Tartu Ülikool

Loodus- ja täppisteaduste valdkond

Matemaatika ja statistika instituut

Kadi Kitsing

Poliisi uuendustõenäosuse sõltuvus erinevatest faktoritest

Finants- ja kindlustusmatemaatika eriala

Magistritöö (30 EAP)

Juhendaja: vanemteadur Tõnu Kollo

Tartu 2020

Poliisi uuendustõenäosuse sõltuvus erinevatest faktoritest

Magistritöö

Kadi Kitsing

Lühikokkuvõte. Kindlustusseltside üks olulisemaid ülesandeid on klientide hoidmine. Klientide profiil on väga varieeruv ja klientide käitumine sõltub paljudest näitajatest. Seetõttu on tähtis aru saada, mida iga üksik uuendusklient ootab. Antud magistritöös uuritakse, millised faktorid mõjutavad liikluskindlustuse klientide uuendustõenäosust ja milline on erinevate kliendisegmentide šanss poliisi uuendamiseks. Selleks modelleeritakse kliendi uuendamise tõenäosust logistilise regressiooni mudeliga, mis kuulub üldistatud lineaarsete mudelite klassi.

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

Märksõnad: Kahjukindlustus, binaarne uuritav tunnus, logistiline regressioon, šansside suhe.

Policy’s renewal probability dependancy on different factors

Master’s thesis

Kadi Kitsing

Abstract. One of the most important aims for insurance companies is to try to retain customers. The profile of customers vary very much and behaviour of customers depends on several parameters. Therefore it is important to grasp what every single renewal customer expects. Aim of this thesis is to analyze which factors affect renewal probability of the third party liability insurance customers and what are odds ratios of different customer segments to renew a policy. For that customer renewal probability is modelled using logistic regression model, which is a special case of generalized linear models.

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

Keywords: Non-life insurance, binary response variable, logistic regression, odds ratio.

Sisukord

Sissejuhatus

Igal sõidukil on kohustus omada liikluskindlustust. Erinevalt kaskokindlustusest, mis pakub liiklusõnnetuse puhul kaitset enda sõiduki remondikuludeks, katab liikluskindlustus kannatanu sõiduki remondikulud ja nii kannatanu kui liiklusõnnetuse põhjustaja isikukahjude kulud. Liikluskindlustus on seaduse poolt reguleeritud ja kindlustusseltsid ei määra ise liikluskindlustuse tingimusi. Seetõttu liikluskindlustus kui toode on üle kindlustusturu ühesugune.

Tulenevalt sellest, et liikluskindlustus on reguleeritud, on klientide kindlustusseltsi valikul olulisemaks määrajaks liikluskindlustuse hind. Hinnapoliitikat toetab veel ka see, et turul on erinevad hinnakalkulaatorid, mille abil kliendid saavad võrrelda enda sõiduki poliisi hinda erinevates kindlustusseltsides. See teeb liikluskindlustuse kliendid väga tundlikuks ka väiksemagi hinnamuudatuse pärast. Koos liikluskindlustuse turu läbipaistvuse ja pideva dünaamilisusega muutub järjest olulisemaks vajadus hoida olemasolevat klienti.

Käesolevas töös analüüsitakse *If P&C Insurance AS* (edaspidi IF) liikluskindlustuse portfelli. Portfell jagatakse kaheks – uued kliendid ja uuenduskliendid. Uued kliendid on sellised kliendid, kes ostavad liikluskindlustust kindlustusseltsist IF esimest korda või kes ei kindlusta IF-is järjepidevalt. Uuenduskliendid on kliendid, kellel on juba kehtiv liikluskindlustuse poliis IF-is ja kes loodetavasti sõlmivad ka uue poliisi.

Antud magistritöö raames uuritakse, millistest faktoritest sõltub uuendusklientide soov jätkata samas kindlustusseltsis ehk uuritakse millest oleneb uuendustõenäosus. Liikluskindlustuse toote omapära tõttu tekkis vajadus uurida eelkõige hinnamuudatuse mõju uuendustõenäosusele.

Magistritöö on jagatud kolmeks osaks. Esimeses osas antakse ülevaade andmetest, et kirjeldada magistritöös uuritavat probleemi. Teises osas tutvustatakse matemaatilisi meetodeid, mida kasutatakse andmete analüüsimiseks. Töö viimases osas rakendatakse teooriat reaalsetel andmetel ja leitakse statistiline mudel kirjeldamaks klientide uuendustõenäosust. Magistritöö lisades on välja toodud lõpliku mudeli väljundid. Esimeses lisas esitatakse mudeli parameetrite hinnangud koos hinnangute standardvigadega ja statistiliselt olulisuse testi tulemustega. Teine lisa sisaldab seletavate tunnuste tasemetevahelisi šansside suhteid koos usaldusvahemikega.

1 Kirjeldav analüüs

Järgnev kirjeldav analüüs põhineb *If P&C Insurance AS* infosüsteemi andmetel.

Magistritöö eesmärgiks on liikluskindlustuse klientide uuenduskäitumise analüüsimine. Selleks kasutatakse IF-i liikluskindlustuse portfelli andmeid. Andmete homogeensuse huvides uuritakse ainult eraisiku tavasõiduauto liikluskindlustuse poliise. Valiti ainult otsekanali kliendid, välja jäeti maaklerkanali ja partnerkanali kliendid. Maaklerkanal ja partnerkanal on sellised kanalid, kus poliise müüakse IF-i eest ja IF maksab poliiside müümise eest vahendustasu (partnerkanali näide on autoesindused). Samuti jäeti analüüsist välja lühiajalised poliisid. Analüüsitakse kliente, kellel perioodil $01.03.2019 - 29.02.2020$ lõppes aastane poliis.

Olemasolevatest andmetest valiti välja 13 faktorit, mille korral soovitakse näha, kas ja kuidas need faktorid mõjutavad uuendustõenäosust. Tunnuste loetelu koos faktorite tasemete arvuga on toodud tabelis 1.1.

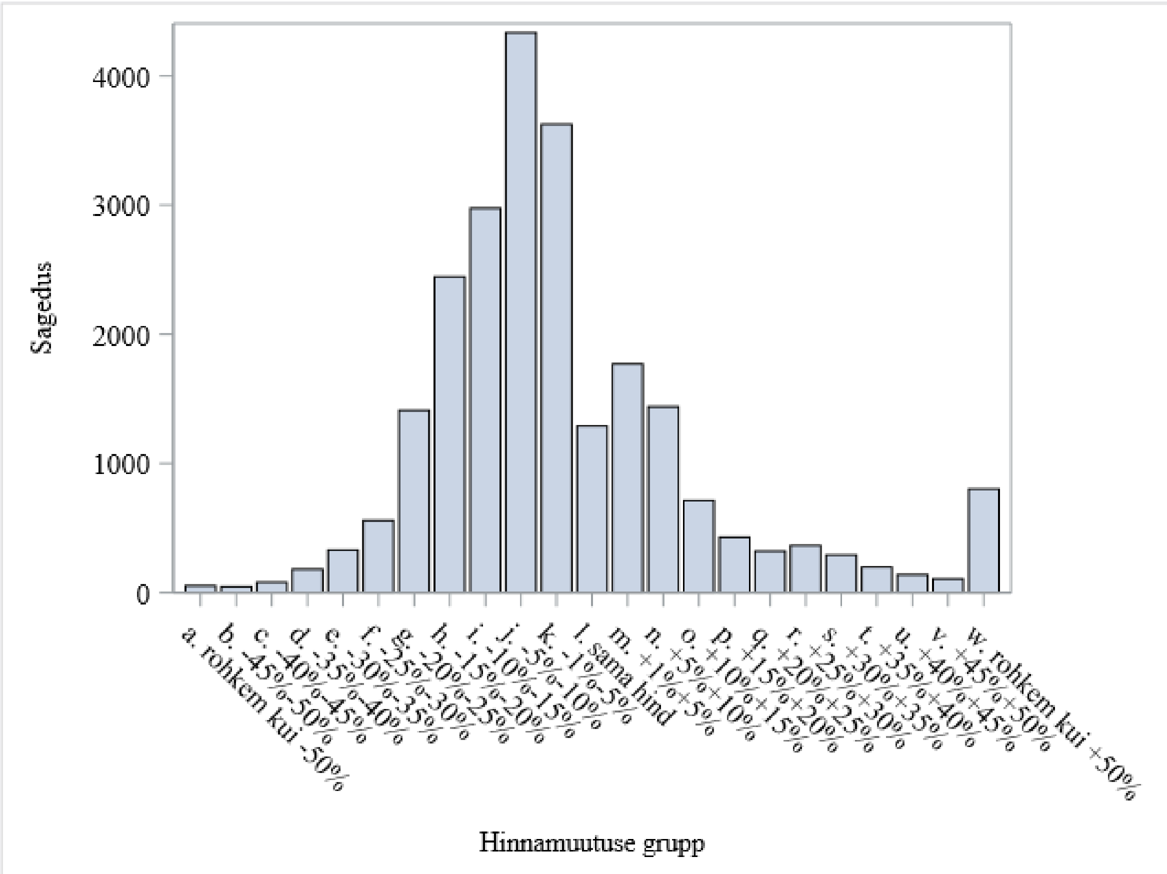
Tabel 1.1: Tunnuste loetelu.

Tunnus	Tunnuse nimi	Tasemete arv
UuenduseIndikaator	Uuenduse indikaator	2
HinnaMuutuseGrupp	Hinnamuutuse grupp	23
HinnaMuutuseGruppEurodes	Hinnamuutuse grupp eurodes	23
PreemiaGrupp	Preemia suuruse grupp	23
PoliisiKanal	Kanal	3
KliendiVanuseGrupp	Kliendi vanuse grupp	9
SõidukiVanuseGrupp	Sõiduki vanuse grupp	10
Regioon	Regioon	20
BoonusMaalusKlass	Boonus-maalus klass	15
KahjuOlemasolu	Kahju olemasolu	2
SõidukiteArvuGrupp	Kliendi sõidukite arv	4
UuendusteArv	Uuenduste arv	11
SõidukiKasutusala	Sõiduki kasutusala	3
SõidukiVõimsuseGrupp	Sõiduki võimsuse grupp	5

Järgnevalt antakse ülevaade analüüsitavatest tunnustest.

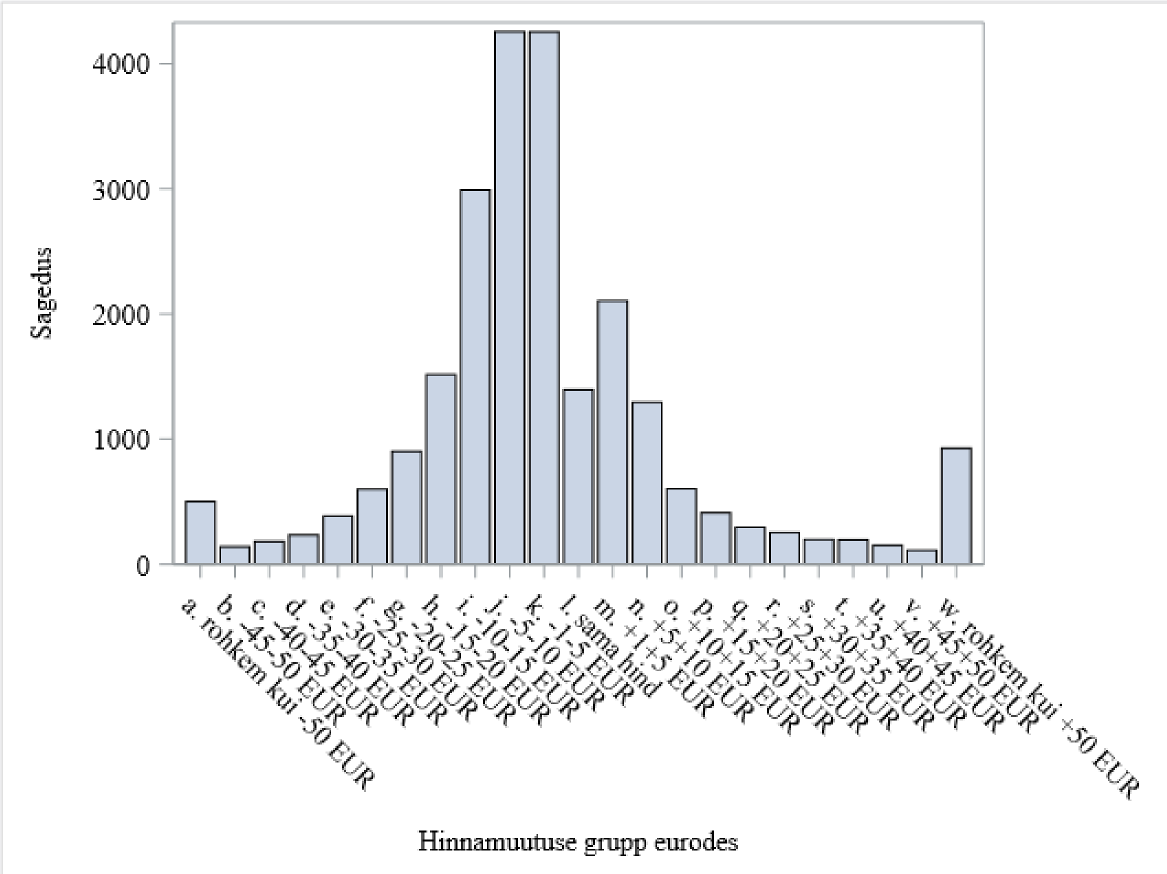
Uuritavaks tunnuseks on tunnus UuenduseIndikaator, millel on kaks taset 0 ja 1. Väärtusega 0 tähistatakse neid kliente, kes ei uuendanud enda poliisi IF-is ja väärtusega 1 neid kliente, kes jätkasid kindlustamist IF-is.

Huvipakkuvaim faktor on hinnamuutus. Hinnamuutus on arvutatud eelmise poliisi ja järgpoliisi või pakkumise hinnamuutusena. Öeldakse, et poliisi hind pole muutunud kui hinna suhteline muutus on kuni 1% (kaasa arvatud). Gruppides „k. -1%-5%", „m. +1%+5%" on kliendid, kelle poliisi hind muutus $1\% - 4\%$, gruppides „a. rohkem kui -50%", „w. rohkem kui +50%" on kliendid, kelle poliisi hind muutus rohkem kui 50%, ülejäänud poliisid on jagatud 5 protsendipunkti hinnamuutuse kaupa gruppidesse. Sageduse graafiku jooniselt 1.1 on näha, et antud andmestikus on enamus klientide poliise saanud soodsama hinna. Soovitakse uurida, kas hinnatõusu saanud kliendid lahkuvad IF-ist ja kas soodsama pakkumise saanud kliendid jäävad suurema tõenäosusega IF-i klientideks.



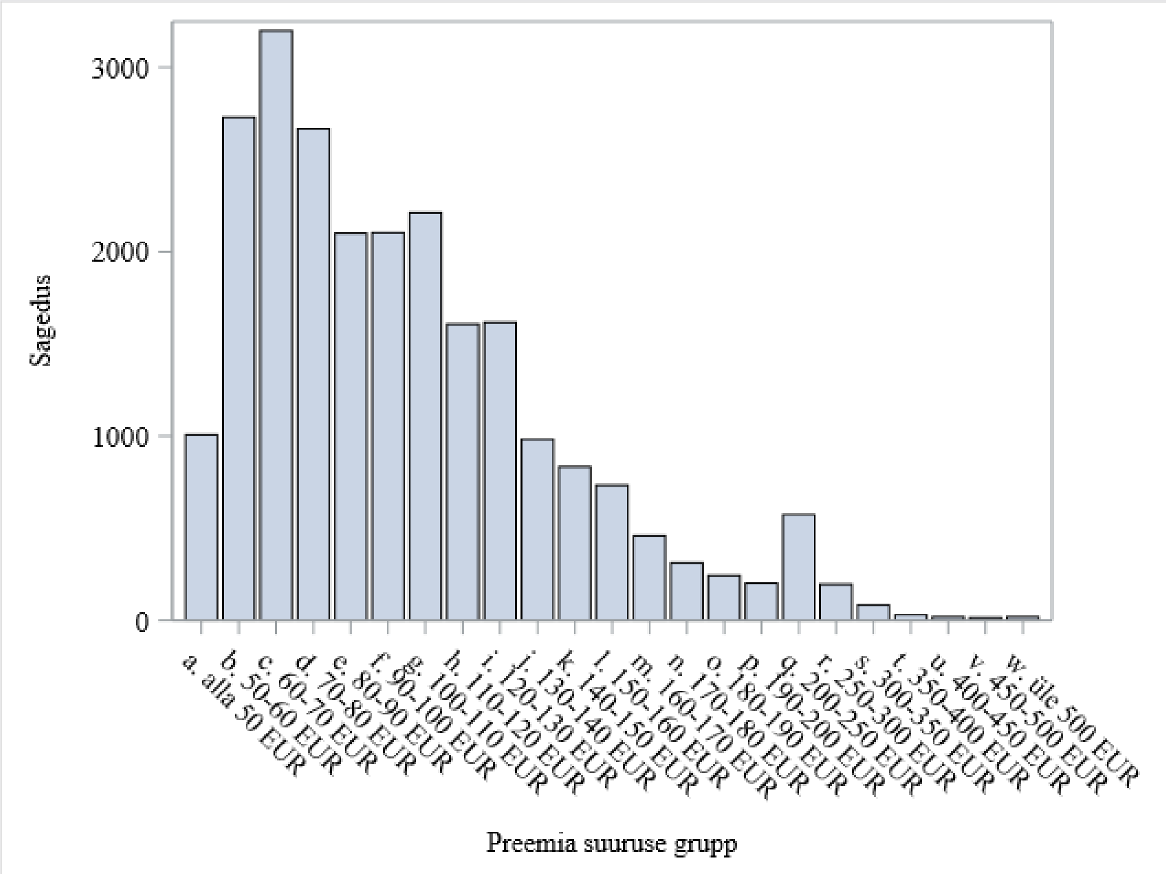
Joonis 1.1: Hinnamuutuse tasemed ja sagedused.

Uuritakse, kas kliendid on rohkem tundlikumad selle suhtes kui palju tõuseb poliisi hind eurodes või protsentuaalselt. Selleks valitakse üheks tunnuseks hinnamuutus eurodes (joonis 1.2). Arvutatakse poliisi hinna absoluutne muutus ja muutused jagatakse gruppidesse. Grupis „l. sama hind“ on kliendid, kelle poliisi hind muutus kuni 1 euro (kaasa arvatud), gruppides „k. -1-5 EUR“, „m. +1+5 EUR“ on kliendid, kelle poliisi hind muutus $1 - 4$ eurot, gruppides „a. rohkem kui -50 EUR“, „w. rohkem kui +50 EUR“ on kliendid, kelle poliisi hind muutus üle 50 euro, ülejäänud poliisid on jagatud 5 euro hinnamuutuse kaupa gruppidesse. Enamiku klientide uuenduspoliisi hind on kuni 10 eurot odavam kui eelmise poliisi hind. Samal ajal on palju kliente, kelle poliisi hind on tõusnud üle 50 euro.



