Uvod v analizo preživetja

Janez Stare

Medicinska fakulteta, Ljubljana, Slovenija

Druga poimenovanja Analize preživetja (Survival Analysis)

- Failure Time Data Analysis
- Reliability Analysis
- Event History Analysis (Analiza zgodovine dogodkov)

Kdaj uporabljamo metode analize preživetja?

Kdaj uporabljamo metode analize preživetja?

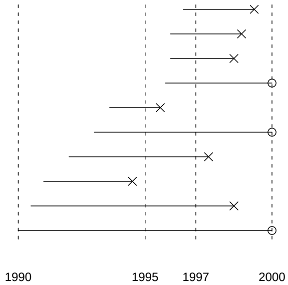
Kadar je izid, ki nas zanima, doba preživetja (ali čas do kakšnega drugega dogodka).

Kdaj uporabljamo metode analize preživetja?

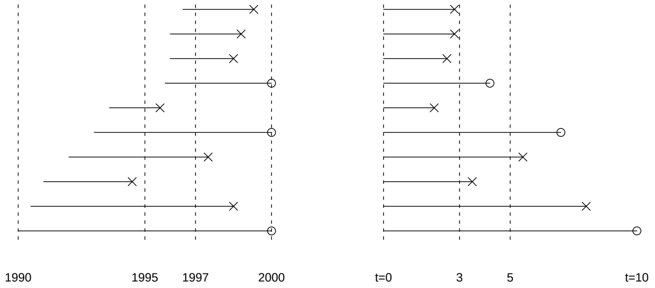
Kadar je izid, ki nas zanima, doba preživetja (ali čas do kakšnega drugega dogodka).

To je sicer precej očiten odgovor, toda **zakaj potrebujemo posebne metode**?

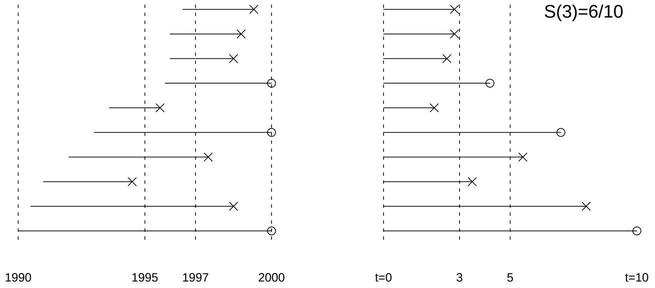
Tipičen primer



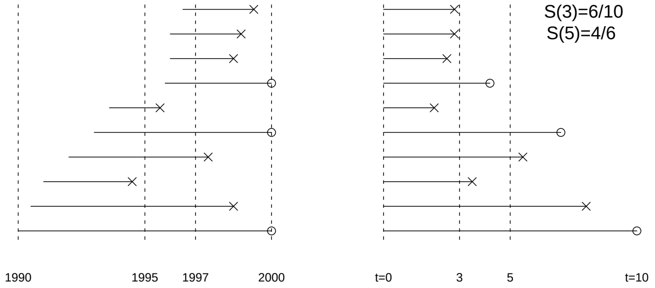
Tipičen primer



Tipičen primer



Tipičen primer



Posebne metode potrebujemo zaradi **krnjenja** (censoring). Za krnjenje obstajajo različni razlogi:

- oseb ne moremo več spremljati (lost to follow up)
- smrt zaradi drugih razlogov
- konec študije (najbolj pogosto)

Z okrnjenimi podatki ne moremo izračunati niti aritmetične sredine ali narisati histograma.

Z okrnjenimi podatki ne moremo izračunati niti aritmetične sredine ali narisati histograma.

Stvar se zdi precej brezupna.

Z okrnjenimi podatki ne moremo izračunati niti aritmetične sredine ali narisati histograma.

Stvar se zdi precej brezupna.

Na srečo ni, čeprav je trajalo kar nekaj časa, da so se razvile metode, ki naredijo, kar želimo.

Z okrnjenimi podatki ne moremo izračunati niti aritmetične sredine ali narisati histograma.

Stvar se zdi precej brezupna.

Na srečo ni, čeprav je trajalo kar nekaj časa, da so se razvile metode, ki naredijo, kar želimo.

Kaj pa želimo?

Cilji analize preživetja (oziroma analize nasploh)

Cilji analize preživetja (oziroma analize nasploh)

1. Ocena porazdelitvene funkcije (funkcije preživetja).

Cilji analize preživetja (oziroma analize nasploh)

1. Ocena porazdelitvene funkcije (funkcije preživetja).
2. Primerjava porazdelitev (funkcij preživetja).

Cilji analize preživetja (oziroma analize nasploh)

1. Ocena porazdelitvene funkcije (funkcije preživetja).
2. Primerjava porazdelitev (funkcij preživetja).
3. Iskanje povezave med izidom (časom preživetja) in napovednimi dejavniki.

Funkcija preživetja

Formalno:

Če je T zvezna nenegativna slučajna spremenljivka z gostoto $f(t)$, potem je njena funkcija preživetja

$$S(t) = P(T > t) = 1 - F(t) = \int_t^\infty f(x)dx,$$

Funkcija preživetja

Formalno:

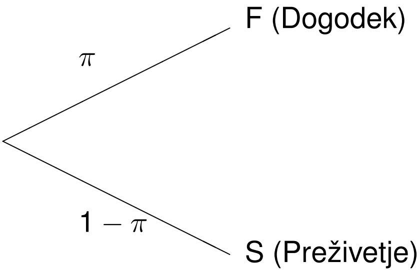
Če je T zvezna nenegativna slučajna spremenljivka z gostoto $f(t)$, potem je njena funkcija preživetja

$$S(t) = P(T > t) = 1 - F(t) = \int_t^\infty f(x)dx,$$

To pomeni:

Vrednost funkcije preživetja v danem času t je **delež ljudi, ki so še živi** v tem času.

