

Työn keston ja sijainnin vaikutus julkisen hallinnon ja maanpuolustuksen avoimien työpaikkojen täyttymiseen vuonna 2018

Tarja Sundqvist
Datatieteen pro gradu -tutkielma
1461635
Matemaattisten tieteiden tutkimusyksikkö
Oulun yliopisto
Kevät 2021

Sisältö

Tiivistelmä	2
1 Johdanto	3
2 Aineisto	4
2.1 Rajaukset	5
2.2 Vastemuuttuja	7
2.3 Seuruaika	8
2.4 Selittävät muuttujat	10
2.5 Aineiston soveltuvuus elinaika-analyysiin	11
3 Elinaikojen funktiot	12
3.1 Kumulatiivinen ilmaantuvuusfunktio	12
3.2 Välttöfunktio	12
3.3 Uhkafunktio	13
3.4 Kumulatiivinen uhkafunktio	14
3.5 Parametriton estimointi	14
4 Coxin regressiomalli	16
4.1 Mallin määrittely	16
4.2 Välttöfunktio ja kumulatiivinen uhkafunktio	18
4.3 β -kertoimien estimointi ja tulkinta	18
4.4 Mallidiagnostiikka	21
5 Coxin regressiomallin sovitukset	23
5.1 Aineiston kuvaileva tarkastelu	23
5.2 Kaikki avoimet työpaikat	27
5.3 Vuorotteluvapaa- ja palkkatyöt	29
6 Yhteenveto	35
7 Pohdinta	37
LÄHTEET	38
Liite: Rekisteriaineiston muuttujat	40

Tiivistelmä

Tutkielmassa analysoidaan työn keston ja sijainnin vaikutusta julkisen hallinnon ja maanpuolustuksen avointen työpaikkojen täyttymiseen vuonna 2018. Analyysin kohteena ovat ne julkisen hallinnon ja maanpuolustuksen avoimet työpaikat, joiden avoinnaolon kesto määräytyy vapaasti työmarkkinoilla eli ns. ei-hakupaikat. Työpaikkojen avoinnaolon kesto käytetään tehokkuuden mittarina, joka kuvaa sitä, miten hyvin potentiaalista työllisyyttä eli avoimia työpaikkoja saadaan otettua käyttöön.

Tutkielman aineistona on työ- ja elinkeinoministeriön keräämä rekisteriaineisto, joka sisältää tietoja julkiseen työnvälitykseen vuonna 2018 ilmoitetuista työpaikoista. Aineistossa on tiedot 381 121 havaintoyksiköstä 41 muuttujan osalta. Aineisto on koottu työpaikkailmoituksista, joista on laskettu kussakin ilmoituksessa avoinna olevien työpaikkojen määrä. Lisäksi mukana on muita tietoja työpaikoista, niitä tarjoavista työnantajista ja työpaikan täyttöprosessista, sekä erilaisia kesto-muuttujia, kuten avoinnaolon kesto. Tutkielmaa varten aineistoa rajattiin siten, että analyysin kohteena oli 3 109 julkisen hallinnon ja maanpuolustuksen avointa työpaikkaa.

Julkisen hallinnon ja maanpuolustuksen avointen työpaikkojen analyysissä sovitettiin työn kesto- ja kuntatyyppi-muuttujat sisältävä Coxin regressiomalli sekä koko aineistoon että työpaikan tyyppiin (palkka, vuorotteluvapaa) mukaan rajattuihin aineistoihin. Avointen työpaikkojen avoinnaolon kesto havaittiin vuorotteluvapaa- ja palkkatöissä niin erilaisiksi, että työpaikan tyyppi -tietoa käytettiin osittavana tekijänä koko aineistoon sovitetussa mallissa.

Mallioletusten paikkaansapitävyyttä arvioitiin pääasiassa skaalattuihin Schoenfeldin residuaaleihin perustuvan zph-testin avulla. Lisäksi työpaikkojen täyttöprosessin tehokkuutta arvioitiin Kaplanin-Meierin menetelmällä laskettujen välttöfunktion estimaattien avulla.

Analyysin tuloksena havaittiin, että työpaikan sijainnilla on vaikutusta avoimen työpaikan täyttöön erityisesti silloin, kun työpaikka sijaitsee maaseutumaisessa tai kaupunkimaisessa kunnassa. Taajaan asutun kunnan vaikutuksesta avoimen työpaikan täyttöön saatiin jonkin verran viitteitä, mutta näihin tuloksiin sisältyi epävarmuutta. Taajaan asutun kunnan vaikutus työpaikan täyttöön näyttäisikin olevan samankaltainen kuin kaupunkimaisen kunnan vaikutus.

Työn kestolla havaittiin olevan vaikutusta avoimen työpaikan täyttöön. Vaikutus kuitenkin vaihtelee sen mukaan, onko kyseessä palkka- vai vuorotteluvapaatyö. Analyysissä havaittiin myös, ettei Coxin regressiomallin oletus verrannollisista uhkasuhteista aina päde työn kesto -muuttujaan.

1 Johdanto

Työpaikkojen avoinnaolon kestoa voidaan käyttää tehokkuuden mittarina, joka kuvaa sitä, miten hyvin potentiaalista työllisyyttä eli avoimia työpaikkoja saadaan otettua käyttöön. (Räisänen, 2016). Pro gradu -tutkielmassani analysoin julkisen työnvälityksen tehokkuutta tutkimalla työn keston ja sijainnin vaikutusta julkisen hallinnon ja maanpuolustuksen avointen työpaikkojen täyttymiseen vuonna 2018. Analyysin kohteena ovat ne julkisen hallinnon ja maanpuolustuksen avoimet työpaikat, joiden avoinnaolon kesto määräytyy vapaasti työmarkkinoilla eli ns. eihakupaikat. Kyseinen toimiala on useana vuonna kuulunut julkisen työnvälityksen suurimpiin työnantajatoimialoihin.

Tutkielmani aineistona on työ- ja elinkeinoministeriön keräämä rekisteriaineisto, joka sisältää tietoja julkiseen työnvälitykseen vuonna 2018 ilmoitetuista työpaikoista. Aineisto on koottu työpaikkailmoituksista, joista on laskettu kussakin ilmoituksessa avoinna olevien työpaikkojen määrä. Aineisto sisältää tietoja avoinna olleista työpaikoista, niitä tarjoavista työnantajista ja työpaikan täyttöprosessista, ja siihen on lisätty erilaisia kestromuuttujia, kuten avoinnaolon kesto. Aineisto on saatavilla Tietoarkisto Ailasta tutkimus-, opetus- ja opiskelukäyttöön (Työ- ja elinkeinoministeriö, 2019).

Työ- ja elinkeinoministeriössä (TEM) on tutkittu samaa aihepiiriä, ja vuonna 2016 julkaistiin analyysi, jossa tutkittiin vuoden 2015 avointen työpaikkojen täyttöprosessia ja erityisesti työtarjousten vaikutusta työpaikan täyttöön (Räisänen, 2016). Tämä TEM-analyysi ovat tärkeä viitekehys omalle tutkielmalleni. Räisäsen (2016) TEM-analyysin kohteena ovat kaikki toimialat, mutta omassa työssäni olen rajannut aiheen huomattavasti suppeammaksi. Aineisto ja siihen tehdyt rajaukset on esitelty luvussa 2.

Aineiston analyysissä hyödynnän elinaika-analyysin menetelmiä. Analysoin työn keston ja sijainnin vaikutusta työpaikan avoinnaoloon Coxin regressiomallin avulla, joten eliaika-analyysin kohteena on yhden työpaikkailmoituksen avoinnaolon aika eli elinaika. Coxin regressiomallia on sovellettu myös Räisäsen (2016) TEM-analyysissä. Elinaika-analyysin tärkeimmät käsitteet on esitelty luvussa 3, ja Coxin regressiomalli luvussa 4. Mallin sovitukset ja sen tulokset on kerrottu luvussa 5, ja tulosten yhteenveto löytyy luvusta 6.

Elinaika-analyysia voidaan käyttää aikamuuttujia sisältävän datan analysointiin. Oman aineistoni soveltuvuutta elinaika-analyysin kohteeksi on arvioitu luvussa 2.5. Elinaika-analyysin historia ja analyysiin liittyvät käsitteet kytkeytyvät tiiviisti epidemiologian kehitykseen, ja käytänkin tutkielmassani epidemiologian suomenkielisiä käsitteistöä ja sanastoa (Läärä, 2018; Läärä ym., 2008).

2 Aineisto

Tässä tutkielmassa sovelletaan elinaika-analyysin käsitteitä ja menetelmiä julkiseen työnvälitykseen vuonna 2018 ilmoitettujen avointen työpaikkojen rekisteriaineistoon (Työ- ja elinkeinoministeriö, 2019). Elinaika-analyysi (*survival analysis*) tarkoittaa yleisellä tasolla sitä, että dataa analysoidaan ajan muodossa. Elinajan seuraaminen alkaa tarkoin määritellystä alkuhetkestä, ja jatkuu kiinnostavan vastetapahtuman eli kohdetapahtuman tai jonkin muun tapahtuman ilmaantumiseen asti tai kunnes jokin muu tietyt kriteetit täyttävä ajanhetki saavutetaan. Elinaika (*survival time*) viittaa näiden määriteltyjen alku- ja loppuhetkien väliseen aikaan, ja tätä ajanjaksoa kutsutaan myös seuruuajaksi. Yksittäisten havaintoyksiköiden sijaan elinaika-analyysissä keskitytään yleensä jonkin ryhmän tai ryhmien elinaikoihin ja niiden vertailuun. (Collett, 2015)

Elinaika-analyysin ja siinä käytettävien menetelmien historia kytkeytyy epidemiologian historiaan, jonka eräs tärkeä varhainen merkkipaalu on vuosi 1662. Tuolloin englantilainen John Graunt (24.4.1620-18.4.1674) julkaisi teoksensa *Natural and Political Observations Made Upon the Bills of Mortality*, jossa hän analysoi 50 vuoden ajalta lontoolaisten viikottaisia kuolintilastoja. Niiden pohjalta Graunt mm. kehitti ensimmäisen elinaikataulun, ja samoja periaatteita käytetään edelleen elinaikojen analyysissä. (Morabia, 2013)

Tutkielmassa käytetty aineisto on koottu työpaikkailmoituksista, joista on laskettu kussakin ilmoituksessa avoinna olevien työpaikkojen määrä. Aineiston on koostanut työ- ja elinkeinoministeriön työllisyys ja toimivat markkinat -osasto, ja se saatavilla tutkimus-, opetus- ja opiskelukäyttöön Tietoarkisto Ailassa. Kyseessä on kvantitatiivinen kokonaisaineisto, jonka kohdealueena on Suomi. Aineisto kuuluu *Julkiseen työnvälitykseen ilmoitettujen avointen työpaikkojen rekisteriaineisto*-sarjaan, jossa on julkaistu myös vastaavat vuosikohtaiset aineistot vuosilta 2008-2017. (Työ- ja elinkeinoministeriö, 2019.)

Aineiston havaintoyksikkönä on työpaikkailmoitus. Ilmoitukseen liittyvien tietojen lisäksi mukana on tietoja työnantajista ja työpaikan täyttöprosessista sekä erilaisia kestopuuttujia. Yhteensä aineisto sisältää tiedot 381 121 havaintoyksiköstä 41 muuttujan osalta. (Työ- ja elinkeinoministeriö, 2019.) Kaikki muuttujat ja niiden koodit on listattu tutkielman liitteessä. Muuttujiin viitataan jatkossa sekä niiden nimillä että koodeilla.

Seuraavissa alaluvuissa on kuvattu aineistoon tehtyjä rajauksia, sekä kerrottu, miten aineistosta on muodostettu vastemuuttuja ja seuruaika elinaika-analyysia varten. Lopuksi kuvataan, miksi aineisto soveltuu elinaika-analyysin kohteeksi.

2.1 Rajaukset

Tutkielman kohteeksi valittiin työpaikkailmoitukset, joissa työnantajan toimialana (muuttuja Q4) on *Julkinen hallinto ja maanpuolustus; pakollinen sosiaalivaikutus*. Valitun toimialan ilmoituksia oli yhteensä 48 320 kpl eli noin 13 % koko aineistosta.

Aineistossa toimialat on ilmoitettu koodilla, ja rajausta varten koodit muutettiin selväkielisiksi Tilastokeskuksen tarjoaman toimialaluokituksen (TOL 2008) avulla. Luokitusta käytetään taloudellista toimintaa kuvaavien tilastojen tiedonkeruu- ja kuvauskehikkona, ja se perustuu EU:n toimialaluokitukseen NACEen (*Nomenclature générale des Activités économiques dans les Communautés européennes*). EU-jäsenmaiden on käytettävä tilastotoimissa NACEa tai siihen perustuvaa kansallista versiota, jollainen TOL 2008 -toimialaluokitus on. (Toimialaluokitus, 2008)

Aineiston työpaikkailmoitukset jakaantuvat hakupaikkoihin ja ei-hakupaikkoihin (Q28). Hakupaikat ovat työpaikkailmoituksia, joihin työnantaja on yhdessä TE-toimiston kanssa määrännyt ennalta avoinnaolon keston. Ei-hakupaikoissa ilmoituksen avoinnaolon kesto sen sijaan määräytyy työmarkkinoilla. (Räisänen, 2016) Taulukossa 1 on esitetty hakupaikkojen ja ei-hakupaikkojen osuus julkisen hallinnon ja maanpuolustuksen työpaikkailmoituksissa:

	<i>n</i>	%
Ei hakupaikka	3364	7
Hakupaikka	44 956	93
Yhteensä	48 320	100

Taulukko 1: Hakupaikkojen ja ei-hakupaikkojen osuus julkisen hallinnon ja maanpuolustuksen työpaikkailmoituksissa vuonna 2018.

Räisänen (2016) toteaa, että myös työnantajan ja TE-toimiston määräämään avoinnaolon keston voi vaikuttaa eri tekijöitä, joten hakupaikkoja ja ei-hakupaikkoja voitaisiin tutkia samoilla menetelmillä. Tässä tutkielmassa on kuitenkin rajattu hakupaikat pois analysoitavasta datasta.

Aineistoon sisältyy täyttötieto-muuttuja (Q15), jonka arvot kuvaavat eri syitä ilmoituksen avoinnaolon päätymiselle. Muuttujasta rajattiin pois työpaikkoja lisätty -kategoria eli ne ilmoitukset, joihin on jälkikäteen lisätty työpaikkoja. Julkisen hallinnon ja maanpuolustuksen työpaikkailmoituksissa tähän kategoriaan kuului 370 ilmoitusta.

