

Analiza cenzuriranih podataka i krivulje preživljenja

Prof. dr. sc. Mladen Petrovečki



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku

Analiza preživljenja

1. temeljni pojmovi
2. izračun vjerojatnosti preživljenja
 - a) tablice preživljenja
 - b) Kaplan-Meierov postupak
3. rizik umiranja
4. programska potpora
5. usporedba podataka o preživljenju
6. statističko zaključivanje
7. regresijska analiza cenzuriranih podataka



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku

Obrada podataka o preživljenju bolesnika

- analiza preživljenja
- *survival analysis*
- ponekad
 - analiza tablica preživljenja
 - analiza osiguravateljskih (aktuarskih) podataka
 - *actuarial analysis*



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku

Analiza preživljenja

- Edmund Halley, 17. st
- engleski astronom, geofizičar, matematičar, meteorolog i fizičar
- http://en.wikipedia.org/wiki/Edmond_Halley



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku

www.aktuari.hr

- aktuar – stručnjak koji se bavi problemima financijske neizvjesnosti i rizika koristeći matematičke metode teorije vjerojatnosti, statistike i financijske matematike
- posao – analiza podataka iz prošlosti, procjenu postojećih rizika i razvoj modela za projekciju budućih događaja
- zaposlenje – osiguranje i mirovinsko osiguranje
- znanja – matematika, ekonomija, praksa i zakoni države u kojoj radi, demografska i financijska kretanja, vještina komunikacije



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku

Aktuarske tablice preživljenja (tablice smrtnosti)

8 National Vital Statistics Reports, Vol. 54, No. 14, April 19, 2006

Table 1. Life table for the total population: United States, 2003

Click here for spreadsheet version

Age	Probability of dying between ages x to $x+1$	Number surviving to age x	Number dying between ages x to $x+1$	Person-years lived between ages x to $x+1$	Total number of person-years lived above age x	Expectation of life at age x
	q_x	l_x	d_x	L_x	T_x	e_x
0-1	0.000885	100,000	887	99,394	7,748,865	77.5
1-2	0.000485	99,313	48	99,280	7,649,471	77.0
2-3	0.000391	99,267	39	99,251	7,550,191	76.1
3-4	0.000259	99,234	26	99,222	7,450,930	75.1
4-5	0.000188	99,209	20	99,189	7,351,709	74.1
5-6	0.000169	99,189	17	99,181	7,252,519	73.1
6-7	0.000151	99,172	15	99,165	7,153,329	72.1
7-8	0.000142	99,158	14	99,150	7,054,164	71.1
8-9	0.000139	99,143	14	99,137	6,955,013	70.2
9-10	0.000134	99,130	13	99,123	6,855,877	69.2
10-11	0.000165	99,116	16	99,108	6,756,754	68.2
11-12	0.000147	99,100	15	99,093	6,657,646	67.2
12-13	0.000176	99,085	17	99,077	6,558,553	66.2



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku

http://en.wikipedia.org/wiki/Actuarial_table

Analiza preživljenja

- psihijatrija – 1%
- patologija – 1%
- kirurgija – 12%
- onkologija – 14%
- izvorni znanstveni radovi u *The New England Journal of Medicine* – 32%
- podaci 1986.-2001., Dawson Saunders & Trapp, Basic and Clinical Biostatistics

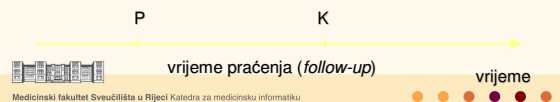


Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku



Analiza preživljenja

- analiza podataka vezanih uz vremensko praćenje događaja
- dvije točke praćenja:
 - početak (P) (*time origin*)
 - kraj (K) (*end point*)

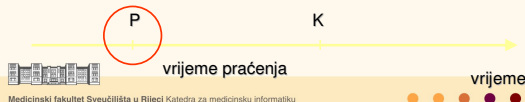


Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku



Početak praćenja

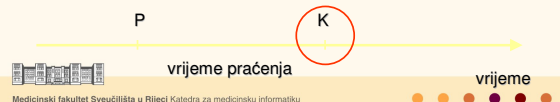
- rođenje
- pojava znaka bolesti
- postavljanje dijagnoze
- početak liječenja
- dan operativnog zahvata



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku

Kraj praćenja

- smrt od osnovne bolesti
- smrt (svi ostali mogući uzroci)
- ponovno javljanje bolesti
- postizanje učinka liječenja
- gubitak iz uzorka (ispitne skupine)



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku



Kraj praćenja

- smrt od osnovne bolesti
- smrt (svi ostali mogući uzroci)

usklađeno preživljenje
engl. *adjusted survival rate*

ukupno preživljenje
engl. *observed survival rate*

Manual for Staging of Cancer
3rd ed., AJCC



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku



Vrijeme praćenja

- raspodjela u pravilu nije simetrična
- podaci su nepotpuni, praćenje je nepotpuno, "cenzurirano" (*censored data*)
- podaci za primjere:
 - istraživanje karcinoma usne šupljine
 - MFK KBD
 - dr. Ivica Lukšić
 - n = 52; 1. siječnja 2000. – 31. prosinca 2004.
 - reprezentativni probirani uzorak
 - dio populacije tog razdoblja
 - prva dg. karcinoma, bez regionalnih metastaza, itd.

