

In 25 Minuten zum SPSS-Profi



In diesem Kapitel

- ▶ Das Programm SPSS starten
- ▶ Daten aus einer Befragung vom Papier in den Computer eingeben
- ▶ Eine Datendatei anlegen und richtig strukturieren
- ▶ Stupide, aber notwendig: Daten eintippen
- ▶ Mit SPSS rechnen und Daten für eine Analyse vorbereiten
- ▶ Eine erste statistische Auswertung
- ▶ Eine erste Grafik mit SPSS erstellen
- ▶ Daten und Ergebnisse speichern
- ▶ SPSS beenden

S PSS dient dazu, mehr oder weniger umfangreiche Datenmengen mit statistischen Methoden zu untersuchen und auszuwerten. Nur wer eine solche statistische Datenanalyse durchführen möchte, wird sich typischerweise überhaupt näher mit SPSS beschäftigen. Dennoch kann das Programm SPSS wesentlich mehr, als »einfach nur Analysen durchzuführen«, und wenn Sie ernsthaft mit SPSS arbeiten, werden Sie auch nicht umhin kommen, neben der eigentlichen Analyse weitere, meistens vorbereitende Arbeitsschritte mit SPSS umzusetzen. Bevor sich irgendwelche Daten mit SPSS analysieren lassen, müssen diese nämlich erst einmal in eine Datendatei von SPSS eingegeben oder aus anderen Programmen eingelesen und so strukturiert werden, dass sie von SPSS interpretiert und analysiert werden können. In diesem Kapitel werden an einem kleinen Beispiel alle Arbeitsschritte vorgestellt, die eine vollständige Analyse mit SPSS typischerweise umfasst – vom Starten des Programms über die Dateneingabe und die Durchführung statistischer Analysen bis zum Beenden von SPSS. Wenn Sie dieses Beispiel nachvollzogen haben, werden Sie nahezu alle Arten von Arbeiten kennen, die Sie mit SPSS durchführen können. Natürlich gibt es von jedem Arbeitsschritt sehr, sehr viele Varianten (zum Beispiel sehr viele unterschiedliche Verfahren der statistischen Analyse oder viele verschiedene Grafiktypen), die Sie in dem folgenden Beispiel nicht alle kennen lernen werden, aber in Bezug auf den generellen Ablauf der Arbeit mit SPSS sollten Sie nach diesem Kapitel keine großen Überraschungen mehr erleben.

Eine »typische Aufgabenstellung für SPSS«

Sollten Sie irgendwann einmal ein Buch in die Hände bekommen, in dem eine Überschrift großspurig eine »typische Aufgabenstellung für SPSS« ankündigt, können Sie das Buch getrost ungelesen zurücklegen, denn offensichtlich hat der Autor keine Ahnung, worüber er schreibt:

SPSS wird in der Praxis für derart viele und unterschiedliche Fragestellungen zum Beispiel aus der Marktforschung, den Naturwissenschaften, der Ökonomie oder den Sozialwissenschaften eingesetzt, dass es eine »typische Aufgabenstellung für SPSS« gar nicht geben kann. Allerdings gibt es so etwas wie »typische Arbeitsschritte«, die unabhängig von der konkreten Aufgabenstellung bei der Datenanalyse mit SPSS anfallen. So lässt sich nahezu jedes Analyseprojekt, das mit SPSS durchgeführt wird, in die folgenden Teilaufgaben untergliedern, wobei je nach Art und Umfang des Projekts einige dieser Schritte auch entfallen können:

- ✓ Zu Beginn wird das Programm SPSS gestartet.
- ✓ Die zu analysierenden Daten müssen irgendwie in eine Datendatei von SPSS gelangen. Hierzu können die Daten entweder direkt in SPSS eingetippt oder aus einem anderen Programm übernommen werden.
- ✓ Bevor die Daten analysiert werden können, müssen sie häufig noch aufbereitet oder umstrukturiert werden.
- ✓ Anschließend kann die eigentliche Analyse der Daten durchgeführt werden. Das Ergebnis solcher Analysen sind meistens Tabellen oder Grafiken, in denen die Resultate präsentiert werden.
- ✓ Wenn die Ergebnisse der Analyse in Dokumentationen oder Präsentationen weiterverwendet werden sollen, lassen sie sich zuvor mit SPSS in ihrem Aussehen präsentationsreif gestalten, so dass man die Ergebnisse hoffentlich nicht nur inhaltlich, sondern auch äußerlich herzeigen mag.
- ✓ Damit Daten und Ergebnisse nicht verloren gehen, müssen sie vor dem Beenden des Programms gespeichert werden.
- ✓ Zum Abschluss wird das Programm SPSS wieder beendet.

Ein erstes kleines Beispiel

Im Folgenden wird eine kleine Analyse durchgeführt, die alle wesentlichen Arbeitsschritte umfasst. Die Aufgabenstellung besteht darin, die Ergebnisse einer kleinen Kundenbefragung auszuwerten. Eine solche Aufgabe ist »ein Klassiker« für alle »Newbies«, die anfangen, mit SPSS zu arbeiten. Das Beispiel geht davon aus, irgendein beliebiges Unternehmen, wie zum Beispiel ein Kaufhaus, habe seine Kunden anhand der »Gewinnspielkarte« aus Abbildung 1.1 über deren Produktinteressen befragt und gleichzeitig versucht, Informationen über das Alter der Kunden zu gewinnen und das Einverständnis für weitere Werbezusendungen zu erhalten. Natürlich ist die hier abgebildete Karte simplifiziert, denn wesentliche Informationen wie die Abfrage von Adressdaten, einer Kundennummer oder der E-Mail-Adresse fehlen. Gehen Sie einfach davon aus, die Daten würden auf der Rückseite der Karte abgefragt und sind im ersten Schritt nicht relevant.



Antwortkarte

Vorname: Oskar

Nachname: Schwarz

Geburtsdatum: 14.8.1964

Ja, ich möchte an dem Gewinnspiel teilnehmen.

Bitte informieren Sie mich über neue Angebote im Bereich

Computer Audio/TV/Photo Hausgeräte

Abbildung 1.1: Antwortkarte aus einer fiktiven Kundenbefragung

Um die mit diesen Antwortkarten gewonnenen Daten auszuwerten, werden im Folgenden alle oben aufgeführten Arbeitsschritte durchgeführt. Sie können die folgende Analyse Schritt für Schritt an Ihrem PC nacharbeiten. Je nachdem, wie viel Erfahrung Sie bereits mit anderen Programmen wie Excel, Word oder Access haben, werden Ihnen dabei einige Schritte vielleicht sehr vertraut vorkommen. Bei anderen Schritten haben Sie möglicherweise das Gefühl, man müsste noch wesentlich mehr Details erfahren, um alle Möglichkeiten kennen zu lernen. Dieses Gefühl wird dann vermutlich richtig sein, denn jeder der folgenden Arbeitsschritte wird in den weiteren Kapiteln dieses Buches noch sehr viel detaillierter vorgestellt. Das folgende Beispiel sollte vor allem ein guter Start für die erste Begegnung mit SPSS sein:

- ✓ Sie lernen den »typischen« Ablauf einer Analyse mit SPSS kennen.
- ✓ Sie werden mit der Oberfläche von SPSS und verschiedenen Fenstern vertraut (SPSS hat nämlich jeweils eigene Fenster zur Darstellung und Bearbeitung von Daten, Ergebnissen und Grafiken).
- ✓ Sie können hinterher ohne rot zu werden behaupten, bereits vollständige Analysen mit SPSS durchgeführt zu haben, also quasi ein SPSS-Profi zu sein.

SPSS starten

Wie für nahezu jedes Programm unter Windows gilt auch für SPSS: Viele Wege führen zum Starten des Programms. Glücklicherweise gibt es aber überhaupt keinen Grund, alle Wege zu lernen – es genügt vollkommen, den einfachsten Weg zu kennen.

Der einfachste Weg zu SPSS

Der einfachste Weg zum Starten von SPSS ist meistens das Icon auf der Oberfläche von Windows. Wenn Sie auf dieses Icon doppelklicken, starten Sie das Programm.



Statt dieses Icons auf der Windows-Oberfläche können Sie auch das START-Menü verwenden, um SPSS aufzurufen. Die verschiedenen Wege zum Starten von SPSS werden im folgenden Kapitel systematisch dargestellt.

Verknüpfung zu SPSS: Doppelklicken auf das Symbol startet das Programm

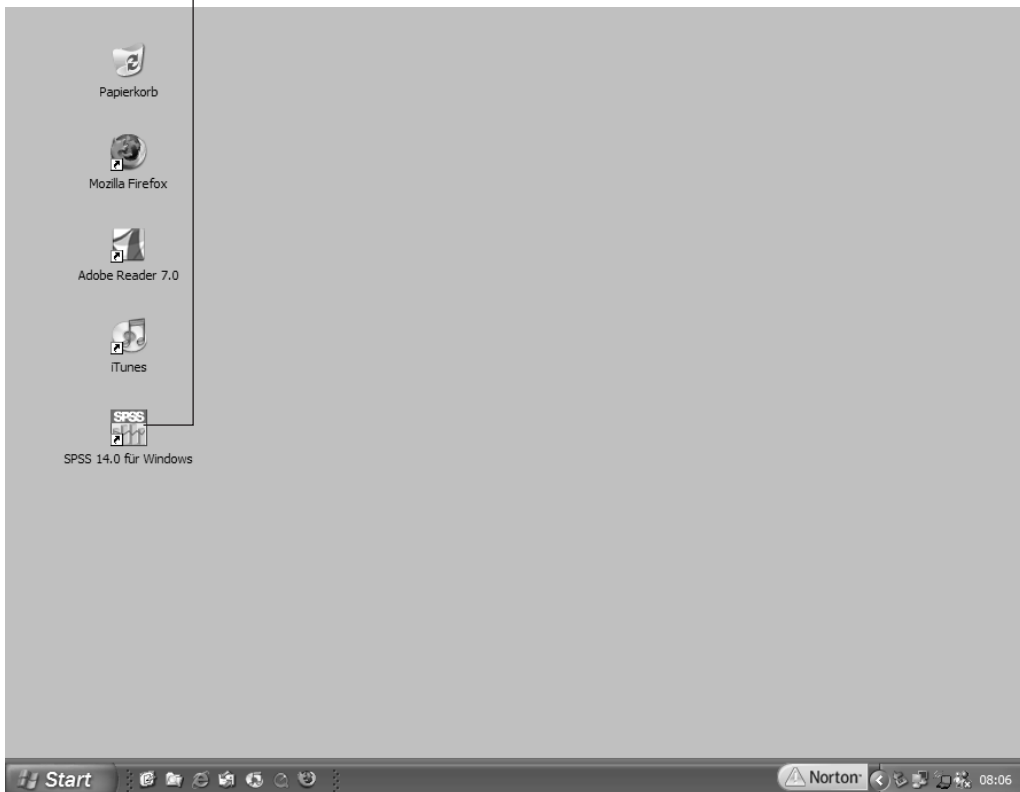


Abbildung 1.2: Windows-Oberfläche mit dem Icon von SPSS zum Starten des Programms

Die erste Begrüßung durch SPSS

Die Begrüßung durch SPSS fällt traditionell sehr nüchtern aus. Je nach den bei Ihnen vorgenommenen Einstellungen zeigt SPSS zur Begrüßung entweder nur ein vollkommen leeres Tabellenblatt oder, wie in Abbildung 1.3 dargestellt, ein leeres Tabellenblatt mit einem Dialogfeld, das Sie fragt, was Sie denn jetzt eigentlich tun möchten. Sollte dieses Dialogfeld bei Ihnen angezeigt werden, ist es in den meisten Fällen das Zweckmäßigste, dieses Dialogfeld einfach mit der Schaltfläche **ABBRECHEN** zu schließen, um ohne Umwege mit der Arbeit beginnen zu können. Danach sieht der Bildschirm ziemlich aufgeräumt aus und zeigt neben einer leeren Tabelle nur einige wenige Symbole und Menübefehle.

