

Maisteriuraseuranta-aineistojen analyysi logistisen regression avulla

Tuukka Kangas

Helsingin yliopisto

Valtiotieteellinen tiedekunta

Tilastotiede

Pro gradu -tutkielma

Toukokuu 2018



HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

MATEMAATTIS-LUONNONTIETEELLINEN TIEDEKUNTA
MATEMATISK-NATURVETENSKAPLIGA FAKULTETEN
FACULTY OF SCIENCE

Tiedekunta – Fakultet – Faculty		Koulutusohjelma – Utbildningsprogram – Degree programme	
Valtiotieteellinen		Tilastotiede	
Tekijä – Författare – Author			
Tuukka Kangas			
Työn nimi – Arbetets titel – Title			
Maisteriuraseuranta-aineistojen analyysi logistisen regression avulla			
Työn laji – Arbetets art – Level	Aika – Datum – Month and year	Sivumäärä – Sidoantal – Number of pages	
Pro gradu -tutkielma	Toukokuu 2018	67 + 16 sivua liitteitä	
Tiivistelmä – Referat – Abstract			
<p>Maistereiden uraseuranta on vuonna 2005 alkanut kyselytutkimus. Kyselytutkimuksen kohdejoukkona ovat viisi vuotta aikaisemmin valmistuneet ylemmän korkeakoulututkinnon tai päättävän alemman korkeakoulututkinnon suorittaneet henkilöt. Kysely toteutetaan kokonaistutkimuksena. Kyselytutkimusta koordinoi yliopistojen työelämä- ja urapalveluiden Aarresaari-verkosto. Helsingin yliopisto on ollut mukana kyselytutkimuksissa alusta saakka. Vuonna 2017 maisteriuraseurantoja esitettiin osaksi yliopistojen rahoitusmallia. Tämä on lisännyt tutkittavan aineiston kiinnostavuutta sekä valtakunnallisesti että yliopiston sisällä. Koska kyselytutkimus on muuttunut vuosien varrella, tässä tutkielmassa on keskitytty vain vuonna 2007, 2009 ja 2011 valmistuneiden aineistoihin. Nämä kyselyt on toteutettu vuosina 2012, 2014 ja 2016.</p> <p>Tutkittava kyselytutkimus on suhteellisen laaja ja sisältää useita muuttujia. Tutkielmassa on hyödynnetty binääristä ja multinomiaalista logistista regressiota. Vastemuuttujaksi valittiin <i>kuinka tyytyväinen vastaaja oli kokonaisuudessaan suorittamaansa tutkintoonsa työuran kannalta</i>. Tämä on myös yksi muuttujista, jota on esitetty osaksi rahoitusmallia. Muita muuttujia hyödynnettiin selittävinä muuttujina. Lopputuloksena saatiin kaksi tilastollista mallia, joista ensimmäinen oli kolmen aineiston yhdistelmästä tehty malli ja toinen vuonna 2011 valmistuneiden aineistosta tehty malli. Mallit analysoitiin hyödyntäen eri tapoja, kuten ROC-käyriä ja Hosmer-Lemeshow'n testiä.</p> <p>Yhdistelmäaineiston perusteella tutkintotyytyväisyyden todennäköisyyttä kasvattaa muun muassa, jos pystyy hyödyntämään oppimaansa nykyisessä työssään, työpaikan vaatavuustaso vastaa koulutusta ja ei ole kohdannut työttömyyttä valmistumisen jälkeen. Samoin jos vastaaja kokee, että yliopisto-opiskelu on kehittänyt hyvin esimerkiksi teoreettista osaamista, opetustaitoja ja tietoteknisiä taitoja, niin tämä parantaa tyytyväisyyden todennäköisyyttä. Yhdistelmäaineiston perusteella sellaiset henkilöt, jotka pätevoityvät suoraan johonkin ammattiin, kuten lääkäriksi, lakimieheksi tai farmaseutiksi, ovat hieman todennäköisemmin tyytyväisempiä. Vastaavasti vuonna 2011 valmistuneiden aineistosta ns. generalistialoilta valmistuneet henkilöt ovat hieman tyytyväisempiä suhteessa ammattialoihin. Uusin aineisto tukee suurelta osin yhdistelmäaineiston havaintoja, mutta sen perusteella merkittävämpiä muuttujia tyytyväisyyden todennäköisyyden selittämiseksi ovat riittävien valmiuksien saaminen työelämään ja että työnantaja arvostaa tutkintoa. Näitä kahta muuttujaa ei ole käytettävissä aikaisemmissa aineistoissa.</p>			
Avainsanat – Nyckelord – Keywords			
uraseuranta, logistinen regressio, tutkintotyytyväisyys, tilastotiede, kyselytutkimus			
Säilytyspaikka – Förvaringställe – Where deposited			
Muita tietoja – Övriga uppgifter – Additional information			

Sisältö

1	Johdanto	4
2	Logistinen regressio	6
2.1	Lineaarinen malli	6
2.2	Yleistetty lineaarinen malli	7
2.2.1	Määrittely	7
2.2.2	Uskottavuusfunktiot	8
2.2.3	Linkkifunktiot	10
2.3	Logistinen regressio	10
2.3.1	Logistisen regression parametrien määrittely	11
2.3.2	Uskottavuusfunktio ja parametriestimointi	13
2.4	Devianssi	15
2.5	Waldin testisuure	16
2.6	Regressiomallin diagnostiikka	16
2.6.1	Mallin jäännösvirhe	17
2.6.2	Mallin ennustavuus	18
2.7	Multinomiaalinen logistinen regressio	19
2.8	Mallinvalintakriteeri	20
2.9	Usean muuttujan korrespondenssianalyysi	21

3	Opiskelijatytyväisyyden mittaus	23
3.1	Opiskelijatytyväisyyteen vaikuttavat tekijät	24
4	Aineiston kuvaus	26
4.1	Aineiston edustavuus ja laatu	28
4.2	Muuttujat ja niiden muokkaus	30
5	Tutkimusasetelma	35
6	Regressiomallit	38
6.1	Mallien rakentaminen	39
6.2	Lopullisen mallin rakentaminen ja valinta	42
6.2.1	Yhdistelmäaineisto	42
6.2.2	Vuonna 2011 valmistuneiden aineisto	46
6.2.3	Multinomiaaliset logistisen regression mallit	49
6.2.4	Mallien visualisointi	53
7	Johtopäätökset	57
8	Lähteet	64
	Liitteet	68
A	Kyselylomake	69
B	Tytyväisyys koulutusaloittain	76
C	Logistinen regressio: koulutusalat	77
D	Tutkintotytyväisyys koulutusaloittain ja työttömyyden suhteen	78

Luku 1

Johdanto

Helsingin yliopisto on mukana kansallisesti toteutettavissa uraseuranta-kyselytutkimuksissa. Näiden tutkimusten perusjoukkona ovat viisi vuotta aikaisemmin ylemmän korkeakoulututkinnon suorittaneet henkilöt. Näiden lisäksi perusjoukkoon kuuluvat lastentarhanopettajat ja farmaseutit. Uraseuranta-tutkimusten tarkoituksena on selvittää sekä valmistumisen jälkeistä uraa että tutkintotyytyväisyyttä. Tämän tutkielman tarkoituksena on hyödyntää logistista regressiota uraseuranta-aineistojen analyysissä. Erityisesti kiinnostuksen kohteena on se kuinka tyytyväinen vastaaja on suorittamaansa tutkintoon työuran kannalta. Uraseuranta-aineistojen kiinnostavuus on kasvanut viime aikoina, sillä opetus- ja kulttuuriministeriölle on ehdotettu, että ne tuotaisiin osaksi yliopistojen rahoitusmallia. Samaan aikaan yliopiston strategiassa korostetaan laadukkaita tutkintoja, joissa korostetaan työelämärelevanssia.

Tässä tutkielmassa käsitellään ensin logistisen regression teoriaa yleisemmin. Logistinen regressio johdetaan yleistetystä lineaarisesta mallista. Tämän jälkeen suoritetaan aineistolle lyhyt kuvaus sekä osa muuttujista pitää uudelleen luokitella. Tämän jälkeen määritetään tutkimusasetelma ja suoritetaan varsinainen logistinen regressioanalyysi, jonka lopputuotteena saadaan kaksi logistista regressiomallia. Toinen malleista tehdään kolmen

edellisen kyselyn yhdistelmäaineistolle. Toinen malleista rakennetaan uusimmalle, vuonna 2016 kerätylle aineistolle. Mallit rakennetaan binäärisen logistisen regression avulla, mutta analyysiä tuetaan multinomiaalisella logistisella regressiolla. Lopuksi regressiomallit analysoidaan ja niistä tehdään joitakin johtopäätöksiä.

Luku 2

Logistinen regressio

Tässä osiossa johdetaan logistisen regression malli. Johtaminen aloitetaan lineaarisesta mallista, joka on yleistetyn lineaarisen mallin erikoistapaus. Tämän jälkeen lineaarinen malli yleistetään yleistetyksi lineaariseksi malliksi. Logistinen regressio voidaan tämän jälkeen esittää yleistetyn lineaarisen mallin erikoistapauksena. Tämän jälkeen esitellään joitakin yleisempiä testejä ja tunnuslukuja. Lopuksi esitellään logistisen regression laajennos multinomiaalinen logistinen regressio sekä mallinvalintakriteerejä.

2.1 Lineaarinen malli

Oletetaan, että havainnot y_1, \dots, y_i voidaan tulkita satunnaismuuttujien Y_1, \dots, Y_i havaituiksi arvoiksi. Tavallisesti oletetaan, että satunnaismuuttujat ovat riippumattomia. Oletetaan, että satunnaismuuttujien odotusarvo $E(Y_i) = \mu_i$ riippuu ei-satunnaisista muuttujista x_i . Tällöin tilanne voidaan kuvata yhden muuttujan lineaarisena mallina: (Demaris, 2004)

$$(2.1) \quad Y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

Yhtälössä 2.1 β_0 ja β_1 ovat tuntemattomia parametreja, joille pätee $\beta_0, \beta_1 \in \mathbb{R}$. $\epsilon_1, \dots, \epsilon_i$ ovat ei-havaittavia satunnaismuuttujia, joille pätee

$$(2.2) \quad \epsilon_1, \dots, \epsilon_i \perp\!\!\!\perp \epsilon_i \sim N(0, \sigma^2)$$

Nyt $Y_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2)$.

Jos malliin 2.1 lisätään selittäviä muuttujia $p - 1$ kappaletta, saa se muodon:

$$(2.3) \quad Y_i = \beta_0 + \beta_1 x_i + \dots + \beta_p x_p + \epsilon_i$$

Yhtälö 2.3 voidaan esittää myös matriisimuodossa: (Graybill, 1976)

$$(2.4) \quad \mathbf{Y} = \boldsymbol{\beta}\mathbf{X} + \boldsymbol{\epsilon}$$

$$\begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} \begin{bmatrix} x_{11} & \dots & x_{1p} \\ \vdots & \vdots & \vdots \\ x_{n1} & \dots & x_{np} \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_p \end{bmatrix}$$

Kun käytetään matriisimuotoa, niin virhetermi noudattaa multinormaalijakaumaa eli $\boldsymbol{\epsilon} \sim N(0, \sigma^2 \mathbf{I}_n)$, jolloin $\mathbf{Y} \sim N(\boldsymbol{\beta}\mathbf{X}, \sigma^2 \mathbf{I}_n)$

2.2 Yleistetty lineaarinen malli

2.2.1 Määrittely

Kuten aikaisemmin on todettu, niin lineaarinen malli on yleistetyn lineaarisen mallin (YLM) erikoistapaus. Yleistetty lineaarinen malli perustuu kolmeen oletukseen. Oletetaan, että satunnaismuuttujat Y_1, \dots, Y_i kuuluvat samaan eksponenttiperheeseen. Satunnaismuuttujia voidaan kutsua myös *vastemuuttujiksi*. Oletetaan, että satunnaismuuttujat

ovat toisistaan riippumattomia. Viimeisenä oletuksena on, että löytyy sellainen monotoninen ja derivoituva funktio $g(\cdot)$, joka kuvaa satunnaismuuttujien odotusarvot $E(y_i) = \mu_i$ lineaariseksi ennusteeksi η_i . $g(\cdot)$:ta kutsutaan linkkifunktioksi. Lineaarinen ennuste on määritelty:

$$(2.5) \quad \eta_i = \sum_{j=1}^p \beta_j x_{ij}$$

Yhtälössä 2.5 $i = 1, \dots, n$. Yhtälön 2.5 voi myös esittää matriisimuodossa:

$$(2.6) \quad \boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta}$$

Nyt yleistetty lineaarinen malli voidaan määrittää linkkifunktion ja lineaarisen ennusteen avulla:

$$(2.7) \quad g[E(\mathbf{y})] = \mathbf{X}\boldsymbol{\beta}$$

Huomataan, että lineaarisessa mallissa linkkifunktio on identiteettifunktio eli $g(\mu_i) = \mu_i$ tai toisin ilmaisten $\mu_i = \eta_i$.

2.2.2 Uskottavuusfunktiot

Oletetaan, että satunnaisvektorin \mathbf{Y} komponentit noudattavat eksponenttiperheen jakaumaa. Tällöin tiheysfunktio on:(McCullagh & Nelder, 1989)

$$(2.8) \quad f_{\mathbf{Y}}(\mathbf{y}; \boldsymbol{\theta}, \phi) = \exp\{(\mathbf{y}\boldsymbol{\theta} - b(\boldsymbol{\theta}))/a(\boldsymbol{\theta}) + c(\mathbf{y}, \boldsymbol{\theta})\}$$

Yhtälössä 2.8 $a(\cdot)$, $b(\cdot)$ ja $c(\cdot)$ ovat tunnettuja funktioita. ϕ on dispersioparametri ja θ on luonnollinen parametri. Jos ϕ on tunnettu, niin tällöin tiheysfunktio kuvaa eksponenttiperheen mallin kanoonisella parametrilla θ . Tällöin yhtälö 2.8 voidaan ilmaista helpommin:(Agresti, 2002)

$$(2.9) \quad f_{\mathbf{Y}}(\mathbf{y}; \boldsymbol{\theta}) = a(\theta)b(\mathbf{y})\exp\{\boldsymbol{\theta}/a(\boldsymbol{\theta})\}$$

Usein $a(\phi) = \phi/w_i$, jossa w_i on tunnettu painokerroin.

Eksponenttiperheen uskottavuusfunktiot ilmaistaan tyypillisesti logaritmien avulla. Logaritminen uskottavuusfunktio määrittellään:(Agresti, 2002)

$$(2.10) \quad l(\boldsymbol{\theta}, \phi; \mathbf{y}) = \log(f_{\mathbf{Y}}(\mathbf{y}; \boldsymbol{\theta}, \phi))$$

Merkitään $l_i = \log(f_i(y_i; \theta_i, \phi))$, jolloin uskottavuusfunktio $l = \sum_i l_i$. Nyt

$$(2.11) \quad l_i = (y_i\theta_i - b(\theta_i))/a(\phi) + c(y_i, \theta)$$

Otetaan uskottavuusfunktioista θ :n suhteen ensimmäinen ja toinen derivaatta:

$$(2.12) \quad \left\{ \begin{array}{l} \frac{\partial l_i}{\partial \theta_i} = (y_i - b'(\theta_i))/a(\phi) \\ \frac{\partial^2 l_i}{\partial \theta_i^2} = -b''(\theta_i)/a(\phi) \end{array} \right.$$

$$(2.13)$$

Yleisesti tiedetään(Agresti, 2002), että uskottavuusfunktioille L pätee:

$$(2.14) \quad \left\{ \begin{array}{l} E\left(\frac{\partial L}{\partial \theta}\right) = 0 \\ -E\left(\frac{\partial^2 L}{\partial \theta^2}\right) = E\left(\frac{\partial L}{\partial \theta}\right)^2 \end{array} \right.$$

$$(2.15)$$

Kun yhdistetään ylläolevat yhtälöryhmät saadaan:

$$(2.16) \quad \begin{cases} E(y_i - b'(\theta_i))/a(\phi) = 0 \rightarrow \mu_i = E(y_i) = b'(\theta_i) \end{cases}$$

$$(2.17) \quad \begin{cases} b''/a(\phi) = E[(y_i - b'(\theta_i))/a(\phi)]^2 = \text{var}(y_i)/(a(\phi))^2 \rightarrow \text{var}(y_i) = b''(\theta_i)a(\phi) \end{cases}$$

2.2.3 Linkkifunktiot

Linkkifunktion valintaan vaikuttaa pitkälti vastemuuttujan jakauma. Jos vastemuuttuja on normaalijakautunut, käytetään linkkifunktiona identiteettifunktiota. Tällöin yleistetty lineaarinen malli on (yleinen) lineaarinen malli. Jos vastemuuttuja on lukumääräinen, niin tällöin vastemuuttuja noudattaa Poisson-jakaumaa, jolle luonnollinen parametri on $\log(\mu_i)$. Tällöin 2.7 muuntuu:

$$(2.18) \quad \log(\boldsymbol{\mu}) = \mathbf{X}\boldsymbol{\beta}$$

Jos taas vastemuuttuja on binäärinen, niin tällöin se on binomijakautunut. Tällöin luonnollinen parametri on $\log(\frac{\mu_i}{1-\mu_i})$, jolloin 2.7 muuntuu:

$$(2.19) \quad \log\left(\frac{\mu_i}{1-\mu_i}\right) = \sum_{j=1}^p \beta_j x_{ij}$$

Yhtälöä 2.19 kutsutaan logistiseksi regressioksi.

2.3 Logistinen regressio

Yhtälössä 2.19 esitetty logistinen regressio on tärkein malli, jota käytetään luokitteluasteikollisten selittävien muuttujien kanssa. (Agresti, 2002) Dikotomisen toistokokeen lopputuloksen jakauma esitetään tyypillisesti binomijakaumana. Binomijakauman tiheysfunktio on esitetty yhtälössä 2.20.

$$(2.20) \quad f(y_i; \pi_i, n_i) = \binom{n_i}{n_i y_i} \pi_i^{n_i y_i} (1 - \pi_i)^{n_i - n_i y_i}$$

2.3.1 Logistisen regression parametrien määrittely

Otetaan yksinkertaisuuden vuoksi vain vastemuuttuja Y ja yksi selittävä muuttuja X . Olkoon $\pi(x) = P(Y = 1|X = x)$, jolloin logistinen regressio malli voidaan määrittää:

$$(2.21) \quad \pi(x) = \frac{e^{\beta_0 + \beta x}}{1 + e^{\beta_0 + \beta x}}$$

Jos selittäviä muuttujia on p kappaletta, niin yhtälö 2.21 saa muodon:

$$(2.22) \quad \pi(x) = \frac{e^{\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p}}{1 + e^{\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p}}$$

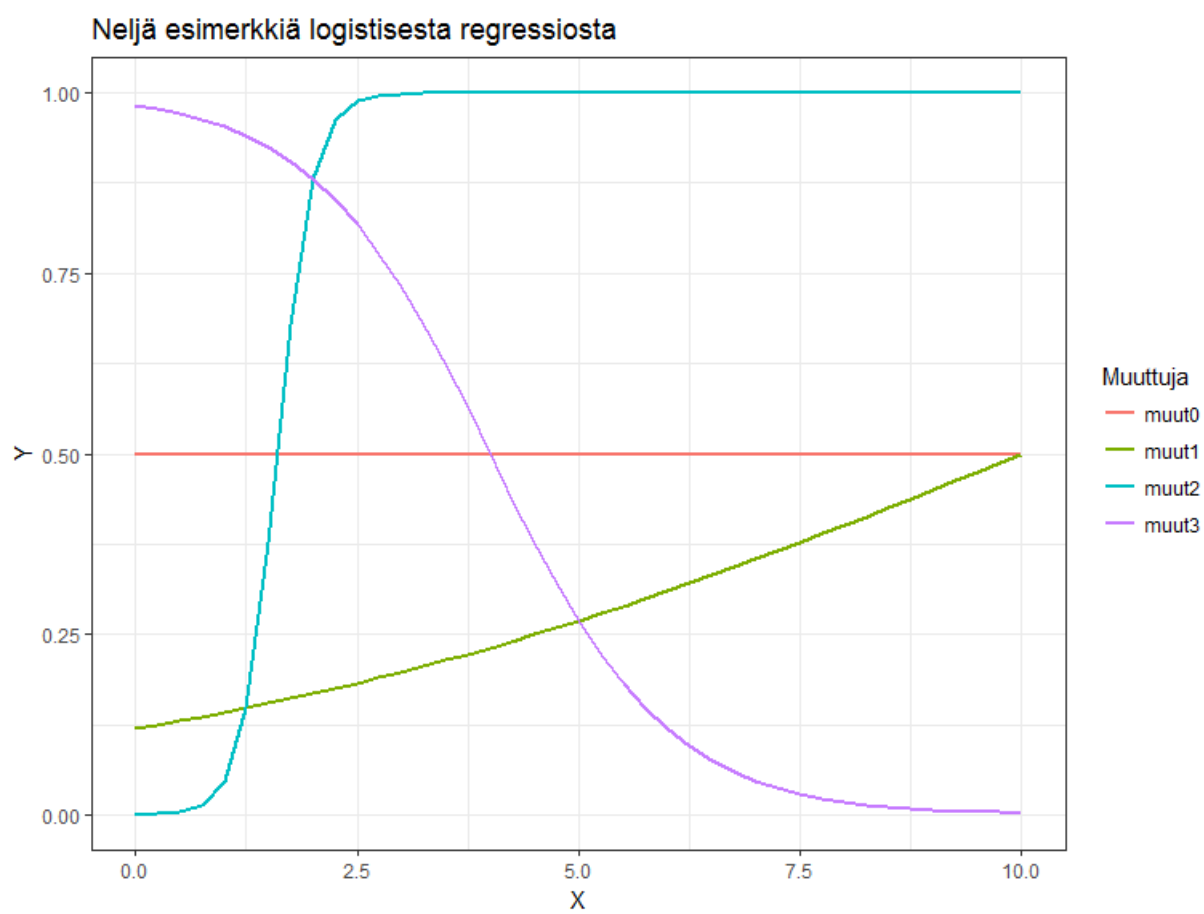
Määritellään myös termi *log odds* eli *logit* yhden muuttujan tapauksessa.

$$(2.23) \quad \text{logit}[\pi(x)] = \log\left(\frac{\pi(x)}{1 - \pi(x)}\right) = \beta_0 + \beta x$$

Suhdetta $\frac{\pi(x)}{1 - \pi(x)}$ kutsutaan myös vetosuhteeksi tai riskiksi.

