PDDS Molekularne bioznanosti
Počela biostatistike, akad. god. 2009/10.

Analiza cenzuriranih podataka i krivulje preživljenja

Mladen Petrovečki

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

Obrada podataka o preživljenju bolesnika

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

Analiza preživljenja

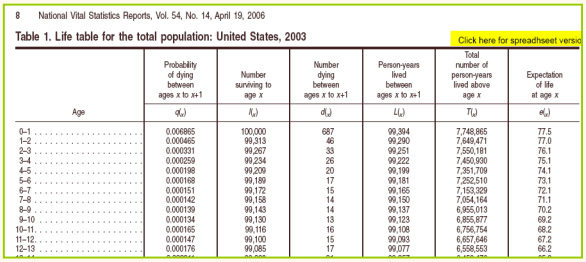
- Edmund Halley, 17. st
- engleski astronom, geofizičar, matematičar, meteorolog i fizičar
- http://en.wikipedia.org/wiki/Edmond_Halley

komet, 1986.
(sljedeći: 2061.)

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

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 spreadhseet versio

Age	Probability of dying between ages x to $x+1$ (q_x)	Number surviving to age x (l_x)	Number dying between ages x to $x+1$ (d_x)	Person-years lived between ages x to $x+1$ (L_x)	Total number of person-years lived above age x (T_x)	Expectation of life at age x (e_x)
0–1	0.006865	100,000	687	99,394	7,748,865	77.5
1–2	0.000465	99,313	46	99,290	7,649,471	77.0
2–3	0.000331	99,267	33	99,251	7,550,181	76.1
3–4	0.000259	99,234	26	99,222	7,450,930	75.1
4–5	0.000198	99,209	20	99,199	7,351,709	74.1
5–6	0.000168	99,189	17	99,181	7,252,510	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

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

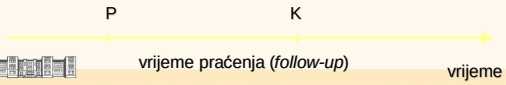
Analiza preživljenja

- psihijatrija – 1%
- patologija – 1%
- kirurgija – 12%
- onkologija – 14%
- izvorni znanstveni radovi u *The New England Journal od Medicine* – 32%

- podaci 1986.-2001., Dawson Saunders & Trapp, Basic and Clinical Biostatistics

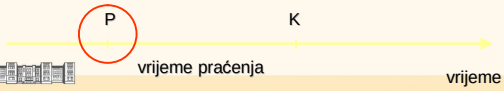
Analiza preživljenja

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



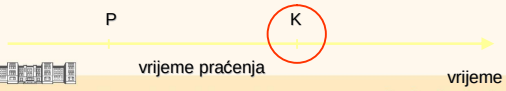
Početak praćenja

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



Kraj praćenja

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



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

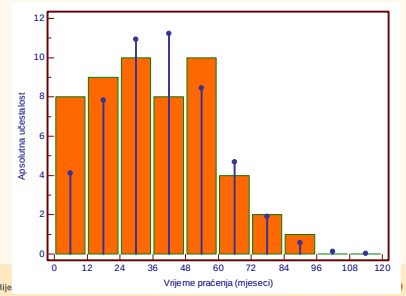
Vrijeme praćenja

- 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 probrani uzorak
    - dio populacije tog razdoblja
    - prva dg. karcinoma, bez regionalnih metastaza, itd.

