

Datenaufbereitung

Wir wollen in diesem Kapitel die Schritte der Datenaufbereitung anhand eines kleinen Beispiels vorstellen. Wir gehen dabei von einer fiktiven Befragung aus, bei der Studierende der Universität Heidelberg Wählern und Wählerinnen die »Sonntagsfrage« gestellt haben:

»Was würden Sie wählen, wenn am Sonntag Bundestagswahlen wären?«

Mit dem folgenden Fragebogen sind 30 Telefon-Interviews durchgeführt worden. Wir beschränken die Zahl der Interviews bewusst, um Ihnen unnötige Tipparbeit zu ersparen.

Institut für Soziologie Universität Heidelberg Seminar »Wahlforschung« Sommersemester 2002			
»Sonntagsfrage«			
Fragebogen			
Fragebogen-Nr.:	(wird vom Interviewer ausgefüllt)	
Geschlecht	<input type="checkbox"/>	weiblich	(bitte ankreuzen)
	<input type="checkbox"/>	männlich	
	<input type="checkbox"/>	keine Angabe	
Alter	(bitte Zahl eingeben)	
	<input type="checkbox"/>	keine Angabe	
Welche Partei würden Sie wählen, wenn am Sonntag Bundestagswahlen wären?	<input type="checkbox"/>	CDU/CSU	
	<input type="checkbox"/>	FDP	
	<input type="checkbox"/>	SPD	
	<input type="checkbox"/>	Grüne/Bündnis 90	
	<input type="checkbox"/>	PDS	
	<input type="checkbox"/>	Republikaner	
	<input type="checkbox"/>	Sonstige	
	<input type="checkbox"/>	keine Angabe	

Liegen die Fragebögen ausgefüllt vor, so müssen die erhobenen Daten EDV-gerecht aufbereitet werden, damit sie mit Hilfe des Programms SPSS für Windows ausgewertet werden können.

3.1 Kodierung und Kodeplan

Um die erhobenen Daten auswerten zu können, muss zunächst ein Kodeplan erstellt werden. Der Kodeplan ordnet den einzelnen Fragen des Fragebogens Variablenamen zu; so kann z.B. die Frage nach dem Geschlecht den Variablenamen sex erhalten.

Variablen sind Speicherstellen, die über die Tastatur eingegebene Werte aufnehmen können. Wir haben den Variablenamen sex gewählt, da Variablenamen bei SPSS für Windows aus maximal acht Zeichen bestehen dürfen. Der Variablenname geschlecht wäre somit zu lang. Variablenamen dürfen aus Buchstaben, Ziffern und einigen Sonderzeichen bestehen; das erste Zeichen eines Variablenamens muss ein Buchstabe sein.

Variablen können unterschiedliche Merkmalsausprägungen besitzen. Die Variable sex besitzt zwei Merkmalsausprägungen, die Merkmalsausprägung »weiblich« sowie die Merkmalsausprägung »männlich«. Der Kodeplan ordnet den Merkmalsausprägungen der einzelnen Variablen Kodenummern zu, so kann z.B. die »1« für weiblich stehen und die »2« für männlich.

Fassen wir die Aufgaben eines Kodeplans zusammen:

- ▶ Ein Kodeplan ordnet den einzelnen Fragen eines Fragebogens Variablenamen zu.
- ▶ Ein Kodeplan ordnet den Merkmalsausprägungen einer Variablen Kodenummern zu.

Bezogen auf unsere Befragung können wir das folgende Kodierschema entwerfen. Das Kodierschema ist dabei in den Fragebogen eingearbeitet.

Institut für Soziologie Universität Heidelberg Seminar »Wahlforschung« Sommersemester 2002			
»Sonntagsfrage«			
Fragebogen			
fragebnr:	Fragebogen-Nr.:	(wird vom Interviewer ausgefüllt)
sex:	Geschlecht	<input type="checkbox"/> 1: weiblich <input type="checkbox"/> 2: männlich <input type="checkbox"/> 0: keine Angabe	(bitte ankreuzen)
alter:	Alter <input type="checkbox"/> 0: keine Angabe	(bitte Zahl eingeben)
partei:	Welche Partei würden Sie wählen, wenn am Sonntag Bundestagswahlen wären?	<input type="checkbox"/> 1: CDU/CSU <input type="checkbox"/> 2: FDP <input type="checkbox"/> 3: SPD <input type="checkbox"/> 4: Grüne/Bündnis 90 <input type="checkbox"/> 5: PDS <input type="checkbox"/> 6: Republikaner <input type="checkbox"/> 7: Sonstige <input type="checkbox"/> 0: keine Angabe	

3.2 Datenmatrix

Wir gehen davon aus, dass die 30 Fragebögen wie folgt ausgefüllt vorliegen:

	fragebnr	sex	alter	partei
1	W-001	weiblich	45	CDU/CSU
2	W-002	männlich	22	SPD
3	W-003	männlich	19	SPD
4	W-004	weiblich	42	CDU/CSU
5	W-005	männlich	34	Grüne/Bündnis 90
6	W-006	weiblich	72	FDP
7	W-007	männlich	38	SPD
8	W-008	weiblich	56	SPD
9	W-009	männlich	61	CDU/CSU
10	W-010	weiblich	77	CDU/CSU
11	W-011	weiblich	23	Grüne/Bündnis 90
12	W-012	männlich	67	Republikaner
13	W-013	männlich	79	Sonstige
14	W-014	weiblich	26	SPD
15	W-015	männlich	59	CDU/CSU
16	O-001	weiblich	34	Grüne/Bündnis 90
17	O-002	männlich	18	Republikaner
18	O-003	weiblich	44	CDU/CSU
19	O-004	männlich	68	CDU/CSU
20	O-005	weiblich	33	PDS
21	O-006	männlich	66	CDU/CSU
22	O-007	weiblich	22	keine Angabe
23	O-008	männlich	keine Angabe	SPD
24	O-009	weiblich	67	SPD
25	O-010	männlich	33	FDP
26	O-011	männlich	44	CDU/CSU
27	O-012	weiblich	22	SPD
28	O-013	weiblich	19	Sonstige
29	O-014	weiblich	55	CDU/CSU
30	O-015	männlich	39	SPD

Bei der obigen Tabelle spricht man von einer Datenmatrix. Daten, welche mit SPSS für Windows ausgewertet werden, müssen in Form einer solchen Matrix vorliegen. Eine Datenmatrix besteht aus einer Anzahl von Zeilen sowie aus einer Anzahl von Spalten. Zeilen und Spalten ergeben ein rechteckiges Format. Eine Zeile steht dabei für einen Fragebogen, eine Spalte für eine Variable. Da bei unserer kleinen Befragung 30 Interviews durchgeführt wurden, liegen 30 Zeilen vor. Jede Zeile besteht aus vier Spalten, und zwar den vier Variablen `fragebnr`, `sex`, `alter` und `partei`.

Wir gehen bei der Erhebung davon aus, dass diese sowohl in den alten als auch in den neuen Bundesländern durchgeführt wurde. Die Interviewer wurden angewiesen, diesen Sachverhalt durch ein Kürzel bei der Fragebogennummer kenntlich zu machen. Das Kürzel »W-« soll für die alten Bundesländer stehen, das Kürzel »O-« für die neuen Bundesländer. W-001 steht z.B. für die erste Befragung, welche in den alten Bundesländern durchgeführt wurde, O-005 für die fünfte Befragung in den neuen Bundesländern.

3.3 Start von SPSS

Wir wollen mit der Dateneingabe des kleinen Auswertungsbeispiels beginnen.

- Starten Sie SPSS für Windows, indem Sie mit der linken Maustaste doppelt auf das SPSS-Symbol klicken.

Es öffnet sich der SPSS-Daten-Editor (siehe Bild 3.1).

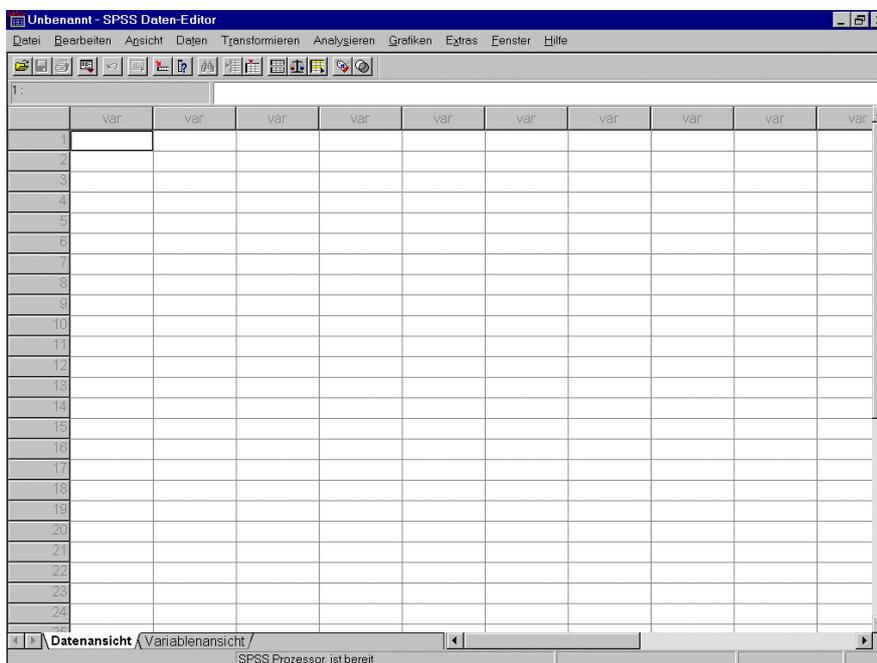


Bild 3.1: Daten-Editor: Datenansicht

Der Daten-Editor ist eines von mehreren Fenstern in SPSS. Sie können hier neue Daten eingeben oder bestehende Daten mit Hilfe der Menüwahl

*Datei
Öffnen*

aus einer Datendatei laden. Da beim Start von SPSS keine solche Datendatei geladen ist, wird der Daten-Editor mit »Unbenannt« betitelt. Der Daten-Editor wird am oberen Bildrand durch eine Menüzeile und eine Symbolleiste ergänzt.