Koska $\pi(x)$ kuvaa onnistumistodennäköisyyttä, niin usein mielenkiinto liittyy parametriin β : kasvaako onnistumistodennäköisyys, kun x kasvaa. Erityisesti kun $\beta = 0$, niin Y on riippumaton X :stä.

Kuvassa 2.1 on neljä esimerkkiä β :n suuruuden vaikutuksesta logistiseen regressioon. Punaisella (muut0) on tilanne, jossa $\beta = 0$. Vihreällä (muut1) on hyvin pieni, positiivinen β :n arvo, sinisellä (muut2) on suurehko β :n arvo, joka on positiivinen. Violetilla (muut3) β :n arvo on negatiivinen ja itseisarvoltaan pienempi kuin muut2.



Kuva 2.1: Neljä esimerkkiä logistisesta regressiosta eri β :n arvoilla.

2.3.2 Uskottavuusfunktio ja parametriestimointi

Binomijakauman tiheysfunktio on esitetty yhtälössä 2.20. Yleistettyjen lineaaristen mallien uskottavuusfunktioiden muodostus on esitetty aikaisemmin osiossa 2.2.2. Logistisen regression uskottavuusfunktio $l(\boldsymbol{\pi}; \mathbf{y})$ on:(McCullagh & Nelder, 1989)

$$(2.24) \quad l(\boldsymbol{\pi}; \mathbf{y}) = \sum_{i=1}^n [y_i \log\left(\frac{\pi_i}{1 - \pi_i}\right) + n_i \log(1 - \pi_i)]$$

Kuten yhtälöstä 2.24 nähdään, voidaan binomikertoimeen liittyvä termi jättää pois, jolloin uskottavuusfunktio saa yksinkertaisemman muodon. Tämä johtuu siitä, että binomikerroin ei sisällä termejä, jotka riippuisivat π :stä.

Koska yleisille lineaarisille malleille pätee osiossa 2.2.1 esitetty yhteys:

$$(2.25) \quad g(\pi_i) = \eta_i = \sum_j x_{ij} \beta_j$$

Niin voidaan yhtälössä 2.24 esitetty uskottavuusfunktio esittää parametrien β_1, \dots, β_p suhteen. Tällöin uskottavuusfunktio saa muodon:(McCullagh & Nelder, 1989)

$$(2.26) \quad l(\boldsymbol{\beta}; \mathbf{y}) = \sum_i \sum_j y_i x_{ij} \beta_j - \sum_i m_i \log(1 + \exp \sum_j x_{ij} \beta_j)$$

Uskottavuusfunktiota voidaan hyödyntää parametrien estimointiin. Estimointi voidaan suorittaa esimerkiksi Newton-Raphsonin menetelmän (NR-menetelmä, myös Newtonin menetelmä) avulla. NR-menetelmä soveltuu sellaisten ongelmien ratkaisuun, jossa etsitään funktion maksimikohtaa. Metodi on iteratiivinen, jonka alkuaskeleena on tehdä ensimmäinen arvaus. Tämän jälkeen funktio korvataan tangentillaan ja etsitään uusi arvaus maksimikohdan löytämiseksi sen avulla. Tämä toistetaan, kunnes tulos ei muutu enää merkittävästi.(McCullagh & Nelder, 1989)(Agresti, 2002)

Derivoidaan ensin yhtälö 2.24 π_i :n suhteen, jonka jälkeen voidaan hyödyntää ketjusääntöä, jolloin uskottavuusfunktio derivoidaan estimoitavan parametrin suhteen.

$$\begin{aligned}\frac{\partial l}{\partial \pi_i} &= \frac{y_i - n_i \pi_i}{\pi_i(1 - \pi_i)} \\ \frac{\partial l}{\partial \beta_r} &= \sum_{i=1}^n \frac{y_i - n_i \pi_i}{\pi_i(1 - \pi_i)} \frac{\partial \pi_i}{\partial \beta_r} \\ \frac{\partial \pi_i}{\partial \beta_r} &= \frac{d\pi_i}{d\eta_i} x_{ir} \\ \frac{\partial l}{\partial \beta_r} &= \sum_{i=1}^n \frac{y_i - n_i \pi_i}{\pi_i(1 - \pi_i)} \frac{d\pi_i}{d\eta_i} x_{ir}\end{aligned}$$

Nyt saadaan β :n Fisherin informaatioksi:(McCullagh & Nelder, 1989)

$$\begin{aligned}(2.27) \quad -E\left(\frac{\partial^2 l}{\partial \beta_r \partial \beta_s}\right) &= \sum_i \left(\frac{n_i}{\pi_i(1 - \pi_i)} \frac{\partial \pi_i}{\partial \beta_r} \frac{\partial \pi_i}{\partial \beta_s}\right) \\ &= \{\mathbf{X}^T \mathbf{W} \mathbf{X}\}_{rs}\end{aligned}$$

Matriisi \mathbf{W} on diagonaalimatriisi, joka sisältää painot.

$$(2.28) \quad \mathbf{W} = \text{diag}\left\{n_i \left(\frac{d\pi_i}{d\eta_i}\right)^2 / \pi_i(1 - \pi_i)\right\}$$

Yhtälö 2.28 supistuu muotoon:(McCullagh & Nelder, 1989)

$$(2.29) \quad \mathbf{W} = \text{diag}\{n_i \pi_i(1 - \pi_i)\}$$

Nyt saadaan matriisimuotoisen uskottavuusfunktion osittaisderivaataksi β :n suhteen:

$$(2.30) \quad \frac{\partial l}{\partial \beta} = \mathbf{X}^T (\mathbf{Y} - \boldsymbol{\mu})$$

Seuraavaksi voidaan hyödyntää Newton-Raphsonin menetelmää. Ensin annetaan ensiarvaukset $\hat{\beta}_0$, jolloin voidaan määrittää vektorit $\hat{\eta}_0$ ja $\hat{\pi}_0$. Seuraavaksi määritellään avustettu riippuvainen muuttuja (*adjusted dependent variate*) \mathbf{Z} , jolle pätee:

$$(2.31) \quad z_i = \hat{\eta}_i + \frac{y_i - n_i \hat{\pi}_i}{n_i} \frac{d\eta_i}{d\pi_i}$$

Suurimman uskottavuuden estimaateille pätee: (McCullagh & Nelder, 1989)

$$(2.32) \quad \mathbf{X}^T \mathbf{W} \mathbf{X} \hat{\boldsymbol{\beta}} = \mathbf{X}^T \mathbf{W} \mathbf{Z}$$

Yhtälö 2.32 voidaan ratkaista iteratiivisesti pienimmän neliösumman menetelmällä. Estimaatiksi saadaan:

$$(2.33) \quad \hat{\boldsymbol{\beta}}_1 = (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W} \mathbf{Z}$$

2.4 Devianssi

Devianssi on testisuure, jonka avulla voidaan tarkastella kuinka hyvin malli suhteutuu dataan. Merkitään devianssia $D(\mathbf{y}; \hat{\boldsymbol{\mu}})$. Olkoon uskottavuusfunktio $L(\boldsymbol{\mu}; \mathbf{y})$, jolloin suurimman uskottavuuden estimaatti on $L(\hat{\boldsymbol{\mu}}; \mathbf{y})$. Määritellään kyllästetty malli (*eng. saturated model*), jossa $\hat{\boldsymbol{\mu}} = \mathbf{y}$. Kyllästetyillä malleilla ei ole käytännön hyötyä, sillä ne eivät tiivistä aineistoa mitenkään, mutta tällöin saadaan täydelliset sovitearvot aineistoon. (Isotalo, 2012)

Devianssi määritellään seuraavasti:

$$(2.34) \quad D(\mathbf{y}; \hat{\boldsymbol{\mu}}) = -2 \log \frac{L(\hat{\boldsymbol{\mu}}; \mathbf{y})}{L(\mathbf{y}; \mathbf{y})}$$

Ja logaritmisten uskottavuusfunktioiden tapauksessa

$$(2.35) \quad D(\mathbf{y}; \hat{\boldsymbol{\mu}}) = 2(l(\mathbf{y}; \mathbf{y}) - l(\hat{\boldsymbol{\mu}}; \mathbf{y}))$$

Luonnollisesti devianssi ei voi saada negatiivisia arvoja. Mitä suurempi devianssi on, sitä kehnommin malli selittää dataa.(Agresti, 2015)

2.5 Waldin testisuure

Olkoon nollahypoteesi $H_0 : \beta = \beta_0$. Kun estimaatin $\hat{\beta}$ keskivirhe merkitään SE , niin Waldin testisuure z :(Agresti, 2015)

$$(2.36) \quad z = \frac{\hat{\beta} - \beta_0}{SE}$$

Kun H_0 pätee, niin z noudattaa standardinormaalijakaumaa. Toisaalta tällöin z^2 noudattaa χ^2 -jakaumaa vapausasteella 1. Waldin testisuure voidaan yleistää monimuuttujatapauksessa:(Agresti, 2015)

$$(2.37) \quad W = (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)'[\text{cov}(\hat{\boldsymbol{\beta}})]^{-1}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$$

W noudattaa χ^2 -jakaumaa vapausasteella $r(\text{cov}(\hat{\boldsymbol{\beta}}))$

Waldin testisuureta voidaan hyödyntää, kun arvioidaan selittävän muuttujan vaikutusta malliin. Pienet arvot viittaavat siihen, että kyseisillä muuttujalla ei ole tilastollista merkitsevyyttä kyseisessä mallissa.

2.6 Regressiomallin diagnostiikka

Kun regressiomalli on luotu, on syytä tarkastella kuinka hyvin malli kuvaa ilmiötä. Tarkempana tarkastelunkohteena voi olla myös mallin jäännösvirheet eli residuaalit.

2.6.1 Mallin jäännösvirhe

Kategorisilla muuttujilla jäännösvirheet voidaan määrittää vertaamalla havaittuja ja sovitettuja lukumääriä. Pearsonin residuaali e_i voidaan määrittää yhtälön 2.38 avulla. (Agresti, 2002)

$$(2.38) \quad e_i = \frac{y_i - n_i \hat{\pi}_i}{\sqrt{n_i \hat{\pi}_i (1 - \hat{\pi}_i)}}$$

Yhtälössä 2.38 $\hat{\pi}_i$ kuvaa onnistumisen todennäköisyyden suurimman uskottavuuden estimaattia, jolloin $n_i \hat{\pi}_i$ kuvaa sovitettuja lukumääriä. Residuaaleille on määritetty testisuure X^2 yhtälön 2.39 mukaan. (Agresti, 2015)

$$(2.39) \quad X^2 = \sum_{i=1}^N e_i^2$$

Kun n_i on riittävän suuri, niin X^2 noudattaa likimain χ^2 -jakaumaa. Pearsonin residuaali voidaan standardoida hattumatriisin avulla. Standardoitu Pearsonin residuaali noudattaa likimain normaalijakaumaa keskiarvolla 0 ja varianssilla 1. Pearsonin jäännösvirheen voi korvata devianssiresiduaalilla ϵ_i . Tätä varten määritetään ensin d_i , joka määritetään yhtälön 2.40 mukaisesti.

$$(2.40) \quad d_i = 2 \left(y_i \log \left(\frac{y_i}{n_i \hat{\pi}_i} \right) + (n_i - y_i) \log \left(\frac{n_i - y_i}{n_i - n_i \hat{\pi}_i} \right) \right)$$

$$(2.41) \quad \epsilon_i = \sqrt{d_i} \cdot \text{sign}(y_i - n_i \hat{\pi}_i)$$

Yhtälössä 2.41 funktio sign kääntää aina suluissa olevan erotuksen epänegatiiviseksi. Devianssiresiduaaleille voidaan muodostaa samanlainen testisuure kuin Pearsonin residuaaleille. Merkitään tätä testisuuretta G^2 , joka on määritelty samankaltaisesti kuin X^2 . (Agresti, 2002)

$$(2.42) \quad G^2 = \sum_{i=1}^N \epsilon_i^2$$

Kun testisuureiden arvot lasketaan, niin suuret arvot viittaavat siihen, että muodostettu malli kuvaa ilmiötä kehnosti.

2.6.2 Mallin ennustavuus

Lineaarisisessa regressiossa ennustavuutta voidaan kuvata suureen R^2 avulla, joka mittaa kuinka hyvin malli selittää selitettävän muuttujan varianssia. Kategorisille muuttujille ei ole pystytty luomaan täysin vastaavaa suretta. Tässä tutkielmassa esitetään kaksi tapaa mitata ennustavuutta. Ensimmäinen tapa on ROC-käyrä ja toinen on Hosmer-Lemeshow'n testi.

ROC-käyrä (*eng. ROC-curve, receiver operating characteristic curve*) mittaa mallin sensitiivisyyttä ja spesifisyyttä. Sensitiivisyys voidaan ilmaista ehdollisena todennäköisyydenä $P(\hat{y} = 1|y = 1)$ ja spesifisyys $P(\hat{y} = 0|y = 0)$. Käyrä piirretään tyypillisesti siten, että x-akselilla on arvo $(1 - \text{spesifisyys})$ ja y-akselilla spesifisyys . Mitä lähempänä käyrän alle jäävä pinta-ala on yhtä, sitä parempi mallin ennustavuus on. Jos käyrän alle jäävän alueen pinta-ala on 0,5, tällöin mallin erottelukyky on heikko. (Agresti, 2002)

Hosmer-Lemeshow'n testi perustuu kuinka hyvin ennustetut osuudet vastaavat mitattuja osuuksia eri alipopulaatioissa. Testisuure H on hyvin lähellä yhtälössä 2.38 esitettyä Pearsonin residuaalia, kuten yhtälöstä 2.43 nähdään. Yhtälössä 2.43 vastaukset on jaettu g osaan. (Agresti, 2002)

$$(2.43) \quad H = \sum_{i=1}^g \left(\frac{(\sum_j y_{ij} - \sum_j \hat{\pi}_{ij})^2}{(\sum_j \hat{\pi}_{ij})[1 - (\sum_j \hat{\pi}_{ij})/n_i]} \right)$$

Hosmer-Lemeshow'n testin nollahypoteesina on, että aineisto on yhdenmukainen, jolloin testisuure noudattaa khi toiseen jakaumaa vapausasteella $g - 2$. Jos testisuureen laskettu p-arvo on hyvin pieni, niin tällöin malli ei selitä hyvin aineistoa. (Demaris, 2004) Hosmer-Lemeshow'n testiin liittyy joitakin puutteita, sillä se ei huomaa kaikkia puutteita ennustavuudessa. (Agresti, 2002)

2.7 Multinomiaalinen logistinen regressio

Aikaisemmin tässä osiossa logistista regressiota on käsitelty binääriseen logistiseen regressioon kautta. Jos vastemuuttujalla on enemmän kuin kaksi vastausvaihtoehtoa, voidaan hyödyntää multinomiaalista logistista regressiota. Toinen vaihtoehto on supistaa vastemuuttuja binääriseksi uudelleenluokittelulla. Jos vastausvaihtoehdot eivät ole selkeässä järjestyksessä, voidaan hyödyntää baseline-category logit-mallia. Tässä hyödynnetään vertailuryhmää c . Baseline-category logit-malli voidaan ilmaista yhtälön 2.44 avulla (Agresti, 2015)

$$(2.44) \quad \text{logit}[P(Y_{ij} = 1 | y_{ij} = 1 \text{ tai } y_{ic} = 1)] = \log\left(\frac{\pi_{ij}}{\pi_{ic}}\right) = \mathbf{x}_i \boldsymbol{\beta}_j = \sum_{k=1}^p \beta_{jk} x_{ik}$$

Estimaatit voidaan määrittää suurimman uskottavuuden avulla. Uskottavuusfunktiot ovat: (Agresti, 2015)

$$(2.45) \quad \sum_{i=1}^N x_{ij} y_{ij} = \sum_{i=1}^N x_{ik} \pi_{ij}$$

Kuten binäärisessä logistisessa regressiossa, myös multinomaalisessa logistisessa regressiossa voidaan määrittää esimerkiksi devianssi G^2 .

$$(2.46) \quad G^2 = 2 \sum_{i=1}^N \sum_{j=1}^c n_i y_{ij} \log \frac{n_i y_{ij}}{n_i \hat{\pi}_{ij}}$$

Jos taas vastaukset voidaan järjestää luonnollisesti kasvavaan järjestykseen, voidaan hyödyntää kumulatiivista logistista regressiota. Tämä voidaan ilmaista yhtälön 2.47 avulla (Agresti, 2015)

$$(2.47) \quad \begin{aligned} \text{logit}[P(y_i \leq j)] &= \log\left(\frac{\pi_{i1} + \dots + \pi_{ij}}{\pi_{i,j+1} + \dots + \pi_{ic}}\right), j = 1, \dots, c - 1 \\ &= \alpha + \mathbf{x}_i\beta \end{aligned}$$

2.8 Mallinvalintakriteeri

Samaan aineistoon luodaan usein useita eri malleja. Jotta malleja voidaan verrata toisiinsa, on syytä käyttää joitakin tunnuslukuja. Erilaisia tapoja on useita, mutta tässä tutkielmassa käytetään kolmea tapaa, jotka ovat yleisesti käytettyjä metodeja. (Dziak *et al.*, 2012)

Ensimmäinen tapa on Akaiken informaatiokriteeri (*Akaike information criteria, AIC*). Akaike julkaisi tämän tunnusluvun vuonna 1974. (Akaike, 1974) Olkoon k muuttujien määrä mallissa ja \hat{L} uskottavuusfunktion maksimiarvo, niin tällöin:

$$(2.48) \quad AIC = 2k - 2\ln(\hat{L})$$

Tyypillisesti etsitään mallia, joka minimoi AIC:n. Kuten nähdään, niin AIC:ta kasvattaa muuttujien lisääminen ja pienentää mitä paremmin malli kuvaa aineistoa. (Agresti, 2015)(Agresti, 2002) Jos mallissa on liikaa muuttujia, niin tällöin ongelmaksi muodostuu se, että malli on kehnosti yleistettävissä. (Dziak *et al.*, 2012)

AIC:sta on olemassa muunnelma AICc (Corrected Akaike information criteria), jota voidaan hyödyntää, kun aineiston koko on pieni. AICc:n etu suhteessa AIC:hen on se, että AICc säilyttää asymptotiset ominaisuutensa suhteessa uskottavuusfunktion maksimi-

miarvoon paremmin kuin AIC, jos otoskoko on pieni.(Wong & Li, 1998) Olkoon otoskoko n , jolloin: (Cavanaugh, 1997)

$$(2.49) \quad AIC_c = AIC + \frac{2k(k+1)}{n-k-1}$$

Huomataan, että kun $n \rightarrow \infty$, niin $AIC_c \rightarrow AIC$.

Kolmas tapa arvioida mallin valintaa on bayesilainen informaatiokriteeri (*Bayesian information criterion, BIC*). BIC on approksimaatio bayesilaisista faktoreista.(Raftery, 1995) Tämä on hyvin lähellä kahta edellistä mallia, joskin sen lähtökohta on erilainen: BIC olettaa, että mallien joukossa on yksi oikea malli. BIC:n esitteli ensimmäisen kerran 1978 Gideon Schwarz.(Schwarz, 1978) BIC rankaisee voimakkaammin ylimääräisistä muuttujista kuin AIC tai AICc.(Agresti, 2015) BIC korjaa samalla AIC:n ongelmaa eli ylisovitusta (*overfitting*), jolloin mallissa on liikaa selittäviä muuttujia. BIC:n ongelmana voidaan pitää taas alisovitusta (*underfitting*).(Dziak *et al.*, 2012)(Kass & Raftery, 1995)

$$(2.50) \quad BIC = \ln(n)k - 2\ln(\hat{L})$$

AIC:ta, AICc:ta ja BIC:ta tulkitaan samalla tavalla eli pyritään löytämään malli, joka minimoi kyseisen arvon. Jos kahden mallin AIC:n arvojen erotus on vähemmän kuin 2, niin tällöin malleja voidaan pitää tämän mittarin suhteen yhdenvertaisina. Jos ero on yli kymmenen, niin tällöin mallien välillä on tilastollisesti merkitsevä ero.(Dziak *et al.*, 2012) Samankaltaisia raja-arvoja voidaan hyödyntää myös BIC:n kanssa.(Raftery, 1995)

2.9 Usean muuttujan korrespondenssianalyysi

Koska tässä tutkielmassa päätellään usean muuttujan vaikutusta selitettävään muuttujaan, on analyysiä syytä tukea selittävien muuttujien välisillä yhteyksillä. Sopiva tapa

tähän on korrespondenssianalyysi.(Nenadic & Greenacre, 2007)(Sourial *et al.*, 2010) Korrespondenssianalyysin lähtökohtana on usein lukumäärämatriisi N ($I \times J$). Korrespondenssimatriisi M muodostetaan siten, että $m_{ij} = n_{ij}/n$, jossa n viittaa otoskokoon. Tämän jälkeen korrespondenssimatriisiin M lasketaan rivi- ja sarakesummat, joita merkitään vektorimuodossa \mathbf{r} ja \mathbf{s} . Tämän jälkeen matriisi keskitetään ja normalisoidaan diagonaalimatriisien \mathbf{D}_r ja \mathbf{D}_s avulla. Tällöin saadaan matriisi \mathbf{S} .(Greenacre & Blasius, 2006)

$$(2.51) \quad \mathbf{S} = \mathbf{D}_r^{-1/2}(\mathbf{M} - \mathbf{r}\mathbf{s}^T)\mathbf{D}_s^{-1/2}$$

Matriisi S on lähtökohtana, kun määritetään korrespondenssianalyysissä piirrettävän hajontakuvan koordinaatit. Tätä varten määritetään vielä S uudestaan singulaariarvojen avulla. Olkoon Σ diagonaalimatriisi, joka sisältää singulaariarvot. Matriisit \mathbf{U} ja \mathbf{V} ovat järjestyksessä vasen ja oikea singulaarivektori. Tällöin S voidaan ilmaista yhtälön 2.52 avulla.(Greenacre & Blasius, 2006)

$$(2.52) \quad \mathbf{S} = \mathbf{U}\Sigma\mathbf{V}^T$$

Tarvittavat asteikot \mathbf{X} ja \mathbf{Y} voidaan ilmaista(Greenacre & Blasius, 2006)

$$(2.53) \quad \left\{ \begin{array}{l} \mathbf{X} = \mathbf{D}_r^{-1/2}\mathbf{U} \\ \mathbf{Y} = \mathbf{D}_s^{-1/2}\mathbf{V} \end{array} \right.$$

Luku 3

Opiskelijatytyväisyyden mittaus

Korkeakoulutuksen laatu on ollut kiinnostuksen kohteena erityisesti 1980-luvulta lähtien. Laadun mittaaminen voidaan jakaa kahteen osaan: parantamiseen ja varmentamiseen. Näistä jälkimmäinen on historiallisesti merkittävämpi, sillä korkeakouluihin on kohdistunut laadunvalvontapaineita. Tämä on johtanut siihen, että koulutussektori on nähty palvelualana. Samoin korkeakoulut ovat tyypillisesti vain pyrkineet täyttämään muualta määrättyt kriteerit ja opiskelijoiden toiveet. On kuitenkin myös ilmeisiä tavoitteita parantaa korkeakoulutuksen laatua. Tässä parannustyössä on tärkeää hyödyntää opiskelijoiden ja valmistuneiden henkilöiden palautetta.(Nair *et al.*, 2011)

Vuonna 1991 esitettiin ajatus, että opiskelija on asiakas. Tämä ajattelutapa ei ole ongelmaton, sillä opiskelijalla on muitakin rooleja korkeakouluissa: esimerkiksi opiskelija on oppimismielessä yhteistyökumppani. Tällöin on oleellista erottaa palvelun laatu ja asiakastytyväisyys. Näiden kahden yhteydestä on erilaisia konsepteja. Yksi suosituista tavoista on ajatella asiakastytyväisyyden olevan palvelun laadun esiaste, kun taas toisen ajatusmallin ajatuksena on, että palvelun laatu johtaa asiakastytyväisyyteen.(Douglas *et al.*, 2008)

3.1 Opiskelijatyytyväisyyteen vaikuttavat tekijät

Useassa tutkimuksessa on pystytty osoittamaan, että oppimis- ja opettamisympäristöllä on vahva yhteys tyytyväisyyteen. (Nair *et al.*, 2011) Samoin on pystytty osoittamaan, että opetuksen hyödyllisyys johtaa tyytyväisyyteen, mutta sillä ei ole selkeää yhteyttä tyytymättömyyteen. (Douglas *et al.*, 2008)

Australiassa on huomattu, että valmistuneilla on tavallisesti hyvät käytännön taidot, mutta heidän kriittisen ajattelun, ryhmätyö- ja ongelmanratkaisutaidot eivät ole kehittyneet riittävästi opintojen aikana. (Nair *et al.*, 2011) Taloustieteilijöille toteutetussa tutkimuksessa huomattiin, että tärkeimmät yleiset taidot olivat kirjallinen osaaminen sekä esiintymistäidot. Samassa tutkimuksessa todetaan, että yleiset taidot eivät riitä, vaan tämän lisäksi pitää olla vahva oman alan substanssiosaaminen. (Andrews & Higson, 2008) Suomalaisille vastavalmistuneille tehdystä kyselytutkimuksesta, jossa tutkittiin työllistyvyyttä, osoitettiin, että yleiset taidot olivat työllistyvyydelle hieman tärkeämpiä kuin ammatilliset taidot. Tutkimuksessa eriteltyjä taitoja olivat muun muassa ongelmanratkaisutaidot, sosiaaliset taidot, yhteistyö- ja ryhmätaidot. (Puhakka *et al.*, 2010) Alankomaalaisessa tutkimuksessa on tutkittu yleisien ja alalle spesifien taitojen yhteyttä työelämän vaatimukseen. Tässä tutkimuksessa osoitettiin, että spesifeillä taidoilla työllistettiin koulutusta vastaavaan työhön. Korkeamman tason työpaikoissa korostui yleisten työelämätaitojen merkitys. Erityisesti ryhmätyötaitojen merkitys korostui. (Semeijn *et al.*, 2006)

Euroopan laajuisessa tutkimuksessa on tutkittu miten koulutusala vaikuttaa uratyytyväisyyteen uran alkuvaiheessa. Tässä tutkimuksessa korostetaan, että alan valinta on monesti mieltymyskysymys. Tutkimuksessa on kuitenkin huomattu, että kun työuraansa aloittava työskentelee työssä, jonka vaatimustaso vastaa koulutustasoa, on hän myös tyytyväisempiä työuraansa. Vastaavasti jos vastaajat kokivat, että he olivat ylipäteviä työhönsä, aiheutti tämä tyytymättömyyttä. Samassa tutkimuksessa havaittiin, että jos

vastaaaja sai hyödyntää oppimiaan taitoja ja tietoja, kasvoi tyytyväisyys. Kun työhön liittyvät muut muuttajat kuin koulutusala oli vakioitu, ei koulutusalojen välillä huomattu eroa tyytyväisyyteen.(Vila *et al.*, 2007)

Sukupuoli on usein kiinnostava muuttuja, kun pyritään löytämään ryhmien välisiä eroja. Australiassa tehdyssä tutkimuksessa on huomattu, että naiset ovat vaativampia opintojen laadun suhteen kuin miehet. Grebennikovin ja Skainesin tutkimuksessa on pyritty selittämään tätä muun muassa alempiarvoisilla töillä ja kehnommalla palkalla(Grebennikov & Skaines, 2007)

Luku 4

Aineiston kuvaus

Korkeakoulujen uraseurantakyselyt ovat vuosittain järjestettäviä kyselytutkimuksia, joissa selvitetään valmistuneiden henkilöiden urapolkua sekä heidän näkemystään saamastaan koulutuksesta. Suomalaisissa yliopistoissa, kuten Helsingin yliopistossa, on yleisesti käytössä kaksi uraseurantaa, joista toinen on suunnattu maistereille ja toinen tohtoreille. Tässä tutkielmassa keskitytään vain maisteriuraseurantaan. Maisteriuraseurannan perusjoukkoon kuuluvat he, jotka ovat suorittaneet joko maisterin tutkinnon, päättyvän alemman korkeakoulututkinnon tai päättyvän lisensiaatin tutkinnon. Helsingin yliopistossa päättyviä alempia korkeakoulututkintoja ovat lastentarhanopettajan ja farmaseutin tutkinnot. Samoin lääketieteen, hammaslääketieteen ja eläinlääketieteen lisensiaatit kuuluvat maisteriuraseurannan perusjoukkoon. Kysely suoritetaan viisi vuotta valmistumisen jälkeen: viimeisin aineisto kerättiin syksyllä 2016, jolloin perusjoukkona oli vuonna 2011 valmistuneet. Ensimmäinen maisteriuraseuranta toteutettiin vuonna 2005. Tämän jälkeen kysely on toteutettu pääsääntöisesti kahden vuoden välein. Kyselyn kehittämistä ja raportointia koordinoi yliopistojen ura- ja työelämäpalvelujen Aarresaari-verkosto. Kysely toteutetaan kokonaistutkimuksena. (OKM, 2017)

Aineisto koostuu taustamuuttujista ja kyselytutkimuksessa kysytyistä muuttujista.

Taustamuuttujia ovat muun muassa tutkintoon liittyvät tiedot, sukupuoli ja ikä valmistuessa. Kyselytutkimuksessa kysytään hieman tietoja ennen tutkinnon suorittamista. Pääpaino on kuitenkin tutkinnon arvioinnissa sekä siinä mitä vastaaja on tehnyt valmistumisen jälkeen työelämässä. Tutkintoa arvioidaan erityisesti sisällön ja sen tuottaman osaamisen kautta. Työurasta kysytään muun muassa työttömyydestä, yrittäjyydestä sekä tietoja ensimmäisestä ja nykyisestä työpaikasta. Tutkimuksessa painotetaan työllistymisen laatua. Kyselyn lopussa arvioidaan useita työelämässä keskeisiä taitoja: niiden tarvetta nykyisessä työssä ja kuinka hyvin yliopisto-opiskelu kehitti näitä taitoja. Tämän lisäksi kysytään mitkä taidot tulevat olemaan tärkeitä tulevaisuudessa. Vuonna 2016 käytetty kyselytutkimuslomake on liitteenä A.

Vuonna 2017 uraseuranta-aineistoja ehdotettiin osaksi rahoitusmallia, jonka perusteella opetus- ja kulttuuriministeriö (OKM) jakaa rahaa korkeakouluille. Vuonna 2016 asetettu työryhmä esitti, että rahoitusmalliin tuotaisiin uraseurantakyselyistä indikaattoreita, jotka kuvaavat tyytyväisyyttä tutkintoon, työn vaativuutta suhteessa koulutukseen, miten vastaaja pystyy hyödyntämään oppimaansa sekä koulutuksen antamia työelämävalmiuksia. Näistä käytetään tässä tutkielmassa nimitystä *tulosohjausmallikysymykset*.(OKM, n.d.)(Sainio *et al.*, 2017)

4.1 Aineiston edustavuus ja laatu

Aineiston edustavuutta on arvioitu tarkemmin erillisessä tutkielmassa.(Kangas, 2017) Tutkielmassa on tarkasteltu erityisesti yksikkö- ja eräkatoa sekä niiden suhdetta tulosohjausmallikysymyksiin. Yksikkökato on yleinen ongelma kyselytutkimuksissa. Vuoden 2016 aineiston perusjoukon koko oli 2403 henkilöä, joista vastauksen antoi 917. Vastauskato on tällöin noin 62 %. Yleisesti naiset ovat vastanneet kyselyihin paremmin kuin miehet eron ollessa korkeintaan muutamia prosenttiyksikköjä verrattuna perusjoukon sukupuoliosuuteen. Ikäryhmien välillä ei ole suurta eroa. Koulutusalojen välillä on merkittäviä eroja vastausalttiudessa. Esimerkiksi lääketieteen, farmasian ja oikeustieteen aloilla vastausalttius on kehnompaa kuin kasvatustieteellisellä, luonnontieteellisellä ja yhteiskuntatieteellisellä alalla. Tutkielmassa on korjattu erityisesti koulutusaloista eroa jälkiosituksella. Jälkiositus ei muuttanut vastausjakaumia kuin korkeintaan muutaman prosenttiyksikön verran, joten voidaan päätellä, että aineiston edustavuus on riittävä. Joissakin kysymyksissä, kuten onko ollut työttömänä valmistumisen jälkeen, on melko suuriakin eräkatoja, jotka kasvattavat epävarmuutta.

Taulukkoon 4.1 on koottu kolmen viimeisimmän aineiston koulutusaloittaiset vastausprosentit. Kuten taulukosta nähdään, niin vastausalttiudet poikkeavat koulutusaloittain melko paljon.

Koulutusala	2007	2009	2011
Humanistinen	49,2	47,1	37,6
Kasvatustieteellinen	51,7	47,3	43
Luonnontieteellinen	50,8	44,9	41,8
Lääketieteellinen	46,2	37,0	21,0
Oikeustieteellinen	42,6	41,5	34,5
Psykologian	55	45,5	47,6
Teologinen	50,3	43,0	42,3
Yhteiskuntatieteellinen	54,0	47,7	41,4
Farmasian	47,4	40,0	26,1
Maatalous-metsätieteellinen	52,7	44,4	38,7
Eläinlääketieteellinen	63,4	53,3	55,3
Hammaslääketieteellinen	46,9	32,3	25,9
Kaikki	50,1	44,5	38,2

Taulukko 4.1: Koulutusalaakohtaiset vastausprosentit

4.2 Muuttajat ja niiden muokkaus

Aineisto koostuu useasta kymmenestä muuttujasta. Tässä osiossa ei esitellä kaikkia muuttujia tai niiden jakaumia, vaan pääpaino on sellaisissa muuttujissa, jotka ovat potentiaalisia selittäviä muuttujia logistisessa regressiomallissa. Vuonna 2007 valmistuneista kyselyyn vastasi 1432 henkilöä, vuonna 2009 valmistuneista 794 ja vuonna 2011 valmistuneista 917. Kuten edellisessä osiossa näytettiin, vastausprosentti on laskenut melko paljon kyselyiden välillä.

Muuttujia on melko paljon ja alkutilanteena voidaan pitää sitä, että myös logistinen regressio tulee sisältämään paljon muuttujia. Osa muuttujista pitää uudelleen luokitella, jotta näistä tulisi helpommin käsiteltäviä ja malliin ei tulisi kovinkaan paljoa sellaisia muuttujaluokkia, joissa on vähän vastauksia.

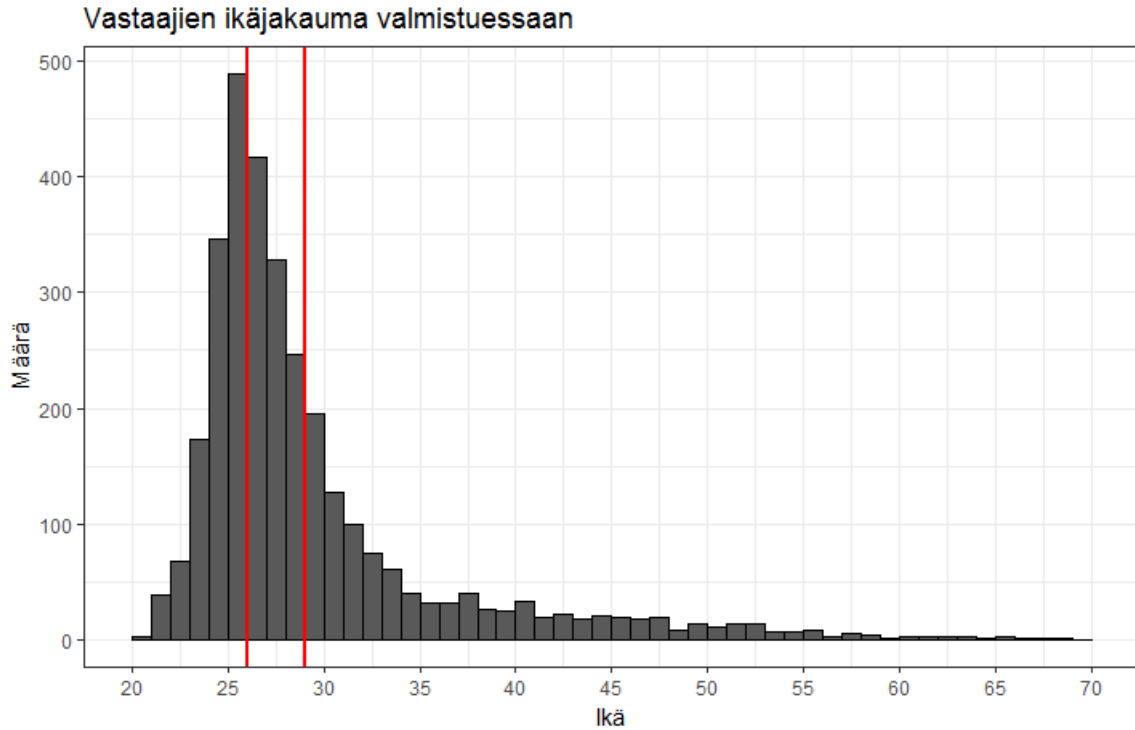
Vastaaajille on tuotu Virta-järjestelmästä taustamuuttujat. Näitä hyödynnetään myöhemmin mallin rakentamisessa. Sukupuolien osuudet ovat pysyneet likimain samoina eri kyselyjen välillä. Taulukkoon 4.2 on koottu sukupuolien jakaumat.

Valmistumisvuosi	Miehet [%]	Naiset [%]
2007	25,0	75,0
2009	23,2	76,8
2011	28,2	71,8

Taulukko 4.2: Sukupuolien jakaumat kolmessa uusimassa maisteriuraseuranta-aineistossa

Aineistossa on myös ikä valmistuessa. Kuvaan 4.1 on kuvattu vastaajien ikäjakauma. Vastajaat on jaettu kolmeen ikäluokkaan: alle 27-vuotiaat, 27-29-vuotiaat ja yli 30-vuotiaat. Vastajia on ryhmissä 35,6, 31,6 ja 32,7 % nousevassa järjestyksessä. Kuvaan on merkitty punaisella pystyviivalla analyysissä käytetyt ikäryhmät.

Taustamuuttujiksi voidaan laskea myös tutkimukseen liittyviä tietoja. Tutkinnosta on ra-



Kuva 4.1: Vastaajien ikäjakauma. Pystyviivoilla on merkitty luokkarajat.

portoitu muun muassa kansallinen kouluskoodi, tutkintonimike sekä Helsingin yliopiston omia organisaatiohierarkisia tietoja, kuten tiedekunta ja pääaine. Tässä tutkielmassa käytetään vain yhtä tutkintoon liittyvää muuttujaa, joka on kansallisesti ja kansainvälisesti käytetty koulutusluokitus (ISCED 2011). Taulukkoon 4.3 on koottu vastausjakaumat koulutusaloittain ja valmistumisvuosittain. Huomataan, että koulutusluokitukseen liittyi joitakin ongelmia, joista ilmeisin on, että osa ryhmistä on hyvin pieniä.

Koulutusala	2007	2009	2011
Humanistinen	18,4	19,8	17,1
Kasvatustieteellinen	15,2	14,5	14,7
Luonnontieteellinen	16,0	12,7	17,3
Lääketieteellinen	3,9	5,9	2,7
Oikeustieteellinen	7,8	9,6	8,9
Psykologian	1,5	1,3	2,2
Teologinen	6,0	5,4	5,7
Yhteiskuntatieteellinen	13,1	11,7	14,9
Farmasian	7,0	8,8	5,1
Maatalous-metsätieteellinen	8,2	5,0	8,2
Eläinlääketieteellinen	1,8	4,0	2,3
Hammaslääketieteellinen	1,1	1,3	0,8

Taulukko 4.3: Vastaaajien koulutusala jakauma valmistumisvuosittain. Solussa on kyseisestä koulutusalaista valmistuneiden osuus kyseisen vuoden valmistuneista.

Kyselyssä on kysytty jonkin verran kysymyksiä koskien ensimmäistä ja nykyistä työpaikkaa. Tässä tutkielmassa ei käsitellä ensimmäiseen työpaikkaan liittyviä kysymyksiä. Nykyisestä työstä ei käsitellä muun muassa toimialaan tai palkkaan liittyviä kysymyksiä. Aineistossa olevat palkkatiedot ovat hyvin puuttelliset. Nykyisestä työstä on otettu mukaan kysymykset *pystyykö nykyisessä työssä hyödyntämään yliopistossa opittuja asioita (oppimansa hyödyntäminen)* ja *vastaako nykyisen työn vaativuustaso koulutusta (koulutustasovastaavuus)*. Oppimansa hyödyntäminen on kysytty kolmiportaisella asteikolla *ei ole voinut hyödyntää juuri lainkaan, hyödynsi jonkin verran/osittain ja hyödynsi jatkuvasti*. Näistä kaksi ensimmäistä on luokiteltu samaan vastausluokkaan, jolloin saadaan binäärinen dummy-muuttuja. Vastaajista 62 % pystyy hyödyntämään jatkuvasti oppimaansa. Koulutustasovastaavuus on kysytty neliportaisella asteikolla *selvästi alhaisempi, osittain alhaisempi, vastaa hyvin ja vaativampaa*. Näistä kaksi viimeistä vastausta on uudelleenluokituksessa laitettu samaan luokkaan, jolloin saadaan binäärinen dummy-muuttuja. Vastaajien nykyinen työ vastaa koulutustasoa tai on vaativampaa 86 % tapauksista.

Tutkinnosta on kysytty joitakin kysymyksiä. Tarkemmassa analyysissä on mukana väite *koulutus antoi riittävät valmiudet työelämään*, johon vastaaja on voinut vastata kuusiportaisella asteikolla *täysin eri mieltä - eri mieltä - hieman eri mieltä - hieman samaa mieltä - samaa mieltä ja täysin samaa mieltä*. Tämä kysymys on ollut mukana vain viimeisimmässä, vuonna 2016 tehdyssä tutkimuksessa. Tästä muuttujasta on luotu binäärinen siten, että kaksi viimeisintä vastausvaihtoehtoa muodostavat yhden luokan ja muut toisen luokan. Tällöin noin 31 % vastaajista on täysin samaa tai samaa mieltä, että koulutus antoi riittävät valmiudet työelämään. Samoin kuin edellinen kysymys, niin myös kysymys *työnantajat arvostavat tutkintoani* on luokiteltu samalla tavalla. Yleisestä tutkintotyytyväisyydestä on oma kysymys, mutta tämä käsitellään tarkemmin luvussa 5. Tutkinnosta on kysytty tarkemmin tietoja liittyen erilaisiin työelämätaitoihin. Väite on esitetty muodossa *Miten yliopisto-opiskelu kehitti kyseisiä työelämävalmiuksia? Esimies-*

taidot. Vastusvaihtoehdot ovat *ei lainkaan- vain vähän - jonkin verran - melko paljon - paljon - erittäin paljon*. Näistä on luotu binäärinen muuttuja laittamalla kolme viimeistä vastausvaihtoehtoa yhteen luokkaan.

Muille muuttujille ei ole tehty uudelleenluokittelua. Tämä koskee muun muassa muuttujia koskien työttömyyttä ja onko suorittanut jatko-opintoja. Nämä ovat luonteeltaan muutenkin binäärisiä, eivätkä näin vaadi luokkien määrän vähentämistä.

Luku 5

Tutkimusasetelma

Urauseuranta-aineistojen kiinnostus on kasvanut, sillä OKM:lle on esitetty, että siitä tuodaan indikaattoreita yliopistojen rahoitusmalliin. Asetettu työryhmä esittää, että seuraavat kysymykset tulisivat osaksi rahoitusmallia:

Arvioi vuonna 201x suorittamaasi tutkintoa seuraavien väittämien perusteella. Valitse sopivin vaihtoehto. Koulutus antoi riittävät valmiudet työelämään.

Miten tyytyväinen olet kokonaisuudessaan vuonna 201x suorittamaasi tutkintoon työurasi kannalta?

Miten hyvin pystyt (tällä hetkellä) hyödyntämään yliopistossa oppimiasi tietoja ja taitoja?

Miten hyvin työsi vastaa (tällä hetkellä) vaativuustasoltaan yliopistollista koulutustasi.

Miten yliopisto-opiskelu kehitti kyseisiä työelämävalmiuksia?

Analyttiset, systemaattisen ajattelun taidot

Tiedonhankintataidot

Kyky oppia ja omaksua uutta

Tieteidenvälisyys/moniammatillisissa ryhmissä toimiminen

Itseohjautuvuus/oma-aloitteisuus

Yrittäjyystaidot

Viides indikaattori koostuisi kuudesta osa-alueesta, joita vastaa eri työelämätaidot. Kaikken viiden indikaattorin painoarvo on sama, joten yhden työelämävalmiuden painoarvo rahoitusmallissa olisi yksikolmaskymmenesosa. Rahoitusosa laskettaisiin absoluuttisista vastauksista, jolloin sekä vastausten määrä että laatu vaikuttaa rahanjakoon.

Koska uraseuranta-aineistot mahdollisesti tulevat muodostamaan osan yliopiston rahoituksesta, on tämä kasvattanut kiinnostusta kyseiseen aineistoon. Samoin yksi Helsingin yliopiston vuosien 2017-2020 strategian osuuksista on *opiskelijat keskiöön*, jossa painotetaan kilpailukykyisiä ja laadukkaita tutkintoja sekä tuodaan työelämä vahvemmin esille. (HY, 2016) Valitut tulosohjausmallikysymykset korostavat yliopiston omaa strategiaa, sillä niissä keskitytään laadukkaaseen tutkintoon sekä siihen onko valmistunut henkilö saanut riittävät työelämävalmiudet vaatimaan asiantuntijatyöhön.

Staddon ja Standish ovat tutkineet opiskelijoiden kokemuksen mittaamista. He ovat arvioneet, että yliopistot, jotka pystyvät tuottamaan opiskelijoiden tyytyväisyyttä kasvatavia palveluja, selviytyvät kilpailukykyisempinä verrattuna muihin yliopistoihin. He perustelevat tätä erityisesti sillä, että keho palaute heikentää heidän mainettaan suhteessa korkeakouluihin, jotka saavat hyvää palautetta. Staddon ja Standish laajentavat opiskelijatyytyväisyyden kestävästi työelämäänsä saakka. Tätä voidaan pitää hyvänä motivaationa

myös tälle tutkielmalle.(Staddon & Standish, 2012)

Koska tulosohjausmallikysymyksiä on viisi, rajattiin tämän tutkielman tarkastelu yhteen vastemuuttujaan, sillä logistinen regressio soveltuu hyvin yhden vastemuuttujan selittämiseen. Vastemuuttujaksi valittiin tutkintotyytyväisyys eli kuinka tyytyväinen vastaaja on suorittamaansa tutkintoonsa työuran kannalta. Tätä valintaa puoltaa se, että tätä on kysytty useana vuonna toisin kuin kysymystä *riittävät valmiudet työelämään*. Kaksi muuta yksittäistä kysymystä soveltuvat paremmin selittäviksi tekijöiksi eivätkä suoraan keskity tutkintoon ja näin ollen ne eivät ole yhtä selkeästi yliopiston hallittavissa olevia muuttujia. Tutkintotyytyväisyyttä on kysytty kyselyssä kuusiportaisella asteikolla. Nousevassa järjestyksessä vastausvaihtoehdot ovat: *erittäin tyytymätön, tyytymätön, hieman tyytymätön, hieman tyytyväinen, tyytyväinen ja erittäin tyytyväinen*. Analyysiä varten vastaukset on ryhmitelty siten, että kaksi viimeistä vastauskategoriaa on tulkittu tyytyväiseksi. Tätä muuttujaa kutsutaan tässä tutkielmassa *tutkintotyytyväisyydeksi*.

Luku 6

Regressiomallit

Analyysin suoritus voidaan jakaa kahteen osaan. Ensimmäisessä osassa käsitellään vuosien 2007, 2009 ja 2011 valmistuneiden yhdistelmäaineistoa. Toisessa osassa käytetään vain vuoden 2011 aineistoa. Tämä johtuu siitä, että uusimmassa kyselyssä on jonkin verran poikkeavuuksia verrattuna kahteen edelliseen kyselyyn. Osa näistä muutoksista on kiinnostavia muun muassa rahoitusmallin suhteen. Toisaalta kolmen aineiston yhdistelmällä saadaan pitkäaikaisempaa ja lukumäärältään suurempaa tietoa. Tämä myös häivyttää vuosien välistä vaihtelua pois analyysistä.

Osiassa 2.8 on esitelty tässä tutkielmassa käytettyjä mallinvalintakriteerejä. Hyödynän tutkielmassani AIC:ta, AICc:ta ja BIC:ta. Näiden tunnuslukujen perusteella pystytään tekemään arvioita siitä onko malleissa liikaa muuttujia. Pyrin näiden avulla löytämään sellaiset mallit, jotka selittävät hyvin tutkintotyytyväisyyttä, mutta eivät sisällä turhia muuttujia. Kun mallit ovat valmiit tulkitse näitä tarkemmin erillisessä osiossa muiden tunnuslukujen, kuten vetosuhteiden avulla.

Logistiset regressiot on muodostettu R:llä hyödyntäen sen glm-funktiota tai SPSS:llä (versio 24) GENLIN-funktion avulla.

6.1 Mallien rakentaminen

Agresti esittelee kirjassaan kaksi tapaa rakentaa regressiomalli: joko lisäämällä muuttujia yksitellen, kunnes lisääminen ei enää paranna mallia tai eliminaation, jossa turhia muuttujia poistetaan kokonaismallista. Tässä tutkielmassa hyödynnetään hybridiä näistä kahdesta tavasta eli kokonaismalliin verrataan pienempiä kokonaisuuksia. (Agresti, 2002)

Analyysissä käsitellään ensin malli, jossa on käytetty vuosien 2007, 2009 ja 2011 yhdistelmäaineistoa. Mallissa, jossa on käytetty kaikkia sopivia muuttujia, on 33 muuttujaa, kun huomioidaan vain päävaikutukset eli interaktiotermejä ei ole huomitu. Koska muuttujia on niin paljon, on selvyuden vuoksi helpompi käyttää muuttujista lyhenteitä tai lyhenneryhmiä. Lyhenteet on esitelty taulukossa 6.1. Tarkemmat sanamuodot kyselyssä voi katsoa liitteestä A.

Lyhenne	Selite	Muuttujamäärä	2007-2011	2011
S	Sukupuoli	1	On	On
I	Ikäryhmät	2	On	On
K	Koulutusala	12	On	On
J	On tehnyt jatko-opintoja	1	On	On
Ton	On ollut valmistumisen jälkeen työtön	1	On	On
H	Pystyy hyödyntämään jatkuvasti oppimaansa	1	On	On
V	Nykyinen työ vastaa koulutustasoa	1	On	On
Va	Riittävät valmiudet työelämään	1	Ei	On
T	Työnantaja arvostaa tutkintoani	1	Ei	On
Ta3	Työelämätaitojen kehittyminen	15	On	Ei
Ta1	Työelämätaitojen kehittyminen	23	Ei	On

Taulukko 6.1: Mallien rakentamisessa käytetyt muuttujat ja niiden lyhenteet tai lyhenne-ryhmät

Taulukkoon 6.2 on koottu erilaisten mallien tuloksia. Jos mukana on koulutusala, niin tällöin vertailuryhmänä on käytetty luonnontieteellistä koulutusala, sillä se on lähimpänä kaikkien vastauksien keskiarvoa.

Muuttujat	Muuttujien lukumäärä	N	AIC	AICc	BIC
S + I + K	14	3120	4022	4022	4113
S+ I + K + J + Ton + H + V	18	2676	3191	3191	3302
Ta3	15	2622	3863	3863	3959
S+ I + K + J + Ton + H + V + Ta3	33	2622	3051	3052	3251

Taulukko 6.2: Erilaisia logistisia regressiomalleja yhdistelmäaineistosta.

Seuraavaksi tarkastellaan vain vuoden 2011 aineistoa. Taulukkoon 6.3 on koottu erilaisten mallien tuloksia. Jos mukana on koulutusala, niin vertailuryhmänä pidetään luonnontieteellistä koulutusala.

Muuttujat	Muuttujien lukumäärä	N	AIC	AICc	BIC
S + I + K	14	914	1226	1227	1299
S+ I + K + J + Ton + H + V + Va + T	20	701	739	741	835
Ta1	23	826	1068	1070	1186
S+ I + K + J + Ton + H + V + Ta1	43	634	687	693	887

Taulukko 6.3: Erilaisia logistisia regressiomalleja vuoden 2011 aineistosta.

6.2 Lopullisen mallin rakentaminen ja valinta

Taulukoista 6.2 ja 6.3 voidaan tarkastella muuttujien vaikutusta malliin. Yleisesti voidaan päätellä, että pelkät taustamuuttujat sukupuoli, ikä ja koulutusala eivät luo hyvää mallia. Saman voi päätellä myös muuttujista, jotka koskevat työelämätaitojen kehittymistä yliopisto-opiskelun aikana. Vuosien 2007-2011 yhdistelmäaineistossa pienimmän BIC-arvon ja AIC-arvon saa malli, jossa on kaikki muuttujat mukana. Vuoden 2011 tilanne muuttuu siten, että pienimmän BIC-arvon saa malli, jossa on mukana taustamuuttujat sekä työuraa kuvaavat muuttujat. Pienimmän AIC-arvon saa malli, jossa on mukana kaikki arvot. Tämä antaa viitteen siitä, että yhdistelmäaineistossa on hyvä lähtökohta pitää mukana paljon muuttujia, ja mahdollinen mallin parantaminen kannattaa tehdä poistamalla joitakin muuttujia täydestä mallista. Vuoden 2011 aineistossa optimointia kannattaa suorittaa siten, että on lähtökohtana malli, jossa ei ole mukana työelämätaitoja, vaan lähteä liikkeelle pienemmästä mallista ja lisätä tähän joitakin työelämätaitoja kuvaavia muuttujia. Jotta lopullisten mallien rakentaminen olisi mahdollisimman helppoa, niin liitteeseen E on kirjattu täysien mallien vetosuhteet, sen luottamusväli (95 %), devianssit ja devianssien p-arvot.

6.2.1 Yhdistelmäaineisto

Lähtökohta on taulukon 6.2 mukaan BIC:n arvo 3251, joka saavutetaan silloin, kun mallissa käytetään kaikkia muuttujia. Liitteen E mukaan kuitenkin mallissa on paljon muuttujia, joiden vetosuhde on lähellä yhtä ja joiden devianssi on hyvinkin pieni. Koska BIC rankaisee melko voimakkaasti ylimääräisistä muuttujista, kiinnostava lähtökohta mallin parantamiseen on ylimääräisten muuttujien poistaminen mallista. Samalla voidaan huomata, että koulutusalojen välillä on suuriakin eroja miten ne vaikuttavat malliin. Koulutusaloja on melko paljon ja näiden kesken on melko selvä yhteys: professioalat, joilta

työllistyy suoraan johonkin ammattiin, kuten lääkäriksi, asianajajaksi tai eläinlääkäriksi, ovat tyytyväisempiä kuin niin sanotut generalistialat, kuten humanistinen ja maatalousmetsätieteellinen ala. Tämän perusteella yhdistelmäaineistoon luotiin uusi muuttuja, joka kuvaa onko ala selkeästi professioala tai ei. Professioaloiksi tulkittiin lääketieteellinen, hammaslääketieteellinen, eläinlääketieteellinen, psykologinen, farmasian ja oikeustieteen alat kokonaisuudessaan. Kasvatustieteelliseltä alalta voidaan erottaa lastentarhanopettajat ja luokanopettajat muista kasvatustieteilijöistä. Teologisessa tiedekunnassa ei voida erottaa pappeja, opettajia tai generalisteja toisistaan kaikissa kolmessa aineistossa. Muuttujan tekee ongelmalliseksi se, että vanhemmissa aineistoissa ei ole eroteltu aineenopettajia generalisteista. Vuoden 2011 aineistolle on pystytty tekemään parempi genearilisti-professio-muuttuja.

Seuraavaksi mallista poitettiin muuttujia sen perustella, että mallin BIC-arvo laski. Tässä vaiheessa BIC-arvoksi saatiin 3185. Poistetuissa muuttujissa ei ollut koulutusaloja koskevia muuttujia. Kun korvasin koulutusaloja kuvaavat dummy-muuttujat uudella professiota kuvaavalla muuttujalla, BIC-arvo putosi 3133. Tämän muuttujan vetosuhteeksi tuli noin 1,8 ja devianssiksi 39. Tämän perusteella on mallin kannalta hyvä idea korvata useita dummy-muuttujia yhdellä yleistasoisemmalla muuttujalla, jos mittarina käytetään BIC-arvoa. Ylimääräisistä parametreista vähemmän rankaiseva AIC-arvo on hieman korkeampi, kun käytetään professiota kuvaavaa muuttujaa verrattuna yksittäisiin koulutusalamuuttujiin.

Ennen lopullista mallia tutkittiin kaksitasoiset interaktiotermit. Malli parani lisäämällä mukaan interaktiotermin *oppimansa hyödyntäminen* * *koulutustasovastaavuus*, jonka β :n estimaatiksi tuli noin -1 . Muilla interaktiotermeillä ei ollut tilastollisesti merkittävää vaikutusta malliin.

Taulukkoon 6.4 on koottu yhdistelmäaineiston logistiseen regressiomalliin liittyvät tunnusluvut samalla tavalla kuin liitteessä E on tehty. Devianssien p-arvoja ei ole raportoitu

erikseen, mutta todettakoon, että kaikki ovat reilusti alle 0,01.

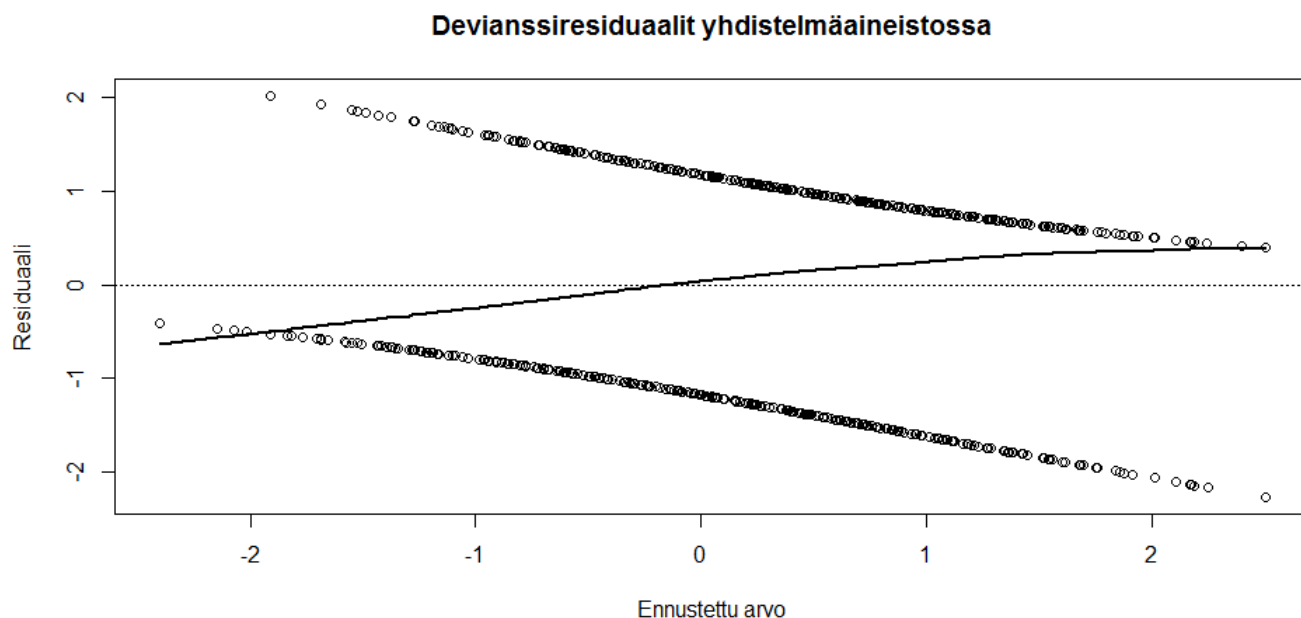
Muuttuja	OR ¹	OR:n luottamusväli	Devianssi
Ikä yli 29	1,62	1,34 - 1,97	14
Professioala	1,84	1,52 - 2,24	105
Työtön	0,61	0,50 - 0,74	34
opp. hyöd.	6,66	3,87 - 11,8	183
vaat.	2,33	1,71 - 3,18	16
keh. teor.	1,47	1,10 - 1,99	20
keh. orat.	1,27	1,03 - 1,55	24
keh. org.	1,33	1,16 - 1,69	21
keh. viest.	1,38	1,16 - 1,67	16
keh. opet.	1,31	1,13 - 1,65	11
opp. hyöd. * vaat.	-1,08	-1,68 - -0,50	14

Taulukko 6.4: Tunnuslukuja yhdistelmäaineiston lopullisesta mallista. ¹ interaktiotermin on raportoitu estimaatti ja sen luottamusväli.

