Sterbetabellenanalyse

1. Allgemeines

Die primäre Aufgabe der Sterbetabellenanalyse (oder auch *Ereigniszeitanalyse*) besteht in der Berechnung der Überlebenswahrscheinlichkeiten für ein Zeitintervall, d.h. dass eine Person dieses Zeitintervall überlebt (bzw. das untersuchte Ereignis nicht eintritt), z.B. das Überleben nach Erkennen (und Behandlung) eines bestimmten Tumors. Andere Beispiele: Analyse der Dauer der Arbeitslosigkeit nach einem Jobverlust oder die Analys der Dauer nach einer Heirat bis zur Scheidung.

Das Datenmodell sieht vor, dass es für jede Person einen Anfangszeitpunkt und einen Endzeitpunkt der Beobachtung gibt sowie eine Ereignisvariable, die den Zustand zum Endzeitpunkt enthält: typischerweise entweder „gestorben“ oder „lebendig“ („censored“), bzw. geschieden oder noch verheiratet. Letzteres ist dann der Fall, wenn die Person nicht weiter beobachtet werden konnte oder aber aus einem anderen Grund verstorben ist, als es das primäre Untersuchungsziel ist, etwa durch einen Unfall. SPSS verlangt anstatt der beiden Zeitpunkte eine Variable, die die Zeitdifferenz enthält, die gegebenenfalls selbst errechnet werden muss.

2. Datumsfunktionen in SPSS

Wichtige Datumsfunktionen in SPSS:

- Tag/Monat/Jahr sind getrennte Variablen:
  DATE.DMY(*tag, monat, jahr*)
  DATE.MOYR(*monat, jahr*)
  DATE.WKYR(*woche, jahr*)
  DATE.YRDAY(*jahr, tag*)
  fassen jeweils die Variablen in eine zusammen, die an schließend vom Typ *Datum* deklariert werden muss, wobei das Format irrelavant ist, allerdings tt-mmm-jjjj das verständlichste ist.
- Tag/Monat/Jahr sind in nur einer Variable gespeichert:
  Diese muss vom Typ *Datum* deklariert sein, wobei das verwendete Format, z.B. tt.mm.jj , spezifiziert werden muss.
- Berechnung einer Zeitdifferenz:
  Die Funktion DATEDIFF emöglicht die Berechnung der Zeitdifferenz aus zwei Datumsvariablen, wahlweise in Anzahl Tage, (volle) Monate bzw. (volle) Jahre:
  DATEDIFF(*Datum2*, *Datum1*,“days“)
  DATEDIFF(*Datum2*, *Datum1*,“months“)
  DATEDIFF(*Datum2*, *Datum1*,“years“)

  Alternativ kann die Zeitdifferenz auch “elementar“ ermittelt werden: Zunächst muss immer für jedes Datum die Zeitspanne zum 14.10.1582 errechnet werden und zwar mittels
  CTIME.DAYS(*datum*)
  Aus der Differenz der beiden Zeitspannen wird dann die Zeitdifferenz in Tagen errechnet. Die Umrechnung z.B. in Monaten oder Jahren muss dann selbst erfolgen.
- Berechnung von Tag/Monat/Jahr aus einem Datum:
  XDATE.MDAY(*datum*)
  XDATE.MONTH(*datum*)
  XDATE.YEAR(*datum*)
  errechnen Tag, Monat und Jahr aus einer Variable vom Typ *Datum.*

- Hilfreiche Webseiten:
  Übersicht der wichtigsten Datums- und Zeitfunktionen:
  http://pascal.kgw.tu-berlin.de/gnom/Lehre/spss/funktionen.html
  Beispiele hierzu:
  http://www.ats.ucla.edu/stat/spss/library/dates.htm

3. Wichtige Funktionen

Überlebensfunktion *S(t)*: Wahrscheinlichkeit dafür, dass die Zeit bis zum Eintreten des Ereignisses (z.B. Tod) länger als *t* ist.

Median der Überlebenszeit: Zeitpunkt, an dem die Überlebenswahrscheinlichkeit 0,5 beträgt, d.h. zu diesem Zeitpunkt sollte bei 50% der Personen das Ereignis bereits eingetreten sein.

Hazardfunktion $\lambda(t)$: Rate, mit der ein Ereignis zum Zeitpunkt *t* eintritt unter der Voraussetzung, dass es bis zum Zeitpunkt *t* noch nicht eingetreten ist.

Kumulative Hazardfunktion $\Lambda(t)$: Summe der Hazard-Risiken bis zum Zeitpunkt *t*. ($\Lambda(t)$ kann beliebig groß werden.)

Wahrscheinlichkeitsdichte: Schätzung der Wahrscheinlichkeit mit der das Ereignis zum Zeitpunkt *t* eintritt (im Gegensatz zur Hazardfunktion *ohne* die Prämisse, dass es bis zum Zeitpunkt *t* noch nicht eingetreten ist).

Kaplan-Meier-Überlebensfunktion: Alternative Berechnung (Produkt-Limit-Methode) der Überlebenswahrscheinlichkeiten, die - im Gegensatz zu den o.a. Methoden - keine Einteilung in Intervalle erfordert.

Cox-Regression: eine der logistischen Regression ähnliche Methode, primär zur Analyse von Einflussgrößen (*Kovariate*) auf die Hazard-Funktion (und damit auf die Überlebenswahrscheinlichkeiten). Dabei können die Einflussgrößen auch zeitabhängig sein.

4. Methoden

Überleben -> Sterbetafeln:
Der zu betrachtende Zeitbereich wird in eine Anzahl von z.B. sieben oder zehn gleich lange Intervalle unterteilt. Die Berechnungen, insbesondere Schätzwerte für die Überlebenszeitfunktion, werden dann jeweils zu den Endpunkten der Intervalle durchgeführt. Bezüglich der Überlebenszeitverteilung sind Gruppenvergleiche möglich.

Überleben -> Kaplan-Meier:
Die Produkt-Limit-Methode nach Kaplan und Meier erlaubt differenziertere Analysen. Sie stellt Berechnungen zu jedem Zeitpunkt an, an dem ein Response erfolgte. Auch hier sind Gruppenvergleiche, darüber hinaus Trend- und geschichtete Analysen möglich. Da bei dieser Methode keine Einteilung in Zeitintervalle vorgenommen wird, empfiehlt sie sich nur für kleinere Fallzahlen.

Überleben -> Cox-Regression:
Unter diesem Ansatz wird untersucht, ob die Hazardfunktion und damit die Überlebenszeitfunktion von Kovariaten beeinflußt wird, d.h. ob vermutete Einflußgrößen einen prognostischen Wert für die Überlebenszeitverteilung besitzen.

Überleben -> Cox mit zeitabhängigen Kovariaten:
Es ist möglich, beim Cox-Modell die zeitliche Veränderlichkeit der Kovariaten zu berücksichtigen und die mögliche Abhängigkeit der Überlebenszeitverteilung auch von solchen Kovariaten zu untersuchen. Dies ist sinnvoll, wenn eine Kovariate (z.B. Blutdruck) im Laufe der Zeit mehrfach erhoben wurde. SPSS stellt eine interne Variable T_ zur Verfügung, die das jeweilige Zeitintervall angibt und mit deren Hilfe die zeitabhängige Kovariate über arithmetische oder logische Abfragen berechnet werden kann.

Hinweise zur Cox-Regression:

- Kategoriale Kovariate müssen *kodiert* werden, d.h. in mehrere dichotome Variablen zerlegt werden. Hierfür wird die „Abweichung“ (Deviation)-Kodierung empfohlen. Die Regressionskoeffizienten für jede Kategorie geben die Veränderung der Sterberate in der Gruppe dieser Ausprägung gegenüber dem Durchschnitt aller Gruppen an. Für die Gruppe der Ausprägung, die nicht für die Dichotomisierung verwendet wurde, ist der Regressionskoeffizient berechenbar als Summe aller *m -1* Regressionskoeffizienten, jedoch mit umgekehrtem Vorzeichen.
- Positive Regressionskoeffizienten bedeuten eine Verringerung der Überlebenswkt.
- *exp(b)* gibt an, um das wievielfache die Hazardfunktion ansteigt bzw. abfällt, wenn man die Kovariate um eine Einheit vergrößert.
- Die möglichen Varianten der Auswahl der Kovariaten entsprechen denen der Regression, z.B. Einschluss und auf- bzw. absteigend schrittweise.
- Die Tests sind mit denen der logistischen Regression identisch.
- Falls der Einfluss einer Kovariaten über die Zeit als nicht proportional anzusehen ist, sollte eine geschichtete Berechnung erfolgen. Dazu ist eine solche Kovariate unter Schichten einzutragen. Die Berechnungen werden für die Schichten getrennt durchgeführt, die Tests allerdings gepoolt.

6. Gruppenvergleiche

Die Kaplan-Meier-Methode erlaubt Vergleiche der Überlebensfunktion für Gruppen. Folgende Methoden stehen zur Verfügung, die sich in der Gewichtung der einzelnen Zeitpunkte unterscheiden:

- Log Rank:
  alle Zeitpunkte werden gleich gewichtet.
- Breslow:
  Gewichtung proportional zur Anzahl der Fälle, die zum Zeitpunkt noch dem Risiko ausgesetzt sind.
- Tarone-Ware:
  Gewichtung proportional zur Wurzel aus der Anzahl der Fälle, die zum Zeitpunkt noch dem Risiko ausgesetzt sind.

Bei der Methode von Breslow werden die Unterschiede im „hinteren“ Zeitbereich, in dem nur noch weniger Fälle zur Verfügung stehen, gering berücksichtigt. Die Methode von Tarone-Ware stellt quasi einen Kompromiss zwischen den beiden anderen dar.

Standardmäßig wird ein globaler Test (ähnlich einer Varianzanalyse) durchgeführt, wahlweise können aber auch paarweise Vergleiche angefordert werden. Diese Tests können auch getrennt für die Schichten einer weiteren Gruppierungsvariablen durchgeführt werden.

Bei der „elementaren“ Sterbetabellen-Analyse kann ebenfalls ein Gruppenvergleich der Überlebensfunktion durchgeführt werden, und zwar ein Wilcoxon-Test.

6. Literatur

http://www.mh-hannover.de/fileadmin/institute/biometrie/Scripte/ausgew/survival.pdf

Haiko Lüpsen
Regionales Rechenzentrum der Universität zu Köln

21.8.2012