

Pregledni znanstveni članek ■

## Individualno relativno preživetje

## Individual Relative Survival

---

Institucija avtorjev: Medicinska fakulteta, Univerza v Ljubljani.

Kontaktna oseba: Maja Pohar Perme, Medicinska fakulteta, Univerza v Ljubljani, Vrazov trg 2, 1000 Ljubljana. email: [maja.pohar@mf.uni-lj.si](mailto:maja.pohar@mf.uni-lj.si).

### Maja Pohar Perme, Janez Stare

**Izvleček.** V analizi preživetja proučujemo čas od začetka opazovanja do nekega dogodka (smrt, ponovni pojav bolezni ...). Pri tem predpostavljamo, da je končni dogodek natančno določen, kar pa v praksi ni vedno res. Pogosto nas namreč zanima preživetje, povezano z določeno boleznijo, a je vzrok smrti težko oziroma celo nemogoče določiti. V takih primerih standardne tehnike analize preživetja ne zadostujejo več. Rešitev problema ponujajo metode relativnega preživetja, ki primerja preživetje opazovane skupine pacientov s preživetjem, ki bi ga pričakovali, če bi ti pacienti živeli enako kot populacija. V članku opisujemo nedavno predstavljeni pristop, ki omogoča primerjavo s populacijo na individualni ravni, in R knjižnico *relsurv*, s pomočjo katere lahko tovrstne podatke tudi analiziramo.

**Abstract.** In survival analysis we are interested in time from the beginning of an observation until certain event (death, relapse). We assume that the final event is well defined, however, that is not always true in practice. Although the cause of death is often hard or even impossible to obtain, our main interest lies in cause specific survival. In such cases, the standard survival techniques can not be used. The cure to the problem is the relative survival analysis, that compares the observed survival to the one expected should the patients follow the population survival rates. This paper describes a recently introduced approach, that compares the observed and population data on an individual level, and the R package *relsurv*, that provides functions for straightforward use of all the relative survival methods.

■ **Infor Med Slov:** 2007; 12(1): 15-23

## Uvod

V analizi preživetja proučujemo čas med dvema dogodkoma, na primer čas med diagnozo in smrtjo, čas hospitalizacije, čas nezaposlenosti ... Razlog, da podatkov ne moremo analizirati z običajnimi metodami za kvantitativne podatke, je prisotnost krnjenja, torej dejstvo, da vseh posameznikov ne moremo spremljati do končnega dogodka. Tako imamo namesto časa  $T$  na voljo le podatek o času  $T^* = \min(T, C)$ , kjer  $C$  označuje čas krnjenja.

Najosnovnejša naloga analize preživetja je oceniti krivuljo preživetja  $S(t) = P(T > t) = 1 - F(t)$ , kjer je  $F(t)$  kumulativna porazdelitvena funkcija spremenljivke  $T$ . Pogosto nas zanima tudi pogojna funkcija  $S(t | X)$ , torej vpliv vektorja spremenljivk  $X$  na preživetje. Zaradi krnjenja funkcije  $S(t | X)$  ne moremo oceniti neposredno, zato je večina regresijskih modelov definiranih preko funkcije ogroženosti

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

Ker v gornjem izrazu nastopajo le pogojne verjetnosti, to funkcijo lahko ocenjujemo navkljub krnitvam. Funkciji ogroženosti in preživetja sta povezani s formulo

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

če poznamo eno, lahko torej vedno izračunamo drugo. Najpogosteje je uporabljen Coxov model,<sup>1</sup> ki predpostavlja, da je ogroženost enaka

$$\lambda(t, x) = \lambda_0(t) e^{\beta' x},$$

kjer je  $x$  vektor spremenljivk za vsakega posameznika,  $\beta$  vektor koeficientov,  $\lambda_0(t)$  pa osnovna ogroženost, ki je enaka za vse, lahko pa se spreminja v času.

