TARTU ÜLIKOOL

LOODUS- JA TÄPPISTEADUSTE VALDKOND

MATEMAATIKA JA STATISTIKA INSTITUUT

Nora Roosileht

Swedbank P&C Insurance AS-i reisikindlustuse preemia analüüs

Magistritöö (30 EAP)

Finants- ja kindlustusmatemaatika eriala

Juhendaja Meelis Käärik

Tartu

2016

Swedbank P&C Insurance AS-i reisikindlustuse preemia analüüs

Käesoleva magistritöö eesmärk on uurida ja analüüsida Swedbank P&C Insurance AS-i kasutatavat reisikindlustuse preemia valemit. Antud töö mõte on teha kindlaks, kas selle valemi abil leitud preemiad katavad kahjude kulud ning vajadusel teha ettepanek selle korrigeerimiseks. Analüüsi teaostamiseks rakendatakse lineaarseid mudeleid ning fikseeritud naabruses keskmise leidmist, kus piirkonnad on määratud k-lähima naabri meetodit kasutades. Et seda paremini mõtestada, tehakse ülevaade ka eel mainitud meetodite teoreetilisest taustast.

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

Märksõnad: Reisikindlustus, kahjukindlustusmatemaatika

The Premium Analysis of Swedbank P&C Insurance AS

The purpose of this thesis is to investigate and analyse the premium calculations in Swedbank P&C Insurance AS. It is aimed to make sure if the current premium equation realises in its objective, which is to gain more premium than pay out loss amounts. In case of finding malfunction in the equation, requests for correction will be made to the company. To conduct the analysis, there is used linear models and finding means in fixed area that are determined by k-nearest neighbours method. To make the analysis carried out more reasonable, an overview of theoretical background of used methods is done as well.

CERCS research specialisation: P160 Statistics, operation research, programming, actuarial mathematics

Keywords: Travel insurance, non-life insurance mathematics

Sisukord

Sissejuhatus

Iga kindlustusselts on äriettevõte, mille eesmärgiks on teenida kasumit. See saadakse suures osas klientide makstavatelt preemiatelt, kas regulaarselt või ühekorde maksena olenevalt kindlustusliigist. Samuti olenevalt kindlustatavast objektist võivad kahjud osutuda väikesteks või vastupidi väga suurteks. Võiksime siis ajaloolistele juhtumitele tuginedes küsida igalt inimeselt maksimaalse preemia, et korvata võimaliku suurima kahju ning seal juures säilitada kasumlikkus. Siin aga tuleb mängu konkurents turul. Kuna Eestis on tänaseks päevaks juba arvestatav hulk erinevaid kindlustusseltse, küsides jäigalt maksimaalset preemialt, võime kaotada kliendi väga lihtsal põhjusel - konkurent pakub samasugust toodet parema hinnaga. Seega ongi väga oluline kindlustusseltsil leida optimaalne preemia, et oleks puhver korvamaks kliendil tekkinud kahjusid ning samal ajal teenida ärilist kasumit.

Antud töös uuritaksegi, kui hästi küsib Swedbank P&C Insurance preemiat reisikindlustuse eest. Töötatakse läbi preemia arvutamise valem, testides seda ajaloolistel andmetel. Kui õnnestub leida koht, kus saaks kliendile pakkuda paremat hinda ning seejuures võita neid juurde või, kus oleks vajalik hoopis preemia osa suurendada, tehakse need ettepanekud eel mainitud ettevõttele.

Analüüsiks mõeldud andmed on antud kasutamiseks kindlustusseltsi enda poolt, millele on rakendatatud arvutused ja kasutatavad meetodid statistika tarkvarapaketis R. Töö on koostatud kasutades tekstitöötlusprogrammi LaTex. Graafikud on valminud kasutades R-i või tabeltöötlusprogrammi Excel.

Autor tänab Swedbank P&C Insurance AS-i koostöö eest ning juhendajat dotsent Meelis Käärikut asjakohaste märkuste ja konsultatsioonide eest.

1 Swedbank P&C Insurance AS-i reisikindlustuse tutvustus

Swedbank P&C Insurance AS-i näol on tegu kindlustusseltsiga ehk ettevõttega, mis pakub kaitset riski vastu, et mingi juhuslik sündmus häirib kindlustuvõtja igapäevaelu ning, mille tagajärjeks on rahaliselt välja käidav mitte planeeritud kulutus.

Lisaks liiklus- ja kasko-, kodu-, laenumakse- ning krediitkaardi kindlustusele pakub käesolev ettevõte ka reisikindlustuse toodet. Reisi all mõeldakse ajutist viibimist ühes või mitmes välisriigis, mis algab kodust või töölt lahkumisega ning lõpeb koju või tööle tagasijõudmisega. Reisikindlustuse poolt kindlustatu saab olla vaid see isik, kelle alaline elukoht on Eesti Vabariik [1]. Varasemalt oli selles seltsis reisikindlustus omavastutuseta kindlustusliik, kuid alates 17.09.2015 on lisandunud antud liigi tingimustesse ka omavastutuse punkt, mida küsitakse vaid varalise kahju, reisitõrke-, vastutus- ja õigusabikindlustuse korral. Kuna antud töös kasutamiseks on andmed enne eel mainitud kuupäeva, on kõik kahjujuhtumid omavastutuseta.

Swedbank P&C Insurance'i reisikindlustuse toode sisaldab õnnetusjuhtumi ja meditsiiniabi kindlustust (ravikulud ja sellega seonduvad lisakulud). Samuti on selle toote sees ka reisitõrke kaitse, mille all peetakse silmas reisi ärajäämist, katkemist, transpordivahendile hilinemist või lennureisi edasilükkumist, ja pagasikindlustus ehk varakindlustus, milleks on reisile kaasa võetud esemed. Olenevalt reisi eesmärgist, on võimalik soetada reisikindlustuses sisalduvatele kaitsetele ka lisakaitseid, kui tegu on spordireisiga, ohtliku tegevusega reisil ning välisriigis töötamisega.

Kliendile pakutava reisikindlustuse preemia ehk kindlustusmakse hind sõltub kaitse suurusest, mida kindlustatu saab lepingu sõlmides ise valida.

Alates 2015. aasta septembrist on meditsiiniabi kindlustuskaitse fikseeritud 500 000 € suuruse summaga. Varasemalt sai kindlustatu valida kolme erineva suurusega kaitse vahel: 80 000 €, 50 000 € ja 32 000 €. Kuna meditsiiniabi, mis rakendub nii haigestumise kui õnnetusjuhutmi korral, on olnud enim vaja minev kaitse Swedbanki reisikindlustust ostnute seas (vt ptk 4), siis on likvideeritud võimalus valida liiga väike kaitse. Paraku sageli on kahju kulud suuremad kui eel mainitud kolm kaitset. Näiteks ka If kindlustus kasutab fikseeritud summat maditsiiniabi kaitse puhul. Küll aga on kliendil võimalus valida kaitse suurus pagasile/varale ning reisitõrkele. Nendeks summadeks on vastavalt 0 €, 500 €, 1 000 € ja 2 000 € varakindlustuse ning 0 €, 500 €, 1 500 € ja 2 500 € reisitõrke kindlustuse eest. Enne 2015. aasta septembrit olid summad 0 €, 500 €, 1 000 € varakindlustuse ning 0 €, 500 €, 1 000 € ja 2 000 € reisitõrke eest. Antud töös on kasutada andmed varasemate kaitsetega.

Näide.

Oletame, et reis algab 1. juunil ja lõpeb 8. juunil ning reisile minejal on sõlmitud reisikindlustusleping. Juhtumisi 3. juunil tekib reisil olles vältimatu vajadus arstiabi järele, mis kujutab endast rahalist väljaminekut vastavalt kohalikule hinnakirjale. Kuna reisikindlustus katab haigestumise reisi toimumise ajal, korvab selts kindlustatule tema raviks kulunud vajaliku summa (kaitse ulatuses) reisilt naastes.

Mis juhtudel kindlustatule kahjusid ei hüvitata, võib leida Swedbank P&C Insurance AS-i reisikindlustuse tingimustest [1]. Antud töös nendele aspektidele ei keskenduta. Samuti vajab märkimist, et ei võeta arvesse neid preemiaid ning kahjusid, mis puudutab krediitkaardiga kaasnevat reisikindlustust.

2 Preemia arvutamise ülesanne

Kahjukindlustusseltsi üheks suurimaks probleemiks saab pidada õiglase preemia arvutamist kindlustusvõtjale.

2.1 Klassifitseerimisülesanne

Staatilised klassid

Kõige lihtsam viis suurte andmemahtude ja vahemike puhul on jagada andmed staatilistesse klassidesse. Jagamisi ei pea tegema mingi kindla algoritmi alusel. Seejärel teostatakse saavutatud klassides vajalikuks peetavad analüüsid ning tehakse järeldused just konkreetse grupi jaoks.

Nii nagu üldse kindlustuses, on ka reisikindlustuses levinud jagada kindlustusvõtjaid gruppidesse: Euroopasse ja väljapoole reisijad, tava- või tööreisile suundujad jms. Selliste nominaaltunnuste puhul on mõistetav, et kindlustusvõtjate preemiad on personaalselt erinevad, kui kuulutakse erinevatesse klassidesse. Näiteks mujal maailmas väljaspool Euroopat võivad meditsiiniabikulud olla eurooplasele kordades kallimad kui Euroopa riikides. Sellest võib tuleneda kallim poliisi hind Euroopast välja reisides. Kui nüüd kasutada staatilistesse klassidesse jagamisel pidevaid tunnuseid (nt vanus või reisi kestus päevades), võib näiliselt väike muutus tuua sisse väga suure erinevuse poliisi hinnas. Kuna mudelid leitakse igas klassis eraldi, siis ei saa garanteerida, mis juhtub ühest klassist teise liikumisel ning kindlustusmakse suurus võib tõusta mitmekordselt. Sellist olukorda nimetame hinnašokiks. Näiteks, kui klassi piir jookseb 25. ja 26. eluaasta vahelt, erineb küsitav preemia selles vanuses olevatel inimestel, kuigi tegelikult ei saa öelda, et 25- ja 26-aastane erineksid teineteisest suurel määral. Oletame, et tegu on lojaalse kliendiga, kes reisib sagedasti, näiteks aastas korra, ning soetab kõik reisikindlustused alati sellest seltsist. Kui aasta vanemaks saanuna tabab teda kindlustuskaitse soetamisel põhjendamatult suur hinnatõus, võime ta sootuks klientide hulgast kaotada.

Lineaarne mudel

Staatilistest klassidest vabanemiseks oleks vaja andmetele lähenemiseks dünaamilisemat meetodit. Üheks võimaluseks on kasutada lineaarset mudelit, mis eemaldab piirid klasside vahel sootuks ning likvideerib hinnašoki probleemi. Tunnuste vahel leitakse lineaarne funktsioon, mis muutuseid edasi annab. Nominaalsete tunnuste jaoks suures osas ei muutu midagi. Mõte jääb samaks - liikudes ühest grupist teise, on y-tunnuse tulemus erinev. Samas on ka sellisel lähenemisel puudused preemia leidmiseks kindlustuse praktikas. Arvesse võetakse kõik andmepunktid valitud tunnuste seast. Ka väga erineva taustaga ajaloolised objektid avaldavad võrdselt mõju uue poliisi tarbeks preemia leidmisel. Kindlustusvaldkonna puhul ei ole selline lähenemine kuigi mõistlik, sest kahjujuhtumid erinevad paljuski inimeste elustiilist, -korraldustest, harjumustest. Näiteks õnnetusjuhtumite sageduse ja rahalise ulatuse puhul ei pruugi olla otstarbekas 25-aastasele kliendile leitava mudeli jaoks arvesse võtta 60-aastase kliendi andmeid ning vastupidi, sest nende käitumine võib olla väga erinev. Nooremad inimesed on hulljulgemad ja ka nende reisil ettevõetavad tegevused erinevad sageli vanemaealiste reisijate plaanidest. Selline tendents on üldtuntud teave ka näiteks liikluskindlustuse praktikas. Paraku reisil juhtuvad õnnetused on sagedamini esinenud nooremate inimeste seas, samas haigestumise puhul võib täheldada vastupidist olukorda.