3.4 Daten-Editor

Wir wollen mit Hilfe des Daten-Editors eine Datendatei erstellen. Der Daten-Editor ist ein spreadsheet-ähnliches Arbeitsmittel. Unter Spreadsheet versteht man ein in Zeilen und Spalten aufgeteiltes Arbeitsblatt, welches eine einfache und effiziente Dateneingabe ermöglicht. Die einzelnen Zeilen entsprechen den einzelnen Fällen. In einer Umfrage enthält eine Zeile z. B. die Angaben einer befragten Person. Die einzelnen Spalten entsprechen den einzelnen Variablen. Bei einem Fragebogen speichert eine Variable die Antworten auf eine einzelne Frage. Die einzelnen Zellen enthalten die Werte der jeweiligen Variablen des jeweiligen Falles; jede Zelle speichert einen einzelnen Variablenwert.

3.4.1 Definition der Variablen

Beginnen wir mit der Definition von Variablen. Eine Variable können Sie wie folgt definieren:

- Doppelklicken Sie in der Datenansicht auf eine mit var betitelte Zelle, oder klicken Sie auf die Registerkarte *Variablenansicht* am unteren Rand des Spreadsheets.

In beiden Fällen sehen Sie die Variablenansicht des Daten-Editors (siehe Bild 3.2).

Sie können nunmehr damit beginnen, zeilenweise die einzelnen Variablen zu definieren.

Variablenname

Zur Vergabe eines Variablennamens gehen Sie wie folgt vor:

- Geben Sie in das Textfeld *Name* den gewünschten Variablennamen ein. Bezogen auf unseren Beispieldatensatz wollen wir zunächst die Variable *fragebnr* definieren. Geben Sie daher »fragebnr« in das Textfeld *Name* ein.

Die Vergabe von Variablennamen unterliegt bestimmten Festlegungen. Die folgenden Regeln müssen beachtet werden.

- ▶ Variablennamen können aus Buchstaben und Ziffern gebildet werden. Erlaubt sind ferner die Sonderzeichen *_* (underscore), *.* (Punkt) sowie die Zeichen *@* und *#*. Nicht erlaubt sind z. B. Leerzeichen, Umlaute sowie spezifische Zeichen, wie *!*, *?*, *»* und ***.
- ▶ Der Variablenname muss mit einem Buchstaben beginnen.

- ▶ Das letzte Zeichen eines Variablennamens darf kein Punkt und sollte kein _ (underscore) sein.
- ▶ Der Variablenname darf nicht länger als acht Zeichen sein.
- ▶ Variablennamen sind nicht case-sensitive, d.h., die Groß- und Kleinschreibung ist nicht relevant.

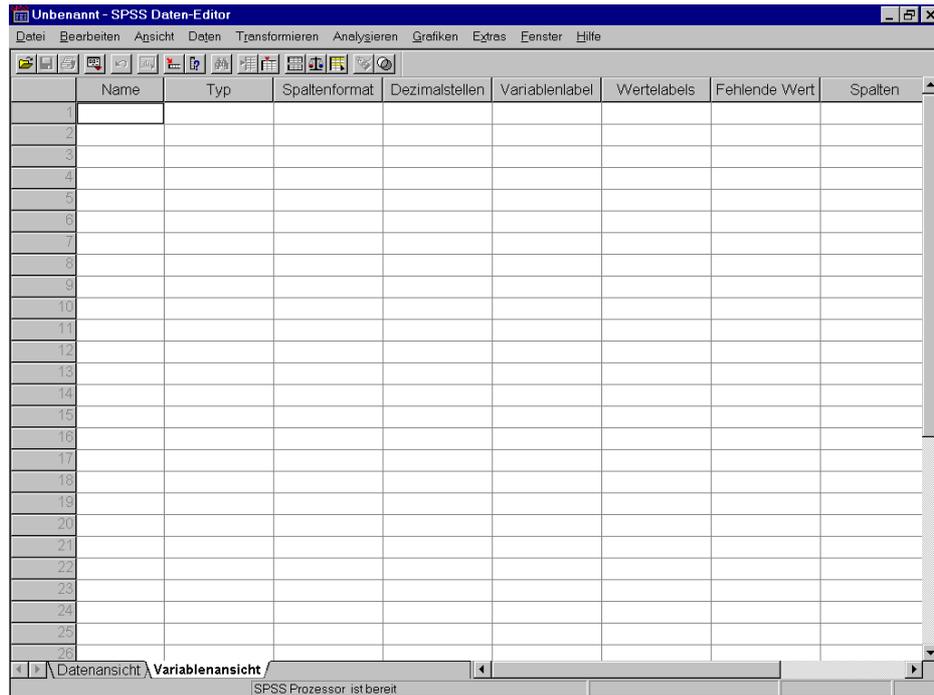


Bild 3.2: Daten-Editor: Variablenansicht

Beispiele für gültige Variablennamen:

haushalt
sex
gehalt93
frage_13
var3_1_2

Beispiele für ungültige Variablennamen:

1mal1	Name beginnt nicht mit einem Buchstaben.
Geschlecht	Name ist länger als acht Zeichen.
Stand 94	Name enthält ein Leerzeichen.
Wagen!	Das Sonderzeichen »!« ist nicht erlaubt.

Umlaute und »ß« sind zwar mittlerweile in Variablenamen zugelassen, allerdings sind dann die betreffenden Datendateien in älteren Programmversionen nicht problemlos zu verwenden.

- Drücken Sie die Tabulator-Taste, um Ihre Eingabe zu bestätigen und zur Festlegung des Variablentyps überzugehen.

Variablentyp

Neu angelegte Variablen sind, wie Sie dem Spreadsheet entnehmen können, bei SPSS per Voreinstellung numerisch mit der Maximallänge von acht Zeichen, wobei zwei Nachkommastellen voreingestellt sind (Format 8.2).

- Wollen Sie den Variablentyp ändern, so klicken Sie in der Zelle auf die Schaltfläche mit den drei Punkten:



- Es öffnet sich die Dialogbox *Variablentyp definieren*.

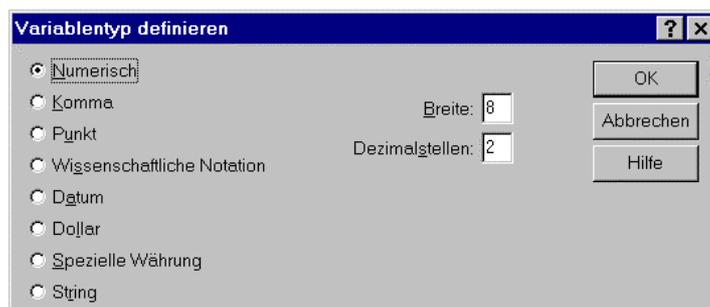


Bild 3.3: Dialogbox Variablentyp definieren (numerische Variable)

SPSS kennt folgende Variablentypen:

Numerisch	Gültige Werte enthalten Ziffern, ein vorangestelltes Plus- oder Minuszeichen und ein Dezimaltrennzeichen. Im Unterschied zum Minuszeichen wird das vorangestellte Pluszeichen nicht angezeigt. Im Textfeld <i>Breite</i> wird die maximale Anzahl der Zeichen eingetragen, einschließlich einer Position für das Dezimaltrennzeichen. Im Textfeld <i>Dezimalstellen</i> wird die Anzahl der Dezimalstellen angegeben, die angezeigt werden sollen.
Komma	Gültige Werte enthalten Ziffern, ein vorangestelltes Plus- oder Minuszeichen, einen Punkt als Dezimaltrennzeichen sowie ein oder mehrere eingebettete Kommas als Tausendertrennzeichen. Werden die Kommas bei der Dezimaleingabe weggelassen, werden sie automatisch eingefügt. Als Länge wird die maximale Zeichenzahl eingegeben, einschließlich des Dezimaltrennzeichens sowie der eingebetteten Kommas.

