

Sozialwissenschaftliche Datenverarbeitung

Datenanalyse mittels SPSS[®]



Inhaltsverzeichnis

1	EINLEITUNG	4
2	DAS CODEBOOK	5
2.1	Identifikationsnummer.....	5
2.2	Kodierung von Variablen mit einer Antwortmöglichkeit	6
2.3	Kodierung von Mehrfachantworten	7
2.4	Kodierung von offenen Fragen.....	9
2.5	Kodieren von Kategorien „Sonstige“	9
2.6	Das Vorkodieren offener Fragen.....	11
3	SPSS: DIE OBERFLÄCHE	13
3.1	Fenster	13
3.1.1	Der Dateneditor	13
3.1.2	Der Viewer.....	14
3.1.3	Der Syntax-Editor.....	15
3.2	Menüs.....	16
3.3	Dialogfelder	16
4	DIE DATENEINGABE	17
4.1	SPSS-Dateneditor	17
4.2	Excel.....	19
4.3	ASCII.....	19
4.4	Das Programmieren einer Eingabemaske mittels Access.....	21
5	DER DATENSATZ	27
5.1	Die Umwandlung der Daten in SPSS-Format	27
5.2	Datendefinition.....	28
5.2.1	Variable Labels	29
5.2.2	Value Labels	29
5.2.3	Missing Values	30
5.3	Datenmodifikation	31
5.3.1	Umkodieren (Recode)	31
5.3.2	Variablen Modifizieren (If).....	31
5.3.3	Variablen einfügen oder löschen	32
5.3.4	Berechnen neuer Variable (Compute)	32
5.3.5	Filter setzen (Select if)	32

6	DIE DATENANALYSE	34
6.1	Univariate Datenanalyse	34
6.1.1	Häufigkeitsverteilungen	34
6.1.1.1	Häufigkeitsverteilungen und Datenbereinigung	36
6.1.1.2	Umkodieren von offenen Fragen und Kategorien „Sonstige“	36
6.1.2	Deskriptive Statistiken	37
6.2	Bivariate Datenanalyse.....	38
6.2.1	Kreuztabellen	38
6.2.1.1	Plausibilitätschecks in der Datenbereinigung	40
6.2.2	Korrelationen	41
6.2.3	T-Test	43
6.2.4	Mann-Whitney-U-Test	45
7	LITERATURVERZEICHNIS	48

1 Einleitung

Viele Studenten beschließen, als Diplom- oder Masterarbeit eine empirische Arbeit zu schreiben. Zur Überprüfung ihrer Hypothesen entwickeln sie Fragebögen und schicken diese raus. Irgendwann kommen diese Fragebögen zurück.... und dann? Ich habe festgestellt, dass viele Studenten nicht genügend Einblick in die Datenverarbeitung haben, um die Ergebnisse dieser Fragebögen selbständig zu verarbeiten. Das Ziel dieses Kurses ist es, Sie in die Lage zu versetzen, diese Fragebögen auf eine professionelle Art und Weise zu verarbeiten und Ihre Daten mittels SPSS zu analysieren. Im Kurs werden die wichtigsten Aspekte des Datensammelns, der Datenbereinigung und der Datenanalyse, wie das Codebook, die Eingabemaske, Plausibilitätschecks, Häufigkeitstabellen, Korrelationen etc., besprochen. Der Umgang mit dem Programm SPSS wird durch verschiedene Übungsaufgaben gelernt. Der Abschnitt „Analyse“ ist mit Absicht beschränkt gehalten, da die teilnehmenden Studenten unterschiedliche Vorkenntnisse haben. Wenn Sie aber verstanden haben, wie das Programm SPSS funktioniert, wird es nicht schwierig sein, die Ihnen bekannten datenanalytischen Prozeduren durchzuführen.

Ich habe versucht, diesen Kurs möglichst wenig „akademisch“ zu gestalten und möglichst wenig Jargon zu benutzen, damit er auch für Studenten mit wenigen Vorkenntnissen bezüglich Methoden und Statistik verständlich ist. In diesem Kurs sind viele Aspekte verarbeitet, auf die ich während meiner Arbeit gestoßen bin.

Viel Spaß!

Marleen Brinks

2 Das Codebook

Bevor Sie anfangen können, Daten einzugeben oder zu analysieren, brauchen Sie ein Codebook. Im Codebook steht genau beschrieben, welche Variable welchen Namen trägt, welche Kategorien sie hat und eventuell in welchen Spalten in der Datenmatrix sie zu finden ist. Die Spaltenangabe braucht man, wenn man die Daten z.B. in ASCII-Format eingibt (SPSS: DATA LIST) oder Daten aus einem SAS-Job auf SPSS-Format bringen möchte. Bei den modernen Eingabemethoden kann man allerdings sehen, wo die Daten im File stehen. Die Basis für das Codebook ist der Fragebogen.

Auch beschreibt man im Codebook, wie man während der Dateneingabe mit Fehlern umgehen wird. Dies ist vor Allem sehr wichtig, wenn die Daten von mehreren Personen eingegeben werden. Auch ist es für jemand, der die Daten zur sekundären Datenanalyse benutzt, wichtig zu wissen wie Fehler im Fragebogen behandelt wurden. Wenn Sie selbst eine sekundäre Datenanalyse vornehmen, sind sie auf ein Codebook, das jemand anderes verfaßt hat, angewiesen.

2.1 Identifikationsnummer

Sehr wichtig: Die erste Variable in der Datenmatrix, also auch die erste Variable in dem Codebook sollte eine eindeutige Nummer sein, anhand derer die Fragebögen einzeln identifiziert werden können. Diese Nummer wird gut sichtbar auf dem Fragebogen notiert. Wenn Sie während Ihrer Datenbereinigung Ungereimtheiten feststellen, versetzt diese Nummer Sie in die Lage, die Fragebögen mit den eingegebenen Daten zu vergleichen.

2.2 Kodierung von Variablen mit einer Antwortmöglichkeit

Mit Variablen mit einer Antwortmöglichkeit sind Variablen gemeint, die Kategorien haben, die sich gegenseitig ausschließen. Man ist entweder männlich oder weiblich und man kann nicht gleichzeitig katholisch und evangelisch sein. Mehrfachnennungen kann es bei solchen Variablen per Definition nicht geben.

Beispiel:

Frage 1: Wo ist ihr erster Wohnsitz? Bitte kreuzen Sie das zutreffende Kästchen an.

- In den alten Bundesländern
- In den neuen Bundesländern
- Weiß nicht

Der Abschnitt im Codebook, der die obere Frage beschreibt, könnte wie folgt aussehen:

Frage 1: V2 - Wohnsitz

1 – in den alten Bundesländern

2 – in den neuen Bundesländern

9 – Weiß nicht

0 – Keine Angabe

Mehrfache Ankreuzung wird mit 0 (keine Angabe) kodiert.

Es ist vernünftig, die verschiedenen Arten von fehlenden Informationen unterschiedlich zu kodieren. Nicht zu wissen, wo man seinen ersten Wohnsitz hat, ist etwas anderes, als keine Angabe machen zu wollen.

2.3 Kodierung von Mehrfachantworten

Es gibt zwei verschiedene Arten von Fragestellungen die Mehrfachantworten erzeugen, die jeweils auf einer anderen Art und Weise kodiert werden müssen.

In der ersten Fragestellung wird zum Beispiel gefragt:

7. Welche Krimis schauen sie häufig (mehr als 1 mal pro Quartal)? Bitte kreuzen Sie die zutreffenden Kästchen an.

Tatort
Derrick
Der Alte
Adelheid und ihre Mörder
Polizeiruf 110
Kommissar Rex
Die Autobahnpolizei
Helicops
Sonstige

Da man pro Respondent auf eine Variable nur EINEN Wert eingeben kann, hat diese Fragestellung zur Folge, dass zur Kodierung 9 Variablen (zu jeder Sendung eine) angelegt werden müssen mit zwei Antwortkategorien. Man hat die Sendung entweder genannt oder nicht genannt. Falls Sie den Respondenten die Option „Sonstige“ durch einen zusätzlichen Text spezifizieren lassen (Sonstige, nämlich _____), haben Sie zusätzlich die Möglichkeit, diese Spezifikationen im Nachhinein zu kodieren. Siehe dazu Abschnitt 2.5. Im Codebook sieht diese Variable wie folgt aus:

V7a: Tatort
1=genannt
0=nicht genannt

V7c: Der Alte
1=genannt
0=nicht genannt

V7b: Derrick
1=genannt
0=nicht genannt

Etc.

Die zweite Möglichkeit zur Kodierung von Mehrfachnennungen ist vor Allem geeignet, wenn man eine kleine Menge aus einer großen Anzahl von Wahlmöglichkeiten wählen soll. Man legt dann nur soviel Variablen an, wie es Wahlmöglichkeiten gibt.

Beispiel:

Frage 29: Bitte kreuzen Sie die drei Fernsehsender an, die Sie am häufigsten schauen

ZDF
WDR
MDR
Radio Bremen
NDR
BR
Sender Freies Berlin
Pro 7
SAT 1
VOX
TV5
RAI UNO
RTL
RTL2

Hier legt man drei Variablen an mit den gleichen Kategorien. Im Codebook sieht das so aus:

V29A: Fernsehsender 1

1=ZDF
2=WDR
3=MDR
4=Radio Bremen
5=NDR
6=BR
7=Sender Freies Berlin
8=Pro 7
9=SAT 1
10=VOX
11=TV5
12=RAI UNO
13=RTL
14=RTL2

V29B: Fernsehsender 2

1=ZDF
2=WDR
3=MDR
4=Radio Bremen
5=NDR
6=BR

7=Sender Freies Berlin
8=Pro 7
9=SAT 1
10=VOX
11=TV5
12=RAI UNO
13=RTL
14=RTL2

V29C: Fernsehsender 3

1=ZDF
2=WDR
3=MDR
4=Radio Bremen
5=NDR
6=BR
7=Sender Freies Berlin
8=Pro 7
9=SAT 1
10=VOX
11=TV5
12=RAI UNO
13=RTL
14=RTL2

Missing = 99 .

Bei Nennung von einer höheren Zahl als 3 Fernsehsendern werden alle drei Variablen mit „0“ kodiert.

2.4 Kodierung von offenen Fragen

Eine offen gestellte Frage ist eine Frage, bei der keine Antwortmöglichkeiten (Kategorien) vorgegeben sind. Sie erzeugt im Gegensatz zu den oben aufgeführten Beispielen keine numerische Variable, sondern eine Textvariable. Bei der Kodierung von offenen Fragen muß beim Erstellen der Eingabemaske berücksichtigt werden, dass diese Variable aus Text besteht und nicht aus Zahlen. Die Antwort, die der Respondent zu dieser Variable gegeben hat, wird entweder vollständig oder zusammengefaßt abgetippt. Wie man offene Fragen weiterverarbeitet und analysiert wird im Abschnitt 6.1.1.2 behandelt.

Frage 16: Haben Sie für die Firma XYZ Verbesserungsvorschläge in Bezug zu dem Call Center?

Codebook:

V16: Verbesserungsvorschläge

Textvariable, wird vollständig abgetippt.

2.5 Kodieren von Kategorien „Sonstige“

Häufig gibt man eine bestimmte Zahl von Kategorien vor, und gibt den Respondenten die Möglichkeit, unter „Sonstige“ Antworten zu geben, die nicht in den vorgegebenen Kategorien definiert wurden. Zur Kodierung der Kategorie „Sonstige“ nehmen wir eine abgeänderte Form des Beispiels aus Abschnitt 2.2.

Frage 17: Welche Krimi schauen Sie *am Liebsten*? Bitte kreuzen Sie nur ein Kästchen an.

Tatort
Derrick
Der Alte
Adelheid und ihre Mörder
Polizeiruf 110
Kommissar Rex
Die Autobahnpolizei
Helicops
Sonstige, bitte spezifizieren: _____

Zur Verarbeitung dieser Variable werden zwei Variablen angelegt, eine numerische Variable und eine Textvariable. In der numerische Variable wird die Kategorie zu der Frage kodiert. Falls nun mit „sonstige“ geantwortet wird, wird die Spezifikation in der Textvariable erfaßt. Wenn die Daten vollständig erfaßt sind, kann man sich die verschiedenen unter „sonstige“ genannten Sendungen auflisten lassen und kategorisieren. Die technische Ausführung wird im Abschnitt 6.1.1.2. beschrieben.

Beispiel

V17: Krimi am Liebsten

1=Tatort
2=Derrick
3=Der Alte
4=Adelheid und ihre Mörder
5=Polizeiruf 110
6=Kommissar Rex
7=Die Autobahnpolizei
8=Helicops
9=Sonstige

0=missing

v17a: Krimi am Liebsten – sonstige

Textvariable.

Wenn hier mehrere Sendungen genannt sind oder wenn die Kategorie nicht spezifiziert wurde („Sonstige“ ausgewählt aber nichts eingetragen), wird die Variable leer gelassen.

2.6 Das Vorkodieren offener Fragen

Wenn Sie eine große Befragung durchführen (z.B. 1000 Respondenten), oder wenn in den Antworten auf eine offene Frage mehrere Themen angesprochen werden, empfiehlt es sich, die Fragen vorzukodieren und nur die Zahlen zu erfassen.

Gehen wir noch mal zurück zu unserem Beispiel. Stellen Sie sich vor, dies ist der erste Fragebogen den Sie kodieren möchten.

Frage 16: Haben Sie für die Firma XYZ Verbesserungsvorschläge in Bezug zu dem Call Center?

Ich möchte nicht mehr auf die Anzahl der Sekunden pro Gespräch beurteilt werden,

sondern danach, wie zufrieden die Kunden im Endeffekt sind.

Außerdem ist es hier viel zu kalt und die Luft ist stickig. Es wäre gut, wenn im auf den

Toiletten nicht mehr heimlich geraucht wird.

Sie schreiben jetzt auf ein Blatt Papier:

1. ich möchte nicht mehr auf die Anzahl der Sekunden pro Gespräch beurteilt werden
2. ich möchte nach Kundenzufriedenheit beurteilt werden
3. Temperatur ist zu niedrig
4. Luft ist stickig
5. Auf den Toiletten wird heimlich geraucht

Sie schreiben jetzt 1, 2, 3, 4, 5 in dem Fragebogen.

Danach nehmen Sie den zweiten Bogen. Darin steht:

Frage 16: Haben Sie für die Firma XYZ Verbesserungsvorschläge in Bezug zu dem Call Center?