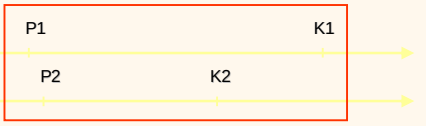
Vrijeme praćenja (1)

- raspodjela u pravilu nije simetrična



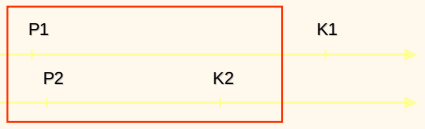
Vrijeme praćenja (2)

- potpuni podaci (potpuno praćenje)



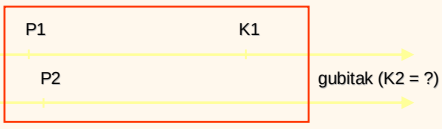
Vrijeme praćenja (2)

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



Vrijeme praćenja (3)

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



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

"Izgubljen iz praćenja"

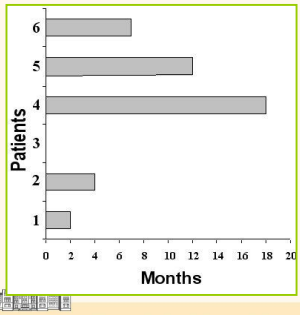
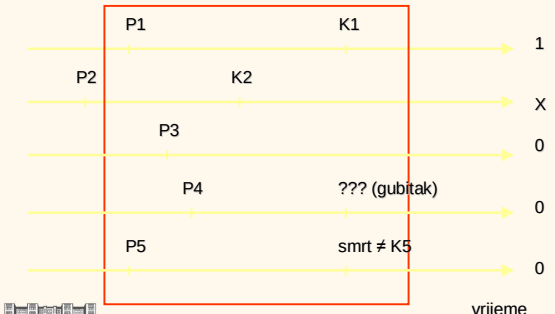


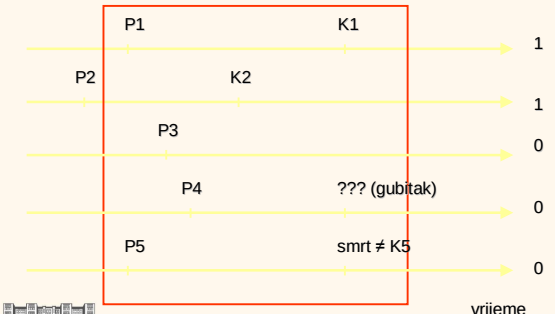
Figure 3. Outcome

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

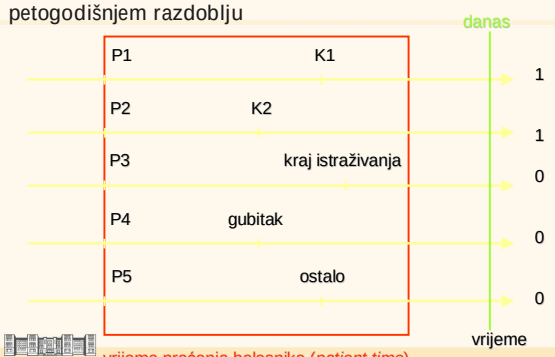
Cenzuriranje: bolesnici s postavljenom dijagnozom (P) u zadanih pet godina



Cenzuriranje: bolesnici liječeni u zadanom petogodišnjem razdoblju



Cenzuriranje: bolesnici liječeni u zadanom petogodišnjem razdoblju



Cenzuriranje

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

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 jedinka preživjeti ili točno doživjeti vrijeme od *t* jedinica praćenja, ili
- vjerojatnost preživljenja jedinke u rasponu od početka praćenja do trenutka praćenja *t*

A sada – još veće veselje!

$$h(t) = \lim_{\triangle t \to 0} \frac{R(t) - R(t + \triangle t)}{\triangle t \cdot R(t)}$$

$$F(t) = \int_0^t f(x)\,dx$$

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

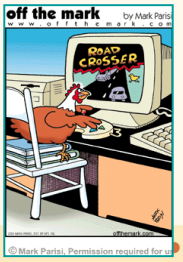
h(t):
- vjerojatnost da će jedinka umrijeti u trenutku *t*, uz uvjet da je preživjela do toga trenutka, i uz uvjet da je
- F(t) funkcija gustoće

Zaključak: podaci o preživljenju

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

Izračunavanje preživljenja

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



I. Tablice preživljenja

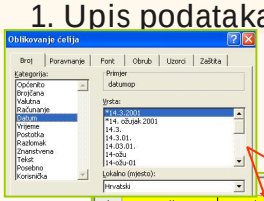
- osiguravateljske tablice
- tablice smrtnosti

vrijeme praćenja	vjerojatnost smrti q=d/(n-w/2)	vjerojatnost preživljenja p=1-q	kumulativna vjerojatnost preživljenja S(t)=∏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

Kako do preživljenja?