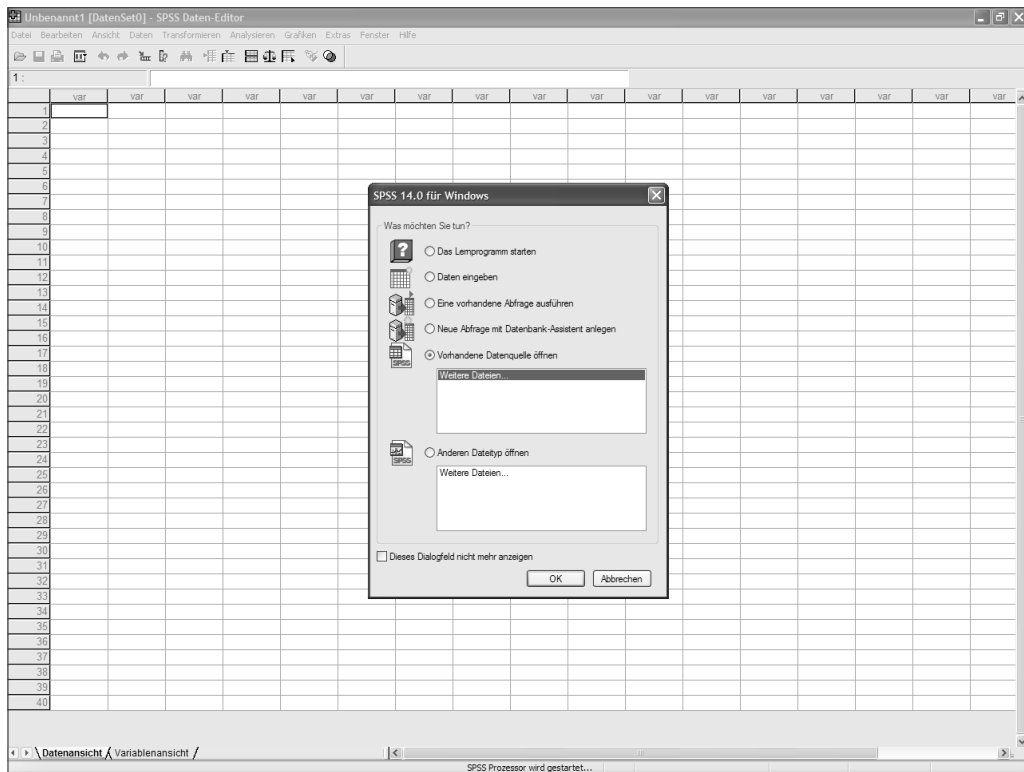


Abbildung 1.3: Erster Bildschirm nach dem Start von SPSS – hier mit Begrüßungdialogfeld

Datendatei anlegen

Die leere Tabelle in Abbildung 1.3 ist das Gerüst einer Datendatei. Die ersten Schritte bei der Arbeit mit SPSS bestehen nahezu immer darin, diese Tabelle mit den Daten zu füllen, die dann im Folgenden analysiert werden sollen. Das Füllen dieser Datei kann dadurch geschehen, dass die zu untersuchenden Daten tatsächlich einzeln »mit der Hand«, also über die Tastatur in die leere Tabelle eingegeben werden. In vielen Fällen liegen die Daten aber auch schon in elektronischer Form als Datei vor und können dann je nach dem Format dieser Datei einfach als fertige SPSS-Datei geöffnet oder aus einer anderen Anwendung wie Excel oder Access eingelesen werden. Dieses Glück haben Sie mit den Antwortkarten aus der Kundenbefragung leider nicht, so dass Sie im nächsten Schritt tatsächlich alle Daten mit der Hand eingeben müssen.

Ordnung schaffen: Daten brauchen eine Struktur

Bevor Sie mit der Eingabe der Antworten von den Antwortkarten in die Datendatei beginnen können, sollten Sie sich die Frage stellen, in welcher Form diese Daten in das leere Tabellen-

blatt von SPSS übertragen werden sollen. Diese Frage mag trivial erscheinen, in Wirklichkeit ist sie aber von zentraler Bedeutung. Vor jeder Dateneingabe sollte man sich zwei Punkte sehr gut überlegen:

- ✓ Wie sollen die Daten angeordnet werden? Die richtige Antwort auf diese Frage lautet: Jeder Datensatz wird als Zeile und jede Variable als Spalte angelegt. Was dies genau bedeutet, wird im folgenden Abschnitt deutlich.
- ✓ Wie können die Antworten so eingegeben werden, dass man möglichst wenig tippen muss, die Daten gut mit dem Computer auswerten kann und die Inhalte der Datendatei auch in Zukunft noch versteht? Hier werden Sie bei der Datenanalyse sehr schnell merken, dass Codierungen (statt der Eingabe vollständiger Antwortsätze) die Arbeit ganz erheblich erleichtern. Unten wird sich zeigen, wie sich dies auf die vorliegenden Antwortkarten übertragen lässt.

Von Fällen und Variablen in Zeilen und Spalten

Die Aufgabe des Beispiels besteht darin, die Ergebnisse der kleinen Kundenbefragung in die Datendatei einzugeben. Es werden also viele einzelne Kunden (beziehungsweise deren Antwortkarten) betrachtet, und für jeden dieser Kunden liegen mehrere Merkmale wie der Name, das Geburtsdatum, der Wunsch nach Werbezusendungen etc. vor. Jedes dieser Merkmale wird als eine *Variable* bezeichnet, denn es handelt sich um eine veränderliche (also *variable*) Größe, die an verschiedenen Stellen (hier bei verschiedenen Kunden) beobachtet wurde. So umfasst eine Variable alle Vornamen, die von den Kunden abgefragt wurden, eine zweite Variable enthält alle Nachnamen, eine dritte die Geburtstage und so weiter.

Für jeden Kunden aus der Kundenbefragung liegt damit in jeder Variablen genau ein Wert vor, also ein Wert in der Vornamensvariablen, ein Wert in der Nachnamensvariablen etc. Die Gesamtheit aller Werte, die sich auf denselben Kunden (generell auf dieselbe Beobachtungseinheit) beziehen, wird als ein *Datensatz* oder in der statistischen Analyse häufig auch als ein *Fall* bezeichnet.

Bei der Eingabe der Daten in eine Datendatei ist es allgemein üblich und bei SPSS zwingend notwendig, dass jeder Fall als eine Zeile und jede Variable als eine Spalte eingegeben wird. Das heißt im Ergebnis, dass sämtliche Werte einer Variablen wie zum Beispiel alle Nachnamen in einer Spalte untereinander stehen, während gleichzeitig alle Daten eines Falles, wie hier alle Informationen zu einem Kunden, in einer Zeile nebeneinander aufgeführt werden.

Was heißt das für das Antwortkarten-Beispiel?

Die Anwendung dieser Regeln auf die Kundenbefragung aus dem Beispiel ist sehr einfach: Jede Antwortkarte bildet einen Fall und wird damit zu einer Zeile in der Datendatei. Jede auf den Karten abgefragte Information bildet eine Variable und wird damit zu einer Tabellenspalte. Es werden also insgesamt acht Spalten (Variablen) benötigt, denn jede Antwortkarte fragt acht Informationen ab:

1. Vorname
2. Nachname
3. Geburtsdatum
4. Möchte der Kunde an dem Gewinnspiel teilnehmen?
5. Möchte der Kunde über Angebote informiert werden?
6. Interessiert sich der Kunde besonders für Computer?
7. Interessiert sich der Kunde besonders für Audio/TV/Photo?
8. Interessiert sich der Kunde besonders für Hausgeräte?

Warum das Ganze?

Wichtig ist: Diese Unterscheidung zwischen Fällen und Variablen dient nicht nur der Befriedigung irgendwelcher formaler Anforderungen oder theoretischer Überlegungen, sondern ist von höchster praktischer Bedeutung. Die Fälle sind stets die Untersuchungsgegenstände, über die man Erkenntnisse gewinnen möchte. In diesem Beispiel handelt es sich um Personen, ebenso könnten es aber auch bestimmte Gegenstände, Länder oder auch unterschiedliche Zeitpunkte, an denen Messungen vorgenommen wurden, sein. Die Variablen enthalten hingegen die unterschiedlichen Größen, die gemessen wurden. In der späteren Analyse werden diese Größen betrachtet, um Aussagen über die Untersuchungsgegenstände (hier die Kunden) zu gewinnen. Beispielsweise könnte in dem Beispiel mit den Antwortkarten aus dem Geburtsdatum der Kunden ermittelt werden, wie alt die Kunden im Durchschnitt sind. Ebenso ließe sich auszählen, welcher Anteil der Kunden sich für Computer interessiert und welcher Anteil über Hausgeräte auf dem Laufenden bleiben möchte. Eine solche Auswertung sämtlicher Daten einer Variablen über alle Fälle hinweg ist in umgekehrter Weise in aller Regel nicht sinnvoll. Auch in dem Beispiel mit den Antwortkarten lassen sich nicht alle Antworten eines Kunden gemeinsam betrachten, um daraus übergeordnete Erkenntnisse oder Durchschnittswerte abzuleiten, denn die Daten einer Untersuchungseinheit (hier eines Kunden) sind zum einen zu speziell (eben nur auf die eine Einheit bezogen) und zum anderen zu heterogen (hier Namensangaben, Datumsangaben, Interessensgebiete), um mit statistischen Methoden sinnvolle Erkenntnisse gewinnen zu können.

SPSS die Datenstruktur erklären

Bevor Sie nun mit der eigentlichen Dateneingabe beginnen können, müssen Sie SPSS erklären, wie die Datendatei aufgebaut sein soll. Dies geschieht, indem Sie die erforderlichen Variablen für die Datei beschreiben. Sie legen dabei fest, wie viele Variablen Sie benötigen, wie diese Variablen heißen sollen und welche Art von Daten (Texte, Datumswerte, Zahlen) in die Variablen eingegeben werden.

Die Variablenansicht der Datendatei

Eine Datendatei lässt sich bei SPSS in zwei unterschiedlichen Ansichten betrachten, der *Datenansicht* und der *Variablenansicht*. Um die Beschreibung der Variablen vornehmen zu können, muss die bisher leere Datendatei in der *Variablenansicht* angezeigt werden. Klicken Sie hierzu am unteren Rand der Datendatei auf die Registerkarte VARIABLENANSICHT (siehe Abbildung 1.3). Anschließend hat die Datendatei das Aussehen aus Abbildung 1.4. In dieser Ansicht dient jede Zeile der Tabelle dazu, eine Variable für die Datendatei anzulegen und näher zu beschreiben. Jede Variable, die Sie hier anlegen, wird anschließend in der Datenansicht der Datei als eine Spalte angezeigt.

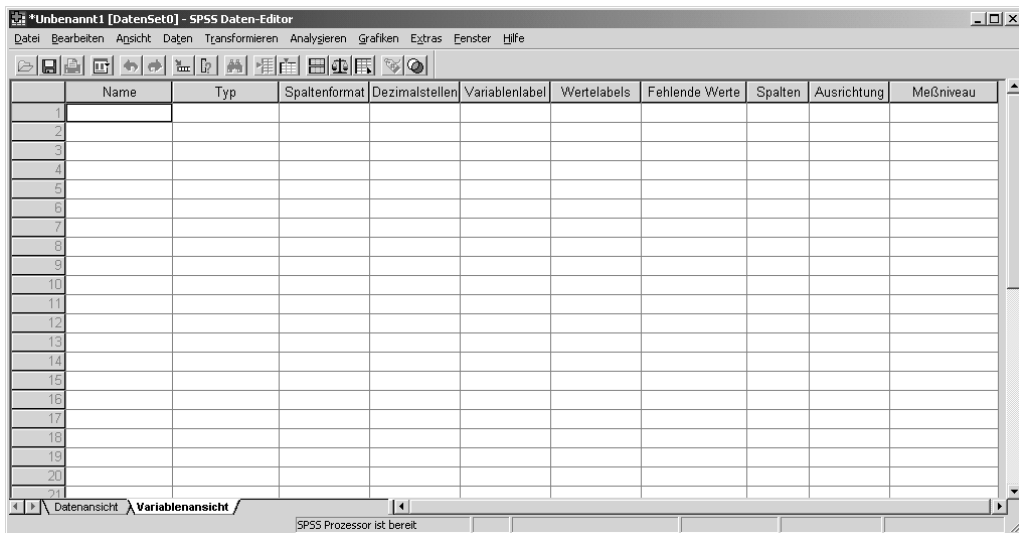


Abbildung 1.4: Leere Datendatei in der Variablenansicht

Den Variablen Namen geben

Das Anlegen einer Variablen beginnt immer mit der Vergabe eines Namens. Sie können daher im ersten Schritt acht Variablenamen vergeben, um die acht Variablen, die Sie zum Erfassen der Daten aus den Antwortkarten benötigen, zu definieren. Die Namen können im Rahmen gewisser Syntaxregeln frei gewählt werden, es empfiehlt sich aber in vielen Fällen, sprechende Namen zu verwenden. So könnte zum Beispiel die erste Variable *Vorname* heißen, die zweite *Nachname*, die dritte *Geburtsdatum* etc. Um die gewünschten Variablen mit diesen Namen zu erstellen, werden die Namen einfach untereinander in die Spalte *Name* eingetragen. Nach Eingabe der acht Namen, die Sie in diesem Beispiel verwenden, hat die Datendatei in der Variablenansicht das Aussehen aus Abbildung 1.5, wobei Sie auch beliebig andere Namen als die hier vorgeschlagenen verwenden können.