Datasta rajattiin pois ne ilmoitukset, joissa on mukana useampi kuin yksi työpaikkailmoitus (Q16). Näin elinajan analyysin kohteena ovat yksittäiset työpaikat. Taulukossa 2 on esitetty avointen työpaikkojen määrä julkisen hallinnon ja maanpuolustuksen työpaikkailmoituksissa vuonna 2018. Lähes 90 % ilmoituksista sisälsi vain yhden avoimen työpaikan.

	<i>n</i>	%
1 työpaikka	42 994	89
2-10 työpaikkaa	4 998	10
11-100 työpaikkaa	312	1
yli 100 työpaikkaa	16	0
Yhteensä	48 320	100

Taulukko 2: Avointen työpaikkojen määrä julkisen hallinnon ja maanpuolustuksen työpaikkailmoituksissa vuonna 2018.

Työpaikan tyyppi -muuttujan (Q10) arvoista otettiin mukaan vain palkka- ja vuorotteluvapaatyöt. Provisio- ja yritys-tyyppisiä töitä oli niin vähäinen määrä (yhteensä 14 kpl), että ne rajattiin analyysin ulkopuolelle.

Alueellisia rajauksia tehtiin sekä tilastollisen kuntatyyppin (Q37) että sijaintikunnan (Q2) perusteella. Jos tilastollisena kuntatyyppinä oli ”ulkomaat” (202 kpl) tai työpaikka sijaitti Ahvenanmaalla (563 kpl), ilmoitus rajattiin pois.

Datassa sijaintikunnat on ilmoitettu koodeilla. Kullekin kunnalle haettiin Tilastokeskuksen luokitteluavaimen (Luokitusavain, 2018) avulla tieto siitä, mihin NUTS-aluejaon (*Nomenclature des Unités Territoriales Statistiques*) mukaiseen suuralueeseen kunta kuuluu. NUTS on Euroopan unionin (EU) alueluokitusjärjestelmä, jota käytetään kaikissa EU:n yhteisissä alueellisissa tilastoissa. Suositus on, että tilastoissa käytetään pääasiallisena aluejakona virallista NUTS-alueluokitusta. Suuraluejako on NUTS 2 -tason luokitus, ja siinä Suomi on jaettu viiteen alueeseen: Ahvenanmaa, Etelä-Suomi, Helsinki-Uusimaa, Länsi-Suomi ja Pohjois- ja Itä-Suomi. (NUTS-aluejako, 2021) Näistä alueista Ahvenanmaa rajattiin pois.

Taulukossa 3 on esitetty julkisen hallinnon ja maanpuolustuksen avointen työpaikkojen sijainti suuralueen mukaan vuonna 2018.

Suuralue	<i>n</i>	%
Etelä-Suomi	7 669	16
Helsinki-Uusimaa	19 914	41
Länsi-Suomi	9 324	19
Pohjois- ja Itä-Suomi	10 648	22
Åland	563	1
Tieto puuttuu	202	0
Yhteensä	48 320	100

Taulukko 3: Julkisen hallinnon ja maanpuolustuksen avointen työpaikkojen sijainti suuralueen mukaan vuonna 2018.

Julkisen hallinnon ja maanpuolustuksen työpaikkailmoituksissa oli 202 ilmoitusta, josta puuttui sijaintikunnan tiedot eikä näin ollen myöskään tietoa suuralueesta voitu liittää ilmoituksen tietoihin. Tilastollisen kuntatyyppin mukaan näiden ilmoitusten avoin työpaikka sijaitsee ulkomailla, eli ilmoitukset rajautuivat pois muuttujan Q37 rajausten myötä.

Rajausten jälkeen aineistoon jäi 3 109 havaintoyksikköä eli työpaikkailmoitusta. Koska aineisto rajattiin siten, että mukaan otettiin vain ne ilmoitukset, joissa on yksi avoin työpaikka, havaintoyksikkönä voidaan pitää myös avointa työpaikkaa.

2.2 Vastemuuttuja

Tutkielman vastemuuttujana eli kiinnostavana tapahtumana on tieto siitä, onko työpaikkailmoituksessa avoinna ollut työpaikka täytetty vai ei (0 = Ei täytetty, 1 = Työpaikka täytetty). Aineistoon kuuluu luokitteluasteikollinen täyttötietomuuttuja (Q15), jonka luokat kuvaavat eri syitä työpaikan täyttymiselle sekä sille, että työpaikka on jäänyt täyttymättä. Tutkielman vastamuuttuja muodostettiin täyttötieto-muuttujasta siten, että kaikki muuttujan välittämä informaatio työpaikan täyttymisen tai täyttymättömyyden syistä on keskenään samanarvoista. Täyttymisen eri syitä ei siten pidetä kilpailevina tapahtumina.

Vastemuuttuja muodostettiin aineiston täyttötieto-muuttujan (Q15) luokkien perusteella seuraavasti:

Täyttötieto	<i>n</i>	%	Vaste
Täytetty työtarjouksesta	2104	68	1
Täytetty muutoin toimiston hakijalla	278	9	1
Täytty muutoin	323	10	1
Peruutettu	149	5	0
Ehdokkaita riittävästi	120	4	1
Työpaikkoja lisätty	0	0	-
Hakuaika päättynyt	0	0	-
Täytetty verkon kautta	73	2	1
Poistettu/peruutettu verkon kautta	4	0	0
Täyttämätön työpaikka	58	2	0
Yhteensä	3109	100	-

Taulukko 4: Vastemuuttujan muodostaminen täyttötieto-muuttujasta.

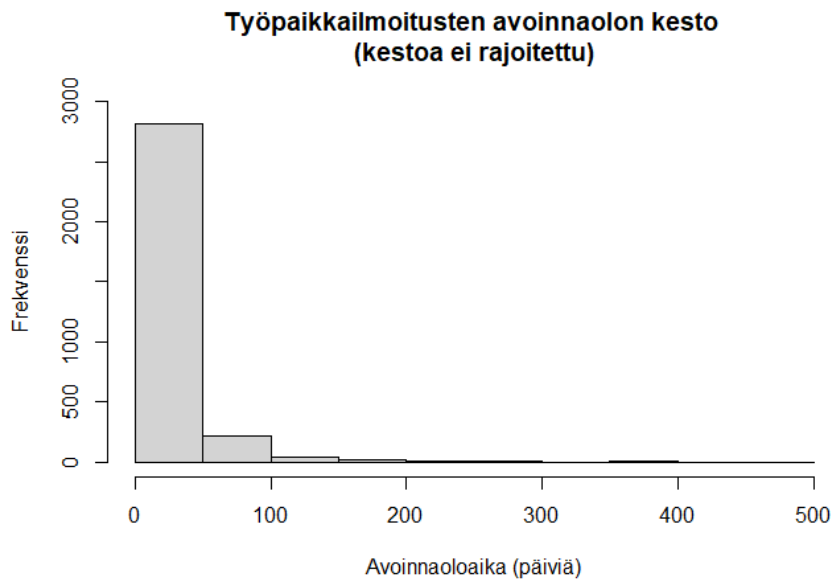
Täyttyneitä työpaikkoja on yhteensä 2 898 kpl eli 93 %, ja täyttymättömiä 211 kpl eli 7 %.

Ehdokkaita riittävästi -kategoria tarkoittaa, että työnantaja on saanut vaatimansa määrän ehdokkaita, eikä edellytä TE-toimistolta muita toimenpiteitä. Varsinainen täyttötieto näistä paikoista toimitetaan TE-toimistolle jälkikäteen. Räisänen (2016) on luokitellut nämä havainnot täyttyneeksi työpaikaksi, ja tässä tutkielmassa on toimittu samoin. Työhallinnon tilastoissa kyseisiä havaintoja ei kuitenkaan enää lueta täyttyneiksi paikoiksi, vaan ne lasketaan rekrytointiprosessi käynnissä -ryhmään. (Räisänen, 2016)

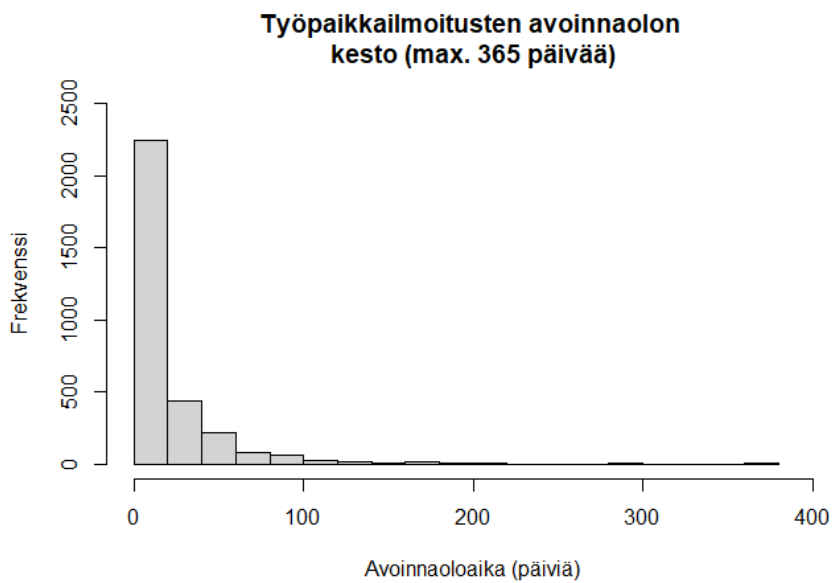
2.3 Seuruaika

Aineisto sisältää useita julkisen työvölityksen työpaikkailmoituksiin liittyviä kesto-muuttujia. Niiden avulla voidaan ilmaista erilaisia näkökulmia työpaikan täyttymiseen. Ilmoituksen avoinnaolon alku- ja loppuhetken välinen aika kuvaa työnhakijan näkökulmaa. Työnantajan näkökulmaan puolestaan sisältyy myös rekrytointiin kuuluva aika, joka päättyy työn aiottuun aloittamishetkeen. (Räisänen, 2016) Tässä tutkielmassa ei käsitellä rekrytointiin kuluva aikaa, vaan keskitytään työnhakijan näkökulmaan eli työpaikkailmoituksen avoinnaolon keston.

Kuvassa 1 on esitetty kahden avoinnaolon keston liittyvän muuttujan jakaumat julkisen hallinnon ja maanpuolustuksen työpaikkailmoituksissa:



(a) Muuttuja Q20: Avoinnaolon kesto



(b) Muuttuja Q21: Avoinnaolon kesto (maksimiksi laskettu 365 päivää)

Kuva 1: Julkisen hallinnon ja maanpuolustuksen työpaikkailmoitusten avoinnaolon kesto vuonna 2018.

Avoinnaolon kesto -muuttuja (Q20) kertoo, montako päivää työpaikkailmoitus on

ollut avoinna (kuva 1a). Osa ilmoituksista on ollut auki yli vuoden, ja muuttujassa Q21 (kuva 1b) näiden ilmoitusten aukioloajaksi on korjattu 365 päivää. Tutkielmassa käytetään seuruaikana muuttujaa Q21 eli seuruu kestää enintään 365 päivää.

Kuvan 1 histogrammeista havaitaan, että suurin osa työpaikkailmoituksista sulkeutuu 100 päivän kuluessa niiden avaamisesta. Kun seuruaika on enintään 365 päivää, avoinnaolon keston keskiarvo on 18 päivää, ja mediaani 7 päivää. Vuoden mittaisia seuruaikoja sisältyy aineistoon vain muutamia, mutta niillä on merkittävä vaikutus keskiarvoon.

2.4 Selittävät muuttujat

Tutkielmassa käytetään selittävinä muuttujina työpaikan tyyppiä (Q10), tilastollista kuntatyyppiä (Q37) sekä työn kestoa (Q7). Aineiston muuttujissa on tilastollisen kuntatyyppin lisäksi muitakin työpaikan sijaintiin liittyviä muuttujia: työpaikan sijaintikunta (Q2), seutukunta (Q36) ja suurkaupunki (Q38). Niitä ei ole huomioitu tutkielman analyysissä, sillä sijaintikunnan ja seutukunnan sijaan päätettiin käyttää karkeampaa sijaintiin liittyvää luokittelua eli tilastollista kuntatyyppiä. Suurkaupunki-muuttujasta puolestaan puuttui lähes 70 % muuttujan arvoista.

Aineistoon tehtyjen rajausten jälkeen työpaikan tyytit jakautuvat palkkatöihin ja vuorotteluvapaisiin. Vuorotteluvapaa kestää vähintään 100 päivää ja enintään 180 päivää, ja vapaalle jäävän tilalle on palkattavat työtön työnhakija. (Vuorotteluvapaalaki, 2002)

Tilastollinen kuntaryhmitys on virallinen tilastoinnissa käytetty alueluokitusstandardi, jossa luokittelukriteereinä ovat kunnan taajamaväestön osuus ja suurimman taajaman väkiluku.

Maaseutumaisissa kunnissa joko alle 60 % väestöstä asuu taajamissa ja suurimman taajaman väkiluku on alle 15 000 tai väestöstä vähintään 60 % mutta alle 90 % asuu taajamissa ja suurimman taajaman väkiluku on alle 4 000.

Taajaan asutuissa kunnissa vähintään 60 % mutta alle 90 % asuu taajamissa ja suurimman taajaman väkiluku on vähintään 4 000 mutta alle 15 000.

Kaupunkimaisissa kunnissa väestöstä vähintään 90 % asuu taajamissa tai suurimman taajaman väkiluku on vähintään 15 000. (Tilastollinen kuntaryhmitys, 2011)

2.5 Aineiston soveltuvuus elinaika-analyysiin

Elinaika-analyysissä dataa analysoidaan ajan muodossa. (Collett, 2015) Julkiseen työväilytykseen ilmoitettujen avointen työpaikkojen rekisteriaineistossa on valmiina elinaika-analyysin vaatimat aikamuuttujat, sillä työpaikkailmoitusten avoinnaololle ja rekrytoinnin eri vaiheille on ilmoitettu tarkat alkamis- ja päättymispäivät. Jos avoinnalon päättymispäivä ei ole ollut tiedossa rekisteriaineiston koontihetkellä, päättymispäiväksi on merkitty 11.11.2022. Näitä havaintoja ei sisältynyt julkisen hallinnon ja maanpuolustuksen työpaikkailmoituksiin.

