

EINHEIT 1

Überblick über SPSS für Windows

Inhalt:

1 FRAGEBOGEN UND KODIERUNG	3
2 DEFINIEREN VON VARIABLEN UND ZUORDNEN VON BESCHREIBENDEN NAMEN	5
3 ERFASSEN BZW. EINGABE DER KODIERTEN DATEN, DATENKONTROLLE	8
DATENKONTROLLE	9
ABSPEICHERN DER DATEN	10
4 TRANSFORMIEREN DER DATEN BZW. ERZEUGEN VON NEUEN VARIABLEN	11
BERECHNEN NEUER VARIABLEN	11
UMKODIEREN VON VARIABLEN	14
5 AUSWÄHLEN VON FÄLLEN	17
6 GEWICHTEN VON DATEN	18
7 ANALYSIEREN VON DATEN	19
8 BEARBEITUNG UND VERWENDEN DES OUTPUTS	23
MODIFIKATION VON TABELLEN	23
MODIFIKATION VON GRAFIKEN	26
ÜBERNAHME VON TABELLEN UND GRAFIKEN IN TEXTVERARBEITUNGS- BZW. PRÄSENTATIONSSOFTWARE BZW. EXPORTIEREN DES OUTPUTS	29
9 EINLESEN EXTERNER DATEN	30
IMPORT VON EXCEL DATEIEN:	30
IMPORT VON TEXT DATEIEN:	30
10 VERWENDEN DER PROGRAMMSPRACHE („SYNTAX“)	33
11 HILFE ZU EINZELNEN PROZEDUREN, TUTORIALS UND LINKS	36

Überblick über SPSS für Windows

In diesem Kapitel erhalten Sie einen kurzen Überblick über die Bedienung von SPSS für Windows sowie über typische Arbeitsschritte.

Typische Arbeitsschritte

Sie führen bei einer statistischen Datenanalyse in der Regel die folgenden Schritte durch:

1. Erstellen eines **Kodierungsschemas** – Vergabe von Variablennamen und ev. Zahlen für Kategorien
2. **Definieren von Variablen** und **Zuordnen von beschreibenden Namen** (Etiketten, Umschreibungen) für Variablen (*variable labels*) und Daten-Werte (*value labels*), um die spätere Text- und Grafik-Ausgabe aussagekräftiger zu gestalten. **Speichern der Definitionen** in eine SPSS Datendatei.
3. **Erfassen** der kodierten Daten, **Kontrollieren** auf Eingabefehler.
4. (*Optional*) **Transformieren** der Daten in eine zweckmäßigere Form bzw. **Erzeugen** von neuen Variablen.
5. (*Optional*) **Auswählen** von Fällen für die folgende Analysen
6. (*Optional*) **Gewichten** von Daten (notwendig bei der Analyse von kategorialen Variablen die in Tabellenform und nicht als Einzelbeobachtungen gegeben sind)
7. **Analysieren** der Daten:
 1. **Grafisches Darstellen (Visualisieren)** der Daten zur grafischen Beschreibung und ev. Vorbereitung von statistischen Analysen
 2. **Tabellarisches Darstellen** des Datenmaterials und **Berechnen von Kennzahlen** zur numerischen Beschreibung der Daten und ev. Vorbereitung von statistischen Analysen
 3. **eigentliche Analyse** (mit Verfahren der induktiven oder Inferenz-Statistik)
8. **Nachbearbeitung des Outputs** und **Kopieren in ein Textverarbeitungsprogramm** zur Anfertigung eines Berichts.

Im folgenden soll auf jeden dieser Schritte überblicksmäßig eingegangen werden um die grundlegende Arbeitsweise mit SPSS zu erläutern und die wichtigsten Eigenschaften und Bedienungsweisen vorzustellen.

Darüber hinaus sollen noch weitere Aspekte kurz behandelt werden. Dies betrifft

9. **Einlesen externer**, bereits vorhandener **Daten**, die **nicht im SPSS Format** gespeichert sind, wie z.B. Excel-Dateien oder Text-Dateien.
10. Verwenden der **Programmsprache („Syntax“)**, die hinter der grafischen Benutzeroberfläche läuft und deren Verwendung das Leben manchmal sehr erleichtern kann.
11. **Bedienen der Hilfe** zu einzelnen Prozeduren, Hinweise auf **Tutorials** und **Links**

1 Fragebogen und Kodierung

Als Beispiel für dieses Kapitel soll der folgende Fragebogen dienen.

ID : _____							
Fragebogen: (bitte ankreuzen)							
mein Geschlecht? <table border="1" style="display: inline-table; border-collapse: collapse;"> <tr> <td style="padding: 2px;">weiblich</td> <td style="padding: 2px;">männlich</td> </tr> </table>	weiblich	männlich					
weiblich	männlich						
mein liebstes Alter wäre ?							
<table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="padding: 2px;">egal</td> <td style="padding: 2px;">jünger</td> <td style="padding: 2px;">älter</td> <td style="padding: 2px;">genau so wie jetzt es ist</td> <td style="padding: 2px;">gleich wie jetzt oder jünger</td> <td style="padding: 2px;">gleich wie jetzt oder älter</td> </tr> </table>	egal	jünger	älter	genau so wie jetzt es ist	gleich wie jetzt oder jünger	gleich wie jetzt oder älter	
egal	jünger	älter	genau so wie jetzt es ist	gleich wie jetzt oder jünger	gleich wie jetzt oder älter		
meine Größe ? _____ cm							
mein Alter ? _____ Monate							
bei einem ersten Date bevorzuge ich ?							
<table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="padding: 2px;">romantischer Spaziergang</td> <td style="padding: 2px;">Kino</td> <td style="padding: 2px;">Abendessen</td> <td style="padding: 2px;">Disco</td> <td style="padding: 2px;">spontane Entscheidung</td> <td style="padding: 2px;">Video ausleihen</td> <td style="padding: 2px;">auf der Couch knuddeln</td> </tr> </table>	romantischer Spaziergang	Kino	Abendessen	Disco	spontane Entscheidung	Video ausleihen	auf der Couch knuddeln
romantischer Spaziergang	Kino	Abendessen	Disco	spontane Entscheidung	Video ausleihen	auf der Couch knuddeln	
die meisten Entscheidungen, die ich treffe, beruhen auf ?							
<table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="padding: 2px;">Logik</td> <td style="padding: 2px;">Intuition</td> <td style="padding: 2px;">Moral</td> <td style="padding: 2px;">Freunde</td> <td style="padding: 2px;">Entscheidung ... Was ist das ?</td> </tr> </table>	Logik	Intuition	Moral	Freunde	Entscheidung ... Was ist das ?		
Logik	Intuition	Moral	Freunde	Entscheidung ... Was ist das ?			
wenn ich an einem Projekt arbeite							
<table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="padding: 2px;">ich beginne sofort und bin frühzeitig fertig</td> <td style="padding: 2px;">ich schiebe es vor mir her, dann mache ich eine Nacht durch</td> <td style="padding: 2px;">ich kaufe mir einen Hund und sage, er hat die Arbeit gefressen</td> </tr> </table>	ich beginne sofort und bin frühzeitig fertig	ich schiebe es vor mir her, dann mache ich eine Nacht durch	ich kaufe mir einen Hund und sage, er hat die Arbeit gefressen				
ich beginne sofort und bin frühzeitig fertig	ich schiebe es vor mir her, dann mache ich eine Nacht durch	ich kaufe mir einen Hund und sage, er hat die Arbeit gefressen					
ich würde mich beschreiben als:							
bei Terminen oder Verabredungen bin ich nie zu spät							
<table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="padding: 2px;">trifft sehr zu</td> <td style="padding: 2px;">trifft eher zu</td> <td style="padding: 2px;">weder/ noch</td> <td style="padding: 2px;">trifft eher nicht zu</td> <td style="padding: 2px;">trifft überhaupt nicht zu</td> </tr> </table>	trifft sehr zu	trifft eher zu	weder/ noch	trifft eher nicht zu	trifft überhaupt nicht zu		
trifft sehr zu	trifft eher zu	weder/ noch	trifft eher nicht zu	trifft überhaupt nicht zu			
ich kann gut zuhören, lasse andere zuerst sprechen							
<table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="padding: 2px;">trifft sehr zu</td> <td style="padding: 2px;">trifft eher zu</td> <td style="padding: 2px;">weder/ noch</td> <td style="padding: 2px;">trifft eher nicht zu</td> <td style="padding: 2px;">trifft überhaupt nicht zu</td> </tr> </table>	trifft sehr zu	trifft eher zu	weder/ noch	trifft eher nicht zu	trifft überhaupt nicht zu		
trifft sehr zu	trifft eher zu	weder/ noch	trifft eher nicht zu	trifft überhaupt nicht zu			
ich versuche oft mehrere Dinge gleichzeitig zu tun – " und was kommt als nächstes ?"							
<table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="padding: 2px;">trifft sehr zu</td> <td style="padding: 2px;">trifft eher zu</td> <td style="padding: 2px;">weder/ noch</td> <td style="padding: 2px;">trifft eher nicht zu</td> <td style="padding: 2px;">trifft überhaupt nicht zu</td> </tr> </table>	trifft sehr zu	trifft eher zu	weder/ noch	trifft eher nicht zu	trifft überhaupt nicht zu		
trifft sehr zu	trifft eher zu	weder/ noch	trifft eher nicht zu	trifft überhaupt nicht zu			
ich kann gelassen warten							
<table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="padding: 2px;">trifft sehr zu</td> <td style="padding: 2px;">trifft eher zu</td> <td style="padding: 2px;">weder/ noch</td> <td style="padding: 2px;">trifft eher nicht zu</td> <td style="padding: 2px;">trifft überhaupt nicht zu</td> </tr> </table>	trifft sehr zu	trifft eher zu	weder/ noch	trifft eher nicht zu	trifft überhaupt nicht zu		
trifft sehr zu	trifft eher zu	weder/ noch	trifft eher nicht zu	trifft überhaupt nicht zu			
es macht mir nichts aus, wenn Dinge nicht ganz fertig sind							
<table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="padding: 2px;">trifft sehr zu</td> <td style="padding: 2px;">trifft eher zu</td> <td style="padding: 2px;">weder/ noch</td> <td style="padding: 2px;">trifft eher nicht zu</td> <td style="padding: 2px;">trifft überhaupt nicht zu</td> </tr> </table>	trifft sehr zu	trifft eher zu	weder/ noch	trifft eher nicht zu	trifft überhaupt nicht zu		
trifft sehr zu	trifft eher zu	weder/ noch	trifft eher nicht zu	trifft überhaupt nicht zu			

In der Regel empfiehlt es sich, ein paar Vorbereitungen für die spätere Dateneingabe und Analyse zu treffen, bevor man überhaupt SPSS startet (bei kleinen, klar strukturierten Datensätzen mag das überflüssig sein).

Bei sozialwissenschaftlichen Untersuchungen wird man die Rohdaten meist in Form von ausgefüllten Fragebögen oder Protokollblättern vorliegen haben, die man dann in irgendeiner Form in den Computer bringen will. Bei der Dateneingabe ist es sehr nützlich, wenn man schon weiß, welche Namen man für die Variablen und welche Zahlen man für die Beobachtungen (Antwortkategorien) verwenden will. Bei Fragebögen sollte einem klar sein, dass jede Frage (auch Subfrage, manchmal sogar auch nur eine Antwortkategorie) eine Variable darstellt.

Die Umsetzung dieser Überlegungen nennt man ein **Kodierungsschema** entwickeln.

Am einfachsten ist es, sich einen leeren Fragebogen zu nehmen und **für jede Frage einen Kurznamen** zu definieren (den wir dann für SPSS benötigen) und dazuzuschreiben. Für kategoriale Variablen (also Fragen mit mehreren Antwortkategorien) schreibt man noch dazu, **welche Zahlen man verwenden will**. Der Fragebogen könnte dann etwa so aussehen:

ID : _____

Fragebogen: (bitte ankreuzen)

sex
mein Geschlecht?
weiblich männlich

mein liebstes Alter wäre ? **liebtalt**

egal 1	jünger 2	älter 3	genau so wie jetzt es ist 4	gleich wie jetzt oder jünger 5	gleich wie jetzt oder älter 6
---------------	-----------------	----------------	------------------------------------	---------------------------------------	--------------------------------------

meine Größe ? _____ cm **groesse**

mein Alter ? _____ Monate **monate**

bei einem ersten Date bevorzuge ich ? **date**

romantischer Spaziergang 1	Kino 2	Abendessen 3	Disco 4	spontane Entscheidung 5	Video ausleihen 6	auf der Couch knuddeln 7
-----------------------------------	---------------	---------------------	----------------	--------------------------------	--------------------------	---------------------------------

Diese **Kurznamen** dürfen aus maximal 64 Zeichen bestehen (ab SPSS Version 12, früher maximal 8). Das erste Zeichen muss ein Buchstabe oder ein @-Zeichen sein. Alle weiteren Zeichen im Namen können Buchstaben, Zahlen oder eines der fünf Sonderzeichen: _ . @ # \$ sein (. und _ nicht als letztes Zeichen verwenden). Leerzeichen dürfen nicht verwendet werden. Groß- und Kleinbuchstaben werden (ab Version 12) unterschieden. Schlüsselwörter, die nicht als Variablennamen verwendet werden dürfen, sind: ALL, NE, EQ, TO, LE, LT, BY, OR, GT, AND, NOT, GE und WITH.

Auch wenn man (jetzt) 64 Buchstaben und/oder Ziffern verwenden darf, sollte man sich auf möglichst kurze Namen beschränken, da dies die Übersichtlichkeit erhöht. Damit man später auch noch weiß, was die Variable eigentlich bedeutet und um die Ausgabe lesbar zu gestalten, vergibt man längere Beschreibungen (sogenannte **Variablenlabel**). Dazu später.

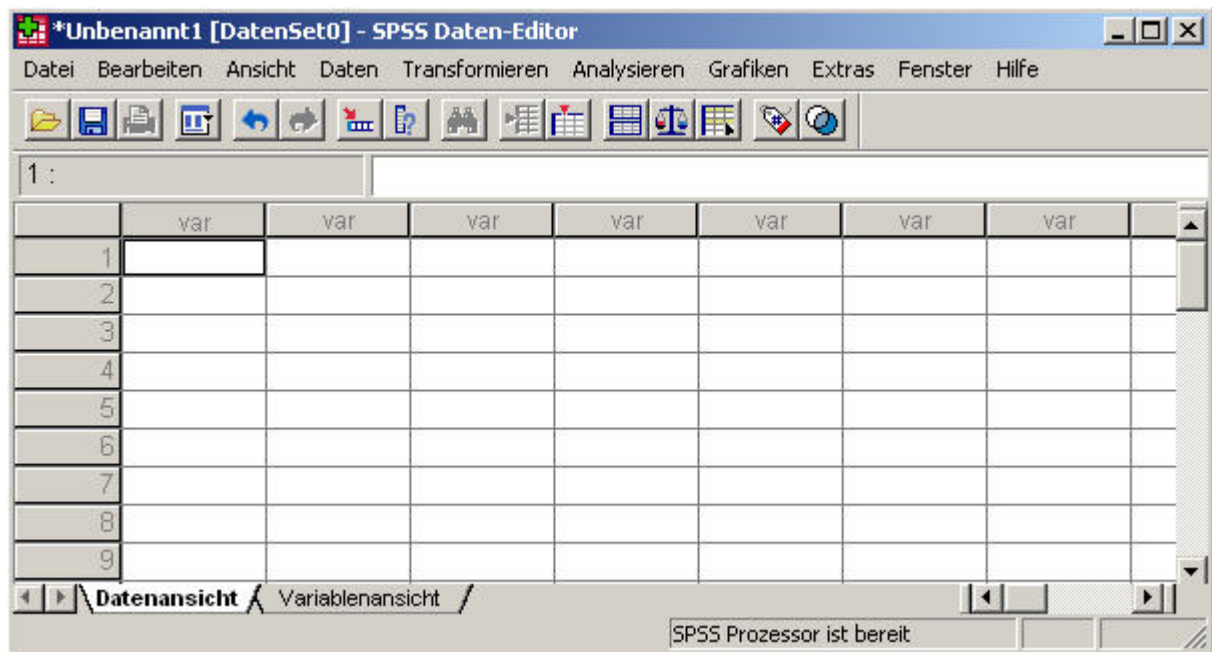
Welche **Zahlen (Codes)** man verwendet ist eigentlich egal (man könnte auch „Strings“ verwenden, wovon aber abzuraten ist), es ist aber sinnvoll eine natürliche Ordnung zu verwenden, also bei 0 oder 1 zu beginnen und dann fortlaufend zu nummerieren. Für metrische Variablen (wie hier z.B. „Alter“) braucht man natürlich keine Codes.



