

Larissa Pankova

# FAKTORIANALYYSIN SOVELTAMINEN URASEURANTAKYSELYAINEISTOON

Informaatioteknologian ja viestinnän tiedekunta  
Kandidaattitutkielma  
Elokuu 2025

# Tiivistelmä

Larissa Pankova: Faktorianalyysin soveltaminen uraseurantakyselyaineistoon

Kandidaattitutkielma

Tampereen yliopisto

Tilastollinen data-analyysi

Elokuu 2025

---

Tässä tutkielmassa tarkastellaan Tampereen yliopiston vuoden 2014 uraseurantakyselyaineistoa eksploratiivisen faktorianalyysin avulla. Saatua faktorimallia hyödynnetään varianssianalyyseissä (ANOVA ja MANOVA) koulutusalojen välisten erojen tarkasteluun.

Uraseurantakyselyn tavoitteena on kartoittaa valmistuneiden opiskelijoiden tyytyväisyyttä tutkintoonsa, tähänastiseen työuraansa sekä työelämän laatuun suhteessa saatua yliopistokoulutukseen. Lisäksi kyselyssä selvitetään, mitkä tiedot ja taidot ovat työelämässä tärkeitä ja kuinka hyvin yliopisto-opinnot ovat näitä kehittäneet. Aineisto koostuu 508 vastaajasta yhdeksältä eri koulutusosalalta.

Faktorianalyysi on tilastollinen menetelmä, jolla pyritään selvittämään liittyvät-  
kö muuttujat piileviin taustatekijöihin eli faktoreihin. Menetelmän avulla voidaan yksinkertaistaa laajoja aineistoja ryhmittelemällä muuttujia niiden välisten yhteisten piirteiden perusteella.

Analyysi aloitettiin aineiston muuttujien ja faktorianalyysin teorian tarkastelulla sekä aineiston soveltuvuuden arvioinnilla. Lopulliseen malliin valittiin viisi faktoria, jotka selittivät 40 prosenttia aineiston kokonaisvaihtelusta. Jatkoanalyysissä ANOVA ja MANOVA osoittivat merkittäviä eroja kolmella faktorilla sekä yleisesti koulutusalojen välillä.

Avainsanat: eksploratiivinen faktorianalyysi, varianssianalyysi, ANOVA, MANOVA

Tämän julkaisun alkuperäisyys on tarkastettu Turnitin OriginalityCheck -ohjelmalla.

# Tekoälyn käyttö opinnäytteessä

Opinnäytteessäni on käytetty tekoälysovelluksia:

Kyllä,

Ei.

## **Tekoälysovellusten nimi ja versio**

ChatGPT, GPT-5 -malli

## **Käyttötarkoitus**

Olen käyttänyt tekoälyä kieliopin tarkistamiseen, lauserakenteiden paranteluun sekä käännöstyökaluna yksittäisten englanninkielisten termien ja ilmausten ymmärtämiseen.

## **Osiot, joissa tekoälyä on käytetty**

Tekoälyä on käytetty tutkielmassa yleisellä tasolla.

Olen tietoinen siitä, että olen täysin vastuussa koko opinnäytteeni sisällöstä, mukaan lukien tekoälyllä tuotetut osat, ja hyväksyn vastuun mahdollisista eettisten ohjeiden rikkomuksista.

# Sisällys

<b>1</b>	<b>Johdanto</b>	<b>1</b>
<b>2</b>	<b>Tutkimusaineisto</b>	<b>2</b>
2.1	Aineiston kuvailu . . . . .	2
2.2	Aineiston muuttujat . . . . .	3
<b>3</b>	<b>Faktorianalyysin teoria ja estimointi</b>	<b>6</b>
3.1	Faktorianalyysin perusmalli . . . . .	6
3.2	Estimointi pääakselimenetelmällä . . . . .	9
3.3	Faktoripisteiden estimointi . . . . .	11
<b>4</b>	<b>Aineiston analysointi</b>	<b>13</b>
4.1	Faktorikelpoisuus ja faktoreiden valinta . . . . .	13
4.2	Faktorimallin rakentaminen . . . . .	15
4.3	Faktoripisteiden tarkastelu ja jatkoanalyysit . . . . .	17
<b>5</b>	<b>Yhteenveto ja pohdinta</b>	<b>25</b>
	<b>Lähteet</b>	<b>26</b>
	<b>Liite A: R-koodit</b>	<b>27</b>
	<b>Liite B: Lataukset</b>	<b>30</b>

# 1 Johdanto

Tässä tutkielmassa sovelletaan eksploraatiivista faktorianalyysiä Tampereen yliopiston uraseurantakyselyaineistoon sekä verrataan opiskelijoiden vastauksia koulutusaloittain. Tutkielman toisessa luvussa esitellään tarkemmin aineiston muuttujat, luvussa kolme perehdytään faktorianalyysin teoriaan. Tutkielman neljännessä luvussa tarkastellaan aineiston faktorikelpoisuutta, rakennetaan faktorimalli, ja sen pohjalta suoritetaan varianssianalyysi sekä muut tarvittavat tilastolliset testit koulutusalojen erojen tutkimiseen.

Varsinaisen faktorianalyysin kehitti vuonna 1904 Charles Spearman. Menetelmää käytetään erityisesti yhteiskuntatieteissä sekä psykologian alalla. Yksi tunnetuimmista faktorianalyysin sovelluksista on Spearmanin työ älykkyyden mittaamisesta [5, s. 1]. Spearman esitti, että älykkyyden taustalla vaikuttaa yksi yleinen tekijä, niin sanottu g-faktori. Tulosta pidetään yhä yleisesti hyväksyttynä. [3, s.650]

Faktorianalyysin tavoitteena on tunnistaa, mitkä aineiston muuttujat ryhmittyvät yhteen niiden välisten korrelaatioiden perusteella, sekä mitkä muuttujat eivät toisaalta korreloi keskenään eli ovat toisistaan riippumattomia. Muuttujat, joilla on keskenään samankaltaista vaihtelua, kuvataan faktoreiden avulla [3, s. 650]. Päämääränä on tiivistää muuttujien määrää käyttäen pienempää määrää faktoreita, jotka ovat niin sanotusti piileviä eli latentteja tekijöitä aineiston muuttujien taustalla [7, s. 25].

Faktorianalyysissä on olemassa kaksi lähestymistapaa: eksploraatiivinen faktorianalyysi (EFA) ja konfirmatorinen faktorianalyysi (CFA). EFA:ta käytetään tilanteissa, joissa ei ole tarkkaa tietoa siitä, kuinka monta faktoria aineistosta tulisi löytyä, vaan tavoitteena on tunnistaa piileviä rakenteita aineistosta. Tämä menetelmä perustuu siihen, että faktorien lukumäärä ja rakenne ovat avoimia kysymyksiä. CFA puolestaan lähtee liikkeelle selkeistä teoreettisista oletuksista. Siinä on jo hypoteesi siitä, kuinka monta faktoria aineistossa on ja mitkä muuttujat liittyvät kuhunkin faktoriin. CFA:ta käytetään usein arvioimaan, toimiiko jo kehitetty mittari samalla tavalla uudessa aineistossa ja vastaako se odotuksia. [2, s. 608]

## 2 Tutkimusaineisto

### 2.1 Aineiston kuvailu

Tutkimusaineistona käytetään Tampereen yliopiston opiskelijoille tarkoitettua uraseurantakyselyä. Uraseurantakysely lähetetään kaikille yliopistosta valmistuneille opiskelijoille tietyn ajan kuluttua valmistumisesta. Uraseurantakyselyillä kerätään tietoa valmistumisen jälkeisen työuran ensimmäisistä vuosista. Kyselyssä kartoitetaan valmistuneiden opiskelijoiden tyytyväisyyttä tutkintoonsa ja työuraansa sekä työelämän laatua suhteessa heidän saamansa yliopistokoulutukseen. Lisäksi kyselyssä selvitetään, mitkä tiedot ja taidot ovat työelämässä tärkeitä ja kuinka hyvin yliopisto-opinnot ovat näitä kehittäneet. [10]

Aineisto jakautuu kuuteen kysymysalueeseen: taustakysymykset, tyytyväisyys tutkintoon, työuran kokonaisuus, työtilanne, työ ja osaaminen sekä työllistymisen laatu. Jokaisessa kysymysalueessa on 1-27 alakysymystä, ja yhteensä kysymyksiä on 77. Tässä tutkielmassa tarkastellaan vain 1-6 Likert-asteikolla vastattuja kysymyksiä, minkä vuoksi analyysissä käytettäviä muuttujia on 63. Asteikossa 1 tarkoittaa "täysin eri mieltä" ja 6 "täysin samaa mieltä". Kyselyyn on vastannut 508 opiskelijaa.

Tässä tutkielmassa kysymykset on ryhmitelty viiteen pääteemaan, jotka on esitetty taulukossa 2.2. Viimeiset kaksi teemaa käsittelevät samoja kysymyksiä: opiskelijat ovat ensin vastanneet, kuinka tärkeitä tiedot ja taidot ovat heidän nykyisessä työssään, ja sen jälkeen, kuinka hyvin yliopisto-opiskelu on kehittänyt näitä taitoja. Teemojen tunnuslukuja tarkastellaan seuraavassa taulukossa 2.1.

**Taulukko 2.1.** Tunnusluvut teemoittain

<i>Asteikko: 1 = täysin eri mieltä, 6 = täysin samaa mieltä</i>				
<b>Teema</b>	<b>Keskiarvo</b>	<b>Mediaani</b>	<b>Keskihajonta</b>	<b>Moodi</b>
Tyytyväisyys tutkintoon	4,15	4	1,38	5
Työtilanne	4,61	5	1,12	5
Työllistymisen laatu	4,62	5	1,26	5
Työelämässä tärkeät tiedot ja taidot	4,28	5	1,66	6
Yliopisto-opiskelun kehittämät valmiudet	3,51	4	1,52	4

## 2.2 Aineiston muuttujat

Tutkielmassa käytetyt kysymykset on esitetty taulukossa 2.2. Analyysin selkeyttämiseksi ne on koodattu muuttujiksi v1-v63.