Elinaikoihin liittyy tiettyjä erityispiirteitä, jotka on huomioitava analyysissä. Elinaikat eivät yleensä ole symmetrisesti jakautuneita, sillä homogeenisen ryhmän yksilöiden seuruaajoista koostettu histogrammi on useimmiten oikealle vino (Collett, 2015). Väestötasolla elinaikojen jakauma on kuitenkin yleensä vasemmalle vino (Robertson ja Allison, 2012). Normaalijakaumaoletus ei siis ole kovin usein realistinen elinaika-analyysissä. (Collett, 2015) Kuvasta 1 näkyy, että tutkielman aineiston elinaikojen eli työpaikkailmoitusten avoinnaoloaikojen jakauma on oikealle vino.

Havaintojen sensuroituminen on toinen tärkeä elinaika-analyysin piirre. Sensuroituminen viittaa tilanteeseen, jossa havaintoyksikölle ei voida havaita kohdetapahtumaa. Sensuroitumisessa on useita eri alalajeja: oikealta ja vasemmalta sensuroituminen sekä välisensuroituminen, ja nämä jakautuvat edelleen alaryhmiin. (Collett, 2015) Tutkielman aineistossa on mukana oikealta sensuroituneita havaintoja.

Oikealta sensuroituminen (*right censoring*) tarkoittaa, että sensuroituminen tapahtuu sen jälkeen, kun havaintoyksikkö on otettu mukaan tutkimukseen. Oikealta sensuroitunut elinaika on lyhyempi kuin todellinen, tuntemattomaksi jäävä elinaika. Sensuroitumisen syynä voi esimerkiksi olla se, ettei kohdetapahtuma ole sattunut seuruaajan päättymiseen mennessä. (Collett, 2015) Tutkielman aineistossa täyttötieto-muuttujan perusteella tiedetään, että osa ilmoituksista peruttiin ja osa avoimista työpaikoista jäi täyttymättä eli tapahtui oikealta sensuroitumista.

3 Elinaikojen funktiot

Elinaika-analyysiin liittyy muutamia funktioita, joiden avulla analyysin tuloksia esitetään tiivistetysti. Näitä ovat kumulatiivinen ilmaantuvuusfunktio (*cumulative incidence function*), välttöfunktio, jota voidaan kutsua myös elossaolofunktioksi (*survival function*), uhkafunktio (*hazard function*) sekä kumulatiivinen uhkafunktio (*cumulative hazard function*). (Collett, 2015)

Seuraavissa alaluvuissa kuvataan lyhyesti edellä luetellut funktiot ja niiden keskinäiset yhteydet. Kumulatiiviselle ilmaantumiskäyrälle ja välttöfunktiolle esitetään lisäksi parametrin estimointimenetelmä. Tutkielman vastemuuttuja (2.2) on dikotominen, joten elinaikojen funktioiden esittelyssä ei ole huomioitu kilpailevia tapahtumia.

Kaikissa esiteltävissä funktioissa satunnaismuuttuja T kuvaa aikaa, joka on kulunut ajanlaskun määritellyn alkuhetken ja kohdetapahtuman havaitsemisen välillä. T voi saada vain positiivisia arvoja. (Collett, 2015)

3.1 Kumulatiivinen ilmaantuvuusfunktio

Elinaikaa ilmaisevalle satunnaismuuttujalle T oletetaan yleensä jatkuva jakauma, jonka tiheysfunktiona on $f(t)$ ja kertymäfunktiona $F(t)$ (Collett, 2015):

$$F(t) = \mathbb{P}(T < t) = \int_0^t f(u)du. \quad (1)$$

Kertymäfunktio $F(t)$ kuvaa riskiä eli ilmaantumistodennäköisyyttä (*incidence probability*), että kiinnostava tapahtuma havaitaan hetkeen t mennessä eli riskijakson $], t]$ aikana eikä aineistossa ole kilpailevia tapahtumia. Tätä kertymäfunktiota kutsutaan myös kumulatiiviseksi ilmaantuvuusfunktioksi (*cumulative incidence function*). (Collett, 2015; Läärä, 2018)

Epidemiologiassa kumulatiivinen ilmaantuvuus (1) kuvaa riskiä. Riski on siten todennäköisyys, joka ilmaistaan ajan funktiona. (Cole et al., 2015)

3.2 Välttöfunktio

Välttöfunktio $S(t)$ kuvaa todennäköisyyttä, että kiinnostava tapahtuma havaitaan vasta hetken t jälkeen eli tapahtuma on vältetty hetkeen t asti (Collett, 2015):

$$S(t) = \mathbb{P}(T \geq t) = 1 - F(t). \quad (2)$$

Kertymäfunktio $F(t)$ (1) voidaan siis ilmaista myös välttöfunktion avulla eli välttöfunktio on ilmaantumistodennäköisyyden komplementti (Läärä, 2018):

$$F(t) = \mathbb{P}(T < t) = 1 - S(t). \quad (3)$$

Välttöfunktion yhteydessä oletetaan, että ajanhetkellä 0 pätee $S(0) = 1$, ja siten kertymäfunktio $F(0) = 0$ (Kalbfleisch, 1985).

3.3 Uhkafunktio

Uhkafunktio eli riskitiheysfunktio $h(t)$ kuvaa kohdetapahtuman hetkellistä ilmaantumistiheyttä (*failure rate*) hetkellä t kaikkien niiden havaintoyksiköiden joukossa, jotka ovat välttäneet kohdetapahtuman hetkeen t asti. Jos kohdetapahtumana on kuolema, uhkafunktiota kutsutaan myös ikäryhmittäiseksi kuolleisuustiheydeksi (*age-specific mortality rate*) tai kuolemanvaaraksi (*force of mortality*) niiden henkilöiden joukossa, jotka ovat siihen asti säilyneet hengissä. (Kalbfleisch, 1985)

Kun havaintoaineistoon ei sisälly kilpailevia tapahtumia, uhkafunktio $h(t)$ hetkellä t määritellään tapahtuman muutosnopeuden raja-arvona (Collett, 2015):

$$h(t) = \lim_{\Delta t \rightarrow 0} \left\{ \frac{\mathbb{P}(t \leq T < t + \Delta t | T \geq t)}{\Delta t} \right\}. \quad (4)$$

Raja-arvo (4) voidaan ehdollisen todennäköisyyden määritelmän nojalla ilmaista myös muodossa (Collett, 2015):

$$\lim_{\Delta t \rightarrow 0} \left\{ \frac{F(t + \Delta t) - F(t)}{\Delta t} \right\} \frac{1}{S(t)}. \quad (5)$$

Uhkafunktion, välttöfunktion (2) ja tiheysfunktion $f(t)$ välillä on siten seuraavanlainen yhteys, kun ei huomioida kilpailevia tapahtumia (Collett, 2015):

$$h(t) = \frac{F'(t)}{S(t)} = \frac{f(t)}{S(t)} = \frac{-S'(t)}{S(t)} = -\frac{d}{dt} \ln(S(t)). \quad (6)$$

Jos uhkafunktio on kasvava funktio, kohdetapahtuman hetkellinen ilmaantumistiheys kasvaa iän myötä eli mitä vanhemmasta havaintoyksiköstä on kyse, sitä todennäköisemmin se kohtaa kiinnostavan tapahtuman riskijaksolla Δt . Laskeva uhkafunktio ilmaisee, että kiinnostavan tapahtuman kohtaamisen todennäköisyys

riskijaksolla Δt pienenee iän myötä. Jos uhkafunktio on vakioarvoinen, ajan kuluminen ei vaikuta kiinnostavan tapahtuman kohtaamisen todennäköisyyteen millään Δt :n arvolla. (Kalbfleisch, 1985)

3.4 Kumulatiivinen uhkafunktio

Kumulatiivinen uhkafunktio $H(t)$ kuvaa havaintoyksikön kumulatiivista riskiä kohdata kiinnostava tapahtuma hetkeen t mennessä, kun ei oteta huomioon kilpailevia tapahtumia (Collett, 2015):

$$H(t) = \int_0^t h(u) du. \quad (7)$$

Välttöfunktion (2) ja kumulatiivisen uhkafunktion välinen yhteys saadaan selville integroimalla yhtälöä (6) (Kalbfleisch, 1985):

$$\begin{aligned} \ln(S(t)) &= - \int_0^t h(x) dx + C \Leftrightarrow \\ S(t) &= \exp\left\{- \int_0^t h(x) dx\right\} = \exp\{-H(t)\}. \end{aligned} \quad (8)$$

Integroimisvakion C arvoksi tulee 0, sillä ajanhetkellä 0 pätee (Kalbfleisch, 1985):

$$S(0) = \mathbb{P}(T > 0) = 1. \quad (9)$$

Kumulatiivinen uhkafunktio voidaan siis ilmaista välttöfunktion avulla (Collett, 2015):

$$H(t) = - \ln\{S(t)\}. \quad (10)$$

3.5 Parametriton estimointi

Kumulatiivista ilmaantuvuusfunktiota ja välttöfunktiota voidaan estimoida parametrittomalla Kaplanin-Meierin menetelmällä. Estimaattori välttöfunktiolle $S(t)$ (2) hetkestä 0 lähtien on:

$$\hat{S}(t) = \prod_{a_k \leq t} \left(1 - \frac{D_k}{n_k}\right), \quad (11)$$

jossa D_k on hetkellä a_k havaittujen tapahtumien määrä, ja n_k on riskijoukon koko kyseisellä hetkellä, kun $k = 1, 2, \dots$. Riskijoukko sisältää ne havaintoyksiköt, jotka ovat seuruussa juuri ennen hetkeä a_k ja jotka ovat siihen asti vältäneet kiinnostavan tapahtuman. (Läärä, 2018)

Estimaattorin $\hat{S}(t)$ kuvaaja on porrassfunktio, joka laskeutuu tasolle 0 hetkellä a_k , jos kaikki riskijoukkoon vielä kuuluvat kokevat kyseisellä hetkellä kiinnostavan tapahtuman. Muutoin välttökäyrä jää tasolle, jossa se on viimeisen tapahtuman hetkellä ja jatkuu tällä tasolla hetkeen, joka on pisin katoon päättyvä seuruu-aika. (Läärä, 2018)

Kumulatiivisen ilmaantuvuusfunktion $F(t)$ (1) estimaattori saadaan $\hat{S}(t)$:n komplementtina (Läärä, 2018):

$$\hat{F}(t) = 1 - \hat{S}(t). \quad (12)$$

Välttöfunktioita ja kumulatiivista ilmaantuvuusfunktioita voidaan estimoida myös muilla parametrittömillä menetelmillä, mutta niitä ei käsitellä tässä tutkielmassa.

4 Coxin regressiomalli

Nykyään elinaika-analyysissä käytetään useita parametrittomia ja parametrisia menetelmiä, sekä näiden molempien piirteitä sisältäviä semiparametrisia menetelmiä (Collett, 2015). Tässä tutkielmassa käytetään semiparametrista Coxin regressiomallia (Cox, 1972), jonka avulla voidaan mm. tutkia elinaikojen ja niitä selittävien muuttujien välisiä suhteita.

Englantilainen tilastotieteilijä D. R. Cox julkaisi vuonna 1972 *Journal of the Royal Statistical Society* -kausijulkaisussa artikkelin *Regression Models and Life-Tables*, johon on Web of Sciencen tilastojen mukaan viitattu yli 34 000 kertaa. Artikkelissa Cox ilmoitti jatkavansa E.L. Kaplanin ja P. Meierin viitoittamaa tietä elinajan taulujen (*life table*) tilastollisen teorian parissa, ja erityisesti Cox kertoi pyrkivänsä tuomaan regressioanalyysin kaltaista lähestymistapaa elinajan taulujen analyysiin. Cox esitti artikkelissaan tilastollisen mallin, jolla voidaan estimoida uhkafunktiota (4). (Cox, 1972)

Coxin (1972) kehittämä malli perustuu ajatukselle, että kahden ryhmän uhkasuhde on jokaisella ajanhetkellä vakio. Mallia kutsutaankin usein nimellä suhteellisten tai verrannollisten uhkien malli (*proportional hazards model*), mutta se tunnetaan myös Coxin regressiomallina (Collett, 2015), ja tätä nimeä käytetään tässä tutkielmassa.

4.1 Mallin määrittely

Coxin regressiomalli havaintoyksikön i kokemalle uhalle ajanhetkellä t voidaan esittää yleisessä muodossa seuraavasti (Collett, 2015):

$$h_i(t) = \psi(\mathbf{x}_i)h_0(t). \quad (13)$$

Vektori \mathbf{x}_i sisältää havaintoyksikölle i havaitut selittävien muuttujien arvot $x_{1i}, x_{2i}, \dots, x_{pi}$. Perustason uhkafunktio (*baseline hazard function*) $h_0(t)$ on uhkafunktion arvo sellaisella havaintoyksiköllä, jolla kaikki selittävät muuttujat saavat arvon nolla. $\psi(\mathbf{x}_i)$ tulkitaan siten uhkasuhteena (*hazard ratio*) ajanhetkellä t (Collett, 2015):

$$\psi(\mathbf{x}_i) = \frac{h_i(t)}{h_0(t)}. \quad (14)$$

$\psi(\mathbf{x}_i)$ saa vain positiivisia arvoja. Se voidaan parametrisoida useammalla eri tavalla, kun selittävien muuttujien kerroinvektori on $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_p)$ (Cox ja Oakes, 1984), kuten esimerkiksi:

- Log-lineaarinen parametrisointi: $\psi(\mathbf{x}_i) = \exp(\boldsymbol{\beta}^T \mathbf{x}_i)$
- Lineaarinen parametrisointi: $\psi(\mathbf{x}_i) = 1 + \boldsymbol{\beta}^T \mathbf{x}_i$.

Näistä log-lineaarinen parametrisointi on kaikkein suosituin (Cox ja Oakes, 1984), ja tätä parametrisointitapaa käytetään tässä tutkielmassa Coxin regressiomallin esittelyyn.