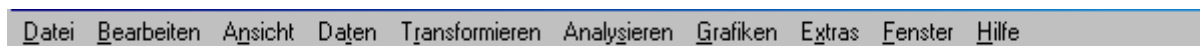
Man sollte keine Umgestaltungen vor der Dateneingabe vornehmen, da das die Dateneingabe nur erschwert, fehleranfällig ist und man im Nachhinein, wie wir noch sehen werden, in SPSS alles ganz leicht nach Belieben ändern kann. Also am besten alles so definieren und die Daten dann so eingeben wie der Fragebogen aussieht. (So ist z.B. die Anordnung bei der Variable „mein liebstes Alter“ nicht wirklich sinnvoll und sollte eigentlich geändert werden, man macht das aber besser später.)

2 Definieren von Variablen und Zuordnen von beschreibenden Namen

Nach diesen Vorbereitungen können wir die Daten in SPSS eingeben. Nach dem Aufruf von SPSS erscheint ein Fenster, in dem wir den Menüpunkt **Daten eingeben** und danach das **OK** Feldes anklicken. Man kommt dann zu folgendem Fenster:



Dieses Fenster, der **SPSS Daten-Editor**, besteht aus einem Arbeitsblatt zum Erstellen und Bearbeiten von SPSS-Datendateien (so ähnlich, wie es aus Tabellenkalkulationsprogrammen bekannt ist), sowie einer Menüleiste



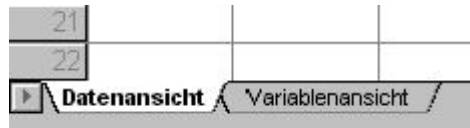
und einer Symbolleiste:



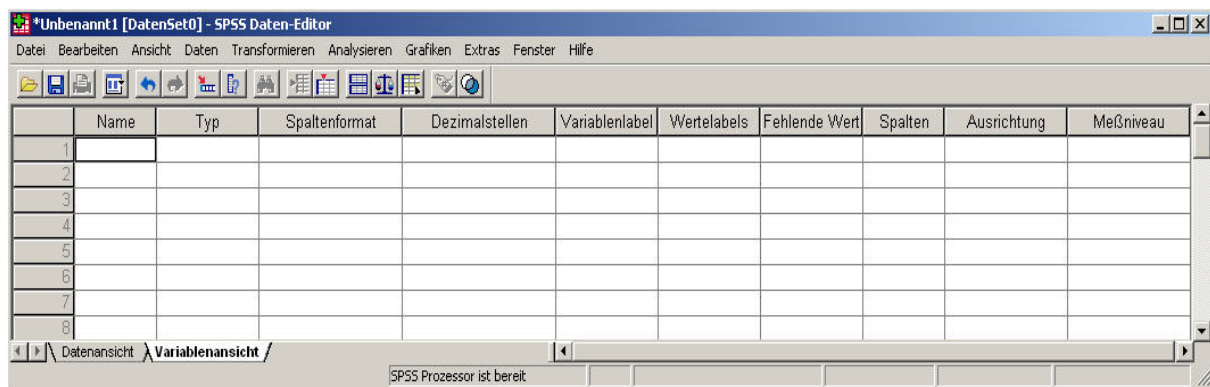
Wie in vielen Windows-Anwendungen üblich, erhält man eine Beschreibung des jeweiligen Symbols, wenn man mit der Maus – ohne zu klicken – auf das Symbol zeigt und kurz wartet.

Die für ein erstes kennen lernen von SPSS wichtigsten Punkte dieser beiden Leisten werden später noch genauer beschrieben.

Im Prinzip könnte man hier die Daten gleich eingeben, aber es empfiehlt sich zunächst die Struktur der Analysedaten zu definieren. Dies erfolgt in einem anderen Fenster. Das obige Bild zeigt die sogenannte **Datenansicht** des Daten-Editors. Der Daten-Editor enthält aber noch eine zweite Ansicht, nämlich die **Variablenansicht**. Zu dieser gelangt man durch entsprechendes Anklicken von **Variablenansicht** am unteren linken Rand der Datentabelle.



Es öffnet sich nun folgendes Fenster:



Hier gibt man zeilenweise für jede Variable die gewünschten Definitionen ein:

1. Name (der Variable)

Hier werden die (vorher im Codeblatt festgelegten) Kurzbezeichnungen eingegeben. Diese Kurzbezeichnungen (gemeinsam mit den Langbezeichnungen – Variablenlabels, siehe weiter unten) werden dann im weiteren benutzt, um SPSS mitzuteilen, welche Variablen für bestimmte Analyseschritte verwendet werden sollen.

Sobald man den Namen eingegeben und „Eingabe“ gedrückt hat (bzw. mit den Cursortasten in eine andere Zelle gewechselt ist) füllt sich der Rest der Zeile mit voreingestellten Werte auf. Diese kann man nun entsprechend ändern (für schnelle Analysen, bei denen es auch nicht auf besondere Form des Outputs ankommt genügen diese Default-Werte meistens).

Die weiteren Felder sind der Reihe nach:

2. - 4. Typ, Spaltenformat und Dezimalstellen

voreingestellt sind Dezimalzahlen („Numerisch“) mit 8 Stellen, hiervon 2 Dezimalstellen, das Komma wird in die 8 Stellen inkludiert. Nach Anklicken des Typs öffnet sich das Fenster **Variablentyp definieren**. Hier kann man verschiedene Formate auswählen (z.B. Numerisch, Zahl, Datum, Währung, Zeichenkette-„Strings“,...), außer in Spezialfällen sollte man aber bei **Numerisch** bleiben. Um die Daten lesbarer zu machen, kann man noch die Anzahl der Dezimalstellen ändern. Dies betrifft aber nur die Darstellung in der Datenansicht. Die interne Repräsentation wird dadurch nicht geändert.

5. Variablenlabels (Etikett für den Variablenname)

Oft reicht der Kurzname nicht aus, um eine Variable sinnvoll zu bezeichnen (so dass man später auch noch weiß, was die Variable eigentlich bedeutet hat). Hier kann man sich helfen,

indem man eine längere Beschreibung (einen sogenannten **Variablenlabel** bzw. *variable labels*) vergibt. (z.B. Liebstes Alter statt liebalt)

6. Wertelabels (Etiketten für einzelne Werte)

Ebenso können für kategoriale Variablen (d.h. solche, die nur wenige Kategorien - z.B. weiblich/ männlich, haben) sogenannte **Wertelabels** bzw. *value labels* vergeben werden (z.B. „weiblich“ für 1, „männlich“ für 2). Diese Labels werden dann im Output (meistens) statt der Zahlen oder zusätzlich zu diesen ausgegeben. Wertelabels und die entsprechenden Werte werden in einem kleinen Dialogfenster eingegeben und dann mit "Hinzufügen" übernommen (das Dialogfenster öffnet sich nach Anklicken des Wertelabelfeldes für die zu definierende Variable und dort ...).



Variablennamen und Variablenlabels in Listen von Dialogfeldern: In Listen von Dialogfeldern können Sie entweder Variablennamen oder Variablenlabels anzeigen lassen. Da man sich bei den Variablennamen oft auf Kurzbezeichnungen beschränkt, bieten Variablenlabels meist aussagekräftigere Informationen. Allerdings ist in manchen Situationen, wie bei der Auswahl von Variablen für bestimmte Analysen, Grafiken oder Datentransformationen, die Verwendung der langen Variablenbezeichnungen sehr unübersichtlich. Man kann wählen, ob in diesen Listen Variablennamen oder -beschriftungen angezeigt werden sollen. Die Auswahl erfolgt im Menü **Bearbeiten > Optionen > Allgemein**, wo man unter dem Punkt **Variablenlisten** entweder **Labels anzeigen** oder **Namen anzeigen** auswählt. Falls kein Variablenlabel definiert ist, wird der Variablenname angezeigt.

7. (benutzerdefinierte) Fehlende Werte

Definiert man nur, wenn man bestimmte Werte (z.B. -999) oder Bereiche als ungültig oder fehlend festlegen will. Sonst verwendet man besser systemdefinierte fehlende Werte, für numerische Variablen ist dies das Dezimalzeichen (, .). Für Zeichenketten existiert kein systemdefinierter fehlender Wert, da jede Zeichenkette (auch die leere Zeichenkette) prinzipiell einen gültigen Datenwert darstellen könnte.

8. - 9. Spaltenformat (Spaltenbreite und Spaltenausrichtung bei der Ausgabe)

z.B. 8 Zeichen, rechtsbündig.

10. Messniveau


Hier kann man zwischen **metrisch**, **ordinal** und **nominal** wählen. Trotz besseren Wissen ist es aus verschiedenen Gründen (deren Erläuterung hier zu weit führen würde) sinnvoll und einfacher nur **metrisch** und **nominal** zu verwenden. Im Zweifelsfall verwendet man überhaupt nur metrisch.

Für unseren Fragebogen könnten die ersten beiden Zeilen nach den Festlegungen in der Variablenansicht folgendermaßen aussehen:

	Name	Typ	Spaltenformat	Dezimalstellen	Variablenlabel	Wertelabels	Fehlende Wert	Spalten	Ausrichtung	Meßniveau
1	id	Numerisch	5	0		Kein	Kein	8	Rechts	Metrisch
2	sex	Numerisch	5	0	Geschlecht	{1, weiblich}...	Kein	8	Rechts	Nominal

Hat man alle Variablen definiert, kann man mit der Dateneingabe beginnen. Sicherheitshalber empfiehlt es sich aber vorher noch das bisherige abzuspeichern.

Speichern der Variablendefinitionen (und später auch der Daten)

Wir erzeugen dabei eine sogenannte SPSS-Datei, auf die wir später auch wieder zugreifen können. Dies geschieht, wie bei Windows-Programmen üblich, über den Menüpunkt **Datei** und dort bei **Speichern unter...**. Als Voreinstellung gibt SPSS als **Dateityp** die Erweiterung **.SAV** an. (Diese Erweiterung behält man am besten bei, da dadurch die Verbindung der Datei zu SPSS hergestellt ist und man am im Windows Explorer am Icon  erkennt, dass es sich bei der Datei um eine SPSS-Datendatei handelt.). Für unser Beispiel könnten wir den Namen **uebungsfragebogen** verwenden. Nachdem die Datei abgespeichert wurde, ist das aktuelle Fenster nicht mehr mit **Unbenannt** sondern **uebungsfragebogen.sav[DatenSet0]** betitelt. Somit wurden

sämtliche Informationen in eine Datei gesichert, d.h. die Definition der Variablen (und später auch die erfassten Datenwerte) aus dem temporären Arbeitsblatt wurden in eine permanente SPSS Arbeitsdatei auf Festplatte oder ein sonstiges Medium gespeichert.

3 Erfassen bzw. Eingabe der kodierten Daten, Datenkontrolle

Man könnte nun, nachdem man von der Variablenansicht zur Datenansicht gewechselt hat, unmittelbar mit der Dateneingabe beginnen. Übungshalber wollen wir aber die vorher gespeicherte **Datei neu einlesen**. Dies erfolgt über die Menüpunkte **Datei > Öffnen > Daten**.

In den meisten Statistikpaketen und so auch in SPSS ist es üblich die Daten in Tabellenform organisiert zu haben und zwar so, dass die **Zeilen** immer den **Beobachtungseinheiten** (Personen, Objekte, etc.) und die **Spalten** den **Variablen** (Alter, Geschlecht, etc.) entsprechen.

Wenn wir Daten aus (unseren) Fragebögen eingeben, werden wir normalerweise die Fragebogen der Reihe nach abarbeiten, d.h. für jeden Fragebogen werden wir eine Zeile eingeben.

Die erste Zahl, die erfasst werden soll ist die ID-Nummer des Fragebogens, nehmen wir an diese sei 109. Die linke obere Ecke des Daten-Editor Fensters ist schwarz umrandet, d.h. diese Zelle ist bereit, einen Wert aufzunehmen. Wenn nun die Zahl 1 eingetippt wird, erscheint sie über dem oberen Tabellenrand und wird erst dann in die dafür bereite (schwarz umrandete) Zelle eingefügt, wenn die Eingabe-Taste gedrückt wird. Die Tabelle (Ausschnitt) sollte dann folgendes Aussehen haben:

	id	sex	liebalt	groesse	monate	da
1	109
2						
3						

Das umrandete Feld ist eine Zeile tiefer gerutscht und wir könnten dort die erste Zahl des zweiten Fragebogens eingeben. Wir wollen aber die Fragebögen der Reihe nach abarbeiten, d.h. es soll nun die zweite Zahl der ersten Zeile eingegeben werden. Dazu bewegen wir die Umrandung zum entsprechenden Feld mittels der Richtungstasten: **→**, **↑**, **↓**, oder **←**. Dann geben wir die entsprechende Zahl ein. Nun kann wie vorher wieder die Return-Taste gedrückt werden (sollte ein Tippfehler passiert sein, kann dieser vorher korrigiert werden), man kann aber ebenso die Richtungstasten zur nächsten Zelle bewegen, in die man Daten eingeben will, und der Inhalt des Eingabefeldes wird genauso in die Tabelle eingefügt (man kann sich also die Return-Taste sparen). Nach der Eingabe einiger Fragebögen könnte der Tabellenausschnitt jetzt folgende Form haben:

	id	sex	liebalt	groesse	monate	date
2	16	1	3	166	241	5
3	17	2	2	178	231	3
4	18	1	3	154	265	3
5	19	1	1	164	225	2
6	111	2	1	181	247	4
7						
8						

Sollte ein **Eingabefehler** (eine falsche Zahl) entdeckt werden, bewegt man das markierte Feld einfach an die betreffende Stelle und bessert die falsche Zahl aus. Hat man eine Zeile vergessen, so kann man die entsprechende Zeile auch am Ende anfügen (für viele statistische Anwendungen ist die Reihenfolge der Datenzeilen egal, außer Daten sind in einer natürlichen Reihenfolge geordnet, wie z.B. nach Jahreszahlen). Will man aber die vergessene Zeile an der richtigen Stelle haben und daher eine **ganze Zeile einfügen**, dann klickt man mit der linken Maustaste auf die Zeilennummer am linken Rand der Tabelle, vor der man die Zeile einfügen will. Die entsprechende Zeile ist dann markiert dargestellt. Danach hat man 2 Möglichkeiten: entweder man verwendet (1) die Symbolleiste, indem man auf das Symbol für **Fall einfügen** klickt, oder man aktiviert (2) den Menüpunkt: **Bearbeiten** in der Menüleiste und klickt auf das Feld: **Fall einfügen..** Alle Zeilen ab der vorher markierten rutschen um eine Zeile nach unten und es können nun die entsprechenden Zahlen in die freigewordene Zeile wie gewünscht eingegeben werden.

Das **Löschen einer Zeile** erfolgt auf ähnliche Weise. Zuerst wird wie oben die Zeilennummer angeklickt. Dann löscht man die entsprechende Zeile entweder durch Drücken der **Entf** - Taste oder man klickt im aktivierten Menü: **Editor** das Feld: **Löschen** an. Das Löschen ganzer Zeilen ist dann wichtig, wenn irrtümlich leere Zeilen generiert wurden (in solchen gibt es keine Einträge, sondern man sieht nur die Kommas, bzw. Dezimalpunkte, ohne Zahlen). SPSS würde solche Zeilen fälschlicherweise als *fehlend* auffassen (am Ende dieses Kapitel finden sich noch einige Bemerkungen zu fehlenden Werten oder *Missing Values*).



