

Dr. German Angele

# SPSS Statistics 20

(IBM® SPSS® Statistics 20)

## Eine Einführung



Not everything that can be counted counts, and not everything that counts can be counted. Albert Einstein

1. Auflage  
Februar 2012

Rechenzentrum der  
Otto-Friedrich-Universität Bamberg



## Vorwort

Bei der vorliegenden Schrift handelt es sich um eine Einführung zu IBM® SPSS® Statistics, die aus zahlreichen Veranstaltungen zum Programmpaket SPSS hervorgegangen ist. Der Einfachheit halber wird das Programm bei uns SPSS oder SPSS Statistics genannt. Es werden hier wichtige SPSS-Befehle und ihre Entsprechung im Menüsystem vorgestellt. Über den vielen, schönen Menüs gerät die Möglichkeit, SPSS für Windows direkt Befehle zu vermitteln, häufig in Vergessenheit. Andererseits ist es meines Erachtens für umfangreichere Analysearbeiten unerlässlich, sich mit SPSS-Befehlen zu beschäftigen. Zur praktischen Umsetzung anspruchsvoller datenanalytischer Problemstellungen ist fast immer das nötig, was wir früher als SPSS-Programm bezeichnet haben, also eine Abfolge von SPSS-Befehlen. Ganz davon abgesehen, dass manche Möglichkeiten von SPSS über die Menüs überhaupt nicht zugänglich sind und den Einsatz der Befehle erzwingen.

Es ist darauf hinzuweisen, dass diese Einführung kein Ersatz für ein SPSS Statistics-Lehrbuch ist. Sie wurde als begleitendes Lehrmaterial zur Vorlesung „SPSS für Windows“ konzipiert.

Diese Einführungsschrift könnte ohne Unterstützung durch meine, im Zeitablauf wechselnden, studentischen Mitarbeiterinnen und Mitarbeiter nicht erscheinen. Nina Baur, Maren Moisl, Oliver Herweg, Olga Lutterbach, Agnes Huber, Martina Kohlhuber, Lisa Kreuzer, Karolina Piber, Charlotte Kellermann, Simone Zdrojewski, Jasmin Eichelsdorfer, Annamaria Pfeffer und Eva Hank haben durch äußerst konstruktive Mitarbeit die Qualität der Arbeit positiv beeinflusst.

Diese erste Auflage des Skripts für SPSS Statistics 20 verwendet für die Übungsbeispiele wieder „Allbus-Daten“. Die letzte Auflage wurde noch für die Version 19 angepasst, SPSS Statistics-Fenster wurden aktualisiert und, wie immer, Unstimmigkeiten korrigiert. Eva Hank hat diese (nicht einfache) Aufgabe übernommen und dafür gesorgt, dass das Skript, wie ich meine, erneut an Qualität gewonnen hat. Die aktuelle Auflage ist, bis auf kleinere, notwendige Berichtigungen, unverändert.

Wir haben uns bemüht, männliche und weibliche Begriffe gleichwertig zu verwenden. Falls dies an einzelnen Stellen noch nicht konsequent umgesetzt ist, bitten wir um Entschuldigung.

Das Skript, alle „Folien“ der Veranstaltung, ein Verweis zu den Übungsdaten und Übungsaufgaben sowie weitere Materialien werden auch im WWW zum Selbststudium bereitgestellt. Hier sind auch Informationen zur Einführungsveranstaltung (Termine, Räume, etc.) zu finden.

**<http://www.uni-bamberg.de/rz/lehre/angele/spss/>**

Zum Nachdenken über eine wohl allgemein vorherrschende „Zahlengläubigkeit“ sollen das Zitat auf dem Titelblatt und die Verse auf dieser Seite anregen.

*Wenn nicht mehr Zahlen und Figuren  
Sind Schlüssel aller Kreaturen,  
Wenn die, so singen oder küssen,  
Mehr als die Tiefgelehrten wissen,  
Wenn sich die Welt ins freie Leben  
Und in die Welt wird zurückbegeben,  
Wenn dann sich wieder Licht und Schatten  
Zu echter Klarheit wieder gatten,  
Und man in Märchen und Gedichten  
Erkennt die ew'gen Weltgeschichten,  
Dann fliegt vor einem geheimen Wort  
Das ganze verkehrte Wesen fort.*

Novalis, „Heinrich von Ofterdingen“

German Angele  
Februar, 2012

# Gliederung

<b>1</b>	<b>VORBEMERKUNGEN ZU SPSS-STATISTICS</b> .....	<b>5</b>
<b>2</b>	<b>GRUNDLAGEN</b> .....	<b>7</b>
2.1	ARBEITEN MIT DER WINDOWS-OBERFLÄCHE VON SPSS STATISTICS.....	7
2.1.1	<i>Aufruf von SPSS Statistics</i> .....	7
2.1.2	<i>Fenster in SPSS Statistics</i> .....	7
2.1.2.1	Daten-Editor .....	8
2.1.2.2	Viewer .....	9
2.1.2.3	Syntax-Editor .....	9
2.1.3	<i>Speichern von Daten, Syntax-Editor und Viewer-Inhalten</i> .....	10
2.1.4	<i>Laden von Daten, Syntax-Editor und Viewer-Inhalten</i> .....	10
2.1.5	<i>Anmerkung zu Inhalten der Fenster</i> .....	11
2.1.6	<i>Die Hilfestellungen</i> .....	11
2.1.6.1	Weiterführende Literatur und Handbücher .....	11
2.1.6.2	Hilfsfunktionen im Programm.....	12
<b>3</b>	<b>DER BEISPIELDATENSATZ</b> .....	<b>15</b>
3.1	DATENERHEBUNG.....	15
3.2	VERKODUNG DES FRAGEBOGENS.....	16
3.3	EINFACHE STRUKTUR EINES DATENSATZES.....	17
<b>4</b>	<b>DATENDEFINITION</b> .....	<b>18</b>
4.1	DATENQUELLEN FÜR SPSS STATISTICS .....	18
4.2	DATENEINGABE DIREKT IN SPSS STATISTICS .....	19
4.2.1	<i>Variablen definieren in der Variablenansicht</i> .....	19
4.2.2	<i>Spalte Variablentyp (Typ)</i> .....	19
4.2.3	<i>Variablen-Labels - Werte-Labels - Fehlende Werte</i> .....	20
4.2.4	<i>Variablenbeschreibung über Befehle im Syntax-Editor</i> .....	21
4.2.4.1	Variablen-Labels .....	22
4.2.4.2	Werte-Labels .....	22
4.2.4.3	Fehlende Werte (missing values) .....	23
4.3	DEFINITION VON ASCII-TEXT-EINGABEDATEN .....	25
<b>5</b>	<b>ANALYSEN IN SPSS STATISTICS (PROCEDURES)</b> .....	<b>26</b>
5.1	DER WEG ZUR ANALYSE .....	26
5.1.1	<i>Analyse und Modifikationen über Menüs</i> .....	26
5.1.2	<i>Analyse und Modifikationen über Befehle</i> .....	30
5.1.3	<i>Befehle oder Menüs – eine Frage des Geschmacks?</i> .....	31
5.2	AUFBAU UND SYNTAX DER BEFEHLE .....	31
5.2.1	<i>Namen (Names)</i> .....	32
5.2.2	<i>Schlüsselworte (Keywords)</i> .....	32
5.2.3	<i>Zahlen und Zeichenketten (Numbers and Strings)</i> .....	32
5.2.4	<i>Arithmetische Operatoren und Begrenzungszeichen (Arithmetic Operators and Delimiters)</i> .....	33
5.3	DARSTELLUNG VON BEFEHLEN IN DIESER EINFÜHRUNGSSCHRIFT.....	34
5.4	PROZEDUREN IM ÜBERBLICK.....	35
5.5	WICHTIGE PROZEDUREN.....	36
5.5.1	<i>LIST</i> .....	36
5.5.2	<i>Häufigkeiten (FREQUENCIES)</i> .....	37

5.5.3	<i>Kreuztabellen (CROSSTABS)</i> .....	39
5.5.4	<i>Deskriptive Statistik (DESCRIPTIVES)</i> .....	41
5.5.5	<i>Korrelationen (CORRELATIONS)</i> .....	43
5.5.6	<i>Mittelwertvergleiche (MEANS)</i> .....	44
<b>6</b>	<b>DATENMODIFIKATIONEN</b> .....	<b>47</b>
6.1	MODIFIKATION NUMERISCHER VARIABLEN .....	47
6.1.1	<i>Rekodierung (RECODE)</i> .....	47
6.1.2	<i>Das COMPUTE-Kommando</i> .....	51
6.1.2.1	Die Befehlsstruktur .....	51
6.1.2.2	Variablenamen und Konstanten .....	51
6.1.2.3	Arithmetische Operatoren .....	51
6.1.2.4	Funktionen .....	52
6.1.2.5	Missing Values bei arithmetischen Ausdrücken.....	53
6.1.2.6	Missing Values bei Funktionen.....	54
6.1.2.7	Berechnen über Menü .....	54
6.1.3	<i>Das COUNT-Kommando</i> .....	54
6.1.4	<i>TEMPORARY</i> .....	56
6.1.5	<i>Das WEIGHT-Kommando: Fälle gewichten</i> .....	57
6.2	MODIFIKATIONEN IN ABHÄNGIGKEIT VON BEDINGUNGEN.....	59
6.2.1	<i>Einfaches IF und logische Ausdrücke</i> .....	59
6.2.2	<i>Erweiterung des IF-Kommandos - DO IF/END IF</i> .....	63
6.3	ABSPEICHERN VON MODIFIKATIONSERGEBNISSEN (SAVE).....	66
<b>7</b>	<b>DATENSELEKTIONEN</b> .....	<b>68</b>
7.1	FILTER .....	68
7.2	SELECT IF.....	70
7.3	DIE SORTIERUNG VON FÄLLEN MIT SORT CASES .....	73
7.4	DAS SPLIT FILE-KOMMANDO .....	73
<b>8</b>	<b>GRAFIK IN SPSS STATISTICS FÜR WINDOWS</b> .....	<b>76</b>
8.1	ARBEITEN MIT DIAGRAMMEN .....	76
8.2	EINFACHES BALKENDIAGRAMM.....	76
8.3	HISTOGRAMM MIT NORMALVERTEILUNGSKURVE.....	79
8.4	STREUDIAGRAMM (SCATTERPLOT) MIT REGRESSIONSGERADE .....	81
8.5	TRANSFER VON GRAFIKEN UND ANALYSEERGEBNISSEN.....	81
8.5.1	<i>Kopieren von Ergebnissen oder Befehlen in ein Textverarbeitungsprogramm</i> .....	81
8.5.2	<i>Umwandlung von Ergebnissen in .htm/.jpg/.txt Dateien</i> .....	82
<b>9</b>	<b>ANHANG</b> .....	<b>83</b>
9.1	WICHTIGE VARIABLEN DES BEISPIELDATENSATZES .....	83
9.2	WICHTIGE SPSS STATISTICS BEFEHLE IM ÜBERBLICK.....	84
9.3	WEITERE WICHTIGE PROZEDUREN .....	85
9.4	ÜBERSICHT: VOR- UND NACHTEILE DER DATENSELEKTIONSVERFAHREN .....	85
9.5	BIVARIATE ASSOZIATIONSMAßE .....	86
<b>10</b>	<b>LITERATURVERZEICHNIS</b> .....	<b>87</b>
<b>11</b>	<b>INDEX</b> .....	<b>89</b>

# 1 Vorbemerkungen zu SPSS-Statistics

Dieses Skript beschäftigt sich ausführlich mit dem Datenanalysesystem IBM SPSS Statistics (kurzzeitig im Jahr 2009 PASW Statistics genannt!). Warum gerade SPSS Statistics und nicht irgendein anderes Statistikpaket? SPSS-Statistics ist wohl das Statistiksystem mit der weltweit größten Verbreitung. Die vielen Auswertungen, die zu Forschungszwecken in den Sozialwissenschaften und anderen Disziplinen mit SPSS Statistics durchgeführt werden, belegen die allgemeine Akzeptanz des Systems.<sup>1</sup> Bei einem ersten Kontakt mit Datenanalysesystemen ist es sinnvoll, einem weit verbreiteten und oft verwendeten universellen System den Vorzug vor anderen Programmen zu geben. Dies versetzt die Leserinnen und Leser in die Lage, eine Vielzahl von Auswertungen durchzuführen, die in der empirischen Praxis gebräuchlich sind. Trotzdem könnte mit einer ähnlichen Argumentation auch noch ein Produkt wie STATA, BMDP, SAS oder P-STAT in Frage kommen.<sup>2</sup> Den Ausschlag für SPSS Statistics gibt letztlich auch die langjährige Erfahrung des Verfassers mit dem Programm.

Seit der Veröffentlichung der ersten Programmbeschreibung zu SPSS im Jahr 1970 hat das System einen großen Aufschwung erfahren. Es liegt in zahlreichen Versionen für die unterschiedlichsten Rechner und Betriebssystemvarianten vor. Kurze Zeit hieß das Programm nicht mehr SPSS sondern PASW-Statistics, wobei PASW für „Predictive Analytics Software“ stand. Seit die Firma SPSS Inc., im Oktober 2009 vollständig von IBM übernommen wurde, heißt das Produkt IBM SPSS Statistics. Der alte Name ist also (fast) wieder da. Wir verwenden aber weiterhin überwiegend den Namen SPSS Statistics, der dann immer für „IBM SPSS Statistics“ steht. Die Firma SPSS ist seit 01.10.2010 komplett in IBM eingegliedert.

Wer setzt nun solch ein Programmsystem ein? Anwender sind alle Institutionen, die sich mit der Auswertung und Darstellung von Daten befassen. Es ist daher nicht verwunderlich, dass mit SPSS, außer an Universitäten, auch in der privaten Wirtschaft (zum Beispiel AUDI, Henkel, Karstadt, Otto-Versand) und den öffentlichen Verwaltungen (zum Beispiel Bundesagentur für Arbeit, Bundeskriminalamt) gearbeitet wird.

Der universitäre Einsatz des Analysesystems erstreckt sich auf viele Disziplinen:

Soziologie, Politikwissenschaft, Psychologie, Betriebswirtschaftslehre, Volkswirtschaftslehre, Biologie, Medizin, Geographie, Geschichtswissenschaften, Pädagogik usw.

Einige Beispiele aus der Praxis sollen demonstrieren, wo bereits erfolgreich mit SPSS gearbeitet wurde:

**Bibliothekswesen:**<sup>3</sup> Buchausstellungen werden als wichtiger Teil der Öffentlichkeitsarbeit von Bibliotheken in der bibliothekswissenschaftlichen Theorie allgemein positiv beurteilt. Eine Befragung von 95 wissenschaftlichen Universalbibliotheken in Deutschland sollte Aufschluss darüber geben, wie es um die Ausstellungsaktivitäten in der Praxis bestellt ist. Die Ergebnisse der Umfrage wurden mittels des SPSS-Programms analysiert und tabellarisch dargestellt. Es zeigt sich, dass wissenschaftliche Universalbibliotheken, teilweise unter hohem Aufwand, ein reichhaltiges Spektrum von Ausstellungsleistungen erbringen. Ausstellungen werden als wirksames Mittel gesehen, die Bibliothek als Kulturträger und -vermittler einer breiteren Öffentlichkeit ins Bewusstsein zu bringen.

**Volkskunde:**<sup>4</sup> Eine interessante Studie liegt zur Veränderung der Volkskultur durch Migration in das Armenviertel Lucero von Bogota vor. In der Hauptstadt Kolumbiens (Bogota) wurden Bewohner des Stadtviertels Lucero, die aus zwei anliegenden Departments zugezogen sind, befragt, wie sich ihr Leben durch den Umzug im Hinblick auf Volkskultur (Kleidung, Tracht, Bräuche, Feste usw.) und Lebensqualität (Arbeitsplatz, Essen, Trinken, Wohnung, Finanzen) verändert habe. Eine Analyse der Daten mit SPSS belegt, dass sich die Volkskultur durch die Migration signifikant verringert, die Lebens-

---

<sup>1</sup> So schreibt Wilke (1986), S. 3: "Man übertreibt sicher nicht, wenn man SPSS (das meistgenutzte System) zum größten Einflussfaktor auf die Forschungspraxis der letzten 10 Jahre erklärt.", und Matiaske (1990), S. 21 führt aus: "Im Bereich der Sozialwissenschaften ist SPSS zu einer Art Standard geworden."

<sup>2</sup> Es muss betont werden, dass noch eine Vielzahl anderer Programme im praktischen Einsatz sind, wie zum Beispiel: LISREL, MSP, PML und eine Reihe von Eigenentwicklungen. So beschreiben Woodward/Elliott/Gray/Matlock (1988), in ihrem "Dictionary of Statistical Microcomputer Software" bereits über 200 Statistikprogramme alleine für Personal Computer. Einen Überblick zu vielen Statistikprogramm Paketen gibt auch Angele (1989a).

<sup>3</sup> Siehe zu diesem Beispiel Unger (1986).

<sup>4</sup> Siehe zu diesem Beispiel Ulmer (1989).

qualität jedoch deutlich zunimmt. Um die Landflucht von staatlicher Seite zu bremsen, wäre die Lebensqualität im ländlichen Bereich durch geeignete Maßnahmen zu verbessern.

**Betriebswirtschaftslehre:**<sup>5</sup> Kaufleute müssen zum Schluss eines jeden Geschäftsjahres ein Inventar aufstellen, in dem sie Vermögensgegenstände und Schulden verzeichnen (§ 240 HGB). Nach § 241 Abs. 1 HGB darf der Bestand der Vermögensgegenstände auch auf Grund von Stichproben ermittelt werden. Mit Hilfe von SPSS ist es möglich, auf der Basis entsprechender Daten, die notwendigen Vorarbeiten zu leisten, Stichproben zu ziehen und nach der Erhebung der Vermögensgegenstände eine Schätzung des Totalwerts eines Vermögenskollektivs durchzuführen. Dabei sind je nach vorhandenen Vorinformationen unterschiedliche Schätzverfahren einzusetzen.

Diese wenigen Beispiele verdeutlichen das breite Einsatzspektrum von SPSS Statistics.

Die aktuelle Version des Programms ist *SPSS-Statistics 19*. In dieser Einführungsschrift werden daher grundlegende Kommandos von SPSS-Statistics 19 für Windows vermittelt. Die Kommandos sind auch auf SPSS Statistics unter anderen Betriebssystemen anwendbar.

SPSS Statistics ist kein monolithisches Programmpaket, sondern besteht aus dem Modul "*SPSS Statistics Base 19*" und diversen Zusatzprodukten (Module).

Bei der vorliegenden Schrift handelt es sich um eine Einführung in SPSS Statistics. Da nicht alle Leserinnen und Leser über die gleichen Voraussetzungen im Umgang mit SPSS Statistics verfügen, wird hier eine Orientierung für die verschiedenen Lesergruppen angeboten.

Zielgruppe	Empfohlene Passagen
Geringe oder keine Grundkenntnisse im Umgang mit Windows und SPSS Statistics	Einführend mit Kapitel 2 beginnen
Windowskenntnisse vorhanden	Zur Orientierung Kapitel 3 lesen, Vertiefung ab Kapitel 4
Grundkenntnisse in SPSS-Statistics vorhanden	Kapitel 5, 6, 7
Fundiertes Wissen im Umgang mit SPSS-Statistics	Gezielte Orientierung am Inhaltsverzeichnis oder Index bzw. Sekundärliteratur

<sup>5</sup> Siehe zu diesem Beispiel Angele (1989).

## 2 Grundlagen

### 2.1 Arbeiten mit der Windows-Oberfläche von SPSS Statistics

Dieses Kapitel ist für Anwender, die noch wenig Erfahrung mit Microsoft® Windows® Systemen und den grundlegenden Funktionen von SPSS Statistics haben. Der Aufruf von SPSS kann je nach Installation variieren. Folgendes sollte (fast) immer funktionieren:

#### 2.1.1 Aufruf von SPSS Statistics


Über die Windowsschaltfläche  (START) in der Taskleiste

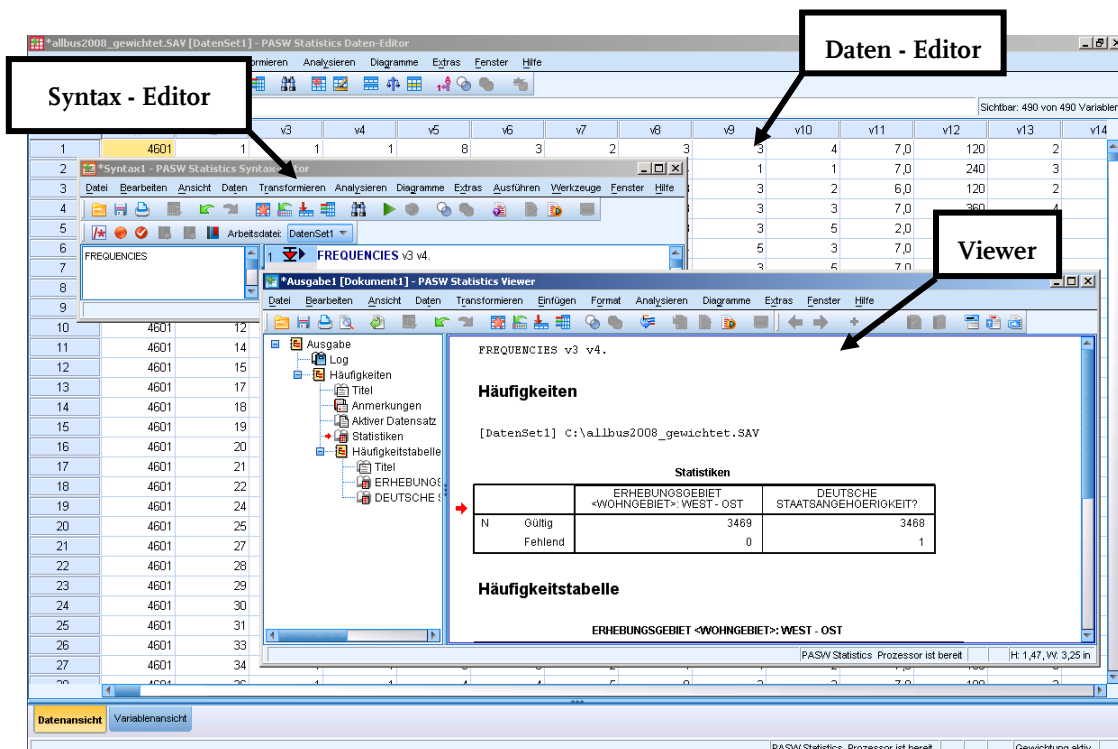
ALLE PROGRAMME ⇒ IBM SPSS Statistics ⇒  IBM SPSS Statistics 20

anklicken und warten, bis SPSS Statistics geladen wird. Im ersten Fenster, das dann erscheint, wird über die weitere Vorgehensweise entschieden. Die Arbeit mit SPSS Statistics kann beginnen.

#### 2.1.2 Fenster in SPSS Statistics

In SPSS Statistics wird in der Regel mit drei Fenstern gearbeitet: *Daten-*, *Syntax-Editor* und *Viewer*:

- *Daten-Editor* (erscheint immer bei Aufruf des Programms): Hier wird in der Datenansicht der Datensatz angezeigt und in der Variablenansicht die verschiedenen Variablenattribute. Beim Speichern von SPSS Statistics-Daten wird die Dateiendung *.sav* vorgegeben.
- *Viewer* (erscheint automatisch nach einer Analyse): Hier werden Ergebnisse angezeigt. Dieses Fenster wird automatisch geöffnet, wenn eine Prozedur ausgeführt wird, die eine Ausgabe erzeugt. *Viewer*-Dateien besitzen die Endung *.spv*.
- *Syntax-Editor* (Aufruf über Menü DATEI ⇒ NEU ⇒ SYNTAX): Hier werden Befehle eingegeben, die SPSS Statistics mitteilen, wie die Daten weiterverarbeitet werden sollen. Die Befehle werden ausgeführt, indem man die eingegebenen Befehle markiert und in der Icon-Leiste das Symbol  anklickt oder gleichzeitig die Tasten **Strg** und **R** antippt. Die Endung von Dateien mit SPSS Statistics-Befehlen lautet *.sps*.



Jeweils das Fenster, dessen Name farbig (in der Standardeinstellung ist dies blau) unterlegt ist, ist **aktiv**, das heißt, man arbeitet gerade in oder mit diesem Fenster. Es gibt drei Möglichkeiten, zwischen SPSS-Fenstern zu wechseln:

- Anklicken des Fensters, mit dem man arbeiten möchte, sofern es sichtbar ist.
- In der Menüleiste eines SPSS-Fensters **FENSTER** anklicken und aus der Liste dasjenige Fenster auswählen, mit dem gearbeitet werden soll.
- Mit Hilfe der Tastenkombination **Alt** und der **Tabulator**-Taste zwischen Fenstern wechseln.

In jedem dieser drei Fenster sind alle Menüpunkte **DATEI**, **BEARBEITEN**, **ANSICHT**, **DATEN**, **TRANSFORMIEREN**, **ANALYSIEREN**, **DIAGRAMME**, **EXTRAS**, **FENSTER** und **HILFE** verfügbar. In der Menüleiste des *Viewers* haben wir die zusätzlichen Menüpunkte **EINFÜGEN** und **FORMAT**. Im *Syntax-Editor* wird die Möglichkeit **AUSFÜHREN** und **WERKZEUGE** angeboten. Die Schaltflächen (Icons) sind den jeweiligen Funktionen eines Fensters angepasst.

### 2.1.2.1 Daten-Editor

Der *Daten-Editor* liefert in SPSS Statistics, neben der Datenansicht, in der Variablenansicht auch eine komplette Aufzählung aller Variablen und deren Eigenschaften.

	v3	v4	v5	v6	v7	v8	v9	v10	v11
1	1	1	8	3	2	3	3	4	7,0
2	1	1	8	4	3	4	1	1	7,0
3	1	1	6	4	3	3	3	2	6,0
4	1	1	5	3	3	3	3	3	7,0
5	1	1	8	3	2	3	3	5	2,0

Registerkarten

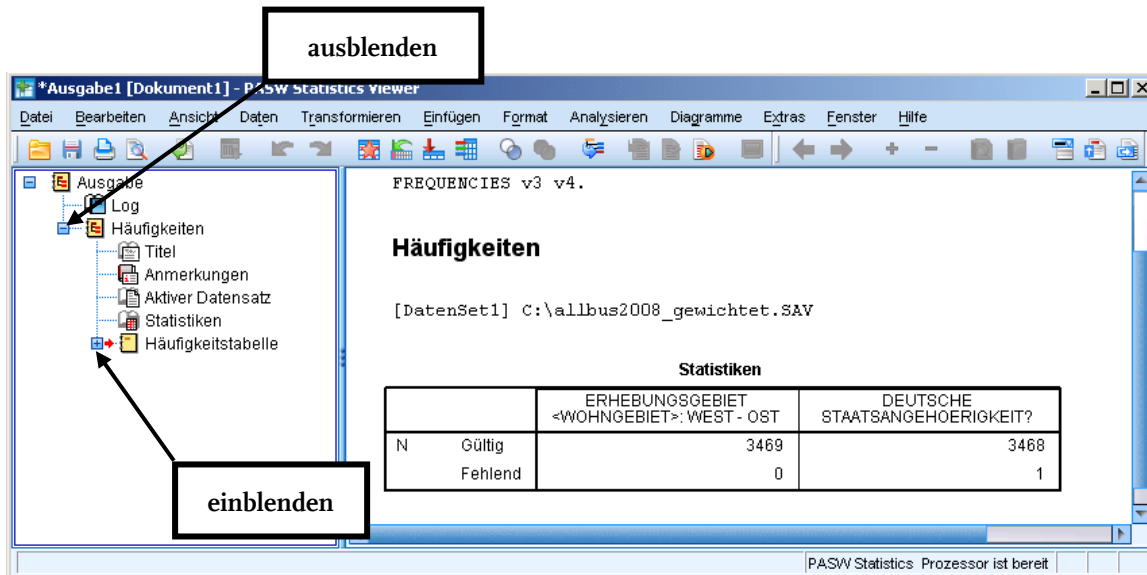
	Name	Typ	Spaltenf...	Dezimal...	Variablenlabel	Wertelabels	Fehlende W...	Spalten	Ausrichtung
1	v3	Numerisch	1	0	ERHEBUNGS...	{1, ALTE B...	Keine	8	Rechts
2	v4	Numerisch	1	0	DEUTSCHE ST...	{1, JA}...	7 - HI	8	Rechts
3	v5	Numerisch	2	0	ATTRAKTIVITA...	{1, UNATTR...	Keine	8	Rechts
4	v6	Numerisch	1	0	WIRTSCHAFT...	{1, SEHR G...	7 - HI	8	Rechts
5	v7	Numerisch	1	0	WIRTSCHAFT...	{1, SEHR G...	7 - HI	8	Rechts
6	v8	Numerisch	1	0	WIRTSCHAFT...	{1, WESEN...	7 - HI	8	Rechts

Der *Daten-Editor* besteht dazu aus zwei Ebenen. Der Benutzer kann über die Registerkarten **DATENANSICHT** und **VARIABLENANSICHT** zwischen diesen Ebenen wählen. In der Datenansicht wird der Datensatz mit den einzelnen Fällen und Variablen angezeigt (siehe Abschnitt 3.3, Seite 17). Die Variablenansicht ermöglicht das Betrachten und Definieren von Variablenattributen wie Datentyp sowie Variablen- und Wertelabels (siehe Abschnitt 4, Seite 18 ff.).



### 2.1.2.2 Viewer

Eine besondere Bedeutung bei der Arbeit mit SPSS Statistics besitzt der *Viewer*. Dort werden die Ergebnisse aller Berechnungen angezeigt, die ja das zentrale Anliegen der Arbeit mit SPSS sind.



Der *Viewer* zeigt in der linken Hälfte eine Gliederung der vorhandenen Ergebnisse. Einzelne Gliederungspunkte können durch Klick auf das Symbol  aus- und durch Klick auf das Symbol  eingeblendet werden. In der rechten Hälfte des *Viewers* sind die Ergebnisse in Tabellenform dargestellt. Jede Tabelle wird als Objekt bezeichnet, das sich wiederum aus einzelnen Bestandteilen (Objekten), wie Spalten, Zellen, Texten zusammensetzt. Durch Anklicken eines Gliederungspunktes erscheint im rechten Fenster unmittelbar das dazugehörige Objekt. Die Gliederung dient somit als Navigationshilfe für eine schnelle Orientierung.

Grundsätzlich ist es möglich, alle Objekte oder Tabellen im *Viewer* zu bearbeiten, wenn sie doppelt angeklickt werden. Dann können die Bestandteile eines Objekts den individuellen Bedürfnissen angepasst werden.


In Textfeldern kann zum Beispiel der Text daraufhin verändert werden. Innerhalb einer Tabelle ist hierbei zusätzlich das zu verändernde Textfeld doppelt anzuklicken. Daraufhin wird der Text innerhalb dieses Textfeldes bearbeitbar.

Zusätzlich zur Bearbeitung des Textes kann auch die Breite der Spalten und der einzelnen Textfelder verändert werden. Um die Breite der Spalten zu korrigieren, wird der Mauszeiger an eine der Begrenzungslinien geführt. Wenn er sich in einen Querpfedel verwandelt, kann, während die linke Maustaste gedrückt wird, die Breite der Spalte verändert werden. Soll dies auch mit einzelnen Textfeldern geschehen, so müssen diese zuvor durch einfaches Anklicken markiert werden, damit sie anschließend auf die gleiche Weise bearbeitet werden können.

### 2.1.2.3 Syntax-Editor

Im Syntax-Editor, der nicht automatisch startet, werden SPSS-Kommandos eingegeben und zur Ausführung gebracht. Der Syntax-Editor wird über DATEI ⇒ NEU ⇒ SYNTAX aufgerufen:



Befehle werden über  oder den Menüpunkt AUSFÜHREN (Tastenkombination Strg + R) aktiviert (Einzelheiten siehe auch Punkt 5.1.2, S. 30).

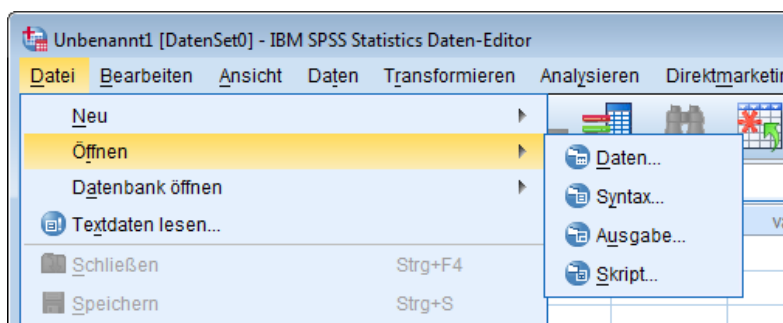
### 2.1.3 Speichern von Daten, Syntax-Editor und Viewer-Inhalten

Änderungen in den Fenstern (*Daten*, *Syntax-Editor*, *Viewer*) sollten öfters (und einzeln) abgespeichert werden, damit keine Inhalte während der Arbeit verloren gehen. Um dies zu tun, muss das gewünschte Fenster aktiviert und mit der Maus das Menü DATEI ⇒ SPEICHERN (Tastenkombination Strg+S) gewählt werden.

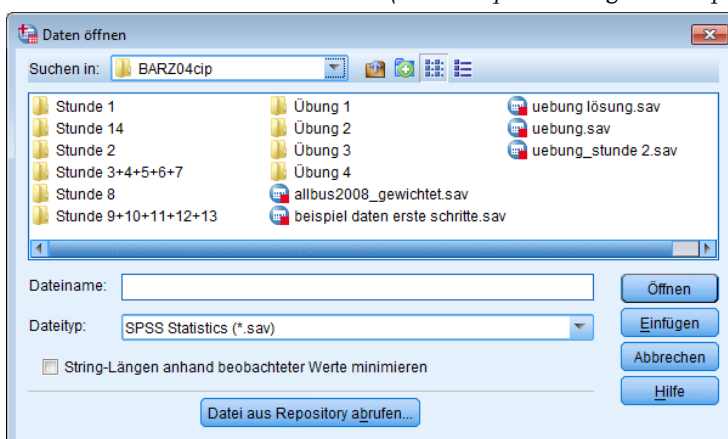
### 2.1.4 Laden von Daten, Syntax-Editor und Viewer-Inhalten

Wenn eine Datei mit Daten-Editor (Dateiname.sav), Syntax-Editor (Dateiname.sps) oder Viewer-Inhalten (Dateiname.spv) vorhanden ist, wird diese folgendermaßen in SPSS Statistics aufgerufen:

1. Mit der Maus in einem SPSS-Fenster (*Daten*, *Ausgabe*, *Syntax*) DATEI ⇒ ÖFFNEN wählen.
2. Den gewünschten Dateityp auswählen, *Daten...*, *Syntax...*, *Ausgabe...*, oder *Skript...*-Datei:



3. Es erscheint das Fenster *Datei (Daten, Syntax, Ausgabe, Skript) öffnen*:



Das richtige Laufwerk oder Verzeichnis auswählen: Durch Anklicken des kleinen Pfeils rechts unter der Überschrift *Datei öffnen*. Es erscheint eine Auflistung aller Laufwerke (Festplatte, CD, USB, ...). Das gewünschte Laufwerk ist dann auszuwählen.

Im Fenster *Datei öffnen* erscheinen nun alle Dateien des entsprechenden Typs (das heißt, alle Daten-Dateien, alle SPSS-Syntax-Dateien oder alle SPSS-Ausgabe-Dateien), die auf dem Datenträger abgespeichert sind. Den gewünschten Dateinamen markieren, dann **Öffnen** anklicken.

Seit der Version SPSS 14 können mehrere Datensätze innerhalb einer SPSS-Sitzung geöffnet werden. Hierzu werden die oben genannten Schritte so oft wiederholt, bis alle benötigten Datensätze geöffnet sind. Analysen beziehen sich, bei mehreren geöffneten Datensätzen, auf den so genannten „aktiven Datensatz“. Dies ist der Datensatz, dessen Fenster aktuell im Vordergrund steht. Alternativ kann im Syntax-Fenster über den Befehl

→ Hinweis!

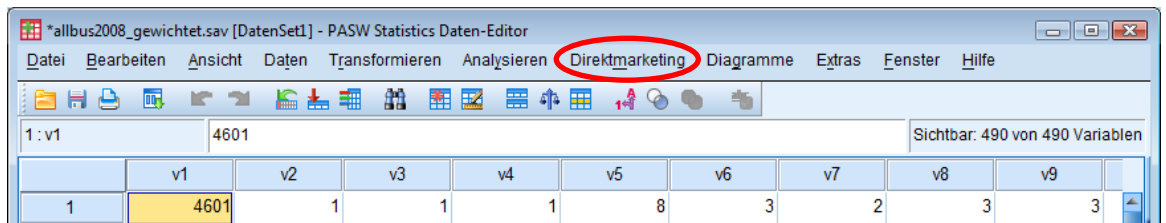
```
DATASET ACTIVATE    datensetname
```

→ Syntax

ein bestimmter Datensatz zum „aktiven Datensatz“ gemacht werden, wobei n für den Namen des „DatenSets“ in der Kopfzeile des Datensatzfensters steht (wie DatenSet1).

### 2.1.5 Anmerkung zu Inhalten der Fenster

SPSS kann mit verschiedenen Modulen lizenziert werden. Dies beeinflusst unter Umständen das Erscheinungsbild der SPSS-Fenster. Bei Lizenzierung des Moduls „SPSS Direct Marketing“ beispielsweise, besitzen Daten-Editor, Syntax-Editor und Viewerfenster einen zusätzlichen „Einstiegspunkt“:



### 2.1.6 Die Hilfestellungen

Grundsätzlich bietet das Programm SPSS Statistics zahlreiche Hilfsmöglichkeiten an. Auf diese wird im Kapitel 2.1.6.2 näher eingegangen. Weiterhin stehen dem Benutzer von SPSS eine Vielzahl von Handbüchern und begleitender Literatur zur Verfügung. Im folgenden Kapitel werden daher kurz die wichtigsten aufgeführt, um dem Leser einen Überblick zu liefern.

#### 2.1.6.1 Weiterführende Literatur und Handbücher

Das gesamte SPSS Statistics-System ist in Handbüchern dokumentiert. Grundlegende Informationen liefert das Buch "IBM SPSS Statistics 19 Core System User's Guide". Ergänzend dazu gibt es Beschreibungen zu verschiedenen Modulen des Systems, wie "Advanced Statistics", "Categories", "Forecasting", "Custom Tables". Das Referenzwerk zur Kommandosprache von SPSS Statistics (SPSS) ist der "IBM SPSS Statistics 19 Command Syntax Reference", der allerdings nicht als einführendes Werk gedacht ist. Allgemein sei in diesem Zusammenhang auf die aktuellen Versionen dieser Handbücher zum Programm verwiesen, die die Firma SPSS über WWW zur Verfügung stellt:

<http://support.spss.com/ProductsExt/Statistics/Documentation/19/clientindex.html>

Weiterführende Literatur (ohne jeden Anspruch auf Vollständigkeit):

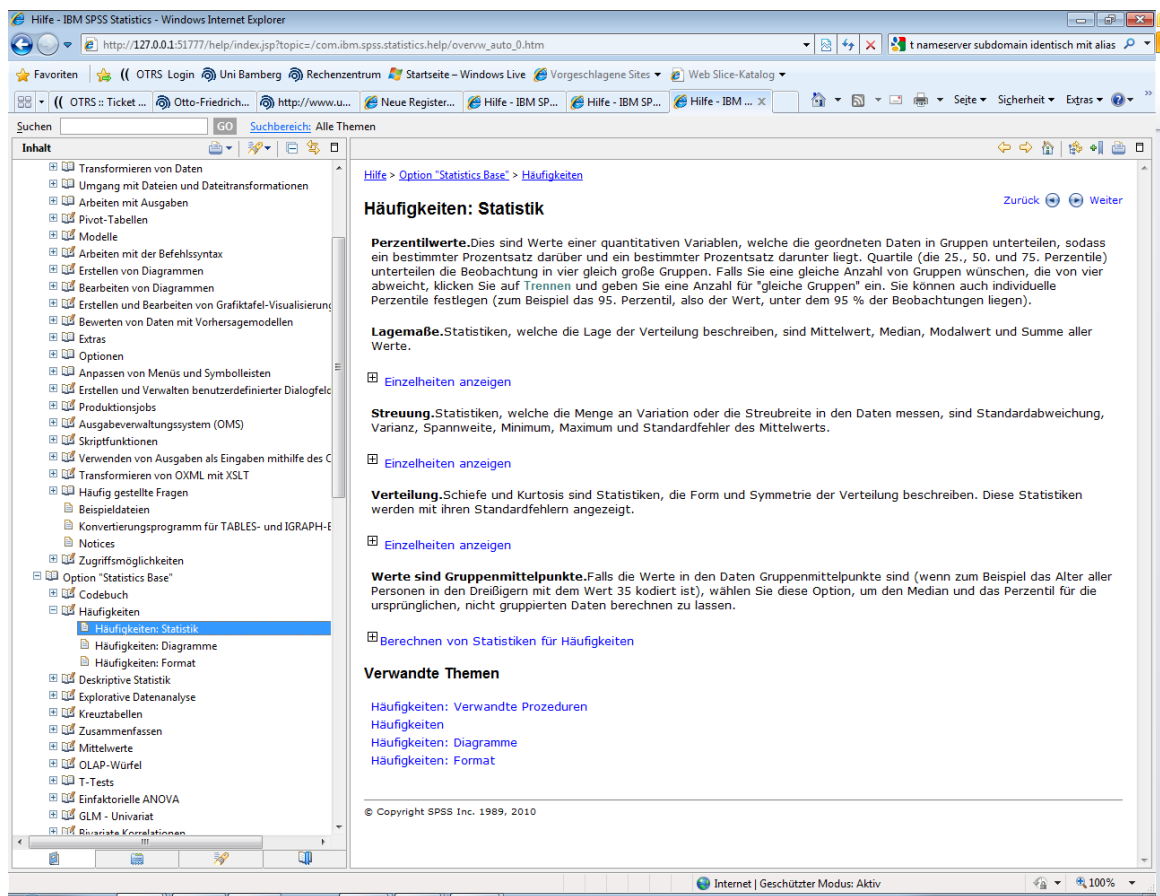
- Bühl, A., SPSS 20, Einführung in die moderne Datenanalyse, 13. Auflage, München 2012.
- Brosius, F., SPSS 20 für Dummies, Wiley-VCH Verlag, 2012.
- Brosius, F., SPSS 19, Verlagsgruppe Hütig Jehle Rehm, Heidelberg 2011.
- Akremi, L./Baur, N./Fromm, S., Datenanalyse mit SPSS für Fortgeschrittene, 3. Auflage, Wiesbaden 2011.
- Fromm, S., Datenanalyse mit SPSS für Fortgeschrittene 2: Multivariate Verfahren für Querschnittsdaten, Wiesbaden 2010.

