



Universitätsklinikum
Hamburg-Eppendorf

Der SPSS- (Wieder-) Einstieg leicht gemacht

Eine Kurzeinführung in die statistische
Datenauswertung mit SPSS



IBM® SPSS® Statistics
Version 20

IBM® SPSS® Statistics
Version 21

Lizenziertes Material - Eigentum von IBM Corp. © Copyright IBM Corporation und anderen 1989, 2012. IBM, das IBM-Logo, ibm.com und SPSS sind Marken oder eingetragene Marken von International Business Machines Corp., die in vielen Ländern weltweit eingetragen sind. Eine aktuelle Liste von Marken von IBM ist im Internet unter www.ibm.com/legal/copytrade.shtml verfügbar. Java und alle Java-basierten Marken und Logos sind Marken oder eingetragene Marken von Oracle und/oder seiner angeschlossenen Unternehmen. Andere Produkt- und Servicennamen können Marken von IBM oder anderen Unternehmen sein. Dieses Programm ist gemäß den Bestimmungen der zum Programm gehörenden Lizenzvereinbarung lizenziert. Diese Lizenzvereinbarung befindet sich ggf. entweder in einem Ordner des Programmverzeichnis oder in einer Bibliothek mit der Kennzeichnung "License" bzw. "Lizenz" oder "Non_IBM_License" bzw. "Nicht-IBM-Lizenz" oder wird als gedruckte Lizenzvereinbarung bereitgestellt. Lesen Sie die Vereinbarung sorgfältig durch, bevor Sie das Programm verwenden. Durch die Verwendung des Programms stimmen Sie diesen Bestimmungen zu.



ab Version 18

Institut für Medizinische Biometrie und Epidemiologie
(IMBE)



Inhaltsverzeichnis

1. Überblick	3
1.1 Allgemeines	3
1.2 Der erste Eindruck.....	4
1.3 Fenster in SPSS.....	6
2. Datenübermittlung an SPSS	8
2.1 Datenerfassung direkt in SPSS.....	8
2.2 Datenimport	10
3. Datenweiterverarbeitung.....	12
3.1 Variablendefinition und formatierung in der Variablenansicht.....	12
3.2 Neuberechnung von Variablen	15
4. Datenauswertung	18
4.1 Statistischer Vergleich mittels geeigneter Maßzahlen	18
4.2 Das SPSS-Ausgabefenster	21
4.3 Grafische Darstellung mittels Box-Plot	22
5. Zusammenfassung und Ausblick.....	25
5.1 SPSS-Syntaxfenster	25
5.2 Mehr Lernen:	26
6. Literaturverzeichnis SPSS	27

SEMINAR: Einführung in die Anwendung statistischer Software (SPSS unter Windows)

<http://www.uke.de/institute/biometrie/>

Anmeldung: grosser@uke.de (Andrea Großer)



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. In der Version 18.0 fand eine Umbenennung zu **PASW® Predictive Analytics Software** statt (Umbenennung 2009 für die Produktfamilie). Nach Zusammenschluss mit IBM heißt das Programm nun

IBM® SPSS® Statistics.

Im UKE stehen Ihnen in der ÄZB im MediTreFF zwei Computerräume mit je 10 Computerarbeitsplätzen (N60 1.OG) zur Verfügung (<http://elearning.uke.uni-hamburg.de/moodle/course/view.php?id=462#section-2>)

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.de/institute/biometrie/> unter Lehre und Fortbildung). Die Kurzeinführung erfolgt am Beispiel der an der Uni Hamburg aktuell zur Verfügung stehenden SPSS-Version, wobei alle Inhalte auch für die Vorversionen ab SPSS 18 und die Nachfolgeversionen gelten. In den neueren SPSS-Versionen haben oftmals die Symbolleiste bzw. Symbolfarben sowie die Programm – und Dialogfenster in kleinen Details ein abweichendes Layout bekommen.

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://jumbo.uni-muenster.de/>. SPSS selbst bietet Online-Seminare zu verschiedenen praktischen Fragestellungen an: <http://www.spss.com/DE/events/>
<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

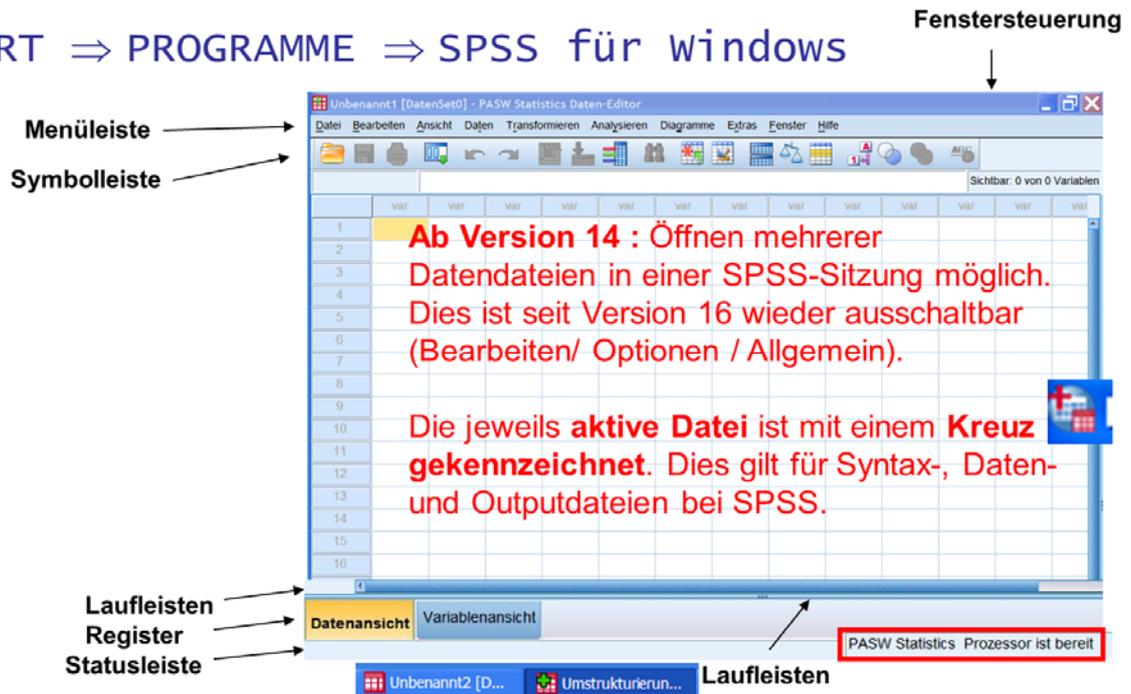


Abbildung 1: SPSS-Datenfenster in der Datenansicht

Der Großteil des Bildschirms wird dabei von den einzelnen Zellen bestimmt, in welche sich Daten importieren 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 importieren 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 Optionen können Sie die Standardeinstellungen von SPSS verändern und Ihre eigenen Präferenzen festlegen.