Sie werden bei der Eingabe der Variablenamen sehr schnell merken, dass dadurch nicht nur die Spalte *Name* gefüllt wird, sondern automatisch auch die übrigen Felder in der jeweiligen

Zeile Einträge zugewiesen bekommen. Dies hat folgende Bedeutung: Jede durch die Vergabe eines Namens definierte Variable wird von SPSS automatisch mit bestimmten Eigenschaften wie einem Datentyp und einem Format versehen, die in den übrigen Feldern der jeweiligen Zeile angezeigt werden. So ist zum Beispiel abzulesen, dass SPSS für jede neu definierte Variable zunächst unterstellt, sie solle numerische Daten (Zahlen) aufnehmen (Typ NUMERISCH) und für diese Zahlen zwei Dezimalstellen ausweisen. Diese Voreinstellungen sind natürlich in vielen Fällen unzutreffend und können daher für jede einzelne Variable geändert werden. Dies geschieht auch für dieses Beispiel in den folgenden Schritten, indem zum Beispiel für die Variablen Vorname und Nachname der numerische Datentyp in den Datentyp für eine Textvariable geändert wird.

	Name	Typ	Spaltenformat	Dezimalstellen	Variablenlabel	Wertelabels	Fehlende Werte	Spalten	Ausrichtung	Metrisch
1	Vorname	Numerisch	8	2		Kein	Kein	8	Rechts	Metrisch
2	Nachname	Numerisch	8	2		Kein	Kein	8	Rechts	Metrisch
3	Geburtsdatum	Numerisch	8	2		Kein	Kein	8	Rechts	Metrisch
4	Gewinnspiel	Numerisch	8	2		Kein	Kein	8	Rechts	Metrisch
5	Infos	Numerisch	8	2		Kein	Kein	8	Rechts	Metrisch
6	Info_Comp	Numerisch	8	2		Kein	Kein	8	Rechts	Metrisch
7	Info_Hifi	Numerisch	8	2		Kein	Kein	8	Rechts	Metrisch
8	Info_Haus	Numerisch	8	2		Kein	Kein	8	Rechts	Metrisch
9										
10										
11										
12										
13										
14										
15										
16										
17										
18										
19										
20										
21										

Abbildung 1.5: Datendatei in der Variablenansicht mit acht definierten Variablen

Der Datentyp bestimmt den Inhalt der Variablen

Die wichtigste Eigenschaft einer Variablen ist der Datentyp. Dieser legt fest, welche Art von Daten in diese Variable eingegeben werden kann. SPSS kennt im Wesentlichen drei verschiedene Datentypen:

- ✓ Der Typ NUMERISCH ist für jede Art von Zahlen geeignet.
- ✓ Variablen vom Typ STRING können beliebige Textwerte aufnehmen.
- ✓ Für Datumswerte gibt es den speziellen Typ DATUM.

Per Voreinstellung erhält jede Variable von SPSS automatisch den Datentyp NUMERISCH zugewiesen. Dies ist auch in Abbildung 1.5 zu sehen: Jede der acht Variablen, für die bisher lediglich die Namen festgelegt wurden, wird in der Spalte TYP als NUMERISCH ausgewiesen. Für die ersten drei Variablen ist dies natürlich Unsinn. In die Variablen Vorname und Nachname sollen Texte

eingetragen werden, also brauchen Sie hier den Datentyp `STRING`. Die Variable `Geburtsdatum` soll natürlich Datumswerte enthalten, so dass hier der Datentyp `DATUM` angebracht ist.

Nomenklatura: Namenslisten brauchen Textvariablen

Für beide Namensvariablen aus Abbildung 1.5 muss der Datentyp also so geändert werden, dass die Namen als Texte eingegeben werden können. Beginnen Sie mit der Variablen `Vorname`:

1. Markieren Sie in der Zeile der Variablen `Vorname` das Feld in der Spalte `Typ`.

Daraufhin erscheint am rechten Rand dieses Feldes eine Schaltfläche, auf der drei Punkte dargestellt sind (siehe Abbildung 1.6).

	Name	Typ	
1	Vorname	Numerisch ...	8
2	Nachname	Numerisch	8
3	Geburtsdatum	Numerisch	8

Abbildung 1.6: Ein Klick auf die Schaltfläche mit den drei Punkten öffnet das Dialogfeld zur Auswahl des Datentyps.

2. Klicken Sie auf die Schaltfläche mit den drei Punkten.

Sie öffnen damit das Dialogfeld aus Abbildung 1.7. In diesem Dialogfeld wählen Sie den Datentyp für die Variable aus. In Abbildung 1.7 ist noch die von SPSS verwendete Voreinstellung für eine numerische Variable zu sehen.



Abbildung 1.7: Dialogfeld zur Auswahl des Datentyps

3. Da Sie die Variable `Vorname` als Textvariable verwenden möchten, wählen Sie den Datentyp `STRING` (Text).

Sobald dieser Typ ausgewählt wird, erscheint ein neues Eingabefeld, in dem die Anzahl der Zeichen abgefragt wird.

4. Geben Sie hier die Anzahl an Zeichen ein, die die Textwerte der Variablen maximal umfassen können.

Gehen Sie für das Beispiel einfach einmal davon aus, dass keiner der Befragten Kunden einen Vornamen hat, der länger als 16 Zeichen ist. Dieser Wert ist in Abbildung 1.8 bereits eingetragen.

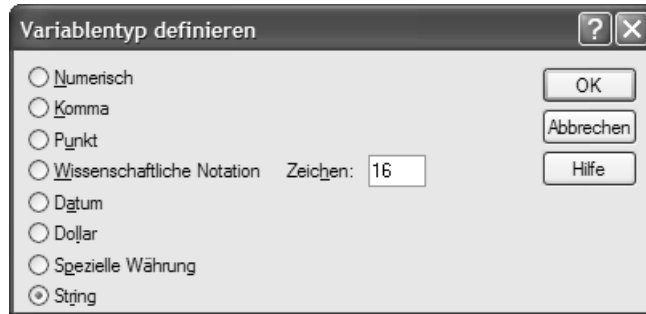


Abbildung 1.8: Datentyp `STRING` für eine Textvariable mit maximaler Länge von 16 Zeichen

5. Wenn Sie den Datentyp `STRING` ausgewählt und die Zeichenanzahl 16 eingegeben haben, können Sie das Dialogfeld mit der Schaltfläche `OK` schließen.

Daraufhin erhalten Sie wieder die Auflistung der bisher definierten Variablen mit ihren Eigenschaften, in der nun wie in Abbildung 1.9 die Angaben für die Variable `Vorname` aktualisiert sind. Die zusammen mit dem Variablentyp definierte Zeichenzahl wird hier in der Spalte `SPALTENFORMAT` aufgeführt.

	Name	Typ	Spaltenformat	Dezimalstellen	Var
1	Vorname	String ...	16	0	
2	Nachname	Numerisch	8	2	
3	Geburtsdatum	Numerisch	8	2	

Abbildung 1.9: Darstellung der Variablen `Vorname` nach Änderung des Datentyps

Ebenso wie die Variable `Vorname` benötigt auch die Variable `Nachname` den Datentyp `STRING`, und auch hier scheint eine Zeichenanzahl von 16 für die meisten Namen ausreichend. Um dieses Beispiel nachzuarbeiten, ändern Sie den Datentyp auf die gleiche Weise wie bei der Variablen `Vorname`.

Wie speichert man ein Geburtsdatum? Als Datum!

Um SPSS zu erklären, dass Sie in die Variable `Geburtsdatum` Datumswerte schreiben möchten, müssen Sie hier ebenfalls den Datentyp ändern:

1. Markieren Sie in der Zeile für die Variable Geburtsdatum das Feld **Typ** und öffnen Sie mit der daraufhin angezeigten Schaltfläche das Dialogfeld für die Auswahl des Datentyps.
2. In diesem Dialogfeld wählen Sie nun – wenig überraschend – den Typ **DATUM**.

Damit ist es jedoch noch nicht getan, denn SPSS kennt viele verschiedene Datumsformate. Diese werden nach der Auswahl des DATUM-Typs alle in einer Liste aufgeführt, aus der Sie nun Ihr Lieblingsformat auswählen können. In Abbildung 1.10 ist das Format **DD.MM.YYYY** ausgewählt, mit dem ein Datum in der Form 19.03.2007 dargestellt wird.

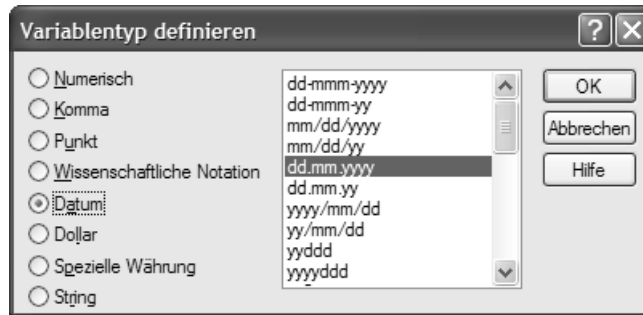


Abbildung 1.10: Datentyp **DATUM** für eine Datumsanzeige in der Form 19.03.2007

3. Nachdem der Datentyp **DATUM** und das gewünschte Format ausgewählt sind, kann das Dialogfeld mit der Schaltfläche **OK** geschlossen werden.

Danach wird auch für die Variable **Geburtsdatum** der neue Datentyp wie in Abbildung 1.11 in der Variablenliste aufgeführt.

	Name	Typ	Spaltenformat	Dezimalstellen	V
1	Vorname	String	16	0	
2	Nachname	String	16	0	
3	Geburtsdatum	Datum	10	0	
4	Gewinnspiel	Numerisch	8	2	

Abbildung 1.11: Darstellung der Variablen **Vorname**, **Nachname** und **Geburtsdatum** mit neuem Datentyp

Ein bisschen schwanger gibt es nicht: Der richtige Datentyp für Ja/Nein-Variablen

Neben dem Namen und dem Geburtsdatum werden auf der Antwortkarte aus Abbildung 1.1 nur »Fragen zum Ankreuzen« abgefragt. Die Antworten auf diese Fragen lauten immer entweder **Ja** (wenn ein Kästchen angekreuzt wurde) oder **Nein** (wenn das Kästchen nicht angekreuzt wurde). Dementsprechend wird zum Beispiel die Variable **Gewinnspiel** nach der Dateneingabe auch nur die zwei unterschiedlichen Werte **Ja** und **Nein** enthalten und damit kennzeichnen, ob die jeweilige Person an dem Gewinnspiel teilnehmen möchte oder nicht.

Es stellt sich nun die Frage, wie derartige Ja/Nein-Daten am sinnvollsten in die Datendatei eingegeben werden sollten. Denkbar sind verschiedene Wege, naheliegender wären zum Beispiel die folgenden Alternativen:

- ✓ Sie erstellen eine Textvariable, in die Sie tatsächlich die Werte *Ja* und *Nein* eintragen; diese Textvariable müsste dann mindestens vier Zeichen für die Datenwerte zulassen.
- ✓ Um Speicherplatz zu sparen, verwenden Sie eine Textvariable mit einer Breite von nur einem Zeichen und tragen die Antworten in der abgekürzten Form *J* und *N* ein oder verwenden die »Anglizismen« *Y* und *N*.
- ✓ Da Speicherplatz heutzutage kein großes Problem mehr darstellt und Sie klar verständliche Datenstrukturen zu schätzen wissen, erstellen Sie eine Textvariable mit einer Breite von 26 Zeichen und tragen hier die Werte *Gewinnspielteilnahme* und *Keine Gewinnspielteilnahme* ein.
- ✓ Sie übersetzen die Ja/Nein-Antworten in Codierungen wie zum Beispiel 1 für *Ja* und 0 für *Nein*. Dementsprechend verwenden Sie den Datentyp *NUMERISCH*, um die Zahlencodes abzuspeichern. Gleichzeitig sollten Sie sich dann aber »gut merken«, welche Codierung für welche Antwort steht.

Von diesen Alternativen ist streng genommen keine richtiger oder falscher als die anderen und tatsächlich werden in der Praxis auch alle vier Varianten benutzt. Allerdings arbeiten »echte Profis« in der Praxis nahezu ausschließlich mit der letzten Variante, also mit numerischen Codierungen wie 1 und 0 für *Ja* und *Nein*. Diese Art der Codierung hat zahlreiche Vorteile. Unter anderem erlaubt sie eine sehr bequeme und schnelle Dateneingabe, spart Speicherplatz, ermöglicht sehr schnelle Auswertungen und – das ist das Wichtigste – ist zwingende Voraussetzung dafür, dass die Daten in statistischen Analyseverfahren wie einer Cluster- oder Regressionsanalyse verwendet werden können, denn alle diese Verfahren setzen mathematische Algorithmen ein und brauchen deshalb Zahlen und keine Texte oder Buchstaben als Input.

