

Korkeakouluopiskelijan
koulutustasotoiveeseen yhteydessä
olevat tekijät

LuK-tutkielma
Susanna Pätsi
2583444
Matemaattisten tieteiden tutkinto-ohjelma
Oulun yliopisto
Kevät 2021

Sisällys

1	Aineisto ja menetelmät	3
1.1	Aineisto	3
1.2	Mittausmenetelmät ja muuttujien käsittely	3
1.3	Tilastollisen analyysin menetelmät	7
1.3.1	Ristiintaulukointi ja χ^2 -testaus	7
1.3.2	Logistinen regressio	9
2	Tulokset	11
2.1	Vanhempien koulutustason yhteys jälkeläisen koulutustasotoiveeseen	11
2.1.1	Isän koulutustaso	11
2.1.2	Äidin koulutustaso	14
2.2	Tahdonvoiman, iän ja ahkeruuden yhteydet opiskelijan koulutustasotoiveeseen	16
3	Pohdinta	18

Johdanto

Nykyään opiskelijoilla on valittavana monia eri koulutustasoja aina peruskoulutuksesta toisen asteen koulutukseen, siitä korkeakoulutukseen sekä ylimpänä tohtoritason tutkintoon. Tähän valintaan vaikuttavia tekijöitä on monia, joista yksi on vanhempien tausta. Vanhemmilta opitut arvot, näkymykset ja asenteet vaikuttavat ympäristön lisäksi nuoren ajatusmaailmaan, joka ohjaa jälkeläisen opintopolkua. Tietokeskuksen tutkimuksen mukaan vanhemmilta periytyy koulutusalojen mukana myös mahdollisesti koulutustaso [1]. Aiemmissä tutkimuksissa on seurattu jälkeläisten matkaa korkeimpaan koulutustasoon asti. Tämän tutkimuksen yhtenä tarkoituksena onkin pohtia, onko vanhempien korkeimmalla koulutustasolla yhteyttä jälkeläisen koulutustasotoiveeseen, kun tarkastellaan korkeakouluopiskelijoita. Ensimmäinen tutkimuskysymys on jaettu kahteen osaan:

1.a Onko isän korkein koulutustaso yhteydessä jälkeläisen korkeimpaan koulutustasotoiveeseen?

1.b Onko äidin korkein koulutustaso yhteydessä jälkeläisen korkeimpaan koulutustasotoiveeseen?

Toisaaltaan voidaan miettiä, vaikuttaako opiskelijassa itsessään olevat ominaisuudet koulutustasotoiveeseen. Tehdyissä tutkimuksissa on havaittu, että opiskelijan aiempi koulumenestys on ollut yhteydessä opiskelijan opintopolkuun vanhempien taustan lisäksi [2]. Opintomenestykseen liittyy keskeisesti motivaatio ja tahdonvoima, sillä ne vaikuttavat oppimiseen [3]. Olisi siis järkevää miettiä, onko näillä ominaisuuksilla suoranaista yhteyttä opiskelijan koulutustasotoiveeseen.

Kuinka paljon tahdonvoimalla on vaikutusta opiskelijan toiveeseen? Miten ikä vaikuttaa tähän? Toivovatko nuoremmat enemmän korkeaa koulutusta ja vanhemmat alempaa koulutusta? Entä jos opiskelija uskoo saavuttavansa menestyksen työnteolla? Onko siis ahkeruudella vaikutusta toiveeseen? Vai viekö onni opiskelijan menestykseen? Näitä kysymyksiä tarkastellaan toisessa tutkimuskysymyksessä:

2. Miten tahdonvoima, ikä ja ahkeruus ovat yhteydessä opiskelijan koulutustasotoiveeseen?

Tutkimuksen ideana on siis selvittää erilaisten ominaisuuksien ja taustatekijöiden yhteyttä korkeakouluopiskelijan korkeimpaan koulutustasotoiveeseen käyttämällä erilaisia tilastollisia menetelmiä.

1 Aineisto ja menetelmät

1.1 Aineisto

Tutkielmassa on käytetty Suomalaisten korkeakouluopiskelijoiden identiteettihorisontit 2015 -aineistoa [4]. Aineisto on osa kansainvälistä hanketta nimeltä "The role of identity horizons in education-to-work transitions - A cross-cultural study of students in higher education in Finland, Japan and Northern America".

Aineiston perusjoukkona on suomalaiset 18-24-vuotiaat korkeakouluopiskelijat. Aineiston asettelu on epäkokeellinen. Otantamenetelmässä ei käytetty todennäköisyysotantaa. Kutsu kyselyyn on lähetetty kaikkiin suomalaisiin ammattikorkeakoulujen sekä yliopistojen opiskelijaverkkoihin, jolloin jokainen opiskelija on voinut halutessaan vastata kyselyyn. Aikaulottuvuus aineistossa on poikittainen ja sen ajankohtana on kevät 2015. Aineiston muuttajat koskevat korkeakouluopiskelijoiden identiteettiä, arvoja, kokemuksia ja tulevaisuuden suunnitelmia. Näistä aiheista kysyttiin 263 muuttujalla 14 eri patterissa. Kokonaisuudessaan havaintoyksiköitä on 730 ja vastauskato on koko aineistossa 14 %.

1.2 Mittausmenetelmät ja muuttujien käsittely

Tutkimukseen on valittu kuusi muuttujaa 263:n joukosta: ikä, opiskelijan korkein koulutustasotoive, isän korkein koulutustaso, äidin korkein koulutustaso, sekä kaksi mielipidettä koskevaa muuttujaa "Minulla on paljon tahdonvoimaa" sekä "Menestys saavutetaan kovalla työllä. Onnella on vähän tai ei mitään tekemistä sen kanssa".