K-lähima naabri meetod

Kahe eelneva meetodi kombinatsioonina dünaamilist klassifitseerimist pakub k-lähima naabri meetod. See võimaldab jagada varasemad juhtumid iga uue poliisi korral uutesse klassidesse. Samuti ei ole selle lähenemise puhul kindlaks määratud üks ja ainus meetod hinnangute leidmiseks. Moodustunud klassile võib rakendada regressioonanalüüsi, fikseeritud keskmise leidmist või mõnda muud asjakohast meetodit. Klassifitseerimisülesande korral on tulemuseks grupp, kuhu uuritav punkt võiks kuuluda, võttes arvesse talle lähedasi andmepunkte mingi tingimuse alusel. Sellisesse gruppi kaasatakse vaadeldava punkti k lähimat naabrit.

2.2 Hinnangu leidmise ülesanne

Edasises kasutame järgmiseid tähistusi

- n - poliiside arv,
- l - argumenttunnuste arv,
- t_i - i-nda poliisi reisi kestus päevades, $i = 1, \ldots, n$.

Edaspidi on samaväärne tähendus kindlustusvõtjal, inimesel, objektil ja poliisil, sest isegi, kui on korduvaid kliente andmestikus, on nad määratud erinevate poliisidega. See tähendab, et ühe reisi korral ühele inimesele kehtib üks poliis.

Staatilised klassid

Staatilisteks klassideks klassifitseerimise (klasterdamise) meetodeid on väga palju. Milline konkreetne meetod valida, sõltub ülesande ja andmete iseloomust. Saab sobitada nii lihtsamaid kui keerulisemaid mudeleid klassides. Näiteks, võib leida lineaarseid mudeleid, mille hinnangud kujunevad nii, nagu on kirjeldatud järgnevas lõigus. Samuti võib proovida sobitada kahjudele mõnda jaotust ning leida hinnangud vastava jaotuse parameetrite ja omaduste järgi. Kuna aga selles töös staatilistesse klassidesse andmeid ei jagata ning nendes mudeleid ei leita, siis pikemalt siin alapunktis ei peatuta.

Lineaarsed mudelid

Tugineme selles osas allikale [2, lk 231-237]. Olgu meid huvitav uuritav tunnus Y, mida soovime kirjeldada argumenttunnuste abil. Tähistame i-nda poliisi jaoks vektori $\mathbf{x}_i' = (1, x_{i1}, ..., x_{il})$, kus $x_{i1}, ..., x_{il}$ on vastavad argumenttunnuste väärtused. Eeldame, et see vektor sisaldab ka vabaliiget. See tähendab, et $\mathbf{x}_{i_1} = 1$. Tähistagu y_i uuritava tunnuse Y väärtust i-nda poliisi korral, siis lineaarne mudel y_i kirjeldamiseks vektori x_i põhjal on järgmisel kujul:

$$y_i = \boldsymbol{\beta}'\mathbf{x}_i + e_i, \tag{1}$$

kus e_i on nullkeskmisega ning konstantse dispersiooniga σ^2 ja $\boldsymbol{\beta}' = (\beta_0, \beta_1, ..., \beta_l)$ on tundmatute parameetrite vektor, millele on vaja leida hinnanguid andmete põhjal. Koefitsientide β jaoks on vaja leida hinnangud vaadeldavate andmete $(y_i, \mathbf{x}_i)$ abil. Selle jaoks kasutatakse vähimruutude meetodit ja leitakse hinnangud $\hat{\beta}$, mille abil saame esitada hinnangud $\hat{y}_i$ (mudeli prognoosid) kujul

$$\hat{y}_i = \hat{\boldsymbol{\beta}}'\mathbf{x}_i. \tag{2}$$

Saadud parameetrite hinnanguid $\hat{\beta}$ rakendame peatükis 5.1 ning prognoose $\hat{y}_i$ tulemuste võrdlemisel peatükis 5.3.

K-lähima naabri meetod

Dünaamilisemat lähenemist klasterdamiseks pakub k-lähima naabri meetod [3, lk 459-481]. Selle korral hinnatakse uue poliisi preemia kindlustusvõtja parameetrite järgi, kasutades temale parameetrite poolest sarnaseid objekte, uskudes, et uue kliendi võimalikud kahjud on lähedased temaga sarnaste klientide kahju suurustega. Tuletame meelde, et Eukleidiline kaugus kahe punkti $\mathbf{x} = (x_1, \ldots, x_l)$ ja $\mathbf{z} = (z_1, \ldots, z_l)$ vahel ruumis $\mathbb{R}^l$ on defineeritud kui

$$d(\mathbf{x},\mathbf{z}) = \sqrt{\sum_{i=1}^{l}(x_i - z_i)^2}. \tag{3}$$

Seda kaugust saame kasutada naabruse määramisel arvuliste argumenttunnuste korral.

K-lähima naabri meetodi rakendamine koosneb peamiselt kahest etapist:

1) Leiame argumenttunnuste ruumis etteantud punktile $\mathbf{x} = (x_1, \ldots, x_l)$ kauguse d mõttes lähedaste poliiside hulga, sh sellise, et selle hulga kindlustuspäevade arv on (vähemalt) k. Nende poliiside indeksite hulga tähistame $U_{\mathbf{x}}$-ga.

2) Fikseeritud naabruse keskmise leidmine hinnanguna poliisile y:

$$\hat{y}(\mathbf{x}) = \frac{\sum_{i \in U_{\mathbf{x}}} y_i}{\sum_{i \in U_{\mathbf{x}}} t_i}, \tag{4}$$

kus t_i on i-nda poliisi kestus päevades. See summa on punktis 1 fikseeritud vektori $\mathbf{x}$ naabruses asuvate poliiside kestuste summa.

Meetodi rakendamiseks on vaja leida optimaalne k. Seda on hea teha ristvalideerimise abil, kus andmestik jagatakse osadeks. Alati jääb üks osa kasutamiseks testandmestikuna ning teised grupeeritakse treeningandmestikuks, mille pealt leitakse hinnangud, mis omakorda implementeeritakse testgrupile. Sedasi tuleb teha läbi kõik kombinatsioonid jagatud andmestiku osadega. Iga kombinatsiooniga leitakse hinnanguid mitme k korral leidmaks optimaalse k. Näiteks võib valida alustuseks $k = 50$ ning igal sammul leida uus k eelmise kahekordsena [4]. Maksimaalne k võiks olla selline, mille korral k elemendist koosnev naabrus on saavutanud koguvalimi suuruse. Iga k korral hinnatakse iga kombinatsiooni mudeli headust valitud veakarakteristiku abiga. Seejärel saab leida kõikide kombinatsioonide korral leitud vigade ruutkeskmise iga

k korral ning valida parimaks k-ks sellise, mille korral see ruutkeskmine on vähim.

Selles töös kasutatakse $k \cdot 365$ kombinatsiooni leidmaks lähimaid vaatluseid, sest naabreid valitakse reisiks kulunud päevade summa seast, mida vaadeldakse aastates. Olgu $T = \sum_{i=1}^{n} t_i$ reiside pikkuste kogusumma treeningandmestikus ning $k \cdot 365 \to T$. Et T on poliiside kestuste päevade kogusumma ümber teisendatuna üle 1 000 aasta, anname k korrutamisel 365-ga sellele sama mõõtme. Kuna $k \cdot 365$ kasvaks väga kiiresti kasutades eel mainitud k sammu suurendamise võimalust, valime antud töös k sammuks 50 ehk iga järgmine k on eelmisest 50 võrra suurem. Eelmises lõigus kirjeldatud algoritmi alusel leitaksegi see üks optimaalne k iga hinnatava riski jaoks.

Veahinnang

Kõikide mudelite, ka Swedbanki enda põhivalemi, headuse mõõtmiseks kasutatakse veakarakteristikut

$$e = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \cdot t_i}{\sum_{i=1}^{n} t_i}}, \tag{5}$$

kus n on vaatluste arv ja y_i on i-nda poliisi tegelik kahju ning $\hat{y}_i$ hinnang i-nda poliisi kahjule.

3 Swedbank P&C Insurance AS-i reisikindlustuse preemia valem

Selles peatükis on põgus kirjeldus Swedbank P&C Insurance'i enda preemia leidmiseks kasutatavast valemist. Põgus seepärast, et tegemist on konfidentsiaalsete andmetega. Küll aga mainitakse ära üldtuntud elemendid valemis, mida kasutatakse kindlustuspraktikas.

Valem koosneb 37. komponendist ehk riskist ning igaühele leitakse oma osalus preemiast, mille kokkuliitmisel saadakse kätte kogupreemia enne kululisa lisandumist. Arvutamisel võetakse arvesse kõik need andmed, mida internetipangas kindlustuslepingu sõlmimisel inimeselt küsitakse - vanus, reisi toimumise aeg, reisi sihtkoht ning võimalikud lisakaitsed. Neid tunnuseid kõiki korraga iga komponendi leidmisel ei rakendata, vaid vastavalt neil riskidel, mille kindlustusselts on heaks arvanud. Näiteks reisitõrke riskide korral, mis on detailselt valemis jaotatud lahti viieks, ei võeta arvesse kindlustusvõtja vanust. Samas, kui õnnetusjuhtumi ja haigestumise kaitsete korral on see tunnus üliolulin. Mis tunnuseid ja millal arvesse võtta, on üsna subjektiivne küsimus ning ilmselt kõige täpsemini hinnatav järjepidevale praktikale tuginedes. Seepärast selles töös analüüs läbi viiaksegi, et näha, kas peaks kusagil midagi teisiti kalkuleerima või vähemasti seda kaaluma.

Hinnangud riskidele leitakse kasutades keskmist kahju suurust ning kahju esinemise sagedust. See on üldtuntud meetod kindlustuspraktikas mitte ainult reisikindlustuses, vaid ka teiste kindlustusliikide puhul. Nende riskide puhul, kus on tegemist valitavate kindlustukaitse suurustega, leitakse riski keskmised sama kaitsega poliiside seast, mida klient endale soovib. Antud töös on see nüanss teadlikult kõrvale jäetud, et parendada tulemusi nende tunnuste abil, mis võiksid mängida rolli just kahju tekkimiseks.

4 Andmete kirjeldus

Preemia arvutamise valemi analüüsimiseks on Swedbank P&C Insurance AS andnud kasutada 89 108 poliisi andmed vahemikus 2010.-2014. aastal. Kõikide poliiside puhul on kahju andmed teada riskide kaupa, kui kahju on toimunud. Kahjujuhtumeid on kokku 2 601 sellel ajavahemikul. Preemiaid koguti 1 744 831.9 € eest ning kahjusid maksti välja 1 005 670.3 € eest. Sellest on näha, et kogu portfelli ulatuses preemia arvutamine toimib õigesti. Kindlustatuid esineb vanusevahemikus 0-100 ehk noorimad kindlustatud on alla aastased ja kõige vanem kindlustatu antud perioodil oli 100-aastane. Andmetabel sisaldab järgnevaid tunnuseid objektide kohta:

- vanus kindlustusperioodil,
- kindlustusperioodi algusaeg,
- kindlustusperioodi lõpuaeg,
- preemia kogusumma (€),
- preemia jaotatud riskide kaupa (€),
- kahju suurus riski tasemel (€),
- piirkond (0 - Euroopa, 1 - muu maailm),
- pakett (0 - standardne, 1 - suusareis),
- ohtlik tegevus (0 - ei, 1 - jah),
- tööreis (0 - ei, 1 - jah).

Tabelist 1 näeme poliiside jaotumist vastavalt reisi eesmärgile. Suurem osa portfellist moodustub standardse paketi ostjatest. Osadel neist võib olla ohtliku või tööreisi lisakaitse. Tabelis kahes esimeses veerus näidatud paketti ei ole kattuvad, kahes keskmises veerus olevad lisakaitsed võivad olla kombineeritud nii standardse paketi, suusareisiga kui ka omavahel. Reisi sihtkohta määratakse kahes grupis, mis on esitatud kahes viimases veerus Tabelis 1.

