

Überlebenszeitanalyse

brunner & hess software ag

Lorenz Gygax & Christian Schmidhauser

Version 1.0, März 2002

1 Einführung

Eine weitere Klasse von Analysemethoden. Was ist an ihr anders und besonders?

1. Es geht um die Analyse von Zeiträumen bis zum Eintreten eines gewissen Ereignisses. Solche Zeiten sind oft sehr schief verteilt, d. h. es gibt einige wenige Beobachtungen (Patienten) mit ausserordentlich langen Zeiten. Wegen dieser schiefen Verteilung können wir nicht direkt mit unseren klassischen Methoden arbeiten. Wenn wir Beobachtungen aller Patienten hätten, dann könnte man allenfalls noch mit den logarithmus transformierten Daten und den klassischen Regressionsverfahren arbeiten.
2. Was ist mit denjenigen Patienten zu tun, bei denen das Ereignis bei Ablauf der Studie noch nicht eingetreten ist oder die man im Verlaufe der Studie aus den Augen verliert? Man spricht hier von zensurierten Daten. In den klassischen Methoden haben wir keine Möglichkeit diese zu berücksichtigen (wir brauchen Werte aller Variablen für alle Beobachtungen). Die Überlebenszeitanalyse bietet uns auch hier eine Strategie wie mit solchen zensurierten Daten umzugehen ist. Allerdings muss die Annahme erfüllt sein, dass die Zensurierung keinen Zusammenhang mit der Fragestellung hat, d. h. alle Patienten haben die gleiche Wahrscheinlichkeit zensuriert zu werden.

Die Fragestellungen, welche wir mit diesen Methoden angehen können sind divers. Ursprünglich ging es um die Analyse von Zeiträumen bis zum Tod (daher der Name). Andere mögliche Analysegrößen sind die Zeit bis zum Eintreffen eines medizinischen Ereignisses wie das Auftreten eines Symptomes, das Verschwinden eines Symptomes, das Erreichen eines Status' (z. B. Responder) und viele mehr.

2 Kaplan-Meier Verfahren

2.1 Schätzung und deskriptive Statistik

Das grundsätzliche Vorgehen der Kaplan-Meier Methode ist es, zu jedem Zeitpunkt eines Ereignisses die Wahrscheinlichkeit desselbigen zu schätzen. Nehmen wir an, wir betrachten Patienten und wollen wissen, wie lange es dauert, bis eine bestimmte Krebsart nach der Behandlung wieder ausbricht. Wir können die kumulative Überlebenswahrscheinlichkeit wie folgt berechnen:

Zeit	Status	Gesund vorher	Gesund	Anteil gesund	kumulative Überl'wahrsch.
2	Ereignis	10	9	9/10	9/10 = 0.90
5	Ereignis	9	8	8/9	0.90 · (8/9) = 0.80
12	zensuriert	8	7		
35	Ereignis	7	6	6/7	0.80 · (6/7) = 0.69
40	zensuriert	6	5		
43	Ereignis	5	4	4/5	0.69 · (4/5) = 0.55
49	Ereignis	4	3	3/4	0.55 · (3/4) = 0.41
64	Ereignis	3	2	2/3	0.41 · (2/3) = 0.27
69	zensuriert	2	1		
72	Ereignis	1	0	0/1	0.27 · (0/1) = 0.00

Die kumulative Überlebenswahrscheinlichkeit schätzt den Anteil aller Patienten, welche zu dem gegebenen Zeitpunkt noch gesund sind. Es erfolgt keine Schätzung zu den zensurierten Zeiten, da sich hier der Anteil der gesunden nicht ändert, sondern nur die Anzahl der Patienten unter Beobachtung.

Der Output aus SPSS zu diesen Daten sieht sehr ähnlich aus, wie unsere Tabelle von oben. Es kommen zusätzlich einige weitere Schätzgrößen dazu:

Survival Analysis for Zeit

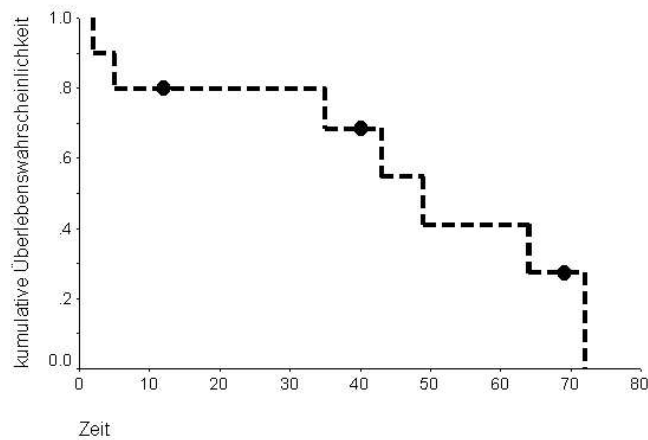
Time	Status	Cumulative Survival	Standard Error	Cumulative Events	Number Remaining
2.00	1.00	.9000	.0949	1	9
5.00	1.00	.8000	.1265	2	8
12.00	.00			2	7
35.00	1.00	.6857	.1515	3	6
40.00	.00			3	5
43.00	1.00	.5486	.1724	4	4
49.00	1.00	.4114	.1756	5	3
64.00	1.00	.2743	.1620	6	2
69.00	.00			6	1
72.00	1.00	.0000	.0000	7	0

Number of Cases: 10 Censored: 3 (30.00%) Events: 7

	Survival Time	Standard Error	95% Confidence Interval
Mean:	45.84	8.73	(28.74, 62.95)
Median:	49.00	8.96	(31.43, 66.57)

Die Berechnung des Mittelwertes erfolgt unter Berücksichtigung der zensurierten Daten. Wenn die längste beobachtete Zeit zensuriert ist, so spricht man davon, dass die mittlere Überlebenszeit auf den längsten beobachteten Wert eines Ereignisses beschränkt ist. Der Median ist derjenige Zeitpunkt, an dem die kumulative Überlebenswahrscheinlichkeit unter 50% sinkt.

Die kumulative Überlebenswahrscheinlichkeit wird oft auch noch graphisch dargestellt (die Stufen liegen bei den Zeitpunkten der Ereignisse, die Punkte stellen die zensurierten Patienten dar):



2.2 Statistische Tests

Mit der Kaplan-Meier Methode ist es nun auch noch möglich kumulative Überlebenswahrscheinlichkeiten verschiedener (Behandlungs-)gruppen zu vergleichen.

Nehmen wir an, die erhobenen Daten stammen aus zwei Gruppen von Patienten, solchen die mit Placebo und solchen welche mit einem wirksamen Stoff behandelt wurden. Der Output aus SPSS schaut dann wie folgt aus (analog wie bisher, aber für beide Gruppen getrennt):

Survival Analysis for Zeit

Factor Behandlung = Placebo

Time	Status	Cumulative Survival	Standard Error	Cumulative Events	Number Remaining
2.00	1.00	.8000	.1789	1	4
12.00	.00			1	3
35.00	1.00	.5333	.2483	2	2
43.00	1.00	.2667	.2257	3	1
49.00	1.00	.0000	.0000	4	0

	Survival Time	Standard Error	95% Confidence Interval	
Mean:	34.27	8.87	(16.87,	51.66)
Median:	43.00	17.35	(8.99,	77.01)

Factor Behandlung = Medikation

Time	Status	Cumulative Survival	Standard Error	Cumulative Events	Number Remaining
5.00	1.00	.8000	.1789	1	4
40.00	.00			1	3
64.00	1.00	.5333	.2483	2	2
69.00	.00			2	1
72.00	1.00	.0000	.0000	3	0

	Survival Time	Standard Error	95% Confidence Interval		
Mean:	56.47	14.26	(28.53,	84.41)	
Median:	72.00	.	(. ,	.)	
	Total	Number	Number	Percent	
		Events	Censored	Censored	
Behandlung	Placebo	5	4	1	20.00
Behandlung	Medikation	5	3	2	40.00
Overall		10	7	3	30.00

Test Statistics for Equality of Survival Distributions for BEH

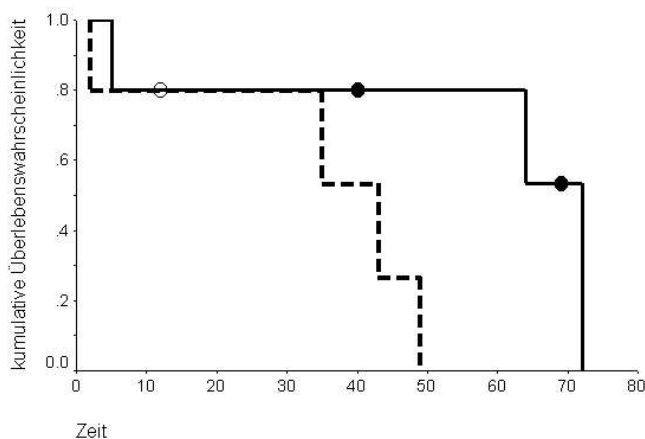
	Statistic	df	Significance
Log Rank	3.34	1	.0675
Breslow	1.83	1	.1757
Tarone-Ware	2.52	1	.1121

Zusätzlich kommen noch die Resultate des Gruppenvergleiches dazu. Drei verschiedene Tests werden hier angeboten. Alle diese Tests vergleichen die Anzahl erwarteter und beobachteter Erkrankungen zu jedem der Zeitpunkte. Der Log Rank Test gewichtet alle Ereignisse gleich, während der Breslow (auch generalisierter Wilcoxon Test genannt) und der Tarone-Ware Test die früheren Ereignisse stärker gewichten.

Wenn die Ereignisraten in den verschiedenen Gruppen ein vielfaches voneinander sind (Multiplikation), so ist der Log Rank Test am mächtigsten, sonst der Breslow Test. Der Breslow Test ist jedoch problematisch, wenn viele zensierte Beobachtungen vorliegen, da die frühen Ereignisse einen sehr starken Einfluss haben.

In unserem Beispiel sehen wir, dass keiner der Tests die Signifikanzgrenze erreicht (was wohl hauptsächlich an der sehr kleinen Stichprobengrösse liegt).

Die graphische Darstellung dieser Ergebnisse:



2.3 Erweiterungen

- Kategorische Variablen mit mehr als 2 Stufen (z. B. Placebo und zwei verschiedene Behandlungen): Die Tests des letzten Abschnittes sagen uns bloss, ob es zwischen irgendwelchen zwei Gruppen einen Unterschied gibt. Um herauszufinden welche Gruppen sich tatsächlich unterscheiden gibt es sogenannte paarweise Vergleiche (Achtung: multiples Testen!).

- Strata: Wenn wir an einem Unterschied verschiedener Medikationen interessiert sind, aber bereits wissen, dass verschiedene Typen der Krankheit zu unterschiedlichen Zeitintervallen bis zum Wiederausbruch führen, so müssen wir beide Faktoren ins Modell einfließen lassen. Wir sind dann an den Unterschieden der Medikationen (Faktor) innerhalb der Krankheiten (Strata) interessiert.

3 Cox Regression / ‘Proportional hazard’

Die Kaplan-Meier Methode ist auf kategoriale Variablen beschränkt und kann mit Hilfe der Strata auf maximal zwei solche Variablen erweitert werden. Die Cox Regression gibt uns hier zusätzliche Flexibilität, indem es möglich wird kontinuierliche Variablen zu berücksichtigen und die Anzahl der Variablen nicht beschränkt ist.

Die Ansätze der Cox Regression sollen hier nur kurz skizziert werden. Als Ansatz benutzen wir die kumulative Überlebensfunktion, d. h. wir schätzen wie bisher den Anteil der Patienten ohne Ereignis:

$$S(t) = [S_0(t)]^{e^{(\beta_1 x^{(1)} + \beta_2 x^{(2)} + \dots + \beta_m x^{(m)})}}$$

$S_0(t)$ ist die Grundüberlebensfunktion und entspricht der Konstanten in der multiplen Regression. Für jede Wertekombination in den erklärenden Variablen lässt sich mit dieser Gleichung eine ganze Überlebenszeitfunktion berechnen.

Für eine erste Interpretation der β 's können wir wieder deren Vorzeichen verwenden, dazu müssen wir jedoch wissen, dass $0 < S_0(t) < 1$. Ist nun z. B. $\beta_1 < 0$ und wir erhöhen den Wert von $x^{(1)}$ so wird der Wert von $S(t)$ erhöht, was gleichbedeutend mit einer Verlängerung der Zeit bis zum Ereignis ist. Umgekehrt verkürzt sich die Zeit bis zum Ereignis bei steigenden Werten von $X^{(1)}$, wenn $\beta_1 > 0$ ist (gleiche Argumentation).

Da es mathematisch einfacher ist, wird oft die kumulative ‘Hazard’ Funktion anstelle der Überlebensfunktion verwendet. Leider ist die Vorstellung dieses Masses im Gegenzug dazu schwieriger. Der Hazard ist ein Mass für die Ereignisrate zum Zeitpunkt t (Anzahl Ereignisse pro Zeitintervall), gegeben, dass der Patient diesen Zeitpunkt erreicht hat. (Der Hazard kann somit Werte grösser als 1 annehmen.) Das Modell kann dann geschrieben werden als:

$$h(t) = [h_0(t)] \cdot e^{(\beta_1 x^{(1)} + \beta_2 x^{(2)} + \dots + \beta_m x^{(m)})}$$

Da der Quotient der Hazards zweier Patienten mit gegebenen Eigenschaften zu jedem Zeitpunkt den gleichen Wert ergibt, spricht man auch von ‘proportional hazards’ Modellen. Falls die Hazards nicht proportional sind, d. h. die kumulativen Hazards oder die kumulativen Überlebenskurven nicht gleichmässig auseinanderlaufen oder sich gar überschneiden, darf man diesen Modellansatz nicht verwenden, da Annahmen verletzt werden. (Es gibt zusätzliche Instrumente zur Überprüfung dieser und weiterer Annahmen, aber darauf soll hier nicht eingegangen werden.)

Die statistischen Tests benutzen wie die logistische Regression das Verfahren des Maximum-Likelihood und die dazugehörigen Likelihood-Ratio Tests. Modelle können in analoger Weise wie bisher in stepwise forward oder backward Verfahren gesucht werden.