

TARTU ÜLIKOOL
LOODUS- JA TÄPPISTEADUSTE VALDKOND
MATEMAATIKA JA STATISTIKA INSTITUUT

Joonas-Sander Tamm
**Covid-19 hospitaliseerimise prognoosimine
nakatunute andmetelt**

Matemaatiline statistika

Bakalaureusetöö(9 EAP)

Juhendaja: prof. Krista Fischer, PhD

TARTU 2022

COVID-19 HOSPITALISEERIMISE PROGNOOSIMINE NAKATUNUTE ANDMETELT

Bakalaureusetöö
Joonas-Sander Tamm

Lühikokkuvõte

Bakalaureuse töös proovitakse leida lineaarse regresiooni ja Ridge regresiooni abil mudelit, mille abil saab ennustada covidiga hospitaliseeritute arvu 7 päeva varasemate nakatunute andmete pealt. Töö käigus kasutatakse mõlemat meetodit ning analüüsitakse nende eeliseid ja puuduseid. Kõige paremini seletas hospitaliseerituse muutumist üleüldine nakatumine ja vaktsineerimine ning samad näitajad eraldi 80 aastaste ja vanemate seas. Kõik töös kasutatud andmed on pärit Terviseameti avaandmete seast.

CERCS teaduseriala: P160 Statistika, operatsioonianalüüs, programmeerimine, finants- ja kindlustusmatemaatika

Märksõnad: Prognoos, Covid-19, hospitaliseerimine.

PREDICTION OF HOSPITALIZATION BASED ON THE DATA OF THE INFECTED

Bachelor's thesis
Joonas-Sander Tamm

Abstract

The objective of this bachelor thesis is using Ridge regression and linear regression to find a model, which predicts the amount of people hospitalised with covid based on data about the infected from 7 days ago. Both methods previously mentioned are used to fit many models, which all have different pros and cons. The best variables, which explained the variance of hospitalisation the most, were the overall infection rate and vaccination rate and the same metrics among people aged over 80 years. All data used in this thesis is from Republic of Estonia Health Board open data.

CERCS research specialisation: P160 Statistics, operations research, programming, financial and actuarial mathematics

Key Words: Prediction, Covid-19, Hospitalisation

Sisukord

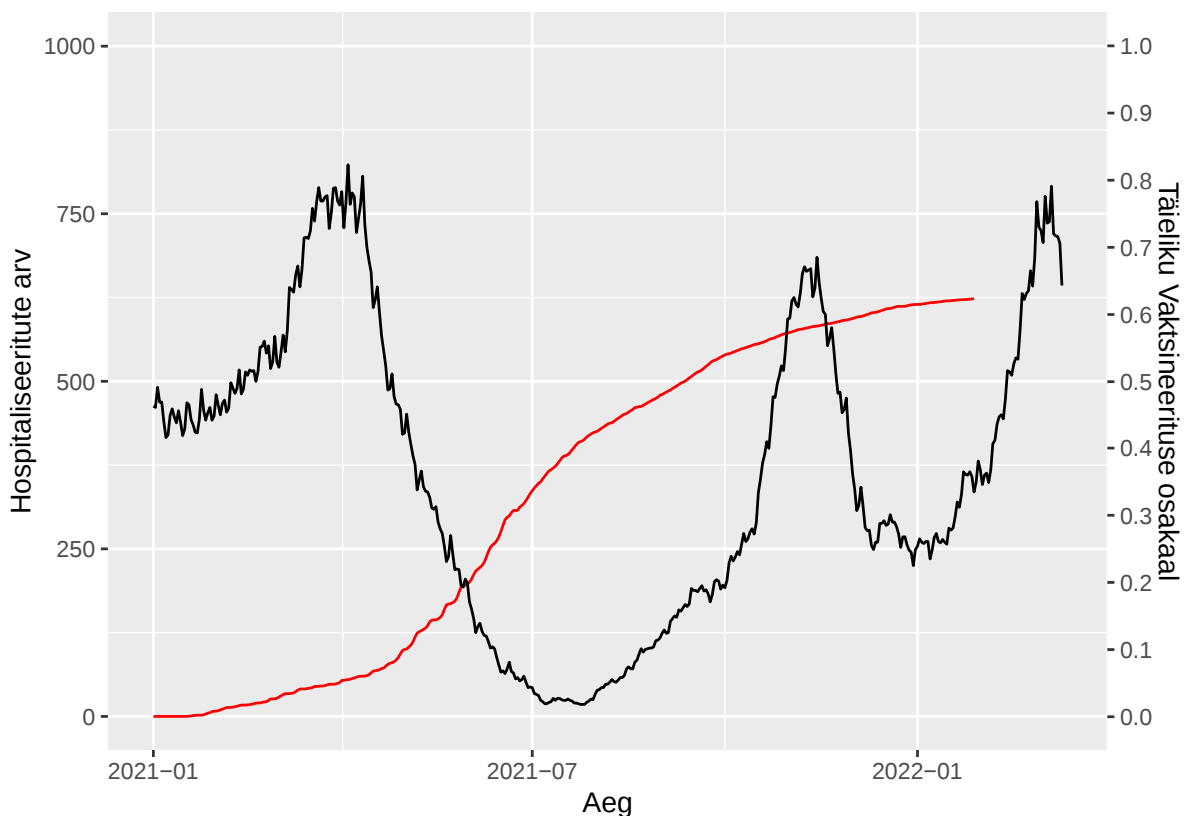
Sissejuhatus	3
Andmed	4
Statistilised meetodid	6
Mitmene lineaarne regressioon	6
Multikollineaarsus	7
Ridge regresioon	9
Andmeanalüüs	11
Ridge regressioon	20
Edasisuunas mudeli valik	24
Kokkuvõte	30
Kasutatud kirjandus	31

Sissejuhatus

Eestis tuvastati esimene koroonajuhtum 26.veebruari 2020. Esimene vaktsineerimine koroonaviiruse vastu toimus 27. detsembril 2020. Kõige esimesest tuvastatud koroonaviiruse juhtumist alates on Terviseamet kogunud andmeid viiruse leviku ja vaktsineerimise kohta Eestis. 15.veebruari 2022 seisuga on Eestis surnud 2122 inimest koroonaviirusesse (Terviseamet koroonakaart, 2022).

Viiruse levikut on proovitud piirata erinevate piirangutega valitsuse poolt. Piirangute tõhusus on vastuoluline teema. Seda ei saa otseselt mõõta, saab vaid kaudselt hinnata efektiivsust laialdase testimisega ja haiglate täituvuse mõõtmisega. Vähem inimesi haiglates ja väiksem nakatunute osakaal testitute seas tähendab, et piirangud saavutasid oma eesmärgi. Üleüldine hospitaliseeritus ja vaktsineeritus on ära toodud joonisel 1.

Järgnevas analüüsis on proovitud mudeldada hospitaliseeritute arvu ajast sõltuvalt. Proovitakse leida mudeleid, mille põhjal saab ennustada koroonaviiruse tõttu haiglas viibivate inimeste arvu sõltuvalt näitajatest, mida saab otseselt mõõta, näiteks vaktsineerituse tase vanuse lõikes, hospitaliseeritute arv nädal või kaks tagasi, positiivsete testide osakaal vanuse lõikes nädal tagasi jne.



Joonis 1. Hospitaliseerituse (must joon) ja vaktsineerituse (punane joon) aegrida

Andmed

Töös kasutatakse Terviseameti avaandmeid Tervise ja Heaolu infosüsteemide keskuse kodulehelt (TEHIK, 2022). Üleüldiste mudelite jaoks on kasutatud andmetabeleid “Hospitaliseerimise statistika aegrida”, “üleriigiline nakatumine” ja vaktsineerimise “üleriigiline andmestik”. Vanusegrupe eristavate mudelite jaoks on kasutatud tabeleid “Hospitaliseerimise statistika aegrida”, “Testitulemused” ja “Vanuserühmade hõlmatus”.

Andmestikes on palju tunnuseid, millel on väga sarnased nimed, kuid kujutavad tähenduselt erinevaid asju. Välja toodud mõned (rohkem segi aetavad ja) rohkem kasutatud :

- LastLoadStatisticsDate (Kuupäev) - Viimane andmestiku uuendamise kuupäev
- StatisticsDate (Kuupäev) - Andmerea sisestamise kuupäev
- Aegfaktor (Diskreetne) - Faktortunnus, mis näitab andmerea sisestamise perioodi (1-(2021.1-2021.2),2-(2021.3-2021.5),3-(2021.6-2021.12),4-(2022.1-2022.2))
- Hospitalised (Täisarv) - Vastaval päeval haiglas olnud inimeste arv (sisaldab ka neid, kes vastaval päeval välja kirjutati)
- ActivelyHospitalised (Täisarv) - Vastava päeva lõpuks haiglas olevad inimesed (ei sisalda neid, kes vastaval päeval välja kirjutati)
- VaccinationSeries (Täisarv) - Vaktsineerimise kuur (1-klassikaline senine süsteem (1 või 2 doosi), 2- esimene tõhustusdoos, 3-teine tõhustusdoos, jne.)
- MeasurementType (Diskreetne) - Vaktsineerimise tabelites real oleva numbriga tähendus (DosesAdministered - sel päeval administreeritud vaktsiinidooside hulk, Vaccinated - Inimeste arv, kes said doosi, FullyVaccinated - Inimeste arv, kes said doosi ja vaktsineerimiskuur loeti lõpetatuks)
- PopulationCoverage (Numbriline) - Vaktsineerituse tase vastavas inimeste gruppis (varieerub nullist üheni)
- PerPopulation (Numbriline) - Viimase 14 päeva nakatunute suhtarv 100000 elaniku kohta vastavas inimeste gruppis.
- ResultValue (Binaarne) - Testi tulemus (N-negatiivne, P-positiivne)
- AgeGroup (Diskreetne) - Inimese Vanusegrupp (0-19,20-39,40-59,60-79,80+)

Kuna koroonaviiruse ja sellega hospitaliseeritud inimeste arvu kohta tuleb iga päev andmeid juurde, ning mudelid muutuksid konstanselt, siis otsustati valida fikseeritud ajavahemik (2021 jaanuari algus kuni 2022 veebruari lõpp) ja kasutada andmeid vaid sealt.

Käesolevas töös on nende andmete põhjal defineeritud järgmised tunnusused :

- Täieliku vaktsineerituse osakaal - Näitab, kui suur osa kogu populatsioonist või osapopulatsioonist on loetud täielikult vaktsineerituks (Arvutatakse järgmiselt : (Täielikult vaktsineeritud inimeste arv populatsioonis)/(Populatsiooni arv))
- Nakatumise suhtarv - Viimase 14 päeva nakatumise suhtarv 100000 elaniku kohta (Suhtarv arvutatakse valemi alusel: viimase 14 päeva positiivsete esmakordsete testide koguarv * $100\,000 / 1\,328\,976$ (elanike arv)) (Osapopulatsioonis hindamiseks kasutatakse osapopulatsioonile vastavaid näitajaid (Osapopulatsioonis positiivsete testide arv ja osapopulatsiooni suurus))

Statistilised meetodid

Mitmene lineaarne regressioon

Peamiseks käesolevas töös kasutatud meetodiks on lineaarne regressioon. Lineaarse regressiooni puhul eeldatakse mudelit, kus ennustatava tunnuse keskväärus avaldub seletavate tunnuste lineaarkombinatsioonina. Kokku võetakse mudelisse p erinevat seletatavat tunnust. Iga ennustatava vaatluse Y_i jaoks saab välja kirjutada järgmise võrrandi:

$$Y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \dots + \beta_p x_{pi} + e_i$$

kus

Y_i - i -nda vaatluse ennustatava tunnuse väärtus

β_0 - vabaliige

$x_{1i,2i,3i,\dots,pi}$ - seletavate tunnuste tegelikud väärtused i -ndal vaatlusel

$\beta_{1,2,3,\dots,p}$ - Lineaarse regressiooni kordajad

e_i - Viga i -ndal vaatlusel, kus $e_i = N(0, \delta^2)$

Lineaarse regressiooni eesmärk on leida sellised kordajad $\beta_{0,1,2,3,\dots,p}$, et e_i ruutude summa kõigist vaatlustest oleks väikseim. See saadakse peamiselt Vähimruutude meetodiga, ehk järgneva võrrandi lahendamisega :

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$$

kus

$\mathbf{X}'\mathbf{X}$ - seletavate tunnuste korrelatsioonimaatriks (dimensioonid : $(p-1) \times (p-1)$)

$\mathbf{X}'\mathbf{Y}$ - Y tunnuse ja kõigi seletatavate tunnuste vahelise lihtsa korrelatsiooni maatriks (dimensioonid : $(p-1) \times 1$)

$\mathbf{b}$ - lineaarse regressiooni kordajate maatriks (dimensioonid : $(p-1) \times (1)$) (Kutner,2005)

R-s kasutatakse mudelite hindamiseks funktsiooni "lm()".

Multikollineaarsus

Multikollineaarsuseks nimetatakse nähtust, kus seletavad tunnused on omavahel tugevalt korreleeritud, mis tähendab, et maatriks $\mathbf{X}'\mathbf{X}$ pole pööratav või on lähedal singulaarsusele. Multikollineaarsuse olemasolu põhjustab järgmiseid probleeme ning nende probleemide olemasolu võib viidata multikollineaarsuse eksisteerimisele andmetes:

- Seletatava tunnuse lisamine või eemaldamine mudelist tunduvalt muudab teiste seletatavate tunnuste kordajaid või nende olulisust.
- Seletatavate tunnuste kordajate standardvead mudelis on suured.

Need meetodid ei ole formaalsed meetodid multikollineaarsuse hindamiseks, kuid võivad anda aimu, et andmetes leidub multikollineaarsust. Formaalsemaks meetodiks on suurus VIF (Variance Inflation Factor) leidmine. Üldiselt näitab VIF, kui palju mingisuguse seletava tunnuse kordaja hinnangu dispersioon mudelis on suurem võrreldes olukorraga, kui see kordaja ei oleks lineaarselt koreleeritud teiste tunnustega. Kõrge VIF-i väärtuse korral saab öelda, et see tunnus on rohkem seotud teiste seletavate tunnustega, kui ennustatava tunnusega.

Kui mingi tunnuse VIF kordaja on täpselt 1, siis ta ei ole lineaarselt seotud ühegi teise tunnusega. Mida suuremaks läheb VIF kordaja, seda suurem on antud tunnusel lineaarne korrelatsioon teiste tunnustega. VIF-i lähenemine lõpmatusele näitab seda, et antud tunnuse saab täielikult ära seletada teiste tunnustega, ehk teiste tunnuste lineaarkombinatsioonina. Multikollineaarsuse hindamiseks mingis kindlas mudelis kasutatakse kõige suuremat VIF väärtust kõigi argumentide seas mudelis. Võib ka kasutada VIF väärtuste keskmist. Öeldakse, et kui VIF väärtus on suurem, kui 5 või 10, siis võivad olla tugevad kõrvalekaldad seletavate tunnuste kordajates mudelis. (Kutner 2005)

VIF arvutatakse järgmiselt :

$$(\text{VIF})_k = (1 - R_k^2)^{-1}$$

kus

R_k^2 - on mitmene regreesiooni kordaja mudelile, kus k-ndat argumenttunnust ennustatakse p-2 teise muutuja pealt (Vabaliige ja k-s tunnus on puudu)

Olgu r_{XX} korrelatsioonimaatriks, kus X on meie mudelis olevad tunnused. r_{XX}^{-1} on selle maatriksi pöördmaatriks. VIF_k näitab k -ndat diagonaalelementi selles pöördmaatriksis, mis näitab ka k -nda muutuja sõltuvust teistest selles maatriksis olevatest muutujatest.

Üks oluline märkus on see, et multikollineaarsuse olemasolu mudelis ei mõjuta selle mudeli võimekust teha ennustusi (ennustuste täpsust, keskmist, jne...), kui eeldada seda, et ka uutes andmetes eksisteerib samasugune multikollineaarsus. Ehk, kui eeldada, et andmete vahel eksisteerib multikollineaarsus ka tulevikus, siis sobiksid eelnevad mudelid ennustamiseks piisavalt hästi. Samas on siis keeruline võrrelda erinevate argumenttunnuste mõju suurust ja olulisust. (Kutner 2005)

Ridge regresioon

Ridge regresiooni kasutatakse siis, kui eksisteerib multikollineaarsus mudelis, mis teeb tavalise lineaarse regressiooniga saadud mudelid ebausaldusväärseks. Ridge regressioon on sarnane tavalise lineaarse regressiooniga, kuid lisab juurde ka normeerimiskonstandi. Kui kõiki tunnuseid teisendada järgmiselt: $x_i^* = \frac{1}{\sqrt{n-1}} \frac{x_i - E(x)}{s_x}$, kus $E(x)$ ja s_x on tunnuse keskvärtus ja standardhälve, siis avaldub vähimruutude hinnang kui $b = M_{xx}^{-1} * M_{xy}$. Siin M_{xx} on seletavate tunnuste korrelatsiooni maatriks ja M_{xy} on y tunnuse ja kõigi seletatavate tunnuste vahelise lihtsa korrelatsiooni maatriks. Ridge regressiooni puhul lisatakse tavalisse vähimruutude meetodisse nihke konstant $c \geq 0$ järgmisel kujul:

$$(\mathbf{X}'\mathbf{X} + cI)b^R = \mathbf{X}'\mathbf{Y}$$

kus

$\mathbf{X}'\mathbf{X}$ - seletavate tunnuste korrelatsioonimaatriks $((p-1) \times (p-1))$

$\mathbf{X}'\mathbf{Y}$ - Y tunnuse ja kõigi seletatavate tunnuste vahelise korrelatsioonimaatriks $((p-1) \times 1)$

I - Ühikmaatriks $((p-1) \times (p-1))$

c - nihke konstant

b^R - Ridge regressiooni kordajate maatriks $((p-1) \times (1))$

Ridge regressiooni kordajate hinnangud leitakse, kui lahendada järgnev võrrand :

$$b^R = (\mathbf{X}'\mathbf{X} + cI)^{-1} * \mathbf{X}'\mathbf{Y}$$

Konstant c kannab nimetust nihke konstant, mis näitab, kui palju on kordajates nihet. Kui $c=0$, siis taandub meetod tavalise standardiseeritud vähimruutude meetodi peale. Mida suuremaks võtta c, seda vähema variatsiooniga on regressioonikordajad, kuid neis kordajates on olemas nihe.

Kõige keerulisem ridge regressiooni juures on nihke konstant c valik. Tavaline meetod on Ridge-i jälje hindamine. Ridge jälg on graafik, kus on näha ridge-i meetodil saadud kordajate väärtused konstandi c kasvades. Tavaliselt tehakse graafik c väärtustele vahemikus nullist üheni. Regressiooni kordajad muutuvad väga kiiresti ja võivad isegi märke vahetada, kui konstanti c hakata suurendama väärtusest 0. Mida suuremaks võtta c, seda väiksemaks muutuvad ridge

regressiooni kordajate muutused, kuni regressiooni kordajad lähenevad mingitele väärtustele. VIF kordajad vähenevad kiiresti, kui hakkata c -d suurendama 0-st, ning suuremate c väärtuste korral muutuvad VIF-i väärtused vaid keskmiselt. c väärtus valitakse ridge jälje järgi. Vaadatakse jälge ja võetakse vähim c väärtus, kus kordajate muutused on piisavalt väikesed, ning VIF väärtused ka piisavalt väikesed. Konstandi c valik on lõpuks subjektiivne (Kutner 2005).