Damit auch Sie dieses Codierungsverfahren für die Daten aus den Antwortkarten verwenden können, müssen Sie die entsprechenden Variablen *Gewinnspiel*, *Infos* etc. als numerische Variablen definieren. Da dies schon per Voreinstellung der Fall ist, brauchen Sie eigentlich nichts mehr zu tun. Aus »ästhetischen« Gründen sollten Sie aber doch noch zwei Anpassungen an den Eigenschaften der Variable vornehmen:

1. Da in die Variablen nur die Codierungen 0 und 1 eingegeben werden, können Sie festlegen, dass die Variablen nur einstellige Zahlen ohne Dezimalstellen enthalten sollen. Um dies für die Variable *Gewinnspiel* vorzugeben, schreiben Sie einfach in der entsprechenden Zeile in die Spalte *DEZIMALSTELLEN* den Wert 0 und anschließend in die Spalte *SPALTENFORMAT* (damit ist die Spaltenbreite gemeint) den Wert 1. Wichtig ist hierbei, dass Sie erst den Wert für die Dezimalstelle und dann die Spaltenbreite ändern!
2. Damit Sie nicht vergessen, wofür die Codierungen 0 und 1 inhaltlich stehen, können Sie deren Bedeutung als so genannte Wertelabels speichern. Ein Wertelabel »klebt ein Etikett« (ein Label) an einen Wert, um so eine inhaltliche Erläuterung des Wertes zu notieren. Um für die Variable *Gewinnspiel* solche Wertelabels zu definieren, gehen Sie folgendermaßen vor:

- ◆ Markieren Sie in der Zeile der Variablen `Gewinnspiel` das Feld in der Spalte `WERTE-LABELS`. Daraufhin erscheint am rechten Rand dieses Feldes eine Schaltfläche, auf der drei Punkte abgebildet sind.
- ◆ Klicken Sie auf die Schaltfläche mit den drei Punkten. Damit öffnen Sie das Dialogfeld aus Abbildung 1.12, das zunächst vollkommen leer ist.
- ◆ Tragen Sie in dem Dialogfeld in das Feld `WERT` die Zahl `0` und in das Feld `LABEL` den Text `Nein` ein. Klicken Sie anschließend auf die Schaltfläche `HINZUFÜGEN`, um so festzulegen, dass der Wert `0` mit dem Label `Nein` versehen wird.
- ◆ Legen Sie anschließend auf die gleiche Weise für den Wert `1` das Label `Ja` fest. Danach hat das Dialogfeld das Aussehen aus Abbildung 1.12. Klicken Sie nun auf die Schaltfläche `OK`, um diese Angaben zu speichern und das Dialogfeld zu schließen.



Abbildung 1.12: Festlegen von Wertelabels für die zwei Codierungen 0 und 1

Diese beiden Schritte zum Ändern der Dezimalstellen und Spaltenbreite sowie zum Festlegen der Wertelabels können Sie in identischer Form für die übrigen Ja/Nein-Variablen `Infos`, `Info_Comp`, `Info_Hifi` und `Info_Haus` wiederholen. Danach sollte die Datendatei wie in Abbildung 1.13 dargestellt aussehen.

Richtfest: Das Gerüst steht

Bis hierher hatten Sie schon jede Menge Arbeit allein für das Anlegen und die Beschreibung der Variablen, ohne dass auch nur eine einzige Antwort aus der Kundenbefragung in die Datendatei eingegeben wurde. Die gute Nachricht ist: Die Arbeit war nicht überflüssig. Sie haben sich damit ein Tabellengerüst geschaffen, das exakt auf Ihre Daten, die Sie im nächsten Schritt eingeben wollen, zugeschnitten ist. Das Erstellen eines solchen Tabellengerüsts ist immer der erste Schritt, und es lohnt sich, hierfür ein wenig Zeit zu investieren, denn die richtige Tabel-

lenstruktur erleichtert nicht nur die Dateneingabe, sondern auch die spätere Arbeit mit den Daten erheblich. Die schlechte Nachricht: Sie können nun mit der eigentlichen Dateneingabe beginnen – und das ist mindestens so viel Arbeit wie das Bauen des Gerüsts.

	Name	Typ	Spaltenformat	Dezimalstellen	Variablenlabel	Wertelabels	Fehlende Werte	Spalten	Ausrichtung	Meßniveau
1	Vorname	String	16	0		Kein	Kein	8	Links	Nominal
2	Nachname	String	16	0		Kein	Kein	8	Links	Nominal
3	Geburtsdatum	Datum	10	0		Kein	Kein	8	Rechts	Metrisch
4	Gewinnspiel	Numerisch	1	0		{0, Nein}...	Kein	8	Rechts	Metrisch
5	Infos	Numerisch	1	0		{0, Nein}...	Kein	8	Rechts	Metrisch
6	Info_Comp	Numerisch	1	0		{0, Nein}...	Kein	8	Rechts	Metrisch
7	Info_Hifi	Numerisch	1	0		{0, Nein}...	Kein	8	Rechts	Metrisch
8	Info_Haus	Numerisch	1	0		{0, Nein}...	Kein	8	Rechts	Metrisch
9										
10										
11										
12										
13										
14										
15										
16										
17										
18										

Abbildung 1.13: Datendatei in der Variablenansicht mit allen acht fertig definierten Variablen

Daten eingeben

Um mit der Dateneingabe zu beginnen, müssen Sie zunächst die Datendatei wieder in der DATENANSICHT anzeigen. Klicken Sie hierzu am unteren Rand der Datendatei auf die Registerkarte DATENANSICHT. Damit ändern Sie die Darstellung der Datendatei, die nun wie in Abbildung 1.14 dargestellt aussieht. Jede der bisher definierten acht Variablen bildet hier, wie es sich gehört, eine Spalte. Die Namen der Variablen sind in den Spaltenköpfen angegeben, und die vielen leeren Zeilen warten darauf, dass Sie hier die Daten aus der Kundenbefragung eingeben.

Die Spaltenbreite ist veränderbar

Möglicherweise werden die einzelnen Spalten bei Ihnen schmaler dargestellt als in Abbildung 1.14, so dass die Variablennamen im Spaltenkopf nicht vollständig ausgeschrieben oder zweizeilig dargestellt werden. Um dies zu korrigieren, können Sie die Breite jeder einzelnen Spalte verändern, indem Sie den rechten Rand des Spaltenkopfes mit der Maus nach links oder rechts ziehen.

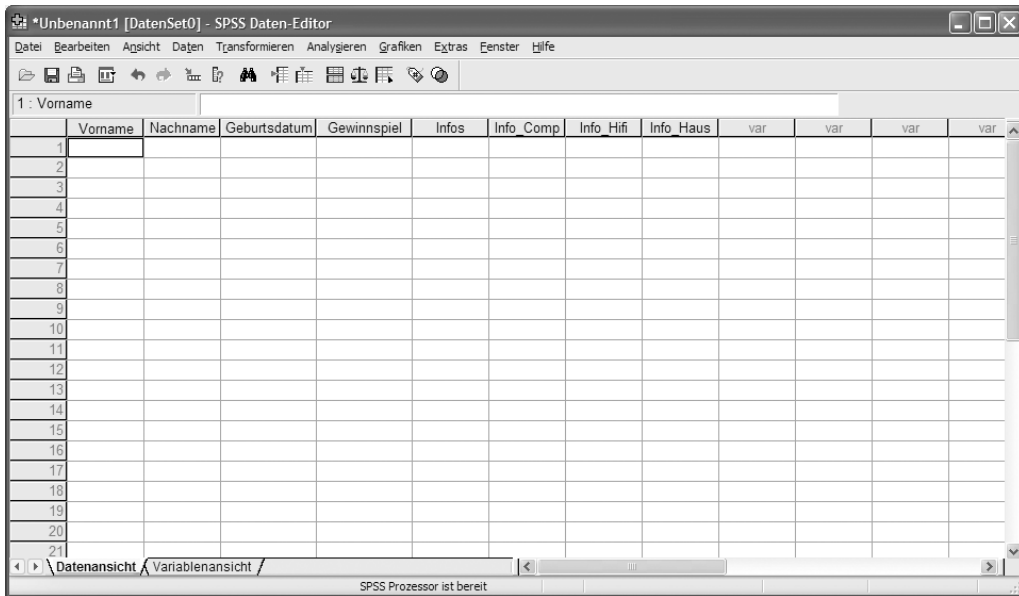


Abbildung 1.14: Datendatei in der Datenansicht nach der Definition von acht Variablen

Die nun beginnende Dateneingabe ist sehr, sehr einfach – und ziemlich laaangweilig. Aber leider unvermeidbar. In Abbildung 1.15 ist noch einmal die erste Antwortkarte aus der Kundenbefragung wiedergegeben.

Abbildung 1.15: Erste Antwortkarte aus der Kundenbefragung

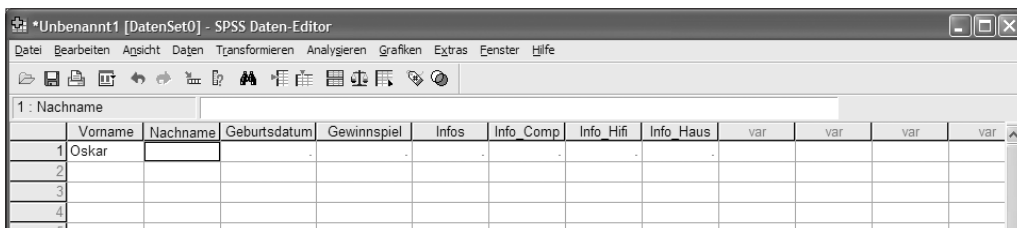
Um die Antworten von dieser Karte in die erste Zeile der Datendatei einzugeben, gehen Sie folgendermaßen vor:

1. Markieren Sie in der ersten Zeile das erste Feld der Spalte Vorname (zum Beispiel indem Sie dieses Feld einfach mit der Maus anklicken) und tippen Sie den Vornamen ein, den

Sie eingeben möchten, in diesem Fall also Oskar. Anschließend drücken Sie die Taste .

Damit bestätigen Sie die Eingabe und wechseln automatisch zum nächsten Feld in der ersten Zeile.


Abbildung 1.16 zeigt das Ergebnis der ersten Dateneingabe. Hier sehen Sie, dass nun nicht nur in der Spalte Vorname der eingegebene Name steht, sondern auch die Variablen Geburtsdatum bis Info_Haus in der ersten Zeile einen Punkt enthalten. Diesen Punkt können Sie zunächst einfach mal ignorieren; er soll lediglich anzeigen, dass in diesen Variablen noch die Werte fehlen, um den ersten Fall vollständig zu beschreiben. Einen solchen Hinweis zeigt SPSS außer bei Textvariablen immer an, wenn Felder in einem Fall (einer Zeile), die aus Sicht von SPSS Daten enthalten müssten, leer sind.



	Vorname	Nachname	Geburtsdatum	Gewinnspiel	Infos	Info_Comp	Info_Hifi	Info_Haus	var	var	var	var
1	Oskar	
2												
3												
4												
5												

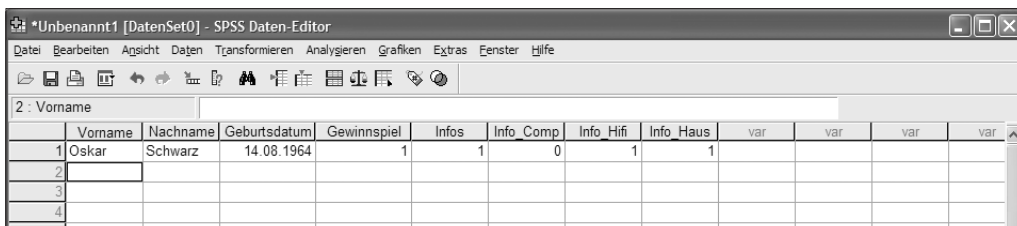
Abbildung 1.16: Datendatei nach der Eingabe des ersten Wertes in der Variablen Vorname

2. Setzen Sie die Dateneingabe für alle weiteren Felder des ersten Falles in der gleichen Weise fort.

Geben Sie also in das Feld Nachname den Namen Schwarz ein und drücken Sie anschließend die Taste . Danach geben Sie in die Variable Geburtsdatum das Datum 14.8.1964 ein, in die Variable Gewinnspiel den Wert 1 (für »Kästchen wurde angekreuzt«), in die Variable Infos ebenfalls den Wert 1 etc. Das Ergebnis dieser Eingabe des ersten Falles sollte nach Möglichkeit so aussehen, wie in Abbildung 1.17 wiedergegeben.




