

Quantitative Datenanalyse: Arbeiten mit SPSS Statistics¹

Eine Einführung

Dr. oec. HSG Sabine Hoidn
sabine.hoidn@unisg.ch

St. Gallen, 12.08.2013

¹ Übungsdateien und -blätter sind bei der Dozierenden erhältlich (Sabine.Hoidn@unisg.ch)

Zitierte Literatur – WAFT

- Bortz, J. & Döring, N. (2006). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler*. 4. Auflage. Heidelberg: Springer.
- Bühl, A. (2008). *SPSS 16 - Einführung in die moderne Datenanalyse*. 11. Auflage. München et al.: Pearson Studium.
- Bühner, M. (2006). *Einführung in die Test- und Fragebogenkonstruktion*. 2., aktualisierte Auflage. München: Pearson. (S. 46-76).
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. 2. Aufl., Hillsdale, NJ: Lawrence Erlbaum.
- Droß, C. & Team (2007). *Skript zum SPSS / Statistik Kompaktkurs*. Berlin.
- Eckstein, P. P. (2008). *Angewandte Statistik mit SPSS. Praktische Einführung für Wirtschaftswissenschaftler*. 6. Auflage. Wiesbaden: Gabler.
- Friese, S. (2008). *Kurzeinführung in atlas ti - the knowledge workbench*. Abgerufen am 20. 11. 2011 von http://www.atlasti.com/uploads/media/QuickTour_a6_de_01.pdf
- Gurtner, A., Tschan, F. & Bogenstätter, Y. (2007). *Deutsche Übersetzung des Team Diagnostic Surveys (TDS) von Wageman, Hackman und Lehman (2005)*, unveröffentlichter Fragebogen. Bern: BFH.
- Kuckartz, U. (2007). *Einführung in die computergestützte Analyse qualitativer Daten*. 2. Auflage. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Kuckartz, U., Grunenberg, H. & Dresing, T. (Hrsg.) (2007). *Qualitative Datenanalyse: computergestützt. Methodische Hintergründe und Beispiele aus der Forschungspraxis*. Wiesbaden: VS Verlag.
- Langer, I., Schulz von Thun, F. & Tausch, R. (2006). *Sich verständlich ausdrücken*. 8. Auflage. München: Reinhardt.
- Lewins, A. & Silver, C. (2007). *Using Software in Qualitative Research: A Step-by-Step Guide*. London: Sage.
- Mayer, H. O. (2009). *Interview und schriftliche Befragung. Entwicklung, Durchführung und Auswertung*. 5. Auflage. München: Oldenbourg.
- Müller-Benedict, V. (2007). *Grundkurs Statistik in den Sozialwissenschaften*. 4. Auflage. Wiesbaden: VS Verlag für Sozialwissenschaften. (Online-Buch)
- Muhr, T. (2004). *User's Manual for ATLAS.ti 5.0. 2nd edition*. Gefunden am 20.11.2011 unter http://www.atlasti.com/uploads/media/atlman_02.pdf
- Quatember, A. (2008). *Statistik ohne Angst vor Formeln. Das Studienbuch für Wirtschafts- und Sozialwissenschaftler*. 2. Auflage. München: Pearson. (Online-Buch)
- Rasch, B., Friese, M., Hofmann, W. & Naumann, E. (2008). *Quantitative Methoden*. 2. Auflage. Heidelberg: Springer. (Online-Buch)

- Rasch, B., Frieze, M., Hofmann, W. & Naumann, E. (2008). *Quantitative Methoden* 2. 2. Auflage. Heidelberg: Springer. (Online-Buch)
-> vgl. auch Onlinematerialien unter www.quantitative-methoden.de
- Rohrmann, B. (1978). Empirische Studien zur Entwicklung von Antwortskalen für die sozialwissenschaftliche Forschung. *Zeitschrift für Sozialpsychologie*, (9)1, S. 222-245.
- Schnell, R., Hill, P. B. & Esser, E. (1999). *Methoden der empirischen Sozialforschung*. 6. Auflage. München et al.: Oldenbourg.
- SPSS Inc. (2009). *Einführungskurs. Arbeiten mit IBM SPSS Statistics 18*. Zürich: SPSS (Schweiz) AG.
- SPSS Inc. (2010). *IBM SPSS Statistics 19 – Schneller Einstieg*. Gefunden am 07.01.2012 unter http://www.id.uzh.ch/cl/dl/sw/statmath/spss/spsskurs/SPSS19_Schnelleinstieg.pdf
- Ziegler, M. & Bühner, M. (2009). *Statistik für Psychologen und Sozialwissenschaftler*. München: Pearson.



Quantitative Datenanalyse

Arbeiten mit SPSS Statistics

Bei diesem Buch handelt es sich um eine verständliche Einführung in das Arbeiten mit dem quantitativen Datenanalyseprogramm SPSS Statistics. Sie setzen sich mit den Grundlagen zum Arbeiten mit SPSS Statistics auseinander, können selbständig eine SPSS Statistics-Datei erstellen, Daten einlesen, zusammenfügen und modifizieren, erfasste Daten managen und die gebräuchlichsten tabellarischen und grafischen Grundausswertungen sowie Analyseverfahren durchführen und interpretieren.

Inhaltsverzeichnis

Grundlagen zum Arbeiten mit SPSS Statistics	8
1. Starten von SPSS Statistics.....	8
2. Daten in den Daten-Editor laden.....	9
3. Die acht Fenster bzw. Editoren von SPSS Statistics	12
4. Die Menüstruktur von SPSS Statistics	13
5. Dialogfelder in SPSS Statistics	14
6. Das Hilfesystem in SPSS Statistics	16
7. Arbeiten mit mehreren Daten-Dateien	17
8. SPSS Statistics-Sitzung beenden	18
 Dateneingabe in SPSS Statistics.....	19
1. Definition von Variablen	19
2. Variablendefinitionen kopieren.....	25
3. Beispiel zur Eingabe von neuen Variablen	26
4. Erstmaliges Speichern der Daten	29
 Ausgabe in SPSS Statistics (SPSS Statistics Viewer)	31
1. Erzeugen von Ausgaben.....	31
2. Anzeigen und Verbergen von Ausgaben	33
3. Umstellen und Kopieren von Ausgaben.....	34
4. Editieren von Tabellen und Grafiken.....	35
5. Speichern der Ausgabe	36
 Datenübernahme: Dateien einlesen	37
 Dateien zusammenfügen.....	40
 Variablen- und Datenübersicht.....	49
1. Variablenübersicht	49
2. Datenübersicht.....	54
2.1 Kategoriale Variablen.....	54
2.2 Metrische Variablen	58
2.3 Codebuch.....	61

Modifikation von Daten	64
1. Umkodieren von Variablen.....	64
1.1 Umkodieren in dieselben Variablen (manuell)	64
1.2 Umkodieren in andere Variablen (manuell)	66
1.3 Automatisch umkodieren	70
2. Variablen berechnen.....	73
Datenmanagement am Beispiel eines Fragebogens	75
1. Konsistenzprüfungen zwischen Variablen	75
2. Mehrfachantworten-Sets	79
3. Fälle auswählen	83
4. Werte umkodieren.....	88
5. Werte berechnen	91
6. Kategorisierung von metrischen Variablen	108
7. Daten aggregieren	113
Datenanalyse	120
1. Allgemeines	120
2. Kreuztabellen und Chi-Quadrat-Test	123
2.1 Vorgehen in SPSS Statistics:.....	124
2.2 Grafische Veranschaulichung von Kreuztabellen	141
2.3 Der Chi-Quadrat-Test	143
3. T-Tests (Mittelwertvergleiche)	146
3.1 Abhängigkeit versus Unabhängigkeit von Stichproben.....	147
3.2 T-Test für unabhängige Stichproben	147
3.3 T-Test für abhängige (gepaarte/verbundene) Stichproben.....	155
3.4 Einstichproben-t-Test.....	158
4. Korrelationen.....	161
4.1 Korrelationsrechnung nach Pearson (Produkt-Moment-Korrelation)	162
4.2 Korrelationsrechnung bei einer Rangkorrelation.....	166
4.3 Korrelationsmasse – Beispiele.....	169
4.4 Partielle Korrelationen.....	174

5. Lineare Regression.....	178
5.1 Einfache lineare Regression	179
5.2 Vorgehen in SPSS Statistics.....	179

Grundlagen zum Arbeiten mit SPSS Statistics

1. Starten von SPSS Statistics

SPSS Statistics aus dem Windows Start-Menü öffnen: Klicken Sie in der Taskleiste am unteren Bildschirmrand auf **Start** und wählen Sie **Alle Programme** aus:

IBM SPSS Statistics | IBM SPSS Statistics 20

Es erscheint ein Startdialog mit verschiedenen Optionen:

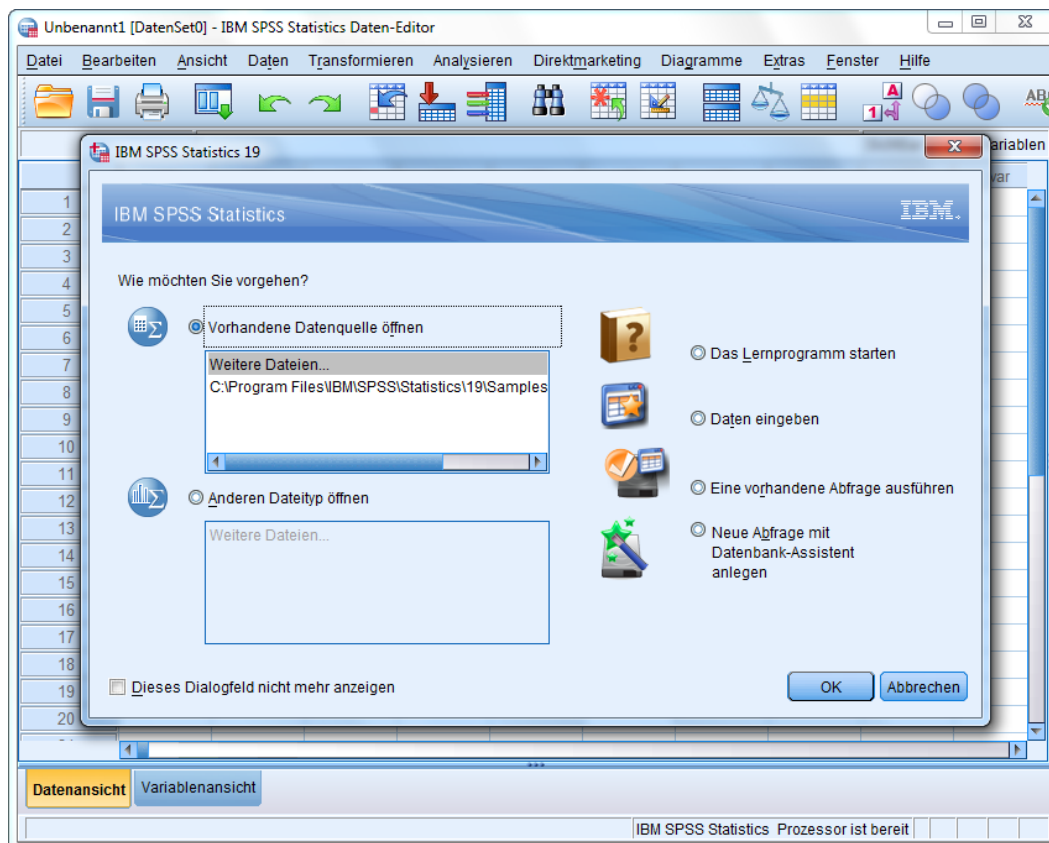


Abbildung 1: Ansicht "SPSS Statistics nach dem Start"

Wählen Sie **Abbrechen** um direkt zum Daten-Editor zu gelangen.

2. Daten in den Daten-Editor laden

Gehen Sie auf

Datei | Öffnen | Daten

und wählen Sie die Datei **demo.sav** aus (diese Datei finden Sie auch im SPSS Statistics 20 Ordner unter Beispieldateien).² Die **Datenansicht** öffnet sich und es werden die einzelnen Datensätze angezeigt:

	alter	heirat	adresse	einkomm	eink_kl	auto	auto_kl	schulab	arbeit
1	55	1	12	72.00	3.00	36.20	3.00	1	23
2	56	0	29	153.00	4.00	76.90	3.00	1	35
3	28	1	9	28.00	2.00	13.70	1.00	3	4
4	24	1	4	26.00	2.00	12.50	1.00	4	0
5	25	0	2	23.00	1.00	11.30	1.00	2	5
6	45	1	9	76.00	4.00	37.20	3.00	3	13
7	42	0	19	40.00	2.00	19.80	2.00	3	10
8	35	0	15	57.00	3.00	28.20	2.00	2	1
9	46	0	26	24.00	1.00	12.20	1.00	1	11
10	34	1	0	89.00	4.00	46.10	3.00	3	12
11	55	1	17	72.00	3.00	35.50	3.00	3	2
12	28	0	3	24.00	1.00	11.80	1.00	4	4
13	31	1	9	40.00	2.00	21.30	2.00	4	0
14	42	0	8	137.00	4.00	68.90	3.00	3	3
15	35	0	8	70.00	3.00	34.10	3.00	3	9
16	52	1	24	159.00	4.00	78.90	3.00	4	16
17	21	1	1	37.00	2.00	18.60	2.00	3	0
18	32	0	0	28.00	2.00	13.70	1.00	1	2
19	42	0	9	109.00	4.00	54.70	3.00	3	20
20	40	1	12	117.00	4.00	58.30	3.00	2	19

Abbildung 2: Datenansicht im Daten-Editor

² Bei **demo.sav** handelt es sich um eine hypothetische Datendatei, bei der es um eine Kundendatenbank geht, die zum Zwecke der Zusendung monatlicher Angebote erworben wurde. Neben verschiedenen demografischen Informationen ist erfasst, ob der Kunde auf das Angebot geantwortet hat.

Die **Datenansicht** im **Daten-Editor** öffnet sich automatisch nach dem Starten von SPSS Statistics. Es handelt sich um ein Spreadsheet-ähnliches Fenster zum Erzeugen und Editieren von SPSS Statistics Datendateien:

- Jede **Zeile** entspricht einem Fall oder einer Beobachtung. In einer Umfrage ist z.B. jede befragte Person ein Fall.
- Jede **Spalte** entspricht einer Variablen, d.h. einem Merkmal des Falles. Beispielsweise dient jede Frage eines Fragebogens zur Erfassung eines Merkmals der befragten Person.
- Eine **Zelle** ist der Schnittpunkt von Fall und Variable. Jede Zelle enthält einen einzelnen Wert einer Variablen für den jeweiligen Fall. Die Zellen des Daten-Editors enthalten nur Datenwerte.
- In einer Datendatei enthalten alle Zeilen die gleiche Anzahl Zellen. Die Dimensionen der Datendatei werden von der Anzahl der Fälle und Variablen bestimmt. In alle Zellen können Daten eingegeben werden. Wenn Daten in eine Zelle ausserhalb der Grenzen der definierten Datendatei eingegeben werden, erweitert SPSS Statistics das Datenfeld, so dass es alle Zeilen und/oder Spalten einschliesst, die zwischen dieser Zelle und den Grenzen der Datendatei liegen. Innerhalb der Grenzen der Datendatei gibt es keine "leeren" Zellen. Bei numerischen Variablen werden leere Zellen zum systemdefiniert fehlenden Wert konvertiert. Bei String-Variablen gelten leere Felder als gültiger Wert.

Im Daten-Editor stehen Symbole als Arbeitserleichterungen zur Verfügung, so dass viele Funktionen direkt über die Symbolleiste ausgeführt werden können:



1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19

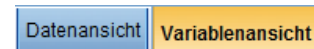
- | | |
|-----------------------------------|-----------------------------|
| 1 Datei öffnen | 11 Fall einfügen |
| 2 Datei speichern | 12 Variable einfügen |
| 3 Drucken | 13 Datei aufteilen |
| 4 Zuletzt verwendete Dialogfelder | 14 Fälle gewichten |
| 5 Rückgängig machen | 15 Fälle auswählen |
| 6 Wiederholen | 16 Werte-Labels anzeigen |
| 7 Gehe zu Fall | 17 Variablen-Sets verwenden |
| 8 Gehe zu Variable | 18 Alle Variablen anzeigen |
| 9 Variablenliste | 19 Rechtschreibprüfung |
| 10 Suchen | |

Um in der **Datenansicht** statt der Codes (Zahlen) Wertelabels anzeigen zu lassen, wählen Sie beispielsweise

Ansicht | Wertelabels oder

das Symbol für Wertelabels  in der Symbolleiste aus.

Neben der Datenansicht können Sie sich auch die Variablenansicht anzeigen lassen, indem Sie auf die Registerkarte **Variablenansicht** links unten im Fenster klicken. In der Datenansicht wird der Datensatz mit den einzelnen Fällen und Variablen angezeigt, in der Variablenansicht sind die Variablenattribute definiert.



3. Die acht Fenster bzw. Editoren von SPSS Statistics

Neben dem Daten-Editor, den Sie gerade geladen haben, beinhaltet SPSS Statistics noch sieben weitere Editoren. Im Folgenden werden die acht Fenster bzw. Editoren kurz dargestellt:

(1) Daten-Editor

Beim Daten-Editor handelt es sich um ein Datenblatt zum Definieren, Eingeben, Editieren und Anzeigen der zu analysierenden Daten; Er muss innerhalb einer Statistics-Sitzung immer geöffnet sein.

(2) Viewer

Der Viewer zeigt alle Ergebnisse, d.h. Statistiken, Grafiken und Tabellen an. Diese Objekte können mittels Doppelklick im Viewer aktiviert und bearbeitet werden.

(3) Grafik-Editor

Mithilfe des Grafik-Editors können Grafiken bearbeitet werden.

(4) Pivot-Tabellen-Editor

Die Zeilen und Spalten einer Tabelle können bspw. verschoben oder gelöscht sowie alle Elemente der Tabelle formatiert werden (z.B. Schriftarten, Farben, Beschriftungen).

(5) Text-Ausgabe

Dieser Editor öffnet sich automatisch sobald ein Text bearbeitet wird.

(6) Syntax-Editor

Zum Bearbeiten und Speichern von Befehlssyntaxen, die dann erneut verwendet werden können.

(7) Skript-Editor

Mittels Python oder Visual-Basic kann SPSS Statistics umfassend an die Anwenderbedürfnisse angepasst werden (z.B. eigene Dialogfelder gestalten, den

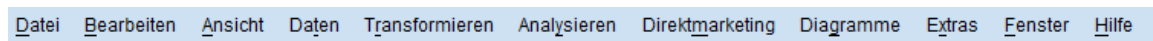
Output von SPSS Statistics automatisch manipulieren, eigene statistische Prozeduren erstellen).

(8) Hilfe-Fenster

Dieser Editor bietet eine Online-Hilfe an.

4. Die Menüstruktur von SPSS Statistics

SPSS Statistics ist menügesteuert. Im Hauptmenü des Daten-Editors stehen elf Untermenüs zur Verfügung:



- (1) **Datei:** Erstellen einer neuen Statistics-Datei, Öffnen einer vorhandenen Datei, Importieren von Daten
- (2) **Bearbeiten:** Befehle wie Ausschneiden, Kopieren oder Einfügen
- (3) **Ansicht:** Verschiedene Optionen zur Darstellung des Datenblattes und zur individuellen Anpassung der Symbolleisten
- (4) **Daten:** Für globale Änderungen an Datendateien, z.B. Verbinden von Dateien, Transponieren der Datenmatrix (= Vertauschen von Variablen und Fällen), Teilmengen von Fällen für bestimmte Analysen erzeugen.
- (5) **Transformieren:** Änderungen an ausgewählten Variablen vornehmen und neue Variablen basierend auf den Werten von vorhandenen Variablen berechnen
- (6) **Analysieren:** Statistische Verfahren auswählen und anwenden, z.B. Kreuztabellen, Varianzanalyse, Korrelation und lineare Regression
- (7) **Direktmarketing:** Speziell auf ein effektives Direktmarketing zugeschnittene Methoden
- (8) **Diagramme:** Verschiedene Diagramme erzeugen (z.B. Balken- und Kreisdiagramme, Histogramme, Streudiagramme und andere Farbgrafi-

- ken). Die Grafiken können mit dem Diagramm-Editor individuell gestaltet werden
- (9) **Extras:** Funktionen wie z.B. Informationen über den Inhalt von Datendateien anzeigen oder ein Skript ausführen
- (10) **Fenster:** Die Anordnung, Auswahl und Eigenschaften der verschiedenen Fenster steuern
- (11) **Hilfe:** Verschiedene Hilfemöglichkeiten

5. Dialogfelder in SPSS Statistics

Die verschiedenen Dialogfelder in SPSS Statistics enthalten im Allgemeinen folgende Komponenten:³

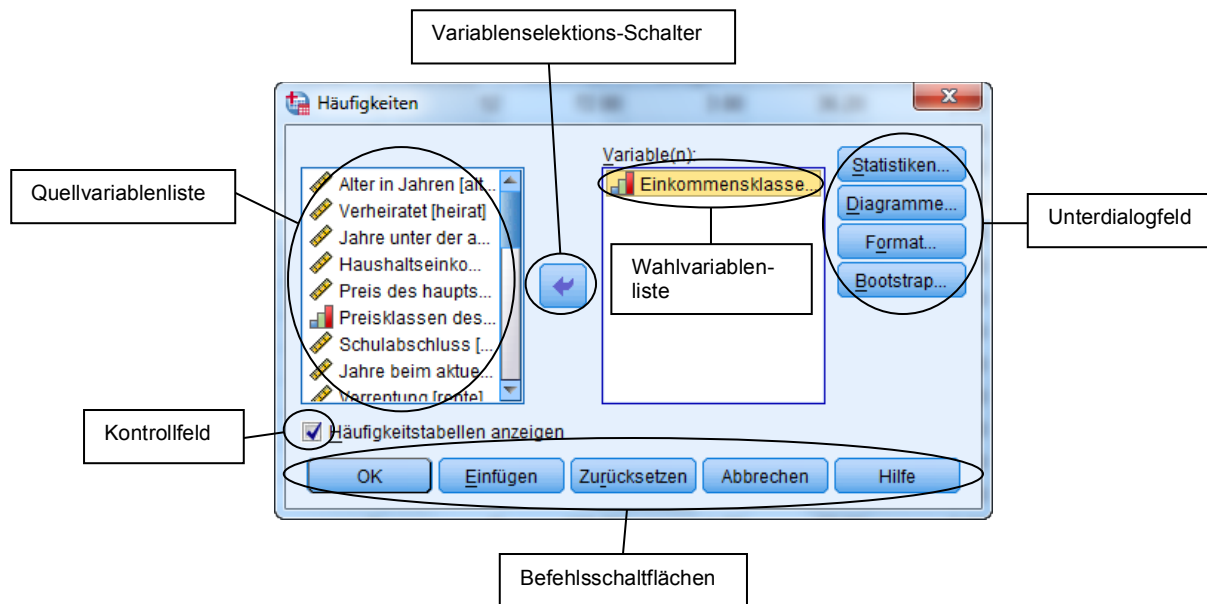


Abbildung 3: Ansicht "Komponenten von Dialogfeldern in SPSS Statistics"

³ Dieses Dialogfeld erhalten Sie bspw. wenn Sie im Menü **Analysieren** auswählen und dann auf **Deskriptive Statistiken** und **Häufigkeiten...** klicken.

Quellvariablenliste: Liste der Variablen in der aktiven Datendatei; Abhängig von der gewählten Prozedur erscheinen alle oder nur bestimmte geeignete Variablen.

Variablen-Selektionsschalter: Zum Verschieben der Variablen, die für die Analyse benötigt werden, in die **Wahlvariablenliste**.

Unterdiaologfelder: Um zusätzliche Spezifikationen zu tätigen; Im Hauptdialogfeld weisen drei Punkte hinter dem Namen einer Schaltfläche darauf hin, dass ein Unterdialogfeld aufgerufen werden kann. SPSS kann mit verschiedenen Modulen lizenziert werden, sodass sich das Erscheinungsbild von Unterdialogboxen ändern kann (z.B. bei Lizenzierung des Moduls "SPSS Bootstrapping" haben einige Dialogboxen für Analysen einen zusätzlichen Unterpunkt).

Über **Kontrollfelder** kann bei vielen Anwendungen die Ausführung des Verfahrens beeinflusst und auch deaktiviert werden; Viele Parameter besitzen eine Voreinstellung (z.B. Häufigkeitstabellen anzeigen).

Befehlsschaltflächen: Aktionsflächen, um Prozeduren auszuführen, Hilfe bereitzustellen oder ein Unterdialogfeld für die Angabe zusätzlicher Parameter zu öffnen. Die fünf Standard-Befehlsschaltflächen 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.
- **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.
- **Hilfe** öffnet ein Hilfefenster, das Informationen über die aktuelle Dialogbox enthält.

6. Das Hilfesystem in SPSS Statistics

In SPSS Statistics stehen folgende drei Hilfesysteme zur Verfügung:

(1) Hilfemenü (Menüzeile)

- **Thematische Hilfe** mit den Registerkarten "Inhalt", "Index" und "Suchen".
- **Lernprogramm** zur anschaulichen, schrittweisen Anleitung für die Verwendung zahlreicher Grundfunktionen von SPSS Statistics. Das modularisierte Lernprogramm ermöglicht die unabhängige und flexible Bearbeitung von einzelnen Themen.
- **Fallstudien**, d.h. praktische Beispiele für die Erstellung verschiedener Arten von statistischen Analysen und für die Interpretation der Ergebnisse.
- **Statistik-Assistent** zur Unterstützung bei der Suche nach einer geeigneten Prozedur und deren Ausführung.
- **Befehlssyntax-Referenz** mit detaillierten Informationen zur Befehlssyntax.

(2) **Kontextsensitive Hilfe** kann an zahlreichen Stellen der Benutzeroberfläche abgerufen werden:

- **Schaltflächen für Hilfe in Dialogfeldern:** Die meisten Dialogfelder verfügen über die Schaltfläche "Hilfe", mit der das entsprechende Hilfethema für das Dialogfeld direkt aufgerufen werden kann (vgl. z.B. Abbildung drei, rechts unten).
- **Hilfe zu Pivot-Tabellen über das Kontextmenü:** Mit der rechten Maustaste auf Begriffe einer im Viewer aktivierten Pivot-Tabelle klicken und dann Direkthilfe aus dem Kontextmenü auswählen, um eine Definition dieser Begriffe zu erhalten.

- **Befehlssyntaxdiagramme:** In einem Befehlssyntaxfenster auf eine beliebige Position innerhalb eines Syntaxblocks für einen Befehl zeigen und auf F1 drücken, um das vollständige Befehlssyntaxdiagramm für diesen Befehl einzublenden.

(3) **Sonstige Ressourcen,** falls die gesuchten Informationen nicht im Hilfesystem zu finden sind:

- **Webseite des technischen Supports:** <http://www-01.ibm.com/software/analytics/spss/support/techsupport.html>
- **Manuals,** die online auf den SPSS-Seiten zur Verfügung stehen (vgl. auch die Unterlagen auf der Lernplattform).

7. Arbeiten mit mehreren Daten-Dateien

In SPSS Statistics können mehrere Datenquellen gleichzeitig geöffnet sein. Das vereinfacht das Wechseln zwischen Datenquellen, das Vergleichen der verschiedenen Datenquellen, das Kopieren und Einfügen von Daten zwischen Datenquellen, das Erstellen von Teilmengen der Fälle und/oder Variablen für die Analyse und das Zusammenführen von verschiedenen Datenquellen mit unterschiedlichen Datenformaten (z.B. Tabellenkalkulationsblätter, Datenbanken, Textdaten), ohne dass zuerst jede Datenquelle im SPSS Statistics-Format gespeichert werden muss.

Die Arbeitsdatei ist links oben im Daten-Editor mit einem roten Kreuz  gekennzeichnet.

8. SPSS Statistics-Sitzung beenden

Zum Beenden von SPSS Statistics wählen Sie aus dem Menü

Datei | Beenden aus.

SPSS Statistics fragt für jedes geöffnete Fenster, ob dessen Inhalt vor dem Verlassen des Programms gespeichert werden soll. Um die Sitzung ohne Speichern zu beenden, auf **Nein** klicken. Um den Inhalt des Fensters unter dem entsprechenden Dateityp abzuspeichern, **Ja** anklicken oder die Eingabetaste drücken.

SPSS Statistics verwendet für die Dateitypen standardmässig folgende Endungen:

- **.sav**: Statistics-Datendatei;
- **.por**: Portable Statistics-Datendatei (Dateien, die von anderen Versionen von SPSS Statistics und Versionen unter anderen Betriebssystemen eingelesen werden können, also bpsw. auf Macintosh).
- **.spv**: Statistics Viewer-Datei;
- **.sps**: Statistics Syntax-Datei.

Dateneingabe in SPSS Statistics

1. Definition von Variablen

Für jede Variable können bestimmte Eigenschaften definiert werden. Betrachten Sie die Variable **Jahre beim aktuellen Arbeitgeber in Klassen [arbei_kl]** in der **Variablenansicht** des Daten-Editors der Datendatei **demo.sav**. Falls Sie sich in der Datenansicht befinden, klicken Sie auf

die Schaltfläche **Variablenansicht** um in die Variablenansicht zu gelangen. Alternativ gelangen Sie auch durch einen Doppelklick auf den Variablenkopf **[arbei_kl]** auf die Ansicht dieser Variable.

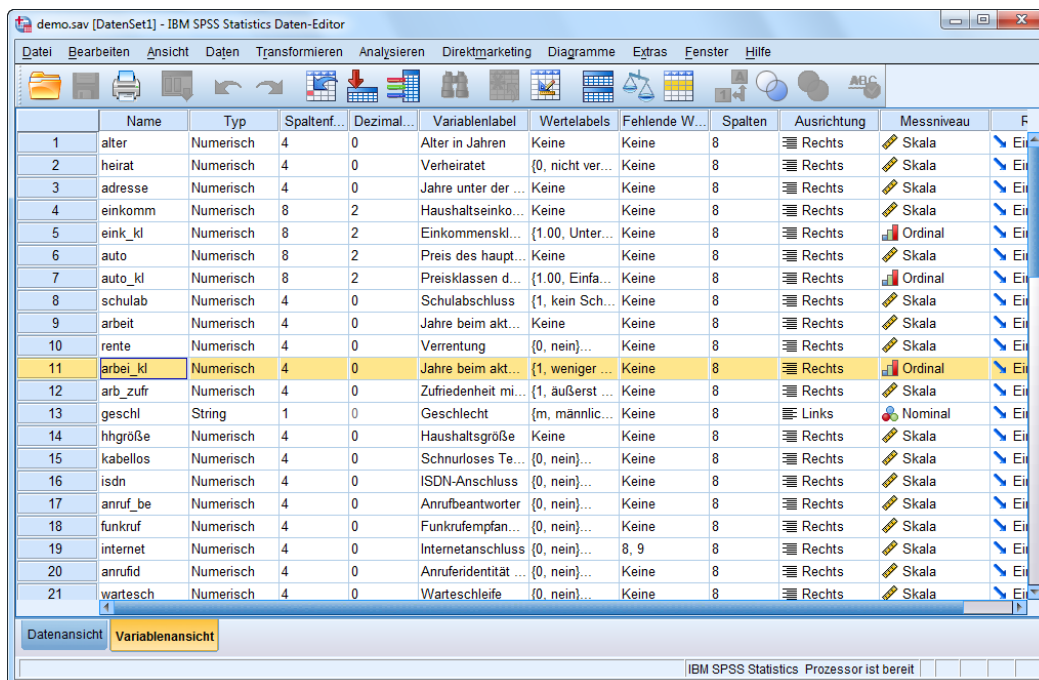


Abbildung 4: Variablenansicht der Daten-Datei demo.sav

Falls Sie selbst eine neue Datendatei erstellen möchten, können Sie Ihre Variablen in der ersten Zeile der **Variablenansicht** im Daten-Editor erfassen. Bei der Vergabe von Variablennamen gelten die folgenden Konventionen:

- Namen müssen mit Buchstaben beginnen. Für die übrigen Zeichen können Buchstaben, Ziffern, Punkte und Symbole @, _ und \$ verwendet werden
- Das letzte Zeichen eines Variablennamens darf kein Punkt sein.
- Namen dürfen max. 64 Zeichen lang sein.
- Leer- und Sonderzeichen, wie bspw. !, ?, ` und *, dürfen nicht verwendet werden.
- Variablennamen müssen eindeutig sein (unabhängig von Gross- und Kleinschreibung).
- Reservierte Schlüsselwörter können nicht als Variablen verwendet werden: ALL, AND, BY, EQ, GE, GT, LE, LT, NE, NOT, OR, TO and WITH.
- Bei älteren SPSS-Versionen sind Umlaute (ä, ö, ü) im Variablennamen nicht zulässig.

In den Spalten der Variablenansicht können folgende **Variableneigenschaften** definiert werden:

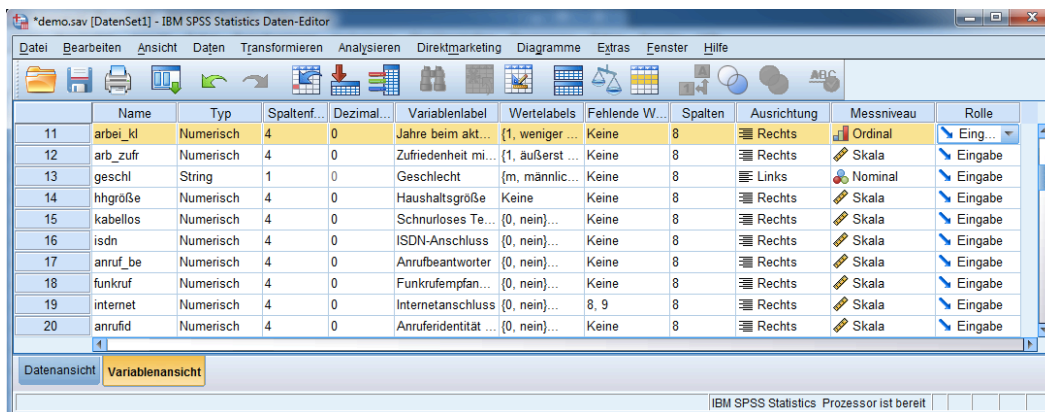


Abbildung 5: Variablenansicht der Variablen arbei_kl

a) Variablentyp

Alle neuen Variablen sind standardmässig als numerisch festgelegt. Es kann jedoch auch ein anderer Datentyp festgelegt werden. Der Inhalt dieses Dialogfeldes ist von dem jeweils ausgewählten Datentyp abhängig. Bei einigen Daten-

typen gibt es Textfelder für die Breite und die Anzahl der Dezimalstellen, bei anderen Datentypen kann einfach ein Format aus einer Liste mit Beispielen ausgewählt werden. Grundsätzlich unterscheidet SPSS Statistics zwischen numerischen Variablen und Stringvariablen. Die Werte numerischer Variablen bestehen nur aus Zahlen (Ziffern 0 bis 9, Dezimalzahlen wie z.B. 0.5). Stringvariablen nehmen beliebige Zeichen auf (Buchstaben A-Z, Ziffern 0-9, Leer- und Sonderzeichen wie z.B. !@%&()).

Wenn Sie auf der Zeile **[arbei_kl]** das Feld **Typ** aktivieren und auf ... klicken, sehen Sie, dass diese Variable als numerisch definiert ist, vier Stellen beansprucht und keine Dezimalstellen besitzt. Die Variable **[arbei_kl]** kann somit als Wert ausschliesslich Zahlen enthalten (und keine Buchstaben).

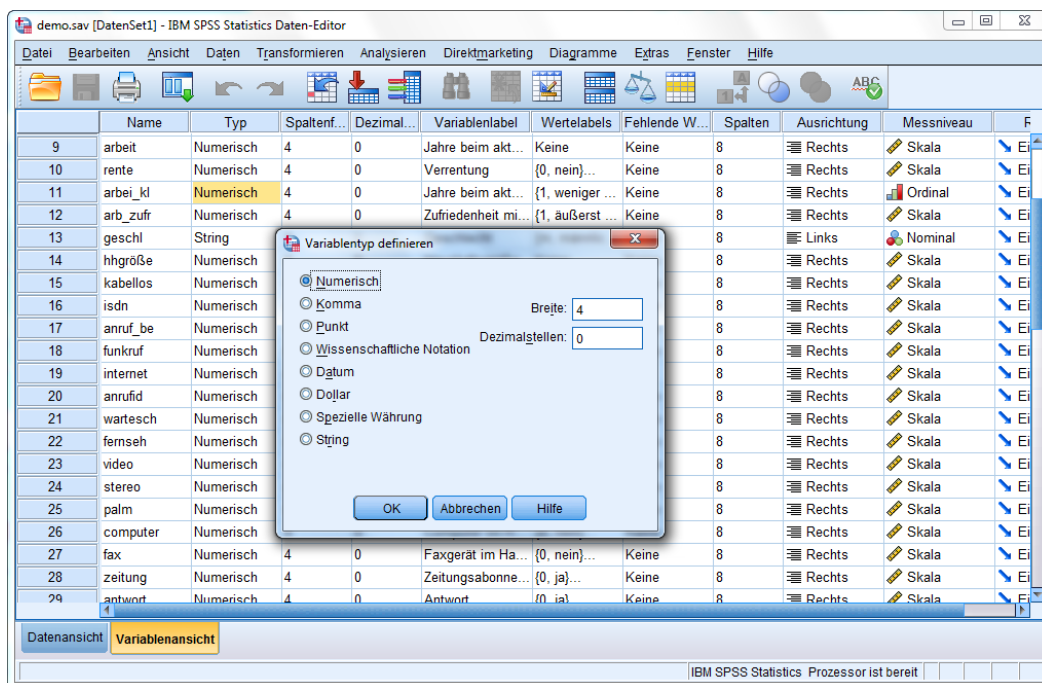


Abbildung 6: Ansicht "Variablentyp definieren"

b) Spaltenformat und Dezimalstellen

Das Spaltenformat (Breite) und die Dezimalstellen können direkt in der **Variablenansicht** innerhalb der Kolonnen **Spaltenformat** und **Dezimalstellen** einge-

stellt bzw. geändert werden. Der Wert für das Spaltenformat muss immer um mindestens 1 höher sein als die Anzahl Dezimalstellen.

c) Variablen- und Wertelabels

SPSS Statistics verwendet Labels als beschreibende Variablennamen. Labels ersetzen in der Ausgabe den Variablennamen bzw. die tatsächlichen Werteausprägungen (z.B. Ziffern) durch ausführliche Bezeichnungen. **Variablen-Labels** können bis zu 255 Zeichen, **Werte-Labels** bis zu 120 Zeichen lang sein.

Wenn Sie beispielsweise auf der Zeile [arbei_kl] das Feld **Wertelabels** aktivieren und auf ... klicken, sehen Sie die definierten Wertelabels und können Ergänzungen bzw. Änderungen vornehmen.

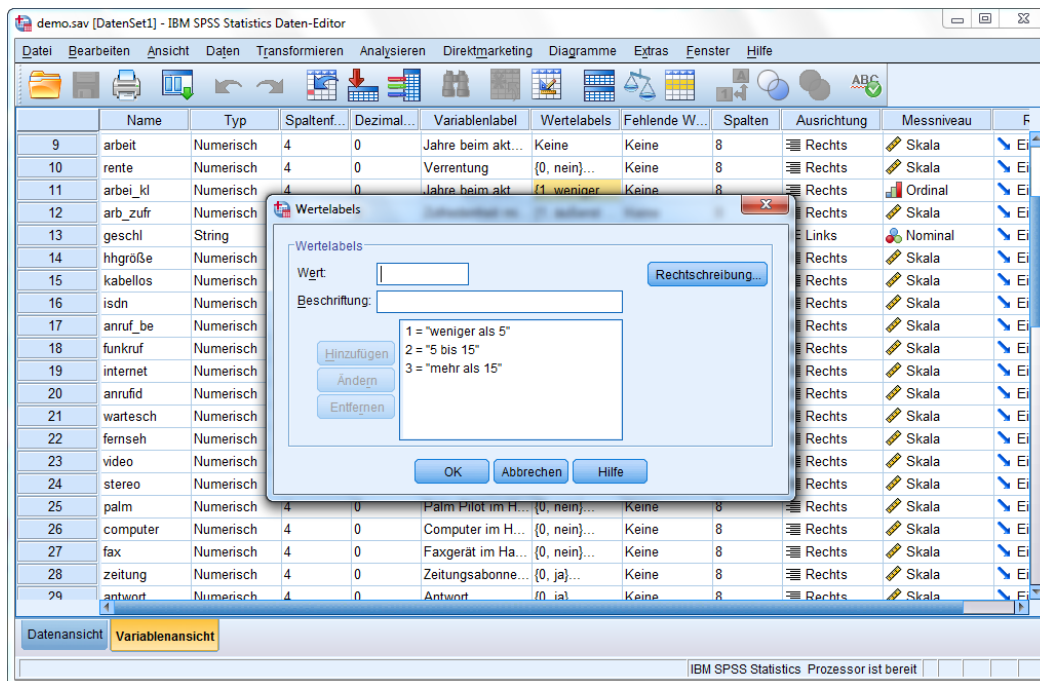


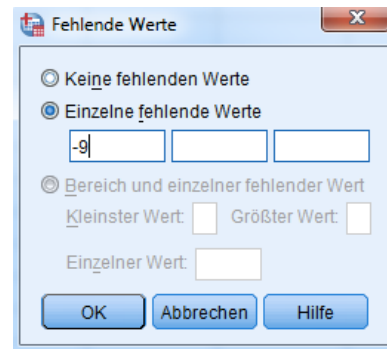
Abbildung 7: Ansicht "Wertelabels erfassen"

d) Fehlende Werte

In SPSS Statistics gibt es zwei Arten von fehlenden Werten:

(1) Allen leeren numerischen Zellen in der Datenmatrix wird ein systemdefinierter fehlender Wert (System-Missings) zugewiesen. Es erscheint ein Punkt im Feld in der Datenansicht. Solche Fälle gehen nicht in Berechnungen und Auswertungen ein.

(2) Um benutzerdefinierte Missing-Werte festzulegen, gehen Sie zur Zeile für **Fehlende Werte** und drücken Sie auf ... Wählen Sie die Option **Einzelne Fehlende Werte** und geben Sie in das erste Textfeld z.B. die Zahl **-9** ein.⁴ Sie können damit Werte ("user-missing values") vorgeben, die signalisieren, dass bei einem Fall keine Angabe über die tatsächliche Ausprägung einer Variablen möglich ist.



Zur Unterscheidung, warum ein Wert fehlend ist, können maximal drei fehlende Werte bei einer Variablen vergeben werden. Der Wert -9 könnte dann bspw. dafür stehen, dass jemand nicht antworten wollte.

Anders als bei den gängigen SPSS Anleitungen üblich sollte der Wert 9 oder 99 zur Kennzeichnung von fehlenden Werten vermieden werden. Dies hängt damit zusammen, dass einige häufig verwendete Skalen (z.B. Net Promoter Score) Skalenpunkte von 1 bis 10 beinhalten bzw. Kundenzufriedenheitsindices typischerweise Werte von 0 bis 100 umfassen. In diesem Fall ist die Zahl 9 bzw. die Zahl 99 ein tatsächlicher Wert. Daher sollte die Zahl -99 zur Kennzeichnung fehlender Werte verwendet werden.

e) Spaltenbreite und Ausrichtung

Zum Einstellen der Spaltenbreite im Daten-Editor oder zur Änderung der Ausrichtung der Daten innerhalb einer Spalte, ändern Sie die Werte der Kolonnen **Spalten** und **Ausrichtung**. Sie ändern die Spaltenbreite, indem Sie einen neuen Wert

⁴ Zur Kennzeichnung des fehlenden Wertes darf keine mögliche Ausprägung der betreffenden Variablen gewählt werden. Als fehlende Werte werden deshalb solche Werte definiert, die außerhalb des Wertebereichs der Variablen liegen (z.B. 9 bei dichotomen Variablen mit 0 und 1 als Ausprägungen).

eingeben. Für die Ausrichtung in der Spalte haben Sie die Optionen *Links*, *Rechts* und *Mitte*.

f) Messniveau

Das Messniveau kann als metrisch, ordinal oder nominal angegeben werden. Nominale und ordinale Daten können entweder aus einem String (alphanumerisch) oder aus Zahlen bestehen:

- **Nominal:** Die Kategorien der Variable lassen sich nicht in eine natürliche Reihenfolge bringen (z.B. Unternehmensabteilung).
- **Ordinal:** Die Kategorien der Variable lassen sich in eine natürliche Reihenfolge bringen (z.B. Zufriedenheit mit Kategorien von sehr unzufrieden bis sehr zufrieden).
- **Metrisch:** Man kann sinnvolle Aussagen über die Abstände zwischen den Werten machen (z.B. Alter in Jahren, Einkommen in Geldeinheiten).

g) Rollen

Manche Dialogfelder unterstützen vordefinierte Rollen, die zur Vorauswahl von Variablen zur Analyse verwendet werden können. Wenn Sie eines dieser Dialogfelder öffnen, werden in der/den Zielliste(n) automatisch Variablen angezeigt, die die Rollenbedingungen erfüllen. Verfügbare Rollen sind:

- **Eingabe:** Die Variable wird als Eingabe verwendet (z.B. Einflussvariable, unabhängige Variable)
- **Ziel:** Die Variable wird als Ausgabe oder Ziel verwendet (z.B. abhängige Variable)
- **Beides:** Die Variable wird sowohl als Eingabe als auch als Ausgabe verwendet
- **Keine:** Der Variablen wird keine Rolle zugewiesen

- **Partitionieren:** Die Variable wird verwendet, um die Daten in separate Stichproben zum Training, zum Test und zur Validierung zu partitionieren
- **Splitten:** Variablen mit dieser Rolle werden in SPSS Statistics nicht als Dateiaufteilungs-Variablen verwendet.

2. Variablendefinitionen kopieren

Sie können dieselbe Information zur Variablendefinition mehreren Variablen zuweisen, indem Sie die **ganze Zeile einer bereits definierten Variablen aktivieren** und diese kopieren mit dem Menü

Bearbeiten | Kopieren (oder Strg + C),

eine neue Zeile aktivieren und dort die Variablendefinitionen mit dem Menü

Bearbeiten | Einfügen (oder Strg + V) einfügen.

Wenn Sie die Variablendefinition erfolgreich eingefügt haben, müssen Sie noch den Variablennamen und die Beschreibung der Variable anpassen. Standardmässig erhält die neue Variable den Namen var00001 (und fortlaufend var00002 etc.). Den Variablennamen können Sie dann manuell verändern/überschreiben.

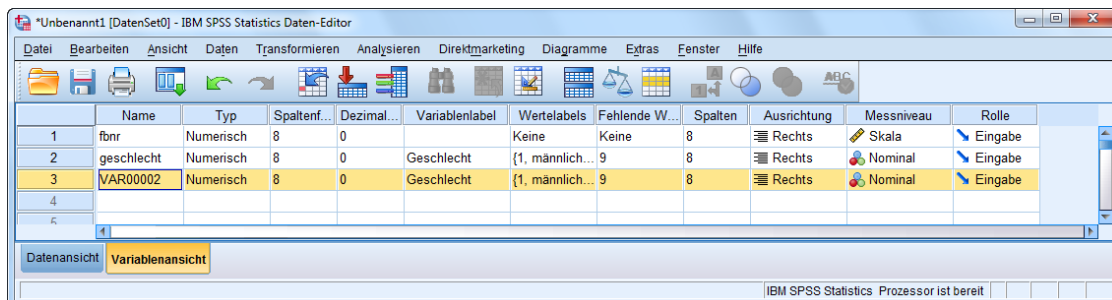


Abbildung 8: Beispiel "Kopieren von Variablendefinitionen"

Wenn Sie die Definition einer bestehenden Variablen **nur teilweise** übernehmen wollen, aktivieren Sie einfach die Zelle der Variablenansicht, die Sie übernehmen wollen (z.B. *Wertelabels* oder *Fehlende Werte*) und gehen Sie für das Kopieren und Einfügen der Definitionen analog vor.

3. Beispiel zur Eingabe von neuen Variablen

Erstellen Sie ein neues Datenblatt, indem Sie selbst die drei nachfolgenden Variablen in der Spalte Variable der **Variablenansicht** erfassen: **alter**, **famstatu**, **einkomme**. Wählen Sie

Datei | Neu | Daten

aus und geben Sie den ersten Variablennamen ein. Nach Drücken der **Return-Taste** wird die Zeile automatisch ausgefüllt. Verfahren Sie ebenso mit den anderen beiden Variablen. Stellen Sie die Dezimalstellen dieser drei Variablen auf Null.

Wechseln Sie dann zur Datenansicht, indem Sie links unten im Daten-Editor auf die Schaltfläche **Datenansicht** klicken. Erfassen Sie für die drei Variablen folgende Werte:

alter	famstatu	einkomme
55	1	72000
43	0	153000

Erfassen Sie eine weitere Variable mit dem Variablennamen **geschl**. Definieren Sie diese Variable als **String-Variable**, indem Sie auf die Pfeile in der Zelle **Typ** klicken, **String** auswählen und auf **OK** klicken. Stellen sie die Dezimalstellen ebenfalls auf Null.

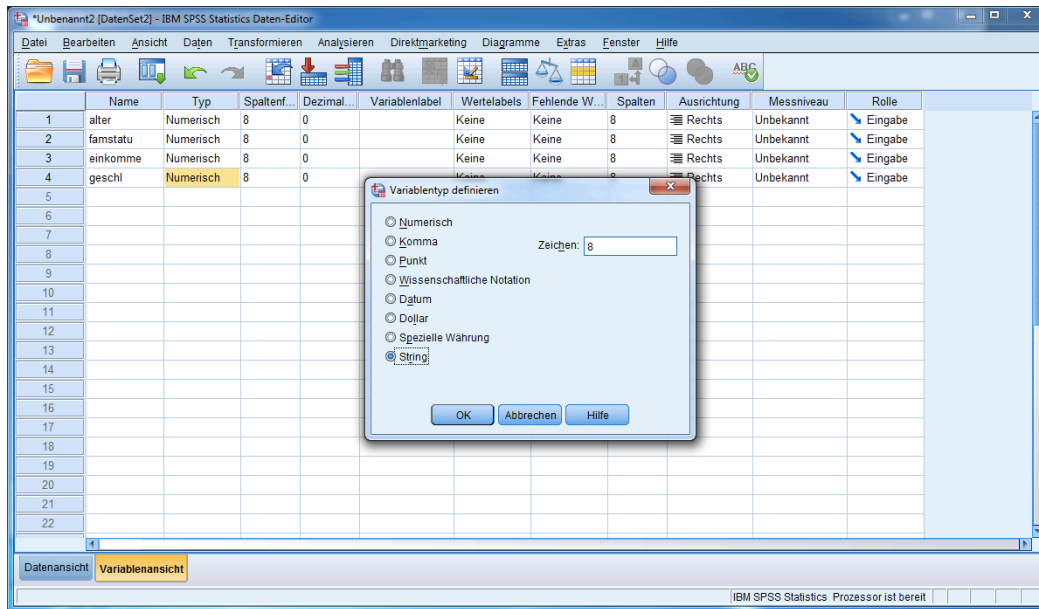


Abbildung 9: Ansicht "Variablentyp definieren"

Erfassen Sie nun folgende **Variablenlabels** in der Variablenansicht, indem Sie die Variablenlabels in der jeweiligen Zelle eingeben:

Variable	Variablenlabel
alter	Alter der befragten Person
famstatu	Familienstand
einkomme	Haushaltseinkommen
geschl	Geschlecht

Definieren Sie **Wertelabels** für die Variablen **famstatu** und **geschl**. Klicken Sie auf die Pfeile in der Zelle Wertelabels der Variable **famstatu** und fügen Sie in der folgenden Ansicht die beiden Wertelabels **0** für **unverheiratet** und **1** für **verheiratet** hinzu. Klicken Sie anschliessend auf **OK** (vgl. Abb. 10). Erfassen Sie für die Variable **geschl** die Wertelabels **f** für **weiblich** und **m** für **männlich**. Beachten Sie, dass bei Stringvariablen zwischen Gross- und Kleinschreibung unterschieden wird (f oder F). Falls Sie bestehende Wertelabels verändern oder löschen wollen, benutzen Sie die Schaltflächen **Ändern** oder **Entfernen**.

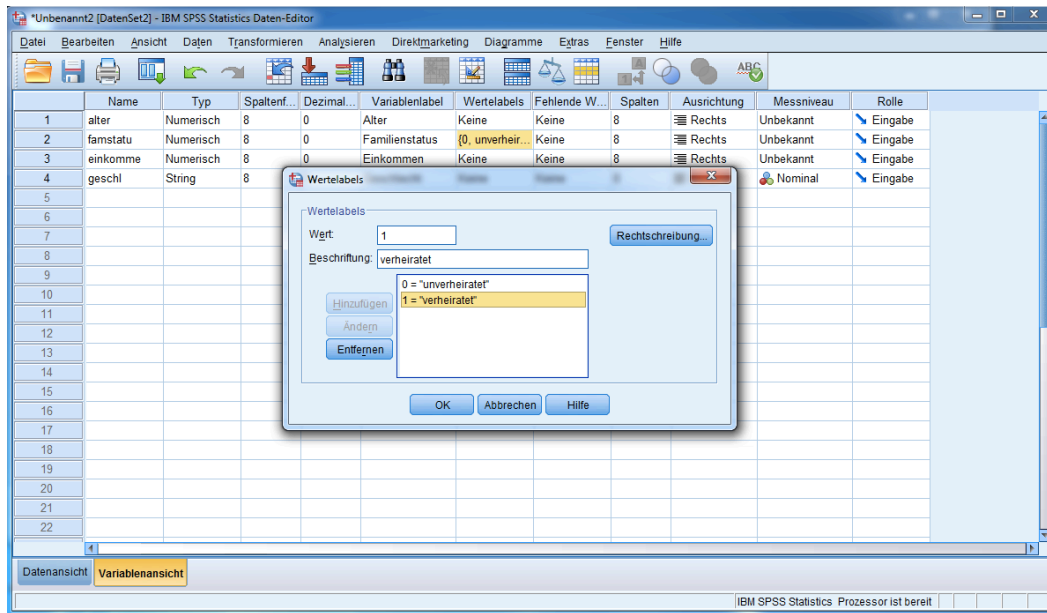


Abbildung 10: Ansicht "Wertelabels"

Wechseln Sie nach der Eingabe zur **Datenansicht** und schalten Sie die Wertela-

bels ein, indem Sie in der Symbolleiste das Symbol für Wertelabels  anklicken. Wählen Sie dann in der ersten Zeile der Datenansicht die **Zelle geschl** aus und klicken Sie zwei Mal auf die Zelle, um mithilfe der **Pfeile rechts** aus der Dropdown-Liste **weiblich** auszuwählen. Wiederholen Sie dasselbe für die **zweite Zeile** und wählen Sie **männlich** aus.

Definieren Sie nun für die numerische Variable **alter** einen fehlenden Wert. Klicken Sie in der Zeile **alter** in der **Variablenansicht** auf die Zelle **Fehlende Werte** und wählen Sie im Fenster **Fehlende Werte** die Option **Einzelne fehlende Werte** aus. Geben Sie in das erste Feld die Zahl **-9** ein und bestätigen Sie die Eingabe mit **OK**. Definieren Sie den Wert **-9** in der Zelle **Wertelabels** der Variable **alter** als **"Keine Angabe"** und bestätigen Sie mit **OK**. Wiederholen Sie die Definition eines fehlenden Wertes für die Stringvariable **geschl**.

4. Erstmaliges Speichern der Daten

Für das erstmalige Speichern der Daten wählen Sie das Menü

Datei | Speichern aus.