Punkt	Gültige Werte enthalten Ziffern, ein vorangestelltes Plus- oder Minuszeichen, ein Komma als Dezimaltrennzeichen und einen oder mehrere eingebettete Punkte als Tausendertrennzeichen. Werden Punkte bei der Dateneingabe weggelassen, werden sie automatisch eingefügt.
Wissenschaftliche Notation	Erlaubt sind bei der Dateneingabe alle gültigen numerischen Werte einschließlich der wissenschaftlichen Notation, die durch ein eingebettetes E, D, Plus- oder Minuszeichen gekennzeichnet ist.
Datum	Gültige Werte sind Datums- und/oder Zeitangaben.
Dollar	Gültige Werte enthalten ein Dollarzeichen, einen Punkt als Dezimaltrennzeichen und Kommas als Tausendertrennzeichen. Werden bei der Dateneingabe das Dollarzeichen oder die Kommas weggelassen, werden sie automatisch eingefügt.
Spezielle Währung	Der Benutzer hat die Möglichkeit, eigene Währungsformate festzulegen. Bei <i>Breite</i> wird dann die maximale Zeichenanzahl eingegeben, einschließlich aller benutzerspezifischen Zeichen. Die Währungsbezeichnung darf bei der Dateneingabe nicht eingetippt werden; sie wird automatisch eingefügt.
String	Zeichenkette. Gültige Werte umfassen Buchstaben, Ziffern und Sonderzeichen. Stringvariablen dürfen bis zu 255 Zeichen lang sein.

Folgende Besonderheiten bei der Eingabe und Ausgabe von Daten müssen beachtet werden:

- ▶ *Numerische Formate:* Bei numerischen Formaten kann das Dezimaltrennzeichen entweder ein Punkt oder ein Komma sein. Die Art des Dezimaltrennzeichens ist abhängig von der Spezifikation in der Dialogbox *Ländereinstellungen* der Windows-Systemsteuerung. Der vollständige Wert wird intern gespeichert. Der Daten-Editor zeigt stets nur die definierte Zahl der Dezimalstellen an. Werte mit mehr Dezimalstellen werden gerundet. Für die Berechnungen wird der vollständige Wert verwendet.
- ▶ *String-Formate:* Bei String-Formaten werden alle Werte bis zur maximalen Länge mit Leerzeichen aufgefüllt. Bei einer String-Variablen der Länge 10 wird z.B. der eingegebene Wert »Mueller« intern als »Mueller » abgespeichert.
- ▶ *Datums- und Zeitformate:* Bei Datumsformaten können Schrägstriche, Trennstriche, Leerzeichen, Kommas oder Punkte als Trennzeichen zwischen den Werten für Tag, Monat und Jahr verwendet werden. Es kann zwischen verschiedenen Datumsformaten (tt-mm-jjjj, tt-mmm-jj, mm/tt/jjjj u.a.) gewählt werden. Datumsangaben im Format tt-mmm-jj werden mit Trennstrichen als Trennzeichen und dreibuchstabigen Abkürzungen für Monate angezeigt. Datumsangaben im Format tt/mm/jj und mm/tt/jj werden mit Schrägstrichen als Trennzeichen und Zahlen für Monate angezeigt.
- ▶ Es stehen insgesamt 27 verschiedene Datums- und Zeitformate zur Verfügung, die in einer Dropdown-Liste angezeigt werden. Bei Zeitformaten können als Trennzeichen zwischen Stunden, Minuten und Sekunden Doppelpunkte, Punkte oder Leerzeichen verwendet werden.

- *Spezielle Wahrung*: Die angezeigten Wahrungsformate CCA, CCB, CCC, CCD und CCE werden nach der Menuwahl

Bearbeiten

Optionen...

mit Hilfe der Registerkarte *Wahrung* definiert.

- Legen Sie fur die Variable *fragebnr* den Variablentyp String fest sowie eine Lange von funf Zeichen, und bestatigen Sie mit OK.



Bild 3.4: Dialogbox Variablentyp definieren (String-Variable)

Die Variable *fragebnr* ist eine Variable vom Typ »Zeichenkette«. Mit solchen Variablen konnen keine Rechenoperationen ausgefuhrt werden; moglich sind jedoch z.B. Haufigkeitsauszahlungen. Die Eingabe von Buchstaben wie »W« fur die alten Bundeslander und »O« fur die neuen Bundeslander soll moglich sein. Wir wahlen eine Lange von funf Zeichen, um fur die alten und neuen Bundeslander jeweils maximal 999 Interviews kodieren zu konnen. Bei einer Lange von funf Zeichen konnen fur die alten Bundeslander Interviews mit den Fragebogennummern von »W-001« bis »W-999« kodiert werden sowie fur die neuen Bundeslander von »O-001« bis »O-999«.

- Drucken Sie die Tabulator-Taste, um nunmehr das Spaltenformat festzulegen.

Spaltenformat

- Das Spaltenformat der Variablen *fragebnr* lautet »5«. Der Wert ergibt sich aus der Variablenlange, die in der Dialogbox *Variablentyp definieren* festgelegt wurde.
- Um das auf der Basis der Eintrage in der Dialogbox *Variablentyp definieren* ubernommene Spaltenformat einer Variablen nachtraglich zu andern, klicken Sie auf die Schaltflache



- Bestatigen Sie in diesem Falle die Voreinstellung mit der Tabulator-Taste.

Dezimalstellen

- Da es sich bei der Variablen `fragebnr` um eine Stringvariable handelt, lautet die Dezimalstellenanzahl »0«. Eine Erhöhung bzw. Herabsetzung des aus der Dialogbox *Variablentyp definieren* übernommenen Wertes können Sie auch hier mit Hilfe der Schaltfläche



- vornehmen. Bestätigen Sie den Wert »0« durch Drücken der Tabulator-Taste.

Variablenlabel

Ein Variablenlabel ist ein Etikett, mit dessen Hilfe Sie eine Variable näher beschreiben können. Ein Variablenlabel kann bis zu 256 Zeichen lang sein. Bei einem Variablenlabel wird die Groß- und Kleinschreibung beachtet. Der Eingabe entsprechend werden diese auch angezeigt. Geben Sie als Variablenlabel für die Variable `fragebnr` im Textfeld *Variablenlabel* »Fragebogen-Nr.« ein.

Wertelabels

Wertelabels sind Etiketten, mit deren Hilfe Sie die Merkmalsausprägungen einer Variablen näher beschreiben können. So können Sie zum Beispiel bei der Variablen `sex` für den Wert »1« das Etikett »weiblich«, für den Wert »2« das Etikett »männlich« vergeben. Ein Wertelabel kann bis zu 60 Zeichen lang sein.

Bestätigen Sie bei der Stringvariablen `fragebnr` die Voreinstellung *Kein* mit der Tabulator-Taste. Übrigens können Sie zur Bestätigung Ihrer Eingaben alternativ auch die Return-Taste drücken.

Fehlende Werte

SPSS kennt zwei Arten von fehlenden Werten:

- ▶ *Systemdefinierte fehlende Werte*: Werden in einer Datenmatrix numerische Zahlen nicht ausgefüllt, so weist SPSS ihnen den systemdefinierten fehlenden Wert zu. Dieser wird in der Datenmatrix mit einem Komma (,) angezeigt.
 - ▶ *Benutzerdefinierte fehlende Werte*: Liegen in bestimmten Fällen bei Variablen keine Antwortvorgaben vor, sei es wegen einer Antwortverweigerung, aus Unwissenheit oder anderen Gründen, kann der Benutzer an diesen Stellen im Daten-Editor einen Zahlenwert eintragen und diesen über den Schalter *Fehlende Werte* als fehlend deklarieren. Dieser Zahlenwert darf natürlich unter den auftretenden Werten nicht vorkommen. Fehlende Werte werden aus weiteren Berechnungen ausgeschlossen.
- Bestätigen Sie die Voreinstellung *Kein* mit Hilfe der Return-Taste.

Spalten

Das Feld *Spalten* bezieht sich auf die Breite, die dem jeweiligen Datenfeld im Spreadsheet bei der Anzeige der Werte eingeräumt wird. Die Spaltenbreite kann auch im Fenster des

Daten-Editors verändert werden. Positionieren Sie hierfür den Mauszeiger auf den Rand zwischen zwei Variablennamen. Der Mauszeiger verändert seine Gestalt. Ein Doppelpfeil weist darauf hin, dass die entsprechende Spalte mit Hilfe der Klicken- und Ziehen-Technik vergrößert oder verkleinert werden kann.

- Bestätigen Sie die Voreinstellung »8« durch Drücken der Return-Taste, um den Variablennamen fragebnr als Spaltenüberschrift im Daten-Editor vollständig darstellen zu können.

Ausrichtung

Sie können hier die Ausrichtung der Werte festlegen, d. h., wie diese zukünftig im Spreadsheet angezeigt werden sollen. Zur Verfügung stehen die Optionen Rechts (»rechtsbündig«), Links (»linksbündig«) und Mitte (»zentriert«). Die Festlegung der Art der Anzeige können Sie mit Hilfe der Schaltfläche vornehmen.

- Bestätigen Sie die Voreinstellung *Links* mit Return.

Bei numerischen Variablen ist *Rechts*, bei Stringvariablen *Links* voreingestellt.

Messniveau

Es kann das Messniveau (Skalenniveau) der Variablen angegeben werden, wobei zwischen nominal, ordinal und metrisch unterschieden wird (siehe Kap. 5.1.1). Bei numerischen Variablen ist das Messniveau *Metrisch*, bei Stringvariablen das Messniveau *Nominal* voreingestellt. Allerdings ist diese Unterscheidung nur bei der Erstellung interaktiver Grafiken relevant, wo nominales und ordinales Messniveau zu »kategorial« zusammengefasst wird.