Unter **Ansicht** kann die optische Gestaltung des Bildschirms an persönliche 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 Datenaufbereitung und -umbereitung. Besonders hervorzuheben sind die Möglichkeiten:



- ⇒ neue Variablen zu berechnen (Befehl **Berechnen**) (vgl. S. 15)
- ⇒ Variablen in vorgegebene Kategorien (neu) zu kodieren (Befehl **Umkodieren**) (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 16.0 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.

1.3 Fenster in SPSS

(1) Datenfenster

Datenfenster = Datendatei

Dateiendung: *.sav

Das Datenfenster ist ein in Zeilen und Spalten aufgeteiltes „Arbeitsblatt“.

Über ein Register erfolgt der Wechsel zwischen zwei Arbeitsmodi:

1. **Datenansicht:** Eingeben und Anzeigen von Daten
2. **Variablenansicht :** Definieren und Bearbeiten der Variablen

(2) Ausgabefenster

Ausgabedatei= Viewer/Output

bis Version 15.0 Dateiendung: *.spo

ab Version 16.0 Dateiendung: *.spv

Hier werden die Ergebnisse ausgegeben und es ist möglich die Ausgabe durch Ein- bzw. Ausblenden der einzelnen Ergebnisse oder durch ändern der Anzeigereihenfolge zu bearbeiten. Des Weiteren können Tabellen und Diagramme in Präsentationsqualität von SPSS zu anderen Anwendungen übertragen werden.

(3) Syntaxfenster

Syntaxfenster = Syntaxdatei = wichtige Steuerdatei Dateiendung: *.sps

Die meisten Aufgaben könnten einfach mit der Maus durchgeführt werden. SPSS bietet aber eine **leistungsfähige Befehlssprache**, mit deren Hilfe Sie fast alle häufig durchzuführenden Aufgaben speichern und automatisieren können und außerdem die gesetzlichen Vorgaben für die Dokumentation von Datenmanagement und statistischer Auswertung erfüllen können.

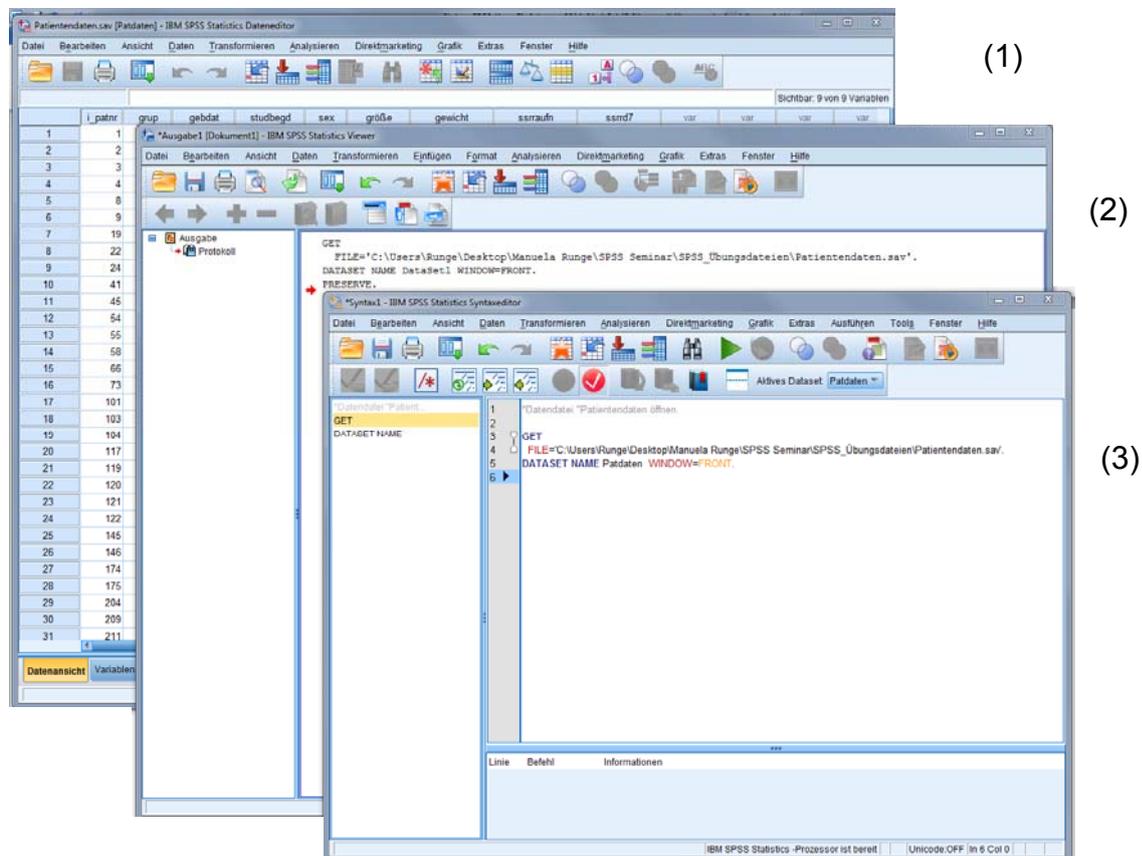


Abbildung 2: Fenster in SPSS



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.

- ⇒ Starten Sie zunächst mit den Variablennamen in der Variablenansicht in der 1. Spalte unter Einhaltung der grundlegenden Richtlinien für Variablennamen (siehe 3.1 S.10).
- ⇒ Dann unter Datenansicht die gewünschte Datenzeile mit der Maus anwählen und die Werte in die Bearbeitungszeile eingeben. Die Bewegung innerhalb der Datenzellen kann mittels Pfeiltasten erfolgen.

Dies wird hier beispielhaft anhand einer Übungsdatei dargestellt [Datendatei **Stamdat.sav**]

Das Resultat kann dann im vorliegenden Beispiel wie folgt aussehen (siehe Abb. 3, S. 9):

*Stamdat.sav [DatenSet3] - IBM SPSS Statistics Daten-Editor

39 : gebdat

	i_patnr	grup	gebdat	sex	grö...	gewic ht	studbegd	alter
1	2	0	14.11.1934	1	169	82	07.12.1997	63,06
2	3	0	01.12.1920	0	162	63	18.01.1998	77,13
3	7	0	07.01.1950	1	164	90	07.02.1998	48,08
4	8	0	04.03.1926	1	184	95	12.02.1998	71,95
5	10	0	18.09.1913	0	165	69	28.02.1998	84,45
6	12	0	13.11.1924	0	150	56	14.04.1998	73,42
7	13	0	27.12.1914	0	.	82	15.05.1998	83,38
8	14	0	15.11.1937	0	165	86	28.05.1998	60,53
9	18	0	18.12.1944	1	166	85	09.07.1998	53,56
10	19	0	01.06.1941	1	176	95	17.07.1998	57,13
11	23	0	02.02.1929	0	168	76	18.08.1998	69,54
12	24	0	04.01.1926	0	163	63	03.12.1998	72,91
13	25	0	05.06.1941	1	178	90	31.05.2000	58,99
14	37	0	26.05.1917	0	166	82	05.02.1999	81,70
15	43	0	20.09.1953	1	180	80	05.05.1999	45,62
16	44	0	06.06.1945	1	187	99	04.06.1999	53,99
17	45	0	01.03.1930	0	174	78	22.05.1998	68,22
18	50	0	18.10.1925	0	178	92	30.08.1999	73,86
19	1	1	25.04.1917	1	162	72	09.12.1997	80,62
20	4	1	09.09.1925	1	168	82	30.12.1997	72,31

Abbildung 3: Datenerfassung in der Datendatei Stamdat.sav

Hier erkennt man bereits wichtige Richtlinien:

- ⇒ Jeder Patient wird hier 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.
Tipp: Erstellen Sie sich einen Codeplan für Ihre Variablen in einem separaten Tabellenblatt!
- ⇒ Fehlende Werte (hier Körpergröße von Patient 13) könnten einfach freigelassen werden. SPSS setzt für systemdefinierte fehlende Werte automatisch ein „.“.



Liegen bei Variablen keine Antworten vor, ist es sinnvoll hierfür **Kodierungen zu vergeben** (z. B. Antwortverweigerung, Unwissenheit, Verzogen, andere Gründe). Im Datenmanagement sind feste Zahlenbereiche vorgesehen: z.B. -9 oder 9 oder 999 (je nach Variablentyp und Format), für fast alle metrische Werte (z. B. Größe, Gewicht) setzen Sie eine -9 (Ausnahme: Laborwerte, hier 999 erfassen). **Dieser Zahlenwert darf unter den natürlich auftretenden Werten nicht vorkommen.**

- ⇒ Die Variablennamen sind seit Version 12 schon längst 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 (10-15 Zeichen).**

Umfassende Informationen zum Dateiaufbau mittels Excel und/oder SPSS erhalten Sie auf unserer Website unter Dienstleistungen/Download-Bereich:

[SOP Eingabe von Daten unter Excel \(pdf\)](#)

Wichtig für Variablennamen:

- bis zu 64 Zeichen möglich. **Sinnvoll: kurz halten!** (10-15 Zeichen sollten reichen)
- jeder Variablenname muss **eindeutig** sein
- **kein** Leerzeichen erlaubt!
- darf **nicht** mit einer Zahl beginnen!
- können aus Buchstaben und Ziffern bestehen
- keine Sonderzeichen, Ausnahmen: `_ . @ #`
- das letzte Zeichen sollte ein Buchstabe oder eine Zahl sein
- Groß- und Kleinschreibung ist nicht relevant
- `$`: steht bei SPSS für Systemvariablen! Nicht nutzen! (z.B. Filter_`$`)
- Bsp. für gültige Variablennamen:

sex gebdat var3_1_2 nuebz sysRRd7 hdl_log

2.2 Datenimport

In der Praxis kommt es häufig vor, dass man die Daten in einem anderen Programm eingegeben hat und nun in SPSS importieren möchte. Aufgrund seiner weiten Verbreitung findet dabei Microsoft Excel besonders häufig 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.de/institute/biometrie/downloads/institut-medizinische-biometrie-epidemiologie/SOP_Datenerfassung_in_Excel_SPSS.pdf).

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

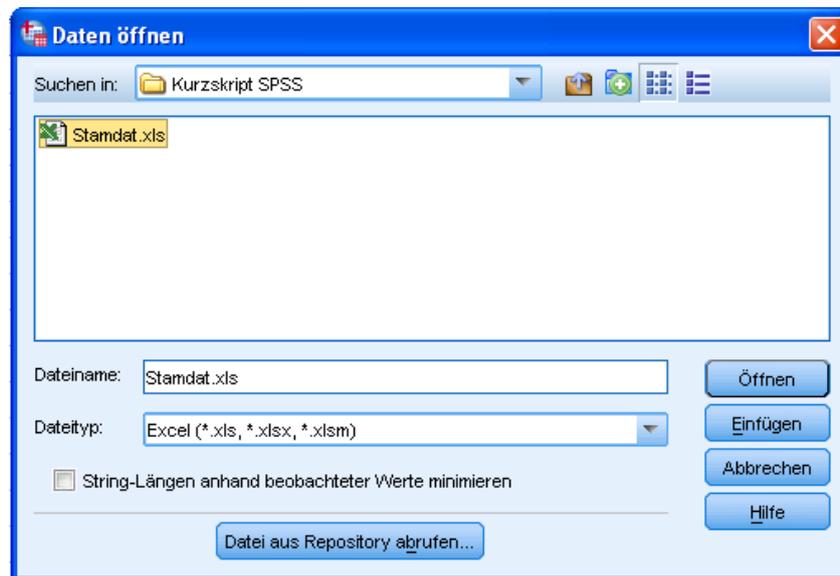


Abbildung 4: Dialogfenster Daten öffnen

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 aus Ihrem Verzeichnis aus und betätigen **Öffnen**, so erhalten Sie folgenden Dialog:

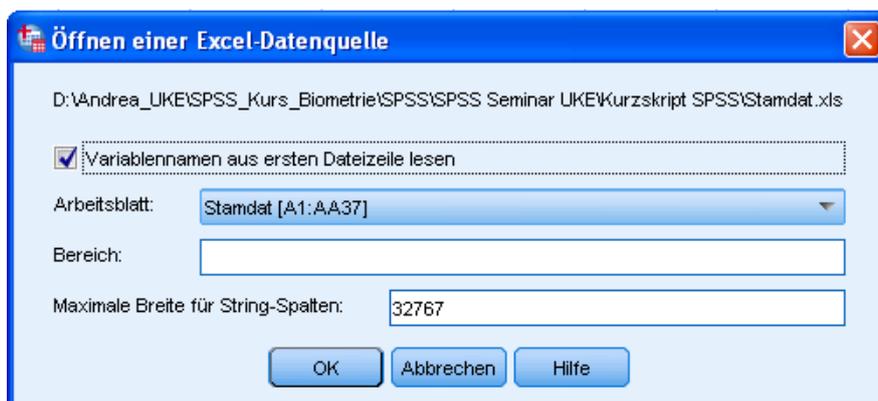


Abbildung 5: Dialogfenster Öffnen einer Excel-Datenquelle



Sie können hierbei zusätzlich anwählen, welche Tabelle der Excel-Datei Sie importieren wollen, und ob die Variablennamen 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, insbesondere bei Datumsvariablen.

Wichtig: Aus der Exceltabelle muss vor Import alle Layouteinstellungen auf die Standardeinstellung zurück gestellt werden (keine Farben, keine Rahmen, kein Fettdruck, keine Schattierungen u.a.). Außerdem muss die zu importierende Exceldatei geschlossen sein!

3. Datenweiterverarbeitung

Sind die Daten in das SPSS-Datenfenster importiert, so ist es in der Regel zunächst nötig, einige weitere Vorbereitungen für die eigentliche Datenauswertung zu treffen.

3.1 Variablendefinition und -formatierung in der Variablenansicht

Der erste Schritt der Variablendefinition sollte in der Vergabe von **Labels** (Etiketten) bestehen. Wie bereits zuvor erwähnt können für SPSS-Variablennamen 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.



Auszugsweise sieht diese am Beispiel folgendermaßen aus (siehe Abb. 6, S. 13):

	Name	Typ	Spaltenf...	Dezimal...	Variablenlabel	Wertelabels	Fehlende W...	Spalten	Messniveau
1	i_patnr	Numerisch	3	0	i_PatNr	Keine	Keine	7	Skala
2	grup	Numerisch	2	0	Therapie	{0, Placebo}...	Keine	8	Nominal
3	gebdat	Datum	10	0	GebDat	Keine	Keine	16	Skala
4	sex	Numerisch	2	0	Geschlecht	{0, weiblich}...	Keine	8	Nominal
5	größe	Numerisch	3	0	Größe	Keine	Keine	3	Skala
6	gewicht	Numerisch	3	0	Gewicht	Keine	Keine	3	Skala
7	studbegd	Datum	10	0	StudBegD	Keine	Keine	14	Skala
8	alter	Numerisch	8	2		Keine	Keine	8	Skala

Abbildung 6: Variablenansicht am Beispieldatensatz Stamdat.sav

Die einzelnen Variablen des Datensatzes sind zeilenweise aufgelistet, jedes einzelne Feld kann durch Anklicken mit der Maus bearbeitet und verändert werden. Alternativ ist dies durch die Menüführung **Daten** → **Variableneigenschaften definieren** möglich.

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

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). Relevant ist die Dezimalstellenangabe.

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 10-15 Zeichen sein können („möglich“ sind bis zu 256 Zeichen) und auch Sonderzeichen enthalten dürfen. Auch hier gilt: so kurz wie möglich und lang wie nötig!

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 grauer Button, mit welchem folgender Dialog aufgerufen wird (siehe Abb. 7, S. 14):

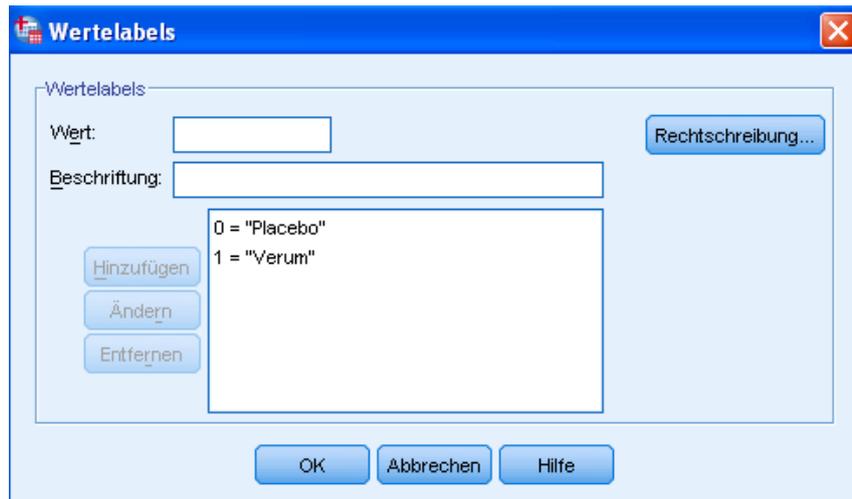


Abbildung 7: Dialogfenster Wertelabels

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 kategoriemer Variablen (also hier z. B. Geschlecht und Behandlungsgruppe) ist zwar manchmal etwas zeitaufwändig, aber durchaus 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 fehlender Wert**, 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 oder 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 (nominal, ordinal, metrisch). Bei numerischen Variablen ist das Messniveau **Metrisch**, bei alphanumerischen Variablen **Nominal** voreingestellt.

3.2 Neuberechnung von Variablen

Um im vorliegenden Beispiel die ursprüngliche Fragestellung (vgl. S. 8) 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 (siehe Abb. 8, S. 15):

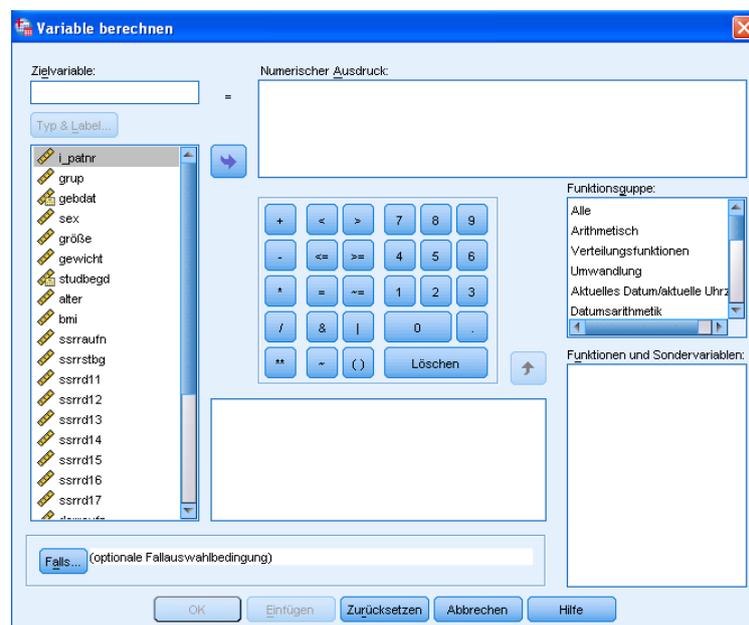


Abbildung 8: Dialogfenster Variable berechnen



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 in Funktionsgruppen unterteilte (alphabetisch geordnete) große Auswahl an mathematischen und statistischen Funktionen, deren einzelne Erläuterung hier zu weit führen 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 blau aufleuchtenden Pfeils in das Feld **numerischer Ausdruck** befördert werden.

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

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 wie folgt aussehen (siehe Abb.9, S. 17):



Abbildung 9: Dialogfenster Berechnung des BMI

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

	Name	Typ	Spaltenf...	Dezimal...	Variablenlabel	Wertelabels	Fehlende W...	Spalten	Messniveau
1	i_patnr	Numerisch	3	0	i_PatNr	Keine	Keine	7	Skala
2	grup	Numerisch	2	0	Therapie	{0, Placebo}...	Keine	4	Nominal
3	gebdat	Datum	10	0	GebDat	Keine	Keine	8	Skala
4	sex	Numerisch	2	0	Geschlecht	{0, weiblich}...	Keine	8	Nominal
5	größe	Numerisch	3	0	Größe	Keine	Keine	4	Skala
6	gewicht	Numerisch	3	0	Gewicht	Keine	Keine	5	Skala
7	studbegd	Datum	10	0	StudBegD	Keine	Keine	10	Skala
8	alter	Numerisch	8	2	Alter	Keine	Keine	6	Skala
9	bmi	Numerisch	8	2	BMI	Keine	Keine	8	Skala

Abbildung 10: Neue Variable „bmi“ in der Variablenansicht

Die neue Variable „bmi“ befindet sich nun an letzter Stelle der Variablenansicht in der Datei. In der Datenansicht existiert für Probandin 13 kein BMI, dieser konnte nicht errechnet werden, da die Angabe ihrer Körpergröße fehlt.

i_patnr	grup	gebdat	sex	größe	gewicht	studbegd	alter	bmi
13	0	27.12.1914	0	.	82	15.05.1998	83,38	.

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 welchen 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:

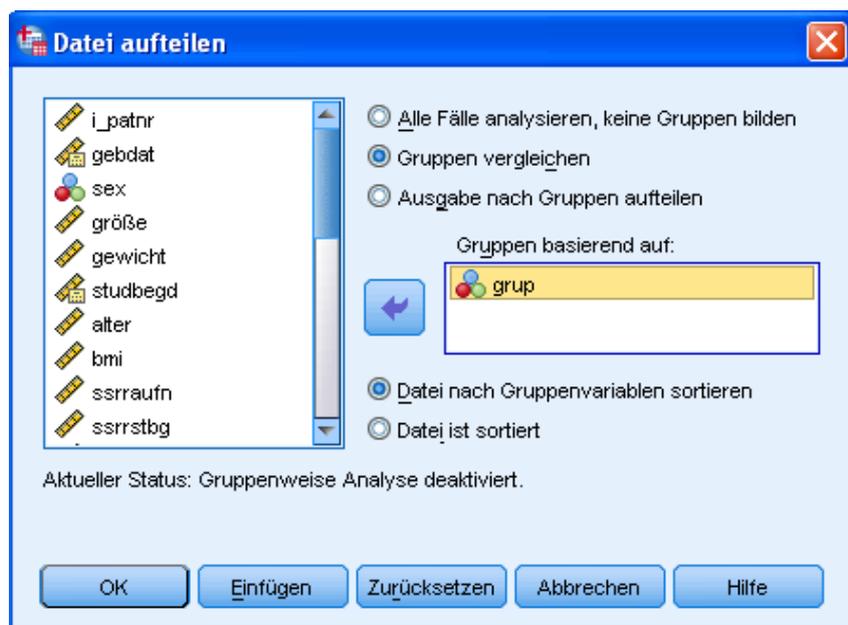
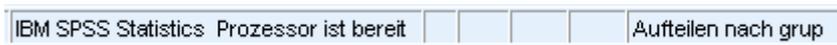


Abbildung 11: Dialogfenster Datei aufteilen

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 wird die Variable „*grup*“, in das Feld **Gruppen basierend auf** gesetzt.

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



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 (Alle Fälle analysieren, keine Gruppen bilden)!

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

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



Abbildung 12: Dialogfenster Deskriptive Statistik

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 gezielt bestimmte Funktionen und Maßzahlen anzufordern.

Die Standardeinstellung besteht in der Angabe von Mittelwert, 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, und auch Fehlermeldungen schreibt. Dieses Fenster wird im folgenden Abschnitt kurz erläutert.

Eine größere Auswahl an statistischen Maßzahlen bekommt man unter **Deskriptive Statistiken** und dann **Häufigkeiten**.

Mittels des Buttons **Diagramme** kann außerdem ein Histogramm ausgewählt werden. Über den Button **Statistiken** können zusätzlich Perzentilwerte, Lage-, Streuungs- und Verteilungsmaße ausgewählt werden (u.a. Median, Quartile, Standardabweichung (siehe Abb. 14, S. 21).



Abbildung 13: Dialogfeld Häufigkeiten

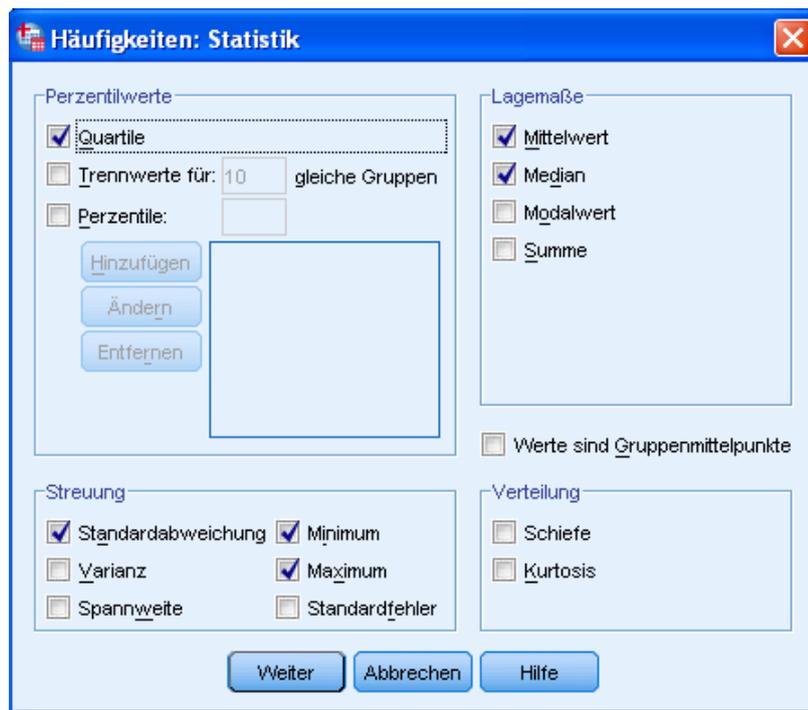


Abbildung 14: Dialogfeld Häufigkeiten: Statistik

4.2 Das SPSS-Ausgabefenster

Das sich neu öffnende **SPSS-Ausgabefenster** sieht wie folgt aus:

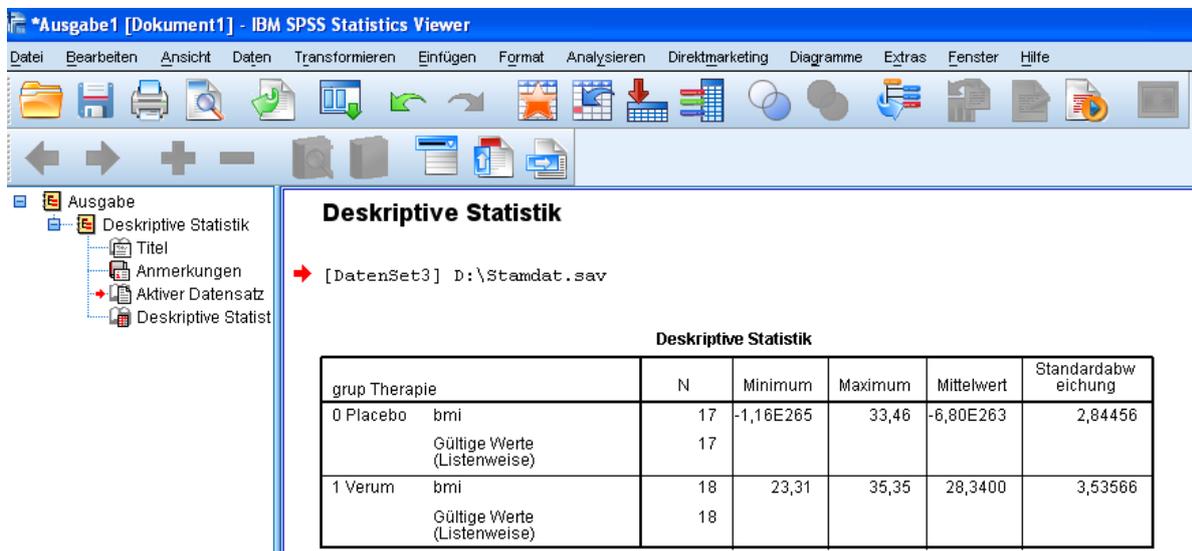


Abbildung 15: SPSS-Ausgabefenster

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.

Um eine **Tabelle** an persönliche Präferenzen (Schriftart, Rahmenlinien etc.) anzupassen, kann mittels Doppelklick in den **Bearbeitungsmodus** gewechselt werden. Hier gibt es zahlreiche Bearbeitungsmöglichkeiten, mit welchen man sich am einfachsten durch selbständiges Probieren vertraut machen kann.

Tipp: unter **Tabelleneigenschaften**→**Drucken** lassen sich große Tabellen passend auf die Seitenbreite/-länge verkleinern.

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

Objekt-Einbindung in Word oder Power Point:

Unter dem Menüpunkt **Daten** mittels des Befehl **Export** besteht die Möglichkeit die Ausgabedatei als Portable Document Format (*.pdf) zu exportieren. Hier bietet SPSS außerdem die Möglichkeit Diagramme als JPEG-Datei (*.jpg) zu speichern.

Um eine Tabelle wie die vorliegende z. B. in ein Word-Dokument oder eine Power-Point-Präsentation einzubinden, besteht natürlich auch die Möglichkeit von „copy and paste“. Klicken Sie die Tabelle mit der rechten Maustaste an und wählen **Kopieren** (oder Kopieren spezial). Damit wird die Tabelle in die Zwischenablage übernommen und kann in andere Windows-Anwendungen direkt mittels **Bearbeiten/Einfügen** übernommen werden.

4.3 Grafische Darstellung mittels Boxplot

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 **Boxplot** erstellt werden kann, welcher die Aussage der vorher erstellten Tabelle unterstützt und ergänzt.

Ein Boxplot dient vor allem zur kompakten Visualisierung der Verteilung numerischer Variablen (Details s. u.).

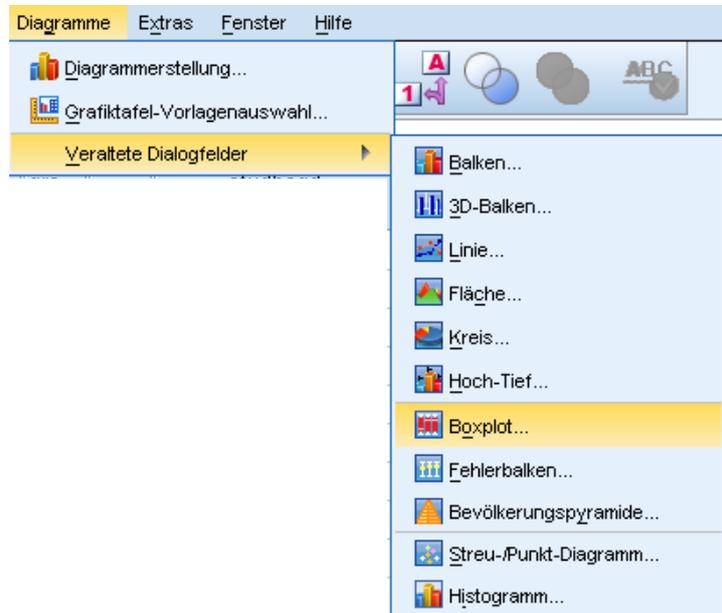


Abbildung 16: Auswählen des Box-Plots über die Menüführung

Nach der Wahl von **Diagramme** → **Veraltete Dialogfelder** → **Boxplot** und Anklicken von **Definieren** erscheint wiederum ein SPSS-Dialogfenster (siehe Abb. 17, S. 23):

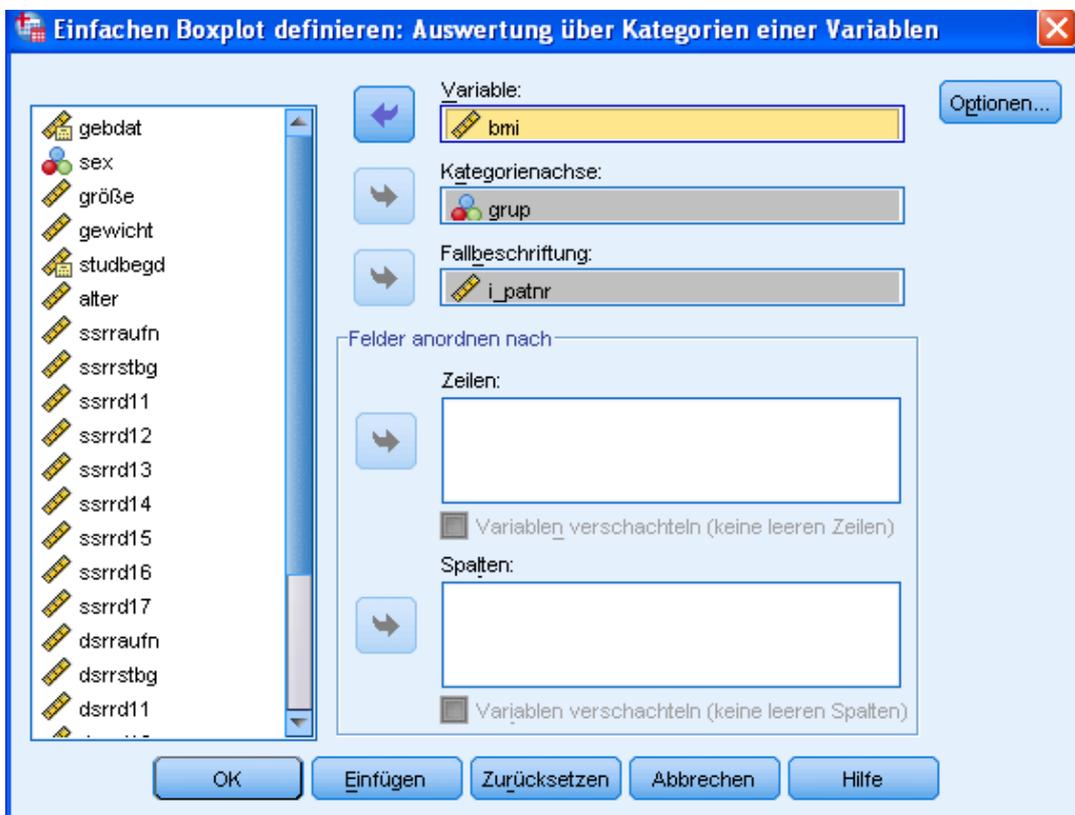


Abbildung 17: Dialogfeld Einfachen Boxplot definieren

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 (siehe Abb.18 S. 24).

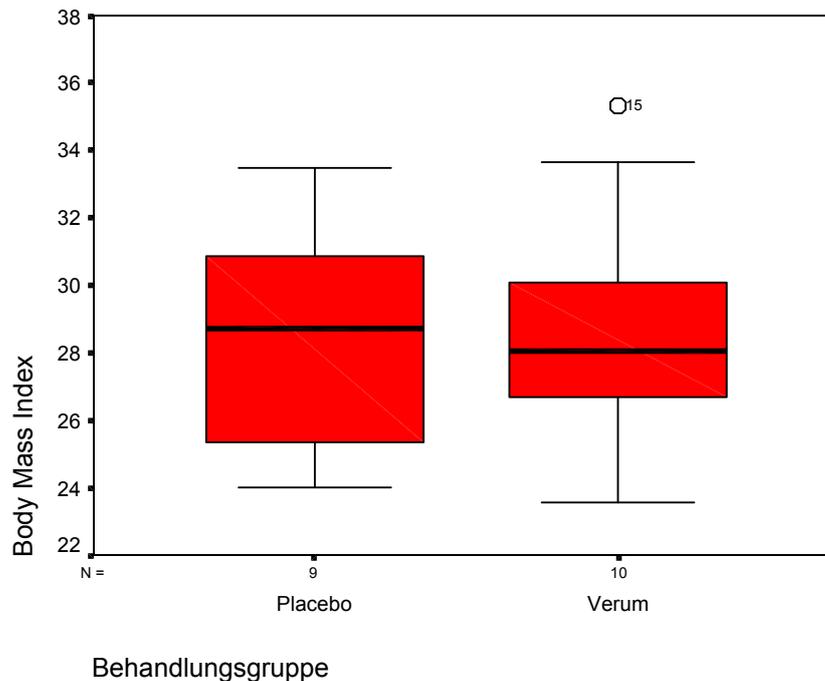


Abbildung 18: Box-Plot BMI im Vergleich zwischen den Behandlungsgruppen

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 Box beschreibt die mittlere Hälfte der Daten und der schwarze Strich innerhalb dieser Box stellt den Median dar. Dieser wird auch „zentraler Wert“ genannt und ist definiert als der Wert, für den die eine Hälfte der Probanden einen größeren und die andere Hälfte einen kleineren BMI hat.
- ⇒ Des Weiteren erkennen Sie, dass in der Verumgruppe eine Person (Pat_ID = 15) einen ungewöhnlich hohen BMI-Wert aufweist (sog. Ausreißer).

Auch diese Grafik kann durch Doppelklick in vielen Einzelheiten bearbeitet werden (z. B. Achsenskalierung und Beschriftung, Titel, Farbe etc.). Der Transfer in andere Windows-Anwendungen funktioniert exakt wie für die Tabellen beschrieben (siehe S. 22).

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.

5.1 SPSS-Syntaxfenster

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

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 ausgeführt wird, sondern in die zugehörige **SPSS-interne Syntax** (das sog. **Syntax-Fenster/Syntax-Datei**) geschrieben wird. Diese Syntax kann dann mittels der im Syntax-Fenster befindlichen „Playtaste“ an SPSS übermittelt und ausgeführt werden.



Warum Syntax?

In der Wissenschaft wird mit reproduzierbaren Ergebnissen gearbeitet!

Wie? → Nur durch die Nutzung von Syntaxprogrammen!

Der Vorteil des Arbeitens mit der Syntax, besteht weiterhin darin, dass:

- alle erforderlichen Arbeitsschritte einer Auswertung übersichtlich und **nachvollziehbar zusammengestellt und gespeichert** werden können
- mehrmaliges Wiederholen des gleichen Befehls deutlich **beschleunigt** wird
- diese bei Bedarf schnell mit **neuen Daten** oder **korrigierten Daten** erneut laufen kann
- **alle Bearbeitungsschritte automatisch dokumentiert** sind.
- die gesetzliche Pflicht in der Klinischen Forschung / Medizinischen Dokumentation erfüllt wird (**GCP, GCDMP, GEP, Guidelines der FDA ...**)!



Eine kurze Warnung sei an dieser Stelle abschließend hinzugefügt. Selbst die beste statistische Software bewahrt den Benutzer nicht davor, Ergebnisse falsch zu interpretieren, wenn nur mangelndes Verständnis über die hinter den einzelnen Prozeduren stehenden Methoden vorhanden ist. Eine Vermittlung dieser Inhalte ist u.a. Ziel der Biometrie-Veranstaltung im Studium Humanmedizin (<http://www.uke.de/institute/biometrie>).

5.2 Mehr Lernen:

- Wahlfachangebot des IMBE (jeweils im 2. Trimester):
„**Biometrie und Statistik für Studie und Doktorarbeit**“
- Wahlfachangebot mit dem ZEM (jedes Trimester):
„**Statistische Auswertungskonzepte für Laborversuche**“
- Vertiefende Statistische Weiterbildungsangebote:
im Rahmen des Phd-Programms für Nichtmediziner bietet das IMBE zwei Themen an (jährlich im Wechsel):
 - „**Analysis of Events and Time-to-Event - Logistic Regression and Survival Models**“
 - „**ANOVA, ANCOVA, Regression**“

Termine und Anmeldung unter: imbe-sekretariat@uke.de

- **Biometrische Beratung** im IMBE (W34) (Voranmeldung erforderlich!)
 - statistische Beratungen nach individueller Terminvereinbarung
 - kostenfreie Beratung für alle UKE-Angehörige, Doktoranden und Studierende
 - Zur Vereinbarung eines Beratungstermins Anmeldung ausfüllen, innerhalb von aktuell 14 Tagen erhalten Sie einen Terminvorschlag.

Das Formular für die Anmeldung zur statistischen Beratung finden Sie auf unserer Website unter Dienstleistungen, dort rechts unter Downloads

http://www.uke.de/institute/biometrie/index_2148.php.

Bei Fragen wenden Sie sich bitte an unser Institutssekretariat:

Renate Reche

Tel. (040) 7410 – 57485

imbe-sekretariat@uke.de



6. 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, wurden auch entsprechende Werke mit aufgenommen. Die Unterschiede zwischen den SPSS-Versionen 18.0, 19.0 und 20 sind primär nicht struktureller Natur, sondern betreffen weitestgehend Detailfragen, deshalb ist es durchaus möglich, auch ein Buch, welches sich auf eine der zwei Vorgängerversionen von SPSS bezieht, als Referenzwerk zu nutzen.

SPSS

- Brosius F. [2013]: *SPSS 21 (mitp Professional)*, ed. mitp; 1. Auflage, gebunden
- Bühl A. [2011]: *SPSS 20: Einführung in die moderne Datenanalyse*, Pearson Studium; 13. Auflage, aktualisierte Auflage, gebunden
- Brosius F. [2011]: *SPSS 19 (mitp Professional)*, mitp, 1. Auflage, gebunden
- Brosius F. [2010]: *SPSS 18 (Für Dummies)*, Wiley-VCH; 1. Auflage
- Bühl A. [2009]: *SPSS 18 (ehemals PASW). Einführung in die moderne Datenanalyse*, Pearson Studium, München, 12. Auflage, gebunden.
- Janssen, J. und Laatz, W., [2013] *Statistische Datenanalyse mit SPSS: Eine anwendungsorientierte Einführung in das Basissystem und das Modul Exakte Tests*, Springer-Gabler-Verlag, 8. Auflage
- Eckstein P. [2012]: *Angewandte Statistik mit SPSS. Praktische Einführung für Wirtschaftswissenschaftler*. Springer-Gabler-Verlag, 7. Auflage
- Tutorials im Internet, z.B. Statistik und SPSS Tutorial online: <http://www.statistik-tutorial.de/>

SPSS-Syntax:

- Sarstedt M., Schütz T., Raithel S., [2010]: *IBM SPSS Syntax: Eine anwendungsorientierte Einführung*, Vahlen Franz GmbH, 2. Auflage
- Brosius F., [2008]: *SPSS-Programmierung - Effizientes Datenmanagement und Automatisierung mit SPSS-Syntax*, Mitp-Verlag, Bonn, 2. Auflage



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

Zusätzliche Hilfestellung bei praktischen Problemen mit SPSS finden Sie auf der Homepage <http://www-01.ibm.com/software/de/analytics/spss/>.

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 <http://campus.uni-muenster.de/3645.html>.

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

- Gaus W., Muche R., [2013]: *Medizinische Statistik: Angewandte Biometrie für Ärzte und Gesundheitsberufe*, Schattauer, 1. Auflage
- Harms V. [2012]: *Medizinische Statistik*, 8. völlig neu bearbeitete Auflage
- Sachs L.; Hedderich J., [2012]: *Angewandte Statistik: Methodensammlung mit R*, Springer, Berlin, 14.Auflage
- Weiß C. [2013]: *Basiswissen Medizinische Statistik*, Springer –Verlag, 6.Auflage
- Altman DG. [2009]: *Practical Statistics for Medical Research*. Chapman & Hall, London., 2.Auflage
- Hilgers R.; Bauer P. Scheiber V., [2006]: *Einführung in die Medizinische Statistik*, Springer Berlin, 2.Auflage
- Clauß G. und Finze F.-R. und Partsch L., [2011]: *Statistik für Soziologen, Pädagogen, Psychologen und Mediziner, Band I*. Harri Deutsch, 6. Auflage
- Armitage P. und Berry G. [2001]: *Statistical Methods in Medical Research*. Blackwell, 4.Auflage
- Trampisch H. und Windeler J.[2000]: *Medizinische Statistik*. Springer, 2.Auflage

Autoren: IMBE

Erstellt: 18.03.2002; aktualisiert 11.08.2014 Kontakt: grosser@uke.de (Andrea Großer)