Es ist hier immer kalt. Die Getränke im Automat sind völlig überteuert. Eine Raucherecke wäre gut.

Hier kodieren Sie „3“ (Temperatur zu niedrig). Für die anderen beiden Nennungen müssen neue Codes vergeben werden.

Sie schreiben jetzt weiter:

4. Luft ist stickig
5. Auf den Toiletten wird heimlich geraucht
6. Getränke sind zu teuer
7. Raucherecke fehlt

In diesem Fragebogen schreiben Sie die Zahlen 3, 6 und 7.

So gehen Sie alle Bögen durch.

Um diese Nennungen zu Erfassen legen Sie z.B. acht Variablen an:

Resp	V16a	V16b	V16c	V16d	V16e	V16f	V16g	V16h
1	1	2	3	4	5			
2	3	6	7					

Die Labels sind für jede Variable die gleichen. Ausgewertet werden diese Variablen mit einer speziellen Prozedur für das Verarbeiten von Mehrfachnennungen in SPSS.

3 SPSS: Die Oberfläche

3.1 Fenster

Im SPSS gibt es viele verschiedene Fenster. Die wichtigsten werden hier beschrieben.

3.1.1 Der Dateneditor

Der Dateneditor besteht aus zwei Fenstern, die Datenansicht und die Variablenansicht

Abbildung 1: Die Datenansicht

id	customer	revenue	support	region	industry
1	1	\$3.787	3	4	1
2	2	\$1.734	3	2	2
3	3	\$2.126	3	4	1
4	4	\$2.269	3	4	2
5	5	\$1.267	4	1	1
6	6	\$0	4	4	3
7	7	\$1.838	4	2	1
8	8	\$1.847	4	4	1
9	9	\$1.714	2	3	3
10	10	\$1.718	4	2	3
11	11	\$4.396	1	4	1
12	12	\$3.156	3	1	2
13	13	\$3.834	4	4	1
14	14	\$2.291	4	1	2
15	15	\$4.140	2	2	2
16	16	\$2.194	2	2	3
17	17	\$2.938	2	2	1
18	18	\$3.313	4	3	2
19	19	\$3.327	4	3	3
20	20	\$1.449	2	3	1
21	21	\$2.696	2	1	3
22	22	\$3.272	3	2	2
23	23	\$4.086	4	1	3
24	24	\$4.340	2	2	3
25	25	\$2.462	2	4	1
26	27	\$3.630	3	3	2
27	28	\$465	3	4	3
28	29	\$1.871	2	1	2
29	30	\$1.377	1	4	3
30	31	\$3.797	1	2	1
31	33	\$4.253	2	1	3

Abbildung 2: Die Variablenansicht

Name	Typ	Spaltenformat	Dezimalstellen	Variablenlabel	Wertelabels	Fehlende Wert	Spalten	Ausrichtung	Maßstab
id	Numerisch	4	0	RESPONDEN	Kein	Kein	8	Rechts	Metrisch
customer	Numerisch	1	0	Customer Stat	{1, Regular cu	Kein	8	Rechts	Ordinal
revenue	Dollar	8	0	Revenue	Kein	\$0, \$8, \$9	8	Rechts	Metrisch
support	Numerisch	8	0	Time on Hold	{1, < 1 Minute}	Kein	8	Rechts	Ordinal
region	Numerisch	2	0	Territory	{1, North}...	Kein	8	Rechts	Ordinal
industry	Numerisch	2	0	Industry	{1, Governmen	Kein	8	Rechts	Ordinal

3.1.2 Der Viewer

Im Viewer werden alle statistischen Ergebnisse, Tabellen und Diagramme angezeigt. Die Ausgaben können bearbeitet und abgespeichert werden. Das Fenster wird automatisch geöffnet. Es empfiehlt sich, den Viewer so einzustellen, dass die SPSS-Kommandos im Output miterscheinen. Anhand von diesem Output kann man immer kontrollieren, ob die richtigen Cases ausgewählt sind und ob die richtigen Prozeduren angesteuert wurden. Auch kann man die Kommandostruktur in das Syntax-Fenster kopieren und so seine Analysen abspeichern.

Abbildung 3: Der Viewer

The screenshot shows the SPSS Viewer window titled 'Ausgabe1 - SPSS Viewer'. The menu bar includes 'Datei', 'Bearbeiten', 'Ansicht', 'Einfügen', 'Format', 'Analysieren', 'Grafiken', 'Extras', 'Fenster', and 'Hilfe'. The toolbar contains various icons for file operations and analysis. The left pane shows a tree view with 'Ausgabe' expanded to 'Log', 'Häufigkeiten', 'Titel', 'Anmerkungen', 'Statistiken', 'Territory', and 'Kreisdiagramm'. The main area displays the following content:

```
FREQUENCIES
  VARIABLES=region
  /PIECHART FREQ
  /ORDER= ANALYSIS .
```

Häufigkeiten

Statistiken

Territory		
N	Gültig	1500
	Fehlend	0

Territory

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	North	382	25,5	25,5	25,5
	South	345	23,0	23,0	48,5
	East	375	25,0	25,0	73,5
	West	398	26,5	26,5	100,0
Gesamt		1500	100,0	100,0	

Territory

The pie chart is titled 'Territory' and shows four segments. The 'West' segment is the largest, followed by 'North', 'East', and 'South'. The x-axis is labeled 'West' and 'North'.

SPSS Prozessor ist bereit

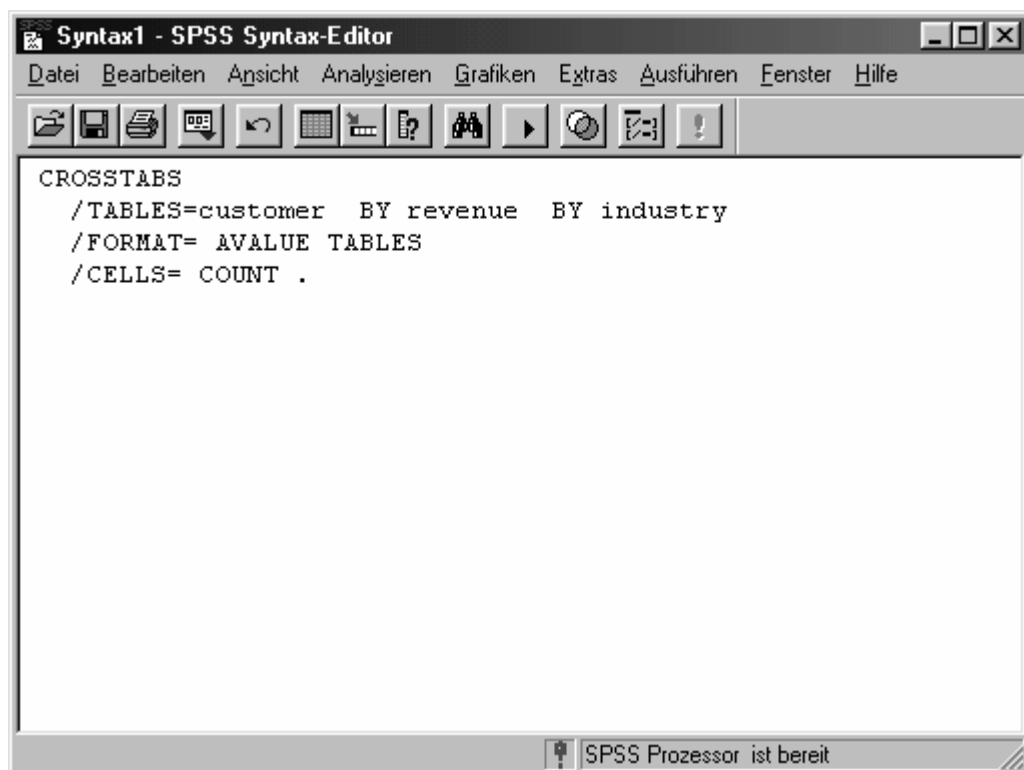
Die Ausgaben im Viewer kann man fast beliebig editieren. Wenn man auf das Chart doppelklickt, kommt man in dem Diagramm-Editor. Hier kann man Farben und Schriftarten abändern, Achsen vertauschen oder den Diagramm-Typ ändern.

Wenn man auf den Tabellen doppelklickt, kommt man in den Pivot-Tabellen-Editor. Hier kann man Text bearbeiten, die Daten in Zeilen und Spalten austauschen, Farben hinzufügen etc.

3.1.3 Der Syntax-Editor

Der Syntax-Editor wird beim Öffnen des Programmes nicht automatisch geöffnet. Man hat allerdings die Möglichkeit, SPSS so einzustellen, dass der Syntax-Editor beim Starten des SPSS automatisch geöffnet wird. Arbeiten mit dem Syntax-Editor hat zwei entscheidene Vorteile. Der Erste ist, dass man seine Analysen im Vorfeld schon planen und programmieren kann, wodurch man im jeden Zeitpunkt im Datenerfassungsprozess schon mal vorläufige Ergebnisse rechnen kann, ohne sich jedesmal neu durch das Menü klicken zu müssen. Der zweite Vorteil ist, dass man über den Syntax Funktionen in SPSS nutzen kann, die im Menü nicht aufgeführt sind. Wenn man die Syntax im Viewer mit ausgeben läßt, kann man ihn direkt in den Syntax-Editor hineinkopieren und damit abspeichern für zukünftige Analysen. Im Syntaxfenster kann man diese Jobs weiterbearbeiten. SPSS hat zur Syntax eine ausführliche Online-Hilfe.

Abbildung 4: Der Syntax-Editor



3.2 Menüs

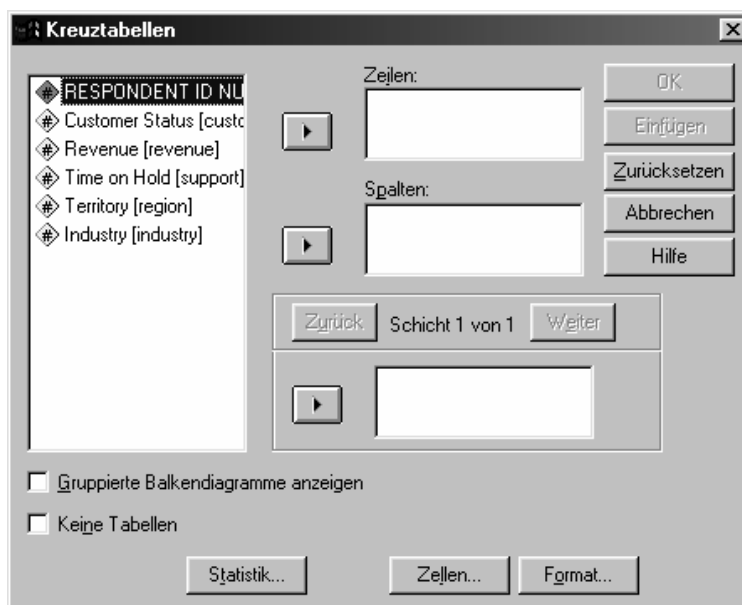
Viele Ausgaben die SPSS ausführen kann, können über Menübefehle angesteuert werden. Jedes Fenster in SPSS besetzt eine eigene Menüleiste mit verschiedenen Menüs für dieses Fenster. Die Menüs „Analysieren“ und „Grafiken“ sind in allen Menüs verfügbar.

3.3 Dialogfelder

Wenn Sie in SPSS eine Option aus dem Menü auswählen, wird in den meisten Fällen ein Dialogfeld geöffnet. In diesen Dialogfeldern können Sie Variablen und Optionen für die Analyse auswählen. Die Hauptdialogfelder der Prozeduren und Diagramme bestehen aus den folgenden Elementen:

1. Liste der Quellvariablen. In diese Liste werden die Variablen in der Arbeitsdatei gezeigt
2. Liste der Zielvariablen. Hier werden die Variablen angezeigt, die Sie für die Analyse ausgewählt haben
3. Befehlsschaltflächen. Hiermit können Sie Optionen einstellen oder Prozeduren starten

Abbildung 5: Dialogfeld der Prozedur Kreuztabellen



4 Die Dateneingabe

Es gibt verschiedene Möglichkeiten, die Daten aus den Fragebögen in einen Datensatz zu bringen. Ein paar von diesen Möglichkeiten werden hier besprochen. Legen Sie bei der Dateneingabe z.B. ein Buch unter Ihre Handgelenke, damit diese auf gleicher Höhe wie die Tastatur sind. Arbeiten Sie bei der Dateneingabe nie länger als drei Viertelstunden nacheinander und machen Sie dann eine Viertelstunde Pause. Nach dieser Zeit läßt die Konzentration erheblich nach. Hiermit stellen Sie sicher, dass Sie möglichst wenig Fehler machen und geben Ihre Handmuskulatur Zeit, sich zu erholen. Hiermit beugen Sie z.B. einer Sehnenscheidenentzündung oder einer Repetitive Strain Verletzung vor.

4.1 SPSS-Dateneditor

Eine Möglichkeit ist es, die Daten direkt in SPSS einzugeben. Man startet das SPSS und geht in das Fenster „Variablenansicht“, wo man die Möglichkeit hat, die Variablen, die man eingeben möchte, zu definieren. Über das Zuweisen von Labels, fehlenden Werten etc. finden Sie mehr in Abschnitt 5.2.

Abbildung 6: Der SPSS-Dateneditor: Variablenansicht

	Name	Typ	Spaltenformat	Dezimalstellen	Variablenlabel	Wertelabels	Fehlende Wert	Spalten	Ausrichtung	Maßniveau
1	nr	Numerisch	8	0	ID	Kein	Kein	8	Rechts	Nominal
2	v1	Numerisch	8	0	Geschlecht	{1, Mann}...	8	8	Rechts	Nominal
3	v2	Numerisch	8	0	ZDF	{1, sehr gut}...	9	8	Rechts	Ordinal
4	v3	Numerisch	8	0	Pro 7	{1, sehr gut}...	9	8	Rechts	Ordinal
5	v4	Numerisch	8	0	MTV	{1, sehr gut}...	9	8	Rechts	Ordinal
6	v5	Numerisch	8	0	RTL	{1, sehr gut}...	9	8	Rechts	Ordinal
7										
8										
9										
10										
11										
12										
13										
14										
15										
16										
17										
18										
19										
20										
21										
22										
23										
24										
25										
26										
27										
28										
29										
30										
31										
32										

Wenn dieser Schritt abgeschlossen ist, wechselt man zu dem Fenster „Datenansicht“, wo man die Daten direkt eintippen kann. Man wechselt mit der „TAB-Taste“ oder mit den „Cursortasten“ zur nächsten Zelle.

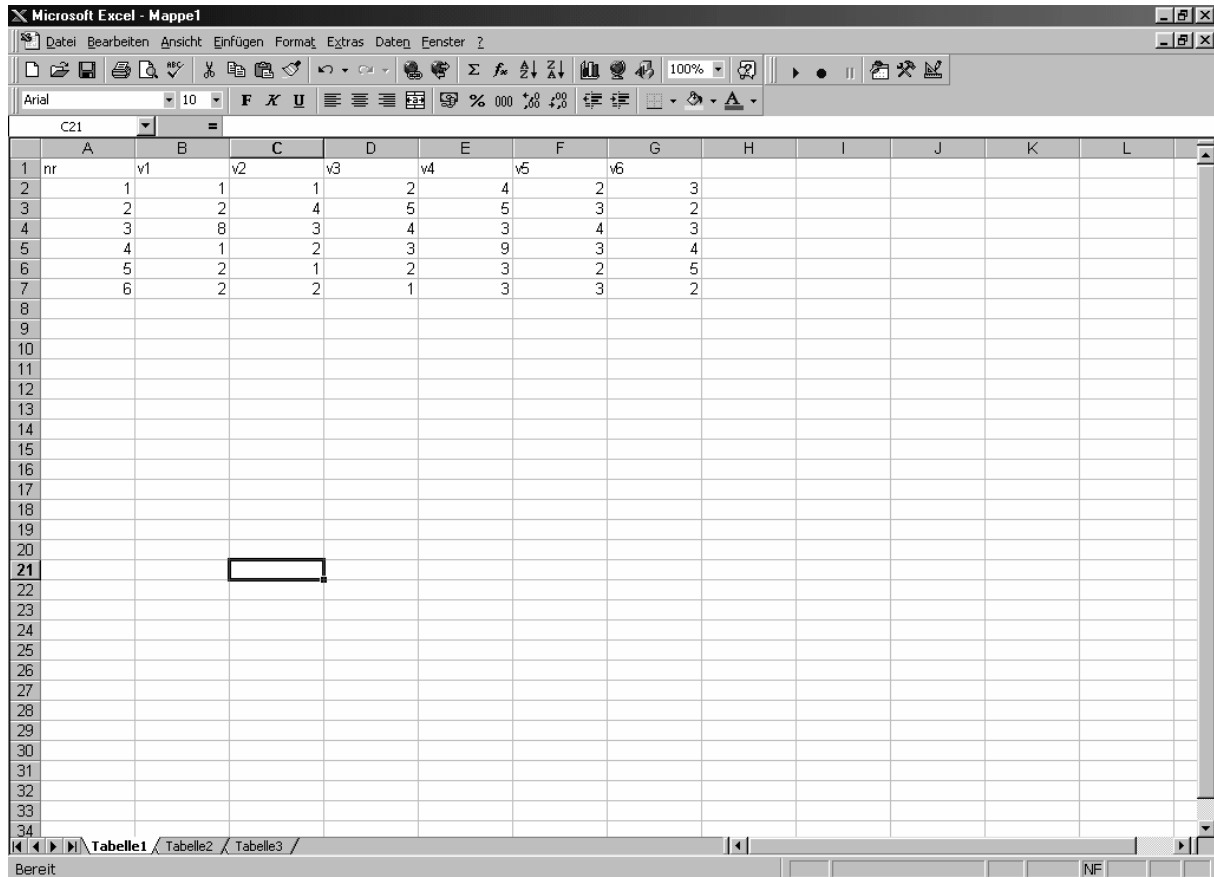
Abbildung 7: SPSS Daten-Editor: Datenansicht

	id	v1	v2	v3	v4	v5	v6	var	var	var	var	var	var
1	1	2	3	2	5	4	18						
2	2	3	3	0	4	3	34						
3	3	1	2	4	3	3	38						
4	4	2	2	2	3	1	62						
5	5	1	2	1	4	3	73						
6	6	3	3	2	5	4	43						
7													
8													
9													
10													
11													
12													
13													
14													
15													
16													
17													
18													
19													
20													
21													
22													
23													
24													
25													
26													
27													
28													
29													
30													
31													

4.2 Excel

Auch in Excel hat man die Möglichkeit Daten einzugeben. Die eigentliche Dateneingabe erfolgt analog zu der in SPSS.

Abbildung 8: Der Datenmatrix in Excel



The screenshot shows a Microsoft Excel spreadsheet with a data matrix. The columns are labeled 'nr', 'v1', 'v2', 'v3', 'v4', 'v5', and 'v6'. The rows are numbered 1 through 7. The data is as follows:

nr	v1	v2	v3	v4	v5	v6
1	1	1	1	2	4	2
2	2	2	4	5	5	3
3	3	8	3	4	3	4
4	4	1	2	3	9	3
5	5	2	1	2	3	2
6	6	2	2	1	3	3
7						

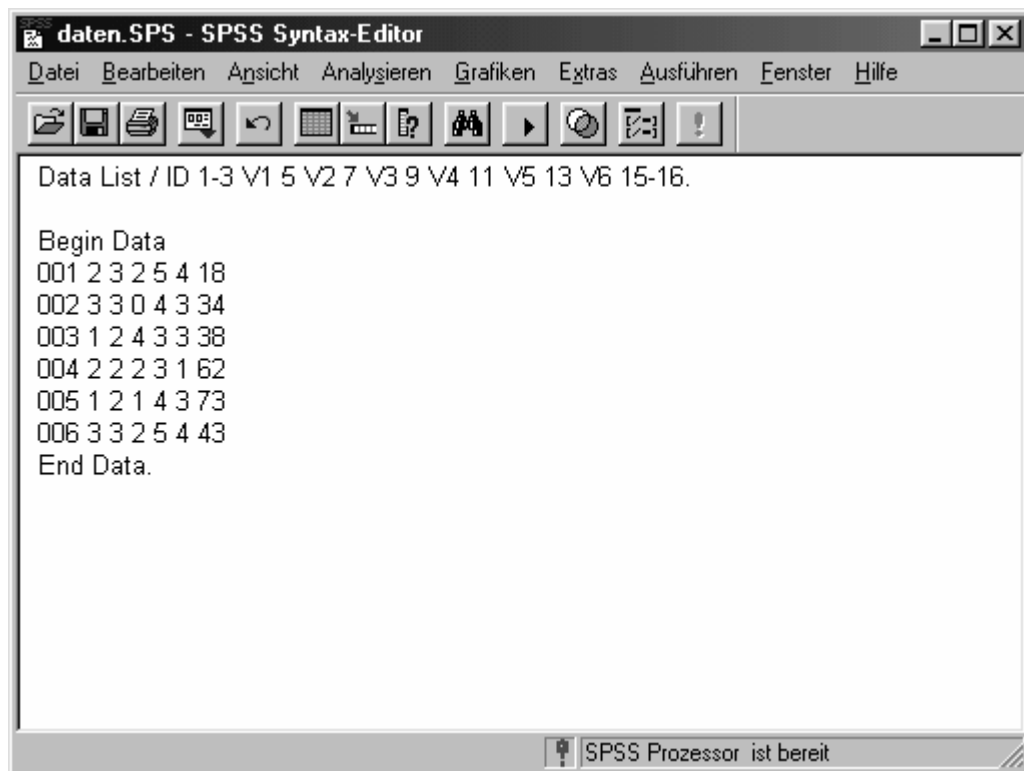
Obwohl man offiziell die Möglichkeit hat, einen Excel-Datensatz direkt in SPSS einzulesen, klappt dies sehr oft nicht. Speichern Sie in diesem Fall den Datensatz in dBase4-Format. Dieses Format liest SPSS in jedem Fall. Siehe hierzu auch Abschnitt 5.1.

4.3 ASCII

Wenn man seine Daten auf ASCII-Format eingibt, fängt man „einfach an zu tippen“. Hierbei muß ganz genau dokumentiert werden, auf welche Spalten welche Variablen kodiert sind. Um dieses ASCII-File in SPSS einzulesen, benutzt man das SPSS-Kommando „Data List“ oder „Data List File“ oder man benutzt den Assistenten für Textimport.

Das Data List-Kommando führt man ein in dem Syntax-Editor von SPSS und sieht aus wie folgt:

Abbildung 9: Der SPSS Syntax-Editor



Im Kommando Data List wird beschrieben, wie die einzelnen Variablen heißen und in welchen Spalten der Datenmatrix man sie findet. Denken Sie daran, dass Sie für mehrstelligen Variablen wie Alter und ID genügend Platz reservieren. Wenn man die Zeilen markiert im Syntax-Fenster und auf das Symbol ▶ klickt, arbeitet SPSS die Syntax ab und wird der Datenmatrix erstellt.

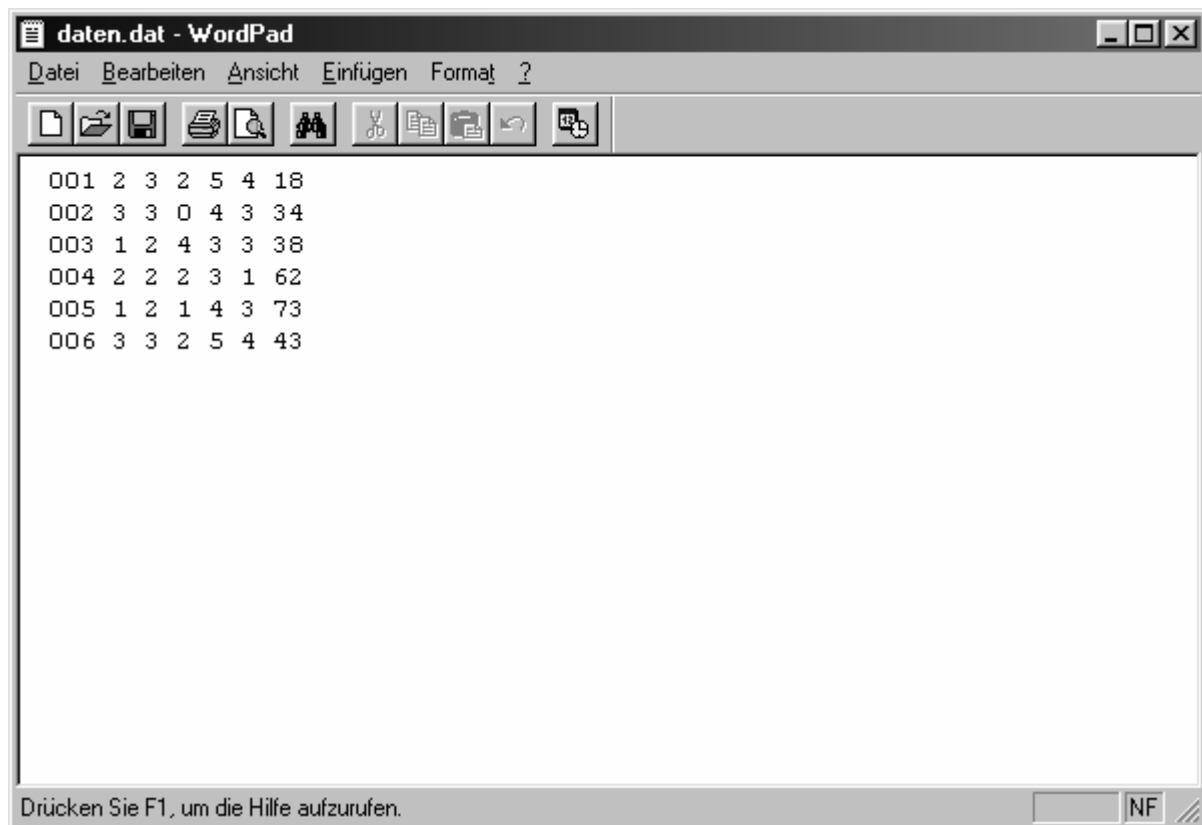
Eine andere Möglichkeit ist es, die Daten in einem Texteditor einzugeben und mit der Endung „.dat“ abzuspeichern. Zum Einlesen dieser Daten in SPSS wendet man das Kommando „Data List File“ an.

Beispiel:

```
Data List File = daten.dat / ID 1-3 V1 5 V2 7 V3 9 V4 11 V5 13 V6 15-16.
```

Das File „.daten.dat“ wird mit einem Texteditor erstellt bzw. bearbeitet. In diesem File stehen nur die Daten (siehe Abbildung 10).

Abbildung 10: Daten.dat



Die dritte Möglichkeit zum Einlesen von ASCII-Daten ist über das Menü. Man wählt in SPSS die Option „Datei öffnen“ und daraus die Option „Daten“. Als Datei-Typ wählt man „Daten“ (*.dat) und selektiert das File, das man einlesen möchte. Hierauf öffnet sich der Assistent für Textimport, worin man Angaben über seine Daten machen kann. Hier kann man sich wie in anderen Windows-Anwendungen durchklicken, wonach die Datenmatrix fertiggestellt werden kann und in SPSS-Format (*.sav) abgespeichert werden kann.

4.4 Das Programmieren einer Eingabemaske mittels Access

Wenn Sie große Mengen Daten zu verarbeiten haben, empfiehlt es sich sehr, eine Eingabemaske zu programmieren. Die oben beschriebenen Methoden sind sehr anfällig für Fehler und bieten nicht die Möglichkeit, Fehleingaben im Vorfeld auszuschließen.

Zur Programmierung von Eingabemasken gibt es verschiedene Hilfsmitteln, wie das SPSS Data Entry Modul und Access. Da die Wahrscheinlichkeit, dass Sie ein Access zu Verfügung haben, wesentlich größer ist als die Wahrscheinlichkeit, dass Sie SPSS Data Entry haben (vor

allem weil diese Komponente separat gekauft werden muß), werde ich hier das Programmieren von Eingabemasken mittels Access beschreiben.

Erstellen Sie eine neue Datenbank, in diesem Falle daten.mbd. Danach wählen Sie im Blatt „Tabellen“ die Option „Neu – Entwurfsansicht“.

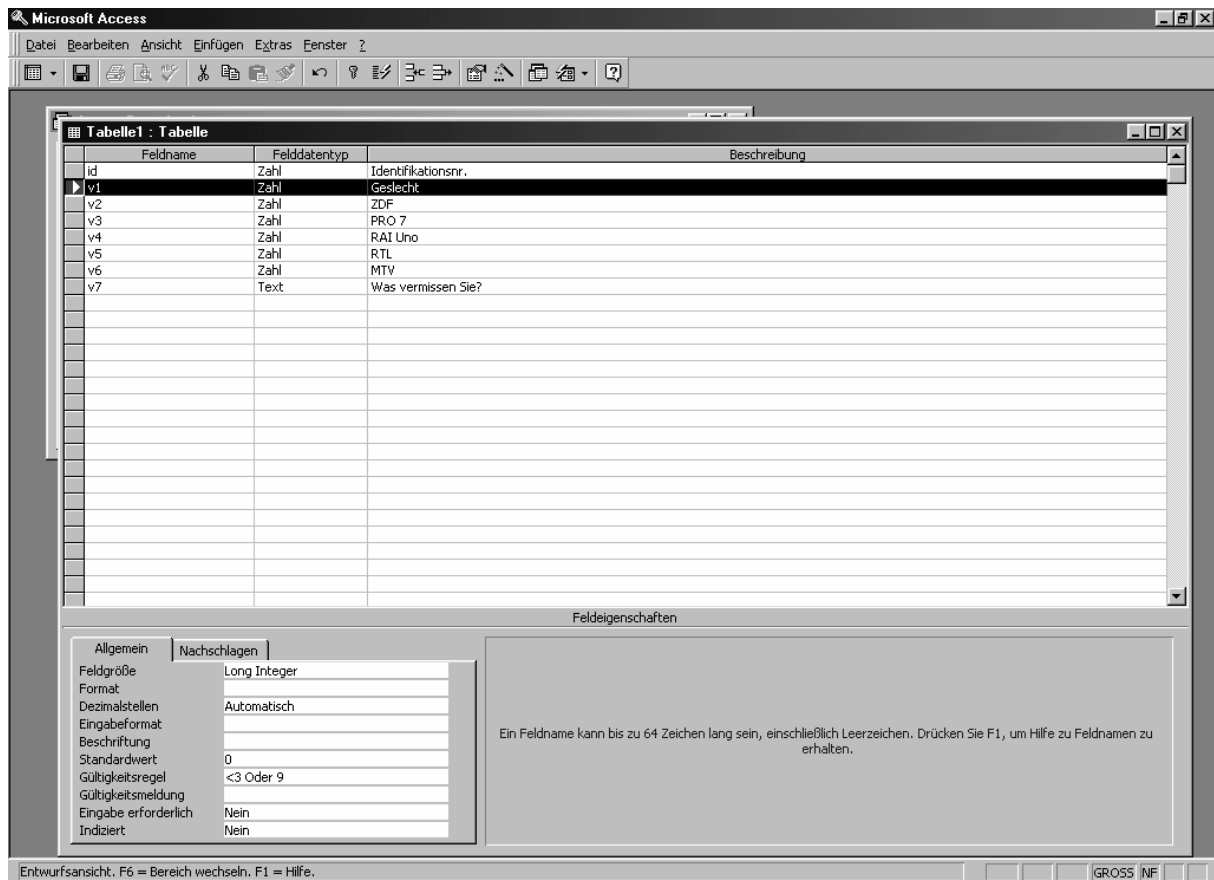
Abbildung 11: Access: Neue Tabelle



Daraufhin öffnet sich folgendes Fenster (Abbildung 12), in dem Sie die Angaben aus Ihrem Codebook verarbeiten können. In diesem Fenster haben Sie auch die Möglichkeit, die Eingabemöglichkeiten zu beschränken. In diesem Beispiel (Abbildung 12) sind die gültigen Eingabecodes für die Variable V1 (Geschlecht) 1 (Mann), 2 (Frau) und 9 (Keine Angabe).

Wenn man die Tabelle abspeichert, fragt Access, ob man einen primären Schlüssel definieren möchte. Es empfiehlt sich, die Variable ID nicht als primären Schlüssel zu definieren. Access numeriert die Schlüsselvariable selbst durch, und dies ist nicht erwünscht, da die ID-Nummer mit der Nummer auf den Fragebögen übereinstimmen muß. Manchmal ist es sinnvoll, an bestimmte Gruppen von Respondenten verschiedene Nummernkontingente zu vergeben, z.B. dann, wenn es für bestimmten Gruppen Zusatzfragen im Fragebogen gibt. Ein Beispiel hierfür ist eine Mitarbeiterbefragung, wo die Führungskräfte einen Abschnitt „Fragen zur Führungsverantwortung“ in ihrem Fragebogen haben, der bei den Mitarbeitern nicht vorhanden ist. Hier könnte man – um bei der Datenbereinigung eine extra Plausibilitätscheck zu haben – die Identifikationsnummern 1 bis 50 für die Führungskräfte reservieren und für die Mitarbeiter die ID-Nummern 51 bis x.

Abbildung 12: Access: Entwurfsmodus Tabelle



Nachdem man die Tabelle abgespeichert hat, schließt man das Fenster. Man kommt jetzt in einen Dialog, wo man das Formular für die Dateneingabe erstellen kann. Hier wählt man im Menü „Formulare“ die Option „Entwurfsansicht“ und man stellt die Tabelle ein, aus der die Daten kommen (Abbildung 13). Man klickt auf „OK“ und kommt im folgenden Fenster, wo man seine Eingabemaske erstellen kann (Abbildung 14). Über den Menüpunkt „Ansicht“ blendet man die Feldliste ein, aus der man die Variablen auf das Formular ziehen kann. Auf dem Formular werden dann der Variablenname und ein Textfeld zur Eingabe der Werte erzeugt. Mit Tools wie Labels etc. kann man das Formular nach eigenem Wunsch gestalten (Abbildung 15).

Abbildung 13: Neues Formular

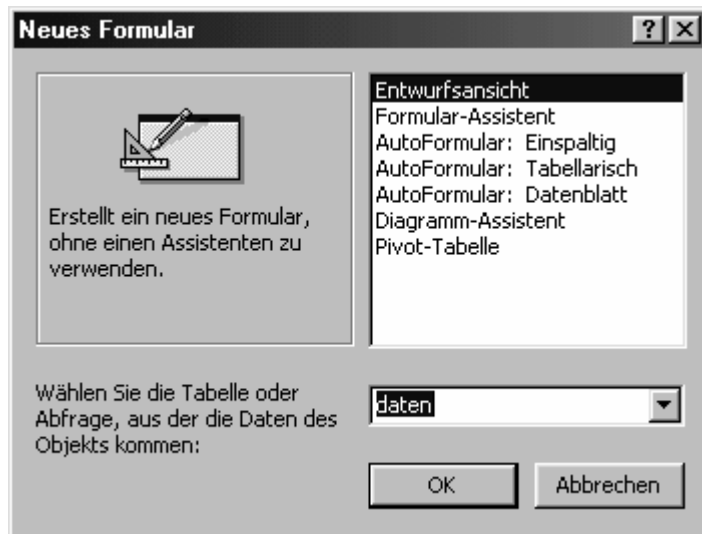


Abbildung 14: Formular Entwurfsansicht

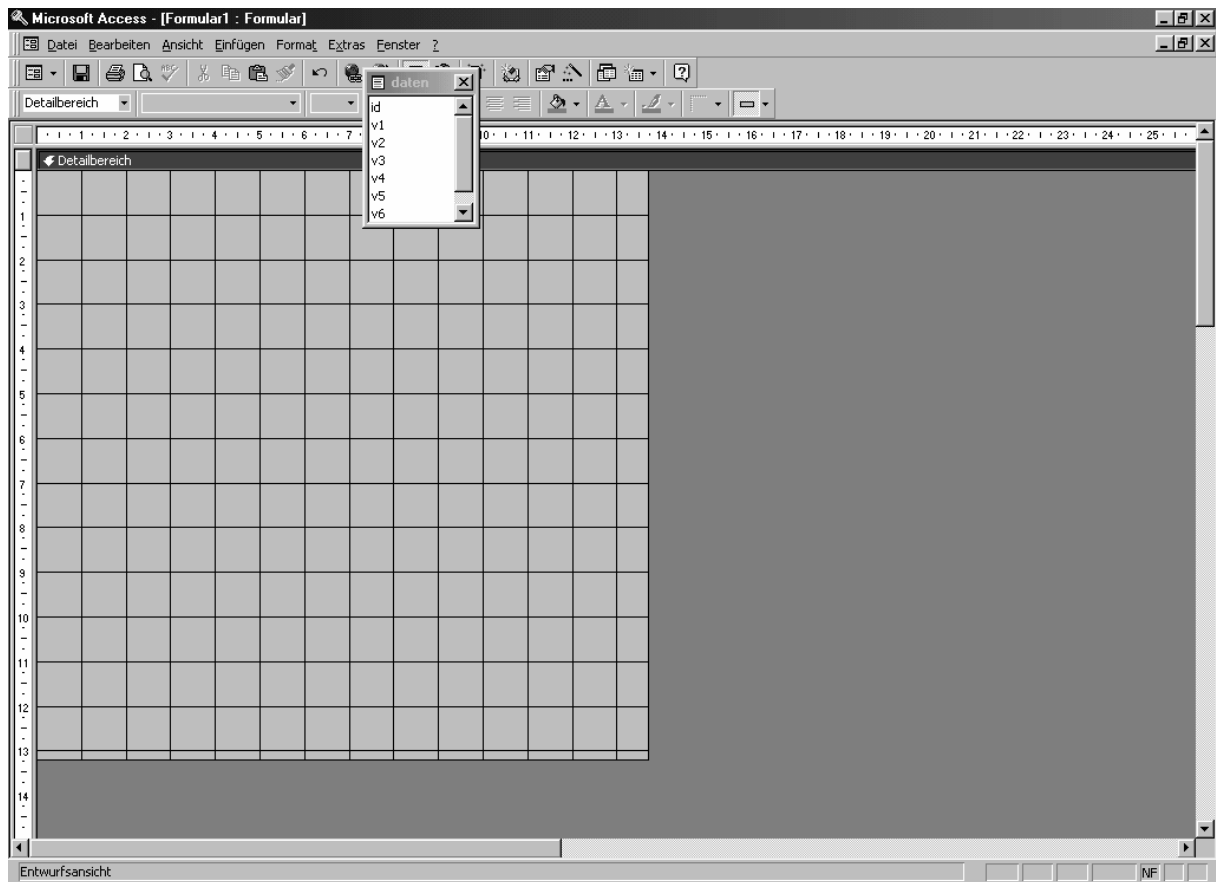
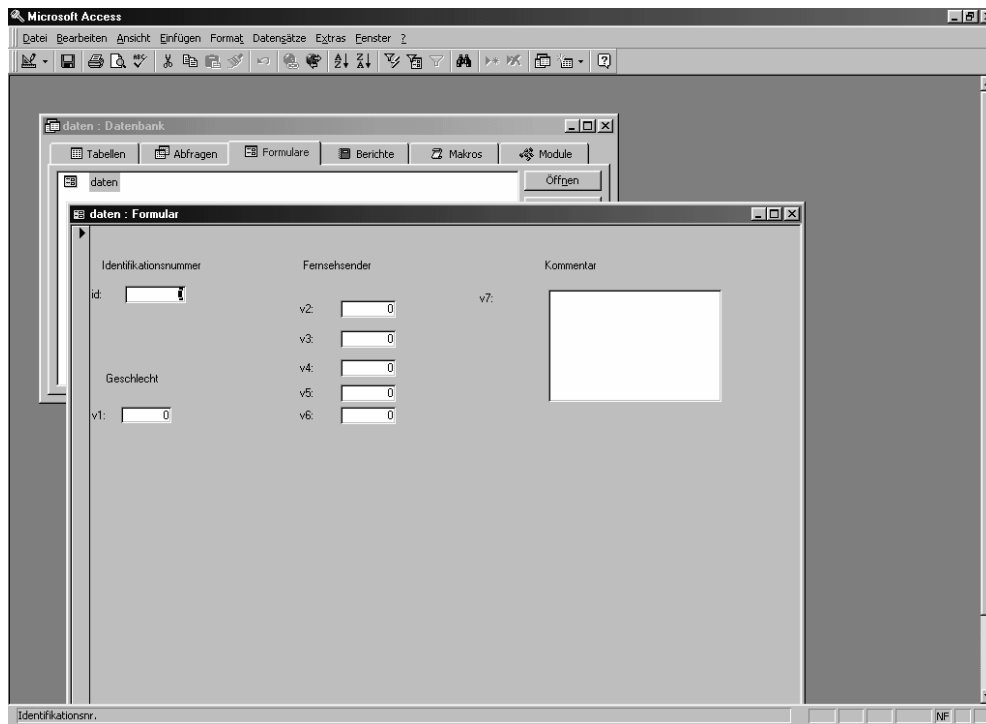
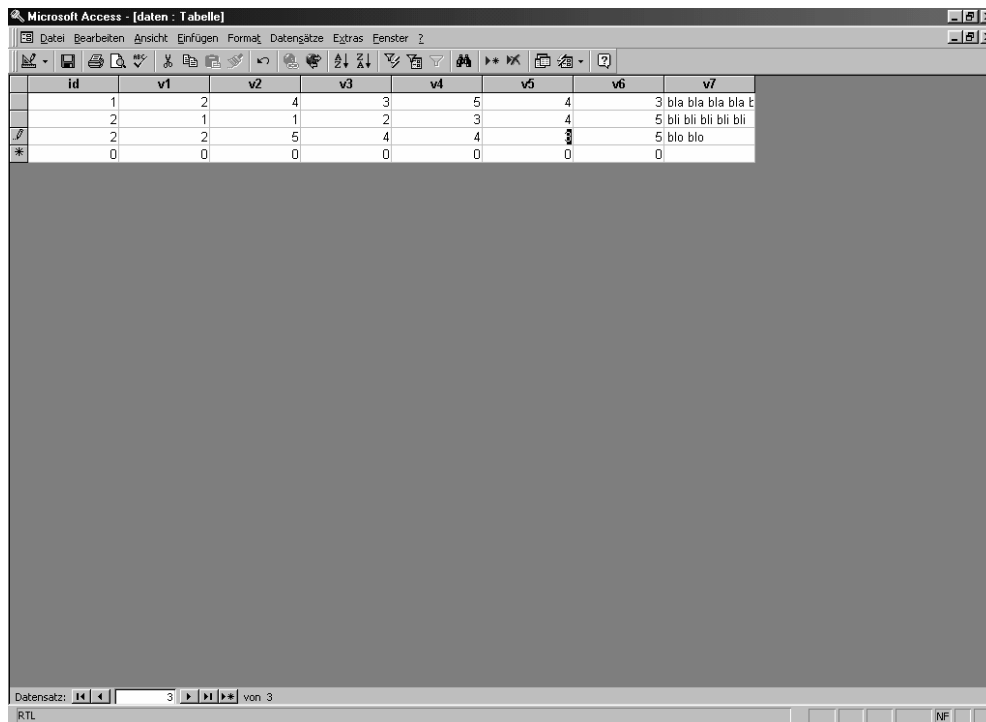


Abbildung 15: Die Eingabemaske



Wenn man nun von der Entwurfsansicht in die Formularansicht wechselt (Icon links oben in der Iconleiste), kann die Eingabemaske verwendet werden (speichern nicht vergessen!). In die Textfelder können nun die Daten eingegeben werden. Bei den numerischen Variablen muß die Eingabe mit der Return-Taste bestätigt werden, bei Textvariablen mit der TAB-Taste. Der Cursor springt dann zum nächsten Feld. Nach dem letzten Feld werden die Daten in der Tabelle gespeichert und die Eingabemaske wird für einen neuen Case initialisiert.

