

## Biometrie – Übungen in SPSS

### Veranstaltung im Rahmen der Junior Class des Studiengangs Experimentelle Medizin

Univ.-Prof. Dr. rer. nat. et med. habil. Andreas Faldum  
Institut für Biometrie und Klinische Forschung (IBKF)  
Schmeddingstr. 56  
48149 Münster  
[www.ibkf.uni-muenster.de](http://www.ibkf.uni-muenster.de)

Ihre Ansprechpartnerin: Dipl.-Stat. Maria Eveslage  
[maria.eveslage@ukmuenster.de](mailto:maria.eveslage@ukmuenster.de)

### Organisatorisches

Das vorliegende Skript soll Ihnen helfen, grundlegende Analysen mit SPSS eigenständig durchzuführen (Stand: IBM® SPSS® Statistics Version 27.0<sup>1</sup>). Die Software SPSS können Sie über das ZIV kostenfrei beziehen. Nähere Informationen finden Sie unter

<https://www.uni-muenster.de/IT/services/arbeitsplatz/software/spss.html>.

Darüber hinaus bieten sowohl das ZIV als auch das IBKF SPSS-Schulungen an.

Es werden Termine zur betreuten Bearbeitung von Übungsaufgaben oder der Analyse eigener Daten angeboten, bei denen ein Biometriker Sie beraten und Ihre Fragen beantworten wird. Die Termine finden i.d.R. im Schulungsraum des IBKF (Raum 314) statt. Es stehen 15 Rechnerarbeitsplätze bereit, Sie benötigen zur Anmeldung Ihre ZIV-Nutzerkennung.



<sup>1</sup> IBM SPSS® Statistics 27 for Windows (IBM Corporation, Somers, NY, USA)

**Inhaltsverzeichnis**

- 1. Tipps zur Dateneingabe und Datenaufbereitung ..... 3**
  - 1.1 Aufbau des Datensatzes ..... 3**
  - 1.2 Plausibilitätschecks ..... 4**
- 2. Benutzeroberfläche von SPSS..... 5**
  - 2.1 Daten-Editor ..... 5**
  - 2.2 Variablenansicht ..... 5**
  - 2.3 Der Ausgabe-Viewer ..... 6**
  - 2.4 Der Syntaxeditor ..... 6**
  - 3.1 Öffnen und Speichern einer Datendatei ..... 7**
  - 3.2 Öffnen und Speichern einer Ausgabedatei ..... 9**
  - 3.3 Zusammenfügen von Datendateien (fall- und variablenweise) ..... 9**
    - 3.3.1 Fallweises Zusammenfügen ..... 10**
    - 3.3.2 Variablenweises Zusammenfügen ..... 11**
  - 4.1 Auswahl von Fällen ..... 13**
  - 4.2 Fälle sortieren ..... 14**
  - 4.3 Aufteilung der Fälle in Gruppen ..... 14**
  - 5.1 Berechnung von neuen Variablen ..... 15**
  - 5.2 Rechnen mit Datumsangaben ..... 15**
  - 5.3 Umkodieren und Klassieren von Werten ..... 18**
  - 5.4 Aggregieren von Daten ..... 20**
- 6. Deskriptive Statistik ..... 22**
  - 6.1 Häufigkeitstabellen kategorialer Merkmale ..... 22**
  - 6.4 Graphische Darstellungen eines qualitativen Merkmals ..... 29**
  - 6.5 Graphische Darstellungen eines quantitativen Merkmals ..... 30**
  - 6.6 Streudiagramme und Korrelationsanalysen ..... 32**
- 7. Statistische Signifikanztests zum Vergleich zweier Verteilungen ..... 36**
  - 7.1 Chi-Quadrat-Test und exakter Test nach Fisher ..... 36**
  - 7.2 Parametrische Tests bei normalverteilten Zielgrößen ..... 38**
    - 7.2.1 t-Test für unverbundene Stichproben ..... 38**
    - 7.2.2 t-Test für verbundene Stichproben ..... 39**
  - 7.3 Nichtparametrische Tests ..... 41**
    - 7.3.1 Mann-Whitney-U Test (Wilcoxon-Rangsummen-Test) ..... 41**
    - 7.3.2 Wilcoxon-Vorzeichen-Rang Test und Vorzeichentest ..... 43**

<b>8.</b>	<b>Lineare Regression</b> .....	45
<b>9.</b>	<b>Logistische Regression</b> .....	47
<b>10.</b>	<b>Time-to-event Analysen</b> .....	53
<b>10.1</b>	<b>Kaplan-Meier Methode und Log-Rank Test</b> .....	53
<b>10.2</b>	<b>Cox-Regression</b> .....	57

## 1. Tipps zur Dateneingabe und Datenaufbereitung

Folgende Hinweise sollen das Einlesen und Auswerten der Daten mit SPSS erleichtern. Sie können beispielsweise Ihre Daten in Excel erheben und diese bei geeignetem Format in SPSS einlesen und dann dort aufbereiten und weiterverarbeiten. Sie können die Daten auch direkt in SPSS erheben.

### 1.1 Aufbau des Datensatzes

- Stets Datum des Datenstandes vermerken, z.B. im Dateinamen
- In der Regel eine Zeile pro Patient/Proband/Probe (Ausnahme: longitudinale Daten / Messwiederholungen, dann abklären) und eine Spalte pro Variable
- Eindeutige ID in separater Spalte
  - o **DON'T**: sich auf Zeilennummer (z.B. in Excel) verlassen. Immer ID-Spalte anlegen.
  - o **DON'T**: Namen von Probanden / Akronym verwenden (die Daten müssen mindestens pseudonymisiert sein)
  - o **DON'T**: die gleiche ID für mehrere Patienten (z.B. in verschiedenen Therapiegruppen) verwenden
  - o ID muss in allen Dateien / Tabellenblättern eindeutig sein
- Bei Erhebung in Excel nach Möglichkeit nicht zu viele Tabellenblätter anlegen, da beim Einlesen in SPSS stets nur ein einzelnes Tabellenblatt eingelesen wird.

### Variablennamen:

- Variablennamen kurz und eindeutig, Sonderzeichen und Umlaute vermeiden, Ausnahme: \_ Unterstrich
- In Excel: Variablennamen in erster Zeile
- **DON'T**: Codierung in Variablennamen einbauen
- Eindeutige Variablennamen verwenden (nicht den gleichen Namen mehrfach verwenden, Messwiederholungen geeignet durch Endung (z.B. \_0, \_1, \_2 usw.) kennzeichnen
- **DON'T**: Gleiche Variablen in unterschiedlichen Dateien / Tabellenblättern / Dateiversionen unterschiedlich benennen
- Variablenlabels vergeben (SPSS): kurze, prägnante Beschreibung der Variablen, eindeutig und nicht zu lang

### Codierung:

- Codierung in separater Datei aufführen bzw. Wertelabels vergeben (in SPSS)
- Kategorien möglichst mit Zahlen (nicht mit Buchstaben! schon gar nicht mit Farben!) codieren
- Falls Codierung doch mit Text durchgeführt werden sollte, Schreibweise beachten (Groß-/ Kleinschreibung / Leerzeichen usw. ), nach Möglichkeit vordefinierte Listen verwenden
- Logische Codierung, z.B. bei ja/nein-Variablen Code „0 = nein, ja = 1“ verwenden, bei ordinalen Variablen Zahlencodes gemäß der Ordnung der Kategorien verwenden
- Falls fehlende Werte codiert werden sollen: „eindeutig andere Zahl“ verwenden (Bsp: -999, falls dieser Wert in Variablen nicht möglich ist), dann diesen Wert in SPSS als benutzerdefinierten fehlenden Wert einstellen

### Dateneingabe:

- **DON'T**: In Variablen Zahlen und Text mischen
- **DON'T**: Freitexte verwenden („keine Prosa“, nur als Ausnahme zusätzlich zu codierter Variante angeben)
- **DON'T**: mehrere Angaben / Zahlen / Werte pro Feld

- **DON'T:** Fälle / Kontrollen durch Farben kennzeichnen. Farben werden in Programmen wie SPSS nicht gelesen, die Information geht verloren. Die Codierung muss durch eigenständige Variable dokumentiert werden
- Datumsangaben einheitlich erfassen, auf korrektes Format achten

## 1.2 Plausibilitätschecks

Es empfiehlt sich, die Daten vor Beginn der eigentlichen Analyse zu prüfen, um Fehler vor Beginn der Analysen zu korrigieren und so häufiges Wiederholen / Korrigieren der Analysen zu vermeiden.

- In SPSS: Variablen der Größe nach sortieren, größte und kleinste Werte prüfen (alternativ: Minimum, Maximum berechnen, Boxplot machen, ...)
- Bei Umcodieren / Bilden von Kategorien: Kreuztabelle mit ursprünglicher Variable anlegen und prüfen
- Variablen mit Datumsangaben: der Größe nach sortieren und prüfen, ob Werte zu weit in Vergangenheit / Zukunft liegen
- Kategoriale Variablen (und evtl. auch metrische Variablen): Häufigkeitstabelle anlegen und prüfen. Ist die Anzahl fehlender Werte ok?
- Falls Werte einer Variablen von anderer Variablen abhängen, dann Kreuztabelle anlegen und prüfen (Bsp. Raucher ja/nein vs. Anzahl Zigaretten / Tag – Nichtraucher sollten einen fehlenden Wert/eine Null aufweisen)

## 2. Benutzeroberfläche von SPSS

SPSS bezeichnet eine Software für statistische Datenanalysen. Es gibt zwei Auswahlsichten: den Daten-Editor (Datenansicht) und die Variablenansicht. Bei Starten des Programms öffnet sich zudem ein Ausgabe-Fenster.

**Hinweis:** Im Folgenden werden verschiedenen Beispieldatensätze verwendet, um Prozeduren und Ausgaben anhand verschiedener Variablen zu veranschaulichen.

### 2.1 Daten-Editor

Der Dateneditor, ein in Zeilen und Spalten aufgeteiltes Arbeitsblatt, ermöglicht das Öffnen einer Datei, das Eingeben, Bearbeiten und das Auswerten von Daten. Die einzelnen Zeilen entsprechen den Fällen, die einzelnen Spalten den Variablen. Die Zellen enthalten den Wert der jeweiligen Variable innerhalb eines entsprechenden Falles. Die Auswertung der Daten erfolgt in der Regel über Menübefehle (ggf. Programmsyntax) und diesen zugeordnete Dialogfelder. Die Menüzeile im oberen Bildrand wird durch eine darunter befindliche Symbolleiste ergänzt.

	PID	Therapie	Geschlecht	alter	KGW	KGR	Zustand	Zustand_bin	Fieber	log_Leuko_0	HB_0	
1	P1	...	1	2	60	64	162	1	0	37,41	7,11	6,88
2	P2	...	1	1	39	71	163	1	0	36,89	9,35	8,12
3	P3	...	1	1	71	89	185	1	0	36,03	10,88	8,47
4	P4	...	1	1	61	103	178	1	0	36,53	8,06	15,22
5	P5	...	2	1	65	72	173	1	0	39,08	7,04	8,30
6	P6	...	1	1	62	98	172	0	0	36,91	8,41	9,36

### 2.2 Variablenansicht

Die Variablenansicht dient zur Vergabe von Variablennamen. Dabei unterliegt die Vergabe von Variablennamen bestimmten Festlegungen (keine Leerzeichen, kein Beginn mit einer Ziffer, keine Sonderzeichen wie !, ?, \*, & und .). Weiterhin werden Variablentyp, Spaltenformat, Variablenlabel, Wertelabels und das Messniveau (Skalenniveau) der Variablen definiert.

	Name	Typ	Breite	Dezimals...	Beschriftung	Werte	Fehlend	Spalten	Ausrichtung	Messniveau	Rolle
1	PID	Zeichenfolge	18	0		Ohne	Ohne	8	Links	Nominal	Eingabe
2	Therapie	Numerisch	8	0	Behandlungsgr...	{1, TAD-HA...	Ohne	8	Rechts	Nominal	Eingabe
3	Geschlecht	Numerisch	8	0	Geschlecht	{1, Männlich...	Ohne	8	Rechts	Nominal	Eingabe
4	alter	Numerisch	2	0	Alter bei Diagn...	Ohne	Ohne	10	Rechts	Metrisch	Eingabe
5	KGW	Numerisch	3	0	Körpergewicht i...	Ohne	Ohne	8	Rechts	Metrisch	Eingabe
6	KGR	Numerisch	3	0	Körpergröße in ...	Ohne	Ohne	8	Rechts	Metrisch	Eingabe
7	Zustand	Numerisch	8	0	Allgemeinzusta...	{0, Normal}...	Ohne	8	Rechts	Ordinal	Eingabe
8	Zustand_bin	Numerisch	8	0	Zustand, binär ...	{0, Normal o...	Ohne	8	Rechts	Ordinal	Eingabe
9	Fieber	Numerisch	5	2	Körpertemperat...	Ohne	Ohne	8	Rechts	Metrisch	Eingabe

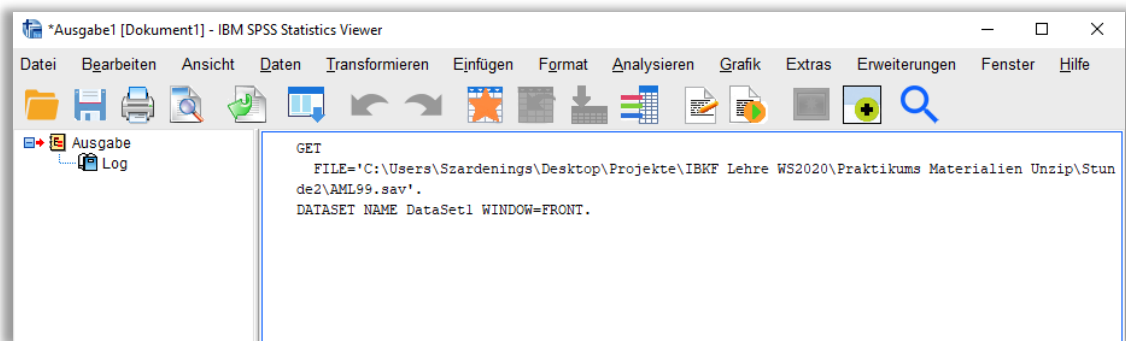
Es empfiehlt sich, im Variablenlabel eine kurze Beschreibung der Variablen zu ergänzen, da Variablenamen nicht immer selbsterklärend sind. Handelt es sich um eine ordinale / nominale Variable, sollten in der Variablen die Kategorien möglichst mit Zahlen codiert erfasst werden. In der Spalte Wertelabels können Beschriftungen für die Codes eingetragen werden. In der Ausgabe werden Variablen- und Wertelabel verwendet.

Wurden fehlende Werte ebenfalls mit einem Code versehen (z.B. -999), kann dieser Code in der Spalte *Fehlende Werte* gekennzeichnet werden. Entsprechende Werte werden bei Analysen dann als fehlend berücksichtigt.

Manche Analysen lassen sich nur mit Variablen eines bestimmten Messniveaus durchführen, daher sollte das Messniveau korrekt angegeben werden. (In einigen Menüfenstern gibt es aber die Möglichkeit, dass Messniveau einer Variablen per Rechtsklick temporär zu ändern.)

### 2.3 Der Ausgabe-Viewer

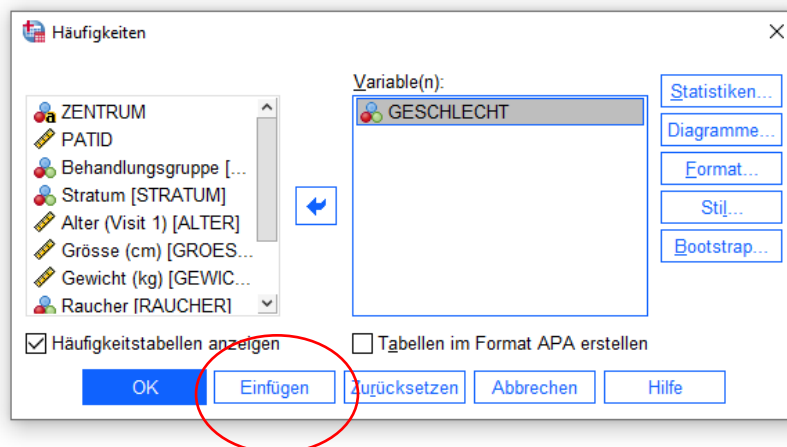
Der Ausgabe-Viewer zeigt die Auswertungsergebnisse an (Tabellen, Statistiken, Diagramme). Der Darstellungsbereich besteht aus zwei Teilen. Im linken Teil wird eine Gliederung aller bisher ausgeführten Analysen als Inhaltsübersicht aufgelistet; der rechte Teil enthält die Ergebnisse mit den Tabellen bzw. Graphiken. Der Ausgabe-Viewer sieht wie folgt aus:

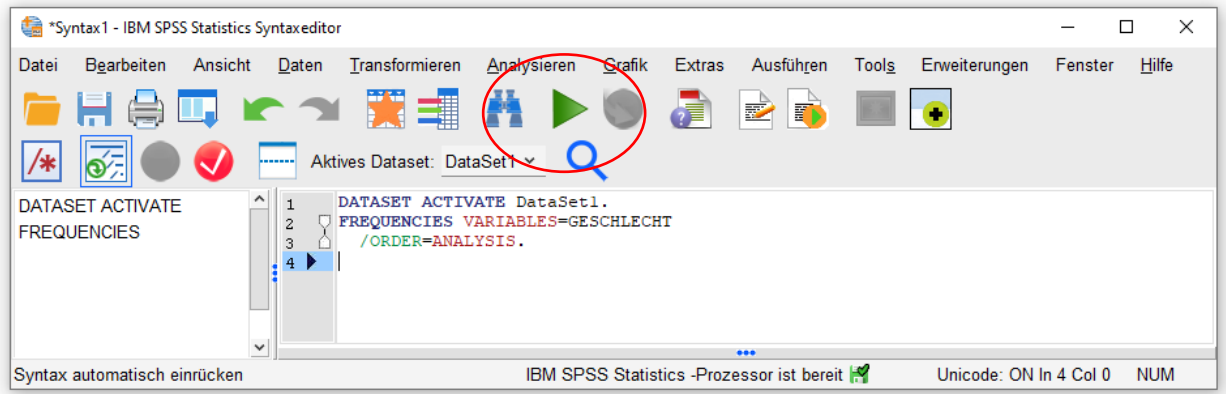


### 2.4 Der Syntaxeditor

SPSS kann über die Menüs (per „Klick“) oder über Programmsyntax (per „Programmiercode“) bedient werden. Über die Befehlsfolge **Datei > Neu > Syntax** öffnet man einen leeren Syntaxeditor.

Syntax kann auch ohne genaue Kenntnis der Programmiersprache recht einfach über die Menüs erzeugt werden. Wählt man per Menü eine Analyse oder eine Datenmodifikation und klickt auf **Einfügen**, so wird der für die Aktion notwendige Programmcode in den Syntaxeditor eingefügt. Der Befehl kann ausgeführt werden, indem der entsprechende Programmcode **markiert** und mit dem **grünen Pfeil** in der Menüzelle ausgeführt wird. Auf diese Weise lassen sich Auswertungen bequem zusammenstellen und in Form der Syntax speichern, so dass sie beispielsweise auf aktualisierte Datensätze mit wenig Aufwand wiederholt werden können.

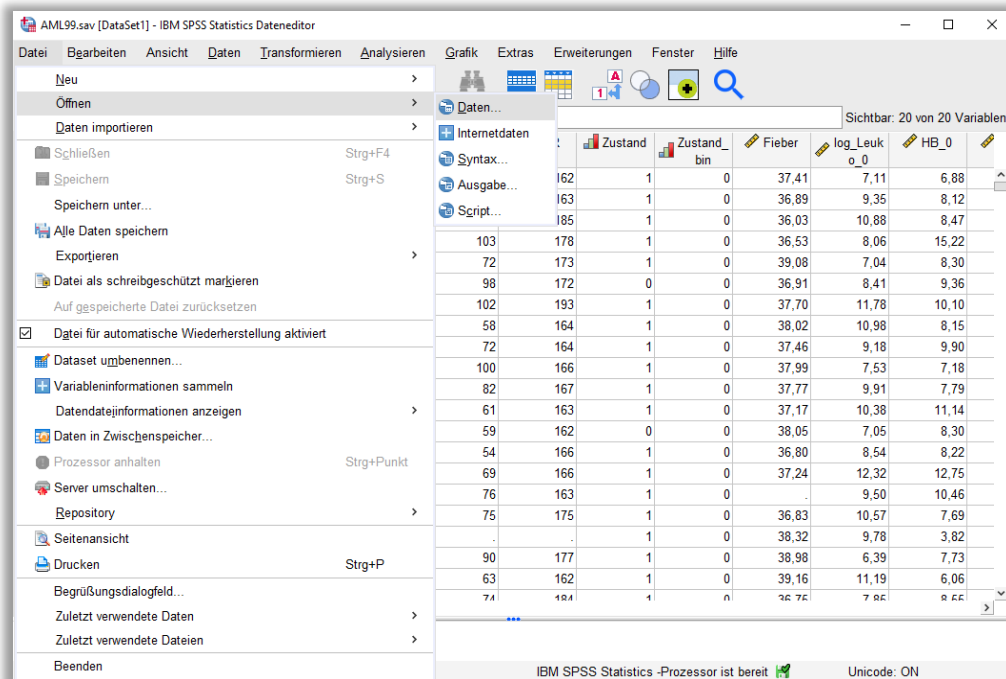




### 3. Datenaufbereitung

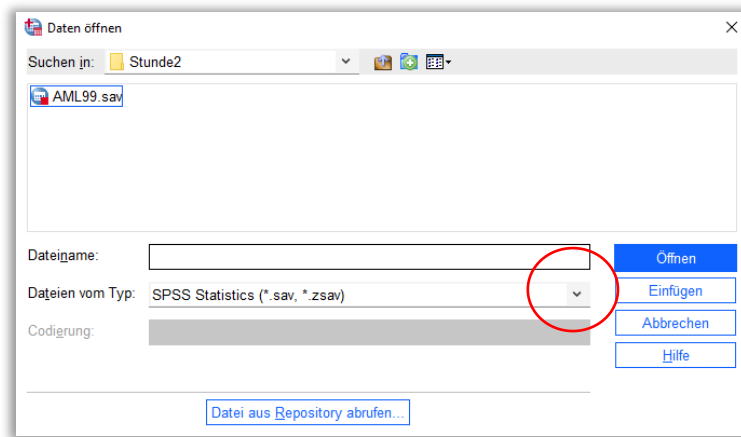
#### 3.1 Öffnen und Speichern einer Datendatei

Starten Sie SPSS für Windows, indem Sie doppelt auf das SPSS-Symbol klicken. Es öffnet sich der SPSS-Daten-Editor. Um eine SPSS-Datendatei zu öffnen, betätigen Sie die Befehlsfolge **Datei > Öffnen > Daten**. Sie gelangen in das Dialogfeld *Datei öffnen*. Der voreingestellte Datei-Typ ist SPSS Statistics mit der Erweiterung **.sav**.

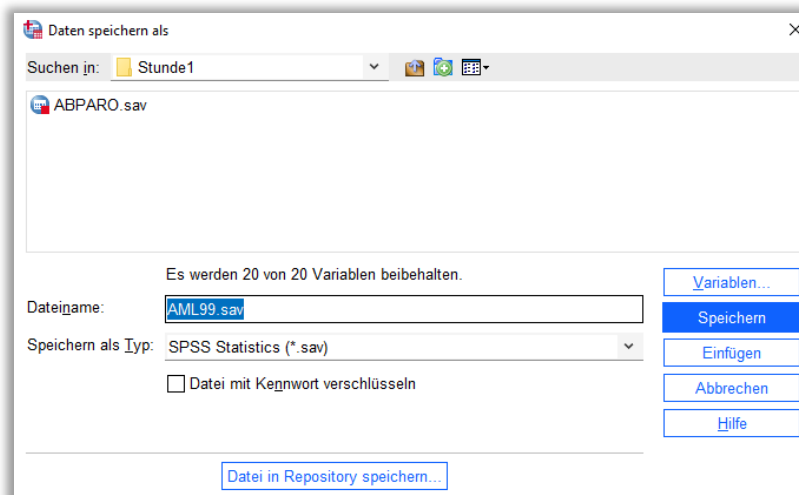
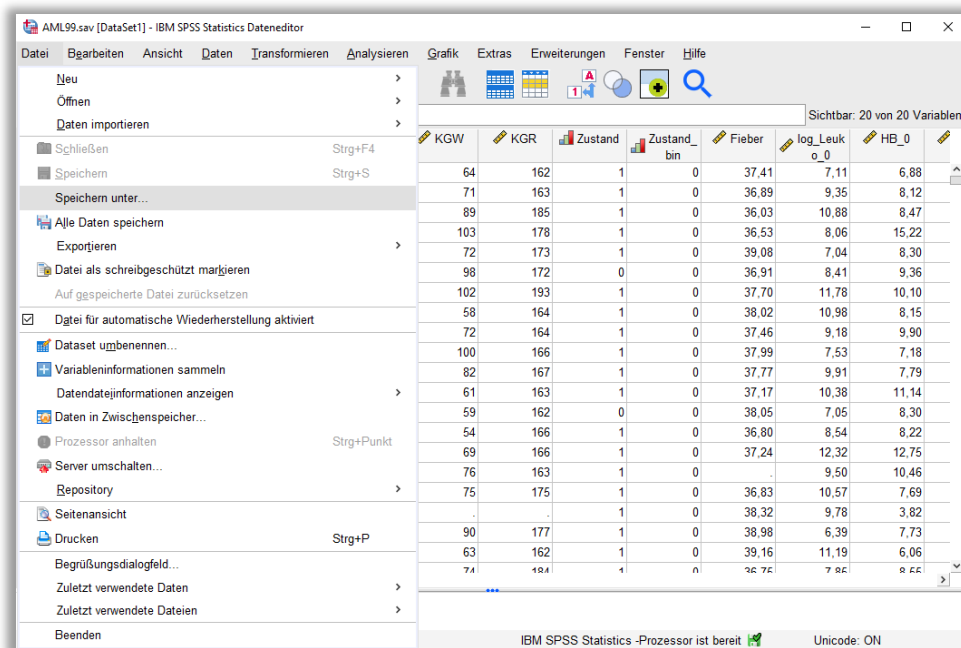


Durch Auswahl eines anderen Dateityps können auch Daten eingelesen werden, die in einem anderen Format vorliegen (z.B. Excel-Dateien). Diese sollten zuvor für das Einlesen in SPSS geeignet aufbereitet werden.



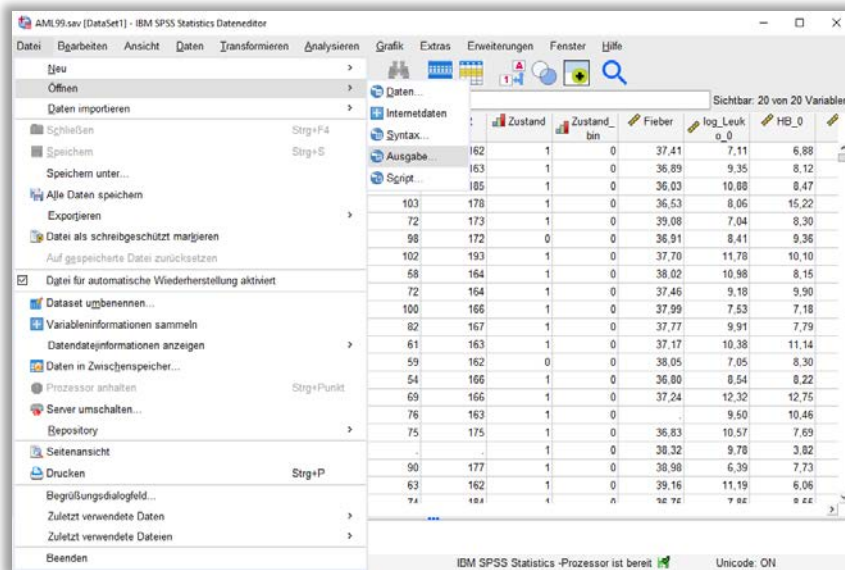


Um eine SPSS-Datei abzuspeichern, betätigen Sie die Befehlsfolge **Datei > Speichern unter**. Im Dialogfeld *Daten speichern als Typ* können Sie die Datei mit der Endung **.sav** in Ihrem gewünschten Verzeichnis ablegen.

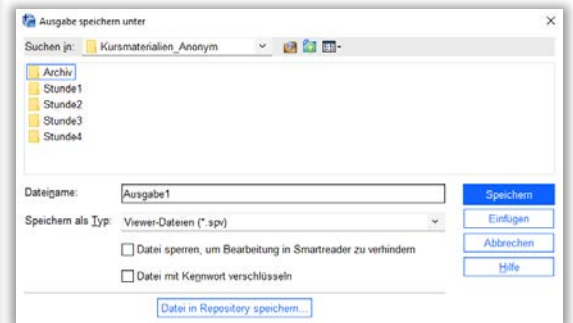
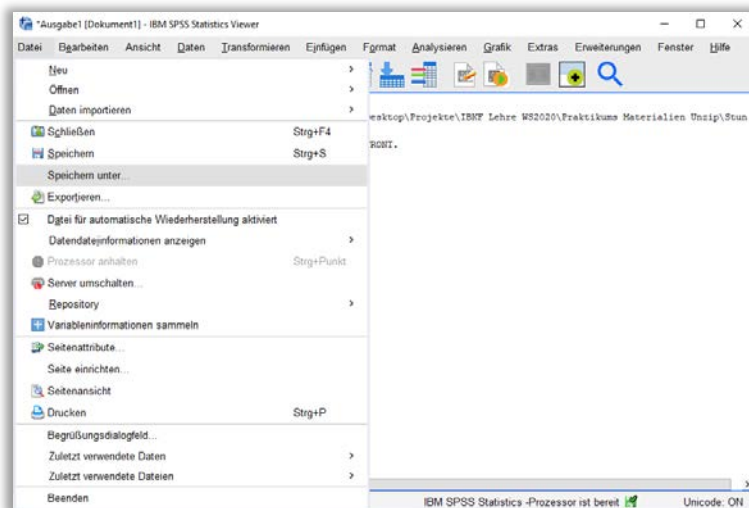


### 3.2 Öffnen und Speichern einer Ausgabedatei

Über die Befehlsfolge **Datei > Öffnen > Ausgabe** gelangen Sie in das Dialogfeld *Ausgabe öffnen*. Der Dateityp ist die Viewer-Datei mit der Erweiterung **.spv**.



Um den gesamten Inhalt des Ausgabefensters abzuspeichern, muss man sich im Ausgabefenster befinden. Betätigen Sie die Befehlsfolge **Datei > Speichern unter**. Sie gelangen in das Dialogfeld *Ausgabe speichern unter*. Der Dateityp ist die Viewer-Datei mit der Erweiterung **.spv**.



Ausgaben lassen sich auch exportieren, so dass sie ohne das Öffnen von SPSS verwendet werden können (Befehlsfolge **Datei > Exportieren**, geeigneten Dokumententyp (z.B. .pdf / .doc) und Pfad wählen, **OK**).

### 3.3 Zusammenfügen von Datendateien (fall- und variablenweise)

In SPSS besteht die Möglichkeit, zwei Datendateien (z.B. *SPSS1.sav* und *SPSS2.sav*) zusammenzuführen. Dabei gibt es zwei Möglichkeiten: fallweises Zusammenfügen und variablenweises Zusammenfügen.

### 3.3.1 Fallweises Zusammenfügen

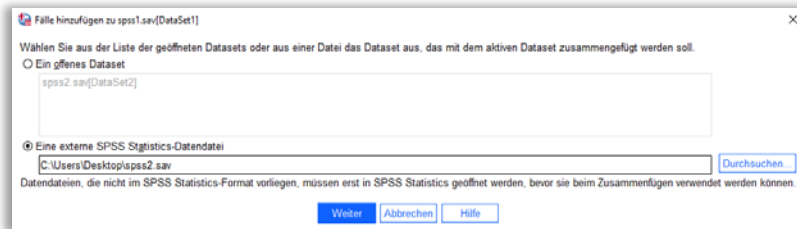
Zusammenfügen zweier Datendateien mit unterschiedlichen Fällen, deren Variablen identisch oder unterschiedlich sind.

	PID	Therapie	Geschlecht	alter	KGW	KGR	Zustand	Zustand_bin
1	P1	1	2	60	64	162	1	0
2	P2	1	1	39	71	163	1	0
3	P3	1	1	71	89	185	1	0
4	P4	1	1	61	103	178	1	0
5	P6	1	1	62	98	172	0	0
6	P7	1	1	44	102	193	1	0
7	P8	1	2	60	58	164	1	0
8	P9	1	2	66	72	164	1	0
9	P10	1	2	61	100	166	1	0
10	P13	1	2	62	59	162	0	0

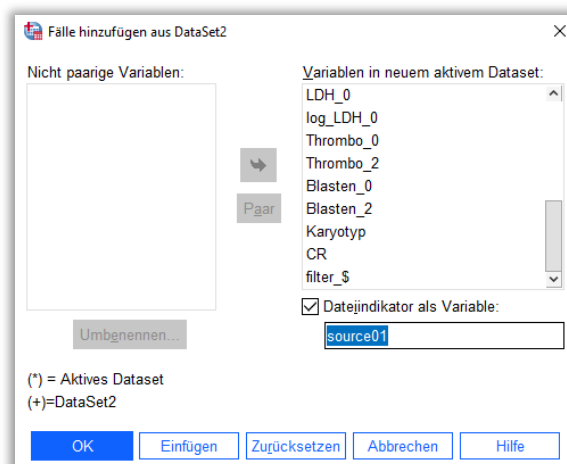
  

	PID	Therapie	Geschlecht	alter	KGW	KGR	Zustand	Zustand_bin	LDH_0	log_LDH_0	Thrombo_0	Thrombo_2	Blasten_0	Blasten_2	Karyotyp	CR	filter_\$
1	P5	2	1	65	72	173	1	0									
2	P11	2	2	64	82	167	1	0									
3	P12	2	2	67	61	163	1	0									
4	P14	2	1	56	54	166	1	0									
5	P15	2	1	72	69	166	1	0									
6	P17	2	1	41	75	175	1	0									
7	P20	2	2	66	63	162	1	0									
8	P22	2	2	63	68	163	1	0									
9	P23	2	1	60	87	171	1	0									
10	P30	2	1	46	128	173	1	0									

Hierfür gehen Sie wie folgt vor: Öffnen Sie zunächst die Datei spss1.sav. Wählen Sie nun aus dem Menü die Befehlsfolge **Daten > Dateien zusammenfügen > Fälle hinzufügen**. Es öffnet sich die Dialogbox *Fälle hinzufügen zu spss1.sav*. Laden Sie die entsprechende zweite SPSS-Datei, aus der Fälle hinzugefügt werden sollen und gehen Sie auf **Weiter**.



Wenn die Variablen in beiden Dateien übereinstimmen, sind alle Variablen im Feld *Variablen in neuer Arbeitsdatei* aufgeführt und das Feld *Nicht gepaarte Variablen* bleibt leer. Um in der späteren zusammengefassten Datei zu kennzeichnen, welche Daten aus welcher Datei stammen, aktivieren Sie zusätzlich die Option *Datei-Indikator als Variable*. Die Variable *source01* als Datei-Indikator hat den Wert 0 für die zuerst geladene Datei *spss1.sav* und den Wert 1 für die hinzugefügte Datei *spss2.sav*. Bestätigen Sie mit **OK**.

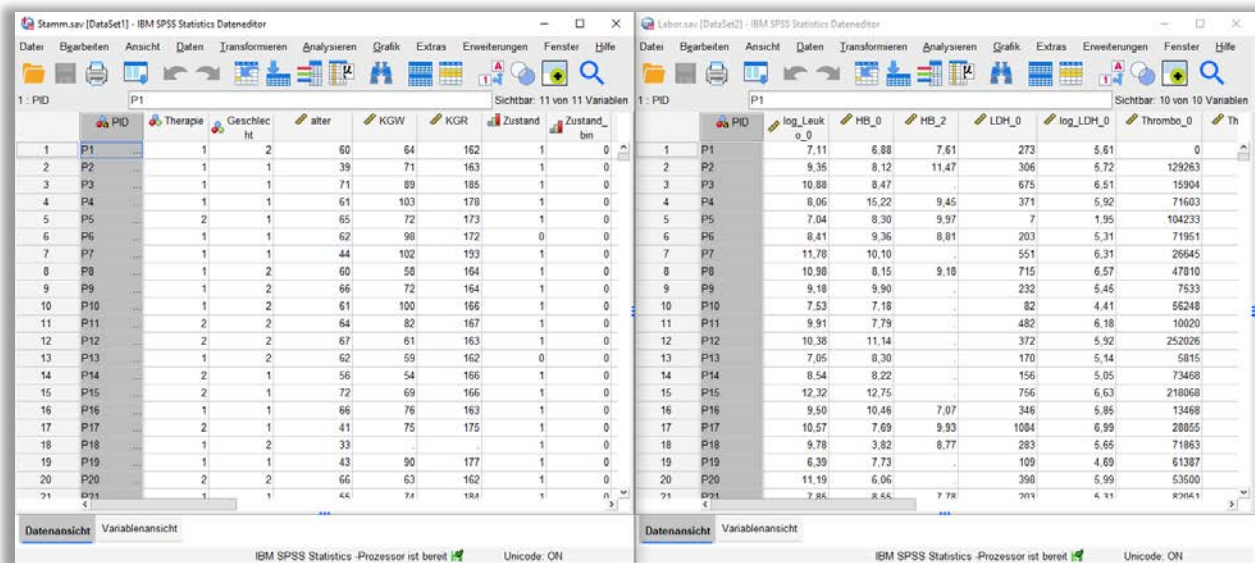


Die fertige, zusammengefügte Datei hat folgende Struktur:

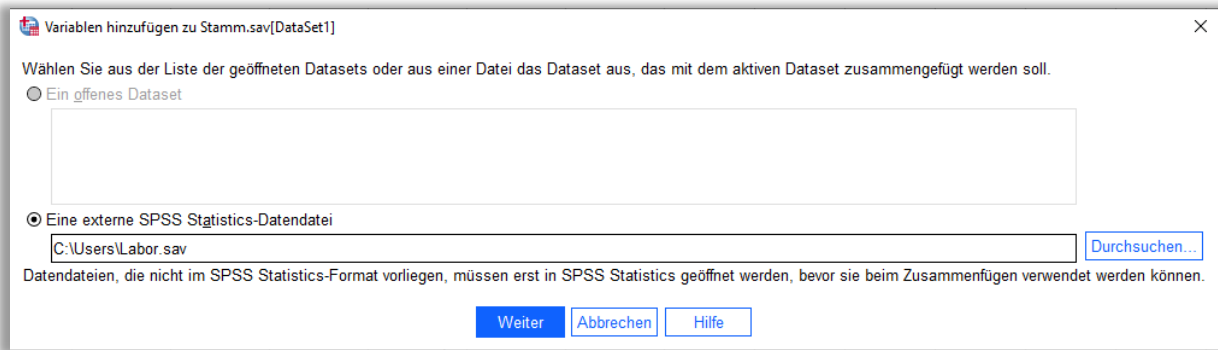
	PID	Therapie	Geschlecht	alter	KGW	KGR	Zustand	Zustand_bin	source01
1		1	2	60	64	162	1	0	0
2	P2	...	1	39	71	163	1	0	0
3	P3	...	1	71	89	185	1	0	0
4	P4	...	1	61	103	178	1	0	0
5	P6	...	1	62	98	172	0	0	0
6	P7	...	1	44	102	193	1	0	0
7	P8	...	1	60	58	164	1	0	0
8	P9	...	1	66	72	164	1	0	0
9	P10	...	1	61	100	166	1	0	0
10	P13	...	1	62	59	162	0	0	0
11	P16	...	1	66	76	163	1	0	0
12	P18	...	1	33	.	.	1	0	0
13	P19	...	1	43	90	177	1	0	0
14	P21	...	1	55	74	184	1	0	0
15	P24	...	1	30	72	176	0	0	0
16	P25	...	1	55	74	175	0	0	0
17	P26	...	1	64	100	176	1	0	0
18	P27	...	1	63	87	172	1	0	0
19	P28	...	1	41	98	184	0	0	0
20	P29	...	1	62	76	173	0	0	0

### 3.3.2 Variablenweises Zusammenfügen

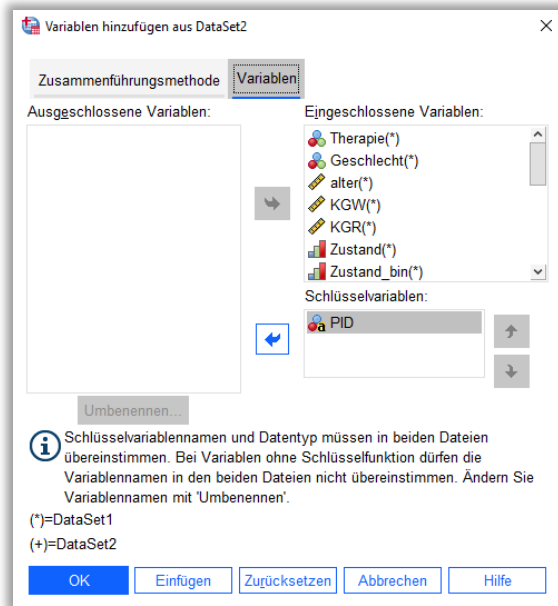
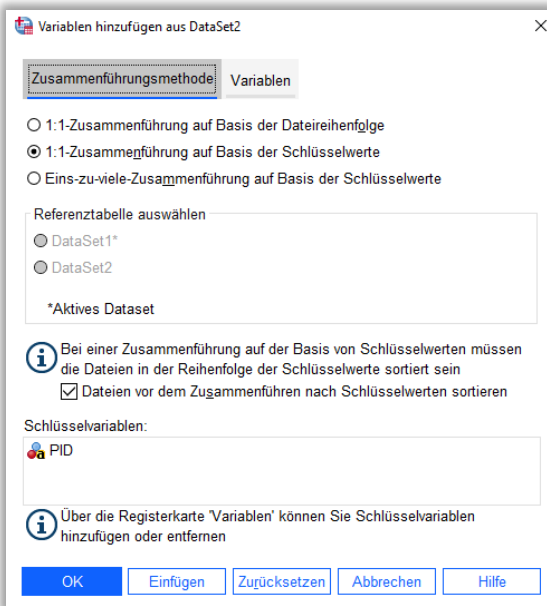
Zusammenfügen zweier Datendateien mit gleichen Fällen, deren Variablen unterschiedlich sind, z.B. eine Datei enthält die Stammdaten der Patienten (*Stamm.sav*), die andere Datei die Laborparameter (*Labor.sav*) dieser Patienten. Empfehlenswert ist die Benutzung einer Schlüsselvariablen in beiden Dateien, da hier die Richtigkeit des Zusammenfügens beider Dateien sichergestellt ist.



Um beide Dateien zu einer Datei zusammenzuführen, gehen Sie wie folgt vor: Öffnen Sie zunächst die Datei *Stamm.sav* und wählen aus dem Menü die Befehlsfolge: **Daten > Dateien zusammenfügen > Variablen hinzufügen**. Es öffnet sich die Dialogbox *Variablen hinzufügen zu Stamm.sav*. Klicken Sie auf den Button **Durchsuchen** und wählen die Datei *Labor.sav* aus.



Klicken Sie auf den Button **Öffnen** und bestätigen Sie anschließend mit **Weiter**. Es öffnet sich die Dialogbox *Variablen hinzufügen aus...* Unter *Variablen* wählen Sie als Schlüsselvariablen Variablen, die in beiden Dateien enthalten sind (z.B. die ID des Patienten *PID*). Variablen, die in beiden Dateien auftreten und nicht zur Identifikation von Fällen dienen, können im Feld *Ausgeschlossene Variablen* vom Hinzufügen ausgeschlossen werden. Aktivieren Sie unter *Zusammenführungsmethode* die Option *1:1-Zusammenführung auf Basis der Schlüsselwerte*. Bestätigen Sie mit **OK**. (Vorsicht: Stellen Sie vorher sicher, dass das Häkchen bei *Dateien vor dem Zusammenführen nach Schlüsselwerten sortieren* gesetzt ist.



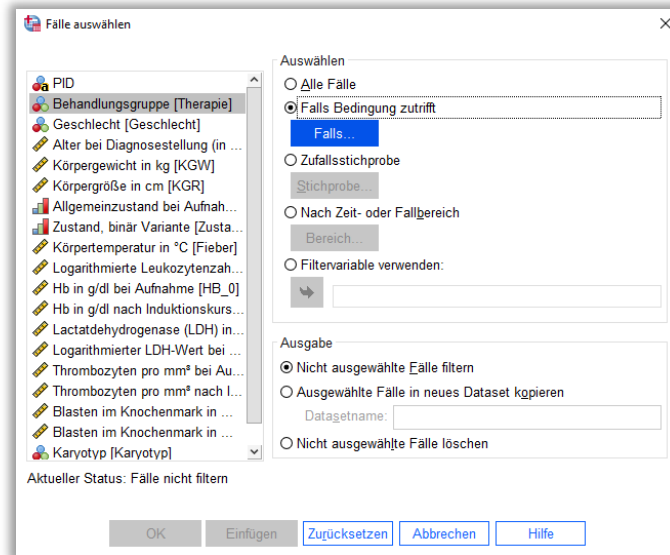
Die fertige zusammengeführte Datei hat folgende Struktur:

	PID	Therapie	Geschlecht	alter	KGW	KGR	Zustand	Zustand_bin	Fieber	Karyotyp	CR	log_Leuk_o_0	HB_0	HB_2	LDH_0	log_LDH_0	Thromb
1	P1	TAD-HAM	Weiblich	60	64	162	Leicht beei...	Normal od...	37,41	Intermediär	Nein	7,11	6,88	7,61	273	5,61	
2	P10	TAD-HAM	Weiblich	61	100	166	Leicht beei...	Normal od...	37,99	Intermediär	Ja	7,53	7,18		82	4,41	
3	P100	TAD-HAM	Männlich	39	93	196	Leicht beei...	Normal od...	36,24	Intermediär	Ja	11,22	9,18	10,71	1019	6,93	
4	P1000	TAD-HAM	Männlich	72	76	172	Leicht beei...	Normal od...	38,28	Ungünstig	Nein	9,46	9,48		177	5,18	1
5	P1001	TAD-HAM	Männlich	62			Leicht beei...	Normal od...	36,72	Ungünstig	Ja	7,38	6,02		284	5,65	
6	P1002	TAD-HAM	Männlich	73	82	170				Intermediär	Nein		8,09		238	5,47	
7	P1003	TAD-HAM	Männlich	69	99	175	Leicht beei...	Normal od...	37,15	Intermediär	Ja	8,09	8,50	9,19	265	5,58	
8	P1004	TAD-HAM	Weiblich	60			Leicht beei...	Normal od...	38,44	Intermediär	Nein	12,21	7,31		2457	7,81	
9	P1005	HAM-HAM	Männlich	73	64	175	Leicht beei...	Normal od...	37,09	Intermediär	Ja	11,31	8,07	10,63	902	6,80	
10	P1006	HAM-HAM	Weiblich	85	48	158	Leicht beei...	Normal od...	36,84	Intermediär	Nein	9,41	9,51	8,94	372	5,92	

## 4. Datenselektion

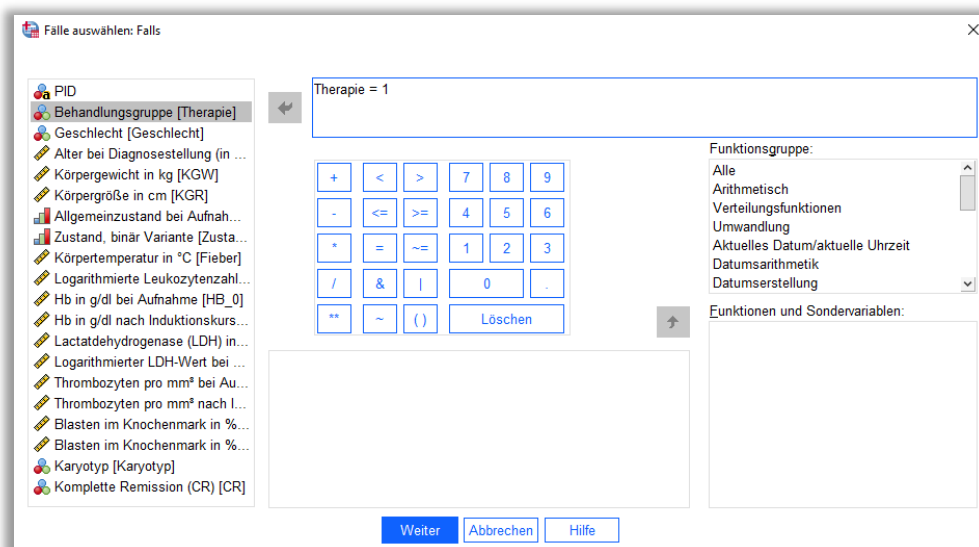
### 4.1 Auswahl von Fällen

Unter Datenselektion versteht man die Auswahl von Fällen nach bestimmten Kriterien. So sollen z.B. bei der Auswertung des für die Demonstration verwendeten AML-Datensatzes nur Patienten berücksichtigt werden, die der Behandlungsgruppe TAD-HAM (*Therapie* = 1) zugeteilt wurden. Dazu wählen Sie aus dem Menü die Befehlsfolge **Daten > Fälle auswählen > Falls Bedingung zutrifft > Falls**.



Es öffnet sich das Dialogfeld *Fälle auswählen: Falls*.

Übertragen Sie die Variable *Therapie* über die Pfeiltaste ⇨ in das rechte Dialogfeld. Nun kann der angeforderte logische Ausdruck über die Buttons der Rechentastatur in das dafür vorgesehene Dialogfeld geschrieben werden. Die Bedingung lautet *Therapie = 1*, d.h. es sollen alle Fälle ausgewählt werden, für welche die Variable *Therapie* den Wert »1« (Codierung für TAD-HAM) hat. Die Dialogbox stellt sich nun wie folgt dar:



Bestätigen Sie mit **Weiter**. Belassen Sie die Voreinstellung *Nicht ausgewählte Fälle filtern*.

Bestätigen Sie mit **OK**.

Sie gelangen zurück zum Dateneditor. Der Filter ist nun eingeschaltet, d.h. alle Patienten, die nicht die Therapie TAD-HAM bekommen haben, sind gefiltert und werden aus allen weiteren Analysen

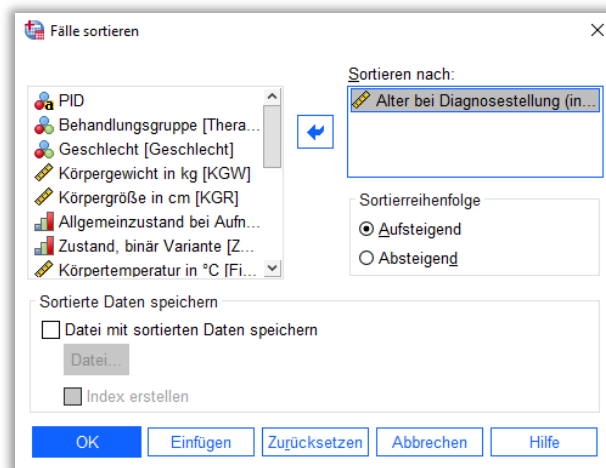


ausgeschlossen. Die Selektion ist erkennbar an den durchgestrichenen Fallnummern, sowie am unteren Rand des Daten-Editors: Meldung »Filter aktiv«.

Bei Bedarf müssen Sie den Filter wieder deaktivieren.

### 4.2 Fälle sortieren

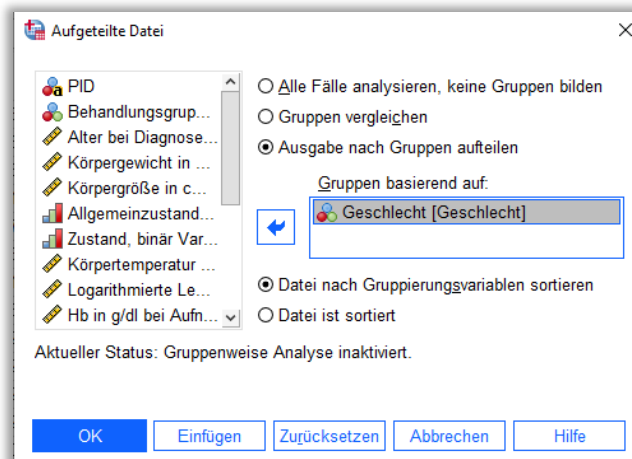
In SPSS können die Daten nach den Werten von einer oder mehreren Variablen in auf- oder absteigender Reihenfolge sortiert werden. Um z.B. die Daten der Datei *AML99.sav* nach dem Alter in aufsteigender Reihenfolge zu sortieren, gehen Sie wie folgt vor: **Daten > Fälle sortieren**. Es öffnet sich das Dialogfeld *Fälle sortieren*. Übertragen Sie die Variable *Alter* über die Pfeiltaste ⇨ in das rechte Dialogfeld *Sortieren nach*. Im Auswahlkasten *Sortierreihenfolge* ist die Sortierreihenfolge *Aufsteigend* vorgegeben. Bestätigen Sie mit **OK**.



Im Daten-Editor sehen Sie nun die Datendatei *AML99.sav* aufsteigend sortiert nach den Werten der Variable *Alter*.

### 4.3 Aufteilung der Fälle in Gruppen

Die Datenanalyse kann in SPSS getrennt nach Gruppen erfolgen. Eine solche Gruppenvariable kann z.B. die Variable *Geschlecht* sein. So können z.B. Häufigkeitstabellen nach den Merkmalen männlich und weiblich getrennt durchgeführt werden. Wählen Sie aus dem Menü **Daten > Datei aufteilen**. Es öffnet sich die Dialogbox *Datei aufteilen*. Aktivieren Sie die Option *Ausgabe nach Gruppen aufteilen* und bestätigen Sie mit **OK**. So erhalten Sie die Ausgabe nachfolgender Analysen getrennt nach Gruppen.

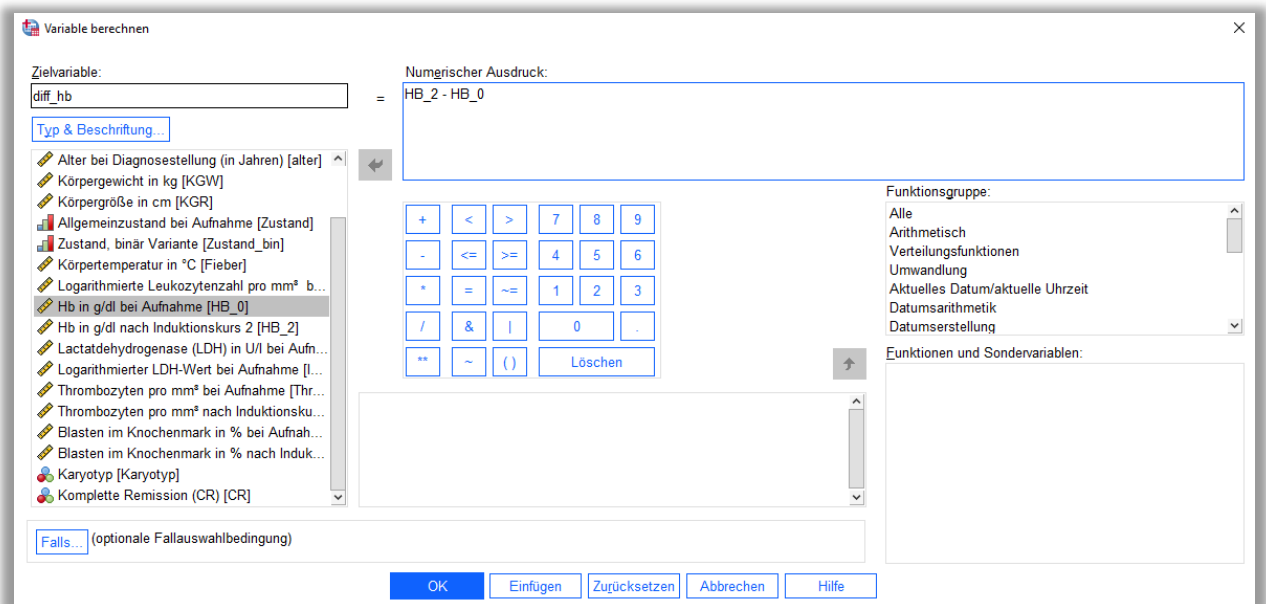


## 5. Datenmodifikation

### 5.1 Berechnung von neuen Variablen

Es gibt in SPSS die Möglichkeit, neue Variablen zu berechnen und der Datendatei anzufügen. So sind etwa in der AML-Studie zu zwei Zeiten (vor und nach Therapie) Messungen des Hämoglobins erfolgt (Variablen *HB\_0* und *HB\_2*). Interessieren Sie sich für die Änderung der Hämoglobinkonzentration zwischen den beiden Zeitpunkten, so kann die Berechnung einer neuen Variable *diff\_hb* von SPSS ausgeführt werden. Wählen Sie dazu im Menü **Transformieren > Berechnen**. Es öffnet sich die Dialogbox *Variable berechnen*. Im Feld *Zielvariable* wird der Name der Variablen eingegeben, die den neu berechneten Wert zugewiesen bekommt. Im Feld *Numerischer Ausdruck* wird der numerische Ausdruck eingetragen, der zur Berechnung des Wertes der Zielvariablen verwendet wird.

Tragen Sie im Feld *Zielvariable* den Namen *diff\_hb* ein, im rechten Dialogfeld Numerischer Ausdruck die Formel »*HB\_2 – HB\_0*«. Die Formel kann entweder per Hand eingegeben werden oder durch Benutzung der Variablenliste über die Pfeiltaste ⇨ und die Buttons im Feld mit der Rechentastatur. Bestätigen Sie mit **OK**. Sie finden die neue Variable *diff\_hb* im Datenblatt hinten angefügt.



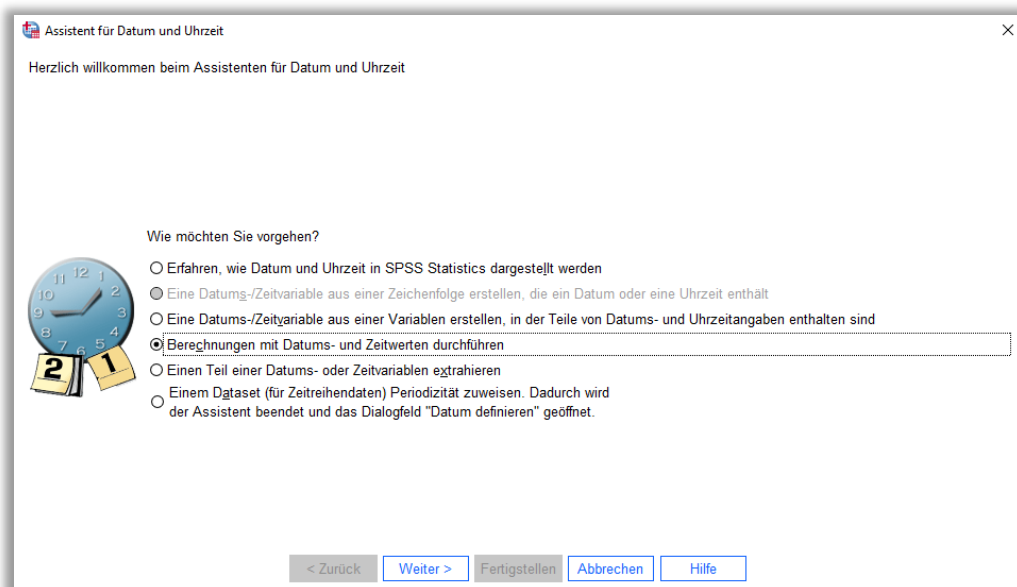
### 5.2 Rechnen mit Datumsangaben

In SPSS ist es möglich, die Differenz aus zwei Datumsvariablen, die bereits in der Datei vorkommen, zu berechnen. Z.B. möchten Sie bestimmen, wie alt jeder Patient zum Zeitpunkt der Untersuchung (in Jahren) ist. Es liegen dafür das Geburtsdatum (*birth\_dt*) und der Untersuchungszeitpunkt (*accept\_dt*) vor (siehe Abbildung unten):



patnr	birth_dt	accept_dt
1	10-Jan-1937	15-Nov-1967
2	02-Mar-1916	02-Jan-1968
3	19-Sep-1913	06-Jan-1968
4	23-Dec-1927	28-Mar-1968
5	28-Jul-1947	10-May-1968
6	08-Nov-1913	13-Jun-1968
7	29-Aug-1917	12-Jul-1968
8	27-Mar-1923	01-Aug-1968
9	11-Jun-1921	09-Aug-1968
10	09-Feb-1926	11-Aug-1968
11	22-Aug-1920	15-Aug-1968
12	09-Jul-1915	17-Sep-1968
13	22-Feb-1914	19-Sep-1968


Um das Alter jedes Patienten in Jahren zu berechnen, gehen Sie wie folgt vor: Wählen Sie im Menü die Befehlsfolge **Transformieren > Assistent für Datum und Uhrzeit**. Sie gelangen in das Dialogfeld *Assistent für Datum und Uhrzeit*. Klicken Sie hier auf **Berechnungen mit Datums- und Zeitwerten durchführen** und gehen auf **Weiter**. Sie gelangen ins nächste Dialogfeld *Schritt 1 von 3*. Aktivieren Sie hier den Befehl *Berechnen der Anzahl der Zeiteinheiten zwischen zwei Datumswerten*. Gehen Sie auf **Weiter**. Sie gelangen ins nächste Dialogfeld *Schritt 2 von 3*. Bringen Sie nun die Variable *accept\_dt* über die Pfeiltaste ⇨ in das Feld *Datum1* und die Variable *birth\_dt* in das Feld *minus Datum2*. Belassen Sie unter *Einheit* die Voreinstellung *Jahre*. Unter *Ergebnisbehandlung* belassen Sie die Voreinstellung *Auf ganze Zahl kürzen*. Gehen Sie auf **Weiter**. Sie gelangen ins nächste Dialogfeld *Schritt 3 von 3*. Unter *Ergebnisvariable* geben Sie den Namen der neuen Variable ein (z.B. *alter*). Optional können Sie noch ein Variablenlabel vergeben (z.B. *Alter in Jahren zum Zeitpunkt der Untersuchung*). Bestätigen Sie Ihre Eingaben durch **Fertigstellen**.



Assistent für Datum und Uhrzeit - Schritt 1 von 3

Berechnungen mit Datumswerten durchführen

Wählen Sie eine der folgenden Aufgaben aus und klicken Sie auf 'Weiter'.

  Addieren bzw. Subtrahieren einer Dauer zu bzw. von einem Datum (z. B. Hinzufügen eines Monats zu einem Alter oder Addieren einer Zeitvariablen zu einer Datums-/Zeitvariablen)

Berechnen der Anzahl der Zeiteinheiten zwischen zwei Datumswerten (z. B. Berechnen eines Alters in Jahren aus dem Geburtsdatum und einem weiteren Datum)


Subtrahieren zweier Werte für Dauer (z. B. Arbeitszeit - Anfahrtszeit)

< Zurück Weiter > Fertigstellen Abbrechen Hilfe

Assistent für Datum und Uhrzeit - Schritt 2 von 3

Berechnung der Anzahl der Zeiteinheiten zwischen zwei Datums- oder Datums-/Zeitvariablen.

Das Ergebnis ist eine ganzzahlige Variable. Alle Bruchteile einer Einheit werden verworfen. Das Ergebnis ist eine Variable für die Dauer. In der unten stehenden Variablenliste werden nur Variablen für die Dauer angezeigt.

 Variablen:  Aktuelles Datum und aktuelle Uhrzeit [...]

Datum1:  acceptance into program [accept\_dt]

minus Datum2:  birth date [birth\_dt]

Einheit: Jahre

Ergebnisbehandlung

Auf ganze Zahl kürzen

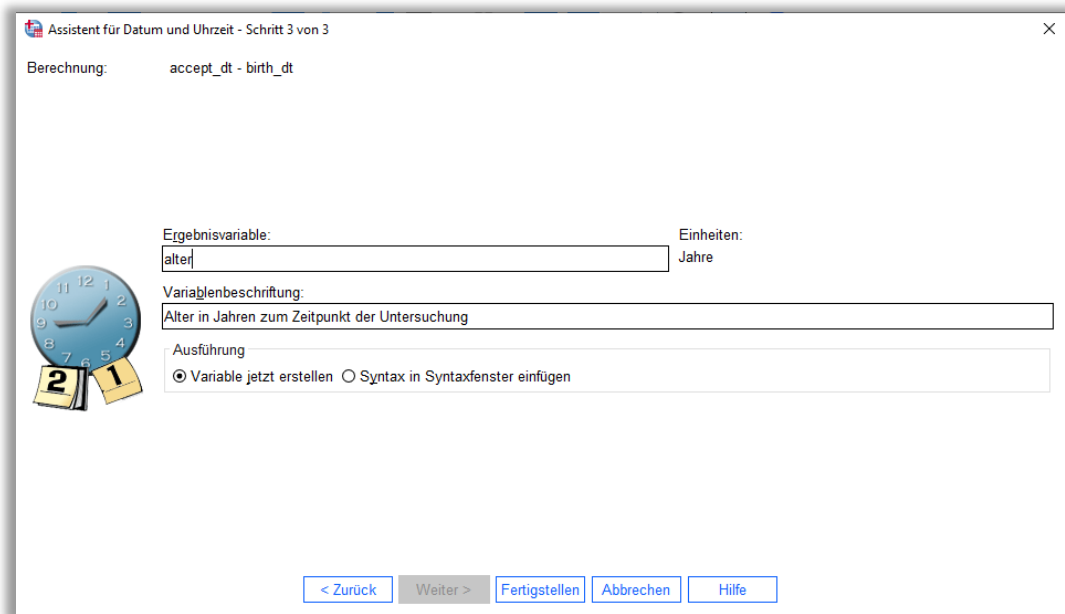
Auf ganze Zahl runden

Bruchteil beibehalten

Für Einheiten von Monat und Jahr basiert das Ergebnis auf durchschnittlichen Einheitenlängen, es sei denn, es wird abgeschnitten

STIME ist das aktuelle Datum und die aktuelle Uhrzeit.

< Zurück Weiter > Fertigstellen Abbrechen Hilfe



Als Ergebnis erscheint folgende Datendatei mit der neu berechneten Variable *alter*, die an den bisherigen Datensatz hinten angefügt wurde:

patnr	birth_dt	accept_dt	alter
1	10-Jan-1937	15-Nov-1967	30
2	02-Mar-1916	02-Jan-1968	51
3	19-Sep-1913	06-Jan-1968	54
4	23-Dec-1927	28-Mar-1968	40
5	28-Jul-1947	10-May-1968	20
6	08-Nov-1913	13-Jun-1968	54
7	29-Aug-1917	12-Jul-1968	50
8	27-Mar-1923	01-Aug-1968	45
9	11-Jun-1921	09-Aug-1968	47
10	09-Feb-1926	11-Aug-1968	42
11	22-Aug-1920	15-Aug-1968	47
12	09-Jul-1915	17-Sep-1968	53
13	22-Feb-1914	19-Sep-1968	54

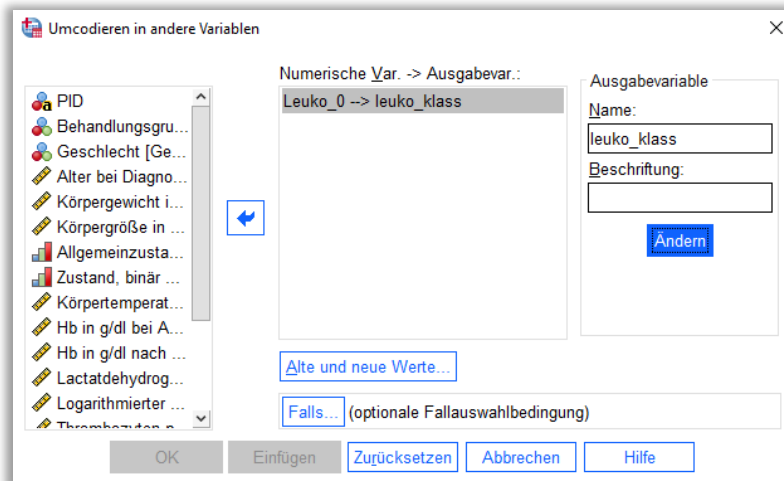
### 5.3 Umkodieren und Klassieren von Werten

In SPSS können ursprünglich erfasste Daten umkodiert werden. Eine Umkodierung numerischer Daten ist z.B. dann sinnvoll, wenn die original erfassten Rohdaten für die weitere Analyse im folgenden Sinn vereinfacht werden sollen. Man möchte den Wert der Leukozyten bei Aufnahme in 3 Klassen einteilen und so die Leukozytenklasse jedes Patienten bestimmen. Die Klassierung soll folgendermaßen aussehen:

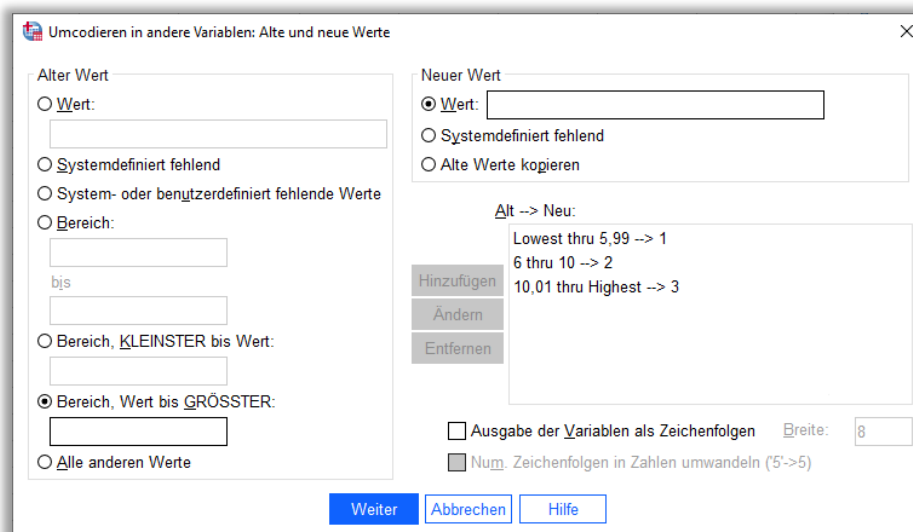
Leukozytenwert	Klasse	Wertelabel
< 6	1	Leukopenie
6-10	2	Normbereich
> 10	3	Leukozytose

Um obiges Verfahren in SPSS umzusetzen, gehen Sie wie folgt vor: Über die Befehlsfolge **Transformieren > Umkodieren in andere Variablen** gelangen Sie in das Dialogfeld *Umkodieren in andere Variablen*. Übertragen Sie die ursprüngliche Variable *Leuko\_0* über die Pfeiltaste ⇨ in das Feld *Numerische Var.* -> *Ausgabevar.*: Anschließend geben Sie der neu zu erstellenden Ausgabevariable einen Namen im Feld

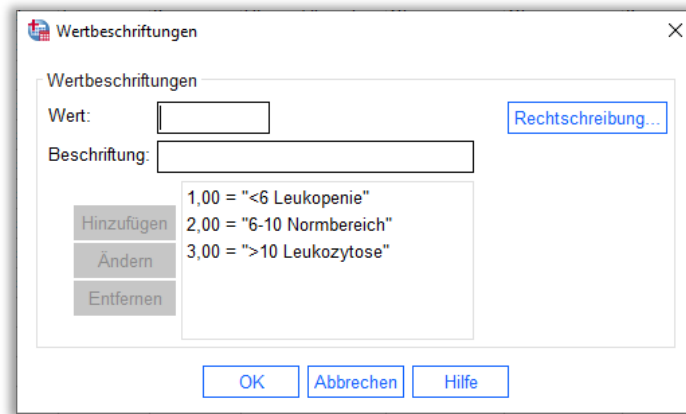
Ausgabevariable (z.B. *leuko\_klass*). Danach klicken Sie auf den Button **Ändern**. So wird der Name der neuen Ausgabevariable festgelegt (*leuko\_klass*). Betätigen Sie nun den Button **Alte und neue Werte**.



Sie gelangen in das Dialogfeld *Umkodieren in andere Variablen: Alte und neue Werte*. Im linken Fensterbereich unter *Alter Wert* geben Sie bei *Bereich*, *KLEINSTER bis Wert* den Wert »5,99« ein. Im rechten Fensterbereich geben Sie unter *Neuer Wert* den Wert »1« ein und bestätigen Ihre Eingaben durch den Button **Hinzufügen**. Dem *Bereich* »6« bis »10« wird der neue Wert »2« zugeordnet. Und im Bereich *Wert bis GRÖSSTER* tragen Sie »10,01« ein und weisen diesem Bereich den Wert »3« zu.



Bestätigen Sie die Eingaben mit **Weiter** und **OK**. In der Datendatei wurde nun die neue umkodierte Variable *leuko\_klass* hinten angefügt. Zusätzlich können in der Variablenansicht **Wertelabels** der neuen Variable vergeben werden.

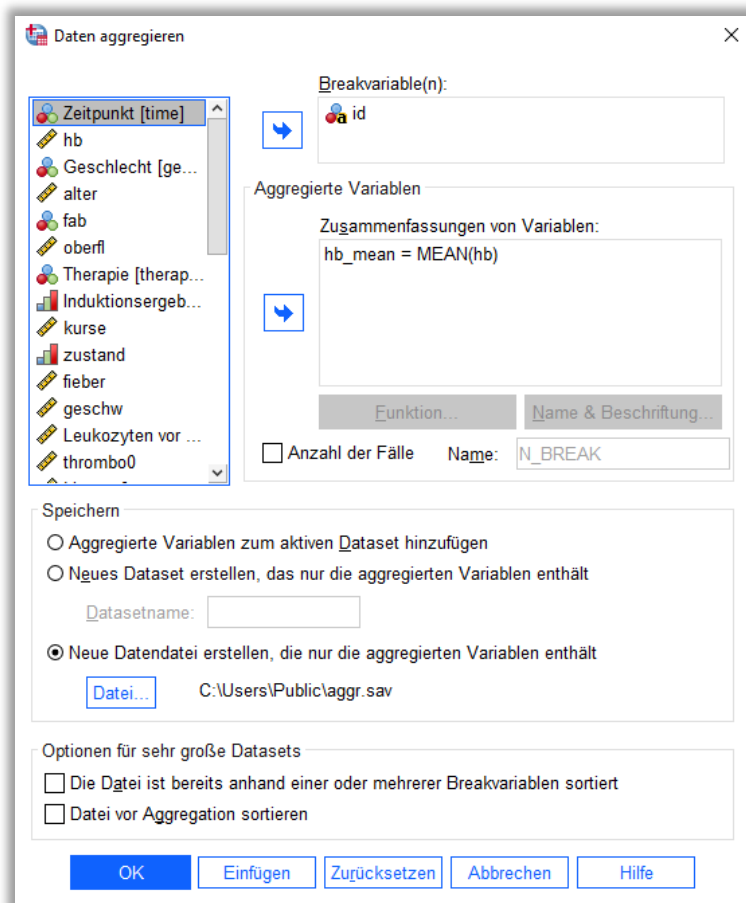


### 5.4 Aggregieren von Daten

Basierend auf den Werten von einer oder mehreren Gruppenvariablen (sog. Break-Variablen) können Fälle zusammengefasst (aggregiert) und eine neue Datendatei erstellt werden, die für jede Break-Gruppe einen Fall enthält. Z.B. wurde in der AML99-Studie der Hb-Wert (in g/dl) jedes Patienten insgesamt drei Mal erfasst: bei Aufnahme sowie nach dem 1. und 2. Induktionskurs. Die drei Werte sind in der Datendatei als drei Fälle untereinander aufgeführt:

	id	time	hb
1	P1202	1	3,40
2	P1202	2	.
3	P1202	3	.
4	P1204	1	8,60
5	P1204	2	8,40
6	P1204	3	.
7	P1205	1	6,60
8	P1205	2	.
9	P1205	3	9,00
10	P1207	1	8,50
11	P1207	2	10,00
12	P1207	3	.
13	P1208	1	10,80
14	P1208	2	12,20
15	P1208	3	8,80

Es soll für jeden Patienten der mittlere Hb-Wert über alle drei Messzeitpunkte bestimmt und eine neue Datendatei erstellt werden, die nur die aggregierten Variablen enthält. Wählen Sie dazu **Daten > Aggregieren**. Es öffnet sich das Dialogfeld *Daten aggregieren*. Übertragen Sie die Variable *id* in das Feld *Breakvariable(n)*. Übertragen Sie die Variable *hb* in das Feld *Zusammenfassung von Variablen*. Als Aggregierungsfunktion ist der Mittelwert bereits voreingestellt. Alternativ können Sie über den Button **Funktion** andere Kenngrößen anfordern. Wenn die aggregierten Daten in einer neuen Datei gespeichert werden sollen, markieren Sie die Option *Neue Datendatei erstellen, die nur die aggregierten Variablen enthält*. Über den Button **Datei** können Sie einen neuen Dateinamen eingeben und die Datei in einem Verzeichnis Ihrer Wahl ablegen. Bestätigen Sie Ihre Wahl mit **Speichern**. Anschließend sollte die Dialogbox wie folgt aussehen:



Bestätigen Sie Ihre Einstellungen mit **OK**. Die neu erzeugte Datei enthält nun die neu gebildete aggregierte Variable *hb\_mean* mit dem mittleren Hb-Wert pro Patient über alle 3 Messzeitpunkte:

	Zentr	Lfdnr	HB_Wert_mean
1	3	2	43,85
2	3	5	43,70
3	3	11	10,13
4	3	12	10,20
5	3	15	54,05

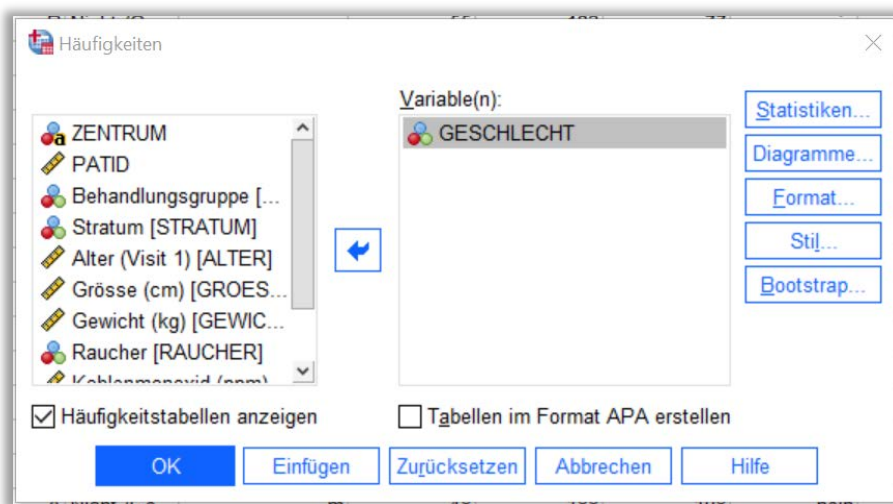
## 6. Deskriptive Statistik

In diesem Kapitel wird erläutert wie deskriptive Analysen mittels Häufigkeitstabellen, Kreuztabellen, Lage- und Streuungsmaßen, graphische Darstellungen qualitativer / quantitativer Merkmale (Kreisdiagramm, Balkendiagramm, Histogramm, Box-and-Whisker-Plot) und Korrelationsanalysen mit SPSS durchgeführt werden können.

### 6.1 Häufigkeitstabellen kategorialer Merkmale

Zur Analyse der Verteilung eines qualitativen Merkmals eignet sich eine Häufigkeitstabelle.

Über die Befehlsfolge **Analysieren > Deskriptive Statistiken > Häufigkeiten** gelangt man in das Dialogfeld *Häufigkeiten*. Zur Bestimmung der absoluten Häufigkeit einer Variablen überträgt man diese in das Feld *Variable(n)*. Möchten Sie weiterhin bestimmen, wie viele Beobachtungen die jeweiligen Variablenausprägungen absolut und relativ aufweisen, so aktivieren Sie zusätzlich die Option *Häufigkeitstabellen anzeigen*. Bestätigen Sie mit **OK**.