Ocenjevanje funkcije preživetja



Ocenjevanje funkcije preživetja

Spomnimo se sedaj formule za verjetnost produkta dogodkov.

Ocenjevanje funkcije preživetja

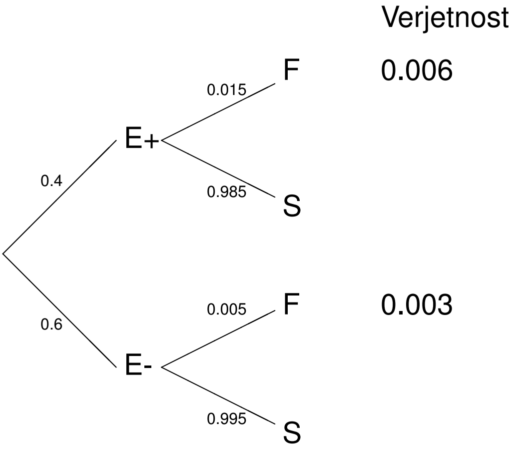
Spomnimo se sedaj formule za verjetnost produkta dogodkov.

Če sta A in B dogodka, potem je verjetnost produkta AB

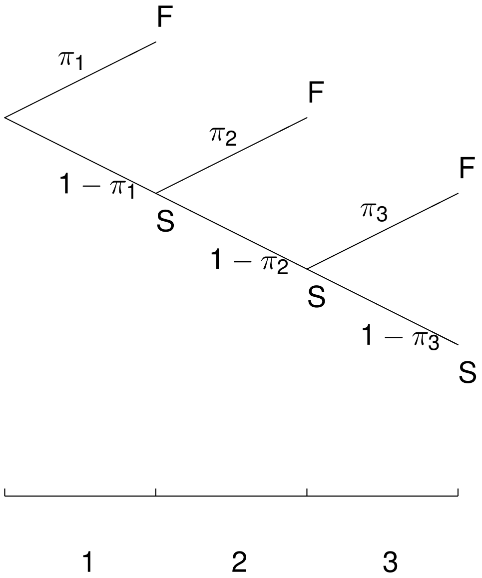
$$P(AB) = P(A)P(B|A)$$

kjer je $P(B|A)$ pogojna verjetnost B pri pogoju A.