1. upis podataka
2. preuređenje podataka
3. izračun podataka

1. Upis podataka, Excel®



=(C2-B2)/365*12

	A	B	C	D	E
1	pacijent	datumop	datumkraj	cenzus	mjeseci
2	23456	23.6.2000	15.4.2007	0	81,8
3	24485	15.10.2003	8.11.2005	0	24,8
4	23080	25.7.2000	29.8.2004	0	49,2
5	23511	28.12.2001	15.2.2007	0	61,6
6	24188	20.2.2002	29.10.2004	0	32,3
7	22701	17.12.2003	8.6.2005	1	17,7
8	24241	17.7.2002	29.4.2007	0	57,4
9	23480	15.5.2003	20.8.2007	0	51,2
10	22823	5.10.2000	26.9.2002	1	23,7

2. Preuređenje podataka

	A	B	C	D	E
1	pacijent	datumop	datumkraj	cenzus	mjeseci
2	24485	15.10.2003	8.11.2005	0	24,8
3	23080	25.7.2000	29.8.2004	0	49,2
4	24188	20.2.2002	29.10.2004	0	32,3
5	22701	17.12.2003	8.6.2005	1	17,7
6	23998	27.4.2000	2.2.2004	1	45,2
7	24544	9.1.2002	29.9.2003	1	20,6
8	23859	10.10.2000	16.11.2003	0	37,2
9	22819	1.3.2001	6.2.2002	1	11,2
10	22921	26.4.2004	7.12.2004	0	7,4
11	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

3. Izračun podataka

vrijeme praćenja	vjerojatnost smrti q=d/(n-w/2)	vjerojatnost preživljenja p=1-q	kumulativna vjerojatnost preživljenja S(t)=∏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

3. Izračun podataka

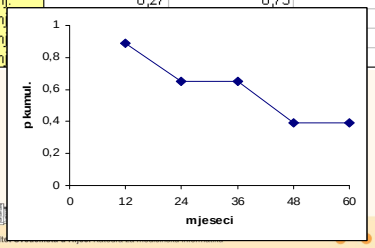
vrijeme praćenja	vjerojatnost smrti q=d/(n-w/2)	vjerojatnost preživljenja p=1-q	kumulativna vjerojatnost preživljenja S(t)=∏p
0-12 mj.	0,11	0,89	0,89
13-24 mj.	0,27	0,73	0,65
25-36 mj.	0	1	0,65
37-48 mj.	0,4	0,6	0,39
49-60 mj.	0	1	0,39

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

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

3. Izračun podataka

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



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

KM podaci o preživljenju

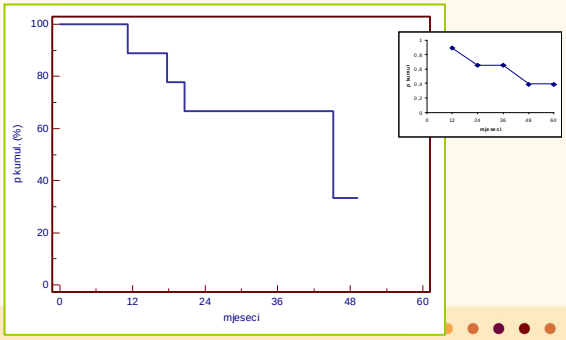
	A	B	C	D	E
1	pacijent	datumop	datumkraj	cenzus	mjeseci
2	24486	15.10.2003	8.11.2005	0	24,8
3	23880	25.7.2000	29.8.2004	0	49,2
4	24188	20.2.2002	29.10.2004	0	32,3
5	22701	17.12.2003	8.6.2005	1	17,7
6	23998	27.4.2000	2.2.2004	1	45,2
7	24544	9.1.2002	29.9.2003	1	20,6
8	23859	10.10.2000	16.11.2003	0	37,2
9	22819	1.3.2001	6.2.2002	1	11,2
10	22921	26.4.2004	7.12.2004	0	7,4
11	23309	9.7.2004	18.7.2006	0	24,3