Abbildung 16: Die Datenbankansicht



Wenn man im Nachhinein eine Variable hinzufügen oder an einer anderen Stelle einfügen möchte (zwischen zwei bestehenden Variablen), muss man nicht nur die Kästchen in der Entwurfsansicht neu positionieren, sondern auch die Aktivierreihenfolge (Unter: „Ansicht“, „Aktivierreihenfolge“) abändern (Abbildung 17). Sonst wird die neue Variable als letzte aktiviert, unabhängig davon, wo die Variable im Entwurfsformular dargestellt ist. Somit ist die Chance sehr groß, dass man die Variablen nicht in der erwarteten Reihenfolge und damit falsch eingibt.

Abbildung 17: Anpassen Aktivierreihenfolge



5 Der Datensatz

5.1 Die Umwandlung der Daten in SPSS-Format

Es gibt viele Methoden Daten, die in einem anderen Format erzeugt wurden, in SPSS-Format zu konvertieren. Bestimmte Formate, wie das Systat-Format kann SPSS ohne weiteres lesen. Mit anderen Formaten (wie z.B. die neuesten Excel-Versionen) hat SPSS Schwierigkeiten, obwohl das Programm die Konvertierung offiziell vornehmen kann. Wenn es Ihnen geschieht, dass SPSS den von Ihnen ausgewählten Datensatz nicht lesen kann, speichern Sie Ihre Daten am Besten als dBase 4 File. Dieses Format liest SPSS garantiert. Die Chance ist allerdings relativ groß, dass SPSS den Datensatz zwar importiert, aber nicht die dazugehörige Labels (Namen von Variablen und Kategorien). Es empfiehlt sich daher, die Labels erst zuzukennen, wenn die Daten auf SPSS-Format gebracht sind. Rohdatensätze, in der die Daten für jeden Respondent in einer Zeile stehen, lassen sich, wenn die Spalten durch ein Trennzeichen wie ein Komma, ein Leerzeichen oder ein Semikolon getrennt sind, über das SPSS-Menü leicht importieren. Bei Daten, in den ein Respondent über mehreren Zeilen definiert wurde, oder in den zwischen den Spalten kein Trennzeichen steht, wie in unterstehendem SAS-Job, ist man allerdings gezwungen, die Ansteuerung in SPSS-Code umzuschreiben (über die Prozedur Data List).

Beispiel SAS-Code:

```
DATA XYZ;
  TITLE 'Beispiel';
  INPUT ID $ 1-3 EVERCONT $4 SATOVER 5 SATORDER 6 SATSTATU 7 SATDEL1 8 SATCONTR 9 SATPRISS 10
  SATREPAI 11 SATCRED 12 SATBILL 13 SATTECHQ 14 SATOTHER 15 ECONT $16 TCONT $17 FCONT $18
  CONTOTH $19 PECONT $20 PTCONT $21 PFCONT $22 PCONTOTH $23 ECONV 24 EFIND 25 ETIMESAT 26 TCONV
  27 TFIND 28 NOBUSY 29 TTIMESAT 30 FCONV 31 FFIND 32 FTIMESAT 33 PININFO 34 HOURS 35 SATREP 36
  REPKNOW 37 REPHelp 38 REPFRIEN 39 REPCLPRO 40 REPDELPR 41 REPOWN 42 REPNEXT 43 SATSOL 44
  SOLCOMPL 45 SOLEXCEE 46 EXPLONCE 47 EFFICIE 48 PROACTIV 49 PROACTOR 50 PRIDISCR 51 SOLQUICK
  52/

  SATINV 1 INVACCQ 2 INVACCP 3 INVDISC 4 INVAMOUN 5 INVLEGAL 6 INVTIME 7 SATDEL2 8 DELREC 9
  DELCHANG 10 DELCHCON 11 DELCOMPL 12 DELACCU 13 DELTIME 14 CLAIM $15 SATCLAIM 16 CLFAIR 17
  CLQUICK 18 FREQCONT $19 COUNTRY $20-21 SINCE $22 EMPLOYEE $23 CRORDER $24 CRORDST $25 CRPRESAL
  $26 CRCONTR $27 CRPRICE $28 CRREPAIR $29 CRCRED $30 CRBILL $31 CROTHER $32 CONTACT $33;
  CARDS;
  0011462272002301100110077777770007565777777641131333
  111111721111110...203251100100111
  002167576507777110100077757776677767776757666565557
  77777766667760...204231010100001
  00314731530360.1110110074177747747413471111111111111
  77777711111111147432122110011110.
  00415544530540.111110007757735775576466555555445555
  4434554533343143342122100010100.
  0051777774657.110011007777477777770777775777767
  76777766665461674821350000000011
  00612662665223.11001100777666723563727776.6222666466
  225555511111111131022110001001.
  00711322046444.1110100077744762761456772264433342164
  33666631111211664303130100010000
  00814643555556.111010007766666...55776466453643454
  .....1111211666125351100000001
  ;
run;
```

Diese Code sieht in SPSS-Format umgeschrieben so aus:

```
Data List records = 2
/ ID 1-3 EVERCONT 4 SATOVER 5 SATORDER 6 SATSTATU 7 SATDEL1 8 SATCONTR 9 SATPRISS 10 SATREPAI
11 SATCRED 12 SATBILL 13 SATTECHQ 14 SATOTHER 15 ECONT 16 TCONT 17 FCONT 18 CONTOTH 19 PECONT
20 PTCONT 21 PFCONT 22 PCONTOTH 23 ECONV 24 EFIND 25 ETIMESAT 26 TCONV 27 TFIND 28 NOBUSY 29
TTIMESAT 30 FCONV 31 FFIND 32 FTIMESAT 33 PININFO 34 HOURS 35 SATREP 36 REPKNOW 37 REPHelp 38
REPFRIEN 39 REPCLPRO 40 REPDELPR 41 REPOWN 42 REPNEXT 43 SATSOL 44 SOLCOMPL 45 SOLEXCEE 46
EXPLONCE 47 EFFICIE 48 PROACTIV 49 PROACTOR 50 PRIDISCR 51 SOLQUICK 52
/SATINV 1 INVACCQ 2 INVACCP 3 INVDISC 4 INVAMOUN 5 INVLEGAL 6 INVTIME 7 SATDEL2 8 DELREC 9
DELCHANG 10 DELCHCON 11 DELCOMPL 12 DELACCU 13 DELTIME 14 CLAIM 15 SATCLAIM 16 CLFAIR 17
CLQUICK 18 FREQCONT 19 COUNTRY 20-21 SINCE 22 EMPLOYEE 23 CRORDER 24 CRORDST 25 CRPRESAL 26
CRCONTR 27 CRPRICE 28 CRREPAIR 29 CRCRED 30 CRBILL 31 CROTHER 32 CONTACT 33 .

Begin data
0011462272002301100110077777770007565777777641131333
111111721111110...203251100100111
0021675765077771110100077757776677767776757666565557
77777766667760...204231010100001
00314731530360.1110110074177747747413471111111111111
7777777111111147432122110011110.
00415544530540.111110007757735775576466555555445555
4434554533343143342122100010100.
00517777774657.1100110077774777777770777775777767
76777766665461674821350000000011
00612662665223.11001100777666723563727776.6222666466
225555511111111131022110001001.
00711322046444.1110100077744762761456772264433342164
336666311111211664303130100010000
00814643555556.111010007766666...55776466453643454
.....11111211666125351100000001
end data.
```

Nach dem Data-List-Kommando ist definiert, dass die Daten sich per Respondent über zwei Zeilen ausstrecken (records=2). Das Zeichen „/“ leitet die jeweilige Zeile ein.

5.2 Datendefinition

Nachdem die Daten in SPSS-Format gebracht wurden, kann man an den Variablen und deren Kategorien Labels zuweisen und Werten als „Missing“ definieren. Dies geht sowohl über den Syntax als über das Menü. Beide Methoden werden hier besprochen. Ohne das Zukennen von Labels sieht die Häufigkeitsverteilung der Variable „Satover“ wie folgt aus:

SATOVER					
		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	1	1	10,0	10,0	10,0
	2	1	10,0	10,0	20,0
	4	3	30,0	30,0	50,0
	5	1	10,0	10,0	60,0
	6	2	20,0	20,0	80,0
	7	2	20,0	20,0	100,0
	Gesamt	10	100,0	100,0	

5.2.1 Variable Labels

Ein „Variable Label“ ist der Name der Variable. Der Name der Variable „ID“ kann zum Beispiel „Identifikationsnummer“ sein. Der Name der Variable „Satover“ ist in diesem Beispiel „Overall Satisfaction“. Im Menü kann man in dem Daten-Editor im Fenster „Variablenansicht“ dieses Label definieren. Dazu klicken Sie in der Spalte „Variablenlabel“ und geben den Namen ein.

Im Syntax sieht es so aus:

```
Variable Label ID 'Identifikationsnummer' Satover 'Overall Satisfaction' .
```

Nach dem Zukennen des Variablenlabel sieht die Häufigkeitsverteilung der Variable „Satover“ aus wie folgt:

Overall satisfaction

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	1	1	10,0	10,0	10,0
	2	1	10,0	10,0	20,0
	4	3	30,0	30,0	50,0
	5	1	10,0	10,0	60,0
	6	2	20,0	20,0	80,0
	7	2	20,0	20,0	100,0
	Gesamt	10	100,0	100,0	

Sie sehen, dass SPSS im Viewer den Variablennamen ausgibt.

5.2.2 Value Labels

Value Labels, auf Deutsch Wertelabels, sind die Namen der einzelnen Kategorien der Variable. Das Zukennen der Wertelabels machen Sie im gleichen Fenster wie die Variable Labels. Im Syntax sieht die Zukennung von Wertelabels an der Variable „Satover“ aus wie folgt:

```
Value labels Satover 1 'Very unsatisfied' 2 'Unsatisfied' 3 'A bit unsatisfied' 4 'Neutral' 5 'A bit satisfied' 6 'Satisfied' 7 'Very satisfied' .
```

Die Ausgabe der Prozedur "Häufigkeiten" nach dem Zukennen der Wertelabels steht im folgenden Beispiel.

Overall Satisfaction

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	Very unsatisfied	1	10,0	10,0	10,0
	Unsatisfied	1	10,0	10,0	20,0
	Neutral	3	30,0	30,0	50,0
	A bit satisfied	1	10,0	10,0	60,0
	Satisfied	2	20,0	20,0	80,0
	Very satisfied	2	20,0	20,0	100,0
	Gesamt	10	100,0	100,0	

5.2.3 Missing Values

Das Zukennen von fehlenden Werten, was man auch in dem Daten-Editor machen kann, ist extrem wichtig. Falls Sie in einer Variable die fehlenden Werte oder die Kategorie „Weiß nicht“ mit „9“ kodiert haben, und dem SPSS nicht mitteilen, dass es sich hier um einen fehlenden Wert handelt, bekommen Sie falsche Ergebnisse.

Ein Beispiel:

Die Frage AB7 „Alles im Allen bin ich zufrieden mit meinen Arbeitsbedingungen“ wird von 5 Respondenten auf folgender Art und Weise beantwortet:

- 1 (sehr zufrieden) 2 Respondenten
- 3 (teils/teils) 1 Respondent
- 4 (unzufrieden) 1 Respondent
- 9 (keine Angabe) 1 Respondent

Falls Sie nun SPSS nicht mitteilen, dass der Wert „9“ ein fehlender Wert ist, und Sie geben das Programm den Auftrag, den Mittelwert zu berechnen, bekommen Sie folgender Wert:

$$(2 \times 1 + 1 \times 3 + 1 \times 4 + 1 \times 9) / 5 = 3,6$$

obwohl der eigentliche Mittelwert $(2 \times 1 + 1 \times 3 + 1 \times 4) / 4 = 2,25$ ist.

Im Syntax definieren Sie Missing Values auf folgender Weise:

Missing Values AB7 (9) .

5.3 Datenmodifikation

Manchmal ist es notwendig, neue Variablen anzulegen oder existierende Variablen abzuändern. Für solche Prozeduren verfügt SPSS über eine Vielzahl von Kommandos.

5.3.1 Umkodieren (Recode)

Das Ziel des Kommandos „Recode“ ist das Ersetzen von bestimmten Werten auf schon existierenden Variablen durch neue Werten oder das Zusammenrechnen von bestimmten Kategorien einer schon bestehenden Variable.

Das „Recode“ Kommando entspricht in der Prozedur „Transformieren“ der Option „Umkodieren in dieselben Variablen“.

Ein Beispiel:

```
Recode V2 (5=10) (4=20) (3=30) (2=40) (1=50) .
```

```
Recode V2 (10=1) (20=2) (30=3) (40=4) (50=5) .
```

5.3.2 Variablen Modifizieren (If)

Mit dem “IF” Kommando können neue Variablen generiert oder bestehende Variablen modifiziert werden. Diese Variablen müssen bestimmte Bedingungen erfüllen. Das If-Kommando entspricht in der Prozedur „Transformieren“ die Option „Umkodieren in andere Variablen“.

Ein Beispiel:

```
If (Bildung >= 1 and Bildung <= 5) RBildung=1.
```

```
If (Bildung > 5 and Bildung <= 10) RBildung=2.
```

Die relationale und logische Operatoren, die bei diesem Kommando angewendet werden dürfen, finden Sie im Syntax-Guide, in der Online-Hilfe und im Menü „Transformieren“ unter der Option „Umkodieren in andere Variablen“.

