

# Studienbrief



Psychologie

## Empirisches Forschungsprojekt

Grundlagen der Datenanalyse mit SPSS (I)

Yvonne Ziert

**1** EMF

### **Verfasserin**

Dipl. Sozialwiss. Yvonne Ziert, MPH

Studium der Sozialwissenschaften an der Leibniz Universität Hannover sowie an der John Moores University in Liverpool. Berufsbegleitender Master of Public Health (MPH) an der Medizinischen Hochschule Hannover (MHH). Aktuell arbeitet sie als wissenschaftliche Mitarbeiterin am Institut für Biometrie der MHH. Im Rahmen ihrer beruflichen Tätigkeit hat sie viel Erfahrung in der Planung und Umsetzung von empirischen Forschungsprojekten im Gesundheitswesen erworben. Darüber hinaus hält sie seit mehreren Jahren Lehrveranstaltungen im Bereich der empirischen Forschungsmethoden (v. a. Biometrie und SPSS-Kurse).

### **Lektorat**

Wissenschaftliche Mitarbeiterinnen und Mitarbeiter der Hamburger Fern-Hochschule

### **Satz/Repro**

Haussatz

### **Redaktionsschluss**

Dezember 2019

1. Auflage 2019

© HFH · Hamburger Fern-Hochschule, Alter Teichweg 19, 22081 Hamburg

Das Werk ist urheberrechtlich geschützt. Die dadurch begründeten Rechte, insbesondere das Recht der Vervielfältigung und der Verbreitung sowie der Übersetzung und des Nachdrucks, bleiben, auch bei nur auszugsweiser Verwertung, vorbehalten. Kein Teil des Werkes darf in irgendeiner Form ohne schriftliche Genehmigung der Hamburger Fern-Hochschule reproduziert oder unter Verwendung elektronischer Systeme verarbeitet, vervielfältigt oder verbreitet werden.

Gedruckt auf 100 % chlorfrei gebleichtem Papier.

# Inhaltsverzeichnis

<b>Abkürzungsverzeichnis .....</b>	<b>5</b>
<b>Einleitung .....</b>	<b>6</b>
<b>1 Vorstellung des Anwendungsbeispiels .....</b>	<b>8</b>
1.1 Szenario .....	8
1.2 Fragebogen .....	9
1.3 Routinedaten .....	10
1.4 Übersicht zu den verfügbaren Datendateien .....	10
Übungsaufgaben .....	11
<b>2 Vorstellung des Statistikprogramms SPSS .....</b>	<b>12</b>
2.1 Starten und Beenden von SPSS .....	12
2.2 SPSS-Grundstruktur .....	14
2.2.1 Daten-Editor .....	15
2.2.2 Viewer .....	16
2.2.3 Syntax-Editor .....	17
2.3 Prinzipielles zur Verwendung der Syntax .....	18
2.3.1 Vor- und Nachteile der Syntax .....	18
2.3.2 Grundlagen zur Befehlsstruktur und Anwendung der Syntax .....	19
2.4 Arbeiten mit dem Viewer .....	22
2.5 Überblick über die allgemeine Menüleiste .....	23
2.6 Allgemeine Voreinstellungen .....	24
Übungsaufgaben .....	25
<b>3 Aufbereitung der Daten .....</b>	<b>26</b>
3.1 Kodierung und Kodeplan .....	26
3.2 Anlegen und Speichern einer neuen Datendatei .....	27
3.3 Anlegen von Variablen .....	27
3.3.1 Variablenlabels .....	30
3.3.2 Wertbeschriftungen .....	30
3.3.3 Fehlende Werte .....	31
3.3.4 Messniveau .....	32
3.3.5 Sonstige Felder .....	33
3.3.6 Verschieben und Löschen von Variablen .....	33
3.4 Manuelle Dateneingabe in den Daten-Editor .....	34
3.5 Einlesen von Datendateien .....	35
3.5.1 Interne Datendateien .....	35
3.5.2 Externe Formate .....	36
3.6 Zusammenfügen von Datendateien .....	37
3.6.1 Fallweise Zusammenfügen .....	37
3.6.2 Variablenweise Zusammenfügen .....	38
3.7 Datenprüfung .....	39
3.7.1 Prüfung der Daten gegen den Urbeleg .....	40
3.7.2 Häufigkeitsauszählungen .....	41
3.7.3 Berechnung von Kennwerten .....	42
Übungsaufgaben .....	44

<b>4 Modifikation von Variablen .....</b>	<b>45</b>
4.1 Umcodieren von Variablen.....	45
4.1.1 In dieselbe Variable.....	45
4.1.2 In andere Variable .....	48
4.1.3 Umcodieren von String-Variablen .....	50
4.2 Berechnung neuer Variablen .....	51
4.2.1 Numerische Operatoren und Funktionen .....	52
4.2.2 Rechnen mit speziellen Formaten .....	53
Übungsaufgaben.....	56
<b>5 Sortierung und Auswahl von Fällen .....</b>	<b>57</b>
5.1 Sortierung von Fällen .....	57
5.1.1 Sortierung über die Menüauswahl.....	57
5.1.2 Sortierung über die Datenansicht im Daten-Editor .....	58
5.2 Fallauswahl unter einer Bedingung .....	59
5.2.1 Temporäre Auswahl .....	59
5.2.2 Permanente Auswahl.....	60
5.2.3 Zufallsstichprobe .....	60
Übungsaufgaben.....	62
<b>Zusammenfassung.....</b>	<b>63</b>
<b>Anhang 1 - Patientenfragebogen .....</b>	<b>64</b>
<b>Anhang 2 - Kodeplan .....</b>	<b>65</b>
<b>Anhang 3 – Beispielfall.....</b>	<b>66</b>
<b>Glossar.....</b>	<b>67</b>
<b>Lösungen zu den Übungsaufgaben .....</b>	<b>68</b>
<b>Literaturverzeichnis .....</b>	<b>71</b>

## Abkürzungsverzeichnis

BMI	Body Mass Index
ICD	International Statistical Classification of Diseases and Related Health Problems
ID	Identifikationsnummer
kg	Kilogramm
m	Meter
SPSS	Statistical Package for the Social Sciences
ZUF-8	Fragebogen zur Patientenzufriedenheit

## Einleitung

Die Analyse komplexer Datensätze spielt für Psychologinnen und Psychologen in vielen Bereichen eine wichtige Rolle – sei es in der psychologischen Grundlagenforschung, bei der Entwicklung von Tests, bei der Analyse von Qualitätssicherungsdaten oder bei der Untersuchung der Effektivität von Interventionen. Allerdings ist der Umgang mit großen Datenmengen komplex und setzt sowohl Kenntnisse im Bereich der statistischen Methoden als auch die Fähigkeit zur Verwendung von spezifischen Softwareprogrammen zum Verwalten und Analysieren der Daten voraus.

### Ziel des Moduls

Ziel des Moduls „Empirisches Forschungsprojekt“ ist es, neben der weiteren Vertiefung des theoretischen Wissens im Bereich der Statistik und empirischen Forschung, diese Methoden praktisch anzuwenden, also konkret mit Daten zu arbeiten, diese zu generieren, aufzubereiten, zu analysieren, zu interpretieren und zu präsentieren. Es ist für eine effiziente Aufbereitung, Analyse und grafische Darstellung von Daten in der heutigen Arbeitswelt der Standard, spezifische Softwarepakete, wie z. B. R, SAS, STATA oder SPSS zu verwenden. Diese unterscheiden sich u. a. hinsichtlich der Benutzeroberfläche, der Programmiersprache (Syntax), der Verfügbarkeit von statistischen Auswertungsmodulen und der Nutzungskosten. Die Wahl der jeweiligen Software hängt neben der Verfügbarkeit bestimmter Rechenprozeduren auch von speziellen Vorlieben des Nutzers ab.

### Statistikprogramm SPSS

Das Statistikprogramm SPSS ist besonders gut für Einsteiger geeignet, da neben der Programmiersprache auch eine anwenderfreundliche Benutzeroberfläche verfügbar ist. SPSS bietet vielfältige Möglichkeiten der Datengenerierung, -eingabe und -überprüfung sowie deskriptive, bi- und multivariate Analysemethoden und grafische Darstellungsmöglichkeiten. Die Verfügbarkeit unterschiedlicher Fenster, wie es bei den gängigen Microsoft-Office-Programmen der Fall ist, macht das nahezu intuitive Erlernen des Programms möglich. Zudem kann es mit unterschiedlichen Betriebssystemen genutzt werden und ist mit einer Reihe weiterer Programme kompatibel. Darüber hinaus weist das Programm einen hohen Verbreitungsgrad auf, sodass SPSS-Kenntnisse für unterschiedlichste Tätigkeitsfelder von Nutzen sein können. Letztlich erleichtern die Kenntnisse eines Statistikprogramms das Einarbeiten in andere Statistikprogramme.

### Bezug zu anderen Studienbriefen des Moduls

Ziel dieses Studienbriefs ist es, die Grundfunktionen des Programms, die insbesondere für die unterschiedlichen Aspekte der Datenverwaltung und -aufbereitung zentral sind, kennenzulernen. Die Datenaufbereitung stellt eine wichtige Basis für die weiteren Schritte der deskriptiven, bivariaten und multivariaten Datenanalyse dar. Die Bearbeitung dieses Studienbriefs setzt keinerlei Vorkenntnisse in Bezug auf das Programm SPSS voraus. Bei seiner Gestaltung sowie bei den anderen Studienbriefen dieses Moduls ist auf eine hohe Anwendungsorientierung bei der Vermittlung der Lerninhalte geachtet worden. So werden alle statistischen Methoden und Prozeduren anhand eines typischen Beispiels zur Datenerfassung aus dem Gesundheitswesen veranschaulicht.

### Syntaxdatei zum Studienbrief

Um den Einstieg in das Programm SPSS insbesondere für Anfänger möglichst einfach zu gestalten, werden die einzelnen Prozeduren über das bedienungsfreundliche Menü (Benutzeroberfläche) erklärt. SPSS ist so aufgebaut, dass alle Prozeduren alternativ über die sog. Syntax (Programmiersprache) gesteuert werden können. Der vorliegende Studienbrief beschränkt sich darauf, grundlegende Prinzipien der Verwendung der Syntax zu erläutern. Ergänzend werden für alle Prozeduren, für die die Steuerung über das Menü gezeigt wird, die Befehle in einer separaten Syntaxdatei aufgeführt (*EMF\_SBI\_Syntax.sps*).

Im Detail gliedert sich der Studienbrief in die folgenden Kapitel:

In *Kapitel 1* wird das Anwendungsbeispiel „Patientenfragebogen und Routinedaten des Klinikverbunds Nordsterne“ ausführlich vorgestellt. Anhand dieses Beispiels werden alle Prozeduren des vorliegenden Studienbriefes vermittelt sowie alle Prozeduren und Methoden der folgenden Studienbriefe des Moduls.

**Kapitel 1**

In *Kapitel 2* erfolgt eine Einführung in den Aufbau und Gebrauch des Statistikprogramms SPSS. Ein spezieller Fokus liegt auf der Vermittlung von grundlegenden Prinzipien zur Verwendung der Programmiersprache (Syntax) von SPSS.

**Kapitel 2**

In *Kapitel 3* werden die Basisfunktionen in SPSS vorgestellt, die erforderlich sind, um die Daten so aufzubereiten, dass sie für eine statistische Analyse nutzbar werden. Dabei wird sowohl die Möglichkeit aufgezeigt, wie ein Kodeplan erstellt wird und Daten manuell in SPSS eingegeben werden. Darüber hinaus wird erläutert, wie Daten eines anderen Dateiformats in SPSS eingelesen und auf ihre Korrektheit und Plausibilität geprüft werden können.

**Kapitel 3**

Die in *Kapitel 4* und *Kapitel 5* dargestellten Prozeduren gehen über die Basisfunktionen zur Datenaufbereitung hinaus und adressieren spezifische Funktionen zur Variablenmodifikation und zur Fallauswahl. In *Kapitel 4* wird dargestellt, wie bereits bestehende Variablen in einer SPSS-Datendatei modifiziert werden können.

**Kapitel 4**

In *Kapitel 5* wird dargestellt, wie bestimmte Fälle aus der Datei aufgenommen bzw. ausgeschlossen werden können.

**Kapitel 5**

Alle im Studienbrief verwendeten Abbildungen (*Screenshots*) wurden mit dem Programm *IBM SPSS Statistics 25* erzeugt.

Nach der Bearbeitung des ersten Studienbriefs sollen die Studierenden in der Lage sein

**Studienziele**

- ⇒ die Grundfunktionen des Statistikprogramms SPSS zu bedienen (Fenster, Menüführung etc.),
- ⇒ den Kodeplan für einen Fragebogen zu erstellen,
- ⇒ Daten manuell einzugeben und automatisch einzulesen,
- ⇒ Daten zu modifizieren (Variablen zu verändern und neu zu berechnen) sowie
- ⇒ Fälle auszuwählen (temporäre und permanente Auswahl, Zufallsstichprobe).

# 1 Vorstellung des Anwendungsbeispiels

Um über das gesamte Modul hinweg einen direkten praktischen Anwendungsbezug zu gewährleisten, werden die unterschiedlichen Prozeduren zur Datenverwaltung und Datenanalyse in SPSS durch ein typisches Beispiel der Datenerhebung aus dem Gesundheitswesen vermittelt. Das fiktive Beispiel wurde so ausgewählt, dass dadurch die vielfältigen Herausforderungen, die es bei einem empirischen Forschungsprojekt von der Datenaufbereitung über die deskriptive Auswertung bis zur umfassenden multivariaten statistischen Analyse zu bewältigen gibt, adressiert werden können.

## 1.1 Szenario

### Beispiel

Der **Klinikverbund „Nordsterne“** betreibt drei Rehabilitationszentren, die jeweils im nördlichen Teil Deutschlands küstennah gelegen sind. Dabei haben sich die Zentren auf die Behandlung von älteren Patientinnen und Patienten, die an Typ-II-Diabetes erkrankt sind und Übergewicht haben, spezialisiert. Ihr Erfolgsmodell ist eine durchschnittlich 4-wöchige medizinische Rehabilitationsmaßnahme (im folgenden Reha-Maßnahme genannt) mit dem primären Ziel der Gewichtsreduktion. Als modernes Dienstleistungsunternehmen hat der Verbund ein großes Interesse daran, die Zufriedenheit seiner Patientinnen und Patienten zu evaluieren, um auf dieser Grundlage die Reha-Maßnahme stetig zu optimieren. Zu diesem Zweck hat die Geschäftsleitung einen Patientenfragebogen entwickeln lassen. Dieser Fragebogen wird den Patientinnen und Patienten am Ende der 4-wöchigen Reha-Maßnahme im Rahmen des ärztlichen Abschlussgesprächs ausgehändigt. Um die Befragungsergebnisse mit anderen Daten der Patientinnen und Patienten aus dem Krankenhausinformationssystem (z. B. Hauptdiagnose, Verweildauer) zu kombinieren, erfragt der Arzt bei der Aushändigung des Fragebogens, ob sich der Patient damit einverstanden erklärt, dass die in der Klinik intern verwendete Identifikationsnummer (ID) des Patienten auf dem Fragebogen vermerkt werden darf. Sofern der Patient zustimmt, vermerkt der Arzt die ID auf dem hierfür vorgesehenen Feld. Da durch dieses Vorgehen die Anonymität der Befragung nicht mehr gegeben ist, werden die Patientinnen und Patienten explizit darauf hingewiesen, dass ihre Daten vertraulich behandelt werden. Anschließend füllt der Patient bzw. die Patientin (freiwillig) den Fragebogen eigenständig aus und wirft ihn in einen hierfür vorgesehenen Briefkasten.

Mittlerweile liegen aus den drei Zentren Fragebögen von jeweils 30 Patientinnen und Patienten vor, die sich auch damit einverstanden erklärt haben, dass ihre ID auf dem Bogen vermerkt wird. Da es für die späteren Analysen wichtig ist, dass die Routinedaten und die Befragungsdaten für jeden Fall über die ID zusammengeführt werden können, werden ausschließlich diese Patientinnen und Patienten im vorliegenden Studienbrief betrachtet.

Die Befragungs- und Routinedaten sollen so aufbereitet werden, dass sie einer statistischen Analyse zugänglich sind, um u. a. die folgenden Fragen zu beantworten:

### Forschungsfragen

- Welche Charakteristika weisen die Patientinnen und Patienten auf, die an der durchschnittlich 4-wöchigen Reha-Maßnahme teilgenommen haben?
- Gibt es Unterschiede hinsichtlich der Patientenzufriedenheit zwischen den drei Zentren? Und falls ja, wie lassen sie sich erklären?
- Welche Faktoren beeinflussen die Patientenzufriedenheit?



## 1.2 Fragebogen

Der Fragebogen wurde speziell zur Evaluation der Patientenzufriedenheit der drei Rehabilitationszentren entwickelt. Je nach Frage waren unterschiedliche Antwortmöglichkeiten vorgegeben (siehe Anhang 1). Die unterschiedlichen Fragen lassen sich inhaltlich den folgenden vier Bereichen zuordnen:

### a. Administrative Angaben

0. Nr. des Zentrums
1. Patientenidentifikationsnummer

### Aufbau des Fragebogens

### b. ZUF-8: Fragebogen zur Patientenzufriedenheit

Die folgenden acht Fragen bilden zusammen den sog. Fragebogen zur Patientenzufriedenheit (ZUF-8). Dieser stellt ein erprobtes und validiertes psychometrisches Instrument zur Erhebung der Patientenzufriedenheit dar und wurde von Schmidt et al. (1989, 1994) bzw. von Schmidt und Nübling (2002) entwickelt. Die Codes zu den Merkmalsausprägungen der Fragen werden zu einem Summenscore aufaddiert (für mehr Informationen vgl. Gesellschaft für Qualität im Gesundheitswesen, o.J.).

2. Wie würden Sie die Qualität der Behandlung, welche Sie erhalten haben, beurteilen?
3. Haben Sie die Art von Behandlung erhalten, die Sie wollten?
4. In welchem Maße hat unsere Klinik Ihren Bedürfnissen entsprochen?
5. Würden Sie einem Freund/einer Freundin unsere Klinik empfehlen, wenn er/sie eine ähnliche Hilfe benötigen würde?
6. Wie zufrieden sind Sie mit dem Ausmaß der Hilfe, welche Sie hier erhalten haben?
7. Hat die Behandlung, die Sie hier erhielten, Ihnen dabei geholfen, angemessener mit Ihren Problemen umzugehen?
8. Wie zufrieden sind Sie mit der Behandlung, die Sie erhalten haben, im Großen und Ganzen?
9. Würden Sie wieder in unsere Klinik kommen, wenn Sie eine Hilfe bräuchten?

### c. Demografische Angaben

10. Geschlecht
11. Alter
12. Körpergröße

### d. Angaben zu unterschiedlichen Aspekten des Aufenthalts

13. Körpergewicht (vor und nach der Reha)
14. Raucherstatus (vor und nach der Reha)
15. Sportliche Betätigung während des Aufenthalts
16. Wahrgenommene Freizeitangebote während des Aufenthalts
17. Einbezug von Angehörigen in die Behandlung
18. Chronische Erkrankung(en)

### Antwortformate

Die Fragen 1–17 weisen im Fragebogen (siehe Anhang 1) ein geschlossenes bzw. ein halb-offenes Antwortformat auf. Das heißt, die Befragten müssen sich entweder auf eine vorgegebene Antwortkategorie festlegen (z. B. Fragen 1–9) oder eigenständig einen Wert in einem fest vorgegebenen Format (z. B. Frage 13) angeben. Bei Frage 18 handelt es sich um ein Item mit offenem Antwortformat. Hier wird den Befragten abverlangt, dass sie Angaben zu ihren chronischen Erkrankungen in Textform machen. Solche Fragetypen wählt man insbesondere dann, wenn man sich im Vorfeld unsicher bezüglich möglicher Antworten ist.

### 1.3 Routinedaten

Die personenspezifischen Daten, die aus dem Krankenausinformationssystem der drei Zentren extrahiert werden, liegen als Excel-File vor und werden im Folgenden als Routinedaten bezeichnet, da sie u. a. ICD-Kodierungen beinhalten, die von den Zentren routinemäßig zu Abrechnungszwecken mit den Krankenkassen erhoben werden. Der Excel-File besteht aus einer Tabelle (siehe **Tabelle 1.1**) mit den folgenden Spalten:

### Datenstruktur der Routinedaten

- *v1\_ID*: Identifikationsnummer des Patienten oder der Patientin
- *ICD\_HD*: ICD-Kodierung (Hauptdiagnose)
- *AUFN\_DATUM*: Datum der Aufnahme in die Klinik
- *ENTL\_DATUM*: Datum der Entlassung aus der Klinik

**Tabelle 1.1:** Aufbau der Datendatei „Routinedaten.xlsx“

<i>v1_ID</i>	<i>ICD_HD</i>	<i>AUFN_DATUM</i>	<i>ENTL_DATUM</i>
1	E11.95	28.06.2012	22.08.2012
...	...	...	...

### 1.4 Übersicht zu den verfügbaren Datendateien

Um die unterschiedlichen SPSS-Prozeduren anwendungsorientiert zu präsentieren, wird in den unterschiedlichen Studienbriefen des Moduls auf die in **Tabelle 1.2** aufgelisteten Datendateien zurückgegriffen. Diese basieren entweder auf den Fragebogendaten oder den Routinedaten des Anwendungsbeispiels.

### Übungsdateien

Insgesamt stehen sieben Übungsdateien zur Verfügung. Die ersten drei SPSS-Dateien beinhalten jeweils die Fragebogendaten für jedes einzelne Rehabilitationszentrum (*Zentrum1.sav*, *Zentrum2.sav*, *Zentrum3.sav*). In der vierten Datendatei sind die Fragebogendaten aller Zentren enthalten (*Zentren\_gesamt.sav*). Diese Dateien umfassen vollständige und plausible Befragungsdaten. Allerdings ist für jeden Patienten für die ZUF-8-Skala ausschließlich ein Summenscore in den Daten enthalten – d. h. dort sind keine Werte zu den Einzelitems der Skala verfügbar. Ausschließlich in der fünften Datendatei *Zentrum1\_Spieldaten.sav* sind Angaben zu Einzelitems der ZUF-8-Skala für das Zentrum 1 (30 Fälle) enthalten. In dieser Datei sind zudem zu Übungszwecken unplausible Daten und Eingabefehler programmiert worden. In der sechsten und siebten Datendatei sind zu den 90 Patienten der drei Zentren die Routinedaten enthalten (*Routinedaten.sav* bzw. *Routinedaten.xlsx*). Diese liegen sowohl im SPSS-Format und als auch im Excel-Format vor.

**Tabelle 1.2:** Übersicht zu den im Studienbrief verwendeten Datendateien

Nr.	Datendatei	Beschreibung
1	<i>Zentrum1.sav</i>	vollständige und plausible Befragungsdaten von Zentrum 1 (n=30 Patienten) im SPSS-Dateiformat
2	<i>Zentrum2.sav</i>	vollständige und plausible Befragungsdaten von Zentrum 2 (n=30 Patienten) im SPSS-Dateiformat
3	<i>Zentrum3.sav</i>	vollständige und plausible Befragungsdaten von Zentrum 3 (n=30 Patienten) im SPSS-Dateiformat
4	<i>Zentren_gesamt.sav</i>	vollständige und plausible Befragungsdaten von allen drei Zentren (n=90 Patienten) im SPSS-Dateiformat
5	<i>Zentrum1_Spieldaten.sav</i>	unvollständige und implausible Befragungsdaten von Zentrum 1 (n=30 Patienten) im SPSS-Dateiformat; für die Summenskala zur Patientenzufriedenheit sind Einzelwerte verfügbar
6	<i>Routinedaten.sav</i>	Routinedaten für alle drei Zentren (n=90) im SPSS-Dateiformat
7	<i>Routinedaten.xlsx</i>	Routinedaten für alle drei Zentren (n=90) im Excelformat

Die verwendeten Datendateien sowie die Syntaxdatei finden Sie im WebCampus im Materialbereich des Moduls.

**Hinweis**

### Übungsaufgaben

- 1.1) Welche unterschiedlichen Antwortformate weist der Patientenfragebogen aus dem Anwendungsbeispiel auf? Geben Sie jeweils ein Beispiel.
- 1.2) Worin unterscheiden sich die beiden Dateien *Routinedaten.sav* und *Routinedaten.xlsx*?

## 2 Vorstellung des Statistikprogramms SPSS

SPSS ist ein modular aufgebautes Programmpaket zum **Verwalten, Analysieren und Visualisieren von Daten**. Das sog. Basismodul ermöglicht sowohl das grundlegende Datenmanagement als auch die Durchführung umfangreicher statistischer Datenanalysen mit entsprechenden grafischen Darstellungen. Für spezielle statistische Verfahren wie z. B. Zeitreihenanalysen existieren ergänzende Zusatzmodule.

### Hintergrund

Die Bezeichnung SPSS stand anfänglich für *Statistical Package for the Social Sciences*. Um die Eignung von SPSS über den sozialwissenschaftlichen Bereich hinaus deutlich zu machen, kam es zwischenzeitlich zu einer Umbenennung des Programms in *Superior Performing Software Systems*. Seit 2011 lautet die offizielle Bezeichnung der Software **IBM SPSS Statistics** (Kurzbezeichnung: SPSS). In der Regel gibt es jährlich ein Programmupdate.

### Anwendungskontext

SPSS wird in den unterschiedlichsten Gesellschaftsbereichen zum Zweck der Datenverwaltung und statistischen Analyse eingesetzt. Insbesondere in empirisch ausgerichteten Wissenschaften gehört es mittlerweile fast zur Standardausstattung und ist häufig elementarer Bestandteil der Methodenausbildung im Studium. Die allgemeine Beliebtheit des Programms ergibt sich sicherlich durch seine hohe Anwenderfreundlichkeit, die sich u. a. durch die folgenden Aspekte auszeichnet (vgl. Zöfel, 2002, S. 11):

- Möglichkeit zur Übernahme und Weiterverarbeitung einer Vielzahl von Formaten,
- diverse Möglichkeiten zur Anwendung des Programms (Menüs, Dialogboxen und Programmiersprache),
- Möglichkeiten der deskriptiven Darstellung von Daten sowie der Durchführung von hochkomplexen statistischen Analysen sowie.

Um den Einstieg insbesondere für erstmalige SPSS-Nutzer möglichst einfach zu gestalten, wird in den folgenden Abschnitten ein Überblick zur Funktionalität und Handhabung des Programms gegeben. Hierbei handelt es sich um allgemeine Hinweise, die für ein effizientes Arbeiten mit dem Programm hilfreich sind. Für SPSS-Einsteiger bzw. Personen, die bisher noch mit keinem statistischen Auswertungsprogramm gearbeitet haben, empfiehlt es sich, die Prozeduren, über das Menü bzw. die Dialogboxen anstatt über die Syntax zu steuern. Für die Personen, die langfristig mit dem Programm arbeiten möchten, ist ein Erlernen der Programmiersprache (Syntax) ratsam, da es ein effizienteres Arbeiten mit dem Programm ermöglicht. Vor diesem Hintergrund wird in Abschnitt 2.3 eine kurze Einführung in grundsätzliche Prinzipien zur Erstellung und Handhabung der Syntax gegeben.

### 2.1 Starten und Beenden von SPSS

Nach dem Starten von SPSS erscheint zunächst eine Dialogbox (siehe **Abb. 2.1**), die die am häufigsten benötigten Aktionen (z. B. *Neues Dataset*, *Zuletzt verwendete Dateien*) zusammenfasst. Darüber hinaus gibt es die Möglichkeiten, sich über Neuerungen in der aktuellen Version zur Vorgänger Version zu informieren (*Neuerungen*), die SPSS Community zu besuchen, erste Schritte mit Hilfe des SPSS-Lernprogramms zu tätigen sowie die Möglichkeit für Hilfe und Support zu nutzen (*Erste Schritte*). Dieses Dialogfeld kann optional auch deaktiviert werden, indem die Checkbox am unteren Rand (*Dieses Dialogfeld nicht mehr anzeigen*) betätigt und mit **OK** bestätigt wird. Um fortzufahren, muss entweder eine der vorgeschlagenen

Aktionen ausgewählt und ausgeführt werden oder das Fenster kann durch *Abbrechen* geschlossen werden.

Sofern eine bereits erstellte und abgelegte (d. h. gespeicherte) SPSS-Datei (Datenda-  
tei, Syntaxdatei oder Ausgabedatei) geöffnet werden soll, kann dies durch einen  
Doppelklick mit der linken Maustaste auf die jeweilige Datei erfolgen. Dies hat zur  
Folge, dass SPSS automatisch geöffnet wird.

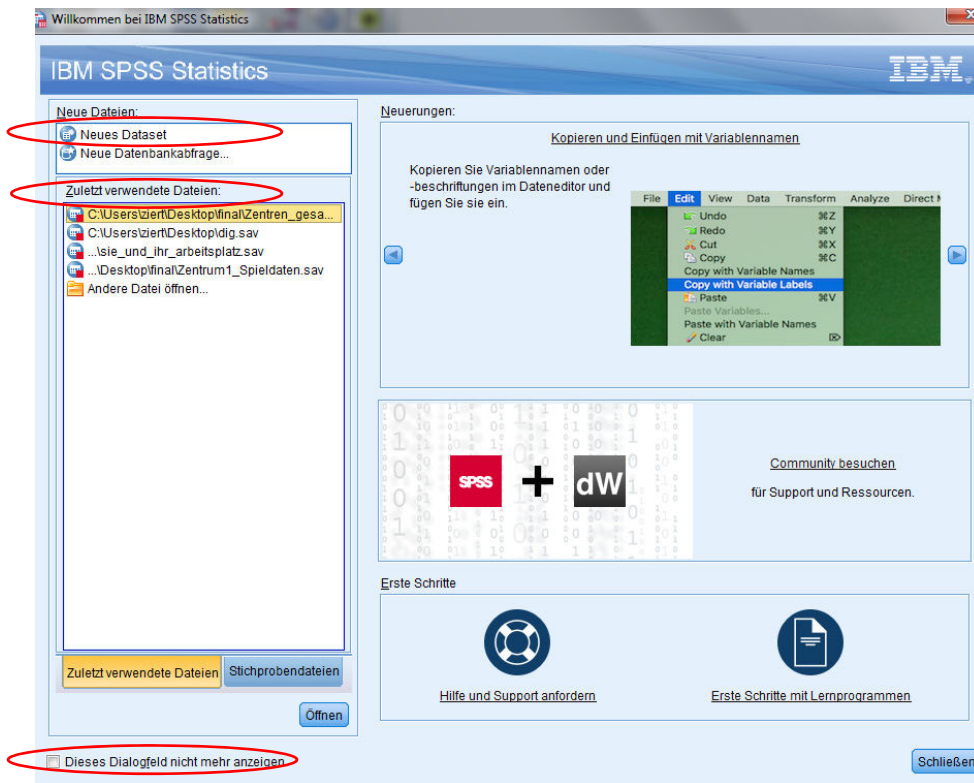


Abb. 2.1: Starten von SPSS

Für die Beendigung von SPSS stehen die folgenden Optionen zur Verfügung. Entweder man wählt über das Menü die Aktion *Datei* → *Beenden* aus oder man schließt die in SPSS geöffneten Fenster jeweils einzeln durch Klicken auf das Symbol oben rechts im Fenster (☒). Für jedes geöffnete Fenster wird von SPSS beim Schließen der jeweiligen Datei, sofern Änderungen an der Datei vorgenommen wurden, separat erfragt, ob die jeweilige Datei gespeichert werden soll.

## Beenden von SPSS

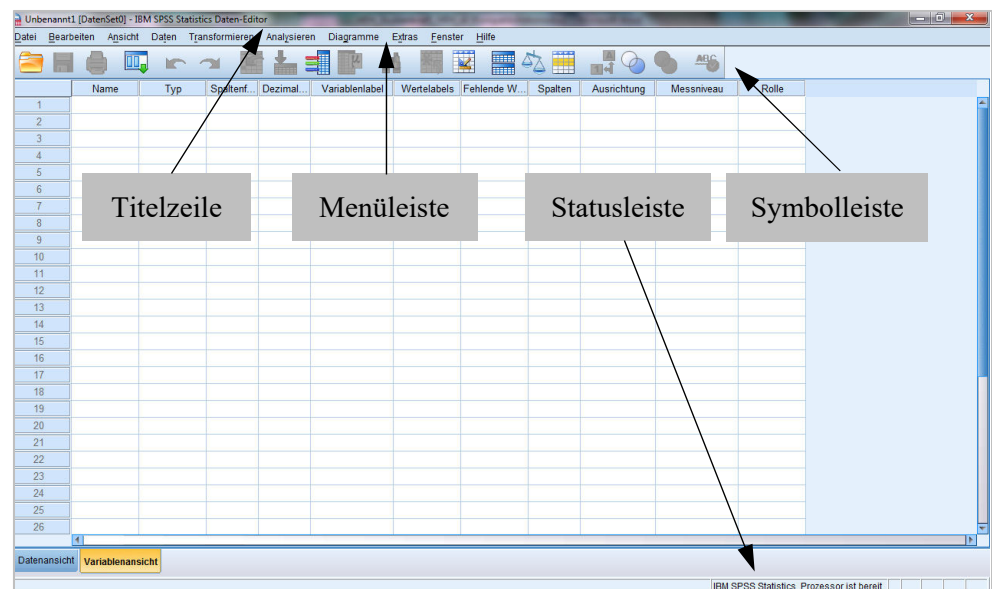
### 2.2 SPSS-Grundstruktur

Für die Bedienung von SPSS für Windows stehen dem Anwender – wie es bei den gängigen Microsoft-Office-Programmen der Fall ist – unterschiedliche Menüleisten und Fenster zur Verfügung. Die Grundstruktur des Programms erschließt sich dem Benutzer über die Betrachtung der folgenden drei Fenster, die das Programm umfasst:

#### Fenster in SPSS

- Daten-Editor
- Viewer
- Syntaxfenster.

Alle drei Fenster erfüllen innerhalb des Programms unterschiedliche Funktionen, die jeweils detailliert in den folgenden Abschnitten erläutert werden. Den Fenstern ist gemeinsam, dass ihr grundsätzlicher Aufbau dem von Microsoft-Office-Programmen entspricht. Ganz oben befindet sich immer die **Titelzeile**, die den Namen der Datei enthält. Darunter befindet sich die allgemeine **Menüleiste** mit allen wichtigen Funktionen des Programms (siehe **Abb. 2.2**), auf die von jedem Fenster aus zugegriffen werden kann (Einzelheiten zur allgemeinen Menüleiste siehe Abschnitt 2.4). Unterhalb der allgemeinen Menüleiste befindet sich die **Symbolleiste**, die in Abhängigkeit von den spezifischen Fenstern, die am häufigsten benötigten Funktionen enthält. Im unteren Teil des Fensters befindet sich die **Statusleiste**, die den aktuellen Status des SPSS-Prozessors angibt und welcher Befehl gerade ausgeführt wird.



**Abb. 2.2:** Titelzeile, Menüleiste, Statusleiste, Symbolleiste

Ein weiteres wichtiges Merkmal aller drei Fenster ist, dass ihre jeweiligen Inhalte gespeichert werden können. Die Speicherung muss jedoch für jedes Fenster bzw. dessen Inhalt separat erfolgen, da das Speicherformat für die Inhalte der drei Fenster jeweils unterschiedlich ist. Näheres hierzu findet sich ebenfalls in dem jeweiligen Abschnitt.

Alle drei Fenster – sowie unterschiedliche Daten-, Ausgabe- und Syntaxdateien – können in SPSS parallel geöffnet werden und es kann problemlos zwischen ihnen hin und her gewechselt werden.

Um die **Funktionen und den Aufbau der drei Fenster** anschaulich zu präsentieren, werden z.T. Beispiele präsentiert, die auf der Datendatei *Zentrum1\_Spieldaten.sav* des Anwendungsbeispiels basieren.

## Beispieldatensatz

### 2.2.1 Daten-Editor

Beim Starten von SPSS wird automatisch immer der Daten-Editor mit geöffnet. Mit Hilfe des Daten-Editors können entweder Daten manuell neu eingegeben (siehe Abschnitt 3.4) oder bereits bestehende Daten aus einer Datendatei geladen werden (siehe Abschnitt 3.5). Wenn beim Start von SPSS keine Datendatei geladen wird, dann wird der Daten-Editor mit *unbenannt* betitelt. Unter der allgemeinen Menüleiste befinden sich diverse Schalter, die spezifische Funktionen des Editors unterstützen. Die Funktion des jeweiligen Schalters wird dem Benutzer angezeigt, wenn man mit dem Mauszeiger auf dem Schalter verweilt.

## Funktion

Der Daten-Editor verfügt über zwei Ansichten:

1. Datenansicht
2. Variablenansicht

Die **Datenansicht** (siehe *Abb. 2.3*) ist wie ein sog. *Spreadsheet* aufgebaut, was bedeutet, dass sie über Zeilen und Spalten verfügt, wodurch eine effiziente und einfache Dateneingabe (siehe Abschnitt 3.4) realisiert werden kann (vgl. Bühl, 2010). Das *Spreadsheet* ist so aufgebaut, dass in den Zeilen jeweils die Fälle (z. B. Patientinnen und Patienten) und in den Spalten jeweils die Variablen (z. B. Fragen im Fragebogen) stehen. Die unterschiedlichen Ausprägungen, die eine Variable annehmen kann, werden als Merkmalsausprägungen bezeichnet. Das heißt, in einer einzelnen Zelle befindet sich die jeweilige Ausprägung einer Variablen für den jeweiligen Fall. Pro Zeile wird ein Wert bzw. im Fall von Textangaben der entsprechende Text abgespeichert.

## Datenansicht

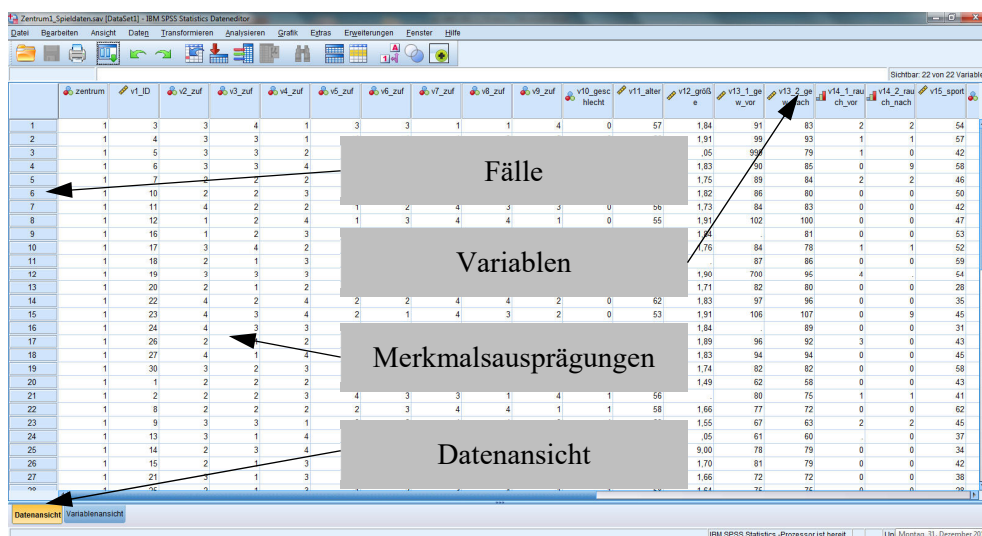


Abb. 2.3: Datenansicht im Daten-Editor

Um auf die sog. **Variablenansicht** (siehe *Abb. 2.4*) zu gelangen, kann entweder auf die Registerkarte *Variablenansicht* geklickt werden oder mit Doppelklick auf eine mit einem Variablennamen betitelte Zelle (z. B. *v1\_ID*). Die Variablenansicht ist so aufgebaut, dass jede Zeile eine einzelne Variable darstellt. Unter der Verwendung der vorgegebenen Funktionen der einzelnen Spalten kann eine Variable neu angelegt

## Variablenansicht

werden (siehe Abschnitt 3.3) bzw. bei einem bereits existierenden Datensatz können aus den spezifischen Spalten wichtige Informationen zu der Variablen abgelesen werden.

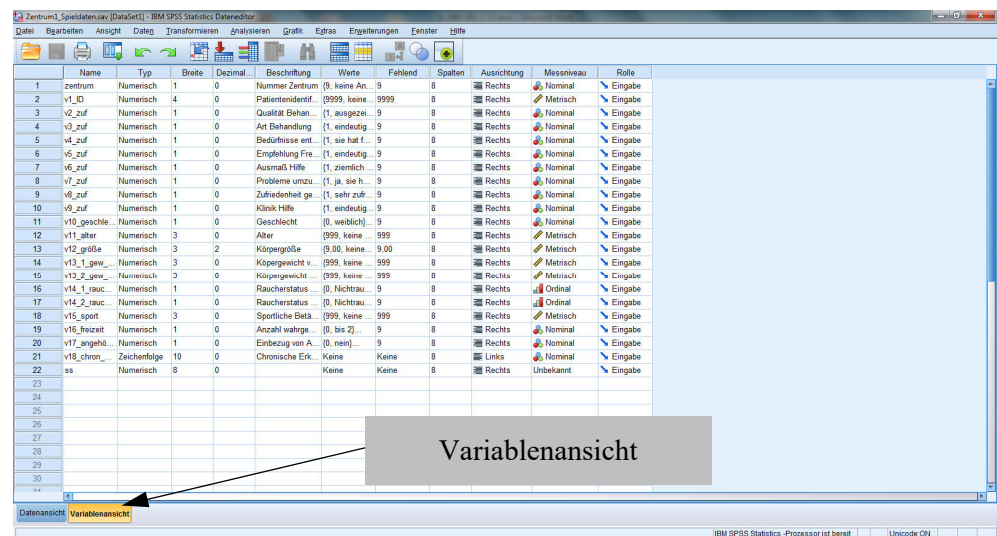


Abb. 2.4: Variablenansicht im Daten-Editor

Um wieder zurück zur Datenansicht zu gelangen, kann entweder auf die Registerkarte *Datenansicht* geklickt werden oder mit Doppelklick auf eine komplette Zeile in der Variablenansicht.

Wenn die Inhalte des Daten-Editors abgespeichert wurden (siehe Abschnitt 3.2), wird eine **Datendatei** mit der **Endung .sav** erzeugt, die benannt werden muss und separat geöffnet werden kann.

### 2.2.2 Viewer

Der Viewer (auch Ausgabe genannt) ist ebenfalls ein separates Fenster in SPSS und kann über die folgende Menüfunktion geöffnet werden: *Datei* → *Neu* → *Ausgabe*. Der Viewer erfüllt im Wesentlichen die folgenden zwei Funktionen:

- Funktion**
1. Dokumentation der durchgeführten Datenaufbereitungs- und Analyseschritte in Form der Programmiersprache (Syntax)
  2. Anzeigen der Ergebnisse der statistischen Prozeduren (z. B. Tabellen und Grafiken).

Darüber hinaus werden hier auch Fehlermeldungen angezeigt, wenn z. B. bestimmte Prozeduren nicht durchgeführt werden können.

**Aufbau**

Um ein effizientes Arbeiten mit dem Viewer zu ermöglichen, besteht er aus zwei Teilen (siehe Abb. 2.5). Im linken Teil befindet sich eine Gliederung der durchgeführten Prozeduren. Im rechten Teil werden die durchgeführten Prozeduren (Syntax) und die entsprechenden Ergebnistabellen und Grafiken dargestellt. Die Breite der beiden Fenster kann durch Ziehen mit der Maus am Trennungsbalken variiert werden.



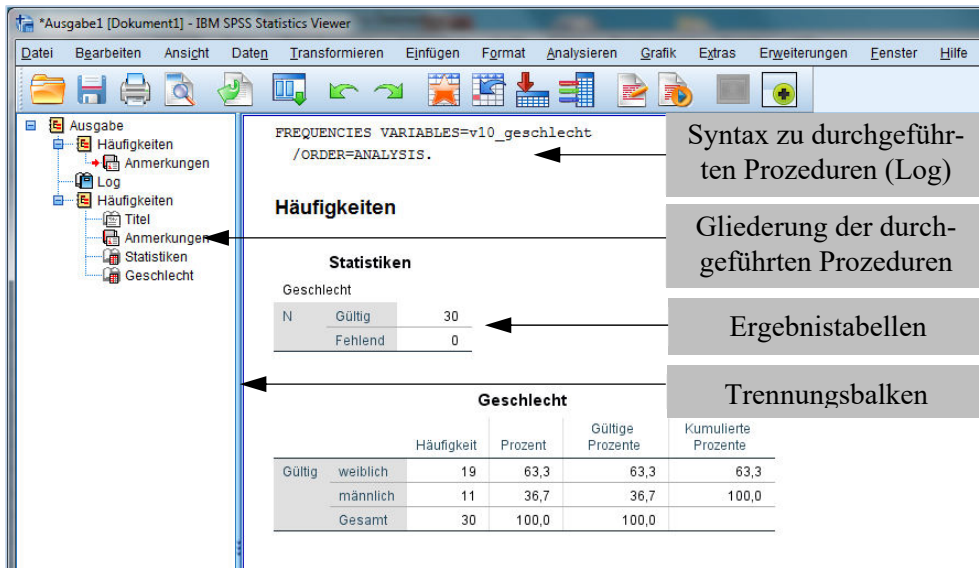


Abb. 2.5: Viewer (Ausgabe)

Die Darstellung der Dokumentation der gerechneten Prozedur (Syntax) und das Anzeigen von Tabellen und Grafiken erfolgt in Form von Blöcken, d.h. jede Prozedur bzw. Tabelle/Grafik ist als eigenes Objekt definiert. Für jede durchgeführte Prozedur wird zunächst ein eigener Block als *Log* angelegt, in dem für die durchgeführte Prozedur die dafür erforderliche Syntax dokumentiert und durch das „Log-Symbol“ im rechten Fenster gekennzeichnet wird. Darunter befindet sich dann der Block mit den tatsächlichen Ergebnissen bzw. Grafiken der durchgeführten Prozedur. Dieser Block ist so gestaltet, dass er mit dem jeweiligen Prozedurnamen beginnt (z.B. Häufigkeiten) und mit dem „Blocksymbol“ gekennzeichnet ist. Diesem ist wiederum ein kleines Minuszeichen vorangestellt, wodurch der gesamte Block geschlossen bzw. geöffnet werden kann (genau wie im Windows-Explorer). Ein einzelner Prozedurblock ist so aufgebaut:

- Titel und Anmerkungen
- Ausgabe
- Unterpunkte (in Abhängigkeit von der Prozedur)

#### Aufbau des Prozedurblocks

Diese Konstruktion der Gliederung ermöglicht ein komfortables Suchen, Umstellen, Kopieren und Löschen im Ausgabeteil (siehe Abschnitt 2.4).

Die Inhalte des Viewers können als separate Ausgabedatei abgespeichert werden, indem die folgende Menüfunktion verwendet wird: *Datei* → *Speichern*. Dann wird eine **Ausgabedatei** mit der **Endung .spv** erzeugt, die benannt werden muss und separat geöffnet werden kann.

### 2.2.3 Syntax-Editor

Der Syntax-Editor ermöglicht es dem Benutzer alle Prozeduren, die in SPSS umsetzbar sind, über die programmeigene Programmiersprache (Syntax) zu steuern. Der Syntax-Editor kann durch die folgenden Menüpunkte geöffnet werden: *Datei* → *Neu* → *Syntax*.

#### Funktion

Ähnlich wie der Viewer besteht der Syntax-Editor aus den folgenden zwei Teilen (siehe Abb. 2.6): Links befindet sich eine Gliederung, die den auskommentierten Text und die Syntaxbefehle beinhaltet. Durch Anklicken der einzelnen Glieder-

#### Aufbau

rungspunkte kann innerhalb der Syntax hin und her gesprungen werden. Der Aufbau der rechten Seite entspricht einem einfachen Textfenster. In diesen Teil kann die Syntax eingegeben bzw. eingefügt werden (siehe Abschnitt 2.3). Die Inhalte des Syntax-Editors können ebenfalls in einer eigenen Syntaxdatei abgespeichert werden, indem die folgende Menüfunktion verwendet wird: *Datei* → *Speichern*. Dann wird eine **Syntaxdatei** mit der **Endung .sps** erzeugt, die benannt werden muss und separat geöffnet werden kann.

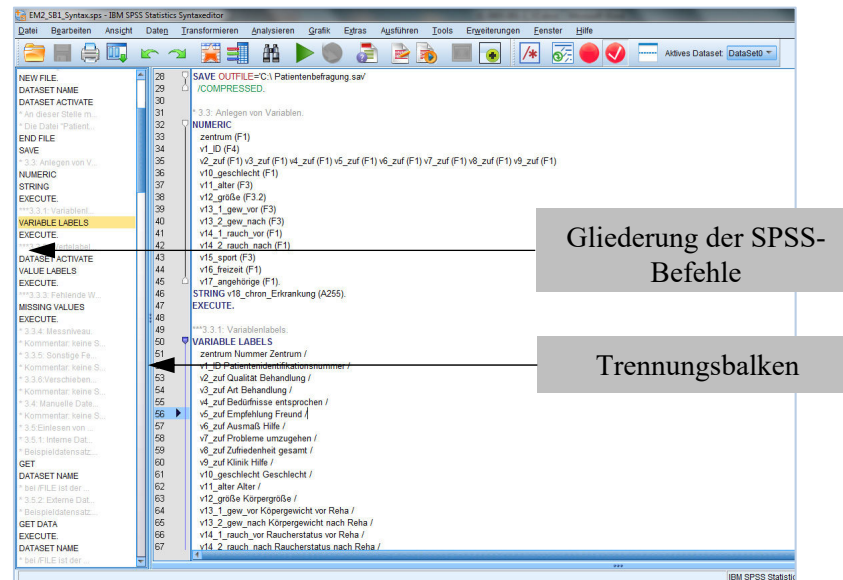


Abb. 2.6: Syntax-Editor

### 2.3 Prinzipielles zur Verwendung der Syntax

Das Arbeiten mit der Programmiersprache (Syntax) von SPSS ist mit einer Reihe von Vorteilen verbunden (siehe Abschnitt 2.3.1). Allerdings ist das Erlernen der Syntaxsprache gerade für Einsteiger, die noch nie mit einem Statistikprogramm gearbeitet haben, oft schwieriger als die im Vergleich intuitive Menüführung von SPSS. Deshalb empfiehlt sich für Einsteiger, das Programm zunächst über das Menü und die Fenster zu erlernen.

#### Syntaxdatei zum Studienbrief

Denjenigen, die bereits mit anderen Statistikprogrammen gearbeitet haben (z. B. SAS, R), fällt es möglicherweise leichter, direkt mit dem Erlernen der Syntaxsprache von SPSS zu beginnen. Auch für Personen, die langfristig mit SPSS arbeiten wollen, ist das Erlernen der Programmiersprache zu empfehlen. Um den Einstieg zu erleichtern, werden die Grundlagen in Abschnitt 2.3.2 vermittelt. Darüber hinaus finden sich für alle Prozeduren, die in Kapitel 3 bis 6 vorgestellt werden, die entsprechenden Syntaxbefehle in der separaten Syntaxdatei *EM2\_SBI\_Syntax.sps*.

#### 2.3.1 Vor- und Nachteile der Syntax

Das Arbeiten mit der Syntax bietet dem Benutzer die folgenden Vorteile (vgl. Zöfel, 2002, S. 11 ff.):

- eine erhöhte Effizienz beim Aufbereiten und Auswerten der Daten,
- die Möglichkeit zum mehrmaligen Verwenden und Modifizieren einer Syntax,

**Vorteile**

- die Syntax ist ein Protokoll der Auswertungsschritte, wodurch die durchgeführten Analysen nachprüfbar werden und repliziert werden können,
- Kenntnisse der SPSS-Syntax-Sprache sind eine gute Vorbereitung auf andere Analysesoftwaretypen, die sich ausschließlich über Programmiersprachen bedienen lassen,
- die Syntax eröffnet neue Möglichkeiten bezüglich spezifischer Prozeduren, die nicht über die Menüpunkte gesteuert werden können.

Mit der Verwendung der Syntax gehen für den Benutzer auch einige Nachteile einher:

- der Aufwand für das Erlernen der Programmiersprache,
- eine mangelnde Anwenderfreundlichkeit,
- die Fehleranalyse in der Syntax ist aufwendig.

#### Nachteile

### 2.3.2 Grundlagen zur Befehlsstruktur und Anwendung der Syntax

Um mit der Syntax in SPSS arbeiten zu können, muss das Textfenster des Syntax-Editors verwendet werden (siehe Abschnitt 2.2.3). Hier besteht zum einen die Möglichkeit, die SPSS-Befehle direkt in den Texteditor einzugeben. Zum anderen können die Prozeduren zunächst wie gewohnt über die Menüs und Dialogboxen in SPSS zusammen geklickt werden. Jede Prozedur endet mit einer Dialogbox, in der der Benutzer dann seine Auswahl mit *OK* bestätigen muss. Statt mit *OK* zu bestätigen, kann auch auf *Einfügen* geklickt werden. Dann wird der Programmcode, der für das Ausführen der ausgewählten Prozedur notwendig ist, in das Syntaxfenster eingefügt. Diese Funktion eignet sich insbesondere für SPSS-Anfänger, da dadurch die Auswertungsschritte dokumentiert werden und repliziert werden können. Es ist auch mittlerweile in der psychologischen Forschung üblich, bei Veröffentlichung einer Forschungsarbeit sowohl die Datensätze als auch die Syntax-Dateien zu veröffentlichen, so dass jemand anhand der Syntax genau die vorgenommene Analyse mit allen Schritten nachvollziehen kann. Allein aus Gründen der methodischen Sorgfalt ist daher die Verwendung der Syntax ratsam.

#### Erzeugung der Syntax über Dialogboxen

Wie in Abschnitt 2.2.2 dargestellt wurde, wird jede durchgeführte Prozedur auch im Viewer in Form der Syntaxsprache dokumentiert. Allerdings ist es nicht möglich, mittels der im Viewer befindlichen Syntax die Prozedur erneut zur Ausführung zu bringen. Dies ist nur möglich, wenn sich die Syntax im Syntax-Editor befindet.

Sofern der Syntaxcode über das Betätigen der Taste *Einfügen* in einer Dialogbox in den Syntax-Editor eingefügt wurde, wird dadurch die entsprechende Prozedur NICHT automatisch ausgeführt. Dazu muss die Syntax erst aktiv markiert und dann zur Ausführung gebracht werden (siehe unten).

Als Beispiel für das **Erstellen der Syntax aus den Dialogboxen** heraus, soll eine Häufigkeitstabelle für die Variable „Geschlecht“ auf Basis der Datendatei *Zentrum1\_Spieldaten.sav* erstellt werden. Dazu muss zunächst die folgende Menüauswahl getroffen werden: *Analysieren* → *Deskriptive Statistiken* → *Häufigkeiten*. Daraufhin öffnet sich die Dialogbox *Häufigkeiten*. Dort ist aus der Quellvariablenliste die Variable *v10\_geschlecht* mithilfe der Transportschaltfläche in das Fenster *Variable(n)* zu bringen. Zudem ist darauf zu achten, dass die Checkbox *Häufigkeitstabellen anzeigen* aktiviert ist. Statt auf *OK* muss jetzt auf *Einfügen* geklickt werden. Dann wird der Syntaxbefehl in das Textfeld des Syntaxfensters geschrieben (siehe **Abb. 2.7**).

#### Beispielaufgabe

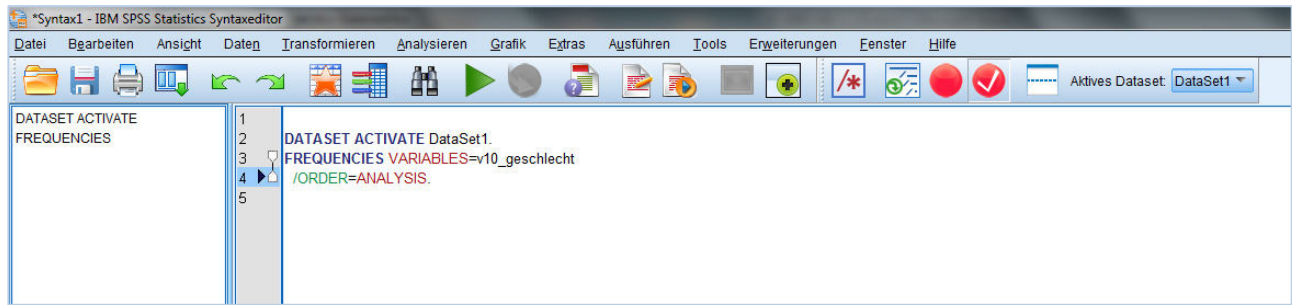




Abb. 2.7: Syntaxbefehl „Häufigkeiten“ im Syntax-Editor

Alternativ hätte dieser Befehl auch manuell direkt in das Syntaxfenster eingegeben werden können. Die eingefügte Syntax kann editiert werden, um z. B. individuelle Anpassungen vorzunehmen oder zusätzliche Möglichkeiten der Syntax zu nutzen. Ein bestehender SPSS-Befehl kann durch das Ausführen der folgenden beiden Schritte zur Ausführung gebracht werden:

### Ausführen der Syntax

1. Markieren der Syntax, indem die linke Maustaste gedrückt und mit dem Cursor der Maus über die entsprechenden Zeilen gefahren wird („Klicken- und Ziehen-Methode“).
2. Klicken mit der linken Maustaste auf den Pfeil  im spezifischen Menü des Syntax-Editors.

Um zu kontrollieren, ob die Ausführung des Befehls erfolgreich war, sollte ein Blick in den Viewer erfolgen. Ein erfolgreiches Ausführen lässt sich daran erkennen, dass der Befehl im Viewer wiederholt wird und die Tabellen bzw. Grafiken im Viewer angezeigt werden. Falls die Ausführung nicht erfolgt ist, erscheint im Viewer eine entsprechende Fehlermeldung.

Um die Häufigkeitstabelle in der Ausgabedatei zu erzeugen, muss die erzeugte Syntax markiert und der Pfeil  betätigt werden. Anschließend erscheint die Tabelle im Ausgabeteil (siehe Abb. 2.8).

**Häufigkeiten**

**Statistiken**

Geschlecht			
N	Gültig	90	
	Fehlend	0	

**Geschlecht**

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	weiblich	64	71,1	71,1	71,1
	männlich	26	28,9	28,9	100,0
	Gesamt	90	100,0	100,0	

Abb. 2.8: Ausgabe – Häufigkeitstabelle zu der Variable „Geschlecht“

Die unterschiedlichen Bausteine der Syntax sollen anhand der im obigen Beispiel erzeugten Syntax erläutert werden (siehe Abb. 2.7). Eine Anweisung (Synonyme: Befehl, Kommando) setzt sich aus den folgenden Komponenten zusammen (vgl. Zöfel, 2002, S. 14):

1. **Anweisung:** Eine Anweisung steuert den Ablauf von SPSS. In dem obigen Beispiel kommen die folgenden beiden Anweisungen vor:
  - a. *DATASET ACTIVATE:* Mithilfe dieser Anweisung wird definiert, auf welches Datenset (Synonym für Datendatei) die Anweisung angewendet werden soll. Dies ist insbesondere dann wichtig, wenn parallel mit mehreren geöffneten Datensets gearbeitet wird. Wenn die Syntax aus den Dialogboxen heraus erzeugt wird, dann fügt SPSS diesen Befehl immer automatisch ein und definiert den zuletzt aufgerufenen Daten-Editor als aktives Datenset. Wenn nur ein Datenset geöffnet ist, dann ist es automatisch Datenset 1. Wenn mehrere Befehle zu einem Datenset ausgeführt werden sollen, reicht es, wenn dieser Befehl einmal zu Beginn auftaucht.
  - b. *FREQUENCIES:* Anweisung für die Erstellung einer Häufigkeitstabelle.
2. **Unteranweisung:** Dies ist eine Zusatzanweisung zu einer Anweisung. Grundsätzlich kann jede Anweisung mehrere Unteranweisungen haben. In dem Beispiel ist z. B. *VARIABLES* eine Unteranweisung.
3. **Spezifikationen:** Hierbei handelt es sich um Angaben, wie Anweisungen bzw. Unteranweisungen hinzugefügt werden können. Dies können Schlüsselwörter, Zahlen, arithmetische Operatoren, Variablennamen und spezielle Trennzeichen enthalten. In dem Beispiel stellt *v10\_geschlecht* eine Spezifikation innerhalb des Unterbefehls *VARIABLES* dar.
4. **Schlüsselwörter:** Dies sind vordefinierte Wörter, die in Spezifikationen Anwendungen finden. In dem Beispiel ist *ANALYSIS* ein Schlüsselwort.

**Bausteine der Syntax**

Beim Erstellen einer Syntax sind die folgenden Regeln einzuhalten bzw. Aspekte zu beachten (vgl. Zöfel, 2002, S. 14):


**Syntaxregeln**

- Jede Anweisung muss in einer neuen Zeile beginnen und muss mit einem Punkt beendet werden.
- Die Trennung von Unteranweisungen erfolgt mit einem Schrägstrich. Dieser kann vor der ersten Unteranweisung weggelassen werden.
- Die Länge einer Syntaxzeile darf höchstens 80 Zeichen umfassen.
- Die Länge einer Anweisung ist unbegrenzt und kann sich über mehrere Zeilen erstrecken.
- Das Einfügen von Leerzeichen oder der Beginn einer neuen Zeile ist an allen Stellen möglich, an denen auch einzelne Leerzeichen erlaubt sind – wie vor und nach Schrägstrichen, Klammern, arithmetischen Operatoren oder zwischen Variablen.
- Wenn Text in Apostrophe gesetzt wird (z. B. bei Variablenlabels), muss sich dieser Text auf einer Zeile befinden.
- In Spezifikationen muss als Dezimaltrennzeichen ein Punkt verwendet werden – unabhängig von der Voreinstellung im Windows-System.
- Grundsätzlich wird bei der Interpretation der Syntax nicht zwischen Groß- und Kleinschreibung unterschieden (Ausnahme: in Apostrophe gesetzte Variablenlabels).

SPSS-Anweisungen lassen sich in die folgenden beiden Gruppen unterteilen:

**Gruppen von SPSS-Anweisungen**

1. Statistische Prozeduren
2. Sonstige Anweisungen

SPSS-Befehle zu statistischen Prozeduren werden direkt nach dem Markieren und durch das Klicken auf den Pfeil  ausgeführt. Die meisten der sonstigen Anweisungen, z. B. zur Datenaufbereitung, werden erst nach der nachfolgenden Prozeduranweisung ausgeführt. Sofern eine sonstige Anweisung einzeln ausgeführt werden soll, kann die spezifikationslose EXECUTE-Anweisung verwendet werden. Beim Arbeiten mit Dialogboxen wird diese bei den Prozeduren immer automatisch mit eingefügt.

### Arbeiten mit Kommentaren

Es ist möglich und empfehlenswert, Kommentare des Benutzers in einer Folge von SPSS-Anweisungen zu platzieren, um die Auswertungsschritte zu erläutern. Damit diese von SPSS als Kommentar erkannt werden, müssen sie mit dem Schlüsselwort COMMENT oder einem \* beginnen. Kommentare sind in ihrer Länge unbegrenzt und können sich über mehrere Zeilen erstrecken. Um SPSS zu signalisieren, dass ein Kommentar zu Ende ist, muss dieser mit einem Punkt beendet werden (siehe die Kommentare in der Syntaxdatei *EM2\_SB1\_Syntax.sps*).

In den folgenden Kapiteln des Studienbriefs wird die Durchführung aller Prozeduren mittels der Benutzeroberfläche bzw. Menüpunkten erläutert. Dem Studienbrief ist jedoch eine Syntaxdatei beigelegt, in der zu allen Prozeduren der entsprechende Syntaxbefehl zu finden ist (vgl. *EM2\_SB1\_Syntax.sps*).

## 2.4 Arbeiten mit dem Viewer

Der Aufbau des Viewers (auch Ausgabeteil, Ausgabefenster genannt) ermöglicht ein effizientes Suchen, Löschen und Umstellen von Inhalten (z. B. Tabellen, Grafiken und Texte). Zudem können Inhalte editiert werden. Welche Möglichkeiten hier zur Verfügung stehen und wie diese genutzt werden können, wird im Folgenden erläutert.

### Suchen

Für das Suchen bestimmter Inhalte im Viewer gibt es drei Möglichkeiten, wobei die Zielgenauigkeit bei der dritten Option am größten ist:

1. Betätigen der Pfeiltaste/Schiebeleiste am rechten Rand
2. Durch Klicken in der Gliederung (links) auf das entsprechende Symbol
3. Gebrauch der Suchfunktion (Menüleiste *Bearbeiten* → *Suchen*)

### Löschen

Für das Löschen bestimmter Teile in der Ausgabe, müssen diese zunächst entweder durch Klicken auf das entsprechende Symbol in der Gliederung oder durch Anklicken der Parts im rechten Teil der Ausgabe markiert werden. Anschließend muss die Taste [Entf] betätigt werden oder die Menüauswahl *Bearbeiten* → *Löschen*.

### Umstellen

Für das Umstellen bestimmter Teile der Ausgabe, müssen diese zunächst entweder durch Klicken auf das entsprechende Symbol in der Gliederung oder durch Anklicken der Parts im rechten Teil der Ausgabe markiert werden. Anschließend können diese bei gedrückter Maustaste an die entsprechende Stelle gezogen werden. Alternativ kann nach der Auswahl die folgende Menüoption verwendet werden: *Bearbeiten* → *Ausschneiden*. Anschließend muss dann auf das Symbol in der Gliederung bzw. den Teil der Ausgabe geklickt werden, hinter dem die Einfügung erfolgen soll, um dann aus dem Menü Folgendes auszuwählen: *Bearbeiten* → *Einfügen nach*.

### Kopieren

Für das Kopieren bestimmter Teile der Ausgabe (d. h. der Teil verbleibt auch an entsprechender Stelle), müssen diese zunächst entweder durch Klicken auf das entsprechende Symbol in der Gliederung oder durch Anklicken der Parts im rechten Teil der Ausgabe markiert werden. Anschließend können diese bei gedrückter Maustaste und gedrückter Kontrolltaste an die entsprechende Stelle gezogen werden. Alternativ

können nach der Auswahl die folgenden Menüoptionen verwendet werden: *Bearbeiten* → *kopieren*. Anschließend muss dann auf das Symbol in der Gliederung bzw. den Teil der Ausgabe geklickt werden, hinter dem die Einfügung erfolgen soll, um dann aus dem Menü folgendes auszuwählen: *Bearbeiten* → *Einfügen nach*.

Für das Bearbeiten bestehender Titel, Texte, Tabellen und Grafiken im Viewer muss im rechten Teil der Ausgabe der jeweilige Titel oder Text per Doppelklick aktiviert werden. Dann kann dieser mit den üblichen Funktionen (z.B. Tasten [Entf] oder [Einfg]) bearbeitet werden. Für die Veränderung der Schriftart und -größe steht sogar eine eigene Menüleiste zur Verfügung.

## Bearbeiten

Für das Bearbeiten von Tabellen und Grafiken müssen diese ebenfalls per Doppelklick aktiviert werden. Die einzelnen Inhalte von Tabellen und Grafiken können genau wie Texte und Titel direkt editiert werden, indem man auf den jeweiligen Text klickt und ihn entsprechend verändert. Darüber hinaus kann das spezifische Design von Tabellen über den Pivot-Tabellen-Editor geändert werden. Die Aktivierung dieses Editors erkennt man durch die geänderte Menüleiste, sodass man dort *Pivot* auswählen kann. Für das Editieren von Grafiken steht ebenfalls ein eigener Diagramm-Editor zur Verfügung. Dieser öffnet sich automatisch, wenn eine Grafik per Doppelklick aktiviert wurde. Eine Anleitung zum Arbeiten mit dem Diagramm-Editor wird Studienbrief 2 „Deskriptive Analysen mit SPSS“ gegeben.

Für das Einfügen von zusätzlichen Titeln und Texten muss zunächst auf dasjenige Objekt geklickt werden, hinter dem der Text bzw. Titel eingefügt werden soll. Anschließend ist folgende Menüfunktion zu verwenden: *Einfügen* → *Neuer Titel* oder *Einfügen* → *Neuer Text*. Daraufhin erscheint ein Textfeld im Editor, in das Text eingegeben werden kann.

## Einfügen von Titeln und Texten

## 2.5 Überblick über die allgemeine Menüleiste

Auf die allgemeine Menüleiste kann aus jedem Fenster des Programms zugegriffen werden. Neben den üblichen Menüpunkten eines Microsoft-Office-Programms, wie *Datei* und *Bearbeiten* gibt es auch SPSS-spezifische Menüpunkte, wie *Daten*, *Transformieren* und *Analysieren*. **Abb. 2.9** gibt einen Überblick zu den zentralen Funktionen der einzelnen Menüpunkte. Für ein effektives Arbeiten mit SPSS sind insbesondere die hier hervorgehobenen Menüpunkte von **zentraler Bedeutung**. Ausgewählte Funktionen innerhalb dieser Menüpunkte, werden in den in Klammern angegebenen Kapiteln dieses Studienbriefs erläutert.

## Zentrale Menüpunkte



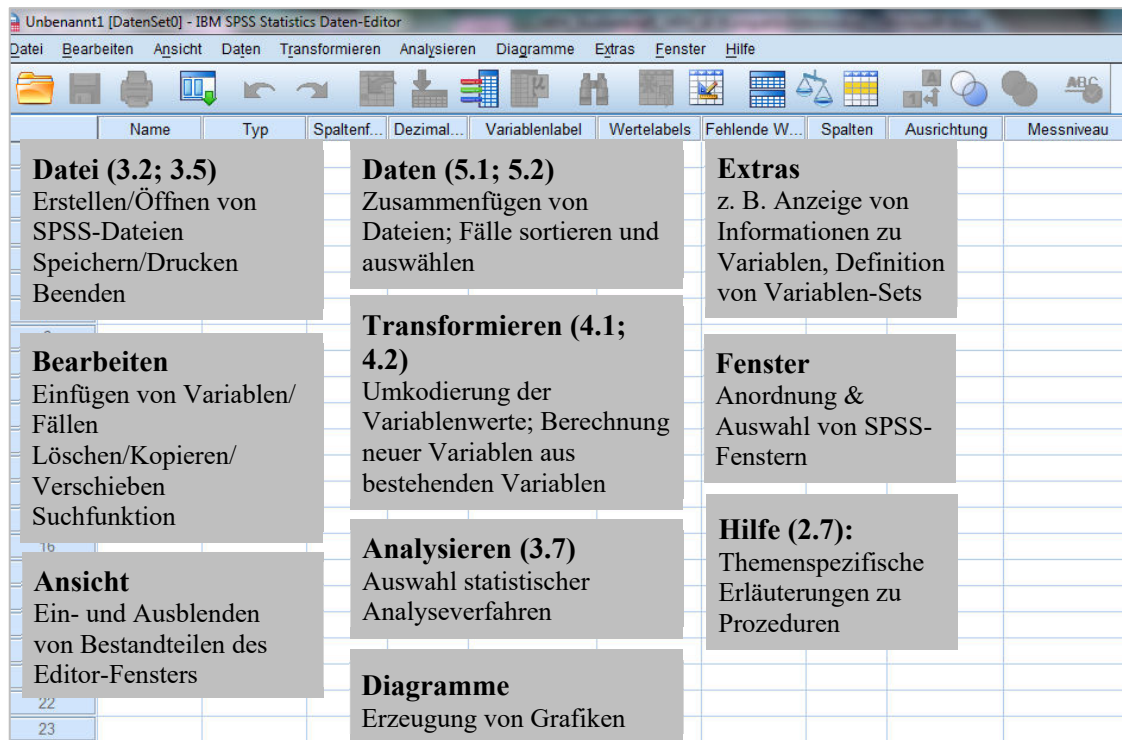


Abb. 2.9: Funktionen der allgemeinen Menüpunkte im Überblick

### 2.6 Allgemeine Voreinstellungen

Für Einsteiger in SPSS wird empfohlen, die allgemeinen Voreinstellungen anfangs unverändert zu lassen. Das heißt, dieses Kapitel kann von Einsteigern zunächst übersprungen werden. Nach dem Einarbeiten in das Programm kann es jedoch sein, dass man als Benutzer bestimmte Vorlieben entwickelt, z. B. hinsichtlich der Schriftgröße in der Ausgabe oder welche Informationen genau in den Ergebnistabellen angezeigt werden sollen. Über den Menüpunkt *Bearbeiten* → *Optionen* öffnen sich eine Reihe von Registerkarten (siehe Abb. 2.10), die es ermöglichen, bestimmte Voreinstellungen nach den eigenen Wünschen zu modifizieren. Hier wird nur eine Auswahl an möglichen Optionen dargestellt:

#### Einstellungsoptionen

##### Registerkarte *Allgemein*

Hier gibt es die Möglichkeit, die Anzeige der Variablenliste in den Dialogboxen anzupassen, z. B. kann bei *Auswahl* eingestellt werden, wie in den Dialogboxen die Variablen angezeigt werden sollen (z. B. der volle Name oder nur das Label der Variablen oder auch die Reihenfolge der Variablen). Zudem kann die Sprache der Benutzeroberfläche und der Ausgabe eingestellt werden.

##### Registerkarte *Viewer*

Hier können z. B. für die unterschiedlichen Blöcke in der Ausgabedatei Schriftart und Schriftgröße festgelegt werden.

##### Registerkarte *Ausgabe*

Hier ist es möglich, einzustellen, welche Informationen zu den Variablen (Name oder Variablenlabel) in der Gliederung des Viewers oder in den (Pivot-)Tabellen (Werte und/oder Wertelabels) in der Ausgabedatei angezeigt werden sollen.



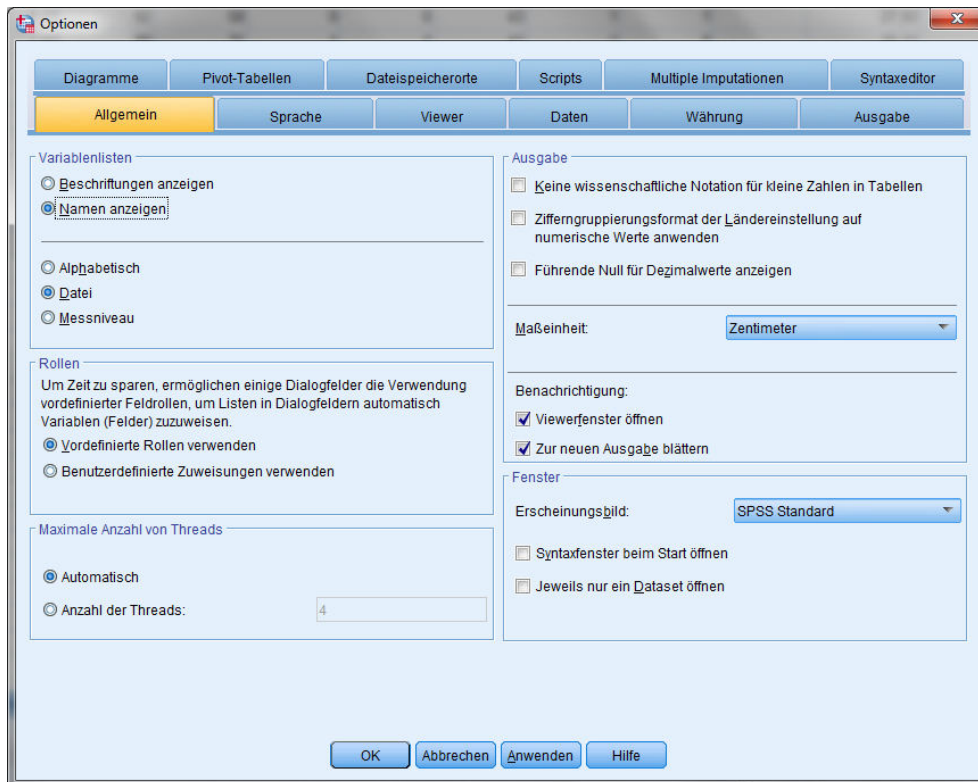


Abb. 2.10: Registerkarten

## Übungsaufgaben

- 2.1) Welche drei Fenster sind für die Arbeit mit dem Programm SPSS zentral? Bitte benennen Sie stichpunktartig ihre Funktion.
- 2.2) Auf welches Element der Benutzeroberfläche kann von jedem Fenster in SPSS zugegriffen werden?
- 2.3) Benennen Sie die Vor- bzw. Nachteile, die mit dem Arbeiten der SPSS-Syntax verbunden sind.

### 3 Aufbereitung der Daten

Bevor mit der eigentlichen statistischen Analyse der Daten begonnen werden kann, müssen diese zunächst entsprechend aufbereitet werden. In Abhängigkeit davon, wie die Daten vorliegen, ist der Prozess der Datenaufbereitung häufig zeitlich aufwendiger als die anschließende Datenanalyse.

Wenn die Daten – wie in dem Anwendungsbeispiel die Befragungsdaten der Patientinnen und Patienten – ursprünglich in Papierform (Fragebogen) vorliegen, besteht ein wesentlicher Schritt der Datenaufbereitung darin, die Daten zunächst in das Auswertungsprogramm SPSS zu transferieren. Alternativ können die Daten bereits digital in Form einer Datendatei vorliegen, sodass diese nur noch in SPSS eingelesen werden müssen. Eine besondere Herausforderung im Rahmen der Datenaufbereitung kann auch darin bestehen, unterschiedliche Datendateien zusammenzufügen. In den folgenden Kapiteln werden die einzelnen Schritte der Datenaufbereitung für die unterschiedlichen Situationen erläutert.

#### 3.1 Kodierung und Kodeplan

Damit Daten in SPSS verwaltet und analysiert werden können, müssen sie eine ganz bestimmte Form aufweisen. Diese Form ergibt sich aus dem Aufbau der Datenansicht des Daten-Editors. Demnach können Daten nur dann sinnvoll weiterverarbeitet werden, wenn die Daten zu einem Fall in einer Zeile stehen und jede Spalte eine Variable umfasst.

Insbesondere dann, wenn die Daten von einzelnen Papierbögen manuell in den Daten-Editor übertragen werden müssen, besteht der erste Schritt darin, einen sog. Kodeplan zu erstellen. Laut Bühl (2010) hat ein Kodeplan die folgende Aufgabe:

##### Aufgaben des Kodeplans

- Ein Kodeplan ordnet den einzelnen Fragen des Fragebogens Variablennamen zu.
- Ein Kodeplan ordnet den Merkmalsausprägungen einer Variablen Kodenummern zu.

Durch die Kodierung der Fragen wird sichergestellt, dass jede Frage eindeutig in der späteren Datendatei zu identifizieren ist. Da insbesondere statistische Prozeduren ausschließlich Ziffern weiterverarbeiten können, ist es zudem erforderlich, den Merkmalsausprägungen Ziffern zuzuordnen – sofern sie nicht sowieso schon als solche im Rahmen von halb-offenen Fragen erhoben werden (wie z.B. bei Frage 11 „Alter“). Die Zuordnung von Ziffern zu Merkmalsausprägungen ist natürlich nur bei geschlossenen Antwortformaten möglich. Grundsätzlich können auch Fragen mit offenen Antwortformaten, die eine Antwort in Textform erforderlich machen, in SPSS eingegeben und weiterverarbeitet werden. Allerdings sind die Möglichkeiten ihrer Auswertung mit SPSS eher begrenzt.

In **Anhang 2** ist ein solcher Kodeplan für den Fragebogen des Anwendungsbeispiels dargestellt. Hierzu eine kurze Erläuterung:

Frage 2 „Wie würden Sie die Qualität der Behandlung, welche Sie erhalten haben, beurteilen?“ wurde z. B. der Variablenname *v2\_zuf* zugeordnet. Die Merkmalsausprägungen dieser Frage erhalten die folgenden Kodierungen: „ausgezeichnet“ – 1, „gut“ – 2, „weniger gut“ – 3 und „schlecht“ – 4. Die Ziffern sind als Index an den Kästchen dargestellt. Bei Frage 13 zum Körpergewicht ist es notwendig, sowohl für die Gewichtsangabe vor der Reha als auch für die Gewichtsangabe nach der Reha eine separate Variable zu definieren. Allerdings braucht den Merkmalsausprägungen keine spezifische Kodierung mehr zugeordnet werden, da die Ausprägungen per se schon Kodierungen darstellen. Frage 18 „Chronische Erkrankung“ wurde mit *v18\_chron\_Erkrankung* kodiert und ist ein Beispielitem mit offenem Antwortformat. Da die zu erwartenden Ausprägungen unbekannt sind, erhält diese Frage keine spezifische Kodierung, sodass die Antworten einfach so in SPSS übertragen werden können.

### Erläuterung des Kodesplans

Grundsätzlich ist ein Kodeplan nicht nur für Befragungsdaten sinnvoll, sondern er sollte auch für bereits existierende Datendateien erstellt werden, sodass der Benutzer weiß, welche Variable sich hinter einer Kodierung verbirgt bzw. wie bestimmte Merkmalsausprägungen definiert sind.

## 3.2 Anlegen und Speichern einer neuen Datendatei

Sofern Daten von Papierbögen in das Programm SPSS manuell übertragen werden sollen, besteht der erste Schritt darin, eine neue Datendatei anzulegen. Dies kann auf zwei unterschiedliche Weisen umgesetzt werden:

1. Beim Starten von SPSS ist in der Dialogbox auf *Daten eingeben* zu gehen (siehe Abschnitt 2.1).
2. Sofern SPSS bereits geöffnet ist, kann man eine neue Datei folgendermaßen anlegen: *Datei* → *Neu* → *Daten*.

### Anlegen einer Datendatei

In beiden Fällen erscheint ein leerer Daten-Editor und die Datei wird mit *unbenannt* betitelt. Die Datendatei kann in SPSS erst abgespeichert werden, wenn sowohl Variablen angelegt (siehe Abschnitt 3.3) als auch Daten eingegeben worden sind (siehe Abschnitt 3.4).

Zum Abspeichern der Datendatei ist im Menü auf *Datei* → *Speichern unter* zu gehen und ein entsprechender Name für die Datei z. B. „Patientenbefragung“ zu vergeben. Zudem ist auszuwählen, in welchem Unterverzeichnis die Datei abzulegen ist.

### Speichern einer Datendatei

## 3.3 Anlegen von Variablen

Für das Anlegen von neuen Variablen ist zunächst im Daten-Editor auf die Variablenansicht zu gehen (siehe Abschnitt 2.2.1). In jeder Zeile (siehe **Abb. 3.1**) kann nun eine Variable angelegt werden, indem die unterschiedlichen Spalten ausgefüllt werden. Es wird empfohlen, links in der Tabelle mit der Vergabe des Namens für die Variable zu beginnen und sich dann nach rechts vorzuarbeiten. Der Name der Variablen sollte der Kodierung der Variablen auf dem Kodeplan entsprechen. Zum Anlegen einer neuen Variablen geht man in die Zelle der ersten Spalte („Name“), um dort die entsprechende Kodierung (entsprechend des Kodeplans) für die Variable einzugeben.

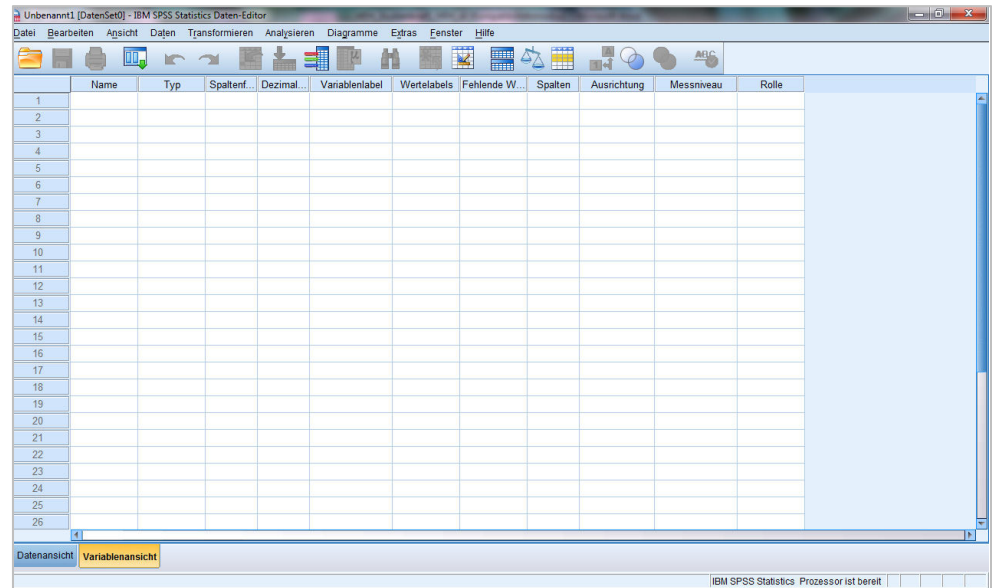


Abb. 3.1: Leere Variablenansicht

Bei der Benennung von Variablen in SPSS sind die folgenden Regeln zu beachten, welche sinnvollerweise auch bereits beim Erstellen des Kodeplans zu berücksichtigen sind (vgl. Bühl, 2010, S. 37):

#### Regeln für Variablennamen

- Für die Variablenbenennung können sowohl Buchstaben als auch Ziffern verwendet werden. Zudem sind die folgenden Sonderzeichen erlaubt:
  - (underscore – Unterstrich),
  - . (Punkt)
  - @, # und \$
- Leerzeichen sowie spezifische Sonderzeichen wie !, ? und \* sind **nicht** erlaubt.
- Die Variablennamen müssen mit einem Buchstaben beginnen.
- Das letzte Zeichen eines Variablennamens darf kein Punkt (.) und sollte kein Unterstrich ( \_ ) sein.
- Die maximale Zeichenlänge für Variablennamen ist 64.
- Die Groß- und Kleinschreibung spielt bei den Variablennamen keine Rolle, d. h. es sind auch beliebige Kombinationen aus Groß- und Kleinschreibung innerhalb eines Variablennamens möglich. Die gewählte Kombination bleibt auch in der Anzeige erhalten.
- Es ist nicht möglich, Variablennamen doppelt zu vergeben.
- Variablennamen dürfen nicht die gleiche Bezeichnung wie die folgenden reservierten Schlüsselwörter haben:
  - ALL, AND, BY, EQ, GE, GT, LE, LT, NE, NOT, OR, TO, WITH

Nachdem man einen Variablennamen vergeben hat, kann man zur nächsten Spalte weitergehen, um seine Angabe zu bestätigen – entweder mithilfe der Tabulator-Taste oder mit der Maus. Dies bewirkt, dass automatisch von SPSS alle weiteren Zellen zu der Variablen aufgefüllt werden. Diese können und sollten aber entsprechend der Empfehlungen in den folgenden Kapiteln weiter angepasst und spezifiziert werden.

Es werden im Folgenden die **Schritte des Anlegens der Variablen** für die folgenden Items des Fragebogens präsentiert:


**2. Frage:** Wie würden Sie die Qualität der Behandlung, welche Sie erhalten haben, beurteilen?

**11. Frage:** Alter

**18. Frage:** Leiden Sie neben Typ-II-Diabetes unter weiteren chronischen Erkrankungen? Falls ja, welche?

Zum Anlegen von drei neuen Variablen für die Fragen 2, 11 und 18 müssen die Variablennamen laut Kodeplan jeweils in ein separates Feld eingetragen werden, d.h. in der ersten Zeile ist für die 2. Frage `v2_zuf` einzutragen. Entsprechend ist `v11_alter` für die 11. Frage und `v18_chron_Erkrankung` für die 18. Frage einzutragen.

## Anwendungsbeispiel

In SPSS werden neuangelegte Variablen automatisch als numerische Variablen mit einer Länge von acht Zeichen angelegt. Diese Voreinstellung kann man ändern, indem man in der Spalte *Typ* mit der linken Maustaste auf die Schaltfläche mit den drei Punkten  in der jeweiligen Zelle der entsprechenden Variablen klickt. Dann öffnet sich eine Dialogbox, in der die unterschiedlichen Variablentypen angezeigt werden (siehe *Abb. 3.2*). Die drei wichtigsten Variablentypen sind die Folgenden:

## Variablentypen

- **Numerisch:** Bei diesem Variablentyp sind als gültige Werte Ziffern mit vorangestelltem Plus- oder Minuszeichen mit oder ohne Dezimaltrennzeichen möglich. Es wird, sofern vorhanden, nur das Minuszeichen angezeigt. In das Textfeld *Breite* kann die maximale Anzahl an Zeichen eingetragen werden und an welcher Stelle sich das Dezimaltrennzeichen befinden soll, indem man im Textfeld *Dezimalstellen* die Anzahl an Dezimalstellen angibt, die angezeigt werden sollen. Das Dezimaltrennzeichen kann entweder ein Punkt oder Komma sein und ist abhängig von der Spezifikation in der Dialogbox *Ländereinstellungen* der Windows-Systemsteuerung. Im Daten-Editor wird immer nur die hier definierte Anzahl an Dezimalstellen angezeigt, wobei Werte mit mehr Dezimalstellen gerundet werden. Für weiterführende Berechnungen wird der vollständige Wert verwendet. In der Regel wählt man für alle Variablen, die als Variablenausprägung im Kodeplan Ziffern haben, diesen Variablentypen.
- **Datum:** Gültige Werte sind bei diesem Variablentyp Datums- und/oder Zeitangaben. Insgesamt stehen 27 verschiedene Datums- und Zeitformate zur Verfügung. Bei Datumsvariablen können Schrägstriche, Trennstriche, Leerzeichen, Kommas oder Punkte als Trennzeichen zwischen den Werten für z. B. Tag, Monat und Jahr verwendet werden. Es ist möglich, zwischen verschiedenen Datumsformaten zu wählen (z. B. `tt-mm-jjjj`, `mm/tt/jjjj`). Bei der Wahl von `tt-mm-jj` als Datumsformat werden Trennstriche als Trennzeichen und Abkürzungen mit drei Buchstaben für Monate angezeigt. Bei der Wahl `mm/tt/jj` und `mm/tt/jjjj` werden Schrägstriche als Trennzeichen und Zahlen für Monate angezeigt. Bei Zeitformaten können als Trennzeichen zwischen Stunden, Minuten und Sekunden Doppelpunkte, Punkte oder Leerzeichen verwendet werden.
- **String:** Dieser Variablentyp ist geeignet für jegliche Art von Zeichenketten bzw. Freitextangaben. Diese dürfen Buchstaben, Ziffern, Sonderzeichen (z. B. auch Leerzeichen) umfassen. In der Dialogbox *Zeichen* kann man die gewünschte Anzahl an Zeichen definieren. Die maximale Zeichenlänge liegt bei 255. Wenn man weniger als die maximale Zeichenlänge eingibt bzw. einliest, werden die Angaben bis zur definierten maximalen Länge mit Leerzeichen aufgefüllt. Mit diesem Variablentyp können keine Rechenoperationen durchgeführt werden.

## Ausgewählte Variablentypen

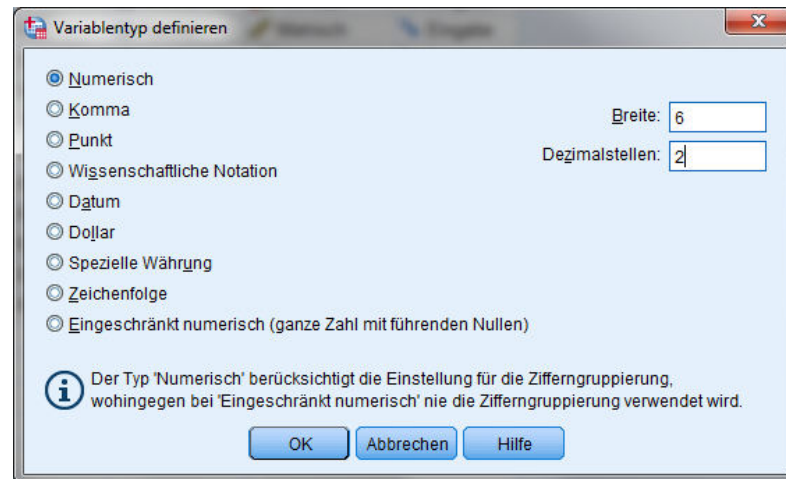


Abb. 3.2: Dialogbox – Variablentyp definieren

#### Anwendungsbeispiel

##### Anpassung von Voreinstellungen

Für die Beispielvariablen sind unter *Typ* die Voreinstellungen folgendermaßen anzupassen:

*v2\_zuf*: Numerisch, Breite = 1, Dezimalstellen = 0

*v11\_alter*: Numerisch, Breite = 3, Dezimalstellen = 0

*v18\_chron\_Erkrankung*: String, Zeichen 255 (maximal mögliche Zeichenanzahl)

Die folgenden Spalten *Spaltenformat* und *Dezimalstellen* ergeben sich aus den Einstellungsmöglichkeiten der Dialogbox *Variablentyp definieren*, der *Variablenlänge* und *Dezimalstellen*. Diese können aber auch über die separaten Spalten entsprechend angepasst werden.

#### 3.3.1 Variablenlabels

Mithilfe des Variablenlabels kann eine Variable näher beschrieben werden, wobei eine maximale Länge von 255 Zeichen zulässig ist. Im Gegensatz zu dem Variablennamen wird bei den Variablenlabels die Groß- und Kleinschreibung beachtet. Entsprechend der Eingabe werden diese auch angezeigt.

#### Anwendungsbeispiel

##### Vergabe von Variablenlabels


Für die Beispielvariablen sind für die Variablenlabels die folgenden Labels zu vergeben:

*v2\_zuf*: Qualität Behandlung

*v11\_alter*: Alter

*v18\_chron\_Erkrankung*: Chronische Erkrankung

#### 3.3.2 Wertbeschriftungen

Mithilfe der Wertbeschriftungen können Merkmalsausprägungen näher beschrieben werden, wobei ein Wertelabel maximal 60 Zeichen lang sein darf. Ein Wertelabel kann vergeben werden, indem man mit der linken Maustaste auf die Schaltfläche mit den drei Punkten klickt . Dann öffnet sich eine Dialogbox (siehe Abb. 3.3:). Hier kann man bei *Wert* den Wert der Merkmalsausprägung eingeben und bei *Beschriftung* die nähere Beschreibung der Merkmalsausprägung. Wichtig ist, dass man seine

Eingabe mit *OK* bestätigt, damit die Angaben gespeichert werden. Bereits bestehende Variablenlabels können angepasst werden, indem man das entsprechende Label anklickt, die gewünschten Änderungen vornimmt und anschließend auf *Ändern* klickt. Wichtig ist abschließend wieder auf *OK* zu klicken, um Eingaben zu bestätigen. Bestehende Wertelabels kann man löschen, indem man sie durch Anklicken markiert, auf *Entfernen* klickt und abschließend mit *OK* bestätigt.

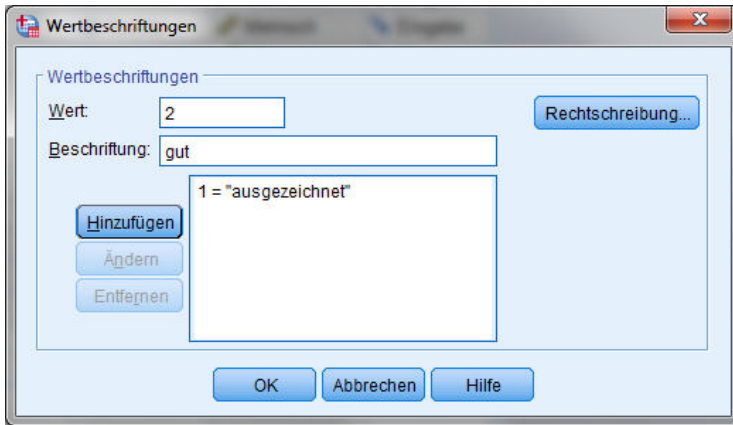


Abb. 3.3: Dialogbox – Wertbeschriftungen

#### Vergabe von Wertelabels

Für die Beispielsvariablen sind für die Wertelabels die folgenden Labels zu vergeben:

*v2\_zuf:* 1 = ausgezeichnet

2 = gut

3 = weniger gut

4 = schlecht

9 = keine Angabe

*v11\_alter:* 999 = keine Angabe


*v18\_chron\_Erkrankung:* keine Wertelabels erforderlich

#### Anwendungsbeispiel

### 3.3.3 Fehlende Werte

Bei fast jeder empirischen Erhebung gibt es fehlende Werte. Damit ist gemeint, dass für einzelne Fälle bei einzelnen Variablen Angaben fehlen. Für das Auftreten von fehlenden Werten kann es unterschiedliche Gründe geben. Bei Befragungen kann es z. B. vorkommen, dass Personen einfach eine Frage übersehen haben oder bei sensiblen Fragen auch eine Antwort verweigern (z. B. bei Fragen nach dem Einkommen). Wenn man ganze Datendateien übernimmt, kann es ebenfalls sein, dass bei einzelnen Variablen Werte fehlen. Insbesondere beim Anlegen neuer Variablen mit anschließender Dateneingabe sollte man sich überlegen, wie man damit umgeht, wenn eine Person zu einer Frage keine Antwort gegeben hat bzw. die Antwort ungültig ist. Prinzipiell gibt es zwei Möglichkeiten mit fehlenden Angaben umzugehen, die sich auch beide in SPSS abbilden lassen:

#### Umgang mit fehlenden Werten

1. **Benutzerdefinierte fehlende Werte:** Der Benutzer legt einen Wert fest, der einzugeben ist, wenn eine Person keine Angabe gemacht hat. Üblich ist bei einstelligen Kodierungen die 9, bei zweistelligen Kodierungen die 99 usw. als fehlende Werte zu definieren. Wichtig ist, dass man eine Ziffer wählt, die nicht bereits durch eine Antwortmöglichkeit vergeben wurde bzw. die nicht zur möglichen Zahlenmenge gehört. Damit diese Kodierung von SPSS als fehlend erkannt wird, ist es erforderlich, den entsprechenden Wert als fehlend zu definieren, indem man in der entsprechenden Zelle mit der linken Maustaste auf die Schaltfläche mit den drei Punkten  klickt. Dann öffnet sich eine Dialogbox, in der man die Möglichkeit hat, einzelne Werte oder ganze Wertbereiche als fehlende Werte zu definieren. Nachdem man die fehlenden Werte definiert hat, muss man seine Eingaben mit *OK* bestätigen. Insbesondere bei einer manuellen Dateneingabe ist es empfehlenswert, für jede Frage Werte für fehlende Angaben zu definieren. Denn nur so ist es möglich, zu kontrollieren, ob man selbst die Angabe vergessen hat oder ob es eine fehlende Angabe im Fragebogen war. Zudem kann man bei den Wertelabels auch eine Bezeichnung für die fehlenden Werte vergeben. Dies bietet sich insbesondere dann an, wenn es verschiedene Arten von fehlenden Werten gibt (z.B. „keine Angabe“ oder „Angabe ungültig“). Falls der Benutzer keine Werte als fehlend definieren möchte, kann er diese Zelle auch einfach mit der Tabulatortaste überspringen, da „keine“ voreingestellt ist.
2. **Systemdefinierte fehlende Werte:** Zudem hat der Benutzer die Möglichkeit für Personen, die keine Angabe gemacht haben, die entsprechende Zelle einfach in der Datenansicht frei zu lassen. Sofern in der Datenansicht Zellen von numerischen Werten nicht aufgefüllt werden, dann wird dies SPSS intern als systemdefinierter fehlender Wert abgespeichert. Diese Art von Werten werden in der Datenansicht mit einem Punkt (.) angezeigt.

#### Ausschluss von der Auswertung

Sowohl das Vorhandensein von benutzer- als auch systemdefinierten fehlenden Werten hat zur Folge, dass die entsprechenden Fälle, von der Auswertung in Bezug auf diese Variable ausgeschlossen werden.

#### Anwendungsbeispiel

##### Definieren von benutzerdefinierten fehlenden Werten

Für die Beispielvariablen sind die folgenden benutzerdefinierten fehlenden Werte zu definieren:

*v2\_zuf:* 9

*v11\_alter:* 999

*v18\_chron\_Erkrankung:* nicht erforderlich

### 3.3.4 Messniveau

Es besteht die Möglichkeit, das Messniveau (d. h. Datenniveau) für die jeweilige Variable zu definieren. Wie bekannt ist, bestimmt das Messniveau darüber, welche statistischen Verfahren anwendbar sind. SPSS unterscheidet zwischen nominalem, ordinalem und metrischem (= intervallskaliert, verhältnisskaliert) Messniveau. Beim Anlegen einer neuen Variablen wird das Messniveau automatisch auf *Unbekannt* gesetzt. Zum Ändern des Datenniveaus ist es erforderlich, die Schaltfläche in der entsprechenden Zelle zu bedienen und das passende Messniveau auszuwählen.



Diese Einstellungen haben jedoch keine Konsequenzen für die Anwendbarkeit bestimmter statistischer Prozeduren. Hier muss der Benutzer immer selbst kritisch hinterfragen, ob die jeweiligen Prozeduren sinnvoll sind.

#### Festlegen des Messniveaus

Für die Beispielvariablen sind die folgenden Messniveaus festzulegen:

*v2\_zuf*: Metrisch

*v11\_alter*: Metrisch

*v18\_chron\_Erkrankung*: Nominal

#### Anwendungsbeispiel

### 3.3.5 Sonstige Felder

Das Feld *Spalten* gibt an, welche Breite dem jeweiligen Datenfeld in der Datenansicht bei dem Anzeigen der Werte eingeräumt wird. Die Voreinstellung von „8“ kann entweder direkt durch Veränderung des Wertes und anschließende Betätigung der Return-Taste verändert werden. Alternativ kann die Variation der Spaltenbreite in der Datenansicht verändert werden, indem der Mauszeiger auf dem Rand zwischen zwei Variablennamen positioniert wird. Sobald ein Doppelpfeil auftaucht, ist dies ein Hinweis dafür, dass die entsprechende Spalte mithilfe der Klicken- und Ziehen-Technik vergrößert und verkleinert werden kann. Das Anpassen der Spaltenbreite (insbesondere die Verkleinerung) kann sinnvoll sein, um im Viewer möglichst viele Variablen auf einen Blick zu erfassen oder um bei längeren Variablennamen den kompletten Variablennamen lesen zu können.

Das Feld *Ausrichtung* ermöglicht die Einstellung, ob die Werte im Daten-Editor linksbündig, rechtsbündig oder zentriert angezeigt werden sollen. Dies kann über die Schaltfläche in der jeweiligen Zelle festgelegt werden und durch Betätigung der Return-Taste bestätigt werden. Über das Feld *Rolle* gibt es die Möglichkeit, für jede Variable festzulegen, welchen Zweck sie in der Auswertung erfüllt (z. B. Ziel, um eine abhängige Variable kenntlich zu machen). Einige Dialogfelder in SPSS unterstützen vordefinierte Rollen.

### 3.3.6 Verschieben und Löschen von Variablen

In der Variablenansicht können angelegte Variablen verschoben werden, indem die gesamte Zeile zu einer Variablen mit der linken Maustaste markiert wird. Dies wird dadurch deutlich, dass sie sich gelb einfärbt. Indem die linke Maustaste gedrückt bleibt und man sie hoch bzw. herunter bewegt (Klicken- und Ziehen-Technik), kann die Variable verschoben werden. In der Datenansicht können ebenfalls Variablen nach links oder rechts verschoben werden.

Zum Löschen einer Variablen muss ebenfalls die gesamte Zeile zu einer Variablen mit der linken Maustaste markiert werden. Anschließend muss der Mauszeiger auf der Zeile verweilen und durch Klicken der rechten Maustaste öffnet sich ein Fenster. Dort kann man dann die Option *Löschen* auswählen.

### 3.4 Manuelle Dateneingabe in den Daten-Editor

Sofern eine Datenerhebung mit einem papierbasierten Fragebogen durchgeführt worden ist, erfolgt nach dem erfolgreichen Anlegen aller Variablen (siehe Abschnitt 3.3) die Eingabe der Daten auf Basis der ausgefüllten Fragebögen. Dafür ist es erforderlich, in die *Datenansicht* zu gehen, sodass der Daten-Editor genutzt werden kann. Im Kopf der Tabelle sieht man die Namen der angelegten Variablen. Pro Zeile können nun manuell die Daten vom Fragebogen in die jeweiligen Zellen übertragen werden. Es wird eine **zeilenweise Dateneingabe** empfohlen. Das heißt, es sollte immer ein kompletter Fragebogen für einen Fall/Patienten am Stück eingegeben werden. Beispielhaft wird dies im Folgenden für Patient 0 demonstriert (siehe Anhang 3).

#### Anwendungsbeispiel

##### Dateneingabe in SPSS

Für die Variable *zentrum* ist in die erste Zeile in der Spalte *zentrum* der Wert 0 einzugeben. Entsprechend ist für die anderen Variablen wie folgt zu verfahren (siehe **Abb. 3.4**):

```

v1_ID → 0
v2_zuf → 2
v3_zuf → 4
v4_zuf → 1
v5_zuf → 1
v6_zuf → 3
v7_zuf → 1
v8_zuf → 2
v9_zuf → 3
v10_geschlecht → 0
v11_alter → 55
v12_größe → 1,64
v13_1_gew_vor → 85
v13_2_gew_nach → 80
v14_1_rauch_vor → 2
v14_2_rauch_nach → 1
v15_sport → 60
v16_freizeit → 9 (Kein Kreuz im Fragebogen, deshalb als „keine Angabe“ kodieren)
v17_angehörige → 1
v18_chron_Erkrankung → COPD

```

	zentrum	v1_ID	v2_zuf	v3_zuf	v4_zuf	v5_zuf	v6_zuf	v7_zuf	v8_zuf	v9_zuf	v10_geschlecht	v11_alter	v12_groesse	v13_1_gewicht	v13_2_gewicht	v14_1_rachen	v14_2_rachen	v15_sport
1	0	0	2	4	1	1	3	1	2	3	0	5	1,64	85	80	2	1	60
2																		
3																		
4																		
5																		
6																		
7																		
8																		
9																		
10																		
11																		
12																		
13																		
14																		
15																		
16																		
17																		
18																		
19																		
20																		
21																		
22																		
23																		
24																		
25																		
26																		
27																		
28																		

Abb. 3.4: Beispiel – Manuelle Dateneingabe

Um die Eingabe für den nächsten Fall (Patienten) fortzusetzen, ist in die zweite Zeile zu wechseln und wieder mit der Eingabe für die Variable *zentrum* zu beginnen.

### 3.5 Einlesen von Datendateien

Viele Daten, die zum Zweck der statistischen Auswertung im Gesundheitswesen genutzt werden können, stehen bereits als Datendatei zur Verfügung. Im Folgenden wird dargestellt, wie bestehende Dateien in SPSS eingelesen werden können. Dabei muss unterschieden werden, ob es sich um eine interne Datendatei im SPSS-Dateiformat oder um eine Datendatei im externen Datenformat (z. B. Excelfile) handelt.

#### 3.5.1 Interne Datendateien

Eine interne Datendatei von SPSS erkennt man daran, dass die Datei die **Endung .sav** aufweist. In diesem Fall kann man einfach im allgemeinen Menü auf *Datei* → *Öffnen* → *Daten* klicken, und wie man es von den Microsoft-Office-Programmen gewohnt ist, die entsprechende Datei auswählen und öffnen. In **Abbildung 3.5** ist dies beispielhaft für das Öffnen der Datei *Zentrum1.sav* dargestellt.

#### Beispieldateien

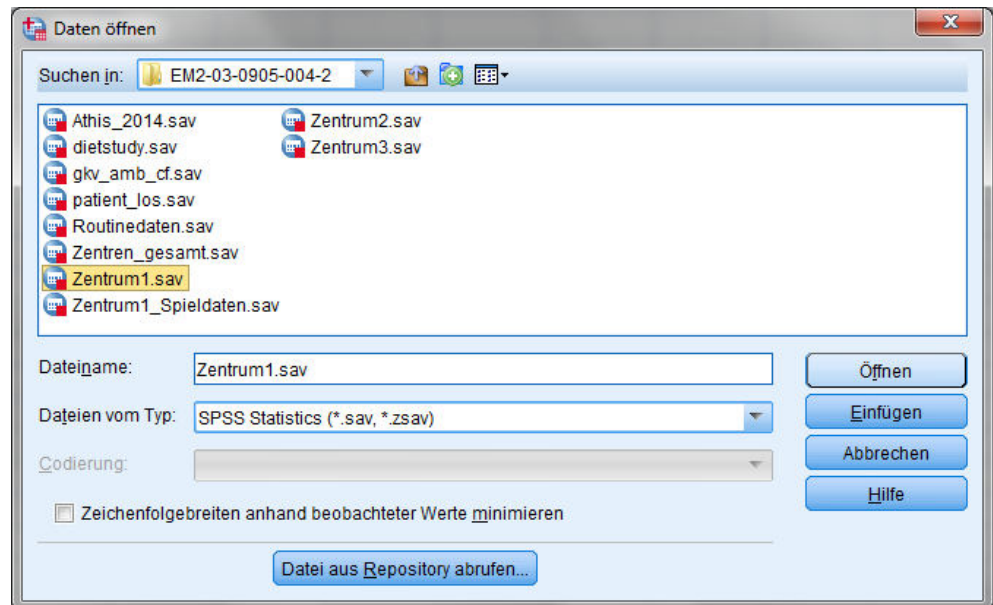


Abb. 3.5: Einlesen einer SPSS-Datendatei

#### 3.5.2 Externe Formate

SPSS kann eine Reihe von externen Datenformaten einlesen. Dabei sollte beachtet werden, dass die Variablenbezeichnungen mit den Notationen in SPSS kompatibel sind, um Probleme zu vermeiden. Zudem ist zu empfehlen, dass Datentabellen, die man einliest, so strukturiert sind, dass in den Zeilen die Fälle stehen und in den Spalten die Variablen.

##### Einlesen einer Excel-Datei

Das Einlesen einer externen Datendatei soll beispielhaft anhand der Datei *Routinedaten.xls* erfolgen. Dazu ist zunächst wieder im allgemeinen Menü auf *Datei* → *Öffnen* → *Daten* zu gehen. Damit die externe Datendatei angezeigt wird, ist in der Schaltfläche unten der entsprechende Dateityp, der eingelesen werden soll, auszuwählen (siehe Abb. 3.6). Anschließend ist mithilfe der Schaltfläche *Suchen in* der Ort auf dem Computer zu bestimmen, an dem die Datei *Routinedaten.xls* liegt und der Schalter *Öffnen* zu klicken. Daraufhin erscheint eine Dialogbox, in der man die Möglichkeit hat, auszuwählen, welches *Arbeitsblatt* der Exceldatei geöffnet werden soll. Außerdem ist voreingestellt, dass die erste Zeile der Datentabelle die Variablennamen beinhaltet. Sofern dies bei vorliegenden Daten nicht der Fall sein sollte, kann diese Voreinstellung auch deaktiviert werden, indem der Schalter *Variablennamen aus der ersten Dateizeile lesen* deaktiviert wird. Durch das Klicken auf den Schalter *OK* öffnet sich dann das Datenblatt im SPSS-Daten-Editor. Um die Datendatei im SPSS-Format abzuspeichern ist im Menü auf *Datei* → *Speichern unter* zu gehen und ein entsprechender Name für die Datei (z. B. „Routinedaten“) zu vergeben. Zudem ist auszuwählen, in welchem Unterverzeichnis die Datei abzulegen ist. Die Datei erhält von SPSS automatisch die entsprechende **Endung .sav**.

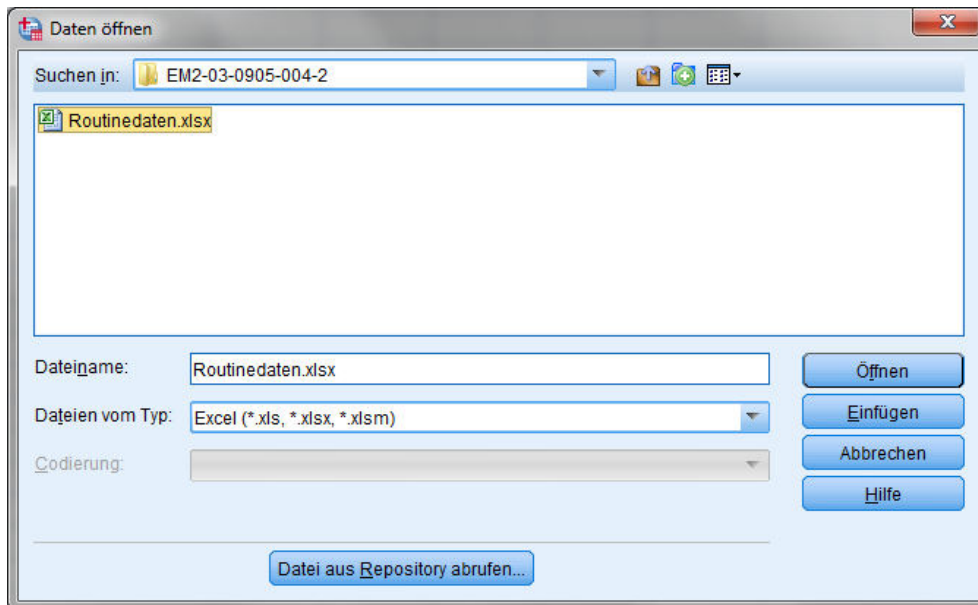


Abb. 3.6: Einlesen externer Daten (Daten öffnen)

### 3.6 Zusammenfügen von Datendateien

Wie in dem Anwendungsbeispiel (siehe Kapitel 1) kommt es in der praktischen Umsetzung einer Datenerhebung häufig vor, dass eine Befragung an unterschiedlichen Orten bzw. Einrichtungen stattfindet. Um die Daten aus den unterschiedlichen Zentren in SPSS einzugeben und zusammenzuführen, können die Papierversionen der Fragebögen entweder alle an ein Zentrum geschickt werden, wo dann die Dateneingabe zentral erfolgt. Alternativ können die Daten jeweils vor Ort in jedem einzelnen Zentrum in SPSS eingegeben werden und die Datendateien können dann z. B. per E-Mail an die auswertende Stelle geschickt werden. Im letzten Szenario steht man nun vor der Herausforderung die sav-Dateien „**Fallweise zusammenfügen**“. Darüber hinaus kann es die Situation geben, dass Variablen, die man für die spätere Analyse braucht, auf unterschiedliche Art und Weise erfasst wurden und somit in unterschiedlichen Datendateien zur Verfügung stehen. Damit alle Variablen später in einer Datendatei zur Verfügung stehen, müssen die Dateien „**variablenweise zusammengefügt**“ werden.

#### 3.6.1 Fallweise Zusammenfügen

Im Folgenden wird dargestellt, wie Datendateien zusammengeführt werden, die exakt die gleichen Variablen aufweisen, jedoch unterschiedliche Fälle beinhalten. Die praktische Umsetzung wird anhand der Befragungsdaten von Zentrum 1 (Datei: *Zentrum1.sav*) und Zentrum 2 (Datei: *Zentrum2.sav*) demonstriert. Um diese Dateien zusammenzuführen ist folgendermaßen zu verfahren:

Öffnen der Datei *Zentrum1.sav* mithilfe der Menüwahl: *Datei* → *Öffnen* → *Daten*. Zum Anfügen der Datei *Zentrum2.sav* ist aus dem Menü folgende Aktion zu wählen: *Daten* → *Dateien zusammenfügen* → *Fälle hinzufügen*. Daraufhin öffnet sich die Dialogbox *Fälle hinzufügen zu* (siehe Abb. 3.7). Durch Klicken auf den Schalter *eine externe SPSS Statistics-Datendatei* wird das Feld *Durchsuchen* aktiviert und die Datei *Zentrum2.sav* kann ausgewählt werden. Anschließend muss noch auf *Öffnen* und *Weiter* geklickt werden.

#### Beispieldatensätze

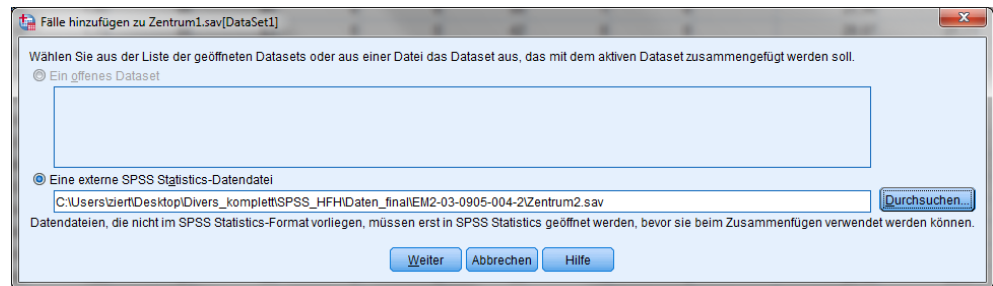


Abb. 3.7: Dialogbox – Fälle hinzufügen

Es öffnet sich eine weitere Dialogbox *Fälle hinzufügen aus* (siehe Abb. 3.8). Da sich die Variablen in den beiden Dateien exakt entsprechen, befinden sich alle Variablen im Feld *Variablen in neuer Arbeitsdatei*. Hingegen ist das Feld *Nicht gepaarte Variablen* leer. Damit SPSS eine Variable als identisch in beiden Dateien ansieht, ist es erforderlich, dass sich die Variablen 1:1 entsprechen – d.h. in Bezug auf alle Komponenten, die man in der Variablenansicht für eine Variable festlegen kann. Die Ausführung der Prozedur ist abschließend mit der Taste *OK* zu bestätigen.

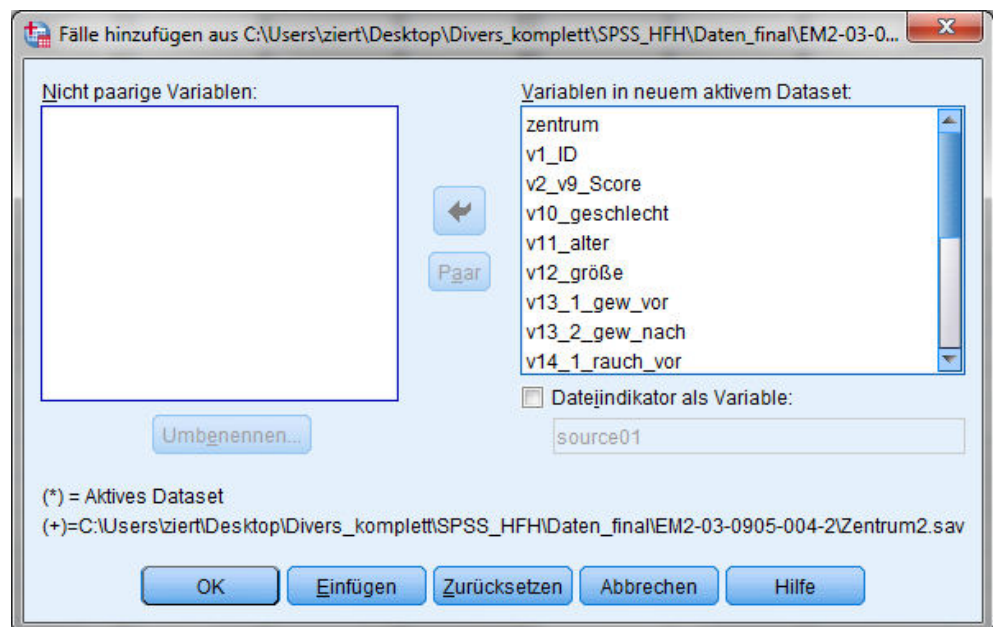


Abb. 3.8: Dialogbox – Fälle hinzufügen aus

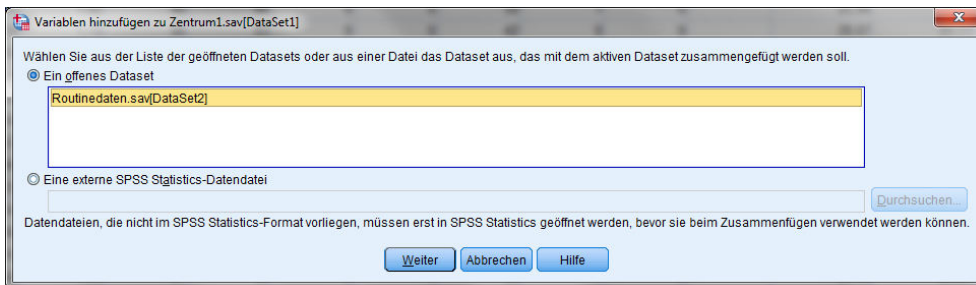
### 3.6.2 Variablenweise Zusammenfügen

#### Beispieldatensätze

Im Folgenden wird dargestellt, wie zwei SPSS-Datendateien zusammengefügt werden können, die identische Fälle beinhalten, jedoch unterschiedliche Variablen aufweisen. Für die praktische Umsetzung wird dies daran demonstriert, wie die Dateien *Zentren\_gesamt.sav* und *Routinedaten.sav* zusammengefügt werden. Dazu müssen beide Dateien zunächst in SPSS eingelesen werden (siehe Abschnitt 3.5).

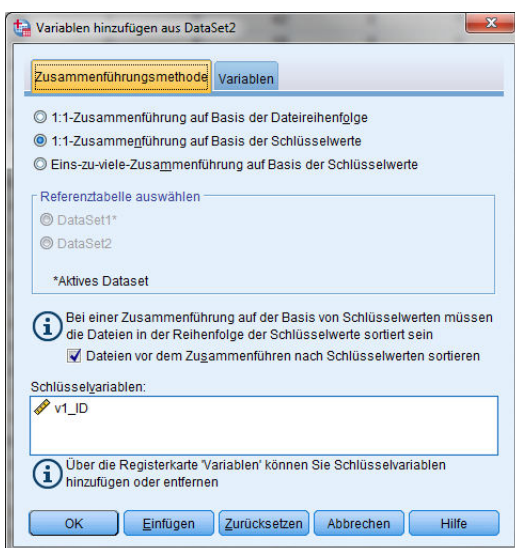
Für ein erfolgreiches „Variablenweise zusammenfügen“ von Datendateien wird empfohlen, dies über eine sog. **Schlüsselvariable** umzusetzen. Dabei handelt es sich um eine Variable, die die **eindeutige Identifikation eines Falls** in beiden Datendateien ermöglicht. In den Datendateien des Anwendungsbeispiels ist es die Variable *ID*. Bevor das Zusammenfügen erfolgen kann, müssen **beide** Datendateien aufstei-

gend nach der Schlüsselvariablen sortiert und abgespeichert werden (siehe Abschnitt 5.1). Anschließend ist in die Datei *Zentren\_gesamt.sav* zu wechseln und über das Menü die folgende Aktion auszuwählen: *Daten* → *Dateien zusammenfügen* → *Variablen hinzufügen*. Daraufhin öffnet sich die Dialogbox *Variablen hinzufügen zu*. Da die Datei *Zentren\_gesamt.sav* bereits geöffnet ist, kann diese ausgewählt werden, indem sie einfach angeklickt wird (siehe **Abb. 3.9**). Anschließend muss der Schalter *Weiter* betätigt werden.



**Abb. 3.9:** Dialogbox – Variablen hinzufügen

Es erscheint die Dialogbox *Variablen hinzufügen aus DataSet2* (siehe **Abb. 3.10**). Die voreingestellte Option *1:1-Zusammenführung auf Basis der Schlüsselwerte* unter der Registerkarte *Zusammenführungsmethode* ist für den Anwendungsfall die Methode der Wahl. Die Schlüsselvariable ist die Variable *v1\_ID*, da sie in beiden Datendateien auftaucht. SPSS hat diese Variable automatisch identifiziert, weshalb sie in dem Feld *Schlüsselvariablen* aufgeführt ist.



**Abb. 3.10:** Dialogbox – Variablen hinzufügen aus

Um weitere Schlüsselvariablen (hier nicht erforderlich) hinzuzufügen oder zu entfernen, sind diese über die Registerkarte auszuwählen. Abschließend ist die Prozedur mit *OK* zu bestätigen.

### 3.7 Datenprüfung

Im Anschluss an die manuelle Dateneingabe und/oder das Dateneinlesen bzw. das Zusammenfügen von Dateien sollte immer eine **Prüfung der Daten auf Plausibilität, Korrektheit und Vollständigkeit** erfolgen. Denn insbesondere bei der

### Prüfen der Daten



manuellen Dateneingabe können Eingabefehler auftreten. Aber auch beim Dateneinlesen sind Formatfehler möglich oder Fälle bzw. Variablen werden von einer Datendatei abgeschnitten. Ziel der Datenprüfung ist es, Fehler in den Daten aufzudecken und sie vor dem Beginn der umfassenden statistischen Analysen entsprechend zu korrigieren.

#### Beispieldatensatz

Die Datenprüfung soll beispielhaft anhand ausgewählter Variablen der Datendatei *Zentrum1\_Spieldaten.sav* durchgespielt werden. In dieser Version wurden bewusst Fehler eingebaut, die im Rahmen der Plausibilitätsprüfung identifiziert werden sollen.

#### Überblick verschaffen

Um sich schnell einen ersten Überblick über die Daten zu verschaffen (z. B. Häufigkeit fehlender Angaben), kann man einfach in der Datenansicht mithilfe der Pfeiltasten und/oder „Schieberegler“ am rechten Rand des Daten-Editor-Fensters innerhalb der Datentabelle hoch bzw. hinunter scrollen. Zudem ist es möglich, in der Datenansicht die Merkmalsausprägungen zu einer Variablen auf- bzw. absteigend zu sortieren (siehe Abschnitt 5.1). Dieses Vorgehen funktioniert insbesondere bei kleinen Datensätzen ( $n < 100$ ). Bei größeren Datensätzen ist dieses Vorgehen nicht zu empfehlen, da das Hin- und Herscrollen bzw. das Sortieren sehr viel Kapazität des Arbeitsspeichers fordert und deshalb das Ergebnis z. T. sehr verzögert angezeigt wird. Zudem ist dieses Verfahren nicht dazu geeignet, systematische Eingabefehler zu identifizieren, da diese bei dieser Methode schnell übersehen werden können.

#### 3.7.1 Prüfung der Daten gegen den Urbeleg

Eine sehr aufwendige Methode, um die Daten auf Eingabefehler zu prüfen, besteht darin, jeden einzelnen Wert in den Daten gegen den Urbeleg (z. B. Fragebogen) zu prüfen (vgl. Bühl, 2010). Um eine vollständige Liste der eingegebenen Daten auf Fallebene im Viewer zu erhalten, ist folgende Menüauswahl zu treffen *Analysieren* → *Berichte* → *Fälle zusammenfassen*. Daraufhin erscheint die Dialogbox *Fälle zusammenfassen* (siehe Abb. 3.11).

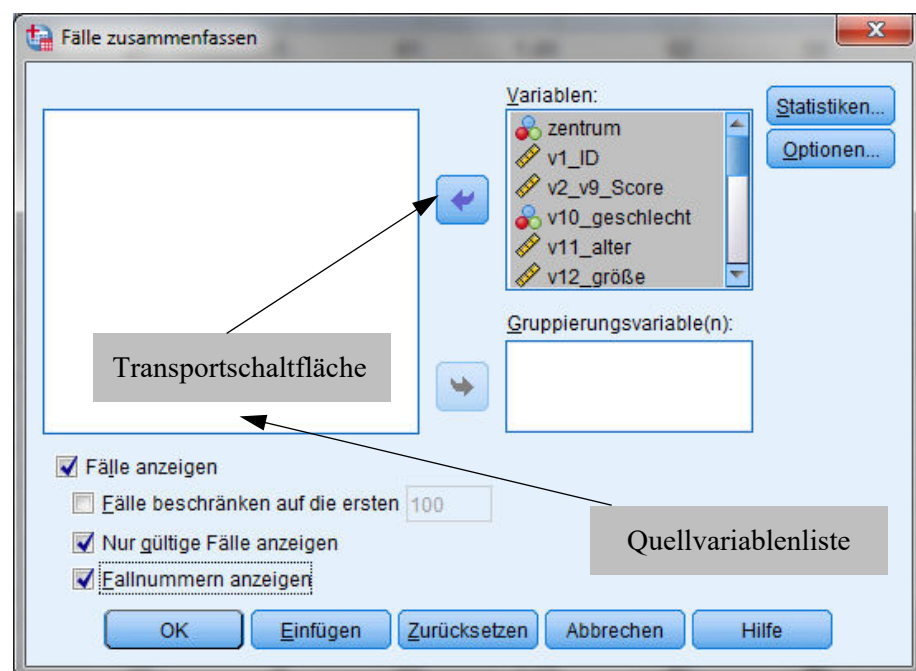


Abb. 3.11: Dialogbox – Fälle zusammenfassen



In der Dialogbox *Fälle zusammenfassen* ist es möglich, die Variablen, für die man eine Auflistung der Merkmalsausprägungen im Viewer angezeigt haben möchte, aus der Quellvariablenliste mit der Transportschaltfläche in das Fenster *Variablen* zu bringen. Um sich die Merkmalsausprägungen für jeden Fall anzeigen zu lassen, ist die Checkbox *Fälle anzeigen* zu aktivieren. Zudem ist es möglich, weitere Spezifizierungen vorzunehmen, wie *Fälle beschränken auf die ersten [100]*, *nur gültige Fälle anzeigen* oder *Fallnummern anzeigen*. Nachdem alle Einstellungen getroffen wurden, ist der Schalter *OK* zu aktivieren. Ein kleiner Nachteil der dann erzeugten Tabelle besteht darin, dass systemdefiniert fehlende Werte (d. h. keine Angabe in einer Zelle) mit einer 0 in der Tabelle angezeigt werden, sodass es nicht möglich ist, eine korrekt eingegebene 0 von einem fehlenden Wert zu unterscheiden.

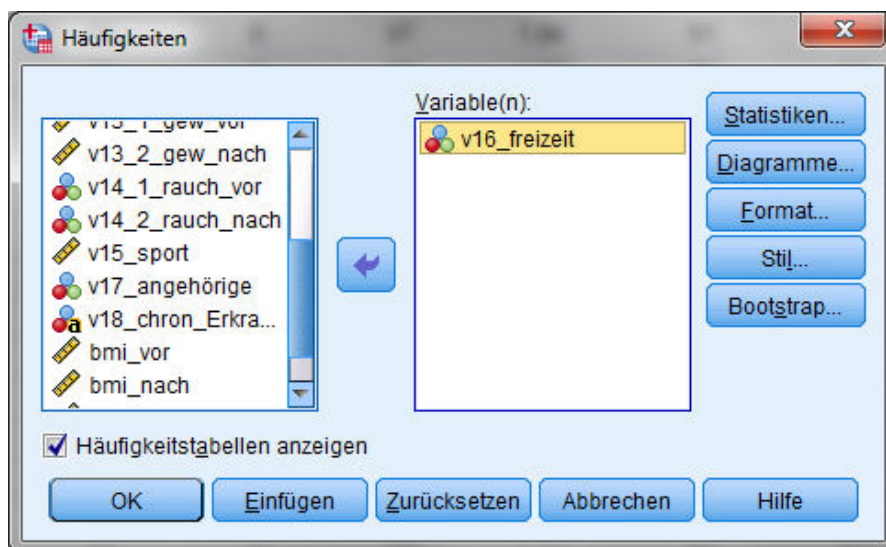
Da der Abgleich der eingegebenen Daten mit dem Urbeleg sehr zeitaufwendig ist, wird man dieses Vorgehen gerade im Falle von sehr großen Datendateien nur in Ausnahmefällen bzw. für eine Stichprobe von Fällen (siehe Abschnitt 5.2.3) durchführen können. Für das Aufdecken von unplausiblen Werten in den Daten und Eingabefehlern stehen in SPSS jedoch noch weitere Verfahren zur Verfügung, die in den beiden folgenden Kapiteln vorgestellt werden.

**Vorsicht: systemdefinierte fehlende Werte**

### 3.7.2 Häufigkeitsauszählungen

Das Erstellen von sog. Häufigkeitstabellen zum Zweck der Häufigkeitsauszählungen von Merkmalsausprägungen ist insbesondere bei kategorialen Variablen (nominales oder ordinales Messniveau) zu empfehlen. Ziel dieser Auszählung ist die Prüfung, ob die Anzahl der Angaben pro Merkmalsausprägung plausibel ist und ob Merkmalsausprägungen eingegeben wurden, die laut Kodeplan gar nicht vorkommen. Die Erstellung von Häufigkeitstabellen soll anhand der Variable *v16\_freizeit* demonstriert werden. Dazu ist zunächst die folgende Menüauswahl zu treffen: *Analysieren* → *Deskriptive Statistiken* → *Häufigkeiten*. Daraufhin öffnet sich die Dialogbox *Häufigkeiten*. Dort ist aus der Quellvariablenliste die Variable *v16\_freizeit* mithilfe der Transportschaltfläche in das Fenster *Variable(n)* zu bringen. Zudem ist darauf zu achten, dass die Checkbox *Häufigkeitstabellen anzeigen* aktiviert ist. Abschließend ist die Prozedur mit *OK* zu bestätigen, damit im Viewer die Häufigkeitstabelle angezeigt wird (siehe *Abb. 3.12*).

**Variablen mit Nominalniveau oder Ordinalniveau**



**Erkennen von Eingabefehlern**

Abb. 3.12: Dialogbox – Häufigkeiten

#### Interpretation

Die erstellte Häufigkeitstabelle (siehe **Abb. 3.13**) zeigt in der Spalte *Häufigkeit* an, wie viele Personen Angaben zu den jeweiligen Merkmalsausprägungen gemacht haben.

**Anzahl wahrgenommener Freizeitangebote**

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	bis 2	12	40,0	46,2	46,2
	3-6	10	33,3	38,5	84,6
	>6	2	6,7	7,7	92,3
	5	1	3,3	3,8	96,2
	7	1	3,3	3,8	100,0
	Gesamt	26	86,7	100,0	
Fehlend	keine Angabe	2	6,7		
	System	2	6,7		
	Gesamt	4	13,3		
Gesamt		30	100,0		

**Abb. 3.13:** Häufigkeitstabelle für die Variable *v16\_freizeit*

So haben insgesamt 26 von 30 Personen überhaupt Angaben bei dieser Variablen gemacht, wobei 12 Personen angegeben haben bis zu 2 Freizeitangebote wahrgenommen zu haben. 10 Personen haben angegeben, 3–6 Freizeitangebote wahrgenommen zu haben und 2 Personen haben angegeben, mehr als 6 Freizeitangebote in Anspruch genommen zu haben. Zudem liegt für eine Person die Ausprägung 5 und für eine andere Person die Ausprägung 7 vor. Da die letzten beiden Merkmalsausprägung laut Kodeplan keine möglichen Ausprägungen sind, handelt es sich hierbei höchstwahrscheinlich um Eingabefehler. Für diese Fälle sollte eine Prüfung der Angaben gegen den Urbeleg (Fragebogen) erfolgen. Darüber hinaus ist der Tabelle zu entnehmen, dass insgesamt 4 Personen gar keine Angabe zu dieser Frage gemacht haben, wobei es sich bei 2 Angaben um systemdefinierte fehlende Werte handelt und bei 2 Angaben um benutzerdefinierte fehlende Werte (keine Angabe). Hier ist ebenfalls kritisch zu prüfen, ob die systemdefinierten fehlenden Werte tatsächlich in den Fragebögen vorliegen oder nicht. Falls es sich tatsächlich um fehlende Angaben handelt, sollte hier auch der Wert für die benutzerdefinierten fehlenden Werte verwendet werden. Die Interpretation der übrigen Spalten wird in Studienbrief 2 dieses Moduls („Deskriptive Analysen mit SPSS“) erläutert (siehe Kapitel 2).

#### 3.7.3 Berechnung von Kennwerten

##### Variablen mit Intervallniveau oder Verhältnisniveau

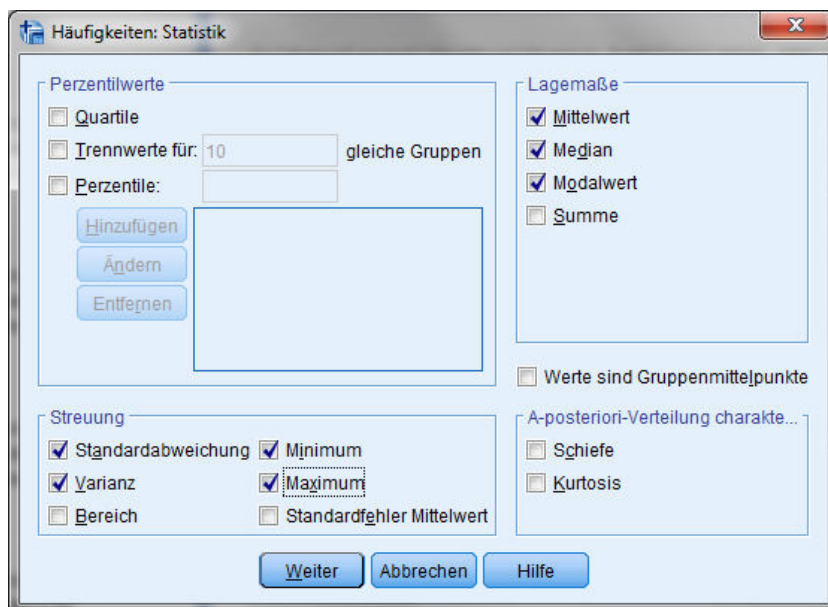
Um Variablen mit Intervallniveau oder Verhältnisniveau auf Plausibilität zu prüfen, ist es zielführender sich bestimmte Kennwerte berechnen zu lassen, da häufig beliebig viele unterschiedliche Werte vorkommen, sodass Häufigkeitstabellen gerade bei großen Datensätzen sehr groß und unübersichtlich werden. Folgende Kennwerte bieten sich an:

##### Kennwerte

- Minimum
- Maximum
- Modus
- Median
- Mittelwert

Die Berechnung dieser Kennwerte soll anhand der Variablen *v13\_1\_gew\_vor* (Körpergewicht vor Reha) und *v13\_2\_gew\_nach* (Körpergewicht nach Reha) präsentiert werden. Die korrekte Interpretation dieser Kennwerte wird genauer im Studienbrief „Deskriptive Analysen mit SPSS“ dieses Moduls erläutert.

Für die Berechnung der Kennwerte ist folgende Menüauswahl zu treffen: *Analysieren* → *Deskriptive Statistiken* → *Häufigkeiten*. Daraufhin öffnet sich die Dialogbox *Häufigkeiten*. Dort sind aus der Quellvariablenliste die Variablen *v13\_1\_gew\_vor* und *v13\_2\_gew\_nach* mithilfe der Transportschaltfläche in das Fenster *Variable(n)* zu bringen. Zudem ist darauf zu achten, dass die Checkbox *Häufigkeitstabellen anzeigen* **deaktiviert** ist. Anschließend ist auf den Schalter *Statistiken* zu klicken, sodass sich die Dialogbox *Häufigkeiten: Statistik* öffnet (siehe **Abb. 3.14**). Dort sind die Checkboxen für die Kennwerte *Minimum*, *Maximum*, *Modus*, *Median* und *Mittelwert* zu aktivieren. Die Auswahl in der Dialogbox ist mit *Weiter* zu bestätigen, sodass man wieder in die Dialogbox *Häufigkeiten* gelangt. Die Prozedur ist mit *OK* zu bestätigen, damit im Viewer die Kennwerte angezeigt werden.



**Abb. 3.14:** Dialogbox – Häufigkeiten: Statistik

Die Tabelle zu den Kennwerten im Viewer ist folgendermaßen zu lesen (siehe **Abb. 3.15**): Insgesamt haben bei der Variablen *v13\_1\_gew\_vor* von den 30 befragten Personen 27 Personen gültige Angaben gemacht. Das Minimum des Gewichts vor der Reha liegt bei 61 und das Maximum bei 700. Da eine Gewichtsangabe von 700 unplausibel ist, sollte der Fall mit der Gewichtsangabe von 700 in den Daten identifiziert und gegen den Urbeleg (Fragebogen) geprüft und korrigiert werden. Bei der Variablen *v13\_2\_gew\_nach* haben alle 30 befragten Personen gültige Angaben gemacht. Das Minimum zur Gewichtsangabe liegt bei 58 und das Maximum bei 107. Beide Werte erscheinen plausibel.

### Interpretation und Erkennen von Eingabefehlern

**Statistiken**

		Körpergewicht vor Reha	Körpergewicht nach Reha
N	Gültig	27	30
	Fehlend	3	0
Mittelwert		106,41	81,53
Median		84,00	80,50
Modus		78 <sup>a</sup>	79
Std.-Abweichung		119,175	11,322
Varianz		14202,712	128,189
Minimum		61	58
Maximum		700	107

a. Mehrere Modi vorhanden. Der kleinste Wert wird angezeigt.

**Abb. 3.15:** Kennwerte für die Variablen *v13\_1\_gew\_vor* und *v13\_2\_gew\_nach*

Um Fehler in den Daten zu identifizieren, lohnt es sich zudem einen genaueren Blick auf die Kennwerte Modus, Median und Mittelwert zu werfen. Diese Werte liegen dann nahe beieinander, wenn die betrachtete Variable annähernd normalverteilt ist. Da bei Gewichtsangaben davon ausgegangen werden kann, kann dies eine zusätzlich sinnvolle Prüfung in Bezug auf Eingabefehler sein. So legen die Werte bei der Variablen *v13\_2\_gew\_nach* alle relativ eng zusammen, was für plausible Werte spricht. Im Gegensatz dazu weichen die Werte bei der Variablen *v13\_1\_gew\_vor* stark voneinander ab (insbesondere Median und Mittelwert). Dies liegt daran, dass der Mittelwert stark von dem Ausreißer (Maximum: 700) verzerrt wird.

Zum Zweck der Datenprüfung von Variablen mit Intervallniveau oder Verhältnissniveau können neben der Berechnung von Kennwerten auch grafische Darstellungsmöglichkeiten, wie Histogramme und Boxplots genutzt werden. Mithilfe von Grafiken ist es möglich, sich schnell einen Überblick über die Verteilung der Daten zu verschaffen und dahingehend zu prüfen, ob diese plausibel ist oder ob es z. B. extreme Ausreißer gibt. Die Möglichkeiten der grafischen Veranschaulichung werden in Studienbrief 2 dieses Moduls ausführlich dargestellt.

### Übungsaufgaben

- 3.1) Welchen Zweck erfüllt ein Kodeplan?
- 3.2) Fügen Sie auf Basis der Dateien *Zentrum2.sav* und *Zentrum3.sav* die Befragungsdaten von Zentrum 2 und Zentrum 3 zusammen.
- 3.3) Führen Sie auf Basis der Datei *Zentrum1\_Spieldaten.sav* eine Datenprüfung (Häufigkeitsauszählung bzw. Berechnung von Kennwerten) für die Variablen *v14\_1\_rauch\_vor* und *v12\_größe* durch.

## 4 Modifikation von Variablen

Nachdem die zu analysierenden Daten erfolgreich kodiert, eingepflegt bzw. eingelesen und eingehend geprüft worden sind, sind häufig weitere Datenmodifikationen, wie das Umcodieren von Variablen (Abschnitt 4.1) oder die Berechnung neuer Variablen (Abschnitt 4.2) erforderlich, um die statistischen Analysen zielgerichtet durchführen zu können.

### 4.1 Umcodieren von Variablen

In SPSS ist es möglich, einmal erfasste Variablen umzucodieren.

Unter **Umcodieren** wird verstanden, dass bestehende Variablenausprägungen nach vorab definierten Regeln verändert werden.

#### Definition

Umcodieren könnte beispielsweise notwendig sein, wenn eine Variablenausprägung in der statistischen Auswertung nicht in dem hohen Detaillierungsgrad gebraucht wird, wie sie primär erfasst worden ist. So werden häufig aus Altersangaben (in Jahren) Altersklassen gebildet. Ein solcher Schritt bedeutet immer eine Informationsreduktion und in dem Anwendungsbeispiel auch eine Veränderung des Messniveaus – von Intervall- bzw. Verhältnissniveau zu Ordinalniveau.

#### Gründe für Umcodieren

Eine andere Notwendigkeit bestimmte Variablen umzukodieren umzucodieren, besteht dann, wenn Einzelitems z. B. für die Berechnung einer psychometrischen Summenskala (Score) genutzt werden und diese auf der Ebene der Einzelitems nicht gleichgerichtet kodiert sind, so wie es bei der Summenskala zur Patientenzufriedenheit (ZUF-8) der Fall ist. Die Skala ist so konstruiert, dass hohe Werte bei dem Summenscore für eine hohe Patientenzufriedenheit stehen sollen. Damit alle Items gleichgerichtet sind und eine hohe Ausprägung für eine hohe Patientenzufriedenheit steht, müssen die folgenden Variablen umcodiert werden: *v2\_zuf*, *v4\_zuf*, *v7\_zuf*, *v8\_zuf*.

#### Umcodieren von Items einer Skala

Die Bildung von Altersklassen, das Umcodieren von Einzelitems der Skala zur Patientenzufriedenheit sowie die Besonderheiten beim Umcodieren im Falle von spezifischen Formaten werden in den folgenden Abschnitten dargestellt.

#### Bildung von Alterskategorien

Grundsätzlich bietet SPSS die Möglichkeit, eine bestehende Variable in die gleiche Variable (siehe Abschnitt 4.1.1) oder in eine neue Variable (siehe Abschnitt 4.1.2) umzucodieren. Da ein Umcodieren in die gleiche Variable zur Konsequenz hat, dass die ursprüngliche Variable überschrieben wird, wird in der Regel empfohlen, in eine neue Variable umzucodieren. Aufgrund der Vollständigkeit werden hier aber beide Möglichkeiten dargestellt.

Die folgenden Prozeduren werden anhand der Datendatei *Zentrum1\_Spieldaten.sav* demonstriert.

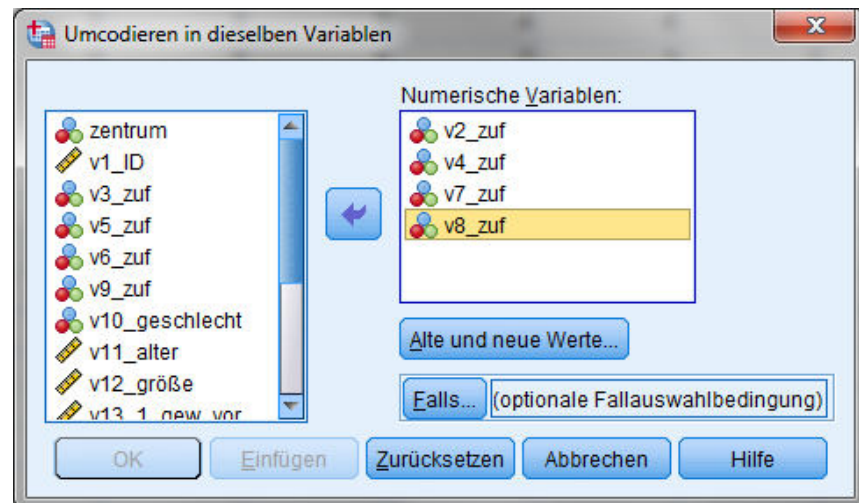
#### Beispieldatensatz

#### 4.1.1 In dieselbe Variable

Die SPSS-Prozedur „Umcodieren in dieselbe Variable“ soll anhand des Umcodierens der Einzelitems *v2\_zuf*, *v4\_zuf*, *v7\_zuf*, *v8\_zuf* der Skala zur Patientenzufriedenheit demonstriert werden. Um am Ende eine aussagekräftige Summenskala zur

Patientenzufriedenheit bilden zu können (hohe Werte = hohe Patientenzufriedenheit), sollen die ursprünglichen Variablenausprägungen folgendermaßen umcodiert werden:  $1 \rightarrow 4$ ,  $2 \rightarrow 3$ ,  $3 \rightarrow 2$  und  $4 \rightarrow 1$ .

Da die ursprünglichen Variablen zur Bildung der Summenskala nicht mehr benötigt werden, können sie mit den veränderten Werten überschrieben werden. In SPSS ist dies wie folgt umzusetzen: Aus dem Menü ist die Prozedur *Transformieren*  $\rightarrow$  *Umcodieren*  $\rightarrow$  *Umcodieren in die dieselben Variablen* auszuwählen. Es öffnet sich die Dialogbox *Umcodieren in dieselben Variablen* (siehe **Abb. 4.1**).



**Abb. 4.1:** Dialogbox – Umcodieren in dieselben Variablen

Die Quellvariablenliste enthält alle numerischen Variablen der Datendatei. Es ist möglich, eine oder mehrere Variablen gleichzeitig auszuwählen. Bei der Wahl von mehreren Variablen müssen diese vom gleichen Typ sein. Da dies bei den Items *v2\_zuf*, *v4\_zuf*, *v7\_zuf*, *v8\_zuf* der Skala zur Patientenzufriedenheit der Fall ist, können alle Items mithilfe der Transportschaltfläche in das Feld *Numerische Variablen* übertragen werden. Anschließend ist für die Festlegung der umzucodierenden Werte auf die Schaltfläche *Alte und neue Werte* zu klicken, wodurch sich die Dialogbox *Umcodieren in dieselben Variablen: Alte und neue Werte* (siehe **Abb. 4.2**) öffnet.

Für ein erfolgreiches Umcodieren ist es erforderlich, dass für jeden zu recodierenden Wert ein Wert bzw. Wertebereich für die Eingabevariable (*Alter Wert*) sowie ein neuer Wert für die Ausgabevariable (*Neuer Wert*) festzulegen ist. Im Detail gliedert sich die Dialogbox in die beiden Eingabekästen *Alter Wert* und *Neuer Wert*. In dem Eingabekasten *Alter Wert* ist es möglich, zwischen den folgenden Alternativen zu wählen (Bühl, 2010, S. 220):

### alte und neue Werte

- *Wert:* Hier besteht die Option, einen einzelnen Wert als Eingabewert einzugeben.
- *Systemdefiniert fehlend:* Hier besteht die Option, systemdefinierte fehlende Werte als Eingabewert festzulegen. In der Werteliste erscheint er als SYSMIS.
- *System- oder benutzerdefinierte fehlende Werte:* Hier besteht die Option, system- oder benutzerdefinierte fehlende Werte als Eingabewert festzulegen. In der Werteliste erscheinen benutzerdefinierte fehlende Werte als MISSING.
- *Bereich bis:* Hier ist es möglich, einen abgeschlossenen Wertebereich als Eingabe festzulegen.
- *Bereich, KLEINSTER bis Wert:* Hier ist es möglich, alle Werte vom kleinsten beobachteten Wert bis zum angegebenen Wert als Eingabe festzulegen.

- *Bereich:* Wert bis GRÖSSTER: Hier ist es möglich, alle Werte vom angegebenen Wert bis zum größten beobachteten Wert als Eingabe festzulegen.
- *Alle anderen Werte:* Diese Option steht für alle Werte zur Verfügung, die nicht im Rahmen der anderen Optionen angegeben wurden. In der Werteliste erscheint der Wert mit ELSE.

In dem Eingabekasten *Neuer Wert* kann zwischen den folgenden Alternativen gewählt werden:

- *Wert:* Hier besteht die Option, einen einzelnen Wert als Ausgabewert einzugeben.
- *Systemdefiniert fehlend:* Hier besteht die Option, systemdefinierte fehlende Werte als Ausgabewert festzulegen. In der Werteliste erscheint er als SYSMIS.
- *Alte Werte kopieren:* Hier kann festgelegt werden, dass die Werte der Eingabewariablen beibehalten werden.

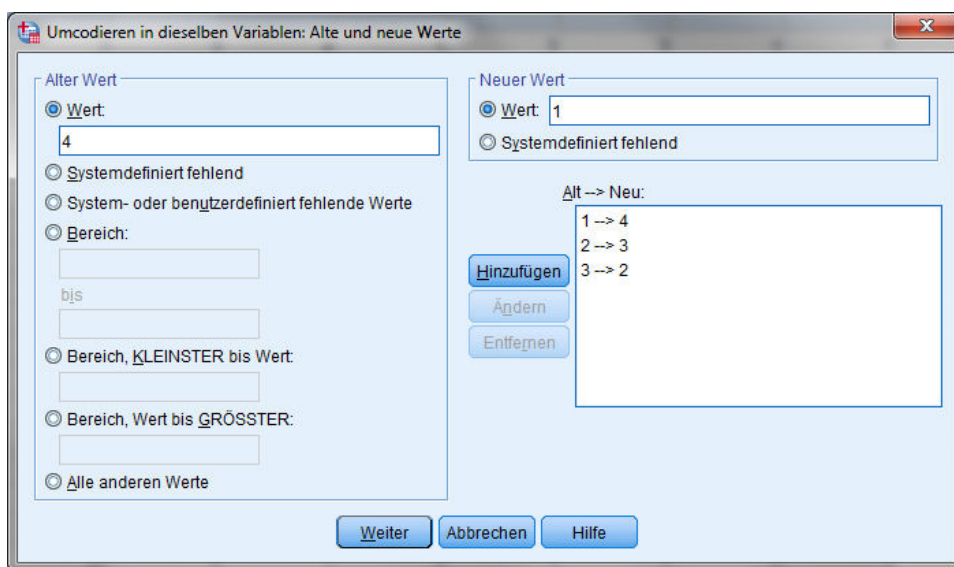


Abb. 4.2: Dialogbox – Umcodieren in dieselben Variablen: Alte und neue Werte

Um nun die entsprechenden Umcodierungen für die Items 1, 3, 6 und 7 vorzunehmen, müssen nacheinander die folgenden Werte in das Editierfeld *Wert* des Eingabekastens *Alter Wert* bzw. im Editierfeld *Wert* des Eingabekastens *Neuer Wert* eingegeben werden:  $1 \rightarrow 4$ ,  $2 \rightarrow 3$ ,  $3 \rightarrow 2$  und  $4 \rightarrow 1$ .

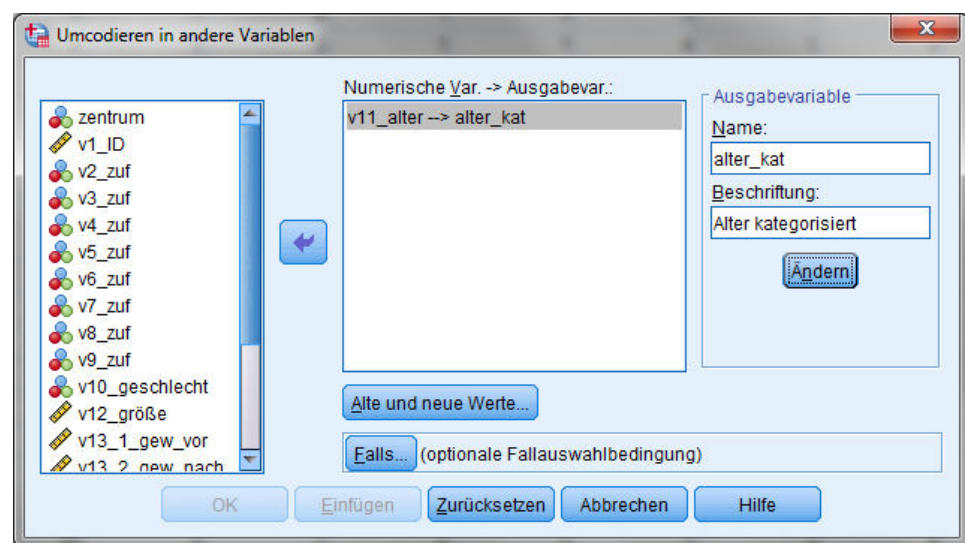
Die Recodierung eines einzelnen Wertes ist jeweils mit dem Schalter *Hinzufügen* zu bestätigen, sodass die Syntax für die Recodierung im Kasten *Alt → Neu* erscheint. Abschließend sind die Schalter *Weiter* (in der Dialogbox *Alte und Neue Werte*) und *OK* (in der Dialogbox *Umcodieren in dieselben Variablen*) zu aktivieren.



### 4.1.2 In andere Variable

Die Prozedur „Umcodieren in andere Variable“ soll anhand des folgenden Beispiels veranschaulicht werden: In der Datendatei *Zentrum1\_Spieldaten.sav* liegt die Variable *v11\_alter* als Variablen mit Verhältnisniveau vor, d.h. für jeden Patienten ist das Alter in Jahren angegeben. Für einige der weiterführenden Auswertungen wird die Variable nicht in diesem Detaillierungsgrad gebraucht. Stattdessen sollen die folgenden Alterskategorien <50, 51–55, 56+ gebildet und bezüglich unterschiedlicher Parameter verglichen werden. Um die ursprüngliche Variable nicht zu überschreiben, soll eine neue Variable gebildet werden. In SPSS ist dies wie folgt umzusetzen:

Aus dem Menü ist die Prozedur *Transformieren* → *Umcodieren* → *Umcodieren in andere Variablen* auszuwählen. Es öffnet sich die Dialogbox *Umcodieren in andere Variablen* (siehe **Abb. 4.3**).



**Abb. 4.3:** Dialogbox – Umcodieren in andere Variablen

Die Quellvariablenliste enthält alle Variablen der Datendatei. Es ist möglich, eine oder mehrere Variablen gleichzeitig auszuwählen. Bei der Wahl von mehreren Variablen müssen diese vom gleichen Typ sein. Für das Umcodieren der Variable *v11\_alter* muss diese Variable mithilfe der Transportschaltfläche in das Editierfeld *Numerische Var. -> Ausgabevar.*: gebracht werden. Ein Fragezeichen zeigt an, dass ein Name für die neue Ausgabevariable definiert werden muss. Der neue Name *alter\_kat* ist in das Editierfeld *Name* einzugeben. Dadurch wird das Fragezeichen im Eingabefeld *Numerische Var. -> Ausgabevar.*: durch *alter\_kat* ersetzt. Zusätzlich ist es zur näheren Bezeichnung der Variablen möglich, eine Eingabe im Feld *Beschriftung* zu tätigen – hier ist „Alter kategorisiert“ einzugeben. Die Eingabe des Namens und der Beschriftung ist mit *Ändern* zu bestätigen.

Für die Festlegung der umzucodierenden Werte ist auf die Schaltfläche *Alte und neue Werte* zu klicken, sodass sich die Dialogbox *Umcodieren in andere Variablen: Alte und neue Werte* öffnet (siehe **Abb. 4.4**). Diese ist genauso aufgebaut, wie die Dialogbox *Umcodieren in dieselben Variablen: Alte und neue Werte* (Erläuterungen zu den Optionen siehe Abschnitt 4.1.1). Zusätzlich gibt es die folgenden Schaltflächen:

#### alte und neue Werte

- *Alte Werte kopieren*: Die alten Werte können mit dieser Option komplett in die neue Variable kopiert werden.



- *Ausgabe der Variablen als String*: Es ist möglich, sich die neue Variable als String ausgeben zu lassen.
- *Num. Strings in Zahlen umwandeln ('5' > 5)*: Die Option macht es möglich, String-Variablen, die ausschließlich Werte beinhalten in numerische Variablen umzucodieren.

Zum Umcodieren der Variablen *v11\_alter* ist nun folgendermaßen zu verfahren:

- Im Eingabekasten *Alter Wert* ist die Option *Bereich KLEINSTER Wert bis*: anzuklicken und in das Eingabefeld ist der Wert 50 einzugeben. Anschließend ist im Eingabekasten *Neuer Wert* der Wert 1 einzugeben. Die Angaben sind mit dem Schalter *Hinzufügen* zu bestätigen.
- Im Eingabekasten *Alter Wert* ist die Option *Bereich*: anzuklicken und im ersten Eingabefeld ist der Wert 51 einzugeben und im zweiten Eingabefeld der Wert 55. Anschließend ist im Eingabekasten *Neuer Wert* der Wert 2 einzugeben. Die Angaben sind mit dem Schalter *Hinzufügen* zu bestätigen.
- Im Eingabekasten *Alter Wert* ist die Option *Bereich Wert bis GRÖSSTER*: anzuklicken und in das Eingabefeld ist der Wert 56 einzugeben. Anschließend ist im Eingabekasten *Neuer Wert* der Wert 3 einzugeben. Die Angaben sind mit dem Schalter *Hinzufügen* zu bestätigen.
- Da weder system- noch benutzerdefinierte fehlende Werte auftreten, ist der Umcodierungsbefehl damit abgeschlossen.

Abschließend sind die Schalter *Weiter* (in der Dialogbox *Alte und neue Werte*) und *OK* (in der Dialogbox *Umcodieren in dieselbe Variable*) zu aktivieren.

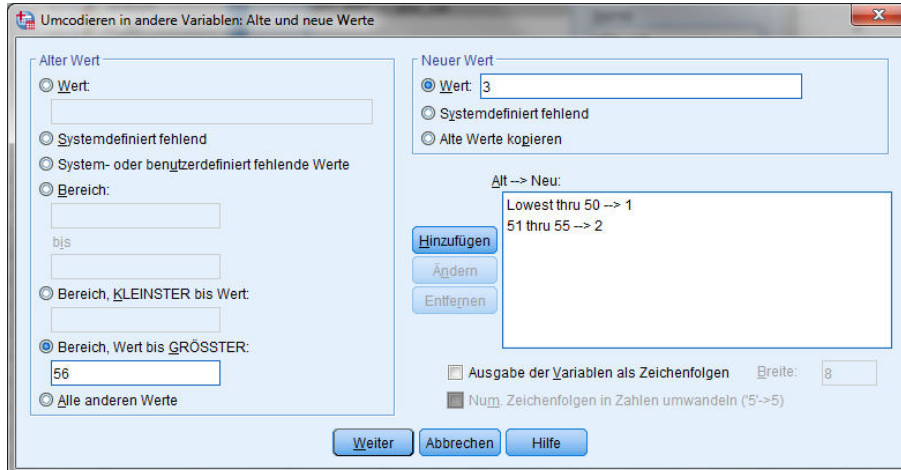


Abb. 4.4: Dialogbox – Umcodieren in andere Variablen: Alte und neue Werte

Um zu kontrollieren, ob das Umcodieren erfolgreich war, wird empfohlen zunächst einen Blick in die Variablenansicht zu werfen, ob die Variablenliste um die Variable *alter\_kat* ergänzt wurde. Sofern dies der Fall ist, besteht nun die Möglichkeit den Variablenausprägungen Wertelabels zuzuweisen (siehe Abschnitt 3.3.3). Zur Überprüfung, ob tatsächlich auch Kategorien erstellt worden sind, kann durch Doppelklick auf die Variable *alter\_kat* in der Variablenansicht zur Datenansicht gewechselt werden. Zur weiteren Kontrolle ist es zudem möglich, sich eine Häufigkeitstabelle zu erzeugen (siehe Abschnitt 3.7.2).

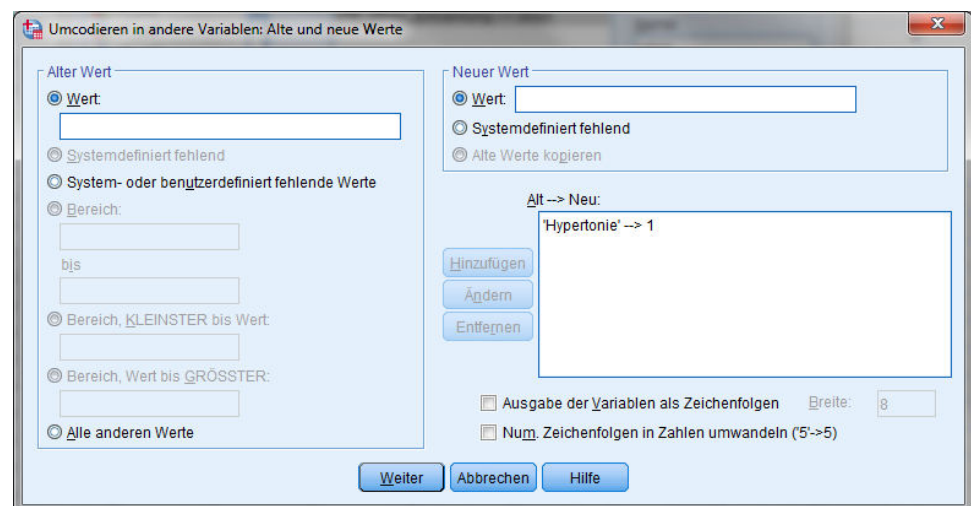
## Kontrolle

### Hinweis

Um zu überprüfen, ob das Umcodieren erfolgreich war, sollte man sich **immer** eine **Häufigkeitstabelle für die neu erstellte Variable** ausgeben lassen und diese mit der Häufigkeitstabelle für die Ursprungsvariable vergleichen.

### 4.1.3 Umcodieren von String-Variablen

Das Umcodieren von Variablen, die als **String-Variable** definiert sind und in der Regel Textangaben beinhalten, ist ebenfalls über die Menüauswahl *Transformieren* → *Umcodieren* → *Umcodieren in die dieselben Variablen* bzw. *Transformieren* → *Umcodieren* → *Umcodieren in andere Variablen* und der Bedienung der entsprechenden Dialogboxen möglich. Das heißt, wenn in der Datendatei *Zentrum1\_Spiel daten.sav* z. B. alle Patienten mittels einer neuen numerischen Variable identifiziert werden sollen, die bei der Variablen *v18\_chron\_Erkrankung* Hypertonie angegeben haben, können hierfür genau dieselben Felder (*Alter Wert* → *Hypertonie* und *Neuer Wert* → *1*) der Dialogbox genutzt werden (siehe **Abb. 4.5**):



**Abb. 4.5:** Dialogbox – Umcodieren von String-Variablen

Abschließend sind die Schalter *Weiter* (in der Dialogbox *Alte und Neue Werte*) und *OK* (in der Dialogbox *Umcodieren in dieselbe Variable*) zu aktivieren. Beim Umcodieren von String-Variablen ist darauf zu achten, dass die Begriffe, die umcodiert werden sollen, exakt so geschrieben werden, wie sie eingegeben wurden (inklusive Groß- und Kleinschreibung), damit SPSS diese korrekt erkennt.

### Der Schalter „Falls“

Sowohl bei der Prozedur *Umcodieren in dieselbe Variable* als auch bei der Prozedur *Umcodieren in andere Variable*, besteht die Möglichkeit, über den Schalter *Falls* die Modifikation von Variablen nur für bestimmte Fälle vorzunehmen. Indem der Schalter *Falls* geklickt wird, öffnet sich die Dialogbox *Fälle auswählen*. Dort ist voreingestellt, dass die ausgewählte Prozedur auf alle Fälle *anzuwenden ist* (*Alle Fälle einschließen*). Dies kann geändert werden, indem die Checkbox *Fall einschließen, wenn Bedingung erfüllt ist* aktiviert wird. Für die Formulierung der Bedingungen stehen dem Anwender die gleichen Operatoren und Funktionen wie in der Dialogbox *Variablen berechnen* zur Verfügung. Wie mit dieser Dialogbox im Detail zu arbeiten ist und wie die Operatoren und Funktionen anzuwenden sind, wird detailliert im folgenden Kapitel erläutert.

### Exkurs

## 4.2 Berechnung neuer Variablen

Basierend auf bestehenden numerischen Variablen können zudem neue Variablen berechnet werden. Diese Funktionalität in SPSS kann im Rahmen unseres Anwendungsbeispiels z. B. dazu genutzt werden, die Summenskala zur Patientenzufriedenheit auf Basis der Einzelitems *v2–v9* zu berechnen. Zudem kann es im Fall von übergewichtigen Patienten interessant sein, neben dem Körpergewicht den Body Mass Index (BMI) zu betrachten. Dieser bezieht neben dem Gewicht die Körpergröße mit ein und kann mittels der folgenden Formel berechnet werden: Körpergewicht (in kg)/Körpergröße (in m)<sup>2</sup>. Die folgenden Prozeduren werden anhand der Datendatei *Zentrum1\_Spieldaten.sav* demonstriert.

Für das Berechnen von neuen Variablen ist in der Menüleiste folgende Prozedur auszuwählen: *Transformieren* → *Variable berechnen*. Daraufhin öffnet sich die Dialogbox *Variable berechnen* (siehe **Abb. 4.6**). Diese Dialogbox stellt eine Reihe von (mathematischen) Ausdrücken und Funktionen zur Berechnung von neuen Variablen zur Verfügung. Die wichtigsten Datenformate werden in Abschnitt 4.2.2 erläutert.

Um die Summenskala im Anwendungsbeispiel zu berechnen, ist wie folgt zu verfahren: Für die Angabe des Namens der zu berechnenden Variable ist in dem Feld *Zielvariable* *v2\_v9\_Score* einzugeben. Für die Berechnung der Summenskala sind alle Einzelitems des Scores zu addieren. Dementsprechend sind alle acht Items, die in den Summenscore einfließen mithilfe der Transportschaltfläche in den Konditional-Editor *Numerischer Ausdruck* zu bringen und jeweils mit einem Plus-Zeichen zu verbinden (siehe **Abb. 4.6**). Praktisch ist dies so umzusetzen, dass zunächst Item *v2\_zuf* per Transportschaltfläche in den Konditional-Editor *Numerischer Ausdruck* gebracht wird. Anschließend ist mit der linken Maustaste das Plus der Rechentastatur zu betätigen. Anschließend ist das Item *v3\_zuf* wieder mit der Transportschaltfläche in das Fenster *Numerischer Ausdruck* zu bringen. Darauf folgt wieder ein Pluszeichen. Letzteres kann natürlich auch mithilfe der Eingabe über die entsprechende Tasten der Computertastatur erfolgen. Diese Vorgehensweise ist solange zu wiederholen bis alle Items in den Konditional-Editor *Numerischer Ausdruck* gebracht worden sind. Abschließend ist die Berechnung der neuen Variable mit dem Schalter *OK* zu bestätigen.

### Berechnung Summenskala

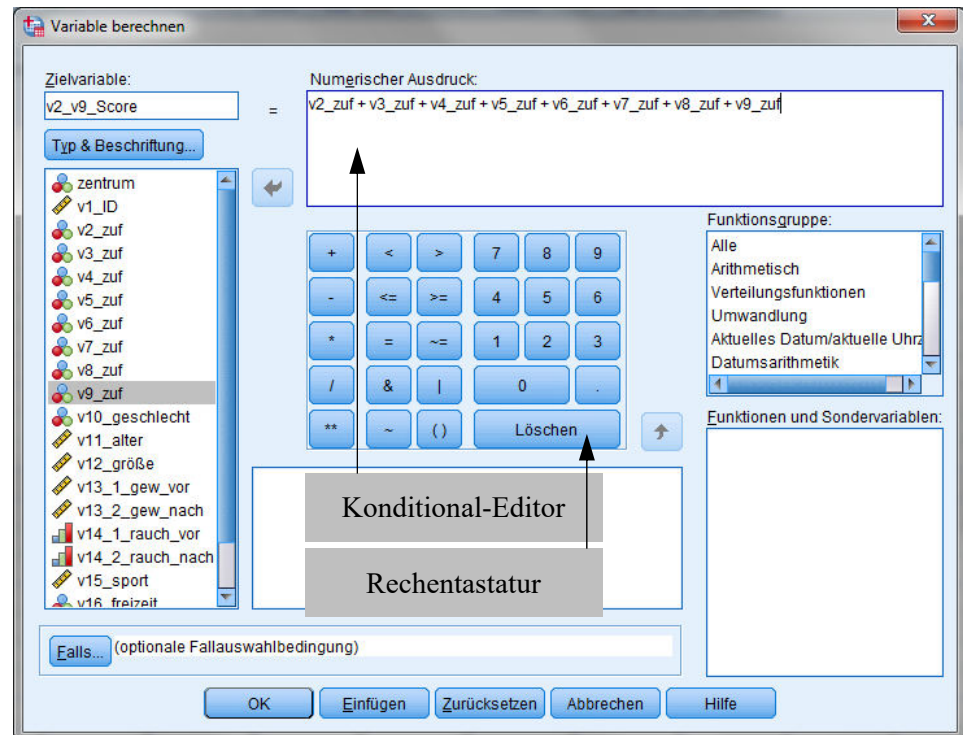


Abb. 4.6: Dialogbox – Variable berechnen: Summenscore ZUF 8

### Berechnung BMI

Für die Berechnung des personenspezifischen BMI vor der Reha ist folgendermaßen vorzugehen: Für die Angabe des Namens der zu berechnenden Variable ist in dem Feld *Zielvariable* *bmi\_vor* einzugeben. Um den BMI zu berechnen, ist zunächst die Variable *v13\_1\_gew\_vor* mithilfe der Transportschaltfläche in das Feld *Numerischer Ausdruck* zu bringen. Anschließend ist ein „/“ mithilfe des entsprechenden Schalters in das Feld *Numerischer Ausdruck* einzufügen. Dann ist die Variable *v12\_größe* in das Feld *Numerischer Ausdruck* zu bringen, die numerischen Zeichen **\*\*** (Potenzierung) und die **(,)** (Klammer-Operator) mithilfe der Funktionstasten entsprechend einzufügen, so dass am Ende die folgende Formel im Feld *Numerischer Ausdruck* steht:  $v13\_1\_gew\_vor / (v12\_größe ** 2)$ . Abschließend ist die Berechnung der neuen Variable mit dem Schalter **OK** zu bestätigen.

Die Variablen *bmi\_nach*, *diff\_gew* und *rauchen\_change* in den Datendateien 1–4 (siehe **Tabelle 1.2**) sind durch die Anwendung der Prozedur *Variable berechnen* entstanden. Das Vorgehen wird in den Übungsaufgaben zu diesem Kapitel erläutert.

### Hinweis

Abschließend noch ein wichtiger Hinweis in Bezug auf die Berechnung von neuen Variablen: Neue Variablen werden nur für die Fälle berechnet, bei denen es keine fehlenden Angaben in den Variablen gibt, die zur Berechnung der neuen Variablen verwendet werden.

### 4.2.1 Numerische Operatoren und Funktionen

Zur Formulierung von numerischen Ausdrücken stehen in SPSS unterschiedliche arithmetische und logische Operatoren zur Verfügung (siehe **Tabelle 4.1** und **Tabelle 4.2**). Diese werden dem Benutzer in unterschiedlichen Dialogboxen zur Verfügung gestellt (z.B. Dialogbox: Variable berechnen). Da die Operatoren dazu

verwendet werden können, komplexe Ausdrücke zu bilden, ist bei ihrer Anwendung auf ihre Priorität zu achten. Insgesamt gibt es vier Prioritätsstufen, wobei Operatoren mit höherer Priorität vor denen mit niedrigerer Priorität ausgeführt werden.

#### Prioritätsstufen

**Tabelle 4.1:** *Arithmetische Operatoren*

Operator	Priorität	Bedeutung
()	1	Klammer-Operator
**	2	Potenzierung
*	3	Multiplikation
/	3	Division
+	4	Addition
-	4	Subtraktion

**Tabelle 4.2:** *Logische Operatoren*

Operator	Priorität	Bedeutung
<	1	Kleiner
>	1	Größer
>=	1	Größer gleich
<=	1	Kleiner gleich
=	1	Gleich
~ =	1	Ungleich
~	2	Logisches NICHT
&	3	Logisches UND
	4	Logisches ODER

Neben den Operatoren stehen in SPSS eine Reihe von Funktionen zur Verfügung, die nach Funktionsgruppen (siehe **Abb. 4.6**) sortiert sind. Diese umfassend vorzustellen, würde den Rahmen der Einführung in das Statistikprogramm SPSS überschreiten. Neben den Funktionsgruppen Datumsarithmetik und Datumsextraktion, die v. a. für die Datenaufbereitung interessant sind und von denen in Abschnitt 4.2.2 einige vorgestellt werden, sind für weiterführende Analysen v. a. die Funktionsgruppen „Arithmetisch“ und „Statistisch“ von Interesse. Wie die Funktionen in SPSS anzuwenden sind, wird jeweils beim Anklicken der jeweiligen Funktion in Form eines kleinen Texts erläutert.

#### Funktionsgruppen

### 4.2.2 Rechnen mit speziellen Formaten

Datenformate, die häufig in gesundheitswissenschaftlichen Datenbeständen vorhanden sind, sind Datumsvariablen (z. B. Aufnahme- und Entlassungsdatum) und ICD-Kodierungen.

Das Rechnen mit diesen speziellen Formaten, macht die Anwendung spezieller Funktionen in SPSS erforderlich, die im Folgenden anhand der Datendatei *Routine-daten.sav* präsentiert werden.

#### Beispieldatensatz

### Datumsvariablen

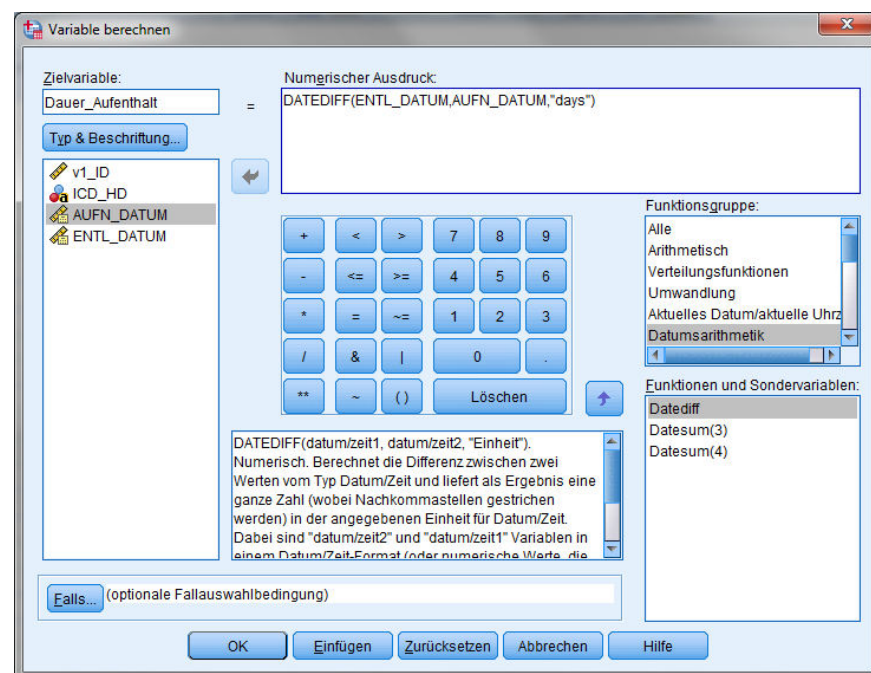
In dem Datensatz *Routinedaten.sav* sind zwei Datumsvariablen vorhanden – das Aufnahmedatum in die Klinik und das Entlassungsdatum aus der Klinik. Um die Länge des Aufenthalts (in Tagen) eines jeden Patienten zu ermitteln, muss die Differenz zwischen den beiden Daten gebildet werden. Dies ist in SPSS folgendermaßen umzusetzen: Über die allgemeine Menüleiste ist – wie für die Berechnung von neuen Variablen erforderlich – die Prozedur *Transformieren* → *Variable berechnen* auszuwählen. Daraufhin öffnet sich die *Dialogbox Variable berechnen* (siehe **Abb. 4.6**). Nachdem in das Feld *Zielvariable* *Dauer\_Aufenthalt* eingegeben wurde, ist in dem Feld *Funktionsgruppe* *Datumsarithmetik* auszuwählen. Anschließend ist in dem Feld *Funktionen und Sondervariablen* „Datediff“ per Doppelklick mit der linken Maustaste so zu aktivieren, dass in dem Feld *Numerischer Ausdruck* folgendes erscheint: *DATEDIFF(?,?,?)*. In dem Feld unter den Schaltflächen mit den unterschiedlichen Funktionen steht eine Erläuterung, die angibt, welche Angaben an den Stellen der Fragezeichen zu tätigen sind:

### Funktionsgruppe Datumsarithmetik

1. Fragezeichen: Datum 2 → *ENTL\_DATUM*
2. Fragezeichen: Datum 1 → *AUFN\_DATUM*
3. Fragezeichen: Einheit → ‚days‘

Um die Variable *ENTL\_DATUM* an die Stelle des ersten Fragezeichens zu bekommen, ist dieses zunächst mithilfe der Maus zu markieren. Anschließend ist die Variable *ENTL\_DATUM* aus der Quellvariablenliste auszuwählen und mithilfe der Transportschaltfläche an die Stelle des ersten Fragezeichens in das Feld *Numerischer Ausdruck* zu transportieren. Genauso kann das zweite Fragezeichen durch die Variable *AUFN\_DATUM* ersetzt werden. Da es das Ziel der Berechnung ist, sich die Aufenthaltsdauer in Tagen angeben zu lassen, ist das letzte Fragezeichen mit dem Ausdruck ‚days‘ zu ersetzen, indem das dritte Fragezeichen markiert wird und ‚days‘ manuell mithilfe der Tastatur eingegeben wird. Abschließend ist die Berechnung der neuen Variable mit dem Schalter *OK* zu bestätigen.

Weitere hilfreiche Funktionen finden sich in der Funktionsgruppe *Datumsextraktion*, die es ermöglichen, Bestandteile der Datumsangabe zu extrahieren (z. B. Monat, Jahr, Woche).



**Abb. 4.6:** Berechnung der Differenz zwischen zwei Datumsangaben

In Routinedaten von Kliniken bzw. Krankenkassen liegen häufig sog. ICD-Kodierungen vor (auch ICD-Klassifikation genannt), um die Diagnosen von Patientinnen und Patienten mittels eines einheitlichen Schemas zu verschlüsseln. ICD-Kodierungen sind dadurch charakterisiert, dass sie aus einer Buchstaben-Zahlen-Kombination bestehen (z. B. E11.91). Dabei ist die erste Stelle immer ein Buchstabe, darauf folgt eine zweistellige Ziffer, dann ein Punkt und wieder eine ein- bzw. zweistellige Kodierung. Die ersten drei Stellen des ICD-Kodes (Kombination aus Buchstabe plus zweistellige Ziffer) geben die allgemeine Krankheitsgruppe an – z. B. steht E10–E14 für den Diabetestyp. Die letzten beiden Ziffern spezifizieren die entsprechende Krankheit. Im Rahmen der statistischen Auswertungen kann es z. B. von Interesse sein, mithilfe der Routinedaten zu verifizieren, ob alle Patienten als Hauptdiagnose Typ-II-Diabetes haben (E11). Dazu ist es sinnvoll, diese Patienten mittels einer Variablen zu identifizieren, indem die ersten drei Ziffern aus der Variablen *ICD\_HD* extrahiert und in eine neue Variable geschrieben werden. Dazu ist über die allgemeine Menüleiste die Prozedur *Transformieren* → *Variable berechnen* auszuwählen. Daraufhin öffnet sich die *Dialogbox Variable berechnen* (siehe **Abb. 4.7**). Nachdem in das Feld *Zielvariable* „E11“ eingegeben wurde, ist in dem Feld *Funktionsgruppe* „Zeichenfolge“ auszuwählen. Anschließend ist in dem Feld *Funktionen und Sondervariablen* *Char.Substr(3)* per Doppelklick mit der linken Maustaste so zu aktivieren, dass in dem Feld *Numerischer Ausdruck* folgendes erscheint: *CHAR.SUBSTR (?,?,?)*. In dem Feld unter den Schaltflächen mit den unterschiedlichen Funktionen steht eine Erläuterung, die angibt, welche Angaben an den Stellen der Fragezeichen zu tätigen sind:

1. Fragezeichen: Ausgangsvariable → *ICD\_HD*
2. Fragezeichen: Start der Zeichenposition in der Ausgangsvariablen → 1
3. Fragezeichen: Ende der Zeichenposition in der Ausgangsvariablen → 3

Um die Variable *ICD\_HD* an die Stelle des ersten Fragezeichens zu bekommen, ist dieses zunächst mithilfe der Maus zu markieren. Anschließend ist die Variable *ICD\_HD* aus der Quellvariablenliste auszuwählen und mithilfe der Transportschaltfläche an die Stelle des ersten Fragezeichens in das Feld *Numerischer Ausdruck* zu transportieren. Das zweite Fragezeichen ist manuell durch die Ziffer 1 und das dritte Fragezeichen manuell durch die Ziffer 3 zu ersetzen.

Zusätzlich ist über die Schaltfläche *Typ & Label* die *Dialogbox Variable berechnen: Typ und Label* als *Typ String* zu definieren. Abschließend ist die Berechnung der neuen Variablen mit *OK* zu bestätigen. Sofern die neu berechnete String-Variablen für Rechenprozeduren genutzt werden soll, die über die Erstellung von Häufigkeitstabellen hinausgeht, kann diese in eine numerische Variable transformiert werden (siehe Abschnitt 4.1.3).

**Funktionsgruppe  
String**

**ICD-Kodierungen**



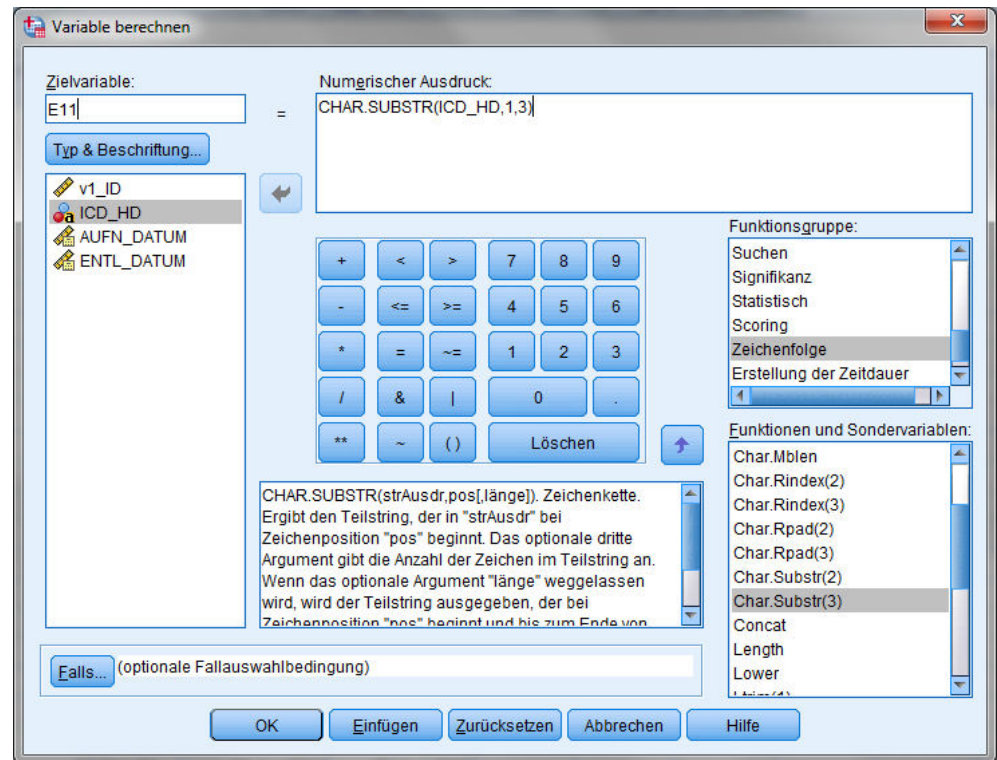


Abb. 4.7: Extraktion von Angaben aus String-Variablen

### Übungsaufgaben

Für die folgenden Übungsaufgaben ist die Datei *Zentrum1\_Spieldaten.sav* zu verwenden.

- 4.1) Berechnen Sie für jeden Patienten seinen BMI nach der Reha unter der Verwendung der Variablen *v12\_größe* und *v13\_2\_gew\_nach* und benennen Sie die neue Variable mit „*bmi\_nach*“ (Formel für BMI: Körpergewicht (in kg)/Körpergröße (in m)<sup>2</sup>).
- 4.2) Berechnen Sie die Variable „*diff\_gew*“, indem Sie die Differenz zwischen den beiden Variablen *v13\_1\_gew\_vor* und *v13\_2\_gew\_nach* bilden (*v13\_1\_gew\_vor* – *v13\_2\_gew\_nach*).
- 4.3) Berechnen Sie die Variable „*rauchen\_change*“, indem Sie die Differenz zwischen den beiden Variablen *v14\_1\_rauch\_vor* und *v14\_2\_rauch\_nach* bilden (*v14\_1\_rauch\_vor* – *v14\_2\_rauch\_nach*).



## 5 Sortierung und Auswahl von Fällen

In diesem Kapitel steht zunächst das Sortieren von Fällen (Abschnitt 5.1) im Fokus. So ist es für einige Prozeduren in SPSS erforderlich (z. B. Variablenweise Zusammenfügen), die Fälle im Daten-Editor in Abhängigkeit von bestimmten Merkmalsausprägungen zu sortieren. Diese Funktion ist zudem nützlich, um sich bei einem kleinen Datensatz ( $n < 100$ ) einen schnellen Überblick zu bestimmten Merkmalsausprägungen (z. B. fehlende Werte, Minimum, Maximum) zu verschaffen. Anschließend wird dargestellt, wie die Auswahl von Fällen in einer Datendatei erfolgt (Abschnitt 5.2). Denn zu Beginn einer jeden Auswertung ist es notwendig, zu definieren, auf Basis welcher Fälle diese durchgeführt werden soll. So erfolgt die Auswertung entweder auf Basis aller Fälle, die in einer Datendatei vorhanden sind oder die Auswertung wird nur für eine Subgruppe von Fällen durchgeführt.

Alle Beispiele in diesem Kapitel werden anhand der Datei *Zentrum1\_Spieldaten.sav* präsentiert.

**Beispieldatensatz**

### 5.1 Sortierung von Fällen

In SPSS ist es möglich, Fälle in Abhängigkeit von bestimmten Merkmalsausprägungen von einer oder mehreren Variablen zu sortieren. Dazu stehen zwei Möglichkeiten zur Verfügung, die im Folgenden anhand der Variablen *v16\_freizeit* und *v10\_geschlecht* demonstriert werden.

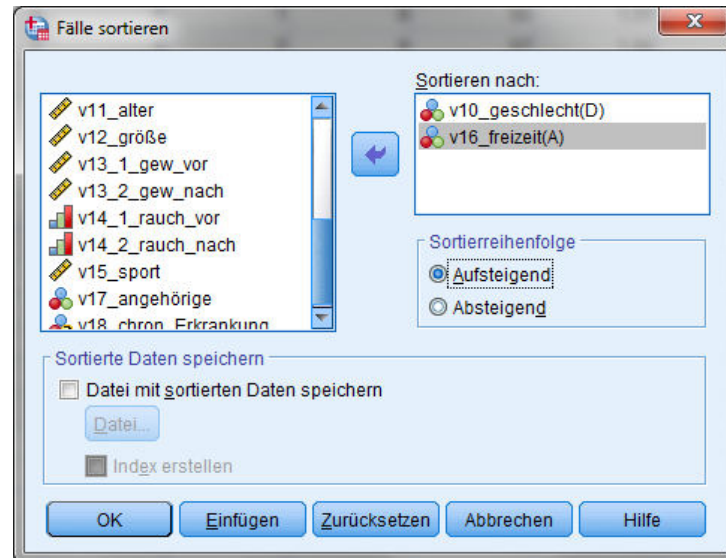
#### 5.1.1 Sortierung über die Menüauswahl

Zunächst ist es das Ziel, die Fälle in der Datendatei aufsteigend nach der Merkmalsausprägung der Variablen *v16\_freizeit* zu sortieren, d. h. die Fälle mit den geringsten Merkmalsausprägungen sollen ganz oben in der Datentabelle und die Fälle mit der höchsten Merkmalsausprägung ganz unten stehen. Dazu ist in dem allgemeinen Menü die Prozedur *Daten → Fälle sortieren* auszuwählen, sodass sich die Dialogbox *Fälle auswählen* (siehe **Abb. 5.1**) öffnet. Um die Variable *v16\_freizeit* aufsteigend zu sortieren, ist diese aus der Quellvariablenliste mit der Transportschaltfläche in das Feld *Sortieren nach* zu bringen. Im Auswahlkasten *Sortierreihenfolge* ist die Option *Aufsteigend* bereits voreingestellt, sodass die Einstellungen nur noch mit *OK* bestätigt werden müssen. In der Datenansicht des Daten-Editors ist nun zu erkennen, dass die Variablen aufsteigend nach den Merkmalsausprägungen der Variablen *v16\_freizeit* sortiert sind. Da es bei der Variable systemdefinierte fehlende Werte gibt, werden die entsprechenden Fälle an den Anfang gestellt.

**Aufsteigende Sortierung**

Angenommen, man möchte nun die beiden Variablen *v10\_geschlecht* und *v16\_freizeit* zum Sortieren verwenden, wobei in der Datendatei zunächst die Frauen (Merkmalsausprägung = 0) stehen sollen und dann die Männer (Merkmalsausprägung = 1). Innerhalb der Frauen bzw. Männer, sollen die Fälle dann aufsteigend nach der Menge der in Anspruch genommenen Freizeitangebote sortiert werden. Da die Sortierreihenfolge von der Reihenfolge der Variablen in dem Fenster *Sortieren nach* bestimmt wird, ist wie folgt zu verfahren: In der Dialogbox *Fälle sortieren* ist die Variable *v10\_geschlecht* mit der Transportschaltfläche in das Fenster *Sortieren nach* zu bringen und im Auswahlkasten *Sortierreihenfolge* ist *Absteigend* auszuwählen. Anschließend ist die Variable *v16\_freizeit* mit der Transportschaltfläche in das Fenster *Sortieren nach* zu bringen und in dem Auswahlkasten ist *Aufsteigend* zu

aktivieren (siehe **Abb. 5.1**). Hinter der Variablen *v10\_geschlecht* taucht ein D auf, was für Descending (= absteigend) steht und hinter der Variablen taucht ein A auf, was für Ascending (= aufsteigend) steht. Abschließend sind die Einstellungen mit der Taste *OK* zu bestätigen. In der Datenansicht des Daten-Editors ist das Ergebnis der Prozedur zu begutachten.

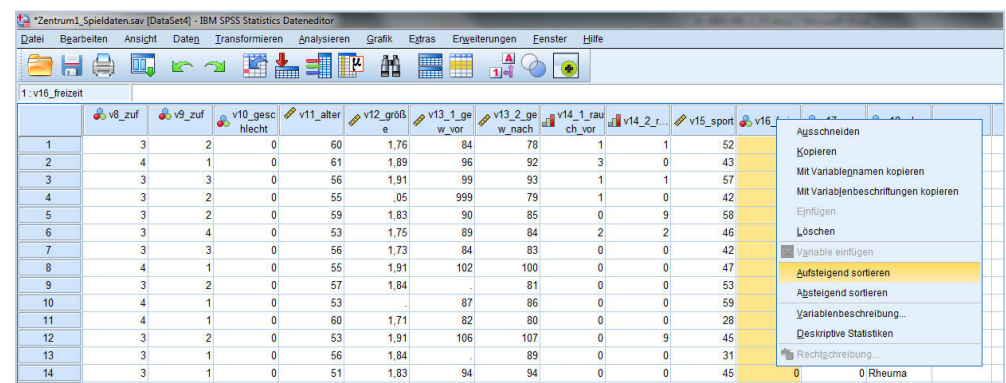


**Abb. 5.1:** Dialogbox – Fälle sortieren

### 5.1.2 Sortierung über die Datenansicht im Daten-Editor

#### Datenansicht

Um die Fälle aufsteigend nach der Variablenausprägung der Variable *v16\_freizeit* zu sortieren, ist im Daten-Editor in die *Datenansicht* zu gehen und die Spalte *v16\_freizeit* ist mit der Maus so zu markieren, dass sie sich gelb einfärbt. Mit dem Mauszeiger ist dann in der Kopfzeile auf *v16\_freizeit* zu gehen und die rechte Maustaste ist zu drücken und *Aufsteigend* auszuwählen (siehe **Abb. 5.2**). Direkt nach Betätigung der Auswahl sortiert das Programm die Daten aufsteigend nach den Merkmalsausprägungen der Variablen *v16\_freizeit*. Sofern zum Sortieren mehrere Variablen herangezogen werden sollen, sind diese alle zu markieren und entsprechend zu verfahren. Bei der Wahl von mehreren Sortiervariablen wird die Reihenfolge der Fälle von der Reihenfolge der Variablen im Daten-Editor bestimmt.



**Abb. 5.2:** Sortierung von Fällen in der Datenansicht

## 5.2 Fallauswahl unter einer Bedingung

Für die Auswahl von Fällen stehen in SPSS unterschiedliche Möglichkeiten zur Verfügung. Diese unterscheiden sich zum einen darin, ob sich die Auswahl der Fälle an bestimmten Kriterien orientiert (z. B. Auswertung soll nur für Männer durchgeführt werden) oder ob die Fälle zufällig aus einer bestehenden Datendatei ausgewählt werden sollen. Sofern die Auswahl anhand bestimmter Kriterien erfolgen soll, ist zudem zu unterscheiden, ob die Auswahl der Fälle nur temporär (Abschnitt 5.2.1) oder permanent (Abschnitt 5.2.2) erfolgen soll. Wie die Ziehung einer Zufallsstichprobe in SPSS umgesetzt werden kann, wird in Abschnitt 5.2.3 erläutert.

### 5.2.1 Temporäre Auswahl

Um die Möglichkeit der temporären Auswahl von Fällen in SPSS zu demonstrieren, soll im Folgenden davon ausgegangen werden, dass die Auswertung der Daten vorübergehend nur für die männlichen Patienten erfolgen soll. Für die temporäre Auswahl der männlichen Patienten ist im allgemeinen Menü die Auswahl *Daten* → *Fälle auswählen* vorzunehmen, sodass sich die Dialogbox *Fälle auswählen* öffnet.

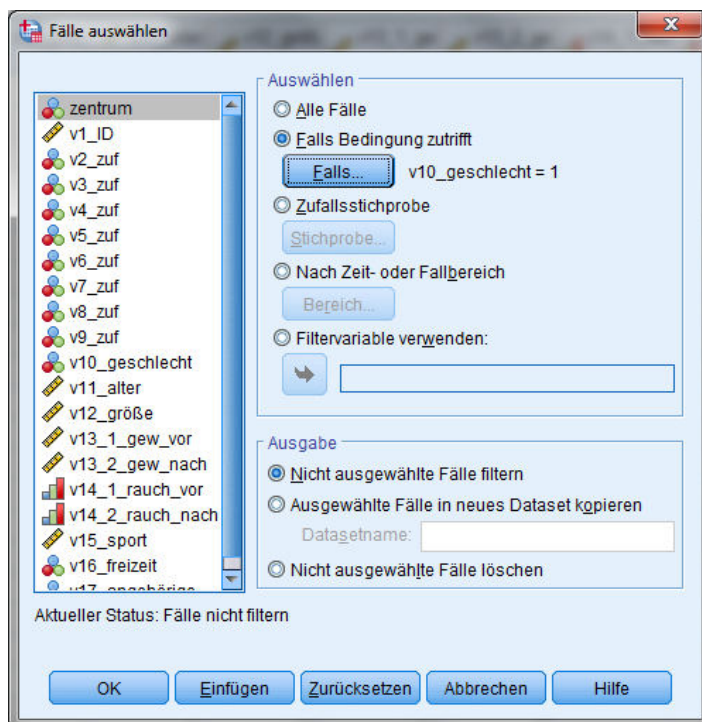


Abb. 5.3: Dialogbox – Fälle auswählen

In der Dialogbox *Fälle auswählen* ist die Auswahl *Alle Fälle* voreingestellt. Um die Fallauswahl auf die männlichen Patienten zu beschränken, ist die Einstellung *Falls Bedingung zutrifft* zu aktivieren und die Schaltfläche *Falls* zu betätigen. Daraufhin öffnet sich die Dialogbox *Fälle auswählen: Falls*. Diese Dialogbox ist vom Grundprinzip genauso aufgebaut wie die Dialogbox *Variable berechnen* (siehe Abb. 4.6). Die Variable *v10\_geschlecht* ist mithilfe der Transportschaltfläche aus der Quellvariablenliste in den Konditional-Editor zu bringen. Anschließend ist auf der Rechen tastatur das „=“ und die 1 zu aktivieren, sodass der folgende Ausdruck im Konditional-Editor steht: *v10\_geschlecht = 1*. Abschließend ist in der Dialogbox *Fälle*

### Visualisierung der temporären Auswahl

*auswählen*: Falls auf *Weiter* zu klicken, sodass sich die Dialogbox *Fälle auswählen* öffnet, in der die Einstellungen mit *OK* zu bestätigen sind.

Die temporäre Auswahl der Fälle wird an unterschiedlichen Stellen im Daten-editor sichtbar:

1. In der Datenansicht wird die Zeilennummer für die Fälle, die keine 1 als Merkmalsausprägung aufweisen, durchgestrichen.
2. Es wird eine neue Variable mit dem Namen *filter\_\$* angelegt.
- 3 In der Statusleiste steht *Filter aktiv*.

### Aufheben des Filters

Die temporäre Auswahl der Fälle bleibt solange erhalten, bis sie wieder deaktiviert wird oder die Datei geschlossen und wieder neu geöffnet wird. Zum aktiven Aufheben des Filters ist über die Menüfunktion *Daten → Fälle auswählen* wieder die Dialogbox *Fälle auswählen* zu öffnen und die Einstellung von *Falls Bedingung zutrifft* auf *Alle Fälle* zu ändern. Die Folge ist, dass das *Filter aktiv* aus der Statusleiste verschwindet und das in der Datenansicht keine Fälle mehr durchgestrichen sind. Die neugebildete Variable bleibt allerdings erhalten.

### 5.2.2 Permanente Auswahl

### Dauerhafter Ausschluss von Fällen

Anstatt eine bestimmte Fallauswahl nur temporär vorzunehmen, ist es auch möglich, diese als permanente Auswahl umzusetzen. Um die männlichen Patienten permanent auszuwählen, ist über das Menü *Daten → Fälle auswählen* zu öffnen. Dann ist in der Dialogbox *Fälle auswählen* die Einstellung auf *Falls Bedingung zutrifft* zu ändern. In der Dialogbox *Fälle auswählen*: *Falls* ist die gleiche Einstellung vorzunehmen wie im Fall der temporären Auswahl (siehe Abschnitt 5.2.1). Um nun eine permanente Auswahl zu bewirken, bietet SPSS die folgenden beiden Möglichkeiten an:

1. In der Dialogbox *Fälle auswählen* ist in dem Auswahlfeld *Ausgabe* die Einstellung *Ausgewählte Fälle in neues Datenblatt kopieren* zu aktivieren und ein Name für die neue Datei (z. B. *zentrum1\_männlich*) anzugeben.
2. In der Dialogbox *Fälle auswählen* ist in dem Auswahlfeld *Ausgabe* die Einstellung *Nicht ausgewählte Fälle löschen* zu aktivieren. Bei dieser Option ist darauf zu achten, dass **zuvor eine Sicherheitskopie der ursprünglichen Datendatei** angelegt wurde.

### 5.2.3 Zufallsstichprobe

Im Rahmen von empirischen Erhebungen kann es Situationen geben, die es erforderlich machen, aus einer vorhandenen Datendatei eine Zufallsstichprobe zu ziehen. Beispielsweise dann, wenn man nur für einen Teil der empirisch gewonnenen Daten einen Abgleich mit dem Urbeleg machen möchte. In diesem Fall ist es sinnvoll, eine Zufallsstichprobe aus der bestehenden Datendatei zu ziehen.

Wie die Ziehung einer Zufallsstichprobe in SPSS praktisch umzusetzen ist, wird im Folgenden daran demonstriert, wie zufällig 10 Patienten aus der Datei *Zentrum1\_Spieldaten.sav* gezogen werden. Dazu ist im allgemeinen Menü die Auswahl *Daten → Fälle auswählen* vorzunehmen, sodass sich die Dialogbox *Fälle auswählen* öffnet. In der Dialogbox *Fälle auswählen* ist die Einstellung *Zufallsstichprobe* vorzunehmen. Anschließend ist auf den Schalter *Stichprobe* zu klicken, sodass sich die Dialogbox *Fälle auswählen: Zufallsstichprobe* öffnet (siehe **Abb. 5.4**). Um genau 10

Fälle aus der Datendatei zufällig zu ziehen ist die Einstellung *Exakt* zu aktivieren und rechts daneben in das Feld die Zahl 10 einzutragen. In das zweite Feld *Fälle aus den ersten ... Fällen* ist für das Anwendungsbeispiel die Anzahl der Fälle der Datei einzugeben, d. h. die Zahl 30. Die vorgenommenen Einstellungen sind mit der Taste *Weiter* zu bestätigen, sodass man wieder in die Dialogbox *Fälle auswählen* zurückkommt. Jetzt muss man sich entscheiden, ob man die Zufallsstichprobe als temporäre oder permanente Auswahl realisieren möchte.

- Sofern eine **temporäre Auswahl** gewünscht ist, ist die Einstellung *Nicht ausgewählte Fälle filtern* vorzunehmen.
- Im Falle einer **permanenten Auswahl** ist in dem Auswahlfeld *Ausgabe* entweder die Einstellung *Ausgewählte Fälle in neues Datenblatt kopieren* zu aktivieren und ein Name für die neue Datei (z. B. *Zentrum1\_weiblich*) anzugeben oder die Einstellung *Nicht ausgewählte Fälle löschen* zu aktivieren.

**Temporäre oder  
permanente Auswahl**

Abschließend sind die vorgenommenen Einstellungen mit der Taste *OK* zu bestätigen.

Alternativ zu der Ziehung einer Zufallsstichprobe mit exakt definierter Größe ist es auch möglich, eine ungefähre Prozentzahl an Fällen festzulegen, die aus der Datendatei gezogen werden soll. Dazu ist die Einstellung *Ungefähr ... % aller Fälle* zu aktivieren und in das entsprechende Feld ist eine Prozentzahl einzutragen.

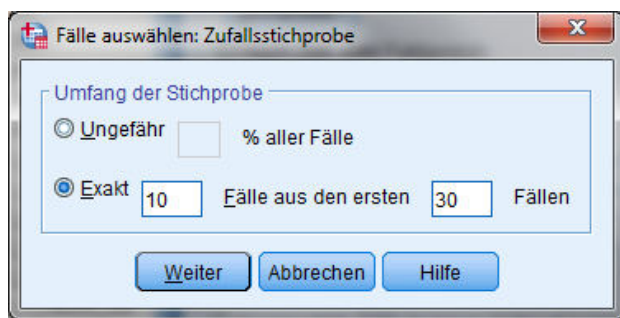


Abb. 5.4: Dialogbox – Fälle auswählen: Zufallsstichprobe

Für die Ziehung der Zufallsstichprobe wird in SPSS als Voreinstellung ein Zufallszahlengenerator verwendet, bei dem immer ein unterschiedlicher Startwert generiert wird. Dies hat zur Folge, dass sich die Zufallsstichprobe stets aus anderen Fällen zusammensetzt. Möchte man sicherstellen, dass die Ziehung der Stichprobe reproduzierbar ist – was zu empfehlen ist – dann ist ein fester Startwert festzulegen. Dazu ist im Menü die folgende Auswahl zu treffen *Transformieren → Zufallszahlengenerator*. Daraufhin öffnet sich die Dialogbox *Zufallszahlengenerator* (siehe Abb. 5.5). In dem Auswahlkasten *Aktiver Generator* stehen die folgenden beiden Zufallszahlengeneratoren mit entsprechenden Eigenschaften zur Verfügung (Bühl, 2010, S. 194):

**Zufallszahlengenerator**

1. *Mit SPSS 12 kompatibel*: Hierbei handelt es sich um einen Zufallszahlengenerator, der bis zur Version 12 von SPSS verwendet wurde. Sofern zufallsbedingte Ergebnisse erzeugt werden sollen, die mit älteren Versionen von SPSS reproduziert werden sollen, ist dieser Generator zu verwenden.
2. *Mersenne-Twister*: Hierbei handelt es sich um einen neueren Zufallszahlengenerator, der für Simulationszwecke eine höhere Zuverlässigkeit aufweist und folglich zu präferieren ist.

Einer der beiden Generatoren ist auszuwählen, indem die Checkbox *Aktiven Generator festlegen* aktiviert und der entsprechende Generator angeklickt wird.

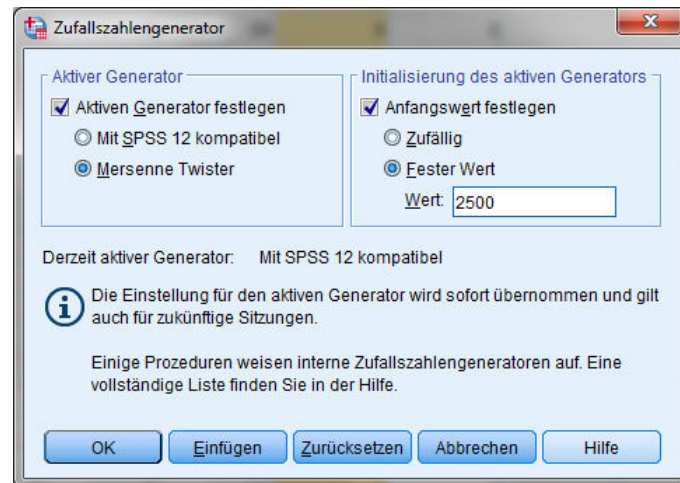


Abb. 5.5: Dialogbox - Zufallszahlengenerator

Um nun einen festen Startwert in SPSS festzulegen, ist in der Dialogbox *Zufallszahlengenerator* im Auswahlkasten *Initialisierung des aktiven Generators* die Check-box *Anfangswert festlegen* zu aktivieren und ein beliebiger ganzzahliger positiver Wert einzutragen (z. B. 2500). Nachdem ein fester Startwert für den Zufallszahlengenerator festgelegt worden ist, muss die Prozedur der Ziehung der Zufallsstichprobe wiederholt werden, damit die vorgenommenen Einstellungen bei der Ziehung der Zufallsstichprobe Berücksichtigung finden.

### Übungsaufgaben

Für die folgenden Übungsaufgaben ist die Datei *Zentrum1\_Spieldaten.sav* zu verwenden.

- 5.1) Wählen Sie die Patienten temporär in der Datei aus, bei denen die Angehörigen nicht mit in die Behandlung einbezogen wurden. Bei wieviel Patienten werden die Angehörigen nicht mit in die Behandlung einbezogen?
- 5.2) Speichern Sie die Patienten, die älter als 60 Jahre alt sind, in einer neuen Datendatei mit der Bezeichnung *Patienten\_älter60.sav* ab. Wie viele Personen enthält der neue Datensatz?
- 5.3) Ziehen Sie eine Zufallsstichprobe von 8 Patienten mit dem Zufallszahlengenerator *Mersenne-Twister* und einem Startwert von 5000 und speichern Sie diese in einer neuen Datei mit der Bezeichnung *Zufallsstichprobe\_8Patienten.sav*.

## Zusammenfassung

Das Ziel dieses Studienbriefs bestand darin, das Programm SPSS in seinen Grundfunktionen kennenzulernen. Dabei lag der besondere Fokus darauf, die Prozeduren zu vermitteln, die insbesondere für die Datenaufbereitung zentral sind.

Nachdem das Anwendungsbeispiel und die daraus resultierenden Beispieldateien ausführlich vorgestellt wurden, zielten die folgenden Kapitel zunächst darauf ab, die Grundstruktur des Programms von SPSS zu präsentieren. Ein wichtiges Lernziel bestand darin, den grundsätzlichen Aufbau von SPSS, der durch die drei unterschiedlichen Fenster Daten-Editor, Viewer und Syntaxfenster gekennzeichnet ist, nachzuvollziehen. Darüber hinaus wurden Grundkenntnisse zur Syntaxsprache von SPSS vermittelt. Zudem wurde dargestellt, wie individuelle Einstellungen am Programm vorgenommen werden können und wie die ausführliche Hilfe von SPSS genutzt werden kann.

Die folgenden Kapitel orientierten sich dann an den unterschiedlichen Schritten der Datenaufbereitung, die unter Bezugnahme auf das Anwendungsbeispiel demonstriert wurden. So bestand bei der Patientenbefragung der erste Schritt der Datenaufbereitung darin, einen Kodeplan zu erstellen. Dieser Kodeplan hatte zum Ziel, den Variablennamen und Merkmalsausprägungen im Fragebogen eindeutige Kodierungen zuzuweisen. Auf Grundlage des Kodeplans wurden dann neue Variablen in der Variablenansicht in SPSS angelegt. Anschließend wurde anhand eines Beispielfalls die Dateneingabe in den Daten-Editor demonstriert.

Da es häufig der Fall ist, dass man keine primäre Datenerfassung durchführt, sondern die Daten zur statistischen Datenanalyse digital in unterschiedlichen Datenformaten erhält, wurde dargestellt, wie bereits existierende Datendateien in SPSS eingelesen werden können. Zudem wurde präsentiert, wie unterschiedliche Datendateien zusammengefügt werden können.

Nachdem alle Daten in SPSS manuell eingegeben bzw. eingelesen wurden, bestand der nächste Schritt der Datenaufbereitung in der ausführlichen Datenprüfung. Diesbezüglich wurden einige Methoden vorgestellt, die angewendet werden können, um Eingabefehler in den Daten zielgerichtet aufzudecken.

Im Anschluss an die Datenprüfung wurde gezeigt, wie die Daten modifiziert werden können. Die Modifikation von Daten zielt darauf ab, bestehende Variablen in der Datendatei so zu verändern, dass sie sinnvoll in der statistischen Auswertung genutzt werden können. Beispielsweise wurde aus den Variablen Körpergröße und Körpergewicht der BMI der Patienten berechnet. Weiterhin lag ein besonderer Fokus auf der Modifikation von Datumsvariablen und ICD-Kodierungen, da diese Datenformate häufig in gesundheitswissenschaftlichen Datenbeständen zu finden sind. Der Studienbrief schließt mit einem Kapitel ab, das die Auswahl von Fällen als einen weiteren wichtigen Bestandteil der Datenaufbereitung thematisiert.

Die in dem Studienbrief dargestellten Grundfunktionen von SPSS und Prozeduren zur Datenaufbereitung sind für ein effizientes Arbeiten mit SPSS zentral und spielen auch für die Anwendung komplexer statistischer Methoden, die in den folgenden Studienbriefen des Moduls präsentiert werden, eine wichtige Rolle.

### Grundstruktur von SPSS

### Kodeplanerstellung und Dateneingabe

### Einlesen von Daten

### Datenprüfung

### Datenmodifikation

### Ausblick

# Anhang 1 - Patientenfragebogen

## Patientenfragebogen

Liebe Patientinnen, liebe Patienten,  
wir hoffen, Sie haben Ihren Aufenthalt bei uns genossen. Bevor wir Sie nach Hause entlassen, würden wir Sie darum bitten, uns die folgenden Fragen zu Ihrer Zufriedenheit mit Ihrem Aufenthalt bei und zu Ihrer Person zu beantworten. Damit unterstützen Sie uns dabei, unsere Angebote und unseren Service stetig zu verbessern. Ihre Daten werden selbstverständlich vertraulich behandelt.

<b>0. Nr. des Zentrums</b>			
(wird von Klinik ausgefüllt)			
<b>1. Patientenidentifikationsnummer</b>			
(wird bei Zustimmung des Patienten von dem Arzt ausgefüllt)			
<b>2. Wie würden Sie die Qualität der Behandlung, welche Sie erhalten haben, beurteilen?</b>			
ausgezeichnet <input type="checkbox"/>	gut <input type="checkbox"/>	weniger gut <input type="checkbox"/>	schlecht <input type="checkbox"/>
<b>3. Haben Sie die Art von Behandlung erhalten, die Sie wollten?</b>			
eindeutig nicht <input type="checkbox"/>	eigentlich nicht <input type="checkbox"/>	im Allgemeinen ja <input type="checkbox"/>	eindeutig ja <input type="checkbox"/>
<b>4. In welchem Maße hat unsere Klinik Ihren Bedürfnissen entsprochen?</b>			
sie hat fast allen meinen Bedürfnissen entsprochen <input type="checkbox"/>	sie hat den meisten meiner Bedürfnisse entsprochen <input type="checkbox"/>	sie hat nur wenigen meiner Bedürfnisse entsprochen <input type="checkbox"/>	sie hat meinen Bedürfnissen nicht entsprochen <input type="checkbox"/>
<b>5. Würden Sie einem Freund / einer Freundin unsere Klinik empfehlen, wenn er / sie eine ähnliche Hilfe benötigen würde?</b>			
eindeutig nicht <input type="checkbox"/>	ich glaube nicht <input type="checkbox"/>	ich glaube ja <input type="checkbox"/>	eindeutig ja <input type="checkbox"/>
<b>6. Wie zufrieden sind Sie mit dem Ausmaß der Hilfe, welche Sie hier erhalten haben?</b>			
ziemlich unzufrieden <input type="checkbox"/>	leidlich oder leicht unzufrieden <input type="checkbox"/>	weitgehend zufrieden <input type="checkbox"/>	sehr zufrieden <input type="checkbox"/>
<b>7. Hat die Behandlung, die Sie hier erhielten, Ihnen dabei geholfen, angemessener mit Ihren Problemen umzugehen?</b>			
ja, sie half eine ganze Menge <input type="checkbox"/>	ja, sie half etwas <input type="checkbox"/>	nein, sie half eigentlich nicht <input type="checkbox"/>	nein, sie hat mir die Dinge schwerer gemacht <input type="checkbox"/>
<b>8. Wie zufrieden sind Sie mit der Behandlung, die Sie erhalten haben, im Großen und Ganzen?</b>			
sehr zufrieden <input type="checkbox"/>	weitgehend zufrieden <input type="checkbox"/>	leidlich oder leicht unzufrieden <input type="checkbox"/>	ziemlich unzufrieden <input type="checkbox"/>
<b>9. Würden Sie wieder in unsere Klinik kommen, wenn Sie eine Hilfe bräuchten?</b>			
eindeutig nicht <input type="checkbox"/>	ich glaube nicht <input type="checkbox"/>	ich glaube ja <input type="checkbox"/>	eindeutig ja <input type="checkbox"/>
<b>10. Geschlecht</b>			
<input type="checkbox"/> weiblich <input type="checkbox"/> männlich			
<b>11. Alter</b>			
_____ Jahre			
<b>12. Körpergröße</b>			
_____ m			
<b>13. Körpergewicht</b>			
vor Reha _____ kg			
nach Reha _____ kg			
<b>14. Raucherstatus</b>			
vor Reha <input type="checkbox"/> Nichtraucher <input type="checkbox"/> Gelegenheitsraucher (d.h. durchschnittlich weniger als eine Zigarette am Tag) <input type="checkbox"/> Raucher			
nach Reha <input type="checkbox"/> Nichtraucher <input type="checkbox"/> Gelegenheitsraucher (d.h. durchschnittlich weniger als eine Zigarette am Tag) <input type="checkbox"/> Raucher			
<b>15. Durchschnittliche Anzahl an Minuten, die Sie sich während Ihres Aufenthalts täglich sportlich betätigt haben:</b>			
_____ Minuten (pro Tag)			
<b>16. Anzahl an wahrgenommenen Freizeitangebote während Ihres Aufenthalts:</b>			
<input type="checkbox"/> bis 2 <input type="checkbox"/> 3-6 <input type="checkbox"/> >6			
<b>17. Einbezug von Angehörigen in Ihre Behandlung:</b>			
<input type="checkbox"/> ja <input type="checkbox"/> nein			
<b>18. Leiden Sie neben Typ-II-Diabetes unter weiteren chronischen Erkrankungen? Falls ja, welche?</b>			
_____			



Vielen Dank für Ihre Unterstützung und eine gute Heimreise!  
Ihr Team des Verbunds der Reha-Zentren „Nordstern“



Anmerkung: Fragen 2 bis 9 = ZUF-8 (Schmidt & Nübling, 2002; Schmidt et al. 1989, 1994).



## Anhang 2 - Kodeplan

### Patientenfragebogen

Liebe Patientinnen, liebe Patienten,  
wir hoffen, Sie haben Ihren Aufenthalt bei uns genossen. Bevor wir Sie nach Hause entlassen, würden wir Sie darum bitten, uns die folgenden Fragen zu Ihrer Zufriedenheit mit Ihrem Aufenthalt bei und zu Ihrer Person zu beantworten. Damit unterstützen Sie uns dabei, unsere Angebote und unseren Service stetig zu verbessern. Ihre Daten werden selbstverständlich vertraulich behandelt.

<b>zentrum 0. Nr. des Zentrums</b>			
_____ (wird von Klinik ausgefüllt)			
<b>v1_ID 1. Patientenidentifikationsnummer</b>			
_____ (wird bei Zustimmung des Patienten von dem Arzt ausgefüllt)			
<b>v2_zuf 2. Wie würden Sie die Qualität der Behandlung, welche Sie erhalten haben, beurteilen?</b>			
ausgezeichnet <input type="checkbox"/> _1	gut <input type="checkbox"/> _2	weniger gut <input type="checkbox"/> _3	schlecht <input type="checkbox"/> _4
<b>v3_zuf 3. Haben Sie die Art von Behandlung erhalten, die Sie wollten?</b>			
eindeutig nicht <input type="checkbox"/> _1	eigentlich nicht <input type="checkbox"/> _2	im Allgemeinen ja <input type="checkbox"/> _3	eindeutig ja <input type="checkbox"/> _4
<b>v4_zuf 4. In welchem Maße hat unsere Klinik Ihren Bedürfnissen entsprochen?</b>			
sie hat fast allen meinen Bedürfnissen entsprochen <input type="checkbox"/> _1	sie hat den meisten meiner Bedürfnisse entsprochen <input type="checkbox"/> _2	sie hat nur wenigen meiner Bedürfnisse entsprochen <input type="checkbox"/> _3	sie hat meinen Bedürfnissen nicht entsprochen <input type="checkbox"/> _4
<b>v5_zuf 5. Würden Sie einem Freund / einer Freundin unsere Klinik empfehlen, wenn er / sie eine ähnliche Hilfe benötigen würde?</b>			
eindeutig nicht <input type="checkbox"/> _1	ich glaube nicht <input type="checkbox"/> _2	ich glaube ja <input type="checkbox"/> _3	eindeutig ja <input type="checkbox"/> _4
<b>v6_zuf 6. Wie zufrieden sind Sie mit dem Ausmaß der Hilfe, welche Sie hier erhalten haben?</b>			
ziemlich unzufrieden <input type="checkbox"/> _1	leidlich oder leicht unzufrieden <input type="checkbox"/> _2	weitgehend zufrieden <input type="checkbox"/> _3	sehr zufrieden <input type="checkbox"/> _4
<b>v7_zuf 7. Hat die Behandlung, die Sie hier erhielten, Ihnen dabei geholfen, angemessener mit Ihren Problemen umzugehen?</b>			
ja, sie half eine ganze Menge <input type="checkbox"/> _1	ja, sie half etwas <input type="checkbox"/> _2	nein, sie half eigentlich nicht <input type="checkbox"/> _3	nein, sie hat mir die Dinge schwerer gemacht <input type="checkbox"/> _4
<b>v8_zuf 8. Wie zufrieden sind Sie mit der Behandlung, die Sie erhalten haben, im Großen und Ganzen?</b>			
sehr zufrieden <input type="checkbox"/> _1	weitgehend zufrieden <input type="checkbox"/> _2	leidlich oder leicht unzufrieden <input type="checkbox"/> _3	ziemlich unzufrieden <input type="checkbox"/> _4
<b>v9_zuf 9. Würden Sie wieder in unsere Klinik kommen, wenn Sie eine Hilfe bräuchten?</b>			
eindeutig nicht <input type="checkbox"/> _1	ich glaube nicht <input type="checkbox"/> _2	ich glaube ja <input type="checkbox"/> _3	eindeutig ja <input type="checkbox"/> _4
<b>v10_geschlecht 10. Geschlecht</b>			
<input type="checkbox"/> _0 weiblich <input type="checkbox"/> _1 männlich			
<b>v11_alter 11. Alter</b>			
_____ Jahre			
<b>v12_größe 12. Körpergröße</b>			
_____ m			
<b>13. Körpergewicht</b>			
v13_1_gew_vor vor Reha _____ kg			
v13_2_gew_nach nach Reha _____ kg			
<b>14. Raucherstatus</b>			
v14_1_rauch_vor vor Reha		<input type="checkbox"/> _0 Nichtraucher <input type="checkbox"/> _1 Gelegenheitsraucher <input type="checkbox"/> _2 Raucher	
v14_2_rauch_nach nach Reha		<input type="checkbox"/> _0 Nichtraucher <input type="checkbox"/> _1 Gelegenheitsraucher <input type="checkbox"/> _2 Raucher	
<b>v15_sport 15. Durchschnittliche Anzahl an Minuten, die Sie sich während Ihres Aufenthalts täglich sportlich betätigt haben:</b>			
_____ Minuten (pro Tag)			
<b>v16_freizeit 16. Anzahl an wahrgenommenen Freizeitangebote während Ihres Aufenthalts:</b>			
<input type="checkbox"/> _0 bis 2 <input type="checkbox"/> _1 3-6 <input type="checkbox"/> _2 >6			
<b>v17_angehörige 17. Einbezug von Angehörigen in Ihre Behandlung:</b>			
<input type="checkbox"/> _1 ja <input type="checkbox"/> _0 nein			
<b>v18_chron_Erkrankung 18. Leiden Sie neben Typ-II-Diabetes unter weiteren chronischen Erkrankungen? Falls ja, welche?</b>			
_____			



Vielen Dank für Ihre Unterstützung und eine gute Heimreise!  
Ihr Team des Verbunds der Reha-Zentren „Nordstern“



Anmerkung: Fragen 2 bis 9 = ZUF-8 (Schmidt & Nübling, 2002; Schmidt et al. 1989, 1994).

## Anhang 3 – Beispielfall

### Patientenfragebogen

Liebe Patientinnen, liebe Patienten,  
wir hoffen, Sie haben Ihren Aufenthalt bei uns genossen. Bevor wir Sie nach Hause entlassen, würden wir Sie darum bitten, uns die folgenden Fragen zu Ihrer Zufriedenheit mit Ihrem Aufenthalt bei und zu Ihrer Person zu beantworten. Damit unterstützen Sie uns dabei, unsere Angebote und unseren Service stetig zu verbessern. Ihre Daten werden selbstverständlich vertraulich behandelt.

<b>zentrum 0. Nr. des Zentrums</b>			
0 (wird von Klinik ausgefüllt)			
<b>v1_ID 1. Patientenidentifikationsnummer</b>			
0 (wird bei Zustimmung des Patienten von dem Arzt ausgefüllt)			
<b>v2_zuf 2. Wie würden Sie die Qualität der Behandlung, welche Sie erhalten haben, beurteilen?</b>			
ausgezeichnet <input type="checkbox"/>	gut <input checked="" type="checkbox"/>	weniger gut <input type="checkbox"/>	schlecht <input type="checkbox"/>
<b>v3_zuf 3. Haben Sie die Art von Behandlung erhalten, die Sie wollten?</b>			
eindeutig nicht <input type="checkbox"/>	eigentlich nicht <input type="checkbox"/>	im Allgemeinen ja <input type="checkbox"/>	eindeutig ja <input checked="" type="checkbox"/>
<b>v4_zuf 4. In welchem Maße hat unsere Klinik Ihren Bedürfnissen entsprochen?</b>			
sie hat fast allen meinen Bedürfnissen entsprochen <input checked="" type="checkbox"/>	sie hat den meisten meiner Bedürfnisse entsprochen <input type="checkbox"/>	sie hat nur wenigen meiner Bedürfnisse entsprochen <input type="checkbox"/>	sie hat meinen Bedürfnissen nicht entsprochen <input type="checkbox"/>
<b>v5_zuf 5. Würden Sie einem Freund / einer Freundin unsere Klinik empfehlen, wenn er / sie eine ähnliche Hilfe benötigen würde?</b>			
eindeutig nicht <input checked="" type="checkbox"/>	ich glaube nicht <input type="checkbox"/>	ich glaube ja <input type="checkbox"/>	eindeutig ja <input type="checkbox"/>
<b>v6_zuf 6. Wie zufrieden sind Sie mit dem Ausmaß der Hilfe, welche Sie hier erhalten haben?</b>			
ziemlich unzufrieden <input type="checkbox"/>	leidlich oder leicht unzufrieden <input type="checkbox"/>	weitgehend zufrieden <input checked="" type="checkbox"/>	sehr zufrieden <input type="checkbox"/>
<b>v7_zuf 7. Hat die Behandlung, die Sie hier erhielten, Ihnen dabei geholfen, angemessener mit Ihren Problemen umzugehen?</b>			
ja, sie half eine ganze Menge <input checked="" type="checkbox"/>	ja, sie half etwas <input type="checkbox"/>	nein, sie half eigentlich nicht <input type="checkbox"/>	nein, sie hat mir die Dinge schwerer gemacht <input type="checkbox"/>
<b>v8_zuf 8. Wie zufrieden sind Sie mit der Behandlung, die Sie erhalten haben, im Großen und Ganzen?</b>			
sehr zufrieden <input type="checkbox"/>	weitgehend zufrieden <input checked="" type="checkbox"/>	leidlich oder leicht unzufrieden <input type="checkbox"/>	ziemlich unzufrieden <input type="checkbox"/>
<b>v9_zuf 9. Würden Sie wieder in unsere Klinik kommen, wenn Sie eine Hilfe bräuchten?</b>			
eindeutig nicht <input type="checkbox"/>	ich glaube nicht <input type="checkbox"/>	ich glaube ja <input checked="" type="checkbox"/>	eindeutig ja <input type="checkbox"/>
<b>v10_geschlecht 10. Geschlecht</b>			
<input checked="" type="checkbox"/> weiblich <input type="checkbox"/> männlich			
<b>v11_alter 11. Alter</b>			
55 Jahre			
<b>v12_größe 12. Körpergröße</b>			
1,64 m			
<b>13. Körpergewicht</b>			
v13_1_gew_vor Reha 85 kg			
v13_2_gew_nach Reha 80 kg			
<b>14. Raucherstatus</b>			
v14_1_rauch_vor Reha <input type="checkbox"/> Nichtraucher <input type="checkbox"/> Gelegenheitsraucher <input checked="" type="checkbox"/> Raucher			
v14_2_rauch_vor nach Reha <input type="checkbox"/> Nichtraucher <input checked="" type="checkbox"/> Gelegenheitsraucher <input type="checkbox"/> Raucher			
<b>v15_sport 15. Durchschnittliche Anzahl an Minuten, die Sie sich während Ihres Aufenthalts täglich sportlich betätigt haben:</b>			
60 Minuten (pro Tag)			
<b>v16_freizeit 16. Anzahl an wahrgenommenen Freizeitangebote während Ihres Aufenthalts:</b>			
<input type="checkbox"/> bis 2 <input type="checkbox"/> 3-6 <input type="checkbox"/> >6			
<b>v17_angehörige 17. Einbezug von Angehörigen in Ihre Behandlung:</b>			
<input checked="" type="checkbox"/> ja <input type="checkbox"/> nein			
<b>v18_chron_Erkrankung 18. Leiden Sie neben Typ-II-Diabetes unter weiteren chronischen Erkrankungen? Falls ja, welche?</b>			
COPD			



Vielen Dank für Ihre Unterstützung und eine gute Heimreise!  
Ihr Team des Verbunds der Reha-Zentren „Nordstern“



Anmerkung: Fragen 2 bis 9 = ZUF-8 (Schmidt & Nübling, 2002; Schmidt et al. 1989, 1994).

# Glossar

**Bivariat:** Verfahren, in denen zwei Merkmale (Variablen) statistisch analysiert werden.

**Fall:** Erhebungseinheit in einer empirischen Untersuchung.

**Grundgesamtheit:** Bezeichnung der gesamten Population, die sich in der Regel auf Grund der Größe und Dimension nicht erfassen lässt und deswegen geschätzt werden muss

**Intervallskaliert:** Variablen mit Intervallskala.

**Intervallskala:** Variablen, deren Merkmalsausprägungen nicht nur eine Rangordnung, sondern zudem gleiche Abstände zwischen den numerischen Ausprägungen aufweisen.

**Median:** Wert, der in der Mitte aller beobachteten Werte einer Variablen liegt.

**Merkmalsausprägung:** unterschiedliche Ausprägungen einer Variablen.

**Mittelwert:** entspricht hier dem arithmetischen Mittelwert und ist die Summe der beobachteten Werte geteilt durch die Anzahl der beobachteten Werte einer Variablen.

**Modus:** der am häufigsten auftretende Wert einer Variablen.

**Multivariat:** Verfahren, in denen mindestens drei Merkmale (Variablen) statistisch analysiert werden.

**Nominal:** Eigenschaft einer Variablen, bei der sich deren Ausprägungen keiner Ordnung unterziehen lassen (z. B. Haarfarbe).

**Ordinal:** Eigenschaft einer Variablen, bei der sich deren Ausprägungen einer Ordnung unterziehen lassen (z. B. Schulnoten).

**Spreadsheet:** (dt. *Arbeitsblatt*). Bezeichnung für das in Zeilen und Spalten eingeteilte Bildschirmfenster einer Software (z. B. Excel, SPSS). Je nach Programm bzw. Bedienungskonzept heißt dieser Bereich zum Beispiel *Arbeitsblatt*, *Worksheet* oder *Spreadsheet*

**Stichprobe:** zufälliges Ziehen aus der Grundgesamtheit.

**Syntax:** Programmiersprache.

**Variable:** ein bestimmtes Merkmal, das im Rahmen einer empirischen Untersuchung erhoben wird und das von Fall zu Fall unterschiedliche Ausprägungen aufweisen kann.

**Verhältnisskaliert:** Variablen mit Verhältnisskala.

**Verhältnisskala:** Variablen, deren Ausprägungen nicht nur in eine Rangreihe zu bringen sind und gleiche Abstände aufweisen, sondern zusätzlich einen natürlichen Nullpunkt aufweisen (auch Ratioskala genannt).

**Zufallsstichprobe:** zufällige Ziehung von Fällen aus einer definierten Gruppe von Fällen.

## Lösungen zu den Übungsaufgaben

### Lösungen zu Kapitel 1

- 1.1) Im Fragebogen sind die folgenden drei Antwortformate zu finden:
- Geschlossenes Antwortformat (z.B. Frage 6): Hier sind die Antwortkategorien mit den Ausprägungen „ziemlich unzufrieden“, „leidlich oder leicht unzufrieden“, „weitgehend zufrieden“ und „sehr zufrieden“ fest vorgegeben.
  - Halboffenes Antwortformat (z.B. Frage 11): Hier müssen die Befragten ihr (unterschiedliches) Alter in Jahren angeben.
  - Offenes Antwortformat (z.B. Frage 18): Hier ist grundsätzlich jede Antwort möglich.

### Lösungen zu Kapitel 2

- 1.2) Die beiden Dateien *Routinedaten.sav* und *Routinedaten.xls* unterscheiden sich inhaltlich nicht voneinander, sondern nur in Bezug auf das Dateiformat. So liegen die Daten in der Datei *Routinedaten.sav* im SPSS-Format und in der Datei *Routinedaten.xls* im Excelformat vor.
- 2.1) In SPSS stehen drei Fenster mit den folgenden spezifischen Funktionen zur Verfügung:
- Daten-Editor: Eingabe von Daten und Anlegen von Variablen; Daten- und Variablenansicht
  - Viewer: Dokumentation der durchgeführten Analyseschritte; Ausgabe der Ergebnisse der durchgeführten statistischen Prozeduren (z.B. Tabellen und Grafiken); Anzeigen von Fehlermeldungen
  - Syntax-Editor: Eingabe und Ausführen der Programmiersprache; Kommentieren von Analyseschritten.
- 2.2) Von allen drei Fenstern in SPSS ist es möglich, auf das allgemeine Menüfenster (Menüleiste) zuzugreifen.
- 2.3) Die Vorteile beim Arbeiten mit der Syntax bestehen in den folgenden Aspekten:
- Es besteht eine erhöhte Effizienz beim Aufbereiten und Auswerten der Daten.
  - Es gibt die Möglichkeit, die Syntax mehrmals zu verwenden und zu modifizieren.
  - Die Syntax ist ein Protokoll der Auswertungsschritte, wodurch die Analysen nachprüfbar und repliziert werden können.
  - Die Arbeit mit der Syntax ist eine gute Vorbereitung auf andere Analysesoftwaretypen, die sich ausschließlich über Programmiersprachen bedienen lassen.
  - Die Verwendung der Syntax eröffnet neue Möglichkeiten bezüglich spezifischer Prozeduren, die nicht über die Menüpunkte möglich sind.
- Die Nachteile beim Arbeiten mit der Syntax bestehen in den folgenden Aspekten:
- Der Aufwand für das Erlernen der Programmiersprache ist relativ hoch.
  - Die Anwenderfreundlichkeit ist gering.
  - Die Fehleranalyse in der Syntax ist aufwendig.

### Lösungen zu Kapitel 3

- 3.1) Ein Kodeplan ordnet den einzelnen Fragen des Fragebogens Variablennamen zu, damit diese eindeutig in der späteren Datendatei identifiziert werden können. Ein Kodeplan ordnet den Merkmalsausprägungen einer Variablen Kodenummern zu, damit diese innerhalb einer Variablen eindeutig gekennzeichnet sind und damit sie für die statistischen Auswertungsprozeduren von SPSS zugänglich sind.
- 3.2) Um die Daten der beiden Zentren zusammenzufügen, ist die Prozedur „Fallweise zusammenfügen“ auszuwählen. Dazu ist zunächst die Datei *Zentrum2.sav* mithilfe der Menüwahl *Datei* → *Öffnen* → *Daten* zu öffnen. Zum Anfügen der Datei *Zentrum3.sav* ist aus dem Menü folgende Aktion zu wählen: *Daten* → *Dateien zusammenfügen* → *Fälle hinzufügen*. Daraufhin öffnet sich die Dialogbox *Fälle hinzufügen zu*. Durch Klicken auf den Schalter *eine externe SPSS Statistics-Datendatei* wird das Feld *Durchsuchen* aktiviert und die Datei *Zentrum3.sav* kann ausgewählt werden. Anschließend muss noch auf *Öffnen* und *Weiter* geklickt werden. Es öffnet sich eine weitere Dialogbox *Fälle hinzufügen aus*, in der die Prozedur abschließend mit der Taste *OK* zur Ausführung zu bringen ist.
- 3.3) Für die variable *v14\_1\_rauch\_vor* ist für die Datenprüfung eine Häufigkeitstabelle zu erstellen, da es sich um eine kategoriale Variable mit ordinalem Messniveau handelt. Bei Betrachtung der Häufigkeitstabelle fällt auf, dass nur 28 Fälle von insgesamt 30 Fällen gültige Angaben bei der Variablen gemacht haben. Zudem gibt es zwei fehlerhafte Dateneingaben in Form der Ziffern 3 und 4.

Raucherstatus vor Reha

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	Nichtraucher	19	63,3	67,9	67,9
	Gelegenheitsraucher	4	13,3	14,3	82,1
	Raucher	3	10,0	10,7	92,9
	3	1	3,3	3,6	96,4
	4	1	3,3	3,6	100,0
	Gesamt	28	93,3	100,0	
Fehlend	System	2	6,7		
Gesamt		30	100,0		

Für die Variable *v12\_größe* sind für die Datenprüfung die Kennwerte Mittelwert, Modus, Median, Minimum und Maximum zu berechnen, da es sich um eine Variable mit Verhältnisniveau handelt. Bei Betrachtung der Kennwerte fällt auf, dass nur 27 von 30 Personen gültige Angaben gemacht haben. Zudem liegt der Wert für das Minimum mit 0,03 außerhalb der Menge an gültigen Werten für die Körpergröße in Metern, d.h. hier liegt vermutlich eine fehlerhafte Dateneingabe vor, die sich auch verzerrend auf den Mittelwert auswirkt.

### Statistiken

Körpergröße

N	Gültig	27
	Fehlend	3
Mittelwert		1,5756
Median		1,7500
Modus		1,83 <sup>a</sup>
Minimum		,03
Maximum		1,91
<sup>a</sup> Mehrere Modi vorhanden. Der kleinste Wert wird angezeigt.		

### Lösungen zu Kapitel 4

- 4.1) Für das Berechnen der Variablen „*bmi\_nach*“ ist in der Menüleiste folgende Prozedur auszuwählen: *Transformieren* → *Variable berechnen*. Daraufhin öffnet sich die Dialogbox *Variable berechnen*. Hier ist mithilfe der Transportschaltfläche die Variablen *v12\_größe* und *v13\_2\_gew\_nach* in den Konditional-Editor zu bringen und auf der Rechentastatur sind die entsprechenden Schalter zu betätigen, sodass am Ende der folgende Ausdruck im Konditional Editor steht:  $v13\_2\_gew\_nach / (v12\_größe * 2)$ . In das Feld *Zielvariable* ist *bmi\_nach* einzutragen. Abschließend ist die Prozedur mit der Taste *OK* zur Ausführung zu bringen. Für das Erstellen einer Häufigkeitstabelle für die neu erstellte Variable ist die SPSS-Prozedur *Analisieren* → *Deskriptive Statistiken* → *Häufigkeiten* entsprechend Kapitel 3.7.2 auszuführen.
- Hinweis zur Häufigkeitstabelle: Aus der Häufigkeitstabelle geht hervor, dass die neu erstellte Variable unplausible Werte enthält. Dies lässt sich damit erklären, dass bei der Variable *v12\_größe* (Körpergröße) bereits unplausible Werte auftreten (siehe Aufgabe 3.3).
- 4.2) Für das Berechnen der Variablen „*diff\_gew*“ ist in der Menüleiste folgende Prozedur auszuwählen: *Transformieren* → *Variable berechnen*. Daraufhin öffnet sich die Dialogbox *Variable berechnen*. Hier ist mithilfe der Transportschaltfläche die Variable *v13\_1\_gew\_vor* in den Konditional-Editor zu bringen. Anschließend ist auf der Rechentastatur das „-“, (minus) zu betätigen. Dann ist die Variable *v13\_2\_gew\_nach* in den Konditional-Editor zu bringen. Zum Ende muss der folgende Ausdruck im Konditional-Editor stehen:  $v13\_1\_gew\_vor - v13\_2\_gew\_nach$ . In das Feld *Zielvariable* ist *diff\_gew* einzutragen. Abschließend ist die Prozedur mit der Taste *OK* zur Ausführung zu bringen.
- Hinweis zur Häufigkeitstabelle: Aus der Häufigkeitstabelle geht hervor, dass die neu erstellte Variable unplausible Werte enthält. Dies lässt sich damit erklären, dass bei der Variable *v13\_1\_gew\_vor* (Gewicht vor Reha) unplausible Werte auftreten (siehe Abschnitt 3.7.3).

Lösungen zu Kapitel 5

- 4.3) Für das Berechnen der Variablen „*rauchen\_change*“, ist in der Menüleiste folgende Prozedur auszuwählen: *Transformieren* → *Variable berechnen*. Daraufhin öffnet sich die Dialogbox *Variable berechnen*. Hier ist mithilfe der Transportschaltfläche die Variable *v14\_1\_rauch\_vor* in den Konditional-Editor zu bringen. Anschließend ist auf der Rechentastatur das „-“, zu betätigen. Dann ist die Variable *v14\_2\_rauch\_nach* in den Konditional-Editor zu bringen. Zum Ende muss der folgende Ausdruck im Konditional-Editor stehen: *v14\_1\_rauch\_vor - v14\_2\_rauch\_nach*. In das Feld *Zielvariable* ist *rauchen\_change* einzutragen. Abschließend ist die Prozedur mit der Taste *OK* zur Ausführung zu bringen.

Hinweis zur Häufigkeitstabelle: Aus der Häufigkeitstabelle geht hervor, dass die neu erstellte Variable unplausible Werte enthält. Dies lässt sich damit erklären, dass bei der Variable *v14\_1\_rauch\_vor* bereits unplausible Werte auftreten (siehe Aufgabe 3.3).

- 5.1) Über das allgemeine Menü ist die Auswahl *Daten* → *Fälle auswählen* vorzunehmen, sodass sich die Dialogbox *Fälle auswählen* öffnet. In der Dialogbox ist dann in dem Auswahlbereich *Ausgabe* → *Nicht ausgewählte Fälle filtern* anzuklicken und die Einstellung *Falls Bedingung zutrifft* zu aktivieren. Anschließend ist die Schaltfläche *Falls* zu betätigen, sodass sich die Dialogbox *Fälle auswählen: Falls* öffnet. Die Variable *v17\_angehörige* ist mithilfe der Transportschaltfläche aus der Quellvariablenliste in den Konditional-Editor zu bringen. Anschließend ist auf der Rechentastatur das „=“ und die „0“ zu aktivieren, sodass der folgende Ausdruck im Konditional-Editor steht: *v17\_angehörige = 0*. Abschließend ist in der Dialogbox *Fälle auswählen: Falls* auf *Weiter* zu klicken, sodass sich wieder die Dialogbox *Fälle auswählen* öffnet, in der die Einstellungen mit *OK* zu bestätigen sind.

Für das Erstellen einer Häufigkeitstabelle für die neu erstellte Variable ist die SPSS-Prozedur *Analysieren* → *Deskriptive Statistiken* → *Häufigkeiten* entsprechend Kapitel 3.7.2 auszuführen.

Bei 25 Patienten wurden keine Angehörigen in die Behandlung einbezogen.

- 5.2) Über das allgemeine Menü ist die Auswahl *Daten* → *Fälle auswählen* vorzunehmen, sodass sich die Dialogbox *Fälle auswählen* öffnet. In der Dialogbox ist dann in dem Auswahlbereich *Ausgabe* → *Ausgewählte Fälle in neues Datenblatt kopieren* und in dem Feld ist der Dateiname „*Patienten\_älter60*“ einzugeben. Dann ist die Einstellung *Falls Bedingung zutrifft* zu aktivieren. Anschließend ist die Schaltfläche *Falls* zu betätigen, sodass sich die Dialogbox *Fälle auswählen: Falls* öffnet. Die Variable *v11\_alter* ist mithilfe der Transportschaltfläche aus der Quellvariablenliste in den Konditional-Editor zu bringen. Anschließend ist auf der Rechentastatur das „>2“ und die „60“ zu aktivieren, sodass der folgende Ausdruck im Konditional-Editor steht: *v11\_alter > 60*. Abschließend ist auf *Weiter* zu klicken, sodass sich wieder die Dialogbox *Fälle auswählen* öffnet, in der die Einstellungen mit *OK* zu bestätigen sind. Daraufhin öffnet sich eine neue Datendatei, in der nur die Fälle > 60 Jahre enthalten sind. Wenn Sie diese Auswahl behalten wollen, müssen Sie den Datensatz nun entsprechend abspeichern.

Für das Erstellen einer Häufigkeitstabelle für die neu erstellte Variable ist die SPSS-Prozedur *Analysieren* → *Deskriptive Statistiken* → *Häufigkeiten* entsprechend Kapitel 3.7.2 auszuführen.

Der neue Datensatz enthält 5 Personen.

- 5.3) Im Menü ist die folgende Auswahl zu treffen: *Transformieren* → *Zufallszahlengeneratoren*. Daraufhin öffnet sich die Dialogbox *Zufallszahlengenerator*. In dem Auswahlkasten *Aktiver Generator* ist die Checkbox *Aktiven Generator festlegen* zu aktivieren und der *Mersenne Twister* auszuwählen. Zudem ist im Auswahlkasten *Initialisierung des aktiven Generators* die Checkbox *Anfangswert festlegen* der Wert 5000 einzutragen. Die Einstellungen sind mit *OK* zu bestätigen. Anschließend ist im allgemeinen Menü die Auswahl *Daten* → *Fälle auswählen* vorzunehmen, sodass sich die Dialogbox *Fälle auswählen* öffnet. In der Dialogbox *Fälle auswählen* ist die Einstellung *Zufallsstichprobe* vorzunehmen. Anschließend ist auf den Schalter *Stichprobe* zu klicken, sodass sich die Dialogbox *Fälle auswählen: Zufallsstichprobe* öffnet. Um genau 8 Fälle aus der Datendatei zufällig zu ziehen, ist die Einstellung *Exakt* zu aktivieren und rechts daneben in das Feld die Zahl 8 einzutragen. In das zweite Feld *Fälle aus den ersten ... Fällen* ist für das Anwendungsbeispiel die Anzahl der Fälle der Datei einzugeben, d. h. die Zahl 30. Die vorgenommenen Einstellungen sind mit der Taste *Weiter* zu bestätigen, sodass man wieder in die Dialogbox *Fälle auswählen* zurückkommt. In dem Auswahlbereich *Ausgabe* ist die Einstellung *Ausgewählte Fälle in neues Datenblatt kopieren* zu aktivieren und als Name für die neue Datei „*Zufallsstichprobe\_8Patienten.sav*“ anzugeben. Abschließend sind die vorgenommenen Einstellungen mit der Taste *OK* zu bestätigen. Wenn Sie diese Auswahl behalten wollen, müssen Sie den Datensatz nun entsprechend abspeichern.

## Literaturverzeichnis

- Bühl, A. (2010). *SPSS 18. Einführung in die moderne Datenanalyse* (12. Auflage). München: Pearson Studium.
- Gesellschaft für Qualität im Gesundheitswesen (o.J.). *ZUF-8. Fragebogen zur Patientenzufriedenheit*. URL: <https://www.gfqg.de/assessment/zuf-8.html> [Stand: 09.01.2019].
- Schmidt, J. & Nübling, R. (2002). ZUF-8. Fragebogen zur Messung der Patientenzufriedenheit. In E. Brähler, J. Schumacher & B. Strauß (Hrsg.), *Diagnostische Verfahren in der Psychotherapie* (S. 392-396). Göttingen: Hogrefe.
- Schmidt, J., Lamprecht, F. & Wittmann, W. W. (1989). Zufriedenheit mit der stationären Versorgung. Entwicklung eines Fragebogens und erste Validitätsuntersuchungen. *Psychother. med. Psychol.*, 39, 248–255.
- Schmidt, J.; Nübling, R.; Lamprecht, F.; Wittmann, W. W. (1994). Patientenzufriedenheit am Ende psychosomatischer Reha-Behandlungen. In F. Lamprecht & R. Johnen (Hrsg.), *Salutogenese. Ein neues Konzept in der Psychosomatik?* (S. 271-283). Frankfurt/Main: VAS.
- Zöfel, P. (2002). *SPSS-Syntax: Die ideale Ergänzung für effiziente Datenanalyse*. München: Pearson Studium.
- Zöfel, P. & Bühl, A. (2000). *Statistik verstehen: Ein Begleitbuch zur computerunterstützten Anwendung*. München: Addison-Wesley.