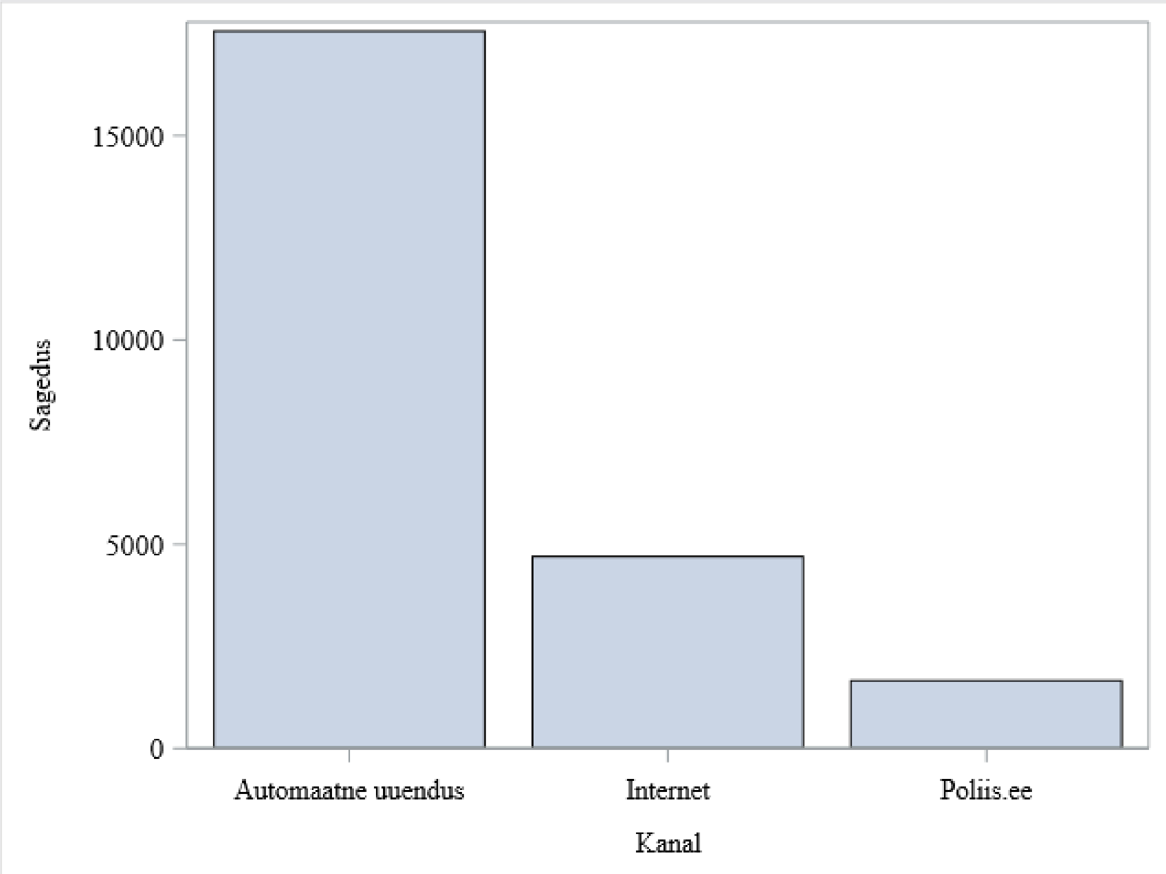
Joonis 1.2: Hinnamuutuse (eurodes) tasemed ja sagedused.

Veel pakub huvi, kas hüpotees, et need kliendid, kellel juba on kõrgem poliisi hind, on tundlikumad hinnamuudatuse suhtes, peab paika. Preemia suurust arvestatakse uuendatava poliisi hinna järgi ja preemiad jagatakse gruppidesse (joonis 1.3). Kliendid, kelle poliis maksab alla 50 euro või üle 500 euro on eraldi gruppides, preemiad suurustega $50 - 200$ jagatakse 10 euro kaupa gruppidesse, preemiad suurustega $200 - 500$ jagatakse 50 euro kaupa gruppidesse. Antud valimis on aastase poliisi keskmine hind 101 eurot.



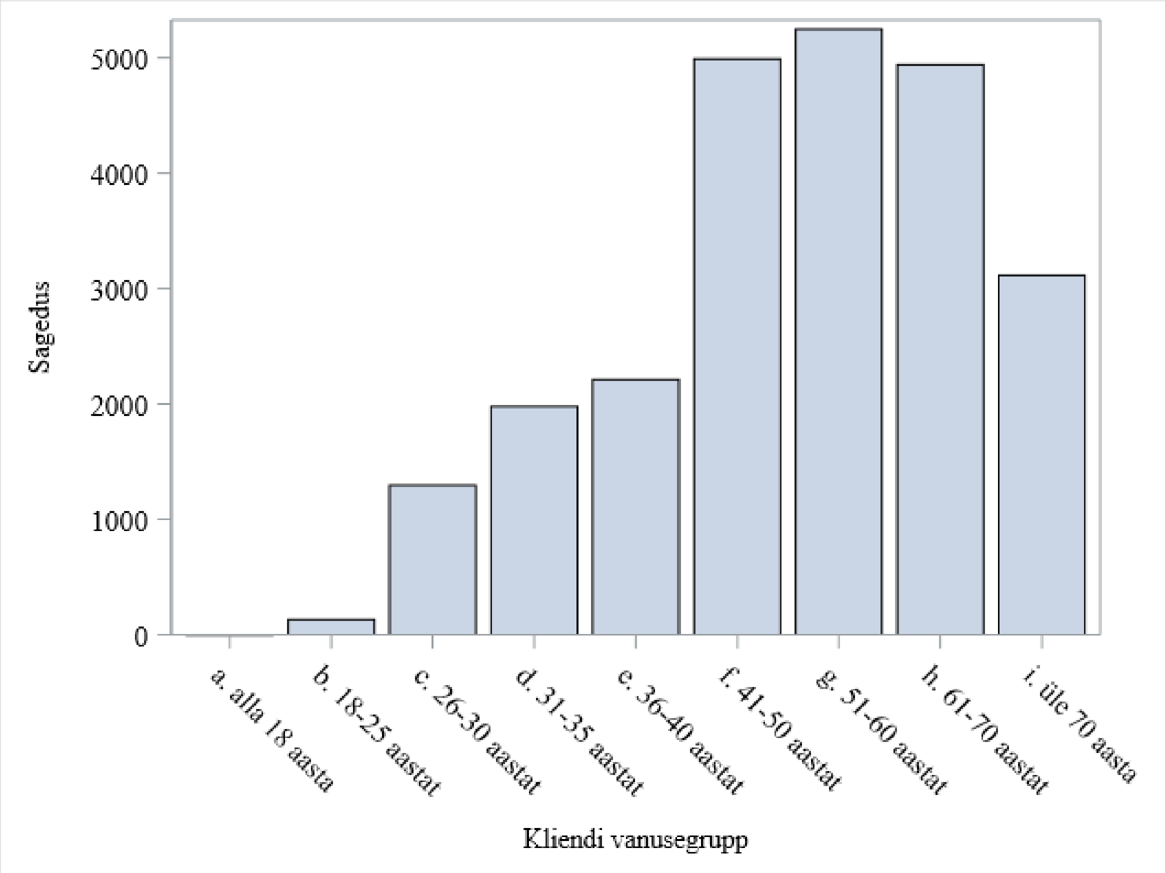
Joonis 1.3: Preemia suuruse tasemed ja sagedused.

Joonisel 1.4 kujutatakse kanalite jaotumist liikluskindlustuse portfellis. Faktor PoliisiKanal on eelmise poliisi ostmise koht. Automaatne uuendus on kanal, kus klientidele saadetakse automaatselt järgpoliisi pakkumine ilma, et klient peaks ise vaeva nägema. Arvatavasti internetikanalites (kanalid Internet ja Poliis.ee) on kliendid aktiivsemad teiste kindlustusandjate pakkumisi jälgima ja seega võib oletada, et nende klientide poliiside uuendustõenäosus on väiksem.



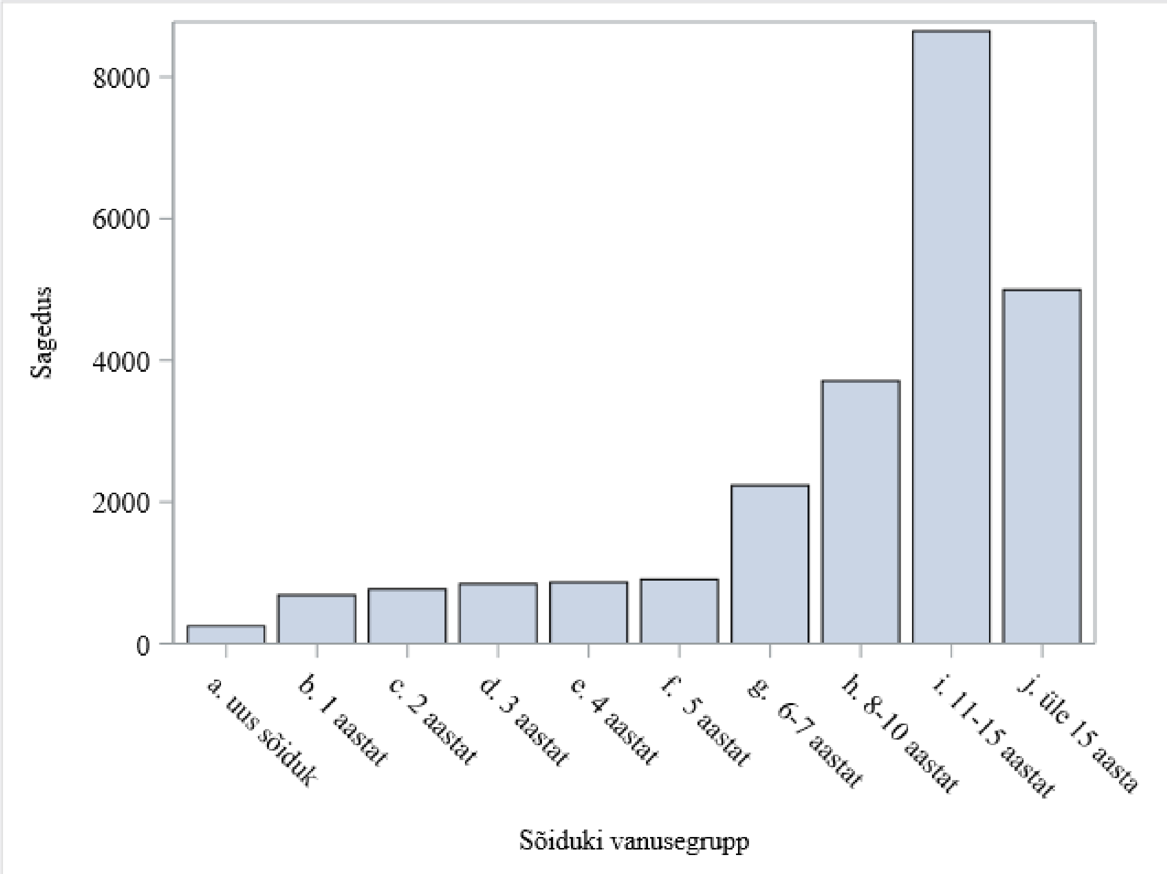
Joonis 1.4 Kanali tasemed ja sagedused.

Samuti soovitakse uurida kliendi käitumist sõltuvalt kliendi vanusest ja kliendi sõiduki vanusest. Kliendid on vanuse järgi jagatud gruppidesse „a. alla 18 aasta“, „b. 18-25 aastat“, „c. 26-30 aastat“, „d. 31-35 aastat“, „e. 36-40 aastat“, „f. 41-50 aastat“, „g. 51-60 aastat“, „h. 61-70 aastat“ ja „i. üle 70 aasta“. Jooniselt 1.5 on näha, et enamus liikluskindlustuse portfelli kliente on keskealised ja vaid väike osa on noori kliente. Magistritööga üritatakse selgusele jõuda, kas nooremad kliendid on tundlikumad hinnamuudatuse suhtes ja kas mõni vanusegrupp uuendab suurema tõenäosusega.



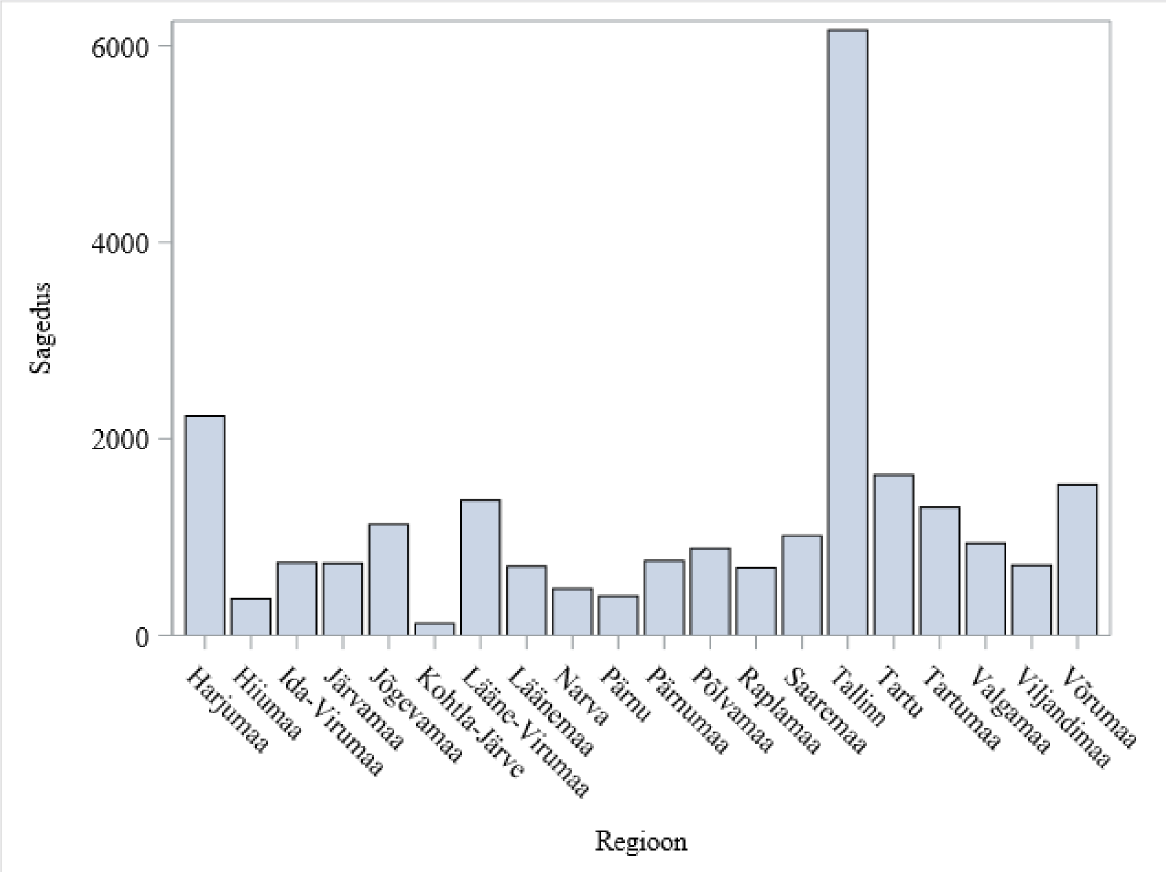
Joonis 1.5: Kliendi vanuse tasemed ja sagedused.

Sõiduki vanuse tasemed on „a. uus sõiduk“, „b. 1 aastat“, „c. 2 aastat“, „d. 3 aastat, „e. 4 aastat“, “f. 5 aastat“, „g. 6-7 aastat“, „h. 8-10 aastat“, „i. 11-15 aastat“ ja „j. üle 15 aasta“. Liikluskindlustuse portfellis on suurem osa üle 10 aasta vanuseid sõidukeid (joonis 1.6), sest enamus Eestis arvelolevaid sõidukeid on vanemapoolsemad. Võib arvata, et kliendid, kes omavad uuemaid sõidukeid, ei ole nii hinnatundlikud.



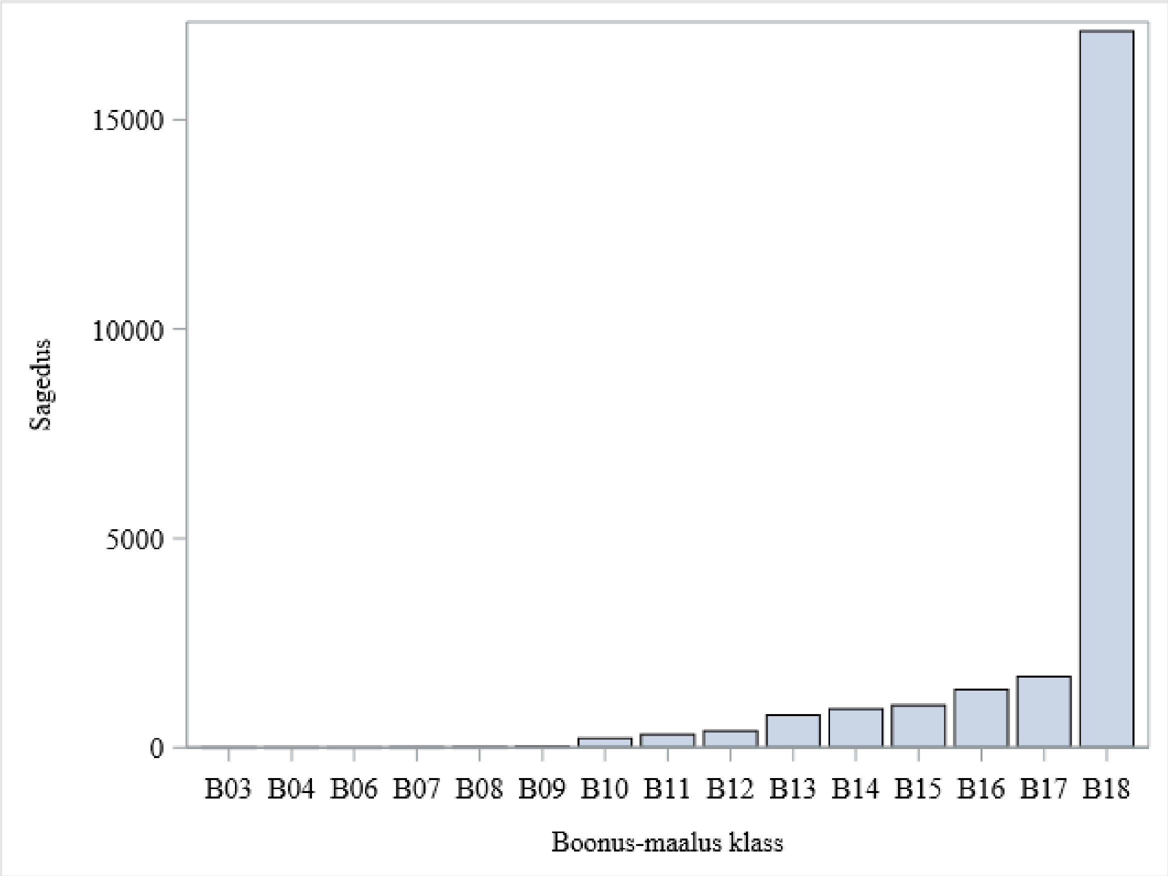
Joonis 1.6: Sõiduki vanuse tasemed ja sagedused.

Erinevate piirkondade elatustaseme, kultuurilise, religioosse jm erinevuse tõttu on oodata mõju kliendi käitumisele ka olenevalt piirkonnast, kus klient elab. Suuremad linnad (Tallinn, Tartu, Pärnu, Narva, Kohtla-Järve) on andmestikus eraldi piirkonnana, ülejäänud piirkonnad on maakonna täpsusega. Joonisel 1.7 kujutatakse IF-i portfelli jaotust piirkonniti. Tallinn, Harjumaa ja Tartu on liikluskindlustuse portfellis suurimad piirkonnad.