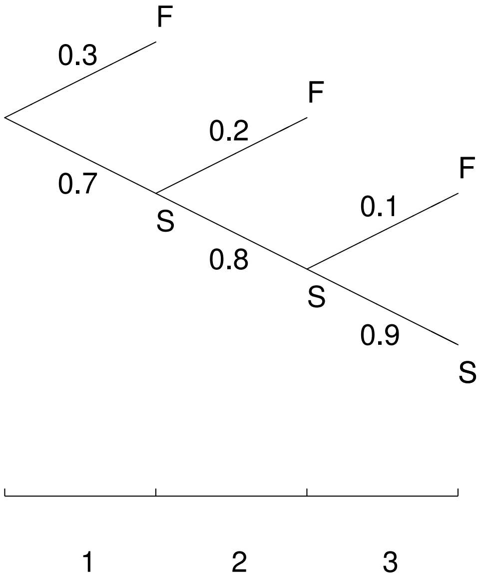
Ocenjevanje funkcije preživetja



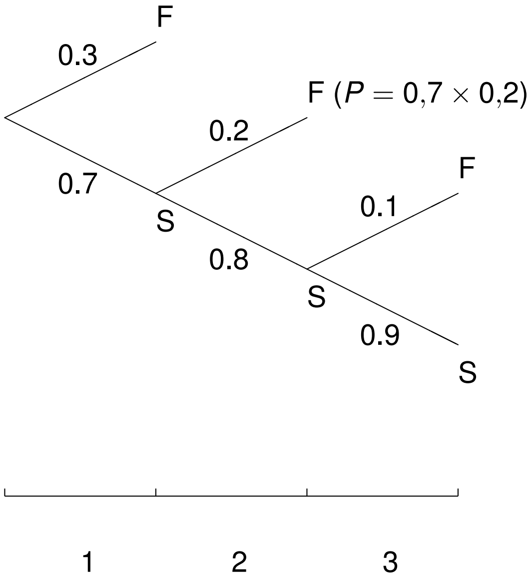
Ocenjevanje funkcije preživetja



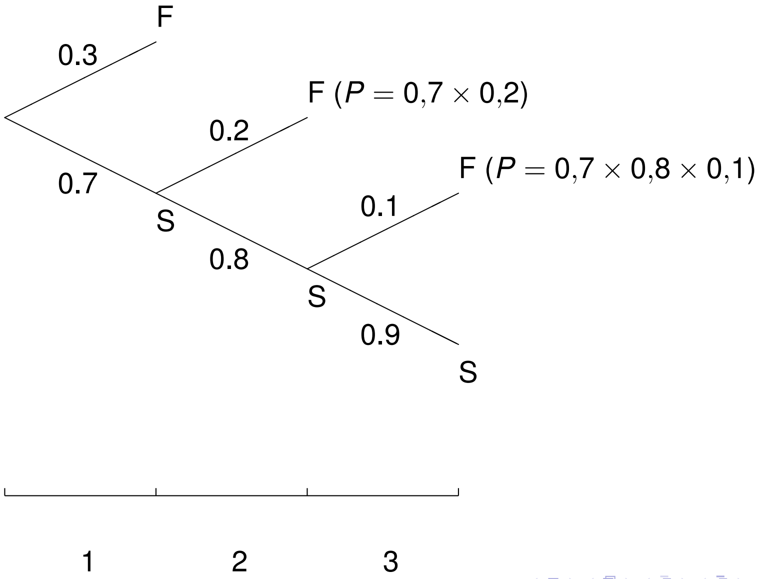
Ocenjevanje funkcije preživetja



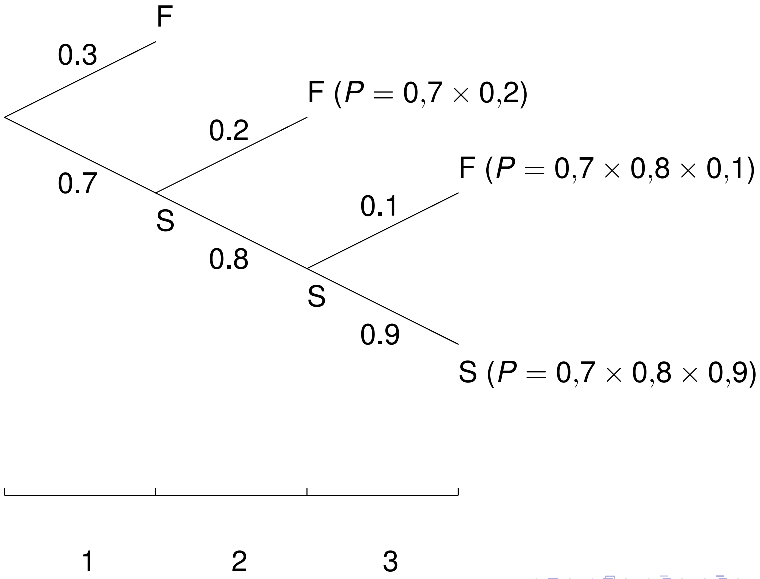
Ocenjevanje funkcije preživetja



Ocenjevanje funkcije preživetja



Ocenjevanje funkcije preživetja



Ocenjevanje funkcije preživetja

Ta princip lahko uporabimo tudi, če imamo **okrnjene podatke**.

Ocenjevanje funkcije preživetja

Ta princip lahko uporabimo tudi, če imamo **okrnjene podatke**.

Časovno lestvico najprej razdelimo na intervale tako, da so časi dogodkov ali krnitev na mejah intervalov.

Ocenjevanje funkcije preživetja

Ta princip lahko uporabimo tudi, če imamo **okrnjene podatke**.

Časovno lestvico najprej razdelimo na intervale tako, da so časi dogodkov ali krnitev na mejah intervalov.

Potem izračunamo (pogojne) verjetnosti preživetja v vsakem intervalu in iz njih dobimo verjetnost preživetja za poljubno obdobje z množenjem pogojnih verjetnosti do danega trenutka.

Ocenjevanje funkcije preživetja

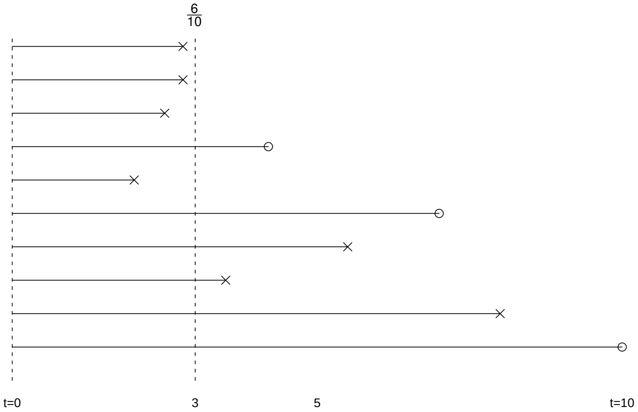
Ta princip lahko uporabimo tudi, če imamo **okrnjene podatke**.

Časovno lestvico najprej razdelimo na intervale tako, da so časi dogodkov ali krnitev na mejah intervalov.

Potem izračunamo (pogojne) verjetnosti preživetja v vsakem intervalu in iz njih dobimo verjetnost preživetja za poljubno obdobje z množenjem pogojnih verjetnosti do danega trenutka.

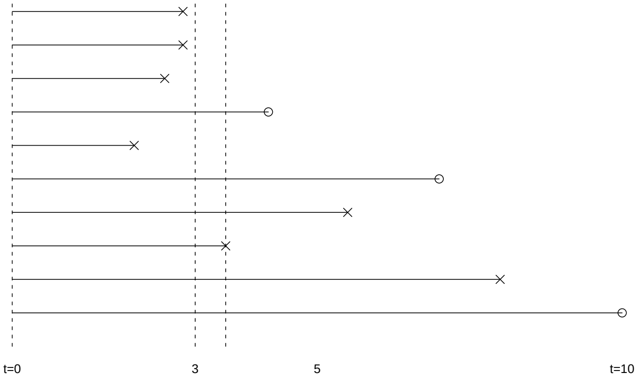
Metodo imenujemo **Kaplan - Meierjeva metoda**.

