

Liikenneonnettomuuksien määrä suhteessa liikennevirtaan valtateillä

Tilastotieteen pro gradu -tutkielma

Niko Into

Jyväskylän yliopisto
Matematiikan ja tilastotieteen laitos
Heinäkuu 2023

JYVÄSKYLÄN YLIOPISTO, Matematiikan ja tilastotieteen laitos

Niko Into

Liikenneonnettomuuksien määrä suhteessa liikennevirtaan valtateilla

Tilastotieteen pro gradu -tutkielma

36 sivua

Heinäkuu 2023

Tiivistelmä

Tieliikenneonnettomuudet ovat yleinen uutisoinnin ja tutkimuksen kohde. Onnettomuuksiin liittyviä taustatekijöitä on monia, kuten esimerkiksi liikenteen määrä, kuljettajan veren alkoholipitoisuus, väsymys, nopeus, sää, valoisuus ja liikennejärjestelyt. Liikennemäärä on selittäjä, johon ei ole suuremmin kiinnitetty huomiota tieliikenneonnettomuustutkimuksissa, vaikka sillä on suuri vaikutus liikenneolosuhteisiin.

Tässä tutkielmassa pyritään estimoimaan tieliikenneonnettomuuksien todennäköisyyksiä eri tilanteissa liikenteen automaattisten mittaustietojen (LAM) avulla. Aineiston laajuuden vuoksi tutkielmassa keskitytään vain valteilla 1–10 tapahtuneisiin onnettomuuksiin. Valtatiet jaetaan 29 tapaukseen, jotta saadaan aluekohtaista tietoa liikennemäärästä. Aineisto on vuosilta 2017–2021 ja valmiin aineiston laajuus on 1 268 544 riviä.

Aluksi aineistot täytyy saada yhtenevään muotoon sekä määrittää tarkalleen, miten onnettomuuksien todennäköisyyksiä estimoidaan. Mallinnetaan yhden tunnin aikaista onnettomuuden todennäköisyyttä tieosuudella, jossa selittäjinä ovat liikennemäärä, suhteellinen ruuhkaisuus, tieosuuden pituus, vuorokaudenaika sekä vuorokaudenaika. Mallinnus toteutetaan logistisella regressiomallilla.

Aineiston laajuuden ansiosta muuttujakohtaiset p-arvot ovat hyvin pieniä. Suurimman kokonaisvaikutuksen mallissa aiheuttaa liikennemäärä. Tien suhteellisen ruuhkaisuuden vaikutus on myös selvä. Tapauskohtaisesti aineistossa havaitut onnettomuusriskit poikkeavat mallin antamista estimaateista, joissain tapauksissa huomattavasti. Tapauksista osa on siis onnettomuusalttiimpia kuin toiset. Mikäli mallista haluaisi tapauskohtaisesti yhä tarkemman, malliin tarvittaisiin lisää selittäjiä, joita ei käytetyissä aineistossa ollut saatavilla.

Avainsanat: liikenne, valtatie, tieliikenneonnettomuus, liikennemäärä, logistinen regressiomalli, tieliikenteen automaattiset mittaustiedot.

Sisällys

1 Johdanto	1
1.1 Onnettomuuksista muualla	3
1.2 Tutkielmassa käytettyjä termejä	4
2 Aineisto	5
2.1 Tieliikenteen automaattiset mittaukset (LAM)	5
2.2 Tieliikenneonnettomuusdata	7
2.3 Lopullisen aineiston muodostaminen	7
2.4 Lopullisen aineiston kuvailua	12
2.5 Esimerkitapaukset	20
3 Menetelmät	23
3.1 Lineaarinen regressiomalli	23
3.2 Logistinen regressiomalli	23
3.3 Akaiken informaatiokriteeri	25
4 Tulokset	25
5 Johtopäätökset ja pohdinta	34

1 Johdanto

Monissa tutkimuksissa tarkastellaan liikenneonnettomuuksia useista eri näkökulmista. Esimerkiksi kuolleisuutta tai raskaan liikenteen osuutta onnettomuuksissa pyritään selittämään. Liikennemäärän vaikutusta ei kuitenkaan ole juurikaan tutkittu. Tämä johtuu varmasti osittain siitä, että riittävän laajaa tietoa liikennemäärästä ei ole aiemmin ollut saatavilla.

Tässä tutkielmassa hyödynnetään liikenteen automaattisten mittauspisteiden (LAM-pisteiden) keräämää liikennemääräaineistoa, jota ei ole toistaiseksi käytetty mallintamaan onnettomuuksien määrää Suomessa. Ensimmäiset LAM-pisteet otettiin käyttöön jo vuonna 1988, mutta vuodesta 2010 eteenpäin kerätty aineisto on ladattavissa sivulta [3], jossa voi myös valita minkä tyyppistä LAM-pisteiden keräämää dataa haluaa hyödyntää. LAM-pisteiden keräämiä liikennemääriä vuosilta 2017–2021 hyödynnetään tässä tutkielmassa mallintamaan liikenteen määrän vaikutusta tieliikenneonnettomuuksiin.

Tämän tutkielman ensisijaisena tarkoituksena on tutkia tieliikenteen liikennevirran vaikutusta onnettomuuksien aiheutumiseen Suomen tärkeimmillä valtateillä. Onnettomuuden mallinnuksessa käytetään liikennemäärää, vuorokaudenaikaa, vuodenaikaa, tieosuuden pituutta sekä liikennemäärän suhteellista ruuhkaisuutta. Onnettomuustodennäköisyyden mallinnus toteutetaan logistisella regressiomallilla ja vaste muuttujana on tuntikohtainen tieliikenneonnettomuuden tapahtumisen todennäköisyys kyseisellä tieosuudella.

Mallinnus sekä aineiston käsittely toteutetaan R-ohjelmistolla. Aineiston käsittelyssä isossa roolissa on R-ohjelman paketti *sqldf*. Kyseisen paketin avulla pystytään käyttämään R-ohjelmistossa SQL (Structured Query Language) tyyppistä syntaksia.

Tutkielmassa selvitetään liikennemäärän vaikutusta tieliikenneonnettomuuden todennäköisyyteen. Tämän lisäksi selvitetään, tapahtuuko onnettomuuksia suhteessa enemmän ruuhkaisina ajankohtina tai ruuhkaisemmilla teillä verrattuna liikenteen kannalta rauhallisempiin teihin ja ajankohtiin. Tarkastellaan, tapahtuuko ruuhka-aikoina enemmän onnettomuuksia kuin rauhallisimpina aikoina liikennemäärään suhteutettuna. Esimerkiksi jos liikenne on vilkasta kyseisen tien tavanomaiseen liikennemäärään nähden, kuinka suuri ero on onnettomuuden todennäköisyydessä.

Liikennemäärän sekä suhteutetun ruuhkaisuuden lisäksi tarkastellaan myös muiden muuttujien vaikutusta onnettomuuden todennäköisyyteen. Erityisesti tarkastellaan, miten ajankohta eli vuorokaudenaika sekä vuodenaika vaikuttavat onnettomuuksien

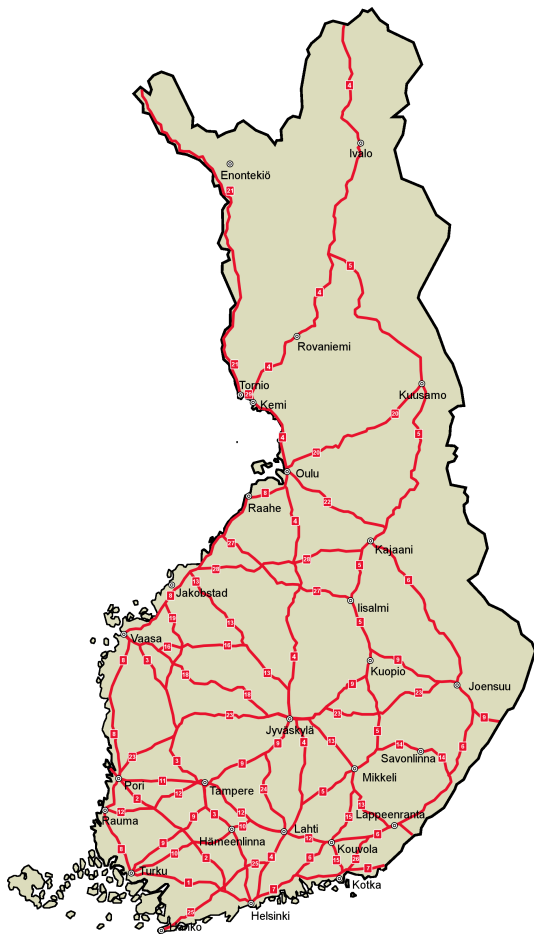
todennäköisyyksiin. Lisäksi tarkastellaan tieosuuden pituuden vaikutusta mallin estimoimaan onnettomuuden todennäköisyyteen.

Tutkielmassa käytetään kahta aineistoa: LAM-aineistoa sekä tieliikenneonnettomuusaineistoa. Näiden yhdistämisessä hyödynnettiin kolmatta LAM-metadata-aineistoa (saatavilla Fintrafficin sivulta [4]), jonka avulla yhdistämiseen tarvittavia tietoja liitettiin LAM-aineistoon. Aikaväliksi valittiin 2017–2021, sillä viiden vuoden aineisto koettiin riittäväksi aineiston koon vuoksi. Lisäksi tämä oli viimeisin ajanjakso, jolta saatiin molemmat aineistot.

Tämän tutkielman jaottelun mukaisesta tieosuudesta käytetään tässä tutkielmassa usein myös nimitystä tapaus. Kaikki tapaukset koostuvat useammasta kunnasta ja tietystä valtatiestä. Pääosa valtateista jaettiin useampaan tapaukseen sekä useammat kunnat esiintyvät myös useassa eri tapauksessa, mikäli kunnan läpi kulkee useampi valtatie. Tapauksesta esimerkkinä tapaus 6 3, joka on valtatie kuusi kolmas osuus tämän tutkielman mukaisen jaon perusteella. Tapaus 6 3 sisältää kunnat Kontiolahti, Juuka sekä Sotkamo.

Tutkielmassa hyödynnetään valtateita 1–10 (https://fi.wikipedia.org/wiki/Valtatiet_Suomessa). Syynä tähän rajoitukseen on teiden lisäämisen työläys. Tiet jaetaan Suomen kartan ja aineiston perusteella vähintään kahden kunnan osiin, jotta lopulliseen aineistoon ja mallinnukseen saadaan käyttöön muuttujaksi myös kriittinen selittäjä tieosuuden pituus. Kuvassa 1 näkyy kaikki Suomen valtatie.

Seuraavissa alaluvuissa esitellään lyhyesti useita eri lähteistä löydettyjä onnettomuustutkimuksia sekä -artikkeleita. Näiden lisäksi esitellään valittuja termejä tutkielmassa esiintyvillä asioilla sekä kokonaisuuksilla. Toisessa luvussa kuvataan LAM-pisteiden toimintaa ja aineiston keruuta sekä lyhyesti tieliikenneonnettomuusaineiston muodostumista. Toisessa luvussa kuvaillaan myös LAM-pisteiden liikennemääräaineistoa sekä tieliikenneonnettomuusdataa. Kappaleen lopussa selitetään myös, miten lopullinen aineisto muodostettiin sekä kuvaillaan tätä lopullista aineistoa tarkemmin kuvaajilla, taulukoilla sekä tunnusluvuilla. Kolmannessa luvussa avataan mallinnuksessa käytettyä logistista regressiomallia. Neljännessä luvussa selitetään mallinvalintaprosessia sekä tulkitaan valittua mallia. Lopuksi viidennessä luvussa käydään läpi johtopäätöksiä sekä haasteellisia kohtia tutkielman toteutuksessa ja mainitaan, kuinka tieliikenneonnettomuuksia voisi mallintaa laajemmin sekä tarkemmin, mikäli käytettävissä olisi enemmän resursseja.



Kuva 1: Kaikki valtatiet Suomessa. Kuvan lähde:

https://fi.wikipedia.org/wiki/Valtatiet_Suomessa#/media/Tiedosto:Suomi_Valtatiet.png

1.1 Onnettomuuksista muualla

Onnettomuuksien ja liikennemäärän yhteyttä ei ole aiemmin tutkittu Suomessa. Onnettomuuksien vakavuudesta sekä erityyppisistä onnettomuuksista on olemassa tietoa useilla sivustoilla. Liikenneturva-sivustolta löytyy lukuisia tutkimuksia onnettomuuksista, kuten esimerkiksi väsymyksestä [9] sekä raskaasta liikenteestä [6]. Väsymykseen liittyvässä ”Väsymys liikenteessä” -tutkimuksessa raportoidaan väsymyksen yleisyydestä liikenteessä sekä onnettomuuksissa. Tämän lisäksi tutkimuksessa avataan väsymyksen syitä ja ehkäisymenetelmiä. Raskaasta liikennettä koskeva tutkimus kuvailee raskaiden ajoneuvojen yleisyyttä onnettomuuksissa sekä näiden onnettomuuksien vakavuutta vuosina 2013–2022. Tämän lisäksi tutkimuksessa raportoitii myös osallisten jakaumia tilanteissa, joissa yhtenä osallisena on ollut linja-auto tai kuorma-auto. Artikkelissa ”Vauhti lisää turmariskiä myös päätteillä” [8] raportoidaan nopeuksien muutosten vaikutuksesta onnettomuuksien ilmenevyy-

teen sekä niiden vakavuuteen.

Arto Luoman tutkimuksessa (2022) [7] tarkastellaan onnettomuuksia erityisesti kuolemaan johtaneiden onnettomuuksien aiheuttajien sekä osallisten näkökulmasta. Tutkimuksessa kiinnitetään huomiota erityisesti kuljettajan ikään, juopumuksen määrään, ylinopeuteen sekä henkilöauton vuosimalliin. Tutkimuksessa tarkastellaan, mitkä tekijät vaikuttivat siihen, että kolarista aiheutui kuolonuhri. Arto Luoman tutkimuksessa käytettiin myös LAM-pisteiden keräämää tietoa, mutta liikennemäärien sijasta niiden keräämää dataa hyödynnettiin arvioimaan autojen nopeusjakaumia.

Edellä esiteltyt artikkelit sekä raportit ovat vain eräitä esimerkkejä onnettomuuksiin liittyvistä tutkimuksista. Yhteisenä tekijänä on kuitenkin havaittavissa, että vaikkakin osassa tutkimuksista saatetaan selittää onnettomuuksien ilmenemistä, niissä tarkastellaan pääosin vain tapahtuneita onnettomuuksia. Jotta saataisiin parempi kuva onnettomuuksista sekä niiden syistä, on syytä myös tarkastella tilanteita, joissa ei ole tapahtunut onnettomuutta. Tämän tutkielman ideana on hyödyntää LAM-pisteiden keräämää liikennedatata selittämään onnettomuuksien ilmenemistä Suomen valtateillä.

1.2 Tutkielmassa käytettyjä termejä

Tässä tutkielmassa käytettyjä termejä sekä niiden tarkemmat määritelmät:

1. **Tapaus:** Tarkoittaa tietyn tien ja kuntien muodostamaa tieosuutta. Kaikki tapaukset ovat esillä taulukossa 1.
2. **Vastemuuttuja:** Tämän tutkielman vastemuuttujana käytetään tuntikohtaista onnettomuuden todennäköisyyttä tietyillä selittäjien arvoilla.
3. **Tilanne:** Mallia estimoitaessa on kyseessä aina tietyt arvot selittäjille, jolloin mallilla estimoidaan kyseistä tilannetta.
4. **Onnettomuusriski:** Aineistosta laskettu onnettomuuden todennäköisyys esimerkiksi tietyssä tilanteessa tai tapauksessa.
5. **Onnettomuusestimaatti:** Mallin avulla laskettu onnettomuuden todennäköisyys esimerkiksi tietyssä tilanteessa tai tapauksessa.
6. **Liikennevirta/Liikennemäärä:** Kuvaa liikenteen tiheyttä. Tässä tutkielmassa useimmiten viitataan kulkevien ajoneuvojen määrään tunnin aikana.

2 Aineisto

Tässä luvussa kuvaillaan käytettyjä aineistoja. Aineisto on jakautunut kahteen osaan: tieliikenneonnettomuusdataan sekä tieliikenteen automaattisiin mittaustietoihin (LAM). Aineistot täytyy yhdistää mallinnusta varten yhdeksi aineistoksi. LAM-mittaukset on muodostettu raakadatasta valmiiksi raportteihin Digitrafficin toimesta. Aineistot täytyy yhdistää siten, että otetaan huomioon vain niillä teillä tapahtuneet onnettomuudet, joilla on LAM-mittauspisteitä. Tämän lisäksi rajoitetaan aineisto valtateihin 1–10, kuten johdannossa mainittiin.

Seuraavissa alaluvuissa kuvaillaan erikseen molempia käytettyjä aineistoja. Lopulta kuvaillaan, millainen aineisto yhdistämisen jälkeen jää jäljelle. Molemmista aineistoista kerrotaan hieman yleistietoa sekä kuvaillaan aineistojen yhdistämiseen valmistavia muokkauksia. Lisäksi havainnollistetaan tärkeimmät vaiheet aineistojen käsittelystä. Lopuksi kuvaillaan valmista aineistoa tunnusluvuilla, kuvaajilla sekä taulukoilla.

2.1 Tieliikenteen automaattiset mittaustiedot (LAM)

Tässä tutkielmassa tärkeässä osassa ovat LAM-pisteet (liikenteen automaattiset mittaustiedot). LAM-pisteisiin liittyviä tietoja löytyy Digitraffic-palvelusta [2].

LAM-pisteillä tien päällysteen sisälle on upotettu silmukka, jossa sähkömagneettinen induktio aiheuttaa muutoksen magneettikentässä, kun ajoneuvon metallinen massa ylittää pisteen. LAM-piste koostuu jokaisella kaistalla olevasta kahdesta induktiosilmukasta sekä tiedonkeruuyksiköstä. LAM-laite kirjaa kaikki pisteen ylittävät ajoneuvot sekä tallentaa tapahtuman kellonajan, ajosuunnan, ajokaistan, ajonopeuden, ajoneuvon pituuden, peräkkäisten ajoneuvojen aikaeron ja ajoneuvoluokan.

LAM-pisteillä lasketaan ohiajavien ajoneuvojen määrä sekä muita tietoja näistä tapahtumista, kun ajoneuvot ylittävät maahan upotetun LAM-pisteen. LAM-asemien historiadataa on kahdenlaista: LAM-raakadata, joissa on yksittäisen LAM-pisteen ylitykseen liittyvää tietoa, sekä LAM-raportit, joita hyödynnetään tässä tutkimuksessa. Molemmat datat ovat saatavilla vain muutaman päivän viiveellä.

Aiemmin LAM-pisteiden tietoja ylläpiti Väylävirasto (ennen vuotta 2019 Liikennevirasto). LAM-pisteitä nykyisin ylläpitävää Digitraffic-palvelua ylläpitää Liikenteenohjausyhtiö Fintraffic Oy, jonka omistaa Suomen valtio. Fintrafficin vastuualue on liikenteenohjauksen ja -hallinnan palvelut (<https://fi.wikipedia.org/wiki/>

Fintraffic). Fintraffic tarjoaa avointa liikennetietoa Suomen tie-, rautatie- sekä vesiliikenteestä tuottamansa Digitraffic-palvelun kautta.

Käytettävä aineisto on saatavilla sivulta [3]. Tämän tutkielman tarpeisiin valittiin sivustolta aineistoksi liikennemäärät sekä raportin tarkkuudeksi tuntiliikenne-raportti. Tämän lisäksi aikaväliksi valittiin 2017–2021. Aineistoon valittiin kaikki teille sijoitetut LAM-pisteet sekä kaikki ajoneuvoluokat summattuna. Lisäksi aineistoon valittiin molemmat ajosuunnat ja kaistat summattuina.

Aineiston suuren koon takia se täytyi ladata kahden kuukauden paloissa, jotta sivusto ei jumittuisi. Ladatussa LAM-aineistossa on 33 saraketta, joista 5 on tyhjiä tai täysin identtisiä. Loput 28 saraketta sisältävät LAM-pisteen pistetunnuksen, LAM-pisteen sijainnin, päivämäärän, 24 saraketta tuntikohtaisia liikennemääriä sekä LAM-pisteen päivän liikenteen totaalin (esimerkiksi 1, vt7_Rita, 20170101, 85, ..., 111, 10239).

Tämän liikennemääräaineiston lisäksi tutkielmassa käytetään myös LAM-metadata-aineistoa, joka on saatavilla Fintrafficin sivulta [4]. LAM-metadata-aineistossa kerrotaan kaikista LAM-pisteistä lisätietoja. Näitä lisätietoja ovat esimerkiksi tien numero, kunta, laskinlaite, koordinaatit, käyttöönottopäivä, päättymispäivä sekä lisätietoa molemmista ajosuunnista (kuten kesänopeus, talvinopeus sekä molempien ajosuuntien seuraavat kunnat).

Vuosien 2017–2019 aikana käytettyjen asemien määrä kasvoi vuosittain vain muutamalla pisteellä vuodessa 481 pisteestä 487:n pisteeseen. Vuoteen 2020 mennessä pisteitä oli 495 ja vuonna 2021 käytössä oli jo 515 pistettä. Edeltävät määrät on laskettu sen mukaan, kuinka monta asemaa on käytetty kyseisenä vuonna. Vuosina 2017–2021 on otettu käyttöön 53 uutta pistettä ja 15 vanhaa pistettä on poistettu kokonaan käytöstä. Näiden lisäksi joitakin pisteitä on otettu väliaikaisesti pois käytöstä. Tässä tutkielmassa käytettäviä valtateitä 1–10 koskevia LAM-pisteitä on aineistossa 294.

LAM-pisteiden aineiston muokkaus aloitetaan yhdistämällä paloittain ladatut, vuosien 2017–2021 aineistot, jolloin koko aineiston kooksi saadaan 875 238 riviä. Seuraavaksi erotellaan jokaisen tunnin havainto omaksi rivikseen, jolloin aineistoon tulee 21 005 712 riviä. Tämän jälkeen liikennemäärän tuntikohtainen aineisto liitetään LAM-pisteiden taustamuuttuja-aineiston kanssa pistetunnuksen avulla. Näin saadaan jokaiselle riville tarpeellista lisäinformaatiota, kuten esimerkiksi kunta sekä tie. Tämän jälkeen LAM-aineistossa on kaikki vaadittavat muuttujat aineistojen yhdistämistä varten.

2.2 Tieliikenneonnettomuusdata

Toinen osa tarvittavasta aineistosta on tieliikenneonnettomuusaineisto. Tämä aineisto on julkinen ja saatavilla Väyläviraston sivuilta [11]. Aineiston on koontanut Väylävirasto [10] ja se on koottu kolmesta lähteestä: poliisin PATJA-tietojärjestelmästä, joka on myös ”Tilastokeskuksen ylläpitämän virallisen tilaston perusta”, vakuutusyhtiöiden liikennevahinkorekisteristä (LVK) ja sairaaloiden hoitoilmoitusrekisteristä (HILMO). Väyläviraston mukaan aineisto sisältää yksityiskohtaista tietoa onnettomuudesta, onnettomuuden osapuolista ja henkilöistä. Väylävirastossa onnettomuustietoihin liitetään tierekisteristä tapahtumapaikan tie- ja liikenneoloja kuvaavat tiedot, kuten Väyläviraston sivuilla sanotaan.

Tieliikenneonnettomuusdatoissa on 102–106 saraketta, mutta tässä tutkielmassa näistä tarvitaan vain muuttujat: tie, kunta, päivämäärä ja tunti. Edellä mainittujen muuttujien suhteen vajaan rivit jätettiin kokonaan pois. Muiden muuttujien poisjättäminen johtuu siitä, että niistä on tietoa ainoastaan onnettomuustapauksissa, eikä onnettomuuden vakavuutta pyritä tässä tutkielmassa kuvaamaan.

Tieliikenneonnettomuusdatojen vuosittaiset koot vuodesta 2017 alkaen vuoteen 2021 laskevat hieman seuraavasti: 17350, 16220, 15297, 11347 sekä 10726. Täten kyseisellä viiden vuoden ajanjaksolla onnettomuuksia tapahtui yhteensä 70 940. Aiemmin mainituista tarvittavista muuttujista puuttuvuutta esiintyi tie-muuttujassa 23 413 riviä ja tunti-muuttujassa 794 riviä. Muuttujien päivä ja kunta kohdalla ei ilmennyt yhtään puuttuvaa arvoa. Näitä muuttujia käytetään yhdistettäessä onnettomuusdata lopulliseen aineistoon.

Valtateillä 1–10 tapahtui 10 593 onnettomuutta tarkasteltavalla aikavälillä. Kun aineistosta poistetaan myös aiemmin mainitut rivit, joilta puuttuu tunti-muuttuja, onnettomuuksien määräksi jää 10 570. Näiden vaiheiden jälkeen aineiston koko laskee vielä, sillä onnettomuuksien todennäköisyyksiä ei voida mallintaa, jos kunnassa kyseisellä tiellä ei ole LAM-pistettä. Näiden rajoitusten jälkeen aineistossa on jäljellä 8162 onnettomuutta.

2.3 Lopullisen aineiston muodostaminen

Edellisissä alaluvuissa valmisteltiin onnettomuusaineistoa sekä LAM-aineistoa aineistojen yhdistämistä varten. Tässä luvussa muodostetaan lopullinen aineisto, jonka avulla mallinnetaan tieliikenneonnettomuuksien todennäköisyyttä. Ensimmäisenä

korjataan molempien aineistojen päiväykset, sillä ne olivat eri muodoissa. Muutetaan ne yhtenevään muotoon.

Onnettomuusaineistoon lisätään vielä onnettomuusmuuttuja, jonka avulla saadaan vastemuuttuja, joka kuvaa tapahtuuko onnettomuus vai ei. Seuraavaksi otetaan liikennemäärien keskiarvot tilanteista, joissa on samat arvot muuttujille kunta, tunti, päivämäärä sekä tien numero. Rivien määrä on tämän vaiheen jälkeen 12 787 715. Tämä tehdään, jotta onnettomuudet eivät kopioidu useampaan kertaan, kun onnettomuusaineisto yhdistetään tuntikohtaiseen aineistoon. Tämän lisäksi saadaan tarkempia arvioita tien liikennemääristä alueilta, joissa on useampia LAM-pisteitä.

Valtateihin 1–10 rajoittuessa aineiston rivimäärä on 5 457 783. Yhdistetään sitten tuntikohtainen liikennemääräaineisto sekä onnettomuusaineisto päivämäärän, tunnin, kunnan sekä tien numeron perusteella. Lopuksi jaetaan vielä aineisto teiden ja kuntien mukaan tapauksiin.

Valtateiden jaon toteutuksen ideana on ottaa peräkkäisiä kuntia ja tarkastella kartalta, kuinka pitkän matkan kyseinen valtatie kulkee näiden kuntien alueella. Tavoitteenä on 100–200 kilometriä, mutta tässä joustetaan, mikäli se on aineiston puolesta tarpeellista. Jaon tavoitteena on saada pituusmuuttuja malliin sekä paloitella pidemmät tiet useampaan osaan, jotta saadaan kohdistetumpaa tietoa liikennevirrasta ja onnettomuuksista.

Tähän liittyviä haasteita ovat esimerkiksi se, että tie saattaa käydä useampaan kertaan saman kunnan alueella. Tapausta muodostettaessa täytyy näissä tilanteissa sisällyttää myös seuraava kunta, jotta saadaan katettua koko väli. Tämän lisäksi kuntien vaihtelevat koot sekä LAM-pisteiden isot välit ja uusien pisteiden käyttöönotto vaikuttavat myös osien pituuksiin. Tapauskohtaiset pituudet sekä niihin sisältyvät kunnat löytyvät taulukosta 1 sekä tapauskohtaisesti vaikuttaneiden LAM-pisteiden määrät taulukosta 2. Jaon jälkeen saadaan 29 eri tapausta, jolloin valmiin aineiston rivimääräksi tulee 1 268 544 riviä.

Aineiston käsittelyn vaiheet tiivistettynä:

1. LAM-aineiston käsittely.

- (a) LAM-pisteiden tietojen lataus paloittain, LAM-metadatan lataaminen
- (b) LAM-pisteaineistojen koonti paloista, 875 238 riviä
- (c) Aineiston jako tuntikohtaiseksi, 21 005 712 riviä

- (d) Otetaan liikennemäärän keskiarvo tilanteista, joissa identtiset arvot muuttujilla kunta, tunti, päivämäärä sekä tien numero, 12 787 715 riviä
 - (e) Tuntikohtaisen aineiston yhdistäminen LAM-metadatan kanssa (vain valtatie 1–10), 5 457 783 riviä
 - (f) Päivämäärän muokkaaminen käsiteltävään sekä yhtenevään muotoon
2. Onnettomuusaineiston käsittely.
- (a) Vuosikohtaisten tieliikenneonnettomuusaineistojen lataaminen
 - (b) Vuosikohtaisten aineistojen yhdistäminen 70 940 riviä
 - (c) Vajaiden rivien poistaminen (puuttuu tie ja/tai tunti), 47 371 riviä
 - (d) Teiden rajoittaminen valtateille 1–10, 10 570 riviä
 - (e) Päivämäärän muokkaaminen käsiteltävään sekä yhtenevään muotoon
 - (f) Onnettomuus-muuttujan lisääminen kaikkiin riveihin
3. Yhdistetään liikennemäärä- sekä onnettomuusaineistot, 5 457 783 riviä. Kaikki LAM-aineiston valtateille 1–10 kuuluvat rivit säilytetään. Yhdistämisessä avainmuuttujina käytetään päivämäärää, tuntia, kuntaa sekä tietä.
4. Asetetaan kaikille riveille NA arvon tilalle arvoksi 0 onnettomuusmuuttujaan
5. Kuntien kerääminen sekä mittaaminen 29 tapaukseksi
6. Tapauskohtaisen aineiston kokoaminen, 1 268 544 riviä
7. Lisämuuttujien johtaminen aineistossa käytettävissä olevien tietojen perusteella (esimerkiksi vuodenaika, ruuhkaisuus, viikonpäivä)

Taulukko 1: Teiden ja kuntien jako. Ensimmäisenä tapaus, joka koostuu tien numerosta sekä numerosta, joka kuvaa monesko osa se on kyseistä tietä. Tapauksen pituus kertoo kyseisen tapauksen pituuden kilometreinä. Kunnat ovat lista, joka sisältää kyseisen tapauksen kaikkien käytettävissä olevien kuntien nimet tien kulkusuunnan mukaisessa järjestyksessä (yleisimmin suunta Etelä-Pohjoinen tai Länsi-Itä).

Tapaus	Tapauksen pituus (km)	Kunnat
1 1	112	Helsinki, Espoo, Kirkkonummi, Vihti, Lohja, Salo
1 2	53	Paimio, Kaarina, Turku
2 1	204	Vihti, Karkkila, Humppila, Nakkila, Pori
3 1	158	Helsinki, Vantaa, Nurmijärvi, Riihimäki, Janakkala, Hämeenlinna, Lempäälä
3 2	199	Tampere, Pirkkala, Nokia, Hämeenkyrö, Ikaalinen, Kurikka
3 3	77	Ilmajoki, Laihia, Mustasaari, Vaasa
4 1	137	Helsinki, Vantaa, Järvenpää, Mäntsälä, Hollola, Lahti, Heinola
4 2	141	Hartola, Joutsa, Toivakka, Jyväskylä
4 3	217	Laukaa, Äänekoski, Viitasaari, Pyhäjärvi, Kärämäki
4 4	121	Siikalatva, Liminka, Kempele, Oulu
4 5	224	Ii, Kemi, Keminmaa, Tervola, Rovaniemi
4 6	327	Sodankylä, Inari, Utsjoki
5 1	213	Pertunmaa, Mäntyharju, Mikkeli, Joroinen, Leppävirta, Kuopio
5 2	179	Siilinjärvi, Iisalmi, Kajaani
5 3	380	Ristijärvi, Suomussalmi, Kuusamo, Kemijärvi
6 1	149	Loviisa, Kouvola, Luumäki, Lappeenranta
6 2	239	Imatra, Rautjärvi, Kitee, Joensuu
6 3	193	Kontiolahti, Juuka, Sotkamo
7 1	91	Helsinki, Vantaa, Sipoo, Porvoo, Loviisa
7 2	98	Pyhtää, Kotka, Hamina, Virolahti
8 1	134	Raisio, Masku, Pyhäranta, Eurajoki, Pori
8 2	229	Kristiinankaupunki, Närpiö, Mustasaari, Vaasa, Vöyri
8 3	148	Uusikaarlepyy, Kokkola, Kalajoki
8 4	108	Raahe, Liminka
9 1	141	Lieto, Akaa, Tampere
9 2	173	Kangasala, Orivesi, Jämsä, Jyväskylä, Muurame, Laukaa
9 3	123	Suonenjoki, Kuopio
9 4	209	Tuusniemi, Liperi, Joensuu, Tohmajärvi
10 1	125	Lieto, Marttila, Tammela, Hämeenlinna

Taulukko 2: Tieosuuskohtaiset liikennemäärien keskiarvot sekä maksimit kuvaavat liikennevirtaa tunnin aikana kyseisellä osuudella. Tapahtuneiden onnettomuustapausten summa kyseisen tapauksen alueella koko tarkastellulla aikavälillä. LAM-pisteiden määrä kertoo, kuinka monta pistettä sisältyy kyseiseen tieosuuteen. Järjestetty liikennemäärän keskiarvon perusteella laskevaan järjestykseen.

Tapaus	Liikennemäärän keskiarvo	Liikennemäärän maksimi	Onnettomuuksien määrä	LAM-pisteitten määrä
3 1	1346.30	4367.20	618	23
4 1	1288.21	4247.12	701	15
1 1	1185.11	3746.56	607	42
3 2	984.03	3435.67	338	10
1 2	892.21	2959.61	104	10
7 1	890.73	3155.60	243	17
4 4	796.53	2585.48	321	14
9 1	722.01	2325.00	122	4
8 1	585.94	1779.40	292	6
5 1	508.21	1905.67	384	11
2 1	446.52	1401.40	257	5
4 2	446.29	1597.92	297	7
9 2	429.24	1645.08	402	7
7 2	406.80	1348.28	98	21
4 5	392.44	1447.23	276	17
6 1	367.79	1469.29	287	19
9 4	366.66	1221.10	111	6
5 2	360.57	1263.94	182	6
3 3	339.80	1397.33	178	4
6 2	333.91	1175.38	170	5
10 1	291.83	986.50	213	5
8 3	279.37	911.00	267	4
9 3	263.00	999.00	135	2
4 3	260.34	1128.00	323	6
8 2	246.45	829.80	474	8
8 4	243.07	781.00	77	3
6 3	233.63	785.33	168	3
5 3	89.60	478.50	206	7
4 6	43.80	251.67	176	4

Lopuksi lisätään muuttujat ruuhkaisuus, päivän aika sekä vuodenaika. Ruuhkaisuutta kuvataan siten, että otetaan tarkasteltavan tunnin liikennemäärä ja jaetaan se tapauskohtaisen liikennemäärän maksimilla. Näin saadaan muuttuja väliltä 0–1, jonka arvo riippuu sekä tapauksesta että liikennemäärästä. Tämän muuttujan idea on ottaa huomioon eri tapauksien erilaisuudet. Tapaukset poikkeavat toisistaan esimerkiksi teiden leveyksissä, kaistamäärissä, näkyvyydessä, tiejärjestelyissä sekä muissa asioissa, jotka vaikuttavat siihen, miten paljon tiellä pystyy kulkemaan liikennettä ja millä nopeudella liikenne etenee. Edellä mainituista muuttujista ei ole helposti saatavilla tietoa, jonka avulla pystyttäisiin kuvaamaan tien liikennemäärän kapasiteettia muutoin.

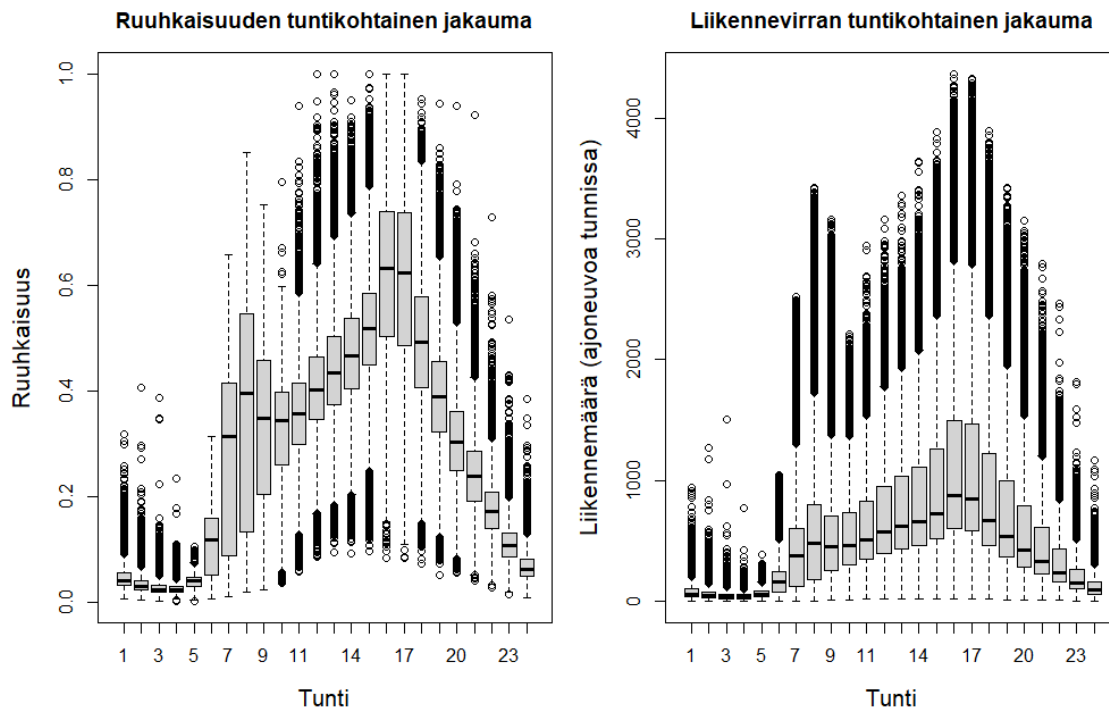
Edellä mainituista tiedoista vain osa on saatavilla, mutta niiden kerääminen tapahtuisi manuaalisesti. Tämän lisäksi samassa tapauksessa saattaa olla huomattavaa vaihtelua esimerkiksi kaistamäärissä sekä nopeusrajoituksissa, joten näitä ei käsitellä tässä tutkielmassa. Lisäksi osalle listatuista piirteistä on erittäin vaikea saada tai määrittää arvoa. Näiden haasteiden sekä tämän tyyllisen ruuhkaisuuden määrittelyn yksinkertaisen tulkinnan ja käytön takia tässä tutkielmassa käytetään edellä mainittua suhteellista tapaa kuvailla ruuhkaisuutta.

2.4 Lopullisen aineiston kuvailua

Tässä alaluvussa kuvaillaan valmiin aineiston muuttujien jakaumia taulukoiden sekä kuvien avulla. Tarvittaessa raportoidaan myös kuvaavia lukuja sekä tulkitaan esitettyjä taulukoita ja kuvaajia.

Lopullisen aineiston liikennemäärän tuntikohtainen laatikkokuvio on kuvassa 2. Kuvasta 2 pystyy erottamaan kiireisimmät tunnit sekä aamulla että iltapäivällä. Vuorokaudenaikamuuttujan tuntivälit on myös jaettu tämän mukaan, jotta tunnin mallintaminen olisi havainnollistettavampaa sekä tehokkaampaa. Tuntimuuttujan ja tässä tapauksessa sen perusteella luodun vuorokaudenaikamuuttujan ideana on ottaa huomioon päivän liikennesyklin lisäksi myös sellaisia selittäjiä, joita ei voida käyttää suoraan, kuten esimerkiksi valoisuus. Päivän aikamuuttujan tuntivälit ovat aamu 7–9, päivä 10–14, iltapäivä 15–18, ilta 19–23 ja yö 0–6. Vuodenaikamuuttujan arvot jaetaan päivämäärien 15.3, 15.6, 15.9 sekä 15.12 mukaan.

Liikennevirran lisäksi kuvassa 2 on myös aineiston ruuhkaisuuden tuntikohtainen jakauma laatikkokuviona. Ruuhkaisuuden kuvaaja on hyvin samanlainen, mutta siinä laatikkokuviot ovat hieman selvempiä. Tämä siksi, että kaikkien tapauksien arvot vaihtelevat juuri samalla skaalalla, jolloin pystytään tarkastelemaan tuntikohtais-

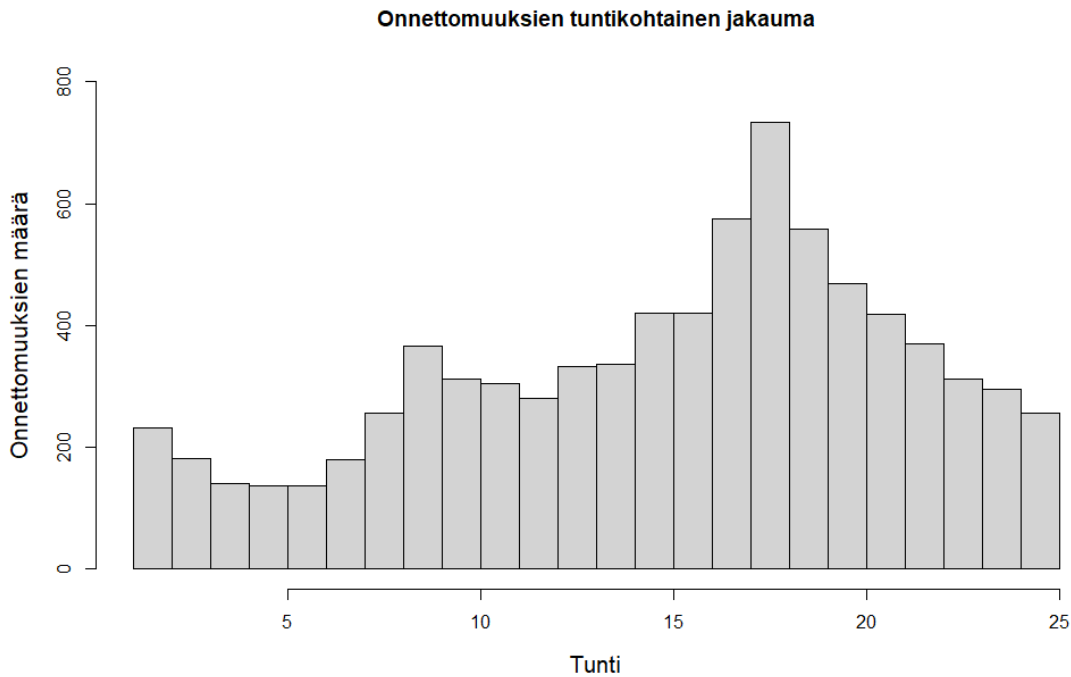


Kuva 2: Vasemmalla ruuhkaisuuden tuntikohtainen jakauma aineistossa laatikkokuviona ja oikealla aineiston tuntikohtainen liikennevirta laatikkokuviona.

ta liikennevirtaa paremmin sekä helpommin. Liikenteessä on huomattavissa selkeitä tuntikohtaisia eroja (kuva 2), kuten esimerkiksi tuntien 16–17 selkeä huippu sekä aamuruuhkan jälkeinen lasku liikenteessä. Kuvasta 3 nähdään, kuinka onnettomuuksien tuntikohtaisessa jakaumassa on havaittavissa hyvin samanlainen jakauma kuin ruuhkaisuuden ja liikennevirran laatikkokuvioissa 2.

Onnettomuustilanteita on lopullisessa aineistossa 8027, sillä muutamissa tilanteissa saman tunnin aikana samassa tapauksessa on tapahtunut useampi onnettomuus (2 onnettomuutta 119 tapauksessa ja 3 onnettomuutta 8 tapauksessa). Useamman onnettomuuden tapauksia on hyvin vähän. Tämän tutkielman tavoitteena on tarkastella onnettomuuden ilmenemistä eikä mallintamisessa täten kiinnitetä huomiota onnettomuuksien määrään.

Onnettomuuksien jakaumia ajankohtaan liittyvien muuttujien suhteen aineistossa on esillä taulukoissa 3 ja 4 sekä kuvassa 4. Taulukossa 3 voidaan tarkastella millä ajankohdilla on korkeimmat onnettomuusriskit, kun tarkastellaan onnettomuuksien määriä ainoastaan vuorokaudenajan sekä vuodenaajan suhteen. Taulukon 3 ylemmällä puoliskolla esitetään suoraan aineistosta otetut onnettomuusmäärät ja alemmalla puoliskolla onnettomuudet suhteutettuna tuntimäärään. Vuorokaudenaikojen tuntimäärät ovat yö 7, aamu 3, päivä 5, iltapäivä 4 ja ilta 5.



Kuva 3: Vaaka-akseli kuvaa tuntia ja pystyakseli aineiston onnettomuuksien määrää kyseisellä tunnilla.

Vuorokaudenajoista iltapäivä on selvästi onnettomuusrikkain. Tuntikohtaisten onnettomuuksien summa on 572. Seuraavaksi eniten onnettomuuksia tapahtui iltana (374) sekä päiväntuntien (334) aikana. Tuntikohtaisesti mitattuna selvästi vähiten onnettomuuksia tapahtui yöllä (180) ja toiseksi vähiten aamulla (313).

Ilman tuntikohtaista tarkastelua saattaisi taulukon perusteella luulla esimerkiksi, että talviöinä (256 onnettomuutta) sekä -aamuina (260 onnettomuutta) olisi yhtä todennäköistä joutua kolariin. Kuitenkin suhteuttaessa onnettomuudet eri vuorokaudenaikojen tuntien määrillä ja tarkastelemalla tuntikohtaista onnettomuusriskiä havaitaan, että talviaamuisin on yli kaksi kertaa enemmän onnettomuuksia tuntia kohden, sillä yöllä tapahtuu keskimäärin 37 ja aamulla keskimäärin 87 tuntikohtaista onnettomuutta. Edellä esitetyissä onnettomuuslukemissa on otettu huomioon kaikki aineiston viiden vuoden aikana tapahtuneet onnettomuudet kaikilla LAM-mittauspisteiden kattavilla alueilla. Koko aineiston onnettomuudet on suhteutettu jokaisen vuorokaudenajan tuntimäärän suhteen, jotta pystytään havainnollistamaan, millä vuorokaudenajalla on korkein onnettomuusriski.

Vuodenajan sekä vuorokaudenajan suhteen aamulla, iltapäivällä sekä illalla tapahtuu syksyisin eniten onnettomuuksia. Yöllä sattuu selvästi eniten onnettomuuksia kesäisin. Päivätunteina eniten onnettomuuksia tapahtuu talvella. Aamuisin tapah-

tuu selvästi vähiten onnettomuuksia keväällä sekä kesällä. Päivätunteina keväällä sattuu hieman vähemmän onnettomuuksia kuin muina vuodenaikoina. Vastaavasti iltapäivisin sekä iltaisin onnettomuuksia tapahtuu huomattavasti vähemmän. Yöllä vähiten onnettomuuksia tapahtuu talvisin.

Taulukko 3: Onnettomuuksien jakautuminen vuodenaajan sekä vuorokaudenaajan mukaan aineistossa. Jälkimmäisen puoliskon rivit (suht) tarkoittaa, että ne on suhteutettu kyseisen vuorokaudenaajan tuntien määrään sekä pyöristetty lähimpään kokonaislukuun. Ne ovat tuntikohtaisia onnettomuusmääriä kyseisenä vuorokaudenaikana. Esimerkiksi keväällä aamulla (tuntien 7–9) tuntikohtainen keskiarvo on 59.

	yö	aamu	päivä	iltapäivä	ilta	yhteensä
kevät	296	176	352	410	340	1574
kesä	400	176	411	533	438	1958
syksy	311	324	429	691	603	2358
talvi	256	260	482	654	485	2137
yhteensä	1263	936	1674	2288	1866	8027
kevät (suht)	42	59	70	102	68	341
kesä (suht)	57	59	82	133	88	419
syksy (suht)	44	108	86	173	121	532
talvi (suht)	37	87	96	164	97	481
yhteensä (suht)	180	313	334	572	374	1773

Viikonpäivän ja vuorokaudenaajan suhteen tarkasteltaessa perjantaisin onnettomuuksien lukumäärä on korkein lähes jokaisena vuorokaudenaikana. Poikkeuksena aamut, jolloin maanantaisin on suurin määrä onnettomuuksia. Toinen poikkeus on yöaika, jolloin lauantaisin tapahtuu eniten onnettomuuksia. Lauantaisin sekä sunnuntaisin aamulla tapahtuu selvästi vähemmän onnettomuuksia kuin muina viikonpäivinä. Sama pätee myös iltapäivällä, vaikkakin ero ei ole yhtä suuri. Sunnuntaina myös päivällä on onnettomuuksien kannalta rauhallista. Öisin rauhallisimmat päivät järjestettynä ovat keskiviikko, tiistai, torstai sekä maanantai. Iltaisin onnettomuuksien kannalta turvallisoin päivä on tiistai.

Kuvasta 4 on huomattavissa selkeä yhteys viikonpäiväkohtaisen liikennemäärän ja onnettomuuksien määrän välillä. Viikonpäiväkohtaisen liikennemäärän kasvaessa myös onnettomuuksien määrä kasvaa. Perjantaisin sekä liikennemäärä että onnettomuuksien määrä ovat selvästi korkeimmat. Pienimmän neliösumman avulla piirretty viivan perusteella perjantaisin tapahtuu myös liikennemäärään nähden suhteessa enemmän onnettomuuksia kuin muina viikonpäivinä.

Taulukko 4: Onnettomuuksien jakautuminen viikonpäivän sekä vuorokaudenajan perusteella aineistossa.

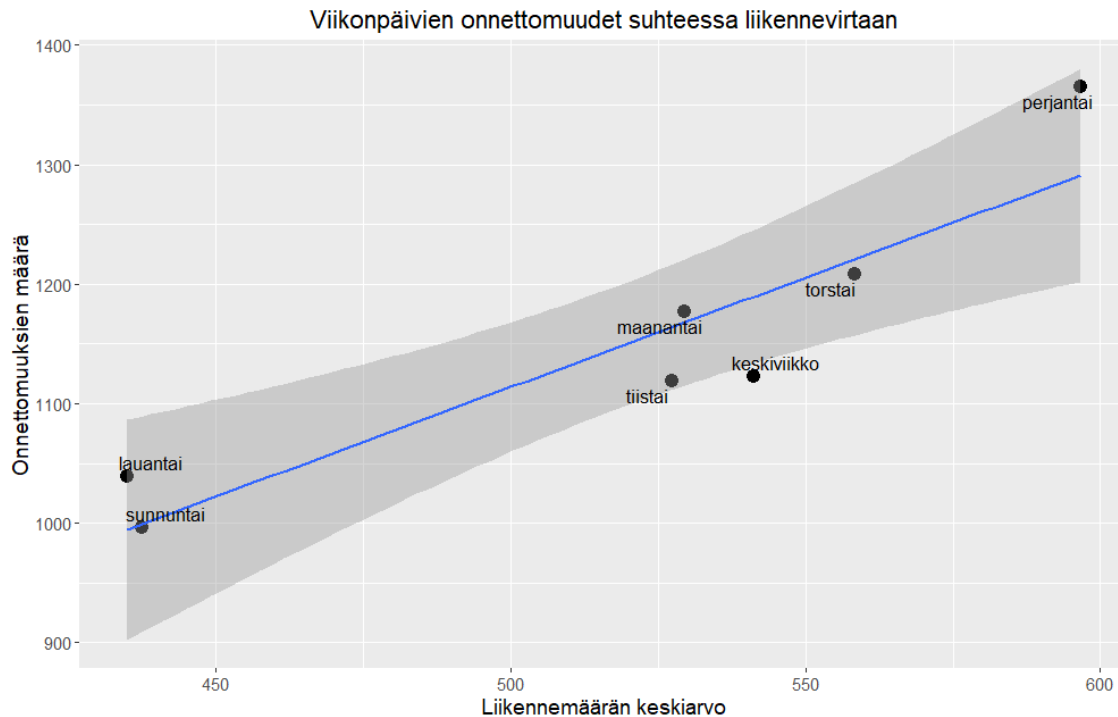
	yö	aamu	päivä	iltapäivä	ilta
maanantai	162	184	263	312	256
tiistai	151	154	233	345	236
keskiviikko	144	168	226	327	258
torstai	153	155	237	376	287
perjantai	203	148	282	425	307
lauantai	243	68	243	240	245
sunnuntai	207	59	190	263	277

Aineiston tienosuuden pituuden jakauma on nähtävissä histogrammina kuvassa 5. Kuvasta 5 nähdään, että tavoiteltuun tienosuuden pituuteen 100–200 kilometriä osuu pääosa havainnoista ja vain muutaman tapauksen pituus poikkeaa tästä huomattavasti (katso myös taulukko 1, jossa tapauskohtaiset pituudet). Onnettomuus-tilanteiden jakauma on pääosin myös samanlainen, mutta ääritapauksissa onnettomuuksia on tapahtunut vähemmän.

On hyvä muistaa, että pisimpien tieosuuksien tapauksissa on selkeästi vähiten liikennettä, joten pelkästään tästä kuvaajasta ei ole helppo havaita pituuden vaikutusta onnettomuuden todennäköisyyteen. Lyhyimpien tieosuuksien tapauksien liikennevirran on korkea ja niillä on suhteessa huomattavasti matalampi frekvenssi verrattaessa koko aineiston vastaaviin. Verratessa alle sadan kilometrin tapauksia on tien pituuden vaikutus havaittavissa.

Aineiston liikennemäärän tapauskohtaiset keskiarvot sekä maksimi-arvot ovat taulukossa 2. Tapauskohtaiset minimi- ja maksimiarvot kaikissa tapauksissa ovat välillä 0–24. Minimien pienimmät erot johtuvat siitä, että yöllä liikenne on rauhallisimmillaan ympäri Suomen. Tästä syystä niitä ei sisällytetä taulukkoon. Minimien pienistä eroista huolimatta myös ne ovat pääosin samassa suuruusjärjestyksessä kuin tapauskohtaiset maksimit sekä keskiarvot. Tapauskohtaiset maksimi-arvot vaihtelevat välillä 251–4367 ja jakautuvat suhteellisen tasaisin välein lukuun ottamatta muutamaa pienintä arvoa, jotka ovat Pohjois-Suomesta. Keskiarvot ovat lähes samassa suuruusjärjestyksessä ja niitä koskee sama ilmiö, jossa tapaukset 5 3 (keskiarvo 90) sekä 4 6 (keskiarvo 44) ovat huomattavasti pienempiliikenteisiä, kun taas muuten keskiarvot saavat arvoja tasaisesti väliltä 234–1346.

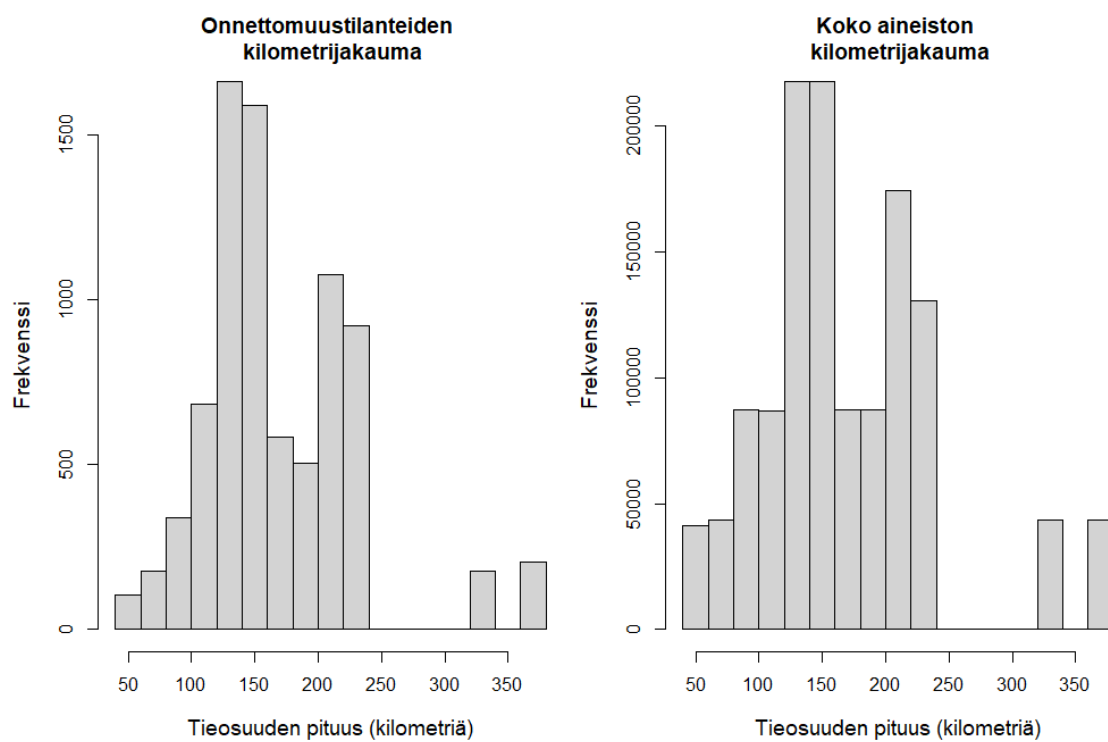
Taulukossa 2 on näkyvissä liikennemäärälukujen lisäksi jokaisen tapauksen alueella tapahtuneiden onnettomuuksien lukumäärät sekä tapaukseen vaikuttaneiden LAM-



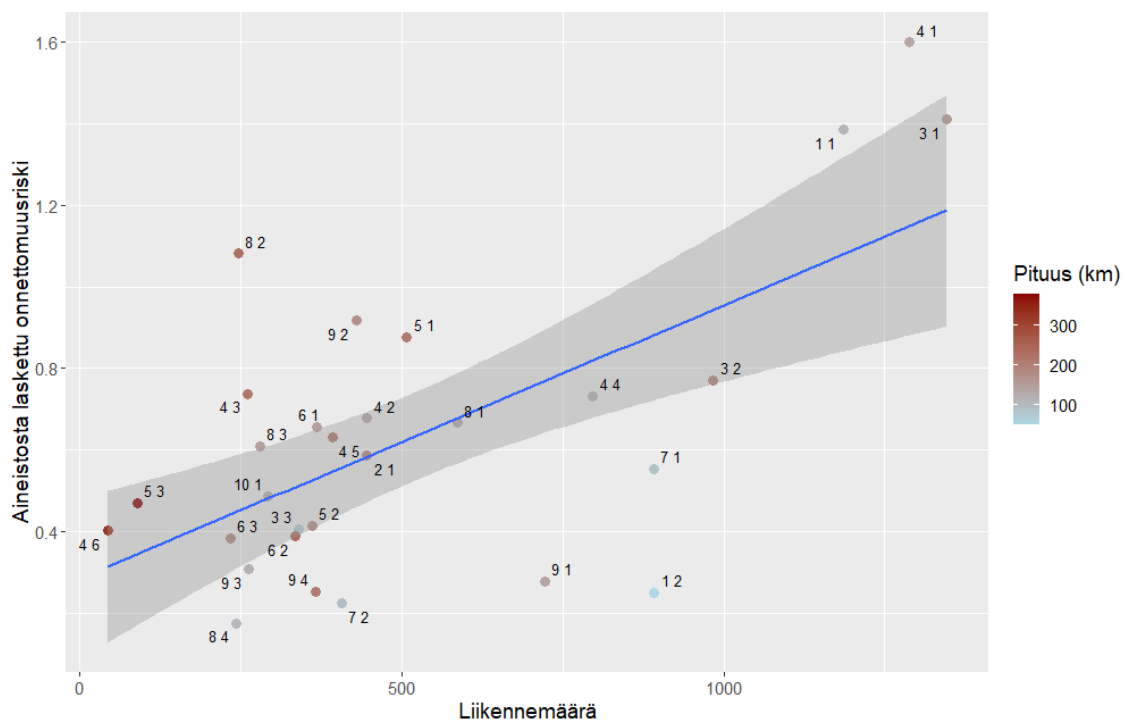
Kuva 4: Pystyakselilla aineistosta laskettu viikonpäiväkohtainen onnettomuuksien määrä ja vaaka-akselilla päiväkohtainen liikennemäärän keskiarvo. Kuvan viiva esitöimöitu pienimmän neliösumman metodilla sekä tummemmalla taustalla 95 % luotamusväli tälle estimaatille.

pisteiden lukumäärät. Kolmessa liikenteeltään suurimmassa tapauksessa tapahtui yli 600 onnettomuutta ja tapauksessa 4 1, jopa 701 onnettomuutta. Neljänneksi eniten onnettomuuksia tapahtui liikennemäärän keskiarvon mukaan vasta 25. ruuhkaisimmassa tapauksessa (tapaus 8 2), jossa tapahtui 474 onnettomuutta, ja vaikka kyseisen tapauksen tieosuuden pituus onkin 229 kilometriä, on tämä silti poikkeuksellisen korkea luku.

Aineistossa tapahtuneiden onnettomuuksien perusteella lasketut tapauskohtaiset onnettomuusriskit on piirretty liikennemäärää vastaan kuvassa 6. Kuvassa jokaista tapausta edustavat pisteet on värjätty tapauksen pituuden mukaan. Pääosa tapauksista on suhteellisen lähellä regressioviivaa. Viivan päällä on enimmäkseen tieosuudeltaan pidempiä ja vastaavasti alla tieosuudeltaan lyhyempiä tapauksia. Vaikka tapauskohtaista vaihtelua on myös kuvaajassa käytettyjen selittäjien lisäksi muilla selittäjillä, näiden muuttujien suhteen yhteys on havaittavissa myös tässä kuvaajassa.



Kuva 5: Vasemmalla onnettomuustilanteiden kilometrijakauman histogrammi ja oikealla vastaavasti koko aineiston kilometrijakauman histogrammi.



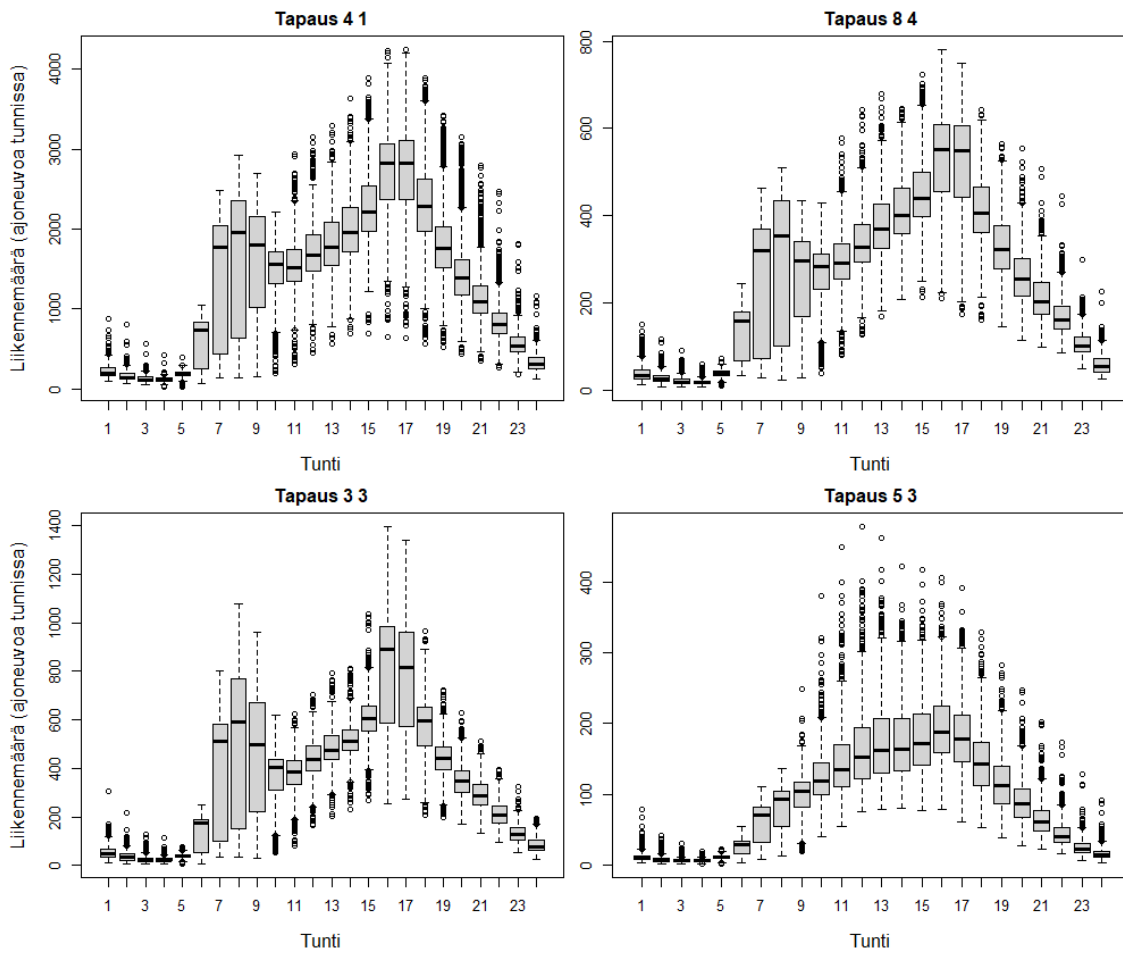
Kuva 6: Pystyakselilla aineistosta laskettu onnettomuuden riski ja vaaka-akselilla tapauskohtainen liikennemäärän keskiarvo. Kuvan viiva estimoitu pienimmän neliösumman mukaan sekä tummemmalla taustalla sen 95 % luottamusväli. Pisteiden väritys tapauksien pituuden mukaan.

2.5 Esimerkkitapaukset

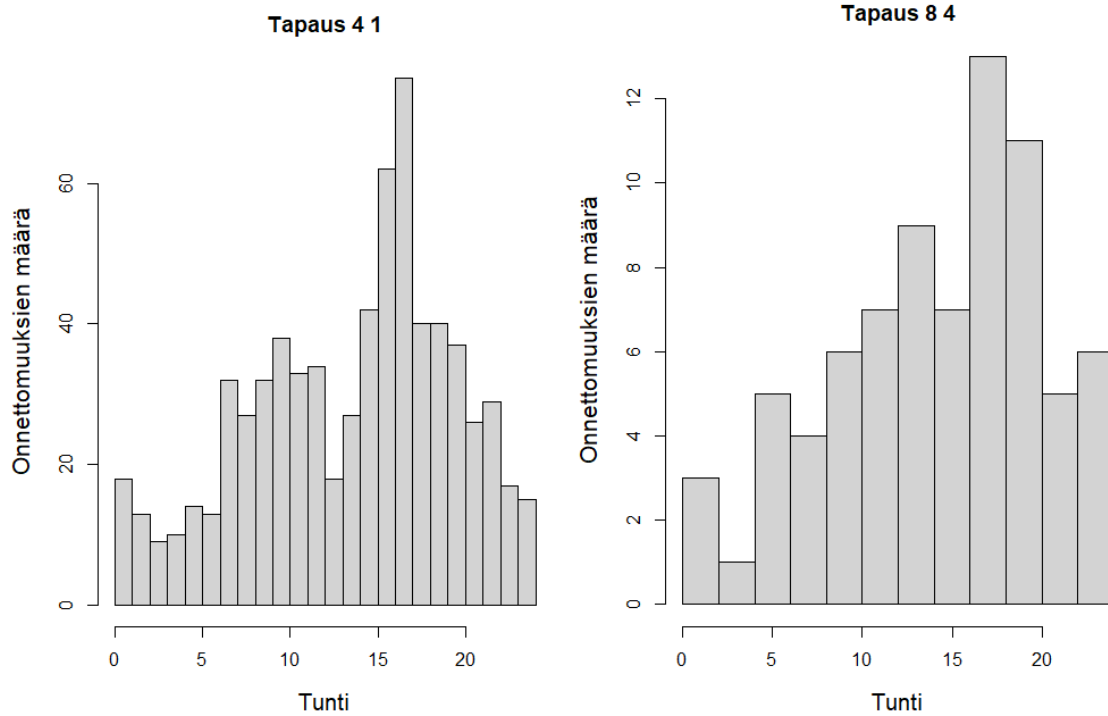
Tarkastellaan sekä vertaillaan hieman kahden esimerkkitapauksen jakaumia eri muuttujilla. Käytetään esimerkkitapauksina tapausta 4 1 sekä tapausta 8 4. Tapaukseen 4 1 sisältyy 7 kuntaa ja 15 LAM-pistettä, kun taas tapaus 8 4 sisältää 2 kuntaa ja 3 LAM-pistettä. Tapaus 4 1 on onnettomuusrikkain 701 onnettomuudella ja tapauksessa 8 4 on tapahtunut vähiten onnettomuuksia (77).

Kuvassa 7 on otettu tarkasteluun lisäksi tapaukset 3 3 sekä 5 3, joista jälkimmäisen jakauma on hyvin samanlainen kuin tapauksen 4 6 jakauma. Aineiston matalimman liikennemäärän tapausten jakaumat ovat siis ainoat merkittävästi poikkeavat, sillä loppujen tilanteiden laatikkokuviot näyttävät erittäin samanlaisilta. Pienenä erona esitetyistä tapauksista on se, että tapauksessa 3 3 on suhteessa suuremmat huiput ruuhkaisimpien aikojen eli aamun ja iltapäivän kohdilla suhteutettuna loppupäivän liikennemääriin muissa tapauksissa.

Kuvassa 8 esillä esimerkkitapausten 4 1 sekä 8 4 tuntikohtaiset onnettomuudet. Kuvasta 8 havaitaan, että yksittäisten tapausten onnettomuudet ovat jakautuneet aikaan nähden pääosin samalla tavalla kuin koko aineiston onnettomuudet. Esimerkkitapauksista tapauksessa 4 1 tapahtuu aineiston suurin määrä onnettomuuksia (701) ja vastaavasti tapauksessa 8 4 niitä on tapahtunut vähiten. Molempien jakaumat ovat kuitenkin pääpiirteittäin samanlaisia kuin koko aineiston tuntikohtainen histogrammin jakauma kuvassa 3.



Kuva 7: Esimerkkitapausten tuntikohtaisten liikennemäärien laatikkokuviot. Ylärivissä vasemmalla tapaus 4 1 ja oikealla tapaus 8 4. Alarivissä vasemmalla tapaus 3 3 ja oikealla tapaus 5 3.



Kuva 8: Esimerkkitapausten histogrammi onnettomuuksien määristä eri kellonai-
koina. Vasemmalla tapaus 4 1 ja oikealla tapaus 8 4. Vasemmanpuoleisessa jokaisella
tunnilla on oma palkki ja oikealla palkit kahden tunnin ryhmissä.

3 Menetelmät

Tutkielman tavoitteena on mallintaa onnettomuuksien tapahtumistodennäköisyyttä. Mallinnukseen käytetään logistista regressiota. Pohjustetaan logistisen regression teoria lineaarisen mallin avulla. Avataan myös vetokerrointen teoriaa, jotta voidaan tulkita tuloksia hieman eri näkökulmasta. Tämän luvun regressiomallien teoria pohjautuu teokseen Hamilton (1992) [5].

3.1 Lineaarinen regressiomalli

Esitetään ensin esimerkki lineaarisesta mallista muuttujien \mathbf{X} ja Y , jossa $K - 1$ selittäjää

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_{K-1} X_{K-1}. \quad (1)$$

Tällöin kaavassa 1 kaikki β_i ovat vakioita ja X_i selittäjien arvoja, joilla pyritään ennustamaan Y :n odotusarvoa. β_0 kuvaa siis Y :n odotusarvoa tilanteessa, jossa $\mathbf{X} = 0$. Kun taas jokainen β_k kertoo kyseisen X_k :n yhden yksikön muutoksen keskimääräisen vaikutuksen Y :hyn, olettaen että muut X :n arvot pysyvät muuttumattomina. Mallista saatavaan estimaattiin sekä havaittuun arvoon Y_i liittyy virhetermi

$$\varepsilon_i = Y_i - (\beta_0 + \beta_1 X_{i1} + \dots + \beta_{K-1} X_{i,K-1}) = Y_i - E[Y_i].$$

Näiden virheiden neliösumma $\sum_{i=1}^n \varepsilon_i^2$, missä n on havaintojen lukumäärä, kuvaa mallin tuottamien ennusteiden tarkkuutta, sekä myös koko mallin tarkkuutta estimoida kyseistä aineistoa tai tilannetta. (Hamilton, 1992, s.30-33)

3.2 Logistinen regressiomalli

Logistisessa regressiomallissa Y voi saada arvokseen vain 0 tai 1. Mallintaessa estimoidaan tilanteen $Y = 1$ todennäköisyyttä, eli tässä tutkielmassa sitä, että onnettomuus tapahtuu kyseisellä hetkellä ja tieosuudella. $P(Y = 1)$ kertoo mallin todennäköisyyden sille, että arvoksi saadaan 1. Vastaavasti vastakohtaan todennäköisyys on tällöin

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

Vetokerroin (engl. odds) sille, että $Y = 1$ on

$$O(Y = 1) = \frac{P(Y = 1)}{1 - P(Y = 1)}.$$

Vetokerrointen on mahdollista saada arvoja nolasta (kun $P(Y = 1) = 0$) äärettömään (kun $P(Y = 1) = 1$). Olkoon $P(Y = 1)$ vain P ja vastaavasti $O(Y = 1)$ pelkästään O . Tällöin saadaan logit-muunnos, kun otetaan luonnollinen logaritmi vetokertoimesta

$$L = \log O = \log\{(P)/(1 - P)\}.$$

Tällöin logit voi saada arvoja aina negatiivisesta äärettömästä (kun $P = 0$) äärettömään (kun $P = 1$). Logit-mallin tapauksessa estimointi toteutetaan useimmiten suurimman uskottavuuden avulla, toisin kuin lineaarisissa malleissa, joissa käytetään pienintä neliösummaa mallin sovittamiseen. Uskottavuusfunktiolla saadaan todennäköisyys saada havaittu otos havaittujen parametrien funktiona. Vektori \mathbf{X}_i sisältää i :n tilastoyksikön selittäjien arvot. Logit-mallin ehdollinen todennäköisyys i :n havainnolle

$$P_i = P(Y_i = 1|\mathbf{X}_i) = \frac{1}{1 + e^{-L_i}},$$

missä

$$L_i = \beta_0 + \sum_{k=1}^{K-1} \beta_k X_{ki},$$

missä $K - 1$ on taas selittävien muuttujien lukumäärä. β :t edustavat malliparametreja, jossa β_0 edustaa vakioarvoa.

Yksittäisen tilastoyksikön vaikutus uskottavuusfunktiioon on $P_i = P(Y_i = 1|\mathbf{X}_i)$ ja $1 - P_i = P(Y_i = 0|\mathbf{X}_i)$. Tämä voidaan esittää myös seuraavasti

$$P_i^{Y_i} (1 - P_i)^{1 - Y_i}.$$

Kun voidaan olettaa havaintojen riippumattomuus, on uskottavuusfunktio tällöin näiden tulo

$$\mathcal{L} = \prod_{i=1}^n \{P_i^{Y_i} (1 - P_i)^{1 - Y_i}\}. \quad (2)$$

β -parametreille pyritään löytämään estimaatit, joilla maksimoidaan uskottavuusfunktion arvot (kaava 2). Tämä toteutuu maksimoitaessa uskottavuuden logaritmi (engl. log likelihood)

$$\log \mathcal{L} = \sum_{i=1}^n \{Y_i \log P_i + (1 - Y_i) \log (1 - P_i)\}.$$

Otetaan ensimmäinen derivaatta log-uskottavuusfunktiosta siten, että asetetaan estimoitujen parametrien derivaatat nolliksi, jotta löydetään suurimman uskottavuuden estimaatit. Tästä saadaan seuraavat yhtälöt

$$\sum_{i=1}^n (Y_i - P_i) = 0 \text{ ja } \sum_{i=1}^n (Y_i - P_i) X_{ki} = 0, \text{ kun } k = 1, 2, \dots, K - 1.$$

Näiden yhtälöiden epälineaarisuuden vuoksi niitä ei useimmiten pystytä ratkaisemaan suoraan ja siksi niiden ratkaisuun käytetään iteratiivisia menetelmiä. (Hamilton, 1992, s.220-224)

Logit-malleja tai tarkkaan ottaen mallin β -kertoimia voidaan tulkita seuraavalla kolmella tavalla: vetokertoimilla tai vetokerrointen suhteilla; logit tai log vetokertoimilla (L) sekä todennäköisyyksillä $\hat{P} = 1/(1 + e^{-\hat{L}})$. (Hamilton, 1992, s.230)

3.3 Akaiken informaatiokriteeri

Mallien vertailu toteutetaan Akaiken informaatiokriteerin eli AIC:n (englanniksi Akaike's information criterion) avulla. Akaiken informaatiokriteerissä vertaillaan, kuinka hyvin mallit kuvaavat aineistoa. Akaiken informaatiokriteerin esitteli Hirotugu Akaike vuonna 1973. Siitä on myös johdettu eri versioita. AIC:n arvo lasketaan uskottavuuden logarimin avulla seuraavasti:

$$\text{AIC} = 2\{-\log\mathcal{L} + K - 1\},$$

missä $K - 1$ edustaa selittäjien lukumäärää. Malleja verrattaessa AIC:n arvon tulee olla mahdollisimman pieni. (Davison, 2003, s. 152)

4 Tulokset

Tässä luvussa käsitellään mallin valintaa sekä tulkintaa. Tarkastellaan, miten valitun mallin estimoidut onnettomuustodennäköisyydet muuttuvat arvojen vaihtuessa. Mallin valinnassa haasteellista on useiden mallien samantasoiset AIC-arvot sekä merkitsevyytasot pääosassa muuttujista. Yleisenä muistutuksena kaikista onnettomuusestimaatteja edustavista prosentiarvoista, että kyseessä on yksi tilanne. Kyseessä on siis annetuilla muuttujilla saatava tuntikohtainen riski, että onnettomuus sattuu kyseisellä tieosuudella.

Lopullinen malli valitaan tulkintojen käytännöllisyyden sekä kiinnostavuuden perusteella. Aineiston koon ansiosta muuttujat saavat mallista toiseen lähes poikkeuksetta p-arvoiksi jopa alle 10^{-5} (katso taulukko 6) ja pääosin huomattavasti pienempiä arvoja. Tutkimuskysymyksen sekä ongelman takia malliin valittiin muuttujat liikennemäärä (ajoneuvoja tunnissa) sekä tapauskohtainen pituus kilometreinä. Tämä siksi, että tieosuuden pituus täytyy sisällyttää malliin, koska onnettomuus tapahtuu todennäköisemmin pidemmällä tieosuudella, mikäli tilanteet ovat muutoin identtisiä.

Näiden muuttujien jälkeen siirrytään tarkastelemaan muiden muuttujien sopivuutta malliin. LAM-pisteiden taustatiedoissa oli saatavilla kaistamäärä, mutta tässä on haasteena se, että jo samassa kunnassa ja tiellä saattaa olla suurta vaihtelua kaistamäärissä. Esimerkiksi Espoossa tiellä 1 kaistamäärä on 2–9. Lopullisessa aineistossa on siis hyvin haastavaa koota selkeää muuttujaa kaistamäärästä kyseisessä tapauksessa erillisten kuntien välillä, kun jo yhden kunnan alueella vaihtelu saattaa olla hyvinkin suurta.

Samantyyppinen ongelma on myös tiettyyppimuuttujassa, jossa vaihtoehtoina on maantie, moottoriliikennetie sekä moottoritie. Käytettävissä on siis loppujen lopuksi liikennemäärä, tieosuuden pituus (kilometreinä), ruuhkaisuus, vuorokaudenaika ja vuodenaika. Useiden mallien sekä interaktioiden kokeilujen jälkeen malli valitaan kiinnostavuuden, tulkittavuuden sekä sopivuuden perusteella.

Onnettomuutta kuvaavan mallin selittäjiksi valikoitui lopulta liikennemäärä, ruuhkaisuus, tieosuuden pituus kilometreinä ja interaktio vuodenaajan sekä vuorokaudenaajan välillä. Mallin β -kertoimet ovat taulukossa 6. Samassa taulukossa on esitelty myös kaikkien β -estimaattien keskivirhe, muuttujakohtaiset z-arvot sekä p-arvot. Suurin osa β -arvoista koostuu vuodenaajan sekä vuorokaudenaajan interaktion kaikkista yhdistelmistä. On huomioitava, että talvi-ilta puuttuu listasta, sillä se toimii referenssitapauksena.

Aloitetaan tarkastelemalla, millaisia eroja todennäköisyyksissä saadaan, kun muutetaan yksittäisen muuttujan arvoa. Muuttujakohtaisten tarkastelujen muuttujien arvot sekä mallin tuottamat todennäköisyydet on esitetty taulukossa 7.

Taulukko 5: Mallilla lasketut estimaatit onnettomuustodennäköisyyksille, kaikkien tapauksien maksimi sekä keskiarvo liikennemäärille. Näiden lisäksi myös aineistosta laskettu tapauskohtainen onnettomuusriski.

Tapaus	Onnettomuuden estimaatti keskimääräiselle liikenteelle (%)	Onnettomuuden estimaatti maksimaaliselle liikenteelle (%)	Aineistosta laskettu onnet- tomuusriski (%)
1 1	0.781	5.527	1.385
1 2	0.561	2.961	0.251
2 1	0.594	1.521	0.586
3 1	0.959	8.896	1.410
3 2	0.822	5.492	0.771
3 3	0.402	1.141	0.406
4 1	0.879	7.922	1.600
4 2	0.506	1.500	0.678
4 3	0.520	1.312	0.737
4 4	0.616	2.709	0.732
4 5	0.587	1.638	0.630
4 6	0.564	0.950	0.402
5 1	0.616	2.148	0.876
5 2	0.523	1.315	0.415
5 3	0.659	1.240	0.470
6 1	0.483	1.405	0.655
6 2	0.588	1.421	0.388
6 3	0.500	0.995	0.383
7 1	0.605	3.645	0.554
7 2	0.452	1.159	0.224
8 1	0.559	1.661	0.666
8 2	0.546	1.110	1.082
8 3	0.467	0.976	0.609
8 4	0.417	0.819	0.176
9 1	0.615	2.397	0.278
9 2	0.533	1.661	0.917
9 3	0.428	0.977	0.308
9 4	0.566	1.368	0.253
10 1	0.445	0.973	0.486

Taulukko 6: Mallin muuttujakohtaiset regressiokertoimien arvot, keskivirheet, z-arvot sekä p-arvot. Kaikki pyöristetty viiden desimaalin tarkkuudelle.

	beta	keskivirhe	z	p
vakio	-5.66659	0.06061	-93.49458	<0.00001
liikennemäärä	0.00066	0.00002	31.05166	<0.00001
ruuhka	0.47154	0.10846	4.34739	1e-05
pituus (km)	0.00227	0.00018	12.89348	<0.00001
kevät · yö	-0.56188	0.07564	-7.42818	<0.00001
kevät · aamu	-0.82371	0.08963	-9.18979	<0.00001
kevät · päivä	-0.63979	0.07194	-8.89336	<0.00001
kevät · iltapäivä	-0.62075	0.07470	-8.31026	<0.00001
kevät · ilta	-0.44435	0.07106	-6.25295	<0.00001
kesä · yö	-0.27399	0.06940	-3.94774	8e-05
kesä · aamu	-0.78048	0.08924	-8.74536	<0.00001
kesä · päivä	-0.61687	0.07101	-8.68681	<0.00001
kesä · iltapäivä	-0.48351	0.07380	-6.55189	<0.00001
kesä · ilta	-0.30430	0.06692	-4.54747	1e-05
syksy · yö	-0.49661	0.07457	-6.65999	<0.00001
syksy · aamu	-0.23747	0.07418	-3.20116	0.00137
syksy · päivä	-0.44223	0.06844	-6.46149	<0.00001
syksy · iltapäivä	-0.14083	0.06894	-2.04283	0.04107
syksy · ilta	0.15937	0.06134	2.59819	0.00937
talvi · yö	-0.68010	0.07912	-8.59596	<0.00001
talvi · aamu	-0.35955	0.07827	-4.59378	<0.00001
talvi · päivä	-0.25602	0.06594	-3.88256	1e-04
talvi · iltapäivä	-0.05650	0.06675	-0.84642	0.39732

Aluksi valitaan taulukon 7 vertailukohteeksi tilanne, jossa tunnin liikennemäärä on 400 ajoneuvoa, ruuhkaisuus 0.3, tieosuuden pituus 200 kilometriä, vuodenaikana kevät ja vuorokaudenaikana aamu. Liikennemääräksi valittiin 400, sillä useiden tapauksien liikennemäärän keskiarvo on suhteellisen lähellä tätä lukua. Vastaavasti ruuhkaisuusarvoksi valittiin 0.3, sillä se on lähellä ruuhkaisuuden keskiarvoa koko aineistossa (0.282). Myös tieosan pituus valittiin keskiarvon (169 km) sekä sen takia, että se oli tapauskohtaisen tavoitepituuden yläraja. Vuodenajaksi sekä vuorokaudenajaksi valittiin matalimman onnettomuustodennäköisyyden perusteella kevätaamu. Näiden avulla saadaan vertailtua todennäköisyyksiä, kun muutoksia tapahtuu aineiston perusteella yleisempien arvojen läheisyydessä.

Kaikki onnettomuustodennäköisyydet kuvaavat sitä, kuinka todennäköisesti onnettomuus tapahtuu tunnin aikana tietyillä selittäjien arvoilla. Vertailutapauksessa tarkastellaan siis 200 kilometrin pituista tieosuutta, jossa tapauksen keskimääräinen liikennevirta on 400 ajoneuvoa tunnissa, tapauksen liikennevirran maksimi on 1333 ajoneuvoa tunnissa ja ajankohtana on kevätaamu.

Huomiona liikennemäärän arvon muodostamisesta, että ensin lasketaan liikennemäärät kunnan kaikkien LAM-pisteiden välillä. Seuraavaksi otetaan vielä keskiarvo tapaukseen sisältyvien kuntien välillä. Liikennemäärän arvo ei siis tarkoita, että koko tiellä kulkisi vain 400 ajoneuvoa vaan sitä, että kyseisellä osuudella liikennevirran keskiarvo kaikissa mittauspaikoissa on keskimäärin 400 ajoneuvoa. Tässä oletetaan, että LAM-pisteet on sijoitettu liikennevirtaan nähden satunnaisesti. Tämän lisäksi tulee muistaa, että malli ei pyri kuvailemaan onnettomuuksien määrää vaan esiintyvyyttä. Aineistossa samaan tilanteeseen eli samaan tapaukseen ja samalle tunnille osui useampi onnettomuus vain 127 kertaa (onnettomuustilanteita 8027).

Onnettomuuden todennäköisyys vertailutilanteessa on mallin mukaan 0.36 %. Kasvattaessa liikennemäärän arvo 900:aan onnettomuuden todennäköisyys kasvaa huomattavasti (0.50 %). Ruuhkaisuusmuuttujaa nostamalla 0.6:een onnettomuustodennäköisyys kasvaa 0.41 prosenttiin. Toisaalta tavallisessa tilanteessa näiden molempien arvot nousisivat samalla, joten näiden molempien vaikutukset vaikuttaisivat samaan suuntaan, jolloin kolarin todennäköisyys olisi yhä suurempi. Ruuhkaisuuden tuplaantuminen voidaan tulkita yksinään siten, että jos tapauksessa olisi tavallisesti vähemmän liikennettä, kuinka sama liikennemäärä vaikuttaisi tällöin. Pituutta tarkasteltaessa puolitetaan tieosuuden kilometrimäärä, jolloin todennäköisyys laskee 0.28 %:iin.

Taulukossa 7 on myös erikseen vuodenaajan sekä vuorokaudenaajan muutosten vaikutus, sillä vuodenaajan sekä vuorokaudenaajan β -kertoimet ovat pääosin samassa

järjestyksessä. Tällöin interaktion avulla saadaan kaikkiin tilanteisiin kohdistetut estimaatit sen sijaan, että mallissa olisi interaktion sijasta vuodenaika- sekä vuorokaudenaikamuuttujat erikseen. Erityisen kiinnostavaa on, millaiset vaikutukset milläkin yhdistelmällä on. Taulukossa 7 on vertailtu onnettomuuden kannalta riskialtuinta tilannetta vähiten riskialtimpaan tilanteeseen. Kuten taulukosta 7 huomataan, vuodenaajan sekä vuorokaudenaajan interaktiolla saadaan aikaan hyvinkin suuri ero onnettomuuden todennäköisyydessä muuttamatta muiden selittäjien arvoja.

Vaikkakin edeltävistä esimerkeistä suurin vaikutus oli vuodenaajalla, vuorokaudenaajalla sekä erityisesti näiden yhdistelmällä, oli kyseessä onnettomuuden riskitason näkökulmasta ajankohdista vertailussa minimi- sekä maksimitilanne. Muiden muuttujien kohdalla valittiin suhteellisen yleinen muutos. Toisaalta esimerkiksi tieosuuden pituusmuuttujassa vain kahden tapauksen pituus on huomattavasti yli 200 kilometrin, joten pituuden vaikutus on harvoin suurempi kuin annetussa esimerkkitapauksessa. Liikennemäärän ja ruuhkaisuuden sekä erityisesti näiden yhteisvaikutuksen tapauksessa vaikutus onnettomuusestimaattiin voi olla vielä huomattavasti suurempi.

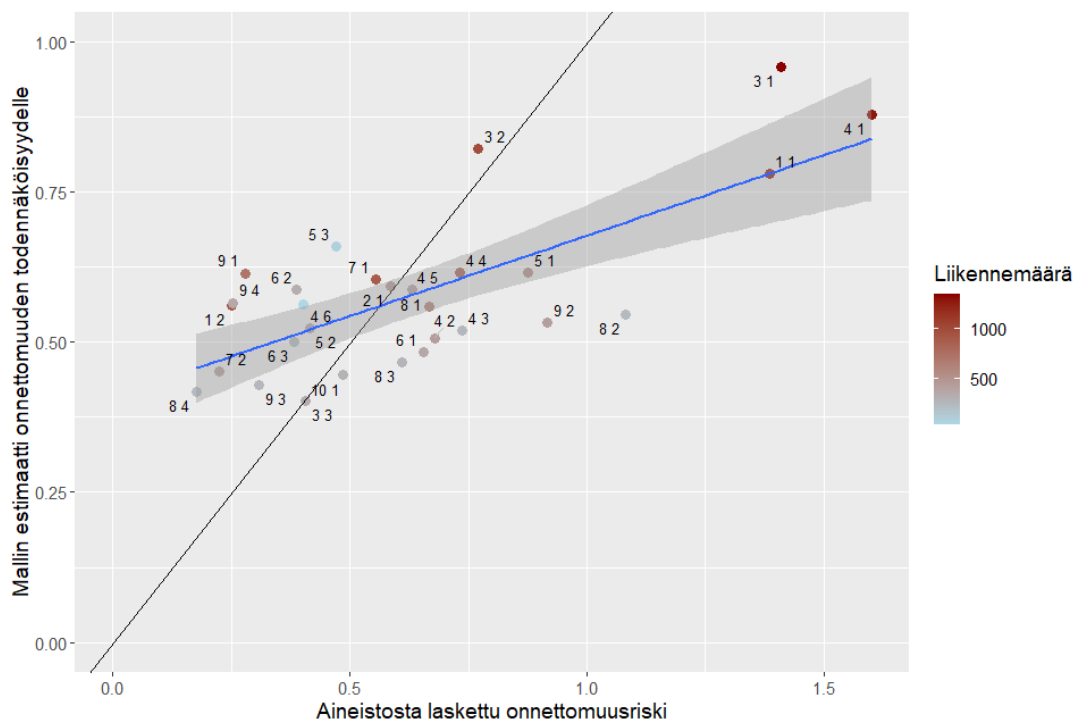
Taulukko 7: Mallilla lasketut todennäköisyydet yksittäisen selittäjän arvon muuttuessa.

Liikennemäärä	Ruuhkaisuus	Pituus (km)	Vuodenaika	Päivänaika	Todennäköisyys (prosenttia)
400	0.3	200	Kevät	Aamu	0.3569
900	0.3	200	Kevät	Aamu	0.4950
400	0.6	200	Kevät	Aamu	0.4109
400	0.3	100	Kevät	Aamu	0.2846
400	0.3	200	Syksy	Aamu	0.6396
400	0.3	200	Kevät	Ilta	0.5207
400	0.3	200	Syksy	Ilta	0.9482

Jokaisen 29 tapauksen liikennemäärän keskiarvolle sekä maksimille on laskettu mallin estimaatit taulukossa 5. Näiden lisäksi taulukkoon 5 on laskettu aineistosta tapahtuneiden onnettomuuksien perusteella riski tieliikenneonnettomuudelle kyseisellä tieosuudella tunnin aikana. Pituutena on käytetty jokaisen tapauksen omaa pituutta. Lisäksi ruuhkaisuusmuuttujan arvo on myös laskettu tapauskohtaisesti. Ajankohtana on käytetty talviaamua, sillä tämän β -arvo on lähinnä vuodenaajan ja vuorokaudenaajan interaktion β -arvojen puoliväliä ja näin ajankohdan vaikutus saadaan minimoitua. Käytetään tätä ajankohtaa vertailuun muissakin kuvaajissa samasta

syystä, mikäli ajankohta ei ole keskeisessä roolissa. Näin tapauskohtaiset estimaatit saadaan nollan molemmin puolin, jolloin luvut ovat paremmin vertailtavissa.

Kuvasta 9 on havaittavissa, että malli estimoii tapauskohtaisia arvoja suhteellisen hyvin, sillä suurin osa pisteistä on melko lähellä viivaa. Muutama piste karkaa suhteellisen kauas, mutta pääosin pisteet pysyvät lähellä regressioviivaa. Pisteistä itessään on havaittavissa trendi, vaikkakin tapauksia erikseen mallintaessa saataisiin varmasti tarkempia estimaatteja. Tämä malli on kuitenkin helpommin yleistettävissä, sillä tarvittavia tietoja ovat ainoastaan ajankohta, liikennemäärä ja liikennemäärän maksimi muutaman vuoden ajalta, jotta saadaan jo laskettua estimaatteja onnettomuuksien todennäköisyydelle.



Kuva 9: Pystyakselilla mallin perusteella lasketut estimaatit onnettomuuksien todennäköisyyksille kaikille tapauksille. Vaaka-akselilla aineistosta lasketut tapauskohtaiset todennäköisyydet onnettomuuksille. Pallojen värit määrittyvät tapauksen keskimääräisen tuntikohtaisen liikennemäärän perusteella. Kuvan viiva on piirretty pienimmän neliösumman perusteella ja tummemmalla taustalla näkyy 95 % luottamusväli tälle viivalle. Origin kautta kulkeva viiva on kulmakertoimella yksi piirretty suora, jonka tarkoituksena on havainnoida tilannetta, jos estimaatit olisivat onnistuneet täydellisesti.

Tarkastellaan liikennemäärän sekä ruuhkaisuuden vaikutusta kuvaajan 10 avulla. Kuvaaajasta voi tarkastella eri liikennemäärien todennäköisyyksiä eri ruuhkaisuusmäärillä. Ylin viiva edustaa 2000 ajoneuvoa tunnissa talviaamuna kahdensadan ki-

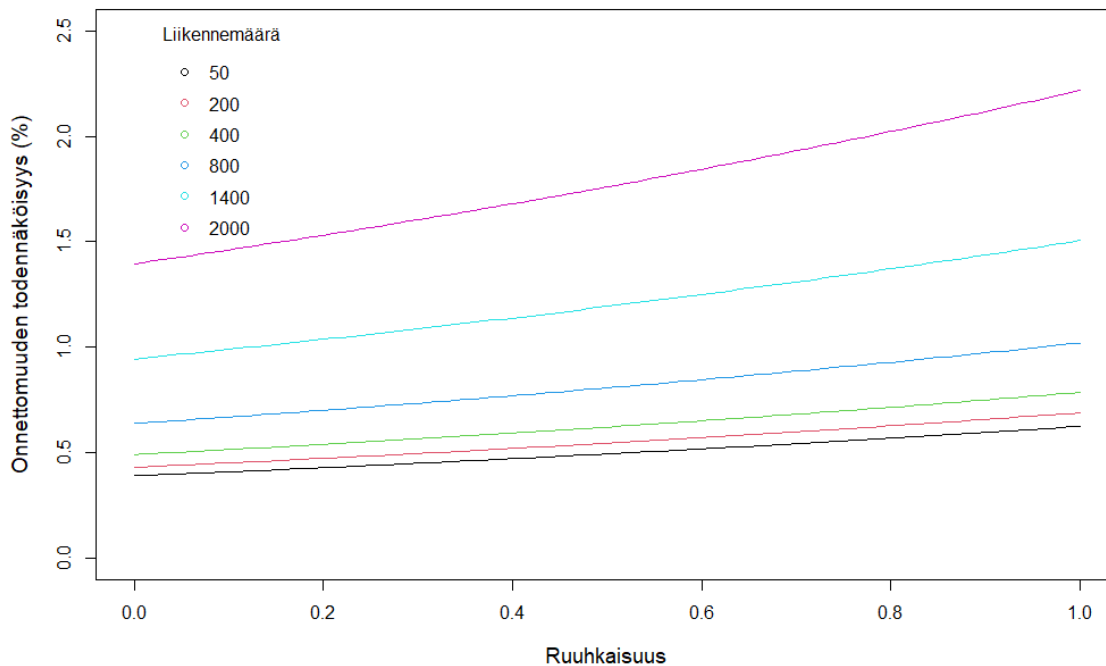
lometrin pituisella valtatieosuudella. Pääosin visualisoidut tilanteet ovat saavutettavissa muutamaa ääritapausta lukuunottamatta. Esimerkiksi 2000 ajoneuvon tilanteen onnettomuusestimaatin arvoja on käytännössä mahdoton saavuttaa Suomessa huomattavasti alle 0.5 ruuhkaisuudella, sillä Suomen liikenneerikkaimpien teiden liikennemäärän volyyymi on 4000 ajoneuvon luokkaa.

Tämän kuvaajan avulla pystytään havainnoimaan, miten mallin mukaan onnettomuuden riski kasvaa, kun lähestytään tarkasteltavan tapauksen liikennemäärän maksimia. Tätä voidaan käytännössä tulkita siten, että tavallisesti tiet on suunniteltu käsittelemään tiettyä liikennemäärää ja mikäli lähestytään tämän tien liikennekapasiteetin ylärajaa, onnettomuuksien riski kasvaa. Siispä tieosuuksilla, joilla on tavallisesti vähemmän liikennettä, sattuu samalla liikennemäärällä todennäköisemmin onnettomuuksia, kuin suuremman liikennekapasiteetin teillä. Vaikka liikennemäärän vaikutus estimaatteihin onkin enimmillään suurempi, myös ruuhkaisuudella eli tien liikennemäärän kapasiteetilla on tämän mallin mukaan selvä merkitys onnettomuuden todennäköisyyteen.

Vertaillaan, miten suuria liikennemäärän sekä ruuhkaisuuden vaikutukset onnettomuusestimaatteihin ovat suhteessa toisiinsa. Mallin mukaan lähes sama onnettomuusestimaatti saavutetaan seuraavissa tilanteissa: liikennemäärä 781, ruuhkaisuus 1, onnettomuuden todennäköisyys 1.01 % sekä liikennemäärä 1274, ruuhkaisuus 0.3, onnettomuuden todennäköisyys 1.00 %. Edellä mainitussa tilanteessa on siis tarkasteltu tapauksia, joista ensimmäisessä liikennemäärän maksimi on 781 (katso taulukko 2 tapaus 8 4) ja toisessa 4247 (tapaus 4 1). Näissä tapauksissa siis noin 500 ajoneuvon ero liikennemäärässä vastaa 0.7 eroa ruuhkaisuusmuuttujassa.

Pyritään siis selvittämään ajankohtien vaikutusta, eli miten suuri riski vuodenaikojen ja vuorokaudenaikojen interaktioilla on. Tämän lisäksi tarkastellaan myös molempien muuttujien vaikutusta kauttaaltaan. Taulukossa 8 on esillä mallista saatavat estimaatit eri ajankohdille, kun liikennemäärä on 400, ruuhkaisuus 0.3 ja pituus 200. Muiden muuttujien arvoina käytetään siis samoja arvoja kuin taulukon 7 vertailutapauksessa. Vaikka mallissa ei olekaan vuodenaajan ja vuorokaudenaajan muuttujia eriteltynä interaktiosta, voidaan taulukon 8 avulla tarkastella rivin sekä sarakkeen perusteella eri ajankohtien trendejä.

Vuodenaajoista onnettomuusestimaatteihin suurimman riskin aiheuttaa syksy. Suhteellisen lähellä syksyn riskitasoja on talvi. Keväällä riskit ovat kauttaaltaan matalimmat tai toiseksi matalimmat. Ero kevään ja kesän sekä syksyn ja talven onnettomuusestimaattien välillä on huomattava. Esimerkiksi keskimääräinen ero kevään ja syksyn onnettomuusestimaattien välillä on 0.22 %.



Kuva 10: Kuvassa näkyy, kuinka onnettomuuden todennäköisyys muuttuu liikennemäärän sekä ruuhkaisuuden muuttuessa. Pystyakselilla estimoitu todennäköisyys onnettomuudelle ja vaaka-akselilla ruuhkaisuuden arvo. Jokainen viiva on piirretty eri liikennemäärän arvolla, jotka ovat esillä kuvassa. Pituusmuuttujan arvona kuvaajassa käytettiin 200 km ja ajankohtana talviaamua.

Vuorokaudenajoista matalimmat onnettomuusestimaatit kasvavassa järjestyksessä ovat aamulla, yöllä sekä päivällä (vertailtaessa estimaattien keskiarvoja). Iltapäivän onnettomuusestimaatit ovat selvästi korkeammat ja iltaisin riski on yhä suurempi. Iltaisin riskit ovat systemaattisesti kaikista korkeimmat. Ainoana poikkeuksena on kesäaamu, joka on hieman korkeampi kuin vastaava ilta. Muistutuksena, että iltapäivisin on korkeimmat liikennemäärät, joten mallin mukaan iltaisin tapahtuu suhteessa eniten onnettomuuksia suhteutettuna muihin muuttujiin.

Selkeästi koko taulukon 8 korkeimmat riskit ovat edellä mainittujen riskialttiimpien ajankohtien interaktioissa eli talvisin sekä syksyisin iltapäivällä sekä illalla. Trendit ovat pääosin aika selkeitä. Poikkeuksina kesäaamu, jonka onnettomuusriski on oletettua suurempi, sekä talviaamu, jonka onnettomuusriski on oletettua alhaisempi.

Taulukko 8: Mallin estimaatit eri ajankohdille. Muiden muuttujien arvot ovat kaikissa ajankohdissa samat: liikennemäärä 400, ruuhkaisuus 0.3 ja pituus 200. Estimaatit on esitetty prosentteina.

	Yö	Aamu	Päivä	Iltapäivä	Ilta
Kevät	0.4632	0.3569	0.4286	0.4368	0.5207
Kesä	0.6168	0.3726	0.4385	0.5008	0.5985
Syksy	0.4943	0.6396	0.5218	0.7040	0.9482
Talvi	0.4118	0.5665	0.6279	0.7655	0.8096

5 Johtopäätökset ja pohdinta

Tässä kappaleessa mainitaan joitain tutkielman vaiheita sekä käsitellään erityisesti sitä, mitä mallin avulla voidaan tulkita onnettomuuksista ja tarkalleen ottaen onnettomuuksien syistä. Tämän lisäksi pohditaan myös, miten tieliikenneonnettomuustutkimusta voisi laajentaa tai tarkentaa tutkielman aikana tehtyjen havaintojen perusteella.

Aineiston laajuuden ansiosta voidaan testata myös tarkempia muuttujia ja muuttujille saadaan luotettavia arvoja. Aineiston laajuus on viisi vuotta ja tapauksia on 29. Dataa on eri puolilta Suomea ja se on kerätty tarpeeksi pitkältä aikaväliltä. Valtateiden suhteen isolla skaalalla mallinnusta voidaan pitää suhteellisen tarkkana.

Mallinnus toteutettiin logistisella regressiolla, sillä sen avulla pyritään estimoimaan ilmenemistodennäköisyyttä (0 ei onnettomuutta ja 1 onnettomuus). Mallin perusteella liikennemäärällä sekä ruuhkaisuudella on selvä vaikutus onnettomuuksien ilmenemiseen. Tämä oli tutkielman ensisijainen tarkastelun kohde ja aineiston sekä mallin perusteella tutkielma antaa hyvin vahvan yhteyden liikennemäärän sekä onnettomuustodennäköisyyden välille.

Tämän lisäksi myös vuorokaudenaika ja vuodenaika vaikuttavat selvästi onnettomuustodennäköisyyteen. Interaktion avulla saatiin myös selville, mitkä yhdistelmät ovat riskialtimpia ajankohtia. Interaktiota käyttäen saatiin myös tarkempia estimaatteja, minkä ansiosta esimerkiksi kevätaamu nousi esille suhteellisen riskialttiina tapauksena, vaikka yleisesti ottaen kevät sekä aamu ovat matalan riskin ajankohtia.

Onnettomuuksien estimaattien mukaan vuodenaikojen järjestys korkeimmaista riskistä matalimpaan on syksy, talvi, kesä ja kevät. Vastaava järjestys vuorokaudenajoille on ilta, iltapäivä, päivä, yö ja aamu. Molemmissa kategorioissa kahden riskialttimeimman tilanteen jälkeen estimaateissa on huomattava ero.

Mallin muotoilun mukaan myös tieosuuden vaikutus on havaittavissa, vaikkakin vaikutus on odotettua pienempi. Tähän osatekijänä on varmasti se, että samantyyppisillä tapauksilla on yhtä pitkät osuudet, sillä tapauksissa, joista on paljon tietoa, voidaan ottaa lyhyempi tieosuus. Vastaavasti Pohjois-Suomeen osuivat pisimmät tapaukset, joissa on vähiten liikennettä. Tämän lisäksi LAM-pisteitä on harvemmassa, jolloin muut muuttujat selittävät jotakuinkin samaa ilmiötä. Siispä muuttujien vaihtelevuus on keskenään useimmiten samantyylistä.

Edellä esitettyä tieosuuden pituuden haastetta voisi pyrkiä ratkaisemaan vaihtelemalla pituuksia myös samantyylisten tapausten kesken. Toisaalta tieosuuksien pituuden tarpeetonta kasvattamista tulee välttää, sillä tällöin liikennevirran tiedot sekä liikenneolosuhteet muuttuvat, jolloin kyseessä ei ole aivan samanlainen tapaus.

Tapauskohtaisesti onnettomuuksien määrät vaihtelevat huomattavasti ja tämän lisäksi malli myös yliarvioi tai aliarvioi joidenkin tapausten onnettomuustodennäköisyyksiä, jopa huomattavia määriä. Mallien onnettomuuksien estimaattien summa on sama kuin oikeiden onnettomuuksien. Mikäli vastaavanlaisen tutkimuksen haluaisi toteuttaa uudestaan, olisi hyvä keskittyä muuttujiin, joilla saataisiin tapauskohtaisia onnettomuusestimaatteja lähemmäksi tapauskohtaisia onnettomuusriskejä. Aineiston perusteella mallissa käytettyjen muuttujien lisäksi näyttää siltä, että jotkut tapaukset ovat turvallisempia ja toiset riskialttiimpia. Varmaa syytä tälle on vaikea löytää ilman dataa, mutta mahdollisia syitä ovat esimerkiksi nopeusrajoitukset, kaistamäärät, tien tyypit (moottoritie, moottoriliikennetie jne.) sekä muut liikennejärjestelyt.

Näiden muuttujien sisällyttäminen malliin olisi hyvin kiinnostavaa. Kaistamäärän luulisi olevan hyvä muuttuja kuvaamaan tien liikennekapasiteettia nopeusrajoituksen lisäksi. Tietyypeistä erityisesti tiet, joiden kaistojen välissä on kaide luulisi vähentävän onnettomuuksien riskiä. Tämä siksi, että kaide todennäköisesti ehkäisee nokkakolareita, jotka aiheutuisivat esimerkiksi rattiin nukahtamisen tai vaarallisen ohituksen takia.

Toisaalta haasteena edellä mainituissa lisättävissä muuttujissa on se, kuinka ne saataisiin sovitettua malliin järkevästi, sillä samalla tiellä sadan kilometrin aikana on varmasti vaihteleva määrä kaistoja sekä todennäköisesti myös useampia tietyyppejä.

Tiehen liittyvien muuttujien lisäksi erityisesti lämpötilan vaikutuksen tarkastelu olisi hyvin kiinnostavaa, mutta se vaatisi sopivan aineiston löytämistä sekä sisällyttämistä mallinnukseen. Koko onnettomuusaineiston lämpötilajakaumaa tarkasteltaessa on havaittavissa, että nollan asteen paikkeilla on tapahtunut paljon onnettomuuksia. Nyt LAM-aineistoon tulisi liittää lämpötilatiedot myös muihin kuin onnettomuusti-

lanteisiin, jotta saataisiin mallinnettua lämpötilan vaikutusta onnettomuusestimaatteihin.

Myös näkyvyys sekä valoisuus ovat mahdollisia tekijöitä, joita voi olla vaikea sisällyttää malliin. Pimeämpinä vuodenaikoina on korkeammat onnettomuusestimaatit ja tämä voisi itsessään olla kiinnostava aihe. Valoisuudessa myös esimerkiksi aurinkoa kohti ajaminen voi olla riskialtista. Tämä erityisesti mikäli kuskilla ei ole käytössä aurinkolaseja, jolloin pelkkä valoisuusaste ei riitä kuvaamaan riskialttiutta. Tässä tutkielmassa käsitellään valtateitä. Pienemmällä teillä on todennäköisemmin huonompi valaistus sekä enemmän kevyen liikenteen väyliä, jolloin onnettomuuksien estimaatit saattaisivat vaihdella verrattuna suurempiin teihin. Myös tien liukkaus kasvattaa sateisella kelillä sekä talvella onnettomuusriskiä.

Suurimpana laajenuksena tutkielmaan olisi pyrkiä mittaamaan myös pienempien teiden osuuksia, jotta malliin saataisiin valtateiden lisäksi pienempien teiden estimaatteja. Näin saataisiin mahdollisesti parempi kuva Suomen teillä tapahtuvista onnettomuuksista.

Viitteet

- [1] A. C. DAVISON: *Statistical Models*. (2003)
- [2] DIGITRAFFIC: *LAM-dokumentaatio*. Luettu: 29.7.2023, <https://www.digitraffic.fi/tieliikenne/lam/#mittauspisteiden-toiminta>
- [3] DIGITRAFFIC: *LAM-tilastohaku*. Luettu: 29.7.2023, <https://tie.digitraffic.fi/ui/tms/history/>
- [4] FINTRAFFIC: *Tieliikennemäärätiedon tuottaminen*. Luettu: 29.7.2023, <https://www.fintraffic.fi/fi/tie/tieliikennemaaratiedon-tuottaminen>
- [5] LAWRENCE C. HAMILTON : *Regression with graphics : a second course in applied statistics*. (1992)
- [6] PETRI JÄÄSKELÄINEN: *Onnettomuudet, joissa raskas ajoneuvo on osallisena*. Liikenneturva, Luettu: 29.7.2023, (<https://www.liikenneturva.fi/tutkimukset/raskaan-liikenteen-liikenneonnettomuudet/>)
- [7] ARTO LUOMA: *Kuolemaan johtaneisiin liikenneonnettomuuksien vaikuttaneita taustatekijöitä 2005–2014*. Onnettomuustietoinstituutti (OTI), (2022). <https://www.lvk.fi/document/318710/2CAFA6300EEB9C65FE7EFA1D90729276024F44EDE293FEB6435E3C075A4E1EEC>
- [8] JUHA LUOMA, HARRI PELTOLA: *Vauhti lisää turmariskiä myös päätteillä*. VTT, (2016), <https://liikennevilkku.fi/2016/10/25/vauhti-lisaa-turmariskia-mys-paateilla/>
- [9] MIA NYHOLM, TUULA TASKINEN: *Väsymys liikenteessä*. Liikenneturva, Luettu: 29.7.2023, <https://www.liikenneturva.fi/liikenteessa/vasymys-liikenteessa/>
- [10] VÄYLÄVIRASTO: *Onnettomuustilastot*. Luettu: 29.7.2023, <https://vayla.fi/vaylista/aineistot/tilastot/tietilastot/liikenneonnettomuudet-maanteilla>
- [11] VÄYLÄVIRASTO: */Tie/Tieliikenneonnettomuudet*. Luettu: 29.7.2023, <https://ava.vaylapilvi.fi/ava/Tie/Tieliikenneonnettomuudet>