

Lineaarsed mudelid
2. praktikum
Suur raharööv

Loe sisse andmestik roov.txt, kättesaadav aadressilt www-1.ms.ut.ee/mart/linmud2025/roov.txt.

Sisselugemist saab teha näiteks järgmise käsu abil:

```
and=read.table(url("http://www-1.ms.ut.ee/mart/linmud2025/roov.txt"),
               header=T)
attach(and)
```

Andmestiku taustast:

Ühes linnas (New York) võitis üks firma (X) avalikul konkursil õiguse korraldada parkimistasude kogumist. Veidi aega hiljem tekkis aga linnaametnikel kahtlus, et sissetulekud parkimisest pole need, mis nad olema peaksid. Kontrollimise käigus õnnestus varguselt tabada 5 firma X töötajat, kõik kuritöölt tabatud mõisteti ka kohtu poolt süüdi. Seepeale otsustas linn parkimistasusid koguvalt firmalt nõuda sisse saamata jäänud tulu (firma X kohustuseks oli ka töötajate järelevalve). Kohus leidis linna nõude põhjendatud olevat ning käskis linnal esitada kalkulatsioon kahju võimaliku suuruse kohta.

Kahju suuruse leidmiseks ongi kasutada andmefail roov.txt, mis sisaldab järgmiseid tunnuseid:

AEG	– aeg kuudes: 1=Mai 1977, 47 = Märts 1981
RAHA	– linnale antud parkimistasu summa
LINN	– raekoja esiselt parkimisplatsilt usaldusväärsete linnaametnike poolt kogutud parkimistasu suurus
FIRMAX	– tunnus, mis näitab, kas antud kuul kogus parkimistasusid firma X või mitte (0-kogus keegi teine; 1- parkimistasusid kogus firma X).

Pilguheit andmetele:

```
plot(AEG, RAHA, type="l", xlab="Aeg", ylab="Laekunud raha", xaxt="n")
axis(1, at=seq(-5.5, max(AEG), 12), 1977:1981)
abline(v=range(AEG[FIRMAX==1]), lty=2)

points(AEG, RAHA, col=c("red", "blue")[2-FIRMAX], pch=21,
       bg=c("yellow", "skyblue")[2-FIRMAX], cex=1.2)

text(mean(AEG[FIRMAX==1]), 900000, "röövimiste \n ajajärk")
arrows(min(AEG[FIRMAX==1]), 800000, max(AEG[FIRMAX==1]), 800000,
       code=3, length=0.15)
```