Da bi razumeli potrebo po analizi relativnega preživetja, predpostavljajmo, da želimo analizirati

slovenske podatke o bolnicah z rakom na dojki. Recimo, da je analiza pokazala, da je preživetje žensk v vzhodnem delu države slabše. To bi nas lahko napeljalo k iskanju napovednih dejavnikov, ki bi lahko razložili to razliko. Vendar pa se izkaže, da se razlikuje tudi populacijsko preživetje žensk iz vzhoda in zahoda, ženske iz vzhodnega dela države imajo tako približno 4 leta krajšo pričakovano življenjsko dobo. To pomeni, da razlika, ki jo opazamo pri raku na dojki, morda le odseva populacijske razlike, tako umrljivost zaradi raka na dojki ni odvisna od geografske lokacije oziroma je odvisnost manjša, kot bi to lahko sklepali iz naše analize. Da bi lahko sklepali o tovrstnih podatkih, moramo zato uporabiti analizo relativnega preživetja, ki primerja opazovane krivulje preživetja s pričakovanimi (populacijskimi).

Metode relativnega preživetja nam zaradi primerjave s populacijskim preživetjem omogočajo, da analiziramo podatke o preživetju po določeni bolezni, čeprav nimamo podatkov o vzroku smrti. Najpogosteje se uporabljajo pri proučevanju preživetja bolnikov z rakom ter so nepogrešljive pri vseh raziskavah, ki spremljajo paciente skozi daljše časovno obdobje in ne morejo zagotoviti zanesljivih podatkov o vzroku smrti.

V članku bomo opisali pristop k relativnemu preživetju,<sup>2</sup> ki omogoča primerjave s populacijo na individualni ravni in s tem ponuja dodaten, s standardnimi metodami nedosegljiv vpogled v podatke. Da bi bil pristop uporabljen tudi v praksi, smo v programskem okolju R<sup>3</sup> razvili knjižnico funkcij relsurv,<sup>4</sup> ki je opisana v nadaljevanju članka. Njen namen je čimbolj poenostaviti delo z metodami relativnega preživetja. Pomemben del knjižnice so tudi funkcije za lažjo vključitev populacijskih tablic umrljivosti, saj je ta korak v analizi podatkov sicer pogosto najbolj zamuden.

## Relativno preživetje

Funkcija relativnega preživetja je definirana kot kvocient opazovanega ( $S_O(t)$ ) in pričakovanega preživetja ( $S_P(t)$ ):

$$S_R(t) := \frac{S_O(t)}{S_P(t)}.$$

Opazovano preživetje pri tem ocenimo z metodo Kaplan-Meier, pričakovano preživetje pa izračunamo s pomočjo populacijskih tablic umrljivosti po Hakulinenovi<sup>5</sup> metodi. Pri tem je pomembno poudariti, da funkcija relativnega preživetja nima nujno lastnosti funkcije preživetja, saj je lahko večja od 1 in ni nujno padajoča.

Če bolezen, ki jo proučujemo, ne poslabša preživetja bolnikov, je opazovana krivulja zelo podobna pričakovani, zato je krivulja relativnega preživetja konstantno blizu 1. Če pa se verjetnost smrti zaradi bolezni poveča, bo opazovana krivulja pod pričakovano in zato krivulja relativnega preživetja nižja od 1.

Poleg same krivulje relativnega preživetja nas pogosto zanima njena odvisnost od spremenljivk; v ta namen obstaja več regresijskih modelov, ki se razlikujejo glede na osnovne predpostavke o razmerju med opazovno in populacijsko ogroženostjo. Najpogosteje uporabljen je aditivni model,<sup>6</sup> lepe teoretične lastnosti pa ima multiplikativni model.<sup>7</sup>

## Individualno relativno preživetje

Krivulja relativnega preživetja, opisana v prejšnjem razdelku, nam da dober vpogled v preživetje opazovane skupine, vendar na podlagi te krivulje ne moremo delati zaključkov na individualni ravni.

Tako ne moremo odgovoriti na vprašanje, kako dolgo je nekdo živel glede na splošno populacijo, oziroma ali je A živel relativno dlje kot B. Kot primer si pogledjmo Franka Sinatro, ki je umrl leta 1996 star 81 let. Francoski matematik Adrien Marie Legendre je dosegel enako starost in umrl leta 1833. Je Legendre živel relativno dlje? Da bi odgovorili na to vprašanje, lahko iz populacijskih tabel izračunamo povprečno starost ob smrti za

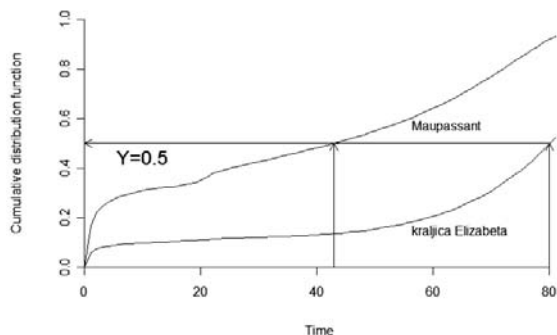
obe koledarski leti. Razlika je vse prej kot zanemarljiva - medtem ko je bila v času Sinatrove smrti povprečna starost umrlih v Ameriki 73,2 leti, je bila povprečna starost umrlih v Franciji leta 1833 le 38 let! Ta podatek nazorno pokaže, da je Legendre živel relativno bistveno dlje glede na takratne razmere. Podobno vprašanje si lahko zastavimo tudi za posameznike različnih starosti, vendar za odgovor potrebujemo bolj občutljivo mero. Recimo, da želimo primerjati preživetje francoskega pisatelja Guya de Maupassanta, ki je umrl leta 1893 star 42,9 let, in angleške kraljice Elizabete, ki je leta 2006 praznovala 80. rojstni dan. Tako Maupassant kot kraljica Elizabeta sta preživela povprečno starost umrlih, ki je bila v Franciji takrat 42,03, v Angliji pa (podatek je za leto 2003) 77,3, torej je njuno relativno preživetje podobno. Želimo torej vpeljati mero, ki bi nam natančneje govorila o teh razmerjih.

Naj  $T$  označuje čas preživetja,  $F_p$  pa kumulativno porazdelitveno funkcijo populacijskega časa smrti za posameznika določene starosti in spola v določenem koledarskem letu. Individualno mero relativnega preživetja<sup>2</sup>  $Y$  definiramo kot

$$Y := F_p(T).$$

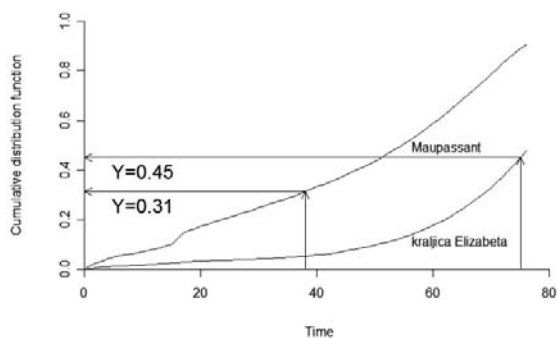
Vrednost  $y = F_p(t)$  nam za nekega posameznika, ki je preživel čas  $t$ , poda delež njegovih vrstnikov (torej populacije iste starosti in spola), ki so umrli do tega časa.

Slika 1 prikazuje izračun vrednosti  $y$  za Maupassanta in angleško kraljico. Krivulji predstavljata populacijski kumulativni porazdelitveni funkciji za Francoze, rojene leta 1850, in Angležinje, rojene leta 1926. Čeprav je njuna dosežena starost različna, vidimo, da sta glede na svoje vrstnike dosegla enako vrednost  $y$ , oba sta preživela natanko polovico svoje populacije.



**Slika 1** Populacijska kumulativna porazdelitvena funkcija.

Na sliki 1 je vidno, da je pomemben delež k velikemu umiranju Maupassantovih vrstnikov prispevala predvsem velika umrljivost dojenčkov. Slika 2 prikazuje kumulativni porazdelitveni funkciji populacijskega preživetja za vrstnike, ki so preživeli prvih pet let. Relativni preživetji se sedaj razlikujeta; kraljica je preživela 45 % vrstnikov, ki so doživeli peto leto starosti, medtem ko jih je pred Maupassantom umrlo le 31 %.



**Slika 2** Populacijska kumulativna porazdelitvena funkcija pogojno glede na preživetih prvih 5 let.

Mera  $Y$  ima torej zanimivo intuitivno razlago, zares uporabno pa jo naredijo njene teoretične lastnosti. S pomočjo preslikave časov  $T$  dobimo nove čase  $Y$ , ki so med 0 in 1. Status krnjenja ter vse ostale spremenljivke pri transformaciji ostanejo enaki in tako lahko vse metode standardne analize preživetja uporabljamo tudi s transformiranimi

časi. Za ocenjevanje krivulje preživetja lahko uporabimo standardne metode, le da namesto časov  $t$  uporabljamo čase  $y$ . Hkrati vedno poznamo populacijsko krivuljo preživetja, ki jo tako lahko uporabljamo za primerjavo. Če so časi  $T$  namreč porazdeljeni tako, kot narekujejo populacijske tabele, torej če opazovani posamezniki živijo enako kot populacija, potem so časi  $Y$  enakomerno porazdeljeni. To pomeni, da je populacijska krivulja preživetja enaka

$$S(y) = 1 - y.$$

Regresijski modeli v transformiranem času  $Y$  imajo pomembno prednost pred standardnimi modeli relativnega preživetja. S transformacijo smo namreč upoštevali vse populacijske razlike med posamezniki, tako da lahko razlike, ki ostanejo v času  $Y$ , neposredno pripisemo bolezni, ki nas zanima. S tem smo se izognili vsakršnim predpostavkam o povezavi med opazovanim in populacijskim preživetjem, ki so potrebne v standardnih modelih relativnega preživetja. V času  $Y$  tako lahko uporabimo poljubni model, podobno kot v standardni analizi preživetja bo verjetno najpogosteje izbran Coxov model sorazmernih tveganj.

## Tablice umrljivosti na spletu

Populacijske tablice umrljivosti, ki jih izdajajo nacionalni statistični uradi, so sestavni del analize relativnega preživetja. Ponavadi so tablice ločene po spolu in starosti ter podane za vsako koledarsko leto posebej, vendar so formati od države do države različni, ZDA imajo na primer dodatno tablice ločene še po rasi. Razlikuje se tudi časovno obdobje, za katerega so tablice na voljo - pri tem prednjači Francija, ki zbira tabele že od leta 1806. Raznovrstnost formatov je lahko ovira pri uporabi tablic v praksi, zato je poenostavitev prevedbe tablic v ustreznem formatu v R-u eden od osnovnih namenov knjižnice *relsurv* in je podrobneje opisana v naslednjem razdelku.

K razpoložljivosti tablic veliko pripomorejo spletne baze, ki objavljajo tablice za različne države v enakem formatu. Ena izmed takih strani je Human Mortality Database (HMD: <http://www.mortality.org>), ki vključuje podatke za 28 držav. Za vsako državo je na voljo več vrst tabel, za uporabo v relativnem preživetju potrebujemo tiste, ki ločeno po spolu in starosti vsebujejo verjetnosti smrti za vsako koledarsko leto (ta podatek je v stolpcu označenem s  $q(x)$ ). Vse tabele so splošno dostopne v formatu txt.

Enotnost formata tabel HMD je seveda pogojena z enotnostjo zbiranja podatkov in metodologije izračunavanja. Nekoliko manj enotna, a zato toliko večja je zbirka Human Lifetable Database (HLD: <http://www.lifetable.de>), kamor prispevajo svoje podatke inštitucije in posamezniki iz 43 držav. Časovna obdobja zajeta v teh tabelah so precej daljša, tabele pa so prav tako dosegljive v formatu txt. Neenotnost tabel seveda pomeni slabšo medsebojno primerljivost in več težav pri prevajanju v enotni format, ki ga potrebujemo za izračune.

Dostopnost tabel preko spleta je pomemben korak k lažji uporabi relativnega preživetja. Kljub temu je izračun vrednosti  $F_p(t)$  še vedno zamuden, saj moramo za vsakega posameznika posebej upoštevati različne podatke v tabeli glede na njegovo spreminjajočo se starost in spreminjanje koledarskega leta. Tu se pokažejo bistvene prednosti statističnega paketa R,<sup>7</sup> v katerem je v knjižnici survival za populacijske tabele definiran poseben objekt, imenovan ratetable, ki poenoti različne formate in omogoča preprosto ravnanje s tabelami. Ker sorodni statistični paketi (SAS, Stata) ne ponujajo tovrstnih poenostavitev, je izbira R-a za relativno preživetje dokaj samoumevna.

Opisani viri tabel ter zmogljivosti R-a so združeni v ilustraciji ideje individualnega pristopa k relativnemu preživetju na strani <http://www.mf.uni-lj.si/ibmi> (Biostatistični center, zanimivosti). S pomočjo preprostega vmesnika lahko uporabnik nariše populacijsko kumulativno

porazdelitveno funkcijo (sliki 1 in 2) ter odčita delež umrlih vrstnikov za osebo poljubne starosti in spola za kar 29 držav. Dodatno so na voljo tudi vse tabele v formatu ratetable, ki ga zahteva R.

## Knjižnica relsurv

Knjižnica relsurv je trenutno najobsežnejša zbirka funkcij za relativno preživetje. Poleg tega da vključuje vse standardne regresijske modele in metode za preverjanje njihovega prileganja, je njen glavni namen poenostavitev analize s pomočjo individualnega pristopa. Poleg prilagodljivosti različnim formatom populacijskih tabel je njena prednost predvsem v enotni sintaksi, ki tesno sledi ukazom iz sorodne bazične knjižnice survival. Knjižnica je splošno dostopna na spletnem zbirališču CRAN.

Prvi korak v analizi relativnega preživetja je zagotovitev populacijskih tabel. Najpreprostejše je delo s slovenskimi podatki, saj so tabele (trenutno segajo do leta 2003) že vključene v knjižnico in jih naložimo z ukazom

```
> data(slopop)
```

Kot je opisano v prejšnjem razdelku, so preko spletne strani <http://www.mf.uni-lj.si/ibmi> v formatu R dosegljive tabele za 29 držav, ki trenutno večinoma zajemajo obdobje do leta 2002 ali 2003. Da bi čimbolj olajšali prevod poljubnih tabel in s tem omogočili njihovo sprotno posodabljanje, smo v knjižnico vključili zbirko funkcij, ki ta postopek močno poenostavijo:

- `transrate.hmd`: funkcija pretvori tabele, dosegljive na spletni strani HMD (glej prejšnji razdelek) v format `ratetable`. Spletno zbirališče HMD ponuja več različnih tabel, za analizo relativnega preživetja so bistvene momentne tablice umrljivosti, urejene po letu smrti (period life tables). Tabele so dosegljive ločeno po spolu, funkcija `transrate.hmd` pa uporablja tiste, ki so podane v letnih intervalih tako za starost kot koledarsko leto (na spletu

poimenovane 1x1). V tabelah odstranimo naslovno vrstico ter jih shranimo v formatu txt. Nato jih prevedemo v format R z ukazom

```
> poptab <- transrate.hmd(male="
mltper_1x1.txt", female="fltper_1x1.txt")
```

Enotnost formatov v HMD nam zagotavlja, da so tabele ločene po starosti (od 0 do 110), spolu in koledarskem letu.

- `transrate.hld`: funkcija poenostavlja uporabo tabel, dosegljivih na strani HLD. Tudi te tabele so na voljo v formatu txt, vsebujejo pa oba spola hkrati. Pogosto so ločene po koledarskih obdobjih, saj vsebujejo podatke iz različnih popisov prebivalstva. Glavni namen te funkcije je zato združevanje teh obdobj. Finske tabele od leta 1975 do 1995 tako lahko dobimo z ukazom

```
finpop <- transrate.hld(file=c("FIN_1971-
75.txt", "FIN_1976-80.txt", "FIN_1981-
85.txt", "FIN_1986-90.txt", "FIN_1991-
95.txt"))
```

Večjo prilagodljivost funkcije omogočata dva dodatna argumenta. Prvi je imenovan `cut.year` in omogoča uporabniku, da določi obdobja, ki jih zajemajo tabele, kadar si ta ne sledijo zaporedoma. Drugi argument, imenovan `race`, omogoča vključitev dodatne dimenzije, na primer `rase`. Je enake dolžine kot argument `file` in povezuje datoteke z ustrežno vrednostjo dodatne dimenzije. Ta argument lahko uporabimo tudi za združevanje dveh tabel različnih držav, vendar le pod pogojem, da so enako urejene in imajo enak časovni razpon.

Kot primer uporabe funkcije si pogledjmo združevanje tabel za Novo Zelandijo, kjer vodijo tabele ločeno za belce in Maore (argument `cutyear=c(1980,1985)` določa, da želimo uporabiti tabele 1980-1982 do leta 1985):

```
nzpop <- transrate.hld(c("NZL_1980-
82_Non-maori.txt", "NZL_1985-87_Non-
```

```
maori.txt", "NZL_1980-82_Maori.txt",
"NZL_1985-87_Maori.txt"), cut.year=
c(1980,1985), race=rep(c("nonmaori",
"maori"), each=2))
```

- `joinrate`: funkcija pomaga pri združevanju dveh ali več tabel, ločenih po starosti, spolu in koledarskem letu v nov objekt tipa `ratetable` z dodatno dimenzijo. Tako tabelo lahko na primer uporabljamo v raziskavi, v katero so vključeni posamezniki iz več držav. Da bo primerjava med državami smiselna, morajo tabele pokrivati enaka časovna razdobja, v nasprotnem primeru funkcija vrne opozorilo.
- `transrate`: funkcija pomaga pri pretvorbi v format `ratetable`, kadar imamo podatke, podane v dveh po spolu ločenih tabelah.

Z nekoliko več truda lahko uporabnik tvori tudi poljubne splošnejše tabele, saj je objekt `ratetable` zelo prilagodljiv.<sup>8,9</sup> Tabeli, ki vsebuje verjetnosti preživetja, je potrebno le dodati opis (`attributes`), ki določa njeno organiziranost.

Ko so tabele urejene v pravo obliko (to lahko preverimo z ukazom `is.ratetable`), si lahko njihovo organizacijo ogledamo z ukazom `summary` oziroma podrobneje z `attributes`.

Pomembna zahteva formata `ratetable` je, da morajo biti vse vrednosti ogroženosti podane v enoti 1/dan. To tudi pomeni, da morajo biti vsi časi in starosti, uporabljeni v funkcijah za relativno preživetje, izraženi v dnevih. Datumski format, uporabljen v knjižnici `relsurv`, je enak formatu v knjižnici `survival` (`date`).

Krovna funkcija za analizo podatkov z individualnim pristopom je `rstrans`. Čeprav je njen primarni namen prileganje Coxovega modela v transformiranem času  $Y$ , jo lahko hkrati uporabimo tudi za samo pretvorbo časov. Njeni osnovni argumenti so:

- `formula`: leva stran formule je objekt tipa `Surv`, ki je definiran v knjižnici `survival` in vsebuje čas opazovanja ter status krnjenja. Desna stran

(od leve strani je ločena z znakom  $\sim$ ) vsebuje spremenljivke, ki jih želimo upoštevati v regresijskem modelu. Če sta čas opazovanja in status krnjenja podana v spremenljivkah z imenom *cas* in *status* ter bi radi v model vključili spremenljivko *x*, napišemo

```
>rstrans(Surv(cas, status)~x)
```

V objektu, poimenovanem *y*, funkcija vedno shrani tudi transformirane čase *Y*, če nas zanimajo slednji, zadostuje ukaz

```
>rstrans(Surv(cas, status))$y
```

- *data*: opazovani podatki v obliki *data.frame*.
- *ratetable*: populacijska tabela v ustreznem formatu.