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	-	-
24,8	-	-
32,3	-	-
37,2	-	-
45,2	0,333	0,248
49,2	-	-

MedCalc® Version 9.3.0.0
Windows 98/NT/Me/2000/XP/Vista
http://www.medcalc.be
Copyright © 1993-2007
Frank Schoonjans
This product is licensed to:
Mladen Petrovecki

KM krivulja preživljenja



Kaplan, Meier

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

NONPARAMETRIC ESTIMATION FROM INCOMPLETE OBSERVATIONS*

E. L. KAPLAN
University of California Radiation Laboratory
AND
PAUL MEIER
University of Chicago

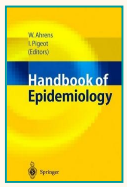
CONTENTS

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

Rizik umiranja

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

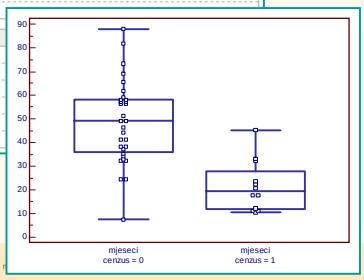


http://www.bips.uni-bremen.de/handbook_of_epidemiology.php

Primjer...

Sample 1	
Variable	mjeseci
Select	cenzus = 0
Sample 2	
Variable	mjeseci
Select	cenzus = 1

Sample size		
Arithmetic mean		
95% CI for the mean		
Variance		
Standard deviation		
Standard error of the mean		



Preživljenje, MedCalc®

Survival time	mjeseci
Endpoint	cenzus
Sample size	43
Median survival	-

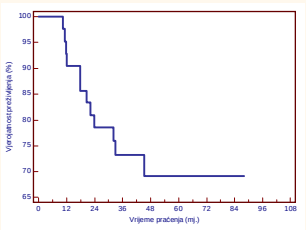
Survival time	Survival Proportion	Standard Error
7,3973	-	-
10,3562	0,976	0,024
11,2438	0,952	0,033
11,6384	0,929	0,040
12	0,905	0,045
17,7205	0,881	0,050
17,8849	0,857	0,054
20,6466	0,833	0,058
22,1918	0,810	0,061
23,7041	0,786	0,063
24,2959	-	-
24,8219	-	-
32,1863	0,760	0,066
45,2384	0,692	0,077
46,3233	-	-
49,1836	-	-
49,4466	-	-
51,2219	-	-
56,3507	-	-
56,9096	-	-
57,3041	-	-
57,4356	-	-
58,0274	-	-
58,126	-	-
58,9479	-	-
61,6438	-	-
65,6219	-	-
69,074	-	-
73,2164	-	-
81,7644	-	-
88,0438	-	-

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

Krivulja preživljenja, 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
11,6384	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

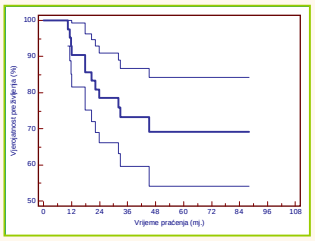


npr., prosječno petogodišnje preživljenje: 69,2±7,7%

Granice pouzdanosti

Survival time	mjeseci
Endpoint	cenzus
Sample size	43
Median survival	-

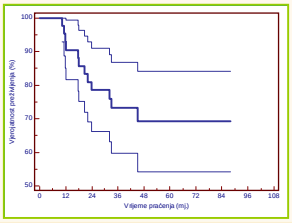
Survival time	Survival Proportion	Standard Error
7,3973	-	-
10,3562	0,976	0,024
11,2438	0,952	0,033
11,6384	0,929	0,040
12	0,905	0,045
17,7205	0,881	0,050
17,8849	0,857	0,054
20,6466	0,833	0,058
22,1918	0,810	0,061
23,7041	0,786	0,063
24,2959	-	-
24,8219	-	-
32,1963	0,760	0,066



