

Elukestvus

Iseseisev praktikum ja
kodune töö (lõpus)

Paljud elukestvusanalüüsi jaoks vajalikud käsud on R-is koondatud lisamoodulisse survival (see lisamoodul paigaldatakse arvutisse enamasti juba R-i enda installeerimise ajal). Võtame vajaliku lisamooduli kasutusse:

```
library(survival)
```

Loeme sisse ka näidisandmestiku (kuna patsientide andmeid ei tohi enamasti niisama avalikult jagada, siis piirdume antud näites R-i endaga kaasatuleva näiteandmestikuga):

```
data(cancer)  
cancer[1:3,]
```

Mõnede tunnuste tähendused antud andmestikus (kopsuvähiga patsiendid):

time: Aeg päevades diagnoosist
status: Mis juhtus patsiendiga: 1=tsenseerimine, 2 = surm
age: Patsiendi vanus aastates diagnoosi saamise ajal
sex: Patsiendi sugu: mees=1, naine=2

Hindame Kaplan-Meieri elukestvuskõvera kopsuvähiga patsientidele (koos usaldusintervalliga):

```
attach(cancer)  
m1=survfit(Surv(time, status==2)~1)  
plot(m1)
```

```
summary(m1)
```

jälgimisaeg

Sündmus jälgimisaja lõpus:

0 – tsenseerimine

1 – surm

Teisendamaks kahed (algandmetes 2=surm)

ühteks kasutame käsku status==2.

Elukestvuskõvera hinnang ka kenasti
numbriliselt väljatrükitud

Ülesanne 1

[enese kontroll – tulemust esitama ei pea](#)

Eelmises praktikumis arvasime järgmiste eluigade jaoks Kaplan-Meieri elukestvuskõvera:

5 9+ 12 12 14+ 18+ 18 19 22 25

Leia ja joonista samade andmete põhjal elukestvuskõver ka R-is!

Logrank-testi abil sai võrrelda, kas elukestvused kahes (või enamas) grupis on samasugused (H_0) või mitte. Võrdleme näiteks, kas nais- ja meespatsientide elukestvused on kopsuvähi korral samasugused või mitte:

```
m2=survfit(Surv(time, status==2)~sex)
plot(m2, col=c("blue", "red"))
legend("topright", c("Mehed", "Naised"), lwd=2,
      col=c("blue", "red"))
survdif(Surv(time, status==2)~sex)
```

Ülesanne 2

enesekontroll – tulemust esitama ei pea

Võrdle, kas kopsuvähi korral patsientidel kes on nooremad kui 60a võiks olla teistsugune elukestvuskõver võrreldes patsientidega kes on vanemad kui 60a?

Cox'i võrdeliste riskide mudel

Millise jaotusega on eluead T ? Seda on vahel väga raske oletadagi. Sestap armastatakse elulemusanalüüsi tehes mitteparameetrilisi meetodeid (näiteks Kaplan-Meieri meetod elulemuskõvera hindamiseks). Keerukamate mudelite hindamisel tuleb siiski mingeid eelduseid teha. Üheks armastatud elulemusanalüüsis kasutatavaks mudeliks on Cox'i võrdeliste riskide mudel (Cox proportional hazards model):

$$h(t) = h_0(t) \exp(c_1x_1 + c_2x_2 + \dots)$$

mis kirjeldab, kuidas mõjutavad surmariski $h(t)$ erinevad selgitavad tunnused x_1, x_2 jne. Baasriskifunktsioon $h_0(t)$ kirjeldab surmariski võrdlusgrupis (kontrollgrupis) ja see võib olla suvaline. Sõltumatud tunnused aga võivad muuta surmariski kontrollgrupi omast suuremaks või väiksemaks. Näiteks kui meil on ainult üks sõltumatu tunnus x , siis Cox'i mudel omandab kuju

$$h(t) = h_0(t) \exp(cx)$$

Kui $x=0$ (kontrollgrupp) siis risk surra ajahetkel t on selle mudeli järgi

$$h(t) = h_0(t) \exp(c \cdot 0) = h_0(t)$$

ja töötlust saanute grupis ($x=1$) oleks samal ajal risk surra

$$h(t) = h_0(t) \exp(c \cdot 1) = h_0(t) \exp(c)$$

ehk $\exp(c)$ korda suurem. Cox'i mudel eeldab, et surmarisk on teises grupis alati, igal ajahetkel, täpselt $\exp(c)$ korda suurem.

Hindame ühe Cox'i mudeli naiste ja meeste surmariskide erinevuste kirjeldamiseks:

```
m3=coxph(Surv(time, status==2)~factor(sex))
summary(m3)
```

Mitu korda on meespatsientide surmarisk naiste surmariskist (igal ajamomendil) suurem?

Lisaks peaksime kontrollima, kas võrdeliste riskide eeldus on ikka täidetud:

```
cox.zph(m3)
```

Antud testi puhul arvab nullhüpotees, et kehtib võrdeliste riskide eeldus (suur p-väärtus GLOBAL real näitab, et kõik võrdeliste riskide eeldusega/eeldustega on mudelis kõik hästi).

Vahel kohtame olukorda, kus võrdeliste riskide eeldusega pole kõik korras, vaata näiteks sellist mudelit:

```
> m3a=coxph(Surv(time, status==2)~factor(sex)+meal.cal)
> cox.zph(m3a)
```

	chisq	df	p
factor(sex)	1.45	1	0.229
meal.cal	4.76	1	0.029
GLOBAL	6.49	2	0.039

Pisike GLOBAL p-väärtus (0,039) näitab, et Cox'i võrdeliste riskide eeldus pole täidetud. Sellistes situatsioonides on abiks eelnevatest tunnuste kaupa tehtud testidest – näeme, et võrdeliste riskide eeldus ei kehti tunnuse meal.cal (lõunasöögi ajal tarbitud kalorid, p-väärtus: 0,029) puhul. Võrdeliste riskide mudel eeldab, et rohkem söövatel patsientidel on alati (igal päeval) väiksem risk surra (1kcal rohkem süües on umbes 20% väiksem risk surra mingil konkreetsel päeval). Tegelik seos pole aga selline – enne surma kaob inimestel lihtsalt söögiisu, ja seega viitab väike söögiisu võimalusele, et inimene sureb lähipäevadel. Pikas perspektiivis – mis juhtub näiteks kahe aasta pärast – on mõõdukalt sööjatel hoopis paremad võimalused elus püsida võrreldes praeguste õgarditega. Seega rohkem sööjate risk pole mitte alati – igas „vanuses“ – sama arv kordi väiksem võrreldes vähemsööjatega. Seega tuleks antud juhul tuleks kasutada alterntiivseid meetodeid sellise mudeli hindamiseks.

Märkus:

Cox'i mudel üldiselt ei võimalda välja joonistada elukestvusfunktsiooni/riskifunktsiooni graafikuid – mudel tunneb huvi vaid selle vastu, kelle surmarisk on suurem või väiksem (ja mitu korda). Soovi korral on võimalik siiski (teatud trikke kasutades) hinnata ka Cox'i mudelit kasutades elukestvuskõveraid. Selleks on siiski tarvis veel ühte täiendavat lisamoodulit (pec):

```
# install.packages("pec")
library(pec)

aeg=0:1000
m3=coxph(Surv(time, status==2)~factor(sex), x=TRUE)

progN=predictSurvProb(m3, newdata=data.frame(sex=2), times=aeg)
progM=predictSurvProb(m3, newdata=data.frame(sex=1), times=aeg)

plot(aeg, progN, col="red", type="l", ylim=c(0,1))
lines(aeg, progM, col="blue")
```

Saadud elukestvuskõverad on mõnevõrra erinevad võrreldes Kaplan-Meieri meetodil leitud elukestvuskõveratest võrdle saadud tulemust varem saadud (m2) jonnisega. Antud joonis on koostatud eeldades, et võrdeliste riskide eeldus kehtib (mehed surevad igal päeval peale diagnoosi saamist veidi suurema tõenäosusega kui naised). Kui antud eeldus tegelikult kehtib, võiksid siintoodud joonised olla veidi täpsemad võrreldes Kaplan-Meieri meetodil saadud joonistega. Kui võrdeliste riskide eeldus aga tegelikult ei kehti, siis on siin saadud joonised pigem eksitavad.

Kodune töö
Iseseisvad ülesanded (saada vastused õppejule!)

Loe sisse kodutöö andmestik:

```
print(load(url("https://www-1.ms.ut.ee/mart/biostat/biostat1.RData")))
```

Tunnuste tähendused:

aeg	- aeg diagnoosist (päevades)
syndmus	- kas eelmises tunnuses märgitud aja möödudes inimene suri (1) või tsenseeriti (0)
sugu	- kodeeritud: 0 – naine; 1 – mees;

1. Hinda elukestvuskõver Kaplan-Meieri meetodil nii naiste kui meeste jaoks (lisa joonis).
2. Milline on elukestvuse mediaan naiste jaoks, milline meeste jaoks?
3. Kas Kaplan-Meieri meetodil hinnatud meeste ja naiste elukestvuskõverad on tõestatavalt erinevad?
4. Hinda Cox'i võrdeliste riskide mudel kus sõltumatu tunnuseks on inimese sugu. Interpreteeri hinnatud mudeli parameetreid, selgita nende tähendust.
5. Võrdle saadud tulemust ka järgneva analüüsi tulemustega:

```
surmaajad=andmed[andmed$syndmus==1,]  
t.test(aeg~sugu, data=surmaajad)
```

Kas ja kuidas varem leitud tulemused ja siinse t-testi tulemused omavahel klapivad?
Kas tulemused on sarnased või mitte? Miks?