Wenn das Ergebnis bei Ihnen nicht so aussieht wie in Abbildung 1.17 dargestellt, sondern statt der eingegebenen Codierungen die Wertelabels angezeigt werden, ist bei Ihnen die so genannte Wertelabelansicht aktiviert. Um diese auszuschalten, wählen Sie den Menübefehl ANSICHT|WERTELABELS.



	Vorname	Nachname	Geburtsdatum	Gewinnspiel	Infos	Info_Comp	Info_Hifi	Info_Haus	var	var	var	var
1	Oskar	Schwarz	14.08.1964	1	1	0	1	1				
2												
3												
4												
5												

Abbildung 1.17: Datendatei nach der Eingabe aller Daten der ersten Antwortkarte

Wenn Sie den letzten Wert für diese Zeile in die Variable Info_Haus eingegeben haben und die Taste  drücken, markiert SPSS automatisch das erste Feld in der zweiten Zeile, da SPSS nun die Eingabe des zweiten Falles erwartet. Unabhängig davon können Sie natürlich jederzeit jedes beliebige Feld der Tabelle durch einfaches Anklicken mit der Maus markieren und so zum Beispiel auch bereits eingegebene Werte wieder überschreiben oder korrigieren.

3. Nun müssen die Daten der übrigen Antwortkarten eingegeben werden.

Damit Sie damit nicht den Rest des Tages verbringen, gehen Sie einfach mal davon aus, Sie hätten bisher nur 20 Antwortkarten erhalten. Die Daten von diesen 20 Antwortkarten können Sie in Abbildung 1.18 ablesen. Dort ist die Datendatei zu sehen, wie sie nach der Eingabe aller Daten aussieht.

Eine Besonderheit ist im neunten Fall zu erkennen: Herr Julius Ricke war ein wenig geizig mit seinen persönlichen Daten und hat das Geburtsdatum nicht mit angegeben. In einem solchen Fall bleibt das entsprechende Feld einfach leer. Der Punkt in dem Feld wurde nicht mit eingegeben, sondern wird automatisch von SPSS angezeigt als Hinweis darauf, dass in diesem Feld ein Wert fehlt.

	Vorname	Nachname	Geburtsdatum	Gewinnspiel	Infos	Info_Comp	Info_Hifi	Info_Haus	var	var	var	var
1	Oskar	Schwarz	14.08.1964	1	1	0	1	1				
2	Frank	Wenzel	12.11.1954	1	0	0	0	0				
3	Tim	Meyer	07.06.1981	1	0	0	0	0				
4	Lina	Rube	03.05.1978	0	1	1	1	1				
5	Susanne	Schumache	05.07.1966	1	0	0	0	0				
6	Christian	Höck	06.10.1983	1	0	0	0	0				
7	Nina	Gotze	08.08.1967	1	1	1	1	0				
8	Anna	Ballhorn	10.05.1975	0	1	0	0	1				
9	Julius	Ricke	.	0	1	0	0	0				
10	Mark	Schulze	11.12.1971	1	0	0	0	0				
11	Lars	Richter	16.01.1980	1	1	1	0	0				
12	Manfred	Scholz	22.01.1974	1	0	0	0	0				
13	Nils	Barr	31.03.1986	1	1	1	0	0				
14	Arne	Rauh	14.08.1977	1	1	1	1	1				
15	Nicola	Schmitt	28.06.1972	0	1	1	0	0				
16	Susanne	Baur	05.08.1979	1	0	0	0	0				
17	Christine	Gärtner	23.04.1957	0	1	0	0	1				
18	Markus	Feldmann	07.07.1967	1	1	0	1	0				
19	Alexander	Kreisler	15.05.1970	1	0	0	0	0				
20	Sophia	Schwaiger	26.02.1982	1	0	0	0	0				

Abbildung 1.18: Datendatei mit 20 Fällen und den Daten von 20 Antwortkarten

Ergebnisse der Dateneingabe speichern

Wenn Sie die Daten aus Abbildung 1.18 alle abgetippt haben, haben Sie möglicherweise ein gewisses Interesse daran, diese Daten zu speichern, um sie in den nächsten Schritten des

Beispiels und auch später jederzeit wiederverwenden zu können, ohne alle Daten noch einmal abtippen zu müssen. Und ehrlich gesagt ist das Speichern der Datei nach so viel mühsamer Arbeit eine ziemlich gute Idee!



Um die Datei zu speichern, wählen Sie den Menübefehl DATEI|SPEICHERN oder klicken auf die Schaltfläche DATEI SPEICHERN. Dies öffnet das Dialogfeld aus Abbildung 1.19. Dieses Dialogfeld entspricht den unter Windows üblichen Dialogfeldern zum Speichern von Dateien. Nehmen Sie hier folgende Angaben vor:

1. Wählen Sie in der Dropdown-Liste SPEICHERN IN den Ordner aus, in dem die Datei gespeichert werden soll.
2. Geben Sie im Feld DATEINAME einen Namen für die Datendatei an.

Diesen Namen können Sie frei wählen. In Abbildung 1.19 wird der Name Antwortkarten verwendet.

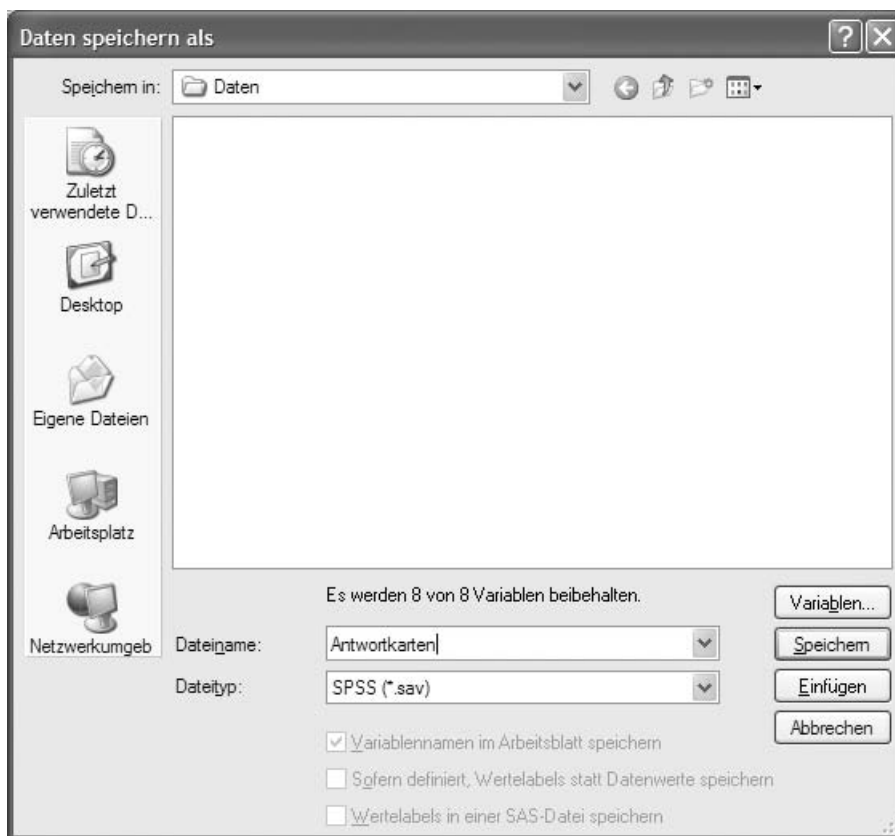


Abbildung 1.19: Dialogfeld zum Speichern der Datendatei

Datendateien von SPSS erhalten die Namenserweiterung `.sav`. Es ist egal, ob Sie diese Namenserweiterung mit eingeben oder nicht. Sie können also den Namen sowohl in der Form `Antwortkarten.sav` oder auch einfach als `Antwortkarten` in das Feld `DATEINAME` schreiben. Wenn Sie die Namenserweiterung nicht mit angeben, wird sie von SPSS automatisch angehängt.

3. Wenn Sie alle Angaben vorgenommen haben, klicken Sie auf die Schaltfläche `SPEICHERN`.

Erst damit wird die Datendatei tatsächlich gespeichert. Gleichzeitig wird das Dialogfeld wieder geschlossen, und Sie können mit der Datendatei weiterarbeiten. Da Sie der Datei beim Speichern auch einen Namen gegeben haben, wird dieser nun in der Titelleiste der Datei angezeigt, wo bisher wie zum Beispiel in Abbildung 1.18 noch die Bezeichnung `Unbenannt` stand.

Altbekanntes zum Befehl `SPEICHERN`

Je nachdem, mit welchem Betriebssystem Sie arbeiten und welche Einstellungen Sie dabei verwenden, kann das Dialogfeld zum Speichern der Datei anders als das in Abbildung 1.19 gezeigte aussehen. Sie werden aber in jedem Fall das Dialogfeld wiedererkennen, das in gleichartiger Form auch bei anderen Programmen auf Ihrem Computer zum Speichern von Dateien erscheint.

Wenn Sie eine Datendatei bereits einmal gespeichert haben und damit der Ordner und Name der Datei festgelegt sind, erscheint das Dialogfeld aus Abbildung 1.19 nicht mehr, wenn Sie den Befehl `DATEI|SPEICHERN` aufrufen, um beispielsweise Änderungen an der Datei zu sichern. Stattdessen wird die Datei dann einfach unter ihrem bisherigen Namen und in dem bisherigen Ordner mit den veränderten, aktuellen Inhalten gespeichert.

Neue Variablen berechnen

Nachdem die Daten nun endlich eingegeben und gespeichert sind, können Sie damit alle möglichen Analysen und Berechnungen durchführen. Unter anderem ist es sehr einfach möglich, aus den vorhandenen Daten weitere, abgeleitete Variablen zu berechnen. Dies ist sehr hilfreich, weil man häufig an Informationen interessiert ist, die zwar »irgendwie in den Daten drinstecken«, sich aber nicht direkt daraus ablesen, sondern nur über mehr oder weniger aufwendige Berechnungen ermitteln lassen. Zum Beispiel können Sie sich jetzt einmal anschauen, für wie viele der drei Produktgruppen `Computer`, `Audio/TV/Photo` und `Hausgeräte` sich die Befragten im Durchschnitt interessieren. Diese Information ist in den drei Variablen `Info_Comp`, `Info_Hifi` und `Info_Haus` enthalten und soll in einer Variablen zusammengefasst werden.

An dieser Stelle können Sie sich erst einmal selbst dazu gratulieren, dass Sie die Informationen, ob ein Kunde Werbung zu einer bestimmten Produktgruppe erhalten möchte oder nicht, als numerische 0/1-Codierungen abgespeichert haben, denn dies macht die Berechnung der Anzahl an Produktgruppen, für die sich ein Kunde interessiert, zu einem Kinderspiel: Sie addieren einfach für jeden Kunden (für jeden Fall und damit jede Zeile in der Datendatei) die Werte der

drei Variablen Info_Comp, Info_Hifi und Info_Haus und schreiben das Ergebnis in eine neue Variable, die Sie zum Beispiel Interessen nennen. Natürlich müssen Sie dies nicht »mit der Hand« berechnen, sondern können hierzu SPSS verwenden:

1. Wählen Sie den Menübefehl TRANSFORMIEREN|BERECHNEN.

Dieser Befehl öffnet das Dialogfeld aus Abbildung 1.20, in dem die beiden oberen Felder ZIELVARIABLE und NUMERISCHER AUSDRUCK jedoch zunächst leer sind.