Vanhempien koulutustasomuuttujissa on otettu huomioon se, että vastaaja on valinnut itselleen sopivat isä- ja äitihahmot. Tutkimuksen selittämisen helpottamista varten valitaan muuttujalle "Minulla on paljon tahdonvoimaa" lyhempi kuvaus *tahdonvoima* ja muuttujalle "Menestys saavutetaan kovalla työllä. Onnella on vähän tai ei mitään tekemistä sen kanssa" kuvaus *ahkeruus*. Tutkimuksessa käytetään ilmaisua *koulutustasomuuttajat*, kun puhutaan isän ja äidin korkeimmasta koulutustasoista sekä jälkeläisen korkeimmasta koulutustasotoiveesta.

Kaikki muut, paitsi ikä, ovat luokiteltuja muuttujia. Ikä on suhdeasteikon muuttuja. Koulutustasomuuttajat on luokiteltu samoihin 17 eriin luokkaan, tahdonvoima on viisiluokkainen ja ahkeruus kuusiluokkainen taulukon 1 mukaisesti.

Luokiteltujen muuttujien luokat

Koulutustaso	Tahdonvoima	Ahkeruus
1.Ei koske minua	1.Täysin väärin	1.Täysin samaa mieltä
2.En osaa sanoa	2. Pääosin väärin	2.Hyväksyn
3.Vähemmän kuin peruskoulu	3. Osittain oikein, osittain väärin	3.Osittain samaa mieltä
4.Osittain suoritettu peruskoulutaso	4.Pääosin oikein	4.Osittain erimieltä
5.Peruskoulu	5.Täysin oikein	5.Erimieltä
6.Osia lukiosta		6.Vahvasti erimieltä
7.Lukio/ylioppilastutkinto		
8.Osittain suoritettu toisen asteen ammatillinen tutkinto		
9.Toisenasteen ammatillinen tutkinto		
10.Joitakin ammattikorkeakouluopintoja		
11.Ammattikorkeakoulututkinto		
12.Joitakin yliopisto-opintoja		
13.Kandidaattitaso yliopistossa		
14.Ylempi ammattikorkeakoulututkinto tai maisteriopinnot yliopistossa		
15.Suoritettu yliopistollinen ammattitutkinto		
16.Suoritettu tohtorin tutkinto		
17.Joku muu tutkinto?		

Taulukko 1: Luokiteltujen muuttujien alkuperäiset luokat: koulutusaste, tahdonvoima ja ahkeruus

Muuttujien käsittely

Koulutustasomuuttujista käsitellään puuttuvina havaintoina ne vastaukset, jotka ovat 1."ei koske minua", 2."en osaa sanoa" ja 17."muu tutkinto", koska vastaukset 1. ja 2. voidaan tulkita eri tavalla jälkeläisen korkeimmasta koulutustasotoiveesta että vanhempien korkeimmasta koulutustasosta. Vastaa- jien muista tutkinnoista on hankala päätellä, minkä tasoisesta tutkinnosta on kyse.

Ikämuuttajassa on tarpeen poistaa 99-vuotiaat, koska ne ovat mahdollisesti vastaamatta jättäneiden yksilöiden vastauksia tai virheellisiä vastauksia, sillä ne näkyvät oudokkeina 18-24-vuotiaiden joukossa. Aineiston selostuksessa ei ole mainintaa näiden poikkeusten alkuperästä.

Muuttujien uudelleenluokittelu

Tässä tutkimuksessa tutkimuskysymyksissä 1a. ja 1b. äidin ja isän korkein koulutustaso on luokiteltu viiteen luokkaan. Jälkeläisen korkeimmasta

koulutustasotoiveesta ne havainnot, joissa toivottiin toisen asteen ammatillista tutkintoa, lukiota, toisen asteen osittaisia opintoja, peruskoulua ja sitä alempia, on yhdistetty samaan luokkaan alemman korkeakoulutuksen kanssa. Yhdistäminen tehdään, koska havaintoja alemmissa koulutustasoissa kuin alempi korkeakoulututkinto on hyvin pieni määrä [5]. Tällä vähennetään analyysissä syntyvää epäluotettavuutta. Voidaan myös pohtia, miksi korkeakouluopiskelija toivoisi korkeakoulua alemman tutkinnon olevan korkein koulutustasonsa. Tätä muutosta ei kuitenkaan tarvitse tehdä vanhempien osalta.

Vanhempien koulutustaso -muuttujan luokittelu viiteen luokkaan

1. Peruskoulu tai alempi
 - Vähemmän kuin peruskoulu
 - Osittain suoritettu peruskoulutaso
 - Peruskoulu
2. Toisen asteen ammatillinen tutkinto tai lukio
 - Osia lukiosta
 - Lukio/ylioppilastutkinto
 - Osittain suoritettu toisen asteen ammatillinen tutkinto
 - Toisen asteen ammatillinen tutkinto
3. Alempi korkeakoulututkinto
 - Joitakin ammattikorkeakouluopintoja
 - Ammattikorkeakoulututkinto
 - Joitakin yliopisto-opintoja
 - Kandidaattitaso yliopistossa
4. Ylempi korkeakoulututkinto
 - Ylempi ammattikorkeakoulututkinto tai maisteriopinnot yliopistossa
 - Suoritettu yliopistollinen ammattitutkinto
5. Suoritettu tohtorin tutkinto

Jälkeläisen koulutustasotoive on jaettu samoihin luokkiin kuin edellä, mutta luokat 1, 2 ja 3 ovat yhdessä luokassa, jonka nimi on *1-3. Alempi korkeakoulututkinto tai alempi*.

Tutkimuskysymyksessä 2 muutetaan koulutustasomuuttuja dikotomiseksi eli kaksiluokkaiseksi. Luokkaan 0 tulee alemmat tutkinnot ja luokkaan 1 ylemmät tutkinnot seuraavan listan mukaisesti.

Koulutustasomuuttujan luokittelu kahteen luokkaan

- 0. Alempi tutkinto
 - Vähemmän kuin peruskoulu
 - Osittain suoritettu peruskoulutaso
 - Peruskoulu
 - Osia lukiosta
 - Lukio/ylioppilastutkinto
 - Osittain suoritettu toisen asteen ammatillinen tutkinto
 - Toisen asteen ammatillinen tutkinto
 - Joitakin ammattikorkeakouluopintoja
 - Ammattikorkeakoulututkinto
 - Joitakin yliopisto opintoja
 - Kandidaattitaso yliopistossa
- 1. Ylempi tutkinto
 - Ylempi ammattikorkeakoulututkinto tai maisteriopinnot yliopistossa
 - Suoritettu yliopistollinen ammattitutkinto
 - Suoritettu tohtorin tutkinto

Kato

Alkuperäisistä muuttujista koulutustasomuuttujissa ja iässä ei ollut lainkaan vastauskatoa. Ikämuuttujassa 99-vuotiaita on 32 havaintoyksikköä. Tahdonvoimamuuttujassa on 18 yksikön kato ja ahkeruusmuuttujassa 15 yksilön kato. Alkuperäisissä muuttujissa kato on 1.3% ja uudelleenluokittelun jälkeen kato on tasan 5%.

1.3 Tilastollisen analyysin menetelmät

Tässä osiossa käsiteltyyn teoriaan on käytetty luentomonisteita Uskottavuus- ja Bayespäätely ja Johdatus monimuuttujamenetelmiin -kurseilta, kirjoja [6] ja [5] sekä nettisivua [7].

1.3.1 Ristiintaulukointi ja χ^2 -testaus

Ristiintaulukointia käytetään luokiteltujen muuttujien jakaumien ja selittäjien erojen tarkasteluun. Tässä tutkimuksessa ristiintaulukoinnissa rivimuuttujana toimii selittävä ja sarakemuuttujana selitettävä. Taulukon soluun ij asetetaan havaittu frekvenssi O_{ij} niistä havainnoista, joilla on selittäjän luokan i ja selitettävän luokan j ominaisuudet.

Ristiintaulukoinnin merkitsevyydestä testauksena toimii χ^2 -testi, joka testaa riippumattomuutta muuttujien välillä. Jokaiselle havaitulle solufrekvenssille lasketaan *odotettu solufrekvenssi* E_{ij} kaavalla 1. Odotettu frekvenssi kuvaa frekvenssiä, joka tulisi soluun ij , jos muuttujien välillä ei olisi riippuvuutta.

$$E_{ij} = \frac{m_i \cdot n_j}{N}, \quad (1)$$

E_{ij} = odotettu frekvenssi solussa ij , m_i = rivin reunafrekvenssi, n_j = sarakkeen reunafrekvenssi ja N = kaikkien havaintojen lukumäärä.

Odotetut frekvenssit E_{ij} asetetaan vastaavien havaittujen frekvenssien O_{ij} kanssa samaan soluun ij taulukon 2 mukaisesti. Näin voidaan vertailla jokaisen solun havaitun ja odotetun frekvenssin eroa.

Jotta selittäjän luokkien välisiä eroja voitaisiin tarkastella paremmin, tarkastellaan ristiintaulukon suhteellisia jakaumia eli prosenttijakaumia. Prosenttiosuus p_{ij} lasketaan solufrekvensseistä O_{ij} , $i = 1, \dots, k$ suhteessa rivin i reunajakaumaan kaavan 2 tavalla. Tällöin rivin i prosenttiosuuksien yhteenlaskettu prosenttimäärä on 100%.

$$p_{ij} = \frac{O_{ij}}{m_i}, \quad (2)$$

jossa O_{ij} = havaittu frekvenssi solussa ij , m_i = rivin i reunajakauma.

	sarakeluokka1	..	sarakeluokka k	Yhteensä
riviluokka 1	O_{11}, E_{11}	..	O_{1k}, E_{1k}	m_1
riviluokka 2	O_{21}, E_{21}	..	O_{2k}, E_{2k}	m_2
.
.
riviluokka h	O_{h1}, E_{h1}	..	O_{hk}, E_{hk}	m_h
Yhteensä	n_1	..	n_k	N

Taulukko 2: Ristiintaulukointi: h = riviluokkien määrä , k =sarakeluokkien määrä , O_{ij}, E_{ij} = havaitut frekvenssit o ja odotetut frekvenssit e solussa ij , jossa $i = 1, \dots, h$, $j = 1, \dots, k$, n_j = sarakkeen j reunajakauma ja m_i = rivin i reunajakauma, N = kaikkien havaintojen lukumäärä

χ^2 -testi

χ^2 -testissä verrataan odotettuja frekvenssejä havaittuihin frekvensseihin laskemalla testisuureen arvo kaavalla 3. Testisuureen arvo, χ^2 - luku, kuvaa odotettujen ja havaittujen frekvenssien eroa toisistaan.

$$\chi^2 = \sum_{i=1}^h \sum_{j=1}^k \frac{(O_{ij} - E_{ij})^2}{E_{ij}}, \quad (3)$$

jossa O_{ij} = havaittu frekvenssi solussa ij , E_{ij} = odotettu frekvenssi solussa ij

P-arvo ja sen tulkinta

P-arvo haetaan χ^2 -jakaumasta vapausasteella df testisuureen avulla kohdasta, jossa $P(\chi^2 \geq x^2 | H_0 \text{ on tosi})$. Vapausaste saadaan rivi- ja sarakeluokkien määristä kaavalla 4. Saatu p-arvo antaa suuntaa antavan kuvan, onko muuttujilla todellisuudessa riippuvuutta.

Nollahypoteesina H_0 ristiintaulukoinnissa on yleensä selittäjän ja selitetävän välinen riippumattomuus. Vastahypoteesina H_1 toimii selittäjän ja selitetävän välinen riippuvuus. Jos p-arvo on pienempää kuin 0.01, aineistolla on näyttöä nollahypoteesia vastaan. P-arvon ollessa suurempaan kuin 0.1, aineisto on kohtalaisen sopusoinnussa nollahypoteesin kanssa. Toisalta, jos p-arvo jää 0.01 ja 0.1 väliin, aineistolla on heikkoa näyttöä nollahypoteesia vastaan.

$$df = (k - 1) \cdot (h - 1) \quad (4)$$

, jossa h = riviluokkien lukumäärä ja k = sarakeluokkien lukumäärä.

1.3.2 Logistinen regressio

Logistisen regression pääideana on kuvata, kuinka todennäköisesti havaintoyksikkö kuuluu vasteen Y luokkaan k selittäjien X_i avulla, jossa $i = 1, \dots, r$. Tässä tutkimuksessa vasteen luokkia on kaksi, eli vaste on dikotominen. Vasteen Y vaihtoehtoisina arvoina toimii 0 ja 1 tavalla 5.

$$Y = \begin{cases} 0, & \text{kun ei a} \\ 1, & \text{kun a.} \end{cases} \quad (5)$$

Logistisen regression pohjana toimii π , joka kuvaa todennäköisyyttä, että vaste Y on luokassa 1. Sen vastakkaisena arvona toimii todennäköisyys, että vaste on luokassa 0. Nämä merkitään kaavojen 6 ja 7 tavalla.

$$P(Y = 1) = \pi \quad (6)$$

$$P(Y = 0) = 1 - \pi \quad (7)$$

Logistisessa regressiossa käytetään mallia, joka on kaavan 8 muotoinen. Todennäköisyys π saa arvoja väliltä $[0, 1]$. Jokaisella selittäjällä X_i on oma kerroin β_i . Kerroin kuvaa, kuinka hyvin vaste $Y = 1$ kohdataan selittäjällä X_i . Poikkeuksena on kerroin β_0 , joka ei ole minkään selittäjän kertoimena, vaan toimii skaalaparametrina, eli vakiona. Jokaiselle selittäjälle X_i estimoidaan *suurimman uskottavuuden menetelmän* avulla kerroin $\hat{\beta}_i$ niin, että ne tuottavat mahdollisimman suurella uskottavuudella havaitun aineiston.

$$\pi = P(Y = 1 | X_1, \dots, X_r) = \frac{e^{\beta_0 + \beta_1 X_1 + \dots + \beta_r X_r}}{1 + e^{\beta_0 + \beta_1 X_1 + \dots + \beta_r X_r}} \quad (8)$$

,jossa $\beta_i =$ selittäjän X_i kerroin, $i = 1, \dots, r$ ja $\beta_0 =$ vakio.

Kun kertoimen β_i arvo on positiivinen, selittäjällä X_i on kasvava vaikutus vasteen kohtaamiselle. Vastaavasti negatiivinen kerroin kertoo vähenevästä vaikutuksesta vasteen kohtaamiselle. Kertoimen suuruus viittaa myös, kuinka paljon selittäjällä on vaikutusta.

Kertoimien testaus

Logistisen regressiossa käytetään Z -testisuuretta testaamaan nollahypoteesia $H_0: \beta_i = \beta_i^0$ kaavan 9 mukaisesti. Z -suure noudattaa likimääräisesti normaalijakaumaa $N(0, 1)$.

$$Z_{hav} = \frac{\hat{\beta}_i - \beta_i^0}{SE(\hat{\beta}_i)} \sim N(0, 1), \text{ kun } H_0 \text{ on tosi} \quad (9)$$

,jossa $\hat{\beta}_i$ = estimoitu kerroin β_i , β_i^0 =kertoimen β_i nollahypoteesi, $SE(\hat{\beta}_i)$ = kertoimen $\hat{\beta}_i$ keskivirhe.

Kertoimen $\hat{\beta}_i$ keskivirhe SE saadaan suurimman uskottavuuden menetelmän avulla informaatiomatriisin käänteismatriisin diagonaalialkioista. Z -suuretta käytetään hakemaan p-arvo normaalijakauman tiheysfunktioista paikasta $P(|Z| \geq |z_{hav}|)$. P-arvolla voidaan arvioida nollahypoteesin sopusointua aineiston kanssa. P-arvo tulkitaan samalla tapaa kuin edellisen luvun osassa *p-arvo ja sen tulkinta*.

P-arvon lisäksi kertoimien todellisuutta tutkitaan kertoimien luottamusvälien avulla. Luottamusväli antaa kuvan, missä todellinen β_i -kerroin voisi sijaita saatujen havaintojen avulla. Kertoimen β_i $100(1 - \gamma)\%$ luottamusvälin laskemiseen käytetään kaavaa 10. Tässä tutkimuksessa $\gamma = 0.05$, eli käytetään 95 % luottamusväliä.

$$\hat{\beta}_i \pm z_{1-\gamma/2} \cdot SE(\hat{\beta}_i) \quad (10)$$

, jossa $z_{1-\gamma/2}$ = normaalijakauman $1 - \gamma/2$ -fraktiili, $SE(\hat{\beta}_i)$ = estimoidun kertoimen $\hat{\beta}_i$ keskivirhe.

Tuloksien laskemiseen on käytetty R-kielen valmiiksi rakennettuja funktiota.

