationsmöglichkeiten erhalten Sie in Lehrbüchern für Statistik oder Datenanalyse.

Korrelation zwischen Variablen

Im Menü “Analysieren” sind über das Untermenü “Korrelation” Zusammenhangsmaße für zwei oder mehr numerische Variablen verfügbar. Für die Beispiele dieses Themas wird die Datendatei *Employee data.sav* verwendet.

Bivariate Korrelationen

Mit der Prozedur “Bivariate Korrelationen” werden Statistiken wie der Korrelationskoeffizient nach Pearson berechnet. Mit Korrelationen werden die Beziehungen zwischen Variablen oder deren Rängen gemessen. Der Wertebereich für Korrelationskoeffizienten reicht von –1 (perfekter negativer Zusammenhang) bis +1 (perfekter positiver Zusammenhang). Der Wert 0 bedeutet, daß kein linearer Zusammenhang besteht.

Mithilfe des Korrelationskoeffizienten nach Pearson können Sie feststellen, ob zwischen *Gehalt [gehalt]* und *Anfangsgehalt [agehalt]* in der Datendatei *Employee data.sav* ein enger linearer Zusammenhang besteht.

Partielle Korrelationen

Partielle Korrelationskoeffizienten beschreiben die Beziehung zwischen zwei Variablen. Diese Koeffizienten werden mit Hilfe der Prozedur “Partielle Korrelationen” berechnet, wobei die Effekte von einer oder mehreren zusätzlichen Variablen einbezogen werden.

Sie können die Korrelation zwischen *Gehalt [gehalt]* und *Anfangsgehalt [agehalt]* schätzen, indem Sie die linearen Effekte von *Beschäftigungsdauer [dauer]* und *Berufserfahrung in Monaten [erfahr]* berücksichtigen. Die Anzahl der Kontrollvariablen bestimmt die Ordnung des partiellen Korrelationskoeffizienten.

Um die Prozedur “Partielle Korrelationen” durchzuführen, führen Sie folgende Schritte aus:

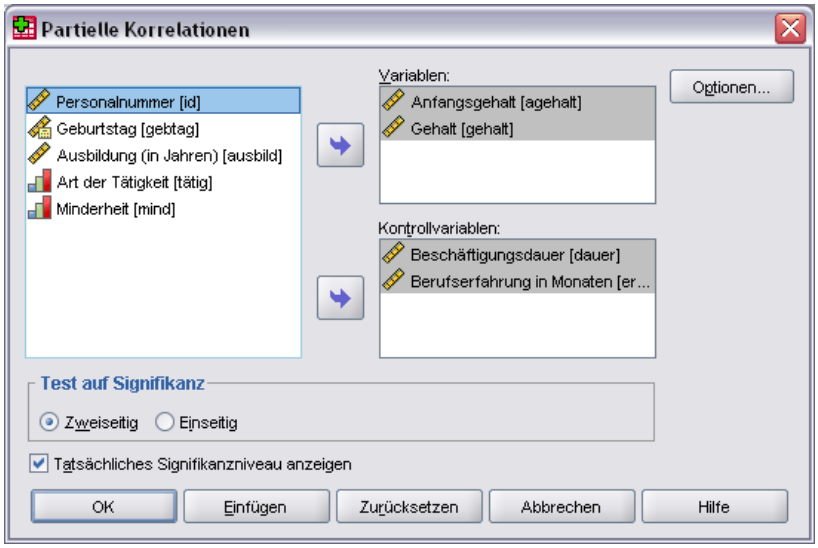
- Öffnen Sie die Datei *Employee data.sav*. Diese Datei befindet sich normalerweise im Installationsverzeichnis.

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Analysieren
  Korrelation
    Partiell

Dadurch wird das Dialogfeld “Partielle Korrelationen” geöffnet.

Abbildung 11-6
Dialogfeld “Partielle Korrelationen”



▶ Wählen Sie die Variablen *Gehalt [gehalt]* und *Anfangsgehalt [agehalt]* aus, und verschieben Sie sie in die Liste “Variablen”.

▶ Wählen Sie die Variablen *Beschäftigungsdauer [dauer]* und *Berufserfahrung in Monaten [erfahr]* aus, und verschieben Sie sie in die Liste “Kontrollvariablen”.

▶ Klicken Sie auf OK, um die Prozedur auszuführen.

Die Ausgabe enthält eine Tabelle mit partiellen Korrelationskoeffizienten, Freiheitsgraden und dem Signifikanzniveau für das Paar *Gehalt [gehalt]* und *Anfangsgehalt [agehalt]*.

Regressionsanalyse

Im Menü “Analysieren” stehen Ihnen über das Untermenü “Regression” verschiedene Regressionsverfahren zur Verfügung.

Lineare Regression

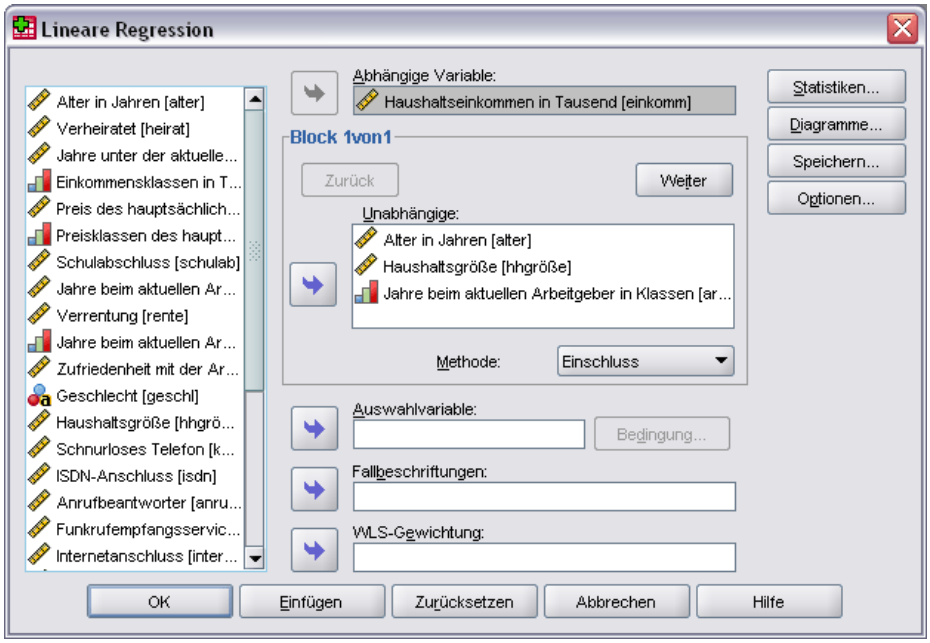
Mit der Prozedur “Lineare Regression” wird die Beziehung zwischen einer abhängigen Variablen und einer Reihe unabhängiger Variablen untersucht. Mit der Prozedur können Sie das Haushaltseinkommen (abhängige Variable) einer Person anhand von unabhängigen Variablen wie Alter, Haushaltsgröße und Jahre beim aktuellen Arbeitgeber vorhersagen.

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Analysieren
  Regression
    Linear...

Dadurch wird das Dialogfeld “Lineare Regression” geöffnet.

Abbildung 11-7
Dialogfeld “Lineare Regression”



▶ Wählen Sie die Variable *Haushaltseinkommen in Tausend [einkomm]* aus, und verschieben Sie sie in die Liste “Abhängige Variable”.

▶ Wählen Sie die Variablen *Alter in Jahren [alter]*, *Haushaltsgröße [hhgröße]* und *Jahre beim aktuellen Arbeitgeber [arbeit]* aus, und verschieben Sie sie in die Liste “Unabhängige Variable(n)”.

▶ Klicken Sie auf OK, um die Prozedur auszuführen.

Die Ausgabe enthält eine Anpassungsstatistik sowie die partiellen Regressionskoeffizienten für die Variablen.