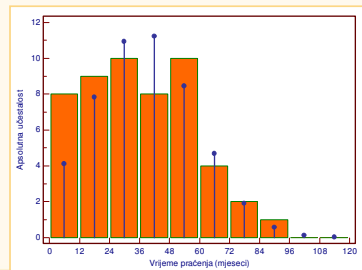


Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku



Vrijeme praćenja (1)

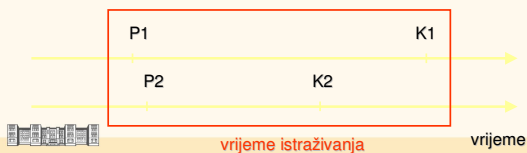
- raspodjela u pravilu nije simetrična



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku

Vrijeme praćenja (2)

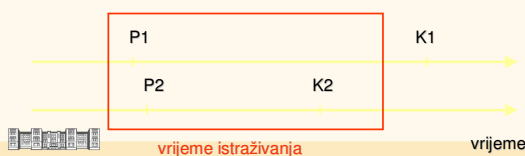
- potpuni podaci (potpuno praćenje)



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku

Vrijeme praćenja (2)

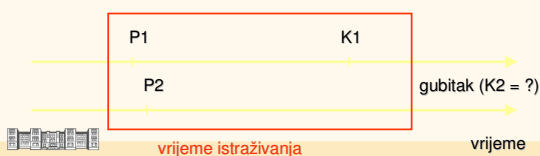
- podaci su nepotpuni, praćenje je nepotpuno, "cenzurirano"
 - cenzurirano vrijeme praćenja = jedinka tijekom praćenja ne dostiže očekivani događaj



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku

Vrijeme praćenja (3)

- podaci su nepotpuni, praćenje je nepotpuno, "cenzurirano"
 - cenzurirano vrijeme praćenja = jedinka tijekom praćenja ne dostiže očekivani događaj



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku

Cenzuriranje

- događaj se ostvaruje = 1
- sve ostalo = 0 (cenzurirani podaci)
 - kraj istraživanja (end of the study)
 - gubitak iz praćenja (lost to follow-up)
 - ostali događaji

Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku

"Izgubljen iz praćenja"

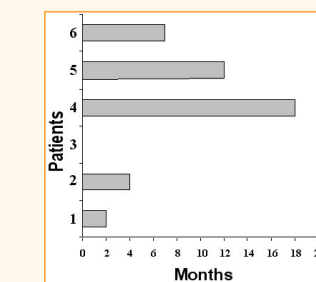
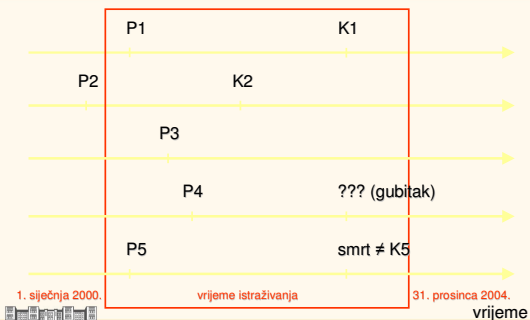


Figure 3. Outcome

The mean length of survival for our patients was 7 months. Patient 3 was lost to follow up.

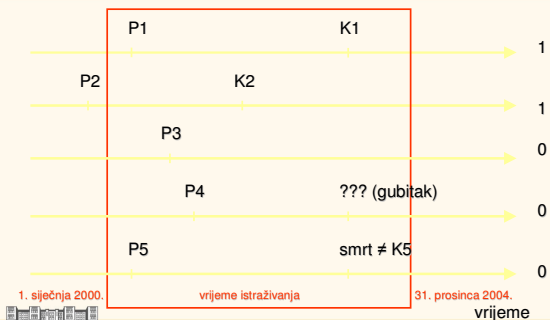
Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku

Cenzuriranje: bolesnici s postavljenom dijagnozom (P) u zadanih pet godina



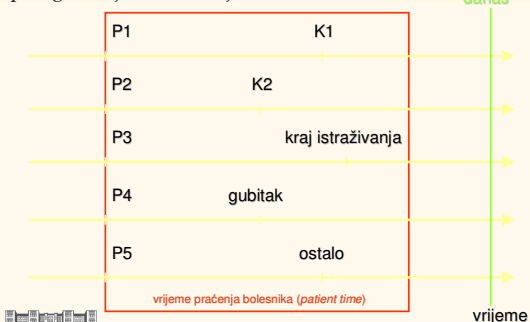
Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku

Cenzuriranje: bolesnici liječeni u zadanom petogodišnjem razdoblju



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku

Cenzuriranje: bolesnici liječeni u zadanom petogodišnjem razdoblju



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku

Cenzuriranje

- desno (*right censoring*)
 - nije lijevo
 - nije intervalno
- neinformativno (*non informative*)
 - nije "informativno"
 - vrijeme preživljenja neke jedinice neovisno je od svakog mogućeg mehanizma koji može uzrokovati da ista jedinica bude cenzurirana
 - npr. cenzuriranje bolesnika s pogoršanjem zdravstvenog stanja

Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku

A sada – veselje!

$$R(t) = P\{T > t\} = \int_t^{\infty} f(u) du = 1 - F(t).$$

- funkcija preživljenja
 - biomedicina
 - *survival function*
- funkcija pouzdanosti
 - inženjerstvo
 - *reliability function*

S(t) ili R(t):

- vjerojatnost da će jedinica preživjeti ili točno doživjeti vrijeme od t jedinica praćenja, ili
- vjerojatnost preživljenja jedinice u rasponu od početka praćenja do trenutka praćenja t

Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku

A sada – još veće veselje!

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{R(t) - R(t + \Delta t)}{\Delta t \cdot R(t)}, \quad F(t) = \int_0^t f(x) dx$$

- funkcija rizika
 - *hazard function*
- kumulativna funkcija rizika umiranja
 - $H(t) = -\log S(t)$

h(t):

- vjerojatnost da će jedinica umrijeti u trenutku t , uz uvjet da je preživjela do toga trenutka, i uz uvjet da je
- F(t) funkcija gustoće

Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku

Zaključak: podaci o preživljenju

- vjerojatnost preživljenja
 - $S(t)$
- rizik umiranja
 - $H(t) = -\log S(t)$



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku



Izračunavanje preživljenja

- neparametrijski postupci
 - Cutler-Edererov postupak (tablice preživljenja)
 - Kaplan-Meierov postupak
- parametrijski



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku



I. Tablice preživljenja

- osiguravateljske tablice
- tablice smrtnosti

vrijeme praćenja	vjerojatnost smrti	vjerojatnost preživljenja	kumulativna vjerojatnost preživljenja
	$q = d/(n-w/2)$	$p = 1 - q$	$S(t) = \prod p$
0-12 mj.	0,11	0,89	0,89
13-24 mj.	0,27	0,73	0,65
25-36 mj.	0	1	0,65
37-48 mj.	0,4	0,6	0,39
49-60 mj.	0	1	0,39



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku



Kako do preživljenja?

- upis podataka
- preuređenje podataka
- izračun podataka



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku



1. Upis podataka, Excel®

Oblikovanje ćelija

Formula: $= (C2-B2)/365 * 12$

	B	C	D	E
	pacijent	datumop	datumkraj	cenzus
1	23456	23.6.2000	15.4.2007	0
2	24485	15.10.2003	8.11.2005	0
3	23080	25.7.2000	29.8.2004	0
4	23511	28.12.2001	15.2.2007	0
5	24188	20.2.2002	29.10.2004	0
6	22701	17.12.2003	8.6.2005	1
7	24544	9.1.2002	29.9.2003	1
8	23859	10.10.2000	16.11.2003	0
9	22819	1.3.2001	6.2.2002	1
10	22921	26.4.2004	7.12.2004	0
11	23309	9.7.2004	18.7.2006	0

Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku

2. Preuređenje podataka

	A	B	C	D	E
	pacijent	datumop	datumkraj	cenzus	mjeseci
1	23456	15.10.2003	8.11.2005	0	24,8
2	23080	25.7.2000	29.8.2004	0	49,2
3	24188	20.2.2002	29.10.2004	0	32,3
4	22701	17.12.2003	8.6.2005	1	17,7
5	23998	27.4.2000	2.2.2004	1	45,2
6	24544	9.1.2002	29.9.2003	1	20,6
7	23859	10.10.2000	16.11.2003	0	37,2
8	22819	1.3.2001	6.2.2002	1	11,2
9	22921	26.4.2004	7.12.2004	0	7,4
10	23309	9.7.2004	18.7.2006	0	24,3

vrijeme praćenja	živi na početku intervala n	smrtni ishod u intervalu d	cenzurirani u intervalu w
0-12 mj.	10	1	1
13-24 mj.	8	2	1
25-36 mj.	5	0	2
37-48 mj.	3	1	1
49-60 mj.	1	0	1



