

3. Model proporcjonalnych hazardów

Model Coxa

Model proporcjonalnych hazardów

Szacowanie współczynników

Wnioskowanie

Model warstwowy

Modele w analizie przeżycia

- ◆ Najczęściej używany *model proporcjonalnych hazardów*, często nazywany *modelem Coxa*.
- ◆ Alternatywna rodzina *modeli przyśpieszonego czasu do niepowodzenia* (*accelerated failure time model, AFT*).

Hazard

- ◆ „Chwilowy współczynnik intensywności zdarzeń”, jako funkcja czasu
 - Oznaczany przez $\lambda(t)$
- ◆ Jednostką są zdarzenia na jednostkę czasu
 - „Prędkość” występowania zdarzeń
 - $\lambda(t) \geq 0$
 - $\lambda(t)\Delta t \approx$ prawdopodobieństwo zdarzenia w $(t, t + \Delta t]$, pod warunkiem niewystąpienia zdarzenia przed t

Funkcje hazardu: przykłady

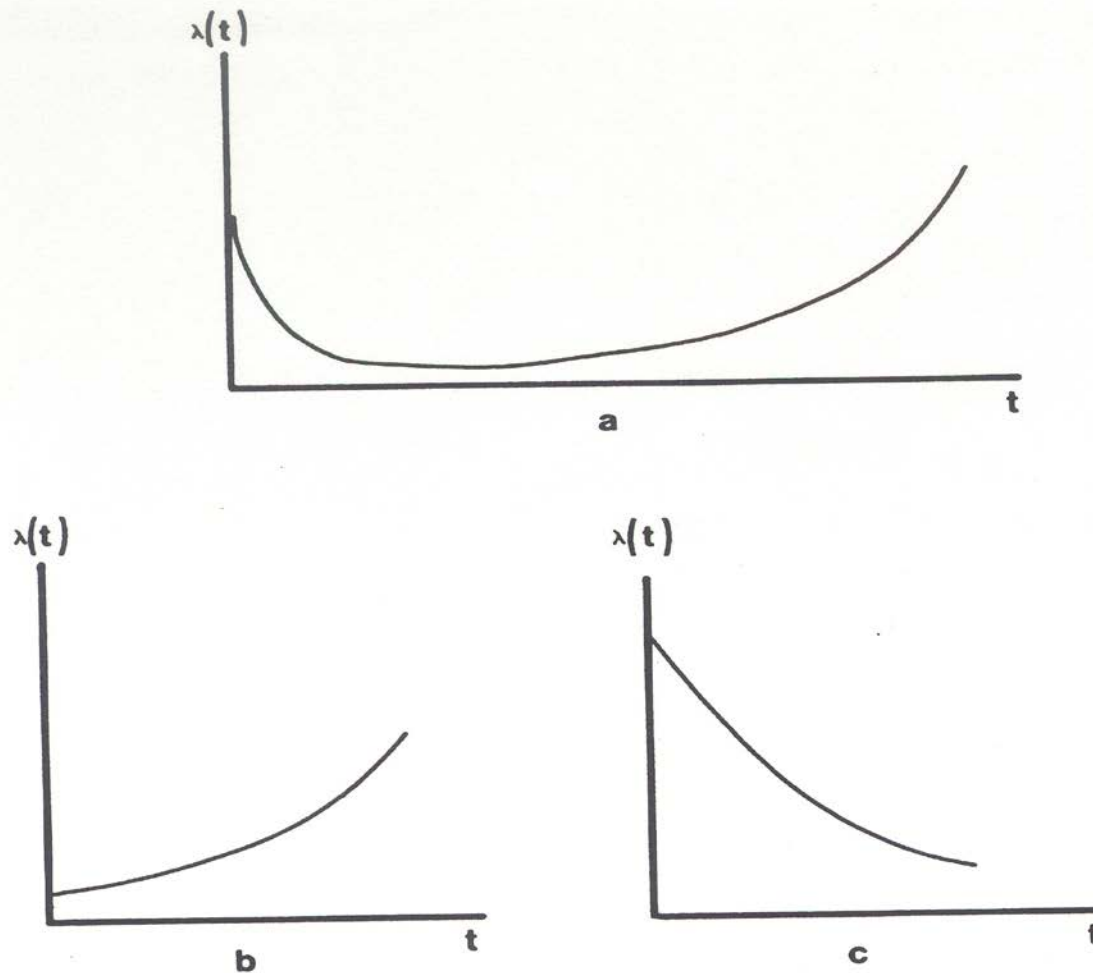


Figure 1.1 Some types of hazard functions: (a) hazard for human mortality; (b) positive aging; (c) negative aging.

Funkcja hazardu i przeżycia

- ◆ Funkcja przeżycia $S(t)$ opisuje prawdopodobieństwo wystąpienia zdarzenia po chwili czasu t
- ◆ Funkcja hazardu $\lambda(t)$ opisuje możliwość wystąpienia zdarzenia tuż po chwili czasu t pod warunkiem, że nie wystąpiło do chwili t
- ◆ Obie funkcje zawierają pełną informację o rozkładzie czasu do wystąpienia zdarzenia:

$$S(t) = e^{-\int_0^t \lambda(u) du}$$

Funkcja skumulowanego hazardu

- ◆ Zdefiniowana jako

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

- ◆ Nieujemna i niemalejąca (gdyż $\lambda(t) \geq 0$)

- ◆ Zachodzi

$$-\ln S(t) = \int_0^t \lambda(u) du = \Lambda(t)$$

- ◆ Funkcja skumulowanego hazardu jest związana z $\lambda(t)$ i $S(t)$, więc również zawiera pełną informację o rozkładzie czasu do wystąpienia zdarzenia

Modelowanie funkcji hazardu

- ♦ Załóżmy, że mamy wektor zmiennych niezależnych

$$\mathbf{Z}(t)' = (Z_1(t), Z_2(t), \dots, Z_p(t))$$

- w szczególnym przypadku, $\mathbf{Z}(t)$ może nie zależeć od czasu

- ♦ Model „względnego hazardu” (“względnego ryzyka”) otrzymujemy przez przyjęcie, że

$$\lambda(t; \mathbf{Z}(t)) = \lambda_0(t) \cdot r(\mathbf{Z}(t))$$

gdzie $r()$ jest ustaloną funkcją, np.

- $r(\mathbf{Z}(t)) = 1 + \beta' \cdot \mathbf{Z}(t)$
 - $r(\mathbf{Z}(t)) = \{1 + \beta' \cdot \mathbf{Z}(t)\}^{-1}$
 - $r(\mathbf{Z}(t)) = e^{\beta' \cdot \mathbf{Z}(t)}$

Model Coxa

- ♦ Zakłada, że

$$\lambda(t; \mathbf{Z}(t)) = \lambda_0(t) \cdot e^{\beta' \mathbf{Z}(t)}$$

- ♦ W szczególności, dla $\mathbf{Z}(t) \equiv \mathbf{Z}$, otrzymujemy *model proporcjonalnych hazardów*:

$$\lambda(t; \mathbf{Z}) = \lambda_0(t) \cdot e^{\beta' \mathbf{Z}}$$

Model proporcjonalnych hazardów

- ◆ Zakłada, że wpływ zmiennej niezależnej Z na czas do wystąpienia niepowodzenia wyraża się poprzez przemnożenie funkcji hazardu przez stałą wartość liczbową
- ◆ Przykład:
 - Funkcja hazardu dla kobiet: $\lambda_F(t)$
 - Funkcja hazardu dla mężczyzn: $\lambda_M(t) = \lambda_F(t) \cdot c$
- ◆ Interpretacja: hazard dla mężczyzn jest c razy mniejszy/wiekszy niż dla kobiet, *niezależnie od chwili czasu*

Iloraz hazardów

- ◆ Równoważne sformułowanie modelu proporcjonalnych hazardów:

$$\frac{\lambda_M(t)}{\lambda_F(t)} = c$$

Iloraz hazardów $HR(t)$

- ◆ Model proporcjonalnych hazardów (PH) zakłada, że iloraz hazardów dla dwóch wartości zmiennej niezależnej Z jest stały w czasie
 - często iloraz jest (niepoprawnie) nazywany *ryzykiem względnym* (*relative risk*)

Iloraz hazardów w modelu PH (1)

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

♦ Przykład:

- Przyjmijmy $Z=0, 1, 2$, dla TNM= I, II, III

- Funkcja hazardu dla TNM I: $\lambda_I(t) = \lambda_0(t) \cdot e^{0 \cdot \beta} = \lambda_0(t)$
- Funkcja hazardu dla TNM II: $\lambda_{II}(t) = \lambda_0(t) \cdot e^{1 \cdot \beta} = \lambda_0(t) \cdot e^{\beta}$
- Funkcja hazardu dla TNM III: $\lambda_{III}(t) = \lambda_0(t) \cdot e^{2 \cdot \beta} = \lambda_0(t) \cdot e^{2\beta}$

