STATYSTYCZNE MODELOWANIE DANYCH BIOLOGICZNYCH

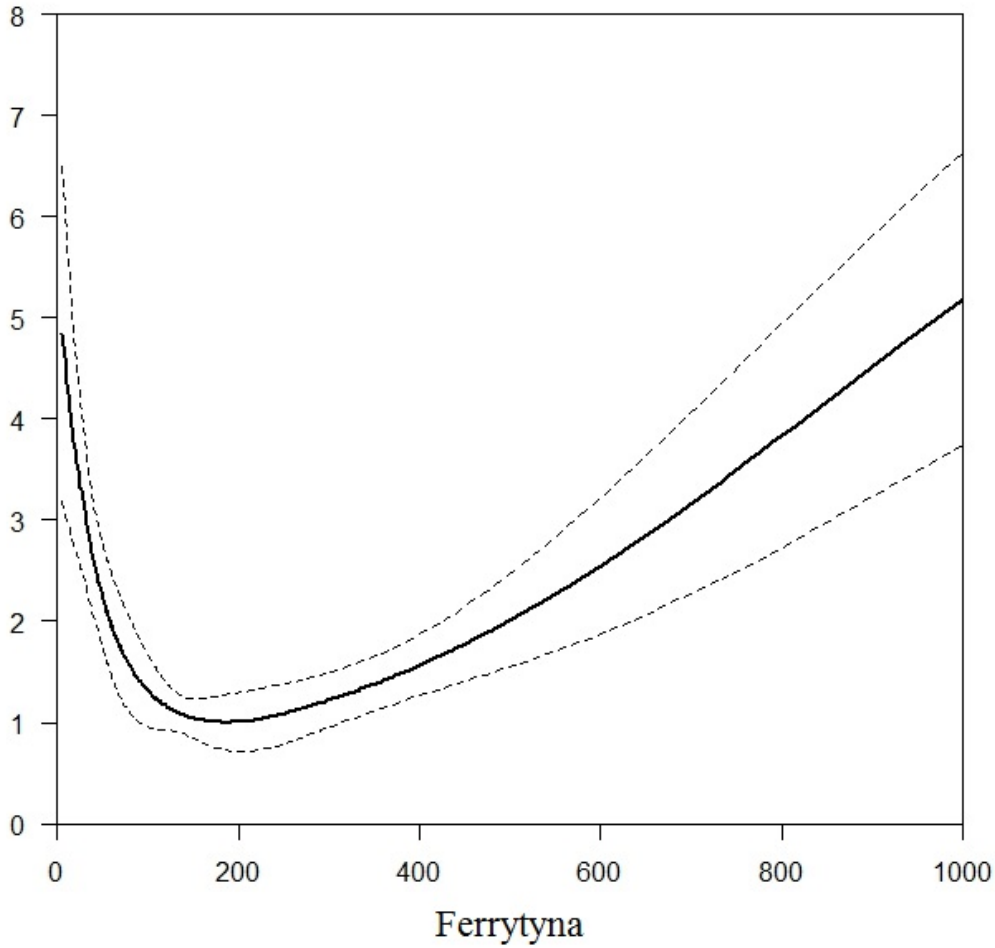
WYKŁAD 1

13 październik 2014

Plan wykładu

1. Analizy przeżycia na przykładach
2. Podstawowe idee statystyki matematycznej wykorzystywane w analizie przeżycia
3. Krzywe przeżycia
4. Pakiet R

ANALIZA PRZEŻYCIA - PRZYKŁADY

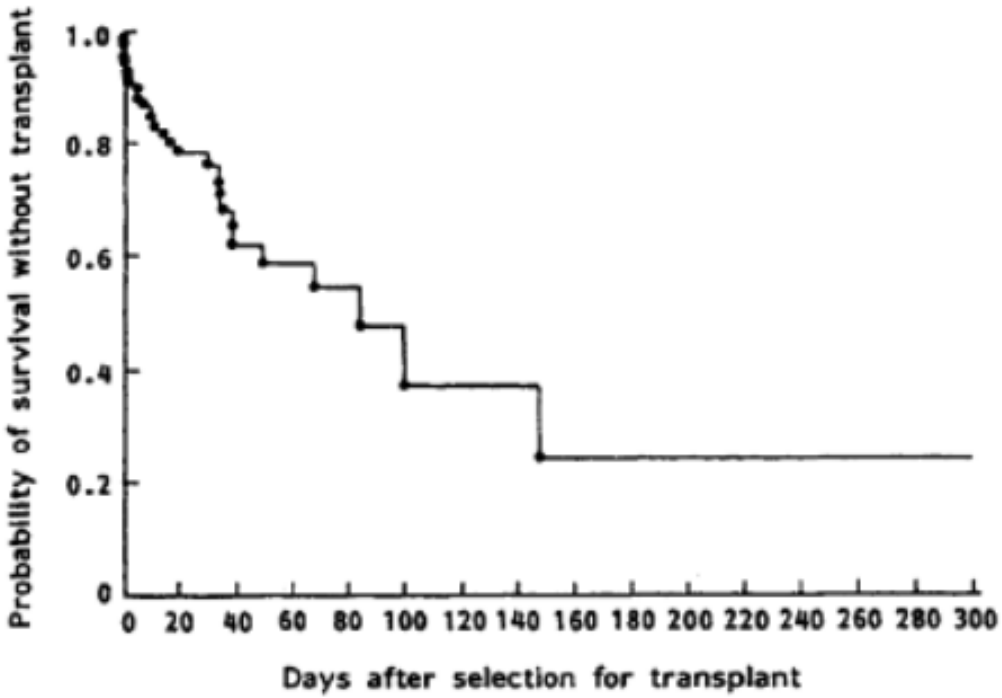


Przykład 1 - transplantacja serca

Patients not receiving a transplant		Patients receiving a transplant					
X	State*	Y	Z	State	Y	Z	State
49	*D*	0	15	*D*	82	663	*A*
5	*D*	35	3	*D*	31	253	*D*
17	*D*	50	624	*D*	40	147	*D*
2	*D*	11	46	*D*	9	51	*D*
39	*D*	25	127	*D*	66	479	*A*
84	*D*	16	61	*D*	20	322	*D*
7	*D*	36	1350	*D*	77	442	*A*
0	*D*	27	312	*D*	2	65	*D*
35	*D*	19	24	*D*	26	419	*A*
36	*D*	17	10	*D*	32	362	*A*
1400	*A*	7	1024	*D*	13	64	*D*
5	*D*	11	39	*D*	56	228	*D*
34	*D*	2	730	*D*	2	65	*D*
15	*D*	82	136	*D*	9	264	*A*
11	*D*	24	1379	*A*	4	25	*D*
2	*D*	70	1	*D*	30	193	*A*
1	*D*	15	836	*D*	3	196	*A*
39	*D*	16	60	*D*	26	63	*D*
8	*D*	50	1140	*A*	4	12	*D*
101	*D*	22	1153	*A*	45	103	*A*
2	*D*	45	54	*D*	25	60	*A*
148	*D*	18	47	*D*	5	43	*A*
1	*D*	4	0	*D*			
68	*D*	1	43	*D*			
31	*D*	40	971	*A*			
1	*D*	57	868	*A*			
20	*D*	0	44	*D*			
118	*A*	1	780	*A*			
91	*A*	20	51	*D*			
427	*A*	35	710	*A*			

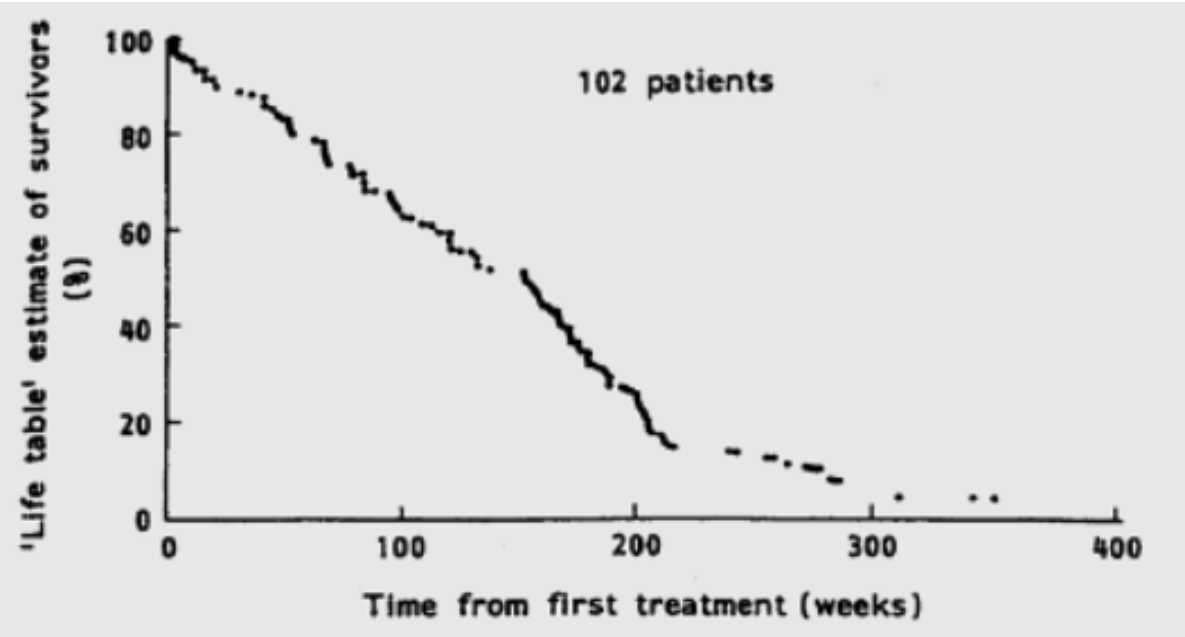
Rysunek 1: Czas do transplantacji serca oraz czas przeżycia dla 82 pacjentów wybranych z Programu Transplantacji Serca Stanford (Turnbull *et al.*, 1974).

Przykład 1 - transplantacja serca



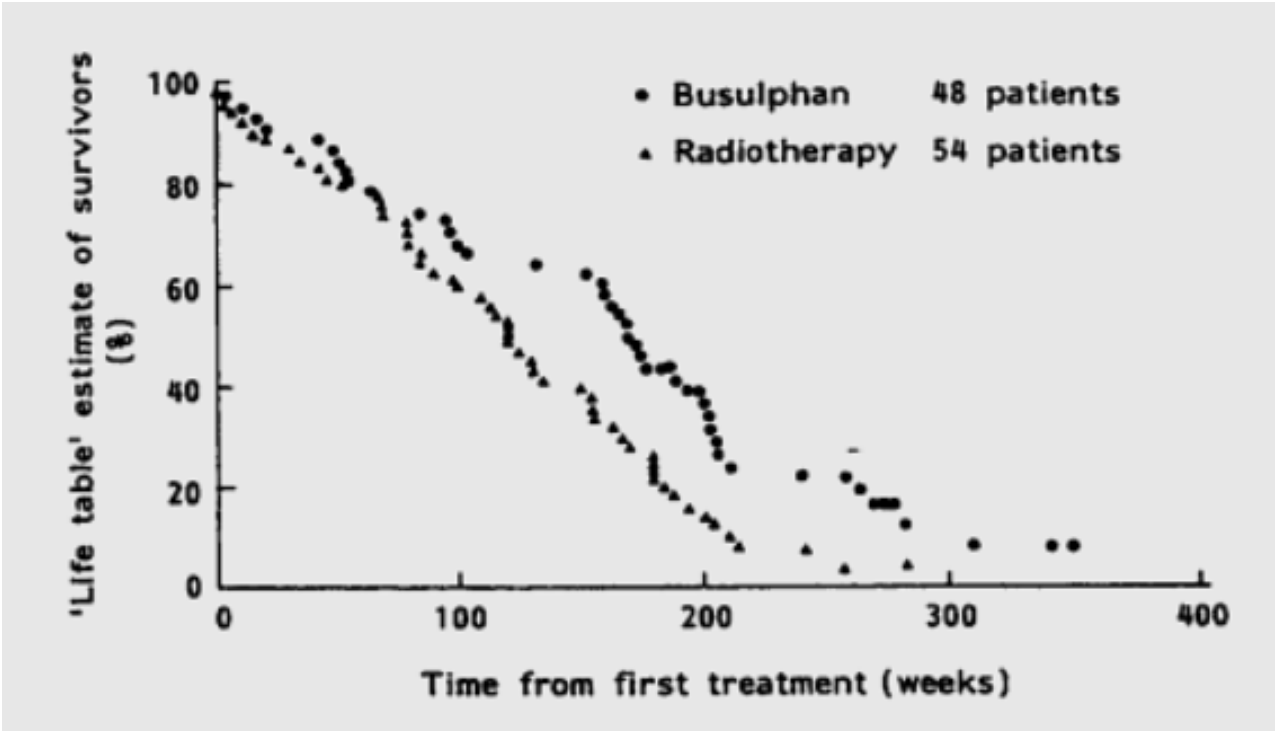
Rysunek 2: Prawdopodobieństwo przeżycia osób zakwalifikowanych do przeszczepu serca estymowane przy pomocy krzywych Kaplana-Meiera (Turnbull *et al.*, 1974).

Przykład 2 - przewlekła białaczka szpikowa



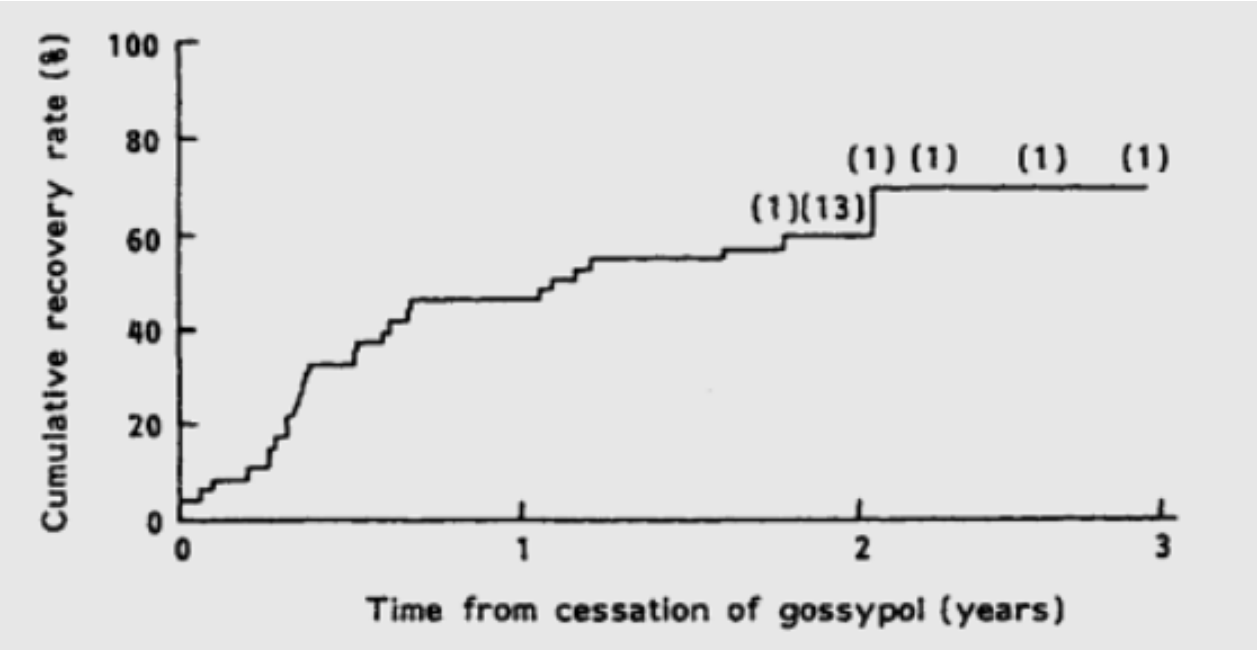
Rysunek 3: Krzywa przeżycia dla pacjentów z przewlekłą białaczką szpikową (Peto *et al.*, 1977).

Przykład 2 - przewlekła białaczka szpikowa



Rysunek 4: Krzywe przeżycia dla pacjentów z przewlekłą białaczką szpikową poddawanych chemioterapii (Busulphan) lub radioterapii (Peto *et al.*, 1977).

Przykład 3 - antykoncepcja



Rysunek 5: Skumulowana stopa "regeneracji" spermy po zaprzestaniu zażywania gossypolu (Meng *et al.*, 1988).

PODSTAWOWE IDEE STATYSTYKI MATEMATYCZNEJ WYKORZYSTYWANE W ANALIZIE PRZEŻYCIA

Mediana czasu przeżycia

Mediana czasu przeżycia - wartość dla której 50% osób ma dłuższy czas przeżycia i 50% osób ma krótszy czas przeżycia.

Dlaczego nie używamy średniego czasu przeżycia?

Zazwyczaj rozkład czasu przeżycia jest skośny (mała ilość osób chorych żyje długo).

Mediana czasu przeżycia

Mediana czasu przeżycia - wartość dla której 50% osób ma dłuższy czas przeżycia i 50% osób ma krótszy czas przeżycia.

Dlaczego nie używamy średniego czasu przeżycia?

Zazwyczaj rozkład czasu przeżycia jest skośny (mała ilość osób chorych żyje długo).

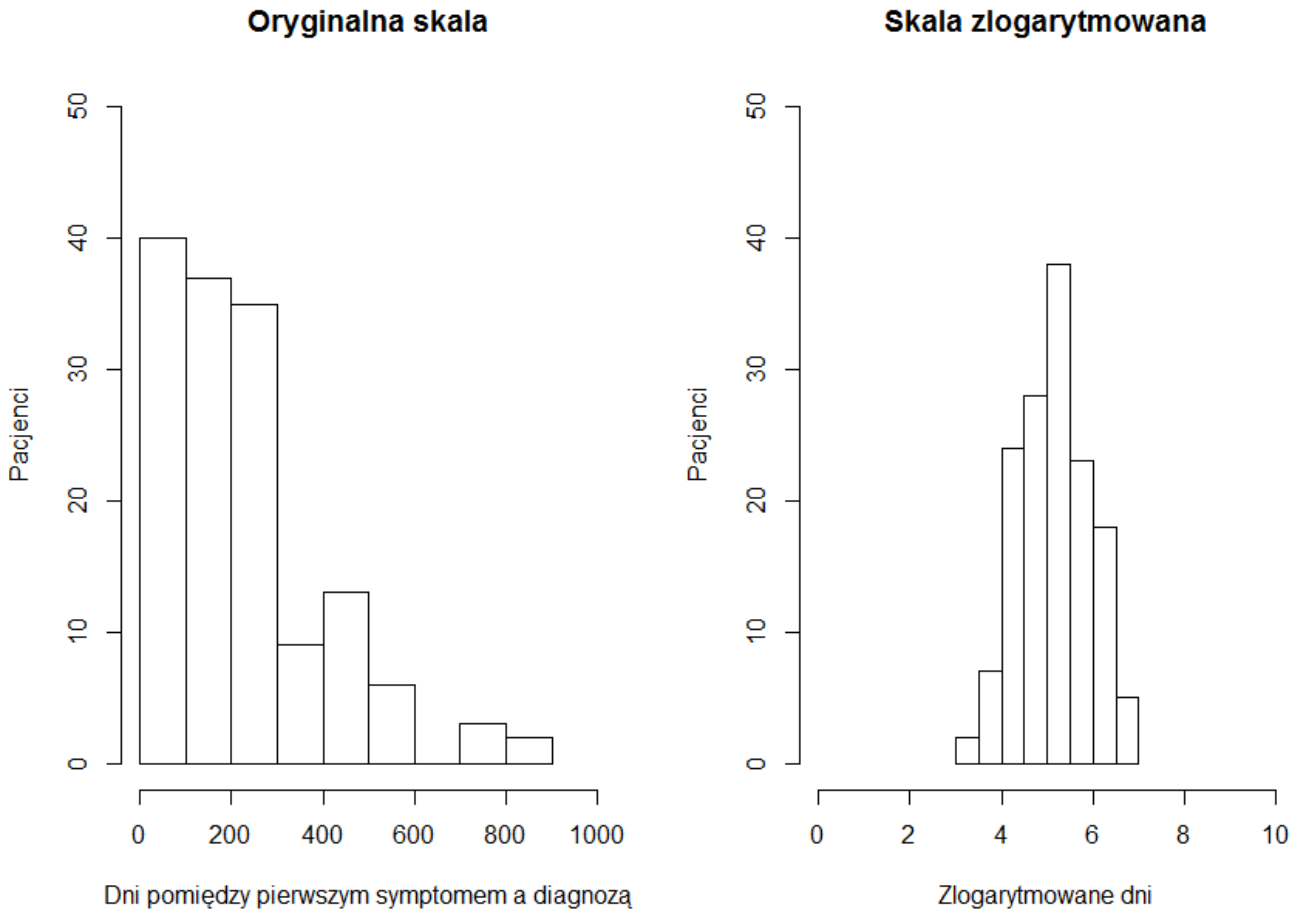
Mediana czasu przeżycia

Mediana czasu przeżycia - wartość dla której 50% osób ma dłuższy czas przeżycia i 50% osób ma krótszy czas przeżycia.

Dlaczego nie używamy średniego czasu przeżycia?

Zazwyczaj rozkład czasu przeżycia jest skośny (mała ilość osób chorych żyje długo).

Mediana czasu przeżycia

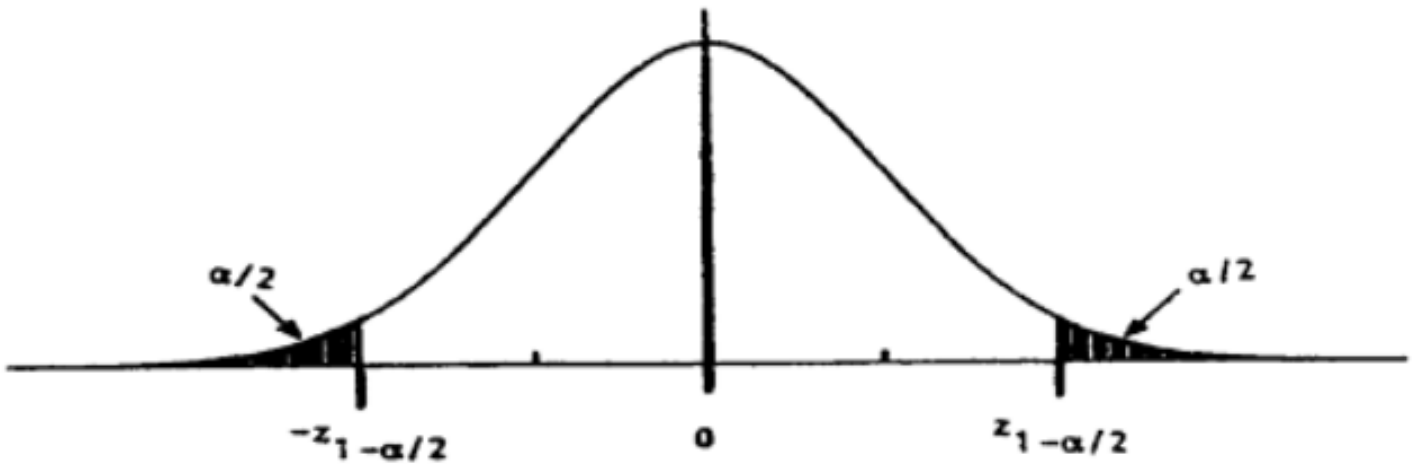


Rysunek 6: Rozkład czasu pomiędzy pierwszym symptomem a diagnozą raka piersi dla 145 kobiet. Średnia wynosi 225.1, a mediana 185 dla skali oryginalnej i odpowiednio 5.220 i 5.151 dla skali zlogarytmowanej.

Przedziały ufności

W analizie przeżycia ważną rolę odgrywają przedziały ufności:

$$\overline{X} - z_{1-\frac{\alpha}{2}} \cdot \frac{S}{\sqrt{n}} < \mu < \overline{X} + z_{1-\frac{\alpha}{2}} \cdot \frac{S}{\sqrt{n}}$$



Rysunek 7: Dolny i górny $\frac{\alpha}{2}$ punkt standardowego rozkładu normalnego

Przedziały ufności

Statystyka	Parametr	Estymator	Błąd standardowy
średnia	μ	$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} x_i$	$\frac{S}{\sqrt{n}}$
różnica średnich	$\delta = \mu_A - \mu_B$	$d = \overline{X}_A - \overline{X}_B$	$\sqrt{\frac{S_A^2}{n_A} + \frac{S_B^2}{n_B}}$
proporcja	π	$p = \frac{r}{n}$	$\sqrt{\frac{p(1-p)}{n}}$
różnica proporcji	$\delta = \pi_A - \pi_B$	$d = p_A - p_B$	$\sqrt{\frac{p_A(1-p_A)}{n_A} + \frac{p_B(1-p_B)}{n_B}}$

Tablica 1: Statystyki i ich błędy standardowe.

Przedziały ufności - przykład

Lekarstwo	nie wyleczony	wyleczony	łącznie	procent niewyleczonych
pirenzepina	7 (a)	23 (c)	30 (a+c)	23.33
trithiozina	13 (b)	18 (d)	31 (b+d)	41.94
łącznie	20 (r)	41 (s)	61 (n)	

Tablica 2: Pacjenci chorujący na wrzody żołądka z podziałem na grupy leczone pirenzepiną (P) i trithioziną (T) (Familiari *et al.*, 1981).

Mamy $n_P = 30$, $p_P = 0.2333$, $n_T = 31$ i $p_T = 0.4194$. Obserwowana różnica pomiędzy sposobami leczenia wynosi $d = p_T - p_P = 0.1861$. Błąd standardowy dla estymatora d wynosi:

$$SE = \sqrt{\frac{p_T(1-p_T)}{n_T} + \frac{p_P(1-p_P)}{n_P}} = 0.1175.$$

Przedział ufności dla parametru δ wynosi:

$$-4\% < \delta < 42\%.$$

Iloraz ryzyka (Hazard ratio)

Załóżmy, że O_A oznacza obserwowaną ilość zgonów w grupie A oraz E_A oznacza oczekiwaną ilość zgonów w tej grupie. Wtedy $\frac{O_A}{E_A}$ oznacza relatywną śmiertelnność w grupie A. Podobnie $\frac{O_B}{E_B}$ oznacza relatywną śmiertelnność w grupie B. Iloraz ryzyka definiujemy jako:

$$HR = \frac{\frac{O_A}{E_A}}{\frac{O_B}{E_B}} = \frac{O_A E_B}{E_A O_B}.$$

Interpretacja:

- $HR = 1$ - nie ma różnic pomiędzy grupami A i B,
- $HR > 1$ - ryzyko w grupie A jest większe niż w grupie B,
- $HR < 1$ - ryzyko w grupie B jest większe niż w grupie A.

Przykład: Jeśli $HR = 3$ to znaczy, że ryzyko śmierci w grupie A jest trzykrotnie wyższe niż w grupie B. Jeśli $HR = 0.5$ to znaczy, że ryzyko śmierci w grupie A jest dwukrotnie niższe niż w grupie B.

Ryzyko względne (Relative risk)

Azbest	zdiagnozowany międzybłoniak	niezdiagnozowany międzybłoniak	łącznie
Typ I	a	c	a+c
Typ II	b	d	b+d
łącznie	r	s	n

Tablica 3: Zestawienie robotników narażonych na dwa rodzaje azbestu oraz informacja, czy zachorowali na międzybłoniaka czy nie.

Na podstawie powyższej tabeli ryzyko względne definiujemy jako:

$$RR = \frac{\frac{a}{a+c}}{\frac{b}{b+d}} = \frac{a(b+d)}{(a+c)b},$$

czyli ryzyko wystąpienia międzybłoniaka u robotników narażonych na azbest typu I dzielimy przez ryzyko wystąpienia międzybłoniaka u robotników narażonych na azbest typu II. Interpretacja RR jest analogiczna jak w przypadku HR.

Ryzyko względne - Przykład

Na podstawie danych z Tablicy (2) mamy:

$$RR = \frac{7 \cdot 31}{13 \cdot 30} = 0.5564.$$

Oznacza to, że pacjenci leczeni trithioziną mają około dwa razy większe ryzyko, iż nie zostaną wyleczeni z wrzodów niż ci, którzy byli leczeni pirenzepiną.

Iloraz szans (Odds ratio)

Używając danych z powyższej tabeli (Tablica 3) iloraz szans jest definiowany jako:

$$OR = \frac{\frac{a}{c}}{\frac{b}{d}} = \frac{ad}{bc}$$

Na podstawie danych z Tablicy (2) mamy:

$$OR = \frac{7 \cdot 18}{13 \cdot 23} = 0.42.$$

Oznacza to, że pacjenci leczeni trithioziną mają około 2.4 razy większe szanse na niewyleczenie, niż pacjenci, którzy byli leczeni pirenzepiną.

Zależność między RR i OR

Zakładając, że w Tablicy (3) zdarzenie pojawia się z bardzo małym prawdopodobieństwem tzn. z dużej liczby robotników tylko kilku zachoruje na międzybłoniaka. Wtedy

$$\frac{a}{(a+c)} \approx \frac{a}{c} \quad \text{i} \quad \frac{b}{(b+d)} \approx \frac{b}{d}.$$

Na podstawie powyższych przybliżeń mamy zależność:

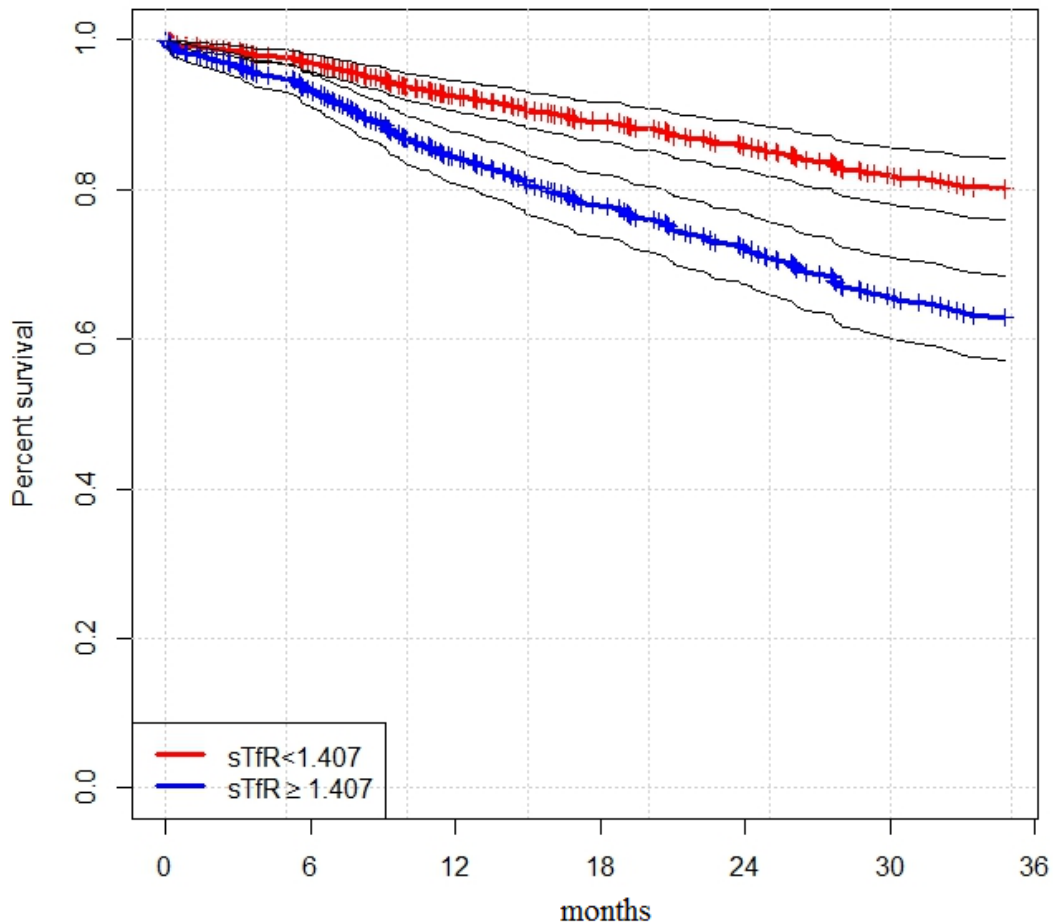
$$OR \approx RR.$$

Testy istotności

Należy przypomnieć sobie następujące testy:

- z test
- χ^2 test
- test oparty na ilorazie wiarygodności

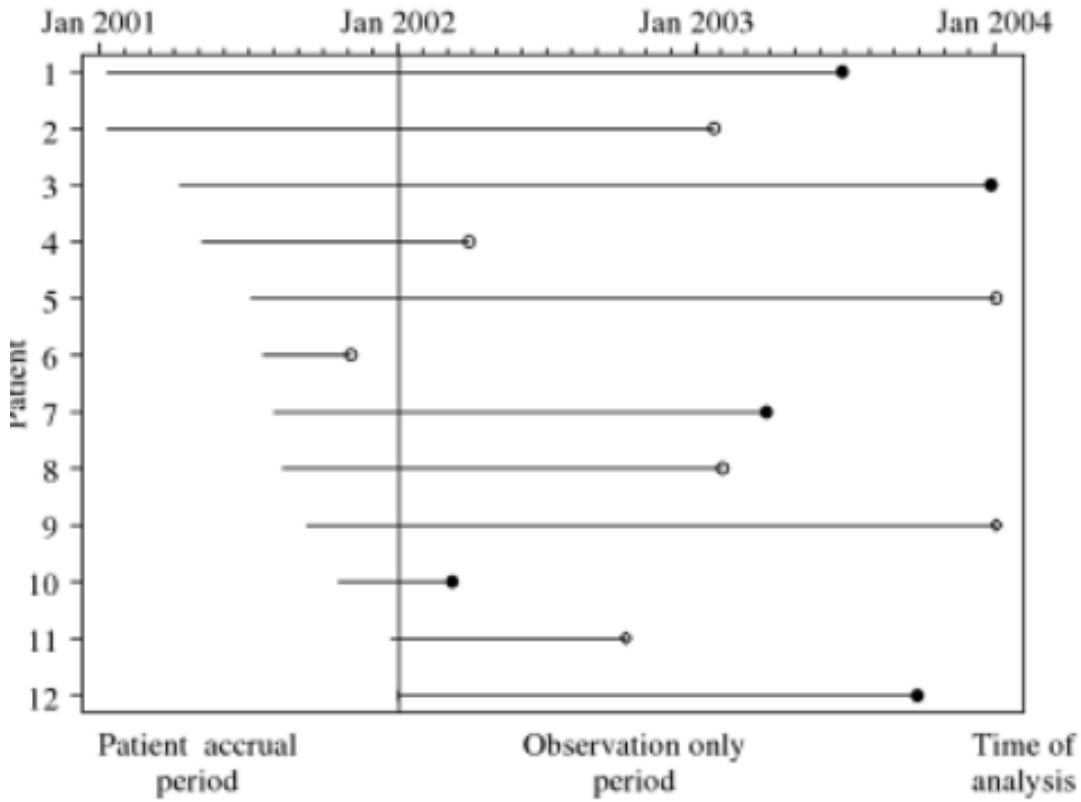
KRZYWE PRZEŻYCIA



Czas przeżycia

Czas przeżycia jest mierzony od jednego zdarzenia (np. start leczenia) do innego zdarzenia (np. śmierć). Pacjentów, dla których nie zaobserwowaliśmy danego zdarzenia nazywamy zmiennymi cenzorowanymi. Wszystko co wiemy to, że taki pacjent dożył do danego punktu czasu. Przyczyną cenzorowania może być również przerwanie leczenia przez pacjenta.

Czas przeżycia



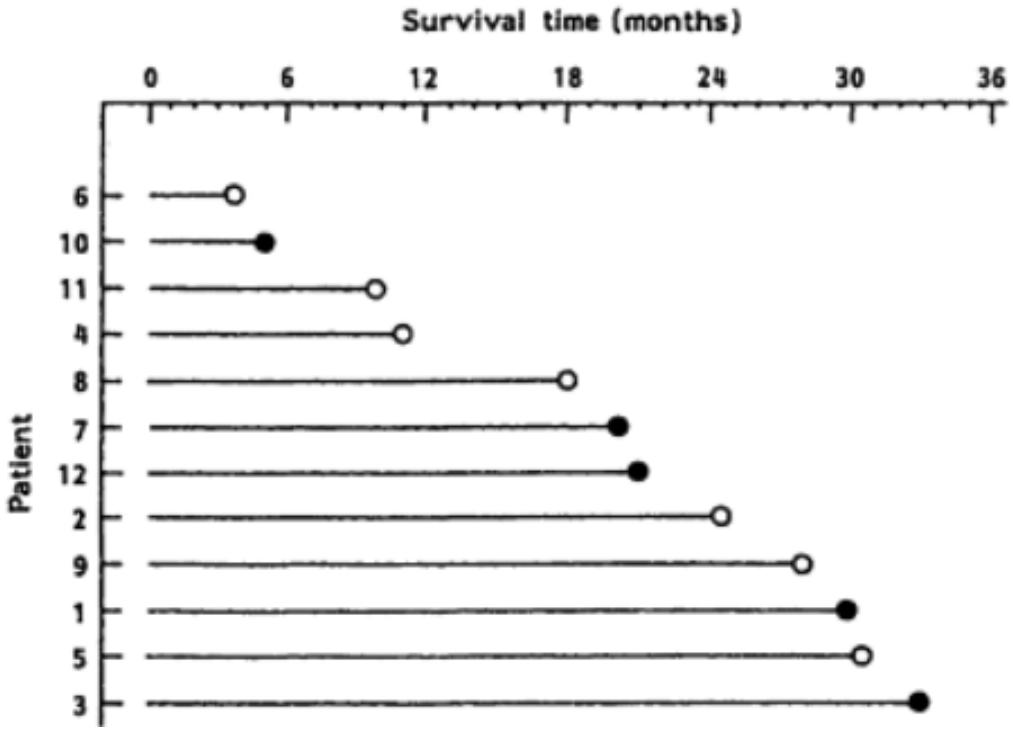
Rysunek 8: Pacjenci włączeni do badań w różnym czasie ze znanym (•) lub cenzorowanym (○) czasem przeżycia.

Czas przeżycia

Patient	Date of entry (dd.mm.yyyy)	Time at entry (days)	Time at death or censoring since 01.01.2001 (days)	Death (D) or censored (C)	Survival time t or $T+$ (days)
1	01.01.2001	0	910	D	910
2	01.01.2001	0	752	C	752+
3	26.03.2001	84	1090	D	1006
4	26.04.2001	115	451	C	336+
5	23.06.2001	173	1096	C	923+
6	09.07.2001	189	307	C	118+
7	22.07.2001	202	816	D	614
8	02.08.2001	213	762	C	549+
9	01.09.2001	243	1097	C	854+
10	07.10.2001	279	431	D	152
11	14.12.2001	347	644	C	297+
12	26.12.2001	359	1000	D	641

Rysunek 9: Tabela z czasem przeżycia dla pacjentów z Rysunku (8).

Czas przeżycia



Rysunek 10: Dane z Rysunku (8) uszeregowane ze względu na czas przeżycia.

Czas przeżycia

Niech T będzie (nieujemną) zmienną losową oznaczającą czas do interesującego nas zdarzenia (survival time). Wtedy dystrybuanta

$$F(t) = \mathbb{P}(T \leq t), \quad t > 0$$

oznacza prawdopodobieństwo, że nasze zdarzenie wydarzy się przed czasem t.

W bioinformatyce (oraz biomedycynie) używa się funkcji przeżycia (survival function)

$$S(t) = \mathbb{P}(T \geq t) = 1 - F(t^-)$$

wtedy $S(t)$ oznacza, że obiekt przeżyje do czasu t.

Czas przeżycia

Własności $S(t) = \mathbb{P}(T \geq t) = 1 - F(t^-)$:

- w $t = 0$ mamy $S(t) = 1$, czyli nikt jeszcze nie umarł
- w $t = +\infty$ mamy $S(t) = 0$, czyli każdy musi kiedyś umrzeć

Estymator Kaplana-Meiera

Rank	Survival time t (months)	Number at risk n_t	Observed deaths d_t	$p_t = 1 - \frac{d_t}{n_t}$	Survival proportion $S(t)$
–	0	24	0	1.0000	1
1	3+	24	0		
2	6				
3	6	23	4	0.8261	0.8261
4	6				
5	6				
6	8	19	2	0.8947	0.7391
7	8				
8	12	17	2	0.8824	0.6522
9	12				
10	12+	15	0		
11	15+	14	0		
12	16+	13	0		
13	18+	12			
14	18+				
15	20	10	1	0.9000	0.5780
16	22+	9	0		
17	24	8	1	0.8750	0.5136
18	28+				
19	28+	7	0		
20	28+				
21	30	4	1	0.7500	0.3852
22	30+	3	0		
23	33+	2	0		
24	42	1	1	0.0000	0

Rysunek 11: Czas do transplantacji serca oraz czas przeżycia dla 82 pacjentów wybranych z Programu Transplantacji Serca Stanford (Turnbull *et al.*, 1974).

Metoda Kaplana-Meiera

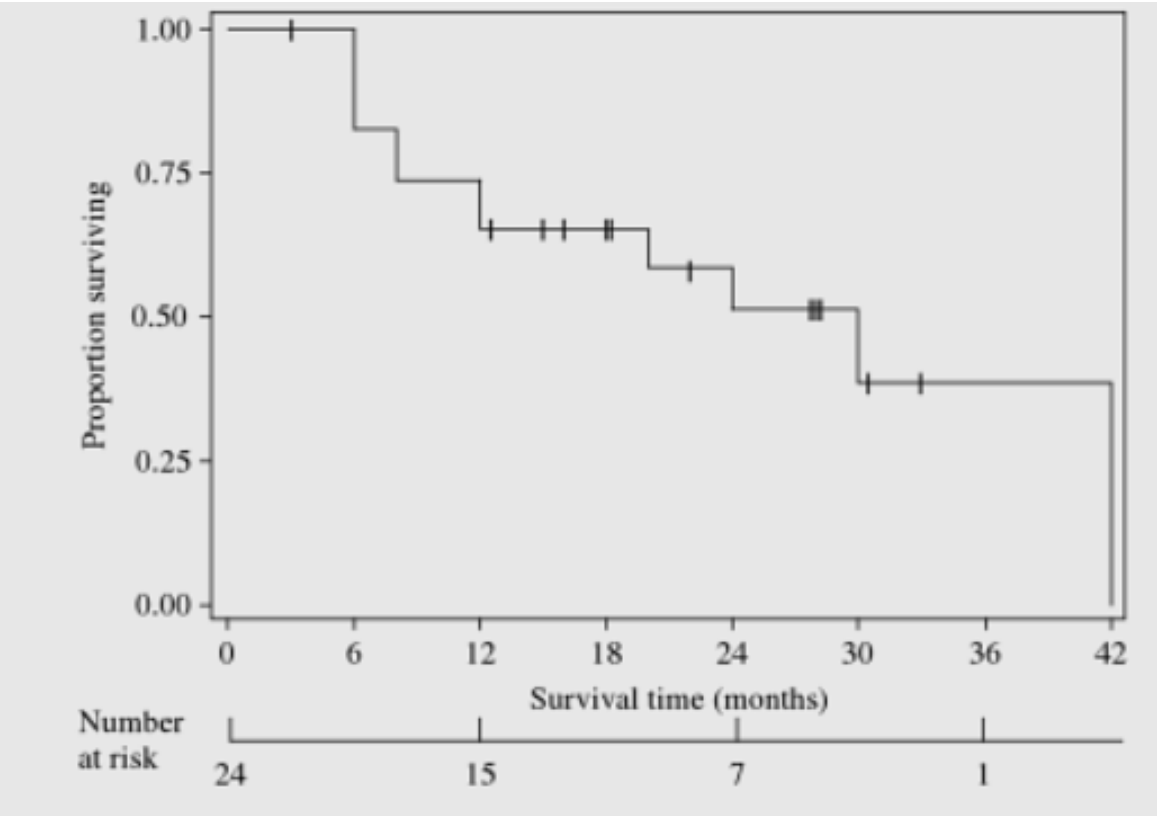
Założenie: czas przeżycia jest niezależny od czynnika powodującego cenzorowanie danych.
Estymator Kaplana-Meiera funkcji przeżycia:

$$\widehat{S}(t) = \prod_{j:t_j \leq t} \frac{r_j - d_j}{r_j}, \quad \text{dla} 0 \leq t \leq t^+$$

gdzie

- $\{t_j : j = 1, 2, \ldots, n\}$ - zbiór wszystkich momentów wystąpienia zdarzenia
- d_j - liczba wystąpień zdarzenia w chwili t_j
- $r_j = n_j - w_j$ - liczba obserwowanych obiektów zagrożonych w chwili t_j
- n_j - liczba obserwowanych obiektów w chwili t_j
- w_j - liczba obiektów uciętych w okresie (t_{j-1}, t_j) - np. pacjent wypisany ze szpitala
- t^+ - moment zakończenia badania

Krzywe przeżycia



Rysunek 12: Estymator krzywej przeżycia Kaplana-Meiera dla 24 pacjentów chorych na raka jelita grubego (McIllmurray and Turkie, 1987).

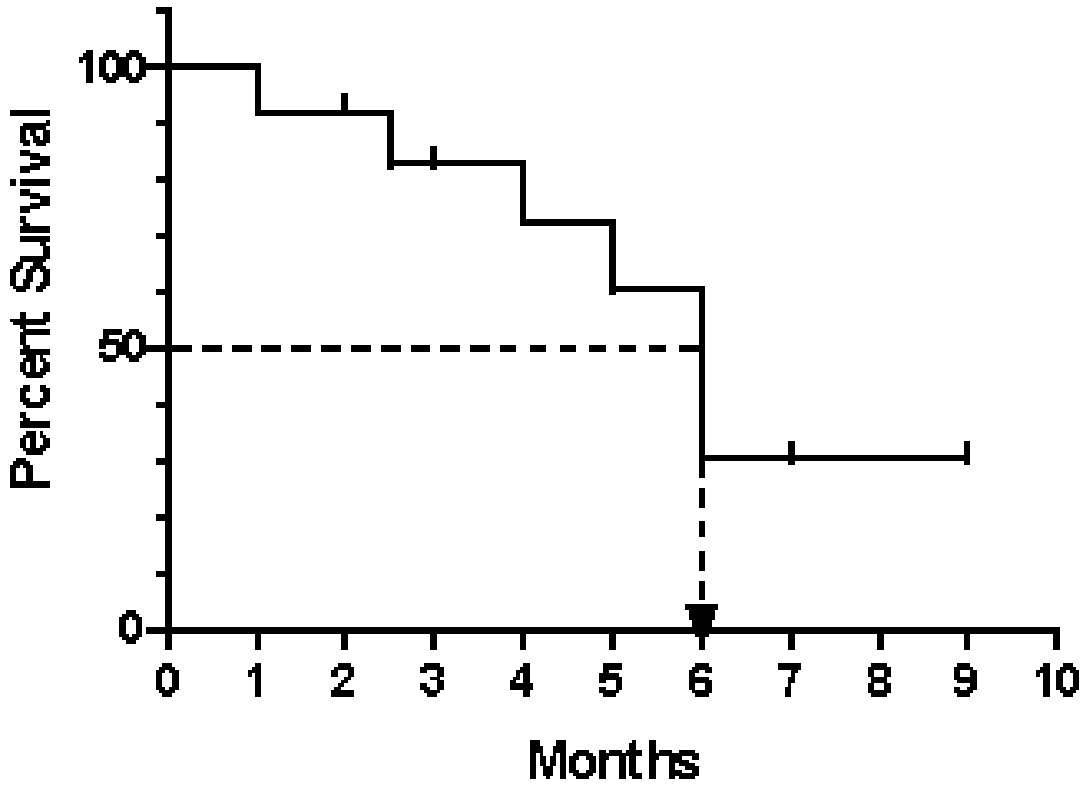
Mediana czasu przeżycia

- W zbiorze danych nie ma obserwacji cenzorowanych

$$M = \begin{cases} t_{([n+1]/2)} & n \text{ jest nieparzyste} \\ 0.5 \cdot \left(t_{(n/2)} + t_{(n/2+1)} \right) & n \text{ jest parzyste} \end{cases}$$

- W zbiorze danych występują obserwacje cenzorowane
  - Wyznaczyć krzywą przeżycia $S(t)$ Kaplana-Meiera
  - Znaleźć wartość M dla której $S(M) = 0.5$

Mediana czasu przeżycia - przykład



Rysunek 13: Mediana przeżycia (M=6) dla pacjentów chorych na stwardnienie rozsiane.

Przedziały ufności dla $S(t)$

Metoda Greenwooda

$$SE_{GR}\left[S(t)\right] = S(t)\left(\sum_{j=0}^{t-1}\frac{d_j}{r_j(r_j-d_j)}\right)^{\frac{1}{2}}$$

Przykład dla danych z Rysunek (11) w punkcie czasu $t = 12$:

$$SE_{GR}\left[S(12)\right] = 0.6522\left[\frac{4}{23\cdot(23-4)}+\frac{2}{19\cdot(19-2)}\right]^{\frac{1}{2}} = 0.0808$$

$$\begin{aligned} CI(95\%) &= (S(t) - 1.96\cdot SE_{GR}[S(t)]; S(t) + 1.96\cdot SE_{GR}[S(t)]) = \\ &= (0.4938; 0.8106) \end{aligned}$$

Przedziały ufności dla $S(t)$

Metoda Peto

$$SE_P\left[S(t)\right] = \left(\frac{S(t)[1-S(t)]}{R(t)}\right)^{\frac{1}{2}},$$

gdzie $R(t) = n - c_t$, c_t jest ilością cenzorowanych danych do chwili t.

Przykład dla danych z Rysunek (11) w punkcie czasu $t = 12$:

$$SE_P\left[S(12)\right] = \left(\frac{0.6522[1-0.6522]}{23}\right)^{\frac{1}{2}} = 0.0993$$

$$\begin{aligned} CI(95\%) &= (S(t) - 1.96 \cdot SE_P[S(t)]; S(t) + 1.96 \cdot SE_P[S(t)]) = \\ &= (0.4575; 0.8468) \end{aligned}$$

Przedziały ufności dla $S(t)$

Metoda transformacyjna

$$SE_{Tr}\left[S(t)\right] = \frac{\left(\sum_{j=0}^{t-1} \frac{d_j}{r_j(r_j-d_j)}\right)^{\frac{1}{2}}}{\left(-\sum_{j=0}^{t-1} \log\left(\frac{r_j-d_j}{r_j}\right)\right)}$$

Przykład dla danych z Rysunek (11) w punkcie czasu $t = 12$:

$$SE_{GR}\left[S(12)\right] = \frac{\sqrt{\frac{4}{23(23-4)} + \frac{2}{19(19-2)}}}{-\log\left(\frac{23-4}{23}\right) - \log\left(\frac{19-2}{19}\right)} = 0.4098$$

$$CI(95\%) = \left(S(t)^{\exp(+1.96 SE_{Tr})}; S(t)^{\exp(-1.96 SE_{Tr})}\right) =$$
$$= (0.3851; 0.8258)$$

Przedziały ufności dla M

Błąd standardowy dla mediany przeżycia wyznacza się z poniższego wzoru:

$$SE_M = SE_{GR}\left[S(M)\right] \cdot \frac{(t_{small} - t_{large})}{S(t_{large}) - S(t_{small})},$$

gdzie t_{small} jest najmniejszym czasem przeżycia z krzywej Kaplana-Meiera dla której $S(t)$ jest mniejsze bądź równe 0.45, natomiast t_{large} jest największym czasem przeżycia z krzywej Kaplana-Meiera dla której $S(t)$ jest mniejsze bądź równe 0.55.

Przedziały ufności dla M - Przykład

Na podstawie Rysunków (11) i (12) mamy $M = 30$, $t_{small} = 30$ i $t_{large} = 20$. Następnie należy wyestymować błąd standardowy metodą Greenwooda dla $S(30)$.

$$SE_{GR}\left[S(30)\right] = 0.5 \left[\frac{4}{23(23-4)} + \ldots + \frac{1}{8(8-1)}\right]^{\frac{1}{2}} = \\ = 0.5 \cdot 0.228377 = 0.114188$$

$$SE_M = 0.114188 \cdot \frac{(30-20)}{(0.5780 - 0.3852} = 5.92$$

$$CI(95\%) = (S(t) - 1.96 \cdot SE_M[S(t)]; S(t) + 1.96 \cdot SE_M[S(t)]) = \\ = (18.40; 41.60)$$

Hazard rate

Podstawowe pojęcia:

mortality rate - procent populacji, dla której będzie miało miejsce oczekiwane zdarzenie (np. śmierć) w okresie między t oraz $t+1$, dla których zdarzenie nie zaszło do czasu t.

$$m(t) = \mathbb{P}(t \leq T < t+1 | T \geq t)$$

hazard rate

$$\lambda(t) = \lim_{\Delta t \to 0} \frac{\mathbb{P}(t \leq T < t + \Delta t | T \geq t)}{\Delta t}$$

Hazard rate

Wzór:

$$\lambda(t) = \lim_{\Delta t \to 0} \frac{\mathbb{P}(t \leq T < t + \Delta t | T \geq t)}{\Delta t}$$

można zapisać jako

$$\lambda(t) = \frac{\lim_{\Delta t \to 0} \frac{\mathbb{P}(t \leq T < t + \Delta t)}{\Delta t}}{\mathbb{P}(T \geq t)} = \frac{f(t)}{S(t)} = -\frac{S'(t)}{S(t)} = -\frac{d \log[S(t)]}{dt}$$

Całkując obie strony otrzymujemy skumulowaną funkcję hazardu

$$\Lambda(t) = \int_0^t \lambda(u) du = \log[S(0)] - \log[S(t)] = -\log[S(t)]$$

Hazard rate

Ze wzoru:

$$\Lambda(t) = -\log[S(t)]$$

można wyliczyć $S(t)$

$$S(t) = \exp\left(-\Lambda(t)\right) = \exp\left(-\int_0^t \lambda(u)du\right)$$

Zależność wariancji funkcji przeżycia od wariancji hazardu skumulowanego:

$$var(\widehat{S}(t)) = \widehat{S}^2(t)var(\widehat{\Lambda}(t))$$

UWAGA!!!
hazard rate nie jest prawdopodobieństwem, może być większy od 1.

Hazard rate - przykład

Na podstawie Rysunku (11) wyznaczymy roczną funkcję hazardu dla pacjentów:

1. Sześcioro pacjentów zmarło (czworo w 6 i dwoje w 8 miesiącu), czyli przeżyli łącznie $f_1 = 6 \cdot 4 + 2 \cdot 8 = 40$ miesięcy. Z pacjentów, którzy przeżyli jeden był cenzorowany w trzecim miesiącu, a siedemnaścioro dożyło do końca roku, czyli przeżyli łącznie $F_1 = 1 \cdot 3 + 17 \cdot 12 = 207$ miesięcy. Wszyscy pacjenci przeżyli $f_1 + F_1 = 247$ miesięcy. Funkcja hazardu dla pierwszego roku wynosi $h_1 = \frac{d_1}{f_1 + F_1} = \frac{6}{247} = 0.02429$ na miesiąc lub $0.02429 \cdot 12 = 0.29$ na rok.

2. $h_2 = 0.019608$ na miesiąc lub 0.24 na rok.

3. $h_2 = 0.06061$ na miesiąc lub 0.73 na rok.

4. $h_2 = 0.16667$ na miesiąc lub 2.00 na rok.

Co zrobić ze zdarzeniami na brzegach przedziałów (np. zgon nastąpił w 12 miesiącu)?

Metoda Flemingtona-Harringtona

Zależność między funkcją przeżycia, a skumulowanym hazardem ma postać:

$$S(t) = \exp\left(-\Lambda(t)\right)$$

Estymator Nelsona-Aalena skumulowanego hazardu ma postać:

$$\widehat{\Lambda}(t) = \sum_{j:t_j \leq t} \frac{d_j}{r_j}, \quad 0 \leq t \leq t^+$$

gdzie t_j, r_j, d_j - jak w przypadku estymatora Kaplana-Meiera.
Estymator Flemingtona-Harringtona funkcji przeżycia otrzymamy przez podstawienie estymatora Nelsona-Aalena do funkcji skumulowanego hazardu

$$\check{S}(t) = \exp\left(-\widehat{\Lambda}(t)\right)$$

Porównanie estymatorów KM i FH

Dla dowolnego t mamy oszacowanie

$$\check{S}(t) = \exp\left(-\widehat{\Lambda}(t)\right) = \exp\left(-\sum_{j:t_j \le t} \frac{d_j}{r_j}\right) =$$

$$= \prod_{j:t_j \le t} \exp\left(-\frac{d_j}{r_j}\right) \ge \prod_{j:t_j \le t} \left(1 - \frac{d_j}{r_j}\right) = \widehat{S}(t)$$

Pakiet R

Pakiet R

```
1 library(survival)
2 data(stanford2)
3 attach(stanford2)
4 head(stanford2)
```

```
>
> library(survival)
Ładowanie wymaganego pakietu: splines
> data(stanford2)
> attach(stanford2)
> head(stanford2)
     id time status age   t5
139 139   86      1  12 1.26
159 159   10      1  13 1.49
181 181   60      0  13   NA
119 119 1116      0  14 0.54
74   74 2006      0  15 1.26
120 120 1107      0  18 0.25
>
```

Rysunek 14: Dane "Stanford2" zaimportowane do pakietu R.

Pakiet R

```
1 Surv(time,status)
2 survfit(Surv(time,status)~1)
```

```
> Surv(time,status)
  [1]   86.0    10.0    60.0+ 1116.0+ 2006.0+ 1107.0+ 1232.0    42.0    98.0+  227.0   793.0+    5.0   119.0+
 [14]  534.0+  781.0+    1.0   149.0+ 1634.0   468.0+ 1193.0+  631.0  2313.0+    1.0+   22.0    56.0+   89.0+
 [27]  391.0+ 1778.0+    7.0    48.0    12.0   176.0+  548.0+  993.0+ 2025.0+  274.0  1150.0  1271.0  2723.0
 [40] 2984.0+   31.0  1961.0    30.0   328.0  1262.0+   68.0   221.0  2127.0    44.0    47.0   322.0+  382.0
 [53]  663.0+ 1478.0    90.0   464.0+  731.0  1202.0  1945.0+ 3021.0+    2.0+  195.0   406.0+ 1102.0+ 1718.0+
 [66]    3.0   14.0   549.0+ 1722.0+ 2106.0+ 3695.0+    0.5   138.0   278.0+  563.0+  738.0+ 1247.0  1373.0+
 [79] 1378.0+   39.0    40.0    46.0   191.0   297.0   660.0+  161.0   292.0+  293.0   752.0+ 1024.0  1040.0+
 [92]  836.0  1534.0  1846.0+   22.0    36.0    65.0   121.0   146.0   381.0  2000.0+ 3410.0+   44.0    50.0
[105]  107.0+  132.0   897.0   950.0+ 1384.0  1393.0+   51.0   148.0   374.0+  431.0   541.0+  619.0+ 1261.0+
[118] 1995.0+ 2734.0+   45.0   126.0   202.0   254.0   323.0   550.0   841.0+  994.0  1318.0  1638.0+ 2805.0+
[131]   54.0    66.0   263.0   265.0   279.0   538.0   547.0   729.0   834.0+ 1866.0+ 1996.0  2878.0    50.0
[144]   51.0   145.0   618.0+  928.0  1341.0+ 1547.0+   48.0    66.0    68.0   122.0   139.0   169.0   623.0
[157] 1612.0+   25.0    26.0    51.0   136.0   188.0+  231.0+  544.0  1264.0+ 2474.0    25.0    29.0    48.0
[170]  576.0+  626.0     1.0    15.0    64.0  1350.0  1352.0    65.0   136.0    10.0    23.0    63.0   730.0
[183]   47.0    60.0
> survfit(Surv(time,status)~1)
Call: survfit(formula = Surv(time, status) ~ 1)

records   n.max n.start  events  median 0.95LCL 0.95UCL
    184     184     184     113     631     328    1232
> |
```

Rysunek 15: Prezentacja funkcji "Surv()" i survfit" w pakiecie R.

Pakiet R

```
1 summary(survfit(Surv(time,status)~1))
```

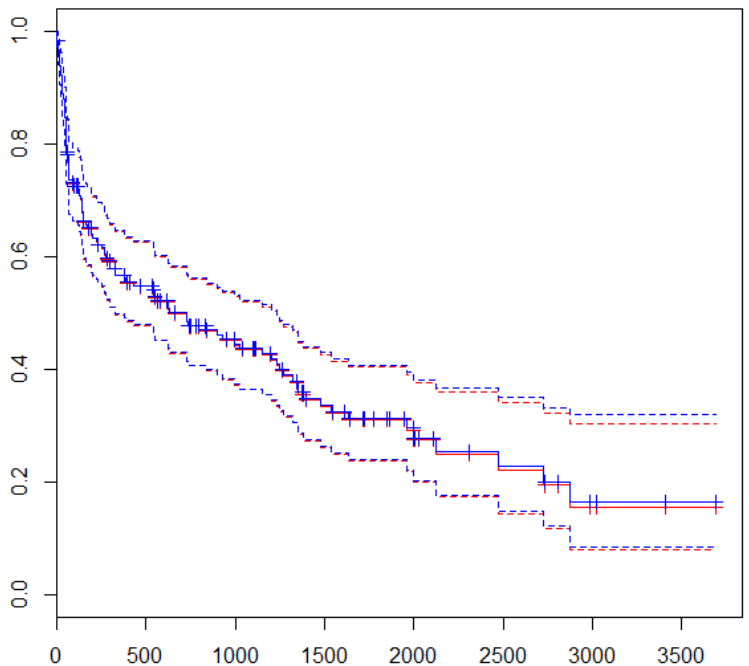
```
> summary(survfit(Surv(time,status)~1))
Call: survfit(formula = Surv(time, status) ~ 1)

   time n.risk n.event survival std.err lower 95% CI upper 95% CI
    0.5    184       1    0.995 0.00542        0.9840        1.000
    1.0    183       2    0.984 0.00934        0.9656        1.000
    3.0    179       1    0.978 0.01078        0.9573        1.000
    5.0    178       1    0.973 0.01204        0.9494        0.997
    7.0    177       1    0.967 0.01317        0.9417        0.993
   10.0    176       2    0.956 0.01514        0.9270        0.986
   12.0    174       1    0.951 0.01602        0.9198        0.983
   14.0    173       1    0.945 0.01684        0.9128        0.979
   15.0    172       1    0.940 0.01762        0.9058        0.975
   22.0    171       2    0.929 0.01905        0.8921        0.967
   23.0    169       1    0.923 0.01971        0.8854        0.963
   25.0    168       2    0.912 0.02095        0.8721        0.954
   26.0    166       1    0.907 0.02154        0.8655        0.950
   29.0    165       1    0.901 0.02210        0.8590        0.946
   30.0    164       1    0.896 0.02263        0.8525        0.941
   31.0    163       1    0.890 0.02315        0.8460        0.937
   36.0    162       1    0.885 0.02365        0.8396        0.932
   39.0    161       1    0.879 0.02414        0.8332        0.928
   40.0    160       1    0.874 0.02460        0.8269        0.923
   42.0    159       1    0.868 0.02505        0.8205        0.919
   44.0    158       2    0.857 0.02591        0.8080        0.910
   45.0    156       1    0.852 0.02632        0.8017        0.905
   46.0    155       1    0.846 0.02672        0.7955        0.900
   47.0    154       2    0.835 0.02748        0.7832        0.891
   48.0    152       3    0.819 0.02854        0.7648        0.877
```

Rysunek 16: Podstawowe statystyki dla funkcji “survfit” w pakiecie R.

Pakiet R

```
plot(survfit(Surv(time,status)~1),col='red')
lines(survfit(Surv(time,status)~1,type="fleming-harrington"),col='blue')
```



Rysunek 17: Krzywe przeżycia Kaplana-Meiera (czerwony) i Flemingtona-Harringtona (niebieski).

Bibliografia

Wykład opracowany na podstawie książki:

- Mahesh K. B. Parmer and David Machin “Survival Analysis - A Practical Approach”