

# Analiza zgodovine dogodkov - Seminarska naloga

## 3.1 Srčno žilne bolezni - preživetje brez srčno žilnih zapletov

Urh Peček

2021/22

### Kazalo

<b>0. Uvod</b>	<b>2</b>
0.1 Opis podatkov . . . . .	2
0.2 Navodilo . . . . .	2
<b>1. Predstavitev podatkov</b>	<b>3</b>
1.1 Spol . . . . .	3
1.2 Starost . . . . .	3
1.3 Diabetes . . . . .	4
1.4 Kap . . . . .	5
1.5 Število pregledov in čas med pregledi . . . . .	6
1.6 Čas spremljanja . . . . .	6
1.7 Čas smrti in vzrok smrti . . . . .	7
1.8 Povezanost vzroka konca spremljanja in časa spremljanja . . . . .	7
<b>2. Priprava podatkov za analizo</b>	<b>9</b>
<b>3. Čas preživetja</b>	<b>10</b>
3.1 Čas preživetja glede na smrt iz katerega koli vzroka . . . . .	10
3.2 Povezanost spola s preživetjem . . . . .	12
3.3 Dodatno upoštevanje starosti . . . . .	14
<b>4. Čas preživetja glede na vzrok smrti</b>	<b>17</b>
4.1 Preživetje glede na smrt iz srčno-žilnih vzrokov . . . . .	17
4.2 Čas preživetja brez srčno žilnih zapletov . . . . .	19
<b>5. Povezanost diabetesa s časom preživetja brez srčno-žilnih zapletov</b>	<b>22</b>
<b>6. Zaključek</b>	<b>25</b>

## 0. Uvod

### 0.1 Opis podatkov

Pri seminarski nalogi bomo uporabili podatke iz podane datoteke `data_sz.r`. Dani so podatki o bolnikih, pri katerih je zdravnik odkril simptome srčno žilnih bolezni. Podatki vsebujejo naslednje spremenljivke:

- `mainid`: id bolnika
- `spol`: moski / zenski
- `d.pregled0`: datum začetnega pregleda
- `d.pregled1`: datum trenutnega pregleda
- `d.roj`: datum rojstva
- `d.diab`: datum diagnoze diabetesa (NA, če diabetesa nima ob začetku raziskave)
- `kaj`: trenutni kadilski status
- `d.kajod`: datum začetka kajenja
- `d.kajdo`: datum konca kajenja
- `alko`: pitje alkohola (0=ne pijem, 1= manj kot 3 merice tedensko, 2=3-7 meric tedensko, 3=več kot 7 meric tedensko; 1 merica je 1dcl vina)
- `bmi`: indeks telesne mase
- `tlak.sis`: sistolični tlak
- `l.hol`: vrednost celokupnega holesterola
- `l.ldl`: vrednost holesterola
- `l.hdl`: vrednost holesterola (slabi holesterol)
- `d.kap`: datum kapi
- `d.smt.sz`: datum smrti iz srčno-žilnih vzrokov
- `d.smrt`: datum smrti iz drugih vzrokov
- `h.sgl`: vrednost glukoze v krvi
- `gluk`: prisotnost glukoze v urinu

### 0.2 Navodilo

Zanima nas preživetje brez srčno žilnih zapletov.

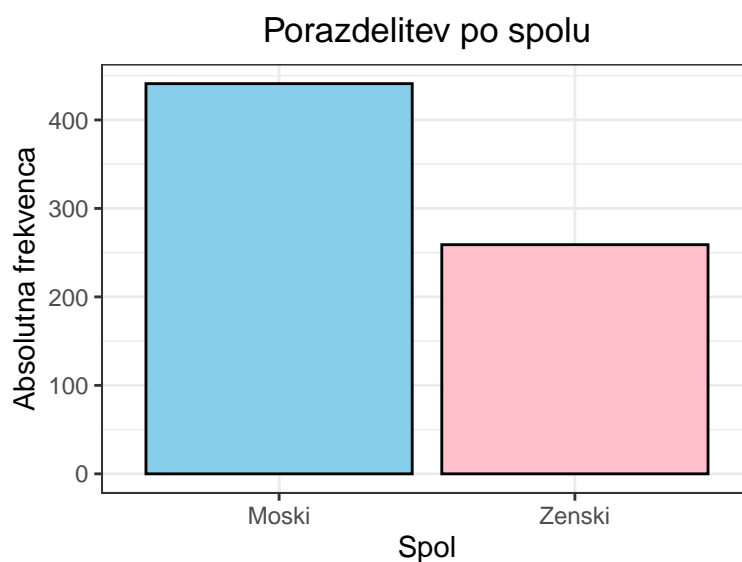
1. Predstavitev podatke.
2. Podatke pretvorite v ustrezen format (pazite na to, kakšne so spremenljivke, ki vas zanimajo).
3. Zanima nas čas preživetja (dogodek je smrt iz katerega koli vzroka). Raziščite povezanost spola s preživetjem. Kaj pa, če upoštevate še starost?
4. Zdravnike zanima le smrt iz srčno žilnih vzrokov. Razložite, zakaj ne moremo gledati preživetja, kjer obravnavamo le ta dogodek?
5. Raziščite čas preživetja brez srčno žilnih zapletov (brez kapi, smrti, infarkta). Primerjajte z gornjimi rezultati in pojasnite razlike v interpretaciji.
6. Kako bi preverili povezanost diabetesa s časom preživetja brez srčno žilnih zapletov?

# 1. Predstavitev podatkov

Podatki vsebujejo informacije o 700 bolnikih predstavljenimi z identifikacijskimi številkami, ki jih opisuje 20 v uvodu opisanih spremenljivk v 3739 vrsticah. Predstavimo najprej nekatere lastnosti domnevno pomembnejših spremenljivk, ki v podatkih nastopajo in bi nam v nadaljevanju lahko pomagale odgovoriti na vprašanja, ki nas zanimajo ter nam dale boljši pregled nad podatki.

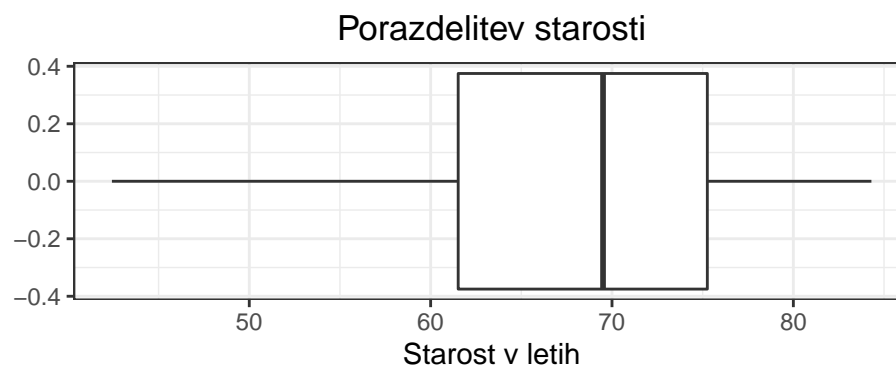
## 1.1 Spol

Moškega spola je 441 posameznikov (63%) in 259 posameznic je ženskega spola (37%).



## 1.2 Starost

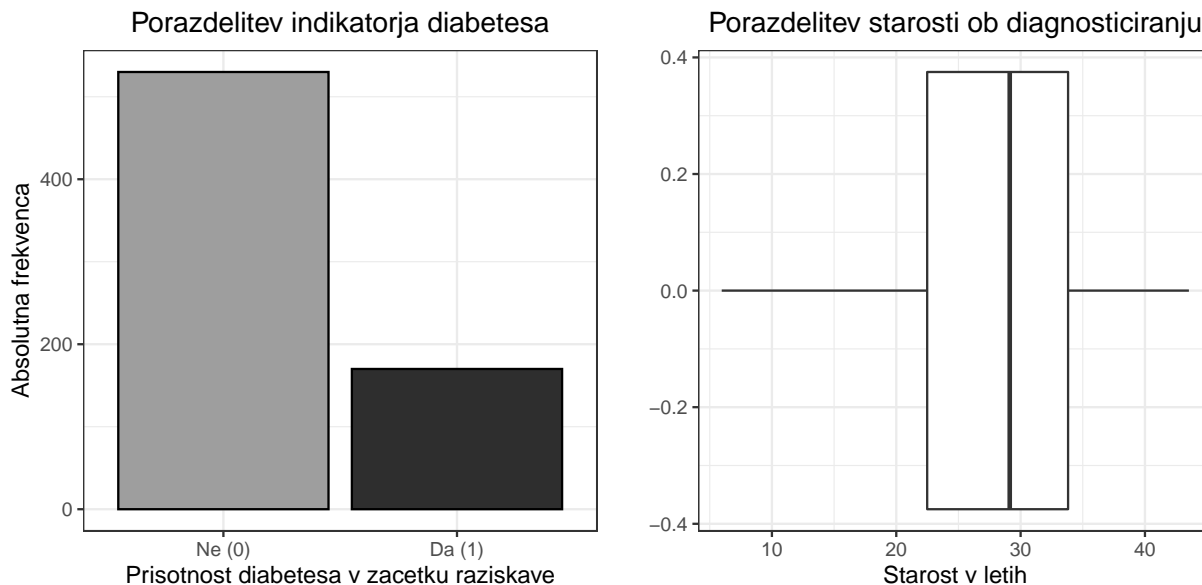
Povprečna starost posameznika ob začetku raziskave je bila 68.1 let. Najmlajši posameznik je bil star 42.4 let, najstarejši pa 84.3 let. Prvi kvartil znaša 61.5 let, četrti kvartil pa 75.3 let.



### 1.3 Diabetes

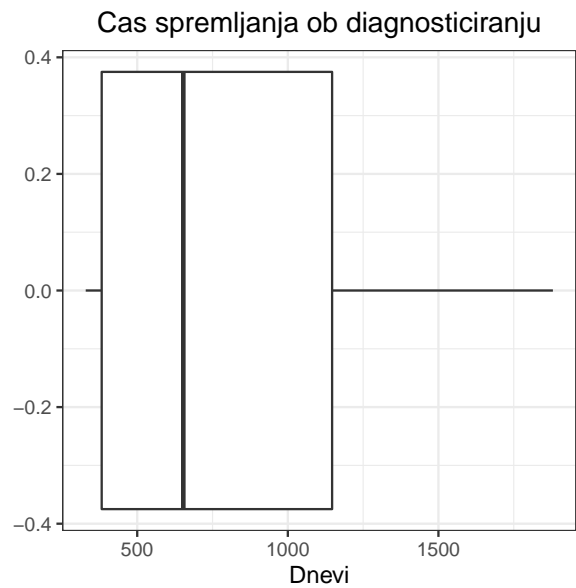
V začetku raziskave je bil diabetes diagnosticiran pri 170 posameznikih (24.3%). Od teh, je povprečna starost ob diagnozi diabetesa znašala 28 let, standardni odklon je bil 8.25 let. Pri bolnikih, ki nastopajo v raziskavi je bil diabetes najhitreje odkrit pri skoraj 6 letih in najkasneje pri 43.5 letih.

Če pogledamo čas, ki je minil od diagnosticiranja diabetesa pa do začetnega pregleda, kjer so bili ugotovljeni simptomi srčno žilnih bolezni velja, da je v povprečju minilo 40.5 let, standardni odklon znaša zgolj 0.48 let. To pomeni, da je povprečni čas od diagnosticiranja diabetesa pa do prvih simptomov srčno žilnih bolezni pri vseh posameznikih zelo podoben in je nekaj več 40 let.



Tekom raziskave je bil diabetes odkrit pri 28 posameznikih, kar je 4%. Pogoji za odkritje diabetesa je bil bodisi vrednost glukoze v krvi nad 7 bodisi prisotnost glukoze v urinu in diagnoza se postavi, če najdemo znak diabetesa na dveh zaporednih pregledih.