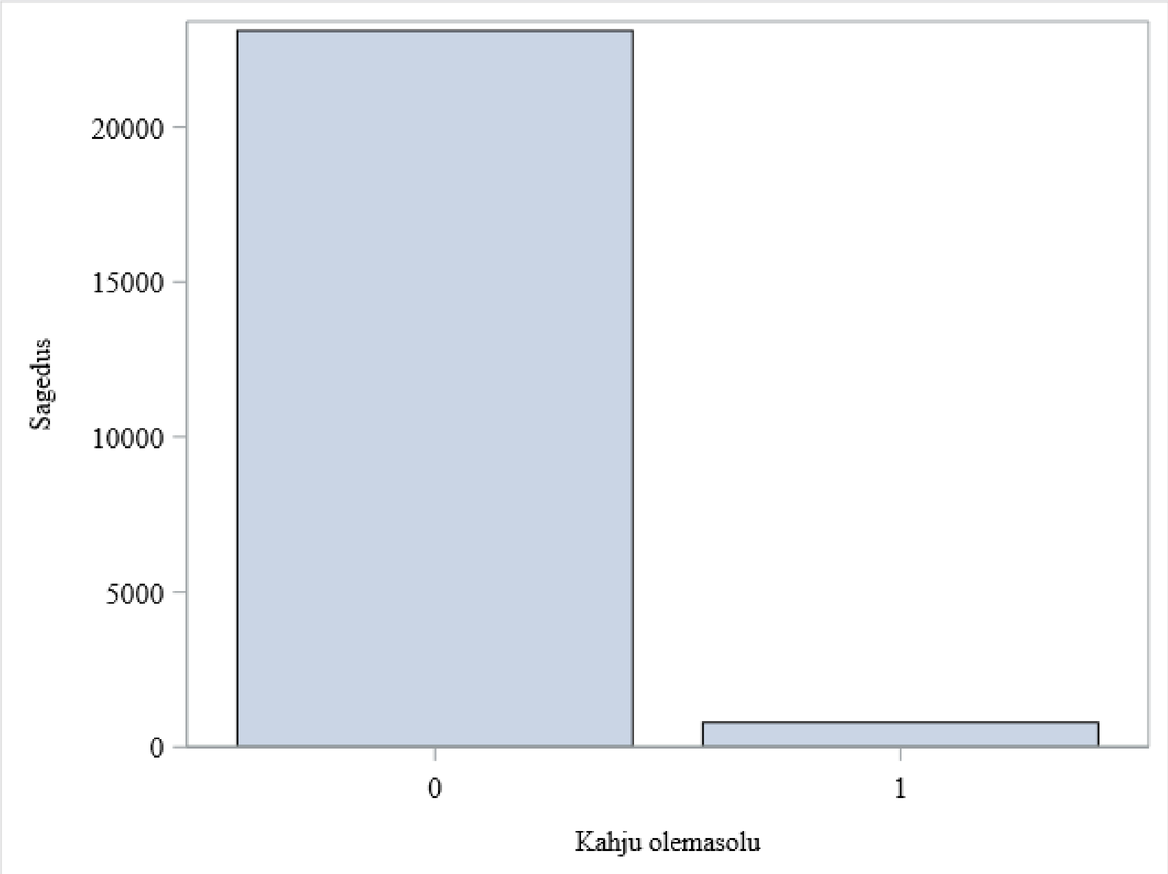
Joonis 1.7: Regiooni tasemed ja sagedused.

Lisaks pakub huvi uuendustõenäosuse sõltuvus klientide erinevast riskitasemest. Kõrgema riskiga klientide preemia on palju suurem kui kõige madalama riskiga klientide preemia. Boonus-maalus klass on kliendi kindlustus- ja kahjuajaloost sõltuv kliendi riskiklass. Boonus-maalus klassis „B18“ on kliendid, kes on kõige madalama riskiga ja boonus-maalus klass „B03“ tähistab kõige kõrgema riskiga kliente. Joonisel 1.8 on toodud boonus-maalus klasside tasemed ja klientide jaotumine tasemete lõikes.



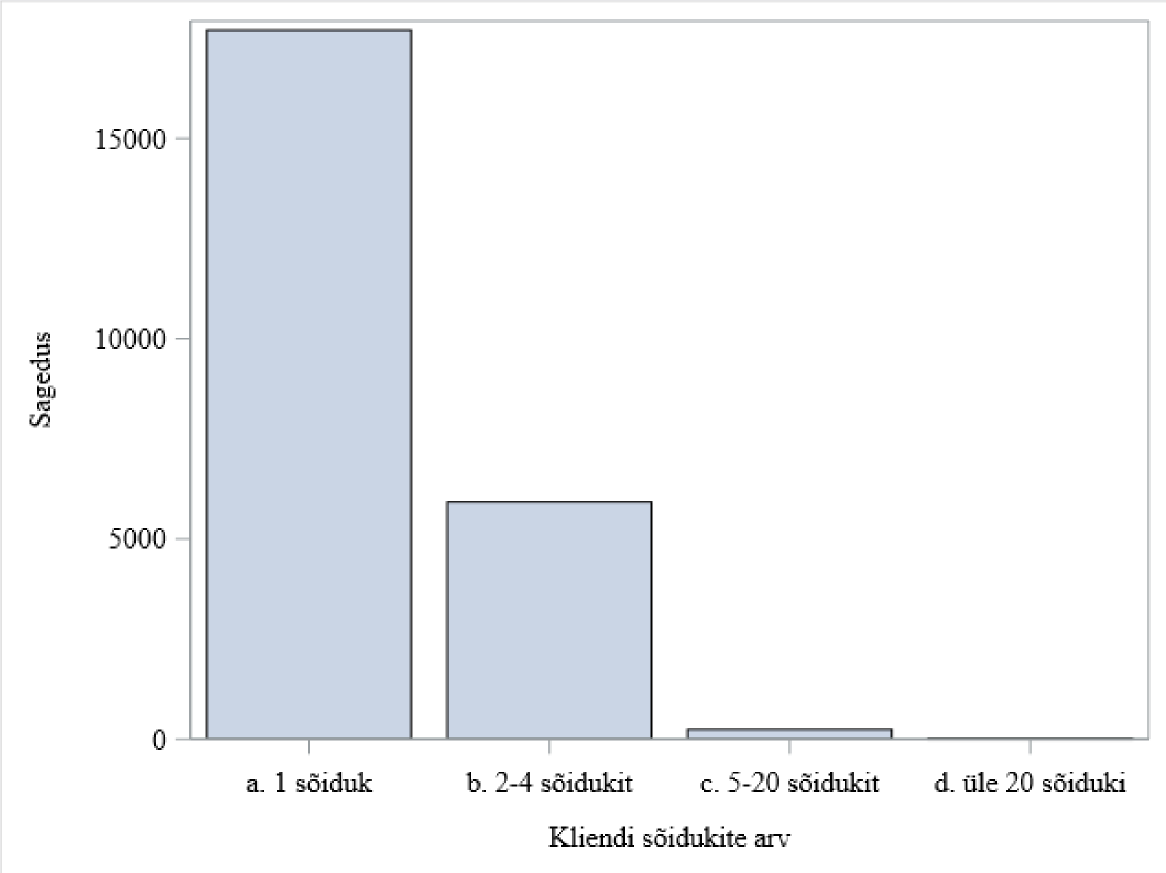
Joonis 1.8: Boonus-maalus klasside tasemed ja sagedused.

Faktor KahjuOlemasolu on indikaator liiklusõnnetuse põhjustamise kohta. Faktori tase on väärtusega 0, kui klient ei põhjustanud eelneva jõusoleva poliisi jooksul ühtegi liiklusõnnetust ja 1 kui klient põhjustas liiklusõnnetuse. Ainult 3.3% IF-i sõiduauto eraklientidest on põhjustanud liiklusõnnetuse (joonis 1.9). Liiklusõnnetuse põhjustamisel kliendi poliisi hind tõuseb märkimisväärselt, seega võib eeldada, et kahju põhjustanud klientide uuendustõenäosus on madalam kui kahjudeta klientidel.



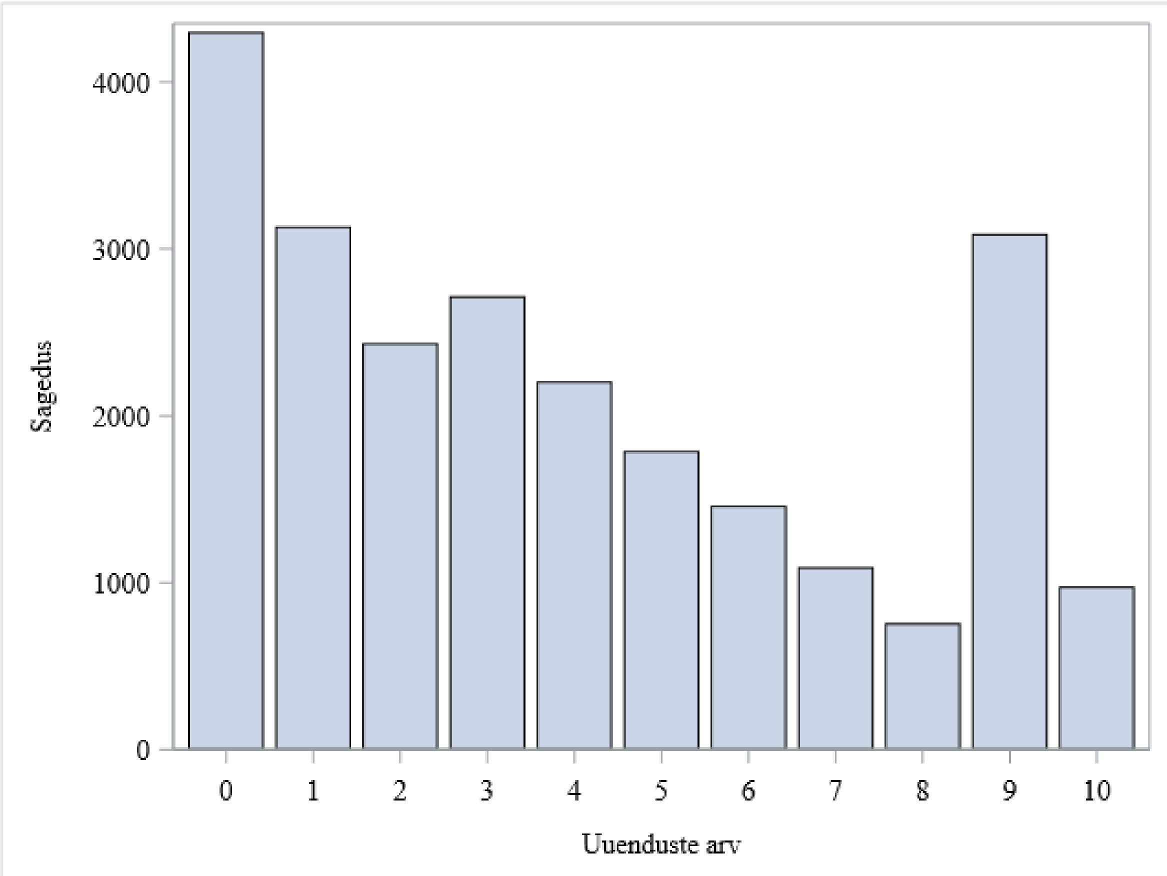
Joonis 1.9: Kahju olemasolu tasemed ja sagedused.

Üheks faktoriks valiti kliendi sõidukite arv. Enamus erakliente omab ainult ühte sõidukit, rohkem kui 20 sõidukit omab ainult 26 klienti (joonis 1.10). Võiks arvata, et mitme sõidukiga kliendid on tundlikumad hinnamuudatuse suhtes, sest mitmele sõidukile liikluskindlustuse poliisi ostmine on kulukam.



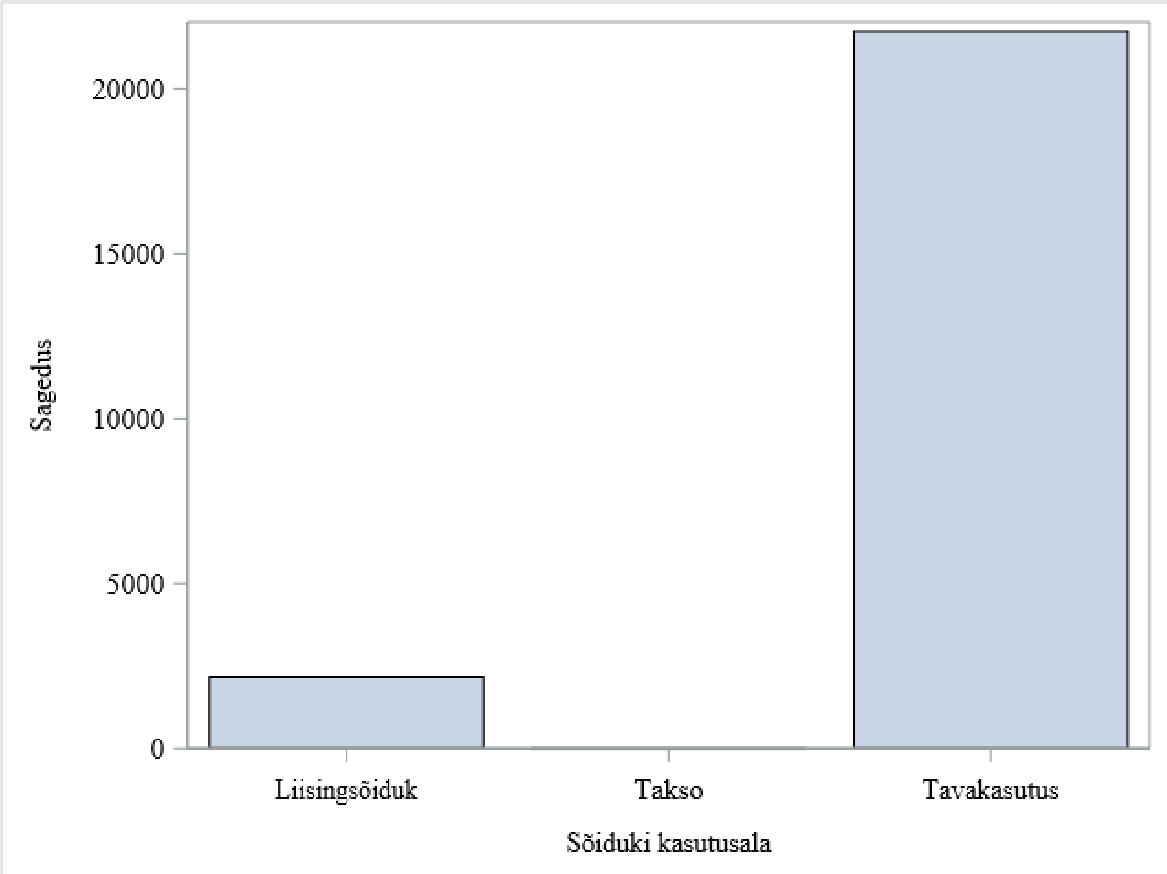
Joonis 1.10: Kliendi sõidukite arvu tasemed ja sagedused.

Faktor UuendusteArv näitab mitu korda eelnevalt on klient poliisi uuendanud (joonis 1.11). Faktori taseme 0 puhul on tegemist eelmise aasta uute klientidega, kes uuendavad esimest korda. Suurema osa portfellist moodustavad uuemapoolsemad kliendid, mis on põhjendatav liikluskindlustuse klientide tundlikkusega hinnamuudatustele. Magistritöös analüüsitakse, kas on erinevust pikaajaliste ja uute klientide uuendustõenäosuste vahel.



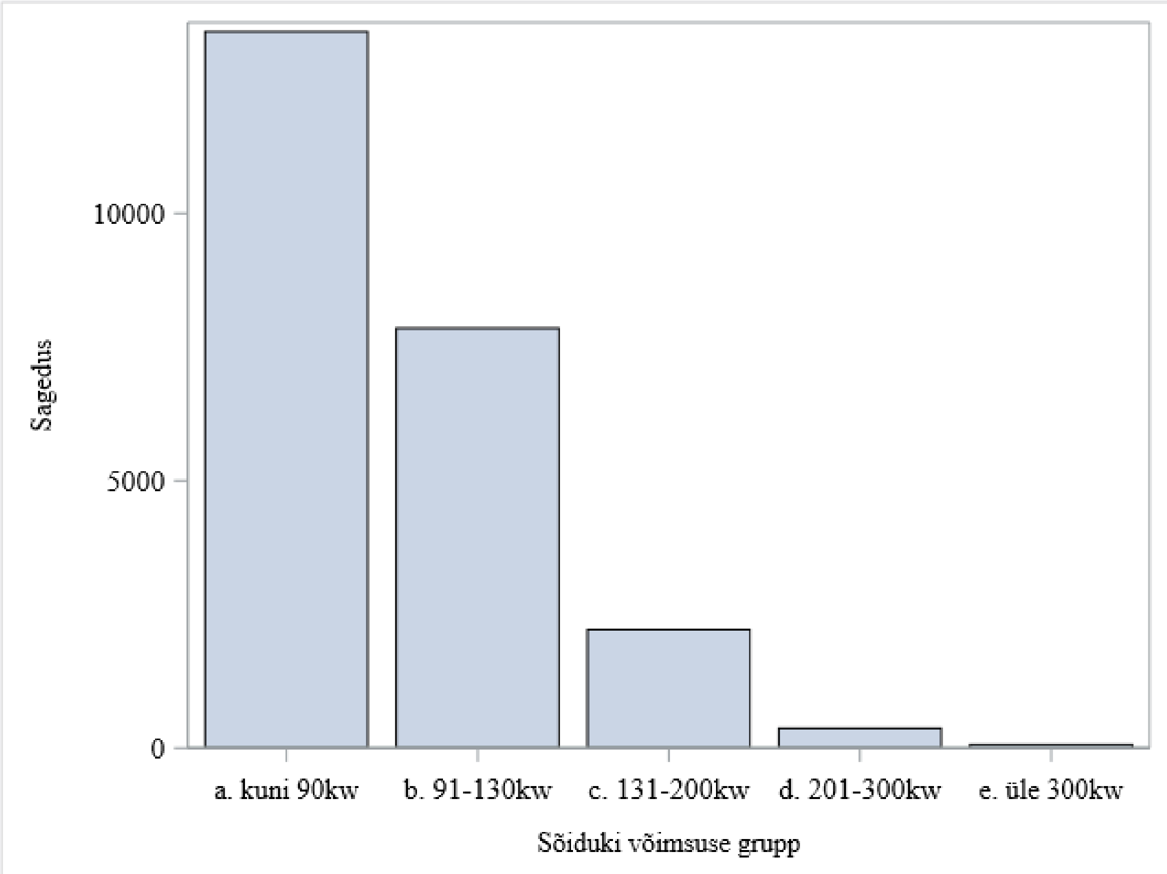
Joonis 1.11: Uuenduste arvu tasemed ja sagedused

Veel huvitab küsimus, kas sõiduki kasutusalal ja võimsusel on mõju uuendustõenäosusele. Faktor SõidukiKasutusala jagab sõidukid 3 gruppi – tavakasutus, liisingsõiduk ja takso (joonis 1.12). Liisitud sõidukid moodustavad väikese osa portfellist, kuid kuna liisingsõidukite omanikud on arvatavasti parema elatustasemega, siis nendelt oodatakse kõrgemat uuendustõenäosust. Valimis on ainult 3 takso poliisi, seda seetõttu, et enamus taksosid kuuluvad juriidilistele klientidele ning taksode preemia on väga kõrge.



Joonis 1.12: Sõiduki kasutusala tasemed ja sagedused.

Sõiduki võimsus on jaotatud tasemeteks „a. kuni 90kw“, „b. 91-130kw“, „c. 131-200kw“, „d. 201-300kw“ ja “e. üle 300kw“ (joonis 1.13). Enamus sõidukite võimsus on kuni 90 kilovatti. Kuna võimsamate sõidukite liikluskindlustuse poliis on kallim kui madalakilovatiliste sõidukite poliis, siis võiks eeldada, et kliendid, kes omavad väiksema võimsusega sõidukit, uuendavad suurema šansiga.



Joonis 1.13: Sõiduki võimsuse tasemed ja sagedused.

2 Mudeli kirjeldus

2.1 Eksponentsiaalne jaotuste pere

Öeldakse, et juhusliku suuruse Y jaotus kuulub eksponentsiaalsesse jaotuste perre, kui tihedusfunktsioon (pideval juhul) või tõenäosusfunktsioon (diskreetsel juhul) avaldub kujul

$$f(y, \theta) = a(\theta) b(y) \exp[yQ(\theta)],$$

kus

y – funktsiooni argument,

θ – jaotuse parameeter,

$a(\theta)$ – parameetri θ funktsioon, mis on diferentseeruv,

$b(y)$ – funktsioon, mis ei sõltu parameetrist θ,

$Q(\theta)$ – parameetri funktsioon (nimetatakse ka loomulikuks parameetriks).

Eksponentsiaalse jaotuste pere korral kehtivad järgnevad tingimused:

- väärtuste piirkond $\{y\colon\ f(y, \theta)\} > 0$ ei sõltu parameetrist θ,
- tihedusfunktsioonil leidub vähemalt teist järku tuletis,
- parameetri järgi diferentseerimise operatsioon ja keskväärtuse leidmine on vahetatavad (Agresti, 2013, lk 114).