5.3.3 Variablen einfügen oder löschen

Das Einfügen einer neuen Variable in SPSS ist sehr leicht. Sie klicken im Daten-Editorfenster (Blatt „Datenansicht“) auf der Variable, vor der Sie eine neue Variable einfügen möchten. Daraufhin klicken Sie auf der rechten Maustaste und wählen die Option „Variable einfügen“. Wenn man eine Variable löschen möchte, markiert man die Variable und drückt auf der Taste „Entfernen“.

5.3.4 Berechnen neuer Variable (Compute)

Mit „Compute“ erzeugt man eine neue Variable oder modifiziert man eine bestehende Variable. Das Compute-Kommando erzeugt eine Variable, die für jeden Respondent aus einer arithmetischen oder logischen Modifikation von schon vorhandenen Variablen oder Konstanten gebildet wird. Im Menü finden Sie die Prozedur unter „Transformieren“ (die Option „Berechnen“).

Im Syntax sieht der Befehl z.B. aus wie folgt:

```
Compute VarNeu = V1 + V2
```

5.3.5 Filter setzen (Select if)

Manchmal möchte man Analysen nur auf einer Teilgruppe der Respondenten ausführen, wie nur auf Männer oder nur auf Führungskräfte. Dafür verwendet man den Befehl „Select If“. Im Menü finden Sie diese Option unter „Daten“ und dann „Fälle Auswählen“. Auch treffen Sie dort die Wahl, ob man die nicht ausgewählte Cases nur herausfiltert oder aus dem Datensatz herauslöscht. Der Filter bleibt aktiv, bis man ihm in der Datenansicht herauslöscht oder wenn man über das Menü alle Fälle selektiert.

Wenn Sie über den Syntax arbeiten, empfiehlt es sich sehr, vor dem Select If –Kommando das Kommando „Temporary“ aufzunehmen. Wenn man diesen Befehl verwendet, wird der Filter nur vor der unmittelbar darauffolgenden Prozedur angewendet. Wenn man diesen Befehl nicht aufnimmt, löscht SPSS die nicht selektierten Cases aus dem Datensatz.

Dieses Beispiel illustriert das oben Beschriebene:

```
data list / v1 1 v2 2 v3 3 v4 4-5.  
begin data  
12345  
24352  
34532  
45432  
55454  
62345  
71234  
end data.  
temporary.  
select if v1=1.  
fre var v2.  
fre var v2.  
select if v1=1.  
fre var v2.  
fre var v2.
```

Hier werden in der Häufigkeitsverteilung von V2 nur die Respondenten aufgelistet, die die Bedingung „V1=1“ erfüllen.

```
temporary.  
select if v1=1.  
fre var v2.
```

Da das Temporary über den überstehenden Prozedur stand, werden hier alle Respondenten aufgelistet.

```
fre var v2.
```

Auch hier werden in der Häufigkeitsverteilung von V2 nur die Respondenten aufgelistet, die die Bedingung „V1=1“ erfüllen.

```
select if v1=1.  
fre var v2.
```

Kein Temporary. Jetzt würde man also erwarten, dass der Filter stehen bleibt. Das ist auch so. Nur, SPSS hat alle andere Cases herausgelöscht, und man hat nicht die Möglichkeit dies rückgängig zu machen! Hier hilft es nur, den Datensatz neu zu laden.

```
fre var v2.
```

6 Die Datenanalyse

Der Abschnitt Datenanalyse ist in diesem Kurs mit Absicht beschränkt gehalten, damit er auch für Studenten mit wenigen statistischen Vorkenntnissen verständlich ist. Ziel dieses Abschnitts ist vor Allem, Sie mit der Menüführung und die einfachste Analysemethoden vertraut zu machen. Wenn Sie das Prinzip der Datenanalyse mittels SPSS verstanden haben, werden Sie auch in der Lage sein, eventuell von Ihnen bekannten weiterführenden Analysen durchzuführen. Für mehrere Informationen über den verschiedenen Analysemethoden empfehle ich das Buch „Sozialwissenschaftliche Datenanalyse“ von Engel Möhring und Troitzsch, dass man sich unter [„http://www.uni-koblenz.de/~kgt/SDA.html“](http://www.uni-koblenz.de/~kgt/SDA.html) als pdf-Datei downloaden kann.

6.1 Univariate Datenanalyse

6.1.1 Häufigkeitsverteilungen

Die Prozedur „Häufigkeitsverteilungen“ gibt die absolute Häufigkeitsverteilungen, Prozentsätze und kumulierte Prozentsätze für jede Variable einzeln aus. Zusätzlich haben Sie die Möglichkeit, allerhand statistische Maße mit ausgeben zu lassen (Mittelwert, Standarddeviation). Auch das Erstellen von Graphiken gehört zu den Möglichkeiten. Im Menü finden Sie die Häufigkeitsverteilungen unter „Analysieren“, „Deskriptive Statistiken“ „Häufigkeiten“. Die Optionen für Statistiken und Diagramme finden Sie unten im Fenster.

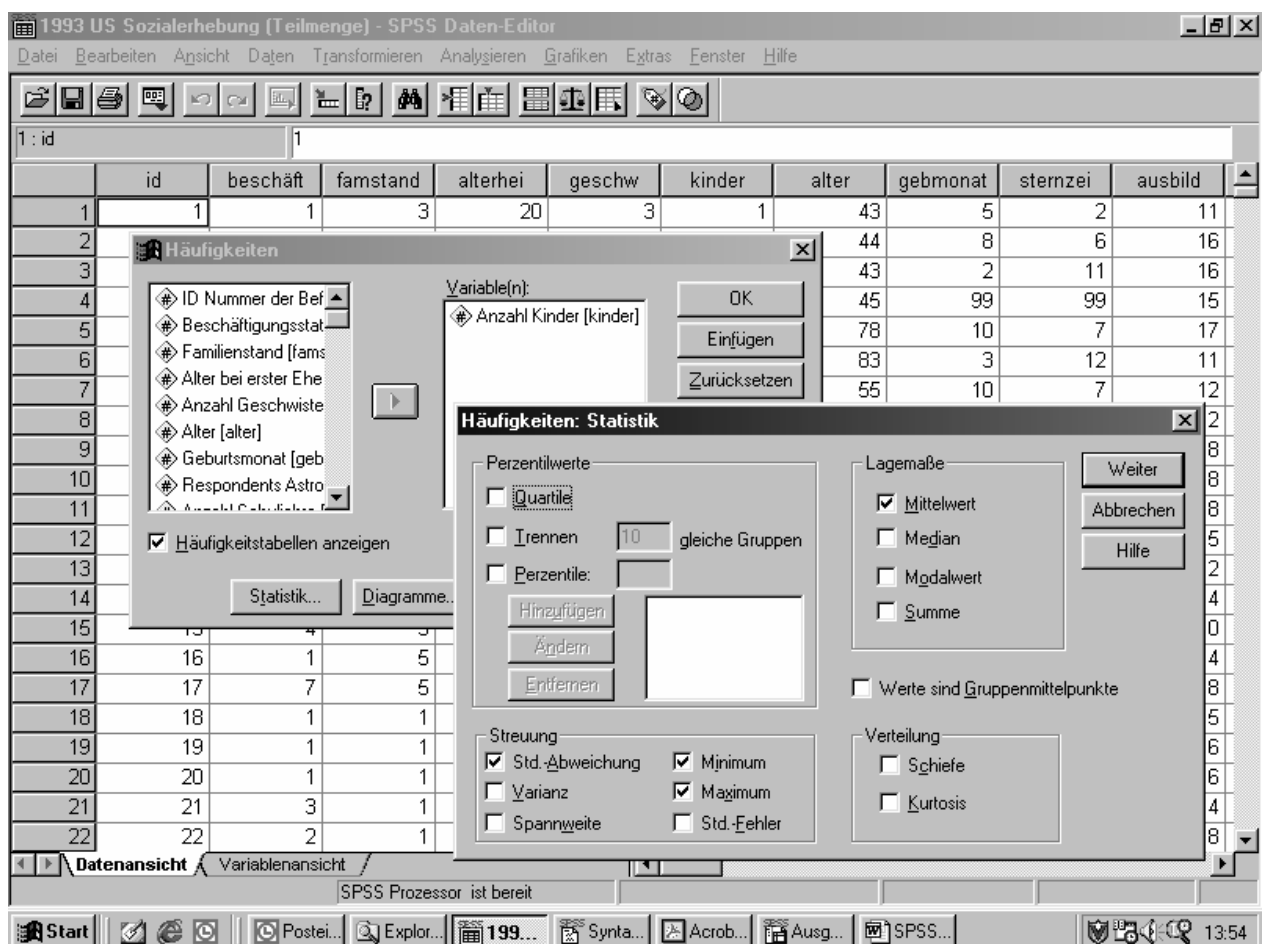
Die Ausgabe folgt in Tabellenform. Die Tabellen sind einfach zu bearbeiten und können direkt in MS-Office-Anwendungen hineinzukopiert werden. Auch die Graphiken lassen sich ohne Probleme kopieren und in Word einfügen.

Beispiel: Ausgabe der Prozedur Frequencies

Familienstand

		Häufigkeit	Prozent	Gültige Prozepte	Kumulierte Prozepte
Gültig	verheiratet	795	53,0	53,0	53,0
	verwitwet	165	11,0	11,0	64,0
	geschieden	213	14,2	14,2	78,3
	getrennt lebend	40	2,7	2,7	80,9
	ledig	286	19,1	19,1	100,0
	Gesamt	1499	99,9	100,0	
Fehlend	keine Angabe	1	,1		
Gesamt		1500	100,0		

Abbildung 18: Menüstruktur Häufigkeitsverteilungen



In der Syntax-Sprache sieht das Kommando aus wie folgt:

```
FREQUENCIES
  VARIABLES=famstand
  /ORDER= ANALYSIS .
```

6.1.1.1 Häufigkeitsverteilungen und Datenbereinigung

Im Prozess der Datenbereinigung sind die Häufigkeitsverteilungen sehr wichtig. An den Tabellen sieht man, ob es irgendwo Bereichsüberschreitungen gibt (z.B. ein Wert „6“ auf einer Fünferskala) und ob es bei bestimmten Variablen ungewöhnlich viele fehlende Werte gibt. Auch eignen sich Häufigkeitsverteilungen für einfache Plausibilitätschecks. Wenn man z.B. weiß, dass 150 Fragebögen an Männer geschickt wurden, können keine 170 männliche Respondenten im Datensatz anwesend sein. Wenn dies doch der Fall ist, steht man vor einer ernsthaften Herausforderung.

6.1.1.2 Umkodieren von offenen Fragen und Kategorien „Sonstige“

In einer chronologischen Reihenfolge der Datenverarbeitung hätte dieser Abschnitt weiter oben stehen müssen, da das Umkodieren der offenen Fragen direkt nach Abschluss der Dateneingabe durchgeführt wird. Für dieses Umkodieren braucht man allerdings bestimmte Fähigkeiten, die unmittelbar vor diesem Abschnitt erst besprochen wurden.

1. Man selektiert alle Respondenten, die auf einer bestimmten Frage mit „Sonstige“ geantwortet haben (z.B: de7=9).
2. Man lässt sich für diese Gruppe die Häufigkeitsverteilung der Variable geben, in der die Textantworten (also die Spezifizierung der Kategorie „Sonstige“) erfasst wurde (z.B: de7text). In der Häufigkeitsverteilung kann man feststellen, wie oft welche Antworten in der Kategorie „Sonstige“ genannt wurden. Achten Sie darauf, dass bei unterschiedlichen Schreibweisen unterschiedliche Kategorien ausgegeben werden.
3. Stellen Sie fest, für welche Antworten es sich lohnt, separate Kategorien anzulegen. Viele statistische Prozeduren setzen eine Minimumanzahl an Cases oder eine minimale Zellfüllung voraus. Es macht also keinen Sinn, für Einzelnennungen neue Kategorien anzulegen.
4. Selektieren Sie alle Cases (löschen Sie den Filter). Legen Sie eine neue, an der ursprünglichen Variable identische, Variable an (Compute v7rec=v7), und kodieren Sie die Kategorie „Sonstige“, wenn die Variable durch die neuen Kategorien zweistellig wird, um auf „99“ (Recode v7rec (9=99)). Nicht vergessen die Missing Values neu zu definieren!
5. Wir gehen jetzt mal davon aus, das die Originalvariable V7 6 gültige Kategorien hat. Selektieren Sie eine Kategorie der Option „Sonstige“ die sie umkodieren möchten:

```

If v7text="Teletubbies" v7rec=7
If v7text="He-Man" v7rec=8
If v7text="Biene Maja" v7rec=9
If v7text="Pokemon" v7rec=10

```

SPSS unterscheidet hier zwischen Groß- und Kleinschreibung, wenn irgendwo in der Matrix „teletubbies“ oder „Teletubbys“ steht, ordnet SPSS diese Antworten nicht automatisch zu.

Da die Kategorien der Variable v7text nicht aus Ziffern sondern aus Text bestehen, muss der Name in Gänzefüßchen stehen.

6. Selektieren Sie erneut v7=9, und kreuztabellieren Sie v7text und v7rec miteinander. In die Kreuztabellen können Sie sehen, ob sie alle Schreibweisen berücksichtigt haben. Wenn nicht: Wiederholen Sie die gesamte Prozedur ab Schritt 5.
7. Vergessen Sie nicht, die neue Variable im Codebook aufzunehmen

6.1.2 Deskriptive Statistiken

Die Option „Deskriptive Statistiken“ finden Sie im Menü unter der gleichnamigen Beschreibung. Dieses Kommando listet bestimmte Merkmale der ausgewählten Variablen in Tabellenform untereinander auf. Die Prozedur ist in erster Linie gemeint für kontinuierliche Variablen (wie Alter oder Temperatur), aber man kann sie auch dazu verwenden, sich einen ersten Eindruck von Bereichen, Streuung etc. zu bilden. Zur Datenbereinigung ist diese Option weniger geeignet. Wenn man z.B. eine Variable hat, die 5 Kategorien hat (von sehr gut bis sehr schlecht), und den Code „9“ für Missing, wird man die Eingabefehler im Bereich 6 bis 8 nicht sehen.