Kaplan-Meierjeva metoda



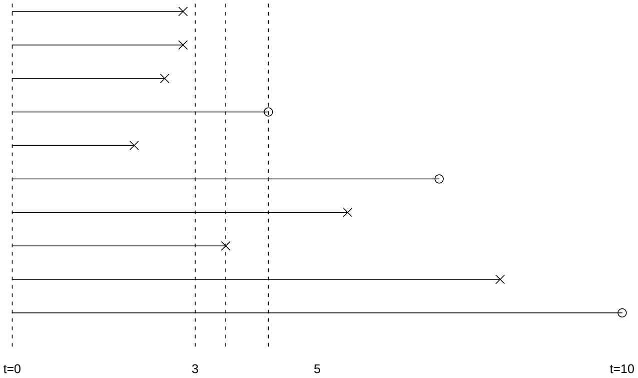
Kaplan-Meierjeva metoda

$$\frac{6}{10} \cdot \frac{5}{6} = \frac{5}{10}$$



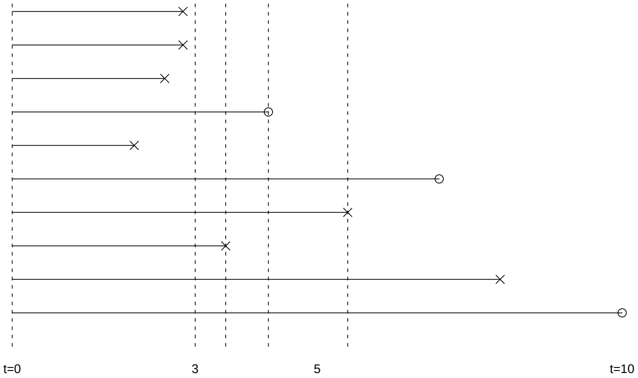
Kaplan-Meierjeva metoda

$$\frac{5}{10} \cdot \frac{5}{5} = \frac{5}{10}$$

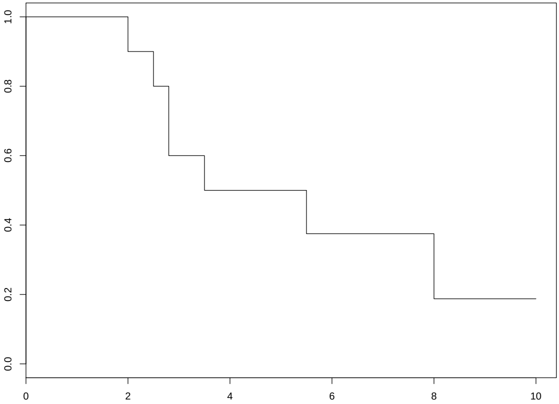


Kaplan-Meierjeva metoda

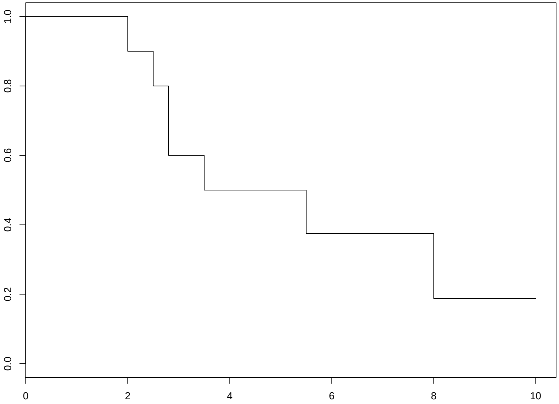
$$\frac{5}{10} \cdot \frac{3}{4} = \frac{3}{8}$$



Kaplan - Meierjeva krivulja za naš primer



Kaplan - Meierjeva krivulja za naš primer



Kaj pomenijo vodoravni deli krivulje?

Še en primer

Podatki

Čas	Izpostavljeni
55	12
61+	11
74	10
81	9
93+	8
122+	7
138	6
151	5
168	4
202+	3
220+	2
238	1

Še en primer

Podatki

Čas	Izpostavljeni
55	12
61+	11
74	10
81	9
93+	8
122+	7
138	6
151	5
168	4
202+	3
220+	2
238	1

Računanje

$$\hat{S}(55) = \frac{11}{12} = 0{,}917$$

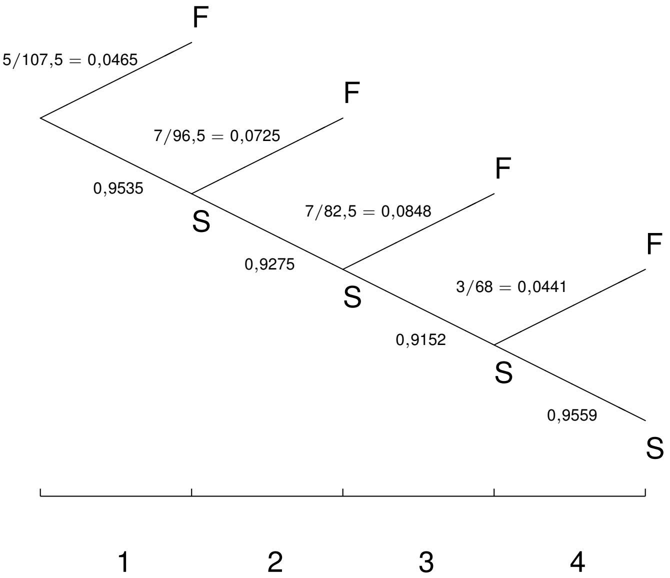
$$\hat{S}(61) = \frac{11}{12} \cdot \frac{11}{11} = 0{,}917$$