- Backhaus, K./ Erichson, B./ Plinke, W./ Weiber, R. (Hg.), Multivariate Analysemethoden. Eine anwendungsorientierte Einführung, 13. Auflage, Berlin 2011.
- Janssen, J./Laatz, W., Statistische Datenanalyse mit SPSS für Windows, Berlin 2010.
- Sarstedt, M./ Schütz, T./ Raitchel, S., IBM SPSS Syntax - Eine anwendungsorientierte Einführung. 2. Auflage, Vahlen Verlag, München 2010.
- Hatzinger, R./Nagel, H., PASW Statistics: Statistische Methoden und Fallbeispiele, München 2009.
- Kinnear, P.R./Gray, C.D., PASW Statistics 17 Made Simple, 2009 Psychology Press.

### 2.1.6.2 Hilfsfunktionen im Programm

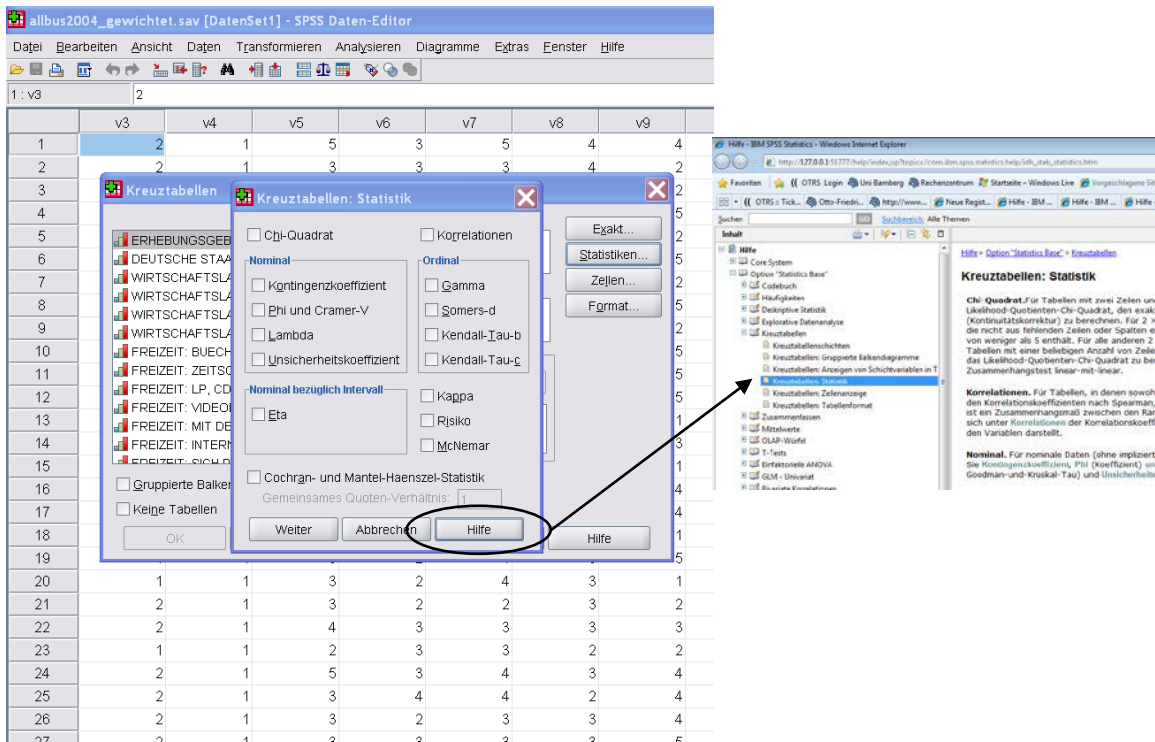
Wenn es Probleme bei der Arbeit mit SPSS Statistics gibt, so stellt das Programm eine umfangreiche Hilfefunktion zur Verfügung. Das Menü HILFE ⇒ THEMEN in der Menüleiste aktiviert diese Hilfefunktion.

Die Registerkarte **INHALT** zeigt eine Übersicht der wesentlichen Themenkomplexe an, wie in einem Inhaltsverzeichnis. So kann man inhaltlich gezielt nach einer Antwort auf das jeweilige Problem suchen.

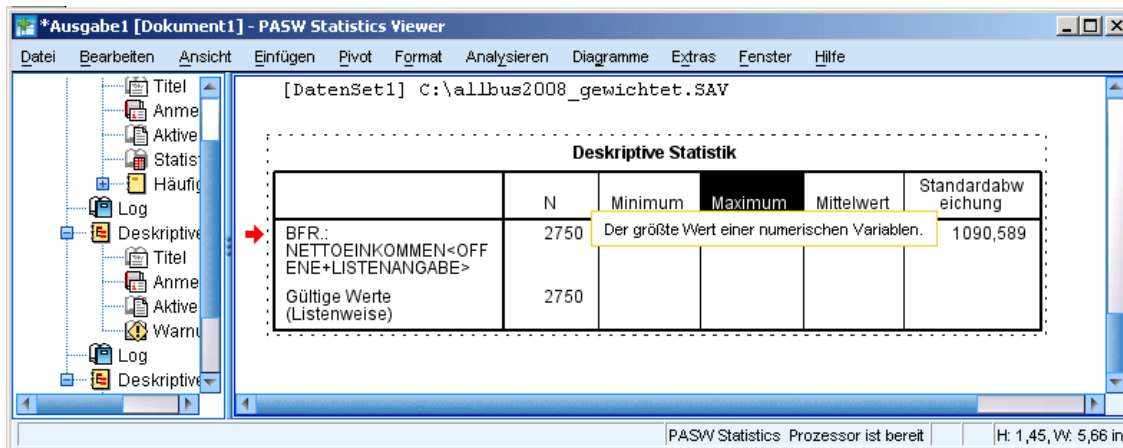


Eine Möglichkeit zur Volltextsuche bietet die Option **SUCHEN**. Hier kann in der ersten Zeile ein Begriff (als vollständiges Wort) eingegeben werden. SPSS Statistics durchsucht dann die Hilfedateien nach dem entsprechenden Begriff. Im mittleren Feld erhält man einige Wortentsprechungen, die durch Doppelklick nochmals eingeschränkt werden können. Im unteren Feld werden die Themen aufgelistet. Die Verwendung der Volltextsuche bietet sich jedoch nur an, wenn die Suche in den Registerkarten Inhalt und Index erfolgreich war.

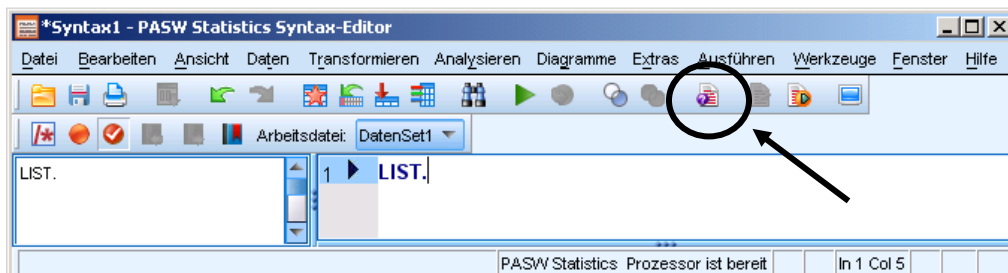
Führt man Prozeduren über die Menüs aus, so bietet SPSS Statistics weitere Informationsmöglichkeiten. In jedem Dialogfeld befindet sich ein Knopf mit der Aufschrift **HILFE**, mit dem eine kontextbezogene Information abgerufen werden kann.



Entsprechend kann auch im Viewer Hilfe für unbekannte Begriffe in Ausgabetabellen angefordert werden. Dazu wird zunächst auf die betreffende Tabelle zweimal geklickt (Doppelklick) und anschließend der unbekannte Begriff markiert. Durch einen Klick mit der rechten Maustaste und nach Auswahl des Menüpunktes **Direkthilfe** bekommt der Nutzer eine Erklärung des Begriffs (im Beispiel: Maximum – Der größte Wert einer numerischen Variablen).



Der *Syntax-Editor* bietet neben der normalen Menühilfe und den automatischen Befehlsergänzungen die Möglichkeit, über eine Schaltfläche Informationen zur Syntax einzelner Befehle abzurufen. Dazu muss der gewünschte Befehl markiert und die unten gezeigte Schaltfläche angeklickt werden (siehe Kapitel 5.2 Aufbau und Syntax der Befehle).



Daraufhin erscheint folgendes SPSS Statistics -Hilfenfenster, das Auskunft über die Struktur des markierten Befehls (Beispiel: LIST) gibt.

The screenshot shows a web browser window displaying the help page for the **LIST** command in IBM SPSS Statistics. The browser's address bar shows the URL: `http://127.0.0.1:51777/help/index.jsp?topics=/com.ibm.spss.statistics.help/syn_list.htm`. The page title is "Hilfe - Command Syntax Reference (Befehlsyntax: Referenz)".

On the left side, there is a navigation pane titled "Inhalt" (Table of Contents) with a search bar. The "LIST" command is highlighted in the list.

The main content area displays the following information:

- LIST**

```
LIST [VARIABLES={ALL* } ] [/FORMAT={ (WRAP* ) } (UNNUMBERED* ) ]
      (varlist) (SINGLE) (NUMBERED )

[/CASES={FROM (1* ) } [TO (eo* ) ] [BY (1* ) ]
          (n ) (n ) (n ) }
```
- \*\*Default if the subcommand is omitted.**
- This command reads the active dataset and causes execution of any pending commands. Under Thema [Command Order](#) finden Sie weitere Informationen.
- Example**

```
LIST VARIABLES=V1 V2 .
```
- Verwandte Themen**
  - [Overview \(LIST command\)](#)
  - [Examples \(LIST command\)](#)
  - [VARIABLES Subcommand \(LIST command\)](#)
  - [FORMAT Subcommand \(LIST command\)](#)
  - [CASES Subcommand \(LIST command\)](#)
  - [SUMMARIZE](#)
- © Copyright SPSS Inc. 1989, 2010

## 3 Der Beispieldatensatz

Um die Einführung in das Statistikpaket SPSS Statistics möglichst anschaulich zu gestalten, werden in diesem Skript viele Beispiele zum praktischen Arbeiten mit SPSS Statistics verwendet. Diesen Beispielen liegen Daten zugrunde, die der „Allgemeinen Bevölkerungsumfrage der Sozialwissenschaften“ (ALLBUS 2008) entstammen.<sup>6</sup>

Dieser Datensatz wurde von uns soweit gekürzt, dass er immer noch vielfältige Analysemöglichkeiten zulässt, aber nicht durch Fragen (Variablen) überfrachtet ist, die wir nicht benötigen. Er wird im Skript auch als „unser Datensatz“ bezeichnet, wobei sich dieser Ausdruck auf den von uns verwendeten Datensatz bezieht. Im Folgenden wird skizziert, wie die ALLBUS - Daten entstanden sind:

### 3.1 Datenerhebung

Die „Allgemeine Bevölkerungsumfrage der Sozialwissenschaften“ enthält die Antworten von Befragten einer repräsentativen Stichprobe der Bevölkerung. Zur Stichprobenbildung wurde ein personenbezogenes, zweistufiges Ziehungsverfahren eingesetzt. In der ersten Auswahlstufe wurde eine größenproportionale Stichprobe von Gemeinden gezogen. In der zweiten Auswahlstufe wurden Personenadressen aus den Einwohnermelderegistern der betreffenden Gemeinden zufällig und gleichwahrscheinlich gezogen. Interviewer befragten die so bestimmten Personen. Ihnen lag ein Fragebogen vor, dem beispielsweise folgender Auszug entstammt:

**V6 WIRTSCHAFTSLAGE IN DER BRD HEUTE**

Position: 6  
Format: F1 Missing: 7 thru highest

Beginnen wir mit einigen Fragen zur wirtschaftlichen Lage.  
Benutzen Sie für Ihre Antworten bitte die Liste.

(Int.: Liste 1 vorlegen und bis Frage 2 liegen lassen!)

F001 Wie beurteilen Sie ganz allgemein die heutige wirtschaftliche Lage in Deutschland?

Sehr gut

Gut

Teils gut / teils schlecht

Schlecht

Sehr schlecht

67.06.056274 • Hauptfragebogen
Liste 1

Es wurden 3469 Personen befragt. Der Fragebogen enthält eine Reihe von Fragen zu persönlichen Daten und Verhältnissen der Befragten (Alter, Geschlecht, Nettomonatseinkommen, ...). Außerdem gibt es Fragen, die sich auf die Einstellung zur persönlichen und wirtschaftlichen Lage, zur Freizeitaktivität und Mediennutzung, zur politischen Einstellung, zum technischen Fortschritt und Computer, zur Gesundheit und viele weitere interessante Eigenschaften beziehen.<sup>7</sup> Mit Ergänzungen enthält der Datensatz 800 Variablen (zur Terminologie „Variable“ siehe S. 17).

<sup>6</sup> Die in diesem Beitrag benutzten Daten entstammen der "Allgemeinen Bevölkerungsumfrage der Sozialwissenschaften" (ALLBUS 2008). Das ALLBUS-Programm ist 1980-86 sowie 1991 von der DFG gefördert worden. Die weiteren Erhebungen wurden von Bund und Ländern über GESIS – Leibniz-Institut für Sozialwissenschaften – finanziert. ALLBUS wird innerhalb von GESIS an den Standorten Mannheim und Köln in Zusammenarbeit mit dem ALLBUS-Ausschuß realisiert. Die vorgenannten Institutionen und Personen tragen keine Verantwortung für die Verwendung der Daten in diesem Beitrag.

<sup>7</sup> Zu Einzelheiten siehe: ZA4601.

Ziele der Erhebung und Bereitstellung der ALLBUS-Daten für Geisteswissenschaftler sind sowohl die Untersuchung von sozialen Lagen, Einstellungen, Werten und Verhaltensweisen in Deutschland, als auch die Beschreibung des sozialen Wandels im Zeitverlauf. Außerdem soll so eine Datenbereitstellung für Forscher und Lernende ermöglicht werden, die keinen unmittelbaren Zugang zu entsprechenden Primärdaten haben.

In der empirischen Praxis ist die Konstruktion der Fragen mit größter Sorgfalt vorzunehmen.<sup>8</sup> Hier wird auf dieses Problem allerdings nicht näher eingegangen.<sup>9</sup>

### 3.2 Verkodung des Fragebogens

Nach Beendigung der Befragung liegen 3469 Fragebögen vor, welche die entsprechenden Antworten der befragten Personen enthalten. Jede Frage misst ein Merkmal der Befragten, also eine Eigenschaft der Untersuchungsobjekte, die im Sachzusammenhang interessiert. Bei der Messung wird jedem Untersuchungsobjekt bei jeder Frage (Merkmal) je nach Antwort anhand einer Skala (also nach bestimmten Regeln) eine Merkmalsausprägung zugeordnet.<sup>10</sup> Merkmalsausprägungen können Zahlen oder Symbole sein. Die Skala ist die Menge der möglichen Ausprägungen eines Merkmals.<sup>11</sup> Man unterscheidet dabei:

#### → Hinweis!

- klassifikatorische (nominalskalierte),
- komparative (ordinalskalierte),
- und metrische (intervall-, verhältnisskalierte) Merkmale.

Bei klassifikatorischen Merkmalen herrscht lediglich eine "Gleich- oder Ungleichbeziehung" zwischen den Ausprägungen (zum Beispiel Mann/Frau: im ALLBUS-Datensatz v151). Bei komparativen Merkmalen hingegen handelt es sich um eine Ordnungsrelation (größer, kleiner oder besser, schlechter) unter den Ausprägungen (zum Beispiel Gesundheitszustand: v247). Die Ausprägungen metrischer Merkmale sind reelle Zahlen. Intervallskalierte Merkmale besitzen dabei keinen natürlichen Nullpunkt, daher kann die Skala positive und negative Zahlen annehmen (zum Beispiel Temperatur in Grad Celsius). Verhältnisskalierte Merkmale besitzen hingegen einen natürlichen Nullpunkt (zum Beispiel Einkommen: v388) und können daher nur aus den positiven reellen Zahlen bestehen.

Von der Skala hängen die erlaubten (das heißt auch sinnvollen) Operationen ab, die mit einem Merkmal durchgeführt werden können. Es ist beispielsweise nicht zulässig, aus klassifikatorischen oder komparativen Merkmalen Mittelwerte zu berechnen. Die Skala determiniert somit auch die Auswertungsmöglichkeiten und sollte daher mit Sorgfalt bestimmt werden.

Für eine maschinelle Auswertung mit Hilfe eines Statistikprogramms ist es günstig, wenn die Antworten (Merkmalsausprägungen) durch einen numerischen Wert repräsentiert werden. Dies ist bei einigen Merkmalen auf Grund des Skalenniveaus bereits von vornherein der Fall (Geburtsjahr, Einkommen). Da es aber die Möglichkeit gibt, Skalen mit nichtnumerischen Werten zu bilden, müssen solche Werte zunächst „numerisch verkodet“ werden. Jeder Merkmalsausprägung wird dabei eine Zahl zugeordnet und zwar so, dass die im Merkmal enthaltenen Informationen nicht verändert werden. Es wird also lediglich eine zulässige Transformation der Skala eines Merkmals durchgeführt.<sup>12</sup> Die Ausprägungen klassifikatorischer Merkmale werden dazu meist durch ganze Zahlen dargestellt.

<sup>8</sup> Auf Probleme, die im Zusammenhang mit der Datengewinnung auftreten, wie:

- richtiges Festlegen der Variablen (Merkmale),
- Messen der Ausprägungen,
- Festlegung der Grundgesamtheit,

kann hier nicht näher eingegangen werden. Das bleibt speziellen Schriften vorbehalten, vgl. zum Beispiel Friedrichs (1982).

<sup>9</sup> Siehe dazu zum Beispiel Friedrichs (1982), S. 192 ff..

<sup>10</sup> Vgl. Vogel (1989), S. 2.

<sup>11</sup> Vgl. Dobbener (1983), S. 13.

<sup>12</sup> Siehe dazu Dobbener (1983), S. 13 ff..



Ebenso die Ausprägungen von komparativen Merkmalen, wobei hier darauf zu achten ist, dass die Ordnungsrelation der Ausprägungen erhalten bleibt.

Es ist ausdrücklich zu betonen, dass sich durch dieses Zuordnen von Zahlen zu Merkmalsausprägungen das Skalenniveau einer Variablen nicht verändert. Es wird lediglich eine zulässige Abbildung der Skala durchgeführt, die den Informationsgehalt des Merkmals nicht verändert.

Bei empirischen Untersuchungen wird bei jeder Frage zusätzlich (mindestens) ein Wert festgelegt, der zu vergeben ist, wenn bei einer Frage keine Antwort vorliegt, zum Beispiel durch Antwortverweigerung (siehe dazu den Abschnitt über fehlende Werte („missing values“) Seite 23 ff.).

### 3.3 Einfache Struktur eines Datensatzes

Ein Datensatz besteht aus Zeilen. Jede Zeile wird in Spalten eingeteilt, wobei jede Spalte ein oder mehrere Zeichen (Ziffer, Buchstabe) aufnehmen kann.

→ Hinweis!

In unserem Beispieldatensatz wird nun die Antwort zu jeder Frage durch einen (in der Regel numerischen) Wert repräsentiert. Diese Werte werden für jeden Fragebogen (Untersuchungsobjekt) zeilenweise in die Spalten eingegeben. Die Spalten einer Zeile werden also den Fragen (Merkmalen) zugeordnet und zwar so, dass ein Merkmal immer in derselben Spalte steht. Ist der Fragebogen eines Untersuchungsobjektes vollständig erfasst, beginnt eine neue Zeile mit dem Fragebogen des nächsten Untersuchungsobjektes. Jede Zeile repräsentiert also einen Fragebogen und somit ein Untersuchungsobjekt. Alle Zeilen besitzen eine feste Struktur, das heißt, die Merkmale befinden sich in den einzelnen Zeilen jeweils an derselben Stelle.

Es ist nun zu beachten, dass sich SPSS Statistics für die Bezeichnung der Untersuchungsobjekte und der Merkmale einer besonderen Terminologie bedient:

Die Untersuchungsobjekte werden **Fälle** und die Merkmale **Variablen** genannt!

Jeder Fragebogen gehört somit zu einem Fall ("case") und jede Frage zu einer Variablen ("variable"). Zukünftig werden fast ausschließlich die Begriffe Fall und Variable verwendet.

Die folgende Abbildung zeigt die grundsätzliche Struktur eines Datensatzes (NxM Datenmatrix):

Fälle	Variable 1	Variable 2	Variable 3	Variable 4	...	Variable M
1	1	2	1964	12	...	1
2	2	3	1972	3	...	1
3	3	1	1960	3	...	1
4	4	3	1976	9	...	1
5	5	2	1962	8	...	1
...	...	...	...	...	...	..
N	N	3	1961	1	...	1

## 4 Datendefinition

### 4.1 Datenquellen für SPSS Statistics

Wie bereits in Kapitel 1 deutlich wurde, ist ein universelles Statistikprogramm in der Lage, empirisches Datenmaterial aus den unterschiedlichsten Anwendungsbereichen zu analysieren. Von sich aus weiß das Statistiksystem SPSS Statistics folglich nichts von den Daten, die in einem konkreten Fall zur Auswertung vorliegen. Es ist daher zunächst zu klären, in welcher Form die Daten vorliegen, und wie sie in SPSS Statistics einzubringen sind.

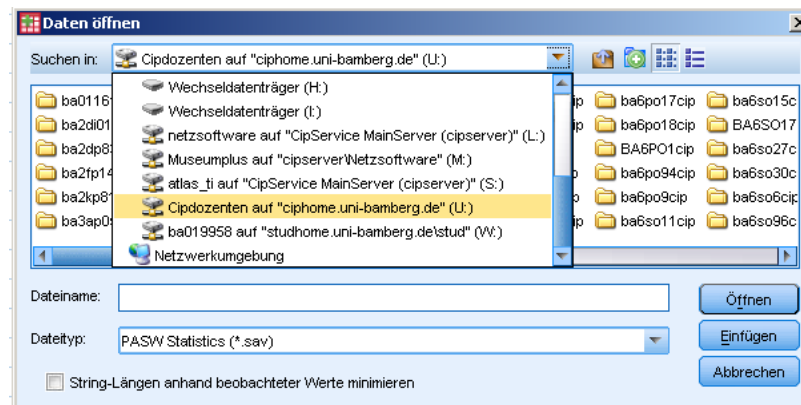
Im Abschnitt 3.3, Seite 17, wurde relativ knapp ausgeführt, wie die Struktur eines einfachen Datensatzes aufgebaut ist. Diese zu analysierenden Daten können "mit Hilfe eines geeigneten Programmsystems" (Word, Excel u. a.) in einer Datei abgelegt werden. Bei solchen Programmsystemen wird zwischen Programmen unterschieden, welche die Daten in einem programmspezifischen Format ablegen und solchen, die Daten als reine Textdateien (ASCII-Dateien) speichern. Geeignet heißt in diesem Zusammenhang, dass die betreffenden Programme die Daten so ablegen müssen, dass sie von der Struktur her für SPSS Statistics verwertbar sind. Das heißt, die grundsätzliche Struktur Anforderung "Fälle in Zeilen und Variablen in Spalten" muss erfüllt sein. Dennoch bleibt anzumerken, dass eine transponierte Sichtweise grundsätzlich auch denkbar ist. Ein Programm, das den genannten Voraussetzungen ideal gerecht wird, ist (wie zu erwarten) SPSS Statistics für Windows.

Wir unterscheiden damit (grob) zwei wesentliche Datenquellen:

⇒ **SPSS Statistics-Datensätze und sonstige programmspezifische Datendateien**

In der Datenansicht des *Daten-Editors* von SPSS Statistics können unmittelbar Daten eingegeben und über DATEI ⇒ DATEN SPEICHERN... als SPSS Statistics-Systemdatei mit der Endung ".sav" abgelegt werden. Die Daten sind dann in einem für SPSS Statistics spezifischen Format gespeichert. Wie im folgenden Kapitel noch gezeigt wird, ist die Definition der Daten in SPSS Statistics relativ einfach. Von der Datenstruktur her sind Tabellenkalkulationsprogramme ähnlich wie SPSS Statistics aufgebaut. SPSS Statistics kann Daten, die in gängigen Tabellenkalkulationsprogrammen, wie etwa Excel, abgespeichert wurden, lesen.

SPSS Statistics-Datendateien, also Datendateien, die über Eingabe der Daten in SPSS Statistics selbst oder ein Programm, das solche Dateien erstellen kann, entstanden sind, besitzen immer die Endung ".sav". Um eine SPSS Statistics-Datendatei zu bearbeiten, wird DATEI ⇒ ÖFFNEN ⇒ DATEN angeklickt. Es kommt das Fenster *Daten öffnen*:



Eine Datei wird ausgewählt und über **Öffnen** bestätigt. Anschließend erscheinen die Daten im *Daten-Editor* von SPSS Statistics.

Ist die Datei in einem anderen Format, das SPSS Statistics lesen kann, abgespeichert, so wird unter Dateiformat das entsprechende Format (.sys, .por, .xls, .w, .slk, .dbf, .dta) gewählt, die Datei gesucht und anschließend mit **Öffnen** bestätigt. Daraufhin erscheinen die Daten im Daten-Editor. Hier ist zu überprüfen, ob das Ursprungsprogramm Datenformate (Datum, Formeln, ...) verwendet, die SPSS Statistics nicht bekannt sind.

Direktzugriff auf Excel-Daten:

Dateien aus Excel 5 oder Nachfolgeversionen können direkt in SPSS Statistics eingelesen werden. Man wählt dazu im Menü *Datei öffnen* den Dateityp Excel (\*.xls).

⇒ ASCII-Textdateien

Eine ganz andere Situation liegt vor, wenn die Werte der Variablen einfach Zahl für Zahl in Dateizeilen übertragen, und dann im "reinen" Textformat (ASCII-Text) abgelegt werden, so wie dies zum Beispiel mit dem Microsoft®-Editor möglich ist. Auch mit Microsoft Office Word® können Daten in dieser Form abgelegt werden, indem beim Speichern als Dateityp "Text" angegeben wird. Sind die Daten so gespeichert, dann enthält die Datei mit den Daten keinerlei Information, welche Spalten zu welcher Variablen gehören. Dies muss innerhalb des SPSS Statistics-Programms definiert werden (siehe Abschnitt 4.3, Seite 25).

## 4.2 Dateneingabe direkt in SPSS Statistics

### 4.2.1 Variablen definieren in der Variablenansicht

Wie bereits oben ausgeführt wurde, können Daten direkt in ein "leeres" Datenfenster von SPSS Statistics eingegeben werden. Dazu sind zunächst die erforderlichen Variablen zu definieren. Dies geschieht mittels der Registerkarte **VARIABLENANSICHT** im *Daten-Editor*.

Jeder Variablen (siehe Abschnitt 3.3, Seite 17) muss in der Spalte „Name“ einen eindeutigen Variablennamen zugewiesen werden.

Die gewünschten Variablennamen trägt man in die Spalte „Name“ ein. Sofern kein Name eingegeben wird, nennt SPSS Statistics die Variablen "VAR" und ergänzt eine fortlaufende fünfstellige Zahl (VAR00001, VAR00002, ...). Bei der Vergabe von Variablennamen sind bestimmte Regeln einzuhalten, die in Abschnitt 5.2.1, Seite 32, beschrieben sind.

	Name	Typ	Spaltenf...	Dezimal...	Variablenlabel	Wertelabels	Fehlende W...	Spalten	Ausrichtung	Messniveau	Rolle
1	v3	Numerisch	1	0	ERHEBUNGSGEBIET ...	{1, ALTE B...	Keine	8	Rechts	Nominal	Eingabe
2	v4	Numerisch	1	0	DEUTSCHE STAATSA...	{1, JA}...	7 - HI	8	Rechts	Nominal	Eingabe
3	v5	Numerisch	2	0	ATTRAKTIVITAET DE...	{1, UNATTR...	Keine	8	Rechts	Nominal	Eingabe
4	v6	Numerisch	1	0	WIRTSCHAFTSLAGE ...	{1, SEHR G...	7 - HI	8	Rechts	Nominal	Eingabe
5	v7	Numerisch	1	0	WIRTSCHAFTSLAGE,...	{1, SEHR G...	7 - HI	8	Rechts	Nominal	Eingabe
6	v8	Numerisch	1	0	WIRTSCHAFTSLAGE ...	{1, WESEN...	7 - HI	8	Rechts	Nominal	Eingabe
7	v9	Numerisch	1	0	WIRTSCHAFTSLAGE,...	{1, WESEN...	7 - HI	8	Rechts	Nominal	Eingabe
8	v10	Numerisch	1	0	REGIERUNGSVERAN...	{1, SEHR G...	7 - HI	8	Rechts	Nominal	Eingabe
9	v11	Numerisch	3	1	HAEUFIGKEIT VON F...	{0, NIE}...	9,7 - HI	8	Rechts	Nominal	Eingabe
10	v12	Numerisch	4	0	FERNSEHGESAMTD...	{0, SIEHT K...	9997 - HI, 0	8	Rechts	Skala	Eingabe
11	v13	Numerisch	1	0	FERNSEHGESAMTD...	{0, SIEHT K...	7 - HI, 0	8	Rechts	Nominal	Eingabe
12	v14	Numerisch	1	0	NACHRICHTENKONS...	{0, NIE}...	7 - HI, 0	8	Rechts	Nominal	Eingabe
13	v15	Numerisch	3	1	KONSUMHAEUFIGK...	{0, NIE}...	9,7 - HI	8	Rechts	Nominal	Eingabe
14	v16	Numerisch	1	0	NACHRICHTENKONS...	{0, SIEHT N...	7 - HI, 0	8	Rechts	Nominal	Eingabe
15	v17	Numerisch	3	1	KONSUMHAEUFIGK...	{0, NIE}...	9,7 - HI	8	Rechts	Nominal	Eingabe
16	v18	Numerisch	3	1	HAEUFIGKEIT TAGES...	{0, NIE}...	9,7 - HI	8	Rechts	Nominal	Eingabe
17	v19	Numerisch	1	0	INTERNET FUER INF...	{1, JA}...	7 - HI	8	Rechts	Nominal	Eingabe
18	v20	Numerisch	1	0	HAEUFIGKEIT: INTER...	{0, NUTZT K...	7 - HI, 0	8	Rechts	Nominal	Eingabe

**Beispiel:**

Das Erhebungsgebiet: West-Ost wird zur Variablen v3.

Eine Übersicht über die SPSS Statistics-Variablen des Beispieldatensatzes befindet sich im Anhang auf Seite 83.

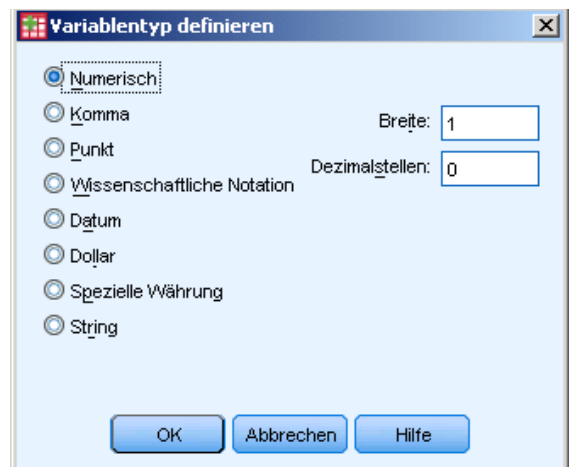
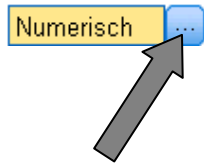
→ Hinweis!

### 4.2.2 Spalte Variablentyp (Typ)

SPSS Statistics unterscheidet bei der Dateneingabe numerische ("numeric") und Zeichenketten- ("string-") variablen. Die Werte numerischer Variablen bestehen nur aus Zahlen ("numbers"). Zeichenketten oder Stringvariablen nehmen beliebige Zeichen (Buchstaben, Ziffern, Sonderzeichen) auf (zur Definition von Zahlen und Zeichenketten siehe Abschnitt 5.2.3, Seite 32).

Der Typ wird folgendermaßen definiert:

Anwählen der neuen Variable in der Spalte Typ. Es erscheint eine grau schattierte Fläche innerhalb der Zelle (siehe Abbildung). Per Mausklick auf diese Fläche gelangt man in das Menü *Variablentyp definieren*.



Die Variablentypen Komma, Punkt, Wiss. Notation, Datum, Dollar und Spezielle Währung sind spezielle Varianten numerischer Variablen und müssen hier nicht diskutiert werden.

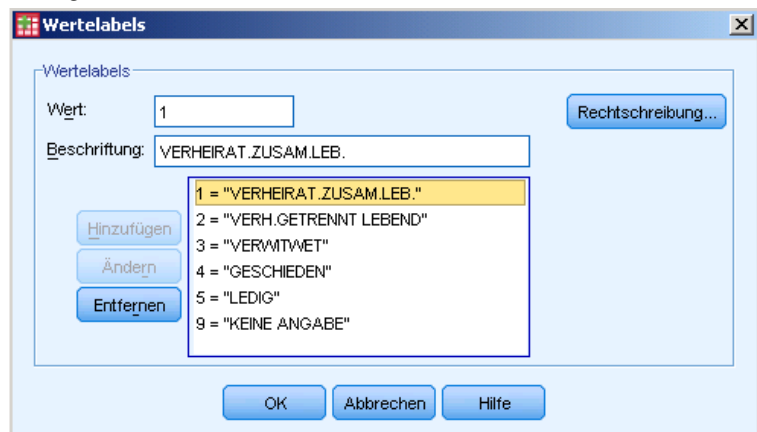
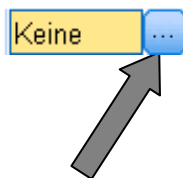
**Beispiel:** Die Variable Geschlecht ist als numerische Variable bestimmt, die eine Stelle beansprucht und keine Dezimalstellen besitzt.

### 4.2.3 Variablen-Labels - Werte-Labels - Fehlende Werte

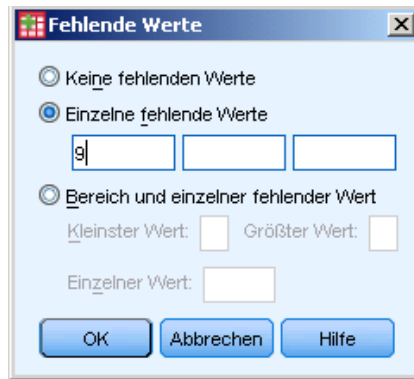
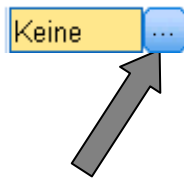
Durch die Vergabe von Labels besteht die Möglichkeit, Variablen oder Variablenwerte genauer zu beschreiben. Im Kapitel 4.2.4.1, Seite 22 und Kapitel 4.2.4.2, Seite 22, wird dies näher erläutert.

Das Variablen-Label der neuen Variable (im Beispiel v248) wird einfach in der Spalte Variablenlabel eingetragen (Familienstand).

Die Werte-Labels der neuen Variable v248 werden durch Anwählen der Zelle in der Spalte Wertelabels vergeben. Man gelangt daraufhin in folgendes Menü:



Fehlende Werte sind ein Problem aller empirischen Untersuchungen. Sie werden in der Spalte „Fehlende Werte“ durch Anklicken der Zelle definiert. SPSS Statistics kennt zwei Arten von fehlenden Werten: systemdefinierte fehlende Werte (**system-missing values**) und benutzerdefinierte fehlende Werte (**user-missing values**). Näheres hierzu in Kapitel 4.2.4.3, Seite 23.



**Beispiel:**

Um die fehlenden Werte zu definieren, klickt man in der Spalte „Fehlende Werte“ in der entsprechenden Zelle auf das graue Kästchen. Daraufhin öffnet sich obiges Menü. Durch Aktivieren des Kontrollkästchens „Einzelne fehlende Werte“ können nun ein bis maximal drei Werte definiert werden (siehe Abschnitt 4.2.4.3, Seite 23).

Nachdem die Definition der Variablen abgeschlossen ist, kann mit der Dateneingabe begonnen werden. Hierzu werden die Werte aus den Fragebögen einfach zeilenweise bei den entsprechenden Variablen eingegeben. Das Ergebnis der Dateneingabe sieht dann folgendermaßen aus:

	v3	v4	v5	v6	v7	v8	v9	v10	v11	v12	v13
1	1	1	8	3	2	3	3	4	7,0	120	
2	1	1	8	4	3	4	1	1	7,0	240	
3	1	1	6	4	3	3	3	2	6,0	120	
4	1	1	5	3	3	3	3	3	7,0	360	
5	1	1	8	3	2	3	3	5	2,0	150	
6	1	1	7	4	4	4	5	3	7,0	180	
7	1	2	4	4	3	4	3	5	7,0	330	
8	1	1	6	3	3	4	4	1	7,0	60	
9	1	1	7	3	2	3	3	2	7,0	300	
10	1	1	8	2	1	4	3	1	5,0	60	
11	1	1	8	2	3	3	3	2	7,0	60	
12	1	1	7	3	2	4	3	3	7,0	240	
13	1	1	8	4	3	4	4	1	7,0	150	
14	1	1	8	5	3	4	3	2	7,0	180	
15	1	1	8	3	5	4	5	2	4,0	240	
16	1	1	8	3	2	2	3	2	7,0	120	
17	1	1	8	3	3	4	3	4	7,0	120	

Der Inhalt des *Daten-Editors* wird als SPSS Statistics-Datei abgespeichert (siehe Abschnitt 4.1, Seite 18).

**4.2.4 Variablenbeschreibung über Befehle im Syntax-Editor**

Nachdem in Abschnitt 4.2.3, Seite 20 ff. bereits die Variablenbeschreibung über Menüs diskutiert wurde, soll in den folgenden Abschnitten auf die Variablenbeschreibung über Befehle eingegangen werden. Diese Befehle sind über den *Syntax-Editor* (siehe dazu Abschnitt 5.1.2, Seite 30 f.) auszuführen. Zum besseren Verständnis des Aufbaus und der Syntax von Befehlen, sowie deren Darstellung in dieser Einführungsschrift wird an dieser Stelle auf die Kapitel 5.2 und 5.3 verwiesen.