Iloraz hazardów w modelu PH (2)

♦ Przykład (cd.):

- Funkcja hazardu dla TNM I: $\lambda_I(t) = \lambda_0(t) \cdot e^{0 \cdot \beta} = \lambda_0(t)$
- Funkcja hazardu dla TNM II: $\lambda_{II}(t) = \lambda_0(t) \cdot e^{1 \cdot \beta} = \lambda_0(t) \cdot e^{\beta}$
- Funkcja hazardu dla TNM III: $\lambda_{III}(t) = \lambda_0(t) \cdot e^{2 \cdot \beta} = \lambda_0(t) \cdot e^{2\beta}$

- Iloraz hazardów dla $Z=1$ vs. $Z=0$:

$$\lambda_{II}(t)/\lambda_I(t) = \lambda_0(t) \cdot e^{\beta} / \lambda_0(t) = e^{\beta}$$

- Dla $Z=2$ vs. $Z=1$:

$$\lambda_{III}(t)/\lambda_{II}(t) = \lambda_0(t) \cdot e^{2\beta} / \lambda_0(t) \cdot e^{\beta} = e^{\beta}$$

♦ Model zakłada, że wzrost Z o 1 zwiększa hazard e^{β} razy

- iloraz hazardu (HR) dla $Z=z+1$ vs. $Z=z$ wynosi e^{β}

Efekt zmiennej niezależnej w modelu PH

- ♦ Wniosek: w modelu PH z jedną zmienną niezależną:

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

e^{β} wyraża proporcjonalną zmianę hazardu wynikającą ze wzrostu Z o 1 jednostkę

- ♦ $\beta > 0 \rightarrow$ hazard wzrasta; $\beta < 0 \rightarrow$ maleje

- ♦ Przykład:

- Przyjmijmy $Z=0, 1, 2$ dla TNM= I, II, III
- Dla $\beta=1$, otrzymujemy $e^{\beta} = e^1 = 2.73$
- Wzrost TNM o jeden stopień zwiększa hazard (np. śmiertelność) 2.73 razy

Poziom odniesienia zmiennej niezależnej (1)

♦ Przykład:

- $Z=0, 1, 2$, dla TNM= I, II, III
 - Funkcja hazardu dla TNM I: $\lambda_I(t) = \lambda_0(t) \cdot e^{0\beta} = \lambda_0(t)$
 - Funkcja hazardu dla TNM II: $\lambda_{II}(t) = \lambda_0(t) \cdot e^{1\beta} = \lambda_I(t) \cdot e^{\beta}$
 - Funkcja hazardu dla TNM III: $\lambda_{III}(t) = \lambda_0(t) \cdot e^{2\beta} = \lambda_I(t) \cdot e^{2\beta}$
- ♦ Wniosek: funkcje hazardu dla TNM II i III są zdefiniowane *względem* funkcji hazardu dla TNM I

Poziom odniesienia zmiennej niezależnej (2)

♦ Przykład (cd.):

- $Z=2, 1, 0$, dla $TNM= I, II, III$
- $TNM III: \lambda_{III}(t) = \lambda_0(t) \cdot e^{0\beta} = \lambda_0(t)$
- $TNM II: \lambda_{II}(t) = \lambda_0(t) \cdot e^{1\beta} = \lambda_{III}(t) \cdot e^{\beta}$
- $TNM I: \lambda_I(t) = \lambda_0(t) \cdot e^{2\beta} = \lambda_{III}(t) \cdot e^{2\beta}$

♦ Wniosek: funkcje hazardu dla $TNM I$ i II są zdefiniowane względem funkcji hazardu dla $TNM III$

♦ Poziom odniesienia ($Z=0$) uległ zmianie!

Częściowa funkcja wiarygodności

- ◆ Przedstawmy dane \mathbf{Y} jako zbiór zmiennych $A_1, B_1, \dots, A_m, B_m$
- ◆ Niech $A^{(j)} = (A_1, \dots, A_j)$ and $B^{(j)} = (B_1, \dots, B_j)$
- ◆ Załóżmy, że gęstość łącznego rozkładu $(A^{(m)}, B^{(m)})$ można przedstawić jako

$$\prod_{j=1}^m f(b_j | b^{(j-1)}, a^{(j-1)}; \theta, \beta) \prod_{j=1}^m f(a_j | b^{(j)}, a^{(j-1)}; \beta)$$

częściowa funkcja
wiarygodności

- ◆ Nie jest funkcją wiarygodności w „zwykłym” znaczeniu. Ale może być użyta do wnioskowania o β
 - z pewną stratą na efektywności szacowania

Częściowa funkcja wiarygodności dla modelu Coxa

$$\lambda(t; \mathbf{Z}(t)) = \lambda_0(t) e^{\beta' \mathbf{Z}(t)}$$

- ◆ K obserwacji pełnych (zdarzeń) dla $t_{(1)} < \dots < t_{(K)}$ (bez „zlepień”)
- ◆ B_k = cenzurowanie w $[t_{(k-1)}, t_{(k)})$ i jedno zdarzenie w $[t_{(k)}, t_{(k)} + h_k)$
- ◆ A_k = zdarzenie dla jednostki k w $[t_{(k)}, t_{(k)} + h_k)$
- ◆ $Y_i(t) = 1$ jeśli jednostka pod obserwacją dla t , 0 w p.p.
- ◆ Dla **niezależnego** cenzurowania

$$\begin{aligned} L(\beta) &= \prod_{k=1}^K f(a_k | b^{(k)}, a^{(k-1)}; \beta) = \prod_{k=1}^K \frac{\lambda(t_{(k)}; \mathbf{Z}_{(k)}(t_{(k)})) h_k}{\sum_{l=1}^n Y_l(t_{(k)}) \lambda(t_{(k)}; \mathbf{Z}_l(t_{(k)})) h_k} \\ &= \prod_{k=1}^K \frac{\lambda_0(t_{(k)}) h_k e^{\beta' \mathbf{Z}_{(k)}(t_{(k)})}}{\sum_{l=1}^n Y_l(t_{(k)}) \lambda_0(t_{(k)}) h_k e^{\beta' \mathbf{Z}_l(t_{(k)})}} = \prod_{k=1}^K \frac{e^{\beta' \mathbf{Z}_{(k)}(t_{(k)})}}{\sum_{l=1}^n Y_l(t_{(k)}) e^{\beta' \mathbf{Z}_l(t_{(k)})}} \end{aligned}$$

Jednoczesne zdarzenia (1)

- ◆ Niech $r_i(t) = \exp\{\beta' \mathbf{Z}_i(t)\}$
- ◆ Wkład pojedynczej obserwacji pełnej (zdarzenia) do $L(\beta)$ to

$$\frac{r_i(t)}{\sum_{l=1}^n Y_l(t) r_l(t)}$$

- ◆ Rozważmy 3 jednoczesne zdarzenia w t wśród 5 obserwacji
- ◆ Gdybyśmy znali dokładne czasy (kolejność) zdarzeń, to wkład do $L(\beta)$ wynosiłoby

$$\begin{aligned} & \frac{r_1}{r_1 + r_2 + r_3 + r_4 + r_5} \cdot \frac{r_2}{r_2 + r_3 + r_4 + r_5} \cdot \frac{r_3}{r_3 + r_4 + r_5} \quad \text{lub} \quad \frac{r_1}{r_1 + r_2 + r_3 + r_4 + r_5} \cdot \frac{r_3}{r_2 + r_3 + r_4 + r_5} \cdot \frac{r_2}{r_2 + r_4 + r_5} \\ & \frac{r_2}{r_1 + r_2 + r_3 + r_4 + r_5} \cdot \frac{r_1}{r_1 + r_3 + r_4 + r_5} \cdot \frac{r_3}{r_3 + r_4 + r_5} \quad \text{lub} \quad \frac{r_2}{r_1 + r_2 + r_3 + r_4 + r_5} \cdot \frac{r_3}{r_1 + r_3 + r_4 + r_5} \cdot \frac{r_1}{r_1 + r_4 + r_5} \\ & \frac{r_3}{r_1 + r_2 + r_3 + r_4 + r_5} \cdot \frac{r_1}{r_1 + r_2 + r_4 + r_5} \cdot \frac{r_2}{r_2 + r_4 + r_5} \quad \text{lub} \quad \frac{r_3}{r_1 + r_2 + r_3 + r_4 + r_5} \cdot \frac{r_2}{r_2 + r_1 + r_4 + r_5} \cdot \frac{r_1}{r_1 + r_4 + r_5} \end{aligned}$$

Jednoczesne zdarzenia (2)

