

Praktikum 7

Teema: hinnang, hinnangufunktsioon ja selle jaotus, hinnangute omadused

Näide Olgu antud juhuslik valim mahuga $n = 15$ jaotusest $N(\mu, 1)$, kus μ on tundmatu parameeter. Parameetri μ hindamiseks on pakutud 4 hinnangut:

$$\hat{\mu}_1 = \bar{x}, \quad \hat{\mu}_2 = Med, \quad \hat{\mu}_3 = \frac{x_{max} + x_{min}}{2}, \quad \hat{\mu}_4 = \frac{al.kvartiil + \text{ül.kvartiil}}{2}.$$

Milline nendest hinnangutest on nihketa, milline on kõige väiksema varieeruvusega? Lahendame R abil.

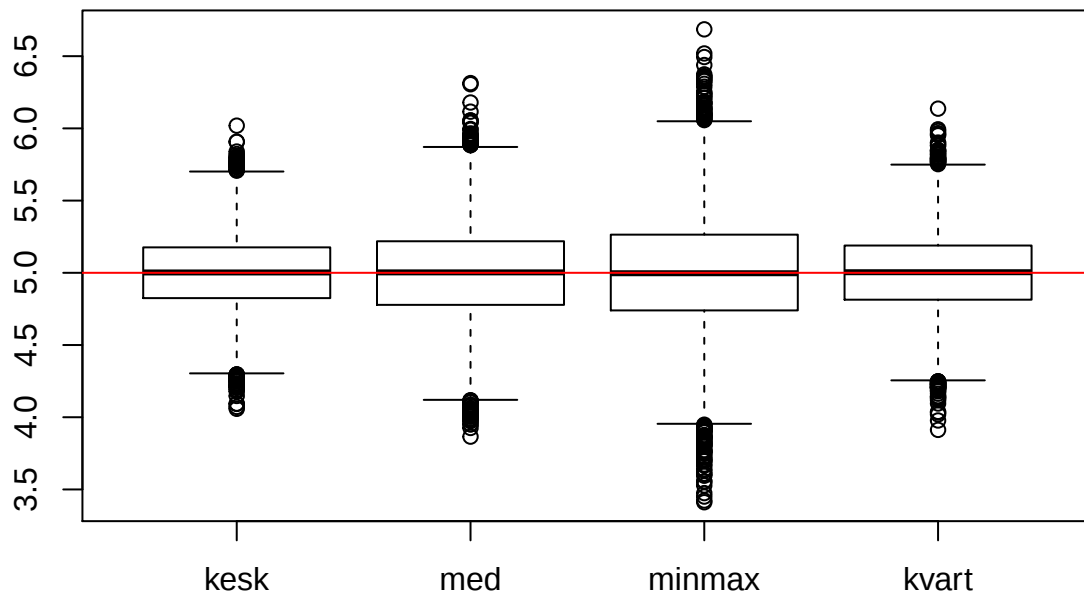
```
R=10000 #korduste arv
#Defineerime vajalikud vektorid pikkusega R
kesk=rep(NA,R) #valimikeskmiste jaoks
med=rep(NA,R) #valimimediaanide jaoks
minmax=rep(NA,R) #maksimaalse ja minimaalse elemendi poolsumma jaoks
kvart=rep(NA, R) #kvartiilide poolsumma jaoks

#Fikseerime mu väärtuse, olgu 5 ja simuleerime R valimit.
#Igas valimis arvutame 4 hinnangut

for(i in 1:R){
  v=rnorm(15,5,1) #juhuslik valim jaotusest N(5, 1) mahuga n=15
  kesk[i]=mean(v)
  med[i]=median(v)
  minmax[i]=(max(v)+min(v))/2
  kvart[i]=(quantile(v, prob=c(0.25)) +quantile(v, prob=c(0.75)) )/2
}

#Karakteristikute vektorid ühte tabelisse:
kar=data.frame(kesk, med, minmax, kvart)

#karpdiagrammina graafikule:
boxplot(kar)
#mu=5 samale graafikule (online R-is ei õnnestu)
abline(h=5, col="red")
```