R-s kasutatakse sellise mudeli hindamiseks funktsiooni `glmnet()`. (Statology 2022)

Andmeanalüüs

Esiteks proovime hinnata mudeli, kus hospitaliseeritute arv sõltub vaid hospitaliseeritusest nädal varem. Kuna haiglas viibitakse koroonaviirusega tavaliselt kauem, kui nädal, siis võib eeldada, et suur osa inimesi, kes praegu haiglas on, on ka seal 7 päeva pärast. (Terviseamet Koroonakaart, 2022)

Lineaarse regressiooniga mudeli üldkuju on :

$$y = \beta_0 + \beta_1 x_1$$

kus

y - hospitaliseeritute arv praegu

β_0 - vabaliige

x_1 - hospitaliseeritute arv 7 päeva tagasi

Tabel 1. Hospitaliseeritusega mudeli kordajad

Tunnus	b(se)	p-väärtus
Vabaliige	11.41(5.48)	0.0379
7 päeva tagune hospitaliseeritus	0.98(0.01)	<0.0001

$$R^2 = 0.9302$$

Mudeli põhjal leiame, et hospitaliseeritute arv on väga tugevalt seotud sama arvuga nädal varem. Samas on see mudel väheinformatiivne, sest ta ei ole võimeline ennustama hospitaliseerituse muutumist. Lisaks ei saa selle mudeliga näidata hospitaliseerituse korreleeritust ühegi teise seletatava tunnusega.

Teiseks proovime prognoosida hospitaliseeritute arvu nakatumisnäitaja ja seitsme päeva taguse vaksineerituse osakaalu põhjal. Seitse päeva on keskmiselt periood, peale mida hakkavad ilmnema koroonaviiruse sümptomid pärast nakatumist. See võiks olla suurus, mis seletab uute inimeste hospitaliseerimist. Vaksineeritus vähendab haiglasse sattumise tõenäosust. Samamoodi on 7

päeva ligikaudu aeg, peale mida hakkab immuunsüsteem vaktsiini abil antikehi tootma. See võiks otseselt vähendada haiglasse sattunud inimeste arvu. Vaktsineerituse all kasutame täieliku vaktsineerituse osakaalu.

Hinnatava regressioonimudeli üldkuju :

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2$$

kus

y - hospitaliseeritute arv praegu

$\beta_0, \beta_1, \beta_2, \beta_3$ - hinnatavad parameetrid

x_1 - nakatumise suhtarv

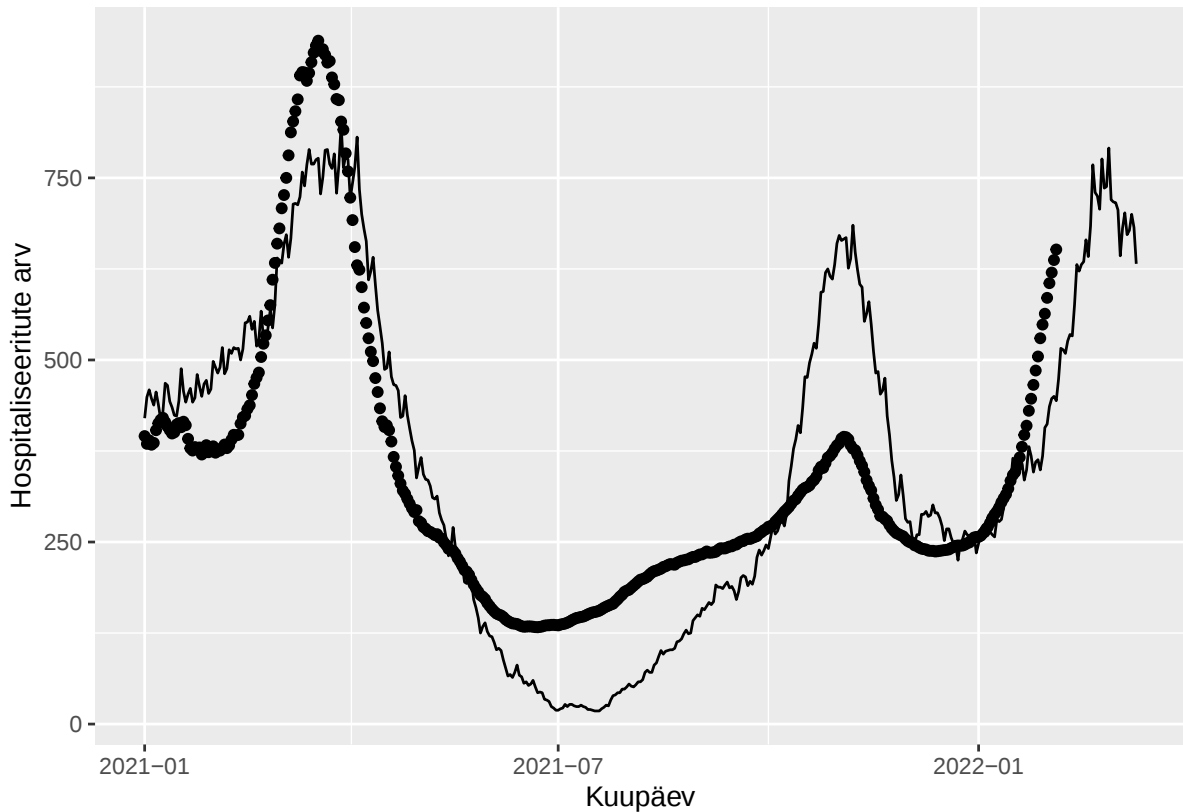
x_2 - 7 päeva tagune täieliku vaktsineerituse osakaal

$\beta_3 x_1 x_2$ - nakatumise suhtarvu ja vaktsineerituse osakaalu koosmõju

Tabel 2. Regressioonimudeli kordajad

tunnus	b(se)	p-väärtus
Vabaliige	57.06(18.02)	0.002
Nakatunute suhtarv	0.60(0.02)	<0.0001
Vaktsineeritus 7 päeva tagasi	2.25(0.41)	<0.0001
Koosmõju	-0.01(0.0003)	<0.0001

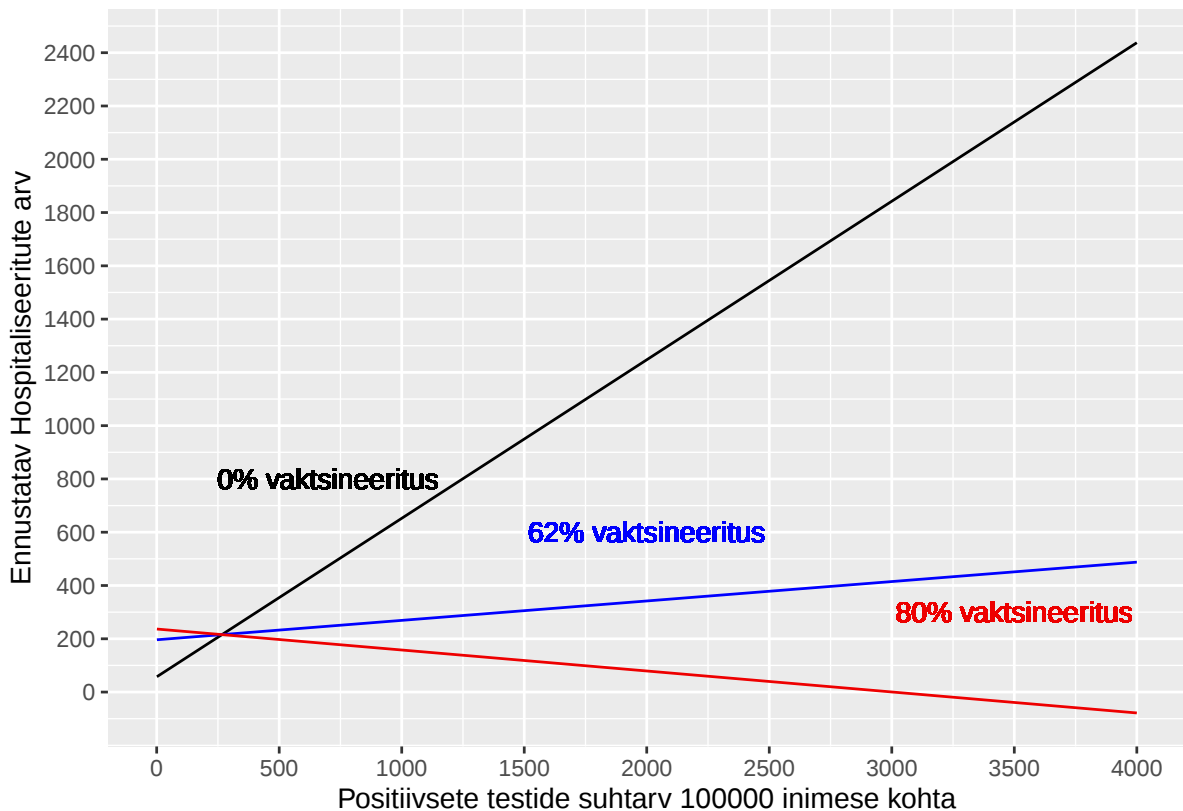
$$R^2 = 0.748$$



Joonis 2. Mudeli ennustus

Joonisel 2 on pidev joon tegelikud väärtused ja punktid on mudeli põhjal ennustatavad väärtused. Näeme, et selle mudeli täpsus on kaheldav. Ajaperioodi alguses mudel ülehindab muutuseid (ennustatav hospitaliseeritute arv tõuseb veidi kiiremini, kui tegelik). Ajatelje keskel mudel alahindab muutuseid mõlemas suunas (ennustab hospitaliseeritute kasvu ja langust väiksemaks, kui tegelik).

Vaadates programmi väljundit, näeme, et kõik tunnused on statistiliselt olulised. Üleüldiselt võrreldes praegust mudelit eelmise mudeliga kirjeldab see mudel vähem hospitaliseerituse muutusest (0.75 vs 0.93). See mudel ennustab ka vaikimisi keskmiselt hospitaliseerituks 57 inimest. Kordajate poole pealt: eksisteerib koosmõju kahe seletava tunnuse vahel, mis tähendab, et ei saa otseselt mõõta ühe teguri mõju ilma teist arvestamata.



Joonis 3. Vaksineerituse mõju hospitaliseeritusele (Hinnatud mudeli põhjal)

Joonis 3 ennustab hospitaliseerituse taset sõltuvalt viimase 14 päeva positiivsete testide suhtarvst erinevate vaksineerituse tasemete korral. Näeme, et kõrgemate vaksineerituse tasemete korral tõuseb hospitaliseeritute arv aeglasemalt positiivsete testide osakaalu tõustes, kui madalama vaksineerituse taseme korral. Mida väiksem on positiivsete testide suhtarv, seda vähem mõjutab vaksineerituse tase hospitaliseeritud inimeste arvu. See tähendab, et nende kahe tunnuse vahel võiks olla koosmõju, mida ka meie mudel tegelikult näitas.

Mudel ei ole ideaalne: Positiivsete testide arvu lähenemisel 0-le peaks ka hospitaliseeritute arv lähenema 0-le, kuid mitte ühegi vaksineerituse taseme juures ei lange see alla 50. Lisaks, kui võtta vaksineerituse tase väga suur, siis väga suurte positiivsete testide suhtarvu juures hakkab mudel andma negatiivseid väärtuseid, mis ei ole võimalikud. Seega oleks vaja järgmise mudeli jaoks rohkem informatsiooni kaasata.

Hindame nüüd mudeli, kus vaatame nakatumist vanuseti eraldi gruppides. Jagame inimesed vanuse järgi gruppidesse (0-19,20-39,40-59,60-79,80+) ja lisame mudelisse 5 tunnust iga vanusegrupi nakatumise kohta. On eelnevalt tõestatud, et erineva vanusega inimestel on erinev tõenäosus koroonaviirusega haiglasse sattuda, seega võiks vaadata nakatumist vanusegruppides.

Jätame välja need inimesed, kelle vanust pole teada, kuid neid inimesi on teistega võrreldes väga vähe. Siinkohal on oluline mainida, et nakatumist saab mõõta kahel erineval viisil: viimase 14 päeva nakatumise suhtarv 100 000 elaniku kohta või lihtsalt positiivsete testide osakaal kõigist testidest. Järgnevates mudelites on kasutatud vaid nakatumise suhtarvu, kuna see seletab paremini nakatunute sattumist haiglasse.

Mudelist on ükshaaval välja jäetud tunnuseid, mis ei ole statistiliselt olulised. Tulemusena leitud mudeli kordajad on toodud tabelis 3.

Tabel 3. Vanusegruppidega regresioonmudeli kordajad

tunnus	b(se)	p-väärtus
Vabaliige	88.11(4.35)	<0.0001
Nakatumisnäitaja vanusegrupis 0-19	-0.19(0.02)	<0.0001
Nakatumisnäitaja vanusegrupis 20-39	-0.35(0.02)	<0.0001
Nakatumisnäitaja vanusegrupis 60-79	0.48(0.03)	<0.0001
Nakatumisnäitaja üldiselt	0.69(0.06)	<0.0001
Koosmõju nak. ja 7.päeva taguse vak. vahel	-0.001(0.0002)	<0.0001

R-ruut kordaja on 0.9428

Näeme, et siin mudelis on mõningaid anomaaliaid. Üleüldine nakatumine suurendab hospitaliseeritute arvu, kuid kui vaadata nakatumist vanusegruppides, siis vanusegruppides 0-19 ja 20-39 on kordaja negatiivse märgiga ja vanusegrupis 60-79 positiivse märgiga. Üleüldine R-kordaja selle mudeli jaoks on 0.94, mis on hea, kuid mudelisse jäi suhteliselt palju olulisi tunnuseid ja mõningad anomaaliad.

Eelnevalt mudelilt sai leitud, et vaksineeritus on oluline tunnus. Lisame kõige viimasesse saadud mudelisse 5 tunnust, kus iga tunnus näitab vaksineeritust vanusegruppides.

Jällegi on ükshaaval jäetud mudelist välja ebaolulised tunnused, kus $p > 0.005$. Kasutatud on rangemaid p -väärtuseid, kui 0.05 kuna testitakse väga paljusid erinevaid tunnuseid ja tahetakse vähendada mitmese-testimise probleemi. Lõpuks jäidi tabelis 4 kirjeldatud mudeli juurde.

Tabel 4. Vanusegruppidega regresioonmudeli kordajad

tunnus	b(se)	p-väärtus
Vabaliige	201.5(8.82)	<0.0001
Vaktsineerituse osakaal vanusegrupis 0-19	716.5(120.9)	<0.0001
Vaktsineerituse osakaal vanusegrupis 40-59	494.9(111.8)	<0.0001
Vaktsineerituse osakaal vanusegrupis 60-79	-941.5(108.9)	<0.0001
Vaktsineerituse osakaal vanusegrupis 80	263.9(36.77)	<0.0001
Nakatumisnäitaja vanusegrupis 0-19	-0.11(0.01)	<0.0001
Nakatumisnäitaja vanusegrupis 20-39	-0.30(0.02)	<0.0001
Nakatumisnäitaja vanusegrupis 60-79	0.41(0.02)	<0.0001
Nakatumisnäitaja üldiselt	0.52(0.04)	<0.0001
Koosmõju nak. ja 7.päeva taguse vak.vahel	-0.001(0.0002)	<0.0005

R-ruut kordaja on 0.9726

Mudelisse jäi väga palju olulisi tunnuseid. Võrreldes eelmise mudeliga on sellel mudelil vabaliige tunduvalt teistsugune (202 vs 88). Tunnustel, mis esinevad mõlemas mudelis, on vaid väikesed muutused. Siiski on sellel mudelil samasugused probleemid, nagu eelmisel mudelil. Nakatumist näitavatel tunnustel on vastuolulised kordajad ning mudelisse jäi väga palju olulisi tunnuseid. R-kordaja on sellel mudelil 0.972.

Jätkame nüüd sellise mõtteviisiga: Vaktsineeritus vähendab haigestunute tõenäosust sattuda haiglasse. Vaktsineerimas käimine iseeneslikult ei vähenda hospitaliseeritute arvu, vaid ta vähendab tõenäosust sattuda sinna, kui juba olla nakatunud. Lisaks vähendab vaktsineeritus veidi ka nakkuse levimist, kuid see peaks kajastuma juba haigestunute arvus. Proovime sellist mudelit, kus vaktsineerituse asemel kasutame vaid vaktsineerituse koosmõjusid sama vanusegrupi nakatumise näitajaga. Selline mudel parandaks ka järgmise fakti: Kui nakatumine on 0 kõigis vanusegruppides, siis on ka hospitaliseeritus 0.

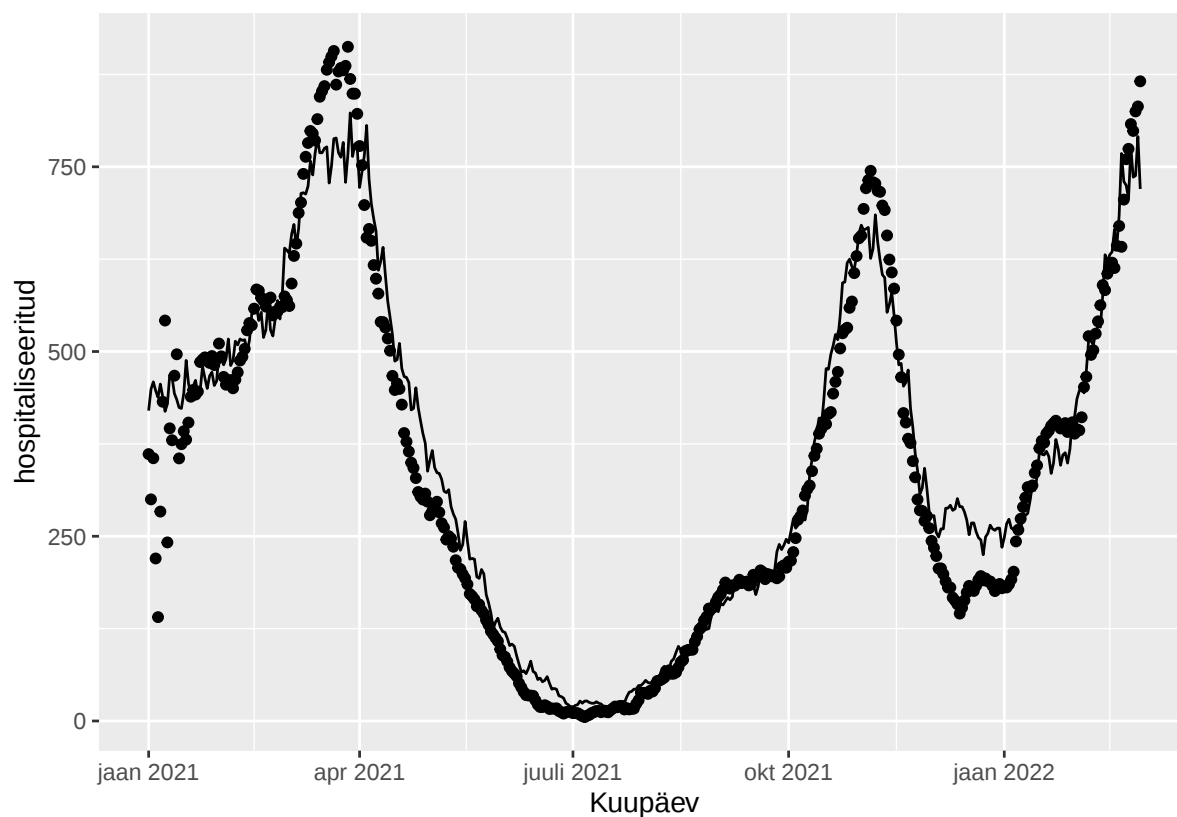
Pärast ebaoluliste tunnuste eemaldamist ($p > 0.005$) jõuame järgmiste kordajateni, mis on toodud tabelis 5

Tabel 5. Vanusegruppidega regresioonmudeli kordajad

tunnus	b(se)	p-väärtus
Nakatumisnäitaja vanusegrupis 0-19	0.66(0.04)	<0.005
Nakatumisnäitaja vanusegrupis 20-39	-0.99(0.05)	<0.005
Nakatumisnäitaja vanusegrupis 60-79	0.93(0.057)	<0.005
Nakatumisnäitaja üldiselt	0.625(0.066)	<0.005
koosmõju vanusegruppis 0-19	-3.39(0.18)	<0.005
koosmõju vanusegrupis 20-39	1.51(0.09)	<0.005
koosmõju vanusegrupis 40-59	-0.96(0.07)	<0.005
koosmõju vanusegrupis 80	-0.167(0.05)	0.002

R-ruut kordaja on 0.9818

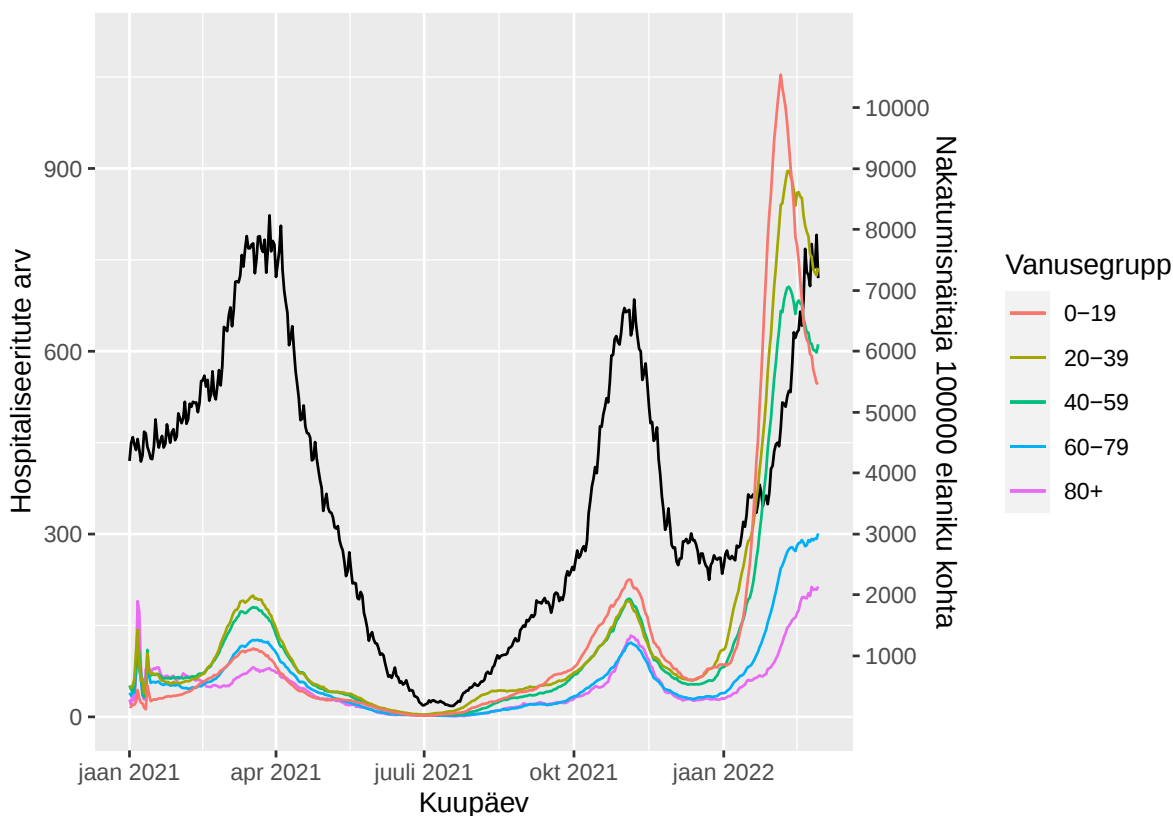
Kui võrrelda selle mudeli tunnuseid eelmise mudeliga, siis kordajad erinevad suhteliselt palju eelmistest mudelitest. Nakatumine vanusegrupis 0-19 vahetas märki ja osad tunnused muutusid olulisteks.



Joonis 4. Mudeli ennustus (Tabelis 5 toodud mudeli põhjal leitud prognoosid (punktid) koos tegelike väärtustega (pidev joon))

Mudeli tulemused on esitatud joonisel 4. Pidev joon on tegelikud väärtused ja punktid on eelmise mudeli pealt ennustatud väärtused. Kui vaadata graafikut, siis mudeli ennustused on väga lähedal tegelikele väärtustele. Samas saame jällegi samasuguse probleemi, nagu eelmiste mudelite korral. Nakatumise näitajad on vastupidiste märkidega ning tunnuste kordajad on kuidagi oluliselt muutunud võrreldes eelmiste graafikutega. Võimalik, et seda saab seletada multikollinearsusega.

Vaatame joonis 5, mis iseloomustab meie mudelit :



Joonis 5. Nakatumine vanusegrupiti ja hospitaliseeritus

Joonisel 5 kujutab must joon hospitaliseeritute arvu ja värvilised jooned kujutavad nakatumise näitajat vanusegruppides. On selgelt näha, et kui nakatumisnäitajad tõusevad, siis tõuseb ka hospitaliseeritute arv, seega võiks eeldada seose olemasolu nakatumise ja hospitaliseeritute vahel. Samas on ka näha, et need jooned käituvad sarnaselt: kui ühe grupi nakatumine tõuseb, siis tõuseb see ka teistel gruppidel, ning kui ühel langeb, siis see langeb ka teistel. Kollineaarust võiks ka seletada pärismaailmale tuginedes: nakatumine ei toimu ainult vanusegrupi siseselt. Näiteks 20-39 aastaste vanusegrupi liige ei nakata ainult teisi oma vanusegrupi liikmeid, ta nakatab kõiki

inimesi. Sellest võiks järeldada multikollineaarsuse olemasolu meie andmetes. Vaatame ka täpselt VIF kordajaid meie mudeli jaoks.

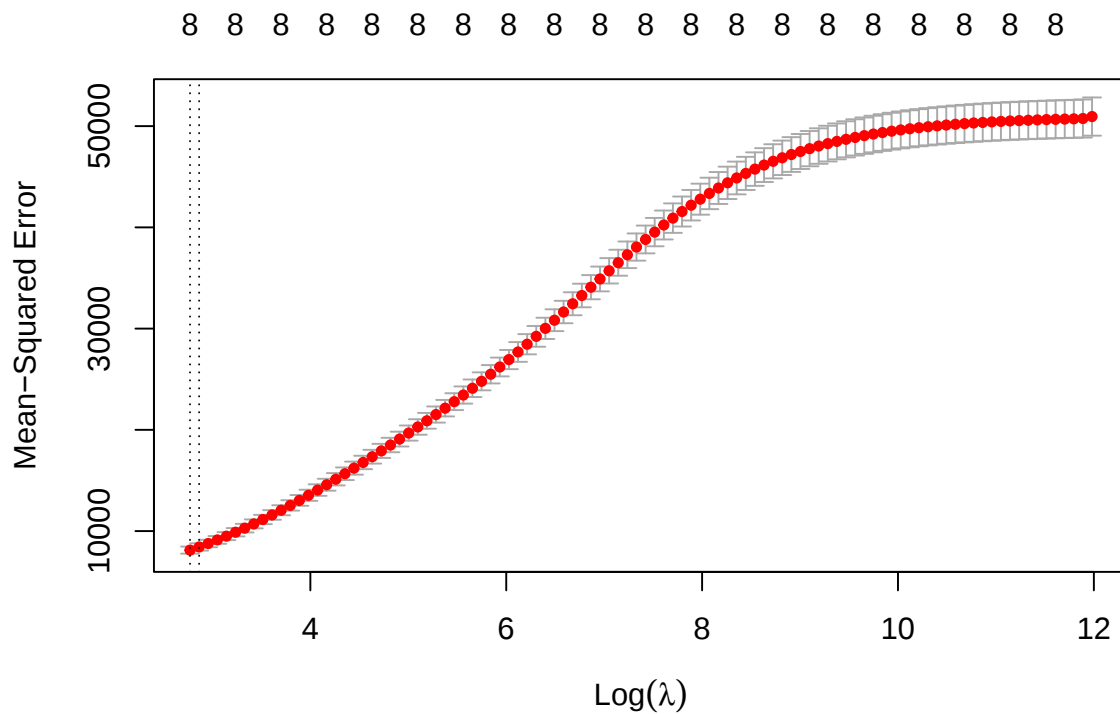
Tabel 6. Vanusegruppidega regresioonmudeli VIF kordajad

Tunnus	VIF
Nakatumisnäitaja vanusegrupis 0-19	1378.18
Nakatumisnäitaja vanusegrupis 20-39	1879.84
Nakatumisnäitaja vanusegrupis 60-79	342.87
Üleüldine nakatumisnäitaja	1995.37
Koosmõju nakatumise ja vaktsineerituse vahel vanusegrupis 0-19	1407.67
Koosmõju nakatumise ja vaktsineerituse vahel vanusegrupis 20-39	2373.40
Koosmõju nakatumise ja vaktsineerituse vahel vanusegrupis 40-59	1230.91
Koosmõju nakatumise ja vaktsineerituse vahel vanusegrupis 80+	65.31

Meie eelmisel mudelil on peaaegu kõik VIF-i väärtused üle 10-e. See tähendab, et meie mudelis eksisteerib väga tugev multikollineaarsus, mis teeb meie kordajate hinnangud mitteusaldusväärseteks. Selle vähendamiseks tuleks läheneda mudeli hindamisele teist moodi. Üks võimalus oleks kasutada ridge regressiooni, mis on mõeldud just mudelite hindamisele olukorras, kus tunnused on tugevalt korreleeritud. Teine võimalus oleks proovida hinnata mudelit mitte tunnuseid ükshaaval eemaldades, vaid ükshaaval lisades. Esiteks võtta mudelisse üks tunnus, mis on kas testiga, või eelnevalt teadaolevalt hospitaliseeritusega korreleeritud. Seejärel lisada ühekaupa tunnuseid juurde, ning iga tunnuse lisamisel teha VIF-test, et vaadata, kas uus lisatud tunnus on tõesti vajalik. Esiteks proovime Ridge regressiooni.

Ridge regressioon

Rakendame ridge regressiooni meetodit kõige viimase mudeli peal (vaata tabel 5) ja vaatame, kas saame teistsugused kordajad.

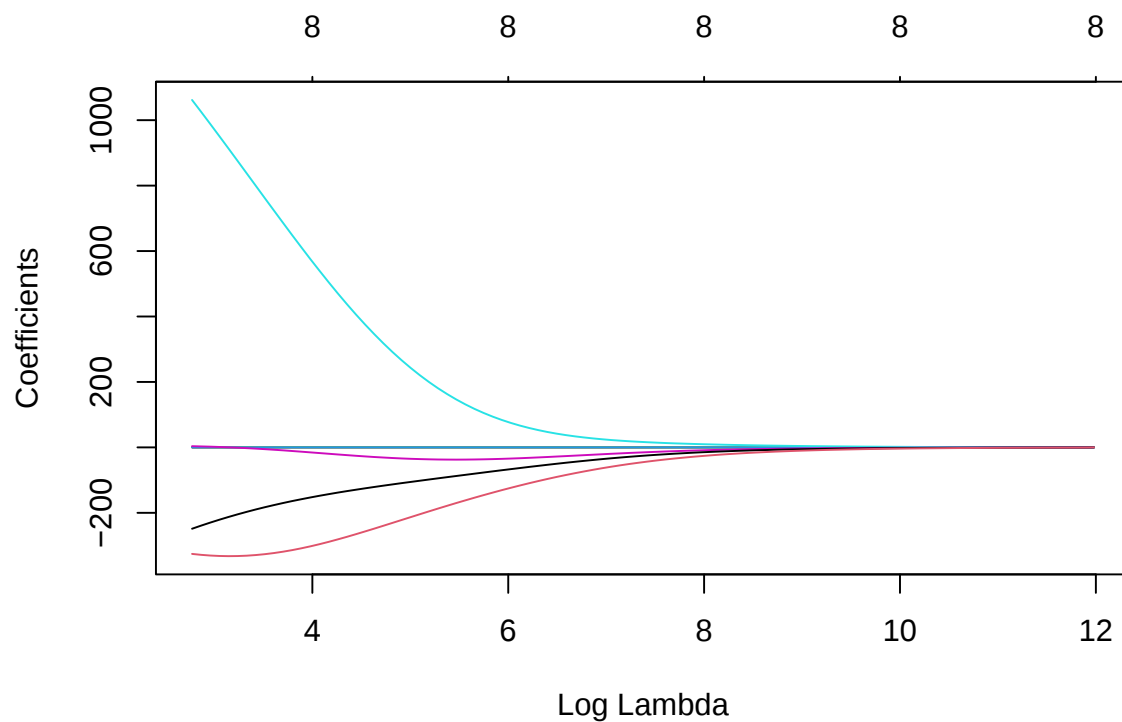


Joonis 6. Tabelis 5 kirjeldatud mudeli ruutkeskmise vea muutus λ väärtuse suurenemisel.

Jooniselt 6 näeme, kuidas käitub testi keskmine viga, kui lamba väärtus suureneb. Optimaalsed (Väikseima ruutkeskmise veaga) konstandi väärtused on graafiku algus punktis, kahe vertikaalse joone vahel. Y-telg näitab MSE (ruutkeskmise viga) suurust ja x näitab konstandi λ suurust. Oluline on märkida, et x-telg on logaritmilise skaalaga. Tuleb välja, et parim λ väärtus on 15.97766. Kasutame seda λ väärtust, et hinnata mudelit.

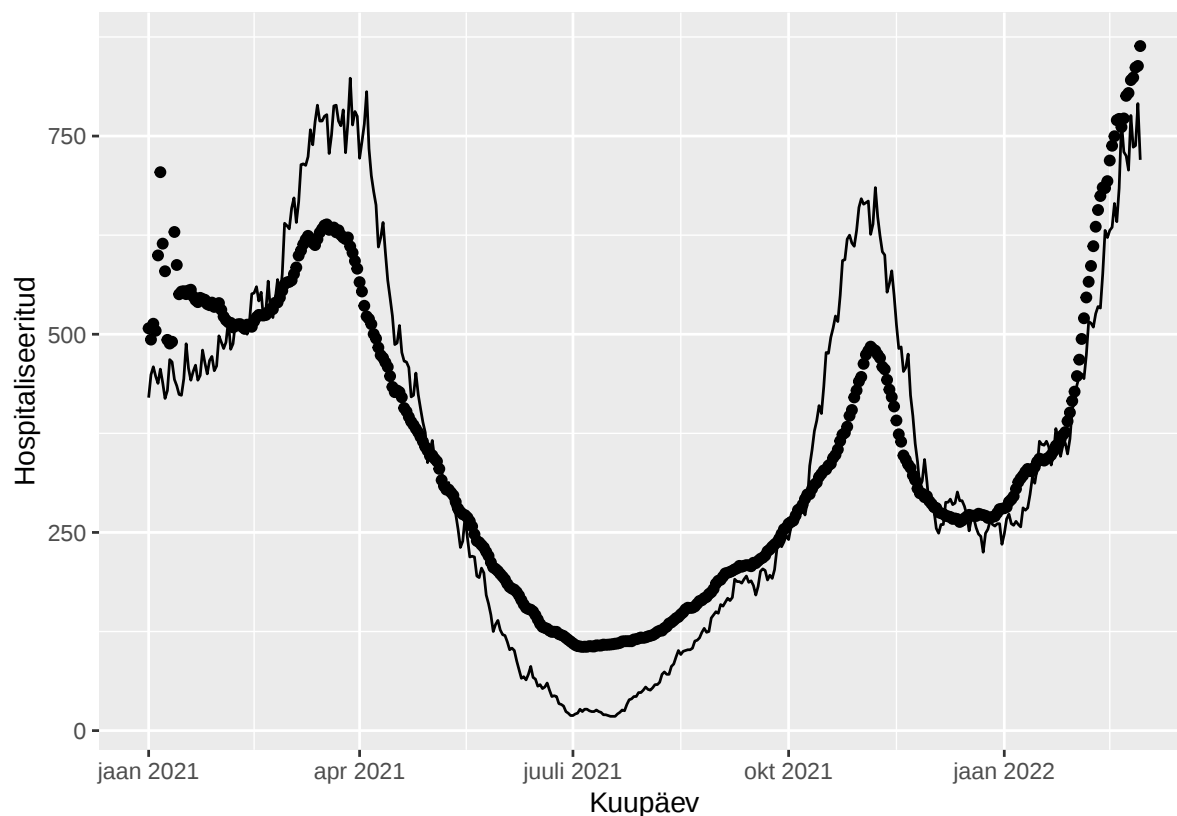
Tabel 7. Vanusegruppidega Ridge mudeli kordajad, kus $\lambda = 15.97766$.

Tunnus	Kordaja
Vabaliige	400.7
Nakatumisnäitaja vanusegrupis 0-19	-0.02
Nakatumisnäitaja vanusegrupis 20-39	-0.03
Nakatumisnäitaja vanusegrupis 60-79	0.31
Üleüldine nakatumisnäitaja	0.002
Koosmõju nakatumise ja vaksineerituse vahel vanusegrupis 0-19	1061.74
Koosmõju nakatumise ja vaksineerituse vahel vanusegrupis 20-39	4.15
Koosmõju nakatumise ja vaksineerituse vahel vanusegrupis 40-59	-0.025
Koosmõju nakatumise ja vaksineerituse vahel vanusegrupis 80+	-0.033



Joonis 7. Ridge jälg tabelis 5 kirjeldatud mudeli jaoks

Joonise 7 pealt näeme, kuidas seletavate tunnuste koefitsendid muutuvad, kui λ väärtust suurendada. Mingist hetkest alates lähenevad kõigi tunnuste koefitsendid ühele ja samale väärtusele nulli lähedal. Tunnuste lähenemine mingile kindlale väärtusele ei tähenda muidugi seda, et mudeli MSE väheneks. Siin kohal on samuti oluline märkida, et x-telg on logaritmilise skaalaga. Arvutatud on ka uue mudeli R^2 kordaja, mis on võrdne $R^2 = 0.8426$.



Joonis 8

Joonis 8 näitab Ridge regressiooni mudeli ennustusi võrreldes tegelike väärtustega, kui võtta $\lambda = 15.977$. Vaadates graafikut, näeb ta suhteliselt samasugune välja modeliga, kus me kasutasime üleüldiseid nakatumisnäitajaid (Vaata joonis 2), kuid suurte muutuste korral on uus mudel veidi reaktiivsem, ehk ennustab muutusi veidi paremini. Kui võrrelda nende mudelite R^2 kordajaid, siis uuel mudelil on see veidi kõrgem (0.843 vs 0.75). Samas on joonisel 4 toodud ennustused märksa täpsemad.

Sellel mudelil on siiski samasuguseid probleeme, mis eelnevatel mudelitel. Osad koefitsendid on vastuoluliste märkidega. Vabaliige on üllatavalt suure väärtusega, mis ei rahulda seda tingimust, et nakatumise puudumise korral oleks hospitaliseeritud ka null. Proovime järgnevalt mudeli peal meetodit, kus sunnime vabaliikme nulliks.

Tabel 8. Vabaliikmeta Ridge mudeli kordajad

Tunnus	Kordaja
Vabaliige	.
Nakatumisnäitaja vanusegrupis 0-19	-0.04
Nakatumisnäitaja vanusegrupis 20-39	-0.09
Nakatumisnäitaja vanusegrupis 60-79	0.68
Üleüldine nakatumisnäitaja	-0.02
Koosmõju nakatumise ja vaktsineerituse vahel vanusegrupis 0-19	373.37
Koosmõju nakatumise ja vaktsineerituse vahel vanusegrupis 20-39	-30.59
Koosmõju nakatumise ja vaktsineerituse vahel vanusegrupis 40-59	-268.85
Koosmõju nakatumise ja vaktsineerituse vahel vanusegrupis 80+	338.92s

Selle mudeli koefitsendid tulid väga erinevad eelnevate mudelite omadest. Koefitsentide märgid on ikka anomaalsed. Lisaks on selle mudeli R^2 kordaja ainult 0.4517 . Vabaliikme ärajätmine mudelist ei ole seega mõttekas tegu.

Ridge regressioon ei lahendanud probleeme, mis meil tekkisid. Mudeli kordajad jäid vastupidiste märkidega, mis ei ole loogiline. Proovime järgmisena mudeli ehitamist teisel meetodil, ehk edasisuunas mudeli valik.

Edasisuunas mudeli valik

Kuna Ridge regressioon ei andnud soovitud tulemust, siis proovime nüüd ehitada mudelit ülesse nii, et lisame sinna ühekaupa olulisi tunnuseid. Kõigepealt võtame mudelisse tunnuseid, mis on juba teadaolevalt korreleeritud hospitaliseeritusega, ning siis ühekaupa proovime lisada juurde tunnuseid. Kui tunnuse lisamisega peaks mudel ebasobivaks muutuma või tunnus ebaoluline olema, siis jätame selle välja. Uute tunnuste valikul lähtume kas eelnevatest testidest või teadmistest.

Alustuseks võtame mudeli, kus olid üldised nakatumis ja vaktsineerimisnäitajad (vaata tabel 2). Ennem, kui lähme ühekaupa tunnuste lisamise juurde, võiks selle mudeliga teha seda, mida tegime vanusegruppidega mudelite juures: Jätame ära vaktsineerimise peamõju, ning vaatame vaid koosmõju haigestumisega.

Tabel 9. Üldise regressioonmudeli kordajad

tunnus	b(se)	p
Vabaliige	145.1(8.557)	<0.00001
Nakatumisnäitaja 100000 elaniku kohta	0.507(0.016)	<0.00001
Nakatamise ja vaktsineerituse koosmõju	-0.0067(0.00024)	<0.00001

R-ruut kordaja on 0.7287

Sellest mudelist on eriti hästi näha, et üldiselt nakatumine suurendab hospitaliseeritust ning vaktsineeritus vähendab hospitaliseeritust, mille mõju suureneb nakatamise suurenemisel. Nüüd tuleks vaadata, missuguste tunnuste lisamine mudelisse annaks meile võimaliult palju infot selle kohta, kuidas hospitaliseeritus muutub. Võime testimiseks teha 5 mudelit, kus igas mudelis vaatame eraldi vanusegrupi haigestumise ja vaktsineerituse mõju lisamise mõju mudelile.

Tabel 10. 0-19 vanusegrupiga mudeli kordajad

tunnus	b(se)	p
Vabaliige	63.41(9.88)	<0.00001
Nakatumisnäitaja 100000 elaniku kohta	0.71(0.025)	<0.00001
Nakatumisnäitaja vanusegrupis 0-19	-0.17(0.031)	<0.00001
Vaktsineerituse osakaal vanusegrupis 0-19	736.8(101.0)	<0.00001
Nakatumise ja vaktsineerituse koosmõju	-0.0061(0.00059)	<0.00001

R-ruut kordaja on 0.8179

Tabel 11. 20-39 vanusegrupiga mudeli kordajad

tunnus	b(se)	p
Vabaliige	42.07(12.23)	<0.00001
Nakatumisnäitaja 100000 elaniku kohta	1.15(0.042)	0.00064
Nakatumisnäitaja vanusegrupis 20-39	-0.44(0.0031)	<0.00001
Vaktsineerituse osakaal vanusegrupis 20-39	279.1(28.52)	<0.00001
Nakatumise ja vaktsineerituse koosmõju	-0.0089(0.0003)	<0.00001

R-ruut kordaja on 0.8419

Tabel 12. 40-59 vanusegrupiga mudeli kordajad

tunnus	b(se)	p
Vabaliige	93.888(17.355)	<0.00001
Nakatumisnäitaja 100000 elaniku kohta	-0.23(0.099)	<0.02136
Nakatumisnäitaja vanusegrupis 40-59	0.67(0.08)	<0.00001
Vaktsineerituse osakaal vanusegrupis 40-59	99.3(33.07)	0.003
Nakatumise ja vaktsineerituse koosmõju	-0.0054(0.0005)	<0.00001

R-ruut kordaja on 0.7822

Tabel 13. 60-79 vanusegrupiga mudeli kordajad

tunnus	b(se)	p
Vabaliige	146.7(13.24)	<0.00001
Nakatumisnäitaja 100000 elaniku kohta	-0.20(0.03)	<0.00001
Nakatumisnäitaja vanusegrupis 60-79	0.80(0.028)	<0.00001
Vaktsineerituse osakaal vanusegrupis 60-79	-94.29(21.28)	<0.00001
Nakatamise ja vaktsineerituse koosmõju	-0.000070(0.00036)	0.845

R-ruut kordaja on 0.9142

Tabel 14. 80+ vanusegrupiga mudeli kordajad

tunnus	b(se)	p
Vabaliige	5.027(16.75)	0.764
Nakatumisnäitaja 100000 elaniku kohta	0.36(0.016)	<0.00001
Nakatumisnäitaja vanusegrupis 80+	0.41(0.016)	<0.00001
Vaktsineerituse osakaal vanusegrupis 80+	113.3(24.00)	<0.00001
Nakatamise ja vaktsineerituse koosmõju	-0.0056(0.00026)	<0.00001

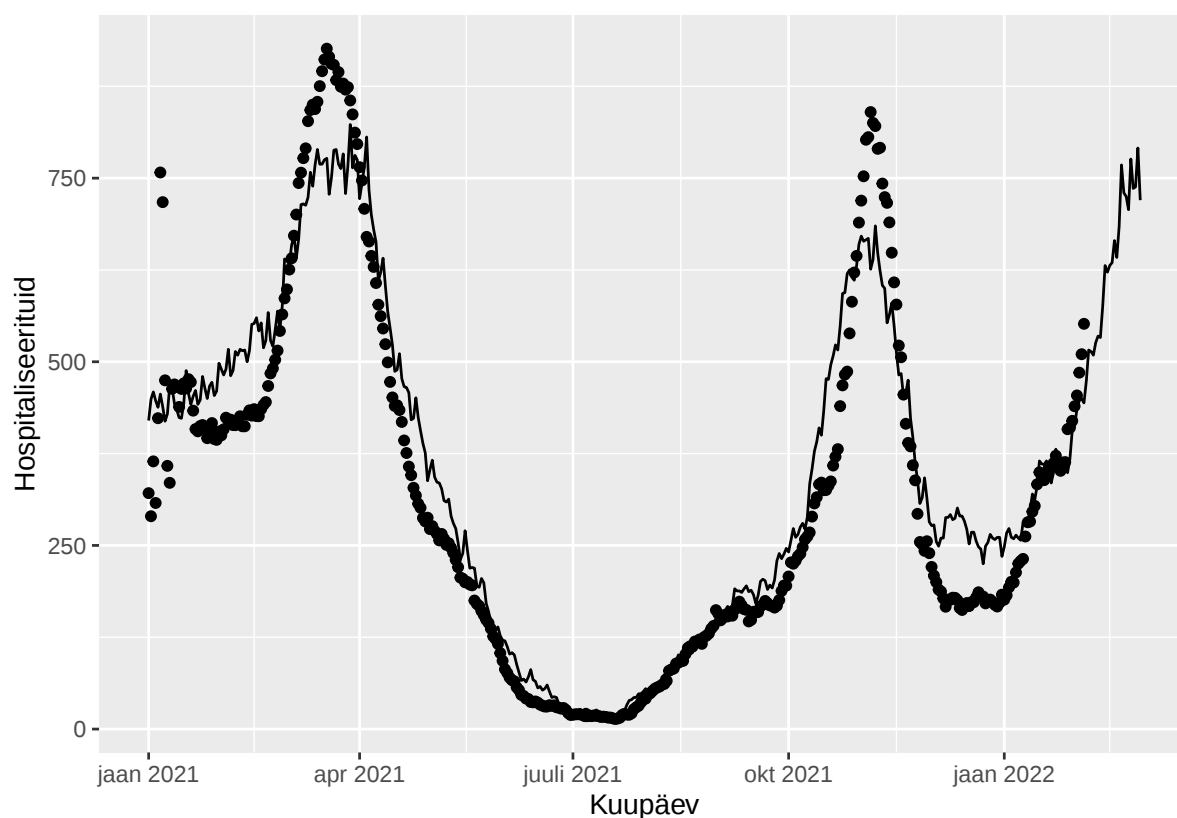
R-ruut kordaja on 0.8957

Vaadates nende mudelite R-ruut kordajaid, siis 60-79 vanusegrupi ja 80+ vanusegrupi lisamised suurendasid seda enim. Lisaks on 80+ vanusegrupis vabaliige ebaoluline, mis aitab meil selle ära jätta ja lahendada 0 nakatamise probleemi. Vaatame 80+ vanusegrupiga mudelit veidi edasi, kui jätame vabaliikme välja.

Tabel 15. 80+ vanusegrupiga mudeli kordajad

tunnus	b(se)	p
Nakatumisnäitaja 100000 elaniku kohta	0.43(0.014)	<0.00001
Nakatumisnäitaja vanusegrupis 80+	0.27(0.021)	<0.00001
Üldine vaktsineerituse ja nakatumise koosmõju	-0.007(0.0002)	<0.00001
Nakatumise ja vaktsineerituse koosmõju vanusegrupis 80+	0.47(0.03)	<0.00001

R-ruut kordaja on 0.9695



Joonis 9. Tabelis 15 kirjeldatud mudeli ennustused (punktid) ja tegelikud väärtused (pidev joon)

Joonis 9 on viimase mudeli võrdlus tegelike väärtustega. See mudel on väga lähedal tegelikule hospitaliseeritusele (R-ruut kordaja on 0.97). Ainuke probleem selle mudeli juures on kordajate vastuolulised väärtused. Üldisel vaktsineerituse koosmõjul üldise nakatumisega on negatiivne väärtus, aga samal koosmõjul vanusegrupis 80+ on positiivne väärtus. On võimalik, et seda saab parandada teise vanusegrupi juurde lisamisega. Lisame järgmiselt juurde vanusegrupi 60-79, kuna eelnevalt sai nähtud, et see suurendas R-ruutu enim.

Tabel 16. Kahe vanusegrupiga mudeli kordajad

tunnus	b(se)	p
Nakatumisnäitaja 100000 elaniku kohta	0.73(0.057)	<0.00001
Nakatumisnäitaja vanusegrupis 80+	0.33(0.03000)	<0.00001
Nakatumisnäitaja vanusegrupis 60-79	-0.37(0.083)	<0.00001
Vaktsineerituse ja hospitaliseerituse koosmõju	-0.014(0.00096)	<0.00001
Nakatumise ja vaktsineerituse koosmõju vanusegrupis 80+	-0.066(0.109)	0.545
Nakatumise ja vaktsineerituse koosmõju vanusegrupis 60-79	1.18(0.18)	<0.00001

R-ruut kordaja on 0.9733

Vanusegruppi 60-79 juurde lisamisel muutusid kordajad väga palju, ning R-ruut ei läinud väga paremaks. Samamoodi tekib meil uuesti multikollineaarsuse probleem, kui tuua sisse liiga palju muutujaid. Vaatame kõige viimase mudeli VIF kordajaid.

Tabel 17. Kahe vanusegrupiga mudeli VIF kordajad

tunnus	VIF
Nakatumisnäitaja 100000 elaniku kohta	439.98
Nakatumisnäitaja vanusegrupis 80+	22.19
Nakatumisnäitaja vanusegrupis 60-79	256.97
Vaktsineerituse ja hospitaliseerituse koosmõju	382.84
Nakatumise ja vaktsineerituse koosmõju vanusegrupis 80+	84.31
Nakatumise ja vaktsineerituse koosmõju vanusegrupis 60-79	379.58

Näeme, et VIF- kordajad tõusevad kuni 500-ni, mis on loogiline, kuna on eelnevalt teada, et nakatumine on vanusegruppide vahel korrelatsioonis, ning me lisasime vanusegruppe juurde. Kui vaadata vaid 80+ vanusegrupi mudeli VIF kordajaid, siis saame järgmised arvud:

Tabel 18. 80+ vanusegrupiga mudeli VIF kordajad

tunnus	VIF
Nakatumisnäitaja 100000 elaniku kohta	23.77
Nakatumisnäitaja vanusegrupis 80+	9.66
Vaktsineerituse ja hospitaliseerituse koosmõju	18.62
Nakatamise ja vaktsineerituse koosmõju vanusegrupis 80+	7.04

Kuigi siit keskmine tuleb üle 10, mis teadaolevalt oli piiriks multikollineaarsuse suurele mõjule, on need võrreldes kõigi eelmiste mudelitega, mis me oleme teinud, väga väikesed. Ükskõik millise vanusegrupi andmete mudelisse juurde lisamisel tõusevad VIF kordajad kohe väga kõrgeks, mis teeb mudeli täpsemaks tegemise keeruliseks. Seega osutus parimaks mudeliks, kus hospitaliseerimise prognoosimisel oli lisaks üldistele näitajatele võetud arvesse ka nakatamine ning nakatamise ja vaktsineerituse koosmõju vanusegruppis 80+.

Kokkuvõte

Esmalt leidsime, et hospitaliseerimine on ootuspäraselt seda suurem, mida suurem on üldine nakatumine. Seejärel nägime ka, et vaktsineeritusel oli vähendav mõju, mis esines koosmõjuna nakatumisega - mida kõrgem oli nakatumine, seda rohkem vähendas vaktsineeritus hospitaliseeritute arvu.

Kui eesmärgiks on hospitaliseeritute arvu prognoosimine, siis ainult üldist nakatumist ja vaktsineeritust arvestava mudeli prognoosid on suhteliselt ebatäpsed. Mudel ei suuda väga hästi muutusi ennustada, ning jääb muutuste suhtes konservatiivseks. Tuues sisse erinevad vanusgrupid, said mudelid näiliselt täpsemaks, kuid siis tekkis multikollineaarsuse probleem.

Üldiselt jäid mudelitele järgnevad probleemid:

- Vanusegrupid mudelitele jäid vastuolulised kordajad (Kui üldised nakatumis või vaktsineerimis kordajad olid ühe märgiga, siis vanusegrupid samad kordajad olid vastupidise märgiga.)
- Mudelite kordajad muutuvad suhteliselt palju, kui mudelile informatsiooni juurde anda
- Vabaliige peaks hea mudeli korral olema 0, mis näitaks hospitaliseerimise puudumist nakatumise puudumise korral, kuid enamustel mudelitel esines 0-st erinev oluline vabaliige.

Kõige viimane mudel, milles olid üldised näitajad ja vanusegrupi 80+ näitajad, prognoosis hospitaliseerimist üldiselt hästi. Kui eesmärgiks on ühe nädala võrra ette hospitaliseeritute arvu prognoosimine, siis sobiks eelpool mainitud mudel hästi. Mudelit tuleks siis iga nädal uuesti hinnata, et mudel täpsust ei kaotaks. Kui eesmärgiks on andmete põhjal järelduste tegemine, siis tuleks kaasata nakatumise ja vaktsineeritusega mittekollineaarset infot (näiteks sugu, elukoht, jne.) Mittekollineaarne info vältib eelpool arutelus tekkinud probleeme.

Kõige viimaste mudelite juures sai näha, et 60-79 ja 80+ vanusegrupiga mudelid olid kõige suurema R-ruut kordajaga. Üks võimalus oleks liita need vanusegrupid kokku ja vaadata, kas nende koos esinemine ennustab paremini hospitaliseeritust.

Kasutatud kirjandus

- (1) Terviseameti koroonakaart (<https://www.terviseamet.ee/et/koroonaviirus/koroonakaart>)
- (2) Terviseameti avaandmed (<https://www.terviseamet.ee/et/koroonaviirus/avaandmed>)
- (3) Tervise ja Heaolu infosüsteemide keskuse avaandmed (TEHIK) (<https://opendata.digilugu.ee/docs/#/>)
- (4) M.H.Kutner, C.J.Nachtsheim, J.Neter, W.Li (2005). Applied Linear Statistical Models. 5.väljaanne. McGraw-Hill Irwin ,lk 279-289 , 406-408, 431-432 , loetud veebist : (http://users.stat.ufl.edu/~winner/sta4211/ALSM_5Ed_Kutner.pdf)
- (5) Statology leht (<https://www.statology.org/ridge-regression-in-r/>)

Lihtlitsents lõputöö reprodutseerimiseks ja üldsusele kättesaadavaks tegemiseks

Mina, Joonas-Sander Tamm,

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) minu loodud teose "Covid-19 hospita-liseerimise prognoosimine nakatunute andmetelt", mille juhendaja on Krista Fischer, reprodutseerimiseks eesmärgiga seda säilitada, sealhulgas lisada digitaalarhiivi DSpace kuni autoriõiguse kehtivuse lõppemiseni.
2. Annan Tartu Ülikoolile loa teha punktis 1 nimetatud teos üldsusele kättesaadavaks Tartu Ülikooli veebikeskkonna, sealhulgas digitaalarhiivi DSpace kaudu Creative Commonsi litsentsiga CC BY NC ND 3.0, mis lubab autorile viidates teost reprodutseerida, levitada ja üldsusele suunata ning keelab luua tuletatud teost ja kasutada teost ärieesmärgil, kuni autoriõiguse kehtivuse lõppemiseni.
3. Olen teadlik, et punktides 1 ja 2 nimetatud õigused jäävad alles ka autorile.
4. Kinnitan, et lihtlitsentsi andmisega ei riku ma teiste isikute intellektuaalomandi ega isikuandmete kaitse õigusaktidest tulenevaid õigusi.

Joonas-Sander Tamm

10.05.2022