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku

3. Izračun podataka

vrijeme praćenja	vjerojatnost smrti $q=d/(n-w/2)$	vjerojatnost preživljenja $p=1-q$	kumulativna vjerojatnost preživljenja $S(t)=\prod p$
0-12 mj.	0,11	0,89	0,89
13-24 mj.	0,27	0,73	0,65
25-36 mj.	0	1	0,65
37-48 mj.	0,4	0,6	0,39
49-60 mj.	0	1	0,39

vrijeme praćenja	živi na početku intervala n	smrtni ishod u intervalu d	cenzurirani u intervalu w
0-12 mj.	10	1	1
13-24 mj.	8	2	1
25-36 mj.	5	0	2
37-48 mj.	3	1	1
49-60 mj.	1	0	1



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku

3. Izračun podataka

vrijeme praćenja	vjerojatnost smrti $q=d/(n-w/2)$	vjerojatnost preživljenja $p=1-q$	kumulativna vjerojatnost preživljenja $S(t)=\prod p$
0-12 mj.	0,11	0,89	0,89
13-24 mj.	0,27	0,73	0,65
25-36 mj.	0	1	0,65
37-48 mj.	0,4	0,6	0,39
49-60 mj.	0	1	0,39

$$q = \frac{d + \frac{1}{2}wq}{n}$$

d – smrtni ishod u intervalu
 n – živi na početku intervala
 w – izgubljeni u intervalu

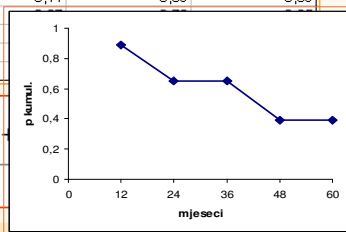


Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku

3. Izračun podataka

vrijeme praćenja	vjerojatnost smrti $q=d/(n-w/2)$	vjerojatnost preživljenja $p=1-q$	kumulativna vjerojatnost preživljenja $S(t)=\prod p$
0-12 mj.	0,11	0,89	0,89
13-24 mj.			
25-36 mj.			
37-48 mj.			
49-60 mj.			

$$q = \frac{d}{n}$$



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku

II. Kaplan-Meierov postupak

- nema zadane intervale praćenja
- vjerojatnost preživljenja se izračunava za svakog bolesnika koji umre
- cenzurirani ispitanici nisu dio izračunavanja vjerojatnosti
- dugotrajno izračunavanje kod velikih skupina ispitanika

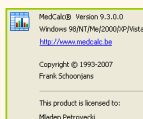


Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku

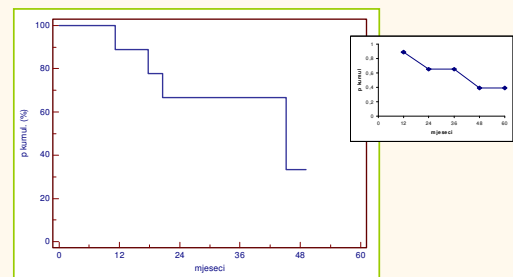
KM podaci o preživljenju

Survival time	mjeseci
Endpoint	cenzus
Sample size	10
Median survival	45,2

Survival time	Survival Proportion	Standard Error
7,4	-	-
11,2	0,889	0,105
17,7	0,778	0,139
20,6	0,667	0,157
24,3	-	-
24,3	-	-
32,3	-	-
37,2	-	-
45,2	0,333	0,248
49,2	-	-



KM krivulja preživljenja



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku

Kaplan, Meier

- Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. J Am Stat Assoc 1958;53:457-81.



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku

Kaplan, Meier

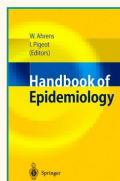
- među 5 najcitiranijih radova u znanosti od trenutka objavljivanja (M. Zhou, Kentucky University; <http://www.ms.uky.edu/~mai/>)
- prikaz krivulje u zavisnosti od N <http://www.ms.uky.edu/~mai/java/stat/KapMei.html>



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku

Rizik umiranja

- $H(t) = -\log S(t)$
- snaga mortaliteta (epidemiologija)
- ...

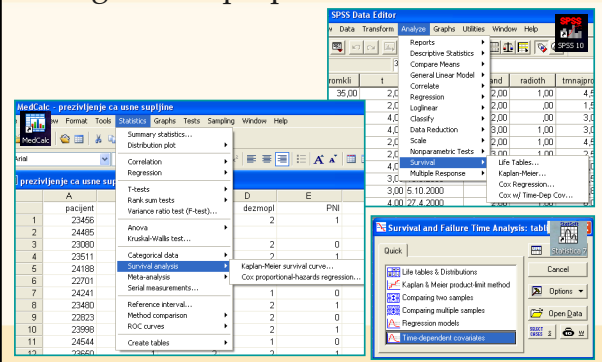


http://www.bips.uni-bremen.de/handbook_of_epidemiology.php

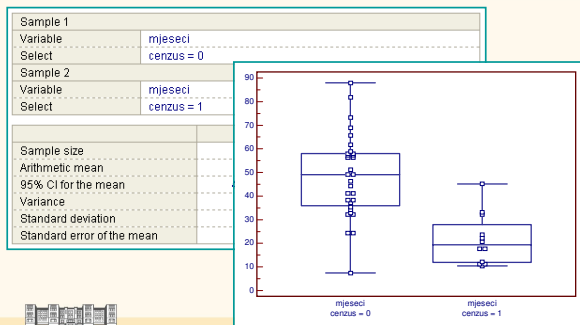


Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku

Programska potpora



Primjer...



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku

Preživljenje, MedCalc®

Survival time	mjeseci		
Endpoint	cenzus		
Sample size	43		
Median survival			
Survival time	Survival Proportion	Standard Error	
7,3973			
10,3562	0,976	0,024	
11,2438	0,952	0,033	
11,8384	0,929	0,040	
12	0,905	0,045	
17,7205	0,881	0,050	
17,8849	0,857	0,054	
20,6466	0,833	0,058	
22,1918	0,810	0,061	
23,7041	0,786	0,063	
24,2959			
24,8219			
32,1863	0,760	0,066	
45,2384			
46,3233	0,692	0,077	
49,1836			
49,4466			
51,2219			
56,3507			
56,9096			
57,3041			
57,4356			
58,0274			
58,126			
58,9479			
61,6438			
65,6219			
68,074			
73,2184			
81,7644			
88,0438			

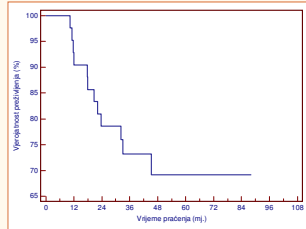
npr., prosječno petogodišnje preživljenje: 69,2±7,7%



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku

Krivulja preživljenja, MedCalc®

Survival time	Endpoint	mjeseci	cenzus
Sample size			43
Median survival			
Survival time	Survival Proportion	Standard Error	
7,3973			
10,3582	0,976	0,024	
11,2438	0,952	0,033	
11,6384	0,929	0,040	
12	0,905	0,045	
17,7205	0,881	0,050	
17,8949	0,857	0,054	
20,8466	0,833	0,059	
22,1918	0,810	0,061	
23,7041	0,786	0,063	
24,2959	-	-	
24,8219	-	-	
32,1863	0,760	0,066	



npr., prosječno petogodišnje preživljenje: 69,2±7,7%

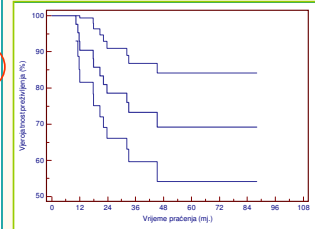


Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku



Granice pouzdanosti

Survival time	Endpoint	mjeseci	cenzus
Sample size			43
Median survival			
Survival time	Survival Proportion	Standard Error	
7,3973			
10,3582	0,976	0,024	
11,2438	0,952	0,033	
11,6384	0,929	0,040	
12	0,905	0,045	
17,7205	0,881	0,050	
17,8949	0,857	0,054	
20,8466	0,833	0,059	
22,1918	0,810	0,061	
23,7041	0,786	0,063	
24,2959	-	-	
24,8219	-	-	
32,1863	0,760	0,066	



npr., prosječno petogodišnje preživljenje: 69,2±7,7%



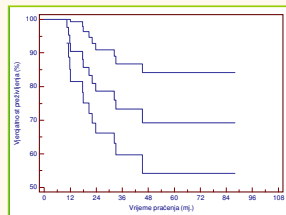
Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku



Izračun granica pouzdanosti

- granice pouzdanosti (Confidence Intervals)
- CI = $x \pm z \text{ SE}(x)$
 - 95%CI = $x \pm 1,96 \text{ SE}(x)$
 - 99%CI = $x \pm 2,56 \text{ SE}(x)$

$$P(Z \leq z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du$$



<http://www.fourmilab.ch/rpkp/experiments/analysis/zCalc.html> (pazi: p/2!)



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku



Granice pouzdanosti

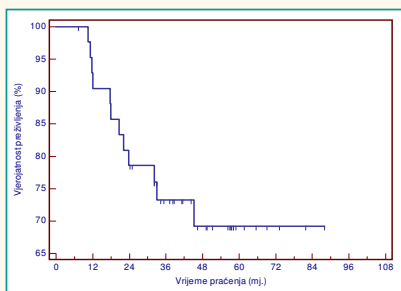
Results from life table analysis are usually presented in a survival curve rather than in tables. The solid line in Fig 11-3 is a survival curve for the kidney transplant data. The dashed lines on either side of the survival curve represent 95% **confidence bands** for the curve. Although confidence bands are often not presented in journal articles, they should be included because they help readers interpret the amount of variability in the results. Typically, as the time interval from entry into the study becomes longer, the number of patients who have been in the study that long becomes increasingly smaller. The confidence



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku



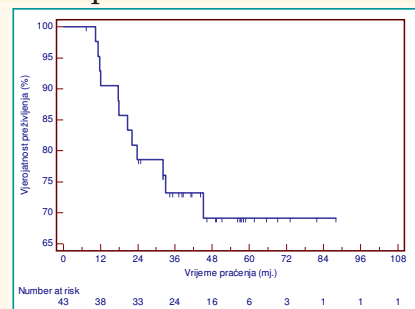
Cenzurirani podaci



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku



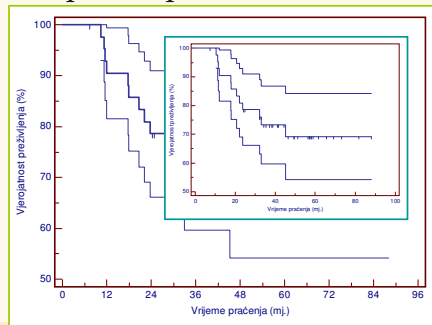
Broj bolesnika pod rizikom



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku

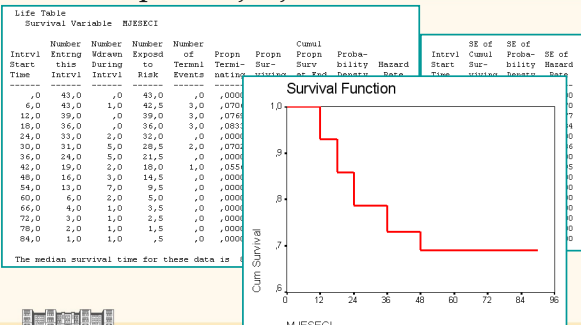


Standardni prikaz podataka



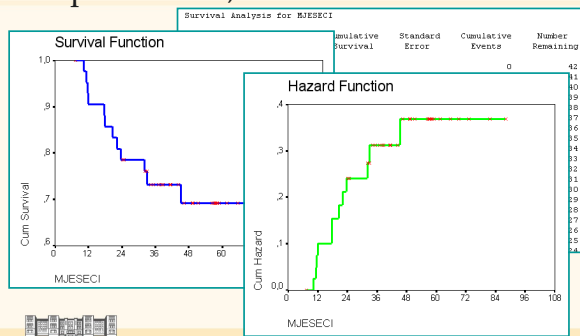
Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku

Tablice preživljenja, SPSS®



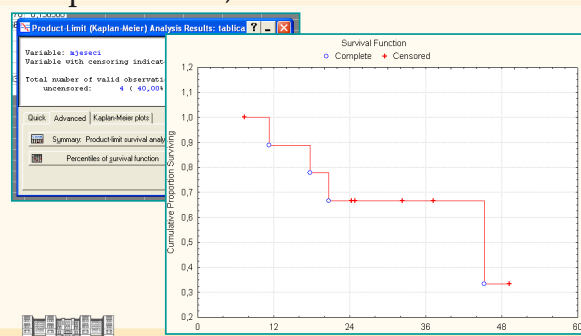
Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku

Kaplan-Meier, SPSS®



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku

Kaplan-Meier, Statistica®



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku

"Preživljenje"

- generalno: **preživljenje**
 - survival
 - praćenje
 - početak
 - smrt od osnovne bolesti
- posebno: **preživljenje bez znakova bolesti**
 - disease-free survival (DFS)
 - praćenje
 - početak
 - relaps bolesti ili smrt od osnovne bolesti



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku

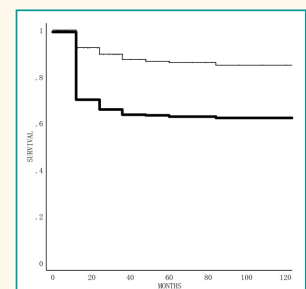
DFS, primjer

Survival curves for 277 patients with T₂ tongue carcinoma treated with an ¹⁹²Ir, ¹³⁷Cs or ²²⁶Ra needle;
thin – regional disease-free survival
thick – distant metastasis-free survival

<http://www.royal-society.com/content/1/1/21/figure/F5>



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku



Usporedba podataka o preživljenju

- usporedba dvije skupine podataka
 - *log-rank (logrank)* test
 - Mentelov ili Mentel-Coxov test
 - Wilcoxonov test
 - generalizirani Wilcoxonov test
 - Gehanov test
 - Gehan-Breslowljev test
 - opći Kruskal-Wallisov test za cenzurirane podatke
 - Mantel-Haenszelov test
 - Tarone-Wareov test
- usporedba triju i više skupina



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku

Usporediti dvije skupine...