2 Tulokset

Tässä osiossa tarkastellaan saatuja tuloksia. Tuloksissa ei mennä tarkemmin suoritettuihin laskuihin, joten laskennat ovat tehty R-ohjelmiston versiolla 3.6.3. Tuloksien laskemiseen on käytetty `epi`-paketin funktioita, kuten `stat.table` taulukoiden tekemiseen ja `ci.lin` logistisen regression tulosten laskemiseen, sekä `r`-kielen sisäisiä funktioita, kuten `dchisq` χ^2 -luvun hakemiseen ja `glm` logit-mallin tekemiseen. Muuttujien käsittelyä varten käytettiin `naniar`-paketin `place_with_na` -funktioita.

Ensin tarkastellaan tutkimuskysymyksiin 1a ja 1b liittyviä prosentti- ja frekvenssitaulukkoja 4, 5, 6 ja 7. Taulukoissa 5 ja 7 on ilmoitettu havaitut ja odotetut frekvenssit. Prosenttijakaumat on erotettu erillisiin taulukoihin 4 ja 6. Kumpaankin alakysymykseen on laskettu χ^2 -luku ja sitä vastaava *p*-arvo. Tutkimuskysymyksen 2 osalta tarkastellaan taulukkoa 8, johon on ilmoitettu estimoidut beta-kertoimet, kertoimien keskivirheet, *Z*-luvut, *p*-arvot sekä 95% luottamusvälit.

2.1 Vanhempien koulutustason yhteys jälkeläisen koulutustasotoiveeseen

Tässä luvussa käsitellään ensimmäistä tutkimuskysymystä, joka on jaettu kahteen osaan: isän koulutustason yhteyteen ja äidin koulutustason yhteyteen. Tarkastellaan kuitenkin ensimmäisenä, miten isän ja äidin koulutustasot ovat jakautuneet korkeimman koulutustason mukaan ja miten jälkeläisen toive on jakautunut korkeimman koulutustasotoiveen mukaan taulukossa 3 vanhempien luokittelun mukaisesti.

Taulukosta huomataan, että suurin osa (65%) jälkeläisistä toivoo ylempää korkeakoulututkintoa, mutta suurin osa kummastakin vanhemmasta on käynyt toisen asteen koulutuksen. Hyvin pieni osa jälkeläisistä on toivonut toisen asteen koulutusta. Vanhempien koulutustasot ovat painottuneet riviluokkiin 2, 3 ja 4. Huomataan myös, että äidit ovat hieman isää korkeakoulutetumpia.

2.1.1 Isän koulutustaso

Taulukosta 4 havaitaan eroja jälkeläisten jakautumisesta sen mukaan, mikä on isän korkein koulutustaso. Ylempi korkeakoulututkinto on kaikista toivottuin jälkeläisten keskuudessa riippumatta isän koulutustasosta. Lisäksi kaikista eniten (70.8%) jälkeläinen on toivonut ylempää korkeakoulututkintoa, jos isän korkein koulutustaso on myös ylempi korkeakoulututkinto. Huomataan, että jakautumiset ovat hyvin samanlaiset, jos isä on käynyt peruskoulun

Taulukko 3: Jälkeläisen korkeimman koulutustasotoiveen, isän ja äidin korkeimman koulutustason jakautuminen(%)

Koulutustaso	Jälkeläinen	isä	äiti
1.Peruskoulu tai alempi	0.0	13.9	4.5
2.Toisen asteen koulutus	0.6	43.3	43.5
3.Alempi korkeakoulututkinto	27.0	25.4	28.4
4.Ylempi korkeakoulututkinto	64.6	15.0	22.0
5.Suoritettu tohtorin tutkinto	7.9	2.3	1.6
Yhteensä %	100	100	100
n	710	646	669

tai toisen asteen koulutuksen. Havaitaan myös opiskelijoiden toiveiden siirtymistä alemmista korkeakoulututkinnoista ylempiin tutkintoihin, mitä korkeampi isän koulutustaso on. Tohtorin tutkintoa toivotaan selvästi enemmän, mitä korkeakoulutetumpi isä on. Kaikista eniten alempaa korkeakouluastetta tai sitä alempaa koulutustasoa toivovat ne, joiden isä on käynyt peruskoulun tai alemman koulutuksen.

Taulukko 4: Suhteelliset jakaumat(%): Isän korkeimman koulutustason ja jälkeläisen korkeimman koulutustasotoiveen ristiintaulukointi

Koulutustaso	Jälkeläinen			
	Isä	1-3.	4.	5. yht.
1.Peruskoulu tai alempi		31.8	62.5	5.7 100
2.Toisen asteen koulutus		30.5	62.2	7.3 100
3.Alempi korkeakoulututkinto		23.8	67.5	8.8 100
4.Ylempi korkeakoulututkinto		18.8	70.8	10.4 100
5.Suoritettu tohtorin tutkinto		26.7	53.3	20.0 100
Yhteensä %		27.5	64.6	7.9 100
n		172	410	52 634

Taulukosta 5 huomataan, että havaittujen ja odotettujen frekvenssien välillä ei ole suuria eroja. Suurin ero on vain noin kymmenen yksikön kokoinen. Tämä löytyy jälkeläisen toiveen ollessa alempi korkeakoulututkinto tai sitä alempi ja isän koulutustason ollessa toiseen asteen koulutus. Koska hyvin

pieniä odotettuja frekvenssejä ei ole paljon, voidaan todeta, että tuloksiin ei tule liiallista epäluotettavuutta.

Taulukosta 5 löytyvien odotettujen ja havaittujen frekvenssien avulla on laskettu testisuureen arvo χ_{IJ}^2 kaavalla 3. P-arvo on haettu χ^2 -tiheysfunktioista vapausasteella $(5 - 1) \cdot (3 - 1)$ eli 8. Nollahypoteesina H_0 on isän korkeimman koulutustason ja jälkeläisen korkeimman koulutustasotoiveen välinen riippumattomuus, eli isän koulutustaso ei ole yhteydessä jälkeläisen toiveeseen.

$$\chi_{IJ}^2 = 10.59212 \quad (11)$$

$$p - arvo_{IJ} = 0.06203375 \approx 0.06 \quad (12)$$

P-arvo sijoittuu 0.01 ja 0.1 väliin. Tämä viittaa siihen, että havainnoilla on heikkoa näyttöä nollahypoteesia vastaan, eli isän koulutustasolla olisi mahdollisesti yhteys jälkeläisen korkeimpaan koulutustasotoiveeseen. Näyttö on kuitenkin heikkoa, joten todellisuudessa isän korkeimmalla koulutustasolla ei välttämättä ole vaikutusta jälkeläisen korkeimpaan koulutustasotoiveeseen.

Taulukko 5: Havaitut ja odotetut frekvenssit: Isän korkeimman koulutustason ja jälkeläisen korkeimman koulutustasotoiveen ristiintaulukointi

Isä	Jälkeläinen			
	1-3.	4.	5.	Yht.
1.Peruskoulu tai alempi	28	55	5	88
	23.9	56.9	7.2	
2.Toisen asteen koulutus	84	171	20	275
	74.6	177.8	22.6	
3.Alempi korkeakoulututkinto	38	108	14	160
	43.4	103.5	13.1	
4.Ylempi korkeakoulututkinto	18	68	10	96
	26.0	62.1	7.9	
5.Suoritettu tohtorin tutkinto	4	8	3	15
	4.1	9.7	1.2	
Yhteensä	168	410	52	634

2.1.2 Äidin koulutustaso

Taulukosta 6 havaitaan äidin ja jälkeläisen välillä samanlaisia eroja kuin isän ja jälkeläisen välillä. Ylempi korkeakoulututkinto on kaikista yleisin toive jälkeläisille riippumatta äidin koulutustasosta. Havaitaan samanlainen ilmiö kuin isän koulutustason vaihtelussa: mitä enemmän koulutettu äiti on, sitä enemmän jälkeläinen toivoo ylempää korkeakoulututkintoa tai tohtorin tutkintoa. Vastaavasti, mitä vähemmän äiti on kouluttautunut, sitä enemmän jälkeläinen toivoo alempaa korkeakoulututkintoa. Suurin suhteellinen osuus löytyy myös jälkeläisen toivoessa ylempää korkeakoulututkintoa ja äidin ollessa ylempään korkeakoulututkintoon käynyt.

Taulukko 6: Suhteelliset jakaumat(%): Äidin korkein koulutustaso ja jälkeläisen korkein koulutustasotoive

Koulutustaso	Jälkeläinen			
	1-3.	4.	5.	yht
Äiti				
1.Peruskoulu tai alempi	33.3	60.0	6.7	100
2.Toisen asteen koulutus	34.5	59.2	6.3	100
3.Alempi korkeakoulututkinto	22.0	68.3	9.7	100
4.Ylempi korkeakoulututkinto	17.8	73.3	8.9	100
5.Suoritettu tohtorin tutkinto	18.2	63.6	18.2	100
Yhteensä %	27.5	64.6	7.9	100
n	174	427	53	657

Taulukosta 7 huomataan, että odotetut ja havaitut frekvenssit poikkeavat toisistaan vähän enemmän verrattuna isän koulutustason tapauksessa. Suurin poikkeama löytyy äidin toisen asteen koulutuksen ja jälkeläisen alemman korkeakoulututkintotoiveen solusta, jossa ero on noin 21 yksikköä. Kolmessa muussa solussa on hieman yli kymmenen yksikön poikkeama ja muissa soluissa alle kymmenen yksikköä. Odotettu frekvenssi E_{55} on hieman alle 1. Näin pieni odotettu frekvenssi voi aiheuttaa pientä epäluotettavuutta analyysivaiheessa.

Taulukosta 7 lasketaan testisuure χ^2_{AJ} kaavalla 3. Vapausasteena toimii myös $(5 - 1) \cdot (3 - 1)$ eli 8. Nollahypoteesina H_0 on äidin korkeimman koulutustason ja jälkeläisen korkeimman koulutustasotoiveen välinen riippumat-

Taulukko 7: Havaitut ja odotetut frekvenssit: Äidin korkein koulutustaso ja jälkeläisen korkein koulutustasotoive

Koulutustaso	Jälkeläinen			
	1-3.	4.	5.	Yht.
Äiti				
1.Peruskoulu tai alempi	10 8.1	18 19.5	2 2.4	30
2.Toisen asteen koulutus	98 76.5	168 184.6	18 22.9	284
3.Alempi korkeakoulututkinto	41 50.1	127 120.9	18 15.0	186
4.Ylempi korkeakoulututkinto/	26 39.3	107 94.9	13 11.8	146
5.Suoritettu tohtorin tutkinto	2 3.0	7 7.1	2 0.9	11
Yhteensä	174	427	53	657

tomuus, eli äidin koulutustasolla ei ole yhteyttä jälkeläisen toiveeseen.

$$\chi_{AJ}^2 = 19.68675 \quad (13)$$

$$p - arvo_{AJ} = 0.004220142 \approx 0.004 \quad (14)$$

P-arvo on pienempi kuin 0.01. Havainnoilla on siis näyttöä nollahypoteesia vastaan. Äidin korkeimmalla koulutustasolla voi siis olla yhteys jälkeläisen koulutustasotoiveeseen.

2.2 Tahdonvoiman, iän ja ahkeruuden yhteydet opiskelijan koulutustasotoiveeseen

Tässä osiossa tarkastellaan toista tutkimuskysymystä. Vasteena toimii opiskelijan korkein koulutustasotoive, joka on jaettu *muuttujien uudelleen luokittelu* kappaleen mukaisesti kahteen luokkaan, kuten kaavassa 15.

$$Y = \begin{cases} 0, & \text{kun opiskelija toivoo alempaa koulutustasoa} \\ 1, & \text{kun opiskelija toivoo ylempää koulutustasoa.} \end{cases} \quad (15)$$

Taulukossa 8 on kirjattu kaikki saadut tulokset. Z -suureen testaukseen käytetty nollahypoteesi on $H_0 : \beta_i = 0$ kullakin $i = 0, 1, 2, 3$. Toisin sanoen, kyseisellä selittäjällä X_i ei olisi mitään vaikutusta vasteeseen Y .

i.Muuttuja	$\hat{\beta}_i$	SE	z	p	2.5%	97.5%
0.Vakio	0.7418	1.3986	0.5304	0.5959	-1.9942	3.4961
1.Tahto	0.2823	0.0961	2.9377	0.0033	0.0940	0.4714
2.Ahkeruus	-0.0887	0.0749	-1.1855	0.2358	-0.2355	0.0584
3.Ikä	-0.0238	0.0600	-0.3959	0.6922	-0.1419	0.0937

Taulukko 8: Logit-mallin su-estimoidut β -kertoimet $\hat{\beta}_i$, keskivirhe, Z -luku, p-arvo, 95 % luottamusvälin ala- ja yläraja β_i :lle

Tahdonvoiman estimoitu beta-kerroin $\hat{\beta}_1$ on positiivinen, eli mitä enemmän opiskelijalla on mielestään tahdonvoimaa, sitä todennäköisimmin opiskelija toivoo ylempää koulutustasoa. Tahdonvoima-kertoimen luottamusvälit näyttävät ilmaisevan samaa: 0 ei kuulu luottamusvälien sisälle ja väli on kokonaan positiivisen puolella. On kuitenkin huomioitava, että on myös pieni mahdollisuus, että todellinen beta-kerroin sijaitsee luottamusvälin ulkopuolella. Kun tarkastellaan beta-kertoimen $\hat{\beta}_1$ p-arvoa, voidaan todeta sen olevan hyvin pieni. Aineistolla on siis näyttöä nollahypoteesia vastaan, kun $i = 1$. Voidaan siis todeta, että tahdonvoimalla olisi hyvin mahdollisesti yhteys opiskelijan toiveeseen.

Ahkeruuden ja iän estimoidut beta-kertoimet, $\hat{\beta}_2$ ja $\hat{\beta}_3$, ovat negatiivisia. Tämä voi tarkoittaa sitä, että mitä enemmän opiskelija on eri mieltä siitä, että menestys saavutetaan työnteolla ja onnella ei ole merkitystä ja mitä vanhempi opiskelija on, sitä todennäköisemmin opiskelija ei toivo ylempää koulutustasoa. Toisin sanoen, mitä enemmän opiskelija on samaa mieltä ahkeruusmuuttujan kanssa ja mitä nuorempi hän on, sitä enemmän hän toivoo

ylempää koulutustasoa. Näillä muuttujilla ei välttämättä ole kuitenkaan mitään vaikutusta opiskelijan korkeimpaan koulutustasotoiveeseen, koska kummankin luottamusvälit menevät negatiiviselta positiiviselle, jolloin väliin jää myös 0. On siis mahdollista, että todellisuudessa menestyksellä ja iällä olisi positiivinen yhteys opiskelijan toiveeseen tai että niillä ei ole ollenkaan yhteyttä koulutustasotoiveeseen. Tämän voi myös huomata beta-kertoimille saaduista p-arvoista, jotka ovat suurempia kuin 0.1. Tämä tarkoittaisi aineiston olevan kohtuullisesti sopusoinnussa nollahypoteesin kanssa, kun $i = 2, 3$.

Huomataan myös, että kaikki kertoimet ovat $[-1, 1]$ välissä. Tahdonvoimalla on hieman voimakkaampi yhteys vasteeseen kuin ahkeruus- ja ikämuuttujalla, sillä sen kerroin on itseisarvollisesti suurin näistä kolmesta. Koska vakio-arvo on positiivinen, vasteen kohtaamiselle on valmiiksi hieman suurempi todennäköisyys, olkoon yksilön ominaisuudet muissa muuttujissa mitä vain. On kuitenkin huomioitava, että vakion luottamusvälit menevät myös negatiiviselta positiiviselle, joten todellisuus voi olla mitä vain. Vakion p-arvo myös viittaa siihen, että painokerrointa ei välttämättä olisi lainkaan. Kaiken kaikkiaan muuttujien luottamusvälin pituudet ja keskivirheet ovat pienet. Tämä voi johtua siitä, että havaintojen määrä on jo suhteellisen iso.

3 Pohdinta

Vanhempien koulutustason yhteydestä jälkeläisen koulutustason valitsemiseen on havaittu tietokeskuksen tekemissä tutkimuksissa samanlaisia havaintoja. Vuonna 2009 tehdyssä tutkimuksessa on todettu korkeakoulutuksen pe-riytyvän selvästi vanhemmalta jälkeläiselle sekä äidin korkealla koulutuksella olevan voimakkaampi vaikutus jälkeläisen korkeakoulututkinnon hankintaan kuin isän koulutuksella [8]. Tässä tutkimuksessa havaittiinkin hieman vahvempaa näyttöä äidin koulutustason yhteydestä jälkeläisen koulutustasotoiveeseen kuin isän koulutustason yhteyden kohdalla. Tämä voi myös selittyä sillä, että tässä aineistossa äidit olivat hieman enemmän korkeakoulutettumpia kuin isät.