#### 4.2.4.1 Variablen-Labels

Die Kennzeichnung der Variablen lediglich durch deren Variablennamen ist für den Anwender häufig ziemlich unbefriedigend. In den Ergebnislisten sollten nicht nur die kurzen SPSS Statistics-Variablennamen aufgeführt werden, sondern zusätzlich eine Erläuterung zu den Variablen in der Fachterminologie des Verarbeiters erfolgen. Auch wenn Ausprägungen der Variablen in der Ausgabeliste erscheinen, bedürfen nichts sagende Verkodungen einer Erklärung. Ein Problem für viele Anwender ist darüber hinaus der Umgang mit "fehlenden Werten" in den Daten.

Mit dem `VARIABLE LABELS`-Kommando wird Variablen eine Erklärung in der Sprache des Benutzers zugewiesen (zugehöriges Menü siehe Abschnitt 4.2.3, Seite 20 f.). Der Befehl hat folgende Struktur:

##### → Syntax

```
VARIABLE LABELS varname 'label' [[/]varname ... ].
```

**varname** ist der Name einer bereits definierten Variablen.

**label** ist eine nähere Bezeichnung, ein "Etikett" für diese Variable. Das "label" erläutert, was die Variable inhaltlich bedeutet. Ein "label" muss den Konventionen für Zeichenketten entsprechen. Die Länge eines "labels" beträgt maximal 255 Zeichen. Analyseergebnisse zeigen mindestens 40 Zeichen des "labels", somit wird nicht immer das gesamte "label" ausgegeben. Jede Variable darf nur ein "label" haben. Durch die Etikettierung können die Variablen in der Ausgabeliste so gekennzeichnet werden, dass ihre Bedeutung transparent wird. Für unser Beispiel wurden unter anderen folgende `VARIABLE LABELS` vergeben:

##### → Beispiel

```
VARIABLE LABELS
  V154 'Alter Befragte<r>'
  /v151 'Geschlecht Befragte<r>'
  /v524 'Konfession Befragte<r>'
  /v248 'Familienstand'
  /v173 'Allgemeiner Schulabschluss'
  /v386 'Befr: Nettomonatseinkommen'.
```

#### 4.2.4.2 Werte-Labels

Der Befehl `VALUE LABELS` hingegen erlaubt es, Etiketten auch für Werte (Ausprägungen) von Variablen zu vergeben (zugehöriges Menü siehe Abschnitt 4.2.3, Seite 20 f.):

##### → Syntax

```
VALUE LABELS varliste wert 'label' wert 'label' ...
  [[/]varliste ... ].
```

**varliste** ist eine Liste bereits definierter Variablen (mindestens eine Variable). Besteht die "varliste" aus mehreren Zeichenkettenvariablen ("strings"), so müssen alle die gleiche Länge besitzen.

**wert** ist eine mögliche Ausprägung der Variablen der "varliste". Werte von Zeichenkettenvariablen sind in Hochkommata einzuschließen.

**label** ist eine bis zu 120 Zeichen lange Erläuterung (Etikett) zur Bedeutung dieses Wertes. Das "label" erscheint meist dann, wenn der zugehörige "wert" in der Ergebnisliste steht. Nicht alle Prozeduren drucken das Etikett in seiner ganzen Länge aus.

##### → Beispiel

```
VALUE LABELS
  V151 1 'Mann' 2 'Frau'.
```

### 4.2.4.3 Fehlende Werte (missing values)

Ein besonderes Problem aller empirischen Untersuchungen stellen die so genannten missing values dar. SPSS Statistics unterscheidet zwischen „system-missing value“ und „user-missing value“, behandelt bei Analysen aber beide Varianten gleich, sofern dies nicht explizit geändert wird.

#### ⇒ system-missing values

Findet SPSS Statistics als Wert einer numerischen Variablen ein Zeichen, das keine Zahl ist oder nur ein Leerzeichen, so vergibt SPSS Statistics für den betreffenden Fall den so genannten „system-missing value“. Auch das Ergebnis unzulässiger Berechnungen, wie Division durch Null, ist der „system-missing value“. SPSS Statistics kennzeichnet den Fall dadurch als "defekt". In der Ausgabeliste erscheint für diesen Wert ein •.<sup>13</sup> Solche Fälle gehen nicht in Berechnungen und Auswertungen ein. „System-missing values“ werden im Dateneditor durch ein Komma und im Viewer durch einen Punkt repräsentiert. Leerzeichen am Anfang und Ende eines numerischen Wertes "ignoriert" SPSS Statistics.

#### ⇒ user-missing values

Es ist bei empirischen Datenerhebungen nicht ungewöhnlich, dass bei einem Teil der Fälle für einzelne Variablen die Ausprägungen fehlen. Das kann vielfältige Gründe haben. Einige davon seien kurz skizziert:

Bei einer Befragung von Personen mögen manche Fragen einen sensiblen Bereich betreffen, zu dem sich ein Teil der Befragten nicht äußern möchte (beispielsweise Fragen nach dem Einkommen, Sexualverhalten, Videokonsum).

Eine andere Möglichkeit des Zustandekommens von fehlenden Werten ist, dass die Werte bestimmter Variablen an manchen Untersuchungsobjekten nicht messbar sind. Eine Analyse zeitgenössischer Aufzeichnungen über Krankheiten der Patienten eines Krankenhauses im 15. Jahrhundert wird nicht für jeden Patienten vollständige Daten ergeben, da die Aufzeichnungen teilweise unleserlich sind.

Was ist hier zu tun? Im ersten Beispiel liegen die Gründe, warum eine Frage unbeantwortet bleibt, vermutlich im (unbekannten) wahren Wert. Mancher Mensch möchte Dritten keine Angaben über seinen Verdienst machen. Vielleicht weil dieser nach eigener Einschätzung sehr hoch oder sehr niedrig ist. Im zweiten Beispiel (Krankenhaus im 15. Jahrhundert) fehlt ein solcher Grund für die Abwesenheit eines Wertes. Durch zufällige Einflüsse kommen hier keine Messungen zu Stande.

Darüber hinaus können Werte auch einfach nur falsch erfasst (gemessen) sein. Eine Temperatur von 40° Celsius am 24. Dezember in Bamberg wird vermutlich ein Erhebungsfehler sein.

SPSS Statistics bietet nun mit dem *MISSING VALUES*-Kommando (zugehöriges Menü siehe Abschnitt 4.2.3, Seite 20) die Möglichkeit der Kennzeichnung, wenn bei einem Fall ein fehlender Wert vorliegt. Dazu muss der Anwender Werte vorgeben, die an Stelle des unbekanntes (wahren) Wertes stehen. Diese Werte heißen „user-missing values“. Solche Werte in den Daten signalisieren, dass bei einem Fall keine Angabe über die tatsächliche (wahre) Ausprägung einer Variablen möglich ist. Um unterscheiden zu können, warum ein Wert fehlend ist, dürfen maximal drei Werte bei einer Variablen zu "missing values" erklärt werden. Es bedeutet dann zum Beispiel "9" als Variablenwert, dass sich jemand geweigert hat zu antworten, "99", dass sie/er nicht zu Hause war und "999" dass sie/er vergessen wurde.

SPSS Statistics stellt mit den Befehlen *MVA* (Menü: ANALYSIEREN ⇒ ANALYSE FEHLENDER .WERTE...) und *MULTIPLE IMPUTATION* (Menü: ANALYSIEREN ⇒ MULTIPLE IMPUTATION) Werkzeuge zum Ersetzen fehlender Werte durch geschätzte Werte zur Verfügung.

Das Kommando zur Definition von „user-missing values“ lautet:

<b>MISSING VALUES varliste (werteliste) [[/]varliste ... ].</b>
---

<b>varliste</b>	ist der Name einer oder mehrerer Variablen, die bereits definiert sind. Die Variablen einer Liste müssen vom gleichen Typ (" <i>numeric</i> " oder " <i>string</i> ") sein. Die Länge von Zeichenkettenvariablen einer "varliste" darf unterschiedlich sein.
-----------------	--

→ Syntax

<sup>13</sup> Leerzeichen in den Daten können als beliebige Zahlen eingelesen werden, wenn dies durch das SPSS-Kommando *SET BLANKS=zahl* gefordert wird. Siehe dazu SPSS (1990), S. 666.

Durch das Schlüsselwort *ALL* an Stelle einer Variablenliste werden für alle Variablen dieselben fehlenden Werte vereinbart. Dann müssen jedoch alle Variablen vom gleichen Typ sein.

**werteliste** kann aus folgenden Elementen bestehen:

Aus bis zu drei Einzelwerten, die durch ein Komma oder Leerzeichen getrennt sind. Werte von Zeichenkettenvariablen sind dabei in Hochkommata einzuschließen. Bestehen die aufgeführten (fehlenden) Werte von Zeichenkettenvariablen aus weniger Zeichen als die Variable aufnehmen darf, so wird rechts mit Leerzeichen aufgefüllt. Ist ein Wert länger als eine Zeichenkettenvariable, wird dieser rechts abgeschnitten.

Aus einem Wertebereich, der durch das Schlüsselwort *THRU* gebildet wird:

*wert1 THRU wert2*

*wert1* und *wert2* sind Zahlen und gehören dem Bereich der fehlenden Werte an. Alle Zahlen von *wert1* bis *wert2* werden also zu „missing values“ erklärt. Der größte und der kleinste Wert werden durch *HIGHEST* und *LOWEST* angesprochen (Abkürzung *HI*, *LO*). Wertebereiche sind für Zeichenkettenvariablen nicht erlaubt.

Insgesamt dürfen innerhalb der Klammern nur drei Einzelwerte stehen, einschließlich derer links und rechts vom *THRU*. *HI* und *LO* werden wie Einzelwerte behandelt.

Folgende Kombinationen von Werten und Bereichen sind denkbar:

<i>MISSING VALUES</i>	<i>varliste (wert1).</i>
<i>MISSING VALUES</i>	<i>varliste (wert1, wert2).</i>
<i>MISSING VALUES</i>	<i>varliste (wert1, wert2, wert3).</i>
<i>MISSING VALUES</i>	<i>varliste (wert1 THRU wert2).</i>
<i>MISSING VALUES</i>	<i>varliste (wert1, wert2 THRU wert3).</i>
<i>MISSING VALUES</i>	<i>varliste (wert1 THRU wert2, wert3).</i>

Die fehlenden Werte, die durch die "werteliste" eines *MISSING VALUES*-Kommandos definiert werden, heißen, wie oben bereits ausgeführt wurde, in der SPSS Statistics-Terminologie „user-missing values“. Diese sind nur mit dem *MISSING VALUES*-Kommando vereinbar. „System-missing values“ hingegen weist das Programm, wie bereits oben geschildert, selbst zu, wenn ein Wert in den Daten nicht zum Typ der Variablen passt (eine numerische Variable zum Beispiel Buchstaben aufnehmen soll) oder eine Transformation zu einem undefinierten Wert führt (Division durch null).

Zunächst ein fiktives Beispiel, um einige Möglichkeiten des *MISSING VALUES*-Kommandos zu demonstrieren:

### → Beispiel

<i>MISSING VALUES</i>	<i>parks (9999).</i>
<i>MISSING VALUES</i>	<i>sex ('X') /nme (-9, -99, -9999).</i>
<i>MISSING VALUES</i>	<i>gew (1 THRU 20) /kurz, lang (LO THRU 0, 99999).</i>
	<i>/vx (-1, 200 THRU HI).</i>

Bei unseren Beispieldaten haben wir nahezu bei allen Variablen fehlende Werte. Beispielsweise wurden für "v6" alle Werte, die größer sind als der Wert „7“, vorgesehen, falls jemand nicht geantwortet hat oder keine Antwort auf die Frage weiß. Diese Werte werden durch folgendes Kommando als "fehlend" gekennzeichnet:

<i>MISSING VALUES</i>	<i>v6 (7 THRU HI).</i>
-----------------------	------------------------

Mehrere *MISSING VALUES*-Kommandos sind zulässig. Wird eine Variable in mehreren Kommandos aufgeführt, so ist nur das letzte Kommando wirksam. Alle früheren Vereinbarungen sind "überschrieben". Die folgende Anweisung hebt für die "varliste" alle "missing values" Vereinbarungen auf:

<i>MISSING VALUES</i>	<i>varliste ( ).</i>
-----------------------	----------------------



Zwei Dinge sind bislang ungeklärt:

Wie werden fehlende Werte verarbeitet, und welche Werte sollten als "user-missing values" herangezogen werden?

Fälle mit fehlenden Werten werden häufig einfach aus der Berechnung ausgeschlossen. Über Spezifikationen lässt sich in der Regel davon abweichend steuern, ob und wie fehlende Werte in Auswertungen einbezogen werden. Wie die fehlenden Werte zu behandeln sind, ist vor der Analyse zu klären.

Zur Kennzeichnung eines fehlenden Wertes darf keine mögliche Ausprägung der betreffenden Variablen genommen werden. Dies hätte zur Folge, dass nicht mehr unterscheidbar ist, welcher Fall einen fehlenden Wert besitzt.

Meist wird daher als fehlender Wert ein sehr großer oder sehr kleiner Wert definiert, der außerhalb des Wertebereiches der Variablen liegt (zum Beispiel 9 bei dichotomen Variablen mit 0 und 1 als Ausprägungen). Vorsicht ist bei Variablen geboten, deren Wertebereich von vornherein nicht bekannt ist. Es wäre gefährlich, bei der Frage nach dem Nettomonatseinkommen 9999,99 als fehlenden Wert zu nehmen, weil das ein mögliches Einkommen eines Befragten sein könnte. In diesen Fällen ist unter Umständen ein negativer Wert als fehlender Wert sinnvoll (wie -1).

### 4.3 Definition von ASCII-Text-Eingabedaten

Eingabedaten im "reinen" Textformat enthalten keine Informationen über die Zuordnung der Spalten der Datenzeilen zu Variablen. Die Daten müssen daher dem SPSS Statistics-Programm bekannt gemacht werden. Zu dieser „Bekanntmachung“ oder „Beschreibung“ der Daten (Datendefinition) sind mindestens folgende Informationen notwendig:

- In welcher Datei steht unser Datensatz?
- Wie heißen die Variablen?
- Wo stehen welche Variablen im Datensatz?
- Wie sind die Variablen im Datensatz getrennt (Leerzeichen, Komma, ...)?

Die Daten könnten, wie ausgeführt wurde, im Datenfenster von SPSS Statistics direkt eingegeben werden. Dies ist für größere Datensätze allerdings nicht praktikabel, da der "Dateneditor" von SPSS Statistics recht unkomfortabel ist. Häufig werden die Daten daher mit speziellen Eingabeprogrammen in einfachen Textdateien, so genannten ASCII-Dateien, abgelegt. In diesem Fall liegen sie als reine Zahlenkolonnen (ohne jegliche Formatierungsangabe) vor. Wir gehen im Folgenden davon aus, dass Letzteres der Fall ist.

Diese so genannten „Rohdaten“ gewinnen nur durch die Datendefinition an Bedeutung. Ohne Beschreibung bleiben sie lediglich Kolonnen von Zahlen. Erst die Beschreibung füllt diese Kolonnen inhaltlich aus. Dies kann entweder mit Hilfe der Menüsteuerung (DATEI ⇒ ÖFFNEN ⇒ DATEN...) oder mit dem nachstehend aufgezeigten Syntax-Befehl *DATA LIST* geschehen.

```
DATA LIST FILE='datei' [FIXED] [RECORDS=n]
          /[satz#] varliste spalte[-spalte] [(format)]
          [varliste ...][ / ...] .
```

→ Syntax

Ein Beispiel mit unserem Datensatz wäre dann (... steht für Zeilen, die im Beispiel der Übersichtlichkeit halber nicht aufgeführt wurden):

```
DATA LIST FILE = 'allbus2008.dat' RECORDS=1
          /1 v1 1-4 v2 5-8 v3 9 v4 10 v5 11-12 v6 13 v7 14
          ...
          v794 697-707 (F11.9) v795 708-718 (F11.0) v796 719-720 v800 721-728.
```

→ Beispiel

Der Befehl hat einen sehr komplexen Aufbau, der hier nicht erläutert werden soll. Weitere Informationen zur Kommandosprache von SPSS Statistics sind daher dem „IBM SPSS Statistics 19 Command Syntax Reference“ zu entnehmen.

## 5 Analysen in SPSS Statistics (procedures)

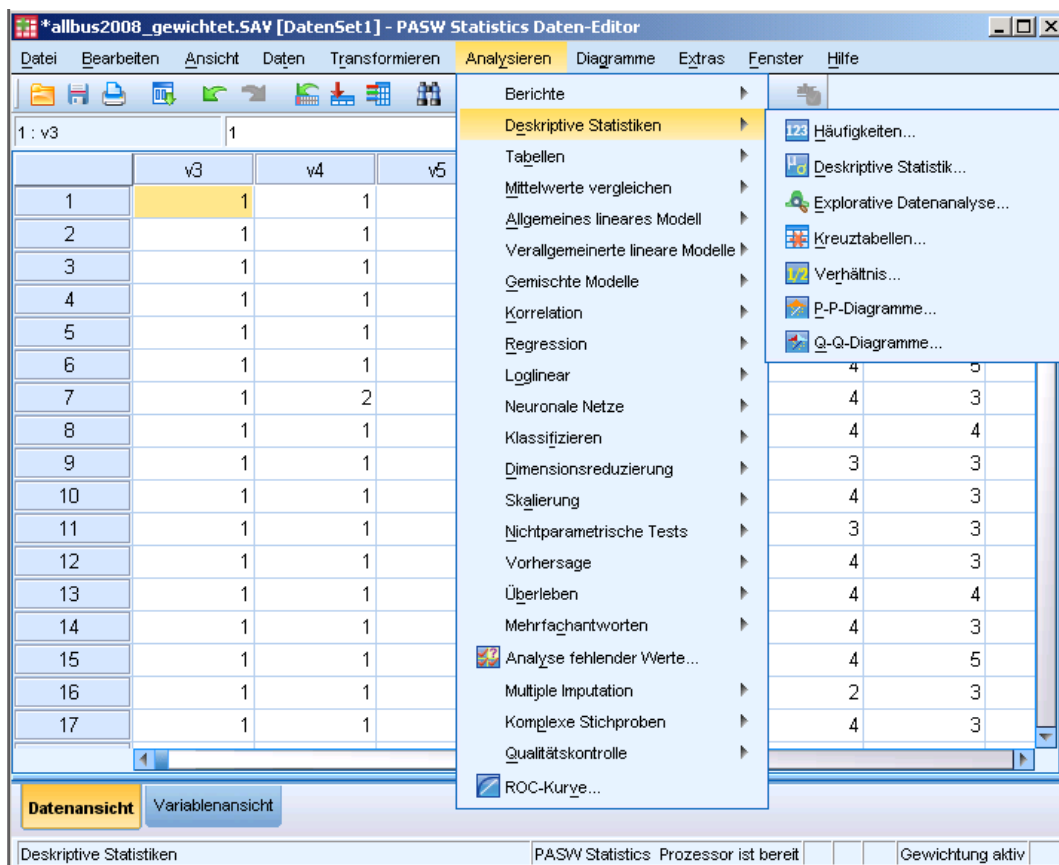
### 5.1 Der Weg zur Analyse

Der Anwender muss seine Analysewünsche in einer für das Statistiksystem verständlichen Weise konkretisieren. Dazu gibt es in SPSS Statistics grundsätzlich zwei Möglichkeiten:

1. Eine Analyse wird über ein Menü angefordert.
2. Eine Analyse wird über einen Befehl im *Syntax-Editor* angefordert.

#### 5.1.1 Analyse und Modifikationen über Menüs

Die Menüleiste von SPSS Statistics für Windows gliedert sich in die Menüpunkte DATEI, BEARBEITEN, ANSICHT, DATEN, TRANSFORMIEREN, ANALYSIEREN, DIAGRAMME, EXTRAS, FENSTER und HILFE. Das Menü ANALYSIEREN beinhaltet die statistischen Verfahren, die mit SPSS Statistics durchgeführt werden können. Es enthält eine Liste von Oberbegriffen, die dann zu den eigentlichen Menüs zum Anfordern von Statistiken führen. Jedem Oberbegriff-Eintrag folgt daher ein Pfeil, der auf eine weitere Menüebene verweist. Mit den Oberbegriffen wurde versucht, die Statistikverfahren in SPSS Statistics inhaltlich zu gliedern, damit den Nutzerinnen und Nutzern nicht lediglich eine Liste von Verfahren angeboten wird. Es ist trotzdem ein gewisses Maß an Sachkenntnis nötig, um zu einem bestimmten Verfahren zu finden.



SPSS kann mit verschiedenen Modulen lizenziert werden. Bestimmte Module können weitere Einstiegspunkte unter ANALYSIEREN erzeugen!

Als Beispiel für die Vorgehensweise mit der Menüsteuerung soll das Menü zur Erstellung einer absoluten, relativen und kumulierten Häufigkeitsverteilung dienen. Aus den Menüs wird dazu

ANALYSIEREN ⇒ DESKRIPTIVE STATISTIKEN ⇒ HÄUFIGKEITEN...

ausgewählt. Es erscheint die Dialogbox *Häufigkeiten*.



Alle Dialogboxen für statistische Verfahren enthalten folgende oder ähnliche Komponenten:

- **Variablenliste:** Die Variablenliste befindet sich auf der linken Seite der Dialogbox und enthält eine Liste aller Variablen der Datendatei.
- **Variable(n):** Das Feld Variable(n) enthält die Variable(n), die für die ausgewählte statistische Analyse gebraucht werden. Um die entsprechende Variable aus der Variablenliste dorthin zu übertragen, muss man zuerst die ausgewählte Variable markieren (die Auswahl erfolgt schneller durch Eintippen der ersten Buchstaben) und auf die Pfeil (↗)-Schaltfläche klicken. Alternativ kann die Variable per Doppelklick in das Feld übertragen werden.
- **Schaltflächen:** Durch das Anklicken dieser Buttons rechts und unten im Fenster werden bestimmte Befehle bzw. Aktionen ausgeführt. In der Dialogbox Häufigkeiten befinden sich die Schaltflächen **OK**, **Einfügen**, **Zurücksetzen**, **Abbrechen** und **Hilfe** sowie **Statistiken...**, **Diagramme...** und **Format...**. Die drei Punkte (...) deuten darauf hin, dass hinter diesen Schaltflächen noch Unterdialogboxen existieren.

Die fünf Standard-Befehlsschaltflächen eines Dialogfeldes, die sich immer rechts oder unten im Fenster befinden, haben die folgende Bedeutung:

- **OK:** Startet die entsprechende Prozedur und schließt gleichzeitig das Dialogfeld.
- **Einfügen:** Überträgt einen SPSS Statistics-Befehl, der den gewählten Menü-Einstellungen entspricht, in den Syntax-Editor (siehe hierzu Abschnitt 5.1.2, S.30). Dort können sie gespeichert und bearbeitet werden.
- **Zurücksetzen:** Macht eine Auswahl aus der Variablenliste rückgängig.
- **Abbrechen:** Macht alle Änderungen, die seit dem Öffnen des Dialogfeldes getätigt wurden, rückgängig und schließt das Dialogfeld wieder.
- **Hilfe:** Öffnet ein Hilfefenster, das Informationen über die aktuelle Dialogbox enthält.

Fast alle Statistikverfahren in SPSS Statistics liefern ohne besondere Zusatzanforderungen bestimmte Ergebnisse, die voreingestellt sind. So liefert folgendes *Häufigkeiten*-Fenster nach einem Klick auf **OK**

...



... ohne besondere Zusatzangabe folgende Tabelle:

**KONFESSION, BEFRAGTE<R>**

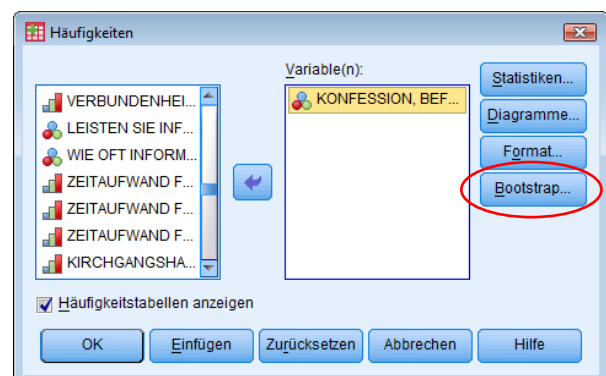
		Häufigkeit	Prozent	Gültige Prozen- te	Kumulierte Pro- zente
Gültig	EVANG.OHNE FREIKIRCH	1088	31,4	31,6	31,6
	EVANG.FREIKIRCHE	37	1,1	1,1	32,6
	ROEMISCH-KATHOLISCH	1180	34,0	34,2	66,9
	AND.CHRISTL.RELIGION	79	2,3	2,3	69,1
	AND.NICHT-CHRISTLICH	132	3,8	3,8	73,0
	KEINER RELIGIONSGEM.	932	26,9	27,0	100,0
	Gesamt	3448	99,4	100,0	
Fehlend	VERWEIGERT	16	,5		
	KEINE ANGABE	5	,2		
	Gesamt	21	,6		
Gesamt		3469	100,0		

Über verfahrensspezifische Schaltflächen, die sehr vielfältig sind und im weiteren Verlauf des Kapitels 5 bei der Diskussion einzelner Statistikverfahren genauer erörtert werden, können, abhängig vom gewählten Verfahren, mehr oder weniger Ausgabelemente angefordert werden, als von SPSS Statistics per Voreinstellung erscheinen.

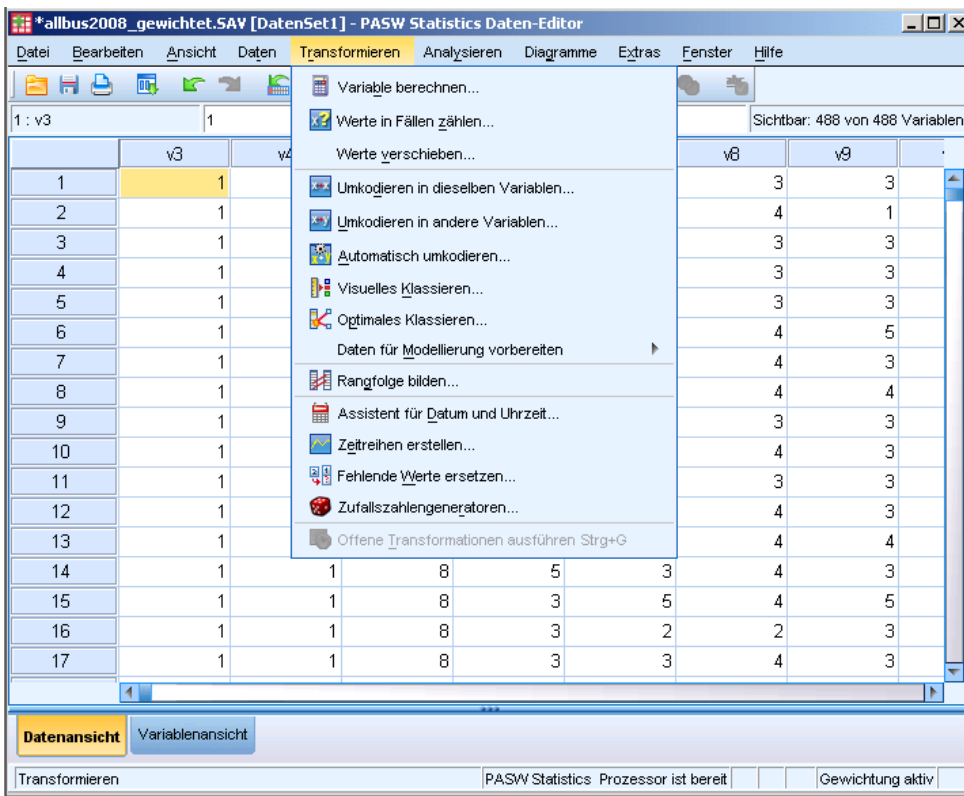
Im obigen Menü sind beispielsweise **Statistiken...**, **Diagramme...**, **Format...** solche verfahrensspezifischen Schaltflächen. Über **Diagramme...** kann so zum Beispiel zusätzlich eine Grafik erzeugt werden.

#### Anmerkung zu Unterdialogboxen in „Analysefenstern“:

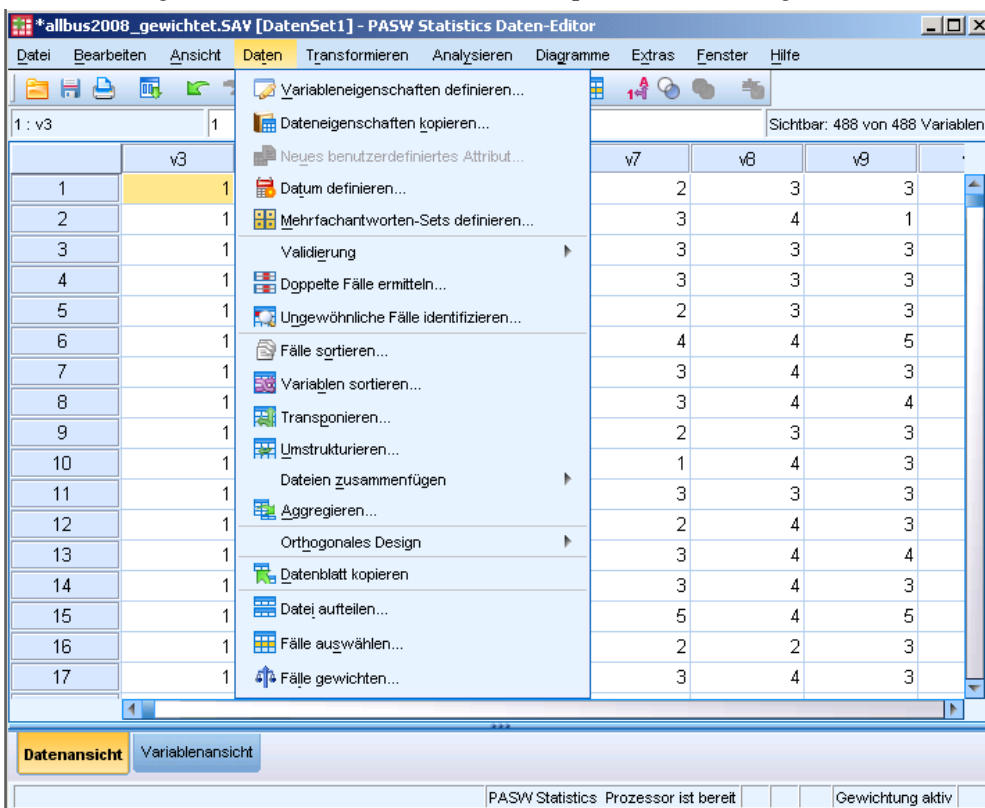
SPSS kann mit verschiedenen Modulen lizenziert werden. Bestimmte Module können das Erscheinungsbild von Unterdialogboxen ändern. Bei Lizenzierung des Moduls „SPSS Bootstrapping“ beispielsweise, haben einige Dialogboxen für Analysen einen zusätzlichen Unterpunkt:



Neben den Analysen spielen Modifikations- und Selektionsmöglichkeiten der Daten eine große Rolle in SPSS Statistics (siehe dazu Abschnitt 6, Seite 47 ff.). Modifikationen finden sich unter dem Menüpunkt TRANSFORMIEREN:



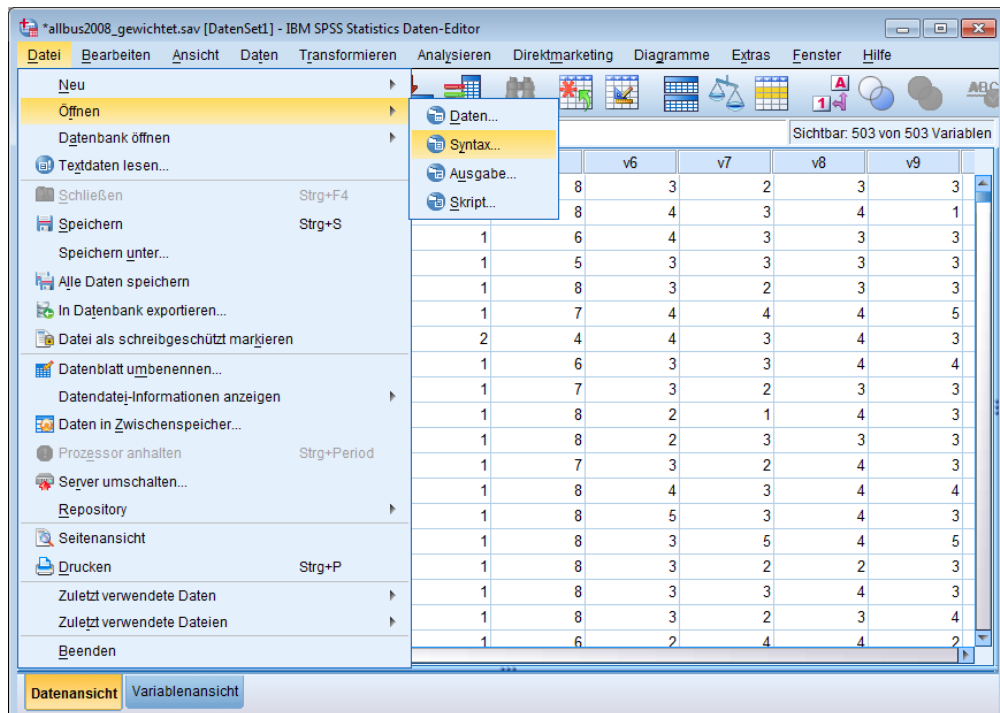
Selektionsmöglichkeiten können unter dem Menüpunkt DATEN ausgewählt werden:



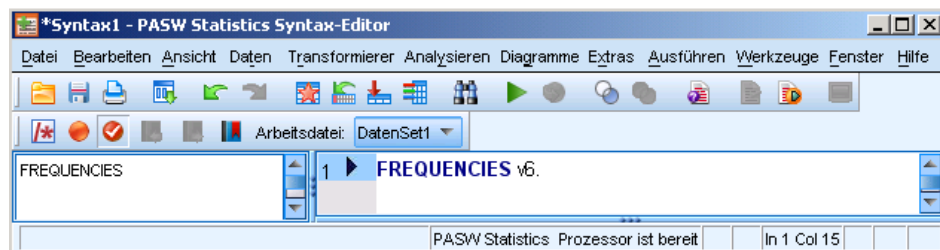
## 5.1.2 Analyse und Modifikationen über Befehle

Der Anwender kann seine Analyse- und Modifikationswünsche auch direkt durch Befehle oder Kommandos an SPSS Statistics konkretisieren. Befehle sind Anweisungen, die dem Programm sagen, was genau womit gemacht werden soll. Letztlich erstellen die oben diskutierten Befehlsmenüs lediglich einen Befehl für den Anwender und vermitteln diesen, ohne dass der Anwender dies merkt, an SPSS Statistics. Ein Beleg hierfür ist die Möglichkeit, aus den Menüpunkten von ANALYSIEREN über die Schaltfläche **Einfügen**, direkt Befehle zu erzeugen.



Befehle für SPSS Statistics werden im *Syntax-Editor* formuliert. Der *Syntax-Editor* muss in SPSS Statistics in der Regel erst über das Menü DATEI ⇒ NEU ⇒ SYNTAX angefordert werden:



Im *Syntax-Editor*, der daraufhin erscheint, werden dann die Befehle eingegeben:



Der Syntax-Editor bietet Unterstützung beim Erstellen der Befehle. Sobald ein Buchstabe im Editor getippt wird, schlägt SPSS Statistics alle Befehle mit dem entsprechenden Anfangsbuchstaben vor. Ist ein Befehl ausgewählt, können mit Strg+Leertaste Spezifikationen angerufen werden. Solange ein erkannter Befehl unvollständig ist, wird er rot angezeigt. Bei einem erkannten, syntaktisch korrekten Befehl wechselt die Farbe auf Blau. Unerkannte Kommandos sind schwarz. Syntaktisch korrekte Befehlsabkürzungen werden ebenfalls schwarz angezeigt.

Ein Befehl kann über beliebig viele Zeilen fortgesetzt werden und endet immer mit einem Punkt. Soll ein Befehl ausgeführt werden, wird der Cursor innerhalb des Befehls positioniert, oder aber der Befehl wird markiert. Ein Klick auf  oder den Menüpunkt AUSFÜHREN (Tastenkombination Strg + R) bringt den Befehl zur Ausführung. Mehrere hintereinander stehende Befehle werden nacheinander ausgeführt, indem alle Befehle markiert werden und anschließend auf  oder den Menüpunkt AUSFÜHREN geklickt wird.

SPSS Statistics unterscheidet zwar zwischen Groß- und Kleinschreibung, es ist allerdings für die Interpretation der Befehle ohne Bedeutung, ob sie im *Syntax-Editor* groß- oder kleingeschrieben werden. Bei den Befehlen ist eine bestimmte Struktur (*Syntax*) einzuhalten, damit SPSS Statistics sie interpretieren kann. Diese wird im übernächsten Abschnitt (Abschnitt 5.2, Seite 31 ff.) ausführlich dargestellt.

### 5.1.3 Befehle oder Menüs – eine Frage des Geschmacks?

Eine viel diskutierte Frage ist, ob Analysewünsche mit Menüs oder Befehlen vermittelt werden sollten. Menüs bieten den Vorteil, schnell und ohne große Grundkenntnisse Analysen anzufordern. Damit ermöglichen sie unerfahrenen Anwendern einen guten Einstieg in SPSS Statistics. Auch erfahrene Nutzerinnen und Nutzer greifen inzwischen für ad hoc-Abfragen gerne auf Menüs zurück.

Komplexere Studien erfordern allerdings in der Regel, dass eine Reihe von Abläufen sukzessiv von SPSS Statistics abgearbeitet wird, damit das gewünschte Ergebnis erzeugt wird. Obwohl dies meist auch mit Menüs gemacht werden könnte, ist hier Vorsicht geboten: Wenn in vielen Menüs nacheinander gearbeitet wird, geht leicht der Überblick verloren. Es wird fast unmöglich, die durchgeführten Arbeiten nachzuvollziehen, oder diese Dritten (Kollegen, Prüfern) transparent zu machen. So werden ähnliche Arbeitsabläufe bei der Arbeit mit Menüs oft jedes Mal mit großem Aufwand neu erstellt. Hier liegt aus unserer Sicht der Vorteil der Befehle. Sie lassen sich speichern und sind damit jederzeit reproduzierbar. Bei der Arbeit mit Befehlen wird für Dritte nachvollziehbar, was gemacht wurde, weshalb sich auch Fehler schneller entdecken lassen. Außerdem erfordern „fortgeschrittene“ Analyseanforderungen manchmal auch den Einsatz von Befehlen, weil sie über Menüs überhaupt nicht realisierbar sind.

Unserer Meinung nach sind umfangreichere Analysearbeiten nur unter Kenntnis der SPSS Statistics-Befehle zu bewältigen. Daher werden in diesem Skript Befehle ausführlich dargestellt. Allerdings wird, soweit möglich, auch auf die jeweils zugehörigen Menüs eingegangen.

Im Zusammenhang mit komplexeren Analyseverfahren kann das Menü vor allem dann als eine geeignete Hilfestellung verwendet werden, wenn man das vollständige Kommando nicht kennt. Dazu aktiviert man im Menü alle Felder und klickt anschließend auf den Button „Einfügen“ (statt auf „Ok“). So wird der Befehl in ein Syntaxfenster eingefügt, und kann von dort aus abgespeichert, nachbearbeitet und auch ausgeführt werden.

## 5.2 Aufbau und Syntax der Befehle

Aufgrund der großen Bedeutung der Befehle, die über das Syntax-Menü vermittelt werden, wird an dieser Stelle ausführlich erörtert, wie Befehle zu formulieren sind.

Ein SPSS Statistics-Kommando kann sowohl mit großen als auch mit kleinen Buchstaben geschrieben werden. (In diesem Skript werden Großbuchstaben für die Kommandos benutzt.) Ein Befehl beginnt immer mit einem **Kommandoschlüsselwort** ("**command keyword**"), das auch aus mehreren einzelnen Worten bestehen kann. Das Schlüsselwort sagt dem Programm, welche Aktion durchgeführt werden soll, ist also der eigentliche Befehlsname.

Der Befehl *SHOW* zum Beispiel zeigt SPSS Statistics-Systemeinstellungen an.

Zu fast allen Kommandos gibt es so genannte **Spezifikationen** ("**specifications**"), die bei einem Kommando genau festlegen, was, wie, womit zu machen ist. Spezifikationen sind immer durch mindestens ein Leerzeichen vom Kommandoschlüsselwort getrennt. Viele Spezifikationen beinhalten ein oder mehrere **Subkommandos** ("**subcommands**"), die ihrerseits wieder Spezifikationen benötigen.

Beim *SHOW*-Kommando gibt es unter anderem die Spezifikationen *ALL*, *BOX*, *ERRORS*:

```
SHOW ALL.
SHOW BOX.
SHOW ERRORS.
```

Diese dürfen einzeln oder kombiniert aufgeführt werden.