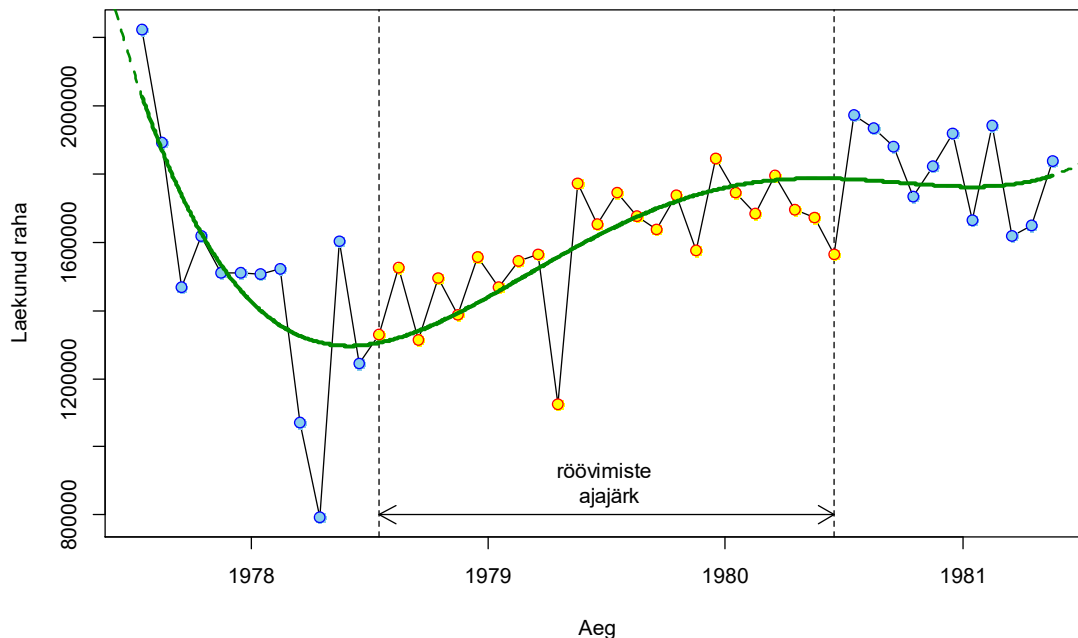
Tänane ülesanne

Leia hinnang röövitud parkimistasude kogusummale!

Vihjed

NB! Hea vastuseni on võimalik liikuda ka mõnda teist teed pidi!

1. Hinda model 0, mis kirjeldab, kuidas sõltuvad laekuvad parkimistasud ajast! Model ei pea olema perfektne – esialgne model ju ei arvesta nagunii röövimistega. Aga ta peaks ligilähedaselt kirjeldama seda, kuidas laekuvad parkimistasud ajas muutuvad. Võid näiteks kasutada polünoomi (mistahes lõigul pidevat funktsiooni saab kuitahes täpselt kirjeldada polünoomi abil... hästi, antud juhul pole laekumiste puhul tegemist pideva funktsiooniga, aga kui varguseid poleks olnud...)



Kõigi mudelite puhul soovitan mudeli prognoosid ka välja joonistada. Kui tahaksime tunnust y modelleerida polünoomi abil (sõltuvust x -i ja y -i vahel kirjeldab polünoom) siis saaksime seda näiteks teha nii:

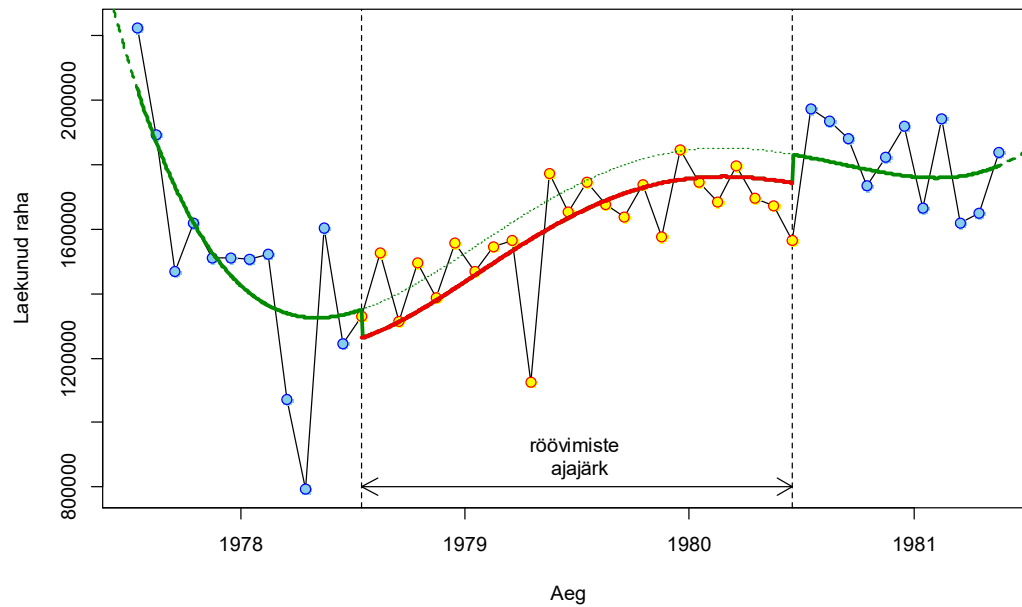
```
# Ruutpolünoom (parabool)
Mudel2=lm(y~x+I(x^2))
Mudel2a=lm(y~poly(x, 2, raw=TRUE))

# Kuuppolünoom
Mudel3=lm(y~x+I(x^2) +I(x^3))
Mudel3a=lm(y~poly(x, 3, raw=TRUE))
```