2.2 Bernoulli jaotus

Paljudel kvalitatiivsetel funktsioontunnustel on ainult 2 väärtust, selliseid juhuslikke suuruseid nimetatakse binaarseteks juhuslikeks suurusteks. Binaarse funktsioontunnuse korral saab iga vaatluse klassifitseerida kahte gruppi jah/ei, on/ei ole, esineb/ei esine. Need grupid kodeeritakse tavaliselt $\{1, 0\}$ väärtusteks, kus 1 on „edu“.

Bernoulli jaotusega binaarse juhuliku suuruse $Y \sim (1, \pi)$ keskväärtus on võrdne „edu" tõenäosusega

$$\pi = E(Y) = P(Y = 1),$$

ja seega

$$P(Y = 0) = 1 - \pi.$$

Juhusliku suuruse Y dispersioon avaldub

$$V(Y) = E(Y^2) - [E(Y)]^2 = \pi(1 - \pi).$$

Kui Y on Bernoulli jaotusega parameetriga π, siis tõenäosusfunktsioon avaldub kujul

$$f(y, \pi) = P(Y = y) = \pi^y (1 - \pi)^{(1-y)} = (1 - \pi) \left[\frac{\pi}{(1-\pi)}\right]^y = (1 - \pi) \exp \left[y \ln \frac{\pi}{(1-\pi)}\right]. \quad (2.1)$$

Järelikult Bernoulli jaotus on eksponentsiaalsest jaotuste perest. Loomulik parameeter $Q(\pi)$ avaldub seega

$$Q(\pi) = \ln \left[\frac{\pi}{(1 - \pi)}\right].$$

Sündmuse $Y = 1$ šanss (*odd*) on

$$\gamma(\pi) = \frac{\pi}{(1 - \pi)},$$

järelikult $Q(\pi)$ on sündmuse $Y = 1$ šansi naturaallogaritm, mida kutsutakse ka π *logit*-iks (Agresti, 2013, lk 114-115).

2.3 Üldistatud lineaarsed mudelid

Üldistatud lineaarsed mudelid on väga lai mudelite klass. Üldistatud lineaarseid mudeleid kirjeldatakse kolme komponendiga:

- juhuslik komponent, millega määratakse funktsioontunnuse jaotus,
- süstemaatiline komponent, millega määratakse argumenttunnuste lineaarne funktsioon,
- seosefunktsioon, millega määratakse suhe süstemaatilise komponendi ja funktsioontunnuse keskväärtuse vahel.

Olgu n vaatlust ja m argumenttunnust $X_0, X_1, \dots, X_m$, kus X_0 määrab mudeli vabaliikme. Seletavate tunnuste väärtused vaatluse i korral on $\boldsymbol{x_i} = (x_{i0}, x_{i1}, \dots, x_{im})$, $i = 1, \dots, n$. Andmemaatriks tähistatakse $\boldsymbol{X} = \left(x_{ij}\right)$, $i = 1, \dots n, j = 0,1, \dots, m$, kus $x_{i0} = 1$.

Üldistatud lineaarsete mudelite juhuslikuks komponendiks on juhuslik tunnusvektor $\boldsymbol{y} = (Y_1, \dots, Y_n)^T$, kus Y_i jaotus on eksponentsiaalsest jaotuste perest parameetriga θ_i ja juhuslikud suurused $Y_i, i = 1, \dots, n$ on sõltumatud. Sealjuures juhusliku suuruse Y_i väärtus sõltub seletavatest tunnustest, $y_i = Y_i|\boldsymbol{x_i}$. Edaspidi tähistab $\boldsymbol{y}$ ka juhusliku vektori realisatsiooni $(y_1, \dots, y_n)^T$.

Süstemaatiliseks komponendiks on vektor $\boldsymbol{\eta} = (\eta_1, \dots, \eta_n)^T$, mis seob argumenttunnused mudelisse lineaarselt

$$\boldsymbol{\eta} = \boldsymbol{X\beta},$$

kus $\boldsymbol{X}$ on mudeli maatriks (plaanimaatriks) ja $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_m)^T$on mudeli parameetrite vektor. Parameeter β_0 on mudeli vabaliige. Üldistatud lineaarse mudeli rakendamisel hinnatakse parameetervektor $\boldsymbol{\beta}$, kordajad β_j näitavad, kuidas seletavad tunnused mõjutavad uuritavat tunnust.

Kolmandaks komponendiks on seos juhusliku komponendi ja süstemaatilise komponendi vahel. Tähistades $\mu_i = E(Y_i|\boldsymbol{x_i}), i = 1,$ ...,n, saadakse μ_i ja süstemaatilise komponendi η_i vaheline seos järgnevalt

$$\eta_i = g(\mu_i),$$

kus g on monotoonne diferentseeruv funktsioon. Seega üldistatud lineaarne mudel seob funktsioontunnuse keskväärtuse argumenttunnustega läbi võrdsuse

$$g(\mu_i) = \sum_{j=0}^{m} \beta_j x_{ij}, \quad i = 1, \dots, n$$

ehk

$$g(\mu_i) = \beta_0 + \beta_1 x_{i1} + \dots + \beta_m x_{im}, \qquad i = 1, \dots n.$$

Elemendiviisilise esitusena

$$\begin{bmatrix} g(\mu_1) \\ \vdots \\ \vdots \\ g(\mu_n) \end{bmatrix} = \begin{bmatrix} 1 x_{11} \dots x_{1m} \\ 1 x_{21} \dots x_{2m} \\ \vdots \ \vdots \ \cdots \ \vdots \\ 1 x_{n1} \cdots x_{nm} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_m \end{bmatrix}$$

ehk maatrikskujul

$$g(\boldsymbol{\mu}) = \boldsymbol{X\beta}.$$

Kokkuvõtlikult öeldes pakub huvi seos uuritava tunnuse keskväärtuse ja seletavate tunnuste vahel. Sealjuures argumenttunnused on mudelis alati lineaarse kombinatsioonina, millest tuleneb ka üldistatud lineaarsete mudelite nimetus. Kuid mudeliga ei pruugita lähendada uuritava tunnuse keskväärtust, vaid modelleeritavaks võib olla ka mingisugune uuritava tunnuse keskväärtuse funktsioon. Modelleerimise käigus leitakse sellised parameetrite väärtused, mille korral oleks modelleeritava lähend täpseim. Iga lähendiga kaasnevad vead, üldistatud lineaarsete mudelite puhul räägitakse erinevat tüüpi vigadest.

Üldistatud lineaarsete mudelite korral võib funktsioontunnus olla nii pidev kui diskreetne. Järgnevalt on mõned näited üldistatud lineaarsetest mudelitest olenevalt funktsioontunnuse jaotusest, seosefunktsioonist ja seletavate tunnuste tüübist:

Tabel 2.1: Üldistatud lineaarsete mudelite näited.

Juhuslik komponent	**Seosefunktsioon**	**Süstemaatiline komponent**	**Mudeli nimetus**
Normaaljaotus	Samasusseos	Pidev	Regressioonmudel
Poissoni jaotus	Log	Pidev ja/või diskreetne	Log-lineaarne mudel
Bernoulli jaotus	Logit	Pidev ja/või diskreetne	Logit mudel
Bernoulli jaotus	Probit	Pidev ja/või diskreetne	Probit mudel

Käesolevas töös kasutatakse seosefunktsioone *logit* ja *probit*. Seosefunktsioon *logit* on sündmuse šansi naturaallogaritm ja seosefunktsioon *probit* on standardse normaaljaotuse jaotusfunktsiooni pöördfunktsioon. Täpsemalt kirjeldatakse probit ja logit mudeleid järgnevates peatükkides. Logit mudelit nimetatakse ka logistilise regressiooni mudeliks. Binaarse tunnuse modelleerimisel on võimalik kasutada ka teisi seosefunktsioone, näiteks seosefunktsiooni täiend-log-log, kuid antud magistritöös neid ei vaadelda (Agresti, 2013, lk 113-114, Gernard, 2012, lk 33-35, Möls, 2011).

2.4 Mudelid binaarse tunnuse jaoks

Käesoleva magistritöö raames on uuritaval funktsioontunnusel väärtused 1 – klient uuendab poliisi ja 0 – klient ei uuenda poliisi, seega edaspidi räägitakse binaarsest Bernoulli jaotusega funktsioontunnusest. Olgu juhuslike suuruste vektor $\boldsymbol{y} = (Y_1, \dots, Y_n)^T$, kus $Y_i = Y_i|\boldsymbol{x_i} \sim B(1, \pi_i), i = 1, \dots, n$. Eesmärk on kirjeldada poliisi uuendamise tõenäosuse $\boldsymbol{\pi} = (\pi_1, \dots, \pi_n)^T$sõltuvust argumenttunnustest

$$\pi_i = P(Y_i = 1|\boldsymbol{x_i}), i = 1, \dots, n.$$

Lihtne lineaarne regressioonmudel uuritava tunnuse keskväärtuse jaoks

$$\pi_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_m x_{im}$$

ei sobiks Bernoulli jaotusega tunnuse modelleerimiseks, sest väga suurte või väga väikeste seletavate tunnuste väärtuste puhul võib juhtuda, et mõne i korral tõenäosus π_i osutub negatiivseks või suuremaks ühest. Seetõttu on vaja alternatiivset seosefunktsiooni samasusseosele (Agresti, 2013, lk 117-118, Traat, 2018).

2.4.1 Logit mudel

Funktsiooni kujul

$$\pi_i = \frac{\exp(\beta_0 + \beta_1 x_{i1} + \cdots + \beta_m x_{im})}{1 + \exp(\beta_0 + \beta_1 x_{i1} + \cdots + \beta_m x_{im})} \tag{2.2}$$

nimetatakse *logistilise regressiooni funktsiooniks*. Tõenäosusvektori $\boldsymbol{\pi}$ modelleerimiseks on vaja leida seosefunktsioon g, mille korral kehtiks

$$g(\pi_i) = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_m x_{im}, \qquad i = 1, \dots n.$$

Valemist (2.2) tuleneb, et

$$\frac{1}{\pi_i} = \frac{1 + \exp(\beta_0 + \beta_1 x_{i1} + \cdots + \beta_m x_{im})}{\exp(\beta_0 + \beta_1 x_{i1} + \cdots + \beta_m x_{im})},$$

$$\frac{1}{\pi_i} = 1 + \frac{1}{\exp(\beta_0 + \beta_1 x_{i1} + \cdots + \beta_m x_{im})},$$

$$\frac{1 - \pi_i}{\pi_i} = \frac{1}{\exp(\beta_0 + \beta_1 x_{i1} + \cdots + \beta_m x_{im})}.$$

Seega sündmuse $Y_i|\boldsymbol{x_i} = 1$ toimumise šanss avaldub

$$\frac{\pi_i}{1 - \pi_i} = \exp(\beta_0 + \beta_1 x_{i1} + \cdots + \beta_m x_{im}) = e^{\beta_0}\left(e^{\beta_1}\right)^{x_{i1}} \dots \left(e^{\beta_m}\right)^{x_{im}}. \tag{2.3}$$

Valemist (2.3) tuleneb mudeli parameetrite interpretatsioon: kõikide teiste argumenttunnuste fikseerides, argumendi kasvamisel ühiku võrra kasvab sündmuse „klient uuendab poliisi“

šanss $e^{\beta_j}, j = 1, \dots, m$ korda. Kordaja e^{β_0} näitab poliisi uuendamise šanssi kui kõikide teiste argumentide väärtused on nullid. Rakendades seosele (2.3) logaritmilist teisendust saadakse

$$\ln\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_m x_{im}. \tag{2.4}$$

Logaritmiline šanss $\ln\big(\gamma(\pi_i)\big)$ ehk π_i *logit* on sobilik seosefunktsioon mudeli teisendamiseks lineaarsele kujule. Seosefunktsiooni, mis seab keskväärtuse vastavusse loomuliku parameetriga $Q(\pi_i)$ nimetatakse kanooniliseks seosefunktsiooniks (Agresti, 2013, lk 119-120, 182-183).

Kindlustuspoliiside hinnastamisel kasutatakse enamasti tunnuseid, mis on juba grupeeritud. Samuti on kõik analüüsiks valitud seletavad tunnused kvalitatiivsed. Lihtsuse mõttes vaadeldakse ühte kvalitatiivset kahe tasemega seletavat tunnust. Kvalitatiivsete tunnuste paremaks interpreteerimiseks kasutatakse tunnuse ühe taseme suhtes modelleerimist ehk kasutatakse $\{1, 0\}$ kodeeringut selliselt, et seletava tunnuse üks tase määratakse baastasemeks. Kahe tasemega seletava tunnuse X puhul oleks $\{1, 0\}$ kodeering

$$X = \begin{cases} 1, kui\ on\ tegemist\ tasemega\ 1, \\ 0, kui\ on\ tegemist\ tasemega\ 2. \end{cases}$$

Esimese taseme korral sündmuse $Y_i = 1$ (klient uuendab poliisi) tõenäosus tähistatakse π_{i1} ja teise taseme korral sündmuse $Y_i = 1$ tõenäosus tähistatakse π_{i2}. Tõenäosused saab esitada sagedustabelina (tabel 2.2).

Tabel 2.2: Tõenäosuste sagedustabel.

	$\boldsymbol{Y_i = 1}$	$\boldsymbol{Y_i = 0}$
Tase 1	π_{i1}	$1 - \pi_{i1}$
Tase 2	π_{i2}	$1 - \pi_{i2}$

Tasemete šansid avalduvad

$$\gamma(\pi_{i1}) = \frac{\pi_{i1}}{1 - \pi_{i1}},$$

$$\gamma(\pi_{i2}) = \frac{\pi_{i2}}{1 - \pi_{i2}}.$$

Logistilise regressiooni mudeli kujust (valem 2.4) saadakse $\{1, 0\}$ kodeeringu korral

$$\ln\left(\frac{\pi_{i1}}{1 - \pi_{i1}}\right) = \beta_0 + \beta_1 * 1 = \beta_0 + \beta_1,$$

$$\ln\left(\frac{\pi_{i2}}{1-\pi_{i2}}\right) = \beta_0 + \beta_1 * 0 = \beta_0,$$

millest saab avaldada

$$\beta_1 = \ln\left(\frac{\pi_{i1}}{1-\pi_{i1}}\right) - \ln\left(\frac{\pi_{i2}}{1-\pi_{i2}}\right) = \ln[\gamma(\pi_{i1})] - \ln[\gamma(\pi_{i2})] = \ln\left[\frac{\gamma(\pi_{i1})}{\gamma(\pi_{i2})}\right],$$

$$\exp(\beta_1) = \frac{\gamma(\pi_{i1})}{\gamma(\pi_{i2})}.$$

Järelikult parameeter β_0 on teise taseme sündmuse „klient uuendab poliisi“ saamise šansi naturaallogaritm ja $\exp(\beta_0)$ on teise taseme klientide uuendamise šanss. Parameeter β_1 on seletava tunnuse kahe taseme poliisi uuendamise šansside suhte naturaallogaritm ja $\exp(\beta_1)$ on seletava tunnuse tasemete šansside suhe. Kui parameeter $\beta_1 = 0$ või kui $\exp(\beta_1) = 1$, siis pole erisusi kahe taseme uuendustõenäosuse vahel. Kui $\exp(\beta_1)$ on suurem kui 1, siis esimese taseme klientide uuendamise šanss on suurem kui teise taseme klientide uuendamise šanss. Kui $\exp(\beta_1)$ on väiksem kui 1, siis teise taseme kliendid uuendavad suurema šansiga võrreldes esimese taseme klientidega. Kuigi modelleeritakse parameetrit β_1, siis statistika tarkvarad väljastavad ka väärtuse $\exp(\beta_1)$, sest selle kaudu on mudeli tulemusi lihtsam interpreteerida (Gernard, 2012, lk 15-16, 31-33).

2.4.2 Probit mudel

Vaja on leida mudel tõenäosuse $\pi_i = P(Y_i = 1|\boldsymbol{x_i}), i = 1, \dots, n$ modelleerimiseks

$$g(\pi_i) = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_m x_{im}, i = 1, \dots n,$$

mis kasutaks alternatiivset seosefunktsiooni g $logit$ asemel. Seosena (2.2) esitatud logistilise regressiooni funktsioonil on jaotusfunktsiooni omadused. Seetõttu võib arvata, et sobiks mudel

$$\pi_i = F(\beta_0 + \beta_1 x_{i1} + \cdots + \beta_m x_{im}), i = 1, \dots, n,$$

kus F on mingisugune jaotusfunktsioon.

Olgu juhuslikud suurused Y_i^*, millel leidub selline väärtus y_i^*, et kehtib järgmine võrdus

$$\begin{cases} y_i = 0, kui\ y_i^* \leq \tau_i, i = 1, \dots, n, \\ y_i = 1, kui\ y_i^* > \tau_i, i = 1, \dots, n. \end{cases}$$

Olgu $Y_i^* = \pi_i + \epsilon_i$, kus $\pi_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_m x_{im}$ ja ϵ_i on sõltumatud normaaljaotusega juhuslikud suurused jaotusega $N(0, \sigma_i^2)$. Siis kehtib

$$\pi_i = P(Y_i = 1) = P({Y_i}^* > \tau_i) = P(\beta_0 + \beta_1 x_{i1} + \cdots + \beta_m x_{im} + \epsilon_i > \tau_i) =$$

$$= P(-\epsilon_i < \beta_0 + \beta_1 x_{i1} + \cdots + \beta_m x_{im} - \tau_i) =$$

$$= \Phi\left(\frac{\beta_0 + \beta_1 x_{i1} + \cdots + \beta_m x_{im} - \tau_i}{\sigma_i}\right),$$

kus Φ on standardse normaaljaotuse jaotusfunktsioon. Kui $\sigma_i = 1$ ja $\tau_i = 0$, siis

$$\pi_i = \Phi(\beta_0 + \beta_1 x_{i1} + \cdots + \beta_m x_{im}) \tag{2.5}$$

ja

$$\Phi^{-1}(\pi_i) = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_m x_{im},$$

seega standardse normaaljaotuse jaotusfunktsiooni pöördfunktsioon on sobilik seosefunktsioon teisendamaks seose (2.2) üldistatud lineaarsete mudelite kujule. Seosefunktsioon Φ^{-1} seab tõenäosustele vahemikus $(0,1)$ vastavusse väärtused vahemikus $(-\infty, \infty)$.

Valemist (2.5) tuleneb parameetri $\beta_j, j = 1, \ldots, m$ interpretatsioon: ühikuline muutus argumendi väärtuses toob kaasa muutuse standardse normaaljaotuse jaotusfunktsioonis ehk muutuse uuendamise tõenäosuses. Vabaliikme β_0 abil arvutatud väärtus $\Phi(\beta_0)$ näitab, milline on uuendamise tõenäosus kui kõikide teiste argumentide väärtused on nullid.

Ühe kvalitatiivse kahe tasemega seletava tunnuse X korral, kus

$$X = \begin{cases} 1, kui\ on\ tegemist\ tasemega\ 1, \\ 0, kui\ on\ tegemist\ tasemega\ 2 \end{cases}$$

ja poliisi uuendamise tõenäosus esimese taseme korral on tähistatud π_{i1} ning teise taseme korral π_{i2}, saadakse valemist (2.5)

$$\pi_{i1} = \Phi(\beta_0 + \beta_1 * 1) = \Phi(\beta_0 + \beta_1),$$

$$\pi_{i2} = \Phi(\beta_0 + \beta_1 * 0) = \Phi(\beta_0).$$

Seega $\Phi(\beta_0 + \beta_1)$ on esimese taseme uuendamise tõenäosus ja $\Phi(\beta_0)$ on teise taseme uuendamise tõenäosus (Agresti, 2013, lk 121-122, 251-253).