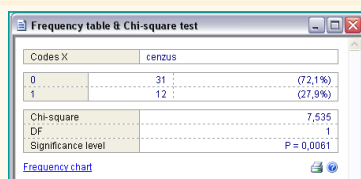
ceiving a kidney in 1984 is above the curve for patients receiving a kidney in 1978, indicating a higher proportion of patients retaining a functioning graft at any one point in time. However, variation in samples may be expected to occur simply by chance, and a reasonable question is whether the differences between the two patient cohorts is greater than expected by chance. To test this hypothesis, we need methods to compare survival distributions. If there are no censored observations, the **Wilcoxon rank-sum test** in-



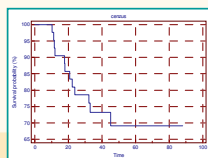
Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku

Primjer

- N = 43

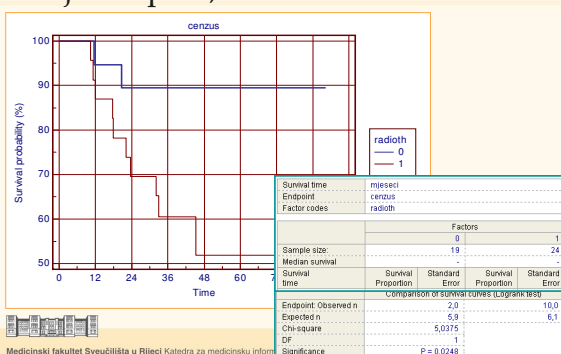


Variable	mjeseci
Select	(cenzus = 1) or (cenzus = 0)
Sample size	43
Lowest value	7.3973
Highest value	88.0438
Arithmetic mean	40.9583
95% CI for the mean	34.8091 to 47.1074



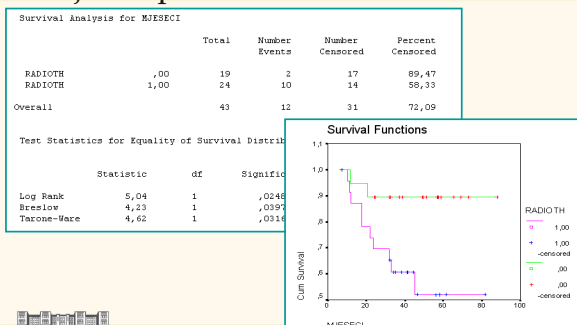
Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku

Dvije skupine, MedCalc®



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku

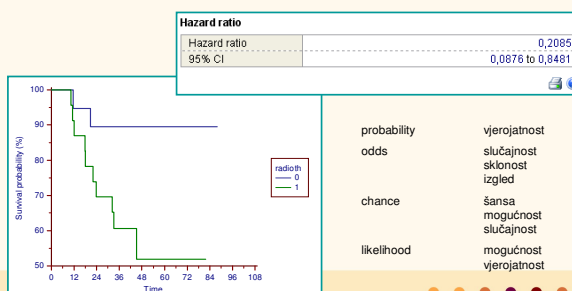
Dvije skupine, SPSS®



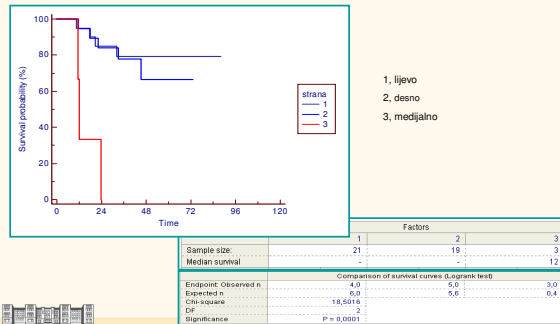
Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku

Omjer izgleda

- odds ratio, hazard ratio



Usporedba triju skupina



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku

Zaključivanje

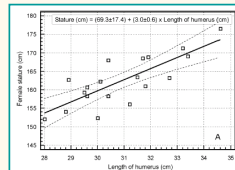
- granice pouzdanosti
- p-vrijednosti



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku

Regresijska analiza podataka

- korelacija
- pravocrtna regresija
- višestruka regresija
- logistička regresija
- Coxova regresija
- Passing-Bablokova regresija



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku

Coxov regresijski test

- multivarijantni postupak
- Cox regression
 - Cox, 1972.
 - proportional hazard model