On kuitenkin otettava huomioon, että tässä tutkimuksessa tarkastellaan pelkästään niitä jälkeläisiä, jotka ovat korkeakouluopiskelijoita vuonna 2015. Tarkastelun ulkopuolelle jää he, jotka eivät ole korkeakouluopiskelijoita kyseisenä vuonna. Lisäksi tarkastelemme vain jälkeläisen korkeinta koulutustasotoivetta. Tämä ei siis välttämättä tarkoita, että opiskelija saavuttaisi korkeimman koulutustasotoiveensa, eikä vanhemman koulutustaso siis välttämättä periydy jälkeläiselle.

Toisen tutkimuskysymyksen kohdalla todettiin, että opiskelijan korkeimpaan koulutustasotoiveeseen selvästi isoin merkitys valituista muuttujista on tahdonvoiman määrällä. European sociological review:n julkaisemassa tutkimuksessa onkin havaittu, että jälkeläisen saaduilla kouluarvosanoilla sekä vanhempien koulutustasolla on vaikutusta, mihin jälkeläinen päätyy opiskelemaan [2]. Tahdonvoimalla onkin motivaation rinnalla vaikuttava tekijä oppimiseen ja tavoitteiden saavuttamiseen, sillä ne vaativat sitoutumista. Sillä on siis vaikutusta opintojen suorittamiseen ja hyvien arvosanojen saamiseen, mikä selittäisi saatuja tuloksia [3].

Toisaaltaan ikä- ja ahkeruusmuuttujien vähäistä yhteyttä voidaan selittää aineistoon liittyvillä seikoilla. Ikämuuttujan kohdalla suppea ikäväli voi olla syynä havaittuihin tuloksiin. Suurin osa aineiston vastaajista on 20-24-vuotiaita. Tällä viiden ikävuoden välillä on hankala arvioida iän yhteyttä, koska mielipiteet eivät välttämättä vaihtele vuoden erolla suuresti. Lisäksi pienen välin avulla on hankalaa arvioida suurempaa kokonaisuutta.

Ahkeruusmuuttujan vastaukset ovat voineet vaihdella riippuen, miten menestys-sana on tulkittu. Joillekin menestys voi tarkoittaa rahallista menestystä, mutta taas toiselle mainetta. Ahkeruusmuuttuja on enemmänkin mielipidekysymys, joten vastaaja ei välttämättä noudata tätä. Lisäksi voidaan pohtia, kuvaako muuttujan kysymys ahkeruutta.

On myös tarpeellista tarkastella harhan määrää aineistossa. Harhaa on voinut syntyä inhimillisistä tekijöistä. Vastajat ovat arvioineet omia ominai-

suuksiaan, ilmoittaneet vanhempien koulutustason oman tiedon perusteella, sekä tulkinneet sanoja omalla tavallaan. Systemaattisia virheitä, kuten vanhingossa väärän vastauksen painaminen, on voinut myös tapahtua. Harhan määrästä ei kuitenkaan voida olla varmoja.

Tämän tutkimuksen pohjalta voidaan päätellä, että vanhempien koulutustaustalla on havaittava yhteys jälkeläisen koulutustasotoiveeseen. Samoin on havaittavissa yhteyttä tahdonvoiman ja koulutustasotoiveen välillä. Iän ja ahkeruuden yhteyttä koulutustasotoiveeseen ei voida päätellä tämän tutkimuksen pohjalta. Näiden muuttujien osalta vaadittaisiinkin lisätutkimusta selkeämpien vastauksien saamiseksi. Saadut tulokset ovat kuitenkin sopuoinnussa muiden tutkimusten kanssa.

Viitteet

- [1] Miina Keski-Petäjä and Mika Witting. Vanhempien koulutus vaikuttaa lasten valintoihin, 2016. Viitattu: 9.3.2021, url: [//www.tilastokeskus.fi/tietotrendit/artikkelit/2016/vanhempien-koulutus-vaikuttaa-lasten-valintoihin](http://www.tilastokeskus.fi/tietotrendit/artikkelit/2016/vanhempien-koulutus-vaikuttaa-lasten-valintoihin).
- [2] L.Heis-kala, J.Erola, and E.Kilpi-Jakonen. Compensatory and multiplicative advantages: Social origin, school performance, and stratified higher education enrolment in finland. *European Sociological Review*, (jcaa046), 2020. DOI: doi.org/10.1093/esr/jcaa046.
- [3] Sanna Sutela. Mistä löytää motivaatio, jos opiskelu ei voisi vähempää kiinnostaa?, 2014. Viitattu: 2.7.2021, url: <https://yle.fi/aihe/artikkeli/2014/10/06/mista-loytaa-motivaatio-jos-opiskelu-ei-voisi-vahempaa-kiinnostaa>.
- [4] H.Helve and A.Svynarenko. *Suomalaisten korkeakouluopiskelijoiden identiteettihorisontit 2015*. 2015.
- [5] D.Yates. *The Practice of Statistics (1st Ed.)*. W.H. Freeman, New York (N.Y.), 1999.
- [6] G.James, D.Witten, T.Hastie, and R.Tibrishari:. *An introduction to statistical learning*. Springer, 2013.
- [7] M.Mattila, E.Paaso, M.Alastalo S.Borg, N.Ellonen, J.Sivonen, H.Keckman-Koivuniemi, S.Antikainen, T.Pasanen, and T.J. Alaterä. Ristiintaulukointi, 2004. Viitattu: 2.7.2021, url: <https://www.fsd.tuni.fi/menetelmaopetus/ristiintaulukointi/ristiintaulukointi.html>.
- [8] Pekka Myrskylä. Koulutus periytyy edelleen. *Hyvinvointikatsaus 1/2009*, 1, 2009. Viitattu: 9.3.2021, url:https://www.stat.fi/artikkelit/2009/art2009-03-16_02.html?s=0.