**Taulukko 2.2.** Tutkimuksessa käytetyt muuttujat

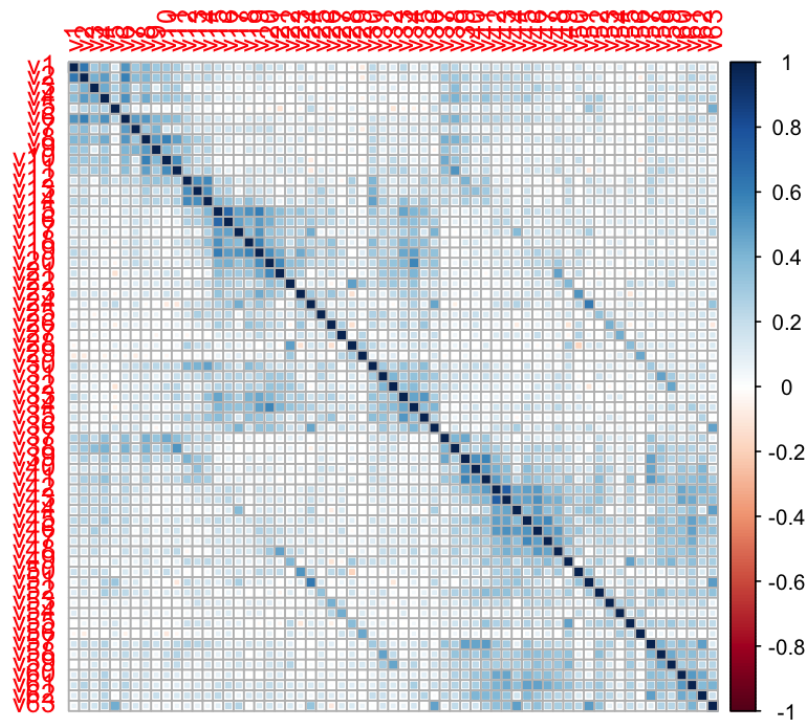
Teema	Muuttujat
Tyytyväisyys tutkintoon	v1: Työnantajat arvostavat tutkintoa v2: Suosittelevat koulutusta muille v3: Osaamistavoitteet tuotiin selkeästi esiin v4: Koulutus antoi riittävät valmiudet työelämään v5: Koulutuksen aikana tuotiin esille yrittäjyys uravaihtoehtona v6: Tyytyväisyys työuran kannalta
Työtilanne	v7: Tyytyväisyys työuraan
Työllistymisen laatu	v8: Opintojen hyödyntäminen nykyisessä työssä v9: Työ vastaa vaatimustasolta koulutusta
Työelämässä tärkeät tiedot ja taidot	v10–v36: Tietojen ja taitojen merkitys nykyisessä työssä (esim. viestintä-, yhteistyö- ja ongelmanratkaisutaidot)
Yliopisto-opiskelun kehittämät valmiudet	v37–v63: Miten yliopisto-opiskelu kehitti edellä mainittuja tietoja ja taitoja

Koulutusalojen koodaukset on esitetty taulukossa 2.3. Aloilta 6 (oikeustieteellinen) ja 9 (teologia) ei ollut vastaajia, joten ne on jätetty analyysistä pois.

**Taulukko 2.3.** Koulutusalat

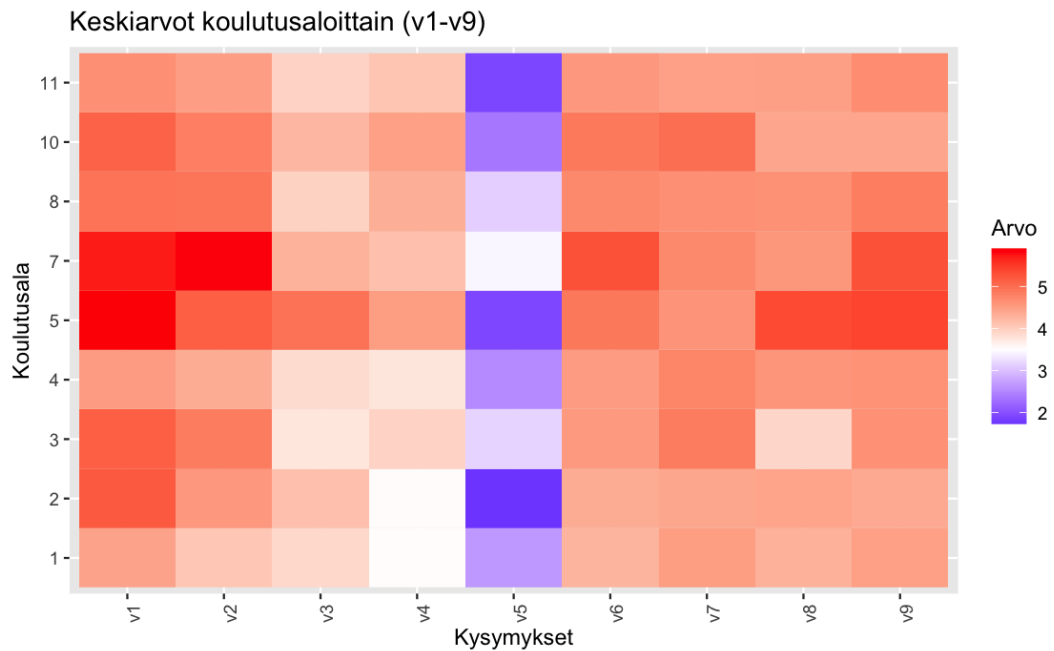
Koodi	Koulutusala	Koodi	Koulutusala
1	Humanistinen	7	Psykologian
2	Kasvatustieteellinen	8	Teknillistieteellinen
3	Kauppateieteellinen	10	Terveystieteiden
4	Luonnontieteellinen	11	Yhteiskuntatieteellinen
5	Lääketieteellinen		

Muuttujien välisiä yhteyksiä tarkastellaan korrelaatiomatriisissa 2.1.

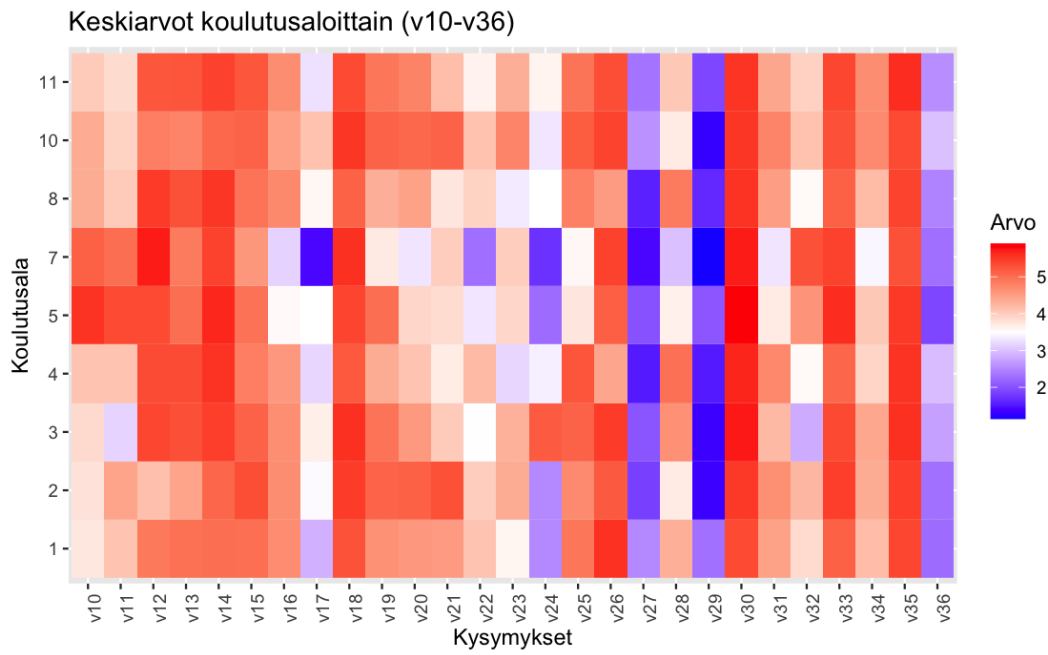


**Kuva 2.1.** Muuttujien korrelaatiot

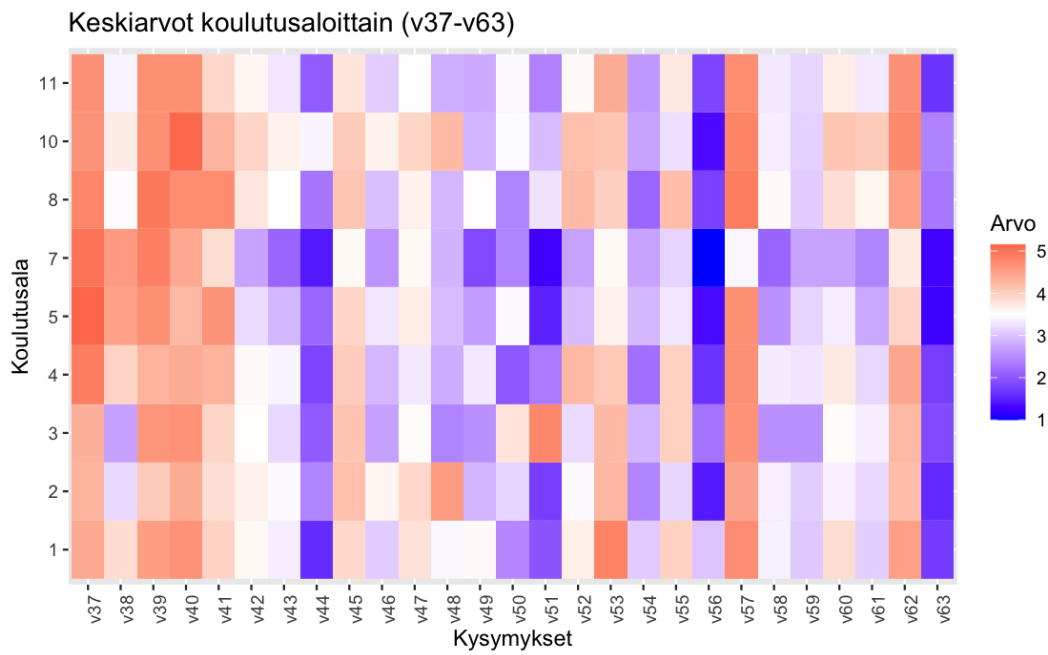
Muuttujien keskiarvot koulutusaloittain on visualisoitu kuvissa 2.2, 2.3, ja 2.4.



**Kuva 2.2.** Tyytyväisyys tutkintoon, työtilanne ja työllistymisen laatu-teemat



**Kuva 2.3.** Kuinka tärkeitä tiedot ja taidot ovat nykyisessä työssä- teema



**Kuva 2.4.** Miten yliopisto- opiskelu kehitti näitä taitoja- teema

# 3 Faktorianalyysin teoria ja estimoiminen

## 3.1 Faktorianalyysin perusmalli

Faktorianalyysissä pyritään esittämään alkuperäiset muuttujat pienemmällä määrällä faktoreita ilman merkittävää tiedon menetystä. Tavoitteena on selvittää, mitkä muuttujat selittävät toistensa vaihtelua ja millainen yhteinen tekijä eli faktori on tämän vaihtelun taustalla, samalla kun malli kuvaa aineiston yhteisvaihtelua mahdollisimman kattavasti. [3, s. 650]

Faktoriratkaisujen keskeiset ehdot ovat siis seuraavat:

1. Faktoreiden pitäisi selittää yhteisvaihtelusta mahdollisimman paljon.
2. Faktoreita pitäisi olla mahdollisimman vähän.
3. Mallin tulisi sisältää mahdollisimman paljon joko itseisarvoltaan suuria tai pieniä latauksia, kun taas keskinkertaisia latauksia tulisi olla mahdollisimman vähän.
4. Faktoreille tulee olla sisällöllisesti mielekäs tulkinta.