Der unten stehende Dialog erscheint. Geben Sie einen beliebigen Namen für die zu speichernde Datei ein. Standardmässig werden die Daten im SPSS Statistics Datendatei-Format abgespeichert (*.sav).

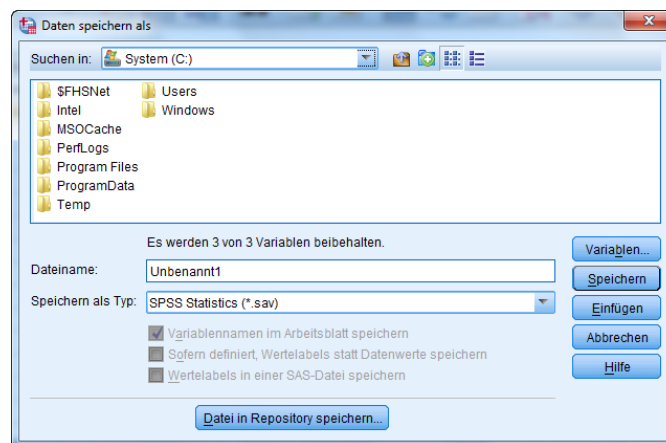


Abbildung 11: Ansicht "Daten speichern als"

Wenn Sie nicht alle Variablen abspeichern wollen, klicken Sie auf die Befehlsschaltfläche **Variablen**. Es erscheint der folgende Dialog:

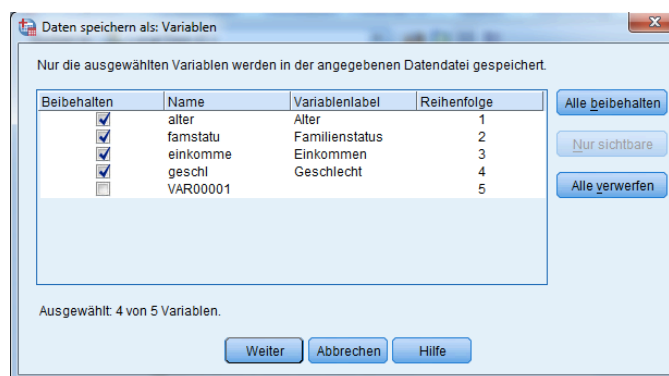


Abbildung 12: Ansicht "Daten speichern als Variablen"

Geben Sie nun an, welche Variablen Sie beibehalten und welche Sie aus der Datendatei ausschliessen wollen. Bestätigen Sie Ihre Eingaben mit **Weiter** und speichern Sie die Datei unter dem gewählten Namen ab.

Ausgabe in SPSS Statistics (SPSS Statistics Viewer)

1. Erzeugen von Ausgaben

Sobald Sie eine Analyse durchführen, wird automatisch ein SPSS Statistics Viewer geöffnet. Dort werden die entsprechenden Tabellen und Grafiken ausgegeben. Zur Berechnung einer Häufigkeitstabelle wählen Sie

Analysieren | Deskriptive Statistiken | Häufigkeiten

aus und übertragen die Variablen **[arbei_kl]** und **[geschl]** ins Fenster **Variablen(n)**. Zusätzlich fordern Sie unter **Diagramme** ein Balkendiagramm an und klicken **OK**.

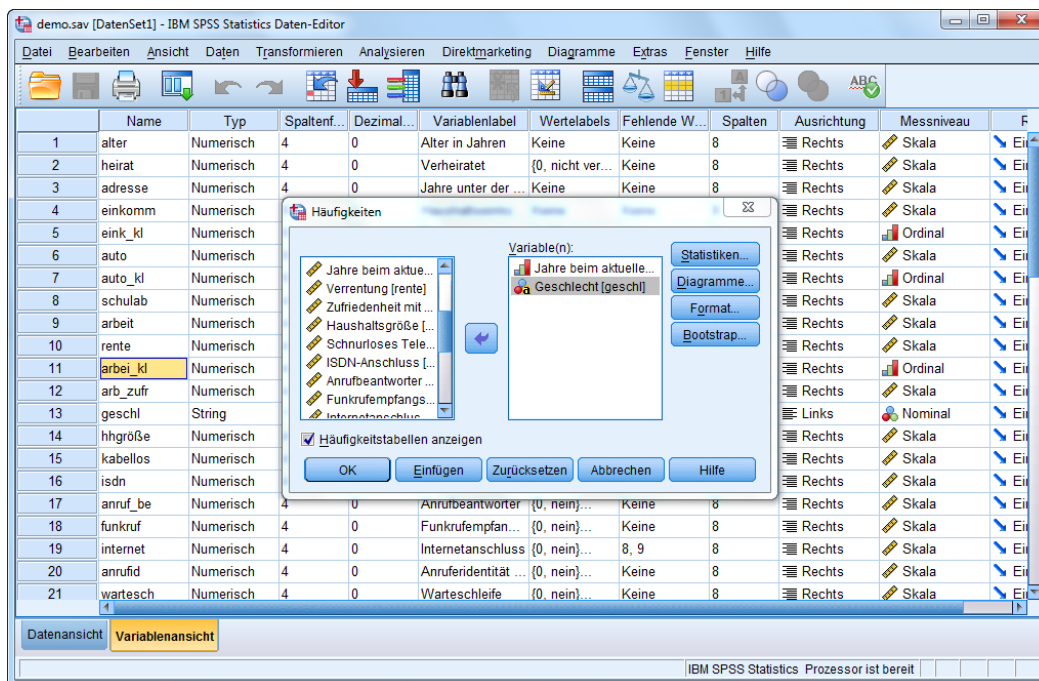


Abbildung 13: Ansicht "Dialogfeld Häufigkeiten"

Der SPSS Statistics Viewer ist in **zwei Fenster** unterteilt:

- Das linke Fenster enthält das Inhaltsverzeichnis des Viewers.

- Das rechte Fenster enthält die statistischen Tabellen, Grafiken und die Textausgabe.

Die jeweils neu erzeugte Ausgabe wird am Ende der letzten Ausgabe angefügt. Wenn Sie sich im linken Fenster durch die Ausgabe bewegen, wird gleichzeitig im rechten Fenster die entsprechende Grafik oder Tabelle mit einem kleinen roten Pfeil markiert und erscheint im Fenster.

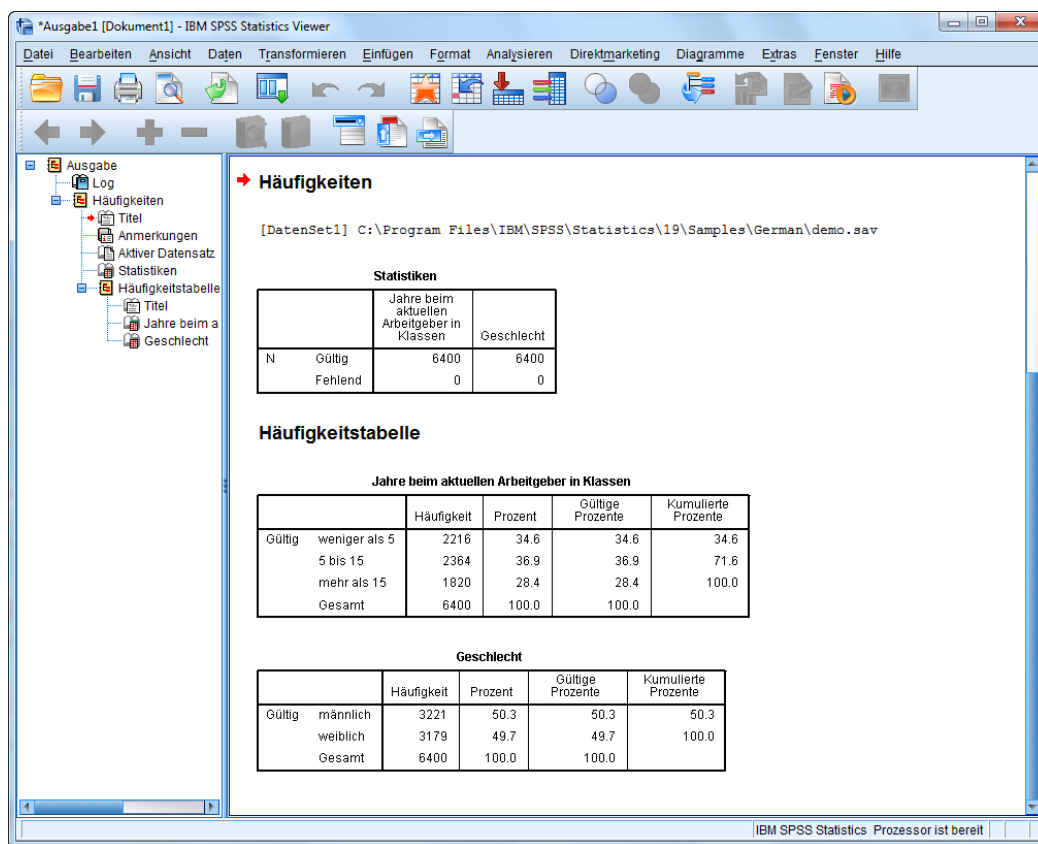


Abbildung 14: Ansicht "Ausgabefenster"

Die Breite der beiden Fenster kann durch Ziehen mit der Maus am Trennungsbalken verändert werden.

2. Anzeigen und Verbergen von Ausgaben

Da die Inhaltsübersicht (siehe linkes Fenster des SPSS Statistics Viewers) sehr umfangreich werden kann, kann es sinnvoll sein, Teile der Ausgabe zu verbergen. Dies verkürzt die Ausgabe und beinhaltet (im Gegensatz zum Löschen) die Möglichkeit, die Ausgabe jederzeit wieder sichtbar zu machen. Um Titel, Anmerkungen, Tabellen oder Grafiken zu verbergen, kann entweder auf das Symbol  doppelgeklickt oder in der Symbolleiste das Minussymbol  benutzt oder – nach Auswahl einzelner Ausgabeobjekte – das Menü

Ansicht | Reduzieren (bzw. erweitern)

verwendet werden. Ebenso kann durch Klicken auf das Pluszeichen ein zusammengefasster Block wieder geöffnet werden.

Zum Sehen bzw. Verstecken von **einzelnen Teilen des Ausgabeblocks** stehen neben dem Doppelklick auf die einzelnen Ausgabeelemente in der Symbolleiste am oberen Rand des Bildschirms die beiden Symbole für das Einblenden bzw.

zum Ausblenden zur Verfügung. 

Eine wichtige Ergänzung der Ausgabe sind die **Anmerkungen** (vgl. Inhaltsverzeichnis, linkes Viewer-Fenster). Sie enthalten Informationen über die Analyse der verwendeten Datei und die getroffenen Einstellungen. Ausserdem wird dort der Syntaxbefehl vermerkt. Gemäss Voreinstellung ist dieses Objekt zunächst ausgeblendet. Zum Ansehen markieren Sie Anmerkungen im linken View Fenster und klicken in der Symbolleiste am oberen Rand des Bildschirms auf das

Symbol  oder wählen über das Menü

Ansicht | Einblenden aus.

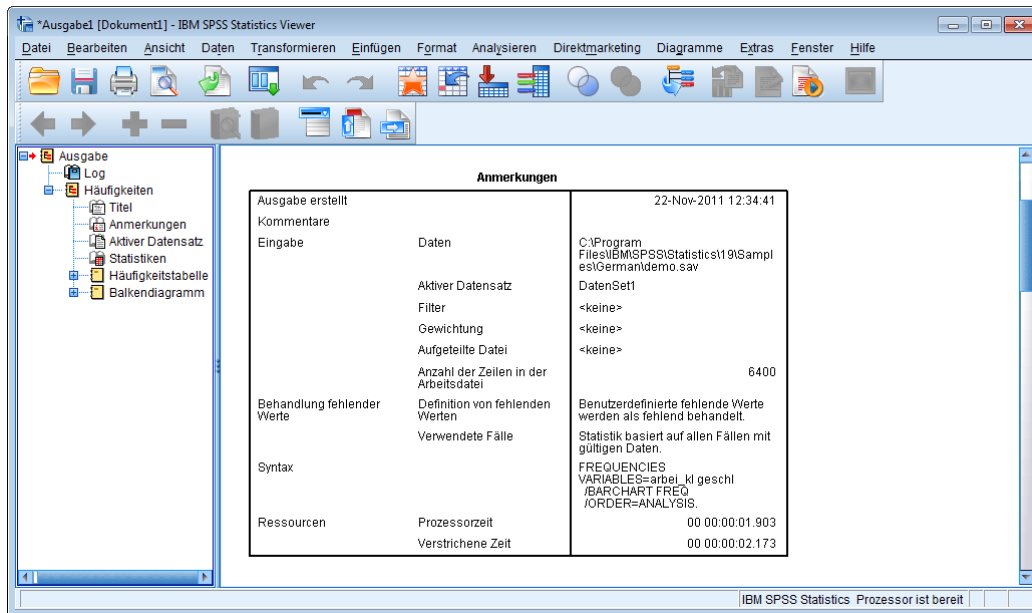
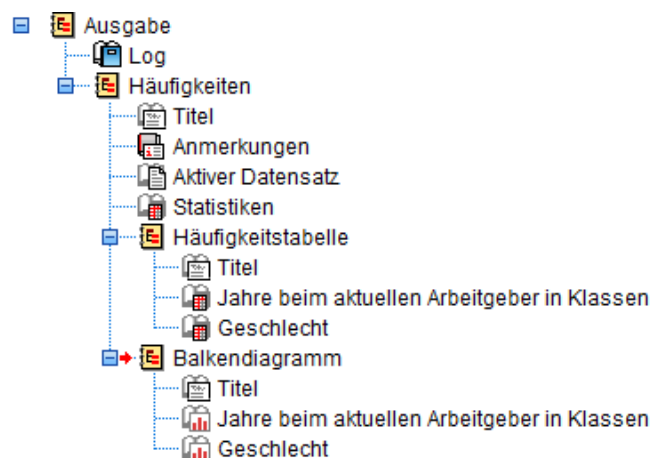


Abbildung 15: Ansicht "Einblenden Anmerkungen im Viewer-Fenster"

3. Umstellen und Kopieren von Ausgaben

Mit dem Navigator ist es möglich, die Anordnung der Ausgabe zu ändern. Man kann Teile kopieren, verschieben oder löschen. Um die Anordnung zu verändern, klicken Sie das entsprechende Objekt in der Inhaltsübersicht an (z.B. Häufigkeitstabelle).

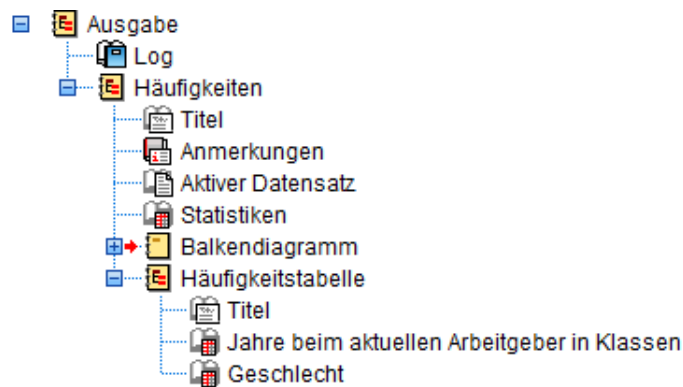


Um Output zu kopieren, klicken Sie bpsw. auf das gewünschte Objekt und wählen aus dem Menü

Bearbeiten | Kopieren (oder Strg + C) aus.

Bewegen Sie die Maus an die gewünschte Stelle im Inhaltsverzeichnis und klicken Sie auf

Bearbeiten | Einfügen (oder Strg + V).



4. Editieren von Tabellen und Grafiken

Sämtliche Elemente des SPSS Statistics-Viewers können im rechten Fenster weiter verändert und editiert werden. Durch Doppelklick auf das Element, das verändert werden soll, gelangen Sie in den entsprechenden Editor. Doppelklicken Sie z.B. auf die Häufigkeitstabelle **Jahre beim aktuellen Arbeitgeber in Klassen**, um in den Tabellen-Editor zu gelangen. Falls die Formatierungs-Symbolleiste nicht offen ist, können Sie diese unter

Ansicht | Symbolleiste

anzeigen lassen.

Formatierungs-Symbolleiste

SansSerif 9 **B** *I* U *A*

Jahre beim aktuellen Arbeitgeber in Klassen

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
→ Gültig	weniger als 5	2216	34.6	34.6	34.6
	5 bis 15	2364	36.9	36.9	71.6
	mehr als 15	1820	28.4	28.4	100.0
	Gesamt	6400	100.0	100.0	

Um nun z.B. einzelne Werte hervorzuheben, klicken Sie auf die entsprechende Zelle und wählen Fettdruck oder eine andere Farbe, um die Formatierung dieser Zelle zu verändern.

5. Speichern der Ausgabe

Um Ausgaben unter dem voreingestellten Namen zu speichern, wählen Sie im Menü

Datei | Speichern aus oder



Sie benutzen das Symbol . Um die Ausgabe unter einem anderen Namen zu speichern, können Sie auch

Datei | Speichern unter

wählen.

Datenübernahme: Dateien einlesen

Daten können entweder direkt eingegeben werden oder aus einer Reihe unterschiedlicher Quellen in SPSS Statistics importiert werden:

(a) Einlesen von SPSS Statistics-Datendateien

Um die SPSS Statistics-Datei **demo.sav** in SPSS Statistics einzulesen, wählen Sie aus dem Menü

Datei | Öffnen aus.

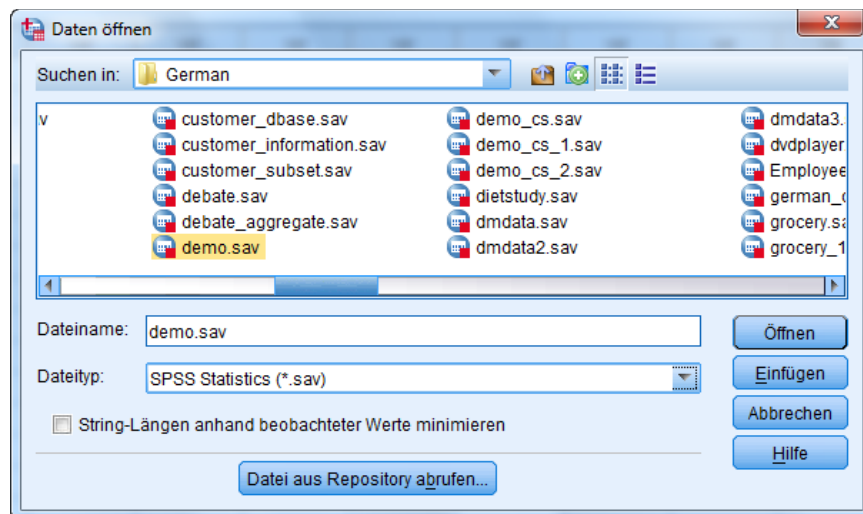


Abbildung 16: Ansicht "Daten öffnen" am Beispiel einer Daten-Datei

Im Dialogfeld **Daten öffnen** wählen Sie den Dateityp SPSS Statistics (*.sav). Dadurch werden die Dateien mit dieser Erweiterung aufgelistet.

(b) Einlesen von Excel-Dateien

Um die SPSS Statistics-Datei **demo.xls** in SPSS Statistics einzulesen, wählen Sie aus dem Menü

Datei | Öffnen aus.

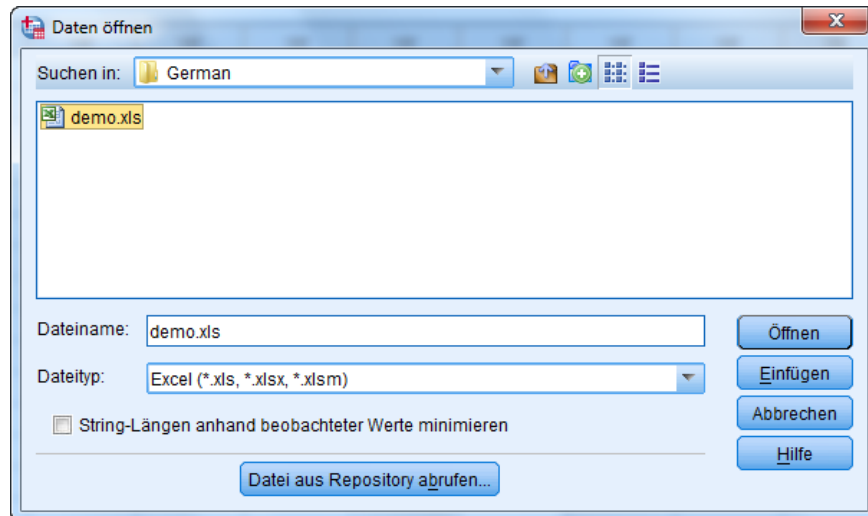
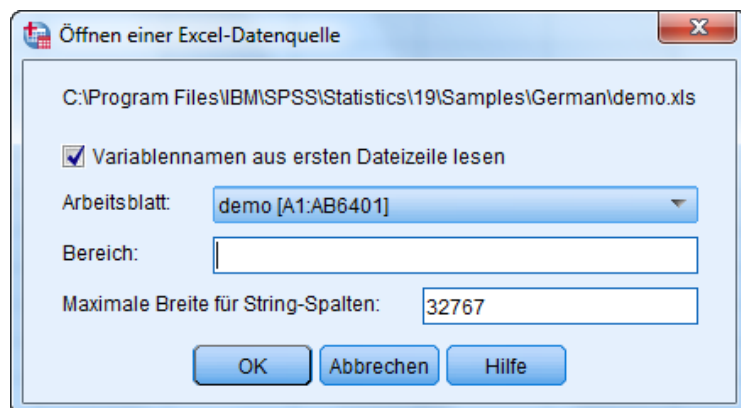


Abbildung 17: Ansicht "Daten öffnen" am Beispiel einer Excel-Datei

Im Dialogfeld **Daten öffnen** wählen Sie den Dateityp Excel aus (*.xls, *.xlsx, *.xlsm). Dadurch werden die Dateien mit dieser Erweiterung aufgelistet.

Markieren Sie dann im nachfolgenden Dialogfenster **Variablenamen aus ersten Datenzeile lesen**. Das bedeutet, SPSS Statistics soll die Variablenamen aus der ersten Zeile des Arbeitsblattes übernehmen. Unter Arbeitsblatt kann das entsprechende Arbeitsblatt ausgewählt werden. In eckigen Klammern wird der Bereich angezeigt, wo SPSS Statistics Daten gefunden hat. Falls der vorgeschlagene Bereich vom tatsächlichen Datenbereich abweicht, können Sie unter **Bereich** den Bereich angeben, in dem sich die Daten tatsächlich befinden. Mit **OK** werden die Daten in den Daten-Editor eingelesen.



IBM SPSS Statistics Daten-Editor

Menü: Datei, Bearbeiten, Ansicht, Daten, Transformieren, Analysieren, Direktmarketing, Diagramme, Extras, Fenster, Hilfe

Sichtbar: 28 von 28 Variablen

	Age	Maritalstatus	Address	Income	IncomeCategory	Carprice	Carpricecategory	Education	Emph
1	55	1	12	72.00	3.00	37.00	3.00	1	
2	56	0	29	153.00	4.00	76.00	3.00	1	
3	28	1	9	28.00	2.00	13.90	1.00	3	
4	24	1	4	26.00	2.00	13.00	1.00	4	
5	25	1	2	23.00	1.00	11.30	1.00	2	
6	45	0	9	76.00	4.00	37.30	3.00	3	
7	44	1	17	144.00	4.00	72.10	3.00	2	
8	46	1	20	75.00	4.00	37.10	3.00	1	
9	41	0	10	26.00	2.00	13.00	1.00	1	
10	29	0	4	19.00	1.00	9.60	1.00	2	
11	34	0	0	89.00	4.00	44.40	3.00	3	
12	55	0	17	72.00	3.00	36.10	3.00	3	
13	28	0	9	55.00	3.00	28.20	2.00	4	
14	21	1	2	20.00	1.00	9.60	1.00	3	
15	55	0	8	283.00	4.00	77.40	3.00	3	
16	35	0	8	70.00	3.00	35.40	3.00	3	
17	45	0	4	48.00	2.00	25.00	2.00	2	
18	21	0	1	37.00	2.00	18.40	2.00	3	
19	32	0	0	28.00	2.00	14.20	1.00	1	
20	42	0	9	109.00	4.00	54.00	3.00	3	
21	40	1	12	117.00	4.00	59.80	3.00	2	
22	36	0	6	39.00	2.00	19.20	2.00	2	

Datenansicht Variablenansicht

IBM SPSS Statistics Prozessor ist bereit

Abbildung 18: Ansicht "Eingelesene Excel-Datei demo.sav"

Dateien zusammenfügen

Mit SPSS Statistics können Sie zwei Dateien auf unterschiedliche Weise zusammenfügen. Welche Methode Sie verwenden, hängt von den gegebenen Daten ab:

- Wenn die Dateien dieselben Variablen, aber unterschiedliche Fälle enthalten, dann fügen Sie Fälle hinzu.
- Wenn die Dateien dieselben Fälle, aber unterschiedliche Variablen enthalten, dann fügen Sie Variablen hinzu.

(1) Dateien mit unterschiedlichen Fällen zusammenführen

Fügen Sie die Dateien der SPSS-Kundenbefragung von 2008 und 2009 zusammen, indem Sie zunächst die Datei SPSS Kundenbefragung 2008.sav in den Daten-Editor laden und wählen Sie aus dem Menü

Daten | Dateien zusammenfügen | Fälle hinzufügen.

Im Dialogfeld **Fälle hinzufügen** wählen Sie als externe SPSS Statistics-Datendatei **SPSS Kundenbefragung 2009.sav** und klicken Sie auf **Weiter**.

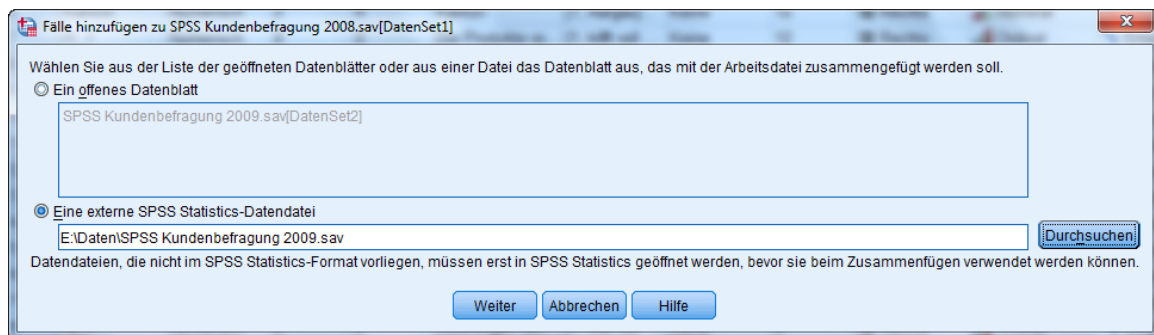


Abbildung 19: Ansicht "Fälle hinzufügen (1)"

Alle Variablen, die in beiden Dateien vorkommen, erscheinen im Feld **Variablen in der neuen Arbeitsdatei**. Variablen, die nur in einer der beiden Dateien vorkommen, erscheinen im Feld **Nicht gepaarte Variablen**.

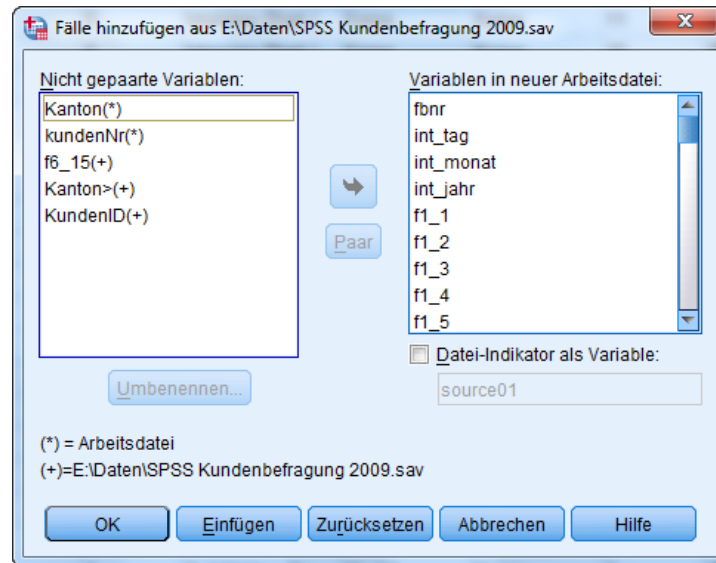


Abbildung 20: Ansicht "Fälle hinzufügen (2)"

In diesem Beispiel sind einige Variablen (kundenNr und KundenID bzw. Kanton und kanton_cod) in den beiden Dateien unter verschiedenen Variablennamen abgespeichert, obwohl sie dieselben Informationen enthalten. Diese erscheinen deshalb unter **Nicht gepaarte Variablen**.

Um aus nicht gepaarten Variablen Variablenpaare zu bilden, markieren Sie **kundenNr**, anschliessend mit gedrückter STRG-Taste **KundenID**. Klicken Sie nun auf **Paar**, um dieses Variablenpaar in die Listen **Variablen in neuer Arbeitsdatei** zu verschieben (Der Variablenname aus der Arbeitsdatei wird in der zusammengeführten Datei als Variablenname verwendet).

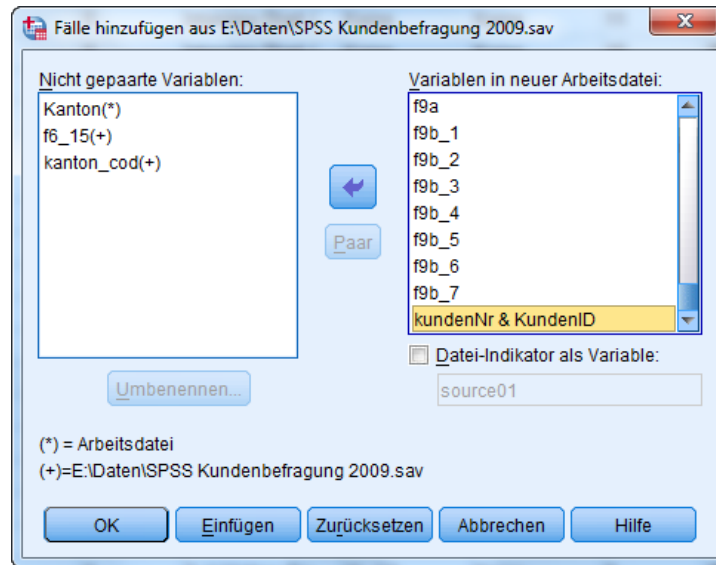


Abbildung 21: Ansicht "Fälle hinzufügen (3)"

Wiederholen Sie diesen Schritt für Kanton (*) und kanton_cod (+). Um auch die Variablen f6_15 in die neue Datei aufzunehmen, markieren Sie diese und klicken

anschliessend auf .

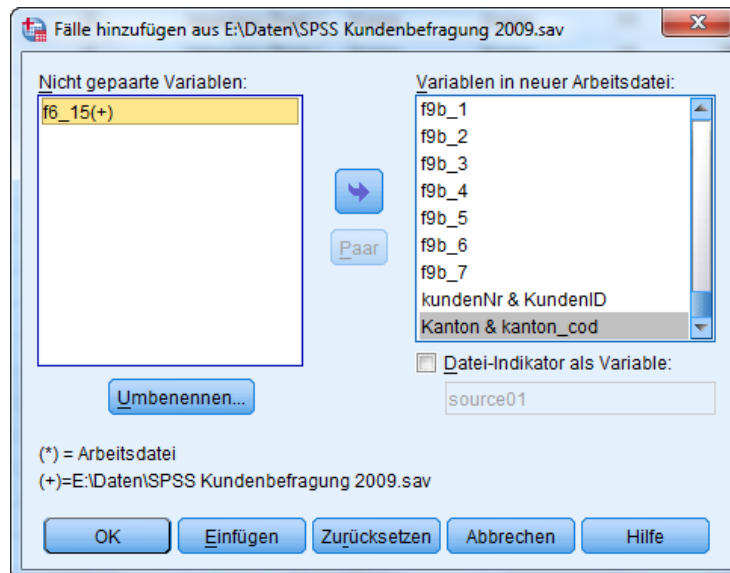


Abbildung 22: Ansicht "Fälle hinzufügen (4)"

Mit der Option **Datei-Indikator als Variable** kann die neue Datendatei um eine Variable mit dem Namen **source01** ergänzt werden, die den Ursprung der Daten angibt. Setzen Sie hier ein **Häkchen**.

Die **Variable f6_15** finden Sie ganz am Ende des neuen Datenfiles. Um diese nach f6_14 zu verschieben, wechseln Sie in die Variablenansicht. Markieren Sie dort die Zeile von f6_15 und verschieben Sie diese Variable mit gedrückter linker Maustaste nach oben, bis die rote Markierung unter f6_14 liegt; lassen Sie die linke Maus dann wieder los. Speichern Sie das File unter **SPSS Kundenbefragung.sav** ab.

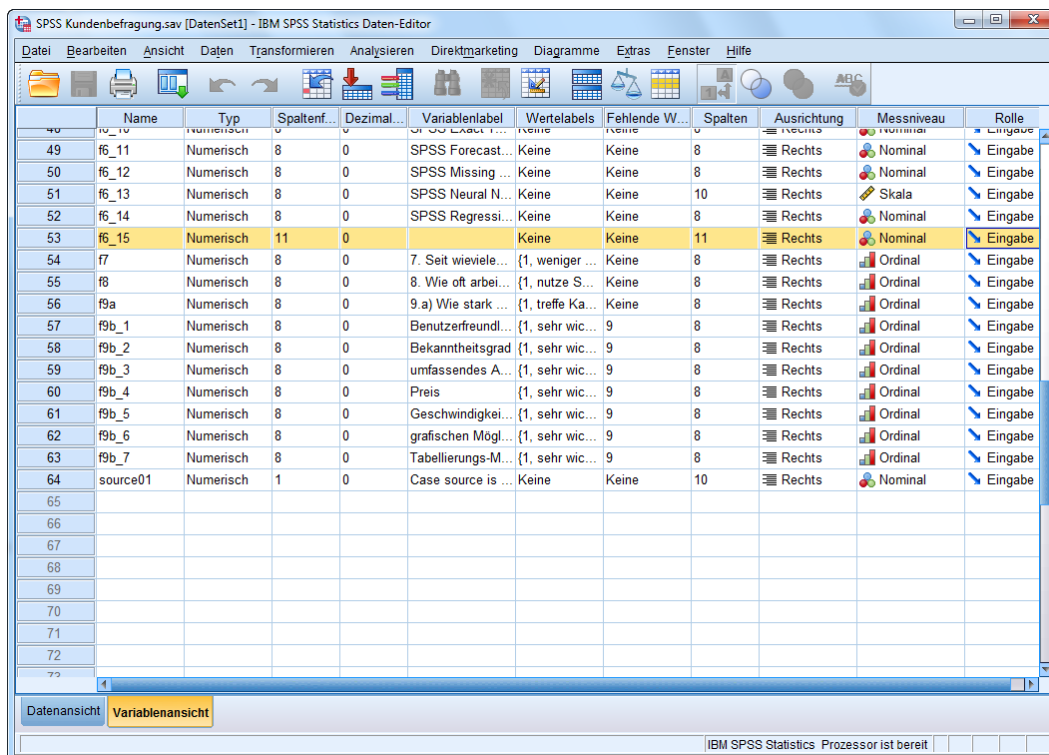


Abbildung 23: Ansicht "SPSS Kundenbefragung.sav"

(2) Dateien mit unterschiedlichen Variablen zusammenführen

Das obige File soll nun zusätzlich durch weitere Kunden-Details aus der Datei **SPSS Kunden.sav** ergänzt werden. Statt neuer Fälle sollen nun somit zusätzliche Variablen hinzugefügt werden.

Es kann vorkommen, dass die beiden Dateien, die zusammengefügt werden sollen, eine unterschiedliche Struktur haben. Eine Datei enthält Fälle, die durch eine Schlüsselvariable gruppiert sind, d.h. die Werte der Schlüsselvariable können mehrfach vorkommen. Die andere Datei ist eine sogenannte Schlüsseltabelle, d.h. sie enthält zu jedem Wert der Schlüsselvariablen nur einen Eintrag. In der Schlüsseltabelle können auch Werte enthalten sein, die in der anderen Datei nicht vorkommen.

In diesem Beispiel sollen die Dateien über die **Kunden-Nummer** miteinander verbunden werden. Während die Kunden-Nummer (kundenNr) in der Datei SPSS Kundenbefragung.sav z.T. mehrfach vorhanden ist, finden Sie in der Datei SPSS Kunden.sav für jede Kunden-Nummer (id) nur 1 Eintrag. Somit ist SPSS Kunden.sav die Schlüsseltabelle, deren Informationen an SPSS Kundenbefragung.sav angehängt werden sollen.

Beim Zusammenführen von Dateien mit unterschiedlichen Variablen ist Folgendes zu beachten:

- Die Identifizierung und richtige Zuordnung der Fälle in den beiden Daten-Sets erfolgt über eine Schlüsselvariable, welche in beiden Dateien vorhanden sein sollte.
- Beide Daten-Sets müssen in aufsteigender Reihenfolge der Schlüsselvariablen sortiert sein.

Öffnen Sie deshalb als Erstes die Datei **SPSS Kundenbefragung.sav** und klicken Sie mit der rechten Maustaste auf den Spaltenkopf von **kundenNr**, um diese aufsteigend zu sortieren. Öffnen Sie anschliessend **SPSS Kunden.sav** und sortieren Sie die **Variable id** auf dieselbe Art und Weise. Um die beiden Dateien zusammenzuführen, gehen Sie zurück zu SPSS Kundenbefragung.sav und wählen dort

Daten | Dateien zusammenfügen | Variablen hinzufügen.

Markieren Sie **SPSS Kunden.sav** in der Liste der geöffneten Daten-Sets und klicken auf **Weiter**.

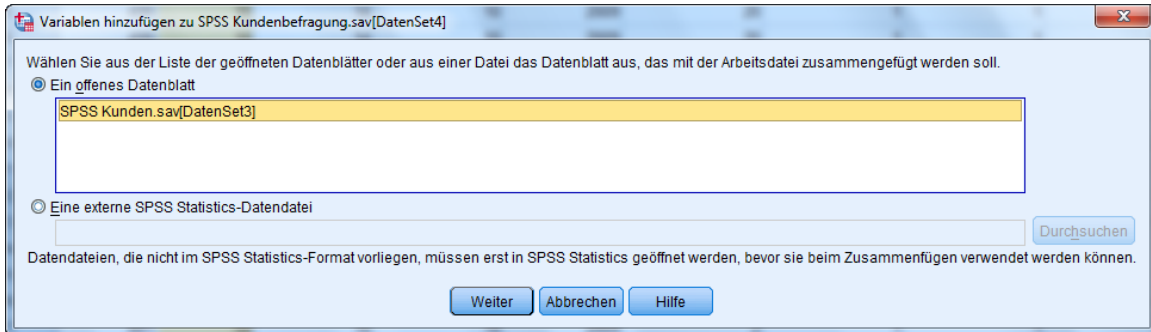


Abbildung 24: Ansicht "Variablen hinzufügen (1)"

Da die Schlüsselvariable (Kunden-Nummer) in den beiden Dateien unter einem unterschiedlichen Namen abgespeichert ist, muss eine der beiden Variablen vor­gängig umbenannt werden, um sie als Schlüsselvariable verwenden zu können. Suchen Sie im Bereich **Neue Arbeitsdatei** nach der **Variable id** und verschieben Sie diese in den Bereich **Ausgeschlossene Variablen**.

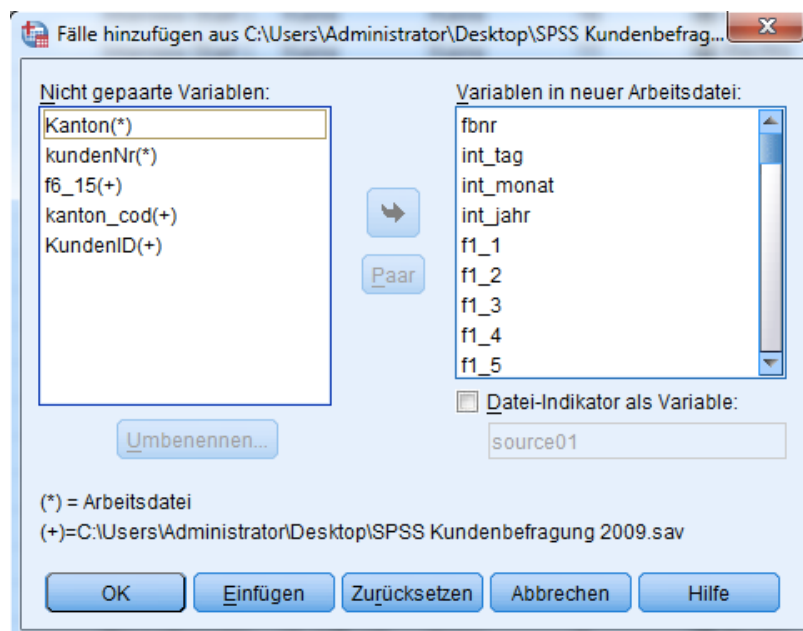


Abbildung 25: Ansicht "Variablen hinzufügen (2)"

Klicken Sie nun auf **Umbenennen** und geben dort als neuen Namen **kundenNr** ein (wichtig ist, dass diese Variable denselben Variablennamen erhält wie die entsprechende Variable in der anderen Datei). Klicken Sie auf **Weiter**.

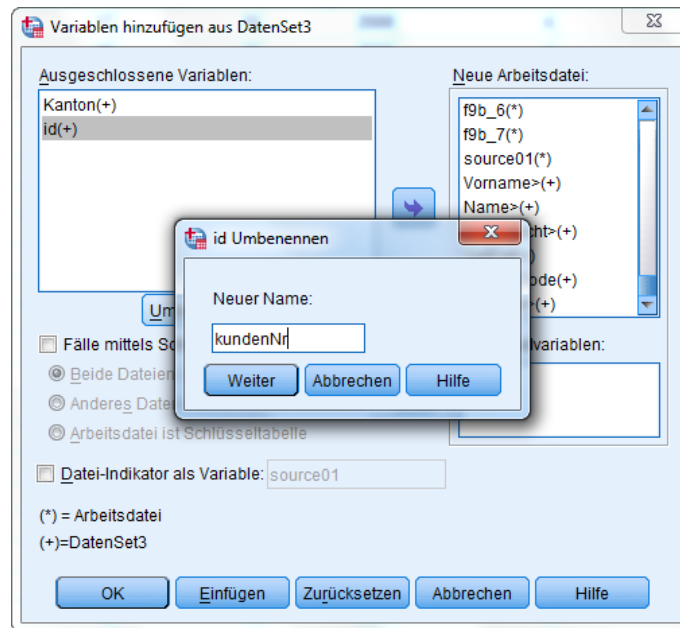


Abbildung 26: Ansicht "Variablen hinzufügen (3)"

Aktivieren Sie nun das Kästchen **Fälle mittels Schlüsselvariablen verbinden** und wählen Sie die Option **Anderes Datenblatt ist Schlüsseltabelle**. Verschieben Sie die Variable **kundenNr** in den Bereich **Schlüsselvariablen**.

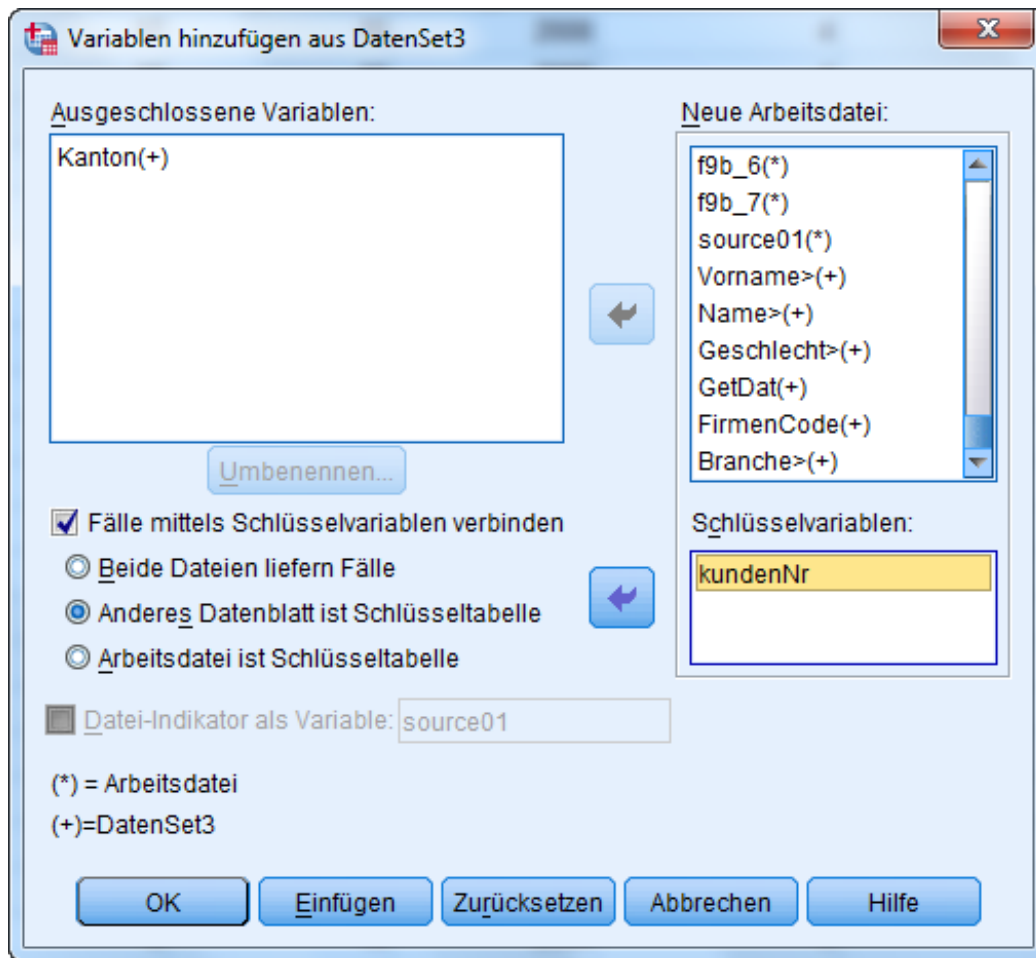
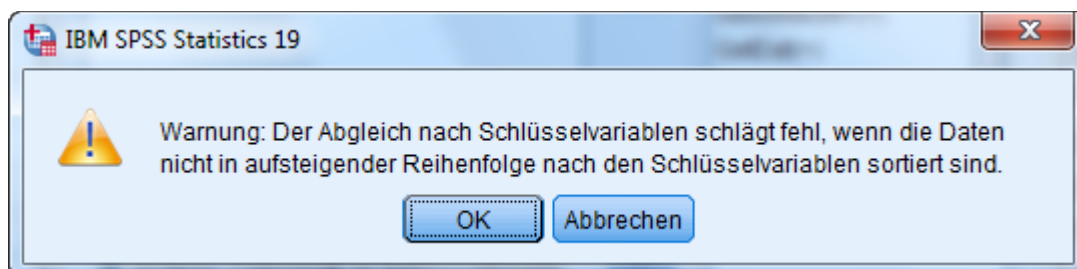


Abbildung 27: Ansicht "Variablen hinzufügen (4)"

Klicken Sie auf **OK**. Es erscheint eine Warnung, die darauf hinweist, dass für ein erfolgreiches Zusammenfügen beide Dateien vorgängig nach der Schlüsselvariable sortiert sein müssen.



Dies ist bei den verwendeten Dateien der Fall. Sie können die Warnung mit OK bestätigen. Die Dateien werden nun über die Schlüsselvariable zusammengefügt. Speichern Sie das File unter **SPSS Kundenbefragung.sav** ab.

SPSS Kundenbefragung.sav [DatenSet1] - IBM SPSS Statistics Daten-Editor																											
Datei Bearbeiten Ansicht Daten Transformieren Analysieren Direktmarketing Diagramme Extras Fenster Hilfe																											
5 : Branche CON_G																											
Sichtbar: 70 von 70 Variablen																											
	f6_8	f6_9	f6_10	f6_11	f6_12	f6_13	f6_14	f6_15	f7	f8	f9a	f9b_1	f9b_2	f9b_3	f9b_4	f9b_5	f9b_6	f9b_7	source01	Vorname	Name	Geschlecht	GetDat	FirmenCo...	Branche		
1	0	1	0	0	1	0	0	1	.	3	4	1	2	2	3	1	1	2	1	0	Gina	Mondher	weiblich	18-Nov-196...	76	CON_M	
2	0	1	1	0	0	1	0	0	0	4	6	1	1	3	1	2	1	2	1	1	Gina	Mondher	weiblich	18-Nov-196...	76	CON_M	
3	0	0	0	0	0	0	0	1	.	3	1	2	9	9	3	2	3	9	4	0	Hanspeter	Masshard	männlich	30-Nov-198...	63	SOZ_M	
4	1	0	0	0	0	1	0	0	0	4	5	2	2	2	2	1	2	2	2	1	Hanspeter	Masshard	männlich	30-Nov-198...	63	SOZ_M	
5	0	0	0	0	1	0	0	1	.	3	6	2	9	2	3	3	9	9	2	0	Jolanda	Rotz	weiblich	5-May-198...	38	CON_G	
6	1	0	0	1	0	0	0	0	0	1	3	1	1	2	3	2	2	4	4	1	Jolanda	Rotz	weiblich	5-May-198...	38	CON_G	
7	0	1	0	0	0	0	0	0	1	.	2	4	1	2	2	9	3	4	2	3	0	Sandra	Vanzetti	weiblich	22-Sep-19...	162	SOZ_G
8	0	0	0	0	0	0	0	0	0	2	4	1	1	4	2	3	3	3	2	1	Sandra	Vanzetti	weiblich	22-Sep-19...	162	SOZ_G	
9	0	1	0	0	1	0	0	1	.	4	5	2	1	1	3	2	2	1	2	0	Jonas	Heering	männlich	14-Sep-19...	173	PRD_M	
10	0	0	0	0	0	0	0	1	0	5	7	1	2	2	1	2	1	2	2	1	Jonas	Heering	männlich	14-Sep-19...	173	PRD_M	
11	0	0	0	1	0	0	0	1	.	5	3	2	2	1	3	1	1	1	4	0	Thea	Tschudin	weiblich	14-Mar-196...	170	MED_K	
12	0	1	0	0	0	1	0	0	0	5	6	1	1	2	1	1	1	2	2	1	Thea	Tschudin	weiblich	14-Mar-196...	170	MED_K	
13	0	1	0	0	0	0	0	1	.	2	5	2	1	1	2	2	2	3	2	0	Maja	Zimprich	weiblich	7-Apr-1979...	114	PRD_M	
14	0	0	1	0	1	1	0	0	0	3	6	1	1	1	1	1	1	1	1	1	Maja	Zimprich	weiblich	7-Apr-1979...	114	PRD_M	
15	0	0	0	1	1	0	0	0	.	2	5	2	2	2	3	2	3	3	5	0	Biancha	Fernánde	weiblich	31-Aug-19...	26	TEL_G	
16	0	0	0	0	0	0	0	1	0	1	7	1	1	2	1	1	2	1	1	1	Biancha	Fernánde	weiblich	31-Aug-19...	26	TEL_G	
17	1	1	0	0	0	0	0	0	.	3	4	2	1	2	9	2	2	3	2	0	Nadia	Loacker	weiblich	2-Jul-1950...	38	CON_G	
18	0	1	1	1	0	0	0	0	0	3	7	1	1	2	1	1	2	1	1	1	Nadia	Loacker	weiblich	2-Jul-1950...	38	CON_G	
19	1	0	0	1	0	0	1	0	.	3	3	2	1	9	1	3	2	2	2	0	Carmelo	Sturzene	männlich	3-Apr-1976...	136	CON_M	
20	0	1	0	0	0	1	0	1	0	3	3	3	.	.	.	.	.	.	.	1	Carmelo	Sturzene	männlich	3-Apr-1976...	136	CON_M	
21	0	0	1	0	0	0	0	1	.	4	6	1	1	1	2	1	3	2	5	0	René	Siems	männlich	22-Apr-198...	73	KRK_M	
22	0	1	0	0	0	1	1	0	0	5	5	1	2	4	1	2	2	2	2	1	René	Siems	männlich	22-Apr-198...	73	KRK_M	
23	0	0	0	1	1	1	0	1	.	2	3	2	3	2	3	9	1	2	4	0	Irene	Sterli	weiblich	4-Oct-1977...	187	RET_M	

Variablen- und Datenübersicht

1. Variablenübersicht

Es gibt verschiedene Möglichkeiten, um sich einen Überblick über die Definition der erfassten Variablen zu verschaffen.

(a) Eine gute Übersicht über die im Datenfile vorhandenen Variablen erhalten Sie über



Extras | Variablen oder über

Markieren Sie eine Variable (z. B. Geschlecht) und Sie erhalten die folgenden Variablen-Informationen:

- Datenformat
- Variablenlabel
- Benutzerdefinierte fehlende Werte
- Wertelabel

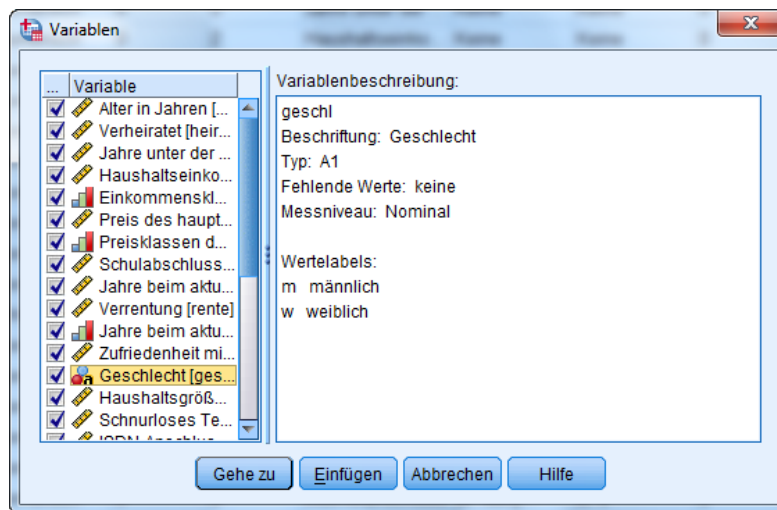


Abbildung 28: Ansicht "Variablenbeschreibung" der Variablen Geschlecht

Über die Schaltfläche **Gehe zu** können Sie im Fenster des Daten-Editors direkt zur ausgewählten Variable wechseln.

(b) Auch die Funktion **“Variableneigenschaften definieren”** bietet eine Übersicht über die Variablendefinitionen. Diese können hier jedoch auch korrigiert werden. Die Funktion finden Sie unter

Daten | Variableneigenschaften definieren...

Markieren Sie im ersten Fenster sämtliche aufgelisteten Variablen und verschieben Sie diese mit der Pfeiltaste in den Bereich **Zu durchsuchende Variablen**. Klicken Sie auf **Weiter** und anschliessend auf eine der aufgelisteten Variablen. Im rechten Bereich werden nun sämtliche Definitionen sowie die Anzahl der Fälle pro Wert dieser Variablen angezeigt (vgl. Abb. 29).