*In unserem Beispiel kommen keine **Dezimalzahlen** vor, allerdings ist hier in SPSS Vorsicht geboten. Zu beachten ist, wie das Dezimalzeichen definiert ist. Dies kann entweder das im Deutschen üblichen Komma oder der im Englischen verwendete Punkt sein. In der Standardeinstellung hängt dies davon ab, wie Windows - in der Systemsteuerung – Ländereinstellungen – Zahlen - Dezimaltrennzeichen - konfiguriert ist. In deutschen Versionen wird dies üblicherweise ein Komma sein und entsprechend wird man bei der Dateneingabe also ein Komma verwenden. Allerdings gibt es Situationen, wie z.B. bei der Berechnung neuer Variablen, wo SPSS einen Punkt als Dezimalzeichen verlangt. Sollten Sie also bei irgendeinem Vorgang eine kryptische Fehlermeldung bezüglich eines Kommas erhalten, versuchen Sie den Punkt als Dezimaltrennzeichen.*

Datenkontrolle

Bevor man mit der statistischen Analyse von Daten beginnt, sollte man diese immer zuerst auf Korrektheit überprüfen. Bei kleinen Datensätzen, die man vielleicht schon auf Papier erfasst hat, geht das leicht, indem man einfach die eingegebenen Daten mit den Rohdaten vergleicht. Bei größeren Datensätzen kann man zu diesem Zweck Statistikprozeduren verwenden, die eine Übersicht der eingegebenen Daten liefern. Dies kann zum Beispiel mittels der Prozeduren: **Häufigkeiten...** bzw. **Kreuztabellen...** erfolgen. (Diese werden aus dem Untermenü: **deskriptive Statistiken** des Hauptmenüpunktes: **Analysieren**, das aus der Menüleiste aktiviert wird, aufgerufen. Darauf wird in Einheit 2 bzw. 3 noch genauer eingegangen.) Die Ergebnisse dieser Prozeduren werden dann darauf überprüft, ob die

aufgelisteten Werte sinnvoll sind. (Haben wir zum Beispiel Geschlecht mit 1 und 2 kodiert und finden wir in einer Ergebnistabelle auch mehrmals die Zahl 8, dann stimmt offensichtlich etwas nicht.)

Abspeichern der Daten

Zur Wiederholung: geht, wie am Ende des vorigen Abschnitts beschrieben, über die Menüpunkte **Datei > Speichern** bzw. **Speichern unter...** bzw. mit Klick auf **.An** dieser Stelle sei (wieder einmal) darauf hingewiesen, dass man lieber zu oft als zuwenig speichern sollte. Ein Tipp daher:



SPEICHERN SIE BEI DER EINGABE GRÖßERER DATENMENGEN (CA.) ALLE 15 MINUTEN (das bisher gespeicherte wird dann ohnehin überschrieben, es wird also kein Speicherplatz verloren, und der Verlust, falls etwas passiert, beträgt maximal eine Viertelstunde).

Noch eine Anmerkung zu fehlenden Werten:

In der Praxis passiert es leider häufig, dass nicht immer alle Daten vorhanden sind. So gibt es etwa bei Fragebogenuntersuchungen immer wieder Personen, die auf bestimmte Fragen keine Antwort geben bzw. keine der vorgesehenen Antwortmöglichkeiten ankreuzen. Die Frage ist, wie wir mit solchen Daten umgehen. Eine radikale Möglichkeit ist es, die Daten der entsprechende Person überhaupt nicht zu berücksichtigen. Allerdings fehlen oft nur einzelne Werte und es wäre schade, auf die restlichen Daten zu verzichten. In diesem Fall spricht man von fehlenden Werten (oder *missing values*). Wird, wie in unserem Beispiel, der Daten-Editor (die Datentabelle) zur Eingabe der Werte verwendet, dann lässt man das entsprechende Feld einfach leer, d.h. in der Datentabelle ist nur das Dezimalzeichen zu sehen. SPSS erkennt dann automatisch, dass es sich hierbei um einen fehlenden Wert handelt und berücksichtigt dies in weiteren Analyseschritten. Man erkennt dies in den jeweiligen Ergebnissen immer anhand eines Ausgabeteils, der anzeigt, wie viele Beobachtungen von SPSS als fehlend aufgefasst wurden. Meist erfolgt dies durch die Angabe der Anzahl fehlender Werte bzw. auch deren Gegenteil, nämlich gültiger Werte.

4 Transformieren der Daten bzw. Erzeugen von neuen Variablen

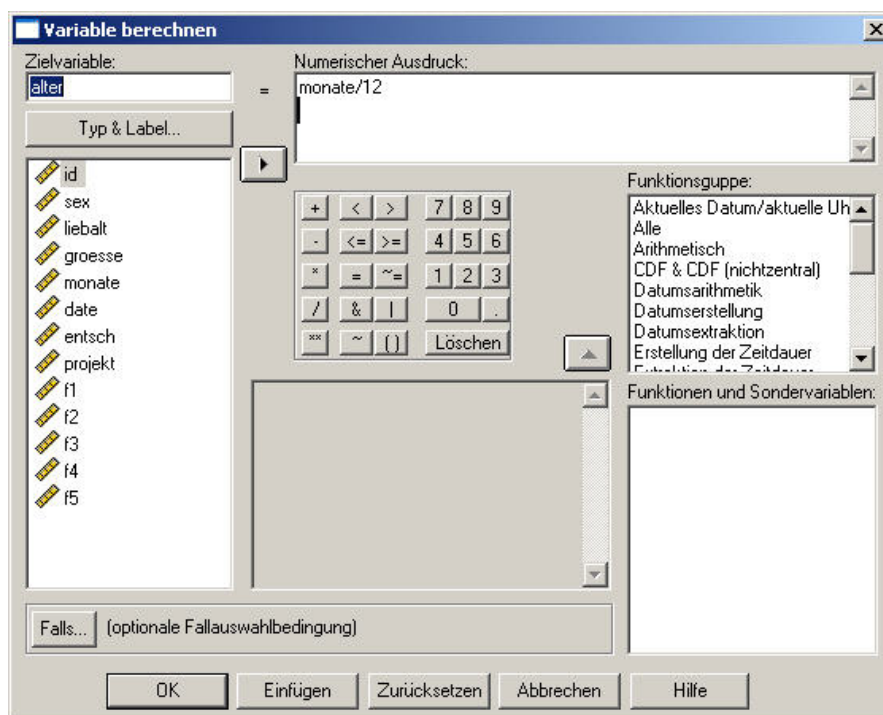
In diesem Abschnitt werden Methoden beschrieben, um neue Variablen zu erzeugen. Hierbei geht es vor allem um das (arithmetische) Neuberechnen (**Berechnen**) von Variablen bzw. das Ersetzen (**Umkodieren**) bestimmter Werte einer (alten) Variable durch neue Werte (in einer neuen Variable). Beide Methoden sind als Unterpunkte im Menü **Transformieren** enthalten

Berechnen neuer Variablen

Dies betrifft meistens Situationen, in denen man Variablen in die Untersuchung einbeziehen will, die sich aus den vorhandenen Variablen ableiten lassen. Zwei Beispiele aus unserem Fragebogen sollen die Vorgangsweise veranschaulichen.

Zunächst wollen wir das Alter der Befragten, das wir in Monaten erfasst haben in Jahre umrechnen. Die neue Variable wollen wir `alter` nennen.

Über das Menü **Transformieren > Berechnen** öffnet sich folgendes Dialogfenster (wir lassen uns der Übersichtlichkeit wegen in der Variablenliste nur die Namen anzeigen, siehe S.7)



Ins linke obere Feld **Zielvariable:** trägt man die Bezeichnung der neuen Variable ein, hier `alter`. Im Feld **Numerischer Ausdruck:** wird die „Formel“ eingegeben, über die der Wert der Zielvariablen für jede Beobachtung errechnet werden soll, in unserem Beispiel also `monate/12`. Hierzu stehen Ihnen arithmetische Operatoren, logische Ausdrücke und diverse Funktionen zur Verfügung. Zu beachten ist, dass für jede Datenzeile (Beobachtung) ein Wert der Zielvariable aus dem entsprechenden Wert, der sich aus dem rechten Feld ergibt, berechnet wird. Ist in einer Zeile ein Wert einer der rechts verwendeten Variablen fehlend, dann ist der Wert der Zielvariable auch fehlend.

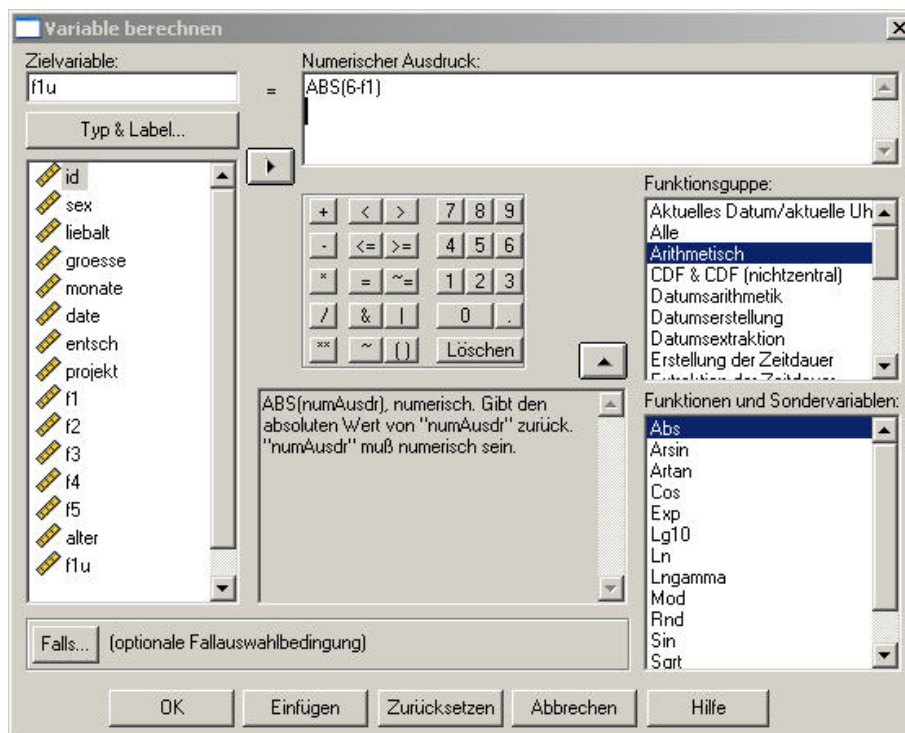
Sobald man OK geklickt hat, kann man die Ergebnisse der Berechnung in der Datenansicht sehen. Neue Variablen werden immer ganz rechts, hinter der letzten Variable, angefügt.

Als zweites Beispiel soll eine etwas komplexere Aufgabe gelöst werden, die sich bei der Analyse von Fragebögen öfters ergibt. In unserem Fragebogen entstammen die letzten fünf Fragen einem längeren Stress-Fragebogen, der messen soll, wie sehr jemand unter Stress leidet und damit umgeht. Die extremen Ausprägungen sind: „Leben unter starkem Zeitdruck, Wettbewerb und Zielgerichtetheit, aber auch Mangel an Erfolgserlebnissen und Freude an geleisteten Tätigkeiten“ vs „keine Getriebenheit, easy going, gute Fähigkeit zu Delegieren, entspannt“. Eine übliche Vorgangsweise bei solchen Fragebogenskalen ist es, die Werte der einzelnen Items (Fragen) zusammenzuzählen und den Summenscore als zusammenfassenden, beschreibenden Wert zu verwenden. Bei Ansicht unseres Fragebogens sehen wir aber, dass das nicht ohne weiteres geht, da das Ergebnis keinen Sinn machen würde. Hat man bei allen 5 Fragen „trifft sehr zu“ mit 1 und „trifft überhaupt nicht zu“ mit 5 kodiert, dann bedeutet bei Frage 1: „bei Terminen oder Verabredungen bin ich nie zu spät“ ein niedriger Wert hohen Stress, während bei Frage 2: „ich kann gut zuhören, lasse andere zuerst sprechen“ ein niedriger Wert auch niedrigen Stress indiziert. Man nennt diese Eigenschaft von Items auch „Polung“. Einfach Zusammenzählen geht also nicht, wir müssen zunächst bei einigen Items die Polung umkehren (d.h. 1 zu 5, 2 zu 4, etc. machen), bevor wir den Summenscore bilden können. Dazu müssen wir zunächst entscheiden was ein hoher Summenscore bedeuten soll. Wenn wir wollen, dass ein hoher Wert auch hohen Stress anzeigt, dann passen Items 2, 4 und 5. Die anderen beiden, nämlich 1 und 3 müssen „umgepolt“ werden. Eine einfache Formel erlaubt diese Transformation:

$$\text{neuer (umgepolter) Wert} = \text{Absolutbetrag}(6 - \text{alter Wert})$$

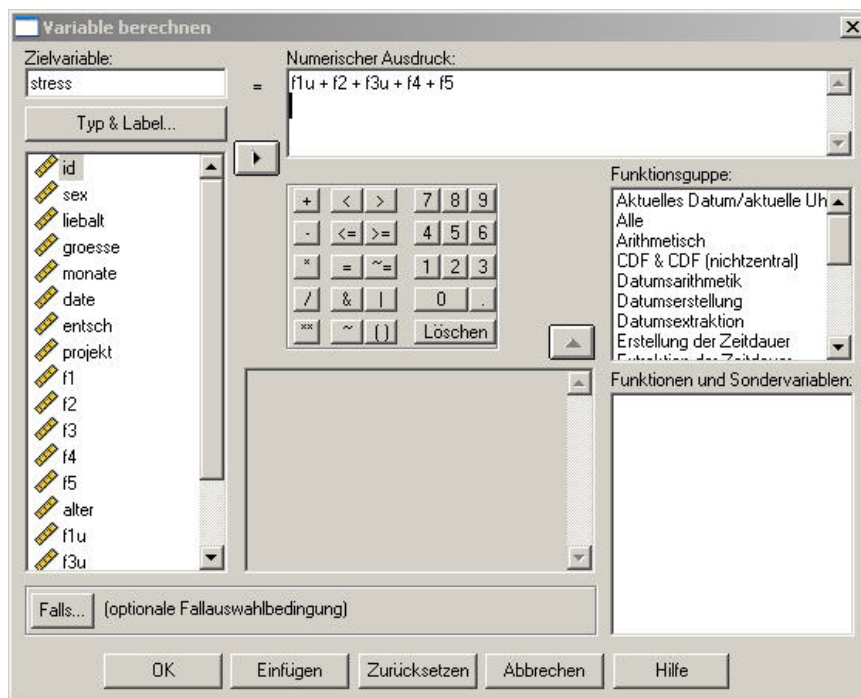
Die Zahl 6 ergibt sich aus dem höchsten Wert, der vorkommen kann plus 1. (Wären die Kategorien nicht von 1 bis 5, sondern von 0 bis 4 definiert, dann wäre die rechte Seite $\text{Absolutbetrag}(4 - \text{alter Wert})$). Wenn man 0 als niedrigste Kategorie verwendet, braucht man also zur höchsten nicht 1 dazuzählen).

In SPSS würde das Berechnungsfenster für Variable f1u also so aussehen (f1u soll der Name der neuen, umgepolten Variable zu Frage 1 sein):



Wie haben hier eine Funktion, nämlich ABS(numerischer Ausdruck) aus der Liste der Funktionen und Sondervariablen verwendet. Dazu wählt man zuerst die Funktionsgruppe aus, zu welcher die jeweilige Funktion gehört (hier: **Arithmetisch**). Durch Klicken auf eine der Funktionen aus der dann erscheinenden Liste erhält man im Fenster links davon eine kurze Beschreibung der jeweiligen Funktion.

Wenn die gleiche Berechnung noch für die zweite „verkehrt“ gepolte Frage, nämlich Frage 3 durchführen, haben wir alle Fragen so gepolt, dass wir den Summenscore, den wir **stress** nennen wollen, bilden können. Das entsprechende Dialogfenster würde dann so aussehen:



Anmerkung zu numerischen Ausdrücken:

Ganz allgemein sind numerische Ausdrücke in SPSS aus einem oder mehreren der folgenden Elemente zusammengesetzt:

- Bereits definierte Variablen
- Zahlen
- arithmetische Operatoren:
 - Addition (+)
 - Subtraktion (-)
 - Multiplikation (*)
 - Division (/)
 - Potenzfunktion (**)
 - Klammern
 - Funktionen

In numerischen Ausdrücken können Sie zahlreiche Funktionen verwenden, die numerische Variablen oder Zahlen als Argumente (*arg*) verarbeiten. Einige wichtige Funktionen sind:

- ABS(*arg*) Absoluter Wert
- EXP(*arg*) Exponentialfunktion
- LN(*arg*) Natürlicher Logarithmus

- $RND(arg)$ auf eine ganze Zahl gerundeter Wert
- $TRUNC(arg)$ auf eine ganze Zahl abgerundeter Wert
- $SQRT(arg)$ Quadratwurzel

Für alle anderen Funktionen sei auf die SPSS - Hilfe verwiesen. Auf diese soll später noch näher eingegangen werden.