[3, s. 657]

Faktorimalli voidaan esittää seuraavalla kaavalla

$$(3.1) \quad \mathbf{y} - \boldsymbol{\mu} = \mathbf{A}\boldsymbol{\eta} + \boldsymbol{\epsilon},$$

missä  $\mathbf{y} = (y_1, y_2, \dots, y_p)'$  on havaintomuuttujien vektori,  $\boldsymbol{\mu} = (\mu_1, \mu_2, \dots, \mu_p)'$  on havaintomuuttujien keskiarvojen vektori,  $\mathbf{A}$  on  $p \times k$ -ulotteinen faktoreiden latausmatriisi,  $\boldsymbol{\eta} = (\eta_1, \eta_2, \dots, \eta_k)'$  on piilevien faktoreiden vektori ja  $\boldsymbol{\epsilon} = (\epsilon_1, \epsilon_2, \dots, \epsilon_p)'$  on jäännöstermien vektori. Aukikirjoitettuna esitys on seuraavanlainen:

$$(3.2) \quad \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_p \end{pmatrix} - \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{pmatrix} = \begin{pmatrix} \lambda_{11} & \lambda_{12} & \cdots & \lambda_{1k} \\ \lambda_{21} & \lambda_{22} & \cdots & \lambda_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{p1} & \lambda_{p2} & \cdots & \lambda_{pk} \end{pmatrix} \begin{pmatrix} \eta_1 \\ \eta_2 \\ \vdots \\ \eta_k \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_p \end{pmatrix}.$$

[8, s. 438]

Malleja (3.1) ja (3.2) havainnollistetaan seuraavalla esimerkillä.

Oletetaan, että tarkastelemme tilannetta, jossa on viisi havaintomuuttujaa ja kaksi piilevää faktoria. Tällöin faktorimalli voidaan kirjoittaa muodossa

$$\begin{aligned}
 y_1 - \mu_1 &= \lambda_{11}\eta_1 + \lambda_{12}\eta_2 + \epsilon_1 \\
 y_2 - \mu_2 &= \lambda_{21}\eta_1 + \lambda_{22}\eta_2 + \epsilon_2 \\
 y_3 - \mu_3 &= \lambda_{31}\eta_1 + \lambda_{32}\eta_2 + \epsilon_3 \\
 y_4 - \mu_4 &= \lambda_{41}\eta_1 + \lambda_{42}\eta_2 + \epsilon_4 \\
 (3.3) \quad y_5 - \mu_5 &= \lambda_{51}\eta_1 + \lambda_{52}\eta_2 + \epsilon_5.
 \end{aligned}$$

Esimerkiksi  $y_1, y_2, \dots, y_5$  voivat edustaa kyselylomakkeen viittä vastausta. Piilevät faktorit  $\eta_1$  ja  $\eta_2$  voivat kuvata esimerkiksi yleistä tyytyväisyyttä koulutukseen ja työuraan sekä palkkatyytyväisyyttä. Latauskertoimet  $\lambda_{ij}$  kuvaavat, kuinka voimakkaasti kukin havaintomuuttuja liittyy piileviin faktoreihin. Faktorianalyysin perusmalli perustuu oletuksiin, että  $E(\mathbf{y}) = \boldsymbol{\mu}$ ,  $E(\boldsymbol{\eta}) = \mathbf{0}$ ,  $E(\boldsymbol{\epsilon}) = \mathbf{0}$ ,  $cov(\boldsymbol{\eta}) = \mathbf{I}_k$ ,  $cov(\boldsymbol{\epsilon}) = \boldsymbol{\Psi}$ ,  $cov(\boldsymbol{\eta}, \boldsymbol{\epsilon}) = \mathbf{0}$ . [8, s. 439]

Seuraavaksi tarkastellaan rotaatioiden vaikutusta faktorimalliin. Aloitetaan kovarianssimatriisin esityksestä. Edellisten oletusten perusteella kovarianssimatriisi voidaan johtaa seuraavasti

$$(3.4) \quad \boldsymbol{\Sigma} = cov(\mathbf{y}) = cov(\mathbf{y} - \boldsymbol{\mu}) = cov(\boldsymbol{\Lambda}\boldsymbol{\eta} + \boldsymbol{\epsilon}).$$

Koska  $\boldsymbol{\Lambda}\boldsymbol{\eta}$  ja  $\boldsymbol{\epsilon}$  ovat oletuksen mukaan korreloimattomia, voidaan kovarianssimatriisi esittää summana

$$(3.5) \quad \boldsymbol{\Sigma} = cov(\boldsymbol{\Lambda}\boldsymbol{\eta}) + cov(\boldsymbol{\epsilon}).$$

Oletusten  $cov(\boldsymbol{\eta}) = \mathbf{I}_k$ ,  $cov(\boldsymbol{\epsilon}) = \boldsymbol{\Psi}$  ja  $E(\boldsymbol{\epsilon}) = \mathbf{0}$  perusteella saadaan kovarianssimatriisille seuraavanlainen muoto

$$\begin{aligned}
 (3.6) \quad \boldsymbol{\Sigma} &= cov(\boldsymbol{\Lambda}\boldsymbol{\eta}) + cov(\boldsymbol{\epsilon}) \\
 &= \boldsymbol{\Lambda}\boldsymbol{\Lambda}' + \boldsymbol{\Psi},
 \end{aligned}$$

jossa  $\boldsymbol{\Lambda}\boldsymbol{\Lambda}'$  on  $p \times p$  -kokoinen matriisi, joka kuvaa muuttujien välistä kovarianssia, joka selittyy yhteisten faktoreiden avulla. [8, s. 439]

Faktorianalyysissä rotaation tarkoituksena on tehdä faktorimallista tulkinnan kannalta selkeämpi. Rotaatiossa muutetaan faktoriratkaisua siten, että yksittäisten muuttujien lataukset maksimoidaan yhteen faktoriin, samalla kun lataukset muille faktoreille pyritään minimoimaan. Tavoitteena on, että yhden muuttujan vaihtelua kuvataisiin mahdollisimman vähällä määrällä faktoreita. [3, s. 660]

Rotaatiotapoja on kaksi: ortogonaalinen, kun faktorit eivät saa korreloida keskenään, ja vinokulmainen, kun faktorit voivat korreloida keskenään. [2, s. 612]

Otetaan tarkasteluun, miten rotaatiot vaikuttavat faktorimalliin. Alkuperäinen faktorimalli oli muotoa  $y - \mu = \Lambda\eta + \epsilon$ . Ortogonaalisessa rotaatiossa  $\Lambda$  muuttuu muotoon

$$(3.7) \quad \Lambda^* = \Lambda T,$$

jossa  $TT' = I$ .

Faktorit muuttuvat myös muotoon

$$(3.8) \quad \eta^* = T'\eta.$$

Rotatoitu faktorimalli on tällöin seuraavanlainen

$$(3.9) \quad y - \mu = \Lambda^* \eta^* + \epsilon$$

$$(3.10) \quad = (\Lambda T)(T'\eta) + \epsilon.$$

Rotaation myötä kovarianssimatriisi  $\Sigma$  ja havaitut muuttujat eivät muutu. Kovarianssimatriisin tapauksessa, joka oli

$$(3.11) \quad \Sigma^* = \Lambda^* \Lambda'^* + \Theta$$

$$(3.12) \quad = (\Lambda T)(\Lambda T)' + \Theta$$

$$(3.13) \quad = \Lambda \Lambda' + \Theta,$$

$TT' = I$  rotaatio ei muuta kovarianssia. Uudet lataukset ja uudet faktorit säilyttävät alkuperäisten faktorien ominaisuuksia  $E(\eta^*) = \mathbf{0}$ ,  $cov(\eta^*) = I$ ,  $cov(\eta^*, \epsilon) = \mathbf{0}$ . Jos ehdot eivät täyty, kovarianssimatriisit pysyvät erilaisina. [8, s. 442]

Rotaatio ei siis muuta faktorien selitysosuutta, vaan rotatoimalla mallista saadaan helposti tulkittava [3, s. 660].

Koska vinokulmaisessa rotaatiossa faktorit voivat korreloida keskenään, ortogonaalisen rotaatiomatriisin sijaan käytetään ei-singulaarista muunnosmatriisia  $Q$ , jolloin faktorit muuttuvat muotoon

$$(3.14) \quad \eta^* = Q'\eta$$

ja

$$(3.15) \quad cov(\eta^*) = Q'Q \neq I.$$

Tällöin malli 3.7 ja 3.8 muuttuvat seuraavasti

$$(3.16) \quad \Lambda^* = \Lambda Q,$$

$$(3.17) \quad \eta^* = Q' \eta.$$

Korrelaation vuoksi muuttujien ja faktorien välisten yhteyksien tulkinta muuttuu. Vinokulmaisen rotaation tuloksena saadaan yleensä kolme erillistä matriisiä: pattern-matriisi  $\Lambda^*$ , joka sisältää faktoreiden regressiokertoimet muuttujille, rakennematriisi  $S = \Lambda^* \Phi$ , joka kuvaa muuttujien ja faktoreiden välisiä korrelaatioita sekä faktoreiden korrelaatiomatriisi  $\Phi = \text{cov}(\eta^*)$ . Pattern-matriisiä pidetään tulkinnan kannalta ensisijaisena, koska sen lataukset edustavat muuttujien luonnollisia koordinaatteja vinokulmaisten akselien suunnassa. Koska vinokulmainen rotaatio ei säilytä faktoreiden välistä ortogonaalisuutta eikä välttämättä yhteisvaihtelun määrää, se voi kuitenkin johtaa paremmin tulkittavaan ratkaisuun. [8, s. 463]

## 3.2 Estimointi pääakselimenetelmällä

Tässä tutkielmassa latausten estimoimiseen käytetään pääakselimenetelmää. Faktoriaaliansalysissä lataukset kuvaavat, kuinka paljon kukin faktori selittää kunkin muuttujan vaihtelua. Toisin sanoen lataukset ilmaisevat, kuinka voimakkaasti kukin havaittu muuttuja liittyy tiettyyn faktoriin. Jokaisella faktorilla on oma lataus jokaista muuttujaa kohtaan, ja mitä suurempi on latauksen itseisarvo, sitä enemmän kyseinen faktori selittää kyseisen muuttujan vaihtelua. Yleisesti ottaen voidaan sanoa, että muuttuja latautuu faktorille, jos sen lataus kyseistä faktoria kohti on suuri verrattuna muihin faktoreihin. [9, s. 416]

Menetelmän tavoitteena on saada hajotelma muotoon  $S - \Psi \approx \hat{\Lambda} \hat{\Lambda}'$ , missä  $\hat{\Lambda}$  on  $p \times m$  latausmatriisi ja  $\Psi$  on yksittäisten muuttujien varianssi [8, s.448]. Tässä yhteydessä  $S$  on otoskovarianssimatriisi, joka lasketaan havaintoaineistosta  $y_1, y_2, \dots, y_n$  [8, s.443].