Laden Sie Daten von Dateien, die mit früheren SPSS-Versionen erstellt wurden, oder definieren Sie das Messniveau nicht explizit, so setzt SPSS automatisch zunächst metrisches Messniveau voraus und ordinales Messniveau dann, wenn die betreffende Variable Wertelabels aufweist oder weniger als 24 verschiedene Werte hat.

- Bestätigen Sie die Voreinstellung *Nominal* durch Drücken der Tabulator-Taste. Auf diese Weise positionieren Sie den Cursor auch wieder in das Feld *Name*, um die nächste Variable zu deklarieren.

Wir fahren fort mit der Deklaration der Variablen *sex*.

- Tragen Sie in das Textfeld *Name* *sex* ein, und bestätigen Sie die Eingabe durch Drücken der Return- oder der Tabulator-Taste.
- Klicken Sie zur Festlegung des Variablentyps im Feld *Typ* auf die Schaltfläche mit den drei Punkten. Es öffnet sich die Dialogbox *Variablentyp definieren*. Übernehmen Sie die Voreinstellung *numerisch*, und legen Sie als Breite »1« und als Dezimalstellenanzahl »0« fest, da nur die Werte 1, 2 oder 0 abgespeichert werden sollen. Bestätigen Sie mit *OK*, und betätigen Sie die Tabulator-Taste.
- Übernehmen Sie als Spaltenformat den Wert »1« und als Dezimalstellenanzahl »0«, da es sich um die Werte handelt, die aus den Einstellungen in der Dialogbox *Variablentyp definieren* übernommen wurden. An dieser Stelle könnten Sie die von Ihnen vorgenommenen Festlegungen nachträglich ändern.

- Vergeben Sie als Variablenlabel die Bezeichnung »Geschlecht«.
- Klicken Sie im Feld *Wertelabels* auf die Schaltfläche . Es öffnet sich die Dialogbox *Wertelabels definieren*.

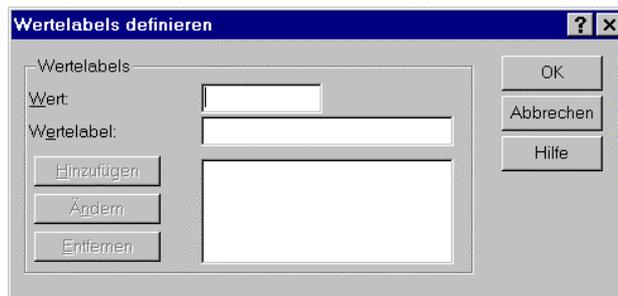


Bild 3.5: Dialogbox Wertelabels definieren

Die Zuweisung der Wertelabels geschieht wie folgt:

- Geben Sie zunächst eine »1« im Textfeld *Wert* ein. Drücken Sie die Tabulator-Taste.
- Tippen Sie in das Wertelabel-Textfeld »weiblich«.
- Klicken Sie auf *Hinzufügen*. Das Wertelabel wird der Liste hinzugefügt. Alternativ zum Klicken auf *Hinzufügen* könnten Sie auch die Tastenkombination <Alt>+<h> drücken.
- Wiederholen Sie die Schritte mit den Merkmalsausprägungen »2« für männlich und »0« für keine Angabe.

Der Inhalt der ausgefüllten Dialogbox *Wertelabels* ist in Bild 3.6 dargestellt.



Bild 3.6: Ausgefüllte Dialogbox Wertelabels definieren

- Bestätigen Sie ihre Angaben mit *OK* und anschließend durch Drücken der Tabulator-Taste.

- Klicken Sie zur Festlegung fehlender Werte im Feld *Fehlende Werte* auf die Schaltfläche mit den drei Punkten . Es öffnet sich die Dialogbox *Fehlende Werte definieren* (siehe Bild 3.7).

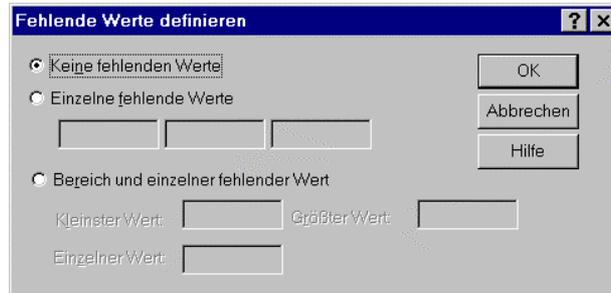


Bild 3.7: Dialogbox Fehlende Werte definieren

Voreingestellt ist die Option *Keine fehlenden Werte*, d.h., alle Werte werden zurzeit als gültig betrachtet.

- Klicken Sie auf die Option *Einzelne fehlende Werte*. Es können bis zu drei einzelne benutzerdefinierte fehlende Werte für eine Variable eingegeben werden. Tragen Sie eine »0« ein.

Es gibt ferner folgende Option:

- ▶ *Bereich und einzelner fehlender Wert*: Alle Werte zwischen den Ausgaben *Kleinsten Wert* und *Größter Wert* einschließlich werden als fehlend deklariert. Festgelegt werden kann ferner ein zusätzlicher *Einzelner Wert* außerhalb des Bereichs.

Fehlende Werte in einer Datenerhebung sind leider in der Regel nicht zu vermeiden. Bei vielen statistischen Verfahren, vor allem bei univariaten, ist die Berücksichtigung fehlender Werte unproblematisch, da sich außer der entsprechenden Verringerung der Fallzahl keine weiteren Konsequenzen ergeben. Bei bivariaten oder gar multivariaten Analysen bringen fehlende Werte in den Variablenlisten größere Probleme mit sich, da ein einziger auftretender fehlender Wert ausreicht, um den ganzen Fall für die betreffende Analyse unbrauchbar zu machen. Bei manchen Analysen werden von SPSS allerdings Auswege aus dieser Situation angeboten.

- Bestätigen Sie den Eintrag fehlender Werte für die Variable *sex* mit *OK*.
- Übernehmen Sie die Voreinstellungen in den Feldern *Spalten* und *Ausrichtung*.
- Klicken Sie im Feld *Messniveau* auf die Schaltfläche, so werden Ihnen drei Skalenniveaus zur Auswahl gestellt:



- Ändern Sie die Voreinstellung *metrisch* auf *nominal*, und drücken Sie die Tabulator-Taste.

Wir fahren fort mit der Deklaration der Variablen *alter*.

- Tragen Sie in das Textfeld *Name* *alter* ein, und bestätigen Sie ihre Eingabe.
- Klicken Sie zur Festlegung des Variablentyps im Feld *Typ* auf die Schaltfläche mit den drei Punkten . Es öffnet sich die Dialogbox *Variablentyp definieren*. Übernehmen Sie die Voreinstellung *numerisch*, und legen Sie als Breite »2« (unterstellt wird dabei, dass kein Befragter älter ist als 99 Jahre) und »0« Dezimalstellen fest. Bestätigen Sie mit *OK*, und drücken Sie anschließend die Tabulator-Taste.
- Übernehmen Sie die Angaben in den Feldern *Spaltenformat* und *Dezimalstellen*.
- Vergeben Sie als Variablenlabel die Bezeichnung »Lebensalter«, und bestätigen Sie die Voreinstellung *Kein* bei Wertelabels mit *Return*.
- Klicken Sie zur Festlegung fehlender Werte im Feld *Fehlende Werte* auf die Schaltfläche mit den drei Punkten . Es öffnet sich die Dialogbox *Fehlende Werte definieren*. Voreingestellt ist die Option *Keine fehlenden Werte*, d.h., alle Werte werden zur Zeit als gültig betrachtet. Tragen Sie als einzelnen fehlenden Wert die »0« ein, und bestätigen Sie mit *OK*.
- Übernehmen Sie die Voreinstellung »8« im Feld *Spalten*, »Rechts« im Feld *Ausrichtung* und »Metrisch« im Feld *Messniveau*.

Wir schließen die Erstellung der Datenmaske mit der Deklaration der Variablen *partei* ab.

- Tragen Sie in das Textfeld *Name* *partei* ein, und bestätigen Sie ihre Eingabe durch Drücken der Tabulator-Taste.
- Klicken Sie zur Festlegung des Variablentyps im Feld *Typ* auf die Schaltfläche mit den drei Punkten. Es öffnet sich die Dialogbox *Variablentyp definieren*. Übernehmen Sie die Voreinstellung *numerisch*, und legen Sie als Breite »1« und als Dezimalstellenanzahl »0« fest, da nur die Werte 1 bis 7 und die Null als fehlender Wert abgespeichert werden sollen. Bestätigen Sie mit *OK*, und betätigen Sie die Tabulator-Taste.
- Übernehmen Sie als Spaltenformat den Wert »1« und als Dezimalstellenanzahl »0«.
- Vergeben Sie als Variablenlabel die Bezeichnung »Partei«.
- Klicken Sie im Feld *Wertelabels* auf die Schaltfläche . Es öffnet sich die Dialogbox *Wertelabels definieren* (siehe Bild 3.5).
- Geben Sie zunächst eine »1« im Textfeld *Wert* ein. Drücken Sie die Tabulator-Taste.
- Tippen Sie in das Wertelabel-Textfeld »CDU/CSU«.
- Klicken Sie auf *Hinzufügen*. Das Wertelabel wird der Liste hinzugefügt.
- Wiederholen Sie die Schritte mit den Merkmalsausprägungen »2« für FDP, »3« für SPD, »4« für Grüne/Bündnis90, »5« für PDS, »6« für Republikaner, »7« für Sonstige und »0« für keine Angabe.