Übungsaufgabe: Versuchen Sie die vorher erstellte Variable **alter** (Alter in Jahren statt Monaten) so zu berechnen, dass die neue Variable **alter** ganzzahlig ist.

Umkodieren von Variablen

Im Gegensatz zu **Berechnen** verwendet man **Umkodieren** vor allem um Werte einer Variable zu ändern bzw. Werte umzugruppieren. Dabei wird an allen Stellen einer Variable, in der ein zu bestimmender Wert auftritt, dieser Wert durch einen anderen ersetzt und das Ergebnis entweder in eine neue Variable gespeichert (man kann das Ergebnis auch in die alte Variable speichern, aber in den meisten Fällen ist das nicht empfehlenswert).

Meistens verwendet man **Umkodieren** um Kategorien neu zu definieren oder Werte zusammenzufassen.

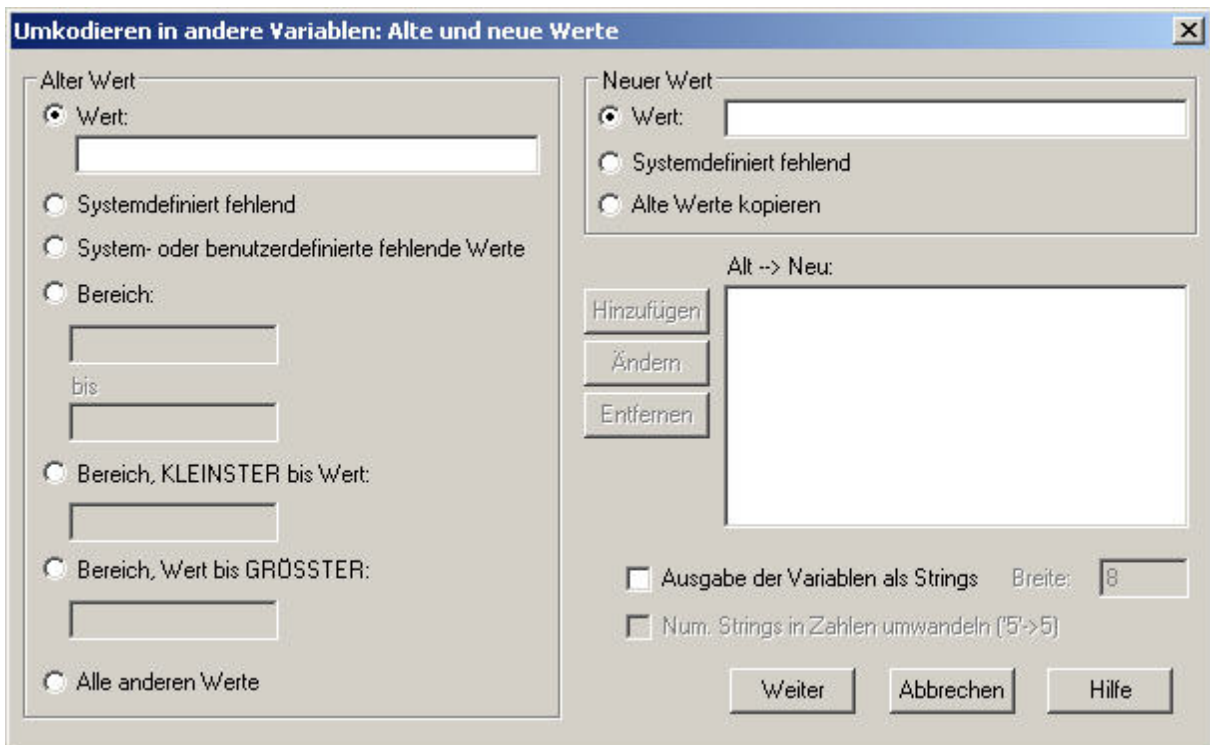
Als Beispiel wollen wir die Variable **sex** umkodieren, und zwar so, dass für männlich 0 und für weiblich 1 verwendet wird (jetzt ist männlich mit 2 und weiblich mit 1 kodiert, wir müssen also nur männlich ändern).

Das erreichen wir so: über **Transformieren > Umkodieren > In andere Variablen...** kommen wir in das Definitionsfenster, wo schon die Ausgangsvariable **sex** in das mittlere Feld bewegt wurde.

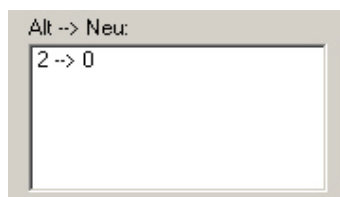


Das Fragezeichen bei **sex -> ?** bedeutet, dass wir einen Namen für die neue Variable angeben sollen. Dazu klicken wir in das Feld: **Ausgabvariable Name:** und geben dort **sex01** ein. Nach klicken auf die Taste: **Zuweisen** wird das Fragezeichen im mittleren Feld durch **sex01** ersetzt. Im Feld **Beschriftung:** könnten wir einen längeren Namen für die Variable angeben. Nun können wir definieren, unter welchen Bedingungen die Variable

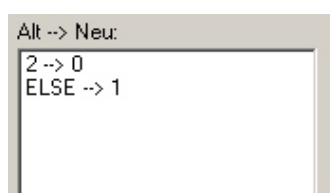
sex01 welchen Wert annehmen soll, in dem wir auf die Taste: **Alte und neue Werte...** klicken. Wir öffnen damit folgendes Fenster:



Wir müssen nun dieses 'Formular' ausfüllen. Zunächst definieren wir den Werte der neuen Variable **sex01**, die den neuen Wert 0, nämlich 'männlich', erhalten sollen. Dies sind alle Werte von **sex**, die jetzt 2 sind. Hierzu markieren wir auf der linken Seite (**Alter Wert**) das Feld: **Wert:** und geben die Zahl 2 ein. Nun können wir den neuen Wert definieren, der für diese Zahl soll. Dazu schreiben wir im rechten Teilfenster: **Neuer Wert** die Zahl 0 in das leere Feld neben **Wert:**. Schließlich müssen wir noch im Teilfenster: **Alt -> Neu:** das Feld: **Hinzufügen** anklicken (nicht vergessen!). Das Resultat sieht folgendermaßen aus:



Damit ist das Wichtigste schon erledigt, wir müssen nur noch die Werte für weiblich definieren, die mit 1 aber gleich bleiben sollen. Hierzu können wir entweder genauso vorgehen wie oben (also den alten Wert 1 durch den neuen Wert 1 ersetzen) oder wir verwenden links die Option: **Alle anderen Werte** und setzen ebenso wie oben 1 für den neuen Wert ein. Dann wieder **Hinzufügen** klicken. Das Ergebnis ist dann:



Nach **Weiter** und dann **OK** sollte die so erzeugte neue Variable wieder am Ende in der Datenansicht angefügt sein.



*Manchmal werden Daten-Transformationen nicht sofort ausgeführt, d.h. man sieht das Ergebnis nicht unmittelbar nach **Berechnen** oder **Umkodieren**. Das kann auf eine Voreinstellung zurückzuführen sein. Sollte das bei Ihnen der Fall sein, dann können Sie diese Einstellung ändern über **Bearbeiten > Optionen** . Im Fenster **Daten** markiert man **Werte sofort berechnen** links oben.*

Anmerkung zum Umkodieren:

Umkodieren eine sehr einfache Datentransformation, sie ist aber oft sehr nützlich und leichter handzuhaben als Berechnen (mit der man alle Umkodierungsoperationen auch durchführen könnte). Folgendes ist bei der praktischen Anwendung zu beachten:

- o Sie können bei einem Einsatz der Dialogbox **Umkodieren in andere Variablen** beliebig viele Variablen gleichzeitig umkodieren.
- o Bei der Spezifikation der alten Werte, die auf einen neuen Wert abgebildet werden sollen, können Sie angeben:
 - Einen einzelnen **Wert**
 - **Systemdefiniert fehlend** (SYSMIS) – das entspricht dem oben erwähnten Dezimalzeichen (So ist es also möglich, den automatischen Missing-Data-(MD) Indikator auf einen anderen Wert umzusetzen).
 - System- oder benutzerdefinierte fehlende Werte
 - Den **Bereich** von einem ersten Wert bis zu einem zweiten Wert (jeweils inklusive)
 - Den **Bereich** vom niedrigsten Wert in der Stichprobe bis zu einem bestimmten Wert (inklusive)
 - Den **Bereich** von einem bestimmten Wert bis zum höchsten Wert in der Stichprobe (inklusive)
 - **Alle anderen Werte** (damit sind alle in keiner anderen Ersetzungsvorschrift genannten Werte angesprochen , inklusive MD-Indikatoren, auch System-Missing).
- o Als neuen Wert, auf den die alten Werte einer Ersetzungsvorschrift abgebildet werden sollen, können Sie angeben:
 - Einen **Wert**
 - Systemdefiniert fehlend (SYSMIS , dann werden alle zugehörigen alten Werte auf SYSMIS umgesetzt.
 - **Alte Werte kopieren** (Diese Möglichkeit steht nur beim Umkodieren in *andere* Variablen zur Verfügung und bewirkt dann für die zugehörigen alten Werte eine unveränderte Übernahme. Dies ist besonders nützlich, wenn die alten Werte mit **Alle anderen Werte** spezifiziert worden sind).
- o Sie können beliebig viele Ersetzungsvorschriften festlegen. SPSS bringt diese automatisch in eine sinnvolle Ordnung.

5 Auswählen von Fällen

Es kommt durchaus vor, dass man sich bei einer statistischen Analyse auf eine Teilstichprobe beschränken möchte. Man möchte z.B. bei einer bestimmten Analyse getrennt Ergebnisse für Frauen und Männer. Dies lässt sich auf einfache Weise dadurch erreichen, dass bei dieser Analyse jeweils eine Geschlechts-Teilstichprobe ausgewählt wird. Wir wollen als Beispiel alle Frauen in unserem Datensatz auswählen.

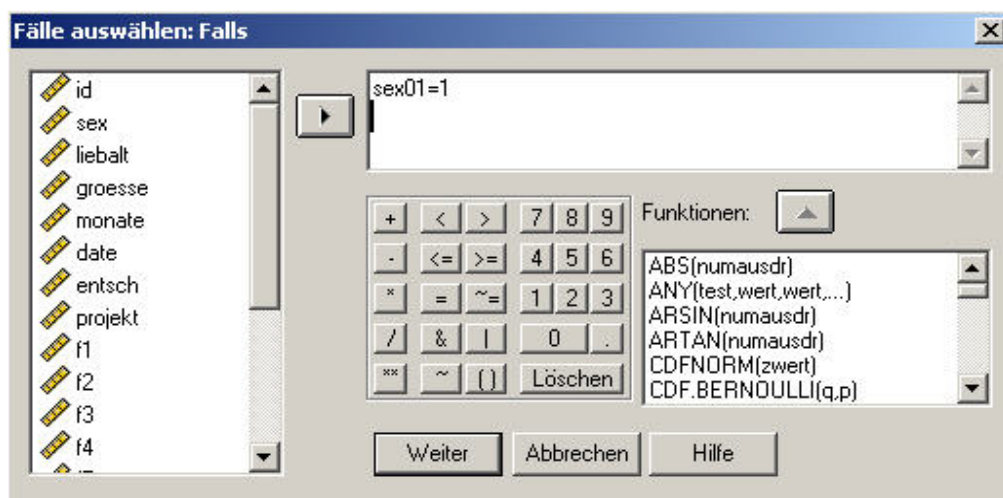
SPSS erlaubt es, Fälle in Abhängigkeit von einer Bedingung, die man festlegen muss, temporär oder permanent aus der Datendatei auszuwählen.

Das entsprechende Dialogfenster erhält man über **Daten > Fälle auswählen...**

Zur Festlegung der Bedingung für die Teilnahme an den weiteren Auswertungen markiert man die Option **Falls Bedingung zutrifft** und klickt auf **Falls...** .



Es öffnet sich



wo schon die Bedingung **sex01 = 1** eingegeben ist. Diese Bedingung besagt, dass ab jetzt für alle weiteren Arbeitsschritte nur mehr solche Fälle herangezogen werden, in denen die Variable **sex01** den Wert 1, also weiblich, hat. Wenn man wieder alle Fälle verwenden möchte, muss man diese Bedingung wieder aufheben (dazu später).

Nach Klicken auf **Weiter** kommt man wieder in das vorherige Dialogfenster und hier kann man sich noch entscheiden, was man mit den **nicht ausgewählten** Fällen machen will. Üblicherweise wird man diese nur temporär aus weiteren Analysen ausschließen, in diesem Fall markiert man **Nicht ausgewählte Fälle filtern**. Will man diese Fälle aber überhaupt „loswerden“ dann kann man sie hier auch **Löschen**. Zusätzlich besteht auch noch die Möglichkeit, ein neues Datenfile zu erzeugen, welches nur die ausgewählten Fälle beinhaltet.

Mit OK wird der Vorgang abgeschlossen.

In der Datenansicht ergibt sich jetzt folgendes Bild (wenn man mit der Laufleiste den Ausschnitt ganz nach rechts verschiebt):

	sex01	filter_\$	var	var	var	var
1	1,00	1				
2	1,00	1				
3	,00	0				
4	1,00	1				
5	1,00	1				
6	,00	0				
7						

Erstens sieht man, an den durchgestrichenen Zeilennummern links, welche Fälle ausgeschlossen wurden, nämlich jene bei denen die Bedingung **sex01 = 1** nicht zutrifft. Zweitens sehen wir, dass SPSS eine Variable **filter_\$** angelegt hat, die immer dann, wenn die Bedingung zutrifft, den Wert 1, sonst aber den Wert 0, hat.

Diese Variable bleibt erhalten und wird nur dann überschrieben, wenn man eine neue Bedingung einführt. Man kann sie also auch später noch verwenden (SPSS legt zusätzlich einen Variablenlabel für diese Filtervariable an, für unser Beispiel „sex01 = 1 (FILTER)“, damit man auch später noch die im Falls-Fenster festgelegte Bedingung weiß, auf der der Filter beruht).

Wichtig ist, dass man nicht vergisst, diese **Auswahl wieder aufzuheben**, wenn man sie nicht mehr braucht. Diese geschieht, in dem man im **Fälle auswählen – Fenster** (über **Daten > Fälle auswählen...**) die Option **alle Fälle** markiert.

Die Schaltflächen im Dialogfenster **Fälle auswählen: Falls** haben folgende Bedeutung:

	<	kleiner als		>	größer als
	<=	gleich oder kleiner		>=	gleich oder größer
	=	gleich		~=	ungleich
	&	UND			ODER
	~	NICHT			

Bei zusammengesetzten logischen Ausdrücken unter Verwendung von NICHT, ODER bzw. UND sollte man immer genau auf die Regeln für logische Operationen achten und lieber zu viele als zu wenig Klammern verwenden. Eine detaillierte Beschreibung dieser Regeln würde hier zu weit führen, eine ausführlichere Beschreibung kann man im (englischen) Syntax-Guide über die SPSS Hilfe finden.

6 Gewichten von Daten

Manchmal werden in statistischen Analysen Fällen gewichtet um bestimmte Mankos bei Stichprobenerhebungen auszugleichen. Meist sind solche Gewichtungen methodisch fragwürdig und darauf soll hier auch nicht eingegangen werden.

Es gibt aber Situationen, wo wir die Gewichtungsfunktionen von SPSS benötigen um bestimmte Analysen überhaupt durchführen zu können. Dies betrifft vor allem kategoriale Variablen, die in Tabellenform und nicht als Einzelbeobachtungen gegeben sind. Dies wird in den Kapiteln 2 und 3 behandelt und wir wollen hier nicht näher darauf eingehen.

Jedenfalls erfolgt die Gewichtung über die Menüpunkte **Daten > Fälle gewichten...** wo man eine Gewichtsvariable angeben kann. Wie schon beim Auswählen von Daten sollte man auch hier nicht vergessen, die Gewichtung wieder „auszuschalten“ (im gleichen Dialogfenster), wenn man sie nicht mehr benötigt.

7 Analysieren von Daten

Da die eigentliche Datenanalyse in den weiteren Kapiteln dieses Skriptums im Vordergrund steht, sollen hier nur ganz kurz die wesentlichen Aspekte überblicksmäßig dargestellt werden. Alle Details werden dann später ausführlich behandelt.

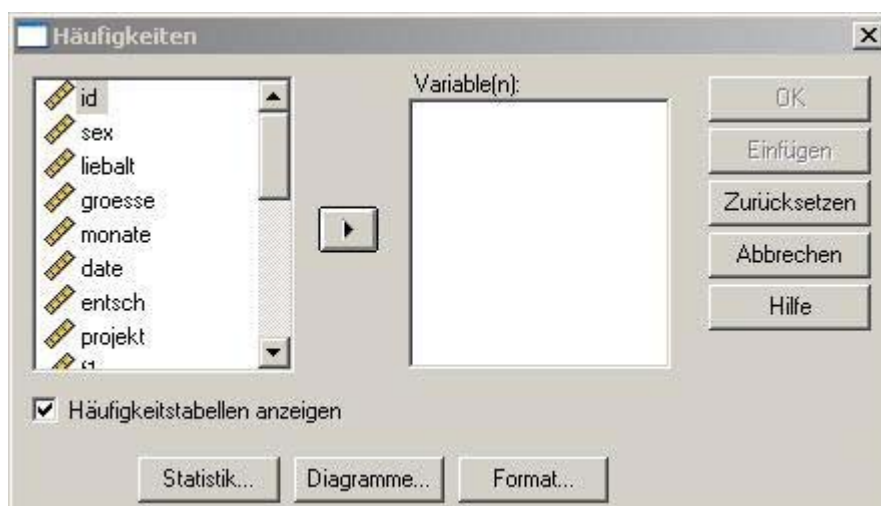
Wie wollen anhand des vorher angesprochenen Punktes **Datenkontrolle** kurz folgendes behandeln:

- wie werden numerische und grafische Prozeduren in SPSS aufgerufen
- wie sieht das Ergebnis aus
- wie kann man das Ergebnis bearbeiten und speichern

Ein wesentlicher Aspekt von Datenkontrolle ist es, zu überprüfen ob die Werte der eingegebenen Daten in einem sinnvollen Bereich. Gibt es beispielsweise für Körpergröße einen Wert mit 389 cm, dann wird wohl ein Fehler vorliegen. Nun kann es bei größeren Datensätzen relativ schwierig sein, solche Fehler zu entdecken (um sie dann gegebenenfalls ausbessern zu können). Am besten verwendet man in solchen Fällen Methoden der beschreibenden Statistik (Häufigkeitsauszählungen, grafische Darstellungen der Verteilungen der Variablen) um ungewöhnliche Werte identifizieren zu können.

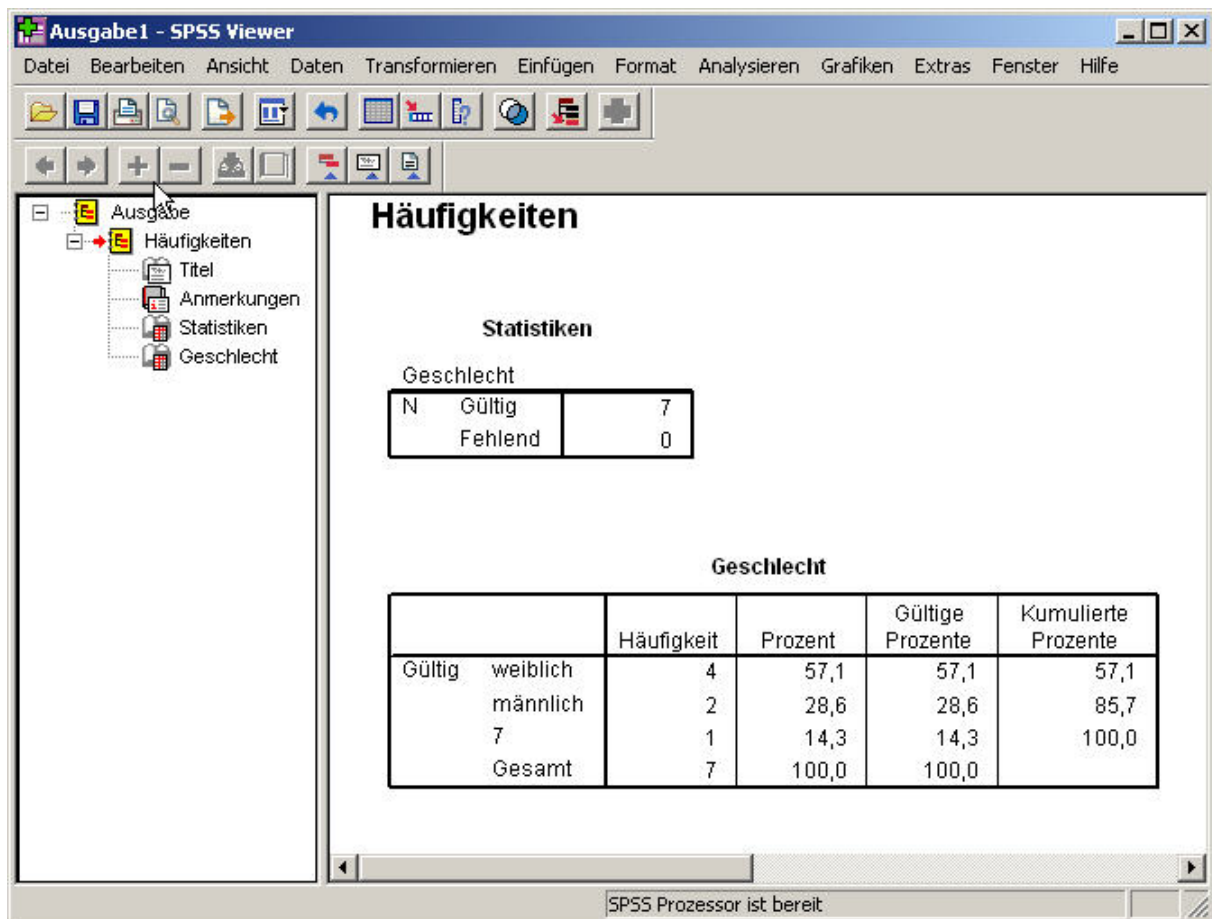
Als Illustration sollen zwei Analyseprozeduren – **Häufigkeiten** als numerisches und **Balkendiagramme** als grafisches Beispiel – für die Variable **sex** vorgestellt werden.

Eine Übersicht über alle im Datenfile vorkommenden Werte einer bestimmten Variable bekommt man über **Analysieren > deskriptive Statistiken > Häufigkeiten...**



Jene Variablen, für die man Häufigkeitsauszählungen will (das kann durchaus mehr als eine sein) bewegt man in das Feld **Variable(n)**;, die Box **Häufigkeitstabellen anzeigen** sollte markiert sein.

Nach Klicken von **OK** öffnet sich das Ausgabefenster (**SPSS Output Viewer**):



Das Fenster des SPSS Viewers besteht aus einer Menüleiste (einige Menüpunkte sind die gleichen wie beim Daten-Editor), zwei Symbolleisten (ebenfalls mit einigen gleichen Symbolen), sowie aus zwei darunter liegenden Fenstern. Im rechten, größeren Fenster ("Inhaltsfenster") finden sich die Ergebnisse der verwendeten Prozeduren (jetzt das von uns angeforderte Ergebnis der Prozedur: Häufigkeiten), das linke ("Gliederungsfenster") gibt eine Übersicht über alle Inhalte des rechten Ausgabefensters und dient zur Organisation und zum schnellen Zugriff (Navigieren) auf diese Inhalte. Der Inhalt des rechten Fensters bedeutet nun folgendes:

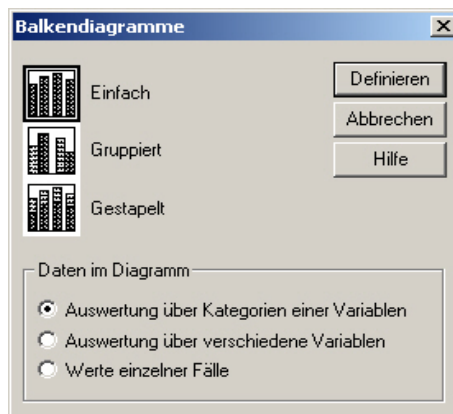
Zuerst, als Titel, steht der Name der verwendeten Prozedur (**Häufigkeiten**). Darunter findet sich eine kleine Tabelle (**Statistiken**), in der die Anzahl der fehlenden Werte für die berechnete(n) Variable(n) angegeben wird. Wie Sie sehen, findet sich in dieser Tabelle nicht die Bezeichnung für die Variable, nämlich **sex**, sondern der von uns vergebene Variablenlabel *Geschlecht*. (Da es in unserem Beispiel keine fehlenden Daten gibt, ist die entsprechende Zahl 0.)

Die nächste Tabelle enthält nun das eigentliche, uns interessierende Ergebnis, nämlich die Häufigkeiten (und diverse Prozentsätze) für jeden einzelnen Wert, der für die Variable **sex** im Datensatz vorkommt. Man sieht das anstatt der Zahlen 1 und 2 die Labels *weiblich* und *männlich* angegeben werden, aber es kommt auch der Wert 7 vor, und zwar mit Häufigkeit 1.

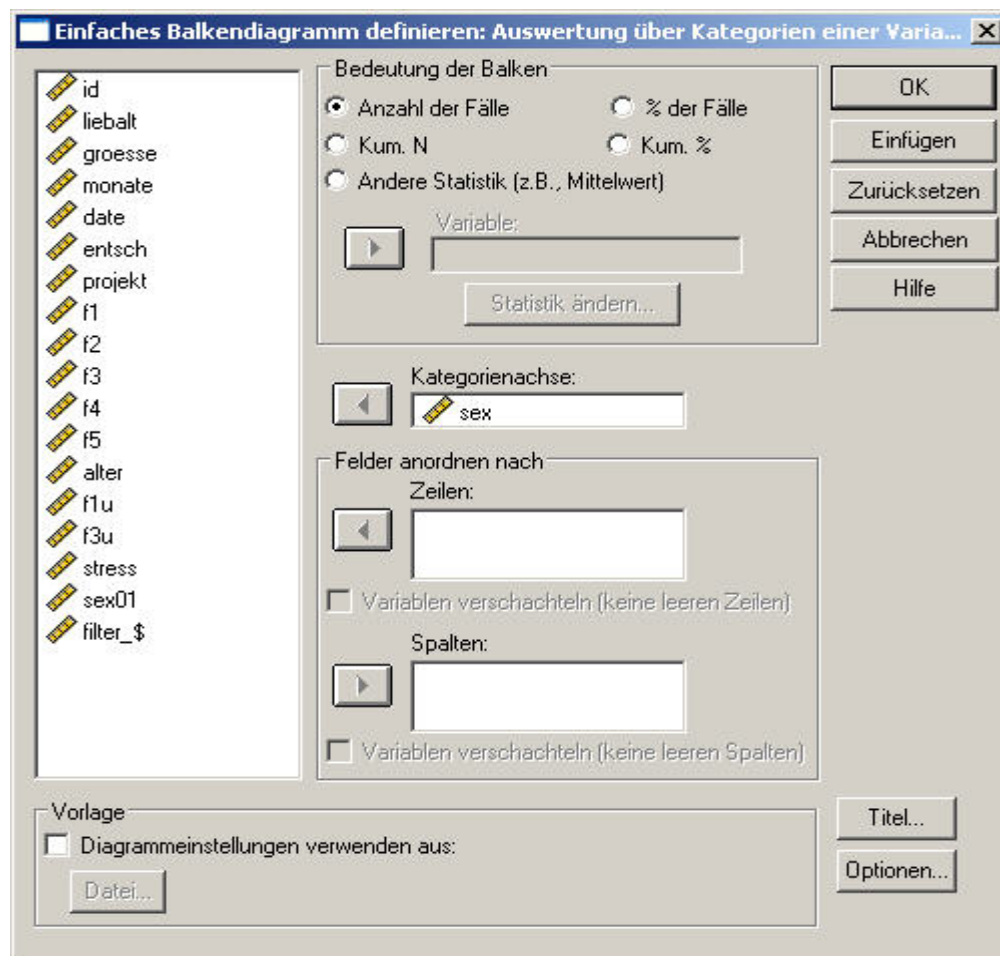
Irgendwo in unserem Datensatz wurde also einmal für Geschlecht 7 eingegeben und das kann offensichtlich nicht stimmen.

Das Gleiche hätten wir auch über eine Grafikprozedur erkennen können. Zur Illustration wollen wir ein Balkendiagramm verwenden. Zum Aufruf muss man nicht in das Datenfenster wechseln, da im Menü des Output Viewer auch der Punkt **Grafiken** (wie auch **Analysieren**) vorkommt.

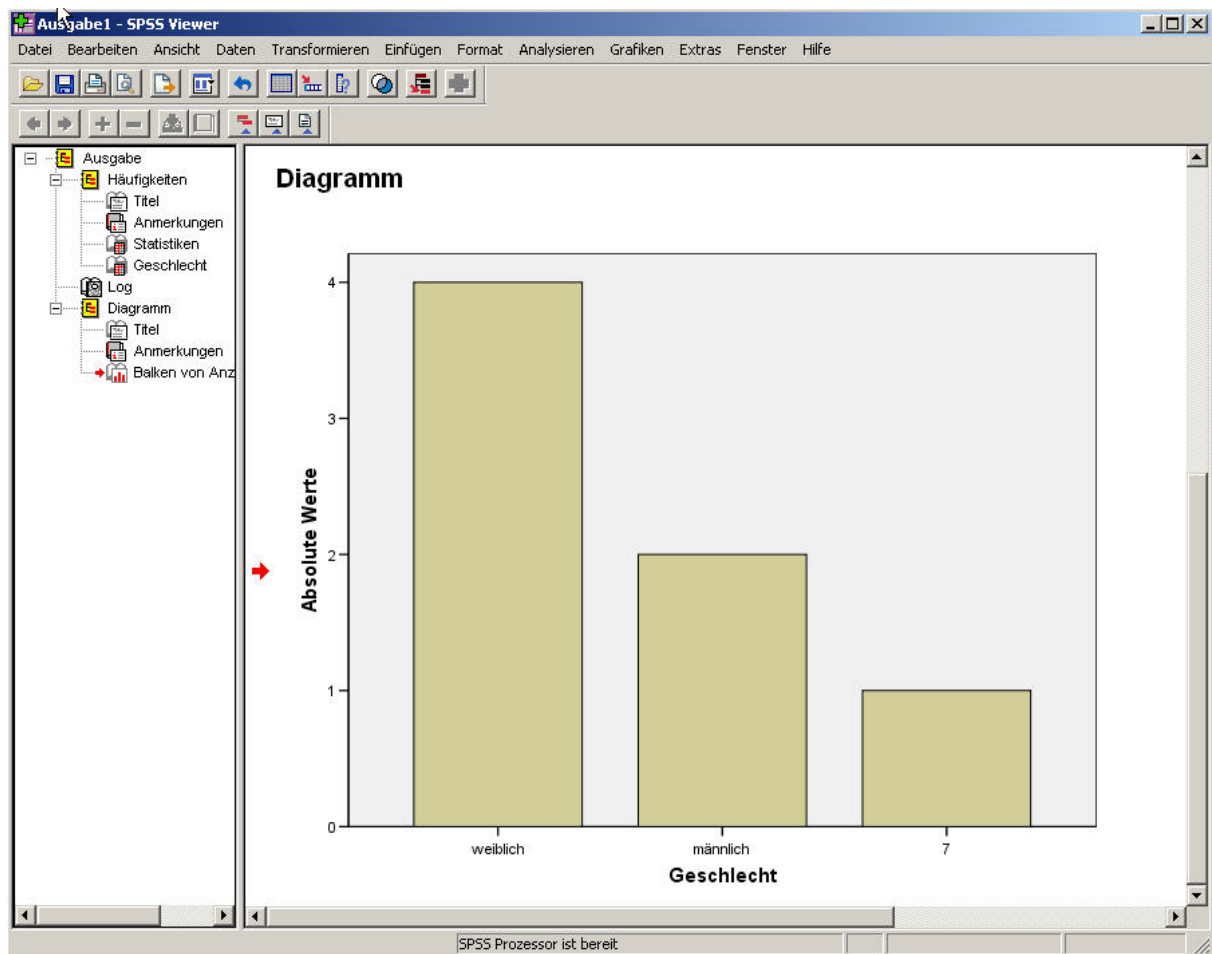
Über **Grafiken > Balken...** kommen wir ins erste Definitionsfenster



wo wir **Einfach** und **Auswertungen über Kategorien einer Variablen** anklicken. Mittels **Definieren** gelangen wir zur Dialogbox



wo wir schon **sex** in das Feld **Kategorienachse:** bewegt haben. Nach OK erhalten wir im Ausgabefenster das Balkendiagramm, in dem man natürlich auch den Datenfehler sehen kann.



An diesem einfachen Beispiel sollte nur gezeigt werden, wie man fehlerhafte Werte identifizieren kann und wie der Aufruf einer Statistik- bzw. einer Grafikprozedur abläuft. Da SPSS in den letzten Jahren sehr viel Energie in die Gestaltung der grafischen Benutzeroberfläche gesteckt hat, gäbe es hier eine Menge anzumerken. So kann der Output völlig umgestaltet werden (ein paar wichtige Dinge werden im nächsten Abschnitt beschrieben), es gibt jede Menge voreingestellter Optionen, die man ändern kann, über den Navigationsbaum links lässt sich die Ausgabe auf unterschiedlichste Weise darstellen und gruppieren, etc. Eine ausführliche Übersicht erhält man über das **Hilfe-Menü** unter **Lernprogramm** und dort im Kapitel **Arbeiten mit Ausgaben**.

8 Bearbeitung und Verwenden des Outputs

In diesem Abschnitt soll gezeigt werden, wie man den Standard-Output von SPSS modifizieren und an andere Softwareprodukte übergeben kann. Auf diese Möglichkeiten wird man meistens im Zusammenhang mit Präsentationen und/oder der Erstellung von Berichten zurück greifen. Im wesentlichen geht es dabei um die

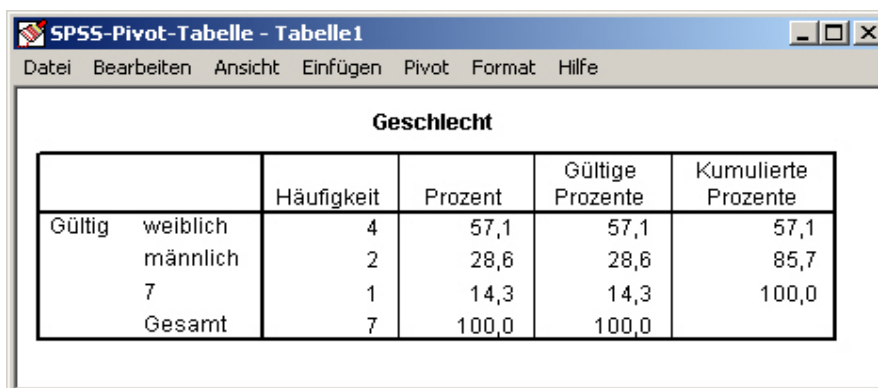
- Modifikation von Tabellen
- Modifikation von Grafiken
- Übernahme von Tabellen und Grafiken in Textverarbeitungs- bzw. Präsentationssoftware

Wie schon erwähnt, gibt es eine Vielzahl an Möglichkeiten die Form des vorhandenen Output zu ändern, und selbst eine Beschreibung, die über Grundlegendes hinausgeht, würde den Rahmen dieses Skripts sprengen. Daher soll im folgenden nur die grundsätzliche Vorgehensweise dargestellt werden, auf alles weiterführende wird an den entsprechenden Stellen verwiesen.

Grundsätzlich erfolgen **Modifikationen von Outputobjekten** (Überschriften/Text, Tabellen oder Grafiken) in eigenen Fenstern (Editoren), die man öffnet, indem man entweder irgendwo im Element **mit der rechten Maus** klickt und dann im untersten Menüpunkt das entsprechende Objekt öffnet (**oder**: es geht auch mit **Doppelklick** auf das zu verändernde Element, dann öffnet sich aber nur bei Grafiken der entsprechende Editor, ansonsten gibt es nur veränderte Menüs). Jedenfalls stehen dann je nach Art des zu verändernden Elements verschiedene Menüleiste zur Verfügung mittels derer man die Modifikationen durchführen kann.

Modifikation von Tabellen

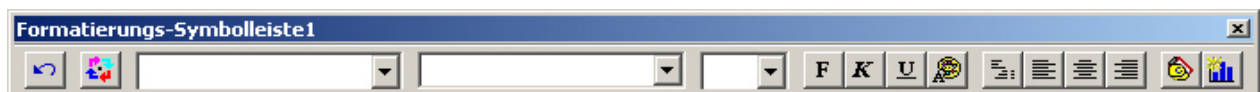
Zur Illustration verwenden wir die im vorigen Abschnitt erzeugte Häufigkeitstabelle für Geschlecht. Zur Modifikation klickt man also mit der rechten Maustaste irgendwo in die Tabelle. Aus dem sich öffnenden Fenster wählt man **Objekt: SPSS Pivot-Tabelle** und dort **Öffnen**. Die zu bearbeitende Tabelle ist jetzt in einem neuen Fenster: **SPSS-Pivot-Tabelle** dargestellt.



The screenshot shows a window titled "SPSS-Pivot-Tabelle - Tabelle1" with a menu bar containing "Datei", "Bearbeiten", "Ansicht", "Einfügen", "Pivot", "Format", and "Hilfe". The main area displays a pivot table with the following data:

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	weiblich	4	57,1	57,1	57,1
	männlich	2	28,6	28,6	85,7
	7	1	14,3	14,3	100,0
Gesamt		7	100,0	100,0	

Zusätzlich erhält man eine Formatierungsleiste



Es gibt nun verschiedenste Möglichkeiten, die Tabelle zu bearbeiten. Wir wollen hier:

- Zeilen und Spalten vertauschen
- einige Zeilen bzw. Spalten löschen,
- einige Elemente der Tabelle umformatieren.

Zeilen und Spalten vertauschen:

Dies ist nützlich wenn die Ausgabe sehr breit wird und man nicht mehr alles auf einem Bildschirm sieht, bzw. in einem Bericht die Seitenbreite nicht ausreicht. Vertauschen von Zeilen und Spalten geht ganz einfach über: **Pivot > Zeilen und Spalten vertauschen**. Man erhält

The screenshot shows a window titled "SPSS-Pivot-Tabelle - Tabelle1" with a menu bar (Datei, Bearbeiten, Ansicht, Einfügen, Pivot, Format, Hilfe). The main content is a table titled "Geschlecht".

	Geschlecht			
	Gültig			Gesamt
	weiblich	männlich	7	
Häufigkeit	4	2	1	7
Prozent	57,1	28,6	14,3	100,0
Gültige Prozente	57,1	28,6	14,3	100,0
Kumulierte Prozente	57,1	85,7	100,0	

Zeilen bzw. Spalten löschen:

Zeilen bzw. Spalten löscht man, in dem man die Inhalte löscht, dabei wird auch der Titel der Zeile bzw. Spalte mitgelöscht. Zunächst wollen wir die Spalten mit der Überschrift „7“ und „Gesamt“ beseitigen. Wenn man die **Inhalte dieser Spalten mit der Maus markiert** (auf linkes oberes Element – hier 1 – klicken und sie mit gedrückter linker Taste ins rechte untere Feld – hier kein Inhalt – ziehen) dann werden diese Zellen invertiert angezeigt. Löschen erfolgt über die Menüpunkte **Bearbeiten > Löschen**. Die beiden rechten Spalten sind jetzt verschwunden.

Zeilen löschen funktioniert analog. Löschen Sie bitte die beiden unteren Zeilen, so dass nur mehr die Zeilen „Häufigkeit“ und „Prozent“ überbleiben.

Geschlecht

	Geschlecht	
	weiblich	männlich
Häufigkeit	4	2
Prozent	57,1	28,6

Zellen formatieren:

Nun wollen wir noch ein bisschen Kosmetik betreiben. Die **Darstellung der Zahlen**, insbesondere wie viele Kommastellen angezeigt werden sollen, ändert man durch Markieren der zu ändernde(n) Zelle(n) und dann über **Format > Zelleneigenschaften... > Wert**. Dort kann man verschiedene Einstellungen vornehmen. Im wesentlichen entspricht dieses Dialogfenster ähnlichen in Windows üblichen Dialogfenstern zur Zahlenformatierung (wie z.B. in Excel). Wollen wir z.B. die Prozentangaben in unserer Tabelle ohne Dezimalstellen,

dann wählen wir dort als **Format**: den obersten Eintrag der Liste: "#,#" und setzen die **Dezimalstellen**: auf Null. Mit **OK** ist die Operation abgeschlossen. Diese Änderungsmöglichkeit kann man auch dann anwenden, wenn im SPSS Output eine Zahl in Exponentialdarstellung (z.B. 9,890E-02) ausgegeben wird und man eine ‚normale‘ Darstellung möchte (also z.B. ",0989).

Ersetzen von Text:

Das geht ganz einfach, in dem man auf das Textelement **doppelklickt**, den neuen Text darüberschreibt, und mit der Eingabetaste bestätigt.

Ändern von Spaltenbreiten:

Oft sind Beschriftung ist zu lang für die Breite einer Spalte. Wenn man die Maus auf eine vertikale Begrenzungslinie platziert, ändert sich der Mauszeiger zu ↔. Nun kann man die Linie nach links oder rechts ziehen, bis der Text in die Spalte passt.

Ändern der Schriftart, Größe und Farbe:

Solche Änderungen lassen sich leicht über die Formatierungssymbolleiste durchführen und funktionieren wie auch in anderen Windowsanwendungen (z.B. Word). Nach Markieren des gewünschten Elements führt man die Änderungen über die Symbolleiste durch. Hier kann man auch Spalteninhalte zentrieren oder rechts- bzw. linksbündig anordnen. (Wie bei anderen Symbolleisten erhält man eine kurze Beschreibung wenn man die Maus über dem Symbol platziert und ein wenig wartet).

Ändern der Zellenhintergrundfarbe:

In das entsprechende Dialogfenster kommt man über **Format > Zelleneigenschaften... > Schattierung** nachdem man die zu ändernde Zelle durch anklicken markiert hat. Unter **Hintergrund** kann man die gewünschte Farbe auswählen. Auch hier ist die Bedienung analog wie in anderen Windows Produkten (wie z.B. Word), im speziellen wenn man sich nicht mit den vordefinierten Farben begnügen will.

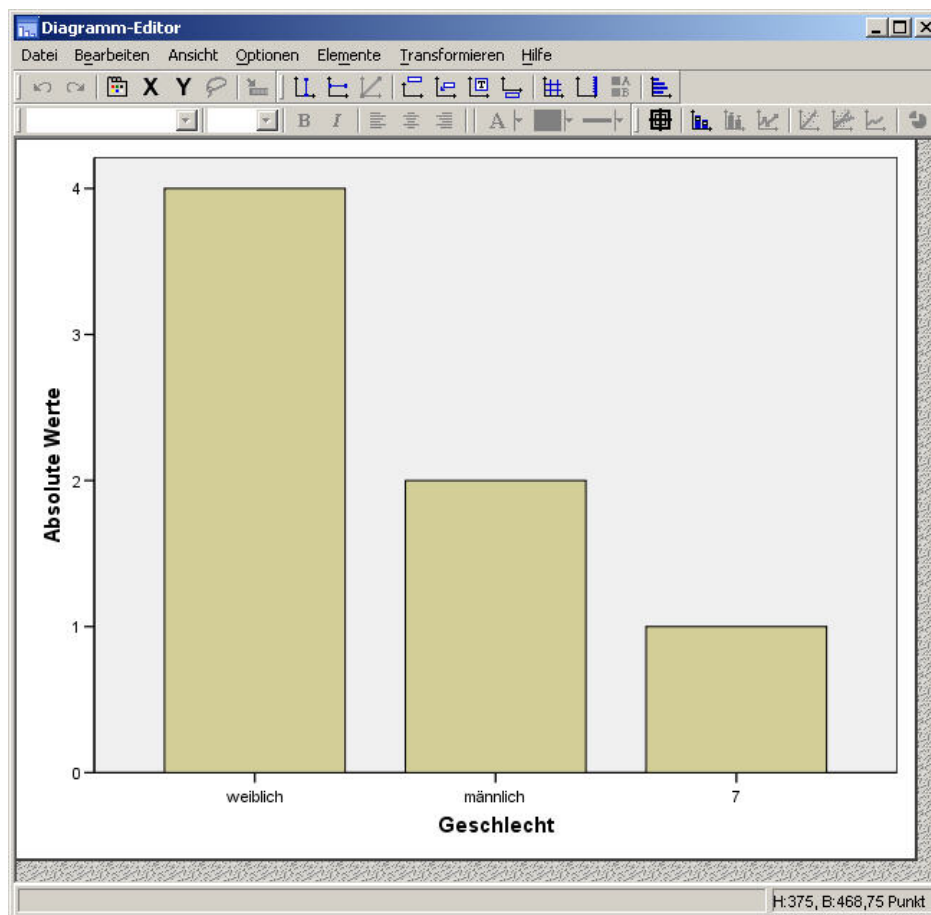
Abschließen der Tabellenmodifikationen

Hat man alle gewünschten Änderungen durchgeführt, schließt man einfach das Fenster: **SPSS-Pivot-Tabelle**. Das modifizierte Objekt ist nun in den Output-Viewer übernommen worden. Zu erwähnen wäre noch, dass Änderung nur beschränkt rückgängig gemacht werden können. Wenn man also mit seinen Modifikationen nicht zufrieden ist, muss man das ganze Objekt (Tabelle oder Grafik) neu erzeugen.

Modifikation von Grafiken

Ebenso wie Textoutput lässt sich grafischer Output auf vielfältige Weise nachbearbeiten. Dazu muss angemerkt werden, dass es zwei verschiedene Grafiktypen in SPSS gibt. Die (älteren) einfachen Diagramme und die (neueren, mit einigen wenigen Typen beginnend ab SPSS Version 8.0) interaktiven Diagramme. Der für den Anwender sichtbare Unterschied ist, dass bei interaktiven Grafiken auch noch im Nachhinein Variablen und Diagrammtypen geändert werden können und die Spezifikation ein wenig anders erfolgt. Dementsprechend unterschiedlich sind auch die Modifikationsmöglichkeiten, und besonders bei interaktiven Diagrammen sind die Möglichkeiten nahezu unüberschaubar. Wir beschränken uns hier deshalb nur auf ein paar wichtige Möglichkeiten.

Wie bei Tabellen gelangt man in einen Editor, in dem man **mit der rechten Maustaste** auf eine Grafik klickt und dort ganz unten **Objekt: SPSS Diagramm** bzw. **SPSS Interaktives Grafikobjekt** öffnet. Wir verwenden das vorher erzeugte Balkendiagramm und erhalten:



Die grundsätzliche Vorgehensweise für Modifikationen ist:

- (1) Markieren Sie durch Doppelklicken das Element der Grafik, das bearbeitet werden soll (z.B. Achsenbeschriftung, Linie, Fläche, Überschrift, Legende, ...)
- (2) Wählen Sie im sich öffnenden **Eigenschaften**-Fenster neue Einstellungen unter der jeweiligen Registerkarte und klicken Sie auf **Zuweisen**, um die neuen Einstellungen wirksam werden zu lassen

Alternativ kann man auch auf das zu ändernde Element rechts-klicken, und aus dem sich öffnenden Menü entweder **Eigenschaften** oder direkt eine Änderungsoption wählen

Exemplarisch soll gezeigt werden wie man Farben von Elementen verändert und wie Beschriftungen angepasst werden können:

Ändern der Farbe (von Grafikelementen und Beschriftungen/Texten):

Nach Doppelklicken auf irgendeinen Balken werden alle Balken umrandet dargestellt und es öffnet sich das **Eigenschaften**-Fenster. Zur Änderung der Farbe müssen wir die Registerkarte **Füllung & Rahmen** wählen.