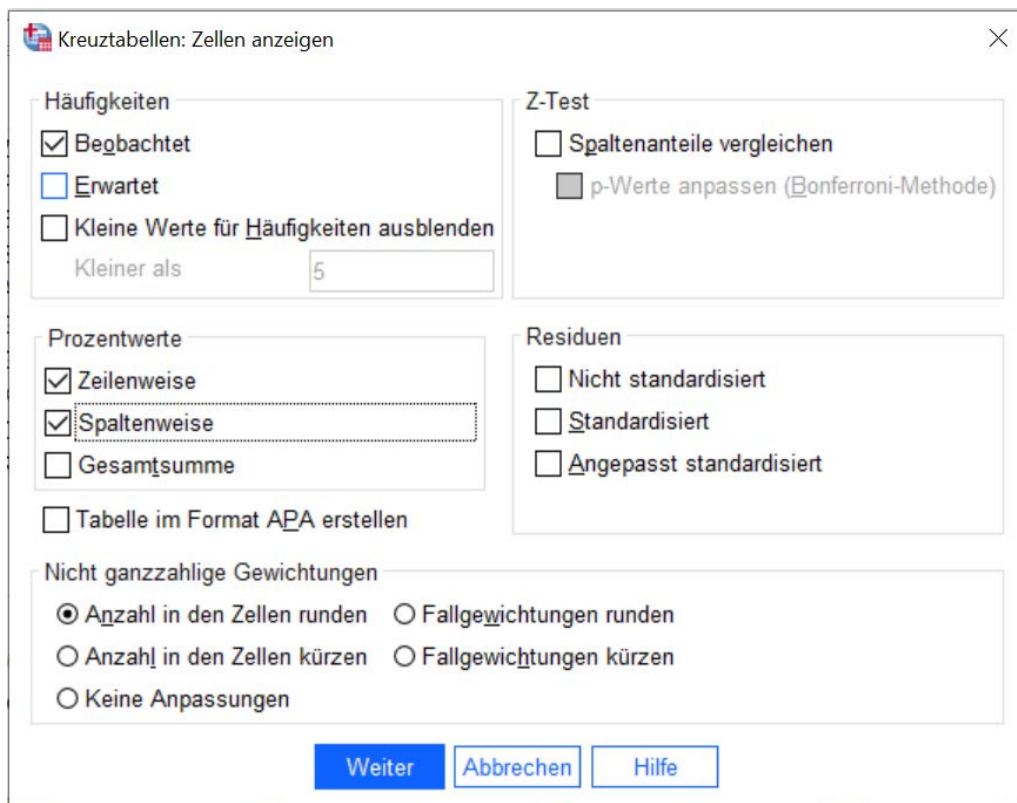
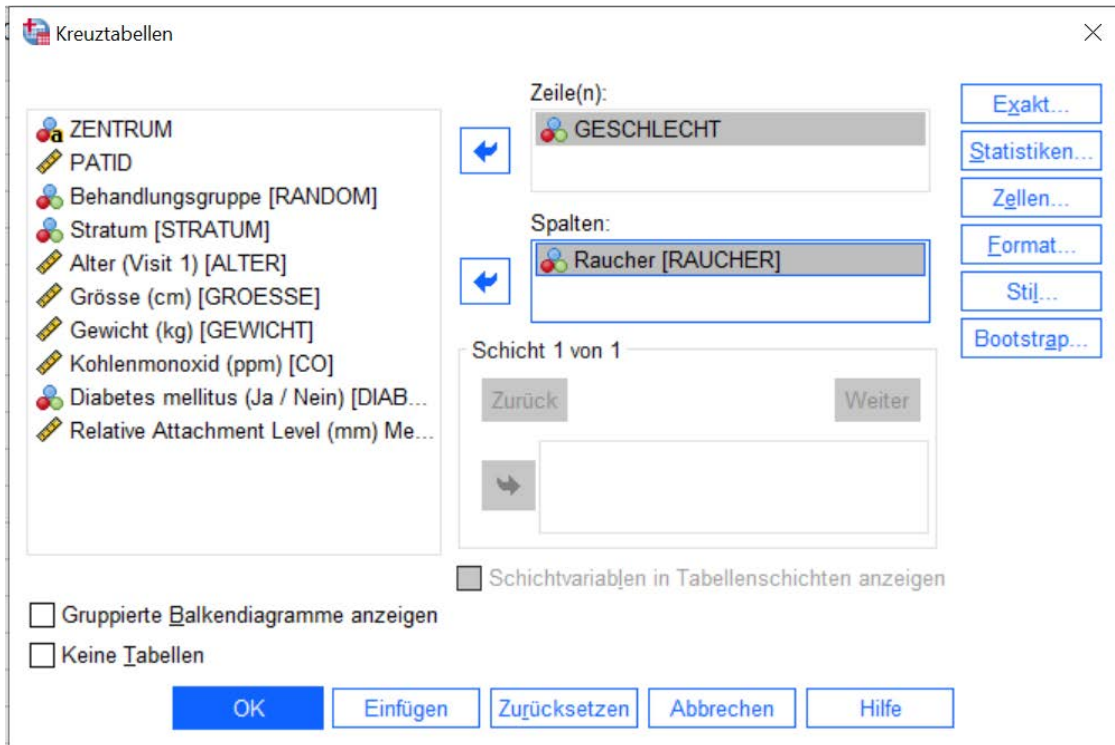
### 6.2 Kreuztabellen

Möchte man nicht nur die Verteilung *eines* qualitativen Merkmals analysieren (univariat), sondern den Zusammenhang *zweier* qualitativer Merkmale (bivariat), so kann man eine Kontingenztafel (Kreuztabelle) erzeugen. Im einfachsten Fall analysiert man zwei qualitative Merkmale mit jeweils 2 Ausprägungen. Dafür erzeugt man eine 2x2 Kreuztabelle, auch Vierfeldertafel genannt.

Über die Befehlsfolge **Analysieren > Deskriptive Statistiken > Kreuztabellen** gelangt man in die Dialogbox *Kreuztabellen*, in der das erste interessierende Merkmal in das Feld *Zeile(n)* und das weitere Merkmal in das Feld *Spalten* übertragen wird.

Anmerkung: In der Regel wird das unabhängige Merkmal (Einflussgröße, in unserem Beispiel das GESCHLECHT) der *Zeile*, das abhängige Merkmal (Zielgröße, in unserem Beispiel das Merkmal RAUCHEN) der *Spalte* zugewiesen.

Per Default werden die absoluten Häufigkeiten ausgegeben. Zusätzlich können die Zeilen- und Spaltenprozentage angefordert werden, indem man unter der Schaltfläche **Zellen...** die Dialogbox *Kreuztabellen: Zellen anzeigen* aktiviert und hier unter *Prozentwerte* die Optionen *Zeilenweise* oder *Spaltenweise* mit einem Häkchen aktiviert. Über **Weiter** gelangen Sie wieder in die übergeordnete Dialogbox *Kreuztabellen*. Bestätigen Sie mit **OK**.



### 6.3 Lagemaße, Streuungsmaße, Verteilungsmaße und -formen bei stetigen Variablen

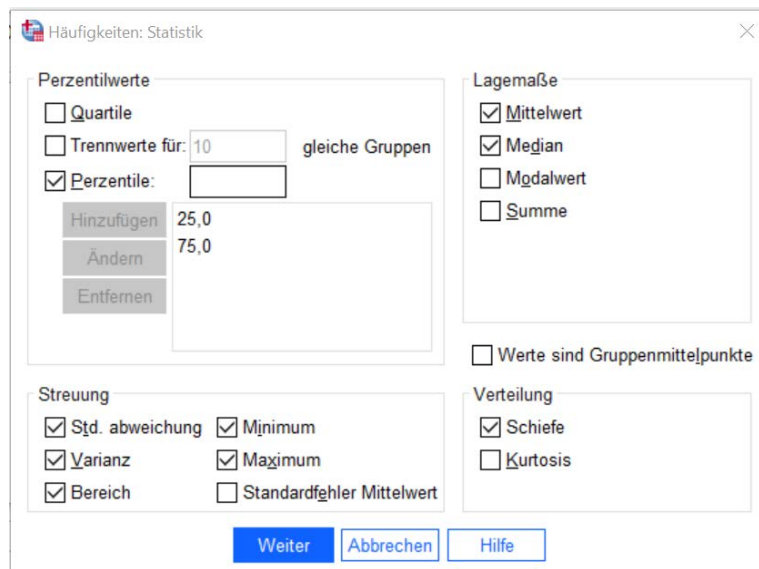
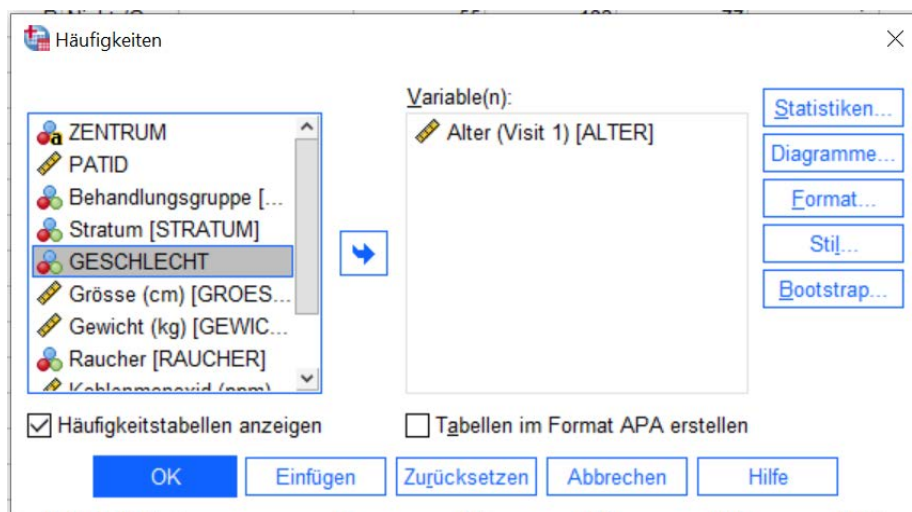
Maßzahlen dienen zur Charakterisierung empirisch gewonnener Daten von quantitativen Merkmalen. Man unterscheidet Lagemaße und Streuungsmaße. Lagemaße sind Maßzahlen zur Charakterisierung des Zentrums einer Häufigkeitsverteilung (z.B. Mittelwert, Median). Streuungsmaße sind Maßzahlen zur



Charakterisierung der Variabilität (z.B. Varianz, Standardabweichung). Mit Hilfe von Verteilungsmaßen wird die Form der Verteilung charakterisiert (z.B. Schiefe).

Option A:

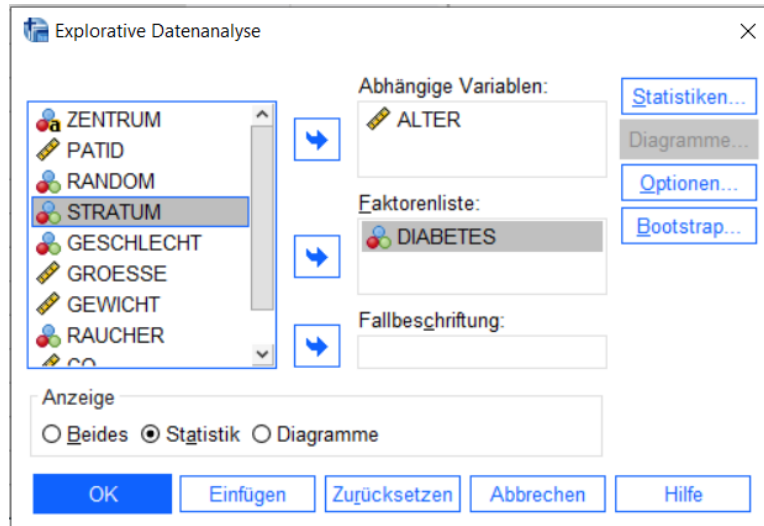
Über die Befehlsfolge **Analysieren > Deskriptive Statistiken > Häufigkeiten** gelangt man in die Dialogbox *Häufigkeiten*. Man überträgt die interessierende Variable in das Feld *Variable(n)*. Um zu unterbinden, dass für ein stetiges Merkmal eine Häufigkeitstabelle für ausgegeben wird, inaktiviert man das Häkchen unter *Häufigkeitstabellen anzeigen*. Klicken Sie auf die Schaltfläche **Statistiken...**. Es öffnet sich die Dialogbox *Häufigkeiten: Statistik*. Hier können die gewünschten Parameter über ein Häkchen angefordert werden. Über **Weiter** gelangen Sie wieder zur übergeordneten Dialogbox *Häufigkeiten*. Bestätigen Sie Ihre angeforderten Analysen mit **OK**.



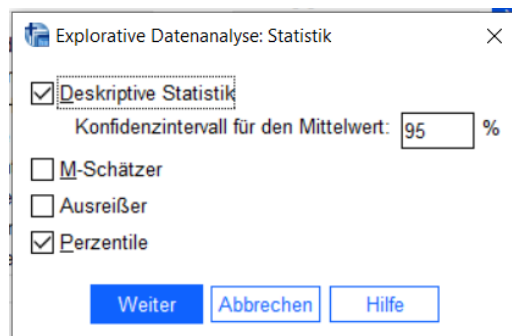
Option B:

Über die Befehlsfolge **Analysieren > Deskriptive Statistiken > Explorative Datenanalyse** gelangen Sie in die die Dialogbox *Explorative Datenanalyse*. Hier können Sie Maßzahlen stetiger Variablen (z.B. ALTER) getrennt nach Gruppen einer kategorialen Variablen (z.B. DIABETES „ja/ nein“) aufteilen.

Übertragen Sie die stetige Variable ALTER per Drag & Drop in das Feld *Abhängige Variablen* und die kategoriale Variable DIABETES in das Feld *Faktorenliste*. Unter *Anzeige* wählen Sie zur besseren Übersichtlichkeit die Option *Statistik* aus.



Um sich die Perzentile anzeigen zu lassen, gehen Sie über den Reiter **Statistiken...**, wählen Sie hier die Option *Perzentile* aus und bestätigen Sie mit **Weiter**.

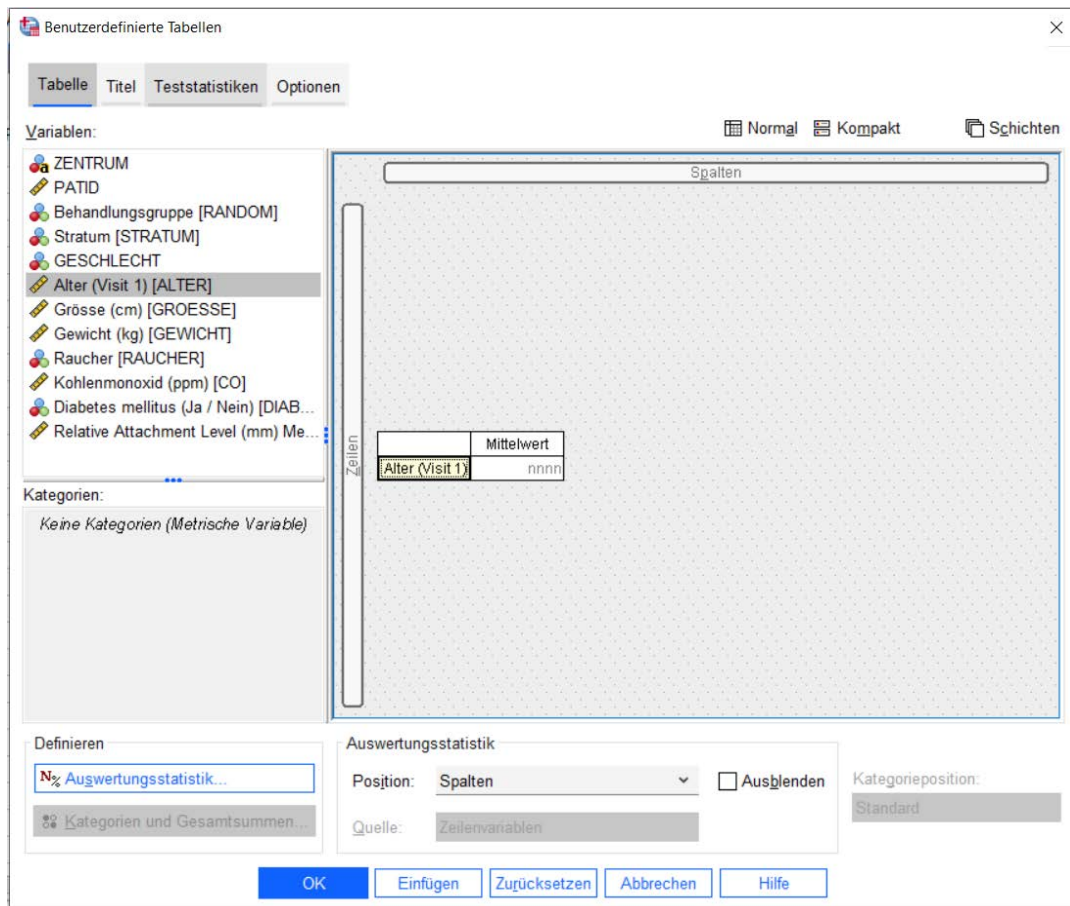


Zurück im Dialogfenster *Explorative Datenanalyse* bestätigen Sie mit **OK**. Die so entwickelte Ausgabe hat die unten gezeigte Form. Die erste Tabelle zeigt die statistischen Maßzahlen aufgeteilt nach dem Wert in der Variable DIABETES. Kategoriale Variablen mit mehr als zwei möglichen Wertbelegungen produzieren eine längere Tabelle, in der alle Möglichkeiten weiter unten angehängt werden.

**Option C:** (Bei dieser Variante keine Statistiken zur Schiefe möglich)

Noch variabelere Tabellen können wie folgt erstellt werden. Über die Befehlsfolge **Analysieren > Tabellen > Benutzerdefinierte Tabellen** gelangen Sie in die Dialogbox *Benutzerdefinierte Tabellen*. Übertragen Sie die interessierende Variable per Drag & Drop in das Feld *Zeilen*. Über die Schaltfläche **Auswertungsstatistik** gelangen Sie in die Dialogbox *Auswertungsstatistik*. Hier können die interessierenden Parameter unter *Statistik* angefordert werden und über die Pfeiltaste ⇨ in das entsprechende Anzeigefeld *Anzeigen* übertragen werden. Bestätigen Sie Ihre Anforderungen über die Schaltfläche **Der Auswahl zuweisen**. Sie gelangen wieder in die übergeordnete Dialogbox *Benutzerdefinierte Tabellen*. Zur besseren Übersichtlichkeit der Ergebnisse können Sie unter *Position* die Kenngrößen zeilen- bzw. spaltenweise anzeigen lassen (Empfehlung: zeilenweise Anzeige). Zusätzlich besteht die Möglichkeit, die Ergebnisse getrennt nach Gruppen aufzuteilen. Dazu übertragen Sie die Gruppenvariable in das Feld *Spalten*. Bestätigen Sie mit **OK**.

Auswertung über das Gesamtkollektiv:



Benutzerdefinierte Tabellen

Tabelle
  Titel
  Teststatistiken
  Optionen

Normal
  Kompakt
  Schichten

Variablen:

- ZENTRUM
- PATID
- Behandlungsgruppe [RANDOM]
- Stratum [STRATUM]
- GESCHLECHT
- Alter (Visit 1) [ALTER]
- Grösse (cm) [GROESSE]
- Gewicht (kg) [GEWICHT]
- Raucher [RAUCHER]
- Kohlenmonoxid (ppm) [CO]
- Diabetes mellitus (Ja / Nein) [DIAB...]
- Relative Attachment Level (mm) Me...

Kategorien:

Keine Kategorien (Metrische Variable)

Spalten		
Zeilen	Alter (Visit 1)	
	Mittelwert	nnnn
	Varianz	nnnn
	Standard ...	nnnn
	Minimum	nnnn
	Maximum	nnnn
	Median	nnnn
	25. Perzentil	nnnn
75. Perzentil	nnnn	

Definieren

N% Auswertungsstatistik...

Kategorien und Gesamtsummen...

Auswertungsstatistik

Position: Zeilen  Ausblenden

Quelle: Zeilenvariablen

Kategorieposition: Standard

**Auswertung getrennt nach Gruppe:**

Benutzerdefinierte Tabellen

Tabelle
  Titel
  Teststatistiken
  Optionen

Normal
  Kompakt
  Schichten

Variablen:

- ZENTRUM
- PATID
- Behandlungsgruppe [RANDOM]
- Stratum [STRATUM]
- GESCHLECHT**
- Alter (Visit 1) [ALTER]
- Grösse (cm) [GROESSE]
- Gewicht (kg) [GEWICHT]
- Raucher [RAUCHER]
- Kohlenmonoxid (ppm) [CO]
- Diabetes mellitus (Ja / Nein) [DIAB...]
- Relative Attachment Level (mm) Me...

Kategorien:

- m
- w

		GESCHLECHT	
		m	w
Alter (Visit 1)	Mittelwert	nnnn	nnnn
	Varianz	nnnn	nnnn
	Standard ...	nnnn	nnnn
	Minimum	nnnn	nnnn
	Maximum	nnnn	nnnn
	Median	nnnn	nnnn
	25. Perzentil	nnnn	nnnn
	75. Perzentil	nnnn	nnnn

Definieren

N% Auswertungsstatistik...

Kategorien und Gesamtsummen...

Auswertungsstatistik

Position:   Ausblenden

Kategorieposition:

Quelle:

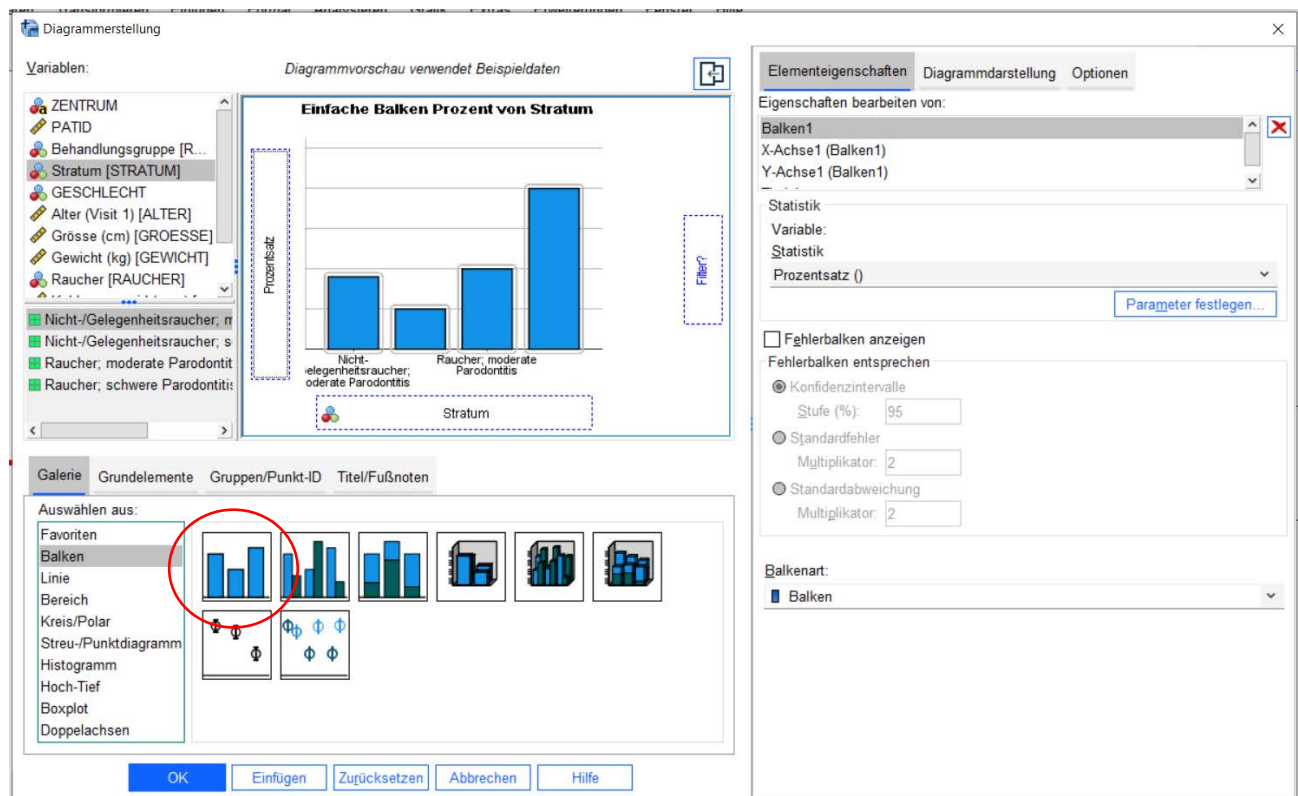
### 6.4 Graphische Darstellungen eines qualitativen Merkmals

Es gibt zwei Möglichkeiten, die Verteilung eines qualitativen Merkmals darzustellen: Balken- und Kreisdiagramme.

#### a) Balkendiagramm

Ein Balkendiagramm eignet sich grundsätzlich für alle qualitativen Merkmale. Die Längen der einzelnen Balken entsprechen den absoluten bzw. relativen Häufigkeiten.