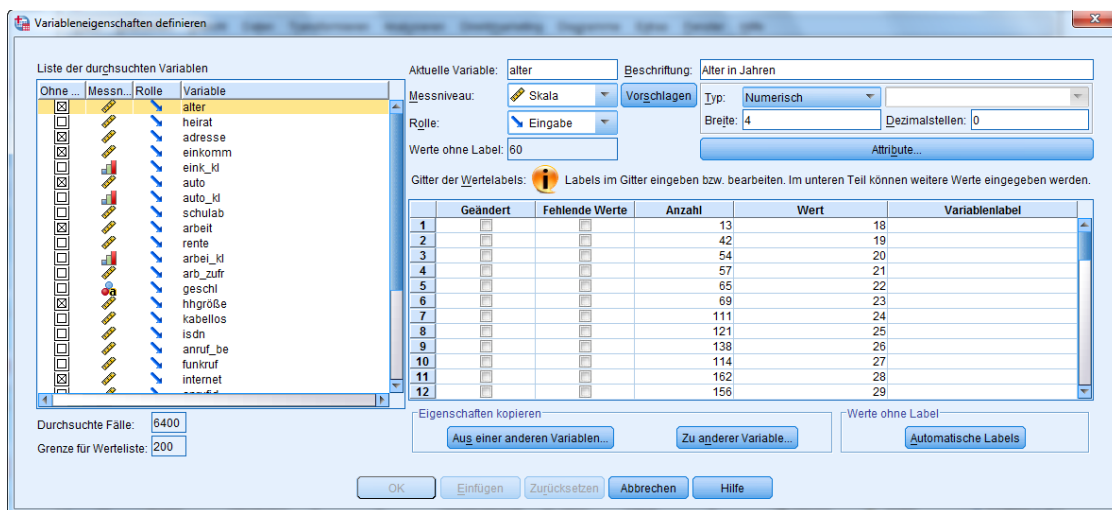
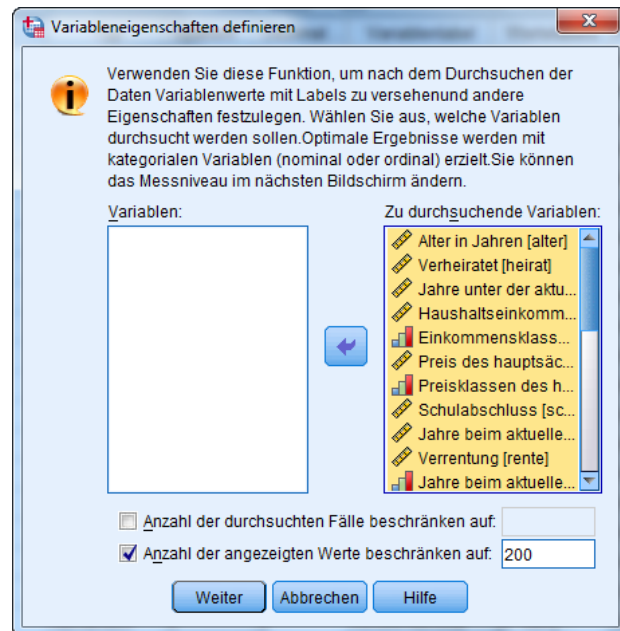


Abbildung 29: Ansicht “Variableneigenschaften definieren”

Ergänzungen bzw. Korrekturen zu den jeweiligen Variablendefinitionen können nun direkt in diesem Fenster vorgenommen werden.

Wenn Sie sich nicht sicher sind, welches Messniveau einer Variablen zugewiesen werden soll, klicken Sie auf die Schaltfläche **Vorschlagen**. Die aktuelle Variable wird nun anhand der durchsuchten Fälle und der definierten Wertelabels bewertet. Anschliessend wird das Dialogfeld **Messniveau vorschlagen** mit einem Vorschlag für das Messniveau eingeblendet. Im Bereich **Erklärung** finden Sie eine kurze Beschreibung der Kriterien, nach denen das vorgeschlagene Messniveau ausgewählt wurde.

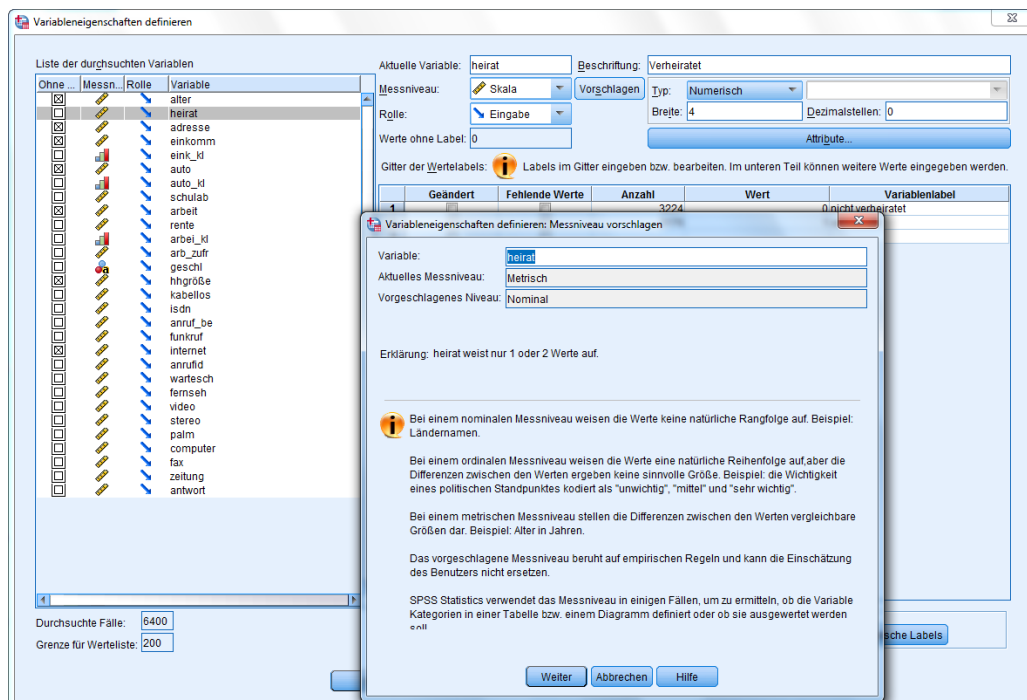


Abbildung 30: Ansicht "Messniveau vorschlagen"

(c) Eine andere Möglichkeit, um die **Datendatei-Informationen** jederzeit zur Hand zu haben, finden Sie unter

Datei | Datendatei-Informationen anzeigen | Arbeitsdatei

Die Informationen zur Datendatei werden nun in den **SPSS Statistics Viewer** geschrieben, den Sie gegebenenfalls als separates File abspeichern, in Word oder Excel exportieren oder ausdrucken können. Der so erzeugte Output enthält die Variablenbeschreibungen sowie die Variablenwerte der in der Datei enthaltenen Variablen.

DISPLAY DICTIONARY.

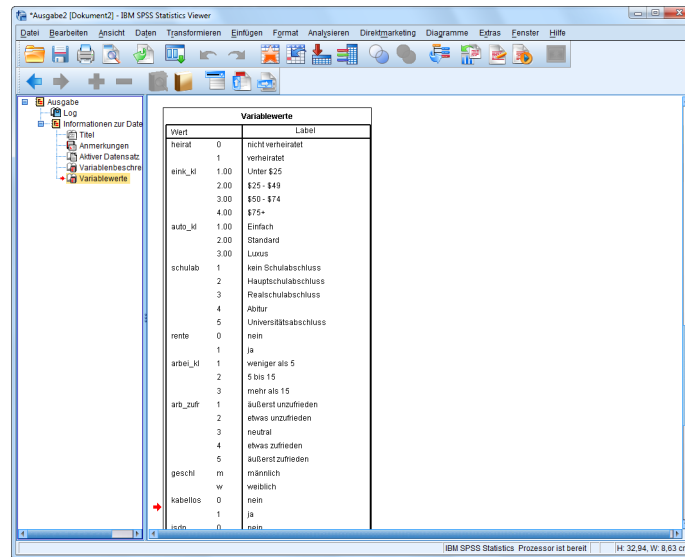
Informationen zur Datei

[DatenSet1] C:\Program Files\IBM\SPSS\Statistics\19\Samples\German\demo.sav

Variable	Position	Label	Meßniveau	Rolle	Spaltenbreite	Ausrichtung	Druckformat	Speicherformat	Fehlende Werte
alter	1	Alter in Jahren	Metrisch	Eingabe	8	Rechts	F4	F4	
heirat	2	Verheiratet	Metrisch	Eingabe	8	Rechts	F4	F4	
adresse	3	Jahre unter der aktuellen Adresse	Metrisch	Eingabe	8	Rechts	F4	F4	
einkomm	4	Haushalts einkommen in Tausend	Metrisch	Eingabe	8	Rechts	F8.2	F8.2	
eink_kl	5	Einkommensklassen in Tausend	Ordinal	Eingabe	8	Rechts	F8.2	F8.2	
auto	6	Preis des hauptsächlich genutzten Autos	Metrisch	Eingabe	8	Rechts	F8.2	F8.2	
auto_kl	7	Preisklassen des hauptsächlich genutzten Autos	Ordinal	Eingabe	8	Rechts	F8.2	F8.2	
schulab	8	Schulabschluss	Metrisch	Eingabe	8	Rechts	F4	F4	
arbeit	9	Jahre beim aktuellen Arbeitgeber	Metrisch	Eingabe	8	Rechts	F4	F4	
rente	10	Verrentung	Metrisch	Eingabe	8	Rechts	F4	F4	
arbeit_kl	11	Jahre beim aktuellen Arbeitgeber	Ordinal	Eingabe	8	Rechts	F4	F4	

Abbildung 31: Ansicht "Informationen zur Datei" im Viewer

Falls für Sie einzelne Spalten der erzeugten Ausgabe nicht von speziellem Interesse sind, können Sie diese im Output auch ausblenden.



Doppelklicken Sie auf die Tabelle und verschieben Sie die rechte Spaltenmarkierung mit gedrückter linker Maustaste soweit nach links, bis die Meldung **Ausblenden** erscheint.

Pivot-Tabelle Variablenbeschreibungen									
Variablenbeschreibungen									
Variable	Position	Label	Meßniveau	Rolle	Spaltenbreite	Ausrichtung	Druckformat	Speicherformat	Fehlerwert
alter	1	Alter in Jahren	Metrisch	Eingabe	8	Rechts	F4	F4	
heirat	2	Verheiratet	Metrisch	Eingabe	8	Rechts	F4	F4	
adresse	3	Jahre unter der aktuellen Adresse	Metrisch	Eingabe	8	Rechts	F4	F4	
einkomm	4	Haushaltseinkommen in Tausend	Metrisch	Eingabe	8	Rechts	F8.2	F8.2	
eink_kl	5	Einkommensklassen in Tausend	Ordinal	Eingabe	8	Rechts	F8.2	F8.2	
auto	6	Preis des hauptsächlich genutzten Autos	Metrisch	Eingabe	8	Rechts	F8.2	F8.2	
auto_kl	7	Preisklassen des hauptsächlich genutzten Autos	Ordinal	Eingabe	8	Rechts	F8.2	F8.2	
schulab	8	Schulabschluss	Metrisch	Eingabe	8	Rechts	F4	F4	
arbeit	9	Jahre beim aktuellen Arbeitgeber	Metrisch	Eingabe	8	Rechts	F4	F4	

Ebenso können Sie einzelne Spalten verbreitern, indem Sie die Spaltenmarkierung weiter nach rechts ziehen.

2. Datenübersicht

Ein möglichst umfassender Überblick über die Verteilung der erfassten Variablen ist sehr wichtig bei der Vorbereitung für eine Datenanalyse. Folgende Funktionen dienen der Datenübersicht:

- **Kategoriale Variablen** (z.B. ja/nein oder sehr gut, gut, befriedigend ...)
Auszählung der jeweiligen Ausprägungen via Häufigkeitstabelle sowie einer grafischen Darstellung mit einem Balken- oder Kuchendiagramm.
- **Metrische Variablen** (z.B. Alter, Einkommen)
Es bieten sich statistische Masszahlen wie Mittelwert, Median und Standardabweichung an.
- **Codebuch**
Sie können in einem Schritt die Datenlexikoninformation – wie Variablennamen, Variablenlabels, Wertelabels, fehlende Werte – sowie Auswertungsstatistiken für ausgewählte Variablen im aktiven Daten-Set ausgeben.

2.1 Kategoriale Variablen

In SPSS Statistics können Sie parallel zu den Häufigkeitstabellen auch gleich entsprechende Grafiken erzeugen, die diese Werte visualisieren. Wählen Sie für die Datei **demo.sav** aus dem Menü

Analysieren | Deskriptive Statistiken | Häufigkeiten aus.

Im Dialogfeld Häufigkeiten markieren Sie in der Variablenliste die Variablen **geschl** (Geschlecht) und **eink_kl** (Einkommensklassen in Tausend) und übertragen sie in das Feld **Variable(n)**. Achten Sie darauf, dass die Option **Häufigkeitstabellen anzeigen** aktiviert ist.

Hinweis: Mit einem Rechtsklick auf eine Variable in der Quellvariablenliste des Dialogfeldes Häufigkeiten können Sie die Variablennamen oder Variablenlabels anzeigen und

ändern sowie eine Variablenbeschreibung ansehen. Falls Sie die kürzeren Variablennamen in den Dialogfeldern standardmässig angezeigt haben wollen, können Sie dies über

Bearbeiten | Optionen

im Register **Allgemein** unter **Variablenlisten** einstellen, indem Sie **Namen anzeigen** auswählen.

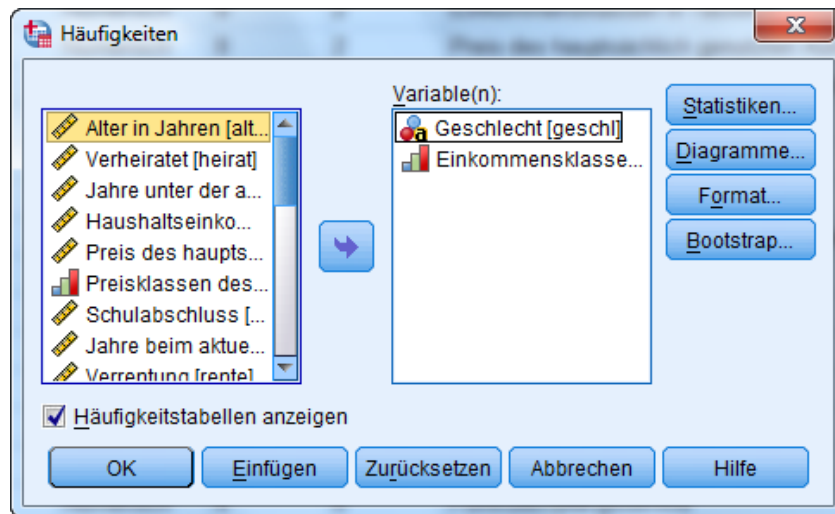


Abbildung 32: Ansicht Dialogfeld "Häufigkeiten"

Klicken Sie auf die Schaltfläche **Diagramme**. Im Dialogfeld **Häufigkeiten: Grafiken** wählen Sie **Balkendiagramme**⁵ und unter **Diagrammwerte** **Prozente**, um die Prozentwerte grafisch darzustellen, und wählen dann **Weiter**.

⁵ Beachten Sie: In SPSS werden Säulendiagramme als Balkendiagramme bezeichnet, was etwas irreführend ist.

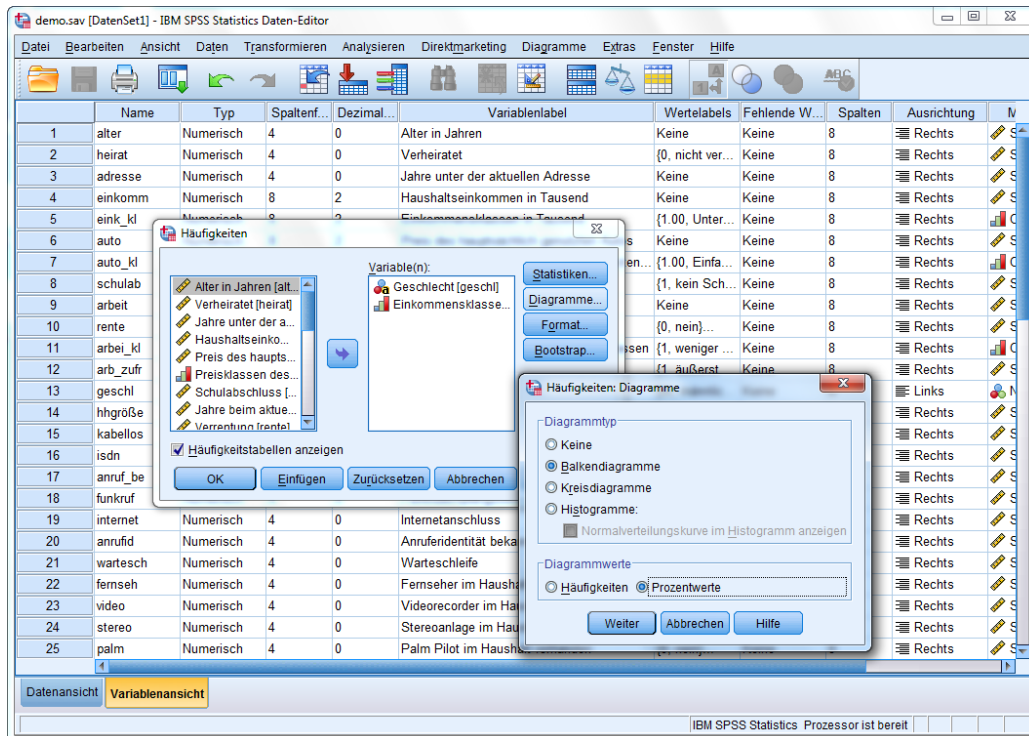
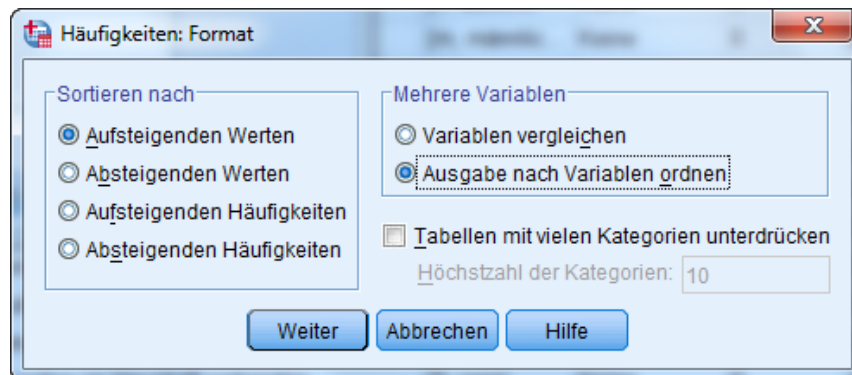


Abbildung 33: Ansicht "Häufigkeiten" und "Häufigkeiten Diagramme"

Klicken Sie schliesslich auf die Schaltfläche **Format** und wählen die Option **Ausgabe nach Variablen ordnen** aus, um für jede Variable das Balkendiagramm direkt nach der entsprechenden Häufigkeitstabelle anzuzeigen.



Klicken Sie auf **Weiter** und dann auf **OK**. Statistiken, Häufigkeitstabelle sowie Diagramm werden für jede Variable im Viewer angezeigt.

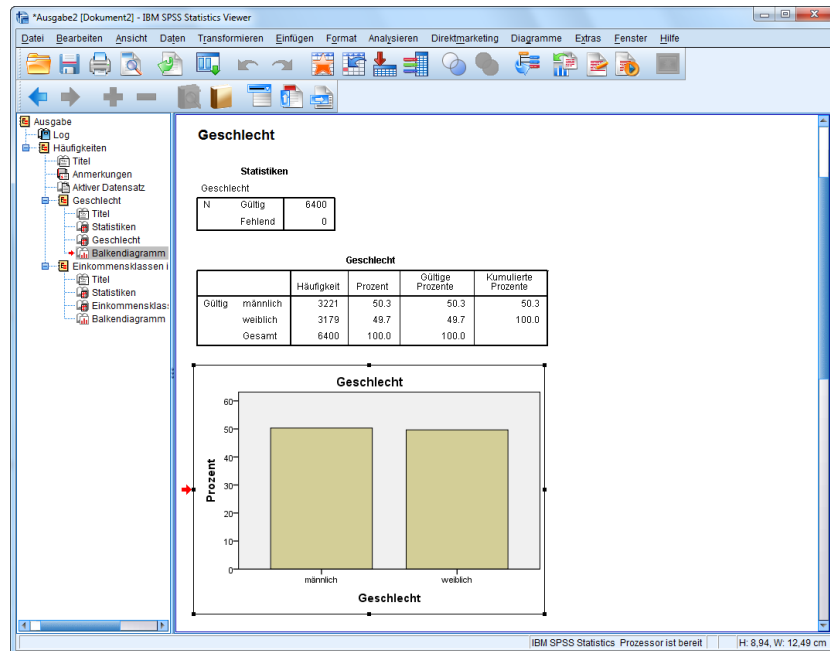


Abbildung 34: Ansicht "Ausgabe Variable Geschlecht (Statistik, Häufigkeitstabelle und Balkendiagramm)"

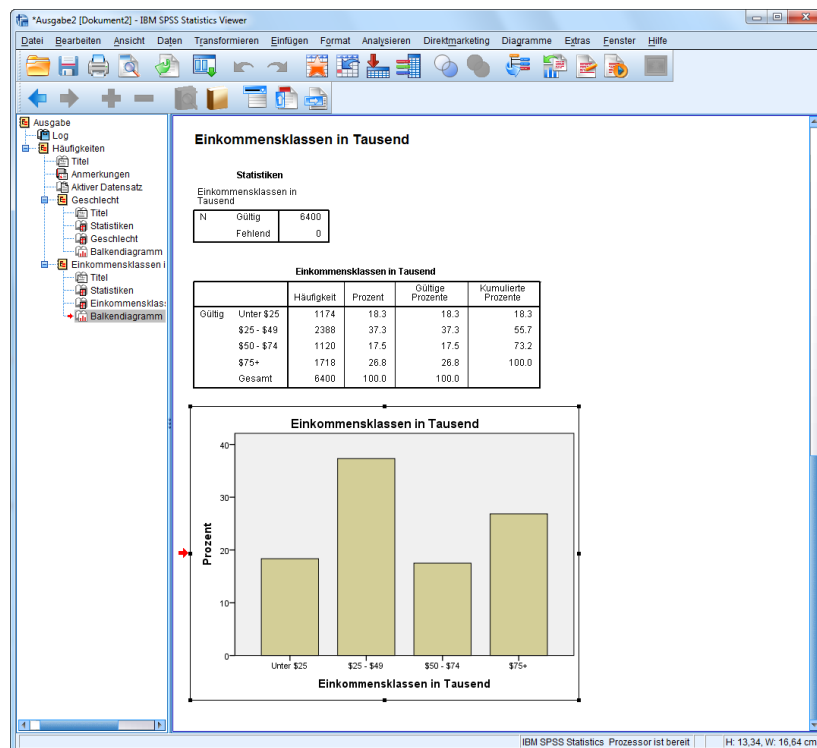


Abbildung 35: Ansicht "Ausgabe Variable Einkommensklassen in Tausend (Statistik, Häufigkeitstabelle und Balkendiagramm)"

2.2 Metrische Variablen

Bei diesen Variablen interessieren Masse der zentralen Tendenz sowie Streuungsmasse. Gehen Sie auf

Analysieren | Deskriptive Statistiken | Häufigkeiten.

Klicken Sie ausserdem auf **Zurücksetzen** um vorherige Auswahlen im Dialogfenster rückgängig zu machen. Wählen Sie dann die Variable **einkomm** (Haushaltseinkommen in Tausend) aus und geben Sie unter **Statistiken** an, welche statistischen Masszahlen zur Beschreibung der Lage, Streuung und Verteilung der Werte dieser Variablen ausgegeben werden sollen. Wählen Sie z.B. Mittelwert, Median, Standardabweichung, Minimum und Maximum aus.

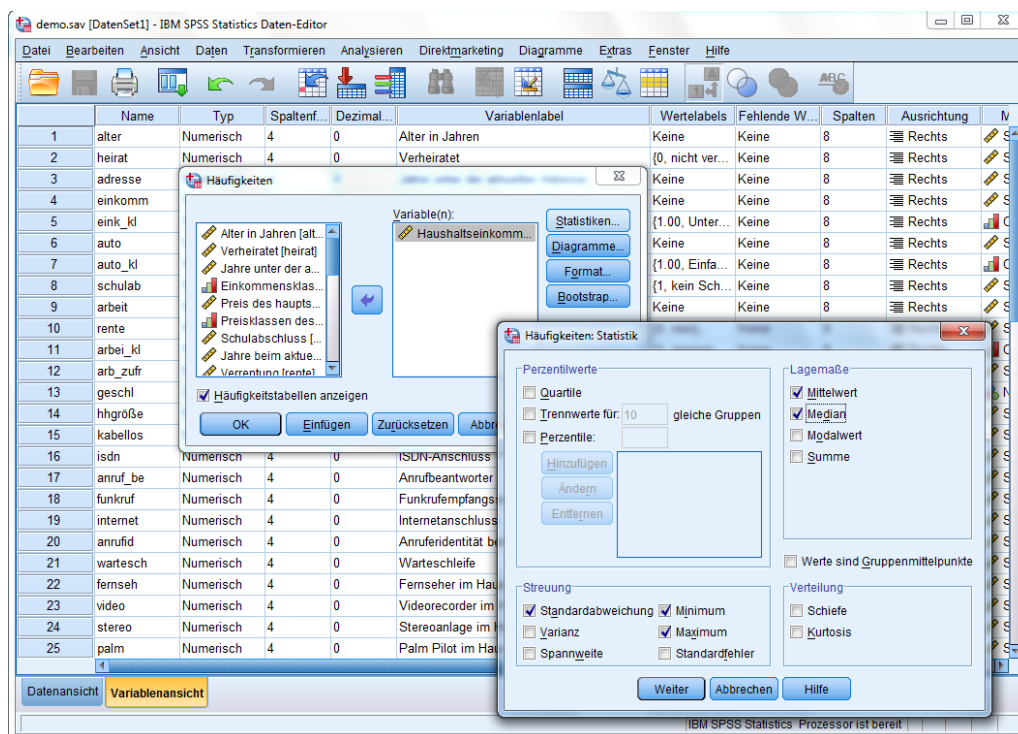


Abbildung 36: Ansicht "Häufigkeiten: Statistik"

Klicken Sie dann auf **Weiter** und achten Sie darauf, dass bei “Häufigkeitstabellen anzeigen” das Häkchen deaktiviert ist. Klicken Sie auf **OK**, um die Prozedur auszuführen.

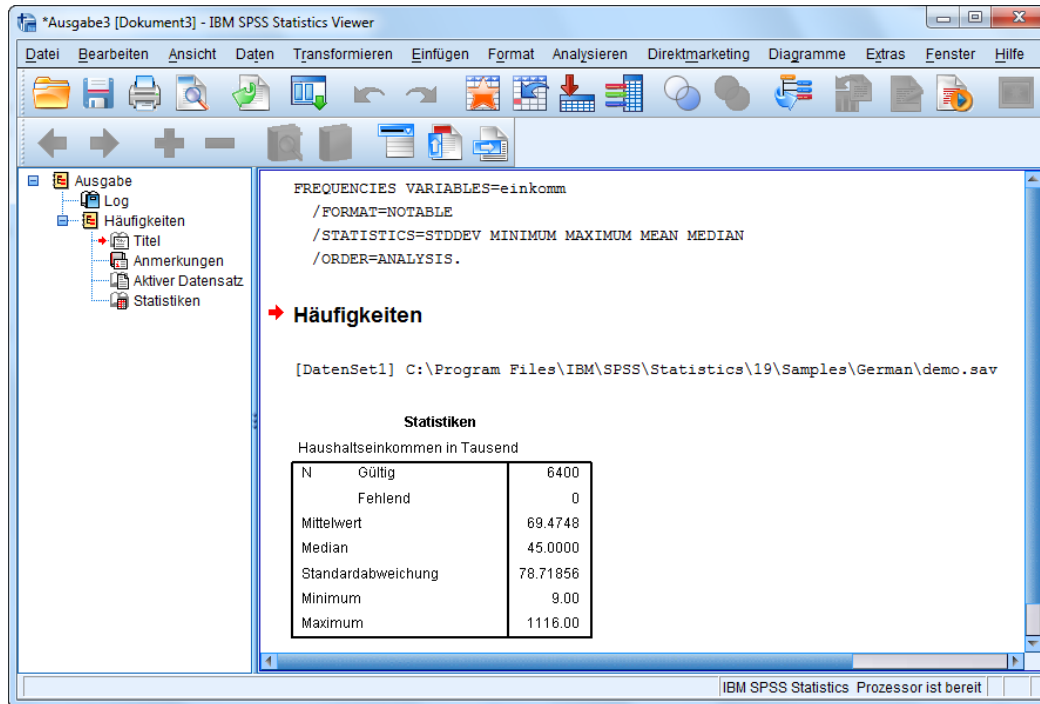


Abbildung 37: Ansicht “Ausgabe Variable Haushaltseinkommen in Tausend”

Die Daten in obiger Abbildung zeigen bspw. eine grosse Differenz zwischen Mittelwert und Median. Der Mittelwert ist um fast 25000 grösser als der Median. Die Werte sind demnach nicht normalverteilt. Die Verteilung kann durch ein Histogramm anschaulich dargestellt (überprüft) werden.

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



Um ein **Histogramm** auszugeben, öffnen Sie wieder das Dialogfeld **Häufigkeiten** (Analysieren -> Deskriptive Statistik -> Häufigkeiten...) und wählen die Option **Diagramme**. Wählen Sie dann **Histogramm** und **Normalverteilungskurve im Histogramm anzeigen**. Klicken Sie auf **Weiter** und **OK**.

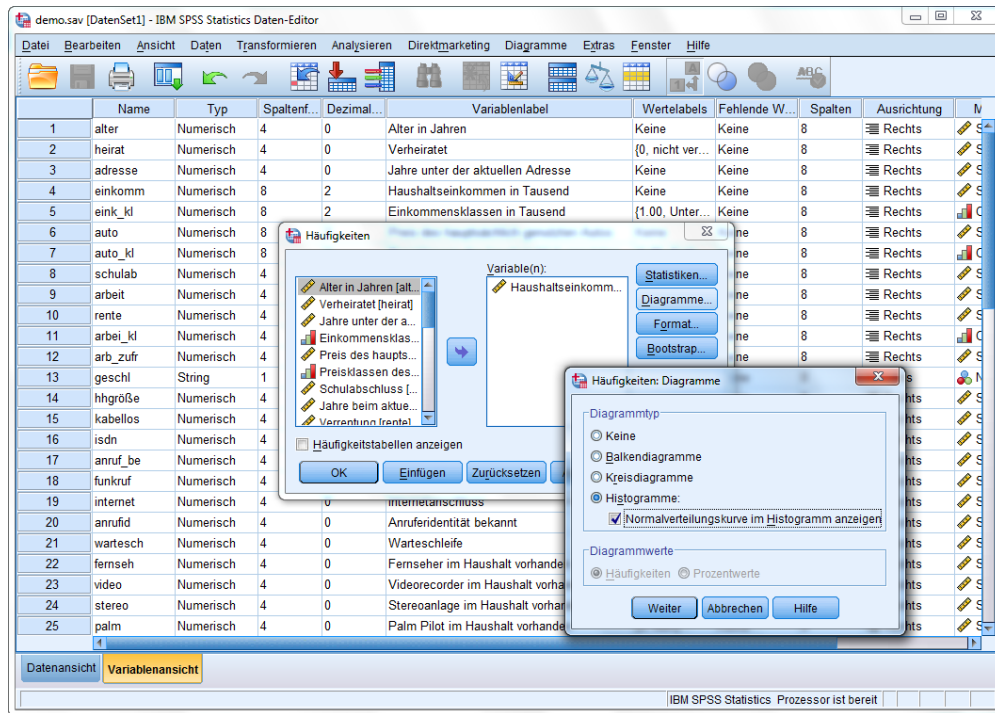


Abbildung 38: Ansicht "Häufigkeiten" und "Häufigkeiten: Diagramme"

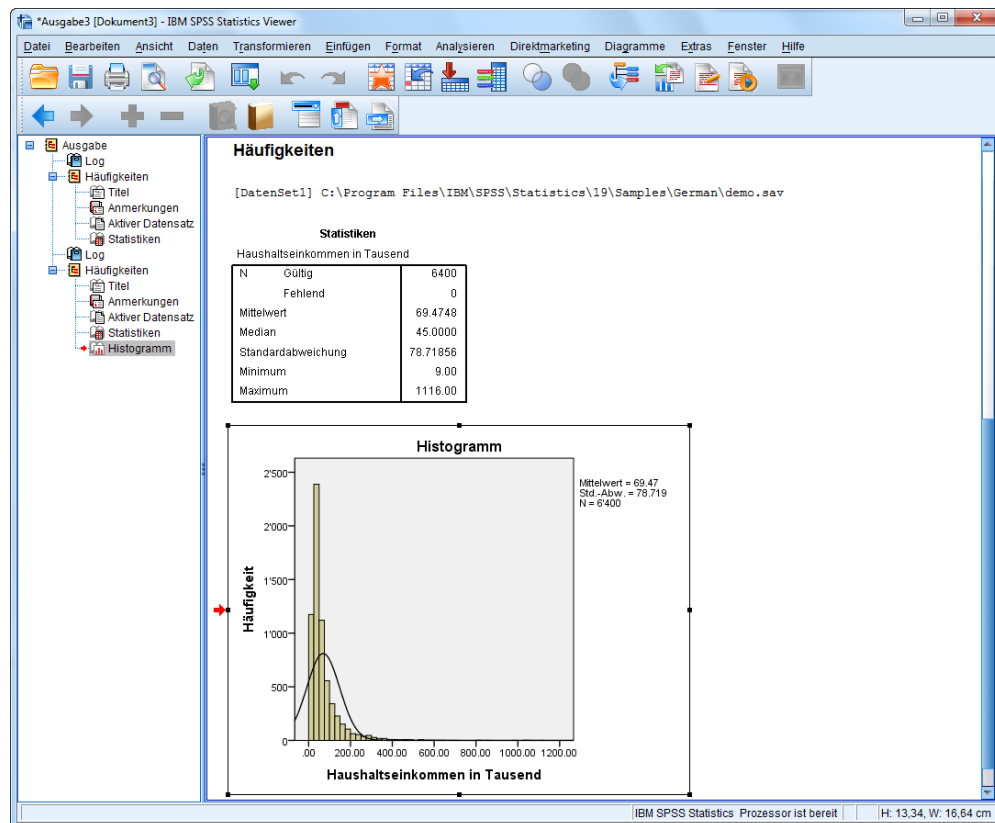


Abbildung 39: Ansicht "Ausgabe Variable Haushaltseinkommen in Tausend (Statistiken, Histogramm)"

Interpretation Abbildung 39:

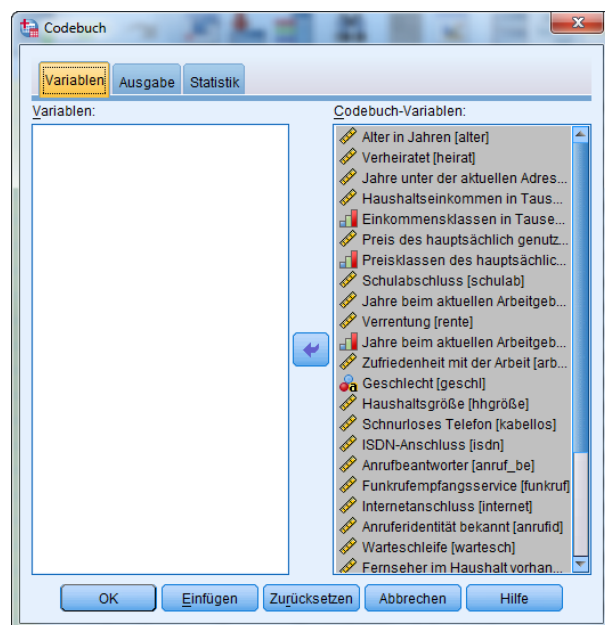
- Die meisten Fälle liegen unter 100.000, ein paar Fälle im Bereich um 500.000 und darüber.
- Hohe Werte bei einigen wenigen Fällen haben grosse Auswirkungen auf den Mittelwert, aber nur geringe oder gar keine auf den Median. Der Median ist in diesem Beispiel ein besserer Indikator für die zentrale Tendenz.

2.3 Codebuch

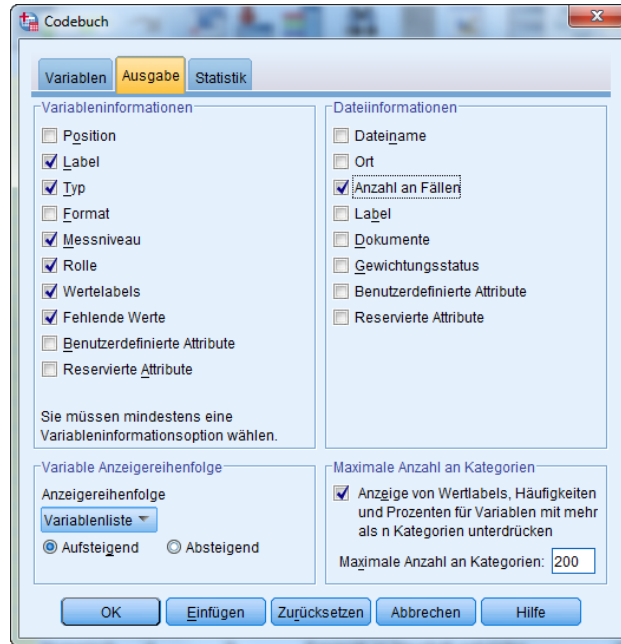
Sie können auch ein Codebuch ausgeben lassen. Das Codebuch gibt neben den Datenlexikoninformationen – wie Variablennamen, Variablenlabels, Wertelabels, fehlende Werte – auch Auswertungsstatistiken für alle oder bestimmte Variablen im aktiven Daten-Set aus. Für nominale und ordinale Variablen enthalten die Auswertungsstatistiken Häufigkeiten und Prozentangaben, für metrische Variablen werden standardmässig Mittelwert, Standardabweichung und Quartile angezeigt. Um ein Codebuch zu erstellen, wählen Sie im Menü

Analysieren | Berichte | Codebuch... aus.

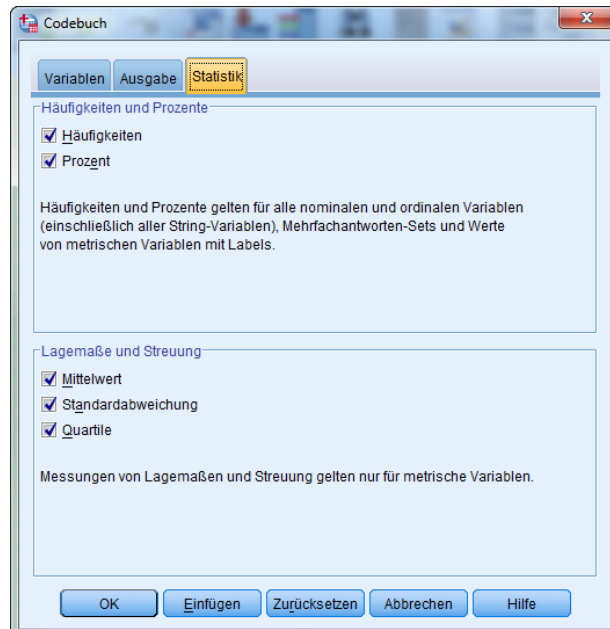
Klicken Sie auf die Registerkarte **Variablen** und wählen Sie alle Variablen aus.



Die Registerkarte **Ausgabe** steuert die Variablenbeschreibungen, die für jede Variable enthalten sind, die in der Reihenfolge, in der die Variablen angezeigt werden, und den Inhalt der optionalen Dateiinformationstabelle. Deaktivieren Sie **Position**, **Format**, **Benutzerdefinierte Attribute** und **Reservierte Attribute** unter den Variableninformationen. Als Dateiinformation wählen Sie **Anzahl an Fällen** aus.



Über die Registerkarte **Statistik** können Sie die Auswertungsstatistiken steuern, die in die Ausgabe aufgenommen werden, oder die Anzeige von Auswertungsstatistiken komplett unterdrücken. Belassen Sie die Standardeinstellung und klicken sie auf **OK**.



Für jede ausgewählte kategoriale Variable erhalten Sie im Output eine Häufigkeitsauszählung der einzelnen Kategorien bzw. die angeforderten Lage- und Streuungsmasse für alle metrischen Variablen.

*Ausgabe3 [Dokument3] - IBM SPSS Statistics Viewer

Datei Bearbeiten Ansicht Daten Transformieren Einfügen Format Analysieren Direktmarketing Diagramme Extras Fenster Hilfe

← → + - [Icons]

Codebuch

[DatenSet1] C:\Program Files\IBM\SPSS\Statistics\19\Samples\German\demo.sav

Informationen zur Datei

Anzahl der Fälle	Ungewichtet	6400
	Gewichtet	6400

alter

		Wert
Standardattribute	Label	Alter in Jahren
	Typ	Numerisch
	Messung	Metrisch
	Rolle	Eingabe
N	Gültig	6400
	Fehlend	0
Zentrale Tendenz und Streuung	Mittelwert	42.06
	Standardabweichung	12.290
	Perzentil 25	33.00
	Perzentil 50	41.00
	Perzentil 75	51.00

heirat

		Wert	Anzahl	Prozent
Standardattribute	Label	Verheiratet		
	Typ	Numerisch		
	Messung	Metrisch		
	Rolle	Eingabe		
N	Gültig	6400		
	Fehlend	0		
Zentrale Tendenz und Streuung	Mittelwert	.50		
	Standardabweichung	.500		
	Perzentil 25	.00		
	Perzentil 50	.00		
	Perzentil 75	1.00		
Bezeichnete Werte	0	nicht verheiratet	3224	50.4%

Zum Bearbeiten von Pivot-Tabelle doppelklicken

IBM SPSS Statistics Prozessor ist bereit | H: 8,92, W: 13,68 cm

Abbildung 40: Ansicht "Ausgabe Codebuch"

Modifikation von Daten

1. Umkodieren von Variablen

Datenmodifikationen, d.h. die Veränderung von Daten vor der Analyse durch bestimmte Operationen, können unter dem Menüpunkt **Transformieren** gewählt werden. Im Folgenden soll das Umkodieren von Variablen näher betrachtet werden. Umkodieren bedeutet, Werte in vorhandene Variablen umzukodieren oder neue Variablen auf der Grundlage der umkodierten Werte zu erzeugen. Es gibt drei Möglichkeiten:

1.1 Umkodieren in dieselben Variablen (manuell)

Bei dieser Option können Werte vorhandener Variablen erneut zugewiesen werden oder Bereiche vorhandener Werte in einen neuen Wert zusammengefasst werden (z.B. Löhne in Kategorien von Lohnbereichen zusammenfassen).

Numerische Variablen und String-Variablen können umkodiert werden. Bei der Auswahl mehrerer Variablen müssen diese vom gleichen Typ sein. Öffnen Sie die Datendatei **demo.sav** und wählen Sie

Transformieren | Umkodieren in dieselben Variablen aus.

Wählen Sie die Variable **Haushaltseinkommen in Tausend [einkomm]** und klicken Sie auf die Schaltfläche **Alte und neue Werte**.

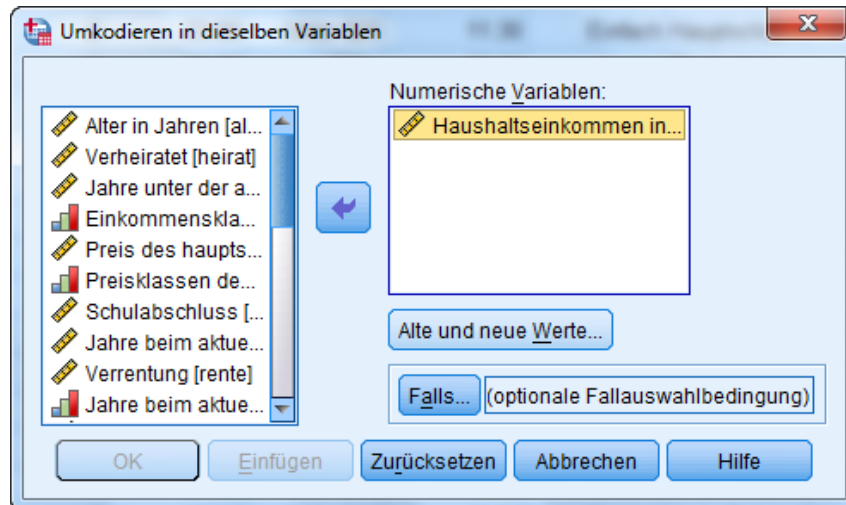


Abbildung 41: Ansicht "Umkodieren in dieselben Variablen"

Erfassen Sie folgende **alte und neue Werte**:

Bereich	1-50	->	50
	51-100	->	100
	101-200	->	200

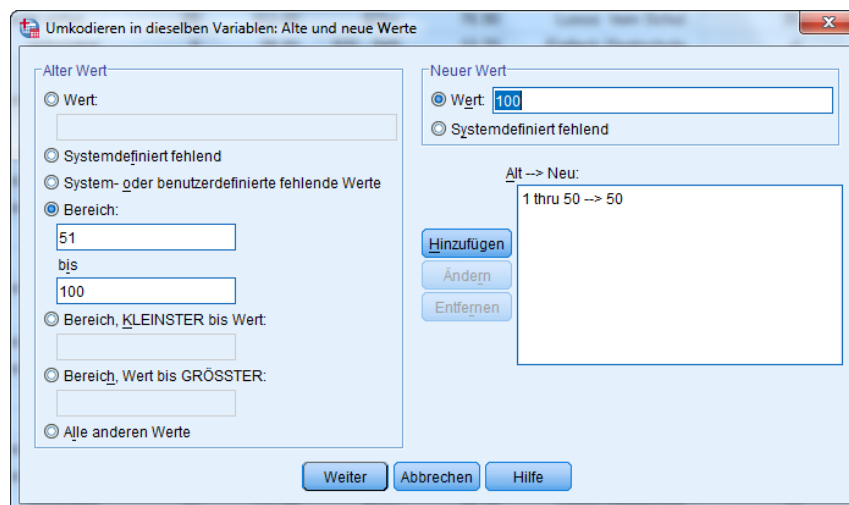


Abbildung 42: Ansicht "Umkodieren in dieselben Variablen: Alte und neue Werte"

Geben Sie jeweils einen Wertebereich ein (z.B. 1-50 -> 50) und klicken Sie auf **Hinzufügen**. Nach Abschluss der Eingabe klicken Sie auf **Weiter** und auf **OK**.

1.2 Umkodieren in andere Variablen (manuell)

Erstellen Sie ein neues Arbeitsblatt mit dem Titel **Sonntagsbefragung**. Das Ergebnis einer Sonntagsbefragung soll manuell erfasst und in eine andere Variable umkodiert werden. Erfassen Sie die Variable **partei** (vgl. Abbildung 43).

Parteien: SPD, Grüne, Linkspartei, CDU/CSU, FDP, Republikaner, Sonstige, Keine Antwort

Skalenniveau: nominal, keine Dezimalstellen, Variablenlabel: Was würden Sie wählen, wenn am Sonntag Bundestagswahlen wären? Definieren Sie die Wertelabels (von 1 bis 7 für SPD, Linkspartei etc.).

Der Variablen **partei** sollen folgende Wertelabels zugewiesen werden:

Variable partei	
Werte	Wertelabels
-99	Keine Antwort
1	CDU/CSU
2	FDP
3	SPD
4	Grüne
5	Linkspartei
6	NPD
7	Sonstige

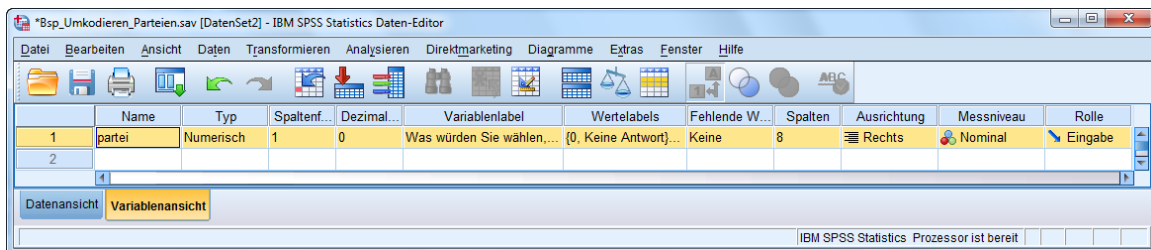


Abbildung 43: Variablenansicht der Variablen **partei**

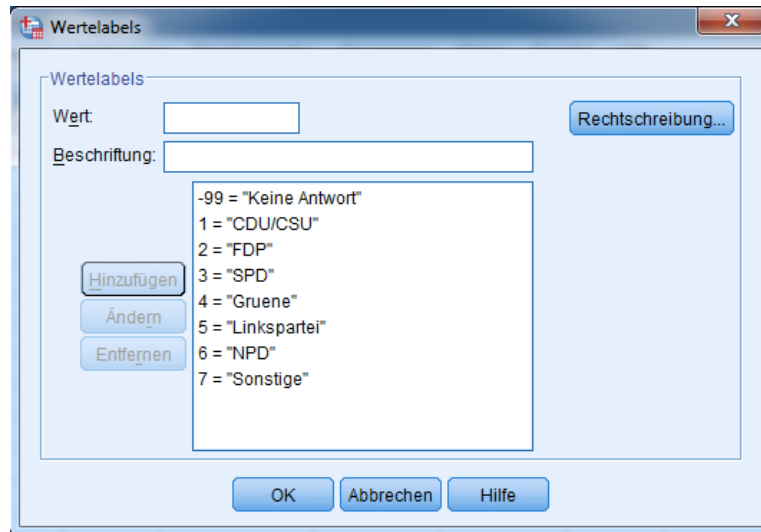


Abbildung 44: Ansicht "Wertelabels" der Variablen partei

Erfassen Sie in der Datenansicht 20 fiktive Angaben:

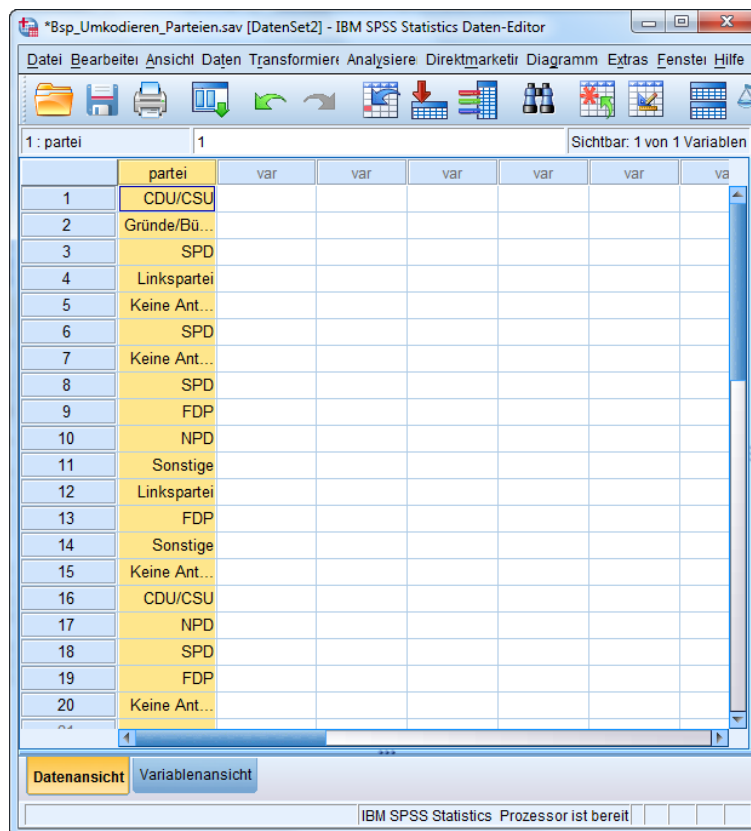


Abbildung 45: Datenansicht der Variable partei

Aufgabe: Die prozentuale Aufteilung gemäss des klassischen Links-Rechts-Spektrums soll ermittelt werden. Dazu soll die Variable **partei** in eine andere Variable **lire** (links-rechts) umkodiert werden.

Linksparteien:	Rechtsparteien:	Nicht definiert:
SPD	CDU/CSU	Keine Antwort
Grüne	FDP	Sonstige
Linkspartei	Republikaner	

Wählen Sie

Transformieren | Umkodieren in andere Variablen aus.

Verschieben Sie die Variable **partei** von der Quell- in die **Wahlvariablenliste**. Geben Sie bei der Ausgabevariablen im Editierfeld **Name** den neuen Variablennamen **lire** ein sowie bei Beschriftung **Parteienspektrum** und klicken Sie auf **Ändern** (Das Editierfeld Name bezieht sich dabei auf den max. 8-stelligen Variablennamen, die Beschriftung auf das Variablenlabel).

Legen Sie die umzukodierenden Werte fest und klicken Sie dazu auf die Schaltfläche **Alte und neue Werte**. Die Kodierung der neuen Variable **lire** sollte folgendermassen aussehen:

Alter Wert Variable partei	Neuer Werte Variable lire
1	2
2	2
3	1
4	1
5	1
6	2

Alle anderen Werte [ELSE] sollen -99 sein.

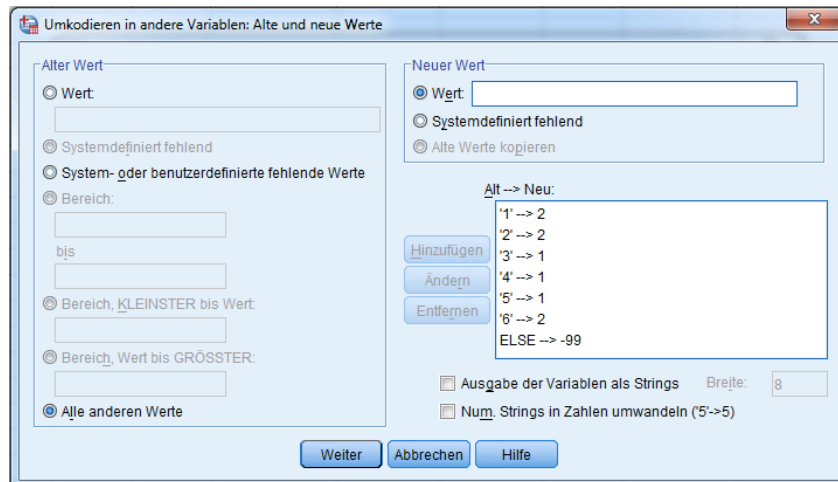


Abbildung 46: Ansicht "Umkodieren in andere Variablen: Alte und neue Werte"

Beachten Sie: Auch Werte, die gleich bleiben sollen, müssen definiert werden, sodass sie in der neuen Variablen erscheinen.

Klicken Sie auf **Weiter** und **OK**. Die neue Variable **lire** erscheint in der **Variablenansicht**.