- ◆ Metoda Breslowa:
$$\frac{r_1 r_2 r_3}{(r_1 + r_2 + r_3 + r_4 + r_5)^3}$$
- ◆ Metoda Efrona:
$$\frac{r_1 r_2 r_3}{(r_1 + r_2 + r_3 + r_4 + r_5) \left(\frac{2r_1}{3} + \frac{2r_2}{3} + \frac{2r_3}{3} + r_4 + r_5 \right) \left(\frac{r_1}{3} + \frac{r_2}{3} + \frac{r_3}{3} + r_4 + r_5 \right)}$$
- ◆ Metoda „dokładna”: rozważyć wszystkie 3! konfiguracje i użyć średniej odpowiednich wkładów do $L(\beta)$.
 - Problem: ogólnie, $d! \cdot d$ mnożeń; np. $8! \cdot 8 = 322560$
- ◆ Idea:

$$\frac{r_1}{r_1 + r_2 + r_3 + r_4 + r_5} \cdot \frac{r_2}{r_2 + r_3 + r_4 + r_5} + \frac{r_2}{r_1 + r_2 + r_3 + r_4 + r_5} \cdot \frac{r_1}{r_1 + r_3 + r_4 + r_5}$$

$$= \int_0^\infty \left\{ 1 - \exp\left(-\frac{r_1 t}{r_3 + r_4 + r_5}\right) \right\} \left\{ 1 - \exp\left(-\frac{r_2 t}{r_3 + r_4 + r_5}\right) \right\} e^{-t} dt$$
 - Całkujemy numerycznie; złożoność $O(d^2)$

Jednoczesne zdarzenia (3)

- ◆ Potraktujmy dane jako „dyskretne”. Wówczas możemy rozważyć

$$\frac{r_1 r_2 r_3}{\sum_{(i_1, i_2, i_3)} r_{i_1} r_{i_2} r_{i_3}}$$

- ◆ Problem numeryczny: dla 10 jednoczesnych zdarzeń wśród 1000 obserwacji, mianownik zawiera $2.6 \cdot 10^{23}$ składników.
- ◆ Problem teoretyczny: odpowiedni model to

$$\frac{\lambda_i(t)}{1 - \lambda_i(t)} = \frac{\lambda_0(t)}{1 - \lambda_0(t)} e^{\beta' Z_i}$$

Jednoczesne zdarzenia (4)

- ◆ Metoda Breslowa: najszybsza, lecz najmniej dokładna.
 - Bardzo często domyślna w różnych pakietach statystycznych!
- ◆ Metoda „dokładna”: złożona numerycznie.
- ◆ Metoda Efrona: szybka i daje dobre przybliżenie „dokładnej”.
 - O ile liczba jednoczesnych zdarzeń nie jest zbyt duża.
- ◆ Metoda dla danych dyskretnych: nie dla modelu PH.

Pochodna logarytmu (częściowej) funkcji wiarygodności i macierz informacji

♦ Średnia ważona $\mathbf{Z}(t)$:
$$\bar{Z}(\beta, t) = \sum_{i=1}^n Z_i(t) \frac{Y_i(t) e^{\beta' Z_i(t)}}{\sum_{j=1}^n Y_j(t) e^{\beta' Z_j(t)}} \equiv \sum_{i=1}^n w_i(t) Z_i(t)$$

♦ *Gradient („score”)*
$$U(\beta) \equiv \frac{\partial \ln L(\beta)}{\partial \beta} = \sum_{k=1}^K \{Z_{(k)}(t_{(k)}) - \bar{Z}(\beta, t_{(k)})\}$$

♦ *(Obserwowana) Macierz informacji*
$$I(\beta) = -\frac{\partial^2 \ln L}{\partial \beta \partial \beta^T} = \sum_{k=1}^K V(\beta, t_{(k)})$$

gdzie
$$V(\beta, t) = \sum_{i=1}^n w_i(t) \{Z_i(t) - \bar{Z}(\beta, t)\} \{Z_i(t) - \bar{Z}(\beta, t)\}^T$$

- ważona wariancja \mathbf{Z} w chwili t

Szacowanie parametrów modelu Coxa

- ♦ Oszacowanie β uzyskujemy rozwiązując $U(\hat{\beta}) = 0$
- ♦ Można pokazać, że asymptotycznie $\hat{\beta} \sim N[\beta, \{E(I(\beta))\}^{-1}]$
- ♦ Estymator wariancji: $I(\hat{\beta})^{-1}$

Testowanie hipotez (1)

♦ Testowanie hipotez $H_0: \beta = \beta_0$ (np. $\beta = 0$)

- Test ilorazu wiarygodności: $2\{\ln L(\hat{\beta}) - \ln L(\beta_0)\} \equiv 2\{l(\hat{\beta}) - l(\beta_0)\}$

- Test Walda: $(\hat{\beta} - \beta_0)^T I(\hat{\beta})(\hat{\beta} - \beta_0)$

- Test „score”: $U^T(\beta_0)I(\beta_0)^{-1}U(\beta_0)$

♦ Rozkład χ^2 z p (liczba elementów β) stopniami swobody

Testowanie hipotez (2)

- ◆ Test Walda, ilorazu wiarygodności i „score” równoważne asymptotycznie, ale nie dla „małych” próbek
 - Preferowany test ilorazu wiarygodności
- ◆ Test „score” dla jednej zmiennej kategorycznej równoważny testowi logrank.
- ◆ Uwaga na braki danych przy teście ilorazu wiarygodności! Modele powinny być szacowane na tym samym zbiorze danych.

Równoważność asymptotyczna testów

- ◆ Test Walda, ilorazu wiarygodności i „score” równoważne asymptotycznie.

- Test Walda: przybliżenie $l(\beta)$ funkcją kwadratową w $\hat{\beta}$

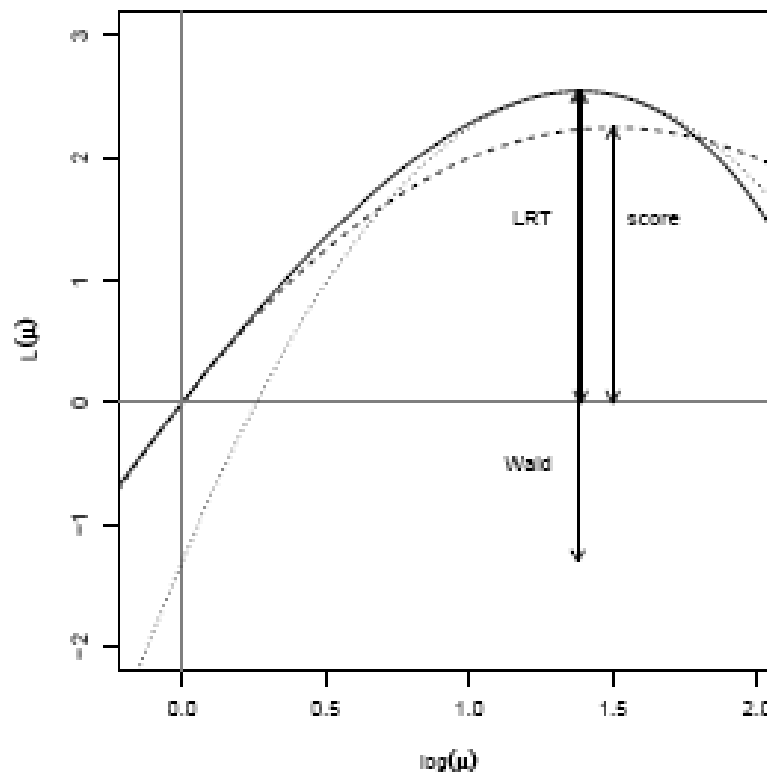
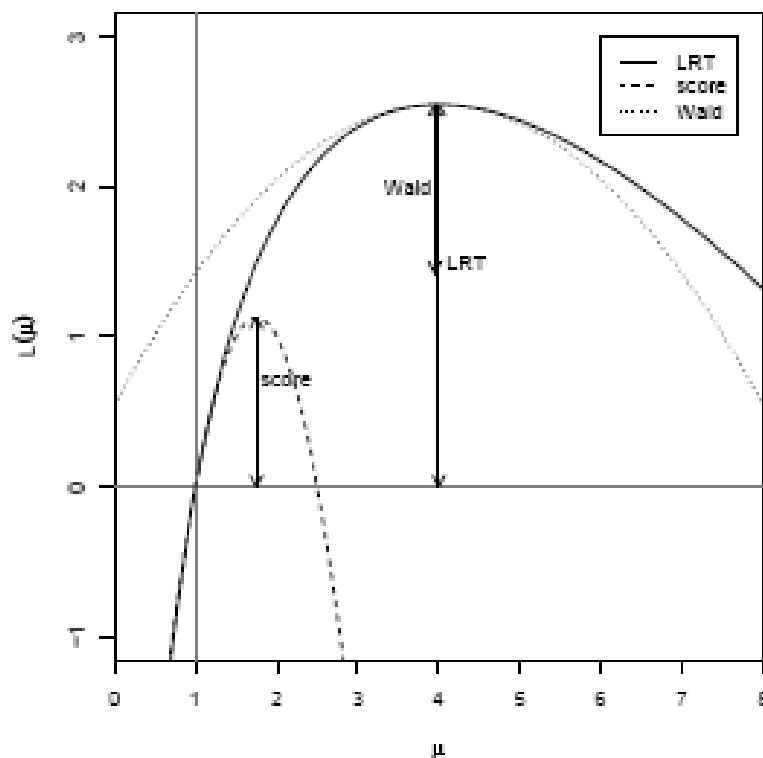
$$l(\beta) \approx l(\hat{\beta}) + (\beta - \hat{\beta})^T U(\hat{\beta}) - (\beta - \hat{\beta})^T I(\hat{\beta}) (\beta - \hat{\beta}) / 2$$
$$l(\hat{\beta}) - l(\beta_0) \approx (\hat{\beta} - \beta_0)^T I(\hat{\beta}) (\hat{\beta} - \beta_0) / 2$$

- Test „score”: przybliżenie $l(\beta)$ funkcją kwadratową w β_0

$$l(\beta) \approx l(\beta_0) + (\beta - \beta_0)^T U(\beta_0) - (\beta - \beta_0)^T I(\beta_0) (\beta - \beta_0) / 2$$
$$U(\hat{\beta}) \approx U(\beta_0) - I(\beta_0) (\hat{\beta} - \beta_0) \Rightarrow I(\beta_0)^{-1} U(\beta_0) \approx (\hat{\beta} - \beta_0)$$
$$l(\hat{\beta}) - l(\beta_0) \approx U(\beta_0)^T I(\beta_0)^{-1} U(\beta_0) - U(\beta_0)^T I(\beta_0)^{-1} U(\beta_0) / 2 = U(\beta_0)^T I(\beta_0)^{-1} U(\beta_0) / 2$$

Własności testów dla „małych” próbek

- ◆ Test Walda, ilorazu wiarygodności i „score” równoważne asymptotycznie, ale nie dla „małych” próbek
 - Przykład: test $H_0: \mu=1$ dla obserwacji $Y=4$ z rozkładu Poissona



Testowanie hipotez liniowych

◆ Testowanie liniowych hipotez $H_0: C\beta = \beta_0$

- Np. $(0,0,1,-1) (\beta_1, \beta_2, \beta_3, \beta_4)^T = 0 \Leftrightarrow \beta_3 = \beta_4$

◆ Test Walda: $(C\hat{\beta} - \beta_0)^T \left\{ CI(\hat{\beta})^{-1} C^T \right\}^{-1} (C\hat{\beta} - \beta_0) \sim \chi^2_{rank(C)}$

◆ Test ilorazu wiarygodności:

- Szacujemy model „alternatywny” (β)
- Szacujemy model odpowiadający H_0 (β^*)

- Statystyka testowa: $2\{l(\hat{\beta}^*) - l(\hat{\beta})\} \sim \chi^2_{rank(C)}$

Przedziały ufności dla współczynników modelu

- ◆ Dopasowanie modelu do danych pozwala na uzyskanie oszacowania współczynników β i ich błędów standardowych (SE)
- ◆ Przy ich użyciu możemy skonstruować 95% przedział ufności dla β_i : $[\hat{\beta}_i - 1.96 \cdot SE(\hat{\beta}_i), \hat{\beta}_i + 1.96 \cdot SE(\hat{\beta}_i)]$
- ◆ Przedział ufności dla ilorazu hazardów:

$$[e^{\hat{\beta}_i - 1.96 \cdot SE(\hat{\beta}_i)}, e^{\hat{\beta}_i + 1.96 \cdot SE(\hat{\beta}_i)}]$$

Przedział ufności oparty na „profilowanej” funkcji wiarygodności

- ◆ Użyteczne jeśli problem z przybliżeniem rozkładem normalnym (Wald)
- ◆ Idea: dla różnych wartości β_i dopasowujemy model do danych i wykreślamy wartość $l(\hat{\beta})$
- ◆ 95% przedział ufności dla β_i zawiera wartości β_i , dla których statystyka $LRT \leq 3.84$, czyli
$$l(\beta_i) \geq l(\hat{\beta}_i) - 3.84/2$$
 - może wymagać „ręcznego” programowania

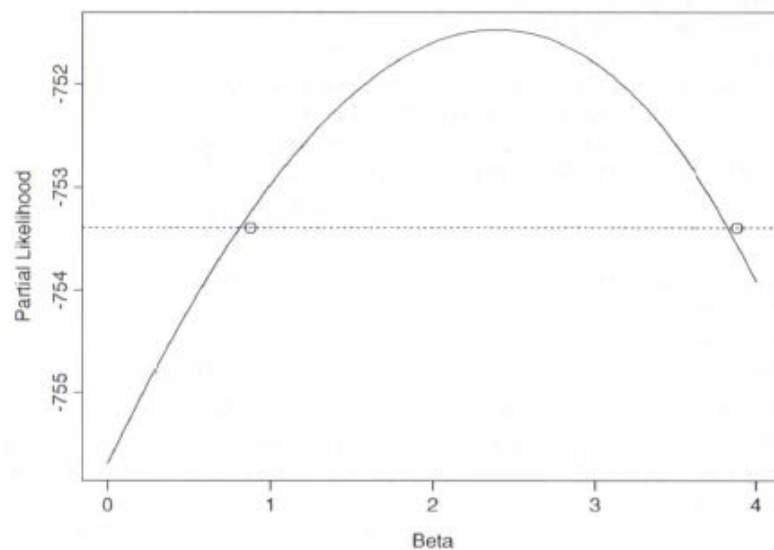


FIGURE 3.2: Profile likelihood curve and 95% confidence interval cutoff for prothrombin time

„Nieskończone” współczynniki

- ◆ Gdy np. w jednej z grup nie było zdarzeń:

Grupa	Żyjący	Zgony
Leczona	40	0
Kontrolna	30	10

- ◆ Przykład (S+):

```
>fit <- coxph(Surv(futime, fustat) ~ rx + fustat, ovarian)
Warning messages:
  Loglik converged before variable 2 ; beta may be infinite.
```

```
>summary(fit)
n= 26

      coef exp(coef) se(coef)      z      p
rx    -0.557  5.73e-01    0.62 -0.898 0.37
fustat 10.582  3.94e+04   34.38  0.308 0.76
```

```
Likelihood ratio test= 30.8  on 2 df,   p=2.05e-07
Wald test               = 0.9  on 2 df,   p=0.637
Efficient score test = 29.1  on 2 df,   p=4.83e-07
```

Testy dla „nieskończonych” współczynników

◆ Test ilorazu wiarygodności:

$$l(10.582) - l(0) \approx l(\infty) - l(0)$$

- wynik ma sens

◆ Test „score” : przybliżenie funkcją kwadratową w 0

- max dla 3.4
- wynik ma sens

◆ Test Walda : przybliżenie funkcją kwadratową w 10.6

- B. złe, wynik nie ma sensu

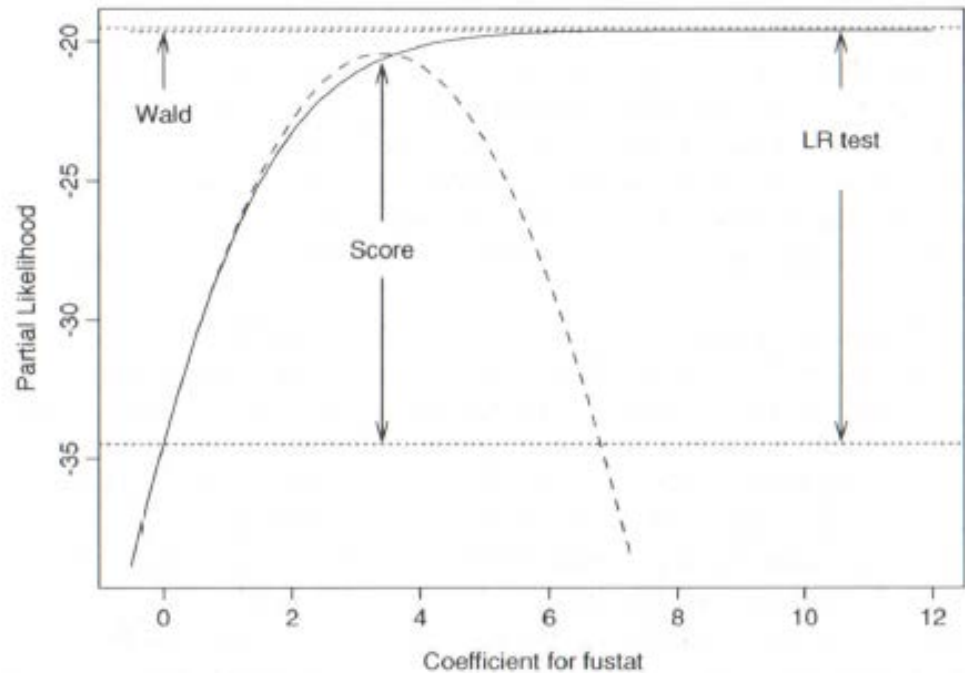
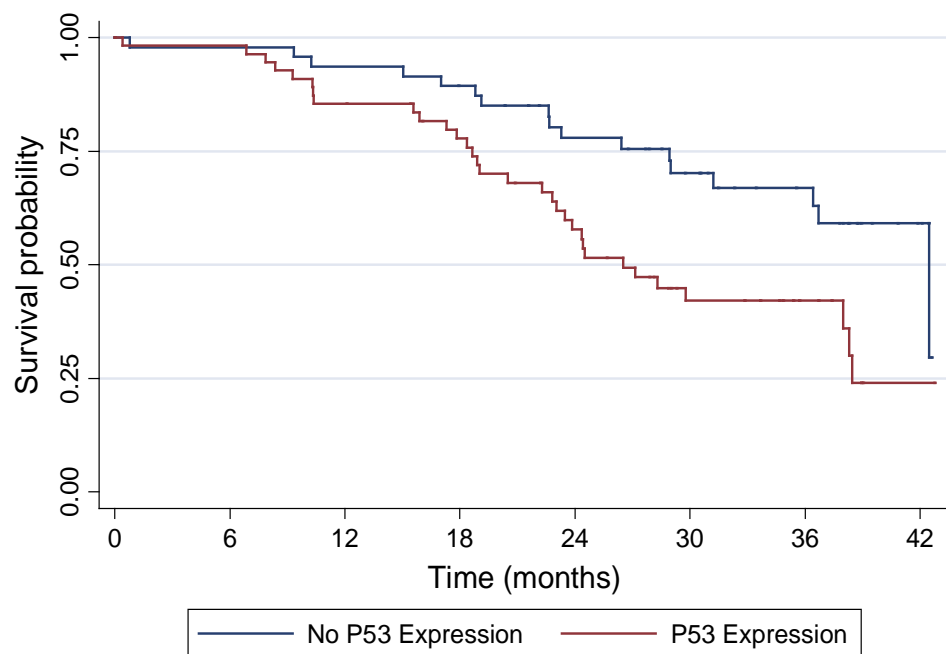


FIGURE 3.3: Likelihood ratio, score, and Wald tests for an infinite coefficient

Model PH: niedrobnokomórkowy rak płuca (NSCLC) (1)

- ♦ Laudański *et al.*, *Eur Respir J* (2001)
- ♦ Wpływ ekspresji białka P53 na czas przeżycia 102 pacjentów
- ♦ Test logrank: $p=0.008 \rightarrow$



Model PH: NSCLC (2)

- ♦ Zmienna „P53e” = 0 jeśli brak ekspresji, 1 jeśli ekspresja

_t	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
P53e	.7859412	.3021812	2.601	0.009	.1936769	1.378205
	Haz. Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
P53e	2.194471	.663128	2.601	0.009	1.213704	3.967775

- ♦ Wniosek: ekspresja P53 zwiększa hazard zgonu (śmiertelność) w przybliżeniu 2 razy

- ♦ Efekt jest istotnie statystyczny na poziomie istotności 5%

- Uwaga: dla testu logrank, $p=0.008$

Alternatywna postać modelu PH (1)

- ♦ Zauważmy, że

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

implikuje

$$\Lambda(t) = \Lambda_0(t) \cdot e^{\beta \cdot Z}$$

- ♦ W konsekwencji

$$e^{-\Lambda(t)} = e^{-\Lambda_0(t) \cdot \exp(\beta \cdot Z)}$$

oraz

$$S(t) = S_0(t)^{\exp(\beta \cdot Z)}$$

gdzie $S_0(t)$ jest funkcją przeżycia odpowiadającą bazowej funkcji hazardu

Alternatywna postać modelu PH (2)

$$S(t) = S_0(t)^{\exp(\beta \cdot Z)}$$

- ♦ Wzrost $Z = z \rightarrow Z = z + 1$ powoduje:

$$S_z(t) = S_0(t)^{\exp(\beta \cdot z)} \quad \rightarrow \quad S_{z+1}(t) = S_0(t)^{\exp(\beta \cdot z + \beta)} = S_z(t)^{\exp(\beta)}$$

- ♦ $\beta > 0$ oznacza zmniejszenie wartości p-stwa przeżycia
 - wyższą wartość hazardu

Model PH na skali logarytmicznej

- ◆ Ponieważ

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

$$\Lambda(t) = \Lambda_0(t) \cdot e^{\beta \cdot \mathbf{Z}}$$

to

$$\ln \lambda(t) = \ln \lambda_0(t) + \beta \cdot \mathbf{Z}$$

$$\ln \Lambda(t) = \ln \Lambda_0(t) + \beta \cdot \mathbf{Z}$$

- ◆ Na skali logarytmicznej otrzymujemy model liniowy

- interpretacja β jak w modelu prostej regresji liniowej

- ◆ $\Lambda(t) = -\ln S(t)$, skąd

$$\ln \{ -\ln S(t) \} = \ln \{ -\ln S_0(t) \} + \beta \cdot \mathbf{Z}$$

- Związek użyteczny dla graficznego sprawdzenia założenia PH

Szacowanie bazowej funkcji hazardu

- ◆ Dla zaobserwowanych czasów zdarzeń $t_{(1)} < \dots < t_{(K)}$ kładziemy

$$\hat{\lambda}_0(t_{(k)}) = 1 - \hat{\xi}_k$$

gdzie

$$\sum_{j: t_j = t_{(k)}} \frac{\delta_j e^{\hat{\beta}' Z_j(t_{(k)})}}{1 - \left(\hat{\xi}_k\right)^{e^{\hat{\beta}' Z_j(t_{(k)})}}} = \sum_{i=1}^n Y_i(t_{(k)}) e^{\hat{\beta}' Z_i(t_{(k)})}$$

- ◆ Jeśli nie ma jednoczesnych czasów zdarzeń, to

$$\hat{\xi}_k = \left\{ 1 - \frac{e^{\hat{\beta}' Z_{(k)}(t_{(k)})}}{\sum_{i=1}^n Y_i(t_{(k)}) e^{\hat{\beta}' Z_i(t_{(k)})}} \right\} e^{-\hat{\beta}' Z_{(k)}(t_{(k)})}$$

Szacowanie bazowej funkcji przeżycia

- ◆ Podobnie jak w metodzie Kaplana-Meiera, używamy

$$\hat{S}_0(t) = \prod_{t_{(k)} \leq t} \hat{\xi}_k$$

- Bez zmiennych niezależnych mamy

$$\sum_{j: t_j = t_{(k)}} \frac{\delta_j}{1 - \hat{\xi}_k} \equiv \frac{d_k}{1 - \hat{\xi}_k} = \sum_{i=1}^n Y_i(t_{(k)}) \equiv n_k \Rightarrow \hat{\xi}_k = \frac{n_k - d_k}{n_k}$$

czyli estymator Kaplana-Meiera

Szacowanie bazowej funkcji skumulowanego hazardu

- ◆ Korzystając z oszacowania funkcji przeżycia, mamy

$$\hat{\Lambda}_0(t) = -\ln \hat{S}_0(t) = -\sum_{t_{(k)} \leq t} \ln \hat{\xi}_k$$

- ◆ Alternatywnie, kładąc $(\xi_k)^{e^{-\hat{\beta}'Z_j(t_{(k)})}} \approx 1 + e^{-\hat{\beta}'Z_j(t_{(k)})} \ln \xi_k$

dostajemy

$$\sum_{j:t_j=t_{(k)}} \frac{\delta_j}{-\ln \tilde{\xi}_k} = \sum_{i=1}^n Y_i(t_{(k)}) e^{\hat{\beta}'Z_i(t_{(k)})} \Rightarrow \tilde{\xi}_k = \exp \left(-\frac{d_k}{\sum_{i=1}^n Y_i(t_{(k)}) e^{\hat{\beta}'Z_i(t_{(k)})}} \right)$$

i w konsekwencji estymator Breslowa:

$$\tilde{\Lambda}_0(t) = -\ln \tilde{S}_0(t) = \sum_{t_{(k)} \leq t} \frac{d_k}{\sum_{i=1}^n Y_i(t_{(k)}) e^{\hat{\beta}'Z_i(t_{(k)})}}$$

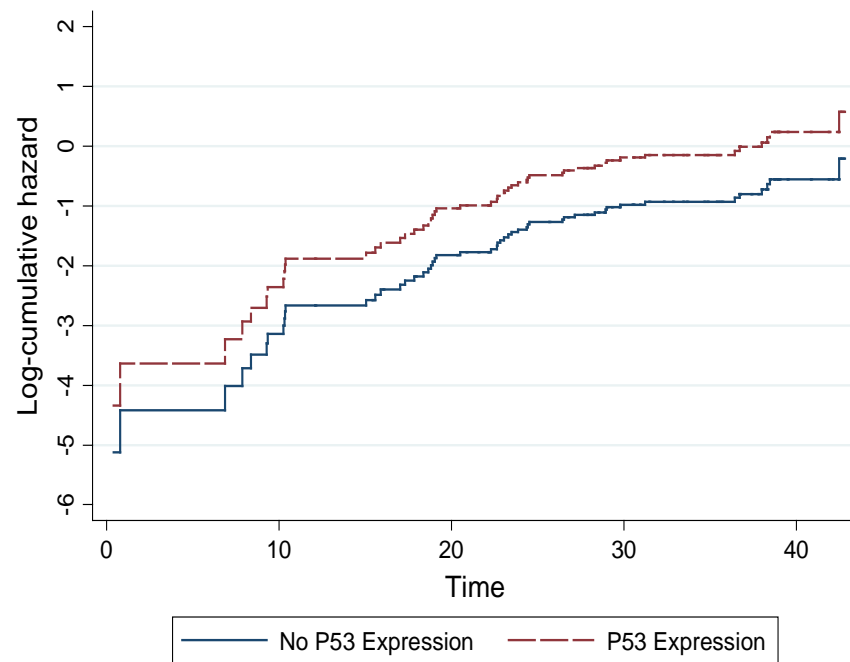
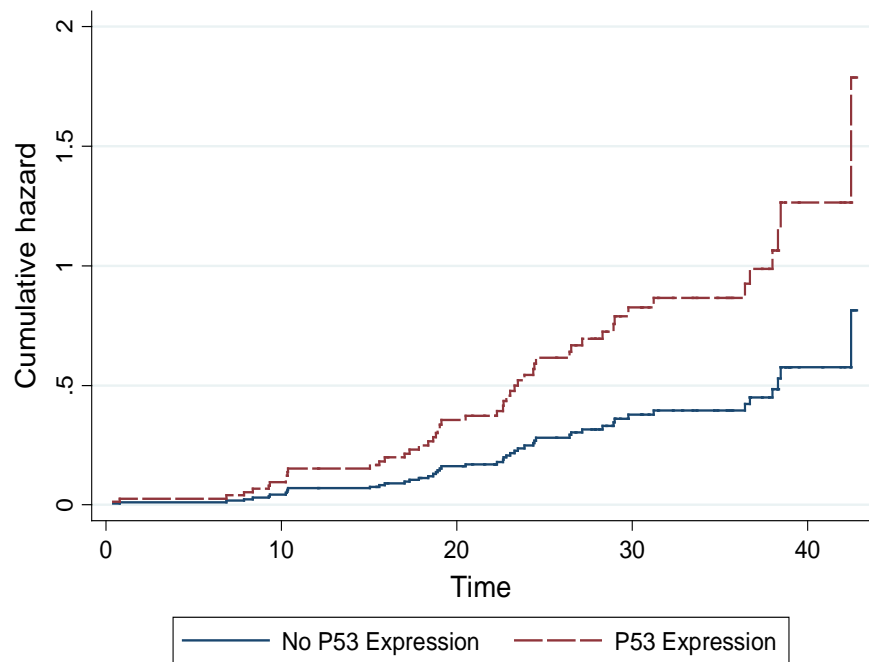
Model PH: NSCLC (3)

- ♦ Zmienna „P53e” = 0 jeśli brak ekspresji, 1 jeśli ekspresja

_t	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
P53e	.7859412	.3021812	2.601	0.009	.1936769	1.378205
	Haz. Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
P53e	2.194471	.663128	2.601	0.009	1.213704	3.967775

- ♦ Wniosek: ekspresja P53 zwiększa śmiertelność w przybliżeniu 2 razy w porównaniu do braku ekspresji
 - logarytm hazardu wzrasta o 0.78

Model PH: NSCLC, ekspresja P53



◆ Oszacowanie (logarytmu) funkcji skumulowanego hazardu

Model PH dla wielu zmiennych niezależnych

- ♦ Model dla trzech zmiennych Z_1 , Z_2 , i Z_3 przyjmuje postać:

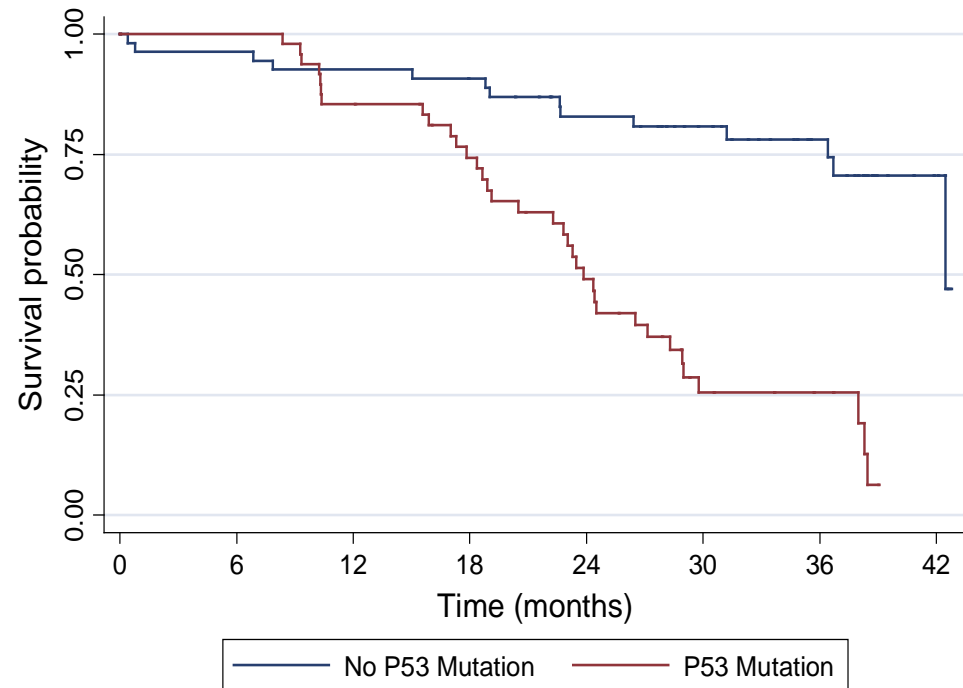
$$\lambda(t) = \lambda_0(t) e^{\beta_1 Z_1 + \beta_2 Z_2 + \beta_3 Z_3}$$

- ♦ $\exp(\beta_1)$ opisuje zmianę hazardu wynikającą ze wzrostu Z_1 o 1 jednostkę dla ustalonych wartości Z_2 i Z_3
- ♦ Model umożliwia skorygowanie efektu Z_1 ze względu na możliwy efekt Z_2 i Z_3

Model PH: NSCLC (4)

- ♦ Laudański *et al.*, *Eur Respir J* (2001)
- ♦ Mutacja genu P53 (analiza DNA)
- ♦ Wpływ na czas przeżycia 102 chorych
- ♦ Test logrank: $p < 0.001 \rightarrow$

mutacja	ekspresja +
-	27.8%
+	82.3%



Model PH: NSCLC (5)

- ♦ Zmienna „P53m” = 0 jeśli brak mutacji, 1 jeśli mutacja

_t	Haz. Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
P53m	5.670694	2.308899	4.262	0.000	2.553043	12.59547
P53e	.8191471	.3038457	-0.538	0.591	.3959371	1.694718

- ♦ Wniosek: skorygowanie ze względu na efekt mutacji „usuwa” efekt ekspresji białka P53 (współliniowość?)
- ♦ Ocena ekspresji białka nie „dodaje informacji” jeśli wiemy czy występuje mutacja

Model PH: NSCLC (6)

_t	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
P53m	1.735312	.4071634	4.26	0.000	.937286	2.533337
P53e	-.1994916	.3709294	-0.54	0.591	-.9264999	.5275166

_t	Haz. Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
P53m	5.670694	2.308899	4.262	0.000	2.553043	12.59547
P53e	.8191471	.3038457	-0.538	0.591	.3959371	1.694718

- ◆ Obecność mutacji zwiększa śmiertelność ponad 5.5 razy
- ◆ Teoretycznie, model definiuje 4 „grupy ryzyka”:

mutacja (P53m)	ekspresja (P53e)	iloraz hazardów
- (0)	- (0)	1
- (0)	+ (1)	$e^{1.74 \cdot 0 - 0.20 \cdot 1} = 1 \cdot e^{-0.20} = 0.82$
+ (1)	- (0)	$e^{1.74 \cdot 1 - 0.20 \cdot 0} = e^{1.74} \cdot 1 = 5.67$
+ (1)	+ (1)	$e^{1.74 \cdot 1 - 0.20 \cdot 1} = 5.67 \cdot 0.82 = 4.65$

Model PH: marskość wątroby

- ◆ Christiansen *et al.*, *Gastroenterology* (1985)
- ◆ Próba kliniczna z 248 chorymi na marskość wątroby
- ◆ Czas przeżycia
- ◆ W oparciu o model PH, zdefiniowano indeks prognostyczny

Marskość wątroby: indeks prognostyczny

	Coeff.	Std. Err.	P> z
bilirubin [log (μmol/L)]	2.51	0.32	<0.001
age [exp {(years-20)/10}]	0.0069	0.0016	<0.001
cirrhosis [no=0, yes=1]	0.88	0.22	<0.001
albumin [g/L]	-0.50	0.018	<0.006
cholestiasis [no=0, yes=1]	0.68	0.27	0.01
drug [azathioprine=0, placebo=1]	0.52	0.21	0.01

♦ indeks prognostyczny:

$$\text{PI} = 2.51 \times \text{bilirubin} + 0.0069 \times \text{age} + 0.88 \times \text{cirrhosis} - 0.50 \times \text{albumin} + 0.68 \times \text{cholestiasis} + 0.52 \times \text{drug}$$

Marskość wątroby: walidacja indeksu

85 chorych z
grupy placebo
w innej próbie
klinicznej

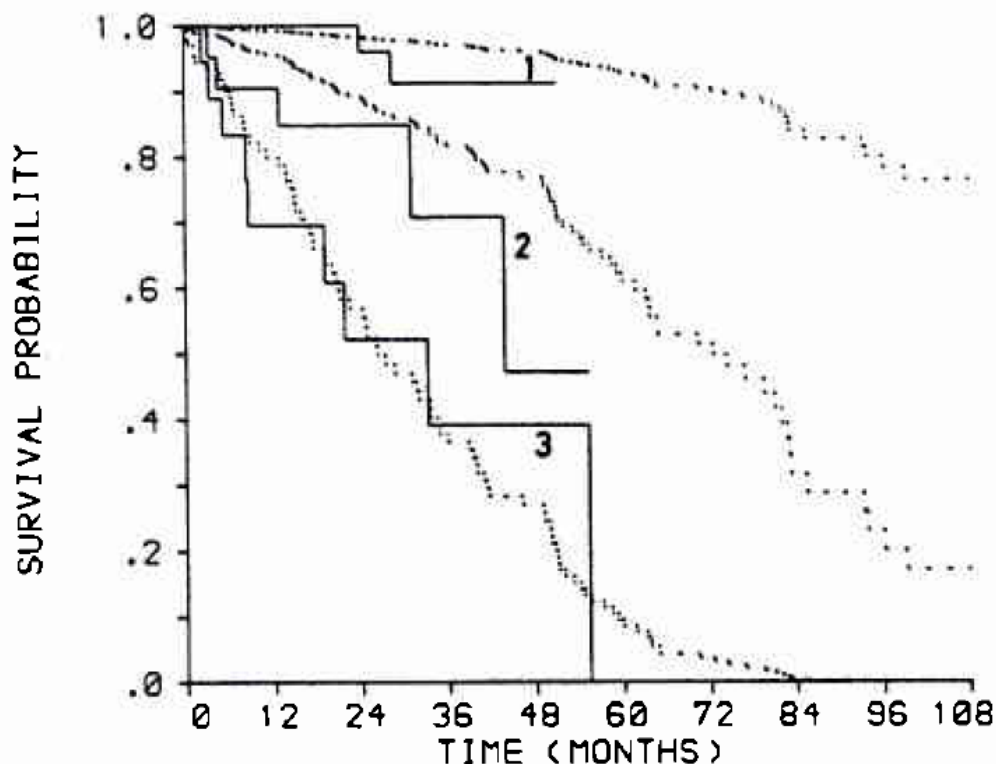


Figure 7. Observed (—) and estimated (-----) survival functions of three groups of placebo-treated patients from a controlled trial of D-penicillamine (Reference 4): group 1: $PI < 2.5$, $n = 44$, $O = 2$, $E = 0.9$; group 2: $2.5 < PI < 3.75$, $n = 22$, $O = 5$, $E = 2.7$; group 3: $PI > 3.75$, $n = 19$, $O = 9$, $E = 9.0$. O and E are the observed and expected numbers dying in each group ($\chi^2_3 = 3.2$, $p = 0.4$) (Reference 9).

Warstwowy model PH (1)

- ◆ W niektórych sytuacjach założenie PH może nie być spełnione dla wszystkich zmiennych niezależnych
- ◆ Można wówczas użyć modelu PH dla zmiennych, dla których założenie jest spełnione, przy odpowiedniej modyfikacji bazowych funkcji hazardu

Warstwowy model PH (2)

- ◆ Załóżmy PH dla zmiennej binarnej Y
- ◆ Przyjmijmy, że założenie nie jest spełnione dla X
- ◆ Możemy wówczas zdefiniować następujący model:

$$\lambda(t) = \lambda_1(t) \cdot e^{\beta \cdot Y} \quad \text{if } X=x_1$$

$$\lambda(t) = \lambda_2(t) \cdot e^{\beta \cdot Y} \quad \text{if } X=x_2$$

- ◆ Jest to *model warstwowy*.
 - bazowe funkcje hazardu są różne dla różnych *warstw*, zdefiniowanych wartościami zmiennej niezależnej X , ale
 - efekt Y jest nadal wyrażony jako proporcjonalna zmiana funkcji hazardu (dla ustalonej wartości X)

Interpretacja warstwowego modelu PH

$$\lambda(t) = \lambda_1(t) \cdot e^{\beta \cdot Y} \quad \text{jeśli } X=x_1$$

$$\lambda(t) = \lambda_2(t) \cdot e^{\beta \cdot Y} \quad \text{jeśli } X=x_2$$

- ◆ Gdy $X=x_1$, wzrost Y o 1 zmienia funkcję hazardu e^{β} razy. Podobnie dla $X=x_2$.

- ◆ Ale iloraz hazardów dla x_2 vs x_1 wynosi

$$\{\lambda_2(t) \cdot e^{\beta \cdot Y}\} / \{\lambda_1(t) \cdot e^{\beta \cdot Y}\} = \lambda_2(t) / \lambda_1(t)$$

- ◆ Wniosek: zależy od t !

Szacowanie parametrów warstwowego modelu PH

- ◆ Częściowa funkcja wiarygodności (s indeksuje warstwę):

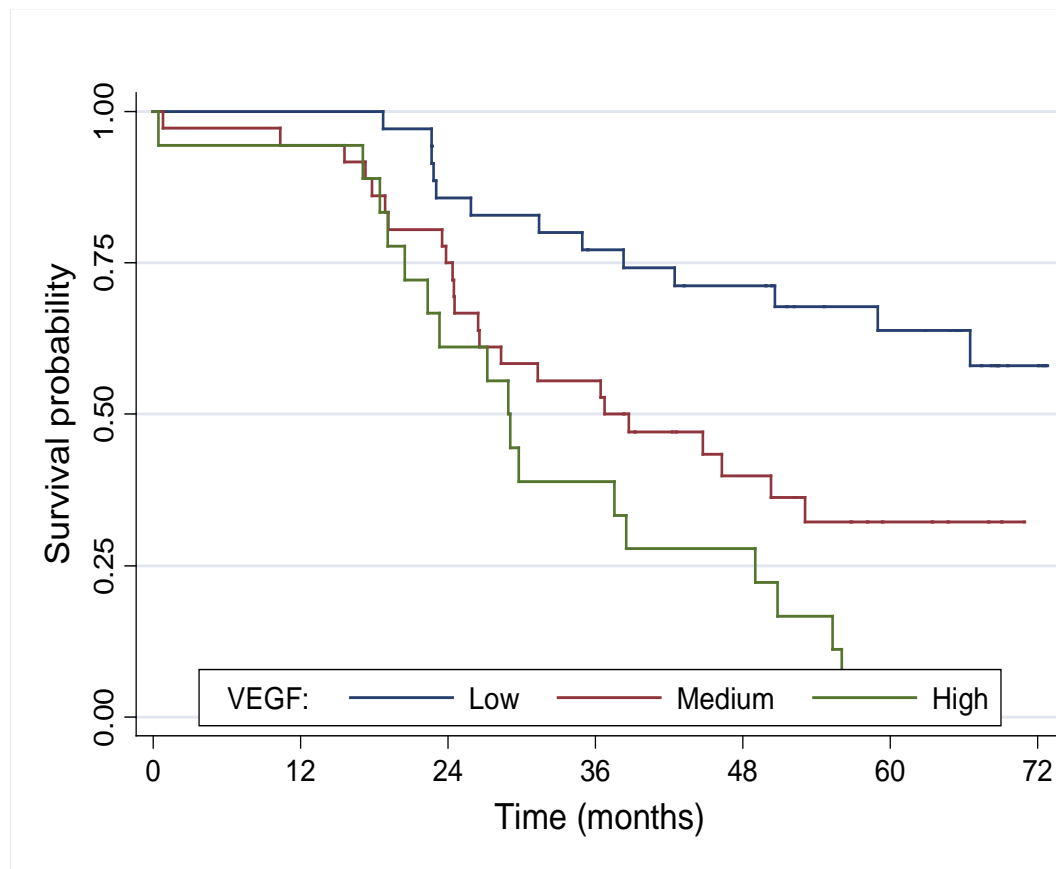
$$L(\beta) = \prod_{s=1}^S \prod_{k=1}^{K_s} \frac{\lambda_{0s}(t_{(k)}) h_k e^{\beta' Z_{(k)}(t_{(k)})}}{\sum_{l=1}^{n_s} Y_l(t_{(k)}) \lambda_{0s}(t_{(k)}) h_k e^{\beta' Z_l(t_{(k)})}} = \prod_{s=1}^S \prod_{k=1}^{K_s} \frac{e^{\beta' Z_{(k)}(t_{(k)})}}{\sum_{l=1}^{n_s} Y_l(t_{(k)}) e^{\beta' Z_l(t_{(k)})}}$$

gdzie K_s – liczba czasów zdarzeń w warstwie s , n_s – liczba obserwacji w warstwie s

- ◆ Oszacowanie β uzyskujemy rozwiązując $U(\hat{\beta}) = 0$

Warstwowy model PH: przykład (1)

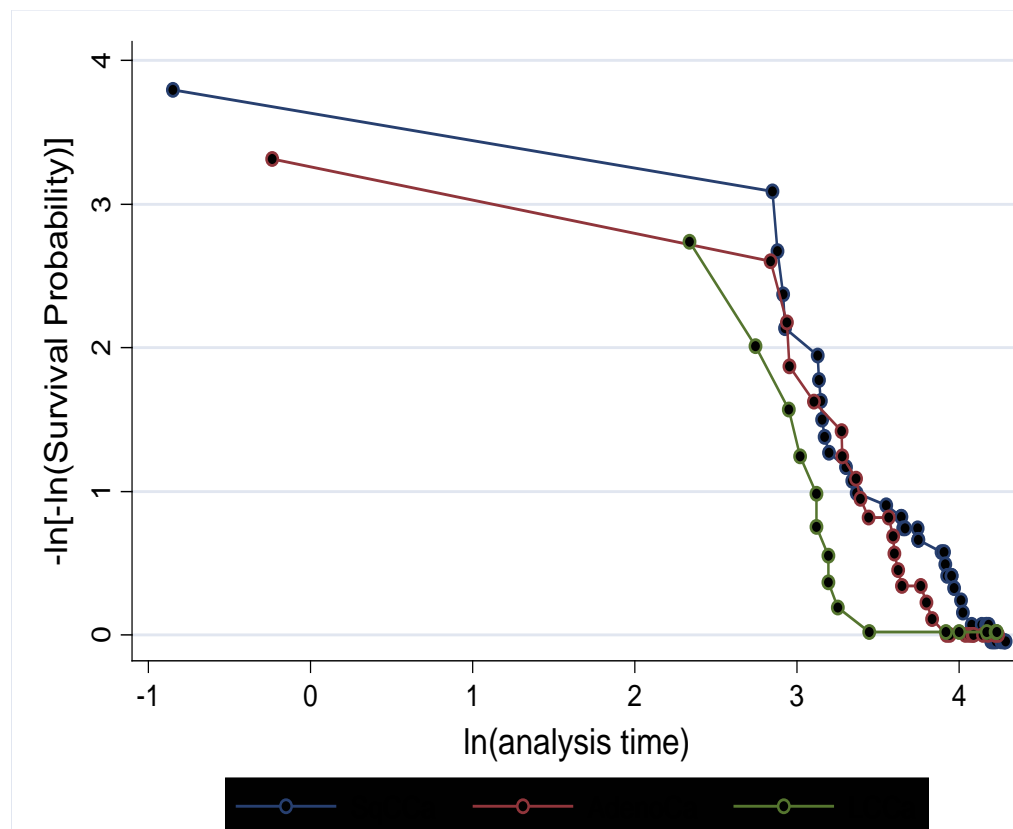
- ◆ Niklińska *et al.*, *Lung Cancer* (2001)
- ◆ Ekspresja czynnika wzrostu VEGF
- ◆ Wpływ na przeżycie 89 chorych leczonych operacyjnie
- ◆ Test logrank (dla trendu): $p < 0.001 \rightarrow$



Wastwowy model PH: przykład (2)

- ♦ Trzy typy histpat: SqCCa, AdenoCa, LCCa
- ♦ Graficzna ocena założenia proporcjonalnych hazardów:

wykresy powinny być równoległe →



(Nie-)Warstwowy model PH: przykład (3)

◆ Model ze wieloma zmiennymi niezależnymi:

- „vegf” = 1, 2, 3 dla niskiej, średniej, i wysokiej ekspresji VEGF
- „p53m” = 0 jeśli brak mutacji P53, 1 jeśli mutacja
- „sex” = 0 dla kobiet, 1 dla mężczyzn
- „tnm” = 1, 2, 3 dla TNM I, II and IIIA
- „adCC” = 0 dla SqCCa/LCCa, 1 dla AdenoCa
- „lcc” = 0 dla SqCCa/AdenoCa, 1 dla LCCa

_t	Haz. Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
vegf	1.428507	.3145729	1.619	0.105	.9277663	2.199512
p53m	2.767612	1.039047	2.712	0.007	1.325983	5.776602
sex	1.114986	.471841	0.257	0.797	.4864711	2.555535
tnm	3.208635	.8751451	4.274	0.000	1.88	5.476245
adCC	1.699829	.617637	1.460	0.144	.8339101	3.464906
lcc	3.330073	1.384905	2.893	0.004	1.473864	7.524021

Warstwowy model PH: przykład (4)

- ◆ Warstwowy model z warstwami zdefiniowanymi typem histopatologicznym:

_t	Haz. Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
vegf	1.548393	.3493506	1.938	0.053	.99502	2.409519
p53m	2.738736	1.061083	2.600	0.009	1.281636	5.852422
sex	1.00601	.4357343	0.014	0.989	.4304479	2.351171
tnm	3.095804	.8460452	4.135	0.000	1.811968	5.289278

Stratified by histpat

- ◆ Efekt typu histpat nie jest oszacowany
- ◆ Jest uwzględniony poprzez bazowe funkcje hazardu

Warstwowanie i liczebność próbki (1)

- ◆ Przyjmijmy $\beta=0$ i $V(\beta, t) \equiv V$ dla ważonej wariancji Z
- ◆ Oczekiwana wartość oszacowania wariancji wynosi $V(n-1)/n$
- ◆ Załóżmy brak jednoczesnych zdarzeń i cenzurowania dla $n=10$
- ◆ Oczekiwana wartość informacji wynosi
$$E\{I(\beta)\} = V(9/10 + 8/9 + \dots + 1/2 + 0/1)$$
 - I^{-1} daje wariancję oszacowania β
- ◆ Niech $n=200$. Wówczas
$$E\{I(\beta)\} = V(199/200 + 198/199 + \dots + 1/2 + 0/1) \approx 194.1 V$$
- ◆ Rozważmy 4 warstwy, każdą z 50 zdarzeniami. Wówczas
$$E\{I(\beta)\} = 4V(49/50 + 48/49 + \dots + 1/2 + 0/1) \approx 182 V$$
 - $(194.1 - 182) / 194.1 = 6\%$ strata efektywności

Warstwowanie i liczebność próbki (2)

- ◆ Dla $n=200$, $E\{I(\beta)\} \approx 194.1 \text{ V}$
- ◆ Dla 4 warstw z 50 zdarzeniami, $E\{I(\beta)\} \approx 182 \text{ V}$
 - $(194.1 - 182) / 194.1 = 6\%$ strata efektywności
- ◆ Dla 50 warstw z 4 zdarzeniami każda,
 $E\{I(\beta)\} = 50V(3/4 + 2/3 + 1/2 + 0/1) \approx 95.8 \text{ V}$
 - $(194.1 - 95.8) / 194.1 = 51\%$ strata efektywności

Warstwowanie i liczebność próbki (3)

- ♦ Załóżmy $d=200$ dla $n=1400$ i obserwacje cenzurowane > czas ostatniego zdarzenia
- ♦ Wówczas

$$E\{I(\beta)\} = V \sum_{i=1}^{200} \frac{1200 + i - 1}{1200 + i} \approx 199.85V$$

- ♦ Dla 50 równolicznych warstw z 4 zdarzeniami każda:

$$E\{I(\beta)\} = 50V \sum_{i=1}^4 \frac{24 + i - 1}{24 + i} = 50V \left(\frac{27}{28} + \frac{26}{27} + \frac{25}{26} + \frac{24}{25} \right) \approx 192.44V$$

- ♦ Strata efektywności: $(1 - 192.44/199.85) = 3.7\%$