Karpiagrammilt on näha, et kõik hinnangud on ligikaudu nihketa. Varieeruvus tundub hinnangutel $\hat{\mu}_1$ ja $\hat{\mu}_4$ väiksem võrreldes hinnangutega $\hat{\mu}_3$ ja $\hat{\mu}_4$. Kuid täpsema tulemuse saame leides hinnangute aritmeetilised keskmised ja dispersioonid üle simulatsioonide:

```
apply(kar,2, mean) #2 vastab keskmistele veergudes ja 1 keskmistele ridades
```

```
##      kesk      med    minmax     kvart
## 5.000580 5.000141 4.997570 5.001482
```

```
apply(kar,2, var) #var dispersiooni leidmiseks
```

```
##      kesk      med    minmax     kvart
## 0.06753337 0.10418021 0.15921409 0.07735158
```

Näeme, et kõige väiksem varieeruvus on siiski hinnangul $\hat{\mu}_1$, ehk valimikeskmisel.

Ülesanded

Ülesanne 1 Olgu antud juhuslik valim mahuga $n = 10$ jaotusest $N(3, \sigma)$, kus σ on tundmatu parameeter. Näidata nii analüütiliselt kui ka R-abil, et hinnang

$$\hat{\sigma}_1^2 = \frac{1}{n} \sum_{i=1}^n (x_i - 3)^2$$

on nihketa hinnang parameetrile σ^2 , ja hinnang

$$\hat{\sigma}_2^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - 3)^2$$

on nihkega hinnang parameetrile σ^2 .

Kumma hinnangu varieeruvus on suurem (vaadata R abil)?

Ülesanne 2 Olgu p uue presidendi pooldajate osakaal riigs A. See on tundmatu parameeter. Kaks erinevat agentuuri viisid läbi sõltumatud uuringud, et seda osakaalu ära hinnata. Esimese agentuuri nihketa hinnangu $\hat{p}_1$ kohta on teada, et selle standardhälve on 0,4. Teise uuringu agentuuri nihketa hinnangu $\hat{p}_2$ standardhälve on 0,6. Nendest hinnangutest on omakorda kombineeritud uus hinnang,

$$\hat{p} = \alpha \hat{p}_1 + (1 - \alpha) \hat{p}_2,$$

kus $0 \leq \alpha \leq 1$.

1. Tõestada, et $\hat{p}$ on nihketa hinnang parameetrile p .
 2. Leida selline α , mille korral on hinnangu $\hat{p}$ dispersioon minimaalne nii analüütiliselt kui ka R abil.
-

Ülesanne 3 Olgu $x_1, x_2, \dots, x_n$ juhuslik valim jaotusest, mille tihedusfunktsioon on järgmine:

$$f_X(x) = \begin{cases} e^{-(x-\theta)} & , \text{ kui } x \geq \theta; \\ 0, & \text{ muidu;} \end{cases}$$

kus $\theta \geq 0$ on tundmatu parameeter. Selle hindamiseks on pakutud hinnang

$$\hat{\theta}_1 = \min(x_1, \dots, x_n) - \frac{1}{n}.$$

Kontrollida, kas tegemist on nihketa hinnanguga või mitte.

Kodused ülesanded 7

Kodu 1. Ülesandes 3 pakutakse veel üks hinnang parameetrile θ kujul $\hat{\theta}_2 = \bar{x} + c$, kus $\bar{x}$ on valimikeskmine ja c on mingi konstant.

1. Leida konstandi c väärtus nii, et $\hat{\theta}_2$ oleks nihketa.
2. Leitud c väärtusega vastata küsimusele: kumb hinnangutest ($\hat{\theta}_1$ või $\hat{\theta}_2$) on efektiivsem?

Kodu 2. Olgu antud valim $x_1, \dots, x_n$ jaotusest $Po(\lambda)$, kus λ on tundmatu parameeter. Poissoni jaotuse kohta on teada, et parameeter λ on nii jaotuse keskväärtus kui ka dispersioon. Seetõttu on loomulik pakkuda kaks järgmist hinnangut λ hindamiseks:

$$\hat{\lambda}_1 = \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i, \quad \hat{\lambda}_2 = s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2.$$

Viia läbi simuleerimiseksperiment, et välja selgitada, kumb hinnangutest on efektiivsem (kontrollida ka nihketuse suhtes). Lahendus peab sisaldama kommentaari, R-koodi ning kasutatud väljundeid. *Poissoni jaotusest simuleerimiseks kasutada funktsiooni rpois.*