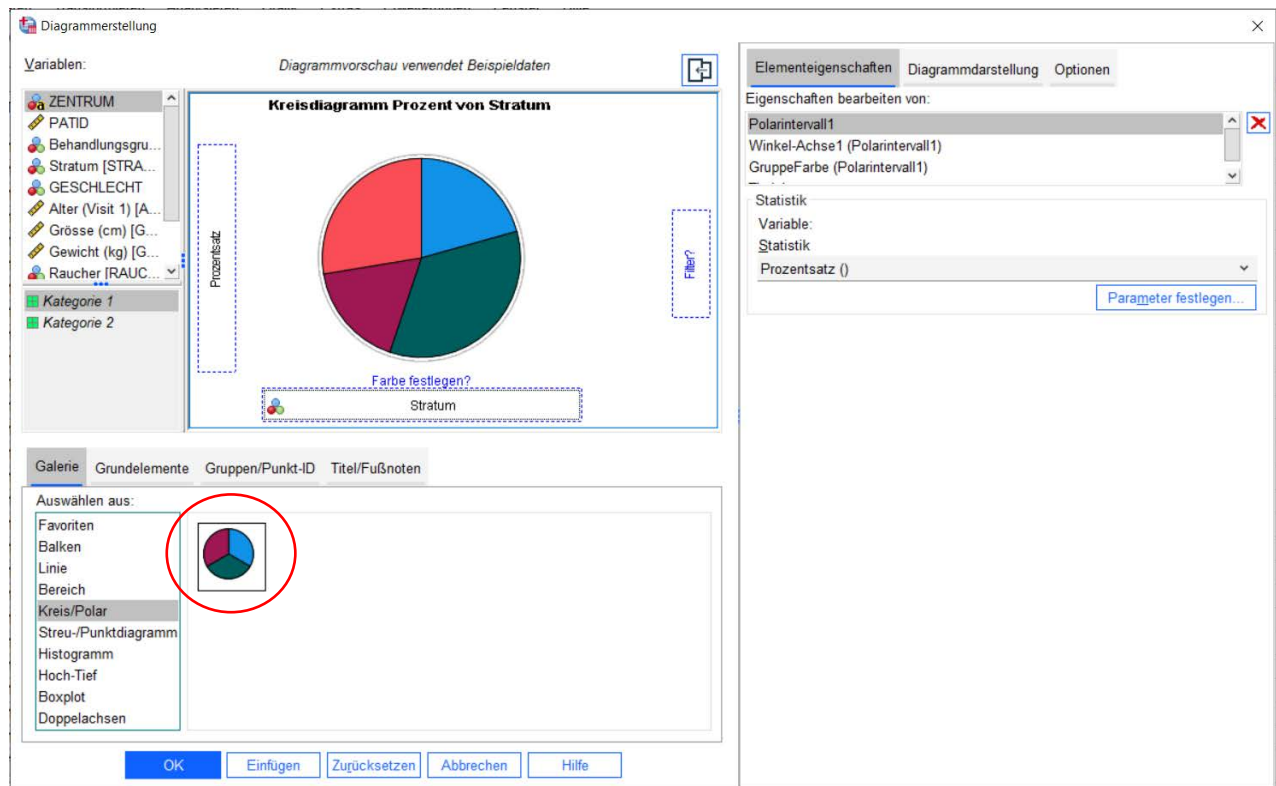
Über die Befehlsfolge **Graphik > Diagrammerstellung** gelangt man in die Dialogbox *Diagrammerstellung*. Unter *Galerie* im Feld *Auswählen aus* kann die entsprechende Graphik angefordert werden. Als Beispiel wählen wir zunächst die Graphik *Balken* und dann *Einfache Balken*. Dieses Galeriediagramm ziehen Sie per Drag & Drop in das Vorschaufenster. Nun müssen die *X-Achse* und *Y-Achse* definiert werden. Zur Definition der *X-Achse* ziehen Sie die interessierende Variable per Drag & Drop in das Feld *X-Achse*. Zur Definition der *Y-Achse* wählen Sie zunächst die Schaltfläche **Elementeigenschaften** aus. Es öffnet sich das Dialogfeld *Elementeigenschaften*. Hier wählen Sie zunächst unter *Eigenschaften bearbeiten* von die Auswahl **Balken1** aus. Aktivieren Sie unter *Statistik* (per default steht die Statistik auf „Anzahl“) **Prozentsatz (?)**. Unter *Parameter festlegen...* ist als *Nenner für die Berechnung des Prozentsatzes* das *Gesamtergebnis* voreingestellt. Nachdem alle Einstellungen getroffen wurden, klicken Sie auf die Schaltfläche **Weiter**. Bestätigen Sie Ihre angeforderten Analysen mit **OK**.





b) Kreisdiagramm (Befehlsfolge analog zum Balkendiagramm)

In einem Kreisdiagramm entspricht der zentrale Winkel des zugeordneten Kreissegmentes der relativen Häufigkeit einer Merkmalsausprägung. Die relative Häufigkeit wird in Prozent angegeben.



Weitergehende Möglichkeiten zur Diagramm-Editierung:

Um die Prozente der einzelnen Merkmalsausprägungen innerhalb der jeweiligen Balken bzw. Kreissegmente anzeigen zu lassen, tätigen Sie einen Doppelklick in die Graphik der Ausgabedatei. Es öffnet sich das Dialogfenster *Diagramm-Editor*. Aktivieren Sie die Schaltfläche **Elemente > Datenbeschriftungen einblenden** und schließen die *Dialogbox* wieder.

Weitere Modifikationen der erzeugten Graphiken können Sie nach dem Doppelklick in die Graphik im *Diagrammeditor* unter *Eigenschaftsfenster anzeigen* vornehmen.

**6.5 Graphische Darstellungen eines quantitativen Merkmals**

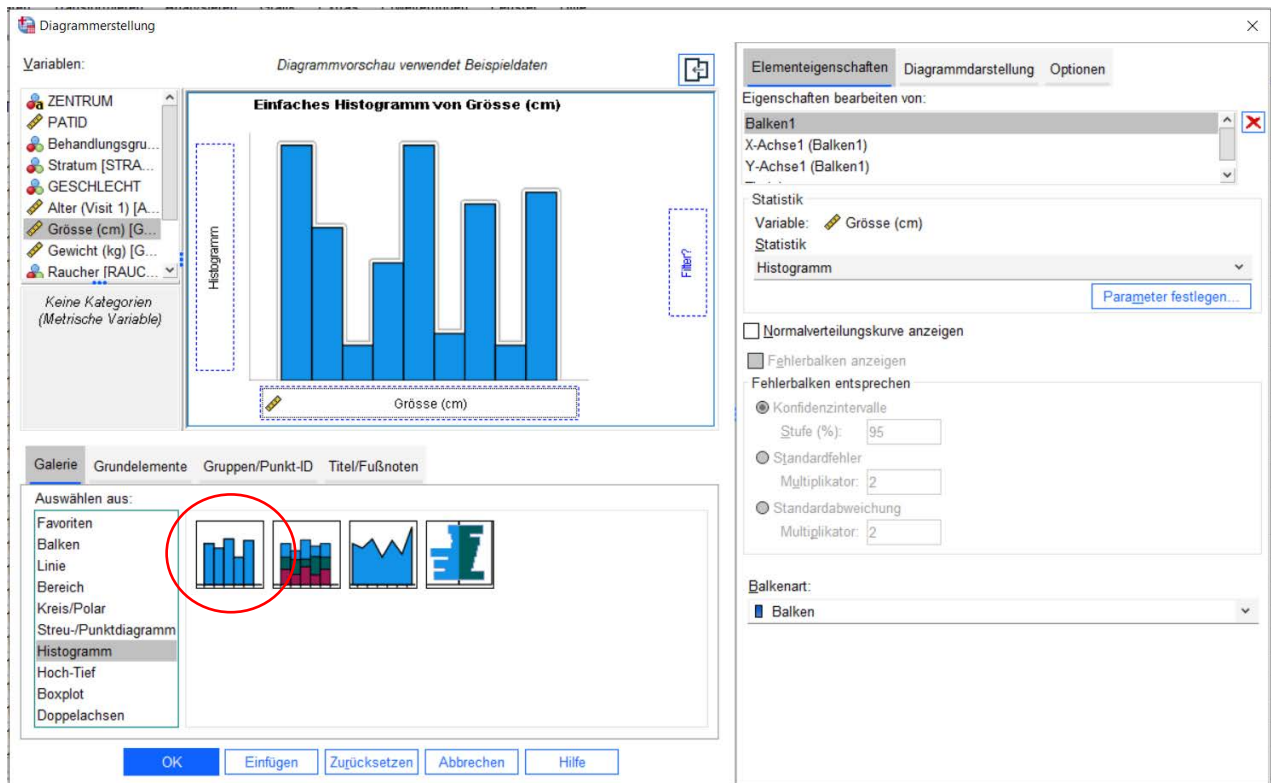
Für die Darstellung quantitativer Merkmale eignen sich Histogramme und Boxplots.

a) Histogramm

Das Histogramm dient zur graphischen Darstellung der Häufigkeitsverteilung quantitativer Merkmale. Die Daten werden der Größe nach in Klassen eingeteilt und jede Klasse durch ein Rechteck dargestellt, dessen Fläche proportional zu den jeweiligen Klassenhäufigkeiten ist.

Über die Befehlsfolge **Graphik > Diagrammerstellung** gelangen Sie in die Dialogbox *Diagrammerstellung*. Unter der Schaltfläche *Galerie* im Feld *Auswählen aus* kann die entsprechende Graphik angefordert werden. Wählen Sie die Graphik *Histogramm* und anschließend *Einfaches Histogramm*. Ziehen Sie dieses

Galeriediagramm per Drag & Drop in das Vorschaufenster. Zur Definition der *X-Achse* ziehen Sie die interessierende Variable per Drag & Drop in das Feld *X-Achse*. Bestätigen Sie Ihre angeforderten Analysen mit **OK**.

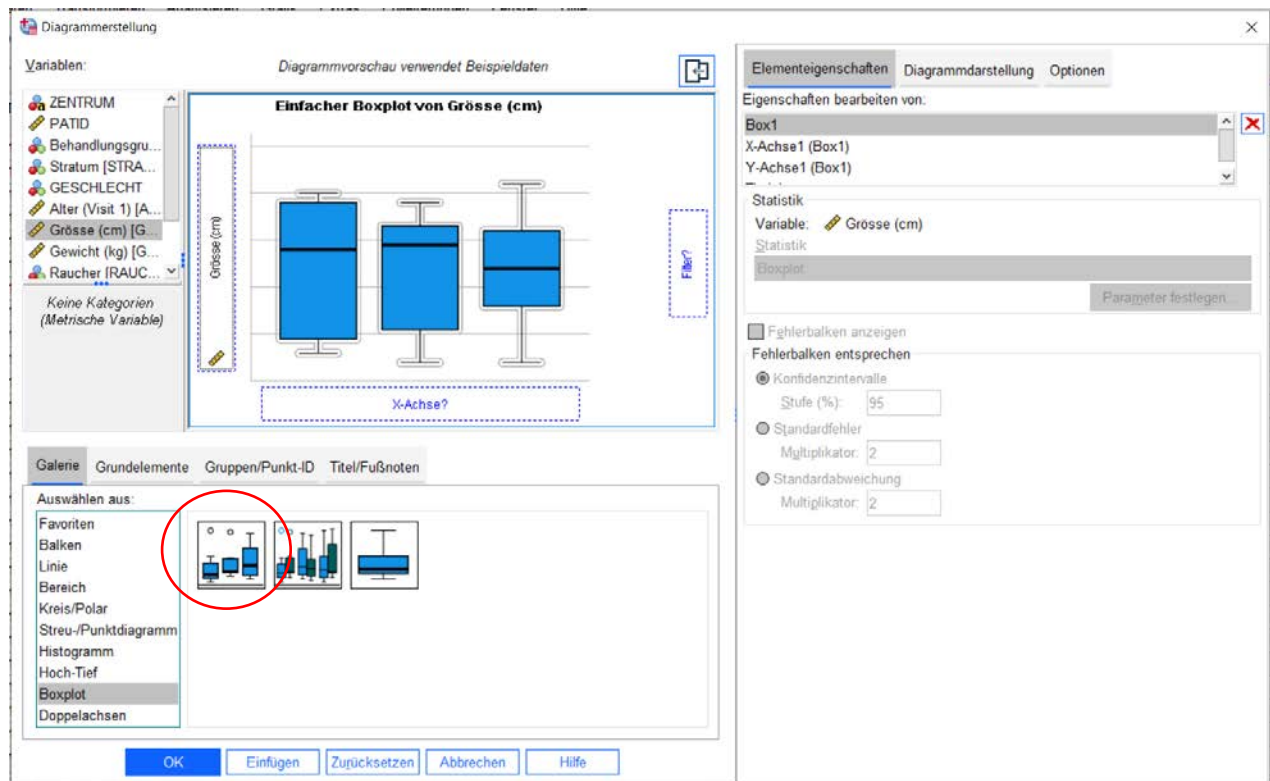


b) Box-and-Whisker-Plot (Boxplot)

Der Boxplot ist eine graphische Darstellung zur Charakterisierung der Verteilung stetiger Merkmale, die auf den empirischen Quartilen beruht. Die Quartile werden als Kasten (engl.: box) dargestellt, der unten und oben vom 1. bzw. 3. Quartil begrenzt wird und 50 % der Stichprobenwerte beinhaltet. Innerhalb der Box wird der Median gekennzeichnet. Die Länge der von der Box ausgehenden Striche (whiskers) ist üblicherweise beschränkt auf das maximal 1,5-fache des Interquartilsabstands (interquartile range, IQR).

Über die Befehlsfolge **Grafik > Diagrammerstellung** gelangen Sie in die Dialogbox *Diagrammerstellung*. Unter *Galerie* im Feld *Auswählen aus* kann die entsprechende Graphik angefordert werden. Wählen Sie die Graphik *Boxplot* und anschließend *Einfacher Boxplot*. Ziehen Sie dieses Galeriediagramm per Drag & Drop in das Vorschaufenster. Zur Definition der *Y-Achse* ziehen Sie die interessierende Variable per Drag & Drop in das Feld *Y-Achse*. Bestätigen Sie Ihre Eingaben mit **OK**.





Weitergehende Möglichkeiten zur Diagramm-Editierung:

Um den Box-and-Whisker-Plot getrennt nach einer Gruppierungsvariablen darzustellen, ziehen Sie zusätzlich zu den obigen Schritten im Dialogfenster *Diagrammerstellung* die Gruppierungsvariable per Drag & Drop in das Feld *X-Achse*.

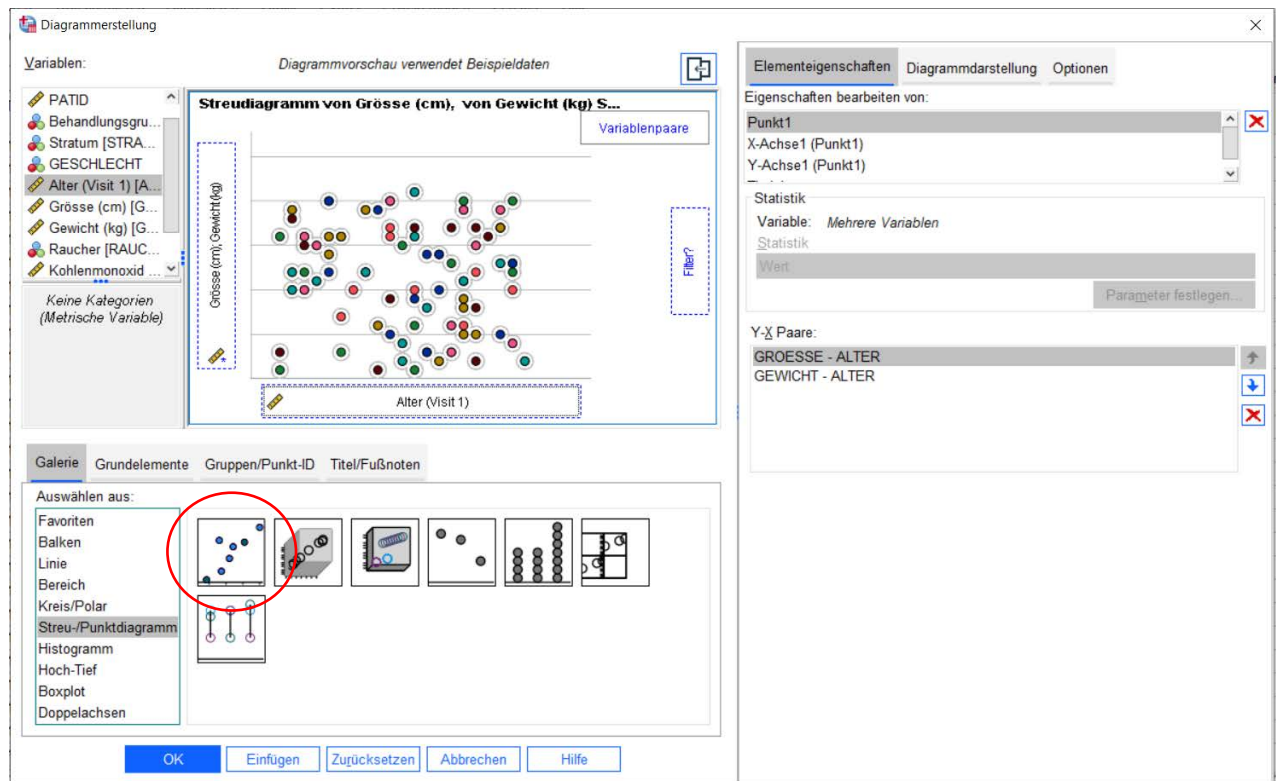
**6.6 Streudiagramme und Korrelationsanalysen**

Die Korrelation beschreibt den Zusammenhang zwischen zwei quantitativen Merkmalen. Um eine erste Übersicht zu erhalten, sollte man eine graphische Darstellung anfertigen. Ein Scatterplot (Streudiagramm) dient zur graphischen Veranschaulichung des Zusammenhangs zweier stetiger Merkmale, die an *n* Beobachtungseinheiten erfasst wurden. Jede Beobachtungseinheit liefert genau einen Punkt für die Punktwolke. Der Korrelationskoeffizient ist eine Maßzahl für die Stärke des Zusammenhangs zwischen zwei Merkmalen und nimmt Werte zwischen -1 und +1 an.

Zur Erzeugung des Scatterplots wählen Sie über die Befehlsfolge **Grafik > Diagrammerstellung** das Dialogfeld *Diagrammerstellung*. Unter *Galerie* im Feld *Auswählen aus* fordern Sie zunächst die Graphik *Streu-/Punktdiagramm* an und wählen dann *Einfaches Streudiagramm*. Dieses Galeriediagramm ziehen Sie per Drag & Drop in das Vorschauenfenster.

*Anmerkung:* Üblicherweise enthält die *X-Achse* das *unabhängige Merkmal (Einflussgröße)* und die *Y-Achse* das *abhängige Merkmal (Zielgröße)*, wenn eine diesbezügliche Unterscheidung möglich ist.

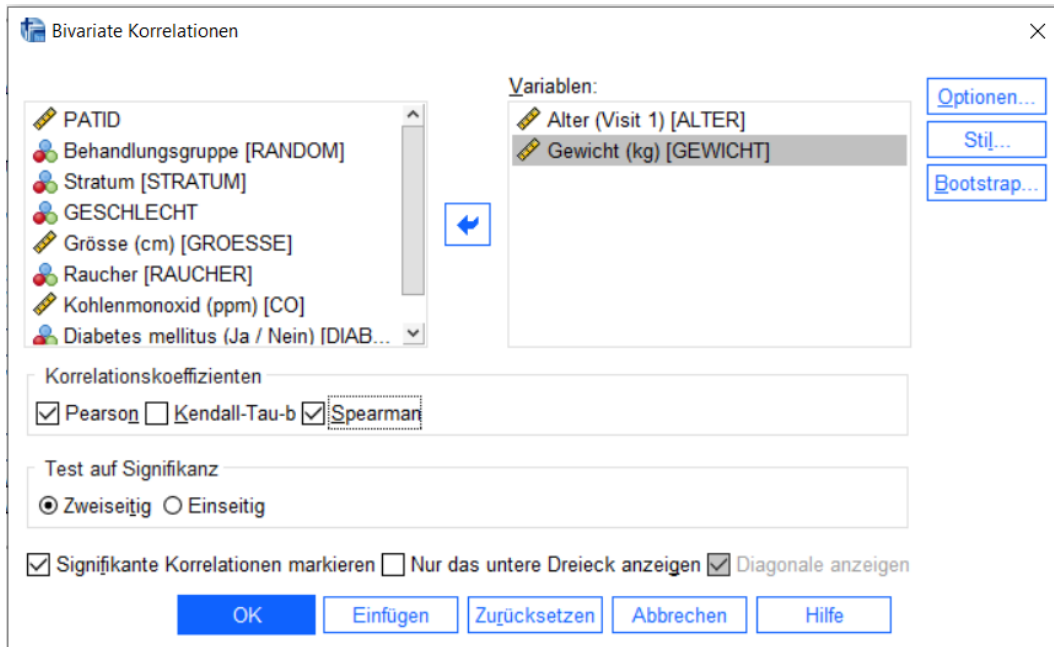
Zur Definition der *X-Achse* ziehen Sie die unabhängige Variable per Drag & Drop in das Feld *X-Achse* und die abhängige Variable in das Feld *Y-Achse*. Bestätigen Sie Ihre Eingaben mit **OK**.



**Hinweis:**

Durch **Doppelklick auf die Abbildung** im Ausgabefenster öffnet sich die Abbildung für eine mögliche Bearbeitung. So lassen sich beispielsweise die Größe der Abbildung oder die verwendeten Farben verändern oder eine Regressionsgerade einzeichnen.

Zur Bestimmung des Korrelationskoeffizienten wählen Sie die Befehlsfolge **Analysieren > Korrelation > Bivariat**. Es öffnet sich das Dialogfenster *Bivariate Korrelationen*. Markieren Sie mit Hilfe der Strg-Taste beide interessierenden Variablen und ziehen Sie diese per Drag & Drop in das Feld *Variablen*. Unter *Korrelationskoeffizienten* aktivieren Sie die gewünschten Koeffizienten (z.B. Pearson oder Spearman). Bestätigen Sie Ihre Eingaben mit **OK**.



Im Viewer werden die Ergebnisse angezeigt:

**Korrelationen**

**Korrelationen**

		Alter (Visit 1)	Gewicht (kg)
Alter (Visit 1)	Pearson-Korrelation	1	<b>-,007</b>
	Sig. (2-seitig)		,876
	N	505	504
Gewicht (kg)	Pearson-Korrelation	-,007	1
	Sig. (2-seitig)	,876	
	N	504	505

**Nichtparametrische Korrelationen**

**Korrelationen**

		Alter (Visit 1)	Gewicht (kg)
Spearman-Rho	Alter (Visit 1)	Korrelationskoeffizient	1,000
		Sig. (2-seitig)	,838
		N	505
	Gewicht (kg)	Korrelationskoeffizient	<b>,009</b>
		Sig. (2-seitig)	,838
		N	504

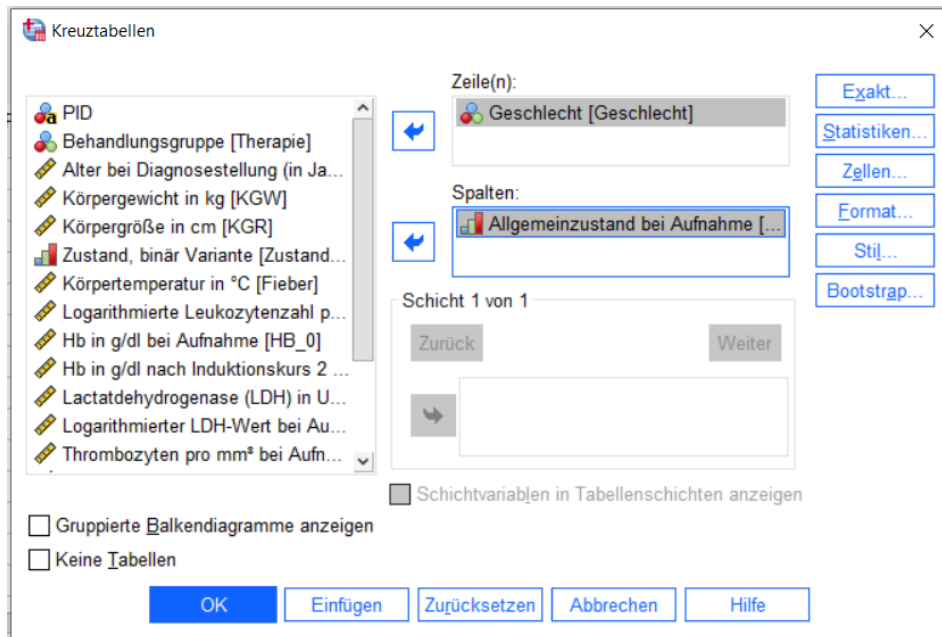
In der Tabelle kann der Korrelationskoeffizient nach Pearson (bzw. Spearman) abgelesen werden. Der Korrelationskoeffizient kann Werte zwischen -1 und 1 annehmen. Je größer der Absolutbetrag der Korrelation, desto stärker ist der zugrunde liegende (Pearson: lineare / Spearman: monotone) Zusammenhang. Dabei zeigt ein positiver Korrelationskoeffizient einen positiven (linearen / monotonen) Zusammenhang und ein negativer Korrelationskoeffizient einen negativen (linearen / monotonen) Zusammenhang an.

## 7. Statistische Signifikanztests zum Vergleich zweier Verteilungen

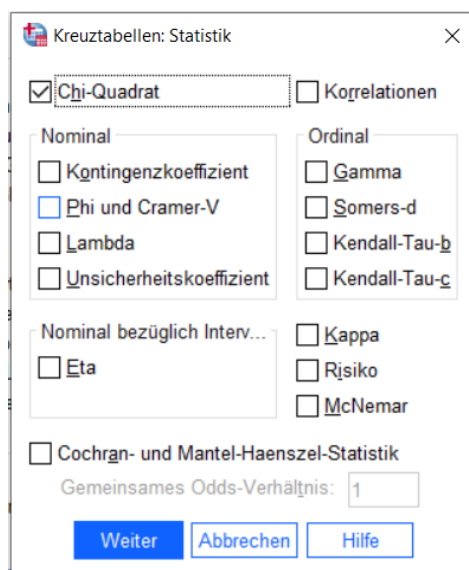
### 7.1 Chi-Quadrat-Test und exakter Test nach Fisher

Zur Analyse des Zusammenhangs zweier kategorialer Merkmale wird der Chi-Quadrat-Test bzw. der exakte Test nach Fisher verwendet.

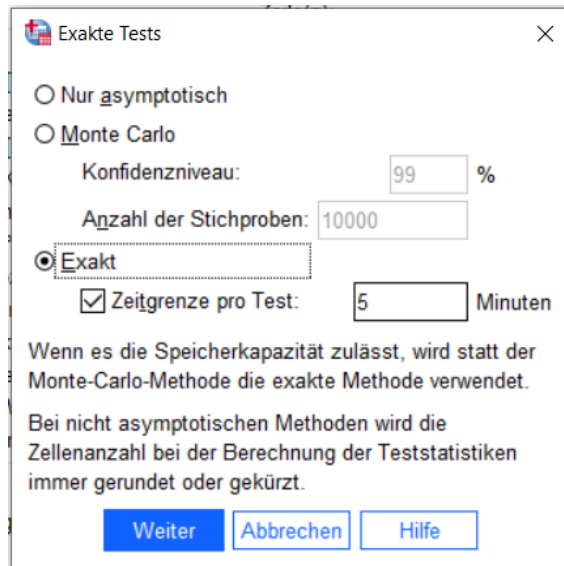
Über die Befehlsfolge **Analysieren > Deskriptive Statistiken > Kreuztabellen** gelangt man in das Dialogfeld Kreuztabellen. Hier ist die Einflussvariable in das Feld *Zeile(n)*, die Zielvariable in das Feld *Spalten* einzutragen.



Unter **Statistik...** wird der **Chi-Quadrat-Test** gewählt. In 2x2-Kreuztabellen kann mit der Option *Risiko* das Relative Risiko berechnet werden.



Im Dialogfeld **Exakt...** ist die Option **Exakt** zu wählen, um im Fall von Kreuztabellen größer als 2x2 exakte Statistiken (Fisher-Test) anzufordern.



Unter **Zellen...** können die Zeilen- bzw. Spaltenprozentage angefordert werden. Die Eingaben werden abschließend mit **Weiter** und **OK** beendet.

In der Ausgabe steht der p-Wert des Chi-Quadrat-Tests unter Chi-Quadrat nach Pearson und asymptotische Signifikanz (2-seitig). Der p-Wert des exakten Tests nach Fisher steht unter Exakter Test nach Fisher und Exakte Signifikanz (2-seitig).

Verwenden Sie in der Regel den exakten Test nach Fisher. Wenn der exakte Test nach Fisher nicht durchgeführt werden kann, verwenden Sie den Chi-Quadrat-Test.

Hinweis: Ein möglicher Eintrag der Form „Sig. = ,000“ ist so zu lesen, dass der p-Wert  $p < 0,001$  ist.

**Chi-Quadrat-Tests**

	Wert	df	Asymptotische Signifikanz (zweiseitig)	Exakte Sig. (zweiseitig)	Exakte Sig. (einseitig)	Punkt-Wahrscheinlichkeit
Pearson-Chi-Quadrat	1,050 <sup>a</sup>	2	,591	,592		
Likelihood-Quotient	1,046	2	,593	,592		
Exakter Test nach Fisher-Freeman-Halton	1,062			,592		
Zusammenhang linear-mit-linear	,210 <sup>b</sup>	1	,647	,669	,346	,044
Anzahl der gültigen Fälle	977					

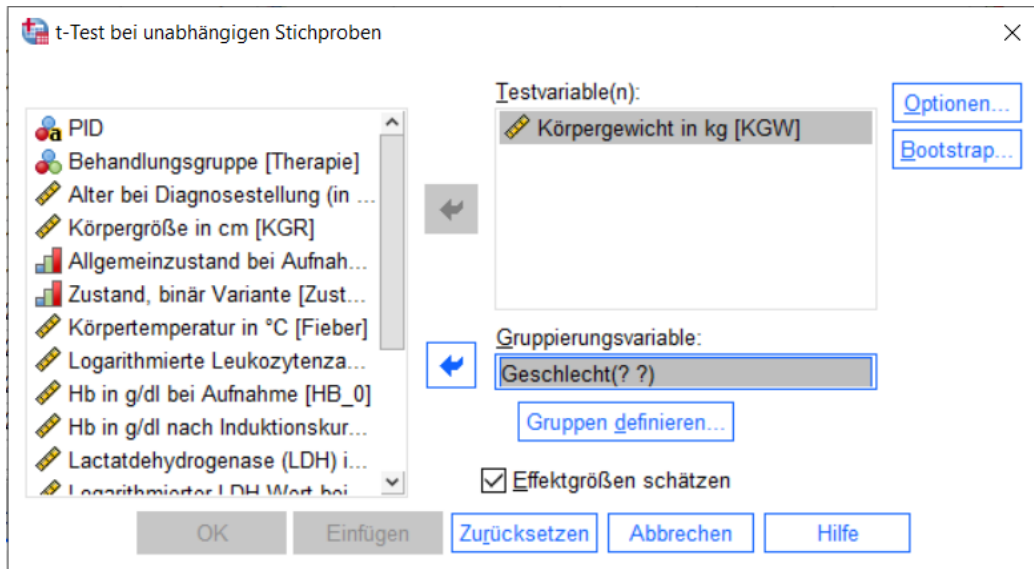
a. 0 Zellen (0,0%) haben eine erwartete Häufigkeit kleiner 5. Die minimale erwartete Häufigkeit ist 33,85.

b. Die standardisierte Statistik ist ,458.

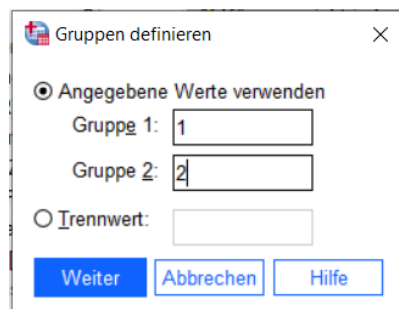
## 7.2 Parametrische Tests bei normalverteilten Zielgrößen

### 7.2.1 t-Test für unverbundene Stichproben

Wählen Sie **Analysieren > Mittelwerte vergleichen > t-Test bei unabhängigen Stichproben**.



Die *Testvariable* ist die Zielgröße und die *Gruppenvariable* die Einflussgröße. Unter **Gruppen def. ...** wird die Kodierung der Gruppenvariablen eingetragen.



Die Eingaben werden mit **Weiter** und **OK** abgeschlossen.

Ausgabe: Der p-Wert des t-Tests und das Konfidenzintervall des Mittelwertunterschieds sind im Screenshot rot markiert. Wir gehen standardmäßig davon aus, dass die Varianzen der Zielgröße in den beiden zu vergleichenden Gruppen nicht gleich sind, weshalb die Ergebnisse in der unteren Zeile abzulesen sind.

Hinweis: Ein möglicher Eintrag der Form „Sig. = ,000“ ist so zu lesen, dass der p-Wert  $p < 0,001$  ist.

**t-Test**

**Gruppenstatistiken**

	Geschlecht	N	Mittelwert	Standardabweichung	Standardfehler des Mittelwertes
Körpergewicht in kg	Männlich	527	82,61	13,770	,600
	Weiblich	442	72,12	15,220	,724

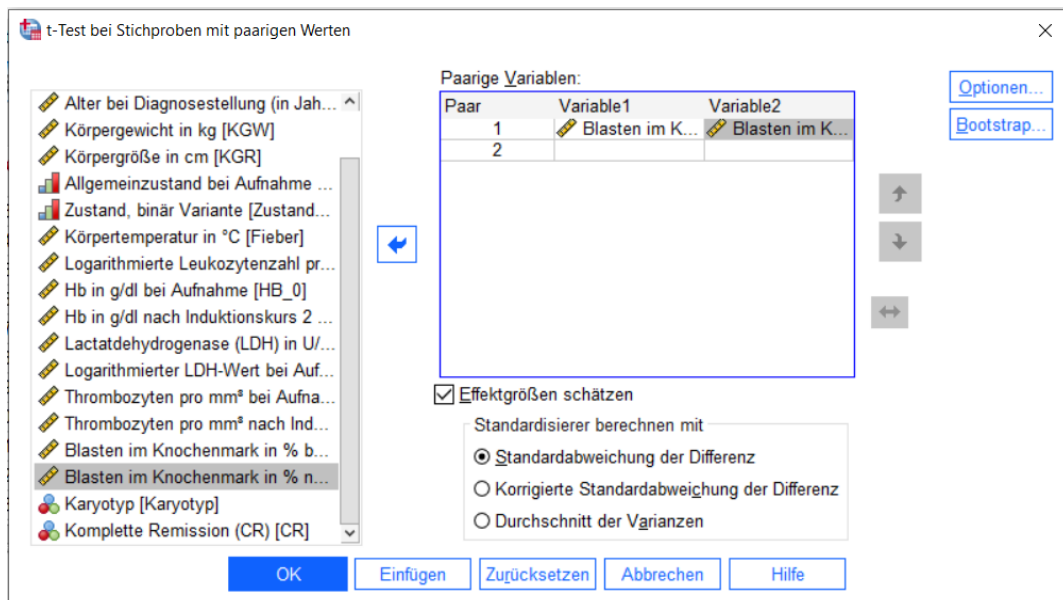
  

**Test bei unabhängigen Stichproben**

		Levene-Test der Varianzgleichheit		t-Test für die Mittelwertgleichheit					95% Konfidenzintervall der Differenz	
		F	Sig.	T	df	Sig. (2-seitig)	Mittlere Differenz	Standardfehler der Differenz	Unterer Wert	Oberer Wert
Körpergewicht in kg	Varianzen sind gleich	4,737	,030	11,264	967	,000	10,497	,932	8,668	12,326
	Varianzen sind nicht gleich			11,165	899,070	,000	10,497	,940	8,652	12,342

**7.2.2 t-Test für verbundene Stichproben**

Wählen Sie **Analysieren > Mittelwerte vergleichen > t-Test bei verbundenen Stichproben**. Wählen Sie als *Variable1* die erste Messung und als *Variable2* die zweite und bestätigen Sie mit **OK**.



Die wichtigsten Werte der Ausgabe sind rot markiert.

Hinweis: Ein möglicher Eintrag der Form „Sig. = ,000“ ist so zu lesen, dass der p-Wert  $p < 0,001$  ist.



**t-Test**

**Statistik bei gepaarten Stichproben**

		Mittelwert	N	Standardabweichung	Standardfehler des Mittelwertes
Paaren 1	Blasten im Knochenmark in % bei Aufnahme	71,73	395	22,561	1,135
	Blasten im Knochenmark in % nach Induktionskurs 2	8,36	395	16,578	,834

**Korrelationen bei gepaarten Stichproben**

		N	Korrelation	Sig.
Paaren 1	Blasten im Knochenmark in % bei Aufnahme & Blasten im Knochenmark in % nach Induktionskurs 2	395	,057	,256

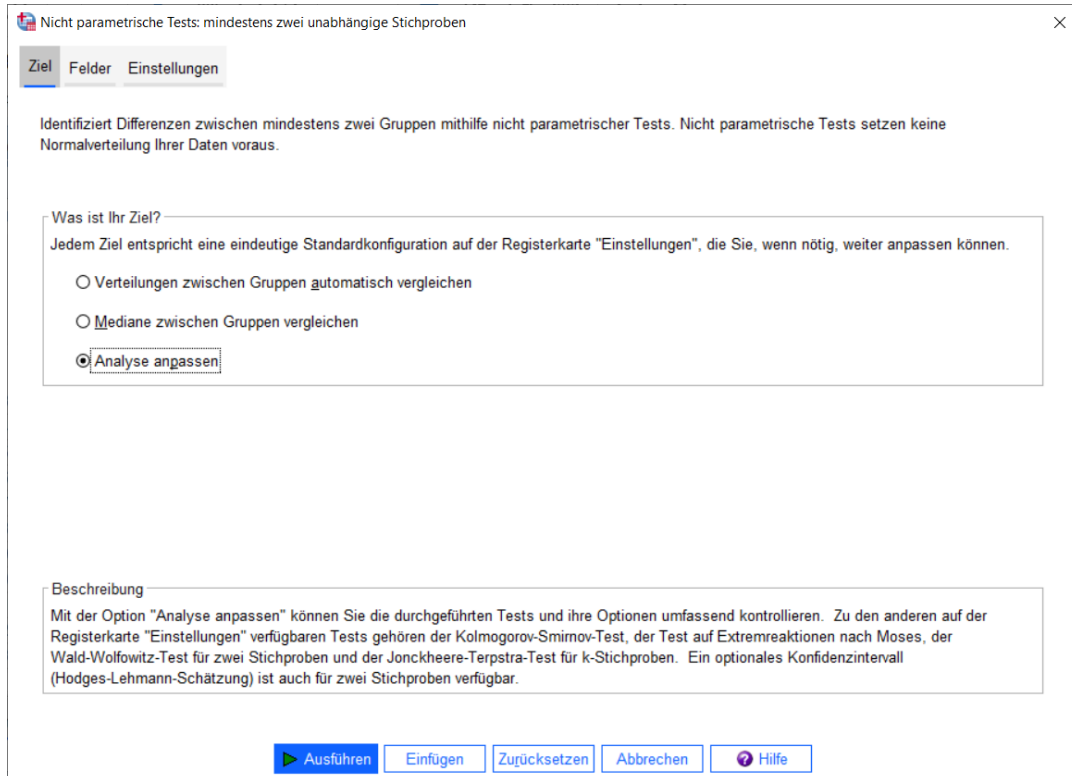
**Test bei gepaarten Stichproben**

		Mittelwert	Standardabweichung	Standardfehler des Mittelwertes	Gepaarte Differenzen		T	df	Sig. (2-seitig)
					95% Konfidenzintervall der Differenz				
					Unterer Wert	Oberer Wert			
Paaren 1	Blasten im Knochenmark in % bei Aufnahme - Blasten im Knochenmark in % nach Induktionskurs 2	63,377	27,220	1,370	60,685	66,070	46,274	394	,000

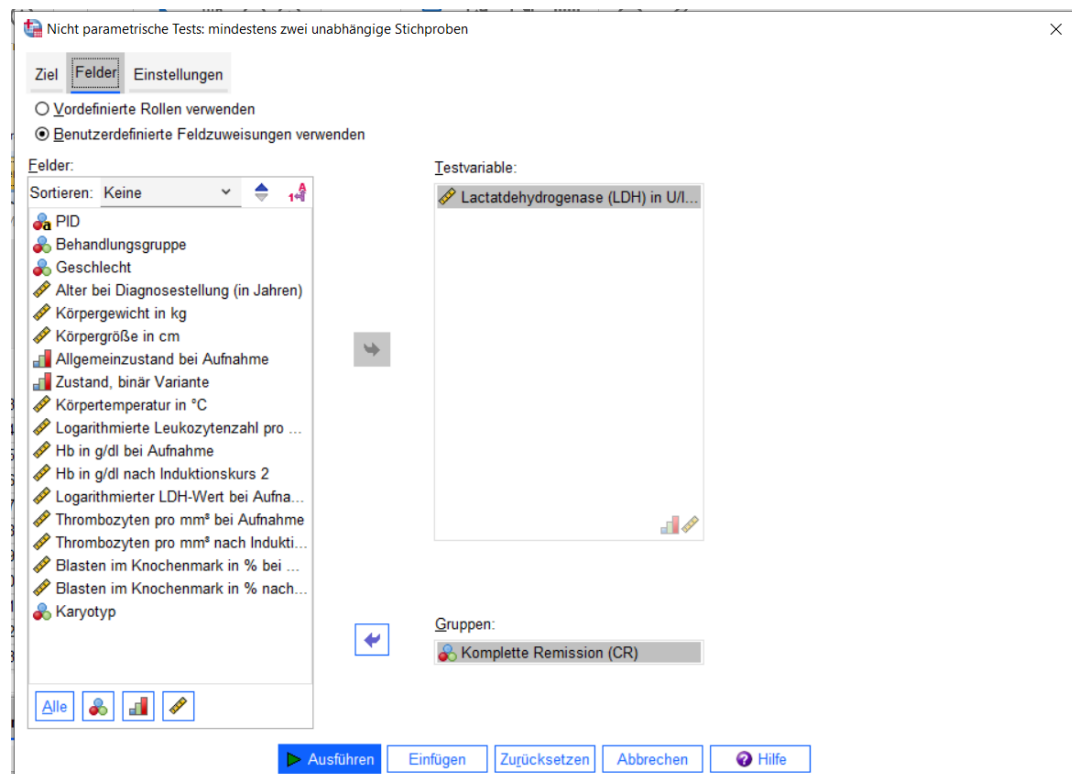
### 7.3 Nichtparametrische Tests

#### 7.3.1 Mann-Whitney-U Test (Wilcoxon-Rangsummen-Test)

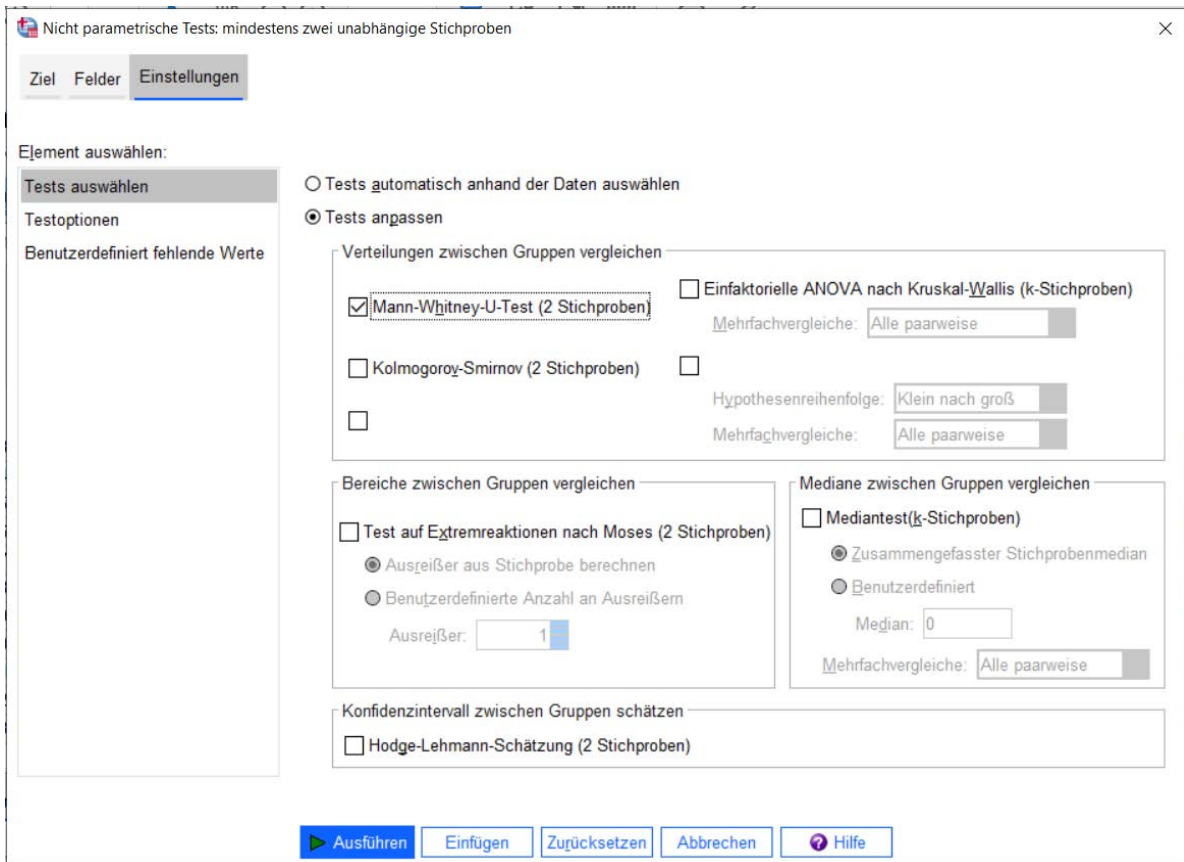
Wählen Sie **Analysieren > Nichtparametrische Tests > Unabhängige Stichproben**. Wählen Sie im Reiter *Ziel* die Option **Analyse anpassen**.



Im Reiter *Felder* wird im Feld *Testvariable* die Zielgröße eingetragen und als Gruppe die Gruppierungsvariable.



Im Reiter *Einstellungen* wird als Test der **Mann-Whitney-U-Test** gewählt.



Die Eingabe wird abgeschlossen durch den Button **Ausführen**.

**Nicht parametrische Tests**

**Hypothesentestübersicht**

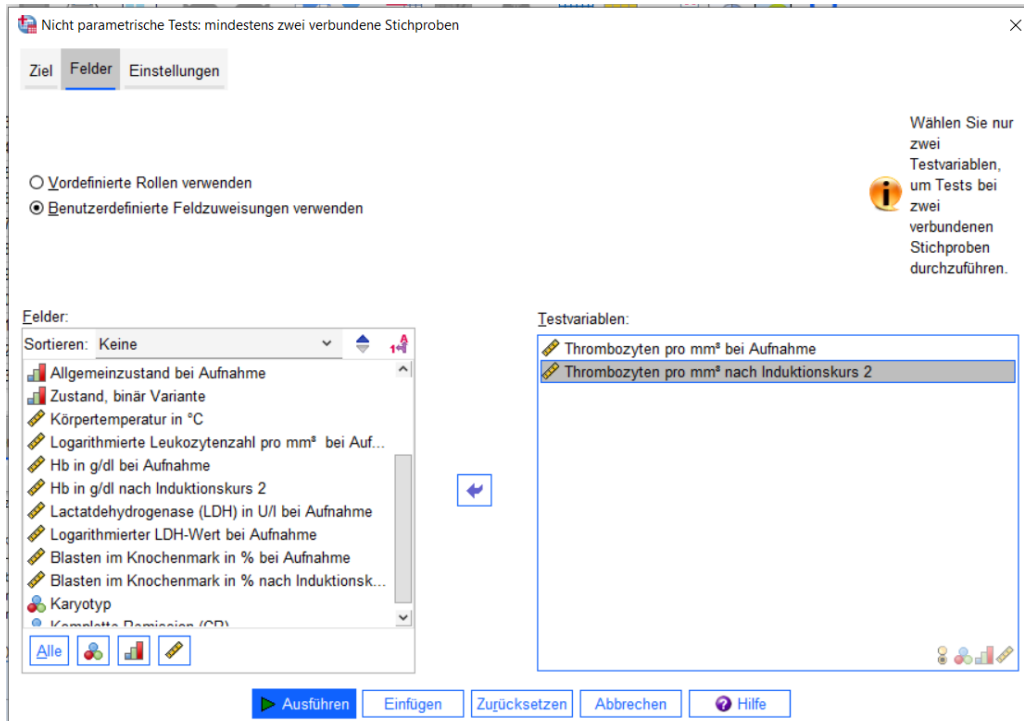
	Nullhypothese	Test	Sig. <sup>a,b</sup>	Entscheidung
1	Die Verteilung von Lactatdehydrogenase (LDH) in U/I bei Aufnahme ist über die Kategorien von Komplette Remission (CR) identisch.	Mann-Whitney-U-Test bei unabhängigen Stichproben	,575	Nullhypothese beibehalten

a. Das Signifikanzniveau ist ,050.

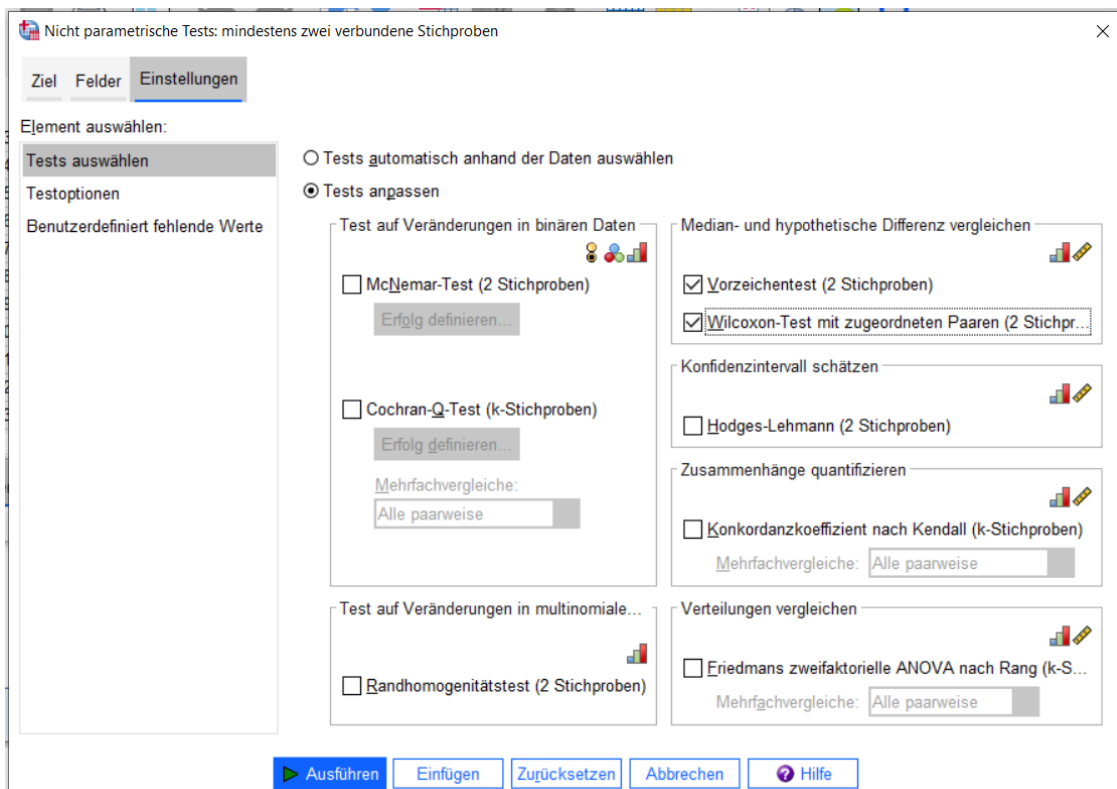
b. Asymptotische Signifikanz wird angezeigt.

### 7.3.2 Wilcoxon-Vorzeichen-Rang Test und Vorzeichentest

Wählen Sie **Analysieren > Nichtparametrische Tests > Verbundene Stichproben**. Wählen Sie im Reiter *Ziel* die Option **Analyse anpassen**. Im Reiter *Felder* werden als *Testvariablen* die beiden zu vergleichenden Variablen eingetragen.



Im Reiter *Einstellungen* wird als Test der *Wilcoxon-Test* bzw. der *Vorzeichentest* gewählt.



Die Eingabe wird mit **Ausführen** beendet. In der Ausgabe wird unter „Sig.“ der p-Wert angegeben und unter „Entscheidung“ die zugehörige Testentscheidung abhängig vom Signifikanzniveau (vgl. Anmerkung a).

Hinweis: Ein möglicher Eintrag der Form „Sig. = ,000“ ist so zu lesen, dass der p-Wert  $p < 0,001$  ist.

**Nicht parametrische Tests**

<b>Hypothesentestübersicht</b>				
	Nullhypothese	Test	Sig. <sup>a,b</sup>	Entscheidung
1	Der Median der Differenzen zwischen Thrombozyten pro mm <sup>3</sup> bei Aufnahme und Thrombozyten pro mm <sup>3</sup> nach Induktionskurs 2 ist gleich 0.	Vorzeichentest bei verbundenen Stichproben	,000	Nullhypothese ablehnen
2	Der Median der Differenzen zwischen Thrombozyten pro mm <sup>3</sup> bei Aufnahme und Thrombozyten pro mm <sup>3</sup> nach Induktionskurs 2 ist gleich 0.	Wilcoxon-Test bei verbundenen Stichproben	,000	Nullhypothese ablehnen

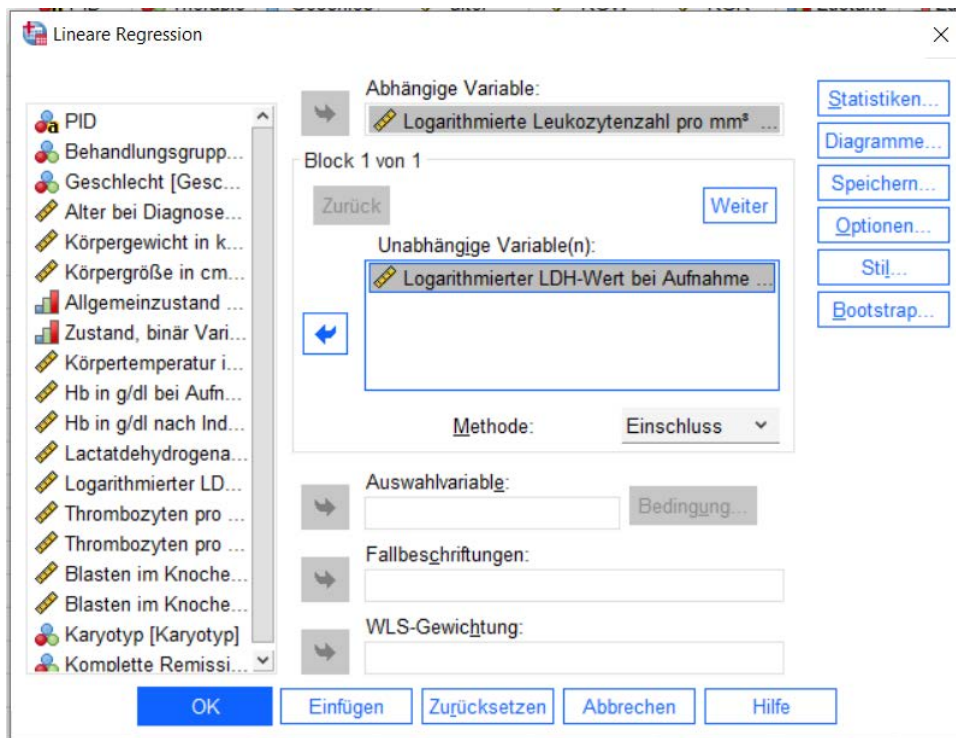
a. Das Signifikanzniveau ist ,050.

b. Asymptotische Signifikanz wird angezeigt.

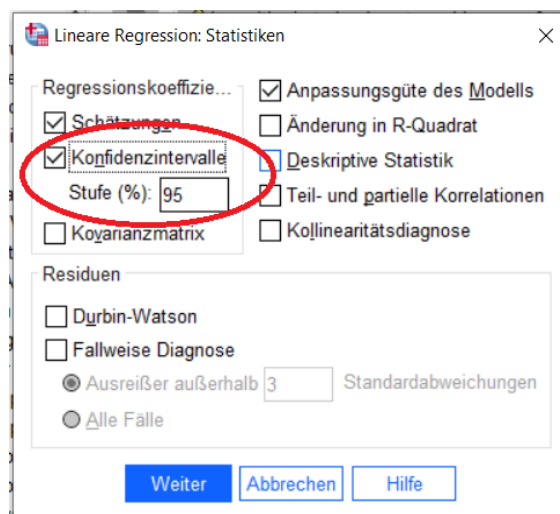
### 8. Lineare Regression

Die lineare Regression eignet sich zur Bestimmung des Einflusses einer oder mehrerer Einflussgrößen auf eine stetige Zielgröße.

Um eine lineare Regressionsanalyse durchzuführen, wählt man **Analysieren > Regression > Linear**. Aus der Variablenliste werden die *abhängige Variable* und *unabhängigen Variable(n)* ausgewählt und in die entsprechenden Felder eingetragen. Die abhängige Variable ist die Zielgröße und die unabhängige Variable die Einflussgröße.



Um zusätzlich zu den Regressionskoeffizienten des Regressionsmodells die zugehörigen Konfidenzintervalle ausgeben zu lassen, klickt man auf den Button **Statistiken** und setzt im folgenden Fenster ein Häkchen bei Konfidenzintervalle.



Nach Bestätigung mit **Weiter** und **OK** wird der folgende Output erzeugt, in dem relevante Kenngrößen rot hervorgehoben sind:

**Regression**

**Aufgenommene/Entfernte Variablen<sup>a</sup>**

Modell	Aufgenommene Variablen	Entfernte Variablen	Methode
1	Logarithmierter LDH-Wert bei Aufnahme <sup>b</sup>		Einschluß

a. Abhängige Variable: Logarithmierte Leukozytenzahl pro mm<sup>3</sup> bei Aufnahme

b. Alle gewünschten Variablen wurden eingegeben.

**Modellzusammenfassung**

Modell	R	R-Quadrat	Korrigiertes R-Quadrat	Standardfehler des Schätzers
1	,528 <sup>a</sup>	,279	,278	1,36961

a. Einflußvariablen : (Konstante), Logarithmierter LDH-Wert bei Aufnahme

**ANOVA<sup>a</sup>**

Modell		Quadratsumme	df	Mittel der Quadrate	F	Sig.
1	Regression	712,768	1	712,768	379,973	,000 <sup>b</sup>
	Nicht standardisierte Residuen	1840,200	981	1,876		
	Gesamt	2552,968	982			

a. Abhängige Variable: Logarithmierte Leukozytenzahl pro mm<sup>3</sup> bei Aufnahme

b. Einflußvariablen : (Konstante), Logarithmierter LDH-Wert bei Aufnahme

**Koeffizienten<sup>a</sup>**

Modell		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten	T	Sig.	95,0% Konfidenzintervalle für B	
		Regressionskoeffizient B	Std.-Fehler				Untergrenze	Obergrenze
1	(Konstante)	3,096	,331		9,368	,000	2,448	3,745
	Logarithmierter LDH-Wert bei Aufnahme	1,046	,054	,528	19,493	,000	,940	1,151

a. Abhängige Variable: Logarithmierte Leukozytenzahl pro mm<sup>3</sup> bei Aufnahme

In der Tabelle *Modellzusammenfassung* ist das Bestimmtheitsmaß R<sup>2</sup> angegeben, das beschreibt, welcher Anteil der Varianz der abhängigen Variable durch das Modell erklärt wird. Das Bestimmtheitsmaß entspricht dem quadrierten Korrelationskoeffizienten nach Pearson. Im Gegensatz zum Korrelationskoeffizienten gibt das Bestimmtheitsmaß nicht die Richtung des Zusammenhangs (d.h. positiv oder negativ) an.

In der Tabelle *Koeffizienten* werden die Regressionskoeffizienten angegeben, die zugehörigen p-Werte sowie die oberen und unteren Grenzen der 95%-Konfidenzintervalle. Der geschätzte Regressionskoeffizient entspricht der Steigung der Geraden, die in Abschnitt 1 in das Streudiagramm gezeichnet wurde. Anhand des zugehörigen p-Wertes lässt sich ermitteln, ob sich der Koeffizient von 0 unterscheidet.

Im Falle eines positiven linearen Zusammenhangs ist der Regressionskoeffizient positiv, bei einem negativen Zusammenhang ist er negativ. Anders als beim Bestimmtheitsmaß und dem

Korrelationskoeffizienten ist der Wertebereich des Regressionskoeffizienten nicht auf ein bestimmtes Intervall eingeschränkt. Der Regressionskoeffizient kann theoretisch jede mögliche reelle Zahl annehmen.

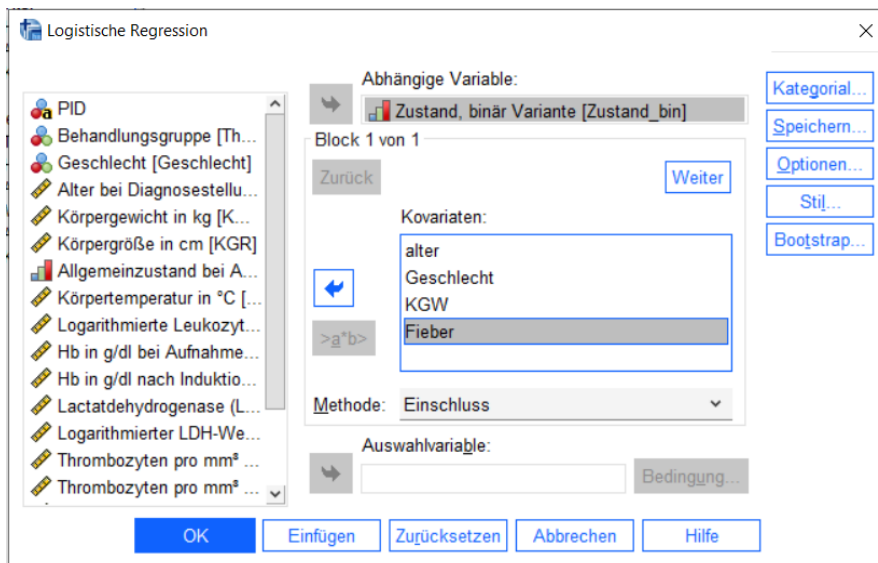
Der Wert des Y-Achsenabschnitts gibt an, wo die Regressionsgerade die Y-Achse schneidet. Dies entspricht dem geschätzten Wert der abhängigen Variablen, wenn die unabhängige Variable den Wert 0 annimmt, d.h.  $x = 0$ . Inhaltlich ist der Y-Achsenabschnitt oft nicht interpretierbar.

### 9. Logistische Regression

In einer Regressionsanalyse ergibt sich die zu verwendende Methode aus dem Messniveau der Zielgröße. Im Falle einer stetigen Zielgröße wird eine lineare Regressionsanalyse durchgeführt. Ist die Zielgröße binär, so führt man eine logistische Regression durch. In einer logistischen Regressionsanalyse wird der Einfluss verschiedener Variablen auf die Wahrscheinlichkeit (bzw. die Odds, d.h. die Chance), stark beeinträchtigt zu sein, modelliert.

Um eine logistische Regressionsanalyse durchzuführen, wählt man **Analysieren > Regression > Binär logistisch**.

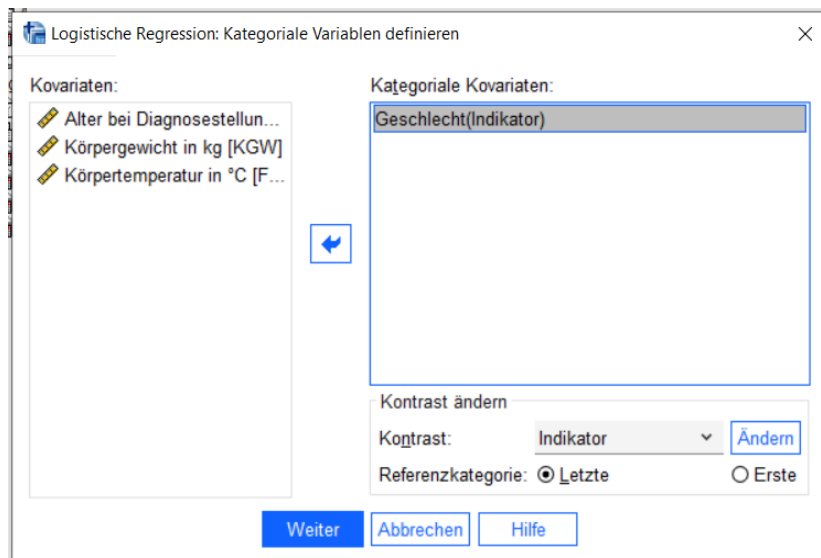
Als *abhängige Variable* wird die binäre Zielgröße aus der Variablenliste ausgewählt und eingetragen, die Einflussgrößen werden als *Kovariaten* eingetragen.



Kategoriale (d.h. nominale und ordinale) Merkmale müssen als solche gekennzeichnet werden. Dies geschieht durch Klicken auf den Button **Kategorial...**

Es öffnet sich ein neues Dialogfenster, in dem die kategorialen Variablen definiert werden können. Dazu werden ordinale und nominale Merkmale aus der angezeigten Liste der Kovariaten ausgewählt und in die Liste der *kategorialen Kovariaten* eingefügt. Für die einzelnen kategorialen Variablen kann anschließend die verwendete Codierung festgelegt werden, indem entweder die *erste* oder die *letzte* Kategorie der Variable als Referenz ausgewählt wird. Das Umcodieren muss durch Klicken auf den Button **Ändern** bestätigt werden.



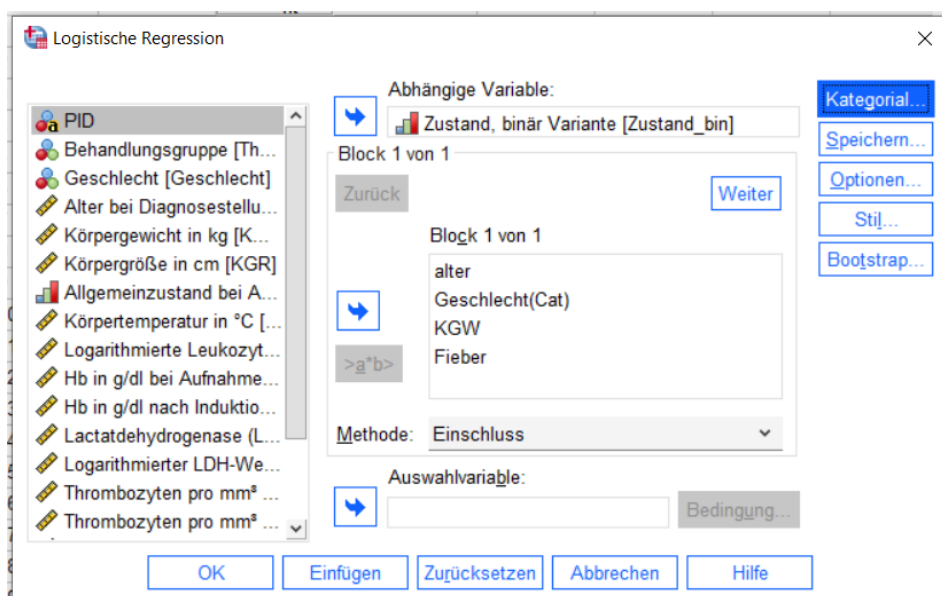


Die aktuelle Kodierung der Variable kann über einen Rechtsklick auf die Variable Geschlecht und **Variablenbeschreibung** angezeigt werden.



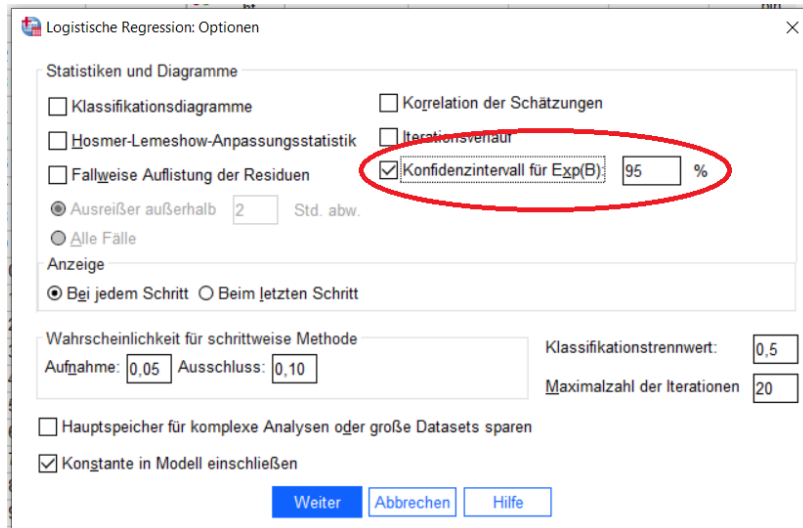
Per Voreinstellung wird die letzte Kategorie als Referenz gewählt.

Durch Bestätigen mit **OK** wird die Definition der kategorialen Variablen in das Modell übernommen. Dass die kategorialen Variablen erfolgreich definiert wurden, ist daran zu erkennen, dass sie in der Liste der Kovariaten anschließend durch die zusätzliche Angabe (*Cat*) gekennzeichnet sind.



Die interessierenden Kenngrößen in einer logistischen Regressionsanalyse sind die so genannten Odds Ratios. Sie geben den Faktor an, mit dem sich eine Einflussvariable auf die abhängige Variable auswirkt. Die zugehörigen Konfidenzintervalle zeigen die Präzision des entsprechenden Schätzwerts an und sollten in der Regel zusätzlich zum Odds Ratio angegeben werden. Enthält das Konfidenzintervall den Wert 1, so ist das zugehörige Odds Ratio nicht signifikant.

Um die Konfidenzintervalle der Odds Ratios ausgeben zu lassen, klickt man auf den Button **Optionen...** Anschließend setzt man ein **Häkchen** bei dem *Konfidenzintervall für Exp(B)* und bestätigt diese Auswahl mit **Weiter**.



Wenn alle notwendigen Einstellungen vorgenommen wurden, bestätigt man im Hauptdialogfenster mit **OK**, dass die logistische Regressionsanalyse durchgeführt werden soll.

Im Folgenden werden die wichtigsten Tabellen des SPSS-Outputs erläutert.

**Logistische Regression**

**Zusammenfassung der Fallverarbeitung**

Ungewichtete Fälle <sup>a</sup>		N	Prozent
Ausgewählte Fälle	Einbezogen in Analyse	847	84,1
	Fehlende Fälle	160	15,9
	Gesamt	1007	100,0
Nicht ausgewählte Fälle		0	,0
Gesamt		1007	100,0

a. Wenn die Gewichtung wirksam ist, finden Sie die Gesamtzahl der Fälle in der Klassifizierungstabelle.

**Codierung abhängiger Variablen**

Ursprünglicher Wert	Interner Wert
Normal oder leicht beeinträchtigt	0
Stark beeinträchtigt	1

**Codierungen kategorialer Variablen**

		Häufigkeit	Parametercodierung (1)
Geschlecht	Männlich	465	1,000
	Weiblich	382	,000

Die erste Tabelle gibt an, wie viele Fälle in der Regressionsanalyse verwendet werden und wie viele auf Grund fehlender Werte in der abhängigen oder den unabhängigen Variablen nicht verwendet werden können. In der zweiten Tabelle wird die Codierung der abhängigen Variablen angegeben. Der Wert „0“ entspricht dem Referenzzustand und „1“ dem Alternativzustand.

Anschließend wird die Codierung der kategorialen Variablen angegeben. Die Angaben in der Tabelle „Codierungen kategorialer Variablen“ sind wichtig für das Verständnis der zu den kategorialen Variablen gehörenden Odds Ratios (siehe unten).

Der als Omnibus-Test der Modellkoeffizienten bezeichnete Test ist ein Likelihood-Quotienten-Test, der die Gesamtanpassung des Modells überprüft. Ein p-Wert von  $p \leq 0,05$  zeigt an, dass die gewählten Einflussvariablen bzw. das erstellte Modell grundsätzlich geeignet zur Beschreibung der abhängigen Variable ist.

**Omnibus-Tests der Modellkoeffizienten**

		Chi-Quadrat	df	Sig.
Schritt 1	Schritt	63,743	4	,000
	Block	63,743	4	,000
	Modell	63,743	4	,000

In der Tabelle *Variablen in der Gleichung* befinden sich die geschätzten Odds Ratios des angepassten Modells in der mit  $Exp(B)$  überschriebenen Spalte. Die Bezeichnung beruht darauf, dass die Odds Ratios die exponierten Regressionskoeffizienten darstellen.

**Variablen in der Gleichung**

		Regressions koeffizient B	Standardfehler	Wald	df	Sig.	Exp(B)	95% Konfidenzintervall für EXP (B)	
			r					Unterer Wert	Oberer Wert
Schritt 1 <sup>a</sup>	Alter bei Diagnosestellung (in Jahren)	,027	,012	5,563	1	,018	1,028	1,005	1,051
	Geschlecht(1)	-,223	,284	,616	1	,433	,801	,459	1,396
	Körpergewicht in kg	-,001	,009	,018	1	,892	,999	,981	1,017
	Körpertemperatur in °C	,867	,116	55,668	1	,000	2,379	1,895	2,987
	Konstante	-36,618	4,657	61,830	1	,000	,000		

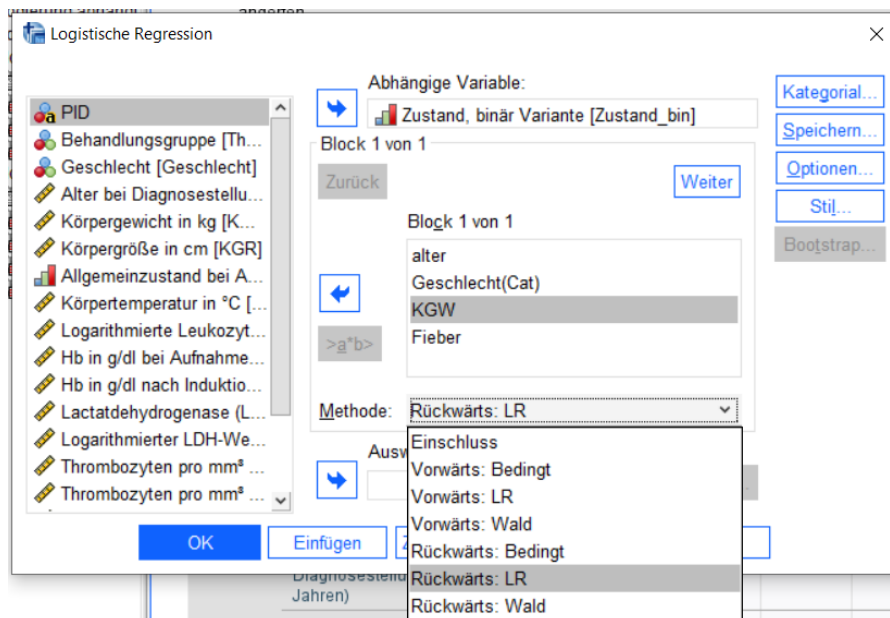
a. In Schritt 1 eingegebene Variablen: Alter bei Diagnosestellung (in Jahren), Geschlecht, Körpergewicht in kg, Körpertemperatur in °C.

In den beiden Spalten hinter *Exp(B)* werden die unteren und oberen Grenzen der entsprechenden 95%-Konfidenzintervalle angegeben. Außerdem werden die Regressionskoeffizienten mittels Wald-Tests überprüft. Die jeweiligen p-Werte werden in der Spalte *Sig.* angegeben.

Bei den kategorialen Variablen ist es wichtig die gewählte Codierung zu beachten, um die Odds Ratios richtig zu interpretieren.

Die in der Tabelle aufgeführte *Konstante* hat keine Bedeutung für die angegebenen Odds Ratios.

Im Rahmen der so genannten Modellwahl trifft man die Entscheidung, welche Einflussvariablen in ein erstelltes Regressionsmodell aufgenommen werden. Es gibt viele verschiedene Möglichkeiten der Modellwahl. Die Auswahl der Einflussvariablen kann inhaltlich begründet sein oder auf univariablen logistischen Modellen beruhen. Ziel ist die Erstellung eines Modells, das die beobachteten Ausprägungen der Zielgröße möglichst gut beschreibt. Im Hauptdialogfenster der **logistischen Regression** kann im Feld *Methode* eine automatisierte Modellwahl angefordert werden. Häufig führt man eine Rückwärts-Selektion per Likelihood-Quotienten-Test (*Rückwärts: LR*, LR für Likelihood Ratio) durch:



Nach Bestätigen mit **OK** wird die angeforderte Variablenselektion durchgeführt.

Bei der Rückwärts-Selektion wird zunächst ein logistisches Regressionsmodell mit allen eingetragenen Kovariaten angepasst. Anschließend wird folgendes schrittweises Verfahren durchgeführt: In jedem

Schritt wird für jede Variable überprüft, ob sich die Modellanpassung relevant verschlechtert (z.B.  $p \leq 0,10$ ; dieser Wert kann unter **Optionen** geändert werden), wenn die jeweilige Variable aus dem Modell entfernt wird. Ist dies für keine Variable der Fall, so ist jede Variable offenbar wichtig und das schrittweise Verfahren stoppt. Andernfalls wird die Variable mit dem größten p-Wert aus dem Modell entfernt und es folgt der nächste Selektionsschritt.

Im Ausgabefenster wird die Modellbildung in mehreren Tabellen dargestellt. Die Tabelle *Variablen in der Gleichung* fasst die schrittweise angepassten Modelle, die Odds Ratios sowie die zugehörigen p-Werte und Konfidenzintervalle zusammen. Im letzten Schritt ist das endgültige Modell angegeben, das sich aus der Selektion ergibt. Die Tabelle *Modellieren, wenn Term entfernt* gibt die p-Werte des Likelihood-Quotienten-Tests an. Bei einer Rückwärts-Selektion wird anhand dieser p-Werte die Entscheidung getroffen, ob eine Variable in dem Model verbleibt oder entfernt wird.

		Variablen in der Gleichung					95% Konfidenzintervall für EXP (B)		
		Regressionskoeffizient B	Standardfehler	Wald	df	Sig.	Exp(B)	Unterer Wert	Oberer Wert
Schritt 1 <sup>a</sup>	Alter bei Diagnosestellung (in Jahren)	,027	,012	5,563	1	,018	1,028	1,005	1,051
	Geschlecht(1)	-,223	,284	,616	1	,433	,801	,459	1,396
	Körpergewicht in kg	-,001	,009	,018	1	,892	,999	,981	1,017
	Körpertemperatur in °C	,867	,116	55,668	1	,000	2,379	1,895	2,987
	Konstante	-36,618	4,657	61,830	1	,000	,000		
Schritt 2 <sup>a</sup>	Alter bei Diagnosestellung (in Jahren)	,027	,011	5,699	1	,017	1,028	1,005	1,051
	Geschlecht(1)	-,236	,265	,797	1	,372	,790	,470	1,326
	Körpertemperatur in °C	,867	,116	55,731	1	,000	2,380	1,895	2,988
	Konstante	-36,728	4,587	64,109	1	,000	,000		
Schritt 3 <sup>a</sup>	Alter bei Diagnosestellung (in Jahren)	,028	,011	5,924	1	,015	1,028	1,005	1,051
	Körpertemperatur in °C	,866	,116	55,749	1	,000	2,377	1,894	2,984
	Konstante	-36,840	4,581	64,679	1	,000	,000		

a. In Schritt 1 eingegebene Variablen: Alter bei Diagnosestellung (in Jahren), Geschlecht, Körpergewicht in kg, Körpertemperatur in °C.

		Log-Likelihood des Modells	Änderung der -2 Log-Likelihood	df	Signifikanz der Änderung
Schritt 1	Alter bei Diagnosestellung (in Jahren)	-210,346	6,168	1	,013
	Geschlecht	-207,570	,614	1	,433
	Körpergewicht in kg	-207,272	,018	1	,892
	Körpertemperatur in °C	-236,723	58,921	1	,000
Schritt 2	Alter bei Diagnosestellung (in Jahren)	-210,422	6,300	1	,012
	Geschlecht	-207,670	,797	1	,372
	Körpertemperatur in °C	-236,770	58,998	1	,000
Schritt 3	Alter bei Diagnosestellung (in Jahren)	-210,957	6,574	1	,010
	Körpertemperatur in °C	-237,152	58,964	1	,000

		Wert	df	Sig.	
Schritt 2 <sup>b</sup>	Variablen	Körpergewicht in kg	,018	1	,892
	Gesamtstatistik		,018	1	,892
Schritt 3 <sup>b</sup>	Variablen	Geschlecht(1)	,799	1	,371
		Körpergewicht in kg	,199	1	,655
	Gesamtstatistik		,818	2	,664

a. In Schritt 2 entfernte Variablen: Körpergewicht in kg.

b. In Schritt 3 entfernte Variablen: Geschlecht.

Grundsätzlich sollten in einer Regressionsanalyse sowohl inhaltliche Überlegungen als auch automatisierte Methoden der Modellwahl angewendet werden, um ein geeignetes Modell zur Beschreibung der Zielgröße zu entwickeln.

## 10. Time-to-event Analysen

### 10.1 Kaplan-Meier Methode und Log-Rank Test

Für den Vergleich der Überlebenszeiten zwischen zwei (oder mehr) Gruppen benötigt man drei Variablen: eine Zeitvariable, eine Statusvariable und eine Gruppenvariable. Die Zeitvariable gibt die Zeit von Diagnose bis zum Tod oder bis zum letzten Follow-up, falls der Patient nicht verstorben ist, an. Die Statusvariable gibt des Status des Patienten zum Zeitpunkt der letzten Information an. Als Gruppenvariable kann eine Variable gewählt werden, deren Einfluss auf die Überlebenszeit überprüft werden soll.

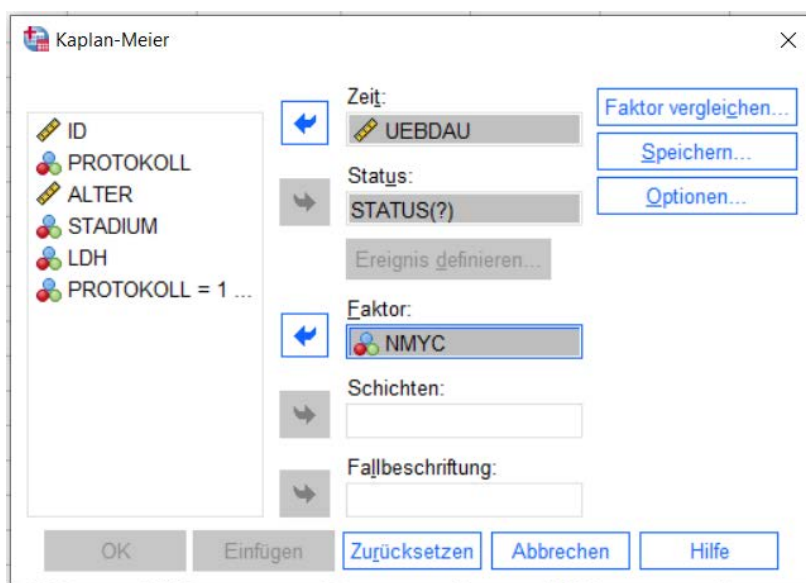
Exemplarisch werden folgende Variablen verwendet:

Zeitvariable: UEBDAU (Zeit von Diagnose bis zum Tod oder bis zum letzten Follow-up, falls der Patient nicht verstirbt)

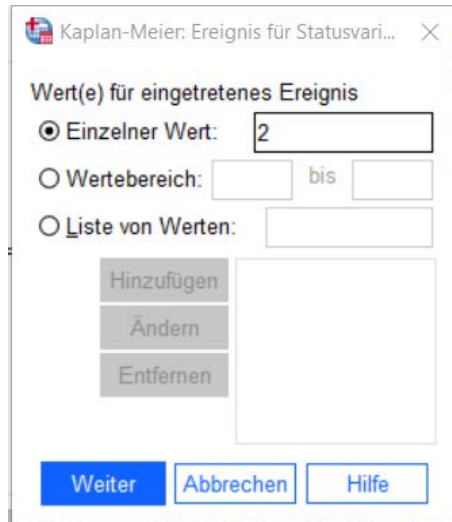
Statusvariable: STATUS  
 Lebt ohne Event / Loss to Follow-up = 1 = zensiert  
 Verstorben = 2 = Event  
 Lebt mit Rezidiv / Progress = 3 = zensiert  
 Lebt mit Zweitmalignom = 4 = zensiert

Faktor: NMYC (amplifiziert 1 = ja, 0 = nein)

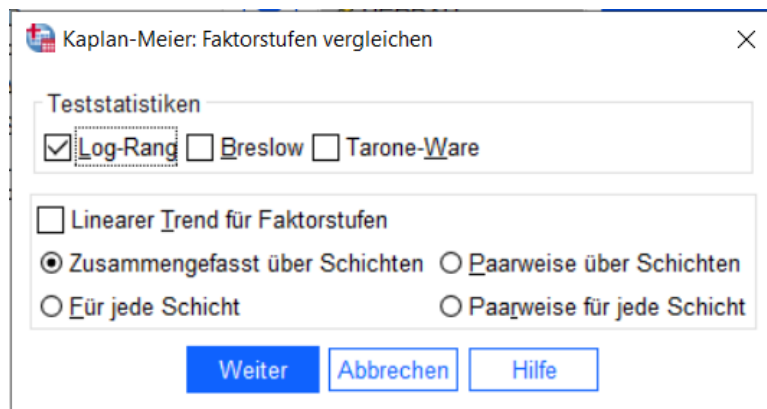
Über die Befehlsfolge **Analysieren > Überleben > Kaplan-Meier** gelangt man in das folgende Dialogfeld *Kaplan-Meier*, in dem die Zeitvariable in das Feld *Zeit*, die Statusvariable in das Feld *Status* und die Gruppenvariable in das Feld *Faktor* einzutragen sind.



Über die Schaltfläche **Ereignis definieren** muss die Ausprägung des interessierenden Ereignisses eingetragen werden, in diesem Falle die Ausprägung 1 für „Patient verstorben“

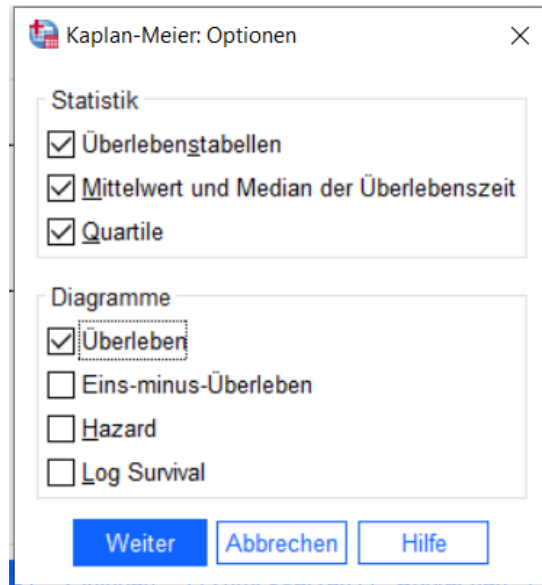


Ein Klick auf den Button **Faktor vergleichen** führt in das Dialogfeld *Kaplan-Meier: Faktorstufen vergleichen*, in dem ein Häkchen bei *Log-Rang* gesetzt wird.



Damit wird der Log-Rank-Test angefordert. Über **Weiter** kehren wir zurück ins übergeordnete Dialogfeld. Oft ist es von Interesse, den Kaplan-Meier-Schätzer graphisch darzustellen. Dazu klicken wir auf **Optionen**. In dem sich öffnenden Dialogfeld *Kaplan-Meier: Optionen* setzen wir folgende Häkchen:





Damit sind alle Einstellungen getroffen. Über **Weiter** geht es zurück zum übergeordneten Dialogfeld *Kaplan-Meier* und nach Bestätigung mit **OK** wird folgender Output ausgegeben.

Den p-Wert des Log-Rank-Tests entnimmt man der Tabelle „Gesamtvergleiche“ in der Spalte *Sig.*

**Kaplan-Meier**

**Zusammenfassung der Fallverarbeitung**

NMYC	Gesamtzahl	Anzahl von Ereignissen	Zensiert	
			N	Prozent
nicht NMYC amplifiziert	163	117	46	28,2%
NMYC amplifiziert	123	89	34	27,6%
<b>Gesamt</b>	<b>286</b>	<b>206</b>	<b>80</b>	<b>28,0%</b>

**Überlebenstabelle**

NMYC	Zeit	Status	Kumulierter Anteil Überlebender zum Zeitpunkt		Anzahl der kumulativen Ereignisse	Anzahl der verbliebenen Fälle	
			Schätzer	Standardfehler			
nicht NMYC amplifiziert	1	,016	verstorben	,994	,006	1	162
	2	,025	verstorben	,988	,009	2	161
	3	,030	verstorben	,982	,011	3	160
	4	,036	verstorben	,975	,012	4	159
	5	,088	verstorben	,969	,014	5	158
	6	,142	verstorben	,963	,015	6	157
	7	,244	verstorben	,957	,016	7	156
	8	,307	verstorben	,951	,017	8	155
	9	,309	verstorben	,945	,018	9	154
	10	,381	verstorben	,939	,019	10	153
	11	,386	verstorben	,933	,020	11	152
	12	,405	verstorben	,926	,020	12	151
	13	,435	verstorben	,920	,021	13	150
	14	,441	verstorben	,914	,022	14	149
	15	,441	verstorben	,908	,022	15	148



**Perzentile**

NMYC	25,0%		50,0%		75,0%	
	Schätzer	Standardfehler r	Schätzer	Standardfehler r	Schätzer	Standardfehler r
nicht NMYC amplifiziert	15,458	.	3,332	,452	1,440	,194
NMYC amplifiziert			1,667	,278	,652	,090
Gesamt	15,458	.	2,631	,273	1,043	,114

**Gesamtvergleiche**

	Chi-Quadrat	df	Sig.
Log Rank (Mantel-Cox)	3,466	1	,063

Test auf Gleichheit der Überlebensverteilungen für die verschiedenen Stufen von NMYC.

Erläuterungen:

**Tabelle Überlebenstabelle:**

Die Tabelle zeigt zu verschiedenen Zeitpunkten (Spalte ‚Zeit‘) eine Auflistung der geschätzten Überlebenswahrscheinlichkeiten (Spalte ‚Schätzer‘). In der Spalte ‚Standardfehler‘ ist der Standardfehler der Schätzung der Überlebenswahrscheinlichkeit angegeben. Die Auflistung erfolgt ggf. nach Gruppen getrennt.

**Tabelle Perzentile:**

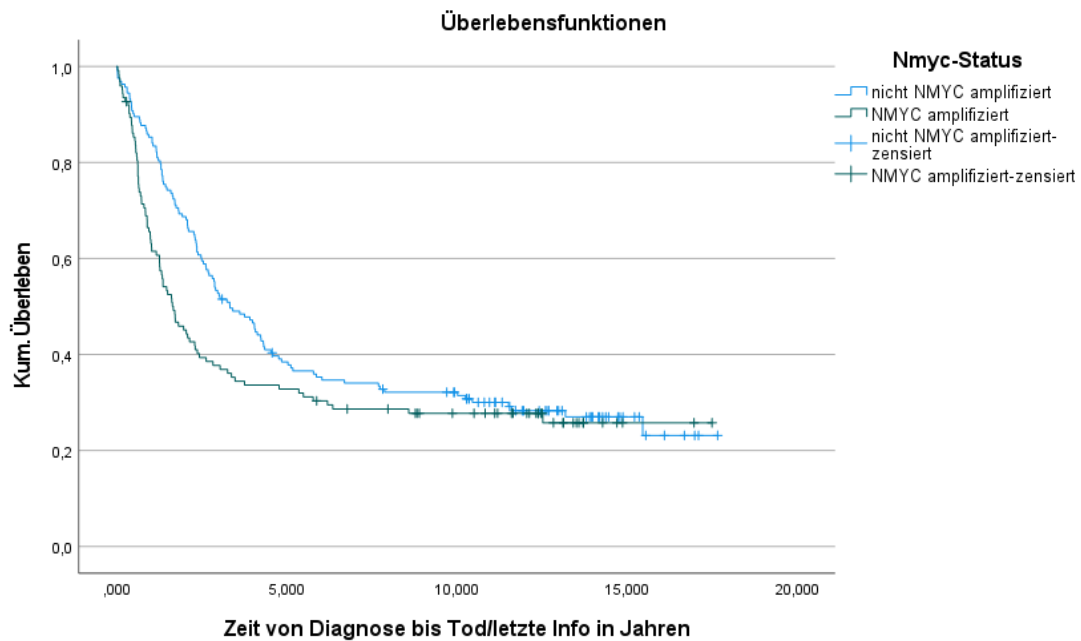
In dieser Tabelle sind die Zeitpunkte angegeben, zu denen die geschätzte Überlebenswahrscheinlichkeit 25%, 50% bzw. 75% beträgt.

**Tabelle Gesamtvergleiche:**

In dieser Tabelle stehen die Ergebnisse des Log-Rank-Tests in der Zeile „Log Rank (Mantel-Cox)“. In der Spalte *Chi-Quadrat* ist der Wert der Teststatistik angegeben und in der Spalte *Sig.* der p-Wert des zweiseitigen Log-Rank-Tests. SPSS gibt den p-Wert auf die dritte Nachkommastelle gerundet an.

Hinweis: Ein möglicher Eintrag der Form „Sig. = ,000“ ist so zu lesen, dass der p-Wert  $p < 0,001$  ist.

Wegen ihrer guten Anschaulichkeit ist die graphische Darstellung der Überlebensfunktionen nach Kaplan-Meier oft von besonderem Interesse. Durch **Doppelklick auf die Graphik im Ausgabefenster** kann die Abbildung bei Bedarf nachbearbeitet werden. Die Linienfarben, die Achsenbeschriftungen und die Überschrift können durch nochmaligen Doppelklick auf das zu editierende Objekt verändert werden. Darüber hinaus können Textfelder eingefügt sowie die Legende verschoben und editiert werden. Per Default sieht die Graphik wie folgt aus.

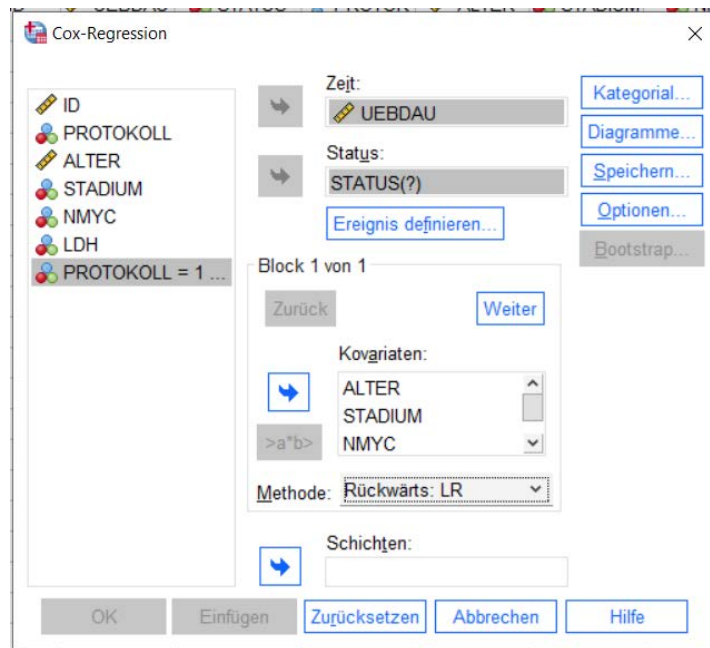


### 10.2 Cox-Regression

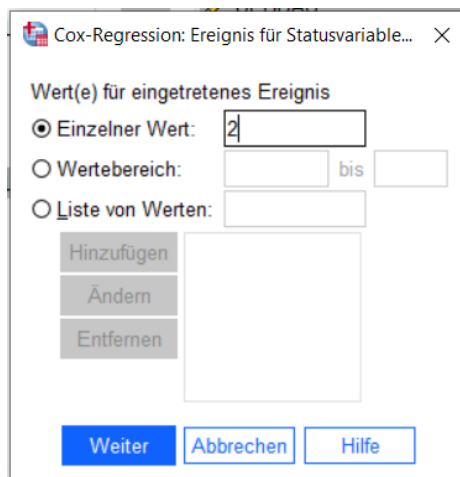
Bei der Kaplan-Meier-Methode kann nur ein einzelner Einflussfaktor berücksichtigt und so die Überlebenszeiten zwischen zwei (oder mehr) Gruppen verglichen werden. Mit einer Cox-Regression können gleichzeitig mehrere so genannte Kovariablen (z.B. Geschlecht, Alter, usw.) in der Analyse berücksichtigt werden.

Für eine Cox-Regression müssen i.d.R. folgende Variablen spezifiziert werden: eine Zeitvariable, eine Statusvariable und Kovariablen, die sowohl stetig als auch kategorial sein dürfen.

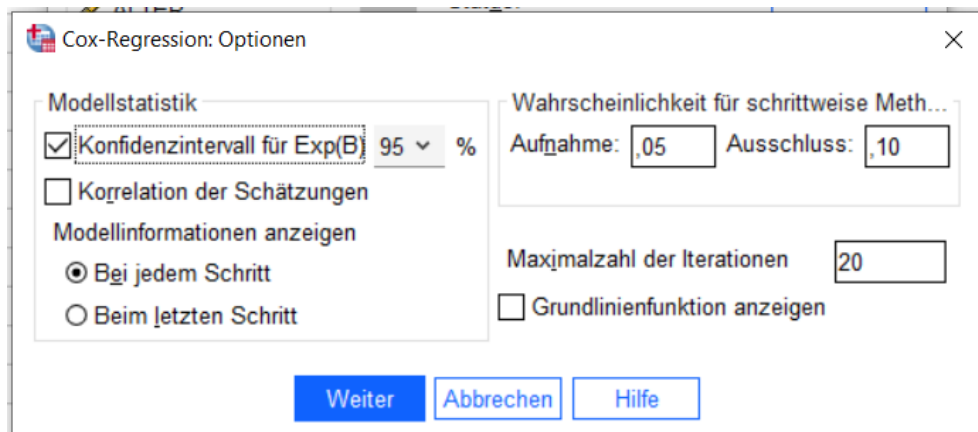
Mit der Befehlsfolge **Analysieren > Überlebensanalyse > Cox-Regression** gelangt man in das Dialogfeld *Cox-Regression*, in dem die Zeitvariable in das Feld *Zeit*, die Statusvariable in das Feld *Status* und sämtliche Kovariablen in das Feld *Kovariaten* eingetragen werden. Unter *Methode* kann in der Scroll-Leiste die Methode der Variablenselektion ausgewählt werden. Sollen keine Selektion erfolgen, wird als Methode **Einschluss** gewählt. Im Beispiel wählen wir die schrittweise Rückwärts-Selektion per Likelihood-Quotienten-Test **Rückwärts: LR** (vgl. Anleitung zur logistischen Regression).



Unter der Schaltfläche **Ereignis definieren** muss die Ausprägung des interessierenden Ereignisses in der Statusvariable eingetragen werden, in diesem Falle die Ausprägung 2 für „Patient ist verstorben“.



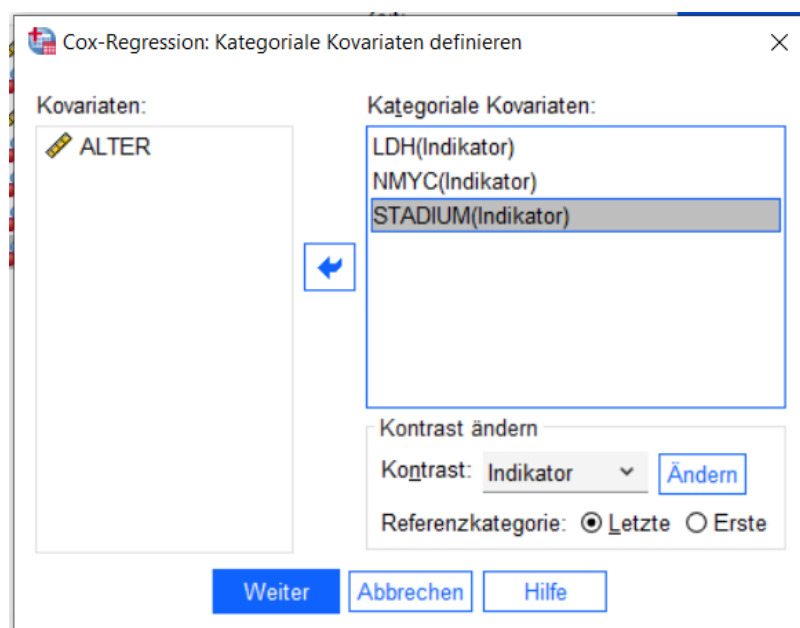
Unter **Optionen** ist durch Anklicken das *95%-Konfidenzintervall für Exp(B)* anzufordern. Der Ausdruck *Exp(B)* steht für das Hazard Ratio, das eine der zentralen Größen in einer Cox-Regression darstellt.



Über **Weiter** kommt man zurück zum übergeordneten Dialogfeld *Cox-Regression*. Über die Schaltfläche **Kategorial** werden die kategorialen Variablen gekennzeichnet. Alle kategorialen Variablen werden markiert und über den Pfeil in das Feld *Kategoriale Kovariaten* überführt. Der Kontrast sollte als *Indikator* verbleiben. Anschließend muss für jede kategoriale Variable die Referenzkategorie festgelegt werden, auf die die berechneten Hazard Ratios bezogen werden. Per Default steht die *Referenzkategorie* auf *Letzte* (vgl. logistische Regression).

Um die Referenzkategorie zu ändern, setzt man den Punkt in das gewünschte Feld und klickt **Ändern**. Eine Definition der *Referenzkategorie*, die weder der kleinsten noch der größten von mehreren numerischen Ausprägungen entspricht, ist über die Syntax möglich.

Nachdem alle Einstellungen getroffen sind, sieht das Dialogfeld wie folgt aus:



Über **Weiter** gelangt man zurück zum übergeordneten Dialogfeld *Cox-Regression*. Nach Bestätigung mit **OK** wird folgender Output ausgegeben. Die Hazard Ratios stehen in der Tabelle *Variablen in der Gleichung* in der Spalte *Exp(B)*. Den zugehörigen p-Wert des Hazard Ratios entnimmt man der Spalte *Signifikanz*. Das 95%-Konfidenzintervall des jeweiligen Hazard Ratios  $\text{Exp}(B)$  ist in der Spalte *95,0%-Konfidenzintervall für Exp(B)* angegeben. Zur Kontrolle empfiehlt es sich in der Tabelle *Codierung für kategoriale Variablen* zu überprüfen, dass alle Referenzkategorien der kategorialen Variablen korrekt definiert wurden.

Vorsicht: Bitte beachten Sie, dass die Bezeichnung der Variablen in der Tabelle *Variablen in der Gleichung* u.U. von der ursprünglichen Kodierung abweicht bzw. nicht intuitiv ist. Die verwendete Codierung ist der Tabelle *Codierung für kategoriale Variablen* zu entnehmen.

**Cox-Regression**

**Auswertung der Fallverarbeitung**

		N	Prozent
Für Analyse verfügbare Fälle	Ereignis <sup>a</sup>	202	51,8%
	Zensiert	77	19,7%
	Insgesamt	279	71,5%
Nicht verwendete Fälle	Fälle mit fehlenden Werten	111	28,5%
	Fälle mit negativer Zeit	0	0,0%
	Zensierte Fälle vor dem frühesten Ereignis in einer Schicht	0	0,0%
	Insgesamt	111	28,5%
Insgesamt		390	100,0%

a. Abhängige Variable: UEBCAU

**Codierungen für kategoriale Variablen<sup>a,c,d</sup>**

		Häufigkeit	(1)	(2)	(3)	(4)
STADIUM <sup>b</sup>	1=Stadium 1	3	1	0	0	0
	2=Stadium 4S	8	0	1	0	0
	3=Stadium 2, 2a oder 2b	6	0	0	1	0
	4=Stadium 3	21	0	0	0	1
	5=Stadium 4	241	0	0	0	0
NMYC <sup>b</sup>	1=nicht NMYC amplifiziert	156	1			
	2=NMYC amplifiziert	123	0			
LDH <sup>b</sup>	1=normal	24	1			
	2=pathologisch	255	0			

a. Kategorie-Variablen: STADIUM

b. Kodierung für Indikatorparameter

c. Kategorie-Variablen: NMYC

d. Kategorie-Variablen: LDH

**Variablen in der Gleichung**

		B	SE	Wald	df	Signifikanz	Exp(B)	95,0% Konfidenzinterv. für Exp(B)	
								Untere	Obere
Schritt 1	ALTER	,031	,024	1,678	1	,195	1,031	,984	1,081
	STADIUM			11,669	4	,020			
	STADIUM(1)	-1,747	1,009	2,998	1	,083	,174	,024	1,259
	STADIUM(2)	,168	,446	,141	1	,707	1,183	,493	2,835
	STADIUM(3)	-1,301	,719	3,276	1	,070	,272	,067	1,114
	STADIUM(4)	-,784	,326	5,791	1	,016	,457	,241	,865
	NMYC	-,474	,158	9,044	1	,003	,623	,457	,848
	LDH	-,720	,309	5,430	1	,020	,487	,265	,892
Schritt 2	STADIUM			11,510	4	,021			
	STADIUM(1)	-1,807	1,008	3,213	1	,073	,164	,023	1,184
	STADIUM(2)	,083	,441	,035	1	,852	1,086	,458	2,576
	STADIUM(3)	-1,306	,719	3,301	1	,069	,271	,066	1,108
	STADIUM(4)	-,775	,326	5,668	1	,017	,461	,243	,872
	NMYC	-,440	,155	8,065	1	,005	,644	,475	,873
	LDH	-,727	,309	5,527	1	,019	,484	,264	,886

**Modell bei entferntem Term**

Term entfernt		Verlust-Chi-Quadrat	df	Signifikanz
Schritt 1	ALTER	1,574	1	,210
	STADIUM	15,609	4	,004
	NMYC	8,707	1	,003
	LDH	6,548	1	,011
Schritt 2	STADIUM	15,650	4	,004
	NMYC	7,759	1	,005
	LDH	6,679	1	,010

**Variablen nicht in der Gleichung<sup>a</sup>**

		Score	df	Signifikanz
Schritt 2	ALTER	1,680	1	,195

a. Chi-Quadrat-Residuen= 1,680 bei 1 df Sig. = ,195

Die wesentlichen Angaben in allen Tabellen sind diejenigen nach dem letzten Schritt der Variablenselektion.

In der Tabelle *Variablen in der Gleichung* sind die geschätzten Hazard Ratios in der Spalte *Exp(B)* angegeben. Die Bezeichnung beruht darauf, dass die Hazard Ratios die exponierten Regressionskoeffizienten (Spalte *B*) darstellen. In den Spalten *95% Konfidenzintervall für Exp(B)* findet sich die untere und obere Schranke des Konfidenzintervalls des jeweiligen Hazard Ratios. Außerdem werden die Hazard Ratios mittels Wald-Tests überprüft, d.h. es wird geprüft, ob das jeweilige Hazard Ratio in der Grundgesamtheit ungleich 1 ist. Die jeweiligen p-Werte sind in der Spalte *Signifikanz* angegeben.

In der Tabelle *Modell bei entferntem Term* sind in der Spalte *Signifikanz* die p-Werte eines Likelihood-Ratio-Tests angegeben, in dem geprüft wird, ob die jeweilige Kovariable einen Einfluss auf die Überlebenszeit hat.

In der Tabelle *Variablen nicht in der Gleichung* sind die p-Werte von Score-Tests derjenigen Kovariablen angegeben, die im Rahmen der Variablenselektion aus dem Modell ausgeschlossen wurden.

Hinweise:

1. Für die im endgültigen Model eingeschlossenen Variablen können die Ergebnisse der Wald-Tests in Tabelle *Variablen in der Gleichung* und der entsprechenden Likelihood-Ratio-Tests in Tabelle *Modell bei entferntem Term* voneinander abweichen. In diesem Fall wird in der Regel der Likelihood-Ratio-Test bevorzugt.
2. Für die ausgeschlossenen Kovariablen sind die relevanten p-Werte in der Tabelle *Variablen nicht in der Gleichung* angegeben, nicht in der Tabelle *Modell bei entferntem Term*.