Ker so populacijske vrednosti izračunane glede na vrednosti demografskih spremenljivk (npr. starost, spol, koledarsko leto), ki jih vsebujejo opazovani podatki, morajo biti imena teh spremenljivk enaka tistim v populacijskih tabelah. Če temu ni tako, moramo v formulo dodati člen, v katerem razložimo svojo specifično urejenost. Enako postopamo, kadar časovne spremenljivke niso izražene v dnevih. Če imamo torej spremenljivke z imeni *spol*, *starost* in *leto* in je *starost* merjena v letih, moramo v desni del formule dodati člen

```
ratetable(age=starost*365.24, sex=spol,
year=leto)
```

Funkcija *rstrans* vrne objekt razreda *coxph*, kar omogoča uporabo poljubnih funkcij, ki so v knjižnici *survival* na voljo za Coxov model. Dodatno lahko preverimo tudi prilaganje modela s pomočjo testov, ki temeljijo na Brownovem mostu,<sup>9</sup> in sicer z ukazom *rs.br*. Tudi ta ukaz sprejme več argumentov: *test* določa, kateri test želimo uporabiti, z argumentom *rho* pa lahko izberemo uteži. Na voljo je tudi metoda za risanje dobljenega procesa, ki jo priključimo z ukazom *plot*.

Poleg funkcij za individualno relativno preživetje vsebuje knjižnica *relsurv* tudi vse potrebno za

uporabo ostalih regresijskih modelov v relativnem preživetju. Za aditivni model tako uporabimo ukaz *rsadd*, medtem ko multiplikativni model priključimo z ukazom *rsmul*. Sintaksa vseh funkcij je popolnoma enaka, dodani so le argumenti, ki ponujajo možnosti, ki so za določeni model specifične.

## Primeri

Kot primer osnov ravnanja s populacijskimi tabelami v R-u si pogledjmo ukaze, potrebne za sliki 1 in 2. Iz baze HLD lahko dobimo ustrezne populacijske tabele, ki jih nato s pomočjo funkcije *transrate* pretvorimo v zahtevani format. Za francoske tabele torej napišemo

```
frpop <- transrate.hld("FRA18061997.txt")
```

Da bi izračunali kumulativno porazdelitveno funkcijo za 80 let po Maupassantovem rojstvu, napišemo:

```
y <- 1-survexp(~ratetable(age=0, sex="male",
year=as.date("5Aug1850")),
times=(0:80)*365.24, ratetable=frpop)$surv
```

ter nato narišemo *y* glede na starost z ukazom

```
plot(0:80, y, type='l')
```

Kot drugi primer si pogledjmo raziskavo o preživetju po miokardnem infarktu, ki je bila izvedena v Kliničnem centru v Ljubljani. Vanjo je bilo vključenih 1040 pacientov, ki so imeli infarkt med leti 1982 in 1986, spremljali pa smo jih do 1997. V tem obdobju je 547 bolnikov umrlo. Ker je v večini primerov nemogoče določiti, ali je bila smrt posledica infarkta ali ne, je ta raziskava dober primer potrebe po metodologiji relativnega preživetja.

Podatki o bolnikih z miokardnim infarktom so v okleščeni obliki vključeni knjižnico *relsurv* pod imenom *rdata*. Populacijski podatki za Slovenijo, ki so dosegljivi v knjižnici (imenovani *slopop*) so

ločeni glede na starost (age), spol (sex) in koledarsko leto (year). Podatke naložimo z ukazi

```
> library(relsurv)
> data(slopop)
> data(rdata)
```

Cilj raziskave je ugotavljanje vpliva spola, starosti in koledarskega leta na preživetje po infarktu. Pogled v podatke nam pove, kako so spremenljivke urejene in poimenovane (tabela 1).

**Tabela 1** Imena spremenljivk in njihovo urejenost, ki jih izpišemo z ukazom `> rdata[1:2,]`.

	time	cens	age	sex	year	agegr
1	2657	1	68	2	24Jun82	61-70
2	1097	1	63	2	31Aug82	61-70

Vidimo, da je čas preživetja poimenovan time in podan v dnevih, kot to zahtevajo funkcije za relativno preživetje. Status krnjenja je v spremenljivki cens. Starost, spol in leto so v spremenljivkah z enakim imenom kot v populacijskih tabelah, potrebno je le spremeniti starost v letih v starost v dnevih. To spremembo določimo v formuli:

```
> rstrans(Surv(time, cens)~age + sex + year +
  ratetable(age=age*365.24, sex=sex, year=year),
  data=rdata, ratetable=slopop)
```

Rezultati so podani v tabeli 2.

**Tabela 2** Rezultati Coxovega modela v transformiranem času za paciente z miokardnim infarktom.

Spremenljivka	Koeficient	Standardna napaka	P
age	-0.0139	0.0049	0.005
sex	0.5287	0.1010	<0.001
year	-0.0002	0.0001	0.012

Vidimo lahko, da vse tri spremenljivke značilno vplivajo na relativno preživetje. Relativno preživetje je boljše za moške in se izboljšuje s koledarskim letom, kar nakazuje, da se je zdravljenje bolnikov z infarktom z leti izboljševalo.

Smiselno je tudi negativni koeficient za starost, ki pomeni, da starejši pacienti z infarktom izgubijo relativno manj kot mlajši.

## Zaključki

Analiza relativnega preživetja nam odgovarja na vprašanja, na katera ne moremo odgovoriti s pomočjo klasičnih metod preživetja. Trenutno se uporablja predvsem za podatke iz registrov raka in je manj znana med ostalimi raziskovalci v medicini. Dejstvo, da ne obstaja knjiga o analizi preživetja, ki bi vsebovala poglavje o tej temi, nazorno ponazarja trenutno stanje na tem področju. Delni razlog za to je zagotovo tudi dejstvo, da je analiza relativnega preživetja vezana na populacijske tabele, kar pogosto pomeni oviro pri njeni uporabi. S čedalje boljše dosegljivostjo tabel se bo stanje spremenilo in tu je obstoj ustreznega programskega orodja ključnega pomena. Ker knjižnica relsurv s svojimi funkcijami pokriva vse faze analize in ravnanja s podatki, je pomemben prispevek k splošni uporabi teh metod.

## Literatura

1. Cox DR. Regression models and life-tables (with discussion). *Journal of the Royal Statistical Society – Series B* 1972; 34:187-220.
2. Stare J, Henderson R, Pohar M. An individual measure of relative survival. *Journal of the royal statistical society – Series C* 2005; 54:115-126.
3. Foundation for Statistical Computing, Vienna, Austria, 2006. <http://www.R-project.org>.
4. Pohar M, Stare J. Relative survival analysis in R. *Computer methods and programs in biomedicine* 2006; 81:272-278.
5. Ederer F, Axtell LM, Cutler SJ. The relative survival rate: a statistical methodology. *National Cancer Institute Monograph* 1961; 6:101-121.
6. Hakulinen T, Tenkanen L. Regression analysis of relative survival rates. *Journal of the Royal Statistical Society – Series C* 1987; 36:309-317.
7. Andersen PK, Borch-Johnsen K, Deckert KT et al. A Cox regression model for relative mortality and its application to diabetes mellitus survival data. *Biometrics* 1985; 41:921-932.



8. Therneau T, Offord J. Expected survival based on hazard rates (update). Technical Report 63, Section of Biostatistics, Mayo Clinic, 1999.
9. Stare J, Pohar M, Henderson R. Goodness of fit of relative survival models. *Statistics in Medicine* 2005; 24:3911-3925.