Analiza przeżycia

Survival Analysis

2022

Analiza przeżycia

- Analiza takich zdarzeń jak zachorowanie, wyzdrowienie, śmierć, ciąża, …
- Ważne jest nie tylko wystąpienie zdarzenia, ale również czas do momentu wystąpienia zdarzenia
- Przykład: Na początku wszystkie osobniki są chore (stan początkowy) i zagrożone śmiercią (zdarzenie).

Analiza przeżycia

- Wszystkie osobniki mają ten sam stan początkowy

- Przykłady:
  - Wszystkie osobniki są chore (stan) i zagrożone śmiercią (zdarzenie).
  - Wszystkie samice są ciężarne (stan) i z czasem może wystąpić poród (zdarzenie)

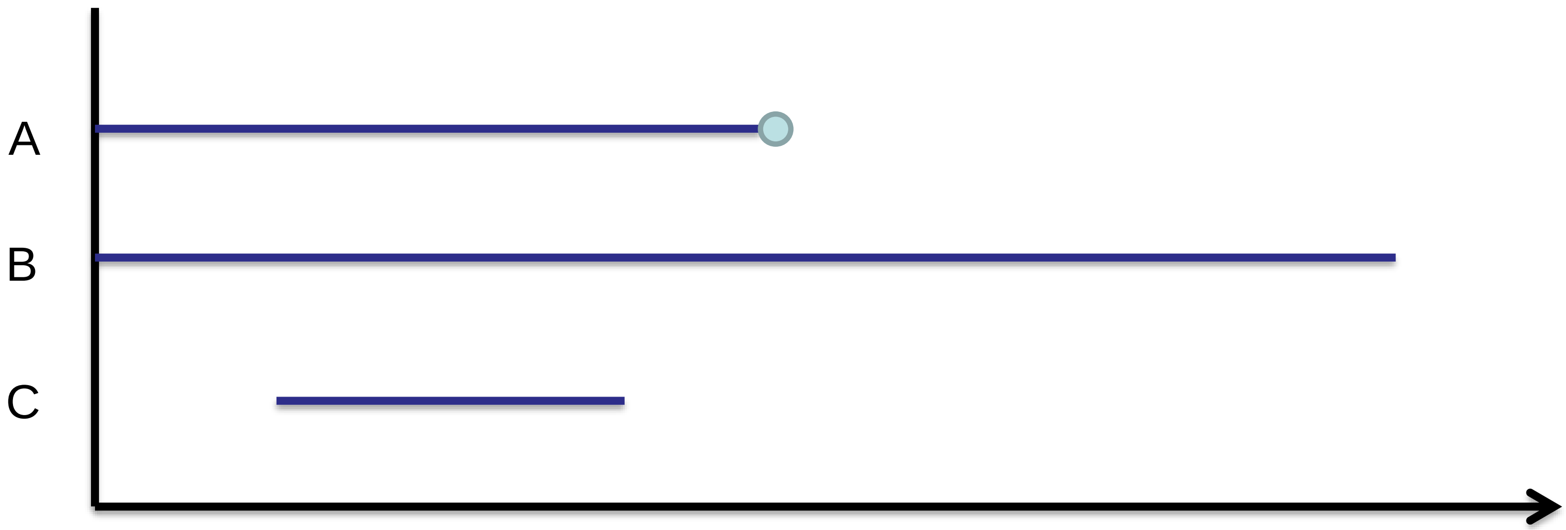
Analiza przeżycia

- AP uwzględnia, że dane są recenzowane
- Zwierzęta znikają w czasie obserwacji
- Tylko niektóre znikają z przyczyn, którymi się faktycznie interesujemy (np. zdarzenie-śmierć)

Dane cenzorowane (ucięte)

- Czas jest znany tylko dla niektórych osobników
- Dla innych czas jest nieznany:
  - Zdarzenie nie zaszło podczas trwania obserwacji
  - Zwierzę padło z innych przyczyn
  - Farmer wycofał zgodę, itp

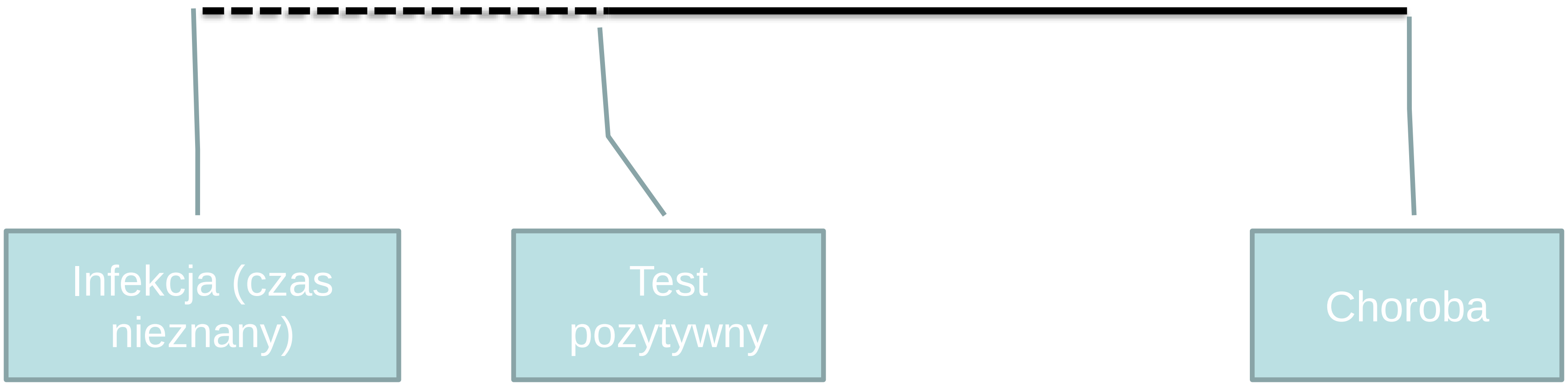
Dane cenzorowane (ucięte)



Osobnik	Czas przeżycia	Zdarzenie
A	5	1 (śmierć)
B	10	0 (ucięte)
C	3	0 (ucięte)

Dane ucięte także z lewej strony

Rzadziej spotykane



Cele analizy przeżycia

- **Porównanie funkcji przeżycia/hazardu między grupami**
- Opis przeżycia znanym rozkładem teoretycznym (np. Weibull)
- Opis relacji między przeżyciem a potencjalnymi zmiennymi objaśniającymi

Funkcja przeżycia

Jakie jest prawdopodobieństwo, że osobnik przeżyje co najmniej 5 jednostek czasu (np. 5 dni)? $P(T > t=4)$

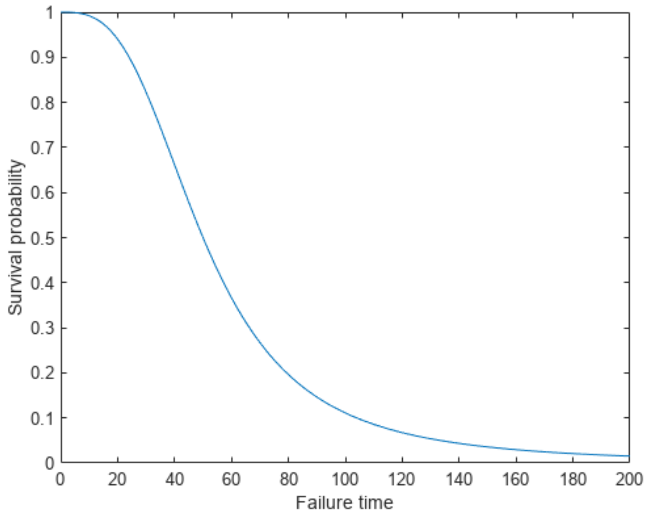
T = czas przeżycia, t = specyficzna wartość T

Ogólnie: $S(t) = P(T > t)$

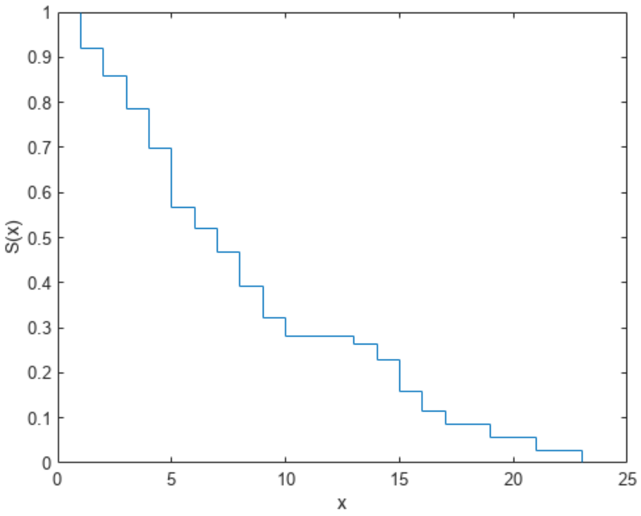
Inaczej mówiąc, S to prawdopodobieństwo, że u osobnika zdarzenie nie wystąpi do czasu t

Funkcja przeżycia

Teoretyczny przebieg



Praktyczny przykład



S(t=0) = 1

S nigdy nie rośnie

Funkcja hazardu

h(t) – warunkowe prawdopodobieństwo zdarzenia (np. śmierci) w następnej jednostce czasu przypadające na jednostkę czasu

$$h(t) = \lim_{\Delta t \to 0} P(\; t \le T < (t+\Delta t) \;\; | \;\; T \ge t \;) \; / \; \Delta t$$

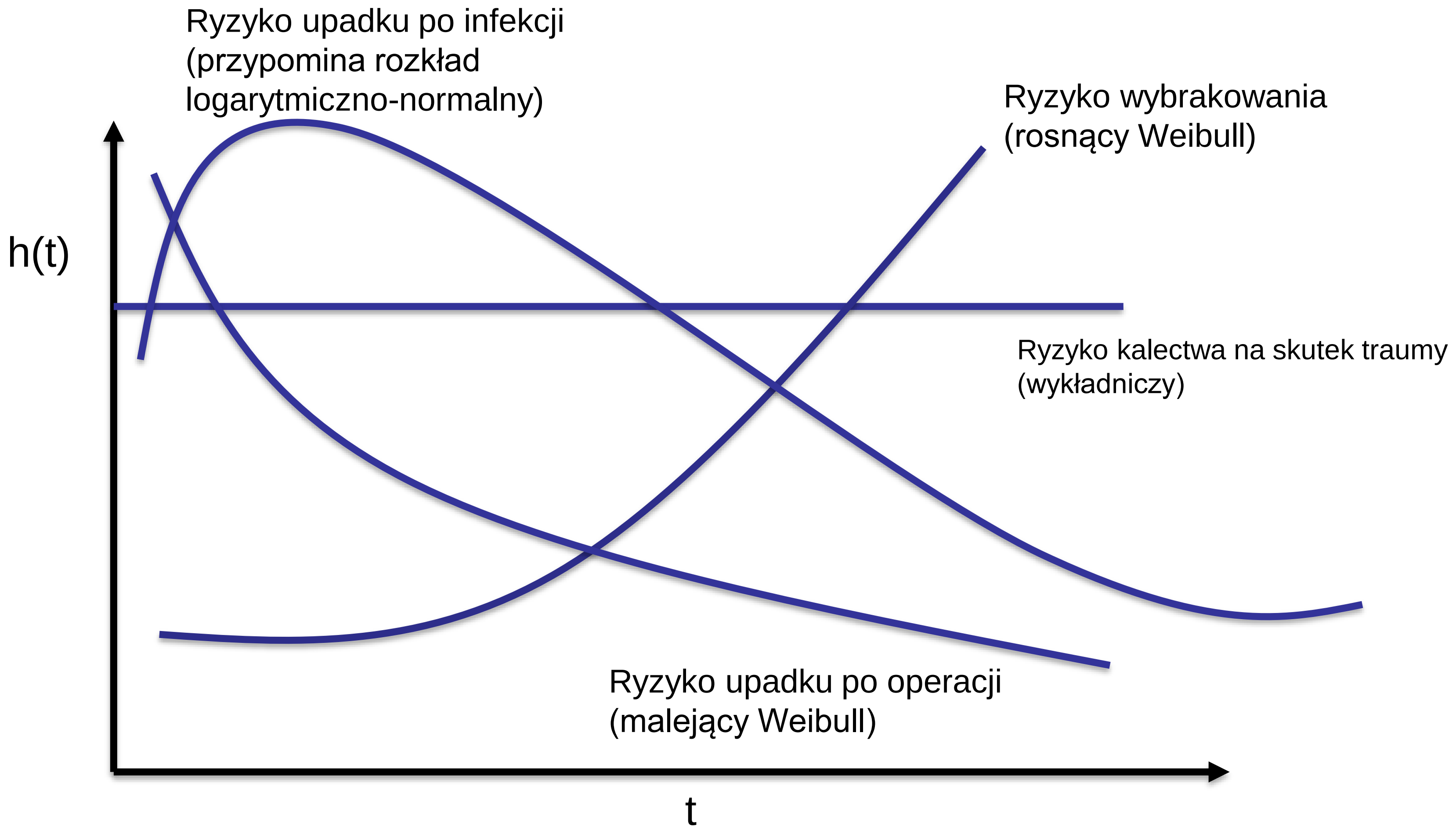
Myślimy tu raczej o gęstości prawdpopodobieństwa (h może być większe od 1).

Przeżycie a hazard

$h(t) = [\,-1 / S(t)\,] \times$ pierwsza pochodna $S(t)$

Komputer łatwo przelicza między funkcjami

Typowe przebiegi funkcji hazardu



Estymator Kaplana-Meiera

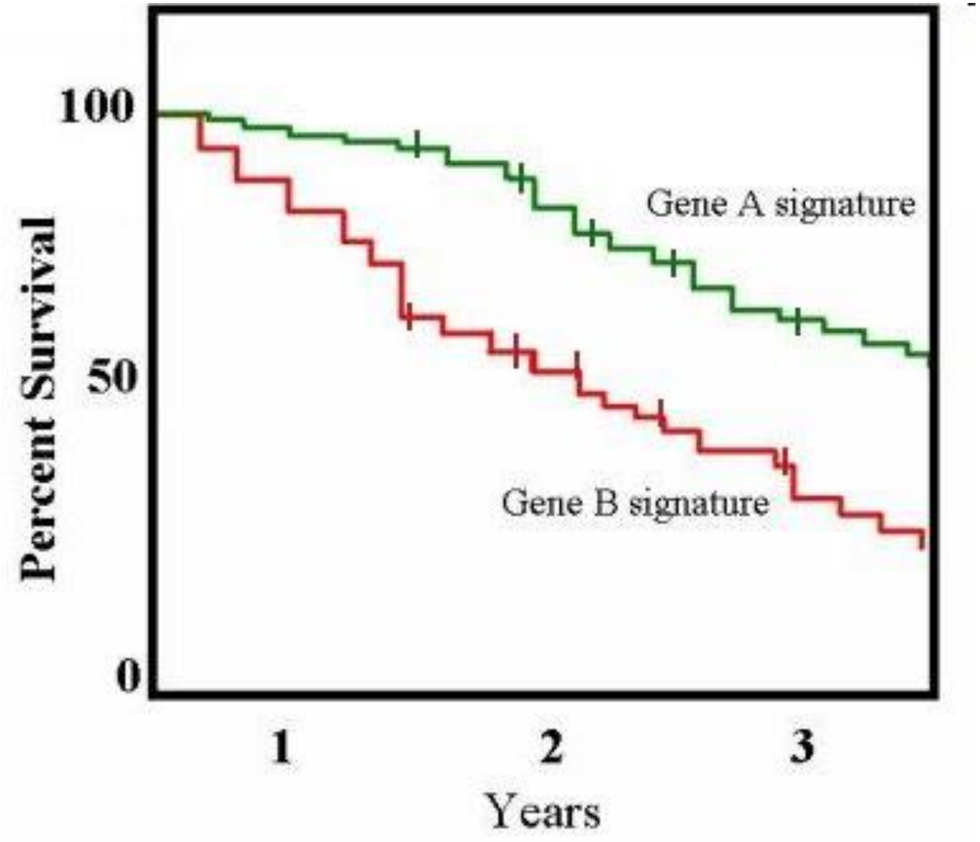
•Uwzględnia cenzorowane dane
•Najpopularniejszy

$$\hat{S}(t) = \prod_{t_i \leq t} \frac{n_i - d_i}{n_i}$$

S(0)=1

Przedział (dzień)		Liczba zwierząt na początku (n)	Śmierć (d)	Utrata z innych powodów	S(t)
12	13	116	0	1	1,0
14	15	115	1	0	0,9913
55	56	85	1		0,7435
60	61	84	0	84	0,7435

Estymator Kaplana-Meiera



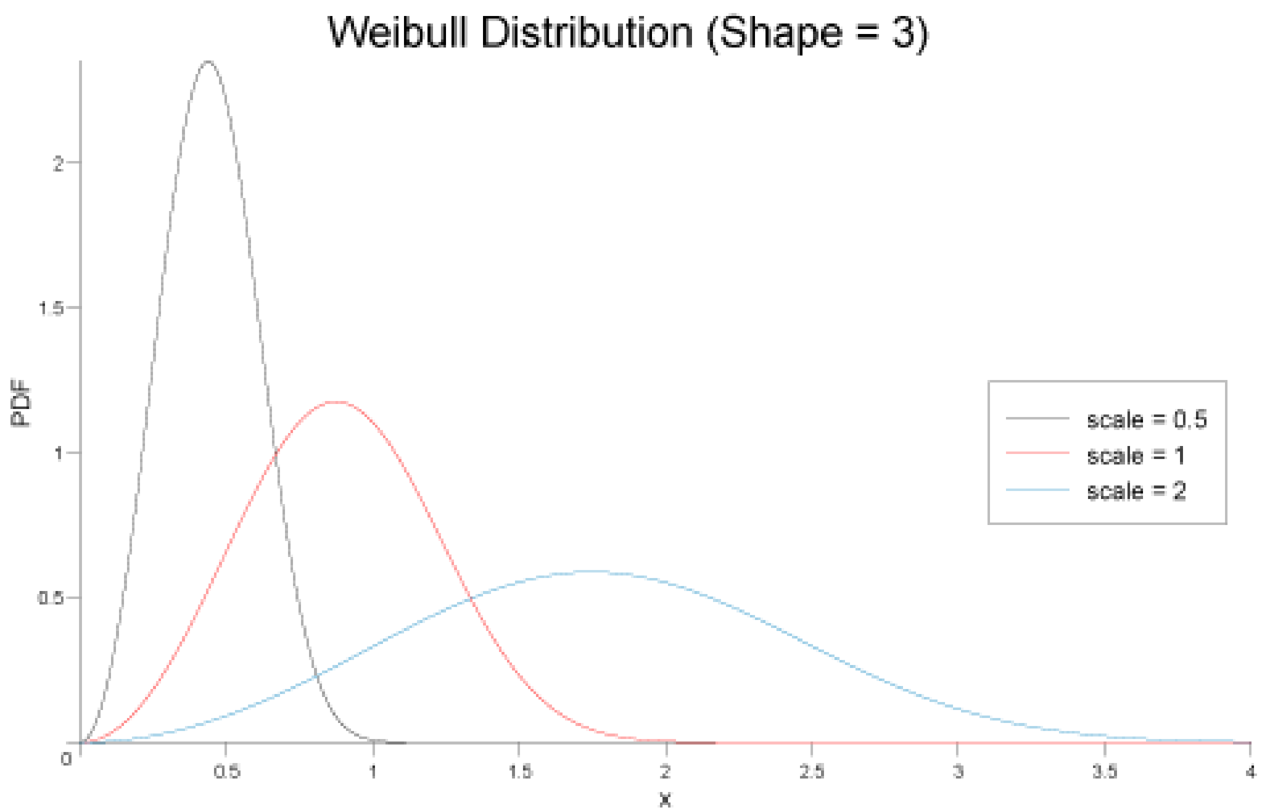
Cele analizy przeżycia

- Porównanie funkcji przeżycia/hazardu między grupami
- **Opis przeżycia znanym rozkładem teoretycznym (np. Weibulla)**
- Opis relacji między przeżyciem a potencjalnymi zmiennymi objaśniającymi

Rozkład Weibulla

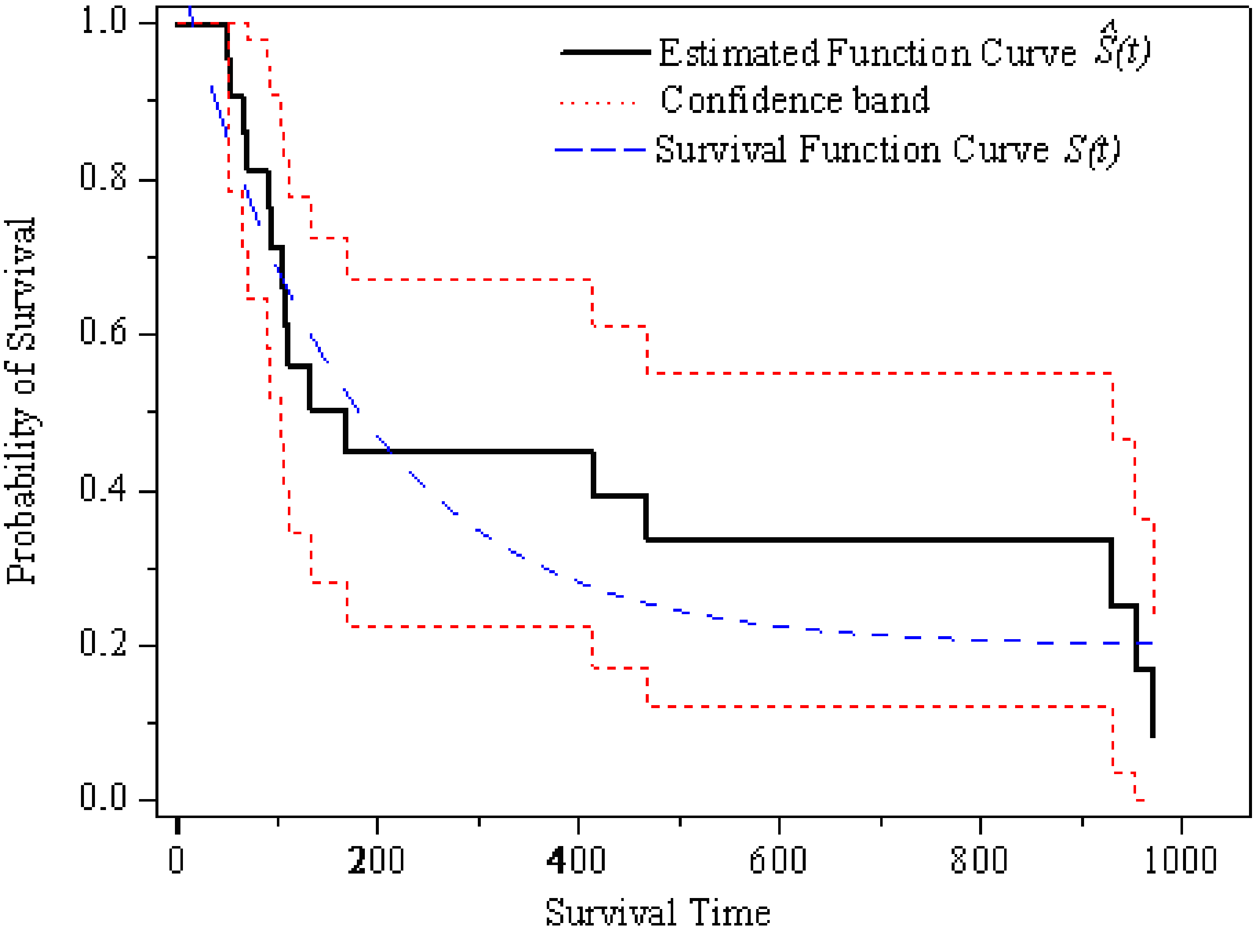
- $S(t) = \exp(- (a \times t)^{p})$
- $h(t) = a \times p \times (a \times t)^{p-1}$

- a = parametr skali
- p = parametr kształtu
- a i p muszą być oszacowane z danych
- Może przypominać rozkład wykładniczy lub normalny


Weibull Distribution (Shape = 3)
PDF
x
0
0.5
1
1.5
2
2.5
3
3.5
4
scale = 0.5
scale = 1
scale = 2

-

Kaplan-Meier oraz Weibull



Cele analizy przeżycia

- Porównanie funkcji przeżycia/hazardu między grupami
- Opis przeżycia znanym rozkładem teoretycznym (np. Weibull)
- **Opis relacji między przeżyciem a potencjalnymi zmiennymi objaśniającymi**

Modelowanie funkcji hazardu rozkładem wykładniczym

$$h(t, x) = h_0(t) \times \exp(\beta_1 x_1 + \ldots + \beta_k x_k)$$

h_0 to zerowa linia hazardu (gdy każdy x=0)

- Model regresji wykładniczej: $h_0 = 1$ (mniej elastyczny)
- Regresja Weibulla: $h_0(t) = a \times p \times (a \times t)^{p-1}$
- Model proporcjonalnych hazardów Coksa: nie ma potrzeby definicji h_0 (bardzo praktyczne)

Model proporcjonalnych hazardów Coksa

$$\frac{\dfrac{h(t, x=1)}{h_0(t)}}{\dfrac{h(t, x=0)}{h_0(t)}} = e^{\beta_1 x_1}$$

- Podobne do regresji logistycznej

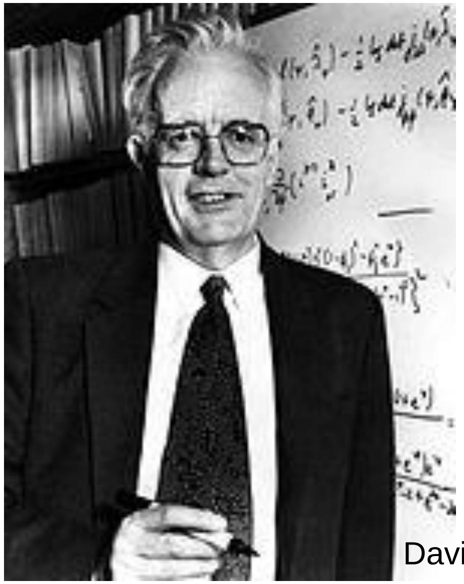
- Nie trzeba estymować parametrów funkcji zerowej h_0

- Ale warunek: HR musi być stałe w czasie, sprawdź rysując dla każdej grupy ln[-ln(S(t))]) względem ln(t), przebiegi powinny być równoległe

Model proporcjonalnych hazardów Coksa

Hazard ratio

$$= HR = h(\, t, \; x{=}1 \,) \,/\, h(\, t, \; x{=}0 \,) = \exp(\, \beta_1 x_1 \,)$$



David Cox, ur. 1924

HR=1: czynnik nie ma znaczenia

HR>1: czynnik pozytywnie powiązany ze zdarzeniem

HR<1: czynnik negatywnie skorelowany

Przykład danych uciętych z prawej strony

Porównywane grupy

Czas zdarzenia lub cenzury (np. dni)

Zdarzenie tak=1, nie=0

plec	czas	smierc
1	131	0
1	150	1
1	231	0
1	240	0
1	400	0
2	1	1
2	3	1
2	4	0
2	5	1

Dane ucięte także z lewej strony

Czas pozytywnego testu

Czas zdarzenia lub cenzury (np. dni)

plec	czas1	czas2	smierc
2	51	52	1
2	58	59	1
2	55	57	1
2	28	50	1
1	21	51	0
1	19	28	1
2	25	32	1

Analiza przeżycia w R

Zainstaluj i przygotuj do działania pakiet **survival**

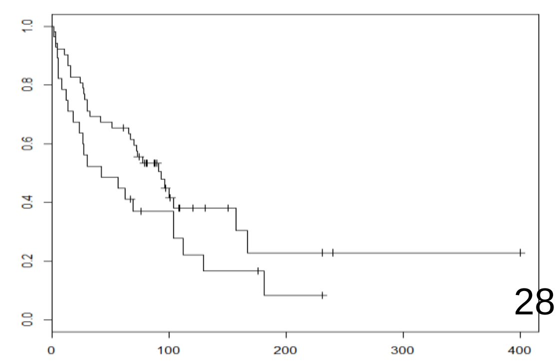
```
install.packages( ’survival’)
library( survival )
```

```
read.table( ‚dane.txt’, header=TRUE ) -> d
attach( d )

survfit( Surv( czas, smierc ) ~ plec )    ->    fit
plot( fit )

summary( fit )
summary( fit )$surv -> s
summary( fit )$time -> t
plot( log(t), log( -log(s) )  )


coxph( formula = Surv( czas, smierc ) ~ plec ) -> wynik
summary( wynik )
```



Zadanie 1

Badano skuteczność szczepionki porównując grupę kontrolną (grupa=0) z zaszczepioną (grupa=1). Dane o poszczególnych osobnikach zebrano w pliku „szczepionka.txt”.

(a) Narysuj wykres Kaplana-Meiera prezentujący prawdopodobieństwa przeżycia w czasie dla obu grup.

(b) Zbadaj wpływ szczepienia modelem proporcjonalnych hazardów Coksa