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

Izračun granica pouzdanosti

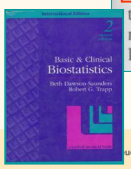
- granice pouzdanosti (*Confidence Intervals*)
- CI = x ± z SE(x)
  - 95%CI = x ± 1,96 SE(x)
  - 99%CI = x ± 2,56 SE(x)



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

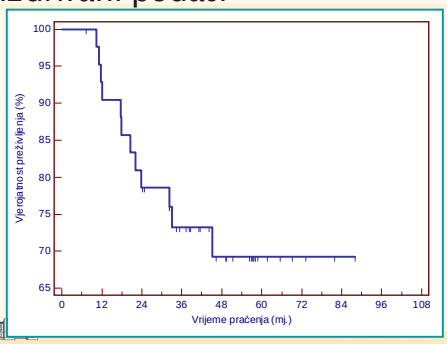
http://www.fourmilab.ch/rpkp/experiments/analysis/zCalc.html (pazi: p/2!)

Granice pouzdanosti

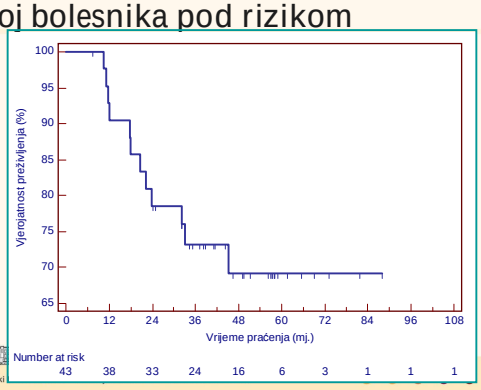


Results from life table analysis are usually presented in a survival curve rather than in tables. The solid line in Fig 11–3 is a survival curve for the kidney transplant data. The dashed lines on either side of the survival curve represent 95% **confidence bands** for the curve. Although confidence 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

