

Studiengang Medizinische Biometrie: Survivalanalyse

Prof. Dr. Vanessa Didelez

BIPS, Universität Bremen

WS 2019/20

Einleitung

Motivation

Beispiele

Grundlegende Definitionen

Zensierung

Nicht-parametrische Verfahren

Schätzen der Survivorfunktion

Schätzen von Median und Quantile

Schätzen der Hazardfunktion

Bias in der Schätzung von $S(t)$

Standardfehler der Schätzungen von $S(t)$

Vergleich von Survival-Kurven

Modellierung: Das Cox Modell

Das Proportional Hazards Modell

Interpretation der Parameter im Cox Modell

Schätzen im Cox Modells

Konfidenzintervalle für β

Vergleich alternativer Modelle

Schätzen von Hazard- und Survivalfunktion

Residuenanalyse

- Residuen im Cox Modell
- Martingal-Residuen
- Deviance-Residuen
- Schoenfeld-Residuen
- Score-Residuen

Erweiterungen des Cox Modells

- Cox Modell und Log-Rank-Test
- Stratifizierte Cox-Regression
- Zeitabhängige Kovariablen
- Kausale Interpretation

Parametrische Modelle

- Exponential- und Weibullverteilung
- Schätzen im parametrischen Modell
- Modell für den Vergleich zweier Gruppen
- Das Weibull Proportionale Hazards Modell
- Das Gompertz Proportionale Hazards Modell

Additive Hazard Modelle

Das additive Hazard Modell
Schätzen im additiven Hazard Modell
Datenbeispiele
Ausblick

Competing Risks

Das Problem bei “Competing Risks”
Cause-specific Hazards
Bias bei naiver Schätzung
Schätzung bei Competing Risks
Modellierung mit Einflußgrößen
Ausblick: Multi-state Modelle

Bestimmung des Stichprobenumfangs

Vergleich von zwei Behandlungsgruppen
Anzahl erwartete Fälle (nach Schoenfeld, 1983)
Benötigte Anzahl Probanden

Zensierungsmechanismen

Zensierungstypen (Wdhlg.)
Likelihoodkonstruktion
Nichtparametrische Schätzung bei Linkstrunkierung

Ausblick

Organisatorisches

Vorlesung und Übung (V2, Ü1)

- ▶ Vorlesung: Mittwochs 8:30-10:00 (Reserve: Dienstags statt Übung)
- ▶ Übung: Dienstags 8:15-9:00
- ▶ Keine Vorlesung: 4.12. (Dies Academicus).
- ▶ Prüfung: Klausur!

Literatur zur Vorlesung

- ▶ Klein, Moeschberger: Survival Analysis. Springer.
- ▶ Collett, D. 2003: Modelling Survival Data in Medical Research, 2nd Edition. Chapman & Hall/CRC, Boca Raton, Florida, USA.
- ▶ Allison, P. 2010: Survival Analysis Using SAS - A Practical Guide, 2nd Edition. SAS Institute Inc., Cary, NC, USA.

Inhalt von Abschnitt 1

Einleitung

Motivation

Beispiele

Grundlegende Definitionen

Zensierung

Was ist “Überlebenszeitanalyse” (Survival Analysis)

Daten: eine oder mehrere der zentralen (Ziel)Größen messen eine “Dauer” bzw. einen Zeitraum von Anfang bis Eintreten eines / verschiedener Ereigniss(e).

Beispiele:

- ▶ Zeit, die ein aortokoronarer Bypass offen bleibt
- ▶ Dauer bis zum ersten Defekt einer neue Waschmaschine
- ▶ Dauer bis Studierende nach MSc Abschluss den ersten Job finden
- ▶ Überlebenszeit nach Nierentransplantation
- ▶ Zeit zwischen zwei epileptischen Anfällen
- ▶ Dauer bis zum nächsten Vulkanausbruch

Warum “Überlebenszeitanalyse”

Besonderheiten: die Dauer bis zu einem Ereigniss kann oft nicht vollständig beobachtet werden, d.h. oft liegt nur die Information vor, dass das Ereignis bis zum Zeitpunkt t noch nicht eingetreten ist.

Dies wird als **Zensierung (censoring)** bezeichnet.

Viele der sonst üblichen Kenngrößen und Inferenzen lassen sich mit zensierten Daten nicht erhalten.

Beispiele: → Diskussion.

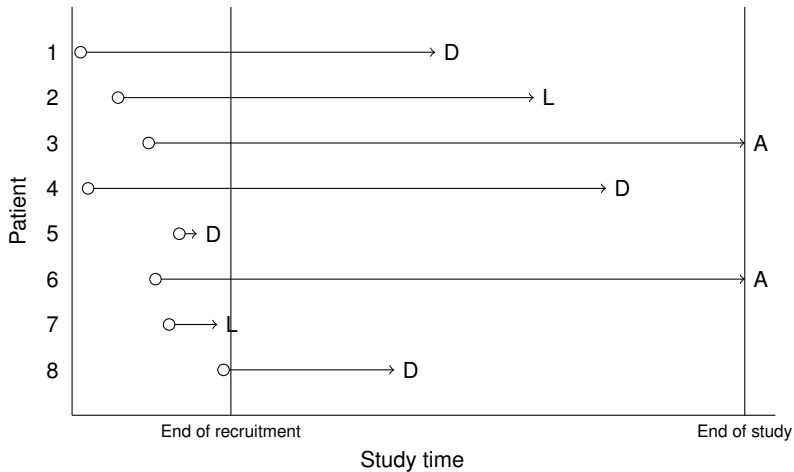
Zensierung

Zensierung: verschiedene Arten

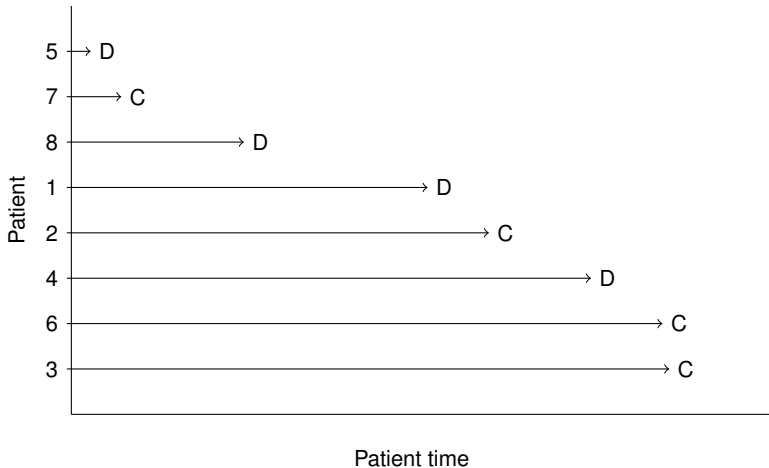
- ▶ Rechtszensierung (drop-out, Studienende)
- ▶ Linkszensierung
- ▶ Intervallzensierung (diskrete Beobachtungszeitpunkte)
- ▶ “Truncation”, Trunkierung (best. Zeiten nicht in der Stichprobe)
- ▶ etc.

Annahme: der Zensierungsmechanismus wird gerne als **uninformativ** oder **unabhängig** angenommen; *informell:* der Grund für die Zensierung sollte nicht systematisch mit der zensierten Dauer zusammenhängen. Das ist leider oft unrealistisch...

Typische Studie



Typische Studie (Forts.)



Beispiel 1.1 (Intra-Uterine Device (IUD))

Tabelle: Time in weeks to discontinuation of the use of an IUD.

| | | | | | | | | |
|-----|-----|-----|----|-----|------|-----|------|------|
| 10 | 13* | 18* | 19 | 23* | 30 | 36 | 38* | 54* |
| 56* | 59 | 75 | 93 | 97 | 104* | 107 | 107* | 107* |

Beobachtungen mit * stellen rechts-zensierte Zeiten dar.

Beispiel 1.2 (Prognose bei Brustkrebs bei Frauen)

Tabelle: Survival times (in months) after mastectomy of women with Grade 2-4 tumours that were negatively or positively stained with HPA.

| | | | | | | | | |
|-------------------|----------------------|----------------------|-------------------------|--------------------------|--------------------------|--------------------------|-------------------------|--------------------------|
| Negative staining | 23 181 | 47 198* | 69 208* | 70* 212* | 71* 224* | 100* | 101* | 148 |
| Positive staining | 5 31 68 118 | 8 35 71 143 | 10 40 76* 154* | 13 41 105* 162* | 18 48 107* 188* | 24 50 109* 212* | 26 59 113 217* | 26 61 116* 225* |

Beobachtungen mit * stellen rechts-zensierte Zeiten dar.

Beispiel 1.3 (Überleben von Patienten mit multiplem Myelom)

Multiple myeloma patients (Collett, p. 9)
(Auszug)

| patient | time | status | age | sex | bun | ca | hb | pcells | protein |
|---------|------|--------|-----|-----|-----|----|------|--------|---------|
| 1 | 13 | 1 | 66 | 1 | 25 | 10 | 14.6 | 18 | 1 |
| 2 | 52 | 0 | 66 | 1 | 13 | 11 | 12.0 | 100 | 0 |
| 3 | 6 | 1 | 53 | 2 | 15 | 13 | 11.4 | 33 | 1 |
| 4 | 40 | 1 | 69 | 1 | 10 | 10 | 10.2 | 30 | 1 |
| 5 | 10 | 1 | 65 | 1 | 20 | 10 | 13.2 | 66 | 0 |
| 6 | 7 | 0 | 57 | 2 | 12 | 8 | 9.9 | 45 | 0 |
| 7 | 66 | 1 | 52 | 1 | 21 | 10 | 12.8 | 11 | 1 |
| 8 | 10 | 0 | 60 | 1 | 41 | 9 | 14.0 | 70 | 1 |
| 9 | 10 | 1 | 70 | 1 | 37 | 12 | 7.5 | 47 | 0 |
| 10 | 14 | 1 | 70 | 1 | 40 | 11 | 10.6 | 27 | 0 |

Biometrie\Examp1-34.sas

Beispiel 1.4 (Vergleich von Behandlungen bei Prostata-Karzinom)

Treatment of prostatic cancer patients (Collett, p. 10)
(Auszug)

| patient | treat | time | status | age | shb | size | index |
|---------|-------|------|--------|-----|------|------|-------|
| 1 | 1 | 65 | 0 | 67 | 13.4 | 34 | 8 |
| 2 | 2 | 61 | 0 | 60 | 14.6 | 4 | 10 |
| 3 | 2 | 60 | 0 | 77 | 15.6 | 3 | 8 |
| 4 | 1 | 58 | 0 | 64 | 16.2 | 6 | 9 |
| 5 | 2 | 51 | 0 | 65 | 14.1 | 21 | 9 |
| 6 | 1 | 51 | 0 | 61 | 13.5 | 8 | 8 |
| 7 | 1 | 14 | 1 | 73 | 12.4 | 18 | 11 |
| 8 | 1 | 43 | 0 | 60 | 13.6 | 7 | 9 |
| 9 | 2 | 16 | 0 | 73 | 13.8 | 8 | 9 |
| 10 | 1 | 52 | 0 | 73 | 11.7 | 5 | 9 |

Biometrie\Examp1-34.sas

Survivalfunktion

Aufgrund der zensierten Daten, werden in der Überlebenszeitanalyse andere grundlegende Größen verwendet.

Die beobachtete Dauer / Überlebenszeit t wird aufgefasst als Realisierung einer positiven (*stetigen*) Zufallsvariablen T . Die Verteilungsfunktion von T ist

$$F(t) = \Pr(T \leq t) = \int_0^t f(u) du.$$

Anstatt der Verteilungsfunktion ist die **Survivalfunktion** (Überlebenszeitfunktion) $S(t)$ von Interesse; sie ist definiert als

$$S(t) = \Pr(T > t) = 1 - F(t). \quad (1.1)$$

Es gilt $f(t) = -\frac{dS(t)}{dt}$. → Warum?

Hazardfunktion

Anstatt der Dichte ist in der Survivalanalyse die **Hazardfunktion** $\lambda(t)$ von Interesse; sie ist definiert als

$$\lambda(t) = \lim_{\Delta t \rightarrow 0} \frac{\Pr(t \leq T < t + \Delta t | T \geq t)}{\Delta t}. \quad (1.2)$$

Und dazu ist die **kumulative Hazardfunktion**

$$\Lambda(t) = \int_0^t \lambda(u) du. \quad (1.3)$$

→ Interpretation?!

Beispiel für Hazardkurven:

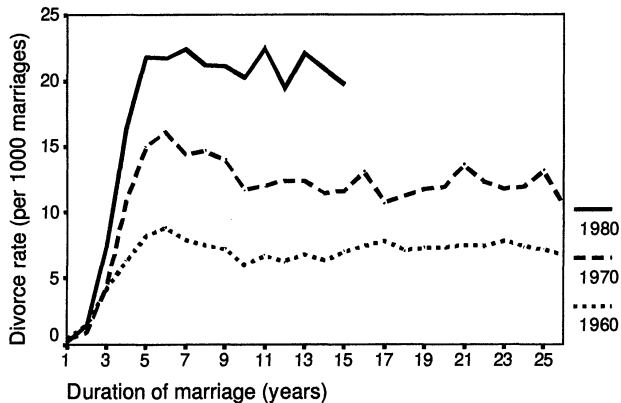


FIG. 2. *Rates of divorce among Norwegian couples married in 1960 (lower curve), 1970 (middle curve) and 1980 (upper curve). (Based on data from Statistics Norway.)*

(aus Aalen & Gjessing (2001). Understanding the shape of the hazard rate. Statistical Science)

Zusammenhänge

Es gilt

$$\lambda(t) = \frac{f(t)}{S(t)} = -\frac{d \log(S(t))}{dt} \quad (1.4)$$

$$S(t) = \exp\{-\Lambda(t)\}. \quad (1.5)$$

Beispiele: → Skizzen.

Bemerkung: Jede der Funktionen f , F , S , λ und Λ legt die Verteilung von T eindeutig fest und kann aus den anderen hergeleitet werden.

Beispiel: → Exponentialverteilung.

Mean Residual Life

- ▶ **Mean residual life** zum Zeitpunkt t : erwartete verbleibende Lebenszeit zum Zeitpunkt t

$$r(t) = E(T - t | T \geq t) = \frac{\int_t^\infty (t - u) f(u) du}{S(t)} = \frac{\int_t^\infty S(u) du}{S(t)}.$$

- ▶ Es gilt $E(T) = r(0) = \int_0^\infty S(u) du$
- ▶ **p -Quantile** ergeben sich aus der Gleichung $S(t_p) = 1 - p$, also im absolut stetigen Fall $t_p = S^{-1}(1 - p)$.

Beispiel: Exponentialverteilung: $t_{med} = \log(2)/\lambda$.

Modell für Rechtszensierung

n Einheiten (z.B. Patienten). Für $i = 1, \dots, n$:

- ▶ Survival-Zeit: T_i (Zeit vom Start, z.B. Rekrutierung, bis zum Ereignis)
- ▶ Zensierungszeit: C_i
- ▶ Beobachtungen: $T_i^* = \min(T_i, C_i)$ und Indikator

$$\delta_i = \begin{cases} 0 & T_i > C_i \\ 1 & T_i \leq C_i \end{cases}$$

Damit enthält (T_i^*, δ_i) die gesamte Information.

- ▶ **Uninformative Zensierung:** für alle i , und t , sowie 'kleines' Δt :

$$\Pr(t \leq T_i < t + \Delta t | T_i \geq t, C_i < t) = \Pr(t \leq T_i < t + \Delta t | T_i \geq t).$$

(Oft ist dies nur unter Einbeziehen von Covariablen plausibel.)

Likelihood

Parameter(vektor) von $F(t)$ sei θ .

Ohne Zensierung: $L(\theta|\text{Daten}) = \prod f(t_i)$.

Mit uninformativer Rechtszensierung zwei Fälle:

- (1) Beobachtung $(t_i, \delta_i = 0)$, d.h. zensiert; damit trägt i nur die Information bei, dass $T_i > t_i$ war $\Rightarrow S(t_i)$.
 - (2) Beobachtung $(t_i, \delta_i = 1)$, d.h. nicht zensiert; damit 'normale' Information $\Rightarrow f(t_i)$.
- Da Zensierung uninformativ, kann die Information in δ_i selbst ignoriert werden.
- \Rightarrow Likelihood $L(\theta|(t_1, \delta_1), \dots, (t_n, \delta_n)) = \prod f(t_i)^{\delta_i} S(t_i)^{1-\delta_i}$;
dies kann umgeschrieben werden zu dem üblichen Ausdruck

$$L(\theta|(t_1, \delta_1), \dots, (t_n, \delta_n)) = \prod_{i=1}^n \lambda(t_i)^{\delta_i} \exp\{-\Lambda(t_i)\}.$$

Inhalt von Abschnitt 2

Nicht-parametrische Verfahren

Schätzen der Survivorfunktion

Kaplan-Meier-Schätzer von $S(t)$

Nelson-Aalen-Schätzer von $S(t)$

Schätzen von Median und Quantile

Schätzen der Hazardfunktion

Nelson-Aalen-Typ-Schätzung von $\lambda(t)$

Schätzen der kumulativen Hazardfunktion

Bias in der Schätzung von $S(t)$

Standardfehler der Schätzungen von $S(t)$

Standardfehler des KM-Schätzers $\hat{S}(t)$

Standardfehler des Nelson-Aalen-Schätzer

Punktweises Konfidenzintervall für $S(t)$

Konfidenzintervalle für Median und Quantile

Vergleich von Survival-Kurven

Vergleich von zwei Survival-Kurven

Vergleich von mehr als zwei Survivalkurven

Stratifizierte Tests

Inhalt von Abschnitt 2 (Forts.)

Log-Rank-Test auf Trend
Darstellung von Survivalkurven

Erster Schritt in der Analyse von Ereigniszeitdaten:
numerische oder graphische Zusammenfassung der individuellen
Überlebenszeiten via

- ▶ Survivalfunktion
- ▶ Hazardfunktion.

Oft: Vergleich des Survivals von zwei oder mehr Gruppen durch nicht-parametrische Tests (Log-Rank- und Wilcoxon-Test).

Schätzen der Survivorfunktion

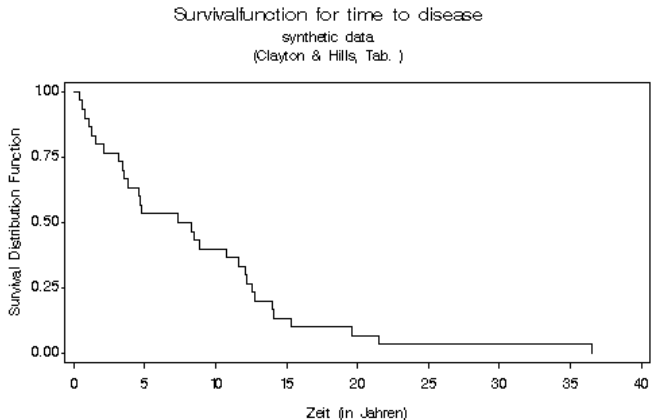
Ohne zensierte Daten wäre eine naheliegende Schätzung der Survivalkurve

$$\hat{S}(t) = \frac{\#\{\text{Personen mit Survivalzeit} > t\}}{\#\{\text{Personen in Stichprobe}\}}, \quad (2.1)$$

$\hat{S}(t) = 1 - \hat{F}(t)$, $\hat{F}(t)$ ist die empirische Verteilungsfunktion.

Beispiel 2.1 (Fiktive Erkrankungszeiten)

Inzidenzdaten von 30 Personen (Clayton & Hills, 1993, p. 43)



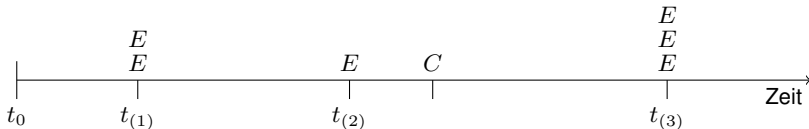
Biometrie\life.sas

Kaplan-Meier-Schätzer von $S(t)$

Die **Kaplan-Meier-Methode** (KM-Methode, auch Produkt-Limit-Schätzer) beruht auf der Kenntnis der genauen individuellen Ereignis- und Zensierungszeiten.

Betrachte eine Gruppe von n Personen mit t_1, \dots, t_n beobachteten Survival- oder Zensierungszeiten.

$t_{(1)} < t_{(2)} < \dots < t_{(r)}$ seien die geordneten **Ereigniszeiten**



Kaplan-Meier-Schätzer von $S(t)$ (Forts.)

Intuitive Begründung für K-M-Schätzer??

Für die j -te Ereigniszeit, $1 \leq j \leq r$, sei

n_j – Anzahl unter Risiko zum Zeitpunkt $t_{(j)}^-$

d_j – Anzahl Fälle in $t_{(j)}$

$\hat{q}_j = \frac{d_j}{n_j}$ – (bedingte) Ereigniswahrsch. in $t_{(j)}$

$\hat{p}_j = 1 - \hat{q}_j = \frac{n_j - d_j}{n_j}$ – (bedingte) Überlebenswahrsch. in $t_{(j)}$

Der Kaplan-Meier-Schätzer $\hat{S}(t)$ für $S(t)$ ergibt sich dann für $t_{(k)} \leq t < t_{(k+1)}$ als

$$\hat{S}(t) = \prod_{j=1}^k \frac{(n_j - d_j)}{n_j} = \prod_{j:t(j) \leq t} \left[1 - \frac{d_j}{n_j} \right] \quad (2.2)$$

Kaplan-Meier-Schätzer von $S(t)$ (Forts.)

- ▶ $\hat{S}(t) = 1$ für $t < t_{(1)}$
- ▶ $\hat{S}(t) = \hat{S}(t_{(k)})$ für $t_{(k)} \leq t < t_{(k+1)}$ (Stufenfunktion)
- ▶ Falls Zensierungs- und Survivalzeitpunkte gleich sind: Zensierungszeit ist später als Ereigniszeit.
- ▶ Daten ohne Zensierung: dann gilt $n_j - d_j = n_{j+1}$ und („Telescoping“)

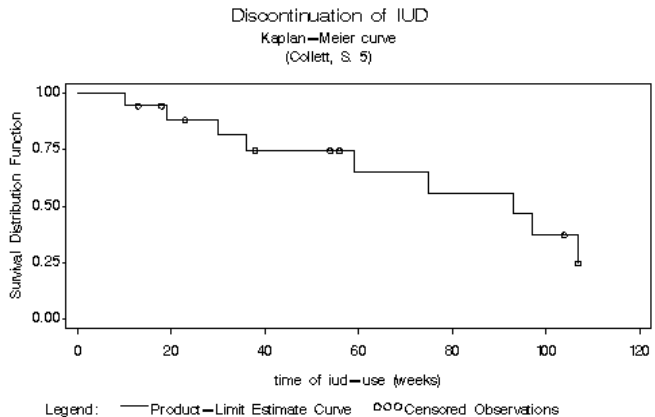
$$\begin{aligned}\hat{S}(t) &= \frac{n_2}{n_1} \cdot \frac{n_3}{n_2} \cdots \frac{n_{k+1}}{n_k} \\ &= \frac{n_{k+1}}{n_1},\end{aligned}\tag{2.3}$$

d.h. ohne Zensierung reduziert sich der K-M-Schätzer auf (2.1).

Beispiel 2.2 (IUD-Daten (Bsp. 1.1))

Kaplan-Meier-Schätzer der Survivalfunktion

| Time interval | n_j | d_j | $(n_j - d_j)/n_j$ | $\hat{S}(t)$ |
|---------------|-------|-------|-------------------|--------------|
| 0– | 18 | 0 | 1,0000 | 1,0000 |
| 10– | 18 | 1 | 0,9444 | 0,9444 |
| 19– | 15 | 1 | 0,9333 | 0,8815 |
| 30– | 13 | 1 | 0,9231 | 0,8137 |
| 36– | 12 | 1 | 0,9167 | 0,7459 |
| 59– | 8 | 1 | 0,8750 | 0,6526 |
| 75– | 7 | 1 | 0,8571 | 0,5594 |
| 93– | 6 | 1 | 0,8333 | 0,4662 |
| 97– | 5 | 1 | 0,8000 | 0,3729 |
| 107 | 3 | 1 | 0,6667 | 0,2486 |



Biometrie\life.sas

Nelson-Aalen-Schätzer von $S(t)$

Ein alternativer Schätzer von $S(t)$, gleichfalls beruhend auf individuellen Ereigniszeiten, ist der **Nelson-Aalen-Schätzer** $\tilde{S}(t)$, für $t_{(k)} \leq t < t_{(k+1)}$ definiert als

$$\begin{aligned}\tilde{S}(t) &= \prod_{j=1}^k \exp(-d_j/n_j) \\ &= \exp\left(-\sum_{j=1}^k d_j/n_j\right).\end{aligned}\tag{2.4}$$

Der Nelson-Aalen-Schätzer basiert auf einer Schätzung der kumulativen Hazardfunktion.

$$\tilde{\Lambda}(t) = \sum_{j:t(j) \leq t} \frac{d_j}{n_j}$$

→ Warum??

Nelson-Aalen-Schätzer von $S(t)$ (Forts.)

- ▶ Der Kaplan-Meier-Schätzer ist eine Approximation des Nelson-Aalen-Schätzers
- ▶ N-A-Schätzer hat bessere Eigenschaften bei kleinen Stichproben als K-M.
- ▶ N-A-Schätzer nützlich speziell bei parametrischer Modellwahl.
- ▶ N-A-Schätzer $\tilde{\Lambda}(t)$ auch Basis für 'grobe' Schätzung der Hazardrate $\lambda(t) \rightarrow \approx$ Steigung.
- ▶ K-M und N-A stimmen asymptotisch überein.

Beispiel 2.3 (IUD)

| Time interval | n_j | d_j | $(n_j - d_j)/n_j$ | $\hat{S}(t)$ | $\exp(-d_j/n_j)$ | $\tilde{S}(t)$ |
|---------------|-------|-------|-------------------|--------------|------------------|----------------|
| 0– | 18 | 0 | 1,0000 | 1,0000 | 1,0000 | 1,0000 |
| 10– | 18 | 1 | 0,9444 | 0,9444 | 0,9460 | 0,9460 |
| 19– | 15 | 1 | 0,9333 | 0,8815 | 0,9355 | 0,8850 |
| 30– | 13 | 1 | 0,9231 | 0,8137 | 0,9260 | 0,8194 |
| 36– | 12 | 1 | 0,9167 | 0,7459 | 0,9200 | 0,7539 |
| 59– | 8 | 1 | 0,8750 | 0,6526 | 0,8825 | 0,6653 |
| 75– | 7 | 1 | 0,8571 | 0,5594 | 0,8669 | 0,5768 |
| 93– | 6 | 1 | 0,8333 | 0,4662 | 0,8465 | 0,4882 |
| 97– | 5 | 1 | 0,8000 | 0,3729 | 0,8187 | 0,3997 |
| 107 | 3 | 1 | 0,6667 | 0,2486 | 0,7165 | 0,2864 |

Schätzen von Median und Quantile

Die geschätzte mediane Ereigniszeit $\hat{t}(50) = \hat{t}_{0,5}$ ist definiert als kleinste beobachtete Survivalzeit, für die $\hat{S}(t) < 0,5$ gilt:

$$\hat{t}(50) = \min\{t_{(j)} \mid \hat{S}(t_{(j)}) < 0,5\}.$$

Falls $\hat{S}(t_{(j)}) > 0,5$ für alle t , so kann die mediane Ereigniszeit nicht geschätzt werden.

Ähnlich zum Median geht man auch bei anderen Quantilen vor.

Es kann alternativ natürlich auch der N-A-Schätzer verwendet werden.

Nelson-Aalen-Typ-Schätzung von $\lambda(t)$

Betrachte wieder $t_{(1)} < t_{(2)} < \dots < t_{(r)}$. Ein Schätzer für $\lambda(t)$ im Intervall $t_{(j)}$ bis $t_{(j+1)}$, $j < r$, ist

$$\hat{\lambda}(t) = \frac{d_j}{n_j \tau_j} \quad \text{mit } \tau_j = t_{(j+1)} - t_{(j)}. \quad (2.5)$$

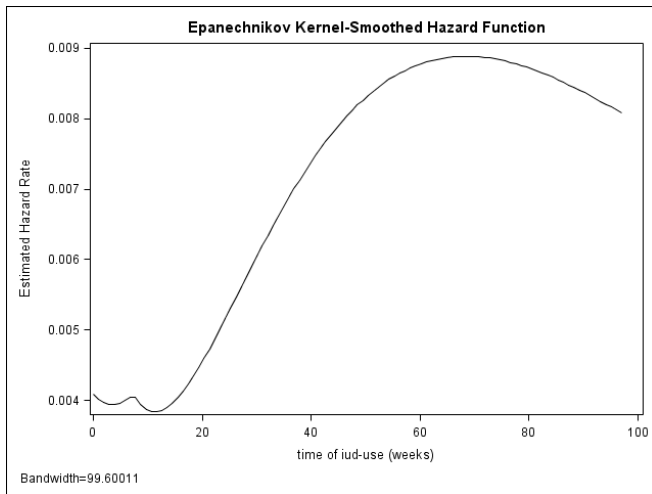
Die Varianz von $\hat{\lambda}(t)$ erhält man daraus, dass man d_j als Realisierungen einer $\text{Bin}(n_j, q_j)$ -verteilten Zufallsvariable auffasst (s. später):

$$\text{SE}(\hat{\lambda}(t)) \approx \hat{\lambda}(t) \left(\frac{n_j - d_j}{n_j d_j} \right)^{\frac{1}{2}}.$$

In der Regel ist das Aussehen von $\hat{\lambda}(t)$ sehr irregulär, deshalb Anwendung von Glättungstechniken, um die Hazardfunktion darzustellen.

Beispiel 2.4 (IUD)

| Time interval | τ_j | d_j | d_j | $\hat{\lambda}(t)$ | $SE(\hat{\lambda}(t))$ |
|---------------|----------|-------|-------|--------------------|------------------------|
| 0– | 10 | 18 | 0 | 0,0000 | – |
| 10– | 9 | 18 | 1 | 0,0062 | 0,0060 |
| 19– | 11 | 15 | 1 | 0,0061 | 0,0059 |
| 30– | 11 | 13 | 1 | 0,0128 | 0,0123 |
| 36– | 23 | 12 | 1 | 0,0036 | 0,0035 |
| 59– | 16 | 8 | 1 | 0,0078 | 0,0073 |
| 75– | 18 | 7 | 1 | 0,0079 | 0,0073 |
| 93– | 4 | 6 | 1 | 0,0417 | 0,0380 |
| 97– | 10 | 5 | 1 | 0,0200 | 0,0179 |



Schätzen der kumulativen Hazardfunktion

Die kumulative Hazardfunktion $\Lambda(t) = \int_0^t \lambda(u) du$ ist wichtig zur Identifikation von Modellen für Survivaldaten (siehe später).

Seien $t_{(1)} < \dots < t_{(r)}$ wieder die geordneten Ereigniszeiten und sei $t_{(k)} \leq t < t_{(k+1)}$, $k = 1, \dots, r$.

► Schätzung aus dem Kaplan-Meier-Ansatz:

$$\hat{\Lambda}(t) = - \sum_{j=1}^k \log \frac{n_j - d_j}{n_j}$$

► Nelson-Aalen-Schätzer:

$$\tilde{\Lambda}(t) = \sum_{j=1}^k \frac{d_j}{n_j}$$

Bias in der Schätzung von $S(t)$

Annahme: unabhängige Beobachtungen und uninformative Rechtszensierung.

Man kann zeigen, dass

$$\mathbb{E}(\tilde{\Lambda}(t)) - \Lambda(t) = - \int_0^t \lambda(u) \Pr(N_R(u) = 0) du$$

wobei $N_R(u)$ die Anzahl ‘at risk’ kurz vor Zeitpunkt u ist.

- ▶ Damit hat $\tilde{S}(t)$ – und aus ähnlichen Gründen auch $\hat{S}(t)$ – einen leicht positiven Bias, insbesondere für ‘spätere’ t , wo die Wahrscheinlichkeit $\Pr(N_R(u) = 0)$ nicht vernachlässigbar ist.
- ▶ Beide Schätzer sind aber auf einem festen Zeitintervall $[0, t]$ consistent.
- ▶ In der Praxis sollte man Schätzer für $S(t)$ für ‘späte’ t , vor allem nach dem letzten beobachteten Ereignis, einfach als undefiniert betrachten.

Standardfehler des KM-Schätzers

Für $t_{(k)} \leq t < t_{(k+1)}$ ist

$$\hat{S}(t) = \prod_{j=1}^k \hat{p}_j \quad \text{mit} \quad \hat{p}_j = \frac{n_j - d_j}{n_j},$$

$$\log \hat{S}(t) = \sum_{j=1}^k \log(\hat{p}_j) \quad \Rightarrow \quad \mathbb{V}(\log \hat{S}(t)) = \sum_{j=1}^k \mathbb{V}(\log \hat{p}_j). \quad (2.6)$$

Nun gilt

$$\hat{\mathbb{V}}(\hat{p}_j) = \frac{\hat{p}_j(1 - \hat{p}_j)}{n_j}. \quad (2.7)$$

Mit der „ δ -Methode“

$$\mathbb{V}(g(X)) \approx \left(\frac{dg(x)}{dx} \right)^2 \cdot \mathbb{V}(X) \quad (2.8)$$

erhält man

$$\hat{\mathbb{V}}(\log \hat{p}_j) \approx \frac{d_j}{n_j(n_j - d_j)}. \quad (2.9)$$

Standardfehler des KM-Schätzers (Forts.)

Mit (2.6) folgt
$$\hat{W}(\log \hat{S}(t)) \approx \sum_{j=1}^k \frac{d_j}{n_j(n_j - d_j)} \quad (2.10)$$

Wegen
$$\hat{W}(\log \hat{S}(t)) \approx \frac{1}{\hat{S}(t)^2} \hat{W}(\hat{S}(t))$$

folgt
$$\hat{W}(\hat{S}(t)) \approx [\hat{S}(t)]^2 \sum_{j=1}^k \frac{d_j}{n_j(n_j - d_j)} \quad (2.11)$$

und damit **Greenwoods Formel**:

$$\text{SE}(\hat{S}(t)) = \hat{S}(t) \left[\sum_{j=1}^k \frac{d_j}{n_j(n_j - d_j)} \right]^{\frac{1}{2}}. \quad (2.12)$$

Standardfehler des KM-Schätzers (Forts.)

Ohne Zensierung in den Daten ergibt sich aus (2.11):

$$\hat{\mathbb{V}}(\hat{S}(t)) = \frac{\hat{S}(t) (1 - \hat{S}(t))}{n_1}. \quad (2.13)$$

Dies ist eine Abschätzung der Varianz von $S(t)$ aus (2.1) unter der Annahme, dass die Anzahl Personen unter Risiko zum Zeitpunkt t einer $\text{Bin}(n_1, S(t))$ -Verteilung folgt.

Standardfehler des Nelson-Aalen-Schätzer

Für den N-A Schätzer $\tilde{\Lambda}(t)$ der kumulativen Hazardfunktion und für $t_{(k)} \leq t < t_{(k+1)}$ ist ein gängiger Varianzschätzer

$$\widehat{\mathbb{V}}(\tilde{\Lambda}(t)) = \sum_{j=1}^k \frac{d_j}{n_j^2}. \quad (2.14)$$

Erneute Anwendung der δ -Methode ergibt dann:

$$\text{SE}(\tilde{S}(t)) \approx \tilde{S}(t) \left[\sum_{j=1}^k \frac{d_j}{n_j^2} \right]^{\frac{1}{2}} \quad (2.15)$$

Beachte: $\mathbb{V}(\tilde{\Lambda}(t)) = \mathbb{V}(\log \tilde{S}(t)).$

→ Warum??

Punktwises Konfidenzintervall für $S(t)$

- Konfidenzintervall basierend auf asymptotischer Normalverteilung von $\hat{S}(t)$:

$$\hat{S}(t) \pm \text{SE}(\hat{S}(t)) \cdot u_{1-\frac{\alpha}{2}}.$$

Probleme, falls $\hat{S}(t) \approx 1$ oder $\hat{S}(t) \approx 0$.

- Alternative: Transformation von $S(t)$, z.B. komplementäre log-log-Transformation $\log(-\log S(t))$.

Es ergibt sich, dass

$$\widehat{\mathbb{W}}\{\log(-\log \hat{S}(t))\} \approx \frac{1}{(\log \hat{S}(t))^2} \sum_{j=1}^k \frac{d_j}{n_j(n_j - d_j)} \quad (2.16)$$

und man erhält ein $100 \cdot (1 - \alpha)\%$ -KI für $S(t)$ als

$$\hat{S}(t)^{\exp[\pm \text{SE}\{\log(-\log \hat{S}(t))\} \cdot u_{1-\frac{\alpha}{2}}]}. \quad (2.17)$$

Beispiel 2.5 (IUD)

Discontinuation of IUD Kaplan-Meier curve (Collett, S. 5)

| time | events | censored | Left | SURVIVAL | StdErr | SDF_LCL | SDF_UCL |
|------|--------|----------|------|----------|---------|---------|---------|
| 0 | . | . | 18 | 1.00000 | 0.00000 | 1.00000 | 1.00000 |
| 10 | 1 | 0 | 17 | 0.94444 | 0.05399 | 0.83863 | 1.00000 |
| 13 | 0 | 1 | 16 | . | . | . | . |
| 18 | 0 | 1 | 15 | . | . | . | . |
| 19 | 1 | 0 | 14 | 0.88148 | 0.07899 | 0.72667 | 1.00000 |
| 23 | 0 | 1 | 13 | . | . | . | . |
| 30 | 1 | 0 | 12 | 0.81368 | 0.09778 | 0.62204 | 1.00000 |
| 36 | 1 | 0 | 11 | 0.74587 | 0.11067 | 0.52896 | 0.96278 |
| 38 | 0 | 1 | 10 | . | . | . | . |
| 54 | 0 | 1 | 9 | . | . | . | . |
| 56 | 0 | 1 | 8 | . | . | . | . |
| 59 | 1 | 0 | 7 | 0.65264 | 0.13032 | 0.39721 | 0.90806 |
| 75 | 1 | 0 | 6 | 0.55940 | 0.14117 | 0.28272 | 0.83608 |
| 93 | 1 | 0 | 5 | 0.46617 | 0.14520 | 0.18158 | 0.75075 |
| 97 | 1 | 0 | 4 | 0.37293 | 0.14299 | 0.09267 | 0.65319 |
| 104 | 0 | 1 | 3 | . | . | . | . |
| 107 | 1 | 0 | 2 | 0.24862 | 0.13925 | 0.00000 | 0.52154 |
| 107 | 0 | 1 | 1 | 0.24862 | . | 0.00000 | 0.52154 |
| 107 | 0 | 1 | 0 | 0.24862 | . | 0.00000 | 0.52154 |

=====

9

=====

9 Biometrie\life.sas

- └ Standardfehler der Schätzungen von $S(t)$
- └ Punktwises Konfidenzintervall für $S(t)$

Konfidenzbänder für $S(t)$

Das punktweise Konfidenzband (zusammengesetzt aus den einzelnen punktweisen Konfidenzintervallen) hat nicht die korrekte Überdeckungswahrscheinlichkeit für die ganze Survivalfunktion.

Die bekanntesten simultanen Konfidenzbänder sind die sog. Hall-Wellner-Bänder. Die Herleitung ist aber schwierig und es existiert keine geschlossene Formel.

Siehe z.B. Klein/Moeschberger Kapitel 4.4.

Konfidenzintervalle für Median und Quantile

Die Konstruktion von approximativen Konfidenzintervallen für Quantile von $S(t)$ über die Normalapproximation von \hat{t}_p ist schwierig, da die asymptotische Varianz von \hat{t}_p von der Dichte $f(\hat{t}_p)$ abhängt.

Alternativ: Ein approximatives $1 - \alpha$ Konfidenzintervall für t_p ist die Menge aller Zeitpunkt t mit

$$\left| \frac{\hat{S}(t) - (1 - p)}{\sqrt{\hat{W}(\hat{S}(t))}} \right| \leq z_{1-\alpha/2} \quad (2.18)$$

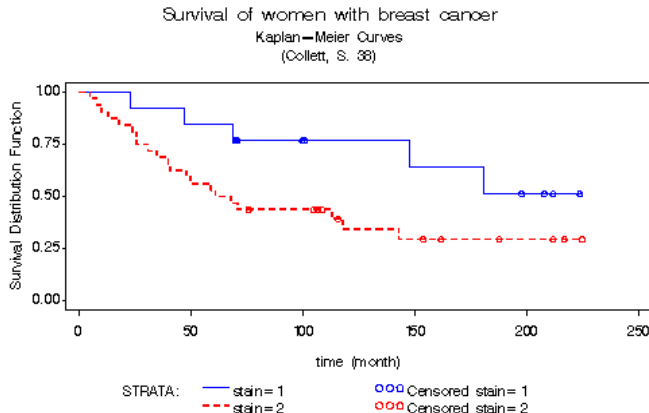
Diese Menge lässt sich einfach bestimmen, da $\hat{S}(t)$ nur endlich viele Werte annimmt.

→ Illustration!

Vergleich von zwei Survival-Kurven

Zunächst sollte man sich durch geeignete Plots einen optischen Eindruck verschaffen, und Unterschiede qualitativ beschreiben. → Unterschied?

Beispiel 2.6



Der Log-Rank-Test

Zwei-Gruppen-Vergleich (Gruppe 1 und 2)

$t_{(1)} < t_{(2)} < \dots < t_{(r)}$ – geordnete *Ereigniszeitpunkte* der kombinierten Stichprobe

d_{ij} – Anzahl Fälle zum Zeitpunkt $t_{(j)}$ in Gruppe i

n_{ij} – Anzahl unter Risiko zum Zeitpunkt $t_{(j)}$ in Gruppe i

Für jeden Failure-Zeitpunkt $t_{(j)}$:

Tabelle: Anzahl Fälle am j-ten Ereigniszeitpunkt nach Gruppe

| Gruppe | Anzahl „failures“ | Anzahl „survived“ | Anzahl „at risk“ |
|--------|----------------------|----------------------|---------------------|
| 1 | d_{1j} | $n_{1j} - d_{1j}$ | n_{1j} |
| 2 | d_{2j} | $n_{2j} - d_{2j}$ | n_{2j} |
| Gesamt | d_j | $n_j - d_j$ | n_j |

Betrachte H_0 : „Die beiden Survivalfunktionen sind gleich.“

Gültigkeit von H_0 wird überprüft, indem für jeden Failure-Zeitpunkt $t_{(j)}$ die beobachtete Anzahl der Failures mit der unter H_0 erwarteten Anzahl verglichen und schließlich aggregiert wird.

Betrachtet man die Ränder von Tabelle 3 als fix, so ist die Tabelle allein durch d_{1j} bestimmt. Unter H_0 ist D_{1j} eine (zentral) hypergeometrischverteilte Zufallsvariable mit

$$\Pr(D_{1j} = d_{1j} \mid \text{Rand}) = \frac{\binom{d_j}{d_{1j}} \cdot \binom{n_j - d_j}{n_{1j} - d_{1j}}}{\binom{n_j}{n_{1j}}}$$

Es ist

$$e_{1j} := \mathbb{E}(D_{1j}) = n_{1j} \cdot \frac{d_j}{n_j} \quad (2.19)$$

$$v_{1j} := \mathbb{V}(D_{1j}) = \frac{n_{1j} n_{2j} d_j (n_j - d_j)}{n_j^2 (n_j - 1)} \quad (2.20)$$

Zusammenfassen der Diskrepanz von „Observed“ und „Expected“ aus den r einzelnen 2×2 -Tafeln liefert

$$U_L = \sum_{j=1}^r (d_{1j} - e_{1j}) \quad \text{mit } \mathbb{E}(U_L) = 0, \quad (2.21)$$

$$V_L := \mathbb{V}(U_L) = \sum_{j=1}^r v_{1j}. \quad (2.22)$$

Unter H_0 ist

$$\frac{U_L}{\sqrt{V_L}} \stackrel{asy}{\sim} \mathcal{N}(0, 1),$$

wenn „hinreichend viele“ Ereignisse vorliegen. Es folgt dann (unter H_0):

$$T_L = \frac{U_L^2}{V_L} \stackrel{asy}{\sim} \chi_1^2.$$

- ▶ Der Test beruhend auf T_L heißt **Log-Rank-Test** (andere Namen: *Mantel-Cox*, *Peto-Mantel-Haenszel-Test*).
- ▶ Der Test ist optimal, wenn die sogenannte proportionale Hazards-Annahme gerechtfertigt ist:

$$\frac{\lambda_1(t)}{\lambda_2(t)} = \text{konstant}$$

notwendig: Survivalkurven dürfen sich nicht schneiden.

→ Diskussion!

Beispiel 2.7 (Brustkrebs-Daten)

Für diesen Datensatz ergeben sich $d_1 = \sum d_{1j} = 5$ und $e_1 = \sum e_{1j} = 9,565$. Man erhält folgende Teststatistiken:

Survival of women with breast cancer

Kaplan-Meier Curves

(Collett, S. 38)

| Obs | Test | ChiSq | DF | Prob ChiSq |
|-----|-----------|--------|----|---------------|
| 1 | Log-Rank | 3.5150 | 1 | 0.0608 |
| 2 | Wilcoxon | 4.1800 | 1 | 0.0409 |
| 3 | -2Log(LR) | 4.3563 | 1 | 0.0369 |

Biometrie\life.sas

Der Wilcoxon-Test

Der Wilcoxon-Test ist gleichfalls ein Test auf H_0 : „Die Survivalfunktionen der zwei Gruppen sind gleich“. Er basiert auf der Teststatistik

$$U_W = \sum_{j=1}^r n_j (d_{1j} - e_{1j})$$

mit e_{1j} definiert wie in (2.19).

Mit

$$V_W = \sum_{j=1}^r n_j^2 v_{1j}$$

ergibt sich die Wilcoxon-Statistik T_W als

$$T_W = \frac{U_W^2}{V_W}.$$

Unter H_0 ist $T_W \stackrel{asy}{\sim} \chi_1^2$.

Vergleich von Log-Rank- und Wilcoxon-Test

- ▶ Wilcoxon-Test legt mehr Gewicht auf frühere Zeitpunkte
- ▶ Log-Rank-Test „trennschärfer“ unter PH-Annahme.
- ▶ Es gibt eine Vielzahl an Abwandlungen dieser Tests. Allgemeine Form: gewichteter Log-Rank Test mit

$$U = \sum_{j=1}^r w_j (d_{1j} - e_{1j}) \quad V = \sum_{j=1}^r w_j^2 v_{1j}$$

und $T = U^2/V$ wobei verschiedene Gewichte w_j möglich sind.

Vergleich von mehr als zwei Survivalkurven

Sowohl der Log-Rank- wie der Wilcoxon-Test lassen sich auf den simultanen Vergleich von $G > 2$ Gruppen verallgemeinern. In der kombinierten Stichprobe betrachtet man dann eine Folge von $G \times 2$ -Tafeln. In Verallgemeinerung des Vorangehenden definiert man für $k = 1, \dots, G - 1$:

$$U_{Lk} = \sum_{j=1}^r \left(d_{kj} - n_{kj} \frac{d_j}{n_j} \right)$$
$$U_{Wk} = \sum_{j=1}^r n_j \left(d_{kj} - n_{kj} \frac{d_j}{n_j} \right),$$

$$\mathbf{U}_L^T = (U_{L1}, \dots, U_{L_{G-1}}), \mathbf{U}_W^T = (U_{W1}, \dots, U_{W_{G-1}}).$$

Für Varianz bzw. Kovarianz ergibt sich

$$V_{Lkk'} = \sum_{j=1}^r \frac{n_{kj}d_j(n_j - d_j)}{n_j(n_j - 1)} \left(\delta_{kk'} - \frac{n_{k'j}}{n_j} \right)$$

$$V_{Wkk'} = \sum_{j=1}^r n_j^2 \frac{n_{kj}d_j(n_j - d_j)}{n_j(n_j - 1)} \left(\delta_{kk'} - \frac{n_{k'j}}{n_j} \right)$$

Tests auf die Nullhypothese ergeben sich dann via

$$U_L^T V_L^{-1} U_L \text{ bzw. } U_W^T V_W^{-1} U_W.$$

Unter H_0 sind diese Statistiken χ_{G-1}^2 -verteilt.

Stratifizierte Tests

Eine Möglichkeit, Kovariable zu berücksichtigen: Schichte (stratifiziere) Daten bzgl. solcher Variablen (nur praktisch möglich, wenn 'wenige' Schichten).

Für solche Situationen gibt es den stratifizierten Log-Rank-Test, der auf den stratumspezifischen Testgrößen beruht. Für den Vergleich von zwei Gruppen lautet die Teststatistik

$$T_S = \frac{\left(\sum_{k=1}^S U_{Lk}\right)^2}{\sum_{k=1}^S V_{Lk}}, \quad (2.23)$$

wobei U_{lk} und V_{Lk} die entsprechenden Testgrößen in Stratum k gemäß Formeln (2.21) und (2.22) sind. Unter H_0 : „kein Survival-Unterschied in den zwei Gruppen“ ist $T_S \stackrel{asy}{\sim} \chi_1^2$.

Beispiel 2.8 (Survival von Melanom-Patienten)

Tabelle: Ereigniszeiten von Melanompatienten in zwei Behandlungsgruppen, stratifiziert nach Altersgruppe.

| 21-40 | | 41-60 | | 60+ | |
|-------|-----------|-------|-----------|-----|-----------|
| BCG | C. parvum | BCG | C. parvum | BCG | C. parvum |
| 19 | 27* | 34* | 8 | 10 | 25* |
| 24* | 21* | 4 | 11 | 5 | 8 |
| 8 | 18* | 17* | 23* | | 11* |
| 17* | 16* | | 12* | | |
| 17* | 7 | | 15* | | |
| 34* | 12* | | 8* | | |
| | 24 | | 8* | | |
| | 8 | | | | |
| | 8* | | | | |

Tabelle: Werte der Log-Rank Statistik für jede Altersgruppe.

| Altersgruppe | U_L | V_L | T_L |
|--------------|---------|--------|-------|
| 21–40 | –0,2571 | 1,1921 | 0,055 |
| 41–60 | 0,4778 | 0,3828 | 0,596 |
| 61– | 1,0167 | 0,6497 | 1,591 |
| Total | 1,2374 | 2,2246 | |

Daher: $T_S = 1,2374^2 / 2,2246 = 0,688$ (nicht signifikant).

Log-Rank-Test auf Trend

Der Log-Rank-Test auf Trend bei G geordneten Gruppen basiert auf

$$U_{\text{Trend}} = \sum_{k=1}^G w_k (d_{k+} - e_{k+}), \quad (2.24)$$

wobei w_k ein Score für die k -te Gruppe, $k = 1, \dots, G$, und

$$d_{k+} = \sum_{j=1}^{r_k} d_{kj}, \quad e_{k+} = \sum_{j=1}^{r_k} e_{kj}.$$
$$V_{\text{Trend}} = \sum_{k=1}^G (w_k - \bar{w})^2 e_{k+} \quad (2.25)$$

$$\text{mit } \bar{w} = \frac{\sum_{k=1}^G e_{k+} w_k}{\sum_{k=1}^G e_{k+}}.$$

Unter H_0 : „kein Trend“ hat die Teststatistik $T_{\text{Trend}} = U_{\text{Trend}}^2 / V_{\text{Trend}}$ eine χ_1^2 -Verteilung.

- ▶ Wahl der Scores w_k :
Meistens äquidistant, z.B. 1, 2, 3 oder -1, 0, 1.
- ▶ Alternative zum Trendtest:
Modellierung im Cox Modell.

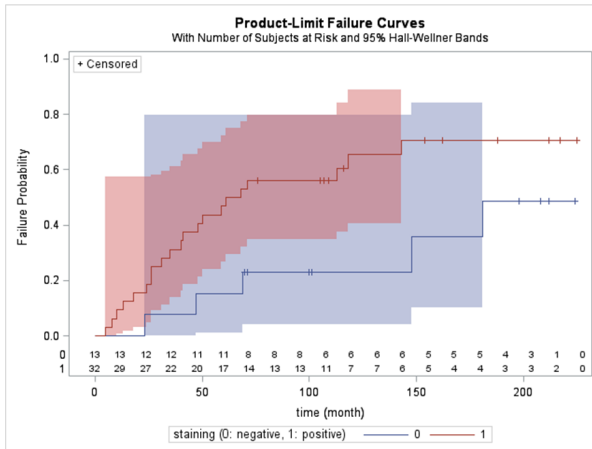
Darstellung von Survivalkurven

Aus: Pocock, Clayton, Altman: Survival plots of time-to-event outcomes in clinical trials: good practice and pitfalls. The Lancet, Vol 359, 2002.

- ▶ Auflistung der “number at risk” unterhalb der Zeitachse in jeder Gruppe;
- ▶ Maß für die stat. Unsicherheit (SE oder Konfidenzintervalle) angeben;
- ▶ Darstellung der geschätzten *kumulativen Inzidenzfunktion* $1 - \hat{S}(t)$ statt $\hat{S}(t)$ manchmal (seltene Ereignisse) besser.

Beispiel 2.9 (Inzidenz: Prognose bei Brustkrebs)

Inzidenzkurven mit Hall-Wellner-Bändern und “# at risk” (vgl. Bsp. 2.6).



Inhalt von Abschnitt 3

Modellierung: Das Cox Modell

Das Proportional Hazards Modell

Interpretation der Parameter im Cox Modell

Schätzen im Cox Modells

Begründung der Likelihoodfunktion

Bindungen (ties)

Konfidenzintervalle für β

Vergleich alternativer Modelle

Schätzen von Hazard- und Survivalfunktion

- ▶ Kaplan-Meier-Kurven, Log-Rank-Test: geeignet, wenige Gruppen miteinander zu vergleichen;
- ▶ Modellierung von Hazardfunktion: Berücksichtigung von mehreren Kovariablen.
- ▶ Welche Faktoren beeinflussen Ereignisrate?
- ▶ Schätzen der (individuellen) Hazardfunktion, um Survival-Prognose für zukünftige Patienten zu treffen.
- ▶ Kausale Interpretation aus verschiedenen Gründen problematisch — Alternative: marginales strukturelles Hazardmodell.

Das Cox Modell

Das **proportionale Hazards (PH) Modell** von Cox (1972): Veränderung von Kovariablen führt zu proportionaler Veränderung der Ereignisrate. Dabei: nicht weiter spezifizierte **Basis-Hazardfunktion** $\lambda_0(t)$ (semi-parametrischer Ansatz).

Einfachstes Beispiel: zwei Gruppen (Kontrolle = 0, Behandlung = 1):

$$\lambda_1(t) = \psi \lambda_0(t), \quad \psi > 0,$$

$\psi = \lambda_1(t)/\lambda_0(t)$ heißt Hazard Ratio (HR), manchmal auch “relative risk”.

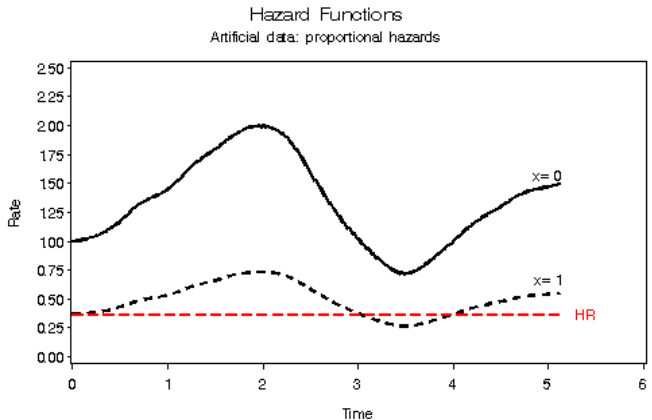
$$\psi : \begin{cases} < 1 \text{ d.h. } \lambda_1(t) < \lambda_0(t) \\ = 1 \text{ etc.} \\ > 1 \end{cases}$$

Es gilt auch: $\psi = \Pr(T_0 > T_1)/(1 - \Pr(T_0 > T_1))$.

—→ Interpretation!?

—→ Vergleiche Survivalfunktionen!?

Die Proportionale Hazards Annahme am Beispiel:



Biometrie, StatMod4, Examp4 → x.sas

Modellierung: n (unabh.) Patienten in Studie, $X_i \in \{0, 1\}$ Indikatorvariable für Behandlung

$$\lambda_i(t) = \lambda_0(t) \cdot \exp(\beta x_i).$$

d.h. hier $HR = \psi = \exp(\beta)$.

→ Warum $\psi(x_i) = \exp(\beta x_i)$?

Allgemein: Berücksichtigung von mehreren Kovariablen X_1, \dots, X_p :
 $\mathbf{x}_i^T = (x_{i1}, \dots, x_{ip})$ gemessen zu Zeitp. $t = 0$ (baseline).
Dann

$$\lambda_i(t) := \lambda(t; \mathbf{X} = \mathbf{x}_i) = \psi(\mathbf{x}_i) \lambda_0(t),$$

$\psi(\mathbf{x}_i)$ ist Hazard Ratio einer Person mit $\mathbf{X} = \mathbf{x}_i$ gegenüber jemanden mit $\mathbf{X} = \mathbf{0}$.

Definition 3.1 (Cox Modell)

Sei \mathbf{x}_i ein p -dimensionaler Kovariablenvektor, gemessen zum Zeitpunkt $t = 0$, und $\boldsymbol{\beta}$ ein entsprechender Parametervektor. Sei $\eta_i = \mathbf{x}_i^T \boldsymbol{\beta} = \beta_1 x_{i1} + \dots + \beta_p x_{ip}$ der lineare Prädiktor. Das proportionale Hazards (PH) Modell besagt:

$$\begin{aligned} \lambda_i(t) &= \lambda_0(t) \cdot \exp(\eta_i) \\ \iff \log \frac{\lambda_i(t)}{\lambda_0(t)} &= \beta_1 x_{i1} + \dots + \beta_p x_{ip}, \end{aligned} \tag{3.1}$$

(Bemerke: kein Intercept-Parameter).

→ Notation auch: $\lambda(t; \mathbf{x}) = \lambda_0(t) \cdot \exp(\mathbf{x}^T \boldsymbol{\beta})$.

Die lineare Komponente des Cox Modells

Modellierung analog zum generalisierten linearen Modell (GLM), d.h. linearer Prädiktor η kann umfassen

- ▶ Faktoren,
- ▶ stetige Kovariable,
- ▶ sowie Wechselwirkungsterme zwischen den Variablen.

Beispiel 3.2 (Prognose für Frauen mit Brustkrebs)

Zunächst: Survivalfunktionen (Bsp. 2.6) kreuzen sich nicht, d.h. PH Annahme nicht ganz abwegig.

„Staining“ wird modelliert via Indikatorvariable

$$X = \begin{cases} 1 & \text{positiv / rot / "stain 2"} \\ 0 & \text{negativ / blau / "stain 1"} \end{cases}$$

Cox Modell

$$\lambda_1(t) = \lambda_0(t) \exp(\beta)$$

Schätzung (siehe später) liefert:

$$\begin{aligned} \hat{\beta} &= 0,908 & \text{SE}(\hat{\beta}) &= 0,501; & 95\text{-KI} &= [-0,074; 1,890] \\ \exp(\hat{\beta}) &= 2,48 & \text{SE}(\exp(\hat{\beta})) &= 1,242; & 95\text{-KI} &= [0,93; 6,62] \end{aligned}$$

→ Interpretation!?

Interpretation bei nicht-binären Kovariablen

- ▶ eine (stetige) Kovariable X

$$\begin{aligned}\lambda(t; X = x) &= \lambda_0(t) \exp(\beta x) \\ \iff \frac{\lambda(t; x)}{\lambda_0(t)} &= \exp(\beta x) \quad (\text{HR})\end{aligned}$$

HR ist Vergleich zu $x = 0$ (sinnvoll?)

→ Vergleiche $\lambda(t; x)$ mit $\lambda(t; x^*)$!

β - „log rate ratio Parameter“

- ▶ Faktor F mit $G + 1$ Faktorstufen

$$\lambda(t; F = j) = \lambda_0(t) \exp(\gamma_j), \quad j = 1, \dots, G,$$

mit $\gamma_0 = 0$ (Referenzkategorie).

Andere Parameterisierung für Faktoren möglich (s. StatMod-II).

Beispiel 3.3 (Nierenkarzinom-Patienten)

Survival of Kidney Cancer Patients
(Collett, p. 91)

| Variable | Estimate | StdErr | ChiSq | Label |
|----------|----------|---------|--------|------------------------------|
| Age2 | -0.06457 | 0.49843 | 0.0168 | Age=2 (≥ 60 , < 70) |
| Age3 | 1.82448 | 0.68184 | 7.1600 | Age=3 (≥ 70) |

andere Parametrisierung;

| | | | | |
|------|---------|---------|--------|---------------------|
| Age1 | 0.06457 | 0.49843 | 0.0168 | Age=1 (< 60) |
| Age3 | 1.88905 | 0.73954 | 6.5247 | Age=3 (≥ 70) |

Biometrie\StatMod4\Examp3-8.sas

► mehrere Variablen

Es gelten dieselben Prinzipien wie in anderen multivariaten Regressionsmodellen: ist $\eta = x_1\beta_1 + \dots + x_p\beta_p$, so stellt β_j den HR-Parameter für Variable j dar, gegeben alle anderen Variablen sind gleich.

→ Vergleiche $\lambda(t; \mathbf{x})$ mit $\lambda(t; \mathbf{x}^*)$!

Anders als in linearen Modellen, hilft die Unabhängigkeit von X_1, \dots, X_p nicht bei der Interpretation der Parameter, da das Cox Modell “nicht kollabierbar” (non-collapsible) ist.

Sprich, auch wenn X_1 randomisiert wurde, sind β_1 im Modell mit $\eta = x_1\beta_1 + \dots + x_p\beta_p$ und $\tilde{\beta}_1$ im Modell mit $\tilde{\eta} = x_1\tilde{\beta}_1$ nicht notwendig gleich.

⇒ Im Cox Modell müssen die β_j *immer* im Kontext der mit einbezogenen Kovariablen interpretiert werden (Ford et al., 1995).

Beispiel 3.4 (Prostata-Krebs-Patienten)

Survival of Prostatic Cancer Patients (Collett, p. 95)

| Variable | Estimate | StdErr | Hazard Ratio | Label |
|----------|----------|---------|-----------------|--------------------------------|
| treat | -1.11276 | 1.20312 | 0.329 | treatment (2: DES, 1: placebo) |
| size | 0.08257 | 0.04746 | 1.086 | tumor size |
| index | 0.71022 | 0.33790 | 2.034 | Gleason index |

----- Biometrie\StatMod4\Examp3-10.sas -----

Beispiel 3.5 (Nierenkrebs-Patienten)

Survival of Kidney Cancer Patients
(Collett, p. 96)

08:04 Tuesday, March 18, 2008

| Variable | Estimate | StdErr | Hazard Ratio | Prob ChiSq | Label |
|-------------|----------|---------|-----------------|---------------|--------------------|
| Nephrectomy | -1.94325 | 0.73052 | 0.143 | 0.0078 | Nephrectomy (0=No) |
| Age2 | 0.00548 | 0.83489 | 1.005 | 0.9948 | Age=2 (>=60, <70) |
| Age3 | 0.06513 | 1.17737 | 1.067 | 0.9559 | Age=3 (>=70) |
| I_N_Age2 | -0.05114 | 0.97067 | 0.950 | 0.9580 | age2*nephrectomy |
| I_N_Age3 | 2.00299 | 1.34276 | 7.411 | 0.1358 | age3*nephrectomy |

Biometrie\StatMod4\Examp3-11.sas

Das geschätzte Modell ist

$$\eta = \beta N + \gamma_2 A_2 + \gamma_3 A_3 + \delta_2 (N A_2) + \delta_3 (N A_3)$$

mit N als Indikator für Treatment (Nephrectomy = 1).

In diesem Fall lassen sich die Hazard Ratios übersichtlich in folgender Tabelle anordnen.

| | Nephrectomy | |
|---------|----------------|-------------------------------|
| | 0 | 1 |
| Age = 1 | 1 | e^{β} |
| Age = 2 | e^{γ_2} | $e^{\beta+\gamma_2+\delta_2}$ |
| Age = 3 | e^{γ_3} | $e^{\beta+\gamma_3+\delta_3}$ |

—→ Interpretation!?

Grafische Überprüfung der PH-Annahme

Es gilt:

$$S(t|\mathbf{X} = \mathbf{x}) = \exp(-\Lambda_0(t) \exp(\boldsymbol{\beta}^T \mathbf{x}))$$

—→ Warum?

Es folgt (siehe 2-24):

$$\log[-\log S(t|\mathbf{X} = \mathbf{x})] = \log[-\log S_0(t)] + \boldsymbol{\beta}^T \mathbf{x}.$$

Das bedeutet, der Abstand zwischen den beiden Kurven

$$t \mapsto \log[-\log S(t|\mathbf{X} = \mathbf{x})]$$

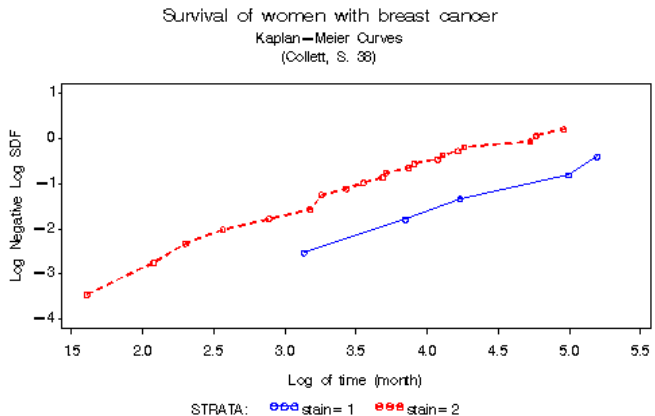
und

$$t \mapsto \log[-\log S_0(t)]$$

ist konstant, d.h. die beiden Kurven sind *parallel*.

⇒ **Visuelle Überprüfung** bei diskreten Kovariablen: wende log-log Transformation auf $\hat{S}(t)$ oder $\tilde{S}(t)$ in den verschiedenen Gruppen an.

Beispiel 3.6



Biometrie\life.sas

Rechtfertigung der PH-Annahme

Typische Situation: Interesse vor allem an Effekt einer (binären) Behandlungsvariable X_1 .

Auch wenn X_1 randomisiert wurde, ist es in der Regel sinnvoll zusätzliche Kovariablen X_2, \dots, X_p in das Modell aufzunehmen, damit die PH-Annahme plausibel ist.

Denn: auch wenn PH bedingt auf Kovariablen gilt, ist dies nicht notwendig marginal (ohne X_2, \dots, X_p) der Fall (siehe Ford et al., 1995).

Die Umkehrung (zu viele Kovariablen) kann auch ein Problem sein.

Problem der “non-collapsibility”!

Schätzen im Cox Modells

Zu schätzen sind

- ▶ $\beta^T = (\beta_1, \dots, \beta_p)$
- ▶ $\lambda_0(t)$.

Es stellt sich heraus, dass β und $\lambda_0(t)$ *getrennt* geschätzt werden können.

Gegeben seien Daten von n (unabh.) Probanden mit r beobachteten Ereigniszeiten und $(n - r)$ (uninformativ) rechtszensierten Zeiten. Dann ist die relevante (*partielle*) Likelihoodfunktion für das Cox Modell (3.1) gegeben als

$$L(\beta) = \prod_{j=1}^r \frac{\exp(\beta^T \mathbf{x}_{(j)})}{\sum_{l \in R(t_{(j)})} \exp(\beta^T \mathbf{x}_l)}, \quad (3.2)$$

wobei $t_{(1)} < \dots < t_{(r)}$ die geordneten Ereigniszeiten (ohne Bindungen) und $R(t_{(j)})$ das „Risk Set“ zum Zeitpunkt $t_{(j)}$.

- ▶ Produkt allein über die Zeitpunkte, an denen Ereignis stattfindet;
- ▶ nur Beiträge von Personen tauchen im Zähler auf, bei denen Ereignis beobachtet wird;
- ▶ zensierte Beobachtungen: Beiträge allein zum Nenner;
- ▶ Inferenz über β hängt vom Rang der Ereigniszeiten ab,
- ▶ (3.2) ist eine sogenannte partielle (engl.: partial) Likelihood.
- ▶ $\lambda_0(t)$ nicht “relevant” für Schätzung von β , da Proportionalität die Zeitabhängigkeit “entfernt”.
- ▶ Schätzen von β : Maximieren der Log-Partial Likelihood:

$$\log L(\beta) = \sum_{j=1}^r \left\{ \beta^T \mathbf{x}_{(j)} - \log \sum_{l \in R(t_j)} \exp(\beta^T \mathbf{x}_l) \right\}$$

Führt auf Lösen der Score-Geichung

$$\begin{aligned} U(\boldsymbol{\beta}) &= \sum_{j=1}^r \left\{ \mathbf{x}_{(j)} - \frac{\sum_{l \in R(t_j)} \mathbf{x}_l \exp(\boldsymbol{\beta}^T \mathbf{x}_l)}{\sum_{l \in R(t_j)} \exp(\boldsymbol{\beta}^T \mathbf{x}_l)} \right\} \\ &= \sum_{j=1}^r \left\{ \mathbf{x}_{(j)} - \sum_{l \in R(t_j)} w_l \mathbf{x}_l \right\} \stackrel{!}{=} 0, \end{aligned}$$

mit Gewichten

$$w_l = \frac{\exp(\boldsymbol{\beta}^T \mathbf{x}_l)}{\sum_{l' \in R(t_j)} \exp(\boldsymbol{\beta}^T \mathbf{x}_{l'})}.$$

Alternative Formulierung:

n Personen mit beobachteten Zeiten t_1, \dots, t_n und Ereignis-Indikatoren

$$\delta_i = \begin{cases} 1, & \text{falls Ereignis beobachtet,} \\ 0 & \text{sonst (Zensierung).} \end{cases}$$

(3.2) kann dann geschrieben werden als

$$L(\boldsymbol{\beta}) = \prod_{i=1}^n \left\{ \frac{\exp(\boldsymbol{\beta}^T \mathbf{x}_i)}{\sum_{l \in R(t_i)} \exp(\boldsymbol{\beta}^T \mathbf{x}_l)} \right\}^{\delta_i},$$

wobei $R(t_i)$ das Riskset zum Zeitpunkt t_i darstellt.

$$\log L(\boldsymbol{\beta}) = \sum_{i=1}^n \delta_i \left\{ \boldsymbol{\beta}^T \mathbf{x}_i - \log \sum_{l \in R(t_i)} \exp(\boldsymbol{\beta}^T \mathbf{x}_l) \right\}$$

Begründung der Likelihoodfunktion

Basis der Likelihood-Konstruktion für das Cox Modell: die Intervalle zwischen den Ereigniszeitpunkten haben keine Information über β (da $\lambda_0(t)$ beliebig).

Zu bestimmen: die Wahrscheinlichkeit dafür, dass Ereigniszeit für Individuum i gleich $t_{(j)}$ ist, bedingt darauf, dass $t_{(j)}$ einer der beobachteten Ereigniszeitpunkte:

$$\begin{aligned} & \Pr\{\text{Zeit(für Person mit } x_{(j)}) = t_{(j)} \mid \text{ein Ereignis in } t_{(j)}\} \\ &= \frac{\Pr\{\text{Zeit(für Person mit } x_{(j)}) = t_{(j)}\}}{\Pr\{\text{ein Ereignis in } t_{(j)}\}} \end{aligned}$$

Da die Ereigniszeiten als unabhängig angenommen werden

$$\begin{aligned}
 &= \frac{\Pr\{\text{Zeit(für Person mit } x_{(j)}) = t_{(j)}\}}{\sum_{l \in R(t_{(j)})} \Pr\{\text{Zeit (für Person } l) = t_{(j)}\}} \\
 &= \lim_{\Delta t \rightarrow 0} \frac{\Pr\{T(x_{(j)}) \in [t_{(j)}, t_{(j)} + \Delta t)\} / \Delta t}{\sum_{l \in R(t_{(j)})} \Pr\{T(x_l) \in [t_{(j)}, t_{(j)} + \Delta t)\} / \Delta t} \\
 &= \frac{\lambda_i(t_{(j)})}{\sum_{l \in R(t_{(j)})} \lambda_l(t_{(j)})}
 \end{aligned}$$

(unter der Annahme, dass Person i im Zeitpunkt $t_{(j)}$ Ereignis hat). Dieser Ausdruck reduziert sich zu

$$= \frac{\exp(\beta^T \mathbf{x}_{(j)})}{\sum_{l \in R(t_{(j)})} \exp(\beta^T \mathbf{x}_{(l)})},$$

Multiplizieren über alle Ereigniszeitpunkte liefert (3.2).

Bindungen (ties)

- ▶ Hazardfunktion des Cox Modells im Prinzip stetig;
- ▶ durch Rundung der Zeiten entstehen Bindungen (engl. ties);
- ▶ falls Zensierungszeiten und Ereigniszeiten gleich, werden Zensierungszeiten als größer angenommen.

Approximationen für „tied observations“:

Angenommen es gäbe d_j Ereignisse in $t_{(j)}$. Sei $s_j = \sum_{l \in D(t_{(j)})} x_l$, wobei $D(t_{(j)})$ die entsprechende Gruppe von Personen mit Ereignis sei,

- ▶ nach Breslow
- ▶ nach Efron
- ▶ nach Cox
- ▶ — hier nur der Vollständigkeit halber:

Breslow-Approximation:

$$L(\beta) = \prod_{j=1}^r \frac{\exp(\beta^T \mathbf{s}_j)}{\left\{ \sum_{l \in R(t_{(j)})} \exp(\beta^T \mathbf{x}_l) \right\}^{d_j}}$$

(gute Approximation, falls wenige Bindungen; schnell zu berechnen, deshalb Standardeinstellung in Software).

Efron-Approximation:

$$L(\beta) =$$

$$\prod_{j=1}^r \frac{\exp(\beta^T \mathbf{s}_j)}{\prod_{k=1}^{d_j} \left[\sum_{l \in R(t_{(j)})} \exp(\beta^T \mathbf{x}_l) - \frac{k-1}{d_j} \sum_{l \in D(t_{(j)})} \exp(\beta^T \mathbf{x}_l) \right]}$$

(bessere Approximation an Likelihoodfunktion, in der Regel ähnlich zur Breslow-Approximation).

Cox-Approximation:

$$L(\beta) = \prod_{j=1}^r \frac{\exp(\beta^T s_j)}{\sum_{l \in R(t_{(j)}; d_j)} \exp(\beta^T s_l)},$$

wobei $R(t_{(j)}; d_j)$ eine Auswahl von d_j Personen aus $R(t_{(j)})$ bezeichnet; die Summation im Nenner ist über alle solche Auswahlen. Diese Approximation beruht auf einem diskreten Hazard Modell.

- ▶ Ohne Bindungen in den Daten reduzieren sich alle Approximationen auf (3.2).

Konfidenzintervalle für β

- ▶ Wie normaler ML-Schätzer hat auch $\hat{\beta}$ asy. Normalverteilung.
- ▶ Die Varianz von $\hat{\beta}$ aus partieller Likelihood wird geschätzt über die Inverse der Informationsmatrix (siehe ML-Schätzung, GLMs).

$$I(\beta)_{jk} = -\frac{\partial^2 \log L(\beta)}{\partial \beta_j \partial \beta_k}, \quad j, k = 1, \dots, p.$$

$$\hat{V}(\hat{\beta}) \approx I^{-1}(\hat{\beta}).$$

- ▶ Wald-Test zum Prüfen von $H_0: \beta_j = 0$ vs. $H_1: \beta_j \neq 0$: lehne H_0 ab, falls $|\hat{\beta}_j / \text{SE}(\hat{\beta}_j)| > u_{1-\frac{\alpha}{2}}$, wobei $u_{1-\frac{\alpha}{2}}$ das Quantil von $\mathcal{N}(0, 1)$.
- ▶ Konfidenz-Intervall für β_j :

$$\hat{\beta}_j \pm \text{SE}(\hat{\beta}_j) \cdot u_{1-\frac{\alpha}{2}}.$$

- Konfidenz-Intervall für Hazard Ratio $\exp(\beta_j)$:

$$\exp\{\hat{\beta}_j \pm \text{SE}(\hat{\beta}_j) \cdot u_{1-\frac{\alpha}{2}}\}$$

- Standardfehler für Hazard Ratio (δ -Methode):

$$\text{SE}(\exp(\hat{\beta}_j)) = \text{SE}(\hat{\beta}_j) \cdot \exp(\hat{\beta}_j)$$

Beispiel 3.7 (Prognose für Frauen mit Brustkrebs)

Schon überprüft: PH Annahme nicht ganz abwegig.

„Staining“ wird modelliert via Indikatorvariable

$$X = \begin{cases} 1 & \text{positiv / rot / "stain 2"} \\ 0 & \text{negativ / blau / "stain 1"} \end{cases}$$

Cox Modell:

$$\lambda_1(t) = \lambda_0(t) \exp(\beta)$$

Schätzung mit partieller Likelihood liefert:

$$\begin{aligned} \hat{\beta} &= 0,908 & \text{SE}(\hat{\beta}) &= 0,501; & 95\text{-KI} &= [-0,074; 1,890] \\ \exp(\hat{\beta}) &= 2,48 & \text{SE}(\exp(\hat{\beta})) &= 1,242; & 95\text{-KI} &= [0,93; 6,62] \end{aligned}$$

Beispiel 3.8 (Survival von Patienten mit multiplem Myelom)

Survival of multiple myeloma patients
(Collett, p. 72)

-----Alphabetic List of Variables and Attributes-----

| # | Variable | Type | Len | Pos | Label |
|----|----------|------|-----|-----|--------------------------------|
| 4 | age | Num | 8 | 24 | age (years) |
| 6 | bun | Num | 8 | 40 | blood urea nitrogen |
| 7 | ca | Num | 8 | 48 | serum calcium |
| 8 | hb | Num | 8 | 56 | haemoglobin |
| 1 | patient | Num | 8 | 0 | |
| 9 | pccells | Num | 8 | 64 | plasma cells in bonemarrow (%) |
| 10 | protein | Num | 8 | 72 | B-J-protein in urine (0: no) |
| 5 | sex | Num | 8 | 32 | sex (1: Male, 2: Female) |
| 3 | status | Num | 8 | 16 | death (0: censored) |
| 2 | time | Num | 8 | 8 | time (month) |

Biometrie\StatMod4\Examp3-2.sas

Beispiel 3.8 (Survival von Patienten mit multiplem Myelom – Forts.)

Survival of multiple myeloma patients
(Collett, p. 72)

| Variable | Estimate | StdErr | Prob ChiSq | Label |
|----------|----------|---------|---------------|--------------------------------|
| age | -0.01936 | 0.02792 | 0.4882 | age (years) |
| sex | -0.25090 | 0.40229 | 0.5328 | sex (1: Male, 2: Female) |
| bun | 0.02083 | 0.00593 | 0.0004 | blood urea nitrogen |
| ca | 0.01312 | 0.13244 | 0.9211 | serum calcium |
| hb | -0.13524 | 0.06889 | 0.0496 | haemoglobin |
| pcells | -0.00159 | 0.00658 | 0.8085 | plasma cells in bonemarrow (%) |
| protein | -0.64044 | 0.42669 | 0.1334 | B-J-protein in urine (0: no) |

Biometrie\StatMod4\Examp3-2.sas

Vergleich alternativer Modelle - LQ-Test

- ▶ Man betrachte geschachtelte (hierarchische) Modelle

Modell 1: $\eta = x_1\beta_1 + \dots + x_p\beta_p$

Modell 2: $\eta = x_1\beta_1 + \dots + x_p\beta_p + \dots + x_{p+q}\beta_{p+q}$

Frage: sind die q zusätzlichen Variablen nötig?

- ▶ Partial-Likelihood-Quotienten-Test (auch: Likelihood-Ratio-Test)

Seien \hat{L}_1, \hat{L}_2 die partiellen Likelihoodfunktionen (3.2) nach Einsetzen der geschätzten Parameter unter Modell 1 und 2.

Unter H_0 : „ $\beta_{p+1} = \dots = \beta_{p+q} = 0$ “ ist

$$X^2 = -2 \log \frac{\hat{L}_1}{\hat{L}_2} \stackrel{\text{asy}}{\sim} \chi^2_q.$$

Survival of Kidney Cancer Patients (Collett, p. 77)

| | | | |
|----------------|--------------------------------|-----------------------|----|
| Data Set Name: | COLL.NEPHROMA | Observations: | 36 |
| Member Type: | DATA | Variables: | 4 |
| Engine: | V8 | Indexes: | 0 |
| Created: | 18:08 Thursday, March 13, 2008 | Observation Length: | 32 |
| Last Modified: | 18:08 Thursday, March 13, 2008 | Deleted Observations: | 0 |
| Protection: | | Compressed: | NO |
| Data Set Type: | | Sorted: | NO |
| Label: | DataCollett.sas, 13MAR08 | | |

| # | Variable | Type | Len | Pos | Label |
|-------|-------------|------|-----|-----|---------------------|
| ----- | | | | | |
| 2 | Age | Num | 8 | 8 | Age group |
| 1 | Nephrectomy | Num | 8 | 0 | Nephrectomy (0=No) |
| 4 | Status | Num | 8 | 24 | death (0: censored) |
| 3 | Time | Num | 8 | 16 | time (months) |

----- Biometrie\StatMod4\Examp3-4.sas -----

Survival of Kidney Cancer Patients (Collett, p. 77)

| Criterion | Without Covariates | With Covariates | Variable |
|-----------|-----------------------|--------------------|--|
| -2 LOG L | 177.667 | 172.172 | age2 age3 |
| -2 LOG L | 177.667 | 170.247 | nephrectomy |
| -2 LOG L | 177.667 | 165.508 | nephrectomy age2 age3 |
| -2 LOG L | 177.667 | 162.479 | nephrectomy age2 age3 i_n_age2 i_n_age3 |

----- Biometrie\StatMod4\Examp3-4.sas -----

Vergleich alternativer Modelle

Andere Ansätze möglich

- ▶ Wald-Test
- ▶ Score-Test
- ▶ andere Kriterien: AIC oder BIC

Schätzen von Hazard- und Survivalfunktion

Motivation

Cox-Modell führt 'nur' zur Schätzung von HR, d.h. relatives Maß für den Vergleich von z.B. Patienten mit unterschiedlichen Kovariablen. Die eigentlichen Überlebenschancen bzw. Dauer zum Ereignis ist daraus nicht zu ersehen.

Beispiel: geschätzte HR für Nierenkrebspatienten mit Nephrectomy ist 0.244 im Vgl. zu Kontrollgruppe. Aber wie ist die Wahrscheinlichkeit, ein / zwei oder drei Jahre zu überleben?

—→ Diskussion.

Schätzen von Hazard- und Survivalfunktion (Forts.)

Interpretation HR

Beachte auch: HR ist ein Maß, das auf $T \geq t$ *bedingt*; dies erschwert die Interpretation auch bei randomisierten Studien: Randomisierung stellt zwar zu $t = 0$ sicher, dass Behandlungs- und Kontrollgruppe vergleichbar sind; aber wenn es außer der Behandlung andere Faktoren gibt, die die Dauer zum Ereignis (Überleben) beeinflussen (und die gibt es immer), dann sind Probanden mit $T \geq t > 0$ nicht mehr vergleichbar, und zwar i.d.R. umso weniger je größer t .

Dies ist nicht weitläufig bekannt und HR werden oft falsch interpretiert!

—→ Diskussion.

Ausweg:

betrachte andere Maße als HR, die sich auf die Survivalfunktion beziehen.

Aber: dafür muss Baselinehazard geschätzt werden.

Schätzen von Hazard- und Survivalfunktion (Forts.)

Da

$$S(t|\mathbf{X} = \mathbf{x}) = \exp(-\Lambda_0(t) \exp(\boldsymbol{\beta}^T \mathbf{x}))$$

und partielle Likelihood Schätzer für $\boldsymbol{\beta}$ liefert, muss noch baseline kumulative Hazardfunktion $\Lambda_0(t)$ geschätzt werden.

Breslow-Schätzer $\hat{\Lambda}_0(t)$:

$$\hat{\Lambda}_0(t) = \sum_{t_{(j)} \leq t} \frac{d_j}{\sum_{l \in R(t_{(j)})} \exp(\hat{\boldsymbol{\beta}}^T \mathbf{x}_l)}, \quad (3.3)$$

⇒ Beide Schätzer einsetzen liefert

$$\hat{S}(t|\mathbf{X} = \mathbf{x}) = \exp(-\hat{\Lambda}_0(t) \exp(\hat{\boldsymbol{\beta}}^T \mathbf{x})) = \hat{S}_0(t)^{\exp(\hat{\boldsymbol{\beta}}^T \mathbf{x})}$$

Schätzen von Hazard- und Survivalfunktion (Forts.)

- ▶ $\hat{S}(t|\mathbf{X} = \mathbf{x})$ kann nun z.B. für verschiedene Werte von \mathbf{x} geplottet werden (möglichst mit KIs).
Nachteil: wenn \mathbf{x} teilweise stetig oder viele Dimensionen, ist es schwierig, einen Überblick zu erhalten.
- ▶ Wenn keine Kovariablen vorhanden sind (d.h. $\beta = 0$), reduziert sich der Breslow-Schätzer (3.3) auf den Nelson-Aalen Schätzer

$$\tilde{\Lambda}(t) = \sum_{j:t(j) \leq t} \frac{d_j}{n_j}.$$

- ▶ Ein ML-basierter Ansatz $\lambda_0(t)$ und somit $S(t|\mathbf{X} = \mathbf{x})$ zu schätzen ähnelt dagegen dem Kaplan-Meier Prinzip (Kalbfleisch & Prentice).

Beispiel 3.9 (Nierenkrebspatienten)

Angepasst wird ein Haupteffekt-Modell (vgl. Bsp. 3.5 mit Interaktionen)

$$\eta = \beta_1 N + \beta_2 Age_2 + \beta_3 Age_3$$

Survival of Kidney Cancer Patients
(Collett, p. 103)

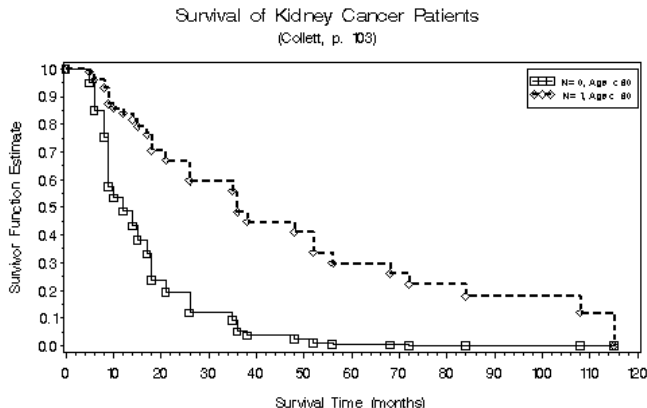
| Variable | Estimate | StdErr | Hazard Ratio | Prob ChiSq | Label |
|----------|----------|---------|-----------------|---------------|-------------------|
| N | -1.41176 | 0.51523 | 0.244 | 0.0061 | Nephrectomy |
| Age2 | 0.01239 | 0.42460 | 1.012 | 0.9767 | Age=2 (>=60, <70) |
| Age3 | 1.34156 | 0.59176 | 3.825 | 0.0234 | Age=3 (>=70) |

Biometrie\StatMod4\Examp3-12.sas

| Obs | N | Age2 | Age3 | Time | s | ls | lls | patt |
|-----|---|------|------|------|---------|----------|----------|------|
| 1 | 0 | 0 | 0 | 0 | 1.00000 | 0.00000 | . | 1 |
| 2 | 0 | 0 | 0 | 5 | 0.95023 | -0.05105 | -2.97499 | 1 |
| 3 | 0 | 0 | 0 | 6 | 0.85154 | -0.16071 | -1.82815 | 1 |
| 4 | 0 | 0 | 0 | 8 | 0.75516 | -0.28083 | -1.27000 | 1 |
| 5 | 0 | 0 | 0 | 9 | 0.57607 | -0.55153 | -0.59505 | 1 |
| . | | | | | | | | |
| . | | | | | | | | |
| 73 | 1 | 0 | 0 | 0 | 1.00000 | 0.0000 | . | 4 |
| 74 | 1 | 0 | 0 | 5 | 0.98764 | -0.0124 | -4.38674 | 4 |
| 75 | 1 | 0 | 0 | 6 | 0.96159 | -0.0392 | -3.23991 | 4 |
| 76 | 1 | 0 | 0 | 8 | 0.93385 | -0.0684 | -2.68176 | 4 |
| 77 | 1 | 0 | 0 | 9 | 0.87423 | -0.1344 | -2.00681 | 4 |
| . | | | | | | | | |
| . | | | | | | | | |

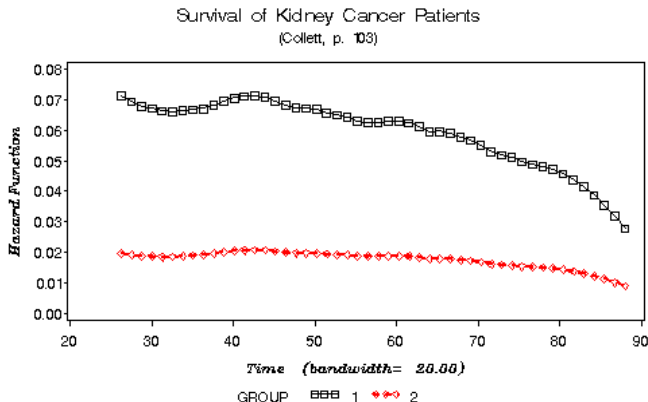
 Biometrie\StatMod4\Examp3-12.sas

Es läßt sich ersehen, dass in dieser Altersgruppe die Wahrscheinlichkeit 3 Jahre zu überleben mit Nephrectomy ca. 50% ist, aber ohne unter 10%.



Biometrie, StatMod4, Examp3 – 12 sas

Beachte: $\hat{\Lambda}_0(t)$ und $\hat{S}(t; \mathbf{x})$ Stufenfunktionen. D.h. $\hat{\lambda}(t; \mathbf{x})$ wird durch Glättung erhalten (hier nicht näher behandelt).



Biometrie, StatMod4, Examp3 – 12 sas

Inhalt von Abschnitt 4

Residuenanalyse

- Residuen im Cox Modell

- Martingal-Residuen

- Deviance-Residuen

 - Plots mit Martingal- und Deviance-Residuen

 - Funktionale Form der Kovariablen

- Schoenfeld-Residuen

 - Zeitabhängigkeit von Effekten prüfen

- Score-Residuen

 - Einfluss auf Parameterschätzung

 - Einfluss auf Gesamtfit

Model Check

Residuen in der Überlebenszeitanalyse etwas anders als vom linearen Modell her bekannt (z.B. nicht symmetrisch verteilt). Überlebenszeitmodelle haben ja keine 'additiven Fehlerterme' sondern sind grundsätzlich anders aufgebaut. Was sich untersuchen lässt ist

- ▶ angemessene funktionale Form der Kovariablen (z.B. x_j oder $\log(x_j)$ oder x_j^2 etc.);
- ▶ gibt es zeitabhängige Effekte (Verletzung der PH-Annahme);
- ▶ Ausreißer / einflussreiche Beobachtungen.

Model Check (Forts.)

Verschiedene Residuen

- ▶ Martingal-Residuen
- ▶ Deviance-Residuen
- ▶ Schoenfeld-Residuen
- ▶ Score-Residuen

Warum so viele?

Es gibt verschiedene Sichtweisen dazu, wie man Abweichungen vom Modell definiert...

Residuen im Cox Modell

- ▶ **Voraus.**: Von n Probanden werden r Ereigniszeiten gemessen, die restlichen $(n - r)$ Zeitpunkte sind uninformativ (rechts)-zensierte Zeiten; zugrunde liegen erklärende Variable $\mathbf{X}_i = (X_{i1}, \dots, X_{ip})^T$.
- ▶ Verschiedene Residuen-Typen basieren auf:
Breslow-Schätzer $\hat{\Lambda}_0(t)$, siehe (3.3), so dass

$$\begin{aligned}\hat{\Lambda}(t|\mathbf{X}_i = \mathbf{x}_i) &= \hat{\Lambda}_0(t) \exp(\hat{\boldsymbol{\beta}}^T \mathbf{x}_i) \\ &= \hat{\Lambda}_0(t) \exp(\hat{\eta}_i)\end{aligned}$$

Martingal-Residuen

Die Martingal-Residuen r_{M_i} sind definiert als (t_i ist der Ereignis- oder Zensierungszeitpunkt für Person i)

$$r_{M_i} = \delta_i - \hat{\Lambda}(t_i; \mathbf{x}_i). \quad (4.1)$$

- ▶ $r_{M_i} \in (-\infty, 1)$, sehr schief verteilt;
für zensierte Beobachtungen sind $r_{M_i} \leq 0$;
- ▶ $\sum_{i=1}^n r_{M_i} = 0$;
- ▶ $\text{Cov}(r_{M_i}, r_{M_j}) < 0, i \neq j$, aber in großen Stichproben sind r_{M_i} approx. unkorreliert.

Martingal-Residuen (Forts.)

- ▶ Man kann die Martingal-Residuen interpretieren als Differenz von Anzahl(b beobachtete Ereignisse bis t_i) und Anzahl(erwartete Ereignisse bis t_i) bei Individuum i .
- ▶ $r_{M_i} \ll 0$ deuten auf Beobachtungen mit sehr langen Ereigniszeiten,
- ▶ $r_{M_i} > 0$ auf solche mit kurzen Zeiten.
- ▶ Bemerkung: $\hat{\Lambda}(t_i; \mathbf{x}_i) = r_{c_i}$ sind die sogenannten Cox-Snell-Residuen, so dass $r_{M_i} = \delta_i - r_{c_i}$. Die r_{c_i} sind bei unzensierten Daten, wenn das Modell korrekt ist, exponentialverteilt mit Rate 1.

Deviance-Residuen

Die Deviance-Residuen sind definiert als

$$r_{D_i} = \text{sgn}(r_{M_i}) [-2\{r_{M_i} + \delta_i \log(\delta_i - r_{M_i})\}]^{\frac{1}{2}} \quad (4.2)$$

- ▶ Die Deviance-Residuen sind Bestandteil der Deviance via

$$D = \sum r_{D_i}^2$$
$$D = -2\{\log \hat{L}_C - \log \hat{L}_f\},$$

wobei \hat{L}_C die geschätzte partielle Likelihood unter dem aktuellen Modell, \hat{L}_f der entsprechende Wert unter dem „vollen Modell“.

- ▶ Die r_{D_i} sind unter dem korrekten Modell *symmetrisch* um Null verteilt (Vorteil gegenüber r_{M_i}), summieren sich jedoch nicht notw. zu Null.
- ▶ Deviance-Prinzip: siehe auch GLMs!

Plots mit Martingal- und Deviance-Residuen

Durch Index-Plots mit r_{M_i} bzw. r_{D_i} kann man Individuen mit relativ langen bzw. kurzen Ereigniszeiten identifizieren (Ausreißer).

- ▶ Aufgrund der schiefen Verteilung der Martingal-Residuen eignen sich Deviance-Residuen besser zur Diagnostik.
- ▶ Beobachtungen mit r_{M_i} nahe 1 haben 'zu kurze' Ereigniszeiten als erwartet; dies wird durch die Transformation zu r_{D_i} nach oben 'gestreckt'.
- ▶ Deviance-Residuen sind unter einem korrekten Modell ungefähr normalverteilt, wenn nicht zu viel Zensierung vorliegt.
- ▶ Bei starker Zensierung gibt es eine Häufung von Werten nahe Null.

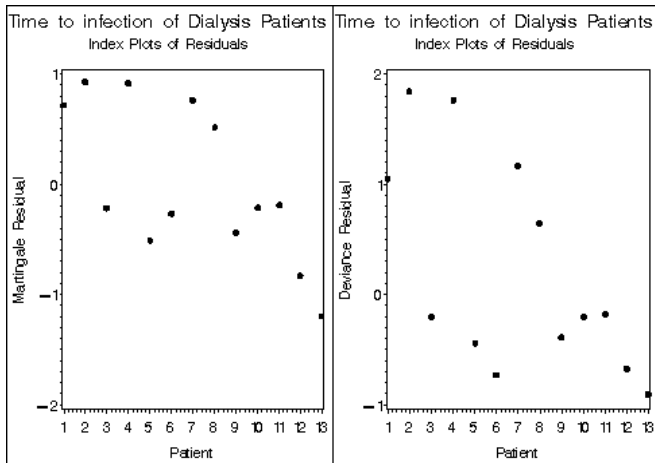
Beispiel 4.1 (Infektion bei Dialyse)

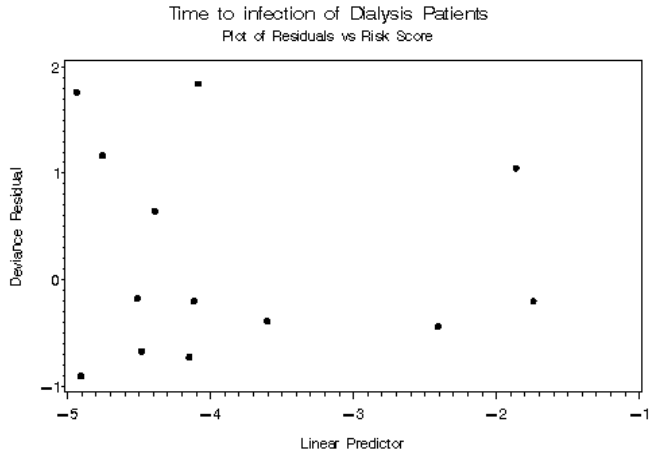
Time to infection of Dialysis Patients (Collett, p. 119)

| Patient | Time | Status | Age | Sex |
|---------|------|--------|-----|-----|
| 1 | 8 | 1 | 28 | 1 |
| 2 | 15 | 1 | 44 | 2 |
| 3 | 22 | 1 | 32 | 1 |
| 4 | 24 | 1 | 16 | 2 |
| 5 | 30 | 1 | 10 | 1 |
| 6 | 54 | 0 | 42 | 2 |
| 7 | 119 | 1 | 22 | 2 |
| 8 | 141 | 1 | 34 | 2 |
| 9 | 185 | 1 | 60 | 2 |
| 10 | 292 | 1 | 43 | 2 |
| 11 | 402 | 1 | 30 | 2 |
| 12 | 447 | 1 | 31 | 2 |
| 13 | 536 | 1 | 17 | 2 |

| Variable | Estimate | StdErr | Hazard Ratio | Prob ChiSq | Label |
|----------|----------|---------|-----------------|---------------|------------------|
| Age | 0.03037 | 0.02624 | 1.031 | 0.2470 | Age (years) |
| Sex | -2.71076 | 1.09590 | 0.066 | 0.0134 | 1/2: male/female |

| Obs | Age | Sex | Status | ResCox | ResMart | ResDev | ResP1 | ResP2 | ResS1 | ResS2 |
|-----|-----|-----|--------|---------|----------|----------|----------|----------|----------|----------|
| 1 | 28 | 1 | 1 | 0.3286 | 0.71996 | 1.05152 | -1.0850 | -0.24163 | -0.7811 | -0.17396 |
| 2 | 44 | 2 | 1 | 0.0785 | 0.92769 | 1.84341 | 14.4930 | 0.66438 | 13.4322 | 0.61350 |
| 3 | 32 | 1 | 1 | 1.4331 | -0.21393 | -0.20034 | 3.1291 | -0.30646 | -0.3223 | 0.05787 |
| 4 | 16 | 2 | 1 | 0.0939 | 0.91573 | 1.76519 | -10.2216 | 0.43406 | -9.2144 | 0.38429 |
| 5 | 10 | 1 | 1 | 1.7736 | -0.50602 | -0.43943 | -16.5882 | -0.55037 | 9.8333 | 0.12952 |
| 6 | 42 | 2 | 0 | 0.3117 | -0.26463 | -0.72751 | . | . | -3.8255 | -0.14516 |
| 7 | 22 | 2 | 1 | 0.2655 | 0.76453 | 1.16760 | -17.8286 | 0.00000 | -15.4014 | -0.07908 |
| 8 | 34 | 2 | 1 | 0.5386 | 0.51632 | 0.64809 | -7.6201 | 0.00000 | -7.0914 | -0.11385 |
| 9 | 60 | 2 | 1 | 1.6523 | -0.43792 | -0.38658 | 17.0910 | 0.00000 | -15.8114 | -0.25077 |
| 10 | 43 | 2 | 1 | 1.4234 | -0.21235 | -0.19894 | 10.2390 | 0.00000 | 1.5643 | -0.14964 |
| 11 | 30 | 2 | 1 | 1.4207 | -0.18662 | -0.17613 | 2.8575 | 0.00000 | 6.5754 | -0.10083 |
| 12 | 31 | 2 | 1 | 2.3927 | -0.82793 | -0.67044 | 5.5338 | 0.00000 | 4.7967 | -0.10394 |
| 13 | 17 | 2 | 1 | . | -1.19482 | -0.90412 | 0.0000 | 0.00000 | 16.2456 | -0.06794 |
| | | | | ===== | ===== | ===== | ===== | ===== | ===== | ===== |
| | | | | 11.7125 | 0.00000 | 2.77232 | 0.0000 | -0.00000 | 0.0000 | -0.00000 |





Funktionale Form der Kovariablen

Es ist oft sinnvoll, eine Transformation $f(x_j)$ einer Kovariablen im Modell zu betrachten.

Beispiel: X_i = Alter von Patient i

$$\lambda_i(t) = \lambda_0(t) \exp(\beta x_i)$$

bedeutet: HR von 45-Jährigen zu 50-Jährigen ist gleich HR von 80- zu 85-Jährigen. Realistisch?

Funktionale Form der Kovariablen (Forts.)

- ▶ Mit den Martingal-Residuen läßt sich die funktionale Form grob überprüfen.
- ▶ Therneau et al. haben gezeigt, dass gilt:

$$\mathbb{E}(r_{M_i} \mid X_{ij} = x_j) \approx c \cdot f(x_j),$$

wenn f die korrekte Transformation der Kovariablen ist.

- ▶ Dies kann also grob mit einem geglätteten Scatterplot (Martingal-Residuum versus x_j) visuell geprüft werden.

→ Beispiel!

Beispiel 4.2 (Leber-Erkrankung (Mayo Liver Disease Data))

The CONTENTS Procedure

| | | | |
|----------------|------------------------------|-----------------------|-----|
| Data Set Name: | WORK.LIVER | Observations: | 418 |
| Member Type: | DATA | Variables: | 10 |
| Engine: | V8 | Indexes: | 0 |
| Created: | 14:04 Friday, March 28, 2008 | Observation Length: | 80 |
| Last Modified: | 14:04 Friday, March 28, 2008 | Deleted Observations: | 0 |
| Protection: | | Compressed: | NO |
| Data Set Type: | | Sorted: | NO |
| Label: | | | |

-----Alphabetic List of Variables and Attributes-----

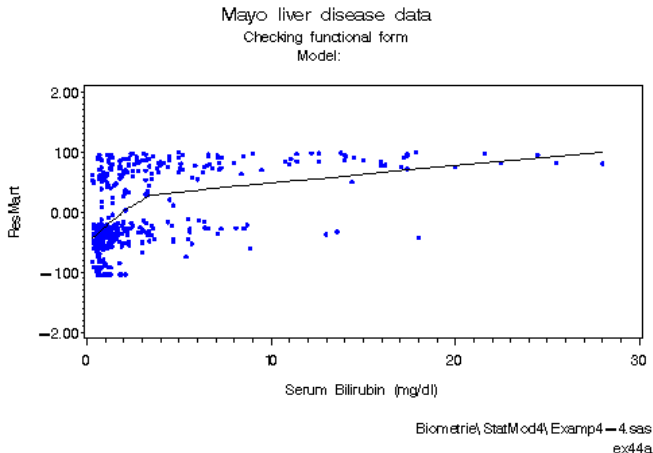
| # | Variable | Type | Len | Pos | Label |
|----|--------------|------|-----|-----|------------------------------|
| 3 | Age | Num | 8 | 16 | Age in years at registration |
| 4 | Albumin | Num | 8 | 24 | Serum Albumin (mg/dl) |
| 5 | Bilirubin | Num | 8 | 32 | Serum Bilirubin (mg/dl) |
| 6 | Edema | Num | 8 | 40 | Edema presence |
| 9 | LogAlbumin | Num | 8 | 64 | |
| 10 | LogBilirubin | Num | 8 | 72 | |
| 8 | LogProtime | Num | 8 | 56 | |
| 7 | Protime | Num | 8 | 48 | Prothrombin time (s) |
| 2 | Status | Num | 8 | 8 | 1: death, 0: censored |
| 1 | Time | Num | 8 | 0 | Follow-up time in years |

Mayo Liver Data

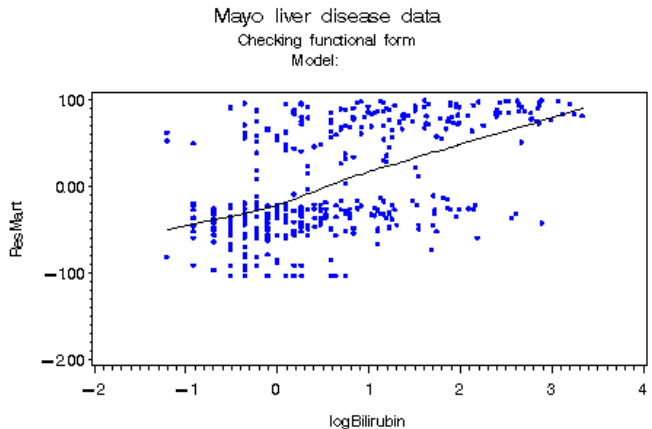
| Obs | Time | Status | Age | Albumin | Bilirubin | Protime | Edema |
|-----|---------|--------|---------|---------|-----------|---------|-------|
| 1 | 1.0951 | 1 | 58.7652 | 2.60 | 14.5 | 12.2 | 1.0 |
| 2 | 12.3203 | 0 | 56.4463 | 4.14 | 1.1 | 10.6 | 0.0 |
| 3 | 2.7707 | 1 | 70.0726 | 3.48 | 1.4 | 12.0 | 0.5 |
| 4 | 5.2704 | 1 | 54.7406 | 2.54 | 1.8 | 10.3 | 0.5 |
| 5 | 4.1177 | 0 | 38.1054 | 3.53 | 3.4 | 10.9 | 0.0 |
| 6 | 6.8528 | 1 | 66.2587 | 3.98 | 0.8 | 11.0 | 0.0 |
| 7 | 5.0157 | 0 | 55.5346 | 4.09 | 1.0 | 9.7 | 0.0 |
| 8 | 6.7515 | 1 | 53.0568 | 4.00 | 0.3 | 11.0 | 0.0 |
| 9 | 6.5708 | 1 | 42.5079 | 3.08 | 3.2 | 11.0 | 0.0 |
| 10 | 0.1396 | 1 | 70.5599 | 2.74 | 12.6 | 11.5 | 1.0 |

-----Biometrie\StatMod4\Examp4-4.sas-----

Plottet man r_{M_i} gegen Bilirubin, ist eine starke Krümmung für kleine Werte erkennbar.

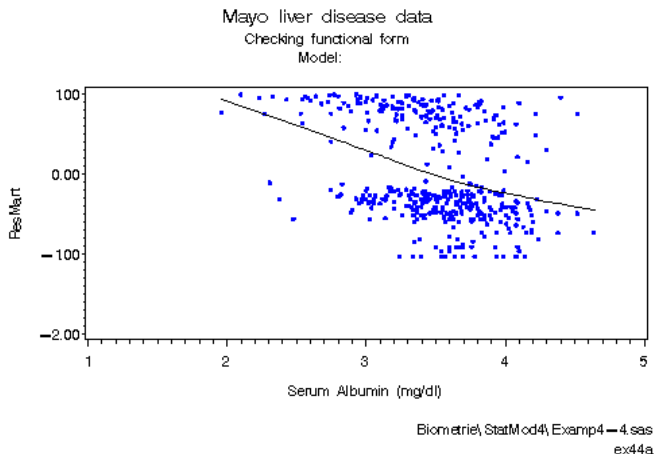


Nach log-Transformation von Bilirubin ist der Plot 'linearer'

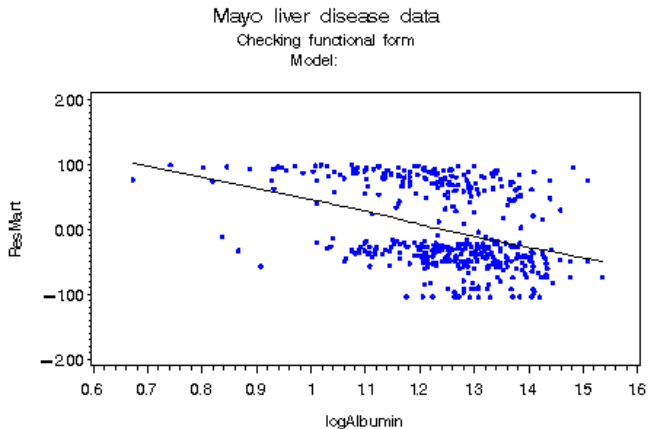


Biometrie\StatMod4\ Examp4 — 4.sas
ex44b

Plottet man r_{M_i} gegen S.Albumin, ist eine leichte Kurve erkennbar.

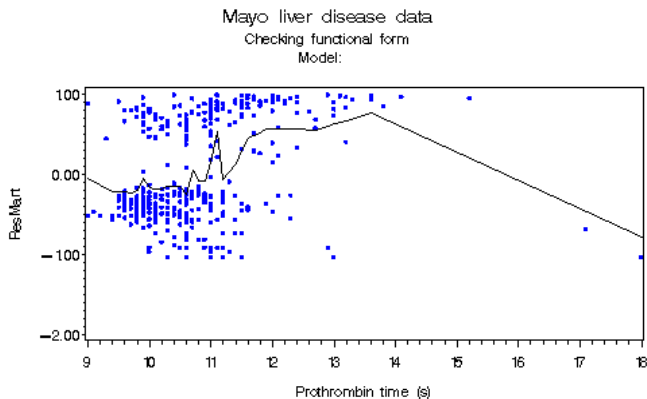


Auch hier hilft die log-Transformation.



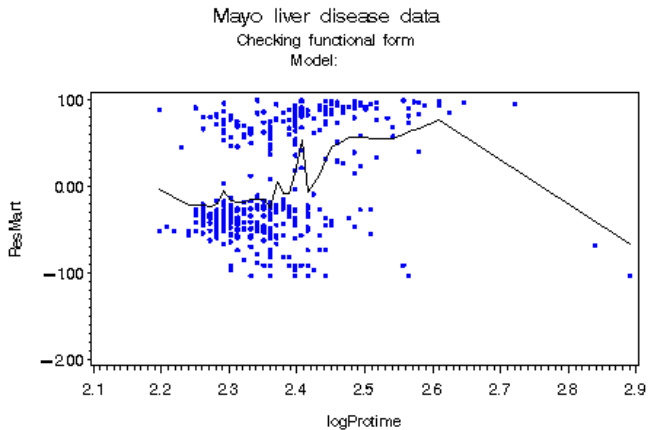
Biometrie\StatMod4\ Examp4 — 4.sas
ex44b

Bei Prothrombin...



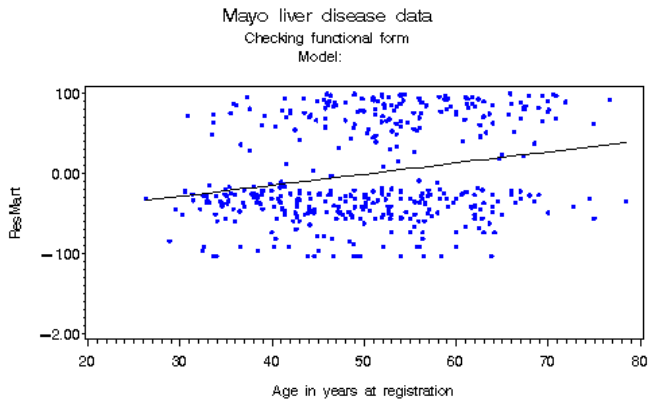
Biometrie\StatMod4\Examp4-4.sas
ex44a

... hilft auch die log-Transformation nicht viel. Evtl. diskretisieren!



Biometrie\StatMod4\ Examp4 — 4.sas
ex44b

Bei der Variable Alter braucht man nicht zu transformieren.



Biometrie\StatMod4\Examp4-4.sas
ex44a

Schoenfeld-Residuen

- ▶ Keine Schätzung von $\Lambda(t)$ benötigt;
- ▶ ein Residuum für jede Komponente von $X = (X_1, \dots, X_p)$ für jede Ereigniszeit;
- ▶ wenn keine Bindungen:

$$r_{P_{ij}} = \delta_i(x_{ij} - \hat{a}_{ij}) \quad (4.3)$$

mit

$$\hat{a}_{ij} = \sum_{l \in R(t_i)} w_l \cdot x_{lj},$$

wobei w_l Gewichte:

$$w_l = \frac{e^{\hat{\eta}_l}}{\sum_{k \in R(t_i)} e^{\hat{\eta}_k}}.$$

- ▶ $r_{P_{ij}}$ sind nicht definiert für zensierte Beobachtungen.

Schoenfeld-Residuen (Forts.)

Die Schoenfeld-Residuen kommen zustande, indem man sich das Schätzen von β via Maximieren der partiellen Likelihood vor Augen führt (Scores $U(\beta)$, siehe 3-20):

$$\frac{\partial \log L(\beta)}{\partial \beta_j} = \sum_{i=1}^n \delta_i (x_{ij} - a_{ij}) \quad (4.4)$$

mit

$$a_{ij} = \sum x_{lj} \frac{e^{\beta^T x_l}}{\sum e^{\beta^T x_l}}$$

Der i -te Term in (4.4), berechnet an der Stelle $\hat{\beta}$, ist das Schoenfeld-Residuum für X_j .

- Wegen $\partial \log L(\hat{\beta}) / \partial \beta_j = 0$ gilt für alle j : $\sum_{i=1}^n r_{P_{ij}} = 0$.

Schoenfeld-Residuen (Forts.)

- ▶ Man kann $r_{P_{ij}}$ auch interpretieren als die Differenz zwischen beobachtetem X_{ij} und unter dem Modell erwartetem Wert dieser Kovariable für Individuum i . → Warum?
- ▶ Skalierte Schoenfeld-Residuen:
Bezeichne $\mathbf{r}_{P_i}^T = (r_{P_{i1}}, \dots, r_{P_{ip}})$ den Vektor der Residuen, so sind die skalierten Schoenfeld-Residuen definiert als

$$\mathbf{r}_{P_i}^* = r \widehat{\mathbf{V}}(\widehat{\boldsymbol{\beta}}) \mathbf{r}_{P_i},$$

wobei r die Anzahl der Ereignisse in der Stichprobe bezeichnet.

Zeitabhängige Effekte?

- ▶ Die skalierten Schoenfeld-Residuen können genutzt werden, um auf zeitabhängige Effekte zu prüfen (Verletzung der PH-Annahme).
- ▶ Angenommen das wahre Modelle hätte zeitabhängige Koeffizienten $\beta(t)$

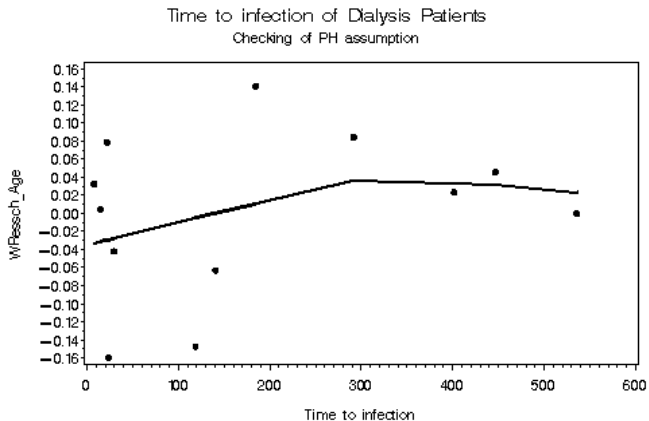
$$\lambda(t; \mathbf{x}) = \lambda_0(t) \exp(\beta(t)^T \mathbf{x})$$

- ▶ Sei $\hat{\beta}$ der geschätzte Koeffizientenvektor aus dem gewöhnlichen Modell $\lambda_0(t) \exp(\beta^T \mathbf{x})$.
- ▶ Dann gilt

$$\mathbb{E}(r_{P_{ij}}^*) + \hat{\beta}_j \approx \beta_j(t_{(i)}).$$

- ▶ Es folgt, dass ein Plot von $r_{P_{ij}}^* + \hat{\beta}_j$ gegen $t_{(i)}$ Art und Ausmaß der Zeitabhängigkeit aufzeigt.

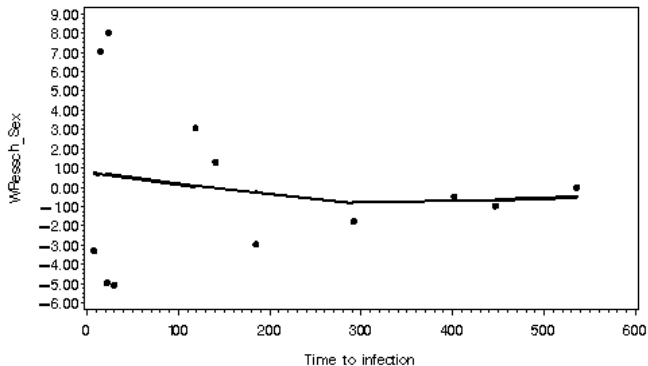
Beispiel 4.3 (a)



Biometrie, StatMod4, Examp4 – 10.sas

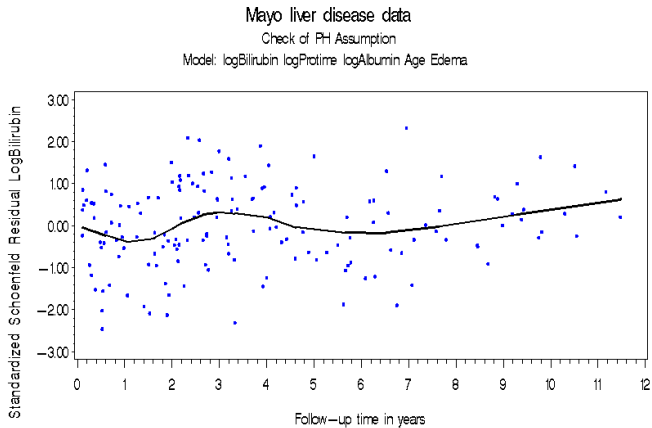
Beispiel 4.3 (b)

Time to infection of Dialysis Patients
Checking of PH assumption

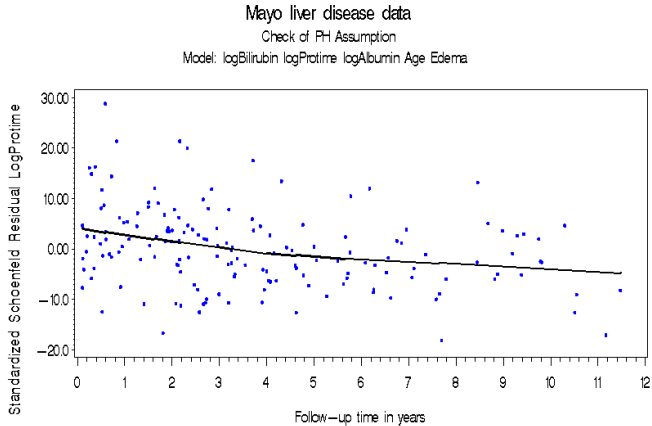


Biometrie, StatMod4, Examp4-10.sas

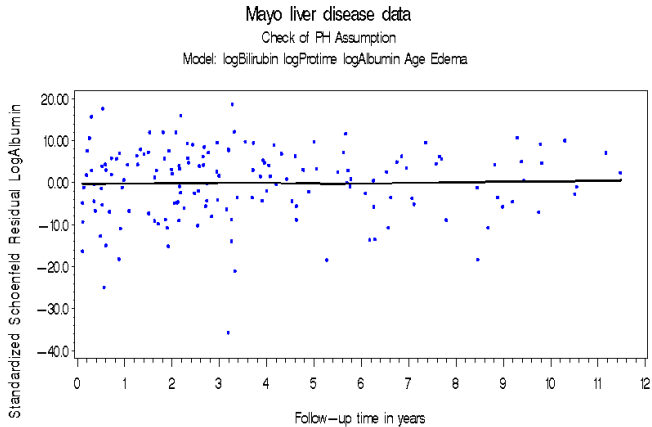
Beispiel 4.3 (c)



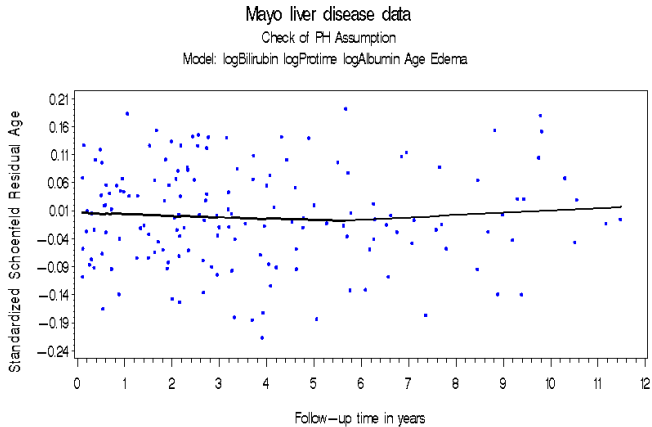
Biometrie\StatMod4\Examp4-10.sas
ex46a



Biometrie\StatMod4\Examp4—10.sas
ex46a



Biometrie\StatMod4\Examp4—10.sas
ex46a



Biometrie\StatMod4\Examp4—10.sas
ex46a

Score-Residuen

- ▶ Auch hier, separates Residuum für jede Kovariable.
- ▶ Die Score-Residuen leiten sich her als partielle Ableitung nach β_j des Beitrags des i -ten Individuums zu der partiellen Likelihood bis zu seiner Ereigniszeit t_i .

Das i -te Score-Residuum für Kovariable X_j ist definiert als

$$\begin{aligned}
 r_{S_{ij}} &= \delta_i(x_{ij} - \hat{a}_{ij}) + e^{\hat{\eta}_i} \sum_{t_r \leq t_i} \frac{(\hat{a}_{rj} - x_{ij})\delta_r}{\sum_{l \in R(t_r)} e^{\hat{\eta}_l}} \\
 &= r_{P_{ij}} + e^{\hat{\eta}_i} \sum_{t_r \leq t_i} \frac{(\hat{a}_{rj} - x_{ij})\delta_r}{\sum_{l \in R(t_r)} e^{\hat{\eta}_l}}
 \end{aligned} \tag{4.5}$$

$$\Rightarrow \sum_{r=1}^n r_{S_{ij}} = 0 \text{ für alle } j.$$

Score-Residuen (Forts.)

- ▶ Die Score-Residuen dienen selbst zur Überprüfung der PH-Annahme und haben den Vorteil, dass sie bei stetigen Einflußgrößen verwendet werden können. Die zugrundeliegende Theorie beruht auf Eigenschaften stochastischer Prozesse.
- ▶ Wenn das Modell korrekt ist, gilt, dass sich $\sqrt{\mathbb{V}(\hat{\beta}_j)} r_{S_{ij}}$ als Funktion der Zeit ungefähr wie eine *brownsche Brücke* verhält.
- ▶ D.h. in einem Plot von $\text{SE}(\hat{\beta}_j) r_{S_{ij}}$ gegen t_i müßten die Punkte innerhalb des ± 1.3581 Intervals liegen.

→ Beispiel!

Einflussreiche Beobachtungen

- ▶ Bezeichne $\hat{\beta}_{j(i)}$ die Schätzung von β_j beim Weglassen der i -ten Beobachtung.
- ▶ Eine Approximation von $\hat{\beta}_j - \hat{\beta}_{j(i)}$ basiert auf den Score-Residuen.

$$\hat{\beta}_j - \hat{\beta}_{j(i)} \approx \hat{\mathbb{V}}(\hat{\boldsymbol{\beta}}) \mathbf{r}_{S_i} =: \Delta_i \beta_j,$$

wobei $\mathbf{r}_{S_i}^T = (r_{S_{i1}}, \dots, r_{S_{ip}})$ der Vektor der Score-Residuen für Beobachtung i .

Beispiel 4.4 (Infektion bei Dialyse-Patienten)

Time to infection of Dialysis Patients

Checking of influential observations

Complete data

| Obs | Variable | Estimate | StdErr | Prob ChiSq | Label |
|-----|----------|----------|---------|---------------|------------------|
| 1 | Age | 0.03037 | 0.02624 | 0.2470 | Age (years) |
| 2 | Sex | -2.71076 | 1.09590 | 0.0134 | 1/2: male/female |

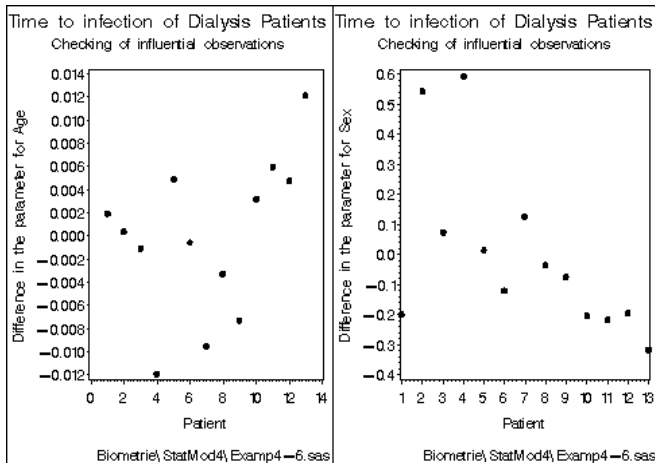
| Patient | Time | Age | Sex | ResMart | ResSco_a | ResSco_s | Dfb_a | Dfb_s |
|---------|------|-----|-----|----------|----------|----------|-----------|----------|
| 1 | 8 | 28 | 1 | 0.71996 | -0.7811 | -0.17396 | 0.001969 | -0.19767 |
| 2 | 15 | 44 | 2 | 0.92769 | 13.4322 | 0.61350 | 0.000406 | 0.54326 |
| 3 | 22 | 32 | 1 | -0.21393 | -0.3223 | 0.05787 | -0.001056 | 0.07414 |
| 4 | 24 | 16 | 2 | 0.91573 | -9.2144 | 0.38429 | -0.011880 | 0.59430 |
| 5 | 30 | 10 | 1 | -0.50602 | 9.8333 | 0.12952 | 0.004903 | 0.01386 |
| 6 | 54 | 42 | 2 | -0.26463 | -3.8255 | -0.14516 | -0.000542 | -0.11922 |
| 7 | 119 | 22 | 2 | 0.76453 | -15.4014 | -0.07908 | -0.009462 | 0.12695 |
| 8 | 141 | 34 | 2 | 0.51632 | -7.0914 | -0.11385 | -0.003241 | -0.03455 |
| 9 | 185 | 60 | 2 | -0.43792 | -15.8114 | -0.25077 | -0.007271 | -0.07335 |
| 10 | 292 | 43 | 2 | -0.21235 | 1.5643 | -0.14964 | 0.003233 | -0.20226 |
| 11 | 402 | 30 | 2 | -0.18662 | 6.5754 | -0.10083 | 0.005979 | -0.21584 |
| 12 | 447 | 31 | 2 | -0.82793 | 4.7967 | -0.10394 | 0.004800 | -0.19394 |
| 13 | 536 | 17 | 2 | -1.19482 | 16.2456 | -0.06794 | 0.012162 | -0.31568 |

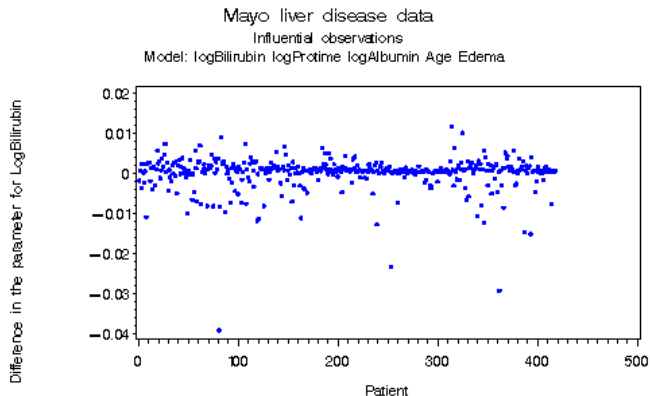
Time to infection of Dialysis Patients
 Checking of influential observations
 Patient 13 discarded

13:37 Friday, March 28, 2008 3

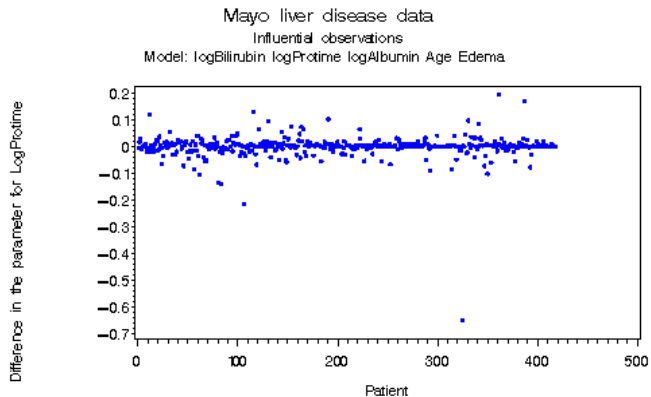
| Obs | Variable | Estimate | StdErr | Prob ChiSq | Label |
|-----|----------|----------|---------|---------------|------------------|
| 1 | Age | 0.01086 | 0.03032 | 0.7202 | Age (years) |
| 2 | Sex | -2.23367 | 1.11459 | 0.0451 | 1/2: male/female |

 Biometrie\StatMod4\Examp4-6.sas

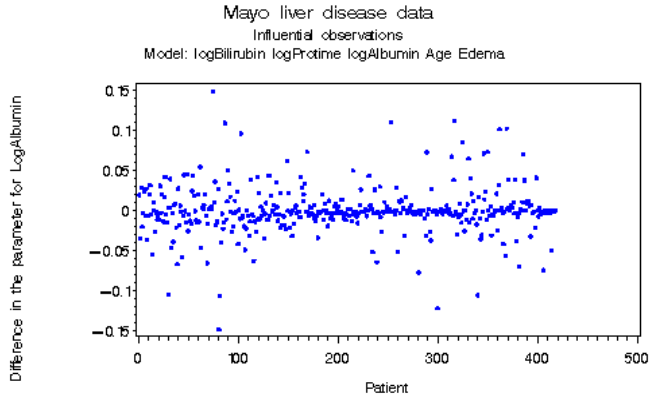




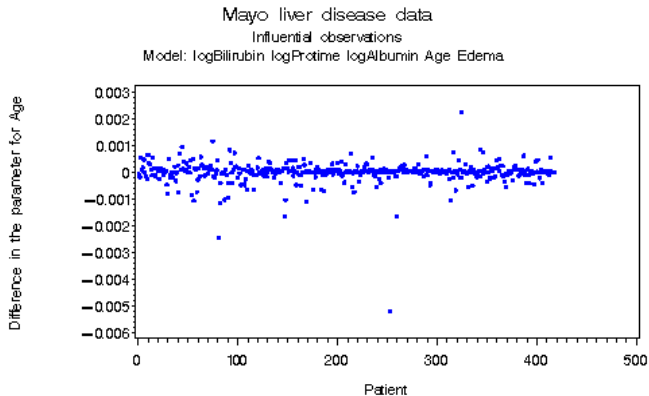
Biometrie\StatMod4\Examp4-6.sas
ex46a



Biometrie\StatMod4\Examp4-6.sas
ex46a



Biometrie\StatMod4\Examp4-6.sas
ex46a



Biometrie\StatMod4\Examp4-6.sas
ex46a

Einfluss auf Gesamt Fit

Die Likelihood-Displacement-Statistik ist

$$2\{\log L(\hat{\beta}) - \log L(\hat{\beta}_{(i)})\}$$

und wird approximiert durch

$$L\Delta_i = \mathbf{r}_{S_i}^T \hat{\mathbf{W}}(\hat{\beta}) \mathbf{r}_{S_i}.$$

Ein weiteres Maß, das den Einfluss einzelner Beobachtungen auf den Gesamtfit wiedergibt, heißt l_{\max} . Es berechnet sich als Betrag des Eigenvektors zum maximalen Eigenwert von

$$B = \Theta^T \hat{\mathbf{W}}(\hat{\beta}) \Theta,$$

wobei Θ^T die $n \times p$ -Matrix der \mathbf{r}_{S_i} , $i = 1, \dots, n$.

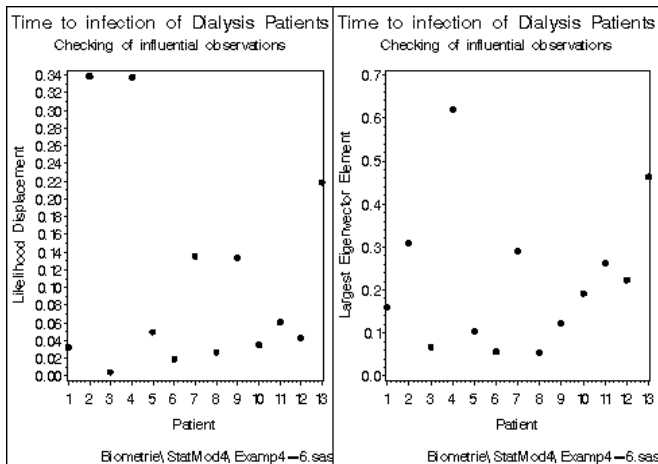
Zur Illustration

Time to infection of Dialysis Patients

Checking of influential observations-Complete data

| Patient | Time | Age | Sex | ResMart | ResSco_a | ResSco_s | L_Displace | Lmax |
|---------|------|-----|-----|----------|----------|----------|------------|---------|
| 1 | 8 | 28 | 1 | 0.71996 | -0.7811 | -0.17396 | 0.03285 | 0.16101 |
| 2 | 15 | 44 | 2 | 0.92769 | 13.4322 | 0.61350 | 0.33875 | 0.30927 |
| 3 | 22 | 32 | 1 | -0.21393 | -0.3223 | 0.05787 | 0.00463 | 0.06766 |
| 4 | 24 | 16 | 2 | 0.91573 | -9.2144 | 0.38429 | 0.33785 | 0.62061 |
| 5 | 30 | 10 | 1 | -0.50602 | 9.8333 | 0.12952 | 0.05001 | 0.10416 |
| 6 | 54 | 42 | 2 | -0.26463 | -3.8255 | -0.14516 | 0.01938 | 0.05750 |
| 7 | 119 | 22 | 2 | 0.76453 | -15.4014 | -0.07908 | 0.13570 | 0.29117 |
| 8 | 141 | 34 | 2 | 0.51632 | -7.0914 | -0.11385 | 0.02692 | 0.05397 |
| 9 | 185 | 60 | 2 | -0.43792 | -15.8114 | -0.25077 | 0.13335 | 0.12352 |
| 10 | 292 | 43 | 2 | -0.21235 | 1.5643 | -0.14964 | 0.03532 | 0.19266 |
| 11 | 402 | 30 | 2 | -0.18662 | 6.5754 | -0.10083 | 0.06108 | 0.26352 |
| 12 | 447 | 31 | 2 | -0.82793 | 4.7967 | -0.10394 | 0.04318 | 0.22366 |
| 13 | 536 | 17 | 2 | -1.19482 | 16.2456 | -0.06794 | 0.21903 | 0.46368 |

-----Biometrie\StatMod4\Examp4-6.sas-----



Inhalt von Abschnitt 5

Erweiterungen des Cox Modells

- Cox Modell und Log-Rank-Test

- Stratifizierte Cox-Regression

- Zeitabhängige Kovariablen

 - Likelihood mit zeitabhängigen Variablen

 - Interpretation von Behandlungseffekten

 - Beispiele

 - Gewichteter Log-rank Test

- Kausale Interpretation

Cox Modell und Log-Rank-Test

Wir betrachten hier noch einmal den Log-Rank-Test (vgl. Folie 2-29) als Hintergrundinformation zu den nächsten Themen. Zusammenhang: Log-Rank-Test zum Vergleich der Survivalkurven zweier Gruppen und Cox Modell.

Seien wiederum

$t_{(1)} < t_{(2)} < \dots < t_{(r)}$ – geordnete Ereigniszeitpunkte der kombinierten Stichprobe

d_{ij} – Anzahl Fälle in $t_{(j)}$ aus Gruppe i

n_{ij} – Anzahl unter Risiko in $t_{(j)}$ in Gruppe i

Sei
$$X = \begin{cases} 1 & \text{für Person aus Gruppe 1} \\ 0 & \text{für Person aus Gruppe 2} \end{cases}$$

Falls es keine Bindungen gibt, d.h. $d_{1j} + d_{2j} = 1$ für alle j , schreibt sich die Likelihood (3.2) als

$$L(\beta) = \prod_{j=1}^r \frac{\exp(\beta x_{(j)})}{\sum_{l=1}^{n_j} \exp(\beta x_l)}.$$

Die Loglikelihood ergibt sich als

$$\log L(\beta) = \sum_{j=1}^r \beta x_{(j)} - \sum_{j=1}^r \log \left\{ \sum_{l=1}^{n_j} \exp(\beta x_l) \right\} \quad (5.1)$$

$$= d_1 \beta - \sum_{j=1}^r \log \{ n_{1j} \exp(\beta) + n_{2j} \} \quad (5.2)$$

Um $H_0: \beta = 0$ zu prüfen, kann man den Score-Test anwenden.

Die Teststatistik ist (vgl. Folien 3-18ff.)

$$X = \frac{U^2(0)}{I(0)},$$

wobei $U(\beta) = \frac{\partial \log L(\beta)}{\partial \beta}$ (efficient score)

$$I(\beta) = -\frac{\partial^2 \log L(\beta)}{\partial \beta^2} \quad (\text{beob. Fisher-Information})$$

Unter $H_0 : \beta = 0$ hat X eine χ_1^2 -Verteilung.

Man berechnet, dass

$$\begin{aligned} \frac{\partial \log L(\beta)}{\partial \beta} &= \sum_{j=1}^r \left(d_{1j} - \frac{n_{1j} \exp(\beta)}{n_{1j} \exp(\beta) + n_{2j}} \right) \\ \frac{\partial^2 \log L(\beta)}{\partial \beta^2} &= - \sum_{j=1}^r \frac{n_{1j} n_{2j} \exp(\beta)}{(n_{1j} \exp(\beta) + n_{2j})^2}. \end{aligned}$$

Unter $\beta = 0$ ergeben sich damit

$$U(0) = \sum_{j=1}^r \left(d_{1j} - \frac{n_{1j}}{n_{1j} + n_{2j}} \right)$$

$$I(0) = \sum_{j=1}^r \frac{n_{1j}n_{2j}}{n_j^2},$$

Im Falle keiner Bindungen sind das exakt die Ausdrücke für U_L und V_L in (2.21) und (2.22) für den Fall $d_j = 1$ für alle j .

⇒ Falls es Bindungen gibt, ist der Log-Rank-Test äquivalent zum Score-Test basierend auf der Cox-Approximation für Bindungen.

Motivation: Stratifizieren

Zur Erinnerung:

Voraussetzungen für das Cox Modell:

$$\lambda_i(t) := \lambda(t; \mathbf{X} = \mathbf{x}_i) = \psi(\mathbf{x}_i)\lambda_0(t),$$

- ▶ Basis-Hazardfunktion $\lambda_0(t)$ gleich für alle Personen mit $\mathbf{X} = \mathbf{0}$.
- ▶ $\lambda(t; \mathbf{X} = \mathbf{x}_i)$ proportional zu $\lambda_0(t)$.

Diskussion:

Studie zu einer neuen Behandlung; PH-Annahme plausibel für Behandlungseffekt für Frauen und separat für Männer, aber PH-Annahme zwischen Frauen und Männern nicht plausibel. Was dann?

Stratifizierte Cox-Regression

Beispiel 5.1 (Myelomstudie (Peto, 1977))

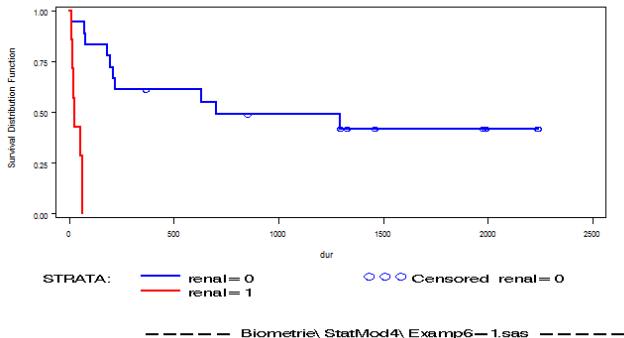
Randomisierte Studie mit 25 Myelom-Patienten, in der die Überlebenszeit für zwei Behandlungsgruppen untersucht wird. Als wichtiger Einflussfaktor auf das Überleben wurde die Nierenfunktion (1=normale Funktion) zum Zeitpunkt der Randomisierung erfasst.

First 10 Observations of Dataset myel

| Obs | id | dur | status | treat | renal |
|-----|----|------|--------|-------|-------|
| 1 | 1 | 8 | 1 | 1 | 1 |
| 2 | 2 | 180 | 1 | 2 | 0 |
| 3 | 3 | 632 | 1 | 2 | 0 |
| 4 | 4 | 852 | 0 | 1 | 0 |
| 5 | 5 | 52 | 1 | 1 | 1 |
| 6 | 6 | 2240 | 0 | 2 | 0 |
| 7 | 7 | 220 | 1 | 1 | 0 |
| 8 | 8 | 63 | 1 | 1 | 1 |
| 9 | 9 | 195 | 1 | 2 | 0 |
| 10 | 10 | 76 | 1 | 2 | 0 |

Beispiel 5.1 (Myelomstudie (Peto, 1977) – Forts.)

Survival Curves for Myeloma Patients Stratified by Renal Function



Vermutung: Form der Hazardfunktion unterschiedlich für Personen mit und ohne Nierenschaden. Proportionalität der Hazardfunktionen ist nicht gegeben.

Lösung: Passe Cox Modelle mit unterschiedlichen Baseline-hazardfunktionen für verschiedene Werte von R Nierenfunktion an \Rightarrow stratifiziert nach R .

► Mit Nierenschaden:

$$\lambda_{i,R=0}(t) = \psi(\mathbf{x}_i)\lambda_{0,R=0}(t)$$

► Ohne Nierenschaden:

$$\lambda_{i,R=1}(t) = \psi(\mathbf{x}_i)\lambda_{0,R=1}(t)$$

Beachte: $\psi(\mathbf{x}_i) = \exp(\boldsymbol{\beta}^T \mathbf{x}_i)$ wird in beiden Modellen als gleich angenommen.

—→ Diskussion!

Stratifizierte Partial Likelihood

Likelihoodbeitrag aus Stratum s :

$$L_s(\beta) = \prod_{i=1}^{n_s} \left\{ \frac{\exp(\beta^T \mathbf{x}_{si})}{\sum_{l \in R(t_{si})} \exp(\beta^T \mathbf{x}_{sl})} \right\}^{\delta_{si}}$$

Multiplizieren liefert stratifizierte Partial Likelihood:

$$L_S(\beta) = \prod_{s=1}^S L_s(\beta)$$

Schätzer werden in Analogie zur nicht stratifizierten Situation bestimmt.

Beachte: Koeffizientenvektor β ist für alle Straten gleich.

Beispiel 5.1 (Myelomstudie (Peto, 1977) – Forts.)

Cox Model Stratified by Renal Function

| Parameter | Estimate | StdErr | Hazard Ratio | Prob ChiSq |
|-----------|----------|---------|--------------|------------|
| treat | 1.46398 | 0.65965 | 4.323 | 0.0265 |

Cox Model Adjusted for Renal Function

| Parameter | Estimate | StdErr | Hazard Ratio | Prob ChiSq |
|-----------|----------|---------|--------------|------------|
| treat | 1.22191 | 0.59779 | 3.394 | 0.0409 |
| renal | 4.27792 | 1.19119 | 72.091 | 0.0003 |

→ Diskussion!

Eigenschaften stratifizierter Modelle

- ▶ Stratifizierung ermöglicht flexible Modellierung aufgrund unterschiedlicher Grundhazardfunktionen in den Straten. Insbesondere werden keine unnötigen Annahmen über die Form der Hazardfunktion getroffen.
- ▶ Der Effekt der Stratifizierungsvariable selbst wird dabei nicht geschätzt — dies ist ein kleiner Preis, wenn hauptsächlich der Behandlungseffekt von Interesse und durch die größere Allgemeinheit Bias vermieden werden kann.
- ▶ Stratifizierte können nicht mit unstratifizierte Modellen auf Basis der Log-Likelihood miteinander verglichen werden.

Stratifizierter Long-rank Test

Wir testen hier $H_0 : \beta = 0$ — da β in allen Strata als gleich angenommen wird, ist die Hypothese automatisch für alle Strata.

Strata $s = 1, \dots, S$. Score in Stratum s

$$U_s(\beta) = \frac{\partial \log L_s(\beta)}{\partial \beta}.$$

Varianz des Scores in Stratum s

$$I_s(\beta) = -\frac{\partial^2 \log L_s(\beta)}{\partial \beta^2}.$$

Log-rank Test in Stratum s :

$$X_s(\beta) = \frac{U_s(0)^2}{I_s(0)}.$$

Stratifizierter Long-rank Test (Forts.)

⇒ Stratifizierter Log-rank Test:

$$X_S(\beta) = \frac{(\sum U_s(0))^2}{\sum I_s(0)}.$$

D.h. analog zum unstratifizierten Fall ist auch der Score Test im stratifizierten Cox-Modell identisch mit dem stratifizierten Log-rank Test (2.23).

Stratifizierter Long-rank Test (Forts.)

Dies impliziert, dass der stratifizierte Log-rank Test besonders “empfindlich” ist für Alternativhypothesen der Form

$$H_1 : \frac{\lambda_{1s}(t)}{\lambda_{2s}(t)} = \dots = \frac{\lambda_{1S}(t)}{\lambda_{2S}(t)} = e^\beta \text{ für } \beta \neq 0,$$

d.h. bei Alternativen bei denen

- ▶ in jedem Stratum der Baseline-Hazard beliebig ist;
- ▶ die Behandlungsgruppen in jedem Stratum proportionale Hazards haben;
- ▶ der Behandlungseffekt $\exp(\beta)$ in jedem Stratum der gleiche ist.
- ▶ \longrightarrow z.B. nicht... ?

Zeitabhängige Kovariablen

In vorigen Kapiteln wurde mit Kovariablen gearbeitet, die zum Beginn ($t = 0$) erhoben waren. Falls Kovariable im Laufe einer Studie regelmäßig erhoben werden, kann die Einbeziehung solcher Information die Prognosegüte des Modells verbessern.

Beachte: Die Interpretation der Parameterschätzer wird allerdings dabei i.d.R. sehr viel komplizierter!

Beispiele

- ▶ Bei Behandlungen / Medikamenten, die immer mal wieder unterbrochen werden: Indikatorvariable $x(t)$ dafür, ob Medikament zum Zeitpunkt t genommen wurde $x(t) = 1$ oder nicht $x(t) = 0$.
- ▶ Bei Leukämiepatienten: $X(t)$ = Anzahl Blutplättchen zum Zeitpunkt t nach Transplantation ($t = 0$); oder Indikator dafür, wann normales Niveau wieder erreicht wurde.
- ▶ Wechselnde äußere Umstände, z.B. wann und wie lange jemand am Arbeitsplatz radioaktiver Strahlung ausgesetzt war.

Typen zeitabhängiger Kovariablen

Man unterscheidet zwei Typen zeitabhängiger Variablen, interne und externe Variable.

Interne Variablen beziehen sich auf ein Individuum der Studie und können nur gemessen werden, solange die Person kein Ereignis hatte (z.B. noch lebt) und nicht zensiert wurde. Beispiele sind

- ▶ Lungenfunktionsmessungen, Anzahl weißer Blutzellen, systolischer Blutdruck, Serum Cholesterin;
- ▶ Immunreaktion nach Transplantation.

Beachte: Survivalfunktion kann nicht auf interne Variablen bedingt werden, da nicht im Vorhinein definiert (z.B. Überleben vorausgesetzt)!

Externe Variablen sind zeitabhängige Variable, die nicht das Überleben des Patienten zur Voraussetzung haben. Beispiele sind

- ▶ Variable, deren Änderung genau vorhersehbar ist, wie z.B. das Alter eines Patienten (gegeben sein Anfangsalter); variable Dosierung eines Medikaments, das einem vorgegebenen Zeitmuster folgt.
- ▶ Ein anderer Typ externer Kovariablen betrifft Merkmale, die unabhängig von den Patienten sind wie z.B. Wetterbedingungen.
- ▶ Zeitabhängige Kovariable entstehen außerdem in Situationen, in denen sich die Wirkung einer konstanten Kovariablen mit der Zeit ändert, also mit t interagiert.

Oft: Zur Überprüfung der PH-Annahme wird ein zeitabhängiger Effekt ins Modell eingefügt, um ihn dann auszuschliessen.

Zeitabhängige Kovariablen im Cox Modell

Überprüfen der PH-Annahme:

Eine (von vielen) Möglichkeiten zur Überprüfen der proportionalen Hazard-Annahme führt eine zeitabhängige Kovariable $x(t)$ folgendermassen ein.

Sei $x_1 = x$, $x_2(t) = x \cdot g(t)$ (z.B. $g(t) = t$ oder oft $g(t) = \log(t)$).

$$\lambda(t \mid \mathbf{X} = \mathbf{x}) = \lambda_0(t) \exp(\beta_1 x_1 + \beta_2 x_2(t))$$

β_1 – Behandlungseffekt zu $t = 0$, und für $g(t) = t$:

$$\beta_2 = \begin{cases} > 0 & : \text{Hazard Ratio der Behandlung steigt mit Zeit} \\ = 0 & : \text{kein Zeiteffekt der Behandlung} \\ < 0 & : \text{Hazard Ratio der Behandlung fällt mit Zeit} \end{cases}$$

Falls Kovariable zeitabhängig ist, also hier $\beta_2 \neq 0$, handelt es sich **nicht** mehr um ein proportionales Hazard Modell.

Beispiel: Chemotherapie bei Ovarial-Krebs — verändert sich die Prognose bei älteren Frauen z.B. schneller als bei jüngeren?

Ein Modell mit zeitabhängigen Variablen

Die Erweiterung des proportionalen Hazards Modells

$$\lambda_i(t) = \exp \left(\sum_j x_{ij} \beta_j \right) \cdot \lambda_0(t)$$

auf zeitabh. Kovariable ist

$$\lambda_i(t) = \exp \left(\sum_j x_{ij}(t) \beta_j \right) \cdot \lambda_0(t), \quad (5.3)$$

wobei die Basis-Hazardfunktion eine Person repräsentiert mit $x_j(t) = 0$ für alle $t, j = 1, \dots, p$.

Binäre zeitabhängige Variable

- ▶ z.B. Behandlung setzt zu einem nicht festgelegten Zeitpunkt ein;
- ▶ Hier wird ein zeitabhängiger Indikator $x_1(t)$ eingeführt:

$$x_1(t) = \begin{cases} 0 & \text{zum Ztp. } t \text{ noch nicht in Behandlung} \\ 1 & \text{zum Ztp. } t \text{ schon in Behandlung,} \end{cases}$$

die Hazardfunktion für Person i ist dann

$$\lambda_i(t) = \exp\{\eta_i + \beta x_{i1}(t)\} \cdot \lambda_0(t),$$

wobei η_i eine Linearkombination zeit-unabhängiger Kovariablen für Person i (zu Beginn der Studie).

Sei $x(t)$ = externer zeitabh. Indikator.

—> Wie sehen Hazardfunktion, Survivalfunktion etc. aus?

Das Hazard Ratio in einem Modell mit zeitabhängigen Variablen ist **nicht konstant**.

Für zwei unabhängige Personen r, s ist

$$\frac{\lambda_r(t)}{\lambda_s(t)} = \exp\{\beta_1(x_{r1}(t) - x_{s1}(t)) + \cdots + \beta_p(x_{rp}(t) - x_{sp}(t))\},$$

so dass e^{β_j} das Hazard Ratio zweier Personen darstellt, die sich zu allen Zeitpunkten um eine Einheit in Kovariable x_j unterscheiden, gegeben dass alle anderen Kovariablen bei beiden zu allen Zeitpunkten gleich sind.

Bemerkung: $x_{ij}(t)$ kann eine Funktion des gesamten Zeitintervalls $(0, t)$ sein, z.B. Dauer der bisherigen Exposition.

Likelihood mit zeitabhängigen Variablen

Mit den hier üblichen Bezeichnungen schreibt sich die log-Likelihood für das Cox-Modell zeitabhängigen Kovariablen als

$$\log L = \sum_{i=1}^n \delta_i \left(\beta^T x_i(t_i) - \log \sum_{l \in R(t_i)} \exp(\beta^T x_l(t_i)) \right). \quad (5.4)$$

Um (5.4) zu berechnen, müssen $x_l(t_i)$ bekannt sein für alle $l \in R(t_i)$ für Ereigniszeitpunkte! Das kann für interne Kovariable ein Problem sein.

Interpretation von Behandlungseffekten

► Behandlungsvergleich

Obacht!

Ein Behandlungseffekt kann **maskiert** werden durch Einbeziehung zeitabhängiger Kovariable:

$$\text{Behandlung} \rightarrow X(t) \rightarrow \text{Gesundung}$$

—→ Diskussion

Beispiel 5.2 (Knochenmark-Transplantation bei Leukämie-Patienten)

In dieser Studie wird ALL- und AML-Leukämie-Patienten neues Knochenmark eingepflanzt. Vorher wurde per Chemotherapie versucht, alle Krebszellen zu zerstören, was auch die Blutplättchen in Mitleidenschaft zog.

- ▶ Die Anzahl der Blutplättchen (Normalniveau ja oder nein) ist ein wichtiger prognostischer Faktor für das Überleben des Patienten.
- ▶ Dieser Zustand ist eine zeitabhängige Kovariable.
- ▶ Die Analyse soll herausfinden, ob es, unter Berücksichtigung prognostischer Faktoren, einen Survival-Unterschied zwischen den betrachteten Gruppen gibt.

Survival of Leukaemia Patients (Collett, p. 262)

Variable Type Len Pos Label

```

-----
6 Dage      Num      8  40 Donor age
4 Group     Num      8  24 Disease group
7 P         Num      8  48 Platelet recovery (1/0: yes/no)
8 PTime     Num      8  56 Time in days to return of platelets to normal
5 Page      Num      8  32 Patient age
1 Patient   Num      8    0
3 Status    Num      8  16 Event (1: event, 0: censored)
2 Time      Num      8    8 Survival time (days)

```

| Patient | Time | Status | Group | Page | Dage | P | PTime | group1 | group2 | group3 |
|---------|------|--------|-------|------|------|---|-------|--------|--------|--------|
| 1 | 1199 | 0 | ALL | 24 | 40 | 1 | 29 | 1 | 0 | 0 |
| 2 | 1111 | 0 | ALL | 19 | 28 | 1 | 22 | 1 | 0 | 0 |
| 3 | 530 | 0 | ALL | 17 | 28 | 1 | 34 | 1 | 0 | 0 |
| 4 | 1279 | 1 | ALL | 17 | 20 | 1 | 22 | 1 | 0 | 0 |
| 5 | 110 | 1 | ALL | 28 | 25 | 1 | 49 | 1 | 0 | 0 |

-----Biometrie\StatMod4\Examp8-1.sas-----

Survival of Leukaemia Patients (Collett, p. 262)

| Patient | Group | Page | Dage | P | group1 | group2 | group3 | Start | Stop | Status |
|---------|-------|------|------|---|--------|--------|--------|-------|------|--------|
| 1 | 1 | 24 | 40 | 0 | 1 | 0 | 0 | 0 | 29 | 0 |
| 1 | 1 | 24 | 40 | 1 | 1 | 0 | 0 | 29 | 1199 | 0 |
| 2 | 1 | 19 | 28 | 0 | 1 | 0 | 0 | 0 | 22 | 0 |
| 2 | 1 | 19 | 28 | 1 | 1 | 0 | 0 | 22 | 1111 | 0 |
| 3 | 1 | 17 | 28 | 0 | 1 | 0 | 0 | 0 | 34 | 0 |
| 3 | 1 | 17 | 28 | 1 | 1 | 0 | 0 | 34 | 530 | 0 |
| 4 | 1 | 17 | 20 | 0 | 1 | 0 | 0 | 0 | 22 | 0 |
| 4 | 1 | 17 | 20 | 1 | 1 | 0 | 0 | 22 | 1279 | 1 |
| 5 | 1 | 28 | 25 | 0 | 1 | 0 | 0 | 0 | 49 | 0 |
| 5 | 1 | 28 | 25 | 1 | 1 | 0 | 0 | 49 | 110 | 1 |

-----Biometrie\StatMod4\Examp8-1.sas-----

Survival of Leukaemia Patients (Collett, p. 262)

| Variable | Estimate | StdErr | ChiSq | Prob ChiSq |
|----------|----------|---------|--------|---------------|
| group1 | 2.07607 | 1.15104 | 3.2532 | 0.0713 |
| group3 | 2.46550 | 1.20765 | 4.1680 | 0.0412 |
| Plate_t | -2.31956 | 1.22938 | 3.5599 | 0.0592 |

-----Biometrie\StatMod4\Examp8-1.sas-----

→ Interpretation?

Beispiel 5.3 (Chemotherapie bei Ovarial-Krebs)

Nimmt der Behandlungseffekt bei älteren Frauen schneller ab?

Untersuchung einer Zeit \times Alterswechselwirkung in der Behandlung von Patientinnen mit Ovarialkarzinom.

Hier: $x_{Alter}(t) = \text{Alter} \cdot t$.

Survival of Ovarian Cancer Patients (Collett, p. 262)

1

| # | Variable | Type | Len | Pos | Label |
|---|----------|------|-----|-----|----------------------------------|
| 5 | Age | Num | 8 | 32 | Age |
| 1 | Patient | Num | 8 | 0 | |
| 7 | Perf | Num | 8 | 48 | |
| 6 | Rdisease | Num | 8 | 40 | |
| 3 | Status | Num | 8 | 16 | Event (1/0: event/censored) |
| 2 | Time | Num | 8 | 8 | |
| 4 | Treat | Num | 8 | 24 | Treatment (1/2: single/combined) |

| Patient | Time | Status | Treat | Age | Rdisease | Perf |
|---------|------|--------|-------|-----|----------|------|
| 1 | 156 | 1 | 1 | 66 | 2 | 2 |
| 2 | 1040 | 0 | 1 | 38 | 2 | 2 |
| 3 | 59 | 1 | 1 | 72 | 2 | 1 |
| 4 | 421 | 0 | 2 | 53 | 2 | 1 |
| 5 | 329 | 1 | 1 | 43 | 2 | 1 |

-----Biometrie\StatMod4\Examp8-2.sas-----

Survival of Ovarian Cancer Patients (Collett, p. 262)

| Variable | Estimate | StdErr | ChiSq | Prob ChiSq | Criterion | With Covariates |
|----------|----------|---------|---------|---------------|-----------|--------------------|
| Age | 0.14657 | 0.04585 | 10.2174 | 0.0014 | | |
| Treat | -0.79593 | 0.63294 | 1.5813 | 0.2086 | | |
| | | | | | -2 LOG L | 54.148 |

```
-----
```

| Variable | Estimate | StdErr | ChiSq | Prob ChiSq | Criterion | With Covariates |
|----------|------------|-----------|--------|---------------|-----------|--------------------|
| Age | 0.21565 | 0.11261 | 3.6673 | 0.0555 | | |
| Treat | -0.66359 | 0.66955 | 0.9823 | 0.3216 | | |
| Age_t | -0.0002031 | 0.0002832 | 0.5141 | 0.4734 | | |
| | | | | | -2 LOG L | 53.613 |

```
-----Biometrie\StatMod4\Examp8-2.sas -----
```

► Fazit: der Alterseffekt scheint sich mit der Zeit nicht zu ändern.

Beispiel 5.4 (Cirrhose-Studie (hypothetisch))

In einer Studie werden 12 Patienten zu zwei Behandlungsgruppen (Placebo und Liverol) zugelost und das Survival wird überwacht.

- ▶ Bilirubin ist ein wichtiger prognostischer Faktor, der bei jedem Follow-up visit in der Klinik gemessen wird.
- ▶ Als zeitabhängige Kovariable könnte man den jeweils zuletzt verfügbaren Wert von log-Bilirubin verwenden.

Cirrhosis Study (Collett, p. 266)

| # | Variable | Type | Len | Pos | Label |
|----|----------|------|-----|-----|-----------------------------|
| 5 | Age | Num | 8 | 32 | Age (years) |
| 6 | Lbr0 | Num | 8 | 40 | Log Bilirubin (t=0) |
| 8 | Lbr1 | Num | 8 | 56 | Log Bilirubin (t=T1) |
| 10 | Lbr2 | Num | 8 | 72 | |
| 12 | Lbr3 | Num | 8 | 88 | |
| 14 | Lbr4 | Num | 8 | 104 | |
| 16 | Lbr5 | Num | 8 | 120 | |
| 1 | Patient | Num | 8 | 0 | |
| 3 | Status | Num | 8 | 16 | Event (1/0: event/censored) |
| 7 | T1 | Num | 8 | 48 | Time 1 (days) |
| 9 | T2 | Num | 8 | 64 | |
| 11 | T3 | Num | 8 | 80 | |
| 13 | T4 | Num | 8 | 96 | |
| 15 | T5 | Num | 8 | 112 | |
| 2 | Time | Num | 8 | 8 | Survival time (days) |
| 4 | Treat | Num | 8 | 24 | Treatment |

----- Biometrie\StatMod4\Examp8-3.sas -----

Cirrhosis Study (Collett, p. 266)

| # | Variable | Type | Len | Pos | Label |
|---|----------|------|-----|-----|-----------------------------|
| 7 | Age | Num | 8 | 48 | Patient age |
| 5 | Event | Num | 8 | 32 | Event (1/0: event/censored) |
| 4 | Lbr | Num | 8 | 24 | Log Bilirubin |
| 1 | Patient | Num | 8 | 0 | |
| 2 | Time1 | Num | 8 | 8 | Start time |
| 3 | Time2 | Num | 8 | 16 | Stop time |
| 6 | Treat | Num | 8 | 40 | Treatment |

| Patient | Time1 | Time2 | Lbr | Event | Treat | Age |
|---------|-------|-------|-----|-------|-------|-----|
| 1 | 0 | 47 | 3.2 | 0 | 0 | 46 |
| 1 | 47 | 184 | 3.8 | 0 | 0 | 46 |
| 1 | 184 | 251 | 4.9 | 0 | 0 | 46 |
| 1 | 251 | 281 | 5.0 | 1 | 0 | 46 |
| 2 | 0 | 94 | 3.1 | 0 | 0 | 57 |
| 2 | 94 | 187 | 2.9 | 0 | 0 | 57 |
| 2 | 187 | 321 | 3.1 | 0 | 0 | 57 |
| 2 | 321 | 604 | 3.2 | 0 | 0 | 57 |

----- Biometrie\StatMod4\Examp8-3.sas -----

- Es werden Modelle verglichen, in denen zum einen nur der anfängliche Wert von log-Bilirubin Lbr0 enthalten ist, mit Modellen, die den zuletzt verfügbaren Wert von log-Bilirubin Lbrt als zeitveränderliche Kovariable enthalten.

Cirrhosis Study (Collett, p. 266)

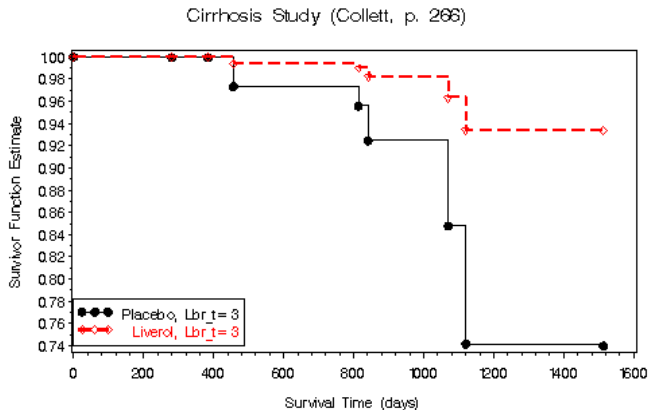
| Obs | Variable | Criterion | Without Covariates | With Covariates |
|-----|----------------|-----------|-----------------------|--------------------|
| 1 | age | -2 LOG L | 25.121 | 22.135 |
| 2 | Lbr0 | -2 LOG L | 25.121 | 21.662 |
| 3 | age Lbr0 | -2 LOG L | 25.121 | 18.475 |
| 4 | age Lbr0 treat | -2 LOG L | 25.121 | 13.293 |
| 5 | Lbr_t | -2 LOG L | 25.121 | 12.050 |
| 6 | age Lbr_t | -2 LOG L | 25.121 | 11.145 |
| 7 | age Lbr0 Lbr_t | -2 LOG L | 25.121 | 11.133 |
| 8 | Lbr_t treat | -2 LOG L | 25.121 | 10.676 |

- ▶ Im Modell 4 ist der Koeffizient von `treat` $-3,052$, d.h.
 $HR(treat) = e^{-3.052} = 0,047$. Der p -Wert zum Vergleich von Modell 3 mit 4 ist 0.023 .
- ▶ Im Modell 8 ist der Koeffizient von `treat` $-1,479$, d.h.
 $HR(treat) = e^{-1.479} = 0,228$. Der p -Wert zum Vergleich von Modell 5 mit 8 ist 0.241 .

—→ Interpretation?

Survivalfunktion?

Die Survivalfunktion kann für eine interne zeitabhängige Kovariable wie Bilirubin nicht sinnvoll betrachtet werden. Es läßt sich allerdings die bedingte Wahrscheinlichkeit $P(T_i > t + h | T_i > t)$ approximieren, indem ähnlich wie zuvor die baseline kumulative Hazardfunktion geschätzt wird. Das Beispiel zeigt dies für zwei (Placebo / Liverol) hypothetische Patienten, für die $Lbrt = 3$ zu allen Zeitpunkten.



Biometrie\StatMod4\Examp8-3.sas

Zusammenhang zum gewichteten Log-rank Test

- ▶ Betrachte binären Gruppenindikator $x_i \in \{0; 1\}$ für Gruppe 1 bzw. 2.
- ▶ Seien $w_j = w(t_{(j)})$ die Gewichte in einem gewichteten Log-rank Test für die Ereigniszeitpunkte $j = 1, \dots, r$.
- ▶ Definiere die zeitabhängige Kovariable

$$x_i(t) = x_i w(t).$$

- ⇒ Es lässt sich zeigen, dass der Score Test im Cox-Modell mit $x_i(t)$ als einziger Kovariable identisch ist mit dem gewichteten Log-rank Test mit Gewichten w_j .
- ⇒ D.h. die Gewichte können im gewichteten Log-rank Test so gewählt werden, dass der Test gegenüber bestimmten zeitabhängigen Abweichungen von der Nullhypothese besonders geeignet ist.

Kausale Interpretation

Speziell in der Survivalanalyse, oft Vermischung zweier Ziele / Aufgaben:

- ▶ best-mögliche **Vorhersage** der Überlebenschancen
- ▶ **kausale** Schlüsse, z.B. welche Behandlung verbessert die Überlebenschancen mehr.

Unterschied?

Vorhersage: wenn zum Zeitpunkt t eine Vorhersage getroffen werden soll, macht es Sinn, alle zum Zeitpunkt t verfügbare 'relevante' (prognostische) Information zu nutzen, also auch zeitabhängige Kovariablen.

Kausale Faktoren: sind solche, in denen eine *Intervention* zu einer Veränderung der Ereignisdauer (z.B. Überleben) führt.

Obacht: soll ein kausaler Effekt einer Behandlung geschätzt werden, darf nicht (implizit oder explizit) auf Folgen/Konsequenzen der Behandlung (z.B. zeitabhängige Kovariablen) bedingt werden (siehe 5-25)!

Ziel: kausaler Behandlungseffekt

Unterscheide:

(i) ist Behandlung¹ eine Baselinevariable (zum Ztp. $t = 0$ gegeben) oder zeit-veränderlich?

(ii) wurde Behandlung randomisiert oder nicht?

⇒ vier Fälle.

(1) Falls **randomisierte Baseline-Behandlung**

⇒ zur Schätzung des Behandlungseffekts werden zeitabhängige Kovariablen ignoriert (sonst Bias!).

Baseline Kovariablen können / sollten ins Modell aufgenommen werden, um es plausibler zu machen, z.B. die PH Annahme.

¹Mit 'Behandlung' muss nicht unbedingt eine Behandlung im medizinischen Sinne gemeint sein; es kann z.B. auch eine Exposition wie 'Rauchen' sein.

(2) Beobachtungsstudie, Baseline-Behandlung

- ▶ Die Behandlung wurde nicht randomisiert
- ▶ Bias durch 'Confounding' kann entgegengewirkt werden, indem geeignete Baseline-Kovariablen (gemessene Confounder) in das Modell aufgenommen werden; dabei ist Plausibilität der PH-Annahme zweitrangig.
- ▶ Zeitabhängige Kovariablen werden wiederum ausgeschlossen.

Methoden zur Confounderadjustierung

- ▶ Kovariablen ins Modell \Rightarrow Koeffizient ist *bedingter* Effekt
- ▶ ... mit 'Standardisierung' \Rightarrow *marginaler* Effekt
- ▶ 'Matching' auf oder 'Gewichtung' mit der Patientenspezifischen Wahrscheinlichkeit, die Behandlung zu erhalten, = 'inverse probability of treatment weighting' (IPTW) \Rightarrow marginaler Effekt.

Nebenbemerkung:

Marginaler (kausaler) Parameter: Kontrast in der Gesamtpopulation, z.B.
“wenn alle rauchen würden versus niemand würde rauchen, dann...”

Bedingter (kausaler) Parameter: Kontrast in einer Teilmenge, z.B. z.B.
“wenn alle nicht-schwangeren Frauen unter 30 rauchen würden versus nicht, dann...”

Wenn (a) keine Interaktionen bestehen und (b) der Parameter ‘kollabierbar’ ist, dann sind beide gleich.

Parameter in linearen Modellen sind kollabierbar; Parameter in Cox- oder logistischen Modellen sind nicht kollabierbar.

IPTW – Motivation

—→ an der Tafel.

Beispiel: Disease-free survival für Ewing Sarcoma Patienten mit neuer versus Standardbehandlung; Serum lactic acid dehydrogenase (LDH) ist starker Prädiktor und wahrsch. Confounder für Behandlung und Survival. (Aus: Cole & Hernan, 2003)

Table 1 Recurrence of Ewing's sarcoma by treatment and lactic acid level ($N = 76$)

| Treatment | LDH ^a | Days to recurrence ^b |
|-----------|------------------|--|
| S4 | 0 | 31, 335, 366, 426, 456, 578, 589, 762+, 792, 913+, 914+, 974, 1005, 1035, 1065+, 1096, 1107+, 1219+, 1250+, 1312+, 1403+, 1461+, 1553+, 1645+, 1706+, 1734+, 1826+, 1948+, 1949+, 1979+, 2222+, 2374+, 2435+, 2465+, 2526+ |
| | 1 | 0, 91, 183, 334, 338, 365, 391, 518, 547, 608, 609, 851 |
| S1–S3 | 0 | 153, 945, 1400, 1887, 2557+, 3134+, 3226+, 3348+, 3501+, 3743+ |
| | 1 | 151, 152, 212, 214, 242, 243, 244, 245, 249, 273, 336, 337, 396, 427, 457, 761, 1249, 1310, 2708+ |

^a Serum lactic acid dehydrogenase, <200 vs. ≥ 200 international units.

^b Censored subjects indicated by +.

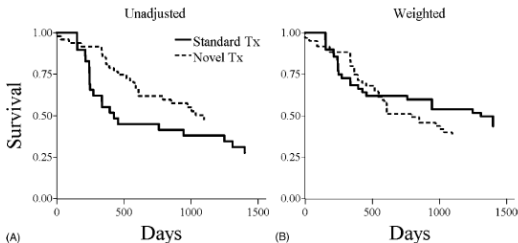


Fig. 1 Panel (A) presents the unadjusted survival curves for 76 Ewing's sarcoma patients. Panel (B) presents the weighted survival curves.

(3) Zeitveränderliche Behandlung, randomisiert

Beispiel: $K + 1$ Gruppen; jeder PatientIn wird zu einer Gruppe randomisiert; Gruppe k : Behandlungsstart ist zum Zeitpunkt t_k ; wobei t_0, \dots, t_K festgelegt sind, z.B. nach k Monaten.

- ▶ Survivalkurven? Skizze!
- ▶ Behandlung: im Prinzip K diskrete Gruppen \Rightarrow betrachte K K-M-Kurven; aber z.B. für $k = 4, 5$ können sich die Kurven zu Zeitpunkten $t \leq t_4$ nicht (systematisch) unterscheiden, also kein PH!
 \Rightarrow zeitabhängige Behandlungsvariable.
- ▶ Confounderadjustierung nicht nötig; Baseline-Kovariablen können / sollten zur Plausibilität des Modells genutzt werden.

Marginales Strukturelles Modell (MSM)

(4) Zeitveränderliche Behandlung, Beobachtungsstudie

Beispiel: frühe Studien (ca. 1990-2000) zur Behandlung von HIV Patienten mit ART waren Beobachtungsstudien; dabei wurden oft erst dann mit der Behandlung angefangen, wenn der CD4-Wert des Patienten als zu niedrig erachtet wurde.

Zudem wurde Behandlungserfolg durch den 'viral load' gemessen, und wenn sich dies nicht besserte, wurde die Behandlung gewechselt ('treatment switch').

Beides, Start der Erstbehandlung und Wechsel der Medikation, sind zeitabhängige Behandlungen. Hierbei sind CD4-Wert und VL zeitabhängige Konfounder.

- ▶ Zeitabhängiges Confounding kann i.d.R. nicht durch die übliche Kovariablenadjustierung vermieden werden, ohne neuen Bias zu schaffen.
- ▶ Eine 'einfache' Methode ist IPTW, wobei die Gewichte nun zeitabhängig sein müssen, um zeitabhängige Kovariablen zu berücksichtigen.

Marginales Strukturelles Modell

Ein MSM ist ein kausales Modell für den Effekt einer (i.d.R. zeitabhängigen) Behandlung.

- ▶ Der Effekt ist so definiert, als ob jedeR PatientIn zu seiner/ihrer Behandlungssequenz randomisiert wurde (siehe Fall (3));
- ▶ Zeitabhängigem Confounding wird durch (zeitabhängige) IPTW entgegen gewirkt.

Sei $x(t)$ die zeitabhängige Behandlung, wobei $\text{do}(\bar{x}(T))$ dafür steht, dass diese durch eine Intervention festgelegt wurde; sei η der lineare Prädiktor bzgl. relevanter Baseline-Kovariablen; das Cox-MSM besagt dann:

$$\lambda^{MSM}(t; \text{do}(x(t))) = \lambda_0(t) \exp\{\eta + \beta x(t)\}.$$

Bemerkungen

- ▶ Ein MSM mit IPTW liefert eine konsistente Schätzung des gewünschten kausalen Effekts, wenn das MSM und das Modell für die Gewichte (i.e. Behandlung in Abhängigkeit der Confounder Variablen zu den versch. Zeitpunkten) korrekt spezifiziert sind.
- ▶ In vielen Anwendungen, wird die Zeit diskretisiert, um die Berechnung der Gewichte zu vereinfachen. Statt des Cox-Modelles wird dann eine gepoolte logistische Regression verwendet, welche eine Approximation des Cox-Modells für diskrete Zeit darstellt.
- ▶ Das Problem der informativen Zensierung läßt sich im Übrigen ähnlich lösen: wenn zeitabhängige Information vorhanden ist, durch die die Zensierung erklärt werden kann, wird der Bias durch 'inverse probability of censoring weighting' IPCW verringert/vermieden.
- ▶ Es können extreme Gewichte auftreten, was zur Instabilität (sehr große Varianz) der Schätzung führt; Methoden: Stabilisierung der Gewichte durch geschickte Wahl des Zählers; Trunkierung der Gewichte, etc.

Beispiel 5.5 (Historisches HIV Beispiel)

Aus Hernan, Brumback, Robins (2000), Epidemiology.

Multizentrische Kohortenstudie in den Jahren 1984-1991, mit 2178 HIV-positiven Männern, anfänglich ohne AIDS und ohne med. Behandlung; bis Ende des follow-up: Zidovudine wurde bei 1296 Männern initiiert, 750 waren verstorben.

Zeitabhängiges Confounding durch CD4 sehr wahrscheinlich, da starker Prädiktor für das Initiieren der Behandlung mit Zidovudine.

Wenn CD4 als 'normale' zeitabhängige Kovariable ins Modell genommen wird (ohne IPTW), ergibt sich ein HR für den Behandlungseffekt von 0.4.

T . For these histories $T_{\bar{a}}$ equals T . For each \bar{a} , we specify the marginal structural Cox proportional hazards model

$$\lambda_{T_{\bar{a}}}(t|V) = \lambda_0(t)\exp(\beta_1 a(t) + \beta_2 V)$$

where $\lambda_{T_{\bar{a}}}(t|V)$ is the hazard of death at t among subjects with baseline covariates V in the source population had, contrary to fact, all subjects followed zidovudine history \bar{a} through time t , the scalar β_1 and the row vector β_2 are unknown parameters, and $\lambda_0(t)$ is an unspecified baseline hazard. We refer to this model as an MSM because,

TABLE 1. Inverse-Probability-of-Treatment Weighted Estimates of the Causal Effect of Zidovudine Therapy on Mortality in the Multicenter AIDS Cohort Study

| Unweighted estimates* | | RR | 95% CI |
|--------------------------|------|------------------------------|-----------------------------------|
| Unadjusted | | 3.55 | 2.95–4.27 |
| Only baseline covariates | | 2.32 | 1.92–2.81 |
| Weighted estimates† | RR | Valid 95% Conservative CI | Invalid Model-Based‡ 95% CI |
| Stabilized weights | 0.74 | 0.57–0.96 | 0.62–0.87 |
| Nonstabilized weights | 0.76 | 0.54–1.05 | 0.71–0.80 |

RR = mortality rate ratio (zidovudine users vs nonusers); CI = confidence interval.

Inhalt von Abschnitt 6

Parametrische Modelle

Exponential- und Weibullverteilung

- Angemessenheit des Weibull Modells

Schätzen im parametrischen Modell

- Schätzen bei Exponentialverteilung

- Schätzen bei Weibull-Verteilung

Modell für den Vergleich zweier Gruppen

- log-kumulativer Hazard Plot

- Schätzen der Modellparameter

Das Weibull Proportionale Hazards Modell

- Parameterschätzung

- Re-Parameterisierung

- Vergleich verschiedener Weibull-Modelle

Das Gompertz Proportionale Hazards Modell

Parametrische Modelle

- ▶ Vorteil des Cox-Modells ist, dass man die zugrunde liegende Hazardfunktion $\lambda_0(t)$ nicht zu spezifizieren braucht, um den Einfluss von Kovariablen auf die Ereignisrate zu schätzen.
- ▶ Wenn man jedoch die Form der Basis-Hazardfunktion kennt (aus substanzwissenschaftlichen Überlegungen oder aus Erfahrung), kann man die Einflussparameter effizienter schätzen, wenn man die spezifizierte Form von $\lambda_0(t)$ benutzt.
- ▶ Andere parametrische Regressionsmodelle die nicht die PH-Annahme machen sind auch möglich.

Zunächst betrachten wir Modelle ohne Einflußgrößen, anschließend mit.

Zur Erinnerung

Sei T eine positive Zufallsvariable (Dauer bis Ereignis).

Ist die Dichte gegeben, dann erhält man:

$$S(t) = 1 - \int_0^t f(u) du$$

und

$$\lambda(t) = \frac{f(t)}{S(t)} = -\frac{d}{dt} \{\log S(t)\}$$

Ist umgekehrt die Hazardfunktion $\lambda(t)$ gegeben, so ergibt sich

$$S(t) = \exp\{-\Lambda(t)\}$$

$$f(t) = \lambda(t)S(t) = -\frac{dS(t)}{dt}$$

mit

$$\Lambda(t) = \int_0^t \lambda(u) du.$$

Exponentialverteilung

Das einfachste parametrische Modell setzt die Hazardfunktion konstant für $0 < t \leq \infty$:

$$\lambda(t) = \lambda, \quad \lambda > 0.$$

Damit ergibt sich

$$\begin{aligned} S(t) &= \exp \left(- \int_0^t \lambda du \right) \\ &= \exp(-\lambda t) \end{aligned} \tag{6.1}$$

und Dichte

$$f(t) = \lambda e^{-\lambda t}.$$

→ $\Lambda(t)$? Plot?

Erwartungswert und Varianz einer exponentialverteilten ZV ergeben sich zu

$$\mathbb{E}(T) = \frac{1}{\lambda}, \quad \mathbb{V}(T) = \frac{1}{\lambda^2}.$$

Das p -Quantil der Ereigniszeitverteilung ist der Zeitpunkt t_p , sodass $S(t_p) = 1 - p$ (siehe Folie 1-15), sodass sich mit (6.1) ergibt

$$t_p = \frac{1}{\lambda} \log \frac{1}{1-p}, \quad (0 < p < 1).$$

Speziell für den Median ($p = 0,5$) ergibt sich

$$t_{med} = \frac{\log 2}{\lambda}.$$

→ andere Eigenschaften?

Weibull-Verteilung $W(\lambda, \gamma)$

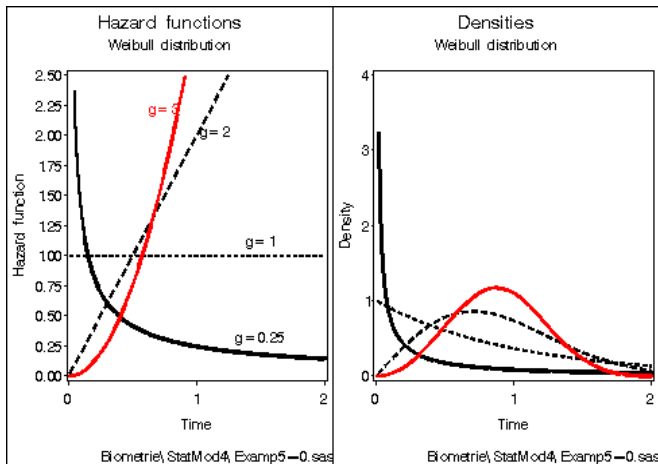
Die Hazardfunktion der Weibull-Verteilung wird durch zwei Parameter spezifiziert — hier $\lambda, \gamma > 0$:

$$\lambda(t) = \lambda \gamma t^{\gamma-1}.$$

- ▶ Es heißen: γ : Formparameter, λ : Skalenparameter.
- ▶ Für $\gamma = 1$ ergibt sich die Exponentialverteilung.
- ▶ Die Hazardfunktion von $W(\lambda, \gamma)$ ist monoton steigend für $\gamma > 1$, monoton fallend für $\gamma < 1$.
- ▶ Die Survivalfunktion $S(t)$ ist gegeben als

$$S(t) = \exp \left(- \int_0^t \lambda \gamma u^{\gamma-1} du \right) = \exp (-\lambda t^\gamma) \quad (6.2)$$

mit Dichte $f(t) = \lambda \gamma t^{\gamma-1} \exp (-\lambda t^\gamma)$.



$\mathbb{E}(T)$ einer $W(\lambda, \gamma)$ -verteilten ZVe T ist

$$\mathbb{E}(T) = \lambda^{-\frac{1}{\gamma}} \Gamma(\gamma^{-1} + 1),$$

wobei

$$\Gamma(x) = \int_0^{\infty} u^{x-1} e^{-u} du$$

die Gamma-Funktion darstellt (für ganzzahlige x : $\Gamma(x) = (x-1)!$).

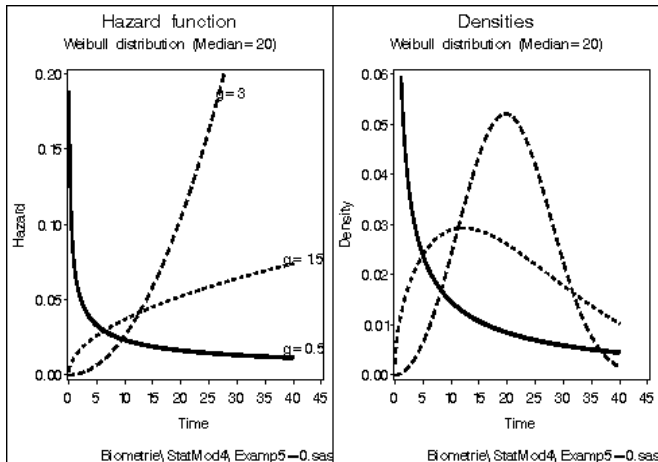
Für das p -Quantil einer Weibull-Verteilung ergibt sich der Zeitpunkt t_p zu

$$t_p = \left\{ \frac{1}{\lambda} \log \frac{1}{1-p} \right\}^{\frac{1}{\gamma}}, \quad 0 < p < 1,$$

Speziell für den Median ($p = 0,5$) ergibt sich

$$t_{med} = \left\{ \frac{1}{\lambda} \log 2 \right\}^{\frac{1}{\gamma}}.$$

Hazardfunktion und Dichte für Weibull-Verteilung mit Median = 20 und Formparameter $\gamma = 0,5; 1,5; 3,0$; $\lambda = 0,15; 0,0078; 0,000\,087$.



Angemessenheit des Weibull Modells

Ist die Ereigniszeit $W(\lambda, \gamma)$ -verteilt, so ist nach (6.2)

$$\begin{aligned}\log(-\log S(t)) &= \log(\lambda t^\gamma) \\ &= \log \lambda + \gamma \log t.\end{aligned}\tag{6.3}$$

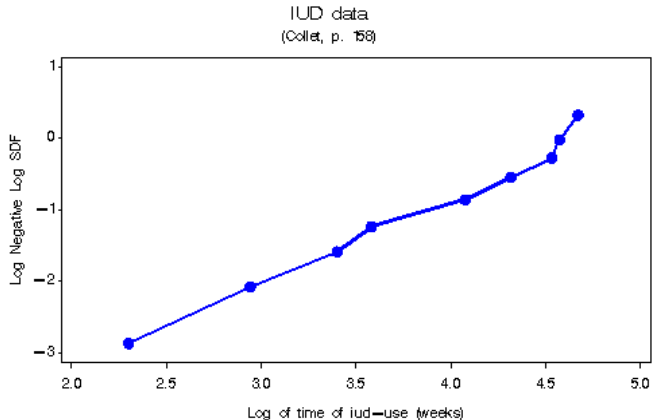
Damit bietet sich für empirische Daten ein Plot $\log(-\log \hat{S}(t))$ gegen $\log(\text{Zeit})$ an.

→ Warum?

→ Spezialfall Exponentialverteilung?

Beispiel 6.1 (IUD-Daten)

Eine entsprechende Grafik zeigt, dass die Ereigniszeit als $W(\lambda, \gamma)$ -verteilt plausibel ist. Man schätzt: $\lambda^* \approx e^{-6} \approx 0,002$ und $\gamma^* \approx 1,25$.



Biometrie\StatMod4\Examp5-1.sas

Schätzen im parametrischen Modell

Wir behandeln zunächst das 1-Stichproben-Problem.

- ▶ Gegeben seien Ereigniszeiten von n unabhängigen Personen, wobei r Ereignisse beobachtet wurden, die restlichen $(n - r)$ Zeiten sind **rechtszensiert**.
- ▶ Es wird angenommen, dass die beobachteten Ereigniszeiten Realisierungen von unabhängigen Zufallsvariablen T_1, \dots, T_n sind, die eine Dichte $f(t)$ besitzen.
- ▶ Die Zensierung sei **uninformativ**.
- ▶ Die Likelihood ist damit voll spezifiziert und wird wie auf Folie 1-17 aufgestellt:

$$L = \prod_{i=1}^n \lambda(t_i)^{\delta_i} S(t_i).$$

Schätzen bei Exponentialverteilung

Mit

$$\lambda(t) = \lambda, \quad S(t) = e^{-\lambda t}$$

ergibt sich

$$L(\lambda) = \prod_{i=1}^n \lambda^{\delta_i} e^{-\lambda t_i}.$$

\Rightarrow

$$\begin{aligned} \log L(\lambda) &= \sum_{i=1}^n \delta_i \log \lambda - \lambda \sum_{i=1}^n t_i \\ &= r \log \lambda - \lambda n \bar{t}, \end{aligned}$$

wobei $r = \#\{\text{Ereignisse}\}$, $\bar{t} = \sum t_i / n$.

Maximieren der Loglikelihood liefert via $\frac{\partial \log L}{\partial \lambda} \stackrel{!}{=} 0$:

$$\hat{\lambda} = \frac{r}{\sum_{i=1}^n t_i} \quad \Rightarrow \quad \hat{\mu} = \frac{1}{\hat{\lambda}} = \frac{\sum_{i=1}^n t_i}{r}.$$

—→ vgl. mit unzensierter Stichprobe?

Die geschätzte Varianz von $\hat{\lambda}$ ergibt sich über die Inverse der Informationsmatrix:

$$\begin{aligned} \frac{\partial^2 \log L(\lambda)}{\partial \lambda^2} &= -\frac{r}{\lambda^2} = I(\lambda) \\ \Rightarrow \quad \mathbb{V}(\hat{\lambda}) &= \left\{ -\mathbb{E} \left(\frac{\partial^2 \log L(\lambda)}{\partial \lambda^2} \right) \right\}^{-1} = \frac{\lambda^2}{r}, \\ \text{SE}(\hat{\lambda}) &= \frac{\hat{\lambda}}{\sqrt{r}}. \end{aligned}$$

Damit kann ein asymptotisches Konfidenzintervall für λ bestimmt werden.

Quantile und ihre KI's erhält man entsprechend, indem man $\hat{\lambda}$ einsetzt:

$$\hat{t}_p = \frac{1}{\hat{\lambda}} \log \left(\frac{1}{1-p} \right)$$
$$\hat{t}_{med} = \frac{1}{\hat{\lambda}} \log 2$$

Der Standardfehler für ein Quantil ergibt sich approximativ via δ -Methode:

$$\text{SE}(\hat{t}_p) = \hat{t}_p / \sqrt{r},$$

Speziell für den Median ergibt sich

$$\text{SE}(\hat{t}_{med}) = \hat{t}_{med} / \sqrt{r}.$$

Konfidenzintervalle für Quantile erhält man besser, indem man die Konfidenzintervalle für die log-Quantile exponiert: Es ist

$$\begin{aligned}\text{SE}(\log \hat{t}_p) &= \hat{t}_p^{-1} \text{SE}(\hat{t}_p) \\ &= \frac{1}{\sqrt{r}}.\end{aligned}$$

Damit ergibt sich ein $(1 - \alpha)$ -KI für das p-Quantil als

$$\exp \left(\log \hat{t}_p \pm u_{1-\frac{\alpha}{2}} / \sqrt{r} \right) = \hat{t}_p \exp \left(\pm u_{1-\frac{\alpha}{2}} / \sqrt{r} \right) .$$

—→ Warum?

Beispiel 6.2 (IUD-Daten)

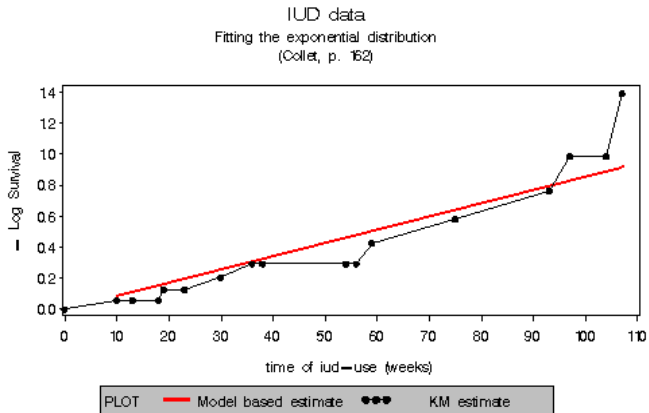
Hier treten 9 Ereignisse auf, die Personenzeit beträgt 1046 Wochen, daher $\hat{\lambda} = 9/1046 = 0,0086/\text{Woche}$. $\hat{\mu} = 1/\hat{\lambda} = 116$ Wochen. Für den Median ergibt sich $\text{Median} \approx 81$ Wochen.

→ Beispiel KI t_{med}

Obacht: Oft wird die Exponential- (und auch die Weibull-) Verteilung anders parameterisiert, in SAS z.B. wird als Scale Parameter $\hat{\mu}$ angegeben, nicht $\hat{\lambda}$.

⇒ Überprüfen Sie immer die von einer Software gewählten Parameterisierung! (Siehe auch später.)

- └ Schätzen im parametrischen Modell
- └ Schätzen bei Exponentialverteilung



Biometrie\StatMod4\Examp5-1sas

Schätzen bei Weibull-Verteilung

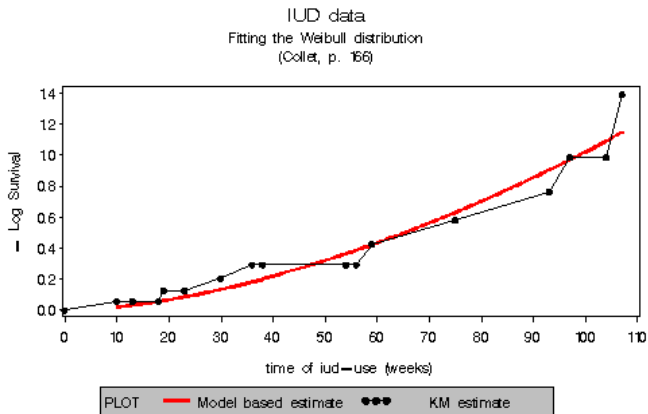
Schätzen der Parameter λ , γ erfolgt gleichfalls via ML, wobei die sich ergebenden Schätzgleichungen ein nichtlineares Gleichungssystem darstellen.

Quantile und ihre SEs sind (mit δ -Methode; siehe später):

$$\begin{aligned}\hat{t}_p &= \left\{ \frac{1}{\hat{\lambda}} \log \left(\frac{1}{1-p} \right) \right\}^{\frac{1}{\hat{\gamma}}} \\ \text{SE}(\hat{t}_p) &= \frac{\hat{t}_p}{\hat{\lambda} \hat{\gamma}^2} \left\{ \hat{\gamma}^2 \mathbb{V}(\hat{\lambda}) + \hat{\lambda}^2 (c_p - \log \hat{\lambda})^2 \mathbb{V}(\hat{\gamma}) \right. \\ &\quad \left. + 2 \hat{\lambda} \hat{\gamma} (c_p - \log \hat{\lambda}) \text{Cov}(\hat{\lambda}, \hat{\gamma}) \right\}^{\frac{1}{2}}\end{aligned}$$

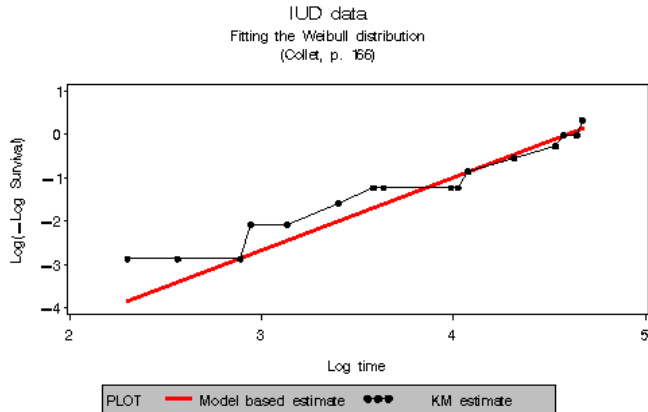
mit $c_p = \log \log \left(\frac{1}{1-p} \right)$.

Beispiel 6.3 (IUD-Daten)



Biometrie\StatMod4\Examp5-1sas

- └ Schätzen im parametrischen Modell
- └ Schätzen bei Weibull-Verteilung



Biometrie\StatMod4\Examp5-1.sas

Modell für den Vergleich zweier Gruppen

Betrachte den einfachen Fall mit einer dichotomen Kovariablen $X \in \{0, 1\}$ und das proportionale Hazards-Modell

$$\lambda(t) = e^{\beta x} \lambda_0(t). \quad (6.4)$$

Ist in Gruppe 0 die Ereigniszeit $W(\lambda, \gamma)$ -verteilt, so ist in Gruppe 1 die Ereigniszeit $W(\psi\lambda, \gamma)$ -verteilt mit $\psi = e^{\beta}$.

Die Verteilungen haben also denselben Formparameter γ .

—→ Warum?

log-Kumulativer Hazard Plot

Gilt die proportionale Hazards Eigenschaft für zwei oder mehr Gruppen $k = 1, \dots, K$, mit einer Basis Weibull-Verteilung, so ergibt sich in einem „log-kumulativen Hazard Plot“ eine Schar paralleler Geraden wegen

$$\begin{aligned} S_k(t) &= \exp(-\psi_k \lambda t^\gamma) \\ &= [\exp(-\lambda t^\gamma)]^{\psi_k} = S_0(t)^{\psi_k}, \end{aligned}$$

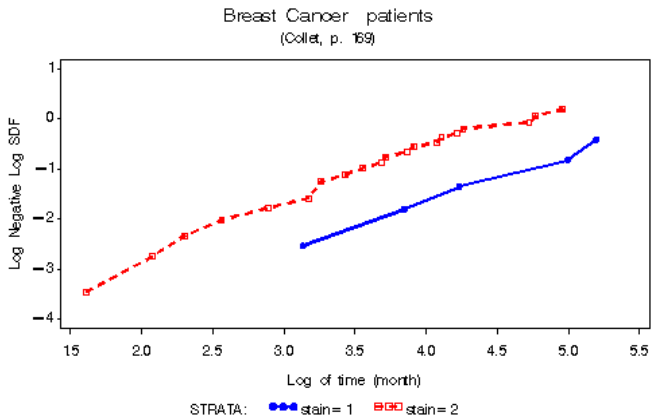
daher

$$\begin{aligned} \log(-\log S_k(t)) &= \log \psi_k + \log \lambda + \gamma \log t \\ &= \log \psi_k + \log(-\log S_0(t)). \end{aligned}$$

→ vgl. frühere log-log-Plots.

- └ Modell für den Vergleich zweier Gruppen
- └ log-kumulativer Hazard Plot

Beispiel 6.4 (Patientinnen mit Brustkrebs)



Biometrie\StatMod4\Examp5-4.sas

Schätzen im Exponentialmodell

Voraus.: 2 Gruppen mit n_i Individuen, (t_{i_j}, δ_{i_j}) seien Ereigniszeiten und Ereignisindikatoren, $i = 1, 2, j = 1, \dots, n_i$.

$$f(t_{i_1}) = \lambda e^{-\lambda t_{i_1}};$$

$$S(t_{i_1}) = e^{-\lambda t_{i_1}}$$

$$f(t_{i_2}) = \psi \lambda e^{-\psi \lambda t_{i_2}};$$

$$S(t_{i_2}) = e^{-\psi \lambda t_{i_2}}$$

Die Likelihood für die $(n_1 + n_2)$ Beobachtungen ergibt sich als

$$L(\lambda, \psi) = \prod_{i_1=1}^{n_1} \lambda^{\delta_{i_1}} e^{-\lambda t_{i_1}} \prod_{i_2=1}^{n_2} (\psi \lambda)^{\delta_{i_2}} e^{-\psi \lambda t_{i_2}}$$

Mit $r_i = \sum_{j=1}^{n_i} \delta_{ji}$, $T_i = \sum_{j=1}^{n_i} t_{ij}$ erhält man die Loglikelihood als

$$\begin{aligned}\log L(\psi, \lambda) &= r_1 \log \lambda - \lambda T_1 + r_2 \log(\psi \lambda) - \psi \lambda T_2 \\ &= (r_1 + r_2) \log \lambda + r_2 \log \psi - \lambda(T_1 + \psi T_2)\end{aligned}$$

Ableiten und Nullsetzen liefert die Schätzgleichungen

$$\frac{r_2}{\widehat{\psi}} - \widehat{\lambda} T_2 = 0 \quad (6.5)$$

$$\frac{r_1 + r_2}{\widehat{\lambda}} - (T_1 + \widehat{\psi} T_2) = 0 \quad (6.6)$$

$$\Rightarrow \widehat{\lambda} = \frac{r_2}{\widehat{\psi} T_2} \quad \widehat{\psi} = \frac{r_2 T_1}{r_1 T_2} \quad \widehat{\lambda} = \frac{r_1}{T_1}. \quad (6.7)$$

Für die Kovarianzmatrix von $(\widehat{\psi}, \widehat{\lambda})$:

$$\frac{\partial^2 \log L}{\partial \psi^2} = -\frac{r_2}{\psi^2}; \quad \frac{\partial^2 \log L}{\partial \lambda \partial \psi} = -T_2; \quad \frac{\partial^2 \log L}{\partial \lambda^2} = -\frac{r_1 + r_2}{\lambda^2}.$$

Daraus
$$I(\psi, \lambda) = \begin{pmatrix} r_2/\psi^2 & T_2 \\ T_2 & (r_1 + r_2)/\lambda^2 \end{pmatrix}$$

und
$$\mathbb{V}(\hat{\psi}, \hat{\lambda}) = \frac{1}{(r_1 + r_2)r_2 - T_2^2\psi^2\lambda^2} \begin{pmatrix} (r_1 + r_2)\psi^2 & -T_2\psi^2\lambda^2 \\ -T_2\psi^2\lambda^2 & r_2\lambda^2 \end{pmatrix}$$

insbesondere
$$\text{SE}(\hat{\psi}) = \sqrt{\frac{(r_1 + r_2)\hat{\psi}^2}{(r_1 + r_2)r_2 - T_2^2\hat{\psi}^2\hat{\lambda}^2}} = \hat{\psi} \sqrt{\frac{r_1 + r_2}{r_1 r_2}}$$

$$\text{SE}(\hat{\lambda}) = \hat{\lambda} / \sqrt{r_1}$$

Die Standardfehler können bei kleinen oder moderaten Stichproben-umfängen nicht direkt zur Konstruktion eines Konfidenzintervalls genutzt werden, da die Normalverteilungsannahme verletzt ist (beide Schätzer haben wegen Positivität schiefe Verteilungen).

Ein Konfidenzintervall für den log-Rate-Ratio Parameter $\beta = \log \psi$ benötigt den Standardfehler von β , den man via δ -Methode erhält:

$$\begin{aligned}\mathbb{V}(\log \hat{\psi}) &\approx \hat{\psi}^{-2} \mathbb{V}(\hat{\psi}) \\ \Rightarrow \quad \text{SE}(\hat{\beta}) &\approx \hat{\psi}^{-1} \text{SE}(\hat{\psi}) = \sqrt{\frac{r_1 + r_2}{r_1 r_2}}\end{aligned}\tag{6.8}$$

Durch Exponieren gewinnt man daraus auch ein KI für ψ .

—→ Das Prinzip läßt sich leicht auf mehr als zwei Gruppen verallgemeinern.

Beispiel 6.5 (Patientinnen mit Brustkrebs)

$$\hat{\psi} = \frac{21 \cdot 1652}{5 \cdot 2679} = 2,59$$

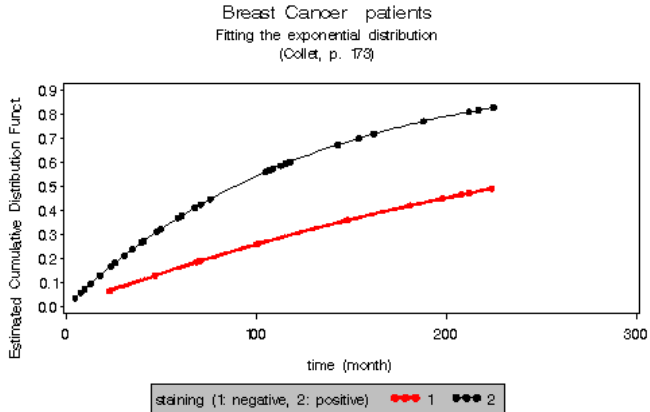
...

In SAS (proc lifereg) wird $\hat{\beta} = 0,9516$ als Schätzwert ausgegeben.

Hier: $SE(\hat{\beta}) = 0,4976 \rightarrow$ nur wenig kleiner als im Cox-Modell (Folie 3-30)!

→ Diskussion!

- └ Modell für den Vergleich zweier Gruppen
- └ Schätzen der Modellparameter



Biometrie\StatMod4\Examp5-4.sas

Das Weibull Proportionale Hazards Modell

Das Seien X_1, \dots, X_p Kovariable, so ergibt sich das Weibull-Modell als

$$\begin{aligned}\lambda(t \mid \mathbf{X} = \mathbf{x}) &= \exp(\beta_1 x_1 + \dots + \beta_p x_p) \lambda_0(t) \\ &= \exp(\boldsymbol{\beta}^T \mathbf{x}) \lambda_0(t) \\ \text{mit} \quad \lambda_0(t) &= \lambda \gamma t^{\gamma-1},\end{aligned}\tag{6.9}$$

$\lambda_0(t)$ ist die Hazardfunktion für jemanden mit $\mathbf{x} = \mathbf{0}$. Die entsprechende Survivalfunktion ergibt sich dann als (mit $S_0(t) = \exp(-\lambda t^\gamma)$)

$$\begin{aligned}S(t \mid \mathbf{X} = \mathbf{x}) &= \exp\left(-\exp(\boldsymbol{\beta}^T \mathbf{x}) \lambda t^\gamma\right) \\ &= S_0(t)^{\exp(\boldsymbol{\beta}^T \mathbf{x})}.\end{aligned}$$

Parameterschätzung

Im Modell (6.9) ist der Parametervektor $\boldsymbol{\theta}^T = (\beta_1, \dots, \beta_p, \lambda, \gamma)$ zu schätzen. Die Likelihood für das hier betrachtete Szenario lautet dann

$$L(\boldsymbol{\theta}) = \prod_{i=1}^n \lambda_i(t_i)^{\delta_i} S_i(t_i).$$

Für die Loglikelihood ergibt sich mit $\eta_i = \boldsymbol{\beta}^T \mathbf{x}_i$

$$\begin{aligned} l(\boldsymbol{\theta}) &= \sum_{i=1}^n [\delta_i \{ \eta_i + \log(\lambda\gamma) + (\gamma - 1) \log t_i \} - \lambda e^{\eta_i} t^\gamma] \\ &= \sum_{i=1}^n [\delta_i \{ \eta_i + \log(\lambda\gamma) + \gamma \log t_i \} - \lambda e^{\eta_i} t^\gamma] - \sum_{i=1}^n \delta_i \log t_i. \quad (6.10) \end{aligned}$$

Weil der letzte Term in (6.10) nicht von dem Parameter θ abhängt, kann er für die Maximierung der log-Likelihood außer Acht gelassen werden, d.h. man maximiert

$$l(\theta) = \sum_{i=1}^n [\delta_i \{\eta_i + \log(\lambda\gamma) + \gamma \log t_i\} - \lambda e^{\eta_i} t^\gamma] \quad (6.11)$$

Hier: keine weiteren Details.

→ Vgl. mit Cox-Modell (3-21)!

→ Weibull, da parametrisch, i.d.R. effizientere Schätzung; aber natürlich Bias, wenn param. Annahme falsch.

Obacht: Computer-Software (inkl. SAS) benutzt meist eine andere Form des Modells, um ein Weibull-PH-Modell anzupassen. Dabei wird ein log-lineares Modell benutzt, das zu einer Survivalfunktion der Form

$$S(t \mid \mathbf{X} = \mathbf{x}) = \exp \left[- \exp \left(\frac{\log t - \mu - \boldsymbol{\alpha}^T \mathbf{x}}{\sigma} \right) \right]$$

führt, d.h. die log-Ereigniszeiten sind “extremwert-verteilt”.

Die Umrechnung der Parameter zwischen den beiden Modellformen geschieht via

$$\lambda = \exp(-\mu/\sigma), \quad \gamma = 1/\sigma, \quad \beta_i = -\alpha_i/\sigma, \quad j = 1, \dots, p.$$

(μ und σ heißen oft „Intercept“ und „Scale“).

—> Warum?

Ist man an der PH Parametrisierung interessiert, so hat man die Parameterschätzungen und insbesondere auch ihre Varianzen zu transformieren. Zur Berechnung der Varianz benutzt man wiederum die δ -Methode: Für eine Funktion $g(\boldsymbol{\theta})$ eines Zufallsvektors $\boldsymbol{\theta}$ mit Kovarianzmatrix $\Sigma_{\boldsymbol{\theta}}$ lautet diese

$$\mathbb{V}(g(\widehat{\boldsymbol{\theta}})) = \left(\frac{\partial g}{\partial \widehat{\boldsymbol{\theta}}} \right)^T \widehat{\Sigma}_{\boldsymbol{\theta}} \left(\frac{\partial g}{\partial \widehat{\boldsymbol{\theta}}} \right),$$

im Fall von zwei Variablen θ_1, θ_2 also

$$\begin{aligned} \mathbb{V}(g(\widehat{\boldsymbol{\theta}})) &= \left(\frac{\partial g}{\partial \widehat{\theta}_1} \right)^2 \mathbb{V}(\widehat{\theta}_1) + \left(\frac{\partial g}{\partial \widehat{\theta}_2} \right)^2 \mathbb{V}(\widehat{\theta}_2) \\ &\quad + 2 \left(\frac{\partial g}{\partial \widehat{\theta}_1} \frac{\partial g}{\partial \widehat{\theta}_2} \right) \text{Cov}(\widehat{\theta}_1, \widehat{\theta}_2). \end{aligned} \tag{6.12}$$

Angewendet auf die Funktion $g(\hat{\alpha}, \hat{\sigma}) = -\hat{\alpha}/\hat{\sigma}$ liefert das

$$\frac{\partial g}{\partial \hat{\alpha}} = -\frac{1}{\hat{\sigma}}, \quad \frac{\partial g}{\partial \hat{\sigma}} = \frac{\hat{\alpha}}{\hat{\sigma}^2}.$$

In (6.12) eingesetzt, ergibt sich

$$\begin{aligned} \mathbb{V} \left(\frac{\hat{\alpha}}{\hat{\sigma}} \right) &\approx \left(-\frac{1}{\hat{\sigma}} \right)^2 \mathbb{V}(\hat{\alpha}) + \left(\frac{\hat{\alpha}}{\hat{\sigma}^2} \right)^2 \mathbb{V}(\hat{\sigma}) \\ &\quad + 2 \left(-\frac{1}{\hat{\sigma}} \right) \left(\frac{\hat{\alpha}}{\hat{\sigma}^2} \right) \text{Cov}(\hat{\alpha}, \hat{\sigma}). \\ &= \frac{1}{\hat{\sigma}^4} \{ \hat{\sigma}^2 \mathbb{V}(\hat{\alpha}) + \hat{\alpha}^2 \mathbb{V}(\hat{\sigma}) - 2\hat{\alpha}\hat{\sigma} \text{Cov}(\hat{\alpha}, \hat{\sigma}) \} \end{aligned}$$

Beispiel 6.6 (Patientinnen mit Brustkrebs)

Die Hazardfunktionen für das Weibull-Modell lauten dann

$$\lambda(t \mid X = x) = \begin{cases} \lambda \gamma t^{\gamma-1} & \text{für } x = 0 \\ \psi \lambda \gamma t^{\gamma-1} & \text{für } x = 1 \end{cases}$$

Dies läßt sich auch schreiben als $\psi = \exp(\beta x)$.

Man erhält:

Breast Cancer patients

Fitting the Weibull distribution (Collet, p. 174)

| Parameter | DF | Estimate | StdErr | LowerCL | UpperCL | ChiSq | Prob ChiSq |
|---------------|----|----------|--------|---------|---------|--------|---------------|
| Intercept | 1 | 5.8544 | 0.4989 | 4.8766 | 6.8321 | 137.71 | <.0001 |
| stain | 1 | -0.9967 | 0.5441 | -2.0631 | 0.0697 | 3.36 | 0.0670 |
| Scale | 1 | 1.0668 | 0.1786 | 0.7684 | 1.4810 | - | - |
| Weibull Shape | 1 | 0.9374 | 0.1569 | 0.6752 | 1.3014 | - | - |

-----Biometrie\StatMod4\Examp5-4.sas-----

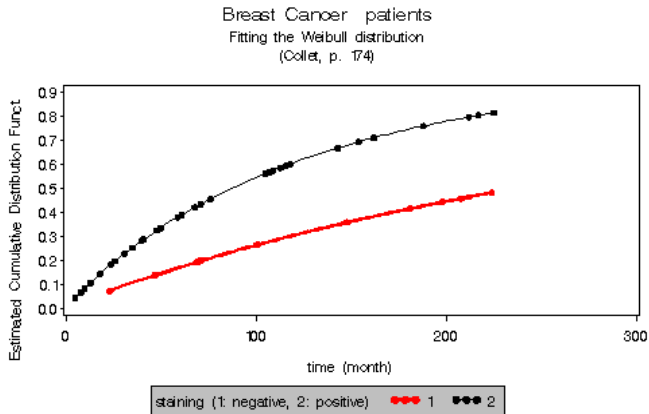
Dabei muss man die Parameter umrechnen:

$$\begin{aligned}
 \hat{\mu} &= 5,8544 & \hat{\lambda} &= 0,00414 = \exp[-\text{„Intercept“}/\text{„Scale“}] \\
 \hat{\alpha} &= -0,9967 & \rightarrow \hat{\gamma} &= 0,937 = [1/\text{„Scale“}] = \text{„Weibull Shape“} \\
 \hat{\sigma} &= 1,0668 & \hat{\beta} &= 0,9343 = -\text{„Stain“}/\text{„Scale“}
 \end{aligned}$$

und $\hat{\psi} = \exp(\hat{\beta}) = 2,55$.

Ausserdem: $\mathbb{V}(\hat{\beta}) \approx 0,2498$.

Beachte: $\gamma = 1$ wäre konstante Hazardfunktion; hier $\hat{\gamma} < 1 \Rightarrow ?$



Biometrie\StatMod4\Examp5-4.sas

Vergleich verschiedener Weibull-Modelle

Beim Vergleich hierarchischer Modelle geht man wie bei anderen Regressionsmodellen vor. Bezeichnet M_1 ein Modell mit p Einflussvariablen, M_2 ein Modell mit zusätzlichen q Variablen, so berechnet man (Devianz bzw. LR Test)

$$X^2 = -2[\log \hat{L}_1 - \log \hat{L}_2],$$

wobei \hat{L}_i die maximierten Likelihoodfunktionen in Modell M_i darstellen, $i = 1, 2$. Unter H_0 : „die zusätzlichen q Parameter sind Null“ ist $X^2 \sim \chi_q^2$.

Vergleich verschiedener Weibull-Modelle

Beispiel 6.7 (Survival of Kidney Cancer Patients)

Die Daten wurden bereits mit dem Cox-Modell ausgewertet, siehe Kapitel 3.

| Modell | Variable |
|--------|--|
| 0 | keine Kovariable |
| 1 | Altersgruppen |
| 2 | Behandlung |
| 3 | Altersgruppen + Behandlung |
| 4 | Altersgruppen + Behandlung + Wechselw. |

Vergleich verschiedener Weibull-Modelle (Forts.)

Survival of Kidney Cancer Patients
(Collett, p. 184)

| Criterion | -2 log Likelihood | Variable |
|-----------|----------------------|---|
| -2 log L | 104.885 | |
| -2 log L | 96.400 | age2 age3 |
| -2 log L | 94.384 | nephrectomy |
| -2 log L | 87.758 | nephrectomy age2 age3 |
| -2 log L | 83.064 | nephrectomy age2 age3 i_n_age2 i_n_age3 |

----Biometrie\StatMod4\Examp5-9.sas----

Vergleich verschiedener Weibull-Modelle (Forts.)

Estimates in model 4

| Parameter | DF | Estimate | StdErr | LowerCL | UpperCL | ChiSq | Prob ChiSq |
|---------------|----|----------|--------|---------|---------|-------|---------------|
| Intercept | 1 | 2.5590 | 0.3727 | 1.8285 | 3.2894 | 47.15 | <.0001 |
| Nephrectomy | 1 | 1.5676 | 0.4095 | 0.7649 | 2.3702 | 14.65 | 0.0001 |
| Age2 | 1 | 0.0548 | 0.5259 | -0.9759 | 1.0855 | 0.01 | 0.9170 |
| Age3 | 1 | -0.0741 | 0.7437 | -1.5317 | 1.3836 | 0.01 | 0.9207 |
| I_N_Age2 | 1 | -0.0781 | 0.6056 | -1.2649 | 1.1088 | 0.02 | 0.8974 |
| I_N_Age3 | 1 | -1.6334 | 0.8283 | -3.2568 | -0.0100 | 3.89 | 0.0486 |
| Scale | 1 | 0.6436 | 0.0938 | 0.4837 | 0.8563 | - | - |
| Weibull Shape | 1 | 1.5538 | 0.2264 | 1.1678 | 2.0674 | - | - |

----Biometrie\StatMod4\Examp5-9.sas----

Das Gompertz Proportionale Hazards Modell

Das Gompertz-Modell (mit Anwendungen insbesondere in der Demographie) hat Hazardfunktion

$$\lambda(t) = \lambda e^{\theta t}, \quad \lambda > 0.$$

Für $\theta = 0$ erhält man wiederum konstanten Hazard, d.h. die Ereigniszeit ist exponentialverteilt. Für die Hazardfunktion der Gompertzverteilung gilt

$$\log \lambda(t) = \log \lambda + \theta t,$$

im Unterschied zur Weibullverteilung, wo

$$\log \lambda(t) = \log \lambda + \log \gamma + (\gamma - 1) \log t.$$

Für Survivalfunktion $S(t)$ und Dichte $f(t)$ erhält man

$$S(t) = \exp\left(\frac{\lambda}{\theta}(1 - e^{\theta t})\right)$$
$$f(t) = \lambda e^{\theta t} \exp\left(\frac{\lambda}{\theta}(1 - e^{\theta t})\right).$$

Auch das Gompertz-Modell kann zu einem Proportionalen Hazards Modell ausgebaut werden via

$$\lambda(t \mid X = x) = \exp(\beta^T x) \lambda e^{\theta t}$$

Die Gompertzverteilung wird nicht von der SAS-Prozedur PROC LIFEREG unterstützt.

Inhalt von Abschnitt 7

Additive Hazard Modelle

- Das additive Hazard Modell

- Schätzen im additiven Hazard Modell

- Datenbeispiele

 - Prognose bei Brustkrebs bei Frauen

 - Leber-Erkrankung (Mayo Liver Disease Data)

- Ausblick

Additives Hazard Modell (Aalen, 1980, 1989)

- ▶ Bisher: Cox- und andere (parametrische) Regressionsmodelle, alle beruhend auf der Annahme, dass Variablen sich multiplikativ auf die Baselinehazardfunktion auswirken;
 - ▶ wenn nicht multiplikativ, dann Stratifizierung oder Zeitabhängigkeit einbauen.
 - ▶ Jetzt: alternative Modellierung: additives Hazard Modell — Variablen wirken sich additiv auf die Hazardfunktion aus.
 - ▶ Additives Hazard Modell: auch semiparametrisch (manchmal als nicht-parametrisch bezeichnet).
- Diskussion: Vor- / Nachteile?

Additives Hazard Modell (Aalen, 1980, 1989) (Forts.)

Datenstruktur (Ereignisse, Zensierungen) wie bisher, seien X_1, \dots, X_p die Einflußgrößen.

Das additive Hazard Modell nimmt an, dass

$$\lambda(t; \mathbf{X} = \mathbf{x}) = b_0(t) + b_1(t)x_1 + \dots + b_p(t)x_p,$$

wobei $b_k(t)$ Koeffizienten sind, die zeitabhängig sein dürfen.

Hier ist $b_0(t)$ wieder eine Baselinehazardfunktion für den Fall, dass $x_1 = \dots = x_p = 0$.

Additives Hazard Modell (Aalen, 1980, 1989) (Forts.)

Noch allgemeiner wird das Modell i.d.R. für zeitabhängige Variablen aufgeschrieben:

$$\lambda(t; \mathbf{X}(t) = \mathbf{x}(t)) = b_0(t) + \sum_{k=1}^p b_k(t)x_k(t),$$

- **Interpretation:** Wenn alle $x_k(t) = 0$ Sinn macht, dann ist $b_0(t)$ die Baseline-Hazardrate, sprich: man erwartet in der Baseline $b_0(t)$ Ereignisse pro Zeiteinheit zum Zeitpunkt t .
- $b_k(t)$ sind die zusätzlich erwarteten Ereignisse pro Zeiteinheit zum Ztp. t , wenn alle anderen Kovariablen zu t gleich sind und $x_k(t)$ für zwei Individuen um eine Einheit verschieden sind.
- Wenn $x_k(t) = 0$ keinen Sinn macht (z.B. $X_k = \text{Blutdruck}$), dann Zentrieren in Betracht ziehen.

Einfaches Beispiel

- ▶ Beispiel: sei X_1 binär; dann:

$$\lambda(t; X_1 = x_1) = b_0(t) + b_1(t)x_1.$$

- ▶ Also, zwei Gruppen:
für $x_1 = 0$ ist $\lambda_0(t) = b_0(t)$;
für $x_1 = 1$ ist $\lambda_1(t) = b_0(t) + b_1(t)$.

→ Diskussion. Survivalfunktionen?

Warum *additiv*?

Hazardfunktion muss positiv (nicht-negativ) sein \Rightarrow additive Modellierung eher ungewöhnlich, da sie negative Werte erlaubt. Aber:

- ▶ Aus inhaltlichen Gesichtspunkten erscheint ein additiver Ansatz manchmal besser geeignet (Risiken addieren sich auf) als ein multiplikativer.
- ▶ Bei multiplikativen Effekten (wie beim Cox Modell) ist es nicht automatisch klar, ob der Effekt wirklich substanziell ist, da dies von dem Baseline-Risiko abhängt. Bei additiven Effekten ('risk differences') ist dagegen direkt klar, um wieviel sich das Risiko erhöht.
- ▶ Statistische Inferenz (Schätzen etc.) ist dafür, dass das additive Hazard Modell sehr allgemein ist, einfach. Eine vergleichbare Verallgemeinerung des Cox Modells ist sehr viel komplizierter.
- ▶ Bei fehlenden / zu vielen Variablen (oder auch additiven Messfehlern) bleibt die additive Struktur erhalten, während die PH Struktur zerstört wird (siehe Folie 3-11; Aalen, 1989).

Kleinste-Quadrate Methode

Zur Erinnerung: für die Hazardfunktion gilt (für 'kleines' Δt)

$$\lambda(t)\Delta t \approx \Pr(t \leq T < t + \Delta t | T \geq t).$$

Sei $N(t)$ eine Indikatorzufallsvariable dafür, dass ein Ereignis vor t stattgefunden hat. Sei $dN(t) = N(t + \Delta t) - N(t)$. Dann gilt (wenn noch 'at risk', also gegeben $N(t) = 0$):

$$\mathbb{E}(dN(t)) \approx \lambda(t)\Delta t,$$

da $\mathbb{E}(dN(t)) = \Pr(dN(t) = 1)$; und mit dem additiven Modell (wenn noch 'at risk')

$$\mathbb{E}(dN(t) | \mathbf{X}(t) = \mathbf{x}(t)) \approx b_0(t)\Delta t + \sum_{k=1}^p b_k(t)\Delta t \cdot x_k(t).$$

D.h. im Erwartungswert folgen die *Sprünge* einem linearen Regressionsmodell mit Koeffizienten $b_k(t)\Delta t$.

Kleinste-Quadrate Methode (Forts.)

Seien die kumulativen Regressionskoeffizienten gegeben als

$$B_k(t) = \int_0^t b_k(u) du,$$

so dass $dB_k(t) = b_k(t)\Delta t$.

Anwendung des Kleinste-Quadrate Prinzips ergibt als Schätzer:

$$d\hat{\mathbf{B}}(t) = \{\mathbf{X}(t)^\top \mathbf{X}(t)\}^{-1} \mathbf{X}(t)^\top d\mathbf{N}(t).$$

Dabei sind:

$d\mathbf{B}(t)$, $d\mathbf{N}(t)$ die entsprechenden Vektoren zu den n Beobachtungen,
 $\mathbf{X}(t)$ die Designmatrix, bestehend aus $R_i(t)x_{ik}(t)$ mit $R_i(t)$ als 'at risk'
Indikatoren (d.h. Indikator dafür, ob Person i das Ereignis bis t noch nicht
erlebt hat und nicht zensiert ist).

Kleinste-Quadrate Methode (Forts.)

Für Ereigniszeitpunkte $t_{(1)}, \dots, t_{(r)}$ und für die kumulierten Zuwächse $B(t)$ läßt sich zeigen:

$$\hat{B}(t) = \sum_{t_{(j)} \leq t} (\mathbf{X}(t_{(j)})^\top \mathbf{X}(t_{(j)}))^{-1} \mathbf{X}(t_{(j)})^\top \Delta \mathbf{N}(t_{(j)})$$

wobei $\Delta \mathbf{N}(t_{(j)})$ ein Vektor ist, der aus Nullen und einer einzelnen Eins für das Individuum mit Ereignis zu $t_{(j)}$ besteht.

Hierbei wird angenommen, dass $\mathbf{X}(t_{(j)})$ invertierbar ist (dies hat damit zu tun, ob hinreichend Individuen 'at risk' sind), und dass es keine Bindungen gibt.

Bemerkungen

- ▶ Während $d\hat{B}(t)$ keine stabilen Schätzer sind (und nicht wirklich benutzt werden), sind die kumulierten Koeffizienten $\hat{B}(t)$ stabil und werden üblicherweise als Schätzer für das additive Modell berichtet.
- ▶ Im Spezialfall einer einzelnen binären Variable $X_1 \in \{0, 1\}$ ist der sich oben ergebende Schätzer für $\hat{B}_0(t)$ der Nelson-Aalen Schätzer $\tilde{\Lambda}_0(t)$ (s. Folie 2-11) für die Referenzgruppe, und $\hat{B}_1(t)$ ist die Differenz der N-A-Schätzer der beiden Gruppen.
- ▶ $\hat{B}_k(t)$ sind Stufenfunktionen, nicht notw. monoton. Diese werden üblicherweise als Ergebnis geplottet. Zur Interpretation ist interessant, ob und wie sich die *Steigung über die Zeit verändert*. → Warum?
- ▶ Bei der Herleitung der (asy.) Verteilung von $\hat{B}(t)$ ist es hilfreich, dass die Zuwächse unabhängig voneinander sind.

Bemerkungen (Forts.)

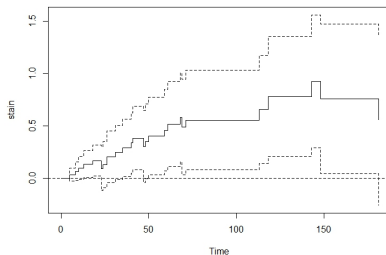
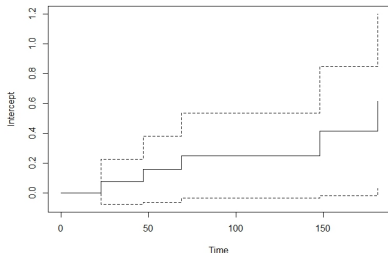
- ▶ $\hat{B}(t)$ ist (beinahe) unverzerrt für $B(t)$; Probleme ergeben sich nur dann, wenn die Wahrscheinlichkeit, dass $X(u)$, $u \in [0, t]$ nicht invertierbar ist, nicht nahe Null ist.
- ▶ Mit der Theorie Stochastischer Prozesse, läßt sich zeigen, dass $\hat{B}(t)$ asymptotisch Normalverteilt ist. Daraus werden z.B. die üblichen (punktweisen) Konfidenzintervalle hergeleitet.
- ▶ Die Schätzung der Koeffizienten $b_k(t)$ beruht auf Glättungsverfahren für $\hat{B}_k(t)$ (im Prinzip: Ableitung von $B_k(t)$, aber da $\hat{B}_k(t)$ Stufenfunktion \rightarrow glätten).

Beispiel 7.1 (Prognose bei Brustkrebs bei Frauen)

“Survival times (in months) after mastectomy of women with Grade 2-4 tumours that were negatively ($X = 0$) or positively ($X = 1$) stained with HPA.”

Dargestellt sind die KQ-Schätzer $\hat{B}_0(t)$ und $\hat{B}_1(t)$ inkl. Konfidenzintervalle.

→ Interpretation?



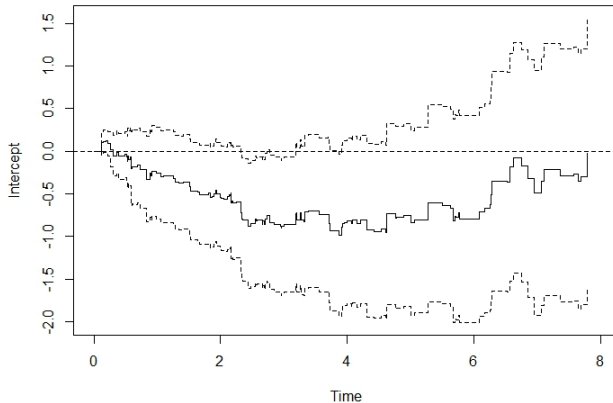
Beispiel 7.2 (Leber-Erkrankung (Mayo Liver Disease Data))

Analyse mit additivem Hazardmodell; fünf Variablen: Age, Albumin, Bilirubin, Protime, Edema. $n = 418$.

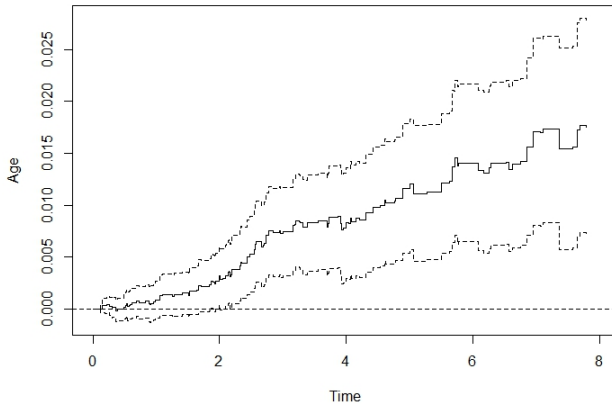
Mayo Liver Data

| Obs | Time | Status | Age | Albumin | Bilirubin | Protime | Edema |
|-----|---------|--------|---------|---------|-----------|---------|-------|
| 1 | 1.0951 | 1 | 58.7652 | 2.60 | 14.5 | 12.2 | 1.0 |
| 2 | 12.3203 | 0 | 56.4463 | 4.14 | 1.1 | 10.6 | 0.0 |
| 3 | 2.7707 | 1 | 70.0726 | 3.48 | 1.4 | 12.0 | 0.5 |
| 4 | 5.2704 | 1 | 54.7406 | 2.54 | 1.8 | 10.3 | 0.5 |
| 5 | 4.1177 | 0 | 38.1054 | 3.53 | 3.4 | 10.9 | 0.0 |
| 6 | 6.8528 | 1 | 66.2587 | 3.98 | 0.8 | 11.0 | 0.0 |
| 7 | 5.0157 | 0 | 55.5346 | 4.09 | 1.0 | 9.7 | 0.0 |
| 8 | 6.7515 | 1 | 53.0568 | 4.00 | 0.3 | 11.0 | 0.0 |
| 9 | 6.5708 | 1 | 42.5079 | 3.08 | 3.2 | 11.0 | 0.0 |
| 10 | 0.1396 | 1 | 70.5599 | 2.74 | 12.6 | 11.5 | 1.0 |

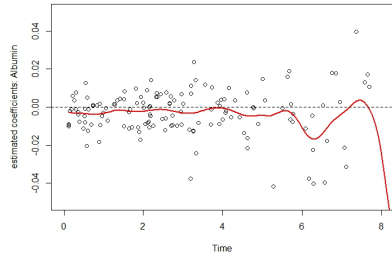
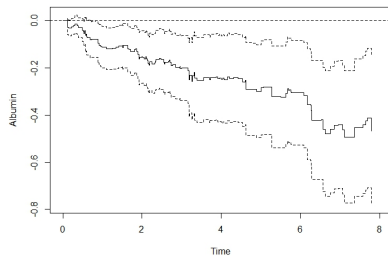
$\hat{B}_0(t)$ = Baseline kum. Hazard — Interpretation?



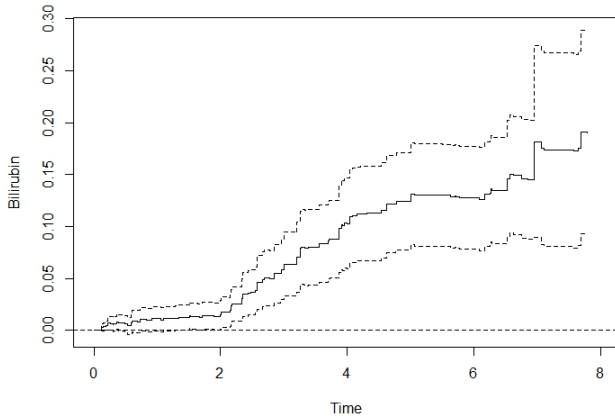
$\hat{B}_1(t)$ = kum. Hazard für 'Age' (in Jahren).



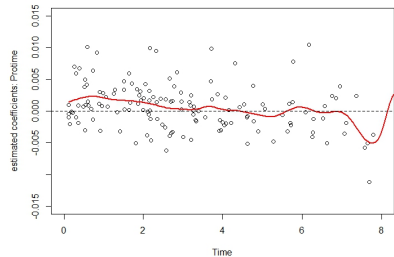
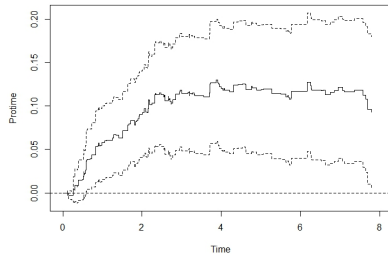
$\hat{B}_2(t)$ = kum. Hazard und $\hat{b}_2(t)$ = für 'Albumin' (mg/dl).



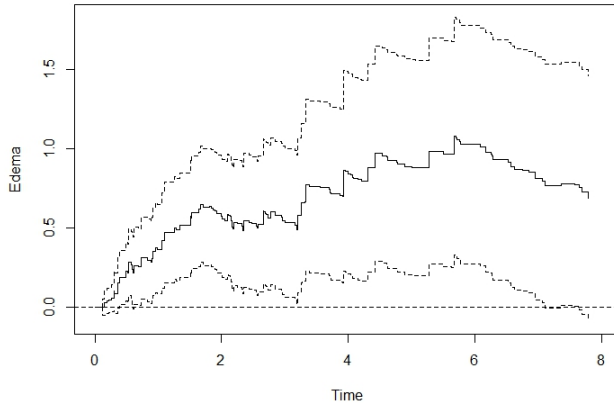
$\hat{B}_3(t)$ = kum. Hazard für 'Bilirubin' (mg/dl).



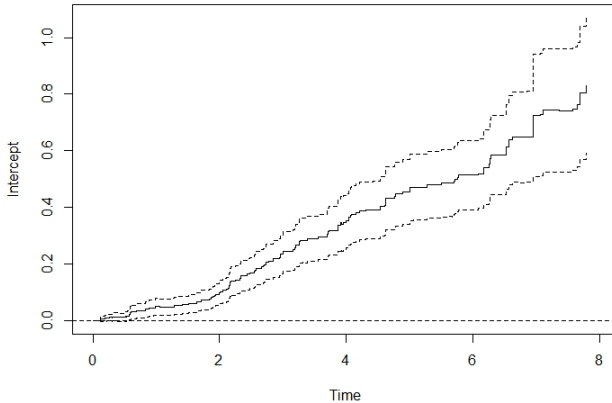
$\hat{B}_4(t)$ = kum. Hazard und $\hat{b}_4(t)$ = für 'Prothrombine time' (in s).



$\hat{B}_5(t)$ = kum. Hazard für 'Edema' (binär).



Nach Zentrierung aller stetigen Variablen: $\hat{B}_0(t)$ = Baseline kum. Hazard.



Ausblick: Gesamthazard

- ▶ Bei fixen Kovariablen läßt sich die kum. Hazardfunktion aus den einzelnen Koeffizienten wie folgt schätzen:

$$\hat{\Lambda}(t|\mathbf{X} = \mathbf{x}) = \hat{B}_0(t) + \sum \hat{B}_k(t)x_k. \quad (7.1)$$

- ▶ Aus $\hat{\Lambda}(t|\mathbf{X} = \mathbf{x})$ läßt sich die entsprechende Survivalfunktion entweder nach dem N-A-Prinzip oder nach dem K-M-Prinzip konstruieren. Aber: $\hat{\Lambda}(t|\mathbf{X} = \mathbf{x})$ ist nicht notw. monoton wachsend; es gibt einfache Verfahren, dies zu beheben.

Ausblick: Tests

- Tests: eine typische Hypothese wäre $H_{0,q} b_q(t) = 0, \forall t \in [0, t^*]$, wobei t^* eine sinnvolle Obergrenze für das zu betrachtende Zeitintervall ist. Eine Teststatistik läßt sich ähnlich wie im linearen Modell herleiten:

$$U(t^*) = \sum_{t_{(j)} \leq t^*} W_q(t_{(j)}) \Delta \hat{B}_q(t_{(j)}),$$

wobei $W_q(t_{(j)})$ Gewichte sind und $\Delta \hat{B}_q(t_{(j)})$ die Zuwächse in $\hat{B}_q(t)$ zu den Zeitpunkten $t_{(j)}$. Mit anderen Worten: die Teststatistik ist ein gewichtetes Mittel der Zuwächse.

Ähnlich wie beim Log-Rank Test, können die Gewichte nach unterschiedlichen Prinzipien gewählt werden, je nachdem, gegen welche Form der Alternative $H_{1,q}$ man testen möchte. Eine typische Wahl ist, die $W_q(t_{(j)})$ proportional zu den entsprechenden inversen

Ausblick: Tests (Forts.)

Diagonalelementen von $(\mathbf{X}(t_{(j)})^\top \mathbf{X}(t_{(j)}))^{-1}$ zu wählen (vgl. gewichtete KQ-Methode).

Unter der Nullhypothese ist die obige Teststatistik, nach Standardisierung, asymptotisch standard-normalverteilt.

- Wenn man eine Menge $q \in Q$ von Parametern testen möchte, $H_{0,Q} b_q(t) = 0, \forall q \in Q, t \in [0, t^*]$, läßt sich obiges Prinzip unter einbeziehung der Kovarianzen erweitern.

Beispiel, Brustkrebs Daten (Forts.): Ein Test, dass

$H_{0,1} b_1(t) = 0, \forall t \in [0, t^*]$, wobei $b_1(t)$ der Koeffizient von `stain` ist, hat einen p -Wert von 0.0271. Dabei wurden die Gewichte proportional zu den entsprechenden inversen Diagonalelementen von $(\mathbf{X}(t_{(j)})^\top \mathbf{X}(t_{(j)}))^{-1}$ gewählt ('Aalen Gewichte'). → Diskussion.

Ausblick: Residuen

- ▶ Zur Erinnerung, die Martingal-Residuen waren definiert als

$$r_{M_i} = \delta_i - \hat{\Lambda}(t_i; \mathbf{x}_i),$$

wobei t_i der Ereignis- oder Zensierungszeitpunkt für Person i ist.

- ▶ Diese können für das additive Hazard Modell ganz genauso definiert werden; dabei entspricht (in der hier verwendeten Notation) δ_i dem $\Delta N(t_i)$ und $\hat{\Lambda}(t_i; \mathbf{x}_i)$ ist der gefittete Wert des Modells wie in (7.1) zu diesem Zeitpunkt.
- ▶ Etwas anders als im PH-Modell gilt hier: wenn das additive Modell korrekt spezifiziert ist, dann sind die Martingal-Residuen und die (korrekten Transformationen) der Variablen *unkorreliert*.

Inhalt von Abschnitt 8

Competing Risks

- Das Problem bei “Competing Risks”

- Cause-specific Hazards

- Bias bei naiver Schätzung

- Schätzung bei Competing Risks

- Modellierung mit Einflußgrößen

 - Proportional cause-specific Hazard Modell

 - Subdistribution Hazard und Fine & Gray Modell

- Ausblick: Multi-state Modelle

Competing Risks: Beispiele

Survival Daten — bisher: Dauer bis zu einer Art von Ereignis.

Competing Risks: es sind unterschiedliche Arten von Ereignissen möglich, von denen nur das zuerst auftretende beobachtet werden kann; z.B.

- ▶ Nach einer OP: Dauer bis entweder Rückfall oder Tod ohne vorherigen Rückfall;
- ▶ zur Beurteilung der Risiken eines Medikaments: verschiedene Todesursachen — Herzversagen, Blutgerinnsel, Nierenversagen etc.
- ▶ Dauer bis die Benutzung eines IUD eingestellt wird: verschiedene Gründe — Schwangerschaft, Ausstoss des IUD, andere Unverträglichkeit, persönliche Gründe.
- ▶ Überleben von Dialysepatienten: Nierentransplantation ist ein 'competing risk'.
- ▶ Krebspatienten: Tod durch die Krebserkrankung, oder aufgrund anderer altersbedingter Gründe.

Competing Risks: Mögliches Vorgehen

1. Manchmal werden die verschiedenen Versionen eines Ereignisses einfach zusammengefasst, z.B. 'relapse-free survival' heißt: Dauer bis entweder Rückfall oder Tod. \Rightarrow möglicherw. wichtige Unterschiede können dann nicht analysiert werden.
2. Gängiges aber falsches Vorgehen: betrachte eine Art von Ereignis, und definiere alle anderen als 'Zensierung'. \Rightarrow Oft läßt sich die Annahme der uninformativen Zensierung dann nicht aufrecht erhalten!! \longrightarrow Diskussion!

Beispiel 8.1 (Kardiovaskuläre versus andere Todesursachen)

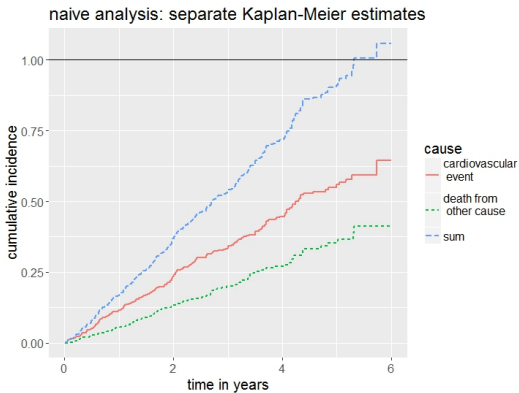
Um die Verzerrung des Kaplan-Meier Schätzers zu demonstrieren, betrachten wir als Beispiel folgende Studie ($n = 636$).

- ▶ Für die Placebo-Gruppe von Patienten mit Diabetes Typ II, die auf Dialyse angewiesen waren, ist die Dauer (gemessen in Jahren) bis zum Ereignis 'Tod durch kardiovaskuläre Gründe' von besonderem Interesse (genau gesagt handelt es sich um den kombinierten Endpunkt 'death from cardiac causes, stroke, and non-fatal myocardial infarction, whichever occurred first');
- ▶ da es sich meist um ältere Patienten handelt, können diese auch durchaus aufgrund von anderen Ursachen sterben — diese stellen konkurrierende Ereignisse dar (werden aber zu einem zusammengefaßt).

Der Graph zeigt aus zwei naiven Analysen: die Inzidenzkurven $\hat{F}_{kardio}(t)$ und $\hat{F}_{andere}(t)$.

→ Interpretation?

⇒ Das Risiko wird jeweils überschätzt — dies ist der typische Fehler bei Competing Risks.



Cause-specific Hazards – Definition

Sei $X_T \in \{1, \dots, K\}$ eine Indikatorvariable dafür, welches von K konkurrierenden Ereignissen (Ursache / Cause) zu der ersten Ereigniszeit T eintritt.

Man definiert als cause-specific Hazard:

$$\lambda_k(t) = \lim_{\Delta t \rightarrow 0} \frac{\Pr(t \leq T < t + \Delta t, X_T = k | T \geq t)}{\Delta t}.$$

Die kumulative Hazardfunktion für Ursache k ist:

$$\Lambda_k(t) = \int_0^t \lambda_k(u) du.$$

Cause-specific Hazards – Definition (Forts.)

- ▶ Es gilt: $\sum_k \lambda_k(t)$ ist die Gesamthazardfunktion (*all-cause hazard*) für die Dauer T bis zum Eintreten irgendeines Ereignisses.
- ▶ Damit ist die Survivalfunktion von T gegeben durch

$$S(t) = \exp\left\{-\sum_k \Lambda_k(t)\right\}. \quad (8.1)$$

Kumulative Inzidenzfunktion (CIF)

Die kumulative Inzidenzfunktion $F_k(t)$ ist die Wahrscheinlichkeit, dass das erste Ereignis vor t stattfindet und vom Typ k ist:

$$F_k(t) = \Pr(T \leq t, X_T = k) = \int_0^t \lambda_k(u) S(u) du.$$

—→ Warum?

—→ Was ist $1 - F_k(t)$?

Eigenschaften der CIF

- ▶ F_k hängt über $S(t)$ von allen cause-specific Hazardfunktionen ab, nicht nur von λ_k .
- ▶ F_k ist keine richtige Verteilungsfunktion:

$$\lim_{t \rightarrow \infty} F_k(t) < 1.$$

Bias bei naiver Schätzung

Wenn konkurrierende Ereignisse als Zensierungen angesehen werden, schätzt man

$$S_k(t) = \exp\{-\Lambda_k(t)\} = 1 - \int_0^t \lambda_k(u) S_k(u) du.$$

Es gilt aber:

$$\begin{aligned} S_k(t) &= 1 - \int_0^t \lambda_k(u) \exp\{-\Lambda_k(u)\} du \\ &\leq 1 - \int_0^t \lambda_k(u) \exp\left\{-\sum_{j=1}^K \Lambda_j(u)\right\} du, \end{aligned}$$

d.h. $S_k(t) \leq 1 - F_k(t)$ — die Survivalwahrscheinlichkeit wird unterschätzt, das Risiko wird überschätzt. Die Größe des Bias hängt von den konkurrierenden Ereignistypen ab.

Bias bei naiver Schätzung (Forts.)

Was ist $S_k(t) = \exp\{-\Lambda_k(t)\}$?

- ▶ Läßt sich interpretieren als die Wahrscheinlichkeit, ein Ereignis des Typs k nicht vor t zu erleben, **wenn** die λ_j 's für alle anderen Ereignistypen *auf Null gesetzt* würden. Dann: $S_k(t) = 1 - F_k(t)$.
- ▶ Wenn andere Ereignistypen tatsächlich nur 'Zensierung' sind, ist dies plausibel — Zensierung ließe sich im Prinzip verhindern.
- ▶ Ob eine hypothetische Situation, in der alle $\lambda_j = 0, j \neq k$, sind und λ_k den gleichen Wert beibehält, sinnvoll ist, hängt im Prinzip vom Kontext ab. Erfahrungsgemäß ist das selten der Fall bzw. wird nicht diskutiert.
- ▶ Das Problem wird in medizinischen Artikeln trotzdem (sehr) oft ignoriert.
- ▶ Man spricht oft vom Bias des Kaplan-Meier Schätzers bei competing risks; der Bias betrifft aber alle Methoden, die $S_k(t)$ statt korrekter cause-specific Größen wie $\lambda_k(t)$, $\Lambda_k(t)$, $F_k(t)$ (oder equiv.) schätzen.

Schätzung von Λ_k

Schätzer für die kum. cause-specific Hazardfunktionen sind durch das Nelson-Aalen Prinzip gegeben (vgl. Folie 2-11):

$$\hat{\Lambda}_k(t) = \sum_{j:t(j) \leq t} \frac{d_{kj}}{n_j}, \quad k = 1, \dots, K. \quad (8.2)$$

wobei

- d_{kj} = Anzahl Ereignisse vom Typ k zu $t(j)$;
- n_j = Anzahl Patienten 'at risk' zu $t(j)$.

Schätzung von F_k

- Grob, läßt sich damit λ_k schätzen als

$$\hat{\lambda}_k(t_{(j)}) = \hat{\Lambda}_k(t_{(j)}) - \hat{\Lambda}_k(t_{(j-1)}) = \frac{d_{kj}}{n_j}.$$

- Mit dem Kaplan-Meier Prinzip ergibt sich somit

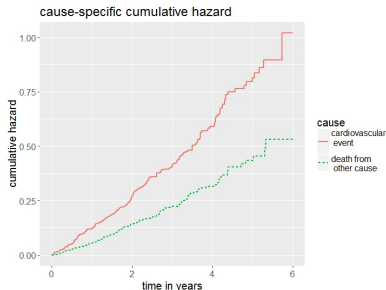
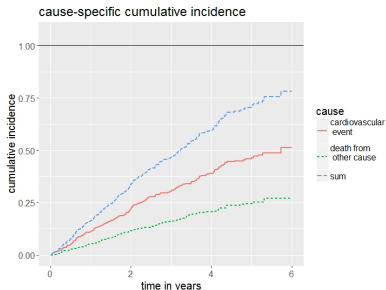
$$\hat{S}(t) = \prod_{j:t(j) \leq t} \left(1 - \frac{d_j}{n_j}\right) = \prod_{j:t(j) \leq t} \left(1 - \sum_{k=1}^K \hat{\lambda}_k(t_{(j)})\right).$$

- Und durch direktes Einsetzen ergibt sich

$$\hat{F}_k(t) = \sum_{j:t(j) \leq t} \hat{\lambda}_k(t_{(j)}) \hat{S}(t_{(j-1)}). \quad (8.3)$$

Beispiel 8.2 (Kardiovaskuläre versus andere Todesursachen, Fts.)

Inzidenzfunktionen ($\hat{F}_{kardio}(t)$ und $\hat{F}_{andere}(t)$) und kum. Hazardfunktionen ($\hat{\Lambda}_{kardio}(t)$ und $\hat{\Lambda}_{andere}(t)$) für dieses Beispiel sind nun plausibler.



→ Interpretation?

Proportional cause-specific Hazard Modell

- ▶ Modell für den Einfluss von (nicht zeitabhängigen) X_1, \dots, X_p auf $\lambda_k(t)$?
- ▶ PH-Modell für $\lambda_k(t)$ möglich:

$$\lambda_k(t; \mathbf{X} = \mathbf{x}) = \lambda_{0;k}(t) \exp(\boldsymbol{\beta}_k^T \mathbf{x}),$$

wobei $\lambda_{0;k}(t)$ cause-specific Baseline-Hazardfunktion und $\boldsymbol{\beta}_k^T = (\beta_{1k}, \dots, \beta_{pk})$ Koeffizienten speziell für Ursache k .

- ▶ Interpretation von $\boldsymbol{\beta}_k^T = (\beta_{1k}, \dots, \beta_{pk})$ wie im üblichen Cox PH Modell, aber bezogen auf cause-specific Hazard Ratios.

→ Diskussion!

Proportional cause-specific Hazard Modell (Forts.)

- ▶ Schätzung: wieder über partial likelihood, alle Ursachen $k = 1 \dots, K$ mit einbeziehend.
- ▶ Wenn separate Parameter β_k angenommen werden (üblich) können die Modelle auch für k separat angepasst werden. Dabei werden die anderen Ereignistypen wie Zensierungen behandelt.
Das darf aber wg. des Bias nicht auf die Schätzung der Survivalfunktion oder kum. Inzidenzfunktion übertragen werden!

Proportional cause-specific Hazard Modell (Forts.)

- ▶ Ähnlich wie im einfachen Fall läßt sich der Breslow-Schätzer (3.3) auf competing risks übertragen:

$$\hat{\Lambda}_{0;k}(t) = \sum_{t_{(j)} \leq t} \frac{d_{kj}}{\sum_{l \in R(t_{(j)})} \exp(\hat{\beta}_k^T \mathbf{x}_l)},$$

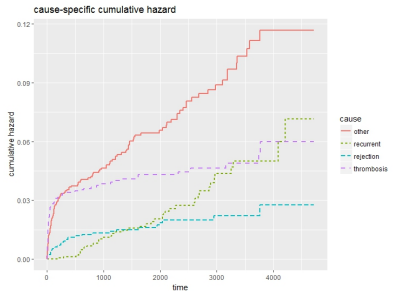
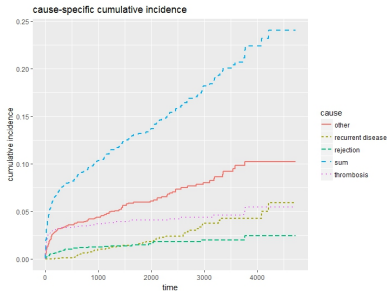
woraus sich die kum. cause-specific Hazardfunktion für $\mathbf{X} = \mathbf{x}$ ergibt:

$$\hat{\Lambda}_k(t; \mathbf{X} = \mathbf{x}) = \hat{\Lambda}_{0;k}(t) \exp(\hat{\beta}_k^T \mathbf{x}).$$

- ▶ Durch Einsetzen in (8.1) oder analog zu (8.3) lassen sich $S(t)$ und $F_k(t)$ für $\mathbf{X} = \mathbf{x}$ schätzen.
Beachte: Während der N-A-Schätzer (8.2) nicht-parametrisch ist, und z.B. bei wenigen Gruppen zum Vergleich berechnet werden kann, ist der Breslow-Schätzer an die PH Annahme mit linearem Prädiktor $\beta_k^T \mathbf{x}$ gebunden.

Beispiel 8.3 (Graft failure times in liver transplant recipients)

Von Interesse ist die Dauer (in Tagen) bis zu einem Versagen des Transplantats, und zwar kann dieses durch Abstoßen (rejection), durch Thrombose (thrombosis), Wiederkehr der Erkrankung (recurrent disease) oder einen anderen Grund (other) passieren. Dies wurde in einer Studie mit 1761 Patienten untersucht. → Handout.



Anmerkungen

- ▶ Im Falle $K = 1$ ist das proportional cause-specific Hazardmodell identisch mit dem üblichen Cox PH Modell.
- ▶ Im Allgemeinen ergibt sich für den all-cause Hazard kein PH Modell, wenn für jeden cause-specific Hazard ein PH Modell angenommen wird.
- ▶ Kein direkter Zusammenhang mit der kum. Inzidenzfunktion, d.h. Einfluss der Kovariablen \mathbf{X} auf kum. Inzidenzfunktion nicht unmittelbar feststellbar.
Insbesondere gibt es für $K > 1$ **keinen** vergleichbaren Zusammenhang wie

$$S(t|\mathbf{x}) = S_0(t)^{\exp(\beta^T \mathbf{x})}$$

im $K = 1$ Fall.

⇒ Subdistribution Hazard und Fine & Gray Modell!

Subdistribution Hazard

Die kumulative Inzidenzfunktion ist keine richtige Verteilungsfunktion (subdistribution function). Der zugehörige “Hazard” heißt deshalb *subdistribution hazard* und ist definiert durch die Gleichung:

$$F_k(t) = \Pr(T \leq t, X_T = k) = 1 - \exp \left(- \int_0^t \alpha_k(u) du \right).$$

Obige Gleichung beruht auf der Analogie zur Beziehung zwischen $\lambda(t)$ und $F(t)$ im einfachen Fall, siehe (1.4) und (1.5) mit $S = 1 - F$.

Daraus folgt unmittelbar

$$\alpha_k(t) = \frac{\lambda_k(t)S(t)}{1 - \Pr(T \leq t, X_T = k)} = \lambda_k(t) \frac{1 - \Pr(T \leq t)}{1 - \Pr(T \leq t, X_T = k)}.$$

—> Warum? Interpretation von $\alpha_k(t)$?

Fine & Gray Modell

Fine & Gray (1999) haben ein sog. *proportional subdistribution* Hazardmodell vorgeschlagen:

$$\alpha_k(t; \mathbf{X} = \mathbf{x}) = \alpha_{0;k}(t) \exp(\gamma_k^T \mathbf{x}),$$

- ▶ wobei γ_k Vektor der cause-specific Koeffizienten; $\alpha_{0;k}(t)$ baseline Subdistributionhazard.
- ▶ Schätzen: wie beim Cox Modell mittels partial likelihood; aber Bestimmung der at-risk-sets komplizierter, da Patienten mit Ereignistyp $\neq k$ unter Risiko bleiben. D.h. es muss “abgeschätzt” werden, wann diese zensiert würden.

→ Beispiel (Fts.): Handout.

Multistate Modelle

- ▶ Competing risks können als Spezialfall von multi-state Modellen angesehen werden: von Zustand “0” wird in einen der Zustände “ $k = 1, \dots, K$ ” gewechselt.
- ▶ Dies läßt sich verallgemeinern: ein stochastischer multi-state Prozeß $Y(t)$ kann zu jedem Zeitpunkt in einem von S Zuständen sein, und von Interesse sind die *Übergangsintensitäten* (vgl. cause-specific Hazards):

$$\lambda_{r,s}(t) = \lim_{\Delta t \rightarrow 0} \frac{\Pr(Y(t + \Delta t) = s | Y(t) = r)}{\Delta t},$$

wobei hier angenommen wird, dass nur der letzte Zustand $Y(t) = r$ relevant ist (Markov-Prozeß); in die Bedingung könnte aber (im Prinzip) zusätzliche Information über die Vergangenheit wie bei zeit-veränderlichen Kovariablen mit aufgenommen werden.

Multistate Modelle (Forts.)

- ▶ Beispiele: **Illness-Death Modelle** mit Zuständen 'gesund', 'krank', 'verstorben', wobei ein mehrfacher Wechsel zwischen 'gesund' und 'krank' möglich ist.
- ▶ Die kum. Übergangshazardfunktionen lassen sich unter der Markov-Annahme nicht-parametrisch nach dem Nelson-Aalen Prinzip schätzen.

Inhalt von Abschnitt 9

Bestimmung des Stichprobenumfangs

Vergleich von zwei Behandlungsgruppen

Anzahl erwartete Fälle (nach Schoenfeld, 1983)

Benötigte Anzahl Probanden

Vorüberlegungen

Wie bei jeder Studie:

- ▶ Einschluss-/Ausschluss-Kriterien
- ▶ Behandlungsdefinition
- ▶ Randomisierung
- ▶ Verblindung
- ▶ Anfangszeitpunkt und Endzeitpunkt der Studie
- ▶ sequentielles Design?
- ▶ Interim-Analyse?

Vergleich von zwei Behandlungsgruppen

- ▶ Voraussetzung: PH-Modell

$$\lambda_1(t) = \psi \lambda_0(t)$$

$\theta := \log \psi$ Log-Hazard Ratio;

Gruppe 0 (Standard / Kontrolle λ_0), Gruppe 1 ((neue) Behandlung λ_1).

- ▶ Test auf

$$H_0 : \theta = 0 \text{ vs. } H_1 : \theta \neq 0$$

basierend auf Log-Rank-Test;

- ▶ zur Bestimmung der Fallzahl muss geklärt werden:
 - ▶ Anzahl benötigter Ereignisse d ,
 - ▶ Wahrscheinlichkeit, dass Ereignis während der Studie eintritt \Rightarrow Anzahl benötigte Fallzahl n .

Anzahl benötigter Ereignisse

Man will einen Effekt von θ_R mit einer Power von $1 - \beta$ entdecken, wobei man den Test auf H_0 mit einem Signifikanzniveau von α durchführt.

- Anzahl benötigter Ereignisse bei gleich großen Gruppen

$$d = \frac{4(u_{1-\frac{\alpha}{2}} + u_{1-\beta})^2}{\theta_R^2}. \quad (9.1)$$

- Falls man ungleiche Anzahlen hat, bezeichne $\pi =$ Anteil Gruppe 0, $1 - \pi =$ Anteil Gruppe 1

$$\Rightarrow d = \frac{(u_{1-\frac{\alpha}{2}} + u_{1-\beta})^2}{\pi(1 - \pi)\theta_R^2}$$

Bemerke, dass ungleiche Verteilung zu einer größeren benötigten Anzahl führt. → Warum?

Herleitung von (9.1) (auf der Basis des Log-Rank-Tests).

Zur Erinnerung:

$$\begin{aligned}
 t_{(1)} &< \cdots < t_{(r)} && - r \text{ Ereigniszeiten} \\
 n_{ij} &&& - \text{Anzahlen unter Risiko} \\
 d_j &= d_{1j} + d_{2j} && - \text{Anzahlen Fälle}
 \end{aligned}$$

Die Log-Rank-Statistik ist

$$U_L = \sum_{j=1}^r (d_{1j} - e_{1j}) \quad \text{mit } e_{1j} = n_{1j} \frac{d_j}{n_j},$$

Die Varianz von U_L ist

$$V_L = \sum_{j=1}^r \frac{n_{1j} n_{2j} d_j (n_j - d_j)}{n_j^2 (n_j - 1)}$$

- ▶ Zur Fallzahlbestimmung ist es geeigneter, das Testproblem über U_L selbst anstatt über $T_L = U_L^2/V_L$ zu formulieren.
- ▶ H_0 wird abgelehnt, falls $|U_L|$ hinreichend groß, d.h. $|U_L| \geq k$ für eine Konstante k (der übliche kritische Wert für eine Teststatistik).
- ▶ Die Bedingungen an k , sowie die zweite Bedingung zur Bestimmung von d , sind dabei

$$\Pr(|U_L| > k; \theta = 0) = \alpha$$

$$\Pr(|U_L| > k; \theta = \theta_R) = 1 - \beta$$

(2-seitiger Test zum Niveau α mit Power $(1 - \beta)$ für $\theta = \theta_R$).

► Speziell wird benutzt:

Unter $\theta = 0$:

$$\Pr(U_L > k; \theta = 0) = \frac{\alpha}{2} \quad (9.2)$$

Unter $\theta = \theta_R < 0$ (oBdA):

$$\begin{aligned} \Pr(|U_L| > k; \theta_R) &= \Pr(U_L > k; \theta_R) + \Pr(U_L < -k; \theta_R) \\ &\approx \Pr(U_L < -k; \theta_R). \end{aligned} \quad (9.3)$$

- Bestimmen von k aufgrund von (9.2), wobei $U_L \stackrel{H_0}{\sim} N(0, V_L)$:

$$\begin{aligned}\Pr(U_L > k; \theta = 0) &= 1 - \Pr(U_L \leq k; \theta = 0) \\ &= 1 - \Phi\left(\frac{k}{\sqrt{V_L}}\right) \stackrel{!}{=} \stackrel{(9.2)}{=} \frac{\alpha}{2} \\ \Rightarrow \quad \Phi\left(\frac{k}{\sqrt{V_L}}\right) &= 1 - \frac{\alpha}{2} \\ \Rightarrow \quad \frac{k}{\sqrt{V_L}} &= u_{1-\frac{\alpha}{2}} \\ \Rightarrow \quad k &= u_{1-\frac{\alpha}{2}} \cdot \sqrt{V_L}\end{aligned}\tag{9.4}$$

- Dabei ist die Varianz V_L eine Funktion von d und n , die noch zu bestimmen sind.

- ▶ **Ohne Beweis:** für kleine θ gilt $U_L \sim \mathcal{N}(\theta V_L, V_L)$
- ▶ Sei nun $\theta = \theta_R$; man benutze (9.3)

$$\begin{aligned}\Pr(U_L < -k; \theta_R) &= \Pr\left(\frac{U_L - \theta_R V_L}{\sqrt{V_L}} < \frac{-k - \theta_R V_L}{\sqrt{V_L}}; \theta_R\right) \\ &= \Phi\left(\frac{-k - \theta_R V_L}{\sqrt{V_L}}\right) \stackrel{!}{=} (1 - \beta) \\ \Rightarrow \quad \frac{-k - \theta_R V_L}{\sqrt{V_L}} &= u_{1-\beta}\end{aligned}$$

k aus (9.4) eingesetzt: $-u_{1-\frac{\alpha}{2}}\sqrt{V_L} - \theta_R V_L = u_{1-\beta}\sqrt{V_L}$

$$\Rightarrow \quad V_L = \frac{(u_{1-\frac{\alpha}{2}} + u_{1-\beta})^2}{\theta_R^2} \quad (9.5)$$

- ▶ Jetzt: nach d auflösen...

- Ann.: Anzahl Fälle klein im Vergleich Anzahl unter Risiko ($d_j \ll n_j$)

$$\Rightarrow V_L \approx \sum_j \frac{n_{1j}n_{2j}d_j}{n_j^2}$$

- Ann.: θ klein und Rekrutierung in beide Gruppen geschieht mit gleicher Rate $\Rightarrow n_{1j} \approx n_{2j}$ für alle j .

$$\Rightarrow \frac{n_{1j}n_{2j}}{n_j^2} = \frac{n_{1j}n_{2j}}{(n_{1j} + n_{2j})^2} \approx \frac{n_{1j}^2}{4n_{1j}^2} = \frac{1}{4}$$

$$\Rightarrow V_L \approx \sum_j \frac{d_j}{4} = \frac{d}{4}$$

- Mit Gleichung (9.5) erhält man als Forderung an die Gesamtzahl der Ereignisse

$$\frac{d}{4} = \frac{(u_{1-\frac{\alpha}{2}} + u_{1-\beta})^2}{\theta_R^2}$$

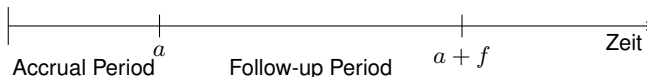
Bemerkung: für „späte“ Ereignisse (j in der Nähe von r) werden n_{1j} und n_{2j} idR verschieden sein und $\frac{n_{1j}n_{2j}}{n_j^2} < \frac{1}{4}$ und die Anzahl der benötigten Ereignisse wird daher tendenziell unterschätzt.

Zusammengefaßt: verwendete Annahmen

- ▶ Cox Modell;
- ▶ θ klein so, dass $U_L \sim \mathcal{N}(\theta V_L, V_L)$;
- ▶ Anzahl Fälle klein im Vgl. zu Anzahl unter Risiko ($d_j \ll n_j$);
- ▶ Rekrutierung in beide Gruppen geschieht mit gleicher Rate
 $\Rightarrow n_{1j} \approx n_{2j}$ für alle j .

Benötigte Anzahl Probanden

- ▶ Mit d gegeben, wieviele Probanden?
- ▶ Man hat auszurechnen die Wahrscheinlichkeit, während der Studie das Ereignis E zu erleben, um die benötigte Anzahl Probanden zu ermitteln — hier kurz mit $\Pr(E)$ bezeichnet.



- ▶ Hat man $\Pr(E)$ ermittelt, so ergibt sich die benötigte Anzahl n von Probanden aus

$$n = \frac{d}{\Pr(E)} \quad (9.6)$$

mit d aus (9.1) (oder der entspr. Formel für ungleiche Gruppengrößen).

- Man berechnet für die Wahrscheinlichkeit eines Ereignisses:

$$\Pr(E) = 1 - \frac{1}{6} [\bar{S}(f) + 4\bar{S}(0,5a + f) + \bar{S}(a + f)] \quad (9.7)$$

mit $\bar{S}(t) = \frac{S_1(t) + S_2(t)}{2}$.

- Dabei müssen die Survivalkurven $S_1(t)$, $S_2(t)$ entweder durch eine parametrische Annahme, oder durch Daten aus einer Pilotstudie approximiert werden.

- Begründung von (9.7), wobei untersch. Anfangszeitpunkte (“Eintritt”) zu berücksichtigen ist:

$$\Pr(E) = \int_0^a \Pr(E, \text{Eintritt in } t) dt$$

mit

$$\Pr(E, \text{Eintritt in } t) = \Pr(E \mid \text{Eintritt in } t) \cdot \Pr(\text{Eintritt in } t)$$

Ann., dass Rekrutierung gleichmäßig in $[0, a]$

Also, $\Pr(\text{Eintritt zu } t) \equiv \frac{1}{a} \quad \forall t \in [0, a]$.

$$\Rightarrow \Pr(E) = 1 - \frac{1}{a} \int_f^{a+f} S(u) du \quad (9.8)$$

► Approximation des Integrals (9.8) mit Simpsons Formel

$$\int_u^v f(x)dx \cong \frac{v-u}{6} \left[f(u) + 4f\left(\frac{u+v}{2}\right) + f(v) \right]$$

$$\Rightarrow \int_f^{a+f} S(u)du \approx \frac{a}{6} [S(f) + 4S(0,5a + f) + S(a + f)]$$

$$\Rightarrow \Pr(E; \text{Gruppe 0}) = 1 - \frac{1}{6} [S_1(f) + 4S_1(0,5a + f) + S_1(a + f)]$$

$$\Pr(E; \text{Gruppe 1}) = 1 - \frac{1}{6} [S_2(f) + 4S_2(0,5a + f) + S_2(a + f)]$$

► Unter der Annahme der zufälligen Zuordnung zur Behandlung hat man

$$\Pr(E) = \frac{\Pr(E; \text{Gruppe 0}) + \Pr(E; \text{Gruppe 1})}{2}$$

$$\Rightarrow \Pr(E) = 1 - \frac{1}{6} [\bar{S}(f) + 4\bar{S}(0,5a + f) + \bar{S}(a + f)]$$

$$\text{mit } \bar{S}(t) = \frac{1}{2}(S_1(t) + S_2(t)).$$

- ▶ Alternative: Berechnung des Integrals (9.8) unter der Annahme einer Exponentialverteilung für die Ereigniszeiten:
Mit Ratenparameter λ ergibt sich, unter obiger Rekrutierungs- und Follow-up Phase, die Wahrsch. für ein Event als

$$\Pr(E; \lambda) = 1 - \frac{e^{-\lambda f} - e^{-\lambda(a+f)}}{\lambda a} \quad (9.9)$$
$$\Rightarrow \Pr(E) = \frac{1}{2}(\Pr(E; \lambda) + \Pr(E; \psi_R \lambda)).$$

- ▶ Hier muss ein plausibler Wert für λ angenommen werden.
- ▶ Eine andere Alternative: es läßt sich auch überlegen, wie lange f eine Studie laufen muss, um die nötige Anzahl an Ereignissen d bei gegebenem n zu erreichen.

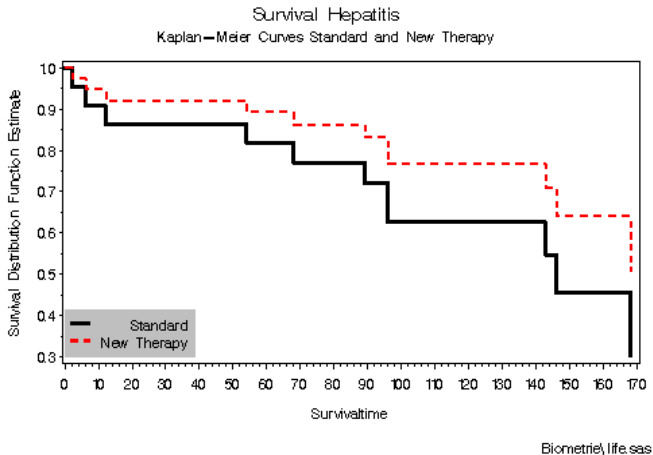
Beispiel 9.1 (Hepatitis-Studie)

- Neue Behandlung (Gruppe 1) vs. Standard (Gruppe 0); abzusichernder Hazardratio sei $\psi_R = 0,57$, $\alpha = 0,05$, $\beta = 0,10$:

$$d = \frac{4 \cdot 10,51}{0,562^2} = 133;$$

Fazit: 140 Ereignisse (aufgerundet) werden benötigt.

Mit Daten aus Pilotstudie berechnet man die Survivalfunktionen.



► Anzahl Probanden:

Rekrutierungsphase: $a = 18$ Monate

Dauer des Follow-up: $f = 24$ Monate

$$S_2(t) = S_1(t)^{\psi_R} \text{ mit } \psi_R = 0,57.$$

Fazit: hier ergäbe sich $\Pr(E) \approx 1 - 0,9$ (siehe vorangehende Graphik), sodass 1400 Patienten zu rekrutieren wären.

- Um Formel (9.9) anzuwenden, würde man zunächst, unter denselben Bedingungen, eine Ereignisrate unter Standardtherapie schätzen: Diese ergibt sich hier zu $\hat{\lambda} = 4,564/1000$. Mit dem abzusichernden Hazard Ratio von 0,57 errechnet sich dann eine Ereigniswahrsch. von ca. 11%. Damit wären etwa 1280 Probanden zu rekrutieren.
- In komplizierteren Situationen: Monte-Carlo Simulationen!

Inhalt von Abschnitt 10

Zensierungsmechanismen

- Zensierungstypen (Wdhlg.)

- Likelihoodkonstruktion

- Nichtparametrische Schätzung bei Linkstrunkierung

- Ausblick

Rechtszensierung

- ▶ **Definition:** es wird $T_i^* = \min(T_i, C_i)$ beobachtet, wobei C_i der Zensierungszeitpunkt ist; C_i kann für verschiedene Individuen variieren.
- ▶ **Typ I** Rechtszensierung: C_i repräsentiert ein festgelegtes zensierendes Ereignis, oft das Studienende. Da Eintritt in die Studie zu unterschiedlichen Zeiten sein kann, können C_i zwischen i variieren.
- ▶ **Typ II** Rechtszensierung: die Studie wird durchgeführt, bis r Individuen das Ereignis gehabt haben. Die Daten stellen dann die r kleinsten Dauern dar, was die Likelihoodkonstruktion vereinfacht.
- ▶ **Zufällige** Rechtszensierung: C_i ist eine durch ein anderes Ereignis entstehende Zensierungszeit, z.B. Versagen eines Messgerätes, oder Tod durch Unfall.
- ▶ Kombinationen und Verallgemeinerungen sind möglich und üblich.

Links- und Intervallzensierung

- ▶ **Linkszensierung Definition:** es wird $T_i^* = \max(T_i, C_i)$ beobachtet, wobei C_i der Zensierungszeitpunkt ist; C_i kann für verschiedene Individuen variieren.
- ▶ **Beispiel:** Studie mit Kleinkindern — wann lernen sie lesen? Manche Kinder können vielleicht bei Studieneintritt schon lesen, diese sind dann linkszensiert.
- ▶ Kombination von Links- und Rechtszensierung:
 $T^* = \max\{\min(T, C^r), C^l\}$, wobei C^l und C^r die (vielleicht zufälligen und individuumspezifischen) linken und rechten Zensierungszeiten sind.
- ▶ **Intervallzensierung:** die genaue Dauer T_i ist nicht bekannt, nur dass $T_i \in (L_i, R_i]$ liegt. Dies kommt oft vor, wenn zu diskreten Zeitpunkten beobachtet / gemessen wird.

Trunkierung (Truncation)

- ▶ **Linkstrunkierung:** T wird überhaupt nur beobachtet, wenn $T \geq U$, wobei U einen Zeitpunkt definiert, der erreicht worden sein muss, um in die Studie zu gelangen.
- ▶ Im Gegensatz zu Zensierung, sind bei Trunkierung Individuen mit $T < U$ gar nicht in der Studie (oft per Design).
- ▶ **Beispiel:** Studie mit Personen, die in einem Altersheim sind — Personen, die vor dem Eintritt ins Altersheim gestorben sind, kommen hier per Design gar nicht vor. U = Eintritt ins Altersheim.
- ▶ **Rechtstrunkierung**, z.B.: Studie mit Personen, die mit einer Krankheit infiziert sind; solche, die noch nicht infiziert sind, kommen dabei nicht vor. Oder: Studien, die auf Mortalitätsregister basieren.

Uninformative Zensierung!

Die verschiedenen Zensierungstypen wirken sich natürlich darauf aus, wie die Likelihood zu konstruieren ist, da die Daten unterschiedliche Informationen enthalten.

Zunächst ist es aber am wichtigsten, zu klären, *ob der Zensierungsmechanismus überhaupt ignoriert* werden kann, oder in irgendeiner Weise informativ ist!

- ▶ **Informativ:** wenn das interessierende Ereignis für eine zensierte Person systematisch früher (oder später) eintritt, z.B. eine Person verläßt die Studie, weil sie sich nicht gesund genug fühlt, um weiter zu machen.
- ▶ **Uninformativ:** Zensierung darf zwar von dem beobachteten Zeitraum abhängen, aber nicht von dem unbeobachteten; z.B. werden lange Dauern eher zensiert als kurze, das ist aber an sich keine informative Zensierung.

Komponenten der Likelihood

Für die verschiedenen Arten von (un)zensierten Beobachtungen, und bei uninformativer Zensierung, setzt sich die Likelihood aus den folgenden Komponenten zusammen (vgl. Folie 1-17):

- ▶ exakt beobachtete Dauer: $f(T_i)$;
- ▶ rechtszensierte Dauer: $S(C_i^r)$;
- ▶ linkszensierte Dauer: $1 - S(C_i^l)$;
- ▶ intervallzensierte Dauer: $S(L_i) - S(R_i)$;
- ▶ linkstrunkierte Dauer: $f(T_i)/S(U_i)$;
- ▶ rechtstrunkierte Dauer: $f(O_i)/(1 - S(O_i))$.

Beachte: die Likelihood ist vor allem für parametrische Modelle (Kapitel 6) von direktem Interesse; gerade bei komplizierten Zensierungsproblemen sind parametrische Ansätze zunächst 'einfacher', da sie alles modellieren.

Uninformative Linkstrunkierung

Sei U_i die Zufallsvariable, die den Trunkierungszeitpunkt für Individuum i beschreibt mit Realisation u_i .

Formal bedeutet uninformative Trunkierung, dass für alle i :

$$\Pr(t \leq T_i < t + \Delta t | T_i \geq t, T_i \geq u_i, U_i = u_i) = \Pr(t \leq T_i < t + \Delta t | T_i \geq t).$$

D.h. der Hazard (oder das Risiko) für ein Ereignis im nächsten Moment hängt nicht davon ab, dass die Person den Zeitpunkt U_i überlebt hat (wenn T_i eh schon größer als U_i ist).

Linkstrunkierung: 'At-Risk' Menge

Unter uninformativer Linkstrunkierung (und Rechtszensierung) lässt sich nicht-parametrisch schätzen indem:

- die Bestimmung der 'at-risk' Mengen und
- die Interpretation der geschätzten Funktion angepasst wird.

Linkstrunkierung: 'At-Risk' Menge (Forts.)

Betrachte eine Gruppe von n Personen mit t_1, \dots, t_n beobachteten Survival- oder Zensierungszeiten; für jede Person sei der Zeitpunkt u_1, \dots, u_n des Eintritts in die Studie bekannt (z.B. Alter bei Eintritt ins Altersheim).

$t_{(1)} < t_{(2)} < \dots < t_{(r)}$ seien die geordneten *Ereigniszeiten*.

Für die j -te Ereigniszeit, $1 \leq j \leq r$, sei nun die Anzahl unter Risiko zum Zeitpunkt $t_{(j)}^-$

$$n_j = \sum_{i=1}^n I\{u_i \leq t_{(j)} \leq t_i\}.$$

Dies ist etwas anders als vorher, siehe Kapitel 2 (Folie 2-7).

→ Beispiel!

Linkstrunkierung: 'At-Risk' Menge (Forts.)

Der Kaplan-Meier-Schätzer schätzt nun $\Pr(T > t | T > \min\{u_i\} = u)$ und ist gegeben, für $t_{(k)} \leq t < t_{(k+1)}$, wie vorher:

$$\hat{S}_u(t) = \prod_{j=1}^k \frac{(n_j - d_j)}{n_j} = \prod_{j:t(j) \leq t} \left[1 - \frac{d_j}{n_j} \right].$$

Da die Schätzung für die frühen $t_{(j)}$ i.d.R. instabil ist wegen kleiner at-risk Mengen, wird der KM-Schätzer oft so modifiziert, dass man erst bei einem hinreichend großen Zeitpunkt u^* anfängt:

$$\hat{S}_{u^*}(t) = \prod_{j:u^* \leq t(j) \leq t} \left[1 - \frac{d_j}{n_j} \right].$$

$\hat{S}_{u^*}(t)$ schätzt $\Pr(T > t | T > u^*), t \geq u^*$. → Beispiel (Handout)!

Linkstrunkierung — Bemerkungen

- ▶ Mit der neuen Definition der n'_j s lassen sich auch alle anderen Größen, die auf d_j/n_j basieren, wie zuvor berechnen (inkl. KIs oder Konfidenzbänder), z.B. der NA-Schätzer $\tilde{\Lambda}(t)$.
- ▶ Der NA-Schätzer $\tilde{\Lambda}(t)$ schätzt nun allerdings den kum. Hazard über das Intervall (u, t) .
- ▶ Die Steigung des NA-Schätzers $\tilde{\Lambda}(t)$ lässt sich wie zuvor als *unbedingte* Hazardfunktion interpretieren (und z.B. durch Glättung schätzen);
→ warum?

Schätzen bei anderen Zensierungstypen

- ▶ Bei 'purer' Linkszensierung (uninf.), lässt sich die Zeitskala '*umdrehen*', so dass alle bisherigen Verfahren 'rückwärts' angewendet werden können. Der KM-Schätzer schätzt dann $\Pr(\tau - T > t) = \Pr(T < \tau - t)$ für ein großes τ . Linkszensierung ohne Rechtszensierung ist allerdings selten.
- ▶ Bei allen anderen Zensierungstypen müssen zur nicht-parametrischen Schätzung *iterative* Verfahren verwendet werden, da geschlossene Formeln nicht möglich sind.