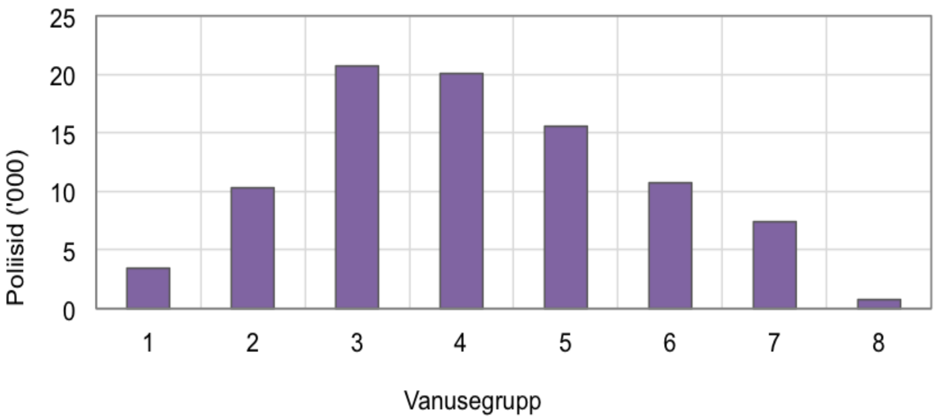
Standarne reis	Suusareis	Ohtlik reis	Tööreis	Euroopa	Muu maailm
85 979	3 129	4 512	1 351	65 647	24 461

Tabel 1: Poliiside jagunemine reisi eesmärgi ja piirkonna järgi

Preemia ja kahjude jaotumine vanusegrupi lõikes on näha Jooniselt 1. Jooniselt 2 nähtub portfelli jagunemine aastate lõikes. Täheldame iga-aastast kasvu poliiside arvus. Vanusegrupid moodustati järgnevalt:

Grupp	1	2	3	4	5	6	7	8
Vanus	0-5	6-18	19-29	30-39	40-49	50-59	60-74	75+

Tabel 2: Vanusegruppide jaotus



Joonis 1: Kindlustuspoliiside jagunemine vanusegruppide lõikes

Joonis 2: Kindlustuspoliiside jagunemine aastate lõikes

Viis kõige enam levinud kahjuliiki on:

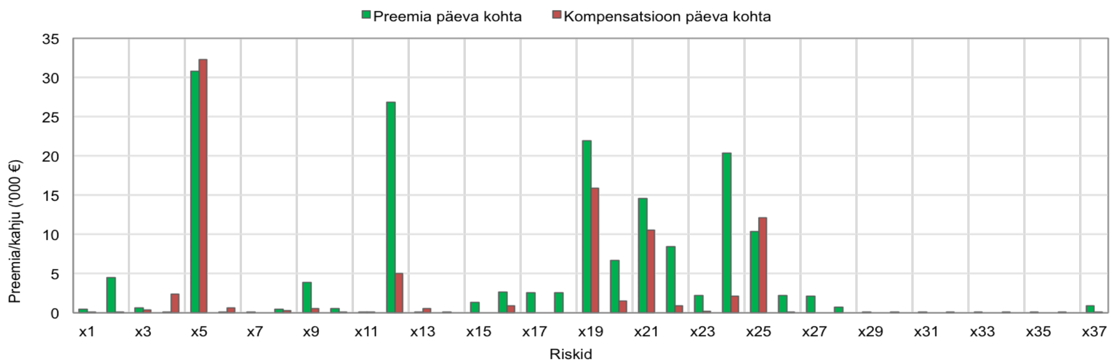
1. Haigestumine - ravikulud, vältimatu arstiabi;
2. Reisi kulgemine - ärajäämine;
3. Pagas - kadumine, purunemine, vargus;
4. Reisi kulgemine - transpordivahendile hilinemine;
5. Õnnetusjuhtum - ravikulud, vältimatu arstiabi.

Antud magistritöös aga ei keskenduta nendele riskidele, mis kõige sagedamini juhtuvad, vaid neile, mille preemia ning kahjumakse suurus on tasakaalust väljas. Selle all mõeldakse juhtu, kus riski puhul toimub suures osas preemia ülekogumine või kahjumaksed osutuvad suuremaks kui kogutud preemia. Teada on, et Swedbank P&C Insurance AS jälgib preemiate-kahjude vahekorda kogu portfelli ulatuses mitte riskide kaupa. Alljärgnev Joonis 3 ilmestab seda mõtet. On näha, kuidas näiteks riskid $x5$ ja $x25$ ei ole preemiatega kaetud. Samas $x12$ ja $x24$ on väga üle kogutud. Ka $x4$, $x6$ ja $x13$ puhul on näha, et pisut on kahjude väljamakseks kulunud rohkem kui preemiat saadud on.

Selles töös on teadlikult kõrvale jäetud väiksemas ulatuses preemiat üle kogunud riskid, sest autori arvates võiks säilida äriline aspekt teenida kasumit ning mitte preemiaid ja kahjusid analüüsi käigus täpselt võrdsustada.

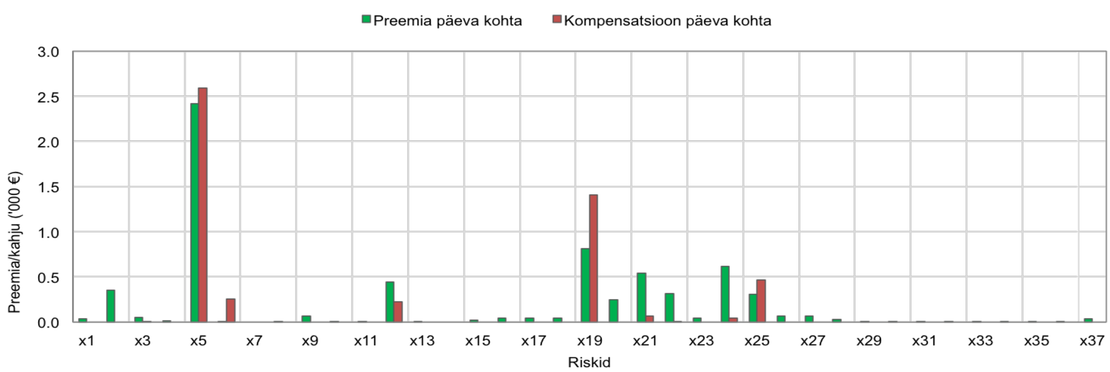
Et mõista riske ja põhianalüüsi paremini, esitame põgusa loetelu, mis kaitsega mingi risk seotud on. Nagu peatükis 3 mainitud, on kõik kaitsed kahju võimalike esinemisviiside kaudu omakorda peensusteni laiali jaotatud. Riskid jagunevad:

- $x1$ - $x7$ on seotud haigestumisega,
- $x8$ - $x18$ on seotud õnnetusjuhtumiga,
- $x19$ - $x23$ on seotud reisi kulgemisega,
- $x24$ - $x25$ on seotud pagasiga,
- $x26$ ja $x27$ on vastavalt vastutus- ja õigusabi,
- $x28$ - $x37$ on seotud terrorismiga.

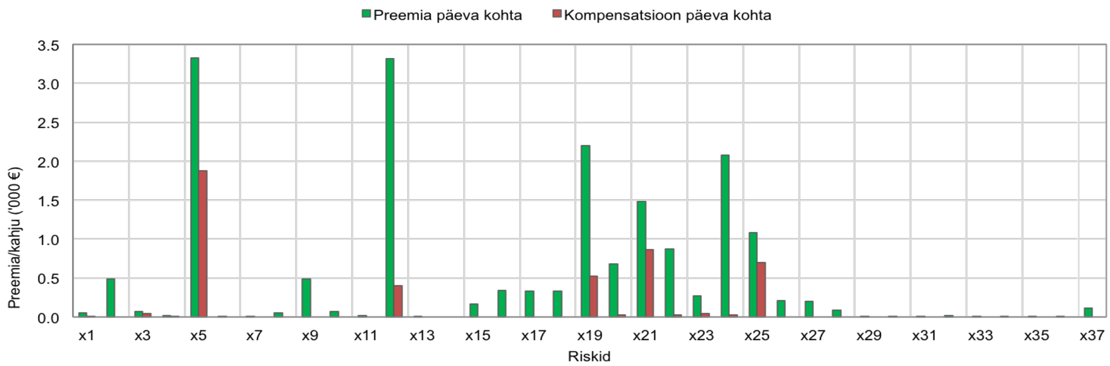


Joonis 3: Preemia ja kompensatsioon riskide kaupa

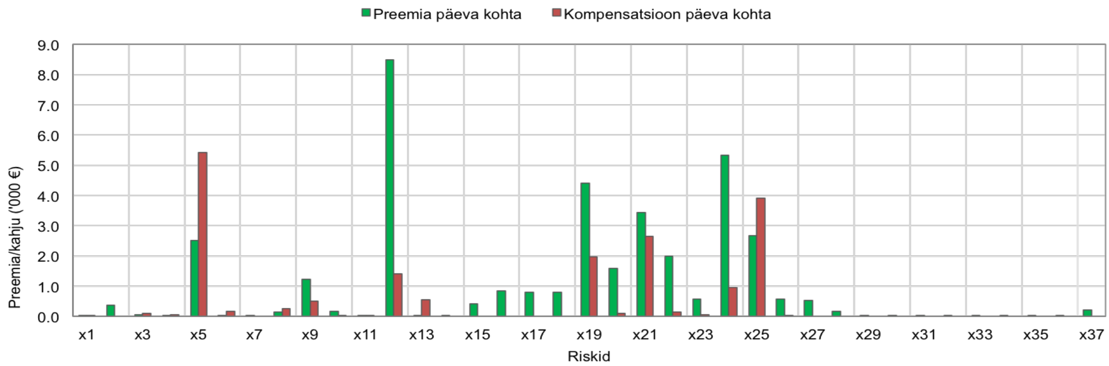
Vaatleme riskide kahjude ja preemiate jagunemist ka vanusegruppide kaupa. See aitab paremini aru saada hetkel kasutuses oleva arvutusvalemi tööst. Nii nagu Joonisel 3, paistab ka alljärgnevalt kaheksalt jooniselt (Joonis 4 - Joonis 11) silma riski $x5$ ebastabiilsus kahjude ja preemiate jagunemise vahel. Ainult Joonisel 8 esineb selline pilt, mida ootaks nii igas vanuserühmas kui ka üle portfelli. Veidikene preemia ülekaal kahjudega võrreldes on oodatav ja vajalik, et teenida kasumit selle riski pealt. Samas, meile huvi pakkuv risk $x12$ on peaaegu igas vanusegrupis tugevalt üle kogutud. Pigem ootaks sellist pilti, mis esineb Joonisel 9. Riskide $x24$ ja $x25$ esineb sarnane tasakaalutus. Kusjuures, riski $x25$ preemia ei sõltu vanusest, kuid preemiad-kahjud esinevad vanusegrupiti erinevalt. Siin muidugi mängib rolli asjaolu, et Joonis 6 ja 7 esindavad vanuserühmi, mis reisivad selle kindlustusseltsi klientide seast kõige sagedamini. Seega nendest tulenevad kahjud nii arvult kui suuruselt moodustavad suurema osa portfelli kahjudest kui teised rühmad. Neid jooniseid vaadates tundub, et sellel riskil võiks esineda seos vanusega.