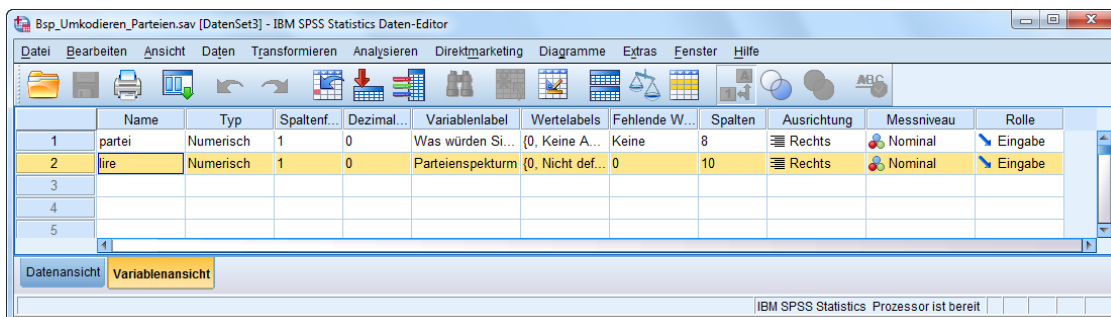
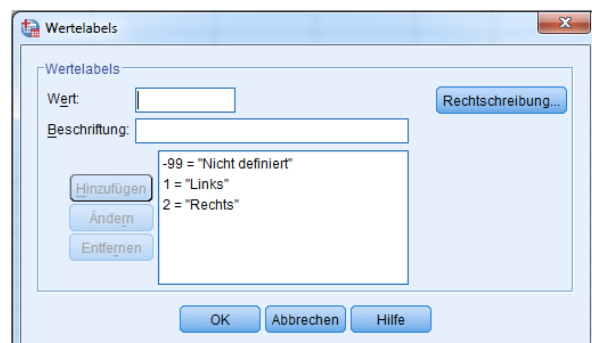


Abbildung 47: Variablenansicht der neuen Variablen lire

Nehmen Sie noch folgende Änderungen vor: Zeichen: 1, 0 Dezimalstellen, Wertelabels: -99 = nicht definiert, 1 = Links, 2 = Rechts, Spalten: 10, Skalenniveau: nominal. Deklarieren Sie -99 als fehlenden Wert.



1.3 Automatisch umkodieren

Öffnen Sie die Datei **Bsp_umkodieren_automatisch**. Es soll folgende Hypothese untersucht werden: Väter von Medizinstudenten sind häufiger in helfenden Berufen tätig als angestellt.

Aufgabe: Aus der String-Variable **berufv** soll eine dichotome Variable generiert werden, die angibt, ob ein helfender Beruf vorliegt oder nicht.

1. Schritt

Kodieren Sie die String-Variable **berufv** in eine numerische Variable um: Wählen Sie

Transformieren | Automatisch umkodieren

Übertragen Sie die Variable **berufv** in das Feld **Variable -> neuer Name** und geben Sie den neuen Namen **berufneu** ein. Klicken Sie auf Neuen Namen hinzufügen und dann auf OK.

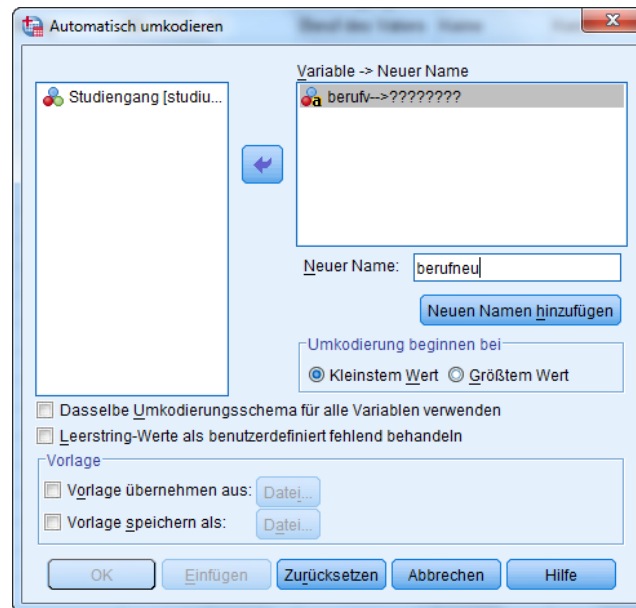


Abbildung 48: Ansicht "Automatisch umkodieren"

Im **Daten-Editor** erscheint in der **Variablenansicht** die neue Variable **berufneu**. Den Strings der String-Variablen **berufv** wurden in der Spalte Wertelabels in alphabetischer Reihenfolge die fortlaufenden natürlichen Zahlen von 1 bis 8 zugeordnet.

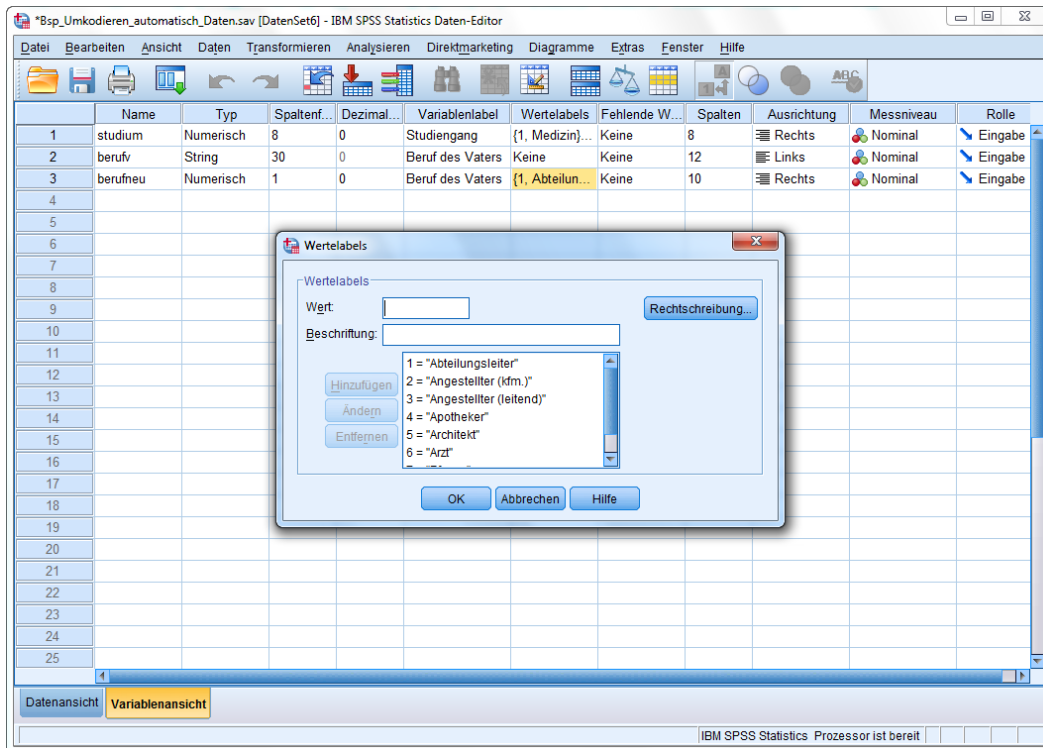


Abbildung 49: Ansicht "Wertelabels" der Variable berufneu

2. Schritt

Kodieren Sie in einem zweiten Schritt manuell um, indem Sie die Codenummern der helfenden Berufe zu einer Kategorie zusammenfassen.

1	Abteilungsleiter	1 (kein helfender Beruf)
2	Angestellter (kfm.)	1 (kein helfender Beruf)
3	Angestellter (leitend)	1 (kein helfender Beruf)
4	Apotheker	2 (helfender Beruf)
5	Architekt	1 (kein helfender Beruf)
6	Arzt	2 (helfender Beruf)
7	Pfarrer	2 (helfender Beruf)
8	Psychologe	2 (helfender Beruf)

Gehen Sie folgendermassen vor: Wählen Sie

Transformieren | Umkodieren in andere Variable aus.

Verschieben Sie die Variable **berufneu** in die Wahlvariablenliste. Geben Sie als neuen Variablennamen **helfber** ein. Als Variablenlabel: **Helfender Beruf**. Klicken Sie auf **Ändern** und dann auf **Alte und neue Variablen**. Kodieren Sie wie folgt um:

Alte Werte Variable	Neue Werte Variable
1 bis 3	1
4	2
5	1
6 bis 8	2

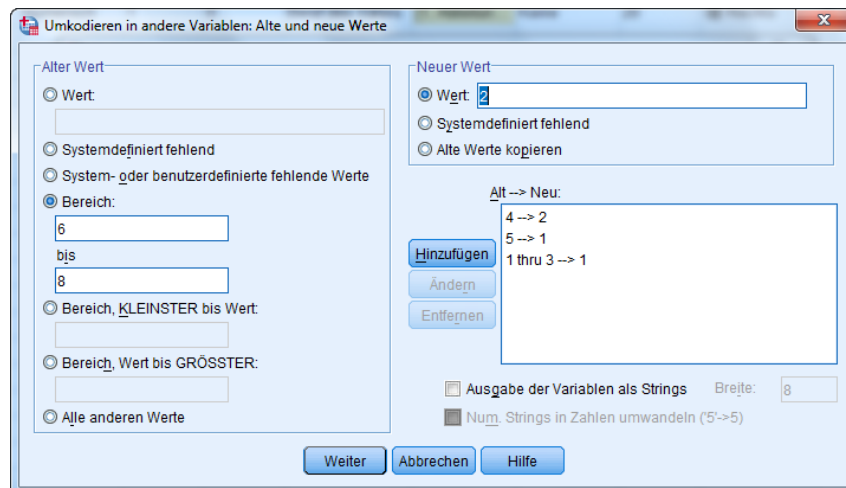


Abbildung 50: Ansicht "Umkodieren in andere Variablen: Alte und neue Werte"

Klicken Sie auf **Weiter** und **OK**. Die neue Variable **helfber** erscheint in der Variablenansicht des Daten-Editors. Setzen Sie die Dezimalstellen der Variable **helfber** auf 0 und erfassen Sie die neuen Wertelabels: 1 = KHB (kein helfender Beruf), 2 = HB (helfender Beruf).

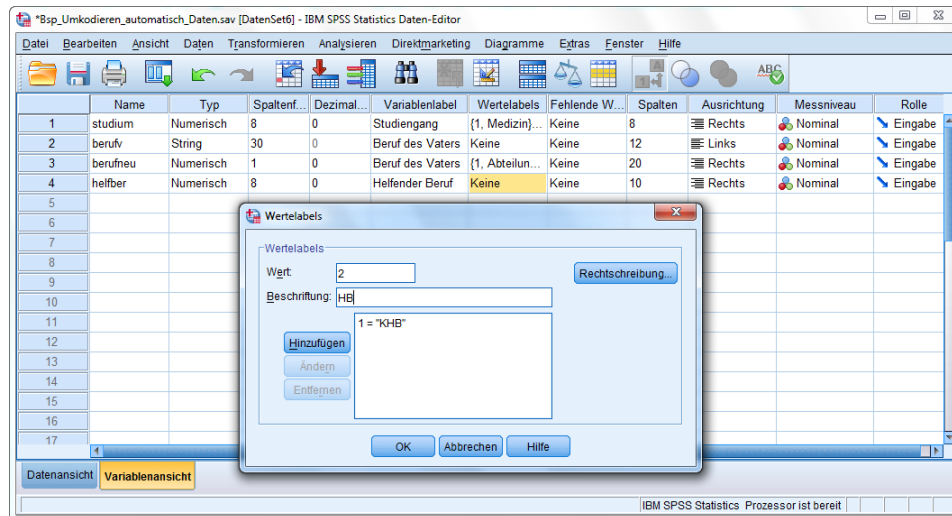


Abbildung 51: Ansicht "Wertelabels" der Variable helfber

2. Variablen berechnen

Im Laufe der Datenanalyse sind häufig Variablen zu bilden, die sich aus arithmetischen Operationen mit vorhandenen Variablen ergeben. Die Funktion "Variable berechnen" ermöglicht es, verschiedene Variablen zu einer neuen (übergeordneten) Variable zusammenzufassen. Es wird eine neue Zielvariable definiert und im Feld "numerischer Ausdruck" die mathematische Funktion eingegeben. Um Berechnungen über das Menü auszuführen, wählen Sie

Transformieren | Variable berechnen

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

Variablennamen sind Namen bereits definierter Variablen.

Konstanten (numerische) sind Zahlen, die sich aus den Ziffern 0 bis 9 und ggf. dem Dezimalpunkt zusammensetzen.

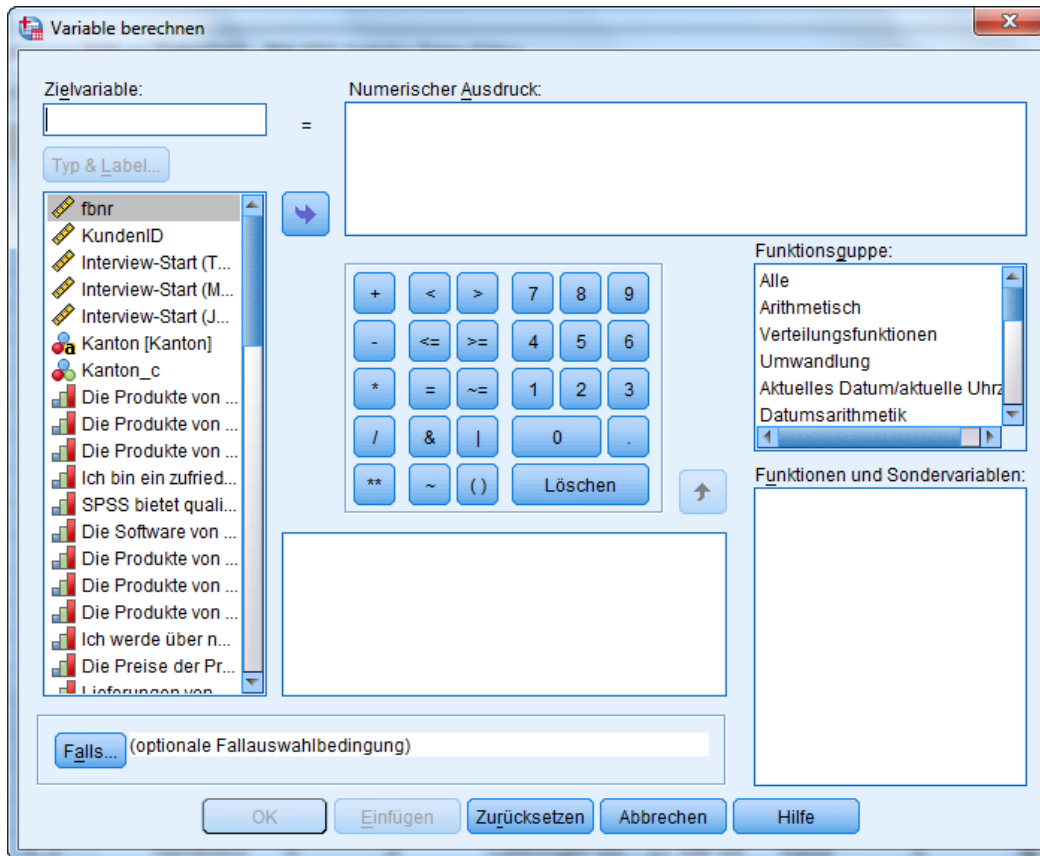


Abbildung 52: Ansicht "Variable berechnen"

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.

Datenmanagement am Beispiel eines Fragebogens

1. Konsistenzprüfungen zwischen Variablen

In der **SPSS Kundenbefragung.sav** sollte z.B. die Leistung der Verkaufsmitarbeiter nur dann beurteilt werden, wenn die befragte Person während dem letzten Jahr auch tatsächlich Kontakt zu einem Verkaufsmitarbeiter von SPSS hatte, wenn **Frage f3a** also mit “**Ja**” beantwortet wurde. Wurde **Frage f3a** mit “**Nein**” beantwortet sollten die Fragen **f3b_1 bis f3b_7** alle leer sein.

Es sollen all jene Fälle “markiert” werden, welche in den Fragen f3b_1 bis f3b_7 Bewertungen zur Leistung der Verkaufsmitarbeiter vorgenommen haben, obwohl Frage f3a mit “Nein” (Code 2) beantwortet wurde. Es soll also eine neue Variable erstellt werden, welche bei all jenen Fällen Code **1** (= Ja) enthält, welche dieses Konsistenzkriterium nicht erfüllen. Wählen Sie

Transformieren | Variable berechnen

Geben Sie als **Zielvariable** f3_check ein, als numerischer Ausdruck “1”, und klicken Sie dann auf **Falls...**, um die Konsistenz-Bedingung zu definieren.

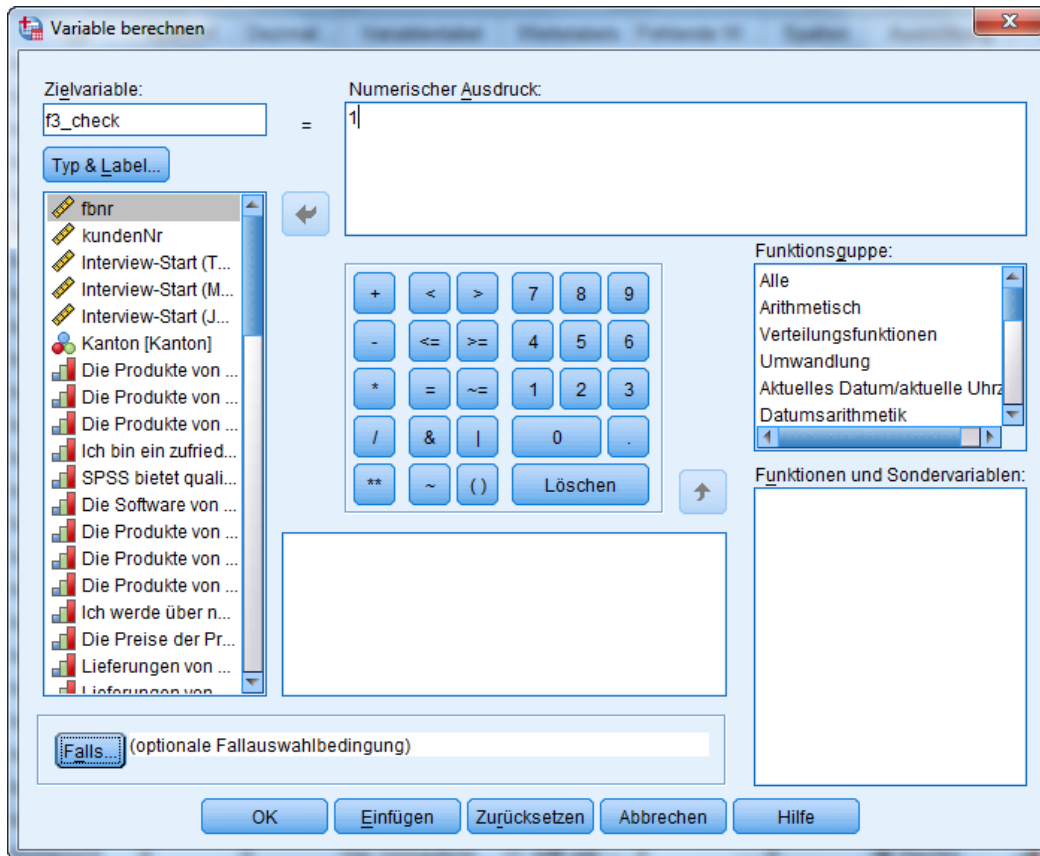


Abbildung 53: Ansicht "Variable berechnen"

Wählen Sie im nachfolgenden Fenster die Option **Fall einschliessen, wenn die Bedingung erfüllt ist** und geben Sie im darunter liegenden Bereich folgenden Ausdruck ein:

$f3a=2 \ \& \ (f3b_1 \geq 1 \mid f3b_2 \geq 1 \mid f3b_3 \geq 1 \mid f3b_4 \geq 1 \mid f3b_5 \geq 1 \mid f3b_6 \geq 1 \mid f3b_7 \geq 1)$

Das Zeichen **&** steht für eine logische Und-Verknüpfung, das Zeichen **|** steht für eine logische Oder-Verknüpfung.

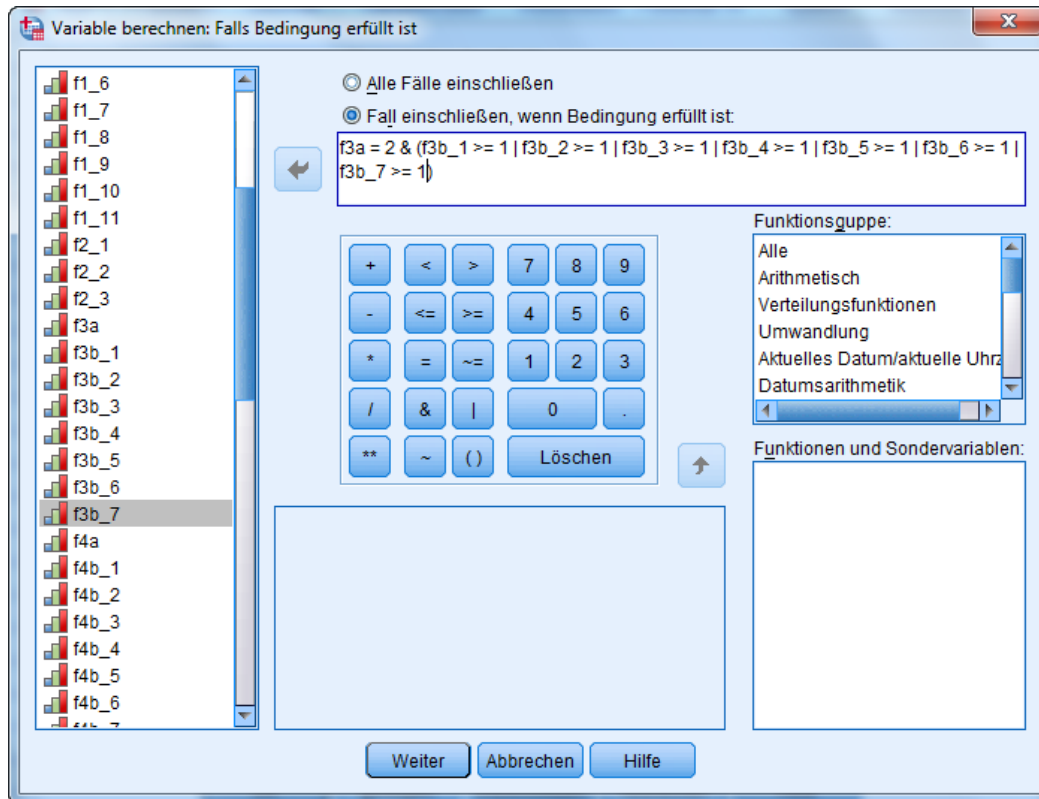


Abbildung 54: Ansicht "Variable berechnen: Falls Bedingung erfüllt ist"

Der Ausdruck **f3a=2** prüft, ob Frage f3a ("Hatten Sie während dem letzten Jahr Kontakt mit einem Verkaufsmitarbeiter von SPSS?") mit "**Nein**" beantwortet wurde, anschliessend wird kontrolliert, ob eine der Aussagen zur Leistung der Verkaufsmitarbeiter (f3b_1 bis f3b_7) einen Code grösser oder gleich 1 enthält.

Klicken Sie auf **Weiter** und **OK**, um die Berechnung laufen zu lassen. Die neu erstellte Variable finden Sie am Ende des Datenfiles. In diesem Beispiel empfiehlt es sich, die neue Variable zur besseren Kontrolle vor die Variable f3a zu verschieben. Wechseln Sie in die **Variablenansicht** und markieren Sie den Zeilenkopf der Variable **f3_check**. Verschieben Sie nun diese Variable mit gedrückter linker Maustaste bis vor die Variable f3a und lassen dort die Maustaste wieder los.

	Name	Typ	Spaltenf...	Dezimal...	Variablenlabel	Wertelabels	Fehlende W...	Spalten	Ausrichtung	Messniveau	Rolle
11	f1_5	Numerisch	8	0	SPSS bietet qu...	{1, trifft voll ...	9	8	Rechts	Ordinal	Eingabe
12	f1_6	Numerisch	8	0	Die Software vo...	{1, trifft voll ...	9	8	Rechts	Ordinal	Eingabe
13	f1_7	Numerisch	8	0	Die Produkte vo...	{1, trifft voll ...	9	8	Rechts	Ordinal	Eingabe
14	f1_8	Numerisch	8	0	Die Produkte vo...	{1, trifft voll ...	9	8	Rechts	Ordinal	Eingabe
15	f1_9	Numerisch	8	0	Die Produkte vo...	{1, trifft voll ...	9	8	Rechts	Ordinal	Eingabe
16	f1_10	Numerisch	8	0	Ich werde über ...	{1, trifft voll ...	9	8	Rechts	Ordinal	Eingabe
17	f1_11	Numerisch	8	0	Die Preise der ...	{1, trifft voll ...	9	8	Rechts	Ordinal	Eingabe
18	f2_1	Numerisch	8	0	Lieferungen von...	{1, trifft voll ...	Keine	8	Rechts	Ordinal	Eingabe
19	f2_2	Numerisch	8	0	Lieferungen von...	{1, trifft voll ...	Keine	8	Rechts	Ordinal	Eingabe
20	f2_3	Numerisch	8	0	Bei Lieferproble...	{1, trifft voll ...	Keine	8	Rechts	Ordinal	Eingabe
21	f3_check	Numerisch	8	2		Keine	Keine	10	Rechts	Nominal	Eingabe
22	f3a	Numerisch	8	0	3.a) Hatten Sie ...	{1, Ja}...	Keine	8	Rechts	Ordinal	Eingabe
23	f3b_1	Numerisch	8	0	Die Verkaufsmi...	{1, trifft voll ...	9	8	Rechts	Ordinal	Eingabe
24	f3b_2	Numerisch	8	0	Die Verkaufsmi...	{1, trifft voll ...	9	8	Rechts	Ordinal	Eingabe
25	f3b_3	Numerisch	8	0	Die Verkaufsmi...	{1, trifft voll ...	9	8	Rechts	Ordinal	Eingabe
26	f3b_4	Numerisch	8	0	Die Verkaufsmi...	{1, trifft voll ...	9	8	Rechts	Ordinal	Eingabe
27	f3b_5	Numerisch	8	0	Die Verkaufsmi...	{1, trifft voll ...	9	8	Rechts	Ordinal	Eingabe
28	f3b_6	Numerisch	8	0	Die Verkaufsmi...	{1, trifft voll ...	9	8	Rechts	Ordinal	Eingabe
29	f3b_7	Numerisch	8	0	Die Verkaufsmi...	{1, trifft voll ...	9	8	Rechts	Ordinal	Eingabe
30	f4a	Numerisch	8	0	4.a) Hatten Sie ...	{1, Ja}...	Keine	8	Rechts	Ordinal	Eingabe
31	f4b_1	Numerisch	8	0	Ich bin mit dem...	{1, trifft voll ...	9	8	Rechts	Ordinal	Eingabe
32	f4b_2	Numerisch	8	0	Der technische ...	{1, trifft voll ...	9	8	Rechts	Ordinal	Eingabe
33	f4b_3	Numerisch	8	0	Meine Anrufe w...	{1, trifft voll ...	9	8	Rechts	Ordinal	Eingabe
34	f4b_4	Numerisch	8	0	Für besondere ...	{1, trifft voll ...	9	8	Rechts	Ordinal	Eingabe
35	f4b_5	Numerisch	8	0	Die Antworten, ...	{1, trifft voll ...	9	8	Rechts	Ordinal	Eingabe

Abbildung 55: Variablenansicht “f3_check”

Mit einem Doppelklick auf f3_check können Sie direkt in die Datenansicht zur entsprechenden Variable wechseln. Um einen schnellen Überblick über die inkonsistenten Fälle zu erhalten, **sortieren Sie f3_check in absteigender Reihenfolge** (so dass die mit “1” markierten Fälle ganz zuoberst angeordnet werden). Klicken Sie mit der **rechten Maustaste** auf f3_check und wählen **Absteigend sortieren**. Insgesamt befinden sich **vier inkonsistente Fälle** im Datenfile. Grundsätzlich liegt es in Ihrem Ermessen zu entscheiden, ob und wie diese Fälle bereinigt werden sollen.

Wird beispielsweise davon ausgegangen, dass diese vier Personen Frage f3a fälschlicherweise mit “Nein” beantwortet haben, können Sie bei diesen vier Fällen Frage f3a deshalb **manuell auf “Ja” (Code 1) setzen**. Die Daten werden nacheditiert.

	f3_check	f3a	f3b_1	f3b_2	f3b_3	f3b_4	f3b_5	f3b_6	f3b_7	f4a	f4b_1
1	1.00	Ja	trifft voll un...	trifft eher ni...	trifft eher zu	weder noch	trifft eher zu	trifft voll un...	trifft eher zu	Ja	trifft eher ni...
2	1.00	Ja	trifft voll un...	trifft voll un...	trifft voll un...	trifft voll un...	trifft voll un...	trifft voll un...	trifft voll un...	Ja	trifft voll un...
3	1.00	Nein	weder noch	weder noch	trifft eher zu	weder noch	trifft eher zu	trifft eher zu	trifft eher zu	Nein	.
4	1.00	Nein	trifft eher zu	trifft eher zu	trifft voll un...	trifft voll un...	trifft voll un...	trifft voll un...	trifft voll un...	Ja	trifft voll un...
5	.	Ja	trifft voll un...	trifft eher ni...	trifft voll un...	weder noch	trifft voll un...	trifft voll un...	weder noch	Nein	.
6	.	Ja	trifft voll un...	trifft voll un...	trifft voll un...	trifft voll un...	trifft voll un...	trifft voll un...	trifft voll un...	Nein	.
7	.	Ja	trifft voll un...	weder noch	trifft eher zu	trifft eher ni...	trifft voll un...	trifft eher zu	trifft eher ni...	Nein	.
8	.	Ja	trifft voll un...	trifft eher zu	trifft eher zu	trifft eher zu	trifft eher zu	trifft eher zu	trifft eher zu	Ja	trifft eher zu
9	.	Nein	.	.	.	.	.	.	.	Nein	.
10	.	Nein	.	.	.	.	.	.	.	Nein	.
11	.	Ja	trifft eher zu	trifft eher zu	trifft eher zu	trifft eher ni...	weder noch	trifft voll un...	trifft voll un...	Nein	.
12	.	Ja	trifft voll un...	weder noch	weiss nicht	weiss nicht	weder noch	trifft eher zu	trifft eher zu	Nein	.
13	.	Nein	.	.	.	.	.	.	.	Ja	trifft eher zu
14	.	Ja	trifft voll un...	trifft voll un...	trifft voll un...	trifft voll un...	trifft voll un...	trifft voll un...	trifft voll un...	Ja	trifft eher zu
15	.	Nein	.	.	.	.	.	.	.	Ja	weder noch
16	.	Ja	trifft voll un...	trifft eher zu	trifft voll un...	weder noch	trifft eher zu	trifft voll un...	trifft voll un...	Ja	weder noch
17	.	Ja	trifft eher zu	weder noch	trifft überha...	trifft eher ni...	trifft voll un...	trifft voll un...	weder noch	Ja	weder noch
18	.	Ja	trifft eher zu	trifft eher zu	trifft eher zu	trifft eher zu	trifft eher ni...	trifft eher zu	trifft eher ni...	Nein	.
19	.	Nein	.	.	.	.	.	.	.	Nein	.
20	.	Ja	trifft voll un...	trifft eher ni...	trifft eher zu	weder noch	trifft überha...	trifft überha...	trifft eher zu	Ja	weder noch
21	.	Nein	.	.	.	.	.	.	.	Nein	.
22	.	.	.	.	.	.	.	.	.	Ja	trifft eher zu
23	.	Ja	trifft eher zu	weder noch	trifft eher zu	trifft eher zu	weiss nicht	trifft eher ni...	trifft eher zu	Nein	.
24	.	Nein	.	.	.	.	.	.	.	Ja	trifft eher ni...

Abbildung 56: Datenansicht "f3_check und f3a"

Die Variable f3_check kann anschliessend wieder gelöscht werden; klicken Sie mit der rechten Maustaste auf den Spaltenkopf von f3_check und wählen **Löschen**. Speichern Sie das korrigierte Datenfile ab.

2. Mehrfachantworten-Sets

Bei Mehrfachantworten-Sets (MFAS) handelt es sich nicht um "Variablen" im üblichen Sinn. MFAS verwenden mehrere Variablen, um Antworten auf Fragen aufzuzeichnen, auf welche der Befragte mehr als eine Antwort geben kann. Sie werden wie kategoriale Variablen behandelt und bieten weitestgehend dieselben

Möglichkeiten wie kategoriale Variablen.⁶ Sie sind nur dann anwendbar, wenn für alle Variablen die gleiche Basis gilt.

In der SPSS Kundenbefragung konnten zu **Frage 6** “Welche Zusatzmodule von SPSS Statistics benutzen Sie?” mehrere Antworten gewählt werden. Im Datenfile sehen Sie, dass im SPSS Statistics für jede dieser Antwortmöglichkeiten eine separate Variable definiert ist, jeweils codiert mit “1” (Ja) und “0” (Nein). Diese Art von Kodierung wird als “dichotom” bezeichnet, da pro Variable jeweils nur zwei unterschiedliche Codes möglich sind (normalerweise “0” für Nein bzw. “nicht genannt” und “1” für Ja bzw. genannt).

	f6_1	f6_2	f6_3	f6_4	f6_5	f6_6	f6_7	f6_8	f6_9	f6_10	f6_11	f6_12	f6_13	f6_14	f6_15
1	ja	ja	nein	nein	ja	nein	ja	nein	ja	nein					
2	ja	ja	nein	ja	nein	nein	nein	nein	nein	nein	ja				
3	ja	ja	ja	nein	nein	ja	nein	ja	nein	ja					
4	ja	ja	nein	nein	nein	ja	nein	nein	nein	nein					
5	ja	ja	nein	nein	nein	ja	nein	ja	nein	nein					
6	ja	ja	nein	nein	nein	ja	nein	ja	ja	nein					
7	ja	ja	nein	nein	nein	ja	nein	nein	nein	nein					
8	ja	ja	nein	nein	nein	nein	ja	nein	nein	nein					
9	ja	nein	nein	nein	nein	nein	nein	nein	nein	nein					
10	ja	ja	nein	nein	nein	nein	ja	nein	nein	ja					
11	ja	ja	nein	ja	nein	nein	nein	ja	nein	nein					
12	ja	ja	nein	ja	ja	nein	nein	nein	nein	nein					
13	ja	ja	ja	nein	nein	nein	nein	ja	nein	nein					
14	ja	ja	nein	nein	nein	ja	nein	nein	nein	nein					
15	ja	ja	nein	nein	ja	ja	nein	nein	nein	nein	ja				
16	ja	ja	nein	ja	nein	ja	nein	ja	nein	nein					
17	ja	nein	nein	nein	nein	ja	nein	ja	nein	nein					
18	ja	nein	nein	nein	nein	ja	nein	nein	ja	nein					
19	ja	ja	ja	nein	nein	nein	nein	nein	nein	ja					
20	ja	ja	nein	nein	nein	ja	nein	nein	nein	nein					
21	ja	ja	nein	nein	nein	nein	ja	ja	nein	nein					
22	ja	ja	nein	nein	nein	ja	nein	ja	ja	ja					
23	ja	ja	ja	ja	nein	ja	ja	nein	nein	ja					
24	ja	nein	nein	nein	nein	nein	nein	ja	nein	nein					

Abbildung 57: Datenansicht “f6_1 bis f6_15”

Um MFAS zu definieren wählen Sie folgenden Befehl aus dem Menü aus:

Analysieren | Mehrfachantworten | Variablen-Sets definieren...

⁶ Um MFAS im Datenfile abzuspeichern, benötigen Sie das Zusatzmodul SPSS Statistics Tables! Mit dem Basismodul von SPSS Statistics lassen sich MFAS zwar definieren und auswerten, die Definitionen gehen jedoch verloren, sobald das Datenfile geschlossen wird.

Wählen Sie die Variablen **f6_1 bis f6_15** aus. Markieren Sie die Option **Dichotomien** und tragen unter **Gezählter Wert** Code “1” ein. Geben Sie als Namen für dieses MFAS “f6” ein, als Beschriftung “6. Welche Zusatzmodule benutzen Sie?” und klicken auf **Hinzufügen** (dem Set-Namen wird automatisch ein Dollarzeichen vorangestellt) und dann auf **Schliessen**.

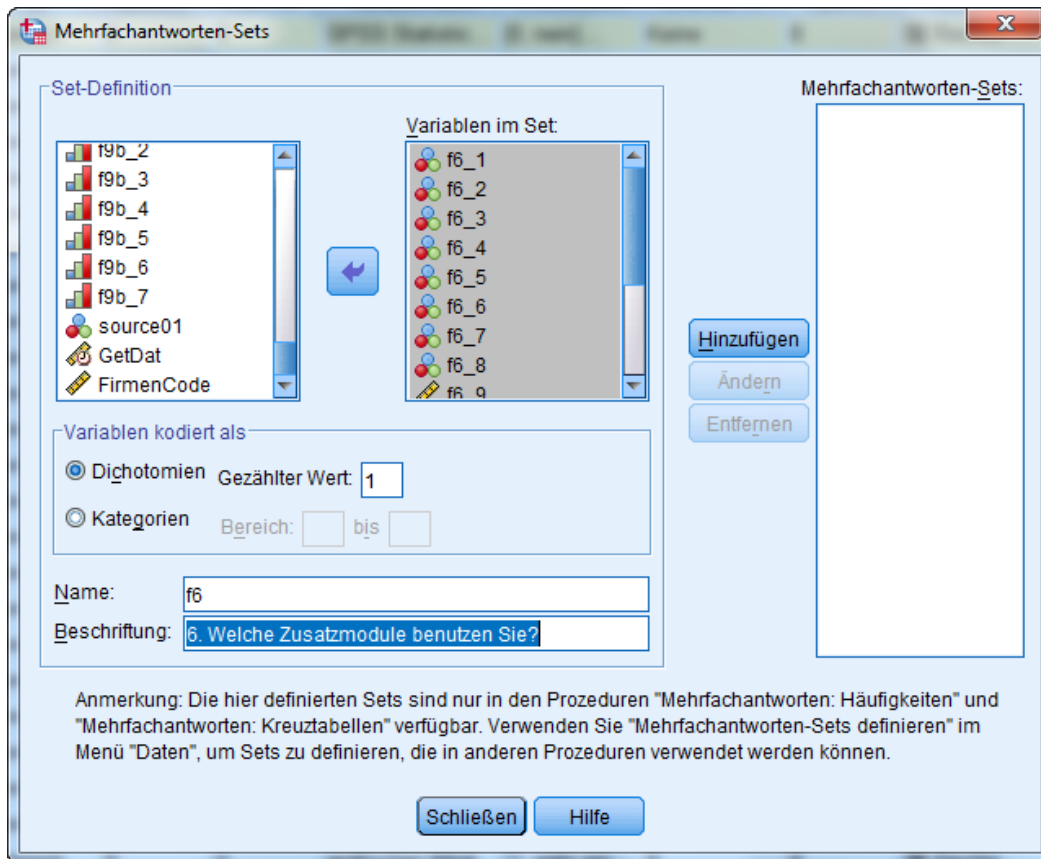


Abbildung 58: Mehrfachantworten-Sets

Um eine Häufigkeitsauszählung über ein MFAS zu erstellen, wählen Sie

Analysieren | Mehrfachantworten | Häufigkeiten...

Verschieben Sie Set **\$f6** nach **Tabelle(n) für** und klicken auf **OK**. Im View Fenster erscheinen die Häufigkeiten von \$f6.

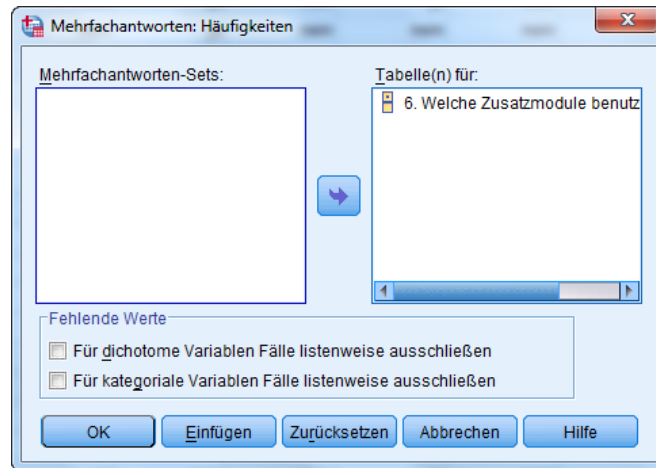


Abbildung 59: Ansicht "Mehrfachantworten: Häufigkeiten"

Häufigkeiten von \$f6				
		Antworten		Prozent der Fälle
		N	Prozent	
6. Welche Zusatz- module benutzen Sie? ^a	SPSS Statistics Base	1424	18.4%	99.7%
	SPSS Advanced Statistics	1051	13.6%	73.6%
	SPSS Categories	528	6.8%	37.0%
	SPSS Complex Samples	276	3.6%	19.3%
	SPSS Conjoint	299	3.9%	20.9%
	SPSS Custom Tables	1010	13.0%	70.7%
	SPSS Data Preparation	418	5.4%	29.3%
	SPSS Decision Trees	544	7.0%	38.1%
	SPSS Direct Marketing	211	2.7%	14.8%
	SPSS Exact Tests	488	6.3%	34.2%
	SPSS Forecasting	316	4.1%	22.1%
	SPSS Missing Value Analysis	625	8.1%	43.8%
	SPSS Neural Networks	94	1.2%	6.6%
	SPSS Regression	425	5.5%	29.8%
	f6_15	33	.4%	2.3%
Gesamt		7742	100.0%	542.2%

a. Dichotomie-Gruppe tabellarisch dargestellt bei Wert 1.

Abbildung 60: Viewerfenster "Häufigkeiten von \$f6"

3. Fälle auswählen

Fälle können anhand einer logischen Bedingung nach bestimmten Kriterien ausgewählt werden. Soll sich eine Auswertung nur auf eine Teilmenge der erfassten Fälle beziehen, bietet SPSS Statistics verschiedene Möglichkeiten, um Fälle auszuwählen. Verwenden Sie dazu den Menüpunkt

Daten | Fälle auswählen

Im Dialogfeld markieren Sie unter Auswählen die Option **Falls Bedingung zutrifft**. Klicken Sie dann die Schaltfläche **Falls...**.

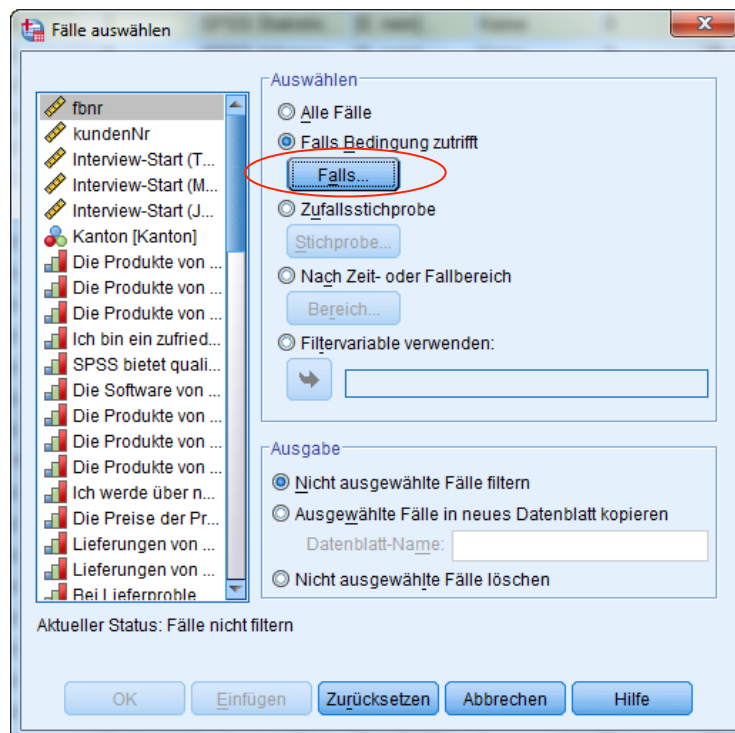


Abbildung 61: Ansicht "Fälle auswählen"

Im Dialogfeld **Fälle auswählen**: **Falls** wird die Bedingung eingegeben, nach der die Fallauswahl erfolgen soll. Es sollen **alle Kunden ausgewählt werden, die in einer Bank oder in einer Versicherung tätig sind und mindestens 1 x pro Monat mit SPSS Statistics arbeiten**. Für den ersten Teil der Bedingung (Tätigkeitsbereich) soll die Funktion **ANY** verwendet werden; diese prüft, ob der Test-

wert einem der nachfolgenden Werte entspricht. Markieren Sie unter Funktionsgruppe **Alle** und dann unter Funktionen und Sondervariablen **ANY** (**test,wert,wert,...**) und klicken Sie dann auf  , um diese Funktion in den oberen freien Bereich zu verschieben.

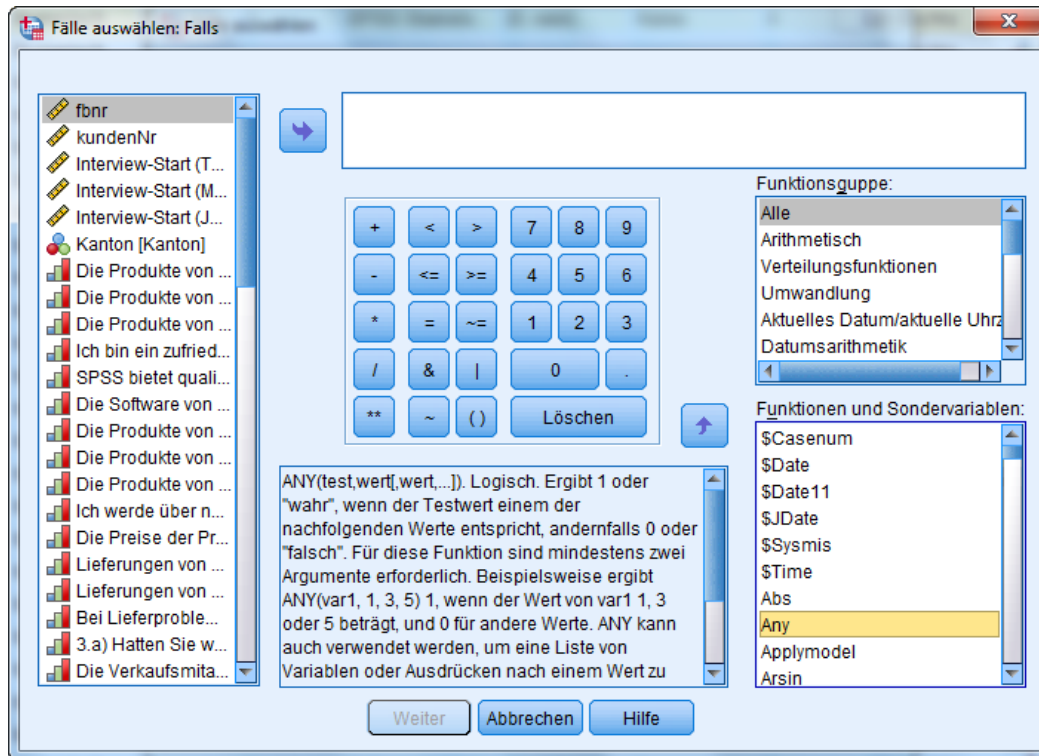


Abbildung 62: Ansicht "Fälle auswählen: Falls (1)"

Als **Testvariable** (1. Argument) wählen Sie die **Variable f5**, als **zu überprüfende Werte** **"2"** und **"3"**. Die erste Bedingung lautet demnach wie folgt:

ANY(f5,2,3)

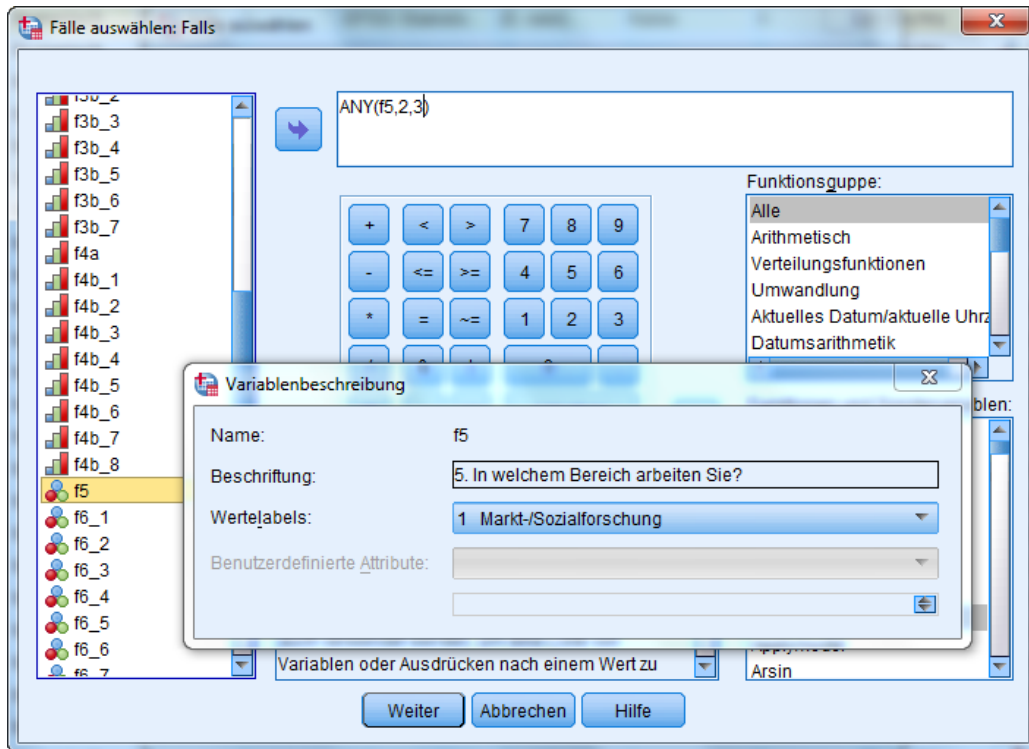


Abbildung 63: Ansicht "Fälle auswählen: Falls (2)"

Um die Kodierung der **Variablen f8** ("Wie oft arbeiten Sie mit SPSS Statistics?") zu erhalten, klicken Sie in der Variablenliste des Dialogfensters mit der rechten Maustaste auf diese Variable. Hier wählen Sie die Option **Variablenbeschreibung** und es erscheint ein Fenster, das Sie über die Labels der Werte und der Variablen informiert. Für den zweiten Teil der Bedingung sollen somit all jene Fälle berücksichtigt werden, welche **bei der Variable f8 einen Wert grösser oder gleich 4** haben.

Übertragen Sie die Variable **f8** in das Formelfeld, wählen Sie dann auf dem Taschenrechner des Dialogfeldes oder von der Tastatur **>=** und geben Sie **4** ein. Um diesen Teil der Bedingung mit dem ersten Teil zu kombinieren, setzen Sie den Cursor vor f8 und klicken auf **&** (d.h. beide Teile der Bedingung müssen erfüllt sein, damit ein Fall ausgewählt wird). Die vollständige Bedingung lautet somit:

ANY(f5,2,3) & f8>=4

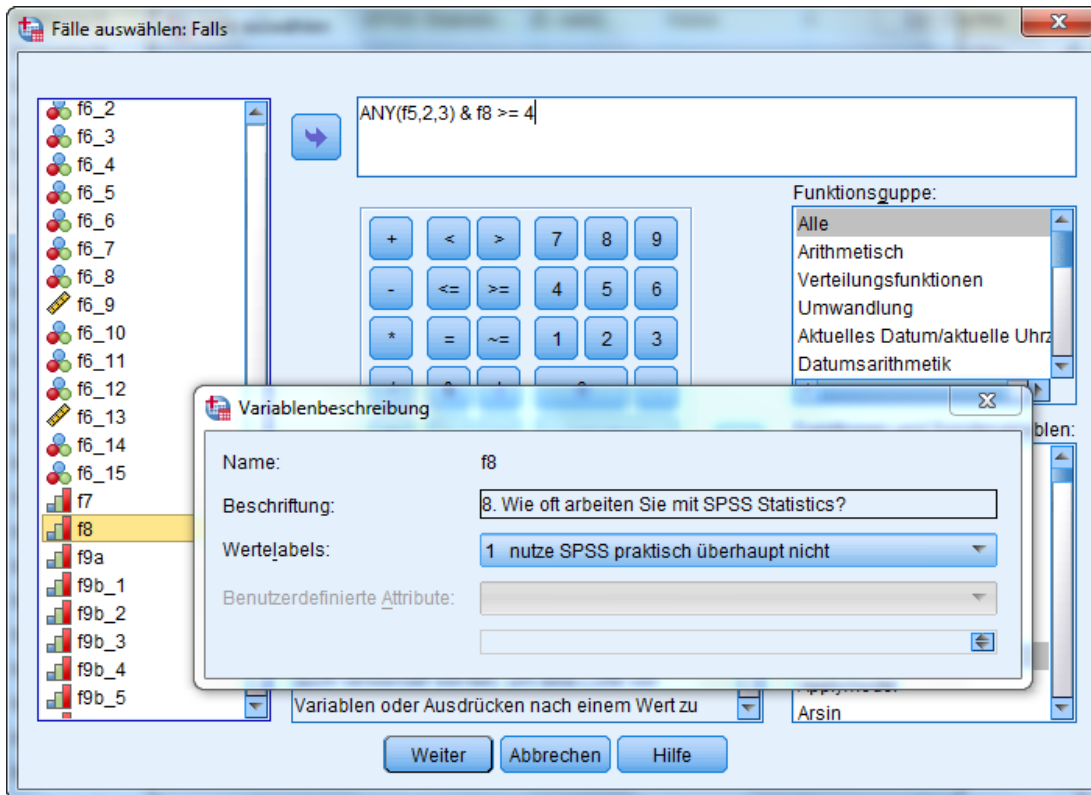


Abbildung 64: Ansicht "Fälle auswählen: Falls (3)"

Klicken Sie auf **Weiter** und **OK**, um den Filter zu setzen. SPSS Statistics erzeugt eine **Filtervariable** mit Namen **filter_\$**, die angibt, welche Fälle die Filterbedingung erfüllen und für weitere Analysen herangezogen werden. Ausserdem wird die Fallnummer der nicht ausgewählten Fälle durchgestrichen und in der Statuszeile angezeigt, dass ein Filter aktiv ist.

1 : fbnr 29 Sichtbar: 71 von 71 Variablen

	fbnr	kundenNr	int_tag	int_monat	int_jahr	Kanton	f1_1	f1_2	f1_3
1	1	1162	8	10	2008	12	3	1	2
2	1	1162	14	10	2009	12	2	2	2
3	2	3586	6	10	2008	4	2	1	3
4	2	3586	7	11	2009	4	1	4	4
5	3	2417	11	10	2008	1	2	1	2
6	3	5089	28	10	2009	23	2	1	3
7	4	2417	2	10	2009	1	1	1	1
8	4	1295	21	10	2008	6	2	3	4
9	5	1295	14	10	2009	6	2	2	3
10	5	2879	26	10	2008	19	2	3	3
11	6	3705	23	10	2008	16	3	2	3
12	6	5815	13	11	2009	23	1	1	2
13	7	438	25	10	2008	21	3	4	2
14	7	2879	24	11	2009	19	4	3	3
15	8	859	12	11	2008	1	3	2	3
16	8	3705	21	10	2009	16	2	2	2
17	9	438	3	11	2009	21	2	2	2
18	9	1174	30	11	2008	9	3	3	4
19	10	358	14	10	2008	15	1	2	3
20	10	859	21	11	2009	1	4	5	3
21	11	1174	25	11	2009	9	1	1	2
22	11	1278	14	10	2008	20	5	1	3
23	12	358	30	11	2009	15	2	3	4
24	12	2177	22	11	2008	20	2	2	4

Datenansicht Variablenansicht

IBM SPSS Statistics Prozessor ist bereit Filter aktiv

Abbildung 65: Datenansicht der ausgewählten Fälle

Sie können die Fallauswahl jederzeit wieder rückgängig machen, indem Sie im Dialogfeld Fälle auswählen die Option **Alle Fälle** markieren. Wenn Sie eine Filtervariable speichern möchten, empfiehlt es sich, den Namen der Variablen zu ändern. Sobald Sie das Datenblatt abspeichern, wird auch die Filtervariable gespeichert. Bereits vorhandene Filtervariablen können unter

Daten | Fälle auswählen

Filtervariable verwenden wieder aktiviert werden.⁷ Falls Sie die nicht ausgewählten Fälle nicht mehr benötigen, wählen Sie im Dialogfeld **Nicht ausgewählte Fälle** statt Filtern die Alternative **Löschen**. Beachten Sie dabei, dass diese Fälle verloren gehen, sobald Sie das Datenblatt unter dem gleichen Dateinamen speichern.

⁷ Beim Arbeiten mit den logischen Operatoren & (und) und | (oder) ist darauf zu achten, dass & stärker bindet als |, d.h. zum Teil sind Klammern erforderlich.

Um zu sehen, wie zufrieden die ausgewählte Gruppe ganz allgemein mit SPSS Statistics ist, gehen Sie zu

Analysieren | Deskriptive Statistiken | Häufigkeiten

und erstellen Sie eine Häufigkeitstabelle der Variable **f1_4** ("Ich bin ein zufriedener SPSS Kunde").

Ich bin ein zufriedener SPSS Kunde					
		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	trifft voll und ganz zu	75	29.3	29.4	29.4
	trifft eher zu	86	33.6	33.7	63.1
	weder noch	63	24.6	24.7	87.8
	trifft eher nicht zu	25	9.8	9.8	97.6
	trifft überhaupt nicht zu	6	2.3	2.4	100.0
	Gesamt	255	99.6	100.0	
Fehlend	weiss nicht	1	.4		
Gesamt		256	100.0		

Abbildung 66: View Fenster "Häufigkeiten f1_4"

Insgesamt haben 256 Personen die Filter-Bedingung erfüllt; 63% dieser Kunden bezeichnen sich dabei als zufriedene oder sehr zufriedene SPSS Kunden.

4. Werte umkodieren

Die meisten Werte der SPSS Kundenbefragung sind auf einer Skala von 1 = trifft voll und ganz zu und 5 = trifft überhaupt nicht zu codiert. Bei bipolaren Skalen (von "sehr schlecht" bis "sehr gut") wird zudem häufig gewünscht, dass diese Bipolarität auch durch positive bzw. negative Mittelwerte dargestellt wird; je negativer ein Mittelwert, desto schlechter, je positiver der Mittelwert, desto besser die Bewertung einer Aussage. Ein Mittelwert nahe bei 0 steht in diesem Fall für eine indifferente Bewertung ("weder noch"). Sämtliche Aussagen der SPSS Kunden-

befragung, welche auf einer Skala von 1 bis 5 bewertet werden mussten, sollen im Folgenden so umkodiert werden, dass

Trifft voll und ganz zu	->	Code 2
Trifft eher zu	->	Code 1
Weder noch	->	Code 0
Trifft eher nicht zu	->	Code -1
Trifft überhaupt nicht zu	->	Code -2

erhält. Wählen Sie aus dem Menü

Transformieren | Umkodieren in dieselben Variablen...

Im Dialogfeld **Umkodieren in dieselben Variablen** markieren Sie in der Variablenliste die Variable f1_1 bis f1_11, f2_1 bis f2_3, f3b_1 bis f3b_7, f4b_1 bis f4b_8 und f9b_1 bis f9b_7 und übertragen sie in das Feld **Numerische Variablen**. Anschliessend klicken Sie auf die Schaltfläche **Alte und neue Werte**, um die Rekodierung festzulegen. Im Dialogfeld **Umkodieren in dieselben Variablen: Alte und neue Werte** klicken Sie im Feld **Alter Wert** die Option **Wert** an und geben den Wert **1** ein. Im Feld **Neuer Wert** klicken Sie den Punkt **Wert** an und geben **2** ein und klicken dann auf **Hinzufügen**. Entsprechend der Abbildung unten nehmen Sie die weiteren Wertzuweisungen vor. Klicken Sie anschliessend auf **Weiter** und **OK**.

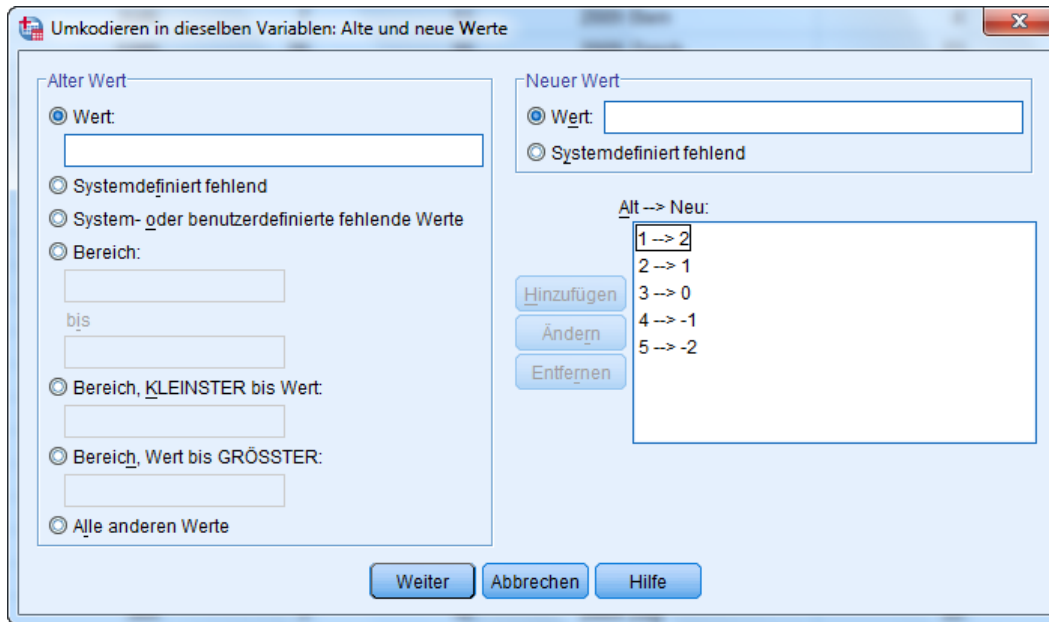


Abbildung 67: Ansicht "Umkodieren in dieselben Variablen: Alte und neue Werte"

Wichtig ist, dass anschliessend die Werte-Labels der rekodierten Variablen angepasst werden. Gehen Sie in der **Variablenansicht** des Daten-Editors zur Variable **f1_1** und klicken dort auf **Wertelabels**. Entfernen Sie zuerst alle alten Labels und setzen neue gemäss neuer Kodierung. Klicken Sie dann auf **OK**.

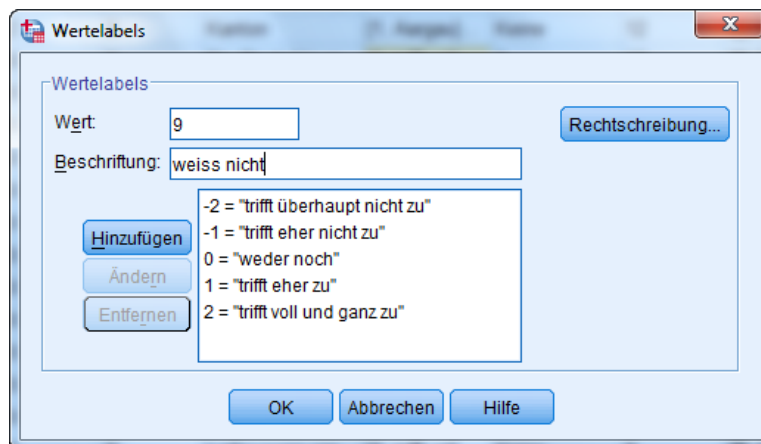


Abbildung 68: Ansicht "Wertelabels (1)"

Kopieren Sie nun die neuen Wertelabels von f1_1 und übertragen Sie diese auf die Variablen f1_2 bis f1_11, f2_1 bis f2_3, f3b_1 bis f3b_7, f4b_1 bis f4b_8. Passen Sie anschliessend auch die Labels für f9b_1 bis f9b_7 noch entsprechend an (vgl. nachfolgende Abbildung).

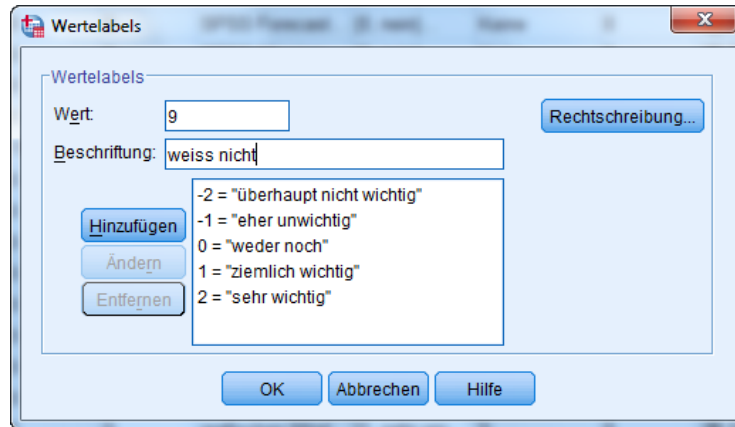


Abbildung 69: Ansicht "Wertelabels (2)"

Speichern Sie dann das File unter **SPSS Kundenbefragung 2009_rec.sav** neu ab.

5. Werte berechnen

Mithilfe von SPSS Statistics können anhand verschiedener Funktionen neue Variablen berechnet werden.

Numerische Funktionen

Im Folgenden soll ein Gesamtzufriedenheits-Index aus den Variablen f1_1 (Die Produkte von SPSS sind ein wichtiger Teil meiner täglichen Arbeit), f1_2 (Die Produkte von SPSS erleichtern meine Arbeit) und f1_4 (Ich bin ein zufriedener SPSS Kunde) berechnet werden. Im Folgenden soll der Mittelwert berechnet werden:

- *Arithmetisches Mittel*

Um eine neue Variable zu berechnen, wählen Sie

Transformieren | Variable berechnen.

Im Dialogfeld Variable berechnen tippen Sie als Namen für die Zielvariable **satis_1** ein. Ein arithmetisches Mittel errechnet sich aus der Summe der Werte dividiert durch die Anzahl der Werte. Der Ausdruck für die Berechnung des Gesamtzufriedenheits-Index lautet somit:

$$(f1_1 + f1_2 + f1_4) / 3$$

Geben Sie diesen Ausdruck im Feld **Numerischer Ausdruck** ein.

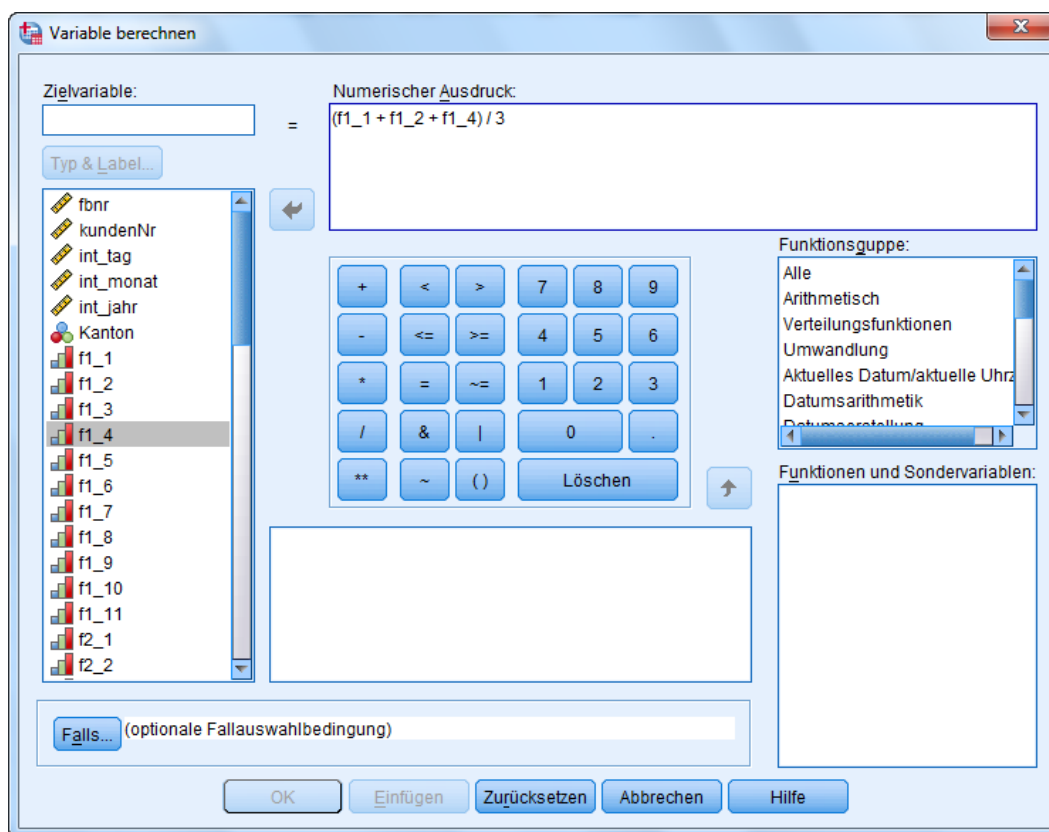


Abbildung 70: Ansicht "Variable berechnen"

Klicken Sie nun auf die Schaltfläche **OK**. Die neue Variable satis_1 wird ganz rechts an Ihre Datendatei angehängt. Verschieben Sie sie in der Variablenansicht vor die Variable f1_1.

Die Formel ist an sich korrekt, sie hat jedoch einen gewichtigen Nachteil: Sobald eine der beteiligten Variablen einen **Missing-Wert** aufweist (z.B. weiss nicht), kann die Berechnung für die neue Variable nicht durchgeführt werden, sie erhält somit ebenfalls ein Missing. In diesem Beispiel konnte deshalb für 54 Fälle kein Index-Wert berechnet werden (satis_1 in der Datenansicht aufsteigend sortieren).

- *Berechnungen mit der MEAN-Funktion*

Wählen Sie wieder

Transformieren | Variable berechnen

Als Namen für die Zielvariable geben Sie **satis_2** ein. Gehen Sie zur Funktionsgruppe **Statistisch** und wählen dort unter **Funktionen und Sondervariablen** die Funktion **Mean** an. Im linken Fenster finden Sie eine kurze Erläuterung zur ausgewählten Funktion. Verschieben Sie die Funktion Mean mit der Pfeil-Taste in den Bereich **Numerischer Ausdruck** und passen Sie sie wie folgt an:

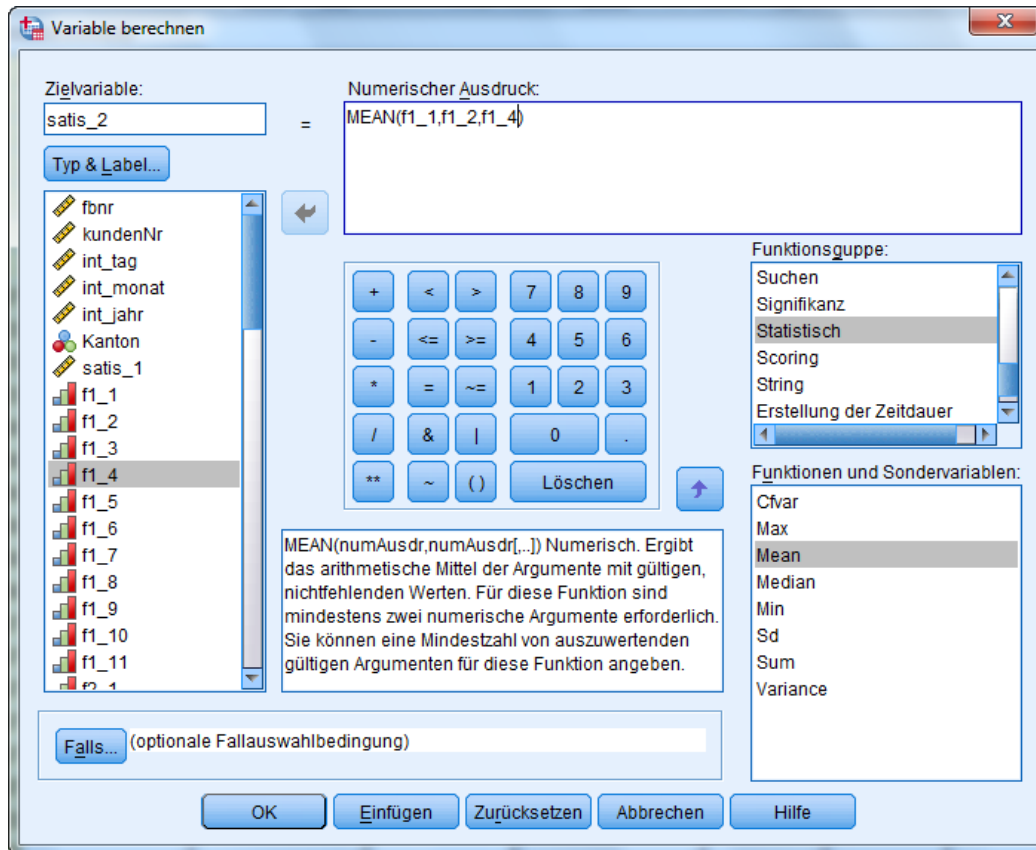


Abbildung 71: Ansicht "Variable berechnen"

Klicken Sie auf **OK** und verschieben Sie die neue Variable `satis_2` zu `satis_1`. Wie Sie sehen, wurde nun für sämtliche Fälle ein Index-Wert berechnet.

Oftmals wollen Sie jedoch einen Mittelwert nur rechnen, falls eine bestimmte Anzahl der beteiligten Variablen einen gültigen Wert ausweisen. Für die SPSS Kundenbefragung **soll z.B. nur dann der Gesamtzufriedenheits-Index berechnet werden, wenn zumindest 2 der 3 Variablen einen gültigen Wert haben**; falls nur 1 Aussage bewertet wurde, soll der Index-Wert jedoch auf Missing gesetzt werden. Sie können dies erreichen, indem Sie die obige Formel wie folgt anpassen:

$$\text{satis_3} = \text{MEAN.2}(\text{f1_1}, \text{f1_2}, \text{f1_4})$$

- *Zählen von bestimmten Codes*

Mit der Prozedur **Zählen...** kann eine Variable erstellt werden, mit welcher das Auftreten bestimmter Werte in einer Variablenliste pro Fall gezählt wird. Für die Auswertung der SPSS-Kundenbefragung ist z.B. von Interesse, **wie viele Zusatzmodule von SPSS Statistics benutzt werden**. Die Prozedur Zählen... finden Sie unter

Transformieren | Werte in Fällen zählen...

Geben Sie als Zielvariablen-Name **anz_module** ein, als Label **Anzahl benutzter Statistics Module**. Markieren Sie nun die Variablen **f6_1 bis f6_15** und verschieben diese in den Bereich **Numerische Variablen**. Klicken Sie anschließend auf **Werte definieren...**

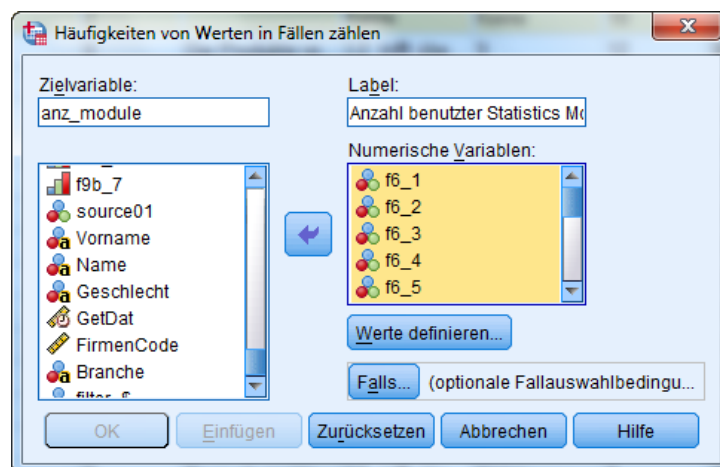


Abbildung 72: Ansicht "Häufigkeiten von Werten in Fällen zählen"

Da die Antwortmöglichkeiten zu Frage 6 dichotom mit 0 (Nein) bzw. 1 (Ja) codiert sind, tragen Sie nach der Auswahl von "Werte definieren" unter Wert Code 1 ein und klicken auf **Hinzufügen**.

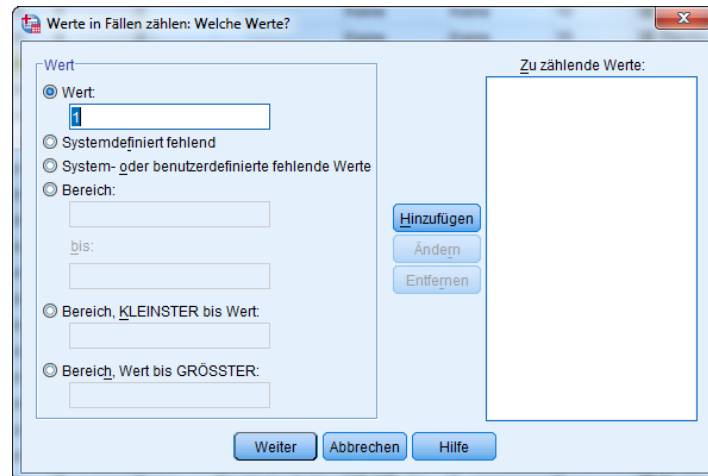


Abbildung 73: Ansicht “Werte in Fällen zählen: Welche Werte?”

Lassen Sie die Prozedur mit **Weiter** und **OK** laufen. Die neue Variable **anz_module** finden Sie wieder am Ende Ihres Datenfiles. Verschieben Sie diese zur besseren Kontrolle vor die Variable **f6_1**.

*SPSS Kundenbefragung_rec.sav [DatenSet1] - IBM SPSS Statistics Daten-Editor

Datei Bearbeiten Ansicht Daten Transformieren Analysieren Direktmarketing Diagramme Extras Fenster Hilfe

Sichtbar: 75 von 75 Variablen

	anz_module	f6_1	f6_2	f6_3	f6_4	f6_5	f6_6	f6_7	f6_8	f6_9
1	6	ja	ja	ja	nein	nein	ja	nein	ja	ja
2	4	ja	nein	ja	nein	nein	nein	ja	nein	nein
3	3	ja	nein	nein	nein	nein	ja	ja	nein	nein
4	3	ja	nein	nein	nein	nein	ja	nein	nein	nein
5	4	ja	nein	nein	nein	nein	ja	nein	ja	nein
6	3	ja	ja	nein	nein	ja	nein	nein	nein	nein
7	5	ja	ja	ja	nein	nein	nein	nein	ja	nein
8	4	ja	nein	nein	nein	nein	nein	nein	ja	nein
9	6	ja	ja	ja	ja	nein	ja	nein	nein	nein
10	4	ja	ja	nein	nein	nein	ja	ja	nein	nein
11	7	ja	nein	ja	nein	nein	ja	ja	ja	nein
12	4	ja	ja	nein	nein	nein	nein	nein	ja	nein
13	3	ja	nein	ja	nein	nein	nein	nein	nein	nein
14	3	ja	nein	nein	nein	nein	ja	nein	ja	nein
15	4	ja	ja	nein	nein	nein	ja	nein	nein	nein
16	5	ja	ja	nein	ja	nein	ja	nein	nein	nein
17	3	ja	nein	nein	nein	nein	nein	nein	nein	nein
18	4	ja	ja	nein	nein	nein	nein	nein	ja	nein
19	5	ja	ja	nein	nein	nein	ja	nein	nein	nein
20	7	ja	ja	nein	nein	nein	ja	ja	ja	nein
21	3	ja	nein	nein	nein	nein	ja	nein	nein	nein
22	3	ja	ja	nein	nein	ja	nein	nein	nein	nein
23	5	ja	nein	ja	ja	ja	ja	nein	nein	nein
24	6	ja	ja	ja	nein	nein	ja	nein	ja	nein

Datenansicht Variablenansicht

IBM SPSS Statistics Prozessor ist bereit

Abbildung 74: Variablenansicht

Berechnungen mit Datumsvariablen

In SPSS steht ein Assistent für Datum und Uhrzeit zur Verfügung. Als Erstes soll aus den Variablen `int_tag` (Tag des Interviews), `int_monat` (Monat des Interviews) und `int_jahr` (Jahr des Interviews) eine Datumsvariable erstellt werden. Öffnen Sie den Assistenten für Datum und Uhrzeit über

Transformieren | Assistent für Datum und Uhrzeit...

Es soll eine **neue Datumsvariable** aus bestehenden Datums-Elementen erstellt werden. Wählen Sie dazu im Willkommens-Fenster die 3. Option **Eine Datums-/Zeitvariable aus seinen Variablen erstellen, in der Teile von Datums- und Uhrzeitangaben enthalten sind**.

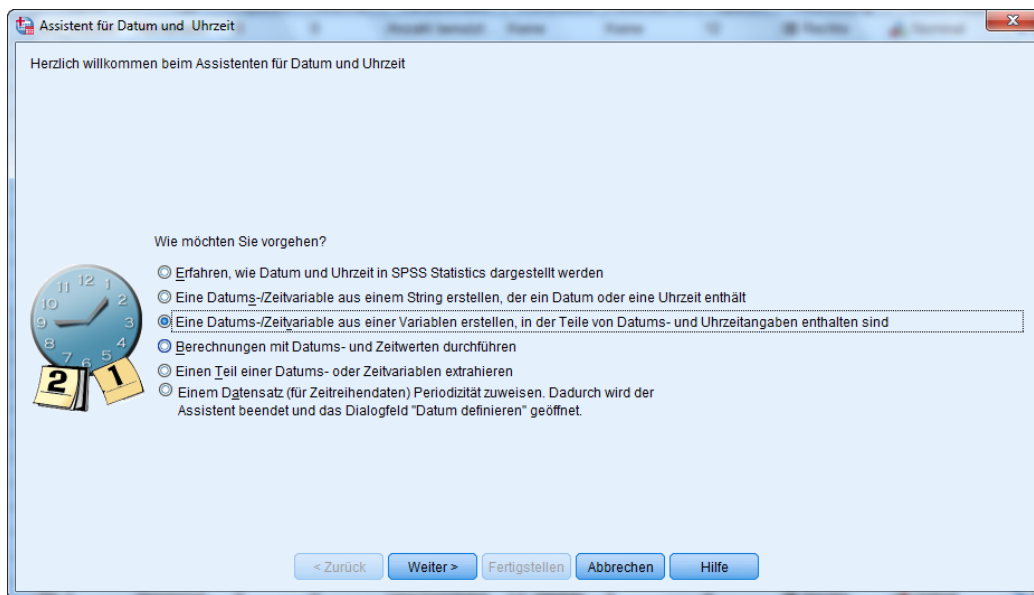


Abbildung 75: Ansicht "Assistent für Datum und Uhrzeit"

Verschieben Sie im folgenden Fenster die Variable `int_jahr` ins Feld Jahr, `int_monat` ins Feld Monat und `int_tag` ins Feld Tag des Monats und klicken dann auf **Weiter**.

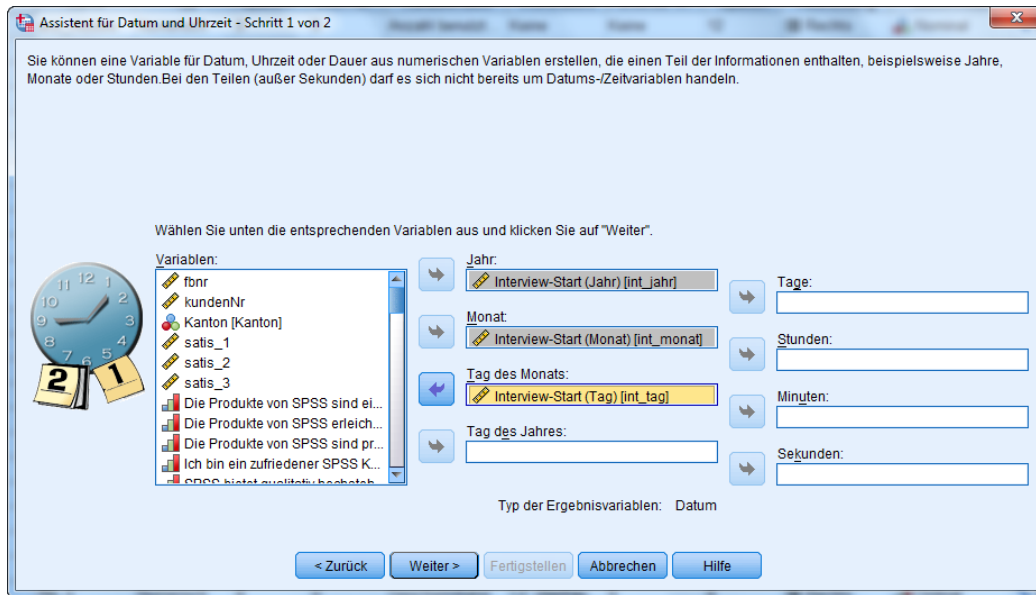


Abbildung 76: Ansicht "Assistent für Datum und Uhrzeit – Schritt 1 von 2"

Geben Sie im nächsten Fenster als Ergebnisvariable **int_dat** ein, als Variablenlabel **Datum des Interviews** und wählen **tt.mm.jjjj** als Ausgabeformat. Klicken Sie auf **Fertig stellen**.

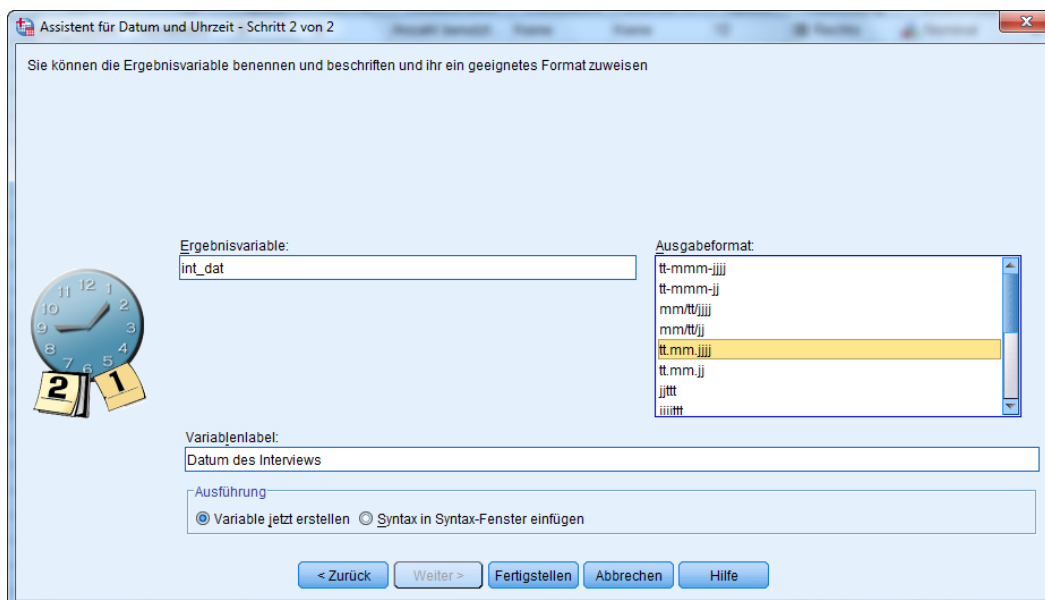


Abbildung 77: Ansicht "Assistent für Datum und Uhrzeit – Schritt 2 von 2"

Verschieben Sie die neu erstellte Variable **int_dat** vor die Variable **int_tag** und vergleichen Sie das Ergebnis mit den Input-Variablen.

	int_dat	int_tag	int_monat	int_jahr	Kanton	satis_1	satis_2	satis_3	f1_1
1	07.10.2009	7	10	2009	Schwyz	.	1.00	1.00	trifft eher zu
2	20.11.2008	20	11	2008	Schwyz	.	.00	.00	trifft eher nicht zu
3	14.11.2009	14	11	2009	Bern	.	1.00	1.00	trifft eher zu
4	02.10.2008	2	10	2008	Luzern	.	.50	.50	weiss nicht
5	04.10.2008	4	10	2008	Zürich	.	.00	.00	weder noch
6	06.10.2008	6	10	2008	Bern	.	1.50	1.50	trifft eher zu
7	21.10.2008	21	10	2008	Zürich	.	.50	.50	trifft eher zu
8	25.10.2008	25	10	2008	Waadt	.	-.50	-.50	weder noch
9	21.10.2008	21	10	2008	Aargau	.	1.00	.	weiss nicht
10	18.10.2009	18	10	2009	Basel (Stadt)	.	.00	.	weiss nicht
11	23.11.2008	23	11	2008	Zürich	.	.50	.50	trifft eher zu
12	14.11.2008	14	11	2008	Luzern	.	.50	.50	weiss nicht
13	14.10.2008	14	10	2008	Basel (Land)	.	1.00	1.00	weiss nicht
14	28.10.2008	28	10	2008	Zürich	.	1.50	1.50	weiss nicht
15	24.10.2009	24	10	2009	Glarus	.	.00	.00	weder noch
16	08.11.2008	8	11	2008	Graubünden	.	.50	.50	weiss nicht
17	01.11.2008	1	11	2008	Zürich	.	.00	.	weiss nicht
18	24.11.2008	24	11	2008	Basel (Stadt)	.	.00	.00	weiss nicht
19	28.10.2008	28	10	2008	Bern	.	.50	.50	trifft eher nicht zu
20	22.10.2008	22	10	2008	Basel (Land)	.	1.50	1.50	trifft voll und ganz
21	04.10.2008	4	10	2008	Zürich	.	1.00	1.00	weder noch
22	28.11.2008	28	11	2008	St. Gallen	.	.50	.50	trifft eher zu
23	28.11.2009	28	11	2009	Zürich	.	1.00	1.00	trifft eher zu
24	22.11.2008	22	11	2008	Thurgau	.	-.50	-.50	trifft eher nicht zu

Abbildung 78: Datenansicht der Datumsvariablen

Als nächstes soll noch das Alter zum Zeitpunkt des Interviews berechnet werden. Dazu verwenden Sie die Variable **GebDat** mit dem Geburtsdatum der befragten Person sowie die neu erstellte Variable **int_dat** mit dem Interview-Datum. Öffnen Sie wieder den Datums-Assistenten über

Transformieren | Assistent für Datum und Uhrzeit...

und wählen die 4. Option **Berechnungen mit Datums- und Zeitwerten durchführen**. Markieren Sie im nachfolgenden Fenster die Option **Berechnen der Anzahl der Zeiteinheiten zwischen zwei Datumswerten** und klicken auf **Weiter**. Verschieben Sie die Variable **int_dat** zu Datum 1 und **GebDat** zu minus Datum2. Als Einheit wählen Sie **Jahre** und klicken auf **Weiter**.

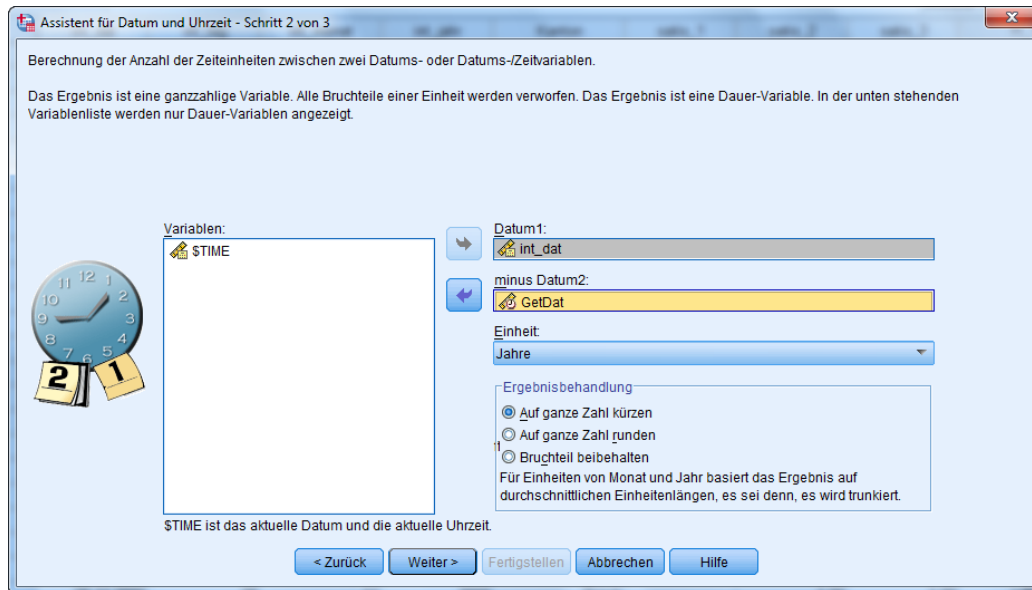


Abbildung 79: Ansicht "Assistent für Datum und Uhrzeit – Schritt 2 von 3"

Geben Sie als Name für die Ergebnisvariable **alter**, als Variablenlabel **Alter** ein.

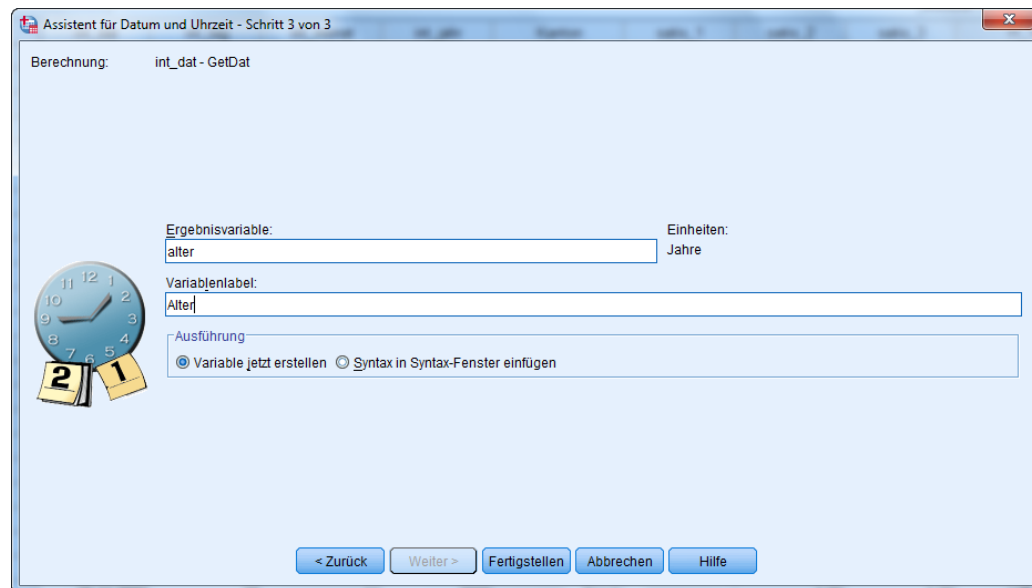


Abbildung 80: Ansicht "Assistent für Datum und Uhrzeit – Schritt 3 von 3"

Klicken Sie auf **Fertig stellen**. Die neue Variable **alter** finden Sie am Ende des Datenfiles in der Variablenansicht.

String-Funktionen

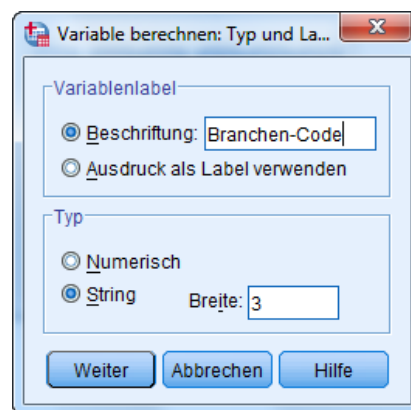
Für Datentransformationen mit Text-Variablen stehen ebenfalls nützliche Funktionen zur Verfügung.

- **SUBSTR**

Diese Funktion ist nützlich, wenn Sie einen **Teilstring aus einer Text-Variable extrahieren** möchten. Die Variable **Branche** enthält z.B. zwei Informationen: Die ersten drei Zeichen enthalten den Branchencode, das letzte Zeichen enthält eine Information über die Grösse des jeweiligen Betriebes. Diese beiden Informationen sollen im Folgenden in zwei unterschiedlichen Variablen abgespeichert werden. Wählen Sie

Transformieren | Variable berechnen

um eine Variable zu erstellen. Geben Sie als Namen für die Zielvariable **Bran**
chenCode ein. Klicken Sie auf **Typ&Label**. Da eine Text-Variable erstellt werden soll, markieren Sie unter Typ die Option **String**; setzen Sie die Breite auf **3 Zeichen**. Geben Sie schliesslich als Label **Branchen-Code** ein und klicken Sie auf **Weiter**.



In der Funktionsgruppe **String** doppelklicken Sie die Funktion **CHAR.SUBSTR(3)**. Diese Funktion benötigt 3 Argumente: 1) Die Text-Variable, aus der ein Teilstring extrahiert werden soll, 2) die Startposition, sowie 3) die

Länge des Teilstrings. Da Sie die ersten drei Zeichen aus der **Variable Branche** extrahieren sollen, lautet der korrekte String-Ausdruck:

BranchenCode = CHAR.SUBSTR(Branche, 1, 3)

Klicken Sie dann auf **OK**.

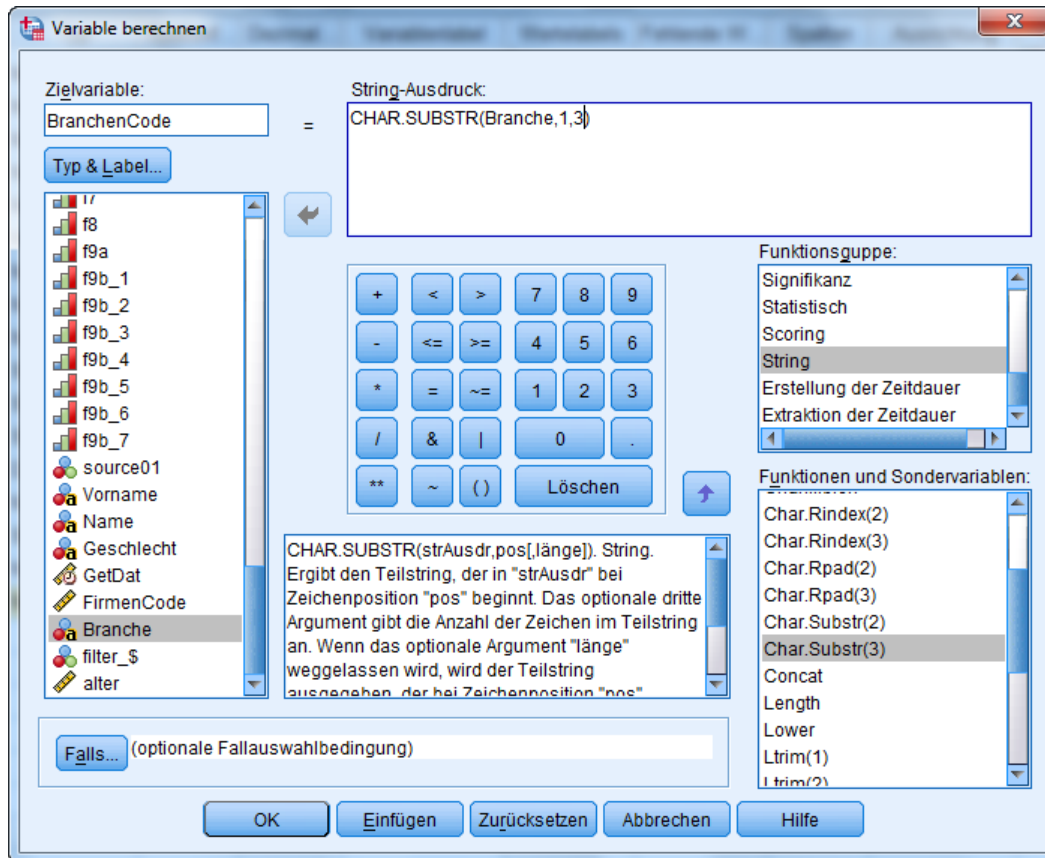


Abbildung 81: Ansicht "Variablen berechnen – BranchenCode"

Extrahieren Sie nun analog die Informationen zur Betriebsgröße (5. Zeichen der Variable Branche): Als Zielvariable wählen Sie **BetriebsGroesse**, definieren Sie unter **Typ&Label** diese Variable als **String** mit **Breite 1** und passen Sie den String-Ausdruck wie folgt an:

BetriebsGroesse = CHAR.SUBSTR(Branche, 5, 1)

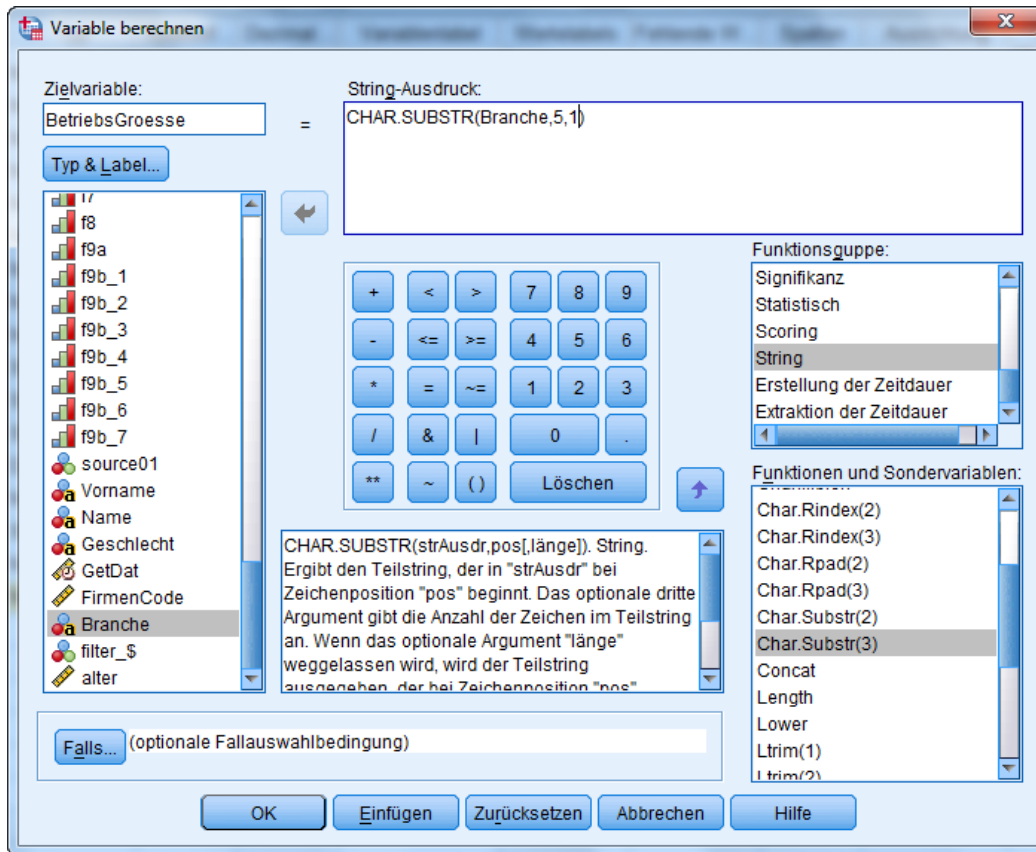


Abbildung 82: Ansicht "Variablen berechnen – BetriebsGroesse"

Klicken Sie auf **OK** und kontrollieren Sie das Ergebnis im Daten-Editor.

	Branche	alter	BranchenCode	BetriebsGroesse
1	NAT_M	55	NAT	M
2	MED_G	41	MED	G
3	KRK_G	45	KRK	G
4	MED_M	45	MED	M
5	CON_M	42	CON	M
6	KRK_K	51	KRK	K
7	KRK_M	39	KRK	M
8	PRD_G	30	PRD	G
9	BNK_K	56	BNK	K
10	PRM_M	29	PRM	M
11	NAT_G	44	NAT	G
12	MED_M	40	MED	M
13	SOZ_K	24	SOZ	K
14	KRK_M	32	KRK	M
15	MED_K	43	MED	K
16	SOZ_M	30	SOZ	M
17	CON_G	27	CON	G
18	NAT_K	54	NAT	K
19	SOZ_G	30	SOZ	G
20	KRK_M	23	KRK	M
21	KRK_M	57	KRK	M
22	SOZ_G	31	SOZ	G
23	BNK_G	41	BNK	G
24	MED_G	36	MED	G

Abbildung 83: Datenansicht der neu berechneten Variablen

- *CONCAT und RTRIM*

Mit der CONCAT-Funktion können Sie mehrere Text-Ausdrücke oder –variablen zu einer Variablen verknüpfen. Die Funktion RTRIM dient dazu, allfällige nachstehende Leerzeichen einer Variable zu entfernen. Sie wird sehr häufig zusammen mit der CONCAT-Funktion verwendet, wenn zwei String-Variablen verkettet werden sollen. In der SPSS Kundenbefragung sind **Vorname** und **Name** in zwei Variablen abgespeichert. Diese sollen nun in einer Variable zusammengefasst werden. Wählen Sie wieder

Transformieren | Variable berechnen

Geben Sie als Zielvariable **FullName** ein. Klicken Sie auf **Typ&Label** und definieren Sie diese Variable als **String** mit einer Breite von **255 Zeichen**. Doppelkli-

cken Sie in der Funktionsgruppe **String** die **CONCAT-Funktion**. Ersetzen Sie die **?** durch die Variablen **Vorname** und **Name**.⁸ Um die nachstehenden Leerzeichen der Variable **Vorname** zu entfernen, können Sie die RTRIM-Funktion verwenden. Ergänzen Sie den String-Ausdruck wie folgt:

FullName = CONCAT(RTRIM(Vorname), RTRIM(Name))

Schliesslich sollen in der neuen Variable Vorname und Name noch durch ein Leerzeichen voneinander getrennt werden. Der korrekte Ausdruck lautet somit:

FullName = CONCAT(RTRIM(Vorname), " ",RTRIM(Name))

⁸ Leerzeichen werden in SPSS Statistics als normale Text-Zeichen betrachtet! Da die Variable Kanton im Datenfile als String-Variable mit einer Breite von 255 Zeichen definiert ist, würde die CONCAT-Funktion – so wie sie jetzt definiert ist – die Variable Kanton erst nach 13 Zeichen an die Variable Kanton_neu anhängen.

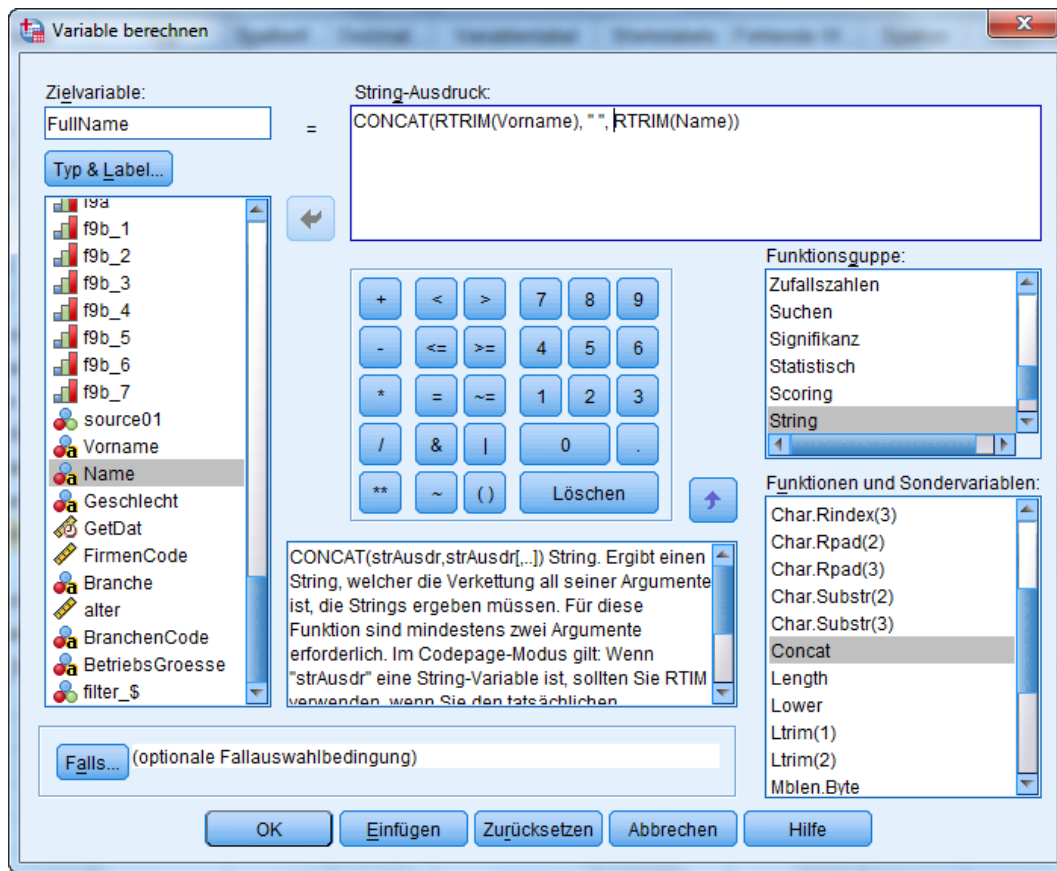


Abbildung 84: Ansicht "Variablen berechnen – FullName"

Klicken Sie auf **OK** und verschieben Sie die neue Variable vor die Variable Vorname und kontrollieren Sie das Ergebnis im Daten-Editor. Speichern Sie das Datenfile mit den neuen Variablen ab.

	FullName	Vorname
1	Jolanda Caduff	Jolanda
2	Michèle Rolla	Michèle
3	Sinja Kraus	Sinja
4	Lea Allmen	Lea
5	Rolf Kiewitz	Rolf
6	Michèle Zurbuche	Michèle
7	Claudia Ebnetter	Claudia
8	David Bury	David
9	Klara Broi	Klara
10	Urs Burgauer	Urs
11	Lorana Kistler	Lorana
12	René Scherrie	René
13	Lorana Retran	Lorana
14	Rolf Bienz	Rolf
15	Dimitri Nu	Dimitri
16	Nadia Weniger	Nadia
17	Jolanda Rotz	Jolanda
18	Leandra Eigenman	Leandra
19	Gertrud Rimoldi	Gertrud
20	Till Köchli	Till
21	Dimitri Gertsch	Dimitri
22	Simon Gertsche	Simon
23	Robin Nogler	Robin
24	Daniel Pock	Daniel

Abbildung 85: Datenansicht “Variable FullName”

Bedingte Berechnungen

Falls Sie die Berechnung der neuen Variablen nur für ausgewählte Fälle durchführen möchten, klicken Sie auf die Schaltfläche **Falls**. Im Dialogfeld **Variable berechnen: Falls Bedingung erfüllt ist** können Sie nun wie bei der allgemeinen Fallauswahl eine Bedingung eingeben. Für Fälle, die die Bedingungen nicht erfüllen, wird kein Wert berechnet. Wenn die Zielvariable eine neue Variable ist, erhalten diese Fälle das System Missing. Falls die Zielvariable eine bereits vorhandene Variable ist, behalten diese Fälle ihren ursprünglichen Wert.⁹

⁹ Die Verwendung des Filters aus Daten – Fälle auswählen hat keinen Einfluss auf die Berechnung und Erstellung von neuen Variablen. Ein Beispiel für bedingte Berechnungen finden Sie im Abschnitt zu den Konsistenzprüfungen.