$$\hat{S}(74) = \frac{11}{12} \cdot \frac{9}{10} = 0{,}825$$

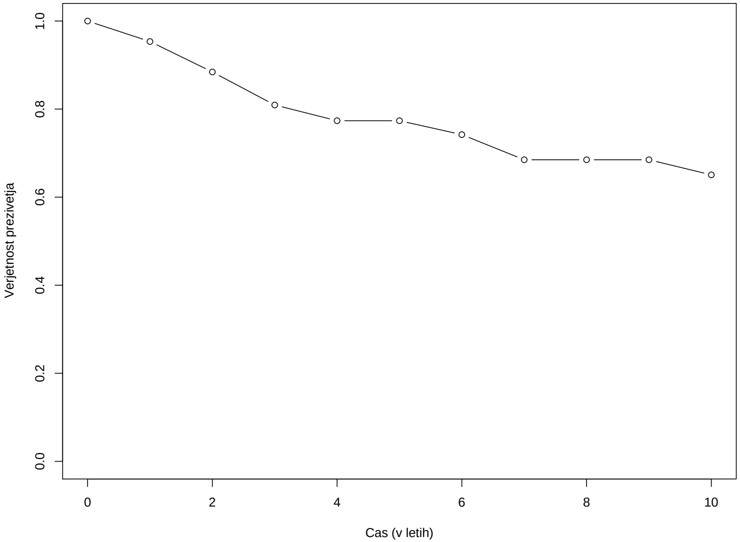
Življenjske tabele

Leto	N	F	L
1	110	5	5
2	100	7	7
3	86	7	7
4	72	3	8
5	61	0	7
6	54	2	10
7	42	3	6
8	33	0	5
9	28	0	4
10	24	1	8

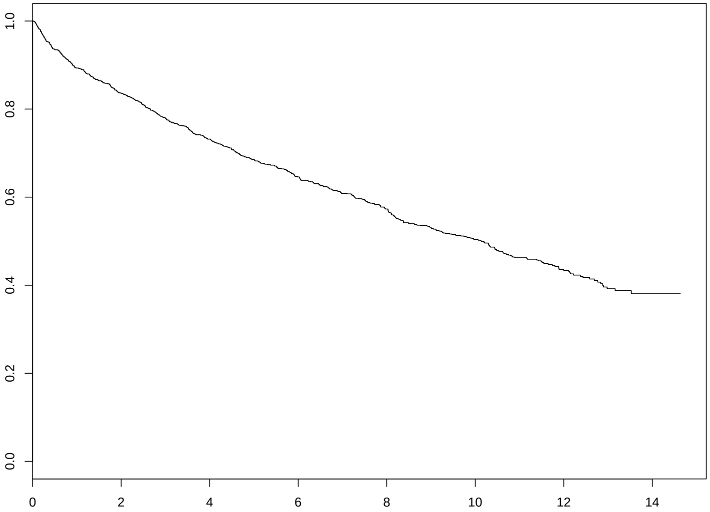
Življenjske tabele



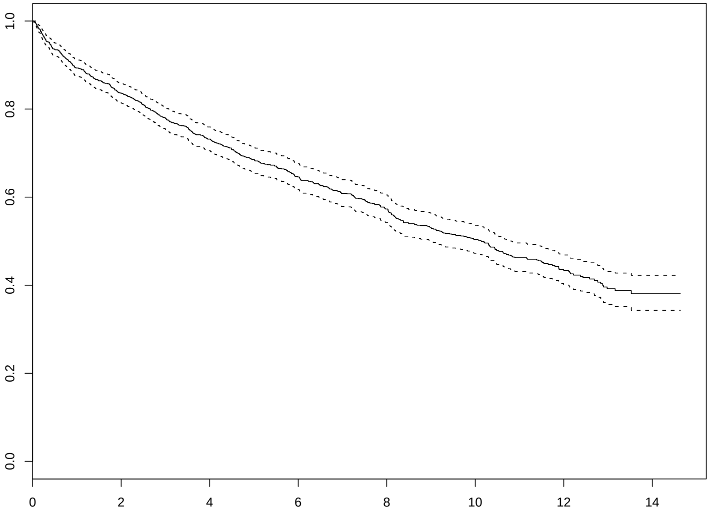
Risanje krivulj preživetja iz tabel preživetja



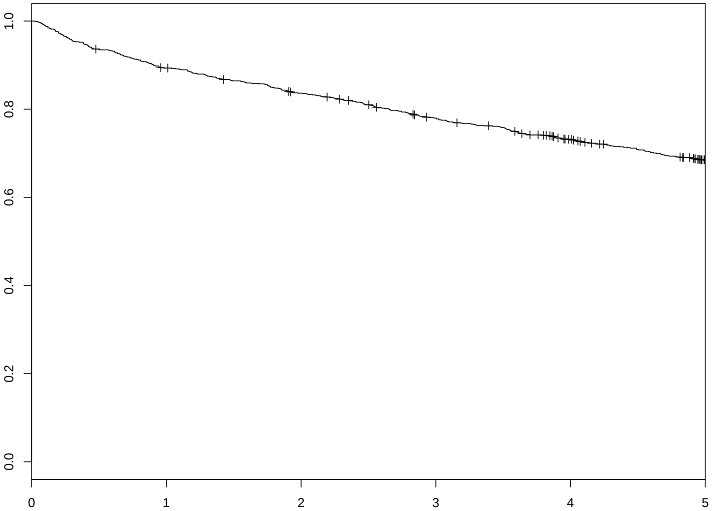
Primer - preživetje po miokardnem infarktu



