4. Model Coxa (cd.)

Zmienne objaśniające zależne od czasu

Ocena dobroci dopasowania

Model regresji Poissona

Zmienne objaśniające zależne od czasu w modelu Coxa (1)

$$\lambda(t) = \lambda_0(t) \cdot e^{\mathbf{X'}\boldsymbol{\alpha} + Z(t)\beta}$$

- Zmienne objaśniające, których wartości zmieniają się w czasie
  - Zmienne wewnętrzne: mierzalne tylko gdy jednostka pozostaje pod obserwacją; np. wielokrotne pomiary parametru krwi chorego
  - Zmienne zewnętrzne: nie związane z obserwowaną jednostką, o wartościach znanych z góry (np. planowane dawki leku, pora roku, itd.)
- e^{β} jest ilorazem hazardu związanym ze wzrostem $Z(t)$ o 1 *dla dowolnego* t (nie tylko dla t=0), przy ustalonej wartości $\mathbf{X}$

Zmienne objaśniające zależne od czasu w programach statystycznych

- Tzw. „counting process input”:

PatID	Start	Stop	Event	Z_1	Z_2
1	0	3	0	1	4.5
1	3	5	0	1	6
1	5	13	0	1	3
1	13	22	1	1	6

- *Start* i *stop* mogą być datą lub wartością czasu
  - Metoda pozwala na uwzględnienie „opóźnionego rozpoczęcia obserwacji” lub wyłączeń ze zbioru ryzyka

PatID	Start	Stop	Event	Z_1	Z_2
1	0	3	0	1	4.5
1	4	5	0	1	6
2	5	8	0	1	3
2	8	11	1	1	7

„Wewnętrzne” zmienne objaśniające zależne od czasu

- Ich użycie wymaga ostrożności

- Przykład (hipotetyczny): leczenie białaczki
  - Przeżycie chorego może zależeć od poziomu płytek krwi
  - Załóżmy, że lek podwyższa poziom płytek
  - Po uwzględnieniu poziomu płytek jako zmiennej objaśniającej zależnej od czasu w modelu, efekt leczenia może nie być istotny

Szacowanie modelu Coxa ze zmiennymi objaśniającymi zależnymi od czasu

$$L(\beta) = \prod_{k=1}^{K} \frac{e^{\beta' Z_{(k)}(t_{(k)})}}{\sum_{l=1}^{n} Y_l(t_{(k)}) e^{\beta' Z_l(t_{(k)})}}$$

- Konieczna znajomość wartości zmiennych dla <u>wszystkich</u> zaobserwowanych czasów zdarzeń
  - potencjalna trudność
    - interpolacja dla zmiennych ciągłych
    - problem dla zmiennych dyskretnych

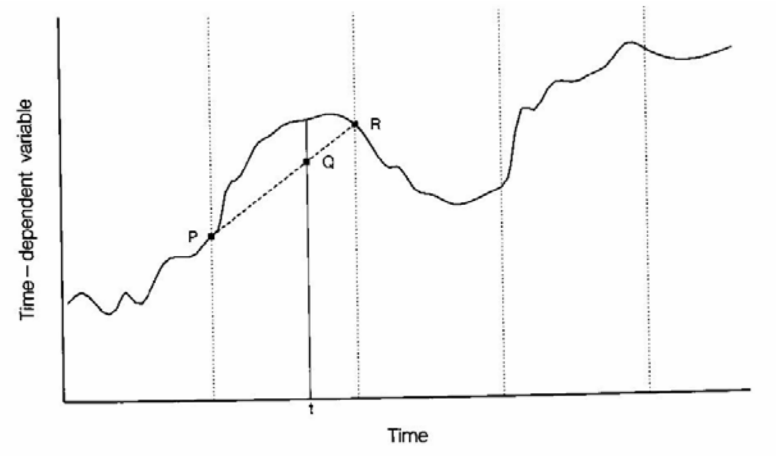

Figure 8.2 *Computation of the value of a time-dependent variable at intermediate times.*

Ocena dobroci dopasowania modelu (1)

- Reszty martyngałowe: $$M(t) = N(t) - \int_0^t Y(u)\lambda_0(u) e^{\beta^T Z(u)} du$$

- Przy „counting style input”, kilka reszt dla obserwowanej jednostki: *M*(*stop*)-*M*(*start*)
  - Suma daje resztę całkowitą

- Wykres reszt martyngałowych (lub opartych na dewiancji) vs. ***Z’(t)β*** daje możliwość ogólnej oceny dopasowania modelu
  - np. obserwacji odstających

Ocena dobroci dopasowania modelu (2)