Yhdistelmäaineiston diagnostiikka

Logistisen regressiomallin diagnostiikkaan liittyviä aiheita on käsitelty tarkemmin osiossa 2.6. Aloitetaan käsittely jäännösvirheistä. Devianssiresiduaalit yhdistelmäaineistolle on esitetty kuvassa 6.1.

Kuvasta 6.1 nähdään, että yhdistelmäaineiston tapauksessa residuaalit ovat melko lähellä nollaa. Kuvaan on piirretty lokaaliregressio, joka vahvistaa tämän päätelmän. Tämä on toivottava tulos ja viittaa siihen, että malli sovituu hyvin aineistoon. Kummassakin

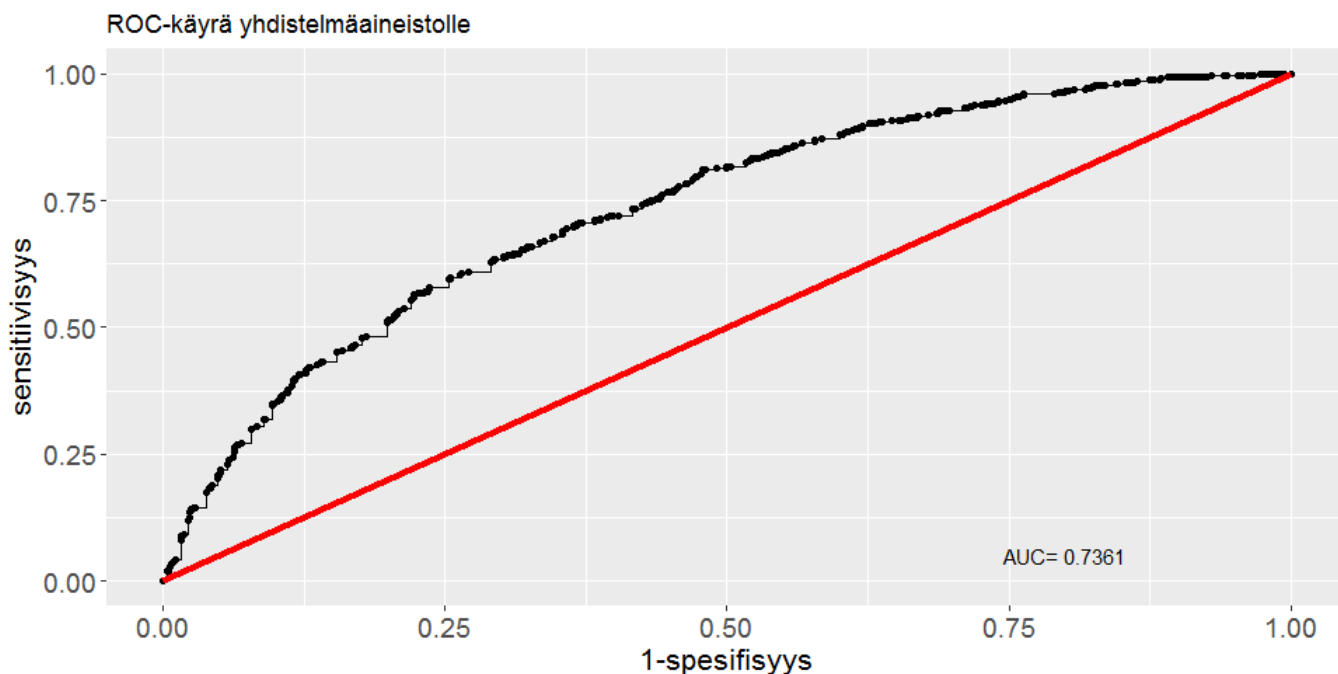


Kuva 6.1: Yhdistelmäaineiston devianssiresiduaalit

ääripäässä on joitakin pisteitä, jotka poikkeavat yleisestä trendistä. Pearsonin residuaalit muodostavat samankaltaisen hajontakuvan. Erona devianssiresiduaaleihin verrattuna on, että suurempi osa jäännösvirheistä on lähempänä nollaa, mutta toisaalta ääriarvojen itseisarvot ovat suurempia kuin devianssiresiduaalien.

Mallin ennustavuutta voidaan arvioida ROC-käyrän avulla. Kuvassa 6.2 on piirretty yhdistelmäaineiston ROC-käyrä.

Nähdään, että käyrän alle jäävän alueen pinta-ala on 0,74 eli malli ennustaa 74 % tapauksista oikein. Tämä on kohtuullinen tulos ja viittaa siihen, että malli ennustaa kohtuullisesti tutkintotyytyväisyyttä. Kohtuullista ennustavuutta tukee Hosmer-Lemeshow'n testiin liittyvän khin neliötestin arvo, joka on hieman yli 10 (p-arvo yli 0,2).



Kuva 6.2: Yhdistelmäaineiston ROC-käyrä.

6.2.2 Vuonna 2011 valmistuneiden aineisto

Vuonna 2011 valmistuneiden aineiston lähtökohtana on BIC:n arvo 835, joka saavutettiin, kun mallissa oli mukana vain taustamuuttujat ja työhön liittyvät muuttujat. Tällöin työelämätaidot on jätetty mallista pois. Kuitenkin edellisessä osiossa nähtiin, että joillakin työelämätaidoilla oli tilastollisesti merkitsevä vaikutus regressiomalliin, joten on kiinnostavaa tutkia onko näillä vaikutusta myös uusimmassa aineistossa. Samoin lopullista mallia varten arvioin samalla tavalla kuin edellisessä osiossa 6.2.1 onko koulutusala kuvaavat dummy-muuttujat syytä korvata professiota kuvaavalla yhdellä muuttujalla. Edellisessä osiossa esiteltiin tämän muuttujan luonti myös vuoden 2011 aineistolle, jossa voidaan tehdä tarkempi jaottelu professio- ja generalistialoille, koska aineenopettajat voidaan erotella aineistossa.

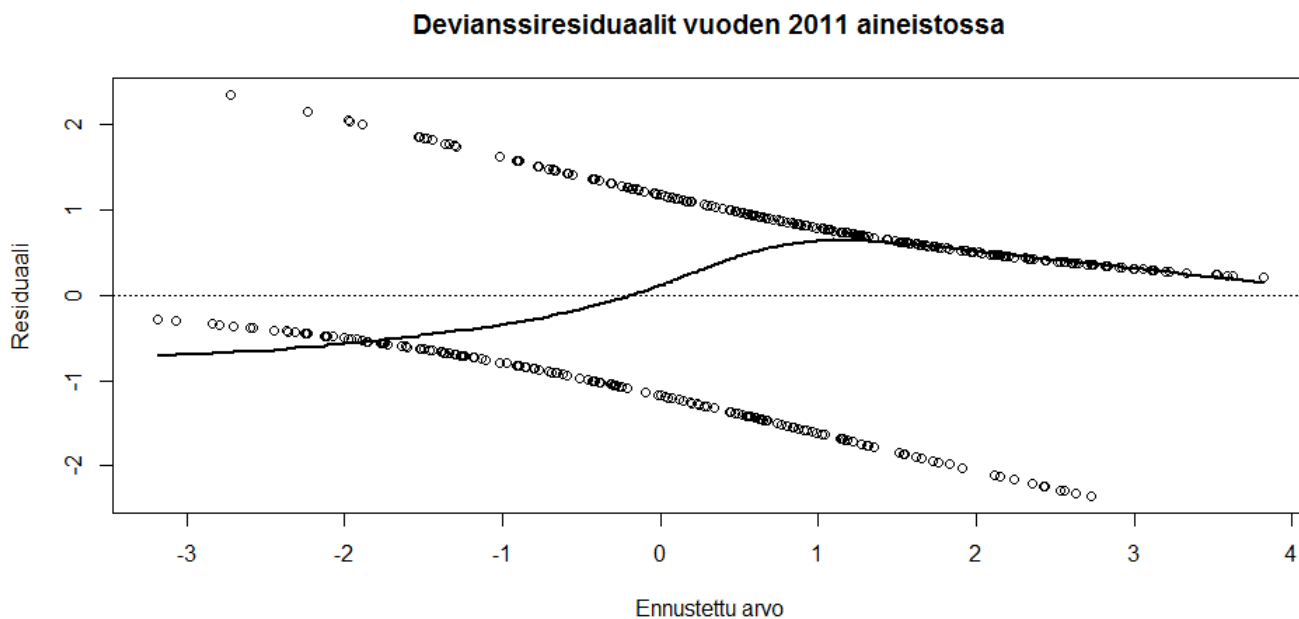
Jälleen koulutusaloja kuvaavien dummy-muuttujien korvaaminen professiota kuvaavalla dummy-muuttujalla pienensi mallin BIC-arvoa useilla kymmenillä. Jos tämän jälkeen lisättiin yksittäisiä koulutusalaan kuvaavia dummy-muuttujia, mallin BIC-arvo kasvoi. Malliin koetettiin lisätä kaikki vuonna 2011 kysytyt työelämätaidot, mutta näistä vain kolmella oli tilastollisesti merkitsevä vaikutus malliin. Malliin lisättiin kaikki mahdolliset kaksitasoiset interaktiotermit päävaikutusten lisäksi, mutta näillä ei ollut vaikutusta malliin. Lopulta mallin BIC-arvoksi saatiin 744. Taulukkoon 6.5 on koottu vuonna 2011 valmistuneiden aineiston logistiseen regressiomalliin liittyvät tunnusluvut samalla tavalla kuin edellisessä osiossa. Toisin kuin edellisessä osiossa, nyt raportoidaan myös devianssien p-arvot, sillä ne vaihtelevat melko paljon.

Muuttuja	OR	OR:n luottamusväli	Devianssi	p-arvo
Sukupuoli (nainen)	0,56	0,36-0,85	5	< 0,05
Ikä yli 29	1,80	1,16-2,83	8	< 0,01
Professio	0,82	0,53-1,26	22	< 0,01
Työtön	0,62	0,41-0,94	12	< 0,01
opp. hyöd.	2,37	1,60-3,52	60	< 0,01
vaat.	1,61	0,93-2,80	7	< 0,05
riit. valm.	2,97	1,91-4,65	40	< 0,01
ta. arv.	6,24	4,24-10,08	77	< 0,01
keh. tieh.	1,40	0,89 - 2,70	4	< 0,01
keh. org.	1,48	1,00 - 2,26	6	< 0,01
keh. opet.	1,69	1,08 - 2,54	6	< 0,01

Taulukko 6.5: Tunnuslukuja vuonna 2011 valmistuneiden lopullisesta mallista. P-arvot on raportoitu < 0,05 ja < 0,01.

Vuona 2011 valmistuneiden aineiston diagnostiikka

Devianssiresiduaalit vuonna 2011 valmistuneiden aineistolle on esitetty kuvassa 6.3.

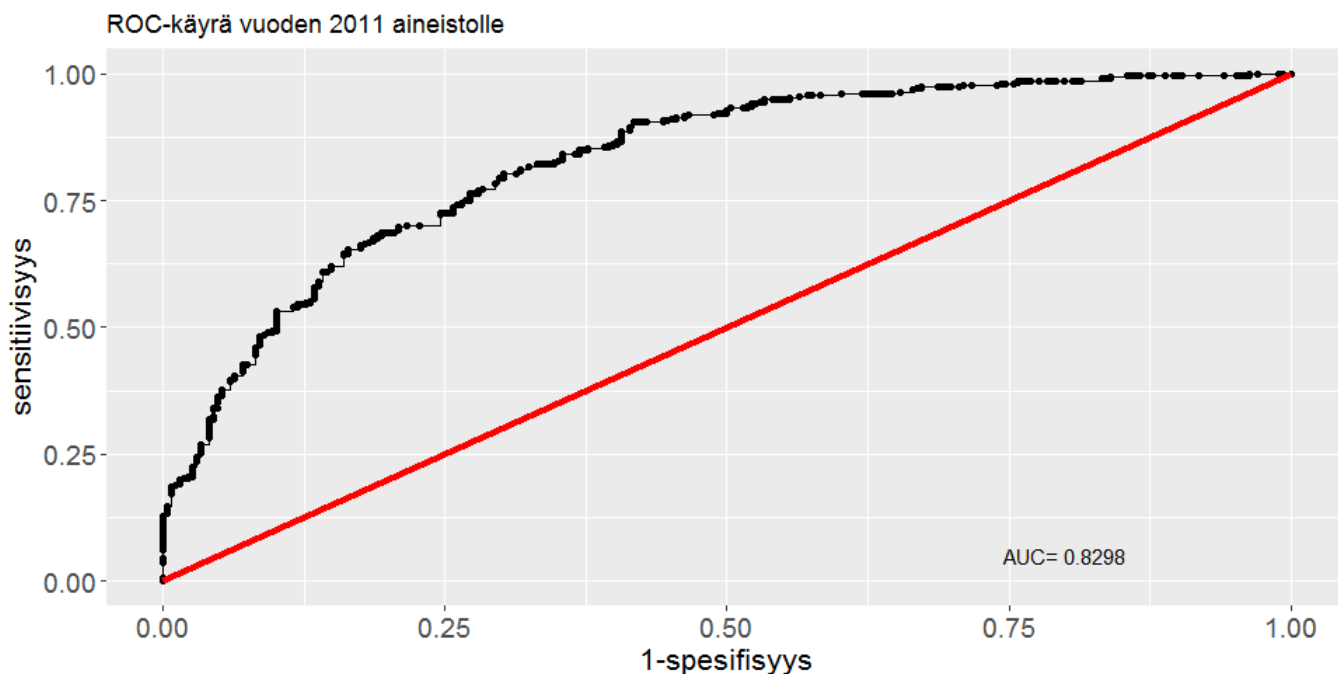


Kuva 6.3: Vuonna 2011 valmistuneiden aineiston devianssiresiduaalit.

Kuvasta 6.3 nähdään, että myös vuonna 2011 valmistuneiden aineistossa jäännösvirheet ovat lähellä nollaa. Erona yhdistelmäaineistoon on, että lokaaliregressio on hieman kauempana nollassa, joskaan ei hälyyttävästi. Samoin ääriarvojen itseisarvot ovat hieman suurempia kuin yhdistelmäaineistossa. Pearsonin residuaalit käyttäytyvät samalla tavalla kuin yhdistelmäaineistossa: ääriarvot ovat hieman suurempia, mutta yleisesti jäännösvirheet ovat lähempänä nollaa.

Kuvassa 6.4 on piirretty yhdistelmäaineiston ROC-käyrä.

Nähdään, että käyrän alle jäävän alueen pinta-ala on 0,83 eli malli ennustaa 83 % tapauksista oikein. Tämä on paremmin kuin yhdistelmäaineistossa ja mallin avulla pys-



Kuva 6.4: Vuonna 2011 valmistuneiden aineiston ROC-käyrä.

tytään ennustamaan melko hyvin lopputulos. Tulosta tukee Hosmer-Lemeshow'n testin hyvin pieni khin neliötestin arvo, ja siihen liittyvän p-arvon hyvin suuri arvo.

6.2.3 Multinomiaaliset logistisen regression mallit

Koska tutkintotyytyväisyys on kysytty alkuperäisessä kyselyssä kuusiportaisella asteikolla, tutkittiin vielä pystytäänkö datan ymmärrystä lisäämään, jos käytetään binomisen logistisen regression tukena multinomiaalista logistista regressiota. Tällöin kysymys tutkintotyytyväisyydestä on luokiteltu kolmeen luokkaan: tyytymätön (1-2), "väli" (3-4) ja tyytyväinen (5-6). Tarkoituksena on selvittää miten hienojakoisempi jaottelu muuttaisi tuloksia. Ryhmät eivät ole samankokoisia, sillä kaikkina kolmena vuotena tyytymättömiä on 6-7 %, kun tyytyväisiä on noin 60 %. Taulukkoon 6.6 on koottu vastausjakaumat

valmistumisvuosittain kolmessa ryhmässä.

Valmistumisvuosi	Tyytymättömiä	Väliryhmä	Tyytyväisiä	Yhteensä
2007	99	447	876	1422
2009	55	239	496	790
2011	53	334	527	914

Taulukko 6.6: Tutkintotyytyväisyys kolmessa luokassa valmistumisvuosittain.

Pohjana on käytetty samoja malleja, jotka on esitetty osioissa 6.2.2 ja 6.2.1. Multinomiaaliset logistiset regressiot on tehty hyödyntäen R:n nnet-pakettia. Vertailukategoriana on käytetty tyytymättömien ryhmää (1-2). Kahteen seuraavaan taulukkoon on koottu vetosuhteet kummankin mallin samoilta muuttujille, jotka on löydetty optimoiduissa malleissa. Mahdollisia interaktiotermiä ei ole huomioitu.

Muuttuja	(3-4)	(5-6)
Ikä yli 29	0,80	1,35
Professioala	1,17	2,16
Työtön	0,77	0,48
opp. hyöd.	2,70	14,93
vaat.	3,61	6,68
keh. teor.	1,76	2,35
keh. orat.	1,05	1,34
keh. org.	1,36	1,84
keh. viest.	1,35	1,82
keh. opet.	1,24	1,65

Taulukko 6.7: Yhdistelmäaineiston vetosuhteet multinomiaalisessa logistisessa regressiossa

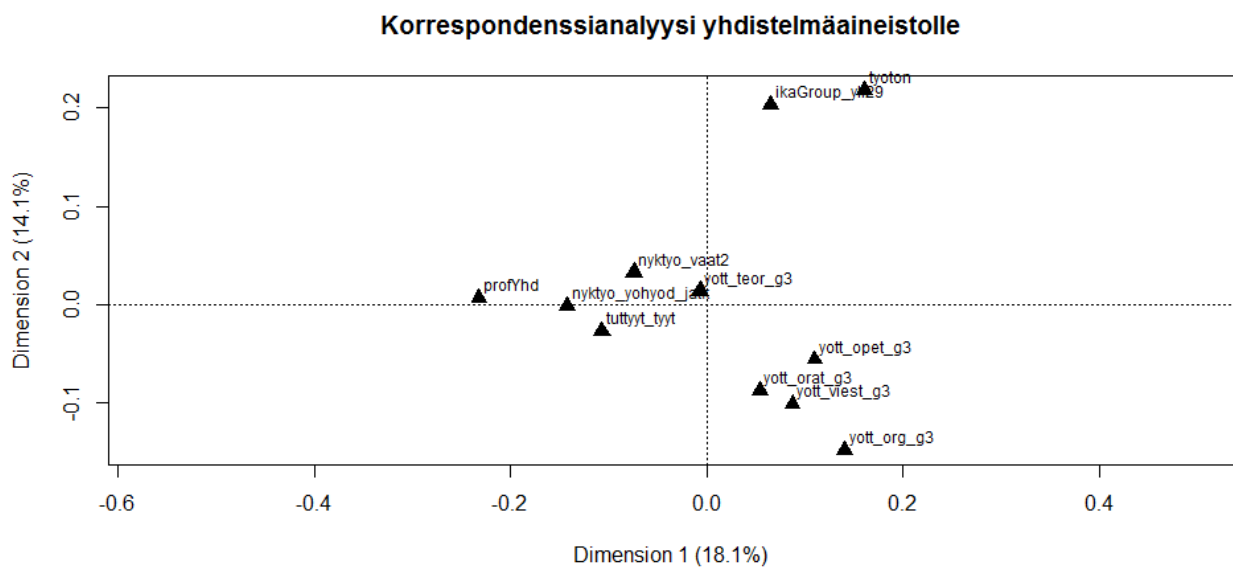
Muuttuja	(3-4)	(5-6)
Sukupuoli (nainen)	0,93	0,52
Ikä yli 29	0,44	0,85
Professio	0,79	0,66
Työtön	0,58	0,37
opp. hyöd.	1,21	2,82
vaat.	4,54	6,05
riit. valm.	0,44	1,38
ta. arv.	6,14	33,82
keh. tieh.	1,74	2,35
keh. org.	1,66	2,37
keh. opet.	0,73	1,26

Taulukko 6.8: Vuonna 2011 valmistuneiden aineiston vetosuhteet multinomiaalisessa logistisessa regressiossa

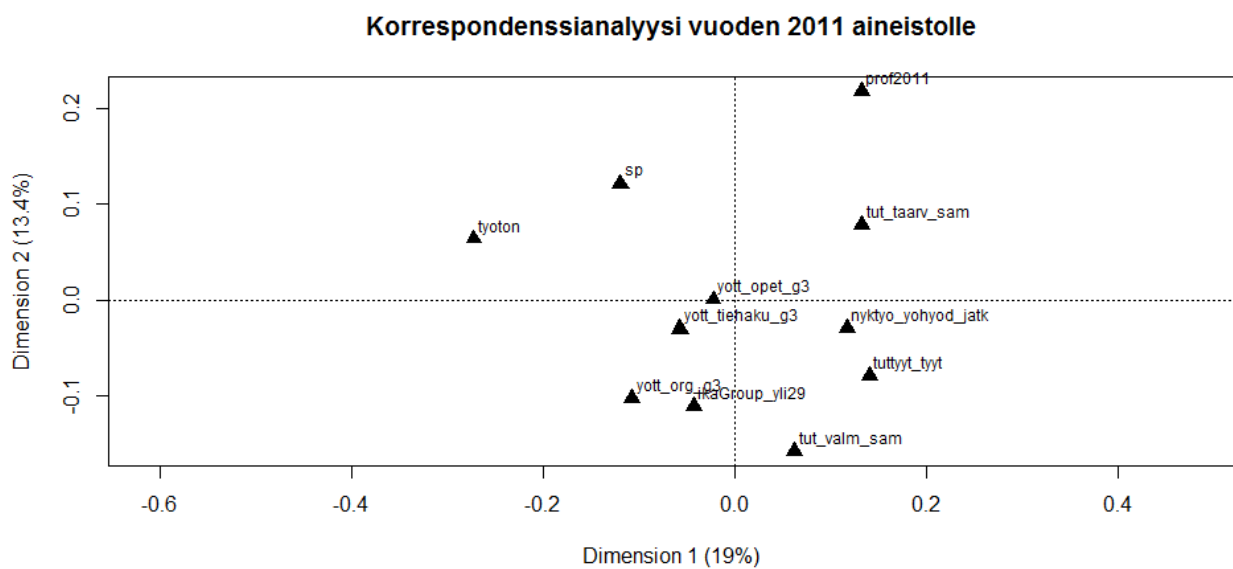
Erityisesti vuoden 2011 aineiston kohdalla on huomioitava, että luottamusvälit ovat melko suuret ja vetosuhteille laskettava p-arvo on merkittävän pieni vain muuttujilla työtön (ryhmä 5-6), pystyy hyödyntää oppimaansa (ryhmä 5-6), koulutustasovastaavuus ja työnantajat arvostavat tutkintoa. Tämä selittyy sillä, että aineiston koko on niin pieni, että multinomiaalinen logistinen regressio ei ole paras vaihtoehto. Toisaalta tilanne on paljon parempi yhdistelmäaineistossa, erityisesti ryhmässä 5-6, jossa tilastollisesti merkittävää eroa ei ole muuttujissa yli 29-vuotiaat ja ongelmanratkaisutaidot (orat.). Vastaavasti 3-4-ryhmässä tilastollisesti merkittävä ero on oppimansa hyödyntämisessä, koulutustasovastaavuudessa ja teoreettisessa osaamisessa.

6.2.4 Mallien visualisointi

Yksi tapa visualisoida monimuuttujamallia on korrespondenssianalyysi. Sen avulla voidaan etsiä luokittelevien muuttujien välisiä yhteyksiä. Tarkempi teoria on esitetty osiossa 2.9.

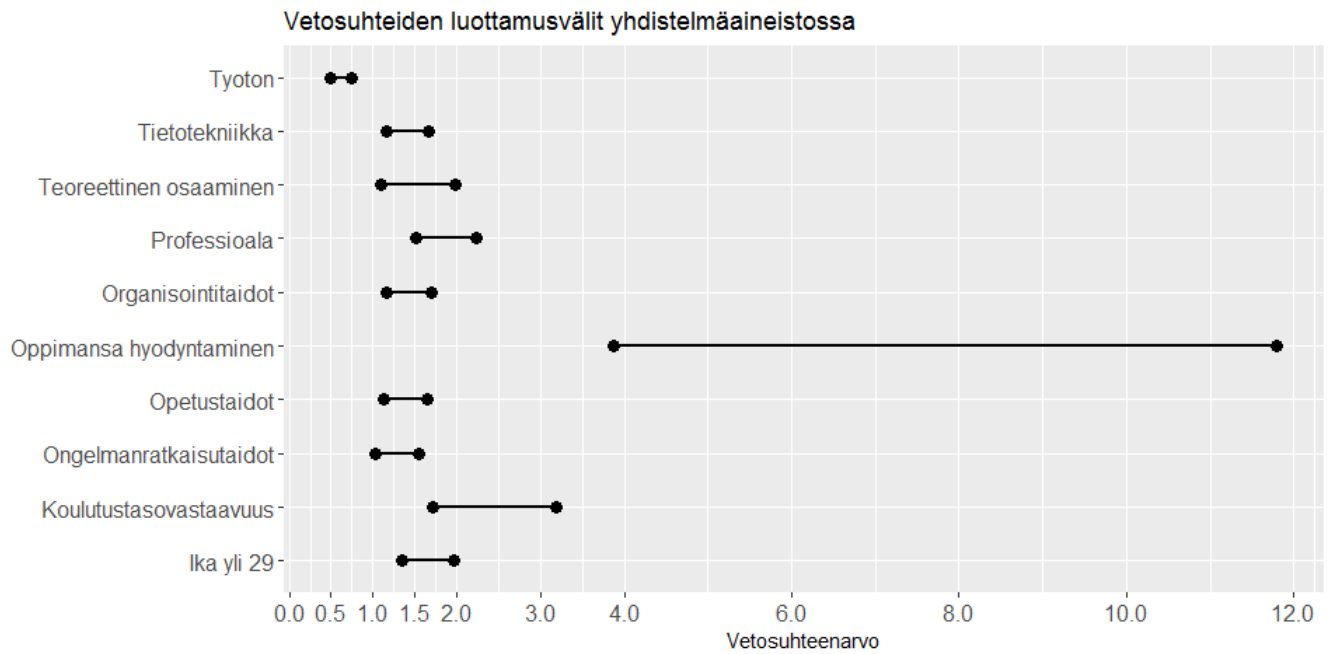


Kuva 6.5: Yhdistelmäaineistosta piirretty korrespondenssianalyysi.

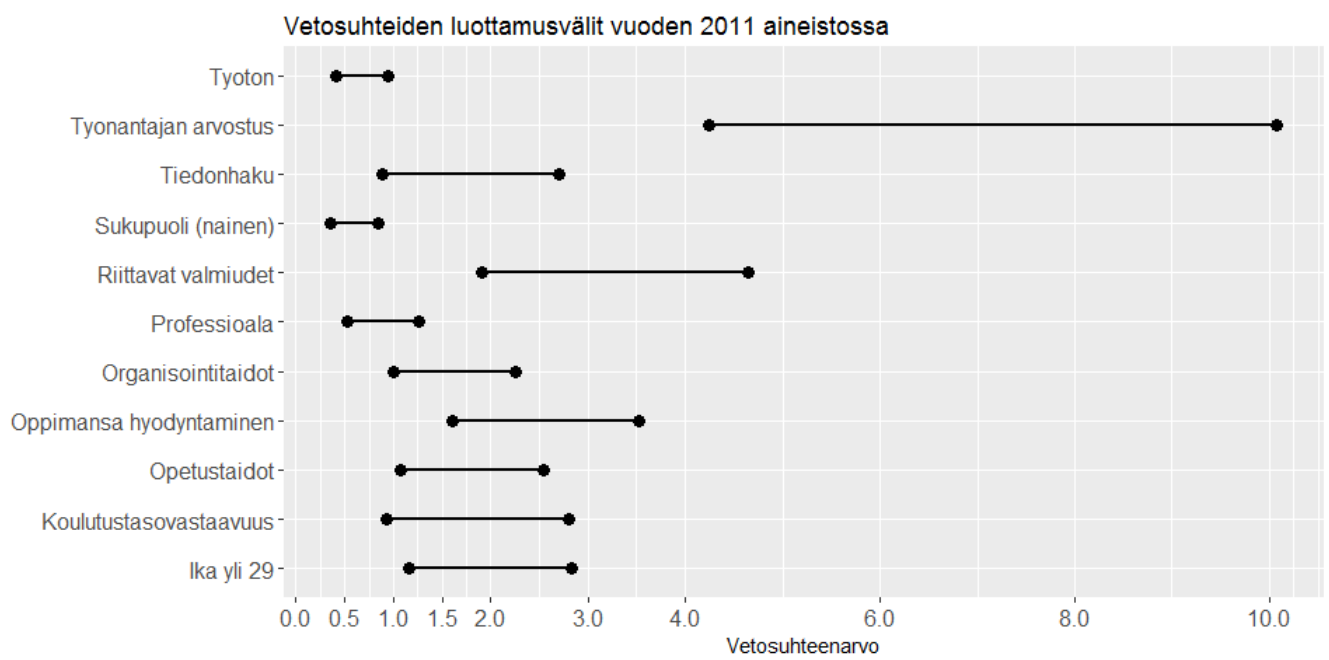


Kuva 6.6: Vuonna 2011 valmistuneiden aineistosta piirretty korrespondenssianalyysi.

Myös vetosuhteita voidaan visualisoida yksinkertaisilla kuvilla. Tämä voi helpottaa vetosuhteiden vertailua keskenään.



Kuva 6.7: Yhdistelmäaineistosta tehdyn logistisen regressiomallin vetosuhteiden luottamusvälit.



Kuva 6.8: Vuonna 2011 valmistuneiden aineistosta tehdyn logistisen regressiomallin vetosuhteiden luottamusvälit.

Luku 7

Johtopäätökset

Johtopäätösten tekemisen pohjana ovat taulukot 6.4 ja 6.5. Analyysiä tuetaan muilla osiossa 6.2 esitetyillä taulukoilla sekä tarvittaessa muilla tiedoilla. Ensin käsitellään yhdistelmäaineisto ja tämän jälkeen vuoden 2011 aineisto. Lopuksi pyritään tekemään yleisiä johtopäätöksiä koko aineistoa koskien. Tuloksia peilataan osiossa 3 esitettyyn lyhyeen kirjallisuuskatsaukseen.

Jos verrataan aluksi tehtyä mallia, jossa oli monta kymmentä selittävää muuttujaa, niin lopuksi on päästy melko yksinkertaiseen malliin. Yhdistelmäaineistosta muodostui kymmenen muuttujan päävaikutusten sekä yhden interaktiotermin muodostama logistinen regressio. Kun tarkastellaan taulukossa 6.4 olevia vetosuhteita, niin huomataan, että ne ovat suhteellisen maltillisia. Tämä koskee erityisesti työelämätaitojen kehittymistä yliopisto-opiskelussa, jossa korkein 95 % luottamusvälin yläraja on 1,99. Kiinnostavaa on, että vain osa työelämätaidoista jäi malliin eikä suurimmalla osalla muuttujista ollut merkitystä. Malliin jäi mukaan teoreettinen osaaminen, ongelmanratkaisutaidot, organisoititaidot, ICT-taidot ja opetustaidot. Kaikkien suunta oli sama: jos vastaaja koki, että yliopisto-opiskelu kehitti kyseisiä taitoja hyvin, niin tällöin hän oli myös todennäköisemmin tyytyväinen tutkintoonsa. Nämä viisi muuttujaa ovat selkeästi yliopiston toimintaan

liittyviä muuttujia. Taustamuuttujista malliin jäi vain ikä eli varttuneemmat vastaajat olivat hieman tyytyväisempiä tutkintoonsa kuin korkeintaan 29-vuotiaat.

Professioala, työttömyys, oppimansa hyödyntäminen ja koulutustasovastaavuus ovat taasen osittain yliopistosta irrallisia muuttujia. Nämä liittyvät esimerkiksi työmarkkinatilanteeseen: professioaloilla on matalampi työttömyys kuin generalistialoilla. (Taulu, 2017) Toisaalta mahdollinen selitys on myös, että näiden alojen koulutus antaa paremmat työelämävalmiudet: 76 % professioalojen vastaajista pystyy hyödyntämään jatkuvasti oppimaansa, kun taas generalistialoilla vastaava luku on 54 %. On kuitenkin selvitetty, että generisiä taitojen oppimista ei tunnusteta niin hyvin kuin käytännön taitoja. (Chan, 2010) Tämä tukisi sitä, että aloilla, joilla on enemmän selkeitä taitovaatimuksia, kuten lääketieteellisellä alalla, yliopisto-opiskelun kehittämä osaaminen tunnustetaan paremmin.

Koska työttömyyden kokemisella on tilastollisesti merkitsevä yhteys vastaajan tyytyväisyyteen, nostaa tämä tutkimus esille työllistymisen ja työllistyvyyden tärkeyden. On väitetty, että tutkinnon merkitys työllistymiselle on, että se mahdollistaa työllistymisen, mutta rekrytointitilanteessa arvostetaan yleisiä taitoja. (Tuominen, 2012) Samassa artikkelissa korostetaan muun muassa viestintä- ja kongnitiivisia taitoja, jotka nousevat myös tässä tutkielmassa esille. On raportoitu, että suurin osa valmistuneista kokee, että yliopiston tärkein tehtävä on tuottaa heille riittävät teoreettiset valmiudet, mutta samalla kaivattiin sosiaalisia taitoja. Samassa raportissa todetaan, että 64 % vastaajista on oppinut tärkeimmät työtaitonsa työelämässä eikä koulutuksen aikana. (Slotte *et al.*, 2006)

Oppimansa hyödyntäminen ja koulutustasovastaavuus ovat yhdistelmäaineistossa merkittävimmät muuttujat. Erityisesti he, jotka voivat hyödyntää jatkuvasti oppimaansa nykyisessä työssään, erottuvat selkeästi ryhmästä, joka ei voi hyödyntää oppimaansa. Tämä kysymys voidaan tulkita sekä yliopiston sisäiseksi muuttujaksi ja toisaalta työmarkkinointa johtuvaksi: opetetaanko yliopistossa oikeita asioita vai onko vastaajilla ollut sellainen tilanne, että on otettava vastaan työ, vaikka siinä ei pystyisikään hyödyntämään oppi-

maansa. Nämä kaksi muuttujaa eivät kuitenkaan ole täysin additiivisia, sillä näillä on tilastollisesti merkittävä negatiivinen yhdysvaikutus. Tämän tulkintaa helpottaa ristiintaulukointi: jos pystyy hyödyntämään oppimaansa jatkuvasti, niin tämän ryhmän sisällä ei koulutustasovastaavuudella ole merkitystä tutkintotyytyväisyyteen, vaan tutkintoonsa tyytyväisiä on 74 %. Jos taas nykyisessä työssä ei voi hyödyntää oppimaansa ja työ ei vastaa koulutustasoa, niin heistä 29 % oli tyytyväisiä tutkintoonsa, mutta jos työ vastasi koulutustasoa tai oli vaativampaa, niin heistä 50 % oli myös tyytyväisiä tutkintoonsa. Tämä myös samalla selittää osittain miksi opintojen hyödyntämisellä on suurempi vaikutus malliin kuin koulutustasovastaavuudella.

Osiossa 6.2.3 esitetty multinomaalinen logistinen regressiomalli antaa lisäinformaatiota tutkitotyytyväisyydestä. Kun vertailuryhmänä on käytetty tyytymättömien ryhmää, niin kaikkien muuttujien kohdalla tyytyväisten ryhmä erottuu väliryhmästä siten, että tyytyväisyyden todennäköisyys kasvaa. Esimerkiksi jos vastaaja pystyy hyödyntämään oppimaansa jatkuvasti nykyisessä työssään, on hänen vedonlyöntisuhteensa noin 15 kertainen suhteessa heihin, jotka eivät voi hyödyntää oppimaansa jatkuvasti, kun väliryhmän vedonlyöntisuhde on noin kolme. Multinomiaalinen logistinen regressio tukee alkuperäistä binääristä jaottelua siinä mielessä, että tyytyväisten ryhmä erottuu selkeästi myös väliryhmästä, joka on lähempänä kaikista tyytymättömimpien ryhmää.

Osiossa 6.2.4 on piirretty korrespondenssianalyysin kaksiulotteinen kuva. Tämän avulla voidaan tehdä päätelmiä muuttujien suhteellisesta läheisyydestä toisiinsa nähden. Yhdistelmäaineiston kohdalla muuttujat muodostavat kolme ryhmää, joista yksi muodostaa selkeän kokonaisuuden. Tähän ryhmään kuuluu sosiaalisia ja kognitiivisia taitoja kuvaavia muuttujia. Keskelle kuvaa kuvautuu suurin osa muista muuttujista, kuten nykyistä työtä kuvaavat muuttujat, teoreettisen osaamisen kehittyminen ja tieto professionaalasta. Ikä ja työttömyys sijoittuvat erilleen muista. Oikeaan alakulmaan sijoittuva työelämätaitojen rypäs on mielenkiintoinen ja viestii siitä, että nämä muodostavat yhdessä ryhmän.

Tätä ryhmää voidaan myös pitää osiona, johon yliopisto voi helpommin vaikuttaa omilla toimillaan, muun muassa ICT-taitoja on helppo kehittää tiettyyn rajaan asti. Ryhmässä on myös yleisempiä taitoja, kuten ongelmanratkaisu- ja organisointitaidot. Työelämätaitoja kuvaavilla muuttujilla on myös pienemmät vetosuhteet kuin muuttujilla, jotka ovat keskellä kuvaajaa.

Vuoden 2011 aineistosta muodostuva logistinen regressiomalli koostuu suurelta osin samoista muuttujista kuin yhdistelmäaineisto. Erona on, että tässä aineistossa korostuu eri työelämätaitojen kehittyminen sekä sukupuolen vaikutus. Tämän lisäksi mukaan on voitu ottaa vain joitakin vuonna 2011 kysytyjä muuttujia, jotka kuvaavat riittävien valmiuksien kehittymistä ja arvostaako työnantaja tutkintoa. Kaksi viimeisintä muuttujaa ovat mallin kannalta tärkeitä muuttujia, sillä näiden muuttujaryhmien välillä on korkeat vetosuhteet. Erityisesti se, että työnantaja arvostaa tutkintoa, kasvattaa myös vastaajan omaa tutkintotyytyväisyyttä. Koulutustasovastaavuuden vetosuhde on likimain sama kuin yhdistelmäaineistossa. Oppimansa hyödyntäminen näyttäisi olevan vuoden 2011 kohdalla vähemmän tyytyväisyyttä erottava muuttuja. Yleisesti aineiston pienempi koko heikentää tulosten raportoinnin luotettavuutta suhteessa yhdistelmäaineistoon.

Suurin muutos koskee professioalaa: yhdistelmäaineistossa vetosuhteen arvo on yli 1, eli professioalalta valmistuneet ovat todennäköisemmin tyytyväisempiä kuin ei-professioalalta valmistuneet, kun muut muuttujat pidetään vakiona. Vuoden 2011 aineistossa asetelma on kääntynyt toisinpäin, joskin vetosuhteen luottamusväli on melko suuri eikä suuria päätelmiä voida tehdä. Vuoden 2011 aineistoon on pystytty hyödyntämään erilaista professio-muuttujaa, jossa on huomioitu myös aineenopettajat, joka selittää osaltaan muutosta. Jos tämä korvataan samalla professiomuuttujalla, jota käytettiin yhdistelmäaineiston kanssa, vetosuhteen arvo kasvaa hieman, mutta on edelleen alle yhden. Aineenopettajat ovat hieman tyytymättömämpiä tutkintoonsa kuin muut professioalalta valmistuneet, mutta ero on vain muutaman prosenttiyksikön. Todettakoon, että mallin BIC-arvo olisi edelleen sa-

ma ilman professiomuuttujaa kuin sen kanssa vuoden 2011 aineiston kohdalla. Tilastollisesti merkittäviä kaksiulotteisia interaktiotermejä ei havaittu, joskin jotkin interaktiot heikensivät professiomuuttujaa koskevaa Waldin testisuuretta.