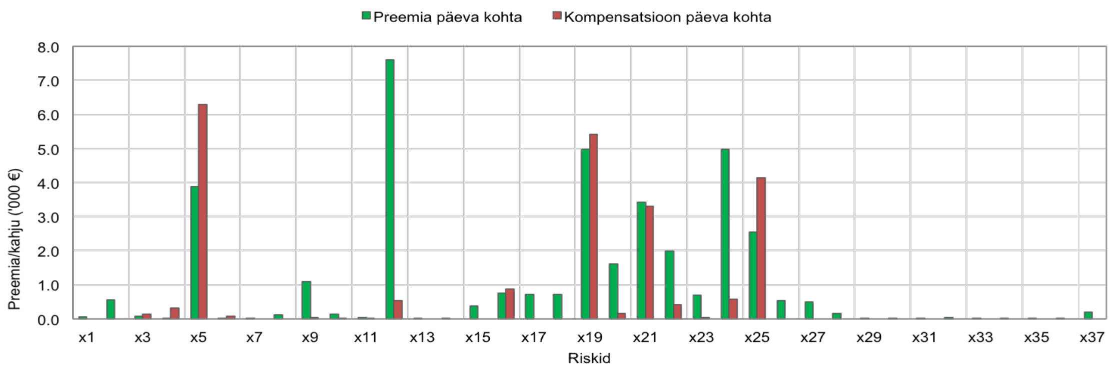
Joonis 4: Preemia ja kompensatsioon riskide kaupa vanusevahemikus 0-5



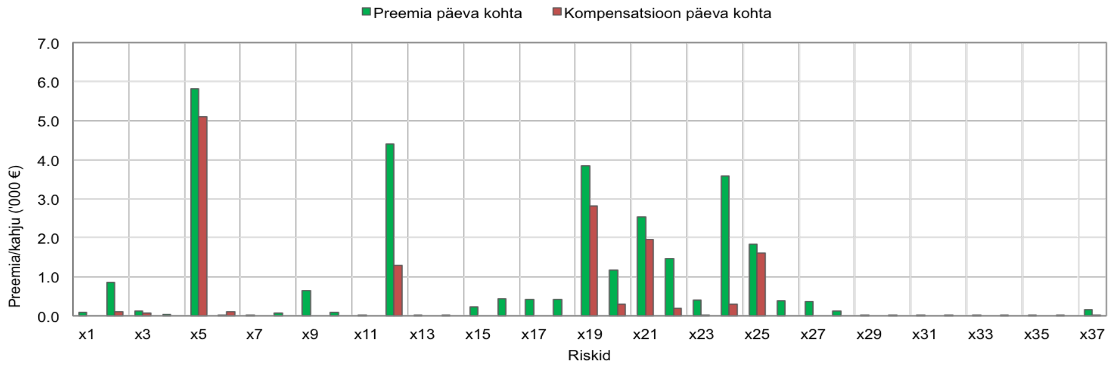
Joonis 5: Preemia ja kompensatsioon riskide kaupa vanusevahemikus 6-18



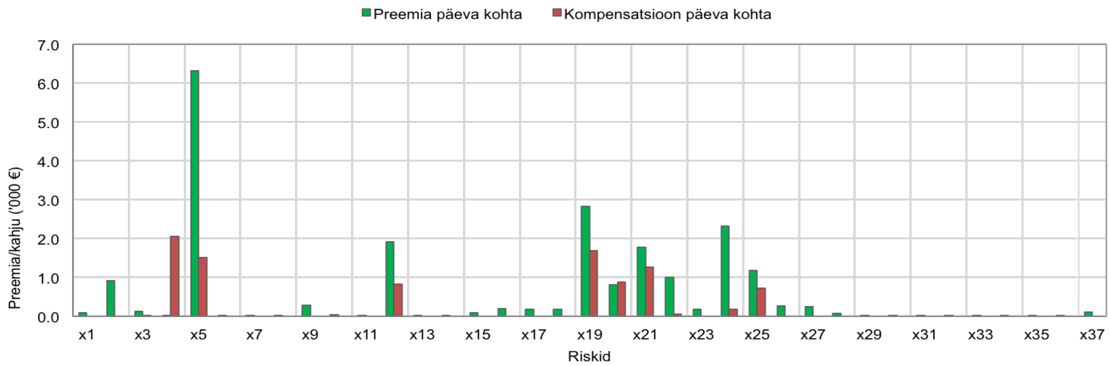
Joonis 6: Preemia ja kompensatsioon riskide kaupa vanusevahemikus 19-29



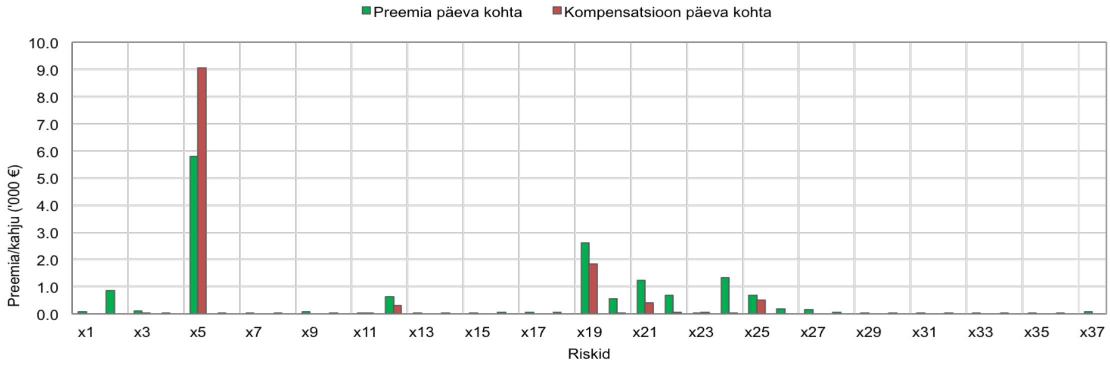
Joonis 7: Preemia ja kompensatsioon riskide kaupa vanusevahemikus 30-39



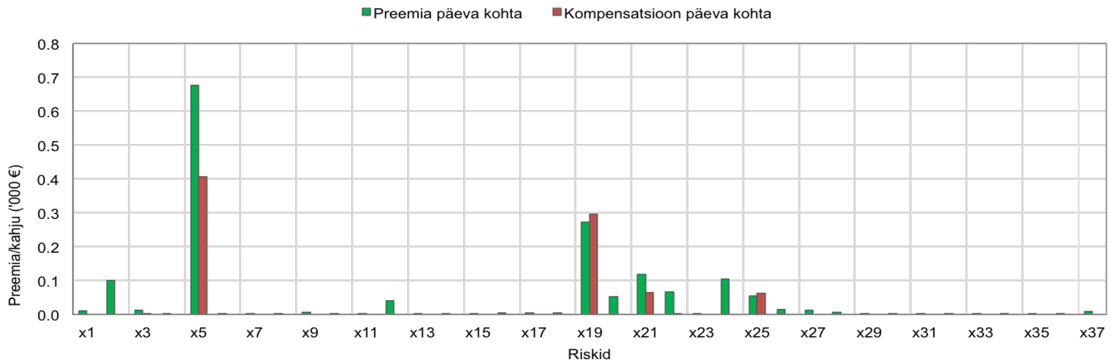
Joonis 8: Preemia ja kompensatsioon riskide kaupa vanusevahemikus 40-49



Joonis 9: Preemia ja kompensatsioon riskide kaupa vanusevahemikus 50-59



Joonis 10: Preemia ja kompensatsioon riskide kaupa vanusevahemikus 60-74

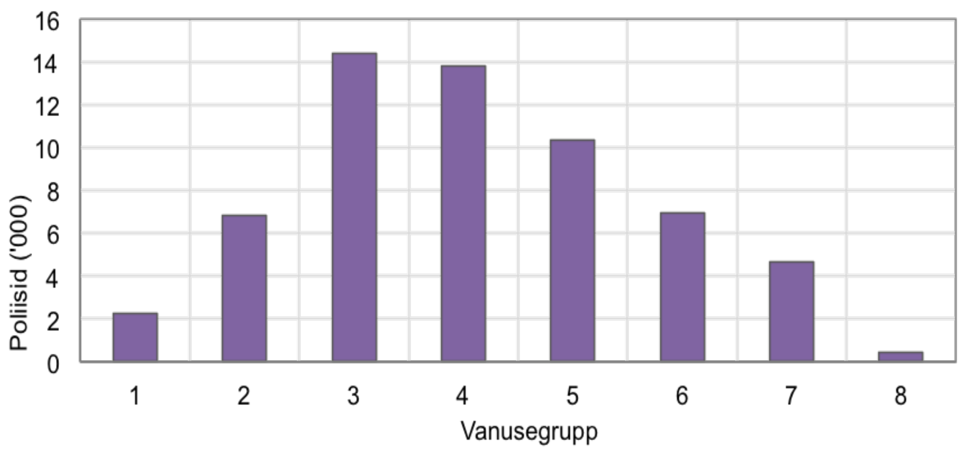


Joonis 11: Preemia ja kompensatsioon riskide kaupa vanuses 75+

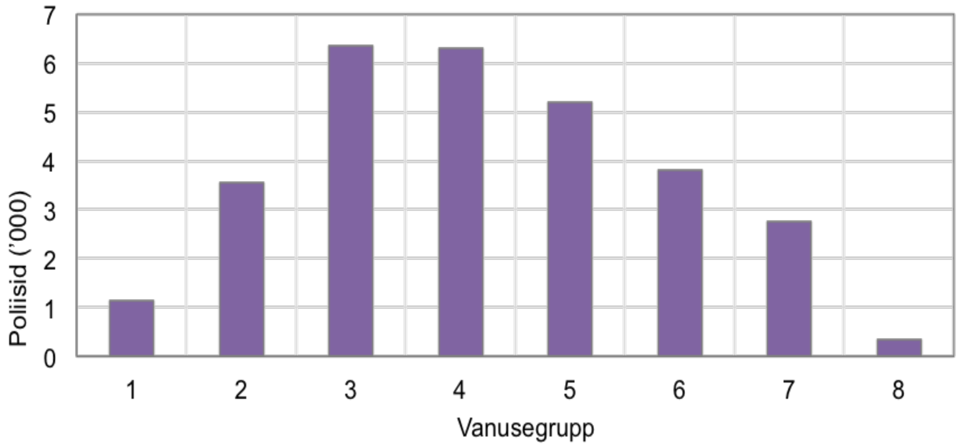
5 Andmete analüüs

5.1 Lineaarse mudeli rakendamine

Ühe võrdleva mudelina rakendatakse andmetele lineaarset mudelit meetodit. Lisaks Swedbank P&C Insuranse AS-i enda valemiga ning k-lähima naabri meetodiga võrdlemiseks, tehakse need mudelid seepärast, et leida olulisi argumente, mida rakendada k-lähima naabri meetodis järgmises peatükis. Kasutamiseks mõeldud andmed jaotatakse treening- ja testandmestikuks. Mudelid koostatakse treeningandmeid ning prognoosid testandmeid kasutades. Esimene andmekogum koosneb poliisidest ja kahjudest täisaastatel 2010-2013, mis on selles töös ajaloolisteks andmeteks, ning teine täisaastast 2014. Vastavalt on neis poliise 59 667 ning 29 441, millest kahjusid on toimunud vastavalt 1 787 ja 814. Vaatlused otsustati jagada kaheks sel viisil seepärast, et ka kindlustusselts ise hindab tänase poliisi preemia hinna eelnenud kahjude pealt. Reisi pikkuste kogusumma päevades treeningandmestikus on 719 379 ja testandmestikus 333 965.



Joonis 12: Poliiside jagunemine treeningandmestikus vanusegruppide kaupa



Joonis 13: Poliiside jagunemine testandmestikus vanusegruppide kaupa

Lineaarsed mudelid tehakse eelmises peatükis mainitud kriitilistele riskidele ajalooliste andmete järgi. See tähendab, et proovitakse leida oluline lineaarne mudel järgmistele riskidele: $x4$, $x5$, $x6$, $x12$, $x13$, $x24$ ja $x25$. Valime need riskid, sest eelneva arutluse tulemusena olid need problemaatilised. Jooniselt 3 on näha, et riskid ja kahjud ei ole tasakaalus. Mudelite puhul on sõltuvaks tunnuseks iga loetletud riski kahju päeva kohta. Argumenttunnusteks on pidevad tunnused vanus ja reisi pikkus ning 0-1-tunnused reisi toimumise piirkond, ohtlik reis, tööreis, reisi pakett. Intuitiivselt tundub ja ka varasemast kindlustuse praktikast on teada, et kahjude toimumisega vanus ei ole seotud lineaarselt. Seepärast on mudelisse lisatud veel ka vanuse ruut.

Enne põhianalüüsi juurde asumist esitame informatiivse tabeli iga huvi pakkuva riski preemia-kahjude seisu nii treening- kui ka testandmestikus. Preemia ja kahjude summa on eurodes ning kindlustusperioodid päevades.

Risk	Kindlustus-periood	Preemiad kokku	Kahjude arv	Kahjusumma kokku	Kindlustusperiood kahjudega poliiside hulgas
$x4$	719 379	916.56	13	40 704.90	351
$x5$	719 379	214 157.60	501	334 632.61	16 785
$x6$	719 379	629.86	6	2 826.19	206
$x12$	719 379	213 298.32	82	41 264.07	2 415
$x13$	719 379	6.68	1	1 867.25	21
$x24$	176 581	51 567.72	100	13 205.43	1 616
$x25$	176 581	31 015.13	375	81 682.36	5 596

Tabel 3: Riskid treeningandmestikus (2010-2013 a)

Risk	Kindlustus-periood	Preemiad kokku	Kahjude arv	Kahjusumma kokku	Kindlustusperiood kahjudega poliiside hulgas
$x4$	333 965	508.51	1	1 407.25	67
$x5$	333 965	119 113.94	222	6 8983.74	5 929
$x6$	333 965	349.64	5	4 631.65	65
$x12$	333 965	106 799.83	45	33 253.51	1 178
$x13$	333 965	3.32	1	2 772.27	6
$x24$	165 812	65 882.69	44	6 663.9	852
$x25$	165 812	38 224.18	193	40 417.27	2 355

Tabel 4: Riskid testandmestikus (2014 a)

Tabelitest 3 ja 4 näeme, et riskidele $x4$, $x6$ ja $x13$ ühtegi statistilist meetodit rakendada pole mõtet, sest kahjud praktiliselt puuduvad. Kuna need on väga harva esinevad kahjud, siis see seletab ka kindlustusseltsi enda raskuseid hinnata vastavaid riske optimaalselt. Seega jätkame analüüsi riskidega $x5$, $x12$, $x24$ ja $x25$.

Esitame ka korrelatsioonimaatriksi huvi pakkuvate riskide ja argumenttunnuste kohta. Lisaks eelmises lõigus mainitud riskidele näitame maatriksis ära ka autori arvates hästi hinnatud riskid nagu $x19$, $x20$ ja $x21$.

	$x5$	$x12$	$x19$	$x20$	$x21$	$x24$	$x25$
$x5$	1.0000	-0.0001	-0.0006	-0.0001	-0.0005	-0.0002	-0.0004
$x12$	-0.0001	1.0000	-0.0004	-0.0001	-0.0004	-0.0002	-0.0003
$x19$	-0.0006	-0.0004	1.0000	-0.0003	-0.0018	-0.0007	-0.0015
$x20$	-0.0001	-0.0001	-0.0003	1.0000	-0.0003	-0.0001	-0.0002
$x21$	-0.0005	-0.0004	-0.0018	-0.0003	1.0000	-0.0007	-0.0014
$x24$	-0.0002	-0.0002	-0.0007	-0.0001	-0.0007	1.0000	-0.0006
$x25$	-0.0004	-0.0003	-0.0015	-0.0002	-0.0014	-0.0006	1.0000
Vanus	0.0070	0.0008	0.0065	0.0043	0.0014	-0.0023	-0.0020
Reisi pikkus	0.0003	0.0024	-0.0047	-0.0010	-0.0032	-0.0018	-0.0033
Piirkond	0.0018	0.0067	0.0079	0.0000	0.0184	0.0036	0.0135
Ohtlik	0.0021	0.0011	0.0003	0.0003	0.0017	-0.0013	0.0047
Töö	-0.0011	-0.0012	-0.0055	-0.0008	0.0168	0.0002	-0.0030
Pakett	0.0010	0.0077	-0.0022	0.0228	0.0003	0.0008	0.0055

Tabel 5: Korrelatsioonimaatriks treeningandmetelt

	Vanus	Reisi pikkus	Piirkond	Ohtlik	Töö	Pakett
$x5$	0.0070	0.0003	0.0018	0.0021	-0.0011	0.0010
$x12$	0.0008	0.0024	0.0067	0.0011	-0.0012	0.0077
$x19$	0.0065	-0.0047	0.0079	0.0003	-0.0055	-0.0022
$x20$	0.0043	-0.0010	0.0000	0.0003	-0.0008	0.0228
$x21$	0.0014	-0.0032	0.0184	0.0017	0.0168	0.0003
$x24$	-0.0023	-0.0018	0.0036	-0.0013	0.0002	0.0008
$x25$	-0.0020	-0.0033	0.0135	0.0047	-0.0030	0.0055
Vanus	1.0000	-0.0290	0.0114	-0.0462	-0.0343	-0.0253
Reisi pikkus	-0.0290	1.0000	0.2905	-0.0009	0.2663	-0.0457
Piirkond	0.0114	0.2905	1.0000	0.0119	0.0722	-0.0960
Ohtlik	-0.0462	-0.0009	0.0119	1.0000	0.0360	0.1876
Töö	-0.0343	0.2663	0.0722	0.0360	1.0000	-0.0158
Pakett	-0.0253	-0.0457	-0.0960	0.1876	-0.0158	1.0000

Tabel 6: Korrelatsioonimaatriks treningandmetelt (jätk)

Tabelitest 5 ja 6 näeme, et riskidel praktiliselt puuduvad seosed argumenttunnustega. Korrelatsioonide p-väärtused on esitatud Lisas 2. Ka nende pealt on näha, et need ülinõrgad seosed on täiesti ebaolulised. See on teada tuntud probleem kindlustuspraktikas teiste liikide korral samuti, et kui tundub intuitiivselt, et mingi objekti puudutava tunnusega on kindlasti seos, siis tegelikkuses enamasti neid tunnuseid ei leidu. Aeg-ajalt kontrollitakse selles kindlustusseltsis, kas olukord on muutunud, kuid kui midagi leitakse, siis mõni üksik tunnus, millega on mingil riskil seos. See aitab omakorda selgitada senini kohati ebaproportsionaalselt kogutud kindlustusmakseid. Kuna aga kuidagi tuleb toote pakkumiseks riskid hinnata, siis vaatame, kas õnnestub edasise analüüsiga seoseid leida või prognoose parandada.

Kõikide alljärgnevate mudelite puhul alustame mudelist, kus kaasame argumenttunnustena vanuse, vanuse ruudu, reisimise piirkonna, reisipaketi, töö- ja ohtliku reisi ning reisi pikkuse päevades. Esialgsete mudelite parameetrite hinnangud on esitatud Lisas 1.

Risk $x5$

Mudelist eemaldatakse mitte olulised tunnused, need, millel $p > 0.05$. Oluliseks tunnuseks jäi ainult vanuse ruut. Tulemused on järgmised:

Tunnus	Parameetri hinnang	p-väärtus
Vabaliige	0.0883	0.69
Vanus^2	0.0002	0.04

Tabel 7: Riski $x5$ lõplik lineaarne mudel

Riski $x5$ jaoks on õnnestunud leida mudel olulisusega $p = 0.04$. Tuleb välja, et sellel riskil on seos vanusega, mida saab k-lähima naabri mudelis arvesse võtta.

Risk $x12$

AJärgnevas tabelis on riski $x12$ lõplik lineaarne mudel.

Tunnus	Parameetri hinnang	p-väärtus
Vabaliige	0.0348	0.05
Pakett	0.1839	0.06

Tabel 8: Riski $x12$ lõplik lineaarne mudel

Mudeli p-väärtus $= 0.06$. Kui võtta $\alpha = 0.1$, siis võiks saadud mudeli lugeda sobivaks. Tabelist 10 on näha, et $x12$ on ainuke risk, mille kahjudel on seos reisipaketiga. Võtamegi edaspidises analüüsis selle arvesse.

Risk $x24$

Nagu ka teiste riskide puhul, toimus siingi tunnuste eemaldamine ebaolulisuse järjekorras p mõttes. Lõplikuks loetud lineaarne mudel, kuhu viimase tunnusena jäi sisse vanuse ruut, on Tabelis 9.

Tunnus	Parameetri hinnang	p-väärtus
Vabaliige	0.0028	$8.62 \cdot 10^{-4}$
Vanus^2	$-4.09 \cdot 10^{-7}$	0.35

Tabel 9: Riski $x24$ lõplik lineaarne mudel

Tulemuseks on mudel, mis osutus ebaoluliseks p-väärtusega 0.35. Sellest hoolimata proovime sobitada k-lähima naabri meetodil saadud mudelit, kus argumendiks on vanus.

Risk $x25$

Tabelis 10 näeme lõplikku lineaarset mudelit riskile $x25$.

Tunnus	Parameetri hinnang	p-väärtus
Vabaliige	-0.0002	0.98
Vanus	0.0008	0.02
Vanus^2	$-1.16 \cdot 10^{-5}$	0.01
Piirkond	0.0114	$1.73 \cdot 10^{-3}$

Tabel 10: Riski $x25$ lõplik lineaarne mudel

Ka lõplik mudel on osutunud oluliseks, mille p-väärtus on $5.52 \cdot 10^{-4}$. Ühtlasi on $x25$ ainuke risk, mille korral on kahjul seosed kahe sõltumatu tunnusega. Järgmises osas võtame arvesse k-lähima naabri meetodi rakendamisel nii kindlustusvõtja vanuse kui reisimise piirkonna.

5.2 K-lähima naabri meetodi rakendamine

K-lähima naabri meetodi rakendamisel võetakse aluseks nii eelmises peatükis saadud tulemused kui kindlustusseltsi enda preemia arvutamise valem. Iga riski puhul on leitud argument, millega antud meetodit rakendada. Seega kasutatakse hindamiseks ühe argumendiga fikseeritud naabruses hindamist keskmise kahju kohta ühes päevas, kahe argumendiga sama meetodi rakendamise järele vajadus puudus selle loogika järgi, millega argumenttunnuseid valiti. Seega saame lihtsustada valemit (3) ning leida $d(\mathbf{x}, \mathbf{z})$ ruumi $\mathbb{R}$ jaoks. See taandub lihtsalt absoluutväärtuseks, mille võime kirja panna

$$d(x_1, z_1) = \sqrt{(x_1 - y_1)^2} = |x_1 - y_1|. \tag{6}$$

Jällegi kasutatakse treeningandmeid mudelite leidmiseks ning kahjude hinnanguteks rakendatakse mudeleid testandmestikule. Kõikide mudelite tulemusi võrreldakse järgmises, 5.3 alapeatükis.

Ülesande teostamiseks on vaja leida optimaalne k iga riski korral. Kui $365 \cdot k \geq N$, kus N on reisi kestuste summa päevades üle kõikide poliiside treeningandmestikus, siis on tegemist globaalse keskmisega ehk mudeli tegemisel võetakse arvesse kõik argumenttunnuse andmepunktid. Sel juhul oleks selle ülesande korral tegu keskmisega üle kõikide objektide. Optimaalse k leidmiseks jagatakse treeningandmestik omakorda kolmeks, seda puhku aga juhuslikult, kronoloogiat arvesse võtmata. Saadakse kolm ühesuurust andmestikku 39 778 poliisiga igaühes. Tähistame need vastavalt o_1, o_2 ja o_3. Teostatakse kolm optimaalse k leidmise protseduuri iga ette antud k korral. Osade o_1 ja o_2 pealt leitakse vähim viga e_1 prognooside, mis on leitud andmestiku osale o_3, ja tegelike andmete vahel. Seejärel rakendatakse treeningandmestikuna osasid o_2 ja o_3 ning prognoosid leitakse osale o_1. Kolmandaks võetakse treeningandmestikuks osad o_1 ja o_3 ning testandmestikuna kasutatakse esialgse treeningandmestiku osa o_2. Leitakse ka e_2 ja e_3 vastavalt.

Arvestades et, e_1, e_2 ja e_3 on ruutjuured keskmisest ruutveast (vt ptk 2.2), leitakse iga k korral karakteristik

$$e_k = \sqrt{\frac{e_1^2 + e_2^2 + e_3^2}{3}}. \tag{7}$$

Seejärel vaadatakse, millise k korral saadi kõige väiksem viga e_k ning vastav k valitaksegi antud riski parimaks, millega meetodit rakendada. Antud töös rakendatakse varianti kus alustatakse $k = 50$, ning iga järgmine suureneb 50 võrra eelmisest (vt ptk 2.2) ning $max(k) = 2000$, mille korral ületatakse päevade arvu summa N.

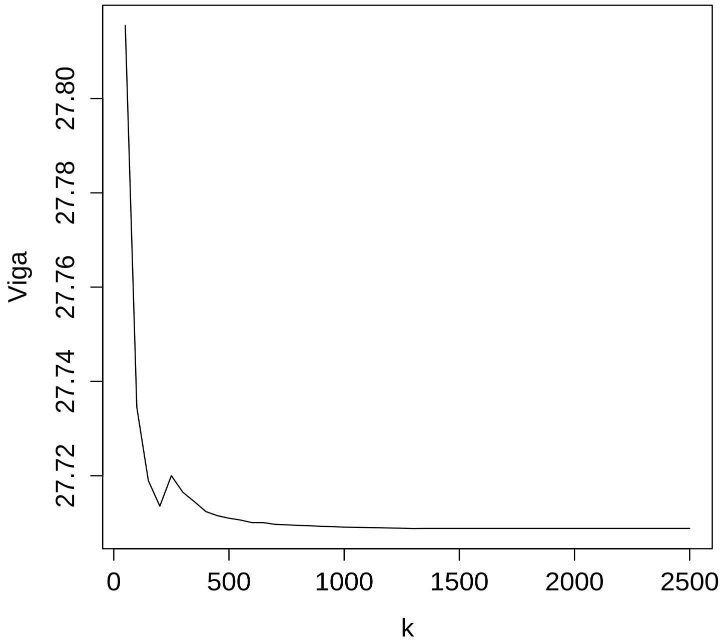
Risk $x5$

Riski $x5$ puhul on kaks võimalust, kuidas leida hinnangud selle riski kahjudele:

1. Lineaarse mudeli sobitamisel jäi viimaseks olulise tunnusena sisse vanuse ruut. Seega oleks üks võimalus rakendada fikseeritud keskmise leidmist ühe argumendiga, milleks on vanus.

2. Kuna põhivalemis sõltub see risk ka piirkonnast, kuhu reisitakse, võib jagada mõlemad andmestikud kaheks ning võtta argumenttunnusena arvesse vanuse.

Et selle juhtumi korral tekib kaks k-lähima naabri meetodit, mille vahel valida, leitakse lisaks vea hinnang nende kahe vahel valimiseks. Esmalt aga on vaja leida optimaalne k. Selleks proovime Joonise 14 abil leida k, mille korral viga e_k on minimaalne.



Joonis 14: Riski $x5$ vead e_k

Vähim viga leidub alates $k = 1350$ ning jääb püsima. Seega valime siin esimese sellise k ehk k suurusega 1350. Et teada, kumb üles loetletud k-lähima naabri meetodist paremini töötab, leiame e, kus võrreldakse treeningandmete saadud prognoose treeningandmestikus olevate tegelike kahjudega. Kumb annab vähima vea, see võrreldavaks meetodiks valitaksegi.

K-lähima naabri meetod	e
Üks valim	27.76681
Valim piirkonniti jaotatud	27.76661

Tabel 11: Riski $x5$ k-lähima naabri meetodite võrdlus

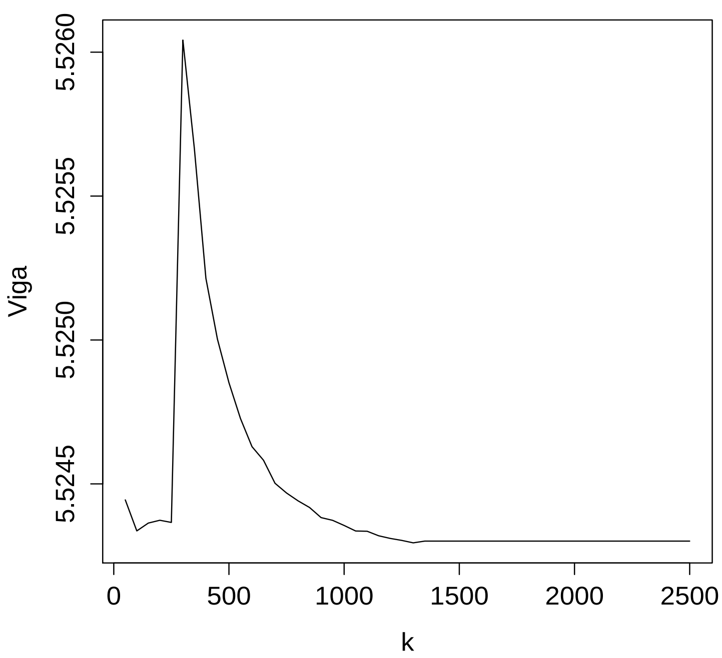
Osutub, et natuke väiksema vea annab meetodi rakendus siis, kui andmed on jaotatud kaheks piirkondade kaupa. Selle tulemused valime riski $x5$ k-lähima naabri meetodi hinnanguteks.

Risk $x12$

Riski $x12$ puhul on samuti kaks võimalust, mil viisil võiksime hinnangud leida:

1. Kuna lineaarse mudeli analüüsi tulemusena jäi peaaegu olulise tunnusena sisse reisipaketi binaarne tunnus, saaksime leida väga lihtsad prognoosid nii, et võtame paketi kaupa keskmised $x12$ kahjud ja omistame need vastavate pakettide kaupa testandmetele.
2. Kuna põhivalemis sõltub see risk ka vanusest, teostame k-lähima naabri meetodi analüüsi reisipaketiti, kus argumendiks on vanus.

Viimase variandi jaoks on vaja leida taas optimaalne k, millega ülesanne läbi viia. Selle leidmist ilmestab all olev joonis.

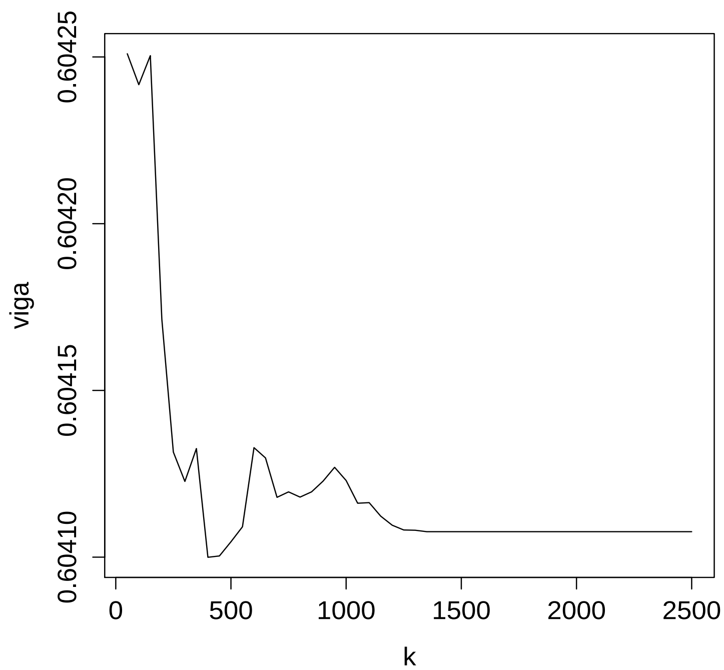


Joonis 15: Riski $x12$ vead e_k

Jooniselt 15 on näha, et vähim viga on vahemikus $1000 < k < 1500$. Lähemal tulemuste vaatlusel selgub, et see leidub, kui $k = 1300$. Valimegi selle, millega peamine analüüs teostada.

Risk $x24$

Osutub, et $x24$ ei sõltu põhivalemis ühestki tunnusest, mille oleme andmete kirjelduse osas analüüsis sõltumatuteks tunnusteks valinud. Seega teostame k-lähima naabri meetodil hinnangute leidmise lineaarse mudeli tulemuse põhjal ehk rakendame argumendina vanust treeningandmetelt hinnangute leidmiseks.



Joonis 16: Riski $x24$ vead e_k

Jooniselt 16 on näha, et minimaalne viga leiab aset enne $k = 500$. Lähemal vaatlusel selgub, et see on $k = 400$ korral. Samas, see on üsna väike piirkond arvestades treeningandmestiku reisi päevade kogusummat.

Vaatleme järgmist madalat lohku. Madalaim punkt seal vea poolest asub kohal $k = 700$. Seega arvame, et see on optimaalne k, millega analüüsis edasi minna.

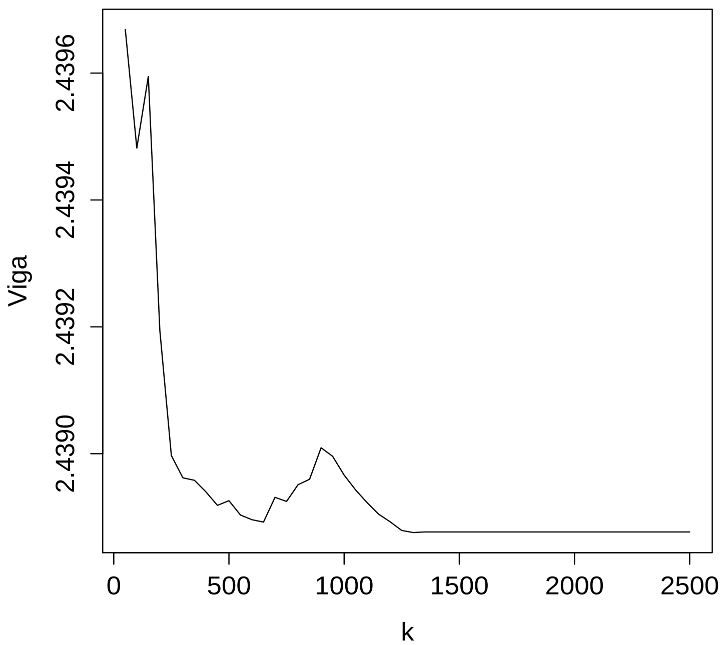
Risk $x25$

Nagu eelmise riski korral, ei sõltu põhivalemis ka $x25$ preemia ühestki meie valitud argumenttunnustest. Seega tekitame kaks k-lähima naabri meetodi varianti, mille vahel valime. Nendeks on:

1. Lineaarse mudeli tulemusena tulid olulisteks tunnusteks nii vanus kui vanuse ruut. Seega teostame k-lähima naabri meetodil analüüsi, kus argumendiks on vanus.

2. Lisaks vanusele osutus ka oluliseks seos piirkonna ja $x25$ kahjude vahel. Seega saame jagada treeningandmestiku kaheks ning teostada mõlemas osas punktis 1. üles märgitud analüüsi.

Nende kahe punkti täitmiseks on vaja leida optimaalne k.



Joonis 17: Riski $x25$ vead e_k

Ka Joonisel 17 huvitab meid kaks piirkonda. Esimeses lohus leidub vähim viga $k = 650$ korral. Otsustame, et see on optimaalne, sest järgmise minimaalse vea korral on tegemist juba üldkeskmisega.. Järgmiseks oleks vaja leida, kumb eel loetletud analüüsist töötab paremini. Seepärast viime läbi treeningandmetel mõlemad analüüsid leitud optimaalse k-ga. See, mis annab vähima vea, hindab kahjud $x25$ jaoks järgmise alapeatüki tarbeks.

K-lähima naabri meetod	e
Üks valim	2.4397
Valim piirkonniti jaotatud	2.4395

Tabel 12: Riski $x25$ k-lähima naabri meetodite võrdlus

Tabelist 14 on näha, et see meetod, kus jagame andmed piirkonna kaupa kaheks ning valime argumendiks vanuse, võiks olla täpsem kui leida hinnangud valimit laiali jagamata.

5.3 Tulemused

Selles alapeatükis võrdleme kolme erinevat mudelit peatükis 2.2 kirja pandud veakarakteristiku (vt valem 5) järgi. Meenutuseks - võrdleme omavahel Swedbank P&C Insurance AS-i valemit, lineaarse mudeli ning iga riski parimalt töötanud k-lähima naabri meetodi ning kahjude üldkeskmise tulemusi. Samuti on asjakohane teha portfelli võrdlus kahjude ja saadud kahjuhinnangutega.

Risk	Kasutatud preemia	Lineaarne mudel	K-lähima naabri meetod	Keskmine jaotatud valimist$*$	Aritmeetiline keskmine
$x5$	3.89299	3.90824	3.89245	-	**3.88989**
$x12$	5.38446	5.37754	**5.37736**	5.37756	5.37810
$x24$	0.79673	**0.75245**	0.75247	-	0.75247
$x25$	**2.45707**	2.46155	2.46128	-	2.460163

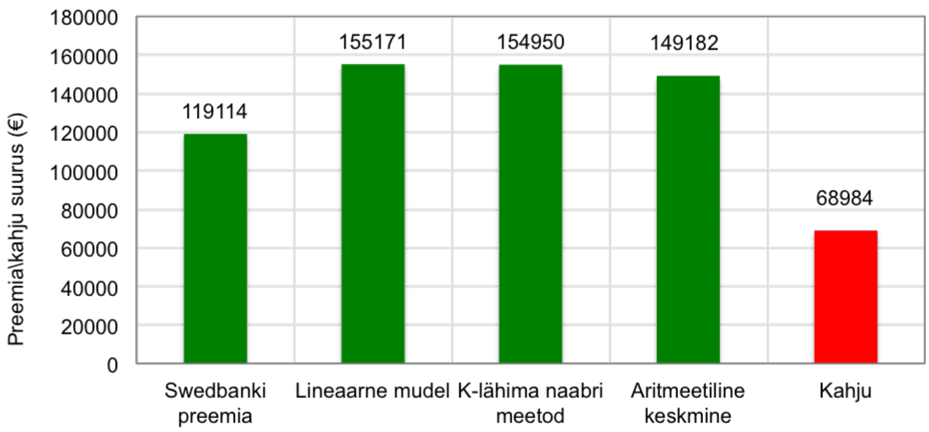
Tabel 13: Viga e meetodite kaupa.

$*$ Keskmine leitud riski $x12$ jaoks, kui valim on jaotatud paketi kaupa kaheks (ptk 5.2).

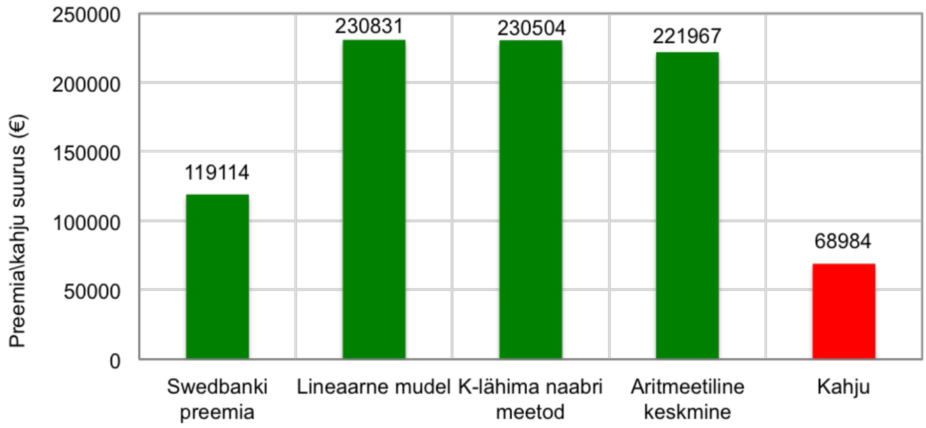
Tabeli 13 järgi on näha, milliste riskide puhul õnnestus saada selle töö tulemusena parem tulemus vea mõttes kui kindlustusseltsil endal. Teise veeru tulemus on viga tegeliku kasutatud preemia ja kahjude vahel. Rõhutatult on märgitud parima mudeli tulemus.

Näeme, et riski $x25$ puhul annab vähima vea prognooside ja tegelike kahjude vahel Swedbank P&C Insurance'i enda preemia. Aga nagu töös eespool mainitud on, ei olnud meil analüüsiks kasutada puhast preemiat, vaid kululisaga. Seega on veidi üllatav, et see meetod saavutas vähima vea. K-lähima naabri meetodiga õnnestus saada kõige parem tulemus riskile $x12$. Riski $x24$ parimaks meetodiks on osutunud lineaarne mudel, kuid meenutame, et saadud mudel ei tulnud oluline p-väärtuse mõttes (vt ptk 5.1) ning võiks kaaluda k-lähima naabri meetodi rakendust, mis annab väiksema vea kui põhivalem. Samas, paneme tähele, et k-lähima naabri meetodi ja aritmeetilise keskmise kaudu hindamise vahel vigade erinevus puudub. See tähendab, et k valikuga jõudsime juba globaalse keskmise leidmiseni. Seda kinnitab ka Tabelis 3 riski $x24$ kindlustusperioodi pikkus. Viimaseks, osutub, et riski $x5$ jaoks annab parima tulemuse vea mõttes aritmeetiline keskmine.

Oluline on lisaks veakarakteristikule vaadata testandmestiku hinnangute pealt ka, kuidas kujuneks portfell ühe või teise meetodi korral. Et Swedbanki valem ei oleks ebavõrdses seisus meie testitud meetoditega, kasutame nende reaalselt kasutatavaid marginaale kululisa arvestamisel. Konfidentsiaalsuse mõttes neid numbreid siin töös ei avaldata.

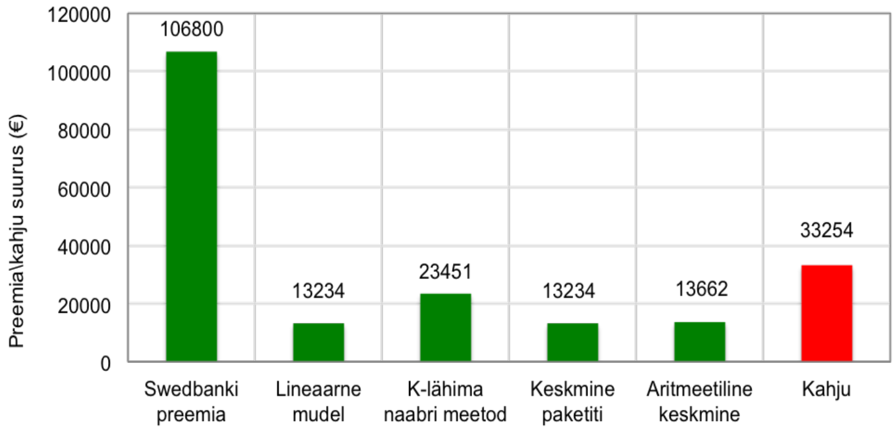


Joonis 18: Riski $x5$ tulemused

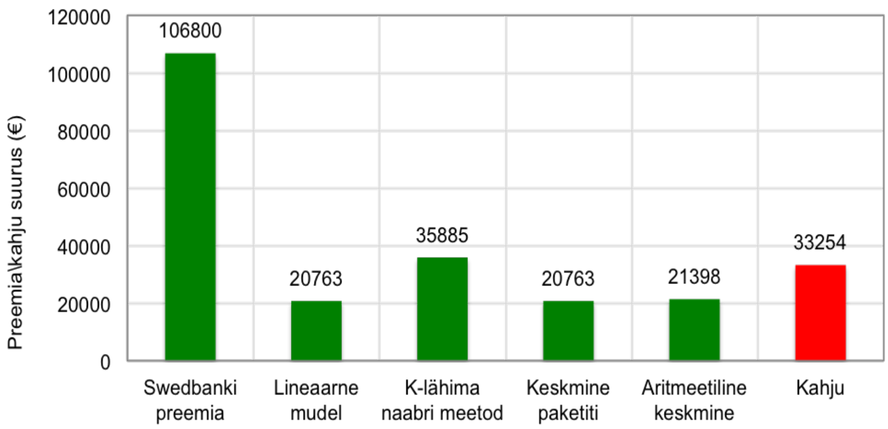


Joonis 19: Riski $x5$ tulemused kululisaga

Jooniselt 18 ja 19 näeme, kuidas kujuneks riski $x5$ kindlustusmakse vastu kahjusid erinevaid meetodeid kasutades. Meenutame, et Joonisel 3 oli kogu kahjusumma suurem kui preemia kogusumma. Siiski neilt jooniselt näib, nagu Swedbanki enda valem saaks siiski riskide õigesti hindamisega hakkama. Paraku on 2014. aastal esinenud kahjud väiksemate suurustega kui treening-projektis (vt Tabel 4), seepärast võib Jooniselt 18 ja 19 tunduda, et probleemi polnudki, mida uurida. Iseenesest, graafikuid vaadates tundub, et ei mängi rolli, millist kolmest katsetatud meetodist võiks rohkem kasu olla, kui teistest. Hoopis torkab silma uus mõttekoht, nimelt kululisa suurus. Kui võtta arvesse, mis suurusjärgus meie kolm testitud meetodit muutusid, võiks öelda, et Swedbanki mudel selle riski jaoks tänu kululisale vähemalt niigi kõrge tulemuse saavutab. Kui siiski oleks vaja mingi meetod valida, siis Tabelile 13 tuginedes, võib arvata, et tavalise aritmeetilise keskmine leidmine koguvalimilt hindab riske teistega võrreldes paremini.

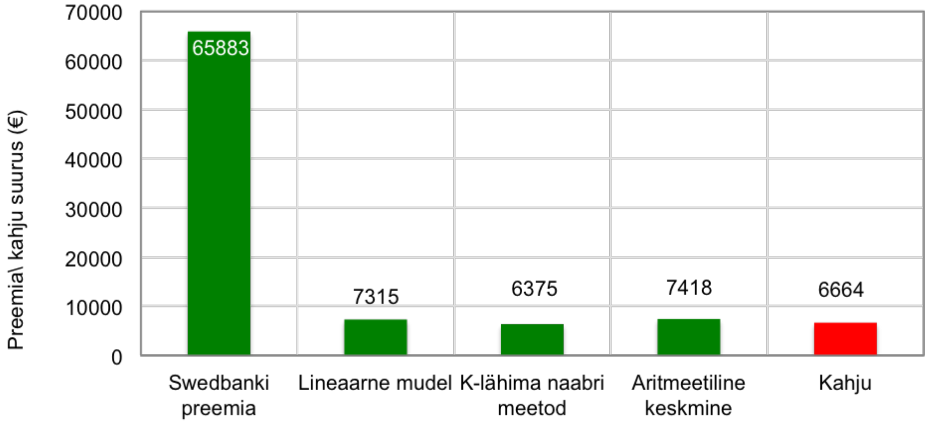


Joonis 20: Riski $x12$ tulemused

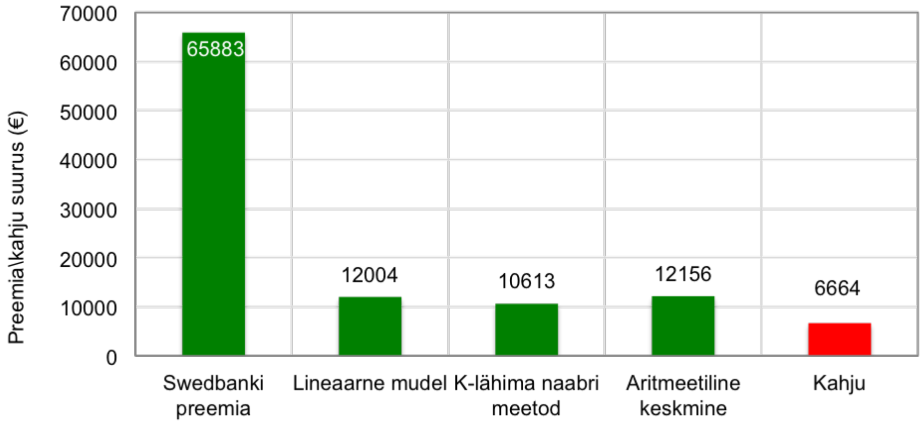


Joonis 21: Riski $x12$ tulemused kululisaga

Risk $x12$ oli see, mille preemiat Joonise 3 järgi soovisime alandada. Jooniselt 21 näeme, et kululisa arvesse võtmisel saavutasime k-lähima meetodiga parema tulemuse kui Swedbank P&C Insurance enda valemiga. Ka Tabeli 13 järgi oli riski $x12$ parimaks meetodiks vähima vea mõttes k-lähima naabri meetod. Seega õnnestus leida ehk paremat hinda pakkuv meetod.



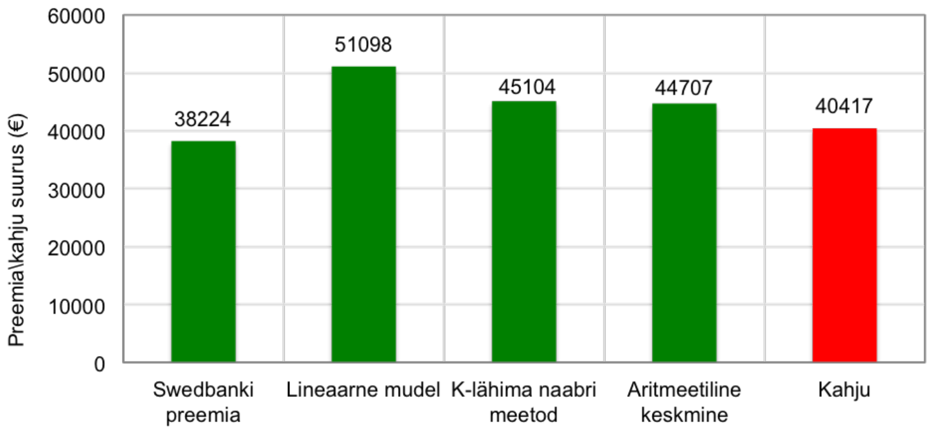
Joonis 22: Riski $x24$ tulemused



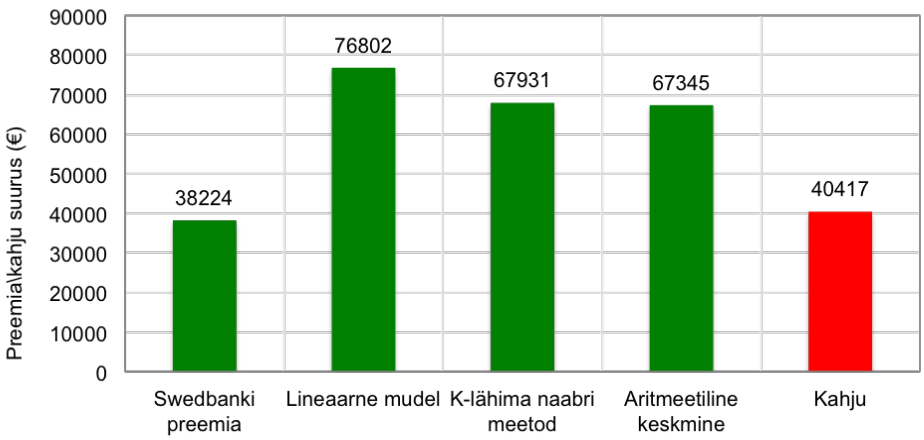
Joonis 23: Riski $x24$ tulemused kululisaga

Ka riski $x24$ kindlustusmakseid soovisime Joonisele 3 tuginedes alandada. Oleme saanud kõigi meetodiga kululisa kaasates Swedbanki enda arvutusvalemist üsna erinevad tulemused. On õnnestunud saavutada mitte nii ülehinnatud preemia, kuid mis katab siiski kahjumaksed. Varasema arutluse tulemusena lineaarse mudeli hinnangute juurde ei soovitaks jääda selle mudeli ebaolulisuse tõttu. Vea mõttes töötavad k-lähima naabri meetod ja aritmeetiline keskmine sarnaselt, samas portfelli kujundamisel annaks aritmeetilise keskmise kasutamine veidi rohkem tulu, hinda hüppeliselt lakke ajamata.

Seega, antud riski puhul oleks hea jääda aritmeetilise keskmise teel saadud hinnangute juurde.



Joonis 24: Riski $x25$ tulemused



Joonis 25: Riski $x25$ tulemused kululisaga

Joonised 24 ja 25 selgitavad, miks kõige parema hinnangu vea mõttes annab Swedbanki enda hinnangud. Isegi koos kululisaga on need kahjudele sarnased, aga seda mitte positiivses pooles. Seega mõlemale graafikule tuginedes võime öelda, et oleme saavutanud selle, mida Joonise 3 järgi soovisime ehk suurendada kindlustusmakseid selle riski korral. Vea järgi headuselt teine meetod on k-lähima naabri meetod. Kuna selle hinnangud annavad kahjudest kõrgema tulemuse, siis võib riski $x25$ korral sellega rahule jääda.

6 Kokkuvõte

Käesolevas magistritöös uuriti Swedbank P&C Insurance AS-i preemia kujunemist ning võeti eesmärgiks testida teisi statistilisi meetodeid parema kindlustusmakse saavutuse eesmärgil. Võrdlemiseks rakendati lineaarseid mudeleid ja k-lähima naabri meetodil fikseeritud keskmise leidmist. Kasutamiseks olid selle kindlustusseltsi poliiside andmed aastatel 2010-2014.

Analüüsi teostamiseks jagati andmed treening- ja testgruppi. Treeningrühma pealt leiti huvi pakkuvatele riskidele hinnangud ning testandmestiku abil leiti uued prognoosid. Hindamismeetodeid võrreldi veakarakteristiku (vt valem 5) järgi lisaks portfelli võimaliku kujunemise võrdlusele.

Põhianalüüs viidi läbi 37 komponendist Joonise 3 järgi maksete-kahjude võrdluse tulemusena välja valitud riskidega $x5$, $x12$, $x24$ ja $x25$. Komponentide $x5$ ja $x25$ puhul oli eesmärgiks kindlustusmakseid suuremaks ning $x12$ ja $x24$ makseid vähemaks saada. Portfelli võrdlemiseks suurendati saadud hinnanguid Swedbanki P&C Insurance'i igapäevases töös kasutatavate kululisadega, et portfelli võrdlemisel tagada eelduste poolest võrdsem olukord nende leitud preemia hinnangutele.

Lineaarseid mudeleid leiti peamiselt sel eesmärgil, et avastada seosed eel mainitud riskidel argumenttunnustega, mida rakendada k-lähima naabri meetodi korral. Oluline mudel töös valitud α järgi tuli riskidel $x5$ ja $x25$. Teiste puhul võeti arvesse, et mingi seos siiski võiks olla viimasena mudelisse jäänud tunnusega. Saadud olulisi või ebaolulisi tulemusi rakendati k-lähima naabri meetodi läbiviimisel.

Tulemuste arutelu põhjal leiti, et k-lähima naabri meetodiga saadi kõikide meetodite seast parim tulemus riskidele $x12$ ja $x25$. Teisele kahele komponendile, $x5$ ja $x24$, võiks sobida aritmeetilise keskmise leidmine. Kindlustuspraktika sage probleem on leida seoseid riskidel mingite kindlustusvõtjat puudutavate tunnustega ja seda mitte ainult reisikindlustuse, vaid ka teiste liikide puhul, nii et kindlustusmaksete optimaalne hindamine ongi ras-kendatud. Töö tulemusena leiti, et vähemalt nende nelja riski puhul on võimalik saada hinnangud paremaks ja valem täpsemini tööle. Samuti hinnaalandust vajavate riskidega saadi hüpe hinnas suurem kui hinna suurendamist nõudvate riskide hinnatõusuga. Seega saaks selles seltsis kehtivate kululisade marginaale säilitades pakkuda paremat hinda, seda enam, et mõndade riskide kahju aset leidmise korral on lisandunud reisikindlustuse tootele omavastutus, mis võiks hinnaalanduse korral tekkivat tulu vähenemist korvata mingis ulatuses.

Kirjandus

[1] Swedbank P&C Insurance AS-i reisikindlustuse tingimused https://www.swedbank.ee/static/pdf/private/insurance/travel/cond_travelins_2015_09_17_est.pdf.

[2] Zelterman, D, 2015. Applied Multivariate Statistics with R.

[3] Hastie, T, Tibshirani, R, Friedman, J, 2008. The Elements of Statistical Learning. Data Mining, Inference, and Prediction.

[4] Pärna, K, Kangro, R, Kaasik, A, Möls, M, 2011. K-nearest neighbors as pricing tool in insurance: a comparative study. *Multivariate Statistics Theory and Applications (Proceedings of IX Tartu Conference on Multivariate Statistics and XX International Workshop on Matrices and Statistics)*. 130-140.

Lisad

1. Lineaarsed mudelid

Tunnus	Parameetri hinnang	p-väärtus
Vabaliige	0.6305	0.26
Vanus	-0.0391	0.21
Vanus^2	0.0007	0.07
Reisi pikkus	0.0005	0.95
Piirkond	0.1787	0.60
Ohtlik reis	0.4385	0.53
Tööreis	-0.2254	0.85
Pakett	0.2467	0.76

Tabel 14: Riski $x5$ esialgne lineaarne mudel

Tunnus	Parameetri hinnang	p-väärtus
Vabaliige	0.0003	0.97
Vanus	$4.60 \cdot 10^{-5}$	0.90
Vanus^2	$-3.32 \cdot 10^{-7}$	0.95
Reisi pikkus	$2.76 \cdot 10^{-5}$	0.78
Piirkond	0.0071	0.92
Ohtlik reis	-0.0009	0.91
Tööreis	-0.0064	0.66
Pakett	0.0203	0.04

Tabel 15: Riski $x12$ esialgne lineaarne mudel

Tunnus	Parameetri hinnang	p-väärtus
Vabaliige	0.0005	0.81
Vanus	0.0001	0.23
Vanus^2	$-2.23 \cdot 10^{-6}$	0.16
Reisi pikkus	$-2.35 \cdot 10^{-5}$	0.45
Piirkond	0.0013	0.31
Ohtlik reis	-0.0013	0.63
Tööreis	0.0005	0.91
Pakett	0.0008	0.81

Tabel 16: Riski $x24$ esialgne lineaarne mudel

Tunnus	Parameetri hinnang	p-väärtus
Vabaliige	0.0009	0.89
Vanus	0.0008	0.03
Vanus^2	$-1.12 \cdot 10^{-5}$	0.02
Reisi pikkus	-0.0001	0.11
Piirkond	0.0139	$2.84 \cdot 10^{-4}$
Ohtlik reis	0.0052	0.50
Tööreis	-0.0091	0.50
Pakett	0.0121	0.19

Tabel 17: Riski $x25$ esialgne lineaarne mudel

2. Korrelatsioonimaatriksi p-väärtuste tabel

	$x5$	$x12$	$x19$	$x20$	$x21$	$x24$	$x25$
$x5$		0.9761	0.8907	0.9841	0.8979	0.9585	0.9145
$x12$	0.9761		0.9183	0.9882	0.9237	0.9690	0.9362
$x19$	0.8907	0.9183		0.9458	0.6604	0.8587	0.7134
$x20$	0.9841	0.9882	0.9458		0.9493	0.9795	0.9576
$x21$	0.8979	0.9237	0.6604	0.9493		0.8680	0.7316
$x24$	0.9585	0.9690	0.8587	0.9795	0.8680		0.8895
$x25$	0.9145	0.9362	0.7134	0.9576	0.7316	0.8895	
Vanus	0.0893	0.8535	0.1109	0.2987	0.7372	0.5745	0.6296
Reisi pikkus	0.9348	0.5523	0.2555	0.8085	0.4406	0.6562	0.4219
Piirkond	0.6563	0.1027	0.0523	0.9976	0.0000	0.3745	0.0010
Ohtlik	0.6093	0.7797	0.9351	0.9504	0.6797	0.7567	0.2506
Töö	0.7858	0.7738	0.1787	0.8450	0.0000	0.9621	0.4615
Pakett	0.8111	0.0600	0.5920	0.0000	0.9412	0.8540	0.1772

Tabel 18: Korrelatsioonide p-väärtused

	Vanus	Reisi pikkus	Piirkond	Ohtlik	Töö	Pakett
$x5$	0.0893	0.9348	0.6563	0.6093	0.7858	0.8111
$x12$	0.8535	0.5523	0.1027	0.7797	0.7738	0.0600
$x19$	0.1109	0.2555	0.0523	0.9351	0.1787	0.5920
$x20$	0.2987	0.8085	0.9976	0.9504	0.8450	0.0000
$x21$	0.7372	0.4406	0.0000	0.6797	0.0000	0.9412
$x24$	0.5745	0.6562	0.3745	0.7567	0.9621	0.8540
$x25$	0.6296	0.4219	0.0010	0.2506	0.4615	0.1772
Vanus		0.0000	0.0054	0.0000	0.0000	0.0000
Reisi pikkus	0.0000		0.0000	0.8184	0.0000	0.0000
Piirkond	0.0054	0.0000		0.0037	0.0000	0.0000
Ohtlik	0.0000	0.8184	0.0037		0.0000	0.0000
Töö	0.0000	0.0000	0.0000	0.0000		0.0001
Pakett	0.0000	0.0000	0.0000	0.0000	0.0001	

Tabel 19: Korrelatsioonide p-väärtused (jätk)

Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks

Mina, Nora Roosileht,

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) enda loodud teose

Swedbank P&C Insurance AS-ireisikindlustuse preemia analüüs,

mille juhendaja on dotsent Meelis Käärik,

1.1. reprodutseerimiseks säilitamise ja üldsusele kättesaadavaks tegemise eesmärgil, sealhulgas digitaalarhiivi DSpace-is lisamise eesmärgil kuni autoriõiguse kehtivuse tähtaja lõppemiseni;

1.2. üldsusele kättesaadavaks tegemiseks Tartu Ülikooli veebikeskkonna kaudu, sealhulgas digitaalarhiivi DSpace'i kaudu alates 12.05.2019 kuni autoriõiguse kehtivuse tähtaja lõppemiseni.

2. olen teadlik, et nimetatud õigused jäävad alles ka autorile.

3. kinnitan, et lihtlitsentsi andmisega ei rikuta teiste isikute intellektuaalomandi ega isikuandmete kaitse seadusest tulenevaid õigusi.

Tartus, 12.05.2016