Aluksi kommunaliteetti  $\hat{h}_i^2 = s_{ii} - \hat{\psi}_i$  eli yhteisvarienssit estimoidaan. Kommunali-teetti mittaa, kuinka hyvin yksittäisen muuttujan vaihtelua voidaan kuvata yhdessä kaikkien faktoreiden avulla [8, s. 448].

Näiden diagonaaliarvojen kanssa  $\mathbf{S} - \hat{\Psi}$  saa seuraavanlaisen muodon

$$(3.18) \quad \mathbf{S} - \hat{\Psi} = \begin{bmatrix} \hat{h}_1^2 & s_{12} & \dots & s_{1p} \\ s_{21} & \hat{h}_2^2 & \dots & s_{2p} \\ \vdots & \vdots & & \vdots \\ s_{p1} & s_{p2} & \dots & \hat{h}_p^2 \end{bmatrix}$$

[8, s. 449].

Menetelmässä oletetaan alustavasti yksinkertaisuuden vuoksi, että  $\Psi = 0$ , jolloin saadaan

$$(3.19) \quad \mathbf{S} = \hat{\Lambda}\hat{\Lambda}'.$$

Koska  $\mathbf{S}$  on symmetrinen ja positiivisemidefiniittinen matriisi, voimme käyttää ominaisarvohajotelmaa

$$(3.20) \quad \mathbf{S} = \mathbf{C}\mathbf{D}\mathbf{C}',$$

missä  $\mathbf{C}$  on ortogonaalinen matriisi, jonka sarakkeina ovat  $\mathbf{S}$ :n normitetut ominaisvektorit ( $\mathbf{c}_i'\mathbf{c}_i = 1$ ), ja  $\mathbf{D}$  on diagonaalimatriisi, jonka lävistäjällä ovat  $\mathbf{S}$ :n ominaisarvot  $\theta_1, \theta_2, \dots, \theta_p$

$$(3.21) \quad \mathbf{D} = \begin{bmatrix} \theta_1 & 0 & \dots & 0 \\ 0 & \theta_2 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & \theta_p \end{bmatrix}.$$

Tässä ominaisarvot merkitään  $\theta$ , jotta niitä ei sekoiteta latauksiin, jotka yleensä merkitään  $\lambda$ -kirjaimella. Jotta  $\mathbf{S}$  voidaan esittää muodossa  $\hat{\Lambda}\hat{\Lambda}'$ , hajotamme diagonaalimatriisin  $\mathbf{D}$  seuraavasti

$$(3.22) \quad \mathbf{D} = \mathbf{D}^{1/2}\mathbf{D}^{1/2},$$

missä

$$(3.23) \quad \mathbf{D}^{1/2} = \begin{bmatrix} \sqrt{\theta_1} & 0 & \dots & 0 \\ 0 & \sqrt{\theta_2} & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & \sqrt{\theta_p} \end{bmatrix}.$$

Korvaamalla tämä alkuperäiseen ominaisarvohajotelmaan, saadaan

$$(3.24) \quad \mathbf{S} = \mathbf{C}\mathbf{D}\mathbf{C}' = \mathbf{C}\mathbf{D}^{1/2}\mathbf{D}^{1/2}\mathbf{C}' = (\mathbf{C}\mathbf{D}^{1/2})(\mathbf{C}\mathbf{D}^{1/2})',$$

jolloin voimme valita latausmatriisiksi  $\hat{\mathbf{A}} = \mathbf{C}\mathbf{D}^{1/2}$ . Dimensioiden vähentämisessä valitaan vain ne ominaisarvot ja ominaisvektorit, jotka selittävät suurimman osan varianssista. Jos valitsemme  $m$  suurinta ominaisarvoa ja niiden vastaavat ominaisvektorit, saadaan approksimointi

$$(3.25) \quad \mathbf{S} \approx \hat{\mathbf{A}}\hat{\mathbf{A}}',$$

missä  $\mathbf{D}_m = \text{diag}(\theta_1, \theta_2, \dots, \theta_m)$  sisältää suurimmat  $m$  ominaisarvot ( $\theta_1 > \theta_2 > \dots > \theta_m$ ),  $\mathbf{C}_m = (\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_m)$  sisältää niiden vastaavat ominaisvektorit. Tällöin estimaattorina voidaan käyttää

$$(3.26) \quad \hat{\mathbf{A}} = \mathbf{C}_m\mathbf{D}_m^{1/2} = (\sqrt{\theta_1}\mathbf{c}_1, \sqrt{\theta_2}\mathbf{c}_2, \dots, \sqrt{\theta_m}\mathbf{c}_m),$$

missä  $\hat{\mathbf{A}}$  on  $p \times m$  -matriisi ja  $\mathbf{D}_m^{1/2}$  on  $m \times m$  -diagonaalimatriisi. [8, s. 443]

### 3.3 Faktoripisteiden estimointi

Faktoripisteillä tarkoitetaan kullekin havainnolle laskettuja arvioita faktoreista. Ne ovat estimaatteja, jotka kuvaavat, miten vahvasti kukin havainto liittyy taustalla vaikuttaviin faktoreihin. Faktoripisteet ovat kiinnostavia varsinkin jatkoanalyysissä, kun halutaan selvittää, kuinka havainnot käyttäytyvät suhteessa faktoreihin. [8, s. 466] Koska faktorit eivät ole suoraan havaittavissa, ne arvioidaan havaituista muuttujista  $y_1, y_2, \dots, y_p$ . Yleisin tapa faktoripisteiden estimointiin on regressiopohjainen menetelmä. Koska faktorianalyysissä oletetaan, että  $E(\boldsymbol{\eta}_i) = \mathbf{0}$ , faktorit voidaan esittää  $\eta_i$  havaittujen muuttujien  $y_i$  avulla seuraavalla regressiomallilla

$$(3.27) \quad \begin{aligned} \eta_1 &= \beta_{11}(y_1 - \bar{y}_1) + \beta_{12}(y_2 - \bar{y}_2) + \dots + \beta_{1p}(y_p - \bar{y}_p) + \epsilon_1 \\ \eta_2 &= \beta_{21}(y_1 - \bar{y}_1) + \beta_{22}(y_2 - \bar{y}_2) + \dots + \beta_{2p}(y_p - \bar{y}_p) + \epsilon_2 \\ &\vdots \\ \eta_m &= \beta_{m1}(y_1 - \bar{y}_1) + \beta_{m2}(y_2 - \bar{y}_2) + \dots + \beta_{mp}(y_p - \bar{y}_p) + \epsilon_m, \end{aligned}$$

mikä voidaan kirjoittaa matriisimuodossa

$$(3.28) \quad \boldsymbol{\eta} = \mathbf{B}'_1(\mathbf{y} - \bar{\mathbf{y}}) + \boldsymbol{\epsilon},$$

missä  $\mathbf{B}'_1$  on tuntematon painotettu matriisi.

Malli pätee jokaiselle havainnolle

$$(3.29) \quad \eta_i = \mathbf{B}'_1(\mathbf{y}_i - \bar{\mathbf{y}}) + \epsilon_i, i = 1, 2, \dots, n.$$

Transpoosi muodossa mallista tulee seuraavanlainen

$$(3.30) \quad \eta'_i = (\mathbf{y}_i - \bar{\mathbf{y}})' \mathbf{B}_1 + \epsilon'_i, i = 1, 2, \dots, n,$$

nämä yhtälöt voidaan yhdistää yhdeksi malliksi

$$(3.31) \quad \boldsymbol{\eta} = \begin{pmatrix} \eta'_1 \\ \eta'_2 \\ \vdots \\ \eta'_n \end{pmatrix} = \begin{pmatrix} (\mathbf{y}_1 - \bar{\mathbf{y}})' \mathbf{B}_1 \\ (\mathbf{y}_2 - \bar{\mathbf{y}})' \mathbf{B}_1 \\ \vdots \\ (\mathbf{y}_n - \bar{\mathbf{y}})' \mathbf{B}_1 \end{pmatrix} + \begin{pmatrix} \epsilon'_1 \\ \epsilon'_2 \\ \vdots \\ \epsilon'_n \end{pmatrix} = \begin{pmatrix} (\mathbf{y}_1 - \bar{\mathbf{y}})' \\ (\mathbf{y}_2 - \bar{\mathbf{y}})' \\ \vdots \\ (\mathbf{y}_n - \bar{\mathbf{y}})' \end{pmatrix} \mathbf{B}_1 + \boldsymbol{\Theta}$$

$$(3.32) \quad = \mathbf{Y}_c \mathbf{B}_1 + \boldsymbol{\Theta}.$$

Estimaatti  $\mathbf{B}_1$  faktoripisteiden laskemiseen on

$$(3.33) \quad \hat{\mathbf{B}}_1 = (\mathbf{Y}'_c \mathbf{Y}_c)^{-1} \mathbf{Y}'_c \boldsymbol{\eta}.$$

Painotusmatriisi  $\mathbf{B}'_1$  valitaan siten, että se minimoi jäännöstermin  $\epsilon$  varianssin. Tämä saadaan regressioestimaatilla, ja matriisi saadaan laskettua kaavasta

$$(3.34) \quad \hat{\mathbf{B}}_1 = \mathbf{S}^{-1} \hat{\boldsymbol{\Lambda}}.$$

Mallista 3.31 estimoidut  $\eta_i$  arvot ovat

$$(3.35) \quad \hat{\boldsymbol{\eta}} = \begin{pmatrix} \hat{\eta}'_1 \\ \hat{\eta}'_2 \\ \vdots \\ \hat{\eta}'_n \end{pmatrix} = \mathbf{Y}_c \hat{\mathbf{B}}_1 = \mathbf{Y}_c \mathbf{S}^{-1} \hat{\boldsymbol{\Lambda}}.$$

[8, s. 467]

## 4 Aineiston analysointi

### 4.1 Faktorikelpoisuus ja faktoreiden valinta

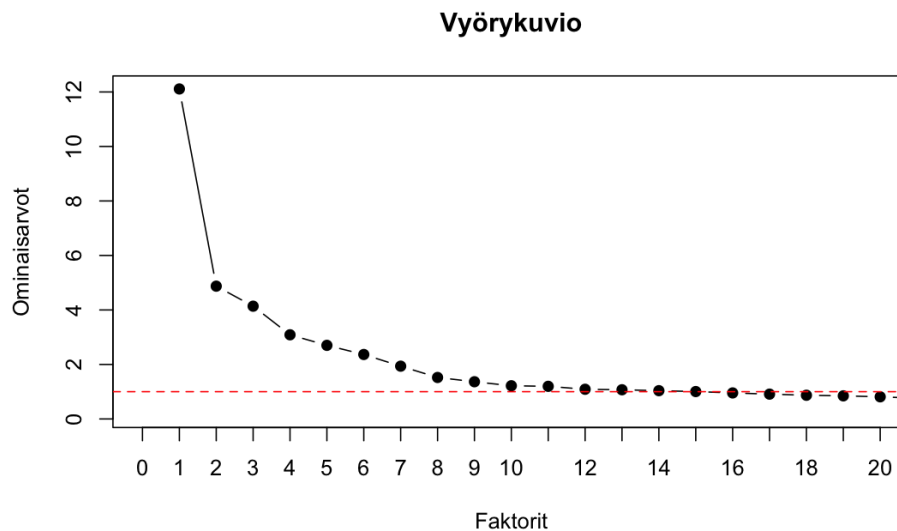
Ennen varsinaista faktorianalyysiä tutkitaan Kaiser-Meyer-Olkin testillä (KMO), onko aineisto sopiva faktorianalyysiin. KMO-testissä tarkastellaan korrelaatioiden ja osittaiskorrelaatioiden välistä suhdetta [4, s.46]. Jos arvo on yli 0,6, aineistoa voidaan pitää soveltuvana faktorianalyysiin.