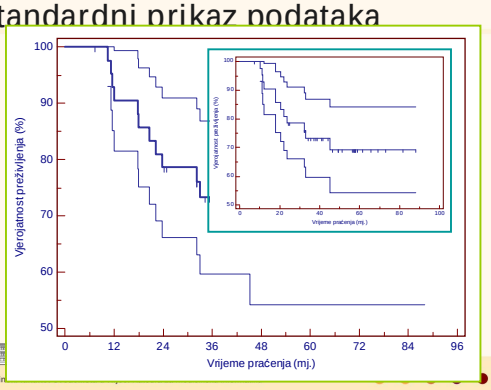
Cenzurirani podaci



Broj bolesnika pod rizikom



Standardni prikaz podataka

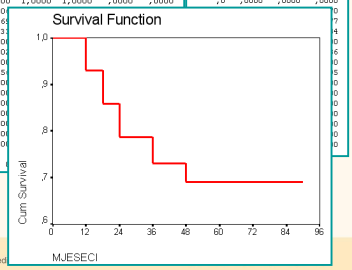


Tablice preživljenja, SPSS®

Life Table
Survival Variable MJESECI

Intrvl Start Time	Number Entrng this Intrvl	Number Wdrawn During Intrvl	Number Exposd to Risk	Number of Termnl Events	Propn Termi-nating	Propn Sur-viving	Cumul Propn Surv at End	Proba-bility Densty	Hazard Rate	SE of Cumul Sur-viving	SE of Proba-bility Densty	SE of Hazard Rate
,0	43,0	,0	43,0	,0	,0000	1,0000	1,0000	,0000	,0000	,0000	,0000	,0000
6,0	43,0	1,0	42,5	3,0	,0706							
12,0	39,0	,0	39,0	3,0	,0769							
18,0	36,0	,0	36,0	3,0	,0833							
24,0	33,0	2,0	32,0	,0	,0000							
30,0	31,0	5,0	28,5	2,0	,0702							
36,0	24,0	5,0	21,5	,0	,0000							
42,0	19,0	2,0	18,0	1,0	,0556							
48,0	16,0	3,0	14,5	,0	,0000							
54,0	13,0	7,0	9,5	,0	,0000							
60,0	6,0	2,0	5,0	,0	,0000							
66,0	4,0	1,0	3,5	,0	,0000							
72,0	3,0	1,0	2,5	,0	,0000							
78,0	2,0	1,0	1,5	,0	,0000							
84,0	1,0	1,0	,5	,0	,0000							

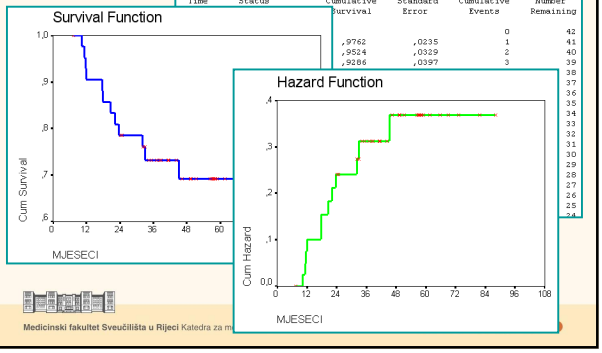
The median survival time for these data is



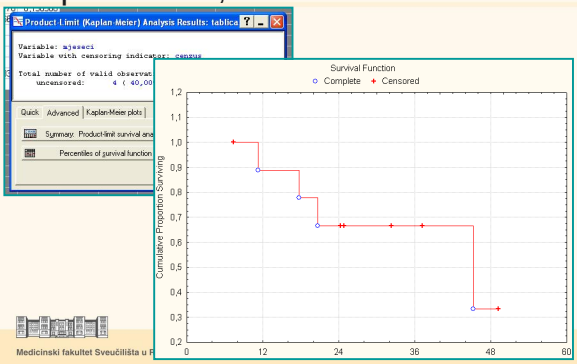
Kaplan-Meier, SPSS®

Survival Analysis for MJESECI

Time	Status	Cumulative Survival	Standard Error	Cumulative Events	Number Remaining
				0	42
		,9762	,0235	1	41
		,9524	,0329	2	40
		,9286	,0397	3	39



Kaplan-Meier, Statistica®



"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

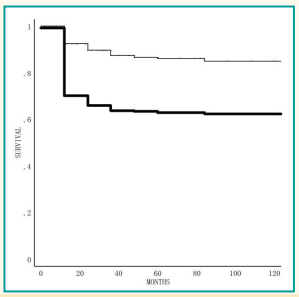
DFS, primjer

Survival curves for 277 patients with T_2 tongue carcinoma treated with an ^{192}Ir, ^{137}Cs or ^{226}Ra needle;

thin – regional disease-free survival

thick – distant metastasis-free survival

http://www.ro-journal.com/content/1/1/21/figure/F5



Usporedba podataka o preživljenju

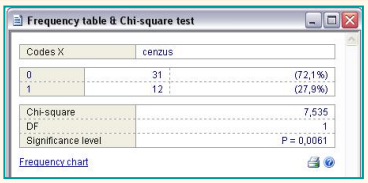
- usporedba dvije skupine podataka
  - *log-rank* (*logrank*) test
    - Mentelov ili Mentel-Coxov test
  - Wilcoxonov test
    - generalizirani Wilcoxonov test
    - Gehanov test
    - Gehan-Breslowljev test
    - opći Kruskal-Wallisov test za cenzurirane podatke
  - Mantel-Haenszelov test
  - Tarone-Wareov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