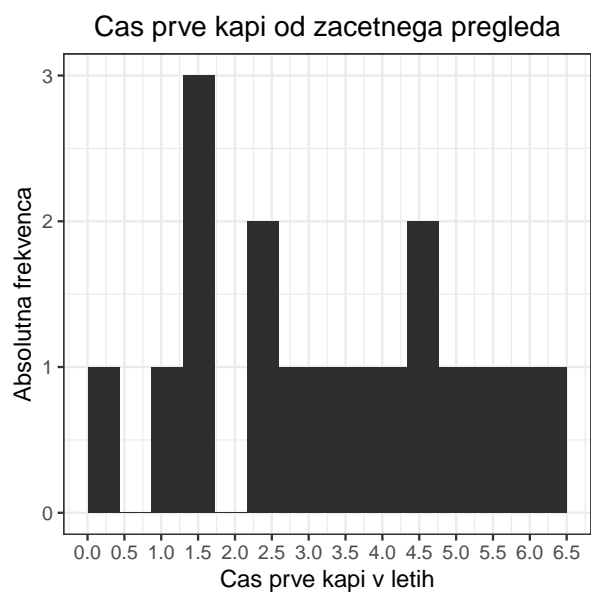
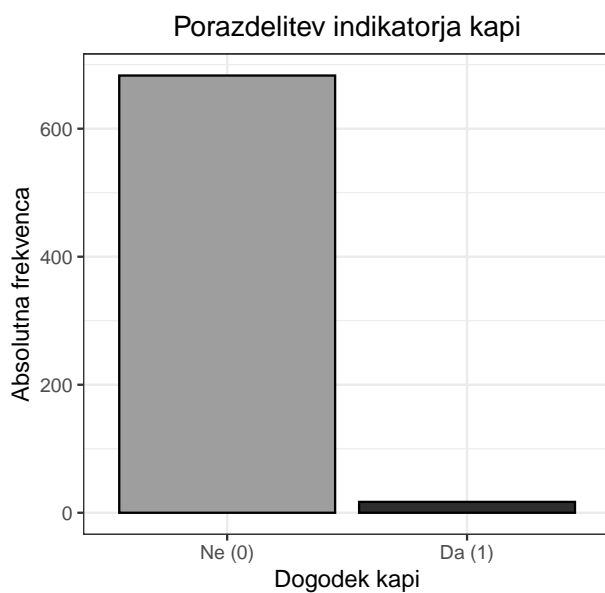
Pri posameznikih, katerim je bil diabetes odkrit tekom raziskave, je od začetnega pregleda, pa do pregleda, kjer je bil diabetes odkrit, v povprečju minilo 811 dni oz. 2.22 let. Prvi kvartil je 382 dni in tretji kvartil je 1147 dni.



Do konca spremljanja je imelo diabetes tako 198 posameznikov oziroma 28.2%.

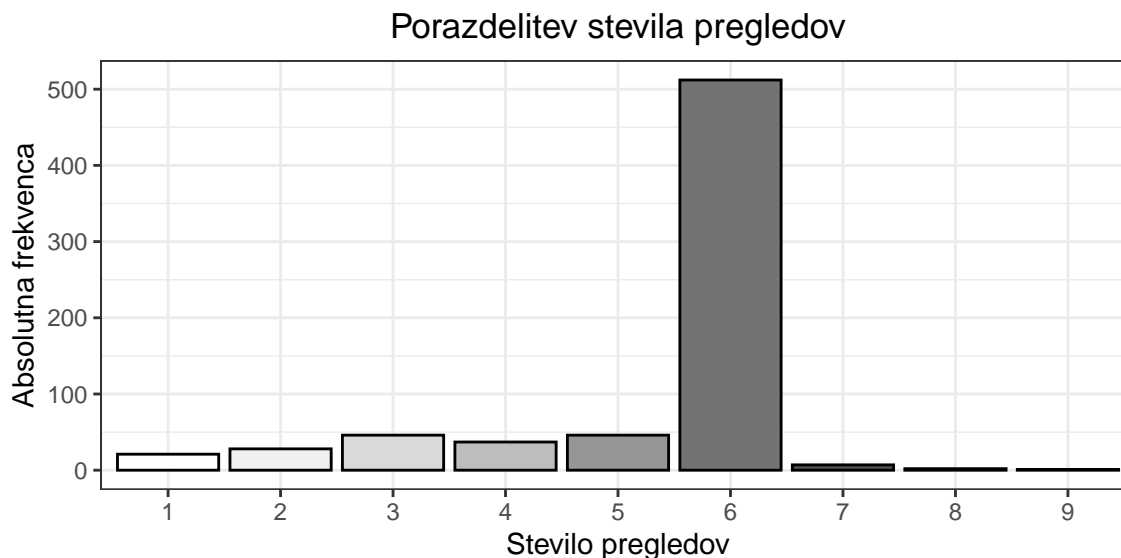
## 1.4 Kap

Od začetnega pregleda pri zdravniku, je do kapi prišlo pri 17 bolnikih, kar predstavlja zgolj 2.43% bolnikov. Povprečen čas spremljanja, od začetnega pregleda pri zdravniku do prve kapi je bil 3.27 let. Najhitrejša kap je bila po 40 dneh, do zadnje kapi pa je prišlo po 6.18 letih. Pri nekaterih bolnikih je tekom spremljanja prišlo tudi do več kapi.

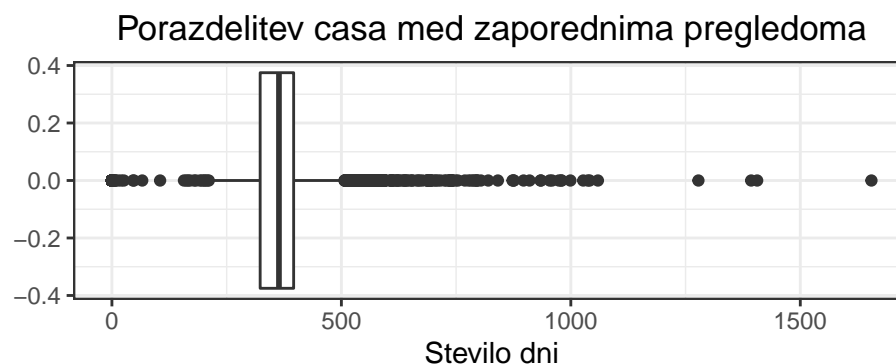


## 1.5 Število pregledov in čas med pregledi

V povprečju je bil posameznik tekom raziskave na 5.34 pregledih pri zdravniku. Največ obiskov posameznika pri zdravniku je bilo 9 in to je bilo le pri 1 posamezniku. Najmanjše število pregledov pri zdravniku je bilo 1 in to je bilo pri 21 posameznikih.



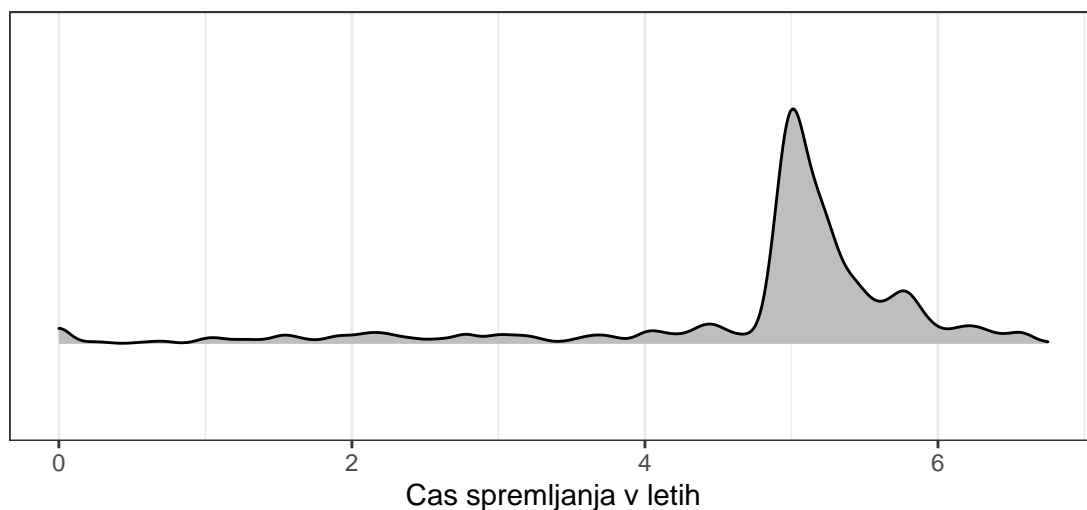
Med zaporednima pregledoma posameznika pri zdravniku je v povprečju minilo 317 dni. Mediana znaša 364 dni in prvi in tretji kvartil sta ji precej blizu, oba tudi večja od povprečja. Na spodnji sliki vidimo, da imamo nekaj zelo izstopajočih vrednosti, kjer je med zaporednima pregledoma posameznika pri zdravniku minilo več kot 3 leta. Taki posamezniki so štirje.



## 1.6 Čas spremljanja

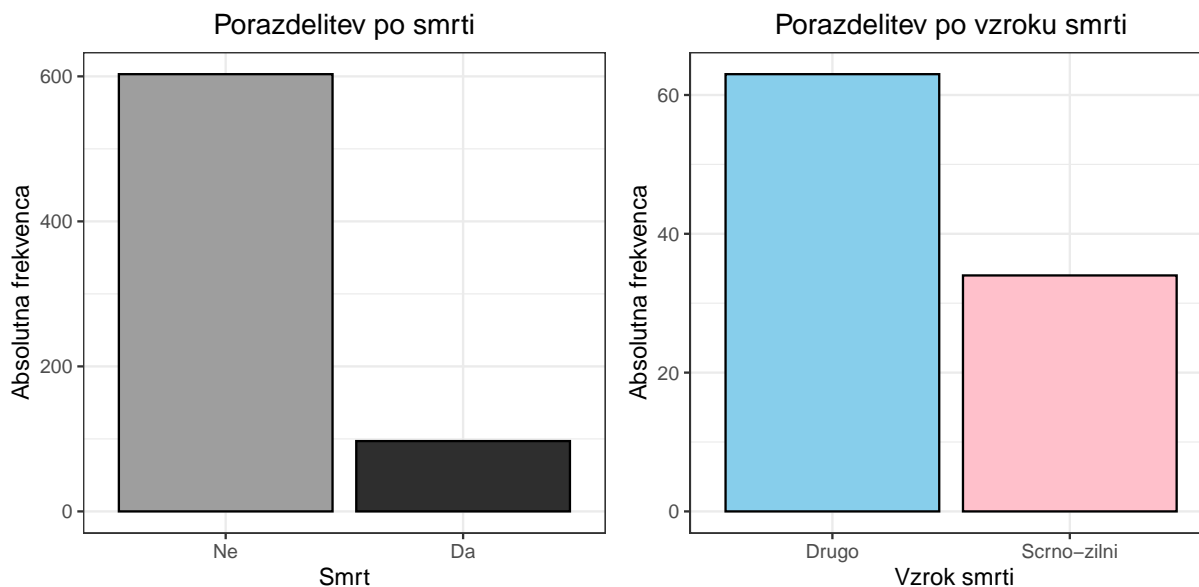
Povprečni čas spremljanja bolnika (čas od prvega do zadnjega pregleda, zadnje kapi v kolikor ne pride do smrti in je ta po zadnjem pregledu, ali smrti iz katerega koli vzroka) je bil 1731 dni oz. 4.74 let. Najdlje časa spremljan bolnik je bil spremljan 2464 dni (6.75 let). Najmanj časa spremljan bolnik je bil spremljan 0 dni, kar pomeni, da je prišel le na prvi pregled, kasneje pa smo ga iz raziskave “izgubili”, takih bolnikov je bilo 13.

## Porazdelitev časa spremljanja



## 1.7 Čas smrti in vzrok smrti

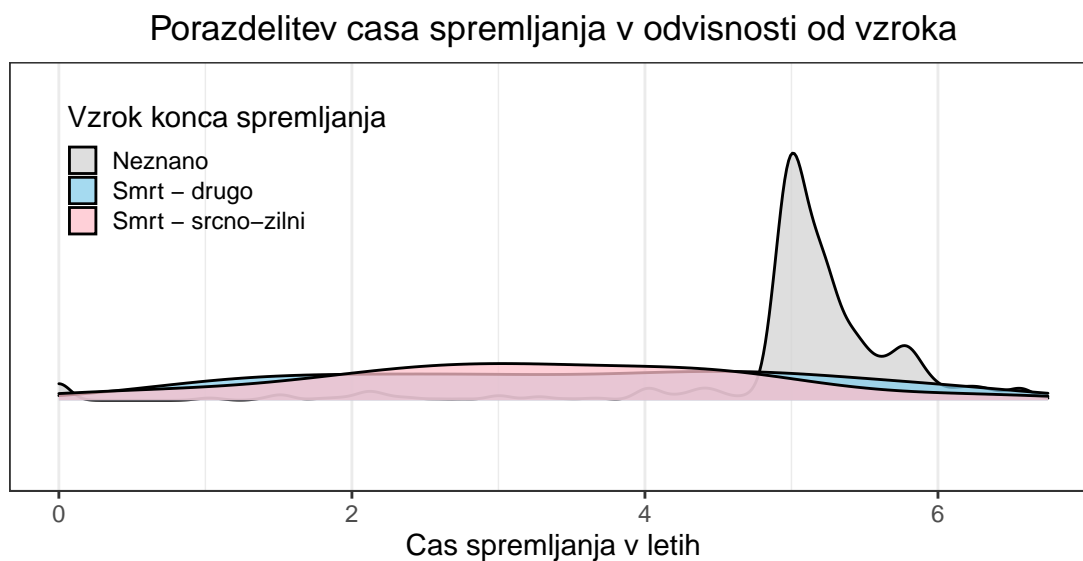
Do konca udejstvovanja posameznika v raziskavi zaradi smrti iz kateregakoli vzroka je prišlo pri 97 posameznikih, kar znaša 13.9%. Od tega je zaradi srčno žilnih vzrokov umrlo 34 posameznikov (4.9% oz. 35%). Zaradi drugih vzrokov je umrlo 63 posameznikov (9% oz. 65%). Posameznikov, s katerimi je bil zadnji stik zadnji zabeležen pregled in nimamo podatka o smrti je bilo 603, kar je 86.1%.



## 1.8 Povezanost vzroka konca spremljanja in časa spremljanja

Če si ogledamo kako je s porazdelitvijo časa spremljanja v primeru smrti iz srčno žilnega vzroka, drugega vzroka ali konca spremljanja, kjer ni bil vzrok smrt vidimo, da zaradi smrti do konca spremljanja pride nekako enakomerno porazdeljeno na celotnem intervalu spremljanja, zaradi drugih vzrokov pa do konca

spremljanja pride v večini v kasnejšem obdobju, okrog petega leta in v času nič, kar je omenjeno že zgoraj. Velikih razlik v času smrti od začetnega pregleda med srčno-žilnimi vzroki in drugimi vzroki ni. Morda je smrt zaradi drugih vzrokov nekoliko bolj pogosta v začetku in koncu spremljanja, torej v prvih dveh letih in po petem letu po prvem pregledu, smrt zaradi srčno žilnih bolezni pa od dveh do petih let od prvega pregleda.



Ostale spremenljivke bomo po potrebi predstavili v nadaljevanju naloge. Prav tako bomo po potrebi podrobneje predstavili že opisane spremenljivke, ter nadalje preverili njihovo povezanost.



## 2. Priprava podatkov za analizo

V namen nadaljnje analize bomo začetne podatke nekoliko priredili oziroma skrčili, da bodo bolj pregledni in vsebovali zgolj informacije, ki bi nas utegnile resnično zanimati. V kolikor bomo tekom raziskave ugotovili, da kakšna informacija manjka oz. podatki niso v formatu kakršnega potrebujemo, bomo to uredili sproti.

Naredimo torej nov podatkovni okvir, ki bo vseboval predvsem informacije, ki smo jih predstavili v uvodnem poglavju. Vsaka vrstica predstavlja enega posameznika (število vrstic je enako številu posameznikov) in vsakega posameznika opišemo s 15 spremenljivkami.

- `mainid` - id bolnika
- `spol` - spol bolnika
- `starost` - starost bolnika (dnevi)
- `diabetes_starost` - starost bolnika ob diagnosticiranju diabetesa (dnevi)
- `diabetes_do_pregled` - čas, ki je minil od diagnosticiranja diabetesa do začetnega pregleda (dnevi)
- `diabetes_ind` - indikator prisotnosti diabetesa ob začetnem pregledu (0/1)
- `cas_diabetes_tekom` - čas, ki je minil od začetnega pregleda pa do diagnosticiranja diabetesa
- `diabetes_tekom` - indikator diagnosticiranja diabetesa tekom raziskave
- `cas_prve_kapi` - čas od začetnega pregleda do prve kapi (dnevi)
- `kap_ind` - indikator kapi (0/1)
- `stevilo_pregledov` - število pregledov pri zdravniku
- `cas_spremljanja` - čas od začetnega do zadnjega pregleda ali kapi v kolikor ne pride do smrti, sicer smrti
- `smrt_sz` - indikator smrti zaradi srčno-žilnih vzrokov (0/1)
- `smrt_drugo` - indikator smrti zaradi drugih vzrokov (0/1)
- `smrt` - indikator smrti zaradi enega od dveh zgornjih vzrokov (0/1)

V kolikor bo potrebno si bomo podobno kot zgoraj preuredili prvotne podatke, vendar tako, da bo vsakega posameznika predstavljalo več vrstic, ki označujejo čas med zaporednima pregledoma. V ta namen smo si pripravili tudi dve novi spremenljivki oziroma stolpca, ki jih bomo po potrebi uporabili v kombinaciji s prvotnimi podatki. Ti dve spremenljivki predstavljata število dni, ki je minilo med zaporednima pregledoma posameznika pri zdravniku tj. `cas_med_pregledoma` in zaporedna številka pregleda posameznika tj. `trenutni_pregled` ter se nahajata v podatkovnem okviru `trenutni_pregled`, ki ga lahko enostavno uparimo z originalnim podatkovnim okvirjem.

Sedaj, ko smo predstavili in preuredili podatke v meri, za katero mislimo da je potrebna, se lahko lotimo same analize in odgovorov na zastavljena vprašanja.

### 3. Čas preživetja

Tekom celotne analize nas bo zanimalo preživetje bolnikov, ki ga bomo obravnavali prek analize preživetja. Sprva bo dogodek na katerega se bomo osredotočili smrt. V nekaterih primerih bo to smrt, kjer nas njen vzrok ne bo zanimal, v drugih primerih pa se bomo osredotočili na smrt pogojno na vzrok. Vzrok smrti je kot smo že videli lahko srčno-žilni ali drugi. Pri nekaterih posameznikih do smrti seveda ne pride in je čas konca njihovega udejstvovanja v raziskavi dan s časom zadnjega pregleda pri zdravniku oz. časom zadnje kapi. Te posameznike, pri katerih nimamo podatka o času smrti bomo v začetku obravnavali kot krnjene.

Glavna predpostavka celotne analize preživetja je predpostavka neinformativnega krnjenja. Ta predpostavka pravi, da imajo krnjeni posamezniki ob času krnitve enako verjetnost preživetja kot tisti, ki ostanejo v vzorcu.

Ta predpostavka je, v primeru obravnave dogodka kot smrt ne glede na vzrok izpolnjena. Krnjeni posamezniki imajo enako verjetnost preživetja kot tisti, ki v raziskavi še nastopajo, saj nimamo nobenih dodatnih informacij zakaj je do krnjenja prišlo in zakaj bi bila verjetnost preživetja krnjenih večja ali manjša od tistih, ki v raziskavi še nastopajo. Krnjenje v našem primeru zgolj pomeni, da posameznik do konca raziskave ni prišel na ponoven pregled pri zdravniku. Nimamo niti dodatnih podatkov o njegovem stanju, torej smrti in njenem času, po zadnjem pregledu pri zdravniku (oz. kapi).

Poleg zgornjega lahko sklepamo tudi tako, da so nekateri prenehali hoditi na preglede, ker jim je šlo z “boleznijo” tako dobro in niso več potrebovali obiskovanja zdravnikov, na drugi strani pa so med krnjenimi verjetno tudi taki, ki so se počutili že zelo slabo in v pregledih niso videli več smisla. V obeh primerih smo za posamezniki izgubili vsako sled in njihov vzrok krnjenja se ravno nekako izenači ter je krnjenje ponovno neinformativno.

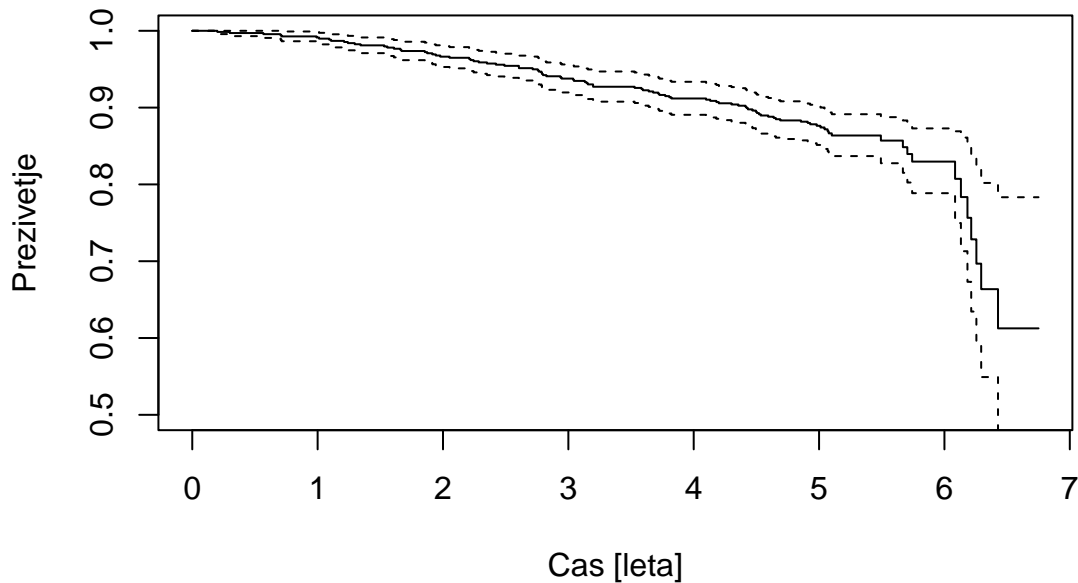
#### 3.1 Čas preživetja glede na smrt iz katerega koli vzroka

Sprva nas zanima čas preživetja posameznika. Dogodek na katerega se bomo osredotočili je smrt posameznika iz katerega koliko vzroka. V ta namen bomo uporabili podatke, kjer je posameznik predstavljen samo z eno vrstico, ter se osredotočili na spremenljivko, ki predstavlja indikator smrti ne glede na vzrok. V kolikor pri posamezniku do konca spremljanja ni prišlo zaradi smrti, temveč je zadnji zabeležen datum čas zadnjega pregleda ali kapi, bomo posameznika krnili.

Ocenimo krivuljo preživetja kot smo opisali zgoraj in jo grafično predstavimo. Za oceno krivulje uporabimo metodo Kaplan-Meier. Kot čas bomo upoštevali spremenljivko `cas_spremljanja` in kot indikator dogodka spremenljivko `smrt`. Krivuljo bomo narisali za celotno skupino bolnikov in jih ne bomo ločili glede na njihove lastnosti.

Vidimo, da je tveganje za smrt v prvih letih po začetnem pregledu in odkritju simptomov srčno-žilnih bolezni relativno nizko in precej konstantno. Tako je vse do nekje petega leta in pol po začetnem pregledu, kjer je verjetnost preživetja bolnika nekaj več kot 85.7%. Po šestem letu krivulja preživetja strmo pade in verjetnost preživetja 6.427 let po začetnem pregledu je 61.3%.

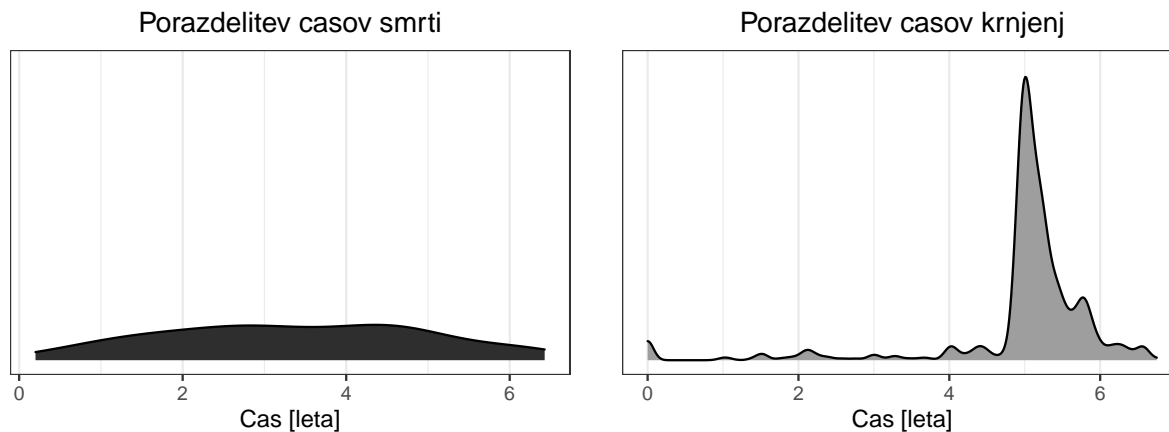
## Krivulja preživetja glede na smrt



Rezultat je glede na porazdelitev časov spremljanja, katere smo spoznali v uvodnem poglavju in jih imamo še enkrat predstavljene spodaj, morda nekoliko presenetljiv. Vidimo, da je porazdelitev časov smrti precej enakomerna tekom celotnega intervala časa raziskave, krivulja preživetja pa na prvi pogled nakazuje, da je tveganje za smrt precej večje v zadnjem obdobju raziskave.

Tu nastopi vloga krnjenih posameznikov s katerimi ta pojav razložimo. V prvem obdobju je krnjenj zelo malo, večina se jih zgodi na intervalu malo pred petim letom pa do šestega leta, kjer je tudi padec krivulje preživetja, z nekolikšnim zamikom, največji. Velikost skokov cenilke krivulje preživetja je odvisna od razmerja med številom posameznikov pod preiskavo in števila posameznikov katerim se v danem trenutku dogodek zgodi. Z vsakim dodatnim krnjenim posameznikom se torej manjša velikost vzorca in s tem vsaka smrt igra večjo vlogo in povzroči večji padec krivulje preživetja.

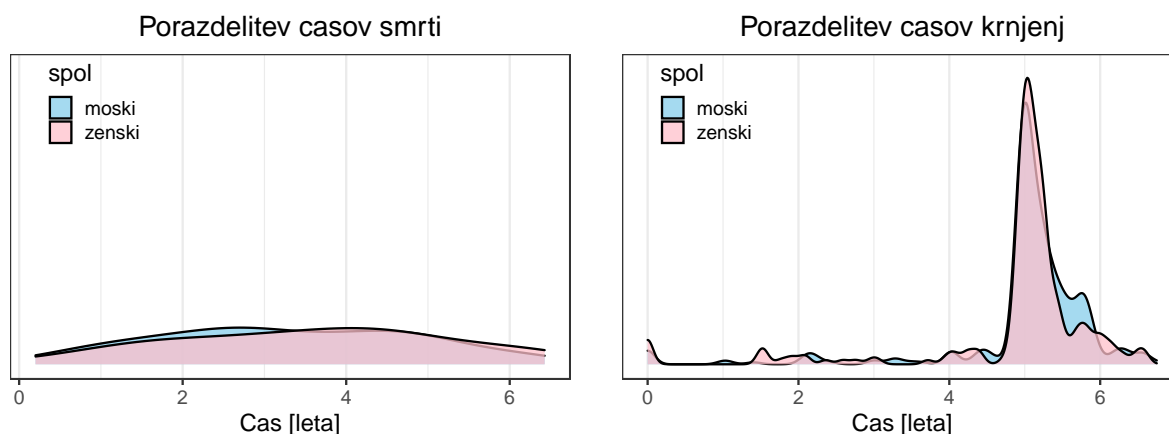
Padec krivulje preživetja po šestem letu je torej posledica povečanega števila krnjenih posameznikov, kar povzroči, da ima vsaka smrt večjo težo in tveganje za smrt je največje po šestem letu, čeprav so časi smrti precej enakomerno porazdeljeni po celotnem intervalu časa raziskave.



### 3.2 Povezanost spola s preživetjem

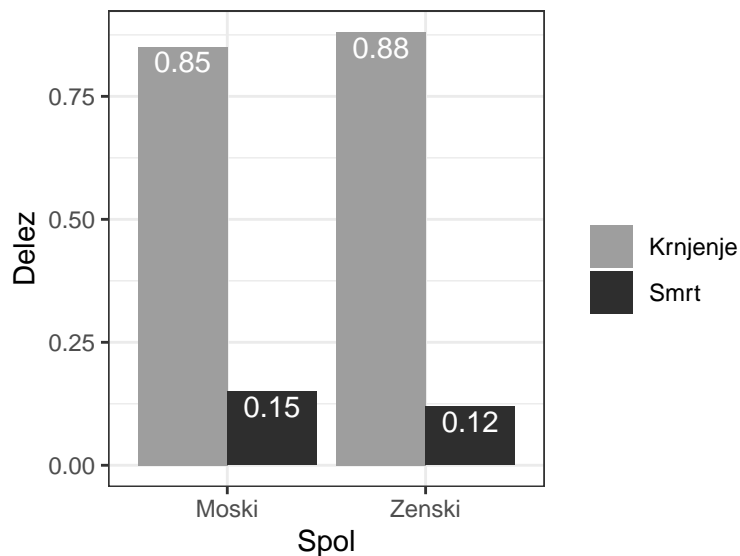
Predstavili in pojasnili smo krivuljo preživetja in tveganje posameznikov za smrt, kjer nismo upoštevali nobene dodatne spremenljivke. Sedaj pa si oglejmo kakšna je povezanost krivulje preživetja s spolom.

Grafično si najprej pogledjmo kako je s časi smrti in krnjenj upoštevajoč spol. Ali morda že tu opazimo kakšno razliko med spoloma, ki bi morda razložila razlike v času preživetij, ki bi jih lahko opazili prek krivulje preživetja in kasnejših testov.



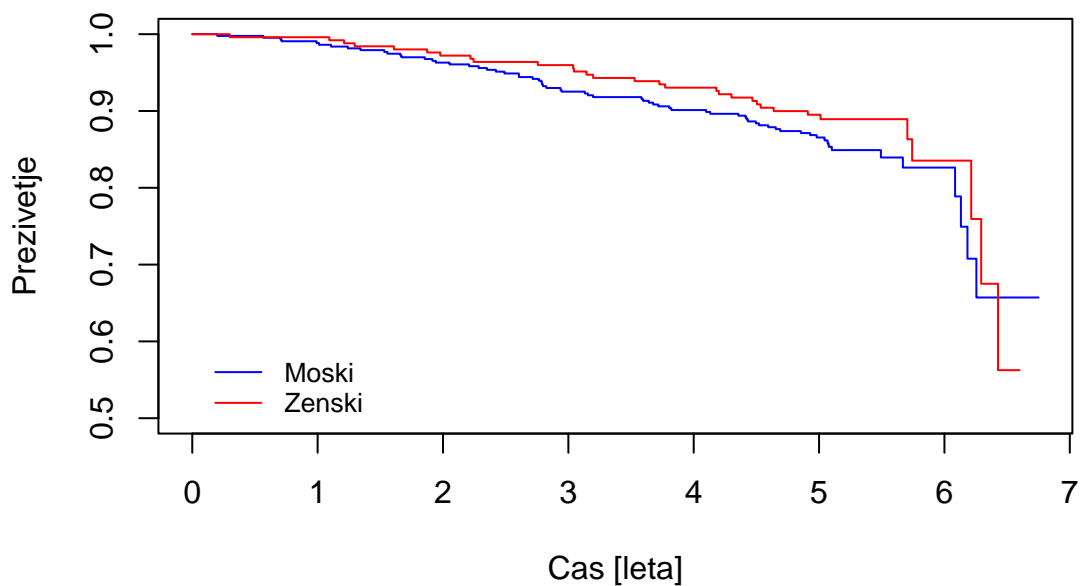
Iz časov smrti in krnjenj ni opaziti večjih razlik med spoloma. Pri moških se za odtenek več smrti zgodi v prvi polovici časovnega intervala in nekoliko več krnjenj proti koncu. Razlike so precej majhne in na podlagi tega ne bi mogli sklepati ničesar.

Kot smo že povedali, je moških 63% in žensk 37%. Delež krnjenj je pri moških za odtenek večji kot pri ženskah in posledično je delež smrti pri moških nekoliko večji. Pričakovati gre torej nekoliko večje tveganje za smrt pri moških kot pri ženskah.



V dodatek zgornjim dognanjem, slika kjer smo krivuljo preživetja izračunali glede na spol in na isti graf narisali krivuljo za moške in ženske kaže nekoliko različne razlike med spoloma. Ker je delež smrti v začetnem obdobju nekoliko večji pri moških in kasneje ni večjih razlik, ter zaradi večjega deleža smrti pri moških, je krivulja preživetja moških skozi celoten časovni interval nekoliko nižja od krivulje žensk.

### Krivulja preživetja glede na spol



Zgleda torej, da je tveganje moških nekoliko večje od tveganja žensk. Razlika med krivuljama je opazna, vendar ni nujno, da je statistično značilna. Ali je preživetje med spoloma statistično značilno različno bomo preverili s testom log-rank.

Test log-rank preverja ničelno domnevo  $H_0 : S_1(t) = S_2(t)$  za vsak  $t$ . Ta domneva pravi, da imata skupini, v našem primeru moški in ženske, enako porazdelitev do časa dogodka in s tem enaki tveganji oz. krivulji preživetja. Preverja torej domnevo o nepovezanosti opisne spremenljivke, v našem primeru spola, s časom smrti. Uporaben je samo za časovno neodvisne spremenljivke, torej spremenljivke, katerih vrednost se tekom raziskave ne spreminja. Pri nas spremenljivka spol temu zadošča.