10.5.3 Importance of the Cox Model

The Cox model is very useful in medicine, and it is easy to see why it is being used with increasing frequency. It provides the only valid method of predicting a time-dependent outcome, and many health-related outcomes are related to time. If the independent variables are divided into two categories (dichotomized), the exponential of the regression coefficient, $\exp(b)$, is the odds ratio, a useful way to interpret the risk associated with any specific factor. In addition, the Cox model provides a method for producing survival curves that are adjusted for confounding variables. The Cox model can be extended to the case of multiple events for a subject, but that



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku

Podatci &

- zavisni pokazatelji
 - vrijeme praćenja
 - podatak o ishodu (cenzuriranje)
- nezavisni pokazatelji
 - prediktori ili kovarijate (covariates)
 - sve mjerene ljestvice dopuštene
- rezultat
 - regresijski koeficijenti
 - ⇒ omjer rizika (hazard ratio (HR), ratio of the hazard function)
 - ⇒ mjera rizika (relativni rizik; relative risk (RR))

$$h_i(t) = h_0(t) \exp(\beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik})$$



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku

Primjer, MedCalc® Coxov regresijski test

- analiza preživljenja (smrt od osnovne bolesti)
- podaci – cenzurirani
- nezavisni pokazatelji
 - spol (M, Ž)
 - zahvaćena strana lica (L, D, M)
 - T-klasifikacija
 - resekcija donje čeljusti (1-5 kao nije, segmentalna, marginalna...)
 - liječenje zračenjem (da, ne)
 - najveći promjer tumora (cm)



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku

Primjer, MedCalc® “svi pokazatelji”

Survival time: mjeseci
Endpoint: cenzus
Method: Enter
Sample size: 43

Overall Model Fit

Null model -2 Log Likelihood	84,67522
Full model -2 Log Likelihood	73,62894
Chi-square	11,0463
DF	6
Significance level	P = 0,0870

Coefficients and Standard Errors

Covariate	b	SE	P	Exp(b)	95% CI of Exp(b)
spol	0,6557	1,2109	0,5801	1,9266	0,1817 to 20,4295
strana	1,1298	0,5442	0,0379	3,0945	1,0709 to 8,9427
T	-0,1311	0,5840	0,7848	0,8771	0,3283 to 2,3438
resmand	-0,2101	0,3444	0,5265	0,8040	0,4108 to 1,5736
raduoh	1,6251	0,8270	0,0484	5,0789	1,0138 to 25,4756
Tmnaprom	0,0919	0,3877	0,7652	1,0982	0,6015 to 1,9874

$$RR = \text{Exp}(b) \text{ ili } \text{Exp}(\beta) = e^b = 2,72^b$$

Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku

Primjer, MedCalc® “postupno biranje, unaprijed”

Survival time: mjeseci
Endpoint: cenzus
Method: Enter
Sample size: 43

Overall Model Fit

Null model -2 Log Likelihood	84,67522
Full model -2 Log Likelihood	79,19430
Chi-square	5,4809
DF	1
Significance level	P = 0,0192

Coefficients and Standard Errors

Covariate	b	SE	P	Exp(b)	95% CI of Exp(b)
spol	0,6557	1,2109	0,5801	1,9266	
strana	1,1298	0,5442	0,0379	3,0945	
T	-0,1311	0,5840	0,7848	0,8771	
resmand	-0,2101	0,3444	0,5265	0,8040	
raduoh	1,6251	0,8270	0,0484	5,0789	
Tmnaprom	0,0919	0,3877	0,7652	1,0982	

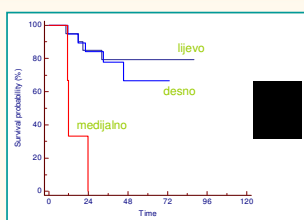
Overall Model Fit

Null model -2 Log Likelihood	84,67522
Full model -2 Log Likelihood	79,19430
Chi-square	5,4809
DF	1
Significance level	P = 0,0192

Coefficients and Standard Errors

Covariate	b	SE	P	Exp(b)	95% CI of Exp(b)
strana	1,1744	0,5045	0,0199	3,2381	1,2100 to 8,6543

Primjer, MedCalc® “postupno biranje, unaprijed”



Survival time: mjeseci
Endpoint: cenzus
Method: Forward
Enter variable if P<: 0,05
Remove variable if P>: 0,1
Sample size: 43

Overall Model Fit

Null model -2 Log Likelihood	84,67522
Full model -2 Log Likelihood	79,19430
Chi-square	5,4809
DF	1
Significance level	P = 0,0192

Coefficients and Standard Errors

Covariate	b	SE	P	Exp(b)	95% CI of Exp(b)
strana	1,1744	0,5045	0,0199	3,2381	1,2100 to 8,6543

Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku

mladenp@kdb.hr



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku