

Lineaarsed mudelid Mitmene võrdlemine I

Usaldusriba (Confidence Band)

Olgu meil väike andmestik,

```
set.seed(4)
n=25
x=runif(n, 0,10)
y=2+0.6*x-0.1*x*x+rnorm(n, sd=1)
plot(x,y)
```

ja soovime seda andmestikku kasutades hinnata regressioonanalüüsi mudelit:

```
m1=lm(y~x+I(x**2))

abi=seq(-1,11, length=200)
yhat=predict(m1, data.frame(x=abi))
lines(abi,yhat, lwd=2)
```

Loomulikult tahame ka iseloomustada seda, kui täpselt me ikka seda mudelit teame. Üheks võimaluseks on lisada joonisele 95%-usaldusintervallid:

```
abi=seq(-1,11, length=200)
yhat=predict(m1, data.frame(x=abi), interval="confidence")

lines(abi,yhat[,2], lty=2)
lines(abi,yhat[,3], lty=2)
```

Aga nende usaldusintervallidega on üks häda: iga x -i väärtuse korral võime küll väita 95%-kindlusega, et tegelik regressioonjoon asub usalduspiiride sees, aga meil on joonistatud praegu väga palju erinevaid usalduspiire (arvutasime tegelikult kakssada usaldusipiiri....). Seega on täiesti võimalik, et joonisel kuskil (näiteks 5% ulatuses) on tegelik regressioonjoon väljaspool usalduspiire...

Kanna joonisele ka tegelik regressioonijoon. Kas regressioonijoon püsib kenasti punktiviisiliste 95%-usaldusintervallide vahel?

Tee 20 simulatsiooni (iga kord uute andmetega, ilma set.seed-käsku R'le andmata). Loe kokku mitme genereeritud andmestiku korral sattus tegelik regressioonijoon joonisel väljapoole 95%-usaldusvahemikku:

..... / 20

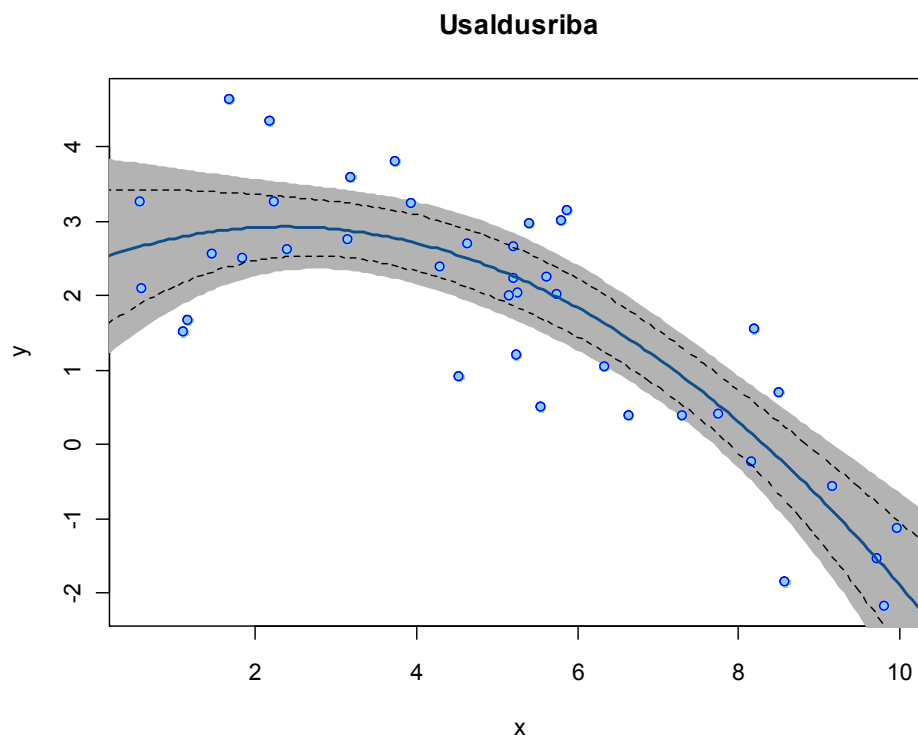
Kuidas saada regressioonjoonele selliseid usalduspiire, et 95% kindlusega saaksime väita – tegelik regressioonjoon asub täies ulatuses joonistatud piiride vahel – joonistatud usaldusriba sees (95% valimitest on sellised, et regressioonjoon jääb täies ulatuses joonisele kantud usaldusriba sisse)?

Appi tuleb Scheffé meetod...:

$$P \left(\bigcap_{\{\lambda | \lambda^T = v^T X; v^T X_0 = 0\}} \left\{ \lambda^T \hat{\beta} - \sqrt{(p - p_0) D(\hat{\lambda}^T \hat{\beta})} f_{1-\alpha} \leq \lambda^T \beta \leq \lambda^T \hat{\beta} + \sqrt{(p - p_0) D(\hat{\lambda}^T \hat{\beta})} f_{1-\alpha} \right\} \right) = 1 - \alpha$$

Toodud võrdus kehtib samaaegselt kõigi hinnatavate lineaarkombinatsioonide $\lambda^T \beta$ jaoks... (mis puudutavad vaid neid parameetreid, mis visatakse välja mudeli maatriksi muutmisel $X \rightarrow X_0$)

Kanna joonisele Scheffé meetodil leitud 95%-usaldusriba (lisaks tavalisele punktiviisilisele 95% , näiteks nii:



Genereeri ka nüüd 20 andmestikku. Mitmel juhul jäi tegelik regressioonseost kirjeldav sirge väljapoole usaldusriba?

..... / 20