Die Ausgabe der Prozedur sieht aus wie folgt:

Deskriptive Statistik	N	Minimum	Maximum	Mittelwert	Standardabweichung
Verhütung bei Teenagern 14-16	974	1	4	2,34	1,07
Sexualkunde im Unterricht	984	1	2	1,15	,36
Eine Tracht Prügel hat noch keinem geschadet	997	1	4	2,11	,83
Einkommen des Befragten	994	1	22	12,80	5,62
Größe der Ortschaft	757	1	10	4,64	2,84
politische Einstellung	1491	0	7	2,91	2,06
Gültige Werte (Listenweise)	303				

Die dazugehörige Syntax lautet:

```
DESCRIPTIVES  
  VARIABLES=pille sexualkd prügel einkbefr ort partei  
  /STATISTICS=MEAN STDDEV MIN MAX .
```

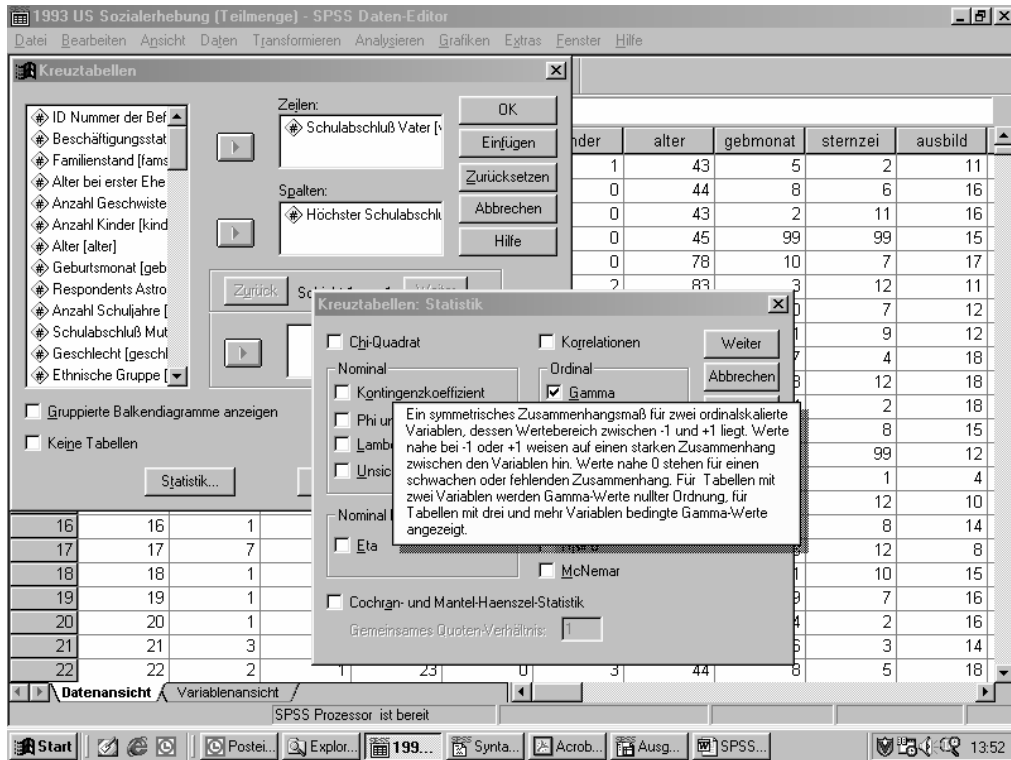
6.2 Bivariate Datenanalyse

6.2.1 Kreuztabellen

In einer Kreuztabelle werden zwei Variablen „gekreuzt“. Im Menü finden Sie die Prozedur unter: „Deskriptive Statistiken“ und „Kreuztabellen“. Es gibt zwei Kriterien dafür, wie die Kreuztabelle angeordnet werden sollte. Normalerweise wird die unabhängige Variable (z.B. Bildungsgrad) in den Zeilen und die abhängige Variable (z.B. Gehalt) in den Spalten dargestellt. Wenn die Variablen nicht voneinander abhängig sind (wie Wohnsitz und Geschlecht) wird die Variable mit den wenigsten Kategorien in den Spalten aufgelistet. Wenn die abhängige Variable sehr viele Kategorien hat, weicht man, um die Leserlichkeit der Tabelle zu verbessern, von der ersten Regel ab und ordnet die abhängige Variable in den Zeilen an.

Wenn gewünscht kann man sich zu der Kreuztabelle Prozentsätze, Erwartungswerte und Assoziationsmaße ausgeben lassen. Bei den Assoziationsmaßen müssen sie allerdings das Messniveau der Variablen berücksichtigen. Informationen über die verschiedenen Tests bekommen Sie, wenn Sie im Menü den Mauszeiger auf dem Maß setzen und auf der rechten Maustaste klicken.

Abbildung 19: Das Menü „Kreuztabelle“



Im Beispiel wird ein Zusammenhang zwischen der Schulbildung der Vater und der Schulbildung des Respondents vermutet. Diese beide Variablen (beide von ordinalem Messniveau) werden in der Kreuztabelle gekreuzt.

Schulabschluß Vater * Höchster Schulabschluß Kreuztabelle

Anzahl		Höchster Schulabschluß					Gesamt
		Hauptschule	Hauptschule (Quali.)	Mittlere Reife	Abitur	Universität/FH	
Schulabschluß Vater	Hauptschule	156	308	29	45	25	563
	Hauptschule (Quali.)	27	248	34	79	37	425
	Mittlere Reife	1	11	2	8	3	25
	Abitur	6	43	7	47	18	121
	Universität/FH	3	22	3	27	16	71
Gesamt		193	632	75	206	99	1205

Da man die Daten nicht nur darstellen möchte, sondern auch wissen möchte ob die Verteilung in der Kreuztabelle zufällig ist oder nicht, kann man sich in diesem Fall den Assoziationsmaß „Gamma“ mit ausgeben lassen. Dieses Maß setzt voraus, dass beide Variablen ein ordinales Messniveau haben.

Symmetrische Maße

		Wert Asymptotischer Standardfehler	Näherungs- weises T	Näherungs- weise Signifikanz
Ordinal- bzgl. Ordinalmaß	Gamma	,511	,031	14,648
Anzahl der gültigen Fälle		1205		,000

a Die Null-Hyphothese wird nicht angenommen.

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

Wie Sie sehen, ist Gamma signifikant ($p < 0,05$). Die Null-Hypothese (der höchste Schulabschluss des Kindes ist unabhängig von der Schulabschluss des Vaters) wird nicht angenommen.

Im Syntax sieht die Prozedur aus wie folgt:

```
GET  
  FILE='C:\Programme\SPSS\1993 US Sozialerhebung (Teilmenge).sav'.  
CROSSTABS  
  /TABLES=vaterab BY abschluss  
  /FORMAT= AVALUE TABLES  
  /STATISTIC=GAMMA  
  /CELLS= COUNT .
```

6.2.1.1 Plausibilitätschecks in der Datenbereinigung

Auch in der Datenbereinigung braucht man oft Kreuztabellen. Vor allem bei Befragungen, bei denen die Population ganz genau bekannt ist. Ich werde dies anhand eines Beispiels illustrieren. Man macht als Unternehmensberater eine Mitarbeiterbefragung in einem großen Unternehmen. Sie wissen, dass das Unternehmen 7 Vorstände hat. Sie wissen, dass diese Vorstände alle männlich sind und eine Betriebszugehörigkeit von mehr als 10 Jahren haben. Außerdem wissen Sie, zu welchen Abteilungen sie gehören. Nach der Dateneingabe stellt sich aber heraus, dass 18 Mitarbeiter sich als Vorstand eingestuft haben. Welche von dieser 18 nach aller Wahrscheinlichkeit die richtigen Vorstände sind, lässt sich herausfinden wenn sie die verschiedenen Variablen miteinander kreuztabellieren. Wenn sie sich zuerst die Cases „Position=Vorstand“ herausfiltern, und diese mit Geschlecht kreuztabellieren, sehen sie, dass z.B. 4 von diesen 18 „Vorstände“ Frauen sind. Nachdem sie diese als „Position=Missing“ umkodiert haben, sind noch 14 Vorstände übrig. Wenn Sie nun die Variable „Position“ mit der Variable „Betriebszugehörigkeit“ kreuztabellieren, stellen sie fest, dass 3 dieser Vorstände

weniger als 10 Jahre dabei sind. Nachdem man diese herausgefiltert hat, hat man noch 11 statt 7 Vorstände. Von diesen 11 sind 4 in Abteilungen, in denen es keinen Vorstand gibt. Nachdem sie bei diesen 4 die Position auf „Missing“ gesetzt haben, haben Sie in Ihrem Datensatz nur noch „wirkliche“ Vorstände. Dieses Beispiel verdeutlicht auch die Wichtigkeit der Datenbereinigung und ein hoher Datenqualität. Stellen Sie sich mal vor, Sie hätten als Unternehmensberater die Vorstandspräsentation abgehalten mit der Aussage:“ Also, die 11 Vorstandsmitglieder haben geantwortet, dass...“.

6.2.2 Korrelationen

In der Prozedur „Korrelationen“ (Unterpunkt: Bivariat) hat man die Möglichkeit, Produkt-Moment-Korrelationen (Pearson) oder nonparametrische Korrelationen (Spearman) zu rechnen.

Ein Korrelationskoeffizient ist ein Maß für den Zusammenhang zwischen zwei Variablen. Typische Fragestellungen, die anhand von Korrelationsanalyse beantwortet werden können, sind z.B: „Gibt es einen Zusammenhang zwischen Körpergröße und Körpergewicht?“ Oder: „Gibt es einen Zusammenhang zwischen Alter und Einkommen?“

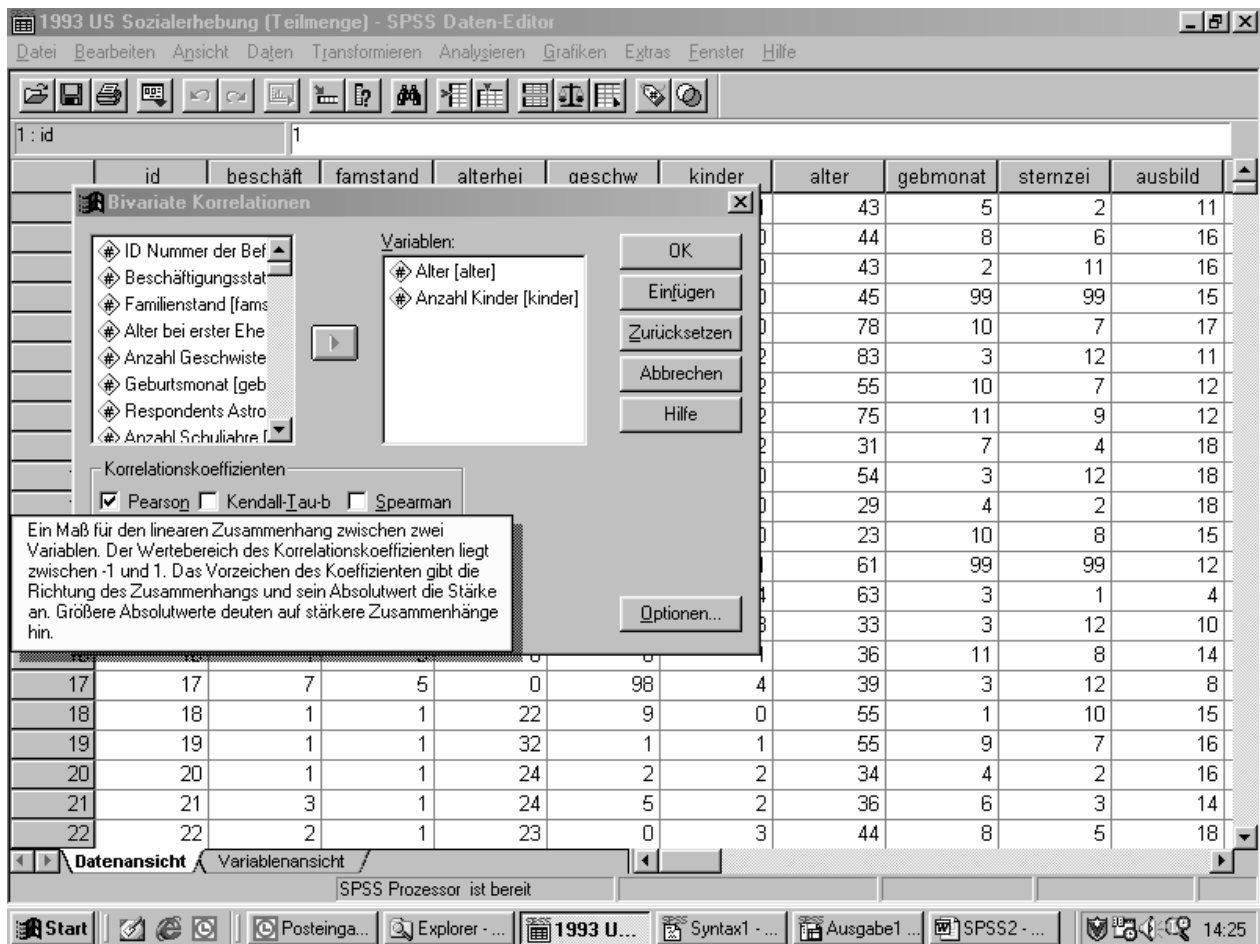
Es gibt verschiedene Arten von Korrelationen, die Pearson-Korrelation für metrischen Daten und die Spearman-Korrelation für ordinale Daten. Voraussetzungen zur Berechnung der Korrelationskoeffizient nach Pearson sind:

- Die Variablen, zwischen denen die Korrelation berechnet wird, sind mindestens von Intervall-Datenniveau
- Zwischen den Variablen gibt es einen linearen Zusammenhang
- Jedes Variablenpaar ist bivariat normalverteilt.

Falls die Daten nicht normalverteilt sind, oder wenn die Daten ein ordinales Datenniveau haben, rechnet man die Rangkorrelationskoeffizient nach Spearman, mit der die Beziehungen zwischen Rangordnungen gemessen werden.

Korrelationskoeffizienten können Werte von -1 bis $+1$ annehmen. Bei $+1$ gibt es einen perfekt positiven linearen Zusammenhang, bei -1 einen perfekt negativen. Bei dem Wert „0“ gibt es keinen Zusammenhang.

Abbildung 20: Menüpunkt „Korrelationen“



Im Beispiel wird die Korrelation berechnet zwischen den Variablen „Alter“ und „Anzahl der Kinder“. Diese Variable haben als Messniveau „Ratio“ und daher wird die Pearson-Korrelation berechnet.

Korrelationen

		Alter	Anzahl Kinder
Alter	Korrelation nach	1,000	,404
	Pearson		
	Signifikanz (2-seitig)	,	,000
	N	1495	1491
Anzahl Kinder	Korrelation nach	,404	1,000
	Pearson		
	Signifikanz (2-seitig)	,000	,
	N	1491	1495

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

Es gibt also einen signifikanten Zusammenhang zwischen dem Alter der Respondenten und der Anzahl der Kinder, die sie haben.