Die Zuweisung einer Farbe erfolgt durch 1) Wählen einer Farbe in der Farbauswahl (mit **Bearbeiten** kann man wie bei Tabellen die Farbauswahlpalette ändern) und 2) **Zuweisen**. Sie sollten nun die Änderungen sehen. Wenn Sie weitere Elemente ändern wollen müssen Sie das **Eigenschaften**-Fenster nicht verlassen sondern können durch Anklicken eines anderen Grafikelements dieses modifizieren. Es ändern sich dann im **Eigenschaften**-Fenster je nach Element die zur Verfügung stehenden Registerkarten. Sie könnten nun z.B. analog die Schriftfarben der Achsenbeschriftungen ändern. Wenn Sie die gewünschten Modifikationen durchgeführt haben **Schließen** Sie das **Eigenschaften**-Fenster.



Verwenden Sie für gefüllte Grafikobjekte (wie Balken, Segmente in Kreisdiagrammen, etc.) nur sehr helle Farben bzw. helle Grautöne. Bei Schwarz-Weiß Ausdrucken werden dunkle, kräftige Farben wie z.B. Rot fast zu Schwarz und sehen dann sehr dominant und grafisch nicht sehr ansprechend aus und überlagern manchmal Grafikelemente wie den Median bei Boxplots (aber das ist Geschmackssache).

Ändern der Schriftgröße:

Geht analog zur Änderung von Farben, nur verwendet man jetzt die Registerkarte **Textstil**.

Ändern von Beschriftungen:

Zunächst Doppelklicken oder einfacher rechter Mausklick auf den zu ändernden Text. Der wird jetzt umrandet dargestellt (und das Eigenschaften-Fenster öffnet sich, das wir aber ignorieren). Nun nochmals ein einfacher Klick und der Cursor ist an der geklickten Stelle. Sie können jetzt den Text (oder die Zahl) ändern.

Ändern von Achsen (Einteilungen):

Man doppelklickt auf irgendein Element der zu modifizierenden Achse (entweder auf ein Zahl oder auf eine Beschriftung). Je nach Type der Achse kann man im Eigenschaften-Fenster **Kategorien** (bei kategorialen Einteilungen) oder **Skala** (bei numerische Einteilungen) ändern.

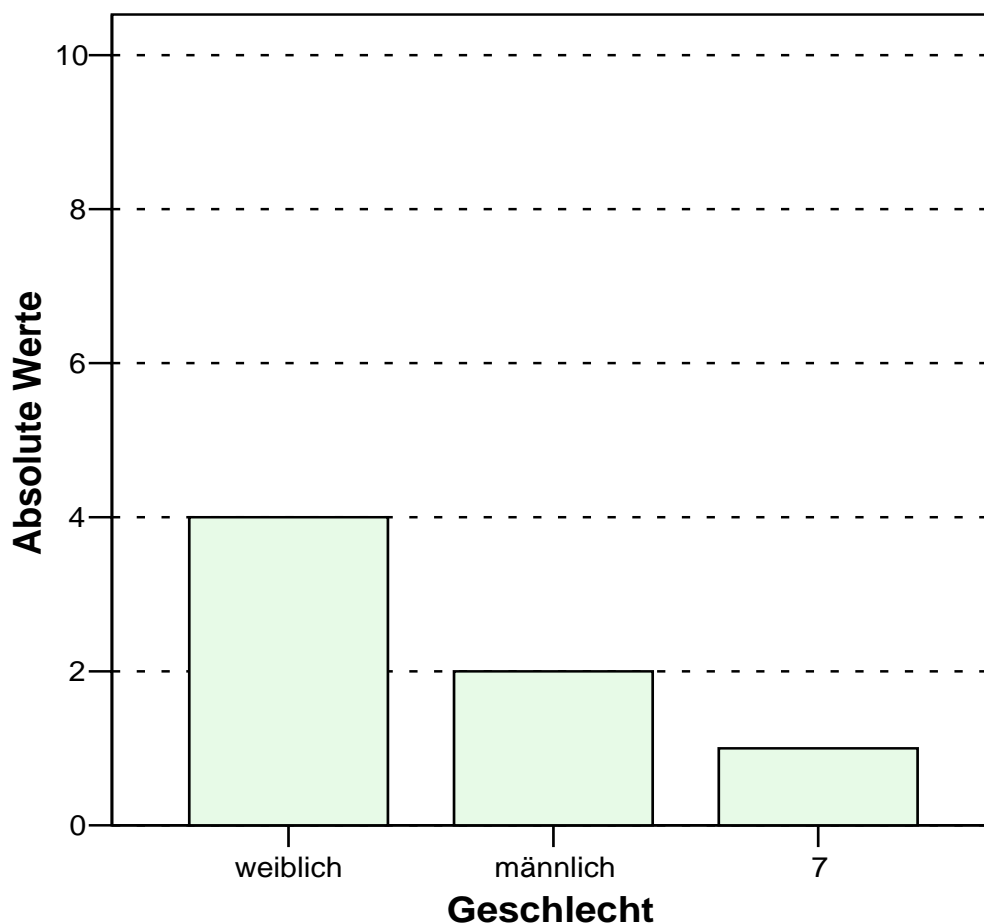
Als Beispiel wollen wir die Darstellung der Y-Achse („Absolute Werte“) und deren Einteilung ändern. Nach dem Doppelklick auf z.B. „Absolute Werte“ kann man unter **Skala** den dargestellten Bereich (**Minimum** bis **Maximum**) sowie die Unterteilung (**Erste Unterteilung**) modifizieren. Dazu muss man zunächst die Markierungen unter **Auto** wegklicken.

Die Balken sollen in einem Bereich von 0 bis 10 dargestellt werden, dazu ändert man den Wert für **Minimum** und **Maximum** (die daneben unter **Daten** angezeigten Werte sind die tatsächlichen Werte aus den Rohdaten, was hier aber leider nicht immer stimmt). Zusätzlich ändern wir noch den Wert für **Erste Unterteilung** auf **2** (die Unterteilung läuft dann in

Zweier-Schritten). Wenn man die Darstellung der Zahlen (Anzahl der Dezimalstellen) ändern will, dann sollte man nicht auf den Achsentitel sondern auf irgendeinen Beschriftungszahl klicken, weil nur dann die Registerkarte **Zahlenformat** im Eigenschaften-Fenster zur Verfügung steht.

Schließlich sollen noch Hilfslinien für die Y-Achse gezeichnet werden. Dazu wählt man im Dateneditor aus dem Menü **Optionen** den Menüpunkt **Gitterlinien einblenden**. Die Hilfslinien sollen strichliert dargestellt werden. Dazu klickt man auf irgendeine Linie, um sie zu markieren. Im Eigenschaften-Fenster wählt man **Linien** und führt dort die gewünschten Änderungen durch.

Die Grafik sieht nach all diesen Schritten so aus (zusätzlich wurden noch die Schriftgrößen geändert):



Abschließen der Grafikmodifikationen

Wie bei Tabellen schließt man nach allen gewünschten Änderungen einfach das Fenster: **Diagramm Editor** und hat damit das modifizierte Objekt in den Output-Viewer übernommen.

(Wenn Sie den Diagramm-Editor nicht schließen, bleibt das Objekt im Output-Viewer dunkel schraffiert. Die Schraffierung ist ein Hinweis, dass das Objekt gerade editiert wird.) Natürlich kann man den Diagramm-Editor mehrmals aufrufen um zusätzliche Änderungen vorzunehmen, da man nicht immer mit einer ersten Nachbearbeitung zufrieden ist.

Übernahme von Tabellen und Grafiken in Textverarbeitungs- bzw. Präsentationssoftware bzw. Exportieren des Outputs

In praktischen Anwendungen wird es wohl am häufigsten vorkommen, dass man den (modifizierten) Output von SPSS in ein anderes Programm kopieren will, sei es bei der Erstellung eines Berichts oder einer Präsentation.

Kopieren mit „cut and paste“

Bei Verwendung von Textverarbeitungsprogrammen (wie z.B. Word) oder Präsentationssoftware (wie z.B. Powerpoint) kann man sowohl Tabellen wie auch Grafiken sehr einfach übernehmen, wobei ein kleiner Unterschied zwischen Tabellen und Grafiken besteht.

Diagramme sollte man so kopieren, dass man im SPSS Output-Viewer **mit der rechten Maustaste** auf die jeweilige Grafik klickt und dort **Objekt kopieren** verwendet. Man kann dann noch Word-Formatierungen für Grafiken verwenden (Größe, Layout, etc.). In der Zielanwendung (Word oder Powerpoint) können sie mit **STRG-V** bzw. mit der **rechten Maustaste > Einfügen** die Grafik an der entsprechenden Stelle einfügen.

Bei **Tabellen** kommt es darauf an, wie sie in der Zielanwendung verwendet werden soll. Verwendet man (nach dem rechten Mausklick auf die Tabelle) **Objekt kopieren**, dann erhält man nach dem Einfügen eine grafische Repräsentation der Tabelle, d.h. sie sieht so aus wie im SPSS Output-Viewer, man kann aber abgesehen von der Größe nichts mehr ändern. Verwendet man **Kopieren**, dann bekommt man eine RTF Version der Tabelle, d.h. sie sieht so aus wie eine Tabelle in Word oder Powerpoint. Man hat dann die Formatierung verloren, kann aber die Inhalte noch ändern.

Export in externe Dateien (Text, HTML bzw. Grafik)

Sowohl Tabellen wie auch Grafiken lassen sich als weiterverwendbare Dateien speichern. Für Texte und Tabellen stehen sowohl das HTML wie auch Textformat zur Verfügung. Grafiken können in gebräuchlichen Formaten (wie Windows-Metadateien, Windows-Bitmap, Encapsulated PostScript, JPEG, TIFF, CGM, PNG und Macintosh-PICT) gespeichert werden.

In das Export Dialogfenster kommt man im Output-Viewer über die Menüpunkte **Datei > Exportieren...** oder mit einem **rechten Mausklick** auf das zu exportierende Element und dann ebenfalls **Exportieren...** Alle gewünschten Optionen (wie Dateiname, welche Elemente, welches Format, etc.) lassen sich dann dort definieren.

Der Export in eine Grafikdatei ist eine alternative Möglichkeit, SPSS Output in anderen Programmen zu verwenden.



Außer beim Export in ein vektorgrafikbasiertes Format, lassen sich die einmal kopierten Objekte nicht mehr (oder nur sehr mühsam) verändern. Führen Sie also besser alle Modifikationen und Nachbearbeitungen in SPSS selbst aus und kopieren Sie die Tabelle oder das Diagramm erst dann, wenn Sie mit dem Resultat zufrieden sind.

9 Einlesen externer Daten

Manchmal kommt es vor, dass man Daten, die nicht im SPSS Format gespeichert sind, wie z.B. Excel-Dateien oder Text-Dateien, in SPSS einlesen möchte. Im wesentlichen gibt es 4 Typen externer Dateiformate, die SPSS lesen kann – andere SPSS Versionen (Macintosh, Unix, DOS) und andere Statistikpakete (SYSTAT), Datenbanksysteme (diverse SQL, Access, Dbase), Spreadsheets (Excel, Lotus, SYLK) sowie einfache Textdateien (mit fixen und variablen Spaltenbreiten, CSV).

Wir wollen hier die beiden in der Praxis wohl am häufigsten auftretenden Situationen behandeln: Excel-Dateien und reine Text-Dateien (Ascii-Dateien). Für den Import aus Datenbanken (hier steht ein eigener Datenbank-Assistent zur Verfügung) sowie anderen Statistikpaketen sei auf die SPSS Hilfe verwiesen (**Hilfe > Lernprogramm > Einlesen von Daten**), wo auch die folgenden beschriebenen Möglichkeiten detaillierter behandelt werden.

Import von Excel Dateien:

Um Excel-Dateien einzulesen, wechseln Sie ins Untermenü **Datei > Öffnen > Daten**. Wählen Sie aus der Dropdown-Liste unter Dateityp: **Excel (*.xls)** aus. Wählen Sie nun eine **Datei** aus der Liste aus. Wenn die erste Zeile der Excel-Datei Spaltenüberschriften bzw. Labels enthält, klicken Sie auf **Variablenamen aus ersten Dateizeile** lesen. Bei Dateien aus Excel 5 oder nachfolgenden Versionen können Sie außerdem bestimmen, aus welchem Arbeitsblatt Sie Daten übernehmen wollen.

Die Spaltenüberschriften aus der Excel-Datei werden als Variablenamen verwendet. Wenn eine Änderung erforderlich ist (z.B. wegen Leerzeichen zwischen Wörtern), wird der Name entsprechend angepasst, die ursprüngliche Spaltenüberschrift bleibt als Variablenlabel erhalten. Wenn die Zelle in der Spaltenüberschrift leer ist, wird ein *Default-Wert* (z.B. v2) zugewiesen.



Achten Sie darauf, dass nur Daten eines bestimmten Typs (am besten nur numerisch) in einer Spalte vorkommen. Wenn eine Spalte in einer Datei aus Excel 5 oder nachfolgenden Versionen Daten mit verschiedenen Typen enthält (beispielsweise sowohl numerische als auch String-Daten), werden die Datenwerte in den Spalten als String-Daten eingelesen. Bei Dateien aus Excel 4 oder Vorgängerversionen wird Spalten mit gemischten Datentypen der Datentyp des ersten Datenwerts zugewiesen. Daten mit anderem Typ werden als systemdefiniert fehlend (oder bei String-Variablen als leer) ausgewiesen.

Import von Text Dateien:

Für den Import von Daten, die in einem einfachen Text-Format gespeichert sind, steht in SPSS der Assistent für Textimport zur Verfügung. Der Aufruf erfolgt über: **Datei > Textdaten einlesen**. Hier wählen Sie die Datei, die den Dateityp: Text (*.txt) hat, aus. Zur Illustration verwenden wir die Daten der ersten 5 Variablen des schon vorher verwendeten Fragebogens. Der Inhalt einer Datei mit Namen `übungsfragebogen_5var.txt`, könnte so aussehen (hier nur die ersten 3 Zeilen):

```
ID, SEX, LIEBALT, GROESSE, MONATE
109, 1, 2, 173, 266
16, 1, 3, 166, 241
...
```

Hat man mit **Öffnen** die Datei gewählt, startet der Assistent für Textimport. Hier werden jetzt 6 Schritte durchlaufen, in denen alles Notwendige festgelegt wird. Der Assistent macht hierbei (mehr oder weniger intelligente) Vorschläge.

Schritt 1:

Die Datendatei wird in einem Vorschauenfenster angezeigt. Sie könnten hier noch ein vordefiniertes (früher im Text-Assistenten gespeichertes) Format festlegen, was für uns aber nicht gilt, da wir ein neues Format definieren werden.

Schritt 2:

Hier wird definiert, in welcher Form die Daten in der Textdatei gespeichert sind. Wichtig ist die Unterscheidung **Festes Format** und **Variables Format**. Festes Format bedeutet, dass jede Variable in der gleichen Spaltenposition für jeden Fall aufgezeichnet wird. Variables Format bzw. **Mit Trennzeichen** bedeutet, dass Leerzeichen, Kommata, Tabulatoren oder andere Zeichen zum Trennen der Variablen eingesetzt werden. Die Variablen müssen für jede Beobachtungseinheit (Fall) in derselben Reihenfolge gespeichert sein, befinden sich aber nicht notwendigerweise an derselben Spaltenposition. In unserem Beispiel ist die Datendatei mit Trennzeichen, nämlich Komma, versehen. Weiters ist anzugeben, ob die erste Zeile die Variablennamen enthält. Wenn das nicht der Fall ist, kann man in Schritt 5 Namen vergeben.

Schritt 3:

In Schritt 3 werden die Informationen zu den Zeilen (Beobachtungseinheiten) definiert. Im vorhergehenden Schritt wurde entschieden, ob die erste Zeile der Datei die Variablennamen enthält. Falls ja, dann geben Sie in diesem Schritt an, dass die Datenwerte in der zweiten Zeile beginnen, sonst beginnen Sie in der ersten. Außerdem legen Sie hier fest, wie viele Variablen für jeden Fall eingelesen werden müssen. Das ist vor allem wichtig, wenn es mehrere Zeile pro Beobachtungseinheit gibt. In unserem Beispiel sind alle Datenwerte für einen Fall in einer Zeile aufgezeichnet (Achtung: Wenn nicht alle Zeilen die gleiche Anzahl von Werten enthalten, schlägt SPSS hier als Anzahl der Variablen die Zahl vor, die sich aus der Zeile mit den meisten Datenwerten ergibt.). Weiters kann man hier entscheiden wie viele Fälle eingelesen werden sollen oder ob man eine bestimmte Anzahl Zeilen, oder eine Zufallsstichprobe importieren möchte.