Multinomiaalinen logistinen regressio antaa myös vuonna 2011 valmistuneiden aineiston kohdalla samanlaisia viitteitä kuin yhdistelmäaineiston kohdalla, joskin erot väliryhmän ja kaikista tyytyväisempien välillä ovat hieman pienempiä. Poikkeuksen muodostaa muuttuja, joka kuvaa työnantajan arvostusta tutkintoa kohtaan, missä ryhmien väliset erot ovat suuret. On kuitenkin huomioitava, että luottamusvälit ovat melko suuria, sillä yhden ryhmän koko ei enää ollut kovinkaan suuri ja liian pitkälle meneviä johtopäätöksiä tulee välttää.

Korrespondenssianalyysin tulokset eivät ole niin helposti tulkittavia kuin yhdistelmäaineiston kohdalla. Muuttujat eivät muodosta samalla tavalla selkeitä ryhmiä, vaan hajautuvat tasaisemmin kummallekin akselille. Varauksella voidaan todeta, että kuten yhdistelmäaineiston kohdalla, työelämätaitojen kehittymistä kuvaavat muuttujat muodostavat jonkinlaisen ryppään.

Aineisto on sinänsä erikoislaatuinen, sillä tyypillisesti opiskelijakyselyt toteutetaan opiskeluvaiheessa oleville opiskelijoille eikä useamman vuoden päästä valmistumisesta. Tämä luo osaltaan hankaluutta yhdistää tutkittava aineisto tutkittuun käsitteistöön ja teoriaan.

Kaiken kaikkiaan mallin kannalta tärkeimmiksi muuttujiksi osoittautuivat nykyisen työn laatua koskevat muuttujat: pystyykö hyödyntämään oppimaansa ja vastaako työ koulutustasoa sekä onko ollut työttömänä. Taustamuuttujien suhteen suuremmassa mitakaavassa sukupuolien välillä ei ollut juuri eroja, mutta uuden aineiston valossa naiset ovat hieman tyytymättömämpiä tutkintoonsa kuin miehet. Samoin yli 29-vuotiaat olivat hieman tyytyväisempiä, mutta tässäkään ero ei ollut kovin suuri. Uuden aineiston valossa voidaan nostaa myös merkittäviksi muuttujiksi riittävien valmiuksen saaminen ja sen, että

työnantaja arvostaa tutkintoa. Kaksi jälkimmäistä muuttujaa olisi syytä tarkastella vielä paremmin joko valtakunnallisen aineiston parissa tai seuraavan kyselykierroksen aineiston avulla. Australialaisissa yliopistoissa toteutetuissa samankaltaisissa tutkimuksissa on todettu, että juuri yleisten taitojen kehittyminen on voimakkaassa yhteydessä opiskelijoiden kokonaistyytyväisyyteen. (Grace *et al.*, 2012) (Brown *et al.*, 2004) Näissä kyselytutkimuksissa tarkasteltu työuran vaikutusta tyytyväisyyteen, mutta näissä on voitu osoittaa, että esimerkiksi hyväksi koettu opettaminen vaikuttaa ensisijaisesti taitojen kehittymiseen, joka on yhteydessä tyytyväisyyteen. Tämä ilmiö korostuu erityisesti naisopiskelijoilla.

Sukupuoli on usein kiinnostava muuttuja, kun pyritään löytämään ryhmien välisiä eroja. Australialaisessa tutkimuksessa on huomattu, että naiset ovat vaativampia opintojen laadun suhteen kuin miehet. Tämä saa hieman tukea tutkielmassa käytetyssä aineistossa, mutta suhteessa muihin muuttujiin, sukupuoli ei näyttele suurta roolia, kun pyritään selittämään tutkintotyytyväisyyttä. Kyseisessä tutkimuksessa pyrittiin selittämään tätä kehnomilla töillä ja palkalla. Ensimmäiselle selittävälle tekijälle ei löydy tukea kerätystä aineistosta. Aineiston palkkatiedot ovat hyvin puutteelliset ja sisältävät paljon epävarmuutta. Tämä huomioiden aineisto antaa osviittaa siihen, että naiset saavat kehnompaa palkkaa kuin miehet. Ero ei kuitenkaan selity koulutusaloilla eroilla.

Uraseuranta-aineistot tukevat tätä ilmiötä, jossa geneeristen taitojen oppiminen on yhteydessä tutkintotyytyväisyyteen. On kuitenkin huomioitavaa, että vain osa näistä taidoista jää malliin ja näiden vaikutus on pienempi kuin yleisemmän *koulutus antoi riittävät valmiudet työelämään* -kysymyksen. Tätä voi osittain selittää tunnettu ilmiö, jonka mukaan vastaajat eivät osaa tarkastella selvästi oppimaansa. (Tuominen, 2012) On mahdollista, että valmistuneet eivät osaa erotella tärkeitä yksittäisiä taitoja, mutta jos he yleisesti kokevat saaneensa riittävät valmiudet työelämään, he ovat tällöin myös tyytyväisiä tutkintoonsa.

Yliopiston kannalta kiinnostava muuttuja on myös koulutusala, joka on pelkistetty lo-

pullisessa mallissa yhdeksi professioalaa koskevaksi muuttujaksi. Täydemmissä malleissa on kuitenkin nähtävissä jonkin verran koulutusalojen välisiä eroja. Näissä malleissa professioalojen lisäksi korostuu humanistisen ja maatalous-metsätieteellisen alan hieman kehnompi tutkintotyytyväisyys. Näitä voidaan kuitenkin jonkin verran selittää huonommalla työmarkkinatilanteella.

Samalla olisi kiinnostavaa tutkia uuden aineiston kanssa kuinka hyvin nyt luodut mallit pystyvät ennustamaan tutkintotyytyväisyyttä. Tässä tutkimuksessa käytetty aineisto oli kuitenkin sen verran pieni, että sitä ei voi osittaa kahteen osaan, jolloin olisi voinut arvioida mallin ennustettavuutta.

Luku 8

Lähteet

Kirjallisuutta

Agresti, Alan. 2002. *Categorical Data Analysis*. Toinen painos edn. Wiley.

Agresti, Alan. 2015. *Foundations of Linear and Generalized Linear Models*. Wiley.

Akaike, Hirotugu. 1974. A New Look at the Statistical Model Identification. *IEEE Transactions on Automatic Control*, **19**(6), 716 – 723.

Andrews, Jane, & Higson, Helen. 2008. Graduate Employability, 'Soft Skills' Versus 'Hard' Business Knowledge: A European Study. *Higher Education in Europe*, **33**(4), 411–422.

Brown, Shane, Kabanoff, Boris, & Richardson, Alan. 2004. Business Graduates' Perceptions of the Quality of their Course: A View from their Workplace. *Journal of Institutional Research*.

Cavanaugh, Joseph. 1997. Unifying the derivations for the Akaike and corrected Akaike information criteria. *Statistics & Probability Letters*, **33**(2), 201–208.

- Chan, Wincy. 2010. Stundets' understanding of generic skills development in a university in Hong Kong. *Procedia Social and Behavioral Sciences*, **2**, 4815–4819.
- Demaris, Alfred. 2004. *Regression with Social Data : Modeling Continuous and Limited Response Variables*. Wiley.
- Douglas, Jacqueline, McClelland, Robert, & Davies, John. 2008. The development of a conceptual model of student satisfaction with their experience in higher Education. *Quality Assurance in Education*, **16**(1), 19–35.
- Dziak, John, Coffmann, Donna, Lanza, Stephanie, & Li, Runze. 2012. Sensitivity and specificity of information criteria. *Technical Report Series (The Pennsylvania State University)*, **12**.
- Grace, D, Weaven, S, Bodey, K, Ross, M, & K, Weaven. 2012. Putting student evaluations into perspective: The Course Experience Quality and Satisfaction Model (CEQS). *Studies in Educational Evaluation*, **38**, 35–43.
- Graybill, Franklin. 1976. *Theory and Application of the Linear Model*. Vol. 1. Wadsworth Publishing Company.
- Grebennikov, Leonid, & Skaines, Ivan. 2007. Gender and higher education experience: a case study. *Higher Education Research & Development*, **28**(1), 71–84.
- Greenacre, Michael, & Blasius, Jörg. 2006. *Multiple Correspondence Analysis and Related Analysis*. Chapman & Hall.
- HY. 2016. *Helsingin yliopiston strategia 2017-2020*.
- Isotalo, Jarkko. 2012. *Johdatus yleistettyihin lineaarisiin malleihin*. Tampereen yliopisto.

- Kangas, Tuukka. 2017. *Edustavuuden arviointi maisteriuraseuranta-aineistossa*. Helsingin yliopisto.
- Kass, Robert, & Raftery, Adrian. 1995. Bayes Factors. *Journal of the American Statistical Association*, **90**(430), 773–795.
- McCullagh, Peter, & Nelder, John Ashworth. 1989. *Generalized Linear Models*. 2. edn. Chapman and Hall.
- Nair, Sid, Arun, Patil, & Mertova, Patricie. 2011. Enhancing the quality of engineering education by utilising student feedback. *European Journal of Engineering Education*, **36**(1), 3–12.
- Nenadic, Oleg, & Greenacre, Michael. 2007. Correspondence Analysis in R, with Two- and Three-dimensional Graphics: The ca Package. *Journal of Statistical Software*, **20**(3).
- OKM. *Ehdotus laadullisen työllistymisen sisällyttämiseksi korkeakoulujen rahoitusmalleihin 1.1.2019 alkaen*. Opetus ja kulttuuriministeriö.
- Puhakka, Antero, Rautopuro, Juhani, & Tuominen, Visa. 2010. Employability and Finnish University Graduates. *European Educational Research Journal*, **9**(1), 45–55.
- Raftery, Adrian. 1995. Bayesian Model Selection in Social Research. *Sociological Methodology*, **25**, 111–163.
- Sainio, Juha, Carver, Eric, & Kangas, Tuukka. 2017. *Eväitä hyvän työuran rakentamiseen*. Aarresaari-verkosto.
- Schwarz, Gideon. 1978. Estimating the Dimension of a Model. *The Annals of Statistics*, **6**(2), 461–464.

- Semeijn, Judith, van der Velden, Rolf, Heijke, Hans, van der Vleuten, Cees, & Boshuizen, Henny. 2006. Competence indicators in academic education and early labour market success of graduates in health sciences. *Journal of Education and Work*, **19**(4), 383–413.
- Slotte, Virpi, Nieminen, Juha, Lonka, Kirsti, & Olkinuora, Erkki. 2006. *Higher Education and Working Life - Collaborations, Confrontations and Challenges*. Elsevier. Chap. From University to Working Life: Graduates' Workplace Skills in Practice, pages 72–88.
- Sourial, N., Wolfson, C., Zhu, B., Quail, J., Fletcher, J., Karunanathan, S., & Bergman, H. 2010. Correspondence analysis is a useful tool to uncover the relationships among categorical variables. *Journal of Clinical Epidemiology*, **63**(6), 638–646.
- Staddon, Elizabeth, & Standish, Paul. 2012. Improving the Student Experience. *Journal of Philosophy of Education*, **46**(4), 631–648.
- Taalu, Heikki. 2017. *Tilastoja korkeasti koulutettujen työttömyydestä*. Akava.
- Tuominen, Visa. 2012. Henkilökohtaiset ominaisuudet maistereiden työllistymisessä. *Tiedepolitiikka*, 33–42.
- Vila, Luis, Garcia-Aracil, Adela, & Mora, Jose-Gines. 2007. The Distribution of Job Satisfaction Among Young European Graduates: Does the Choice of Study Field Matter? *The Journal of Higher Education*, **78**(1), 97–118.
- Wong, S., & Li, W. K. 1998. A note on the corrected Akaike information criterion for threshold autoregressive models. *Journal of Time Series Analysis*, **19**(1), 113–124.

Liitteet

Liite A

Kyselylomake

Vastaa verkossa:

aarresaari.net/uraseuranta

Tarvitset vastaukseen
tämän numeron:

Uraseuranta vuonna 2011 valmistuneille

Vastaa kysymyksiin valitsemalla sopivimman vaihtoehdon numero tai kirjoittamalla vastaus sille varattuun tilaan.

TAUSTAKYSYMYKSET

Saamme opintorekisteristä tiedot 2011 suorittamastasi tutkinnosta. Haluaisimme kuitenkin tietää mahdollisista muista opinnoistasi.

1. Olitko suorittanut muita ammatillisia tai korkeasteen tutkintoja ennen vuonna 2011 suorittamaasi tutkintoa?

- 1 en
- 2 kyllä, tutkinnon/tutkintojen nimet _____

2. Oletko osallistunut koulutukseen vuonna 2011 suorittamasi tutkinnon jälkeen?

Valitse sopivat vaihtoehdot.

- 1 Olen osallistunut/suorittanut ammatillisen erikoistumiskoulutuksen, pätevyitysmiskoulutuksen tms. Minkä? _____
- 2 Olen suorittanut muun tutkinnon. Minkä? _____
- 3 Olen suorittanut opintoja tähtäimessä toinen korkeakoulututkinto. Miltä alalta/mikä tutkinto? _____
- 4 Olen suorittanut tieteellisiä/taiteellisia jatko-opintoja. Miltä alalta? _____
- 5 Olen suorittanut lyhyempiä koulutuksia tai kursseja.
- 6 En ole osallistunut koulutukseen.

TYTYTYVÄISYYS TUTKINTOON

Seuraavien kysymysten avulla pyrimme saamaan kuvan siitä, mitä ajattelet tutkinnostasi nyt, viisi vuotta valmistumisen jälkeen. Käytämme tietoa koulutuksen kehittämiseen.

3. Arvioi vuonna 2011 suorittamaasi tutkintoa seuraavien väittämien perusteella.

Valitse sopivin vaihtoehto. 1 = täysin eri mieltä, 2 = eri mieltä, 3 = hieman eri mieltä, 4 = hieman samaa mieltä, 5 = samaa mieltä, 6 = täysin samaa mieltä

- | | | | | | | |
|--|---|---|---|---|---|---|
| 1. Työnantajat arvostavat tutkintoani. | 1 | 2 | 3 | 4 | 5 | 6 |
| 2. Suosittelisin koulutustani muille. | 1 | 2 | 3 | 4 | 5 | 6 |
| 3. Opintojen osaamistavoitteet tuotiin selkeästi esille. | 1 | 2 | 3 | 4 | 5 | 6 |
| 4. Koulutus antoi riittävät valmiudet työelämään. | 1 | 2 | 3 | 4 | 5 | 6 |
| 5. Koulutuksen aikana tuotiin esille yrittäjyys uravaihtoehtona. | 1 | 2 | 3 | 4 | 5 | 6 |

4. Miten tyytyväinen olet kokonaisuudessaan vuonna 2011 suorittamaasi tutkintoon työurasi kannalta?

- 1 erittäin tyytymätön
- 2 tyytymätön
- 3 hieman tyytymätön
- 4 melko tyytyväinen
- 5 tyytyväinen
- 6 erittäin tyytyväinen

Perustele vastaustasi. Miksi näin on?

Tarvittaessa lisää kirjoitustilaa viimeisellä sivulla.

TYÖURAN KOKONAISUUS

Seuraavat kysymykset liittyvät urapolkusi vaiheisiin. Meille on tärkeää saada tietoa myös työuran katkoksista.

5. Mikä seuraavista vaihtoehdoista kuvaa parhaiten tähänastisen työurasi kokonaisuutta?

- 1 Yhtäjaksoisesti saman työnantajan palveluksessa tai yrittäjänä valmistumisesta lähtien.
- 2 Useita eri työnantajia, määräaikaishuuksia tai toimeksiantoja tai työskentelyä apurahalla. Ei juuri katkoksia.
- 3 Vaihtuvia työnantajia ja tehtäviä, joiden välissä katkoksia, opiskelua tai työttömyysjaksoja.
- 4 Työttömyyttä, jota satunnaiset määräaikaishuudet, harjoittelut sekä keikka- tai freelancetyöt pilkkovat.
- 5 Pääasiassa työvoiman ulkopuolella: esim. opiskelua ja/tai vanhempainvapaata valtaosan ajasta.
- 6 Joku muu, mikä? _____

6. Kuinka monen työnantajan palveluksessa olet ollut valmistumisesi jälkeen (ml. nykyinen työnantajasi)?

_____ kpl.

7. Oletko valmistumisesi jälkeen toiminut yrittäjänä, ammatinharjoittajana tai freelancerina? Valitse parhaiten tilannettasi kuvaava vaihtoehto.

- 1 en
- 2 Kyllä, suurin osa/koko toimeentuloni koostuu yrittäjänä/freelancerina tehdystä työstä.
- 3 Kyllä, suurin osa/koko toimeentuloni koostui aiemmin yrittäjänä/freelancerina tehdystä työstä, mutta nyt olen työsuhteessa.
- 4 Kyllä, tein/teen silloin tällöin toimeksiantoja/freelancesuhteisia töitä, mutta ne eivät ole päätulonlähteeni, sillä olen myös palkkatyössä.
- 5 Kyllä, teen toimeksiantoja/freelancetöitä, mutta ne eivät riitä turvaamaan toimeentuloani, en ole palkkatyössä.

8. Jos olet toiminut yrittäjänä, a. kuinka monessa yrityksessä olet ollut perustajana/osakkaana valmistumisesi jälkeen?

_____ yritykseen.

b. kuinka monta henkilöä yrityksesi ovat yhteensä työllistäneet itsesi lisäksi?

_____ henkilöä.

9. Oletko ollut valmistumisesi jälkeen poissa työelämästä

a. perhevapaiden vuoksi?

- 1 en
- 2 kyllä, kesto _____ v _____ kk.

b. työttömyyden (ml. lomautukset) vuoksi?

- 1 en
- 2 kyllä _____ kertaa, kesto _____ v _____ kk.

c. muun syyn vuoksi?

- 1 en
- 2 kyllä, syy _____, kesto _____ v _____ kk.

10. Kuinka tyytyväinen olet tähänastiseen työuraasi?

- 1 erittäin tyytymätön
- 2 tyytymätön
- 3 hieman tyytymätön
- 4 melko tyytyväinen
- 5 tyytyväinen
- 6 erittäin tyytyväinen

Perustele vastaustasi.

Tarvittaessa lisää kirjoitustilaa viimeisellä sivulla.

TILANNE VALMISTUMISHETKELLÄ JA NYT

Seuraavat kysymykset koskevat ensimmäistä valmistumisen jälkeistä työpaikkaa ja nykyhetkeä. Tietoja käytetään mm. opiskelijoiden uraohjauksessa konkreettisina esimerkkeinä työnantajista ja aloista, joihin eri alojen opiskelijat ovat työllistyneet.

OHJEITA

- Ensimmäinen valmistumisen jälkeinen työ on voinut alkaa jo ennen valmistumista (kysymykset 12.–17.)
- Jos olet tällä hetkellä esim. **perhe- tai opintovapaalla**, vastaa sen työpaikan perusteella, josta jäit vapaalle.
- Jos **et ole tällä hetkellä työssä tai et ollut työssä**, kun jäit esim. perhevapaalle, jätä nykyistä työtä koskevat kohdat tyhjäksi.
- **Jos olet yrittäjä**, ammatinharjoittaja tai freelancer, vastaa kysymyksiin soveltuvin osin.

11. Mikä seuraavista vaihtoehtoista kuvasi/kuvaa parhaiten tilannettasi

a. valmistumishetkellä?

Valitse yksi vaihtoehto.

- 1 vakituinen kokopäivätyö
- 2 määräaikainen kokopäivätyö
- 3 osa-aikatyö
- 4 itsenäinen yrittäjä/ammattiharjoittaja/freelancer (oma Y-tunnus)
- 5 useita rinnakkaisia työsuhteita (työskentely verokortilla)
- 6 työllistetty/töharjoittelu
- 7 työtön työnhakija
- 8 työvoimakoulutus tai vastaava
- 9 päätoiminen opiskelu (johtaa tutkintoon tai arvosanaan)
- 10 perhevapaa (työsuhteesta)
- 11 perhevapaa (ei työsuhdetta)
- 12 työskentely apurahalla
- 13 työvoiman ulkopuolella (asevelvollisuus tms.)
- 14 muu, mikä? _____

b. tällä hetkellä?

Valitse yksi vaihtoehto.