Spezifikationen bestehen allgemein aus den folgenden Elementen:

- Namen (Names),
- Schlüsselworten (Keywords),
- Zahlen (Numbers),
- Zeichenketten (Strings),
- arithmetischen Operatoren (Arithmetic Operators),
- speziellen Begrenzungszeichen (Special Delimiters) und Leerzeichen (Space), um, sofern dies notwendig ist, die einzelnen Spezifikationselemente voneinander abzugrenzen.

Im Folgenden werden diese Elemente einzeln besprochen:

→ Beispiel

### 5.2.1 Namen (Names)

Namen werden benötigt, um die Variablen für SPSS Statistics zu benennen. Das geschieht entweder zu Beginn der Auswertung, um Variablen einmal neu zu definieren oder während der Auswertung, um zusätzliche Variablen zu erzeugen.

Namen dürfen seit der Programmversion SPSS 16.0 maximal 64 Zeichen lang sein (vorher nur 8 Zeichen), und beginnen mit einem der Zeichen A-Z, @, # oder \$. Dem ersten Zeichen eines Namens dürfen beliebige Buchstaben, Ziffern, die Zeichen @, #, \$, ., und \_ folgen. Namen dürfen nicht mit . oder \_ enden.

Variablen, deren Name mit einem # beginnt, sind **Hilfs-, oder Arbeitsvariablen** ("scratch variables"). Sie existieren nur temporär, das heißt, eine bestimmte Zeit lang und können weder ausgewertet noch gespeichert werden. Ein \$-Zeichen am Beginn eines Variablennamens kennzeichnet so genannte **Systemvariablen** ("system variables"). Das sind Variablen, die automatisch von SPSS Statistics gebildet werden und die der Benutzer nicht verändern kann. Sie stellen dem Anwender eine Reihe von Informationen zur Verfügung:

Variable	Information
\$CASENUM	Fallnummer
\$SYSMIS	System-missing value
\$DATE	Datum
\$TIME	Datum und Uhrzeit

Es ist nicht erlaubt, dass **Benutzervariablen** mit einem \$-Zeichen beginnen!

Eine Ausnahmestellung bei der Namensbildung nehmen Namen in SPSS die Statistics-Makros ein, wobei auf diese hier nicht weiter eingegangen wird.<sup>14</sup>

### 5.2.2 Schlüsselworte (Keywords)

Schlüsselworte sind für SPSS Statistics von ganz spezieller Bedeutung. Neben den "command keywords" (wie *SHOW*), die ein Kommando identifizieren, gibt es weitere Schlüsselworte, die Spezifikationselemente sind (bei *SHOW* beispielsweise: *LOCAL* oder *ALL*). Diese Schlüsselworte veranlassen SPSS Statistics zu festgelegten Aktionen.

Die SPSS Statistics-Sprache ist so strukturiert, dass es nicht zu Verwechslungen zwischen Variablennamen und Schlüsselworten kommen kann. Einige Schlüsselworte sind jedoch auch an solchen Stellen erlaubt, wo Variablennamen vorkommen können. Diese Schlüsselworte ("reserved keywords") sind daher reserviert und dürfen nicht als Variablennamen verwendet werden. Solche "reserved keywords" sind:

*ALL AND BY EQ GE GT LE LT NE NOT OR TO WITH.*

### 5.2.3 Zahlen und Zeichenketten (Numbers and Strings)

Zahlen und Zeichenketten sind in SPSS Statistics-Programmen an vielen Stellen erlaubt:

- Beim Bezug auf Werte ("values") von Variablen (wie *nme = 2000*),
- als Konstanten bei numerischen Transformationen (wie *x = 1*)
- oder als Zeichenketten ("strings") bei "String-" Funktionen, also Funktionen mit Zeichenketten als Argument (wie *abt = 'rolag'*)

**Zahlen** setzen sich aus den Ziffern 0 bis 9 zusammen. Sie sind durch ihren numerischen Wert von Bedeutung. Handelt es sich um Dezimalzahlen, so steht an Stelle des bei uns üblichen Kommas ein (Dezimal-) Punkt (Beispiel: statt 0,5 steht 0.5). Als erstes Zeichen einer Zahl sind auch '+' und '-' zu-

<sup>14</sup> Siehe zu Makros SPSS (1990), S. 344 ff.



lässig. Zahlen dürfen beliebig viele Ziffern umfassen. Es sollte aber beachtet werden, dass die Rechengenauigkeit eines PCs begrenzt ist. Erlaubte Zahlen sind entsprechend beispielsweise:

1, 12345, -0.07, 78.54, -457.80, +70, 0.000024

**Zeichenketten** bestehen aus beliebigen Zeichen: den Buchstaben A-Z, den Ziffern 0-9, dem Leerzeichen und Sonderzeichen, wie !@#\$%&/()=?^ . Auch Umlaute sind erlaubt. Zeichenketten müssen in Hochkommata eingeschlossen werden, wenn sie als Spezifikation stehen: 'HANS', '12-35-89', '§ XXX'.

Falls Zeichenketten aufgrund ihrer Länge über mehrere Zeilen fortzusetzen sind, so werden sie zeilenweise in Hochkommata eingeschlossen. Am Schluss jeder Zeile, die fortzusetzen ist, steht ein Pluszeichen:

*TITLE            'Das ist eine SPSS Statistics-Auswertung ' +  
                      'für den Beispieldatensatz'.*

Statt der Hochkommata (') [nicht zu verwechseln mit Akzenten (´ `)] können auch Anführungszeichen (") geschrieben werden, solange man nicht mit einem Hochkomma beginnt und mit einem Anführungszeichen aufhört oder umgekehrt. Schließt man Zeichenketten mit Hochkommata ein, so ist das Anführungszeichen ein erlaubtes Zeichen innerhalb der Zeichenkette. Ebenso ist bei mit Anführungszeichen umschlossenen Zeichenketten das Hochkomma ein erlaubtes Zeichen. Ein verdoppeltes Hochkomma innerhalb einer Zeichenkette, die durch Hochkommata begrenzt ist, wird wie ein (erlaubtes) Hochkomma interpretiert (entsprechendes gilt umgekehrt für Anführungszeichen).

## 5.2.4 Arithmetische Operatoren und Begrenzungszeichen (Arithmetic Operators and Delimiters)

Arithmetische Operatoren werden in arithmetischen Ausdrücken verwendet. Arithmetische Operatoren sind folgende Zeichen:

Operator	Bedeutung
+	Addition
-	Subtraktion
*	Multiplikation
/	Division
**	Potenzierung

Das **Gleichheitszeichen** (=) steht im Zusammenhang mit arithmetischen Ausdrücken bei Wertzuweisungen wie  $A = A+B+C$ . Das bedeutet: Zunächst wird der Wert der rechten Seite ( $A+B+C$ ) bestimmt und dann der Variablen auf der linken Seite des Gleichheitszeichens (A) zugewiesen (sprich: A ergibt sich aus  $A+B+C$ ).

Das Gleichheitszeichen (=) als spezielles **Begrenzungszeichen** steht zwischen einem Subkommando und den dazugehörigen Spezifikationen ( $FILE='DATE1'$ ) und um Äquivalenz zweier Listen auszudrücken ( $V1, V2 = V3, V4$ ).

Arithmetische Operatoren sind "selbstbegrenzend" und erfordern keine besondere Abgrenzung zu anderen Spezifikationselementen. Allerdings darf vor und nach ihnen eine beliebige Anzahl von Leerzeichen stehen.

Manche Elemente von Spezifikationen bedürfen allerdings einer besonderen Abgrenzung zu ihrer "Umgebung". Dafür gibt es **spezielle Begrenzungszeichen** ("special delimiters"). Spezielle Begrenzungszeichen sind:

(	'	"	/	=
---	---	---	---	---

Die speziellen Begrenzungszeichen werden im Folgenden kurz erläutert.

**Klammern** ( ) schließen häufig Argumente von Funktionen ein, aber auch Schlüsselworte, die mit Variablen verwechselt werden könnten, und bestimmte Wertelisten.

**Hochkommata** (') und **Anführungszeichen** (") schließen Zeichenketten ("strings") ein und grenzen diese dadurch ab (Beispiel: "Harry's Liquor Shop").

Der **Schrägstrich** (/) trennt Subkommandos oder wiederholbare Anweisungen voneinander ab.

Um Probleme zu vermeiden, empfiehlt es sich, das Gleichheitszeichen (=) und den Schrägstrich (/) immer genau dort einzusetzen, wo diese Zeichen in den Syntaxdiagrammen zu SPSS Statistics aufgeführt sind.

Beispiel für den Einsatz spezieller Begrenzungszeichen:

```
RECODE   var1 (1,2 = 1)
/ var2 ('X' = 'Y').
```

Der Befehl *RECODE* führt dabei einmal die Anweisung *var1 (1,2 = 1)* durch. Der Schrägstrich bedeutet, dass zudem die Anweisung *var2 ('X' = 'Y')* durchgeführt werden soll. (Die genaue Bedeutung des Befehls *RECODE* wird auf S. 47 ff. näher erläutert.)

Ist zwischen Spezifikationselementen weder ein arithmetischer Operator, noch ein Begrenzungszeichen vorgeschrieben, so werden sie durch Leerzeichen, ein Komma oder eine Kombination von beiden getrennt.

### 5.3 Darstellung von Befehlen in dieser Einführungsschrift

Bei den Kommandos in SPSS Statistics sind meist nur wenige Spezifikationselemente notwendig, um Aktionen des Programms zu erreichen. Die übrigen Spezifikationen eines Kommandos beziehen sich auf besondere Anforderungen und sind durch **Voreinstellungen** ("defaults") abgedeckt.

In dieser Einführungsschrift wird eine Reihe von Befehlen dargestellt, die zum Arbeiten mit SPSS Statistics wichtig sind. Der Leser wird in die Lage versetzt, anhand der vorhandenen Angaben selbstständig Analyseprogramme zu erstellen. Dazu ist es nicht notwendig, alle Befehle mit all ihren Spezifikationen zu erläutern. Dies würde die Übersichtlichkeit der Schrift beeinträchtigen. Es ist daher in der Regel nicht der volle Umfang eines Befehls angegeben, sondern nur die Spezifikationselemente, die für einfache Analysen unabdingbar sind.

An einigen Stellen gibt es bei den Kommandos Hinweise auf zusätzliche Möglichkeiten der Spezifikation, meist ergänzt durch einen Verweis auf das SPSS Statistics-Handbuch.

Hier werden Kommandos folgendermaßen dargestellt:

Mit Großbuchstaben werden Schlüsselworte ("keywords") bezeichnet, die genauso zu übernehmen sind, wie sie aufgeführt werden. Angaben in Kleinbuchstaben sind benutzerspezifisch. Jeder Anwender muss hier die für seine Problemstellung zutreffenden Werte einsetzen. Wenn aus mehreren möglichen Spezifikationen eine ausgewählt werden kann, so stehen die einzelnen (alternativen) Spezifikationselemente in geschweiften Klammern ({...}). Angaben in eckigen Klammern ([...]) sind optional, das heißt, nicht unbedingt erforderlich.

Im Folgenden wird ein Beispiel für die Darstellung von Befehlen gegeben:

#### → Syntax

```
SORT CASES BY   varliste [{(A)}{(D)}] [varliste ...].
```

Die Worte *SORT CASES BY* sind zu übernehmen. An Stelle von "varliste" gibt der Benutzer seine eigenen Variablen an. Alle übrigen Angaben sind entbehrlich. Eine mögliche Zusatzangabe ist (A) oder (D). Die Sequenz "varliste [{(A)}{(D)}]" darf beliebig oft dastehen ([varliste ...]).

Zulässig sind demnach (unter anderem) folgende *SORT CASES* Befehle:

```
SORT CASES BY   sex alter.
SORT CASES BY   sex alter (A).
SORT CASES BY   sex alter (D) einkomm bildung.
SORT CASES BY   sex alter (A) einkomm bildung (D).
SORT CASES BY   sex alter (A) einkomm bildung (D) TV.
```

#### → Beispiel

Eine genauere Erläuterung dieses Befehls erfolgt auf S. 73 ff.

Bei der Darstellung der einzelnen SPSS Statistics-Befehle wird zunächst ein allgemeines Syntaxdiagramm eines Befehls aufgeführt. Anschließend werden die einzelnen Spezifikationselemente ausführlich erörtert. Diese Syntaxdiagramme sind immer eingerahmt und grau unterlegt.

Darauf folgen Beispiele anhand des ALLBUS-Datensatzes, der auf Seite 15 ff. beschrieben ist. Diese Beispiele demonstrieren, wie ein Befehl eingesetzt werden kann. Beispiele werden kursiv gedruckt. Schlüsselworte sind in den Beispielen stets in Großbuchstaben und kursiv, Variablen- und Dateinamen in Kleinbuchstaben geschrieben:

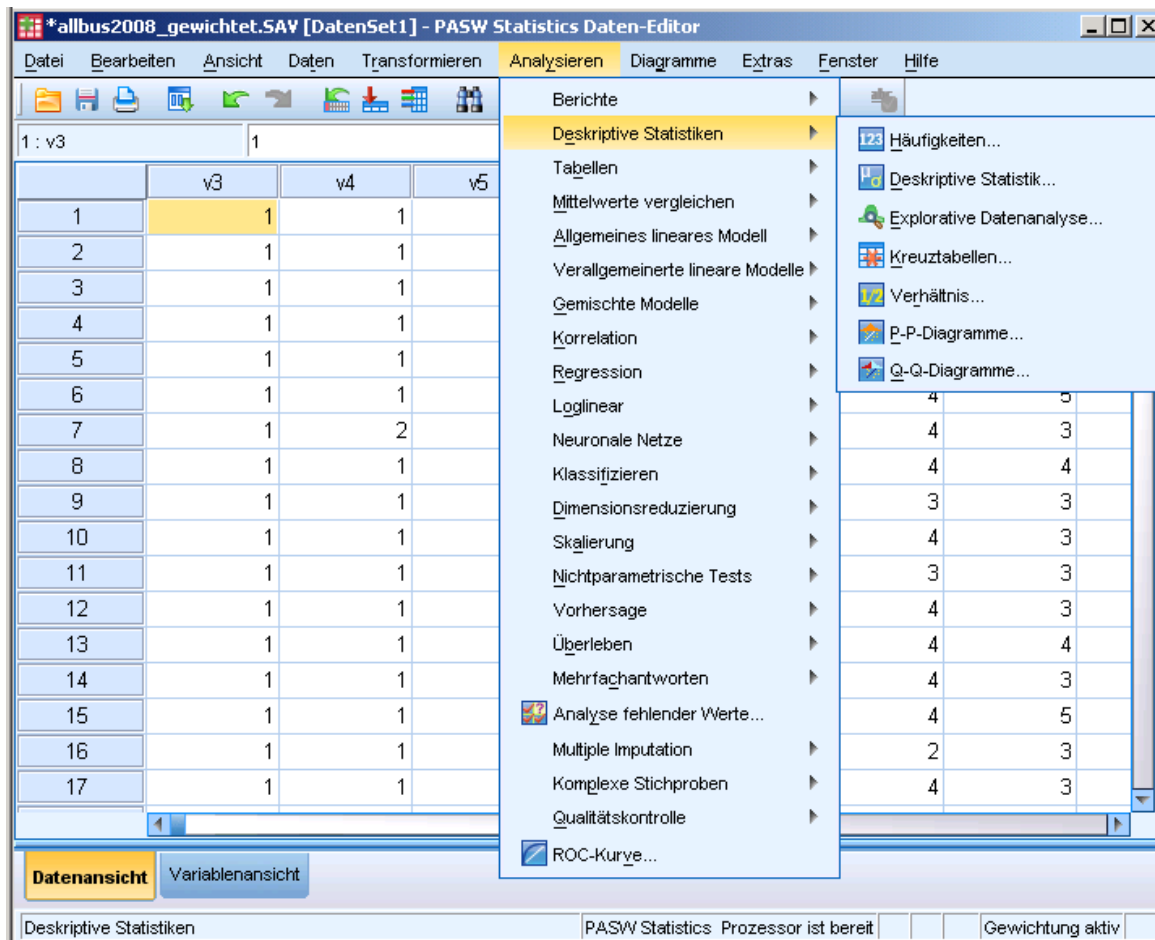
```
SORT CASES BY variable.
```

### 5.4 Prozeduren im Überblick

SPSS Statistics-Kommandos zur Durchführung statistischer Analysen heißen Prozeduren ("procedures"). Die Prozeduren legen fest, welche Auswertungen durchzuführen sind. Vor der Durchführung von Analysen kann es notwendig sein, Variablen zu modifizieren und/oder Fälle zu selektieren.

→ Menü

Die Prozeduren befinden sich bei den Menüs unter dem Menüpunkt ANALYSIEREN:



In Abschnitt 5.1.1, S. 26 ff., wurde bereits ausführlich besprochen, wie Analysen mit Hilfe von Menüs durchzuführen sind. Wie im Abschnitt 5.1.2, S. 30 ff. dargestellt wurde, können Analysen auch über SPSS Statistics-Befehle oder Menüs angefordert werden.

Eine Prozedur ("procedure") ist ein Kommando, das Daten liest. Prozeduren greifen immer auf die Daten zu und führen die angeforderten statistischen Berechnungen durch, deren Ergebnisse dann im Viewer erscheinen. Fast alle übrigen Kommandos (wie Modifikationen) werden dagegen zunächst nur vermerkt und erst mit der ersten ihnen folgenden Prozedur ausgeführt.

Prozedur-Kommandos haben immer folgende Struktur:

**PROZEDURNAME spezifikationen.**

<b>Prozedurname</b>	gibt an, welche Prozedur ausgeführt werden soll.
<b>spezifikationen</b>	regeln genau, womit was zu machen ist. Bei den Prozeduren in SPSS Statistics sind meist nur wenige Spezifikationselemente notwendig, um Aktionen des Programms zu erreichen. Die übrigen Spezifikationen eines Kommandos beziehen sich auf besondere Anforderungen und sind durch Voreinstellungen ("defaults") abgedeckt.

**→ Beispiel**

```
DESCRIPTIVES v388.
```

Die Prozedur *DESCRIPTIVES* berechnet deskriptive Statistiken für (metrische) Variablen. Per Voreinstellung sind das Mittelwert, Standardabweichung, Minimum und Maximum. Fehlende Werte sind von den Berechnungen ausgeschlossen. Diese Voreinstellungen lassen sich durch entsprechende Subkommandos ändern:

```
DESCRIPTIVES v388
/MISSING=INCLUDE
/STATISTICS=MEAN, VARIANCE.
```

*MISSING=INCLUDE* bewirkt, dass fehlende Werte in die Berechnung mit eingehen. *STATISTICS=MEAN, VARIANCE* bedeutet, dass nur der Mittelwert und die Varianz auszugeben sind.

**5.5 Wichtige Prozeduren**

**Hinweis für Leserinnen und Leser, die das Skript zum Selbststudium von SPSS Statistics benutzen:**

Alle Beispiele wurden mit gewichteten Allbus-Daten berechnet, um „repräsentative“ Ergebnisse zu präsentieren. Gewichtet wurde mit dem „Personenbezogenen Ost-West-Gewicht“ aus v792. Für Einzelheiten zur Gewichtung siehe Abschnitt 6.1.5, S. 57.

**5.5.1 LIST**

Die *LIST* Prozedur listet die Werte der in der "varliste" aufgeführten Variablen fallweise in einem Standardformat auf.

**→ Syntax**

```
LIST [[VARIABLES=] {varliste} {ALL}].
```

Bedeutung der Spezifikationselemente:

Steht statt einer "varliste" das Schlüsselwort *ALL*, werden alle Variablen aufgelistet. Das bewirkt auch ein *LIST* ohne Spezifikationen.

Wir lassen für die Beispieldaten die Variablen v3 bis v10 auflisten. Dabei beschränken wir die Auflistung auf die ersten 10 Fälle des Datensatzes mit Hilfe des Unterbefehls *CASES*. Der Syntaxbefehl lautet:

```
LIST VARIABLES = v3 to v10
/CASES = 10.
```

**→ Beispiel****→ Ergebnis**

Im *Viewer* erscheint dann folgendes Ergebnis:

```
v3 v4 v5 v6 v7 v8 v9 v10
1 1 8 3 2 3 3 4
1 1 8 4 3 4 1 1
1 1 6 4 3 3 3 2
1 1 5 3 3 3 3 3
1 1 8 3 2 3 3 5
1 1 7 4 4 4 5 3
1 2 4 4 3 4 3 5
```

```

1 1 6 3 3 4 4 1
1 1 7 3 2 3 3 2
1 1 8 2 1 4 3 1
    
```

Number of cases read: 10      Number of cases listed: 10

Wir sehen, dass die Variablen und ihre Werte spaltenweise aufgelistet sind. In jeder Zeile ist somit ein Fall mit seinen Merkmalsausprägungen aufgeführt.

### 5.5.2 Häufigkeiten (FREQUENCIES)

*FREQUENCIES* berechnet absolute, relative und kumulierte relative Häufigkeiten für die Ausprägungen der aufgeführten Variablen.

```

FREQUENCIES      varliste
                    [/STATISTICS=[MEAN][STDDEV][VARIANCE][MODE][MEDIAN]...[ALL]].
    
```

→ Syntax

<b>STATISTICS</b>	gibt an, welche statistischen Maßzahlen zur Charakterisierung der Häufigkeitsverteilungen zu berechnen sind. Einige wichtige Maßzahlen:	
<b>MEAN</b>		Mittelwert
<b>STDDEV</b>		Standardabweichung
<b>VARIANCE</b>		Varianz
<b>MODE</b>		Modus
<b>MEDIAN</b>		Median
<b>ALL</b>		alle 15 verfügbaren Statistiken

Für die Variable "Kirchgangshäufigkeit" werden aus den Beispieldaten absolute, relative und kumulierte relative Häufigkeiten berechnet. Zusätzlich wird der Modus (Modalwert), also der häufigste Wert, angefordert. Der Befehl dafür sieht folgendermaßen aus:

→ Beispiel

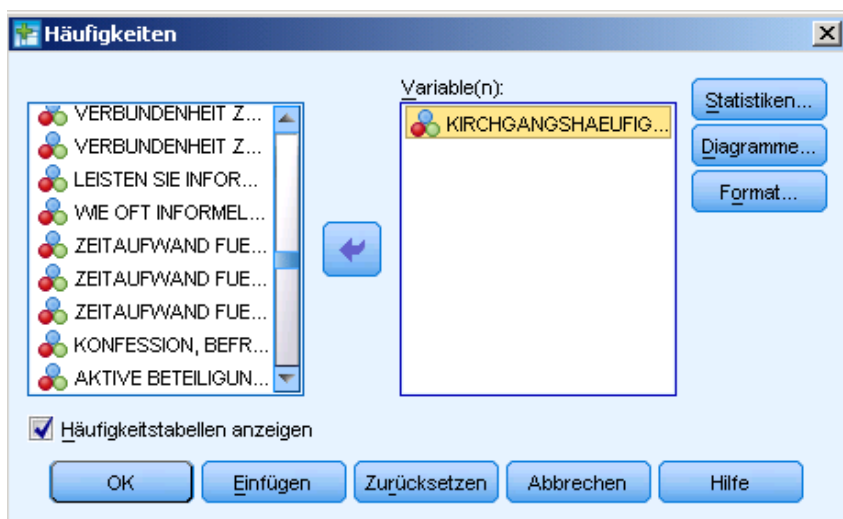
```

FREQUENCIES v526/STATISTICS = MODE.
    
```

Menüentsprechung:

→ Menü

ANALYSIEREN ⇒ DESKRIPTIVE STATISTIKEN ⇒ HÄUFIGKEITEN...

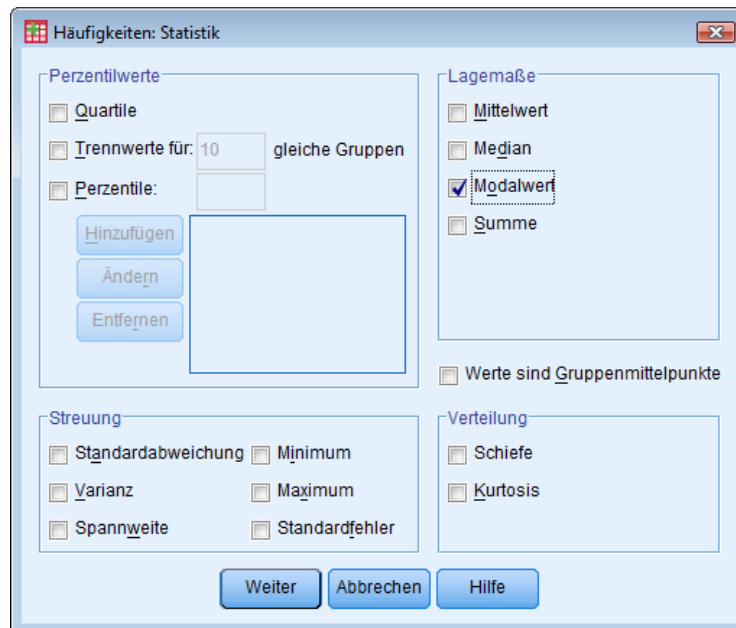


In der Variablenliste werden alle numerischen und String-Variablen angezeigt. Die Variable, für welche die statistischen Maßzahlen berechnet werden sollen, wird aus der Variablenliste ausgewählt und mit der Pfeil (↵)-Schaltfläche in die Liste **Variable(n)**: verschoben.

Um für numerische Variablen statistische Maßzahlen zu berechnen, wird durch Aktivierung der Schaltfläche Statistiken... das Dialogfenster *Häufigkeiten: Statistik* geöffnet.

Durch Ankreuzen werden für die angegebene Variable die entsprechenden Maßzahlen bestimmt. Per Voreinstellung werden keine Maßzahlen berechnet.

Dialogfenster der Schaltfläche: **Statistiken...**



## → Ergebnis

Unabhängig davon, welche Vorgehensweise gewählt wurde, also ob über den Syntax-Befehl oder über die Menüsteuerung, erscheint im *Viewer* folgendes Ergebnis:

### Statistiken

#### KIRCHGANGSHAEUFIGKEIT

N	Gültig	3456
	Fehlend	13
Modus		6

#### KIRCHGANGSHAEUFIGKEIT

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	UEBER 1X DIE WOCHE	57	1,7	1,7	1,7
	1X PRO WOCHE	241	6,9	7,0	8,6
	1-3X PRO MONAT	360	10,4	10,4	19,0
	MEHRMALS IM JAHR	701	20,2	20,3	39,3
	SELTENER	1001	28,9	29,0	68,3
	NIE	1095	31,6	31,7	100,0
	Gesamt		3456	99,6	100,0
Fehlend	KEINE ANGABE	13	,4		
Gesamt		3469	100,0		

Die erste Tabelle (Statistiken) beinhaltet Angaben über die einbezogenen Fälle (Gültig, Fehlend) und den Modus (6 = NIE). In der Ergebnisliste sehen wir die absoluten Häufigkeiten (Häufigkeit), die relativen Häufigkeiten mit Berücksichtigung eventueller "missing values" (Prozent), die relativen Häufigkeiten ohne "missing values" (Gültige Prozente) und die kumulierten relativen Häufigkeiten (Kumulierte Prozente) für die Variable "v526" (Kirchgangshäufigkeit).

Neben diesen Tabellen bietet SPSS Statistics noch eine Vielzahl anderer Möglichkeiten der Tabellenerstellung (Menüpfad: ANALYSIEREN ⇒ TABELLEN). Diese hat jedoch in SPSS Statistics an Bedeutung verloren, seitdem die Ergebnisse von *FREQUENCIES* im *Viewer* bereits in angemessener Tabellenform erstellt werden.

### 5.5.3 Kreuztabellen (CROSSTABS)

*CROSSTABS* erzeugt zwei- oder mehrdimensionale Assoziations- oder Kontingenztabellen zwischen Variablen. Dabei stehen die abhängigen Variablen zumeist in den Zeilen und die unabhängigen Variablen in den Spalten. In der SPSS Statistics-Syntax muss dafür die abhängige Variable vor und die unabhängige nach einem Schlüsselwort, *BY*, geschrieben werden.

```
CROSSTABS varliste BY varliste [/STATISTICS=[PHI]]
[/CELLS=[COUNT] [ROW] [COLUMN] [TOTAL]].
```

→ Syntax

- STATISTICS** liefert zusätzliche Zusammenhangsmaße; PHI bedeutet z.B., dass ein Assoziationskoeffizient (Maß Cramers "V") für klassifikatorische Merkmale berechnet wird.
- CELLS** gibt an, was in den Feldern der Kreuztabelle stehen soll. Einige Möglichkeiten sind:
  - COUNT** absolute Häufigkeiten,
  - ROW** bedingte relative Häufigkeiten bezogen auf die Zeile,
  - COLUMN** bedingte relative Häufigkeiten bezogen auf die Spalte,
  - TOTAL** gemeinsame relative Häufigkeiten bezogen auf alle Fälle

Für unsere Beispieldaten wird eine Kreuztabelle angefordert, welche die zweidimensionalen Häufigkeitsverteilungen zwischen der klassifikatorischen Variablen Geschlecht des Befragten ("v151") und der komparativen Variablen Schulabschluss des Befragten ("v173") darstellt.

→ Beispiel

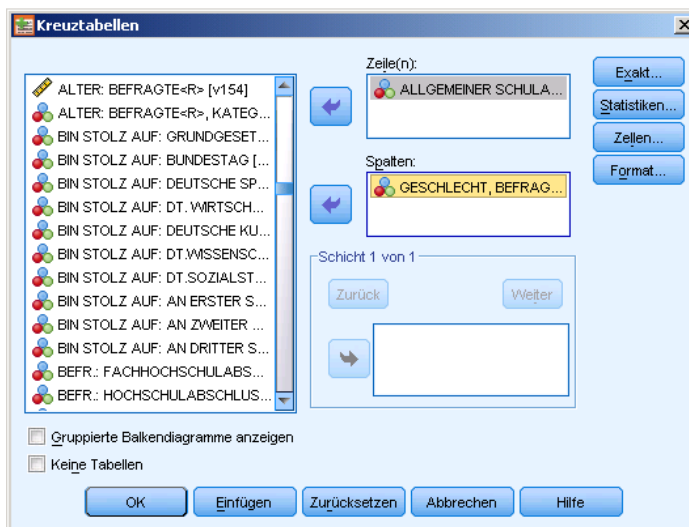
Der Befehl lautet:

```
CROSSTABS v173 BY v151
/CELLS=COUNT ROW COLUMN TOTAL
/STATISTICS=PHI.
```

Menüentsprechung:

ANALYSIEREN ⇒ DESKRIPTIVE STATISTIKEN ⇒ KREUZTABELLEN...

→ Menü

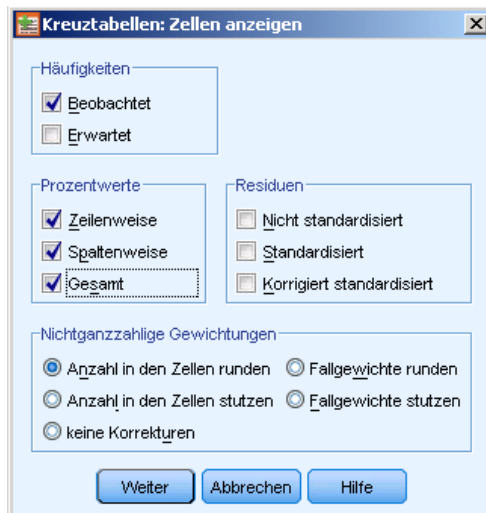


Aus der Variablenliste für die Kreuztabelle wird nun in die Liste **Zeilen** diejenige Variable eingefügt, deren Ausprägungen in der Kreuztabelle in den Zeilen stehen. In die Liste **Spalten** wird dementsprechend die Variable eingefügt, deren Ausprägungen in den Spalten der Kreuztabelle aufgelistet werden.

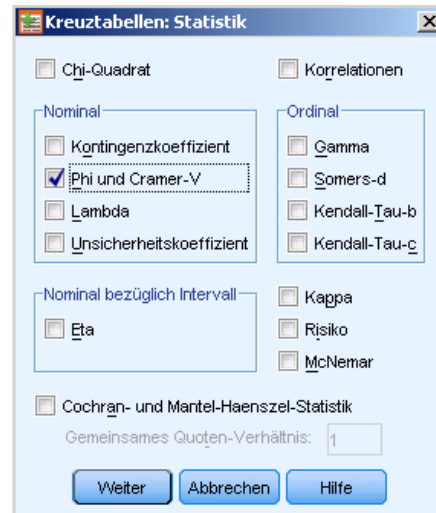
Per Voreinstellung werden in den Feldern einer Kreuztabelle die absoluten Häufigkeiten angegeben. Die Aktivierung der Schaltfläche **Zellen...** öffnet das Dialogfeld *Kreuztabellen: Zellen anzeigen*, in dem weitere Angaben angefordert werden können. In unserem Beispiel werden zusätzlich die Prozentwerte für gemeinsame relative, bedingte relative Häufigkeiten sowie Randverteilungen der Variablen angefordert. Die Optionen  **Zeilweise**,  **Spaltenweise** und  **Gesamt** in dem Bereich **Prozentwerte** müssen aktiviert werden.

Im Dialogfeld *Kreuztabellen: Statistik* der Schaltfläche **Statistik...** können zusätzliche Zusammenhangsmaße berechnet werden. Geeignete Maße für klassifikatorische Daten sind das Maß von Cramer ("V") und die mittlere quadratische Kontingenz ("phi").

Dialogfenster der Schaltfläche: **Zellen...**



Dialogfenster der Schaltfläche: **Statistik...**



Aus einer Kreuztabelle ergibt sich oft ein erster Eindruck davon, ob es einen Zusammenhang zwischen zwei Variablen gibt. Das Ergebnis der Prozedur sieht folgendermaßen aus:

#### Verarbeitete Fälle

	Fälle					
	Gültig		Fehlend		Gesamt	
	N	Prozent	N	Prozent	N	Prozent
ALLGEMEINER SCHULABSCHLUSS * GESCHLECHT, BEFRAGTE<R>	3458,941 <sup>a</sup>	99,7%	10,059	,3%	3469,000	100,0%

a. Die Anzahl der gültigen Fälle unterscheidet sich von der Gesamtzahl in der Kreuztabelle, da die Zellenhäufigkeiten gerundet wurden.

Um den Umfang der Tabelle zu begrenzen, wurde in der folgenden Tabelle ein Stück „ausgeschnitten“.

#### ALLGEMEINER SCHULABSCHLUSS \* GESCHLECHT, BEFRAGTE<R> Kreuztabelle

→ Ergebnis

		GESCHLECHT, BEFRAGTE<R>		Gesamt
		MANN	FRAU	
OHNE ABSCHLUSS	Anzahl	43	50	93
	% innerhalb von ALLGEMEINER SCHULABSCHLUSS	46,2%	53,8%	100,0%
	% innerhalb von GESCHLECHT, BEFRAGTE<R>	2,5%	2,8%	2,7%
	% der Gesamtzahl	1,2%	1,4%	2,7%
VOLKS-,HAUPTSCHULE	Anzahl	614	701	1315
	% innerhalb von ALLGEMEINER SCHULABSCHLUSS	46,7%	53,3%	100,0%
	% innerhalb von GESCHLECHT, BEFRAGTE<R>	36,0%	39,9%	38,0%
	% der Gesamtzahl	17,7%	20,3%	38,0%



MITTLERE REIFE	Anzahl	469	530	999
	% innerhalb von ALLGEMEINER SCHULABSCHLUSS	46,9%	53,1%	100,0%
	% innerhalb von GESCHLECHT, BEFRAGTE<R>	27,5%	30,2%	28,9%
	% der Gesamtzahl	13,6%	15,3%	28,9%
FACHHOCHSCHULREIFE	Anzahl	136	86	222
	% innerhalb von ALLGEMEINER SCHULABSCHLUSS	61,3%	38,7%	100,0%
	% innerhalb von GESCHLECHT, BEFRAGTE<R>	8,0%	4,9%	6,4%
	% der Gesamtzahl	3,9%	2,5%	6,4%
HOCHSCHULREIFE	Anzahl	410	363	773
	% innerhalb von ALLGEMEINER SCHULABSCHLUSS	53,0%	47,0%	100,0%
	% innerhalb von GESCHLECHT, BEFRAGTE<R>	24,1%	20,7%	22,3%
	% der Gesamtzahl	11,8%	10,5%	22,3%

**Symmetrische Maße**

		Wert	Näherungsweise Signifikanz
Nominal- bzgl. Nominalmaß	Phi	,085	,000
	Cramer-V	,085	,000
Anzahl der gültigen Fälle		3460	

Die Kreuztabelle enthält die mit *COUNT* angeforderten absoluten Häufigkeiten (Anzahl). Zum Beispiel sind 614 Befragte männlich und haben einen Haupt- oder Volksschulabschluss. Weiterhin enthält die Kreuztabelle die mit *ROW* angeforderten Zeilenprozent (bedingte relative Häufigkeiten) (% von Allgemeiner Schulabschluss): z. B sind 46,7 % der befragten Hauptschulabsolventen männlich. Die durch *COLUMN* erzeugten Spaltenprozent (% von Geschlecht Befragte<r>) bedeuten z. B., dass 36,0 % der Männer einen Haupt- oder Volksschulabschluss haben. Schließlich sind gemeinsame relative Häufigkeiten angegeben (Prozent der Gesamtzahl), die durch *TOTAL* abgerufen werden. So sind 17,7 % der Befragten gleichzeitig männlich und haben einen Haupt- oder Volksschulabschluss.

### 5.5.4 Deskriptive Statistik (DESCRIPTIVES)

*DESCRIPTIVES* berechnet statistische Maßzahlen für metrische Variablen. Per Voreinstellung sind das: Mittelwert, Standardabweichung, Minimum und Maximum.

```
DESCRIPTIVES varliste
  [/STATISTICS= [MEAN] [STDDEV] [VARIANCE] [MIN]
  [MAX] [ALL]].
```

→ Syntax

**STATISTICS** fordert gezielt Maßzahlen an. Einige wichtige Maßzahlen:

- MEAN** Mittelwert
- STDDEV** Standardabweichung
- VARIANCE** Varianz
- MIN** Minimum
- MAX** Maximum
- ALL** Alle 11 verfügbaren Statistiken

## → Beispiel

Aus den Beispieldaten werden für die Variablen „Alter Befragte<r>“ „v154“ und „Nettoeinkommen Befragte<r>“ „v388“ einige Maßzahlen berechnet.

Der Befehl lautet:

```
DESCRIPTIVES v154 v388
/STATISTICS=MEAN STDDEV VARIANCE MIN MAX.
```

## → Menü

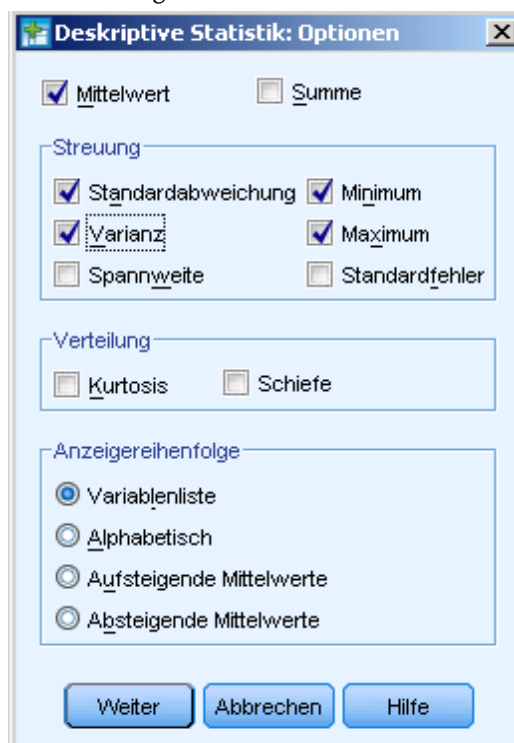
Menüentsprechung:

ANALYSIEREN ⇒ DESKRIPTIVE STATISTIKEN ⇒ DESKRIPTIVE STATISTIKEN...



Um Kennzahlen für die Beschreibung der Lage, der Streuung und der Verteilung von Werten einer Variablen zu berechnen, wird die Variable, für welche die Maßzahlen berechnet werden sollen, mit der Pfeil (↵)-Schaltfläche in die Liste **Variable(n)**: verschoben.

Mit Hilfe der Schaltfläche **Optionen...** gelangt man in das Dialogfeld *Deskriptive Statistik: Optionen*, in dem die zu berechnenden Statistiken ausgewählt werden können.



**Deskriptive Statistik**

	N	Minimum	Maximum	Mittelwert	Standardabweichung	Varianz
ALTER: BEFRAGTE<R>	3455	18	97	49,94	17,797	316,723
BFR.:NETTOEINKOMMEN<OF FENE+LISTENANGABE>	2750	1	8750	1411,47	1090,589	1189384,501
Gültige Werte (Listenweise)	2746					

→ Ergebnis

Die Ergebnisliste gibt für die einzelnen Variablen spaltenweise die Anzahl der zugrunde liegenden validen Fälle (N), den kleinsten Wert (Minimum), den größten Wert (Maximum), den arithmetischen Mittelwert, die Standardabweichung und die Varianz an. Wenn viele Statistiken angefordert werden, erfolgt deren Ausgabe nicht mehr spaltenweise, sondern neben- und untereinander.

### 5.5.5 Korrelationen (CORRELATIONS)

*CORRELATIONS* berechnet für die Variablen der "varliste" eine Korrelationsmatrix (Korrelationskoeffizienten nach Bravais/Pearson) und Signifikanzwerte für zweiseitige Tests der Korrelationskoeffizienten ( $H_0: R=0$  oder  $H_1: R \neq 0$ ). Der Test ist nur unter einer Reihe von Annahmen zulässig (einfache Zufallsstichprobe aus einer zweidimensionalen Normalverteilung).<sup>15</sup> Diese Voraussetzungen wären für unsere Beispieldaten zunächst zu prüfen, darauf wird hier aber nicht näher eingegangen.

```
CORRELATIONS varliste.
```

→ Syntax

Es wird in unserem Beispiel vermutet, dass ein positiver linearer Zusammenhang zwischen dem Alter und der Höhe des Einkommens einer befragten Person besteht. Als Zusammenhangsmaß für metrische Merkmale wird deshalb der Korrelationskoeffizient nach Bravais/Pearson berechnet. Im *Syntax-Editor* wird die Prozedur *CORRELATIONS* eingegeben. Es sind lediglich die Variablen anzugeben, zwischen denen der Koeffizient berechnet werden soll.

Der Befehl lautet:

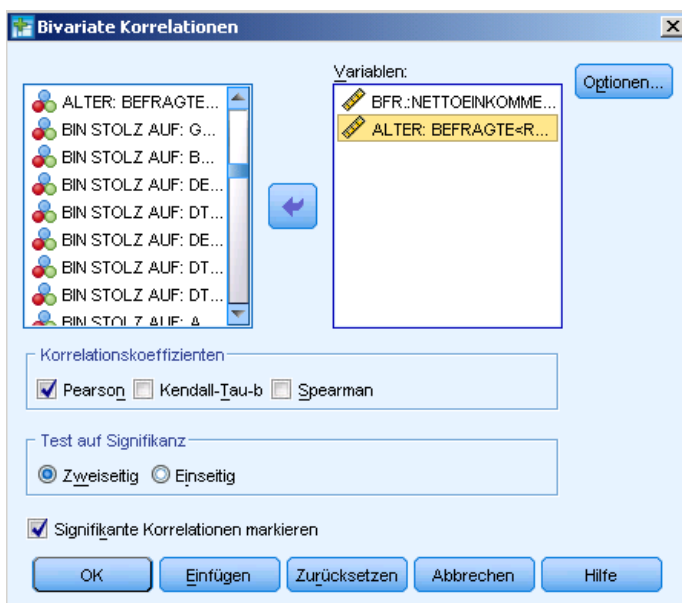
```
CORRELATIONS v388 v154.
```

Menüentsprechung:

ANALYSIEREN ⇒ KORRELATION ⇒ BIVARIAT...

→ Beispiel

→ Menü



<sup>15</sup> Siehe dazu Vogel (1989a), S. 73.

In der Variablenliste werden nur die numerischen Variablen aufgeführt. Um Korrelationskoeffizienten berechnen zu können, müssen mindestens zwei Variablen ausgewählt werden. Per Voreinstellung wird der Pearson'sche Korrelationskoeffizient berechnet.

## → Ergebnis

Das Ergebnis sieht folgendermaßen aus:

Korrelationen			
		BFR.: NETTOEINKOM MEN<OFFENE+L ISTENANGABE>	ALTER: BEFRAGTE<R>
BFR.:NETTOEINKOMMEN<OF FENE+LISTENANGABE>	Korrelation nach Pearson Signifikanz (2-seitig) N	1  2750	,052  2746
ALTER: BEFRAGTE<R>	Korrelation nach Pearson Signifikanz (2-seitig) N	,052  2746	1  3455

Die Korrelationsmatrix zeigt die jeweiligen Korrelationskoeffizienten nach Bravais/Pearson. Sie messen die Stärke und Richtung des linearen Zusammenhangs zwischen dem Nettomonatseinkommen und dem Alter des/der Befragten. Der Wert des Korrelationskoeffizienten liegt immer zwischen -1 und +1. In unserem Fall beträgt er 0,052. Dies bedeutet, dass der lineare Zusammenhang zwischen den beiden Variablen Nettoeinkommen und Alter nur „sehr schwach“ positiv ist. Zusätzlich ist die Anzahl der Fälle (N) angegeben, die zur Berechnung des Korrelationskoeffizienten herangezogen wurden.

## 5.5.6 Mittelwertvergleiche (MEANS)

MEANS berechnet per Voreinstellung Mittelwerte, Standardabweichungen und Anzahl valider Fälle für Subgruppen von metrischen Variablen.

## → Syntax

<b>MEANS</b>	<b>varliste BY varliste [/CELLS= [MEAN] [COUNT][STDDEV][SUM][VARIANCE]].</b>
--------------	--

Links vom Schlüsselwort **BY** sind die metrischen Variablen aufzuführen. Die Merkmalsausprägungen der Variablen rechts vom Schlüsselwort **BY** definieren die Subgruppen.

<b>CELLS</b>	<b>MEAN</b>	Mittelwert
	<b>STDDEV</b>	Standardabweichung
	<b>VARIANCE</b>	Varianz
	<b>SUM</b>	Summe
	<b>COUNT</b>	Anzahl

## → Beispiel

Für die Beispieldaten werden einige über **MEANS** verfügbaren Maßzahlen für die Variablen „Alter Befragte<r>“ („v154“) und „Nettoeinkommen Befragte<r>“ („v388“) berechnet, und zwar getrennt nach Geschlechtern („v151“).

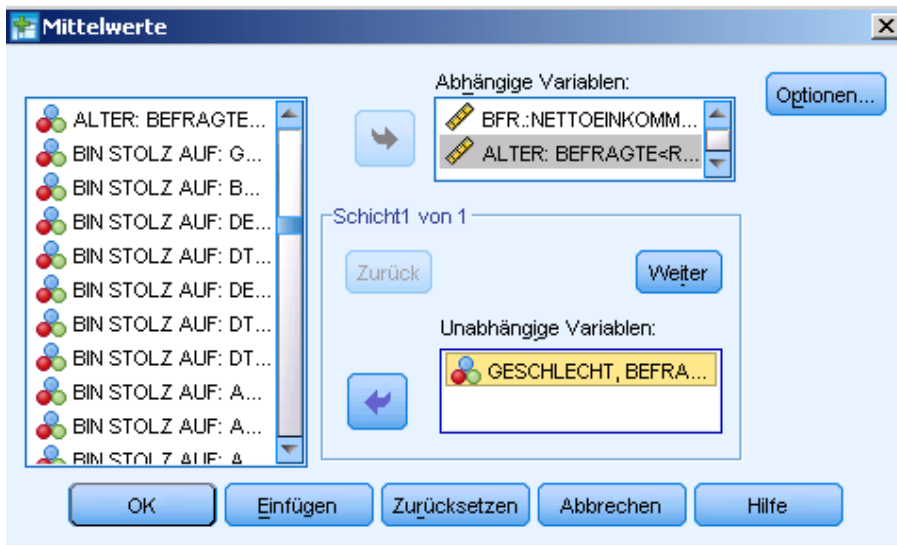
Der Befehl lautet:

```
MEANS v388 v154 BY v151
/CELLS = mean count stddev sum variance.
```

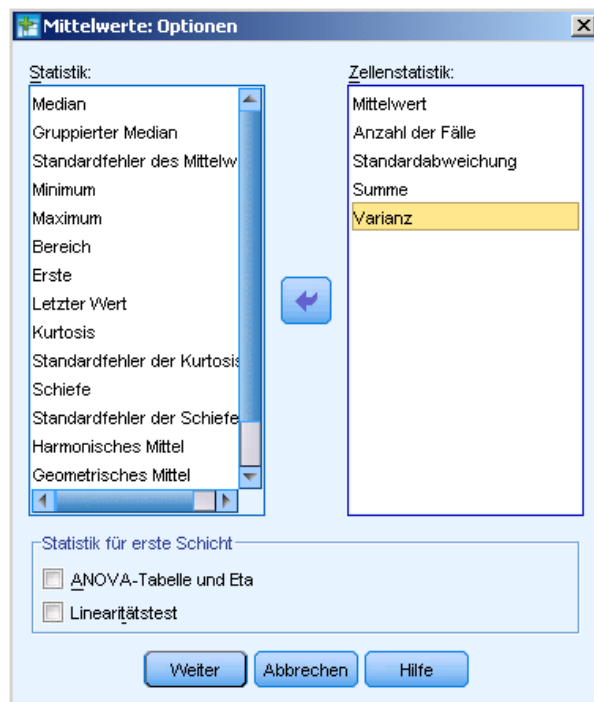
## → Menü

Menüentsprechung:

ANALYSIEREN ⇨ MITTELWERTE VERGLEICHEN ⇨ MITTELWERTE...



In der Variablenliste werden alle numerischen und String-Variablen des Datensatzes aufgeführt. In die Liste **Abhängige Variablen:** werden nun die Variablen verschoben, deren Mittelwerte für die Gruppen berechnet werden sollen. In diesem Beispiel sind es das Alter und das Nettoeinkommen des oder der Befragten. Mit der Pfeil (↵)-Schaltfläche wird in das Feld **Unabhängige Variablen:** die Variable eingefügt, durch deren Merkmalsausprägungen die Gruppen gebildet werden sollen. In diesem Fall teilt sich die Variable "Geschlecht" mit den Subgruppen "weiblich" und "männlich". Neben den Gruppenmittelwerten können noch weitere statistische Maßzahlen berechnet werden, über die Schaltfläche **Optionen...**, die zu dem Dialogfeld *Mittelwerte: Optionen* führt:



Nachdem wir Summe und Varianz als gewünschte Statistiken auf die rechte Seite gebracht haben, bekommen wir folgendes Ergebnis:

**Verarbeitete Fälle**

	Fälle					
	Eingeschlossen		Ausgeschlossen		Insgesamt	
	N	Prozent	N	Prozent	N	Prozent
BFR.:NETTOEINKOMMEN<OFFENE+LISTENANGABE> * GESCHLECHT, BEFRAGTE<R>	2750	79,3%	719	20,7%	3469	100,0%
ALTER: BEFRAGTE<R> * GESCHLECHT, BEFRAGTE<R>	3455	99,6%	14	,4%	3469	100,0%

➔ Ergebnis

## Bericht

GESCHLECHT, BEFRAGTE<R>		BFR.:NETTOEINKOMMEN<OFFENE+LISTENANGABE>	ALTER: BEFRAGTE<R>
MANN	Mittelwert	1751,28	49,32
	N	1421	1705
	Standardabweichung	1204,345	17,619
	Summe	2487754	84107
	Varianz	1450446,411	310,436
FRAU	Mittelwert	1048,37	50,55
	N	1329	1750
	Standardabweichung	809,876	17,952
	Summe	1393746	88471
	Varianz	655898,732	322,287
Insgesamt	Mittelwert	1411,47	49,94
	N	2750	3455
	Standardabweichung	1090,589	17,797
	Summe	3881500	172578
	Varianz	1189384,501	316,723

Zunächst wurden für die männlichen Befragten, dann für die Fälle mit der Ausprägung 'weiblich' und schließlich für alle Fälle (Insgesamt) folgende statistischen Maßzahlen bestimmt: arithmetischer Mittelwert, N und Standardabweichung, Summe und Varianz. 'N' gibt an, wie viele Fälle zur Berechnung herangezogen wurden. In der linken Spalte beziehen sich die Berechnungen auf das Nettoeinkommen, in der Rechten auf das Alter.

## 6 Datenmodifikationen

Voraussetzung für die Datenanalyse ist häufig eine geeignete Modifikation der Daten. Datenmodifikation bedeutet in diesem Zusammenhang, dass die Daten vor der Analyse durch bestimmte Operationen verändert werden. Es seien nur einige Situationen genannt, in denen Datenmodifikationen erforderlich sind:

- Kodierfehler werden beseitigt.
- Aus mehreren Variablen wird ein Index berechnet.
- Eine ungünstige Skalierung von Variablen wird geändert.
- Zur Überprüfung der Korrektheit der Daten werden Kontrollrechnungen durchgeführt.

Für die praktische Arbeit ist nur die Modifikation numerischer Variablen von Bedeutung. Daher wird auf die Darstellung der Modifikationsmöglichkeiten von Zeichenkettenvariablen verzichtet.

Bei der Datenmodifikation kann die menügestützte Eingabe der Befehle nicht ausführlich dargestellt werden, da häufig eine ganze Reihe von Dialogfeldern zu durchlaufen sind, um das gleiche Ergebnis zu erhalten. Es wird generell empfohlen, hier Befehle einzusetzen, da diese einfacher zu handhaben sind. Am Schluss eines jeden Abschnittes wird kurz auf die Dialogfenster verwiesen.

### 6.1 Modifikation numerischer Variablen

Die wichtigsten Sprachelemente im Rahmen der Datenmodifikation numerischer Variablen sind die Kommandos *RECODE* und *COMPUTE*.

Bei der Anwendung beider Modifikationen ist es wichtig zu wissen, dass die Ergebnisse der Befehle in der Datenansicht des *Daten-Editors* erst dann sichtbar werden, wenn im Anschluss die erste Analyse angefordert wird. Dies kann beispielsweise mit Hilfe einer Prozedur, wie etwa *FREQUENCIES* oder *DESCRIPTIVES*, vorgenommen werden oder über das *EXECUTE*-Kommando, das vorangegangene, noch nicht abgearbeitete, Modifikationen lediglich ausführt, aber keinerlei Ausgabe erzeugt. Solange keine Analysen aufgerufen werden, erscheint rechts unten im Datenfenster ein Hinweis auf „Offene Transformationen“, der daran erinnert, dass zu einem früheren Zeitpunkt Variablen verändert wurden. Modifikationen über Menüs werden sofort ausgeführt.

→ Hinweis!

#### 6.1.1 Rekodierung (RECODE)

Das *RECODE*-Kommando ändert die Skalierung der Variablen. Die Werte einer "werteliste" werden durch *RECODE* zu einem neuen "wert" rekodiert.

Das *RECODE*-Kommando hat folgende Struktur:

```
RECODE varliste1 (werteliste = wert) ... (werteliste = wert)
      ... [{werteliste}{MISSING}{SYSMIS}{ELSE} = {wert}{SYSMIS}{COPY}]
      [INTO varliste2] [/varliste ...].
```

→ Syntax

- varliste1** ist der Name einer oder mehrerer bereits definierter Variablen.
- werteliste** ist eine Reihe von Variablenwerten, die durch Komma oder Leerzeichen getrennt sind (mindestens ein Wert). Zur Bildung von Wertebereichen werden, wie beim *MISSING VALUES*-Kommando, die Schlüsselworte *LO* (LOWEST), *HI* (HIGHEST) und *THRU* eingesetzt. Wertebereiche und Einzelwerte lassen sich beliebig kombinieren. Ihre Anzahl ist nicht begrenzt. *LO* und *HI* schließen "user-missing values", aber nicht "system-missing values" ein. Wertebereiche beinhalten den Anfangs- und den Endpunkt. Um zu vermeiden, dass bei kontinuierlichen Variablen Werte nicht rekodiert werden, kann es günstig sein, wenn der Endpunkt einer Werteliste und der Anfangspunkt der nächsten Werteliste identisch sind ((*LO THRU 4000=...*) (*4000 THRU HI=...*)). In den folgenden Beispielen wird deutlich, dass darauf zu achten ist, welcher Wert zuerst genannt wird.

Beispiel 1:

*RECODE v388 (LO THRU 4000=1) (4000 THRU HI=2).*

In diesem Fall wird das Nettoeinkommen von 4000 zu 1 rekodiert.

Beispiel 2:

*RECODE v388 (4000 THRU HI=2) (LO THRU 4000=1).*

In diesem Fall wird das Nettoeinkommen von 4000 zu 2 rekodiert.

Bei der Verwendung von *THRU* muss der niedrigere Wert des Wertebereiches links stehen, zum Beispiel (4 *THRU* 9). (9 *THRU* 4) ist falsch.

Anstelle der "werteliste" sind die Schlüsselworte *ELSE*, *MISSING* und *SYSMIS* zulässig.

*ELSE=* bezieht sich auf alle Werte, die in keiner vorhergehenden Werteliste aufgeführt sind (Auffangkategorie). *ELSE* darf daher nur vor der letzten zu einer "varliste" gehörigen Werteliste stehen. Alle folgenden Wertelisten werden ignoriert. *ELSE* beinhaltet den "system-missing value" und den "user-missing value".

*MISSING=* dient der Rekodierung von "user-missing values" und "system-missing values". Wird mit *MISSING* rekodiert, sind die neuen Werte nicht automatisch fehlende Werte.

*SYSMIS=* dient allein der Rekodierung von "system-missing values".

**wert** ist der neue Variablenwert, in den alle Werte der "werteliste" rekodiert werden. Es ist nur ein einziger numerischer Wert erlaubt. Das Schlüsselwort *SYSMIS* als Wert wandelt die Werte der "werteliste" in den "system-missing value" um. Das Schlüsselwort *COPY* als Werte kopiert die Werte der „werteliste“ einfach.

**INTO** als Spezifikation ermöglicht es, die Werte der Variablen der "varliste1" zu rekodieren und das Ergebnis in den Variablen der "varliste2" abzulegen. Ohne die Angabe von *INTO* werden die Variablen der "varliste1" direkt verändert. Die "varliste2" muss genauso viele Elemente besitzen wie die "varliste1". Die Variablen werden der Reihenfolge nacheinander zugeordnet. Die (Ziel-) "varliste2" kann aus bereits definierten oder neuen Variablen bestehen. Neue Variablen in diesem Sinne sind solche, die vor dem Auftreten des *RECODE*-Kommandos noch nicht definiert sind. Sie werden also erstmals in der "varliste2" aufgeführt.

Fälle mit Werten, die nicht in einer "werteliste" aufgeführt sind, bleiben bei bereits definierten Variablen der "varliste2" unverändert. Bei neuen Variablen der "varliste2" erhalten diese Fälle den "system-missing value".

*COPY* als „wert“ und *ELSE* als "werteliste" bieten allerdings die Möglichkeit, alle nicht zu rekodierenden Werte einfach zu übertragen, ohne dass diese verändert werden. (*ELSE=COPY*) steht dann als letzte Rekodierungsregel. *COPY* transferiert auch "user-missing values" und "system-missing values". Der Missing-Status der "user-missing values" geht dabei verloren.

Zunächst einige fiktive Beispiele, um die Möglichkeiten des *RECODE*-Kommandos zu demonstrieren.

```
RECODE  x1 TO x5 (2=3) (4=5)
        /x7 (LO THRU 10=1).
RECODE  x8 (0=1) (1=0) (ELSE=SYSMIS).
RECODE  x10 (0=1) (2=0) (MISSING=99).
RECODE  y1 TO y3 (2 THRU 7, 8, 10 THRU HI=5) (ELSE=COPY)
        INTO y4 TO y6.
```

→ **Beispiel**

Nun ein Beispiel mit unserem Datensatz. Die Variable "v388" (Nettoeinkommen Befragte<r>) wird klassiert. Es sind fünf Einkommensklassen zu bilden. Die rekodierten Werte werden in der neuen Variablen "nmekl" abgelegt. Diese Variable erhält dann ein *VARIABLE LABEL* und ihre Werte *VALUE LABELS*. Für die neue Variable "nmekl" wird letztlich noch die Häufigkeitsverteilung der Variablenwerte berechnet.

→ **Hinweis!**

Die im folgenden Beispiel verwendeten Schlüsselworte *LOW* und *HI* schließen bei der Rekodierung "user-missing values" ein. Mit dem Schlüsselwort *MISSING* in der ersten Werteliste werden diese



deshalb zunächst zur 9 rekodiert. Damit die 9 in *nmekl* ein fehlender Wert wird, muss ein *MISSING VALUES*-Kommando vorhanden sein!

```

RECODE   v388           (MISSING=9)
                    (LOW THRU 2000 = 1)
                    (2000 THRU 3000 = 2)
                    (3000 THRU 4000 = 3)
                    (4000 THRU 5000 = 4)
                    (5000 THRU HI   = 5)
                    INTO nmekl.
VARIABLE LABEL nmekl 'klassiertes Einkommen'.
VALUE LABELS  nmekl 1 'bis 2000'
                    2 '2000 bis 3000'
                    3 '3000 bis 4000'
                    4 '4000 bis 5000'
                    5 'über 5000'.
MISSING VALUES nmekl (9).
FREQUENCIES  nmekl.
    
```

Das Ergebnis der Rekodierung kann nur mit Hilfe einer Analyse (wie *FREQUENCIES*) oder des *EXECUTE*-Kommandos (*EXECUTE*.) dargestellt werden:

→ Ergebnis

**Statistiken**

klassiertes Einkommen

N	Gültig	2750
	Fehlend	719

**klassiertes Einkommen**

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	bis 2000	2250	64,9	81,8	81,8
	2000 bis 3000	321	9,3	11,7	93,5
	3000 bis 4000	101	2,9	3,7	97,2
	4000 bis 5000	43	1,2	1,5	98,7
	über 5000	36	1,0	1,3	100,0
	Gesamt	2750	79,3	100,0	
Fehlend	9,00	719	20,7		
Gesamt		3469	100,0		

Die Rekodierungen werden erst sichtbar, wenn eine Prozedur oder *EXECUTE* ausgeführt wird oder die Daten gespeichert werden. Vorher sind die rekodierten Daten im *Daten-Editor* nicht zu sehen.

→ Hinweis!

Rekodierungen werden fallweise ausgeführt. Bei jedem Fall wird geprüft, ob der aktuelle Variablenwert in einer "werteliste" vorkommt. Falls ja, wird der Wert rekodiert, falls nein, bleibt er bestehen.

Die Spezifikationen "(werteliste=wert)" werden von links nach rechts ausgewertet. Ein Wert wird pro *RECODE*-Kommando immer nur einmal rekodiert. Die Anweisung

```
RECODE x1 (0=1) (1=0).
```

vertauscht daher bei jedem Fall die Codes 0 und 1. Bei kontinuierlichen Variablen führt der Befehl

```
RECODE laenge (17 THRU HI=2) (LO THRU 17=1).
```

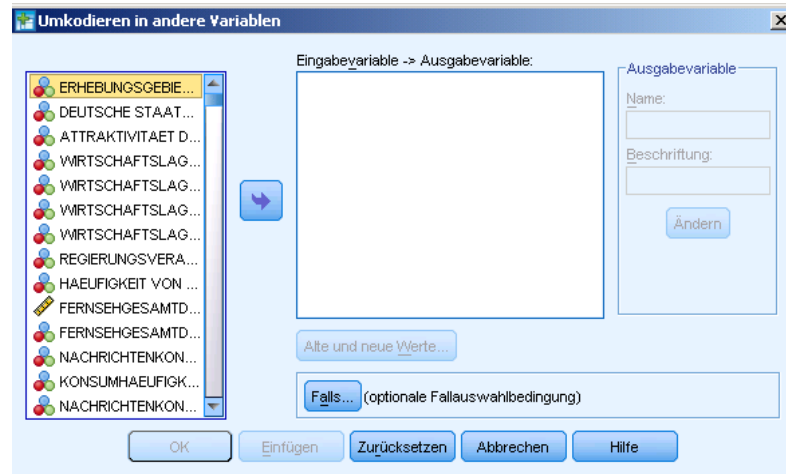
dazu, dass ein Fall mit dem Wert 17 in die Kategorie 2 fällt.

Menüentsprechung:

Sollen die Werte einer Variablen umkodiert werden und zusätzlich auch in ihrer ursprünglichen Form erhalten werden, dann wird hier der Menüpunkt:

→ Menü

TRANSFORMIEREN ⇨ UMKODIEREN IN ANDERE VARIABLEN... gewählt.



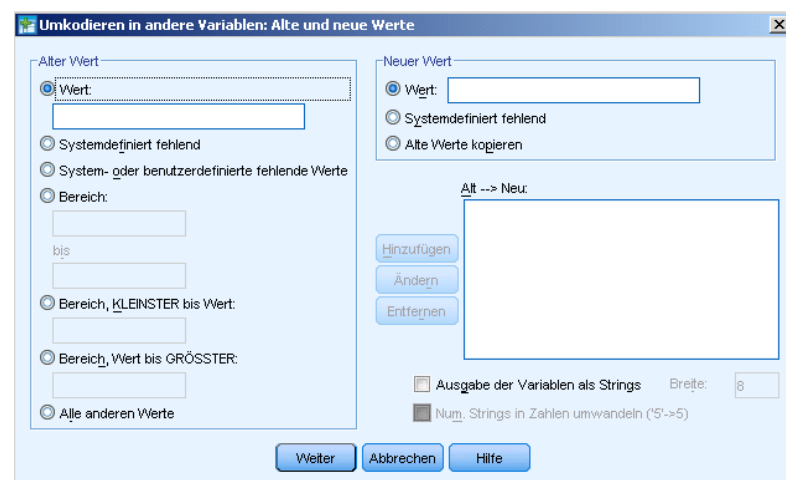
Mit dem Befehl

TRANSFORMIEREN ⇨ UMKODIEREN IN DIESELBEN VARIABLEN...

öffnet sich das Dialogfenster, um Werte innerhalb einer Variablen umzukodieren.



Über die Schaltfläche **Alte und neue Werte...** gelangt man in das Dialogfeld *Umkodieren in ... Variablen: Alte und neue Werte*. Auf der linken Seite sind hierbei die Ausgangswerte einzugeben (als Einzelwerte oder Wertebereich) und rechts die Zielwerte. Über die Schaltfläche **Hinzufügen** werden die Werte in die Liste der zu verändernden Werte übernommen. Die Schaltfläche **Weiter** schließt die Eingabe ab.



→ Hinweis!

Im Gegensatz zur Syntax-Variante werden die vorgenommenen Veränderungen sofort sichtbar, nachdem im Dialogfenster *Umkodieren in ... Variablen* die Schaltfläche **OK** betätigt wurde.

## 6.1.2 Das COMPUTE-Kommando

### 6.1.2.1 Die Befehlsstruktur

Im Laufe einer Datenanalyse sind häufig Variablen zu bilden, die sich aus arithmetischen Operationen mit vorhandenen Variablen ergeben. Hier ist das *COMPUTE*-Kommando dienlich.

→ Syntax

**COMPUTE variable = arithmetischer Ausdruck.**

**variable** (Zielvariable) ist der Name einer bereits definierten oder neuen Variablen. Bei jedem Fall wird der Zielvariablen der Wert zugewiesen, der sich als Ergebnis der Berechnung des arithmetischen Ausdrucks ergibt. Existiert die Zielvariable bereits, werden die "alten" Variablenwerte ersetzt. Handelt es sich um eine neue Variable, so wird sie erzeugt und im Arbeitsspeicher an die Variablenliste hinten angefügt. Alle Fälle neuer Variablen werden zunächst mit dem "system-missing value" vorbesetzt (initialisiert) und dann den Berechnungen entsprechend abgeändert.

**arithmetischer Ausdruck** setzt sich aus folgenden Komponenten zusammen:

- Variablennamen und Konstanten,
- arithmetischen Operatoren (+, -, \*, /, \*\*),
- arithmetischen Funktionen,
- statistischen Funktionen,
- Datums- und Zeitfunktionen,
- logischen Funktionen,
- Funktionen für fehlende Werte,
- Funktionen,
- Klammern.

Um die Übersichtlichkeit zu erhöhen, dürfen die einzelnen Komponenten eines arithmetischen Ausdrucks durch Leerzeichen getrennt werden.

Die (möglichen) Komponenten eines arithmetischen Ausdrucks werden im Folgenden einzeln erörtert.

Es wird an dieser Stelle betont, dass das *COMPUTE*-Kommando fallweise wirkt. Die Berechnungen des arithmetischen Ausdrucks werden also für jeden Fall einzeln ausgeführt. Obwohl das Kommando nur aus einer Zeile besteht, werden die Berechnungen trotzdem für alle Fälle gemacht.

→ Hinweis!

### 6.1.2.2 Variablennamen und Konstanten

Variablennamen sind Namen bereits definierter Variablen. Konstanten (numerische) sind Zahlen ("numbers"). Sie setzen sich aus den Ziffern 0 bis 9 und, sofern notwendig, dem Dezimalpunkt zusammen. Beispiele für Konstanten sind: 1 7 8.9 1234.567.

Im einfachsten Fall besteht ein arithmetischer Ausdruck nur aus einem Variablennamen oder einer Konstanten.

### 6.1.2.3 Arithmetische Operatoren

Arithmetische Operatoren verknüpfen die Komponenten eines arithmetischen Ausdrucks. Sie stehen deshalb "zwischen" den Komponenten. Arithmetische Operatoren und ihre Bedeutung sind:

Operator	Bedeutung
+	Addition
-	Subtraktion
*	Multiplikation
/	Division
**	Potenzierung

SPSS Statistics führt die arithmetischen Operationen in folgender Reihenfolge aus:

\*\*   \*   /   +   -

Operationen auf einer Ebene werden von links nach rechts durchgeführt. Ausdrücke in Klammern werden vorrangig ausgewertet. Das Setzen von Klammern kann daher eine andere Bearbeitungsreihenfolge erzwingen.

Zunächst ein fiktives Beispiel, um einen "längeren" arithmetischen Ausdruck zu zeigen:

```
COMPUTE neu = (a + b - c)*d + e/f**5 - (4.4567 + x**c).
```

## → Beispiel

Für unsere Beispieldaten wollen wir ermitteln, um wie viel das Haushaltseinkommen höher ist als das Nettoeinkommen des oder der Befragten. Dazu subtrahieren wir bei jedem Fall das Nettoeinkommen des/der Befragten („v388“) vom Haushaltseinkommen („v448“). Das Ergebnis weisen wir der neuen Variablen „zeink“ zu. Diese erhält dann noch ein Label, „Zusatzeinkommen“. Zur Demonstration werden anschließend die Korrelationen zwischen den Variablen „v388“ (Nettoeinkommen Befragte<r>), „zeink“ (Zusatzeinkommen) und „v154“ (Alter Befragte<r>) berechnet. Hier fordert *STATISTICS=DESCRIPTIVES* zusätzlich Mittelwert (*MEAN*) und Standardabweichung (*STD DEV*) der Variablen an. Da „v388“ und „v448“ fehlende Werte besitzen, kann „zeink“ für manche Fälle nicht berechnet werden. „zeink“ wird daher für diese Fälle den „system-missing value“ zugewiesen (Vergleiche hierzu auch Seite 53). Die Korrelationen basieren daher auf unterschiedlichen Fallzahlen.

Der Befehl lautet:

```
COMPUTE zeink=v448 - v388.
```

```
VARIABLE LABELS zeink Zusatzeinkommen.
```

```
CORRELATION v388 v154 zeink
```

```
/STATISTICS=DESCRIPTIVES.
```

### Deskriptive Statistiken

	Mittelwert	Standardabweichung	N
BFR.:NETTOEINKOMMEN<OFFENE+LISTENANGABE>	1411,47	1090,589	2750
ALTER: BEFRAGTE<R>	49,94	17,797	3455
Zusatzeinkommen	948,7005	1246,86081	2564

## → Ergebnis

### Korrelationen

		BFR.:NETTOEINKOMMEN<OFFENE+LISTENANGABE>	ALTER: BEFRAGTE<R>	Zusatzeinkommen
BFR.:NETTOEINKOMMEN<OFFENE+LISTENANGABE>	Korrelation nach Pearson	1	,052	-,104
	Signifikanz (2-seitig)		,006	,000
	N	2750	2746	2564
ALTER: BEFRAGTE<R>	Korrelation nach Pearson	,052	1	-,205
	Signifikanz (2-seitig)	,006		,000
	N	2746	3455	2560
Zusatzeinkommen	Korrelation nach Pearson	-,104	-,205	1
	Signifikanz (2-seitig)	,000	,000	
	N	2564	2560	2564

## 6.1.2.4 Funktionen

Alle **Funktionen** besitzen die Struktur:

<b>funktionsname (argument)</b>
---------------------------------

**argument** ist der Ausdruck, den die Funktion "funktionsname" transformiert. Als Ergebnis dieser Transformation liefert die Funktion immer eine Zahl ("number") oder den "system-missing value". Argumente können arithmetische Ausdrücke, Variablennamen, Konstanten oder Funktionen sein.

Eine Funktion besitzt entweder eine feste oder eine variable Anzahl von Argumenten. Um Missverständnisse zu vermeiden, wird dann, wenn nur ein Argument gemeint ist, dieses Argument durch "arg" symbolisiert. "argliste" hingegen, weist auf eine variable Anzahl von Argumenten hin. Argumente einer "argliste" werden durch Komma getrennt.

SPSS Statistics unterscheidet dabei folgende Funktionstypen:

⇒ **Arithmetische Funktionen** führen bestimmte "vorgefertigte" Berechnungen aus. Ein Beispiel wäre die Funktion „ $SQRT(arg)$ “ zum Ziehen der Wurzel.

Fiktives Beispiel zur Demonstration eines *COMPUTE* mit arithmetischen Funktionen:

$COMPUTE\ x = SQRT(x7)$  .

⇒ **Statistische Funktionen** berechnen aus den Argumenten statistische Maßzahlen. Beispielsweise berechnet die Funktion „ $VARIANCE(argliste)$ “ die Varianz der Argumente.

Fiktives Beispiel:

$COMPUTE\ xx = VARIANCE(x1, x2, x7, x9)$ .

⇒ **Logische Funktionen** liefern als Ergebnis den Wert 1 (wahr), 0 (falsch) oder den "system-missing value".<sup>16</sup>

Fiktives Beispiel:

$COMPUTE\ xx = ANY(x1,1,2,3,4)$ .

⇒ **Datums- und Zeitfunktionen** ermöglichen es, mit einem Datum oder Zeitintervallen zu arbeiten.  $YRMODA(Jahr, Monat, Tag)$  beispielsweise berechnet die Anzahl der Tage zwischen dem 15. Oktober 1582, dem ersten Tag des Gregorianischen Kalenders und dem Tag, der durch die Argumente "(Jahr, Monat, Tag)" definiert ist. Diese Funktion kann eingesetzt werden, um die Differenz (in Tagen) zwischen zwei Zeitpunkten zu berechnen.

Ein Datum ist (bei SPSS Statistics) die Anzahl der Sekunden zwischen dem 14. Oktober 1582 Mitternacht (24.00 Uhr) und einem späteren Zeitpunkt. Wegen der Umstellung des Julianischen auf den Gregorianischen Kalender war der Tag vor dem 15. Oktober 1582 der 4. Oktober 1582. Die Umstellung des Kalenders wurde allerdings nicht einheitlich vorgenommen. Da sie von der katholischen Kirche veranlasst war, schlossen sich protestantische Regionen und damit viele Länder, erst später an.

### 6.1.2.5 Missing Values bei arithmetischen Ausdrücken

Zum Abschluss des *COMPUTE*-Kommandos, ist noch die Behandlung von fehlenden Werten im "arithmetischen Ausdruck" zu erörtern:

Wenn eine der im arithmetischen Ausdruck aufgeführten Variablen einen fehlenden Wert hat, ist das Ergebnis dieses arithmetischen Ausdrucks fast immer der "system-missing value", da dann der Ausdruck unbestimmt ist (Ausnahme siehe unten). Ebenso bedingen unerlaubte arithmetische Operationen (Division durch 0) "system-missing values" im Ergebnis. Einige Operationen im Zusammenhang mit der Konstanten 0 führen auch mit fehlenden Werten immer zu den gleichen (validen) Ergebnissen:

$0^*missing$	= 0
$0/missing$	= 0
$missing^{**}0$	= 1
$0^*missing$	= 0
$MOD(0,missing)$	= 0

<sup>16</sup> Diese Funktionen werden hier nicht weiter erörtert, da sie für spezielle Anwendungen vorgesehen sind. Siehe dazu SPSS (1990), S.19f.

### 6.1.2.6 Missing Values bei Funktionen

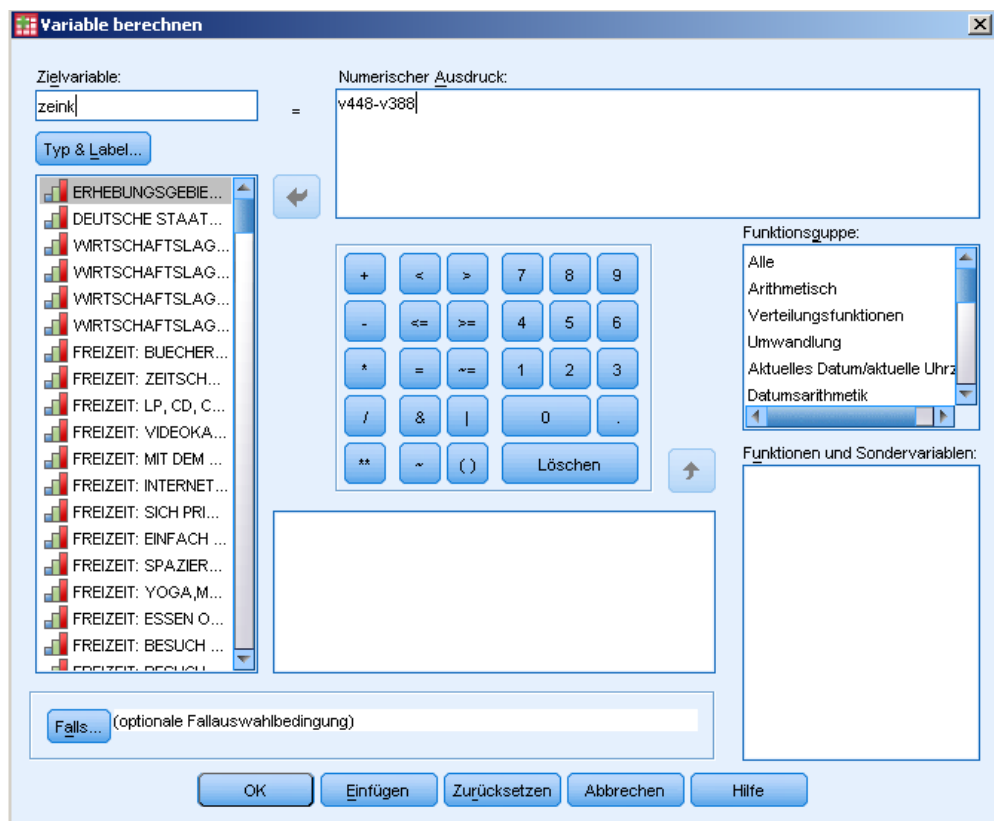
Arithmetische Funktionen liefern für Argumente, die nicht dem Definitionsbereich der Funktion angehören, wie beispielsweise  $SQRT(-1)$ , als Ergebnis "system-missing values".

Funktionen mit mehreren Argumenten werden immer dann ausgewertet, wenn dies mit den Informationen, welche die Argumente liefern, möglich ist. Daher erzeugt eine Funktion unter Umständen selbst dann numerische Ergebnisse, wenn ein Teil der Argumente einen fehlenden Wert annimmt. Es genügt beispielsweise eine Variable mit valider Information, damit die statistische Funktion  $SUM(x1, x2, x3)$  berechnet wird.<sup>17</sup> Es lässt sich bei statistischen Funktionen allerdings steuern, wie viele valide Argumente vorhanden sein müssen, damit die Funktion noch ausgewertet wird.  $SUM.n(argliste)$  bedeutet, dass mindestens  $n$  Argumente valide Informationen liefern müssen, damit die Summe berechnet wird.  $SUM.2(x1, x2, x3)$  heißt demnach: Zumindest zwei Variablen dürfen weder "user-" noch "system-missing values" besitzen, damit ausgewertet wird.

### 6.1.2.7 Berechnen über Menü

Um Berechnungen über ein Menü auszuführen, muss TRANSFORMIEREN ⇒ VARIABLE BERECHNEN angewählt werden, und es erscheint das Fenster *Variable berechnen*. Zielvariable, numerischer Ausdruck und Funktionen wurden bereits oben ausführlich erörtert.