Schritt 4:

In Schritt 4 zeigt der Text-Assistent einen Vorschlag an, wie die Variablen aus der Datendatei eingelesen werden. Sie können diesen Vorschlag ändern. In unserem Beispiel würde der Text-Assistent richtig erkennen, dass Kommas zum Trennen der Datenwerte eingesetzt wurden. Aufeinanderfolgende Trennzeichen würden fehlende Daten kennzeichnen (bei z.B. durch Kommas getrennten Werten bedeutet dies, dass zwei Kommas, die keinen Datenwert umschließen, eine Variable mit einem fehlenden Wert darstellen). Es können hier auch mehrere Zeichen als Trennzeichen definiert werden (z.B. Komma und Leerzeichen).

Schritt 5:

In Schritt 5 legen Sie fest, welches Datenformat der Text-Assistent beim Einlesen für die Variablen verwendet – numerisch, String, etc. Die Festlegungen hier entsprechen den Definitionen, wie sie auch im Abschnitt 2 bei der Beschreibung der Variablenansicht als Variablentyp behandelt wurden. Um das vom Textassistenten vorgeschlagene Format einer Variable zu ändern, klicken Sie in der **Datenvorschau** auf den Variablennamen (erste Zeile) und wählen aus der Dropdown-Liste bei **Datenformat**: dann ein Format (z.B. **Numerisch**)

aus. Man kann im fünften Schritt auch noch entscheiden, welche Variablen in der endgültigen Datendatei enthalten sein werden. Um eine Variable aus der importierten Datendatei auszuschließen, klicken Sie wieder in der **Datenvorschau** auf den Variablennamen und wählen aus der Dropdown-Liste die Option **Nicht importieren** aus.

Schritt 6:

Im letzten Schritt können Ihre Angaben speichern, um eventuell vergleichbare Textdatendateien einlesen zu können. Sie können auch die zugrundeliegende Befehlssyntax einfügen und speichern (siehe dazu auch den nächsten Abschnitt). Wenn Sie mit dem Einlesen der Textdatendatei beginnen möchten, klicken Sie auf **Fertigstellen**.

Abschließende Anmerkung:



Ganz allgemein gilt: Wenn Sie Daten aus anderen Anwendungen übernommen haben, sollten Sie in der Datenansicht zunächst nachsehen, ob alles auch so funktioniert hat wie Sie sich das vorgestellt haben. Wichtig ist es, in der Variablenansicht sicherheitshalber noch einmal zu kontrollieren, ob alle Variablendefinitionen akzeptabel sind. Dies betrifft besonders die letzte Spalte Messniveau !

10 Verwenden der Programmsprache („Syntax“)

Die SPSS Kommandosprache ist das, was im Hintergrund in SPSS abläuft, wenn Sie über Menüs und Dialogfenster bestimmte Arbeitsschritte durchführen. Tatsächlich war die Verwendung dieser Befehle früher, vor der Version 6, die einzige Möglichkeit SPSS zu verwenden. Obwohl man meistens mit der gewohnten Oberfläche auskommt (Ausnahmen werden unten beschrieben) kann es doch ganz nützlich sein, ein paar grundlegende Aspekte zu kennen und sie gegebenenfalls auch anwenden zu können.

Im wesentlichen besteht SPSS aus 2 Komponenten:

Benutzeroberfläche: Hier geben wir unsere Anweisungen ein und bekommen die Ergebnisse bereitgestellt. Wir können der Benutzerschnittstelle unsere Anweisungen in Form von ausgefüllten Dialogboxen oder als Folge von SPSS-Kommandos übergeben.

SPSS-Prozessor: Die Benutzerschnittstelle gibt unsere Anweisungen (indirekt oder direkt) in Form von SPSS-Kommandos an den Prozessor weiter, der im Hintergrund arbeitet. Wir erfahren übrigens in der Statuszeile der SPSS-Fenster (ganz unten), was der Prozessor gerade treibt. Wenn wir ein ausgefülltes Dialogfenster mit **OK** schließen, führt der SPSS-Prozessor also im Hintergrund das entsprechende Kommando aus.

Das Syntaxfenster

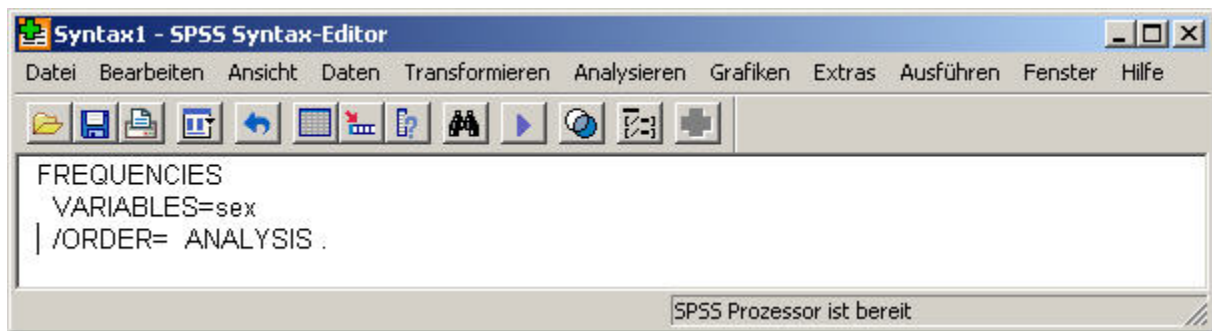
In fast allen SPSS-Dialogboxen können über die Standardschaltfläche **Einfügen** die zugrunde liegenden SPSS-Kommandos produziert werden. Diese werden dann *nicht* ausgeführt, sondern in ein sogenanntes **Syntaxfenster** übertragen, das weitgehend analog zu einem Texteditor funktioniert. Hier können Sie beliebige Kommandos zu einer Sequenz ansammeln, nach Bedarf einzeln oder geschlossen ausführen lassen und schließlich in einer Datei abspeichern. Später können Sie die Kommandos aus dieser Datei wieder laden und, eventuell nach manueller Überarbeitung, erneut ausführen lassen.

Als Beispiel wollen wir nochmals die Prozedur Häufigkeiten verwenden, die wir schon im Abschnitt 7 (Analysieren von Daten) behandelt haben.

Über **Analysieren > deskriptive Statistiken > Häufigkeiten...** kommt man in das Dialogfenster Häufigkeiten, wo schon **sex** als zu beschreibende Variable definiert ist.

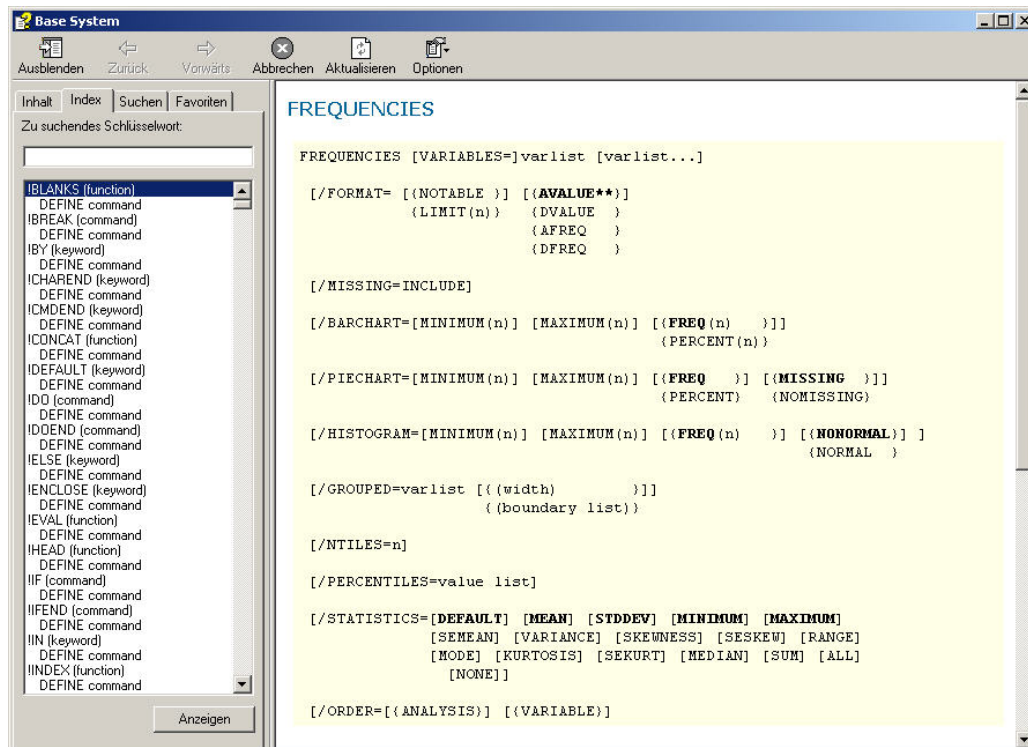


Wenn man jetzt auf **Einfügen** klickt, öffnet sich das sogenannte Syntaxfenster,



in dem das Kommando **FREQUENCIES** (für Häufigkeiten), so wie wir es spezifiziert haben und wie es im Hintergrund ablaufen würde, eingefügt wurde. Das ist das gleiche Kommando, wie wir es im Abschnitt 7 ausgeführt haben. Hierbei ist **FREQUENCIES** das eigentliche Kommando, mit den beiden Subkommandos **VARIABLES** und **ORDER**, getrennt durch **/** und, wie bei allen Kommandos, abgeschlossen mit einem Punkt. Im **VARIABLES** Subkommando werden die Variablen angeführt, für die wir die Häufigkeiten wollen. (Das **ORDER** Subkommando ist hier eigentlich unnötig, mit dem Schlüsselwort **ANALYSIS** wird nur festgelegt, dass im Output zuerst alle Tabellen, und nachher alle Grafiken dargestellt sein sollen.)

Alle möglichen Subkommandos (und das sind mehr als man über die Menüsteuerung angeben könnte) erhält man durch Klick auf



Ausführen eines „Syntax“ Kommandos

Hat man nun ein oder mehrere Kommandos im SPSS Syntax-Editor, kann man diese von SPSS abarbeiten lassen, genauso wie wenn man in einem Dialogfenster auf OK geklickt. Hierzu gibt es mehrere Möglichkeiten.

Die einfachste ist es, jenes Kommando mit der Maus zu markieren, das man ausführen will. Dann klickt man auf



Alternativ dazu gibt es den Menüpunkt **Ausführen**. In jedem Fall erhält man im Output-Viewer die Resultate genauso, als hätte man sich durch die entsprechenden Menüs und Dialogfenster geklickt.

Händische Eingabe von „Syntax“ Kommandos

Neben der eben beschriebenen Möglichkeit, durch **Einfügen** aus einem Prozedur Dialogfenster das jeweilige Kommando im Syntax Editor zu erhalten, kann man natürlich auch selbst die entsprechenden Kommandos eingeben.

Ein leeres Syntaxfenster erhält man über **Datei > Neu > Syntax**. (Das erste Syntaxfenster wird üblicherweise mit Syntax1 bezeichnet). Hier kann man nun Befehle eingeben und ausführen, analog zum oben Beschriebenen.

Je nach Installation ist der sogenannte Command Syntax Reference über die SPSS Hilfe verfügbar, ansonsten gibt es ihn auf der SPSS CD im PDF Format. Dort findet man ausführliche Beschreibungen aller verfügbaren Kommandos und Prozeduren, und wie man mithilfe der SPSS Syntax auch automatisierte Berechnungen, Analysen und Reports erstellen kann.

Speichern und Einlesen von „Syntax“ Dateien

Wie Daten- bzw. Outputdateien können auch Syntaxdateien abgespeichert bzw. wieder eingelesen werden. Beides erfolgt über den Menüpunkt **Datei** und dort unter **Öffnen** bzw. **Speicher** oder **Speichern unter...** Die Standard-Dateierweiterung einer Syntaxdatei ist **.SPS**.

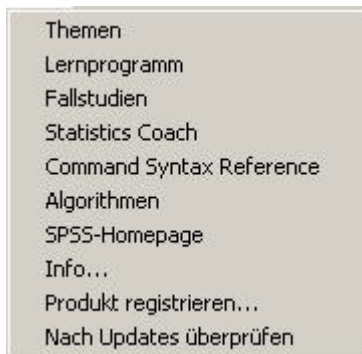
Automatisches Protokollieren der Arbeit in SPSS

Eine sehr nützliche Verwendungsmöglichkeit der „Syntax“ ist es, Protokolle seiner Arbeit automatisch anfertigen zu lassen. Dazu gibt es zwei Möglichkeiten:

1. Man kann zu jedem Output, den man erzeugt, das zugehörige SPSS Kommando dazufügen lassen. Das kann man unter **Bearbeiten > Optionen > Viewer** einstellen, wenn man links unten das Feld: **Befehle im Log anzeigen** markiert.
2. Alles, was man im Verlauf einer SPSS Session tut, kann auch in eine Datei in Form der Syntaxkommandos abgespeichert werden. Das kann man unter **Bearbeiten > Optionen > Allgemein** festlegen, wo man bei **Sitzungsjournal** (links) verschiedene Einstellungen treffen kann (wie Dateiname des Protokolls, ob die Protokollierung jedes Mal neu angelegt wird, etc.). Die Standard-Dateierweiterung einer solchen Protokoll- oder Journaldatei ist **.JNL**

11 Hilfe zu einzelnen Prozeduren, Tutorials und Links

Unter **Hilfe** findet man in SPSS folgende Untermenüpunkte:



- Die SPSS Hilfsfunktionen, die man in jedem SPSS Fenster über **Hilfe > Themen** erreicht, bzw. die als **Schaltflächen** in nahezu jedem **Dialogfenster** verfügbar sind, sind sehr ausführlich und meistens hilfreich. Sie können immer auf die Registerkarten **Inhalt**, **Index** und **Suchen** zugreifen und finden dann allgemeine Informationen und Verknüpfungen zu verwandten Themen.
- Das **Lernprogramm** gibt zu den wichtigsten Funktionen von SPSS kurze, ablaufartige Darstellungen, und ist auch für einen ersten Einstieg gut geeignet.
- Auf der **SPSS-Homepage** (<http://www.spss.com>) muss man sich erst zurechtfinden, es gibt aber Tutorials zu speziellen Inhalten, natürlich einen Downloadbereich, mit einer Demoversion der jeweils aktuellsten SPSS Version und vieles mehr. Ein Besuch dieser Seite lohnt sich immer, für manche Inhalte muss man sich allerdings registrieren (was aber kein Problem ist, da man dort festlegen kann, ob man Werbung etc. von SPSS geschickt bekommen möchte oder nicht).
- Ob die **Command Syntax Reference** (sehr komplett und umfangreich !) direkt verfügbar ist, hängt von der jeweiligen Installation ab. Jedenfalls ist sie auf der SPSS CD verfügbar.
- Der **Statistics Coach** ist der eher missglückte Versuch, einen Laien zur richtigen Analysemethode zu geleiten. Nicht zu empfehlen.

Zusätzlich zum Hilfemenü gibt es:

- **Hilfe zu Dialogfeldern über das Kontextmenü.** Wenn Sie mit der rechten Maustaste auf ein Steuerelement in einem Dialogfeld klicken, erhalten Sie eine Beschreibung des Steuerelements und seiner Verwendung.
- **Hilfe zu Tabellen über das Kontextmenü.** Wenn Sie mit der rechten Maustaste auf Begriffe einer im Viewer aktivierten Tabelle klicken und dann **Direkthilfe** aus dem Kontextmenü auswählen, erhalten Sie eine Definition dieser Begriffe.
- Wenn Sie mit der rechten Maustaste auf eine Tabelle klicken und dann **Ergebnis-Assistent** auswählen, erhalten Sie Informationen, die Sie bei der Interpretation der Ergebnisse unterstützt.
- Ein ausgezeichnetes **Tutorial**, nicht nur zur SPSS Bedienung sondern auch mit einfach verständlichen Beschreibungen wichtiger Statistikprozeduren findet man auf der SPSS CD. Online gibt es eine Kopie unter <http://www.wu-wien.ac.at/inst/stat1/hatz/spss/tutorial/>