- 1 vakituinen kokopäivätyö
- 2 määräaikainen kokopäivätyö
- 3 osa-aikatyö
- 4 itsenäinen yrittäjä/ammattiharjoittaja/freelancer (oma Y-tunnus)
- 5 useita rinnakkaisia työsuhteita (työskentely verokortilla)
- 6 työllistetty/töharjoittelu
- 7 työtön työnhakija
- 8 työvoimakoulutus tai vastaava
- 9 päätoiminen opiskelu (johtaa tutkintoon tai arvosanaan)
- 10 perhevapaa (työsuhteesta)
- 11 perhevapaa (ei työsuhdetta)
- 12 työskentely apurahalla
- 13 työvoiman ulkopuolella (asevelvollisuus tms.)
- 14 muu, mikä? _____

12. Mikä oli/on päätyönantajasi

a. ensimmäisessä työssä valmistumisen jälkeen?

- Valitse yksi vaihtoehto
- 1 kunta tai kuntayhtymä
 - 2 valtio
 - 3 suuri yritys (250 työntekijää tai enemmän)
 - 4 pieni tai keskisuuri yritys (alle 250 työntek.)
 - 5 oma yritys/työllistän itse itseni
 - 6 järjestö, säätiö tai vastaava
 - 7 yliopisto
 - 8 ammattikorkeakoulu
 - 9 muu, mikä?

b. tällä hetkellä?

- Valitse yksi vaihtoehto
- 1 kunta tai kuntayhtymä
 - 2 valtio
 - 3 suuri yritys (250 työntekijää tai enemmän)
 - 4 pieni tai keskisuuri yritys (alle 250 työntek.)
 - 5 oma yritys/työllistän itse itseni
 - 6 järjestö, säätiö tai vastaava
 - 7 yliopisto
 - 8 ammattikorkeakoulu
 - 9 muu, mikä?

13. Mitkä seuraavista kuvasivat/kuvaavat parhaiten työtehtäviesi luonnetta?

a. ensimmäisessä työssä valmistumisen jälkeen?

- Voit valita useita vaihtoehtoja.
- 1 tutkimus
 - 2 opetus tai kasvatus
 - 3 johto- ja esimiestehtävät
 - 4 konsultointi tai koulutus
 - 5 asiakastyö/potilastyö
 - 6 markkinointi ja myynti
 - 7 suunnittelu-, kehitys- tai hallintotehtävät
 - 8 viestintä- ja mediatyö
 - 9 toimistotehtävät
 - 10 taiteellinen työ
 - 11 lainopillinen työ
 - 12 kirkollinen työ
 - 13 rahoituksen ja taloushallinnon tehtävät
 - 14 muu, mikä?

b. tällä hetkellä?

- Voit valita useita vaihtoehtoja.
- 1 tutkimus
 - 2 opetus tai kasvatus
 - 3 johto- ja esimiestehtävät
 - 4 konsultointi tai koulutus
 - 5 asiakastyö/potilastyö
 - 6 markkinointi ja myynti
 - 7 suunnittelu-, kehitys- tai hallintotehtävät
 - 8 viestintä- ja mediatyö
 - 9 toimistotehtävät
 - 10 taiteellinen työ
 - 11 lainopillinen työ
 - 12 kirkollinen työ
 - 13 rahoituksen ja taloushallinnon tehtävät
 - 14 muu, mikä?

14. Valitse edellisen kysymyksen vaihtoehdoista se, joka kuvaa parhaiten pääasiallisinta työtehtävääsi

a. ensimmäisessä työssä valmistumisen jälkeen?

Merkitse vain yhden vaihtoehdon numero

b. tällä hetkellä?

Merkitse vain yhden vaihtoehdon numero

15. Mikä oli/on työsi ammatti-, tehtävä- tai virkanimike

a. ensimmäisessä työssä valmistumisen jälkeen?

b. tällä hetkellä?

16. Mikä oli/on päätyönantajasi nimi

a. ensimmäisessä työssä valmistumisen jälkeen?

b. tällä hetkellä?

Työnantajien nimiä käytetään ainoastaan opiskelijoiden uraohjauksessa esimerkkinä alan työnantajista.

17. Millä paikkakunnalla työpaikkasi/yrityksesi sijaitsee/sijaitsee

a. ensimmäisessä työssä valmistumisen jälkeen?

b. tällä hetkellä?

Jos olet työssä ulkomailla, kirjoita paikkakunnan sijaan maa.

TYÖLLISTYMISEN LAATU

Seuraavien kysymysten avulla seuraamme, miten yliopistoista valmistuneet kokevat työnsä suhteessa koulutukseensa.

18. Oliko vuonna 2011 suorittamasi ylempi (alempi) korkeakoulututkinto vaatimuksena nykyiseen työhösi?

- 1 ei
- 2 kyllä
- 3 en tiedä

19. Miten hyvin pystyit/pystyt hyödyntämään yliopistossa oppimiasi tietoja ja taitoja

a. ensimmäisessä työssä valmistumisen jälkeen?

- Valitse yksi vaihtoehto
- 1 en voinut hyödyntää juuri lainkaan
 - 2 jonkin verran/osittain
 - 3 hyödynsin jatkuvasti

b. tällä hetkellä?

- Valitse yksi vaihtoehto
- 1 en voi hyödyntää juuri lainkaan
 - 2 jonkin verran/osittain
 - 3 hyödynnän jatkuvasti

20. Miten hyvin työsi vastasi/vastaa vaativuustasoltaan yliopistollista koulutustasi

a. ensimmäisessä työssä valmistumisen jälkeen?

Valitse yksi vaihtoehto

- 1 Työn vaativuustaso oli koulutustasoani selvästi alhaisempi.
- 2 Työn vaativuustaso oli koulutustasoani osittain alhaisempi.
- 3 Työ vastasi hyvin koulutustasoani.
- 4 Työ oli koulutukseeni nähden vaativampaa.

b. tällä hetkellä?

Valitse yksi vaihtoehto

- 1 Työn vaativuustaso on koulutustasoani selvästi alhaisempi.
- 2 Työn vaativuustaso on koulutustasoani osittain alhaisempi.
- 3 Työ vastaa hyvin koulutustasoani.
- 4 Työ on koulutukseeni nähden vaativampaa.

21. Mikäli nykyinen työsi ei vastaa koulutustasoasi, mikä oli tärkein syy työn vastaanottamiseen?

- 1 Jatkan työssä, jossa olin jo ennen valmistumistani.
- 2 En ole saanut koulutustasoani vastaavaa työtä.
- 3 Työ on mielenkiintoisempaa.
- 4 Palkka ja/tai työehdot ovat paremmat.
- 5 Muu koulutus on ohjannut minut nykyisiin tehtäviini.
- 6 En ole halunnut toimia koulutustasoani vastaavissa töissä.
- 7 Muu syy, mikä? _____

22. Mikä on keskimääräinen bruttopalkkasi tai kuukausitulosi (säännölliset lisät, luontoisetujen verotusarvot ja ylityökorvaukset mukaan lukien)?

TYÖ JA OSAAMINEN

Tämän osion kysymyksillä selvitämme muun muassa, miten yliopisto-opinnot ovat kehittäneet taitoja, joita tarvitset työssäsi.

OHJEITA:

- Jos olet esim. perhe- tai opintovapaalla, vastaa sen **työpaikan perusteella, josta jäit vapaalle.**
- **Jos olet yrittäjä**, ammatinharjoittaja tai freelancer, vastaa kysymyksiin soveltuvin osin.

23. Arvioi, miten seuraavat tekijät ovat vaikuttaneet työllistymiseesi valmistumisen jälkeen? Vastaa koko valmistumisen jälkeisen työurasi perusteella.

Valitse sopivin vaihtoehto.

- 1 = ei lainkaan tärkeää
2 = vain vähän merkitystä
3 = jonkin verran merkitystä
4 = melko tärkeää
5 = tärkeää
6 = erittäin tärkeää
0 = vaihtoehto ei koske minua.

HUOM! Jos sinulla ei esimerkiksi ole lainkaan kansainvälistä kokemusta, valitse valinta 0.

1. tutkintosi aineyhdistelmä	1	2	3	4	5	6	0
2. harjoittelu	1	2	3	4	5	6	0
3. muu työkokemus	1	2	3	4	5	6	0
4. muut opinnot tai koulutukset (ei vuonna 2011 suorittamaasi tutkintoon kuuluvat)	1	2	3	4	5	6	0
5. järjestötyöhön tai harrastuksiin liittyvä kokemus	1	2	3	4	5	6	0
6. kansainvälinen kokemus	1	2	3	4	5	6	0
7. kontaktit/suhdeverkostot	1	2	3	4	5	6	0
8. kyky kertoa omasta osaamisesta	1	2	3	4	5	6	0
9. aktiivisuus ja profiloituminen sosiaalisessa mediassa	1	2	3	4	5	6	0
10. muu, mikä? _____	1	2	3	4	5	6	0

Perustele halutessasi vastaustasi. Voit kertoa myös muista työllistymiseesi vaikuttaneista tekijöistä.

Tarvittaessa lisää kirjoitustilaa viimeisellä sivulla.

24. Kuinka tärkeitä seuraavat tiedot ja taidot ovat nykyisessä työssäsi? Miten yliopisto-opiskelu kehitti kyseisiä työelämävalmiuksia?

OHJE: Jos et ole tällä hetkellä työelämässä, arvioi viimeisimmän työpaikkasi mukaan. Valitse jokaisesta sopivimmat vaihtoehdot.

	Kuinka tärkeitä seuraavat tiedot ja taidot ovat nykyisessä työssäsi?						Miten yliopisto-opiskelu kehitti kyseisiä työelämävalmiuksia?					
	1	2	3	4	5	6	1	2	3	4	5	6
1. opinnoista saatu teoreettinen osaaminen	1	2	3	4	5	6	1	2	3	4	5	6
2. opinnoista saatu käytännön osaaminen	1	2	3	4	5	6	1	2	3	4	5	6
3. analyttiset, systemaattisen ajattelun taidot	1	2	3	4	5	6	1	2	3	4	5	6
4. tiedonhankintataidot	1	2	3	4	5	6	1	2	3	4	5	6
5. ongelmanratkaisutaidot	1	2	3	4	5	6	1	2	3	4	5	6
6. organisointi- ja koordinoititaidot	1	2	3	4	5	6	1	2	3	4	5	6
7. projektinhallintataidot	1	2	3	4	5	6	1	2	3	4	5	6
8. esimiestaidot	1	2	3	4	5	6	1	2	3	4	5	6
9. yhteistyötaidot	1	2	3	4	5	6	1	2	3	4	5	6
10. neuvottelutaidot	1	2	3	4	5	6	1	2	3	4	5	6
11. esiintymistäidot	1	2	3	4	5	6	1	2	3	4	5	6
12. opetus-, koulutus- ja ohjaustaidot	1	2	3	4	5	6	1	2	3	4	5	6
13. toimiminen monikulttuurisessa ympäristössä	1	2	3	4	5	6	1	2	3	4	5	6
14. lainsäädännön tuntemus	1	2	3	4	5	6	1	2	3	4	5	6
15. liiketoiminnan/taloushallinnon perusteiden tuntemus	1	2	3	4	5	6	1	2	3	4	5	6
16. tieto- ja viestintätekniikan taidot	1	2	3	4	5	6	1	2	3	4	5	6
17. viestintä suomen kielellä	1	2	3	4	5	6	1	2	3	4	5	6
18. viestintä ruotsin kielellä	1	2	3	4	5	6	1	2	3	4	5	6
19. viestintä englannin kielellä	1	2	3	4	5	6	1	2	3	4	5	6
20. viestintä muilla kielillä, millä? _____	1	2	3	4	5	6	1	2	3	4	5	6
21. kyky oppia ja omaksua uutta	1	2	3	4	5	6	1	2	3	4	5	6
22. luovuus	1	2	3	4	5	6	1	2	3	4	5	6
23. tieteidenvälisyys/moniammatillisissa ryhmissä toimiminen	1	2	3	4	5	6	1	2	3	4	5	6
24. stressinsietokyky	1	2	3	4	5	6	1	2	3	4	5	6
25. verkostoitumistaidot	1	2	3	4	5	6	1	2	3	4	5	6
26. itseohjautuvuus/oma-aloitteisuus	1	2	3	4	5	6	1	2	3	4	5	6

25. Minkälaisia taitoja tai osaamista kannustaisit nykyisiä opiskelijoita hankkimaan tulevaisuuden työelämää ajatellen?

Tarvittaessa lisää kirjoitustilaa viimeisellä sivulla.

Liite B

Tyytyväisyys koulutusaloittain

Koulutusala	2007	2009	2011	Yhteensä
Humanistinen	50,6 %	48,7 %	44,9 %	48,5 %
Kasvatustieteellinen	69,3 %	69,6 %	65,9 %	68,4 %
Luonnontieteellinen	61,2 %	56,4 %	57,0 %	58,8 %
Lääketieteellinen	81,8 %	89,4 %	80,0 %	84,3 %
Oikeustieteellinen	78,6 %	76,0 %	72,0 %	75,8 %
Psykologian	72,7 %	90,0 %	80,0 %	78,8 %
Teologinen	63,9 %	39,5 %	58,8 %	56,5 %
Yhteiskuntatieteellinen	49,7 %	52,7 %	53,3 %	51,6 %
Farmasian	69,0 %	80,0 %	61,7 %	71,0 %
Maatalous-metsätieteellinen	48,7 %	57,5 %	45,3 %	49,1 %
Eläinlääketieteellinen	88,5 %	68,8 %	61,9 %	73,4 %
Hammaslääketieteellinen	66,7 %	90,0 %	57,1 %	71,9 %
Kaikki	61,5 %	62,8 %	57,7 %	60,7 %

Taulukko B.1: Vastausjakaumat koulutusaloittain

Liite C

Logistinen regressio: koulutusalat

Muuttuja	Estimaatti	Keskivirhe	OR (95 %)	Wald	p-arvo
Vakio	-4,331	1,215	1,20 - 1,73	12,7	< 0,000
Humanistinen	-0,436	0,130	0,51 - 0,83	11,3	0,001
Kasvatustieteellinen	-0,431	0,143	1,15 - 1,96	9,1	0,003
Oikeustieteellinen	0,733	0,176	1,57 - 3,05	17,4	< 0,000
Psykologia	1,059	0,383	1,34 - 5,40	7,6	0,006
Teologia	-0,47	0,185	0,64 - 1,28	0,065	0,798
Yhteiskuntatieteellinen	-0,335	0,140	0,58 - 0,96	5,73	0,017
Farmasia	0,477	0,180	1,21 - 2,41	7,02	0,008
Maatalous-metsätieteellinen	-0,458	0,168	0,49 - 0,92	7,45	0,006
Eläinlääketieteellinen	0,593	0,279	1,14 - 3,33	4,5	0,034
Hammaslääketieteellinen	0,583	0,426	0,83 - 4,12	1,9	0,171
Lääketieteellinen	1,349	0,280	2,28 - 6,35	23,2	< 0,000

Taulukko C.1: Logistisen regression parametrit. Selittävänä muuttujana koulutusala, vertailuryhmänä on käytetty luonnontieteellistä alaa.

Liite D

Tutkintotyytyväisyys koulutusaloittain ja työttömyyden suhteen

Koulutusala	On ollut työtön	Tyytyväisiä
Humanistinen	Ei	54,4 %
	Kyllä	41,4 %
Kasvatustieteellinen	Ei	72,3 %
	Kyllä	61,4 %
Luonnontieteellinen	Ei	67,4 %
	Kyllä	42,0 %
Lääketieteellinen	Ei	84,5 %
	Kyllä	100 %
Oikeustieteellinen	Ei	78,5 %
	Kyllä	65,6 %
Psykologian	Ei	81,3 %
	Kyllä	80,0 %
Teologinen	Ei	68,0 %
	Kyllä	42,9 %
Yhteiskunnallinen	Ei	54,3 %
	Kyllä	44,8 %
Farmasian	Ei	71,2 %
	Kyllä	57,1 %
Maatalous-metsätieteellinen	Ei	54,0 %
	Kyllä	39,5 %
Eläinlääketieteellinen	Ei	74,2 %
	Kyllä	63,6 %
Hammaslääketieteellinen	Ei	75,0 %
	Kyllä	0,0 %
Yhteensä	Ei	66,6 %
	Kyllä	47,6 %

Taulukko D.1: Vastausjakaumat koulutusaloittain ja työttömyyden mukaan.

Liite E

Täyden mallin parametrit

Muuttuja	Aineisto	OR	OR:n luottamusväli	Devianssi	p-arvo
Sukupuoli (nainen)	yhd	0,94	0,76-1,16	0,089	> 0,1
	2011	0,63	0,41-0,97	5,4	> 0,01
ikä alle 27	yhd	0,99	0,80-1,23	0,22	> 0,1
	2011	0,94	0,61-1,46	1,4	> 0,1
ikä yli 29	yhd	1,58	1,26-1,98	16	≤ 0,01
	2011	1,54	0,96 - 2,49	5,0	> 0,01
kouala (hum.)	yhd	0,60	0,44-0,82	35	≤ 0,01
	2011	0,69	0,36-1,30	4,1	> 0,01
kouala (kasv.)	yhd	0,89	0,62-1,27	3,1	> 0,05
	2011	0,95	0,45-2,01	6,7	≤ 0,01
kouala (lääk.)	yhd	2,19	1,24-4,05	26	≤ 0,01
	2011	0,64	0,17-3,18	4,4	> 0,01
kouala (oik.)	yhd	1,79	1,15-2,81	21	≤ 0,01
	2011	0,78	0,37-1,71	4,8	> 0,01

kouala (psy.)	yhd	1,91	0,89-4,47	8,9	$\leq 0,01$
	2011	1,11	0,26-7,78	8,1	$\leq 0,01$
kouala (teol.)	yhd	0,94	0,61-1,45	0,0	$> 0,1$
	2011	1,04	0,42-2,62	0,77	$> 0,1$
kouala (yhdk.)	yhd	0,76	0,55-1,06	15	$\leq 0,01$
	2011	0,72	0,39-1,32	0,70	$> 0,1$
kouala (farm.)	yhd	1,17	0,76-1,81	6,6	$> 0,01$
	2011	0,47	0,15-0,91	0,41	$> 0,1$
kouala (maamet.)	yhd	0,67	0,46-0,99	12	$\leq 0,01$
	2011	0,55	0,25-1,17	4,5	$> 0,01$
kouala (eläin.)	yhd	1,23	0,65-2,40	3,6	$> 0,05$
	2011	0,34	0,099-1,20	0,015	$> 0,1$
kouala (hammas.)	yhd	1,12	0,46-3,03	1,7	$> 0,1$
	2011	0,21	0,032-1,50	0,231	$> 0,1$
jatko-opin.	yhd	0,79	0,62-1,00	0,0	$> 0,1$
	2011	0,81	0,50-1,33	1,4	$> 0,1$
työtön	yhd	0,63	0,52-0,77	29	$\leq 0,01$
	2011	0,64	0,42-0,97	11	$\leq 0,01$
opp. hyöd.	yhd	2,55	2,10-3,08	173	$\leq 0,01$
	2011	2,60	1,74-3,88	66	$\leq 0,01$
vaat	yhd	1,66	1,29-2,14	13	$\leq 0,01$
	2011	1,67	0,98-2,86	8,1	$\leq 0,01$
riit. valm.	2011	3,48	2,27-5,44	41	$\leq 0,01$
ta. arv.	2011	5,42	3,60-8,26	69	$\leq 0,01$
keh. teor	yhd	1,42	1,03-1,94	18	$\leq 0,01$

keh. anal	yhd	1,05	0,77-1,42	7,9	$\leq 0,01$
keh. tieh	yhd	1,04	0,78-1,38	5,0	$> 0,01$
keh. orat	yhd	1,18	0,94-1,47	13	$\leq 0,01$
keh. neuv	yhd	1,19	0,94-1,50	20	$\leq 0,01$
keh. org	yhd	1,30	1,05-1,62	14	$\leq 0,01$
keh. esim	yhd	1,35	0,92-2,02	5,3	$> 0,01$
keh. proj	yhd	1,05	0,85-1,30	1,5	$> 0,1$
keh. laki	yhd	1,10	0,87-1,41	1,9	$> 0,1$
keh. viest	yhd	1,34	1,10-1,63	12	$\leq 0,01$
keh. suom	yhd	0,93	0,73-1,18	0,0	$> 0,1$
keh. ruots	yhd	1,21	0,98-1,49	3,8	$> 0,05$
keh. eng	yhd	0,92	0,74-1,14	0,4	$> 0,1$
keh. esiin	yhd	1,17	0,95-1,44	4,9	$> 0,01$
keh. opet	yhd	1,35	1,08-1,69	7,1	$\leq 0,01$

Taulukko E.1: Tunnuslukuja logistisen regression ensimmäisistä malleista. Lyhenne *yhd* viittaa yhdistelmäaineistoon, joka koostuu vuosina 2007-2011 valmistuneiden aineistoista. P-arvot on raportoitu tasoilla $>0,1$, $> 0,05$, $> 0,01$ ja $\leq 0,01$