→ Menü



### 6.1.3 Das COUNT-Kommando

Ein weiterer Befehl zur Modifikation numerischer Variablen ist das *COUNT*-Kommando. *COUNT* zählt aus, wie oft bei einem Fall ein Wert der "werteliste" in den Variablen der "varliste" vorkommt, und weist das Ergebnis einer Zielvariablen zu. Es wird demnach bei einem Fall über die Variablen der "varliste" hinweg gezählt.

→ Syntax

```
COUNT   variable = varliste (werteliste) ...
        [/variable = ...].
```

<sup>17</sup> Siehe dazu SPSS (1990), S. 26, 27.

- variable** ist der Name einer existierenden oder neuen Variablen.
- varliste** ist eine Liste von bereits definierten Variablen (mindestens eine Variable). In der Variablenliste kann eine Variable mehr als einmal vorkommen. Sie wird dann mehrfach gewichtet. Wenn zum Beispiel eine Variable in der "varliste" zweimal aufgeführt ist, dann erhält sie doppeltes Gewicht.
- werteliste** ist eine Reihe von durch Komma oder Leerzeichen getrennten Einzelwerten oder Wertebereichen. Wertebereiche werden durch Einsatz des Schlüsselwortes *THRU* definiert (oft in Verbindung mit *LO* oder *HI*). Wertebereiche und Werte dürfen in beliebiger Anzahl vorkommen und beliebig kombiniert werden:  
*COUNT x = vx TO vz (1, 7 THRU 17, 20, 47, 70 THRU HI).*

**Beispiel:**

*COUNT neu = a b c d (1).*

Ergebnis dieses *COUNT*-Befehls für folgende Werte:

A	B	C	D	→ Neu =
0	0	0	0	0
0	1	0	0	1
1	0	0	1	2
1	1	0	1	3
1	1	1	1	4

*COUNT* ignoriert "user-missing values" und behandelt sie als "legale" Werte. *COUNT* liefert als Ergebnis nie einen fehlenden Wert, da, wenn kein Wert der "werteliste" in den Variablen der "varliste" vorkommt, das Ergebnis der Zählung 0 ist. In der Werteliste können die Schlüsselworte *MISSING* oder *SYSMIS* stehen, um alle fehlenden Werte mit *MISSING* ("user-" und "system missing values") oder nur "system-missing values" mit *SYSMIS* auszählen zu lassen.

Anhand der Übungsdaten wird die Wirkungsweise von *COUNT* noch einmal verdeutlicht: Die neue Variable "vcount" erhält bei einem Fall eine 2, wenn sich nach Meinung des/der Befragten die eigene Wirtschaftslage („v9“) und die Wirtschaftslage in der BRD („v8“) in einem Jahr verbessern. "vcount" erhält eine 1 zugewiesen, wenn sich entweder die persönliche Lage oder die wirtschaftlichen Verhältnisse seiner/ihrer Ansicht nach verbessern. Eine 0 erhalten alle Befragten, die bei beiden Fragen keine Verbesserung sehen. Ein *LIST* listet die Variablen dann fallweise auf. *CASES=10* beschränkt diese Auflistung allerdings auf die ersten zehn Fälle.

→ Beispiel

*COUNT verbessert = v9 v8 (1,2).*

*LIST VARIABLES = v9 v8 verbessert/ CASES=10.*

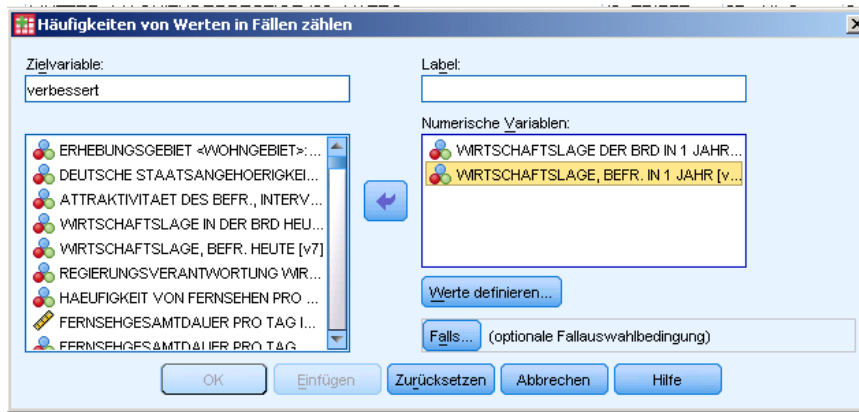
```
v9 v8 verbessert
3 3      ,00
1 4      1,00
3 3      ,00
3 3      ,00
3 3      ,00
5 4      ,00
3 4      ,00
4 4      ,00
3 3      ,00
3 4      ,00
```

Number of cases read: 10      Number of cases listed: 10

Menüentsprechung:

TRANSFORMIEREN ⇒ WERTE IN FÄLLEN ZÄHLEN...

→ Menü



Um die zu zählenden Werte angeben zu können, muss die Schaltfläche **Werte definieren ...** aktiviert werden. Diese Schaltfläche ist erst aktiv, wenn mindestens eine Variable aus der Liste in dem Feld Variablen angegeben ist.

### 6.1.4 TEMPORARY

#### → Syntax

Im Zusammenhang mit Datenmodifikationen ist das *TEMPORARY*-Kommando von großer Bedeutung. Dieser Befehl ist nur über die Syntax und nicht über die Menüsteuerung verfügbar.

#### TEMPORARY.

*TEMPORARY* signalisiert den Beginn einer Folge von so genannten "temporären Modifikationen".<sup>18</sup> Das sind Modifikationskommandos (wie *COMPUTE*, *RECODE*), deren Wirkung nur bis zum nächsten Prozedurkommando anhalten soll. So existieren Variablen, die nach einem *TEMPORARY*-Befehl neu gebildet werden, ebenfalls nur bis zur nächsten Prozedur. Für diese stehen sie allerdings zur Verfügung. Entsprechend sind auch Modifikationen von „alten“ Variablen nur bis zu der dem *TEMPORARY*-Befehl folgenden Prozedur verfügbar. Alle Transformationen, die nicht zwischen einem *TEMPORARY*-Kommando und der folgenden Prozedur stehen, sind "permanente" Transformationen und gelten für den gesamten folgenden Auswertungslauf.

#### → Beispiel

Im folgenden Beispiel wird die Variable "v173" (Allgemeiner Schulabschluss) temporär rekodiert. Beim ersten *FREQUENCIES*-Kommando hat die Variable dann drei Ausprägungen. Beim zweiten *FREQUENCIES*-Kommando hingegen besitzt die Variable wieder sieben Ausprägungen, da das *RECODE*-Kommando, durch das vorgeschaltete *TEMPORARY*, nach der ersten Prozedur unwirksam wird. Dies ist in diesem Fall *FREQUENCIES*. Auch das *VALUE LABELS*-Kommando ist in diesem Fall nur temporär wirksam!

```
TEMPORARY.
RECODE      v173  (2 THRU 3=1) (4 THRU 5=2) (1,6,7=9).
VALUE LABELS v173  1 'niedriger Abschluss'
                2 'höherer Abschluss'
                3 'Weiteres'.
FREQUENCIES VARIABLES = v173.
FREQUENCIES VARIABLES = v173.
```

Ergebnis des ersten *FREQUENCIES*-Kommandos:

#### Statistiken

ALLGEMEINER SCHULABSCHLUSS		
N	Gültig	3459
	Fehlend	10

#### → Ergebnis

<sup>18</sup> SPSS (1990), S. 698 bringt eine Liste von Kommandos, die nach einem *TEMPORARY*-Kommando zulässig sind.



## ALLGEMEINER SCHULABSCHLUSS

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	niedriger Abschluss	2313	66,7	66,9	66,9
	höherer Abschluss	995	28,7	28,8	95,7
	9	150	4,3	4,3	100,0
	Gesamt	3459	99,7	100,0	
Fehlend	99	10	,3		
Gesamt		3469	100,0		

Ergebnis des zweiten *FREQUENCIES*-Kommandos:

## Statistiken

## ALLGEMEINER SCHULABSCHLUSS

N	Gültig	3459
	Fehlend	10

## ALLGEMEINER SCHULABSCHLUSS

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	OHNE ABSCHLUSS	93	2,7	2,7	2,7
	VOLKS-,HAUPTSCHULE	1314	37,9	38,0	40,7
	MITTLERE REIFE	999	28,8	28,9	69,6
	FACHHOCHSCHULREIFE	223	6,4	6,4	76,0
	HOCHSCHULREIFE	773	22,3	22,3	98,3
	ANDERER ABSCHLUSS	16	,5	,5	98,8
	NOCH SCHUELER	41	1,2	1,2	100,0
	Gesamt		3459	99,7	100,0
Fehlend	KEINE ANGABE	10	,3		
Gesamt		3469	100,0		

### 6.1.5 Das WEIGHT-Kommando: Fälle gewichten

Gemäß Voreinstellung von SPSS Statistics geht jeder Fall nur einmal in eine Analyse ein. Jeder Fall wird also für Auswertungen mit dem Wert 1 gewichtet. Teilweise wird es jedoch als sinnvoll angesehen, die Merkmalsträger mit unterschiedlichen Gewichten in die Berechnungen einzubeziehen. Dies ist häufig dann sinnvoll, wenn vermutet wird, dass die Objekte einer Stichprobe bezüglich der Ausprägungen bestimmter Merkmale nicht repräsentativ sind; das heißt, dass bestimmte Merkmalsausprägungen in der Stichprobe unter- und andere überrepräsentiert sind.

Ein Beispiel hierfür ist, dass aufgrund der Stichprobenauswahl die doppelte Anzahl an Männern, verglichen mit der Anzahl der Frauen zu befragen war, obwohl beide Geschlechter in der Grundgesamtheit zu gleichen Teilen vertreten sind. Hier könnte erwogen werden, die Fälle mit den Antworten der Männer mit 0,5 und die der Frauen mit 2 zu gewichten. Jede Frau zählt dann für eine Analyse doppelt, jeder Mann nur die Hälfte. Damit bleibt die Anzahl der Objekte gleich dem Stichprobenumfang. Die Häufigkeitsverteilung des Geschlechts weist nach dieser Gewichtung 50 % Frauen und 50 % Männer aus, genau wie in der Grundgesamtheit.

So genannte Repräsentativität wird aber immer nur für bestimmte gemeinsame Merkmalsverteilungen hergestellt. Dies gewährleistet nicht, dass die gemeinsame Stichprobenverteilung aller Merkmale mit ihrer gemeinsamen Verteilung in der Grundgesamtheit identisch ist. Defizite durch fehlende Werte lassen sich mit dieser Methode ebenfalls nicht beseitigen. Bei vielen Ausfällen („missings“) werden zwangsläufig bestimmte Merkmalausprägungen über- oder unterrepräsentiert.

**WEIGHT {BY varname} {OFF}.**

→ **Syntax**

**varname** ist eine bereits existierende numerische Variable, nach der gewichtet werden soll. Die Gewichtung erfolgt nach der Codierung der jeweiligen Variablen.

Beim *WEIGHT*-Befehl gilt das *TEMPORARY*-Kommando nicht. Um die Gewichtung auszuschalten, muss der Befehl "*WEIGHT OFF.*" ausgeführt werden.

→ **Beispiel**

Bei Allbus 2008 wurden mehr Personen aus den neuen Bundesländern befragt, um eine aussagefähige Untergliederung der ostdeutschen Teilstichprobe zu gewährleisten. Im folgenden Beispiel wird der Datensatz deshalb nach dem „personenbezogenen Ost-West-Gewicht“ (v792) gewichtet. Diese Variable ist so codiert, dass „Alte Bundesländer“ die Ausprägung 1,1833 und „Neue Bundesländer“ die Ausprägung 0,593 zugeteilt bekommen hat. Dies sorgt dafür, dass Befragte aus den alten Bundesländern „häufiger“ gezählt werden. Zum Vergleich wird je eine Kreuztabelle vor und eine nach der Gewichtung aufgeführt. Die ungewichtete Summe der Gewichte ist gleich der Anzahl der ursprünglich Befragten.

*WEIGHT OFF.*

*DESCRIPTIVES v792 /STA=SUM.*

**Deskriptive Statistik**

	N	Summe
PERSONENBEZOGENES OST-WEST-GEWICHT	3469	3469
Gültige Werte (Listenweise)	3469	

→ **Ergebnis**

*WEIGHT OFF.*

*CROSSTABS v524 BY v3.*

*WEIGHT BY v792.*

*CROSSTABS v524 BY v3.*

Die SPSS Statistics-Ausgabe liefert folgende Ergebnisse:

Kreuztabelle Konfession / Erhebungsgebiet ungewichtet:

**KONFESSION, BEFRAGTE<R> \* ERHEBUNGSGEBIET <WOHN- GEBIET>: WEST - OST Kreuztabelle**

Anzahl

		ERHEBUNGSGEBIET <WOHN- GEBIET>: WEST - OST		Gesamt
		ALTE BUN- DESLAENDER	NEUE BUN- DESLAENDER	
KONFESSION, BEFRAG- TE<R>	EVANG.OHNE FREIKIRCH	815	209	1024
	EVANG.FREIKIRCHE	28	7	35
	ROEMISCH-KATHOLISCH	977	40	1017
	AND.CHRISTL.RELIGION	60	13	73
	AND.NICHT-CHRISTLICH	108	7	115
	KEINER RELIGIONSGEM.	388	797	1185
Gesamt		2376	1073	3449

Kreuztabelle Konfession/ Geschlecht mit v792 gewichtet:

**KONFESSION, BEFRAGTE<R> \* ERHEBUNGSGEBIET <WOHN- GEBIET>: WEST - OST Kreuztabelle**

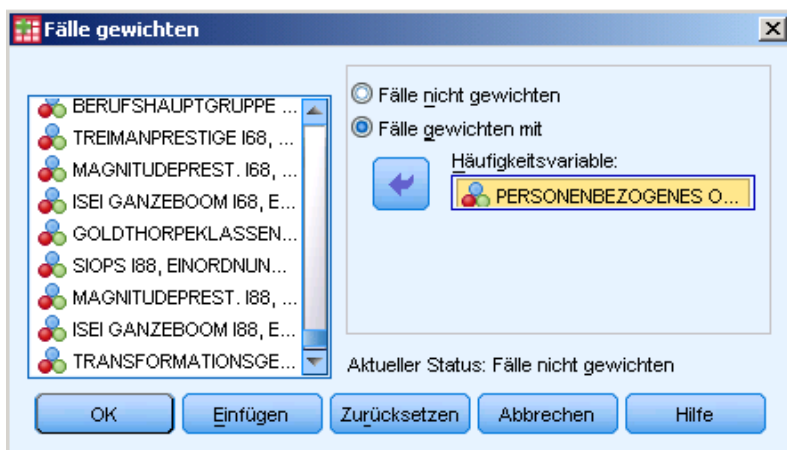
Anzahl

		ERHEBUNGSGEBIET <WOHN- GEBIET>: WEST - OST		Gesamt
		ALTE BUN- DESLAENDER	NEUE BUN- DESLAENDER	
KONFESSION, BEFRAG- TE<R>	EVANG.OHNE FREIKIRCH	964	124	1088
	EVANG.FREIKIRCHE	33	4	37
	ROEMISCH-KATHOLISCH	1156	24	1180
	AND.CHRISTL.RELIGION	71	8	79
	AND.NICHT-CHRISTLICH	128	4	132
	KEINER RELIGIONSGEM.	459	473	932
Gesamt		2811	637	3448

Durch die Gewichtung wird das reale Verhältnis der Bevölkerung in den alten und den neuen Bundesländern wieder hergestellt.

Menüentsprechung:

DATEN ⇒ FÄLLE GEWICHTEN...



→ Menü

## 6.2 Modifikationen in Abhängigkeit von Bedingungen

Im Verlauf einer Datenanalyse ist es oft erforderlich, Modifikationen von Variablen nur dann vorzunehmen, wenn bestimmte Konstellationen bei den Variablenwerten vorliegen. Einer solchen Modifikation in Abhängigkeit von Bedingungen dienen die folgenden Kommandos.

### 6.2.1 Einfaches IF und logische Ausdrücke

Das *IF*-Kommando wird nur für numerische Zielvariablen und numerische Variablen im logischen Ausdruck erörtert. Auf Zeichenkettenvariablen und Zeichenketten im *IF*-Kommando wird nicht gesondert eingegangen. Die *IF*-Anweisung entspricht einer *COMPUTE*-Anweisung, die nur unter ganz bestimmten Bedingungen ausgeführt wird, nämlich immer dann, wenn ein "logischer Ausdruck" "wahr" ist.

**IF (logischer Ausdruck) variable = arithmetischer Ausdruck.**

→ Syntax

Der Spezifikationsbereich "variable = arithmetischer Ausdruck" ist mit dem entsprechenden Teil des *COMPUTE*-Kommandos für numerische Variablen identisch und muss daher nicht mehr dargestellt werden (Einzelheiten zum *COMPUTE*-Kommando siehe Seite 51 ff.).

Ein „logischer Ausdruck“ besteht aus einer oder mehreren Relationen ("relations"). Relationen werden durch einen logischen Operator verknüpft

**logischer Ausdruck:** *Relation* [*logischer Operator* *Relation* ...]

Eine Relation setzt sich aus zwei arithmetischen Ausdrücken zusammen, zwischen denen ein Vergleichsoperator ("relational operator") steht:

**Relation:** *arithmetischer Ausdruck* *Vergleichsoperator* *arithmetischer Ausdruck*

Zunächst ist zu klären, wie Relationen gebildet werden:

Für die Komponente "arithmetischer Ausdruck" gelten die gleichen Regeln wie für den "arithmetischen Ausdruck" beim *COMPUTE*-Kommando (siehe auch "arithmetischen Ausdruck" beim *COMPUTE*-Kommando Seite 51 ff.). Neu ist demnach der "Vergleichsoperator". Operatoren und Ersatzzeichen können dabei synonym verwendet werden. Wir verwenden wegen der Anschaulichkeit die Ersatzzeichen der Vergleichsoperatoren. Vergleichsoperatoren und ihre Bedeutung sind:

Operator	Ersatzzeichen	Bedeutung
EQ	=	gleich
NE	<>	ungleich
LT	<	kleiner als
LE	<=	kleiner gleich
GT	>	größer als
GE	>=	größer gleich

Für jeden Fall werden die arithmetischen Ausdrücke einer Relation anhand des Vergleichsoperators miteinander verglichen. Nach der Auswertung der arithmetischen Ausdrücke, welche die gleiche, wie beim *COMPUTE*-Kommando ist, entspricht dies einem Vergleich von zwei Einzelwerten. Jede Relation kann "wahr", "falsch" oder "missing" sein. Eine Relation ist "wahr", wenn der Vergleich für den betrachteten Fall stimmt, sie ist "falsch", wenn er nicht stimmt und sie ist "missing", wenn aufgrund fehlender Werte in Variablen keine Entscheidung möglich ist.

Mehrere Relationen werden durch logische Operatoren ("logical operator") miteinander verbunden:

**Relation** *logischer Operator* *Relation*

Die logischen Operatoren und ihre Bedeutung sind:

Operator	Ersatzzeichen	Bedeutung
AND	&	UND-Verknüpfung
OR		ODER-Verknüpfung
NOT	~	Umkehrung des Wahrheitswertes des folgenden logischen Ausdruckes

Ob das Ergebnis einer solchen Kombination von Relationen "wahr", "falsch" oder "missing" ist, hängt von den "Wahrheitswerten" der zugehörigen Relationen und der gewählten Verknüpfung ab. Es gilt dabei:

Relation	Operator	Relation	Ergebnis
Wahr	And	Wahr	Wahr
Wahr	And	Falsch	Falsch
Falsch	And	Falsch	Falsch
Wahr	And	Missing	Missing
Falsch	And	Missing	Falsch
Missing	And	Missing	Missing
Wahr	Or	Wahr	Wahr
Wahr	Or	Falsch	Wahr
Falsch	Or	Falsch	Falsch
Wahr	Or	Missing	Wahr
Falsch	Or	Missing	Missing
Missing	Or	Missing	Missing

Die erste Zeile dieser Tabelle ist folgendermaßen zu interpretieren: Wenn bei einer Kombination von zwei Relationen die erste Relation "wahr" ist und gleichzeitig (*AND*) auch die zweite Relation "wahr"

ist, so ist das Ergebnis der Kombination auch "wahr". Die Interpretation der übrigen Zeilen ist analog vorzunehmen.

Stößt SPSS Statistics auf eine *IF*-Anweisung, dann wird für jeden Fall geprüft, ob der gesamte logische Ausdruck "wahr", "falsch" oder "missing" ist. Ist der logische Ausdruck "wahr", so wird die aufgeführte "Rechenanweisung" (variable = arithmetischer Ausdruck) ausgeführt. Ist der logische Ausdruck hingegen "falsch" oder "missing", so wird für den betreffenden Fall keine Berechnung durchgeführt. Die Zielvariable links des Gleichheitszeichens bleibt unverändert. Wenn ein Fall auf diese Weise ausgewertet ist, fährt SPSS Statistics mit der Prüfung des logischen Ausdrucks für den nächsten Fall fort.

Im Folgenden wird die Wirkungsweise des *IF*-Kommandos an einigen Beispielen erörtert.

Im einfachsten Fall besteht ein logischer Ausdruck nur aus einer einzigen Relation, in der die beiden arithmetischen Ausdrücke einzelne Konstanten oder einzelne Variablen sind.

#### Beispiele:

*IF (v154 >= 70) klasse = 1.*

Der arithmetische Ausdruck "v154" (Variable Alter Befragte <r>) und der arithmetische Ausdruck "70" (Konstante) sind durch den Vergleichsoperator ">=" verbunden. Bei jedem Fall wird daher geprüft, ob der Wert der Variablen "v154" größer oder gleich 70 ist. Wenn ja, wird der Variablen "klasse" bei eben diesem Fall der Wert 1 zugewiesen. Wenn nein, bleibt "klasse", falls es sich um eine "alte" Variable handelt, unverändert, beziehungsweise nimmt den "system-missing value" an, falls es sich um eine "neue" Variable handelt.

*IF (v154 >= 70 AND v151 = 1) klasse = 1.*

Bei dieser Bedingung muss bei einem Fall sowohl der Wert der Variablen "v154" größer als 70 als auch der Wert der Variablen "v151" (Geschlecht) gleich 1 (männlich) sein, damit "klasse" den Wert 1 zugewiesen bekommt.

*IF (v151 = 2 OR v248 <> 1) vneu = 2.*

"vneu" erhält bei einem Fall den Wert 2 genau dann zugewiesen, wenn die Variable "v151" (Geschlecht) den Wert 2 (weiblich) besitzt oder "v248" (Familienstand) nicht den Wert 1 (ledig), oder wenn beides zutrifft.

#### Reihenfolge der Befehlsausführung:

Besteht ein logischer Ausdruck aus mehreren Relationen, so werden zunächst die arithmetischen Ausdrücke der einzelnen Relationen ausgewertet. Diese Auswertung entspricht der des *COMPUTE*-Kommandos. Anschließend wird anhand der Vergleichsoperatoren der Wahrheitswert der einzelnen Relationen ermittelt. Um den Wahrheitswert des gesamten logischen Ausdrucks zu bestimmen, wertet SPSS Statistics zunächst den "NOT"-Operator aus, dann "AND"-Verknüpfungen von Relationen und als Letztes "OR"-Verknüpfungen. Durch geeignetes Setzen von Klammern kann eine davon abweichende Auswertungsreihenfolge erzwungen werden, da Ausdrücke in Klammern vorrangig analysiert werden.

Es ist äußerst wichtig, immer genau zu wissen, unter welchen Bedingungen ein logischer Ausdruck "wahr" ist. Wertzuweisungen in Abhängigkeit von der Auswertung komplexer logischer Ausdrücke führen häufig zu unliebsamen Überraschungen bei den folgenden Analysen. Logische Ausdrücke sollten durch das Setzen von Klammern so strukturiert sein, dass die Reihenfolge der Bearbeitung klar ersichtlich ist. Die Anzahl der Klammerebenen ist nicht begrenzt.

Zusätzlich sollten Anwender, die bereits Berechnungen mit alten SPSS Statistics Versionen (wie SPSS-X, SPSS 11.0 und SPSS/PC+) erstellt haben, beachten, dass fehlende Werte im logischen Ausdruck in diesen Programmen unterschiedlich gehandhabt werden. Dies führt unter Umständen zu unterschiedlichen Ergebnissen bei Auswertungen.

Falsche Ergebnisse als Resultat "unkontrollierter" logischer Bedingungen werden unter Umständen nur dadurch entdeckt, dass sie im Sachzusammenhang unplausibel sind. Häufig geschieht dies erst nach vielen Auswertungsschritten, die dann unter entsprechendem Aufwand zu wiederholen sind.

Bei Auswertungen mit großen Fallzahlen sollten die Auswirkungen von "unüberschaubaren" logischen Bedingungen anhand kleiner Beispiele "getestet" werden. Gegebenenfalls müssen Variablenwerte durch ein *LIST*-Kommando ausgedrückt werden, um die Korrektheit der formulierten Bedingung zu überprüfen.

Beispiel mit fiktiven Variablen:

*IF (a = 1 AND b = 1 OR c = 1) z = 1.*

→ Beispiele

Der gesamte logische Ausdruck ist wahr, wenn "a" und "b" gleich 1 sind oder wenn "c" gleich 1 ist oder beides zutrifft. Es genügt demnach, dass "c" gleich 1 ist, um "z" auf 1 zu setzen.

$$IF (a = 1 AND (b = 1 OR c = 1)) z = 1.$$

Hier wird zunächst der logische Ausdruck in der inneren Klammerebene untersucht. Dieser kann wahr oder falsch sein. Der gesamte logische Ausdruck kann wegen der "AND"-Verknüpfung der ersten Relation ( $a = 1$ ) mit dem Klammerausdruck allerdings nur dann wahr sein, wenn "a" gleich 1 und gleichzeitig die Relation in der inneren Klammer "wahr" ist.

## → Beispiel

Aus unseren Beispieldaten wollen wir die Befragten aufteilen in Frauen über / unter 40 und Männer über / unter 40. Anschließend wollen wir anhand eines *FREQUENCIES*-Befehls sehen, mit welcher Häufigkeit die Kombinationen (neue Variable *Typ*) vorkommen.

Die Befehle hierfür lauten:

$$IF (v151=1 AND v154 >= 40) \quad Typ = 1.$$

$$IF (v151=1 AND v154 < 40) \quad Typ = 2.$$

$$IF (v151=2 AND v154 >= 40) \quad Typ = 3.$$

$$IF (v151=2 AND v154 < 40) \quad Typ = 4.$$

VALUE LABELS *Typ* 1 'männlich ab 40'  
 2 'männlich unter 40'  
 3 'weiblich ab 40'  
 4 'weiblich unter 40'.

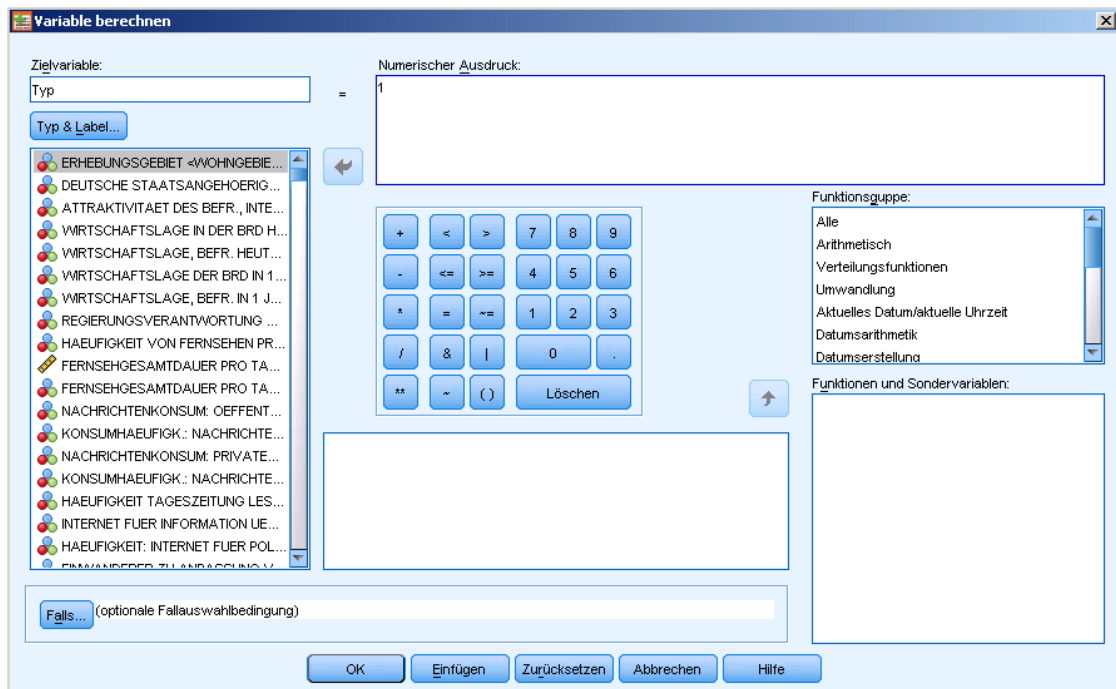
*FREQUENCIES Typ*.

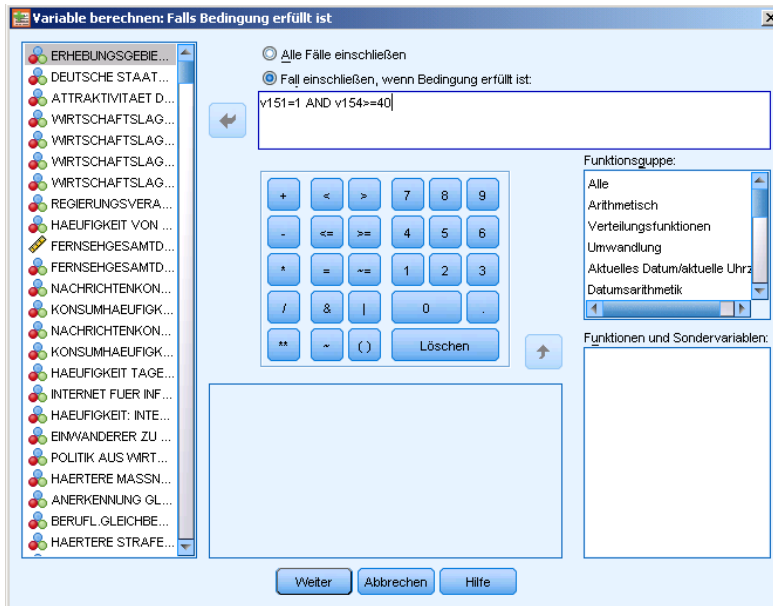
## → Menü

Die Menüberechnung erfolgt ähnlich wie beim *COMPUTE*-Kommando über den Pfad TRANSFORMIEREN ⇒ VARIABLE BERECHNEN, es erscheint auch hier zunächst das Fenster *Variable berechnen*.

Mit Hilfe der Schaltfläche **Falls...** gelangt man in das Fenster *Variable berechnen: Falls Bedingung erfüllt ist* und kann hier die logische Bedingung formulieren. Wichtig dabei ist, die Option **Fall einschließen, wenn Bedingung erfüllt ist** zu aktivieren.

Für unser Beispiel sieht das Menü, das viermal zu „durchlaufen“ ist, dann folgendermaßen aus:





Das Ergebnis unseres Beispiels sieht dann folgendermaßen aus:

→ Ergebnis

		Typ			
		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	'männlich ab 40'	1203	34,7	34,8	34,8
	'männlich unter 40'	504	14,5	14,6	49,4
	'weiblich ab 40'	1235	35,6	35,7	85,1
	'weiblich unter 40'	515	14,8	14,9	100,0
	Gesamt	3457	99,7	100,0	
Fehlend	System	12	,3		
Gesamt		3469	100,0		

### 6.2.2 Erweiterung des IF-Kommandos - DO IF/END IF

Das *IF*-Kommando ist bezüglich der Menge möglicher Modifikationen eingeschränkt auf Modifikationen, wie sie bei einem *COMPUTE*-Kommando zulässig sind. Die Strukturen der Kommandos *RECODE* und *COUNT* stehen nicht zur Verfügung. Mehrere *IF*-Kommandos mit sich ausschließenden logischen Bedingungen erfordern bei jedem Fall für jedes *IF* eine Prüfung der logischen Bedingung, obwohl bei einem Fall unter Umständen nach dem ersten *IF* bekannt ist, dass er die übrigen logischen Bedingungen nicht erfüllt. Diese Vorgehensweise ist zeitraubend. Eine Erweiterung des *IF*-Kommandos, das *DO IF*-Kommando, besitzt die geschilderten ungünstigen Eigenschaften nicht und lässt zudem komplexe Bedingungsstrukturen zu:

→ Syntax

```

DO IF      (logische Bedingung).
    Datenmodifikation.
[ELSE IF  (logische Bedingung).]
    Datenmodifikation.
[ELSE IF  (logische Bedingung).]
...
[ELSE.]
    Datenmodifikation.
END IF.
    
```

Eine solche Sequenz von logischen Bedingungen und Datenmodifikationen beginnt immer mit einem *DO IF* und endet immer mit einem *END IF*. *ELSE IF*-Kommandos dürfen beliebig oft vorkommen.

Das *ELSE*-Kommando hingegen ist nur einmal zulässig und steht hinter allen *ELSE IF*-Kommandos. Wie werden diese Kommandos jetzt abgearbeitet?

Wenn der logische Ausdruck des *DO IF*-Kommandos wahr ist, werden die Modifikationen, die dem *DO IF* unmittelbar folgen, abgearbeitet und zwar bis zum nächsten *ELSE IF* oder *ELSE*. Sind weder *ELSE IF*-Kommandos noch ein *ELSE*-Kommando vorhanden, werden alle Modifikationen zwischen *DO IF* und *END IF* ausgeführt. Anschließend wird mit dem ersten Kommando nach dem *END IF* fortgefahren.

Wenn der logische Ausdruck beim *DO IF* falsch ist, springt das Programm zum ersten *ELSE IF* und wertet dort den logischen Ausdruck aus. Wenn dieser wahr ist, werden die diesem *ELSE IF* folgenden Modifikationen ausgeführt und zwar bis zum nächsten *ELSE IF* oder *ELSE*-Kommando. Anschließend wird mit dem ersten Kommando nach dem *END IF* fortgefahren.

Wenn die Bedingungen des *DO IF* und des ersten *ELSE IF* falsch sind, wird das nächste *ELSE IF* ausgewertet. Falls keine der Bedingungen eines *ELSE IF*-Kommandos wahr ist, werden die Modifikationen zwischen *ELSE* und *END IF* ausgeführt. Anschließend wird die Struktur verlassen.

Wenn kein *ELSE* vorhanden ist und die *DO IF* sowie alle *ELSE IF* Bedingungen falsch sind, findet keine Aktion statt.

Falls irgendein logischer Ausdruck bei der Auswertung den fehlenden Wert (missing value) liefert, wird die Struktur für den betreffenden Fall sofort verlassen und nicht mehr weiter geprüft.

Im einfachsten Fall besteht die *DO IF/END IF*-Struktur nur aus Modifikationskommandos, die von *DO IF* und *END IF* umschlossen sind.

Ein Beispiel für den Übungsdatensatz könnte folgendermaßen lauten:

```
DO IF      (v151 = 1).
  RECODE   v173 (1,2=0) (3,4=1) (ELSE=9).
  COMPUTE  v388 = v388 + v388 * 0.05.
  COUNT    ein = v6 v7 (1,2).
END IF.
```

Die Modifikationen zwischen *DO IF* und *END IF* werden nur dann ausgeführt, wenn die Variable "v151" (Geschlecht) den Wert 1 besitzt. Wenn der Wert der Variablen "v151" bei einem Fall ungleich 1 ist oder einen fehlenden Wert annimmt, erfolgt keine Aktion. "v173", „v388" und "ein" werden nicht verändert.

```
DO IF      (v151 = 1).
  COMPUTE  y = 1.
ELSE.
  COMPUTE  y = 2.
END IF.
```

Immer dann, wenn die Variable "v151" den Wert 1 hat, wird "y" ebenfalls auf 1 gesetzt. Bei allen anderen Fällen erhält "y" den Wert 2. Wenn "v151" allerdings einen fehlenden Wert annimmt, erfolgt wieder keine Aktion. "y" wird nicht verändert.

```
DO IF      (v151 = 1).
  COMPUTE  y = 1.
ELSE IF    (v151 = 2).
  COMPUTE  y = 2.
ELSE.
  COMPUTE  y = 9.
END IF.
```

Für Fälle, deren Ausprägung bei "v151" gleich 1 ist, wird "y" auf 1 gesetzt. Für Fälle, deren Ausprägung bei der Variablen "v151" gleich 2 ist, erhält "y" eine 2 zugewiesen. Bei allen übrigen Fällen erhält "y" eine 9 als Ausprägung. Wenn "v151" einen fehlenden Wert annimmt, bleibt "y" unverändert.

#### Anmerkungen:

*DO IF* - *END IF*-Strukturen können beliebig geschachtelt werden. Eine eher theoretische Grenze stellt der verfügbare Hauptspeicherplatz des ausführenden Rechners dar.

```
DO IF (x = 1).
  DO IF (v151 = 1).
    COMPUTE  y = 1.
  ELSE IF (v151 = 2).
    COMPUTE  y = 2.
```

## → Beispiel



```

        END IF.
    ELSE IF (x <> 1).
        COMPUTE y = 3.
    END IF.

```

Wenn ein logischer Ausdruck für einen Fall wahr ist, wird die *DO IF-END IF*-Struktur immer verlassen, sobald die zugehörigen Modifikationen ausgeführt sind, und zwar auch dann, wenn ein folgender logischer Ausdruck für den betreffenden Fall ebenfalls wahr wäre:

```

DO IF (x = 1).
    COMPUTE y = 1.
ELSE IF (x LE 10).
    COMPUTE y = 2.
END IF.

```

Wenn der Wert von "x" bei einem Fall gleich 1 ist, wird "y" auf 1 gesetzt und anschließend mit dem ersten Kommando hinter dem *END IF*-Kommando fortgefahren. Der Fall wird bezüglich "y" nicht weiter überprüft, obwohl er dann auch die Bedingung des *ELSE IF* erfüllt.

Außer den "reinen" Modifikationskommandos (*RECODE*, *COMPUTE*, *COUNT*) sind in einer *DO IF - END IF* Struktur noch weitere Kommandos (Transformationskommandos) zulässig. Einzelheiten und Besonderheiten beim Einsatz solcher Befehle sind dem Handbuch zu entnehmen.<sup>19</sup>

Zum Abschluss des *DO IF*-Befehls folgt noch ein komplettes Beispiel mit unseren Daten. Für Angestellte mit einem Nettomonatseinkommen ("v388") von unter 2500, die verheiratet sind und mit ihrem Ehepartner zusammen leben, soll das Einkommen um 5 % erhöht werden. Dazu wird zunächst der Betrag der Erhöhung ("nmeerh") berechnet und anschließend diese Erhöhung zum ursprünglichen Nettomonatseinkommen addiert. Dies ergibt dann das neue Einkommen "nmeneu". Die Ergebnisse eines *DESCRIPTIVES*-Kommandos zeigen anschließend, dass lediglich 1324 Personen von der Einkommenserhöhung betroffen sind. Für die übrigen Fälle wurden, da sie die logische Bedingung ("v388 LT 2500 AND v248 EQ 1") nicht erfüllen, die *COMPUTE*-Kommandos nicht ausgeführt.

```

DO IF (v388 LT 2500 AND v248 EQ 1).
    COMPUTE nmeerh = v388*5/100.
    COMPUTE nmeneu = v388 + nmeerh.
END IF.

```

```

VARIABLE LABELS nmeneu 'neues Einkommen' /nmeerh 'Einkommenserhöhung'.
DESCRIPTIVES nmeneu nmeerh/STATISTICS=MEAN.

```

→ Ergebnis

**Deskriptive Statistik**

	N	Mittelwert
neues Einkommen	1324	1191,8850
Einkommenserhöhung	1324	56,7564
Gültige Werte (Listenweise)	1324	

Es sei an dieser Stelle darauf verwiesen, dass es äußerst umständlich ist, die dargestellten Sequenzen von logischen Bedingungen mittels *DO IF*, *ELSE IF* und *END IF* über Menüs zu erzeugen. Aus diesem Grund wird auch hier nicht näher darauf eingegangen.

<sup>19</sup> Siehe dazu SPSS (1990), S. 885, 886.

### 6.3 Abspeichern von Modifikationsergebnissen (SAVE)

Wie bereits ausgeführt wurde, beeinflussen Modifikationen nur die Daten im Dateneditor. Sobald SPSS Statistics auf ein Prozedurkommando trifft, werden alle vor dieser Prozedur (und nach der vorhergehenden Prozedur) stehenden Modifikationskommandos ausgeführt und so der Dateneditor entsprechend modifiziert.