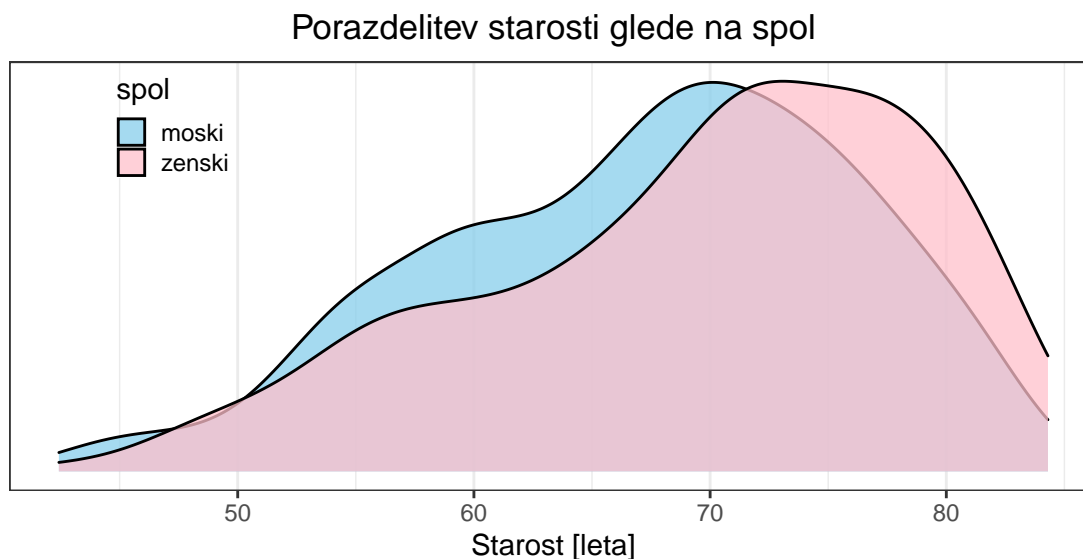
```
## Call:
## survdiff(formula = Surv(cas_spremljanja/365, smrt) ~ spol, data = podatki_ana)
##
##              N Observed Expected (O-E)^2/E (O-E)^2/V
## spol=moski  441         66      60.9      0.421      1.13
## spol=zenski 259         31      36.1      0.712      1.13
##
## Chisq= 1.1  on 1 degrees of freedom, p= 0.3
```

Na podlagi vrednosti  $p$  ( $p = 0.3$ ), na našem vzorcu ne moremo zavrniti ničelne domneve, da sta si krivulji preživetja glede na spol enaki. S tem ne moremo sprejeti sklepa, da je spol povezan s časom smrti. Razlika med krivuljama iz zgornje slike torej ni značilna. Ne moremo zavrniti domneve, da spol ni povezan s časom smrti, razlika med spoloma je premajhna.

### 3.3 Dodatno upoštevanje starosti

V model, kjer smo pri preživetju upoštevali spol bomo dodali še starost. Tvorili bomo torej multivariatni model.

Na spodnji sliki, ki prikazuje porazdelitev starosti ob prvem pregledu pri zdravniku glede na spol, opazimo nekoliko povezanost teh dveh spremenljivk. Velja, da so moški ob prvem pregledu oz. prvih znakih simptomov srčno-žilnih bolezni, v povprečju mlajši kot ženske.



Ta slika bi sicer večjo težo nosila v primeru statistične značilnosti spola, ki bi bila lahko posledica povezanosti s starostjo, vendar pa si jo velja pogledati tudi v splošnem.

Ker je starost številska in ne opisna spremenljivka in ker bomo v modelu upoštevali dve spremenljivki, bomo namesto log-rank testa uporabili Coxov model. Coxov model ima dve predpostavki. Prva je proporcionalnost tveganj, kar pomeni, da se razmerje tveganj v času ne spreminja. Druga predpostavka je linearni

efekt kovariat. Nobena od spremenljivk spola ali starosti ni časovno odvisna, starost se spreminja za vse posameznike enako.

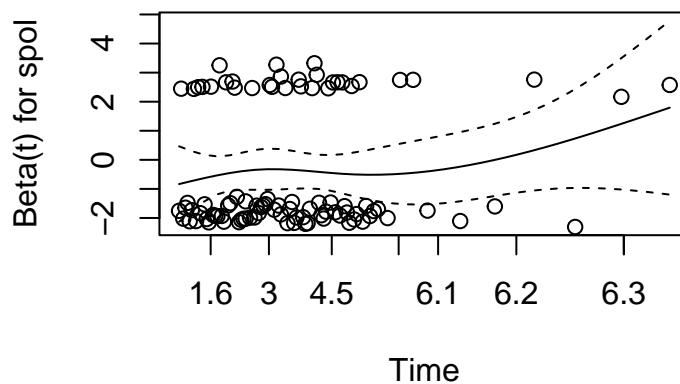
Z multivariatnim Coxovim model torej preverimo torej, ali je starost povezana s tveganjem neodvisno od spola (in obratno).

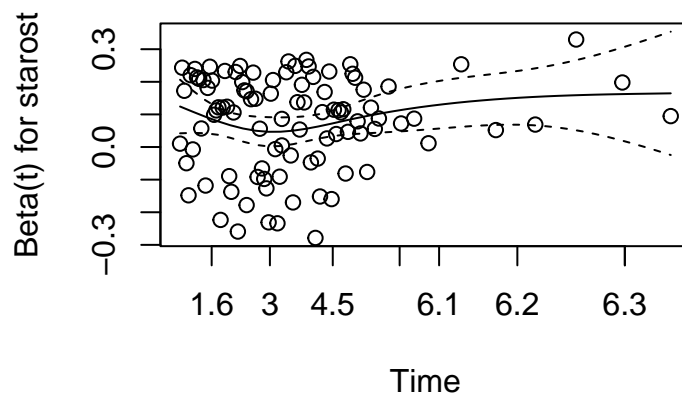
```
## Call:
## coxph(formula = Surv(cas_spremljanja/365, smrt) ~ spol + starost,
##       data = podatki_ana)
##
##               coef exp(coef) se(coef)      z      p
## spolzenski -0.39958   0.67060  0.21909 -1.824  0.0682
## starost      0.07899   1.08219  0.01385  5.704 1.17e-08
##
## Likelihood ratio test=39.12  on 2 df, p=3.193e-09
## n= 700, number of events= 97
```

Ob upoštevanju spola v modelu, je povezanost starosti s tveganjem smrti, na podlagi vrednosti  $p$  ( $p < 0.001$ ), močno statistično značilno. Koeficient starosti je pozitiven, kar pomeni, da se z vsakim letom starosti, ko ugotovijo simptome srčno-žilnih bolezni oz. pridemo na prvi pregled h zdravniku, poveča tveganje za smrt oz. v povprečju posameznik hitreje umre. Z drugimi besedami, tveganje za smrt je pri dveh posameznikih enakega spola ob ugotovitvi srčno-žilnih simptomov značilno večje pri starejšem posamezniku. Koeficient za spol ponovno ni statistično značilen. Pri dveh posameznikih enake starosti, tveganje ni povezano s spolom.

Preverimo še predpostavko o sorazmernem tveganju. Predpostavko bomo preverili prek Schoenfeldovih ostankov, katerih test preverja domnevo, da je razmerje tveganj za spremenljivko sorazmerno tj. skozi čas konstantno.

```
##           chisq df    p
## spol       1.76  1 0.18
## starost    1.31  1 0.25
## GLOBAL     3.02  2 0.22
```





Na podlagi vrednosti  $p$  ( $p > 0.05$ ) ne moremo zavrniti domneve (za vsako spremenljivko posebej in tudi za obe spremenljivki skupaj, kar je tu pomembno), da se razmerje tveganj spreminja v času. Predpostavka sorazmernega tveganja je torej izpolnjena.



## 4. Čas preživetja glede na vzrok smrti

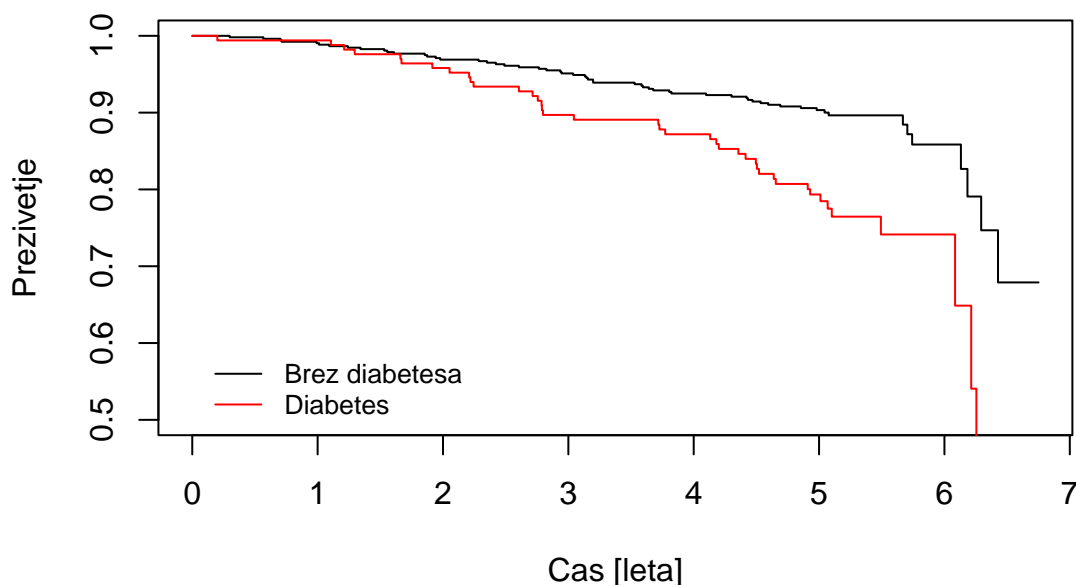
### 4.1 Preživetje glede na smrt iz srčno-žilnih vzrokov

V kolikor nas zanima le smrt iz srčno-žilnih vzrokov, ne moremo gledati preživetja, kjer obravnavamo le ta dogodek. Krnjenje smrti iz drugih vzrokov lahko povzroči pristranske rezultate pri dogodku zanimanja in osnovne metode ne zadostujejo. Težavo bomo sprva poskusili pojasniti nekoliko bolj raziskovalno, kasneje pa tudi teoretično.

V podatkih imamo informacijo o prisotnosti diabetesa pri vsakem posamezniku ob vstopu v raziskavo. Iz prakse vemo, da prisotnost diabetesa znatno poveča tveganje za smrt. To lahko vidimo tudi na spodnji sliki, kjer smo krivuljo preživetja izračunali in narisali glede na indikator prisotnosti diabetesa v začetku raziskave.

Tu smo prisotnost diabetesa interpretirali kot časovno neodvisno spremenljivko, saj smo vzeli zgolj informacijo o prisotnosti diabetesa, ki jo imamo v začetku raziskave in ne upoštevali posameznikov, katerim je diabetes diagnosticiran tekom raziskave.

**Krivulja preživetja glede na diabetes**



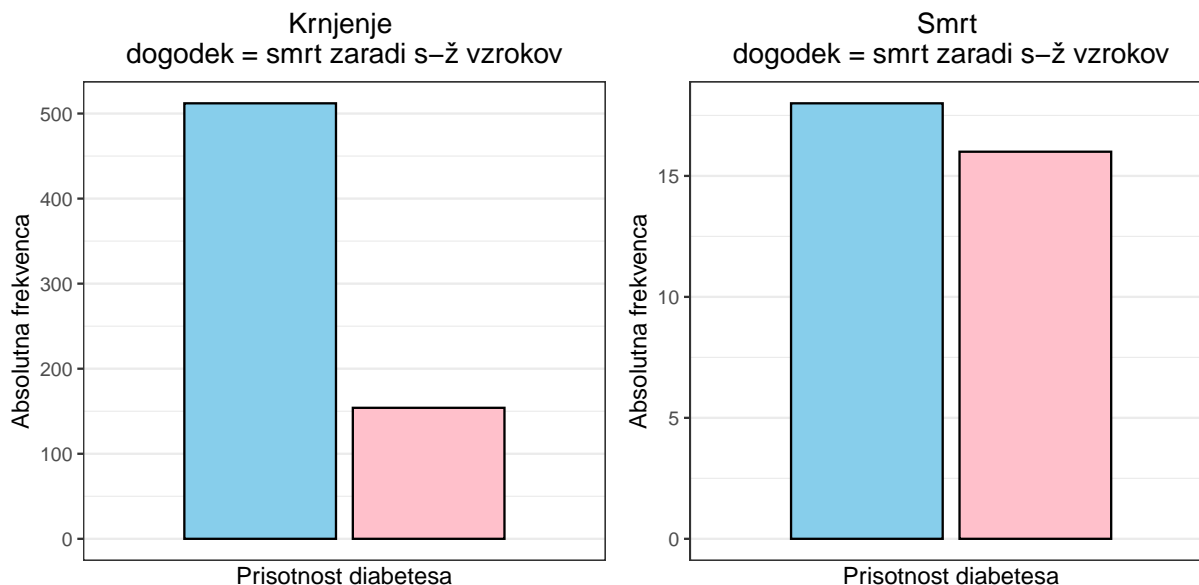
Močno značilno povezanost prisotnosti diabetesa v začetku raziskave s tveganjem kaže tudi test log-rank. V povezavi z zgornjo sliko sprejmemo sklep, da je tveganje posameznikov z diabetesom v začetku raziskave statistično značilno višje od tveganja posameznikov brez diabetesa ob vstopu v raziskavo.

```
## Call:
## survdiff(formula = Surv(cas_spremljanja/365, smrt) ~ diabetes_ind,
##           data = podatki_ana)
##
##               N Observed Expected (O-E)^2/E (O-E)^2/V
## diabetes_ind=0 530      57    74.3      4.04     17.3
## diabetes_ind=1 170      40    22.7     13.24     17.3
```

##

## Chisq= 17.3 on 1 degrees of freedom, p= 3e-05

Poglejmo si sedaj še povezanost diabetesa v začetku raziskave s krnjenji in smrtmi, kjer kot dogodek upoštevamo smrt zaradi srčno-žilnih vzrokov. Pri interpretaciji bomo upoštevali tudi sliko prisotnosti diabetesa ob začetnem pregledu z uvodnega poglavja, kjer smo ugotovili, da je v delež posameznikov, ki v raziskavo vstopijo z diabetesom enak 24% in delež posameznikov brez diabetesa enak 76%. Upoštevajoč spodnjo sliko vidimo, da je, ne glede na majhen delež smrti iz srčno žilnih vzrokov, porazdelitev diabetesa približno enakomerno porazdeljena pri smrtnih primerih, medtem ko je pri krnjenih posameznikih delež tistih brez diabetesa bistveno višji.



Zaključimo lahko, da v kolikor za končni dogodek upoštevamo zgolj smrt iz srčno-žilnih vzrokov je predpostavka o neinformativnem krnjenju kršena. Večji delež krnjenih posameznikov bi prihajal iz skupine brez prisotnosti diabetesa v začetku raziskave, za katere smo dokazali, da imajo manjše tveganje za smrt. Ob krnjenju bi imel tako povprečni krnjeni posameznik večjo verjetnost preživetja, kot tisti, ki v vzorcu ostane, saj bi bila verjetnost, da je krnjeni posameznik brez diabetesa večja kot z diabetesom, obenem pa bi v vzorcu ostajal vedno večji delež tistih z diabetesom.

Diabetes smo res da upoštevali zgolj kot indikator v začetku raziskave, vendar tudi, če bi ga obravnavali kot časovno odvisno spremenljivko in s tem upoštevali tudi posameznike, katerim je diagnosticiran tekom raziskave, menim, da se ne bi nič dramatičnega spremenilo.

Sedaj pa poskusimo odgovoriti še nekoliko bolj teoretično. V naši analizi imamo situacijo sotveganj, torej imamo več možnih različnih izidov, to sta smrt zaradi srčno-žilnih vzrokov in smrt zaradi drugih vzrokov.

Pri sotveganjih med funkcijo tveganja za nek izid in verjetnostjo, da nekdo konča z nekim izidom ni več bi-ektivne preslikave, kar pomeni, da iz funkcije tveganja ne moremo neposredno izračunati funkcije preživetja. Namesto funkcije preživetja moramo obravnavati funkcijo pojavnosti. Kumulativna funkcija pojavnosti za nek izid je vedno odvisna od funkcij tveganja za vse možne izide.

Pri analizi s sotveganji je bistveno razumeti kaj nas zanima oz. kaj želimo z raziskavo pokazati. Pri analizi povezanosti spremenljivk s tveganjem posameznega izida lahko krnimo vse izide, ki niso predmet zanimanja, saj krnjenje, čeprav ni neinformativno, ne vpliva na rezultate. Pri neposrednem sklepanju o verjetnosti danega izida, pa se ne moremo poslužiti krnjenja vseh izidov, ki niso predmet zanimanja. Namesto funkcije preživetja moramo obravnavati funkcijo pojavnosti.

## 4.2 Čas preživetja brez srčno žilnih zapletov

Zanima nas čas preživetja brez srčno-žilnih zapletov, torej brez kapi, smrti ali infarkta. Rezultate želimo primerjati tudi z gornjimi.

V namen odgovora na vprašanje bomo v obstoječ podatkovni okvir dodali spremenljivki `ind_sz` ki bo indikator sestavljenega dogodka in `cas_sz`, v katerem bomo upoštevali podatek o prvem času kapi, smrti (iz drugih vzrokov) oz. infarktu (smrt iz srčno-žilnih vzrokov), vedno je to sicer čas prve kapi. V kolikor do nobenega izmed dogodkov ni prišlo bomo to upoštevali kot krnitev in za čas krnitve vzeli čas zadnjega obiska pri zdravniku. Še enkrat omenimo, da bomo kot čas kapi upoštevali čas prve kapi, saj se kap pri nekaterih posameznikih pojavi tudi večkrat.

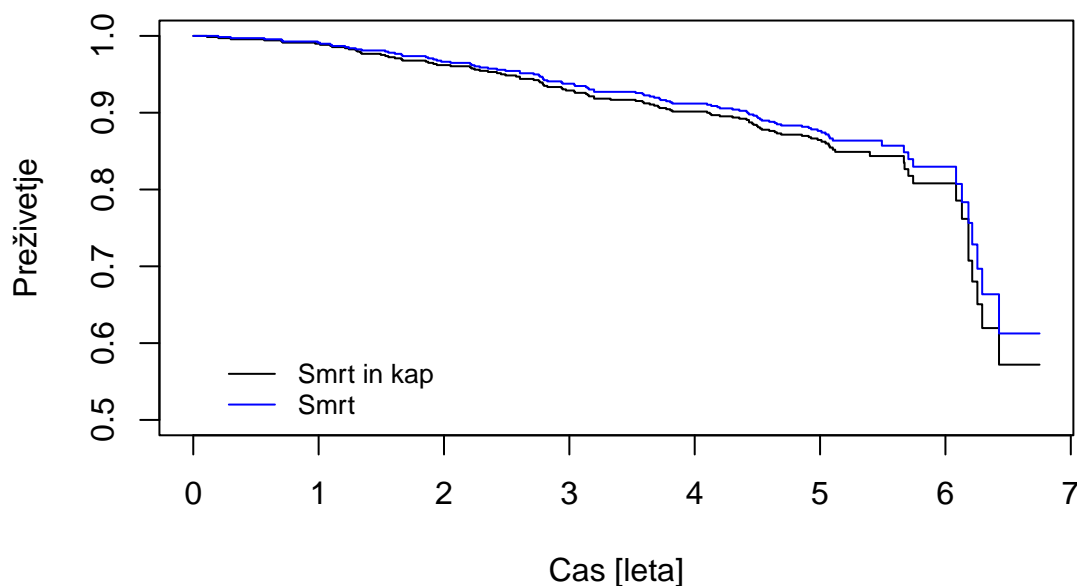
Najprej omenimo še, da tudi v tem primeru prihaja do neinformativnega krnjenja, saj se prav veliko glede na krnjenje, kjer je bil dogodek zgolj smrt ni spremenilo. H dogodkom smo dodali zgolj kapi in s tem odstranili nekaj krnjenj oz. jih premestili. Še vedno je krnjenje neinformativno, saj nimamo nobenih dodatnih informacij zakaj je do krnjenja sploh prišlo.

Dogodek na katerega se bomo osredotočili je torej sestavljeni dogodek, ki zaobjema kap, smrt iz srčno-žilnih vzrokov in smrt iz drugih vzrokov. Za izračun in izris krivulje preživetja po metodi Kaplan-Meier bomo uporabili zgoraj omenjeni spremenljivki `cas_sz` in `ind_sz`. V namen primerjave z rezultati, kjer kot dogodek upoštevamo zgolj smrt, na sliko dodajmo še krivuljo preživetja, kjer kot dogodek upoštevamo zgolj smrt, torej brez kapi.

V kolikor na sliko krivulje preživetja bolnikov brez srčno-žilnih zapletov ne bi dodali krivulje preživetja zgolj glede na smrt, razlike ne bi opazili, saj je ta precej majhna. Prek primerjave obeh krivulj pa vidimo, da se, če kot dogodek upoštevamo še kap, preživetje bolnikov nekoliko zmanjša. Z drugimi besedami, kap poveča tveganje za smrt.

Opomnimo, da to ni krivulja preživetja glede na indikator kapi, ki je med drugim časovno odvisna spremenljivka, ampak je to primerjava krivulj preživetja z različnimi končnimi dogodki.

### Primerjava krivulj preživetja – kap in brez kapi



Ali je povezanost kapi s preživetjem statistično značilna bomo preverili s Cox-ovim modelom. Ker je kap časovno odvisna spremenljivka, kjer njena vrednost v začetku raziskave ni znana, moramo podatke za pravilno analizo nekoliko preurediti.

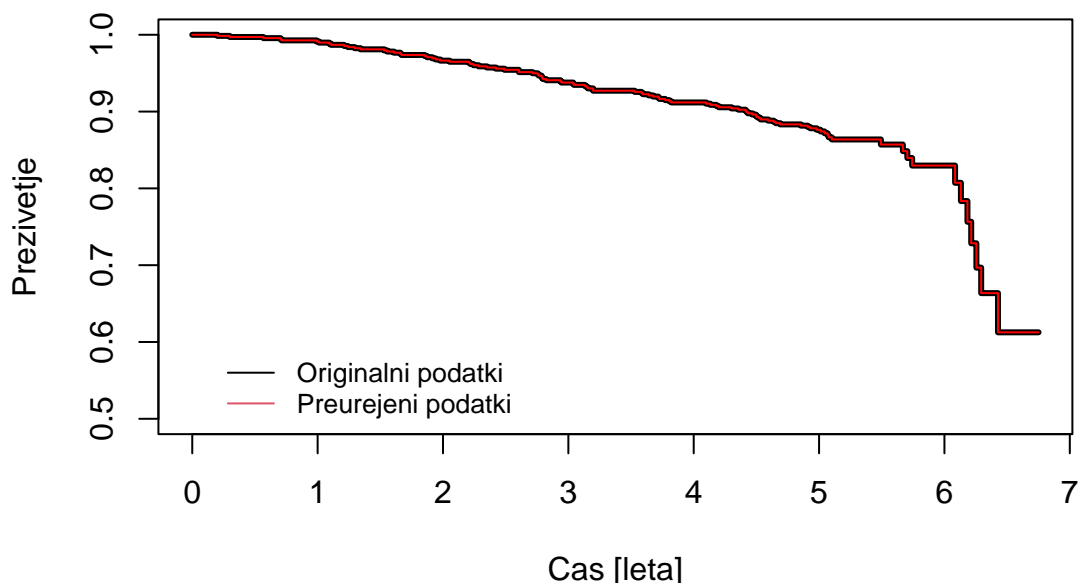
V podatkih bomo upoštevali spremenljivke: `mainid`, `cas_prve_kapi`, `kap_ind`, `cas_spremljanja`, in `smrt`. Podatke bomo razdelili na 3 dele in dodali novi spremenljivko `start` in `end`.

- Prvi del podatkov bo vseboval le tiste bolnike, ki tekom raziskave niso doživeli kapi (`kap_ind = 0`). Spremenljivka `start` je pri teh bolnikih enaka 0, torej čas začetnega pregleda pri zdravniku, spremenljivka `end` pa je enaka času smrti oz krnjenja tj. `cas_spremljanja`. Vsi ti bolniki imajo status kapi na celotnem intervalu  $[0, \text{time})$  enak 0.
- Drugi del podatkov vsebuje bolnike, ki so tekom raziskave doživeli kap (`kap_ind = 1`). Njihov začetni čas (`start`) je enak 0, spremenljivka `end` je enaka času kapi tj. `cas_prve_kapi`, indikator dogodka označuje krnjenje (`smrt = 0`) in spremenljivka `kap_ind` na tem intervalu zavzame vrednost 0, saj ni prišlo do kapi.
- Tretji del podatkov prav tako vsebuje bolnike s kapjo, le da se začne z dnem kapi in konča s koncem opazovanja, torej zavzema obdobje po prvi kapi. Začetni čas `start` je torej enak času prve kapi, spremenljivka `end` pa enaka času konca spremljanja tj. `cas_spremljanja`, spremenljivka `kap_ind` je enaka 1. Spremenljivka `smrt` ohrani vrednosti.

Na koncu dele podatkov združimo v skupni podatkovni okvir in jih uredimo po identifikacijski številki (`mainid`). Podatki za posameznika brez kapi so ostali enaki, posameznike s kapjo pa smo razbili v 2 vrstici. Prva vrstica opisuje interval do kapi in bolnik je ob koncu krnjen, druga vrstica pa opisuje interval po kapi in bolnikov status smrti je enak prvotnem.

Da smo podatke pravilno preoblikovali lahko preverimo, tako da na isto sliko narišemo enostavno Kaplan-Meier krivuljo preživetja na originalnih podatkih in ji dodamo krivuljo preživetja na preoblikovanih podatkih, kjer upoštevamo časovno odvisnost spremenljivke `kap_ind` prek vrednosti `start` in `end`. Krivulji se popolnoma prekrivata saj nismo dodali nobene dodatne informacije.

## Krivulja preživetja



S pravilno preoblikovanimi podatki sedaj lahko preverimo povezanost kapi in preživetja (dogodek je smrt ne glede na vzrok), kjer bomo upoštevali časovno odvisnost kapi. S Coxovim modelom bomo preverili domnevo, da kap in preživetje nista povezana.

```
## Call:
## coxph(formula = Surv(start, end, smrt) ~ as.factor(kap_ind),
##       data = podatki_urejeno_kap)
##
##               coef exp(coef) se(coef)      z      p
## as.factor(kap_ind)1 2.0607     7.8512   0.4271 4.824 1.4e-06
##
## Likelihood ratio test=13.83 on 1 df, p=2e-04
## n= 702, number of events= 97
## (15 observations deleted due to missingness)
```

Na podlagi vrednosti  $p$  ( $p < 0.05$ ) lahko zavrnemo ničelno domnevo, da pojav kapi in tveganje za smrt (ne glede na vzrok) nista povezana in sprejmemo sklep, da pojav kapi značilno poveča tveganje in s tem zmanjša čas preživetja. Razlika med zgornjima krivuljama preživetja, kjer kot dogodek upoštevamo tudi kap in dogodek zgolj s smrtjo je torej značilno različna.

Opazimo tudi izpis “(15 observations deleted due to missingness)”. To so posamezniki, ki imajo enaki vrednosti pri spremenljivkah `start` in `end`. 13 od teh predstavljajo posamezniki, ki so prišli zgolj na začetni pregled in jih kasneje h zdravniku več ni bilo, torej je njihov čas spremljanja enak 0, 2 posameznika pa imata čas prve kapi enak času smrti in je tako njihova dolžina intervala od kapi do smrti enaka 0. Takšne podatke imamo in z njimi moramo operirati, zato se za zgornji izpis ne obremenjujemo. Tega bi se sicer lahko rešili tako, da bi pri teh posameznikih za odtenek (npr.  $+0.00001$ ) podaljšali končni čas, ali pa skrajšali začetni čas. S tem bi se znebili “manjkajočih vrednosti” vendar pa nisem prepričan, če bi s tem res kaj pridobili, poleg tega bi pa spreminjali tudi dane podatke, za kar najverjetneje nismo pristojni.

Če kot dogodek poleg smrti upoštevamo še kap, je preživetje torej značilno nižje od preživetja, kjer je dogodek zgolj smrt. To je posledica značilne povezanosti kapi s preživetjem, ki poveča tveganje za smrt. Glavna razlika v interpretaciji je torej upoštevanje definicije dogodka in s tem opazovanje različnih “zaključkov” posameznika in s tem interpretacije “smrti”.

## 5. Povezanost diabetesa s časom preživetja brez srčno-žilnih zapletov

Preveriti želimo še povezanost diabetesa s časom preživetja brez srčno-žilnih zapletov. Velja, da je prisotnost diabetesa časovno odvisna spremenljivka. Pri nekaterih posameznikih je njena vrednost dana ob času 0, saj diabetes dobijo že pred začetkom raziskave, pri nekaterih posameznikih pa do diagnosticiranja pride tekom raziskave. Diabetes bi tako lahko obravnavali kot časovno odvisno spremenljivko, pri čemer se veliko “premikov vrednosti” zgodi že pred začetkom raziskave, drugi “premiki vrednosti” pa se zgodijo tekom raziskave. Podobno kot v prejšnjem primeru pri spremenljivki `kapi`, je podatke za analizo potrebno preurediti.

Ker nas zanima čas preživetja brez srčno-žilnih zapletov, bomo v sestavljenem izidu kot dogodek upoštevali smrt iz drugih vzrokov, smrt iz srčno-žilnih vzrokov in kap. Indikator tega dogodka je že ustvarjena spremenljivka `ind_sz`, spremenljivka `cas_sz` pa označuje čas krnjenja oziroma dogodka, ki ga določa spremenljivka `ind_sz`.

Kot glavni spremenljivki bomo upoštevali že omenjeni `ind_sz` in `cas_sz`, katerima bomo dodali novo spremenljivko `diabetes_TF`, ki bo indikator prisotnosti diabetesa tekom raziskave. Dodali bomo še spremenljivko `cas_diabetes`, ki bo zavzela čas diagnosticiranja diabetesa po začetnem pregledu, v kolikor je bil ta diagnosticiran med raziskavo.

Podatke bomo ponovno razdelili na 3 dele in dodali spremenljivki `start` in `end`.

- Prvi del podatkov bo vseboval le tiste bolnike, katerim ni bil diagnosticiran diabetes (`diabetesTF = 0`). Spremenljivka `start` je pri teh bolnikih enaka 0, torej čas začetnega pregleda pri zdravniku, spremenljivka `end` pa je enaka času smrti oz krnjenja tj. `cas_sz`. Vsi ti bolniki imajo status diabetesa na celotnem intervalu  $[0, \text{time})$  enak 0.
- Drugi del podatkov vsebuje bolnike, katerim je bil diagnosticiran diabetes (`diabetesTF = 1`). Njihov začetni čas (`start`) je enak 0, spremenljivka `end` je enaka času diagnosticiranja (od začetka raziskave naprej) tj. `cas_diabetes`, indikator dogodka označuje krnjenje (`ind_sz = 0`) in spremenljivka `diabetesTF` na tem intervalu zavzame vrednost 0.
- Tretji del podatkov prav tako vsebuje bolnike z diabetesom, le da se začne z dnem diagnosticiranja in konča s koncem opazovanja, torej zavzame obdobje po diagnosticiranju diabetesa. Začetni čas `start` je torej enak času diagnosticiranja, spremenljivka `end` pa enaka času konca spremljanja tj. `cas_sz`, spremenljivka `diabetesTF` je enaka 1. Spremenljivka `ind_sz` ohrani vrednosti.

Na koncu dele podatkov združimo v skupni podatkovni okvir in jih uredimo po identifikacijski številki (`mainid`). Podatki za posameznika brez diabetesa so ostali enaki, posameznike z diabetesom pa smo razbili v 2 vrstici. Prva vrstica opisuje interval do diagnosticiranja in bolnik je ob koncu krnjen, druga vrstica pa opisuje interval po diagnosticiranju in bolnikov status dogodka je enak prvotnem.

Ves čas sem kolebal kako obravnavati posameznike, katerim je bil diabetes diagnosticiran že pred samim začetkom raziskave. Samega časa prisotnosti njihovega diabetesa v raziskavo ne morem vključiti saj s tem zanemarim tiste, ki so tudi imeli diabetes in že pred začetkom raziskave umrli, tako bi ohranil samo tiste “najmočnejše” kar seveda ni prav, saj bi avtomatsko podaljšal čas preživetja in rezultat bi bil pristranski.

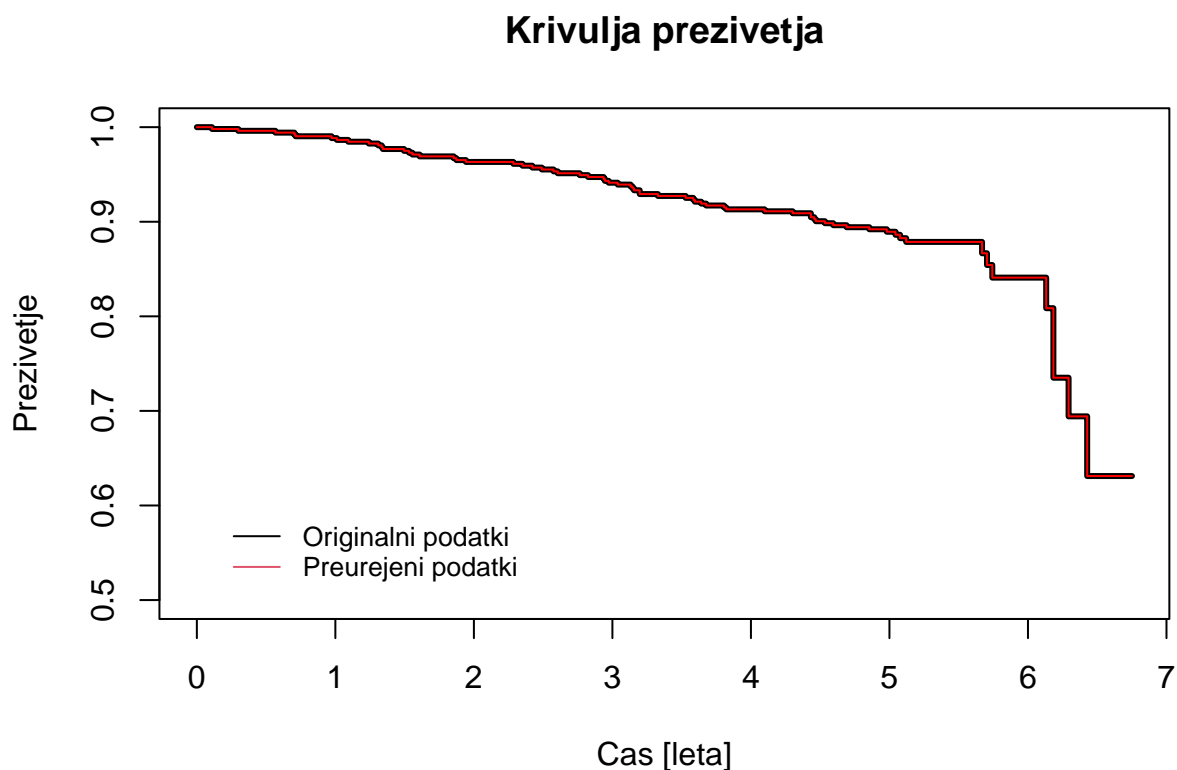
Posameznike z diabetesom že pred raziskavo sem sprva obravnaval, kot da so diabetes dobili v času 0 in jih ohranil v raziskavi. Posledično so bili njihovi intervali vmesnih časov, od začetka raziskave do diagnosticiranja diabetesa enaki 0. Podobno kot pri prejšnji obravnavi povezanosti časa preživetja s kapjo, se je pri obravnavi povezanosti diabetesa s časom preživetja pojavil izpis “185 observations deleted due to missingness”. To je posledica 170 posameznikov, katerim je bil diabetes diagnosticiran pred raziskavo. Algoritem je tako sam izbrisal posameznike z diagnosticiranim diabetesom pred začetkom raziskave (dodani so seveda posamezniki s samo enim pregledom).