2.5 Mudeli parameetrite hindamine

Olgu $Y_i \sim B(1, \pi_i), i = 1, \dots, n$, vaatluste vektor $\boldsymbol{y} = (y_1, \dots, y_n)^T$ ja vaatlustele vastav tõenäosuste vektor $\boldsymbol{\pi} = (\pi_1, \dots, \pi_n)^T$. Seosest (2.1) tuleneb Bernoulli jaotusega juhusliku vektori tõepärafunktsioon

$$L(\boldsymbol{y}, \boldsymbol{\pi}) = \prod_{i=1}^{n} f(y_i, \pi_i) = \prod_{i=1}^{n} \pi_i{}^{y_i}(1 - \pi_i)^{1-y_i}. \tag{2.6}$$

Seosest (2.2) on näha, et tõenäosused π_i sõltuvad parameetervektorist $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_m)^T$, seega sõltub ka tõepärafunktsioon parameetritest. Mudeli parameetrite hindamiseks maksimiseeritakse tõepärafunktsioon $\boldsymbol{\beta}$ suhtes. Maksimumpunkt $\widehat{\boldsymbol{\beta}} = (\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_m)^T$ on parameetervektori $\boldsymbol{\beta}$ suurima tõepära hinnanguks. Kuna tõepärafunktsiooni ja logaritmilise tõepärafunktsiooni maksimumpunkt on sama, siis praktikas maksimiseeritakse logaritmilist tõepärafunktsiooni. Logaritmiline tõepärafunktsioon avaldub seosest (2.6)

$$\begin{aligned} l(\boldsymbol{y}, \boldsymbol{\pi}) &= \ln[L(\boldsymbol{y}, \boldsymbol{\pi})] \\ &= \sum_{i=1}^{n} \ln[\pi_i{}^{y_i}(1 - \pi_i)^{1-y_i}] \\ &= \sum_{i=1}^{n} \ln\left[\pi_i{}^{y_i}(1 - \pi_i)\frac{1}{(1 - \pi_i)^{y_i}}\right] \\ &= \sum_{i=1}^{n} \ln\left[(1 - \pi_i)\left(\frac{\pi_i}{1 - \pi_i}\right)^{y_i}\right] \\ &= \sum_{i=1}^{n} \left[\ln(1 - \pi_i) + y_i \ln\frac{\pi_i}{1 - \pi_i}\right]. \end{aligned}$$

Maksimiseerimiseks leitakse logaritmilise tõepärafunktsiooni tuletis parameetervektori $\boldsymbol{\beta}$ järgi, mis võrdsustatakse 0-ga ehk lahendatakse võrrand

$$s(\boldsymbol{\beta}) = \sum_{i=1}^{n} s_i(\boldsymbol{\beta}) = 0, \tag{2.7}$$

kus

$$s_i(\boldsymbol{\beta}) = \left(\frac{\partial l_i(\boldsymbol{\beta})}{\partial \beta_0}, \frac{\partial l_i(\boldsymbol{\beta})}{\partial \beta_1}, \dots, \frac{\partial l_i(\boldsymbol{\beta})}{\partial \beta_m}\right).$$

Funktsiooni $s(\boldsymbol{\beta})$ nimetatakse skoorifunktsiooniks. Võrrandi (2.7) lahendamiseks kasutatakse iteratiivseid meetodeid. Statistikatarkvara SAS vaikimisi parameetrite hindamismeetod on Fisheri skoorimeetod (Möls, 2011, SAS Institute Inc.).

Mudeli parameetrite hindamisel saadakse parameetervektori $\boldsymbol{\beta}$ suurima tõepära hinnang $\widehat{\boldsymbol{\beta}} = (\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_m)^T$. Suure valimi korral on hinnangud asümptootilise normaaljaotusega. Kui parameetri hinnangud $\hat{\beta}_j,\ j = 0, \dots, m$ on arvutatud, siis kontrollitakse, kas parameetrid on oluliselt erinevad nullist ehk kontrollitakse hüpoteese:

$$\begin{cases} H_0\colon \beta_j = 0, j = 0, \dots, m, \\ H_1\colon \beta_j \neq 0, j = 0, \dots, m. \end{cases}$$

Hüpoteeside kontrollimiseks kasutatakse Waldi teststatistiku väärtust

$$\chi_W^2 = \left(\frac{\hat{\beta}_j}{\hat{\sigma}_j}\right)^2,$$

kus $\hat{\sigma}_j$ on parameetri β_j hinnangu $\hat{\beta}_j$ standardviga. Waldi teststatistik on asümptootiliselt jaotusega $N(0,1)$ eeldusel, et ei suudeta ümber lükata nullhüpoteesi.

Lisaks Waldi statistiku väärtusele väljastab statistikatarkvara SAS parameetrite Waldi usaldusvahemikud. Waldi $100(1-\alpha)\%$ usaldusvahemik parameetri β_j jaoks avaldub

$$\hat{\beta}_j \pm z_{1-\frac{\alpha}{2}} \hat{\sigma}_j,$$

kus

$\hat{\beta}_j$ – parameetri β_j suurima tõepära hinnang,

$z_{1-\frac{\alpha}{2}}$ – normaaljaotuse $N(0,1)$ kvantiil,

$\hat{\sigma}_j$ – parameetri β_j hinnangu $\hat{\beta}_j$ standardviga.

Hinnangute normaaljaotus kehtib ligikaudu suure valimi korral ning väga väikeste või suurte tõenäosuste korral võivad usaldusvahemikud olla ebasümmeetrilised (Agresti, 2013, lk 169-170).

2.6 Mudeli sobivus

Olgu mudeli parameetrite hinnang $\widehat{\boldsymbol{\beta}} = (\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_m)^T$ ja uuritava tunnuse vaatluste vektor $\boldsymbol{y} = (y_1, \dots, y_n)^T$. Seosest (2.2) saab seega iga vaatluspunkti i jaoks arvutada tõenäosuste hinnangud $\hat{\boldsymbol{\pi}} = (\hat{\pi}_1, \dots, \hat{\pi}_n)$, mis Bernoulli jaotusega juhusliku suuruse Y_i korral on ühtlasi vaatluse prognoosid

$$\hat{y}_i = \hat{E}(Y_i|\boldsymbol{x_i}) = \hat{\pi}_i.$$

Kui mudeli prognoosid on sellised, et

$$\hat{\pi}_i = y_i,$$

siis sellist mudelit nimetatakse küllastunud mudeliks. Küllastunud mudelit saab saavutada üleparametriseerimisega, mis tähendab, et mudelis on nii palju seletavaid tunnuseid ja seega hinnatavaid parameetreid, et mudel sobib hästi valitud andmetega, kuid ei sobi prognoosimiseks. Küllastunud mudelit kasutatakse uuritava mudeli sobivuse hindamiseks. Mudeli kooskõla hindamine tähendab järgnevate hüpoteeside kontrolli:

$$\begin{cases} H_0: hinnatud\ mudel\ ei\ erine\ oluliselt\ küllastunud\ mudelist, \\ H_1: mudeli\ sobivusmäär\ pole\ hea, mudel\ pole\ adekvaatne. \end{cases}$$

Logaritmiliste tõepärafunktsioonide abil defineeritakse statistik

$$D = D(\hat{\boldsymbol{\pi}}, \boldsymbol{y}) = -2\left[\ln\left(L(\hat{\boldsymbol{\pi}}, \boldsymbol{y})\right) - \ln\left(L(\boldsymbol{y}, \boldsymbol{y})\right)\right], \tag{2.8}$$

kus $L(\boldsymbol{y}, \boldsymbol{y})$ on küllastunud mudeli juhuslik tõepärafunktsioon, $L(\hat{\boldsymbol{\pi}}, \boldsymbol{y})$ hinnatud mudeli juhuslik tõepärafunktsioon ja $\boldsymbol{y} = (Y_1, \dots, Y_n)^T$. Statistikut D nimetatakse hälbimuseks. Suure valimi korral on hälbimus D ligikaudu hii-ruut jaotusega $D(\hat{\boldsymbol{\pi}}, \boldsymbol{y}) \sim {\chi^2}_{n-p}$, kus p on mudeli parameetrite arv. Kui hälbimus on suur, siis antud mudel ei sobi.

Hälbimust (2.8) kasutatakse ka kahe mudeli sobivuse võrdlemiseks. Olgu $\hat{\boldsymbol{\pi}}_{\boldsymbol{0}}$ lihtsama mudeli tõenäosuste hinnangute vektor ja $\hat{\boldsymbol{\pi}}$ hinnatud mudeli tõenäosuste vektor, siis

$$L(\hat{\boldsymbol{\pi}}_{\boldsymbol{0}}, \boldsymbol{y}) \leq L(\hat{\boldsymbol{\pi}}, \boldsymbol{y}),$$

sest lihtsam mudel sisaldab vähem parameetreid kui hinnatud mudel. Seega valemist (2.8) tuleneb, et

$$D(\hat{\boldsymbol{\pi}}, \boldsymbol{y}) \leq D(\hat{\boldsymbol{\pi}}_{\boldsymbol{0}}, \boldsymbol{y}).$$

Tarkvara väljundis on sageli vabaliikmega mudeli hälbimus. Kui hinnatud mudeli hälbimus ei erine oluliselt vabaliikmega mudeli hälbimusest, siis mudel ei ole sobiv (Agresti, 2013, lk 136-139)

Mudeli sobivust hinnatakse veel Pearsoni χ^2- statistiku järgi, mille väärtus avaldub

$$\chi_P^2 = \sum_{i=1}^{n} \frac{(y_i - \hat{\pi}_i)^2}{\hat{\pi}_i(1 - \hat{\pi}_i)}.$$

Pearsoni χ^2- statistik on valimimahu kasvades asümptootiliselt ${\chi^2}_{n-p}$ jaotusega, kus p on mudeli parameetrite arv (Agresti, 2013, lk 139-140).

Magistritöös vaadeldakse veel järgmisi kooskõlakriteeriume:

- tõepärastatistiku väärtus $-2\ln L$,
- Akaike informatsioonikriteerium $\mathrm{AIC} = -2\ln L + 2p$,
- Schwarzi kriteerium $\mathrm{SC} = -2\ln L + p\ln n$,

kus p on mudeli parameetrite arv. Suurust $-2\ln L$ kasutatakse kriteeriumide AIC ja SC arvutamiseks, AIC ja SC väärtuseid kasutatakse mudelite võrdlemiseks. Mudelite võrdlemisel väiksema AIC ja SC väärtustega mudelid on sobilikumad.

2.7 Mudeli prognoosivõime

Pärast mudeli parameetrite hindamist saab vaatlusvektorile $\boldsymbol{y} = (y_1, \dots, y_n)^T$ vastavusse seada hinnatud tõenäosused $\hat{\boldsymbol{\pi}} = (\hat{\pi}_1, \dots, \hat{\pi}_n)^T$. Kuna binaarse juhusliku suuruse võimalikud väärtused on 0 ja 1, siis oleks loomulik öelda, et vaatluse y_i hinnang $\hat{y}_i$=$\hat{\pi}_i$ kuulub klassi $\{1\}$ kui hinnatud tõenäosus $\hat{\pi}_i$ on ühe lähedal. Hinnatud vaatluse klassifitseerimiseks $\{1, 0\}$ klassi defineeritakse lävi π_0

$$\hat{y}_i = 1, kui\ \hat{\pi}_i > \pi_0\ ja\ \hat{y}_i = 0, kui\ \hat{\pi}_i \leq \pi_0, i = 1, \dots, n.$$

Selliselt saab iga vaatluse korral vaadata, kui hästi mudel prognoosib tegelikku väärtust. Varieerides suurust π_0 ja kandes sündmuse 1 õigesti ja valesti klassifitseerimiste osakaalud graafikule, saadakse kõver, mida nimetatakse ROC (*Receiver Operating Characteristic Curve*) kõveraks (Agresti, 2013, lk 224-225, Traat, 2018).

2.8 Mudeli jäägid

Jäägid näitavad, kuidas sobivad mudeli hinnangud igas vaatluspunktis. Seega jääk on mudeli juhusliku vea hinnang. Üldistatud lineaarsete mudelite puhul ei saa rääkida lihtsalt uuritava tunnuse ja uuritava tunnuse mudeli prognoosi vahest, vaid räägitakse üldistatud jääkidest. Üldistatud lineaarsete mudelite jääke on erinevaid, antud magistritöös vaadeldakse Pearsoni ja hälbimusjääke.

Olgu $\hat{\pi}_i$ tõenäosuse hinnang vaatluse y_i korral. Mudeli Pearsoni jääk i-ndas vaatluspunktis y_i defineeritakse

$$r_{Pi} = \frac{y_i - \hat{\pi}_i}{\sqrt{\hat{\pi}_i(1 - \hat{\pi}_i)}}.$$

Pearsoni jääkide ruutude summa avaldub Pearsoni χ^2-statistiku väärtusena.

Mudeli hälbimusjääk i-ndas vaatluspunktis y_i defineeritakse

$$r_{Di} = \operatorname{sign}(y_i - \hat{\pi}_i)\sqrt{d_i},$$

kus

$$d_i = 2\left(y_i \ln\frac{y_i}{\hat{\pi}_i} + (1 - y_i)\ln\frac{1 - y_i}{1 - \hat{\pi}_i}\right).$$

Hälbimusjääkide ruutude summa annab kokku hälbimuse statistiku väärtuse (Agresti, 2013, lk 140-142, 215-216).

3 Statistiline analüüs

3.1 Andmete töötlus

Lõputöö eesmärk on analüüsida kliendi tundlikkust hinnamuudatustele, seega kliendi sensitiivsusest tervikliku pildi saamiseks on vaja koguda andmeid ka nende poliiside hinna kohta, millest klient keeldus. Lisaks jõustatud poliisidele lisati andmestikku info poliiside kohta, mis saadeti kliendile uuenduse pakkumisena ning info kliendi hinnaarvutuste kohta elektroonilises kanalis. Uuenduspoliisina defineeriti poliis, mille algusaeg on kuni 15 päeva enne või pärast eelmise poliisi lõppu. Kui klient valis järgpoliisi erineva pikkusega kui eelmise poliisi, siis jäeti need poliisid andmestikust välja, sest poliisi pikkuse vahetamisel oleks uuenduse poliisi hinnamuutus kunstlikult tekitatud.

Üks modelleerimise probleeme on puuduvate andmete töötlemine. Puuduvate andmete korral asendatakse tavaliselt lüngad mingisuguse väärtusega, näiteks keskmise väärtusega või kustutatakse. Kuna antud projektis oli info puudulik 2 451 (ca 9% andmestikust) kliendi poliisi uuenduspreemia kohta, siis need poliisid jäeti valimist välja.

Lisaks rakendati erinevaid andmete töötlemise võtteid, et saada toorandmetest analüüsiks kõlbulikud andmed.

Andmete kogumiseks, töötlemiseks ja analüüsimiseks kasutatakse statistikatarkvara SAS.

Analüüsiks kasutatav valimi maht peale andmete töötlust on 23 908 rida. Andmestikus on 16 tunnust, nendest 2 on pidevad tunnused ja 14 diskreetsed tunnused. Uuritavaks tunnuseks on UuenduseIndikaator. Pidevaid tunnuseid EelminePreemia ja UuendusePreemia ei kasutata otseselt modelleerimisel, vaid nende abil defineeritakse tunnused HinnaMuutuseGrupp, HinnaMuutuseGruppEurodes ja PreemiaGrupp. Pidevaid tunnuseid ei lisata mudelisse, sest huvi pakub liikluskindlustuse hinnastamisel juba kasutuses olevate gruppide analüüsimine, lisaks saadud tulemuste rakendamisel kindlustustegevuses ei ole tehniliselt võimalik määrata lõpmata arv hinnamuutuse ja preemia suuruse gruppe. Analüüsitavatest tunnustest annab ülevaate tabel 3.1. Tunnuseid kirjeldati täpsemalt magistritöö esimeses osas.

Tabel 3.1: Tunnuste loetelu.

Tunnuse number	Tunnus	Tunnuse nimi	Tunnuse tüüp
1	EelminePreemia	Eelmine preemia	Pidev
2	UuendusePreemia	Uuenduse preemia	Pidev
3	UuenduseIndikaator	Uuenduse indikaator	Diskreetne
4	HinnaMuutuseGrupp	Hinnamuutuse grupp	Diskreetne
5	HinnaMuutuseGruppEurodes	Hinnamuutuse grupp eurodes	Diskreetne
6	PreemiaGrupp	Preemia suuruse grupp	Diskreetne
7	PoliisiKanal	Kanal	Diskreetne
8	KliendiVanuseGrupp	Kliendi vanuse grupp	Diskreetne
9	SõidukiVanuseGrupp	Sõiduki vanuse grupp	Diskreetne
10	Regioon	Regioon	Diskreetne
11	BoonusMaalusKlass	Boonus-maalus klass	Diskreetne
12	KahjuOlemasolu	Kahju olemasolu	Diskreetne
13	SõidukiteArvuGrupp	Kliendi sõidukite arv	Diskreetne
14	UuendusteArv	Uuenduste arv	Diskreetne
15	SõidukiKasutusala	Sõiduki kasutusala	Diskreetne
16	SõidukiVõimsuseGrupp	Sõiduki võimsuse grupp	Diskreetne

3.2 Mudeli koostamine

Mudeli prognoosivõime hindamiseks jagatakse kõigepealt andmestik selliselt, et ühe andmestiku osa pealt hinnatakse mudeli parameetrid ja teise andmestiku osa pealt kontrollitakse hinnatud mudeli sobivust. Andmestik jagatakse suhtega $\{70\%, 30\%\}$ treeningandmestikuks ja testandmestikuks. Selleks kasutatakse tarkvara SAS protseduuri *surveyselect* (SAS Institute Inc.). Mudeli koostamiseks lõplik valimimaht on seega $16\ 737$.

Koostatakse logistilise regressiooni mudel uuritava tunnuse UuenduseIndikaator modelleerimiseks. Selleks kasutatakse tarkvara SAS protseduuri *proc logistic* (SAS Institute Inc.). Mudel on koostatud uuritava tunnuse Y väärtuse $Y = 1$ hindamiseks ehk modelleeritakse kliendi uuendamise tõenäosust. Seletavateks tunnusteks valitakse HinnaMuutuseGrupp, HinnaMuutuseGruppEurodes, PreemiaGrupp PoliisiKanal, KliendiVanuseGrupp, SõidukiVanuseGrupp, Regioon, BoonusMaalusKlass KahjuOlemasolu, SõidukiteArvuGrupp, UuendusteArv, SõidukiKasutusala ja SõidukiVõimsuseGrupp. Seletavate tunnuste mudelisse määramiseks kasutatakse sammuviisilist (*stepwise*) argumenttunnuste valiku meetodit. Igal sammul valitakse mudelisse statistiliselt kõige olulisem seletav tunnus ja visatakse mudelist välja mitteolulised seletavad tunnused. Nii tunnuste lisamisel kui eemaldamisel kasutatakse olulisusnivood $\alpha = 0.05$, mis on ühtlasi tarkvara SAS vaikeväärtus (SAS Institute Inc.).

Tunnuseid HinnaMuutuseGruppEurodes, SõidukiKasutusala ja SõidukiVõimsuseGrupp ei valitud mudelisse. Suhteline hinnamuutus on olulisem kui absoluutne hinnamuutus, see tähendab, et hinnamuutus protsentides seletab juba suurema osa kliendi käitumisest. Seega võib öelda, et uuendustõenäosuse kirjeldamisel on olulised klienti iseloomustavad tunnused, mitte sõidukiga seotud tunnused (ka erineva vanusega sõiduki omamine iseloomustab kliendi jõukust ja eelistusi).