- Funkcja przeżycia dla jednostki *i* na podstawie modelu:

$$S_i(t) = \exp\left\{-\int_0^t \exp\left(\beta^T Z(u)\right)\lambda_0(u)du\right\}$$

  - Może zależeć od przyszłych (nieznanych) wartości $Z(t)$

- Rozważmy p-stwo przeżycia „krótkiego” przedziału:

$$\mathrm{P}(T_i^* \geq t+h \mid T_i^* \geq t) = S_i(t+h)/S_i(t) =$$

$$= \frac{\exp\left\{-\int_0^{t+h} \exp\left(\beta^T Z(u)\right)\lambda_0(u)du\right\}}{\exp\left\{-\int_0^{t} \exp\left(\beta^T Z(u)\right)\lambda_0(u)du\right\}} \approx \frac{\exp\left\{-\exp\left(\beta^T Z(t)\right)\int_0^{t+h} \lambda_0(u)du\right\}}{\exp\left\{-\exp\left(\beta^T Z(t)\right)\int_0^{t} \lambda_0(u)du\right\}} =$$

$$= \exp\left\{\left(-\Lambda_0(t+h)+\Lambda_0(t)\right)\exp\left(\beta^T Z(t)\right)\right\}$$

Ocena dobroci dopasowania modelu (3)

- 1- P($T^*\geq t+h|T^*\geq t$) to prawdopodobieństwo zdarzenia w ($t,t+h$)
- Możemy oszacować wartości P($T^*\geq t+h|T^*\geq t$) na podstawie oszacowania β i Λ_0 ...
- ... i użyć ich do obliczenia oczekiwanej liczby zdarzeń.
- Porównanie z obserwowaną liczbą zdarzeń w kolejnych przedziałach daje możliowść oceny dopasowania modelu do danych.

Ocena formy funkcjonalnej zmiennej objaśniającej zależnej od czasu (1)

- „Wygładzony" wykres reszt martyngałowych dla „pustego" modelu nie działa
  - kilka reszt dla jednostki – wiele reszt w okolicach 0
  - użycie reszty całkowitej – ale dla jakiej wartości zmiennej?
  - możliwość obciążenia estymacji na podstawie wykresu

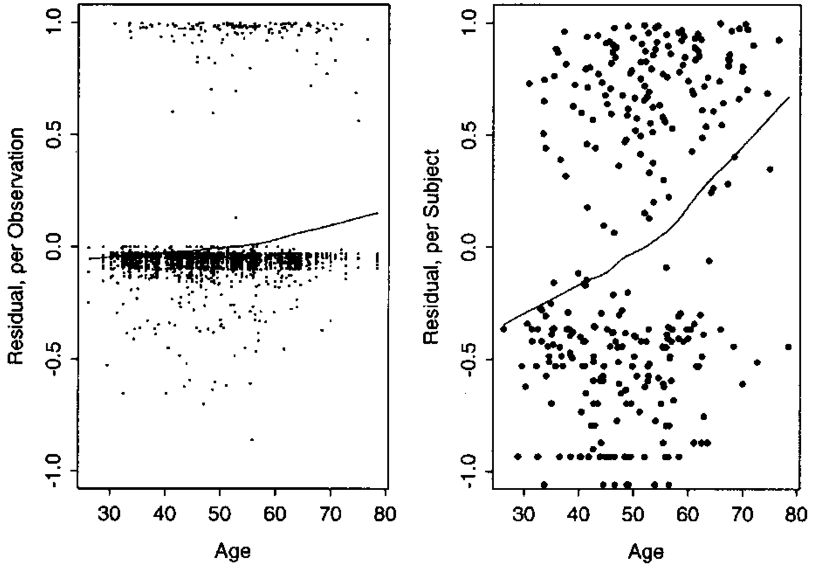

FIGURE 5.17: *Sequential PBC data, null model martingale residuals*

Ocena formy funkcjonalnej zmiennej objaśniającej zależnej od czasu (2)

- Możliwe użycie modelu Poissona z gładką funkcją zmiennej lub zastosowanie spline'ów