6. Kategorisierung von metrischen Variablen

Oft ist es nötig, metrische Variablen wie Alter und Gehalt zu kategorisieren, um diese als Gruppierungsvariable (z.B. für Kreuztabellen) verwenden zu können. Für die Kategorisierung einer kontinuierlichen Variablen steht in SPSS Statistics der **Visual Bänder (Visuelles Klassieren)** zur Verfügung. Als Grundlage dafür werden nebeneinander liegende Werte bestehender Variablen in eine begrenzte Anzahl unterschiedlicher Kategorien gruppiert. Den Visual Bänder finden Sie unter

Transformieren | Visuelles Klassieren...

Wählen Sie **alter** und **satis_2** als zu unterteilende Variable und drücken Sie **Weiter**.

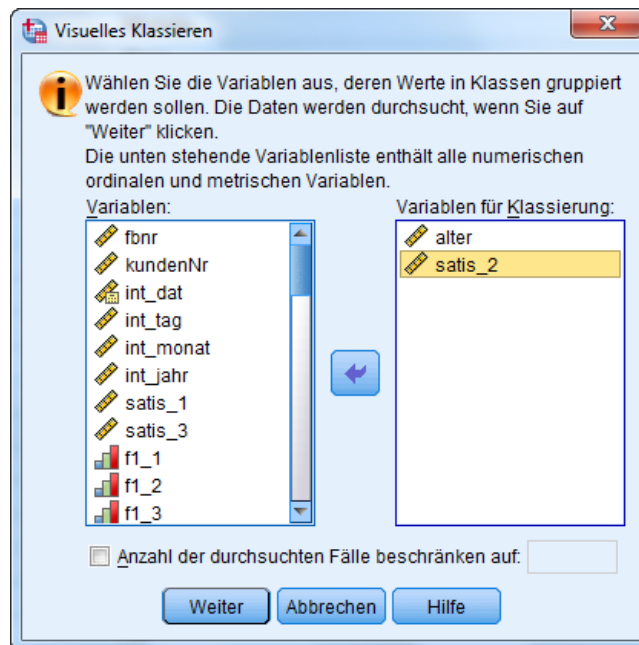


Abbildung 86: Ansicht "Visuelles Klassieren (1)"

Beispiel 1: Zunächst soll die Variable **alter** in vier ungefähr gleich grosse Gruppen unterteilt werden. Markieren Sie **alter**. Die Verteilung der ausgewählten Variable wird rechts in einem Histogramm dargestellt.

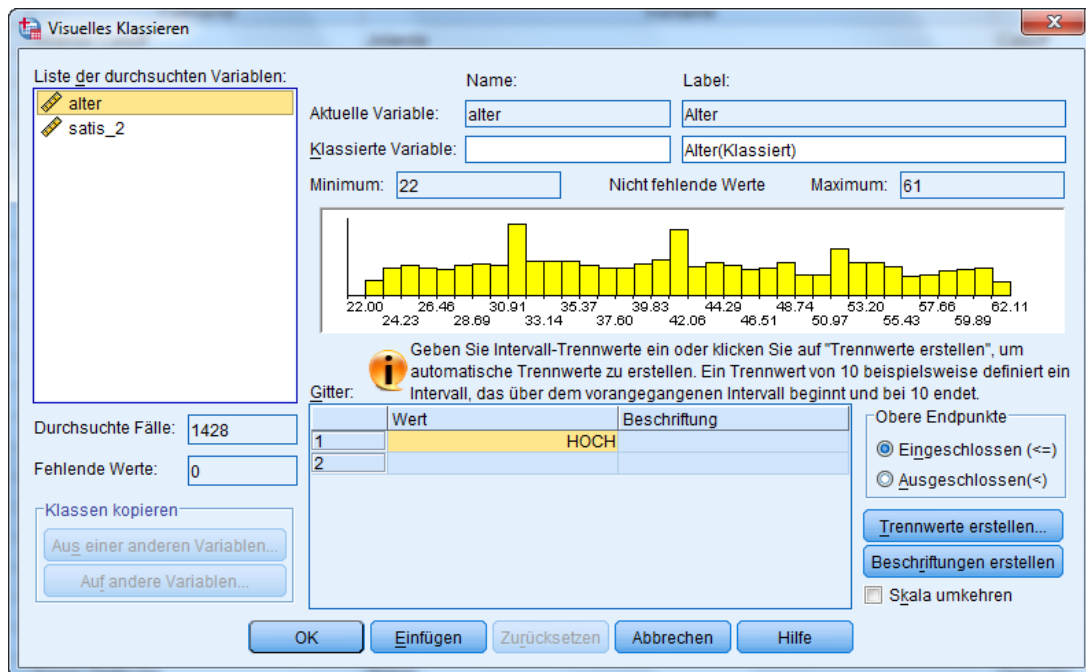


Abbildung 87: Ansicht "Visuelles Klassieren (2)"

Als Ausgabevariable geben Sie im Feld **Klassierte Variable** **alter_kat** ein und ändern nach Wunsch das entsprechende Label. Unter **Wert** können Sie nun die oberen Endpunkte der jeweiligen Kategorien definieren. Für Kategorien, welche nach oben offen sein sollen, setzen Sie den Wert **HOCH** ein. Unter **Obere Endpunkte** können Sie angeben, ob die oberen Endpunkte in die entsprechende Kategorie eingeschlossen oder dem nächst höheren Bereich zugeordnet werden sollen. Im letzteren Fall aktivieren Sie die Option **Ausgeschlossen (<)**. Um Gruppen gleicher Grösse zu bilden, klicken Sie auf **Trennwerte erstellen**.

Wählen Sie im Fenster **Trennwerte erstellen...** die 2. Option **Gleiche Perzentile auf der Grundlage der durchsuchten Fälle**. Für die vier gewünschten Altersgruppen benötigen wir **3 Trennwerte**, jede Gruppe enthält folglich ca. 25% aller Fälle. Tragen Sie unter **Breite (%)** deshalb den Wert **25** ein.

Trennwerte erstellen

☒ **Intervalle mit gleicher Breite**

Intervalle - mindestens zwei Felder ausfüllen

Position des ersten Trennwerts:

Anzahl der Trennwerte:

Breite:

Position des letzten Trennwerts:

☐ **Gleiche Perzentile auf der Grundlage der durchsuchten Fälle**

Intervalle - eines der beiden Felder ausfüllen

Anzahl der Trennwerte:

Breite (%):

☐ **Trennwerte bei Mittelwert und ausgewählten Standardabweichungen auf der Grundlage der durchsuchten Fälle**

☐ +/- 1 Std.-Abw.

☐ +/- 2 Std.-Abw.

☐ +/- 3 Std.-Abw.

Durch "Zuweisen" werden die Trennwertdefinitionen durch diese Spezifikation ersetzt.
Ein letztes Intervall enthält alle übrigen Werte: N Trennwerte führen zu N+1 Intervallen.

Abbildung 88: Ansicht "Trennwerte erstellen"

Klicken Sie auf **Zuweisen**. Unter Wert werden nun die vorgeschlagenen Trennwerte aufgelistet, welche Sie manuell noch anpassen können (z.B. 30, 40 und 50). Klicken Sie schliesslich noch auf **Beschriftungen erstellen**, um die entsprechenden Wertelabels zu generieren.

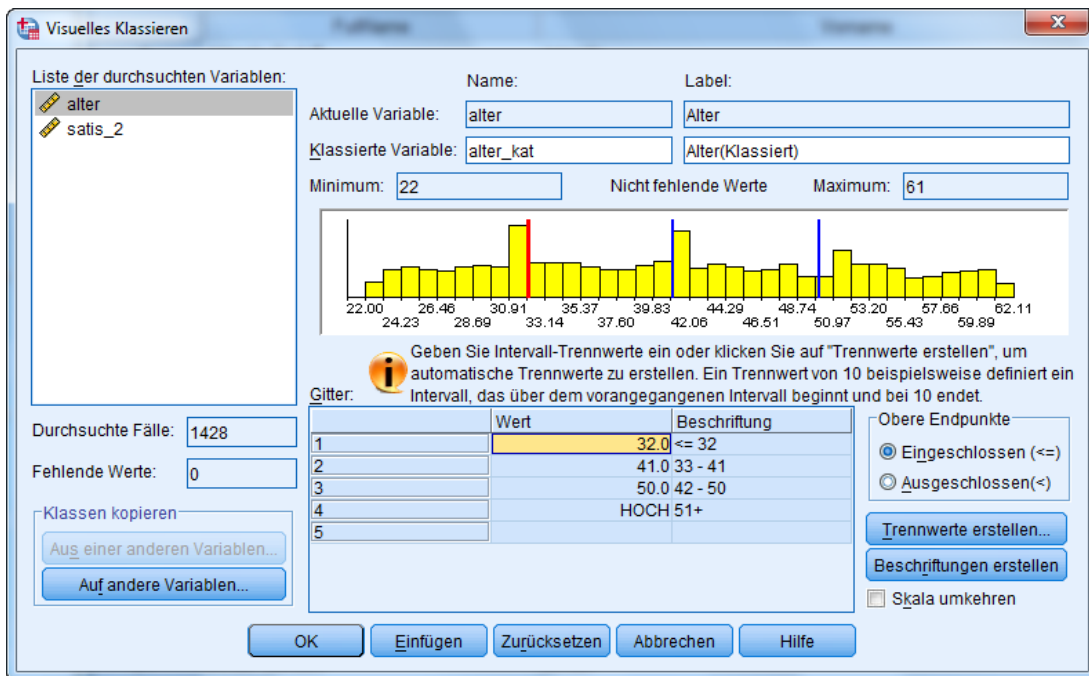


Abbildung 89: Ansicht "Visuelles Klassieren"

Beispiel 2: Markieren Sie nun die Variable **satis_2**. Diese soll nach folgenden Kriterien in 4 Gruppen aufgeteilt werden:

- Gruppe 1: Klar unterdurchschnittlich zufriedene SPSS-Kunden
(Index-Wert liegt mehr als 1 Standardabweichung unter dem allgemeinen Durchschnitt)
- Gruppe 2: Unterdurchschnittlich zufriedene SPSS-Kunden
(Index-Wert liegt weniger als 1 Standardabweichung unter dem allgemeinen Durchschnitt)
- Gruppe 3: Überdurchschnittlich zufriedene SPSS-Kunden
(Index-Wert liegt weniger als 1 Standardabweichung über dem allgemeinen Durchschnitt)
- Gruppe 4: Klar überdurchschnittlich zufriedene SPSS-Kunden
(Index-Wert liegt mehr als 1 Standardabweichung über dem allgemeinen Durchschnitt)

Geben Sie im Feld Klassierte Variable **satis_kat** ein und klicken dann auf **Trennwerte erstellen...** Wählen Sie die 3. Option **Trennwerte bei Mittelwert und ausgewählten Standardabweichungen auf der Grundlage der durchsuchten Fälle** und markieren Sie **+/- 1 Std.-Abw.**

Trennwerte erstellen

☐ Intervalle mit gleicher Breite

Intervalle - mindestens zwei Felder ausfüllen

Position des ersten Trennwerts:

Anzahl der Trennwerte:

Breite:

Position des letzten Trennwerts:

☐ Gleiche Perzentile auf der Grundlage der durchsuchten Fälle

Intervalle - eines der beiden Felder ausfüllen

Anzahl der Trennwerte:

Breite (%):

☒ **Trennwerte bei Mittelwert und ausgewählten Standardabweichungen auf der Grundlage der durchsuchten Fälle**

☒ +/- 1 Std.-Abw.

☐ +/- 2 Std.-Abw.

☐ +/- 3 Std.-Abw.

Durch "Zuweisen" werden die Trennwertedefinitionen durch diese Spezifikation ersetzt.
Ein letztes Intervall enthält alle übrigen Werte: N Trennwerte führen zu N+1 Intervallen.

Zuweisen **Abbrechen** **Hilfe**

Abbildung 90: Ansicht "Trennwerte erstellen"

Klicken Sie auf **Zuweisen**. Geben Sie nun noch aussagekräftige Labels für die einzelnen Gruppen und klicken dann auf **OK**, um die zwei neuen Gruppierungsvariablen zu erstellen.

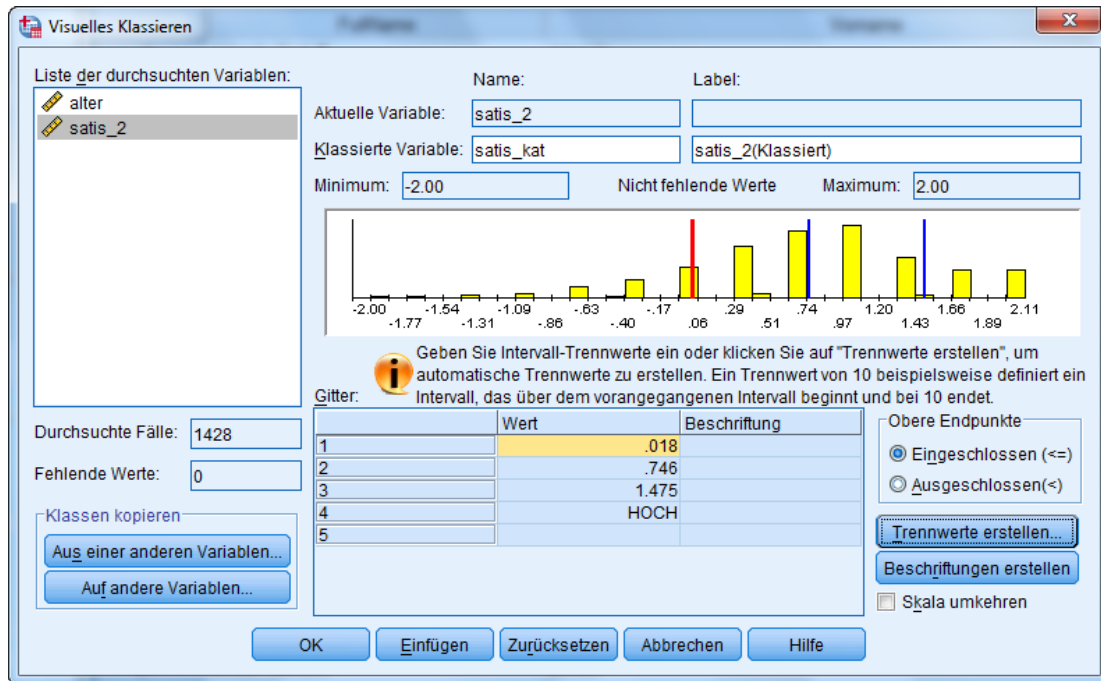


Abbildung 91: Ansicht "Visuelles Klassieren"

7. Daten aggregieren

Mit "Daten aggregieren" werden Fallgruppen in der Arbeitsdatei zu einzelnen Fällen kombiniert; hierbei wird eine neue, aggregierte Datei angelegt, oder es werden neue Variablen in der Arbeitsdatei angelegt, die aggregierte Informationen enthalten.

- Wenn Sie eine neue, aggregierte Datendatei anlegen, enthält diese neue Datei je einen Fall für jede Gruppe, die in den Break-Variablen definiert sind. Liegt beispielsweise eine Break-Variable mit zwei Gruppen vor, enthält die neue Datendatei nur zwei Fälle.
- Wenn Sie Aggregierungsvariablen in die Arbeitsdatei aufnehmen, wird die Datendatei selbst aggregiert. Jeder Fall mit denselben Werten für die Break-Variable(n) erhält dieselben Werte für die neuen Aggregierungsvariablen. Wenn beispielsweise nur eine Break-Variable für das Geschlecht vorliegt, erhalten alle männlichen Personen denselben Wert für eine neue Aggregierungsvariable, die das Durchschnittsalter erfasst.

Als Beispiel soll aus den Daten ein neues Datenfile auf Firmenebene generiert werden, welches für jede Firma folgende Informationen enthält:

- den durchschnittlichen Gesamtzufriedenheits-Index,
- die Anzahl benutzter Module,
- sowie der prozentuale Anteil derjenigen befragten Personen in dieser Firma, welche mit SPSS Statistics im Allgemeinen zufrieden sind.

Um die Daten zu aggregieren, wählen Sie

Daten | Aggregieren

Hier übernehmen Sie die Variable **FirmenCode** als Break-Variable(n) und verschieben die Variable **satis_2** in das Feld Aggregierte Variablen:

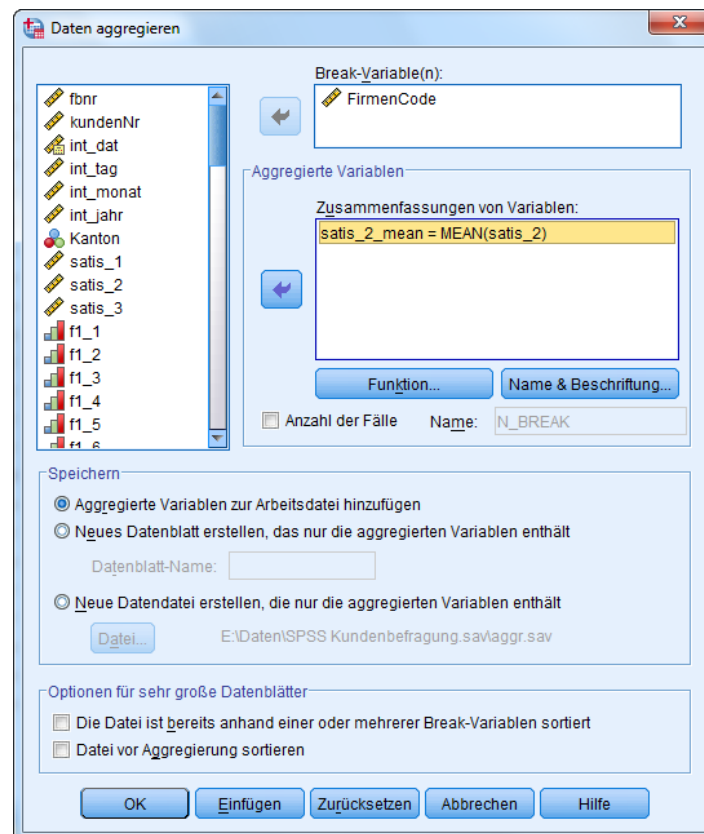


Abbildung 92: Ansicht "Daten aggregieren"

Standardmässig wird beim Aggregieren der Mittelwert der aggregierten Variable gerechnet. Für die Variable **satis_2** ist dies korrekt, Sie können die Funktion somit so stehen lassen. Klicken Sie auf **Name&Beschriftung...**, um den Variablennamen zu ändern und ein Label zu setzen. Tippen Sie als Name **satis_firma** ein, als Beschriftung **Firmen-Zufriedenheitsindex**.

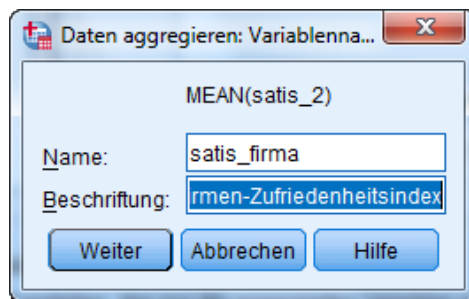


Abbildung 93: Ansicht "Daten aggregieren: Variablenname und -label"

Klicken Sie auf **Weiter**. Verschieben Sie nun die Variable **anz_module** in das Feld **Aggregierte Variablen**. Um die totale Anzahl der Module pro Firma zu erhalten, müssen wir nun die Funktion ändern: Klicken Sie auf **Funktion...** und wählen als **Auswertungsstatistik** statt Mittelwert **Summe**.

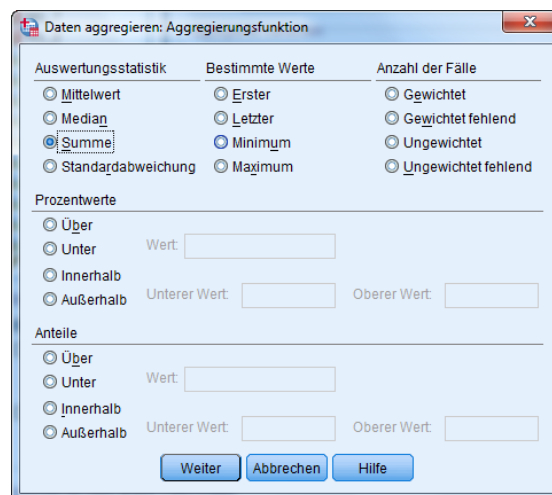


Abbildung 94: Ansicht "Daten aggregieren: Aggregierungsfunktion"

Klicken Sie auf **Weiter** und dann auf **Name&Beschriftung...** Tippen Sie als Name **AnzModule_firma** und als Beschriftung **Anzahl Module innerhalb Firma** ein.

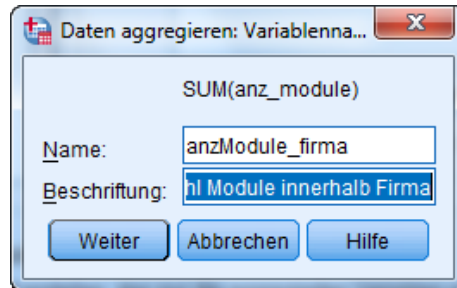


Abbildung 95: Ansicht "Daten aggregieren: Variablenname und -label"

Klicken Sie auf **Weiter**. Als letzte Information wird noch der prozentuale Anteil derjenigen befragten Personen pro Firma, welche mit SPSS Statistics im Allgemeinen zufrieden sind, benötigt. Diese Information erhalten Sie über die im letzten Kapitel erstellte Variable **satis_kat**, mit den 4 Kundensegmenten:

- 1) Klar unterdurchschnittlich unzufriedene Kunden
- 2) Unterdurchschnittlich unzufriedene Kunden
- 3) Überdurchschnittlich zufriedene Kunden
- 4) Klar überdurchschnittlich zufriedene Kunden

Verschieben Sie die Variable **satis_kat** in das Feld **Aggregierte Variablen**. Wählen Sie als Aggregierungsfunktion Prozentwerte **Über** aus und notieren als Wert **2**. Pro Firma erhalten Sie so den Anteil der befragten Personen aus Kundensegment 3 und 4.

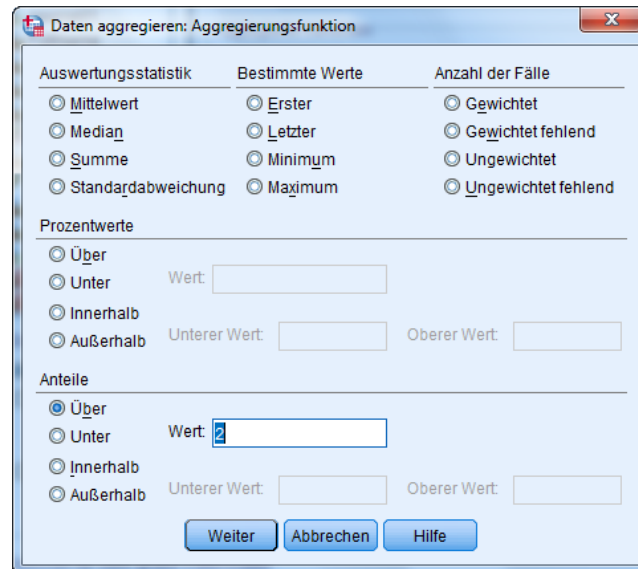


Abbildung 96: Ansicht "Daten aggregieren: Aggregierungsfunktion"

Klicken Sie auf **Weiter**. Unter **Name&Beschriftung...** geben Sie als Name **p_zufrieden** ein, als Beschriftung **%-Anteil überdurchschnittlich zufriedener Kunden pro Firma**. Klicken Sie anschliessend auf **Weiter**.

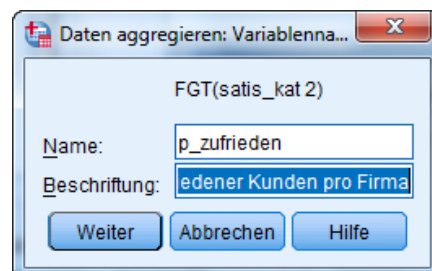


Abbildung 97: Ansicht "Daten aggregieren: Variablenname und -label"

Mit der Option **Anzahl der Fälle** können Sie eine zusätzliche Variable erstellen lassen, welche die Anzahl Fälle pro Breakgruppe (in diesem Fall also Anzahl SPSS Statistics-User pro Firma) enthält. Markieren Sie diese Option, als Name wählen Sie **anz_user**. Als Letztes müssen Sie noch angeben, wo die aggregierten Variablen abgespeichert werden sollen. Sie können diese Ihrer Arbeitsdatei hinzufügen, oder eine neue Datei nur mit den aggregierten Informationen erstel-

len lassen. Wählen Sie die Option **Neues Datenblatt erstellen**, als Datenblatt-Name geben Sie **aggr_firma** ein.

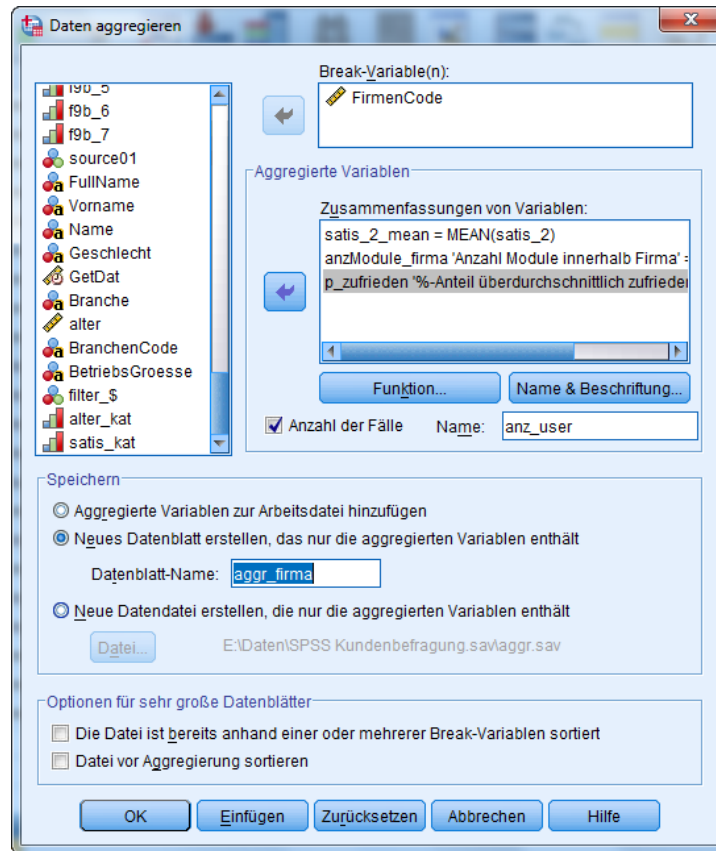


Abbildung 98: Ansicht “Daten aggregieren”

Klicken Sie auf **OK**, um die Aggregation durchzuführen. Es öffnet sich ein neues Datenfile mit den aggregierten Informationen. Speichern Sie das File unter **aggr_firma.sav** ab und schliessen Sie es.

aggr_firma.sav [aggr_firma] - IBM SPSS Statistics Daten-Editor

Menü: Datei Bearbeiten Ansicht Daten Transformieren Analysieren Direktmarketing Diagramme Extras Fenster Hilfe

Sichtbar: 5 von 5 Variablen

	FirmenCode	satis_firma	anzModule_firma	p_zufrieden	anz_user	var	var	var
1	1	.33	3.00	1.000	1			
2	2	.86	31.00	.833	6			
3	3	1.00	11.00	1.000	2			
4	4	.17	14.00	.667	3			
5	5	.36	67.00	.455	11			
6	6	.73	27.00	1.000	5			
7	7	.43	53.00	.700	10			
8	8	.95	40.00	.857	7			
9	9	.60	89.00	.813	16			
10	10	.92	40.00	1.000	6			
11	11	.85	46.00	.889	9			
12	12	.62	42.00	.857	7			
13	13	.87	27.00	1.000	5			
14	14	.59	50.00	.556	9			
15	15	.63	45.00	.875	8			
16	16	.46	21.00	.750	4			
17	17	.62	49.00	.857	7			
18	18	.75	35.00	1.000	6			
19	19	.60	43.00	.857	7			
20	20	.33	22.00	.750	4			
21	21	.67	12.00	1.000	2			
22	22	.22	27.00	.667	6			
23	23	.61	49.00	.778	9			
24	24	.39	58.00	.545	11			

Datenansicht Variablenansicht

IBM SPSS Statistics Prozessor ist bereit

Abbildung 99: Datenansicht neue Datei aggr_firma.sav

Datenanalyse

1. Allgemeines

Im Rahmen der Datenanalyse werden univariate und bi-/multivariate Analysen unterschieden:

Univariat	Bi-/multivariat
Häufigkeitsauszählungen; einzelne Variablen werden durch statistische Kennwerte (z.B. Mittelwert) beschrieben	Zusammenhänge/Unterschiede zwischen zwei oder mehr Variablen

Bivariate Analysemethoden nach Skalenniveaus

Skalenniveau	<i>Nominalskala</i> (dichotom) (z. B. Geschlecht)	<i>Ordinalskala</i> (z. B. Schulnoten)	<i>Intervallskala</i> (z. B. Körpergrösse)
<i>Nominalskala</i> (dichotom) (z. B. Geschlecht)	Kontingenzanalyse Chi ² -Statistik		T-Test (für unabhängige Stichproben)
<i>Ordinalskala</i> (z. B. Schulnoten)			(Rang)Korrelation (Rho-Koeff. oder tau-Koeff.)
<i>Intervallskala</i> (z. B. Körpergrösse)			Korrelation (Pearson-Koeffizient)

Skalen in der quantitativen Sozialforschung



	Skala	Mathematische Eigenschaften	Messwert-Eigenschaften	Beispiele
Nicht-Metrische Skala	Nominalskala	$A = A \nparallel B$	Klassifikation, Messwerte sind identisch oder nicht identisch	Geschlecht Betriebstyp im Handel
	Ordinalskala	$A > B > C$	Rangordnung, Messwerte lassen sich als kleiner / grösser / gleich einordnen	Schulnoten Präferenzurteile Schichtzugehörigkeit
Metrische Skala	Intervallskala	$A > B > C$	Rangordnung und Abstandsbestimmung, Abstände zwischen Messwerten können angegeben werden	Kalenderzeit Intelligenzquotient Celsius-Skala
	Ratioskala	$A = x \cdot B$	Absoluter Nullpunkt, neben der Abstandsbestimmung können auch Messwertverhältnisse berechnet werden	Alter Jahresumsatz

Messen umfasst immer den Versuch, eine Quantifizierung von qualitativen und/oder quantitativen Phänomenen vorzunehmen.

Nominalskalen

- Niedrigstes Skalenniveau;
- Ausprägungen eines Merkmals werden lediglich nach ihrer Gleichartigkeit unterschieden, gleiche Ausprägungen erhalten denselben Zahlenwert;
- Es können Klassifikationen gebildet werden (z. B. Geschlecht, Farben, Freizeitaktivitäten);
- Es sind lediglich einfache Häufigkeitsauswertungen möglich.

Ordinalskalen

- Klassifizierte Merkmale können ausserdem in eine Rangfolge gebracht werden;
- Neben der Gleichartigkeit von Ausprägungen können diese auch in Grösser-/Kleiner-Verhältnisse gesetzt werden (z. B. Benotungen, höherer Bildungsabschluss).

Intervallskala

- Zusätzlich haben auch die relativen Differenzen zwischen den Messwerten eine empirische Bedeutung (z. B. Kalendertage, Temperaturangaben in Celsiusgraden, Messwerte standardisierter Tests, wie z. B. zur Intelligenz oder zu Einstellungen).

Verhältnis- oder Ratioskala

- Höchstes Messniveau;
- Merkmalsausprägungen können sinnvoll auf einen absoluten Nullpunkt bezogen werden (z. B. Alter, Jahresumsatz);
- Kategorisierte Zahlen können ausser im Hinblick auf ihre Rangordnung und ihre Abstände auch in ihren Verhältnissen interpretierbar sein.

Nominal- und Ordinalskalen = nicht-metrische (qualitative) Skalen

Intervall- und Ratioskalen = metrische (quantitative) Skalen

2. Kreuztabellen und Chi-Quadrat-Test

- Frage: Bestehen zwischen zwei oder mehr Variablen Zusammenhänge?
- Beispiel: Sind Geschlecht und Musikgeschmack unabhängig voneinander?
- Zusammenhänge zwischen nichtmetrischen, also nominal- oder ordinalskalierten Variablen
- Chi-Quadrat-Test überprüft, ob sich die beobachteten Häufigkeiten signifikant von den erwarteten Häufigkeiten unterscheiden.

Fall Titanic (Datei: titanic.sav)

Am 10. April 1912 wird die Titanic, die als das grösste, sicherste und luxuriöseste Schiff galt, für ihre Jungfernfahrt nach New York vorbereitet. Streng nach Klassen getrennt, gehen die Passagiere an Bord. Nur vier Tage später rammt das Schiff einen Eisberg und sinkt innerhalb nur weniger Stunden. Der Untergang der Titanic wird zum Sinnbild für die Dreiklassengesellschaft des viktorianischen England und zum Symbol für einen unkritischen Technik- und Fortschrittsglauben.

Aufgabe: Analysieren Sie den Zusammenhang zwischen der Klassenzugehörigkeit der Passagiere und der Tatsache, ob sie die Katastrophe überlebt haben.

2.1 Vorgehen in SPSS Statistics:

Wählen Sie aus dem Menü

Analysieren | Deskriptive Statistiken | Kreuztabellen...

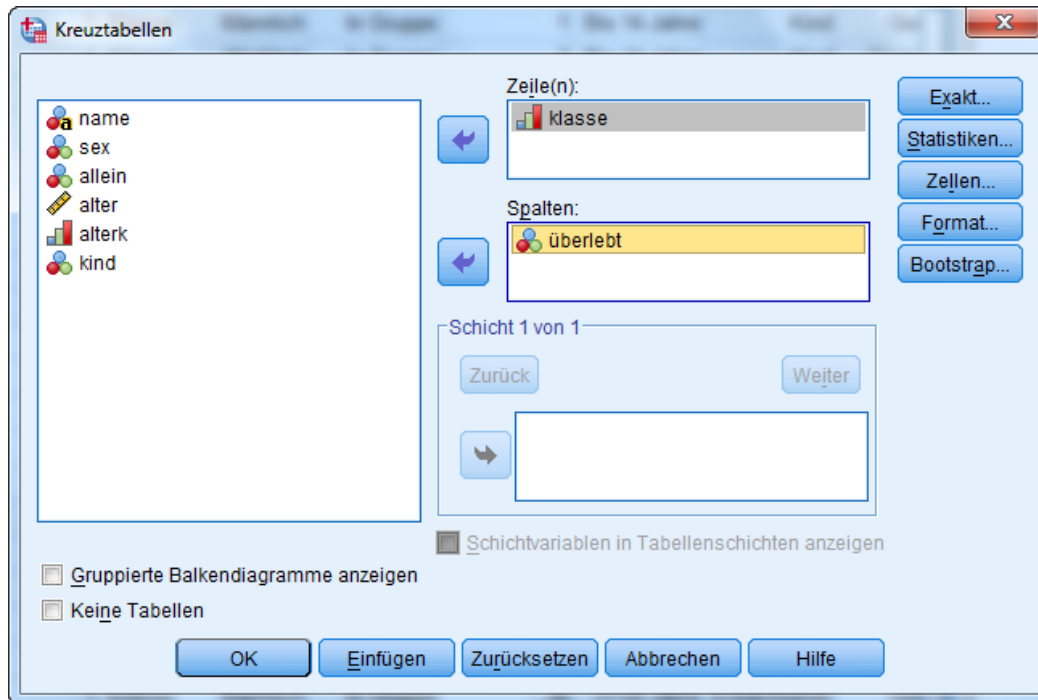


Abbildung 100: Ansicht "Kreuztabellen"

In der Quellvariablenliste erscheinen die Variablen der aktiven Datendatei. Sie wählen hier die Variablen, die Sie als Zeilen- und Spaltenvariablen verwenden wollen. Bilden Sie zunächst eine Kreuztabelle aus den Variablen **klasse** (Klassenzugehörigkeit des Passagiers an Bord des Schiffs) und **überlebt** (Information, ob der Passagier starb oder gerettet wurde). Gehen Sie folgendermassen vor:

Übertragen Sie die Variable **klasse** in die **Zeilenliste** und die Variable **überlebt** in die **Spaltenliste**. Klicken Sie auf **OK**, um die Kreuztabelle im voreingestellten Format zu erhalten. Im Viewer werden die beiden folgenden Tabellen angezeigt:

Verarbeitete Fälle						
	Fälle					
	Gültig		Fehlend		Gesamt	
	N	Prozent	N	Prozent	N	Prozent
Klasse * Überlebt?	1310	100.0%	0	.0%	1310	100.0%

Klasse * Überlebt? Kreuztabelle

Anzahl

		Überlebt?		Gesamt
		Gerettet	Gestorben	
Klasse	1. Klasse	201	123	324
	2. Klasse	118	158	276
	3. Klasse	183	527	710
Gesamt		502	808	1310

Abbildung 101: Viewerfensterausgabe "Verarbeitete Fälle und Kreuztabelle"

Die **erste Tabelle** gibt an, dass sich die Anzahl der Passagiere an Bord der Titanic auf 1310 Personen beläuft. Von allen Personen liegt eine Information vor, ob sie überlebt haben oder verstorben sind, so dass es keine fehlenden Werte gibt.

Die **zweite Tabelle** ist die eigentliche **Kreuztabelle**. Die Variable **überlebt** bildet die Spaltenvariable, die Variable **klasse** die Zeilenvariable. Der Eintrag in jeder Zelle gibt die Anzahl der Fälle an (die absolute Häufigkeit). So haben z. B. 201 Passagiere der ersten Klasse überlebt, 118 Personen aus der zweiten sowie 183 aus der dritten Klasse. Die innerhalb der Kreuztabelle eingetragenen Variablenlabels und Wertelabels entsprechen den Variablendefinitionen der SPSS-Datendatei.

Die Zahlen rechts neben der Tabelle und unterhalb der Tabelle (Gesamt) werden Randsummen genannt, und zwar Zeilen- bzw. Spaltensummen. Die Zeilensummen zeigen in diesem Beispiel an, dass 324 (201 + 123) Personen Passagiere der ersten Klasse waren. Die Spaltensummen zeigen an, dass 502 Passagiere (201 + 118 + 183) das Unglück überlebt haben. Insgesamt können die Angaben wie folgt interpretiert werden:

- Von den 1310 Titanic-Passagieren gehörten 324 zur ersten, 276 zur zweiten und 710 zur dritten Klasse.
- Bei der ersten Klasse haben von 324 Personen 201 Personen überlebt, während es bei der dritten Klasse von 710 Personen lediglich 183 waren.
- 502 Passagiere an Bord der Titanic wurden gerettet, während 808 Personen den Untergang des Schiffs nicht überlebten.

Der erste Eindruck deutet auf einen Zusammenhang zwischen den beiden Variablen **klasse** und **überlebt** hin. Die Überlebenswahrscheinlichkeit der Passagiere der ersten Klasse scheint deutlich höher zu liegen als die der dritten Klasse. Um den Zusammenhang näher zu untersuchen, soll präziser beantwortet werden, inwiefern es überhaupt einen signifikanten Zusammenhang gibt.

Eine Möglichkeit, die Existenz eines Zusammenhangs näher zu untersuchen, ist die Ausgabe der erwarteten Häufigkeit. Um die erwartete Häufigkeit auszugeben, wählen Sie im Menü

Analysieren | Deskriptive Statistiken | Kreuztabellen...

In der Zeilenliste sollte noch die Variable **klasse** eingetragen sein, in der Spaltenliste die Variable **überlebt**. Klicken Sie auf die Schaltfläche **Zellen...** Es öffnet sich die Dialogbox **Kreuztabellen: Zellen anzeigen**.

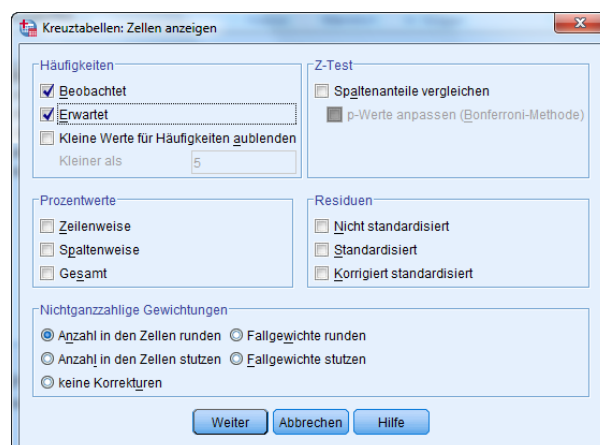


Abbildung 102: Ansicht "Kreuztabellen: Zellen anzeigen"

Im Auswahlkasten **Häufigkeiten** ist die Ausgabe der Anzahl der beobachteten Häufigkeiten voreingestellt. Aktivieren Sie zusätzlich die **erwarteten Häufigkeiten** und bestätigen Sie mit **Weiter** und dann **OK**.¹⁰ Sie erhalten die folgende Kreuztabelle:

			Überlebt?		Gesamt
			Gerettet	Gestorben	
Klasse	1. Klasse	Anzahl	201	123	324
		Erwartete Anzahl	124.2	199.8	324.0
	2. Klasse	Anzahl	118	158	276
		Erwartete Anzahl	105.8	170.2	276.0
	3. Klasse	Anzahl	183	527	710
		Erwartete Anzahl	272.1	437.9	710.0
	Gesamt	Anzahl	502	808	1310
		Erwartete Anzahl	502.0	808.0	1310.0

Abbildung 103: View Fensterausgabe "Kreuztabelle"

Neben den beobachteten Häufigkeiten (Anzahl) sind nun auch die erwarteten Häufigkeiten (erwartete Anzahl) aufgenommen worden. Die **erwarteten Häufigkeiten**, also diejenigen Häufigkeiten, die vorliegen müssen, wenn statistisch betrachtet kein Zusammenhang zwischen den betreffenden Variablen vorläge, lassen sich für die sechs Zellen der obigen Kreuztabelle wie folgt berechnen:

Zelle	Anzahl (beobachtete Werte)	Rechenformel	Ergebnis	Erwartete Anzahl (gerundeter Wert)
1. Zelle	201	$324 * 502 / 1310$	124,15877	124,2
2. Zelle	123	$324 * 808 / 1310$	199,84122	199,8
3. Zelle	118	$276 * 502 / 1310$	105,76488	105,8
4. Zelle	158	$276 * 808 / 1310$	170,23511	170,2
5. Zelle	183	$710 * 502 / 1310$	272,07633	272,1

¹⁰ Die erwarteten Häufigkeiten berechnen sich aus dem Produkt der betreffenden Zeilen- und Spaltensumme, geteilt durch die Gesamtsumme der Häufigkeiten (erwartete Häufigkeit = Zeilensumme * Spaltensumme / Gesamtsumme der Häufigkeit).

6. Zelle	527	$710 * 808 / 1310$	437,92366	437,9
----------	-----	--------------------	-----------	-------

Interpretation der obigen Kreuztabelle: Während bei der ersten und der zweiten Klasse die absolute Häufigkeit bei den „geretteten“ Personen höher ist als die erwartete Häufigkeit (201 statt 124,2; 118 statt 105,8), liegt sie bei der dritten Klasse niedriger (183 statt 272,1). Bei der Merkmalsausprägung „gestorben“ findet sich der entgegengesetzte Trend: Während bei der ersten und zweiten Klasse die absolute Häufigkeit bei den ertrunkenen Personen niedriger ist als die erwartete Häufigkeit (123 statt 199,8; 158 statt 170,2), liegt sie bei der dritten Klasse höher (527 statt 437,9).

Der obige Eindruck, dass die Überlebenswahrscheinlichkeit der Passagiere der ersten und der zweiten Klasse höher liegen als die der dritten Klasse, scheint sich zu bestätigen. Eine weitere Möglichkeit, um die Existenz eines solchen Zusammenhangs näher zu untersuchen, ist die **Berechnung der Residuen**. Diese sind ein Mass dafür, wie stark beobachtete und erwartete Häufigkeiten voneinander abweichen. Um die Residualhäufigkeit auszugeben, wählen Sie

Analysieren | Deskriptive Statistiken | Kreuztabellen...

In der Zeilenliste sollte noch die Variable **klasse** eingetragen sein, in der Spaltenliste die Variable **überlebt**. Klicken Sie auf **Zellen...** Aktiviert sein sollten noch die Optionen **Beobachtet** und **Erwartet**. Im Auswahlkasten Residuen können Sie eine oder mehrere der folgenden Angaben wählen:

- **Nicht standardisiert:** Zeigt nicht standardisierte Residuen an, d.h. beobachtete Zellenhäufigkeiten minus erwartete Zellenhäufigkeiten.
- **Standardisiert:** Zeigt standardisierte Residuen an. Die nicht standardisierten Residuen werden hierbei durch die Quadratwurzel aus der erwarteten Häufigkeit dividiert. Die standardisierten Residuen sind nützlich bei der näheren Analyse des Chi-Quadrat-Tests.

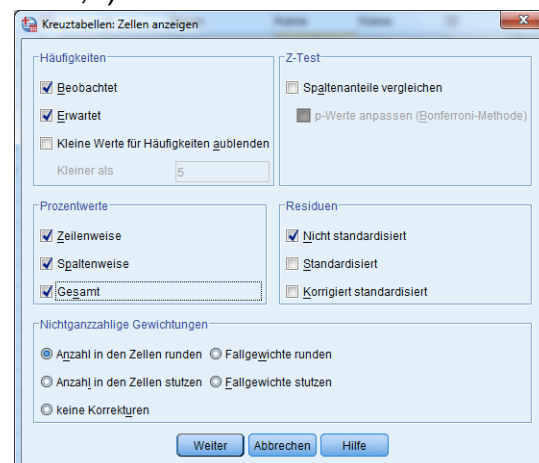
Aktivieren Sie im Auswahlkasten **Residuen** zusätzlich die Option **Nicht standardisiert**. Bestätigen Sie mit **Weiter** und klicken Sie in der Hauptdialogbox auf **OK**. Sie erhalten folgende Kreuztabelle:

			Überlebt?		Gesamt
			Gerettet	Gestorben	
Klasse	1. Klasse	Anzahl	201	123	324
		Erwartete Anzahl	124.2	199.8	324.0
		Residuen	76.8	-76.8	
	2. Klasse	Anzahl	118	158	276
		Erwartete Anzahl	105.8	170.2	276.0
		Residuen	12.2	-12.2	
	3. Klasse	Anzahl	183	527	710
		Erwartete Anzahl	272.1	437.9	710.0
		Residuen	-89.1	89.1	
Gesamt	Anzahl	502	808	1310	
	Erwartete Anzahl	502.0	808.0	1310.0	

Abbildung 104: Viewerfensterausgabe "Kreuztabelle"

An den Zelleninhalten können Sie erkennen, dass die **Residuen** gleich der beobachteten Häufigkeiten minus der theoretisch erwarteten Häufigkeiten sind (bezogen auf die erste Zelle z.B. $201 - 124,2 = 76,8$). Die mit Plus- und Minuszeichen versehenen Residuen lassen den gegenläufigen Trend bei Passagieren der ersten und der dritten Klasse noch deutlicher erkennen.

Um Werte gemäss ihrer Bedeutung zu gewichten, ist die Kenntnis der jeweiligen prozentualen Anteile wichtig. Um zusätzlich **prozentuale Anteile** auszugeben, wählen Sie aus dem Menü



Übernehmen Sie die bisherigen Eintragungen und klicken Sie auf **Zellen...** Es öffnet sich die Dialogbox **Kreuztabellen: Zellen anzeigen**. Im Auswahlkasten Prozentwerte können Sie die Prozentwerte **Zeilenweise**, **Spaltenweise** und/oder **Gesamt** wählen. Klicken Sie alle drei an. Bestätigen Sie mit **Weiter** und klicken Sie in der Hauptdialogbox auf **OK**.

Im Viewer sehen Sie die folgende Kreuztabelle:

Klasse * Überlebt? Kreuztabelle					
			Überlebt?		Gesamt
			Gerettet	Gestorben	
Klasse	1. Klasse	Anzahl	201	123	324
		Erwartete Anzahl	124.2	199.8	324.0
		% innerhalb von Klasse	62.0%	38.0%	100.0%
		% innerhalb von Überlebt?	40.0%	15.2%	24.7%
		% der Gesamtzahl	15.3%	9.4%	24.7%
		Residuen	76.8	-76.8	
	2. Klasse	Anzahl	118	158	276
		Erwartete Anzahl	105.8	170.2	276.0
		% innerhalb von Klasse	42.8%	57.2%	100.0%
		% innerhalb von Überlebt?	23.5%	19.6%	21.1%
		% der Gesamtzahl	9.0%	12.1%	21.1%
		Residuen	12.2	-12.2	
	3. Klasse	Anzahl	183	527	710
		Erwartete Anzahl	272.1	437.9	710.0
		% innerhalb von Klasse	25.8%	74.2%	100.0%
		% innerhalb von Überlebt?	36.5%	65.2%	54.2%
		% der Gesamtzahl	14.0%	40.2%	54.2%
		Residuen	-89.1	89.1	
	Gesamt	Anzahl	502	808	1310
		Erwartete Anzahl	502.0	808.0	1310.0
		% innerhalb von Klasse	38.3%	61.7%	100.0%
		% innerhalb von Überlebt?	100.0%	100.0%	100.0%
		% der Gesamtzahl	38.3%	61.7%	100.0%

Abbildung 105: Viewerfensterausgabe "Kreuztabelle"

Bei den prozentualen Zeilenhäufigkeiten wird eine Prozentuierung auf den jeweiligen Zeilensummenwert vorgenommen, bei der prozentualen Spaltenhäufigkeit eine Prozentuierung auf den jeweiligen Spaltensummenwert; bei der prozentualen Gesamthäufigkeit wird auf die Gesamtzahl der gültigen Fälle (hier: 1310) prozentuiert. Die Werte der ersten Zeilen lassen sich bspw. wie folgt interpretieren:

- Von den insgesamt 324 Passagieren der ersten Klasse haben 201 überlebt, das sind 62%.

- Von 502 Geretteten stammen 201 aus der ersten Klasse, was 40% aller geretteten Passagiere entspricht.
- Unter den 1310 Passagieren kommen 201 aus der ersten Klasse, was 15,3% aller Passagiere der Titanic entspricht.

I.d.R. ist es sinnvoll nur die **unabhängige Variable** als Basis der Prozentuierung zu verwenden. Im gegebenen Beispiel ist die abhängige Variable die Variable **überlebt**. Sie wird als abhängig von der **Klassenzugehörigkeit** an Bord angesehen. Bezogen auf die oben genannte Fragestellung ist vor allem die Prozentuierung auf die Zeilensumme (% von Gerettet) nützlich, die im Folgenden zusammengestellt ist:

	Gerettet	Gestorben
1. Klasse	62,0%	38,0%
2. Klasse	42,8%	57,2%
3. Klasse	25,8%	74,2%

Da sich in diesem Fall die prozentualen Verteilungen erheblich voneinander unterscheiden, kann ein **statistischer Zusammenhang** zwischen den Merkmalen **klasse** und **überlebt** vermutet werden. Wesentlich mehr Passagiere der ersten Klasse haben überlebt als Passagiere der dritten Klasse. Die Passagiere der zweiten Klasse nehmen damit verglichen eine Mittellage ein. Ein klassenspezifischer Unterschied hinsichtlich der Überlebenswahrscheinlichkeit der Passagiere ist somit deutlich erkennbar. Ob dieser Unterschied signifikant ist, wird mit dem Chi-Quadrat-Test entschieden (dazu mehr nach dem folgenden Exkurs).

Die **Sortierreihenfolge der Zeilenvariablen einer Kreuztabelle** (aufsteigend (= Standard) und absteigend) können Sie verändern, indem Sie in der **Dialogbox Kreuztabelle** auf **Format...** klicken. Es öffnet sich die Dialogbox **Kreuztabellen: Tabellenformat**.

Exkurs: Einsatz von Gruppen- und Schichtenvariablen

(1) Schichtenvariable „geschlecht“

Es ist denkbar, dass die Variable **geschlecht** einen Einfluss auf den Zusammenhang zwischen **überlebt** und **klasse** ausübt. Um eventuell bestehende geschlechtsspezifische Unterschiede zu erfassen, ist die Ausgabe getrennter Tabellen erforderlich, und zwar eine Tabelle für jedes Geschlecht. Die Variable **sex** spielt daher in diesem Fall die Funktion einer **Schichtenvariablen**. Durchgeführt wird ein Gruppenwechsel, d.h., für jede Gruppe (hier: weiblich und männlich), wird eine Kreuztabelle ausgegeben. Um eine Schichtenvariable hinzuzunehmen, gehen Sie wie folgt vor:

Analysieren | Deskriptive Statistiken | Kreuztabellen...

In der Zeilenliste sollte noch die Variable **klasse** eingetragen sein, in der Spaltenliste die Variable **überlebt**. Übertragen Sie die Variable **sex** in das **Schichtenvariablenfeld**. Dieses ist das noch freie dritte Feld in der Dialogbox. Die Dialogbox Kreuztabellen sieht wie folgt aus:

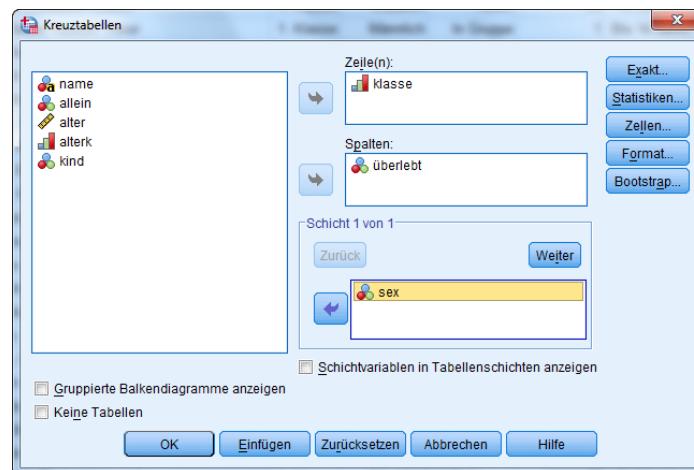


Abbildung 106: Ansicht "Kreuztabellen"

Wahlweise können Sie weitere Ebenen von Schichtenvariablen wählen. Für jede Kategorie jeder Schichtenvariablen wird eine eigene Kreuztabelle erstellt. Weitere Ebenen von Schichtenvariablen können Sie hinzunehmen, indem Sie auf **Weiter** klicken. Jede weitere Ebene unterteilt die Kreuztabellierung in kleinere Unter-

gruppen. Die Schaltflächen Weiter und Zurück können Sie verwenden, um zwischen den verschiedenen Ebenen der Schichtenvariablen zu wechseln. Die Unterteilung der Schichtenvariablen ist beim Einsatz statistischer Kennziffern allerdings nur bei genügend grossen Fallzahlen sinnvoll. Klicken Sie auf den Schalter **Zellen**. Lassen Sie sich die **beobachteten** und die **erwarteten Häufigkeiten** ausgeben sowie die **Prozentwerte zeilenweise, spaltenweise** und **gesamt**. Bestätigen Sie mit **Weiter** und **OK**.

