

Kodune töö
Iseseisvad ülesanded (saada vastused õppejule!)

Ülesanne 1 – Sissejuhatus. Elulemusanalüüsist üldiselt

Loe sisse esimene 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 tunnusena on sees 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?

(jätkub...)

Ajast sõltuvad tunnused ja mudeli eeldused

Alustame ühest näitest. Vaata järgmist andmestikku *pisi*, milles on järgmised tunnused:

id – inimese identifikaator

vanus – inimese vanus uuringu lõpus või eluiga (jälgimisaeg)

suitsetamise_algus – millises vanuses hakkas suitsetama (kui ei hakanudki suitsetama, siis on tunnuse väärtuseks NA)

suri – kas vanus näitab inimese eluiga (*suri*=1) või lihtsalt inimese vanust uuringu lõpus (*suri*=0)

Näiteandmed ise saame luua järgmise käsu abil:

```
pisi=data.frame(id=1:6, vanus=c(25, 24, 22, 20, 32,33),
               suitsetamise_algus=c(NA, 16, NA,18, NA,30),
               sugu=c("N", "N", "M", "M", "N", "M"), suri=c(0,1,1,0,1,1))
```

Soovime inimeste jälgimisaja jagada eraldi perioodiks mil ta ei suitseta ja perioodiks, mil ta suitsetab. Teeme seda näiteks järgmise käsuga:

```
pisi2=tmerge(pisi, pisi, id=id, surm=event(vanus, suri),
            suitsetab=tdc(suitsetamise_algus))
```

Võrdle esialgset andmestikku ja tükkideks lõigatud andmestikku! Kas mõistad, mis on toimunud? Vaata kuidas suitsetajate (*id*=2,4,6) jaoks on tekkinud kaks kirjet!

pisi					
	id	vanus	suitsetamise_algus	sugu	suri
1	1	25	NA	N	0
2	2	24	16	N	1
3	3	22	NA	M	1
4	4	20	18	M	0
5	5	32	NA	N	1
6	6	33	30	M	1

pisi2									
	id	vanus	suitsetamise_algus	sugu	suri	tstart	tstop	surm	suitsetab
1	1	25	NA	N	0	0	25	0	0
2	2	24	16	N	1	0	16	0	0
3	2	24	16	N	1	16	24	1	1
4	3	22	NA	M	1	0	22	1	0
5	4	20	18	M	0	0	18	0	0
6	4	20	18	M	0	18	20	0	1
7	5	32	NA	N	1	0	32	1	0
8	6	33	30	M	1	0	30	0	0
9	6	33	30	M	1	30	33	1	1

Veendu, et mõlemad andmestikud annavad korrektse ja võrreldava analüüsi korral täpselt samad tulemused:

```
library(survival)
coxph(Surv(vanus, suri)~factor(sugu), data=pisi)
coxph(Surv(tstart, tstop, surm)~factor(sugu), data=pisi2)
```

Nende kahe analüüsi tulemused peaksid olema täpselt samad (ainult kasutatud vaatluste/kirjete arv muutub).

Teise andmestiku (kus oleme jälgimisajad jaganud mitmeks perioodiks) eeliseks on võimalus hinnata ka suitsetamise mõju:

```
coxph(Surv(tstart, tstop, surm)~factor(sugu)+suitsetab, data=pisi2)
```

Ülesanne 2

Järgmise käsu abil saad sisse lugeda andmestiku syda:

```
load(url("https://www-1.ms.ut.ee/mart/biostat/Elulemus3.RData"))
syda[1:3,]
```

Andmestikus olevad tunnused:

id – inimese id

vanus – inimese vanus siis, kui tekkis vajadus südant siirdada

jalgimisaeg – jälgimisaeg päevades peale haigestumist

suri – kas jälgimisaja lõpus inimene suri (1) või oli veel elus (0)

siirdamine_ootaeg – kui kaua pidi patsient siirdamiseks sobivat südant ootama
(NA kui südant ei siiratud)

siirdamine – kas patsiendile siirati süda (1) või südamesiirdamist ei toimunud (0)

Andmestiku põhjal tehti esmane analüüs, mille põhjal järeldati, et südamesiirdamine vähendab patsientide surmariski umbes 73% võrra:

```
m1=coxph(Surv(jalgimisaeg, suri)~factor(siirdamine), data=syda)
summary(m1)
```

Esialgset analüüsi kritiseeriti tugevalt – näiteks juhiti tähelepanu, et kehvemas seisundis patsiendid surevad enne, kui suudetakse leida nende jaoks sobiv süda.

Tee sama analüüs paremini! Ära võrdle neid patsiente, kellele siirdatakse kunagi süda nende patsientidega, kes surevad varakult, ilma sobivat doonorsüdant leidmata. Võrdle surmariski neil, kellel parasjagu on juba doonorsüda sees nende inimeste surmariskiga, kellel hetkel veel doonorsüdant sees pole! Kontrolli ka võrdeliste riskide eelduse täidetust ja tee vajadusel parandusi/täiendusi, nii et võrdeliste riskide eelduse pärast me muretsema ei peaks. Interpreteeri saadud mudelit.