Im Syntax sieht die Prozedur „Correlations“ aus wie folgt:

Pearson

```
CORRELATIONS  
  /VARIABLES=alter kinder  
  /PRINT=TWOTAIL NOSIG  
  /MISSING=PAIRWISE .
```

Spearman

```
NONPAR CORR  
  /VARIABLES=abschluss pille  
  /PRINT=SPEARMAN TWOTAIL NOSIG  
  /MISSING=PAIRWISE .
```

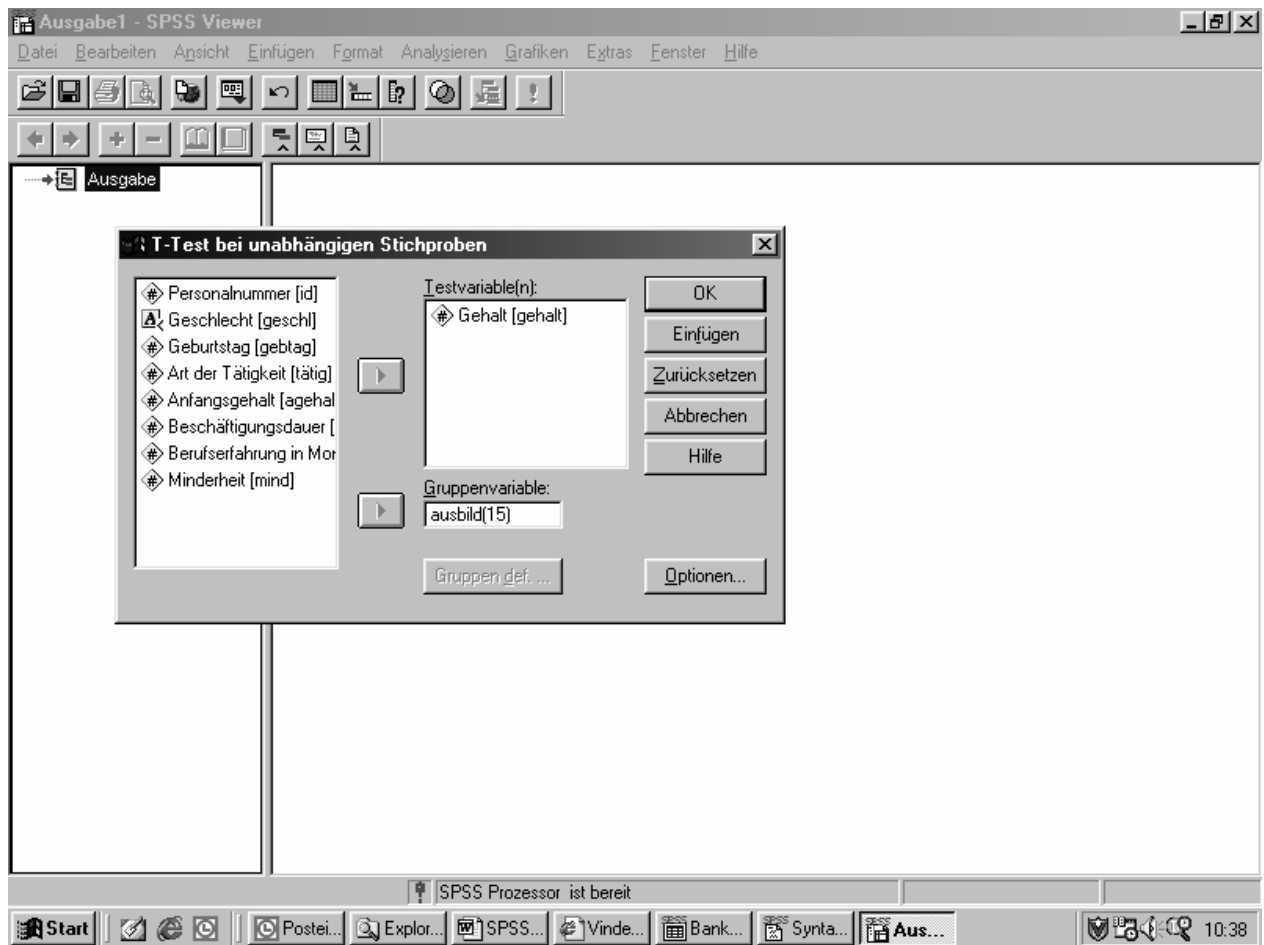
Tipp: Wenn Sie in der Syntax: „V2 with V3 V4 V5 V6“ schreiben, weil Sie an der Korrelation von V2 mit den anderen Variablen interessiert sind, bekommen Sie einen viel leserlicheren Output.

6.2.3 T-Test

Ein T-Test prüft, ob es einen signifikanten Unterschied in den Mittelwerten zweier Variablen gibt, die von Intervall-Niveau und normal verteilt sein müssen. Es gibt zwei Variationen, ein für unabhängige Stichproben (verschiedene Gruppen Respondenten mit Scores auf der gleichen Variable) und ein für abhängige Stichproben (die gleiche Gruppe von Respondenten aber verschiedene Variablen).

So möchte man z.B wissen, ob Personen, die viele Jahre zur Schule gegangen sind oder eine Ausbildung gemacht haben (15 Jahre oder länger), mehr verdienen als Personen, die wenige Jahre Unterricht absolviert haben (weniger als 15 Jahre). Dazu wähle ich im Menü die Prozedur „Mittelwerte vergleichen“ und den Unterpunkt „T-Test bei unabhängigen Stichproben“.

Abbildung 21: Menüpunkt „T-Test bei unabhängigen Stichproben“



Der Output dieser Prozedur besteht aus zwei Tabellen und sieht aus wie folgt:

Gruppenstatistiken

	Ausbil- dung (in Jahren)	N	Mittelwert	Standard- abweichun	Standard- fehler des Mittelwerte
Gehalt	>= 15	225	44.059,62	20.088,57	1.339,24
	< 15	249	25.708,67	5.578,52	353,52

Test bei unabhängigen Stichproben

		Levene-Test	T-Test für		df	Sig. (2-	95%	
		der	die	Untere			Obere	
	Varianzgleichheit	F	Signifikanz		T			seitig)
							der	
							Differenz	
Gehalt	Varianzen sind gleich	209,028	,000	13,838	472	,000	15.745,07	20.956,82
	Varianzen sind nicht gleich			13,249	255,186	,000	15.623,24	21.078,66

Aus der zweiten Tabelle sind aus Darstellungsgründen die Spalten „Mittlere Differenz“ und „Standardfehler der Differenz“ gelöscht. Wie man sieht, wird zuerst geprüft, ob die Varianzen der beiden Gruppen signifikant verschieden sind. Falls dies der Fall ist, interpretieren Sie die Signifikanz des T-Test-Wertes für unterschiedliche Varianzen. Wenn sich die Varianzen nicht signifikant voneinander unterscheiden, interpretiert man den T-Test-Wert für gleiche Varianzen.

Im Syntax sieht die Ansteuerung der oben vorgeführten Test aus wie folgt:

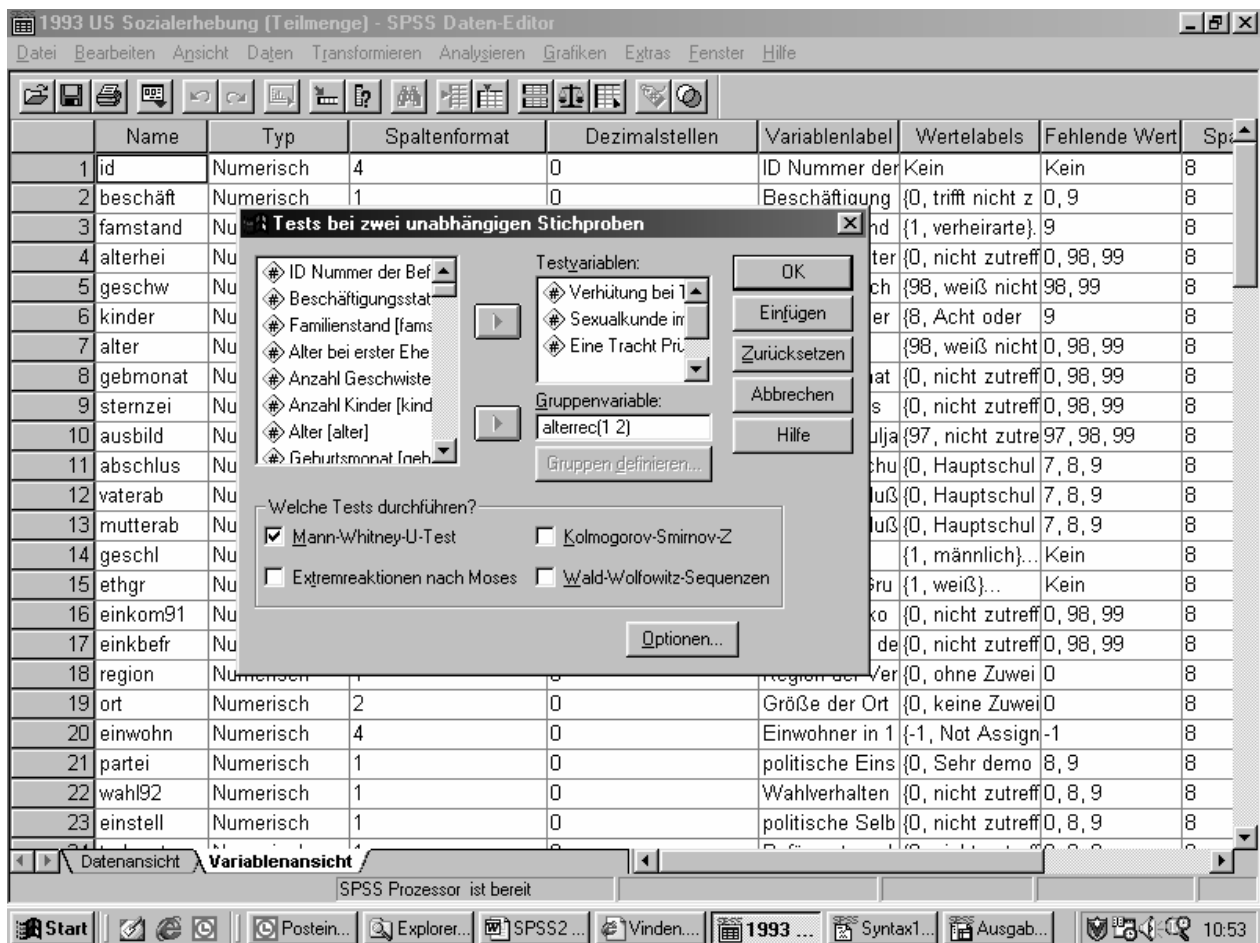
```
T-TEST
  GROUPS=alterrec(1 2)
  /MISSING=ANALYSIS
  /VARIABLES=pille sexualkd prügel
  /CRITERIA=CIN(.95) .
```

6.2.4 Mann-Whitney-U-Test

Wie oben genannt, ist der T-Test gemeint für Variablen auf Intervall-Niveau. Für Variablen auf ordinalem Niveau entspricht der Mann-Whitney-U-Test dem T-Test. Dieser Test stellt keine Bedingungen an die Verteilung der Variablen. Dieser Test ist äquivalent an dem Wilcoxon Rangsummentest, wird aber etwas anders berechnet. Der T-Test testet, ob die Mittelwerte von zwei Gruppen unterschiedlich sind, aber der Mittelwert ist für ordinale Daten kein gültiger Testwert. Der Mann-Whitney-U-Test bringt die Scores beider Gruppen in eine Reihenfolge und zählt sie auf. Darauf werden die beobachteten Summen der Rangscores untersucht und verglichen mit deren erwarteten Summen.

Wenn man z.B wissen möchte, ob junge (jünger als 43, Code 1) und ältere (ab 43, Code 2) Personen eine unterschiedliche Meinung zu bestimmten Items haben, wählt man im Menü die Prozedur „Nichtparametrische Tests“ und den Unterpunkt „Zwei unabhängige Stichproben“.

Abbildung 22: Menüpunkt „Zwei unabhängige Stichproben“



Auch der Output dieser Prozedur besteht aus zwei Tabellen und sieht aus wie folgt:

Ränge

	alter recode	N	Mittlerer Rang	Rangsumme
Verhütung bei Teenagern 14-16	1,00	509	436,62	222238,00
	2,00	465	543,20	252587,02
	Gesamt	974		
Sexualkunde im Unterricht	1,00	518	468,74	242807,00
	2,00	466	518,91	241812,98
	Gesamt	984		
Eine Tracht Prügel hat noch keinem geschadet	1,00	513	511,17	262228,50
	2,00	484	486,10	235274,50
	Gesamt	997		

Statistik für Test

	Verhütung bei Teenagern 14-16	Sexualkunde im Unterricht	Eine Tracht Prügel hat noch keinem geschadet
Mann-Whitney-U	92443,000	108386,000	117904,500
Wilcoxon-W	222238,000	242807,000	235274,500
Z	-6,119	-4,417	-1,489
Asymptotische Signifikanz (2-seitig)	,000	,000	,137

a Gruppenvariable: alter recode

In dieser Tabelle kann man sehen, dass es für die ersten beiden Variablen einen signifikanten Unterschied zwischen den beiden Gruppen gibt ($p < 0,05$), für die dritte Variable nicht.

Im Syntax sieht der Mann-Whitney-U-Test aus wie folgt:

```

NPAR TESTS
  /M-W= pille sexualkd prügel BY alterrec(1 2)
  /MISSING ANALYSIS.

```

7 Literaturverzeichnis

Backhaus, K., Erichson, B., Plinke W., Weiber, R. Multivariate Analysemethoden - Eine anwendungsorientierte Einführung, Berlin, 1996

Bühl, A., Zöfel, P., SPSS – Methoden für die Markt-und Meinungsforschung, München, 2000

Engel, A., Möhring, M., & Troitzsch, K.G., Sozialwissenschaftliche Datenanalyse, Mannheim, 1995

Fernuniversität Hagen, Einführung in SPSS/Win, Hagen, 1999

Rodeghier, M., Marktforschung mit SPSS - Analyse, Datenerhebung und Auswertung, Bonn, 1997

SPSS, SPSS® Base 10.0 Benutzerhandbuch, München, 1999

Voß W., Praktische Statistik mit SPSS, München, 2000