

Kalle Mattila

**TYÖTAPATURMA- JA
AMMATTITAUTIVAKUUTUKSEN
VAHINKOJEN VAKAVUUSLUOKAN
MALLINTAMINEN ZERO-INFLATED- JA
ZERO-ALTERED-MENETELMILLÄ**

Informaatioteknologian ja viestinnän tiedekunta
Pro gradu -tutkielma
Joulukuu 2019

Tampereen yliopisto

Informaatioteknologian ja viestinnän tiedekunta

Mattila, Kalle: Työtapaturma- ja ammattitautivakuutuksen vahinkojen vakavuusluokan mallintaminen zero-inflated- ja zero-altered-menetelmillä

Pro gradu -tutkielma, 40 s., 19 liites.

Computational Big Data Analytics

joulukuu 2019

Tiivistelmä

Työtapaturma- ja ammattitautivakuutus on lähes kaikille yrityksille lakisääteisesti otettava vakuutus työntekijöille. Suomessa sattuu noin 130 000 työtapaturmaa ja ammattitautia vuosittain. Se on monelle paljon työllistävälle yritykselle isoimpia vakuutuksia maksutasoltaan. Tutkielmassa tutustutaan, miten työ- ja työmatkatapaturmien vahinkojen vakavuutta pystytään mallintamaan, kun selittävänä muuttujana on ammattiluokka. Lisäksi tutkitaan, onko toimialaluokkien välillä eroja mallintamisen lopputuloksessa. Menetelmiksi valittiin nolla-arvoja huomioon ottavat zero-inflated ja zero-altered-menetelmät, koska selitettävä diskreettimuuttuja sisälsi paljon nolla-arvoja. Yhtenä tutkimuksen tarkastelun kohteena oli myös, onko eri menetelmien antamissa tuloksissa eroja. Tutkielmassa käytetään aineistona Tapaturmavakuutuskeskukselta saatua aineistoa, jossa on kaikki Suomessa vuosina 2016 - 2017 sattuneet työ- ja työmatkatapaturmat.

Tämän aineiston kohdalla eri nolla-arvoja huomioon ottavilla menetelmillä oli eroavaisuuksia niiden antamissa lopputuloksissa. Tutkielmassa havaittiin menetelmien zero-inflated Poisson ja zero-altered Poisson antavat lähes samanlaiset tulokset. Zero-inflated negatiivinen binomi ja zero-altered negatiivinen binomi-menetelmät antoivat myös keskenään lähes samanlaiset tulokset, mutta nämä erosivat merkittävästi zero-inflated Poisson ja zero-altered Poisson-menetelmillä saaduista tuloksista. Tarkempi tarkastelu tutkielmassa tehtiin zero-inflated negatiivinen binomi-menetelmällä. Tutkielmassa tehty malli antoi ammattiluokkien välille tilastollisesti merkitseviä eroja tapaturman vakavuuden todennäköisyyksille. Malli ei huomionnut tapaturman todennäköisyyttä kussakin ammattiluokassa, vaan ainoastaan eroja niiden välillä, kun tapaturma on jo sattunut. Toimialaluokkien tarkastelu toteutetaan jaettuina aineistoina, koska usein toimialassa ammattiluokat painottuvat tiettyihin luokkiin. Toimialaluokkien välillä havaittiin olevan eroavaisuuksia.

Asiasanat Poisson-jakauma, negatiivinen binomijakauma, zero-inflated-mallit, zero-altered-mallit

Tämän julkaisun alkuperäisyys on tarkastettu Turnitin OriginalityCheck –ohjelmalla.

Sisältö

1 Johdanto	5
2 Työtapaturma- ja ammattitautivakuutus	7
2.1 Vakuuttaminen	7
2.1.1 Vakuutusmaksun määräytyminen	7
3 Aineiston kuvaileminen	10
3.1 Aineistoon tehtävät rajaukset	13
4 Menetelmät	15
4.1 Poisson-regressio	15
4.2 Ylihajonta	15
4.3 Negatiivinen binomijakauma	17
4.3.1 Sekoitettu Poisson-gammajakauma	17
4.4 Nolla-arvojen ylliedustus	19
4.4.1 Hurdle-mallit	19
4.4.2 Zero-inflated-mallit	21
4.5 Mallin yhteensopivuuden testaaminen	22
4.5.1 R^2 ja pseudo- R^2 yhteensopivuusasteet	22
4.5.2 Devianssin yhteensopivuusaste	23
4.5.3 Uskottavuusosamäärätesti	23
5 Tulokset	24
5.1 Nolla-arvomenetelmien tulosten erot	24
5.2 Ammattiluokkien välisiä eroja	27
5.3 Toimialaluokkien välisiä eroja	28
6 Yhteenveto	38
Lähteet	40
Liite 1: Taulukoita joihin viitattu tekstissä	41
Liite 2: R-koodi ja tulosteet	50

1 Johdanto

Työtapaturma- ja ammattitautivakuutus on lakisääteinen vakuutus, jonka ottaminen on työnantajan velvollisuus. Työtapaturma- ja ammattitautivakuutuksesta korvataan työtapaturmia, työmatkatapaturmia sekä ammattitauteja. Vuosittain Suomessa tilastoidaan noin 130 000 työtapaturmaa ja ammattitautia. Varsinkin paljon työllistävillä yrityksillä ja julkisyhteisöillä kyseinen vakuutus aiheuttaa merkittävän osan yrityksen vakuutusmaksuista ja usein se onkin isoin vakuutusmaksun aiheuttaja. Suurimmissa työtapaturma- ja ammattitautivakuutuksissa oma vakuutusmaksu perustuu aiheutuneisiin tapaturmiin eli korvauksiin, maksettaviin palkkoihin ja ammattiluokitukseen.

Ajatus tämän tutkielman tekemisestä on tullut esiin vakuutusmeklariyrityksestä, jossa yksi tärkeä osa asiakkaiden palvelua on osata työtapaturma- ja ammattitautivakuutuksen eri piirteet. On myös tärkeää tiedostaa työtapaturmien riskejä ja todennäköisyyksiä eri ammattiluokkien välillä, jotta asiakasta pystytään neuvomaan työturvallisuuden edistämiseksi. Tässä tutkielmassa tarkoituksena onkin tutkia työtapaturma- ja ammattitautivakuutuksen ammattiluokkien välisiä eroja tapaturmien vakavuuksissa ja niiden todennäköisyyksissä. Lisäksi tarkastellaan minkälaisia eroavaisuuksia on toimialojen välillä. Analysointimenetelmänä on Poisson-regressio ja negatiivinen binomiregressio. Tarkempi mallinnus ja tarkastelu tehdään nollarvot huomioon ottavilla menetelmillä, joita ovat zero-inflated Poisson, zero-inflated negatiivinen binomi, zero-altered Poisson sekä zero-altered negatiivinen binomimenetelmät. Yhtenä tarkoituksena on myös tarkastella, minkälaisia eroja tuloksissa näillä menetelmillä saadaan.

Mallinnuksessa käytetään R-ohjelmistoa, jolla pystytään mallintamaan ja tuottamaan erilaisia tilastollisia kuvioita ja laskentataulukoita. Ohjelma on ladattavissa ilmaiseksi internetistä osoitteesta <https://www.r-project.org/>. R-ohjelmaan pystytään lataamaan erilaisia ohjelmakirjastoja ja tämän työn osalta tärkein ohjelmakirjasto on nimeltään pscl-kirjasto, jonka avulla pystytään tuottamaan zero-inflated- sekä zero-altered-malleja. Edellä mainitun lisäksi tutkimuksessa annetaan yleiskuva työtapaturma- ja ammattitautivakuutuksesta sekä työtapaturmista ja työmatkatapaturmista.

Työtapaturma- ja ammattitautilaki muuttui vuoden 2016 alusta. Tapaturmavakuutuskeskus eli TVK otti uuden kansainvälisen ammattiluokituksen 2010 (TK 2010) käyttöön liukuvasti vuosina 2015-2016. Aikaisemmin käytössä ollut ammattiluokitus oli rakenteeltaan erilainen ja siitä johtuen ammattiluokituksen vaihtuminen aiheuttaa katkoksen aikasarjoihin. Tässä tutkielmassa käytetään TVK:lta saatua aineistoa, jossa on tilastoa vuosien 2016 ja 2017 osalta, jolloin on jo käytetty uutta ammattiluokitusta 2010. (TVK 2019)

Aikaisemmin aiheesta on tehnyt vuonna 2010 Pro Gradu -tutkielman Erkka Holopainen. Hänen tutkielmassaan on tutkittu lakisääteisen tapaturmavakuutuksen erikoismaksujärjestelmiä vakuutuksenottajan näkökulmasta. Tällöin on ollut voimassa vielä vanha laki, mutta pääpiirteittäin vakuuttaminen menee uuden lain jälkeen myös

samalla tavalla.

Tutkielma jakautuu kuuteen eri kappaleeseen, joista toisessa kappaleessa esitellään miten korvaukset ja vakuutusmaksut määräytyvät työtapaturma- ja ammattitautivakuutuksessa. Kolmannessa kappaleessa esitellään tarkemmin tutkielmassa käytettävää aineistoa sekä tarkemmin mallintamisessa käytettäviä muuttujia. Neljäs kappale sisältää teoriaosuuden tutkielman mallintamisessa käytetyistä menetelmistä. Viidennessä kappaleessa esitetään varsinaiset tutkimustulokset ja analysoidaan näitä. Lopuksi kappaleessa kuusi tehdään yhteenveto tutkielman tuloksista ja pohditaan mahdollisen jatkotutkimuksen aiheita.

2 Työtapaturma- ja ammattitautivakuutus

Työtapaturma- ja ammattitautivakuutuksen päätarkoituksena on korvauksen maksaminen työntekijälle vahingosta, joka aiheutuu työtapaturmasta tai ammattitaudista. Työmatkalla sattuneet tapaturmat katsotaan myös tästä vakuutuksesta korvattavaksi. Yrittäjän on myös mahdollisuus ottaa vakuutus työtapaturman ja ammattitautin varalta, jos hänellä on yrittäjän eläkelain mukainen vakuutus eli YEL-vakuutus. Toisin kuin työntekijöille otettava vakuutus on lakisääteinen, YEL-vakuutus on vapaaehtoinen vakuutus. Tämän tutkielman aineistossa ei ole käytössä tilastoa ammattitau- tautiteista eikä yrittäjille sattuneista tapaturmista, eikä näitä sen vuoksi erikseen käsitellä. Nykyistä työtapaturma- ja ammattitautilakia 459/2015 on alettu soveltaa 1.1.2016 alkaen. (Salo 2015, TVK 2019)

2.1 Vakuuttaminen

Vakuutuksen hankkiminen on työnantajan vastuulla ja vakuutus pitää ottaa, jos työnantajan teettämän työn työansio ylittää 1200 euroa vuodessa. Työnantajan pitää ottaa vakuutus työtapaturma- ja ammattitautivakuutuksia myöntävästä vakuutusyhtiöstä. Jotta vakuutusyhtiö voi myöntää kyseisen vakuutuksen, pitää sillä olla vakuutusyhtiölain tai ulkomaisista vakuutusyhtiöistä annetun lain mukainen oikeus myöntää työtapaturma- ja ammattitautivakuutus. Vakuutusyhtiön pitää myös olla Tapaturma- vakuutuskeskuksen eli TVK:n jäsen. TVK:n sivuilta löytyy kaikkien työtapaturma- ja ammattitautivakuutuksia myöntävien yhtiöiden yhteystiedot. Mikäli vakuutusyhtiöllä on toimilupa työtapaturma- ja ammattitautivakuutukseen myöntämiseen, sen on pakko se myöntää, jos sitä siltä haetaan. Vakuutusyhtiö ei voi kieltäytyä myöntämästä kyseistä vakuutusta, vaikka työnantaja olisi laiminlyönyt vakuutusmaksujen maksamisen. (Kirsi Salo 2015)

Työtapaturma- ja ammattitautivakuutus on useimmiten jatkuva vakuutus. Vakuutus on voimassa aina vakuutuskauden kerrallaan, joka yleisimmin on kalenterivuosi. Poikkeuksena on, jos vakuutus on uusi alkava vakuutus tai siirretty toiseen vakuutusyhtiöön. Vakuutuksen voi siirtää vakuutusyhtiöstä toiseen neljä kertaa vuodessa. Siirtohakemus tulee olla tehtynä päivämäärään 1.1., 1.4., 1.7. ja 1.10. mennessä ja siirtohakemuksen tekemisestä vakuutus voi siirtyä uuteen vakuutusyhtiöön aikaisintaan kolmen kuukauden kuluttua. Esimerkiksi, jos vakuutuksen siirto tehdään päivämäärällä 25.3, niin voi vakuutus alkaa aikaisintaan 1.7. alkaen. Vakuutuksen pitää olla ollut voimassa yhden kokonaisen kalenterivuoden, jotta vakuutus voidaan siirtää toiseen yhtiöön. (Kirsi Salo 2015)

2.1.1 Vakuutusmaksun määräytyminen

Jokaisella työtapaturma- ja ammattitautivakuutusta myöntävällä yhtiöllä on oltava hallituksen hyväksymät maksuperusteet. Maksuperusteista ilmenee, miten vakuutusmaksut määrätään, ja niitä on sovellettava yhdenmukaisesti kaikkiin vakuutuk-

senottajiin. Maksuperusteet on laadittava niin, että ne ovat kohtuullisessa suhteessa vakuutuksista aiheutuvien kustannuksiin nähden. Tarkoituksena on turvata vakuutusyhtiön toiminta sekä varmistaa korvausten maksukyky vahingoittuneelle. Maksuperusteet eivät ole julkisia, joten vakuutusyhtiöiden maksuperusteiden tarkemmasta sisällöstä ei ole mahdollista kertoa. (www.finlex.fi; Kirsi Salo 2015.)

Vakuutusmaksu voi määräytyä kahdella eri tavalla joko taulustomaksuperusteisesti tai erikoismaksuperusteisesti. Taulustomaksuperusteisessa maksujärjestelmässä vakuutuksen maksu perustuu yleiseen vahinkotilastoon, jolloin omat sattuneet vahingot eivät suoraan vaikuta vakuutusmaksuun. Erikoismaksuperusteisen maksujärjestelmän maksussa yrityksen omat vahingot vaikuttavat vakuutusmaksun määritymiseen. Mikäli yrityksen teettämä työn määrä on riittävän suuri ja tapaturmariskiä voidaan luotettavasti arvioida tilastollisesti, pitää tällöin vakuutukseen soveltaa erikoismaksujärjestelmää. (Kirsi Salo 2015)

Vakuutusmaksu koostu kahdesta osasta; perusmaksu (perustariffi) ja lakisääteiset lisät. Taulukosta 2.1 nähdään mitä näiden kahden osan sisälle kuuluu.

Taulukko 2.1. Vakuutusmaksun osat.

Perusmaksu (perustariffi)	Lakisääteisten lisien ja muiden yhteisesti jaettavien kustannusten maksuosa
<ul style="list-style-type: none"> - Korvausmenot kattava maksuosa, riskimaksu - Hoitokuluosuus - Pääomatuoton tms. perusteella myönnettävä hyvitys 	<ul style="list-style-type: none"> Jakojärjestelmämaksu, TyTAL 231\$, mm - Indeksikorotukset - 10 vuotta tai sitä vanhempien vahinkotapahtumien sairaanhoitokustannukset - Pientyönantajien vahinkotapahtumien korvaukset Työsuojelumaksu - 1,75% pakollisen vakuutuksen vakuutusmaksusta
n. 80 % vakuutusmaksusta	n. 20 % vakuutusmaksusta

Lähde: Kirsi Salo 2015, s. 192

Työtapaturma- ja ammattitautivakuutuksista korvataan seuraavia kustannuksia:

- sairaanhoidon kustannukset
- palkan korvaaminen tutkimusajalta ja fysikaalisen hoidon ajalta
- matka- ja majoituskulut
- kodinhoidon lisäkustannukset vuoden ajalta
- vahinkotapahtuman yhteydessä särkyneet henkilökohtaisessa käytössä olleet silmälasit, kuulokojeet ja muut laissa luetellut vastaavat esineet
- sairaanhoidon antamisen yhteydessä rikutut vaatteet ja rikutun sormuksen korjaus

Edellä mainitut korvaukset jaetaan ohimeneviin ja pysyviin korvauksiin sekä varauksiin. Yleisimmät korvaukset ovat ohimeneviä, esimerkiksi päivärahat, ja pysyviin korvauksiin lasketaan esimerkiksi tapaturmaeläkkeet. Varauksia tehdään, kun vahingon satuttua vakuutusyhtiö tietää jo etukäteen, että tästä vahingosta luultavasti tullaan maksamaan tapaturmaeläkettä, mutta tarkkoja korvausmääriä ei vielä tiedetä. (Kirsi Salo 2015)

Erikoismaksuperusteisessa maksujärjestelmässä on usein katkaisuraja, jolla tarkoitetaan sitä mihin saakka yrityksen omat vahingot vaikuttavat vakuutusmaksuun. Nykyään erikoismaksuperusteisissa vakuutuksissa käytetään useimmiten vahinkotapahtumakohtaista katkaisurajaa, joka tarkoittaa sitä, että yhdestä vahingosta tulleet kaikki korvaukset huomioidaan katkaisurajaan. Samassa vahinkotapahtumassa voi siis olla useampi vahingoittunut. Käytännössä katkaisuraja kertoo yrityksen riskinottohalusta ja riskinkantokyvystä. Usein katkaisuraja on korkeampi isoimmilla yrityksillä ja nämä pystyvätkin ottamaan omalle riskille suuremman osan vahingoittuneen korvauksista. (www.if.fi)

Erikoismaksuperusteisiin maksujärjestelmiin voi sisältyä jälkiseuranta-aika. Jälkiseurannalla tarkoitetaan korvausten seuraamista vakuutuksen päättymisen jälkeen. Esimerkiksi kun vakuutus siirretään toiseen yhtiöön, on vanhalla yhtiöllä oikeus periä maksuja jo sattuneista vahingoista, vaikka korvaukset maksettaisiinkin vakuutuksen päättymisen jälkeen. Jälkiseuranta-aika on useimmiten yhdestä kolmeen vuotta, jolla siis tarkoitetaan, kuinka kauan vanhalla yhtiöllä on oikeus periä maksua vakuutuksen voimassaoloaikana sattuneiden vahinkojen korvauksista. Jälkiseuranta-aikana perittyjä maksuja kutsutaan jälkiperintämaksuiksi. Jälkiperintämaksu voi olla myös miinusmerkkistä eli vakuutuksenottaja saa palautusta liikaa maksetuista vakuutusmaksuista.

3 Aineiston kuvaileminen

Tässä tutkielmassa käytetään Tapaturmavakuutuskeskukselta saatua aineistoa Suomessa sattuneista työtapaturmista ja työmatkatapaturmista. Tämän tiedon TVK on kerännyt vakuutusyhtiöiltä, ja vakuutusyhtiöiden onkin järjestettävä omat tilastointinsa siten, että ne pystyvät toimittamaan TVK:lle sen tarvitsemat tiedot. Aineistoissa on yhteensä 235 348 tilastoyksikköä eli Suomessa sattuneita työtapaturmia ja työmatkatapaturmia vuosien 2016-2017 välisenä aikana. Suurin osa aineistojen muuttujista on kategorisia ja tässä tutkielmassa käytetään vain osaa TVK:n keräämistä tiedoista. Seuraavaksi tarkastellaan muuttujia, joita käytetään tämän tutkielman analysointimenetelmissä.

Ammattiluokitus

Uuden lain tultua voimaan vuoden 2016 alussa, tilastoinnissa on aloitettu käyttämään ammattiluokitusta TK2010. Luokitus 2010 muodostuu viidestä eri hierarkkisesta tasosta:

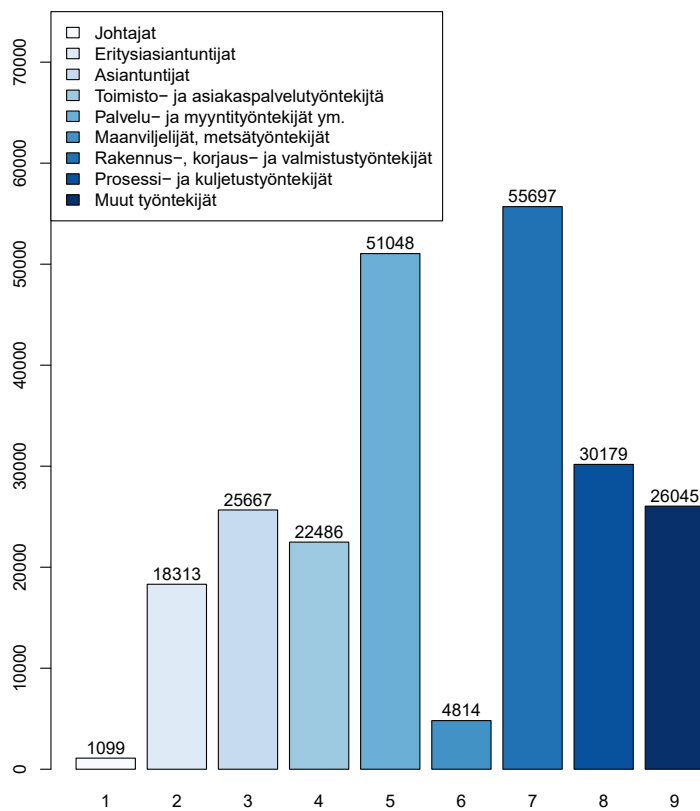
- Pääluokkataso, joka sisältää 10 luokkaa ja johon kuuluvat luokat merkitään 1-numeroisella koodilla
- 2-numerotaso, joka sisältää 43 luokkaa ja johon kuuluvat luokat merkitään 2-numeroisella koodilla
- 3-numerotaso, joka sisältää 130 luokkaa ja johon kuuluvat luokat merkitään 3-numeroisella koodilla
- 4-numerotaso, joka sisältää 436 luokkaa ja johon kuuluvat luokat merkitään 4-numeroisella koodilla
- Kansallinen 5-numerotaso, joka sisältää 103 luokkaa ja johon kuuluvat luokat merkitään 5-numeroisella koodilla.

(Tilastokeskus, <https://www.tilastokeskus.fi>)

Aineistossa kuitenkin on vain neljä eri hierarkkista tasoa. Kansallinen 5-numerotasoa aineistossa ei ole. Muuttujat aineistossa ovat:

- A1LK = pääammattiluokka
- A2LK = ammattiluokka 2-nro kooditaso
- A3LK = ammattiluokka 3-nro kooditaso
- A4LK = ammattiluokka 4-nro kooditaso

Kuviossa 3.1 on havainnollistettu tapaturmien lukumäärä kussakin ammattiluokassa. Voidaan havaita, että ammattiluokissa 1 ja 6, eli johtajat sekä maanviljelijät, metsätyöntekijät ym. sattuu selkeästi vähemmän vahinkoja, johtuen luultavasti näiden tehdystä työn määrästä muihin ammattiluokkiin nähden.



Kuvio 3.1. Vahinkojen lukumäärien jakautuminen ammattiluokkien välillä

Taulukko 3.1. Tapaturmien vakavuuksien luokittelu

Aineiston luokitus	Arvo	Menetelmissä käytetty luokitus
0	0-3 pv	0
1	4-7 pv	5
2	7-14 pv	10
3	15-30 pv	22
4	31-90 pv	60
5	91-180 pv	135
6	yli 180 päivää tai eläke	272
7	kuollut	365

Tapaturmien vakavuuden luokittelumuuttuja

Selitettävänä muuttujana tutkielmassa käytetään vahinkojen vakavuuden ilmaisevaa muuttujaa. Muuttujassa vahingot ovat jaettu kahdeksaan eri luokkaan. Tarkemmin luokitus selviää taulukosta 3.1.

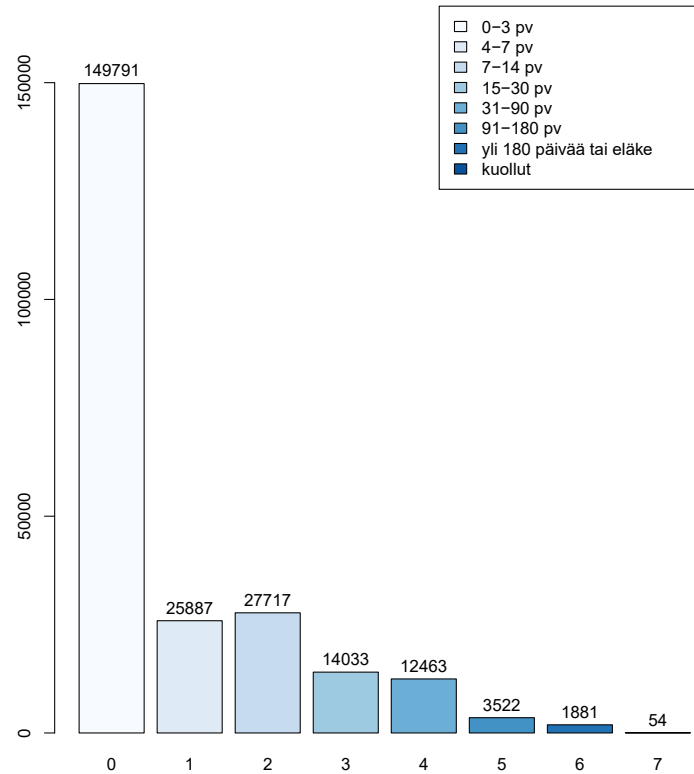
Työtapaturma- tai työmatkatapaturmavahinko voi olla pitkä prosessi, ja vahingoittuneen kuntoutus voi pahimmassa tapauksessa kestää vuosia tai aiheuttaa pysyvän haitan, esimerkiksi peukalon menetyksen. Tämä aiheuttaa tilastoinnissa haasteita, ja TVK seuraakin vahinkojen vakavuutta vuoden ajan ja kiinnittää vakavuuden siten. Aineiston saamisen hetkellä vuoden 2017 vahingot ovat olleet vakavuuksien seurannan osalta hieman keskeneräisiä, jolloin aineistossa vuonna 2017 sattuneet tapaturmat esiintyvät hieman lievempinä ja TVK onkin korjannut lukumääriä selvityskertoimella 1,05.

Kuviossa 3.2 on esitetty aineistossa olevien vahinkojen lukumäärä kussakin luokassa. Kuvioista pystyy selkeästi havaitsemaan, että suurin osa työ- ja työmatkatapaturmista aiheuttaakin ainoastaan vähäisen poissaolon ja kuten luonnollisesti voi ajatella, lukumäärät vähenevät mitä vakavampi vahinko on kyseessä. Tutkielmassa valitut analysointimenetelmät soveltuvat vain rajoitetusti selitettävän muuttujan analysoimiseen, koska se on järjestysasteikollinen. Tutkielmassa kyseinen selitettävä muuttuja onkin tämän vuoksi uudelleen luokiteltu lähemmäs lukumäärää siten

Toimialaluokitus

Toimialaluokituksena käytetään kansainvälistä TOL 2008 luokitusta, joka perustuu EU:n toimialaluokitukseen NACEen. Toimialaluokitus muodostuu viidestä eri hierarkkisesta tasosta

- kirjaintaso, johon kuuluvat kirjainkoodilla merkittävät pääluokat
- 2-numerotaso, johon kuuluvat luokat merkitään 2-numeroisella koodilla
- 3-numerotaso, johon kuuluvat luokat merkitään 3-numeroisella koodilla
- 4-numerotaso, johon kuuluvat luokat merkitään 4-numeroisella koodilla



Kuvio 3.2. Vahinkojen lukumäärien jakautuminen vakavuusluokkien välillä

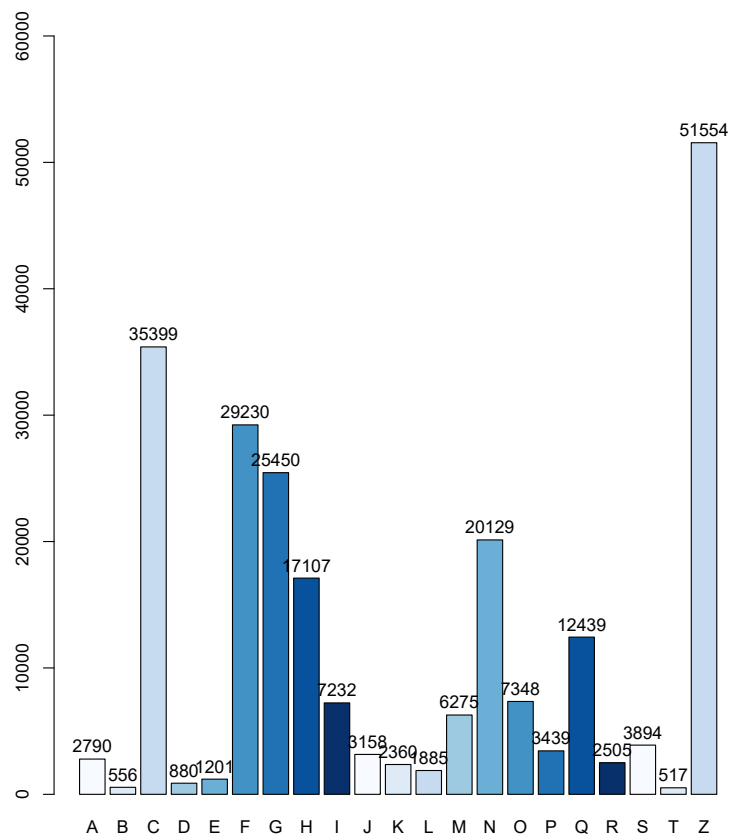
- kansallinen 5-numerotaso, johon kuuluvat luokat merkitään 5-numeroisella koodilla.

(Tilastokeskus, <https://www.tilastokeskus.fi>)

Aineistossa myös toimialaluokkien osalta puuttuu kansallinen 5-numeroinen taso, eli aineistossa on vain 4-hierarkkista tasoa.

3.1 Aineistoon tehtävät rajaukset

Aineistossa on ammattiluokat jaettu neljään eri luokkaan ja näiden luokkien sisälle kuuluu myös ns. erityisryhmiä, kuten ammattikoulujen ja muiden ammattiin tai tutkintoon valmistavien laitosten oppilaat, peruskoulun luokkien 7-10 oppilaat, lukion oppilaat sekä hoito- yms. laitoksiin sisälle otetut henkilöt (ns. hoidokit). Tutkielmassa ei ole mielekästä ottaa erityisryhmiä lainkaan tarkasteluun, joten nämä rajataan pois analyyseissa. Lisäksi ammattiluokka "Sotilaat" rajataan tutkimuksen ulkopuolelle.



Kuvio 3.3. Vahinkojen jakautuminen eri toimialojen välillä

4 Menetelmät

Tässä kappaleessa esitellään tutkielmassa käytettyjen menetelmien teoriaa. Aluksi esitellään Poisson-regressio, Poisson-regressioon liittyvä ylihajonta ja negatiivinen binomijakauma. Nämä ovat suurimmaksi osaksi esitetty, kuten (Hilbe 2011) sen kirjassaan esittää. Tämän jälkeen perehdytään zero-inflated- ja zero-altered-menetelmiin, joissa suurimmaksi osaksi teoria esitetään, kuten (Cameron & Trivedi 1998) sen kirjassaan esittävät.

4.1 Poisson-regressio

Poisson-regressio perustuu Poissonin todennäköisyysjakaumaan ja sitä käytetään yleisesti mallintamaan lukumääräistä dataa. Poissonin todennäköisyysjakauma voidaan esittää muodossa

$$(4.1) \quad f(y; \mu) = \frac{e^{-\mu} \mu^y}{y!},$$

missä $y = 0, 1, \dots$ ja $\mu \geq 0$. Odotusarvo μ_i on nyt selittäjien funktio seuraavasti $\mu_i = \exp(\mathbf{x}'_i \beta)$, missä β on regressiokerroin, ja \mathbf{x} on selittävien muuttujien arvojen muodostama vektori.

Yleisesti Poissonin sekä lukumäärien malleissa μ on odotusarvoparametri. Poissonin jakaumassa keskeisenä oletuksena on odotusarvon ja varianssin yhtäsuuruus $E(y) = \mu = \text{Var}(y)$. Tämä oletus harvoin toteutuu käytännössä, kun tutkitaan oikeaa dataa ja useimmiten näissä tilanteissa varianssi on odotusarvoa suurempi, jolloin kyseessä on ns. ylihajonta. Kun varianssi on pienempi kuin keskiarvo on kyseessä alihajonta. Alihajontaa harvemmin esiintyy käytännössä. (Hilbe 2011, Cameron & Trivedi 1998)

4.2 Ylihajonta

Poissonin malleissa ylihajontaa esiintyy siis, kun varianssi on suurempi kuin odotusarvo. Ylihajontaa aiheutuu positiivisesta korrelaatiosta vasteiden välillä tai ylimääräisestä variaatiosta vasteen todennäköisyyksien tai lukumäärien välillä. Ylihajonta on ongelmallista, koska tuloksissa keskihajonta voi näyttää liian pieneltä ja tällöin ennustava muuttuja voi näyttää tilastollisesti merkitsevältä, mutta todellisuudessa se ei ehkä ole. Tämä taas voi aiheuttaa virheellisen johtopäätelmän. (Hilbe 2011)

Ylihajonta voidaan esimerkiksi havaita testillä, joka perustuu Pearsonin residuaaleihin.

$$(4.2) \quad r_i = \frac{(y_i - \hat{\mu}_i)}{\sqrt{\hat{\mu}_i}}.$$

Mikäli Poisson-mallin oletukset pitävät paikkansa, niin Pearsonin residuaalit pitäisivät tulla standardoidusta normaalijakaumasta. Tästä saadaan testi

$$(4.3) \quad \chi_{n-q}^2 = \sum_{i=1}^n r_i^2,$$

missä q on estimoitujen parametrien määrä. Nollahypoteesin vallitessa testisuure noudattaa nyt χ^2 -jakaumaa vapausastein $n-q$. Mikäli nollahypoteesi hylätään, niin useimmiten testattu malli sisältää ylihajontaa. (Smithson et al. 2013)

Peukalosääntönä ylihajontaa voidaan myös havaita, mikäli otossuure jaettuna vapausasteilla on suurempi kuin 1. Keskipakoisilla malleilla arvon ollessa suurempi kuin 1,25 voi mallin korjaaminen olla perusteltua. Havaintojen lukumäärän ollessa suuri voidaan rajana pitää arvoa 1,05. Ylihajontaa voidaan yrittää korjata erilaisilla tavoilla; malliin voidaan lisätä uusia selittäviä muuttujia, skaalata selittäviä muuttujia uudelleen tai muuttaa käytettyä linkkifunktiota. Useimmiten siirrytään myös käyttämään negatiiviseen binomijakaumaan perustuvaa mallinusta. (Hilbe 2011)

Esimerkki ylihajonnasta

Esitellään seuraavaksi esimerkki Poisson-regressiomallissa esiintyvistä ylihajonnasta. Esimerkki perustuu (Smithson et al. 2013) kirjassaan esittämään esimerkkiin. Esimerkkiaineistona käytetään R-ohjelman paketissa "smdata" löytyvää aineistoa "fixations". Kyseinen data on Owens, Shrestha & Chaparro (2009) tutkimuksesta, jossa tutkittiin tekstin houkuttelevuuden vaikutusta silmien liikkeisiin. Tutkimuksessa nettisivu oli jaettu yhdeksään osioon ja otsikon väri "keskellä, vasen" ja "keskellä, oikea" osiot olivat manipuloituja. Data sisältää kuusi eri muuttujaa ja 48 tilastoyksikköä. Esitellään seuraavaksi R-ohjelmalla toteutettu koodi.

```
#Esimerkki ylihajonnasta
```

```
> library("smdata")
```

```
> data(fixations)
```

```
> malli1 <- glm(countright ~ rt.cond,
                data = fixations, family = poisson)
```

```
> summary(malli1)
```

Call:

```
glm(formula = countright ~ rt.cond,
     family = poisson, data = fixations)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-3.1623	-1.5886	-0.4635	0.9641	4.2427

Coefficients:

Estimate	Std. Error	z value	Pr(> z)
----------	------------	---------	----------


```
(Intercept) 1.37055 0.08909 15.384 <2e-16 ***
rt.condred 0.23889 0.14296 1.671 0.0947 .
```

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

(Dispersion parameter for poisson family taken to be 1)

```
Null deviance: 150.57 on 47 degrees of freedom
Residual deviance: 147.83 on 46 degrees of freedom
AIC: 284.47
```

Number of Fisher Scoring iterations: 5

Esimerkkimallissa selitettävänä muuttujana on muuttuja ”countright” eli tallenteiden lukumäärä keskellä, oikea osio. Selittävänä muuttujana on muuttuja ”rt.cond”, joka on punainen, jos keskellä oikeassa osiossa otsikko oli punainen ja muulloin se saa arvon musta.

Tehdään seuraavaksi aiemmin määritelmässä 4.3 esitetty testi, joka perustuu Pearsonin residuaaleihin.

```
> pearson.residuaalit <- residuals(malli1, type="pearson")
> testi <- sum(pearson.residuaalit^2)
> pchisq(testi, malli1$df.residual, lower.tail = FALSE)
[1] 6.559257e-11
```

Tämän tuloksen seurauksena nollahypoteesi hylätään, jolloin useimmiten testattu malli sisältää ylihajontaa.

4.3 Negatiivinen binomijakauma

Negatiivinen binomijakauman käyttäminen on yleinen ratkaisu tilanteissa, joissa Poissonin regression osalla esiintyy ylihajontaa. Poissonin jakauma on negatiivisen binomijakauman erikoistapaus, jonka vuoksi negatiivisen binomiregression käyttäminen on perusteltua. Nykyään perinteinen negatiivinen binomijakauma johdetaan sekoitetusta Poisson-gamma jakaumasta, jota kutsutaankin NB2 nimellä. NB2 mallin varianssi on muotoa $VAR(\mu) = \mu + \alpha\mu^2$, missä α on ns. ylihajontaparametri.

4.3.1 Sekoitettu Poisson-gammajakauma

Seuraavaksi johdetaan negatiivinen binomijakauma Poissonin ja gammajakauman yhdistelmänä. Poisson-gammajakauman varianssi on muotoa $\mu + \mu^2/\nu$, missä μ on Poissonin varianssi ja μ^2/ν on gamman varianssi. Parametri ν on gamman muoto-parametri. Tiheysfunktioksi saadaan

$$(4.4) \quad f(y; \lambda; u) = \frac{e^{-\lambda_i u_i} (\lambda_i u_i)^{y_i}}{y_i!},$$

missä $i = 1, 2, \dots$. Parametri u noudattaa nyt gammajakaumaa. Tätä voidaan ajatella Poisson mallina, jolla on gammaheterogeenisyys ja jossa on gammankohinan odotusarvona on 1. Tarkastellaan seuraavaksi jakaumaa y ehdolla u . Tämä on Poissonin ehdollisella odotusarvolla ja varianssilla μ , kuten kohdassa 4.4. Y :n ehdollinen odotusarvo gamman heterogeenisyydessä on täten esitetty λu . Tästä saadaan johdettua seuraava

$$(4.5) \quad f(y; u) = \int_0^{\infty} \frac{e^{-(\lambda_i u_i)} (\lambda_i u_i)^{y_i}}{y_i!} g(u_i) \partial u_i.$$

Vasteen Y jakauma määrittelee sen, kuinka määritämme $g(u)$:n. Oletetaan gammajakauman odotusarvoksi 1. Nyt kun integroidaan yhteisjakauma u :n suhteen saadaan johdettua seuraavaa

$$(4.6) \quad \begin{aligned} f(y; u) &= \int_0^{\infty} \frac{e^{-(\lambda_i u_i)} (\lambda_i u_i)^{y_i}}{y_i!} \frac{v^\nu}{\Gamma(\nu)} u_i^{\nu-1} e^{-\nu u_i} du_i \\ &= \frac{\lambda_i^{y_i}}{\Gamma(y_i + 1)} \frac{v^\nu}{\Gamma(\nu)} \int_0^{\infty} e^{-(\lambda_i + \nu) u_i} u_i^{(y_i + \nu) - 1} du_i \\ &= \frac{\lambda_i^{y_i}}{\Gamma(y_i + 1)} \frac{v^\nu}{\Gamma(\nu)} \frac{\Gamma(y_i + \nu)}{(\lambda_i + \nu)^{y_i + \nu}}. \end{aligned}$$

Jatkamalla edellä olevan 4.6 muokkaamista saadaan negatiivisen binomijakauman todennäköisyysfunktio seuraavaanlaiseen muotoon

$$(4.7) \quad \begin{aligned} &= \frac{\lambda_i^{y_i}}{\Gamma(y_i + 1)} \frac{v^\nu}{\Gamma(\nu)} \Gamma(y_i + \nu) \left(\frac{\nu}{\lambda_i + \nu} \right)^\nu \frac{1}{v^\nu} \left(\frac{\lambda_i}{\lambda_i + \nu} \right)^{y_i} \frac{1}{\lambda_i^{y_i}} \\ &= \frac{\Gamma(y_i + \nu)}{\Gamma(y_i + 1) \Gamma(\nu)} \left(\frac{\nu}{\lambda_i + \nu} \right)^\nu \left(1 - \frac{\lambda_i}{\lambda_i + \nu} \right)^{y_i} \\ &= \frac{\Gamma(y_i + \nu)}{\Gamma(y_i + 1) \Gamma(\nu)} \left(\frac{1}{1 + \lambda_i/\nu} \right)^\nu \left(1 - \frac{1}{1 + \lambda_i/\nu} \right)^{y_i}. \end{aligned}$$

Kääntämällä gammafunktion parametrin ν saadaan heterogeenisyys tai ylihaiontaparametri α . Lisäksi merkitsemällä λ ja μ keskenään yhtä suuriksi saadaan tulokseksi negatiivisen binomijakauman tiheysfunktio muotoon

$$(4.8) \quad f(y; \lambda; \alpha) = \frac{\Gamma(y_i + 1/\alpha)}{\Gamma(y_i + 1) \Gamma(1/\alpha)} \left(\frac{1}{1 + \alpha \mu_i} \right)^{\frac{1}{\alpha}} \left(1 - \frac{1}{1 + \alpha \mu_i} \right)^{y_i}.$$

(Hilbe 2011, luku 8)

4.4 Nolla-arvojen yliedustus

Tässä tutkielmassa käytettävässä aineistossa niin sanottujen nolla-arvojen käsittely on hieman poikkeavaa. Käytetyssä esimerkkiaineistossa ei ole täsmällistä nollaluokkaa, vaan luokka sisältää lukumäärät 0, 1, 2 ja 3. Tämä luokka on datassa kuitenkin selvästi yliedustettuna ja sitä on siten perusteltua käsitellä omana nollaluokkana. Näin varsinkin, kun varsinainen kiinnostus on suurissa lukumäärissä.

Monesti nolla-arvojen yliedustus aiheuttaa malleissa ylihajontaa ja niiden käsittelyyn onkin erilaisia tapoja riippuen siitä minkä tyyppisiä nolla-arvoja datassa on. Yleisempiä tapoja niiden käsittelyyn ovat zero-inflated sekä zero-altered menetelmät. Lisäksi on olemassa menetelmä zero-truncated, mikäli selitettävä muuttuja ei voi saada ollenkaan nolla-arvoja, esimerkiksi jos mitataan aikaa kuinka kauan potilas on ollut sairaalassa. Tällöin arvo ei voi olla koskaan nolla. Tässä tutkielmassa ei kuitenkaan ole tällainen tilanne, joten sen teorian käsittelyä ei käydä läpi. Malleja on yhteensä neljä kappaletta; zero-inflated Poisson (ZIP), zero-inflated negative binomial (ZINB), zero-altered Poisson (ZAP) ja zero-altered negative binomial (ZANB). Nämä menetelmät eroavat nolla-arvojen käsittelyn suhteen hieman toisistaan ja käydään niiden erot läpi jäljempänä. Tässä tutkielmassa tarkoituksena on tutkia millaisia eroja ammattiluokkien välillä on tapaturmien vakavuuksien välillä ja yhtenä tutkimuksen kohteena on myös minkälaisia eroja tuloksissa tulee riippuen käytettävästä menetelmästä nolla-arvojen suhteen. (Zuur et al. 2009) Selvennyksen vuoksi jatkossa käytetään edellä mainituista menetelmistä niiden lyhenteitä. Ilman jakaumaa olevat lyhenteet ovat ZI eli zero-inflated ja ZA eli zero-altered. ZA-mallit tunnetaan myös toiselta nimeltä hurdle-mallit. Seuraavaksi tutustutaan tarkemmin ZA ja ZI menetelmiin.

4.4.1 Hurdle-mallit

Ensimmäiset ZA eli toiselta nimeltään hurdle mallit ovat tehneet Mullahy (1986) ja myöhemmin enemmän tietoisuuteen ovat tuoneet Cameron ja Trivedi (1986, 1998).

Hurdle-mallit voidaan tulkita kaksiosaiseksi malliksi. Ensimmäisessä osassa on binäärinen lopputulos ja se kuvaa kynnsarvon ylittymisen todennäköisyyttä binäärisenä mallina eli positiiviset arvot verrattuna nolla-arvoihin. Toinen osa on katkaistu lukumäärien malli, jossa vain positiivisia arvoja (Cameron & Trivedi 1998)

Hurdle-malleissa nolla-arvoja oletetaan tulevan kahdesta eri lähteestä. Esimerkkinä voidaan ottaa, kun kyselytutkimuksessa esitetään kysymys "Kuinka monta kertaa olet käynyt golfaamassa edellisen kuukauden aikana?". Nyt nolla-vastauksia tulee niiltä, jotka eivät koskaan golfaa ja niiltä jotka golfaavat, mutta eivät ole golfanneet edellisen kuukauden aikana. Tämän vuoksi otos on niin sanotusti sekoitettu ja täten voisi olla väärin olettaa, että nolla-arvot ja positiiviset arvot tulevat samasta lähteestä. Seuraava hurdle- ja ZI-mallinnusosio seuraa Cameron ja Trivedin (1998) esitystä. Camenron ja Trivedi esittävät mallinnuksen kaksiosaisena prosessina, missä nolla-arvoja tuottava jakauma on f_1 ja positiivisia arvoja tuottava jakauma on f_2

$$(4.9) \quad \begin{aligned} Pr[y = 0] &= f_1(0) \\ Pr[y = j] &= \frac{1 - f_1(0)}{1 - f_2(0)} f_2(y), \quad j > 0, \end{aligned}$$

Tämä esittää kaksioisaista prosessia, jossa nolla-arvojen ja positiivisten arvojen tuottaminen ei oleteta olevan samoja. Usein nolla-arvojen mallintaminen tapahtuu binomijakaumasta ja positiivisten arvojen mallintaminen tapahtuu Poissoninjakaumasta tai negatiivisesta binomijakaumasta.

Nyt hurdle-mallin odotusarvo voidaan määrittellä kynnsarvon ylittymisen todennäköisyydellä seuraavasti

$$(4.10) \quad E[y|\mathbf{x}] = Pr[y > 0|\mathbf{x}]E_{y>0}[y|y > 0, \mathbf{x}],$$

Varianssi voidaan vastaavasti esittää seuraavalla tavalla

$$(4.11) \quad V[y|\mathbf{x}] = Pr[y > 0|\mathbf{x}]V_{y>0}[y|y > 0, \mathbf{x}] + Pr[y = 0|\mathbf{x}]E_{y>0}[y|y > 0, \mathbf{x}]$$

Käsitellään seuraavaksi hurdlen NB2 mallia. Oletetaan $\mu_{1i} = \exp(\mathbf{x}'_i\beta_1)$ NB2 on odotusarvoparametri nolla-arvoille ja olkoon $\mu_{2i} = \mu_2(\mathbf{x}'_i\beta_2)$ positiivisille arvoille siten että $J = \{1, 2, \dots\}$. Lisäksi määritellään indikaattorifunktio $1[y_i \in J] = 1$, jos $y_i \in J$ ja $1[y_i \in J] = 0$, jos $y_i = 0$. Negatiivisen binomijakauman varianssin neliöfunktion todennäköisyydet voidaan saada:

$$(4.12) \quad Pr[y_i = 0|\mathbf{x}_i] = (1 + \alpha_1\mu_{1i})^{-\frac{1}{\alpha_1}}$$

$$(4.13) \quad 1 - Pr[y_i = 0|\mathbf{x}_i] = \sum_{y_i \in J} h(y_i|\mathbf{x}_i) = 1 - (1 + \alpha_1\mu_{1i})^{-\frac{1}{\alpha_1}}$$

$$(4.14) \quad Pr[y_i|\mathbf{x}_i, y_i > 0] = \frac{\Gamma(y_i + \alpha_2^{-1})}{\Gamma(\alpha_2^{-1})\Gamma(y_i + 1)} \left(\frac{1}{(1 + \alpha_2\mu_{2i})^{-1/\alpha_2} - 1} \right)^{-\alpha_2^{-1}} \times \left(\frac{\mu_{2i}}{\mu_{2i} + \alpha_2^{-1}} \right)^{y_i}$$

missä α on hajontaparametri kuten aikaisemmin esitetty. Määritelmä 4.12 antaa todennäköisyydet nolla-arvoille, kun taas 4.13 todennäköisyyden kynnsarvon ylittämiseksi eli positiivisille arvoille. Määritelmässä 4.14 on esitetty katkaistu lukumäärien NB2-jakauma. Havaintojen logaritmoitu uskottavuusfunktio jakautuu kahteen osaan

$$(4.15) \quad \begin{aligned} \mathcal{L}_1(\beta_1, \alpha_1) &= \sum_{i=1} [(1 - 1[y_i \in J]) \ln Pr[y_i = 0|\mathbf{x}_i]] \\ &+ \sum_{i=1} 1[y_i \in J] \ln(1 - Pr[y_i = 0|\mathbf{x}_i]) \\ \mathcal{L}_2(\beta_2, \alpha_2) &= \sum_{i=1} 1[y_i \in J] \ln[Pr y_i|\mathbf{x}_i, y_i > 0] \\ \mathcal{L}(\beta_1, \beta_2, \alpha_1, \alpha_2) &= \mathcal{L}_1(\beta_1, \alpha_1) + \mathcal{L}_2(\beta_2, \alpha_2) \end{aligned}$$

Nyt $\mathcal{L}_1(\beta_1, \alpha_1)$ on logaritmoitu uskottavuus binääriselle prosessille, jossa havainnot jakautuva nolla-arvoihin ja positiivisiin arvoihin. $\mathcal{L}_2(\beta_2, \alpha_2)$ on negatiivisen binomijakauman katkaistu uskottavuusfunktio positiivisille arvoille. Molempien mekanismien oletetaan olevan itsenäisiä ja täten riippumattomia toisistaan ja sen vuoksi hurdle malli voidaankin estimoida käyttämällä laskentaohjelmaa, missä ei ole sisäänrakennettua valmista hurdle-mallia. (Cameron & Trivedi 1998)

4.4.2 Zero-inflated-mallit

Verrattuna hurdle-malleihin, ZI-mallit sallivat joidenkin nolla-arvojen kuuluvan suoraan lukumäärien regressomalliin ZI-mallit tarjoavat toisenlaisen tavat käsitellä nolla-arvoja. Oletetaan seuraavaksi selkeyden vuoksi Y_i noudattaa Poissonin jakaumaa odotusarvolla μ_i . Nyt voidaan tarkastella seuraavaksi määritelmiä:

$$(4.16) \quad \begin{aligned} Pr[y_i = 0] &= \varphi_i + (1 - \varphi_i)e^{-\mu_i} \\ Pr[y_i = r] &= (1 - \varphi_i)\frac{e^{-\mu_i}\mu_i^r}{r!}, \quad r = 1, 2, \dots \end{aligned}$$

Määritelmässä 4.16 ylempi osa esittää todennäköisyyden $Y_i = 0$. Todennäköisyys nolla-arvon mittaamiseen on yhtäsuuri kuin väärän nolla-arvon todennäköisyys lisätynä todennäköisyydellä ettei se ole väärä nolla-arvo ja kerrottuna tämä todennäköisyydellä, että mittaamme todellisen nolla-arvon. Alemmassa osassa on esitetty todennäköisyys, että Y_i ei ole nolla-arvo. (Zuur et al. 2009)

Osa φ_i arvoista voidaan edelleen parametrisoida muunnoksella $\mathbf{z}_i\gamma$. Tavoitteena on estimoida (β, γ) . Oletetaan identifioituvuus, koska

$$(4.17) \quad V[y_i] = (1 - \varphi_i)(\mu_i + \varphi_i\mu_i^2) > \mu_i(1 - \varphi_i) = E[y_i].$$

Ylimääräiset nollat viittavat ylihajontaan.

Aikaisemmin Lambert (1992) on esitellyt ZIP-mallin, jossa $\mu_i = \mu(\mathbf{x}_i, \beta)$ sekä todennäköisyyden φ_i on parametroidu logistisena funktiona kovarianttien havaintovektorina \mathbf{z}_i ja varmistamalla siten, että φ_i ei ole negatiivinen. Joten

$$(4.18) \quad \begin{aligned} y_i = 0 & \quad \text{todennäköisyydellä } \varphi_i \\ y_i \sim P[\mu_i] & \quad \text{todennäköisyydellä } (1-\varphi_i) \\ \varphi_i &= \frac{\exp(\mathbf{z}_i'\gamma)}{1 + \exp(\mathbf{z}_i'\gamma)} \end{aligned}$$

Oletetaan seuraavaksi $1(y_i = 0)$ tarkoittavan indikaattorimuuttujaa, joka saa arvon 1, jos $y_i = 0$ ja muutoin arvon 0. Nyt yhdistetty uskottavuusfunktio, kun vakiot on jätetty pois on

$$\begin{aligned}
\mathcal{L}(\beta, \gamma) &= \sum_{i=1}^n 1(y_i = 0) \ln(\exp(\mathbf{z}_i' \gamma)) \\
&+ \exp(-\exp(\mathbf{x}_i' \beta)) + \sum_{i=1}^n (1 - 1(y_i = 0)) (y_i \mathbf{x}_i' \beta - \exp(\mathbf{x}_i' \beta)) \\
(4.19) \quad &- \sum_{i=1}^n \ln(1 + \exp(\mathbf{z}_i' \gamma))
\end{aligned}$$

Lambert suosittelee käyttämään EM-algoritmia (Expectation and Maximization algorithm) uskottavuusfunktion maksimointiin. Kuten siis sensuroidussa regressiotapauksessa, tämäkin käyttää odotettavissa olevaa uskottavuutta. Uskottavuusfunktio saavutetaan, mikäli indikaattorimuuttuja korvataan vastaavalla odotetulla arvolla. Odotettu uskottavuusfunktio $\mathcal{E} \mathcal{L}(\beta, \gamma)$ voidaan sitten maksimoida tuntemattomalle parametrille β . Kun huomioidaan komponenttien μ_i ja λ_i riippumattomuus voidaan näiden uskottavuus maksimoida EM-algoritilla. Mahdollisuus on myös käyttää Newton-Raphson algoritmia, joka myöskin suppenee suhteellisen nopeasti. Tämä malli voidaan laajentaa myös käyttämään negatiivisen binomijakauman mallia. Nyt logistinen määritelmä voidaan korvata $\varphi_i = F(\mathbf{z}_i' \gamma)$, missä $F(\cdot)$ on mikä tahansa kertymäfunktio ja \mathbf{z}_i kovarianttien havaintovektori. Tämä lähestymistapa toimii paremmin, mikäli muuttujien \mathbf{x} ja \mathbf{z}_i korrelaatio on pieni. (Cameron & Trivedi 1998)

4.5 Mallin yhteensopivuuden testaaminen

On erittäin tärkeää tehdä käytetylle mallille yhteensopivuustestejä, koska muuten voidaan tehdä vääriä johtopäätöksiä. Voi olla, että luotu malli antaa p-arvoltaan tilastollisesti merkitseviä tuloksia, vaikka todellisuudessa luotu malli ei sovi mallintamaan dataa. Malli, jolle ei ole tehty mitään sopivuustestejä, on tilastollisesti hyödytön. (Hilbe 2011) Seuraavaksi esitetään muutama erilainen testi mallin yhteensopivuuden testaamiseen, kuten (Hilbe 2011) niitä kuvailee.

4.5.1 R^2 ja pseudo- R^2 yhteensopivuusasteet

R^2 -selitysaste on hyvin tunnettu ja se liittyy pääosin pienimmän neliösumman menetelmän käyttöön. Tämä ei kuitenkaan sovellu kaikkien menetelmien sopivuuden testaamiseen, kuten Poisson, negatiivinen binomijakauma tai logistinen regressio. Lukumäärien mallintamisen tilanteissa yleensä käytetäänkin menetelmää, jota kutsutaan pseudo- R^2 yhteensopivuusasteeksi. Tämän yleisin muotoilu on:

$$(4.20) \quad R_p^2 = 1 - \mathcal{L}_F / \mathcal{L}_I$$

Missä \mathcal{L}_F on logaritminen uskottavuus koko mallille ja \mathcal{L}_I logaritminen uskottavuus pelkästään vakiomallille. Pseudo- R^2 -arvoa ei voida kuitenkaan tulkita samalla tavalla

kuin perinteistä R^2 -arvoja. Yleisesti on tiedossa, että matalat arvot voivat osoittaa mallin sopimattomuutta ja korkeat arvot toimivat päinvastoin. Kuitenkaan ei ole asetettu mitään tarkkoja kriteereitä selvyyden lisäämiseksi. (Hilbe 2011)

4.5.2 Devianssin yhteensopivuusaste

Devianssia käytetään yleistettyjen lineaaristen mallien ohjelmistoissa, missä se on normaalisti esitetty mallien lopputuloksissa. Se esitetään:

$$(4.21) \quad D = 2 - \sum_{i=1}^n \{ \mathcal{L}(y_i; y_i) - \mathcal{L}(\mu_i; y_i) \}$$

Missä $\mathcal{L}(y_i; y_i)$ indikoi logaritmista uskottavuusfunktiota jokaiselle arvolle μ_i annettulla arvolla y_i . $\mathcal{L}(\mu_i; y_i)$ on saadun mallin logaritmoitu uskottavuusfunktio.

Devianssin testiä käytetään vertaamaan eri malleja. Testissä käytetään χ^2 -testiä, jossa annetaan devianssin arvo sekä vapausasteet. Mikäli χ^2 -testin tuloksena p-arvo on pienempi kuin 0,05 voidaan ajatella mallin soveltuvan hyvin. Ideana on, että malli, joka antaa pienemmän devianssin on parempi. (Hilbe 2011)

4.5.3 Uskottavuusosamäärätesti

Uskottavuusosamäärätesti on usein käytetty vertailevana sopivuustestinä, mikä soveltuu sisäkkäisiin malleihin. Uskottavuusosamäärätesti voidaan esittää seuraavasti (Hilbe 2011)

$$(4.22) \quad LR = -2\{ \mathcal{L}_{y_{reduced}} - \mathcal{L}_{full} \}$$

Määritelmässä 4.22 $\mathcal{L}_{y_{reduced}}$ on pelkän vakion sisältävä malli ja \mathcal{L}_{full} on selittävät muuttujat sisältävä malli. Uskottavuusosamäärätestillä onkin hyvä testata liittääkö selittäviä muuttujia malliin eli parantaako tämä tehtyä mallian, kun sitä verrataan pelkän vakion sisältävään malliin.

5 Tulokset

Tarkoituksena tutkielmassa on mallintaa, millaisia eroja riskeissä ammattiluokkien välillä on sattuneiden tapaturmien vakavuuksissa. Tässä kappaleessa ensimmäiseksi tarkastellaan eroja edellä esitetyissä menetelmissä ja tarkempi tarkastelu tapahtuu nolla-arvoja huomioon ottavien menetelmien ZIP-, ZAP-, ZINB- ja ZANB-mallien välillä. Tämän jälkeen tulkitaan paremmin minkälaisia eroja saadaan eri ammattiluokkien välillä ja lopuksi tehdään vielä tarkastelua, onko toimialojen välillä minkälaisia eroja.

Esitetyissä malleissa referenssiammattiluokkana on ollut A1LK7 (Rakennus-, korjaus- ja valmistustyöntekijät) sekä A2LK71 (Rakennustyöntekijät ym. (pl. sähköasentajat)). Nämä ammattiluokkia antavat malleissa perustason eli sen mihin kaikkia muita ammattiluokkia verrataan.

Kaikki mallit ja tulosteet on luotu käyttämällä RStudio-ohjelmaa, joka perustuu R-ohjelmaan, mutta on mielestäni visuaalisesti ja toiminnallisesti selkeämpi käyttää. RStudiolla tehdyt mallit ja tulosteet sekä tarkempi tieto ammattiluokkien ja toimialaluokkien nimistä löytyvät tämän tutkielman liitteistä. Tässä kappaleessa esitetään tärkeimmät yhteenvetotaulukot ja tulokset sekä esitetään pohdintaa näiden tietojen perusteella.

5.1 Nolla-arvomenetelmien tulosten erot

Tarkastellaan ensimmäiseksi minkälaisia eroja tuloksissa saadaan, kun tutkitaan nolla-arvot huomioon ottavien menetelmien tuloksia. Tämän jälkeen analysoidaan, minkälaisia eroja ammattiluokkien välillä on.

Taulukkoihin 5.1 ja 5.2 on koostettu yhteenveto eri nolla-arvot huomioon ottavien menetelmien tulokset. Näistä voidaan havaita menetelmien ZIP ja ZAP estimaattien arvot ovat lähes samat ja ZINB ja ZANB menetelmien tulokset ovat keskenään samanlaiset. Ainoastaan viimeisten desimaalien välillä tulee hieman eroja. Vähän suurempi ero estimaateissa on kun verrataan ZIP ja ZAP estimaatteja ZINB ja ZANB estimaatteihin. Suurin ero kuitenkin menetelmissä tulee, kun tarkastellaan tilastollisesti merkitseviä ammattiluokkia eli p-arvo on pienempi kuin 0,05. Näistä tuloksista voidaan havaita, että ZINB ja ZANB menetelmät antavat huomattavasti vähemmän tilastollisesti merkitseviä ammattiluokkia kuin ZIP ja ZAP menetelmät. Eri mallien hyvyyden vertailemiseen voidaan käyttää Akaiken informaatiokriteeriä (AIC). Mitä pienempi AIC arvo on, sitä parempi malli on (Zuur et al. 2009). Menetelmien AIC arvot löytyvät taulukosta 5.3. Voidaan havaita, että menetelmillä ZINB ja ZANB arvot ovat pienemmät ja tämän kertoo näiden mallien selittävän paremmin ammattiluokkien eroja.

Taulukko 5.1. Nolla-arvomenetelmien tuloksia eri ammattiluokille.

Ammattiluokka	ZIP	ZAP	ZINB	ZANB
	Estimaatti	Estimaatti	Estimaatti	Estimaatti
(Intercept)	3,452003 ***	3,452003 ***	3,349096 ***	3,349106 ***
A2LK11	0,189734 ***	0,189718 ***	0,201893 *	0,201707 *
A2LK12	-0,46968 ***	-0,469674 ***	-0,507617 **	-0,507738 **
A2LK13	0,188499 ***	0,188474 ***	0,200537	0,200341
A2LK14	0,217966 ***	0,217942 ***	0,231298	0,231625
A2LK21	-0,002971	-0,00297	-0,00322	-0,003105
A2LK22	-0,014172	-0,014165	-0,015051	-0,014957
A2LK23	-0,026339 ***	-0,026338 ***	-0,028094	-0,028105
A2LK24	-0,016561 .	-0,016558 .	-0,017682	-0,017756
A2LK25	-0,086309 ***	-0,086295 ***	-0,092346	-0,09215
A2LK26	-0,050907 ***	-0,050905 ***	-0,054416	-0,054445
A2LK31	-0,005928	-0,00593	-0,006355	-0,006367
A2LK32	-0,133996 ***	-0,133995 ***	-0,143486 ***	-0,143545 ***
A2LK33	0,010692 .	0,010694 .	0,011485	0,011462
A2LK34	-0,113233 ***	-0,113226 ***	-0,121237 **	-0,121286 **
A2LK35	-0,115836 ***	-0,115836 ***	-0,1241	-0,124114
A2LK41	0,037306 ***	0,037307 ***	0,03986	0,03981
A2LK42	-0,104581 ***	-0,10458 ***	-0,11201 *	-0,11191 *
A2LK43	-0,061764 ***	-0,061751 ***	-0,066056	-0,066055
A2LK44	-0,18909 ***	-0,189088 ***	-0,202823 ***	-0,202816 ***
A2LK51	-0,166161 ***	-0,16616 ***	-0,178137 ***	-0,178183 ***
A2LK52	-0,274648 ***	-0,274648 ***	-0,295307 ***	-0,295269 ***
A2LK53	-0,195966 ***	-0,195964 ***	-0,210219 ***	-0,210225 ***
A2LK54	-0,032061 ***	-0,03206 ***	-0,034291	-0,03428
A2LK61	-0,021699 ***	-0,021698 ***	-0,023143	-0,023142
A2LK62	-0,052586 ***	-0,052579 ***	-0,056188	-0,056176
A2LK72	-0,128123 ***	-0,128122 ***	-0,137079 ***	-0,137225 ***
A2LK73	-0,121875 ***	-0,121856 ***	-0,130383	-0,130453
A2LK74	-0,037343 ***	-0,037342 ***	-0,039922	-0,039881
A2LK75	-0,197233 ***	-0,197233 ***	-0,211609 ***	-0,211643 ***
A2LK81	-0,129408 ***	-0,129409 ***	-0,138672 ***	-0,138578 ***
A2LK82	-0,161165 ***	-0,16117 ***	-0,172759 ***	-0,172685 ***
A2LK83	0,121729 ***	0,121731 ***	0,129641 ***	0,12966 ***
A2LK91	-0,036975 ***	-0,036975 ***	-0,03951	-0,039524
A2LK92	0,281326 ***	0,28132 ***	0,298632	0,298772
A2LK93	-0,185635 ***	-0,185635 ***	-0,199096 ***	-0,199177 ***
A2LK94	-0,390409 ***	-0,390407 ***	-0,421129 ***	-0,421044 ***
A2LK95	-0,37421 ***	-0,374206 ***	-0,403422 .	-0,403557 .
A2LK96	-0,253816 ***	-0,253815 ***	-0,272802 ***	-0,272786 ***

¹Merkitsevyyden selitteet: p-arvon arvot *** - 0, ** < 0,001, * < 0,05, . < 0,1

Taulukko 5.2. Zero-inflation- ja Zero-hurdle-mallien tuloksia ammattiluokille.

Ammattiluokka	ZIP	ZAP	ZINB	ZANB
	Estimaatti	Estimaatti	Estimaatti	Estimaatti
(Intercept)	0,330513 ***	-0,330513 ***	0,146529 ***	-0,330513 ***
A2LK11	0,726835 ***	-0,726856 ***	0,786346 ***	-0,726856 ***
A2LK12	0,912697 ***	-0,91268 ***	0,910985 ***	-0,91268 ***
A2LK13	0,691217 ***	-0,691138 ***	0,749391 ***	-0,691138 ***
A2LK14	0,540412 **	-0,540315 **	0,595182 **	-0,540315 **
A2LK21	0,835901 ***	-0,835896 ***	0,882424 ***	-0,835896 ***
A2LK22	0,899409 ***	-0,899389 ***	0,94693 ***	-0,899389 ***
A2LK23	0,799381 ***	-0,79937 ***	0,842363 ***	-0,79937 ***
A2LK24	1,043288 ***	-1,043288 ***	1,095068 ***	-1,043288 ***
A2LK25	1,103923 ***	-1,103831 ***	1,151063 ***	-1,103831 ***
A2LK26	0,446643 ***	-0,446637 ***	0,471185 ***	-0,446637 ***
A2LK31	0,581327 ***	-0,581332 ***	0,617297 ***	-0,581332 ***
A2LK32	0,239342 ***	-0,239338 ***	0,241203 ***	-0,239338 ***
A2LK33	0,603558 ***	-0,60357 ***	0,642194 ***	-0,60357 ***
A2LK34	0,506889 ***	-0,506895 ***	0,527852 ***	-0,506895 ***
A2LK35	0,908338 ***	-0,908398 ***	0,946709 ***	-0,908398 ***
A2LK41	0,850704 ***	-0,850703 ***	0,901348 ***	-0,850703 ***
A2LK42	0,461068 ***	-0,461074 ***	0,480495 ***	-0,461074 ***
A2LK43	0,490901 ***	-0,490914 ***	0,516619 ***	-0,490914 ***
A2LK44	-0,050913	0,050915	-0,081812	0,050915
A2LK51	0,145117 ***	-0,145116 ***	0,135547 ***	-0,145116 ***
A2LK52	0,429807 ***	-0,429804 ***	0,426818 ***	-0,429804 ***
A2LK53	0,184593 ***	-0,184596 ***	0,174222 ***	-0,184596 ***
A2LK54	0,278647 ***	-0,278646 ***	0,295386 ***	-0,278646 ***
A2LK61	-0,16781 ***	0,167811 ***	-0,186392 ***	0,167811 ***
A2LK62	-0,051821	0,0518	-0,063315	0,0518
A2LK72	0,304667 ***	-0,304669 ***	0,311822 ***	-0,304669 ***
A2LK73	0,231281 **	-0,231298 **	0,233971 **	-0,231298 **
A2LK74	0,387606 ***	-0,387603 ***	0,410353 ***	-0,387603 ***
A2LK75	-0,178924 ***	0,178935 ***	-0,225396 ***	0,178935 ***
A2LK81	-0,004557	0,004557	-0,022211	0,004557
A2LK82	0,267721 ***	-0,267724 ***	0,268182 ***	-0,267724 ***
A2LK83	-0,222616 ***	0,222621 ***	-0,227331 ***	0,222621 ***
A2LK91	-0,071034 **	0,071038 **	-0,082166 **	0,071038 **
A2LK92	-0,037067	0,037165	-0,007581	0,037165
A2LK93	-0,004322	0,004323	-0,030001	0,004323
A2LK94	0,235198 ***	-0,235202 ***	0,201005 ***	-0,235202 ***
A2LK95	-0,097068	0,096898	-0,163668	0,096898
A2LK96	-0,180829 ***	0,180842 ***	-0,236866 ***	0,180842 ***

¹Merkitsevyyden selitteet: p-arvon arvot *** - 0, ** < 0,001, * < 0,05, . < 0,1

5.2 Ammattiluokkien välisiä eroja

Edellä havaittiin, että menetelmillä ZIP ja ZAP verrattuna menetelmiin ZINB ja ZANB saatiin tuloksissa eroja ja otetaankin nyt takemmin tarkasteluun ZINB menetelmällä saadut tulokset positiivisille arvoille, jotka löytyvät taulukosta 5.4. Taulukosta esitetään kyseisen mallin tuloksista Estimaattien kertoimet, keskihajonnat, z-arvot sekä p-arvo. Voidaan havaita, että tuloksissa on yhteensä 19 eri ammattiluokkaa, jotka ovat tilastollisesti merkitseviä, kun asetetaan yleisesti käytettynä p-arvon rajaksi $<0,05$. Nämä ammattiluokat ovat eroavat referenssinä käytetystä ammattiluokasta A2LK71 tilastollisesti merkittävästi.

Mikäli estimaatin arvo on negatiivinen tarkoittaa tämä sitä, että kyseisellä ammattiluokalla on pienempi todennäköisyys vakavampaan tapaturmaa kuin referenssiluokalla. Mikäli taas estimaatin arvo on positiivinen, niin kyseisellä ammattiluokalla on suurempi todennäköisyys vakavampaan tapaturmaan kuin referenssiluokalla. Ainoastaan viidellä ammattiluokalla on positiivinen estimaatin arvo, joka tarkoittaa, että on todennäköisempää kuulua kyseiseen ammattiluokkaan, kuin referenssinä olevaan ammattiluokkaan, kun tapaturman vakavuusluokka kasvaa.

Hieman yllättävästi voidaan todeta, että esimerkiksi johtajat, ylimmät virkamiehet ja järjestöjen johtajat ammattiluokalla (A2LK11) oleva estimaatti on positiivinen (0,2) kun referenssiluokkana on rakennustyöntekijät ym. (pl. sähköasentajat). Tämän tuloksen voidaan selittää sillä, ettei mallintamisessa huomioida mitenkään tehdyn työn määrää tai vahingon sattumisen todennäköisyyttä. Vahingon sattumisen todennäköisyydellä tarkoitetaan sitä, kuinka todennäköistä on tapaturman sattuminen työskennellessä johtotyössä verrattuna rakennustyöntekijän työmaalla. Mikäli vahinko sattuu johtajalle, ja se ei kuulu nollaluokkaan, on mallin mukaan todennäköisempää, että vahingon vakavuusluokka on suurempi kuin rakennustyöntekijällä. Taulukossa 5.7 on esitetty molempien ammattiluokkien tapaturmien lukumäärät vakavuusluokittain. Tästä voidaankin havaita, että molemmissa ammattiluokissa suurin osa tapaturmista luokitellaan vakavuusluokkaan 0 ja johtajilla se on selkeästi suurempi osa kuin rakennustyöntekijöillä. Pitää huomioida, että nyt kun on tarkasteltu positiivisia tuloksia, niin voidaan taulukosta 5.7 havaita, että positiivisilla arvoilla rakennustyöntekijöiden ammattiluokassa suurin osa tapaturmista painottuu pienempiin vakavuusluokkiin verrattuna toimistotyöntekijöihin, jossa luokkien prosenttiosuudet ovat selkeästi tasaisemmin jakautuneet. Tämän vuoksi mallin tulokset ovat edellä olevan kuvatut.

Todellisuudessa kuitenkin johtajat, ylimmät virkamiehet ja järjestöjen johtajat (A2LK11) ammattiluokassa on pienempi riski sattuä työtapaturma kuin rakennustyöntekijöillä. TVK joutuu työtapaturma- ja ammattitautilain mukaan ylläpitämään riskiluokitusta. Kyseisessä riskiluokituksessa ei ole annettuna riskiluokkaa ammat-

Taulukko 5.3. Akaiken informaatiokriteerin arvot

	ZIP	ZAP	ZINB	ZANB
AIC	4 404 656	4 404 656	1 039 086	1 039 086

tiluokalle 71, vaan riskiluokitukset ovat jaettu tarkemmalla tasolla oleviin ammattiluokkiin. Kaikilla ammattikoodinumeroilla 71 alkavat ammattiluokkien riskiluokka on suuri ja ja kaikilla ammattikoodinumeroilla 11 alkavat ammattiluokat kuuluvat riskiluokkaan pieni. Toinen ammattiluokka millä on positiivinen estimaatinarvo ja on tilastollisesti merkitsevä, on ammattiluokalla (A2LK83) eli kuljetustyöntekijät. Yhteensä 15:sta tilastollisesti merkitsevällä ammattiluokalla estimaatin arvo on negatiivinen.

Katsotaan seuraavaksi minkälaisia tuloksia malli antaa nolla-arvoille, jonka tulokset löytyvät talukosta 5.5. Tästä voidaan havaita lähes kaikkien ammattiluokkien arvojen olevan tilastollisesti merkitseviä. Ainoastaan ammattiluokkien 62, 81, 92, 93 ja 95 arvot eivät ole tilastollisesti merkitseviä. Tässäkin positiiviset estimaatit tarkoittavat sitä, että kyseisen ammattiluokan osalta on todennäköisempää kuulua tähän luokkaan kuin referenssinä toimivan ammattiluokan. Negatiiviset estimaatin arvot taas päinvastoin. Voidaankin nyt havaita, että tälläkin kertaa johtajat, ylimmät virkamiehet ja järjestöjen johtajat ammattiluokalla (A2LK11) arvo on positiivinen, eli kun tapaturma sattuu tälle ammattiluokkaan kuuluvalla henkilölle, kuuluu se todennäköisemmin luokkaan 0 verrattuna jos tapaturma sattuu rakennustyöntekijälle. Myös taulukon 5.7 mukaan voidaan havaita, että suurin osa johtajat, ylimmät virkamiehet ja järjestöjen johtajat (A2LK11) ammattiluokan tapaturmista kuuluu nolla-luokkaan, kun verrataan rakennustyöntekijöihin.

Mikäli halutaan saada todelliset ammattiluokittaiset erot esiin, pitää estimaatin arvo korottaa $\exp(\text{estimaatin arvo})$. Taulukossa 5.6 on esitetty tällätavalla kaikki estimaatit niin positiivisille arvoille kuin nolla-arvoille. Pitää huomioida, että taulukossa olevista arvoista ei voi suoraan vetää johtopäätöksiä vaan pitää myös huomioida, onko kyseinen ammattiluokka lisäksi tilastollisesti merkitsevä. Aikaisemmin jo todettiin positiivisia arvoja tarkasteltaessa, että hieman yllättävästi ammattiluokalla 11 (johtajat, ylimmät virkamiehet ja järjestöjen johtajat) riski vakavaan vahinkoon on suurempi kuin referenssinä käytetyllä ammattiluokalla. Toinen ammattiluokka, jolla on suurempi riski vakavaan vahinkoon on ammattiluokka 83 (kuljetustyöntekijät). Kaikilla muilla tilastollisesti merkitsevillä ammattiluokilla riski vakavaan vahinkoon on pienempi kuin referenssinä käytetyllä ammattiluokalla. Suurimmalla osalla nolla-arvoista ammattiluokat olivat tilastollisesti merkitseviä. Yllättävimpänä tuloksena voisi pitää ammattiluokan 44 (muut toimisto- ja asiakaspalvelutyöntekijät), koska tämän osalta kuuluminen nollaluokkaan on pienempi kuin referenssiluokalla.

5.3 Toimialaluokkien välisiä eroja

Seuraavaksi on tarkoitus tarkastella minkälaisia tuloksia saadaan toimialakohtaisesti eri ammattiluokkien osalle. Tarkastelu tapahtuu siten, että koko aineisto on jaettuna erillisiin aineistoihin toimialan mukaan. Tämä ei mahdollista eri toimialojen vertailla tilastollisesti, mutta toimialojen välillä ammattiluokat painottuvat monesti tiettyihin ammattiluokkiin. Toimialat ovat siis niin erilaisia, että niitä on mielekkäämpää käsitellä erikseen.

Otetaan ensimmäisenä tarkasteluun, miten koko aineistossa ZINB-malli antaa tu-

Taulukko 5.4. ZINB menetelmällä saadut tulokset positiivisille arvoille.

	Estimaatti	Keskihajonta	z-arvo	p-arvo
(Intercept)	3,349096	0,012474	268,481	< 2e-16 ***
A2LK11	0,201893	0,100563	2,008	0,04468 *
A2LK12	-0,507617	0,191616	-2,649	0,00807 **
A2LK13	0,200537	0,214004	0,937	0,34872
A2LK14	0,231298	0,213967	1,081	0,2797
A2LK21	-0,00322	0,052466	-0,061	0,95107
A2LK22	-0,015051	0,071117	-0,212	0,83239
A2LK23	-0,028094	0,033828	-0,83	0,40627
A2LK24	-0,017682	0,062143	-0,285	0,77599
A2LK25	-0,092346	0,093088	-0,992	0,32118
A2LK26	-0,054416	0,037416	-1,454	0,14584
A2LK31	-0,006355	0,042538	-0,149	0,88125
A2LK32	-0,143486	0,021836	-6,571	5,00E-11 ***
A2LK33	0,011485	0,043455	0,264	0,79155
A2LK34	-0,121237	0,038135	-3,179	0,00148 **
A2LK35	-0,1241	0,111003	-1,118	0,26357
A2LK41	0,03986	0,025437	1,567	0,11711
A2LK42	-0,11201	0,044506	-2,517	0,01184 *
A2LK43	-0,066056	0,068863	-0,959	0,33744
A2LK44	-0,202823	0,03241	-6,258	3,90E-10 ***
A2LK51	-0,178137	0,022148	-8,043	8,77E-16 ***
A2LK52	-0,295307	0,023394	-12,623	< 2e-16 ***
A2LK53	-0,210219	0,019033	-11,045	< 2e-16 ***
A2LK54	-0,034291	0,03564	-0,962	0,33597
A2LK61	-0,023143	0,031816	-0,727	0,46698
A2LK62	-0,056188	0,077593	-0,724	0,46899
A2LK72	-0,137079	0,01946	-7,044	1,86E-12 ***
A2LK73	-0,130383	0,080454	-1,621	0,10511
A2LK74	-0,039922	0,031641	-1,262	0,20705
A2LK75	-0,211609	0,040387	-5,24	1,61E-07 ***
A2LK81	-0,138672	0,021818	-6,356	2,07E-10 ***
A2LK82	-0,172759	0,03834	-4,506	6,61E-06 ***
A2LK83	0,129641	0,019821	6,541	6,13E-11 ***
A2LK91	-0,03951	0,025824	-1,53	0,12603
A2LK92	0,298632	0,19366	1,542	0,12306
A2LK93	-0,199096	0,022931	-8,682	< 2e-16 ***
A2LK94	-0,421129	0,029656	-14,2	< 2e-16 ***
A2LK95	-0,403422	0,208374	-1,936	0,05286 .
A2LK96	-0,272802	0,047011	-5,803	6,52E-09 ***

Taulukko 5.5. ZINB menetelmällä saaduista estimaateista otettu eksponentti, jossa näkee todellisen eron ammattiluokkien välillä

	Estimaatti	Keskihajonta	z-arvo	p-arvo
(Intercept)	0,146529	0,014081	10,406	< 2e-16 ***
A2LK11	0,786346	0,094792	8,296	< 2e-16 ***
A2LK12	0,910985	0,178899	5,092	3,54E-07 ***
A2LK13	0,749391	0,202442	3,702	0,000214 ***
A2LK14	0,595182	0,207764	2,865	0,004174 **
A2LK21	0,882424	0,049336	17,886	< 2e-16 ***
A2LK22	0,94693	0,065912	14,367	< 2e-16 ***
A2LK23	0,842363	0,032518	25,904	< 2e-16 ***
A2LK24	1,095068	0,056654	19,329	< 2e-16 ***
A2LK25	1,151063	0,083887	13,722	< 2e-16 ***
A2LK26	0,471185	0,037886	12,437	< 2e-16 ***
A2LK31	0,617297	0,041846	14,752	< 2e-16 ***
A2LK32	0,241203	0,023444	10,289	< 2e-16 ***
A2LK33	0,642194	0,04254	15,096	< 2e-16 ***
A2LK34	0,527852	0,038265	13,795	< 2e-16 ***
A2LK35	0,946709	0,102629	9,225	< 2e-16 ***
A2LK41	0,901348	0,024787	36,363	< 2e-16 ***
A2LK42	0,480495	0,044865	10,71	< 2e-16 ***
A2LK43	0,516619	0,068494	7,543	4,61E-14 ***
A2LK44	-0,081812	0,037085	-2,206	0,027381 *
A2LK51	0,135547	0,024199	5,601	2,13E-08 ***
A2LK52	0,426818	0,02441	17,485	< 2e-16 ***
A2LK53	0,174222	0,020762	8,391	< 2e-16 ***
A2LK54	0,295386	0,037313	7,916	2,44E-15 ***
A2LK61	-0,186392	0,037185	-5,013	5,37E-07 ***
A2LK62	-0,063315	0,088204	-0,718	0,472869
A2LK72	0,311822	0,0208	14,992	< 2e-16 ***
A2LK73	0,233971	0,084959	2,754	0,005888 **
A2LK74	0,410353	0,032532	12,614	< 2e-16 ***
A2LK75	-0,225396	0,048183	-4,678	2,90E-06 ***
A2LK81	-0,022211	0,024512	-0,906	0,364883
A2LK82	0,268182	0,040434	6,633	3,30E-11 ***
A2LK83	-0,227331	0,02298	-9,892	< 2e-16 ***
A2LK91	-0,082166	0,029353	-2,799	0,005122 **
A2LK92	-0,007581	0,214704	-0,035	0,971835
A2LK93	-0,030001	0,025849	-1,161	0,245798
A2LK94	0,201005	0,032024	6,277	3,46E-10 ***
A2LK95	-0,163668	0,247268	-0,662	0,508033
A2LK96	-0,236866	0,056511	-4,192	2,77E-05 ***

Taulukko 5.6. ZINB menetelmällä saatujen nolla-arvojen tulokset

Ammattiluokka	Positiiviset exp(estimaatti)	nolla-arvot exp(estimaatti)
(Intercept)	28,4769858	1,1578083
A2LK11	1,223717	2,1953598
A2LK12	0,6019281	2,4867713
A2LK13	1,2220587	2,1157112
A2LK14	1,2602342	1,8133618
A2LK21	0,9967856	2,4167503
A2LK22	0,9850613	2,5777848
A2LK23	0,9722974	2,3218479
A2LK24	0,9824729	2,9893847
A2LK25	0,9117893	3,1615525
A2LK26	0,9470378	1,6018913
A2LK31	0,9936654	1,8539107
A2LK32	0,8663328	1,2727788
A2LK33	1,0115513	1,9006471
A2LK34	0,8858237	1,6952866
A2LK35	0,8832918	2,5772133
A2LK41	1,0406646	2,4629216
A2LK42	0,8940356	1,6168741
A2LK43	0,9360781	1,67635
A2LK44	0,8164229	0,9214455
A2LK51	0,8368274	1,1451634
A2LK52	0,7443027	1,5323744
A2LK53	0,8104071	1,19032
A2LK54	0,9662899	1,3436446
A2LK61	0,9771226	0,8299478
A2LK62	0,9453618	0,9386479
A2LK72	0,8719014	1,3659117
A2LK73	0,8777591	1,2636078
A2LK74	0,9608646	1,5073495
A2LK75	0,8092813	0,7981999
A2LK81	0,8705138	0,9780342
A2LK82	0,8413402	1,3075855
A2LK83	1,1384198	0,7966569
A2LK91	0,9612607	0,9211189
A2LK92	1,3480137	0,9924481
A2LK93	0,8194709	0,970445
A2LK94	0,6563053	1,2226304
A2LK95	0,6680303	0,849024
A2LK96	0,7612432	0,7890971

lokseksi, kun selittävänä muuttujana on tällä kertaa A1LK eli vain ammattiluokkien pääluokat. Tämä tehdään sen vuoksi, koska on toimialoja, joissa selkeästi enemmän tietyiltä ammattiluokilta työntekijöitä ja mikäli aineisto ensin jaetaan toimialoittain ja sen jälkeen vielä n. 40 eri ammattiluokkaan, jää osaan ammattiluokista todella vähän sattuneita tapaturmia. Taulukosta 5.8 voidaan nähdä, millaisia tuloksia saadaan ZINB-menetelmällä, kun selittävänä muuttujana on päätason ammattiluokitus ja referenssi luokkana on 7 eli rakennus-, korjaus- ja valmistustyöntekijät. Selkeästi voidaan nähdä, että positiivisilla arvoilla ammattiluokat 2 (erityisasiantuntijat), 4 (toimisto- ja asiakaspalvelutyöntekijät) ja 6 (Maanviljelijät, metsätyöntekijät ym.) eivät ole tilastollisesti merkitseviä. Tässä voidaan myös tulkita samalla tavalla kuten aikaisemmin eli positiivinen estimaatin arvon omaavalla ammattiluokalla on suurempi todennäköisyys vakavampaan tapaturmaan kuin referenssiluokalla. Negatiivisilla arvoilla päinvastoin. Taulukossa 5.9 on nolla-arvojen tulokset.

Nyt kun tiedetään, minkälainen tulos saadaan koko aineistolla, voidaan siirtyä tutkimaan eri toimialoille saatuja tuloksia. Vertailemalla näitä keskenään ja juuri esiteltyyn koko aineiston tuloksiin, voidaan tehdä johtopäätöksiä, onko minkälaisia eroja toimialojen väleillä. Esitellään seuraavaksi, miten vahingot ovat jakautuneet eri toimialoilla ammattiluokittain. Nämä tulosteet löytyvät liitteistä taulukosta 4 tekstin selkeyttämisen vuoksi. Tästä taulukosta voidaan havaita, että selkeästi ammattiluokkien vahinkoja jakautuu toimialan mukaisesti. Esimerkiksi ammattiluokka 6 eli maanviljelijät, metsätyöntekijät ym. painottuvat vahvasti toimialaluokkaan A, joka on maatalous, metsätalous ja kalatalous. Toimialaluokka Z on poikkeava muista toimialaluokista, koska se ei varsinaisesti ole toimialaluokka vaan sinne on luokiteltu kuntien ja kuntayhtymien työntekijöiden tapaturmat. Jätetään pois tarkastelusta pois kaikki toimialat, joissa on ammattiluokkia, joissa on sattunut alle 10 kpl vahinkoja eli A, B, D, E, H, J, K, L ja T sekä lisäksi kuntien ja kuntayhtymien toimialaluokka Z.

Kaikkien mukana olevien toimialojen tulokset löytyvät tutkielman lopussa ole-

Taulukko 5.7. Ammattiluokkien johtajat, ylimmät virkamiehet ja järjestöjen johtajat ammattiluokalla (A2LK11) ja rakennustyöntekijät (pl. sähköasentajat) (A2LK71) tapaturmien lukumäärät vakavuusluokkien mukaan.

Vakavuusluokka	Johtajat	%	Rakennustyöntekijät	%
0	475	74,22	14891	58,19
5	37	5,78	2956	11,55
10	41	6,41	3580	13,99
22	30	4,69	1805	7,05
60	47	7,34	1556	6,08
135	5	0,78	473	1,85
272	3	0,47	325	1,27
365	2	0,31	5	0,02
	640	100	25591	100

vista liitteistä. Taulukkoon 5.10 on koottu yhteenvetona toimialaluokittaiset tulokset ja esitetty mitkä ammattiluokat ovat missäkin tilastollisesti merkitsevät ja taulukkoon 5.11 on esitetty samalla tavalla estimaatit. Tuloksista ammattiluokka 6 (Maanviljelijät, metsätyöntekijät ym.) ei ole missään tarkasteltavissa toimialaluokissa tilastollisesti merkitsevä, eikä se ollut koko aineistonkaan osalta tilastollisesti merkitsevä. Edellä mainittuja taulukkoja tarkastelemalla voidaan selkeästi havaita toimialaluokkien välisiä eroja ammattiluokkien merkitsevyyksissä sekä myös estimaattien arvoissa. Tarkastellaan seuraavaksi ammattiluokittain eroavaisuuksia toimialojen välillä.

A1LK - Johtajat

Kyseinen ammattiluokka ei ole tilastollisesti merkitsevä missään tarkastelussa olevista toimialaluokista vaikka se on tilastollisesti merkitsevä koko aineiston osalla. Tarkasteltaessa nollaluokan tuloksia, voidaan havaita, että ammattiluokka on tilastollisesti merkitsevä toimialaluokissa G, N, O ja G. Kaikissa näissä toimialoissa estimaatin arvot ovat positiivisia kuten myös tarkasteltaessa koko aineistoa. Estimaattien arvot eroavat suuruudelta toisistaan, mutta pitää myös huomioida, vaikka käytetäänkin samaa referenssiluokkaa voivat nämä erota toimialoittainkin

A2LK - Erityisasiantuntijat

Erityisasiantuntijat ovat positiivisilla arvoilla tilastollisesti merkitseviä toimialaluokissa C, F, Q ja S. Koko aineistossa estimaatin arvo on positiivinen, mutta toimialaluokissa F ja Q eli rakentaminen ja terveys- ja sosiaalipalvelut voidaan havaita estimaattien olevan negatiivisia, joten näiden osalta saadaan päinvastainen tulos verrattuna koko aineistolla saatavaan tulokseen. Pitää kuitenkin huomioida ettei kyseinen ammattiluokka ole tilastollisesti merkitsevä koko aineiston osalta. Kun tarkastellaan nollaluokan tuloksia niin ainoastaan toimialaluokka I ei ole tilastollisesti merkitsevä. Kaikissa toimialaluokissa estimaattien arvot ovat positiivisia kuten myös koko aineistossa on, joten näiden osalta ei ole eroja

A3LK - Asiantuntijat

Tässä ammattiluokassa tilastollisesti merkitseviä toimialaluokkia positiivisille arvoille on C, G ja S. Koko aineiston estimaatti on negatiivinen, mutta näiden toimialaluokkien estimaatin arvo on positiivinen, joten näiden osalta saadaan eroavaisuutta koko aineiston ja kyseisten toimialojen osalta. Nollaluokan osalta tilastollisesti merkitseviä ovat kaikki muut paitsi toimialaluokat I ja R. Estimaattien arvot ovat positiivisia koko aineistolle sekä kaikkien merkitsevien toimialaluokkien osalta.

A4LK - Toimisto- ja asiakaspalvelutyöntekijät

Tilastollisesti merkitseviä toimialaluokkia positiivisille arvoille ovat C, I ja S. Toimialaluokka I eli majoitus- ja ravitsemusalan osalta estimaatin arvo on negatiivinen, vaikka se koko aineiston osalta on positiivinen. Huomioitavaa on ettei ammattiluokka ole tilastollisesti merkitsevä koko aineiston osalta. Nollaluokan osalta kyseinen

ammattiluokka on kaikissa toimialaluokissa tilastollisesti merkitsevä ja estimaatit ovat myös kaikkien osalta positiivisia, kuten on koko aineiston mallissa.

A5LK - Palvelu- ja myyntityöntekijät

Toimialaluokissa G, I, Q ja S tilastollisesti merkitsevä positiivisilla arvoilla. Koko aineistossa estimaatin arvo on negatiivinen, mutta toimialaluokassa S estimaati on positiivinen eli tämän luokan osalta on päinvastainen tulos verrattuna koko aineistoon. Nollaluokka tilastollisesti merkitsevä toimialoissa C, F, M, N ja O eli selkeästi vähemmän toimialaluokkia on tilastollisesti merkitseviä kuin ammattiluokkien A2LK, A3LK ja A4Lk osalta oli huomattavasti useampi toimiala tilastollisesti merkitsevä. Koko aineiston osalta estimaatin arvo on positiivinen, kuten on kaikkien tilastollisesti merkitsevien toimialaluokkienkin osalta.

A6LK - Maanviljelijät, metsätyöntekijät ym.

Edellä jo mainittiin, positiivisten arvojen tarkastelussa olevien toimialaluokkien osalta kyseinen ammattiluokka ei ole missään tilastollisesti merkitsevä. Ammattiluokka ei ole tilastollisesti merkitsevä koko aineistonkaan tarkastelussa. Nollaluokkien osalta ammattiluokka on toimialojen C, M, N ja O osalta tilastollisesti merkitsevä. Koko aineiston osalta estimaatin arvo on negatiivinen, mutta toimialaluokkien C, M ja O:n osalta estimaatin arvo on positiivinen.

A8LK - Prosessi- ja kuljetustyöntekijät

Ainoastaan F ja Q toimialaluokat ovat tilastollisesti merkitseviä ja näistä toimialaluokan Q estimaatti on lisäksi negatiivinen, kun taas koko aineiston osalta se on positiivinen. Nollaluokka on tilastollisesti merkitsevä toimialaluokissa C, F ja G. Näissä kaikissa estimaatin arvo on negatiivinen kuten on myös koko aineiston osaltakin.

A9LK - Muut työntekijät

Toimialaluokat F, G, I, N ja Q ovat tilastollisesti merkitseviä ja toimialaluokassa F estimaatti on myös positiivinen, vaikka koko aineiston osalta se on negatiivinen. Nollaluokan osalta ainoastaan toimialaluokat C, G ja S ovat tilastollisesti merkitsevät tämän ammattiluokan osalta. Kaikissa näissä estimaatit ovat negatiivisia kuten on myös koko aineiston osalta.

Voidaan siis todeta kaikkien tarkastelujen jälkeen, että toimialojen välillä on selkeästi eroja miten vahinkojen vakavuusluokkaa ZINB-menetelmä mallintaa. Koko aineiston osalta ammattiluokkien 1, 3, 5, 7, 8 ja 9 malli antoi tilastollisesti merkitseviä tuloksia positiivisille arvoille. Nollaluokan tuloksissa malli antoi kaikkien ammattiluokkien osalta tilastollisesti merkitsevät tulokset. Kun sama malli tehtiin jaettuna toimialaluokkiin niin osassa toimialaluokista ei tullut ollenkaan tilastollisesti merkitseviä ammattiluokkia, mitkä olisivat eronneet referenssiluokastaan. Näiden osalta

on tärkeä huomata, että osittain saadut tulokset johtuvat siitä, että kun referenssiluokkana oli rakennus- korjaus- ja valmistustyöntekijät, jolloin osassa toimialoista tämä luokka on selkeästi pienemmässä osassa kuin muut ammattiluokat.

Taulukko 5.8. ZINB menetelmällä saadut tulokset päätason ammattiluokka selittävänä muuttujana.

	Estimaatti	Keskihajonta	z-arvo	p-arvo
(Intercept)	3,286632	0,008938	367,696	< 2e-16 ***
A1LK1	0,181776	0,077121	2,357	0,01842 *
A1LK2	0,027855	0,020952	1,33	0,18368
A1LK3	-0,043821	0,016559	-2,646	0,00813 **
A1LK4	0,009458	0,018315	0,516	0,60559
A1LK5	-0,144232	0,012947	-11,14	< 2e-16 ***
A1LK6	0,033035	0,028843	1,145	0,25206
A1LK8	0,067426	0,014246	4,733	2,21E-06 ***
A1LK9	-0,132307	0,015215	-8,696	< 2e-16 ***

Taulukko 5.9. ZINB menetelmällä saatujen nolla-arvojen tulokset päätason ammattiluokka selittävänä muuttujana

	Estimaatti	Keskihajonta	z-arvo	p-arvo
(Intercept)	0,296439	0,009795	30,265	< 2e-16 ***
A1LK1	0,634244	0,072487	8,75	< 2e-16 ***
A1LK2	0,656945	0,020169	32,572	< 2e-16 ***
A1LK3	0,249897	0,017044	14,662	< 2e-16 ***
A1LK4	0,46676	0,018196	25,652	< 2e-16 ***
A1LK5	0,085403	0,013746	6,213	5,20E-10 ***
A1LK6	-0,323363	0,033507	-9,651	< 2e-16 ***
A1LK8	-0,238956	0,015877	-15,05	< 2e-16 ***
A1LK9	-0,155108	0,016806	-9,229	< 2e-16 ***

Taulukko 5.10. Toimialoittain ammattiluokkien tilastolliset merkitsevyydet ZINB malleissa

	C	F	G	I	M	N	O	P	Q	R	S
(Intercept)	***	***	***	***	***	***	***	***	***	***	***
A1LK1											
A1LK2	***	*	.						*		*
A1LK3	**		*	.					.		*
A1LK4	***		.	*	.						**
A1LK5		.	*	*					*		*
A1LK6											
A1LK8		***	.						*		
A1LK9		***	***	*		***			*		
zero-inflation											
(Intercept)	***	***	***		***					**	.
A1LK1		.	**			**	***		*		
A1LK2	***	**	***		***	***	***	**	**	*	**
A1LK3	***	***	***		***	***	***	***	*		*
A1LK4	***	***	***	*	***	***	***	**	*	*	**
A1LK5	**	***			**	***	*				
A1LK6	***				**	*	**				
A1LK8	***	**	***		.						.
A1LK9	**		***								*

¹Merkitsevyyden selitteet: p-arvon arvot *** - 0, ** < 0,001, * < 0,05, . < 0,1

Taulukko 5.11. Toimialoittain ammattiluokkien tilastolliset merkitsevyydet
ZINB malleissa

	C	F	G	I	M	N	O	P	Q	R	S
(Intercept)	0,70	0,78	0,65	0,86	0,75	0,70	0,81	0,78	0,96	0,67	0,58
A1LK1	0,09	0,11	0,15	-0,29	0,13	-0,24	0,14	0,44	0,16	-0,08	0,02
A1LK2	0,17	-0,30	0,17	0,07	0,06	-0,11	-0,02	0,03	-0,32	0,20	0,29
A1LK3	0,11	0,05	0,13	-0,37	0,02	0,08	0,00	0,19	-0,24	0,15	0,24
A1LK4	0,16	-0,01	0,08	-0,36	0,11	-0,08	-0,05	0,04	-0,17	0,07	0,30
A1LK5	0,04	0,13	-0,06	-0,30	-0,06	-0,02	-0,01	-0,06	-0,29	0,12	0,25
A1LK6	0,10	0,03	0,01	0,18	0,00	0,00	-0,05	-0,04	0,31	0,09	0,12
A1LK8	0,00	0,08	0,07	0,29	0,02	0,00	0,05	-0,13	-0,41	0,04	0,00
A1LK9	0,00	0,09	-0,10	-0,28	0,01	-0,08	-0,01	0,02	-0,30	-0,09	-0,04
zero-inflation											
(Intercept)	0,29	0,21	0,48	0,13	0,37	-0,02	-0,01	0,35	0,04	0,49	0,30
A1LK1	0,35	0,68	0,86	0,51	0,29	2,91	0,95	0,51	0,88	0,47	0,02
A1LK2	0,66	0,48	0,57	1,01	0,81	0,94	1,12	0,90	0,78	0,44	0,71
A1LK3	0,29	0,54	0,49	0,31	0,51	0,62	0,60	1,00	0,59	0,18	0,42
A1LK4	0,72	0,82	0,61	0,66	0,88	0,79	0,75	0,84	0,70	0,46	0,57
A1LK5	0,31	0,61	-0,02	0,19	0,40	0,41	0,39	0,27	0,31	0,35	0,05
A1LK6	0,78	-0,08	-0,33	0,63	0,74	0,31	-0,82	-0,14	-0,30	-0,44	0,11
A1LK8	-0,16	-0,16	-0,44	0,49	-0,23	-0,08	-0,36	0,53	0,49	0,17	-0,42
A1LK9	-0,17	0,05	-0,34	0,34	0,15	-0,03	0,09	-0,31	0,17	-0,21	-0,52

6 Yhteenveto

Tässä tutkielmassa tarkoituksena oli tutustua työtaturma- ja ammattitautivakuutukseen ja selvittää minkälaisia tuloksia saadaan mallintamalla työ- ja työmatkataturmien vakavuuksia ammattiluokalla sekä löytyykö eri toimialaluokkien väliltä eroja. Mallintaminen tapahtui RStudio-ohjelmalla ja mallintamisen yhteydessä tarkasteltiin eri nolla-arvoja huomioon ottavien menetelmien eroja lopputuloksissa.

Aluksi tarkasteltiin ZIP-, ZAP-, ZINB- ja ZANB-menetelmillä saatuja tuloksia ja estimaatteja ja todettiin ZIP ja ZAP -menetelmien tulokset olivat lähes samanlaiset sekä ZINB ja ZANB-menetelmien tulokset olivat lähes samanlaiset. AIC arvoja takasteltaessa todettiin, että ZINB ja ZANB-menetelmät soveltuivat mallintamaan paremmin kuin ZIP ja ZAP-menetelmät.

Tarkempi tarkastelu tehtiin ZINB-menetelmällä saatuihin tuloksiin ja positiivisille arvoille saatiinkin tilastollisesti merkitseviksi ammattiluokiksi taulukossa 6.1 olevat tulokset. Referenssi ammattiluokkana oli A2LK71 eli rakennustyöntekijät ym. (pl. sähköasentajat). Suurimmalla osalla tilastollisesti merkitsevistä ammattiluokista estimaatin arvo on negatiivinen mikä tarkoitti, että ammattiluokan vakavan vahingon riski pienenee, mikäli vahinko kuuluu positiivisiin luokkiin.

Taulukko 6.1. ZINB-menetelmällä saadut tilastollisesti merkitsevä tulokset positiivisille arvoille.

	Estimaatti	Keskihajonta	z-arvo	p-arvo
(Intercept)	3,349096	0,012474	268,481	< 2e-16 ***
A2LK11	0,201893	0,100563	2,008	0,04468 *
A2LK12	-0,507617	0,191616	-2,649	0,00807 **
A2LK32	-0,143486	0,021836	-6,571	5,00E-11 ***
A2LK34	-0,121237	0,038135	-3,179	0,00148 **
A2LK42	-0,11201	0,044506	-2,517	0,01184 *
A2LK44	-0,202823	0,03241	-6,258	3,90E-10 ***
A2LK51	-0,178137	0,022148	-8,043	8,77E-16 ***
A2LK52	-0,295307	0,023394	-12,623	< 2e-16 ***
A2LK53	-0,210219	0,019033	-11,045	< 2e-16 ***
A2LK72	-0,137079	0,01946	-7,044	1,86E-12 ***
A2LK75	-0,211609	0,040387	-5,24	1,61E-07 ***
A2LK81	-0,138672	0,021818	-6,356	2,07E-10 ***
A2LK82	-0,172759	0,03834	-4,506	6,61E-06 ***
A2LK83	0,129641	0,019821	6,541	6,13E-11 ***
A2LK93	-0,199096	0,022931	-8,682	< 2e-16 ***
A2LK94	-0,421129	0,029656	-14,2	< 2e-16 ***
A2LK96	-0,272802	0,047011	-5,803	6,52E-09 ***

Nolla-luokan tuloksissa melkein jokainen ammattiluokka oli tilastollisesti mer-

kitsevä ja ne tulokset löytyvät taulukosta 5.5. Näistä ainoastaan kuusi ammattiluokkaa sai negatiivisia arvoja eli näillä oli pienempi todennäköisyys kuulua nolla-luokkaan kuin referenssi ammattiluokalla A2LK71.

Viimeisessä vaiheessa tarkasteltiin, minkälaisia eroja löydetään toimialaluokkien väleillä. Havaittiin, että toimialaluokkien väliltä löydettiin selkeästi eroja ja nämä erosivat monelta osin, myös kun vertailtiin koko aineistolla saatuihin tuloksiin. Tilastollisesti vertailua ei voida kuitenkaan suoraan tehdä, koska mallintaminen toteutettiin jaettuina aineistoina. Osittain toimialaluokkien välisiä eroja selittää paljon se, että toimialojen välillä on paljon eroja siinä, minkälaisia ammattiluokkia toimialaluokassa eniten esiintyy. Tämä aiheuttaa myös se, että referenssi ammattiluokkana toiminut A1LK7 eli rakennus-, korjaus- ja valmistustyöntekijät luokan koko vaihtelee toimialaluokkien välillä.

Tutkielma antoi mielenkiintoisia tuloksia ja käytössä oli laaja aineisto TVK:lta. Kuitenkin nyt tehdyssä mallissa heikkoutena on, se ettei siinä huomioida itse tapaturman sattumisen todennäköisyyttä eri ammattiluokkien välillä. Varsinkin työtapaturmissa ammattiluokkien välillä on eroja riskien määrässä. Mahdollisuus olisi kehittää mallia niin sanotulla offset-muuttujalla, joka skaalaisi mallia yrityksen koon mukaan missä vahinko on sattunut. Tällaista tietoa ei kuitenkaan ollut nyt mahdollisuus käyttää. Lisäksi käytettävissä ollut aineisto ei sisältänyt työtapaturma- ja ammattitautivakuutuksesta korvattavia ammattitautitietoja, jotka kuitenkin vaikuttavat vakuutuksen maksuun. Nämä vahingot kuitenkin ovat luonteeltaan erilaisia niiden pitkäkestoisuuden vuoksi, jonka vuoksi näistä voisi tehdä erillisen tutkielman. Aineistossa vahingot oli luokiteltu monella eri tavalla ja siitä olisikin mahdollista tehdä erilaisia tutkimuksia jatkossa.

Lähteet

- Cameron, A. C. & Trivedi, P. K. (1998), *"Regression Analysis of Count Data"*, Cambridge University Press, s.123-127.
- Hilbe, J. M. (2011), *"Negative binomial regression"*, Cambridge University Press
- Lambert, D. (1992), *"Zero-Inflated Poisson Regression with an Application to Defects in Manufacturing"*, Technometrics, 34
- Owens, J.W., Shrestha, S., & Chaparro, B.S. (2009), Effects of text saliency on eye movements while browsing a web portal. *"Proceedings of the Human Factors and Ergonomics Society Annual Meeting"*, 53, s. 1257-1261.
- Salo, Kirsi. (2015), *"Työtäpaturma- ja ammattitautivakuutus"*, Finva: Vantaa
- Smithson, M. & Merkle E. C. (2013), *"Generalized linear models for categorical and continuous limited dependent variables"*, Chapman and Hall/CRC
- Zuur, A. F., Ieno, E. N., Walker, N. J., Saveliev, A. A., & Smith, G. M. (2009), *"Mixed Effects Models and Extensions in Ecology with R"*, Springer Science,
- Sähköiset lähteet:**
- Tapaturmavakuutuskeskus, palkansaajien työtäpaturmatilasto'*, Tietokanta on muodostettu 27.2.2019 TVK/Janne Sysi-Aho.
- Tapaturmavakuutuskeskus
Saatavilla internetissä: <https://www.tvk.fi/tietopalvelu-ja-julkaisut/tilastot/luokitukset/> 2.10.2019
- *Jäsenlaitokset*
Saatavilla internetissä: <http://www.tvk.fi/tapaturmavakuutuskeskus/jasenlaitokset/> 24.10.2019
- *Avoin data*
Saatavilla internetissä: <https://www.tvk.fi/tietopalvelu-ja-julkaisut/avoin-data/> 24.10.2019
- *Riskiluokitus*
Saatavilla internetissä: <https://www.tvk.fi/tietopalvelu-ja-julkaisut/tulorekisterin-ammattiluokkien-riskiluokitus/> 30.11.2019
- Tilastokeskus, *Ammattiluokat*
Saatavilla internetissä: <https://www.tilastokeskus.fi/meta/luokitukset/ammatti/001-2010/kuvaus.html> 7.6.2019
- *Toimialaluokat*
Saatavilla internetissä: <https://www.stat.fi/meta/luokitukset/toimiala/001-2008/kuvaus.html> 24.10.2019
- IF, *Työtäpaturmavakuutukset ja -korvaukset vakuutusopas*
Saatavilla internetissä : https://www.if.fi/web/fi/sitecollectiondocuments/commercial/henkilovakuutukset/lakisaateinen_tapaturmavakuutus/1401919_50962_lakis_tapaturma_ja_korvaukset_su_01_2015.pdf 24.10.2019
- Finlex, *Työtäpaturma- ja ammattitautilaki 459/2015"*
Saatavilla internetissä: <https://www.finlex.fi/fi/laki/alkup/2015/20150459> 24.10.2019

Liite 1: Taulukoita joihin viitattu tekstissä

Taulukko 2. Toimialaluokat

Koodi	Luokan nimi
A	Maatalous, metsätalous ja kalatalous
B	Kaivostoiminta ja louhinta
C	Teollisuus
D	Sähkö-, kaasu- ja lämpöhuolto, jäähdytysliiketoiminta
E	Vesihuolto, viemäri- ja jätevesihuolto, jätehuolto ja muu ympäristön puhtaanapito
F	Rakentaminen
G	Tukku- ja vähittäiskauppa; moottoriajoneuvojen ja moottoripyörien korjaus
H	Kuljetus ja varastointi
I	Majoitus- ja ravitsemistoiminta
J	Informaatio ja viestintä
K	Rahoitus- ja vakuutustoiminta
L	Kiinteistöalan toiminta
M	Ammatillinen, tieteellinen ja tekninen toiminta
N	Hallinto- ja tukipalvelutoiminta
O	Julkinen hallinto ja maanpuolustus; pakollinen sosiaalivakuutus
P	Koulutus
Q	Terveys- ja sosiaalipalvelut
R	Taiteet, viihde ja virkistys
S	Muu palvelutoiminta
T	Kotitalouksien toiminta työnantajina
Z	Kunnat ja kuntayhtymät

Taulukko 3. Ammattiluokat

Koodi	Ammattiluokka
1	Johtajat
2	Erityisasiantuntijat
3	Asiantuntijat
4	Toimisto- ja asiakaspalvelutyöntekijät
5	Palvelu- ja myyntityöntekijät
6	Maanviljelijät, metsätyöntekijät ym.
7	Rakennus-, korjaus- ja valmistustyöntekijät
8	Prosessi- ja kuljetustyöntekijät
9	Muut työntekijät
11	Johtajat, ylimmät virkamiehet ja järjestöjen johtajat
12	Hallintojohtajat ja kaupalliset johtajat
13	Tuotantotoiminnan ja yhteiskunnan peruspalvelujen johtajat
14	Hotelli- ja ravintola-alan, vähittäiskaupan ja muiden palvelualojen johtajat
21	Luonnontieteiden ja tekniikan erityisasiantuntijat
22	Terveystieteiden erityisasiantuntijat
23	Opettajat ja muut opetusalan erityisasiantuntijat
24	Liike-elämän ja hallinnon erityisasiantuntijat
25	Tieto- ja viestintäteknologian erityisasiantuntijat
26	Lainopilliset, sosiaalialan ja kulttuurialan erityisasiantuntijat
31	Luonnontieteiden ja tekniikan asiantuntijat
32	Terveystieteiden asiantuntijat
33	Liike-elämän ja hallinnon asiantuntijat
34	Lainopilliset avustajat sekä sosiaali- ja kulttuurialan asiantuntijat
35	Informaatio- ja tietoliikenneteknologian asiantuntijat
41	Toimistotyöntekijät
42	Asiakaspalvelutyöntekijät
43	Laskennan ja varastoinnin toimistotyöntekijät
44	Muut toimisto- ja asiakaspalvelutyöntekijät
51	Palvelutyöntekijät
52	Myyjät, kauppiat ym.
53	Hoivapalvelun ja terveystieteiden työntekijät
54	Suojelu- ja vartiointityöntekijät
61	Maanviljelijät ja eläintenkasvattajat ym.
62	Metsä- ja kalatalouden työntekijät
63	Kotitarveviljelijät, -kalastajat ja -metsästäjät
71	Rakennustyöntekijät ym. (pl. sähköasentajat)
72	Konepaja- ja valimotyöntekijät sekä asentajat ja korjaajat
73	Käsityötuotteiden valmistajat, hienomekaanikot sekä painoalan työntekijät
74	Sähkö- ja elektroniikka-alan työntekijät
75	Elintarvike-, puutyö- ja vaatetus- ja jalkinealan valmistustyöntekijät ym.
81	Prosessityöntekijät
82	Teollisuustuotteiden kokoonpanijat
83	Kuljetustyöntekijät
91	Siiivoajat, kotiapulaiset ja muut puhdistustyöntekijät
92	Maa-, metsä- ja kalatalouden avustajat työntekijät
93	Teollisuuden ja rakentamisen avustajat työntekijät
94	Avustajat keittiö- ja ruokatyöntekijät
95	Katunmyyjät, kengänkiillottajat ym.
96	Katujen puhtaanapidon ja jätehuollon työntekijät ym.

Taulukko 4. Ristiintaulukoidut lukumäärät ammattiluokat jaettuna toimialoit-
tain

	1	2	3	4	5	6	7	8	9
A	4	96	137	71	74	1611	139	585	73
B	0	22	17	30	3	2	83	352	47
C	62	817	1051	2354	533	119	16663	12220	1580
D	9	82	57	161	23	1	369	157	21
E	2	44	73	73	25	3	330	464	187
F	34	214	601	693	298	79	23044	2293	1974
G	98	247	1097	1659	11697	81	5884	1039	3648
H	23	129	352	3122	908	9	794	9114	2656
I	28	18	121	331	3675	10	71	33	2945
J	30	1060	454	1005	139	5	228	15	222
K	19	182	114	1783	91	6	95	33	37
L	8	28	380	381	579	27	236	57	189
M	58	1598	934	1389	344	87	1114	486	265
N	24	203	341	1111	5106	248	4661	2117	6318
O	475	1197	1897	949	1982	130	213	54	451
P	21	1433	435	768	466	89	71	16	140
Q	52	1107	3867	1080	5558	11	62	115	587
R	26	517	685	340	434	119	166	61	157
S	48	276	831	794	925	248	251	217	304
T	0	0	4	1	219	75	193	7	18
Z	78	9043	12219	4391	17969	1854	1030	744	4226

Taulukko 5. Toimialojen C ja F tulokset

C	Estimaatti	Keskihajonta	z-arvo	p-arvo
(Intercept)	0,696721	0,0106	65,726	< 2e-16 ***
A1LK1	0,091106	0,171701	0,531	0,595688
A1LK2	0,169554	0,050528	3,356	0,000792 ***
A1LK3	0,112959	0,041382	2,73	0,00634 **
A1LK4	0,164438	0,031546	5,213	1,86E-07 ***
A1LK5	0,038252	0,061006	0,627	0,530643
A1LK6	0,09547	0,143738	0,664	0,506568
A1LK8	-0,002747	0,015143	-0,181	0,856028
A1LK9	0,003586	0,032754	0,109	0,912822
zero-inflation	Estimaatti	Keskihajonta	z-arvo	p-arvo
(Intercept)	0,28974	0,01936	14,966	< 2e-16 ***
A1LK1	0,35227	0,30087	1,171	0,241676
A1LK2	0,66486	0,08711	7,633	2,30E-14 ***
A1LK3	0,29158	0,0743	3,924	8,69E-05 ***
A1LK4	0,72108	0,05408	13,333	< 2e-16 ***
A1LK5	0,31371	0,10543	2,976	0,002925 **
A1LK6	0,7826	0,23385	3,347	0,000818 ***
A1LK8	-0,15621	0,02848	-5,484	4,16E-08 ***
A1LK9	-0,16848	0,06288	-2,679	0,007381 **
F	Estimaatti	Keskihajonta	z-arvo	p-arvo
(Intercept)	0,783797	0,007832	100,078	< 2e-16 ***
A1LK1	0,105125	0,233377	0,45	0,652384
A1LK2	-0,303273	0,120694	-2,513	0,01198 *
A1LK3	0,04844	0,055718	0,869	0,384641
A1LK4	-0,007027	0,059768	-0,118	0,906411
A1LK5	0,1264	0,076179	1,659	0,097064 .
A1LK6	0,027126	0,128281	0,211	0,832531
A1LK8	0,080514	0,023688	3,399	0,000676 ***
A1LK9	0,087496	0,026552	3,295	0,000983 ***
zero-inflation	Estimaatti	Keskihajonta	z-arvo	p-arvo
(Intercept)	0,21488	0,0152	14,14	< 2e-16 ***
A1LK1	0,68003	0,41028	1,657	0,09742 .
A1LK2	0,48272	0,18526	2,606	0,00917 **
A1LK3	0,53893	0,09781	5,51	3,58E-08 ***
A1LK4	0,8201	0,09742	8,419	< 2e-16 ***
A1LK5	0,61414	0,13701	4,482	7,38E-06 ***
A1LK6	-0,08243	0,25765	-0,32	0,74902
A1LK8	-0,15969	0,04938	-3,234	0,00122 **
A1LK9	0,05115	0,0527	0,971	0,33178

Taulukko 6. Toimialojen G ja I tulokset

G	Estimaatti	Keskihajonta	z-arvo	p-arvo
(Intercept)	0,6538	0,01949	33,541	< 2e-16 ***
A1LK1	0,14528	0,17732	0,819	0,412608
A1LK2	0,169	0,09941	1,7	0,089112 .
A1LK3	0,12552	0,05	2,51	0,012066 *
A1LK4	0,08353	0,04482	1,864	0,062346 .
A1LK5	-0,05693	0,0233	-2,443	0,014572 *
A1LK6	0,01496	0,14478	0,103	0,917706
A1LK8	0,07469	0,04162	1,795	0,072708 .
A1LK9	-0,10464	0,03006	-3,481	0,0005 ***
zero-inflation	Estimaatti	Keskihajonta	z-arvo	p-arvo
(Intercept)	0,48254	0,03313	14,563	< 2e-16 ***
A1LK1	0,86355	0,27693	3,118	0,001819 **
A1LK2	0,57473	0,16345	3,516	0,000438 ***
A1LK3	0,48614	0,08209	5,922	3,19E-09 ***
A1LK4	0,60683	0,07144	8,494	< 2e-16 ***
A1LK5	-0,01873	0,03955	-0,474	0,635812
A1LK6	-0,32919	0,27154	-1,212	0,225391
A1LK8	-0,43908	0,08038	-5,463	4,69E-08 ***
A1LK9	-0,33807	0,05335	-6,336	2,35E-10 ***
I	Estimaatti	Keskihajonta	z-arvo	p-arvo
(Intercept)	0,85568	0,1305	6,557	5,49E-11 ***
A1LK1	-0,29092	0,33136	-0,878	0,38
A1LK2	0,07306	0,36524	0,2	0,8415
A1LK3	-0,36781	0,19643	-1,872	0,0611 .
A1LK4	-0,3598	0,1637	-2,198	0,028 *
A1LK5	-0,2997	0,13274	-2,258	0,024 *
A1LK6	0,18164	0,39023	0,465	0,6416
A1LK8	0,28539	0,22235	1,283	0,1993
A1LK9	-0,27523	0,13342	-2,063	0,0391 *
zero-inflation	Estimaatti	Keskihajonta	z-arvo	p-arvo
(Intercept)	0,1325	0,2672	0,496	0,6201
A1LK1	0,5081	0,5555	0,915	0,3604
A1LK2	1,0123	0,6473	1,564	0,1178
A1LK3	0,312	0,3594	0,868	0,3854
A1LK4	0,6632	0,306	2,167	0,0302 *
A1LK5	0,1924	0,2705	0,711	0,4768
A1LK6	0,626	0,7641	0,819	0,4127
A1LK8	0,4929	0,4645	1,061	0,2887
A1LK9	0,3375	0,2712	1,244	0,2134

Taulukko 7. Toimialojen M ja N tulokset

M	Estimaatti	Keskihajonta	z-arvo	p-arvo
(Intercept)	0,754223	0,038092	19,8	<2e-16 ***
A1LK1	0,134673	0,16928	0,796	0,4263
A1LK2	0,055759	0,055302	1,008	0,3133
A1LK3	0,022175	0,061467	0,361	0,7183
A1LK4	0,109298	0,057018	1,917	0,0552 .
A1LK5	-0,057252	0,090074	-0,636	0,525
A1LK6	0,004041	0,178651	0,023	0,982
A1LK8	0,020665	0,065453	0,316	0,7522
A1LK9	0,006576	0,089903	0,073	0,9417
zero-inflation	Estimaatti	Keskihajonta	z-arvo	p-arvo
(Intercept)	0,36921	0,07007	5,269	1,37E-07 ***
A1LK1	0,29306	0,31041	0,944	0,34511
A1LK2	0,80666	0,09546	8,45	< 2e-16 ***
A1LK3	0,50889	0,10665	4,772	1,83E-06 ***
A1LK4	0,8783	0,09895	8,876	< 2e-16 ***
A1LK5	0,40462	0,15099	2,68	0,00737 **
A1LK6	0,74173	0,28637	2,59	0,0096 **
A1LK8	-0,22716	0,12627	-1,799	0,07203 .
A1LK9	0,1541	0,16081	0,958	0,33793
N	Estimaatti	Keskihajonta	z-arvo	p-arvo
(Intercept)	0,699134	0,018375	38,048	< 2e-16 ***
A1LK1	-0,238475	0,926717	-0,257	0,796922
A1LK2	-0,108909	0,12453	-0,875	0,381811
A1LK3	0,076108	0,075584	1,007	0,313965
A1LK4	-0,080682	0,051952	-1,553	0,120417
A1LK5	-0,023096	0,026129	-0,884	0,376751
A1LK6	0,004494	0,084765	0,053	0,957721
A1LK8	-0,004276	0,031319	-0,137	0,891401
A1LK9	-0,080849	0,023902	-3,383	0,000718 ***
zero-inflation	Estimaatti	Keskihajonta	z-arvo	p-arvo
(Intercept)	-0,0192	0,03683	-0,521	0,60214
A1LK1	2,91086	1,10843	2,626	0,00864 **
A1LK2	0,94428	0,18984	4,974	6,55E-07 ***
A1LK3	0,61533	0,1331	4,623	3,78E-06 ***
A1LK4	0,7928	0,08475	9,355	< 2e-16 ***
A1LK5	0,40665	0,04905	8,291	< 2e-16 ***
A1LK6	0,30533	0,15571	1,961	0,04989 *
A1LK8	-0,07743	0,06383	-1,213	0,22511
A1LK9	-0,0329	0,0476	-0,691	0,48935

Taulukko 8. Toimialojen O ja P tulokset

O	Estimaatti	Keskihajonta	z-arvo	p-arvo
(Intercept)	8,14E-01	7,49E-02	10,865	<2e-16 ***
A1LK1	1,41E-01	9,62E-02	1,461	0,144
A1LK2	-1,75E-02	8,77E-02	-0,2	0,841
A1LK3	3,33E-04	8,06E-02	0,004	0,997
A1LK4	-5,17E-02	8,80E-02	-0,588	0,557
A1LK5	-1,32E-02	7,99E-02	-0,166	0,869
A1LK6	-4,63E-02	1,13E-01	-0,409	0,682
A1LK8	5,21E-02	1,52E-01	0,343	0,731
A1LK9	-8,50E-03	9,18E-02	-0,093	0,926
zero-inflation	Estimaatti	Keskihajonta	z-arvo	p-arvo
(Intercept)	-0,01359	0,15743	-0,086	0,931221
A1LK1	0,9479	0,19145	4,951	7,38E-07 ***
A1LK2	1,11558	0,17391	6,415	1,41E-10 ***
A1LK3	0,60303	0,16628	3,627	0,000287 ***
A1LK4	0,7513	0,17585	4,272	1,93E-05 ***
A1LK5	0,39162	0,16571	2,363	0,018112 *
A1LK6	-0,81502	0,2881	-2,829	0,004671 **
A1LK8	-0,35842	0,35614	-1,006	0,314228
A1LK9	0,08752	0,19094	0,458	0,646689
P	Estimaatti	Keskihajonta	z-arvo	p-arvo
(Intercept)	0,78141	0,14701	5,315	1,06E-07 ***
A1LK1	0,43673	0,27508	1,588	0,112
A1LK2	0,0318	0,15331	0,207	0,836
A1LK3	0,1949	0,16398	1,189	0,235
A1LK4	0,03822	0,15797	0,242	0,809
A1LK5	-0,0555	0,16081	-0,345	0,73
A1LK6	-0,03537	0,19599	-0,18	0,857
A1LK8	-0,12898	0,4329	-0,298	0,766
A1LK9	0,01683	0,175	0,096	0,923
zero-inflation	Estimaatti	Keskihajonta	z-arvo	p-arvo
(Intercept)	0,3531	0,2747	1,285	0,198627
A1LK1	0,5144	0,5631	0,913	0,360987
A1LK2	0,8999	0,2834	3,175	0,001498 **
A1LK3	1,0008	0,3021	3,313	0,000922 ***
A1LK4	0,842	0,2903	2,901	0,003721 **
A1LK5	0,2708	0,2965	0,913	0,361125
A1LK6	-0,1428	0,3701	-0,386	0,699664
A1LK8	0,528	0,6976	0,757	0,449084
A1LK9	-0,3079	0,3367	-0,915	0,360403

Taulukko 9. Toimialojen Q ja R tulokset

Q	Estimaatti	Keskihajonta	z-arvo	p-arvo
(Intercept)	0,9579	0,1283	7,467	8,23E-14 ***
A1LK1	0,1592	0,2085	0,764	0,44515
A1LK2	-0,316	0,1372	-2,303	0,02128 *
A1LK3	-0,236	0,1302	-1,812	0,06997 .
A1LK4	-0,1718	0,1352	-1,27	0,20391
A1LK5	-0,2912	0,1295	-2,248	0,02455 *
A1LK6	0,3106	0,263	1,181	0,2376
A1LK8	-0,4061	0,1957	-2,075	0,03799 *
A1LK9	-0,3005	0,1394	-2,155	0,03116 *
zero-inflation	Estimaatti	Keskihajonta	z-arvo	p-arvo
(Intercept)	0,04436	0,27902	0,159	0,87369
A1LK1	0,8844	0,42496	2,081	0,03742 *
A1LK2	0,7805	0,28925	2,698	0,00697 **
A1LK3	0,58787	0,28154	2,088	0,03679 *
A1LK4	0,70078	0,28827	2,431	0,01506 *
A1LK5	0,30729	0,28073	1,095	0,2737
A1LK6	-0,29577	0,6905	-0,428	0,6684
A1LK8	0,48529	0,36817	1,318	0,18746
A1LK9	0,17331	0,29642	0,585	0,55877
R	Estimaatti	Keskihajonta	z-arvo	p-arvo
(Intercept)	0,67064	0,10953	6,123	9,19E-10 ***
A1LK1	-0,07556	0,36191	-0,209	0,835
A1LK2	0,20465	0,12547	1,631	0,103
A1LK3	0,15419	0,12068	1,278	0,201
A1LK4	0,07396	0,13818	0,535	0,593
A1LK5	0,11777	0,1296	0,909	0,364
A1LK6	0,09361	0,15251	0,614	0,539
A1LK8	0,03518	0,2154	0,163	0,87
A1LK9	-0,09274	0,1582	-0,586	0,558
zero-inflation	Estimaatti	Keskihajonta	z-arvo	p-arvo
(Intercept)	0,4941	0,1887	2,618	0,00885 **
A1LK1	0,4714	0,5553	0,849	0,3959
A1LK2	0,4359	0,2166	2,013	0,04413 *
A1LK3	0,1846	0,209	0,883	0,37714
A1LK4	0,4606	0,2329	1,977	0,048 *
A1LK5	0,3504	0,222	1,579	0,11442
A1LK6	-0,435	0,2852	-1,525	0,12717
A1LK8	0,1695	0,3647	0,465	0,64208
A1LK9	-0,2075	0,2757	-0,752	0,45178

Taulukko 10. Toimialan S tulokset

	Estimaatti	Keskihajonta	z-arvo	p-arvo
(Intercept)	5,80E-01	9,04E-02	6,417	1,39E-10 ***
A1LK1	2,18E-02	2,23E-01	0,098	0,92204
A1LK2	2,89E-01	1,25E-01	2,308	0,021 *
A1LK3	2,39E-01	1,02E-01	2,348	0,01885 *
A1LK4	2,97E-01	1,03E-01	2,898	0,00375 **
A1LK5	2,48E-01	9,86E-02	2,511	0,01203 *
A1LK6	1,21E-01	1,24E-01	0,973	0,33069
A1LK8	-1,54E-04	1,26E-01	-0,001	0,99902
A1LK9	-4,42E-02	1,17E-01	-0,377	0,70614
zero-inflation	Estimaatti	Keskihajonta	z-arvo	p-arvo
0 (Intercept)	0,29841	0,1588	1,879	0,06022 .
A1LK1	0,01547	0,39282	0,039	0,96859
A1LK2	0,70954	0,217	3,27	0,00108 **
A1LK3	0,41503	0,17871	2,322	0,02021 *
A1LK4	0,57332	0,17998	3,186	0,00144 **
A1LK5	0,04954	0,17558	0,282	0,77785
A1LK6	0,10678	0,21966	0,486	0,62687
A1LK8	-0,42444	0,23694	-1,791	0,07323 .
A1LK9	-0,51637	0,22167	-2,329	0,01984 *

Liite 2: R-koodi ja tulosteet

```
> zip2testiu <- zeroinfl(EUSEUR ~ A2LK,  
data = uusidata, dist = "poisson")  
> summary(zip2testiu)
```

Call:

```
zeroinfl(formula = EUSEUR ~ A2LK, data = uusidata, dist = "poisson")
```

Pearson residuals:

	Min	1Q	Median	3Q	Max
	-0.9232	-0.7502	-0.6566	-0.2093	26.8720

Count model coefficients (poisson with log link):

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	3.452003	0.001721	2006.116	< 2e-16	***
A2LK11	0.189734	0.012720	14.917	< 2e-16	***
A2LK12	-0.469680	0.033602	-13.978	< 2e-16	***
A2LK13	0.188499	0.027052	6.968	3.22e-12	***
A2LK14	0.217966	0.026658	8.176	2.93e-16	***
A2LK21	-0.002971	0.007297	-0.407	0.6839	
A2LK22	-0.014172	0.009944	-1.425	0.1541	
A2LK23	-0.026339	0.004752	-5.543	2.97e-08	***
A2LK24	-0.016561	0.008699	-1.904	0.0569	.
A2LK25	-0.086309	0.013488	-6.399	1.56e-10	***
A2LK26	-0.050907	0.005315	-9.577	< 2e-16	***
A2LK31	-0.005928	0.005924	-1.001	0.3170	
A2LK32	-0.133996	0.003176	-42.186	< 2e-16	***
A2LK33	0.010692	0.006005	1.780	0.0750	.
A2LK34	-0.113233	0.005573	-20.320	< 2e-16	***
A2LK35	-0.115836	0.016324	-7.096	1.28e-12	***
A2LK41	0.037306	0.003483	10.712	< 2e-16	***
A2LK42	-0.104581	0.006487	-16.120	< 2e-16	***
A2LK43	-0.061764	0.009854	-6.268	3.66e-10	***
A2LK44	-0.189090	0.004884	-38.715	< 2e-16	***
A2LK51	-0.166161	0.003261	-50.953	< 2e-16	***
A2LK52	-0.274648	0.003599	-76.320	< 2e-16	***
A2LK53	-0.195966	0.002803	-69.915	< 2e-16	***
A2LK54	-0.032061	0.005020	-6.387	1.69e-10	***
A2LK61	-0.021699	0.004459	-4.866	1.14e-06	***
A2LK62	-0.052586	0.011055	-4.757	1.97e-06	***
A2LK72	-0.128123	0.002810	-45.597	< 2e-16	***
A2LK73	-0.121875	0.011858	-10.278	< 2e-16	***

A2LK74	-0.037343	0.004464	-8.365	< 2e-16	***
A2LK75	-0.197233	0.006137	-32.138	< 2e-16	***
A2LK81	-0.129408	0.003169	-40.841	< 2e-16	***
A2LK82	-0.161165	0.005726	-28.146	< 2e-16	***
A2LK83	0.121729	0.002655	45.854	< 2e-16	***
A2LK91	-0.036975	0.003638	-10.164	< 2e-16	***
A2LK92	0.281326	0.023376	12.035	< 2e-16	***
A2LK93	-0.185635	0.003407	-54.493	< 2e-16	***
A2LK94	-0.390409	0.004861	-80.315	< 2e-16	***
A2LK95	-0.374210	0.034858	-10.735	< 2e-16	***
A2LK96	-0.253816	0.007353	-34.518	< 2e-16	***

Zero-inflation **model coefficients** (**binomial** with **logit link**):

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	0.330513	0.012673	26.079	< 2e-16	***
A2LK11	0.726835	0.091249	7.965	1.65e-15	***
A2LK12	0.912697	0.169686	5.379	7.50e-08	***
A2LK13	0.691217	0.194781	3.549	0.000387	***
A2LK14	0.540412	0.198917	2.717	0.006592	**
A2LK21	0.835901	0.047284	17.678	< 2e-16	***
A2LK22	0.899409	0.063390	14.188	< 2e-16	***
A2LK23	0.799381	0.030961	25.819	< 2e-16	***
A2LK24	1.043288	0.054675	19.082	< 2e-16	***
A2LK25	1.103923	0.081098	13.612	< 2e-16	***
A2LK26	0.446643	0.035611	12.542	< 2e-16	***
A2LK31	0.581327	0.039670	14.654	< 2e-16	***
A2LK32	0.239342	0.021649	11.056	< 2e-16	***
A2LK33	0.603558	0.040392	14.943	< 2e-16	***
A2LK34	0.506889	0.035987	14.085	< 2e-16	***
A2LK35	0.908338	0.098558	9.216	< 2e-16	***
A2LK41	0.850704	0.023520	36.169	< 2e-16	***
A2LK42	0.461068	0.042154	10.938	< 2e-16	***
A2LK43	0.490901	0.064671	7.591	3.18e-14	***
A2LK44	-0.050913	0.033454	-1.522	0.128038	
A2LK51	0.145117	0.022209	6.534	6.39e-11	***
A2LK52	0.429807	0.022632	18.991	< 2e-16	***
A2LK53	0.184593	0.019066	9.682	< 2e-16	***
A2LK54	0.278647	0.034787	8.010	1.15e-15	***
A2LK61	-0.167810	0.033580	-4.997	5.81e-07	***
A2LK62	-0.051821	0.080217	-0.646	0.518269	
A2LK72	0.304667	0.019242	15.834	< 2e-16	***
A2LK73	0.231281	0.078738	2.937	0.003310	**
A2LK74	0.387606	0.030467	12.722	< 2e-16	***
A2LK75	-0.178924	0.042810	-4.180	2.92e-05	***
A2LK81	-0.004557	0.022330	-0.204	0.838278	

A2LK82	0.267721	0.037449	7.149	8.74e-13	***
A2LK83	-0.222616	0.020907	-10.648	< 2e-16	***
A2LK91	-0.071034	0.026724	-2.658	0.007860	**
A2LK92	-0.037067	0.199594	-0.186	0.852672	
A2LK93	-0.004322	0.023468	-0.184	0.853880	
A2LK94	0.235198	0.029204	8.054	8.03e-16	***
A2LK95	-0.097068	0.217506	-0.446	0.655397	
A2LK96	-0.180829	0.049898	-3.624	0.000290	***

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

Number of iterations in BFGS optimization: 90

Log-likelihood: -2.202e+06 on 78 Df

```
> zap2testiu <- hurdle(EUSEUR ~ A2LK, data = uusidata, dist = "poisson")
> summary(zap2testiu)
```

Call:

`hurdle(formula = EUSEUR ~ A2LK, data = uusidata, dist = "poisson")`

Pearson residuals:

	Min	1Q	Median	3Q	Max
	-0.9232	-0.7502	-0.6566	-0.2093	26.8719

Count model coefficients (truncated poisson with log link):

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	3.452003	0.001721	2006.116	< 2e-16	***
A2LK11	0.189718	0.012720	14.915	< 2e-16	***
A2LK12	-0.469674	0.033602	-13.978	< 2e-16	***
A2LK13	0.188474	0.027053	6.967	3.24e-12	***
A2LK14	0.217942	0.026659	8.175	2.95e-16	***
A2LK21	-0.002970	0.007297	-0.407	0.6840	
A2LK22	-0.014165	0.009944	-1.425	0.1543	
A2LK23	-0.026338	0.004752	-5.543	2.98e-08	***
A2LK24	-0.016558	0.008699	-1.903	0.0570	.
A2LK25	-0.086295	0.013487	-6.398	1.57e-10	***
A2LK26	-0.050905	0.005315	-9.577	< 2e-16	***
A2LK31	-0.005930	0.005924	-1.001	0.3168	
A2LK32	-0.133995	0.003176	-42.186	< 2e-16	***
A2LK33	0.010694	0.006005	1.781	0.0749	.
A2LK34	-0.113226	0.005573	-20.319	< 2e-16	***
A2LK35	-0.115836	0.016324	-7.096	1.28e-12	***
A2LK41	0.037307	0.003483	10.712	< 2e-16	***
A2LK42	-0.104580	0.006487	-16.120	< 2e-16	***

A2LK43	-0.061751	0.009854	-6.267	3.69e-10	***
A2LK44	-0.189088	0.004884	-38.715	< 2e-16	***
A2LK51	-0.166160	0.003261	-50.953	< 2e-16	***
A2LK52	-0.274648	0.003599	-76.319	< 2e-16	***
A2LK53	-0.195964	0.002803	-69.915	< 2e-16	***
A2LK54	-0.032060	0.005020	-6.386	1.70e-10	***
A2LK61	-0.021698	0.004459	-4.866	1.14e-06	***
A2LK62	-0.052579	0.011055	-4.756	1.97e-06	***
A2LK72	-0.128122	0.002810	-45.597	< 2e-16	***
A2LK73	-0.121856	0.011858	-10.277	< 2e-16	***
A2LK74	-0.037342	0.004464	-8.365	< 2e-16	***
A2LK75	-0.197233	0.006137	-32.138	< 2e-16	***
A2LK81	-0.129409	0.003169	-40.842	< 2e-16	***
A2LK82	-0.161170	0.005726	-28.147	< 2e-16	***
A2LK83	0.121731	0.002655	45.854	< 2e-16	***
A2LK91	-0.036975	0.003638	-10.164	< 2e-16	***
A2LK92	0.281320	0.023376	12.034	< 2e-16	***
A2LK93	-0.185635	0.003407	-54.493	< 2e-16	***
A2LK94	-0.390407	0.004861	-80.315	< 2e-16	***
A2LK95	-0.374206	0.034858	-10.735	< 2e-16	***
A2LK96	-0.253815	0.007353	-34.517	< 2e-16	***
Zero hurdle model coefficients (binomial with logit link):					
	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-0.330513	0.012673	-26.079	< 2e-16	***
A2LK11	-0.726856	0.091250	-7.966	1.64e-15	***
A2LK12	-0.912680	0.169685	-5.379	7.50e-08	***
A2LK13	-0.691138	0.194778	-3.548	0.000388	***
A2LK14	-0.540315	0.198913	-2.716	0.006601	**
A2LK21	-0.835896	0.047284	-17.678	< 2e-16	***
A2LK22	-0.899389	0.063390	-14.188	< 2e-16	***
A2LK23	-0.799370	0.030961	-25.818	< 2e-16	***
A2LK24	-1.043288	0.054675	-19.082	< 2e-16	***
A2LK25	-1.103831	0.081096	-13.611	< 2e-16	***
A2LK26	-0.446637	0.035611	-12.542	< 2e-16	***
A2LK31	-0.581332	0.039670	-14.654	< 2e-16	***
A2LK32	-0.239338	0.021649	-11.055	< 2e-16	***
A2LK33	-0.603570	0.040392	-14.943	< 2e-16	***
A2LK34	-0.506895	0.035987	-14.085	< 2e-16	***
A2LK35	-0.908398	0.098559	-9.217	< 2e-16	***
A2LK41	-0.850703	0.023520	-36.169	< 2e-16	***
A2LK42	-0.461074	0.042154	-10.938	< 2e-16	***
A2LK43	-0.490914	0.064672	-7.591	3.18e-14	***
A2LK44	0.050915	0.033454	1.522	0.128024	
A2LK51	-0.145116	0.022209	-6.534	6.40e-11	***
A2LK52	-0.429804	0.022632	-18.991	< 2e-16	***

A2LK53	-0.184596	0.019066	-9.682	< 2e-16	***
A2LK54	-0.278646	0.034787	-8.010	1.15e-15	***
A2LK61	0.167811	0.033580	4.997	5.81e-07	***
A2LK62	0.051800	0.080217	0.646	0.518443	
A2LK72	-0.304669	0.019242	-15.834	< 2e-16	***
A2LK73	-0.231298	0.078738	-2.938	0.003308	**
A2LK74	-0.387603	0.030467	-12.722	< 2e-16	***
A2LK75	0.178935	0.042810	4.180	2.92e-05	***
A2LK81	0.004557	0.022330	0.204	0.838308	
A2LK82	-0.267724	0.037449	-7.149	8.74e-13	***
A2LK83	0.222621	0.020907	10.648	< 2e-16	***
A2LK91	0.071038	0.026724	2.658	0.007857	**
A2LK92	0.037165	0.199592	0.186	0.852282	
A2LK93	0.004323	0.023468	0.184	0.853843	
A2LK94	-0.235202	0.029204	-8.054	8.03e-16	***
A2LK95	0.096898	0.217508	0.445	0.655963	
A2LK96	0.180842	0.049898	3.624	0.000290	***

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

Number of iterations in BFGS optimization: 50

Log-likelihood: -2.202e+06 on 78 Df

```
> zinb2testiu <- zeroinfl(EUSEUR ~ A2LK, data = uusidata, dist = "negbin")
> summary(zinb2testiu)
```

Call:

```
zeroinfl(formula = EUSEUR ~ A2LK, data = uusidata, dist = "negbin")
```

Pearson residuals:

	Min	1Q	Median	3Q	Max
	-0.4884	-0.4283	-0.3840	-0.1144	15.7939

Count model coefficients (negbin with log link):

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	3.349096	0.012474	268.481	< 2e-16	***
A2LK11	0.201893	0.100563	2.008	0.04468	*
A2LK12	-0.507617	0.191616	-2.649	0.00807	**
A2LK13	0.200537	0.214004	0.937	0.34872	
A2LK14	0.231298	0.213967	1.081	0.27970	
A2LK21	-0.003220	0.052466	-0.061	0.95107	
A2LK22	-0.015051	0.071117	-0.212	0.83239	
A2LK23	-0.028094	0.033828	-0.830	0.40627	
A2LK24	-0.017682	0.062143	-0.285	0.77599	

A2LK25	-0.092346	0.093088	-0.992	0.32118	
A2LK26	-0.054416	0.037416	-1.454	0.14584	
A2LK31	-0.006355	0.042538	-0.149	0.88125	
A2LK32	-0.143486	0.021836	-6.571	5.00e-11	***
A2LK33	0.011485	0.043455	0.264	0.79155	
A2LK34	-0.121237	0.038135	-3.179	0.00148	**
A2LK35	-0.124100	0.111003	-1.118	0.26357	
A2LK41	0.039860	0.025437	1.567	0.11711	
A2LK42	-0.112010	0.044506	-2.517	0.01184	*
A2LK43	-0.066056	0.068863	-0.959	0.33744	
A2LK44	-0.202823	0.032410	-6.258	3.90e-10	***
A2LK51	-0.178137	0.022148	-8.043	8.77e-16	***
A2LK52	-0.295307	0.023394	-12.623	< 2e-16	***
A2LK53	-0.210219	0.019033	-11.045	< 2e-16	***
A2LK54	-0.034291	0.035640	-0.962	0.33597	
A2LK61	-0.023143	0.031816	-0.727	0.46698	
A2LK62	-0.056188	0.077593	-0.724	0.46899	
A2LK72	-0.137079	0.019460	-7.044	1.86e-12	***
A2LK73	-0.130383	0.080454	-1.621	0.10511	
A2LK74	-0.039922	0.031641	-1.262	0.20705	
A2LK75	-0.211609	0.040387	-5.240	1.61e-07	***
A2LK81	-0.138672	0.021818	-6.356	2.07e-10	***
A2LK82	-0.172759	0.038340	-4.506	6.61e-06	***
A2LK83	0.129641	0.019821	6.541	6.13e-11	***
A2LK91	-0.039510	0.025824	-1.530	0.12603	
A2LK92	0.298632	0.193660	1.542	0.12306	
A2LK93	-0.199096	0.022931	-8.682	< 2e-16	***
A2LK94	-0.421129	0.029656	-14.200	< 2e-16	***
A2LK95	-0.403422	0.208374	-1.936	0.05286	.
A2LK96	-0.272802	0.047011	-5.803	6.52e-09	***
Log(theta)	-0.512786	0.007167	-71.548	< 2e-16	***

Zero-inflation **model coefficients** (binomial with logit link):

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	0.146529	0.014081	10.406	< 2e-16	***
A2LK11	0.786346	0.094792	8.296	< 2e-16	***
A2LK12	0.910985	0.178899	5.092	3.54e-07	***
A2LK13	0.749391	0.202442	3.702	0.000214	***
A2LK14	0.595182	0.207764	2.865	0.004174	**
A2LK21	0.882424	0.049336	17.886	< 2e-16	***
A2LK22	0.946930	0.065912	14.367	< 2e-16	***
A2LK23	0.842363	0.032518	25.904	< 2e-16	***
A2LK24	1.095068	0.056654	19.329	< 2e-16	***
A2LK25	1.151063	0.083887	13.722	< 2e-16	***
A2LK26	0.471185	0.037886	12.437	< 2e-16	***

A2LK31	0.617297	0.041846	14.752	< 2e-16	***
A2LK32	0.241203	0.023444	10.289	< 2e-16	***
A2LK33	0.642194	0.042540	15.096	< 2e-16	***
A2LK34	0.527852	0.038265	13.795	< 2e-16	***
A2LK35	0.946709	0.102629	9.225	< 2e-16	***
A2LK41	0.901348	0.024787	36.363	< 2e-16	***
A2LK42	0.480495	0.044865	10.710	< 2e-16	***
A2LK43	0.516619	0.068494	7.543	4.61e-14	***
A2LK44	-0.081812	0.037085	-2.206	0.027381	*
A2LK51	0.135547	0.024199	5.601	2.13e-08	***
A2LK52	0.426818	0.024410	17.485	< 2e-16	***
A2LK53	0.174222	0.020762	8.391	< 2e-16	***
A2LK54	0.295386	0.037313	7.916	2.44e-15	***
A2LK61	-0.186392	0.037185	-5.013	5.37e-07	***
A2LK62	-0.063315	0.088204	-0.718	0.472869	
A2LK72	0.311822	0.020800	14.992	< 2e-16	***
A2LK73	0.233971	0.084959	2.754	0.005888	**
A2LK74	0.410353	0.032532	12.614	< 2e-16	***
A2LK75	-0.225396	0.048183	-4.678	2.90e-06	***
A2LK81	-0.022211	0.024512	-0.906	0.364883	
A2LK82	0.268182	0.040434	6.633	3.30e-11	***
A2LK83	-0.227331	0.022980	-9.892	< 2e-16	***
A2LK91	-0.082166	0.029353	-2.799	0.005122	**
A2LK92	-0.007581	0.214704	-0.035	0.971835	
A2LK93	-0.030001	0.025849	-1.161	0.245798	
A2LK94	0.201005	0.032024	6.277	3.46e-10	***
A2LK95	-0.163668	0.247268	-0.662	0.508033	
A2LK96	-0.236866	0.056511	-4.192	2.77e-05	***

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

Theta = 0.5988

Number of iterations in BFGS optimization: 100

Log-likelihood: -5.195e+05 on 79 Df

```
> zanb2testiu <- hurdle(EUSEUR ~ A2LK, data = uusidata, dist = "negbin")
> summary(zanb2testiu)
```

Call:

```
hurdle(formula = EUSEUR ~ A2LK, data = uusidata, dist = "negbin")
```

Pearson residuals:

Min	1Q	Median	3Q	Max
-----	----	--------	----	-----

-0.4884 -0.4283 -0.3840 -0.1144 15.7946

Count model coefficients (truncated negbin with log link):

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	3.349106	0.012474	268.480	< 2e-16	***
A2LK11	0.201707	0.100555	2.006	0.04486	*
A2LK12	-0.507738	0.191606	-2.650	0.00805	**
A2LK13	0.200341	0.213985	0.936	0.34915	
A2LK14	0.231625	0.214003	1.082	0.27910	
A2LK21	-0.003105	0.052469	-0.059	0.95281	
A2LK22	-0.014957	0.071121	-0.210	0.83342	
A2LK23	-0.028105	0.033829	-0.831	0.40608	
A2LK24	-0.017756	0.062141	-0.286	0.77508	
A2LK25	-0.092150	0.093098	-0.990	0.32226	
A2LK26	-0.054445	0.037415	-1.455	0.14562	
A2LK31	-0.006367	0.042538	-0.150	0.88102	
A2LK32	-0.143545	0.021836	-6.574	4.91e-11	***
A2LK33	0.011462	0.043455	0.264	0.79196	
A2LK34	-0.121286	0.038134	-3.181	0.00147	**
A2LK35	-0.124114	0.111003	-1.118	0.26352	
A2LK41	0.039810	0.025436	1.565	0.11757	
A2LK42	-0.111910	0.044508	-2.514	0.01192	*
A2LK43	-0.066055	0.068864	-0.959	0.33745	
A2LK44	-0.202816	0.032410	-6.258	3.90e-10	***
A2LK51	-0.178183	0.022148	-8.045	8.62e-16	***
A2LK52	-0.295269	0.023394	-12.621	< 2e-16	***
A2LK53	-0.210225	0.019033	-11.045	< 2e-16	***
A2LK54	-0.034280	0.035641	-0.962	0.33614	
A2LK61	-0.023142	0.031817	-0.727	0.46701	
A2LK62	-0.056176	0.077594	-0.724	0.46908	
A2LK72	-0.137225	0.019459	-7.052	1.76e-12	***
A2LK73	-0.130453	0.080452	-1.622	0.10491	
A2LK74	-0.039881	0.031642	-1.260	0.20753	
A2LK75	-0.211643	0.040387	-5.240	1.60e-07	***
A2LK81	-0.138578	0.021819	-6.351	2.13e-10	***
A2LK82	-0.172685	0.038342	-4.504	6.67e-06	***
A2LK83	0.129660	0.019822	6.541	6.10e-11	***
A2LK91	-0.039524	0.025824	-1.530	0.12589	
A2LK92	0.298772	0.193675	1.543	0.12292	
A2LK93	-0.199177	0.022931	-8.686	< 2e-16	***
A2LK94	-0.421044	0.029657	-14.197	< 2e-16	***
A2LK95	-0.403557	0.208361	-1.937	0.05277	.
A2LK96	-0.272786	0.047012	-5.802	6.53e-09	***
Log(theta)	-0.512792	0.007167	-71.549	< 2e-16	***

Zero hurdle model coefficients (binomial with logit link):

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-0.330513	0.012673	-26.079	< 2e-16	***
A2LK11	-0.726856	0.091250	-7.966	1.64e-15	***
A2LK12	-0.912680	0.169685	-5.379	7.50e-08	***
A2LK13	-0.691138	0.194778	-3.548	0.000388	***
A2LK14	-0.540315	0.198913	-2.716	0.006601	**
A2LK21	-0.835896	0.047284	-17.678	< 2e-16	***
A2LK22	-0.899389	0.063390	-14.188	< 2e-16	***
A2LK23	-0.799370	0.030961	-25.818	< 2e-16	***
A2LK24	-1.043288	0.054675	-19.082	< 2e-16	***
A2LK25	-1.103831	0.081096	-13.611	< 2e-16	***
A2LK26	-0.446637	0.035611	-12.542	< 2e-16	***
A2LK31	-0.581332	0.039670	-14.654	< 2e-16	***
A2LK32	-0.239338	0.021649	-11.055	< 2e-16	***
A2LK33	-0.603570	0.040392	-14.943	< 2e-16	***
A2LK34	-0.506895	0.035987	-14.085	< 2e-16	***
A2LK35	-0.908398	0.098559	-9.217	< 2e-16	***
A2LK41	-0.850703	0.023520	-36.169	< 2e-16	***
A2LK42	-0.461074	0.042154	-10.938	< 2e-16	***
A2LK43	-0.490914	0.064672	-7.591	3.18e-14	***
A2LK44	0.050915	0.033454	1.522	0.128024	
A2LK51	-0.145116	0.022209	-6.534	6.40e-11	***
A2LK52	-0.429804	0.022632	-18.991	< 2e-16	***
A2LK53	-0.184596	0.019066	-9.682	< 2e-16	***
A2LK54	-0.278646	0.034787	-8.010	1.15e-15	***
A2LK61	0.167811	0.033580	4.997	5.81e-07	***
A2LK62	0.051800	0.080217	0.646	0.518443	
A2LK72	-0.304669	0.019242	-15.834	< 2e-16	***
A2LK73	-0.231298	0.078738	-2.938	0.003308	**
A2LK74	-0.387603	0.030467	-12.722	< 2e-16	***
A2LK75	0.178935	0.042810	4.180	2.92e-05	***
A2LK81	0.004557	0.022330	0.204	0.838308	
A2LK82	-0.267724	0.037449	-7.149	8.74e-13	***
A2LK83	0.222621	0.020907	10.648	< 2e-16	***
A2LK91	0.071038	0.026724	2.658	0.007857	**
A2LK92	0.037165	0.199592	0.186	0.852282	
A2LK93	0.004323	0.023468	0.184	0.853843	
A2LK94	-0.235202	0.029204	-8.054	8.03e-16	***
A2LK95	0.096898	0.217508	0.445	0.655963	
A2LK96	0.180842	0.049898	3.624	0.000290	***

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

Theta: count = 0.5988

Number of iterations in BFGS optimization: 71

Log-likelihood: $-5.195e+05$ on 79 Df