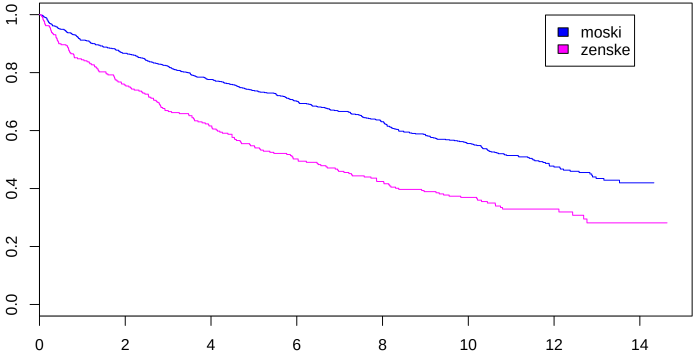
Primer - preživetje po miokardnem infarktu



Primer - preživetje po miokardnem infarktu



Primerjava krivulj preživetja



Statistični test za ničelno hipotezo (da vzorca prihajata iz iste populacije) temelji na običajni ideji:

Statistični test za ničelno hipotezo (da vzorca prihajata iz iste populacije) temelji na običajni ideji:

Če velja ničelna hipoteza, pričakujemo, da bo število dogodkov v posamezni skupini sorazmerno številu enot v skupini.

Statistični test za ničelno hipotezo (da vzorca prihajata iz iste populacije) temelji na običajni ideji:

Če velja ničelna hipoteza, pričakujemo, da bo število dogodkov v posamezni skupini sorazmerno številu enot v skupini.

Na osnovi te predpostavke izračunamo **pričakovano število dogodkov** v vsaki skupini in ga primerjamo z **ugotovljenim številom dogodkov**.

Statistični test za ničelno hipotezo (da vzorca prihajata iz iste populacije) temelji na običajni ideji:

Če velja ničelna hipoteza, pričakujemo, da bo število dogodkov v posamezni skupini sorazmerno številu enot v skupini.

Na osnovi te predpostavke izračunamo **pričakovano število dogodkov** v vsaki skupini in ga primerjamo z **ugotovljenim številom dogodkov**.

Test se imenuje **test log rank**.

Statistični test za ničelno hipotezo (da vzorca prihajata iz iste populacije) temelji na običajni ideji:

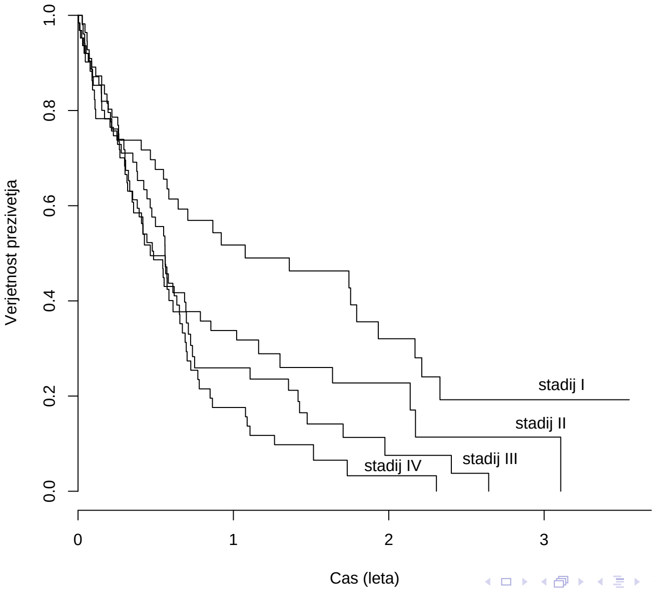
Če velja ničelna hipoteza, pričakujemo, da bo število dogodkov v posamezni skupini sorazmerno številu enot v skupini.

Na osnovi te predpostavke izračunamo **pričakovano število dogodkov** v vsaki skupini in ga primerjamo z **ugotovljenim številom dogodkov**.

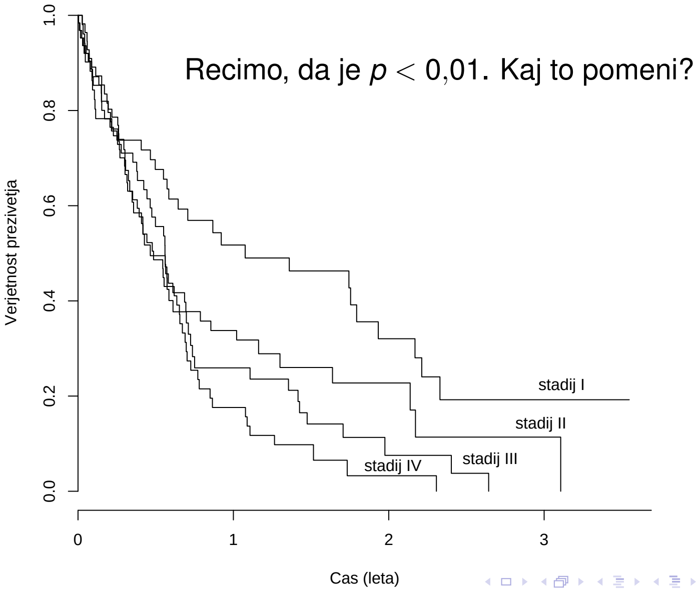
Test se imenuje **test log rank**.

Vrednost p za test log rank za prejšnji primer je $3{,}1 \cdot 10^{-9}$.

Test log rank



Test log rank



Kako pripravimo podatke

Datum začetka	Datum konca	Stanje (dogodek, ni dogodka)

ali

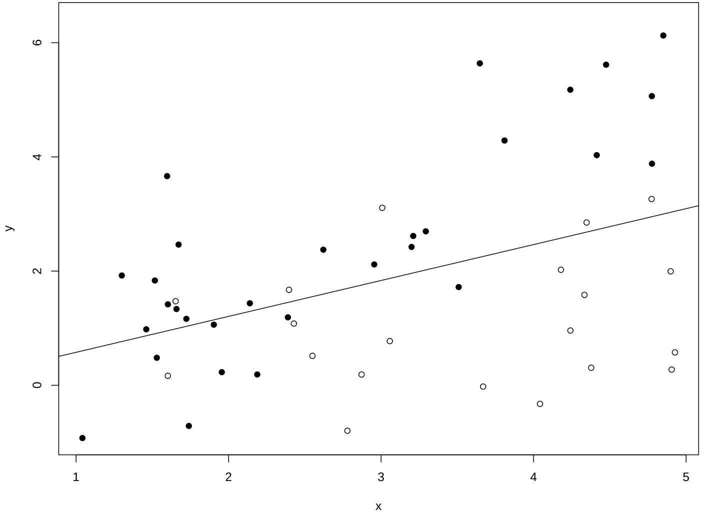
Čas preživetja	Stanje (dogodek, ni dogodka)

Regresijski modeli

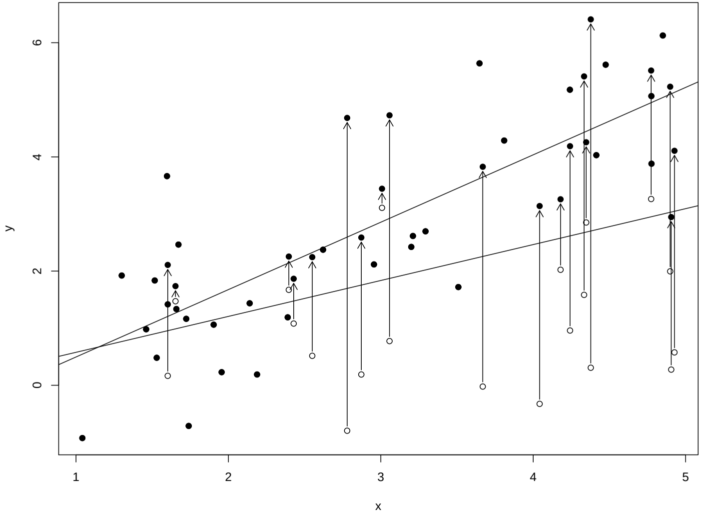
Linearni regresijski model pravi

$$Y \sim \mathcal{N}(\alpha + \sum \beta_i X_i, \sigma^2)$$

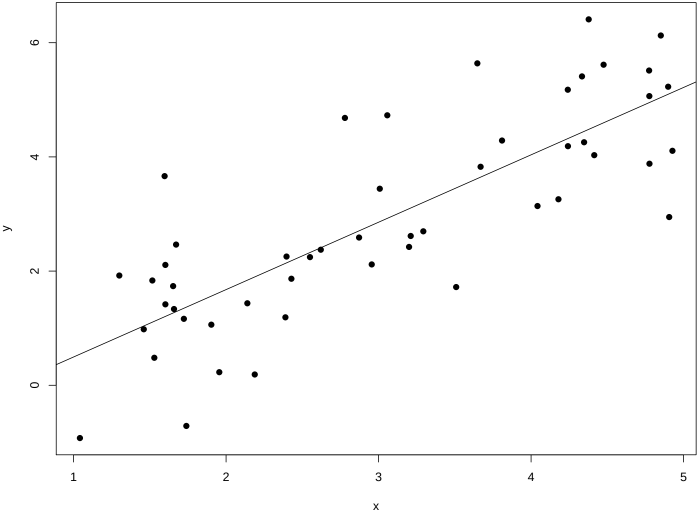
Model povezuje vrednosti spremenljivke Y z vrednostmi spremenljivk X_i. Na okrnjenih podatkih tega ne moremo narediti.



y
x
0
2
4
6
1
2
3
4
5



y
0
2
4
6
x
1
2
3
4
5



y
0
2
4
6
1
2
3
4
5
x

Funkcija ogroženosti

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

Funkcija ogroženosti

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

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

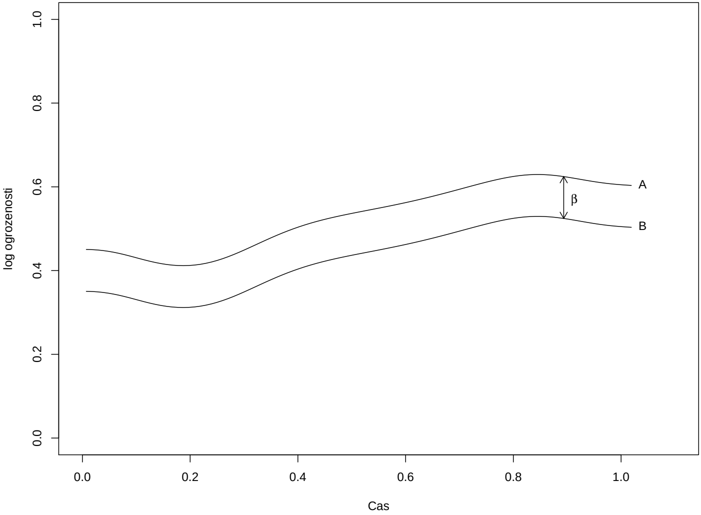
Coxov model

$$h(t,x) = h_0(t)e^{\beta x}$$

$$\frac{h_1(t,x_1)}{h_2(t,x_2)} = e^{\beta(x_1 - x_2)}$$

$$\frac{h(t,x+1)}{h(t,x)} = e^{\beta}$$

Coxov model pogosto imenujemo **model sorazmernih tveganj (proportional hazards model).**


1.0
0.8
0.6
0.4
0.2
0.0
log ogrozenosti
0.0
0.2
0.4
0.6
0.8
1.0
Cas
A
B
β

Tipičen opis metod analize preživetja

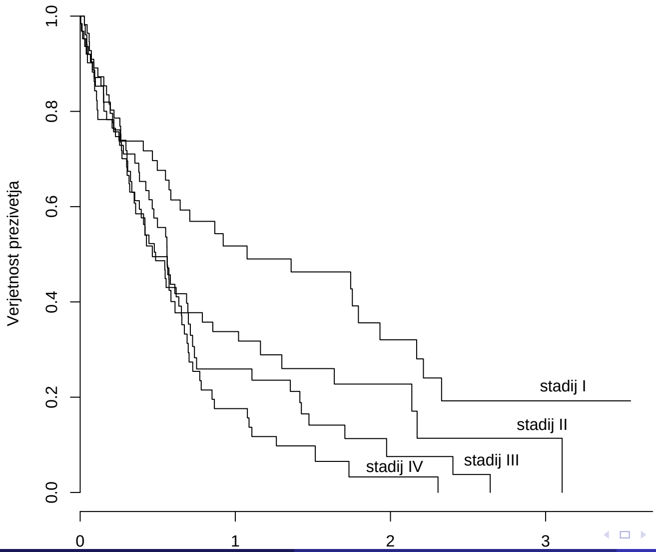
Survival curves were constructed with the Kaplan-Meier method and compared with the log-rank test. Analyses requiring adjustments for potential confounding factors were conducted using the Cox proportional hazards method. The proportional hazards assumption was tested and satisfied for each mathematical model using Cox analysis.

Primer: Coxov model na podatkih o miokardnem infarktu

	coef	exp(coef)	se(coef)	z	p
starost	0.056	1.057	0.004	12.554	0.000
spol	0.004	1.004	0.102	0.036	0.970
leto	-0.081	0.922	0.035	-2.295	0.022
diabetes	0.488	1.630	0.102	4.781	0.000
aspirin	-0.335	0.716	0.094	-3.568	0.000
reinfarkt	0.503	1.653	0.125	4.025	0.000

Likelihood ratio test $= 289$ on 6 df, $p = 0$, $n = 1017$ (23 observations deleted due to missing)

Primer: uporaba Coxovega modela za primerjavo krivulj preživetja



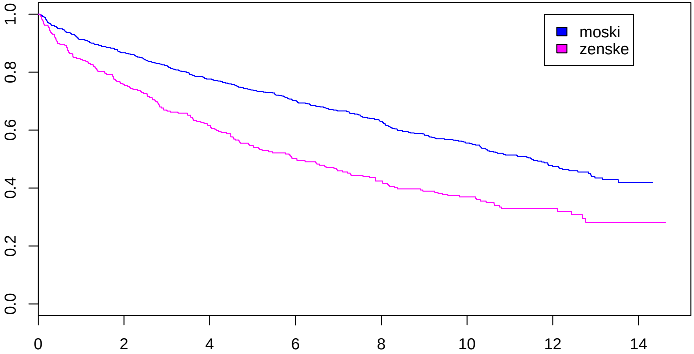
Primer: uporaba Coxovega modela za primerjavo krivulj preživetja

Stadij IV vzamemo za referenčno kategorijo.

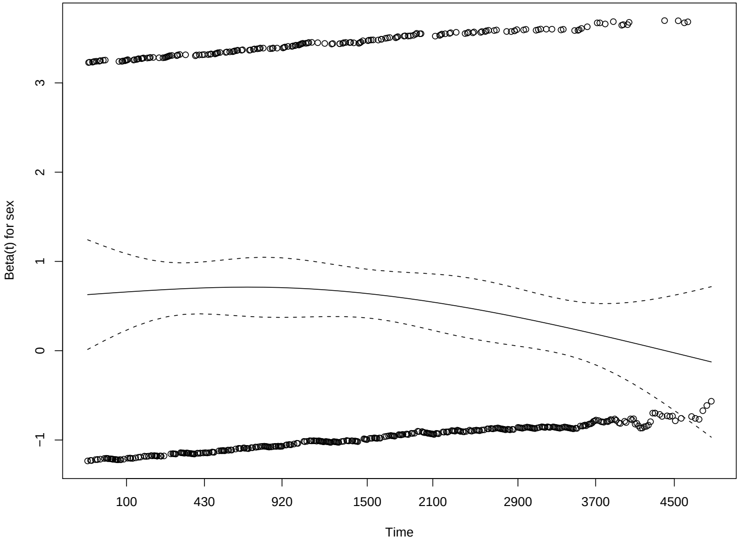
Stadij	Stadij I	Stadij II	Stadij III
I	1	0	0
II	0	1	0
III	0	0	1
IV	0	0	0

	coef	exp(coef)	se(coef)	z	p
Stadij III	-0.316	0.729	0.202	-1.57	0.120
Stadij II	-0.779	0.459	0.199	-3.92	< 0.001
Stadij I	-1.203	0.300	0.213	-5.64	< 0.001

Primer: preverjanje prileganja



Primer: preverjanje prileganja



Druge teme

- Stratificirani model
- Spremenljivke, ki se spreminjajo v času
- Krhkosti (frailties)

Drugi regresijski pristopi

- Parametrični modeli
- Aalenov linearni model
- Linearni model Buckleya in Jamesa

Vendar, vsaj za enkrat, je Coxov model daleč najbolj uporabljan v biomedicinskih aplikacijah.