Untersuchen der Anpassung. Um zu überprüfen, ob sich das Regressionsmodell für Ihre Daten eignet, können Sie die Residuen und andere Diagnosetypen untersuchen, die von dieser Prozedur bereitgestellt werden. Klicken Sie im Dialogfeld “Lineare Regression” auf Speichern, um eine Liste der neuen Variablen anzuzeigen, die Sie der Datendatei hinzufügen können. Wenn Sie eine dieser Variablen erstellen, ohne die Datendatei zu speichern, sind die Variablen in späteren Sitzungen nicht mehr verfügbar.

Methoden. Wenn Sie eine große Anzahl unabhängiger Variablen zusammengestellt haben und ein Regressionsmodell erstellen möchten, in dem nur Variablen berücksichtigt werden sollen, die statistisch auf die abhängige Variable bezogen sind, können Sie in der Dropdown-Liste eine Methode auswählen. Wenn Sie beispielsweise im obigen Beispiel Schrittweise wählen, werden nur

die Variablen in die Gleichung eingefügt, die die Kriterien im Dialogfeld “Lineare Regression: Optionen” erfüllen.

Nichtparametrische Tests

Im Menü “Analysieren” stehen Ihnen über das Untermenü “Nichtparametrische Tests” nichtparametrische Tests für eine Stichprobe oder für mehrere gepaarte bzw. unabhängige Stichproben zur Verfügung. Nichtparametrische Tests erfordern keine Annahmen über die Form der Verteilungen, aus denen die Daten stammen.

Chi-Quadrat

Die Prozedur “Chi-Quadrat-Test” dient zum Testen von Hypothesen über die relative Proportion von Fällen in mehreren sich gegenseitig ausschließenden Gruppen. Sie können die Hypothese testen, daß die Personen, die an der Studie teilgenommen haben, dieselben Geschlechtsanteile wie die Gesamtbevölkerung aufweisen (50% männlich, 50% weiblich).

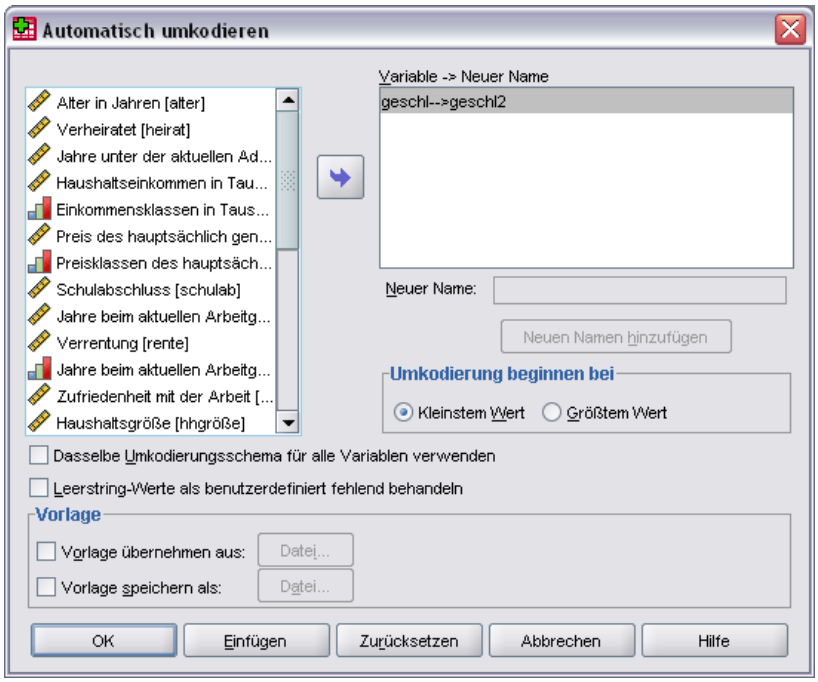
In diesem Beispiel müssen Sie die String-Variable *Geschlecht [geschl]* in eine numerische Variable umkodieren, bevor Sie die Prozedur ausführen können.

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Transformieren
  Automatisch umkodieren

Dadurch wird das Dialogfeld “Automatisch umkodieren” geöffnet.

Abbildung 11-8
Dialogfeld “Automatisch umkodieren”



▶ Wählen Sie die Variable *Geschlecht [geschl]* aus, und verschieben Sie sie in die Liste “Variable -> Neuer Name”.

▶ Geben Sie im Textfeld “Neuer Name” den Text geschl2 ein, und klicken Sie dann auf die Schaltfläche Neuen Namen hinzufügen.

▶ Klicken Sie auf OK, um die Prozedur auszuführen.

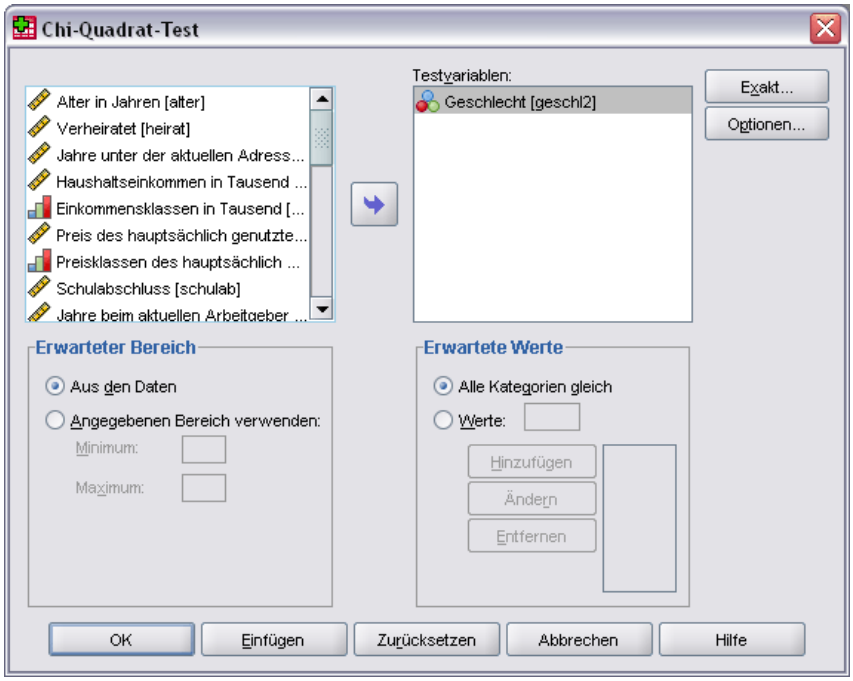
Bei diesem Verfahren wird eine neue numerische Variable mit der Bezeichnung *geschl2* erstellt, die einen Wert von 1 für Frauen und einen Wert von 2 für Männer besitzt. Jetzt kann ein Chi-Quadrat-Test mit einer numerischen Variablen ausgeführt werden.

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Analysieren
  Nichtparametrische Tests
    Chi-Quadrat...

Dadurch wird das Dialogfeld “Chi-Quadrat-Test” geöffnet.

Abbildung 11-9
Dialogfeld “Chi-Quadrat-Test”



▶ Wählen Sie die Variable *Geschlecht [geschl2]* aus, und verschieben Sie sie in die Liste “Testvariablen”.

▶ Wählen Sie die Option Alle Kategorien gleich aus, da die Anzahl von Männern und Frauen im arbeitsfähigen Alter in der Gesamtbevölkerung annähernd gleich ist.

▶ Klicken Sie auf OK, um die Prozedur auszuführen.

Die Ausgabe enthält eine Tabelle mit den erwarteten Werten und Residuen für die Kategorien. Die Signifikanz des Chi-Quadrat-Tests beträgt 0,6. Weitere Informationen zur Interpretation der Statistiken erhalten Sie in Lehrbüchern für Statistik oder Datenanalyse.

Anhang

A

Beispieldateien

Die zusammen mit dem Produkt installierten Beispieldateien finden Sie im Unterverzeichnis *Samples* des Installationsverzeichnisses.

Beschreibungen

Im Folgenden finden Sie Kurzbeschreibungen der in den verschiedenen Beispielen in der Dokumentation verwendeten Beispieldateien:

- **accidents.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um eine Versicherungsgesellschaft geht, die alters- und geschlechtsabhängige Risikofaktoren für Autounfälle in einer bestimmten Region untersucht. Jeder Fall entspricht einer Kreuzklassifikation von Alterskategorie und Geschlecht.
- **adl.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um Bemühungen geht, die Vorteile einer vorgeschlagenen Therapieform für Schlaganfallpatienten zu ermitteln. Ärzte teilten weibliche Schlaganfallpatienten nach dem Zufallsprinzip jeweils einer von zwei Gruppen zu. Die erste Gruppe erhielt die physische Standardtherapie, die zweite erhielt eine zusätzliche Emotionaltherapie. Drei Monate nach den Behandlungen wurden die Fähigkeiten der einzelnen Patienten, übliche Alltagsaktivitäten auszuführen, als ordinale Variablen bewertet.
- **advert.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um die Bemühungen eines Einzelhändlers geht, die Beziehungen zwischen den in Werbung investierten Beträgen und den daraus resultierenden Umsätzen zu untersuchen. Zu diesem Zweck hat er die Umsätze vergangener Jahre und die zugehörigen Werbeausgaben zusammengestellt.
- **aflatoxin.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um Tests von Maisernten auf Aflatoxin geht, ein Gift, dessen Konzentration stark zwischen und innerhalb von Ernteerträgen schwankt. Ein Kornverarbeitungsbetrieb hat aus 8 Ernteerträgen je 16 Proben erhalten und das Aflatoxinniveau in Teilen pro Milliarde (parts per billion, PPB) gemessen.
- **aflatoxin20.sav.** Diese Datendatei enthält die Aflatoxinmessungen aus jeder der 16 Stichproben aus den Erträgen 4 und 8 der Datendatei *aflatoxin.sav*.
- **anorectic.sav.** Bei der Ausarbeitung einer standardisierten Symptomatologie anorektischen/bulimischen Verhaltens führten Forscher eine Studie mit 55 Jugendlichen mit bekannten Ess-Störungen durch. Jeder Patient wurde vier Mal über einen Zeitraum von vier Jahren untersucht, es fanden also insgesamt 220 Beobachtungen statt. Bei jeder Beobachtung erhielten die Patienten Scores für jedes von 16 Symptomen. Die Symptomwerte fehlen für Patient 71 zum Zeitpunkt 2, Patient 76 zum Zeitpunkt 2 und Patient 47 zum Zeitpunkt 3, wodurch 217 gültige Beobachtungen verbleiben.

- **autoaccidents.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um die Bemühungen eines Versicherungsanalysten geht, ein Modell zur Anzahl der Autounfälle pro Fahrer unter Berücksichtigung von Alter und Geschlecht zu erstellen. Jeder Fall stellt einen Fahrer dar und erfasst das Geschlecht des Fahrers, sein Alter in Jahren und die Anzahl der Autounfälle in den letzten fünf Jahren.
- **band.sav.** Diese Datendatei enthält die hypothetischen wöchentlichen Verkaufszahlen von CDs für eine Musikgruppe. Daten für drei mögliche Einflussvariablen wurden ebenfalls aufgenommen.
- **bankloan.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um die Bemühungen einer Bank geht, den Anteil der nicht zurückgezahlten Kredite zu reduzieren. Die Datei enthält Informationen zum Finanzstatus und demografischen Hintergrund von 850 früheren und potenziellen Kunden. Bei den ersten 700 Fällen handelt es sich um Kunden, denen bereits ein Kredit gewährt wurde. Bei den letzten 150 Fällen handelt es sich um potenzielle Kunden, deren Kreditrisiko die Bank als gering oder hoch einstufen möchte.
- **bankloan_binning.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, die Informationen zum Finanzstatus und demografischen Hintergrund von 5.000 früheren Kunden enthält.
- **behavior.sav.** In einem klassischen Beispiel () wurden 52 Schüler/Studenten gebeten, die Kombinationen aus 15 Situationen und 15 Verhaltensweisen auf einer 10-Punkte-Skala von 0 = “ausgesprochen angemessen” bis 9 = “ausgesprochen unangemessen” zu bewerten. Die Werte werden über die einzelnen Personen gemittelt und als Unähnlichkeiten verwendet.
- **behavior_ini.sav.** Diese Datendatei enthält eine Ausgangskonfiguration für eine zweidimensionale Lösung für *behavior.sav*.
- **brakes.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um die Qualitätskontrolle in einer Fabrik geht, die Scheibenbremsen für Hochleistungsautomobile herstellt. Die Datendatei enthält Messungen des Durchmessers von 16 Scheiben aus 8 Produktionsmaschinen. Der Zieldurchmesser für die Scheiben ist 322 Millimeter.
- **breakfast.sav.** In einer klassischen Studie () wurden 21 MBA-Studenten der Wharton School mit ihren Lebensgefährten darum gebeten, 15 Frühstücksartikel in der Vorzugsreihenfolge von 1 = “am meisten bevorzugt” bis 15 = “am wenigsten bevorzugt” zu ordnen. Die Bevorzugungen wurden in sechs unterschiedlichen Szenarien erfasst, von “Overall preference” (Allgemein bevorzugt) bis “Snack, with beverage only” (Imbiss, nur mit Getränk).
- **breakfast-overall.sav.** Diese Datei enthält die Daten zu den bevorzugten Frühstücksartikeln, allerdings nur für das erste Szenario, “Overall preference” (Allgemein bevorzugt).
- **broadband_1.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, die die Anzahl der Abonnenten eines Breitband-Service, nach Region geordnet, enthält. Die Datendatei enthält die monatlichen Abonnentenzahlen für 85 Regionen über einen Zeitraum von vier Jahren.
- **broadband_2.sav.** Diese Datendatei stimmt mit *broadband_1.sav* überein, enthält jedoch Daten für weitere drei Monate.
- **car_insurance_claims.sav.** Ein an anderer Stelle () vorgestelltes und analysiertes Daten-Set bezieht sich auf Schadensansprüche für Autos. Die durchschnittliche Höhe der Schadensansprüche lässt sich mit Gamma-Verteilung modellieren. Dazu wird eine inverse Verknüpfungsfunktion verwendet, um den Mittelwert der abhängigen Variablen mit einer linearen Kombination aus Alter des Versicherungsnehmers, Fahrzeugtyp und

Fahrzeugalter in Bezug zu setzen. Die Anzahl der eingereichten Schadensansprüche kann als Skalierungsgewicht verwendet werden.

- **car_sales.sav.** Diese Datendatei enthält hypothetische Verkaufsschätzer, Listenpreise und physische Spezifikationen für verschiedene Fahrzeugfabrikate und -modelle. Die Listenpreise und physischen Spezifikationen wurden von *edmunds.com* und Hersteller-Websites entnommen.
- **carpet.sav.** In einem beliebten Beispiel möchte einen neuen Teppichreiniger vermarkten und dazu den Einfluss von fünf Faktoren auf die Bevorzugung durch den Verbraucher untersuchen: Verpackungsgestaltung, Markenname, Preis, Gütesiegel, *Good Housekeeping* und Geld-zurück-Garantie. Die Verpackungsgestaltung setzt sich aus drei Faktorenebenen zusammen, die sich durch die Position der Auftragebürste unterscheiden. Außerdem gibt es drei Markennamen (*K2R*, *Glory* und *Bissell*), drei Preisstufen sowie je zwei Ebenen (Nein oder Ja) für die letzten beiden Faktoren. 10 Kunden stufen 22 Profile ein, die durch diese Faktoren definiert sind. Die Variable *Preference* enthält den Rang der durchschnittlichen Einstufung für die verschiedenen Profile. Ein niedriger Rang bedeutet eine starke Bevorzugung. Diese Variable gibt ein Gesamtmaß der Bevorzugung für die Profile an.
- **carpet_prefs.sav.** Diese Datendatei beruht auf denselben Beispielen, wie für *carpet.sav* beschrieben, enthält jedoch die tatsächlichen Einstufungen durch jeden der 10 Kunden. Die Kunden wurden gebeten, die 22 Produktprofile in der Reihenfolge ihrer Präferenzen einzustufen. Die Variablen *PREF1* bis *PREF22* enthalten die IDs der zugeordneten Profile, wie in *carpet_plan.sav* definiert.
- **catalog.sav.** Diese Datendatei enthält hypothetische monatliche Verkaufszahlen für drei Produkte, die von einem Versandhaus verkauft werden. Daten für fünf mögliche Einflussvariablen wurden ebenfalls aufgenommen.
- **catalog_seasfac.sav.** Diese Datendatei ist mit *catalog.sav* identisch, außer, dass ein Set von saisonalen Faktoren, die mithilfe der Prozedur “Saisonale Zerlegung” berechnet wurden, sowie die zugehörigen Datumsvariablen hinzugefügt wurden.
- **cellular.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um die Bemühungen eines Mobiltelefonunternehmens geht, die Kundenabwanderung zu verringern. Scores für die Abwanderungsneigung (von 0 bis 100) werden auf die Kunden angewendet. Kunden mit einem Score von 50 oder höher streben vermutlich einen Anbieterwechsel an.
- **ceramics.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um die Bemühungen eines Herstellers geht, der ermitteln möchte, ob ein neue, hochwertige Keramiklegierung eine größere Hitzebeständigkeit aufweist als eine Standardlegierung. Jeder Fall entspricht einem Test einer der Legierungen; die Temperatur, bei der das Keramikwälzlager versagte, wurde erfasst.
- **cereal.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um eine Umfrage geht, bei der 880 Personen nach ihren Frühstückgewohnheiten befragt wurden. Außerdem wurden Alter, Geschlecht, Familienstand und Vorliegen bzw. Nichtvorliegen eines aktiven Lebensstils (auf der Grundlage von mindestens zwei Trainingseinheiten pro Woche) erfasst. Jeder Fall entspricht einem Teilnehmer.
- **clothing_defects.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um die Qualitätskontrolle in einer Bekleidungsfabrik geht. Aus jeder in der Fabrik produzierten Charge entnehmen die Kontrolleure eine Stichprobe an Bekleidungsartikeln und zählen die Anzahl der Bekleidungsartikel die inakzeptabel sind.

- **coffee.sav.** Diese Datendatei enthält Daten zum wahrgenommenen Image von sechs Eiskaffeemarken (). Bei den 23 Attributen des Eiskaffee-Image sollten die Teilnehmer jeweils alle Marken auswählen, die durch dieses Attribut beschrieben werden. Die sechs Marken werden als “AA”, “BB”, “CC”, “DD”, “EE” und “FF” bezeichnet, um Vertraulichkeit zu gewährleisten.
- **contacts.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um die Kontaktlisten einer Gruppe von Vertretern geht, die Computer an Unternehmen verkaufen. Die einzelnen Kontaktpersonen werden anhand der Abteilung, in der sie in ihrem Unternehmen arbeiten und anhand ihrer Stellung in der Unternehmenshierarchie in Kategorien eingeteilt. Außerdem werden der Betrag des letzten Verkaufs, die Zeit seit dem letzten Verkauf und die Größe des Unternehmens, in dem die Kontaktperson arbeitet, aufgezeichnet.
- **creditpromo.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um die Bemühungen eines Kaufhauses geht, die Wirksamkeit einer kürzlich durchgeführten Kreditkarten-Werbeaktion einzuschätzen. Dazu wurden 500 Karteninhaber nach dem Zufallsprinzip ausgewählt. Die Hälfte erhielt eine Werbebeilage, die einen reduzierten Zinssatz für Einkäufe in den nächsten drei Monaten ankündigte. Die andere Hälfte erhielt eine Standard-Werbebeilage.
- **customer_dbase.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um die Bemühungen eines Unternehmens geht, das die Informationen in seinem Data Warehouse nutzen möchte, um spezielle Angebote für Kunden zu erstellen, die mit der größten Wahrscheinlichkeit darauf ansprechen. Nach dem Zufallsprinzip wurde eine Untergruppe des Kundenstamms ausgewählt. Diese Gruppe erhielt die speziellen Angebote und die Reaktionen wurden aufgezeichnet.
- **customers_model.sav.** Diese Datei enthält hypothetische Daten zu Einzelpersonen, auf die sich eine Marketingkampagne richtete. Zu diesen Daten gehören demografische Informationen, eine Übersicht über die bisherigen Einkäufe und die Angabe ob die einzelnen Personen auf die Kampagne ansprachen oder nicht. Jeder Fall entspricht einer Einzelperson.
- **customers_new.sav.** Diese Datei enthält hypothetische Daten zu Einzelpersonen, die potenzielle Kandidaten für Marketingkampagnen sind. Zu diesen Daten gehören demografische Informationen und eine Übersicht über die bisherigen Einkäufe für jede Person. Jeder Fall entspricht einer Einzelperson.
- **debate.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, die gepaarte Antworten auf eine Umfrage unter den Zuhörern einer politischen Debatte enthält (Antworten vor und nach der Debatte). Jeder Fall entspricht einem Befragten.
- **debate_aggregate.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, in der die Antworten aus *debate.sav* aggregiert wurden. Jeder Fall entspricht einer Kreuzklassifikation der bevorzugten Politiker vor und nach der Debatte.
- **demo.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um eine Kundendatenbank geht, die zum Zwecke der Zusendung monatlicher Angebote erworben wurde. Neben verschiedenen demografischen Informationen ist erfasst, ob der Kunde auf das Angebot geantwortet hat.
- **demo_cs_1.sav.** Hierbei handelt es sich um eine hypothetische Datendatei für den ersten Schritt eines Unternehmens, das eine Datenbank mit Umfrageinformationen zusammenstellen möchte. Jeder Fall entspricht einer anderen Stadt. Außerdem sind IDs für Region, Provinz, Landkreis und Stadt erfasst.

- **demo_cs_2.sav.** Hierbei handelt es sich um eine hypothetische Datendatei für den zweiten Schritt eines Unternehmens, das eine Datenbank mit Umfrageinformationen zusammenstellen möchte. Jeder Fall entspricht einem anderen Stadtteil aus den im ersten Schritt ausgewählten Städten. Außerdem sind IDs für Region, Provinz, Landkreis, Stadt, Stadtteil und Wohneinheit erfasst. Die Informationen zur Stichprobenziehung aus den ersten beiden Stufen des Stichprobenplans sind ebenfalls enthalten.
- **demo_cs.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, die Umfrageinformationen enthält die mit einem komplexen Stichprobenplan erfasst wurden. Jeder Fall entspricht einer anderen Wohneinheit. Es sind verschiedene Informationen zum demografischen Hintergrund und zur Stichprobenziehung erfasst.
- **dietstudy.sav.** Diese hypothetische Datendatei enthält die Ergebnisse einer Studie der "Stillman-Diät" . Jeder Fall entspricht einem Teilnehmer und enthält dessen Gewicht vor und nach der Diät in amerikanischen Pfund sowie mehrere Messungen des Triglyceridspiegels (in mg/100 ml).
- **dischargedata.sav.** Hierbei handelt es sich um eine Datendatei zum Thema *Seasonal Patterns of Winnipeg Hospital Use*, (Saisonale Muster der Belegung im Krankenhaus von Winnipeg) vom Manitoba Centre for Health Policy.
- **dvdplayer.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um die Entwicklung eines neuen DVD-Spielers geht. Mithilfe eines Prototyps hat das Marketing-Team Zielgruppendaten erfasst. Jeder Fall entspricht einem befragten Benutzer und enthält demografische Daten zu dem Benutzer sowie dessen Antworten auf Fragen zum Prototyp.
- **flying.sav.** Diese Datendatei enthält die Flugmeilen zwischen zehn Städten in den USA.
- **german_credit.sav.** Diese Daten sind aus dem Daten-Set "German credit" im Repository of Machine Learning Databases () an der Universität von Kalifornien in Irvine entnommen.
- **grocery_1month.sav.** Bei dieser hypothetischen Datendatei handelt es sich um die Datendatei *grocery_coupons.sav*, wobei die wöchentlichen Einkäufe zusammengefasst sind, sodass jeder Fall einem anderen Kunden entspricht. Dadurch entfallen einige der Variablen, die wöchentlichen Änderungen unterworfen waren, und der verzeichnete ausgegebene Betrag ist nun die Summe der Beträge, die in den vier Wochen der Studie ausgegeben wurden.
- **grocery_coupons.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, die Umfragedaten enthält, die von einer Lebensmittelkette erfasst wurden, die sich für die Kaufgewohnheiten ihrer Kunden interessiert. Jeder Kunde wird über vier Wochen beobachtet, und jeder Fall entspricht einer Kundenwoche und enthält Informationen zu den Geschäften, in denen der Kunde einkauft sowie zu anderen Merkmalen, beispielsweise welcher Betrag in der betreffenden Woche für Lebensmittel ausgegeben wurde.
- **guttman.sav.** Bell () legte eine Tabelle zur Darstellung möglicher sozialer Gruppen vor. Guttman () verwendete einen Teil dieser Tabelle, bei der fünf Variablen, die Aspekte beschreiben, wie soziale Interaktion, das Gefühl der Gruppenzugehörigkeit, die physische Nähe der Mitglieder und die Formalität der Beziehung, mit sieben theoretischen sozialen Gruppen gekreuzt wurden: "crowds" (Menschenmassen, beispielsweise die Zuschauer eines Fußballspiels), "audience" (Zuhörerschaften, beispielsweise die Personen im Theater oder bei einer Vorlesung), "public" (Öffentlichkeit, beispielsweise Zeitungsleser oder Fernsehzuschauer), "mobs" (Mobs, wie Menschenmassen, jedoch mit wesentlich stärkerer Interaktion), "primary groups" (Primärgruppen, vertraulich), "secondary groups"

(Sekundärgruppen, freiwillig) und “modern community” (die moderne Gesellschaft, ein lockerer Zusammenschluss, der aus einer engen physischen Nähe und dem Bedarf an spezialisierten Dienstleistungen entsteht).

- **healthplans.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um die Bemühungen einer Versicherungsgruppe geht, vier verschiedene Pläne zur Gesundheitsvorsorge für Kleinbetriebe zu evaluieren. Zwölf Inhaber von Kleinbetrieben (Arbeitgeber) wurden gebeten, die Pläne danach in eine Rangfolge zu bringen, wie gern sie sie ihren Mitarbeitern anbieten würden. Jeder Fall entspricht einem Arbeitgeber und enthält die Reaktionen auf die einzelnen Pläne.
- **health_funding.sav.** Hierbei handelt es sich um eine hypothetische Datei, die Daten zur Finanzierung des Gesundheitswesens (Betrag pro 100 Personen), Krankheitsraten (Rate pro 10.000 Personen der Bevölkerung) und Besuche bei medizinischen Einrichtungen/Ärzten (Rate pro 10.000 Personen der Bevölkerung) enthält. Jeder Fall entspricht einer anderen Stadt.
- **hivassay.sav.** Hierbei handelt es sich um eine hypothetische Datendatei zu den Bemühungen eines pharmazeutischen Labors, einen Schnelltest zur Erkennung von HIV-Infektionen zu entwickeln. Die Ergebnisse des Tests sind acht kräftiger werdende Rotschattierungen, wobei kräftigeren Schattierungen auf eine höhere Infektionswahrscheinlichkeit hindeuten. Bei 2.000 Blutproben, von denen die Hälfte mit HIV infiziert war, wurde ein Labortest durchgeführt.
- **hourlywagedata.sav.** Hierbei handelt es sich um eine hypothetische Datendatei zum Stundenlohn von Pflegepersonal in Praxen und Krankenhäusern mit unterschiedlich langer Berufserfahrung.
- **insure.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um eine Versicherungsgesellschaft geht, die die Risikofaktoren untersucht, die darauf hinweisen, ob ein Kunde die Leistungen einer mit einer Laufzeit von 10 Jahren abgeschlossenen Lebensversicherung in Anspruch nehmen wird. Jeder Fall in der Datendatei entspricht einem Paar von Verträgen, je einer mit Leistungsforderung und der andere ohne, wobei die beiden Versicherungsnehmer in Alter und Geschlecht übereinstimmen.
- **judges.sav.** Hierbei handelt es sich um eine hypothetische Datendatei mit den Wertungen von ausgebildeten Kampfrichtern (sowie eines Sportliebhabers) zu 300 Kunstturnleistungen. Jede Zeile stellt eine Leistung dar; die Kampfrichter bewerteten jeweils dieselben Leistungen.
- **kinship_dat.sav.** Rosenberg und Kim () haben 15 Bezeichnungen für den Verwandtschaftsgrad untersucht (Tante, Bruder, Cousin, Tochter, Vater, Enkelin, Großvater, Großmutter, Enkel, Mutter, Neffe, Nichte, Schwester, Sohn, Onkel). Die beiden Analytiker baten vier Gruppen von College-Studenten (zwei weibliche und zwei männliche Gruppen), diese Bezeichnungen auf der Grundlage der Ähnlichkeiten zu sortieren. Zwei Gruppen (eine weibliche und eine männliche Gruppe) wurden gebeten, die Bezeichnungen zweimal zu sortieren; die zweite Sortierung sollte dabei nach einem anderen Kriterium erfolgen als die erste. So wurden insgesamt sechs “Quellen” erzielt. Jede Quelle entspricht einer Ähnlichkeitsmatrix mit 15×15 Elementen. Die Anzahl der Zellen ist dabei gleich der Anzahl der Personen in einer Quelle minus der Anzahl der gemeinsamen Platzierungen der Objekte in dieser Quelle.
- **kinship_ini.sav.** Diese Datendatei enthält eine Ausgangskonfiguration für eine dreidimensionale Lösung für *kinship_dat.sav*.
- **kinship_var.sav.** Diese Datendatei enthält die unabhängigen Variablen *gender* (Geschlecht), *gener*(Generation) und *degree* (Verwandtschaftsgrad), die zur Interpretation der Dimensionen einer Lösung für *kinship_dat.sav* verwendet werden können. Insbesondere können sie

verwendet werden, um den Lösungsraum auf eine lineare Kombination dieser Variablen zu beschränken.

- **mailresponse.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, in der es um die Bemühungen eines Bekleidungsherstellers geht, der ermitteln möchte, ob die Verwendung von Briefsendungen für das Direktmarketing zu schnelleren Antworten führt als Postwurfsendungen. Die Mitarbeiter in der Bestellannahme erfassen, wie vielen Wochen nach der Postsendung die einzelnen Bestellungen aufgegeben wurden.
- **marketvalues.sav.** Diese Datendatei betrifft Hausverkäufe in einem Neubaugebiet in Algonquin, Illinois, in den Jahren 1999–2000. Diese Verkäufe sind in Grundbucheinträgen dokumentiert.
- **mutualfund.sav.** Diese Datendatei betrifft Aktienmarktdaten für verschiedene Technologieaktien, die in im Index S&P 500 verzeichnet sind. Jeder Fall entspricht einem Unternehmen.
- **nhis2000_subset.sav.** Die "National Health Interview Survey (NHIS)" ist eine große, bevölkerungsbezogene Umfrage in unter der US-amerikanischen Zivilbevölkerung. Es werden persönliche Interviews in einer landesweit repräsentativen Stichprobe von Haushalten durchgeführt. Für die Mitglieder jedes Haushalts werden demografische Informationen und Beobachtungen zum Gesundheitsverhalten und Gesundheitsstatus eingeholt. Diese Datendatei enthält eine Teilmenge der Informationen aus der Umfrage des Jahres 2000. National Center for Health Statistics. National Health Interview Survey, 2000. Datendatei und Dokumentation öffentlich zugänglich. *ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Datasets/NHIS/2000/*. Zugriff erfolgte 2003.
- **ozone.sav.** Die Daten enthalten 330 Beobachtungen zu sechs meteorologischen Variablen zur Vorhersage der Ozonkonzentration aus den übrigen Variablen. Bei früheren Untersuchungen (,) fanden Wissenschaftler einige Nichtlinearitäten unter diesen Variablen, die die Standardverfahren bei der Regression behindern.
- **pain_medication.sav.** Diese hypothetische Datendatei enthält die Ergebnisse eines klinischen Tests für ein entzündungshemmendes Medikament zur Schmerzbehandlung bei chronischer Arthritis. Von besonderem Interesse ist die Zeitdauer, bis die Wirkung des Medikaments einsetzt und wie es im Vergleich mit bestehenden Medikamenten abschneidet.
- **patient_los.sav.** Diese hypothetische Datendatei enthält die Behandlungsaufzeichnungen zu Patienten, die wegen des Verdachts auf Herzinfarkt in das Krankenhaus eingeliefert wurden. Jeder Fall entspricht einem Patienten und enthält diverse Variablen in Bezug auf den Krankenhausaufenthalt.
- **patlos_sample.sav.** Diese hypothetische Datendatei enthält die Behandlungsaufzeichnungen für eine Stichprobe von Patienten, denen während der Behandlung eines Herzinfarkts Thrombolytika verabreicht wurden. Jeder Fall entspricht einem Patienten und enthält diverse Variablen in Bezug auf den Krankenhausaufenthalt.
- **polishing.sav.** Hierbei handelt es sich um die Datendatei "Nambeware Polishing Times" aus der Data and Story Library. Sie bezieht sich auf die Bemühungen eines Herstellers von Metallgeschirr (Nambe Mills, Santa Fe, New Mexico) zur zeitlichen Planung seiner Produktion. Jeder Fall entspricht einem anderen Artikel in der Produktpalette. Für jeden Artikel sind Durchmesser, Polierzeit, Preis und Produkttyp erfasst.

- **poll_cs.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um Bemühungen geht, die öffentliche Unterstützung für einen Gesetzentwurf zu ermitteln, bevor er im Parlament eingebracht wird. Die Fälle entsprechen registrierten Wählern. Für jeden Fall sind County, Gemeinde und Wohnviertel des Wählers erfasst.
- **poll_cs_sample.sav.** Diese hypothetische Datendatei enthält eine Stichprobe der in *poll_cs.sav* aufgeführten Wähler. Die Stichprobe wurde gemäß dem in der Plandatei *poll.csplan* angegebenen Stichprobenplan gezogen und in dieser Datendatei sind die Einschlusswahrscheinlichkeiten und Stichprobengewichtungen erfasst. Beachten Sie jedoch Folgendes: Da im Stichprobenplan die PPS-Methode (PPS: probability proportional to size; Wahrscheinlichkeit proportional zur Größe) verwendet wird, gibt es außerdem eine Datei mit den gemeinsamen Auswahlwahrscheinlichkeiten (*poll_jointprob.sav*). Die zusätzlichen Variablen zum demografischen Hintergrund der Wähler und ihrer Meinung zum vorgeschlagenen Gesetzentwurf wurden nach der Ziehung der Stichprobe erfasst und zur Datendatei hinzugefügt.
- **property_assess.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, in der es um die Bemühungen eines für einen Bezirk (County) zuständigen Immobilienbewerters geht, trotz eingeschränkter Ressourcen die Einschätzungen des Werts von Immobilien auf dem aktuellsten Stand zu halten. Die Fälle entsprechen den Immobilien, die im vergangenen Jahr in dem betreffenden County verkauft wurden. Jeder Fall in der Datendatei enthält die Gemeinde, in der sich die Immobilie befindet, den Bewerter, der die Immobilie besichtigt hat, die seit dieser Bewertung verstrichene Zeit, den zu diesem Zeitpunkt ermittelten Wert sowie den Verkaufswert der Immobilie.
- **property_assess_cs.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, in der es um die Bemühungen eines für einen US-Bundesstaat zuständigen Immobilienbewerters geht, trotz eingeschränkter Ressourcen die Einschätzungen des Werts von Immobilien auf dem aktuellsten Stand zu halten. Die Fälle entsprechen den Immobilien in dem betreffenden Bundesstaat. Jeder Fall in der Datendatei enthält das County, die Gemeinde und das Wohnviertel, in dem sich die Immobilie befindet, die seit der letzten Bewertung verstrichene Zeit sowie zu diesem Zeitpunkt ermittelten Wert.
- **property_assess_cs_sample.sav.** Diese hypothetische Datendatei enthält eine Stichprobe der in *property_assess_cs.sav* aufgeführten Immobilien. Die Stichprobe wurde gemäß dem in der Plandatei *property_assess.csplan* angegebenen Stichprobenplan gezogen und in dieser Datendatei sind die Einschlusswahrscheinlichkeiten und Stichprobengewichtungen erfasst. Die zusätzliche Variable *Current value* (Aktueller Wert) wurde nach der Ziehung der Stichprobe erfasst und zur Datendatei hinzugefügt.
- **recidivism.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um die Bemühungen einer Strafverfolgungsbehörde geht, einen Einblick in die Rückfallraten in ihrem Zuständigkeitsbereich zu gewinnen. Jeder Fall entspricht einem frühren Straftäter und erfasst Daten zu dessen demografischen Hintergrund, einige Details zu seinem ersten Verbrechen sowie die Zeit bis zu seiner zweiten Festnahme, sofern diese innerhalb von zwei Jahren nach der ersten Festnahme erfolgte.
- **recidivism_cs_sample.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um die Bemühungen einer Strafverfolgungsbehörde geht, einen Einblick in die Rückfallraten in ihrem Zuständigkeitsbereich zu gewinnen. Jeder Fall entspricht einem früheren Straftäter, der im Juni 2003 erstmals aus der Haft entlassen wurde, und erfasst Daten zu dessen demografischen Hintergrund, einige Details zu seinem ersten Verbrechen sowie die

Daten zu seiner zweiten Festnahme, sofern diese bis Ende Juni 2006 erfolgte. Die Straftäter wurden aus per Stichprobenziehung ermittelten Polizeidirektionen ausgewählt (gemäß dem in *recidivism_cs.csplan* angegebenen Stichprobenplan). Da hierbei eine PPS-Methode (PPS: probability proportional to size; Wahrscheinlichkeit proportional zur Größe) verwendet wird, gibt es außerdem eine Datei mit den gemeinsamen Auswahlwahrscheinlichkeiten (*recidivism_cs_jointprob.sav*).

- **salesperformance.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um Bewertung von zwei neuen Verkaufsschulungen geht. 60 Mitarbeiter, die in drei Gruppen unterteilt sind, erhalten jeweils eine Standardschulung. Zusätzlich erhält Gruppe 2 eine technische Schulung und Gruppe 3 eine Praxisschulung. Die einzelnen Mitarbeiter wurden am Ende der Schulung einem Test unterzogen und die erzielten Punkte wurden erfasst. Jeder Fall in der Datendatei stellt einen Lehrgangsteilnehmer dar und enthält die Gruppe, der der Lehrgangsteilnehmer zugeteilt wurde sowie die von ihm in der Prüfung erreichte Punktzahl.
- **satisf.sav.** Hierbei handelt es sich um eine hypothetische Datendatei zu einer Zufriedenheitsumfrage, die von einem Einzelhandelsunternehmen in 4 Filialen durchgeführt wurde. Insgesamt wurden 582 Kunden befragt. Jeder Fall gibt die Antworten eines einzelnen Kunden wieder.
- **screws.sav.** Diese Datendatei enthält Informationen zu den Eigenschaften von Schrauben, Bolzen, Muttern und Reißnägeln ().
- **shampoo_ph.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um die Qualitätskontrolle in einer Fabrik für Haarpflegeprodukte geht. In regelmäßigen Zeitabständen werden Messwerte von sechs separaten Ausgangschargen erhoben und ihr pH-Wert erfasst. Der Zielbereich ist 4,5–5,5.
- **ships.sav.** Ein an anderer Stelle () vorgestelltes und analysiertes Daten-Set bezieht sich auf die durch Wellen verursachten Schäden an Frachtschiffen. Die Vorfallshäufigkeiten können unter Angabe von Schiffstyp, Konstruktionszeitraum und Betriebszeitraum gemäß einer Poisson-Rate modelliert werden. Das Aggregat der Betriebsmonate für jede Zelle der durch die Kreuzklassifizierung der Faktoren gebildeten Tabelle gibt die Werte für die Risikoanfälligkeit an.
- **site.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um die Bemühungen eines Unternehmens geht, neue Standorte für die betriebliche Expansion auszuwählen. Das Unternehmen beauftragte zwei Berater unabhängig voneinander mit der Bewertung der Standorte. Neben einem umfassenden Bericht gaben die Berater auch eine zusammenfassende Wertung für jeden Standort als “good” (gut) “fair” (mittelmäßig) oder “poor” (schlecht) ab.
- **siteratings.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um die Betatests der neuen Website eines E-Commerce-Unternehmens geht. Jeder Fall entspricht einem Beta-Tester, der die Brauchbarkeit der Website auf einer Skala von 0 bis 20 bewertete.
- **smokers.sav.** Diese Datendatei wurde aus der Umfrage “National Household Survey of Drug Abuse” aus dem Jahr 1998 abstrahiert und stellt eine Wahrscheinlichkeitsstichprobe US-amerikanischer Haushalte dar. Daher sollte der erste Schritt bei der Analyse dieser Datendatei darin bestehen, die Daten entsprechend den Bevölkerungstrends zu gewichten.
- **smoking.sav.** Hierbei handelt es sich um eine von Greenacre () vorgestellte hypothetische Tabelle. Die relevante Tabelle wird durch eine Kreuztabelle der Rauchgewohnheiten und der Berufskategorie gebildet. Die Variable *Berufsgruppe* enthält die Berufskategorien *Senior*

Manager, *Junior Manager*, *Angestellter mit Erfahrung*, *Angestellter ohne Erfahrung* und *Sekretariat* sowie die Kategorie *National Average*, die als Ergänzung der Analyse dienen kann. Die Variable *Rauchen* enthält die Rauchgewohnheiten *Nichtraucher*, *Leicht*, *Mittel* und *Stark* sowie die Kategorien *No Alcohol* und *Alcohol*, die als Ergänzung der Analyse dienen können.

- **storebrand.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um die Bemühungen einer Verkaufsleiterin in einem Lebensmittelmarkt geht, die die Verkaufszahlen des Waschmittels der Eigenmarke gegenüber den anderen Marken steigern möchte. Sie erarbeitet eine Werbeaktion im Geschäft und spricht an der Kasse mit Kunden. Jeder Fall entspricht einem Kunden.
- **stores.sav.** Diese Datendatei enthält hypothetische monatliche Marktanteilsdaten für zwei konkurrierende Lebensmittelgeschäfte. Jeder Fall entspricht den Marktanteilsdaten für einen bestimmten Monat.
- **stroke_clean.sav.** Diese hypothetische Datendatei enthält den Zustand einer medizinischen Datenbank, nachdem diese mithilfe der Prozeduren in der Option “Data Preparation” bereinigt wurde.
- **stroke_invalid.sav.** Diese hypothetische Datendatei enthält den ursprünglichen Zustand einer medizinischen Datenbank, der mehrere Dateneingabefehler aufweist.
- **stroke_survival.** In dieser hypothetischen Datendatei geht es um die Überlebenszeiten von Patienten, die nach einem Rehabilitationsprogramm wegen eines ischämischen Schlaganfalls mit einer Reihe von Problemen zu kämpfen haben. Nach dem Schlaganfall werden das Auftreten von Herzinfarkt, ischämischem Schlaganfall und hämorrhagischem Schlaganfall sowie der Zeitpunkt des Ereignisses aufgezeichnet. Die Stichprobe ist auf der linken Seite abgeschnitten, da sie nur Patienten enthält, die bis zum Ende des Rehabilitationprogramms, das nach dem Schlaganfall durchgeführt wurde, überlebten.
- **stroke_valid.sav.** Diese hypothetische Datendatei enthält den Zustand einer medizinischen Datenbank, nachdem diese mithilfe der Prozedur “Daten validieren” überprüft wurde. Sie enthält immer noch potenziell anomale Fälle.
- **tastetest.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um Bewertung der Auswirkungen der Mulchfarbe auf den Geschmack von Pflanzenprodukten geht. Der Geschmack von Erdbeeren, die in rotem, blauem und schwarzem Rindenmulch gezogen wurden, wurde von Testpersonen auf einer ordinalen Skala (weit unter bis weit über dem Durchschnitt) bewertet. Jeder Fall entspricht einem Geschmackstester.
- **telco.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um die Bemühungen eines Telekommunikationsunternehmens geht, die Kundenabwanderung zu verringern. Jeder Fall entspricht einem Kunden und enthält verschiedene Informationen zum demografischen Hintergrund und zur Servicenutzung.
- **telco_extra.sav.** Diese Datendatei ähnelt der Datei *telco.sav*, allerdings wurden die Variablen “tenure” und die Log-transformierten Variablen zu den Kundenausgaben entfernt und durch standardisierte Log-transformierte Variablen ersetzt.
- **telco_missing.sav.** Diese Datendatei entspricht der Datei *telco_mva_complete.sav*, allerdings wurde ein Teil der Daten durch fehlende Werte ersetzt.
- **telco_mva_complete.sav.** Bei dieser Datendatei handelt es sich um eine Teilmenge der Datendatei *telco.sav*, allerdings mit anderen Variablennamen.

- **testmarket.sav.** Diese hypothetische Datendatei bezieht sich auf die Pläne einer Fast-Food-Kette, einen neuen Artikel in ihr Menü aufzunehmen. Es gibt drei mögliche Kampagnen zur Verkaufsförderung für das neue Produkt. Daher wird der neue Artikel in Filialen in mehreren zufällig ausgewählten Märkten eingeführt. An jedem Standort wird eine andere Form der Verkaufsförderung verwendet und die wöchentlichen Verkaufszahlen für das neue Produkt werden für die ersten vier Wochen aufgezeichnet. Jeder Fall entspricht einer Standort-Woche.
- **testmarket_1month.sav.** Bei dieser hypothetischen Datendatei handelt es sich um die Datendatei *testmarket.sav*, wobei die wöchentlichen Verkaufszahlen zusammengefasst sind, sodass jeder Fall einem Standort entspricht. Dadurch entfallen einige der Variablen, die wöchentlichen Änderungen unterworfen waren, und die verzeichneten Verkaufszahlen sind nun die Summe der Verkaufszahlen während der vier Wochen der Studie.
- **tree_car.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, die demografische Daten sowie Daten zum Kaufpreis von Fahrzeugen enthält.
- **tree_credit.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, die demografische Daten sowie Daten zu früheren Bankkrediten enthält.
- **tree_missing_data.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, die demografische Daten sowie Daten zu früheren Bankkrediten enthält und eine große Anzahl fehlender Werte aufweist.
- **tree_score_car.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, die demografische Daten sowie Daten zum Kaufpreis von Fahrzeugen enthält.
- **tree_textdata.sav.** Eine einfache Datendatei mit nur zwei Variablen, die vor allem den Standardzustand von Variablen vor der Zuweisung von Messniveau und Wertelabels zeigen soll.
- **tv-survey.sav.** Hierbei handelt es sich um eine hypothetische Datendatei zu einer Studie, die von einem Fernsehstudio durchgeführt wurde, das überlegt, ob die Laufzeit eines erfolgreichen Programms verlängert werden soll. 906 Personen wurden gefragt, ob sie das Programm unter verschiedenen Bedingungen ansehen würden. Jede Zeile entspricht einem Befragten; jede Spalte entspricht einer Bedingung.
- **ulcer_recurrence.sav.** Diese Datei enthält Teilinformationen aus einer Studie zum Vergleich der Wirksamkeit zweier Therapien zur Vermeidung des Wiederauftretens von Geschwüren. Es stellt ein gutes Beispiel für intervallzensierte Daten dar und wurde an anderer Stelle () vorgestellt und analysiert.
- **ulcer_recurrence_recoded.sav.** In dieser Datei sind die Daten aus *ulcer_recurrence.sav* so umstrukturiert, dass das Modell der Ereigniswahrscheinlichkeit für jedes Intervall der Studie berechnet werden kann und nicht nur die Ereigniswahrscheinlichkeit am Ende der Studie. Sie wurde an anderer Stelle () vorgestellt und analysiert.
- **verd1985.sav.** Diese Datendatei enthält eine Umfrage (). Die Antworten von 15 Subjekten auf 8 Variablen wurden aufgezeichnet. Die relevanten Variablen sind in drei Sets unterteilt. Set 1 umfasst *alter* und *heirat*, Set 2 besteht aus *pet* und *news* und in Set 3 finden sich *music* und *live*. Die Variable *pet* wird mehrfach nominal skaliert und die Variable *Alter* ordinal. Alle anderen Variablen werden einzeln nominal skaliert.
- **virus.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um die Bemühungen eines Internet-Dienstanbieters geht, der die Auswirkungen eines Virus auf seine Netzwerke ermitteln möchte. Dabei wurde vom Moment der Virusentdeckung bis zu dem

Zeitpunkt, zu dem die Virusinfektion unter Kontrolle war, der (ungefähre) prozentuale Anteil infizierter E-Mail in den Netzwerken erfasst.

- **waittimes.sav.** Hierbei handelt es sich um eine hypothetische Datendatei zu den Wartezeiten für Kunden bei drei verschiedenen Filialen einer Bank. Jeder Fall entspricht einem Kunden und zeichnet die Wartezeit und die Filiale.
- **webusability.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um Tests zur Benutzerfreundlichkeit eines neuen Internetgeschäfts geht. Jeder Fall entspricht einer von fünf Testpersonen, die die Benutzerfreundlichkeit bewerten und gibt für sechs separate Aufgaben an, ob die Testperson sie erfolgreich ausführen könnte.
- **wheeze_steubenville.sav.** Hierbei handelt es sich um eine Teilmenge der Daten aus einer Langzeitstudie zu den gesundheitlichen Auswirkungen der Luftverschmutzung auf Kinder (). Die Daten enthalten wiederholte binäre Messungen des Keuchens von Kindern aus Steubenville, Ohio, im Alter von 7, 8, 9 und 10 Jahren sowie eine unveränderlichen Angabe, ob die Mutter im ersten Jahr der Studie rauchte oder nicht.
- **workprog.sav.** Hierbei handelt es sich um eine hypothetische Datendatei zu einem Arbeitsprogramm der Regierung, das versucht, benachteiligten Personen bessere Arbeitsplätze zu verschaffen. Eine Stichprobe potenzieller Programmteilnehmer wurde beobachtet. Von diesen Personen wurden nach dem Zufallsprinzip einige für die Teilnahme an dem Programm ausgewählt. Jeder Fall entspricht einem Programmteilnehmer.

Index

Access (Microsoft), 14
Assistent für Datum und Uhrzeit, 135
Assistent für Textimport, 19
Ausblenden von Zeilen und Spalten in Pivot-Tabellen, 93
Auswählen von Fällen, 144
Auswertungsmaße
  Kategoriale Daten, 56
  Metrische Variablen, 59

Balkendiagramme, 57
Bearbeiten von Pivot-Tabellen, 92
bedingte Ausdrücke, 134
Beispieldateien
  Speicherort, 163
Berechnen von neuen Variablen, 130

Daten-Editor
  Eingeben von nichtnumerischen Daten, 31
  Eingeben von numerischen Daten, 28
  Mehrere geöffnete Datendateien, 51
Daten-Sets
  umbenennen, 53
Datenbank-Assistent, 14
Datenbankdateien
  einlesen, 14
Datendateien
  Mehrere geöffnete Datendateien, 51
Dateneingabe, 28, 31
Datentypen
  für Variablen, 34
Datum- und Uhrzeit-Variablen, 135
Diagramme
  Balken, 57, 63
  Bearbeiten von Diagrammen, 69
  Diagrammoptionen, 82
  Erstellen von Diagrammen, 63
  Histogramme, 61
  Vorlagen, 77

Eingeben von Daten
  nichtnumerisch, 31
  Numerisch, 28
Erstellen von Variablenlabels, 33
Excel (Microsoft)
  Ergebnisse exportieren in, 107
Excel-Dateien
  einlesen, 12
Exportieren von Ergebnissen
  HTML, 118
  in Excel, 107
  in PowerPoint, 107
  in Word, 107

Fälle
  auswählen, 144
  sortieren, 141, 144
Fehlende Werte
  für nichtnumerische Variablen, 40
  für numerische Variablen, 39
  Systemdefiniert fehlend, 38
Funktionen in Ausdrücken, 132

Grafiken
  Balken, 63
  Bearbeiten von Grafiken, 69
  Diagrammoptionen, 82
  Erstellen von Grafiken, 63
  Vorlagen, 77

Häufigkeiten
  Häufigkeitstabellen, 56
Häufigkeitstabellen, 56
Histogramme, 61
HTML
  Exportieren von Ergebnissen, 118

Intervalldaten, 55

Kategoriale Daten, 55
  Auswertungsmaße, 56
Kopieren von Variablenattributen, 41

Mehrere geöffnete Datendateien, 51
  Unterdrücken, 54
Messniveau, 55
metrische Daten, 55
Metrische Variablen
  Auswertungsmaße, 59

Nominale Daten, 55
Numerische Daten, 28

Ordinale Daten, 55

Pivot-Tabellen
  Aufrufen von Definitionen für statistische Begriffe in der Ausgabe, 88
  Ausblenden des Dezimaltrennzeichens, 94

Ausblenden von Zeilen und Spalten, 93
bearbeiten, 92
Formatierung, 92
Pivot-Leisten, 89
Schichten, 91
Transponieren von Zeilen und Spalten, 89
Zellendatentypen, 94
Zellenformate, 94
PowerPoint (Microsoft)
Ergebnisse exportieren in, 107

Qualitative Daten, 55
Quantitative Daten, 55

Schichten
Erstellen in Pivot-Tabellen, 91
Sortieren von Fällen, 141
Stetige Daten, 55
String-Daten
Eingeben von Daten, 31
Symbol “Hilfe zur Syntax”, 122
Syntax, 120
Syntaxdateien
öffnen, 123
Syntaxfenster
Ausführen von Befehlen, 122–123
Bearbeiten von Befehlen, 122
Einfügen von Befehlen, 120
Systemdefiniert fehlende Werte , 38

Tabellenkalkulationsdateien
einlesen, 12
Einlesen von Variablennamen, 12
Teilmengen von Fällen
anhand von Datum und Uhrzeit, 147
auswählen, 144
bedingte Ausdrücke, 145
falls Bedingung zutrifft, 145
Filtern von nicht ausgewählten Fällen, 148
Löschen von nicht ausgewählten Fällen, 148
Zufallsstichprobe, 146
Textdatendateien
einlesen, 19
Transponieren (Vertauschen) von Zeilen und Spalten in Pivot-Tabellen, 89

Übernehmen von Befehlssyntax
aus einem Dialogfeld, 120
Umbenennen von Daten-Sets, 53
Umkodieren von Werten, 124

Variablen, 28
Beschriftungen, 33
Datentypen, 34
Variablenattribute
wiederverwenden, 41
Variablenlabels
erstellen, 33
Verarbeitung von aufgeteilten Dateien, 142
Verhältnisdaten, 55
Verschieben
Elemente in Pivot-Tabellen, 89
Objekte im Viewer, 86
Viewer
Aus- und Einblenden der Ausgabe, 86
Verschieben der Ausgabe, 86

Wertelabels
nichtnumerische Variablen, 36
numerische Variablen, 34
Steuern der Anzeige im Viewer, 34, 36
zuordnen, 34, 36
Word (Microsoft)
Ergebnisse exportieren in, 107