Klasse * Überlebt? * Geschlecht Kreuztabelle						
Geschlecht				Überlebt?		Gesamt
				Gerettet	Gestorben	
Männlich	Klasse	1. Klasse	Anzahl	62	118	180
			Erwartete Anzahl	34.8	145.2	180.0
			% innerhalb von Klasse	34.4%	65.6%	100.0%
			% innerhalb von Überlebt?	38.0%	17.4%	21.4%
			% der Gesamtzahl	7.4%	14.0%	21.4%
		2. Klasse	Anzahl	24	146	170
			Erwartete Anzahl	32.9	137.1	170.0
			% innerhalb von Klasse	14.1%	85.9%	100.0%
			% innerhalb von Überlebt?	14.7%	21.5%	20.2%
			% der Gesamtzahl	2.8%	17.3%	20.2%
		3. Klasse	Anzahl	77	416	493
			Erwartete Anzahl	95.3	397.7	493.0
			% innerhalb von Klasse	15.6%	84.4%	100.0%
			% innerhalb von Überlebt?	47.2%	61.2%	58.5%
			% der Gesamtzahl	9.1%	49.3%	58.5%
		Gesamt	Anzahl	163	680	843
			Erwartete Anzahl	163.0	680.0	843.0
			% innerhalb von Klasse	19.3%	80.7%	100.0%
			% innerhalb von Überlebt?	100.0%	100.0%	100.0%
			% der Gesamtzahl	19.3%	80.7%	100.0%

Weiblich	Klasse	1. Klasse	Anzahl	139	5	144
			Erwartete Anzahl	104.5	39.5	144.0
			% innerhalb von Klasse	96.5%	3.5%	100.0%
			% innerhalb von Überlebt?	41.0%	3.9%	30.8%
			% der Gesamtzahl	29.8%	1.1%	30.8%
		2. Klasse	Anzahl	94	12	106
			Erwartete Anzahl	76.9	29.1	106.0
			% innerhalb von Klasse	88.7%	11.3%	100.0%
			% innerhalb von Überlebt?	27.7%	9.4%	22.7%
			% der Gesamtzahl	20.1%	2.6%	22.7%
		3. Klasse	Anzahl	106	111	217
			Erwartete Anzahl	157.5	59.5	217.0
			% innerhalb von Klasse	48.8%	51.2%	100.0%
			% innerhalb von Überlebt?	31.3%	86.7%	46.5%
			% der Gesamtzahl	22.7%	23.8%	46.5%
		Gesamt	Anzahl	339	128	467
			Erwartete Anzahl	339.0	128.0	467.0
			% innerhalb von Klasse	72.6%	27.4%	100.0%
			% innerhalb von Überlebt?	100.0%	100.0%	100.0%
			% der Gesamtzahl	72.6%	27.4%	100.0%

Gesamt	Klasse	1. Klasse	Anzahl	201	123	324
			Erwartete Anzahl	124.2	199.8	324.0
			% innerhalb von Klasse	62.0%	38.0%	100.0%
			% innerhalb von Überlebt?	40.0%	15.2%	24.7%
			% der Gesamtzahl	15.3%	9.4%	24.7%
		2. Klasse	Anzahl	118	158	276
			Erwartete Anzahl	105.8	170.2	276.0
			% innerhalb von Klasse	42.8%	57.2%	100.0%
			% innerhalb von Überlebt?	23.5%	19.6%	21.1%
			% der Gesamtzahl	9.0%	12.1%	21.1%
		3. Klasse	Anzahl	183	527	710
			Erwartete Anzahl	272.1	437.9	710.0
			% innerhalb von Klasse	25.8%	74.2%	100.0%
			% innerhalb von Überlebt?	36.5%	65.2%	54.2%
			% der Gesamtzahl	14.0%	40.2%	54.2%
		Gesamt	Anzahl	502	808	1310
			Erwartete Anzahl	502.0	808.0	1310.0
			% innerhalb von Klasse	38.3%	61.7%	100.0%
			% innerhalb von Überlebt?	100.0%	100.0%	100.0%
			% der Gesamtzahl	38.3%	61.7%	100.0%

Abbildung 107: Viewerfensterausgabe "Kreuztabelle"

Das Hinzunehmen der Variablen **sex als Schichtenvariable** verdeutlicht die sehr starken geschlechtsspezifischen Effekte der Überlebenswahrscheinlichkeit an Bord der Titanic. Während von den Männern der 1. Klasse 34,4% überlebten, sind es von den Frauen der 1. Klasse 96,5%. Von den Männern der 3. Klasse

überlebten lediglich 15,6% den Untergang der Titanic, während es von den Frauen der 3. Klasse 48,8% waren.

(2) Schichtenvariable kind (als zweite Schichtenvariable)

Wählen Sie im Menü

Analysieren | Deskriptive Statistiken | Kreuztabellen...

In der Zeilenleiste sollte noch die Variable **klasse** eingetragen sein, in der Spaltenliste die Variable **überlebt** und im Schichtenvariablenfeld die Variable **sex**. Übertragen Sie in das Schichtenvariablenfeld zusätzlich die Variable **kind**. Bestätigen Sie mit **OK**. Nach der Ihnen bekannten Kreuztabelle wird nun im Viewer noch eine **zweite Kreuztabelle** mit der Schichtenvariablen **kind** ausgegeben.

Klasse * Überlebt? * Kind oder Erwachsener Kreuztabelle						
Kind oder Erwachsener				Überlebt?		Gesamt
				Gerettet	Gestorben	
Kind	1. Klasse	Anzahl		6	1	7
		Erwartete Anzahl		3.9	3.1	7.0
		% innerhalb von Klasse		85.7%	14.3%	100.0%
		% innerhalb von Überlebt?		9.4%	2.0%	6.1%
		% der Gesamtzahl		5.3%	.9%	6.1%
	2. Klasse	Anzahl		26	0	26
		Erwartete Anzahl		14.6	11.4	26.0
		% innerhalb von Klasse		100.0%	.0%	100.0%
		% innerhalb von Überlebt?		40.6%	.0%	22.8%
		% der Gesamtzahl		22.8%	.0%	22.8%
	3. Klasse	Anzahl		32	49	81
		Erwartete Anzahl		45.5	35.5	81.0
		% innerhalb von Klasse		39.5%	60.5%	100.0%
		% innerhalb von Überlebt?		50.0%	98.0%	71.1%
		% der Gesamtzahl		28.1%	43.0%	71.1%
	Gesamt	Anzahl		64	50	114
		Erwartete Anzahl		64.0	50.0	114.0
		% innerhalb von Klasse		56.1%	43.9%	100.0%
		% innerhalb von Überlebt?		100.0%	100.0%	100.0%
		% der Gesamtzahl		56.1%	43.9%	100.0%

Erwachsener	Klasse	1. Klasse	Anzahl	185	106	291
			Erwartete Anzahl	111.5	179.5	291.0
			% innerhalb von Klasse	63.6%	36.4%	100.0%
			% innerhalb von Überlebt?	44.9%	16.0%	27.1%
			% der Gesamtzahl	17.2%	9.9%	27.1%
		2. Klasse	Anzahl	87	145	232
			Erwartete Anzahl	88.9	143.1	232.0
			% innerhalb von Klasse	37.5%	62.5%	100.0%
			% innerhalb von Überlebt?	21.1%	21.9%	21.6%
			% der Gesamtzahl	8.1%	13.5%	21.6%
		3. Klasse	Anzahl	140	412	552
			Erwartete Anzahl	211.6	340.4	552.0
			% innerhalb von Klasse	25.4%	74.6%	100.0%
			% innerhalb von Überlebt?	34.0%	62.1%	51.3%
			% der Gesamtzahl	13.0%	38.3%	51.3%
		Gesamt	Anzahl	412	663	1075
			Erwartete Anzahl	412.0	663.0	1075.0
			% innerhalb von Klasse	38.3%	61.7%	100.0%
			% innerhalb von Überlebt?	100.0%	100.0%	100.0%
			% der Gesamtzahl	38.3%	61.7%	100.0%

Gesamt	Klasse	1. Klasse	Anzahl	191	107	298
			Erwartete Anzahl	119.3	178.7	298.0
			% innerhalb von Klasse	64.1%	35.9%	100.0%
			% innerhalb von Überlebt?	40.1%	15.0%	25.1%
			% der Gesamtzahl	16.1%	9.0%	25.1%
		2. Klasse	Anzahl	113	145	258
			Erwartete Anzahl	103.3	154.7	258.0
			% innerhalb von Klasse	43.8%	56.2%	100.0%
			% innerhalb von Überlebt?	23.7%	20.3%	21.7%
			% der Gesamtzahl	9.5%	12.2%	21.7%
		3. Klasse	Anzahl	172	461	633
			Erwartete Anzahl	253.4	379.6	633.0
			% innerhalb von Klasse	27.2%	72.8%	100.0%
			% innerhalb von Überlebt?	36.1%	64.7%	53.2%
			% der Gesamtzahl	14.5%	38.8%	53.2%
		Gesamt	Anzahl	476	713	1189
			Erwartete Anzahl	476.0	713.0	1189.0
			% innerhalb von Klasse	40.0%	60.0%	100.0%
			% innerhalb von Überlebt?	100.0%	100.0%	100.0%
			% der Gesamtzahl	40.0%	60.0%	100.0%

Abbildung 108: View Fensterausgabe "Kreuztabelle"

Während 85,7% der Kinder der ersten Klasse überlebt haben sowie alle Kinder der zweiten Klasse, sind es von den Kindern der dritten Klasse lediglich 39,5%. Da nicht zu jedem Passagier der Titanic Angaben über das Alter vorliegen, be-

läuft sich, wie der Tabelle „Verarbeitete Fälle“ zu entnehmen ist, die Anzahl der fehlenden Fälle auf 9,2%.

Verarbeitete Fälle						
	Fälle					
	Gültig		Fehlend		Gesamt	
	N	Prozent	N	Prozent	N	Prozent
Klasse * Überlebt? * Geschlecht	1310	100.0%	0	.0%	1310	100.0%
Klasse * Überlebt? * Kind oder Erwachsener	1189	90.8%	121	9.2%	1310	100.0%

Abbildung 109: Viewerfensterausgabe „Verarbeitete Fälle“

(3) Schichtenvariable kind (als Schichtenvariable der zweiten Ebene)

Im Unterschied zur vorherigen Ausgabe soll nun die Variable kind nicht als eine weitere Schichtenvariable, sondern als **Schichtenvariable der zweiten Ebene** eingesetzt werden. Wählen Sie aus dem Menü

Analysieren | Deskriptive Statistiken | Kreuztabellen...

In der Zeilenliste sollte noch die Variable **klasse** eingetragen sein, in der Spaltenliste die Variable **überlebt** und im Schichtenvariablenfeld die Variablen **sex** und **kind**. Transportieren Sie zunächst die Variable **kind** in die Quellvariablenliste zurück. Klicken Sie auf **Schichtenvariablenfeld** auf den Schalter Weiter. Übertragen Sie nun die Variable **kind** in das Schichtenvariablenfeld.

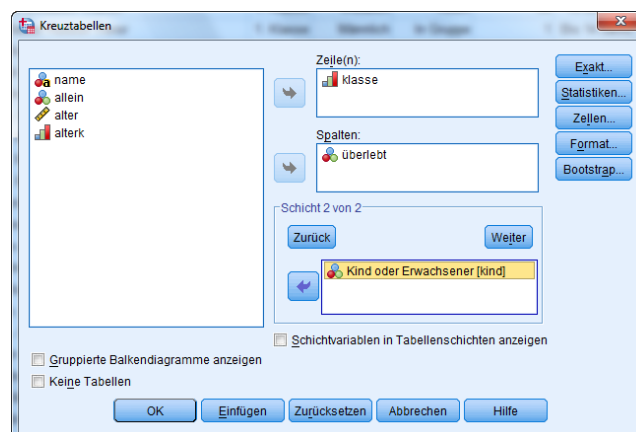


Abbildung 110: Ansicht „Kreuztabellen“

Bestätigen Sie die Eingabe mit **OK**. Im Viewer erscheint die folgende Kreuztabelle:

Klasse * Überlebt? * Geschlecht * Kind oder Erwachsener Kreuztabelle							
Kind oder Erwachsener	Geschlecht	Klasse			Überlebt?		Gesamt
					Gerettet	Gestorben	
Kind	Männlich	Klasse	1. Klasse	Anzahl	5	0	5
				Erwartete Anzahl	2.5	2.5	5.0
				% innerhalb von Klasse	100.0%	.0%	100.0%
				% innerhalb von Überlebt?	16.7%	.0%	8.3%
				% der Gesamtzahl	8.3%	.0%	8.3%
			2. Klasse	Anzahl	11	0	11
				Erwartete Anzahl	5.5	5.5	11.0
				% innerhalb von Klasse	100.0%	.0%	100.0%
				% innerhalb von Überlebt?	36.7%	.0%	18.3%
				% der Gesamtzahl	18.3%	.0%	18.3%
			3. Klasse	Anzahl	14	30	44
				Erwartete Anzahl	22.0	22.0	44.0
				% innerhalb von Klasse	31.8%	68.2%	100.0%
				% innerhalb von Überlebt?	46.7%	100.0%	73.3%
				% der Gesamtzahl	23.3%	50.0%	73.3%
			Gesamt	Anzahl	30	30	60
				Erwartete Anzahl	30.0	30.0	60.0
				% innerhalb von Klasse	50.0%	50.0%	100.0%
				% innerhalb von Überlebt?	100.0%	100.0%	100.0%
				% der Gesamtzahl	50.0%	50.0%	100.0%
	Weiblich	Klasse	1. Klasse	Anzahl	1	1	2
				Erwartete Anzahl	1.3	.7	2.0
				% innerhalb von Klasse	50.0%	50.0%	100.0%
				% innerhalb von Überlebt?	2.9%	5.0%	3.7%
				% der Gesamtzahl	1.9%	1.9%	3.7%
			2. Klasse	Anzahl	15	0	15
				Erwartete Anzahl	9.4	5.6	15.0
				% innerhalb von Klasse	100.0%	.0%	100.0%
				% innerhalb von Überlebt?	44.1%	.0%	27.8%
				% der Gesamtzahl	27.8%	.0%	27.8%
			3. Klasse	Anzahl	18	19	37
				Erwartete Anzahl	23.3	13.7	37.0
				% innerhalb von Klasse	48.6%	51.4%	100.0%
				% innerhalb von Überlebt?	52.9%	95.0%	68.5%
				% der Gesamtzahl	33.3%	35.2%	68.5%
			Gesamt	Anzahl	34	20	54
				Erwartete Anzahl	34.0	20.0	54.0
				% innerhalb von Klasse	63.0%	37.0%	100.0%
				% innerhalb von Überlebt?	100.0%	100.0%	100.0%
				% der Gesamtzahl	63.0%	37.0%	100.0%

Gesamt	Klasse	1. Klasse	Anzahl	6	1	7
			Erwartete Anzahl	3.9	3.1	7.0
			% innerhalb von Klasse	85.7%	14.3%	100.0%
			% innerhalb von Überlebt?	9.4%	2.0%	6.1%
			% der Gesamtzahl	5.3%	.9%	6.1%
	2. Klasse	Anzahl	26	0	26	
		Erwartete Anzahl	14.6	11.4	26.0	
		% innerhalb von Klasse	100.0%	.0%	100.0%	
		% innerhalb von Überlebt?	40.6%	.0%	22.8%	
		% der Gesamtzahl	22.8%	.0%	22.8%	
	3. Klasse	Anzahl	32	49	81	
		Erwartete Anzahl	45.5	35.5	81.0	
		% innerhalb von Klasse	39.5%	60.5%	100.0%	
		% innerhalb von Überlebt?	50.0%	98.0%	71.1%	
		% der Gesamtzahl	28.1%	43.0%	71.1%	
Gesamt		Anzahl	64	50	114	
		Erwartete Anzahl	64.0	50.0	114.0	
		% innerhalb von Klasse	56.1%	43.9%	100.0%	
		% innerhalb von Überlebt?	100.0%	100.0%	100.0%	
		% der Gesamtzahl	56.1%	43.9%	100.0%	

Abbildung 111: Viewervensterausgabe "Kreuztabelle"

Die obige sowie die zuvor wiedergegebene Kreuztabelle lassen sich z.B. nutzen, um eine Rangliste der Überlebenden der Titanic zu erstellen, die im Folgenden wiedergegeben ist:

Rangplatz	Segmentbeschreibung	Anteil der Geretteten in %
1	Kinder, 2. Klasse	100,0
2	Frauen, 1. Klasse	97,1
3	Frauen, 2. Klasse	87,4
4	Kinder, 1. Klasse	85,7
5	Frauen, 3. Klasse	50,9
6	Kinder, 3. Klasse	39,5
7	Männer, 1. Klasse	33,8
8	Besatzung der Titanic	29,5
9	Männer, 3. Klasse	15,0
10	Männer, 2. Klasse	7,6

Die abschliessende Rangliste verdeutlicht noch einmal die starken Effekte, welche von den Variablen Geschlecht, Kind oder Erwachsener und der Klassenzugehörigkeit an Bord der Titanic auf die Überlebensquote ausgehen.

Exkurs: Ende

2.2 Grafische Veranschaulichung von Kreuztabellen

Zur visuellen Darstellung von Kreuztabellen wählen Sie aus dem Menü

Diagramme | Diagrammerstellung...

Es öffnet sich die Dialogbox **Diagrammerstellung**. Wählen Sie aus der **Galerie** die Option **Balken**. Verschiedene Balkendiagramme werden daraufhin angezeigt. Wählen Sie **Gruppierte Balken** und ziehen Sie das Muster auf die Arbeitsfläche der Diagrammvorschau. Ziehen Sie die Variable **überlebt** auf das Feld für die X-Achse, die Variable **klasse** in das Feld für die Cluster- bzw. Gruppierungsvariable.

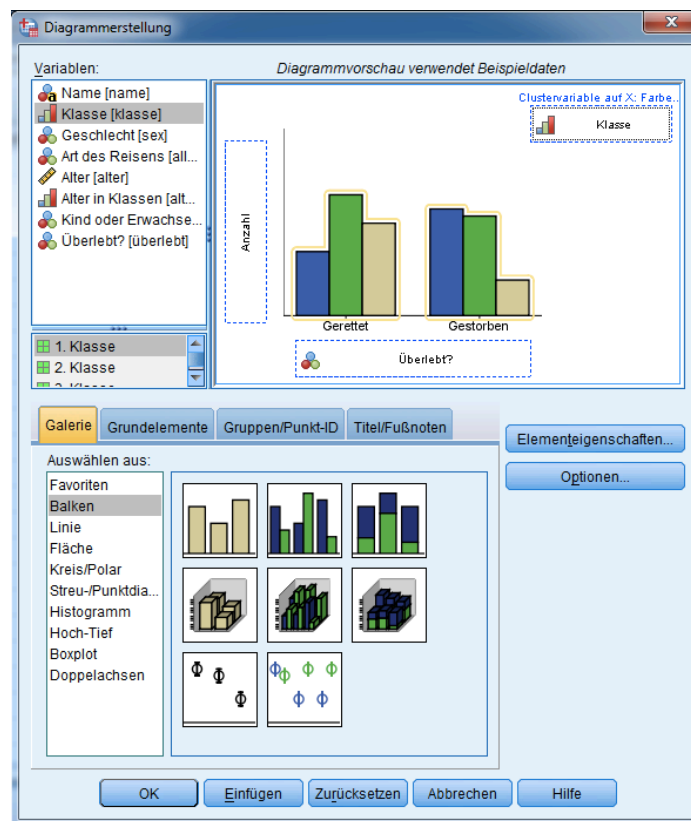


Abbildung 112: Ansicht "Diagrammerstellung"

Klicken Sie auf den Schalter **Elementeigenschaften**, falls die Dialogbox Elementeigenschaften noch nicht geöffnet sein sollte. Wählen Sie unter **Statistik** die Option **Prozentsatz**. Klicken Sie auf den Schalter **Parameter festlegen...** und bestimmen Sie als Nenner für die Berechnung des Prozentsatzes **Gesamt für jede Kategorie der LegendenvARIABLE**. Bestätigen Sie mit **Weiter** und **schließen** Sie die **Dialogbox Elementeigenschaften** per Klick auf den Schalter **Zuweisen**. Ziehen Sie die **Registerkarte Titel/Fussnoten**, klicken Sie auf das **Kästchen Titel 1** und tragen Sie in der sich öffnenden Dialogbox Elementeigenschaften unter Inhalt „Untergang der Titanic“ ein. Bestätigen Sie mit **Zuweisen**. Geben Sie wie oben beschrieben als **Titel 2** „Überlebensquote nach Klassenzugehörigkeit“ ein. Bestätigen Sie erneut mit **Zuweisen** und insgesamt abschließend mit **OK**. Übertragen Sie per Mausklick das Diagramm in den Diagramm-Editor und verschönern Sie ggf. die Grafik noch ein wenig. Sie erhalten folgende grafische Darstellung der Kreuztabelle:

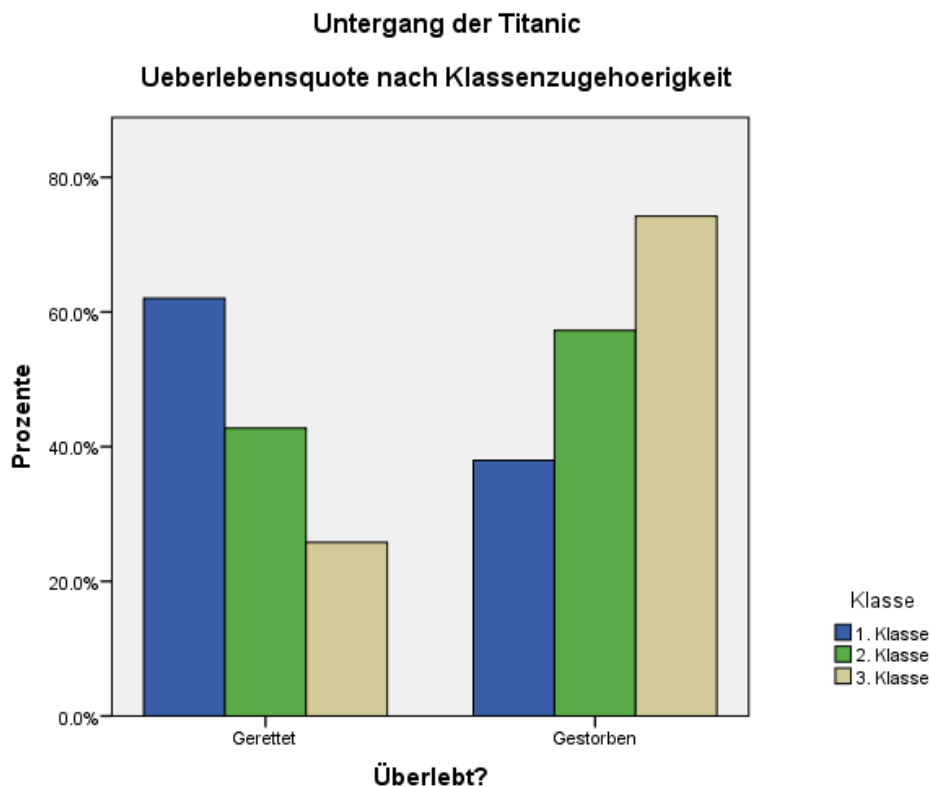


Abbildung 113: Viewerfensterausgabe “Diagramm Untergang der Titanic”

Die Dreiklassengesellschaft des viktorianischen England, welche sich bis in die Rettungsboote der Titanic erstreckte, illustriert die Grafik recht gut.

2.3 Der Chi-Quadrat-Test

Der Chi-Quadrat-Test überprüft die Unabhängigkeit der beiden Variablen der Kreuztabelle und damit indirekt den Zusammenhang der beiden Merkmale. Zwei Variablen einer Kreuztabelle gelten dann als voneinander unabhängig, wenn die beobachteten Häufigkeiten der einzelnen Zeilen mit den erwarteten Häufigkeiten übereinstimmen. Um den Test aufzurufen, wählen Sie

Analysieren | Deskriptive Statistiken | Kreuztabellen...

Übertragen Sie die Variable **klasse** in die **Zeilenliste**, die **Variable überlebt in die Spaltenliste**. Klicken Sie auf die **Schaltfläche Zellen...** Aktivieren Sie in der Dialogbox **Kreuztabellen: Zellen anzeigen** neben der voreingestellten Option **Beobachtet** die Optionen **Erwartet** und **Standardisiert**. Bestätigen Sie mit **Weiter**. Klicken Sie auf die Schaltfläche **Statistik...** Es öffnet sich die bereits beschriebene Dialogbox **Kreuztabellen: Statistik**. Aktivieren Sie die Option **Chi-Quadrat**. Bestätigen Sie anschliessend mit **Weiter** und klicken Sie in der Dialogbox **Kreuztabellen** auf **OK**. Sie erhalten folgende Kreuztabelle:

Klasse * Überlebt? Kreuztabelle

			Überlebt?		Gesamt
			Gerettet	Gestorben	
Klasse	1. Klasse	Anzahl	201	123	324
		Erwartete Anzahl	124.2	199.8	324.0
		Standardisierte Residuen	6.9	-5.4	
	2. Klasse	Anzahl	118	158	276
		Erwartete Anzahl	105.8	170.2	276.0
		Standardisierte Residuen	1.2	-9	
	3. Klasse	Anzahl	183	527	710
		Erwartete Anzahl	272.1	437.9	710.0
		Standardisierte Residuen	-5.4	4.3	
	Gesamt	Anzahl	502	808	1310
		Erwartete Anzahl	502.0	808.0	1310.0

Chi-Quadrat-Tests

	Wert	df	Asymptotisch e Signifikanz (2-seitig)
Chi-Quadrat nach Pearson	126.679 ^a	2	.000
Likelihood-Quotient	126.533	2	.000
Zusammenhang linear- mit-linear	126.465	1	.000
Anzahl der gültigen Fälle	1310		

a. 0 Zellen (.0%) haben eine erwartete Häufigkeit kleiner 5. Die minimale erwartete Häufigkeit ist 105.76.

Die zweite Tabelle zeigt die Ergebnisse des Chi-Quadrat-Tests. Die übliche Formel zur Berechnung des Chi-Quadrat-Werts ist diejenige nach Pearson.¹¹

Der **Chi-Quadrat-Wert** stellt die Summe der Quadrate der standardisierten Residuen dar, die über alle Felder der Kreuztabelle gebildet wird. Die **Felder der Kreuztabelle mit hohen standardisierten Residuen liefern demnach einen hohen Beitrag zum Chi-Quadrat-Wert** und damit zu einem signifikanten Ergebnis. Nach einer Faustregel zeigt ein standardisiertes **Residuum von 2 oder grösser** eine signifikante Abweichung der beobachteten von der erwarteten Häufigkeit an. Je grösser die Abweichung in einem Feld der Tabelle ist, umso grösser fällt auch das Chi-Quadrat aus. Ein grosses Chi-Quadrat ist mit grossen Abweichungen verbunden und deutet auf einen Zusammenhang zwischen den Variablen hin.¹²

¹¹ Handelt es sich bei der Kreuztabelle um eine Vierfeldertafel und ist eine erwartete Häufigkeit kleiner als 5, so wird zusätzlich der exakte Test nach Fisher ausgeführt.

¹² Für jedes Feld der Tabelle werden die quadrierten Abweichungen der erwarteten von den tatsächlichen Häufigkeiten durch die erwarteten Häufigkeiten dividiert. Die Quotienten werden anschliessend addiert. Das Quadrieren sorgt dafür, dass negative wie positive Abweichungen gleichermassen in das Mass eingehen und sich nicht wechselseitig aufheben. Die Division durch die erwarteten Häufigkeiten erfolgt, da sich sonst bei vielen Beobachtungen auch mehr Abweichungen ergeben würden.

Im **obigen Beispiel** ergibt sich nach Pearson ein **höchst signifikanter Chi-Quadrat-Wert von 126.679 ($p < 0,001$)**, d.h. die Variablen **klasse** und **überlebt** sind nicht unabhängig voneinander. Betrachtet man die standardisierten Residuen in den einzelnen Feldern der Kreuztabelle, so erkennt man, dass diese Signifikanz in den Feldern der ersten und zweiten Klasse (jeweils gerettet und gestorben) begründet liegt. Bei Passagieren der ersten Klasse ist die Überlebenswahrscheinlichkeit deutlich erhöht, während sie bei der dritten Klasse deutlich zu niedrig ist. Anhand der standardisierten Residuen zeigt sich noch einmal die Mittellage der zweiten Klasse.

3. T-Tests (Mittelwertvergleiche)

- Frage: Unterscheidet sich der Mittelwert zweier Gruppen in Bezug auf eine Variable?
- Beispiel: Unterscheiden sich Studenten und Dozenten in Bezug auf die durchschnittliche Dauer des Fernsehkonsums?
- Die abhängige Variable (intervallskaliert) wird auf seine unterschiedlichen Mittelwerte in den beiden Gruppen untersucht.

Fall Hypertoniestudie (Datei: hyper.sav)

3.1 Abhängigkeit versus Unabhängigkeit von Stichproben

Zwei Stichproben sind dann voneinander abhängig, wenn jedem Wert der einen Stichprobe auf sinnvolle und eindeutige Weise genau ein Wert der anderen Stichprobe zugeordnet werden kann; Entsprechendes gilt für den Fall von mehreren abhängigen Stichproben. Meist handelt es sich dabei um den Fall, dass eine Messung zu mehreren Zeitpunkten durchgeführt wurde; die Werte zu den verschiedenen Zeitpunkten führen dann zu abhängigen Stichproben. In SPSS Statistics werden abhängige (auch gebundene, gepaarte) Stichproben durch verschiedene Variablen repräsentiert, die am gleichen Kollektiv von Fällen in den entsprechenden Tests einander gegenübergestellt werden.

Ist eine sinnvolle und eindeutige Wertezuordnung zwischen den Stichproben nicht möglich, liegen unabhängige Stichproben vor. In SPSS Statistics beinhalten unabhängige Stichproben unterschiedliche Fälle (z.B. verschiedene Probanden), die in der Regel mit Hilfe einer nominalskalierten Gruppierungsvariablen unterschieden werden.

3.2 T-Test für unabhängige Stichproben

Beim t-Test für unabhängige Stichproben wird untersucht, ob sich zwei Gruppen (zwei Teilgruppen in der Stichprobe) in ihren Mittelwerten überzufällig unterscheiden. Es wird geprüft, ob die Mittelwerte in einer der beiden Gruppen systematisch höher sind als in der anderen.

Der t-Test setzt zwingend Intervallskalenniveau bei der untersuchten (abhängigen) Variable voraus. Theoretisch hat der t-Test noch zwei weitere Voraussetzungen: Normalverteilung des Merkmals in der Grundgesamtheit¹³ und Homogenität der Varianzen in der Stichprobe.

¹³ Praktisch kann diese Bedingung vernachlässigt werden, da man die Verteilung der Grundgesamtheit üblicherweise nicht kennt, sondern nur die Verteilung der Stichprobe. Der t-Test ist sehr

Fallbeispiel:

Zwei Gruppen von Patienten mit unterschiedlicher Herzmedikation (Alphasan und Betasan) werden untersucht, um festzustellen, welches Medikament die Lebensdauer am meisten verlängert. Ein Arzt behauptet, dass viele Patienten der einen Gruppe schon sehr alt seien, was natürlich die Ergebnisse beeinflussen würde. Der zuständige Forscher will das nicht auf sich sitzen lassen.

Unterscheiden sich die beiden Medikamentengruppen signifikant bzgl. des Alters?

Mit dem t-Test für unabhängige Stichproben soll nun untersucht werden, ob sich die beiden Medikamentengruppen signifikant bzgl. des Alters unterscheiden. Falls dies zutrifft, könnten mögliche Wirkungsunterschiede zwischen den beiden Medikamenten auch auf einen unterschiedlichen Altersdurchschnitt zurückgeführt werden. Zunächst soll geprüft werden, ob die Variable **Alter** normalverteilt ist. Mit dem **Kolmogorov-Smirnov-Test** lässt sich die Verteilung einer Variablen auf Normalverteilung¹⁴ überprüfen. Wählen Sie im Menü

Analysieren | Nichtparametrische Tests | Alte Dialogfelder | K-S bei einer Stichprobe...

Die Prüfung auf **Normalverteilung** ist voreingestellt. Bringen Sie die Variable **alter** in das Testvariablenfeld und klicken Sie auf **OK**.

robust gegen Verletzungen dieser Voraussetzung, d.h. er funktioniert auch bei nicht normalverteilten Merkmalen sehr gut und kommt zu identischen Entscheidungen über die Annahme oder Verwerfung von H_0 .

¹⁴ Die Normalverteilung ist eine Verteilung, bei der sich die meisten Werte um den Mittelwert gruppieren, während die Häufigkeiten nach beiden Seiten gleichmässig abfallen.

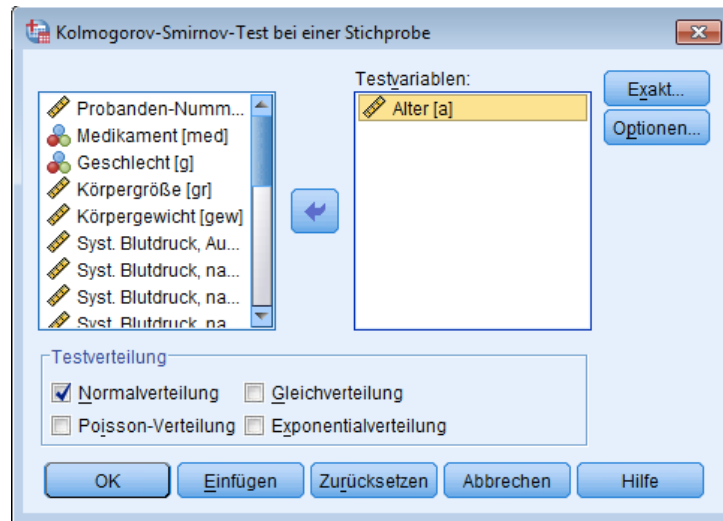


Abbildung 114: Ansicht "Kolmogorov-Smirnov-Test bei einer Stichprobe"

Die Ergebnisse werden im Viewer angezeigt:

Kolmogorov-Smirnov-Anpassungstest		
		Alter
N		174
Parameter der Normalverteilung ^{a, b}	Mittelwert	62.11
	Standardabweichung	11.548
Extremste Differenzen	Absolut	.059
	Positiv	.055
	Negativ	-.059
Kolmogorov-Smirnov-Z		.785
Asymptotische Signifikanz (2-seitig)		.569

a. Die zu testende Verteilung ist eine Normalverteilung.
b. Aus den Daten berechnet.

Die Ausgabe umfasst den Mittelwert (62.11), die Standardabweichung (11.548), die beim K-S-Test anfallenden Zwischenergebnisse sowie die Irrtumswahrscheinlichkeit $p = 0.569$. Mit $p = 0.569 > 0.05$ liegt ein deutlich nicht signifikanter p-Wert vor. Die Werte sind also hinreichend normalverteilt.¹⁵

¹⁵ Eine weitere Möglichkeit zur Überprüfung auf Normalverteilung bietet die Erstellung von Normalverteilungsplots, in denen die beobachteten den unter Normalverteilung zu erwartenden Werten gegenübergestellt werden (siehe Exkurs weiter unten).

Die hinreichende Normalverteilung der Werte der Variable **alter** lässt sich mit einem Histogramm mit Normalverteilungskurve visualisieren. Wählen Sie dazu

Analysieren | Deskriptive Statistiken | Häufigkeiten...

Verschieben Sie die Variable **Alter** in das Variablenfeld und klicken Sie auf die Schaltfläche **Diagramme...** Dort können Sie als Diagrammtyp **Histogramm** auswählen und zudem die **Normalverteilungskurve** anzeigen lassen.

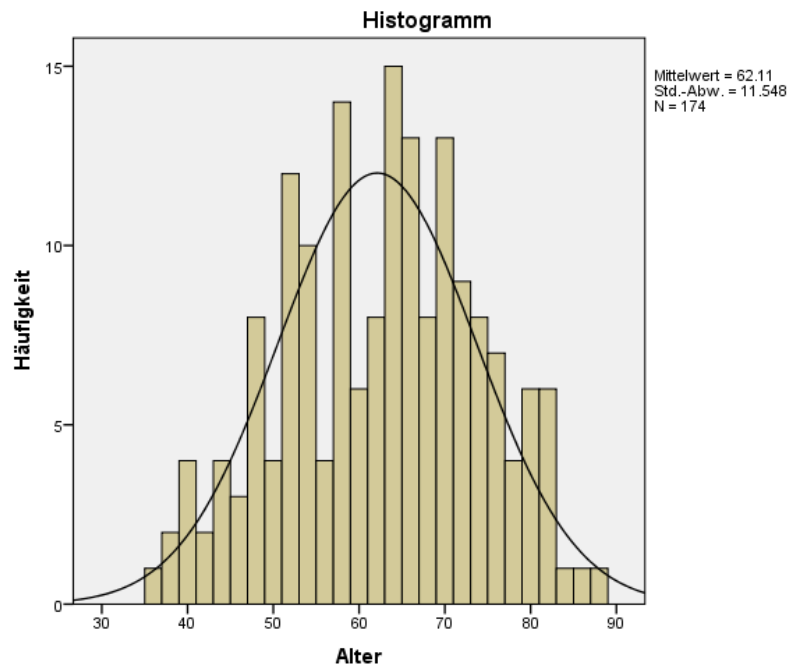


Abbildung 115: Vieweransicht "Histogramm mit Normalverteilungskurve der Variable Alter"

Um den t-Test für unabhängige Stichproben durchzuführen, wählen Sie nun aus dem Menü

Analysieren | Mittelwerte vergleichen | T-Test bei unabhängigen Stichproben...

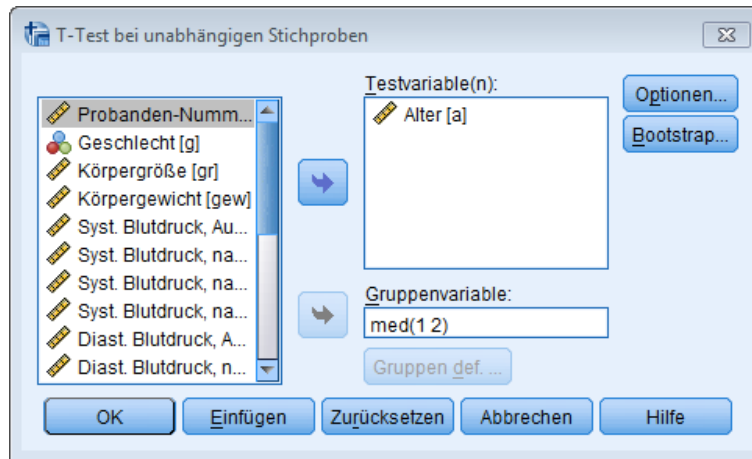


Abbildung 116: Ansicht "T-Test bei unabhängigen Stichproben"

Klicken Sie in der Dialogbox **T-Test bei unabhängigen Stichproben** im Quellvariablenfeld auf die Variable **alter** und bringen Sie diese durch einen Klick auf die Transportschaltfläche in das **Testvariablenfeld**. Übertragen Sie zudem die **Gruppierungsvariable med** in das Gruppenvariablenfeld. Klicken Sie auf **Gruppen def. ...** um die beiden relevanten Werte der Gruppierungsvariablen einzugeben: Geben Sie bei Gruppe 1 die Zahl **1** (steht für med = 1) und bei Gruppe 2 die Zahl **2** (steht für med = 2) ein. Klicken Sie auf **Weiter** und **OK**. Im Viewer erscheinen die folgenden Ergebnisse:

Gruppenstatistiken					
	Medikament	N	Mittelwert	Standardabweichung	Standardfehler des Mittelwertes
Alter	Alphasan	87	62.24	11.191	1.200
	Betasan	87	61.98	11.958	1.282

Test bei unabhängigen Stichproben										
		Levene-Test der Varianzgleichheit		T-Test für die Mittelwertgleichheit						
		F	Signifikanz	T	df	Sig. (2-seitig)	Mittlere Differenz	Standardfehler der Differenz	95% Konfidenzintervall der Differenz	
Alter	Varianzen sind gleich	.543	.462	.151	172	.880	.264	1.756	-3.201	3.730
	Varianzen sind nicht gleich			.151	171.249	.880	.264	1.756	-3.202	3.730

In der ersten Tabelle werden Fallzahl, Mittelwert, Standardabweichung und Standardfehler des Mittelwertes in beiden Gruppen ausgegeben. Die annähernd gleichen Mittelwerte (62.24 bzw. 61.98) deuten auf einen geringen Unterschied zwi-

schen den Stichproben hin. Auch die Standardabweichungen sind nahezu identisch, was auf Varianzhomogenität hindeutet.

Die zweite Tabelle enthält links den Levene-Test auf Gleichheit der Varianzen (Test auf Modellvoraussetzungen). Die Gleichheit der Varianzen (Varianzenhomogenität) wird verworfen, wenn der Levene-Test ein $p < 0,05$ ergibt (Varianzenheterogenität). In diesem Fall ergibt sich ein $p = 0.462 > 0.05$, d.h. es ist von einer Gleichheit der Varianzen auszugehen und es sind im Rahmen des T-Tests für die Mittelwertgleichheit die Werte der ersten Zeile abzulesen. Die Ausgabe umfasst: Das Ergebnis des t-Tests: Prüfgrösse t, Anzahl der Freiheitsgrade (df), Irrtumswahrscheinlichkeit p (Sig. 2-seitig), und die Differenz der beiden Mittelwerte, deren Standardfehler und ein zugehöriges Konfidenzintervall.¹⁶

In diesem Beispiel weist die Prüfgrösse einen Wert von $T = 0.151$ auf. Zwischen den beiden Medikamentengruppen Alphasan und Betasan besteht mit $p = 0.880 > 0.05$ kein signifikanter Altersunterschied (auf dem 5%-Signifikanzniveau).

Exkurs: Ausgabe von Häufigkeitsverteilungen mit Boxplots

Boxplots visualisieren wie die Werte einer Variable verteilt sind und vermitteln einen ersten Eindruck darüber, ob die Verteilung symmetrisch oder schief ist. Erstellen Sie einen Boxplot für die Variable **alter** der Datei **hyper.sav**. Gehen Sie wie folgt vor: Wählen Sie

Analysieren | Deskriptive Statistiken | Explorative Datenanalyse...

und bringen Sie die Variable **alter** in das Feld **Abhängige Variablen**. Klicken Sie auf die Schaltfläche **Statistiken...** und wählen Sie zusätzlich zur Vorauswahl **Perzentile** aus. Klicken Sie dann auf **Weiter** und **OK**.

¹⁶ Die Interpretation steckt in der Signifikanzaussage – alle anderen Werte sind reine Zwischenwerte auf dem Weg zur Signifikanzberechnung und sagen selber eigentlich nichts aus.

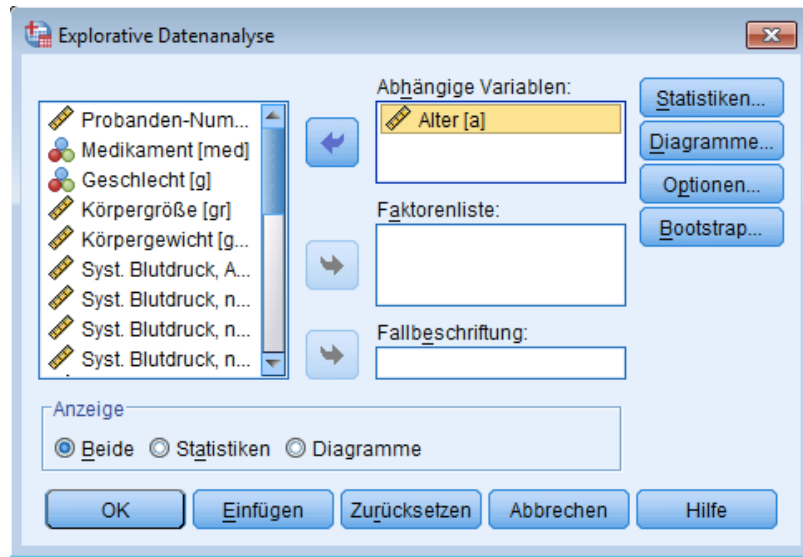


Abbildung 117: Ansicht "Explorative Datenanalyse"

In der Ausgabe erhalten Sie u.a. die **deskriptive Statistik** sowie die **Perzentile** für die Variable **alter** und einen **Boxplot**.

Deskriptive Statistik

			Statistik	Standardfehler
Alter	Mittelwert		62.11	.875
	95% Konfidenzintervall des Mittelwerts	Untergrenze	60.38	
		Obergrenze	63.84	
	5% getrimmtes Mittel		62.25	
	Median		63.00	
	Varianz		133.358	
	Standardabweichung		11.548	
	Minimum		36	
	Maximum		87	
	Spannweite		51	
	Interquartilbereich		17	
	Schiefe		-.143	.184
	Kurtosis		-.635	.366

Perzentile

		Perzentile					
		5	10	25	50	75	95
Gewichtetes Mittel (Definition 1)	Alter	42.00	47.00	53.00	63.00	70.25	81.00
Tukey-Angelpunkte	Alter			53.00	63.00	70.00	

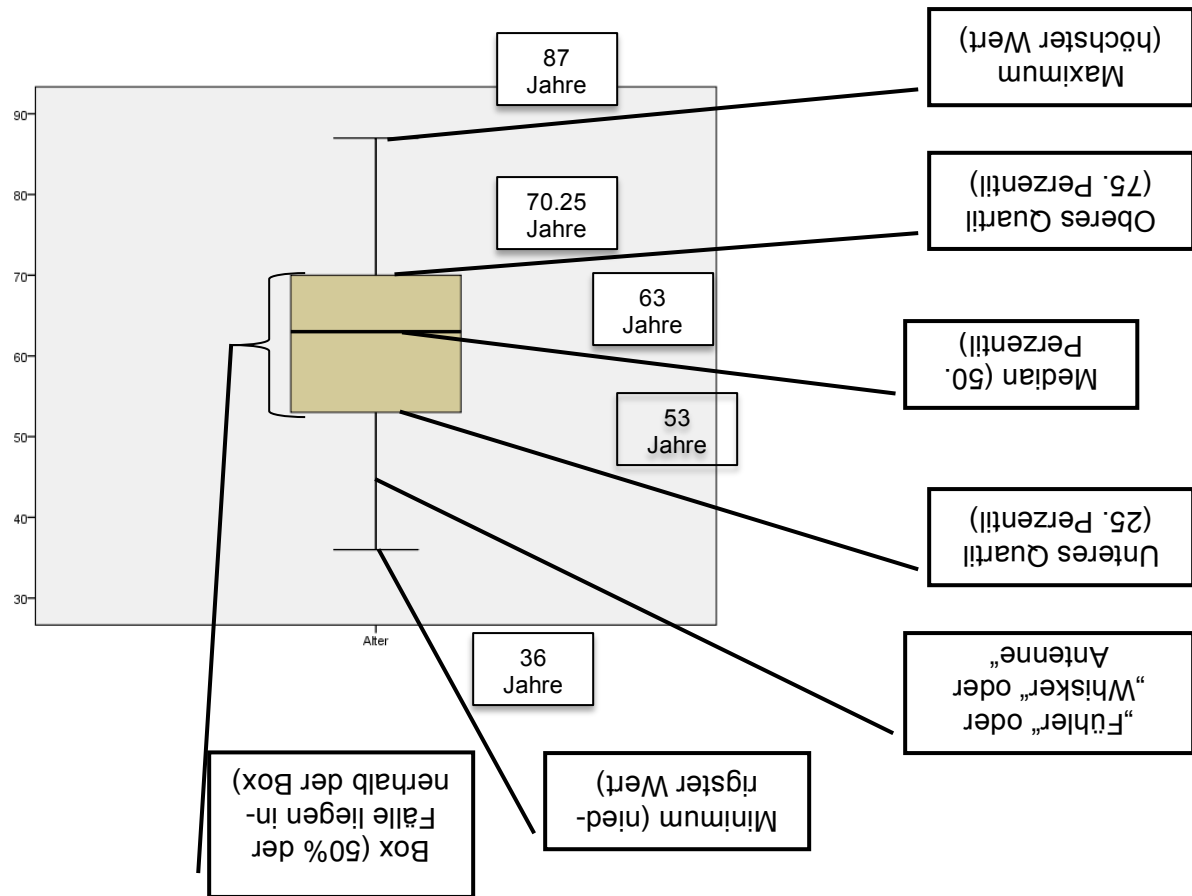


Abbildung 118: Vieweransicht "Boxplot der Variable Alter"

Der Boxplot besteht aus einer Box, die vom ersten und dritten Quartil (25. und 75. Perzentil) begrenzt wird und deren innere Linie den Median repräsentiert. Ausreisser-Werte ober- und unterhalb des Boxplots werden als kleine Kreise angezeigt, Extremwerte als Sternchen.¹⁷ Die Abstände zwischen den Quartilen, dem Minimum und Maximum sowie dem Median sind in diesem Beispiel relativ symmetrisch, was auf eine Normalverteilung der Variable **alter** hindeutet. Bei einer Normalverteilung wird erwartet, dass 95% der Werte innerhalb der Fühler liegen.

In diesem Beispiel sind Modus und Median identisch (jeweils 63 Jahre). Der Mittelwert liegt mit 62.11. Damit fallen alle drei Masse der zentralen Tendenz zu-

¹⁷ Ausreisser-Werte sind Werte, die mehr als 1,5 Kastenlängen ausserhalb liegen. Extremwerte sind Werte, die mehr als 3 Kastenlängen ausserhalb liegen.

sammen und liegen sehr nahe beieinander. Nachfolgende Übersicht verdeutlicht den Einfluss der Lageparameter Modus, Median und Mittelwert auf die Form der Verteilung:

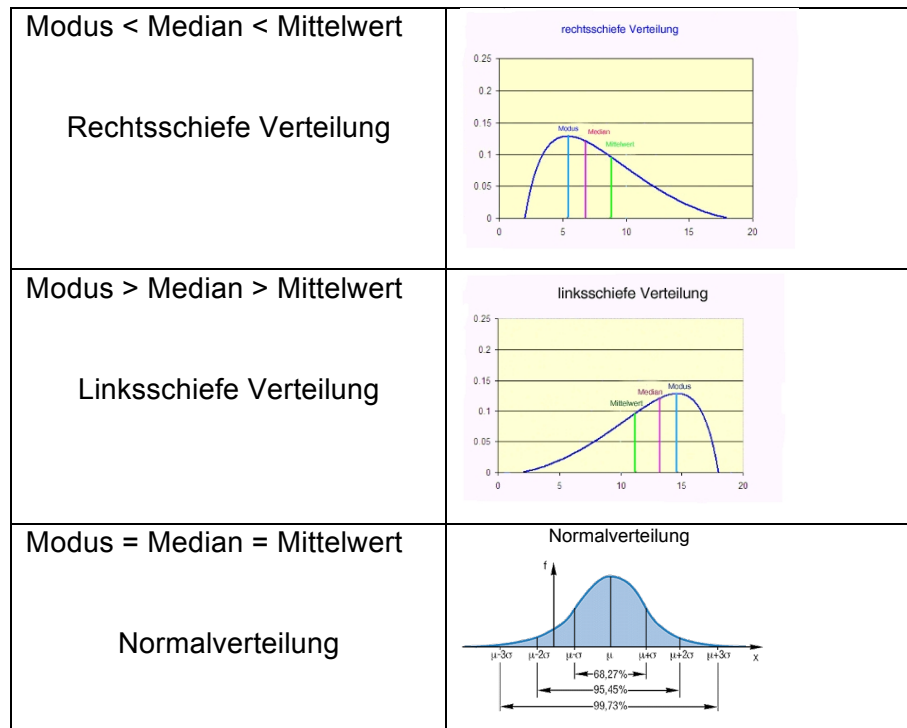


Abbildung 119: Verteilungsformen

Exkurs Ende

3.3 T-Test für abhängige (gepaarte/verbundene) Stichproben

Beim t-Test für abhängige Stichproben geht es um Mittelwertunterschiede zwischen zwei Messwerten desselben Untersuchungsobjekts (z.B. Vorher-Nachher-Vergleich, Untersuchung von Paaren wie Zwillingspaare, Ehepaare). Mit dem t-Test für abhängige Stichproben soll untersucht werden, ob die beiden Messwerte, die gepaart vorhanden sind, sich systematisch voneinander unterscheiden. Die Messwerte müssen auf Intervallskalenniveau vorliegen und die zwei

Variablen müssen dasselbe Merkmal messen und paarweise miteinander verbunden sein.

Fallbeispiel:

Die Patienten mit Herzmedikation (Alphasan und Betasan) werden nach einem Monat untersucht, um festzustellen, ob sich der Cholesterinwert signifikant verändert hat.

- a) Überprüfen Sie, ob sich bei den Patienten mit Herzmedikation eine signifikante Veränderung ergeben hat.
- b) Wiederholen Sie die Berechnung nur für diejenigen Probanden, die das Medikament Alphasan eingenommen haben.

Ad 3a): Wählen Sie aus dem Menü

Analysieren | Mittelwerte vergleichen | T-Test bei gepaarten Stichproben...

Übertragen Sie die Variable **chol0** in das Feld für die erste Variable des ersten Paares und **chol1** in das Feld für die zweite Variable und bestätigen Sie mit **OK**. Im Viewer erscheinen die folgenden Ergebnisse:

Statistik bei gepaarten Stichproben

		Mittelwert	N	Standardabweichung	Standardfehler des Mittelwertes
Paaren 1	Cholesterin, Ausgangswert	237.27	174	49.421	3.747
	Cholesterin, nach 1 Monat	239.20	174	49.510	3.753

Korrelationen bei gepaarten Stichproben

		N	Korrelation	Signifikanz
Paaren 1	Cholesterin, Ausgangswert & Cholesterin, nach 1 Monat	174	.861	.000

Test bei gepaarten Stichproben

		Gepaarte Differenzen				T	df	Sig. (2-seitig)	
		Mittelwert	Standardabweichung	Standardfehler des Mittelwertes	95% Konfidenzintervall der Differenz				
					Untere				Obere
Paaren 1	Cholesterin, Ausgangswert - Cholesterin, nach 1 Monat	-1.925	26.085	1.978	-5.828	1.978	-.974	173	.332

In der ersten Tabelle werden wiederum Mittelwert, Standardabweichung und Standardfehler des Mittelwertes beider Variablen ausgegeben. Die annähernd gleichen Mittelwerte (237.27 bzw. 239.20) deuten auf einen geringen Unterschied zwischen den Stichproben hin. Die Standardabweichungen deuten auf Varianzhomogenität hin. Die zweite Tabelle enthält Angaben zum Korrelationskoeffizienten (Produkt-Moment-Korrelation nach Pearson) zwischen den beiden Variablen und seine Absicherung gegen Null. Dass zwei gepaarte Messwerte hoch miteinander korrelieren, ist fast schon eine Tautologie und deshalb eine ziemlich uninteressante Information im Rahmen eines t-Tests. Die dritte Tabelle umfasst Mittelwert, Standardabweichung, Standardfehler des Mittelwertes, Konfidenzintervall der Wertedifferenzen und das Ergebnis des t-Test: Die Prüfgrösse t , die Anzahl der Freiheitsgrade (df) und die Irrtumswahrscheinlichkeit p (Sig. 2-seitig).

Für die Prüfgrösse ergibt sich ein Wert von $t = 0.974$. Es ergibt sich keine signifikante Änderung des Cholesterins vom Ausgangswert zum Wert nach einem Monat ($p = 0.332 > 0.05$), d.h. der Mittelwertunterschied von 1.93 ist auf dem 5%-Niveau nicht signifikant.