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-

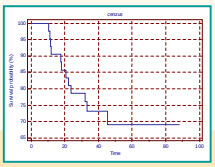
Primjer

- N = 43

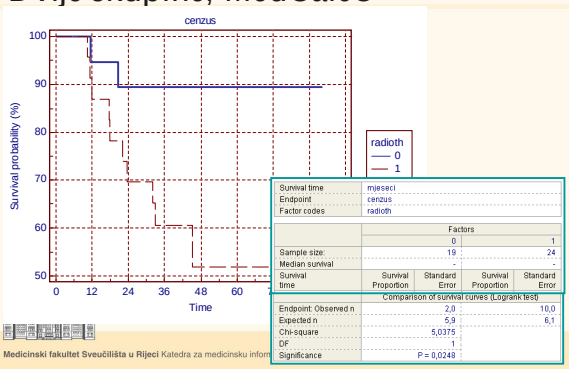


Variable	mjeseci
Select	(cenzus = 1) or (cenzus = 0)

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



Dvije skupine, MedCalc®



Survival time	mjeseci
Endpoint	cenzus
Factor codes	radioth

	Factors	
	0	1
Sample size:	19	24
Median survival	-	-

Survival time	Survival Proportion	Standard Error	Survival Proportion	Standard Error

Comparison of survival curves (Logrank test)

Endpoint: Observed n	2,0	10,0
Expected n	5,9	6,1
Chi-square	5,0375	
DF	1	
Significance	P = 0,0248	

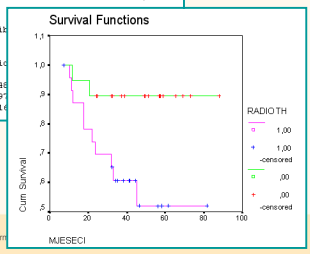
Dvije skupine, SPSS®

Survival Analysis for MJESECI

	Total	Number Events	Number Censored	Percent Censored
RADIOTH ,00	19	2	17	89,47
RADIOTH 1,00	24	10	14	58,33
Overall	43	12	31	72,09

Test Statistics for Equality of Survival Distrib

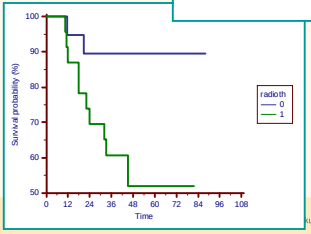
Statistic	df	Signific
Log Rank 5,04	1	,0248
Breslow 4,23	1	,0397
Tarone-Ware 4,62	1	,0316



Omjer izgleda

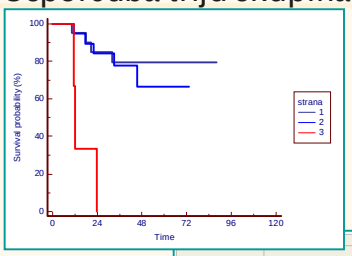
- *odds ratio, hazard ratio*

Hazard ratio	
Hazard ratio	0,2085
95% CI	0,0876 to 0,8481



probability	vjerojatnost
odds	slučajnost sklonost izgled
chance	šansa mogućnost slučajnost
likelihood	mogućnost vjerojatnost

Usporedba triju skupina



1, lijevo
2, desno
3, medijalno

	Factors		
	1	2	3
Sample size	21	19	3
Median survival	-	-	12

Comparison of survival curves (Logrank test)

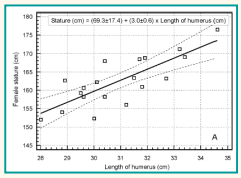
Endpoint: Observed n	4,0	5,0	3,0
Expected n	6,0	5,6	0,4
Chi-square	18,5016		
DF	2		
Significance	P = 0,0001		

Zaključivanje

- granice pouzdanosti
- p-vrijednosti

Regresijska analiza podataka

- korelacija



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

Coxov regresijski test

- multivarijatni 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

Podaci &

- 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)*)

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

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)

Primjer, MedCalc®
"svi pokazatelji"

Cox proportional-hazards regression

Survival time	mjeseci
Endpoint	cenzus
Method	Enter
Sample size	43

Overall Model Fit

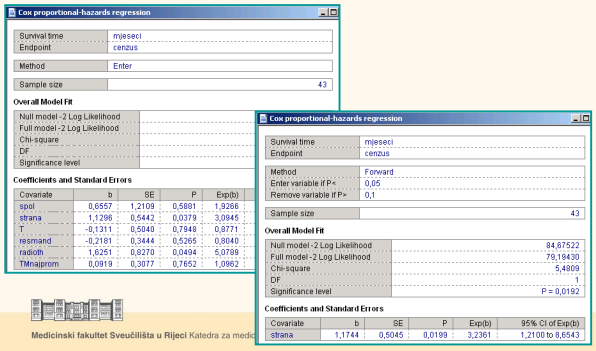
Null model -2 Log Likelihood	84,67622
Full model -2 Log Likelihood	73,62894
Chi-square	11,0463
DF	6
Significance level	P = 0,0870

Coefficients and Standard Errors

Covariate	b	SE	P	Exp(b)	95% CI of Exp(b)
spol	0,6557	1,2109	0,5881	1,9266	0,1817 to 20,4295
strana	1,1296	0,5442	0,0379	3,0945	1,0708 to 8,9427
T	-0,1311	0,5040	0,7948	0,8771	0,3283 to 2,3438
resmand	-0,2181	0,3444	0,5265	0,8040	0,4108 to 1,5738
radioth	1,6251	0,8270	0,0494	5,0789	1,0126 to 25,4756
TMnajprom	0,0919	0,3077	0,7652	1,0962	0,6016 to 1,9974

RR = Exp(b) ili Exp(β) = e^b = $2{,}72^b$

Primjer, MedCalc®
“postupno biranje, unaprijed”



Cox proportional-hazards regression

Survival time	mjeseci
Endpoint	cenzus
Method	Enter
Sample size	43

Overall Model Fit

Null model -2 Log Likelihood	
Full model -2 Log Likelihood	
Chi-square	
DF	
Significance level	

Coefficients and Standard Errors

Covariate	b	SE	P	Exp(b)
spol	0,6557	1,2109	0,5891	1,9266
strana	1,1296	0,5442	0,0379	3,0945
T	-0,1311	0,5040	0,7948	0,8771
resmand	-0,2181	0,3444	0,5265	0,8040
radioth	1,6251	0,8270	0,0494	5,0799
TMnajprom	0,0919	0,3077	0,7652	1,0962

Cox proportional-hazards regression

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

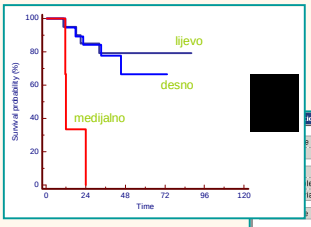
Overall Model Fit

Null model -2 Log Likelihood	84,87522
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	3,2361	1,2100 to 8,6543

Primjer, MedCalc®
“postupno biranje, unaprijed”



Cox proportional-hazards regression

Survival time	mjeseci
Endpoint	cenzus
Method	Enter
Sample size	43

Overall Model Fit

Null model -2 Log Likelihood	
Full model -2 Log Likelihood	
Chi-square	
DF	
Significance level	

Coefficients and Standard Errors

Covariate	b	SE	P	Exp(b)
spol	0,6557	1,2109	0,5891	1,9266
strana	1,1296	0,5442	0,0379	3,0945
T	-0,1311	0,5040	0,7948	0,8771
resmand	-0,2181	0,3444	0,5265	0,8040
radioth	1,6251	0,8270	0,0494	5,0799
TMnajprom	0,0919	0,3077	0,7652	1,0962

Cox proportional-hazards regression

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	84,87522
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	3,2361	1,2100 to 8,6543

mladenp@kbd.hr