

Analyse von Überlebenszeiten mit SPSS

Die für diese Anleitung verwendeten Daten stammen von pädiatrischen Hochrisiko-Patienten mit einem Neuroblastom (NB) aus den Studien NB90 und NB97 (verwendeter Datensatz: NB90_NB97.sav). Dabei betrachten wir für diese Anleitung nur das Subkollektiv jener Patienten, die gemäß dem Protokoll der NB90-Studie behandelt wurden.

Auswahl der Fälle der Studie NB90: Daten > Fälle auswählen > Falls Bedingung zutrifft > Falls > Ins Bedingungsfeld ,PROTOKOLL = 1' (entspricht „NB90“) eintragen > Weiter > OK

Alternativ für die Auswahl der Fälle folgende Syntax ausführen:

```
COMPUTE filter_$=(PROTOKOLL = 1).  
VARIABLE LABELS filter_$ 'PROTOKOLL = 1 (FILTER)'.  
VALUE LABELS filter_$ 0 'Not Selected' 1 'Selected'.  
FORMATS filter_$ (f1.0).  
FILTER BY filter_$.  
EXECUTE.
```

1. Kaplan-Meier-Methode und Log-Rank-Test

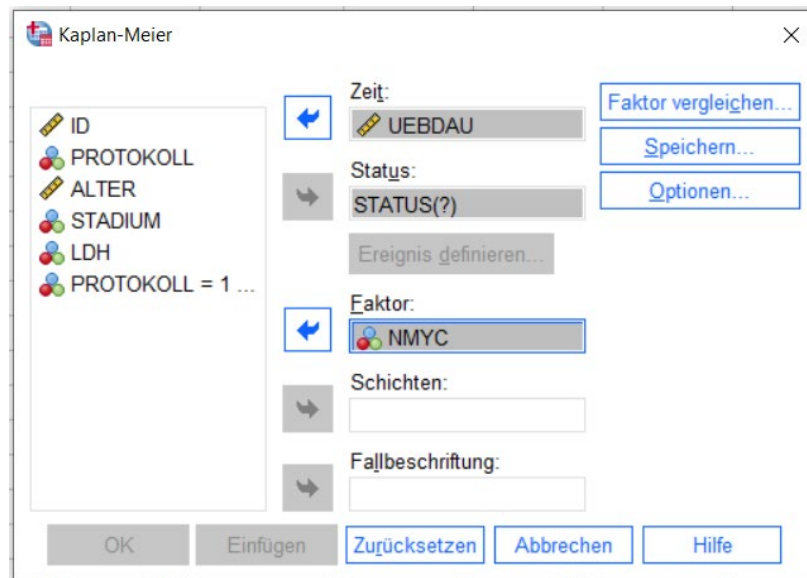
Für den Vergleich der Überlebenszeiten zwischen zwei (oder mehr) Gruppen benötigt man drei Variablen: eine Zeitvariable, eine Statusvariable und eine Gruppenvariable. Es soll untersucht werden, ob der NMYC-Status prognostisch für das Gesamtüberleben ist.

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

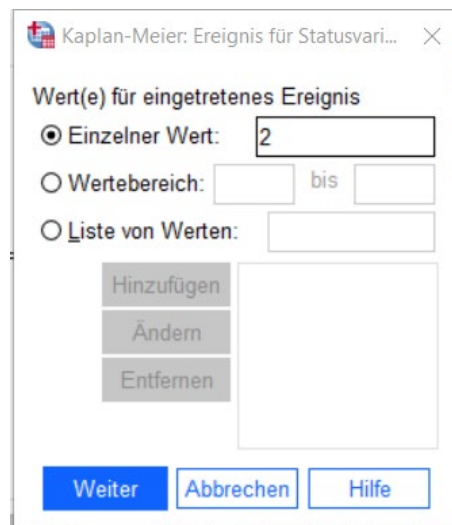
Statusvariable: STATUS
Lebt ohne Event / Loss to Follow-up = 1 = zensiert
Verstorben = 2 = Event
Lebt mit Rezidiv / Progress = 3 = zensiert
Lebt mit Zweitmalignom = 4 = zensiert

Faktor: NMYC (1 = nicht amplifiziert, 2= amplifiziert)