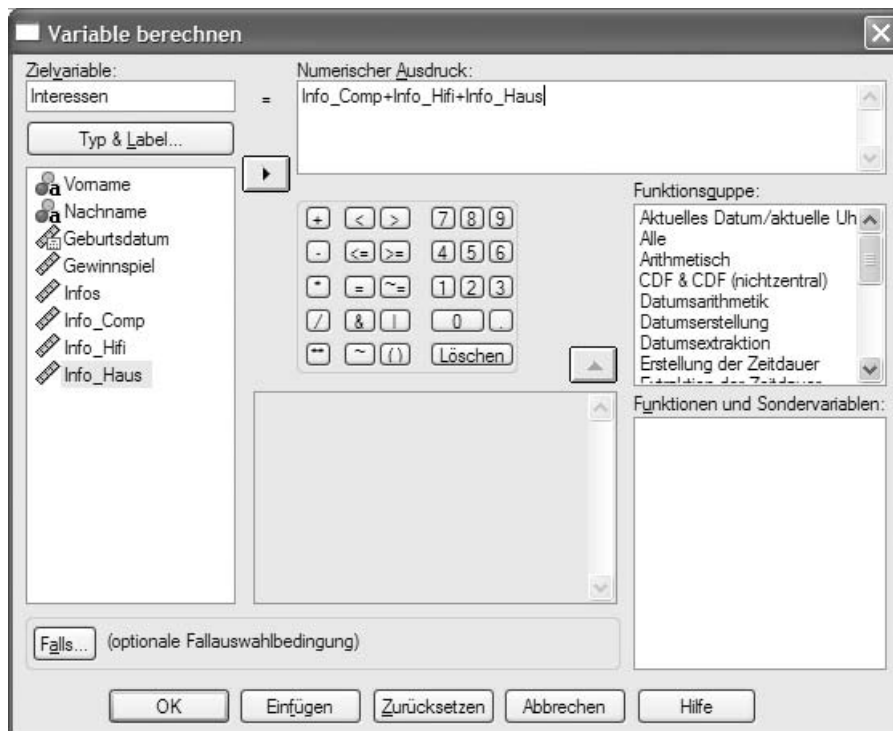


Abbildung 1.20: Dialogfeld des Befehls TRANSFORMIEREN|BERECHNEN zum Berechnen von neuen Variablen

2. Geben Sie in dem Dialogfeld in das Feld ZIELVARIABLE den Namen für die neu zu berechnende Variable ein (zum Beispiel den Namen Interessen) und schreiben Sie in das Feld NUMERISCHER AUSDRUCK die Formel, nach der die Werte für die Variable Interessen berechnet werden sollen.

Hier legt der Ausdruck

Info_Comp + Info_Hifi + Info_Haus

fest, dass in jeder Zeile der Datendatei die Werte der drei aufgeführten Variablen addiert und die Summe in die Zielvariable Interessen geschrieben werden soll.

3. Wenn Sie diese Angaben vorgenommen haben und das Dialogfeld wie in Abbildung 1.20 aussieht, klicken Sie auf die Schaltfläche OK, um die Berechnung der neuen Variablen zu starten.

Die neue Variable wird nun mit den berechneten Werten in die Datendatei eingefügt. Das Ergebnis ist in Abbildung 1.21 wiedergegeben. Es ist unmittelbar zu erkennen, dass alle theoretisch möglichen Werte zwischen 0 (der Kunde wünscht über keine der drei aufgeführten Produktkategorien Informationen) bis 3 (der Kunde hat alle drei Produktkategorien als Interessensgebiete angekreuzt) vertreten sind. Im nächsten Schritt kann nun für die Variable Interessen zunächst eine Häufigkeitstabelle und anschließend eine Grafik erzeugt werden, um darzustellen, wie viele der befragten Kunden sich für keine, für eine, für zwei oder für alle drei Produktkategorien interessieren.

	Vorname	Nachname	Geburtsdatum	Gewinnspiel	Infos	Info_Comp	Info_Hifi	Info_Haus	Interessen	var	var
1	Oskar	Schwarz	14.08.1964	1	1	0	1	1	2,00		
2	Frank	Wenzel	12.11.1954	1	0	0	0	0	,00		
3	Tim	Meyer	07.06.1981	1	0	0	0	0	,00		
4	Lina	Rube	03.05.1978	0	1	1	1	1	3,00		
5	Susanne	Schumache	05.07.1966	1	0	0	0	0	,00		
6	Christian	Höck	06.10.1983	1	0	0	0	0	,00		
7	Nina	Götze	08.08.1967	1	1	1	1	0	2,00		
8	Anna	Ballhorn	10.05.1975	0	1	0	1	1	2,00		
9	Julius	Ricke	.	0	1	0	0	0	,00		
10	Mark	Schulze	11.12.1971	1	0	0	0	0	,00		
11	Lars	Richter	16.01.1980	1	1	1	0	0	1,00		
12	Manfred	Scholz	22.01.1974	1	0	0	0	0	,00		
13	Nils	Barr	31.03.1986	1	1	1	0	0	1,00		
14	Arne	Rauh	14.08.1977	1	1	1	1	1	3,00		
15	Nicola	Schmitt	28.06.1972	0	1	1	0	0	1,00		
16	Susanne	Baur	05.08.1979	1	0	0	0	0	,00		
17	Christine	Gärtner	23.04.1957	0	1	0	0	1	1,00		
18	Markus	Feldmann	07.07.1967	1	1	0	1	0	1,00		
19	Alexander	Kreistler	15.05.1970	1	0	0	0	0	,00		
20	Sophia	Schwaiger	26.02.1982	1	0	0	0	0	,00		
21											

Abbildung 1.21: Datendatei mit der neu berechneten Variablen *Interessen*

Häufigkeitsverteilung einer Variablen darstellen

Wenn Sie wissen wollen, wie viele Kunden sich für keine, eine, zwei oder alle drei Produktgruppen interessieren, brauchen Sie jetzt nur noch die Werte in der Variablen *Interessen* auszuzählen. Genau dies tut eine Häufigkeitstabelle. Das ist eine sehr einfache Tabelle, die alle unterschiedlichen Werte einer Variablen auflistet und zusätzlich zu jedem Wert angibt, wie häufig dieser in der Variablen enthalten ist.

In 60 Sekunden zur Häufigkeitstabelle

Eine Häufigkeitstabelle ist mit SPSS in weniger als 60 Sekunden erstellt:

1. Rufen Sie den Menübefehl ANALYSIEREN\DESKRIPTIVE STATISTIKEN\HÄUFIGKEITEN auf.

Dieser Befehl öffnet das Dialogfeld aus Abbildung 1.22. In diesem Dialogfeld werden auf der linken Seite alle Variablen der aktiven Datendatei aufgeführt.

2. Markieren Sie in dieser Liste die Variable Interessen und klicken Sie anschließend auf die Schaltfläche mit dem Pfeil neben der Variablenliste.

Dadurch wird die Variable *Interessen* wie in Abbildung 1.22 zu sehen in das bisher leere Feld VARIABLE(N) verschoben, und es ist damit festgelegt, dass eine Häufigkeitstabelle für diese Variable erstellt wird.



Abbildung 1.22: Dialogfeld des Befehls ANALYSIEREN\DESKRIPTIVE STATISTIKEN\HÄUFIGKEITEN zum Erstellen von Häufigkeitstabellen

Nach der Auswahl der Variablen *Interessen* könnten Sie bereits auf die Schaltfläche OK klicken, um die Häufigkeitstabelle zu erstellen (damit wäre dann vermutlich auch das 60-Sekunden-Versprechen von oben eingehalten).

3. Wenn Sie jedoch stattdessen zunächst auf die Schaltfläche STATISTIK klicken, können Sie neben der Häufigkeitstabelle weitere Kennzahlen anfordern.

Diese Schaltfläche öffnet das Dialogfeld aus Abbildung 1.23.

4. Kreuzen Sie in diesem Dialogfeld wie in Abbildung 1.23 dargestellt die Option MITTELWERT an.

Dadurch legen Sie fest, dass SPSS nicht nur eine Häufigkeitstabelle für die Variable *Interessen* erstellen, sondern zusätzlich den Mittelwert dieser Variablen berechnen soll.

5. Mit der Schaltfläche WEITER wird das Dialogfeld wieder geschlossen.

6. Wenn Sie anschließend in dem Hauptdialogfeld aus Abbildung 1.22 auf die Schaltfläche OK klicken, startet SPSS mit der Berechnung des Mittelwertes und erstellt die Häufigkeitstabelle.

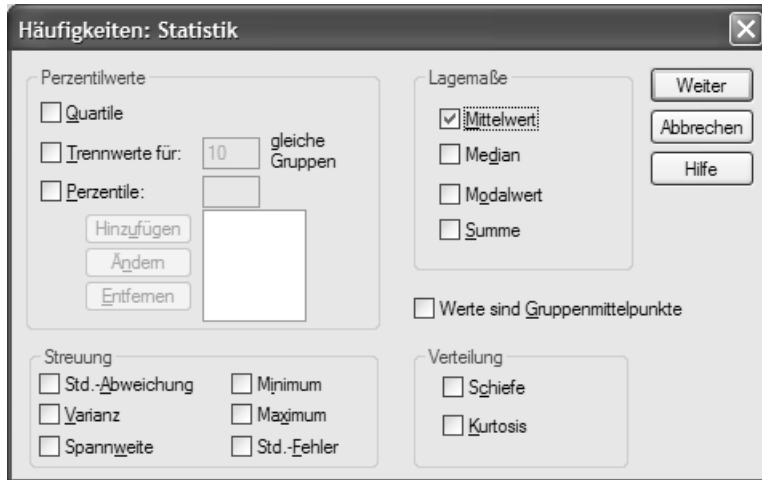


Abbildung 1.23: Dialogfeld der Schaltfläche STATISTIK; hier können statistische Kennzahlen zur Ergänzung der Häufigkeitstabelle angefordert werden.

Ergebnisse werden in eine Ausgabedatei geschrieben

Die Ergebnisse der Berechnung, also die Häufigkeitstabelle und der berechnete Mittelwert, werden von SPSS nicht in die Datendatei geschrieben. Stattdessen öffnet SPSS automatisch eine neue Datei in einem eigenen Fenster. Diese *Ausgabedatei* hat ein vollkommen anderes Format als die Datendatei und dient bei SPSS speziell für die Darstellung von Ergebnistabellen und Grafiken. Die Datei mit den gerade berechneten Ergebnissen ist in Abbildung 1.24 wiedergegeben.

Auf der linken Seite enthält die Datei ein Art Gliederung, die alle Elemente wie Ergebnistabellen, Grafiken und auch Überschriften, die von SPSS erstellt und in die Ausgabedatei geschrieben wurden, aufführt. Die Inhalte selbst werden in dem großen Bereich rechts daneben präsentiert. Die Gliederung auf der linken Seite kann auch zur Navigation innerhalb der Ausgabedatei genutzt werden: Wenn Sie mit der Maus auf eines der dort aufgeführten Elemente klicken, wird dieses Element automatisch in dem großen Bereich auf der rechten Seite angezeigt. Dies ist vor allem hilfreich, wenn mehr als die bisher erstellten zwei Ergebnisse in der Ausgabedatei enthalten sind.

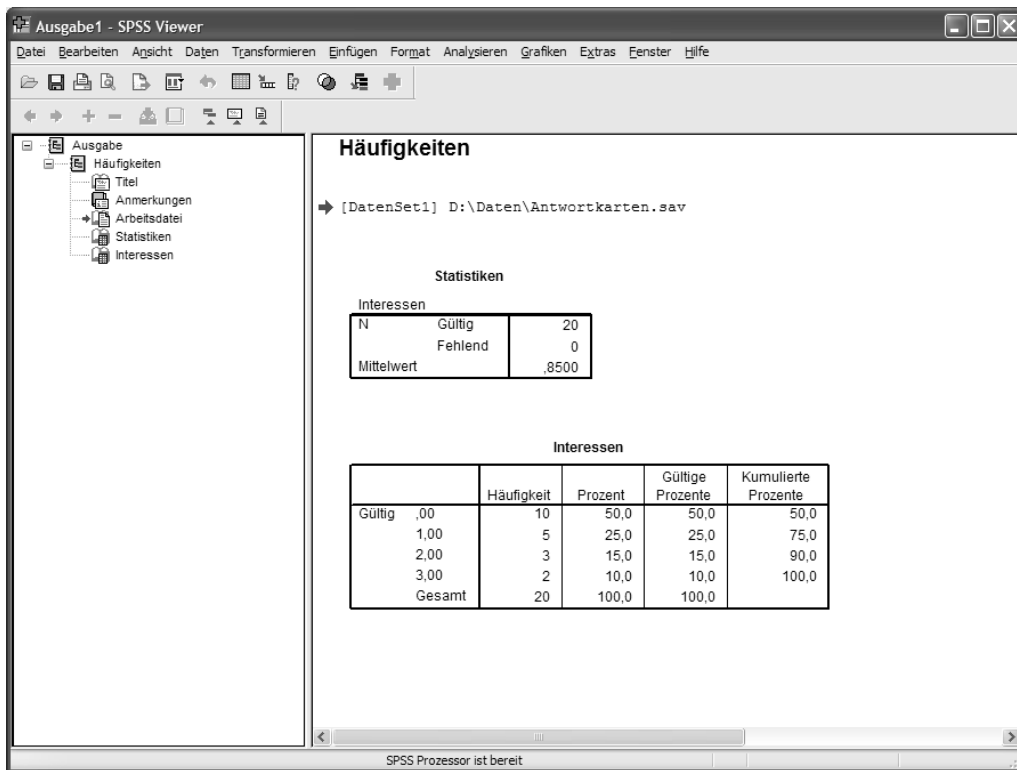


Abbildung 1.24: Ausgabedatei mit statistischen Kennzahlen und einer Häufigkeitstabelle

Ergebnisse richtig lesen

Als Ergebnis Ihrer Anforderung mit dem Befehl HÄUFIGKEITEN hat SPSS zwei Tabellen erstellt. Die obere Tabelle hat die Überschrift *Statistiken* und gibt den angeforderten Mittelwert für die Variable *Interessen* wieder; die zweite Tabelle zeigt die Häufigkeitsverteilung dieser Variablen.

