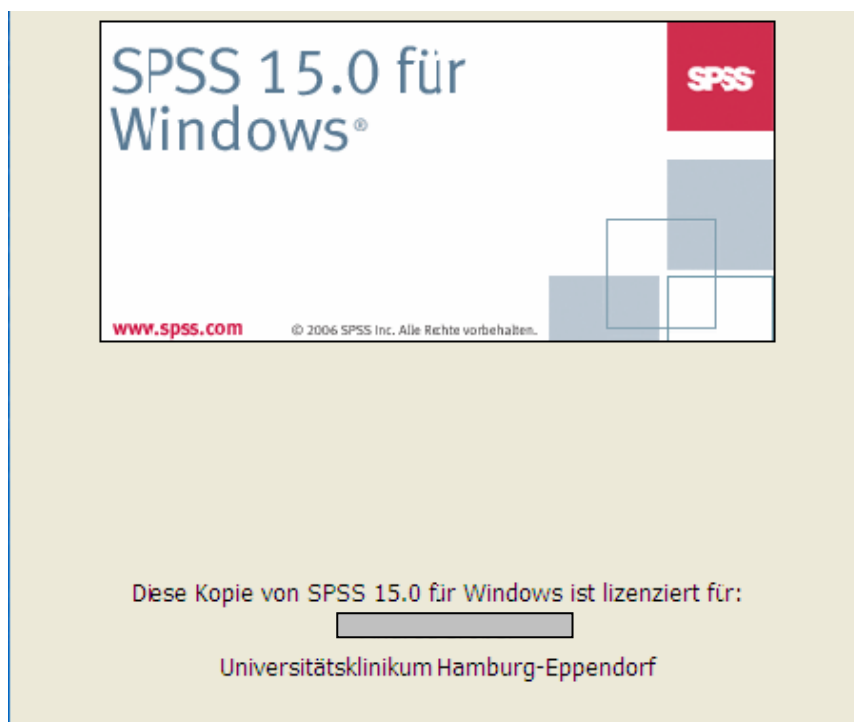




Universitätsklinikum  
Hamburg-Eppendorf

# Der SPSS- (Wieder-) Einstieg leicht gemacht

Eine Einführung in die statistische  
Datenauswertung mit



Universitätsklinikum Hamburg-Eppendorf

Institut für Medizinische Biometrie und Epidemiologie (IMBE)

**SEMINAR:****Einführung in die Anwendung statistischer Software (SPSS unter Windows)**<http://www.uke.uni-hamburg.de/imbe>

Anmeldung: grosser@uke.uni-hamburg.de (Andrea Großer)

# Inhaltsverzeichnis

1. Überblick.....	3
1.1 Allgemeines.....	3
1.2 Der erste Eindruck.....	4
2. Datenübermittlung an SPSS .....	8
2.1 Datenerfassung direkt in SPSS .....	8
2.2 Datenimport.....	10
3. Datenweiterverarbeitung .....	11
3.1 Vergabe von Labels in der Variablenansicht .....	11
3.2 Neuberechnung von Variablen .....	14
4. Datenauswertung .....	17
4.1 Statistischer Vergleich mittels geeigneter Maßzahlen.....	18
4.2 Das SPSS-Ausgabefenster .....	20
4.3 Grafische Darstellung mittels Box-Plot .....	22
5. Zusammenfassung und Ausblick.....	24
SPSS-Syntaxfenster.....	24
Literaturverzeichnis .....	25

# 1. Überblick

## 1.1 Allgemeines

Der Name SPSS stand ursprünglich (Gründung 1968) für **Statistical Package in the Social Sciences**. Heute wird SPSS [**S**uperior **P**erforming **S**oftware **S**ystems] jedoch in sehr vielen Bereichen von Industrie und Wirtschaft und speziell auch in der Medizin zur effizienten und anwenderfreundlichen Datenauswertung genutzt.

Am UKE können Sie nach Voranmeldung im Doktorandenraum im IMBE (S10) SPSS nutzen. Des Weiteren steht zu bestimmten Öffnungszeiten ein Computerpool in Raum 9 und 10 des Rothe-Geussenhainer-Hauses für Ihre Datenauswertung zur Verfügung. Dazu ist dort lediglich ein Einloggen unter dem Benutzernamen „spss“ erforderlich, die Angabe eines Passwortes ist nicht nötig. Die Belegungszeiten erfahren Sie unter <http://bvm.uke.uni-hamburg.de/pclab/week.php?year=2008&month=01&day=30>

Der in diesem Skript als Beispiel verwendete Datensatz steht im dortigen Pool im Verzeichnis „SPSS auf Cip01“ unter **Z:\Seminar\stammdat.sav** zur Verfügung, so dass die hier exemplarisch vorgeführten Auswertungen bei Bedarf direkt am Rechner nachvollzogen können.

Nachfolgend wird in komprimierter Form eine Einführung in die Benutzung der Software gegeben, wobei von Grundkenntnissen im Umgang mit Microsoft Windows ausgegangen wird. Bei weiterführendem Interesse sei hier auch auf den Einführungskurs in SPSS hingewiesen (genaue Termine und Zeiten siehe unter <http://www.uke.uni-hamburg.de/imbe>).

Zur sinnvollen Anwendung statistischer Software ist es natürlich ebenso unerlässlich, die zugrunde liegenden Verfahren und Methoden zu verstehen. Hierfür sei auf die entsprechenden Kurse des IMBE in Biometrie hingewiesen. Außerdem ist die Webpage JUMBO Java-unterstützte Münsteraner Biometrie-Oberfläche zu empfehlen:

<http://imib.uni-muenster.de/fileadmin/template/conf/imib/lehre/skripte/biomathe/jumbo.html>

Eine Lokale Version JUMBO 6.8 mit PDF-Dokumentation ist auch auf CD erhältlich. Auf dieser CD sind außerdem der Vorlesungsskript, das Glossar, die Benutzerdokumentation der Applets und die MC-Fragen und -Antworten enthalten (bei Interesse Email an Prof. Dr. rer. pol. **Wolfgang Köpcke**: [kopcke@uni-muenster.de](mailto:kopcke@uni-muenster.de)). Das Vorlesungsskript steht zum Download (pdf) zur Verfügung..

SPSS selbst bietet Online-Seminare zu verschiedenen praktischen Fragestellungen an:

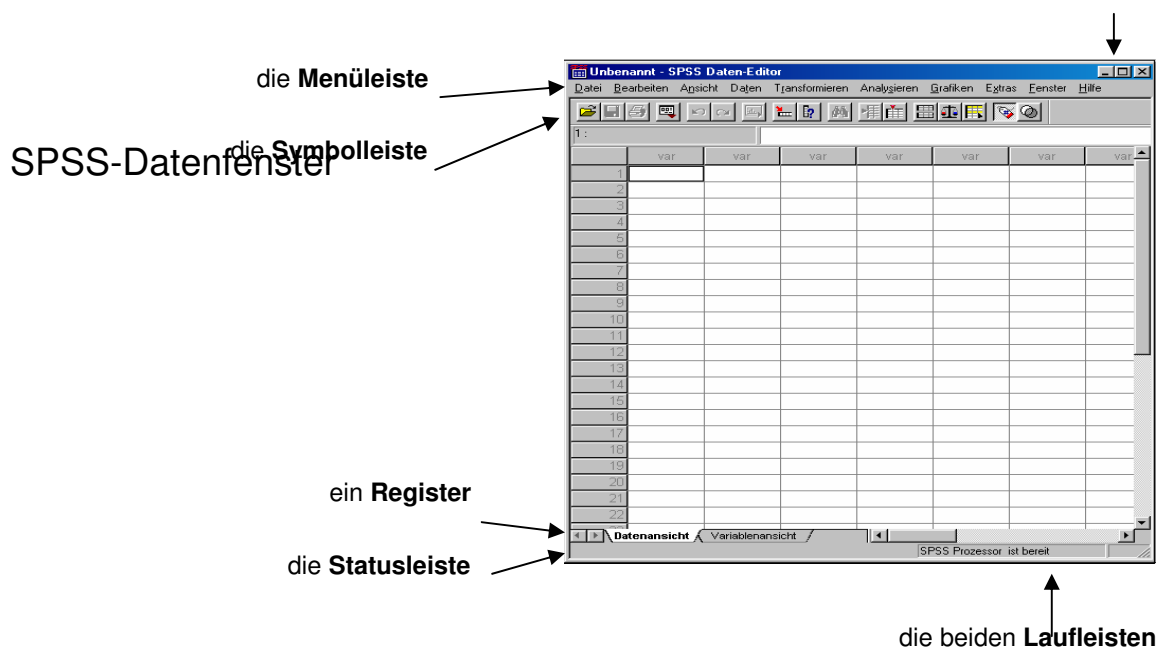
<https://spssevents.webex.com/spssevents>

## 1.2 Der erste Eindruck

Hat man SPSS wie für alle Windows-Applikationen üblich mittels **START – PROGRAMME – SPSS für windows** gestartet, so öffnet sich das sog. **SPSS-Datenfenster** in der Datenansicht.

START – PROGRAMME – SPSS für windows

die Fenstersteuerung



Der Großteil des Bildschirms wird dabei von den einzelnen Zellen bestimmt, in welche sich Daten einlesen lassen. Daneben finden Sie viele, auch aus anderen Windows-Anwendungen bekannte Elemente, so z. B.

- die **Fenstersteuerung** rechts oben

- die beiden **Laufleisten**

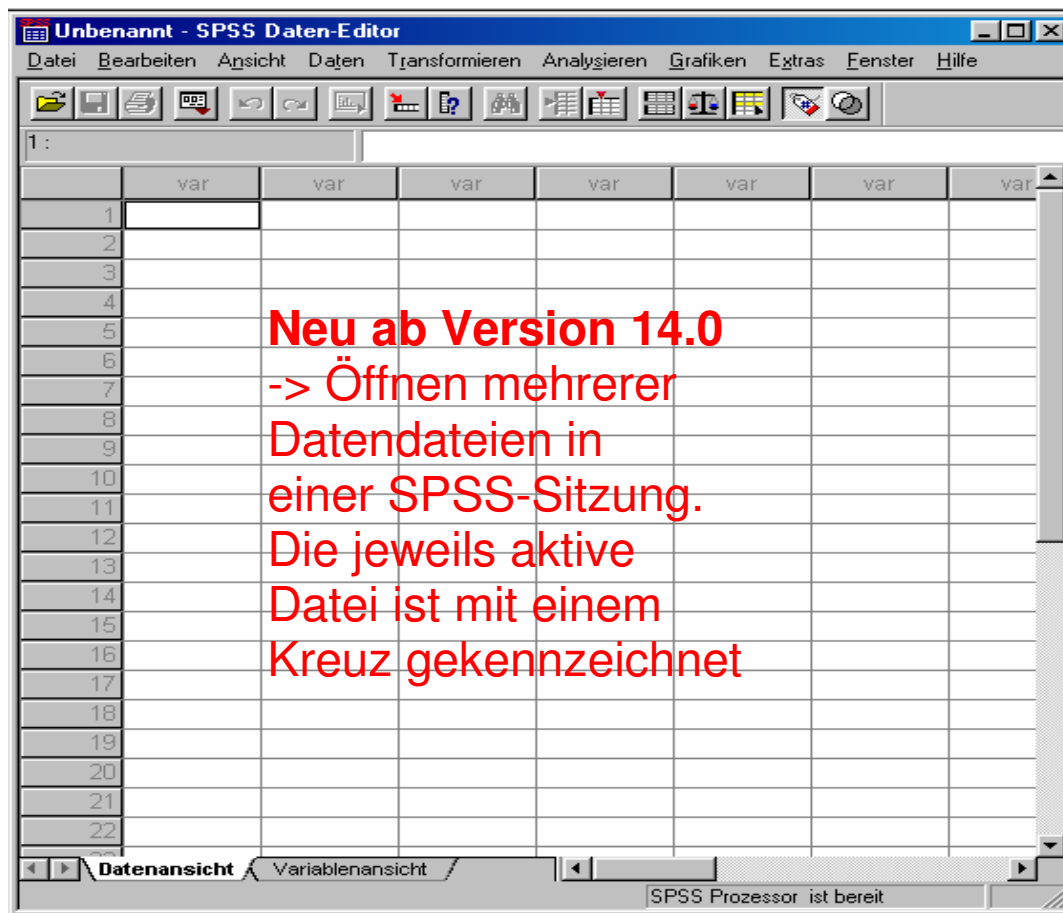
- die **Symbolleiste**

- die **Statusleiste**, welche den aktuellen Status des Programms wiedergibt

- ein **Register**, mittels dessen man zwischen Daten- und Variablenansicht wechseln kann



- Am häufigsten findet sicherlich die sog. **Menüleiste** mit den thematisch geordneten Befehlen Anwendung:



Unter **Datei** finden Sie alles was mit Datenimport und -export zu tun hat, d. h., Sie können Dateien einlesen und als SPSS-Datei in ein Verzeichnis abspeichern. Eine SPSS-Datendatei besitzt dabei immer die charakteristische Endung **.sav**.

Diese Befehle werden hier im Einzelnen kurz angesprochen und in den folgenden praktischen Übungen eingesetzt:

Der Menüpunkt **Bearbeiten** stellt die „üblichen“ Funktionen wie z. B. **kopieren, Ausschneiden, Einfügen, Rückgängig** etc. zur Verfügung.

Unter **Ansicht** kann die optische Gestaltung des Bildschirms persönlichen Präferenzen angepasst werden.

Das Herzstück von SPSS bilden die vier Menüpunkte **Daten, Transformieren, Analysieren und Grafik** mit ihren zahlreichen Unterbefehlen.

Klicken Sie auf **Daten**, so erhalten Sie eine Vielzahl von Möglichkeiten, Daten zu filtern, umzustrukturieren etc. Als die beiden wichtigsten Optionen seien hier genannt

- ⇒ bestimmte Fälle (also in der Regel Patienten) gesondert auszuwählen  
(Befehl **Fälle Auswählen**)
- ⇒ eine Analyse nach Untergruppen zu trennen, also z. B. geschlechtsspezifisch durchzuführen (Befehl **Datei aufteilen**).  
Ein praktisches Beispiel hierzu finden Sie auf S.17f. im Text.

Der Menüpunkt **Transformieren** liefert Ihnen zahlreiche Möglichkeiten der Datenauf- und umbereitung. Besonders hervorzuheben sind die Möglichkeiten:

- ⇒ neue Variablen zu berechnen (Befehl **Berechnen**) (Bsp. Seite 13)
- ⇒ Variablen in vorgegebene Kategorien (neu) zu kodieren (Befehl **Umcodieren**)  
(Bsp. in Übung 2)

Unter **Analysieren** finden Sie die eigentlichen statistischen Verfahren zur Datenauswertung. Hier bieten sich dem Nutzer umfassende Möglichkeiten, angefangen von der Erstellung einfacher Tabellen bis hin zur komplexen statistischen Prozedur.

Der Menüpunkt **Grafiken** bietet die Möglichkeit, verschiedenste visuelle Veranschaulichungen des Datenmaterials zu erzeugen. Ein Blick in die Galerie zeigt Ihnen die

umfangreichen Optionen. Mit der Version SPSS 12.0 steht ein komplett überarbeitetes Grafiksystem für ein deutlich professionelleres Reporting zur Verfügung.

Der Menüpunkt **Extras**, liefert einige Optionen, auf die hier nicht eingegangen werden soll. Mittels **Fenster** kann man zwischen mehreren SPSS-Fenstern wählen und **Hilfe** sollte der erste Schritt bei konkreten praktischen Problemen sein.

Im Folgenden wird anhand eines kleinen Beispieldatensatzes demonstriert, wie SPSS zur Datenanalyse genutzt werden kann. Da jede Fragestellung eigene Verfahren erfordert, kann hier natürlich keine umfassende Anleitung vorgegeben werden. Es werden stattdessen an einem einfachen Beispiel verschiedene Schritte der Datenauswertung und die praktische Umsetzung in SPSS vorgeführt.

## 2. Datenübermittlung an SPSS

In den folgenden Abschnitten sollen Sie einige wesentliche Funktionalitäten von SPSS anhand folgenden Beispiels kennen lernen:

In eine kleine klinische Studie wurden 20 Patienten eingeschlossen, von denen 10 mit einem Verum behandelt wurden und 10 lediglich ein Placebo erhielten. Sie haben von jedem Patienten in einem CRF (Case Report Form) u. a. folgende Merkmale erfasst:

*Behandlungsgruppe, Geburtsdatum, Geschlecht, Körpergröße, Körpergewicht und Alter zum Zeitpunkt des Studieneinschlusses.*

Sie wollen nun wissen, ob sich der Ausgangs-BMI (Body-Maß-Index) der Patienten, die mit Placebo behandelt wurden, von dem der Verumgruppe unterscheidet.

### 2.1 Datenerfassung direkt in SPSS

Zunächst müssen Sie dazu die Daten an SPSS übermitteln. Dies kann dadurch geschehen, dass man die einzelnen Werte in das Datenfenster eintippt.

- ⇒ einfach die gewünschte Datenzeile mit der Maus anwählen und die Zahlen in die Bearbeitungszeile eingeben. Die Bewegung innerhalb der Datenzellen kann mittels Pfeiltasten erfolgen.

Dies wird später anhand einer praktischen Aufgabe geübt. Öffnen Sie nun bitte die Datei **Z:\seminar\stamdat.sav**.