Alle Modifikationsergebnisse können in einer SPSS Statistics-Systemdatei (Endung ".sav") abgespeichert werden. Hierzu gibt es zwei Möglichkeiten:

Dazu wird es entweder über die Menüsteuerung DATEI ⇨ SPEICHERN oder DATEI ⇨ SPEICHERN UNTER... gewählt, wenn der Dateneditor aktiv ist.

Ein SAVE-Kommando erfüllt den gleichen Zweck:

→ Syntax

```
SAVE      OUTFILE='datei'
          [/KEEP=varlist] [/DROP=varlist]
          [/RENAME=(old varlist=new varlist)...].
```

**OUTFILE='datei'** gibt den Namen der Datei an, die die Systemdatei aufnehmen soll (zur Form der Angabe des Dateinamens siehe Seite 32). Dieses Subkommando muss vorhanden sein.

**KEEP=varlist** ist eine Liste von Variablenamen und gibt an, welche Variablen (gezielt) in die Systemdatei übernommen werden sollen. Alle anderen Variablen werden nicht gespeichert. Die übernommenen Variablen besitzen in der Systemdatei die Reihenfolge, in der sie beim *KEEP* aufgeführt sind. Voreinstellung für *KEEP* ist ALL, das heißt, alle Variablen sind zu speichern.

**DROP=varlist** gibt an, welche Variablen nicht in die Systemdatei übernommen werden sollen. Bei *DROP* erhalten die verbleibenden Variablen in der Systemdatei die Reihenfolge, die sie im "active system file" haben.

Bei *KEEP* und *DROP* dürfen Variablenlisten auch mit Hilfe der Spezifikation *TO* gebildet werden, wie beispielsweise: *vara TO varz*.

**RENAME= (altvarlist=neuvarlist)** dient dazu, Variablen beim Abspeichern in die Systemdatei umzubenennen. "altvarlist" und "neuvarlist" sind Variablenlisten. "altvarlist" enthält nur Variablennamen aus dem "active system file". Die Variablennamen in "altvarlist" werden (nacheinander) zu den Variablennamen in "neuvarlist" umbenannt. Die erste Variable auf der linken Seite des Gleichheitszeichens erhält den ersten Variablennamen auf der rechten Seite. Die zweite Variable auf der linken Seite den zweiten Namen der rechten Seite und so weiter. In der Systemdatei werden dann die Variablen unter den Namen aus "neuvarlist" abgespeichert. Die Anzahl der Variablen muss in beiden Variablenlisten identisch sein. Zur Bildung von Variablenlisten ist das Schlüsselwort *TO* sowohl in "altvarlist" als auch in "neuvarlist" zulässig.

Da das Kommando *SAVE* den „Inhalt“ des *Daten-Editors*, nach Auflösung aller offenen Modifikationen, als Systemdatei abspeichert, enthält die so erzeugte Systemdatei alle bis zum *SAVE* durchgeführten Modifikationen und somit auch alle durch Modifikationskommandos erzeugten Variablen und deren *VARIABLE* und *VALUE LABELS*. Um unerwünschten Datenverlusten vorzubeugen, sollte der Name der neuen Systemdatei nicht mit dem Namen einer im selben Auswertungslauf bereits eingelesenen Systemdatei identisch sein.

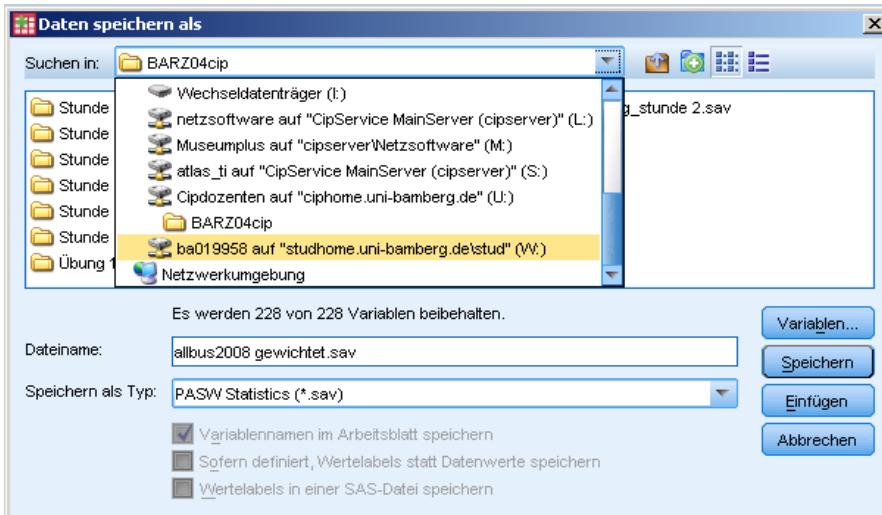
Wir möchten eine ganze Reihe von Variablen, die in unseren Übungsbeispielen gebildet wurden, zusammen mit allen übrigen Variablen (permanent) in einer neuen Systemdatei abspeichern. Es werden dazu alle zur Bildung der Variablen notwendigen Modifikationsbefehle aufgeführt und dahinter ein *SAVE OUTFILE*-Kommando angegeben. Ein Beispiel mit fiktiven Variablen:

```
COMPUTE geburtsjahr = 2009 - v154.
```

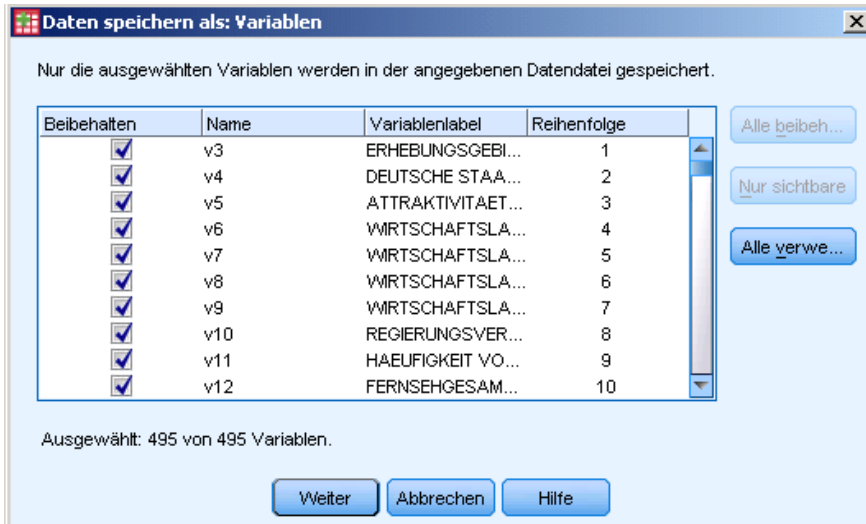
```
SAVE OUTFILE = 'allbusmod.sav'.
```

→ Menü

Das Abspeichern von Daten mit Hilfe des Menüs erfolgt, wie oben ausgeführt, über den Weg DATEI ⇨ SPEICHERN, was den aktiven Datensatz direkt abspeichert oder über DATEI ⇨ SPEICHERN UNTER... . Hier erscheint das Fenster *Daten speichern unter* mit der Möglichkeit einen (neuen) Dateinamen anzugeben.



Mit Hilfe der Schaltfläche **Variablen...** gelangt man in das Fenster *Daten speichern als: Variablen* und hat hier die Möglichkeit, Variablen auszuwählen, die in die zu speichernde Systemdatei übernommen werden sollen. Die entsprechenden Variablen werden mit einem Kreuz markiert, wobei per Voreinstellung zunächst alle Variablen ausgewählt sind. Die übernommenen Variablen besitzen in der Systemdatei die Reihenfolge, in der sie in der Auswahlliste erscheinen.



## 7 Datenselektionen

Im Zuge eines empirischen Forschungsprojekts sind häufig bestimmte Gruppierungen der Untersuchungsobjekte (Fälle) zu analysieren. Hierfür sind die Datenselektionsbefehle *FILTER* und *SELECT IF* dienlich. Sie ermöglichen es, Fälle nach gewissen Kriterien auszuwählen. Die Selektion (Auswahl) der Fälle erfolgt anhand einer logischen Bedingung. Auf eine zufällige Auswahl der Fälle wird hier nicht eingegangen.

### 7.1 Filter

#### → Syntax

Filter sortiert Fälle anhand einer logischen Variablen (Filtervariable) aus. Eine logische Variable hat nur die Werte 0, 1 oder „system-missings“.

```
FILTER [BY variable] [OFF].
```

**variable** logische Variable (mit den Werten 1, 0 oder system-missing value) = Filtervariable.

Wenn die Filtervariable bei einem Fall den Wert 1 annimmt, wird der Fall in die folgenden Analysen mit einbezogen, ansonsten (Wert 0 oder „system-missing value“) geht er nicht in die Analysen ein. *FILTER* wirkt nur auf Prozeduren. **Modifikationen werden auch mit einem Filter für den gesamten Datensatz durchgeführt.** *FILTER* verändert also die Anzahl der Fälle im Datensatz nicht. Filter kennzeichnet lediglich Fälle, die für Analysen zur Verfügung stehen. Sofern dies nicht gewünscht ist, wäre stattdessen ein *SELECT IF*-Kommando angebracht (siehe Seite 70 ff.).

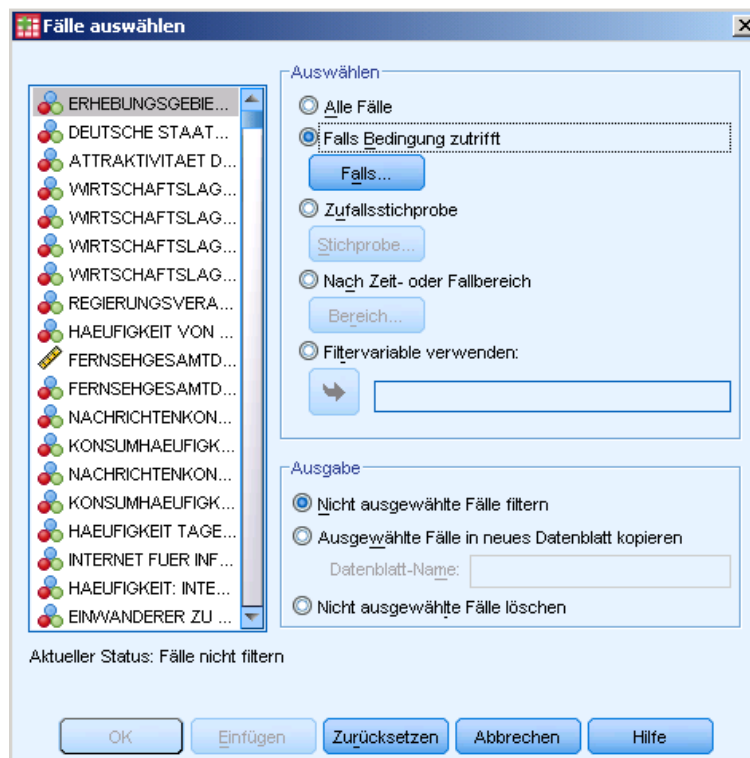
#### → Hinweis!

*FILTER* wirkt solange, bis der Befehl *FILTER OFF* den Filter aufhebt. Dann stehen wieder alle Fälle zur Verfügung.

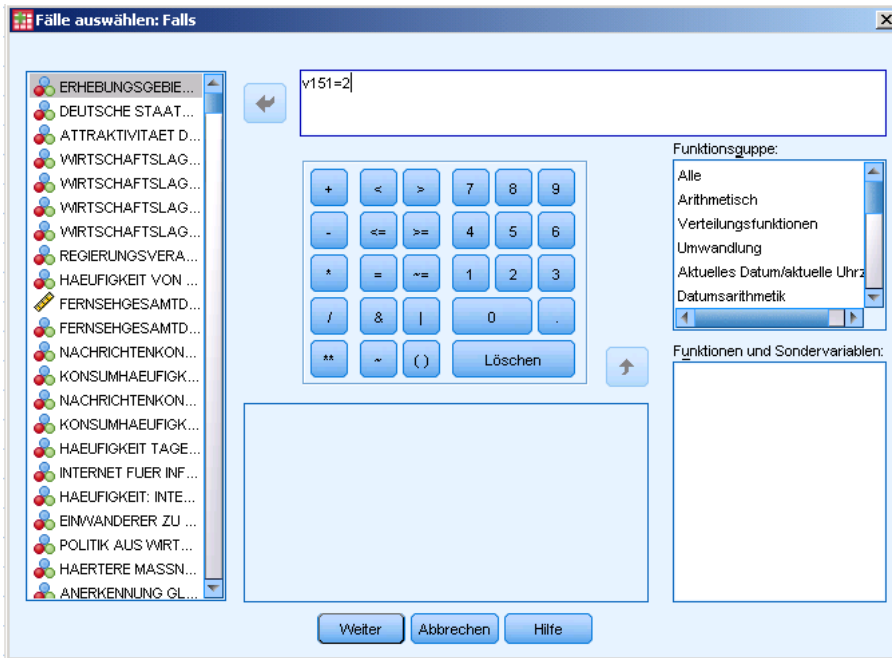
Menüentsprechung:

#### → Menü

DATEN ⇒ FÄLLE AUSWÄHLEN...



Im Dialogfenster *Fälle auswählen* muss die Voreinstellung  **Alle Fälle** deaktiviert werden. Dafür wird der Button  **Falls Bedingung zutrifft** angeklickt. Im Abschnitt *Ausgabe* bleibt der Button  **Nicht ausgewählte Fälle filtern** aktiviert. Dann wird die Schaltfläche **Falls...** angeklickt. Es öffnet sich das folgende Dialogfenster:



Aus der Variablenliste werden die entsprechenden Variablen ausgewählt und mit der (☒)-Schaltfläche in das Editor-Feld übertragen. Ein Anklicken auf die Ziffern der Rechentastatur überträgt diese in das Editor-Feld. Mit einem Doppelklick kann aus der Funktionsliste jede Funktion in das Editor-Feld kopiert werden.

Ein Beispiel, mit einem Filter für die Frauen aus dem Übungsdatensatz, könnte folgendermaßen lauten:

→ Beispiel

```
RECODE v151 (2=1) (1=0) INTO v151filter.
FILTER BY v151filter.
DESCRIPTIVES v388
 / STATISTICS = MIN MAX MEAN.
COMPUTE nme1 = v388 + v388*0.1 .
FILTER OFF.
MEANS v388 nme1 BY v151.
```

**Deskriptive Statistik**

	N	Minimum	Maximum	Mittelwert
BFR.:NETTOEINKOMMEN<OF FENE+LISTENANGABE>	1358	40	8750	1031,47
Gültige Werte (Listenweise)	1358			

→ Ergebnis

**Verarbeitete Fälle**

	Fälle					
	Eingeschlossen		Ausgeschlossen		Insgesamt	
	N	Prozent	N	Prozent	N	Prozent
BFR.:NETTOEINKOMMEN<OF FENE+LISTENANGABE> * GESCHLECHT, BEFRAGTE<R>	2787	80,3%	682	19,7%	3469	100,0%
nme1 * GESCHLECHT, BEFRAGTE<R>	2787	80,3%	682	19,7%	3469	100,0%

## Bericht

GESCHLECHT, BEFRAGTE<R>		BFR.:NETTOEIN KOMMEN<OFFE NE+LISTENANG ABE>	nme1
MANN	Mittelwert	1742,58	1916,8393
	N	1429	1429
	Standardabweichung	1216,521	1338,17289
FRAU	Mittelwert	1031,47	1134,6152
	N	1358	1358
	Standardabweichung	776,106	853,71633
Insgesamt	Mittelwert	1396,08	1535,6910
	N	2787	2787
	Standardabweichung	1085,506	1194,05691

Das Geschlecht wird zunächst rekodiert, um die Filtervariable „v151filter“ zu erhalten. Dann werden mit dem *FILTER*-Befehl die Fälle, für die „v151filter“ den Wert 1 trägt (weibliche Befragte), für die weitere Analyse aus dem Gesamtdatensatz herangezogen. Für diese Fälle werden mit dem Befehl *DESCRIPTIVES* Statistiken über das Nettomonatseinkommen („v388“) angefordert. Es lässt sich erkennen, dass die weiblichen Befragten im Durchschnitt 1031,47 verdienen. Mit einem *COMPUTE*-Kommando wird das Monatseinkommen um 10% erhöht. An einem Mittelwertvergleich (*MEANS*) der Variablen „nme“ und „nme1“ lässt sich erkennen, dass die Einkommenserhöhung für alle Fälle, also auch die Männer, durchgeführt wurde.



## 7.2 Select If

→ Hinweis!

Bei dem folgenden Befehl werden Teile des Datensatzes ausgewählt. Die nicht ausgewählten Teile gehen bei einer Speicherung unter Umständen permanent verloren.

→ Syntax

Das *SELECT IF*-Kommando wählt Fälle nach einem logischen Kriterium aus:

```
SELECT IF (logischer Ausdruck).
```

**logischer Ausdruck** entspricht den Konventionen für einen "logischen Ausdruck", die bereits beim *IF*-Kommando erörtert wurden (siehe auch Seite 59ff.).

Für jeden Fall wird geprüft, ob der logische Ausdruck "wahr" ist. Falls ja, steht der Fall weiterhin für die Analyse zur Verfügung. Fälle hingegen, bei denen der logische Ausdruck "falsch" oder "missing" ist, werden "aussortiert" und gehen nicht mehr in die folgenden Berechnungen ein.

→ Beispiel

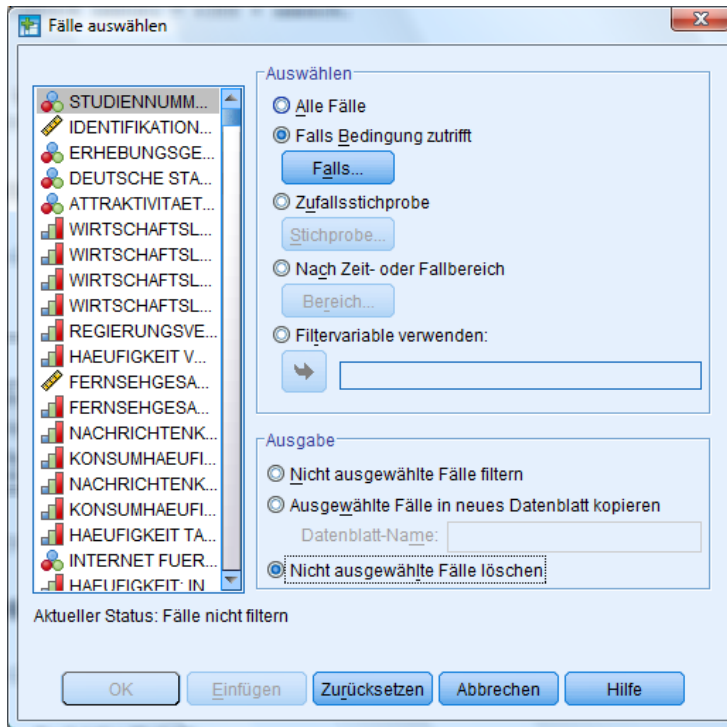
Aus unseren Beispieldaten wollen wir die weiblichen Fälle auswählen und dann nur für diese Fälle eine Häufigkeitsauszählung der Variablen Schulabschluss ("v6") durchführen:

```
SELECT IF (v151 EQ 2).
```

```
VARIABLE LABEL v173 'Schulabschluss der weiblichen Befragten'.  
FREQUENCIES VARIABLES = v173.
```

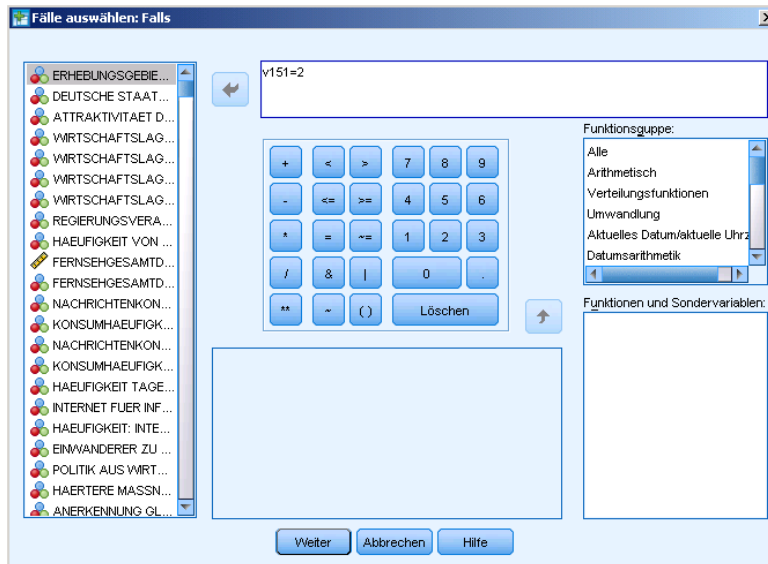
→ Menü

Menüentsprechung:  
DATEN ⇒ FÄLLE AUSWÄHLEN...



Im Dialogfenster *Fälle auswählen* muss die Voreinstellung  **Alle Fälle** deaktiviert werden. Dafür wird der Button  **Falls Bedingung zutrifft** angeklickt. Im Abschnitt *Ausgabe* wird der Button  **Nicht ausgewählte Fälle löschen** aktiviert. Um weiter fortzufahren, wird auf die Schaltfläche **Falls...** geklickt, diese wird erst aktiv, nachdem der Button  **Falls Bedingung zutrifft** angeklickt wurde. Es öffnet sich das folgende Dialogfenster:

Dialogfenster der Schaltfläche: **Falls...**



Die Vorgehensweise ähnelt der bisherigen Handhabung. Aus der Variablenliste werden die entsprechenden Variablen ausgewählt und mit der (👉)-Schaltfläche in das Editor-Feld übertragen. Ein Anklicken auf die Ziffern der Rechentastatur überträgt diese ebenfalls in das Editor-Feld. Mit einem Doppelklick kann aus der Funktionsliste jede Funktion in das Editor-Feld kopiert werden.

**Statistiken**

Schulabschluss der weiblichen Befragten

N	Gültig	1754
	Fehlend	3

➔ Ergebnis

## Schulabschluss der weiblichen Befragten

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	OHNE ABSCHLUSS	46	2,6	2,6	2,6
	VOLKS-,HAUPTSCHULE	678	38,6	38,7	41,3
	MITTLERE REIFE	562	32,0	32,0	73,3
	FACHHOCHSCHULREIFE	81	4,6	4,6	77,9
	HOCHSCHULREIFE	363	20,7	20,7	98,6
	ANDERER ABSCHLUSS	9	,5	,5	99,1
	NOCH SCHUELER	15	,9	,9	100,0
	Gesamt	1754	99,8	100,0	
Fehlend	KEINE ANGABE	3	,2		
Gesamt		1757	100,0		

Werden mehrere *SELECT IF*-Kommandos hintereinandergestellt, so wirken sie wie wenn ihre logischen Bedingungen durch *AND*-Verknüpfungen verbunden wären. Das bedeutet: Damit ein Fall "durchkommt", muss er alle hintereinander stehenden *SELECT IF*-Kommandos passieren, mithin alle logischen Bedingungen erfüllen.

**Beispiel:**

```
SELECT IF (v151 EQ 2).
SELECT IF (v388 GT 1000).
FREQUENCIES VARIABLES = v248.
```

Es werden nur solche Fälle selektiert, die bei der Variablen "v151" (Geschlecht) die Ausprägung 2 (weiblich) und bei der Variablen "v388" (Nettomonatseinkommen) einen Wert haben, der größer als 1000 ist.

Wenn zwischen Selektionskommandos Prozedurkommandos stehen, wird die implizite *AND* Verknüpfung der aufeinanderfolgenden Selektionskommandos von den Anwendern häufig nicht beachtet.

**Beispiel:**

```
SELECT IF (v151 EQ 1).
FREQUENCIES VARIABLES = v248.
SELECT IF (v151 EQ 2).
FREQUENCIES VARIABLES = v248.
```

Beim zweiten *SELECT IF*-Kommando gibt es keine Fälle mehr, für welche die logische Bedingung "wahr" ist. SPSS Statistics meldet deshalb beim *FREQUENCIES*-Kommando:

**Warnungen**

In dieser Prozedur wurden keine Fälle eingegeben. Entweder sind keine Fälle in der Arbeitsdatei vorhanden oder alle Fälle wurden herausgefiltert.  
Execution of this command stops.

Alle weiteren Analysen sind sinnlos. Warum kommt es dazu?

Nach dem ersten *SELECT IF*-Kommando existieren nur noch Fälle, die bei der Variablen "v151" (Geschlecht) die Ausprägung 1 besitzen. Für diese Fälle ist die logische Bedingung des zweiten *SELECT IF*-Kommandos falsch, da keiner der Fälle bei der Variablen "v151" (Geschlecht) gleichzeitig die Ausprägung 1 und 2 besitzen kann.

Dieses Problem lässt sich durch Einsatz des *TEMPORARY*-Kommandos umgehen (siehe Seite 56). Das *TEMPORARY*-Kommando macht Selektionen temporär, das heißt, sie gelten nur bis zur nächsten Prozedur. Nach der ersten Prozedur, die dem *TEMPORARY* folgt, existieren wieder alle Fälle.

**Beispiel:**

```
TEMPORARY.
SELECT IF (v151 EQ 1).
FREQUENCIES VARIABLES = v248.
TEMPORARY.
SELECT IF (v151 EQ 2).
```



`FREQUENCIES VARIABLES = v248.`

In diesem Beispiel sind die `SELECT IF`-Kommandos temporär. Nach dem ersten `FREQUENCIES`-Kommando wird das erste `SELECT IF`-Kommando unwirksam. Das zweite `SELECT IF`-Kommando erzielt daher die gewünschte Wirkung. Nach dem zweiten `FREQUENCIES`-Kommando sind auch wieder alle Fälle verfügbar.

Alle Selektionen, die nicht zwischen einem `TEMPORARY` und einer diesem `TEMPORARY` folgenden Prozedur stehen, sind permanent und gelten für den gesamten restlichen Auswertungslauf.

### 7.3 Die Sortierung von Fällen mit SORT CASES

Für die Präsentation der Analysedaten ist es oft günstig, die Untersuchungsobjekte geeignet anzuordnen, das heißt, sie nach bestimmten Kriterien zu sortieren. Das folgende Kommando ist dabei hilfreich:

→ Syntax

```
SORT CASES BY varliste [{(A)}{(D)}] [varliste ...].
```

Der Befehl `SORT CASES` sortiert die Fälle nach den Werten der Variablen der „varliste“ in aufsteigender (A) oder abfallender (D) Reihenfolge. Wenn nichts angegeben ist, wird in aufsteigender Reihenfolge sortiert.

Die `varliste` kann permanente numerische und alphanumerische Variablen (short strings und long strings) enthalten. Bei alphanumerischen Variablen hängt die Sortierreihenfolge vom Zeichensatz ab, den der Rechner verwendet.

Mit folgendem Befehl sortieren wir unsere Daten nach dem Nettomonatseinkommen (v388):

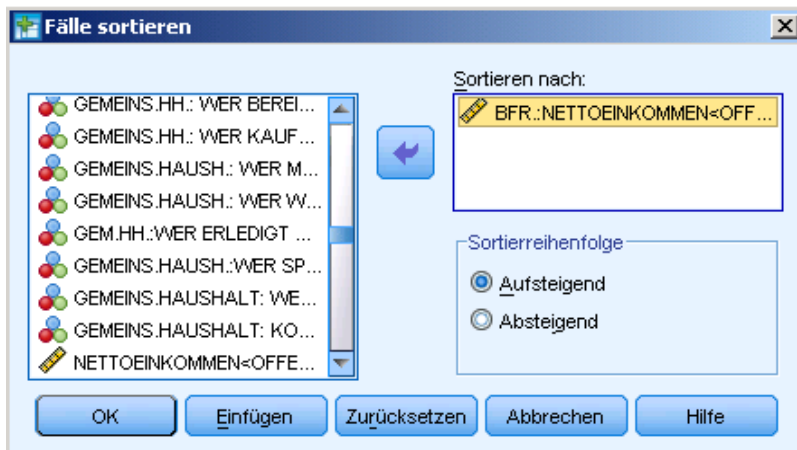
→ Beispiel

`SORT CASES BY v388.`

Menüentsprechung:

→ Menü

DATEN ⇒ FÄLLE SORTIEREN...



### 7.4 Das SPLIT FILE-Kommando

Das `SPLIT FILE`-Kommando teilt den Datensatz in Untergruppen auf, für die jeweils separat eine Analyse zur Verfügung steht.

→ Syntax

```
SPLIT FILE [BY variable] [OFF].
```

**variable** ist eine Variable, nach deren Werten die Unterteilung in Subgruppen vorgenommen wird. Es sollte darauf geachtet werden, dass es sich um diskrete und nicht um stetige Variablen handelt, da es ansonsten passieren könnte, dass Untergruppen mit nur 1 Fall entstehen.

**OFF** schaltet `SPLIT-FILE` wieder aus.

SPSS Statistics prüft, ob ein Fall bei der `SPLIT`-Variablen einen anderen Wert hat als der vorhergehende Fall. Wenn das so ist, wird eine neue Subgruppe gebildet. Der Datensatz muss daher nach den

Werten der *SPLIT FILE*-Variablen sortiert sein, um zu gewährleisten, dass die richtigen Subgruppen gebildet werden. Die Fälle werden dazu mit dem *SORT CASES*-Kommando in die richtige Reihenfolge gebracht (siehe Seite 73).

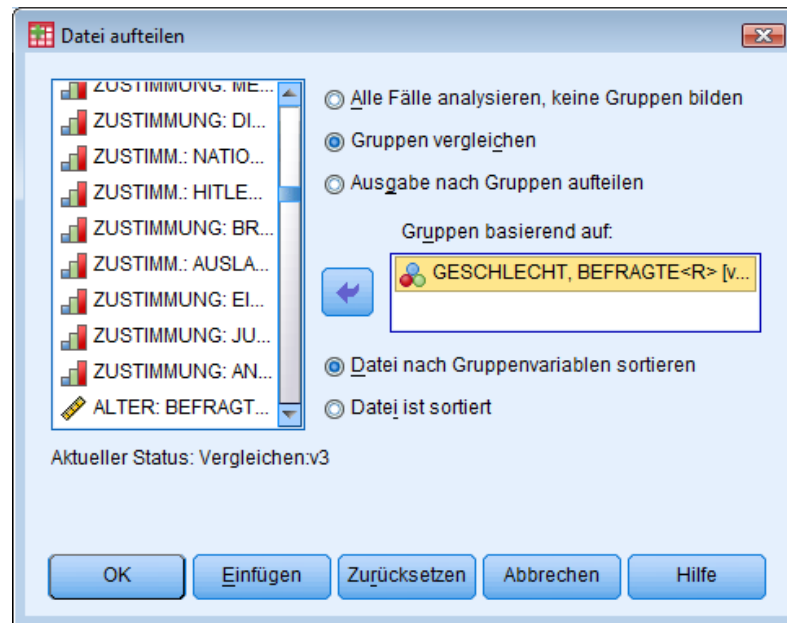
*SPLIT-FILE* wird wirksam, sobald das Kommando ausgeführt wurde und nicht, wie bei Modifikationen üblich, bei der ersten folgenden Prozedur.

Jeder dem *SPLIT-FILE* folgende Befehl wird getrennt für die unterschiedlichen Subgruppen durchgeführt. Es betrifft jedoch immer alle Subgruppen. Modifikationen nach einem *SPLIT-FILE*-Kommando wirken daher immer auf alle Fälle.

Menüentsprechung:

DATEN ⇒ DATEI AUFTEILEN

→ Menü



Der Befehl lautet:

```
SORT CASES BY v151.
SPLIT FILE BY v151.
DESCRIPTIVES v388
  /STATISTICS=MIN MAX MEAN.
SPLIT FILE OFF.
DESCRIPTIVES v388
  /STATISTICS=MIN MAX MEAN.
```

→ Beispiel

→ Ergebnis

Deskriptive Statistik

GESCHLECHT, BEFRAGTE<R>		N	Minimum	Maximum	Mittelwert
MANN	BFR.:NETTOEINKOMMEN<OF FENE+LISTENANGABE> Gültige Werte (Listenweise)	1421	1	8750	1751,28
FRAU	BFR.:NETTOEINKOMMEN<OF FENE+LISTENANGABE> Gültige Werte (Listenweise)	1329	40	8750	1048,37

## Deskriptive Statistik

	N	Minimum	Maximum	Mittelwert
BFR.:NETTOEINKOMMEN<OF FENE+LISTENANGABE>	2750	1	8750	1411,47
Gültige Werte (Listenweise)	2750			

Zuerst werden die Fälle nach dem Geschlecht sortiert. Anschließend wird durch *SPLIT FILE* der Datensatz in die Subgruppen "männlich" und "weiblich" aufgeteilt. Dann werden deskriptive Statistiken angefordert, die getrennt für die Subgruppen ausgeführt werden. Nach einem *SPLIT FILE OFF* werden noch einmal die gleichen Statistiken angefordert, die nun für den Datensatz als Ganzes geliefert werden.

Eine Übersicht zu Vor- und Nachteilen der Datenselektionsverfahren befindet sich im Anhang.

## 8 Grafik in SPSS Statistics für Windows

### 8.1 Arbeiten mit Diagrammen

Mit den Prozeduren des Grafik-Menüs können in SPSS Statistics die Daten zur optischen Darstellung in unterschiedlichen Diagrammtypen veranschaulicht werden. Das folgende Kapitel stellt die Vorgehensweise für einige Diagrammtypen dar und beschreibt, wie das Erscheinungsbild nachträglich verändert werden kann.

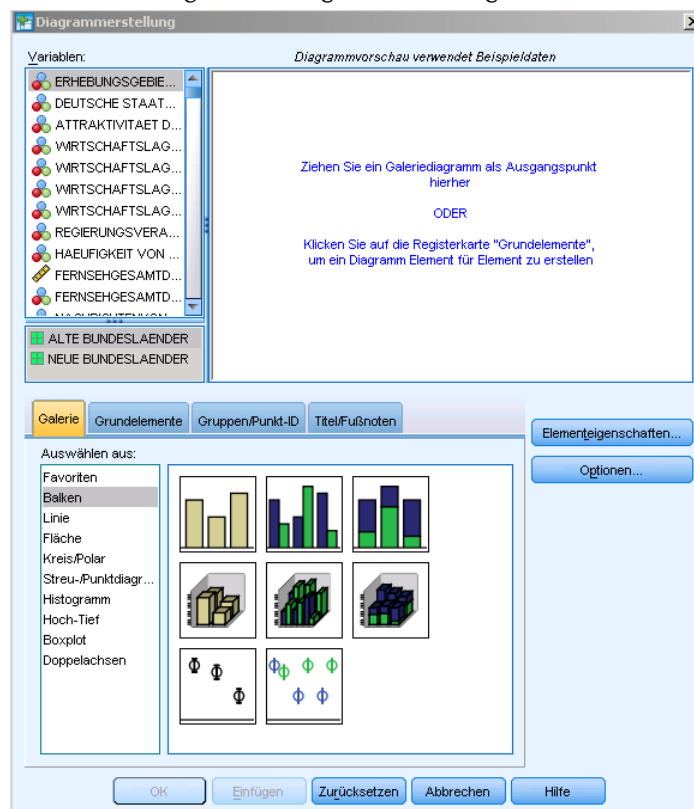
Hierbei wird auf den Menübefehl DIAGRAMME  $\Rightarrow$  DIAGRAMMERSTELLUNG... eingegangen. Diagramme können aber auch über die Abfolge DIAGRAMME  $\Rightarrow$  VERALTETE DIALOGFELDER erstellt werden. Dieser Weg der Diagrammerstellung stammt aus früheren SPSS Statistics-Versionen (SPSS) und wird in den entsprechenden Skripten zu früheren Veranstaltungen beschrieben.

Der Anschaulichkeit halber wird im Folgenden nur die Vorgehensweise über das Menü vorgestellt, da die Grafik-Befehle relativ „aufwändig“ sind. Für jede der Graphiken, die in SPSS Statistics zur Verfügung stehen, besteht selbstverständlich auch die Möglichkeit, sie über die entsprechende Syntax auszuführen. Wie bereits im Kap. 5.1.3 auf S. 31 im Zusammenhang mit der Entscheidung für das Arbeiten mit Befehlen oder Menüs erläutert wurde, kann man sich auch das zugehörige Kommando in ein Syntaxfenster ausgeben lassen, indem man zunächst im Menü alle erforderlichen Felder ausfüllt bzw. aktiviert, und dann anstelle von „Ok“ auf „Einfügen“ klickt.

Das erzeugte Diagramm wird dann zunächst im *Viewer* zur Verfügung gestellt. Nach einem Doppelklick auf das Diagramm öffnet sich der Diagramm-Editor mit weiteren Icons und Menüs zur Bearbeitung der Grafik. Eine Grafik setzt sich aus verschiedenen Objekten (Achsen, Balken, Text, Füllmuster etc.) zusammen, die bearbeitet und verändert werden können.

### 8.2 Einfaches Balkendiagramm

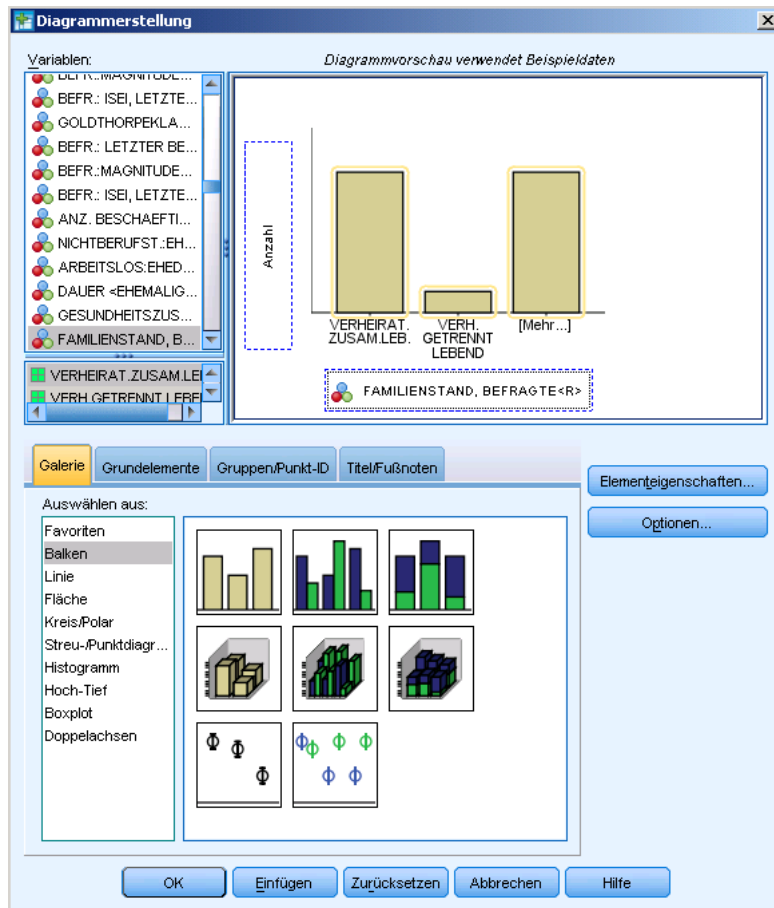
Zur Erstellung eines einfachen Balkendiagramms öffnet der Menübefehl DIAGRAMME  $\Rightarrow$  DIAGRAMMERSTELLUNG... das Dialogfenster *Diagrammerstellung*.



In diesem Fenster können zunächst die Art des Diagramms sowie die zu verwendenden Variablen gewählt werden. Dazu wird mit der Maus auf das gewünschte Diagramm geklickt und dieses in das

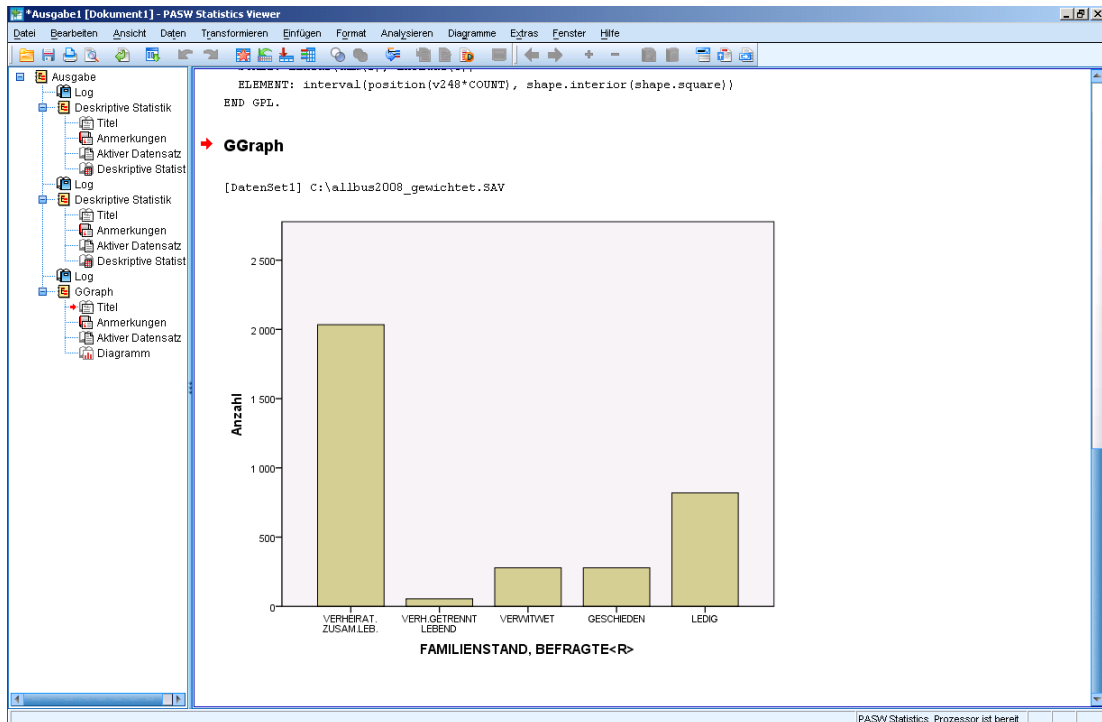
Fenster der Diagrammvorschau gezogen. Analog werden in dem Fenster mit den Variablen diejenigen, die gewünscht sind, angeklickt und auf die Achsen gezogen, auf denen sie abgebildet werden sollen.

In unserem Beispiel soll als Erstes ein Balkendiagramm für die Variable "Familienstand" erstellt werden.

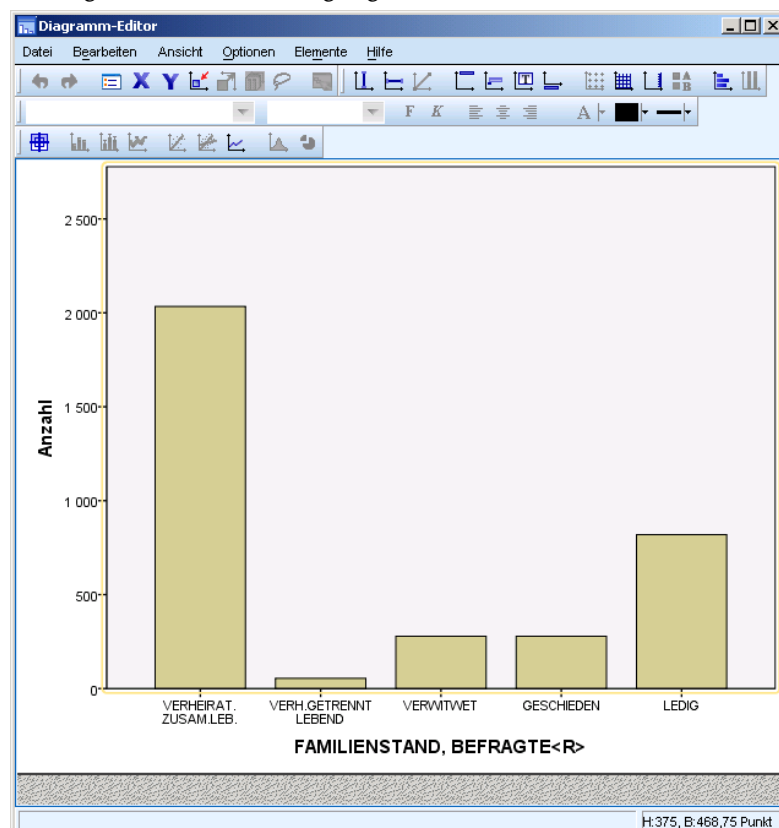


Dazu wird zunächst aus der Galerie das einfache Balkendiagramm in die Diagrammvorschau gezogen. Anschließend wird aus der Variablenliste die entsprechende Variable „FAMILIENSTAND, BEFRAGTE<R>“ ausgewählt und auf das Feld „X-Achse?“ gezogen. Hiermit lassen sich auf der linken Seite die Kategorien der Variablen erkennen.

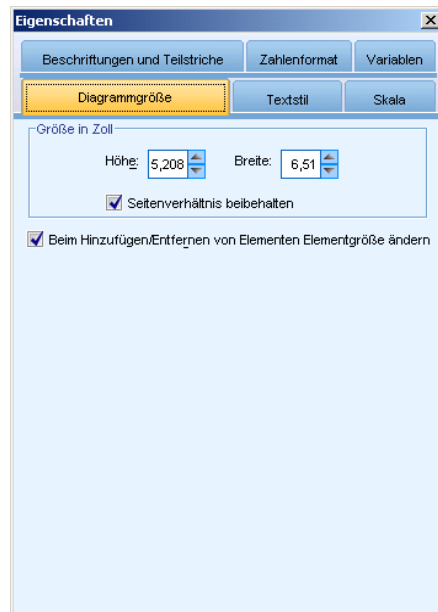
Nachdem nun Diagrammart und Variablen definiert wurden, wird auf die Schaltfläche „OK“ geklickt. Es erscheint folgendes Ergebnis:



SPSS Statistics stellt nun, nach einem Doppelklick auf das neu erstellte Diagramm, im *Diagramm-Editor* verschiedene Möglichkeiten zur Verfügung, um die Grafik individuell zu verändern.



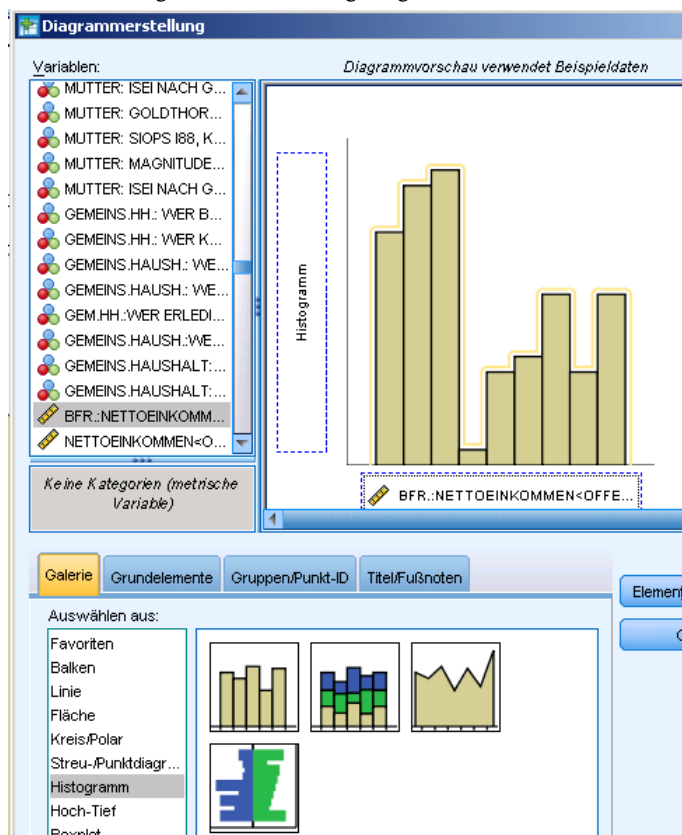
Die verschiedenen Objekte der Grafik, wie Achsen, Balken und Flächen, können nun bearbeitet werden. Durch einen Doppelklick auf die Skalenachse (Y-Achse) öffnet sich beispielsweise das Fenster *Eigenschaften*, das verschiedene Registerkarten enthält, in denen folgende Veränderungen des Diagramms möglich sind.



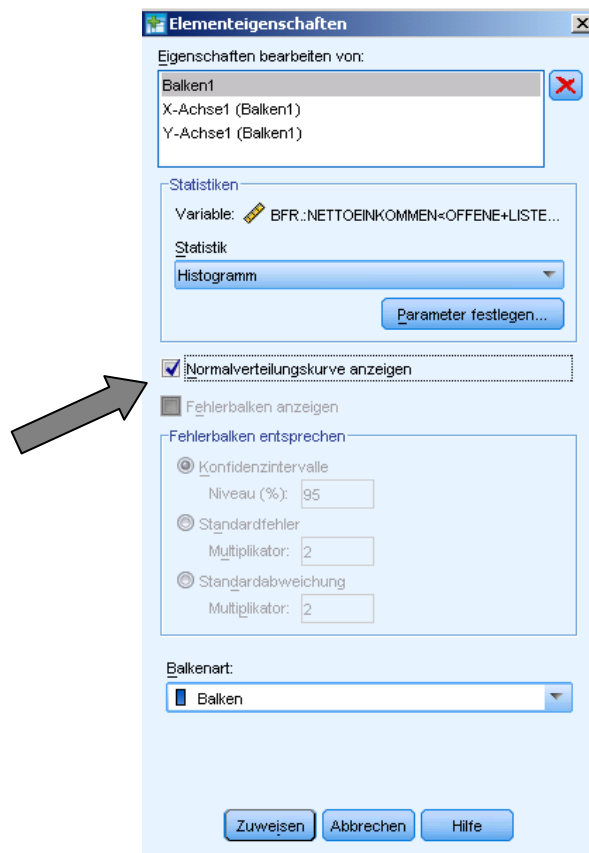
Mit Hilfe der Registerkarte **Diagrammgröße** lässt sich zum Beispiel die Größe des Diagramms bestimmen. Um **Achsentitel** zu ändern, kann dies manuell im Diagramm-Editor vorgenommen werden, indem die bisherige Bezeichnung einfach überschrieben wird.

### 8.3 Histogramm mit Normalverteilungskurve

Zur grafischen Darstellung der Häufigkeitsverteilung von metrischen Variablen eignet sich das Histogramm. In SPSS Statistics kann eine Normalverteilungskurve in ein Histogramm eingefügt werden. Die Erstellung eines Histogramms erfolgt ebenfalls unter dem Menübefehl **DIAGRAMME** ⇒ **DIAGRAMMERSTELLUNG...** Allerdings wird statt eines Balkendiagramms aus der Galerie ein Histogramm ausgewählt und in die Diagrammvorschau gezogen.

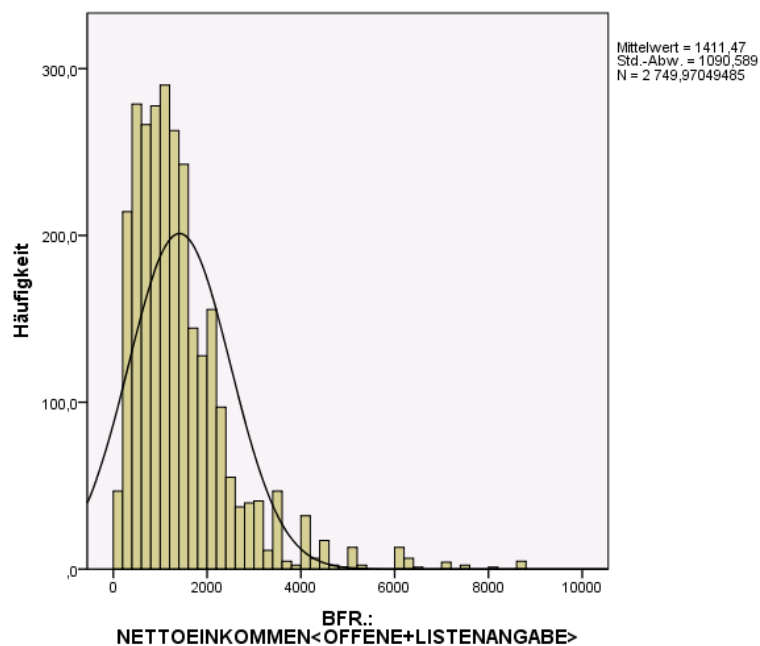


Unser Beispiel zeigt die Verteilung des Nettomonatseinkommens. In dem Fenster „Elementeigenschaften“, das sich mit dem Ziehen des Diagramms in die Diagrammvorschau öffnet, kann „Normalverteilungskurve anzeigen“ gewählt werden.



Die Normalverteilungskurve zeigt die Dichte der Normalverteilung mit dem Mittelwert und der Varianz der Variablen "Nettomonatseinkommen".

Nach der Auswahl der Variablen, der Art des Diagramms und der Normalverteilung öffnet sich die Ausgabe mit dem folgenden Ergebnis:



Durch einen Doppelklick auf das Diagramm im *Viewer* öffnet sich auch hier der Diagramm-Editor, in dem Achsen, Skalen, Kurven, etc. in die gewünschte Form gebracht werden können.

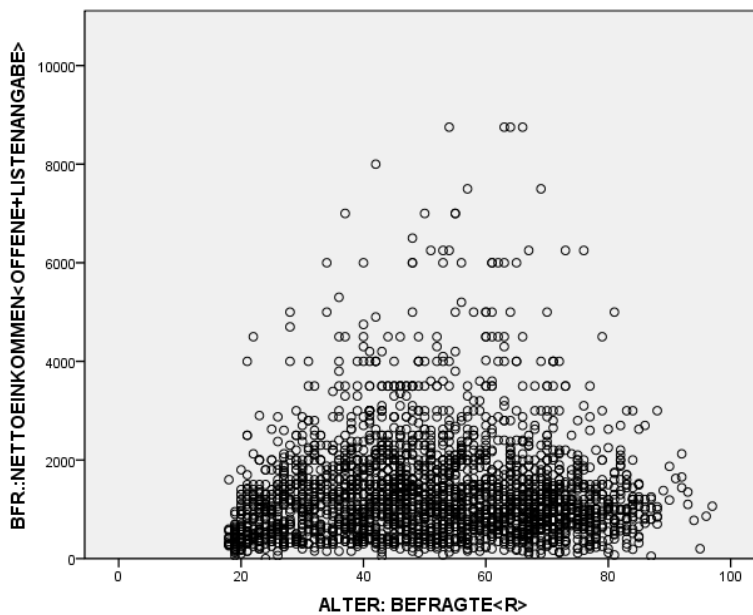


## 8.4 Streudiagramm (Scatterplot) mit Regressionsgerade

Einfache Scatterplots stellen Punkte dar, welche die Objekte repräsentieren. Jedem Objekt ist ein Wertepaar zugeordnet. Jedes Wertepaar ist durch einen Punkt im Diagramm vorhanden. In unserem Beispieldatensatz sind für jedes Objekt Angaben über das Nettomonatseinkommen und über das Alter erfasst. Falls es einen Zusammenhang zwischen dem Alter und der Höhe des Nettomonatseinkommens gibt, dann würden wir erwarten, dass diese Punktwolke eine bestimmte Form annimmt.

Unter dem Menübefehl GRAFIKEN ⇒ DIAGRAMMERSTELLUNG... öffnet sich das oben erläuterte Fenster, aus dem hier ein Streu-/Punktdiagramm ausgewählt wird. Da im vorliegenden Fall ein Zusammenhang untersucht werden soll, müssen zwei Variablen ausgesucht werden. In unserem Beispiel sind diese Alter und Nettoeinkommen der Befragten.

Die X-Achse wird durch die Variable "Alter Befragte<r>" ("v154") und die Y-Achse durch die Variable "Nettomonatseinkommen" ("v388") bestimmt. Das Ergebnis präsentiert sich wie folgt:



Um eine Regressionsgerade einzufügen, bietet der Diagramm-Editor den Menü-Punkt „Elemente“, unter welchem „Anpassungslinie bei Gesamtwert“ gewählt werden kann. Diese kann wiederum im Diagramm-Editor nach subjektiven Wünschen angepasst werden.

## 8.5 Transfer von Grafiken und Analyseergebnissen

Oft steht man nach der Analyse von Daten in SPSS Statistics vor dem Problem, die erzeugten Ergebnisse aus dem *Viewer* in andere Programme transportieren zu wollen. Zum Bearbeiten von SPSS Statistics-Ergebnissen in anderen Programmen reicht meist ein simples „Kopieren“ und „Einfügen“. Oftmals kann es jedoch auch von Vorteil sein, die erzielten Ergebnisse als eigenständige Dateien mit einem bestimmten Dateiformat zu speichern.

### 8.5.1 Kopieren von Ergebnissen oder Befehlen in ein Textverarbeitungsprogramm

Sollen Tabellen aus dem *Viewer* in Word® transportiert werden, können diese einzeln in der linken Hälfte des *Viewers* ausgewählt werden. Um mehrere zu markieren, kann entweder bei fortlaufenden Objekten die Umschalttaste oder bei nicht fortlaufenden Objekten die Steuerungstaste während der Auswahl gedrückt werden. Das gewünschte Objekt wird dann mit der Maus angeklickt.

1. BEARBEITEN ⇒ KOPIEREN wählen, wenn Tabellen als Word-Tabellen zu transportieren sind.
2. Das Textverarbeitungsprogramm aufrufen.

3. Bei Word für Windows, je nach Version, die Funktion EINFÜGEN wählen. Die Tabelle ist nun als „Word-Tabelle“ eingefügt.

Ab SPSS 19.0.0.1 kann über BEARBEITEN ⇒ KOPIEREN SPEZIAL das Format des kopierten Objekts beeinflusst werden.

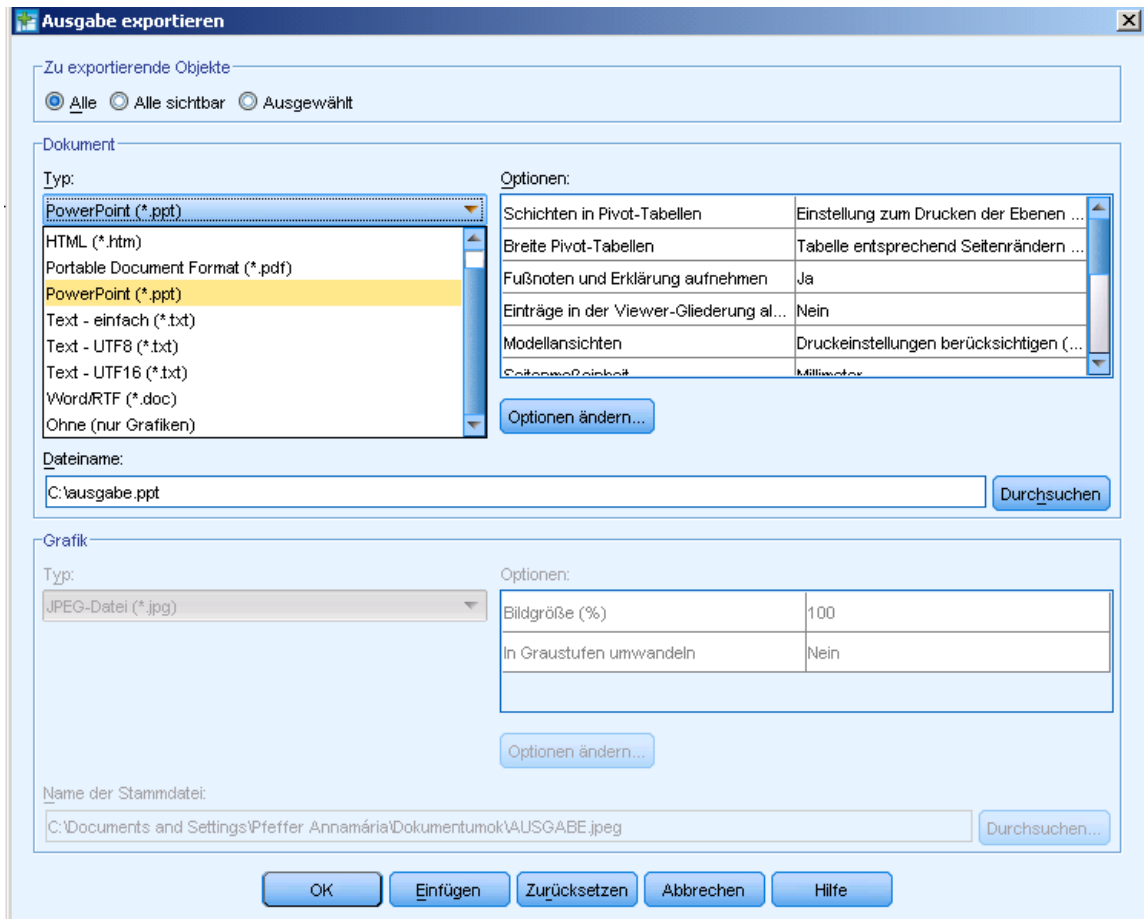
Grafiken können aus dem *Diagramm-Editor* über BEARBEITEN ⇒ DIAGRAMM KOPIEREN kopiert werden. Sie werden aber nur als Bild in Word transportiert. Alternativ könnte ab SPSS 19.0.0.1 zum Kopieren von Grafiken in andere Programme im *Viewer* bei KOPIEREN SPEZIAL Bild gewählt werden.

Darüber hinaus besteht die Möglichkeit, Ergebnisse direkt in andere Formate (Word®, PowerPoint®, Excel®, PDF, HTML, ...) zu exportieren (siehe dazu 8.5.2, S. 82).



## 8.5.2 Umwandlung von Ergebnissen in .htm/.jpg/.txt Dateien

Des Weiteren bietet SPSS Statistics die Möglichkeit, Elemente aus dem *SPSS Statistics Viewer* als Dateien zu exportieren. Zuerst wird das gewünschte Objekt per Mausklick markiert. Die rechte Maustaste öffnet eine Auswahlliste, aus der man durch den Menüschritt EXPORTIEREN in das Fenster *Ausgabe exportieren* gelangt.



In diesem Fenster wird nun unter „Typ:“ eine bestimmte Art der Ausgabe gewählt (Excel, HTML, PowerPoint, Word, ...).

Die so erzeugte Datei kann nun je nach Wunsch weiter bearbeitet werden.

## 9 Anhang

### 9.1 Wichtige Variablen des Beispieldatensatzes

Variablenname	Variablenlabel (Bedeutung der Variablen)	Wertelabel (Bedeutung der Variablenausprägungen)
V3	ERHEBUNGSGEBIET: WEST - OST	1 ‚Alte Bundesländer‘ 2 ‚Neue Bundesländer‘
V6	WIRTSCHAFTSLAGE IN DER BRD HEUTE	1 ‚sehr gut‘ 2 ‚gut‘ 3 ‚teils teils‘ 4 ‚schlecht‘ 5 ‚sehr schlecht‘ 8 ‚weiß nicht‘ 9 ‚keine Angabe‘
V7	WIRTSCHAFTSLAGE, BEFR. HEUTE	1 ‚sehr gut‘ 2 ‚gut‘ 3 ‚teils teils‘ 4 ‚schlecht‘ 5 ‚sehr schlecht‘ 8 ‚weiß nicht‘ 9 ‚keine Angabe‘
V8	WIRTSCHAFTSLAGE DER BRD IN 1 JAHR	1 ‚wesentlich besser‘ 2 ‚etwas besser‘ 3 ‚gleichbleibend‘ 4 ‚etwas schlechter‘ 5 ‚wesentlich schlechter‘ 8 ‚weiß nicht‘ 9 ‚keine Angabe‘
V9	WIRTSCHAFTSLAGE, BEFR. IN 1 JAHR	1 ‚wesentlich besser‘ 2 ‚etwas besser‘ 3 ‚gleichbleibend‘ 4 ‚etwas schlechter‘ 5 ‚wesentlich schlechter‘ 8 ‚weiß nicht‘ 9 ‚keine Angabe‘
V151	GESCHLECHT, BEFRAGTE<R>	1 ‚Mann‘ 2 ‚Frau‘
V154	ALTER: BEFRAGTE<R>	999 ‚keine Angabe‘
V173	ALLGEMEINER SCHULAB- SCHLUSS	1 ‚Ohne Abschluss‘ 2 ‚Volks-, Hauptschulabschluss‘ 3 ‚Mittlere Reife‘ 4 ‚Fachhochschulreife‘ 5 ‚Hochschulreife‘ 6 ‚Anderer Abschluss‘ 7 ‚Noch Schüler‘ 99 ‚Keine Angabe‘
V248	FAMILIENSTAND, BE- FRAGTE<R>	1 ‚verheirat.zusam.leb.‘ 2 ‚verh.getrennt lebend‘ 3 ‚verwitwet‘ 4 ‚geschieden‘

Variablenname	Variablenlabel (Bedeutung der Variablen)	Wertelabel (Bedeutung der Variablenausprägungen)
		5 ,ledig' 9 ,keine Angabe'
V388	BFR.:NETTOEINKOMMEN<OFFENE+LISTENANGABE>	0 ,kein Einkommen' 99997 ,verweigert' 99999 ,keine Angabe'
V448	HAUSHALTSEINKOMMEN <OFFENE+LISTENANGABE>	0 ,kein Einkommen' 99997 ,verweigert' 99999 ,keine Angabe'
V524	KONFESSION, BEFRAGTE<R>	1 ,evang. ohne Freikirch' 2 ,evang. Freikirch' 3 ,roemisch- katholisch' 4 ,and.christl.religion' 5 ,and.nicht-christlich' 6 ,keiner Religionsgem.' 7 ,verweigert' 9 ,keine Angabe'
V526	KIRCHGANGSHAEUEFIGKEIT	1 ,ueber 1x die Woche' 2 ,1x pro Woche' 3 ,1-3x pro Monat' 4 ,mehrmals im Jahr' 5 ,seltener' 6 ,nie' 9 ,keine Angabe'
V792	PERSONENBEZOGENES OST- WEST-GEWICHT	

## 9.2 Wichtige SPSS Statistics Befehle im Überblick

Prozeduren (Kapitel 5)	
Häufigkeitsverteilungen	<i>frequencies v3, v5, v6.</i>
Kreuztabellen	<i>crosstabs v3 by v9.</i>
Deskriptive Kennzahlen (Mittelwerte, ...)	<i>descriptives v12, v388.</i>
Korrelationen	<i>correlations v388, v154.</i>
Mittelwertvergleiche bezüglich Gruppen (v1 ist Gruppierungsvariable)	<i>means v12 v388 by v151.</i>

Modifikationen (Kapitel 6)	
Umkodieren/Rekodieren (Werte ersetzen)	<i>recode v6 (1,2=2) (3,4=1).</i>
Berechnen von Variablen	<i>compute nmeneu = 1.1*v388.</i>
Berechnen in Abhängigkeit von Bedingung	<i>if (v151 = 1 and v388 &lt; 2000) nme1 = 1.2* v388.</i>
Selektionen (Kapitel 7)	
Filtern von Fällen	<i>if (v151 = 1) filter = 0.</i>

(hier auf v1 = 2 ).	<i>if (v151 = 2) filter = 1.</i> ... <i>filter by filter.</i> ... <i>filter off.</i>
Selektieren (löschen) von Fällen	<i>select if (v151 = 2).</i>

### 9.3 Weitere wichtige Prozeduren

Befehl	Menü: ANALYSIEREN	Funktion
ANOVA	⇒ MITTELWERTE VERGLEICHEN ⇒ EINKLASSIFIZIEREN ANOVA ...	Einfaktorielle Varianzanalyse
CLUSTER	⇒ KLASSIFIZIEREN ⇒ HIERARCHISCHE CLUSTER ...	Hierarchische Clusteranalyse (für geringe Fallzahlen)
DISCRIMINANT	⇒ KLASSIFIZIEREN ⇒ DISKRIMINANZ-ANALYSE	Diskriminanzanalyse
EXAMINE	⇒ DESKRIPTIVE STATISTIKEN ⇒ EXPLORATIVE DATENANALYSE ...	Explorative Datenanalyse
FACTOR	⇒ DIMENSIONSREDUZIERUNG ⇒ FAKTORENANALYSE ...	Faktorenanalyse
MULT RESPONSE	⇒ MEHRFACHANTWORTEN ⇒ ...	Analyse von Mehrfachantwortfragen
NPAR TEST	⇒ NICHTPARAMETRISCHE TESTS ⇒ ...	Nichtparametrische Tests
RELIABILITY	⇒ SKALIERUNG ⇒ RELIABILITÄTS-ANALYSE ...	Itemanalysen
SUMMARIZE	⇒ BERICHTE ⇒ FÄLLE ZUSAMMENFASSEN ...	Fälle zusammenfassen
TABLES	⇒ TABELLEN ⇒ ...	Komfortable Tabellen erzeugen
T-TEST	⇒ MITTELWERTE VERGLEICHEN ⇒ T-TEST ...	T-Tests

Da das Gesamtsystem SPSS Statistics, wie auf Seite 6 bereits ausgeführt wurde, aus mehreren Modulen besteht, entstammen die aufgeführten Prozeduren zum Teil verschiedenen SPSS Statistics-Zusatzprodukten (wie ADVANCED STATISTICS). Aufgrund der Komplexität der Prozeduren ist es nicht möglich, diese im Skript detaillierter zu erörtern. Einzelheiten zu den Prozeduren sind daher in den entsprechenden Handbüchern nachzulesen.

### 9.4 Übersicht: Vor- und Nachteile der Datenselektionsverfahren

Verfahren	Wirkung	Vorteile	Nachteile
SELECT IF	wählt Fälle nach einer logischen Bedingung aus	logische Bedingung als Auswahlkriterium möglich folgende Modifikationen	Fälle können "aus Versehen" völlig aus dem Datensatz entfernt werden

Verfahren	Wirkung	Vorteile	Nachteile
		nur für selektierte Fälle Fälle können bei Bedarf völlig aus dem Daten- satz entfernt werden	
FILTER	wählt Fälle anhand ei- ner Filtervariablen aus	Fälle können nicht "aus Versehen" völlig aus dem Datensatz entfernt werden kann einfach ein- und ausgeschaltet werden	Datenmodifikationen können nicht für gefil- terte Fälle getrennt durchgeführt werden logische Variable zum Filtern nötig
SPLIT FILE	teilt die Fälle in Sub- gruppen auf	mit einer Prozedur wer- den alle Subgruppen analysiert Fälle können nicht "aus Versehen" völlig aus dem Datensatz entfernt werden kann einfach ein- und ausgeschaltet werden	Datenmodifikationen können nicht für Sub- gruppen getrennt durchgeführt werden Sortierung der Variab- len vor Unterteilung nötig nur bei diskreten Va- riablen sinnvoll

## 9.5 Bivariate Assoziationsmaße

		UNABHÄNGIGE VARIABLE			
		Skalenniveau	nominal (klassifikato- risch)	ordinal (komparativ)	metrisch (Inter- vall- oder Ratio- skala)
ABHÄNGIGE VARIABLE	nominal (klassifikatorisch)		Chi-Quadrat $\chi^2$ [0;...] Phi $\Phi$ [0;1] Cramers V [0;1]		
	ordinal (komparativ)			Kendalls tau $\tau$ [-1;1] Kontingenzmaß von Goodman/Kruskal $\gamma$ [-1;1]	
	metrisch (Intervall- oder Ratioskala)		Eta $\eta$ [0;1]		Pearson's r [-1;1]

## 10 Literaturverzeichnis

### Literatur dieser Arbeit:

- Angele, G., Anerkannte mathematisch-statistische Methoden zur Stichprobeninventur - Entscheidungshilfen für die Praxis, München 1989.
- Angele, G., Statistikprogramm Pakete - Ein Überblick, Wirtschaftswissenschaftliches Studium (WiSt) Heft 11 1989a, S. 577-579
- Dobbener, R., Grundlagen der numerischen Klassifikation anhand gemischter Merkmale, Göttingen 1983.
- Fahrmeir, L./Kaufmann, H./Kredler, C., Regressionsanalyse, in: Fahrmeir, L./Hamerle, A. (Hrsg.) Multivariate statistische Verfahren, Berlin New York 1984.
- Friedrichs, J., Methoden empirischer Sozialforschung, 10. Auflage, Opladen 1982.
- Hartung, J./Elpelt, B., Multivariate Statistik - Lehr- und Handbuch der angewandten Statistik, München Wien 1984.
- Kähler, W.-M., SPSS X für Anfänger - Eine Einführung in das Datenanalyse-System, Braunschweig/Wiesbaden 1986.
- Küffner, H./Wittenberg, R., Datenanalyse-Systeme für statistische Auswertungen - Eine Einführung in SPSS, BMDP und SAS, Stuttgart New York 1985.
- Little R.J.A./Rubin D.B., Statistical Analysis with Missing Data, New York 1987.
- Matiaske, W., Statistische Datenanalyse mit Mikrocomputern - Einführung in P-STAT und SPSS/PC, München Wien 1990.
- Rowe, B.C./Westlake, A./Rose, P., Software for statistical and social survey analysis 1989-90 prepared for the Study Group on Computers in Survey Analysis, Computational Statistics & Data Analysis 9 (1990), S. 317-340, North-Holland.
- Schaich, E., Schätz- und Testmethoden für Sozialwissenschaftler, München 1977.
- Schubö, W., et al., SPSS - Handbuch der Programmversionen 4.0 und SPSS-X 3.0, Stuttgart New York 1991.
- SPSS Inc., SPSS Reference Guide, Chicago 1990.
- Ulmer, H., Veränderungen der Volkskultur durch Migration in das Armenviertel Lucero von Bogota. Die Hauptstadt Kolumbiens, Bamberg 1989.
- Unger, H., Ausstellungen in wissenschaftlichen Universalbibliotheken der Bundesrepublik Deutschland - Ergebnisse einer Umfrage, in: Karasek, D. (Hrsg.) Buch und Bibliothek in Bamberg - Festschrift zur Einweihung des zentralen Bibliotheksgebäudes der Universitätsbibliothek, Schriften der Universitätsbibliothek Bamberg, Band 3, Bamberg 1986.
- Wilke, H., Eine Analyse der Nutzung Statistischer Programmpakete, Berlin 1986.
- Woodward, W.A./Elliot, A.C./Gray, H.L./Matlock, D.C., Directory of Statistical Microcomputer Software, New York Basel, 1985.
- Vogel, F., Beschreibende und schließende Statistik - Formeln, Definitionen, Erläuterungen, Stichwörter und Tabellen, München Wien 1989.
- Vogel, F., Studienskript: Parametrische und nicht-parametrische (verteilungsfreie) Schätz- und Testverfahren, Bamberg 1989a.
- ZA4601:  
ALLBUS Datenhandbuch 2008 (ZA-Studien-Nr. 4600), Köln und Mannheim: GESIS 2009.

**Originaldokumentation zu SPSS:**

Eine aktuelle Liste mit der gesamten Originaldokumentation zu SPSS finden Leserinnen und Leser über:

[http://www-947.ibm.com/support/entry/portal/Documentation/Software/Information\\_Management/SPSS\\_Statistics](http://www-947.ibm.com/support/entry/portal/Documentation/Software/Information_Management/SPSS_Statistics)

**Weitere Literatur zu SPSS:**

Bühl, A., SPSS 20, Einführung in die moderne Datenanalyse, München 2012

Brosius, F., SPSS 20 für Dummies, Wiley-VCH Verlag, 2012.

Brosius, F., SPSS 19, Verlagsgruppe Hürtig Jehle Rehm, Heidelberg 2011.

Akreimi, L./Baur, N./Fromm, S., Datenanalyse mit SPSS für Fortgeschrittene, 3. Auflage, Wiesbaden 2011.

Backhaus, K./ Erichson, B./ Plinke, W./ Weiber, R. (Hg.), Multivariate Analysemethoden. Eine anwendungsorientierte Einführung, 13. Auflage, Berlin 2011.

Fromm, S., Datenanalyse mit SPSS für Fortgeschrittene 2: Multivariate Verfahren für Querschnittsdaten, Wiesbaden 2010.

Sarstedt, M./ Schütz, T./ Raithel, S., IBM SPSS Syntax - Eine anwendungsorientierte Einführung. 2. Auflage, Vahlen Verlag, München 2010.

Hatzinger, R./Nagel, H., PASW Statistics: Statistische Methoden und Fallbeispiele, München 2009.

Kinney, P.R./Gray, C.D., PASW Statistics 17 Made Simple, 2009 Psychology Press.

Janssen, J./Laatz, W., Statistische Datenanalyse mit SPSS für Windows, Berlin 2010.

Bühl, A., SPSS 18 (ehemals PASW). Einführung in die moderne Datenanalyse, 12. Auflage, Pearson Studium, 2009.

Kinney, P.R./Gray, C.D., PASW Statistics 17 Made Simple, 2009 Psychology Press.

Eckstein, P.P., Angewandte Statistik mit SPSS, Gabler-Verlag 2008.

Backhaus, K./ Erichson, B./ Plinke, W./ Weiber, R. (Hg.), Multivariate Analysemethoden. Eine anwendungsorientierte Einführung, 12. Auflage, Berlin 2008.

Duller, Ch., Einführung in die Statistik mit Excel und SPSS, Physika-Verlag 2005.

Bühl, A./ Zöfel, P., SPSS Version 12. Einführung in die moderne Datenanalyse unter Windows, München 2004.

Martens, J., Statistische Datenanalyse mit SPSS für Windows, Oldenburg 2003.

Wittenberg, R., Cramer, H., Handbuch für computerunterstützte Datenanalyse 9. Datenanalyse mit SPSS für Windows 95/ NT, Stuttgart 2003.



# 11 Index

## A

Advanced Statistics 11  
 aktiver Datensatz 11  
 Allbus 15  
 Anführungszeichen 33  
 ANOVA 85  
 Arithmetische Operatoren 33, 51  
 Assoziationsmaße 86

## B

Balkendiagramm  
 einfach 76  
 Begrenzungszeichen 33

## C

case 17  
 CLUSTER 85  
 Clusteranalyse 85  
 command keyword 31  
 COMPUTE 51  
 CORRELATIONS 43  
 COUNT 54  
 CROSSTABS 39

## D

DATASET ACTIVATE 11  
 Datenansicht 8  
 Datendefinition 18, 25  
 Daten-Editor 7  
 Datenerhebung 15  
 Datenmatrix 17  
 Datenmodifikation 47  
 Datensatz 17, 25  
 Datenselektion 68  
 defaults 34  
 Delimiters 31  
 DESCRIPTIVES 41  
 Deskriptive Statistik 41  
 DISCRIMINANT 85  
 Diskriminanzanalyse 85  
 DO IF 63

## E

ELSE 48  
 Etikett 22  
 Explorative Datenanalyse 85

## F

FACTOR 85  
 Faktorenanalyse 85  
 Fall 17  
 Fehlende Werte  
 siehe auch missing values 23  
 FREQUENCIES 37

Funktion 51, 52

## G

gewichteten 57  
 Grafik 76

## H

Häufigkeiten 37  
 Häufigkeitsverteilung 37  
 HI 24, 47  
 HIGHEST 24, 47  
 Hilfe 11  
 Histogramm 79  
 Hochkommata 33

## I

IF 59  
 Ite-manalysen 85

## K

Kommandos 30  
 Kommandoschlüsselwort 31  
 Kopieren von Ergebnissen oder Befehlen in ein  
 Textverarbeitungsprogramm 81  
 Korrelation  
 -skoeffizient 43  
 -smatrix 44  
 Test des -skoeffizienten 43  
 Kreuztabelle 39

## L

Label 22  
 Laden 10  
 LIST 36  
 LO 24, 47  
 logischer Operator 60  
 LOWEST 24, 47

## M

Maximum 41  
 MEANS 44  
 mehrere Datensätze 11  
 Mehrfachantwortfrage 85  
 Menüsteuerung 26  
 Minimum 41  
 MISSING 47  
 MISSING VALUES 23  
 Mittelwert 36, 37, 41  
 Subgruppen 44  
 -vergleich 44  
 Modifikation  
 der Daten 47  
 in Abhängigkeit von Bedingungen 59  
 numerischer Variablen 47  
 temporär 56

- 
- N**
- Numbers 31
- R**
- RECODE 47  
 Rekodierung 47  
 relational operator 60  
 Relationen 60  
 reserved keywords 32  
 Rohdaten 25
- S**
- SAVE 66  
 Schlüsselwort 31  
 scratch variables 32  
 Select If 70  
 SORT CASES 73  
 Sortierung von Fällen 73  
 special delimiters 33  
 Speichern 10  
 SPLIT FILE 73  
 Standardabweichung 41  
   Subgruppen 44  
 Statistik  
   deskriptive 41  
 Streudiagramm 81  
 string 73  
 Subkommando 31  
 Syntax 31  
 Syntax-Editor 7  
 system-missing value 23  
 Systemvariable 32
- T**
- TEMPORARY 56  
 THRU 24, 47  
 T-TEST 85
- U**
- user-missing value 23
- V**
- Variable  
   Hilfs- 32  
   -nnamen 32  
   numerische erzeugen 19  
 Variablenansicht 8  
 Varianzanalyse 85  
 Vergleichsoperator 60  
 Verkodung 16  
 Voreinstellungen 34
- W**
- Wahrheitswert 60  
 Weight 57  
 Wertebereich 24  
 Wertzuweisungen 33  
 Windows Viewer 7
- X**
- XSAVE 66