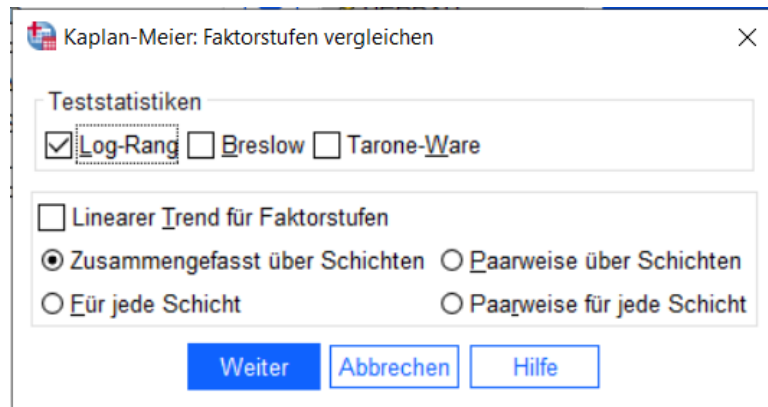
Über die Befehlsfolge **Analysieren > Überleben > Kaplan-Meier** gelangt man in das folgende Dialogfeld *Kaplan-Meier*, in dem die Zeitvariable UEBDAU in das Feld *Zeit*, die Statusvariable STATUS in das Feld *Status* und die Gruppenvariable NMYC in das Feld *Faktor* einzutragen sind.



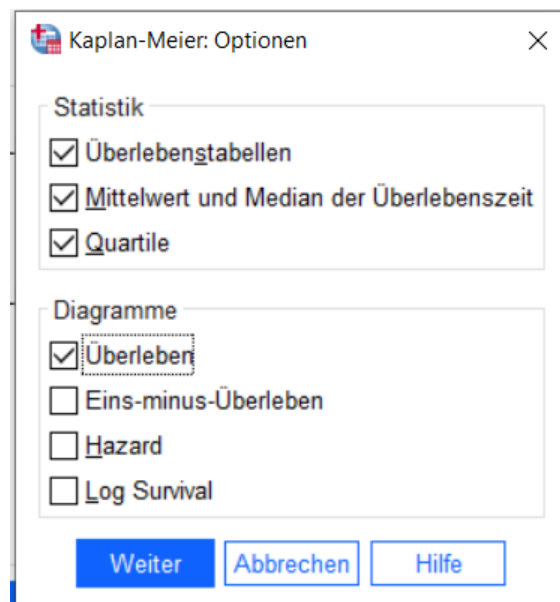
Über die Schaltfläche **Ereignis definieren** muss die Ausprägung des interessierenden Ereignisses eingetragen werden, in diesem Falle die Ausprägung 2 für „Patient verstorben“. Ist der Knopf ausgegraut, muss noch einmal in das Status-Fenster geklickt werden. Die Eingabe wird mit **Weiter** bestätigt.



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%
Gesamt	286	206	80	28,0%

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	,446	verstorben	,908	,023	15	148

Perzentile

NMYC	Schätzer	25,0%	Schätzer	50,0%	Schätzer	75,0%
		Standardfehler		Standardfehler		Standardfehler
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 Überlebensstabelle:**

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:

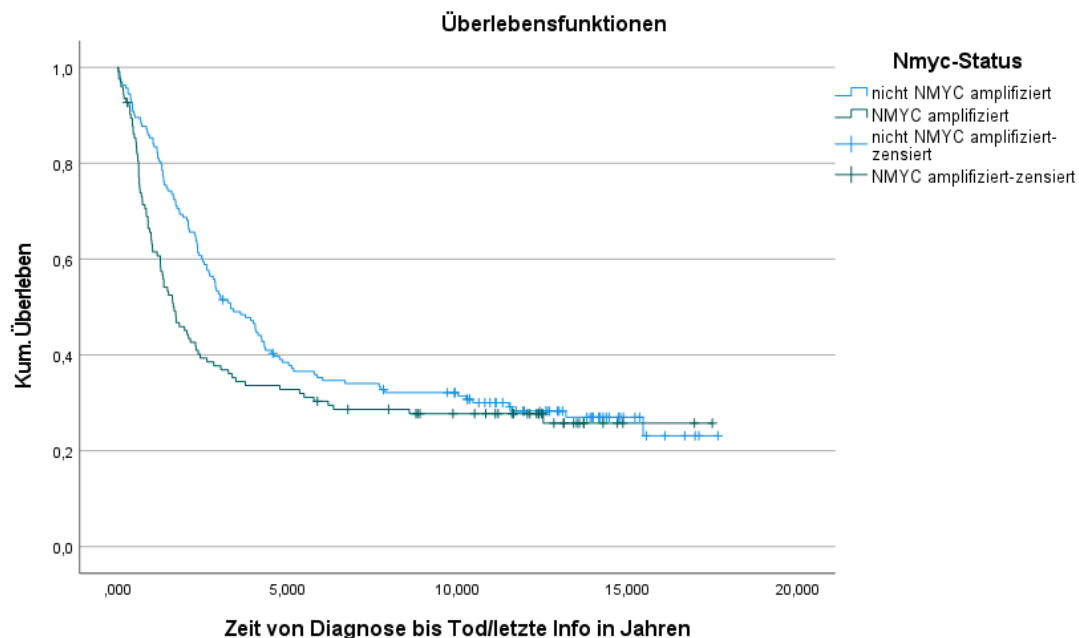
In dieser Tabelle sind die Zeitpunkte angegeben, zu denen die geschätzte Überlebenswahrscheinlichkeit 25%, 50% bzw. 75% beträgt.

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.



2. Cox-Regression (Regression für rechts-zensierte Beobachtungen)

Im Kapitel über die Kaplan-Meier-Methode wurde nur ein einzelner Einflussfaktor (NMYC-Status) berücksichtigt, um die Überlebenszeiten zwischen zwei (oder mehr) Gruppen zu vergleichen. Mit einer Cox-Regression können gleichzeitig mehrere so genannte Kovariablen (z.B. Geschlecht, Alter, usw.) in der Analyse berücksichtigt werden.

An dem oben verwendeten Datensatz soll im Subkollektiv der gemäß NB90-Protokoll behandelten Patienten untersucht werden, ob der NMYC-Status und der LDH-Wert prognostisch für das Gesamtüberleben der Patienten sind. Zusätzlich sollen das Stadium und das Alter bei Diagnose als Kovariablen berücksichtigt werden.

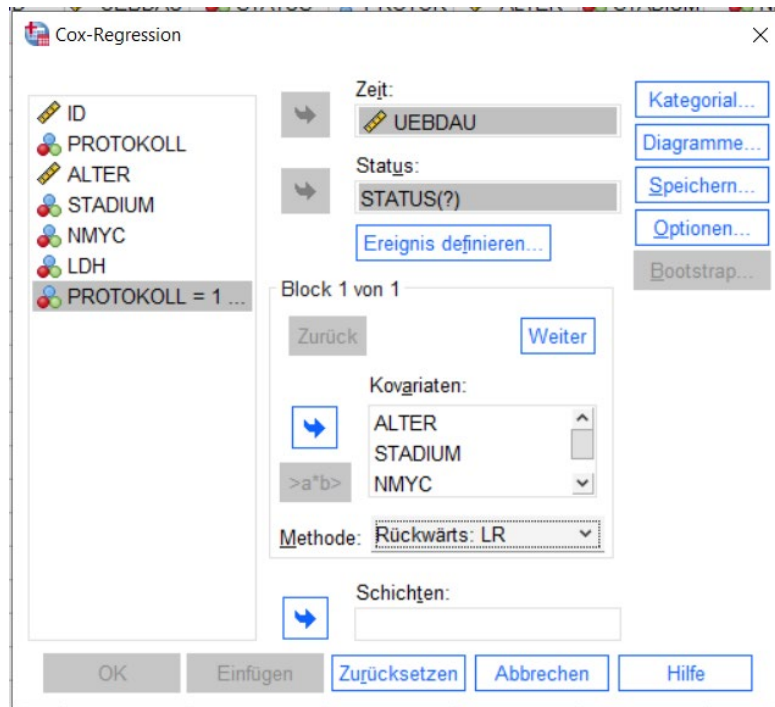
Für eine Cox-Regression müssen i.d.R. folgende Variablen spezifiziert werden: eine Zeitvariable, eine Statusvariable und Kovariablen, die sowohl stetig als auch kategorial sein dürfen. Im vorliegenden Beispiel sind dies die folgenden Variablen:

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

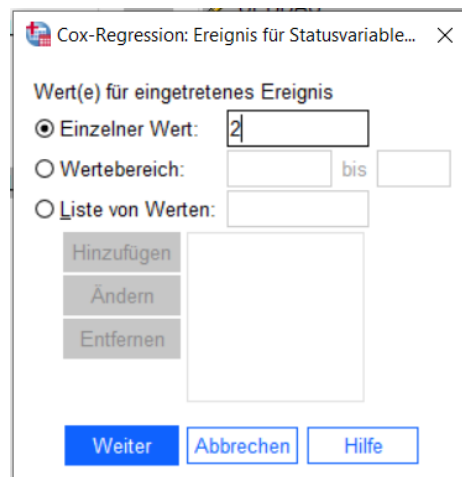
Statusvariable: STATUS
Lebt ohne Event / Loss to Follow-up = 1 = zensiert
Verstorben = 2 = Event
Lebt mit Rezidiv / Progress = 3 = zensiert
Lebt mit Zweitmalignom = 4 = zensiert

Kovariablen: PAT_ALTER (stetig), NMYC (kategorial), LDH (kategorial), STADIUM (kategorial)

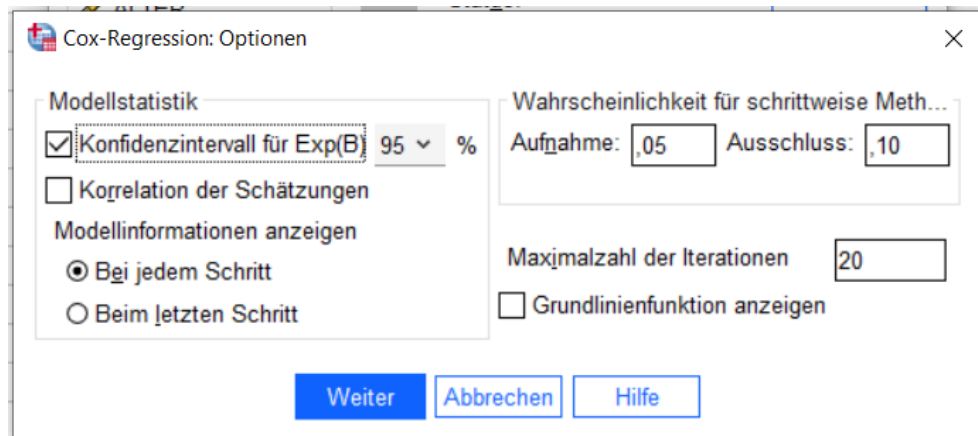
Mit der Befehlsfolge **Analysieren > Überleben > Cox-Regression** gelangt man in das Dialogfeld *Cox-Regression*, in dem die Zeitvariable UEBDAU in das Feld *Zeit*, die Statusvariable STATUS in das Feld *Status* und sämtliche Kovariablen in das Feld *Kovariaten* eingetragen werden. Unter *Methode* kann 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“. Über Weiter kommt man zurück zum übergeordneten Dialogfeld Cox-Regression.



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. Im vorliegenden Beispiel sind dies die Kovariaten LDH, NMYC, STADIUM. 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 welche die berechneten Hazard Ratios bezogen werden. Per Default steht die *Referenzkategorie* auf *Letzte* (vgl. logistische Regression).

Im Beispiel sind die Variablen wie folgt kodiert. Als Referenzkategorie soll die kursiv gesetzte Kategorie gewählt werden.

NMYC:

Nicht NMYC-amplifiziert = 1
NMYC-amplifiziert = 2

LDH:

normal = 1
pathologisch = 2

STADIUM

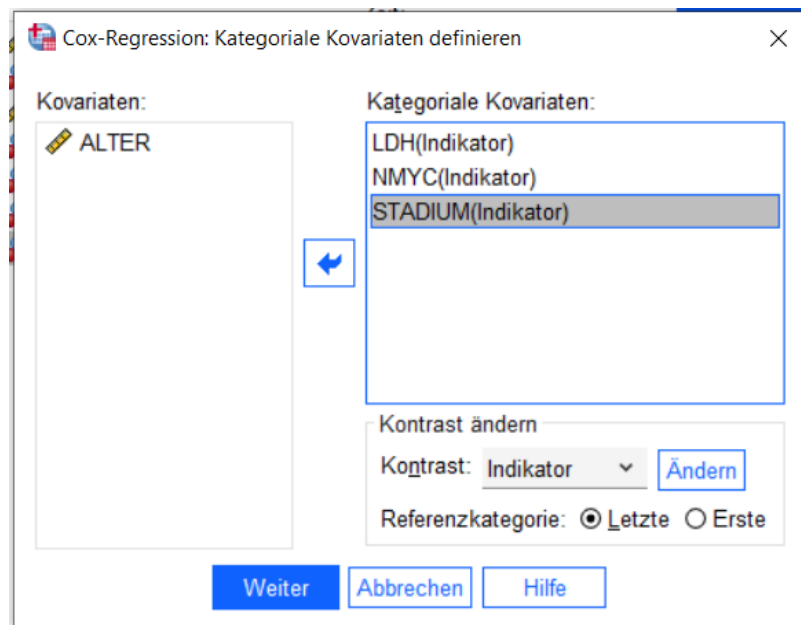
Stadium 1 = 1
 Stadium 2, 2a oder 2b = 3
 Stadium 3 = 4
Stadium 4 = 5
 Stadium 4S = 2

Die Wahl der *letzten* Kategorie als Referenzkategorie für NMYC bedeutet, dass die Ausprägung 2 (NMYC-amplifiziert) die Referenzkategorie darstellt. Würde man die *erste* Kategorie als Referenzkategorie wählen, so wäre die Ausprägung 1 (Nicht NMYC-amplifiziert) die

Referenzkategorie. Gleiches gilt für die Variable STADIUM mit mehr als zwei Ausprägungen. Die Variable STADIUM hat 5 Ausprägungen. *Referenzkategorie: Letzte* bedeutet, dass Stadium 4 die Referenzkategorie darstellt, da Stadium 4 mit der größten der Zahl (=5) codiert ist. Würde man *Referenzkategorie: Erste* wählen, so wäre Stadium 1 die Referenzkategorie, da sie mit der kleinsten Zahl (=1) codiert ist.

Um die Referenzkategorie zu ändern, setzt man den Punkt in das gewünschte Feld und klickt **Ändern**. In unserem Beispiel sind aber alle Referenzkategorien auf *Letzte* zu belassen. 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 $Exp(B)$ 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, ob 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. Im vorliegenden Fall wird die z.B. Ausprägung „Stadium 2, 2a oder 2b“ in der Ausgabe als „STADIUM(3)“ bezeichnet. Dies kann an der Tabelle „Codierung für kategoriale Variablen“ abgelesen werden. In der Tabelle „Codierung für kategoriale Variablen“ steht in der Zeile „3 = Stadium 2, 2a oder 2b“ eine 1 in der Spalte (3). In der Zeile „5 = Stadium 4“ stehen ausschließlich Nullen, womit diese als Referenzkategorie gekennzeichnet ist.

Cox-Regression**Auswertung der Fallverarbeitung**

		N	Prozent
Für Analyse verfügbare Fälle	Ereignis ^a	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: UEBDAU

Codierungen für kategoriale Variablen^{a,c,d}

		Häufigkeit	(1)	(2)	(3)	(4)
STADIUM ^b	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 ^b	1=nicht NMYC amplifiziert	156	1			
	2=NMYC amplifiziert	123	0			
LDH ^b	1=normal	24	1			
	2=pathologisch	255	0			

a. Kategorie-Variable: STADIUM

b. Kodierung für Indikatorparameter

c. Kategorie-Variable: NMYC

d. Kategorie-Variable: LDH

Variablen in der Gleichung							95,0% Konfidenzinterv. für Exp(B)		
		B	SE	Wald	df	Signifikanz	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^a

		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. Die Variablenselektion bricht nach dem 2. Schritt ab. Das endgültige Modell enthält die Kovariablen NMYC, LDH und STADIUM.

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.

Interpretation (exemplarisch für die Zeile „*Stadium(4)*“): Die Hazard-Rate bei Stadium 3 ist um den (geschätzten) Faktor 0,461 niedriger als bei Stadium 4 (Referenzkategorie).

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. Der p-Wert der Kovariable ALTER beträgt $p = 0,195$.

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 (hier ALTER) sind die relevanten p-Werte in der Tabelle „Variablen nicht in der Gleichung“ angegeben, nicht in der Tabelle „Modell bei entferntem Term“.