Ad 3b): Wählen Sie aus dem Menü

Daten | Fälle auswählen...

und aktivieren Sie die Option **Falls Bedingung zutrifft**. Geben Sie als Auswahlbedingung **med=1** ein und klicken Sie auf **Weiter** und **OK**. Starten Sie dann die Berechnung des t-Tests für zwei abhängige Stichproben. Der t-Test wird nunmehr nur für diejenigen Fälle ausgeführt ($N = 87$), die der ersten Medikamentengruppe ($med=1$) angehören. Sie erhalten nachfolgende Ausgabe:

Statistik bei gepaarten Stichproben

		Mittelwert	N	Standardabweichung	Standardfehler des Mittelwertes
Paaren 1	Cholesterin, Ausgangswert	240.67	87	49.832	5.343
	Cholesterin, nach 1 Monat	241.60	87	48.545	5.205

Korrelationen bei gepaarten Stichproben

		N	Korrelation	Signifikanz
Paaren 1	Cholesterin, Ausgangswert & Cholesterin, nach 1 Monat	87	.917	.000

Test bei gepaarten Stichproben

		Gepaarte Differenzen				T	df	Sig. (2-seitig)	
		Mittelwert	Standardabweichung	Standardfehler des Mittelwertes	95% Konfidenzintervall der Differenz				
					Untere				Obere
Paaren 1	Cholesterin, Ausgangswert - Cholesterin, nach 1 Monat	-.931	20.034	2.148	-5.201	3.339	-.433	86	.666

Mit einer Prüfgrösse von $T = 0.433$ und $p = 0.666 > 0.05$ erhalten Sie wieder ein nicht signifikantes Ergebnis.

Um ggf. die nachfolgenden Analysen wieder für alle Fälle ausführen zu können, aktivieren Sie erneut die Dialogbox **Fälle auswählen** und dort die Option **Alle Fälle**.

3.4 Einstichproben-t-Test

Mit dem Einstichproben-t-Test können Sie überprüfen, ob ein aus einer gegebenen Stichprobe gewonnener Mittelwert sich von einem vorgegebenen Testwert unterscheidet.

Fallbeispiel:

Testen Sie, ob sich in der Hypertoniestudie der Ausgangswert des Cholesterins im Mittel von dem Wert 229, der sich in einer anderen Studie ergeben hat, unterscheidet.

Wählen Sie aus dem Menü

Analysieren | Mittelwerte vergleichen | T-Test bei einer Stichprobe...

Bringen Sie die Variable **chol0** in das Testvariablenfeld und tragen Sie als Testwert den Wert **229** ein. Über die Schaltfläche Optionen... kann anstelle des 95-% Konfidenzintervalls auch ein anderes eingegeben werden (zwischen 1 und 99). Lassen Sie die Standardeinstellung unverändert und starten Sie die Berechnungen mit **OK**.



Abbildung 120: Ansicht "T-Test bei einer Stichprobe"

Im Viewer erscheinen folgende Tabellen:

Statistik bei einer Stichprobe

	N	Mittelwert	Standardabweichung	Standardfehler des Mittelwertes
Cholesterin, Ausgangswert	174	237.27	49.421	3.747

Test bei einer Stichprobe

	Testwert = 229					
	T	df	Sig. (2-seitig)	Mittlere Differenz	95% Konfidenzintervall der Differenz	
					Untere	Obere
Cholesterin, Ausgangswert	2.207	173	.029	8.270	.88	15.66

Die im Viewer erscheinenden Ergebnisse zeigen, dass sich in der vorliegenden Studie ein mittlerer Ausgangswert des Cholesterins von 237.27 ergibt, der sich auf dem 5%-Niveau signifikant ($p = 0.029$) vom Vergleichstestwert 229 unterscheidet.

4. Korrelationen

- Frage: Wie stark ist der Zusammenhang von zwei Variablen?
- Beispiel: Bedeutet höheres Einkommen auch gleichzeitig mehr Urlaubsreisen?
- Bei metrischen Variablen: Pearson-Korrelation, bei ordinalen und nominalen Variablen: Kendall-tau-b oder Spearman-Korrelation

Fall ebay-Auktionen von Handys (Datei: ebay.sav)

4.1 Korrelationsrechnung nach Pearson (Produkt-Moment-Korrelation)

Die Korrelationsrechnung nach Pearson sagt anhand eines Kennwertes aus, wie stark zwei eindeutig intervallskalierte Merkmale zusammenhängen (Ausmass des linearen Zusammenhangs).

Korrelationen beschreiben die Enge eines Zusammenhangs anhand des Korrelationskoeffizienten, der Werte zwischen -1 und +1 annehmen kann. Ein Wert von 1 (hundertprozentig gleichgerichtet) gibt einen perfekten Zusammenhang im Sinne von "je höher a desto höher ist b" an. Ein Wert von -1 (hundertprozentig antagonistisch) gibt ebenfalls einen perfekten Zusammenhang im Sinne von "je grösser a, desto kleiner b" an. Ein Wert von 0 bedeutet, die beiden Variablen sind perfekt voneinander unabhängig. Grafisch bedeutet das: Wie eng streut die Punktwolke um die Gerade?

Die üblichen sprachlichen Kennzeichnungen für die Höhe eines Korrelationskoeffizienten (r) sind (siehe Cohen 1988, S. 115):

- Schwacher Zusammenhang: < 0.30
- Mittlerer Zusammenhang: $0.30-0.49$
- Starker Zusammenhang: $0.50-1$

Eine Aussage darüber, wie stark die eine Variable durch die andere determiniert ist, erhält man jedoch nicht durch den Korrelationskoeffizienten r selber, sondern erst mit der Quadrierung des Korrelationskoeffizienten (r^2). Der quadrierte Korrelationskoeffizient r^2 ergibt den Determinationskoeffizienten, auch Varianzaufklärung genannt. Die Varianzaufklärung ist ein direktes lineares Mass für das Ausmass des Zusammenhangs. Beispielsweise bedeutet ein Korrelationskoeffizient von 0.40 nicht eine Varianzaufklärung von 40%, sondern nur von 16%, wie nachfolgende Tabelle veranschaulicht. Die Tabelle zeigt auch, dass Unterschiede

zwischen Korrelationskoeffizienten im Bereich hoher Korrelationen sehr viel bedeutsamer sind als im Bereich geringer Korrelationen.

Korrelationskoeffizient	Varianzaufklärung (Determination)
.10	.01
.20	.04
.30	.09
.40	.16
.50	.25
.60	.36
.70	.49
.80	.64
.90	.81
1.00	1.00

Fallbeispiel:

Untersuchen Sie mithilfe der Datendatei ebay.sav, ob es einen Zusammenhang zwischen dem Startpreis, den der Verkäufer gesetzt hat, und dem erzielten Endpreis gibt. Beide Angaben sind in Euro, also intervallskaliert.

Um eine Korrelationsrechnung durchzuführen wählen Sie

Analysieren | Korrelation Bivariat...

Bringen Sie die Variablen **Startpreis** und **Endpreis** in das Variablenfenster. Alle anderen Einstellungen können unverändert beibehalten werden.

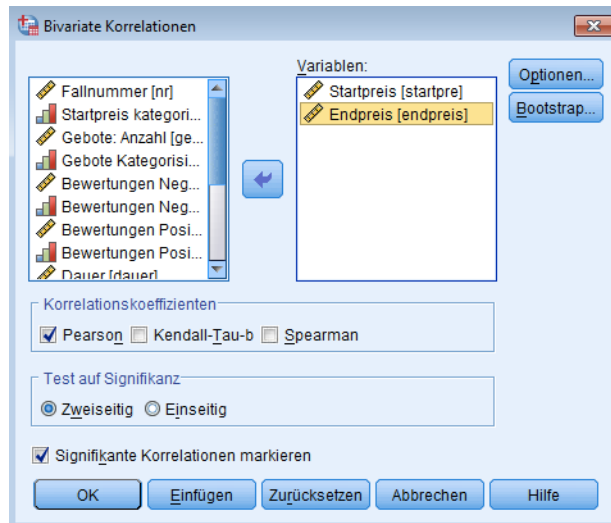


Abbildung 121: Ansicht "Bivariate Korrelationen"

Klicken Sie **OK** um die Berechnung durchzuführen. Als Ergebnis erhalten Sie eine Tabelle, in der in einer Matrix die Korrelationskennwerte und die zugehörigen Signifikanzberechnungen und das N (Anzahl der Fälle mit gültigen Werten in beiden Variablen) ausgegeben werden.¹⁸

Korrelationen		Startpreis	Endpreis
Startpreis	Korrelation nach Pearson	1	-.717**
	Signifikanz (2-seitig)		.000
	N	391	391
Endpreis	Korrelation nach Pearson	-.717**	1
	Signifikanz (2-seitig)	.000	
	N	391	391

** Die Korrelation ist auf dem Niveau von 0,01 (2-seitig) signifikant.

Folgende für die Interpretation wesentlichen Kennwerte werden angezeigt:

¹⁸ Standardmässig wird in SPSS Statistics bei Korrelationen jede Variable mit jeder anderen Variablen korreliert und in der Tabelle der Ergebnisausgabe dargestellt. Daher wird auch die Korrelation der Variablen mit sich selbst in der Diagonalen der obigen Tabelle ausgegeben. Diese Korrelation muss logischerweise immer 1 ergeben. Ausserdem taucht jede Korrelation immer doppelt auf, weil jede Beziehung in jede Richtung untersucht wird (hier: Startpreis mit Endpreis und Endpreis mit Startpreis).

- Der Stichprobenumfang, auf dem die Berechnung der Korrelation und der Signifikanz beruht.
- Die Höhe des Korrelationskoeffizienten und sein Vorzeichen – damit wird die Richtung und die Stärke des Zusammenhangs angegeben.
- Die Signifikanzaussage – damit wird die Wahrscheinlichkeit angegeben, dass diese Korrelation nur rein zufällig von 0 abweicht.

Die Matrix zeigt eine starke negative Korrelation der beiden Parameter Start- und Endpreis in Höhe von -0.72. Je höher der Startpreis angesetzt wird, desto niedriger ist in der Tendenz der Endpreis. Die Korrelation -0.72 ergibt ein r^2 von 0.514, also eine Varianzaufklärung von 51,4%, d.h. der erzielte Endpreis ist zu 51,4% vom Startpreis determiniert, die anderen knapp 50% der Variabilität im Endpreis hängen mit anderen Einflussfaktoren zusammen. Die Wahrscheinlichkeit für reinen Zufall ist $p < 0.001$. Das Ergebnis ist damit höchst signifikant und man kann sagen, dass hier ein systematischer Zusammenhang besteht.

Das folgende Streudiagramm stellt den Zusammenhang grafisch dar (Diagramm | Diagrammerstellung... | Streu-/Punktdiagramme; vgl. auch Seite 176 ff.):

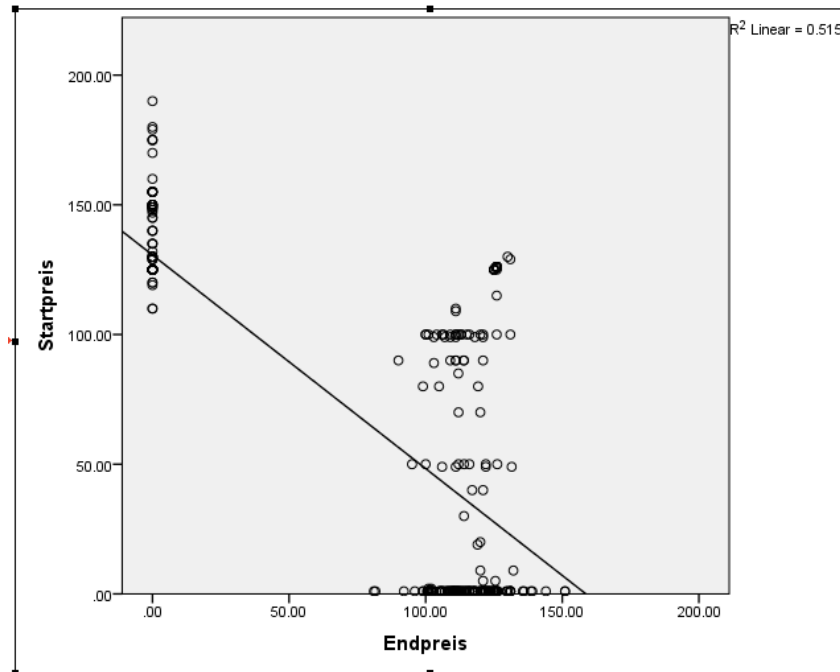


Abbildung 122: Vieweransicht “Streudiagramm der Variablen Start- und Endpreis”

4.2 Korrelationsrechnung bei einer Rangkorrelation

Eine Rangkorrelation ist zu berechnen, wenn eines der beiden Merkmale oder auch beide Merkmale ordinalskaliert sind oder wenn nichtnormalverteilte intervallskalierte Variablen vorliegen. Die Korrelationsrechnung gibt mit einem Kennwert an, wie eng der lineare Zusammenhang zwischen den Merkmalen ist. Die Voraussetzung für die Rangkorrelation ist, dass beide beteiligte Variable mindestens ordinal skaliert sind. Im Gegensatz zur Korrelationsrechnung nach Pearson darf bei der Rangkorrelation nicht von Varianzaufklärung gesprochen werden, da eine Varianz bei reinen Rangdaten nicht definiert ist. Die Bedeutung des r^2 als Determinationskoeffizient ist ansonsten dieselbe wie bei 4.1.

Fallbeispiel:

Untersuchen Sie mithilfe der Datendatei ebay.sav, ob der Endpreis von der Dauer der Auktion abhängt (Variable: dauer_o). Die Dauer einer Auktion bei ebay beträgt entweder 3, 5, 7 oder 10 Tage.

Wählen Sie im Menü

Analysieren | Korrelation Bivariat...

Wählen Sie die Variablen **Endpreis** und **Dauer (ordinal skaliert)** aus und bringen Sie sie in das Variablenfeld. Kreuzen Sie **Spearman** als Korrelationskoeffizienten an und entfernen Sie das Häkchen bei Pearson.¹⁹ Klicken Sie auf **OK**.

Korrelationen			Endpreis	Dauer (ordinal umkod.)
Spearman-Rho	Endpreis	Korrelationskoeffizient	1.000	-.128*
		Sig. (2-seitig)	.	.012
		N	391	384
	Dauer (ordinal umkod.)	Korrelationskoeffizient	-.128*	1.000
		Sig. (2-seitig)	.012	.
		N	384	384

*. Die Korrelation ist auf dem 0,05 Niveau signifikant (zweiseitig).

In diesem Beispiel ergibt sich eine schwache negative Korrelation der beiden Parameter in Höhe von -0.13. Je länger die Auktion läuft, desto niedriger ist in der Tendenz der Endpreis. Es ergibt sich ein Determinationskoeffizient von $r^2 = 0.016$. 16% der Gesamtvariabilität im Endpreis gehen auf die Auktionsdauer zurück, während 84% mit anderen Einflussfaktoren zusammenhängen. Trotz des schwachen Zusammenhangs ist dieser Zusammenhang mit $p = 0.012$ auf dem 5%-Niveau signifikant. Die Wahrscheinlichkeit für reinen Zufall wird mit 1,2% als gering angesehen. Das Ergebnis ist damit signifikant und man kann sagen, dass hier ein – wenn auch schwacher – systematischer Zusammenhang vorliegt.

¹⁹ Der Korrelationskoeffizient nach Spearman ist der gebräuchlichste Kennwert bei Ordinalskalen. Der Kennwert Kendall tau gilt bei kleinen Stichproben mit einem $N < 25$ als genauer in der Signifikanzschätzung. Bei grösseren Stichproben ist er oft etwas „konservativer“ in der Signifikanzschätzung. Zudem ist er immer dann zu empfehlen, wenn man es – unabhängig von der Stichprobengrösse – mit Ausreisserwerten in der Stichprobe zu tun hat. Auch Kendalls tau kann in der Dialogbox Bivariate Korrelation angefordert werden.

Das folgende Streudiagramm stellt den Zusammenhang grafisch dar (Diagramm | Diagrammerstellung... | Streu-/Punktdiagramme; vgl. auch S. 176 ff.):

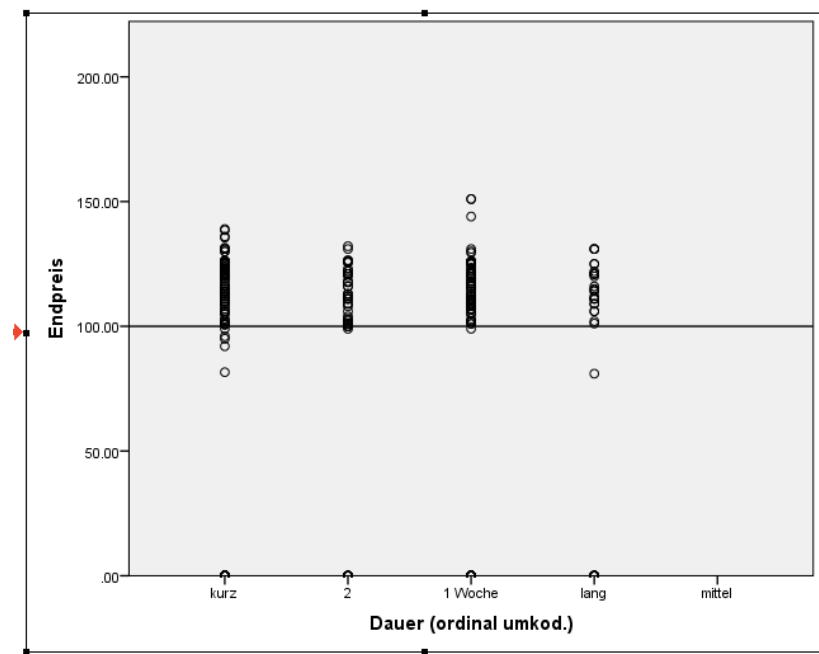


Abbildung 123: Vieweransicht "Streudiagramm der Variablen Endpreis und Dauer"

4.3 Korrelationsmasse – Beispiele

Beispiel 1: Prüfen Sie die Variablen **klasse** und **überlebt** der Datendatei **titanic.sav** auf Korrelation. Erstellen Sie zunächst eine Kreuztabelle und wählen Sie

Analysieren | Deskriptive Statistiken | Kreuztabellen...

Bringen Sie die Variable **klasse** in die Zeilen und die Variable **überlebt** in die Spalte. Gehen Sie auf die **Schaltfläche Statistik** und setzen Sie ein Häkchen bei **Korrelationen**.

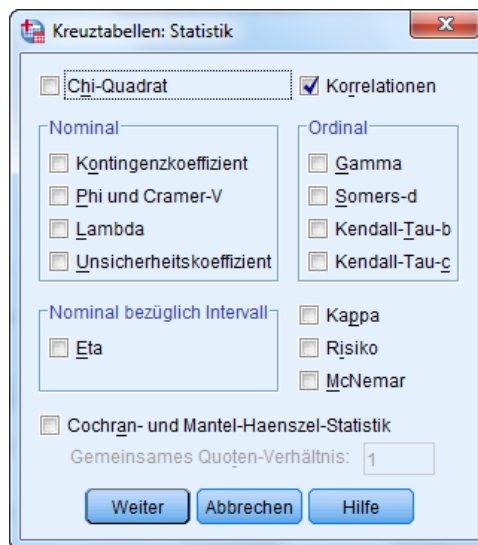


Abbildung 124: Ansicht "Kreuztabellen: Statistik"

Klicken Sie auf **Weiter** und wählen Sie dann im Kreuztabellen-Dialogfeld die Option **Keine Tabellen**. Alle anderen Einstellungen können unverändert beibehalten werden.

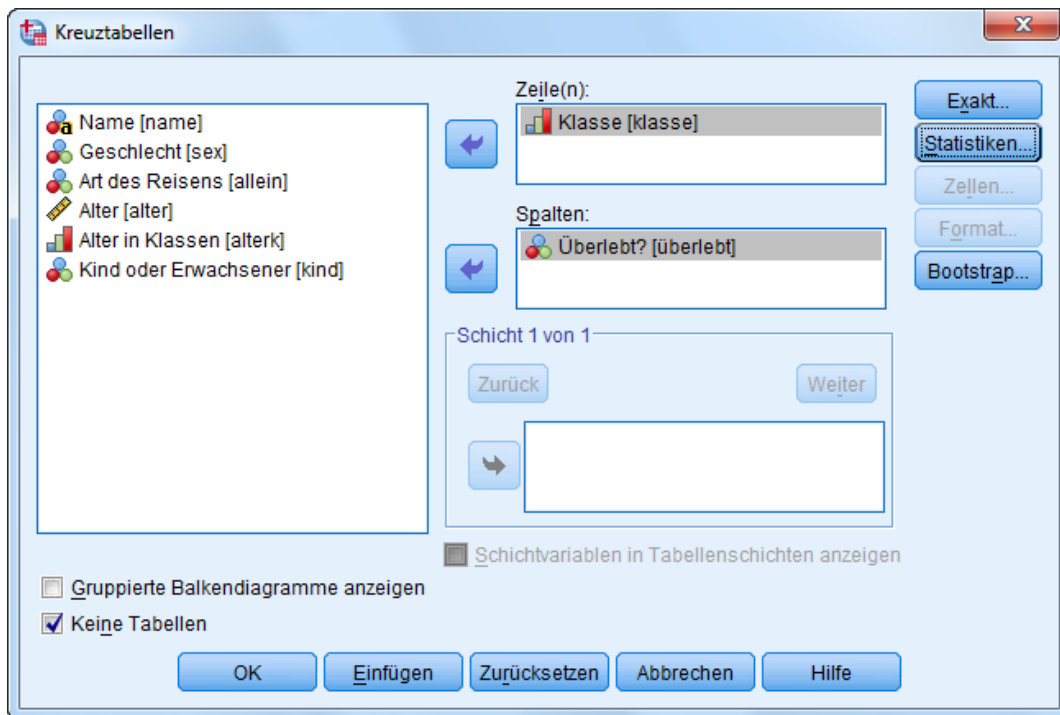


Abbildung 125: Ansicht "Kreuztabellen"

Im Viewer wird folgende Tabelle mit dem Spearman'schen Pearson'schen Korrelationskoeffizienten berechnet sowie die Signifikanzprüfung ausgegeben. Da es sich um nicht-metrische (ordinal und nominal) skalierte Variablen handelt, ist der **Korrelationskoeffizient nach Spearman** zu betrachten.

Symmetrische Maße					
		Wert	Asymptotischer Standardfehler ^a	Näherungsweise T ^b	Näherungsweise Signifikanz
Intervall- bzgl. Intervallmaß	Pearson-R	.311	.027	11.827	.000 ^c
Ordinal- bzgl. Ordinalmaß	Korrelation nach Spearman	.308	.027	11.698	.000 ^c
Anzahl der gültigen Fälle		1310			

a. Die Null-Hypothese wird nicht angenommen.

b. Unter Annahme der Null-Hypothese wird der asymptotische Standardfehler verwendet.

c. Basierend auf normaler Näherung

Abbildung 126: Vieweransicht "Korrelation"

Die Korrelation nach Spearman ergibt einen Koeffizienten r von .308 ($r^2 = .0949$). Das Ergebnis ist höchst signifikant ($p = .0001 < .05$). D. h., zwischen der Variable *klasse* und *überlebt* liegt eine geringe positive Korrelation vor (Aussage über die Beziehung zwischen den beiden Variablen) – die Variablen sind positiv korreliert (Aussage über die Richtung der Beziehung). Die Gleichläufigkeit der Beziehung zwischen den Variablen kann demzufolge so interpretiert werden, dass niedrige Werte der Variable *klasse* (vgl. Wertelabels) mit niedrigen Werten der Variable *überlebt* einhergehen:

Variable <i>klasse</i>		Variable <i>überlebt</i>	
1	1. Klasse	1	gerettet
2	2. Klasse	2	Gestorben
3	3. Klasse		

Passagiere der ersten Klasse zählen häufiger zu den Überlebenden (Wert 1) als Passagiere der dritten Klasse.

Beispiel 2: Prüfen Sie a) die Variablen **sex** und **überlebt** und b) **klasse** und **alter in klasse** auf Korrelation.²⁰

Zu a) Es ergibt sich die folgende Viewer-Ausgabe:

Symmetrische Maße					
		Wert	Asymptotischer Standardfehler ^a	Näherungsweise T ^b	Näherungsweise Signifikanz
Intervall- bzgl. Intervallmaß	Pearson-R	-.525	.024	-22.289	.000 ^c
Ordinal- bzgl. Ordinalmaß	Korrelation nach Spearman	-.525	.024	-22.289	.000 ^c
Anzahl der gültigen Fälle		1310			

a. Die Null-Hyphothese wird nicht angenommen.

b. Unter Annahme der Null-Hyphothese wird der asymptotische Standardfehler verwendet.

c. Basierend auf normaler Näherung

Abbildung 127: Vieweransicht "Korrelation"

²⁰ Dichotom-nominale Variablen werden als "ordinal" behandelt.

Die Korrelation nach Spearmen ergibt einen Koeffizienten r von $-.525$ ($r^2 = .2756$). Das Ergebnis ist höchst signifikant ($p = .0001 < .05$). Zwischen den Variablen **sex** und **überlebt** liegt eine mittlere negative Korrelation vor, d. h. niedrige Werte der Variable **sex** (vgl. Wertelabels) gehen mit hohen Werten der Variable **überlebt** einher und umgekehrt. Die Überlebenswahrscheinlichkeit der Frauen an Bord der Titanic ist damit höher gewesen als die der Männer.

Zu b) Es ergibt sich ebenfalls eine geringe negative Korrelation ($r = -.380$, $r^2 = .14444$). Je grösser die Werte der Variable **klasse**, desto niedriger die Werte der Variable **alter in klasse** und umgekehrt. Berücksichtigt man die Kodierung der beiden Variablen, so ist davon auszugehen, dass ältere Personen (1 = bis 14 Jahre, 2 = 15-30 Jahre, 3 = 31-50 Jahre, 4 = > 50 Jahre) eher in der ersten Klasse zu finden sind sowie jüngere Passagiere häufiger in der dritten Klasse gereist sind, was eine entsprechende Kreuztabelle zwischen der Klassenzugehörigkeit an Bord der Titanic und dem Alter gruppiert nach Altersklassen verdeutlicht.

Erstellen Sie für die beiden Variablen **klasse** (in Zeile) und **alter in klassen** (in Spalte) eine Kreuztabelle. Lassen Sie sich über die Schaltfläche Zellen neben den beobachteten auch die erwarteten Häufigkeiten sowie die standardisierten Residuen anzeigen.

Klicken Sie dann auf **Weiter** und **OK**. Im Viewerfenster erscheint folgende Ausgabe:

Klasse * Alter in Klassen Kreuztabelle

		Alter in Klassen				Gesamt
		Bis 14 Jahre	15-30 Jahre	31-50 Jahre	Über 50 Jahre	
1. Klasse	Anzahl	7	81	143	67	298
	Erwartete Anzahl	28.6	148.4	95.0	26.1	298.0
	Standardisierte Residuen	-4.0	-5.5	4.9	8.0	
Klasse 2. Klasse	Anzahl	26	123	89	20	258
	Erwartete Anzahl	24.7	128.5	82.2	22.6	258.0
	Standardisierte Residuen	.3	-.5	.7	-.5	
3. Klasse	Anzahl	81	388	147	17	633
	Erwartete Anzahl	60.7	315.2	201.8	55.4	633.0
	Standardisierte Residuen	2.6	4.1	-3.9	-5.2	
Gesamt	Anzahl	114	592	379	104	1189
	Erwartete Anzahl	114.0	592.0	379.0	104.0	1189.0

Abbildung 128: Vieweransicht "Kreuztabelle"

Der Anteil der Passagiere in der ersten Klasse, die 30 Jahre und jünger sind, liegt deutlich zu niedrig. Der Anteil der Passagiere, die 31 Jahre und älter sind, liegt deutlich zu hoch. In der dritten Klasse zeigt sich ein gegenläufiger Trend. Besonders stark in das Chi-Quadrat-Mass geht die Kategorie der über 50-Jährigen ein mit einem standardisierten Residuum von +8,0 bei der ersten und -5,2 bei der dritten Klasse.

4.4 Partielle Korrelationen

Fallbeispiel:

Untersuchen Sie den Zusammenhang zwischen der Tatsache, dass ein Passagier allein an Bord der Titanic war oder mit einer anderen Gruppe reiste, und der Tatsache seines Überlebens bzw. Todes.

Öffnen Sie die Datendatei **titanic.sav** und wählen Sie

Analysieren | Korrelation | Bivariat...

Wählen Sie die Variablen **Art des Reisens** (allein) und **Überlebt?** (überlebt) aus und wählen Sie den **Korrelationskoeffizienten von Spearman** aus.

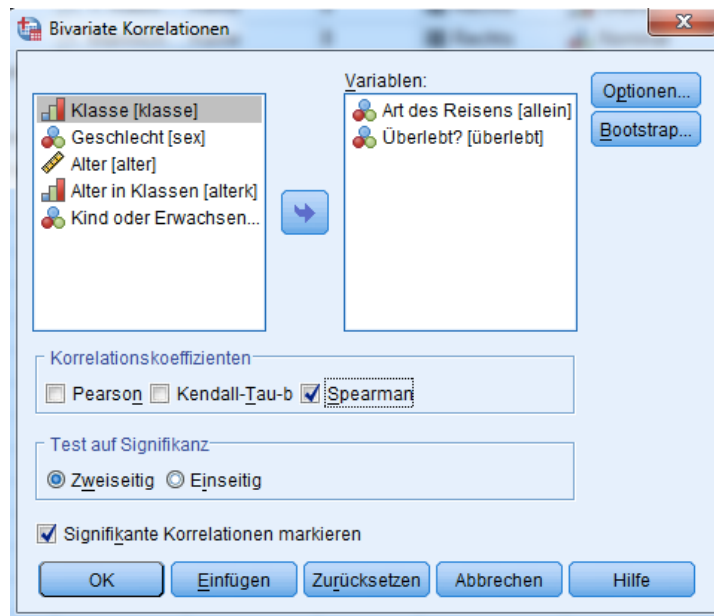


Abbildung 129: VAnsicht "Bivariate Korrelationen"

Klicken Sie auf **OK**. Es erscheint folgende Viewerausgabe:

Korrelationen				
			Art des Reisens	Überlebt?
Spearman-Rho	Art des Reisens	Korrelationskoeffizient	1.000	-.171**
		Sig. (2-seitig)	.	.000
		N	1310	1310
	Überlebt?	Korrelationskoeffizient	-.171**	1.000
		Sig. (2-seitig)	.000	.
		N	1310	1310

** . Die Korrelation ist auf dem 0,01 Niveau signifikant (zweiseitig).

Abbildung 130: Vieweransicht "Korrelationen"

Personen, die mit einer anderen Person an Bord der Titanic waren, verfügen über eine höhere Überlebenswahrscheinlichkeit als allein reisende Passagiere. Die Korrelation nach Spearman ergibt einen Koeffizienten r von $-.171$ ($r^2 = .0292$). Das Ergebnis ist höchst signifikant ($p = .0001 < .05$). Zwischen den Variablen allein und überlebt liegt eine sehr geringe negative Korrelation vor, d. h. niedrige Werte der Variable allein (vgl. Wertelabels) gehen mit hohen Werten der Variable überlebt einher und umgekehrt.

Nehmen Sie nun die Variable **geschlecht** hinzu und prüfen Sie, ob diese mit den Variablen allein und überlebt korreliert.

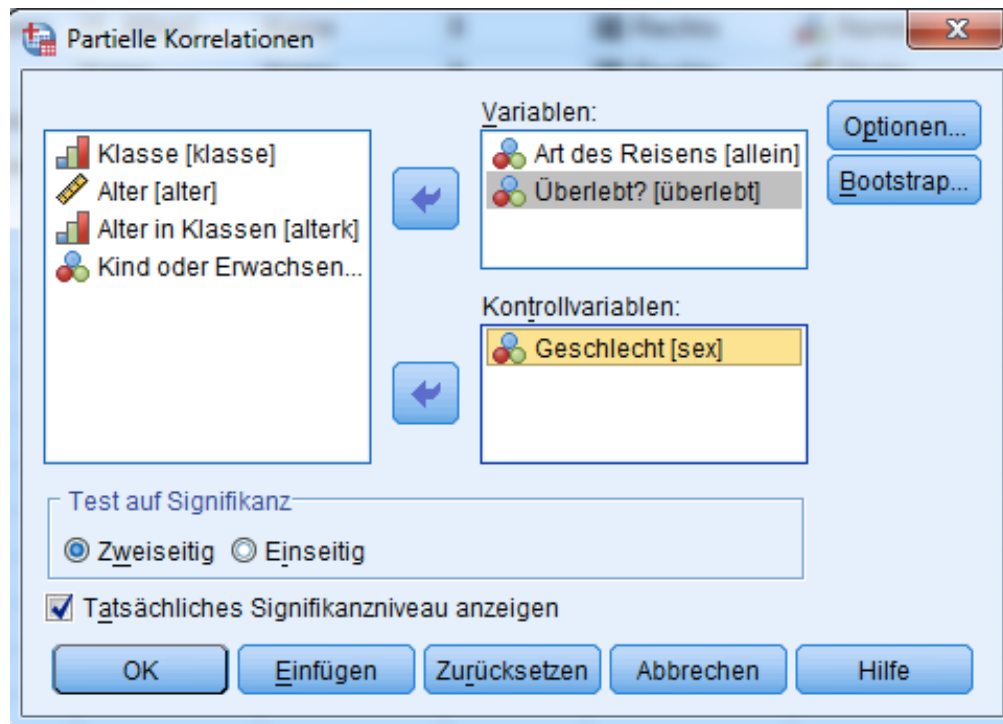


Abbildung 131: VAnsicht "Partielle Korrelationen"

Es ergibt sich folgende Viewerausgabe:

Korrelationen			Art des Reisens	Geschlecht	Überlebt?
Spearman-Rho	Art des Reisens	Korrelationskoeffizient	1.000	.272**	-.171**
		Sig. (2-seitig)	.	.000	.000
		N	1310	1310	1310
	Geschlecht	Korrelationskoeffizient	.272**	1.000	-.525**
		Sig. (2-seitig)	.000	.	.000
		N	1310	1310	1310
	Überlebt?	Korrelationskoeffizient	-.171**	-.525**	1.000
		Sig. (2-seitig)	.000	.000	.
		N	1310	1310	1310

** . Die Korrelation ist auf dem 0,01 Niveau signifikant (zweiseitig).

Abbildung 132: Vieweransicht "Korrelationen"

Die Variable geschlecht korreliert mit der Variable allein signifikant positiv ($r = .272$) und mit der Variable überlebt signifikant negativ ($r = -.525$). Es ergibt sich der Verdacht, dass das Geschlecht eine „Störvariable“ sein könnte, die für eine Scheinkorrelation zwischen der Art des Reisens und der Tatsache des Überlebens verantwortlich ist. Dies soll mithilfe der **Berechnung des partiellen Korrelationskoeffizienten** untersucht werden. Wählen Sie

Analysieren | Korrelation | Partiiell...

Bringen Sie die Variablen allein und überlebt in das Variablenfeld und die Variable geschlecht in das Kontrollvariablenfeld und klicken Sie auf **OK**. Es erscheint nachfolgende Ausgabe im Viewerfenster:

Korrelationen			
Kontrollvariablen		Art des Reisens	Überlebt?
Geschlecht	Korrelation	1.000	-.034
	Art des Reisens	.	.218
	Signifikanz (zweiseitig)	0	1307
	Freiheitsgrade	-.034	1.000
	Überlebt?	.218	.
	Signifikanz (zweiseitig)	1307	0
	Freiheitsgrade		

Abbildung 133: Vieweransicht "Korrelationen"

Bei Ausschluss der Störvariablen Geschlecht liegt keine signifikante Korrelation zwischen der Art des Reisens und dem Überleben der Passagiere mehr vor.²¹ Der ursprünglich als höchst signifikant ausgewiesene Korrelationskoeffizient zwischen der Art des Reisens und der Überlebenswahrscheinlichkeit ist nicht Ausdruck eines unmittelbaren kausalen Zusammenhangs, sondern wird von der Variable Geschlecht entscheidend mitbestimmt.

²¹ Zu jener Zeit war das Alleinreisen von Frauen vor allem höherer Schichten verpönt, sodass Frauen i.d.R. mit ihrem Ehemann oder einem Begleiter an Bord waren.

5. Lineare Regression

- Frage: Welchen Beitrag leistet eine Variable zur Vorhersage einer zweiten Variablen?
- Beispiel: Welchen Einfluss hat die Qualität des Mensaessens auf die Aufmerksamkeit am Nachmittag?
- Bei metrischen Variablen

Fall Hypertoniestudie (Datei: hyper.sav)

5.1 Einfache lineare Regression

Während die Korrelationsrechnung die Stärke des Zusammenhangs zwischen zwei Variablen ermittelt, dient die Regressionsanalyse dazu, die Art dieses Zusammenhangs aufzuzeigen bzw. Möglichkeiten an die Hand zu geben, den Wert einer (abhängigen) Variablen aus den Werten einer anderen (unabhängigen) Variable vorherzusagen. Die Regressionsgerade ist wie folgt definiert:

$$Y = b \cdot x + a$$

Dabei nennt man b den Regressionskoeffizienten und a den Ordinatenabschnitt. Letzterer gibt den Punkt auf der (vertikalen) y -Achse an, an dem diese von der Regressionsgeraden geschnitten wird. Der Regressionskoeffizient b gibt den Steigungswinkel der Geraden an. Aufgabe der einfachen linearen Regression ist es, die Parameter b und a abzuschätzen. Als optimale Lösung gilt dabei diejenige Gerade, für welche die Summe der quadrierten vertikalen Abstände von den einzelnen Messpunkten ein Minimum wird.

5.2 Vorgehen in SPSS Statistics

Überprüfen Sie, welchen Einfluss der Cholesterinwert `chol 0` (unabhängige Variable) auf den Cholesterinwert `chol 1` nach einem Monat hat (abhängige Variable). Wählen Sie im Menü

Analysieren | Regression | Linear...

Es öffnet sich die Dialogbox Lineare Regression. Bringen Sie die Variable **chol1** in das für die **abhängige Variable** vorgesehene Feld und erklären Sie **chol0** entsprechend zur **unabhängigen Variable**. Klicken Sie auf **OK**.

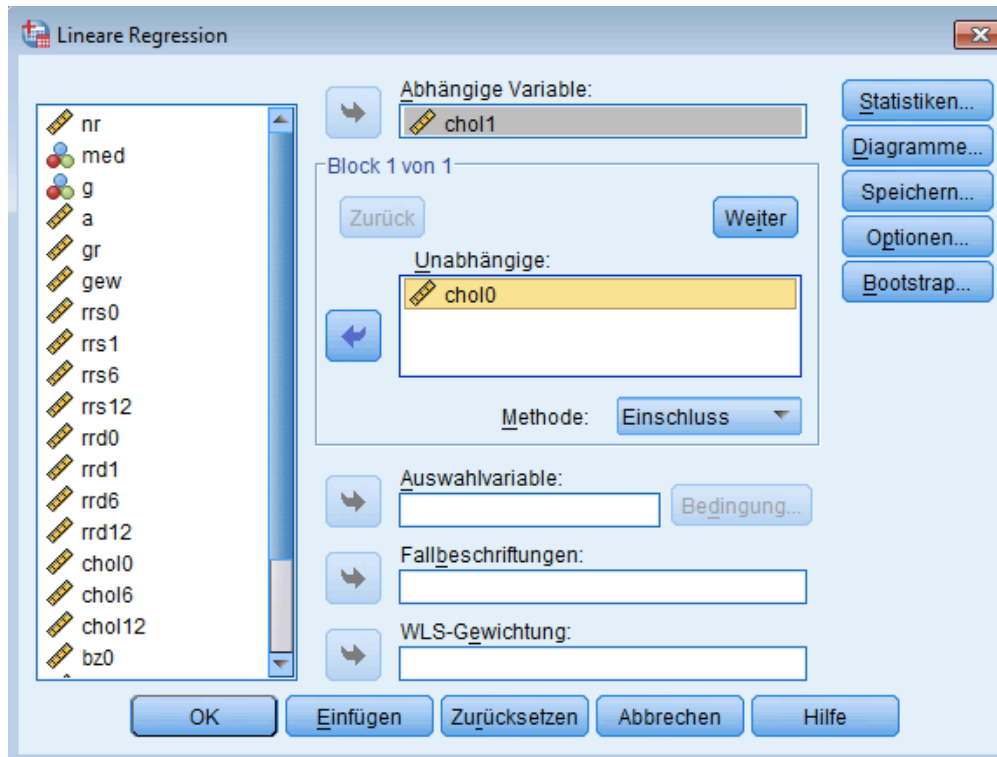


Abbildung 134: Ansicht "Lineare Regression"

Die wesentlichen Ergebnisse sind im Folgenden wiedergegeben. Ausgehend von der Formel für die Regressionsgerade stellt sich die Aufgabe, die Parameter der Beziehung $\text{chol1} = b \cdot \text{chol0} + a$ zu schätzen. In der Folge kann man aus der Kenntnis des Ausgangswerts eine Vorhersage über den Wert nach einem Monat treffen.

Modellzusammenfassung

Modell	R	R-Quadrat	Korrigiertes R-Quadrat	Standardfehler des Schätzers
1	.861 ^a	.741	.740	25.258

a. Einflußvariablen : (Konstante), Cholesterin, Ausgangswert

ANOVA^b

Modell		Quadratsumme	df	Mittel der Quadrate	F	Sig.
1	Regression	314337.948	1	314337.948	492.722	.000 ^a
	Nicht standardisierte Residuen	109729.408	172	637.962		
	Gesamt	424067.356	173			

a. Einflußvariablen : (Konstante), Cholesterin, Ausgangswert

b. Abhängige Variable: Cholesterin, nach 1 Monat

Koeffizienten^a

Modell		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten	T	Sig.
		Regressionskoeffizient B	Standardfehler	Beta		
1	(Konstante)	34.546	9.416		3.669	.000
	Cholesterin, Ausgangswert	.863	.039	.861	22.197	.000

a. Abhängige Variable: Cholesterin, nach 1 Monat

Abbildung 135: Vieweransicht "Lineare Regression"

Tabelle Modellzusammenfassung: Das **korrigierte R-Quadrat** ist das Bestimmtheitsmass, die Varianzaufklärung der Regression. Damit wird gesagt wie viel der Gesamtvariabilität in der Stichprobe durch die unabhängige Variable erklärt wird. Die Quadratwurzel aus dem Bestimmtheitsmass, als "**R**" bezeichnet, ist bei der einfachen Regressionsanalyse gleich dem Korrelationskoeffizienten nach Pearson. Im Beispiel ergibt sich eine Varianzaufklärung von 74%. Diese Varianzaufklärung ist zunächst einmal auf die Stichprobe bezogen. Deswegen wird im nächsten Schritt geprüft, ob das Ergebnis der Bestimmtheit in der Stichprobe auf die Grundgesamtheit übertragbar ist, oder ob es rein zufällig entstanden sein könnte.

Die zweite Tabelle gibt den Anteil der Varianz wieder, der durch die Regressionsgleichung erklärt wird (Quadratsumme "Regression") bzw. nicht erklärt wird (Quadratsumme "Residuen"). Der Quotient aus dem erklärten Teil der Varianz und der Gesamtvarianz, hier also $314337,948 / 424067,356 = 0,741$ wird als "**Bestimmtheitsmass**" bezeichnet und unter der Bezeichnung "**R-Quadrat**" ausgegeben. Diese Grösse ist ein Mass für die Güte der Anpassung durch die Regressionsgerade und immer zwischen 0 und 1 gelegen. Die Absicherung gegen null erfolgt über die Prüfgrösse F und das zugeordnete Signifikanzniveau. Die Tabelle "ANOVA" weist aus, dass das Ergebnis der Bestimmtheit höchst signifikant ist und nicht rein zufällig entstanden ist.²²

In der dritten Tabelle werden der **Regressionskoeffizient b** und unter der Bezeichnung "**Konstante**" der **Ordinatenabschnitt a** ausgegeben. Dies führt zu der Regressionsgleichung $chol1 = 0.863 \cdot chol0 + 34,546$. Ein Ausgangswert von z.B. 280 lässt hiernach einen Wert von 276 nach einem Monat erwarten. Die Quotienten aus den berechneten Koeffizienten und deren Standardfehler bilden die Prüfgrösse t; das zugehörige Signifikanzniveau bezieht sich auf die Absicherung dieser Koeffizienten gegen null. Der Beta-Koeffizient ist bei der multiplen Analyse von Bedeutung.

Die Frage, ob der gegebene Zusammenhang überhaupt als linear angesehen werden kann, ist am einfachsten durch die Betrachtung des entsprechenden Streudiagramms zu entscheiden; ein hoher Anteil der erklärten Varianz deutet auch darauf hin. Um ein Streudiagramm mit Regressionsgerade zu zeichnen, wählen Sie

Diagramme | Diagrammerstellung...

und wählen Sie aus der Galerie die Option **Streu-/Punktdiagramm**. Ziehen Sie das Symbol **Einfaches Streudiagramm** in die Diagrammvorschau. Ziehen Sie

²² Falls sich an dieser Stelle kein signifikantes Ergebnis ergibt, kann man die Untersuchung abbrechen, da der gewählte Prädiktor die abhängige Variable nicht erklären kann.

die Variable **chol1** in das y-Achsenfeld und die Variable **chol0** in das x-Achsenfeld und bestätigen Sie mit **OK**.

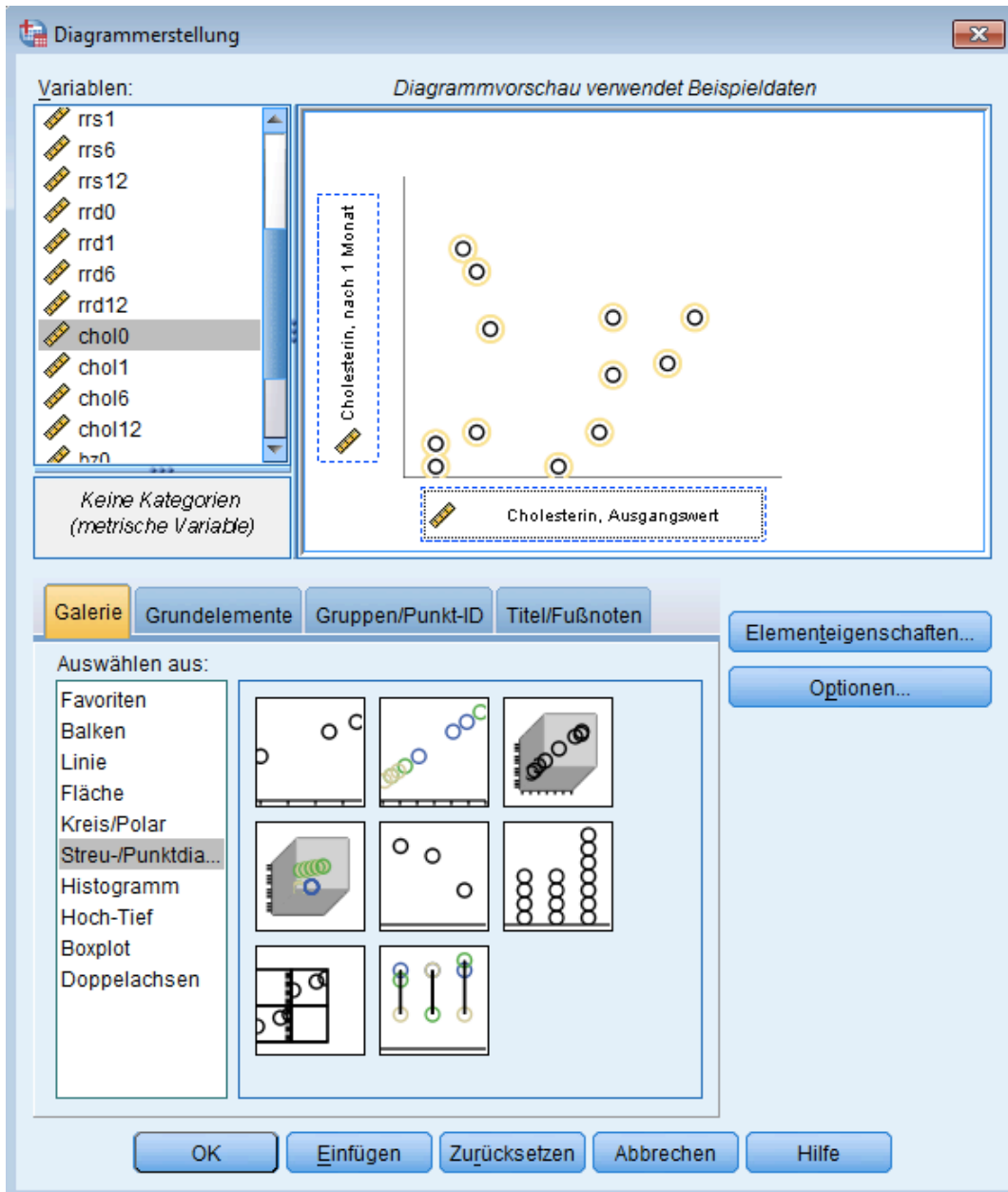


Abbildung 136: Ansicht "Diagrammerstellung"

Das Streudiagramm wird im Viewer angezeigt:

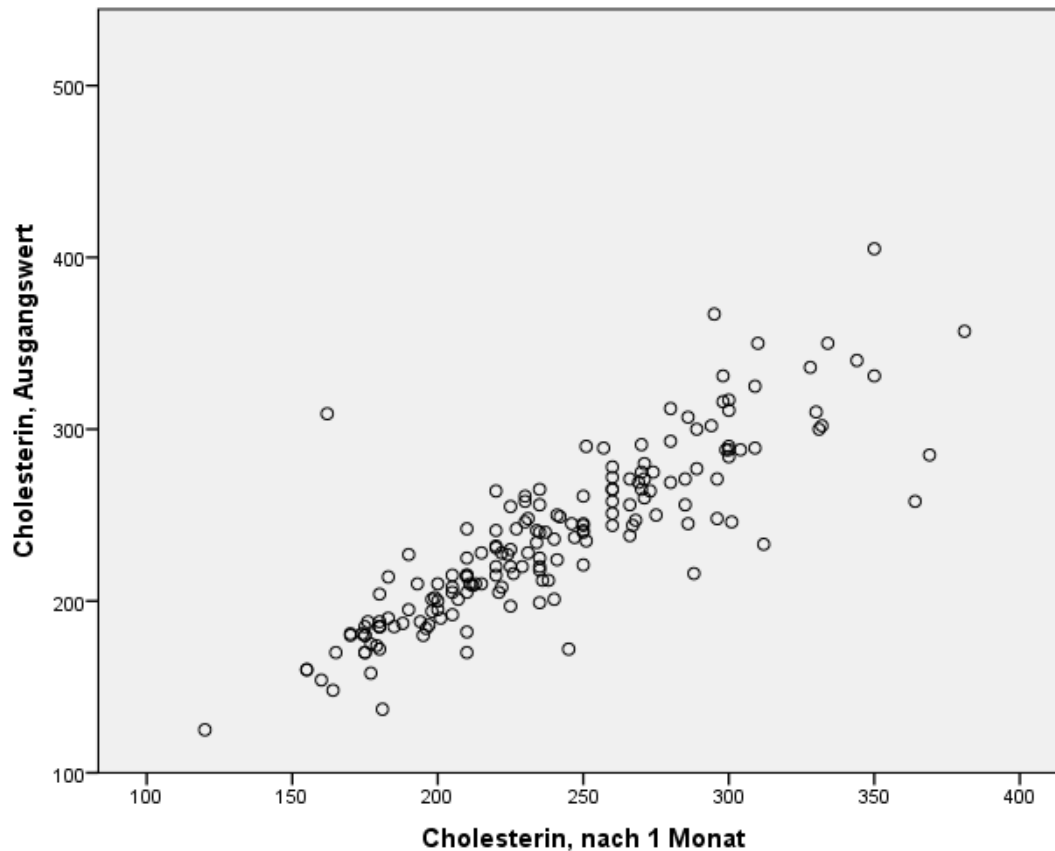


Abbildung 137: Vieweransicht “Streudiagramm – Variablen chol 0 und chol1”

Klicken Sie doppelt auf die Grafik, um sie in den **Diagramm-Editor** zu übertragen. Klicken Sie auf das Icon zum Einzeichnen der

Anpassungslinie . Es öffnet sich die Dialogbox **Eigenschaften**. Wählen Sie hier die Option **Linear** und bestätigen Sie mit **Schliessen**. Schliessen Sie den Diagramm-Editor. Die Regressionsgerade ist nunmehr eingezeichnet.

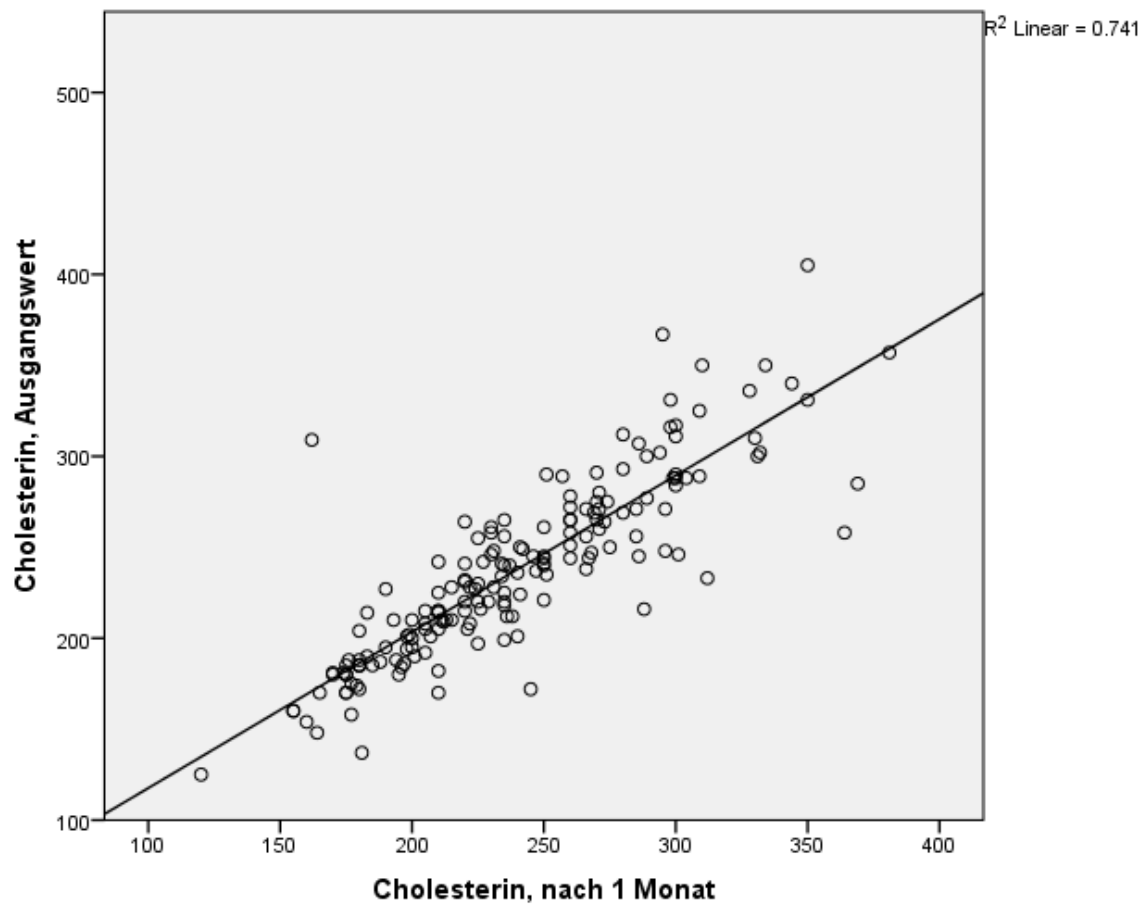


Abbildung 138: Vieweransicht "Streudiagramm – Variablen chol 0 und chol1 mit Regressionsgerade"