

Mittelineaarne regressioon

Allan Sims

31.01.2025

Mittelineaarne regressioon

Mittelineaarne regressioon on statistiline meetod, mida kasutatakse sõltuva muutuja ja ühe või mitme sõltumatu muutuja vahelise mittelineaarse seose modelleerimiseks. Erinevalt lineaarsest regressioonist, mis eeldab sirgjoonelist seost, võimaldab mittelineaarne regressioon paindlikumaid seoseid, näiteks kõverjoonelisi, eksponentsiaalseid või logistilisi seoseid.

Metsanduses esineb sageli mittelineaarseid seoseid. Mittelineaarne regressioon võimaldab neid keerukaid seoseid täpsemalt modelleerida ja paremini mõista.

Mittelineaarne regressioonimudel kirjutatakse üldjuhul kujul:

$$Y = f(X, \beta) + \epsilon$$

kus:

- \mathbf{Y} on sõltuv muutuja.
- \mathbf{X} on sõltumatu(te) muutuja(te) vektor.
- $f(X, \beta)$ on mittelineaarne funktsioon, mis kirjeldab seost X ja Y vahel.
- β on mudeli parameetrite vektor, mida tuleb hinnata.
- ϵ on vealiige, mis esindab juhuslikku varieeruvust.

Mittelineaarse regressiooni parameetreid hinnatakse tavaliselt **vähimruutude meetodil**, kus otsitakse parameetrite väärtusi, mis minimeerivad mudeli ennustatud väärtuste ja tegelike väärtuste vaheliste ruutude summat. See on iteratiivne protsess, mis nõuab arvutitarkvara kasutamist.

R

Näiteks, kui soovime R-is kasutada `nls()` mudelit, mis kirjeldab puu kõrgust sõltuvalt puu diameetrist, võime kasutada järgmist näidiskoodi.

```
# Laadime peatükis kasutatavad paketid
```

```
library(dplyr)
```

```
library(readxl)
```

```
# Laadime andmed sisse failist ning
```

```
# filtreerime välja puud, mille h > 0
```

```
andmed <- read_excel("data/naited.xlsx", "prt_andmed") %>% filter(h > 0)
```

```
# Määratleme mudeli funktsioon
```

```
puu_mudel <- function(D, a, b, c) {
```

```
  1.3 + a * (D / (D + b))^c
```

```
}
```

```

# Omista algväärtused parameetritele
alg_param <- list(a = 1, c = 0.5)

# Sobita mudel andmetele
mudel <- nls(h ~ puu_mudel(d, a, 5, c), data = andmed, start = alg_param)

# Vaata mudeli kokkuvõtet
summary(mudel)

```

Formula: $h \sim \text{puu_mudel}(d, a, 5, c)$

Parameters:

	Estimate	Std. Error	t value	Pr(> t)
a	34.341	1.430	24.013	< 2e-16 ***
c	2.056	0.216	9.515	2.41e-12 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.322 on 45 degrees of freedom

Number of iterations to convergence: 7

Achieved convergence tolerance: 1.141e-07

Esitame selle kohta ka graafiku, millele lisame mudeli joone funktsioonidega `curve()` ja `predict()`, milles kasutame loodud mudeli objekti.

```

plot(h~d, andmed, ylim=c(10, 35), xlim = c(10, 40))
curve(predict(mudel, data.frame(d = x)), 10, 40, add=T, col = 2)

```

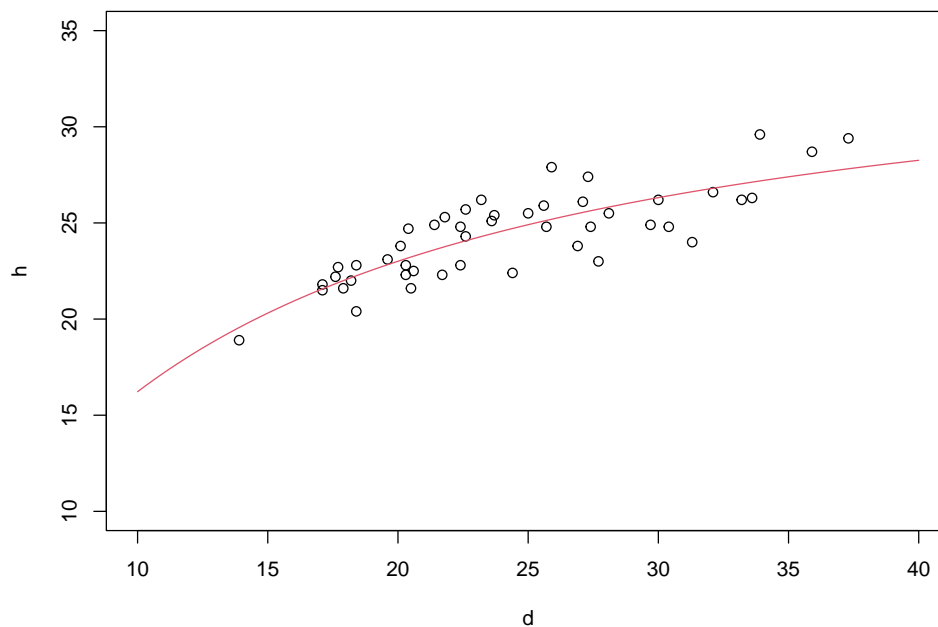


Figure 1: Mittelineaarne regressioon

Kontrollime

```
plot(fitted(mudel), resid(mudel),  
     xlab = "Ennustatud väärtused",  
     ylab = "Jäägid"  
     )  
abline(h = 0, col = "red")
```

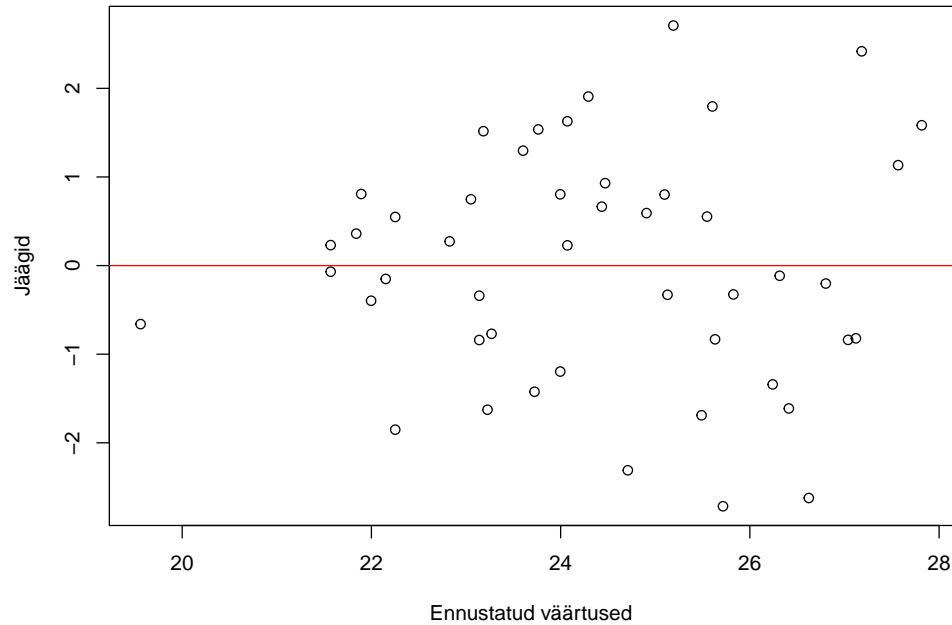


Figure 2: Jäägid vs ennustatud

```
qqnorm(resid(mudel), main = "")  
qqline(resid(mudel), col = "red")
```

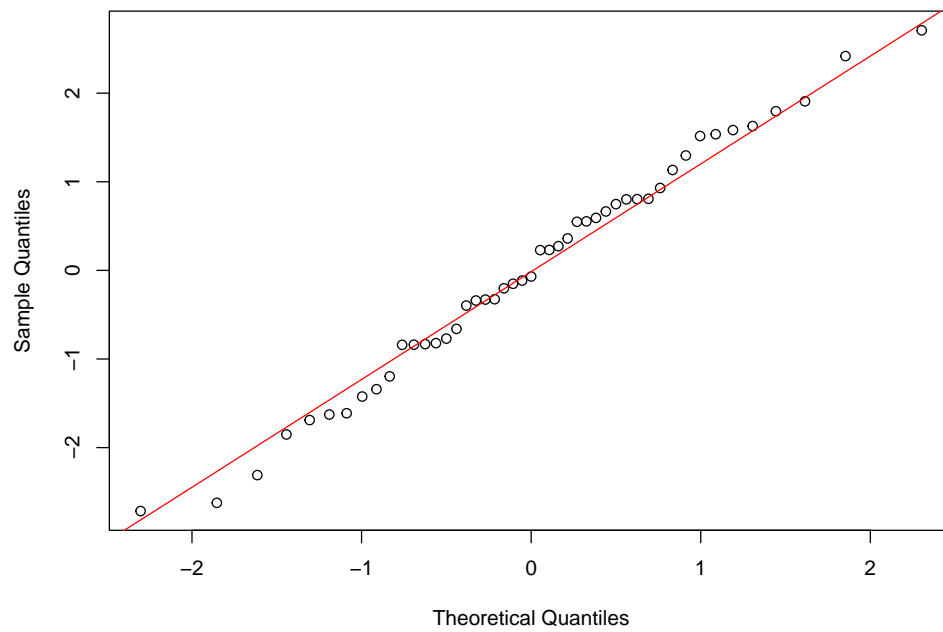


Figure 3: Jääkide normaaljaotuse test

MS Excel

Andmete sisestamine. Avage fail, kus on puude diameetrid ja kõrgused.

Algparameetrite väärtuste sisestamine: Valige kolm vaba lahtrit ja sisestage neisse algparameetrite väärtused a, b ja c jaoks. Need väärtused võivad olla suvalised, aga parema tulemuse saamiseks peaksid need olema lähedased oodatud väärtustele. Näiteks võite alustada $a = 20$, $b = 10$ ja $c = 3$. Nimetage need lahtrid vastavalt “a”, “b” ja “c”.

Arvutatud kõrguse veeru loomine: Looge uus veerg nimega “Arvutatud H”. Esimesse lahtrisse sisestage valem: $=1.3 + \$a * (\$D2 / (\$D2 + \$b))^{\$c}$. Asendage \$a, \$b ja \$c vastavate algparameetrite lahtritega (nt \$F\$1, \$F2, ‘F3’). ‘D2’ on esimese diameetri väärtus. See valem arvutab kõrguse vastavalt sisestatud diameetritele ja algparameetrite väärtustele.

Valemi kopeerimine: Lohistage esimese “Arvutatud H” lahtri paremas alumises nurgas olevat ruudukest alla, et kopeerida valem kõikidele diameetri väärtustele.

Ruutkeskmise vea arvutamine: Valige vaba lahter ja sisestage valem $=\text{SUMXMY2}(H2:Hn, E2:En)/\text{COUNT}(H2:Hn)$. See valem arvutab ruutkeskmise vea (RMSE) mõõdetud kõrguse (H) ja arvutatud kõrguse (Arvutatud H) vahel. H2:Hn on mõõdetud kõrguste vahemik ja E2:En on arvutatud kõrguste vahemik.

Solveri käivitamine: Avage menüüst “Andmed” -> “Solver”. Kui Solverit menüüs ei ole, peate selle lisandmoodulitest aktiveerima (Fail -> Suvandid -> Lisandmoodulid -> Exceli lisandmoodulid -> Mine... -> Märkige ruut “Solver Add-in” ja vajutage “OK”).

Solveri seadistamine: * Määrake “Seadista eesmärk” lahtri RMSE lahter. * Valige “Min”, et minimeerida RMSE. * “Muutuvate lahtrite kaupa” lahtritesse sisestage algparameetrite a, b ja c lahtrid. * Vajutage “Lahenda”.

Tulemuste analüüs: Solver kohandab parameetrite a, b ja c väärtusi, et minimeerida RMSE. Kui Solver on lahenduse leidnud, kuvatakse leitud parameetrite väärtused ja minimaalne RMSE. Nüüd on teil mittelineaarne regressioonimudel, mis kirjeldab puude kõrguse ja diameetri seost teie andmete põhjal.

Lisamärkused:

- Solver ei pruugi alati leida optimaalset lahendust. Proovige erinevaid algparameetrite väärtusi ja käivitage Solver uuesti.
- Saate visualiseerida oma andmeid ja leitud regressioonikõverat, luues hajuvusdiagrammi, kus x-teljel on diameeter ja y-teljel kõrgus. Lisage diagrammile trendijoon, valides “Kohandatud” ja sisestades regressioonivõrrandi.
- Hinnake mudeli sobivust, arvestades RMSE väärtust ja visuaalset vastavust andmetele.