**Taulukko 4.1.** KMO-testin tulokset

Kokonaisarvo	0,89
Muuttujakohtaiset arvot	> 0,64

Taulukon 4.1 tulosten mukaan aineisto on sopiva faktorianalyysiin.

Seuraavaksi tarkastellaan montako faktoria kannattaa harkita faktorimalliin vyörykuvion 4.1 avulla.



**Kuva 4.1.** Vyörykuvio

Kuvassa 4.1 punainen katkoviiiva havainnollistaa ominaisarvon 1 rajan, jota käytetään faktoreiden valintakriteerinä. Tässä aineistossa ominaisarvoltaan yli yhden olevia faktoreita on 1–10. On kuitenkin tärkeää huomata, että faktoreja ei valita pelkästään

ominisarvon perusteella, vaan olennaista on löytää faktoreille selkeä sanallinen selitys. Vyörykuviota tulkitaan usein kyynärpään kohdan perusteella, eli siellä, missä käyrä laskee jyrkästi. Kuvasta 4.1 huomataan, että kyynärpään kohta sijaitsee noin kahden faktorin kohdalla.

Tarkastellaan seuraavaksi kymmenen faktorin mallia. Taulukko 4.2 esittelee faktorimallin, johon on käytetty kovarianssimatriisia sekä pääakselimenetelmää. Mallia ei tässä kohtaa rotaatioitu. Pää tarkoituksena on tarkastella kumulatiivista varianssia ja selvittää monenko faktorin malli selittäisi mahdollisimman suuren osan aineiston kokonaisvaihtelusta.

**Taulukko 4.2.** Kymmenen faktorin mallin tulokset

Faktori	1	2	3	4	5	6	7	8	9	10
Lataukset	12,42	4,36	3,72	2,44	2,43	1,94	1,52	1,13	0,80	0,69
Selitetty varianssi	20 %	7 %	6 %	4 %	4 %	3 %	2 %	2 %	1 %	1 %
Kumulatiivinen varianssi	20 %	27 %	33 %	36 %	40 %	43 %	46 %	48 %	49 %	50 %

Taulukosta 4.2 huomataan, että lataukset eli korrelaatiot alkuperäisten muuttujien ja faktoreiden välillä ovat korkeimmillaan ensimmäisessä faktorissa. Selitetty varianssi kertoo, kuinka hyvin kukin faktori selittää alkuperäisten muuttujien varianssia. Esimerkiksi ensimmäisen faktorin selitetty varianssi selittää 20 prosenttia alkuperäisten muuttujien varianssista. Kumulatiivinen varianssi taas näyttää, kuinka paljon faktorit yhteensä selittävät alkuperäistä aineistoa. Taulukon 4.2 perusteella faktorimallit, joissa on seitsemän, kahdeksan ja yhdeksän faktoria, eivät lisää enää merkittävästi selitettävää varianssia. Voisi olla järkevää valita malli, jossa on vähintään neljä mutta enintään seitsemän faktoria. Tässä tutkielmassa päädytään viiden faktorin malliin, koska aineisto jakautuu viiteen teemaan. Pienempi määrä faktoreita tekee mallista myös helpommin tulkittavan.

## 4.2 Faktorimallin rakentaminen

Faktorimallin rakentamiseen R-ohjelmistoa käyttäen valitaan kovarianssimatriisi, pääakselimenetelmä, vinokulmainen oblimin-rotatio sekä viisi faktoria. Seuraava taulukko 4.3 esittelee kyseisen mallin tuloksia.

**Taulukko 4.3.** Viiden faktorin mallin tulokset

Faktori	1	2	3	4	5
Lataukset	10,89	9,79	9,49	6,46	5,31
Selitetty varianssi	10 %	9 %	9 %	6 %	5 %
Kumulatiivinen varianssi	10 %	20 %	29 %	35 %	40 %

Tehdyn faktorianalyysin perusteella faktorit selittävät 40 prosenttia aineiston kokonaisvaihtelusta eli kohtalaisesti.

Seuraavaksi tarkastellaan faktorimallista saadut lataukset. Taulukko 4.4 esittelee kahdenkymmenen ensimmäisen muuttujan lataukset. Taulukosta on poistettu arvot, jotka ovat alle 0,2.

**Taulukko 4.4.** Faktorimallin lataukset

Muuttuja	1. faktori	2. faktori	3. faktori	4. faktori	5. faktori
v1			0,63		
v2			0,76		
v3	0,27		0,46		
v4		-0,22	0,76	0,22	
v5		-0,43	0,26	0,48	
v6			0,81		
v7			0,52		
v8			0,89		
v9	-0,21		0,86		
v10			0,94		
v11		0,26	0,83	-0,28	
v12	-0,20		0,56	0,26	0,29
v13			0,48	0,22	0,27
v14		0,23	0,40		0,23
v15		0,66	0,20	0,23	
v16		0,65		0,38	
v17		0,77		0,63	
v18		0,55			
v19		0,90		0,31	
v20		0,89			

Taulukosta 4.4 havaitaan, että esimerkiksi muuttaja v3 latautuu kahdelle faktorille, jonka perusteella voidaan sanoa, että "osaamistavoitteet tuotiin selkeästi esiin"kuvaamiseen tarvitaan kaksi faktoria, kun taas muuttajalle v6 eli "tyytyväisyys tutkintoon työuran kannalta"kuvaaminen onnistuu yhden faktorin avulla. Valitaan saaduista tuloksista jokaiselle muuttujalle suurin faktorilataus, jonka mukaan se ryhmitellään. Taulukko 4.5 esittelee muuttujien ryhmittelyn faktorien mukaan.

**Taulukko 4.5.** Ryhmittely suurimman faktorilatauksen mukaan

<b>Faktori</b>	<b>Muuttujat</b>
1.	v40, v41, v42, v43, v44, v45, v46, v47, v51, v52, v53, v54, v57, v58, v59, v60, v61, v62, v63
2.	v15, v16, v17, v18, v19, v20, v21, v23, v25, v27, v31, v32, v33, v34, v35
3.	v1, v2, v3, v4, v6, v7, v8, v9, v10, v11, v12, v13, v14, v30, v37, v38, v39
4.	v5, v24, v36, v50
5.	v22, v26, v28, v29, v48, v49, v55, v56

Taulukon 4.5 mukaan ensimmäinen faktori kuvaa, kuinka hyvin opiskelijat kokevat yliopistokoulutuksen kehittäneen heidän työelämässä vaadittavia taitoja. Tässä faktorissa korostuvat erityisesti tiedonhankinta-, ongelmanratkaisu- ja viestintätaidot suomen kielellä. Nämä taidot näyttäytyvät vahvasti vain tässä faktorissa. Toinen faktori painottaa työssä tarvittavia osaamisalueita, eli niitä taitoja ja tietoja, joita vastaajat pitävät merkityksellisinä omassa nykyisessä työssään. Tähän faktoriin keskittyvät erityisesti opetus-, koulutus- ja ohjaustaidot sekä viestintä ruotsin kielellä, jotka eivät korostu muissa faktoreissa. Kolmas faktori kokoaa yhteen useita eri teemoja. Tässä faktorissa yhdistyvät opiskelijoiden arviot koulutuksen laadusta, sen merkityksestä työuran kannalta sekä opintojen tarjoamasta osaamisesta. Esimerkiksi tutkinnon arvostus työmarkkinoilla, koulutuksen vastaavuus työelämän tarpeisiin, tyytyväisyys tutkintoon ja urakehitykseen sekä opinnoista saatu teoreettinen ja käytännöllinen osaaminen korostuvat tässä kokonaisuudessa. Neljäs faktori liittyy yrittäjyyteen ja liiketoimintaosaamiseen. Se sisältää sekä koulutuksessa esiin tuodun yrittäjyysvalmiuden että työssä tarvittavan liiketoimintaan ja lainsäädäntöön liittyvän osaamisen. Viides faktori muodostuu viestintätaidoista eri kielillä sekä kyvystä

toimia monikulttuurisessa ympäristössä, jotka ovat sekä työelämän tärkeitä taitoja että yliopiston kehittämiä. Viestintä suomen kielellä korostuu ainoastaan työelämässä, kun taas englannin kieli näkyy sekä työelämässä että yliopiston kehittämässä taidoissa. Lisäksi opetus-, koulutus- ja ohjaustaidot liittyvät yliopiston kehittämien taitojen arviointiin.

### 4.3 Faktoripisteiden tarkastelu ja jatkoanalyysit

Faktoripisteet on laskettu kaavan 3.35 mukaisesti. Alla olevassa taulukossa 4.6 on esitetty kymmenen ensimmäisen havaintorivin pisteet viidestäsadastakahdeksasta. Vertailun vuoksi taulukossa 4.7 on esitetty faktoripisteet, jotka on laskettu R:n valmiilla funktiolla. Tässä funktiossa faktoripisteet lasketaan Thurstonen menetelmällä regressiopohjaisina estimaatteina, ja aineisto standardoidaan ennen analyysiiä automaattisesti.

**Taulukko 4.6.** Faktoripisteet  
(kaavan 3.35 mukaan)

	PA1	PA2	PA3	PA4	PA5
1	0,14	-0,62	-2,13	-0,78	-0,68
2	-0,96	-0,21	-0,05	1,73	0,03
3	0,56	0,88	0,13	0,37	0,17
4	-0,48	-2,85	-1,71	-0,61	-1,46
5	0,24	1,04	-0,73	-0,46	1,06
6	-0,82	-0,57	0,53	-1,12	0,62
7	1,69	-0,88	-2,66	-1,60	-0,12
8	0,54	0,21	1,05	-1,10	0,01
9	0,13	-1,43	0,83	-0,66	0,47
10	-0,92	0,33	-0,15	1,60	1,00

**Taulukko 4.7.** Faktoripisteet  
(R:n valmis funktio)

	PA1	PA2	PA3	PA4	PA5
1	-1,16	-1,47	-2,59	-0,97	-1,08
2	-0,82	-0,43	-0,24	2,59	-0,05
3	0,90	1,52	0,49	1,41	0,49
4	-2,15	-2,66	-2,55	-1,14	-1,80
5	1,07	-0,29	-0,05	-0,98	1,06
6	-1,66	-1,19	-1,08	-1,91	0,13
7	0,14	-2,64	-3,57	-2,52	-0,12
8	0,74	0,87	1,40	-1,56	0,47
9	-0,09	-1,19	0,78	-0,63	0,83
10	-0,65	1,26	0,09	1,96	1,41

Taulukoiden 4.6 ja 4.7 mukaan molemmat laskutavat tuottavat samansuuntaisia tuloksia, mutta R:n valmiin funktion tuottamat pisteet ovat laajemmin jakautuneita.

Seuraavaksi perehdytään varianssianalyysiin tutkiaksemme koulutusalojen välisiä eroja. Varianssianalyysillä testataan keskiarvojen eroja [3, s. 306]. Tässä tutkielmassa tarkastellaan, ovatko riippuvan muuttujan eli faktoripisteiden keskiarvot erisuuria eri koulutusalojen muodostamissa ryhmissä.

Varianssianalyyssissä pyritään hajottamaan aineiston yhteisvarianssi osatekijöihin ja arvioimaan näiden osatekijöiden perusteella ryhmien välisiä eroja. Ryhmien sisäinen varianssi kuvaa havaintoarvojen vaihtelua ryhmien sisällä, ja se lasketaan kaavalla

$$(4.1) \quad \frac{\sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2}{k(n-1)},$$

missä  $k$  on ryhmien lukumäärä ja  $n$  yksittäisen ryhmän havaintojen lukumäärä.

Ryhmien välinen varianssi mittaa ryhmien keskiarvojen poikkeamaa kokonaiskeskiarvosta seuraavasti

$$(4.2) \quad \frac{n \sum_{i=1}^k (\bar{y}_i - \bar{y}_{..})^2}{k(n-1)},$$

missä  $\bar{y} = \sum_{i=1}^k \bar{y}_i / k$  on kokonaiskeskiarvo. [8, s. 170]

Varianssianalyyssissä vertaillaan ryhmien välistä ja sisäistä varianssia F-testin avulla seuraavasti

$$(4.3) \quad F = \frac{MSH}{MSE} = \frac{\frac{SSH}{k-1}}{\frac{SSE}{k(n-1)}},$$

missä  $SSH = \frac{\sum \bar{y}_i^2 - \bar{y}^2}{kn}$  on ryhmien välinen neliösumma ja  $SSE = \frac{\sum \sum y_{ij}^2 - \sum \bar{y}_i^2}{n}$  on ryhmien sisäinen neliösumma.  $\frac{MSH}{MSE}$  ovat vastaavat keskineliösummat [8, s.171].

Hypoteesit asetetaan seuraavasti

$$H_0: \mu_1 = \mu_2 = \dots = \mu_k$$

$$H_1: \mu_i \neq \mu_j \quad \text{joillekin } i \neq j.$$

Hypoteeseja testataan vertaamalla laskettua  $F$ -testisuuretta  $p$ -arvoon. Jos  $p$ -arvo on pienempi kuin valittu merkitsevyystaso  $\alpha$ , hylätään nollahypoteesi  $H_0$  ja päätellään, että koulutusalojen välillä on tilastollisesti merkitsevä ero. Jos  $p$ -arvo on suurempi kuin  $\alpha$ , ei ole riittävästi näyttöä ryhmien välisistä eroista. [8, s.171]

Tarkastellaan varianssianalyyssin tuloksia faktorimallista 4.3 seuraavassa taulukossa 4.8.

**Taulukko 4.8.** ANOVA-tulokset

Faktori	Vapausaste	Neliösumma	Keskineliö	F-arvo	p-arvo
1	1	1,30	1,26	1,21	0,27
2	1	5,80	5,79	6,05	0,01
3	1	14,40	14,36	14,47	0,00
4	1	15,90	15,92	18,44	0,00
5	1	0,10	0,10	0,12	0,73

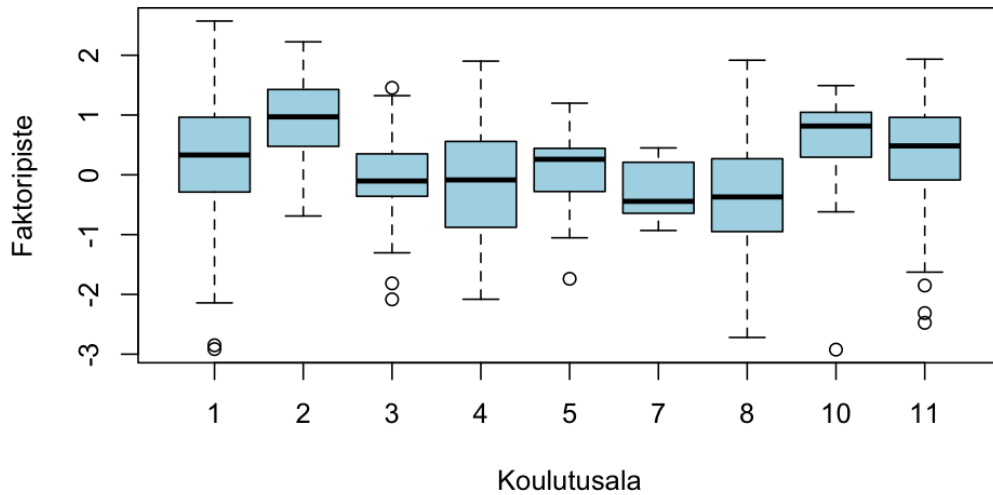
Taulukosta 4.8 huomataan, että toisen, kolmannen ja neljännen faktoripisteen kohdalla p-arvo on pienempi kuin valittu merkitsevyystaso 0,01. Tulos viittaa tilastollisesti merkitsevään eroon koulutusalojen välillä näiden kolmen faktoripisteen tapauksessa.

Tarkastellaan, mitkä koulutusalat eroavat merkitsevästi toisistaan edellä saatujen tulosten mukaan. Taulukko 4.9 esittelee Tukey HSD-testin tilastollisesti vain 0,01 merkitsevyystasolla merkittävät tulokset.

**Taulukko 4.9.** TukeyHSD-testin tulokset

<b>Toisessa faktoripisteessä</b>				
<b>Koulutusalat</b>	<b>Erotus</b>	<b>Alaraja</b>	<b>Yläraja</b>	<b>p-arvo</b>
Kasvatustieteellinen - Humanistinen (2-1)	0,67	0,11	1,24	0,0075
Teknillistieteellinen - Humanistinen (8-1)	-0,60	-0,98	-0,23	0,00003
Kasvatustieteellinen - Kauppatieteellinen (3-2)	-0,95	-1,63	-0,28	0,0004
Luonnontieteellinen - Kauppatieteellinen (4-2)	-1,07	-1,77	-0,37	0,00009
Teknillistieteellinen - Kauppatieteellinen (8-2)	-1,27	-1,78	-0,77	<0,0001
Teknillistieteellinen - Terveystieteiden (10-8)	0,89	0,18	1,60	0,0033
Yhteiskuntatieteellinen - Teknillistieteellinen (11-8)	0,75	0,38	1,12	<0,0001
<b>Kolmannessa faktoripisteessä</b>				
<b>Koulutusalat</b>	<b>Erotus</b>	<b>Alaraja</b>	<b>Yläraja</b>	<b>p-arvo</b>
Lääketieteellinen - Humanistinen (5-1)	1,40	0,66	2,15	0,0000003
Psykologian - Humanistinen (7-1)	1,60	0,43	2,77	0,0009
Teknillistieteellinen - Humanistinen (8-1)	0,67	0,28	1,07	0,000007
Lääketieteellinen - Kasvatustieteellinen (5-2)	1,44	0,61	2,26	0,000003
Psykologian - Kasvatustieteellinen (7-2)	1,64	0,41	2,86	0,0012
Teknillistieteellinen - Kasvatustieteellinen (8-2)	0,71	0,18	1,24	0,0012
Lääketieteellinen - Kauppatieteellinen (5-3)	1,07	0,22	1,91	0,0029
Yhteiskuntatieteellinen - Lääketieteellinen (11-5)	-1,01	-1,76	-0,27	0,0009
<b>Neljännessä faktoripisteessä</b>				
<b>Koulutusalat</b>	<b>Erotus</b>	<b>Alaraja</b>	<b>Yläraja</b>	<b>p-arvo</b>
Kauppatieteellinen - Humanistinen (3-1)	1,77	1,25	2,30	<0,0001
Teknillistieteellinen - Humanistinen (8-1)	0,80	0,46	1,13	<0,0001
Yhteiskuntatieteellinen - Humanistinen (11-1)	0,52	0,11	0,93	0,0024
Kauppatieteellinen - Kasvatustieteellinen (3-2)	2,02	1,42	2,63	<0,0001
Teknillistieteellinen - Kasvatustieteellinen (8-2)	1,04	0,59	1,49	<0,0001
Yhteiskuntatieteellinen - Kasvatustieteellinen (11-2)	0,77	0,27	1,27	0,00009
Luonnontieteellinen - Kauppatieteellinen (4-3)	-1,32	-1,96	-0,68	<0,0001
Lääketieteellinen - Kauppatieteellinen (5-3)	-1,94	-2,66	-1,22	<0,0001
Psykologian - Kauppatieteellinen (7-3)	-2,03	-3,08	-0,98	<0,0001
Teknillistieteellinen - Kauppatieteellinen (8-3)	-0,98	-1,45	-0,50	<0,0001
Terveystieteiden - Kauppatieteellinen (10-3)	-1,32	-2,07	-0,56	0,000003
Yhteiskuntatieteellinen - Kauppatieteellinen (11-3)	-1,25	-1,78	-0,73	<0,0001
Teknillistieteellinen - Lääketieteellinen (8-5)	0,96	0,37	1,55	0,00002

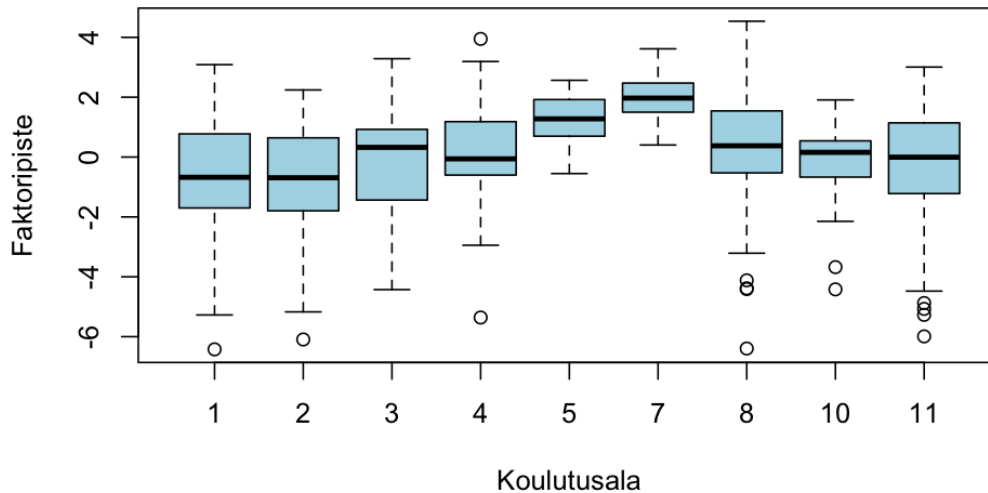
### Toinen faktoripiste koulutusaloittain



**Kuva 4.2.** Laatikkokuvio toisesta faktoripisteestä koulutusaloittain

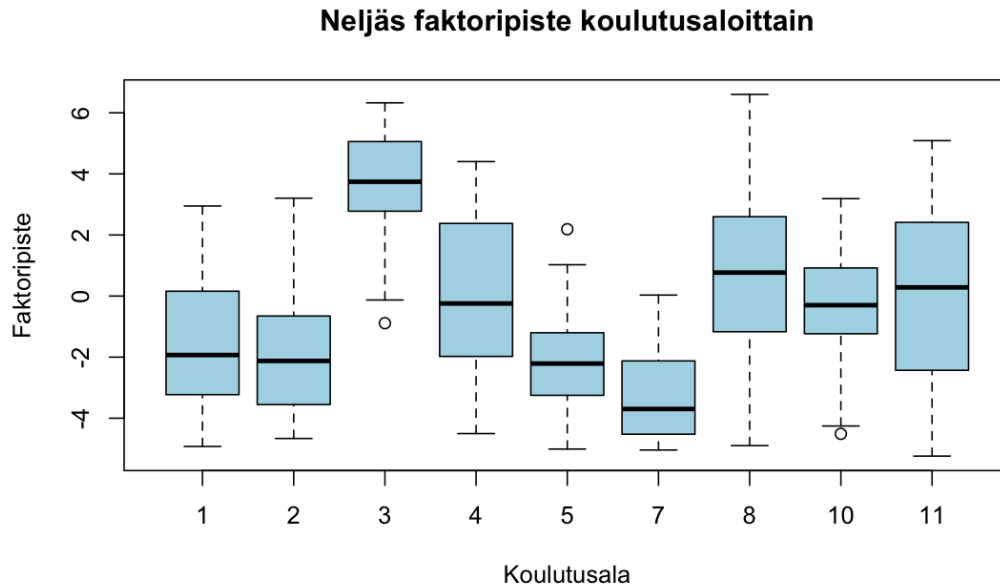
Kuvassa 4.2 havainnollistetaan faktoria, jossa yhdistyvät työssä tarvittavat tiedot ja taidot. Koulutusalojen väliset erot ovat vähäisiä. Suurinta hajontaa esiintyy luonnontieteiden alalla (4). Suurin mediaani on nähtävillä lääketieteellisellä alalla (5), kun taas matalimmat mediaaniarvot näkyvät psykologian alalla (7). Keskimäärin korkeita pisteitä saa kasvatustieteellinen koulutusala (2), mikä viittaa siihen, että opiskelijat näillä aloilla kokevat työssä tarvittavien tietojen ja taitojen olevan merkittäviä tai omaavansa vahvaa osaamista tällä alalla.

### Kolmas faktoripiste koulutusaloittain



**Kuva 4.3.** Laatikkokuvio kolmannesta faktoripisteestä koulutusaloittain

Kuvan 4.3 perusteella koulutusalojen väliset erot opiskelijoiden arvioissa koulutuksen laadusta, työelämävastaavuudesta ja opintojen tarjoamasta osaamisesta ovat vähäisiä. Kauppätieteellisen (3) ja terveystieteiden (10) alan mediaanit sijoittuvat korkealle, mikä viittaa siihen, että näiden alojen opiskelijat arvioivat koulutustaan myönteisemmin kuin muiden alojen opiskelijat. Suurinta hajontaa esiintyy humanistisella (1), kasvatustieteellisellä (2), kauppätieteellisellä (3) ja yhteiskuntatieteellisellä (11) alalla, eli opiskelijoiden kokemukset vaihtelevat paljon näillä aloilla. Luonnontieteellisellä alalla (4) mediaani on pieni, mikä viittaa opiskelijoiden heikompaan tyytyväisyyteen. Psykologian (7), terveystieteiden (10) ja lääketieteellisen koulutusalan (5) vastauksissa hajonta on vähäistä, mikä viittaa myönteisiin ja yhtenäisiin kokemuksiin näillä aloilla. Korkeimmat pisteet saavat psykologian koulutusalan (7) opiskelijat.



**Kuva 4.4.** Laatikkokuvio neljännestä faktoripisteestä koulutusaloittain

Neljäs faktori kokoaa yhteen yrittäjyys- ja liiketoimintaosaamisen, sisältäen muun muassa yrittäjyysvalmiuden ja lainsäädännön tuntemuksen. Kuvassa 4.4 näkyy, että neljännen faktoripisteen koulutusalojen väliset erot ovat selvästi suurempia kuin toisella 4.2 ja kolmannella 4.3. Yhteiskuntatieteellisen koulutusalan (11) hajonta on suurin, samoin luonnontieteiden alalla (4). Opiskelijoiden näkemykset yrittäjyys- ja liiketoimintataidoissa vaihtelevat siis voimakkaasti. Kauppatieteellinen koulutusala (3) erottuu muista korkeammilla pisteillä; opiskelijat arvioivat teeman tärkeäksi ja kokevat saaneensa siihen hyvät valmiudet. Psykologian koulutusalan (7) mediaani ja pisteet sijoittuvat matalimmiksi, mikä viittaa siihen, että opiskelijat ovat vähemmän tyytyväisiä kyseisiin asioihin ja arvioivat ryhmän sisällä keskimäärin huonommin kuin muilla aloilla.

Tarkastellaan seuraavaksi monen muuttujan varianssianalyysiä (MANOVA). Yksisuuntaisessa varianssianalyysissä tarkasteltiin, eroavatko eri koulutusalat yksittäisten faktoripisteiden perusteella. MANOVA:ssa puolestaan tutkitaan, eroavatko koulutusalat toisistaan, kun kaikkia viittä faktoripistettä tarkastellaan yhtäaikaisesti. [3, s.327]

Taulukossa 4.10 on esitetty MANOVA:n tulokset.

**Taulukko 4.10.** MANOVA-tulokset

Faktori	Vapausaste	Ryhmien välinen neliösumma	Jäännöseliösumma	Virhe (SE)
1	8	30,52	497,79	1,00
2	8	81,04	408,93	0,91
3	8	63,03	453,43	0,95
4	8	123,84	328,85	0,81
5	8	83,89	333,62	0,82

Taulukon 4.10 perusteella neljäs faktori erottuu muista, sillä se tuottaa suurimman ryhmien välisen neliösumman sekä pienimmän jäännöseliösumman ja residuaalivirheen (SE). Tämä viittaa siihen, että neljäs faktori selittää parhaiten koulutusalojen välisiä eroja. Myös toisen ja viidennen selittävät varsin merkittävästi vaihtelua, vaikka yksittäiset ANOVA-testit eivät olleet tilastollisesti merkitseviä.

Seuraavaksi testataan koulutusalojen eroja kaikissa faktoreissa. Asetetaan nollahypoteesi  $H_0$ , että koulutusalojen ryhmien keskiarvot kaikissa faktoripisteissä ovat yhtäsuuret, eli

$$\mu_1 = \mu_2 = \dots = \mu_k,$$

missä  $\mu_i$  on  $i$ -nnen ryhmän keskiarvovektori ja  $k$  on ryhmien lukumäärä.

Vaihtoehtoinen hypoteesi  $H_1$  on, että ainakin kahden ryhmän keskiarvot poikkeavat toisistaan, eli

$$\exists i, j \text{ siten, että } \mu_i \neq \mu_j.$$

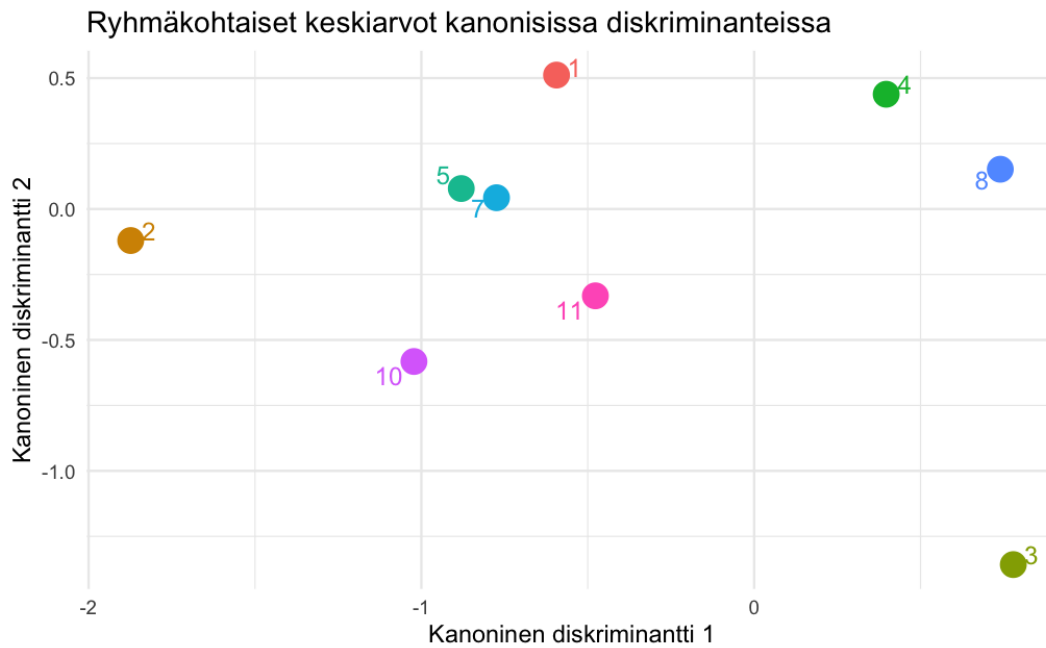
**Taulukko 4.11.** Wilksin testin tulokset

Wilksin testi	F-arvo	p-arvo
0,41	12,31	< 2,2e-16

Wilksin  $\Lambda$ :n arvon vaihteluväli on  $0 \leq \Lambda \leq 1$ , missä arvo 1 tarkoittaa, että ryhmien välillä ei ole eroja, ja arvo 0, että ryhmien välillä on merkittäviä eroja. [8, s.171]

Taulukossa 4.11 tulokseksi saatu arvo 0,41 ja erittäin pieni p-arvo viittaavat siihen, että koulutusalojen välillä on tilastollisesti merkitseviä eroja faktoripisteiden perusteella. Nollahypoteesi siten hylätään, jolloin vaihtoehtoinen hypoteesi astuu voimaan.

Monimuuttujaisten erojen tarkastelua jatketaan diskriminanttianalyysin avulla (kuva 4.5). Diskriminanttifunktiot ovat muuttujien lineaarisia yhdistelmiä, jotka on valittu siten, että ne erottavat ryhmät mahdollisimman hyvin toisistaan [8, s.282].



**Kuva 4.5.** Monimuuttujaiset erot koulutusaloissa

Kuvasta 4.5 nähdään, että koulutusalat kasvatustieteellinen (2) ja kauppatieteellinen (3) eroavat eniten muista ja toisistaan. Lääketieteellinen (5) ja psykologian (7) ala ovat hyvin samankaltaisia ja muodostavat selkeän parin. Luonnontieteellinen (4) ja teknillistieteellinen (8) sijoittuvat lähekkäin, mutta hieman erillään muista koulutusaloista.

## 5 Yhteenveto ja pohdinta

Tutkielman tavoitteena oli selvittää faktorianalyysin soveltuvuutta uraseurantakyselyaineistoon eksploratiivisella lähestymistavalla. Tarkoituksena oli muodostaa mahdollisimman hyvin tulkittava ja yksinkertainen faktorimalli, analysoida kysymysten taustalla vaikuttavia piileviä ilmiöitä sekä tarkastella koulutusalojen välisiä eroja muodostettujen faktoreiden perusteella.

Aineiston soveltuvuus faktorianalyysiin varmistettiin Kaiser-Meyer-Olkin-testillä. Faktorien määrän valinnassa käytettiin vyörykuviota, ja analyysin perusteella valittiin viiden faktorin malli, joka selitti 40 prosenttia aineiston kokonaisvaihtelusta. Faktorilatausten tarkastelussa havaittiin, että useimmat muuttujat latautuivat useammalle kuin yhdelle faktorille, mikä viittaa siihen, että näiden muuttujien kuvaamiseen tarvitaan useampi kuin yksi faktor.

Faktorien tulkinta tuotti seuraavanlaiset piilevät ilmiöt: ensimmäinen faktor liittyi siihen, kuinka hyvin yliopisto-opinnot kehittivät työelämässä tarvittavia tietoja ja taitoja. Toinen faktor kuvasi työssä tarvittavia tärkeitä tietoja ja taitoja. Kolmas faktor yhdisti näkemyksiä koulutuksen laadusta, sen merkityksestä työuran kannalta sekä opintojen tarjoamasta osaamisesta. Neljäs faktor käsitteli yrittäjyyttä, liiketoimintaosaamista ja lainsäädännön tuntemusta. Viides faktor kuvasi viestintäosaamista eri kielillä ja toimimista monikulttuurisessa ympäristössä niin työelämän tärkeyden kuin yliopisto-opintojen kehityksen näkökulmasta.

Faktoripisteiden pohjalta koulutusalojen välisiä eroja analysoitiin ANOVA:lla sekä MANOVA:lla. Tulokset osoittivat, että koulutusalat erosivat merkitsevästi toisistaan toisen, kolmannen ja neljännen faktoripisteen osalta. Myös MANOVA vahvisti, että koulutusalojen väliset erot olivat tilastollisesti merkitseviä.

Yhteenvetona voidaan todeta, että viiden faktorin malli oli sisällöllisesti tulkittava ja tarjosi kiinnostavan näkökulman aineiston rakenteeseen, vaikka se ei selittänyt suurta osuutta kokonaisvaihtelusta. Mallin selitysaste jäi alle 50 prosenttiin, ja faktoreiden päällekkäisyys viittaa siihen, että ilmiöt uraseurantakyselyissä ovat moniulotteisia ja vaativat monifaktoraalista tarkastelua.

# Lähteet

- [1] Finch, W. H. *Exploratory factor analysis*. Thousand Oaks, CA: SAGE Publications, 2020.
- [2] Coolican, H. *Research Methods and Statistics in Psychology*, 7th edition. Taylor & Francis Group, 2018.
- [3] Nummenmaa, L. (2021). *Tilastotieteen käsikirja*. Tammi.
- [4] Luoma, A. *Monimuuttujamenetelmät*. Tampere: Tampereen yliopisto, Kevät 2013.
- [5] Leskinen, A. *Faktorianalyysi*. Jyväskylä: Jyväskylän yliopisto, 1/1987
- [6] Nummi, A. *Multivariate Analysis*. Tampere: Tampereen yliopisto.
- [7] Metsämuuronen, J. *Monimuuttujamenetelmien perusteet*. Helsinki: International Methelp Ky.
- [8] Rencher, A. C., & Christensen, W. F. (2012). *Methods of Multivariate Analysis*. New York: John Wiley & Sons.
- [9] Nummenmaa, L. (2023). *Käyttäytymistieteiden tilastolliset menetelmät*. Tammi.
- [10] Tampereen yliopisto. *Uraseuranta*. <https://www.tuni.fi/fi/palvelut-ja-yhteistyo/alumneille/yliopiston-alumneille/uraseuranta>

## Liite A: R-koodit

```
#Aineiston muokkaukset
(sum(is.na(datatre5v14valm))/(ncol(datatre5v14valm)
*nrow(datatre5v14valm)))*100
# na-arvoja vain 2.73%, poistetaan ne
datak <- na.omit(datatre5v14valm)

datak <- lapply(datak, as.numeric)
datak <- as.data.frame(datak)
datak <- datak[, !names(datak) %in%
"Valmistumisen.jälkeisten.työnantajien.lukumäärä"]
datak <- datak[, !names(datak) %in%
"Tähänastisen.työuran.kokonaisuutta.kuvaa.parhaiten"]

data_subset <- datak[, 6:ncol(datak)] # kysymykset erikseen
# nimetään muuttujat lyhyemmin
rename_variables <- function(data) {
  # nimet muodossa v1, v2, ..., vn
  new_names <- paste0("v", seq_len(ncol(data)))
  colnames(data) <- new_names
  return(data)
}

data_subset <- rename_variables(data_subset)

#KMO
kmo <- KMO(data_subset)

# FAKTORIMALLI
fa10none <- fa(r = data_subset, nfactors = 10, fm = "pa",
rotate = "none", covar=TRUE)

faktorianalyyslop <- fa(r=data_subset, nfactors = 5, fm = "pa",
```

```

rotate = "oblimin",covar=TRUE)

# faktoripisteet
means <- colMeans(data_subset, na.rm = TRUE)
data_subset_kaväh <- sweep(data_subset, 2, means, FUN = "-")
kov <- solve(kov_mat)
lat <- faktorianalyyslop$loadings
data_subset_väh_mat <- as.matrix(data_subset_kaväh)
faktoripisteet_5 <- data_subset_väh_mat %**%kov%**%lat # faktoripisteet
faktoripisteet_5_thurst <- factor.scores(data_subset,faktorianalyyslop,
method="Thurstone")$scores

faktoripisteet_5 <- as.data.frame(faktoripisteet_5)
colnames(faktoripisteet_5) <- paste0("FactorScore_", 1:ncol(faktoripisteet_5))

# indeksit ja koulutusalat
faktoripisteet_5$Person_ID <- datak$Index
faktoripisteet_5$StudyField <- datak$Koulutusala
numeric_columns <- sapply(faktoripisteet_5, is.numeric)

# faktoripisteiden sarakenimet
factor_names <- colnames(faktoripisteet_5)[grepl("FactorScore_",
colnames(faktoripisteet_5))]

# ANOVA: erot koulutusalojen välillä
anova_results <- lapply(factor_names, function(f) {
  aov_formula <- as.formula(paste(f, "~ StudyField"))
  summary(aov(aov_formula, data = faktoripisteet_5))
})

# TukeyHSD -testit
faktoripisteet_5$StudyField <- as.factor(faktoripisteet_5$StudyField)
tukey_results2 <- TukeyHSD(aov(FactorScore_2 ~ StudyField,
data = faktoripisteet_5))
tukey_results3 <- TukeyHSD(aov(FactorScore_3 ~ StudyField,

```

```
data = faktoripisteet_5))
tukey_results4 <- TukeyHSD(aov(FactorScore_4 ~ StudyField,
data = faktoripisteet_5))

# MANOVA + Wilks
manova_formula <- as.formula(paste("cbind(", paste(factor_names,
collapse = ", "), ") ~ StudyField"))
manova_result <- manova(manova_formula,
data = as.data.frame(faktoripisteet_5))
summary(manova_result, test = "Wilks")
```

## Liite B: Lataukset

Loadings:

	PA1	PA3	PA2	PA4	PA5
v1		0.640			-0.181
v2		0.747		0.177	-0.168
v3	0.278	0.448			-0.156
v4	0.192	0.745	-0.219	0.222	-0.140
v5	0.173	0.269	-0.417	0.493	0.123
v6		0.784			-0.110
v7		0.532		0.186	
v8		0.911			
v9	-0.208	0.881	0.114		
v10		0.950	0.104	-0.170	
v11		0.847	0.247	-0.265	
v12	-0.213	0.590		0.265	0.281
v13	-0.114	0.511	0.112	0.211	0.274
v14	-0.118	0.437	0.231	0.153	0.235
v15		0.219	0.658	0.229	
v16		0.187	0.642	0.377	0.182
v17			0.771	0.636	
v18		0.185	0.550	0.149	
v19			0.890	0.315	-0.152
v20			0.887	0.128	
v21	0.193		1.005	-0.178	-0.128
v22			0.623		0.761
v23	0.112	0.171	0.721	0.281	-0.543
v24			0.216	1.396	
v25		0.119	0.245	0.219	0.191
v26		0.138	0.378		-0.379
v27	0.220	-0.166	0.231		
v28	-0.142			0.270	0.946
v29		-0.271	0.182		0.516
v30		0.302	0.256		0.174

v31	0.120	0.195	0.498	-0.116	0.335
v32	0.130	0.198	0.864	-0.412	0.296
v33		0.200	0.559		
v34			0.864	0.128	0.171
v35		0.238	0.375		
v36	0.108		0.468	0.553	0.140
v37		0.763			
v38	0.395	0.684		-0.193	
v39	0.288	0.470	-0.116		0.180
v40	0.486	0.270			
v41	0.558	0.373	-0.237		0.196
v42	0.889			0.152	
v43	0.944			0.178	
v44	0.722			0.283	-0.104
v45	0.700	0.111		0.103	
v46	0.974		0.154		-0.121
v47	0.728				
v48	0.846		0.385	-0.241	
v49	0.532				0.772
v50	0.522		0.253	0.300	-0.545
v51	0.269		-0.255	1.131	
v52	0.452	0.221	-0.230	0.172	0.236
v53	0.395				
v54	0.226			0.112	
v55	0.399		-0.222	0.158	0.671
v56	0.141	-0.194		0.143	0.434
v57	0.376	0.302			0.210
v58	0.616	0.183		-0.183	0.342
v59	0.623		0.250	-0.228	0.407
v60	0.777	0.110		-0.108	
v61	0.784				0.144
v62	0.565	0.196			
v63	0.489			0.446	0.118