Das Resultat kann dann im vorliegenden Beispiel wie folgt aussehen:

	i_patnr	grup	gebdat	sex	gröÙe	gewicht	alter	var
1	1	1	25.04.1917	1	162	72	80.62	
2	2	0	14.11.1934	1	169	82	63.06	
3	3	0	01.12.1920	0	162	63	77.13	
4	4	1	09.09.1925	1	168	82	72.31	
5	5	1	28.09.1926	1	170	87	71.31	
6	6	1	02.04.1914	0	162	70	83.86	
7	7	0	07.01.1950	1	164	90	48.08	
8	8	0	04.03.1926	1	184	95	71.95	
9	9	1	25.01.1945	0	178	112	53.09	
10	10	0	18.09.1913	0	165	69	84.45	
11	11	1	12.03.1938	0	165	74	59.97	
12	12	0	13.11.1924	0	150	56	73.42	
13	13	0	27.12.1914	0	.	82	83.38	
14	14	0	15.11.1937	0	165	86	60.53	
15	15	1	25.01.1914	0	165	78	84.34	
16	16	1	04.05.1926	0	157	83	72.14	
17	17	1	29.04.1922	0	150	53	76.15	
18	18	0	18.12.1944	1	166	85	53.56	
19	19	0	01.06.1941	1	176	95	57.13	
20	20	1	03.05.1923	0	170	70	75.34	
21								
22								

Hier erkennt man bereits wichtige Richtlinien:

- ⇒ Jeder Patient wird durch eine Zeile repräsentiert, jede Variable (Merkmal) durch eine Spalte
- ⇒ Nominale (kategoriale) Variablen werden nicht in Textform, sondern in Zahlen kodiert eingegeben, d. h., bei der Variable Sex (Geschlecht) steht hier z. B. „1“ für männlich und „0“ für weiblich, bei der Variable grup (= Behandlungsgruppe) steht „0“ für Placebo und „1“ für Verum
- ⇒ Fehlende Werte (hier Körpergröße von Patient 13) können einfach freigelassen werden. SPSS setzt für systemdefinierte fehlende Werte automatisch ein „.“.
- ⇒ Die Variablennamen sind ab Version 12.0 nicht mehr nur auf höchstens 8 Zeichen beschränkt, es sind jetzt bis zu 64 Zeichen möglich. Sinnvollerweise sind diese aber kurz und sprechend zu halten. Jeder Variablenname ist verschieden! Sie dürfen zudem keine Sonderzeichen enthalten. Beim Eintippen von Datenwerten vergibt SPSS automatisch durchnummerierte Namen, beginnend mit var00001. Diese können

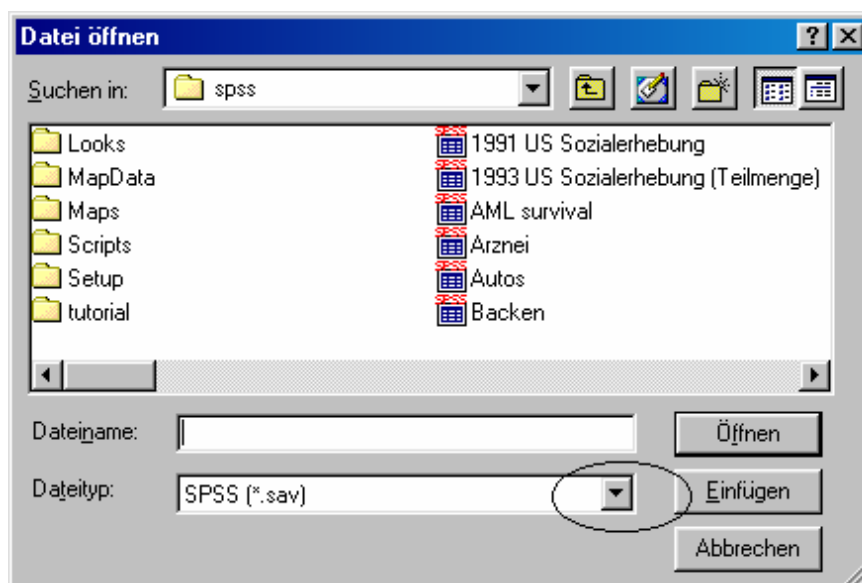
manuell geändert werden (vgl. S. 10f.), was hier im Beispiel schon geschah. Groß- und Kleinschreibung ist nicht relevant.

## 2.2 Datenimport

In der Praxis kommt es häufig vor, dass man die Daten in einem anderen Programm eingegeben hat und nun in SPSS einlesen möchte. Auf Grund seiner weiten Verbreitung findet dabei besonders häufig Microsoft Excel Anwendung. Auf unserer Homepage finden Sie einige grundlegende Hilfen zur Dateneingabe in Excel, die den Transfer in SPSS in der Regel komplikationslos ermöglichen.

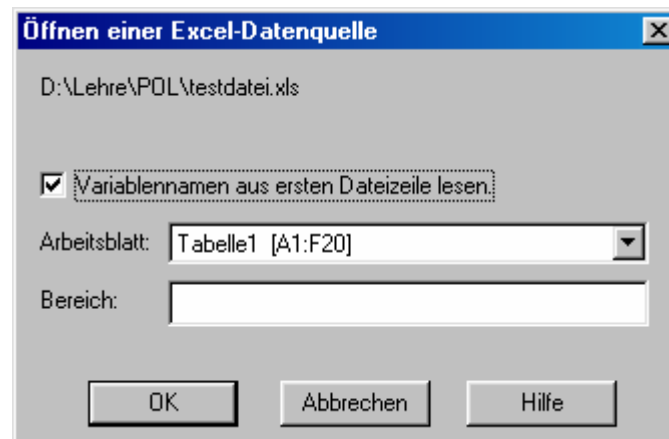
(vgl. [http://www.uke.uni-hamburg.de/zentren/experimentelle\\_medizin/biometrie/leistungen/Dateneingabe\\_unter\\_Excel.html](http://www.uke.uni-hamburg.de/zentren/experimentelle_medizin/biometrie/leistungen/Dateneingabe_unter_Excel.html)).

Wenn Sie eine Exceldatei vorliegen haben und diese mit SPSS nutzen möchten, dann können Sie einfach im **Datei**-Menü **öffnen** und dann **Daten** wählen und erhalten folgendes Dialogfenster:



Die Standardeinstellung besteht dabei in der Suche nach SPSS-Dateien mit der Endung **.sav**. Durch Anwählen des Pull-Down-Menüs Dateityp stehen jedoch diverse andere Formate zur Verfügung, u. a. auch **Excel (\*.xls)**.

Wählen Sie nun die gewünschte Excel-Datei [für den Kurs **Z:\seminar\stamdat.xls**] aus Ihrem Verzeichnis aus und betätigen **öffnen**, so erhalten Sie folgenden Dialog:



Sie können hierbei zusätzlich anwählen, welche Tabelle der Excel-Datei Sie transferieren wollen, und ob die Variablenamen in der ersten Zeile Ihrer Excel-Tabelle stehen und damit direkt von SPSS übernommen werden können.

**Tipp:** Nach einem erfolgten Datentransfer sollte man immer kurz überprüfen, ob die Datei auch wirklich wie geplant übernommen wurde. Dies kann spätere unangenehme Überraschungen vermeiden.

## 3. Datenweiterverarbeitung

Sind die Daten in das SPSS-Datenfenster eingelesen, so ist es in der Regel zunächst nötig, einige weitere Vorbereitungen für die eigentliche Datenauswertung zu treffen. Der erste Schritt sollte dabei in der Vergabe von **Labels** (Etiketten) bestehen.

### 3.1 Vergabe von Label in der Variablenansicht

Wie bereits zuvor erwähnt können für SPSS-Variablenamen bis zu 64 Zeichen vergeben werden (keine Sonderzeichen). Es empfiehlt sich aber kurze sprechende Namen zu vergeben. Die Werte (Ausprägungen) aller kategoriellen Variablen sind zudem als Zahlen kodiert. Dies ist für die Bearbeitung in SPSS notwendig, allerdings möchte man aus verständlichen Gründen in einer Auswertung z. B. nicht „*gebdat*“ sondern „*Geburtsdatum*“ stehen haben.

Dazu ist es nötig, so genannte Label (Etiketten) zu vergeben. Dies geschieht in der Variablenansicht, in welche Sie mittels des Registers links unten



wechseln können. Bitte öffnen Sie nun wieder die Datei **Z:\seminar\stamdat.sav** Auszugsweise sieht diese am Beispiel folgendermaßen aus:

	Name	Typ	Spaltenformat	Dezimalstellen	Variablenlabel	Wertelabels	Fel
1	i_patnr	Numerisch	3	0	Patienten-ID	Kein	Kei
2	grup	Numerisch	2	0	Behandlungsgruppe	Kein	Kei
3	gebdat	Datum	10	0	Geburtsdatum	Kein	Kei
4	sex	Numerisch	2	0	Geschlecht	Kein	Kei
5	größe	Numerisch	3	0	Größe	Kein	Kei
6	gewicht	Numerisch	3	0	Gewicht	Kein	Kei
7	alter	Numerisch	8	2	Alter	Kein	Kei
8							

Die einzelnen Variablen des Datensatzes sind zeilenweise aufgelistet, jedes einzelne Feld kann durch Anklicken mit der Maus bearbeitet und verändert werden.

Spalte 1 enthält den SPSS-internen **Variablennamen** (existiert noch keiner, so nummeriert SPSS einfach durch, beginnend mit var00001). Wichtig für Variablennamen:

- ⇒ bis zu 64 Zeichen lang, aber möglichst kurz und sprechend  
[ab Version 12.0, bis Version 11.5 auf 8 Zeichen beschränkt]
- ⇒ jeder Variablenname muss eindeutig sein
- ⇒ können aus Buchstaben und Ziffern bestehen, müssen mit einem Buchstaben beginnen
- ⇒ das letzte Zeichen sollte ein Buchstabe oder eine Zahl sein
- ⇒ dürfen keine Sonderzeichen enthalten (außer: \_ (Underscore), . (Punkt), @ und #
- ⇒ Groß- und Kleinschreibung ist nicht relevant

Beispiele für gültige Variablennamen: „sex“, „frage\_13“, „var3\_1\_2“

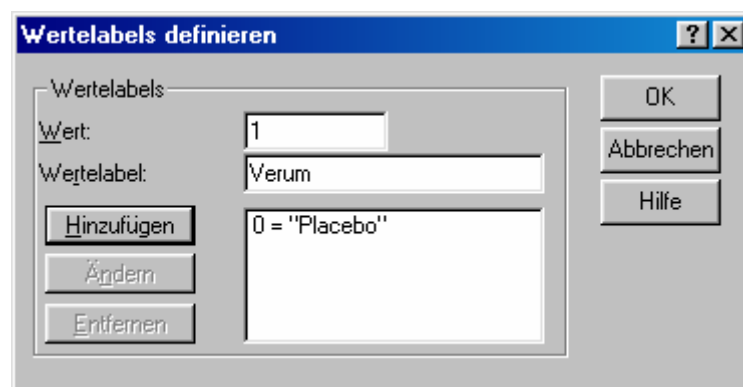
Die zweite Spalte beschreibt den **Variablentyp**. In diesem Beispiel sind alle Variablen numerisch (Voreinstellung von SPSS) mit Ausnahme des Geburtsdatums. Eine ganze Reihe von weiteren Variablentypen steht zur Auswahl. Für Freitexteingaben wird der Variablentyp „String“ genutzt (alphanumerisch).

In Spalte 3 und 4 kann mittels **Spaltenformat** und **Dezimalstellen** die Darstellungsbreite speziell von Zahlen verändert werden. Die Spaltenbreite kann auch im Fenster des Dateneditors verändert werden (analog zu Excel).

Von besonderem Interesse sind die nächsten beiden Spalten 5 und 6. Mittels **Variablenlabel** ist es möglich, jeder Variable ein Label zuzuordnen, welches an Stelle des Variablennamens aus Spalte 1 in jeder Form von Ausgabe (Tabelle, Grafik etc.) erscheint. Hier ist dies durch einfaches Anklicken und Eintippen bereits geschehen und man erkennt, dass diese Label sehr wohl länger als 8 Zeichen sein können („möglich“ sind bis zu 256 Zeichen) und auch Sonderzeichen enthalten dürfen.

Dieselbe Möglichkeit besteht auch für die Werte, welche eine Variable annimmt. Sie werden im vorliegenden Fall sicherlich nicht wollen, dass in einer Tabelle mit der Variablen Behandlungsgruppe „0“ und „1“ auftaucht, sondern „Placebo“ bzw. „Verum“. Eine derartige Zuordnung kann in der Spalte **Wertelabels** geschehen.

Klickt man in der Variablenzeile „grup“ auf die Zelle Wertelabels, erscheint ein (oben markierter) grauer Button, mit welchem folgender Dialog aufgerufen wird:



Hier ist es möglich, zu jedem Wert der Variablen ein Wertelabel (bis zu 60 Zeichen „möglich“) zu vergeben und dieses mittels **Hinzufügen** an SPSS zu übermitteln. Wie üblich beendet **OK** den Dialog.

Eine solche Etikettierung von Variablen und Werten kategorialer Variablen (also hier z. B. Geschlecht und Behandlungsgruppe) ist zwar manchmal etwas zeitaufwändig, aber auf alle Fälle lohnend, damit die erzeugte Ausgabe ohne zusätzliche Hilfsmittel (z. B. Fußnoten, Legenden) verständlich wird.

In Spalte 7 können **Fehlende werte** festgelegt werden. Die Voreinstellung ist **kein**, d. h., SPSS betrachtet alle Werte als gültige Werte. Liegen in bestimmten Fällen bei Variablen keine Antwortvorgaben vor, sei es wegen einer Antwortverweigerung, aus Unwissenheit oder anderen Gründen, kann der Benutzer an diesen Stellen im Daten-Editor einen Zahlen-Wert eintragen und diesen über den Schalter **fehlende werte** als fehlend deklarieren. Dieser Zahlenwert darf natürlich unter den auftretenden Werten nicht vorkommen. Fehlende Werte werden aus weiteren Berechnungen ausgeschlossen (z. B. 9 für keine Angabe bei „sex“, oder -9 für keine Angabe bei „größe“).

In Spalte 8 wird die Breite der **spalten** der jeweiligen Datenfelder beim Anzeigen der Werte in der Datenblattansicht festgelegt.

In Spalte 9 kann die **Ausrichtung** der Werte festgelegt werden (Rechts/ Links/ Mitte).

In Spalte 10 kann das **Messniveau** der Werte festgelegt werden. Bei numerischen Variablen ist das Messniveau **metrisch**, bei alphanumerischen Variablen **nominal** voreingestellt. Allerdings ist die Unterscheidung nur bei der Erstellung interaktiver Grafiken relevant, wo nominales und ordinales Messniveau zu „kategorial“ zusammengefasst wird.

## 3.2 Neuberechnung von Variablen

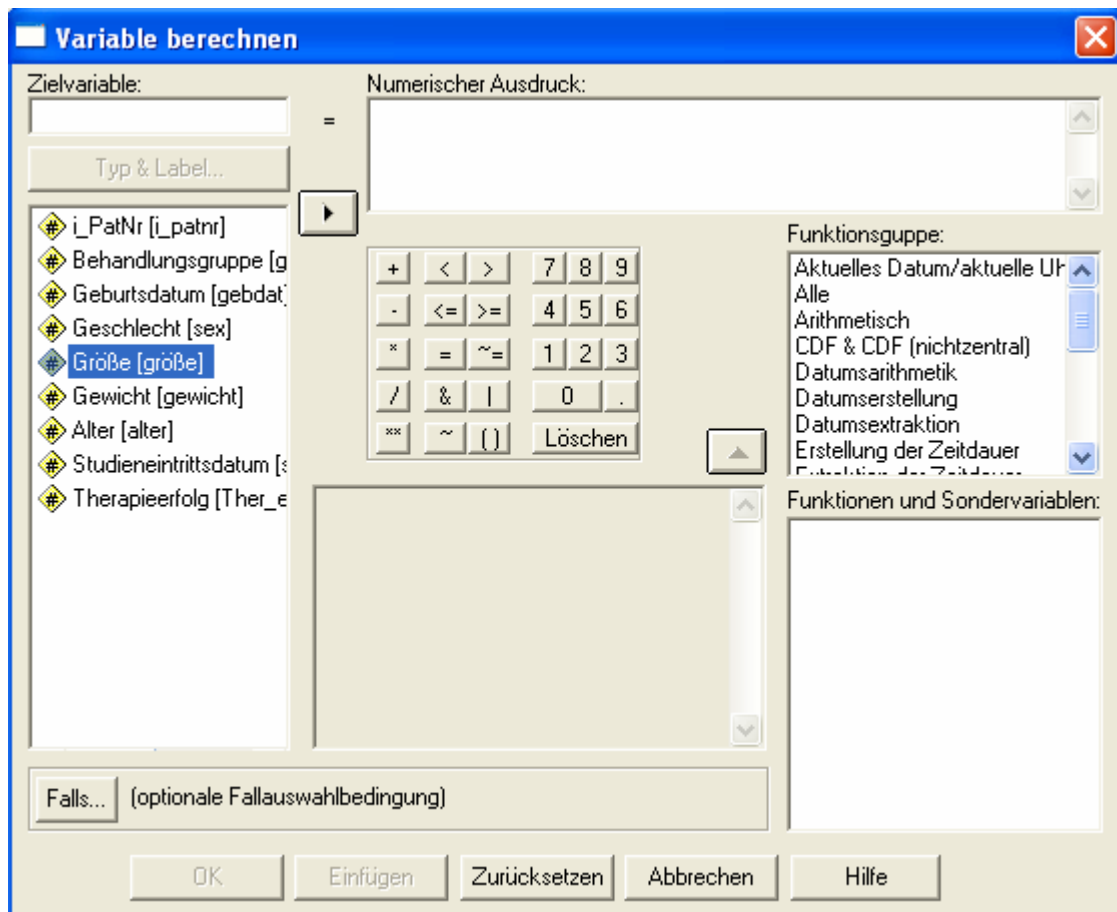
Um im vorliegenden Beispiel die ursprüngliche Fragestellung (vgl. S. 7) beantworten zu können (Vergleich des Body-Mass-Index BMI zwischen Placebo- und Verumgruppe), muss der BMI für jeden der 20 Patienten aus den vorhandenen Daten berechnet werden.

Der BMI ist wie folgt definiert (vgl. Pschyrembel, 258. Auflage, S. 220):

$$BMI = \frac{\text{Gewicht [kg]}}{(\text{Größe [m]})^2} .$$

Da sowohl das Körpergewicht als auch die Körpergröße pro Proband erfasst ist, lässt sich der jeweilige BMI ohne Probleme mit folgenden SPSS-Befehlen errechnen:

Im Menü **Transformieren** findet sich an erster Stelle der Befehl **Berechnen**, womit folgendes Dialogfeld geöffnet wird:



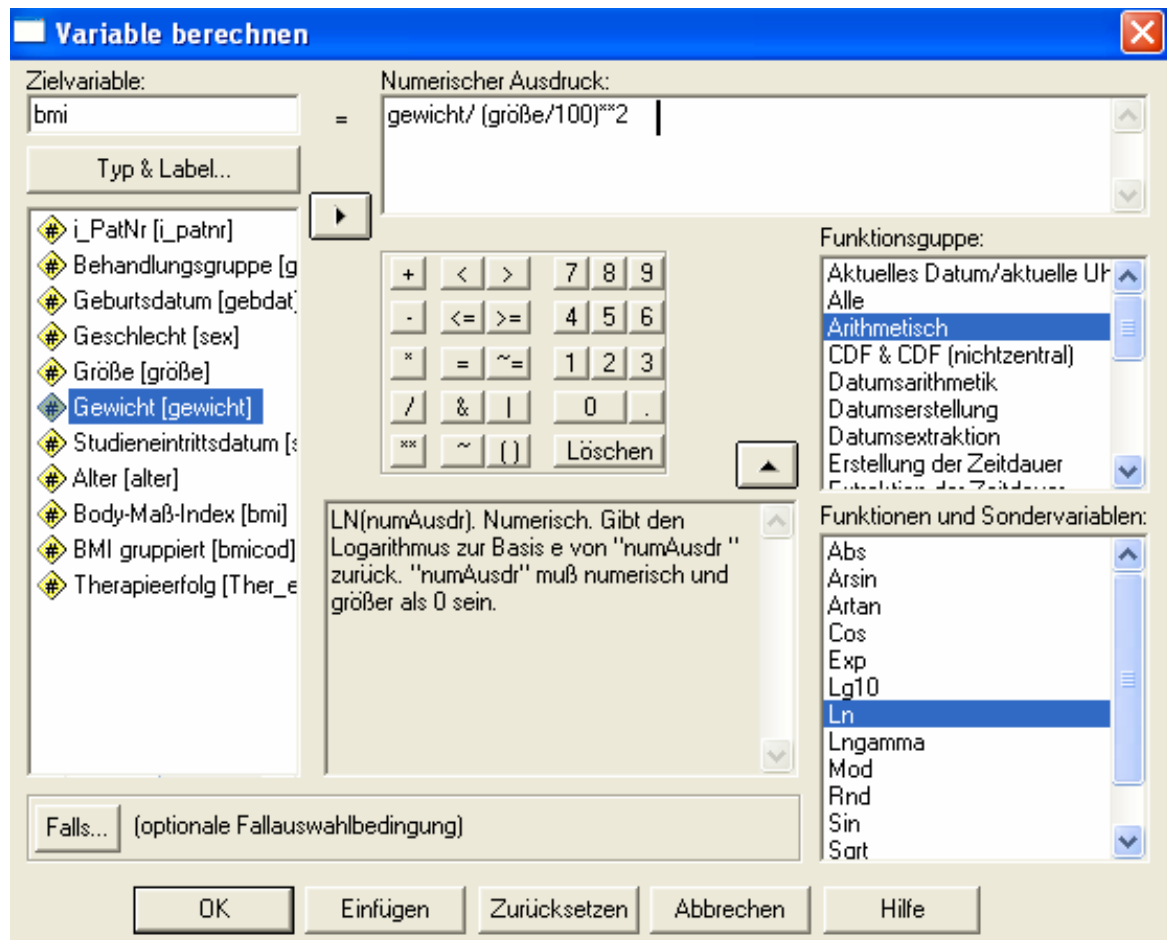
Hier handelt es sich um ein sehr SPSS-typisches Dialogfeld. Man erkennt zunächst links oben das Feld **Zielvariable**, in welche man den Namen der neuen Variablen eintragen kann, also im vorliegenden Fall beispielsweise „bmi“.

Im Feld **numerischer Ausdruck** kann die Formel entwickelt werden, die zur Errechnung der neuen Variablen verwendet werden soll. Dazu hat man drei Komponenten zur Auswahl:

1. In der Mitte eine taschenrechnerähnliche Auswahl an mathematischen und logischen Operatoren, die einfach angeklickt werden können
2. Auf der rechten Seite eine alphabetisch geordnete große Auswahl an mathematischen und statistischen Funktionsgruppen, deren einzelne Erläuterung diesen Rahmen sprengen würde (Erläuterung einzelner Funktionen im mittleren Bereich des Dialogfeldes).
3. Auf der linken Seite die Liste der zurzeit im Datensatz verfügbaren Variablen. Soll eine mathematische Funktion auf eine dieser Variablen angewendet werden, so kann diese Variable mit der linken Maustaste angewählt und mittels des dann schwarz aufleuchtenden Pfeils in das Feld **numerischer Ausdruck** befördert werden.

Die Beendigung des Dialogs mittels **OK** ist erst möglich, wenn eine gültige Formel und ein gültiger neuer Variablenname angegeben sind.

Zur Berechnung des BMI kann dieses Dialogfeld (unter Benutzung von 1. und 2.) wie folgt aussehen:



Dabei ist zu beachten, dass die Körpergröße im Datensatz in cm vorliegt und deshalb eine Umrechnung mittels des Faktors 1/100 nötig ist.

Mittels **OK** kann SPSS nun angewiesen werden, die Berechnung für alle Patienten durchzuführen. Das Ergebnis erkennt man durch einen Blick auf die Datenansicht, welche sich nun folgendermaßen darstellt:

	i_patnr	grup	gebdat	sex	gröÙe	gewicht	alter	bmi	var
1	1	1	25.04.1917	1	162	72	80.62	27.43	
2	2	0	14.11.1934	1	169	82	63.06	28.71	
3	3	0	01.12.1920	0	162	63	77.13	24.01	
4	4	1	09.09.1925	1	168	82	72.31	29.05	
5	5	1	28.09.1926	1	170	87	71.31	30.10	
6	6	1	02.04.1914	0	162	70	83.86	26.67	
7	7	0	07.01.1950	1	164	90	48.08	33.46	
8	8	0	04.03.1926	1	184	95	71.95	28.06	
9	9	1	25.01.1945	0	178	112	53.09	35.35	
10	10	0	18.09.1913	0	165	69	84.45	25.34	
11	11	1	12.03.1938	0	165	74	59.97	27.18	
12	12	0	13.11.1924	0	150	56	73.42	24.89	
13	13	0	27.12.1914	0	.	82	83.38	.	.
14	14	0	15.11.1937	0	165	86	60.53	31.59	
15	15	1	25.01.1914	0	165	78	84.34	28.65	
16	16	1	04.05.1926	0	157	83	72.14	33.67	
17	17	1	29.04.1922	0	150	53	76.15	23.56	
18	18	0	18.12.1944	1	166	85	53.56	30.85	
19	19	0	01.06.1941	1	176	95	57.13	30.67	
20	20	1	03.05.1923	0	170	70	75.34	24.22	
21									

Man erkennt an der rechten Seite zusätzlich die Variable „bmi“ mit den jeweiligen Werten der Patienten. Für Probandin 13 konnte kein BMI errechnet werden, da die Angabe ihrer Körpergröße fehlt.

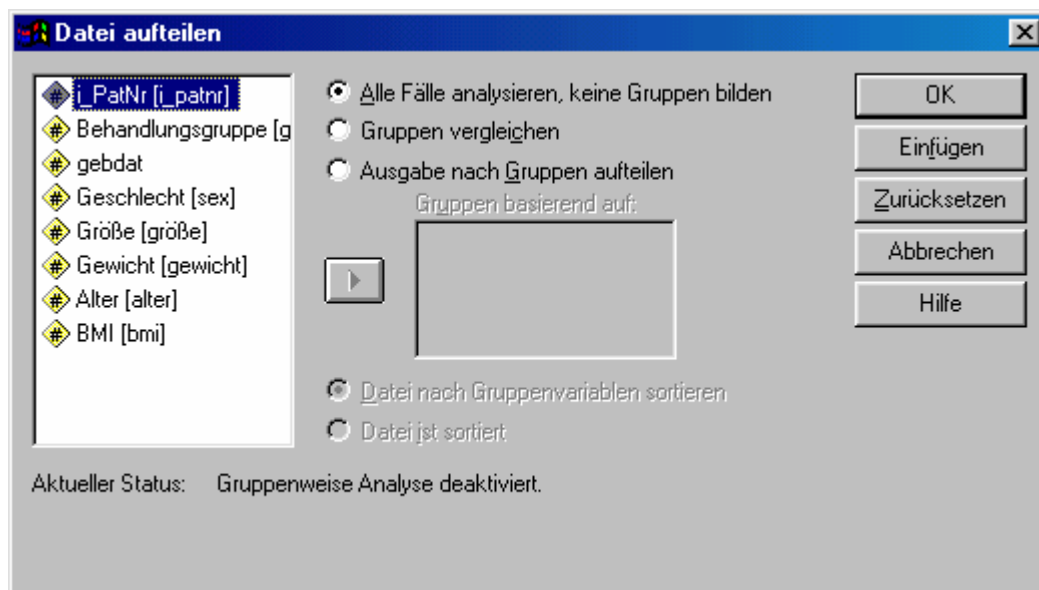
## 4. Datenauswertung

Nach all den vorbereitenden Maßnahmen kann nun zur Beantwortung der Fragestellung übergegangen werden. Um den Rahmen nicht zu sprengen, soll der Vergleich des BMI zwischen Placebo- und Kontrollgruppe lediglich beschreibend (deskriptiv) durch Angabe von geeigneten Maßzahlen und eine adäquate Grafik geschehen.

## 4.1 Statistischer Vergleich mittels geeigneter Maßzahlen

Im Folgenden soll der BMI getrennt nach Verum- und Placebogruppe in einer Tabelle mittels beschreibender Maßzahlen dargestellt werden. Zunächst ist es dafür nötig, SPSS zu übermitteln, nach welcher Variablen die Tabelle unterteilt werden soll.

Derartige Vorgaben sind in dem Menüpunkt **Daten** zu finden. Wählt man dort **Datei aufteilen** an, so erscheint folgendes Dialogfenster:



Ersetzt man die Standardeinstellung (**Alle Fälle analysieren**) durch **Gruppen vergleichen**, kann man wiederum mittels des Pfeils die Variable aus der Liste links auswählen, nach der die Ausgabe getrennt werden soll, in diesem Fall also die Variable „Behandlungsgruppe[grup]“, die in das Feld **Gruppen basierend auf** gesetzt wird.

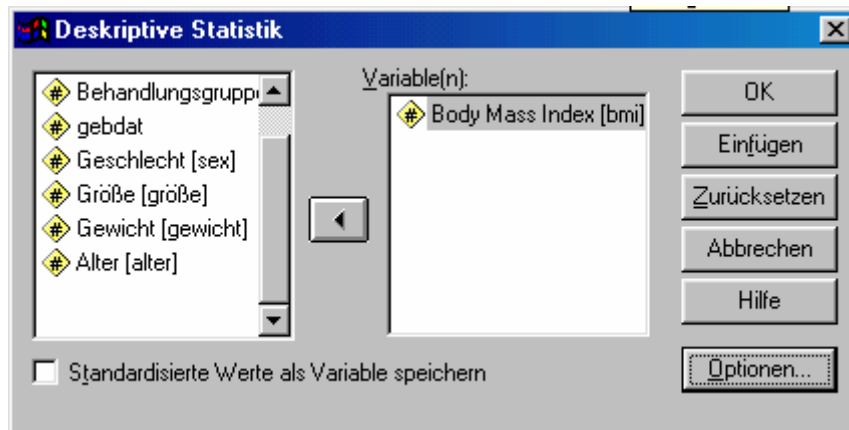
Die Ausführung dieser Auswahl mittels **OK** führt in der Datenansicht zu zwei Effekten:

- ⇒ Zum einen wird die Datei automatisch aufsteigend nach der Gruppenvariable sortiert
- ⇒ zum anderen erscheint in der Statusleiste die Mitteilung Datei aufteilen an

**Warnung:** Von nun an wird alles, was Sie mit SPSS tun, getrennt nach Behandlungsgruppen, durchgeführt. Um die Aufteilung zu deaktivieren, muss wiederum im Menüpunkt **Daten** der Befehl **Datei Aufteilen** angewählt und die Standardeinstellung wieder hergestellt werden!

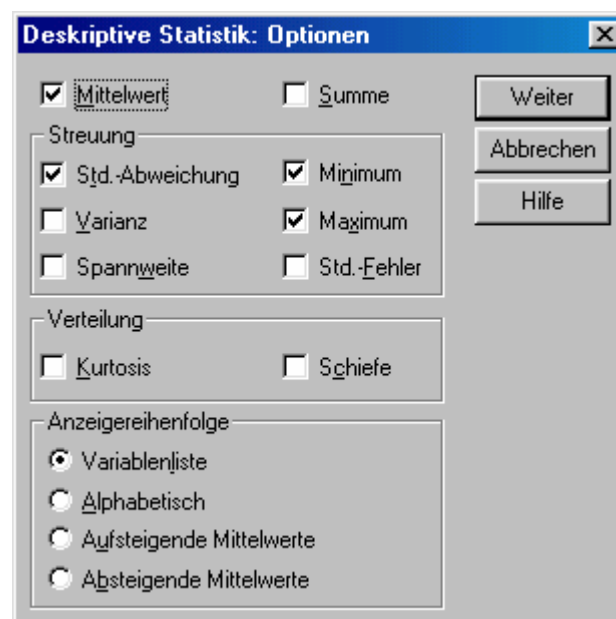
Um nun die gewünschte Tabelle zu erzeugen, wird das Menü **Analysieren** benötigt. Dort lassen sich nochmals, unterteilt nach einzelnen Prozedurgruppen, eine Vielzahl von statistischen Verfahren anwählen und durchführen.

Einfache beschreibende Maßzahlen erhält man unter **Deskriptive Statistiken** und dann wiederum **Deskriptive Statistiken**. Nach dieser Wahl öffnet sich folgendes SPSS-Dialogfenster:



Erneut lässt sich mittels des Pfeils (hier bereits geschehen) die Variable auswählen, für die deskriptive Statistiken erstellt werden sollen. Wie in vielen ähnlichen Dialogfenstern hat man mittels **optionen** die Möglichkeit, bestimmte Funktionen gezielt anzufordern.

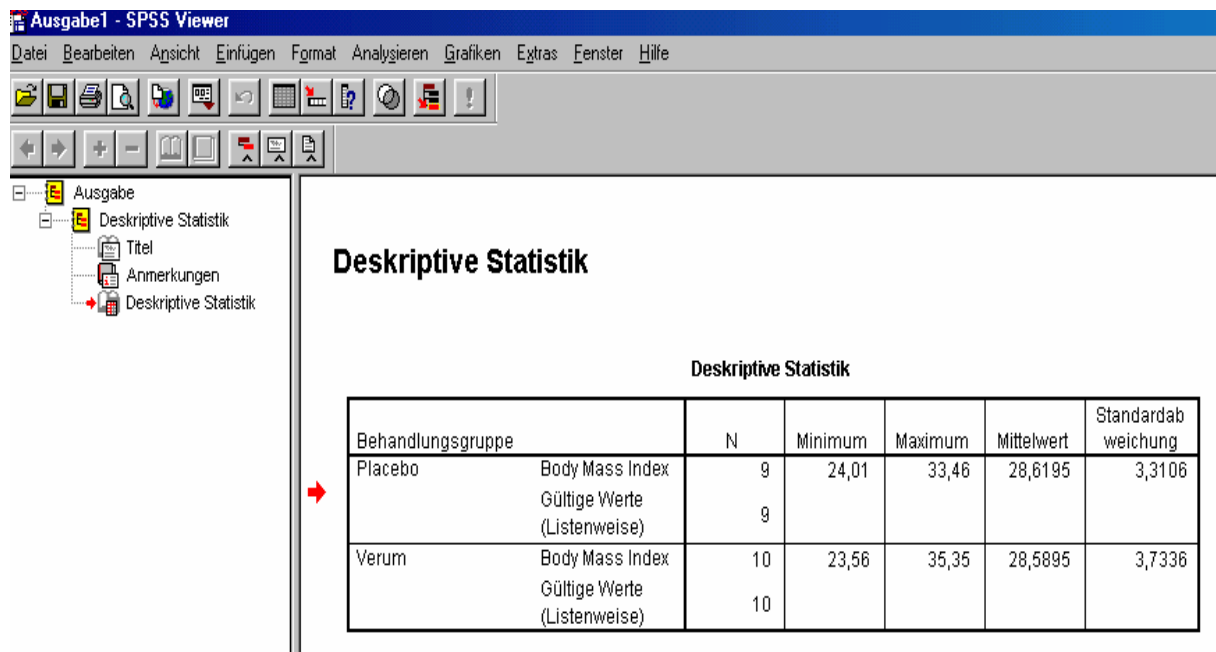
Es öffnet sich ein neues Dialogfenster:



Die Standardeinstellung besteht in der Angabe von Mittelwert und Standardabweichung, Minimum und Maximum der Daten. Ohne darauf näher eingehen zu wollen, kann man erkennen, dass bei Bedarf einige weitere Statistiken angewählt werden können. Nach erfolgter Auswahl kann man mittels **OK** wieder das vorherige Fenster erreichen. Klickt man auch hier auf **OK**, so wird SPSS veranlasst, eine Ausgabe zu erzeugen, die die angeforderten statistischen Maßzahlen enthält. Dazu öffnet sich ein neues Fenster, das sog. SPSS-Ausgabefenster, in welches SPSS sämtliche Ausgaben, aber auch gelegentlich Fehlermeldungen schreibt. Dieses Fenster wird im folgenden Abschnitt kurz erläutert.

## 4.2 Das SPSS-Ausgabefenster

Das sich neu öffnende **SPSS-Ausgabefenster** hat folgende Gestalt:



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 viewing. The left pane shows a tree view with 'Ausgabe' expanded to 'Deskriptive Statistik', which includes 'Titel', 'Anmerkungen', and 'Deskriptive Statistik'. The main area displays the 'Deskriptive Statistik' table.

Behandlungsgruppe		N	Minimum	Maximum	Mittelwert	Standardabweichung
Placebo	Body Mass Index	9	24,01	33,46	28,6195	3,3106
	Gültige Werte (Listenweise)	9				
Verum	Body Mass Index	10	23,56	35,35	28,5895	3,7336
	Gültige Werte (Listenweise)	10				

Man erkennt auf der rechten Seite die gewünschte Ausgabe, eine Tabelle mit den geforderten Maßzahlen getrennt nach den beiden Behandlungsgruppen. Inhaltlich lässt sich dabei erkennen, dass sich die beiden Gruppen nicht unterscheiden, weder in der Größenordnung des BMI (Mittelwert) noch in der Variabilität (Standardabweichung).

Auf der linken Seite kann jedes einzelne Element der gesamten Ausgabe gezielt mit der Maus angesteuert werden. Dies ist speziell bei größeren Arbeiten mit einer umfangreichen Ausgabe äußerst hilfreich. Die Struktur ähnelt dabei der eines Baumes, und es ist z. B. hierdurch möglich, komplette „Äste“ zu markieren und mittels der **Entf**-Taste zu löschen.

Die **Bearbeitung einer Tabelle**, um sie persönlichen Präferenzen (Schriftart, Rahmenlinien etc.) anzupassen, ist mittels Doppelklick möglich. Hier gibt es zahlreiche Bearbeitungsmöglichkeiten, mit welchen man sich am einfachsten durch selbständiges Probieren vertraut machen kann.

**Tip:** unter **Tabelleneigenschaften/Drucken** lassen sich für die Seite zu große Tabellen - passend auf die Seitenbreite/-länge verkleinern.

Zur Dokumentation der Auswertungsergebnisse ist es sinnvoll die SPSS-Outputdatei abzuspeichern (Dateiendung .spo).

#### **Objekt-Einbindung in Word oder Power Point:**

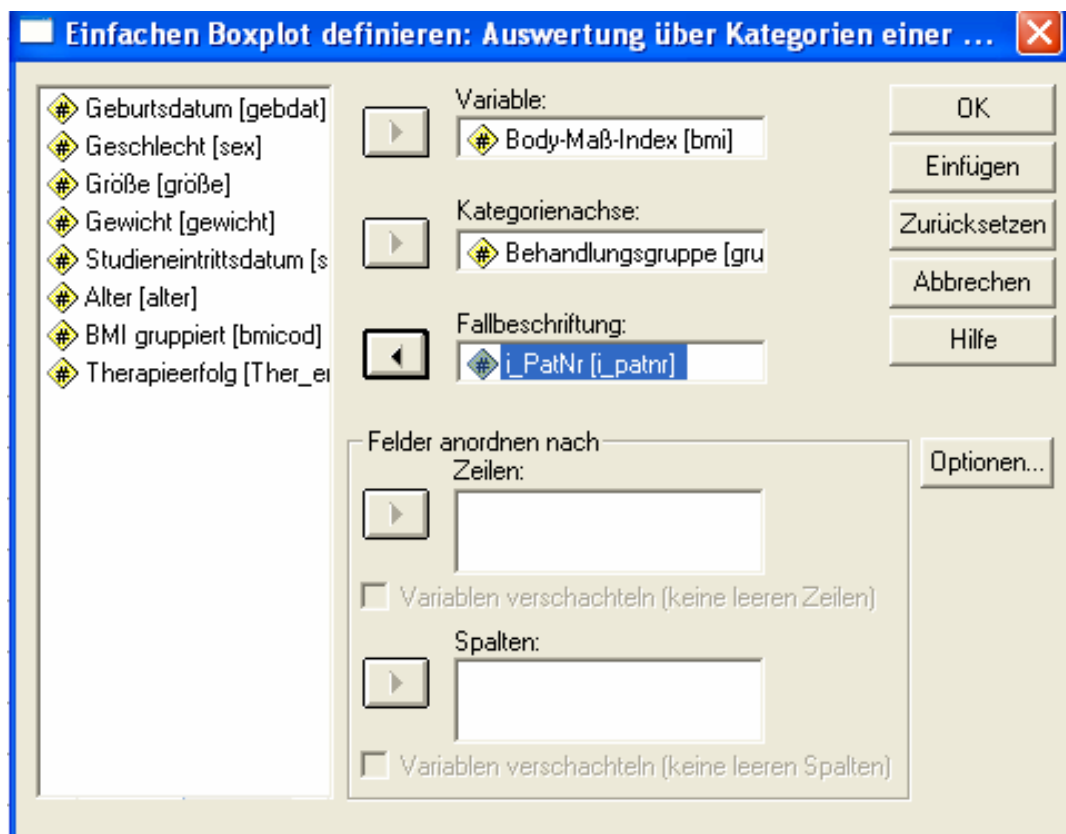
Um eine Tabelle wie die vorliegende z. B. in ein Word-Dokument oder eine Power-Point-Präsentation einzubinden, klicken Sie die Tabelle mit der rechten Maustaste an und wählen **objekte kopieren** an. Damit wird die Tabelle in die Zwischenablage übernommen und kann in andere Windows-Anwendungen direkt mittels **Bearbeiten/Einfügen** eingefügt werden.

### 4.3 Grafische Darstellung mittels Box-Plot

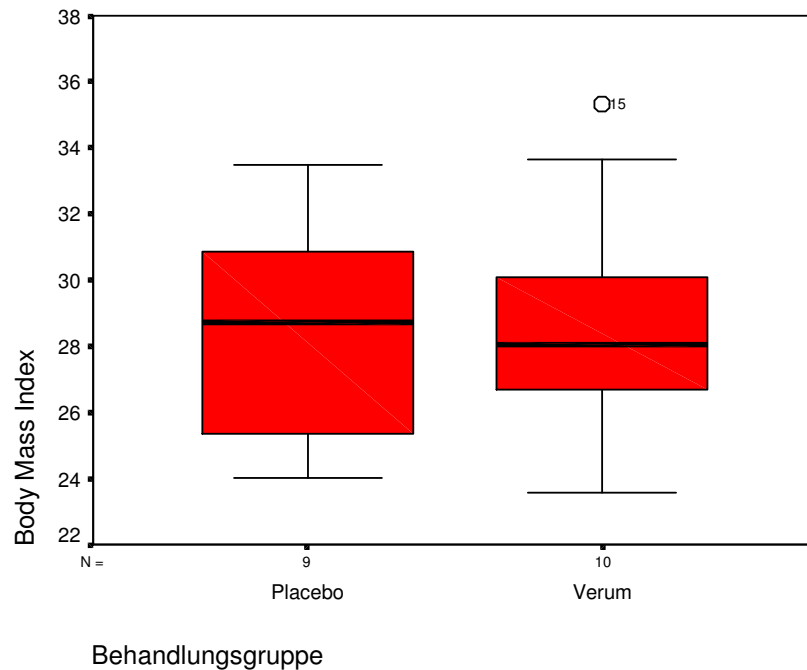
Eine (sinnvolle) Grafik sagt häufig mehr aus als einfache Zahlenkolonnen. Deshalb soll hier zum Abschluss noch kurz darauf eingegangen werden, wie aus den vorhandenen Daten ein sog. **Box-Plot** erstellt werden kann, welcher die Aussage der vorher erstellten Tabelle unterstützt und ergänzt.

