

Lineaarsed mudelid Praktikum

Näide 1 – Raha lõhn

Aktsiatega spekuloides on vahel kasulik tulevikku ette aimata. Näiteks seda, kui palju reisijaid võiks Tallink selle aasta maikuus vedada? Teeme katset soovitud numbrit prognoosida.

Loeme sisse ajaloolised Tallinki reisijate arvud

```
load(url("http://www.ms.ut.ee/mart/linmud2018/Tallink.RData"))
```

ja heidame kiirpilgu andmetele:

```
Tallink[1:4,]  
attach(Tallink)  
plot(aeg, reisijaid, pch=21, bg="gray70")
```

Proovime antud andmeid lähendada esmalt kolmel viisil: polünoomide abil, b-splaine kasutades ja Fourier' teisenduse abil saadud abitunnuseid („sessoonsuse modelleerimine“) kasutades:

```
m1=lm(reisijaid~poly(aeg, ...))  
library(splines)  
m2=lm(reisijaid~bs(aeg, ...))  
  
m3=lm(reisijaid~I(cos(1*2*pi*aeg/12))+I(sin(1*2*pi*aeg/12))+  
      I(cos(2*2*pi*aeg/12))+I(sin(2*2*pi*aeg/12))+ ... )
```

Polünoomi otsimist võid alustada umbes 18.-järku polünoomidest, b-splaini otsimisel sobib ehk samuti alguspunktiks 18 vabadusastet.

Kanna kõigi mudelite prognoosid ka graafikule:

```
xx=seq(0.5, 72.5, length=1000)  
y1=predict(m1, data.frame(aeg=xx))  
...
```

Milline proovitud mudelitest sobib kõige paremini andmeid kirjeldama (AIC väärtuse põhjal otsustades)?

Millised tulevad mainitud mudelite prognoosid Tallinki poolt veetavate reisijate koguarvule aprilliks 2018? Millised neist prognoosidest tunduvad usaldusväärsed, millised mitte?

Võrdle oma prognoosi võrrelda ka tegeliku reisijate arvuga:

<https://cns.omxgroup.com/cdsPublic/viewDisclosure.action?disclosureId=838074&messageId=1053797>

Antud andmete puhul on tegemist aegreaga. Lineaarse mudeli üheks eelduseks on vaatluste sõltumatus. Kas on võimalik, et meie mudelite jäägid on teineteisest sõltumatud? Kasutame Durbin-Watson'i testi selle kontrollimiseks:

```
library(lmtest)
dwtest(m1)
dwtest(m2)
dwtest(m3)
```

Millised on tulemused, kas need on sellised, nagu ootasid?

Näide 2 – tagasi kasvukõverate juurde!

Mitte väga ammu, nii umbes 8. praktikumis joonistasime laste jaoks kasvukõveraid. (Varasemaid loenguid/praktikume saab hetkel näha aadressilt: <http://utitsucks.esy.es/linmud2017/html/>)

Mudeleid sobitades võisime tähele panna, et mudelite jäägid ei olnud päriselt normaaljaotusega. Naase nüüd nendesamade andmete juurde. Hinda mudel lapse kaalu kirjeldamiseks. Veendu, et meil on mudeli eeldustega probleeme (näiteks diagnostilisi graafikuid vaadates või paganatesti kasutades – `bptest(mudel)`), lisamoodulist `lmtest`) ja otsi Box-Cox'i transformatsiooni abil sobivat transformatsiooni uuritavale tunnusele. Proovi joonistada kasvukõverate ja prognoosiintervallide graafik kasutades nii transformeerimata kui ka transformeeritud lapse kaalu.

Mille poolest esialgset, transformeerimata tunnuse abil leitud prognoosiintervallid erinevad transformeeritud tunnuse abil leitud prognoosiintervallidest?