Coxin yleisessä muodossa esitetty regressiomalli (13) havaintoyksikölle i on log-lineaarisesti parametrisoituna (Collett, 2015):

$$\begin{aligned} h_i(t) &= \exp(\boldsymbol{\beta}^T \mathbf{x}_i) h_0(t) \\ &= \exp(\beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_p x_{pi}) h_0(t). \end{aligned} \tag{15}$$

Linearikombinaatio $\beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_p x_{pi}$ muodostaa Coxin regressiomallin mallin lineaarisen ennustimen. (Collett, 2015) Sen sijaan mallissa ei oleteta elinajoille mitään tiettyä jakaumaa eikä siten myöskään perustason uhkafunktion $h_0(t)$ muodosta tehdä oletuksia. (Cox, 1972) Nykyään Coxin regressiomalli luokitellaankin semiparametrisiin malleihin, mutta tämä käsite alkoi yleistyä tilastotieteessä vasta 1980-luvulla, useita vuosia Coxin mallin julkaisun jälkeen (Fan ja Koull, 2006). Coxin regressiomallia voidaan soveltaa myös puhtaasti parametrisena mallina, jolloin elinajoille oletetaan jokin jakauma (Collett, 2015). Mallin parametrisia versioita ei kuitenkaan käsitellä tässä tutkielmassa.

Coxin regressiomallissa sallitaan havaintojen sensuroituminen, mutta sensuroitumisen pitää olla riippumatonta, jolloin havaintoyksikön elinaika ei riipu sensuroitumisen aikaansaavasta mekanismista. Coxin regressiomallin pohjalta on kehitetty malleja myös tilanteisiin, joissa sensuroituminen on riippuvaa tai tapahtuu tietyllä aikavälillä (Collett, 2015) mutta näitä malleja ei käsitellä tässä tutkielmassa.

Uhkafunktion mallissa (15) oletetaan, että uhkafunktio $h_i(t)$ on jatkuva eli jokaisella ajanhetkellä t voi esiintyä korkeintaan yksi sensuroituminen tai yksi kohdetapahtuma. Cox esitti regressiomallistaan myös diskreetin version, jossa sallitaan useampi sensuroituminen tai kohdetapahtuma ajanhetkeä kohden (Cox, 1972), mutta sitä ei käsitellä tässä tutkielmassa.

Uhkafunktio ei aina ole verrannollinen koko aineistossa, mutta se saattaa kuitenkin olla sitä erilaisissa datan osaryhmissä. Malli voidaan tällöin osittaa (*stratified*), ja perustason uhkafunktio voidaan määrittellä kullekin ositteelle erikseen. Tällöin oletetaan, että mallioletukset verrannollisuudesta pätevät selittävälle muuttujille jokaisessa ositteessa erikseen. (Collett, 2015)

Coxin regressiomalli i :lle havaintoyksikölle j :ssä ositteessa on muotoa:

$$h_{ij}(t) = \exp(\boldsymbol{\beta}^T \mathbf{x}_{ij}) h_{0j}(t). \quad (16)$$

Ositusmuuttujaa ei sisällytetä mallin lineaariseen osaan (Collett, 2015).

4.2 Välttöfunktio ja kumulatiivinen uhkafunktio

Log-lineaarisesti parametrisoidusta Coxin regressiomallista (15) saadaan muodostettua kumulatiivinen uhkafunktio integroimalla yhtälö puolittain määritelmän (7) mukaisesti (Collett, 2015):

$$\begin{aligned} \int_0^t h_i(u) du &= \exp(\boldsymbol{\beta}^T \mathbf{x}_i) \int_0^t h_0(u) du \Leftrightarrow \\ H_i(t) &= \exp(\boldsymbol{\beta}^T \mathbf{x}_i) H_0(t). \end{aligned} \quad (17)$$

Kumulatiivisen uhkafunktion avulla voidaan muodostaa välttöfunktion (8) lauseke (Collett, 2015):

$$\begin{aligned} -H_i(t) &= -\exp(\boldsymbol{\beta}^T \mathbf{x}_i) H_0(t) \Leftrightarrow \\ \exp(-H_i(t)) &= \exp(-\exp(\boldsymbol{\beta}^T \mathbf{x}_i) H_0(t)) \Leftrightarrow \\ S_i(t) &= S_0(t)^{\exp(\boldsymbol{\beta}^T \mathbf{x}_i)}. \end{aligned} \quad (18)$$

4.3 β -kertoimien estimointi ja tulkinta

Coxin regressiomallin lineaarinen ennustin $\beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_p x_{pi}$ voi sisältää numeerisia tai luokiteltuja muuttujia, interaktiotermejä ja termejä, joissa on yhdistetty numeerinen ja luokiteltu muuttuja. Lisäksi mallissa voi olla mukana aikariippuvaisia muuttujia. Kaikkiin näihin muuttujiin liittyy oma β -kerroin, jonka estimaatti kuvaa muuttujan vaikutusta uhkafunktiossa. Malliin voidaan ottaa mukaan myös epälineaarisia termejä (Collett, 2015).

Lineaarisen ennustimen β -kertoimet voidaan estimoida ilman, että estimoidaan perustason uhkafunktiota $h_0(t)$ tai uhkafunktion $h_i(t)$ tarkkaa muotoa. (Cox, 1972) Tällöin analyysin päähuomio on regressiokertoimissa eli siinä, millainen vaikutus selittävillä muuttujilla on uhkafunktioon. (Collett, 2015) β -kertoimien estimointi perustuu uskottavuusfunktion maksimointiin, ja Cox esitti uskottavuusfunktion

sekä jatkuva-arvoiselle että diskreetille aika-akselille. Kumpikin uskottavuusfunktio huomioi sensuroitumisen. (Cox, 1975, 1972)

Cox (1972) määritteli regressiomallilleen seuraavan uskottavuusfunktion, kun aika-akseli oletetaan jatkuvaksi:

$$L(\boldsymbol{\beta}) = \prod_{j=1}^r \frac{\exp(\boldsymbol{\beta}^T \mathbf{x}_{(j)})}{\sum_{l \in R(t_{(j)})} \exp(\boldsymbol{\beta}^T \mathbf{x}_{(l)})}. \quad (19)$$

Uskottavuusfunktiossa (19) oletetaan, että n :n havaintoyksikön joukossa on r kohdetapahtumaa ja $n - r$ sensuroitumista. Aika-akselin jatkuvuudesta seuraa, ettei tapahtumien tai sensuroitumisten kesken periaatteessa ole sidoksia eli jokaista ajanhetkeä kohden on havaittu ainoastaan yksi kohdetapahtumana tai sensuroituminen. Kohdetapahtumat voidaan siten järjestää ajan suhteen: $t_{(1)} < t_{(2)} < \dots < t_{(r)}$. Riskijoukko $R(t_{(j)})$ sisältää ne havaintoyksiköt, jotka ovat eivät ajanhetkeen $t_{(j)}$ mennessä ole vielä kokeneet kohdetapahtumaa tai sensuroitumista. (Collett, 2015)

Uskottavuusfunktion (19) osoittajassa $\mathbf{x}_{(j)}$ sisältää selittävien muuttujien arvojen vektorin niille havaintoyksiköille, joille on sattunut kohdetapahtuma hetkellä $t_{(j)}$. Nimittäjässä oleva summa lasketaan kaikkien niiden havaintoyksiköiden yli, joilla on riski kokea kohdetapahtuma ajanhetkellä $t_{(j)}$. Sen sijaan uskottavuusfunktiossa oleva tulo lasketaan kaikkien niiden havaintojen yli, joille on jo havaittu kohdetapahtuma. Sensuroituneita havaintoyksiköitä ei siis huomioida uskottavuusfunktion osoittajaa laskettaessa, mutta ne ovat mukana nimittäjän summalausekkeessa silloin, kun kohdetapahtuman ajanhetki edeltää sensuroitumisen ajanhetkeä. Uskottavuusfunktion maksimointi tapahtuu yleisimmin Newton-Raphsonin menetelmän avulla. (Collett, 2015)

Cox (1972) luonnehti vuoden 1972 artikkelissaan, että (19) on ehdollinen uskottavuusfunktio (*conditional likelihood*). Hänen perusteluitaan kuitenkin kritisoitiin, ja Cox itsekin myöntää artikkelinsa *Discussion*-osiossa, että hänen uskottavuutta koskevat väitteensä kaipaavat selvennystä ja tarkennusta (Cox, 1972). Kalbfleisch ja Prentice (1973) ehdottivat, että Coxin esittämä uskottavuus (19) on itse asiassa sijalukuvektorin (*rank vector*) marginaalinen uskottavuus eli ehdollistamaton todennäköisyys sille, että havainnot esiintyvät juuri siinä järjestyksessä kuin ne kyseisessä aineistossa on havaittu. Aineistossa ei kuitenkaan saa olla mukana sensuroituneita havaintoja eikä aikariippuvaisia muuttujia. (Kalbfleisch ja Prentice, 1973)

Vuonna 1975 Cox nimesi *Partial Likelihood* -artikkelissaan kuvaamansa uskottavuuden (19) osittaisuskottavuudeksi (*partial likelihood*) ja tarkensi sen määritelmän.

mää. Osittaisuskottavuus on Coxin mukaan osa täyden uskottavuuden lauseketta, jota kuitenkin on kokonaisuutena vaikea määritellä ilman parametrissa mallia $h_0(t)$:lle. (Cox, 1972)

Osittaisuskottavuusfunktion määrittely perustuu siihen, että kokonaisuskottavuus hajotetaan tekijöihin todennäköisyyslaskennan tulosäännön avulla, ja kiusaparametreja sisältävät tekijät voidaan jättää huomioimatta, kun estimoidaan kiinnostavia parametreja. (Lawless, 2003)

Cox (1975, 1972) toteaa, että osittaisuskottavuusfunktioon pätevät tavanomaiset parametrusten uskottavuusfunktioiden asymptootiset ominaisuudet. Estimoitu $\hat{\beta}$ -kerroin on siten osittaisuskottavuuden (19) maksimiarvo, ja $n^{1/2}(\hat{\beta} - \beta)$ on asymptoottisesti $N(0, I^{-1}(\beta))$ -jakautunut, kun $I(\beta)$ on osittaisuskottavuuden Fisherin informaatiomatriisi. (Bailey, 1983)

Cox ei kuitenkaan pitävästi todistanut väitteitään yleisten asymptoottisten ominaisuuksien pätevydestä myös osittaisuskottavuuden tapauksessa (Bailey, 1983). Tämä tehtiin vasta myöhemmin muiden tutkijoiden toimesta. Esimerkiksi Tsiatis (1981) osoitti, että asymptoottiset oletukset pätevät silloin, kun havaintoyksiköt valitaan satunnaisotannalla, niiden kovarianssirakenne tiedetään ja aineistossa on satunnaisesti sensuroituneita havaintoja. Bailey (1983) puolestaan osoitti asymptotiikan pätevän siinä kliinisissä kokeissa tavallisessa tilanteessa, että kovariaattien määrä on kiinteä.

Asymptotiikan perusteella β -kertoimen varianssi ja keskivirhe saadaan estimoitua kovarianssimatriisista. Kertoimen luottamusväli on $\hat{\beta} \pm z_{1-\alpha/2} \times \text{SE}(\hat{\beta})$, jossa $z_{1-\alpha/2}$ on luottamustasoa $100(1-\alpha)\%$ vastaava $N(0, 1)$ -jakauman fraktiili (Collett, 2015).

Luottamusväli voidaan laskea myös uhkasuhteelle $\psi = e^\beta$ (13). Uhkasuhteen luottamusväli saadaan ottamalla eksponentti β -kertoimen luottamusvälistä:

$$\begin{aligned} & \exp\{\hat{\beta} \pm z_{1-\alpha/2} \times \text{SE}(\hat{\beta})\} \\ &= \left[\frac{\hat{\psi}}{e^{z_{1-\alpha/2} \times \text{SE}(\hat{\beta})}}, \hat{\psi} e^{z_{1-\alpha/2} \times \text{SE}(\hat{\beta})} \right], \end{aligned} \quad (20)$$

jossa $z_{1-\alpha/2}$ on luottamustasoa $100(1-\alpha)\%$ vastaava $N(0, 1)$ -jakauman fraktiili (Collett, 2015).

Tämän tutkielman selittävät muuttujat (luku 2.4) ovat luokitteluasteikollisia muuttujia. Luokitellulle muuttujalle (*factor*) A on määritelty $p \geq 2$ eri tasoa tai ryhmää, ja regressioparametrit $\alpha_1, \alpha_2, \dots, \alpha_p$ kuvaavat tasojen ja ryhmien päävaiku-

tusta. Yksi muuttujan tasoista valitaan ns. nollassoksi, ja sen regressiokerroin saa arvokseen 0. (Collett, 2015).

Perustason uhkafunktiossa luokitellun muuttujan arvoina ovat nollasson kertoimien arvot. Malliin lisättävä luokiteltu muuttuja on muoto $\alpha_1 x_{1i} + \alpha_2 x_{2i} + \dots + \alpha_p x_{pi}$, jossa indikaattorimuuttuja x_i saa arvon 1 vain silloin, kun havaintoyksikölle i on havaittu luokitellun muuttujan tason α_i mukainen arvo. Perustasolla indikaattorimuuttujan arvo on nolla. (Collett, 2015)

e^{α_i} ilmaisee uhkasuhteen luokitellun muuttujan tietyn tason i ja perustason välillä, ja regressiokerroin puolestaan ilmaisee ko. suhteen logaritmin (Collett, 2015):

$$\begin{aligned} e^{\alpha_i} &= \frac{h_0(t) \exp\{\alpha_i\}}{h_0(t)} \Leftrightarrow \\ \alpha_i &= \ln \left(\frac{h_0(t) \exp\{\alpha_i\}}{h_0(t)} \right). \end{aligned} \tag{21}$$

4.4 Mallidiagnostiikka

Mallidiagnostiikka tarkoittaa dataan sovitetun mallin kriittistä arviointia ja erityisesti mallioletusten paikkaansapitävyyden tutkimista. Coxin regressiomallin tärkein mallioletus on oletus verrannollisista uhista. Jotta dataan sovitettu malli pätee, kahden ryhmän uhkasuhteen pitää olla jokaisella ajanhetkellä t vakio. Tämän mallioletuksen tutkimiseen on kehitetty useita jäännöstermeihin eli residuaaleihin (*residual*) perustuvia menetelmiä, joista tässä tutkielmassa esitellään Schoenfeldin residuaalit (Schoenfeld, 1982) ja niiden pohjalta kehitetty ns. zph-testi.

Coxin regressiomallin verrannollisuusoletusta voidaan testata Schoenfeldin residuaalien avulla, jotka D. Schoenfeld alunperin nimesi osittaisiksi residuaaleiksi (*partial residuals*). Niiden käyttö ei vaadi kumulatiivisen uhkafunktion estimointia eivätkä ne myöskään riipu havaituista elinajoista. (Schoenfeld, 1982) Schoenfeldin residuaaleja ei lasketa jokaiselle havainnolle erikseen, vaan residuaalit tuottavat arvon jokaiselle Coxin regressiomallissa käytetylle muuttujalle. (Collett, 2015)

Schoenfeldin residuaali i mallin selittävälle muuttujalle X_j on

$$r_{Sij} = \delta\{x_{ji} - \hat{\alpha}_{ji}\}, \tag{22}$$

jossa $j = 1, 2, \dots, p$, i on havaintoyksikkö ja

$$\hat{\alpha}_{ji} = \frac{\sum_{l \in R(t_i)} x_{jl} \exp(\hat{\beta}' \mathbf{x}_l)}{\sum_{l \in R(t_i)} \exp(\hat{\beta}' \mathbf{x}_l)}, \quad (23)$$

jossa $R(t_i)$ on kaikkien niiden havaintoyksiköiden i joukko, jotka kuuluvat riskijoukkoon hetkellä t .

Grambsch ja Therneau (1994) esittivät skaalatun version Schoenfeldin residuaaleista, ja niiden käyttäminen on tehokas keino Coxin regressiomallin mallioletusten tarkasteluun. Olkoon Schoenfeldin residuaalivektori i :lle havaintoyksikölle $\mathbf{r}_{Si} = (r_{S1i}, r_{S2i}, \dots, r_{Sp_i})'$. Skaalatut residuaalit saadaan tällöin vektorin

$$\mathbf{r}^*_{Si} = d \text{var}(\hat{\beta}) \mathbf{r}_{Si} \quad (24)$$

komponenteista, missä d on kiinnostavien tapahtumien määrä n havaintoyksikön joukossa, ja $\text{var}(\hat{\beta})$ on sovitetun mallin parametriestimaattien variassi-kovariassimatriisi. (Collett, 2015)

Grambsch ja Therneau (1994) ovat esittäneet skaalattuihin Schoenfeldin residuaaleihin perustuvat testisuureet sekä Coxin regressiomallin yksittäiselle selittäväälle muuttujalle X_j että koko mallille. Mallioletusten vallitessa yksittäisen selittäjän testisuure noudattaa likimäärin χ^2 -jakaumaa vapausasteenaan termin vapaiden parametrien lukumäärä. Koko mallin testisuureen avulla testataan verrannollisuusoletusta yli kaikkien malliin sisältyvien selittäjien yli. Kun verrannollisuusoletus pätee jokaiselle mallin selittäjälle p , globaali testisuure noudattaa likimäärin χ^2 -jakaumaa vapausastein p . Tämä testi tunnetaan nimellä Grambschin ja Therneauin uhkasuhteiden testi (*Grambsch and Therneau test of proportional hazards*), ja siitä käytetään myös nimitystä zph-testi. (Collett, 2015)

Mallidiagnostiikassa voidaan myös tutkia sitä, ovatko jotkin havaintoyksiköt erityisen vaikuttavia β -kertoimien estimoinnin kannalta. $\hat{\beta}$ -kertoimiin vaikuttavien havaintojen esiintymistä datassa voidaan testata ns. delta-beeta-menetelmällä, jossa muutos kertoimessa $\hat{\beta}_j$, kun datasta ollaan poistettu havainto i , merkitään

$$\Delta_i \hat{\beta}_j \approx \hat{\beta}_j - \hat{\beta}_{j(i)}. \quad (25)$$

Delta-beeta-tunnusluku voidaan jakaa $\hat{\beta}$ -kertoimen keskivirheellä, jolloin saadaan standardoitu delta-beeta-tunnusluku. Niiden avulla voidaan tutkia, kuinka paljon yksittäisen havainnon i poistaminen datasta vaikuttaa tietyn β -kertoimen estimaatin merkitsevyyteen. (Collett, 2015)

5 Coxin regressiomallin sovitus

Tutkielman aineistoon sovitettiin Coxin regressiomalli, ja sen tulokset ja mallidiagnostiikka on esitetty tässä luvussa. Selittävien muuttujien kuvailevan tarkastelun tulokset koko aineistossa on kerrottu luvussa 5.1. Malli päädyttiin sovittamaan sekä ositettuna että erikseen palkka- ja vuorotteluvapaatölle, ja nämä tulokset löytyvät luvuista 5.2 ja 5.3.

Kuvailevassa tarkastelussa, mallien sovituksessa ja mallidiagnostiikassa käytettiin R-ohjelmiston versiota 4.0.4 (2021-02-15), R Studio -ohjelman versiota 1.1.423, Epi-paketin versiota 2.43 ja Survival-paketin versiota 3.2.7.

5.1 Aineiston kuvaileva tarkastelu

Aineistoon sovitetuissa malleissa on mukana kolme muuttujaa: työn kesto, työpaikan tilastollinen kuntatyyppi ja työpaikan tyyppi. Taulukossa 5 on esitetty selittävien muuttujien lukumäärät ja prosentuaaliset osuudet koko aineistossa. Muuttujissa ei ole puuttuvia arvoja.

	<i>n</i>	%
Työn kesto		
0-3 kk	839	27
3-6 kk	1823	59
Yli 6 kk	447	14
Kuntatyyppi		
Kaupunkimainen	1787	57
Taajaan asuttu	589	19
Maaseutumainen	733	24
Työpaikan tyyppi		
Palkka	1463	47
Vuorotteluvapaa	1646	53

Taulukko 5: Selittävien muuttujien frekvenssit koko aineistossa.

Kaikki kolme muuttujaa ovat luokiteltuja muuttujia. Työn kesto -muuttuja on järjestysasteikollinen muuttuja, jolla on aineistossa 6 luokkaa: alle 11 päivää, 11 päivää-1 kk, 1-3 kk, 3-6 kk, 6-12 kk sekä yli 12 kk. Kuvailevassa tarkastelussa ja mallien sovituksessa on kuitenkin käytetty karkeampaa jakoa.

Työpaikan tilastollinen kuntatyyppi ja työpaikan tyyppi ovat luokitteluasteikollisia muuttujia. Aineistossa kuntatyyppillä on kolme tasoa (kaupunkimainen, taajaan

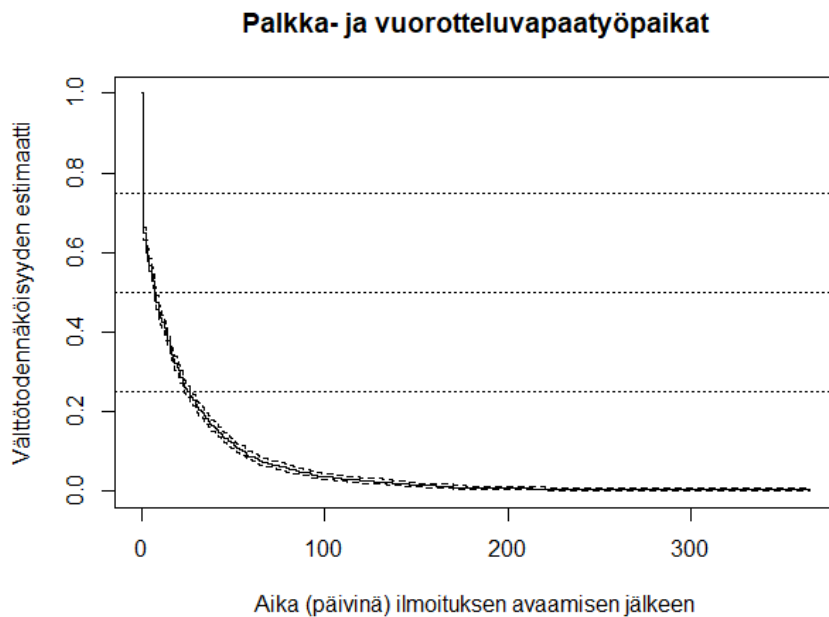
asuttu, maaseutumainen) ja työpaikan tyypillä kaksi tasoa (palkka, vuorotteluvapaa), ja tätä samaa jaottelua on käytetty kuvailevassa tarkastelussa ja mallien sovituksessa.

Lähes 60 % julkisen hallinnon ja maanpuolustuksen avoimista työpaikoista on kestoaltaan 3-6 kuukautta. Pidempiä, yli 6 kuukauden työpaikkoja on vain noin 14 % ja lyhyitä, alle 3 kuukauden työpaikkoja 27 %.

Lähes 60 % prosenttia avoimista työpaikoista sijaitsee kaupunkimaisissa kunnissa. Maaseutumaisissa kunnissa sijaitsee hieman enemmän (24 %) työpaikkoja kuin taajaan asuissa kunnissa (19 %).

Palkkatyön ja vuorotteluvapaiden osuus avoimista työpaikoista jakaantuu aineistossa lähes tasan.

Kuvassa 2 on esitetty kiinnostavan tapahtuman eli avoimen työpaikan täyttymisen välttötodennäköisyyden estimaatti kaikille aineiston avoimille työpaikoille. Estimaatti on laskettu Kaplanin-Meierin menetelmällä. Välttötodennäköisyys ilmaisee todennäköisyyden, että avoinna ollut työpaikka täyttyy vasta hetken t jälkeen eli sen jälkeen, kun työpaikkailmoitus on ollut avoinna t päivää.



Kuva 2: Julkisen hallinnon ja maanpuolustuksen työpaikkojen täyttymisen välttötodennäköisyyden estimaatti riskijaksolla $[0,365]$ päivää vuonna 2018.

Kuvasta 2 havaitaan, että tietty osa työpaikoista täyttyy heti työpaikkailmoituksen avaamisen jälkeen. Noin sadan päivän jälkeen ilmoituksen avaamisesta välttötodennäköisyydessä ei tapahdu kovin suuria muutoksia. Jatkossa välttötodennäköisyyttä tarkastallaankin riskijaksolla $[0,120]$ päivää. Analysoitavassa aineistossa on niin paljon tapahtumia, että välttökäyrän estimaatin luottamusväli vaikuttaa kapealta koko seuruaajalla.

Tämän tutkielman aihepiirissä välttötodennäköisyyttä voidaan pitää julkisen työnvälityksen tehokkuuden mittarina, joka mittaa työpaikan täyttöprosessin tehokkuutta. Tällöin ajatellaan, että mitä alempana välttöfunktion käyrä kulkee, sitä tehokkaampi täyttöprosessi on ollut. (Räisänen, 2016) Eli mitä alempana välttöfunktion käyrä kulkee, sitä pienemmällä todennäköisyydellä työpaikan täytyminen on vältetty hetkeen t asti eli työpaikka on edelleen avoin.

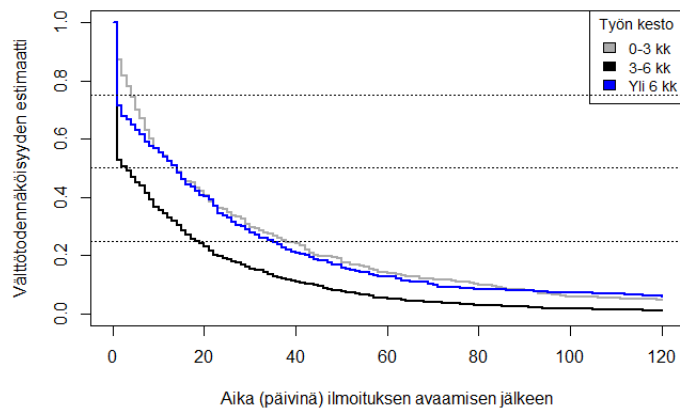
Kuvasa 3 on esitetty Kaplanin-Meierin menetelmällä lasketut välttötodennäköisyyden estimaatit kullekin mallin muuttujalle.

Työn keston (kuva 3a) mukaan lasketuista estimaateista nähdään, että 3-6 kuukautta kestävien työsuhteiden täyttöprosessi on ollut jonkin verran tehokkaampi kuin sitä lyhyempien tai pidempien työsuhteiden täyttöprosessi. Pidempien ja lyhyempien työsuhteiden täyttöprosessin tehokkuus on kuitenkin lähes samalla tasolla.

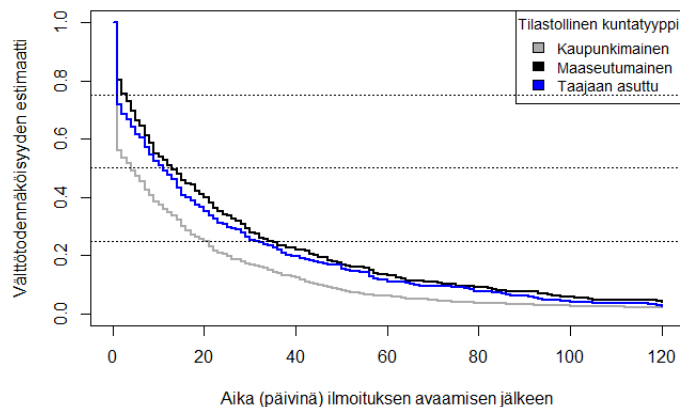
Kaupunkimaisissa kunnissa sijaitsevien työpaikkojen täyttö on ollut tehokkaampaa kuin maaseutumaisissa ja taajaan asutuissa kunnissa (kuva 3b). Maaseutumaisten ja taajaan asuttujen kuntien työpaikkojen täytön tehokkuus sen sijaan ei juurikaan eroa toisistaan.

Palkkatöiden ja vuorotteluvapaiden (kuva 3c) välinen ero työpaikan täyttymisen tehokkuudessa on iso. Vuorotteluvapaissa yli puolet avoimista työpaikoista täyttyy heti ilmoituksen avaamisen jälkeen, palkkatöissä sen sijaan vain pieni osa työpaikoista saadaan täytettyä näin nopeasti.

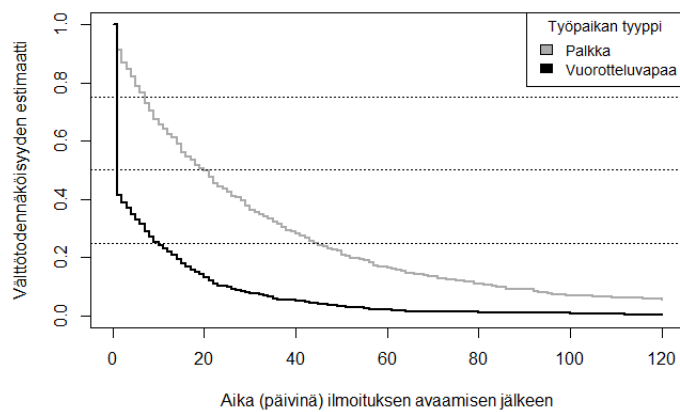
Koska palkkatöissä ja vuorotteluvapaissa havaittiin selkeä ero välttöfunktion estimaatissa, koko aineistoon sovitettu malli päätettiin osittaa työpaikan tyyppin mukaan (luku 5.2). Vertailun vuoksi aineisto jaettiin lisäksi kahteen osaan - palkka- ja vuorotteluvapaisiin -, joihin Coxin regressiomalli sovitettiin erikseen (luku 5.3).



(a) Välttötodennäköisyyden estimaatti työn keston mukaan



(b) Välttötodennäköisyyden estimaatti kuntatyyppin mukaan



(c) Välttötodennäköisyyden estimaatti työpaikan tyyppin mukaan

Kuva 3: Julkisen hallinnon ja maanpuolustuksen työpaikkojen täyttymisen välttötodennäköisyyden estimaatti työn keston, kuntatyyppin ja työpaikan tyyppin mukaan vuonna 2018.

5.2 Kaikki avoimet työpaikat

Julkisen hallinnon ja maanpuolustuksen avoimien työpaikkojen vuoden 2018 dataan sovitettiin Coxin regressiomalli, joka ositettiin työpaikan tyyppin (palkka, vuorotteluvapaa) mukaan. Mallissa selittävinä muuttujina olivat työn kesto ja tilastollinen kuntatyyppi. Malli i :lle havaintoyksikölle j :ssä ositteessa on muotoa:

$$\begin{aligned} h_j(t) &= \exp(\boldsymbol{\beta}^T \mathbf{x}_j) h_{0j}(t) \\ &= \exp(\beta_1 x_{1j} + \beta_2 x_{2j}) h_{0j}(t), \end{aligned} \quad (26)$$

jossa

- x_1 : Työpaikan tilastollinen kuntatyyppi (0 = Kaupunkimainen (viitetaso), 1 = Maaseutumainen, 2 = Taajaan asuttu)
- x_2 : Työn kesto (0 = alle vuosi (viitetaso), 1 = 12 kk tai yli)
- j : Työpaikan tyyppi (0 = Vuorotteluvapaa (viitetaso), 1 = Palkka).

Kunkin muuttujan viitetasoksi on valittu eniten havaintoja sisältävä taso.

Sovitetun mallin selittävien muuttujien estimoidut $\hat{\beta}$ -kertoimet ja niiden keskivirheet, uhkasuhteiden estimaatit ja 95 % luottamusvälit on esitetty taulukossa 6. Mallissa on myös mukana työpaikan tyyppin ja kuntatyyppin interaktiokerroin:

	$\hat{\beta}$ -kerroin	SE	Uhkasuhte	2.5 %	97.5 %
Maaseutumainen	-0.34	0.08	0.71	0.61	0.83
Taajaan asuttu	-0.06	0.07	0.94	0.82	1.09
Kesto: yli 12 kk	-0.44	0.12	0.64	0.51	0.81
Palkka:Maaseutumainen	0.46	0.10	1.58	1.29	1.92
Palkka:Taajaan asuttu	0.10	0.10	1.10	0.90	1.34

Taulukko 6: Selittävien muuttujien $\hat{\beta}$ -kertoimet ja niiden keskivirheet, uhkasuhteiden estimaatit sekä niiden 95 % luottamusvälit.

Estimoitujen uhkasuhteiden perusteella sekä maaseutumaisissa että taajaan asutuissa kunnissa on pienempi riski julkisen hallinnon ja maanpuolustuksen työpaikkojen täyttymiselle kuin kaupunkimaisissa kunnissa. Maaseutumaisissa kunnissa estimoitu riski on kuitenkin kaikkein pienin eli jokaisena päivänä, jolloin työpaikka on edelleen auki, riski sen täyttymiselle maaseutumaisissa kunnissa on 29 % pienempi kuin kaupunkimaisissa kunnissa. Taajaan asutuissa kunnissa vastaava estimoitu riski on vain 6 % pienempi kuin kaupunkimaisissa kunnissa.

Luottamusvälin perusteella taajaan asuttujen kuntien estimoidussa uhkasuhteessa on maaseutuun verrattuna enemmän epävarmuutta, ja on mahdollista, ettei näiden kuntien estimoitu riski juurikaan poikkea kaupunkimaisista kunnista eli estimoitu uhkasuhde saa arvon 1.

Työpaikan tyyppin ja kuntatyyppin interaktiotermin estimoidusta uhkasuhteesta havaitaan, että maaseudulla ja taajamissa sijaitsevilla avoimilla palkkatöillä estimoitu riski työpaikan täyttymiselle on korkeampi vuorotteluvapaisiin verrattuna. Täsäkin kertoimessa maaseutumaisen kunnan vaikutus (estimoitu uhkasuhde 1.58) on voimakkaampi kuin taajaan asutun (estimoitu uhkasuhde 1.10) kunnan.

Jos avoinna oleva työ kestää yli vuoden, estimoitu riski sen täyttymiselle on selvästi matalampi verrattuna alle vuoden mittaisiin työsuhteisiin. Jokaisena päivänä, jolloin yli vuoden mittainen työsuhde on edelleen auki, estimoitu riski sen täyttymiselle on 36 % pienempi kuin alle vuoden mittaisilla työsuhteilla.

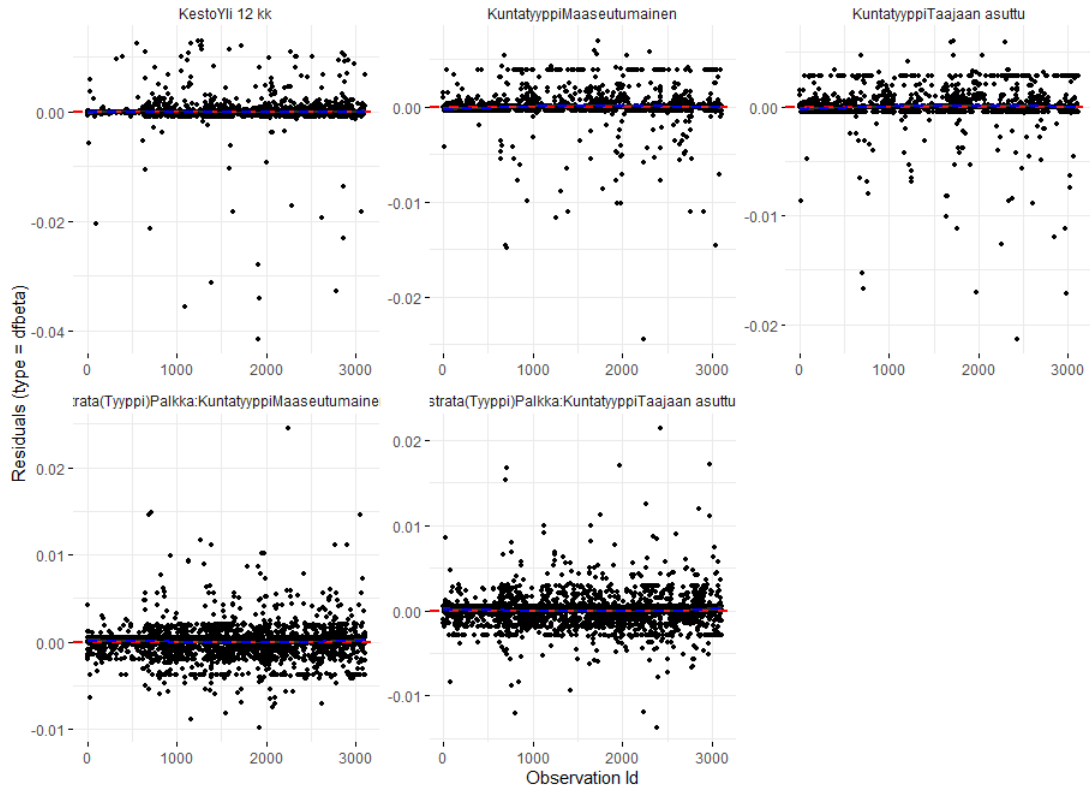
Mallioletusten voimassaoloa arvioitiin Schoenfeldin skaalattujen residuaalien ja `zph`-testin avulla. R:n `cox.zph()`-funktion tuloste näkyy taulukossa 7.

	Testisuure	Vapausasteet	<i>p</i> -arvo
Kuntatyyppi	1.64	2	0.44
Työn kesto	0.91	1	0.34
Työpaikan tyyppi : Kuntatyyppi	5.19	2	0.07
GLOBAL	6.60	5	0.25

Taulukko 7: R:n `cox.zph()`-funktion tuloste koko aineistoon sovitetulle mallille.

`zph`-testi ei viittaa siihen, että sovitettu malli poikkeaisi merkittävästi verrannollisuusoletuksesta. Mallia sovitettaessa kuitenkin huomattiin, että työn kesto -muuttujan luokittelulla oli vaikutusta `zph`-testin tulokseen. Mallin sovituksessa työn kesto jaettiin kahteen luokkaan sen mukaan, kestääkö työ alle vai yli 12 kuukautta. Useampia luokkia sisältävä työn keston luokittelu vaikutti `zph`-testin tulokseen siten, ettei oletus verrannollisesta uhkasuhteesta enää pätenyt. Estimoitujen uhkasuhteiden havaittiin siis olevan työn kesto -muuttujan osalta jonkin verran riippuvaisia ajasta.

$\hat{\beta}$ -kertoimiin vaikuttavien havaintojen olemassaoloa tutkittiin delta-beeta-tunnuslukujen avulla. Muuttujakohtaisten tunnuslukujen kuvaajat on esitetty kuvassa 4, ja ne kuvaavat kertoimissa tapahtuvaa muutosta, kun kukin havainto poistetaan datasta yksi kerrallaan.



Kuva 4: Standardoitujen delta-beeta-tunnuslukujen kuvaajat.

Kuvan 4 kuvajissa näkyy jonkin verran 0-tasosta poikkeavia delta-beeta-tunnuslukuja. Koska käytössä on kohtalaisen paljon dataa, näiden havaintojen vaikutus kertoi-miin arvioitiin niin vähäiseksi, ettei niitä poistettu analysoitavasta aineistosta. Delta-beeta-tunnuslukuja ei myöskään enää tutkittu erikseen datasta, joka sisälsi vain joko vuorotteluvapaa- tai palkkatöiden datan (luku 5.3).

5.3 Vuorotteluvapaa- ja palkkatyöt

Coxin regressiomalli sovitettiin myös koko aineistosta työpaikan tyyppin mukaan rajattuun dataan, eli erikseen julkisen hallinnon ja maanpuolustuksen vuorotteluvapaa- ja palkkatyöpaikkojen datoihin. Taulukossa 8 on esitetty selittävien muuttujien frekvenssit ja %-osuudet kummankin tyypissä työpaikoissa:

	Vuorotteluvapaa		Palkka	
	<i>n</i>	%	<i>n</i>	%
Työn kesto				
0-3 kk	37	2	802	55
3-6 kk	1438	87	385	26
Yli 6 kk	171	10	276	19
Yhteensä	1646	100	1463	100
Kuntatyyppe				
Kaupunkimainen	1220	74	567	39
Maaseutumainen	200	12	533	36
Taajaan asuttu	226	14	363	25
Yhteensä	1646	100	1463	100

Taulukko 8: Selittävien muuttujien frekvenssit vuorotteluvapaa- ja palkkatöissä.

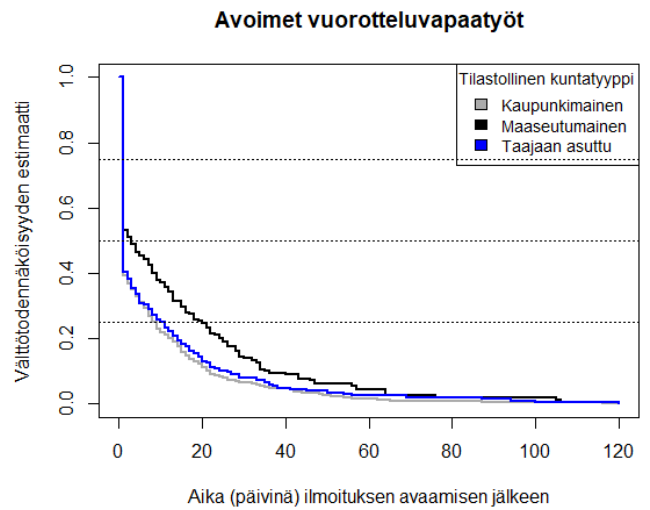
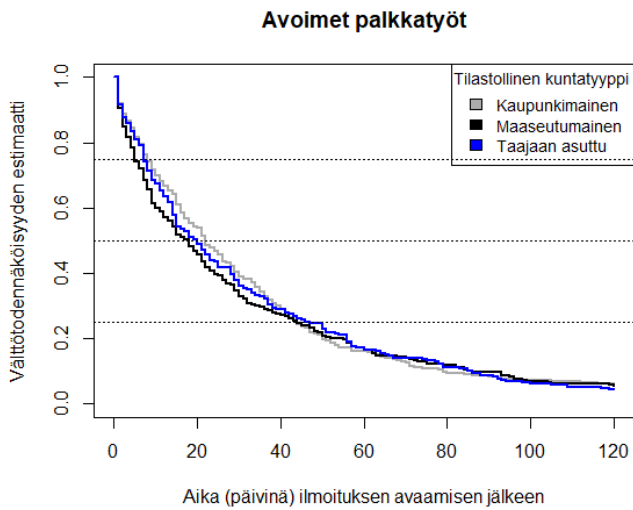
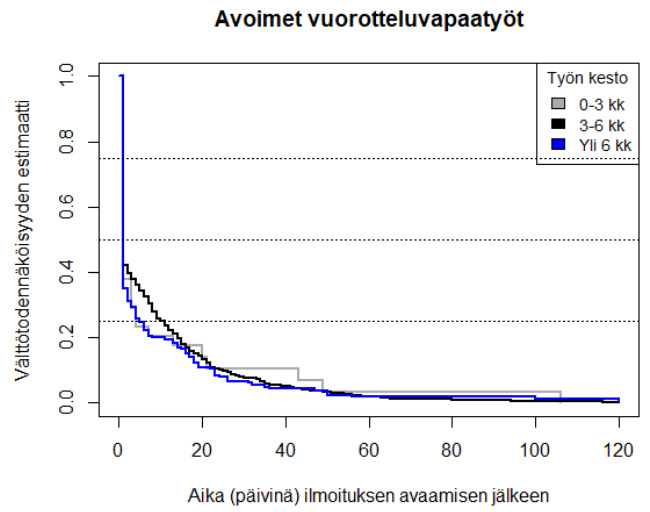
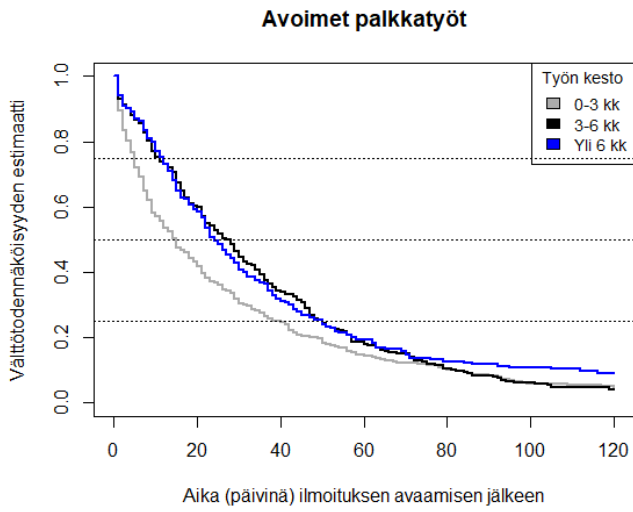
Suurin osa julkisen hallinnon ja maanpuolustuksen vuorotteluvapaaista on kestoltaan 3-6 kuukauden mittaisia työsuhteita. Alle 3 kuukauden mittaisia avoimia työpaikkoja on vain 37 kappaletta (2 %), ja yli puolen vuoden mittaisia 171 kpl (10 %). Vuorotteluvapaaan kesto on enintään 180 päivää eli 6 kuukautta, ja yli 6 kk-ryhmä sisältää nämä kestoltaan pisimmät vuorotteluvapaat.