Ein Box-Plot dient, einfach gesprochen, dazu, die Verteilung numerischer Variablen zu visualisieren (Details s. u.).

Nach der Wahl von **Grafik/Boxplot** und Anklicken von **Definieren** erscheint wiederum ein SPSS-Dialogfenster:



Hier erhalten Sie die Möglichkeit, die metrische Variable auszuwählen, welche Sie grafisch darstellen möchten, und gleichzeitig eine zweite, kategoriale Variable (hier die Behandlungsgruppe) wonach die Grafik getrennt werden soll (sog. Kategorienachse). Betätigen Sie **OK**, so erscheint im Ausgabefenster die angeforderte Grafik.



Dargestellt ist hierbei getrennt nach Behandlungsgruppen die Verteilung des BMI.

- ⇒ Die beiden Box-Plots sind dabei durch den jeweiligen Minimal- und Maximalwert begrenzt.
- ⇒ Die farbige Mitte (Box) beschreibt die mittlere Hälfte der Daten und der schwarze Strich in deren Mitte stellt den sog. Median dar. Dieser wird auch „zentraler Wert“ genannt und ist definiert als der Wert, für den die Hälfte der Probanden einen größeren und die andere Hälfte einen kleineren BMI hat.
- ⇒ Des Weiteren erkennen Sie, dass in der Verumgruppe eine Person einen ungewöhnlich hohen BMI-Wert aufweist (sog. Ausreißer).

Auch diese Grafik kann durch Doppelklick in vielen Einzelheiten bearbeitet und verschönert werden (z. B. Achsenskalierung und Beschriftung, Titel, Farbe etc.). Der Transfer in andere Windows-Anwendungen funktioniert exakt wie für die Tabelle auf S. 21 beschrieben.

## 5. Zusammenfassung und Ausblick

Das vorliegende Skript mit Übungen sollte Ihnen anhand kleiner Beispiele eine kurze Einführung in SPSS liefern. Sicherlich kann diese nicht umfassend sein und viele Dinge nur ganz grob anreißen, ein Softwarepaket wie SPSS lernt man aber ohnehin nur dadurch wirklich kennen, indem man es regelmäßig benutzt.

### SPSS-Syntaxfenster

Unter den zahlreichen Möglichkeiten von SPSS, die in diesem Skriptum unerwähnt blieben, sei hier kurz auf ein Verfahren verwiesen, welches für fortgeschrittene Nutzer speziell dann sinnvoll ist, wenn man die durchgeführten Arbeitsschritte dokumentieren und zu einem späteren Zeitpunkt nachvollziehen will.

Nahezu jede SPSS-Dialogbox bietet die Möglichkeit, anstelle des üblichen **OK** den Dialog mittels **Einfügen** zu beenden. Damit erreicht man, dass der gewünschte Befehl nicht sofort durchgeführt wird, sondern die zugehörige SPSS-interne Syntax in das sog. **Syntax-Fenster** geschrieben wird. Diese Syntax kann dann mittels der im Syntaxfenster befindlichen Pfeiltaste an SPSS übermittelt und ausgeführt werden.

Der Vorteil des Arbeitens mit Syntax besteht darin, dass zum einen mehrmaliges Wiederholen des gleichen Befehls deutlich beschleunigt wird und zum anderen die Möglichkeit gegeben ist, diese Syntax nach beendeter Arbeit gesondert abzuspeichern und als Arbeitsdokumentation zu verwahren.

Eine kurze Warnung sei an dieser Stelle abschließend hinzugefügt. Die beste statistische Software bewahrt den Benutzer nicht davor, die hinter den einzelnen Prozeduren stehenden Methoden zu verstehen und damit auch die Ergebnisse richtig zu interpretieren. Eine Vermittlung dieser Inhalte ist u. a. Ziel der Biometrie-Veranstaltung im Studium Humanmedizin.

# Literaturverzeichnis SPSS

Zur Ergänzung und tieferen Einführung in die Arbeit mit SPSS finden Sie im Folgenden eine kleine Auswahl begleitender Bücher. Ein Teil der Literatur zu SPSS ist thematisch auf Wirtschafts- und Sozialwissenschaftler zugeschnitten. Da jedoch die grundlegenden statistischen Verfahren und deren Umsetzung in allen Wissenschaften identisch sind, wurde auch eines dieser Werke mit aufgenommen. Die Unterschiede zwischen den SPSS-Versionen 9.0, 10.0, 11.0 und 11.5 sind primär nicht struktureller Natur, sondern betreffen weitestgehend Detailfragen, deshalb ist es generell durchaus möglich, auch ein Buch, welches sich auf eine der zwei Vorgängerversionen von SPSS 11.5 bezieht, als Referenzwerk zu nutzen.

## SPSS

- Bühl A. und Zöfel P. [2008]: SPSS 16. Einführung in die moderne Datenanalyse . Pearson Studium, München, 1. Auflage, gebunden.
- Bühl A. und Zöfel P. [2006]: SPSS 14. Einführung in die moderne Datenanalyse . Pearson Studium, München, 10. Auflage.
- Eckstein P. [2006]: *Angewandte Statistik mit SPSS. Praktische Einführung für Wirtschaftswissenschaftler*. Gabler-Verlag, 5. Auflage.]
- Zöfel P. und Bühl A. [2000], *Statistik verstehen. Ein Begleitbuch zur computerunterstützten Anwendung*, . Addison-Wesley, 2. Auflage
- Martens J. [2003]: *Statistische Datenanalyse mit SPSS für Windows*. Oldenbourg, München, 2.Auflage.
- Rinne, K. [2003]: *SPSS - echt einfach*. Franzis Verlag, 1. Auflage (September 2003) ISBN 3-7723-6899-9.
- Brosius F. [2004]: *SPSS 12, m. CD-ROM*. Mitp-Verlag
- Janssen, J. und Laatz, W., [2007] *Statistische Datenanalyse mit SPSS für Windows*, Springer-Verlag, 6. Auflage

SPSS-Syntax:

- Zöfel P., [2002]: *SPSS Syntax. SPSS-Syntax, m. CD-ROM. Die ideale Ergänzung für effiziente Datenanalyse.* Pearson Studium
- Brosius F., [2005]: *SPSS-Programmierung - Effizientes Datenmanagement und Automatisierung mit SPSS-Syntax.* Mitp-Verlag, Bonn, 1.Auflage
- Schendera Christian F.G, [2005]: *Datenmanagement mit SPSS - Kontrollierter und beschleunigter Umgang mit Datensätzen, Texten und Werten.* Springer-Verlag, 1.Auflage

Zusätzliche Hilfestellung bei praktischen Problemen mit SPSS finden Sie auf der Homepage <http://www.spss.com/de>

**Statistische Methoden**

Wie bereits im Text (S.3) erwähnt, finden Sie Basiswissen über statistische Methoden in der Medizin in dem an der Universität Münster entwickelten Skriptum, welches im Copy-Shop am UKE erhältlich ist.

Die einführende Literatur über statistische Methoden in der Medizin ist äußerst umfangreich. Einen guten Überblick finden Sie z. B. in

- Harms V. [1998]: *Biomathematik, Statistik und Dokumentation.* Harms, 7.Auflage
- Trampisch H. und Windeler J.[2000]: *Medizinische Statistik.* Springer, 2.Auflage.
- Armitage P. und Berry G. [2001]: *Statistical Methods in Medical Research.* Blackwell, 4.Auflage.
- Altman DG. [2009]: *Practical Statistics for Medical Research.* Chapman & Hall, London., 2.Auflage
- Clauß G. und Finze F.-R. und Partsch L., [2004]: *Statistik für Soziologen, Pädagogen, Psychologen und Mediziner, Band I.* Harri Deutsch, 5. Auflage
- Sachs L. [2006]: *Angewandte Statistik.* Springer, Berlin, 12.Auflage
- Hilgers R.; Bauer P. Scheiber V., [2007]: *Einführung in die Medizinische Statistik,* Springer Berlin, 2.Auflage

[Autoren: M. Supplith/ V. Schoder / A. Großer IMBE]; erstellt: 18.03.2002; aktualisiert 18.06.2008