V tem primeru, smo na podlagi vrednosti  $p$  ( $p < 0.01$ ) zavrnil domnevo, da diabetes in čas preživetja brez srčno-žilnih vzrokov nista povezana in sprejeli sklep, da diabetes značilno poveča tveganje in s tem zmanjša

čas preživetja. Diabetes, ki smo ga obravnavali enakovredno glede na diagnosticiranje pred raziskavo ali tekom raziskave, tako statistično značilno zmanjša čas preživetja brez srčno-žilnih zapletov.

Naposled sem se odločil, da posameznike z diagnosticiranim diabetesom pred samo raziskavo iz analize izključim kar sam. V obravnavani analizi sem tako ohranil samo posameznike, katerim je bil diabetes diagnosticiran tekom raziskave in posameznike brez diabetesa. Če bi jim tako kot prej, nastavil čas diagnoze na 0, bi podcenil preživetje z diabetesom in pristop ponovno ni pravilen. Pristop, da jih iz analize izključim se mi je zdel še najmanj napačen saj se znebim podcenitve in precenitve preživetja in še večje napake ne morem narediti. Prav tako pa jih v nasprotnem primeru algoritem odstrani sam.

Preverimo ali smo podatke pravilno preoblikovali.



S Coxovim modelom preverimo domnevo o ničelni povezanosti diabetesa in časa preživetja brez srčno-žilnih zapletov.

```
## Call:
## coxph(formula = Surv(start, end, ind_sz) ~ as.factor(diabetesTF),
##       data = podatki_urejeno_diabetes)
##
##               coef exp(coef) se(coef)      z      p
## as.factor(diabetesTF)1 0.4310    1.5388   0.5196 0.83 0.407
##
## Likelihood ratio test=0.61  on 1 df, p=0.4348
## n= 546, number of events= 66
## (12 observations deleted due to missingness)
```

Na podlagi vrednosti  $p$  ( $p = 0.4$ ) ne moremo zavrniti ničelne domneve, da diabetes in čas preživetja brez srčno-žilnih vzrokov nista povezana. Ne moremo sprejeti sklepa, da diabetes značilno poveča tveganje in s tem zmanjša čas preživetja, kakor bi pričakovali.

Poskusimo razložiti. Tekom raziskave je bil diabetes diagnosticiran pri 28 posameznikih, od katerih “umrejo” zgolj štirje. Vsem štirim je bil diabetes diagnosticiran okoli enega leta po začetnem pregledu (380 dni), kasneje pa so v povprečju živeli še 4 leta. Preživeli posamezniki z diagnosticiranim diabetesom so v povprečju po začetnem pregledu živeli še 5 let in diabetes jim je bil diagnosticiran dve leti in pol po začetnem pregledu. Posamezniki brez diagnosticiranega diabetesa so v povprečju po začetnem pregledu “živeli” še 4.7 let.

Torej, če združimo, so posamezniki z diabetesom “preživeli” 5.1 leto in umrlo jih je 15%. Posamezniki brez diabetesa so po začetnem pregledu v povprečju “živeli” še 4.7 let in umrlo jih je 12%. Vidimo torej, da so razlike v preživetju med diabetiki in ne diabetiki resnično majhne, poleg tega pa so umrli diabetiki le štirje oziroma 28 diabetikov. Zaradi majhnih razlik in majhnega vzorca (nadalje podvzorca umrlih) test ni mogel zavrniti ničelne domneve o nepovezanosti diabetesa in časa preživetja brez srčno-žilnih zapletov.



## 6. Zaključek

Raziskovali smo podatke o bolnikih, pri katerih je zdravnik odkril simptome srčno-žilnih bolezni. Zanimal nas je čas preživetja brez srčno-žilnih zapletov. Na zastavljena vprašanja smo odgovarjali s pomočjo metod analize preživetja.

Podatki so vsebovali informacije o 700 bolnikih, ki so bili predstavljeni v 20 spremenljivkah. Začetek spremljanja posameznika je bil z začetnim pregledom, kjer so bili ugotovljeni simptomi srčno-žilnih bolezni. Do konca spremljanja pa je prišlo v primeru smrti ali krnjenja. Informacije o stanju posameznikov so dane ob časih obiskov pri zdravniku in za vsakega posameznika je tako zadnja razpoložljiva informacija dana ob zadnjem pregledu pri zdravniku, zadnji kapi ali nazadnje smrti.

V uvodu smo večino podatkov opisno in grafično kratko predstavili. Osredotočili smo se predvsem na spremenljivke, za katere smo na podlagi vnaprej danih vprašanj menili, da bi nas utegnile zanimati in nam dati boljši pregled na danih podatkih. Veliko dela smo imeli predvsem s čiščenjem podatkov in njihovo preureditvijo v formate in okvirje kakršni so nam tekom raziskovanja ustrezali. Sprva je bil posameznik predstavljen v več vrsticah, vsak njegov pregled je predstavljal svojo vrstico, kasneje pa smo podatke preuredili, tako da je vsakega posameznika predstavljala zgolj ena vrstica.

Podatki so vsebovali podatke o 700 bolnikih, ki so bili predstavljeni v 20 spremenljivkah. Začetek spremljanja posameznika je bil z začetnim pregledom, kjer so bili ugotovljeni simptomi srčno-žilnih bolezni. Do konca spremljanja pa je prišlo v primeru smrti ali krnjenja. Informacije o stanju posameznikov so dane ob časih obiskov pri zdravniku in za vsakega posameznika je tako zadnja razpoložljiva informacija dana ob zadnjem pregledu pri zdravniku, zadnji kapi ali nazadnje smrti.

Nekaj manj kot tretjina vseh posameznikov je bila moškega spola. Stari so bili od 42 do 84 let s povprečjem 68 let. Podani so bili podatki o diagnosticiranju diabetesa, kjer je bil diabetes pri nekaterih diagnosticiran že pred začetkom raziskave, pri nekaterih pa tekom same preiskave. Pri nekaterih posameznikih je tekom raziskave prišlo do kapi, pri nekaterih do več teh. V povprečju je bil posameznik na 5.43 pregledih pri zdravniku v razponu 4.74 let.

Do konca udejstvovanja posameznika v raziskavi zaradi smrti je prišlo pri 97 posameznikov in smrti smo delili na posledice srčno-žilnega vzroka (35%) in drugega vzroka (65%). Velikih razlik v času smrti od začetnega pregleda med srčno-žilnimi vzroki in drugimi vzroki ni in so precej enakomerno porazdeljeni na časovnem intervalu raziskave. Pri 603 posameznikih je zadnji zabeležen stik zadnji pregled pri zdravniku (oziroma kap).

Tekom celotne raziskave je bila glavna predpostavka analize preživetja, predpostavka neinformativnega krnjenja, izpolnjena. Sprva je bil dogodek na katerega smo se osredotočili smrt, ki jo nismo delili na vzrok. Zanimal nas je čas preživetja posameznika. Ocenili smo Kaplan-Meier krivuljo preživetja, kjer smo posameznike, za katere ni bila dana informacija o smrti krnili. Tveganje je bilo v prvih letih relativno nizko in precej konstantno, po nekje petem letu pa je krivulja preživetja strmo padla in verjetnost preživetja 6.5 let po začetnem pregledu je 61%. Na tem primeru smo tudi razložili vpliv krnjenja in videli, da krnjenih posameznikov ne smemo preprosto izločiti iz raziskave.

Sledeče smo preučili povezanost spola s preživetjem. Na hiter pogled je veljalo, da se pri moških napram ženskam za odtenek več smrti zgodi v prvi polovici časovnega intervala in nekoliko več krnjenj proti koncu. Zaradi nekoliko večjega deleža smrti pri moških, je krivulja preživetja moških skozi celoten časovni interval nekoliko nižja od krivulje žensk. S testom log-rank smo preverili domnevo o nepovezanosti spola in preživetja ter je nismo uspeli zavrniti. S tem nismo sprejeli sklepa, da je spol povezan s časom smrti.

Tvorili smo tudi multivariatni model, kjer smo poleg spola pri preživetju upoštevali tudi starost. Velja sicer, da so moški ob prvem pregledu oz. prvih znakih simptomov srčno-žilnih bolezni, v povprečju mlajši kot ženske. Z multivariatnim Coxovim modelom smo preverili ali je starost povezana s tveganjem neodvisno od spola (in obratno). Koeficient starosti v modelu je bil statistično značilen in sprejeli smo sklep, da je tveganje za smrt je pri dveh posameznikih enakega spola ob ugotovitvi srčno-žilnih simptomov značilno večje pri starejšem posamezniku.

V nadaljevanju smo razložili zakaj ne moremo gledati preživetja, kjer nas zanima le smrt iz srčno-žilnih vzrokov. Pokazali smo, da krnjenje smrti iz drugih vzrokov lahko povzroči pristranske rezultate in pride do informativnega krnjenja. To smo skušali razložiti pred spremenljivke prisotnosti diabetesa. Gre za pojav sotveganja, kjer moramo namesto funkcije preživetja obravnavati funkcijo pojavnosti.

V namen odgovora na čas preživetja brez srčno-žilnih zapletov smo kot dogodek upoštevali sestavljeni dogodek, ki je vseboval informacijo o smrti zaradi srčno-žilnih vzrokov, drugih vzrokov ali prve kapi. Ostale posameznike smo krnili. Povezanost kapi s preživetjem smo preverili s Cox-ovim modelom, kjer smo kap upoštevali kot časovno odvisno spremenljivko. V ta namen smo morali podatke nekoliko preurediti, tako da so zajemali informacijo o posameznikovem stanju pred kapjo in stanju po kapi. Zavrnilo smo domnevo, da pojav kapi in tveganje za smrt (ne glede na vzrok) nista povezana in sprejeli sklep, da pojav kapi značilno poveča tveganje in s tem zmanjša čas preživetja.

Nazadnje pa smo poskusili še razložiti povezanost diabetesa s časom preživetja brez srčno-žilnih zapletov. Precej vprašanj se nam je pojavljalo v povezavi s tem kako spremenljivko indikatorja diabetesa sploh obravnavati. Nekateri posamezniki so ga namreč imeli že pred samim začetkom raziskavi, nekaterim pa je bil diagnosticiran tekom raziskave. Odločili smo se, da bomo diabetes obravnavali kot časovno odvisno spremenljivko in posameznike z diagnosticiranim diabetesom že pred raziskavo iz analizi izključili. S Cox-ovim modelom smo preverili povezanost diabetesa s časom preživetja brez srčno-žilnih zapletov in nismo zavrnilo ničelne domneve o nepovezanosti diabetesa s časom preživetja brez srčno-žilnih zapletov. Diabetes torej, po naši raziskavi zanimivo ne zmanjša časa preživetja brez srčno-žilnih zapletov, vsaj tako ne moremo trditi. Razlog je najverjetneje v majhnem vzorcu diabetikov in morda posledica nekoličnega naključja.