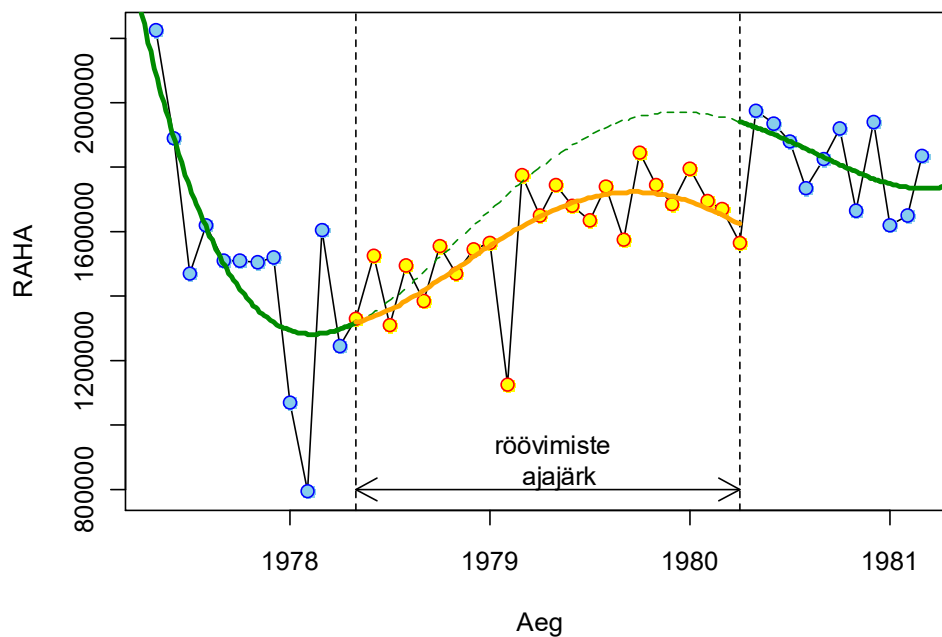
jne

Kuidas regressioonjoont graafikule (predict-käsku kasutades) joonistada, seda võid piiluda eelmise praktikumi materjalist. Hinnatud mudeleid saad näiteks võrrelda kas anova-käsu abil: `anova(Mudel2, Mudel3)` – testib, kas keerukam model on tõestatavalt parem; võid ka kasutada mudeli AIC –väärtust: `AIC(Mudel2, Mudel3)`, väiksem AIC-väärtus näitab parema prognoosivõimega mudelit (mõlemast lähenemisest tuleb pikemalt juttu järgnevatel loengutes)

2. Hinda röövimist kirjeldav mudel (mudel A). Luba laekuvate parkimistasude kõveral olla röövimiste ajajärgul madalamal. Näiteks eeldades, et ajajärgul kui parkimistasusid korjas firma X, läks iga kuu mingi summa kõrvale (osa raha rööviti)!



3. Mõtle, kas hinnatud mudel 2 kirjeldab röövlite tegevust adekvaatselt! Järsku röövlid ei varastanud igal kuul sama palju raha? Äkki röövlite isu kasvas hoopis tasapisi? Hinda mudel, mis lubaks röövlite saagihimul ajapikku suureneda (mudel B), vaata ka lisatud joonist:



4. Võrdle mudeleid A ja B, kumb tundub adekvaatsem? Soovi korral täienda mudelit ka muude tunnustega. Millise mudeli juurde lõpuks jääd?

5. Leia, millised oleksid pidanud olema laekumised siis, kui röövimisi poleks toimunud (koos 95%-usaldusintervalliga). Isegi kui algselt leiad vastuse sellele küsimusele mingil muul moel, siis proovi sellele küsimusel vastata ka estimable-käsu abil (vaata näiteks eelmist praktikumi)

Oletuslik röövitud rahasumma:

95%-usaldusintervall röövitud summale: