

Aegride analüüs

Praktikum nr. 2, 2016

Käsitletavad teemad: Andmete sisselugemine (kordus). Aegrea silumine libisevate keskmiste abil. Aegride osadeks lahutamine.

1 Andmete sisselugemine

Ülesanne 1. Laadige alla kursuse kodulehel olev andmestik `TU121.csv`, mis sisaldab Eestis majutatud turistide andmeid ning lugege need R keskkonda sisse, tekitades aegrea nimega `turistid`.

2 Andmete silumine libisevate keskmistega

Rea (z_t) teisendust kujul

$$y_t = \sum_{i=-q}^r w_i z_{t-i},$$

kus $w_i \geq 0$ ja $\sum_{i=-q}^r w_i = 1$, nimetatakse **libiseva keskmise** leidmiseks. Libisevat keskmist kasutatakse sageli müra eemaldamiseks ning reast trendikõvera väljaeraldamiseks.

Libisevat keskmist saab R keskkonnas arvutada `filter` käsuga. Kui on tegemist lõpliku arvu nullist erinevate kordajatega (nt lihtsa libiseva keskmisega), siis tuleb kaalud anda ajas tagurpidi järjekorras (st esimesena kõige hilisema vaatluse kaal, siis eelmise jne) suvandiga `filter=kaalude_vektor`. Üheks võimaluseks kaalude vektor tekitada on kaalud kirja panna funktsiooni `c()` argumentidena, näiteks `filter=c(1/3,1/3,1/3)`. Ainult minevikuväärtusi kasutava keskmistamise korral tuleb lisaks kasutada veel suvandit `sides=1`, vaikumisi ja suvandi `sides=2` korral loetakse hetkemoment vastavaks kaalude vektori keskmisele elemendile (kui kaalude vektoris on paaritu arv elemente) ning keskmise elemendi puudumisel (st kui kaalude vektor sisaldab paarisarv elemente) kasutatakse tulevikus ühe võrra rohkem vaatluseid kui minevikus. Näiteks `y=filter(z,filter=c(1/3,2/3))` vastab teisendusele $y_t = \frac{1}{3}z_{t+1} + \frac{2}{3}z_t$ ning `y=filter(z,filter=c(1/3,2/3),sides=1)` vastab teisendusele $y_t = \frac{1}{3}z_t + \frac{2}{3}z_{t-1}$.

Ülesanne 2. Kasutage majutatud turistide andmestikku. Leidke 4-kuuline lihtne minevikuandmeid kasutav libisev keskmine `LLK(4)`:

$$y_t = \frac{1}{4} \sum_{i=0}^3 z_{t-i}$$

ja perioodiga 12 sobiv modifitseeritud lihtne sümmeetiline libisev keskmine `mod_LSLK(6)`:

$$y_t = \frac{1}{24}z_{t+6} + \frac{1}{12} \sum_{i=-5}^5 z_{t-i} + \frac{1}{24}z_{t-6}.$$

Kujutage tulemused ühel joonisel koos esialgse reaga.

Ülesanne 3. Arvutame eksponentsiaalselt silutud kinnisvara müügiandmed juhtudel $\alpha = 0.1, 0.3, 0.6, 0.9$ ja kujutame silutud aegridu ühel joonisel. Eksponentsiaalseks silumiseks on kasulik käsk

```
y=filter(x,filter=w,method="recursive",init=algväärtused).
```

Selle käsuga arvutatakse tulemusrea y väärtused valemiga

$$y_i = x_i + w_1 y_{i-1} + \dots + w_p y_{i-p},$$

kus p on kaalude vektori komponentide arv ja algväärtused annavad vektori y esimesed p väärtust. Eksponentsiaalse silumise saamiseks peab olema ainult üks kaal $w = (1 - \alpha)$, algväärtuseks võib võtta keskmistatava rea esimese väärtuse ning x asemel eelpooltoodud käsus `filter` tuleb anda `$\alpha \cdot esialgne\_rida$` .

Ülesanne 4. Leiame turistide majutamise aegrea sesoonse osa käsitsi ning samuti vastava sesoonselt kohandatud rea:

- Eemaldame turistide aegreast trendi (leitud ülesandes 2 modifitseeritud lihtsal silutud keskmise abil)
- Leiame erinevatele kuudele vastavad keskmised ilma trendita reast käsuga
`keskmised=aggregate(trendita~cycle(turistid),FUN=mean)[,2]`
- Lahutame leitud kuistest keskmistest nende aritmeetilise keskmise (et summa tuleks null)
- Moodustame sesoonsele osale vastava aegrea käsuga kujul
`sesoonne=ts(keskmised[cycle(z)],start=start(z),frequency=frequency(z))`
kus eelnevas z tähistab turistide aegrida.
- Leiame sesoonselt kohandatud rea, lahutades esialgsest turistide andmestikust sesoonse osa

Kas vaadeldava rea puhul on ajas muutumatu sesoonse osa eeldus õigustatud?

Ülesanne 5. Vaatleme ka aegrea osadeks lahutamist R käskudega. Klassikaline meetod on R tarkvaras realiseeritud käsuga `decompose()`. Lugege selle kirjeldust (käsuga `help(decompose)`) ning leidke kinnisvara andmestiku ja turistide andmestiku korral nii multiplikatiivne kui ka aditiivne osadeks lahutus.

- Katsetage osadeks lahutamisel ka käsu `stl()` võimalusi. Selleks lugege õpetust ning leidke nii täielikult perioodilise sesoonse osaga lahutus kui ka muutuva sesoonse osaga lahtus kinnisvara andmete jaoks.