Der Inhalt der ausgefüllten Dialogbox *Wertelabels definieren* ist in Bild 3.8 dargestellt.



Bild 3.8: Ausgefüllte Dialogbox Wertelabels definieren für die Variable *partei*

- Bestätigen Sie ihre Angaben mit *OK* und anschließend durch Drücken der Tabulator-Taste.
- Klicken Sie zur Festlegung fehlender Werte im Feld *Fehlende Werte* auf die Schaltfläche mit den drei Punkten . Es öffnet sich die Dialogbox *Fehlende Werte definieren*. Klicken Sie auf die Option *Einzelne fehlende Werte*, und tragen Sie hier eine »0« ein.
- Bestätigen Sie den Eintrag fehlender Werte für die Variable *partei* mit *OK*, und drücken Sie die Tabulator-Taste.
- Übernehmen Sie die Voreinstellungen in den Feldern *Spalten* und *Ausrichtung*.
- Klicken Sie im Feld *Messniveau* auf die Schaltfläche mit den drei Punkten , und ändern Sie die Voreinstellung von *metrisch* auf *nominal* ab.

3.4.2 Dateneingabe

Beginnen wir mit der Eingabe der Daten:

	fragebnr	sex	alter	partei
1	W-001	1	45	1
2	W-002	2	22	3
3	W-003	2	19	3
4	W-004	1	42	1
5	W-005	2	34	4
6	W-006	1	72	2
7	W-007	2	38	3
8	W-008	1	56	3
9	W-009	2	61	1
10	W-010	1	77	1
11	W-011	1	23	4

	fragebnr	sex	alter	partei
12	W-012	2	67	6
13	W-013	2	79	7
14	W-014	1	26	3
15	W-015	2	59	1
16	O-001	1	34	4
17	O-002	2	18	6
18	O-003	1	44	1
19	O-004	2	68	1
20	O-005	1	33	5
21	O-006	2	66	1
22	O-007	1	22	0
23	O-008	2	0	3
24	O-009	1	67	3
25	O-010	2	33	2
26	O-011	2	44	1
27	O-012	1	22	3
28	O-013	1	19	7
29	O-014	1	55	1
30	O-015	2	39	3

Daten können fallweise oder variablenweise eingegeben werden. Gehen Sie wie folgt vor:

- Klicken Sie auf die erste Zelle links oben. Es erscheint ein Rahmen um die Zelle. Die Zelle wird so als aktive Zelle gekennzeichnet.
- Tippen Sie den Wert ein, in unserem Beispiel »W-001«. Der Wert wird im Zellen-Editor am oberen Rand des Daten-Editor-Fensters angezeigt.
- Drücken Sie die Tabulator-Taste. Der Datenwert aus dem Zellen-Editor erscheint in der Zelle.
- Bei fallweiser (zeilenweiser) Eingabe bestätigen Sie mit der Tabulator-Taste, bei variablenweiser (spaltenweiser) Eingabe mit der Eingabe-Taste.

Die folgenden Tabellen zeigen Ihnen, welche Tasten welche Funktion innerhalb des Daten-Editors haben, wobei im Folgenden davon ausgegangen wird, dass die Datenansicht aktiviert ist.

Positionieren

Taste	Funktion
 oder 	Positioniert den Cursor eine Zelle nach rechts.
 oder 	Positioniert den Cursor eine Zelle tiefer.
	Positioniert den Cursor eine Zelle höher
  oder 	Positioniert den Cursor eine Zelle nach links, d.h. auf das vorherige Feld.
Pos1	Positioniert den Cursor auf die erste Zelle einer Zeile bzw. eines Falles
Ende	Positioniert den Cursor auf die letzte Zelle eines Falles
Strg 	Positioniert den Cursor auf den ersten Fall einer Spalte
Strg 	Positioniert den Cursor auf den letzten Fall einer Spalte
Strg Pos1	Positioniert den Cursor auf die erste Zelle des ersten Falles
Strg Ende	Positioniert den Cursor auf die letzte Zelle des letzten Falles
Bild 	Vollzieht einen Bildlauf nach oben um eine Seite
Bild 	Vollzieht einen Bildlauf nach unten um eine Seite

Markieren

Taste	Funktion
 <input type="text"/>	Markiert die ganze Zeile.
Strg <input type="text"/>	Markiert die gesamte Spalte.
    	Auswahl eines Bereichs von Fällen und Variablen. Alternativ: Klicken und Ziehen der Maus von der oberen linken Ecke bis zur rechten unteren Ecke.

Editieren

Taste	Funktion
F2	Schaltet in den Editiermodus um. Ein erneutes Drücken von <F2> schaltet den Editiermodus wieder aus.
	ein Zeichen zum Editieren innerhalb einer Zelle nach rechts bewegen
	ein Zeichen zum Editieren innerhalb einer Zelle nach links bewegen
Pos1	auf den Anfang des Wertes einer Zelle positionieren
Ende	auf das Ende des Wertes einer Zelle positionieren

3.5 Speichern einer Datendatei

Wir wollen nun die Daten speichern. Gehen Sie wie folgt vor:

- Wählen Sie aus dem Menü

Datei

Speichern unter...

Es öffnet sich die Dialogbox *Daten speichern unter*.

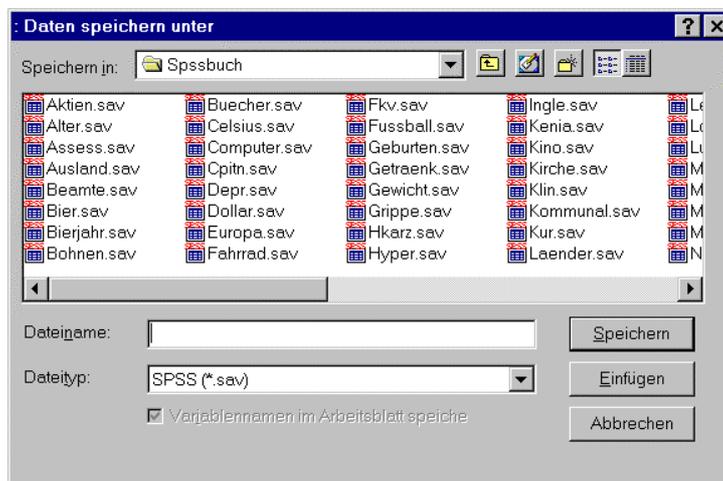


Bild 3.9: Dialogbox Daten speichern unter

SPSS speichert per Voreinstellung die Datendatei im aktuellen Verzeichnis mit der Erweiterung *.sav*. Sind Sie den Installationsanweisungen für die Übungdiskette gefolgt und haben Sie ein Arbeitsverzeichnis namens *\SPSSBUCH* eingerichtet, so ist dieses Verzeichnis voreingestellt.

- Vergeben Sie einen Dateinamen, der die DOS-Namenskonventionen beachtet. Wir schlagen den Dateinamen *btwahl.sav* vor. Die Endung *.sav* vergibt SPSS standardmäßig für Datendateien. Sie brauchen die Extension *.sav* daher nicht einzugeben.

3.6 Variablendeklarationen kopieren

- In einer Studie zum Thema »Gesundheit und Ernährung« wird u. a. nach den Vorlieben für bestimmte Ernährungsarten gefragt. Die Probanden sollten beantworten, inwieweit eine der folgenden Aussagen auf sie zutrifft.

<i>Ich ernähre mich in erster Linie</i>	trifft sehr zu 1	trifft zu 2	trifft zum Teil zu 3	trifft weniger zu 4	trifft nicht zu 5
vegetarisch	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
biologisch- dynamisch	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
fleischarm	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
fast-food	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
kalorienbewusst	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
sättigend	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
preiswert	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

- Da in diesem Falle die Deklaration der sieben anzulegenden Variablen identisch ist, sollten Sie sich auf jeden Fall Arbeit ersparen, indem Sie die Deklarationsangaben für die erste Variable auf die anderen sechs übertragen. Gehen Sie hierzu wie folgt vor.
- Aktivieren Sie die Variablenansicht des Daten-Editors, tragen Sie in das Textfeld *Name* vegetar ein, und bestätigen Sie ihre Eingabe durch Drücken der Tabulator-Taste.
- Klicken Sie zur Festlegung des Variablentyps im Feld *Typ* auf die Schaltfläche mit den drei Punkten. Es öffnet sich die Dialogbox *Variablentyp definieren*. Übernehmen Sie die Voreinstellung *numerisch*, und legen Sie als Breite »1« und als Dezimalstellenanzahl »0« fest, da nur die Werte 1 bis 5 und die Null als fehlender Wert abgespeichert werden sollen. Bestätigen Sie mit *OK*, und betätigen Sie die Tabulator-Taste.
- Übernehmen Sie als Spaltenformat den Wert »1« und als Dezimalstellenanzahl »0«.
- Vergeben Sie als Variablenlabel die Bezeichnung »vegetarisch«.
- Klicken Sie im Feld *Wertelabels* auf die Schaltfläche . Es öffnet sich die Dialogbox *Wertelabels definieren*.
- Geben Sie zunächst eine »1« im Textfeld *Wert* ein. Drücken Sie die Tabulator-Taste.
- Tippen Sie in das Wertelabel-Textfeld »trifft sehr zu«.
- Klicken Sie auf *Hinzufügen*. Das Wertelabel wird der Liste hinzugefügt.
- Wiederholen Sie die Schritte mit den Merkmalsausprägungen »2« für »trifft zu«, »3« für »trifft zum Teil zu«, »4« für »trifft weniger zu«, »5« für »trifft nicht zu« und »0« für »keine Angabe«.
- Bestätigen Sie ihre Angaben mit *OK* und anschließend durch Drücken der Tabulator-Taste.
- Klicken Sie zur Festlegung fehlender Werte im Feld *Fehlende Werte* auf die Schaltfläche mit den drei Punkten . Es öffnet sich die Dialogbox *Fehlende Werte definieren*. Klicken Sie auf die Option *Einzelne fehlende Werte*, und tragen Sie hier eine »0« ein.

- Bestätigen Sie den Eintrag fehlender Werte für die Variable *vegetar* mit *OK*, und drücken Sie die *Tabulator-Taste*.
- Übernehmen Sie die Voreinstellungen in den Feldern *Spalten und Ausrichtung*.
- Definieren Sie das Messniveau als *Ordinal*.
- Positionieren Sie nun den Cursor in die Zelle mit der Nummer 1, d.h. ganz an den Anfang der ersten Zeile, und drücken Sie die linke Maustaste. Die Deklarationsangaben der ersten Variablen sind nun markiert.



Bild 3.10: Markierte Deklaration

- Wählen Sie aus dem Menü

Bearbeiten
Kopieren

- Positionieren Sie den Cursor in die Zelle mit der Nummer 2, d.h. ganz an den Anfang der zweiten Zeile, und drücken Sie die linke Maustaste, so dass die zweite Zeile markiert ist.
- Wählen Sie aus dem Menü

Bearbeiten
Einfügen

Die Deklarationsangaben der ersten Variablen sind nunmehr kopiert worden.



Bild 3.11: Kopierte Deklarationsangaben

- Ändern Sie noch den voreingestellten Variablennamen *var00001* in *biolog* ab, aktualisieren Sie das Variablenlabel, und verfahren Sie mit den übrigen Variablen in gleicher Weise.

- Nach dem Markieren und Kopieren der Variablendeklaration und der Markierung der zweiten Zeile mit der linken Maustaste könnten Sie anstelle der Menüoption

Bearbeiten
Einfügen

auch die Option

Bearbeiten
Variablen einfügen...

wählen.

Es öffnet sich in diesem Falle die Dialogbox *Variablen einfügen*.

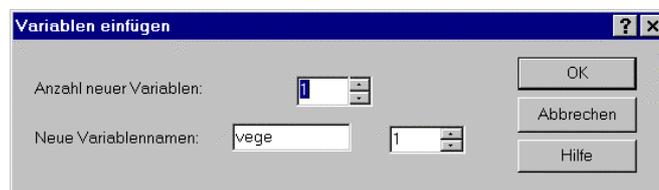


Bild 3.12: Dialogbox Variablen einfügen

- Ersetzen Sie den Vorschlag *vege* für den neuen Variablennamen durch *biolog*, und bestätigen Sie mit *OK*.

Die Dialogbox *Variablen einfügen* (siehe Bild 3.12) bietet Ihnen die Option, die Anzahl der neuen Variablen festzulegen. Sie könnten hier die Anzahl auf 6 festlegen, um die Datendeklaration der Variablen *vegetar* gewissermaßen auf einem Schlag zu übernehmen. In diesem Falle sehe dann das Spreadsheet wie folgt aus:

Bild 3.13: Kopierte Variablendefinitionen

Sie müssten noch die Variablennamen *vege1* bis *vege6* durch die von Ihnen gewünschten ersetzen, wie z. B. *biolog*, *fleisarm*, *fastfood*, *kaloribew*, *satt* und *billig*, ferner die Variablenlabels aktualisieren, und hätten dann bereits alle Variablen der Skala »Ernährungsvorlieben« deklariert.

3.7 Zusammenfügen von Datendateien

In SPSS besteht die Möglichkeit, zwei Datendateien (bei wiederholter Anwendung also auch mehrere) Datendateien zusammenzuführen. Dabei gibt es prinzipiell zwei Möglichkeiten:

- ▶ Beim fallweisen Zusammenfügen werden zwei Datendateien zusammengeführt, die unterschiedliche Fälle haben, deren Variablen aber zumindest teilweise gleiche inhaltliche Bedeutung haben. Im einfachsten Fall sind die Variablen beider Dateien identisch.
- ▶ Beim variablenweisen Zusammenfügen werden zwei Datendateien zusammengeführt, die unterschiedliche Variablen haben, deren Fälle aber zumindest teilweise übereinstimmen. Im einfachsten Fall sind die Fälle beider Dateien identisch.

Beide Möglichkeiten werden im Folgenden anhand kleiner und daher überschaubarer Dateien erläutert.

3.7.1 Fallweises Zusammenfügen

Beim fallweisen Zusammenfügen von Datendateien wollen wir drei Varianten unterscheiden:

- ▶ Die beiden Dateien haben identische Variablen.
- ▶ Die beiden Dateien haben teilweise unterschiedliche Variablen. Variablen mit gleicher inhaltlicher Bedeutung haben in beiden Dateien identische Variablenamen.
- ▶ Die beiden Dateien haben teilweise unterschiedliche Variablen. Variablen mit gleicher inhaltlicher Bedeutung haben zumindest teilweise unterschiedliche Variablenamen.

Diese drei Varianten werden nun anhand kleiner und damit überschaubarer Datendateien erläutert.

Identische Variablen

Dieser einfachste Fall liegt dann vor, wenn die Dateneingabe aus Gründen der Arbeitsteilung von zwei Personen vorgenommen wurde, welche die Daten unter der gleichen Variablenmaske in zwei verschiedene Dateien eingegeben haben. Dazu betrachten wir die beiden Dateien `mix1.sav` und `mix2.sav`, die wir zu einer gemeinsamen Datei vereinigen wollen.

	nr	alter	geschl	gr	gew
1	1,00	56,00	1,00	180,00	78,00
2	2,00	67,00	2,00	167,00	65,00
3	3,00	31,00	2,00	159,00	67,00
4	4,00	22,00	1,00	194,00	95,00

	nr	alter	geschl	gr	gew
1	1,00	45,00	2,00	160,00	58,00
2	2,00	78,00	1,00	184,00	77,00
3	3,00	33,00	1,00	177,00	94,00

Bild 3.14: Dateien mit identischen Variablen

- Laden Sie die Datei mix1.sav mit Hilfe der Menüwahl

Datei
Öffnen
Daten...

- Um die Datei mix2.sav anzufügen, wählen Sie aus dem Menü

Daten
Dateien zusammenfügen
Fälle hinzufügen...

Es öffnet sich die Dialogbox *Fälle hinzufügen: Datei lesen*.

- Markieren Sie in der Dateienliste die Datei mix2.sav, und betätigen Sie den Schalter *Öffnen*.
- Es erscheint die Dialogbox *Fälle hinzufügen aus*.



Bild 3.15: Dialogbox Fälle hinzufügen aus

Da die Variablen in beiden Dateien »eins zu eins« übereinstimmen, sind alle Variablen im Feld *Variablen in der neuen Arbeitsdatei* aufgeführt, während das Feld *Nicht gepaarte Variablen* leer ist.

Beide Dateien haben bei eins beginnende fortlaufende Fallnummern, so dass es aus Gründen der Eindeutigkeit wünschenswert erscheint, in der zusammengefassten Datei eine Kennzeichnung der ursprünglichen Dateien anzufügen.

- Aktivieren Sie daher die Option *Datei-Indikator als Variable*; den vorgeschlagenen Variablenamen *quelle01* können Sie dabei nach Ihren Wünschen abändern. Bestätigen Sie mit *OK*.
- Die zusammengefasste Datei hat das in Bild 3.16 dargestellte Aussehen.

	nr	alter	geschl	gr	gew	quelle01
1	1,00	56,00	1,00	180,00	78,00	0
2	2,00	67,00	2,00	167,00	65,00	0
3	3,00	31,00	2,00	159,00	67,00	0
4	4,00	22,00	1,00	194,00	95,00	0
5	1,00	45,00	2,00	160,00	58,00	1
6	2,00	78,00	1,00	184,00	77,00	1
7	3,00	33,00	1,00	177,00	94,00	1

Bild 3.16: Zusammengefasste Datei

Die Variable `quelle01` als Datei-Indikator hat den Wert 0 für die zuerst geladene Datei `mix1.sav` und den Wert 1 für die daran angehängte Datei `mix2.sav`.

Teilweise unterschiedliche, sonst aber identische Variablen

Die beiden zusammenfassenden Dateien können auch unterschiedliche Variablen enthalten. Einige Variablen sollten aber übereinstimmen, sonst ist eine Zusammenfassung sinnlos. In der zunächst vorgestellten Variante stimmen diese Variablen sowohl in ihrer inhaltlichen Bedeutung als auch in ihrem Variablennamen überein. Als Beispiel betrachten wir die beiden Dateien `mix1.sav` und `mix3.sav`.

	nr	geschl	gr	gew	chol	bz
1	1,00	2,00	170,00	67,00	234,00	115,00
2	2,00	1,00	182,00	78,00	160,00	100,00
3	3,00	2,00	164,00	56,00	183,00	97,00

	nr	alter	geschl	gr	gew
1	1,00	56,00	1,00	180,00	78,00
2	2,00	67,00	2,00	167,00	65,00
3	3,00	31,00	2,00	159,00	67,00
4	4,00	22,00	1,00	194,00	95,00

Bild 3.17: Dateien mit teilweise unterschiedlichen Variablen

- Laden Sie zunächst die Datei `mix1.sav`, und öffnen Sie über die Menüwahl

Daten

Dateien zusammenfügen

Fälle hinzufügen...

die Dialogbox *Fälle hinzufügen: Datei lesen*. Markieren Sie die Datei `mix3.sav` und bestätigen Sie mit *Öffnen*.

Die Dialogbox *Fälle hinzufügen aus* stellt sich nun dar wie in Bild 3.18 wiedergegeben.



Bild 3.18: Dialogbox Fälle hinzufügen aus

Beiden Dateien gemeinsam sind die Variablen nr, geschl, gr und gew, die daher im Feld *Variablen in der neuen Arbeitsdatei* aufgeführt sind. Nur in der ursprünglich geladenen Datei mix1.sav (Arbeitsdatei genannt) enthalten ist die Variable alter, die daher mit einem * markiert und im Feld *Nicht gepaarte Variablen* zu finden ist. Lediglich in der anzuhängenden Datei mix3.sav enthalten sind die Variablen bz und chol; sie sind mit einem + markiert.

- Markieren Sie diese drei Variablen im Feld *Nicht gepaarte Variablen*, und verschieben Sie diese mit Hilfe der Pfeiltaste in das Feld *Variablen in der neuen Arbeitsdatei*. Bestätigen Sie mit **OK**.

Die sich ergebende Gesamtdatei ist in Bild 3.19 dargestellt.

	nr	alter	geschl	gr	gew	chol	bz
1	1,00	56,00	1,00	180,00	78,00	.	.
2	2,00	67,00	2,00	167,00	65,00	.	.
3	3,00	31,00	2,00	159,00	67,00	.	.
4	4,00	22,00	1,00	194,00	95,00	.	.
5	1,00	.	2,00	170,00	67,00	234,00	115,00
6	2,00	.	1,00	182,00	78,00	160,00	100,00
7	3,00	.	2,00	164,00	56,00	183,00	97,00

Bild 3.19: Zusammengefasste Datei

Die jeweils nicht vorhandenen Variablenwerte sind als fehlende Werte markiert.

Variablen mit gleicher Bedeutung, aber unterschiedlichen Namen

Falls die beiden Dateien Variablen mit gleicher inhaltlicher Bedeutung, aber verschiedenen Variablennamen enthalten, so können diese Dateien ebenfalls zusammengeführt werden. Als Beispiel betrachten wir die beiden Dateien mix2.sav und mix4.sav.

	nr	alter	geschl	gr	gew
1	1,00	45,00	2,00	160,00	58,00
2	2,00	78,00	1,00	184,00	77,00
3	3,00	33,00	1,00	177,00	94,00

	nr	geschl	kgr	kgew	chol	bz
1	1,00	2,00	170,00	67,00	234,00	115,00
2	2,00	1,00	182,00	78,00	160,00	100,00
3	3,00	2,00	164,00	56,00	183,00	97,00

Bild 3.20: Dateien mit Variablen gleicher Bedeutung

Die erste Datei (mix2.sav) enthält u.a. die beiden Variablen gr und gew für die Körpergröße bzw. das Körpergewicht. In der zweiten Datei (mix4.sav) heißen diese beiden Variablen kgr bzw. kgew.

- Laden Sie die Datei mix2.sav, und öffnen Sie über die Menüwahl

Daten

Dateien zusammenfügen
Fälle hinzufügen...

- die Dialogbox *Fälle hinzufügen: Datei lesen*.
- Markieren Sie in der Dateienliste die Datei mix4.sav, und bestätigen Sie mit *Öffnen*.

Es öffnet sich die Dialogbox *Fälle hinzufügen aus*.

- Verschieben Sie die Variablen alter, bz und chol, die in nur jeweils einer Datei enthalten sind, mit Hilfe der Pfeiltaste in das Feld *Variablen in der neuen Arbeitsdatei*.

Aus den verbleibenden vier Variablen sind zwei Variablenpaare zu bilden, und zwar gr und kgr für die Körpergröße sowie gew und kgew für das Körpergewicht.

- Um zunächst das Variablenpaar gr – kgr zu bilden, markieren Sie im Feld *Nicht gepaarte Variablen* die Variable gr und drücken dann die Strg-Taste. Bei gedrückter Strg-Taste markieren Sie die Variable kgr und verschieben dann mit Hilfe des Schalters *Paar* dieses Variablenpaar in das Feld *Variablen in der neuen Arbeitsdatei*. Mit den beiden Variablen gew und kgew verfahren Sie entsprechend.
- Bestätigen Sie mit *OK*.

Die zusammengefasste Datei ist in Bild 3.21 dargestellt.

	nr	alter	geschl	gr	gew	chol	bz
1	1,00	45,00	2,00	160,00	58,00	.	.
2	2,00	78,00	1,00	184,00	77,00	.	.
3	3,00	33,00	1,00	177,00	94,00	.	.
4	1,00	.	2,00	170,00	67,00	234,00	115,00
5	2,00	.	1,00	182,00	78,00	160,00	100,00
6	3,00	.	2,00	164,00	56,00	183,00	97,00

Bild 3.21: Zusammengefasste Datei

Für die Körpergröße und das Körpergewicht werden die Variablennamen aus der zuerst geladenen Datei (Arbeitsdatei) verwendet.

3.7.2 Variablenweises Zusammenfügen

Auch beim variablenweisen Zusammenfügen von Datendateien gibt es drei Varianten:

- Jedem Fall der einen Datei entspricht genau ein Fall der anderen Datei. Dabei stimmt die Reihenfolge der Fälle in beiden Dateien genau überein. Eine so genannte Schlüsselvariable ist nicht notwendig.
- Jedem Fall der einen Datei entspricht höchstens ein Fall der anderen Datei. Die einzelnen Fälle werden über eine Schlüsselvariable zugeordnet, die in beiden Dateien enthalten sein muss.
- Jedem Fall der einen Datei (der so genannten Schlüsseldatei) können mehrere Fälle der anderen Datei zugeordnet werden. Auch hier werden die einzelnen Fälle über eine Schlüsselvariable zugeordnet, die in beiden Dateien enthalten sein muss.

Diese drei Varianten sollen nun erläutert werden, wobei aus Gründen der Durchsichtigkeit absichtlich wieder kleine Dateien verwendet werden.

Zusammenfügen ohne Schlüsselvariable

Die Variablen beider Dateien werden zeilenweise eins zu eins zugeordnet. Dieses Vorgehen ohne die Verwendung einer Schlüsselvariablen, sozusagen »ohne Netz und doppelten Boden«, ist insbesondere bei größeren Dateien nicht empfehlenswert, da eine einzige Störung der richtigen Reihenfolge die gesamte Zuordnung fehlerhaft werden lässt.

Ist eine Datei länger als die andere, so werden die überschießenden Zeilen angehängt, wobei für die Variablenwerte der kürzeren Datei fehlende Werte eingesetzt werden. Als Beispiel zweier zusammenzufügender Dateien betrachten wir die Dateien mix5.sav und mix6.sav.

	geschl	alter
1	2,00	66,00
2	2,00	58,00
3	1,00	47,00
4	2,00	52,00

	gr	gew
1	166,00	70,00
2	163,00	63,00
3	173,00	70,00
4	165,00	77,00
5	178,00	82,00
6	161,00	62,00

Bild 3.22: Dateien ohne Schlüsselvariablen

- Um diese beiden Dateien zusammenzufügen, laden Sie zunächst die Datei mix5.sav in den Dateneditor.
- Wählen Sie dann aus den Menüs

Daten

Dateien zusammenfügen

Variablen hinzufügen...

Es öffnet sich die Dialogbox *Variablen hinzufügen: Datei lesen*.

- Markieren Sie die Datei mix6.sav und bestätigen Sie mit dem Schalter *Öffnen*.
- Es erscheint die Dialogbox *Variablen hinzufügen aus*.

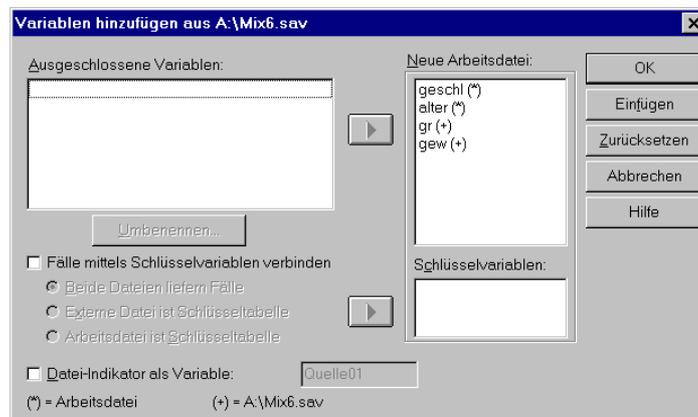


Bild 3.23: Dialogbox Variablen hinzufügen aus

Alle Variablen sind im Feld *Neue Arbeitsdatei* aufgeführt, wobei die Variablen *geschl* und *alter* der zunächst geladenen Datei (Arbeitsdatei) mit einem *, die Variablen *gr* und *gew* der hinzuzufügenden Datei mit einem + gekennzeichnet sind. Das Feld *Ausgeschlossene Variablen* ist dementsprechend leer; hier würden Variablen aufgeführt, die in beiden Dateien enthalten sind.

- Bestätigen Sie mit *OK*.
- Die zusammengefasste Datei ist in Bild 3.24 dargestellt.

	geschl	alter	gr	gew
1	2,00	66,00	166,00	70,00
2	2,00	58,00	163,00	63,00
3	1,00	47,00	173,00	70,00
4	2,00	52,00	165,00	77,00
5	.	.	178,00	82,00
6	.	.	161,00	62,00

Bild 3.24: Zusammengefasste Datei

Empfehlenswerter ist die Verwendung von Schlüsselvariablen, da hier die Richtigkeit der Zusammenführung der beiden Dateien sichergestellt ist.

Zusammenfügen mit Schlüsselvariable

Als Beispiel betrachten wir die beiden Dateien mix7.sav und mix8.sav, die beide die Variable nr enthalten, welche die Fallnummer anzeigt. Diese gemeinsame Variable kann als Schlüsselvariable verwendet werden.

	nr	gr	gew
1	1,00	173,00	70,00
2	3,00	165,00	77,00
3	4,00	166,00	70,00
4	5,00	161,00	62,00
5	6,00	163,00	63,00

	nr	geschl	alter
1	1,00	1,00	47,00
2	2,00	2,00	58,00
3	4,00	2,00	66,00
4	5,00	2,00	52,00

Bild 3.25: Dateien mit Schlüsselvariable

Insgesamt existieren sechs Fälle. Dabei fehlen in der ersten Datei die Fälle 3 und 6, in der zweiten Datei fehlt Fall 2. Aufgrund der Orientierung an der Fallnummer ist trotzdem eine korrekte Zusammenfassung möglich.

- Laden Sie zunächst die Datei mix7.sav, und öffnen Sie über die Menüwahl

Daten

Dateien zusammenfügen

Variablen hinzufügen...

die Dialogbox *Variablen hinzufügen: Datei lesen*.

- Markieren Sie in der Dateienliste die Datei mix8.sav, und bestätigen Sie mit *Öffnen*.

Es öffnet sich die Dialogbox *Variablen hinzufügen aus*.

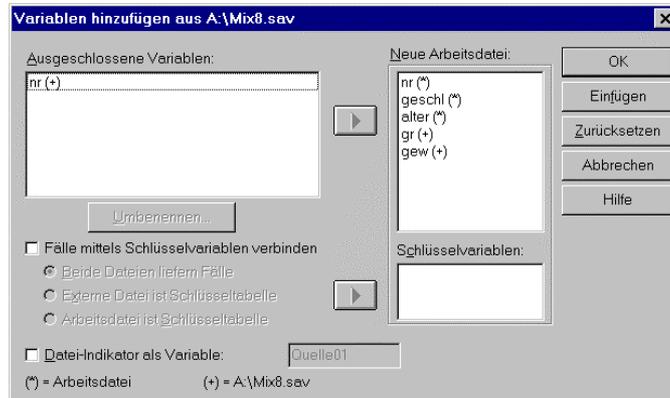


Bild 3.26: Dialogbox Variablen hinzufügen aus

Da die Variable *nr* in beiden Dateien auftritt, wird die Variable *nr* aus der hinzuzufügenden Datei (Markierung mit einem +) im Feld *Ausgeschlossene Variablen* aufgeführt. Diese Variable soll nun als Schlüsselvariable Verwendung finden.

- Aktivieren Sie die Option *Fälle mittels Schlüsselvariablen verbinden*, und belassen Sie es bei der voreingestellten Option *Beide Dateien liefern Fälle*.
- Markieren Sie im Feld *Ausgeschlossene Variablen* die Variable *nr*, und verschieben Sie diese mit Hilfe der Pfeiltaste in das Feld *Schlüsselvariablen*.
- Bestätigen Sie mit *OK*.

Sie werden zunächst mit einer Warnung konfrontiert, dass das Verbinden nach Schlüsselvariablen fehlschlägt, wenn die Dateien nicht in aufsteigender Reihenfolge nach den Schlüsselvariablen sortiert sind. Stellen Sie also sicher, dass dies der Fall ist; gegebenenfalls sortieren Sie die Dateien entsprechend vor dem Zusammenfügen (siehe Kap. 7.3).

- Im gegebenen Beispiel sind beide Dateien entsprechend sortiert; quittieren Sie daher mit *OK*.
- Die zusammengefasste Datei ist in Bild 3.27 dargestellt.

In einer weiteren Variante können den Fällen einer Datei mehrere Fälle der anderen Datei zugeordnet werden.

	geschl	alter	gr	gew
1	1,00	47,00	173,00	70,00
2	2,00	58,00	,	,
3	,	,	165,00	77,00
4	2,00	66,00	166,00	70,00
5	2,00	52,00	161,00	62,00
6	,	,	163,00	63,00

Bild 3.27: Zusammengefasste Datei

Zusammenfügen mit Schlüsselvariable (ein Fall mit mehreren Fällen)

Das folgende Beispiel ist einer zahnmedizinischen Untersuchung entnommen. Die Datei mix9.sav enthält neben einer Fallnummer (Variable nr) drei Variablen zu Alter, Geschlecht und Schulabschluss. Die Datei mix10.sav enthält zu jeder Fallnummer mehrere Zeilen, da von jedem Probanden mehrere Zähne auf Taschentiefe und CPITN-Wert untersucht wurden.

Die Datei mix9.sav ist hierbei der Datei mix10.sav übergeordnet und wird als Schlüssel-tabelle bezeichnet. Die Datei mix10.sav ist die hiervon abhängige Tabelle.

	nr	zahn	tt	cpitn
1	1,00	17,00	4,00	2,00
2	1,00	18,00	6,00	3,00
3	1,00	31,00	5,00	2,00
4	2,00	24,00	7,00	3,00
5	2,00	26,00	5,00	2,00
6	3,00	11,00	3,00	2,00
7	3,00	21,00	4,00	2,00
8	3,00	44,00	5,00	2,00
9	3,00	46,00	4,00	1,00

	nr	alter	geschl	schule
1	1,00	48,00	1,00	2,00
2	2,00	56,00	1,00	1,00
3	3,00	28,00	2,00	3,00

Bild 3.28: Schlüsseltabelle und abhängige Tabelle

- Laden Sie die Datei mix9.sav (Schlüsseltabelle), und wählen Sie aus den Menüs

Daten

Dateien zusammenfügen

Variablen hinzufügen...

Es öffnet sich die Dialogbox *Variablen hinzufügen: Datei lesen*.

- Markieren Sie in der Dateienliste die Datei mix10.sav und bestätigen Sie mit *Öffnen*.
- In der Dialogbox *Variablen hinzufügen aus* aktivieren Sie die Optionen *Fälle mittels Schlüsselvariablen verbinden* und *Arbeitsdatei ist Schlüsseltabelle*. Im Feld *Ausgeschlossene Variablen* markieren Sie die Variable nr; verschieben Sie diese Variable mit Hilfe der entsprechenden Pfeiltaste in das Feld *Schlüsselvariablen*.
- Bestätigen Sie mit *OK*.

Die zusammengefasste Datei ist in Bild 3.29 dargestellt.

	nr	alter	geschl	schule	zahn	tt	cpith
1	1,00	48,00	1,00	2,00	17,00	4,00	2,00
2	1,00	48,00	1,00	2,00	18,00	6,00	3,00
3	1,00	48,00	1,00	2,00	31,00	5,00	2,00
4	2,00	56,00	1,00	1,00	24,00	7,00	3,00
5	2,00	56,00	1,00	1,00	26,00	5,00	2,00
6	3,00	28,00	2,00	3,00	11,00	3,00	2,00
7	3,00	28,00	2,00	3,00	21,00	4,00	2,00
8	3,00	28,00	2,00	3,00	44,00	5,00	2,00
9	3,00	28,00	2,00	3,00	46,00	4,00	1,00

Bild 3.29: Zusammengefasste Datei

Falls eine Variable in der anzufügenden Datei den gleichen Namen hat wie eine Variable in der Arbeitsdatei und diese Variable nicht als Schlüsselvariable Verwendung finden soll, so kann sie im Feld *Ausgeschlossene Variablen* markiert und anschließend mit Hilfe des Schalters *Umbenennen...* umbenannt werden. Unter diesem Namen wird sie dann in der zusammengeführten Datei gespeichert.

3.8 Arbeitssitzung beenden

Wir wollen unsere Arbeitssitzung beenden.

- Wählen Sie aus dem Menü

Datei

Beenden

SPSS fragt zu jedem geöffneten Fenster, ob Sie den Inhalt abspeichern möchten. Wenn Sie »Ja« anklicken oder die Eingabetaste drücken, öffnet SPSS eine spezifische Dialogbox, um den jeweiligen Dateityp (Daten-, Ausgabe- oder Syntaxdateien) abzuspeichern.