

Lineaarsed mudelid
5. praktikum

Surm ja kärbsed.

On aasta 1825. Dekabristide ülestõus ja Hispaania revolutsioon on veriselt maha surutud, Lord Byron on surnud Kreeka vabadust kaitses. B. Gompertz leiab, et inimeste suremus kasvab eksponentsiaalselt. Täpsemalt, Gompertz väidab, et kehtib järgmine seos:

$$Suremus = c_0 e^{c_1 \cdot vanus}$$

Benjamini väide meeldis praktikutele. Elukindlustuse jaoks vajalike arvutuste tegemiseks hakati hoogsalt kasutama Gompertz'i mudelit. Lõpuks otsustas keegi ka Gompertzi teooria paikapidavust kontrollida. Tõsi, see kahtleja otsustas inimeste asemel katsejäänestena äädikakärbsed kasutada. Ta hankis 1 203 646 vastündinud äädikakärbest ja mõõtis kõigi nende eluead. Iga päeva kohta arvutati ka äädikakärbste suremus. Loeme sisse kogutud vaatlusandmed:

```
andmed=read.table("http://www.ms.ut.ee/mart/linmud2021/Surmjakarbes.txt",
                  header=TRUE)
head (andmed)
```

Ülesanne 1

Kuidas võiks hinnata Gompertz'i mudeli parameetreid c_0 ja c_1 lineaarse mudeli (lm-käsu) abil? Millise triki abil saaksime teisendada Gompertz'i mudeli lineaarseks mudeliks?

Gompertz'i Mudeli parameetrite hindamiseks vajalikud R-i käsud:

```
Mudel = lm( ..... )

# c0-i saad kätte nii:
.....

# c1 saad kätte nii:
.....

Ja hinnatud parameetrite väärtused:

c0 = .....

c1 =.....
```

(ülesanne jätkub)

Proovi kirjeldada, millised lineaarse mudeli eeldused võiksid olla rikutud (antud hetkel lihtsalt ignoreerime eelduste rikutust, aga me peame oskama näha probleemseid kohti oma analüüsis). Ära sellele küsimusele vastamiseks vaata diagnostilisi graafikuid või muud sellist (me pole neid nagunii veel kasutama õppinud), proovi arutleda loogiliselt/teoreetiliselt – mis võiks mudeli eeldustega viltu olla.

.....

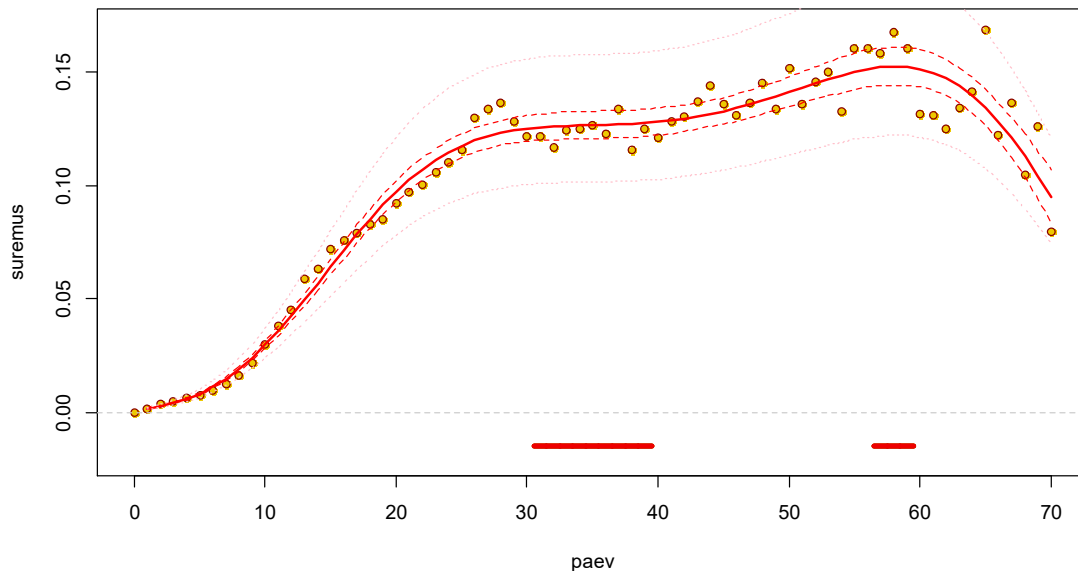
Joonista graafik, kus x -teljel oleks äädikakärbse vanus päevades ja y -teljel suremus. Kanna graafikule tegelikud vaatlused, sinu model ja 95%-usaldusintervallid mudelile (võid lisada veel ka prognoosintervallid). Hinda (visuaalselt) Gompertz'i mudeli sobivust. Kuidas tundub?

```
plot(paev, suremus)
....
```

(1. ülesande lõpp)

Alternatiiviks Gompertz'i mudelile oleks mudel, kus logaritmi suremusest prognoosime polünoomi abil (mistahes lõigul pidevat funktsiooni saab kuitahes hästi lähendada polünoomi abil...). Paku ka enda poolt välja mõni kärbeste suremust hästi kirjeldav polünoomiaalne mudel (m2). Uuri graafikult, kuidas sinu poolt valitud mudel sobib vaatlusandmetega.

Graafikut uurides võib näida, et on teatud vanused, mil suremus püsib muutumatuna – ei suurene ega vähene. Enda poolt hinnatud mudelit kasutades märgi graafikul ära kõik need piirkonnad, kus sinu mudeli arvates võiks suremus püsida muutumatuna. Näiteks nii:



Vajalike arvutuste tegemiseks võid kasutada lisamooduli `gmodels` funktsiooni `estimable`. Kuidas testida, kas y-tunnuse väärtused mingi x-tunnuse väärtuse korral ei muutu – ei suurene ega vähene?

Väga lihtne – meenuta, mida tead funktsiooni tuletise kohta. Seal piirkonnas, kus funktsioon on konstantne on funktsiooni tuletis 0. Seega peame lihtsalt `estimable`- käsu abil testima, kas hinnatud regressioonmudeli tuletis teatud päeva väärtuse korral võiks olla 0 või mitte.

Ülesanne 2

Testi ise (ikkagi R-i abil, aga ilma `estimable`-käsku kasutamata), kas 35. päeval võiks surmafunktsiooni tuletis olla 0.

Ülesande 2 lahenduses tahan näha:

1. Teie poolt hinnatud mudeli m_2 kuju;
2. `estimable` –käsku mille abil otsite esitatud küsimusele lahendust
3. Otsust – kas suremus võib 35 päeva vanuses olla muutumatu – mitte kasvada ega kahaneda?

Vabatahtlik lisaülesanne: joonista sarnane joonis nagu praktikumi materjalis tehtud – kus oleks joonisele ära märgitud piirkonnad, kus surmafunktsioon võiks olla konstantne – milliste vanuste korral suremus ei kasva ega kahane.