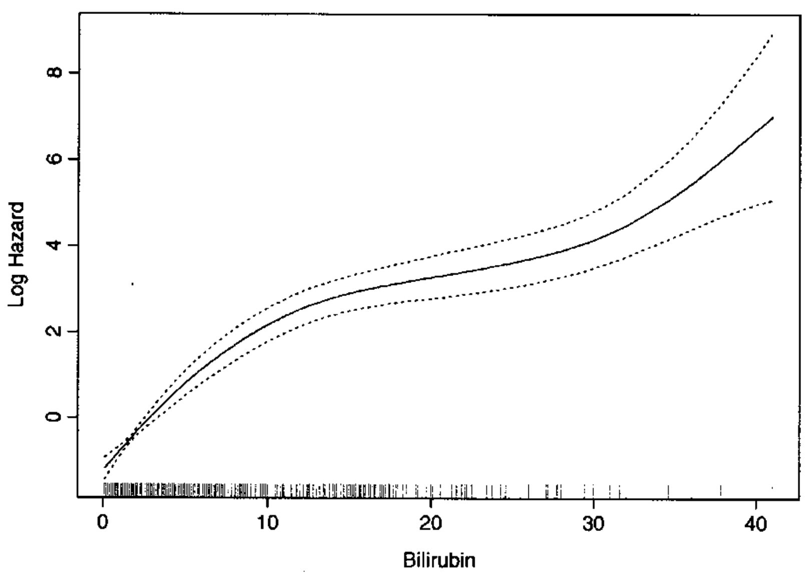

FIGURE 5.18: *Sequential PBC data, smoothing spline fit*

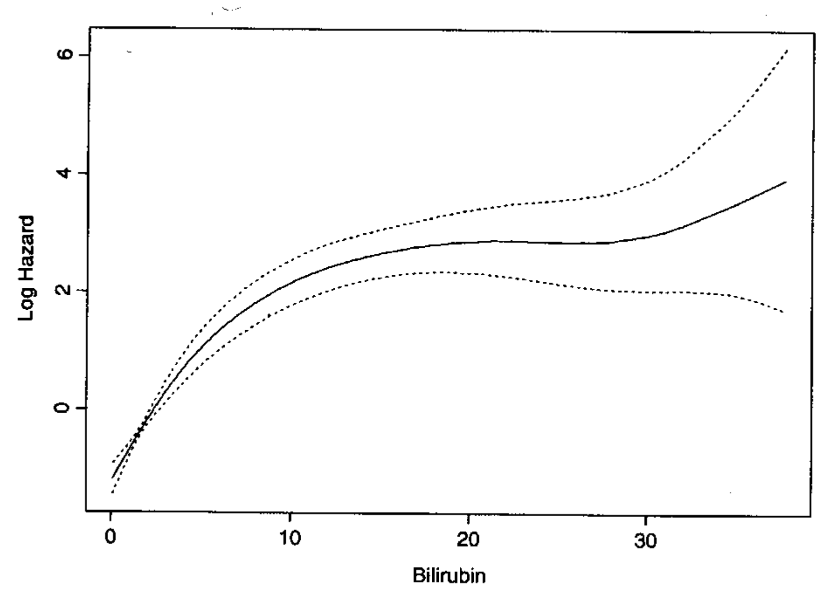

FIGURE 5.19: *Sequential PBC data, protocol visits only*

„Wewnętrzne” zmienne objaśniające zależne od czasu: przykład (1)

- Przykład (hipotetyczny): leczenie marskości wątroby
  - logarytm bilirubiny na początku próby klinicznej

Table 8.3 *Survival times of 12 patients in a study on cirrhosis of the liver.*

Patient	*Time*	*Status*	*Treat*	*Age*	*Lbr*
1	281	1	0	46	3.2
2	604	0	0	57	3.1
3	457	1	0	56	2.2
4	384	1	0	65	3.9
5	341	0	0	73	2.8
6	842	1	0	64	2.4
7	1514	1	1	69	2.4
8	182	0	1	62	2.4
9	1121	1	1	71	2.5
10	1411	0	1	69	2.3
11	814	1	1	77	3.8
12	1071	1	1	58	3.1

Table 8.5 *Values of* $-2\log\hat{L}$ *for models without a time-dependent variable.*

Terms in model	$-2\log\hat{L}$
null model	25.121
Age	22.135
Lbr	21.662
Age, Lbr	18.475

- *Age* + *Lbr* + *Treat*:

LRT = 5.182 (p=0.023)

β_{Treat} =-3.052

„Wewnętrzne” zmienne objaśniające zależne od czasu: przykład (2)

- Logarytm bilirubiny zależny od czasu

Patient	*Time*	*Status*	*Treat*	*Age*	*Lbr*
1	281	1	0	46	3.2
2	604	0	0	57	3.1
3	457	1	0	56	2.2

Patient	Follow-up time	Log(bilirubin)
1	47	3.8
	184	4.9
	251	5.0
2	94	2.9
	187	3.1
	321	3.2
3	61	2.8
	97	2.9
	142	3.2
	359	3.4
	440	3.8

Table 8.6 *Values of* $-2 \log \hat{L}$ *for models with a time-dependent variable.*

Terms in model	$-2 \log \hat{L}$
null model	25.121
Age	22.135
Lbrt	12.053
Age, Lbrt	11.145

- *Lbr* + *Treat*:

LRT = 12.053-10.676=1.377 (p=0.241)

β_{Lbr} =3.605, β_{Treat} =-1.479

- Jeśi leczenie zmienia poziom bilirubiny, włączenie *Lbr* do modelu uwzględnia efekt leczenia

„Wewnętrzne” zmienne objaśniające zależne od czasu: przykład (3)

- Logarytm bilirubiny zależny od czasu

Table 8.7 *Estimated baseline cumulative hazard function, $\tilde{H}_0(t)$, for the cirrhosis study.*

Follow-up time (t)	$\tilde{H}_0(t)$
0	0.000
281	0.009×10^{-6}
384	0.012×10^{-6}
457	0.541×10^{-6}
814	0.908×10^{-6}
842	1.577×10^{-6}
1071	3.318×10^{-6}
1121	6.007×10^{-6}
1514	6.053×10^{-6}

Table 8.8 *Approximate conditional survival probabilities for patients 1 and 7.*

Time interval	$\tilde{P}_1(t, t+h)$	$\tilde{P}_7(t, t+h)$
0–	0.999	1.000
360–	0.000	0.994
720–	0.000	0.969
1080–	0.000	0.457
1440–	0.045	0.364

Interval	Expected	Observed
0-359	0.02	1
360-719	2.46	2
720-1079	5.64	3
1080-1439	6.53	1
1440-1514	3.16	1

- Expected>observed
- Mała liczebność próbki

Model regresji Poissona (1)

- Dla *współczynników intensywności* zdarzeń
- Szacowanie wymaga obserwacji występowania zdarzeń w czasie
- Najbardziej podstawowy estymator (*incidence density, ID*):

$$ID = \frac{I}{PT} = \frac{\text{no. of new cases in the calendar period } (t_0, t_1)}{\text{accrued population time}}$$

- *ID* jest wyznaczany z danych *zgrupowanych*
  - zliczenia zdarzeń

Model regresji Poissona (2)

- Współczynnik intensywności λ dla grupy 1 i λ_0 dla grupy 0
- Rozważmy *iloraz współczynników (incidence rate ratio)*

$$\mathrm{IRR} = \lambda / \lambda_0$$

który można wyrazić jako

$$\lambda = \lambda_0 \cdot \theta$$

gdzie $\theta = \mathrm{IRR}$

- Równoważnie:

$$\ln \lambda = \ln \lambda_0 + \ln \theta$$

lub, kładąc $\beta_0 = \ln \lambda_0$, $\beta = \ln \theta$, i Z=1 (0) dla grupy 1 (0)

$$\ln \lambda = \beta_0 + \beta \cdot Z$$

Model regresji Poissona (2)

- Model

$$\ln \lambda = \beta_0 + \beta \cdot Z$$

jest modelem regresji dla współczynników intensywności

- Parametery są szacowane na podstawie zliczeń zdarzeń i osobo-czasu
  - zakłada się, że zdarzenia mają rozkład Poissona ze średnią λ
  - przykład *uogólnionego modelu liniowego*

Model PH i model regresji Poissona (1)

- Zauważmy, że

$$\ln \lambda = \beta_0 + \beta \cdot Z$$

można zapisać jako

$$\lambda = e^{\beta_0 + \beta \cdot Z} = \lambda_0 e^{\beta \cdot Z}$$

- Załóżmy, że współczynniki intensywności zmieniają się w przedziałach czasu indeksowanych przy pomocy *t*

$$\lambda_t = e^{\beta_{0t} + \beta \cdot Z} = \lambda_{0t} e^{\beta \cdot Z}$$

lub

$$\lambda(t) = \lambda_0(t) e^{\beta \cdot Z}$$

- Bardzo przypomina model…

Model PH i model regresji Poissona (2)

- Oba modele mają podobną formę

$$\ln rate = \beta_0 + \beta \cdot Z$$

- Regresja Poissona wymaga zgrupowanych danych w przedziałach czasu do oszacowania modelu
  - konieczne szacowanie bazowych współczynników intensywności λ_{0t}
  - mała liczebność próbki stanowi problem (mała precyzja oszacowań)

- Model PH używa danych indywidualnych (nieskończenie wąskie przedziały czasu)
  - bazowa funkcja hazardu nie jest szacowana
  - mniejsze wymagania dotyczące liczebności próbki