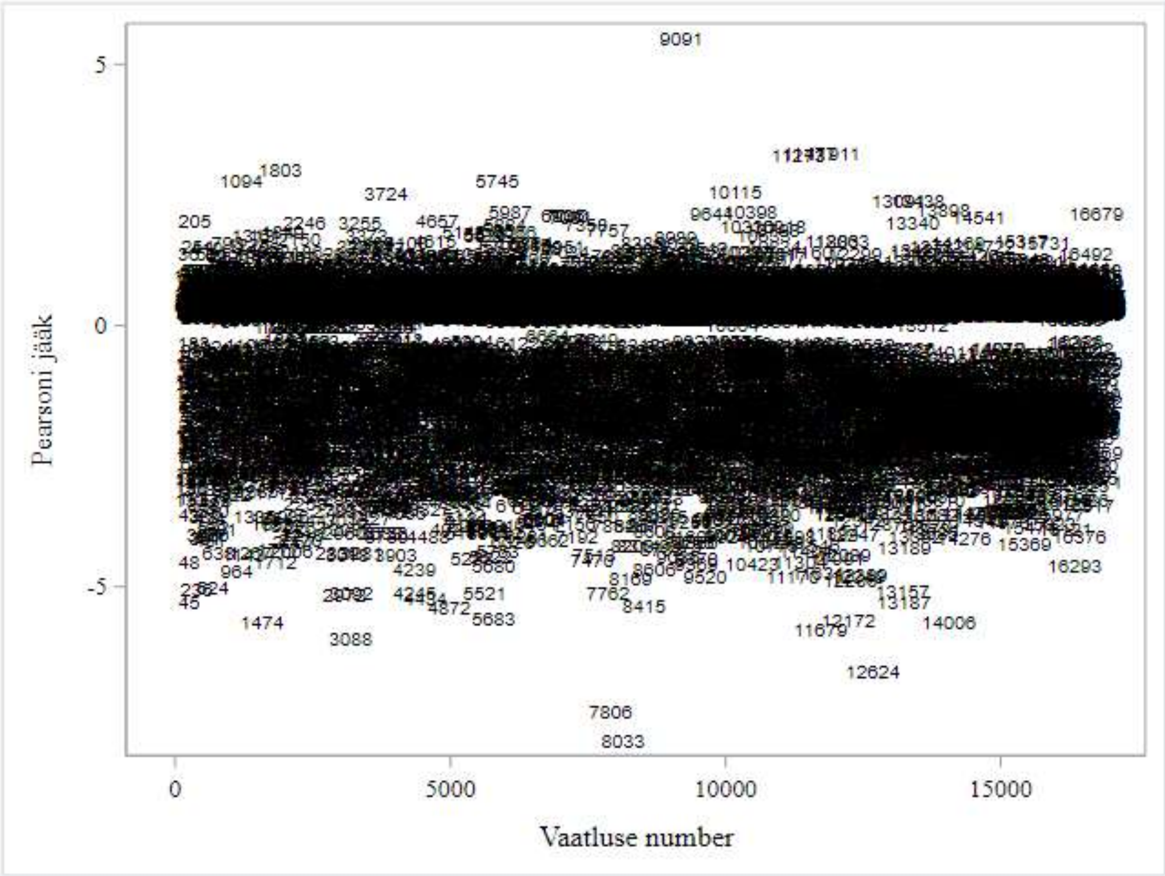
Tunnuste olulisuse testiga (*Type* 3 test) kontrollitakse, kas uuritav tunnus ja seletav tunnus on sõltuvad. Testist selgub (tabel 3.2), et olulisusnivool 0.05 on kõik mudelisse valitud parameetrid oluliselt nullist erinevad $(p < 0.05)$ ehk valitud seletavad tunnused kirjeldavad uuritavat tunnust.

Tabel 3.2: Tunnuste olulisuse test.

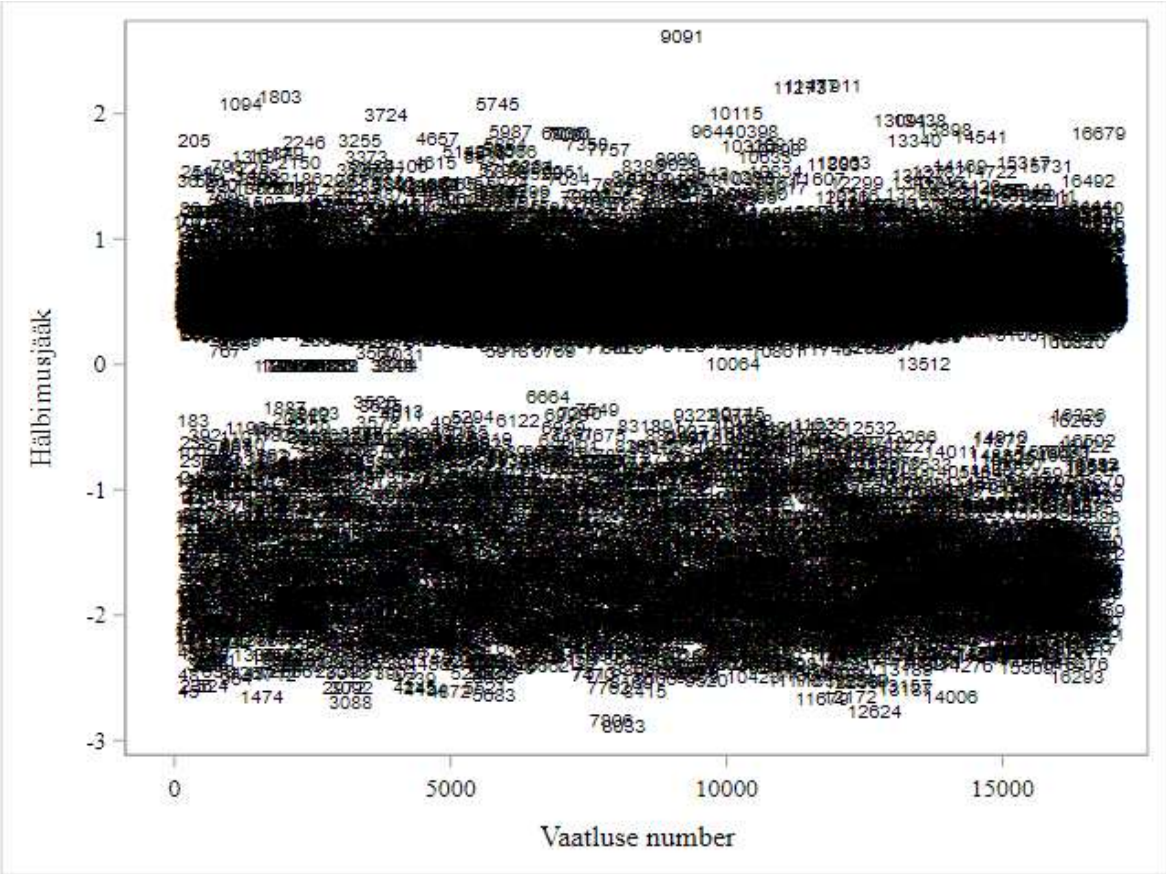
Tunnus	Vabadusastmete arv	Waldi statistik	P
HinnaMuutuseGrupp	22	1013.0851	<.0001
PreemiaGrupp	22	48.6683	0.0009
PoliisiKanal	2	92.0360	<.0001
KliendiVanuseGrupp	7	58.3267	<.0001
SõidukiVanuseGrupp	9	246.2311	<.0001
Regioon	19	164.1041	<.0001
BoonusMaalusKlass	13	29.6117	0.0054
SõidukiteArvuGrupp	3	13.0933	0.0044
KahjuOlemasolu	1	27.9495	<.0001
UuendusteArv	10	59.1342	<.0001

3.3 Jääkide analüüs

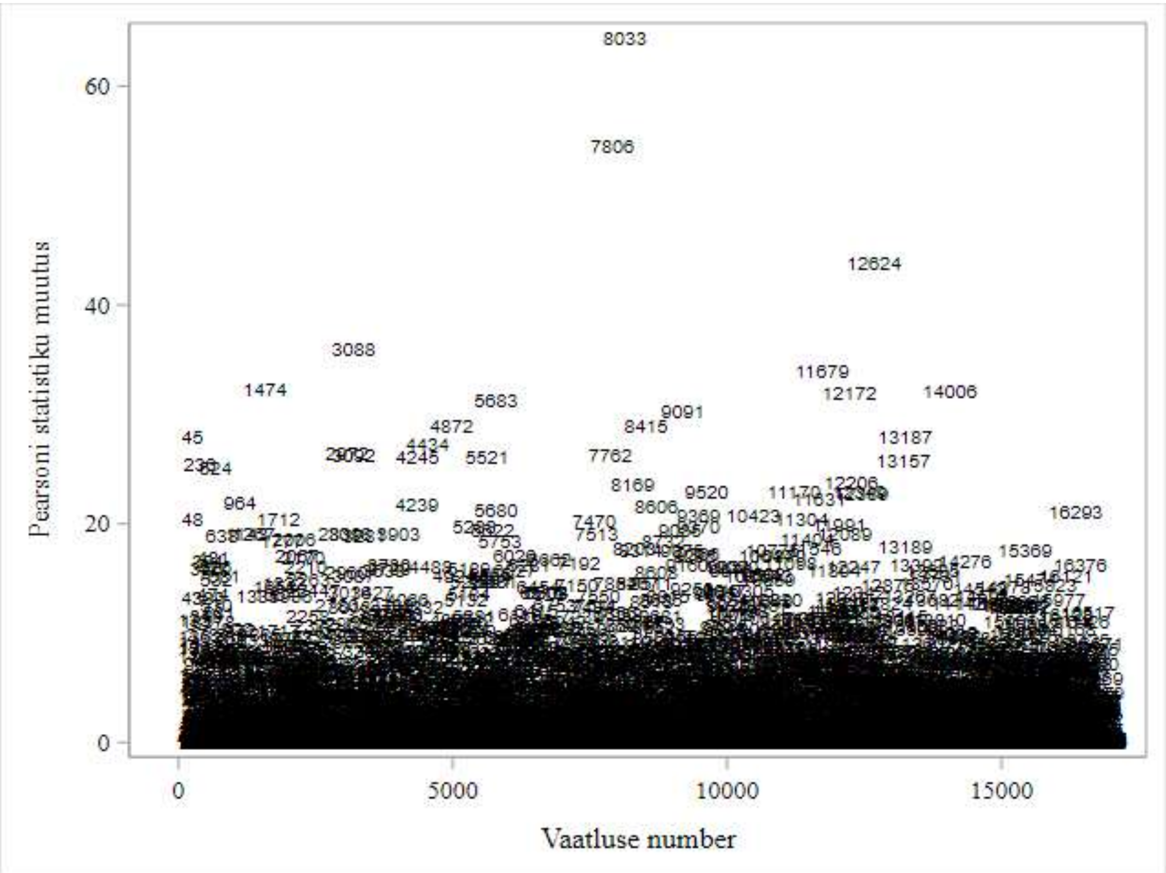
Mudeli koostamise üks olulisemaid samme on erindite väljaselgitamine. Erindite olemasolu võib moonutada uuritava tunnuse jaotust, tekitada ebausaldusväärseid parameetrite hinnanguid ja mõjutada mudeli sobivust. Iga vaatluse mõjususe hindamiseks vaadeldakse Pearsoni jääke (joonis 3.1) ja hälbimusjääke (joonis 3.2). Pearsoni jääkide ja hälbimusjääkide graafikutelt on näha, et vaatlused $(7806, 8033, 9091, 12624)$ ei sobi hästi mudeliga. Kuna Pearsoni jääkide ruutude võrdub Pearsoni χ^2-statistiku väärtusega ja hälbimusjääkide ruutude võrdub hälbimuse statistiku väärtusega, siis on antud erindite mõju oodata ka statistikutele. Vaatluse kustutamise mõju mudeli sobivusele Pearsoni χ^2-statistiku või hälbimuse statistiku järgi näitavad joonised 3.3 ja 3.4. Vaatluste $(7806, 8033, 9091, 12624)$ lähemal uurimisel ei selgunud, et oleks tegemist andmete kvaliteedi veaga või muu anomaaliaga, seega jäetakse nimetatud vaatlused andmestikku.



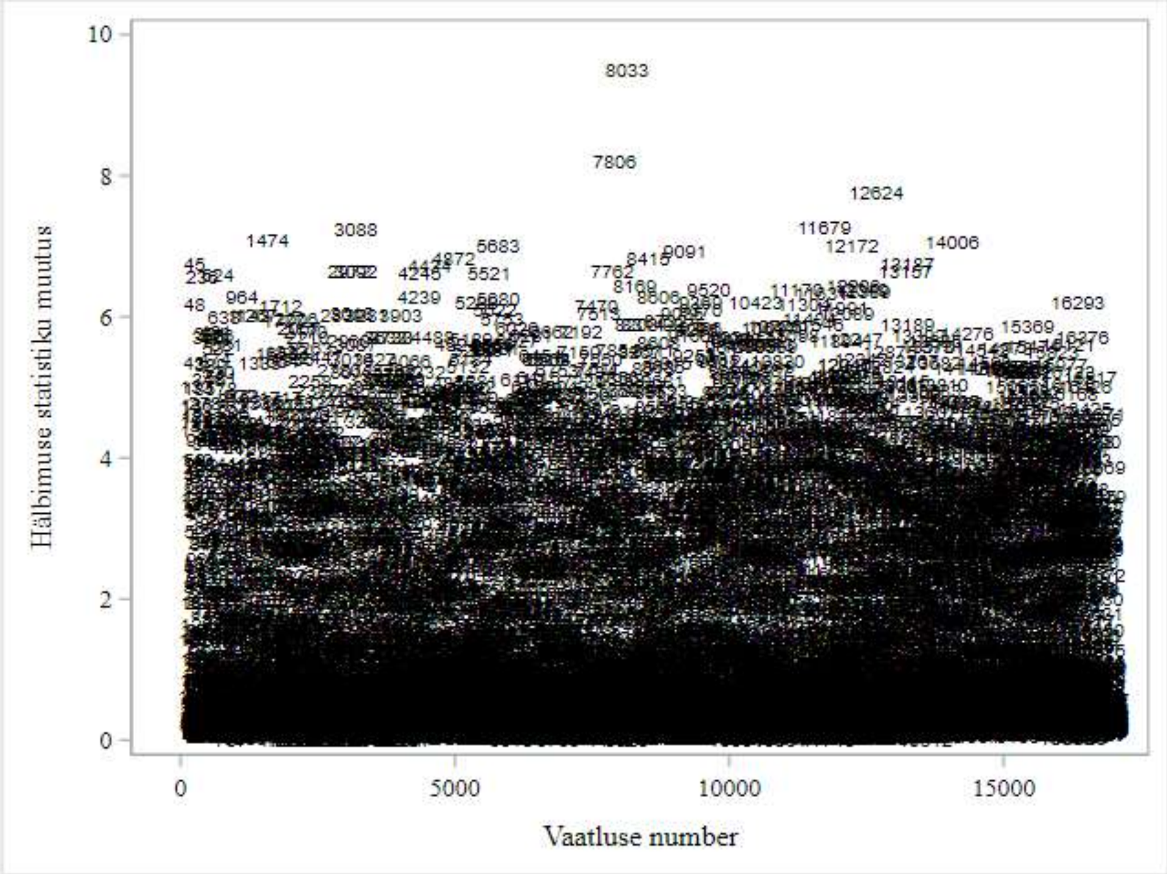
Joonis 3.1: Pearsoni jäägid.



Joonis 3.2: Hälbimusjäägid.



Joonis 3.3: Vaatluse mõju Pearsoni statistikule.

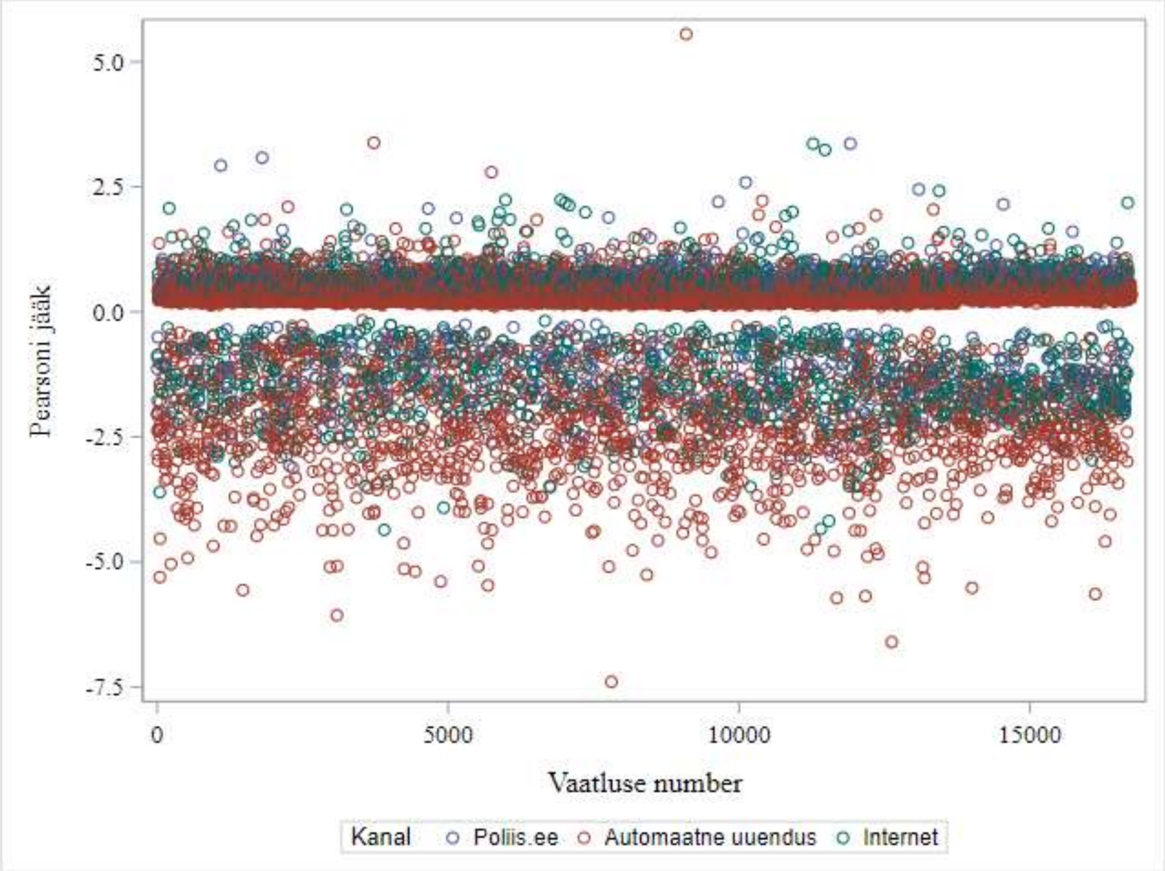


Joonis 3.4: Vaatluse mõju hälbimuse statistikule.

3.4 Argumentide sobivus

Vaatluste vähesuse tõttu grupeeritakse seletava tunnuse HinnaMuutuseGrupp tasemed „a. rohkem kui -50%“, „b. -45%-50%“, „c. -40%-45%“ tasemeks „abc. rohkem kui -40%“, seletava tunnuse PreemiaGrupp tasemed „s. 300-350 EUR“, „t. 350-400 EUR“, „u. 400-450 EUR“, „v. 450-500 EUR“, „w. üle 500 EUR“ tasemeks „stuvw. üle 300 EUR“, seletava tunnuse KliendiVanuseGrupp tasemed „a. alla 18 aasta“, „b. 18-25 aastat“ tasemeks „ab. alla 25 aasta“, seletava tunnuse BoonusMaalusKlass tasemed „B01“, „B02“, „B03“, „B04“, „B05“, „B06“, „B07“, „B08“, „B09“ tasemeks „B01-B09“ ja seletava tunnuse SõidukiteArvuGrupp tasemed „c. 5-20 sõidukit“, „d. üle 20 sõiduki“ tasemeks „cd. üle 5 sõiduki“.

Valitud argumentide sobivuse hindamiseks vaadatakse, kuidas Pearsoni jäägid jaotuvad argumentide tasemete vahel. Argumentide õigele valikule mudelis viitab see kui jäägid moodustavad hajusa pilve olenemata tunnuse tasemest ja jääkide skaala on ühesugune iga argumenttunnuse puhul. Joonisel 3.5 on näitena toodud Pearsoni jääkide jaotumine seletava tunnuse PoliisiKanal korral.



Joonis 3.5: Pearsoni jäägid poliisi kanali tasemete kaupa.

3.5 Mudeli sobivus

Praktikas hinnatakse mudeli sobivust mitme näitaja järgi ja siis otsustatakse, kas jäädakse mudeliga rahule. Kõigepealt vaadatakse Hosmer-Lemeshowi testi tulemust. Hosmer-Lemeshowi test põhineb tegelike ja mudeli prognoositud tõenäosuste hindamisel. Test jagab andmed kümnesse rühma vastavalt hinnatud tõenäosuste protsentiilidele. Tabelist 3.3 on näha, et tegelike ja hinnatud väärtuste vahe on väike.

Tabel 3.3: Hosmer-Lemeshowi testi tulemused.

Grupp	Vaatluste arv grupis	UuenduseIndikaator = 1		UuenduseIndikaator = 0	
		Tegelik	Prognoositud	Tegelik	Prognoositud
1	1674	762	743.93	912	930.07
2	1674	1152	1161.44	522	512.56
3	1674	1298	1294.86	376	379.14
4	1674	1367	1371.24	307	302.76
5	1674	1415	1424.87	259	249.13
6	1674	1451	1467.56	223	206.44
7	1675	1513	1501.57	162	173.43
8	1678	1541	1535.8	137	142.2
9	1674	1567	1563.35	107	110.65
10	1666	1593	1594.36	73	71.64

Ülaloleva tabeli 3.3 põhjal viiakse läbi χ^2-test ja kontrollitakse hüpoteese

$$\begin{cases} H_0: mudel\ on\ koosk\tilde{o}las\ andmetega, \\ H_1: mudel\ ei\ ole\ koosk\tilde{o}las\ andmetega. \end{cases}$$

Testist tulenevalt $p = 0.8266 > 0.05$ (tabel 3.4), mis lubab jääda nullhüpoteesi juurde, seega koostatud mudeli sobivus on hea.

Tabel 3.4: Hosmer-Lemeshowi mudeli sobivuse test.

Statistiku väärtus	Vabadusastmete arv	P
4.3259	8	0.8266

Veel vaadatakse tõepärafunktsioonil põhinevaid näitajaid Akaike informatsioonikriteerium AIC ja Schwarzi kriteerium SC.

Tabel 3.5: Mudeli kooskõlakriteeriumid.

Kriteerium	Ainult vabaliikmega mudel	Kõikide tunnustega mudel
-2LnL	15 976	13 716
AIC	15 978	13 912
SC	15 986	14 669

Statistiku -2lnL väärtust kasutatakse statistikute AIC ja SC väärtuse arvutamiseks. Tabelist 3.5 on näha, et statistikud AIC ja SC on seletavate tunnuste mudelisse lisamisel väiksemad kui ainult vabaliikmega mudeli puhul ehk seletavad tunnused täiendavad mudeli sobivust andmetega.

Mudeli headust hinnatakse veel Pearsoni statistiku abil. Testi tulemusena $p > 0.05$ (tabel 3.6), seega võib öelda, et mudel sobib andmetega.

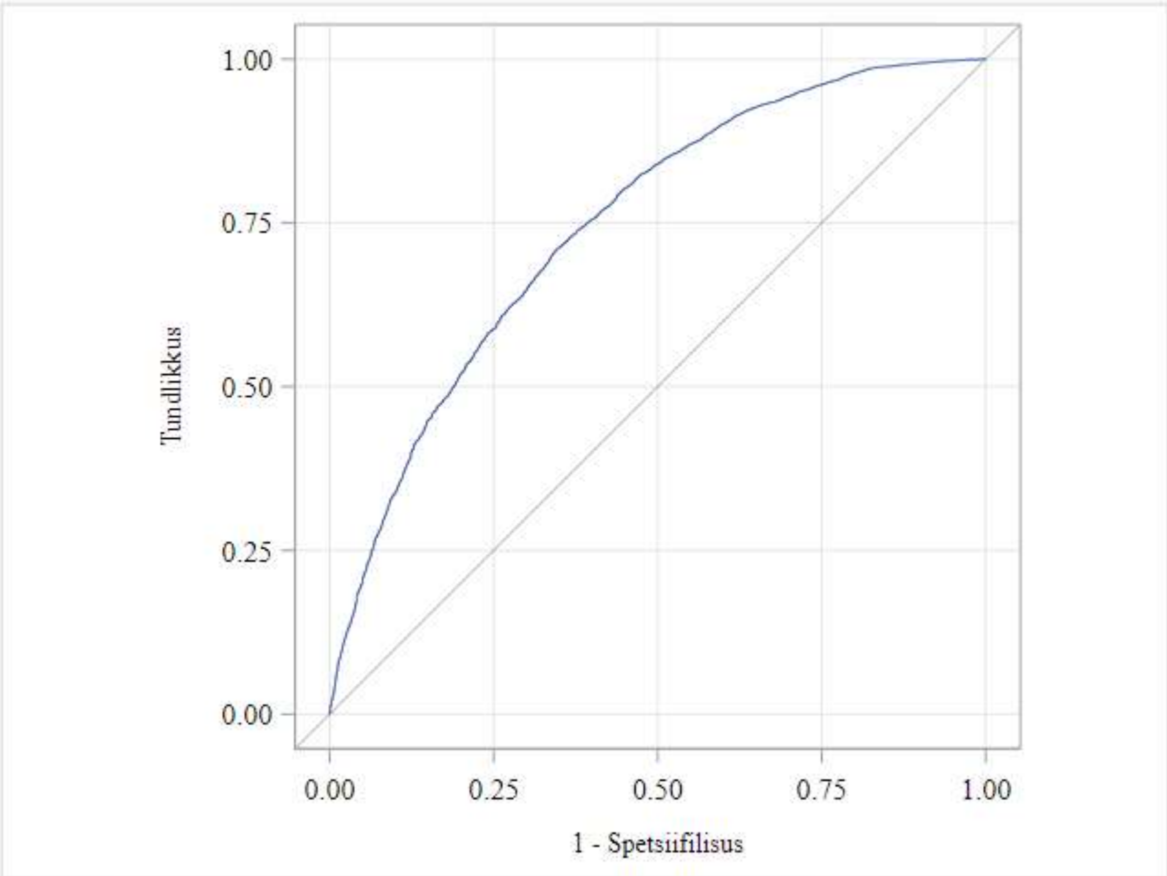
Tabel 3.6: Pearsoni sobivuse test.

Statistik	Väärtus	Vabadusastmete arv	P
Pearsoni χ^2- statistik	15 997	16 639	0.2037

Kokkuvõtlikult saab öelda, et koostatud mudeli sobivusega võib rahule jääda.

3.6 Mudeli prognoosivõime

ROC (*Receiver Operating Characteristic Curve*) kõverat kasutatakse mudeli prognoosivõime hindamiseks. Joonisel 3.6 kujutatud sirge tähistab lävi, mille järgi otsustatakse, kas mudeli poolt prognoositud tõenäosus klassifitseeritakse klassi $\{1\}$ või $\{0\}$. „Tundlikkus“ tähistab olukorda, kus poliisi uuendanud klient määratakse klassi $\{1\}$ ja „1 – Spetsiifilisus“ tähistab olukorda, kus poliisi mitte uuendanud kliendid määratakse siiski klassi $\{1\}$ ehk ROC kõver näitab kuidas antud lävi korral suudab mudel prognoosida uuritava tunnuse väärtust $\{1\}$.



Joonis 3.6: ROC kõver.

ROC graafiku aluse pindala AUC (*Area Under the Curve*) järgi hinnatakse prognoosivõimet: mida suurem on AUC, seda parem on mudeli prognoosivõime. Antud mudeli $\mathrm{AUC} = 0.747$, see tähendab, et mudeli prognoosivõimega võib rahule jääda.

Enne mudeli koostamist jagati andmestik treeningandmestikuks ja testandmestikuks. Tarkvara SAS protseduuri *store* (SAS Institute Inc.) abil prognoositakse uuendustõenäosusi valideerimisandmestikul kasutades selleks koostatud mudelit. Tulemustest, selgub, et valideerimisandmestikul ei ole AUC palju halvemaks läinud ($\mathrm{AUC} = 0.720$), seega mudel sobib ka testandmestikul ning võib öelda, et mudeli prognoosivõime on hea.

3.7 Alternatiivsed mudelid

3.7.1 Koosmõjudega logit mudel

Mudeli sobivust saab parandada lisades mudelisse seletavate tunnuste koosmõjusid. Huvi pakub eelkõige uurida, kas hinnamuutus mõjutab kliente mõne faktori tasemel erinevalt. Tarkvara SAS graafikutelt *interactions* (SAS Institute Inc.) ei ilmnenud, et tunnusel HinnaMuutuseGrupp oleks koosmõjusid teiste seletavate tunnustega. Seetõttu kasutatakse tarkvara SAS protseduuri *stepwise* (SAS Institute Inc.), et testida kõikide argumenttunnuste koosmõjude olulisust mudelis, lubades tunnustel mudelisse siseneda ja väljuda olulisusnivool $\alpha = 0.3$. Selle tulemusel lisati mudelisse tunnuse SõidukiteArvuGrupp koosmõju tunnustega PreemiaGrupp ja KahjuOlemasolu ning tunnuse PoliisiKanal koosmõju tunnusega UuendusteArv

Kokkuvõte *stepwise* protseduurist on toodud tabelis 3.7.

Tabel 3.7: Koosmõjude olulisuse hindamine.

Tunnus	Vabadusastmete arv	Waldi statistik	P
HinnaMuutuseGrupp	20	1002.3854	<.0001
PreemiaGrupp	18	38.6905	0.0031
PoliisiKanal	2	85.9075	<.0001
KliendiVanuseGrupp	7	55.9031	<.0001
SõidukiVanuseGrupp	9	253.1051	<.0001
Regioon	19	168.5444	<.0001
BoonusMaalusKlass	9	25.9570	0.0021
SõidukiteArvuGrupp	2	12.1871	0.0023
PreemiaGrupp*SõidukiteArvuGrupp	35	43.4544	0.1545
KahjuOlemasolu	1	30.1691	<.0001
SõidukiteArvuGrupp*KahjuOlemasolu	2	9.2592	0.0098
UuendusteArv	10	24.9258	0.0055
PoliisiKanal*UuendusteArv	20	44.9932	0.0011

Koosmõjudega logistilise regressiooni mudeli AUC on natuke paranenud võrreldes koosmõjudeta logistilise regressiooni mudeliga (AUC on 0.752 versus 0.747).

3.7.2 Probit mudel

Binaarse tunnuse modelleerimisel saab alternatiivse seosefunktsioonina kasutada *probit* seosefunktsiooni. Hosmer-Lemeshowi testi tulemusena (tabel 3.8) $p = 0.6642$, mis lubab jääda nullhüpoteesi juurde ehk ka probit mudel on kooskõlas andmetega.

Tabel 3.8: Hosmer-Lemeshowi mudeli sobivuse test.

Statistiku väärtus	Vabadusastmete arv	P
5.8489	8	0.6642

Probit mudeli AUC on väga sarnane logit mudeli AUC-le, mõlema mudeli AUC on ligikaudu 0.720. Samamoodi on sarnased valideerimisandmestikul kooskõlakriteeriumid AIC ja SC (tabel 3.9), seega ei saa öelda, et kummagi mudeli prognoosivõime oleks oluliselt parem.

Tabel 3.9: Mudeli kooskõlakriteeriumid valideerimisandmestikul.

Kriteerium	Logit mudel	Probit mudel
AIC	6 143	6 147
SC	6 816	6 819

3.7.3 Lõplik mudel

Mudeli headuse statistikute väärtustest (tabel 3.10) selgub, et logit mudelisse koosmõjude lisamisel AIC väheneb natuke ja SC suureneb. Probit mudeli AIC ja SC on natuke suuremad kui logit mudeli puhul, kuid vahe on väga väike. Kuna alternatiivsete mudelite kasutamine ei paranda mudeli headust ja soovitakse kasutada lihtsamini interpreteeritavat mudelit, siis jäädakse logistilise regressiooni mudeli juurde.

Tabel 3.10: Erinevate mudelite headuse statistikute väärtused.

Kriteerium	Logit mudel	Logit mudel koosmõjudega	Probit mudel
AIC	13 912	13 903	13 917
SC	14 669	15 101	14 674

Kokkuvõtlikult saab öelda, et logistilise regressiooni mudeli sobivusega võib rahule jääda, kuid mudeli headust ja prognoosvõimet saaks veel parandada leides põhjuslikke seoseid klientide karakteristikute ja uuendustõenäosuse vahel.

3.8 Mudeli interpretatsioon

Statistilisest analüüsist selgus, et seletavatel tunnustel HinnaMuutuseGrupp, PreemiaGrupp, PoliisiKanal, KliendiVanuseGrupp, SõidukiVanuseGrupp, Regioon, BoonusMaalusKlass, SõidukiteArvuGrupp, KahjuOlemasolu ja UuendusteArv on mõju uuendustõenäosusele.

Iga faktori puhul määratakse baastase, selleks valitakse argumenttunnuse suurima vaatluste arvuga tase. Tarkvara SAS määrab vaikimisi baastasemeks faktori viimase taseme, kuid kui vaatluseid antud tasemel on vähe, siis parameetrite hinnangud ei ole usaldusväärsed. Seletava tunnuse HinnaMuutuseGrupp baasatasemeks on „j. -5%-10%“, tunnuse PreemiaGrupp baastasemeks on „c. 60-70 EUR“, tunnuse PoliisiKanal baastasemeks on „Automaatne uuendus“, tunnuse KliendiVanuseGrupp baastasemeks on „g. 51-60 aastat“, tunnuse SõidukiVanuseGrupp baastasemeks on „i. 11-15 aastat“, tunnuse Regioon baastasemeks on „Tallinn“, tunnuse BoonusMaalusKlass baastasemeks on „B18“, tunnuse SõidukiteArvuGrupp baastasemeks on „a. 1 sõiduk“, tunnuse KahjuOlemasolu baastasemeks on „1“ ja tunnuse UuendusteArv baastasemeks on „0“.

Lõplik logistilise regressiooni mudel koos kõikide mudeli parameetritega on leitav lisast 1. Mudeli tulemuste interpreteerimisel kasutatakse lisas 2 esitatud šansside suhteid.

Üldistavalt võib öelda, et kliendid, kelle poliisi hind langeb, uuendavad suurema tõenäosusega kui kliendid, kelle poliisi hind tõuseb. Klientidel, kelle uuenduspoliisi hind ei muutu on 34.5% väiksem šanss poliisi uuendada võrreldes klientidega, kelle poliisi hind on $-5\% - 10\%$ soodsam. Kui kliendi poliisi hind tõuseb $+5\% + 10\%$, siis nende klientide šanss uuendada on 35.9% väiksem kui klientidel, kellel hind langeb $-5\% - 10\%$.

Mida väiksem on uuendatava poliisi hind, seda suurem šanss on, et klient jätkab kindlustamist IF-is.

Internetikanalites on uuendamise šanss väiksem kui automaatse uuenduse kanalis: internetis on šanss 32.4% väiksem ja Poliis.ee-s 51.9% väiksem.

Üldkogumis ei suuda kinnitada, et alla 25-aastaste klientide uuenduskäitumine oleks teistsugune, kuid saab öelda, et $26 - 30$-aastaste klientide šanss uuendamiseks on kõige väiksem. Keskealised kliendid (vanusega $31 - 50$) uuendavad kõik ligikaudu sama tõenäosusega. Kõige suurema tõenäosusega uuendavad vanurid (üle 70-aastased kliendid).

Uuemate sõidukite šanss IF-is kindlustamist jätkata on suurem kui vanematel sõidukitel. Kõige väiksem šanss uuendamiseks on üle 15-aastastel sõidukitel, võrreldes $11 - 15$-aastaste sõidukitega on uuendamise šanss 43.6% väiksem.

Võrreldes Tallinnaga on uuenduse käitumine erinev ainult Ida-Virumaal, Narvas ja Võrumaal. Ida-Virumaal on uuendamise šanss 36.4%, Narvas 72.6% ja Võrumaal 50.9% väiksem kui Tallinnas.

Üldkogumis on näha boonus-maalus klassi mõju uuendamisele ainult boonus-maalus klasside „B14“, „B16“ ja „B17“ puhul: kliendid, kes on nendes klassides uuendavad väiksema šansiga kui kliendid, kes on boonus-maalus klassis „B18“. Ühegi teise boonus-maalus klassi vahel statistiliselt olulist erinevust ei leidunud. Samas kahjuga klientide uuendamise šanss on 90.2% kõrgem kui kahjudeta klientidel.

Klientide puhul, kes omavad üle viie sõiduki ei saa öelda, et on erinevusi uuendamise šansis võrreldes klientidega, kellel on ainult üks sõiduk. Kuid kliendid, kes omavad $2 - 4$ sõidukit jätkavad suurema šansiga kindlustamist IF-is kui kliendid, kellel on üks sõiduk.

Üldiselt võib öelda, et esimeste uuenduste puhul (kuni viienda uuenduseni) šanss uuendada tõuseb iga järgneva uuendusega ja pikaajalisemad kliendid on lojaalsemad IF-ile kui uuemad kliendid.

Kokkuvõte

Magistritöö eesmärgiks oli aru saada, mis mõjutab liikluskindlustuse kliendi otsust jätkata samas kindlustusseltsis. Uuritavaks tunnuseks oli Bernoulli jaotusega binaarne juhuslik suurus. Analüüsiks kasutati *If P&C Insurance AS* liikluskindlustuse portfelli andmeid.

Binaarse tunnuse modelleerimiseks on mitmeid võimalusi, antud magistritöö raames kasutati logistilise regressiooni mudelit, et tulemusi saaks otseselt rakendada kindlustustegevuses. Logistiline regressioon kuulub üldistatud lineaarsete mudelite klassi, seega on enamus tuntud võtteid rakendatavad ka uuritud mudeli korral. Modelleerimise protsess on pikk, antud tööga anti ülevaade mudeli koostamise erinevatest sammudest, üritades samal ajal viia statistikatarkvara väljundid kokku matemaatilise teooriaga.

Juba töö alguses püstitati hüpoteese, mida sooviti kontrollida. Analüüsist tuli välja, et uuendustõenäosus sõltub uuenduspoliisi hinnamuutusest, eelmise poliisi hinnast, poliisi ostmise kanalist, kliendi ja sõiduki vanusest, kliendi elamispiirkonnast, riskiklassist ja sellest, kas klient on põhjustanud õnnetuse või mitte, kliendi omandis olevatest sõidukite arvust ning sellest kui mitu korda varem on klient uuendanud poliisi IF-is.

Selge mõju kliendi käitumisele on uuenduspoliisi hinnamuutusel. Hinnalangetuse korral jäävad kliendid suurema šansiga IF-i ja hinnatõusu korral otsitakse turult paremat pakkumist. Mudeli järgi on klientidele olulisem see kui palju muutub hind protsentuaalselt kui see kui palju muutub hind eurodes. Kõrgema poliisi preemiaga kliendid uuendavad halvemini kui odavama poliisi hinnaga kliendid. Nii nagu arvati, siis internetikanalite klientide hoidmine on keerulisem kui automaatse uuenduse kanali klientide hoidmine. Samamoodi on noorte klientide uuendamine väiksema šansiga võrreldes vanemate klientidega, kes on väga lojaalsed. Tõeseks osutusid ka hüpoteesid, et uuemate sõidukite omanike šanss uuendada on kõrge ja uuendustõenäosus sõltub kliendi elamispiirkonnast. Selgus ka, et IF-is pikka aega kindlustanud kliendid jätkavad IF-is kindlustamist suure tõenäosusega ilmselt ka edaspidi. Uuendustõenäosus sõltub ka kliendi riskiklassist, kuid üllatav oli see, et liiklusõnnetuse põhjustanud kliendid jätkavad IF-is kindlustamist suurema šansiga kui kahjudeta kliendid. Vastupidiselt püstitatud hüpoteesile uuendavad $2-4$ sõiduki omanikud suurema tõenäosusega kui ühe sõidukiga omanikud ja uuendustõenäosus ei sõltu sellistest sõiduki omadustest nagu sõiduki võimsus ja kasutusala.

Klientide käitumisest arusaamine on keeruline teema, magistritöös osutusid uuendustõenäosuse kirjeldamisel oluliseks mitmed tunnused, kuid mudeli headuse statistikute järgi on näha, et osa põhjuslikke seoseid on veel leidmata. Magistritöös analüüsiti saadaolevaid andmeid, kuid arvestades uuritava tunnuse iseloomu, siis võiksid sobida veel mitmed erinevad kliendi käitumist seletavad tunnused (näiteks haridustase, palgatase, laste arv peres jne).

Kasutatud kirjandus

[1] **Agresti, A** (2013). *Categorical Data Analysis*. Wiley, Hoboken.

[2] **Gernard, T** (2012). *Regression for Categorical Data*. Cambridge University Press, Cambridge.

[3] **Möls, M**. (2011). *Üldistatud lineaarsed mudelid*. Loengukonspekt. Tartu: Tartu Ülikool, matemaatika ja statistika instituut.

[4] **SAS Institute Inc.**. *SAS/STAT 15.1 User's Guide*. [WWW]

[https://documentation.sas.com].

[5] **Traat, I.** (2018). *Kvalitatiivsete andmete analüüs*. Loengukonspekt. Tartu: Tartu Ülikool, matemaatika ja statistika instituut.

Lisa 1 Mudel

Tunnus	Tase	Para-meeter	Viga	Waldi statistik	P
Intercept		2.0116	0.1450	192.564	<.0001
HinnaMuutuseGrupp	**abc. rohkem kui -40%**	0.1957	0.2461	0.6327	0.4264
HinnaMuutuseGrupp	**d. -35%-40%**	1.6573	0.3995	17.2066	<.0001
HinnaMuutuseGrupp	**e. -30%-35%**	0.8469	0.2244	14.2426	0.0002
HinnaMuutuseGrupp	**f. -25%-30%**	0.9386	0.1897	24.4820	<.0001
HinnaMuutuseGrupp	**g. -20%-25%**	0.5846	0.1251	21.8283	<.0001
HinnaMuutuseGrupp	**h. -15%-20%**	0.4757	0.0986	23.2772	<.0001
HinnaMuutuseGrupp	**i. -10%-15%**	0.3434	0.0895	14.7258	0.0001
HinnaMuutuseGrupp	**k. -1%-5%**	-0.2285	0.0778	8.6325	0.0033
HinnaMuutuseGrupp	**l. sama hind**	-0.4225	0.1043	16.4156	<.0001
HinnaMuutuseGrupp	**m. +1%+5%**	-0.3866	0.0942	16.8266	<.0001
HinnaMuutuseGrupp	**n. +5%+10%**	-0.4442	0.1015	19.1397	<.0001
HinnaMuutuseGrupp	**o. +10%+15%**	-0.8362	0.1218	47.1040	<.0001
HinnaMuutuseGrupp	**p. +15%+20%**	-1.0854	0.1420	58.3903	<.0001
HinnaMuutuseGrupp	**q. +20%+25%**	-1.3082	0.1544	71.8216	<.0001
HinnaMuutuseGrupp	**r. +25%+30%**	-1.0959	0.1560	49.3662	<.0001
HinnaMuutuseGrupp	**s. +30%+35%**	-1.2135	0.1705	50.6474	<.0001
HinnaMuutuseGrupp	**t. +35%+40%**	-1.5737	0.1954	64.8640	<.0001
HinnaMuutuseGrupp	**u. +40%+45%**	-1.7776	0.2277	60.9588	<.0001
HinnaMuutuseGrupp	**v. +45%+50%**	-2.1231	0.2526	70.6376	<.0001
HinnaMuutuseGrupp	**w. rohkem kui +50%**	-2.5611	0.1142	502.5535	<.0001
PreemiaGrupp	**a. alla 50 EUR**	0.3173	0.1355	5.4803	0.0192
PreemiaGrupp	**b. 50-60 EUR**	0.1794	0.0959	3.5019	0.0613
PreemiaGrupp	**d. 70-80 EUR**	0.0623	0.0950	0.4305	0.5118
PreemiaGrupp	**e. 80-90 EUR**	-0.0039	0.1045	0.0014	0.9699
PreemiaGrupp	**f. 90-100 EUR**	-0.1361	0.1117	1.4845	0.2231
PreemiaGrupp	**g. 100-110 EUR**	-0.2693	0.1153	5.4569	0.0195
PreemiaGrupp	**h. 110-120 EUR**	-0.1603	0.1296	1.5286	0.2163
PreemiaGrupp	**i. 120-130 EUR**	-0.2577	0.1331	3.7506	0.0528
PreemiaGrupp	**j. 130-140 EUR**	-0.3956	0.1509	6.8756	0.0087
PreemiaGrupp	**k. 140-150 EUR**	-0.4273	0.1614	7.0126	0.0081
PreemiaGrupp	**l. 150-160 EUR**	-0.4516	0.1678	7.2455	0.0071
PreemiaGrupp	**m. 160-170 EUR**	-0.4228	0.1976	4.5793	0.0324
PreemiaGrupp	**n. 170-180 EUR**	-0.3336	0.2242	2.2132	0.1368
PreemiaGrupp	**o. 180-190 EUR**	-0.1312	0.2520	0.2711	0.6026
PreemiaGrupp	**p. 190-200 EUR**	-0.8987	0.2593	12.0112	0.0005
PreemiaGrupp	**q. 200-250 EUR**	-0.7884	0.2078	14.3916	0.0001
PreemiaGrupp	**r. 250-300 EUR**	-1.3293	0.2851	21.7479	<.0001
PreemiaGrupp	**stuvw. üle 300 EUR**	-0.9941	0.3550	7.8426	0.0051
PoliisiKanal	**Internet**	-0.3922	0.0634	38.2803	<.0001

PoliisiKanal	**Poliis.ee**	-0.7326	0.0813	81.2125	<.0001
KliendiVanuseGrupp	**ab. alla 25 aasta**	-0.2892	0.2671	1.1721	0.2790
KliendiVanuseGrupp	**c. 26-30 aastat**	-0.4731	0.1020	21.5116	<.0001
KliendiVanuseGrupp	**d. 31-35 aastat**	-0.2173	0.0885	6.0325	0.0140
KliendiVanuseGrupp	**e. 36-40 aastat**	-0.2215	0.0842	6.9311	0.0085
KliendiVanuseGrupp	**f. 41-50 aastat**	-0.2148	0.0677	10.0779	0.0015
KliendiVanuseGrupp	**h. 61-70 aastat**	0.0789	0.0720	1.2012	0.2731
KliendiVanuseGrupp	**i. üle 70 aasta**	0.1526	0.0832	3.3661	0.0666
SõidukiVanuseGrupp	**a. uus sõiduk**	0.5134	0.2103	5.9602	0.0146
SõidukiVanuseGrupp	**b. 1 aastat**	0.2923	0.1413	4.2777	0.0386
SõidukiVanuseGrupp	**c. 2 aastat**	0.7323	0.1525	23.0743	<.0001
SõidukiVanuseGrupp	**d. 3 aastat**	0.3072	0.1266	5.8894	0.0152
SõidukiVanuseGrupp	**e. 4 aastat**	0.5432	0.1351	16.1611	<.0001
SõidukiVanuseGrupp	**f. 5 aastat**	0.3632	0.1222	8.8331	0.0030
SõidukiVanuseGrupp	**g. 6-7 aastat**	0.3049	0.0856	12.7020	0.0004
SõidukiVanuseGrupp	**h. 8-10 aastat**	0.2734	0.0702	15.1502	<.0001
SõidukiVanuseGrupp	**j. üle 15 aasta**	-0.5719	0.0573	99.7232	<.0001
Regioon	**Harjumaa**	-0.1371	0.0808	2.8808	0.0896
Regioon	**Hiiumaa**	-0.1162	0.2236	0.2699	0.6034
Regioon	**Ida-Virumaa**	-0.4523	0.1251	13.0591	0.0003
Regioon	**Järvamaa**	0.0230	0.1534	0.0224	0.8810
Regioon	**Jõgevamaa**	0.1811	0.1591	1.2953	0.2551
Regioon	**Kohtla-Järve**	0.0719	0.3399	0.0448	0.8324
Regioon	**Lääne-Virumaa**	-0.2058	0.1289	2.5460	0.1106
Regioon	**Läänemaa**	-0.1904	0.1635	1.3551	0.2444
Regioon	**Narva**	-1.2958	0.1344	92.9441	<.0001
Regioon	**Pärnu**	0.0376	0.1993	0.0357	0.8502
Regioon	**Pärnumaa**	-0.2001	0.1518	1.7373	0.1875
Regioon	**Põlvamaa**	-0.2169	0.1668	1.6918	0.1934
Regioon	**Raplamaa**	-0.1748	0.1451	1.4516	0.2283
Regioon	**Saaremaa**	-0.1103	0.1497	0.5435	0.4610
Regioon	**Tartu**	0.0621	0.1133	0.3006	0.5835
Regioon	**Tartumaa**	-0.1795	0.1208	2.2070	0.1374
Regioon	**Valgamaa**	-0.0967	0.1525	0.4018	0.5262
Regioon	**Viljandimaa**	-0.2581	0.1528	2.8541	0.0911
Regioon	**Võrumaa**	-0.7111	0.1300	29.9006	<.0001
BoonusMaalusKlass	**B01-B09**	-0.0932	0.5055	0.0340	0.8537
BoonusMaalusKlass	**B10**	-0.1159	0.2561	0.2049	0.6508
BoonusMaalusKlass	**B11**	0.0317	0.2258	0.0197	0.8883
BoonusMaalusKlass	**B12**	0.00114	0.1924	0	0.9953
BoonusMaalusKlass	**B13**	-0.1809	0.1339	1.8257	0.1766
BoonusMaalusKlass	**B14**	-0.3355	0.1194	7.9030	0.0049
BoonusMaalusKlass	**B15**	-0.1424	0.1140	1.5594	0.2118
BoonusMaalusKlass	**B16**	-0.3696	0.0999	13.6923	0.0002
BoonusMaalusKlass	**B17**	-0.2222	0.0879	6.3918	0.0115
SõidukiteArvuGrupp	**b. 2-4 sõidukit**	0.1939	0.0530	13.3866	0.0003
SõidukiteArvuGrupp	**cd. üle 5 sõiduki**	-0.2654	0.1909	1.9335	0.1644
KahjuOlemasolu	**1**	0.6430	0.1179	29.7233	<.0001

UuendusteArv	**1**	0.1526	0.0813	3.5257	0.0604
UuendusteArv	**2**	0.2393	0.0924	6.7081	0.0096
UuendusteArv	**3**	0.2850	0.0901	10.0145	0.0016
UuendusteArv	**4**	0.4003	0.0961	17.3680	<.0001
UuendusteArv	**5**	0.4608	0.1057	18.9959	<.0001
UuendusteArv	**6**	0.6417	0.1171	30.0511	<.0001
UuendusteArv	**7**	0.5324	0.1265	17.7007	<.0001
UuendusteArv	**8**	0.4556	0.1487	9.3835	0.0022
UuendusteArv	**9**	0.5492	0.0951	33.3368	<.0001
UuendusteArv	**10**	0.4296	0.1280	11.2705	0.0008

Lisa 2 Šansside suhted

Tunnuste tasemed	**Väär-tus**	**95% Waldi usaldusvahemik**	
HinnaMuutuseGrupp abc. rohkem kui -40% vs j. -5%-10%	1.216	0.751	1.970
HinnaMuutuseGrupp d. -35%-40% vs j. -5%-10%	5.245	2.397	11.477
HinnaMuutuseGrupp e. -30%-35% vs j. -5%-10%	2.332	1.502	3.621
HinnaMuutuseGrupp f. -25%-30% vs j. -5%-10%	2.556	1.763	3.708
HinnaMuutuseGrupp g. -20%-25% vs j. -5%-10%	1.794	1.404	2.293
HinnaMuutuseGrupp h. -15%-20% vs j. -5%-10%	1.609	1.326	1.952
HinnaMuutuseGrupp i. -10%-15% vs j. -5%-10%	1.410	1.183	1.680
HinnaMuutuseGrupp k. -1%-5% vs j. -5%-10%	0.796	0.683	0.927
HinnaMuutuseGrupp l. sama hind vs j. -5%-10%	0.655	0.534	0.804
HinnaMuutuseGrupp m. +1%+5% vs j. -5%-10%	0.679	0.565	0.817
HinnaMuutuseGrupp n. +5%+10% vs j. -5%-10%	0.641	0.526	0.783
HinnaMuutuseGrupp o. +10%+15% vs j. -5%-10%	0.433	0.341	0.550
HinnaMuutuseGrupp p. +15%+20% vs j. -5%-10%	0.338	0.256	0.446
HinnaMuutuseGrupp q. +20%+25% vs j. -5%-10%	0.270	0.200	0.366
HinnaMuutuseGrupp r. +25%+30% vs j. -5%-10%	0.334	0.246	0.454
HinnaMuutuseGrupp s. +30%+35% vs j. -5%-10%	0.297	0.213	0.415
HinnaMuutuseGrupp t. +35%+40% vs j. -5%-10%	0.207	0.141	0.304
HinnaMuutuseGrupp u. +40%+45% vs j. -5%-10%	0.169	0.108	0.264
HinnaMuutuseGrupp v. +45%+50% vs j. -5%-10%	0.120	0.073	0.196
HinnaMuutuseGrupp w. rohkem kui +50% vs j. -5%-10%	0.077	0.062	0.097
PreemiaGrupp a. alla 50 EUR vs c. 60-70 EUR	1.373	1.053	1.791
PreemiaGrupp b. 50-60 EUR vs c. 60-70 EUR	1.197	0.992	1.444
PreemiaGrupp d. 70-80 EUR vs c. 60-70 EUR	1.064	0.884	1.282
PreemiaGrupp e. 80-90 EUR vs c. 60-70 EUR	0.996	0.812	1.223
PreemiaGrupp f. 90-100 EUR vs c. 60-70 EUR	0.873	0.701	1.086
PreemiaGrupp g. 100-110 EUR vs c. 60-70 EUR	0.764	0.609	0.958
PreemiaGrupp h. 110-120 EUR vs c. 60-70 EUR	0.852	0.661	1.098
PreemiaGrupp i. 120-130 EUR vs c. 60-70 EUR	0.773	0.595	1.003
PreemiaGrupp j. 130-140 EUR vs c. 60-70 EUR	0.673	0.501	0.905
PreemiaGrupp k. 140-150 EUR vs c. 60-70 EUR	0.652	0.475	0.895
PreemiaGrupp l. 150-160 EUR vs c. 60-70 EUR	0.637	0.458	0.884
PreemiaGrupp m. 160-170 EUR vs c. 60-70 EUR	0.655	0.445	0.965
PreemiaGrupp n. 170-180 EUR vs c. 60-70 EUR	0.716	0.462	1.112
PreemiaGrupp o. 180-190 EUR vs c. 60-70 EUR	0.877	0.535	1.437
PreemiaGrupp p. 190-200 EUR vs c. 60-70 EUR	0.407	0.245	0.677
PreemiaGrupp q. 200-250 EUR vs c. 60-70 EUR	0.455	0.302	0.683
PreemiaGrupp r. 250-300 EUR vs c. 60-70 EUR	0.265	0.151	0.463
PreemiaGrupp stuvw. üle 300 EUR vs c. 60-70 EUR	0.370	0.185	0.742
PoliisiKanal Internet vs Automaatne uuendus	0.676	0.597	0.765
PoliisiKanal Poliis.ee vs Automaatne uuendus	0.481	0.410	0.564

KliendiVanuseGrupp ab. alla 25 aasta vs g. 51-60 aastat	0.749	0.444	1.264
KliendiVanuseGrupp c. 26-30 aastat vs g. 51-60 aastat	0.623	0.510	0.761
KliendiVanuseGrupp d. 31-35 aastat vs g. 51-60 aastat	0.805	0.677	0.957
KliendiVanuseGrupp e. 36-40 aastat vs g. 51-60 aastat	0.801	0.679	0.945
KliendiVanuseGrupp f. 41-50 aastat vs g. 51-60 aastat	0.807	0.706	0.921
KliendiVanuseGrupp h. 61-70 aastat vs g. 51-60 aastat	1.082	0.940	1.246
KliendiVanuseGrupp i. üle 70 aasta vs g. 51-60 aastat	1.165	0.990	1.371
SõidukiVanuseGrupp a. uus sõiduk vs i. 11-15 aastat	1.671	1.107	2.523
SõidukiVanuseGrupp b. 1 aastat vs i. 11-15 aastat	1.340	1.015	1.767
SõidukiVanuseGrupp c. 2 aastat vs i. 11-15 aastat	2.080	1.543	2.804
SõidukiVanuseGrupp d. 3 aastat vs i. 11-15 aastat	1.360	1.061	1.743
SõidukiVanuseGrupp e. 4 aastat vs i. 11-15 aastat	1.722	1.321	2.244
SõidukiVanuseGrupp f. 5 aastat vs i. 11-15 aastat	1.438	1.132	1.827
SõidukiVanuseGrupp g. 6-7 aastat vs i. 11-15 aastat	1.357	1.147	1.604
SõidukiVanuseGrupp h. 8-10 aastat vs i. 11-15 aastat	1.314	1.145	1.508
SõidukiVanuseGrupp j. üle 15 aasta vs i. 11-15 aastat	0.564	0.504	0.631
Regioon Harjumaa vs Tallinn	0.872	0.744	1.021
Regioon Hiiumaa vs Tallinn	0.890	0.574	1.38
Regioon Ida-Virumaa vs Tallinn	0.636	0.498	0.813
Regioon Järvamaa vs Tallinn	1.023	0.758	1.382
Regioon Jõgevamaa vs Tallinn	1.199	0.877	1.637
Regioon Kohtla-Järve vs Tallinn	1.075	0.552	2.092
Regioon Lääne-Virumaa vs Tallinn	0.814	0.632	1.048
Regioon Läänemaa vs Tallinn	0.827	0.600	1.139
Regioon Narva vs Tallinn	0.274	0.210	0.356
Regioon Pärnu vs Tallinn	1.038	0.703	1.535
Regioon Pärnumaa vs Tallinn	0.819	0.608	1.102
Regioon Põlvamaa vs Tallinn	0.805	0.581	1.116
Regioon Raplamaa vs Tallinn	0.840	0.632	1.116
Regioon Saaremaa vs Tallinn	0.896	0.668	1.201
Regioon Tartu vs Tallinn	1.064	0.852	1.329
Regioon Tartumaa vs Tallinn	0.836	0.660	1.059
Regioon Valgamaa vs Tallinn	0.908	0.673	1.224
Regioon Viljandimaa vs Tallinn	0.773	0.573	1.042
Regioon Võrumaa vs Tallinn	0.491	0.381	0.634
BoonusMaalusKlass B01-B09 vs B18	0.911	0.338	2.454
BoonusMaalusKlass B10 vs B18	0.891	0.539	1.471
BoonusMaalusKlass B11 vs B18	1.032	0.663	1.607
BoonusMaalusKlass B12 vs B18	1.001	0.687	1.460
BoonusMaalusKlass B13 vs B18	0.834	0.642	1.085
BoonusMaalusKlass B14 vs B18	0.715	0.566	0.903
BoonusMaalusKlass B15 vs B18	0.867	0.694	1.084
BoonusMaalusKlass B16 vs B18	0.691	0.568	0.840
BoonusMaalusKlass B17 vs B18	0.801	0.674	0.951
SõidukiteArvuGrupp b. 2-4 sõidukit vs a. 1 sõiduk	1.214	1.094	1.347
SõidukiteArvuGrupp cd. üle 5 sõiduki vs a. 1 sõiduk	0.767	0.528	1.115
KahjuOlemasolu 1 vs 0	1.902	1.510	2.397

UuendusteArv 1 vs 0	1.165	0.993	1.366
UuendusteArv 2 vs 0	1.270	1.060	1.523
UuendusteArv 3 vs 0	1.330	1.115	1.586
UuendusteArv 4 vs 0	1.492	1.236	1.801
UuendusteArv 5 vs 0	1.585	1.289	1.950
UuendusteArv 6 vs 0	1.900	1.510	2.390
UuendusteArv 7 vs 0	1.703	1.329	2.182
UuendusteArv 8 vs 0	1.577	1.178	2.111
UuendusteArv 9 vs 0	1.732	1.437	2.087
UuendusteArv 10 vs 0	1.537	1.196	1.975

Lihtlitsents lõputöö reprodutseerimiseks ja üldsusele kättesaadavaks tegemiseks

Mina, Kadi Kitsing

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

   „Poliisi uuendustõenäosuse sõltuvus erinevatest faktoritest",

   mille juhendaja on Tõnu Kollo,

   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, alates 12.06.2023 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.

Tartus, 25.05.2020