Der Mittelwert

Der Mittelwert wird von SPSS mit ,8500 angegeben. Diese Zahl sieht etwas merkwürdig aus, weil SPSS etwas zu faul war, auch die führende Null vor dem Komma mit darzustellen, dafür aber am Ende noch zwei »überflüssige« Nullen spendiert hat. Sie und ich würden den Mittelwert natürlich üblicherweise in der Form 0,85 schreiben.

Das erste Ergebnis besagt damit, dass der Mittelwert aller Werte aus der Variablen *Interessen* 0,85 beträgt. Die 20 befragten Kunden haben also im Durchschnitt für weniger als eine Produktgruppe Zusatzinformationen angefordert. Es muss somit auf jeden Fall einige Befragte geben, die für keine der drei auf der Antwortkarte genannten Produktgruppen Informationen erhalten möchten.

Über dem Mittelwert werden in der Tabelle noch zwei weitere Werte ausgewiesen. Die Zeile *Fehlend* gibt mit dem Wert 0 an, dass in der Variablen *Interessen* kein Wert fehlt, alle Fälle in der Datendatei also einen (gültigen) Wert enthalten. Die Anzahl der (gültigen) Werte, die bei der Berechnung des Mittelwertes berücksichtigt wurden, beträgt damit 20; dieser Wert wird in der Zeile *Gültig* ausgewiesen.

Der tiefere Sinn von Nullen

Die Tatsache, dass SPSS den Mittelwert ohne führende Null darstellt, hat tatsächlich keinen tieferen Sinn, sondern ist lediglich eine etwas merkwürdige Darstellungsform von spleenigen Statistik-Freaks. Die beiden Nullen am Ende des Wertes ,8500 haben dagegen durchaus eine Bedeutung: Indem SPSS den Wert mit vier Dezimalstellen ausweist, wird implizit deutlich, dass der Wert bis auf vier Dezimalstellen genau angegeben ist. Würde der Wert ohne die beiden letzten Nullen in der Form ,85 wiedergegeben, könnte es sich auch um einen auf zwei Dezimalstellen gerundeten Mittelwert handeln, der in Wirklichkeit 0,854 oder 0,847 beträgt. Durch die beiden zusätzlich ausgewiesenen Dezimalstellen wissen Sie dagegen, dass dies nicht der Fall ist. Allerdings ist es weiterhin möglich, dass der Mittelwert in Wirklichkeit zum Beispiel 0,85004 oder 0,84998 beträgt.

Die Häufigkeitstabelle

Als zweites Ergebnis enthält die Ausgabedatei die Häufigkeitstabelle für die Variable *Interessen*. Diese Tabelle führt in der ersten Spalte alle Werte auf, die in der Variablen *Interessen* mindestens einmal vorkommen. Dies sind die Werte 0 (den SPSS wieder ohne führende Null in der Form ,00 angibt), 1, 2 und 3. Für jeden dieser Werte werden in der jeweiligen Tabellenzeile mehrere Häufigkeitswerte mitgeteilt. So ist in der ersten Zeile abzulesen, dass der Wert 0 mit einer Häufigkeit von 10 in der Variablen *Interessen* enthalten ist. Der Wert 1 kommt fünfmal, der Wert 2 dreimal und der Wert 3 zweimal in der Variablen *Interessen* vor. Die unterste Zeile der Tabelle enthält Angaben über alle Werte zusammen. Hier ist abzulesen, dass die Variable *Interessen* insgesamt 20 Werte enthält.

Inhaltlich bedeutet dies, dass zum Beispiel zehn der befragten Kunden auf der Antwortkarte angegeben haben, dass sie über keine der drei abgefragten Produktgruppen künftig informiert werden möchten. Von den insgesamt 20 vorliegenden Antwortkarten ist dies genau die Hälfte beziehungsweise 50 %. Dieser Anteil von 50 % wird auch in der Häufigkeitstabelle für den Wert 0 angegeben. Er steht dort in der Spalte *Prozent*. Für genau eine Produktgruppe interessieren sich 25 % der Befragten, für zwei Produktgruppen 15 % und für drei Produktgruppen 10 %. (Der unterste Wert in dieser Spalte, 100%, ist eine etwas alberne Angabe, denn er besagt genau genommen, dass alle Werte der Variablen *Interessen* zusammen genau 100 % der Werte aus dieser Variablen darstellen. Auf dieses Ergebnis wären wir auch von selbst gekommen.)

Ein Bild sagt mehr als tausend Worte

Die Informationen aus der Häufigkeitstabelle, nämlich die Verteilung der Werte in der Variablen *Interessen*, lassen sich noch plakativer und anschaulicher in einem Diagramm darstellen. Das Diagramm soll auf einen Blick zeigen, wie viele Kunden sich für keine, eine, zwei oder drei Produktgruppen interessieren. Hierfür eignet sich ein so genanntes Balkendiagramm, das für jeden der vier Werte 0, 1, 2 und 3 aus der Variablen *Interessen* einen Balken darstellt, wobei die Höhe des Balkens die Häufigkeit des jeweiligen Wertes anzeigt.

In 30 Sekunden zum Balkendiagramm

Auch ein Balkendiagramm ist genau wie eine Häufigkeitstabelle bei SPSS mit wenigen Klicks erstellt. Nachdem mit der Häufigkeitstabelle bereits erste Ergebnisse angefordert wurden, sind nun zwei Fenster von SPSS geöffnet, nämlich die Datendatei und die Ausgabedatei.

1. Um nun neue Ergebnisse wie eine Grafik aus den Daten zu erzeugen, stellen Sie zunächst sicher, dass die Datendatei das aktive Fenster darstellt.

Hierzu können Sie dieses Fenster mit den unter Windows üblichen Methoden auswählen. Zum Beispiel können Sie die entsprechende Schaltfläche in der Taskleiste von Windows anklicken oder mit der Tastenkombination **Alt** + **Esc** zwischen allen derzeit geöffneten Fenstern wechseln.

2. Um nun das Balkendiagramm zu erstellen, wählen Sie in dem Fenster der Datendatei den Menübefehl **GRAFIKEN|BALKEN**.



Abbildung 1.25: Dialogfeld zum Auswählen zwischen verschiedenen Arten von Balkendiagrammen

Dieser Befehl öffnet das Dialogfeld aus Abbildung 1.25. In diesem Dialogfeld wird die Art des Balkendiagramms näher bestimmt. Für Ihre Fragestellung sind die Voreinstellungen gerade richtig. Sie legen damit fest, dass Sie ein **EINFACHES** Balkendiagramm erstellen wollen, denn das Diagramm soll nur eine und nicht mehrere Folgen von Balken enthalten.

Außerdem nehmen Sie eine AUSWERTUNG ÜBER KATEGORIEN EINER VARIABLEN VOR, denn Sie betrachten ja nur die eine Variable *Interessen* und wollen deren unterschiedliche Kategorien (0, 1, 2 und 3) auswerten.

3. Deshalb können Sie die Voreinstellungen beibehalten und mit der Schaltfläche DEFINIEREN das nächste Dialogfeld öffnen.

Das daraufhin angezeigte Dialogfeld (siehe Abbildung 1.26) sieht zunächst etwas wild aus, ist aber bei näherer Betrachtung eigentlich ganz harmlos.

4. Sie müssen in diesem Dialogfeld nur zwei Angaben vornehmen, die in Abbildung 1.26 schon wiedergegeben sind:

- ◆ Als Erstes müssen Sie angeben, für welche Variable das Balkendiagramm erstellt werden soll. In Ihrem Fall ist dies die Variable *Interessen*. Diese Variable muss aus der Liste aller Variablen der aktuellen Datendatei in das Feld *KATEGORIENACHSE* verschoben werden. Hierzu markieren Sie die Variable *Interessen* in der Variablenliste, zum Beispiel indem Sie sie einmal mit der Maus anklicken, und klicken anschließend auf die Schaltfläche mit dem Pfeil neben dem Feld *KATEGORIENACHSE*.

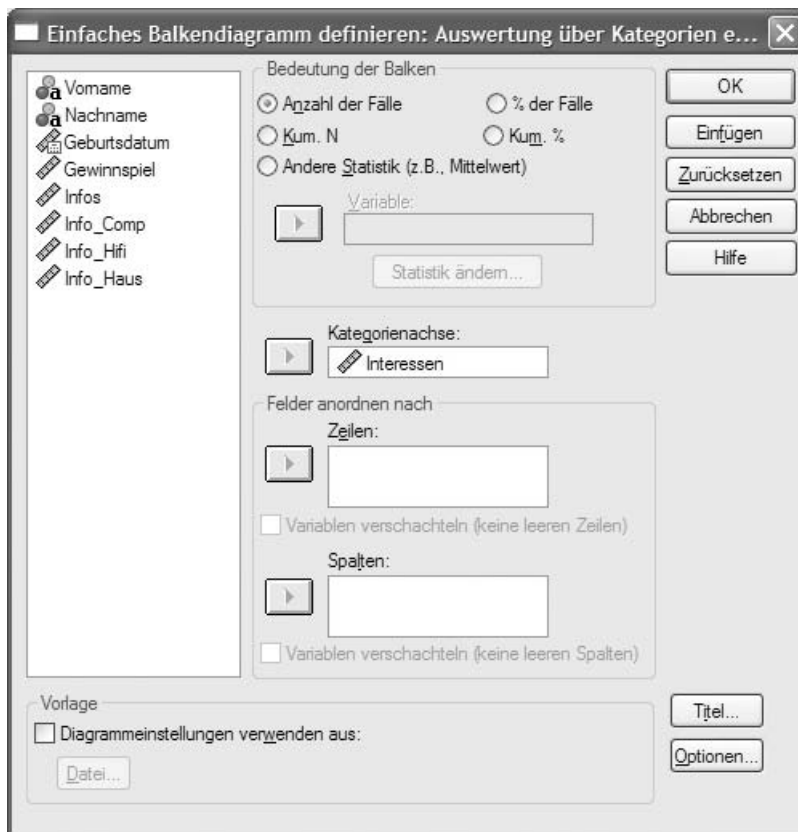


Abbildung 1.26: Dialogfeld zum Erstellen eines einfachen Balkendiagramms

- ◆ Als Zweites legen Sie fest, welche Information durch die Höhe der Balken dargestellt werden soll. In dem aktuellen Beispiel ist dies einfach die Anzahl der Fälle (also die Häufigkeit), in denen die unterschiedlichen Werte der Variablen *Interessen* vorkommen. Deshalb behalten Sie in der Gruppe *BEDEUTUNG DER BALKEN* die voreingestellte Option *ANZAHL DER FÄLLE* einfach bei.

Mit diesen beiden Angaben ist das gewünschte Diagramm bereits vollständig beschrieben.

5. Sie können jetzt auf die Schaltfläche **OK** klicken, um das Diagramm von SPSS erstellen zu lassen.

Die Grafik richtig lesen

Das Diagramm wird genau wie zuvor die Häufigkeitstabelle wieder in die Ausgabedatei geschrieben. Dazu erstellt SPSS nicht schon wieder eine neue Ausgabedatei, sondern fügt das Diagramm der bereits geöffneten Ausgabedatei hinzu. Das Resultat ist in Abbildung 1.27 wiedergegeben. Hier ist zunächst auf der linken Seite zu erkennen, dass die Gliederung nun sowohl die zuvor erstellte Häufigkeitstabelle als auch das neu hinzugekommene Diagramm aufführt. In dem Inhaltsbereich ist nur das Diagramm zu sehen, dies liegt aber lediglich an dem »zufällig« ausgewählten Bildausschnitt. Sie können in diesem Bereich auch die früheren Ergebnisse wie die Häufigkeitstabelle jederzeit wieder anzeigen lassen, zum Beispiel indem Sie mit der Bildlaufleiste am rechten Fensterrand den angezeigten Inhalt nach oben verschieben oder die entsprechenden Einträge in der Gliederung auf der linken Seite anklicken.

Das Diagramm zeigt, wie Sie es sicherlich schon erwartet haben, genau vier Balken, denn die Variable *Interessen* enthält ja vier unterschiedliche Werte (Kategorien). Jeder Balken bezieht sich auf jeweils eine dieser Kategorien aus der Variablen *Interessen*; dies wird unter der horizontalen Achse (der Abszisse) auch noch einmal durch den Variablennamen angezeigt. Die Höhe eines Balkens repräsentiert die *absolute Häufigkeit* (und nicht etwa irgendwelche Prozentwerte oder Ähnliches), mit der die jeweilige Kategorie in der Variablen *Interessen* enthalten ist. So ist zum Beispiel an dem ersten Balken abzulesen, dass die Kategorie 0 zehnmal vorkommt. Der zweite Balken zeigt, dass die Kategorie 1 fünfmal vertreten ist, und am letzten Balken ist abzulesen, dass der Wert 3 nur zweimal in der Variablen *Interessen* auftaucht.

Das Balkendiagramm gibt damit genau die gleichen Informationen wieder wie die Häufigkeitstabelle, die Sie oben erstellt haben. Dabei ist die Verteilung der Häufigkeiten über die unterschiedlichen Kategorien wesentlich plakativer und daher mit einem Blick zu erfassen, denn an der Grafik erkennt man sofort, dass die niedrigen Kategorien deutlich häufiger vorkommen als die hohen. Die meisten Kunden, die eine Antwortkarte ausgefüllt haben, wünschen also für keine oder nur für wenige Produktgruppen Informationen, nur wenige Kunden möchten tatsächlich das Fassungsvermögen ihres Briefkastens austesten und fordern von sich aus für mehrere Produktgruppen Werbematerial an.

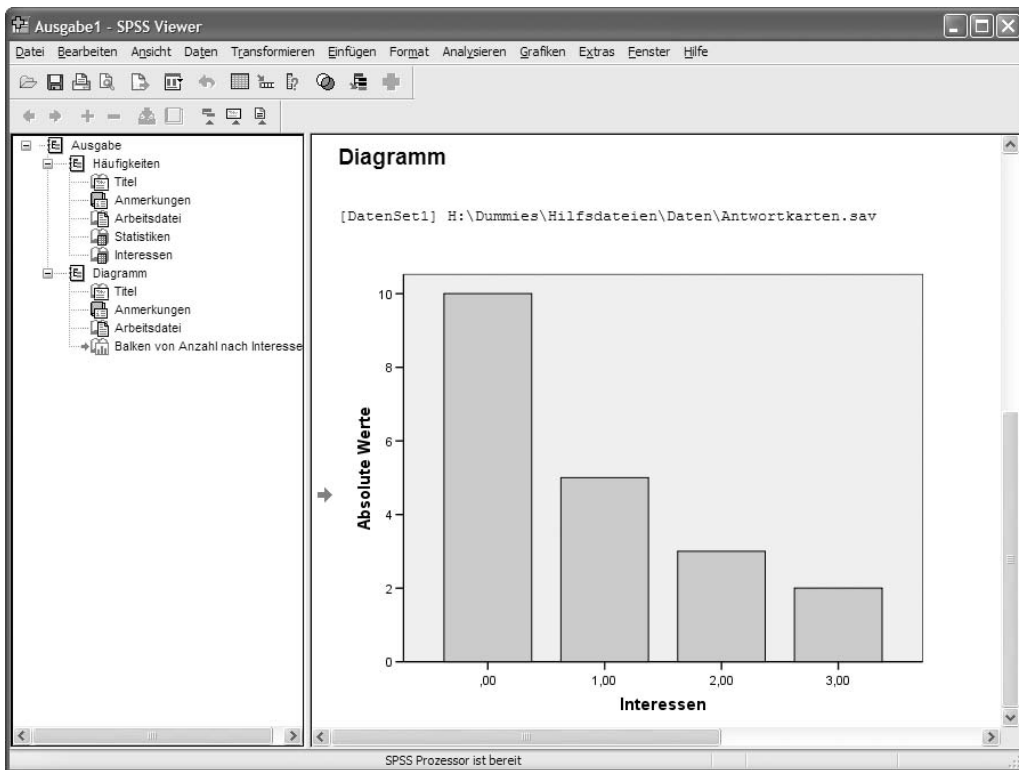


Abbildung 1.27: Ausgabedatei mit dem Balkendiagramm für die Häufigkeitsverteilung der Variablen *Interessen*

Früchte der Arbeit sichern

Alle Daten und Ergebnisse, die Sie mit SPSS erzeugen, werden beim Beenden des Programms SPSS wieder gelöscht und sind damit für immer verloren, wenn Sie sie nicht zuvor ausdrücklich speichern. Die mühsam eingegebenen Daten in der Datendatei wurden bereits oben gespeichert, allerdings haben Sie den Inhalt der Datendatei danach noch einmal verändert, indem Sie die neue Variable *Interessen* berechnet und hinzugefügt haben. Wenn auch diese Änderung erhalten bleiben soll, müssen Sie die Datendatei erneut speichern.

Datendatei erneut speichern



Um die Datendatei mit den Veränderungen neu zu speichern, muss zunächst sichergestellt sein, dass die Datendatei (und nicht etwa die Ausgabedatei) das aktive Fenster darstellt. Wenn dies nicht der Fall ist, können Sie die Taskleiste von Windows oder die Tastenkombination **ALT** + **F5** verwenden, um zur Datendatei zu wech-

seln. Anschließend wählen Sie den Menübefehl DATEI|SPEICHERN oder klicken auf die Schaltfläche DATEI SPEICHERN. Daraufhin wird die Datendatei unmittelbar in der aktuellen Version gespeichert. Es erfolgt keine Abfrage über ein Dialogfeld, wie die Datei heißen oder wo die Datei gespeichert werden soll, denn beides ist ja bereits vom ersten Speichern der Datei bekannt.



Wurde die Datei seit dem letzten Speichern nicht mehr verändert, so dass kein erneutes Speichern erforderlich ist, steht der Befehl DATEI|SPEICHERN bei SPSS gar nicht zur Verfügung, sondern ist ebenso wie die entsprechende Schaltfläche inaktiv.

Ausgabedatei mit den Ergebnissen speichern

Ebenso wie die Datendatei muss auch die Ausgabedatei gespeichert werden, wenn die Ergebnisse wie die Häufigkeitstabelle und das Balkendiagramm beim Beenden von SPSS nicht verloren gehen sollen. Das Vorgehen zum Speichern der Ausgabedatei ist dabei praktisch identisch mit dem Speichern einer Datendatei:

1. Stellen Sie zunächst sicher, dass die Ausgabedatei das aktive Fenster darstellt.

Wenn dies nicht der Fall ist, können Sie dieses Fenster über die Taskleiste von Windows oder die Tastenkombination **Alt** + **↵** anzeigen lassen.



2. Wählen Sie im Fenster der Ausgabedatei den Menübefehl DATEI|SPEICHERN oder klicken Sie auf die Schaltfläche DATEI SPEICHERN.

Beim erstmaligen Speichern der Ausgabedatei öffnen Sie damit das Dialogfeld aus Abbildung 1.28.

3. Wählen Sie in der Dropdown-Liste SPEICHERN IN den Ordner aus, in dem die Ausgabedatei gespeichert werden soll.

4. Geben im Feld DATEINAME einen Namen für die Ausgabedatei an.

Diesen Namen können Sie frei wählen. In Abbildung 1.28 wird der Name *Auswertung* verwendet.

Ausgabedateien von SPSS erhalten die Namenserweiterung *.spo*. Ob Sie diese Namensweiterung mit angeben oder nicht, ist vollkommen irrelevant. Wenn Sie die Namensweiterung nicht mit angeben, wird sie von SPSS automatisch angefügt, so dass die Datei in diesem Beispiel in jedem Fall unter dem vollständigen Namen *Auswertung.spo* gespeichert wird.

5. Wenn Sie den Ordner und den Dateinamen angegeben haben, klicken Sie auf die Schaltfläche SPEICHERN.

Damit wird die Ausgabedatei gespeichert und das Dialogfeld wieder geschlossen. Anschließend wird der Name, den Sie für die Ausgabedatei gewählt haben, auch in der Titelleiste der Datei angezeigt.

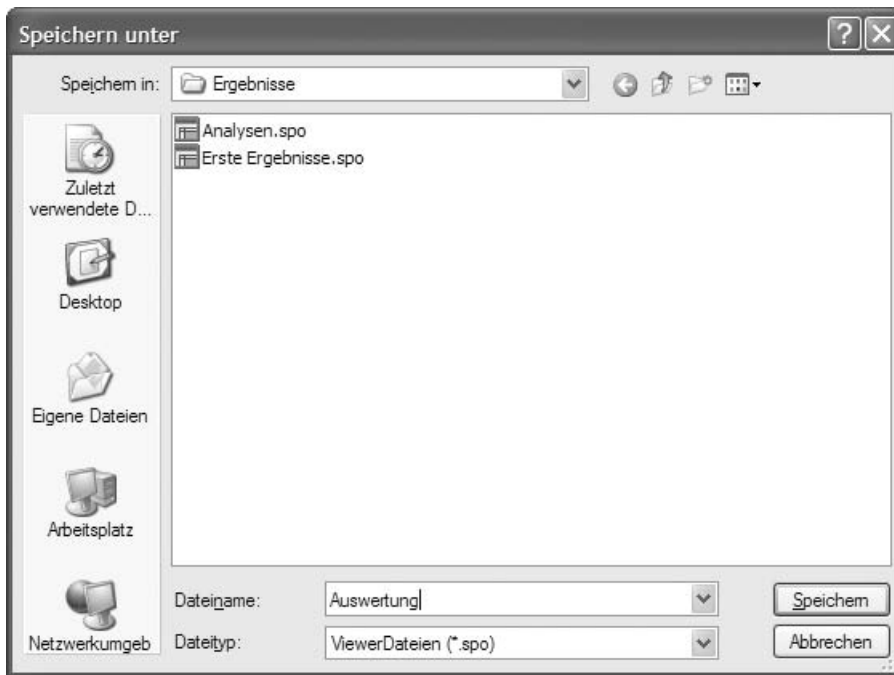


Abbildung 1.28: Dialogfeld zum Speichern der Ausgabedatei

SPSS beenden

Wenn Sie die Arbeit mit SPSS beenden wollen, können Sie das Programm vollständig schließen. Wählen Sie hierzu den Menübefehl **DATEI|BEENDEN**. Dabei ist es egal, in welchem Dateifenster Sie sich gerade befinden, Sie können den Befehl also sowohl in der Datendatei als auch in der Ausgabedatei aufrufen. Wenn Sie diesen Befehl wählen, hat dies folgende Konsequenzen:

- ✓ Alle derzeit geöffneten Fenster von SPSS, wie die Datendatei(en) und Ausgabedatei(en), werden geschlossen.
- ✓ Sollten an einer der derzeit geöffneten Dateien Änderungen vorgenommen worden sein, die bisher noch nicht gespeichert wurden, fragt SPSS vor dem Schließen der jeweiligen Datei, ob diese Änderungen nun gespeichert werden sollen. SPSS zeigt hierzu das Dialogfeld aus Abbildung 1.29 an. Sie können auf diese Frage mit **JA**, **NEIN** oder **ABBRECHEN** antworten:
 - **JA** bedeutet, dass die Änderungen gespeichert werden sollen. Wenn die jeweilige Datei schon einmal gespeichert wurde und damit schon einen Namen erhalten hat, speichert SPSS die Datei automatisch in der aktuellen Version unter ihrem bisherigen Namen und in dem bisherigen Ordner. Dabei wird die zuletzt gespeicherte Version der Datei überschrieben. Wurde die Datei bisher noch nie gespeichert, öffnet SPSS das Dialogfeld zum Speichern einer Datei, in dem Sie den Speicherort (Ordner) und Namen der Datei festlegen können.



Abbildung 1.29: Frage nach dem Speichern von Dateien beim Beenden von SPSS

- NEIN bewirkt, dass die Datei ohne vorheriges Speichern geschlossen wird. Wenn die Datei bisher noch nie gespeichert wurde, gehen alle Inhalte der Datei unwiederbringlich verloren. Wurde die Datei bereits in einer früheren Version gespeichert, bleibt diese Version erhalten und lediglich die Änderungen, die seit dem letzten Speichern an der Datei vorgenommen wurden, gehen verloren.
 - ABBRECHEN führt dazu, dass der gesamte Vorgang, SPSS zu beenden, abgebrochen wird. Wenn Sie ABBRECHEN wählen, bleibt SPSS also geöffnet, und Sie können weiterarbeiten oder die geöffneten Dateien näher betrachten, um zu entscheiden, ob Sie die Inhalte vor dem Beenden von SPSS speichern möchten.
- ✓ Wenn alle Fenster von SPSS geschlossen sind, ist auch das Programm vollständig beendet. Wenn Sie wieder mit SPSS arbeiten möchten, müssen Sie das Programm erneut starten.