Palkkatöissä suurin osa avoista työpaikoista on kestoltaan lyhyitä, alle 3 kuukauden mittaisia työsuhteita. Palkkatöiden avoimet työpaikat painottuvat siis lyhyempiin työsuhteisiin kuin vuorotteluvapaat. Loput avoimista palkkatyöpaikoista jakautuvat hieman tasaisemmin kuin vuorotteluvapaat. Noin neljännes avoimista palkkatöistä on kestoltaan 3-6 kuukautta ja noin viidennes yli puolen vuoden mittaisia.

Sekä vuorotteluvapaaissa että palkkatöissä suurin osa avoimista työpaikoista sijaitsee kaupunkimaisessa kunnassa, mutta vuorotteluvapaaissa painotus on suurempi (74 %) kuin palkkatöissä (39 %).

Vuorotteluvapaaissa seuruuajan eli työpaikkojen avoinnaolon keston keskiarvo on 9 päivää ja mediaani 1 päivä. Palkkatöissä avoinnaolon keston keskiarvo on 30 päivää ja mediaani 17 päivää.

Kuvassa 5 on esitetty Kaplanin-Meierin menetelmällä lasketut estimaatit työpaikan täyttymisen välttötodennäköisyyksille sekä palkkatöiden että vuorotteluvapaiden osalta.



Kuva 5: Julkisen hallinnon ja maanpuolustuksen palkkatöiden ja vuorotteluvapaiden täyttymisen välttödennäköisyyden estimaatit työn keston ja kuntatyyppin mukaan vuonna 2018.

Palkkatöissä tilastollisella kuntatyypillä ei ole suurta vaikutusta työpaikan täyttöprosessin tehokkuuteen. Sen sijaan vuorotteluvapaissa kaupunkimaisissa ja taajaan asutuissa kunnissa täyttöprosessi on jonkin verran tehokkaampaa kuin maaseutumaisissa kunnissa. Koko aineistosta laskettu kuntatyyppin mukainen välttötodennäköisyys (kuva 3) muistuttaa enemmän vuorotteluvapaan välttökäyrää sen suhteen, että maaseutumaisissa kunnissa täyttöprosessi on tehottominta ja kaupunkimaisissa kunnissa tehokkainta.

Työn keston suhteen lasketussa välttötodennäköisyyden estimaatissa ei vuorotteluvapaissa juurikaan ole eroa. Ilmoituksen avaamisen jälkeen 3-6 kuukauden mittaisen työsuhteiden täyttö on jonkin verran tehottomampaa kuin sitä lyhyempien tai pidempien työsuhteiden, mutta erot tasoittuvat noin 20 päivän kuluttua työpaikkailmoituksen avaamisesta.

Palkkatöissä lyhyiden työsuhteiden täyttöprosessi on selvästi tehokkaampi kuin yli 3 kuukauden mittaisten työsuhteiden täyttö. Jos työsuhde on pidempi kuin 3 kuukautta, työn kestolla ei ole enää sen jälkeen suurta merkitystä täyttöprosessin tehokkuuteen.

Kuvan 5 välttöfunktion käyristä huomataan, että suurin osa vuorotteluvapaisten työpaikoista on täyttynyt jo noin 40 päivää työpaikkailmoituksen avaamisen jälkeen. Palkkatöissä sen sijaan osa ilmoituksista on täyttämättä vielä 120 päivän kuluttua ilmoituksen avaamisesta. Vuorotteluvapaissa työpaikan täyttöprosessi on siis paljon tehokkaampi kuin palkkatöissä. Vuorotteluvapaissa myös yli puolet työpaikoista täyttyy heti ilmoituksen avaamisen jälkeen.

Taulukossa 9 on esitetty vain vuorotteluvapaatyöpaikat sisältävään dataan sovitetun Coxin regressiomallin $\hat{\beta}$ -kertoimet ja niiden keskivirheet sekä estimoidut uhkasuhteet ja niiden luottamusvälit. Työn kesto -muuttuja jaettiin kahteen ryhmään sen mukaan, kestäkö työsuhde alle vai yli 3 kuukautta.

	$\hat{\beta}$ -kerroin	SE	Uhkasuhde	2.5 %	97.5 %
Maaseutumainen	-0.35	0.08	0.71	0.61	0.82
Taajaan asuttu	-0.06	0.07	0.94	0.82	1.09
Kesto: alle 3 kk	0.04	0.17	1.04	0.75	1.46

Taulukko 9: Selittävien muuttujien $\hat{\beta}$ -kertoimet ja niiden keskivirheet, estimoidut uhkasuhteet sekä niiden 95 % luottamusvälit vuorotteluvapaatöissä.

Maaseutumaisissa kunnissa estimoitu riski vuorotteluvapaatyöpaikan täyttymiselle on 29 % pienempi kuin kaupunkimaisissa kunnissa. Taajaan asutuissa kunnissa estimoitu riski työpaikan täyttymiselle on lähellä viitetasoa eli noin 6 % viitetasoa

pienempi. Tilastollisen kuntatyypin osalta vuorotteluvapaiden dataan sovitettun mallin estimoidut $\hat{\beta}$ -kertoimet ja uhkasuhteet ovat itse asiassa yhteneväiset koko dataan sovitettun mallin kanssa.

Yli ja alle 3 kuukauden vuorotteluvapaatoissa estimoidut uhkasuhteet on hyvin lähellä toisiaan. Vuorotteluvapaissa alle 3 kuukauden työsuhteiden osuus on vain 2 % avoimista työpaikoista, ja tämä tuo epävarmuutta estimoituun uhkasuhteeseen, ja pidentää sen luottamusväliä. Myös kesto-muuttujan estimoidun $\hat{\beta}$ -kertoimen keskivirhe on isompi kuin kuntatyypin estimoiduissa kertoimissa.

Taulukossa 10 on esitetty R:n `cox.zph()`-funktion tuloste vain vuorotteluvapaiden dataan sovitettulle mallille. Vuorotteluvapaiden osalta `zph`-testin tulos ei viittaa siihen, että sovitettu malli poikkeaisi merkittävästi verrannollisuusoletuksesta.

	Testisuure	Vapausasteet	p -arvo
Kuntatyyppi	0.11	1	0.74
Työn kesto	0.55	1	0.46
GLOBAL	0.64	2	0.73

Taulukko 10: R:n `cox.zph()`-funktion tuloste vain vuorotteluvapaat sisältävään dataan sovitettulle mallille.

Taulukossa 11 on esitetty vain palkkatyöpaikat sisältävään dataan sovitettun Coxin regressiomallin estimoidut $\hat{\beta}$ -kertoimet ja niiden keskivirheet sekä estimoidut uhkasuhteet ja niiden luottamusvälit. Dataan sovitettiin malli, joka ositettiin työn keston mukaan. Kesto-muuttujan arvot jaettiin kahteen luokkaan sen mukaan, kestääkö palkkatyö alle vai yli 3 kuukautta. Malliin sisältyi myös kuntatyypin ja kesto-muuttujan interaktiotermin.

	$\hat{\beta}$ -kerroin	SE	Uhkasuhde	2.5 %	97.5 %
Maaseutumainen	0.18	0.09	1.20	1.00	1.43
Taajaan asuttu	0.07	0.10	1.07	0.88	1.30
Yli 3 kk:Maaseutumainen	-0.36	0.14	0.70	0.53	0.91
Yli 3 kk:Taajaan asuttu	-0.18	0.15	0.84	0.63	1.12

Taulukko 11: Selittävien muuttujien $\hat{\beta}$ -kerroimet ja niiden keskivirheet, estimoidut uhkasuhteet sekä niiden 95 % luottamusväli palkkatöissä.

Työpaikan sijainti maaseutumaisessa ja taajaan asutussa kunnassa nostaa estimoitua riskiä avoimen palkkatyöpaikan täyttymiselle verrattuna kaupunkimaiseen kuntaan. Maaseutumaisten kuntien osalta vaikutus on taajaan asuttuja kuntia

isompi eli siellä estimoitu riski työpaikan täyttymiselle on 20 % korkeampi kuin kaupungeissa kun se taajaan asutuissa on vain 7 % korkeampi.

Estimoitujen uhkasuhteiden 95 % luottamusväleistä huomataan, että maaseutu-
maisten kuntien luottamusväli tukee selvemmin kyseisen kuntatyyppin estimoitua
riskiä eli luottamusväli on suurempi tai yhtä suuri kuin 1, kun taas taajaan asut-
tujen kuntien kohdalla luottamusväli asettuu arvon 1 molemmille puolille.

Työn keston ja kuntatyyppin välisten interaktioiden estimoiduista uhkasuhteis-
ta maaseudun vaikutus erottuu jälleen taajaan asuttuja kuntia voimakkaampana.
Interaktioiden yllä 3 kuukautta kestävä työn keston ja kuntatyyppin yhteisvai-
kutusta vähentää estimoitua riskiä avoimen työpaikan täyttymiselle.

Mallioletusten voimassaoloa arvioitiin Schoenfeldin skaalatuilla residuaaleilla ja
zph-testillä. Taulukossa 12 on esitetty R:n `cox.zph()`-funktion tuloste vain palkka-
työt sisältävään dataan sovitetulle mallille.

	Testisuure	Vapausasteet	<i>p</i> -arvo
Kuntatyyppi	2.51	2	0.29
Kesto:Kuntatyyppi	0.39	2	0.82
GLOBAL	4.91	4	0.30

Taulukko 12: R:n `cox.zph()`-funktion tuloste vain palkkatyöt sisältävään dataan sovitetulle mallille.

zph-testin tulos ei viittaa siihen, että vain palkkatyöt sisältävään dataan sovitet-
tu malli poikkeaisi merkittävästi verrannollisuusoletuksesta. Mallia sovitettaessa
kuitenkin huomattiin, että jollei mallia ositeta työn keston mukaan, verrannolli-
susoletus ei enää päde zph-testin mukaan. Toisaalta työn keston luokat olisi voitu
valita ositukseen myös toisin, ja mallioletukset olisivat silti olleet zph-testin mu-
kaan voimassa.

6 Yhteenveto

Julkisen hallinnon ja maanpuolustuksen avointen työpaikkojen analyysissä sovitettiin työn kesto- ja kuntatyyppi-muuttujat sisältävä Coxin regressiomalli sekä koko aineistoon että työpaikan tyyppin (palkka, vuorotteluvapaa) mukaan rajattuihin aineistoihin. Mallioletusten paikkaansapitävyyttä arvioitiin pääasiassa skaalattuihin Schoenfeldin residuaaleihin perustuvan zph-testin avulla. Lisäksi työpaikkojen täyttöprosessin tehokkuutta arvioitiin Kaplanin-Meierin menetelmällä laskettujen välttöfunktion estimaattien avulla.

Koko aineiston kuvailevassa tarkastelussa havaittiin, että työpaikan tyyppin mukaan estimoitu välttöfunktio on huomattavan erilainen palkkatöille ja vuorotteluvapaille. Vuorotteluvapaissa yli puolet työpaikoista täyttyy heti työpaikkailmoituksen avaamisen jälkeen, ja vain vajaa 10 % vuorotteluvapaista on edelleen täyttämättä, kun ilmoituksen avaamisesta on kulunut noin 40 päivää. Sen sijaan palkkatöissä työpaikkojen täyttöprosessi on tehottomampi. Heti työpaikkailmoituksen avaamisen jälkeen täyttyy alle 10 % työpaikoista, ja 40 päivän kuluttua ilmoituksen avaamisesta yli 20 % työpaikoista on edelleen avoinna.

Myös taulukossa 13 esitetyt seuruaajan eli työpaikan avoinnaolon keston mediaanit ja keskiarvot kertovat palkkatöiden ja vuorotteluvapaiden erilaisuudesta ja kauman sijainnin suhteen.

	Mediaani	Keskiarvo
Vuorotteluvapaa	1	9
Palkka	17	30
Kaikki työpaikat	7	18

Taulukko 13: Työpaikan avoinnaolon keston tunnusluvut julkisen hallinnon ja maanpuolustuksen työpaikoissa vuonna 2018.

Työpaikan tyyppiä päädyttiinkin käyttämään ositustekijänä koko aineistoon sovitetussa mallissa. Vertailun vuoksi Coxin regressiomalli sovitettiin myös erikseen palkka- ja vuorotteluvapaat sisältäviin datoihin.

Koko aineistosta estimoiduista välttöfunktioista havaittiin, että työpaikkojen täyttöprosessi on tehokkain 3-6 kuukauden mittaisissa työsuhteissa sekä kaupunkimaisissa kunnissa sijaitsevilla työpaikoilla.

Koko aineistoon sovitetun Coxin regressiomallin estimoiduista $\hat{\beta}$ -kertoimista ja uhkasuhteista havaittiin, että avoimen työpaikan sijainti maaseutumaisessa kunnassa alentaa estimoitua riskiä työpaikan täyttymiselle lähes 30 % verrattuna siihen, että

työpaikka sijaitisi kaupunkimaisessa kunnassa. Samansuuntainen vaikutus havaittiin sijainnilla taajaan asutussa kunnassa, mutta 95 % luottamusvälin perusteella estimoitu riski työpaikan täyttymiselle on taajaan asutussa kunnassa lähellä kaupunkimaista kuntaa.

Myös vain joko vuorotteluvapaa- tai palkkatöitä sisältäviin datoihin sovitettujen Coxin regressiomallien estimoiduissa $\hat{\beta}$ -kertoimissa ja uhkasuhteissa maaseutumaiden kuntien vaikutus nousi esiin selvemmin kuin taajaan asuttujen kuntien. Vuorotteluvapaissa työpaikan sijainti maaseudulla pienentää estimoitua riskiä työpaikan täyttymiselle, palkkatöissä sijainti maaseudulla sen sijaan kasvattaa riskiä työpaikan täyttymiselle. Tämä havaittiin myös koko aineistoon sovitetun mallin työpaikan tyyppin ja kuntatyyppin estimoiduista interaktiokertoimista.

Julkisen hallinnon ja maanpuolustuksen avointen työpaikkojen täytössä työpaikan sijainti taajaan asutussa kunnassa ei noussut missään aineistoon sovitetussa mallissa voimakkaana työpaikan täyttöön vaikuttavana tekijänä esille. Voidaankin todeta, että taajaan asutuissa kunnissa avoimien työpaikkojen täytön tehokkuus on hyvin lähellä kaupunkimaisia kuntia.

Työn kestolla havaittiin olevan vaikutusta avoimen työpaikan täytön estimoituun riskiin. Työn kesto ei kuitenkaan kaikilla luokkajaoilla noudattanut Coxin regressiomallin verrannollisuusoletusta, ja siksi sovitetuissa malleissa käytettiin erilaisia luokkajakoja ja viitetasoja työn kestolle. Vain palkkatyöt sisältävään dataan sovitetussa mallissa työn kestoa käytettiin myös osittavana tekijänä.

Koko aineistoon sovitetussa mallissa yli 12 kuukautta kestävässä työsuhteissa estimoitu riski työpaikan täyttymiselle oli matalampi kuin alle vuoden mittaisissa työsuhteissa. Vuorotteluvapaissa taas havaittiin, että lyhyissä alle 3 kuukauden mittaisissa työsuhteissa estimoitu riski työpaikan täyttymiselle on aavistuksen korkeampi kuin yli 3 kuukauden työsuhteissa. Tähän estimaattiin sisältyy kuitenkin paljon epävarmuutta, joten vuorotteluvapaiden osalta työn keston vaikutuksesta avointen työpaikkojen täyttöön ei saatu kovinkaan vakuuttavaa näyttöä.

Vain palkkatyöt sisältävään dataan sovitetussa Coxin regressiomallissa työn kesto oli mukana osittavana tekijänä, ja myös sen interaktiota kuntatyyppin kanssa tutkittiin. Palkkatöissä yli 3 kuukauden mittaisissa, maaseudulla sijaitsevista työpaikoissa täytön riski estimoitiin matalammaksi kuin alle 3 kuukauden mittaisissa, kaupungissa sijaitsevista työpaikoissa. Taajaan asuttujen kuntien vaikutus jäi tässäkin tarkastelussa epävarmemmaksi, vaikkakin samansuuntaiseksi maaseutumaiden kuntien kanssa.

7 Pohdinta

Tutkielmassa käytettyyn aineistoon tehtiin paljon rajauksia, ja erityisesti ns. hakupaikkojen pois jättäminen tarkoitti, että kaikista analyysin kohteeksi valitun toimialan vuoden 2018 avoimista työpaikkailmoituksista vain 7 % sisältyi analysoitavaan aineistoon. Räisänen (2016) mainitsee, että myös hakupaikkoihin eli työpaikkoihin, joissa TE-toimisto yhdessä työnantajan kanssa päättää työpaikkailmoituksen avoinnaolon keston, voidaan soveltaa samoja analyysimenetelmiä kuin niihin työpaikkoihin, joissa avoinnaolon kesto määräytyy vapaasti markkinoilla. Työn keston ja sijainnin vaikutusta avoinnaoloon voitaisiin siis tutkia myös hakupaikkojen osalta, jolloin analyysin tulokset antaisivat kattavamman kuvan koko julkisen hallinnon ja maanpuolustuksen toimialan avointen työpaikkojen täytöstä.

Käytetty aineisto sisältää vain vuoden 2018 rekisteriaineiston. Tietoarkisto Ailassa on kuitenkin saatavilla vastaavat vuosikohtaiset aineistot myös vuosilta 2008-2017. Ne yhdistämällä työn keston ja työpaikan sijainnin vaikutuksia avointen työpaikkojen täyttöön voitaisiin tutkia pidemmältä aikajaksolta.

Räisänen (2016) on omassa analyysissään tutkinut työtarjousten vaikutusta avointen työpaikkojen täytön tehokkuuteen. Tutkielmassa käytettyyn rekisteriaineistoon sisältyy muuttuja (Q18), jossa kerrotaan työhönsoitusten määrä kullekin työpaikkailmoitukselle. Työtarjousten vaikutusta oltaisiin siis voitu tutkia myös tässä työssä työhönsoitusten määrä -muuttujan kautta esimerkiksi siten, että siitä oltaisiin muodostettu luokiteltu, dikotominen muuttuja. Toisaalta täyttötietomuuttujaan (Q15) sisältyy täytetty työtarjouksesta -luokka, jota olisi voitu hyödyntää kilpailevana tapahtuvana.

Aineisto, josta tutkielman data rajattiin, sisältää niin paljon havaintoyksiköitä ja erilaisia muuttujia, että sen pohjalta voitaisiin tehdä monenlaisia analyyseja eri menetelmiä käyttämällä. Oma työni on hyvin suppea ja rajattu verrattuna esimerkiksi Räisäsen (2016) analyysiin. Elinaika-analyysin tutkimusmenetelmät havaittiin kuitenkin myös tässä työssä käyttökelpoisiksi, vaikka niiden sovelluskohdeena oli hyvin erilainen ainepiiri kuin menetelmien tavanomainen käyttökohde epidemiologia.

LÄHTEET

- Bailey, K. R. (1983). The asymptotic joint distribution of regression and survival parameter estimates in the Cox regression model. *The Annals of Statistics*, 11(1):39–48.
- Cole, S. R., Hudgens, M. G., Brookhart, M. A., ja Westreich, D. (2015). Risk. *American Journal of Epidemiology*, 181(4):246–250.
- Collett, D. (2015). *Modelling Survival Sata in Medical Research*. Chapman & Hall/CRC, Boca Raton, Florida, 3. painos.
- Cox, D. (1975). Partial likelihood. *Biometrika*, 62(2):269–276.
- Cox, D. R. (1972). Regression models and life-tables (with discussion). *Journal of the Royal Statistical Society. Series B (Methodological)*, 34(2):187–220.
- Cox, D. R. ja Oakes, D. (1984). *Analysis of Survival Data*. Chapman and Hall, London.
- Fan, J. ja Koul, H. (2006). *Frontiers in statistics: Dedicated to Peter John Bickel in honor of his 65th birthday*. <https://search-ebshost-com.pc124152.oulu.fi:9443/login.aspx?direct=true&db=e000xww&AN=516734&site=ehost-live>. Viitattu 20.06.2021.
- Grambsch, P. M. ja Therneau, T. M. (1994). Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika*, 81(3):515–526.
- Kalbfleisch, J. ja Prentice, R. (1973). Marginal likelihoods based on Cox’s regression and life model. *Biometrika*, 60(2):267–278.
- Kalbfleisch, J. G. (1985). *Probability and Statistical Inference. Vol. 1, Probability*. Springer, New York, 2. painos.
- Lawless, J. (2003). *Statistical Models and Methods for Lifetime Data*. John Wiley and Sons, Inc., Hoboken, New Jersey, 2. painos.
- Läärä, E. (2018). *Epidemiologian tilastolliset menetelmät -kurssimateriaali*. Oulun yliopisto.
- Luokitusavain (2018). Vuoden 2018 kuntien ja NUTS 1-3, Suomi hierarkia 2018:n (virallinen NUTS 2016) välinen luokitusavain. https://www.tilastokeskus.fi/fi/luokitukset/corrmaps/kunta_1_20180101%23nuts_2_20180101/. Viitattu 20.06.2021.
- Morabia, A. (2013). Epidemiology’s 350th anniversary: 1662-2012. *Epidemiology*, 24(2):179–183.

NUTS-aluejako (2021). *NUTS-aluejako*. https://www.stat.fi/meta/kas/nuts_aluejako.html. Viitattu 20.06.2021.

Räisänen, H. (2016). *Saavutettiin kylläntymispiste? Työtarjousten lisätyön käytön vaikutus avointen työpaikkojen täyttöön ja rekrytoinnin keskeisyyden vuonna 2015. TEM-analyysijä 70/2016*. <http://urn.fi/URN:ISBN:978-952-327-106-7>. Viitattu 20.06.2021.

Robertson, H. ja Allison, D. (2012). A novel generalized normal distribution for human longevity and other negatively skewed data. *PLoS ONE*, 7(5).

Schoenfeld, D. (1982). Partial residuals for the proportional hazards regression model. *Biometrika*, 69(1):239–241.

Tilastollinen kuntaryhmitys (2011). *Tilastollinen kuntaryhmitys*. https://www.stat.fi/tup/msind/msind_aluetyypit.html. Viitattu 20.06.2021.

Toimialaluokitus (2008). *Toimialaluokitus TOL 2008*. <https://www2.tilastokeskus.fi/fi/luokitukset/toimiala/>. Viitattu 20.06.2021.

Tsiatis, A. A. (1981). A large sample study of Cox's regression model. *The Annals of Statistics*, 9(1):93–108.

Vuorotteluvapaalaki (2002). *Vuorotteluvapaalaki*. <https://www.finlex.fi/fi/laki/ajantasa/2002/20021305>. Viitattu 20.06.2021.

Aineistolähde

Työ- ja elinkeinoministeriö (2019). *Julkiseen työväilytykseen ilmoitettujen avointen työpaikkojen rekisteriaineisto 2018* [sähköinen tietoaaineisto]. Versio 1.0 (2019-04-29). Yhteiskuntatieteellinen tietoaarkisto [jakaja]. <http://urn.fi/urn:nbn:fi:fsd:T-FSD3358>. Viitattu 20.06.2021.

LIITE: Rekisteriaineiston muuttajat

Rekisteriaineistoon sisältyvät seuraavat muuttajat [suluissa muuttujan koodi]:

- Aineistoversio (lisätty FSD:ssä) [FSD_VR]
- Havaintotunnus (lisätty FSD:ssä) [FSD_ID]
- Työ- ja elinkeinotoimisto [Q1]
- Työpaikan sijaintikunta [Q2]
- Työpaikan ammatti [Q3]
- Työnantajan toimiala [Q4]
- Työpaikan koulutusvaatimus [Q5]
- Työaika [Q6]
- Työn kesto [Q7]
- Henkilökunnan koko [Q8]
- Työnantajan sektori [Q9]
- Työpaikan tyyppi [Q10]
- Työpaikan viimeinen hakupäivämäärä [Q11]
- Aiottu työn alkamispäivä [Q12]
- Avoinnaolon alkamispäivä [Q13]
- Avoinnaolon päättymispäivä (päivämäärä 11.11.2222 = paikka edelleen avoinna rekisteriaineistoa poimittaessa) [Q14]
- Täyttötieto [Q15]
- Työpaikkojen määrä ilmoituksessa [Q16]
- Työpaikkailmoitusrivien alkuperäinen määrä ilmoituksessa [Q17]
- Työhönosoitusten määrä [Q18]
- ELY-keskus (Elinkeino-, liikenne- ja ympäristökeskus) [Q19]
- Avoinnaolon kesto [Q20]
- Avoinnaolon kesto (maksimiksi laskettu 365 päivää sitä pidemmillekin havainnoille) [Q21]

- Rekrytointiaika (= työn aiotun alkamispäivän ja avoinnaolon alkamispäivän erotus) [Q22]
- Rekrytointiaika (maksimiksi laskettu 365 päivää sitä pidemmillekin havainnoille) [Q23]
- Täyttöviive (avoinnaolon päättymispäivän ja aiotun työn alkamispäivän erotus + 1, jossa täyttöpäivä > aiottu työn alkamispäivä) [Q24]
- Täyttöviive (maksimiksi laskettu 365 päivää sitä pidemmillekin havainnoille) [Q25]
- Viivekorjattu rekrytointiaika [Q26]
- Viivekorjattu rekrytointiaika (maksimiksi laskettu 365 päivää sitä pidemmillekin havainnoille) [Q27]
- Hakupaikat (Hakupaikat sisältävät mm. julkiset virat ja muut sellaiset paikat, joille työnantaja on asettanut ennakoita viimeisen hakupäivämäärän.) [Q28]
- Avoinnaolon kesto, pl. hakupaikat (maksimiksi laskettu 365 päivää sitä pidemmillekin havainnoille) [Q29]
- Rekrytointiaika, pl. hakupaikat (maksimiksi laskettu 365 päivää sitä pidemmillekin havainnoille) [Q30]
- Täyttöviive, pl. hakupaikat (maksimiksi laskettu 365 päivää sitä pidemmillekin havainnoille) [Q31]
- 0-korjattu täyttöviive, pl. hakupaikat [Q32]
- Viivekorjattu rekrytointiaika, pl. hakupaikat [Q33]
- Viivekorjattu rekrytointiaika, viivepaikat [Q34]
- Työhönsoitusten määrä ilmoitusrivimäärää kohden [Q35]
- Seutukunta [Q36]
- Tilastollinen kuntatyyppi [Q37]
- Suurkaupunki [Q38]