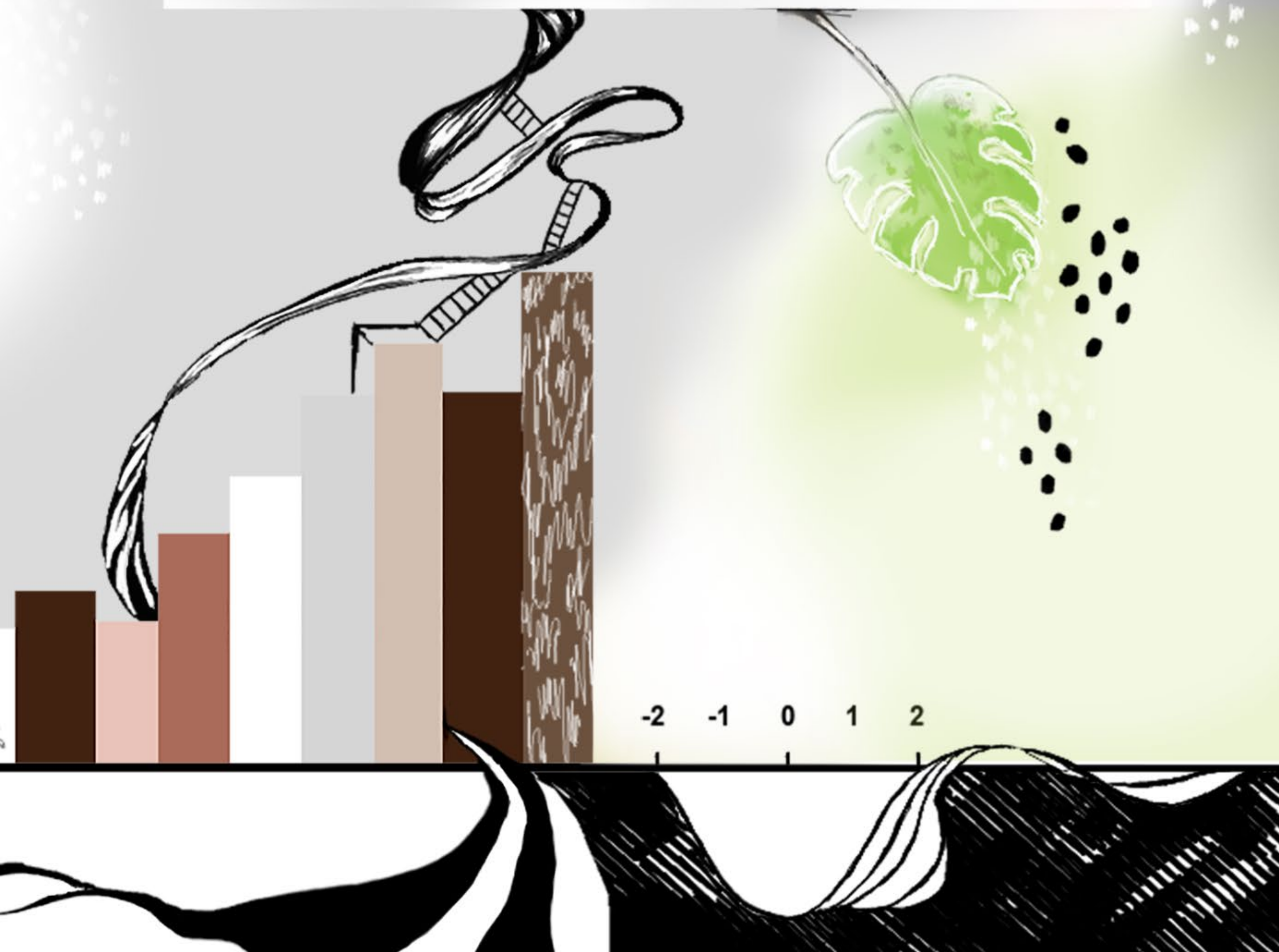
Tilastollisen aineiston käsittelyn ja tulkinnan perusteita

Juhani Tähtinen, Eero Laakkonen & Mari Broberg



TURUN YLIOPISTON KASVATUSTIETEIDEN TIEDEKUNNAN JULKAISUSARJA C,
OPPIMATERIAALIT 22

TILASTOLLISEN AINEISTON KÄSITTELYN JA TULKINNAN PERUSTEITA

Juhani Tähtinen,
Eero Laakkonen
& Mari Broberg

Turku
2020

prof. Arto Jauhiainen (puheenjohtaja)
yliopistotutkija Tuire Palonen
erikoistutkija Juhani Tähtinen

Turun yliopiston kasvatustieteiden
tiedekunnan julkaisuja C: 22,
2. uudistettu painos

Julkaisija:	Turun yliopiston kasvatustieteiden laitos
Kannet ja piirrokset:	Roosa Tähtinen
Taitto ja painatus:	Painosalama Oy
Painovuosi:	2020
ISBN:	978-951-29-8090-1 (PRINT)
	978-951-29-8091-8 (PDF)
ISSN:	1237-7643

Esipuhe

Tämä julkaisu ilmestyi ensimmäisen kerran jo vuonna 1993. Sen jälkeen siitä on suuren kysynnän vuoksi tehty erilaisia versioita, joissa kirjoittajakunta on osittain vaihtunut. Nyt käsissä oleva julkaisu on täydennetty ja uudistettu painos vuoden 2011 teoksesta. Teoksen perusajatus ja myös läpikäytävät analyysimenetelmät ovat samat kuin aiemmin. Olemme kuitenkin liittäneet näihin uusia tulkintaelementtejä, kuten efektikoon ja luottamusvälin tarkastelun, joiden avulla tilastollisesta päättelystä saadaan aiempaa luotettavampaa ja uskottavampaa. Lisäksi olemme muutenkin pyrkineet selkeyttämään ja syventämään eri analyysien toteutukseen ja tulkintaan liittyviä esityksiä. Version täysin uudet tekstiosat liittyvät ennen kaikkea määrällisen tutkimuksen kokonaisprosessin, aineiston analyysien tulkinnan ja tutkimuseettisten kysymysten aiempaa laajempaan ja monipuolisempaan tarkasteluun. Teoksen analyysiesimerkeissä käytetään SPSS-tilasto-ohjelman uusinta versiota: IBM SPSS Statistics 26. Eri analyysimenetelmien perusteiden tarkastelun yhteydessä annetaan konkreettiset ohjeet analyysien toteuttamiseksi edellä mainitulla ohjelmalla. Yleisesti nämä ohjeet soveltuvat ohjelman vanhempiinkin versioihin. Kirjan esimerkkianalyysien perusasiat ja niiden tulosten tulkinnat eivät luonnollisesti ole riippuvaisia käytettävästä tilasto-ohjelmasta. Kirjan verkkosivuilta (https://www.utu.fi/tilastoanalyysiopas) voit kopioida kirjassa käytetyn havaintomatriisin koneellesi. Näin voit harjoitella esitettyjä menetelmiä käytännössä, jolloin asioiden sisäistäminen helpottuu merkittävästi. Sivulta löydät myös erilaista aiheeseen liittyvää lisämateriaalia.

Sen lisäksi, että olemme kokeneet teoksen uudistamisen ja täydentämisen tarpeelliseksi, kirjan päivittämiseen meitä kannusti ajatus julkaista teos perinteisen paperiversion lisäksi ajan trendien mukaisesti verkkoversiona. Verkkoversion etu painettuun kirjaan nähden on siinä, että teoksen saa käsiinsä helposti aina kun sitä tarvitsee. Tämä ei kuitenkaan mielestämme vähennä painetun teoksen arvoa; monesti uutta asiaa on luontevampi ottaa haltuun fyysinen kirja käsissään. Näin varsinkin, jos opittava aineisto sisältää paljon muistettavaa ja erilaisia käytännön ohjeita, kuten tämän teoksen aihepiiri sisältää. Vanhan teoksen julkaiseminen verkossa sellaisenaan ei kuitenkaan innostanut meitä, halusimme samalla päivittää teosta niin, että se palvelisi aikaisempaa paremmin empiiristä ja määrällistä kasvatustieteellistä ja yhteiskuntatieteellistä tutkimusta tekeviä kirjan käyttäjiä. Kaikkein tärkein tarve teoksen uudistamiseen ja tarkastamiseen tulikin juuri tästä näkökulmasta. Haluamme nyt tuoda aiempia teoksia vahvemmin esille sen, kuinka tärkeää empiirisessä tutkimuksessa on omasta aineistosta nousevien tilastollisten analyysien tulosten tulkinta tutkittavaa asiaa tai ilmiötä koskevan teoreettisen tietämyksen (tai aiempien tutkimustulosten ja tulkintojen) varassa. Tämä on alusta lähtien hyvä oppia ja sisäistää! Esimerkiksi niin sanotut ”uuden tilastotieteen” (The New Statistics) edustajat ovat tätä viime vuosina korostaneet. Muutenkin suuntauksen edustajat ovat nostaneet kriittisen tarkastelun kohteeksi perinteisen tilastollisten analyysien hieman mekaanisen, liian vahvasti kategorisoivan ja pinnallisen soveltamisperinteen. Tähän kritiikkiin liittyviä tulkintoja ja varoituksia sisältyi jo kirjamme aikaisempaan versioon, mutta olemme pyrkineet nyt niitä tässä uudistetussa teoksessamme yhä kirkastamaan ja vahvistamaan.

Uusista painotuksista huolimatta kirjan perusidea on edelleen sama kuin aiemmin. Pyrimme tarjoamaan erityisesti ensiaskeleitaan tilastollisten menetelmien maailmaan ottaville kasvatustieteiden ja yhteiskuntatieteiden alan opiskelijoille mahdollisimman käytännönläheisen tien tilastollisten menetelmien käytön perusteiden ja perusanalyysimenetelmien soveltamiseen empiirisessä tutkimuksessa. Toki teos sopii myös vähän pidemmälle edenneiden käsiin, kuten julkaisun aiemman version käyttäjiltä saatu palaute kertoo. Nykyään monet tilasto-ohjelmat ovat hyvinkin helppokäyttöisiä. Vasta-alkajatkin pääsevät suhteellisen nopeasti näiden käytöstä perille, kunhan seuraavat näihin liittyviä opastuksia. Tämän ei saa antaa kuitenkaan hämätä itseään: toteutettiin tutkimus millä menetelmillä tahansa, edellyttävät ne tutkijalta tietoista suhdetta tekemiseensä, kuten jo edellä korostimme. Kirjamme pyrkii tukemaan lukijaansa tässä, mutta se ei ole kaiken kattava tilastollisten analyysien tulkintaan johdattava teos, vaan enemmänkin portteja tähän maailmaan avaava. Teoksemme rinnalla tarvitset varmasti muita alan kirjoja tueksesi. Olemme lisänneet tähän versioon myös aiempaa enemmän myös lähdeviittauksia alan oppikirjoihin ja muihin teksteihin, jotta voit niihinkin tutustumalla päästä eteenpäin. Myös käytetyn tilasto-ohjelman laajoja käsikirjoja sekä ohjelman Help-toimintoa kannattaa hyödyntää kirjamme rinnalla. Uskomme siihen, että jokaisella on mahdollista kehittyä ajan myötä hyvinkin taitavaksi tilastollisten analyysien soveltamisessa, kunhan tämän eteen jaksaa tehdä työtä.

Kirjassa pyritään esittämään tilastolliset perusmenetelmät mahdollisimman selkeästi ja pelkistetysti. Teoksen luonteen vuoksi tekstissä esiintyy väistämättä jonkin verran toistoa, koska ajattelemme osan lukijoistamme käyttävän kirjaa myös hakuteosmaisesti lukien tekstistä itselleen juuri sillä hetkellä ajankohtaisen osuuden. Tällaisilta lukijoilta jäisi ilman näitä uudelleen nostoja helposti huomaamatta tärkeitäkin eri analyyseihin ja niiden tulosten tulkintaan liittyviä tekijöitä, rajoituksia ja mahdollisuuksia. Tämä ei tietysti ole toivottavaa, joten olemme katsoneet tarpeelliseksi toistaa eräitä keskeisiä asioita useammassakin kohdassa kirjaa, vaikka tällainen asioiden toistelu voi häiritä lukijoita, jotka lukevat teoksen kannesta kanteen. Teos on kuitenkin tehty siinä toivossa, että tästä on hyötyä ja apua määrällisen ja tilastoanalyyseihin perustuvan tutkimuksen tekoon.

Olemme saaneet monilta tutkijakollegoiltamme ja opiskelijoiltamme vuosien varrella paljon vinkkejä siitä, miten teosta voisi työstää eteenpäin. Olemme pyrkineet ottamaan niitä mahdollisimman paljon huomioon teosta jälleen kerran työstäessämme. Kiitos Teille kaikille niistä! Erityisesti haluamme kiittää tässä yhteydessä tutkijakollegoitamme Laura Hellettä, Risto Ikosta, Kristiina Ojalaa, Janne Lepolaa, Veli-Matti Ritakalliota ja Jenni Tikkasta, jotka ovat kaikki vaivojaan säästämättä kommentoineet teoksen tämänkertaista viimeistelyvaiheessa ollutta versiota. Saimme heiltä monia tärkeitä kommentteja ja huomioita teoksen viimeistelyyn. Näiden avulla saimme hiottua ja terävöitettyä monia esityksiämme ja tulkintojamme samalla kun teoksen luettavuus parani huomattavasti.

Turussa huhtikuun 15. päivänä 2020
Juhani Tähtinen, Eero Laakkonen ja Mari Broberg

Sisältö

Tietolaarit

Asetelmat

1. Määrällisen tutkimuksen perusteita ja lähtökohtia

1.1 Peruslähtökohtia

Tiede ja tutkimus ovat yksi ihmisten keskeisimmistä tavoista saada tietoa ja lisätä ymmärrystä luonnosta, ihmisestä ja yhteiskunnasta. Tieteelliset tutkimukset tarjoavat meille myös yhden välineen elämämme ja maailmamme kehittämiseen kestävällä tavalla. Tieteellisen tiedon ja ajattelun vahvuus muihin tiedonmuotoihin nähden perustuu tähän liittyvään systemaattisempaan aineiston keruuseen ja tulkitsemiseen. Lisäksi tieteellisen tiedon ja tutkimuksen ideaalit ja lähtökohdat, kuten itsekorjautuvuus, avoimuus, kriittisyys, teoreettisuus, objektiivisuus ja toistettavuus, mahdollistavat omalta osaltaan olemassa olevan tiedon ja tulkintojen kriittisen tarkastelun ja uudistumisen. (Kerlinger 1981, 3–5; Metsämuuronen 2009, 33.) Edellä kuvatut tieteen ominaisuudet ovat yleviä, ja ne jäävät käytännön tutkimuksen tekijälle kaukaisiksi ideaaleiksi. Kuitenkin ne ilmaisevat kiteytetysti tieteellisen toiminnan erityispiirteet, joihin koko tieteenteon mieli perustuu. Niiden voi katsoa sisältyvän tieteen "perussääntöihin", jotka määrittelevät hyväksyttävän tutkimusprosessin eri puolet aineiston keruusta, analysoinnista ja tulkinnasta aina tulosten raportointiin saakka (ks. esim. Kakkuri-Knuutila 1992, 5–7; 2006; Hakala 2017). Niiden perussääntöjen tuntemus on siis hyvälle tutkimukselle välttämätöntä, joskaan ei tutkijoiden luovuuden ja kriittisyyden merkitystäkään tutkimuksen kulun onnistumiselle ja varmistumiselle voi vähätellä (Varto 2011; Ketokivi 2015, 7). Kaiken kaikkiaan tieteellisen tutkimuksen voi määritellä tältä pohjalta erilaisten ilmiöiden ja tekijöiden välisten suhteiden systemaattiseksi, kontrolloiduksi, empiiriseksi ja kriittiseksi tarkasteluksi ja tulkitsemiseksi.

Tilastollisia menetelmiä hyödyntävässä tutkimuksessa ollaan erityisesti kiinnostuneita eri ryhmien ja tekijöiden (muuttujien) välisistä yhteyksistä ja riippuvuuksista, niihin liittyvistä mekanismeista tai yleisesti vain eri ilmiöiden esiintymisestä ja niihin liittyvistä eri tekijöistä. Lähestymistavalle ominaista on aineistopohjaisuus tai muuttujakeskeisyys (variable-centred analysis; ks. enemmän esim. Kerlinger 1981, 8–11; Patton 1999, 37–42; Töttö 2004; Cohen & Manion 2007; Bhattacherjee 2012). Tilastolliset menetelmät antavat siihen tutkijoille hyviä työkaluja. Menetelmien soveltamiseen ja analyysien tulkintaan liittyy erilaisia perussääntöjä ja kriteereitä, jotka perustuvat ennen kaikkea tilastotieteilijöiden mallinnuksiin, todennäköisyyslaskentaan ja muihin matemaattisiin tulkintoihin. Osa näistä sovelluksista ja niihin liittyvistä "säännöistä" ovat hyvinkin vanhoja. Näiden noudattamisen tärkeyttä on alan eri oppikirjoissa perinteisesti korostettu, voisiko sanoa, suhteellisen kaavamaisesti ja pinnallisestikin. Tämä on ehkä yksi syy siihen, että näiden menetelmien käyttö ja niiden tulosten tulkinta on muodostunut jossakin määrin kaavamaiseksi ja pinnalliseksi. Tästä tilastollisiin menetelmiin perustuvaa empiiristä tutkimusta on ainakin usein syytetty. Viime aikoina tilastotieteilijöidenkin piirissä on yhä laajemmin alettu esittää kriittisiä kannanottoja perinteisten menetelmien lähtökohtia ja niiden mekaanista soveltamista kohtaan. On esimerkiksi väitetty perinteisten nollahypoteesin testaukseen (NHST) perustuvien lähtökohtien ja päättelyperinteen ohjanneen menetelmien käyttöä ja saatujen aineistojen tulkintoja hieman väärille urille. Samoin kritiikin kohteena on ollut yksioikoinen ja -puolinen analyysien tulosten tulkintatraditio ja raportointi: kaikki tilastollisesti merkitsevät löydökset eivät suinkaan ole teoreettisesti eivätkä myöskään reaalimaailman hahmottamisen kannalta merkittäviä. Näkyvimmin tätä kritiikkiä ovat esittäneet niin sanotun "uuden tilastotieteen" (The New Statistics) edustajat. Heidän tekstinsä ja tulkintansa ovatkin olleet yhä enemmän esillä muun muassa alan menetelmäkirjallisuudessa ja keskusteluissa.

Sinänsä "uuden tilastotieteen" korostamat lähtökohdat ja analyysimenetelmät eivät ole uusia, eikä toki hypoteesin testauskaan ole mitenkään pois suljettu analysointitapa. Analysoinnin ja tulosten tulkinnan pitäisi olla vain tarpeeksi monipuolista, jotta tutkimustulos kyettäisiin raportoimaan riittävän luotettavasti. "Uuden tilastotieteen" markkinoimassa lähestymistavassa on nostettu esiin erityisesti kolme periaatetta: *1) estimoinnin, 2) meta-analyysin (meta-analysis)* ja *3) tieteen avoimuuden (open science)* korostaminen. *Estimoinnilla* tarkoitetaan sitä, että tilastollisen analyysin perusteella pyritään otoksen analyysituloksesta tekemään mahdollisimman luotettava tulkinta tutkimuksen populaatiosta keskiarvo- tai prosenttisosuustulkinnan avulla. Tässä tulkinnassa esitetään piste-estimaattina esimerkiksi otoksen tuottama keskiarvo, mutta tämän lisäksi muodostetaan myös estimoinnin tarkkuuteen liittyvä virhemarginaali (margin of error) ja sen avulla muodostettava väliestimaatti eli luottamusväli. Virhemarginaalin avulla saadaan käsitystä estimoinnin tarkkuudesta, ja luottamusväli antaa vaihteluvälin, jolle tietyllä luotettavuudella (esim. 95 %) todellinen populaation tuntematon parametri sijoittuu. Tällaisen tulkintatavan ajatellaan olevan havainnollisempi, selkeämpi ja myös moniulotteisempi kuin perinteisen, dikotomiseen nollahypoteesi- ja tilastollinen merkitsevyys -perinteeseen nojaavat tulkinnat. *Meta-analyysilla* tarkoitetaan uudessa tilastotieteessä sitä, että omia tutkimustuloksia verrataan omaa tutkimusta vastaavien muiden tutkimusten tuloksiin. Tällä tavalla katsotaan saatavan tulosten tulkintaan enemmän evidenssiä kuin ainoastaan oman aineiston pohjalta tehtyyn tulkintaan. Kolmannella, *open science* -lähtökohdalla, tarkoitetaan uudessa tilastotieteessä sitä, että koko tutkimusprosessia leimaa avoimuus, aina tutkimuksen tilastollisten analyysien tarkasta suunnittelemisesta ja ennakoimisesta alkaen oman datan avoimeen koodaamiseen ja saatavuuteen sekä toisten suorittamiin varmistusanalyyseihin ja toistettavuuteen saakka. Pyrkimyksenä on siis se, että kaikki tutkimus- ja tulkintavaiheet toteutetaan avoimuutta korostaen. Kaikki mahdolliset virhetekijät, jotka vähentävät tulosten luotettavuutta, pyritään minimoimaan. (Cumming & Calin-Jageman 2017, xvii–xviii, 9–12; Kline 2013; Ketokivi 2015, 42–49.)

Miten tämä sitten eroaa perinteisestä ”vanhasta tilastotieteestä” (The Old Statistics). Perinteisessä tilastotieteessä on nojauduttu pitkälti tilastolliseen päättelyyn, joka perustuu nollahypoteesin testaamiseen ja tilastollisen merkitsevyyden, p-arvon, tulkitsemiseen (p-arvotulkinnasta enemmän luvussa 1.6.1). Uudessa tilastotieteessä korostetaan siis laaja-alaisempaa aineiston analyysia, jossa havaittuja jakaumia ja yhteyksiä tulkitaan myös esimerkiksi efektikoko- ja luottamusvälisuureiden välityksellä. Periaatteessa ”vanha” ja ”uusi” tilastotiede rakentuvat samalle perustalle ja usein myös niiden tulostulkinnat ja johtopäätökset ovat samoja. ”Uuden tilastotieteen” edustajien mukaan laajemmin estimointiin perustuva analyysi auttaa tutkijoita tekemään kuitenkin aineistostaan parempia, luotettavampia ja ilmiöiden moninaisuutta kunnioittavia tulkintoja. Pyrkimyksenä on ennen kaikkea välttää tulkinta-asetelmia, joissa hyvin jyrkästi päätellään esimerkiksi jonkin tilastollisen suureen, esimerkiksi tilastollisten merkitsevyysarvojen, perusteella tutkittavan ilmiön tai tekijöiden välisen suhteen olemassaoloa tai ei-olemassaoloa. Lähestymistavan edustajien mukaan uusi analysointitapa on myös ylipäätään helpompi tutkijoiden oppia ja suuren yleisön ymmärtää. (Cumming & Calin-Jageman 2017, xviii–xix.)

Määrällisen ja tilastollisen tutkimuksen tavoitteeksi voidaan yleisesti määritellä tutkittavan ilmiön kuvailu ja selittäminen, mutta myös ymmärtäminen, joka yleensä on liitetty laadulliseen tutkimukseen. Itse miellämme empiirisen yhteiskunnallisen ja kasvatustieteellisen tutkimuksen ennen kaikkea toiminnaksi, jonka päätavoitteena on tuottaa uutta tietoa tutkittavista ilmiöistä ja yleensäkin tehdä meitä ympäröivää maailmaa ymmärrettäväksi. Tähän liittyy läheisesti myös vallitsevan tietämyksen ja tulkintojen kriittinen tarkastelu ja arviointi. Tähän määrälliset survey-tutkimukset tilastollisine analyyseineen antavat omaltaan osaltaan hyvät mahdollisuudet, voidaanhan näiden avulla tarkastella muun muassa laaja-alaisesti eri tekijöiden yhteyksiä ja eroja (Metsämuuronen 2009, 35–36). Luonnontieteellisessä tutkimuksessa tavoitteena on puhtaimmillaan universaalien lainmukaisuuksien ja säännönmukaisuuksien löytäminen ja testaaminen. Tähän ei juuri päästä kasvatus-, käyttäytymis- ja yhteiskuntatieteellisissä tutkimuksissa. Toki esimerkiksi regressio- ja varianssianalyysiin liittyvä asetelma perustuu syy–seuraus -suhde malliin, joka rakentuu ajatukselle ilmiöiden taustalla vaikuttavista kausaalisista tekijävaikutuksista. Kuitenkin kvantitatiivisen, tilastollisiin analyyseihin nojautuvan tutkimuksen katsotaan yleisesti olevan pikemminkin korrelatiivista kuin kausaalisiin selitysmalleihin pyrkivää tulkintaa. Ihmisiä ja yhteiskuntaa koskevissa tutkimuksissa näin vahvat tulkinnat ovat mahdollisia harvoin, jos koskaan. Toisaalta syy-seuraus- eli kausaalisisia selitysmalleja tai kausaalisija mekanismitulkintoja ei kannata unohtaa kokonaan: onhan niin, että käyttäytymis- ja yhteiskuntatieteissä yksi perimmäisistä tavoitteista on saavuttaa mahdollisimman päteviä tulkintoja eri ilmiöiden taustalla olevista tekijöistä. Tämä tavoite, jos mikä, korostaa sitä, että määrällisissä tutkimuksissa aineiston tilastollisten analyysien tulokset voidaan tulkita tarkoituksenmukaisella tavalla vain tutkittavaa ilmiötä koskevaa tutkimustietoa ja teorioita hyödyntäen.

Olemme kirjan tässä laitoksessa pyrkineet nostamaan aiempaa painotetummin edellä mainituista syistä esiin aineiston tulkinnan ja myös tähän liittyvän teoreettisempaan tulkintaan merkitystä: esimerkiksi kirjan alkulukujen keskeisimmät lisäykset liittyvät juuri siihen, että määrällisessä käyttäytymis- ja yhteiskuntatieteellisessä tutkimuksessa tilastollinen päättely on vasta tulkinnan yksi perusvaihe, jota tulisi aina syventää suhteuttamalla tästä saadut havainnot jo olemassa olevaan tutkimustietoon ja teoreettisiin tulkintoihin. Vasta näin voidaan tulkita aineistosta nousevia tuloksia relevantisti tai arvioida esitetyn tulkintamallin tarkoituksenmukaisuutta.

Määrällisessä tutkimuksessa keskeisellä sijalla on yleisesti erilaisten tilastollisten analyysimenetelmien hyödyntäminen aineiston jäsentämiseksi tulkittavampaan muotoon ja tässä olevien tekijöiden riippuvuuksien tulkitsemiseksi (Fielding & Gilbert 2006, 8–9). Risto Ikonen (2016) on määritellyt osuvasti tutkimusmenetelmät keinovalikoimaksi, ”jolla informaation kokoaminen ja jäsentäminen sekä tutkimuslöydösten esittäminen pyritään toteuttamaan virhetulkinnat minimoivalla tavalla”. Tästä on pitkälti kyse myös niiden tilastollisten

menetelmien valitsemisesta, joita käytämme eri tutkimuksissamme. Tilastomenetelmiä käyttävän tutkimuksen ja aineiston analyysin onnistumisen yksi keskeisimpiä edellytyksiä on se, että tutkija ymmärtää sen, mitä on tekemässä. Tutkijan tulee myös ymmärtää ja hallita käyttämänsä analyysimenetelmät, jos ei täydellisesti, mahdollisimman hyvin kuitenkin. On tärkeää tietää se, millaisia mahdollisuuksia ja rajoituksia eri analyysimenetelmiin liittyy. Nykyiset tilasto-ohjelmat tarjoavat menetelmien suhteen runsaan kattauksen. Niiden peruskäyttö on tullut varsin helpoksi. Nykyään menetelmien käyttöön liittyvä vaikeus onkin enemmän siinä, mitä menetelmää kulloinkin tulisi käyttää ja kuinka saatuja analyyseja tulisi tulkita, kuin itse analyysin tekemisessä. Näihin kun ei ole yksiselitteisiä sääntöjä olemassa. Tällainen tilanne ahdistaa helposti varsinkin vasta-alkajaa. Tämä kirja pyrkii omalta osaltaan antamaan helpotusta tarjoamalla lukijalle osviittoja siihen, kuinka kulkea onnistuneesti määrällisen ja tilastomenetelmiä hyödyntävän tutkimuksen monipolvisessa ja moninaisesti risteilevässä ”menetelmien viidakossa”. Lähtökohtamme, jota haluamme korostaa, on se, ettei ole yhtä oikeaa tapaa tai menetelmää analysoida tutkittavana olevaa kysymystä. Vaihtoehtoja on yleensä useita. Tällainen asenne antaa tutkijalle suuremman liikkumavaran kuin tiukka sitoutuminen jonkin spesifin metodologisen koulukunnan tai menetelmien lähtökohtiin. Eri tilastolliset analyysit antavat tutkijalle hyvät työkalut määrällisen aineiston analyysiin ja käsittelyyn, kunhan hallitsee nämä riittävästi (Fielding & Gilbert 2006, 8–9). Tämä edellyttää ainakin jonkin verran tilastollisten menetelmien perusteiden ja käytön kriteerien hallintaa ja yleensä niin sanottua numerolukutaitoa, jota ilman aineistoamme koskevia analyysitulosteita on vaikea tulkita (Töttö 2012, 9–10). Toivomme, että kirjamme antaa tämänkin kehittymiseen omat eväänsä, vaikka lähestymme empiiriseen tutkimukseen liittyviä kysymyksiä enemmän tutkimuksen teon, erityisesti tähän keskeisesti liittyvän määrällisen aineiston tilastollisten analyysien suorittamisen ja näihin liittyvien tulosten tulkinnan, näkökulmista.

Ennen kuin mennään eteenpäin tutkimuksen teon lähtökohtien ja tutkimuksen eri vaiheiden tarkastelussa, on seuraavalle aukeamalle Tietolaari 1:een koottu tilastolliseen tutkimusaineistoon liittyviä keskeisiä käsitteitä (ks. lisää Tilastokeskus: Tietoa tilastoista, käsitteet). Kirjaan on koottu muitakin tällaisia tietoiskuiksi tarkoitettuja tietolaareja eri asioiden käsittelyn yhteyteen. Huomaa myös käyttää hyväksesi kirjan loppuun koottua asiahakemistoa etsiessäsi vastauksia erilaisiin menetelmällisiin kysymyksiin.

TIETOLAARI 1. KESKEISIÄ TILASTOAINEISTOON LIITTYVIÄ PERUSKÄSITTEITÄ

Populaatio eli perusjoukko

Tutkimuksen taustalla oleva kohdejoukko, esimerkiksi ihmisryhmä, johon tutkimus kohdistuu, esimerkiksi 7.–9. -luokkalaiset oppilaat Suomessa tai tietyn muun rajatun alueen tämän kriteerin täyttävät koululaiset. Perusjoukko voi olla myös hyvinkin rajattu joukko, vaikka jonkun koulun näillä luokilla olevat koululaiset. Tällöin tutkimuksen tulosten ajatellaan edustavan vain tätä tarkoin määriteltyä joukkoa. Yhtälailla populaation kriteerit voivat määrittyä esimerkiksi sukupuolen, äidinkielen tai sosiodemografisten taustatekijöiden mukaan. Tutkimusten perusjoukko määritellään tarkemmin näiden tutkimusasetelmissa.

Havaintoyksikkö eli tilastoyksikkö

Tällä viitataan tutkimuksessa, tutkimuksen otoksessa tai näytteessä, mukana olevaan yksittäiseen kohteeseen, eli esimerkiksi oppilaaseen, opettajaan tai kouluun.

Otos

Yleensä tutkimukseen ei voida ottaa mukaan tutkimuksen koko populaatiota (perusjoukkoa), vaan tilastoyksiköitä poimitaan mukaan tutkimusotokseksi vain tietty osuus. Tilastollisen päättelyn perusteella voidaan tästä otoksesta kuitenkin tehdä koko populaatiota koskevia päätelmiä tietyllä (yleisesti 5 %:n tai 1 %:n) riskitasolla. Otos on siis jonkin tietyn satunnaisuuteen perustuvan valintakriteerin perusteella muodostettu osajoukko populaatiosta (otoksesta puhutaankin silloin, kun jokaisella perusjoukkoon kuuluvalla on ollut yli nollaa suurempi todennäköisyys päätyä tutkimusaineistoon; muussa tapauksessa kyse on näytteestä). Otoksen ja populaation määrittely asianmukaisesti onkin otoksen edustettavuuden kanssa yksi tärkeimpiä ehtoja otoksesta perusjoukkoon tehtävien tulkintojen virheettömyydelle ja uskottavuudelle. Tilastollisen tutkimuksen yksi perustavoite onkin analysoida ja arvioida sitä, kuinka todennäköistä on se, että otoksessa havaittu ilmiö esiintyy myös koko populaatiossa. Otoksen tulisi olla ominaisuuksiltaan ja rakenteeltaan mahdollisimman samanlainen kuin ne ovat tutkimuksen populaatiossa, esimerkiksi sukupuoli- ja ikäjakauman suhteen (otoksen edustavuuskysymys). Otos on sitä edustavampi, mitä lähempänä se on populaation todellista koostumusta.

Otoksen muodostaminen on tekijä, joka osaltaan tuo tutkimukseen ja tutkimusaineistoon satunnaisvaihtelua, jonka vaikutusta tutkijan tulee pyrkiä minimoimaan huolellisella otannan suunnittelulla ja toteutuksella. Otoksen muodostamiseksi on erilaisia tilastollisia otantamenetelmiä, kuten yksinkertainen satunnaisotanta, systemaattinen, ositettu tai ryväsotanta. Se, mikä otosmenetelmä on tarkoituksenmukaisin, riippuu tutkimusasetelmasta ja -tehtävästä. Käytännössä, kuten Pahkinen (2012, 173) toteaa, tutkimuksen otannassa käytetään usein myös eri perusotantamenetelmistä muodostettua ”moniasteista otanta-asetelmaa”. Se, kuinka iso otos on tarpeen, riippuu sekin paljolti tutkimuksen tavoitteista ja tutkimusasetelmasta, mutta myös eri tilastollisten menetelmien tarkoituksenmukainen käyttö asettaa otokselle erilaisia vaatimuksia esimerkiksi otoskoon ja sisäisen jakauman suhteen. Jos emme pysty määrittelemään tutkimuksen populaatiota eli perusjoukkoa tarkkaan tai emme syystä tai toisesta käytä mitään tilastollista otantamenetelmää, käytetään otoksen sijaan käsitettä *näyte*. Kun viittaamme taulukoissa ja kuvioissa yms. koko otokseen, käytetään tämän tunnuksena N:ää, kun taas viitataan osaotokseen, käytetään pientä n:ää.

Näyte

Jos tutkimuksen havaintoyksiköiden valinta tapahtuu muulla tavoin kuin tilastollisella otantamenetelmällä, esimerkiksi harkinnanvaraisesti valikoiden, kutsutaan tällaista tutkimusjoukkoa näytteeksi.

Otoksen edustavuus ja aineistokato

Tilastollisiin menetelmiin perustuvissa tutkimuksissa keskeistä on tulkinta, joka tehdään otoksen perusteella perusjoukosta. Toinen keskeinen tulkinta liittyy tutkittavan ilmiön mallintamiseen. Jotta tulkinnat olisivat näiden suhteen mahdollisimman luotettavia, tulisi otoksessa eli tutkimuksen aineistossa olla riittävän hyvin edustettuna kaikki ihmisryhmät, jotka sisältyvät tutkimuksen perusjoukkoon. Viime vuosina erityisesti yhteiskuntatieteiden puolella on lisääntynyt tienlainen vastaajakato empiiristen tutkimusten aineiston keruun yhteydessä. Tähän on monta syytä. Näin jo tämä tekijä voi tuoda ongelmia otoksien edustavuuteen. Otos voi näin muodostua joltakin osin ongelmalliseksi. Tällöin kyseeseen tulevat yleensä esimerkiksi otoksen ikä-, sukupuoli-, ammatti-, koulutustaso- tai vaikkapa tulotason jakaumien poikkeamat perusjoukon jakaumasta. Tähän on syytä kiinnittää huomiota jo tutkimusasetelmaa, aineiston keruuta ja otoskokoa määriteltäessä. Ennakoimalla mahdollisia otoksen keruuseen liittyviä tulevia ongelmia voidaan tehdä jo paljon näiden välttämiseksi (esim. kyselylomakkeen huolellisella suunnittelulla, otoskokoa kasvattamalla tai tietyn vastaajaryhmän osuutta painottamalla jo aineiston keräysvaiheessa). Tilasto-ohjelmissa voidaan tähän liittyvät ongelmat ratkaista käyttämällä painokerrointa tasaamaan havaittua aliedustavuutta kyseisen muuttujan/tekijän kohdalla. (ks. esim. KvantiMOTV, havaintoaineiston painottaminen). SPSS-ohjelmasta löytyy toiminnot painokertoimien muodostamiseen ja aineiston painotuksen toteutukseen.

Muuttuja

Tutkimuksen havaintoaineisto voidaan kerätä esimerkiksi lomakekyselyllä. Tällöin lomakkeen eri kysymykset ja väittämät tallennetaan havaintomatriisiin omiksi yksiköikseen eli muuttujiksi. Muuttujat muodostavat havaintomatriisin sarakkeet ja niitä on siis yleisesti sama määrä kuin tutkimuslomakkeessa on kysymyksiä. Luonnollisesti tietyt lomakkeen kysymykset/väittämät voi joutua jakamaan aineiston syöttövaiheessa useihin eri muuttujiin, joten muuttujien määrä voi olla tietyissä tapauksissa suurempi kuin kyselylomakkeen kysymysten määrä. Muuttujien arvot vaihtelevat eri tutkittavien kohdalla. Muuttujilla on tutkimuksessa myös erilainen luonne eri analyyseissa, tämän mukaan niitä kutsutaan joko riippuviksi (dependents, selitettäviksi, vastemuuttujat) tai riippumattomiksi (independents, selittäviksi, tekijät) muuttujiksi. Kun olemme esimerkiksi kiinnostuneet siitä, kuinka oppilaiden viriketausta ja sukupuoli vaikuttavat tai heijastuvat heidän opintomenestykseensä lukiossa, on tässä tutkimusasetelmassa opintomenestys lukiossa selitettävä muuttuja (riippuva muuttuja) ja oppilaiden viriketausta ja sukupuoli ovat taas selittäviä (riippumattomia) muuttujia. Kvantitatiivisessa tutkimuksessa ja tilastollisessa analyysissa ollaan usein kiinnostuneita näiden muuttujien (tekijöiden) yhteyksistä tutkimuksesta.

Mittaaminen ja havainnot

Empiirinen tutkimus perustuu tehtyihin havaintoihin ja mittauksiin, joita tutkija tekee eri menetelmien avulla. Tilastollisessa tutkimuksessa kyse on ennen kaikkea tutkittavien ominaisuuksien keräämisestä ja määrittämisestä. Tutkimuksessa on olennaista, että tarvittava tiedonkeruu on systemaattista ja se kohdistuu tutkimusasetelmassa määriteltyihin eri tekijöihin (muuttujiin). Tämän vuoksi puhutaan tutkimuksen yhteydessä yleensä pikemminkin mittaamisesta kuin havainnoinnista. Havainnolla viitataan tässä yhteydessä tutkimuksessa mukana olevaan yksittäiseen ihmiseen tai muuhun tutkittavaan liittyvästä tiedonkeruusta. Mittausmenetelmiä voidaan käyttää useilla eri tavoilla, kuten havainnoimalla ihmisten käyttäytymistä ja toimintaa, haastattelemalla, käyttämällä kyselylomaketta tai käyttämällä erilaisia valmiita testejä.

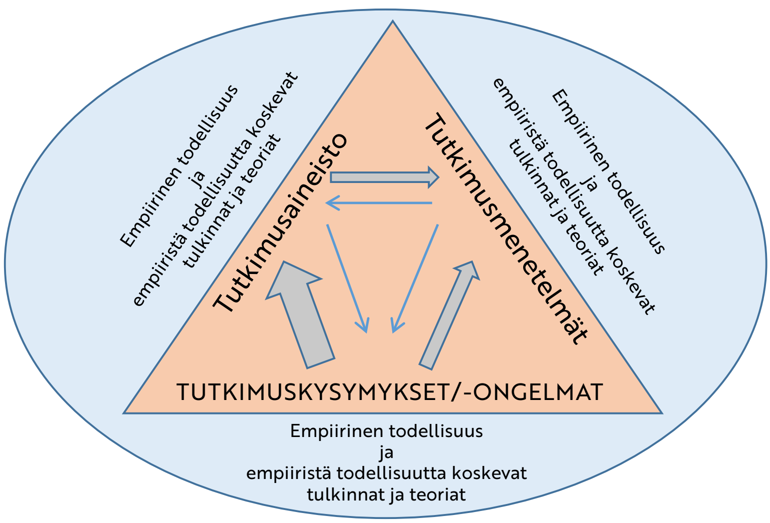
Estimointi/estimaattori/estimaatti

Estimointi on populaation tuntemattomien suureiden (parametrien) arviointia. Kyse on siis tilastollisesta päättelystä, jossa otoksessa havaittujen mallien ja jakaumien perusteella tehdään yleistyksiä tutkimuksen perusjoukossa esiintyvistä jakaumista ja vastaavista suureista. Estimaatti on tilastollisen mallin parametreille aineistossa laskettu arvo. Estimaatti saadaan muodostettua sopivaa estimaattoria käyttäen. Estimaattori voidaan ymmärtää välineenä, jolla estimointia suoritetaan. Käytännössä se on tilastotieteessä kehitetty laskutapa tai laskukaava, jolla estimointia voidaan toteuttaa. Estimaattorin tärkeitä ominaisuuksia ovat harhattomuus, tarkentuvuus ja tehokkuus. Silloin sillä on mahdollista tuottaa laadukkaita arvioita. Estimointi voi olla piste-estimointia (yksittäinen otossuure) tai väliestimointia (luottamusväli, joka huomioi myös estimoinnin tarkkuutta tietyllä luotettavuudella). Tilastollinen päättely ei ole koskaan luotettavaa, jos estimaatit eivät ole luotettavia. Tilastollisten analyysien menetelmien käyttö perustuu kokonaan tälle ajatukselle: päämääränä on estimoida tutkimuksen kannalta tärkeimmät parametrit mahdollisimman luotettavasti, jos näin ei ole, tutkimuksen empiirinen osa ja siitä tehtävät tulkinnat ovat heikolla pohjalla ja vaarassa menettää uskottavuutensa. (Ketokivi 2015, 22–23, 288.)

1.2 Tutkimusmenetelmien valintaan vaikuttavat tekijät

On sitten kyse empiirisen tutkimuksen aineiston tai tutkimusmenetelmien valitsemisesta, tutkimukselle asetetut tutkimuskysymykset määrittävät niiden valinnan perusteet. Tutkimusmenetelmien valintaan vaikuttaa lisäksi esimerkiksi kerätyn tutkimusaineiston luonne ja otoksen suuruus. Omalta osaltaan näihin kaikkiin, myös tutkimuskysymyksiin, vaikuttaa se, mistä perspektiivistä tai teoreettisesta näkökulmasta tutkimuskohdetta tarkastellaan. (Ks. esim. Jokivuori & Hietala 2007, 10–11, 23–24.) Kaikki tutkimuksen teon peruselementit vaikuttavat enemmän tai vähemmän suoraan toisiinsa (ks. kuvio 1).

Kuvion 1 asetelma tähdentää sitä, kuinka tutkimuskysymykset ovat hyvin määräävässä asemassa valittaessa tutkimusaineistoa ja -menetelmää empiirisessä tutkimuksessa. Kuitenkin on tärkeää huomata myös se, että vaikutussuhde ei näiden tekijöiden välillä ole yksisuuntainen: esimerkiksi se, millaista tutkimusaineistoa on mahdollista kerätä, voi joissakin tapauksissa määrittää hyvinkin paljon sitä, millaisia tutkimuskysymyksiä tutkimukselle voidaan asettaa. Tällainen tilanne voi tulla vastaan, jos esimerkiksi suunniteltu tutkimusaihe on siinä määrin arkaluontoinen, ettei siitä saada kerättyä alkuperäisen suunnitelman mukaista ainestoa. Joskus myös tutkimuksen aikataulu voi määrittää sen, millainen tutkimusaineisto on mahdollista kerätä. Tämä ei luonnollisestikaan ole toivottava tilanne, mutta usein tällainen tilanne voi syntyä vaikka opinnäytetöitä tehtäessä. Joskus myös ajatellut tutkimusme-



KUVIO 1. Empiirisen tutkimuksen tutkimuskysymyksiä, -aineistoa ja -menetelmiä määrittäviä tekijöitä.

netelmät voivat määrittää sen, millaisia tutkimuskysymyksiä tutkimukselle on järkevää asettaa tai millainen aineisto tutkimukseen on kerättävä, vaikka yleisesti tämä vaikutussuhde on luonnollisesti toisinpäin. Lisäksi kuviosta ilmenee se, että näiden kolmen tekijän suhteen lisäksi empiirinen todellisuus ja tätä koskevat tulkinnat ja teoriat määrittelevät omalta osaltaan sekä tutkimuskysymysten asettelua kuin kullekin tutkimukselle relevantin tutkimusaineiston ja-menetelmien valintaa. Lisäksi näihin valintoihin vaikuttavat tutkimuksen taustalla olevat yleisemmät metodologiset perustulkinnat, jotka antavat tutkimuksessa tehtäville valinnoille perusteet ja yleisemminkin yhtenäisen tarkastelu- ja tulkintakulman tutkimukselle (Varto 2011).

1.3 Tutkimuksen kulun pääpiirteet

Tutkimuksenteon etenemisen voi hahmottaa parhaiten vaiheittaisen tarkastelun avulla, vaikka käytännössä nämä vaiheet limittyvät yleensä toisiinsa, joskus jopa niin, että tutkimuksen teon jossakin vaiheessa pitää palata aivan alkuvaiheen suunnitelmien uudelleen pohtimiseen. Yleisesti tutkimuksen kulun vaiheet voidaan esittää seuraavan asetelman mukaisesti:

1. TUTKIMUSASETELMAN/-KYSYMYSTEN JA -ONGELMAN MÄÄRITTÄMINEN
   - tutkimusaiheen ja -ongelman määrittäminen
   - aihepiiriin tutustuminen (kirjallisuus, teoriat, keskeiset käsitteet)

2. SUUNNITTELUVAIHE
   - tutkimusongelmien yksityiskohtaisempi määrittäminen (pää- ja alaongelmat yms.)
   - lähtöteorian tai yleisemmin tutkimuksen tarkastelukulman määrittely ja valinta ja teoreettisten väittämien laatiminen (esim. hypoteesien asettaminen)
   - käsitteiden, muuttujien, käytettävien indikaattoreiden ja yksiköiden määrittäminen
   - näihin liittyvien mittarien ja kysymysten laatiminen ja valinta
   - tutkimuskysymysten operationaalistaminen
   - tutkimusmenetelmien valinta
   - aineiston keruun määrittäminen (esim. otoksen ja keruutavan määrittäminen)
   - aineiston analysointitavan tarkempi suunnittelu
   - raportoinnin suunnittelu

3. KENTTÄTYÖVAIHE
   - varsinaisen aineiston keräysvaihe (esim. kyselylomakkeella tai haastattelemalla)

4. AINEISTON ESIKÄSITTELY
   - aineiston koodaus ja siirto tilasto-ohjelmaan (aineiston syöttö, havaintomatriisin teko)

5. AINEISTON ANALYSOINTI
   - varsinaisten tilastoanalyysien suorittaminen
   - johtopäätösten teko

6. TUTKIMUKSEN RAPORTOINTI

ASETELMA 1. Empiirisen tutkimuksen kulkukaavio (mukailen Hirsjärvi ym. 2009, 65; Bhattacherjee 2012, 20).

Tutkimusprosessin eri vaiheita on tarkasteltu monissa tutkimuksen tekemistä käsittelevissä teoksissa (ks. esim. Metsämuuronen 2009, 37–87; Nummenmaa 2011, 22–37; Bhattacherjee 2012, 17–23; Rose 1991; Field 2016, 14–21), joten tässä yhteydessä tätä ei tarvitse käydä lävitse yksityiskohtaisesti. Haluamme kuitenkin korostaa sitä, että paneutuminen omaan tutkimusaiheeseen ja huolellinen tutkimuksen suunnittelu on keskeistä tutkimusprosessin etenemiselle ja sen luotettavuudelle. Sitä ei voi liiaksi korostaa, kuinka tärkeätä on se, että tutkija on perehtynyt tutkimaansa ilmiöön ja alan uusimpaan tutkimukseen. Ilman tätä esimerkiksi tarkoituksenmukaisen ja kestävän tutkimusasetelman ja tutkimusongelmien määrittely on kovin heikolla pohjalla. Tästä aiheutuvia ongelmia voi olla vaikea korjata myöhemmin. Esimerkiksi kyselylomakkeisiin perustuvissa tutkimuksissa aiheeseen perehtymättömyys ja suunnittelemattomuus voivat aiheuttaa suuria hankaluuksia aineiston analyysi- ja tulkintavaiheessa. Tähän tutkimusprosessin vaiheeseen kannattaa siis panostaa. Huolellista suunnittelua edellyttää luonnollisesti myös empiirisen aineiston keruu ja analyysin toteuttaminen. Tutkimuksen perusasetelman, lähtökohtien, käytettyjen menetelmien ja tutkimusaineiston tulee nivoutua yhteen mahdollisimman hyvin.

Tutkimusprosessi lähtee liikkeelle tutkimuksen yleisen tavoitteen ja tarkastelukulman sekä tutkimuskysymysten tai -ongelmien määrittelystä. Tämä on yksi tutkimuksen tärkeimmistä vaiheista, kuten jo edellä todettiin. Tutkimuksen tarkastelukulman ja tavoitteen määrittelyyn onkin syytä varata riittävästi aikaa: jotta osaa asettaa tutkimukselleen mielekkään yleistavoitteen ja kysyä tutkimusaineistoltaan oikeita ja tarkoituksenmukaisia kysymyksiä, on tunnettava riittävästi tutkittavaan ilmiöön tai kysymykseen liittyvää aiempaa tutkimusta, tähän liittyviä tulkintoja ja teorioita. Tätä vaihetta ei kannattaisi laiminlyödä, kuten toisinaan kiireen tai muiden syiden vuoksi esimerkiksi pro gradu -tutkielmissa näytetään usein tehtävän. Jos tämän vaiheen ohittaa liian heppoisesti, tämä kostautuu usein myöhemmin. Tämä ei tarkoita sitä, etteikö tutkimusasetelmaa ja tutkimuskysymyksiä voisi tutkimusprosessin edetessä tarkentaa tai jopa muuttaa radikaalimminkin tarvittaessa. Näin usein käytännössä tapahtuukin. Kuitenkin, mitä paremmin on selvittänyt itselleen oman tutkimuksensa tarkastelukulman ja tutustunut tutkittavana olevaan kysymykseen liittyvään tutkimuskirjallisuuteen, sitä paremmat mahdollisuudet tutkimuksenteon onnistumiselle on. Tältä pohjalta tutkija määrittelee tutkimuk-

selleen spesifimmät tavoitteet, tutkimuskysymykset ja tutkimuksen kohteena olevan populaation ja sen, minkälainen otos on tarkoituksenmukainen kyseisessä tutkimuksessa. Mitä paremmin siis suunnittelet tutkimusprosessin, sitä todennäköisempää on, että tutkimusprosessi etenee suhteellisen kivuttomasti. Ainahan eteen tulee toki kaikenlaisia ongelmia, mutta niistä selviää yleensä paremmin, mitä paremmin on ottanut haltuunsa tutkimuksen aihepiirin ja oman tutkimuksen tarkastelukulman.

Yksi suunnitteluvaiheessa ratkaistava keskeinen kysymys liittyy tutkimusaineiston määrittelyyn. Voimme kerätä harvoin tietoa kaikilta tutkimuksen populaatioon kuuluvilta, joten tässä vaiheessa tulee myös pohdintaan se, millainen otos tutkimuksen perusjoukosta käsillä olevaan tutkimukseen tarvitaan (ks. populaation ja otoksen käsitteistä Tietolaarista 1, s. 16). Kun otos on mahdollisimman edustava ja kattava otoksen perusjoukosta, katsotaan tältä pohjalta voitavan tehdä tietyllä riskitasolla päätelmiä tutkimuksen perusjoukosta. (Helenius 1989, 274–276; Gröönroos 2011, 8, Heinonen 1989, 105–106.)

Otokseen liittyvät päätökset vaikuttavat monella tapaa tutkimuksen tulosten tulkinnan luotettavuuteen. Kaikkiin otantatutkimuksiin liittyy niin sanottu otantavirheongelma. Tällä viitataan siihen, että tutkimuksen otos ei ole riittävän edustava todellisen perusjoukon ja tutkimuksen kysymysten suhteen. Otoksesta päätettäessä tuleekin ottaa huomioon sekä tutkimuksen perusjoukko, tutkimuskysymykset että suunnitellut tutkimusmenetelmät. Se, että otos on valittu tarkoituksenmukaisella tavalla, on hyvin keskeinen tekijä tutkimuksen tulosten yleistettävyyden, ja myös tulosten oikeellisuuden, kannalta. On tärkeää, että tutkimusaineisto sisältää sellaisen otoksen, joka edustaa tutkimuksen perusjoukkoa mahdollisimman kattavasti. Jos otos on epäedustava, saadut tulokset eivät edusta riittävästi perusjoukkoa, jolloin tehdään helposti virhepäätelmiä. Ongelmia tutkimukseen tuo usein myös se, ettei kaikilta otokseen valikoiduilta saada syystä tai toisesta kerättyä tutkimusdataa. Jos olemme esimerkiksi erityisen kiinnostuneista jonkin ryhmän alaryhmästä, emmekä saa tähän ryhmään kuuluvilta riittävästi vastauksia, voi tämä vaikuttaa otoksen edustettavuuteen merkittävästi (tämän tuomaa ongelmaa voidaan korjata painokertoimia käyttäen, mutta tähän liittyy puolestaan omat ongelmansa).

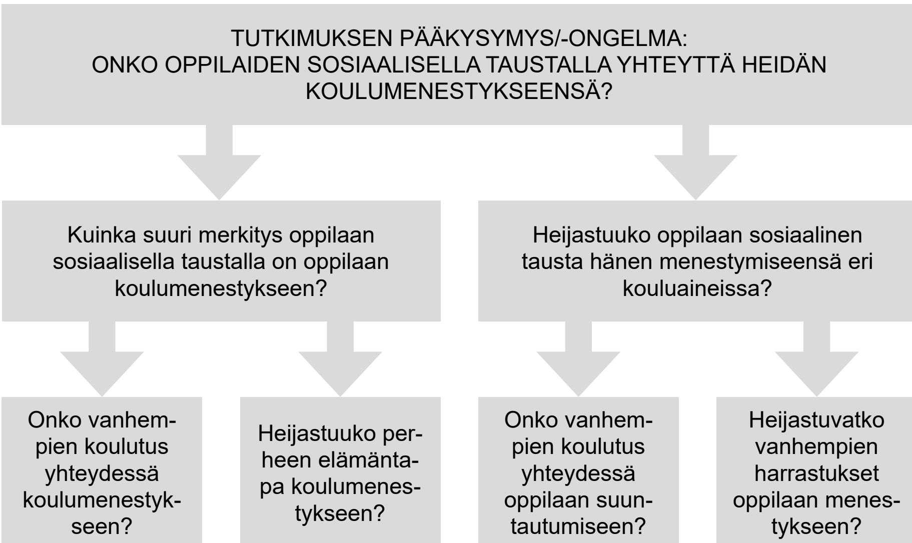
Jotta tutkimuksen otos edustaisi tutkimuksen kohteena olevaa populaatiota mahdollisimman kattavasti, on tärkeää pohtia otannan satunnaisuuden lisäksi, kuinka suuri otoksen tulisi olla. Tämä on tärkeää, koska eri analyysimenetelmien käyttö edellyttää tietyn määrän tapauksia, jotta ne voidaan toteuttaa tarkoituksenmukaisella tavalla. Otoskoko vaikuttaa myös siihen, millaisiksi tilastolliseen päättelyyn liittyvät keskivirheet analyyseissä muodostuvat[1]. Tätä voidaan otantasuunnitelman yhteydessä yleensä huomioida yksinkertaisesti suurentamalla otoskokoa. Toisaalta tulosten tulkinnan yhteydessä on tärkeää muistaa, että suurissa aineistoissa/otoksissa perinteisesti käytettävät tilastollisen päättelyn analyysit ilmoittavat herkästi hyvinkin pienet ryhmien väliset erot tai muut efektit tilastollisesti merkitseviksi, vaikka käytännössä esimerkiksi ryhmien välinen ero tai saatu korrelaatio on itse asiassa merkityksettömän pientä.

Otoksen muodostamista varten on varminta aina käyttää tilastollisia otantatekniikoita, kuten esimerkiksi yksinkertainen satunnaisotanta, systemaattinen otanta, ositettu eli stratifioitu otanta tai ryväs- eli klusteriotanta. Näissä yhteisenä piirteenä on se, että niissä kaikissa yksiköiden valintaan vaikuttaa jossain vaiheessa otantaa sattuma. Yksinkertaisella satunnaisotannalla (YSO) tarkoitetaan otantaa, joka perustuu puhtaasti arvontaan ja jossa kaikilla perusjoukon alkioilla on yhtä suuri, sattumaan perustuva, todennäköisyys tulla valituksi tutkimuksen otokseen. Arvonta voidaan käytännössä nykyisin suorittaa verrattain helposti esimerkiksi taulukkolaskentaohjelman satunnaislukufunktioita käyttäen. Periaatteessa jokin ”hattu & nimilaput” -tyyppinen arvonta kuvastaa hyvin mistä on kyse, ja tällainen toteutuskin olisi satunnaisuuden näkökulmasta mahdollinen. Toteutus vain saattaisi käytännössä osoittautua turhan työlääksi! Yksinkertainen satunnaisotanta esitetään yleensä otoksen muodostamisen perusmenettelyksi. Muita otantatyyppejä ovat yllä mainitut variaatiot satunnaisesta otannasta, esimerkiksi ositettu- ja ryväsotantatekniikat. Esimerkiksi kasvatustieteellisessä tutkimuksessa tutkimusasetelmat ovat usein sellaisia, että ryväsotanta (cluster sampling) tarjoaa tarkoituksenmukaisen perustan tutkimuksen aineiston keruulle. Esimerkiksi laajemmissa kasvatustieteellisissä tutkimuksissa ollaan usein kiinnostuneita erilaisten koulujen, luokkien ja perheiden (kotitalouksien) välillä ilmenevistä eroista, joita ajatellaan voitavan ymmärtää ja selittää joillakin näihin liittyvillä erityispiirteillä. (ks. otantatekniikoista ja niihin liittyvistä ratkaisuista ja toimenpiteistä enemmän esim. Pahkinen 2012; Kerlinger 1981, 117–130; Alkula ym. 2002, 106–117.)

Itse spesifien tutkimuskysymysten tai -ongelmien määrittäminen on tärkeä tutkimuksen tekemisen vaihe, koska nämä määrittelevät sen lisäk-

[1] Otantaan liittyvää epätarkkuutta voidaan arvioida käytettävään estimaattoriin (esim. otoskeskiarvo) liittyvän **keskivirheen** avulla. Keskivirhe voidaan tulkita estimaattorin keskihajonnaksi, jos vastaavaa otantaa voitaisiin toistaa lukuisia kertoja. Keskivirheen avulla voidaan muodostaa estimaattiin liittyvä virhemarginaali ja sen perusteella saatava **luottamusväli**, jolla perinteisen tulkinnan mukaan perusjoukon arvo sijaitsee tietyllä todennäköisyydellä (ks. tarkemmin luku 1.6.1.3).

si, millaista tutkimusaineistoa on tarvetta kerätä, myös sen, miten tätä tulisi analysoida ja millä analyysimenetelmillä tähän liittyviä muuttujia olisi tarkoituksenmukaista tutkia. Tähänkin löytyy eri tutkimusmenetelmäkirjoista ohjeita ja esimerkkejä. Keskeistä näiden asettamisessa on se, että tutkimuskysymyksistä, pää- ja alaongelmista, muodostuu mahdollisimman eheä ja kattava kokonaisuus sen suhteen, mistä tutkimuksessa oikeasti ollaan kiinnostuneita ja mitä on tarkoitus tutkia. Tässä suhteessa voi olla järkevää tehdä itselleen selkeä hierarkkinen asetelma pää- ja alaongelmista (ks. alla oleva esimerkki). Tämän avulla on helppo tarkistaa, että kaikki olennainen tulee kysyttyä. Alla olevassa asetelmassa kuvataan esimerkinomaisesti pää- ja alakysymysjaottelu, kun tutkimuksen pääkysymys liittyy siihen, millainen yhteys oppilaiden sosiaalisilla taustatekijöillä on heidän koulumenestykseensä:



ASETELMA 2. Tutkimuksen ongelmanasetteluesimerkki (Kari & Huttunen 1994, 24).

Yllä olevassa asetelmassa on määritelty ensin tutkimuksen yleiskysymys, johon halutaan hakea vastausta. Tämän jälkeen on asetettu kaksi alakysymystä, joiden avulla ajatellaan voitavan vastata tutkimuksen pääkysymykseen tarkoituksenmukaisella tavalla. Tämän jälkeen tutkija on asettanut lisäkysymyksiä, jotka vuorostaan liittyvät keskeisesti näihin kysymyksiin. Kun näin on saatu asetettua tutkimukselle relevantin tuntuiset ja perustellut tutkimuskysymykset, tutkijan eteen tulee kysymys, millaisia kysymyksiä hän tarvitsee saadakseen tietoa näistä ja tietysti myös oppilaiden koulumenestyksestä. Näitä tutkija joutuu pohtimaan, keräsi hän aineiston sitten haastattelemalla tai kyselylomakkeella. Koulumenestykseen liittyvät tiedot voidaan saada myös kouluilta, jos tähän on saatu lupa.

Määrällisen tutkimuksen peruskysymykset liittyvät ennen kaikkea tutkittavan ilmiön tai tutkittavien henkilöiden ja tapausten ominaisuuksien ja niiden jakaumien tutkimiseen: kysytään esimerkiksi sitä, miten paljon joukossa P on "olioita tai tapauksia, joilla on ominaisuus X", tai haetaan selitystä sille, miksi "joukossa P esiintyy ominaisuutta Y" sekä sitä, onko mitatuilla ominaisuuksilla X ja Y keskinäistä riippuvuutta, ja millaista se on. Lisäksi tutkijoita kiinnostaa yleisesti se, minkälainen mahdollisesti havaittu yhteys on luonteeltaan, voiko sen mieltää todelliseksi tai kausaaliseksi vai onko tämä yhteys vain näennäinen. Pidemmälle menevissä tutkimuksissa voidaan olla kiinnostuneita myös siitä, liittyykö X:n ja Y:n välille jonkinlainen yhdistävä kausaalinen mekanismi, ja siitä, millainen se on. (Töttö 2004, 14; Jokivuori & Hietala 2007, 23.) Empiirinen määrällinen tutkimus on siis samaan aikaan sekä kuvailevaa että tulkitsevaa, induktiivista ja teoreettista.

Tutkimusasetelman ja tutkimuskysymysten asettaminen on yksi aivan keskeisin tutkimuksen tekemisen vaihe. Tässä kohdin on siis pohdittava tarkoin sitä, millaisia mittareita tai millaista kyselypatteristoa kyseisessä tutkimuksessa tarvitaan. Kyse on siis tutkimuskysymysten ja eri tekijöiden operationaalistamisesta mitattavaan muotoon. Toisinaan tässä voi käyttää valmiiksi suunniteltuja ja aikaisemmissa tutkimuksissa hyväksi havaittuja kysymyksiä, lomakkeita tai jopa standardoituja testejä. Standardoitujen testien käyttämiseen vaaditaan yleisesti lupa, ja usein myös niiden käyttö edellyttää käyttäjältä tiettyä koulutusta. Mikäli käyttää jonkun toisen tutkimuksen kyselylomaketta tai tämän osia oman tutkimusaineiston keruussa, on tämä tuotava esiin selkeästi tutkimusraportissa. On kuitenkin yleistä, että esimerkiksi pro gradu -tutkielmissa opiskelijat laativat itse omat kyselylomakkeensa tai haastattelurunkonsa. Koska määrällisessä tutkimuksessa kysymysten ja kysymyslomakkeiden laatiminen on niin keskeinen tekijä tutkimusaineiston analyysin kannalta, käymme vielä lyhyesti läpi tämän laatimiseen liittyviä yleisiä lähtökohtia ennen kuin menemme varsinaiseen asiaan.

1.4 Kyselylomakkeen laadinnan pääpiirteet

Kvantitatiivisessa tutkimuksessa lähtökohtana on empiirisen ilmiön tutkiminen mitattavin määrein. Osa tekijöistä, kuten sukupuoli, ikä tai tulotaso, ovat helposti määriteltävissä. Toisten, kuten älykkyys, luovuus tai kouluviihtyvyys, määrittäminen kvantitatiivisesti mitattavaan muotoon on huomattavasti kompleksisempaa. Tämä edellyttää tutkijalta tutkimusalan ja tutkittavan ilmiön perusteellista tuntemusta. Tutkimusaineisto kerätään yleensä haastattelun, kyselylomakkeen tai observoinnin avulla. Kyselylomakkeen voi laatia ja kyselyn toteuttaa myös sähköisenä esimerkiksi nettiselaimella käytettävällä Webropol-sovelluksella, jonka käyttämiseen on esimerkiksi monissa oppilaitoksissa käyttöoikeus[2]. Kyselylomakkeen vastaukset saa tallennettua sovelluksesta mm. Excel-tiedostona tai valmiina SPSS-ohjelmaan sopivana aineistona.

Kyselylomake sopii hyvin aineiston keruun välineeksi tutkimuksissa, joissa ollaan kiinnostuneita muun muassa tutkittavien mielipiteistä, arvoista, asenteista ja kokemuksista. Näin tutkimusmateriaali saadaan kvantitatiivisessa eli numeerisessa muodossa, jonka avulla päästään käsiksi tutkimusongelmiin esimerkiksi vastausten frekvenssijakaumia tai erilaisia vastauksista muodostettuja skaalamittareita (ns. summamuuttujat) käyttäen. Haastattelulla ja lomakekyselyllä on omat etunsa ja rajoituksensa (ks. asetelma 3). Haastattelun eduiksi luetaan yleensä sen joustavuus; haastattelu ei vaadi vastaajilta erityisiä taitoja (esim. kirjallisia), mahdollisuus muunnella kysymysten järjestystä ja esittää tarkennettuja kysymyksiä. Erityinen etu on se, että haastateltavan oma ääni pääsee paremmin esille. Otoskato jää yleensä selkeästi postikyselyä pienemmäksi. Haastatteluun liittyvistä ongelmista suurimpia ovat sen suuritöisyys ja kalleus. Posti- ja verkkokyselyn suurimpia etuja ovat edullisuus ja keruuvaiheen selkeys ja nopeus. Toisaalta postikyselyssä otoskato jää usein varsin suureksi. Tähän liittyvään riskiin tutkijan on syytä varautua jo kyselylomakkeita lähetettäessä. Sellaisille vastaajaryhmälle, jonka arvelee vastaavan tällaisiin kyselylomakkeisiin syystä tai toisesta vähän laiskasti, voi lähettää kyselyjä vähän enemmän kuin muille ryhmille. Näin on odotettavissa, että heiltäkin saadaan ainakin vähän enemmän vastauksia kuin muutoin saataisiin. Luonnollisesti, jos tutkijan on mahdollista mennä paikan päälle täytättämään lomakkeitaan, vastausprosentti saadaan merkittävästi korkeammaksi kuin postitettujen tai verkossa olevien kyselyiden kohdalla saadaan. Paikalle meneminen ei tule useinkaan kuitenkaan kyseeseen, mutta esimerkiksi koululaisiin tai tietyn alan työntekijöihin kohdistuvissa paikallistutkimuksissa tämä saattaa olla hyvinkin mahdollista.

Lomakekyselyssä saadaan harvoin kerätyksi niin monipuolista aineistoa kuin haastattelulla, vaikka lomakkeen kysymykset olisivat avoimia. Haastattelu on aina vuorovaikutustilanne tutkijan ja tutkittavan välillä, ja siihen voidaan samalla liittää observoinnin kautta saatava tieto. Kyselylomakkeiden ongelmak-

[2] Ks. laajemmin Webropol-kyselyn laatimisesta esim. Ronkainen & Karjalainen 2008. Internetistä löytyy myös hyviä käyttöoppaita kyselyn laatimisen avuksi.

si muodostuu myös niiden vastaajilta edellyttämät perustaidot, esimerkiksi avoimien kysymysten ongelmaksi muodostuu usein se, etteivät vastaajat ole tottuneet ilmaisemaan itseään kirjallisesti. Nuorille tai kielitaidoiltaan heikoille vastaajille voi tuottaa myös vaikeuksia ymmärtää väittämiä, joiden sanoituksissa on käytetty kiertoilmauksia tai mutkikkaita rakenteita. Kyselylomakkeen laadintavaihe on erittäin keskeinen tutkimuksen onnistumisen kannalta; itse tilastoanalysointivaihe on taas jokseenkin selkeä, jos kyselylomake on laadittu tutkimukselle asetetun tutkimusasetelman ja tutkimusongelmien pohjalta.

Asetelma 3 konkretisoi hyvin sen, kuinka erilaisia tiedonkeruutapoja haastattelu ja kyselylomake ovat. Se, kumpaa näistä päämuodoista tutkimuksessa on järkevää käyttää, riippuu paljolti tutkimuksen tavoitteista. Sinänsä molemmilla tavoilla kerättyä aineistoa voidaan analysoida tilastollisin menetelmin, vaikka useimmiten haastatteluaineistoa tulkitaan kvalitatiivisin menetelmin. Se, miten aineistoa käsittelee, ei sinänsä ole tärkeää, vaan tärkeintä on, että ne ovat tutkimuskysymysten ja asetelman suhteen relevantteja.

Kyselylomakkeen laatimiseen on syytä siis varata riittävästi aikaa. Jo lomaketta laadittaessa tulisi ottaa huomioon valitun analyysimenetelmän ja tehtävien tulkintojen vaatimukset. Jos lomake laaditaan itse, eli se ei perustu mihinkään aiemmissa tutkimuksissa käytettyyn valmiiseen lomakkeeseen tai mittariin, tulisi se myös esitestata varsinaista tutkimusryhmää vastaavilla. Tällöin voidaan tarkistaa suunniteltujen kysymysten ja kysymysosioiden muodostamien kokonaisuuksien tarkoituksenmukaisuus ja ymmärrettävyys. Lisäksi voidaan kokeilla, miten lomakkeella saatavat tiedot ovat analysoitavissa tilasto-ohjelmalla. Seuraavassa on luettelomaisesti esitetty kyselylomakkeen laadinnassa huomioonotettavia seikkoja (tarkemmin kyselylomakkeen laadinnasta esim. Tuckman 1988, 212–259; Kerlinger 1981, 479–488; Alkula, Pöntinen & Ylöstalo 2002, 130–137; Vehkalahti 2014, 17–50; Eskola 1975, 158–184; Heikkilä 2014, 45–70):

1. LOMAKKEEN SUUNNITTELUVAIHE

- tutkimuksen tavoitteiden ja ongelmien määrittäminen
- tutkimuksen pää- ja alakysymysten operationalisointi (kysymysten tulee olla suhteutettu tutkimusongelmaan, on mietittävä kunkin kysymyksen kohdalla: onko tämä se, mitä halutaan mitata)
- kysymystyyppien määrittely (esim. suljetut kysymykset, avoimet kysymykset; tosiasia- tai mielipidekysymykset, suorat tai epäsuorat

LÄHTÖKOHTA	HAASTATTELU	KYSELYLOMAKE	VERKKOKYSELY
TUTKIMUSAINEISTON LUONNE	spesifiltä alalta, henkilökohtaisuus, epäsuoruus	faktuaalista, asenne- yms. väittämiä, suoria ja luokitt. kysymyksiä	faktuaalista, asenne- yms. väittämiä, suoria ja luokitt. kysymyksiä
AVOINTEN KYSYMYSTEN KÄYTTÖMAHDOLLISUUS	suuri	voidaan esittää, mutta yleisesti näihin jätetään joko vastaamatta tai vastataan hyvin lyhyesti	voidaan esittää, mutta yleisesti näihin jätetään joko vastaamatta tai vastataan hyvin lyhyesti
TUTKIMUSHENKILÖSTÖN TARVE	edellyttää haastattelijoita	edellyttää kirjaajaa	edellyttää lomakkeen tekijää ja vastausten käsittelijää
PÄÄKULUT	haastattelijoiden palkat + iterointi kulut	painatus-, postitus- ja palkkakulut	nettilomakkeen luonti- ja käsittelykulut
MAHDOLLISUUDET VASTAUSOHJEIDEN ANTAMISEEN	laajat	rajoitetut	rajoitetut, mutta laajemmat kuin kyselylomakkeilla
MAHDOLLISUUDET (LISÄ)KYSYMYSTEN TEKOON	laajat	rajoitetut	rajoitetut
TILAISUUS SYVENTÄÄ VASTAUKSIA (ohjeiden mukaisesti)	mahdollista	ei juuri ole	ei juuri ole
DATAN REDUKTION SUHTEELLINEN SUURUUS	suuri (koodaus)	melko vähäinen	melko vähäinen
OTOKSEN KOKO	rajoitettu	suuri	suuri
KADON SUURUUS	vähäinen, tosin aina ei ole helppo löytää haastateltavia	suurehko	suurehko + lisäksi tutkimuksen perusjoukon määrittely ongelmista
VIRHELÄHTEET	haastattelija, välineet, koodaus, otos tai näyte	mittarin rajoitukset, väärinkäsitysmahdollisuudet, otokseen tai näytteeseen liittyviä	mittarin rajoitukset, väärinkäsitysmahdollisuudet + otokseen tai näytteeseen liittyvät ongelmat – ei kyetä tekemään esim. korrektisti väestötason otanta
KOKONAISRELIAABELIUS (overall reliability)	melko rajoittunut	melko hyvä	melko hyvä
EDELLYTTÄÄKÖ LUKEMISEN JA KIRJOITTAMISEN VALMIUKSIA	ei	edellyttää	edellyttää

ASETELMA 3. Haastattelun, kyselylomakkeen ja verkkokyselyn lähtökohtia (mukaillen Tuckman 1988, 225; Heikkilä 2014, 18; Ronkainen 2008).

kysymykset, kysymys- tai väittämämuoto); se, mikä on oikea tapa kysyä, riippuu paljolti tutkimusongelmasta, vastaajista sekä valituista aineiston analysointitavoista ja niin edelleen
- tässä vaiheessa pohditaan myös vastauksen muoto (esim. avoimet vastaukset, strukturoidut; fill-in response, skaalavastaukset), joka määrittyy tutkimuksen informaation tarpeesta käsin

2. LOMAKKEEN LAATIMINEN

Lomakkeen rakenne

a) taustatiedot (sisältää henkilö- ja taustatiedot)
b) tieto-osa (varsinaiset kysymykset)

- kysymykset on yleensä hyvä järjestää asiakokonaisuuksiksi (voidaan erottaa esim. väliotsikoinnin avulla), kysymysten olisi hyvä olla aikajärjestyksessä ja ne olisi syytä järjestää vaikeusjärjestykseen (ensin helpoimmat kysymykset jne.); tosin kysymysten järjestyksen suhteen käytetään monia erilaisia tapoja, tarkoituksenmukainen ratkaisu riippuu aina tutkimuksen tavoitteista
- kysymysten tulee olla selkeitä ja yksiselitteisiä, eivätkä ne saa olla johdattelevia
- kysymykset saavat mitata vain yhtä asiaa kerrallaan. On tärkeätä varmistaa, että kysymyksissä ei ole kaksiosaisia väittämiä, kuten esimerkiksi ”harrastan joukkueurheilua ja pidän siitä”
- kysymykset eivät saa olla myöskään ”sosiaalisesti ladattuja”, jotta ei saataisi vastaukseksi vain vastauksia, jotka heijastavat enemmän sitä, mitä esimerkiksi yleinen mielipideilmasto odottaa kuin vastaajan todellisia ajatuksia kysymyksestä.
- sosiaalista suotavuutta voidaan vähentää esittämällä osa kysymyksistä käänteisesti: esimerkiksi kouluviihtyvyyttä tutkittaessa väittämien joukossa on hyvä olla myös kielteisiä ilmauksia, kuten ”en viihdy koulussa”.

Vastausvaihtoehdoista

- ei liian monta, eivätkä ne saa olla liian pitkiä. Mielipidettä mittaavissa väittämissä, joita on tarkoitus myöhemmin yhdistää summamuuttujiksi, on hyvä käyttää vastausskaalaa, jossa on riittävästi vastausvaihtoehtoja: esimerkiksi 5-portainen Likert-asteikko toimii yleensä hyvin aikuisilla vastaajilla.

- asteikot samaan, loogiseen suuntaan (esim. Likert-asteikossa 1 tarkoittaa täysin eri mieltä tai vastaavaa, 5 tarkoittaa täysin samaa mieltä, eli pienempi luku viittaa "pienimpään ilmiön olemassaoloon" ja päinvastoin); tosin kysymyksen tai väittämän vastausskaala voi kysymyslomakkeella olla käännetyssä järjestyksessä, esimerkiksi asennemittareissa ja vastaavissa tämä on eräs perustaktiikka kysymyksiä laadittaessa, mutta tilastoanalyysivaiheessa eri osioiden arvojen tulee muodostaa looginen kokonaisuus, eli käänteisten kysymysten kohdalla asteikkoskaala tulee kääntää joko koodausvaiheessa tai suorittaa tämä käännös tilasto-ohjelmalla.
- asteikon keskellä oleva vaihtoehto tulee sanoittaa niin, että se aidosti kuvaa skaalan keskikohtaa tai esimerkiksi neutraalia mielipidettä. Hyvä muoto monesti on ”ei samaa eikä eri mieltä”. Mikäli on mahdollista, että vastaaja ei lainkaan tunne asiaa tai ei yksinkertaisesti pysty ottamaan kantaa, voi varsinaisen skaalan ulkopuolelle sijoittaa lisävaihtoehdon ’en osaa sanoa’ tai ’en tiedä’.
- on vältettävä vastausmuotoja, jotka suosivat jotain tiettyä vastaussuuntaa

Vastausohjeet ja vastausesimerkit

- vastausohjeet lomakkeen alkuun ja aina siinä kohdassa lomaketta, jossa vastaustapa muuttuu (kovin useita erilaisia vastaustapoja ei kannata sisällyttää yhteen kyselyyn)
- ohjeissa tulee ilmetä merkintäpaikka ja -tapa, valintamahdollisuudet

Lomakkeen kieliasu

- lomakkeen laatijalla, käyttäjällä ja vastaajalla tulisi olla "sama kieli" (yleiskieltä, ei epäselviä tai monimerkityksisiä käsitteitä, yksinkertaiset virkkeet)
- kielen selkeys, ymmärrettävyys ja kysymysten yksiselitteisyys

3. LOMAKKEEN ESITESTAUS

- esitestauksella voidaan arvioida ja parantaa kyselylomakkeen kysymysosioita muodon, järjestyksen ja ymmärrettävyyden yms. suhteen
- voidaan parantaa ns. summamuuttujien, useista eri muuttujista muodostuvien muuttujien, sisäistä homogeenisuutta
- parantaa kysymysten erottelukykyä yms.

- voidaan tutkia kysymysten homogeenisuutta, lisätä tai poistaa kysymyksiä, muokata annettuja ohjeita ja testata myös saatujen vastausten tarkoituksenmukaisuutta ja testattavuutta. Lisäksi voidaan testata mittarin eri osioiden homogeenisuutta ja parantaa tarvittaessa osion eri kysymyksiä tai vaikkapa poistaa joitakin mittarin osioita yms.
- kaiken kaikkiaan parantaa siis tutkimuksen reliaabeliutta ja myös validiutta

Hyvän kyselylomakkeen kriteereiksi voi edellisen pohjalta määritellä 1) ulkoasun selkeyden ja siisteyden, 2) kysymysten ymmärrettävyyden (otettava huomioon vastaajien erilaiset taustatekijät) ja yksiselitteisyyden, 3) kysymysten neutraalisuuden eli kysymykset eivät siis saa olla johdattelevia, 4) kysymysten validiuden (operationaalistamiseen liittyvä kysymys; mittaavatko kysymykset todella sitä, mitä niiden on tarkoitus mitata) ja 5) vastausten jatkotyöstämisen vaivattomuuden. Tavoitteena on, että lomake on mahdollisimman vaivaton täyttää ja siitä saatu informaatio on helposti siirrettävissä tilasto-ohjelmaan. Tärkeätä on, että lomakkeessa olevat kohdat ovat oikeassa suhteessa tutkimuskysymyksiin. Lomakkeella kysymyksiä ei saisi olla liikaa – kukaan ei jaksa keskittyä vastaamaan mittaripatteristoon, jonka täyttämiseen kuluu tunteja. On siis tärkeää, että kysymykset liittyvät tutkimuksen keskeisiin käsitteisiin mahdollisimman kattavasti.

Kyselylomakkeita esimerkiksi kotiin lähetettäessä saatekirjeen laadintaan tulee kiinnittää erityistä huomiota. Näin varmistetaan se, että mahdollisimman moni lomakkeen vastaanottajista motivoituu vastaamaan siihen. Riittävä vastausprosentti on yksi merkittävä tekijä, kun pohdimme tutkimuksen tulosten yleistettävyyttä. Tuckmanin (1988, 245) mukaan saatekirjeestä tulisi ilmetä tutkimuksen tavoite, tutkimuksen taustayhteisö sekä tutkijan ja vastaajien yksityisyyden eli heidän anonymiteettinsä turvaaminen. Lisäksi kirjeessä voi antaa yleisiä vastausohjeita (spesifit vastausohjeet annetaan varsinaisessa kyselylomakkeessa) ja luonnollisesti lomakkeen palauttamisohjeet ja maininta mahdollisesta muistutuskirjeen lähettämisestä. Lisäksi vastaajien motivaatiota voi lisätä lupaamalla heille informaatiota tutkimustuloksista. Jos tällaisen lupauksen antaa, on se myös pidettävä. Saatekirje ei saa kuitenkaan olla liian pitkä. Ylipäätään, vaikka tutkija olisi itse paikalla, kun tutkittavat vastaavat kyselylomakkeeseen, on pidettävä huolta heidän motivoinnistaan ja siitä, että he ottavat vakavasti vastaamisen. Tällöinkin on hyvä selvittää heille, miksi tutkitaan, kuka tutkii ja miksi juuri heidän vastauksensa on tärkeä tutkimuksen kannalta.

Kun kyselylomaketta laaditaan, on syytä ottaa huomioon myös se, miten ja millä tasolla aineistoa on tarkoitus analysoida. Tähän liittyy läheisesti kysymys mittauksen asteikkotyypistä. Seuraavassa tarkastellaan lyhyesti tähän liittyviä kysymyksiä.

1.5 Muuttujien mitta-asteikko

Eri tilastotieteellisten analyysimenetelmien käytön yksi keskeisistä kriteereistä on se, millä tasolla tai millä tarkkuudella muuttujat on mitattu. Muuttujat jaetaan yleisesti ensinnäkin kategorisiin ja numeerisiin. Ensin mainitut muuttujat kuvaavat laadullisia ominaisuuksia niin, että niissä käytetyt luokat sulkevat toisensa pois (esim. ammatti-, kotipaikka- ja siviilisäätyluokitukset). Numeeriset muuttujat ovat puolestaan sellaisia muuttujia, joita voidaan kuvata reaaliluvuilla (esim. perheen koko, ikä tai tulotaso). Monesti myös kategorisilla muuttujilla havaintoarvot koodataan numeroilla, mutta nämä toimivat vain vastauskategorioiden tunnuksina, eikä niitä ole tarkoitus käyttää millään lailla laskutoimituksissa. Tämän jaottelun lisäksi muuttujat jaetaan niiden mitta-asteikon ”kehittyneisyyden” mukaisesti neljään ryhmään, eli laatuero- (nominaali), järjestys- (ordinaali), välimatka- (intervalli) ja suhdeasteikkoisiin muuttujiin. Kaksi ensimmäistä asteikkoryhmää kuuluvat kategoristen ja kaksi jälkimmäistä numeeristen muuttujien ryhmään. Muuttujien mitta-asteikko ilmaisee mittauksen ilmaisukykyä ja mittauksen tasoa. Seuraavassa tietolaarissa 2 on vielä kuvattu tiivistetysti eri mitta-asteikkojen peruskriteerit (ks. esim. Helenius 1989, 18–22; Alkula ym. 2002, 84–87; Heikkilä 2014, 81–82).

Muuttujien mitta-asteikko on keskeinen määrittelevä tekijä sen suhteen, mitä tilastollista analyysimenetelmää niiden kuvailussa ja analyysissa voidaan soveltaa. Laatueroasteikko on mittaustasoltaan vaatimattomin (samanlaisuus, erilaisuus) ja suhdeasteikko eksaktein (tällöin käytettävässä asteikossa on absoluuttinen nollapiste, eli nollapisteessä mitattava ominaisuus häviää kokonaan, esim. ikä). Tämä tarkoittaa, että laatuerotason kysymyksistä/muuttujista voidaan laskea eri vastausvaihtoehtojen lukumääriä ja prosenttiosuuksia, joita voidaan kuvailla ja tarkastella frekvenssijakaumien, pylväsdiagrammien ja ristiintaulukointien avulla. Näillä lukumäärillä ei ole keskenään mitään tarkkaa järjestystä, eikä niitä voi verrata mielekkäästi esimerkiksi enemmän–vähemmän -akselilla, vaan ne ovat laadullisia muuttujia, joiden

TIETOLAARI 2. MUUTTUJIEN MITTA-ASTEIKOT.

Kategoriset muuttujat
Kategoriset muuttujat kuvaavat laadullisia ominaisuuksia, jotka ovat toisensa poissulkevia; esim. siviilisääty, ammatti, sukupuoli tai asenteet.

1. Laatuero- eli nominaaliasteikko
- kuvataan kvalitatiivisten muuttujien suhdetta, lähinnä luokittelemalla muuttujat eri ryhmiin
- muuttujien arvoilla ei ole määrättyä paikkaa tai järjestystä
- käytettyjen luokkien välillä ei ole ”tarkkaa” järjestystä
- samanlaisuus/erilaisuus (esim. sukupuolijaottelu, ammattiluokitus)
- havainto voi kuulua vain yhteen luokkaan
- laskutoimitukset ja niihin perustuvat tunnusluvut, esimerkiksi keskiarvo ja keskihajonta, eivät ole mielekkäitä

Suositeltavat kuvailu- ja testausmenetelmät:
- epäparametriset menetelmät (nonparametriset menetelmät)
- frekvenssiesitykset (esim. histogrammi, ristiintaulukointi, moodi)
- khiin neliö -testi (yhteensopivuus- ja riippumattomuustestit)
- log-lineaariset mallit

2. Järjestys- eli ordinaaliasteikko
- mittaustaso ilmaisee havaintoyksiköissä olevaa samanlaisuutta/erilaisuutta sekä näihin liittyvää järjestystä
- kuvataan kvalitatiivisten muuttujien sisältämien ominaisuuksien sijoittumista eri ”tasoille tai luokkiin” esim. suuruuden, paremmuuden tai luokkatason mukaan
- näiden välisen eron suuruutta ei kuitenkaan voida määritellä tarkkaan, koska luokkien väliset erot eivät välttämättä ole yhtä suuria, tasavälisiä,
- voidaan verrata muuttujia tai ryhmiä keskenään (pienempi tai suurempi kuin -asetelman mukaisesti), mutta yleisesti tulkiten laskutoimitukset eivät ole mielekkäitä tässäkään asteikkotyypissä
- kuitenkin esimerkiksi 5-portaisella Likert-asteikolla mitatut asennemuuttujat, jotka ovat periaatteessa järjestysasteikkoisia muuttujia (eri vastausluokkien eron suuruutta ei voida eksaktisti määrittää), tulkitaan käytännössä usein välimatka-asteikollisiksi muuttujiksi, joissa vastausvaihtoehtojen väliset etäisyydet oletetaan samanmittaisiksi; varsinkin näin on, kun näistä muuttujista muodostetaan uusia monen muuttujan sisältämiä summamuuttujia
- tällöin niiden analyyseissa voidaan soveltaa mm. keskiarvoihin perustuvia analyysejä ja testejä; lisäksi näistä voidaan tällöin tehdä summamuuttujia
- se, voidaanko, näin tehdä on aina tutkijan tulkinnasta ja oman tiedeyhteisön käytänteistä kiinni

Suositeltavat kuvailu- ja testausmenetelmät:
- samat kuin nominaaliasteikkoisillakin, lisäksi:
- mediaani, järjestyskorrelaatio
- erilaiset epäparametriset menetelmät

Numeeriset muuttujat
Numeeriset muuttujat ovat määrällisiä muuttujia, joiden arvot ovat reaalilukuja, jotka saadaan joko mittaamalla tai havainnoimalla.

3. Välimatka- eli intervalliasteikko
- voimme määritellä muuttuja-arvojen a ja b eron etäisyyden tai suuruuden, emme voi kuitenkaan määritellä, kuinka monta kertaa a on b:tä suurempi
- intervalliasteikolta puuttuu absoluuttinen nollapiste (se on sopimuksenvarainen), esim. lämpömittarin asteikko
- useimmat käyttäytymistieteiden käyttämät mitat yltävät korkeintaan tälle asteikkotasolle (usein kuitenkin erilaisilla mittareilla todetut arvot lasketaan välimatka-asteikkoisiksi, esim. todistusnumerot yms.)

- muuttujien arvojen järjestys ja etäisyydet ovat määrätyt; muuttujien arvot eivät ole toisiinsa nähden eksaktit, vaikka muuttujien arvojen eksaktia suhdetta ei voida määritellä yleisesti tulkitaan, että useimmat laskutoimitukset ja parametriset tilastoanalyysit ovat tällä mittaustasolla sallittuja

Suositeltavat kuvailu- ja testausmenetelmät:
- Vparametriset ja epäparametriset testit
- tulomomenttikorrelaatio
- regressioanalyysit
- pääkomponenttianalyysi, faktorianalyysi sekä muut monimuuttujamenetelmät
- keskiarvotestit (t-testit, varianssianalyysit)

4. Suhdeasteikko
- mitattavilla arvoilla on absoluuttinen nollapiste, eli kun mittauksen arvoksi tulee nolla, niin mitattavaa ominaisuutta ei esiinny (fysikaaliset ominaisuudet, pituus, paino, tilavuus yms.)
- voidaan verrata saatuja arvojen/pistemäärien suuruuseroja
- lukuarvoja voidaan laskea yhteen ja kertoa

Suositeltavat kuvailu- ja testausmenetelmät:
- samat kuin välimatka-asteikollakin

perusteella havaintoyksiköt voidaan sijoittaa eri ryhmäluokkiin, esimerkiksi sukupuoli-, viriketausta- tai vaikkapa ammattiryhmiin. Kun numeerisilla, välimatka- ja suhdeasteikoilla mitattujen muuttujien sisällä voidaan tehdä erilaisia laskutoimituksia, laatuasteikkoisilla muuttujilla ei näitä voi tehdä. Numeerisille muuttujille on tunnusomaista myös se, että niiden ominaisuuksia voidaan kuvata keskiarvon avulla.

Muuttujien asteikkotarkastelun suhteen kasvatus- ja yhteiskuntatieteen tutkimusaloilla muodostuu erityisen hankalaksi kysymys siitä, voidaanko Likert-asteikollisiin väittämiin perustuvat muuttujat samaistaa välimatka-asteikkoisiin (ks. edellä oleva tietolaari) muuttujiin. Kuten esimerkiksi Helenius (1989, 21–22) toteaa, se, minkä mitta-asteikkoiseksi muuttuja määritellään, on viime kädessä tulkintakysymys, joka tutkijan tulee tehdä. Tässä meitä auttavat luonnollisesti kunkin tieteenalan käytännöt. Se, että Likert-asteikkoisiin väittämiin perustuvien muuttujien tulkitaan usein täyttävän ainakin jossakin suhteessa välimatka-asteikkoisen mittarin kriteerit, johtunee pitkälti siitä, että jälkimäinen mittaustaso antaa tutkijoille mahdollisuuden soveltaa niiden analysointiin enemmän tilastollisia analyysimenetelmiä kuin mitä järjestysasteikollisille mitatuilla muuttujilla on.

Jos muuttujan sisältö voidaan asettaa johonkin järjestykseen sisältönsä suhteen, nimitetään tätä järjestysasteikkoiseksi muuttujaksi. Tälläkin tasolla mittaus jää kovin epätarkalle tasolle, eikä sitä ole mahdollista analysoida tilastollisilla menetelmillä kovinkaan monipuolisesti. Tällaisia järjestysasteikkoisia kysymyksiä ovat muun muassa koulutustausta ja puoluekanta. Tällä tasolla voidaan tarkastella, kuinka paljon jotain ominaisuutta on, mutta ei voida vertailla eksaktisti, kuinka paljon tätä on enemmän kuin toisilla. Näitä muuttujia voidaan käyttää analyysissa luokittelevina tekijöinä kuten laatuasteikollisiakin muuttujia. Välimatka- ja suhdeasteikkoiset mittarit sallivat jo monipuolisempien analyysimenetelmien käytön ja tätä kautta syvälle menevämpien tulkintojen tekemisen. Käyttäytymis-, ihmis- ja yhteiskuntatieteissä harva kysymys on luonteeltaan puhtaasti sellainen, että sitä voitaisiin tarkastella eksaktisti välimatka-asteikosta puhumattakaan suhdeasteikoilla. Likert-asteikkoisten kysymysten, joissa kysytään esimerkiksi mielipidettä viisiasteikkoisen väittämän avulla, katsotaan kuitenkin täyttävän välimatka-asteikolle asetetut kriteerit. Ongelmallista näissä on kuitenkin usein mittarin keskellä olevan neutraalin väittämän tulkinta. (Ks. enemmän esim. Nummenmaa 2011, 40–47; Metsämuuronen 2009, 67–72; Vehkalahti 2014, 27–40.)

Jos asteikkojakoon suhtaudutaan erityisen tiukasti, keskiarvotesteihin perustuvia testejä ei kasvatus- ja yhteiskuntatieteellisessä tutkimuksessa voisi juuri käyttää, vaan näitä soveliaampia olisivat usein epäparametriset testit[3]. Epäparametristen menetelmien käytön etuna on siis niiden laajempi sovellusalue. Niiden käytölle on väljemmät käyttöoletukset kuin parametrisille testeille. Toisaalta niiden rajoituksena on heikompi teho eli selitysvoima parametrisiin menetelmiin verrattuna. (Siegel 1956, 1–34; Helenius 1989, 321–322). Viimeksi mainittu tekijä voi omalta osaltaan selittää sitä, ettei epäparametristen menetelmien käyttö ole yleistynyt kasvatus- ja yhteiskuntatieteissä niin kuin olisi voinut odottaa. Osasyy tälle voi olla myös se, että keskiarvolukutarkastelut ovat meille kaikille tutumpia, ja se, että keskiarvoesityksiä käytetään yleisesti julkisuudessa. Monesti parametriset ja epäparametriset testit tuottavat käytännössä varsin samankaltaiset tulokset. Tämäkin voi osaltaan selittää sitä, etteivät epäparametriset menetelmät ole saavuttaneet nykyistä suurempaa suosiota tutkijoiden keskuudessa.

Vaikka keskiarvotestien käyttö on yleistä, pitäisi niiden käytön tarkoituksenmukaisuutta arvioida aina kunkin tutkimuksen kohdalla erikseen. Muutenkin tutkijoiden kannattaisi käyttää epäparametrisia testejä nykyistä enemmän tai ainakin kokeilla näitä rinnan keskiarvotestien kanssa, kun parametristen testien kriteerit eivät täyty. Epäparametristen (esim. ristiintaulukointi, khiin neliö -testi, Mann-Whitneyn U-testi ja Wilcoxonin

[3] Epäparametrisia menetelmiä kutsutaan myös termeillä nonparametriset, ei-parametriset tai parametrittomat menetelmät.

testi) ja parametristen (esim. t- ja F-testit) menetelmien ero perustuu lähinnä niiden erilaisiin käyttökriteereihin: epäparametrisilla menetelmillä on parametrisia väljemmät käyttökriteerit. Näitä tulisi käyttää silloin, kun tutkimusaineisto ei täytä parametrisille testeille asetettuja kriteereitä (muun muassa tutkittavaa otosta vastaavan populaatiojakauman tulisi olla normaalijakauman mukainen samoin kuin otosta vastaavien populaatiovarianssien tulisi olla mahdollisimman homogeenisia). Lisäksi muuttujien tulisi olla vähintään välimatka-asteikollisia. Useihin epäparametrisiin menetelmiin riittävät jo järjestysasteikon tai jopa nominaaliasteikon mittaukset. Epäparametristen menetelmien käytön etuna on niiden suurempi sovellusalue parametrisiin testeihin verrattuna, mutta samalla niiden rajoituksena on parametrisiä testejä heikompi selitysvoima (Ks. esim. Siegel 1956, 1–34; Helenius 1989, 321–322).

Tässä kirjassa käsitellään eräitä tyypillisiä epäparametrisia vastineita omina alalukuinaan peruskeskiarvotestien esittelyn yhteydessä. Laajempi tutustuminen menetelmiin onnistuu esimerkiksi Marascuilon & McSweeneyn (1977); Reynoldsin (1977); Siegel & Castellan Jr 1988 sekä Metsämuurosen (2004) teosten avulla.

1.6 Analyysien tulkinnasta: kuvailusta ja tilastollisesta päättelystä teoreettisempaan tulkintaan

Edellä käsiteltiin jo yleisellä tasolla tilastollisten analyysien tulkintojen perusteita. Tätä on kuitenkin syytä käsitellä tässä vielä vähän tarkemmin, muodostaahan tulosten tulkinta ja argumentointi kaiken tutkimuksen teon ytimen, oli kyse sitten määrällisestä tai laadullisesta tutkimuksesta. Määrällisessä ja tilastollisessa tutkimuksessa tämä koskee niin otoksesta tehtävää tilastollista päättelyä kuin analyysitulosten laaja-alaisempaakin tulkitsemista. Aineiston tulkintaan on olemassa erilaisia ”perussääntöjä” ja ohjeita, mutta viime kädessä tulkinnan ja argumentoinnin tekee aina tutkija tai tutkijaryhmä. Tämä korostaa ensinnäkin sitä, että tutkijoiden pitää olla tässä suhteessa avoimia ja tuntea mahdollisimman hyvin tutkimaansa ilmiöön tai kysymykseen liittyvät aiemmat tulkinnat ja teoriat. Muutenhan on vaikea tulkita monipuolisesti ja tarkoituksenmukaisella tavalla omasta aineistosta nousevia tuloksia.

Konkreettiselle havaintoaineistolle rakentuvan määrällisen empiirisen tutkimuksen voi jakaa karkeasti ajatellen kuvailevaan/kartoittavaan (deskriptiiviseen) ja ”selittävään”/kausaaliseen tutkimukseen. Ensin mainitussa tutkimuksessa haetaan ja etsitään vastauksia esimerkiksi mikä-, millainen-, missä- ja milloin-kysymyksiin. Jälkimmäisessä tutkimustyypissä puolestaan pyritään selvittämään tutkittaviin ilmiöihin liittyvien eri tekijöiden yhteyksiä ja niihin liittyvien prosessien ja mekanismien luonnetta tilastollisen päättelyn keinoin ja teoreettisen tulkinnan avulla. Vastauksia siis haetaan ennen kaikkea miksi-, kuinka- ja miten-kysymyksiin. Usein tutkimus sisältää sekä kuvailevia että selittäviä tulkintoja. Yleisesti kuitenkin odotetaan tutkimuksen ylittävän pelkän kuvailevan tason. Vastaavasti tilastollisen aineiston ja analyysien tulkintaprosessin voi katsoa jakaantuvan kolmeen eri vaiheeseen (asetelma 4). Asetelma 4:n vaihe-esitys korostaa erityisesti sitä, että tilastollisten menetelmien pitäisi ylittää yhteiskunnallisissa ja kasvatustieteellisissä tutkimuksissa aineistosta nousevan kuvailun ja myös tilastollisen päättelyn tason: aineistosta nousevan tulkinnan tulisi siis aina viime kädessä rakentua olemassa olevaan tutkimustietoon ja tutkimuskohdetta koskevaa teoreettista tulkintaa vasten.

1. Aineiston kuvaus ja tulkinta

* tutkimuksen otoksessa havaittavien muuttujien jakaumien ja yhteyksien kuvaamista
* aineistolähtöistä
  - aineistoanalyysin tulosten kuvaaminen ja tulkinta
  - esim. jakaumien, keskiarvojen, muuttujien/tekijöiden/ryhmien erojen ja yhtäläisyyksien kuvaaminen ja tulkinta
  - muuttujien yhteyksien tulkintaa

2. Tuloksia koskeva päättely ja tulkinta

* perustuu tilastolliseen päättelyyn
  - otoksesta tehdään tutkimuksen perusjoukkoa koskevia päätelmiä, yleistyksiä ja tulkintoja (estimointia)
  - asetettujen hypoteesien testaaminen ja tähän liittyvä tulkinta
  - päättelystä nousevien empiiristen johtopäätösten tulkinta
  * aineistolähtöistä ja analyysimenetelmien yleisiin kriteereihin perustuvaa tulkintaa
  - erilaiset tunnusluvut ja mallien parametrien estimaatit
  - efektikoon, luottamusvälien, merkitsevyystason, p-arvojen tulkinta ja mallien yhteensopivuuden arviointi

3. Tulosten teoreettinen selittäminen ja tulkinta

* analyysin pohjalta nousevien empiiristen tulkintojen tulkitseminen tutkimuskohdetta jäsentävien teorioiden ja jo olemassa olevan tutkimustiedon pohjalta
  - kyse on siis saatujen tulosten teoreettisesta tulkinnasta ja yleistämisestä → syventää tulkintaa, lisää sen uskottavuutta ja tekee tulkinnasta tieteellisen
* tulkinnan perusteet tulevat pääosin oman tutkimusalan teoreettisista ja käytäntöjen traditioista

Lähde: mukaillen Ketokivi 2015, 18–21, 42–49.

ASETELMA 4. Tilastollisen tutkimuksen tulkintavaihekaavio.

Asetelma 4 kuvaa sitä, että tilastomenetelmien antamien tulosten tulkinta etenee kuvailu-päättely-teoreettinen -tulkinta -akselilla. Näin ollen aineistosta nousevien tulosten tulkinta lähtee liikkeelle aineiston kuvailevasta tulkinnasta eli mitattujen muuttujien jakaumien tarkastelusta ja muuttujien välisten yhteyksien kuvailusta.

Aineiston kuvailu perustuu ennen kaikkea konkreettisten jakaumien tarkasteluun, erilaisten mitattuja muuttujia kuvaavien tunnuslukujen, kuten keskiarvojen, prosenttiosuuksien, mediaanien tai arvojen vaihtelua kuvaavien tunnuslukujen, avulla. Tässä yhteydessä kannattaa käyttää hyväksi erilaisia graafisia esityksiä: havainnollisuutensa vuoksi näistä on yleensä merkittävää apua havaittujen jakaumien tarkastelussa ja tulkinnassa. Tässä vaiheessa tutkija ikään kuin ottaa haltuunsa otoksessa ilmenevät havainnot ja yhteydet. Joskus tässä vaiheessa voi nousta esille sellaisiakin keskeisiä huomioita, joita ei tutkimusta suunnitellessa edes tullut mieleen. Tällaisissa tapauksissa tutkija voi vielä jopa tarkistaa tai tarkentaa tutkimusasetelmaansa ja -kysymyksiään.

Aineiston kuvailevaa tulkintaa seuraa tilastollisen päättelyn vaihe. Tutkija tekee tässä vaiheessa edellisen vaiheen ja analyysien tunnuslukujen pohjalta päätelmiä tutkimuksensa populaation suhteen. Tilastollinen päättely on luonteeltaan induktiivista, kerätyn empiirisen aineiston perusteella tehdään yleistyksiä tutkimuksen perusjoukkoon, populaatioon. Kyse on siis tässä vaiheessa empiirisestä yleistyksestä, tilastollisesta päättelystä (asetelma 4), jonka peruskaavan voi tiivistää seuraavaan muotoon:

X on otos A:sta
n % X:n tapauksista on B
➔ otoksesta tehtävän päättelyn perusteella
n % A:n tapauksista on B

Konkreettinen tutkimusaineisto on siis yleisesti jollakin tilastollisella otantamenetelmällä määritelty otos tutkittavana olevasta perusjoukosta. Otoksessa esiintyvien jakaumien perusteella tehdään päätelmiä tietyllä riskitasolla koko perusjoukkoa koskevista jakaumista ja esimerkiksi eri ryhmien välisistä eroista tutkittavana olevan kysymyksen suhteen. Tähän päättelyyn liittyy aina tiettyä epävarmuutta ja virhepäätelmien tekomahdollisuus. Tämän riskin arvioimiseksi tilastoanalyyseihin liittyy erilaisia todennäköisyyslaskentaan perustuvia

menetelmiä, testejä, joiden avulla tutkija voi arvioida sitä, kuinka suuri riski hänen otoksesta perusjoukkoon tekemään päättelyyn ja tulkintoihin liittyy. Vastaavasti näitä menetelmiä käytetään arvioimaan muun muassa eri tekijöiden välisten havaittujen yhteyksien voimakkuutta ja tilastollista riskitasoa. Tilastollinen päättely on perinteisesti ja yleisimmin toteutettu tilastollisin merkitsevyystestein, ns. nollahypoteesin testein (NHST). Testeissä päätöksenteko perustuu testin tuottamaan merkitsevyystasoon, ns. p-arvoa käyttäen (probability value). SPSS-ohjelmassa p-arvo ilmaistaan ja esitetään lyhenteellä sig. (significance). P-arvon mekaaninen käyttö on kuitenkin viime vuosina kohdannut kritiikkiä yhä enemmän. Tämän rinnalla tai tilalla suositellaankin yhä yleisemmin käyttämään erilaisia havaittujen erojen tai yhteyksien efektikokoa mittaavia suureita (effect size estimates) sekä luottamusvälitulkintaa (CI). Nämäkään menetelmät eivät kokonaan poista tehtyihin tulkintoihin liittyvää epävarmuutta. On hyvä muistaa, että erilaiset tilastolliseen päättelyyn liittyvät menetelmät ovat vain apuvälineitä tutkijoille päättelyyn ja tulkintoihin liittyvän epävarmuuden hallitsemiseksi. Missään nimessä yksistään näiden pohjalta ei voi tehdä päätelmiä havaittujen erojen, yhteyksien tai mallien yhteiskunnallisesta merkittävyydestä.

Seuraavassa tarkastellaan tilastolliseen päättelyyn liittyviä tilastollisia testejä sekä muita menetelmiä, minkä jälkeen käsitellään vielä luvussa 1.6.2 lyhyesti asetelmassa 4 esitetyn tulkintaprosessin kolmanteen vaiheeseen, eli tulosten teoreettiseen selittämiseen ja tulkintaan, liittyviä kysymyksiä.

1.6.1 Tilastollista päättelyä tukevat testit ja tunnusluvut

Määrällisissä tutkimuksissa, joissa hyödynnetään tilastollisia analyyseja, tarkastellaan yleisesti tutkittavien muuttujien välisen riippuvuuden tai yhteyden (assosiaation) olemassaoloa. Tähän liittyy olennaisesti sen arviointi, johtuuko havaittu yhteys todellisesta muuttujien yhteydestä vai voiko saatu tulos mahdollisesti johtua otanta- ja mittausvirheen aiheuttamasta satunnaisvaihtelusta. Tässä yhteydessä puhutaan myös aineistossa olevasta signaali- ja kohinasuhteesta, jossa signaali-käsite viittaa tutkittavien muuttujien välillä esiintyvään yhteyteen ja kohina-käsite taas mahdollista otanta- ja mittausvirheestä johtuvan vaikutuksen arviointiin. Tilastolliset menetelmät auttavat näin tutkijoita sen arvioimisessa, johtuuko muuttujien välillä havaittu yhteys todella näiden välillä olemassa olevasta yhteydestä vai onko saatu yhteys vain otoksessamme esiintyvä ominaisuus, jota todellisuudessa ei voida yleistää tutkimuksen taustalla olevaan populaatioon. (Nummenmaa 2005.)

Tilastotieteelliset menetelmät jaetaan yleisesti kahteen pääsuuntaukseen, vaikka näiden rinnalla on kehitelty muitakin menetelmiä. Nämä kaksi pääsuuntausta ovat niin sanottu frekventistinen ja bayesiläinen lähestymistapa. Näiden lähtökohdat ja tarkastelukulmat eroavat toistaan sekä menetelmien perusperiaatteiden että päättelyn suhteen. Varsinkin aiemmin näitä pidettiin lähtökohtaerovaisuuksiensa vuoksi omina koulukuntinaan, mutta viime aikoina nämä on alettu nähdä pikemminkin toisiaan täydentäviksi tai rinnakkaisiksi lähestymistavoiksi (ks. esim. Ketokivi 2015, 315–319). Tässä tulkinnassaan Ketokivi osuu naulan kantaan: käytännön tutkijan ja analyysien toteuttajan näkökulmasta bayesiläisen lähestymistavan aineiston analyysiin ja tulkitsemiseen voisi teknisessä mielessä mieltää yhdeksi vaihtoehdoksi muiden analyyseihin liittyvien estimointimenetelmien joukossa.

Perinteinen frekventistinen lähestymistapa lähtee siitä, että tilastollinen päättely ja todennäköisyystulkinta – laajemmin tutkimus yleensäkin – rakentuu riittävän suuren empiirisen aineiston pohjalta tehtäviin tulkintoihin. Frekventistisen lähestymistavan nimitys tulee juuri siitä, että tulkinta rakentuu empiiriseen aineistoon siten, että tutkittavan ilmiön esiintymisen todennäköisyys ajatellaan tulevan sitä tarkemmin arvioiduksi, mitä useimpia eri mittauksia tästä voidaan tehdä. Jos näitä havaintoja (oikein satunnaistettuja) voitaisiin kerätä ääretön määrä, todellinen populaation tuntematon parametri saadaan määritettyä luotettavasti. Frekventistiselle lähestymistavalle on tyypillistä myös tutkimuksen objektiivisuuden korostus ja tavoittelu. Bayesiläinen lähestymistapa on orientaatioltaan paljon subjektiivisempi. Se lähtee liikkeelle tutkijan rakentaman mallinnuksen, olemassa olevan tiedon (a priori -tiedon tai -todennäköisyyden), pohjalta. Lähestymistavan subjektiivisuutta korostaa se, että mallinnuksen perustana olevan apriorijakauman katsotaan voivan olla ja onkin eri ihmisillä tai ihmisryhmillä (myös tutkijoilla) erilainen, joten tulkinta luonnollisesti rakentuu tältä pohjalta erilaisiksi. Toiseksi analyysi perustuu tutkimusaineistosta nousevaan todellisen (a posteriori -tietoon tai -todennäköisyyteen). Näiden pohjalta rakennetaan erilaisia hypoteeseja, joiden voimassaoloa testillä koitellaan. (Ks. esim. Gill 2002, 2–6, 18–21; Nokelainen 2010.)

Tässä kirjassa käsiteltävät tilastolliset menetelmät ja tulkinnat rakentuvat perinteiselle frekventistiselle lähestymistavalle, vaikka tekijät näkevätkin bayesiläisen lähestymistavan avaavan perinteisen lähestymistavan rinnalla tutkijoille runsaasti uusia mahdollisuuksia tutkimusaineistojen analysointiin ja tulkitsemiseen. Bayes-menetelmien tarjoamista mahdollisuuksista kiinnostuneet SPSS-ohjelman käyttäjät voivat tutustua analyysivaihtoehtoihin, joita on julkaistu ohjelmaversiossa 26. Analyysit löytyvät valikosta *Analyze → Bayesian Statistics*, josta esimerkiksi kahden riippumattoman ryhmän vertailu löytyy otsikolla *Independent Samples Normal*. Analyysien tulosteiden tulkintaan löydät ohjeistusta ja vinkkejä esimerkiksi SPSS:n *Help*-toiminnon takaa. Bayesiläisen tilastotieteen lähtökohdista ja soveltamisesta löytyy maailmalta runsaasti kirjallisuutta (ks. esim. Gill 2002; Koch 2007) suomenkielellä sitä löytyy toistaiseksi vielä melko vähän (ks. esim. Karvanen, Luoma, Penttinen, & Tikka 2019).

Yleisesti tilastolliseen analyysiin nojaavassa määrällisessä tutkimuksessa keskeisellä sijalla on tutkimusotosten pohjalta tutkimuksen perusjoukkoa koskevalla päättelyllä ja tulkinnalla. Ajatuksena on, että otoksen perusteella voidaan tehdä päätelmiä tutkimuksen perusjoukon suhteen. Tällaiseen päättelyyn liittyy aina tietty määrä epävarmuutta. Tätä epävarmuutta pyritään vähentämään ja arvioimaan erilaisin tilastollisin menetelmin (asetelma 4). Otoksesta populaatioon tehtävän päättelyn lisäksi tilastollisia menetelmiä hyödynnetään erilaisten tilastollisten mallien muodostamisessa ja rakentamisessa. Myös mallinnuksien yhteydessä tehtävässä päättelyssä hyödynnetään erilaisia merkitsevyystestejä ja menetelmiä havaittujen yhteyksien ja näiden voimakkuuden arvioimiseksi. Mallien muodostamisen yhteydessä on kyse ennen kaikkea sen arvioimisesta, kuinka saatu tilastollinen mallinnus on yhteensopiva aineistosta olevien yhteyksien kanssa. Tässäkin yhteydessä tutkijalle jää tilastollisen päättelyn jälkeen vielä mallissa havaittujen yhteyksien ja eri tekijöiden vaikutussuuntien ja -mekanismien tulkitseminen ja selittäminen. Näistähän kasvatus- ja yhteiskuntatieteellisessä tutkimuksessa ollaan viime kädessä kiinnostuneita, ei niinkään havaittujen yhteyksien tilastollisista merkitsevyyksistä. Mutta toki tilastolliset merkitsevyys-, luottamusväli- ja efektikokotulkinnat ovat tärkeä osa tutkimuksen tulosten kokonaisprosessia.

1.6.1.1 Todennäköisyys- ja merkitsevyystulkinta, p-arvotulkinta

Tilastollista merkitsevyystasotarkastelua käytetään yleisesti aineistoon liittyvän tulkinnan ja päätöksenteon riskitason arvioimiseen: tilastollisissa testeissä havaittu merkitsevyystaso (p-arvo) ilmoittaa millä todennäköisyydellä tutkimusasetelmaan liittyvä niin sanottu nollahypoteesi (nollahypoteesi H_0 edustaa sitä, ettei tutkittavien muuttujien, tekijöiden tai ryhmien välillä esiinny yhteyttä) hylätään, vaikka se onkin tosi. Mitä lähempänä havaittu merkitsevyystaso p on nollaa (alittaa yleisesti käytetyn/sovitun riskitason 0,05), sitä huonommin havaintoaineisto on sopusoinnussa nollahypoteesin kanssa, jolloin se hylätään. P-arvon liittyvä tulkinta onkin periaatteessa niin sanottu nollahypoteesin merkitsevyystestaukseen (NHST) perustuva menettely. Yleisemmin p-arvotarkastelu liittyy määrällisessä ei-kokeellisessa tutkimuksessa kuitenkin eri tekijöiden välisen riippuvuuden tarkasteluun ja todentamiseen. Hypoteesiasetelma liittyy alkujaan kokeelliseen tutkimusasetelmaan, jossa nollahypoteesille asetetaan vastahypoteesi H_1, jonka mukaan tutkittavilla tekijöillä on yhteyttä tai ryhmien välillä on eroa. Analyysin perusteella tutkija sitten tekee päätöksen, voidaanko nollahypoteesi hylätä vai jääkö se voimaan. Tavallisesti kuitenkin esimerkiksi yhteiskuntatieteellisissä tutkimuksissa tilastollinen merkitsevyystulkinta liittyy ennen kaikkea siihen, millä riskitasolla otoksessa havaittuja tuloksia voidaan yleistää koskemaan perusjoukkoa. Tilastollisten testien lisäksi merkitsevyystaso otetaan huomioon myös luottamusvälien määrittämisessä eli väliestimoinnissa. Siinä tutkija itse asettaa merkitsevyystason ja valinta vaikuttaa muodostettavan luottamusvälin pituuteen.

On hyvä huomata, että tilastollinen merkitsevyys ei itse asiassa ilmaise muuttujien välillä havaittujen erojen tai samanlaisuuden reaalista merkitystä, eikä myöskään kuvaa suoraan näiden välisen yhteyden suuruutta tai voimaa. P-arvotulkinnassa on siis ennen kaikkea kyse siitä, että kuinka iso riski sille on, että otoksesta tehdään virhepäätelmä perusjoukon suhteen. Merkitsevyystasoa ilmaiseva tunnusluku ilmaistaan taulukoissa ja tieteellisessä tekstissä yleensä p- tai sig-merkillä. P on lyhenne sanasta probability (todennäköisyys) ja sig. on puolestaan lyhenne sanasta significance (merkitsevyys). SPSS-tulosteissa p-arvo ilmaistaan jälkimmäisellä lyhenteellä. P-arvo on siis todennäköisyysarvo. Tämä perustuu aineistosta laskettuun testisuureeseen sekä kyseisen testin taustalla olevaan niin sanottuun todennäköisyysjakaumaan. Tällaisia satunnaisilmiön todennäköisyyttä kuvaavia jakaumia ovat mm. normaalijakauma (standardoitu), t-jakauma, F-jakauma ja khiin neliö eli χ^2-jakauma. Tilastolliset testit on yleensä myös nimetty tämän taustalla olevan jakauman mukaan esimerkiksi t- tai F-testeiksi. Toisissa testeissä muun muassa otoskoko, vertailtavien ryhmien lukumäärä tai ristiintaulukon dimensiot vaikuttavat analyysien tuloksiin. Näiden seikkojen perusteella määräytyy testeissä tulostuva vapausasteluku (df eli degrees of freedom), joka myös osaltaan vaikuttaa testin tuloksena saatavaan p-arvoon. Tutkimusasetelmia suunniteltaessakin on tämän vuoksi tärkeää, että tutkija ottaa huomioon näiden merkityksen esimerkiksi analyysimenetelmien valintaa ja tutkimusasetelmaansa pohtiessaan.

Tilastolliseen merkitsevyyteen liittyvä raportointi vaihtelee hieman eri tieteenaloilla ja eri tieteellisissä julkaisuissa. Aiemmin, kun testien tulos perustui taulukoitujen todennäköisyysjakaumien käyttöön, tässä yhteydessä käytettiin raportoinnissa yleisesti tiettyjä riskitasojen rajoja, joiden avulla todennettiin saatujen tulosten tilastollista merkitsevyyttä ($p < 0{,}05$ = havaittu p-arvo on tilastollisesti melkein merkitsevä, $p < 0{,}01$ = havaittu p-arvo on tilastollisesti merkitsevä ja $p < 0{,}001$ = havaittu p-arvo on tilastollisesti erittäin merkitsevä). Toisinaan tilastolliseen päättelyyn liittyvä merkitsevyystaso on ilmaistu tähtisymbolein (yksi tähti *, kun $p < 0{,}05$, kaksi tähteä **, kun $p < 0{,}01$, kolme tähteä ***, kun $p < 0{,}001$). Nykyiset tilasto-ohjelmat antavat näiden sijaan tarkan p-arvon, jonka raportointi onkin yllä esitettyjen raja-arvojen sijaan nykyisin yleisin ja suositeltavinkin tapa esittää testin tuottamaa merkitsevyyttä. Joskus merkitsevyystestillä saadaan niin pieni p-arvo, että ohjelmistossa se tulostuu kolmen desimaalin tarkkuudella muodossa $p = 0{,}000$. Tämä ei kuitenkaan tarkoita, että riskiä ei olisi lainkaan, vaan tällöin se on vain hyvin pieni. Näissä tapauksissa p-arvo kannattaakin esittää muodossa $p < 0{,}001$, vaikka muuten raportoinnissa pyrittäisiin esittämään tarkat p-arvot. Toisinaan esitystapana käytetään näissä tilanteissa desimaalilukujen pyöristyssääntöihin perustuen muotoa $p < 0{,}0005$.

Yleisesti p-arvon riskitasokriteerinä käytetään 0,05:n tasoa, toisin sanoen tätä alemmalla tasolla testin tulosta pidetään edellä mainitulla tavalla tilastollisesti merkitsevänä. Toisin ilmaistuna, kun todetaan esimerkiksi joidenkin vertailtavien osaryhmien välisen eron olevan voimassa myös tutkimuksen populaatiossa $p < 0{,}05$:n tasolla, tämä merkitsee sitä, että riski siitä, että tehty päätelmä on virheellinen, on 1:20:stä, $p < 0{,}01$:n tasolla tämä suhdeluku on 1:100 ja $p < 0{,}001$:n tasolla 1:1000. Toisin sanoen, mitä pienempi p-arvo on, sen suuremmalla varmuudella tutkija voi tulkita aineistossaan esiintyvien erojen esiintyvän myös tutkimuksen perusjoukossa. Tosin on tärkeää huomata, että tulos, jossa p-arvo ylittää kriteeriarvon 0,05, on sekin tärkeä tulos. Monesti nämä tulokset jäävät tutkijoilta usein vähemmälle huomiolle tai kokonaan tulkitsematta. Testituloksen raportoinnissa on tarkan p-arvon lisäksi tapana esittää myös testiin liittyvät testisuure- ja vapausasteluku. Näin lukijalle käy selväksi, mitä testiä on käytetty ja millä testiin liittyvillä arvoilla kyseinen p-arvo on saatu (ks. esimerkki testituloksen raportoinnista t-testin tilanteessa sivulla 123).

P-arvon perusteella tehtävään kategoriseen tulkintaan liittyy monia ongelmia. Yksi näistä, johon jo edellä on viitattukin, on tähän liittyvä jyrkkä on–ei -dikotomia-asetelma eli ajatus siitä, että H_0-hypoteesi tulee automaattisesti hylätä, jos analyysin p-arvosuure alittaa asetetun kriteeriarvon. P-arvotulkintojen kritisoijien mukaan tämän käyttö tutkimusraporteissa on usein aivan liian mekaanista, mikä saa helposti tutkijat ylitulkitsemaan tutkimustuloksiaan. Toinen yleisesti esiintyvä väärinkäsitys p-arvojen tulkinnassa ja raportoinnissa on, että tutkija saattaa virheellisesti päätellä, että mitä pienempi p-arvo on, sitä voimakkaampi on tarkasteltava efekti, kuten esimerkiksi ryhmien välinen ero tai asioiden välinen yhteys. Tätä ei pidä p-arvon perusteella arvioida, vaan tilanteeseen soveltuvalla efektikokoa mittaavalla tunnusluvulla. Näiden lisäksi, osin edellisiin liittyen, p-arvotulkinnan yksi suuri ongelma on myös se, että se on herkkä otoskoon vaikutuksille. (Ks. esim. Nummenmaa 2005; Cumming & Calin-Jageman 2017; Wasserstein ym. 2019.) P-arvotarkastelussa erityisesti suurten aineistojen yhteydessä hyvin pienikin ero tarkasteltavien muuttujien tai ryhmien välillä tulee liian herkästi tilastollisesti merkitseväksi, ja päinvastoin, pienissä otoksissa melko vahvakin yhteys voi testeissä leimautua ei-merkitseväksi. Molemmissa tapauksissa tutkija voi tehdä hätiköityjä ja vääriäkin tulkintoja. Esimerkiksi se, ettei khiin neliö -testin mukaan ryhmien jakaumissa ole tilastollisesti merkitsevää eroa, ei merkitse välttämättä sitä, etteikö havaittu ero, voisi olla tärkeä konkreettisen ilmiön kannalta. Ehkä juuri tästä syystä jotkut näyttävät hyväksyvän merkitsevyystasoksi myös 10 prosentin riskitason, jolloin voidaan puhua tilastollisesti oireellisista eroista – joissakin tapauksissa tällainen kriteerien venyttäminen lienee hyvinkin hyväksyttävää. Tällaisten suhteiden varmistaminen vaatii kuitenkin aina lisätutkimuksia tai -evidenssiä esimerkiksi muiden tutkimusten tuloksista tai asiaa koskevista teorioista. Yleensäkään tilastollista merkitsevyystasoa ei saisi – kuten nykyään paljon tutkimuksissa näkee tehtävän – yhdistää mekaanisesti havaittujen erojen (tai yleisesti saatujen tulosten) sisällölliseen merkitykseen, koska tällöin on vaarana, että tulkitsee esimerkiksi eri tekijöiden ”näennäiset” (sisällöllisesti varsin mitättömät) yhteydet todellisiksi yhteyksiksi.

Yllä olevilla varoituksen sanoilla halutaan tähdentää sitä, että tutkijan on aina tehtävä lopullinen ratkaisu siitä, onko saaduilla tuloksilla todellista merkitystä tutkittavan ilmiön suhteen. Pelkkä p-arvo ei yksistään tätä kerro, eikä myöskään sitä, kuinka voimakasta tutkittavan tekijän vaikutus tai esimerkiksi muuttujien välinen yhteys on. Tilastotieteessä tähän on kehitetty indikaattoreita ja lisämenettelyjä tutkijoiden apuvälineiksi. Näiden avulla tutkijat saavat tilastolliseen päättelyynsä lisää tulkintaperusteita, jotka tekevät tulkinnoista ainakin astetta uskottavampia ja luotettavampia. P-arvoon perustuvan hypoteesintestauksen lisäksi tai jopa vaihtoehtoiseksi lähestymistavaksi sille suositellaankin nykyisin efektikokoa kuvaavien mittalukujen sekä luottamusväli-estimoinnin (confidence intervals, CI) käyttöä ja tarkastelua. Seuraavaksi käydäänkin lyhyesti läpi tällaisia p-arvotarkastelua täydentäviä tai vaihtoehtoja tälle tarjoavia menetelmiä.

1.6.1.2 Efektikoon arviointi tilastollisen päättelyn tukena

Edellisessä luvussa kuvattu tilastollinen hypoteesien testaaminen merkitsevyystasotulkintoineen on perinteisesti ollut pääasiallinen menetelmä tilastollisessa päättelyssä. Tällaisen p-arvon tulkintaan perustuvan päätöksenteon lisäksi tilastotieteessä on kehitelty menetelmiä, joilla voidaan selvittää ja arvioida tutkittavan efektin (tutkittavan tekijän vaikutus tai yhteys) tilastollisen merkitsevyyden lisäksi efektin kokoa (suuruutta/voimakkuutta). Tähän kehitetyillä efektikokoa mittaavilla tunnusluvuilla voidaan arvioida esimerkiksi osaryhmien välisten keskiarvoerojen suuruutta tai muuttujien välisten riippuvuuksien voimakkuutta. Ajatellaan esimerkiksi tilannetta (tätä käydään tarkemmin läpi luvussa 4.1.2), että nuorten naisten ja miesten välillä on havaittu tilastollisesti merkitsevä ero lukion opintomenestyksessä, eli ryhmien välisessä testissä on saatu p-arvoksi 0,05 pienempi arvo. Tuloksesta voidaan päätellä, että ryhmät eroavat toisistaan, mutta se ei kerro mitään siitä, miten voimakas efekti sukupuolella on opintomenestykseen, toisin sanoen miten suurta miesten ja naisten välinen ero on. Kun merkitsevyystestiin liittyvään tulkintaan lisätään efektikokoa ilmaiseva suure, joka kuvaa havaitun eron suuruutta/voimakkuutta/kokoa, saadaan tulkinnasta astetta luotettavampaa ja moniulotteisempi.

Sen lisäksi, että efektikoon arviointi täydentää p-arvon tuottamaa tulosta, efektikoko-mittojen käyttö mahdollistaa eri tutkimuksissa havaittujen yhteyksien vertaamisen toisiinsa. Tämä antaa hyvän työkalun esimerkiksi erilaisten meta-analyysien tekijöille. Toki eri tutkimusten tutkimusasetelmiin ja tutkimusaineiston keruuseen liittyvät seikat vaikuttavat saatuihin efektiarvoihin, joten aivan ongelmatonta näidenkään pohjalta tehty vertailu ei ole. Efektikokoa voidaan arvioida tilanteeseen soveltuvalla mittaluvulla yhtä lailla frekvensseihin, keskiarvoihin, järjestyslukuihin tai korrelaatioihin perustuvissa analyyseissä. Tällaisia efektikokoa mittaavia tunnuslukuja ovat mm. d, r, V, η^2, $\omega 2$, r^2 ja monia muita. Lisäksi monet eri analyysimenetelmien tuottamat suureet, kuten regressioanalyysiin liittyvä standardoitu regressiokerroin tai logistisen regressioanalyysin odds-ratio, voidaan tulkita efektikoko-mittoina.

Efektikoon mittaamiseen on olemassa eri menetelmiä (ks. asetelma 5, s. 49). Nämä voidaan jakaa yleisesti kahteen pääryhmään: 1) Toisen pääryhmän muodostavat mittaluvut, jotka skaalattuna mittaavat keskiarvoerojen suuruutta. Mittaluvut skaalataan, jotta eron suuruus olisi helpompi tulkita yhtenäisellä tavalla ja skaalaaminen perustuu jollakin tavoin muodostettuun keskihajontaan (efektikoko-mitasta riippuen). Tähän ryhmään kuuluvien indeksien sanotaan yleisesti kuuluvan niin sanottuun *d*-perheeseen, ja tällaisia ovat esimerkiksi Cohen's d, Glass's Δ ja Hedges' g. 2) Toiseen ryhmään, eli niin sanottuun *r*-perheeseen, kuuluvat puolestaan tunnusluvut, jotka mittaavat muuttujien tai mittauskertojen välisen yhteyden voimakkuutta korrelaatiokerrointa ja sen tulkintaa hyödyntäen. Lisäksi tähän ryhmään voidaan väljästi tulkiten katsoa kuuluvan myös tunnuslukujen, jotka perustuvat selitysosuuden mittaamiseen. Yleisimmin käytettyjä r-ryhmän tunnuslukuja ovat r, r_s, r_{pb}, η, ω, R^2, η^2, $\eta_p^{\ 2}$ ja ω^2. Näiden lisäksi voidaan mainita kategoristen muuttujien analyyseihin liittyvät, riippuvuuden tai yhteyden voimakkuutta mittaavat tunnusluvut: ristiintaulukoinnin yhteydessä käytettävät V, C, τ tai φ, sekä logistiseen regressioanalyysiin liittyvä OR (Odds-Ratio eli vetosuhde). Tunnuslukuja on toki paljon muitakin, mutta tässä edellä on esitelty joukko yleisimmin käytettyjä. Näitä d- ja r-perheen mittalukuja voidaan myös siksi pitää keskeisinä, että monet muutkin tunnusluvut ovat sopivalla muunnoksella tulkittavissa d- tai r-estimaattien tapaan.

Yleisin d-perheen efektikoko-tunnusluku on Cohenin d, sitä käytetään arvioitaessa kahden ryhmän keskiarvojen eron suuruutta tai voimakkuutta. Keskiarvoeron tilastollista merkitsevyyttä selvitettäisiin tässä tapauksessa tyypillisesti riippumattomien ryhmien t-testillä. Tämän lisäksi eron suuruutta on siis syytä arvioida efektikoko-suuretta käyttäen. Yksinkertaisin tapa kuvata eron suuruutta olisi vain raportoida ryhmien keskiarvojen välinen ero (ka_1-ka_2), jonka suuruutta koetettaisiin tulkita sisällöllisesti muuttujan havaintoarvojen skaalaa hyödyntäen. Tällaisessa varsin alkeellisessa efektikoko-arvioinnissa ongelmaksi voisi muodostua se, että efektikoon suuruusluokka jäisi pelkästään oman tulkinnan varaan, eikä sitä todennäköisesti voitaisi suhteuttaa mihinkään yleisesti käytettyihin kriteereihin. Lisäksi saatua tulosta voisi olla hankala verrata aiempiin aiheesta toteutettuihin tutkimustuloksiin. Tästä johtuen efektikoko-tunnusluku on käytännöllistä skaalata niin, että se tuottaa vertailukelpoisia tuloksia. D-perheen tunnusluvuissa skaalaus tehdään sopivaa hajontalukua käyttäen. Cohenin d:ssä skaalaus tehdään otoksesta laskettua, vertailtavien ryhmien yhdistettyä keskihajontaa käyttäen:

$$d = \frac{ka1 - ka2}{sp}$$

Lausekkeessa ka_1 ja ka_2 ovat vertailtavien ryhmien keskiarvot ja s_p on molemmille ryhmille yhteisesti laskettu niin sanottu yhdistetty (pooled) keskihajonta (ks. esimerkiksi Field 2018, 115). Mikäli ryhmien keskihajonnat ovat yhtä suuret, käytetään kaavassa tätä. Mitä suurempi efektikokoarvo d on, sitä suurempi on tutkittavien ryhmien keskiarvojen ero. Cohen (1988) on itse esittänyt d-suureen tulkintaan suuntaa-antaviksi raja-arvoiksi 0,20 – 0,50 – 0,80 siten, että näissä rajakohdissa efektikoko voitaisiin tulkita vastaavassa järjestyksessä pieni – keskisuuri – suuri (ks. myös asetelma 5, s. 49). Muitakin kriteereitä on annettu, muun muassa neljätasoinen luokitus, jossa efektikoko tulkittaisiin pieneksi, keskikokoiseksi, suureksi tai erittäin suureksi. Yleisesti käytettyjä raja-arvoja hyödyntäen efektikokoa on helppo arvioida ja verrata aiempiin tutkimuksiin, joskin Cohen on myöhemmin korostanut, että tulkinnassa raja-arvot ovat vain suuntaa-antavia ja niitä käytetään monesti liian suoraviivaisesti.

Cohenin d on yleisimmin käytetty keskiarvoerojen mittaamiseen käytetty skaalattu tunnusluku. On kuitenkin osoitettu, että se tietyissä tilanteissa yliarvioi populaation efektikokoa. Erityisesti pienehköissä tai pienissä aineistoissa toinen d-perheen tunnusluku Hedge’s g tuottaa harhattoman efektikoko-arvion. Siinä Cohenin d-arvo kerrotaan korjauskertoimella, jolloin g:n arvoksi saadaan pienissä otoksissa hieman d:tä pienempi lukuarvo. Tulkinnassa voidaan käyttää samoja edellä mainittuja raja-arvoja kuin Cohenin d:lle on esitetty.

Edellä esiteltiin keskiarvojen eroihin liittyviä d-perheen tunnuslukuja. Toinen efektikoko-mittojen pääryhmä on muuttujien välisten yhteyksien voimakkuutta mittaavat r-perheen tunnusluvut, kuten edellä jo mainittiinkin. Nämä perustuvat korrelaatiokertoimen käyttöön. Määrällisten muuttujien tilanteessa yhteyden voimakkuutta voidaan mitata tavallisella Pearsonin korrelaatiokertoimella (r). Mikäli toinen muuttujista on kaksiluokkainen kategorinen, niin efektikokoa mittaavana suureena voidaan käyttää piste-biseriaalista korrelaatiota (r_{pb}). Järjestysasteikollisille muuttujille r-suureeksi soveltuu Spearmanin korrelaatiokerroin. Korrelaatiokerroin soveltuu hyvin efektikoon tunnusluvuksi, koska se on skaalattu itseisarvoltaan välille 0–1 ja lukuarvon suuruus kuvaa yhteyden voimakkuutta. Raja-arvoja r-suureiden tulkintaa on esitetty asetelmassa 5 (s. 49).

D-perheen tunnuslukuja voidaan esittää myös r-perheen suureita käyttäen ja päinvastoin. Esimerkiksi edellä esitettiin ryhmien keskiarvoerojen suuruuden arviointia Cohenin d:n avulla. Vastaavaa tarkastelua voitaisiin toteuttaa r_{pb}:n avulla. Yksi suositus onkin, että efektikoon arviointi pyrittäisiin tekemään yhdenmukaisesti vain r-perheen tunnuslukujen avulla. Silloin tulkinta ja raportointi olisivat selkeätä, koska eri tilanteissa voitaisiin käyttää r-tunnuslukujen raja-arvoja ja lisäksi tulokset olisivat keskenään vertailukelpoisia. (Nummenmaa 2005 ja Field 2017). R-perheen tunnusluvuilla on myös etuna se, että laskettu r-tunnusluku korotettuna toiseen on ymmärrettävissä selitysosuutena r^2 tai R^2. Selitysasteen tulkinta liittyy erityisesti tilanteisiin, joissa yksi tarkasteltavista muuttujista on selitettävän muuttujan roolissa, jolloin selitysaste kuvaa prosenttiosuutta, jonka toinen muuttuja tai toiset muuttujat kykenevät selittämään selitettävän muuttujan vaihtelusta. Tällaisia vaihtelua selittäviä efektikoon tunnuslukuja ovat myös varianssianalyysien yhteyteen soveltuvat η^2, ω^2, $\eta_p{}^2$ tai $\omega_p{}^2$. Nämä tarkastelut sopivat siis tutkimusasetelmiin, joissa selitettävä muuttuja on mitattu vähintään välimatka asteikolla, esimerkiksi eri oppiaineiden arvosanat tai vaikkapa ansiotulot, ja yksi tai useampi selittäjä eli riippumaton muuttuja ovat kategorisia, esimerkiksi sukupuoli tai oppilaiden viriketaustaluokitus. Tunnusluvut η^2, ω^2, $\eta_p{}^2$ tai $\omega_p{}^2$ perustuvat varianssianalyysissä muodostettaviin eri vaihtelujen lähteitä mittaaviin neliösummiin. Tunnusluvuilla saadaan arvioitua päävaikutusten efektikokoa niin, että ne kuvaavat vaihtelun osuutta kokonaisvaihtelusta, jolloin ne saavat arvoja väliltä 0-1. Ensiksi mainittu η^2 soveltuu yhden selittävän tekijän tilanteisiin ja se perustuu varianssianalyysin neliösummahajotelmaan:

$$\eta 2 = \frac{SSe}{SSt}$$

Lausekkeessa SSE on tutkittavaan efektiin liittyvä neliösumma, ja SST on kokonaisneliösumma. Arvot saataisiin varianssianalyysin ANOVA-taulusta. On osoitettavissa, että η^2 tuottaa lievästi harhaisen tuloksen efektikoolle. Tunnusluvussa ω^2 tämä harha on korjattu laskukaavaan sijoitetuilla lisätermeillä. Omega toiseen, eli ω^2 on siis hieman monimutkaisempi laskea, mutta harhattomuuden mielessä suositeltavampi efektikoon tunnusluku yksisuuntaisen varianssianalyysin tilanteisiin. Tulkinnassa voidaan näillä molemmilla tunnusluvuilla käyttää Cohenin esittämiä raja-arvoja, jotka on esitetty kappaleen loppuosan taulukossa (asetelma 5). Efektikoon arvioinnissa on myös hyvä huomata, että tunnuslukujen η^2 ja ω^2 kaltaisten, selitysosuutta kuvaavien tunnuslukujen tulkinta saattaa olla havainnollisempaa, kun saadusta lukuarvosta otetaan neliöjuuri. Näin saatu suure on tulkittavissa r-tunnusluvun tapaan siihen liittyvien raja-arvojen mukaisesti (0,10 – 0,30 – 0, 50: pieni – keskisuuri – suuri). Tilanteissa, joissa varianssianalyysin mallissa on useampia selittäviä tekijöitä, efektikokoa mittaavina tunnuslukuina voidaan käyttää eri tekijöiden erillisten vaikutusten efektikokoa mittaavia tunnuslukuja η_p^2 tai ω_p^2, jotka perustuvat osittaisten selitysosuuksien arvoihin (partial eta squared ja partial omega squared). Lukuarvo kuvaa prosenteissa sitä, miten paljon yksittäinen tekijä selittää vaihtelua, mikä jää selittämättä muilta analyysissä olevilta tekijöiltä. Näiden tulkintaan ei ole esitetty lähdekirjallisuudessa selkeitä viitearvoja, joten saatua tulosta on koetettava arvioida sisällöllisin perustein, tutkittavana olevaan ilmiöön liittyen. Toisinaan näitä osittaisia selitysasteita näkee tulkittavan edellä mainitun η^2 suureelle esitettyjen raja-arvojen avulla, mutta tällaisen arvioinnin käyttökelpoisuudesta ei tuntuisi olevan yksimielisyyttä.

Edellä esitettyjen d- ja r-perheiden efektikokoa mittaavien tunnuslukujen lisäksi eri analyyseissä saadaan tulosteina estimaatteja, kertoimia tai suureita, joita myös voidaan käyttää ja tulkita efektikoon mittoina (osa näistäkin tosin olisi sijoitettavissa d- tai r-perheen tunnuslukuihin). Tällaisina voidaan pitää esimerkiksi regressioanalyysissä mallin standardoitua regressiokerrointa, logistisessa regressioanalyysissä mallin tuottamia OR-vetosuhteita tai vaikkapa ristiintaulukoinnissa riippuvuuden voimakkuutta mittaavaa Cramerin V:tä. Näitä on käsitelty myöhemmissä kappaleissa aina kyseisen analyysimenetelmän yhteydessä.

SPSS-ohjelmassa ei valitettavasti ole erityisen kattavasti tarjolla efektikoko-tarkasteluja eri tilanteisiin. Korrelaatiokertoimet toki saa muodostettua ja varianssianalyysin sekä ristiintaulukointien proseduureissa efektikokoa mittaavia tunnuslukuja saa valittua lisäoptioista (tunnusluvut η^2, η_p^2 ja V, φ). Muuten arvot on laskettava käsin sopivaa laskukaavaa käyttäen analyysien tulosteista saatavilla luvuilla. Usein itse laskutoimitukset ovat varsin helppoja toteuttaa. Lisäksi verkosta on löydettävissä paljon valmiita laskureita, joihin syötetään tarvittavat lukuarvot ja laskuri tuottaa tuloksena efektikoon estimaatin. Tällaisia voi helposti etsiä netin hakukoneella vaikkapa hakusanalla 'effect size calculator'. Kirjan myöhemmissä luvuissa eri analyysien ja esimerkkien yhteydessä palataan uudestaan ja yksityiskohtaisemmin efektikoon arviointiin eri tilanteissa ja paikoin myös sen laskuperiaatteisiin. Yleisajatuksena on se, että I) tunnuslukutietojen ja II) tilastollisen testituloksen lisäksi raportoinnissa olisi hyvä olla näkyvissä myös III) tutkittavaan tilanteeseen soveltuva efektikokoa kuvaava suure (kuten myös seuraavassa kappaleessa esiteltävä IV luottamusvälin raportointi).

Efektikokotarkastelu ei luonnollisestikaan poista kokonaan tulkintaan liittyviä epävarmuustekijöitä. Efektikokoa mittaavia tunnuslukuja käyttäessä tutkija, ja yleisemmin tutkijayhteisö, joutuu ratkaisemaan, mitä saadut efektikokoarvot oikeastaan kertovat tutkittavasta ilmiöstä ja esimerkiksi muuttujiin liittyvistä ryhmäeroista ja yhteyksistä. Lisäksi on tärkeää huomata, ettei

Analyysi	Efektikokoa mittaava tunnusluku	Efektikoko pieni/heikko	 keskisuuri	 suuri/voimakas
Riippumattomien ryhmien keskiarvojen vertailu	Cohenin d Glassin Δ Hedgen g	0,20	0,50	0,80
Suhteellisten osuuksien vertailu	Cohenin g	0,05	0,15	0,25
Korrelaatio[a]	Pearsonin r Järjestyskorrelaatio r_s Piste-biseriaalinen kerroin r_{pb}	0,10	0,30	0,50
Ristiintaulukointi	Cramerin V Phii-kerroin φ Kontigenssikerroin C	0,10	0,30	0,50
ANOVA	Eta toiseen η^2 Omega toiseen (harhaton) ω^2	0,01	0,06	0,14
Regressioanalyysi	Selitysaste (yksi selittäjä) r^2 Selitysaste (useita selittäjiä) R^2	0,01 0,02	0,09 0,13	0,25 0,26

Lähde: Ellis 2010, 41. Huom. taulukossa esitettyjä raja-arvoja ei ole tarkoitus noudattaa orjallisesti, vaan ne on laadittu enemmänkin suuntaa antaviksi ehdotuksiksi, joiden avulla on mahdollista helpottaa tunnuslukujen tulkintaa.

[a] korrelatitivisia, yhteyttä mittaavia efektikoon tunnuslukuja. Soveltuva tunnusluku määräytyy tarkasteltavien muuttujien mitta-asteikon perusteella. Lisäksi moni muukin efektikoon arviointi on palautettavissa sopivalla muunnoksella r-suureeksi, jonka suuruutta voidaan tulkita tässä esitettyjen raja-arvojen avulla.

ASETELMA 5. Suosituksia eräiden efektikokoa mittaavien tunnuslukujen tulkintaan.

tunnusluvun lukuarvoon perustuvalla efektikokotarkastelulla voida päätellä sitä, kuinka todennäköisesti vastaava efekti havaittaisiin toisessa riippumattomassa aineistossa. Tämän vuoksi monet tieteelliset julkaisusarjat kannustavat raportoimaan efektikoko-suureille myös luottamusvälin, jolla saadaan kuvattua efektikoko-arvioinnin tarkkuutta. (Nummenmaa 2005.) Seuraavassa alaluvussa avataankin yleisemmin luottamusvälitulkinnan peruslähtökohtia.

1.6.1.3 Luottamusvälit aineiston tulkinnan perustukivälineeksi

Edellä on käsitelty tilastollisia testejä, joiden käyttö liittyy ennen kaikkea siihen, kuinka luotettavasti otoksesta saadut tulokset olisivat yleistettävissä tutkimuksen taustalla olevaan populaatioon, ja siihen, millaisen riskitason tutkija (ja tutkijayhteisö yleisemmin) tässä suhteessa hyväksyvät. Edellä esitettyjen p-arvoon perustuvan merkitsevyystasotulkinnan sekä efektin voimakkuutta kuvaavien mittalukujen lisäksi tilastollista päättelyä voidaan toteuttaa tai täydentää luottamusvälien avulla. Näiden käytön merkitystä tilastollisessa päättelyssä onkin viime vuosina korostettu ”Uusi tilastotiede” -lähestymistavan piirissä. Tilastollisten testien tapaan luottamusvälitulkinnassakin kyse on siitä, kuinka luotettavasti saadut tulokset kuvaavat populaation vastaavia ominaisuuksia (ks. tietolaari 3). Kuten nollahypoteesien testaamisessa (NHST), tässäkin pyritään huomioimaan otannasta aiheutuva epävarmuus, jota kontrolloidaan merkitsevyystason valinnalla. Yleisimmin tässä käytetään 5 %:n merkitsevyystasoa, jolloin väliestimoinnissa voidaan puhua 95 %:n luottamustasosta ja edelleen 95 %:n luottamusvälin määrittämisestä.

Luottamusvälitarkastelun avulla muodostetaan yksittäisen piste-estimaatin sijaan lukuarvoista muodostuva väli, jolla tutkimuksen kohteena oleva, populaation arvioitavana oleva suure riittävän suurella luotettavuudella sijaitsee. Väli muodostetaan niin, että otoksesta laskettuun piste-estimaattiin lisätään ja siitä vähennetään otannasta johtuva virhemarginaali. Luottamusvälitarkastelun perusajatus kulkee seuraavaan tapaan: kun luottamustasoksi on valittu esimerkiksi 95 %, saadut luottamusvälin ala- ja ylärajat kuvaavat sitä, että jos vastaavanlaiseen otokseen perustuva arviointi toteutettaisiin esimerkiksi 100 kertaa, näistä 95 kertaa todellinen populaation arvioitavana oleva suure asettuisi saadulle luottamusvälille, eli ala- ja ylärajojen väliin. Toisin sanoen vain 5 %:ssa tapauksista arvo asettuisi kyseisen välin ulkopuolelle. (Ks. esim. van de Schoot ym. 2013). Sama voitaisiin myös ilmaista seuraavasti: on uskottavaa (plausible), että otos on peräisin populaatiosta, jossa populaation arvioitavana oleva suure on jotain väliltä [alaraja, yläraja]. Jos tulos halutaan ilmaista tätä suuremmalla varmuudella, valitaan yleisimmin korkeammaksi luottamustasoksi 99 prosenttia, jolloin luottamusväli muodostuu leveämmäksi (pidemmäksi, eli ala- ja ylärajan välinen etäisyys on suurempi verrattuna 95 prosentin tulkintatasoon).

Monille tuttu esimerkki virhemarginaalista ja luottamusvälistä on otokseen perustuva puolueiden kannatusmittaus eli gallup-kysely. Siinä tavoitteena on arvioida puolueen todellista kannatusta eli kannatusprosenttia populaatiossa. Otoksesta saadun kannatusprosentin (piste-estimaatti, esimerkiksi 19 prosenttia) lisäksi tulosten yhteydessä ilmoitetaan nykyisin ilahduttavan usein myös kyselyyn liittyvä virhemarginaali. Tämä voisi olla 95 %:n luottamustasolla laskettuna esimerkiksi kaksi prosenttiyksikköä suuntaansa. Näillä tiedoilla kyse onkin juuri väliestimoinnista: sen sijaan että ilmoittaisimme kannatusprosentiksi 19 %, ilmoitamme kannatusprosentin olevan 95 % luotettavuudella jotain väliltä (17 % , 21 %). Luottamusvälin käytön perusajatus lähtee siitä, että tällainen ”väljempi” tulkinta kuvaisi todellista tilannetta ja arviointiin liittyvää epätarkkuutta tutkittavassa populaatiossa paremmin tai osuvammin kuin perinteinen yksittäiseen (näennäisen tarkkaan) prosenttiosuuteen tai keskiarvoon perustuva tulkinta.

Eräs tärkeä etu luottamusvälien käytössä yksittäiseen piste-estimaattiin verrattuna on, että luottamusväli antaa populaation tuntemattoman arvon arvioinnin lisäksi tietoa arvioinnin tarkkuudesta. Mitä lyhempi luottamusväli on, sitä tarkempaa on tutkittavan parametrin estimointi. Toisin sanoen sitä tarkempaa tilastollista päättelyä kyetään tekemään otoksesta populaatioon. Luottamusvälin pituuteen vaikuttavat 1) käytetty luottamustaso: suurempi luotettavuus saa aikaan pidemmän luottamusvälin. Ja vastaavasti arvioinnin luotettavuudesta tinkimällä saadaan luottamusväliä kavennettua (silloin toisaalta hyväksytään suurempi virheen riski arvioinnissa); 2) otoskoko: suuremmalla otoksella saavutetaan tarkempi estimointi, eli luottamusväli lyhenee; 3) piste-estimaattorin keskivirheeseen liittyvät muut tekijät, esimerkiksi populaation keskiarvoa arvioitaessa havaintoarvoihin liittyvä keskihajonta, vaikuttavat luottamusvälin pituuteen (kun hajontaa on enemmän, luottamusväli muodostuu pidemmäksi, eli arvioinnin tarkkuus heikkenee). Luottamusvälin pituuden muodostumista voitaisiin havainnollistaa esimerkillä populaation keskiarvon estimoinnista. Silloin luottamusvälin ala- ja yläraja määräytyvät seuraavien kaavojen mukaisesti:

$$alaraja: ka - z * \left(\frac{kh}{neliöjuuri(n)}\right)$$

$$yläraja: ka + z * \left(\frac{kh}{neliöjuuri(n)}\right)$$

Lausekkeissa plus- ja miinusmerkkien jälkeiset osuudet muodostavat virhemarginaalin. Termi z liittyy valittuun luottamustasoon, jonka arvo saadaan normaalijakauman todennäköisyysfunktiosta. Lausekkeiden avulla voidaan nähdä edellä mainittuja eri tekijöiden vaikutuksia luottamusvälin leveyteen. Esimerkiksi otoskoko sijaitsee lausekkeissa nimittäjässä, jolloin sen kasvattaminen pienentää virhemarginaalia, toisin sanoen kaventaa luottamusväliä.

Tilastollisen päättelyn välineenä luottamusvälitarkastelu tarjoaa informatiivisen vaihtoehdon tilastolliselle testaamiselle. Nollahypoteesin testaus tuottaa varsin kategorisen tiedon arvioitavasta asiasta (esimerkiksi ryhmäkeskiarvojen ero populaatiossa), kun taas luottamusvälin avulla nähdään varsin havainnollisesti, mille välille arvioitava asia sijoittuu ja miten suuri epävarmuus arviointiin liittyy (esimerkiksi ryhmäkeskiarvojen eroa kuvaavan luottamusvälin sijoittuminen ja välin pituus). Lisäksi, jos arvo nolla ei sisälly luottamusvälille, voidaan eron katsoa olevan tilastollisesti merkitsevä käytetyllä merkitsevyystasolla. Näin ollen luottamusvälitarkastelulla saadaan tarvittaessa tilastollisen testin tapaan tilastolliseen merkitsevyyteen liittyvä tieto sekä sen lisäksi tieto arvioinnin tarkkuudesta.

Edellä on esitetty lyhyesti esimerkkeinä luottamusvälin muodostaminen prosenttiosuudelle ja keskiarvolle. Vastaavalla periaatteella luottamusvälejä voidaan laskea näiden lisäksi muun muassa keskihajonnoille, variansseille, korrelaatiokertoimille, regressiokertoimille, efektikoko-mitoille ja niin edelleen. Käytännön ratkaisuja tarkastellaan muiden merkitsevyyssuureiden lailla edempänä kirjassa eri menetelmien yhteydessä. Luottamusvälejä voi käyttää joko yksinään estimoinnin ja tilastollisen päättelyn välineenä tai yhdessä muiden tässä luvussa esiteltyjen menettelyjen kanssa. Melko tyypillinen suositus nykyisin on, että tilastollisen testin tuottamaa tulosta täydennetään luottamusvälin raportoinnilla yhdessä tilanteeseen soveltuvan efektikoko-mitan kanssa. Joissain tilanteissa todennäköisyysjakaumaan perustuvaa virhemarginaalia ei ole mahdollista ja järkevää muodostaa yllä mainittuun tapaan. Silloin luottamusvälin laskeminen voidaan toteuttaa niin sanotun bootstrap-toteutuksen avulla. Tätä käsitellään tarkemmin seuraavassa luvussa.

1.6.1.4 Bootstrap-estimointi: lähtökohtia, käyttö ja tulkinta

Edellä esitellyt luottamusväli-estimointi ja luottamusvälin pituuden määräytyminen perustuivat käytetyn estimaattorin, esimerkiksi keskiarvon, tilastomatemaattisiin ominaisuuksiin. Tällöin otoksen poimintaa ei tarvitse lähteä toistamaan lukuisia kertoja, jotta nähtäisiin, miten paljon eri otosten tuottamat keskiarvot vaihtelevat otoksesta toiseen. Sen sijaan keskiarvo-estimaattorin ominaisuuksista on tiedossa, että kun ryhmäkoot ovat riittävän suuria (ryhmillä n on vähintään 20, mieluusti yli 30) ja havaintoarvot noudattavat normaalijakaumaa, luottamusväliin liittyvä virhemarginaali saadaan laskettua suoraan otoksen arvoista, käyttäen keskivirhettä (SPSS-tulosteessa oleva Standard Error, s.e.) ja tilanteeseen soveltuvaa todennäköisyysjakaumaa (keskiarvojen yhteydessä t-jakaumaa). Käytännössä hyvä ja monesti riittävän tarkka arvio 95 %:n luottamusvälin virhemarginaaliksi on 2*s.e., jolloin otoksesta lasketun keskiarvon luottamusvälin ala- ja ylärajat asettuvat lukuarvoihin ka ± 2*s.e. On kuitenkin huomattava, että edellä oleva tieto pätee vain, kun ryhmäkokoihin ja normaalijakautuneisuuteen liittyvät oletukset ovat kunnossa. Mikäli nämä oletukset eivät toteudu tai niiden toteutuminen epäilyttää, luottamusvälejä voidaan muodostaa myös Bootstrap-estimointia käyttäen.

Bootstrap-estimoinnissa hyödynnetään tietokoneen laskentatehoa niin, että tutkimusaineistosta muodostetaan arpoen (takaisinpanolla, eli poimittu alkio palautetaan aina takaisin mukaan arvontaan) uusia otoksia, joista kaikista lasketaan tutkimuksen kannalta kiinnostava otossuure, esimerkiksi otoskeskiarvo. Tällaista uudelleenotantaa (re-sampling) toteutetaan lukuisia, monesti satoja tai tuhansia, kertoja ja lopuksi saaduista keskiarvoista muodostetaan jakauma. Tästä jakaumasta voidaan etsiä jakauman ala- ja ylärajoilta kohdat, joiden väliin rajautuu esimerkiksi 95 % eri otosten keskiarvoista. Tätä muodostunutta väliä voidaan pitää populaation keskiarvon 95 % bootstrap-luottamusvälinä. Populaation keskiarvon arvioinnin tapaan Bootstrap-periaate on sovellettavissa aivan vastaavasti lähes mille tahansa arvioitavalle suureelle, esimerkiksi keskihajonnalle, korrelaatiolle, prosenttiosuudelle ja niin edelleen, myös tilanteissa, joissa tilastollinen testaaminen tai luottamusvälien määrittäminen muuten olisi vaikeata tai mahdotonta. Mikäli SPSS-lisenssi sisältää Bootstrap-modulin, bootstrap-painike löytyy useimpien menetelmien yhteydessä vaihtoehtoisena tilastolliseen päättelyyn liittyvänä välineenä. Aiemmin tätä menettelyä on rajoittanut tietokoneiden laskentakapasiteetti, mutta nykyisin esimerkiksi SPSS:n bootstrap-toiminto tuottaa tuloksen yleensä muutamassa sekunnissa, vaikka uusintaotosten määrä olisi asetettu hyvinkin suureksi. Monesti uusinta- eli bootstrap-otosten määränä käytetään vähintään arvoa 1000, jolloin tutkittavan suureen jakauma alkaa yleensä asettua riittävän stabiiliksi. Katso esimerkki bootstrap-luottamusvälin muodostamisesta (ks. luku 4.1.1.1). Samassa luvussa on myös ohjeistus SPSS-toteutuksesta.

Edellä on esitelty erilaisia menetelmiä, joiden avulla voidaan vahvistaa aineistosta nousevia tilastolliseen päättelyyn liittyviä tulkintoja. Käytännössä ei välttämättä ole tarkoituksenmukaista, että tällaisia osin vaihtoehtoisia, osin toisiaan täydentäviä, menetelmiä pyrittäisiin aina kaikkia käyttämään ja raportoimaan. Näiden käyttöä voit kuitenkin harkita aina tutkimuskysymystesi, tutkimusaineistosi ja muuttujien ominaisuuksien mukaan. Useissa viimeaikaisissa tutkimusraportointiin liittyvissä ohjeistuksissa ja tieteellisten julkaisusarjojen kirjoittajaohjeissa kuitenkin suositellaan, että saatua tulosta ei pitäisi esittää pelkästään yhden tulosta kuvaavan arvioinnin avulla (ks. esimerkiksi APA 2020; JARS website, https:/apastyle.apa.org/jars). Yleisohjeena voisi olla, että tunnuslukutietojen ja tilastollisen testituloksen lisäksi raportoinnissa olisi hyvä olla näkyvissä myös tutkittavaan tilanteeseen soveltuvat 95 %:n luottamusväli sekä efektikokoa kuvaava suure. Eräät julkaisusarjat ohjeistavat lisäksi laskemaan luottamusvälit myös efektikoko-arvoille. Se, miten nämä raporteissa ilmaistaan, esitellään tulevilla sivuilla eri menetelmien käsittelyn yhteydessä.

Tämän tilastollisen päättelyn perusteiden lopuksi täytyy vielä kerran korostaa, että vaikka tilastoanalyyseihin opastavissa kirjoissa, kuten tässäkin teoksessa, annetaan runsaasti erilaisia kriteereitä tilastolliseen päättelyyn luotettavuuden ja uskottavuuden takaamiseksi, on tärkeää muistaa se tosiasia, ettei näille ole kuitenkaan viime kädessä olemassa mitään yleispäteviä ja eksakteja kriteereitä. Nämä eivät myöskään takaa, eivät ainakaan yksinään, tutkimuksissa tehtyjen johtopäätösten ja tulkintojen luotettavauutta. Tilastollinen päättely ei siis yksistään ole riittävä perusta havaittujen ilmiöiden selittämiseen eikä sen todentamiseen, miksi havaittu muutos selittävässä tekijässä (muuttujassa) on yhteydessä selitettävässä tekijässä tapahtuviin muutoksiin. (Esim. Ketokivi 2015, 24, 42–43.) Toisin sanoen tilastollisen analyysin tulokset ovat aina perustaltaan probabilistisia, todennäköisyyksiin perustuvaa tilastollista ja induktiivista tulkintaa (ks. esim. Freedman 2010, 3–11), joka kuitenkin tarvitsee rinnalleen myös yleisempää ja teoreettista tulkintaa. Tilastollinen päättely ei ole yksistään riittävä perusta havaittujen ilmiöiden selittämiseen ja esimerkiksi siihen, miksi havaittu muutos on yhteydessä tai vaikuttaa selitettävään tekijään. Tulosten tulkintaa tulisikin näin jatkaa aineiston tulkinnan kolmanteen, teoreettisempaan, vaiheeseen (asetelma 4).

1.6.2 Tulosten laaja-alaisempi tulkinta on tulkinnassa ”a ja o”

Edellä käsiteltiin lyhyesti tilastolliseen päättelyyn liittyviä tilastollisia menetelmiä, joilla voidaan lisätä tähän liittyvien tulkintojen luotettavuutta ja arvioida tutkittavan efektin voimakkuutta. Tutkimuksessa, erityisesti yhteiskunnallisessa ja käyttäytymistieteellisessä tutkimuksessa, aineiston tulkinta ei pääty kuitenkaan tähän. On pyrittävä myös tulkitsemaan ja selittämään havaittujen yhteyksien ja riippuvuuksien luonnetta, syitä ja niihin liittyviä mekanismeja. Tilastollinen päättely antaa tutkijalle eväitä esimerkiksi sen tulkitsemiseen, että X- ja Y-muuttujien välillä on yhteyttä tai ei ole. Mutta sen selittämiseksi, miksi näin on, tilastolliseen päättelyyn liittyvät mittaluvut ja analyysien tuottamat lukuarvot eivät useinkaan anna suoria vastauksia. Tämän tulkitseminen on mahdollista vain teoreettisen tulkinnan ja aiemman tutkimustiedon perusteella tapahtuvan argumentoinnin ja tulkinnan avulla. Jos tulkinta jätetään pelkästään tilastollisen päättelyn tasolle, esimerkiksi listaksi erilaisia tilastollisia tunnuslukuja, jää tulkinta varsin kapeaksi ja latteaksi empiiriseksi kuvaukseksi ja tulkinnaksi tutkittavasta ilmiöstä. Empiirisessä määrällisessäkin tutkimuksessa tarvitaan siis empiirisen aineiston ja teoreettisen tason välistä tiivistä vuorovaikutusta. (Esim. Jokivuori & Hietala 2007, 11–12; Ketokivi 2015, 17–20, 43; Bhattacherjee 2012, 9, 14.)

Tässä yhteydessä ei ole mahdollista tarkastella tarkemmin yhteiskunta- ja käyttäytymistieteiden selitys- ja tulkintamalleja (ks. esim. Niiniluoto 2002; Toivonen 1999, 20–72; Kaidesoja, Kankainen & Ylikoski 2018; tilastolliseen tutkimukseen liittyvästä päättelyn ja selittämisen perinteistä esim. Ketokivi 2015). On kuitenkin syytä pysähtyä hetkeksi sen kysymyksen ääreen, miksi määrällisessä ja tilastollisesti orientoituneessa yhteiskuntatieteellisessä ja käyttäytymistieteellisessä tutkimuksessa laaja-alaisempi ja teoreettisempi tulkinta on tärkeää. Vastaushan tähän on luonnollisesti lähes itsestäänselvyys. Ensiksikin näin mahdollistuu konkreettisesta tutkimusaineistosta havaittujen jakaumien ja eri tekijöiden välisten yhteyksien laaja-alaisempi ja syvempi tulkinta. Teoreettisen argumentoinnin ja selittämisen avulla voidaan siis vastata siihen, miten, miksi ja missä myös kontekstissa havaitut tekijät liittyvät toisiinsa. Kuten edellä jo mainittiin, teoreettiset argumentit ja tulkinnat antavat havainnoille ja tilastollisen päättelyn tuloksille vankemman selitys- ja tulkintaperustan. Tai toisin sanoen näin annetaan tilastolliselle päättelylle ja siihen liittyville tulkinnoille teoreettinen sisältö ja tulkinta. Tällainen laaja-alaisempi tulkinta on tärkeää siksikin, että näin omat tutkimustulokset ja tulkinnat liittyvät osaksi yleistä tieteellistä diskurssia. (Bhattacherjee 2012, 9, 14; Ketokivi 2015, 17–20, 67) Tutkimuksen teoreettinen viitekehys on määrällisessä ja empiirisessäkin tutkimuksessa siis välttämätön osa tutkimuksen aineiston tarkoituksenmukaista tulkintaa. Tämän avulla tutkija kykenee paremmin jäsentämään ja tulkitsemaan tutkittavana olevaa ilmiötä yleisesti ja omaa empiiristä aineistoaan ja siitä nousevia tuloksiaan. Oman tutkimusalueen monipuolinen ja teoreettinen tuntemus auttaa tutkijaa myös haastamaan ensimmäisiä tulkintojaan, hakemaan näille siis vaihtoehtoisia tulkintoja. Omien tulkintojen haastaminen ja koettelu on yleensäkin yksi keskeisimmistä tutkijan kvalifikaatioista, näin tutkija saa tulkintoihinsa syvyyttä ja laajempaa argumentointipintaa. Näin myös määrällisen tutkimuksen avulla voidaan nostaa esiin uusia, ennen tutkimukselta huomaamatta jääneitä, tulkintoja eri tekijöiden yhteyksistä.

Teoreettinen selittäminen ei tässä yhteydessä tarkoita sitä, että meillä pitää olla aina tulkinnan taustalla muodollinen teoreettinen tulkinta tutkittavasta ilmiöstä tai eri tekijöiden suhteiden luonteesta. Useimmiten tutkimusten teoreettinen viitekehys tai tarkastelukulma rakentuukin aikaisemman tutkimuksen pohjalta tehtyyn omaan kokonaisviitekehykseen ja tulkintaan tutkittavan ilmiön eri tekijöiden suhteista. Luonnollisesti, jos tutkittavaan ilmiöön liittyy spesifejä teorioita, nämä tai näihin liittyvä tulkinta ovat keskeisessä asemassa tutkimuksen viitekehyksessä. Teorian mukaisessa tulkinnassa ei myöskään ole kyse siitä, että oman aineiston tulokset pakotettaisiin tiettyyn teoreettiseen tai jonkin tietyn koulukunnan tulkintaan, tämähän on aina vaarana kun omia tuloksia pyritään tulkitsemaan tällaisessa laajemmassa tulkintakehikossa. Teorioiden ei saa antaa myöskään jäykistää aineiston tulkintaa liiaksi, myös määrällisen tilastollisen tutkimuksen tavoitteena on tuottaa uutta tutkimustietoa, kuten kaiken muunkin tutkimuksen. Juuri tämän vuoksi tutkijan onkin tärkeää tarkastella, ja myös koetella, omia tulkintojaan laaja-alaisemmin. On esimerkiksi tärkeää kysyä itseltään sitä, miksi juuri oma tulkinta on uskottava, millaisia vaihtoehtoisia tulkintoja voisi löytyä. Tulee muistaa, että yhteiskunnallisilla ja ihmisten toimintaan liittyvillä ilmiöillä voi olla, ja onkin, useitakin tarkoituksenmukaisia ja uskottavia selityksiä. (Ks. esim. Kakkuri-Knuutila & Heinlahti 2006, 150–156, 216; Jokivuori & Hietala 2007, 18–24; Ketokiven 2015, 283; Ylikoski 2018). Yksi tutkijan perusominaisuus onkin jatkuva ”epäily”, vaihtoehtoisten tulkintojen etsintä, ja valmius asettaa omatkin tulkinnat uudelleen tarkastelun alaisiksi. Tämän avulla omista tulkinnoista ja niihin liittyvästä argumentoinnista tulee aina monipuolisempia, vahvempia ja myös uskottavampia. Samoin kuin tieteelliset tulkinnat saavat todistusvoimansa uusien havaintojen ja tulkintojen ja niitä koskevan kritiikin vuorovaikutuksessa, yksittäisen tutkimuksen havaintoja ja tulkintoja vahvistaa merkittävästi se, että tutkija tai tutkijat tarkastelevat tietoisesti omaan päättelyyn mahdollisesti liittyviä ”heikkoja lenkkejä” ja koettelevat tulkintojaan esittämällä niille vaihtoehtoisia tulkintoja.

Kaikella edellä olevalla tulkintaan liittyvällä olemme halunneet korostaa erityisesti sitä, 1) että aineistoista nousevien päätelmien varassa ei yksinään kyetä yleensä tuottamaan mielenkiintoista eikä adekvaattia tulkintaa ja 2) sitä, että määrällinen empiirinen ja tilastollinen tut-

kimus sisältää tässä mielessä aina tulkintaa. Tulkinta ei siis ole vain laadullisen tutkimuksen ominaisuus, kuten joskus vieläkin jotkut näyttävät tulkitsevan.

Vielä ennen kuin siirrymme tutkimusaineiston käsittelyyn ja analysointiin sekä SPSS-tilasto-ohjelman hyödyntämiseen, tarkastellaan lyhyesti tutkimuksen tekemiseen liittyviä eettisiä kysymyksiä, ovathan ne tulleet viime vuosikymmenten aikana yhä tärkeämmiksi myös kasvatustieteen ja yhteiskuntatieteiden tapaisilla tutkimusaloilla.

1.7 Eettisyys kaikessa: luotettavan tutkimuksen tavaramerkki

Tutkimuksen eettiset kysymykset liittyvät niin tutkimuksen tekemisen eri vaiheisiin kuin tulosten julkaisemiseen. Eettinen tarkastelu jää helposti sivummalle, vaikka se on kestävän ja hyvän tutkimuksen teon yksi välttämätön ehto. Eettisen tarkastelun pitäisikin lävistää koko tutkimusprosessi, tutkijan omien lähtökohtien ja suunniteltavan tutkimuksen perusteiden arvioimisesta aina tutkimusaineistoa ja aineiston tulkitsemista koskeviin perusratkaisuihin asti. Eikä unohdeta myöskään raportointivaihetta. Se, onko tutkimus laadullinen tai määrällinen tutkimus, ei ole ratkaiseva tekijä tutkimuksen eettisille kriteereille (Piispa 2006). Lähtökohtana on se, että kaikissa tutkimuksissa ja jokaisessa eri tutkimusprosessin vaiheissa näitä tulisi tietoisesti punnita. Tutkimuksen tekemiseen liittyvien eettisten kysymysten tärkeyttä kuvaa sekin, että vuonna 1991 perustettiin Tieteellisten seurain valtuuskunnan organisoimana Tutkimuseettinen neuvottelukunta, joka on Opetus- ja kulttuuriministeriön toimintaelin. Sen tavoitteena on edistää hyvää tieteellistä käytäntöä ja estää tieteellisen vilpin tekoa. Tätä edistääkseen neuvottelukunta on antanut ja antaa edelleenkin hyvää tieteellistä käytäntöä normittavia ohjeistoja, viimeisin ohjeistus on vuodelta 2012. Vuonna 2009 neuvottelukunta suositteli humanistista, yhteiskunta- ja käyttäytymistieteellistä tutkimusta koskevat eettiset periaatteet, jotka päivitettiin vuoden 2019 syksyllä. Näihin jokaisen tutkijan ja myös opiskelijoiden, jotka tekevät tutkimusta ja opinnäytetöihin liittyviä tutkielmia, on tärkeää tutustua. Neuvottelukunnan eri ohjeet löytyvät neuvottelukunnan kotisivuilta (https://www.tenk.fi/fi/tenkin-ohjeistot)

Tieteentekoon liittyvä eettisyystarkastelu ei ole tai ei ainakaan saisi olla mikään tutkimukseen liittyvä erityiskysymys, jota tarkastellaan muodon vuoksi, kun sitä meiltä vaaditaan, vaan se koskettaa tutkimuksenteon kaikkia vaiheita ja koko tutkimusprosessia. Se on siis olennainen osa onnistunutta tieteellistä tutkimusprosessia ja yksi keskeinen tae sille, että voimme ylipäätään luottaa tutkimuksen tuloksiin. Tutkimuksen tekoon liittyvä etiikka pyrkiikin antamaan tutkijalle sellaisia tutkimuksellisia periaatteita ja käytäntöjä, yleisesti välineitä, joilla he voivat lisätä tutkimuksensa tieteellistä korrektisuutta ja myös luotettavuutta. Tutkimuksen eettinen tarkastelu tulisikin nähdä ja mieltää erottamattomaksi osaksi hyvää ja luotettavaa tutkimuksen tekoa, yleisemmin hyvän tieteen välttämättömänä ehtona. (Hallamaa & Lötjönen 2002.) Miksi näin on? Jo sen vuoksi, että tämä korostaa sitä, että tutkijalla pitää kaikissa tutkimuksenteon vaiheissa olla tietoinen suhde, reflektoiva suhde, omaan tekemiseensä ja tekemättä jättämiseensä. Kaikkiin tutkimuksenteon vaiheisiin liittyykin vähemmän tai enemmän eettisluontoisia kysymyksiä ja ratkaisuja, joista osa voi olla hyvin vaikeasti havaittavissa, jos tutkija ei ole niille ”herkistynyt” (Atjonen 2008).

Oli kyse sitten tutkimuksen aineiston keruusta, käsittelystä, tulkinnasta, raportoinnista tai tutkimusaineiston säilyttämistä, tieteellisen eetillisen koodiston perusohje on: noudata kaikessa hyvää tieteellistä käytäntöä. Se mitä tämä pitää sisällään vaihtelee jonkin verran eri tieteenaloilla. Kuitenkin perusasiat ovat kaikilla aloilla yhteiset. Näistä saa hyvän peruskuvan vuonna 2012 annetuista eettisestä ohjeistosta. Ohjeistossa on esitetty 7-kohtainen muistilista hyvän ja kestävän tieteellisen käytännön keskeisistä tekijöistä. Pääviesti näissä on se, että tutkimus toteutetaan tiedeyhteisön hyväksymiä toimintatapoja noudattaen rehellisesti ja huolellisesti, minkäänlainen vilpinteko (esim. aineiston vääristelyn tai muiden tutkijoiden tekstien plagioinnin muodossa) ei ole sallittua. Lisäksi tutkijoiden odotetaan arvostavan ja kunnioittavan muiden tutkijoiden työtä ja saavutuksia, mikä on tärkeä osa tieteenteon kommunikatiivista ulottuvuutta. Näiden ja muiden hyvien tieteellisten käytäntöjen noudattamisesta ovat vastuussa niin yksittäinen tutkija, myös opiskelijatutkija, kuin tutkija- ja tiedeyhteisökin. (Tutkimuseettinen neuvottelukunta 2012, 6–9.) Tutkijan eettinen valveutuneisuus ja herkkyys ovat tärkeitä empiirisissä tutkimuksissa myös tutkimukseen osallistuvien oikeusturvan kannalta. Lisäksi eettinen herkkyys antaa tutkijalle vankan perustan sille, että toimii tutkiessaan kaikin puolin oikein ja hyvin. Kyse on siis myös tutkijan ja tutkimuksen teon vahvasta tuesta, jota ei kenenkään tutkijan kannattaisi ohittaa olan kohautuksella.

Tieteellisen toiminnan ja tutkimuksen eettisyys koskee siis kaikkia tutkimusvaiheita tutkimusaiheen asettamisesta ja tutkimuksen viitekehyksen muotoilusta aina tutkimusaineiston ja -menetelmien valintaan ja raportointiin saakka (Atjonen 2008). Seuraavaan asetelmaan on esimerkinomaisesti nostettu esiin eri tutkimusvaiheisiin (asetelmassa 1 esitetyn jaottelun perusteella, ks. s. 16) liittyviä eettisen ja kestävän tutkimuksen kysymyksiä, joita jo-

Hyvän ja kestävän tutkimuksen eettisiä lähtökohtia tutkimusprosessin eri vaiheissa

1. Tutkimusongelman määrittäminen
   - tutkimuksen teon autonomisuuden takaaminen
   - tutkimuksen teon lähtökohtien tulee perustua tutkijan tai tutkijayhteisön valintoihin
   - sen arvioiminen, voiko aiottuja tutkimuskysymyksiä tutkia tai tutkimusasetelmaa toteuttaa tieteellisin menetelmin ja eettisesti kestävällä tavalla
   - pohdittava, asettavatko kysymykset tutkittavien yksityisyyden uhatuksi tai vaativatko ne tutkimukseen osallistujilta tai tutkijalta itseltäänkin jotain epäeettisiä tekoja tms.
   - jos tutkii esim. jonkin erityisryhmän kulttuurisia erityispiirteitä, tulee tehdä itselle selväksi, onko esim. yhteisön tai ryhmän ”sisäisten” normien, joita ei ehkä haluta paljastaa, paljastaminen oikeutettua
   - tutkittavan ilmiön riittävän tuntemuksen hankkiminen (aiemman tutkimuksen ja teorioiden hallinta; myös kenttäkokemuksesta on hyötyä)
   - ja yleensäkin omien aiottujen lähestymiskulmien kriittinen arviointi
2. Suunnitteluvaihe
   - tutkimusasetelman ja tutkimuskysymysten eettisyyden arviointi
   - kysymys liittyy koko tutkimusprosessin vaiheiden arviointiin
   - esim. onko valinnut tutkimuksen perusjoukon ja otoksen tutkimuskysymysten suhteen oikein ja perustellusti, ovatko aiotut aineistonkeruumenetelmät ja analyysimenetelmät eettisesti kestävät, tässä vaiheessa on hyvä miettiä myös tutkimustuloksiin mahdollisesti liittyviä eettisiä kysymyksiä
3. Kenttätyövaihe (aineiston keräysvaihe)
   - esim. kyselylomakkeen ja haastattelun toteutuksen asianmukaisuuden arviointi
   - ei esim. vastaajien johdattelua yms.
   - tutkimukseen osallistumisen pitää olla vapaaehtoista
   - aineisto kerätään aidosti (ei sepittämistä) eikä tätä keräysvaiheessa vääristellä mitenkään
4. Tietojen esikäsittely
   - aineiston huolellinen käsittely
   - tilasto-ohjelmia hyödyntävässä tutkimuksessa tämä tarkoittaa sitä, että syöttövaihe tehdään huolella
   - matriisin alustava työstäminen ja täydentäminen suoritetaan asianmukaisesti
   - ei minkäänlaista vääristelyä tms.
5. Tietojen analysointi ja tulkinta
   - tulkinnan tulee perustua tutkimuksen aineistoon (ja tähän liittyviin teoreettisiin lähtökohtiin)
   - tämä on itsestäänselvä periaate, mutta ei ole harvinaista, että tulkintoja tunnutaan tehtävän esim. omien ennakkoajatusten mukaisesti tai aineiston dataa ali- tai ylitulkintaan tavalla tai toisella
   - tulkinnat tulee suorittaa yleensäkin tiedeyhteisöjen normien mukaisesti
6. Tutkimuksen raportointi ja aineiston säilyttäminen
   - raportissa lähteiden käyttö ja viittaus on tehty asianmukaisesti
   - ei vääristellä tuloksia
   - muistettava turvata osallistujien tietosuoja
   - raportoinnista ei saa aiheutua tutkimukseen osallistujille haittaa
   - muistutetaan lukijaa mahdollisista tutkimukseen ja tulkintaan liittyvistä tulkintaongelmista ja puutteista
   - aineisto tulee säilyttää asiallisesti ja niin, ettei sitä saa käsiin kukaan ulkopuolinen

Eettisesti kestävän tutkimuksen normit yleisesti tutkijaan ja tutkimuksen tekoon liittyen (mukaillen Hirvonen 2006)

1. Epäilyn velvollisuus
2. Rajankäynnin velvollisuus eli tieteellisen tiedon suhde muuhun tietoon
3. Vapauden velvollisuus
4. Ajattelun velvollisuus
5. Avoimuuden velvollisuus
6. Vastuunottamisen velvollisuus
7. ”Säännöttömyyden” vaatimus
   - tieteen tulee ylittää tai ainakin haastaa normaalin ajattelun ja normien asettamat totutut ajattelutavat yms.

ASETELMA 6. Tieteellisen tutkimuksen teon eettisiä lähtökohtia.

kaisen on hyvä pohtia eri vaiheissa tutkimus prosessia. Lisäksi asetelmassa on nostettu esiin Ari Hirvosen (2006) esityksen perusteella hyvän tutkimuksen teon erityisesti tutkijalle asettamia yleisiä odotuksia:

Kuten asetelmasta 6 ilmenee, tutkimusprosessissa eettisen arvioinnin tulisi olla kaikissa vaiheissa yhtenä tutkimuksen keskeisenä johtotähtenä. Asetelman alimman osan vaateet korostavat sitä, että tutkijan odotetaan olevan tarkkana tieteellisen tiedon ja sen tuottamisen suhteen. Tämä edellyttää tutkijalta avoimuutta, kriittisyyttä, ajattelua ja myös vapautta ulkoisista paineista. Periaatteessa tutkimuksenteon yleisiinkin kriteereihin, kuten objektiivisuuden, itseohjautuvuuden, luotettavuuden ja kriittisyyden vaateisiin on sisäänkirjoitettu eettisen ja avoimen suhtautumisen odotteet (Pohjala 2007).

Yksi tärkeä lähtökohta kaikelle tutkimuksen tekemiselle on se, että tutkija on avoin ja herkkä aineistonsa suhteen sekä tarkka tutkijuuteensa liittyvästä itsemääräämisoikeudestaan. Tässä suhteessa on tärkeää, kuten jo edellä on todettu, hallita riittävästi oman tutkimuskohteensa teoreettinen tuntemus, joka antaa mahdollisuuden tarkastella oman aineiston tuloksia mahdollisimman laajasta perspektiivistä. Muuten on vaarana se, että tulkitsee aineistoa jostain ennalta ajatellusta viitekehyksestä käsin, huomaamatta sitä, ettei oma aineisto oikeasti tuekaan tästä näkökulmasta tehtäviä tulkintoja. On siis oltava tarkkana myös tutkimuksen viitekehyksen valinnassa ja tästä mahdollisesti nousevien tulkintavirhetekijöiden suhteen. Tämä sama koskee yleensä jo oman tarkastelukulman valintaa, joka voi pahemmissa tapauksissa johdattaa tarkastelemaan kysymystä tai tutkittavaa ilmiötä varsin kapeasta näkökulmasta käsin. (Pohjola 2007; Atjonen 2008.)

Hyvään tutkimustapaan kuuluu myös tutkimukseen osallistuvien kunnioitus. On esimerkiksi tärkeää, että aineistosta tehtävät tulkinnat tekevät ”oikeutta” vastanneille. Tämä sama koskee myös muita tutkijoita, jotka ovat tutkineet työn alla olevaa aihetta: on oltava tarkkana esimerkiksi sen suhteen, ettei esimerkeissä esitä omissa nimissä toisen esittämiä tulkintoja puhumattakaan toisten tutkijoiden tuloksista. Myös se ei ole hyvän tutkimustavan mukaista, että viittaa vain sellaiseen tutkimukseen tai sellaisiin tulkintoihin, jotka tukevat omia lähtökohtia ja tulkintaa. Hyvässä tieteellisessä keskustelussa ja raportissa otetaan huomioon vastaväitteen esittäjät ja keskustellaan heidän kanssaan. Myös toisten tekstien suora plagiointi on ilman muuta kiellettyä. Kaiken kaikkiaan onkin tärkeää noudattaa raportin kirjoittamisessa tarkkaan lähteiden ja kirjallisuusviitteiden käyttöön liittyviä periaatteita. Tutkimusraportissa tulee pitää huolta myös siitä, että tutkittavat henkilöt eivät tule tunnistetuksi (anonymiteetin turvan periaate). Erityisesti tämä koskee laadullisia tutkimuksia, mutta myös joskus määrällisissäkin tutkimuksissa voi olla sellaisia asetelmia, jolloin tämä tulee ottaa huomioon. (Pohjola 2007.)

Vaikka nämä kriteerit tuntuvat vaativilta, kuten tutkimukselle yleiset itsekorjautuvuuden ja systemaattisuuden vaatimukset yleensäkin, ensisilmäyksellä varsin abstrakteilta ja ahdistavilta, ei kyse oikeasti ole siitä. Yleensä näihin löytyy ratkaisu ihan maalaisjärkeä käyttäen, kuten esimerkiksi Päivi Atjonen (2008) aivan oikein huomauttaa. Kyse on hänen mielestään ennen kaikkea siitä, että tutkijoina sitoudumme rehellisyyteen, avoimuuteen ja kriittisyyteen, ja siitä, että tutkijoina olemme tietoisia siitä, mitä teemme. Kestävän ja hyvän tutkimuksen teko edellyttää meitä jatkuvasti reflektoimaan sitä, mitä teemme. (Ks. myös Pohjola 2007.) Tutkimuksen eettisten kysymysten arvioiminen tukee siis tutkijaa tekemään oikeita ratkaisuja, suorittamaan tutkimuksensa asianmukaisesti ja arvostamaan tutkimuksessa mukana olevien ihmisarvoa ja yksityisyyttä (Kuula 2011, 58–65.)

2. Havaintoaineiston syöttäminen ja matriisin muokkaaminen

Ennen varsinaisiin tilastoanalyyseihin siirtymistä keskitytään vielä havaintomatriisin muodostamiseen ja sen yleiseen esikäsittelyyn. Tilastollisen analyysin kulkua voi kuvata vähän samalla tavalla yleisten vaiheiden avulla kuin tutkimuksen tekoprosessin etenemistä kuvattiin aiemmin asetelma 1:ssä (s. 19).

1. aineiston esikäsittely, syöttäminen tilasto-ohjelmaan ja sen muokkaaminen
2. aineiston alustava tarkastelu frekvenssien ja eri tunnuslukujen avulla
3. käytettävien testien valinta
4. analyysien ja testien suorittaminen
5. saatujen tulosten analysointi ja tulkinta.

Nämä vaiheet, kuten koko tutkimuksenteon eri vaiheet, menevät käytännössä usein päällekkäin. Aineiston käsittelyvaiheessa saattaa nousta esiin uusia aineiston analyysitarpeita, jolloin voidaan toisinaan joutua harkitsemaan uusien menetelmien käyttöönottoa ja joskus palaamaan jopa aineiston alustavaan tarkasteluvaiheeseen uudelleen. Mutta ennen kuin voidaan toteuttaa yhtäkään analyysia tilasto-ohjelmalla, on kyselylomakkeella tai muulla tavalla hankittu aineisto syötettävä ohjelmaan. Seuraavassa paneudutaan havaintomatriisin tekemiseen. Jos havaintoaineisto on kerätty sähköisesti jollakin aineistonkeruujärjestelmällä, säästytään havaintoarvojen syöttämisen vaivalta. Tosin silloinkin ennen varsinaisia analyysejä aineistoa yleensä joudutaan muokkaamaan ja esikäsittelemään.

2.1 Havaintoaineiston ja -matriisin teko ja kirjassa käytetyn matriisin esittely

Julkaisussa hyödynnetään pääosiltaan Karma & Komulaisen (1990) teoksessa käytettyä esimerkkiaineistoa hiukan täydennettynä (liite 1), jotta läpi käydyt eri tilastoanalyysit voidaan käydä tarkoituksenmukaisemmin lävitse. Lisäksi otoksen kokoa on lisätty alkuperäisestä 30:stä 60:een. Näin eri menetelmien, kuten esimerkiksi ristiintaulukointiesimerkin, analyysien käsittely tulee tarkoituksenmukaisemmaksi. Kokonaan ei muuttujamäärän vähäisyydestä ja otoskoosta johtuvia ongelmia ole voitu välttää. Esimerkiksi faktorianalyysi- ja reliaabeliusesitykset ontuvat jossakin määrin näiden ongelmien vuoksi. Toisaalta esityksistä pitäisi kuitenkin ilmetä kunkin analyysin pääpiirteet. Joissakin kohdissa esitykseen on otettu esimerkinomaisesti muuhun aineistoon perustuvia taulukoita ja kuvioita.

Esimerkkiaineistossa (ks. liite 1) on mukana siis 60 opiskelijaa, joista meillä on kerätty seuraavat tiedot:

Muuttujat	Arvot ja luokitukset havaintoaineistossa
1) sukupuoli	1) pojat=1 ja tytöt=2
2) oppilaiden vanhempien koulutustaso	2) kansakoulu=1, peruskoulu=2 ja lukio=3
3) oppilaiden viriketausta	3) matala=1, keskitaso =2 ja korkea=3
4) oppilaiden verbaalitestissä saavuttamat pistemäärät	4) 20–30 (testipisteet)
5) järkeilytestin pistemäärät	5) 30–40 (testipisteet)
6) kieliaineiden keskiarvot	6) 4–10 (koodattu muodossa 40–100)
7) matematiikan numerot	7) 4–10 (kouluarvosanat)
8) opintomenestys lukiossa	8) 10–20 (indeksiarvo)
9) opintomenestys yliopistossa	9) 10–20 (indeksiarvo)
10) KMER_1 eli minkälaisen merkityksen opiskelijat antavat koulutukselle omaa tulevaisuuttaan ajatellen lukiossa	10) ei anna mitään arvoa=1, jonkin verran=2 ja antaa suuren arvon=3
11) KMER_II eli minkälaisen merkityksen opiskelijat antavat koulutukselle omaa tulevaisuuttaan ajatellen yliopistossa	11) ei anna mitään arvoa=1, jonkin verran=2 ja antaa suuren arvon=3

Tutkimusasetelmassa on kaikkiaan 11 tekijää (muuttujaa), jotka voidaan nähdä tyypillisiksi tekijöiksi ja kysymyksiksi, joista kasvatustieteellisissä tutkimuksissa ollaan kiinnostuneita. Näiden tekijöiden avulla voidaan tutkia esimerkiksi sitä, millä tasolla lukion opiskelijat ovat näiden viiden kysymyksen suhteen yleisesti tai vaikka sitä, eroavatko tyttöjen ja poikien taitotaso, arvosanat tai opintomenestys toisistaan (kuvailevia kysymyksiä). Tämän lisäksi voidaan jo näiden muuttujien avulla tutkia, millainen yhteys esimerkiksi kielellisellä ja matemaattisella taitotasolla on oppilaiden koulumenestykseen lukiossa. Voidaan vielä muiden aineistossa olevien muuttujien avulla tarkastella muun muassa sitä, miten lukion opintomenestys heijastuu yliopisto-opiskelussa menestymiseen ja kysyä sitä, millaiseksi opiskelijat näkevät koulutuksen merkityksen lukiossa tai yliopistossa oman tulevaisuutensa kannalta. Entä onko tämä merkityksen anto yhteydessä heidän opintomenetykseensä lukiossa tai yliopistossa? Lisäksi voidaan aineistossa olevien muuttujien avulla pyrkiä mallittamaan esimerkiksi regressio- tai varianssianalyysilla oppilaiden yleiseen opintomenestykseen vaikuttavia tekijöitä. Vaikka esimerkkiaineistossa on vain 11 muuttujaa, voi jo näinkin pienellä muuttujamäärällä tarkastella aihetta monesta näkökulmasta käsin, kunhan valitut tekijät/kysymykset vain on relevantisti määritelty ja kerätty tutkimuskysymysten suhteen.

Nykyisin tilasto- ja taulukkolaskentaohjelmistot ovat niin kehittyneitä, että aineiston voi syöttää niihin suoraan esimerkiksi lomakkeilta. Syötämme aineiston tilasto-ohjelmaan (tässä kirjassa SPSS-versio 26), ja muodostamme tämän jälkeen havaintoaineiston, joka SPSS:ssä on samalla havaintomatriisi (liite 1). Tämän avulla voit vaikka itse syöttää matriisin ohjelmaasi ja tehdä harjoittelumielessä käsiteltäviä analyysejä. Aineiston saat ladattua myös valmiina kirjan verkkosivulta (https://www.utu.fi/tilastoanalyysiopas). Periaatteessa voitaisiin aloittaa nyt varsinainen aineiston analysointi ja käsittely. Yleensä tässä vaiheessa tehdään raakamatriisiin vielä joitakin lisämuokkauksia ja -tarkennuksia, voimme esimerkiksi antaa muuttujien numeerisille arvoille tekstimuotoiset selitteet ja niin edelleen. Seuraavaan alalukuun on koottu lyhyesti havaintomatriisin syötön periaatteet SPSS:ssä ja muut alkutoimet, jotka kannattaa tehdä ennen varsinaisten analyysien tekoa. Voit käyttää myös ohjelman ohjekirjoja tai ohjelman *Help*-toimintoa, joka on varsin kattava esitys. Sen sisältämien osien (*Topics → Tutorials / Case Studies / Statistics Coach*) avulla ohjelman käyttöön pääsee tutustumaan ilman kurssejakin melko kivuttomasti. Lisäksi monesti erittäin käyttökelpoista kohdennettua opastusta löytyy SPSS-analyysiproseduurien valintaikkunoiden yhteydessä löytyvistä Help -painikkeista.

2.1.1 Havaintomatriisin teko SPSS:ssä ja muita matriisinmuokkaamiseen liittyviä vinkkejä

Tietyssä mielessä aineiston raakamatriisi, joka sisältää kunkin havaintoyksikön eri muuttujien koodausarvot, sisältää kaiken havaintoaineiston informaation. Käytännössä tämä matriisi ei kuitenkaan avaudu pelkästään sitä silmäilemällä, vaan tarvitsemme lisäanalysointeja ja -ryhmittelyjä. Tilastollisten menetelmien avulla voidaankin tiivistää alkuperäismatriisissa olevaa informaatiota tulkittavampaan muotoon sekä kuvata ja analysoida eri muuttujien välisiä yhteyksiä. Ennen kuin päästään tähän vaiheeseen, aineisto pitää syöttää kuitenkin ohjelmaan. Lisäksi usein näin muodostunutta matriisia joudutaan vielä jonkin verran työstämän tulevia analyyseja silmällä pitäen. Seuraavassa esitetään lyhyesti muutamia perusasioita havaintoaineiston laatimisesta SPSS:ssä Windows-ympäristössä. (Ohjeet perustuvat ohjelmiston versioon 26.0). Samalla annetaan muutamia yleisimpiä vinkkejä matriisin muokkaamiseen.

Useat toimenpiteet voidaan toteuttaa monella eri tavalla, esimerkiksi yläreunan työkalupalkin painikkeita käyttämällä. Annetut vinkit ovat vain suuntaa antavia alkuun pääsemisen helpottamiseksi. SPSS:n *Help*-valikosta voit katsoa tarkempia ohjeita. SPSS:n grafiikan käytöstä kerrotaan tämän kirjan eri kuvioiden yhteydessä.

Yleisiä lähtökohtia havaintoaineiston laadintaan ja SPSS:n käyttöön:

Ohjelman käynnistys: Käynnistäminen tapahtuu kaksoisnapsauttamalla Windowsin työpöydältä *SPSS*-kuvaketta (*IBM SPSS Statistics 26*). Jos kuvake ei ole työpöydällä, löydät ohjelman: alareunan Windows-kuvake → *IBM SPSS Statistics26* -kansio → *IBM SPSS Statistics26*. Voit myös kirjoittaa Windows-hakukenttään (alareunan suurennuslasi-symboli) alkua ohjelman nimestä, esim. IBM SPSS, jolloin ohjelma löytyy ja on avattavissa.

Syöttö-ikkuna: Ohjelmiston käynnistyttyä valitse aloitusnäkymässä vaihtoehto *New Dataset* → *Open*, mikäli olet aloittamassa aivan uuden havaintomatriisin teon. Jos havaintomatriisi on jo aiemmin tehty ja haluat jatkaa sen muokkaamista tai analysointia, valitse vaihtoehto *Open another file* → *Open*. Valitsitpa kumman tavan tahansa, sen jälkeen olet havaintoaineiston syöttöikkunassa (*IBM SPSS Statistics Data Editor*).

Datan alkutietojen syöttäminen: Syöttöikkunan alapalkissa ovat vaihtoehdot *Data View* ja *Variable View*. Jos valitset jälkimmäisen vaihtoehdon, niin pääset syöttämään havaintoaineiston muuttujien alkutiedot eli nimeämään muuttujat, määrittelemään muuttujien mitta-asteikon, määrittelemään puuttuvien tietojen koodit jne. Nämä määrittelyt tehdään kaikille aineiston muuttujille.

Sarakkeessa *Name* nimeät muuttujat (ensimmäisen merkin on oltava kirjain, yhdysviivaa tai välilyöntiä ei voi käyttää nimissä). Aineiston yleiset taustamuuttujat kannattaa nimetä lyhyillä kuvaavilla nimillä, esim. sukupuoli → SUKUP. Sen sijaan muuttujat, joista muodostetaan esimerkiksi myöhemmin summamuuttujia, tai muuttujat, joita on vaikea nimetä yksittäin, nimetään yleensä kysymyksen lomakkeella olevan numeron mukaan, esim. K8, K9A, K9B. Kun nimeät muuttujan ja painat *Enter*-painiketta (tai siirryt *nuoli alas* -näppäimellä seuraavan muuttujan nimeämiseen), muuttujan muut määrittelysarakkeet täyttyvät oletusarvoillaan. Oletusarvot ovat monesti varsin sopivat. Jos syöttämäsi arvot ovat kokonaislukuja, silloin sarakkeen *Decimals* arvoksi kannattaa antaa arvo nolla (oletusarvon kaksi sijasta), jolloin havaintomatriisin ulkoasu muodostuu selkeämmäksi.

Sarakkeessa *Label* voit antaa muuttujan sisältöä kuvaavan pidemmän sanallisen selitteen, jos *Name*-kohdan nimi ei mielestäsi ole esim. riittävän selkeä. Jos muuttujan arvot ovat luokkia, kuten 1=nainen, 2=mies, silloin sarakkeessa *Values* pystyt määrittelemään muuttujalle tekstiarvot. Se tehdään näin: napsauta vasemmalla hiirinäppäimellä *Values*-saraketta haluamasi muuttujan kohdalla. Napsauta seuraavaksi esiin tullutta sinistä ruutua ja kirjoita avautuvan *Value Labels* -ikkunan kohtaan *Value* numero 1 ja kohtaan *Label* kirjoita *nainen*. Paina painiketta *Add*. Vastaavalla tavalla määrittelet numeron 2 tekstiarvoksi sanan *mies*. Paina lopuksi *OK*-painiketta. Jos useilla muuttujilla on samat tekstiarvot, riittää, että kirjoitat tekstiarvot edellä esitettyyn tapaan yhdelle muuttujalle, kopioit ja liität määrittelyn muihin muuttujien *Values* -kohtiin (hiiren oikealla painikkeella *Copy&Paste* tai painikkeilla *Ctrl&C ja Ctrj&V)* Huom! kopioi-liitä -periaate toimii monessa muussakin kohdassa, joten sen avulla vältyt usein turhalta vaivannäöltä.

Sarakkeessa *Missing* voit määritellä puuttuvien tietojen koodit. Oletusarvo *None* tarkoittaa sitä, että tietojen syöttövaiheessa puuttuvat tiedot jätetään tyhjiksi. (Huom! Puuttuvan tiedon merkkinä SPSS käyttää automaattisesti pientä pistettä). Tarvittaessa puuttuvan tiedon koodiksi voi määritellä itse myös minkä tahansa lukuarvon. Jos esimerkiksi epäselvät tai kelpaamattomat vastaukset halutaan syöttää aineistoon koodilla 999, tämän arvon voi määrittää puuttuvaksi tiedoksi *Missing* -sarakkeen lisämäärityksissä (pieni sininen ruutu muuttujan missing-sarakkeen oikeassa reunassa): *Discrete missing value* -kohtaan annettaisiin arvo 999. Tällöin muuttujassa olevat arvot 999 käsiteltäisiin jatkossa puuttuvana tietona.

Sarakkeessa *Width* voidaan rajata muuttujan arvoissa olevien merkkien, kuten numeroiden lukumäärä. Sarake *Columns* määrittelee sen, kuinka monen merkin levyisenä muuttuja näkyy havaintomatriisissa. Sarake *Align* puolestaan keskittää muuttujan arvot tai tasaa ne jompaankumpaan reunaan. Sarakkeissa *Type* ja *Measure* määritellään muuttujan tyyppi ja mitta-asteikko. Tyyppi tarkoittaa tässä yhteydessä sitä, syötätkö muuttujan arvot numeroina (*Numeric*), kirjaimina (*String*), päivämäärinä (*Date*), tms. Lähestulkoon aina muuttujien arvot pyritään syöttämään numeroina, joten tähän kohtaan valitaan siinä tapauksessa vaihtoehto *Numeric*, joka onkin oletusarvo. Kohdassa *Measure* voit määritellä muuttujan mitta-asteikon. Vaihtoehdot ovat laatu- eli nominaaliasteikko (*Nominal*), järjestysasteikko (*Ordinal*) ja vähintään välimatka-asteikko (*Scale*). Voit käyttää oletusarvoa *Scale*, jos et halua lähteä erikseen miettimään jokaisen muuttujan mitta-asteikkoa. Muista kuitenkin, että muuttujan mitta-asteikko määrää melko pitkälle sen, millaisia tilastomenetelmiä kyseisen muuttujan kohdalla voi käyttää. Vaikka siis antaisitkin esim. sukupuolimuuttujan olla yleismäärittelyissä *Scale*-muotoa, analyyseissa muista, että kyseinen muuttuja on nominaaliasteikollinen, joten esim. keskiarvon laskeminen kyseisenlaisen muuttujan kohdalla ei ole mielekästä. Lisäksi joissain SPSS-ohjelman toiminnoissa (mm. *Chart Builder* -toiminto) on eduksi tai jopa vaatimuksena, että muuttujien mitta-asteikot on täsmällisesti määritelty. *Role*-sarakkeessa voidaan antaa muuttujalle sen tulevaa käyttötarkoitusta kuvaava rooli. Esimerkiksi valinta *Target* = selitettävä muuttuja, jolloin muuttujan arvoja pyritään analyyseissä selittämään muilla tekijöillä. Oletusarvoa *Input* voi hyvin käyttää yleisvalintana.

Datan syöttäminen ja tallentaminen:

Yleensä havaintomatriisin tiedot syötetään lomakkeilta ja lomakkeet on numeroitu juoksevasti. Esimerkkimatriisissa *ID*-niminen muuttuja kertoo vastauslomakkeen (vastaajan) numeron. Lisää myös omaan matriisiisi ensimmäiseksi muuttujaksi tällainen muuttuja, sillä joudut usein tarkistamaan lomakkeelta syöttämiesi arvojen oikeellisuuden huomatessasi esimerkiksi syöttäneesi virheellisen arvon havaintomatriisiin.

ID-muuttujan avulla löydät oikean lomakkeen ja pystyt tekemään korjauksen. ID-muuttuja on tarpeen myös silloin, jos haluat tehdä analyyseja esimerkiksi pienemmistä osa-aineistoista lomakkeiden numeroinnin perusteella. Lisäksi ID-muuttujan avulla aineisto voidaan tarvittaessa järjestää muotoon, jossa vastaajat ovat alkuperäisessä järjestyksessä.

Syöttöikkunan alapalkin vaihtoehdosta *Data View* pääset havaintojen syöttövaiheeseen. Syöttäminen on yksinkertaista: vie kohdistin ensimmäisen havaintoyksikön (koehenkilön yms.) kohdalle eli ensimmäisen rivin alkuun ja aloita havaintoyksikön saamien arvojen syöttäminen. Kunkin arvon syöttämisen jälkeen paina *nuoli oikealle* -näppäintä, jolloin pääset syöttämään havaintoyksikön seuraavan muuttujan arvon sille kuuluvalle sarakkeelle. (*Enterin* painallus tarkoittaisi sitä, että syöttäisit arvot muuttujittain eli sarakkeittain). Virheelliset lyönnit voit korjata menemällä virheelliseen soluun ja antamalla uuden arvon. Tallenna matriisi aika ajoin syöttämisen edistyessä ja aina kun kulloinenkin työrupeama on ohi. Tallentaminen onnistuu valikkorivin *File*-kohdan avulla, josta valitset joko *Save As*:n tai pelkän *Save*n. *Save As* -riviä käytetään, kun matriisi nimetään ensimmäisen kerran tai sille annetaan uusi nimi. Samalla pääset valitsemaan kansion, jonne aineistosi tallentuu. Tallennettu SPSS-tiedosto sisältää sekä muuttujien määrittelyt että havaintomatriisin ja tiedoston nimen lopputunnukseksi tulee automaattisesti *.sav*.

Huom! Varmuuskopiot verkkolevylle, pilvipalveluun tai vaikkapa muistitikulle ja tarvittaessa tiedostosi poistaminen kiintolevyltä on yleensä selkeintä tehdä *Windowsin Resurssienhallinnassa* SPSS:n sulkemisen jälkeen. Resurssienhallinnassa tiedostonimien loppuosat eivät näy, vaan Windows käyttää niitä tiedostojen tyypin tunnistamiseen. Siten *sav*-loppuiset tiedostot tulkitaan SPSS-tiedostoiksi. Opettele Windowsin resurssienhallinnan joustava käyttö.

Liitteessä 1 on käytetty perusmatriisi siinä muodossa, missä se tämän vaiheen jälkeen on. Jos olet antanut muuttujien numeroille myös tekstiarvoja, silloin numerot voivat näkyä matriisissa myös tekstiarvoillaan. SPSS:ssä ylhäällä olevan työkalupalkin oikeassa reunassa sijaitsevan punavalkoisen *nimilappu*-painikkeen (kolmas painike oikealta: *Value Labels*) avulla saat matriisisi näkymään joko pelkkinä numeroina tai määrittelemiesi tekstiarvojen kera.

Nyt voitaisiin periaatteessa aloittaa varsinainen datan analysointi. Melko usein matriisille joutuu kuitenkin suorittamaan ennen analysointia vielä alustavia lisätoimia, joten tarkastellaan niitä seuraavaksi.

Matriisin jatkotyöstö:

A) Muuttujan arvojen muutokset

Joskus on tarpeen muuttaa muuttujien arvoja. Voit esimerkiksi haluta kääntää tietyt käänteiset muuttujat samansuuntaisiksi toisten kanssa, jotta voit esimerkiksi muodostaa näistä uuden summamuuttujan tai haluat jostakin syystä käsitellä muuttujaa jollakin vakiolla yms. Muuttujan arvot käännät seuraavasti. Valitse ylhäältä valikkoriviltä vaihtoehdot *Transform* ja *Compute*. Saat esiin ikkunan, jossa muuttujille voi tehdä monenlaisia muunnosoperaatioita. Olkoon alkuperäisen muuttujan nimi vaikkapa K9B ja sen eri arvot 1,2,3,4 ja 5. Tehdään tämän muuttujan avulla käännetty muuttuja nimeltään K9BK. Tavoitteena on siis kääntää muuttujan vastaus-skaala niin, että muutetaan ykköset viitosiksi, kakkoset nelosiksi, jne. Kirjoita tekstikenttään *Target Variable* käännetyn muuttujan nimi K9BK. Napsauta kohdistin kohtaan *Numeric Expression* ja paina painikkeita 6 ja – (numero 6 ja miinusmerkki) ja valitse muuttujalistasta se muuttuja, jonka haluat kääntää eli tässä K9B ja lisää se muuttujalistan oikealla puolella olevalla nuolipainikkeella muunnoslausekkeen perään. *OK*:n painallus suorittaa muunnosoperaation ja liittää käännetyn muuttujan havaintomatriisin loppuun. Samaan lopputulokseen päästäisiin myös toiminnolla valikkorivin *Transform* ja *Recode into Different Variables*. Avautuvassa valintaikkunassa muuttuja K9B siirretään nuolipainikkeella keskimmäiseen laatikkoon ja uuden muuttujan nimi K9BK annetaan oikean reunan *Output Variable*-osan kohtaan *Name*. Nimi hyväksytään painikkeella *Change*. Arvojen kääntäminen toteutetaan painikkeella *Old and New Values* siten, että *Old Value*-kohtaan annetaan alkuperäinen arvo 1 ja *New Value* kohtaan annetaan muutettu arvo 5. Hyväksy muutos painikkeella *Add*. Tämä toistetaan kaikille muunnettaville arvoille, ja lopuksi hyväksytään muutokset painamalla *Continue* ja *OK*.

B) Poiminnat havaintomatriisista

Joskus tulee vastaan tilanne, jossa haluat tehdä analyysin vain tietylle ryhmälle, esimerkiksi haluat analyysiin mukaan vain tytöt. Tällöin sinun on luotava väliaikaisesti matriisi, jossa on vain tyttöjä. Valitse ylhäältä valikkoriviltä vaihtoehdot *Data* ja *Select Cases*. Valitse seuraavaksi *If condition is satisfied* ja napsauta painiketta *If*. Olet päässyt havaintoyksiköiden valintaikkunaan. Valitse muuttujalistasta sukupuolimuuttuja, ja lisää se listan oikealla puolella olevalla nuolipainikkeella määrittelykenttään, ja valitse sen jälkeen painikkeet = ja 2. *Continue* ja *OK*-painikkeet lopettavat määrittelyn. Nyt analyysiin tulevat mukaan vain tytöt, sillä heitä on merkitty matriisissa kakkosella. Havaintomatriisissa poikien kohdalla näkyy rivinumeron päällä yliviivaus. Toiminto poistetaan valitsemalla uudelleen *data, select cases* ja *analyze all cases*. Tämän jälkeen aineiston kaikki tapaukset ovat jälleen analyyseissä mukana.

C) Puuttuvien tietojen käsittely

Puuttuvia tietoja ei yleensä korvata. Kuitenkin puuttuvien tietojen korvaamisella on merkitystä mm. summamuuttujia muodostettaessa. Summamuuttuja koostuu nimensä mukaisesti useasta muuttujasta, ja jos näistä yhdenkin kohdalla on puuttuva tieto, summamuuttujankin arvoksi tulee puuttuva tieto. Jottei summamuuttujaan tulisi kohtuuttoman paljon puuttuvia arvoja, summamuuttujaan kuuluvien muuttujien puuttuvia tietoja voi koettaa korvata tai käsitellä jollakin menettelyllä. Erilaisia tapoja korvata tai muuten käsitellä puuttuvaa tietoa on useita, monet niistä varsin monimutkaisiakin etenkin aloittelevalle tutkijalle. Eräs melko suoraviivainen tapa käsitellä puuttuvaa tietoa, on puuttuvien tietojen korvaus muuttujassa olevien muiden havaintoarvojen keskiarvoilla. Keskiarvokorvaus onnistuu seuraavasti. Valitse ylhäältä valikkoriviltä vaihtoehdot *Transform* ja *Replace Missing Values*. Valitse muuttujalistasta muuttujat, joiden puuttuvat arvot haluat korvata keskiarvoillaan. SPSS ei hävitä alkuperäisiä muuttujia, vaan tekee niistä uudet muuttujat, jotka ohjelma samalla nimeää automaattisesti. Kohdasta *Name* voit halutessasi vaihtaa ohjelman oletuksena tarjoamaa nimeä. Kohdassa *Method* on korvausmenetelmä. *Series mean* tarkoittaa tässä yhteydessä sitä, että puuttuvat tiedot korvataan kyseisen muuttujan havaintoarvojen keskiarvoilla. *OK*:n painallus lisää uudet muuttujat, joissa siis puuttuvat tiedot on korvattu keskiarvoillaan, havaintomatriisin loppuun. Keskiarvokorvaus on yksinkertainen, mutta ei optimaalinen tapa käsitellä puuttuvia tietoja. Keskiarvokorvauksen lisäksi muita mahdollisia keinoja puuttuvien tietojen käsittelyyn olisivat mm. regressiokorvaus, em-algoritmin tuottama korvausmenettely ja moni-imputointi. Tällaisia kehittyneempiä tapoja käsitellä puuttuvia tietoja löytyy SPSS-valikosta *Analyze* → *Missing Value* Analysis. Huom! Puuttuvan tiedon käsittely vaatii aina perusteellista harkintaa sekä alustavaa perehtymistä ja tarkastelua puuttuvan tiedon esiintymisestä ja luonteesta. Moni korvausmenettely esimerkiksi edellyttää, että puuttuva tieto esiintyy satunnaisesti aineistossa. Jos aineistosi on suuri ja puuttuva tietoa esiintyy vain vähän ja satunnaisesti, puuttuvien tietojen käsittelyyn ei ole välttämättä lainkaan tarvetta. (Ks. tarkemmin Vehkalahti 2014, 81–86; Töttö 2012, 118–140.)

D) Havaintomatriisien yhdistäminen

Kaksi erillistä matriisia voidaan yhdistää joko niin, että perushavaintoaineistoon lisätään toisen matriisin havain toyksiköt (matriisien muuttujien tulee olla samassa järjestyksessä) tai niin, että perusmatriisiin liitetään samoille havaintoyksiköille (koehenkilöille) lisää muuttujia, esim. uusintamittauksen muuttujat, jolloin matriisien havaintoyksiköiden tulee olla samassa järjestyksessä. SPSS:ssä yhdistäminen tapahtuu siten, että ylhäältä valikkoriviltä valitset toiminnot *Data→Merge Files→Add Cases* (jos yhdistät havaintoyksiköitä) tai *Add Variables* (jos yhdistät muuttujia). Kun lisäykset on tehty, on tärkeätä tallentaa uusi matriisi. Tallennus kannattaa tehdä valikkorivin vaihtoehtojen *File* ja *Save As* avulla, jolloin voit tallentaa matriisin uudella nimellä. Näin alkuperäinen perusmatriisi säilyy muuttumattomana ja tarvittaessa voit myös palata siihen.

Analyysien ja grafiikan tallennus: Usein on tarkoituksenmukaista tallentaa tehdyt analyysit ja grafiikat muistiin sopivaan kansioon, josta ne voi tarvittaessa avata uudelleen ja tulostaa tai viedä tekstinkäsittelyyn. Analyysien tulokset ja kuviot menevät automaattisesti ohjelman erilliseen tulosteikkunaan (*SPSS Viewer)*. Tulosteikkuna koostuu itse asiassa kahdesta osasta. Vasemmassa osassa on tulosteiden sisällysluettelo, ja oikeassa osassa ovat varsinaiset tulosteet.

Kummassakin osassa pystyt vaihtamaan tulosteiden paikkoja tai poistamaan haluamiasi tulosteita. (Klikkaa haluamaasi kohdetta kohdistimella, jolloin valitun kohteen eteen tulee punainen nuoli. *Shift*-näppäimen avulla voisit valita useita kohteita; *Ctrl*-näppäimen avulla voisit poimia kohteita sieltä täältä. Vasenta hiirinäppäintä pohjassa pitämällä voit hiiren avulla siirtää kohteen uuteen paikkaan. *Delete*-näppäimen avulla puolestaan poista kohteen).

Oikeanpuoleiseen tulosteosaan voit lisäksi kirjoittaa haluamaasi tekstiä tavallisen tekstinkäsittelyohjelman tapaan valitsemalla valikosta *Insert* kohdan *New Text*. Lisäksi voit muokata tulosteiden (lähes kaikki) tekstit sellaisiksi kuin haluat, esimerkiksi voit vaihtaa (lähes kaikki) englanninkieliset termit suomenkielisiksi. Muokkaamisen teet pääsääntöisesti siten, että ensin valitset kohteen (esim. taulukon) muokkaustilaan kaksoisklikkaamalla. Tämän jälkeen pääset muokkaamaan kyseistä kohdetta. Kannattaa hyödyntää hiiren oikealla painikkeella avautuvia toimintoja sekä ylävalikon muokkaustoimintoja. Kokeilemalla oppii!

Tulosteikkunan sisällön oletusnimi on *Output1*. Tulosteikkunan sisällön voit tallentaa valitsemalla ylhäältä valikkoriviltä vaihtoehdot *File→Save As*. Tällöin voit myös antaa tulostetiedostolle haluamasi nimen. Tulostetiedosto tallentuu .spv-loppuisena (*Viewer Files* -muoto), ja se on vain SPSS:n ymmärtämä tallennusmuoto. Toisin sanoen tämän muotoisena pystyt avaamaan tulostetiedoston vain SPSS:ään. Huom! Vanhemmilla SPSS-versioilla tehdyt .spo -päätteiset tulostetiedostot eivät suoraan avaudu uudempiin versioihin, mutta tiedosto on käännettävissä uudempaan versioon SPSS:n tarjoamalla lisäohjelmalla. Tarkastellaan seuraavaksi sitä, miten saat SPSS-tulosteitasi tekstinkäsittelyohjelmaan.

SPSS:n tulosteen ja grafiikan siirtäminen tekstinkäsittelyyn: Helpoiten teet siirtämisen näin: merkitse SPSS:n tulosteikkunassa siirrettävät kohteet napsauttamalla; jos samanaikaisesti pidät pohjassa *Shift*-näppäintä, voit valita useita peräkkäisiä kohteita. Jos pitäisit *Ctrl*-näppäintä pohjassa, voisit poimia kohteet tulosteikkunasta sieltä täältä. Valitse sen jälkeen ylhäältä valikkoriviltä vaihtoehdot *Edit→Copy*, ja siirry tekstinkäsittelyohjelmaan. Tekstinkäsittelyohjelmassa etsi käsky *Liitä* (*Paste*) tai käytä näppäinyhdistelmää *Ctrl-V*, jolloin siirto käynnistyy. Kyseessä on siis leikkaa–liimaa -menettely.

Aika usein kuitenkin haluat kopioida SPSS:n tulostuksesta jonkin taulukon täsmälleen samanlaisena kuin se siellä on. Silloin edellä esitetty muuttuu hivenen. Ainoa ero itse asiassa on se, että *Copy*-käskyn sijasta käytätkin käskyä *Copy Special*, josta voit valita siirrettävän muodon, esimerkiksi *Metafile*. Tällöin esimerkiksi taulukot siirtyvät tekstinkäsittelyohjelmaan kuvamuodossa, jolloin taulukoiden asettelut säilyvät täsmälleen ennallaan. Toisaalta tässä muodossa siirrettyä kohdetta ei päästä muokkaamaan tekstissä.

Tulosteikkunan voit tallentaa myös suoraan muuhun tiedostomuotoon toiminnolla *File→Export*. Tallentuva tiedosto voi olla esimerkiksi Wordin .docx, Adoben .pdf -muotoa tai verkkoselaimessa luettava Web Report .htm.

Tulostus: SPSS:n tulosteikkunan sisällön tulostaminen paperille ym. on helppoa. Tulosteikkunassa ollessasi valitse ylhäältä valikkoriviltä vaihtoehdot *File → Print* (suoratulostus) tai *Print Preview* (saat tulostuksen ensin esikatseluun). Jos haluat, SPSS:n tulosteikkunaa ei ole pakko tulostaa kokonaan, vaan voit valita ja poimia tulostettavia kohteita edellisessä kappaleessa esitetyllä tavalla (*Shift*- ja *Ctrl*-näppäimet).

Excel-tiedoston siirtäminen SPSS:ään: SPSS lukee ongelmitta Excel-tiedostoja, jos ne ovat muodoltaan sellaisia, että niiden ensimmäisellä rivillä omissa sarakkeissaan ovat muuttujien nimet ja muilla riveillä ovat muuttujien arvot vastaavissa sarakkeissa. Muuttujien nimet on hyvä muokata jo Excelissä sellaisiksi, että ne kelpaavat SPSS:n muuttujanimiksi (alkaen kirjaimella ja ilman välilyöntejä tai -viivoja). Tällaisen tiedoston siirto tapahtuu seuraavasti: valitse SPSS:ssä ylhäältä valikkoriviltä vaihtoehdot *File* → *Open* → *Data* ja vaihda kohtaan *Save as type* tiedostotyypiksi *Excel (*.xlsx tai *.xls)*. Etsi Excel-tiedostosi ja napsauta *Open*-painiketta ja sen jälkeen *OK*-painiketta. Muunto käynnistyy ja tiedosto tulee näkyviin esikatselutilaan, josta se voidaan hyväksyä. SPSS:ssä voit tallentaa aineiston varsinaiseksi SPSS-tiedostoksi (*sav*-loppuinen tiedostonimi).

Tässä yhteydessä on vielä tarkoituksenmukaista käsitellä yleisesti uusien luokkarajojen määrittelyä ja luomista sekä sitä, miten voidaan muuttujia yhdistelemällä uusia summamuuttujia.

2.2 Uusien luokkarajojen määrittely

Käytettävässä havaintoaineistossa on kahdenlaisia muuttujia: kategorisia (esim. sukupuoli ja viriketausta) ja numeerisia, joissa muuttujan arvo voi periaatteessa saada minkä arvon tahansa muuttujan vaihteluvälin sisällä. Näin mahdollisten lukumäärien mahdollisuus on periaatteessa ääretön (esim. verbaalitestin mittausarvot tai kieliaineiden keskiarvomuuttuja). Yleensä jatkuvia muuttujia ei kannata lähteä luokittelemaan harvempiin luokkiin. Oikeaoppista tämä ei ainakaan ole, eivätkä tilastotieteilijät hyväksy tällaista ratkaisua vedoten informaation häviämiseen. Muttei tämä täysin väärinkään ole; joissain tilanteissa luokka-arvoisen muuttujan käyttö voi olla hyvinkin perusteltua.

Kasvatus- ja yhteiskuntatieteellisissä tutkimuksissa tutkimusasetelmat ovat toisinaan sen luonteisia, että numeeristen muuttujien ryhmittäminen 2–5 ryhmään antaa tutkijalle tiedon käyttäjille riittävästi ja tarkoituksenmukaista informaatiota. Luokittelun kriteerien tulee nousta muuttujien luonteesta käsin; myös aiemmat tutkimukset tai teoriat antavat tähän omat kriteerinsä. Kasvatusta tai koulutusta pohtivalle riittää usein esimerkiksi koulumenestyksen suhteen tieto siitä, miten heikosti, keskinkertaisesti tai hyvin menestyvät eroavat eri taustatekijöiden suhteen tai suhteessa esimerkiksi kouluun. Raportissa tulee luonnollisesti eri ryhmien luokkarajojen ilmetä selkeästi. Näin lukija kykenee arvioimaan luokitusten relevanttisuuden. Yhtä lailla yhteiskuntatieteilijä tutkii usein esimerkiksi ansio- ja tulojakaumia luokittelemalla aineiston muutamiin tuloluokkiin (esim. 1500–2000; 2001–3000 jne.), joita sitten hyödynnetään ristiintaulukoinneissa ja muissa analyyseissä. Tällöin hyväksytään ensinnäkin se, että osa aineiston sisältämästä informaatiosta ”menetetään” luokittelusta johtuen, ja toiseksi se, että luokkarajojen asettamiseen liittyy aina tietty määrä harkinnanvaraisuutta, jotka tuovat tulkintaan epätarkkuutta. Luokittelujen käyttöä voi osaksi edesauttaa niiden luoma konkreettisuusvaikutelma, jolloin tulosten raportoinnista tulee helpompaa ja vankemman tuntuista tiedon arkikäyttäjälle. Varsinkin aloittelevalle tukijalle tällaisten luokiteltujen muuttujien tulkinta ja raportointi voi olla helpompaa kuin jatkuvista muuttujista tehdyt tulkinnat.

Uusien luokittelurajojen määrittäminen ei ole helppoa, muutamia perusvinkkejä tähän tarjoavat Kerlingerin (1981, 137–141) esittämät luokkarajojen kriteerit:

1. luokkien (ryhmitysten) tulee perustua tutkimusongelmaan ja tavoitteisiin
2. luokkien tulee olla tyhjentäviä (exhaustive), eli tapauksia ei saa jäädä luokittelematta
3. luokkien on oltava toisensa poissulkevia (eli jos kuuluu yhteen luokkaan, ei voi kuulua toiseen)
4. muuttujien luokitusten tulee olla johdettu samasta luokitteluperiaatteesta
5. luokituksista tulee muodostua yhtenäinen kokonaisuus

Kerlingerin luokkakriteerit ovat valaisevia. Tärkeätä on huomata, että ryhmittelyyn vaikuttaa tutkimuksen konteksti sekä se, että kaikkien tutkimusaineistossa olevien tapausten on yleisesti sijoituttava johonkin uusista luokista. Uusien luokitteluryhmien tulee olla myös rajattu niin, ettei yksikään tapaus voi sijoittua kahteen ryhmään samanaikaisesti (toistensa poissulkevuuden sääntö). (Ks. enemmän Kerlingerin esitystä emt., 137–141). Näiden sääntöjen muistaminen auttaa kuitenkin ryhmien muodostamiseen liittyvien ongelmien ratkomisessa, eikä esimerkiksi luokkarajojen määrittely jää pelkän intuition varaan. Lopuksi voidaan katsoa vielä pääpiirteittäin, miten muuttujien luokittelu toteutetaan SPSS:ssä.

Kategoristen muuttujien suhteen, kuten sukupuolen kohdalla, luokittelu on selkeä; yleisesti jokainen arvo on oma ryhmänsä – naiset ovat naisia, miehet miehiä. Koodauksessa annetaan miehille esimerkiksi arvo 1 ja naisille vastaavasti 2. Näin ne erottelevat tämän muuttujan kaksi luokkaa tai ryhmää toisistaan. Toki kategoristenkin muuttujien suhteen on toisinaan yhdistettävä eri luokkia keskenään, esimerkiksi silloin, kun meille ei tule riittävästi havaintoyksiköitä kaikkiin ryhmiin, mikä on hyvin tavallista pienten aineistojen kanssa toimittaessa. Myöskään sellaisen kaksiluokkaisen muuttujan luomisessa, jossa jaetaan tapaukset sen mukaan, esiintyykö jotain suuretta tai ominaisuutta (1) tai ei (0), ei yleensä liity ongelmia. Esimerkkinä tällaisesta voisi olla luokitus koodauksella: oppilas, jolla on sisaria (1) ja oppilas, joka on perheen ainoa lapsi (0).

Ongelmallisempaa on numeeristen muuttujien uudelleen luokitus, varsinkin kun luokkia on enemmän kuin edellä mainittu kaksiluokkainen luokitus. Meillä täytyy olla joitakin kriteeriarvoja, joiden mukaan luokat muodostetaan. Ne voivat olla kiinteitä, aiemmin määritettyjä raja-arvoja, jos käytetään jotakin standardoitua testiä (esim. useat standardoidut psykologiset testit) tai voidaan asettaa tietyt luokkarajat esimerkiksi koulutodistusarvosanoista (aikaisempien tutkimusten tai tietämyksen perusteella). Luokkarajoja voi muodostaa myös aineistossa olevan numeerisen tiedon perusteella. Voisimme esimerkiksi tehdä muuttujan luokittelun kahteen luokkaan: niin, että asetamme muuttujan keskiarvon uuden muuttujan luokkarajaksi, niin että ne, joiden arvot ovat suuruudeltaan keskiarvon alapuolella sijoittuvat uudessa muuttujassa toiseen ryhmään, ja ne, joilla arvo on keskiarvon mukainen tai suurempi, toiseen ryhmään. Vastaavalla tavalla ryhmittelyn kahteen luokkaan voisi toteuttaa myös mediaanin avulla (kaksi samankokoista luokkaa), tai ala- ja yläkvartiilin avulla, jotka jakaisivat ryhmän kolmeen osaan (ala- ja yläryhmään tulisi kumpaiseenkin 25 % aineistossamme olevista tapauksista, joten keskiryhmään jäisi täten 50 %). Mediaani ja kvartiilit ovat hyvä lähtökohta varsinkin silloin, kun käytössä ei ole absoluuttisia rajoja tiedossa entuudestaan. Ongelmatonta tämä ei kaikissa tapauksissa kuitenkaan ole. Varsinkin sellaisissa tapauksissa tämä on ongelmallista, kun vastaajien saamat arvot pakkautuvat kyseisessä muuttujassa hyvin lähelle toisiaan. Silloin esimerkiksi mediaaniluokitus voi johtaa harhaiseen luokitteluun, koska käytännössä luokat eivät sisällöllisesti juurikaan poikkea toisistaan. Vastaavasti sisällöllisen tulkinnan kannalta käyttökelpoisten rajojen määrittäminen voi johtaa siihen, että toisiin luokkiin voi tulla niin vähän havaintoyksiköitä, ettei tilastollisten analyysien teko niiden avulla ole mielekästä tai edes mahdollista. (Ks. tästä problematiikasta enemmän esimerkiksi Jokivuori & Hietala 2007, 186–189.)

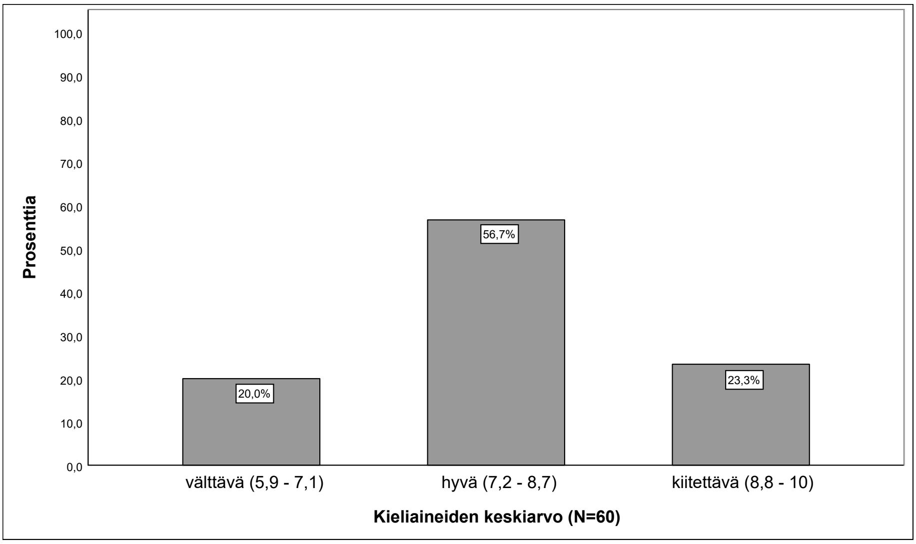
Vaikka tunnuslukujen perusteella muodostettuun luokitteluun liittyy omat ongelmansa, usein näin saatuja raja-arvoja kannattaa ainakin alustavasti tarkastella ala- ja yläkvartiilijaon avulla, jos siis aikoo luokitella muuttujan kolmeluokkaiseksi, eikä itselle ole jo selvää missä luokkien rajat kulkevat (mediaanin avulla voidaan muuttuja luokitella kahteen luokkaan). Näin ainakin saa jonkin lähtökohdan, josta lähteä arvioimaan luokittelurajojen asettamista. Tarkastellaan esimerkkinä käytetyn aineiston määrällistä muuttujaa ’kieliaineiden keskiarvo’ ja sen luokittelua kolmiluokkaiseksi (heikot, keskinkertaiset ja hyvät). Ennen luokittelua muuttujan arvojen jakauma näyttää esimerkissä seuraavalta (SPSS-tulostus):

Kieliaineiden keskiarvo

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	59,00	2	3,3	3,3	3,3
	62,00	2	3,3	3,3	6,7
	63,00	2	3,3	3,3	10,0
	67,00	2	3,3	3,3	13,3
	69,00	2	3,3	3,3	16,7
	71,00	2	3,3	3,3	20,0
	72,00	6	10,0	10,0	30,0
	73,00	2	3,3	3,3	33,3
	75,00	2	3,3	3,3	36,7
	76,00	2	3,3	3,3	40,0
	79,00	6	10,0	10,0	50,0
	81,00	2	3,3	3,3	53,3
	82,00	4	6,7	6,7	60,0
	84,00	2	3,3	3,3	63,3
	85,00	2	3,3	3,3	66,7
	86,00	4	6,7	6,7	73,3
	87,00	2	3,3	3,3	76,7
	89,00	4	6,7	6,7	83,3
	92,00	4	6,7	6,7	90,0
	96,00	4	6,7	6,7	96,7
	97,00	2	3,3	3,3	100,0
	Total	60	100,0	100,0	

Tällaisen lukumääriin perustuvan frekvenssitaulukon avulla emme pysty tekemään kovinkaan paljoa enempää päätelmiä kuin pelkästä havaintomatriisista. Teemme siis seuraavaksi kieliaineiden keskiarvo -muuttujasta kolmiluokkaisen. Samalla tutustutaan, miten ala- ja yläkvartiili määritetään muuttujan arvojen jakauman avulla. Alakvartiili on se muuttujan arvo, jonka kohdalla kumulatiivinen prosenttijakauma (Cumulative Percent) ensimmäisen kerran saavuttaa osuuden 25 prosenttia. Näemme, että alakvartiili asettuu kohtaan 72, sillä tähän arvoon liittyy osuus 30 prosenttia, kun taas edelliseen arvoon 71 liittyy osuus 20 prosenttia. Emme kuitenkaan halua, että luokittelussamme heikoimpaan ryhmään tulee liki kolmannes oppilaista, joten otamme tähän koulumenestykseltään heikoimman viidenneksen ryhmään ne oppilaat, joiden keskiarvot vaihtelevat välillä 59–71. Tarkennettakoon, että tämä siis tarkoittaa kouluarvosanojen skaalassa väliä 5,9–7,1 (todellisessa tutkimuksessa tämän heikoimman ryhmän ylärajaa todennäköisesti alennettaisiin vielä muutama kymmenys, esimerkiksi arvo 6,5 voisi olla sopiva).

Yläkvartiili on se muuttujan arvo, jonka kohdalla kumulatiivinen prosenttijakauma ensimmäisen kerran saavuttaa osuuden 75 prosenttia; siten esimerkissä yläkvartiili on 87. Jos valitaan parhaiden oppilaiden ryhmään heidät, joiden kouluarvosanojen keskiarvo ylittää tämän arvon, ryhmään "kiitettävä" tulevat ne oppilaat, joiden keskiarvot ovat 8,8 tai sen yli. Keskiryhmän muodostavat kaikki loput oppilaat, joten heidän keskiarvonsa vaihtelevat välillä 7,2–8,7. Käytetään tästä keskitasoisten ryhmästä nimeä "hyvä". Uusi kieliaineiden keskiarvo -muuttuja näyttää uudelleen luokiteltuna nyt seuraavanlaiselta (kuvattuna prosenttiosuuksin pylväsdiagrammina):



KUVIO 2. Kieliaineiden keskiarvo -muuttujan jakauma kolmiluokkaisena.

Uudesti luokiteltuun muuttujaan saatiin näin kolme luokkaa, ts. tutkimuksessa mukana olleet opiskelijat on tässä jaettu kolmeen ryhmään (välttävien, hyvien ja kiitettävien) sen mukaan, mikä heidän keskiarvonsa kieliaineissa oli. Kuviosta ilmenevät kunkin ryhmän luokkarajat, johon ryhmien luominen on perustunut. Nyt tutkijan tulisi arvioida ensinnäkin, ovatko muodostetut uudet ryhmät tarkoituksenmukaisia hänen tutkimusasetelmansa suhteen, ja toiseksi miettiä ryhmille sopivat nimet (kuviossa näkyvät luokkarajoihin liittyvät selitteet eivät välttämättä olisi havainnollisin muoto esittää luokat lopullisessa tutkimusraportissa). Näin muodostunutta uutta luokka-arvoista muuttujaa voisi analyysivaiheessa käyttää esimerkiksi ristiintaulukoinneissa (ks. luku 6.1, s. 165). Esimerkissä huomattiin, että tehty luokittelu ei aivan tarkasti toteuta ylä- ja alakvartiilin 25 prosentin osuuksia, mutta tämä ei ole yleensä erityisen ongelmallista. Käytännössä kvartiilit on harvoin mahdollista määrittää niin, että luokittelun tuottamat osuudet olisivat täsmälleen 25 - 50 - 25 prosenttijakauman mukaiset. Yleensä kuitenkin esimerkin kaltainen tarkkuus riittää. Lisäksi tässä saatu ryhmittely on tulkinnaltaan mielekäs ja saadut ryhmät vaikuttavat sisällöllisesti (kielitaidoltaan) toisistaan eroavilta.

Aina ei ala- ja yläkvartiili -jako kuitenkaan anna niin käyttökelpoisia luokkarajoja kuin yllä saatiin. Tällöin tulee käyttää harkintaa tai entisiä mittarikriteereitä tai vastaavia apuna määritettäessä loogisesti perusteltuja luokkarajoja. Aina ei myöskään ole tarkoituksenmukaista jakaa muuttujan arvoja väkisin ala- ja yläkvartiilien mukaisiin luokkiin. Jos aineistossa esimerkiksi on heikkoja tai välttäviä oppilaita kovin vähän, ei välttämättä ole kovin tarkoituksenmukaista tehdä heistä 25 prosentin suuruista ryhmää keinotekoisesti. Luokituksessa kannattaa käyttää siis harkintaa. Kuitenkin mediaanin tai kvartiilien mukainen muuttujan arvojen luokittelu on yksi vaihtoehto, jota voi pohtia siihen liittyvistä ongelmista huolimatta.

SPSS: Määrällisen muuttujan luokittelu kategoriseksi eli luokka-arvoiseksi

Tarkastellaan esimerkiksi, miten luokitellaan määrällisen muuttujan KIELIAKA (kieliaineiden keskiarvo -muuttuja) kolmiluokkaiseksi: muodostetaan luokat 1) 71 tai alle, 2) 72–87 ja 3) 88 tai yli.

1. Valitse valikkoriviltä *Transform → Recode Into Different Variables*
   - näin valittaessa myös alkuperäinen muuttuja jää talteen havaintomatriisiin (näin varmasti yleisesti kannattaa toimia)
   - siirrä valitsemasi numeerinen muuttuja muuttujaluettelon oikealla puolella olevalla nuolipainikkeella kohtaan *Numeric Variable → Output Variable*
   - anna uudelle muuttujalle jokin haluamasi nimi kohdassa *Name* ja paina *Change* -painiketta
2. Painikkeella *Old and New Values* pääset määrittelemään uuden muuttujan luokkarajat
   - rastita siellä valinta *Range: lowest through*, ja kirjoita vieressä olevaan tyhjään tilaan luku 71 – tämä tarkoittaa siis sitä, että kaikki luvut arvoon 71 saakka tulevat sijoittumaan ensimmäiseen luokkaan
   - valitse kohdassa *New Value* vaihtoehto *Value* ja kirjoita sen vieressä olevaan tyhjään tilaan numero 1 ensimmäisen luokan tunnukseksi ja paina *Add*-painiketta
   - ruksaa seuraavaksi valinta *Range: through* ja kirjoita vieressä olevaan ensimmäiseen tyhjään tilaan luku 72 ja toiseen tyhjään tilaan luku 87. Tämä vuorostaan tarkoittaa sitä, että kaikki luvut väliltä 72–87 (kyseiset luvut mukaan lukien) tulevat toiseen luokkaan

ohjetaulukko jatkuu...

- valitse jälleen kohdassa *New Value* vaihtoehto *Value* ja syötä vieressä olevaan tyhjään tilaan numero 2 toisen luokan tunnukseksi ja paina *Add*-painiketta
- rastita vielä samat vaiheet kolmannelle luokalle: valitse *Range: through highest* ja kirjoita vieressä olevaan tyhjään tilaan luku 88. Silloin kaikki alkuperäisen muuttujan luvut arvosta 88 lähtien tulevat kuulumaan kolmanteen luokkaan
- kirjoita lopuksi *New Value* vaihtoehdon *Value* vieressä olevaan tyhjään tilaan numero 3 kolmannen luokan tunnukseksi ja paina *Add*-painiketta
- hyväksy tehdyt valinnat klikkaamalla *Continue*-painiketta

3. *OK*-painikkeella toteutat luokittelun ja saat uuden luokitellun muuttujan havaintomatriisin loppuun (oikeaan reunaan). Samalla avautuu tulosteikkuna, jossa näkyvät tekstimuodossa tiedot toteutetusta luokittelusta.

Tässä on käsitelty muuttujien luokittelu melko yksityiskohtaisesti, koska käytännössä tämä vaihe, kuten muutkin varsinaisen analyysin valmistelutehtävät, tuottavat aloittelijalle usein hankaluuksia. On kuitenkin huomattava, että edellä esimerkinomaisesti esitetty muuttujan luokittelu ei välttämättä ole lainkaan tarpeen, vaan koulumenestysmuuttujaa voisi, ja olisi hyväkin, käyttää sellaisenaan määrällisenä muuttujana varsinaisissa analyyseissä. Muita alustus- tai matriisin muokkaustehtäviä ovat esimerkiksi muuttujien muunnokset, muuttujien arvojen nimeäminen ja mahdollisesti eri havaintoaineistojen ja matriisien yhdistäminen. Lohdutukseksi voidaan taas sanoa, että kun tämän kaltaiset alustavat vaiheet on saatu tehdyksi, melkein koko tilastovaiheen mekaaninen osuus on nyt tehty – kohta kädessäsi on lukematon määrä erilaisia tulosteita, joita voit alkaa tulkita. Seuraavaksi, ennen varsinaista aineiston tarkastelua ja tilastoanalyyseja, katsotaan vielä summa- tai lisämuuttujan muodostaminen, joka tulee lähes jokaisessa tutkimuksessa ajankohtaiseksi jossakin kohden.

2.3 Summa- tai lisämuuttujan muodostaminen

Summamuuttujan laatiminen on hyvin yleistä kasvatus- ja yhteiskuntatieteellisissä tutkimuksissa (ks. esim. Nummenmaa 2011, 161–162; Metsämuuronen 2009, 112, 535–537; ks. myös Vehkalahti 2014, 112–116, jossa hän esittää perinteistä summamuuttujaa kohtaan tiettyä kritiikkiä). Summamuuttajan muodostamisen avulla voidaan vähentää käsiteltävien muuttujien määrää. Summamuuttujan avulla tiivistetään aineistossa olevia muuttujia, jotka mittaavat samaa asiaa tai saman asian eri ulottuvuuksia. Ajatellaan vaikkapa tilannetta, jossa meillä on 40 muuttujaa, jotka mittaavat minäkuvan eri alueita. Olkoon näitä alueita neljä (esim. ammatti-, sosiaalinen, emotionaalinen ja fyysinen minäkuva), ja kutakin aluetta on mitattu kymmenellä muuttujalla. Ei ole tarkoituksenmukaista, että käsitellään kunkin minäkuvan alueella kutakin kymmentä muuttujaa erikseen, vaan on järkevää muodostaa neljä tai oikeastaan viisi summamuuttujaa, joista kukin kuvaa eri minäkuvan ulottuvuutta ja viimeinen, viides, kuvaa vaikkapa yleisminäkuvaa, johon otetaan mukaan kaikki 40 muuttujaa. Näin saadaan tiivistetyksi muuttujamäärän viiteen ja saadaan aineisto tulkittavampaan muotoon, vaikka luonnollisesti hävitetään samalla jonkin verran alkuperäistä informaatiota. Usein aineiston analyysivaiheessa kannattaa vielä taulukoida erikseen summamuuttujan muodostavista muuttujista frekvenssit ja prosenttiosuudet tai keskiarvot, jolloin jokin erityiskysymys voi nousta joissakin tilanteissa – ei siis läheskään aina – tärkeäksi lisäanalyysin kohteeksi, jonka avulla voidaan tehdä syvempiä ja vahvempia tulkintoja kuin yksinomaan summamuuttujaa käyttämällä.

Matriisissa on aina tietysti sellaisia muuttujia, kuten vanhempien koulutustaso ja tutkittavan viriketausta (kategorisia muuttujia), jotka korreloivat keskenään voimakkaasti, mutta silti näitä ei voi eikä kannata yhdistää yleisemmäksi taustamuuttujaksi. Näiden osalta kannattaa joko valita, kumpaa muuttujaa käyttää yleisanalysoinneissa, tai, jos tutkimusasetelma edellyttää näiden molempien taustamuuttujien käyttöä, käytetään niitä analyyseissä erillisinä muuttujina.

Summamuuttujien käyttöä voidaan puolustaa sisällöllisen perustelun lisäksi myös tilastollisin perustein. Keskeisen raja-arvolauseen perusteella voidaan todeta, että mitä useammasta muuttujasta summamuuttuja koostuu, sitä paremmin sen jakauma noudattaa normaalijakaumaa, joka on edellytys monen tilastollisen analyysimenetelmän käytölle. Lisäksi summamuuttujan mahdollisten arvojen lukumäärä on paljon suurempi kuin niiden muuttujien, joista se koostuu. Siten summamuuttuja, joka koostuu viisiluokkaisista Likert-asteikollisista muuttujista, muistuttaa paljon enemmän numeerista muuttujaa kuin yksittäinen viisiluokkainen Likert-tyyppinen muuttuja. Mittarin luotettavuuden eli reliabiliteetin voidaan myös olettaa paranevan, kun sen arvo perustuu useaan samaa asiaa mittaavaan muuttujaan.

Yhden lisämuuttujan sisälle tulevien muuttujien pitää luonnollisesti kuvata samaa ilmiötä. Jos lisämuuttuja sisältää kovin erityyppisiä tai pahimmassa tapauksessa eri asiaa mittaavia muuttujia, ei muodostettu summamittari ole kovin relevantti eikä luotettava. Tätä voidaan tutkia esimerkiksi kunkin summamuuttujan yksittäisten muuttujien välisten korrelaatioiden kautta joko keskenään (korrelaatiomatriisi) tai summamuuttujan summaan nähden (osioanalyysin avulla), katso mittarin reliaabeliuden testaamisesta luku 2.4. Tällaisten tarkastelujen tuloksena voidaan poistaa mittarista sellaiset yksittäiset muuttujat, joiden korrelaatiot jäävät kovin alhaisiksi. Tarkoituksena on siis luoda mahdollisimman homogeeninen summamuuttuja ja tällä tavoin voidaan parantaa mittarin reliabiliteettia.

Esimerkkiaineistosta voidaan muodostaa kaksi summamuuttujaa: järkeilytestin pistemäärät ja matematiikan numerot on yhdistetty loogisuus-matemaattisuus -muuttujaksi, ja kieliaineiden keskiarvot ja verbaalisen testin pistemäärät on yhdistetty kielellisyysmuuttujaksi. Tähän palataan vielä tuonnempana näiden muodostamisen perusteisiin pääkomponenttianalyysin käsittelyn yhteydessä. Näin saadaan typistetyksi muuttujamäärän neljän sijasta kahteen, joita voidaan käyttää jatkoanalyyseissä. Kun muuttujia on alkujaan näin vähän, käytännössä tällaisia summamuuttujia ei kannattaisi luoda, mutta tässä tämä on esitetty esimerkkinä summamuuttujien luomismahdollisuudesta. Sitä ennen on kuitenkin kiinnitettävä huomio kahteen tärkeään seikkaan, joiden kanssa saattaa tulla jatkotarkasteluissa vaikeuksia: A) Koska esimerkissä kumpikin summamuuttuja muodostuu kahdesta erillisestä muuttujasta, tulevat summamuuttujan pistemäärät suuremmiksi kuin yksittäisten muuttujien pistemäärät. Tämä saattaa vaikeuttaa summamuuttujan arvojen tulkintaa. B) Lisäksi tilanne on se, että molemmat uudet summamuuttujat rakentuvat muuttujista, joiden alkuperäiset skaalat eivät ole yhteneväisiä.

> <u>Ongelma A</u>. Tämä ongelma on helppo ratkaista, jos summamuuttujaan kuuluvien muuttujien arvojen vaihteluväli on sama, esimerkiksi jos kaikki summamuuttujaan tulevat muuttujat ovat arvosanoja, jotka voivat saada arvoja väliltä 4...10. Jos summamuuttuja muodostetaan siten, että kyseisten muuttujien summa jaetaan muuttujien lukumäärällä, silloin summamuuttujankin arvot voivat olla vain väliltä 4...10. Näin summamuuttujan arvot on helppo ymmärtää; niiden tulkinta on täsmälleen sama kuin niiden muuttujien, joista se koostuu. Näin muodostettua summamuuttujaa voidaan kutsua myös keskiarvomuuttujaksi.
>
> Jos muuttujien arvojen vaihteluvälit eivät ole samat, silloin ne on skaalattava samoiksi, joten tällöin on siis ratkaistava ongelma B.
>
> <u>Ongelma B</u>. Kuvitellaan, että ollaan tekemässä summamuuttujaa, johon tulevista arvosanamuuttujista osa on mitattu asteikolla 4...10, mutta osassa onkin käytetty asteikkoa 1...5. Jos haluamme, että summamuuttujan tulkinta säilyy järkevänä, silloin jommankumman

asteikon arvot on muunnettava toista asteikkoa vastaaviksi. Jos esimerkiksi halutaan muuttaa arvot 1...5 (muuttuja x_5) asteikolle 4...10 (muuttuja x_{10}), käytämme muunnoskaavaa:

$$x_{10} = \frac{3x_5 + 5}{2}$$

eli jos sijoitetaan x_5:n paikalle luvun väliltä 1...5, saamme sitä vastaavan luvun x_{10} asteikolta 4...10. Käytännössä arvojen laskeminen toteutetaan tilasto-ohjelmistolla, kunhan kerrotaan ohjelmistolle, millaista laskulauseketta muunnoksessa halutaan käyttää. Kun tämän jälkeen toimitaan ongelmassa A esitetyllä tavalla, saataisiin luotua summamuuttuja, jonka arvot voivat vaihdella vain välillä 4...10, joten tulkintaongelma on ratkennut.

Esimerkissä olisi siis menetelty niin, että aluksi olisi muutettu loogisuus-matemaattisuus -summamuuttujaan kuuluvat kaksi muuttujaa yhteismitallisiksi sopivalla muunnoskaavalla ja vasta sen jälkeen olisi tehty itse summamuuttuja näiden kahden siihen kuuluvan muuttujan keskiarvona. Samalla tavoin olisi toimittu kielellisyys-summamuuttujan kanssa.

Kun summamuuttuja on tehty, sen jälkeen sillä voi operoida samaan tapaan kuin millä tahansa numeerisella muuttujalla. Lisäksi sen voi tarvittaessa luokitella sopiviin luokkiin (ks. luokkarajojen määrittelyä edellä), jolloin sitä voi käsitellä myös kategorisen muuttujan tapaan. Summamuuttujien käyttömahdollisuudet ovat siis erittäin laajat.

SPSS: Summamuuttujan muodostaminen aineistoon uudeksi muuttujaksi

1. Valitse valikkoriviltä *Transform → Compute Variable*
2. Kirjoita kohtaan *Target Variable* jokin sopiva nimi uudelle summamuuttujalle
   - nimensä mukaisesti summamuuttuja on joidenkin havaintomatriisissa olemassa olevien muuttujien summa
3. Siirrä ensimmäinen valitsemasi muuttuja muuttujaluettelon oikealla puolella levalla nuolipainikkeella kohtaan *Numeric Expression* ja paina kohdan alla olevan laskinnäppäimistön painiketta +. Samaan tapaan lisäät muut summamuuttujaan kuuluvat muuttujat, paitsi että viimeisen lisäyksen jälkeen ei enää tarvitse painaa + -painiketta
4. *OK*:n painallus aloittaa summamuuttujan teon ja lisää summamuuttujan havaintomatriisin loppuun. Samalla avautuu myös tulosteikkuna, johon tulee tekstimuodossa tiedot uuden muuttujan muodostamisesta

ohjetaulukko jatkuu...

HUOM: Jos haluat, että summamuuttujan arvojen tulkinta säilyy samantapaisena kuin niiden muuttujien, joista se koostuu, tee summamuuttuja olemassa olevien muuttujien keskiarvona. Tällöin muodostamasi summamuuttuja jaetaan sen muodostavien muuttujien lukumäärällä. Varmista tällöin kohdan 3 lausekkeessa kaarisuluilla ( ), että keskiarvo lasketaan oikein eli summattava lauseke lasketaan ensin sulkeiden sisällä ja tämä summa jaetaan väittämien lukumäärällä. Kaarisulutkin löydät kohdassa 3 mainitusta ruudulla näkyvästä laskinnäppäimistöstä. Keskiarvon muodostamisessa kaikilla summamuuttujaan kuuluvilla muuttujilla tulisi olla sama skaala, esimerkiksi että kaikki muuttujat voivat saada arvoja väliltä 1…5. Valintoja tehdessä kannattaa käyttää valintaruudukon alapuolella olevan laskimen toimintonäppäimiä. Näin aiottu laskukaava tulee varmemmin oikein laadituksi. Muista myös, että uuden muuttujan nimessä (Target Variable) ei saa olla välilyöntejä tai erikoismerkkejä.

Valintaikkunassa *Function group* on valittavissa myös valmiita funktioita. Valitse ikkunasta esim. *Statistical*, alapuoliseen ruudukkoon tulee erilaisia valintavaihtoehtoja. Valitse nyt esimerkiksi funktio *Sum*, (ohjelmaa muodostaa muuttujista summaan perustuvan uuden muuttujan) tai funktion *Mean* (ohjelma muodostaa mukaantulevien muuttujien keskiarvon mukaisen uuden muuttujan). Tällöin laskulauseketta ei tarvitse rakennella alusta alkaen itse. Lisäksi valmiisiin funktioihin on lisäoptioita, joilla voidaan vaikuttaa mm. puuttuvien tietojen käsittelyyn.

2.4 Mittarin luotettavuus ja sen arviointi

Tieteellisen tiedon yksi keskeinen perusvaatimus on, että sen hankinta on systemaattista ja kontrolloitua. Näin on erittäin tärkeää, että tutkimuksessa käytetyt mittarit mittaavat sitä, mitä niiden on tarkoitettukin mittaavan (validiteetti) ja ettei mittari tuota sattumanvaraisia tuloksia (reliabiliteetti). Mittarin pätevyys onkin yksi tutkimuksen kulmakivistä. Kasvatustieteellisissä tutkimuksissa on tyypillistä muiden yhteiskuntatieteiden ja käyttäytymistieteiden tapaan se, että tutkimuksen kohteena olevaa ilmiötä ei voi mitata yksiselitteisesti ja suoraan. Usein erilaiset käsitteet, kuten motivaatio tai asennekysymykset, ovat luonteeltaan abstraktisia ja myös teoreettisia käsitteitä. Tämän luontoisten käsitteiden mittaaminen edellyttää niiden muuttamista empiirisiksi, eli ne on operationalisoitava mitattavaan muotoon. Usein joudutaan luomaan myös edellä esitettyjä summamuuttujia, jotka mittaavat eri ulottuvuuksia tutkittavasta ilmiöstä.

Reliabeliteetti, yleistä:

- reliabiliteettiasetelmassa tarkastellaan käytetyn mittarin stabiilisuutta
- reliabiliteettikerroin ilmaisee, kuinka suuri osa mittariin kuuluvien muuttujien arvojen vaihtelusta perustuu tarkasteltavan ilmiön todelliseen vaihteluun (varianssiin) eikä virhevarianssiin
- reliabiliteettikertoimen avulla voidaan arvioida sitä, ovatko esimerkiksi asenne- tai minäkuvamittarin skaalat, jotka muodostuvat useasta eri kysymyksestä, sisäisesti yhteneväisiä eli homogeenisia, kyse on siis tässä summamuuttujan reliabiliteetin arvioimisesta.

- vielä tarkennukseksi reliabiliteetti ilmaisee ennen kaikkea sen, kuinka hyvin mittari mittaa todellisen vaihtelun osuuden, ja toisin päin sen, kuinka suuri osuus tuloksesta johtuu satunnaisista mittausvirheistä (jäännösosuudesta)
- osioanalyysin (item analysis) avulla voidaan arvioida edellä mainitun tapaisten summamuuttujien osioiden korrelaatioita. Mittaria voidaan parantaa esimerkiksi poistamalla siitä kysymykset tai väittämät, jotka eivät korreloi tehtävään kokonaissummamuuttujaan, ja parantaa näin lopullisen mittarin reliabiliteettiä. Oletettavasti tällaiset huonosti korreloivat muuttujat eivät mittaa samaa asiaa kuin muut mittariin kuuluvat muuttujat
- reliabiliteetin arvioimiseksi on kehitetty erilaisia suureita, reliabiliteettikertoimia. Yleisimmin ne saavat arvoja väliltä 0–1. Mitä lähempänä arvo on ykköstä, sen homogeenisempi mittari on, eli sitä vähemmän se sisältää virhevarianssista johtuvaa vaihtelua
- reliabiliteettitarkasteluissa muuttujien tulee olla luonnollisesti numeerisia
- reliabiliteettikertoimeen vaikuttavat esimerkiksi aineiston koko ja mittarin sisältämien kysymysten tai väittämien määrä: kysymysten määrä lisää usein reliabiliteettikertoimen arvoa. Joskus mittarissa (summamuuttujassa) olevien kysymysten vastausarvoissa on hyvin vähän vaihtelua, mikä saattaa laskea reliabiliteettiarvoa alemmaksi kuin se muuten olisi.

Reliabiliteetti liittyy laajasti määritellen mittarin, mittaustilanteen ja mittauskohteen pysyvyyteen. Suppeammin määriteltynä nämä käsitteet viittaavat mittaustuloksen pysyvyyteen eli käytetyn mittarin kykyyn tuottaa mittaustuloksia, jotka perustuvat todelliseen tutkittavaan ilmiöön. Kerlinger (1981, 442) yhdistää reliabiliteetin seuraaviin käsitteisiin, jotka konkretisoivat kompaktisti mistä on kyse: näitä ovat luotettavuus (dependability), pysyvyys (stability), yhdenmukaisuus tai johdonmukaisuus (consistency), ennustuskykyisyys (predictability) ja tarkkuus tai paikkansapitävyys (accuracy). Nämä käsitteet kuvaavat ilman tarkempaa selittämistäkin hyvin, mitä reliaabeliusvaateella tavoitellaan. Toisaalta on tärkeä pitää mielessä, ettei reliabiliteettikerroin ole sama kuin virheettömyys: vaikka mittarin reliaabeliusarvo olisi korkeakin, mit-

tarin skaalaus ei välttämättä ole tutkittavan kysymyksen suhteen ”skaalattu” tarkoituksenmukaisesti, eikä se näin mittaa sitä, mitä sen on tarkoitettu mittaavan. Mittarin reliaabeliuden kannalta on tärkeää pohtia myös sitä, millaisilla kysymyksillä tai havainnoinnilla tullaan keräämään tietoa tutkittavilta. Jos luotetaan esimerkiksi tutkittavan muistiin omasta ansiotasosta 10 vuotta sitten, voi muisti helpostikin pettää yhtälailla kuin havainnoitsijan tekemät virheelliset tulkinnat toimijan tarkoitusperistä tai tuntemuksista (ns. mittausvirheen aiheuttajia). Kyse mittarin reliabeliteetistä ei siis liity pelkästään laadittavan mittarin eri osioiden väliseen korrelaatiotasoon (mittarin osioiden sisäiseen koheesioon), kuten tämä tarkastelu tavallisesti rajataan tutkimusraporteissa. Fielding ja Gilbert (2006, 11) huomauttavatkin osuvasti: ”ovat valitsemamme indikaattorit kuinka valideja ja reliaabeleita tahansa, mittausvirhe on salakavala, joten tämä tulee pitää mielessä, kun analysoimme yhteiskunnallista dataa”. Mittarin reliabiliteettitarkastelussa tuleekin muistaa se, 1) ettei mitään eksaktia raja-arvoa tälle voida määritellä (vaikka yleisesti mittarin sisäisen koheesion kannalta on sitä parempi, mitä korkeampi muodostettavan uuden mittariosion väliset korrelaatiot ovat), 2) että mittarien reliaabeliudella on tutkijalle vain välinearvo, joskin tärkeä sellainen, tutkija on ennen kaikkea kiinnostunut siitä, että käytetyllä mittarilla saadaan esille se, mitä tutkimuksella tavoitellaan ja 3) että on tärkeää, että mittarin reliaabeliuutta ja validiutta tarkastellaan suhteessa tehtäviin tulkintoihin. (Ks. reliaabeliteetista, ja myös validiteetista lisää esim. Nummenmaa, Konttinen, Kuusinen & Leskinen 1997, 201–209; Bhattacherjee 2012, 55–61; Ketokivi 2015, 98–130).

Reliabiliteettikerroin vaihtelee käyttäytymis- tai sosiaalitieteissä tutkimuksesta toiseen muun muassa siitä syystä, että mittariin kuuluvien muuttujien lukumäärä jo sinällään vaikuttaa kertoimen suuruuteen. Siten sen suuruudelle ei voi määritellä absoluuttista ehdottoman oikeaa kriteeriarvoa, kuten edellä jo todettiin. Jonkinlaisia suosituksia tähän voidaan kyllä antaa: yleensä lähdetään siitä, että standardoiduissa testeissä arvo on korkeampi kuin itse laadituissa mittareissa. Näissä tilanteissa kirjallisuudessa useimmin esiintyvä raja-arvo hyväksyttävälle alphalle on, että alphan tulisi olla arvoltaan vähintään 0,70, jotta mittarin reliabiliteetti olisi riittävällä tasolla. Jos omissa, itse laadituissa mittareissa pääsee arvojen 0,60 ja 0,85 välimaastoon, niin ollaan jo aika hyvällä perustalla. Itse asiassa reilusti arvon 0,90 ylittävät kertoimet eivät välttämättä ole tavoittelemisen arvoisia. Hyvin suuri kertoimen arvo voi olla merkki siitä, että kerrointa on hinattu ylöspäin puoliväkisin lisäämällä mittariin erittäin suuri määrä muuttujia. Myös mittarin ekonomisuus on tärkeä! Sekin tilanne, että muuttujia on vähän ja kerroin siitä huolimatta lähentelee arvoa yksi, voi olla ongelma: silloin ollaan yleensä tekemisissä jo liian samanlaisten muuttujien kanssa.

Mittarin reliabiliteettia voidaan arvioida esimerkiksi uusinta- tai rinnakkaismittausten avulla: voidaan verrata, saavatko kaksi tai useampi arvioitsija samanlaisia tuloksia, puhutaan niin sanotusta arvioitsijayksimielisyydestä, tai voidaan verrata eri mittausten tulosten vastaavuutta. Ehkä yleisin kasvatustieteessä käytetty reliaabeliusmittaus liittyy mittarin tai sen osioiden sisäiseen homogeenisuuteen. Tätä voidaan mitata useilla testeillä, joista tärkein lienee Cronbachin alfa -kerroin (α). Jos mittariin kuuluvat muuttujat ovat dikotomisia 0/1-muuttujia, siinä erikoistapauksessa Cronbachin alfa on sama kuin KuderRichardsonin 20 -kerroin. Alfa-kertoimen tulkinta on helppoa. Mitä lähempänä kertoimen arvo on ykköstä, sitä yhdenmukaisempia mittariin kuuluvat muuttujat tai kysymykset keskenään ovat. Sen sijaan kertoimen lähestyessä nollaa muuttujista ei muodostu homogeenista kokonaisuutta. Kerroin voi siis saada arvoja nollasta ykköseen. Mitä parempi mittari on, sitä vähemmän se sisältää sattumaan perustuvaa ainesta eli sen tarkemmin se mittaa sitä, mitä on tarkoituskin.

Cronbachin alfa -kertoimen rinnalla käytetään usein myös puolitusreliabiliteetti (split-half) -kerrointa. Tämäkin kerroin voi saada arvoja nollasta ykköseen. Siinä mittariin kuuluvat muuttujat jaetaan kahteen ryhmään ja kertoimen arvo lasketaan näiden kahden ryhmän välisen korrelaation avulla. Ajatus on siis se, että jos mittarin kaikki muuttujat mittaavat samantyyppistä asiaa, silloin niistä muodostettujen ryhmienkin tulee mitata samaa asiaa eli korreloida keskenään. Puolitus-reliabiliteettikertoimen ongelma on kuitenkin se, että mittariin kuuluvat muuttujat voidaan jakaa ryhmiin monella tavalla. Jokainen jakotapa tuottaa erilaisen puolitusreliabiliteettikertoimen, ja kerroinarvo voi vaihdella paljonkin.

Katsotaan esimerkin avulla, miten mittari rakennetaan. Aikomus on muodostaa matemaattisuus-loogisuutta kuvaava mittari sopivista esimerkkiaineiston muuttujista. Muuttujavalikoima aineistossa on tosin aika rajoitettu, joten tämän vuoksi joudutaan valitsemaan mukaan myös sellaisia muuttujia, joista suoralta kädeltä voi sanoa, etteivät ne kuulu mittariin. Siitä, että voidaan etukäteen aavistella, mitä tuleman pitää, on kuitenkin se etu, että sillä tavoin päästään paremmin selvyyteen tämän menetelmän perusideoista. Reliabiliteettitarkastelun käytännön SPSS-toteutus on kuvattu tämän luvun lopussa. Valitut muuttujat ovat järkeilytestin pistemäärä, kieliaineiden keskiarvo, matematiikan arvosana ja verbaalisuustestin pistemäärä. Koska mittareiden teko perustuu muuttujien välisiin korrelaatioihin, hyvä tapa aloittaa mittareiden rakentaminen on tarkastella mittariin ehdolla olevien muuttujien välisiä korrelaatioita. Käytetyt muuttujat ovat samat, joita tarkastellaan myös faktori- ja pääkomponenttianalyysiesimerkin yhteydessä (luku 8, s. 213), ja siellä on esitetty kyseisiä muuttujia kuvaava korrelaatiomatriisi (s. 220). Näistä korrelaatioista nähdään, että joukossa on myös monta alhaista, lähellä nollaa olevaa, korrelaatiota. Jo tästä voidaan päätellä, että mittariin tulee liittymään joitakin ongelmia.

Seuraavaksi onkin jo itse analyysin vuoro. SPSS:n avulla tehdyt reliabiliteettitarkastelut antavat seuraavanlaisen tulosteen:

Reliability Statistics

Cronbach's Alpha	N of Items
,433	4

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
Järkeilytesti (pisteet)	112,4333	161,945	,177	,446
Kieliaineiden keskiarvo (x10)	67,6667	16,972	,593	,530
Matematiikan numero	140,1667	168,819	,119	,470
Verbaalitesti (pisteet)	121,9333	116,131	,925	,072

Cronbachin alfa -kertoimen arvo on suuruudeltaan $\alpha = 0{,}43$, eli melko matala, joten ainakaan näistä kaikista neljästä muuttujasta ei synny riittävän hyvää mittaria. Onneksi ohjelma on tulostanut myös lisäinformaatiota, jonka avulla pystytään päättelemään, miten mittaria voisi parantaa. Ensinnäkin sarakkeessa *Corrected Item-Total Correlation* on kerrottu, miten kukin neljästä muuttujasta korreloisi siihen summamuuttujaan tai mittariin, joka olisi muodostettu muista kolmesta muuttujasta. Ne paljastavat mittarin muuttujien kaksijakoisuuden; kielellisyysmuuttujilla on korkeat korrelaatiot, matemaattisuus-loogisuus -muuttujilla heikot. Joten jos hyväksyttäisiin mittari sellaisenaan, se olisi kaiken lisäksi enemmänkin kielellisyysmittari kuin tavoiteltu matemaattisuus-loogisuus -mittari.

Mittaria voi parantaa kahdella tapaa: joko lisäämällä siihen uusia tai poistamalla siitä heikosti toimivia muuttujia. Uusia muuttujia ei esimerkkiaineistosta voida tähän lisätä, joten toteutetaan jälkimmäinen vaihtoehto. SPSS-tulosteen sarakkeessa *Alpha if Item Deleted* onkin valmiiksi kerrottu, miten Cronbachin alfa -kertoimen arvo muuttuisi, jos mikä tahansa neljästä muuttujasta poistettaisiin mittarista, eli jos mittarissa olisikin kolme muuttujaa tai osiota (*Item*). Eniten kerroin paranisi, jos mukana olevista muuttujista poistettaisiin ”kieliaineiden keskiarvo” -muuttuja; silloin alfa arvo nousisi jonkin verran (α = 0,53). Poisto kannattaa tehdä, koska se samalla muuttaa jäljelle jäävien muuttujien keskinäisiä suhteita. Tehdään nyt uusi analyysi:

Reliability Statistics

Cronbach's Alpha	N of Items
,530	3

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
Järkeilytesti (pisteet)	32,7000	8,485	,492	,188
Matematiikan numero	60,4333	12,046	,490	,377
Verbaalitesti (pisteet)	42,2000	7,417	,221	,790

Kuten yllä olevasta uudesta reliabiliteettitulosteesta nähdään, Cronbachin alfa todellakin nousi arvoon α = 0,53. Lisäksi uusi asetelma on muuttanut jäljelle jääneiden muuttujien keskinäisiä suhteita. Ensinnäkin sarakkeesta *Corrected Item-Total Correlation* todetaan, että verbaalitesti-muuttuja korreloi heikoimmin muiden muuttujien muodostamaan summamuuttujaan. Nyt mittari mittaisikin jo pääasiassa matemaattisuus-loogisuutta. Saraketta *Alpha if Item Deleted* tarkastelemalla voidaan kuitenkin todeta, että mittarin sisäinen homogeenisuus nousee vielä olennaisesti, jos poistetaan myös verbaalisuustestin pistemäärämuuttuja. Silloin uuden mittarin, joka muodostuu siis järkeilytestin pistemäärästä ja matematiikan arvosanasta, Cronbachin alfa -kerroin nousee jo arvoon α = 0,79, jota voidaan jo pitää varsin hyvänä. Näin on saatu selville lopullisen mittarin rakenne, eli enää sitä ei tarvitse välttämättä parantaa. Tietenkin jos oltaisiin oikean mittarin kanssa liikkeellä, uutta mittaria tuskin rakennettaisiin näiden kahden muuttujan varaan, vaan tutkimusta varten kerättäisiin uusi havaintoaineisto, jossa matemaattisuus-loogisuuteen liittyviä muuttujia olisi enemmän kuin tässä käytetyssä pienessä esimerkkiaineistossa on.

Reliabiliteettikäsite on siis keskeinen mittarin ja täten myös tehtyjen mittausten luotettavuuden käsittelymekanismi, mutta myös tässä suhteessa tulee välttää sen mekaanista käyttöä. Vaikka esimerkissä saatiinkin rakennettua järkevän tuntuinen mittari, tulee muistaa, että ohjelma laskee eri muuttujien välisiä yhteyksiä kaavamaisesti. Joissakin tilanteissa analyysi voi jättää

malliin muuttujan, joka järkevästi ajateltuna sisältönsä puolesta ei mitenkään kuulu kokonaisuuteen (tosin ei tällaisia muuttujia pitäisi ottaa alun alkaenkaan mukaan). Tutkijalle jääkin aina viime kädessä vastuu uuden mittarin tai summamuuttujan loogisuudesta ja tarkoituksenmukaisuudesta. Tilastolliset analyysit tarjoavat tähän vain apuvälineitä. Muutenkin on hyvä muistaa, ettei mittarin luotettavuus yksistään riitä takaamaan tutkimustulosten ja niiden tulkinnan tarkoituksenmukaisuutta. Pätevä mittarikaan ei auta, jos se mittaa jotakin muuta kuin sitä, mitä meidän pitäisi tutkia (validiuskysymys), eli mittauksen reliabiliteetti on kyllä keskeinen ehto validiteetin toteutumiselle, mutta eksaktia taetta tästä tämäkään ei anna. Kuitenkin tutkimuksen validiteetin kannalta on olennaisen tärkeää se, että tutkimus rakentuu mahdollisimman hyvään mittariin tai kyselylomakkeeseen. Näiden luomiseen siis kannattaa panostaa aikaa ja energiaa.

SPSS: Mittarin reliabiliteetin tutkiminen

Cronbachin alfa -suureen laskeminen.

1. Valitse ylhäältä valikkoriviltä *Analyze → Scale → Reliability Analysis*
   - siirrä valitsemasi numeeriset muuttujat muuttujaluettelon oikealla puolella olevalla ylemmällä nuolipainikkeella kohtaan *Items*
2. *Statistics*-painike
   - ruksaa avautuvasta valikosta ainakin kohdan *Descriptives for* vaihtoehto *Scale if item deleted*. Tämä valinta tuottaa tulosteen, jonka avulla pystyt tarkastelemaan yksittäisten muuttujien yhdenmukaisuutta ja toimivuutta muuttujakokonaisuudessa
   - hyväksy valinnat klikkaamalla *Continue*-painiketta
3. Toteuta reliabiliteettianalyysi yllä määrittelemilläsi valinnoilla painamalla *OK*-painiketta ja saat analyysin tulosteet tulosteikkunaan.

Puolitusreliabiliteetin laskeminen:

1. Valitse ylhäältä valikkoriviltä *Analyze → Scale → Reliability Analysis*
   - siirrä valitsemasi numeeriset muuttujat muuttujaluettelon oikealla puolella olevalla ylemmällä nuolipainikkeella kohtaan *Items*
   - valitse seuraavaksi kohtaan *Model* vaihtoehdoksi *Split-half*
2. Toteuta split-half -reliabiliteettianalyysi yllä määrittelemilläsi valinnoilla painamalla *OK*-painiketta ja saat analyysin tulosteet tulosteikkunaan.

3. Aineiston alustava kuvailu ja tarkastelu

Tässä luvussa tarkastellaan aineiston kuvailuun liittyviä yleisiä tilastollisia menetelmiä. Aineiston kuvailuhan on tärkeä osa oman aineiston haltuunottoa ja tulkintaprosessia. Kuvailevilla menetelmillä voidaan ensinnäkin järjestää ja tiivistää havaintoaineistoa ymmärrettävämpään ja selkeämpään muotoon sekä kuvata aineiston muuttujissa havaittavia jakaumia ja eri muuttujien välisiä suhteita. Tutkimuskysymyksissä tyypillisesti on kyse eri muuttujiin (tutkittavien ilmiöiden) liittyvien yhteyksien ja erojen selvittämisestä. Voidaan esimerkiksi olla kiinnostuneita siitä, mikä opetusmenetelmä johtaa parhaaseen oppimistulokseen. Aineiston huolellisen kuvailevan tarkastelun jälkeen sovelletaan monesti lisäksi analyysimenetelmiä, jotka antavat meille välineitä esimerkiksi testattavien hypoteesien hyväksymiseen tai hylkäämiseen, luottamusvälien muodostamiseen ja havaittujen vaikutusten voimakkuuksien eli efektikokojen arviointiin. Lisäksi ne mahdollistavat empiiristen tulosten pysyvyyden ja luotettavuuden arvioimisen tai testaamisen ja erilaisten aikasarjojen tutkimisen. Mutta, kuten edellä jo todettiin, menetelmien hallinta ei yksinään takaa, että tutkija saa esille ”todelliset yhteydet” tutkittavasta ilmiöstä, tätäkin tärkeämpää on tutkijan aito kiinnostus tutkittavaan ilmiöön ja halu pyrkiä monipuolisesti tulkitsemaan aineistossa havaittuja yhteyksiä ja vaikutussuhteita. Jos tyytyy ensimmäiseen tulosteeseen, mitä ohjelmisto ja analyysimenetelmät tuottavat, voi jäädä huomamaatta tulkinnan kannalta hyvinkin keskeisiä tekijöitä, yhteyksiä tai vaikutussuhteita, puhumattakaan omiin tuloksiin liittyvistä tulkinnallisista ja teoreettisista ulottuvuuksista. Katso laajemmin tilastollisen tutkimuksen tulkinnan perusteista ja haasteista (esim. Ketokivi 2015, 232–262).

3.1 Aineiston kuvailu

Ennen varsinaisia analyyseja on siis tärkeää tarkastella tutkimuksen kannalta keskeisten muuttujien jakaumia käyttäen frekvensseihin liittyviä esityksiä ja erilaisia tunnuslukuja. Tämä on jo sen vuoksi tärkeää, että näin tutkija pääsee sisään omaan tutkimusaineistoonsa ja saa hyvän käsityksen omasta tutkimusaineistostaan. Tutkijalle onkin tärkeää tuntea aineistonsa monipuolisesti ennen kuin ryhtyy tekemään varsinaisia tilastollisia analyyseja. Jos ei tätä tee, ”aineiston potentiaali saattaa jäädä hyödyntämättä”, kuten esimerkiksi Ketokivi (2015, 272) osuvasti tämän merkityksen toteaa. Myös korrelaatiomatriisin avulla voi päätellä paljon aineiston keskeisten muuttujien välisistä suhteista. Sen avulla voidaan helposti alustavasti tarkastella aineistossa olevien muuttujien välisiä yhteyksiä ja näiden yhteyksien voimakkuutta. Tästä nähdään nopeasti muun muassa esimerkkiaineistostamme se, kuinka vanhempien koulutustaso ja tutkittavien viriketausta korreloivat keskenään. Tältä pohjalta voidaan tehdä jo alustavia päätelmiä esimerkiksi sen suhteen, kannattaako aiottuihin analyyseihin sisällyttää nämä molemmat tekijät vai olisiko viisasta analysoida tutkittavaa ilmiötä vain toisen tekijän (muuttujan) suhteen. Esimerkiksi tässä tapauksessa voidaan jättää vanhempien koulutustaso vähemmälle huomiolle, koska viriketaustamuuttuja on rakennettu monesta eri tekijästä ja on näin oletettavasti vanhempien koulutustasoa mittaavaa muuttujaa monipuolisempi. Samoin korrelaatiomatriisin avulla voidaan tarkastella alustavasti aiottujen kielellisyys- ja matemaattis-loogisuus -summamuuttujien osioiden keskinäisiä korrelaatioita.

Aineiston kuvaileva tarkastelu on senkin vuoksi tärkeää, että tämän avulla voi hahmottaa esimerkiksi sitä kuinka paljon eri muuttujien kohdalla esiintyy puuttuvia tietoja (siis onko osa vastaajista jättänyt syystä tai toisesta vastaamatta johonkin kysymykseen). Lisäksi tässä yhteydessä on syytä tarkastella sitä, ettei aineistossa ole mukana eri kysymysten kohdalla tapauksia, jotka poikkeavat liiaksi muiden vastaajien vastausarvoista (ei ole mikään vääryys poistaa aineistoista oman tutkimusasetelman kannalta tuloksia merkittävästi vääristeleviä tapauksia; toki tämän suhteen tulee olla tarkkana, ettei ota pois aineistosta sellaisia tapauksia tai vaikka muuttujia, jotka sinne kuuluvat). Kuvailevan tarkastelun yhteydessä on helppo havaita myös mahdolliset syöttövirheet, jotka on syytä korjata ennen varsinaisia analyysiajoja. Kaiken kaikkiaan tutkijan on hyvä tuntea tutkimusaineistonsa hyvin, tähän kuvailevat menetelmät antavat hyvän lähtökohdan.

Aineiston kuvailuun sopivat siis erilaiset frekvenssi- ja prosenttiosuusesitykset, sekä jakauman sijaintia ja hajontaa kuvaavat tunnusluvut, kuten keskiarvo ja keskihajonta. Myös erilaiset diagrammiesitykset ovat avuksi. Määrällisten muuttujien kohdalla on jo tässä vaiheessa syytä tarkastella niiden jakaumia (esimerkiksi sitä, noudattavatko käsiteltävät muuttujat normaalijakaumaa, joka on monissa tilastollisissa analyysimenetelmissä eräs keskeinen oletus). Näin voidaan jo tässä vaiheessa arvioida kunkin muuttujan soveltuvuutta suunniteltuihin analyyseihin. Tilasto-ohjelmissa on omat kuvailuun sopivat analyysimenetelmät ja graafiset esitykset. SPSS:ssä nämä löytyvät *Analyze*-valikon *Descriptive Statistics* -kohdasta, jonka alla on muuttujan yksittäistä kuvausta varten frekvenssi- ja prosenttitaulut sekä histogrammit. Tämän lisäksi tässä valikossa ovat muuttujien normaalijakauman testaamiseen liittyvät testit ja graafiset kuviot.

Graafiset kuviot ovat tärkeä osa tutkimusaineiston kuvailua. Näiden avulla voidaan esittää tutkimustulokset tehokkaasti, havainnollisesti ja ymmärrettävästi. Kuvio on tehokas kommunikointiväline. Hyvin käytettynä kuviot antavat raporteille vaikuttavan ja asiantuntijuutta korostavan ulkoasun, mutta, kuten tunnettua, graafisiin esityksiin liittyy myös tietty manipulointiriski. Käyttämällä erilaisia asteikkoja saadaan tapahtuneet muutokset esimerkiksi näyttämään joko vähäpätöisemmiltä tai merkittävämmiltä kuin ne todellisuudessa ovatkaan. Tutkimustulosten dramatisointia on pyrittävä välttämään, vaikka nykyinen media-aika siihen helposti houkutteleekin. Joka tapauksessa on hyvä pohtia eri tilanteisiin sopivia esitysmuotoja. Yleisesti ottaen raporttien luettavuuden kannalta olisi hyvä, että esitystapa olisi vaihteleva, mutta liiallista kirjavuuttakin tulisi välttää. Yleisesti kuvioiden käytön ohjenuoraksi voidaan listata esimerkiksi Hirsjärven ym. (2009, 348) esittämä kriteeristö: heidän mukaansa kuvioiden tulee olla (1) yksiselitteisiä ja informatiivisia; niissä ei saa olla liiaksi informaatiota, (2) rinnakkaiskuvioiden tulee olla samankokoisia, (3) kuviot tulee numeroida juoksevasti, (4) kuviolla tulee olla selkeä ja informatiivinen nimi, (5) kuvion tulkitsemiseen tarvittavien selitysten tulee olla kuvion yhteydessä ja (6) kuvioiden tulee olla luettavia ja selkeitä. Ylimääräisiä ja usein turhia koristeluja ja tehosteita, kuten esimerkiksi kolmiulotteisuutta kannattaa lähtökohtaisesti välttää. Hieman yksityiskohtaisempi 17 kohdan lista kuvioiden muodostamiseen liittyvistä huomioitavista seikoista on koottu APA-julkaisuoppaaseen (APA 2020, 232). Esimerkkejä ja tarkempaa kuvausta tilastograafien käyttöön liittyvistä ongelmista löytyy myös mm. alan klassikkoteoksesta *How to Lie with Statistics* (D. Huff 1954), joka on julkaisuvuodestaan riippumatta yhä ajankohtainen.

Taulukoiden ja kuvioiden raportoinnissa on hyvä huomata, että tilasto-ohjelmien tuottamat esitykset eivät useinkaan ole muotoiluiltaan ja asetteluiltaan sellaisenaan valmiita ja käyttökelpoisia tutkimusraporttiin, vaan niitä joudutaan yleensä muokkaamaan ja viimeistelemään. Monet tieteelliset julkaisusarjat käyttäytymistieteiden alalla käyttävät ja suosittavat muotoiluihin ns. APA-tyylin mukaista ohjeistusta (APA 2020). Siinä muotoiluissa pyritään yhdenmukaiseen, pelkistettyyn ja selkeään esitystapaan. Taulukoissa viivoituksena riittävät yleensä pelkät vaakaviivat, taulukon sarakkeilla ja riveillä otsikoinnit ovat lyhyet mutta riittävän selkeät, lukuarvoissa riittää kaksi desimaalia (merkitsevyystesteihin liittyvissä p-arvoissa kolme desimaalia). Yksityiskohtaisempi 15 kohdan muistilista taulukointeihin liittyvistä seikoista löytyy APA-julkaisuoppaasta (APA 2020, 206).

Hyvä kuvio eli tilastograafi on mahdollisimman itsenäinen esitys, toisin sanoen sen sisältö on ymmärrettävissä ilman, että lukijan on haettava lisätietoa tutkimusraportin tekstistä. Kuviota laadittaessa on aluksi hyvä harkita, sopiiko käytetty kuviotyyppi tilanteeseen. Eräitä yleisimmin käytettyjä kuviotyyppejä ovat mm. pylväs-, viiva-, laatikko-jana-, sektori- ja pistediagrammi. Mikäli graafi sisältää akseleita, on tärkeätä harkita ja tarkistaa akselien tarkoituksenmukainen skaala ja akselien selkeä otsikointi niin, että myös käytetty mittayksikkö käy ilmi. Pylväskuvioissa on hyvä harkita pylväiden suunta. Esimerkiksi a) Likert-asteikollisissa mielipideväittämissä on tapana käyttää pystysuuntaisia pylväitä eikä pylväiden järjestystä kannata luonnollisestikaan muuttaa, b) puhtaasti nominaaliasteikon muuttujilla luokkien frekvenssejä on luontevaa kuvata vaakapylväillä, joissa pylväät on järjestetty pituusjärjestykseen valitun vaihtoehdon yleisyyden mukaisesti. Akselien skaaloissa on hyvä käyttää tasavälistä luokitusta, eikä asteikkoja kannata katkaista: esimerkiksi asteikon aloittaminen nollan sijaan jostain itse valitusta kohdasta saattaa vääristää kuvattavan asian täysin virheelliseksi!

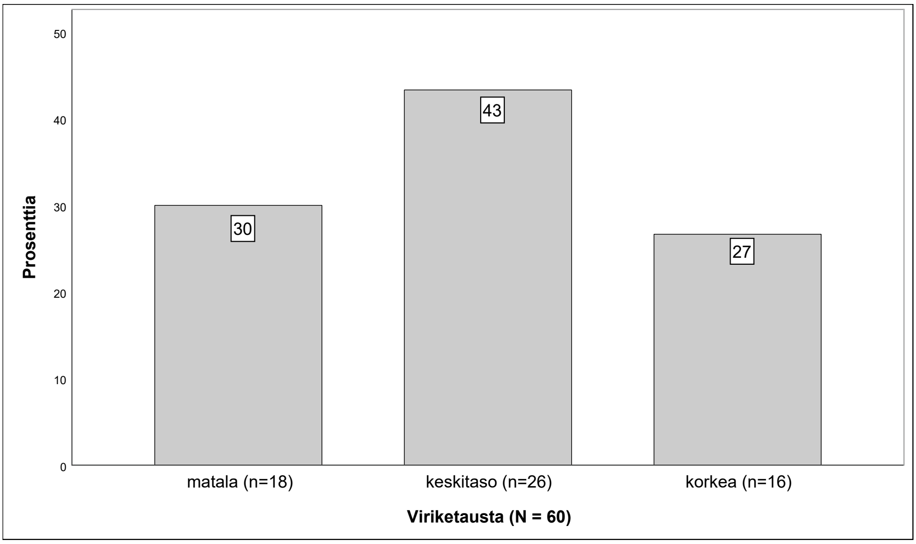
Mihin kohtaan tekstiä kuvio sitten pitäisi sijoittaa? Tämä tuottaa usein päänvaivaa varsinkin aloittelevalle tutkijalle. Yllättävää kyllä, melko usein törmää esimerkiksi pro gradu -opinnäytetöissäkin vielä siihen, että uusi luku tai alaluku aloitetaan suoraan kuviolla. Tämä, samoin kuin luvun päättäminen kuvioon, tuo tekstiin väistämättä atomistisuutta, joka viestii kirjoittajan kiireisestä suhteesta kirjoittamiseen. Kuvioiden yhteyteen tulee aina liittää jonkinlainen johdatusteksti ja tämän jälkeen tulkintateksti, jossa käsitellään kuvion sisältämät pääyhteydet, päätulokset oman tutkimuksen kannalta. Muista myös, että kuviossa tulee käyttää suomenkielisiä termejä, jos kirjoittaa suomeksi. Näin ollen SPSS:n tuottamien kuvioiden otsakkeet ja kaikki selitetekstit olisi aina muokattava englannista suomen kielelle, kuten edellä ja myöhemmin esitettävissä kuvioissa ja taulukoissa on (tosin joihinkin esityksiin on jätetty alkuperäisten tulosten termistöä ja muotoiluja esimerkin vuoksi). Kuvioiden käytön lähtökohdan kriteeriksi sopii se, että ne palvelevat tulosten raportointia lisäämällä havainnollisuutta ja helpottavat tekstin luettavuutta ja vievät esitystä eteenpäin; kriittinen lukija erottaa toisistaan heppoisen ja hyvän työn, oli siinä sitten käytetty hienoja graafisia kuvia tai ei, joten kuvilla kikkailu ei sinänsä kanna kovin pitkälle.

3.2 Aineiston kuvailu pylväsdiagrammilla ja histogrammilla

Tässä yhteydessä käsitellään jakaumien kuvailuun sopivia pylväsdiagrammeja: tavalliset pylväsdiagrammit on tarkoitettu kategoristen muuttujien kuvaukseen, kun taas histogrammit on tarkoitettu jatkuva-arvoisten muuttujien frekvenssien kuvaamiseen. Jos kategorinen muuttuja on nominaaliasteikollinen, pylväsdiagrammi muodostetaan yleensä vaakapylväikkönä. Pylväsdiagrammissa pylväät ovat erillään toisistaan, kun taas histogrammissa pylväät piirretään yhteen. Molemmat ovat paikallaan esimerkiksi tutkimusaineiston muuttujien kuvauksessa. Voidaan kuvata niiden avulla esimerkiksi otoksen sukupuolijakaumaa tai eri muuttujien suhdetta toisiinsa. Kuten edellä jo todettiin, muita paljon käytettyjä esitystapoja ovat esimerkiksi ympyrä- eli sektori-, pinta- ja viivadiagrammiesitykset. Viivadiagrammit sopivat erilaisten aikasarjojen kuvaamiseen, johon tietysti voidaan käyttää myös pylväsdiagrammeja, joiden avulla voidaan kuvata muun muassa ajassa tapahtuneet muutokset. Sektoridiagrammeja ei suositella tutkimusraportin graafeiksi, koska yleensä sama asiasisältö on esitettävissä pylväikkönä selkeämmin ja tarkemmin. (Ks. enemmän esim. Alkula ym. 2002, 191–196; Karjalainen 2000, 29–65; Kuusela 2000; Fieldin & Gilbert 2006, 69–93).

Pylväsdiagrammissa kuvataan joko vaaka- tai pystysuorien pylväiden avulla aineiston eri muuttujien frekvenssijakaumia (eli sitä, kuinka monta tapausta kuhunkin muuttujan luokkaan tulee tai esitys voi nojautua myös prosenttiosuuksiin). Usein pylväsdiagrammissa jakauma kannattaakin esittää prosenttiosuuksina, koska usein eri ryhmien osuudet ovat lukijan kannalta mielekkäämpiä kuin pelkkä frekvenssiesitys. Tosin frekvenssejä voi käyttää sellaisenaankin, esimerkiksi aineiston taustamuuttujia (sukupuoli, viriketausta, ammatti yms.) esiteltäessä. Esimerkeissä käytetään myös tilastotieteilijöiden suosittamaa laatikko-jana -kuviota (ks. kuvio 8a ja 8b, s. 110 ja 111), jonka käytön lisääntyminen voisi olla paikallaan esimerkiksi kasvatusalan ja yhteiskuntatieteellisissä tutkimusraporteissa niiden runsaamman informaation vuoksi. (Ks. enemmän esim. Helenius 1989, 57–60).

Oheen on liitetty muutamia kuvioesimerkkejä. Aluksi kuviossa 3 esitetään otokseen kuuluvien viriketaustajakauma.



KUVIO 3. Tutkimukseen osallistuneiden viriketaustajakauma.

Tästä varsin yksinkertaisesta esimerkkigraafista voidaan todeta 30 prosentilla otokseen kuuluvista henkilöistä viriketaustan olevan matalan, 43 prosentilla keskitasoa ja 27 prosentilla korkean. Kuvioon on lisätty pylväiden yläosaan tarkat prosenttiluvut, jotka helpottavat lukijaa hahmottamaan kuviota. Prosenttilukuja esitettäessä on tärkeätä tietää, mistä lukumäärästä prosenttiosuudet on laskettu, siksi kuvioon on liitetty myös otoskoon osoittava luku N=60. Tietenkin jakauma voitaisiin haluttaessa esittää myös lukumäärinä (frekvensseinä), jolloin pylväiden suhteelliset koot eivät muuttuisi, mutta frekvenssilukujen keskinäinen vertailu on hankalampaa kuin prosenttiosuuksien. Toisinaan näkee käytäntöä, jossa prosenttiluvut esitetään yhdellä desimaalilla ja lukumääriä kuvaavat frekvenssit pelkillä kokonaisluvuilla. Näin luvuista ehkä on helpompi mieltää, kummista frekvensseistä on kyse. Yllä olevassa kuviossa kuitenkin lukuarvot ovat esitetty kokonaislukuina.

SPSS: Pylväsdiagrammin piirtäminen (yksi kategorinen muuttuja):

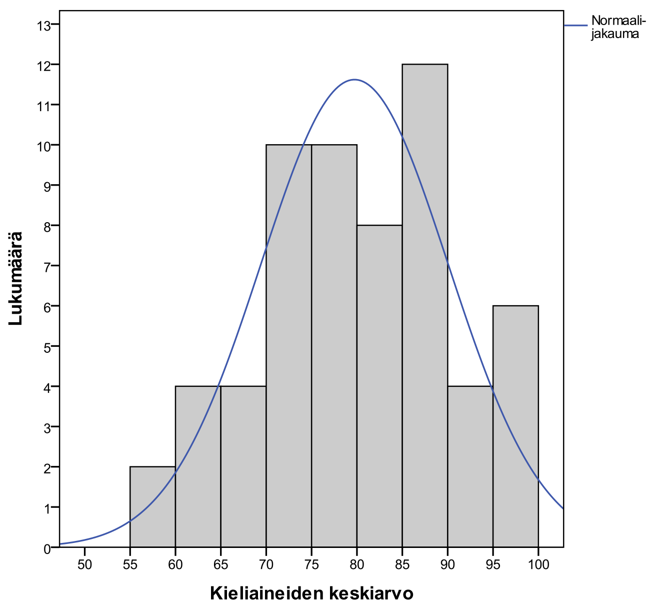
1. Valitse ylhäältä valikkoriviltä *Graphs → Legacy Dialogs → Bar*
   - valitse esiin tulevasta valikosta *Simple*-kuvake ja varmista, että kohdan *Data in Chart Area* vaihtoehto *Summaries for groups of cases* on valittuna, jatka *Define*-painikkeella eteenpäin
   - valitse muuttujalistasta haluamasi kategorinen muuttuja, ja siirrä se listan oikealla puolella olevalla nuolipainikkeella kohtaan *Category Axis*
   - kohdasta *Bars Represent* voit valita, kuvaavatko pylväät esim. frekvenssejä (*N of cases*) vai prosenttiosuuksia (% *of cases*)

ohjetaulukko jatkuu...

2. Tarkista *Options*-painikkeesta, että mahdolliset puuttuvat tiedot eivät tule omaksi pylvääkseen, eli ettei kohdassa *Display groups defined by missing values* ole valintaruksia
   - hyväksy valinnat klikkaamalla *Continue*-painiketta
3. Toteuta pylväsdiagrammin piirtäminen yllä määrittelemilläsi valinnoilla painamalla *OK* -painiketta ja saat kuvion tulosteikkunaan.

Halutessasi voit jatkaa kuvion muokkaamista tulosteikkunassa: muokkaamisen yleisperiaate on se, että kaksoisnapsautat kohteen muokkaustilaan (graafi avautuu erilliseen *Chart Editor* -ikkunaan), jossa pääset muokkaamaan haluamaasi kohtaa hiirellä klikkaamalla tai oikealle avautuvasta *Properties* -valintaikkunasta. Monesti myös yläreunan muokkaamiseen liittyvistä valikoista ja hiiren oikealla painikkeella avautuvasta valikosta saatat löytää kätevästi tarvittavia muokkausmahdollisuuksia. Kun kuvion muokkaus on valmis, sulje *Chart Editor* -ikkuna yläreunan ruksista ja kuvio-osa palautuu muokatussa muodossa tulosteikkunaan.

Jos kyseessä olisi määrällinen numeerinen muuttuja, muuttujan havaintoarvojen jakaumaa voitaisiin kuvata histogrammilla. Se on pylväsdiagrami, joka perustuu muuttujan luokitteluun ja näin muodostettujen luokkien frekvensseihin. Kuvioon voitaisiin lisätä myös normaalijakaumaa osoittava käyrä, josta voitaisiin päätellä silmämääräisesti kyseisen muuttujan jakauman muoto suhteessa normaalijakaumaan. Tätä normaalijakautuneisuutta voidaan selvittää myös erilaisilla tilastollisilla testeillä, joita tulisikin käyttää graafisen kuvailun rinnalla. Seuraavassa histogrammissa (kuvio 4) on kuvattu muuttujaa kieliaineiden keskiarvo sekä muuttujan jakauman muotoa verrattuna normaalijakaumaan.



KUVIO 4. Kieliaineiden keskiarvo -muuttujan jakauma.

Kuviosta 4 voidaan tarkastella kieliaineiden jakaumaa ja sen suhdetta odotettuun normaalijakaumaan. Kuviosta on mahdollista todeta visuaalisesti, että otoksen jakauma noudattaa jokseenkin hyvin normaalijakauman muotoa. Yleensä ei kannata, kuten jo mainittiin, luottaa silmämääräisyyteen tätä arvioitaessa. SPSS:ssä on tätä varten omat testinsä: Kolmogorov-Smirnov -testi Lillieforsin korjauksella ja Shapiro-Wilk -testi. Jälkimmäistä käytetään, jos otoskoko on 50 tai sen alle. Esimerkkiaineisto on tätä suurempi, joten testiksi tulee Kolmogorov-Smirnov. Jos tehdään kyseinen testi, kieliaineiden keskiarvojen jakauma ei poikkea normaalijakaumasta tilastollisesti merkitsevästi ($p \geq 0{,}200$, $N = 60$), kuten jo kuviosta saattoi alustavasti päätellä. Nollahypoteesina testissä on se, ettei otosta vastaava populaatiojakauma poikkea normaalijakaumasta. Perinteisen käytännön mukaisesti, jos merkitsevyystasoksi (p-arvo) tulee alle 0,05, testattava jakauma poikkeaa normaalijakaumasta. Jatkoanalyyseissä ei ainakaan suoralta kädeltä saa käyttää parametrisiä testejä, vaan olisi harkittava epäparametristen testien käyttöä tai edellä esitettyä bootstrap-estimointia (ks. luku 1.6.1.4). Toisaalta kun otoksen tapausten lukumäärä on suuri, ei normaalisuusoletuksista pidetä niin tiukasti kiinni. Tämän ajatuskulun taustalla on todennäköisyyslaskennan keskeisen raja-arvolauseen sovellus. Sen logiikka on seuraavanlainen: koska testeissä p-arvo lasketaan testisuureen jakauman avulla ja testisuureen jakauma lähestyy normaalijakaumaa otoskoon kasvaessa, tarkastelun kohteena olevan muuttujan jakauma ei sinällään olekaan niin tärkeä. Vaikka tarkastelun kohteena olevan muuttujan jakauma olisi hyvinkin kaukana normaalijakaumasta, niin usein jo noin 30:n suuruisesta otoksesta lasketun tarkasteltavan testisuureen jakauma alkaa muistuttaa aika lailla hyvin normaalijakaumaa! Siten testeihin liittyvä p-arvo ei olekaan pahasti pielessä, kunhan käytettävä otoskoko vain on riittävän suuri.

SPSS: Määrällisen muuttujan jakaumaa kuvaavan histogrammin piirtäminen (yksi muuttuja):

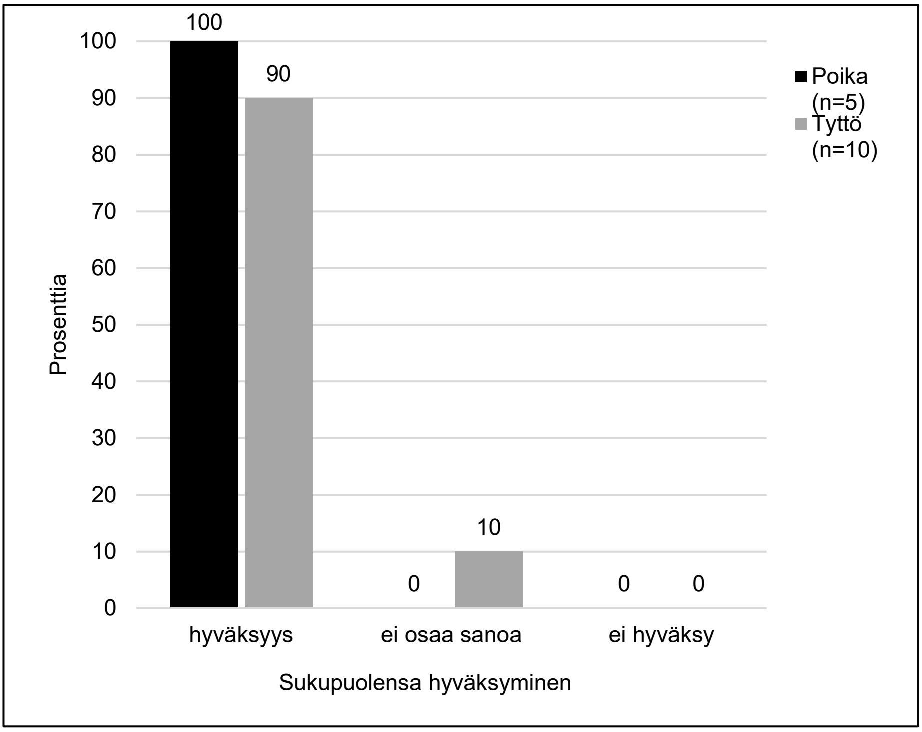
1. Valitse ylhäältä valikkoriviltä *Graphs → Legacy Dialogs → Histogram*.
   - valitse muuttujalistasta haluamasi muuttuja ja siirrä se listan oikealla puolella olevalla nuolipainikkeella kohtaan *Variable*
   - jos haluat kuvioon myös normaalijakaumaa kuvaavan käyrän, niin ruksaa kohta *Display normal curve*
2. Paina *OK*-painiketta ja kuvio piirtyy tulosteikkunaan

Halutessasi voit jatkaa kuvion muokkaamista tulosteikkunassa (ks. aiempi SPSS-ohje sivulla 97)

ohjetaulukko jatkuu...

HUOM:
- tämän kirjan histogrammit on tehty hieman toista reittiä, eli interaktiivisella tavalla, jolloin kuvion ulkoasuun on pystytty vaikuttamaan enemmän. Toteutus olisi tällöin yläreunan valikosta *Graphs* → *Chart Builder* → *Histogram*
- histogrammin saa toteutettua myös muiden SPSS-toimintojen yhteydessä, esimerkiksi kohdasta *Analyze* → *Descriptive Statistics* → *Frequencies* (siellä graafeihin liittyvä painike *Charts*) tai *Analyze* → Descriptive Statistics → *Explore* (siellä graafeihin liittyvä painike *Plots*).

Kuten edellä esitetyt kaksi esimerkkiä osoittavat, diagrammiesitykset ovat hyvin havainnollisia ja niiden käyttöalue on monipuolinen. Valaiskoon tätä vielä yksi esimerkkiaineiston ulkopuolelta tuleva diagrammiesimerkki. Tässä kuvataan sitä, miten 8-vuotiaat haastateltavat lapset samaistuivat omaan sukupuoleensa.



KUVIO 5. Kahdeksanvuotiaiden lasten sukupuolensa hyväksyminen (N=15).

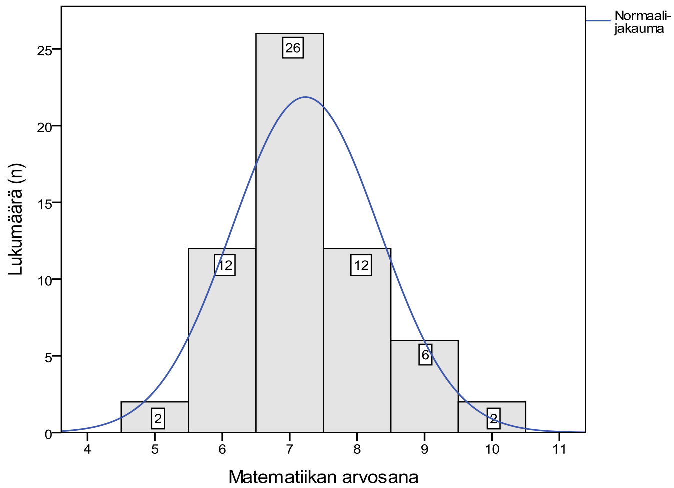
Haastattelun tulos aukenee kuviosta vaivatta: jo pelkkä vilkaisu riittää kertomaan kyseisten lasten sisäistäneen oman sukupuolensa, vain yksi viidestätoista lapsesta (yksi tytöistä) ei haastatteluhetkellä kyennyt sanomaan sitä, miten hän mieltää oman sukupuolensa, minkä esimerkiksi freudilaisen teorian mukaan ei pitäisi olla kovinkaan tavatonta tämän ikäisillä lapsilla. Tässä esimerkkikuviossa tosin olisi harkittava, että ovatko pelkät prosenttiosuudet hyvä ja havainnollinen tapa esittää tällaisen varsin pienen aineiston jakaumaa. Toisin sanoen prosenttiosuuksien sijaan voitaisiin käyttää todellisia lukumääriä eli frekvenssejä. Kuvion oikean yläkulman selitteisiin olisi hyvä lisätä vielä tieto osaryhmien kokojen suuruudesta ('poika', n=5; 'tyttö', n=10).

SPSS: Pylväsdiagrammin piirtäminen (kaksi kategorista muuttujaa):

1. Valitse ylhäältä valikkoriviltä *Graphs → Legacy Dialogs → Bar*.
   - valitse esiin avautuvasta valikosta vuorostaan *Clustered* -kuvake ja alareunan kohdasta *Data in Chart Area* vaihtoehdoista ruksaa *Summaries for groups of cases*. Jatka *Define*-painikkeella eteenpäin
   - valitse muuttujalistasta haluamasi kategorinen muuttuja, ja siirrä se listan oikealla puolella olevalla nuolipainikkeella kohtaan *Category Axis*. Pylväät piirretään siis tämän kategorisen muuttujan eri arvoille. Esimerkissämme kyseinen muuttuja on oman sukupuolensa hyväksyminen
   - siirrä kohtaan *Define Clusters by* toinen kategorinen muuttuja. Pylväät esitetään tämän toisen muuttujan muodostamina ryhminä. Esimerkkikuviossamme tämä muuttuja on sukupuoli.
   - kohdasta *Bars Represent* voit valita, kuvaavatko pylväät esim. frekvenssejä (*N of cases*) tai prosenttiosuuksia (*% of cases*).
2. Tarkista *Options*-painikkeesta, että mahdolliset puuttuvat tiedot eivät tule omaksi pylvääkseen, eli poista valintaruksi kohdasta *Display groups defined by missing values,* hyväksy valinnat klikkaamalla *Continue*-painiketta
3. Toteuta pylväsdiagrammin piirtäminen yllä määrittelemilläsi valinnoilla painamalla *OK* -painiketta ja saat kuvion tulosteikkunaan.

Halutessasi voit jatkaa kuvion muokkaamista tulosteikkunassa (ks. aiempi SPSS-ohje sivulla 97).

Tässä yhteydessä ei käsitellä suoran jakauman yleisiä periaatteita yksityiskohtaisemmin. Näitä esityksiä löytyy useasta tilastotieteen alkeiskirjasta kuin myös pidemmälle menevienkin teosten sivuilta. Kuten kuvion esimerkistä nähdään, SPSS esittää muuttujan, jolla on paljon erisuuruisia numeerisia arvoja, jakauman luokittelemalla aineiston arvot suurempiin luokkiin (esimerkissämme ohjelma yhdistää luokat viiden yksikön suuruisiin ryhmiin, haluttaessa ryhmärajoja voi säädellä itsekin). Kun muuttujalla on vähemmän "arvoluokkia" kuin kymmenen, voidaan siitä tehdä sellainen suora jakauma -esitys, jossa kaikki eri arvot tulevat esiin. Käytetyssä aineistossa tällainen muuttuja on esimerkiksi matematiikan arvosana, koska tässä tutkittavat ovat voineet saada vain kokonaislukuarvoja nelosesta kymmeneen. Kuviona tarkasteltuna tämä näyttää seuraavanlaiselta:



KUVIO 6. Matematiikan arvosana -muuttujan jakauma.

Kuviosta 6 nähdään suoraan jokaisen arvon tapausmäärät ja jakauman vinous ja huipukkuus normaalijakauman suhteen. Lisäksi voitaisiin tulostaa tähän liittyvä frekvenssitaulukko, jolloin saataisiin yksityiskohtaiset prosenttiosuudet yms. Kun tehdään muuttujalle normaalijakaumatesti (Kolmogorov-Smirnov Lillieforsin korjauksella, p < 0,001), vahvistaa tämä kuvastakin havaittavaa seikkaa eli sitä, että muuttujan jakauma poikkeaa jonkin verran normaalijakaumasta, joskaan tämä poikkeama ei ole kovin suuri. Kolmogorov-Smirov -testin ja vastaavien muiden normaalijakaumatestien yksi ongelma on se, että ne antavat etenkin suurissa otoksissa herkästi tuloksen, jonka mukaan testattavan muuttajan jakauma poikkeaa normaalijakaumasta (ks. tietolaari 3, normaalijakauma ja normaalisuuden testaus, s. 103), vaikka ero tässä suhteessa oli suhteellisen vähäinen. Esimerkiksi kuviossa 6 esitetyssä tapauksessa poikkeama on niin pieni, ettei se esimerkiksi estäisi käyttämästä sellaisia analyysimenetelmiä, jotka edellyttävät muuttujien jakauman noudattavan normaalijakaumaa. Yleisimmät tilastomenetelmät ovat muutekin tämän suhteen suhteellisen vakaita, joten suuremmatkaan poikkeamat normaalijakaumasta eivät välttämättä estä käyttämästä näitä analyysimenetelmiä.

3.3 Keskiluvut muuttujan jakauman sijainnin kuvaajana

Yleisimmin käytetyt keskiluvut ovat keskiarvo (mean, average tai $\bar{x}$,) ja mediaani (median, Md). Tämän lisäksi jonkin verran käytetään moodia. Moodi (mode) on se muuttujan arvo, jonka frekvenssi on suurin. Moodiin ei tässä yhteydessä kiinnitä sen enempää huomiota, vaan esityksessä keskitytään lähinnä keskiarvojen ja jossakin määrin myös mediaanien käyttöön. Näiden lisäksi kuvailussa voidaan käyttää luottamusvälitarkastelua, jossa yksittäisen pistearvon sijaan muodostetaan väli, johon kyseinen tunnusluku riittävällä varmuudella sijoittuu. Yksittäinen pistearvo ei ota huomioon lainkaan otoksen sisältämää epävarmuutta (tyypillisesti otos on vain pieni osa tutkimuksen populaatiosta), jota luottamusvälitulkinnalla voidaan arvioida. Luottamusvälejä käsitellään muun muassa luvussa 1.6.1.3. (s. 49) ja tietolaari 3:ssä (s. 103), ja jatkossa käsiteltävien analyysimenetelmien yhteydessä. Keskiarvot ovat käytetyimpiä keskilukuja, vaikka toisinaan mediaanin käyttö olisi aritmeettista keskiarvoa perustellumpaa. Tällaisia ovat esimerkiksi tilanteet, joissa muuttujan jakauma on voimakkaasti vinoutunut tai se sisältää selvästi poikkeavia havaintoja. Keskiarvon suosiota selittänee osaksi sen tuttuus sekä tutkijalle että tutkimustulosten hyödyntäjille. Myös se, että keskiarvo on monen muun tunnusluvun, esimerkiksi varianssin ja korrelaatiokertoimen, laskemisen yksi peruste, selittänee osaltaan keskiarvon suosiota jakauman sijainnin tunnuslukuna.

Keskiarvoihin perustuvien testien käyttö edellyttää välimatka- tai suhdeasteikkoista havaintoaineistoa. Frekvenssiesityksiin nähden keskiarvoesityksissä menetetään jonkin verran informaatiota, mutta toisaalta jakauman sijaintiin liittyvä informaatio kyetään kuvaamaan tiivistetymmin ja aineiston tulkinta sen yleispiirteistä on vaivattomampaa kuin frekvenssiesityksistä. Keskiarvo on muuttujan arvojen summa jaettuna muuttujan arvojen lukumäärällä. Kasvatustieteellisissä, kuten yleisemminkin yhteiskuntatieteellisissä, tutkimuksissa lasketaan välimatka-asteikollisiksi muuttujiksi useimmiten erilaisilla asteikkomittareilla (esim. Likert-asteikko) kerätty aineisto tai esimerkiksi todistusarvosanat, vaikka nämä eivät kriittisesti katsoen välimatka-asteikon kriteereitä täytäkään (ovat paremminkin järjestysasteikollisia).

Keskiarvojen käyttöön liittyy muutamia ongelmia, jotka on hyvä tiedostaa: keskiarvot ovat herkkiä aineiston poikkeaville arvoille (vinot jakaumat), eli jos otoksessa on mukana muutamia vastauksia, joiden arvot poikkeavat muista huomattavasti, voi niiden vaikutus saatuun keskiarvoon olla melkoinen. Mitä pienemmästä tai heterogeenisemmasta otoksesta on kysymys, sitä tärkeämpää

TIETOLAARI 3. JAKAUMAA KUVAAVIA TILASTOLLISIA TUNNUSLUKUJA

Sijaintia kuvaavia tunnuslukuja (keskilukuja)

Moodi (mode)
kuvaa sitä muuttujan havaintoarvoa (tyyppiarvoa), jota on tarkasteltavassa muuttujassa määrällisesti eniten. Mooditarkastelu soveltuu hyvin aineiston karkeaan kuvaamiseen. Tätä keskilukua voidaan käyttää riippumatta siitä, millä mitta-asteikoilla aineisto on kerätty. Aineistossa voi olla kaksi tai useampaakin moodia, jos havaintoarvoja, joita aineistoissa on eniten, on muuttujassa yhtaikaa useampia. Tällöin moodi ei ole yksikäsitteinen, eikä sen käyttö keskilukuna ole mielekästä.

Mediaani (md)
kuvaa muuttujan arvojen suuruusjärjestyksessä keskimmäistä havaintoarvoa, jonka alemmalle puolelle sijoittuu muuttujan arvoista toinen puoli, ja toinen puoli tämän yläpuolelle. Jako perustuu siis muuttujien arvojen suuruusjärjestykseen. Tätä voidaan käyttää, kun aineisto on kerätty vähintään järjestysasteikollisella mittarilla. Jos arvojen määrä jakaantuu tasan, mediaanin voi muodostaa kahden keskimmäisen arvon keskiarvona eli jakamalla kahden keskimmäisen luvun summan kahdella.

Keskiarvo (aritmeettinen, $\bar{x}$, ka)
on arvo, joka saadaan jakamalla muuttujan arvojen summa muuttujan arvojen lukumäärällä. Keskiarvot ovat yksi käytetyimpiä keskilukuja, mikä selittyy sillä, että se on helppo tulkita ja useimmille tuttu. Usein, esimerkiksi voimakkaasti poikkeavien havaintojen takia, mediaanin käyttö olisi tarkoituksenmukaisempaa, mutta tähän liittyvä tulkinnan ”moniasteisuus” ilmeisesti vähentää tämän käyttöä. Myös useat tilastolliset analyysimenetelmät, kuten varianssianalyysit, perustuvat keskiarvojen hyödyntämiseen ja vertailuun. Keskiarvojen käyttö edellyttää, että aineisto on välimatka- tai suhdeasteikkoista. Esimerkiksi kasvatustieteissä ja yhteiskuntatieteellisissä tutkimuksissa Likert-asteikkoisella mittareilla kerätyn aineiston tulkitaan usein täyttävän välimatka-asteikon kriteerit. (Katso enemmän kesiarvosta luvusta 3.3). Keskiarvotulkintaa vahvistetaan keskihajontatulkinnalla (kh, standard deviation, s tai sd) sekä luottamusvälitulkinnalla, joka perustuu tilastollisen päättelyyn liittyvän keskivirheen hyödyntämiseen.

Hajontaa kuvaavia tunnuslukuja

Vaihteluväli
kuvaa havaintoaineiston arvojen minimi- ja maksimiarvojen välistä eroa. Tämä voidaan ilmoittaa, jos aineisto on mitattu vähintään järjestysasteikkoisella mittarilla eli saadut arvot voidaan järjestää suuruusjärjestykseen. Vaihteluväli voidaan ilmaista myös vaihteluvälin pituutena.

Kvartiiliväli
kuvaa havaintoaineiston arvojen ala- ja yläkvartiilin välistä eroa. Kvartiilivälin avulla voidaan jakaa aineisto kolmeen osaan: yleensä kvartiiliväliin kuuluu havainnoista 50 prosenttia. Neljännes (25 %) jää tämän alapuolelle ja neljännes sen yläpuolelle. Kvartiiliväli voidaan ilmaista myös kvartiilivälin pituutena. Kvartiilipoikkeama (Q) on puolet kvartiilivälin pituudesta.

Varianssi (s^2)
on havaintoarvojen vaihteluun liittyvä tunnusluku, joka kuvaa sitä, kuinka saadut havaintoarvot sijoittuvat keskiarvon suhteen. Otoksesta laskettuna varianssista käytetään tunnusta s^2 tai var. Tämä voidaan määrittää, jos mittaus on tehty vähintään välimatka-asteikolla. Tunnusluvun muodostaminen perustuu ns. neliösumman laskemiseen, jossa jokaiselle havaintoarvolle lasketaan etäisyys muuttujan keskiarvosta. Lisäksi varianssin laskemisessa otetaan huomioon otoskoko. Varianssin lukuarvon tulkinta on hankalaa, koska sen yksikkö on muuttujan alkuperäisen mittaskaalan neliö.

Keskihajonta (kh, standard deviation, sd, s)
eli standardipoikkeama, otoksesta laskettuna käytetään tunnusta s tai sd, perusjoukosta laskettuna *σ,* kuvaa muuttujan arvojen hajaantumista keskiarvon suhteen, eli tämä on tietyn muuttujan varianssin neliöjuurta kuvaava tunnusluku, josta voidaan päätellä, miten arvot

ovat keskimäärin sijoittuneet keskiarvon suhteen. Tätä tunnuslukua voi käyttää, jos mittaus on tehty välimatka- tai suhdeasteikolla. Tämä tuo runsaasti informaatiota lisää pelkän keskiarvon käyttöön verrattuna. Esimerkiksi, kun sijoitamme keskiarvon (normaalijakautuman mukaisessa muuttujassa) kummallekin puolelle kaksi kertaa keskihajontaluvun suuruisen luvun, voimme päätellä, että noin 95 prosenttia aineistomme tapauksista sijoittuu tälle alueelle. Jos sijoitamme keskiarvon kummallekin puolelle keskihajontalukua osoittavan alueen, sijoittuu aineistostamme tälle alueelle 68 prosenttia.

Keskiarvon keskivirhe (standard error, Std. Error, s.e.)

voidaan tulkita otoskeskiarvojen keskihajonnaksi, jos otoksia kerättäisiin useita ja niistä kaikista laskettaisiin keskiarvo. Arvo kuvaa sitä, kuinka paljon otannasta aiheutuvaa epävarmuutta otoskeskiarvon estimointiin liittyy. Mitä pienempi keskivirhe on, sitä lähempänä eri otoksista lasketut keskiarvot ovat toisiaan ja populaation keskiarvoa. Silloin otoksen tuottama keskiarvo vastaa tarkemmin populaation arvioitavaa keskiarvoa. Keskivirheen avulla saadaan laskettua arviointiin liittyvä virhemarginaali, jota hyödynnetään luottamusvälien muodostamisessa. (Ks. tarkemmin esimerkiksi Nummenmaa 2011, 141–145.)

Keskiarvon luottamusväli (Confidence Interval for Mean)

Tunnuslukujen käyttöön liittyy tietty riskitaso sen suhteen, kuinka luotettavia otoksesta saadut tulokset ovat koko populaation suhteen. Kyse on siis siitä, kuinka luotettavasti saadut tulokset kuvaavat populaation vastaavia ominaisuuksia. Yleisimmin tässä käytetään 95 prosentin tarkastelutasoa eli tämän tulkinnan mukaan voimme pitää varsin uskottavana (luotettavuus 95 %), että otoksemme on peräisin populaatiosta, jossa keskiarvo sijaitsee luottamusvälin ala- ja ylärajan välissä. Välin pituus ilmaisee arvioinnin tarkkuutta tai vastaavasti epätarkkuutta. Keskiarvon luottamusväli määräytyy otoskeskiarvon (ka), keskivirheen (se) ja todennäköisyysjakaumasta saatavan arvon (suure z tai t, otoskoosta riippuen) perusteella. Esimerkiksi 95 %:n luottamusväli muodostuu seuraavasti: ka ± 1,96*se, jossa lukuarvo 1,96 on standardoidusta normaalijakaumasta saatu 95 % luotettavuuteen liittyvä z-arvo. Jos halutaan tätä varmempi tulkinta, valitaan yleisesti luottamusvälitasoksi 99 prosenttia, jolloin luottamusväli on pidempi verrattuna 95 prosentin tulkintatasoon. Luottamusvälin kokoon voidaan vaikuttaa myös otoskoolla: mitä suurempi otos, sen lyhyempi on luottamusväli. Luottamusväliin vaikuttaa myös tutkittavan muuttujan keskihajonta, mitä pienempi keskihajonta on, sitä tarkempaa tietoa luottamusvälin avulla kyetään tuottamaan eli sitä kapeampi on luottamusvälin alue, ja päinvastoin, mitä suurempi keskihajonta on, sitä suuremmaksi myös luottamusväli venähtää. (Ks. laajemmin luottamusvälitarkastelusta luvusta 1.6.1.3).

Jakauman muotoa kuvaavia tunnuslukuja

Vinous (Skewness)

kuvaa sitä, kuinka symmetrisesti saadut havaintoarvot sijoittuvat aineiston keskiarvon ympärille. Vinous ilmaistaan lukuarvolla ja – tai + -etumerkillä. Vinousarvo kuvaa siis, kuinka tasaisesti tai epätasaisesti annetut arvot ovat suhteessa toisiinsa. Etumerkki kuvaa sitä, mihin suuntaan kerätyn aineiston arvojen jakauma on vinoutunut, ja arvo, kuinka vino tämä jakauma on. Vinouden arvon ollessa negatiivinen, aineistossa on runsaasti arvoja, jotka ovat suurempia kuin aineiston keskiarvo (jakaumakuviossa vasemmanpuoleinen osa laskeutuu loivemmin). Vastaavasti positiivisesti vinossa aineistossa on runsaasti keskiarvoja pienempiä havaintoarvoja. Kun muuttujan vinousarvo asettuu välin -1 ja +1 ulkopuolelle, vinous alkaa olla selvästi normaalijakaumasta poikkeavaa (Bulmer 1979; Nummenmaa 2011, 71). Lisäksi vinousarvon suuruutta voi arvioida sen yhteydessä SPSS-tulosteisiin saatavan keskivirheen (standard error, SE) avulla: jos otantaan liittyvän virhemarginaalin avulla muodostettu luottamusväli (vinousluku ± 2*SE) sisältää arvon nolla, voitaisiin tulkita, että vinous ei merkitsevästi poikkea normaalijakauman symmetrisestä muodosta.

Huipukkuus (Kurtosis)

kuvaa aineiston jakauman huipun ”korkeutta” suhteessa normaalijakaumaan. Normaalijakaumassa huipukkuusluvun arvo on nolla. Mitä terävämpihuippuisempi (eli ”korkeampi”) jakauma on, sitä suuremman positiivisen arvon huipukkuusarvo saa. Ja mitä matalampi jakauman huippu on, sitä suurempi negatiivinen arvo huipukkuusluvuksi tulee. Viimeksi mainittu voi kuvata myös sitä, että jakaumassa on useita huippuja. Jakauman voidaan katsoa noudattavan normaalijakaumaa, jos tämä tunnusluvun arvo sijoittuu välin -1 ja + 1 sisään. Varminta on kuitenkin tunnusluvun lisäksi tarkastella jakauman muotoa myös visuaalisesti histogrammikuvion avulla. Yllä olevan vinousluvun yhteydessä kuvattu luottamusvälitulkinta on käyttökelpoinen aivan vastaavaan tapaan myös tässä yhteydessä (eli kuuluuko lukuarvo nolla luottamusvälille huipukkuusluku ± 2*SE).

Yhdessä vinous ja huipukkuus kuvaavat sitä, millainen tarkasteltavan muuttujan jakauma on normaalijakaumaan verrattuna.

Normaalijakauma ja normaalisuuden testaus

Normaalijakauma (Gaussin jakauma) on tilastotieteessä yksi tärkeimmistä muuttujien jakaumaa kuvaavista käyristä, jota yleisesti käytetään käytännön ilmiöiden matemaattisena mallina, koska monet reaalimaailman ilmiöt noudattavat vähänkään laajemmassa otannassa tätä. Normaalijakauman muoto on symmetrinen, jossa eniten havaintoja keskittyy jakauman keskikohtaan ja jossa havaintojen määrä pienenee asteittain reunoja kohti mentäessä. (Ks. Cumming & Calin-Jagerman 2017, 72–78; Heikkilä 2014, 99–102.) Normaalijakauma on käytännön tutkimuksen kannalta tärkeä muun muassa siksi, että monien tilastollisten menetelmien ja testien käytön edellytys on, että muuttujan jakauma noudattaa normaalijakaumaa (parametriset testit). Mikäli normaalijakautuneisuus ei ole voimassa, käytetään epäparametrisia menetelmiä. Jakaumaa ja sen muotoa voidaan tarkastella tulostamalla SPSS-ohjelmasta histogrammi (ks. sivu 98) ja arvioimalla silmämääräisesti havaintojen jakautumista. Muuttujan mediaanin, moodin ja keskiarvon ollessa lähellä toisiaan muuttuja on ainakin hyvin lähellä normaalijakaumaa. Jakauman normaalisuutta voi arvioida myös vinoutta ja huipukkuutta kuvaavien lukujen avulla (ks. yllä). Jakauman poikkeavuutta normaalijakaumasta voi selvittää myös Kolmogrov-Smirnovin tai Shapiro-Wilkin testeillä (ks. s. 108). Näiden tulkintaan liittyy hiemankin isompien aineistojen kohdalla kuitenkin ongelmana se, että nämä tulkitsevat varsin herkästi jakaumien poikkeavan normaalijakaumasta ($p < 0{,}05$). Tästä johtuen suuremmissa aineistoissa arviointi kannattaa monesti tehdä keskiarvoa ja mediaania vertailemalla, vinous- ja huipuukkuuslukujen avulla sekä visuaalisesti graafisten kuvioiden avulla.

on selvittää, onko aineiston muuttujissa tuloksia vääristäviä ääriarvoja tai selvästi poikkeavia havaintoja (outliers). Ennen kuin siis tekee tällaisissa tilanteissa keskiarvojen perusteella pitkälle meneviä päätelmiä, kannattaa kokeilla ääritapauksien rajaamista tutkimusaineistosta sen selvittämiseksi, kuinka paljon niiden poissulkeminen muuttaisi keskiarvoa. Varmimmin poikkeavien tapausten olemassaolon havaitsee tutkimalla muuttujan frekvenssijakauman pienimpiä ja suurimpia arvoja.

Keskiarvoon perustuvia tuloksia esitettäessä on syytä ilmaista aina myös niihin liittyvä jakauman hajontaa kuvaava tunnusluku keskihajonta. Tilastolliset ohjelmat tulostavat tämän yleensä automaattisesti, mutta esimerkiksi graafisiin diagrammiesityksiin, jos ne eivät liity niihin automaattisesti, ne kannattaa usein lisätä. Keskihajontaluku ei suoraan ilmaise muuttujassa ilmenevää varianssin suuruutta keskiarvoon nähden, vaan se on muuttujan varianssin neliöjuurta kuvaava tunnusluku, josta voidaan päätellä, miten arvot ovat keskimäärin sijoittuneet keskiarvon ympärille. Keskihajontalukuja käyttämällä voidaan siis jossakin määrin paikata keskilukujen käyttöön liittyvää informaatiokatoa (katso oheinen tietolaari 3). Näin saadaan tärkeää informaatiota, joka auttaa hahmottamaan käsiteltävää asiaa laaja-alaisemmin kuin pelkän keskiarvon varassa.

Keskiarvojen ja muiden keskeisten tunnuslukujen avulla saadaan havaintoaineiston numeerisista muuttujista luotua yleiskuva sekä arvioitua eri analyysimenetelmien soveltuvuutta aineistomme analysointiin. Muodostetaan havaintomatriisiesimerkin numeerisista muuttujista jakauman tunnusluvut ja normaalijakaumatesti (jätetään pois sukupuoli, vanhempien koulutustaso ja viriketaustamuuttuja, koska niiden keskiarvokäsittely ei ole tarkoituksenmukaista), joista saadaan seuraavanlainen taulukko:

TAULUKKO 1. Havaintoaineiston numeeristen muuttujien jakaumien tunnuslukuja ja normaalijakaumatesti.

	N	keski-arvo	mini-mi	mak-simi	keski-hajonta	vino-us	huipuk-kuus	Kolmogorov-Smirnov (p)
Verbaalisuustesti	60	25,47	20	30	2,55	-0,41	-0,53	p=0,002
Järkeilytesti	60	35,00	31	40	1,81	0,26	0,70	p=0,002
Kieliaineiden ka	60	79,73	59	97	10,30	-0,16	-0,77	p≥0,200
Matematiikan nro	60	7,23	5	10	1,09	0,48	0,20	p<0,001
Opintomenestys 1	60	15,83	13	19	1,82	-0,06	- 1,07	p=0,004
Opintomenestys 2	60	16,00	13	20	2,54	0,34	- 1,56	p<0,001

Tunnuslukutaulukosta voidaan todeta kunkin muuttujan keskiarvo ja muut keskeiset tunnusluvut. Näiden lisäksi tunnuslukuihin liittyen voitaisiin muodostaa luottamusvälit (ks. luku 1.6.1.4; luottamusväliä tarkastellaan myös jatkossa eri analyysimenetelmien yhteydessä). Tarkastellaan esimerkin vuoksi yleistä lukioaikaista opintomenestystä kuvaavaa muuttujaa ('Opintomenestys1'). Keskiarvon perusteella voidaan todeta otoksen keskiarvon (15,8) sijoittuvan hieman käytettävän skaalan (minimi 10, maksimi 20) puolenvälin paremmalle puolelle. Keskihajontaluku (1,8) ilmaisee hajonta-alueen vaihtelevan noin kaksi yksikköä keskiarvon molemmin puolin. Keskihajonnan avulla voidaan arvioida myös eri ryhmien heterogeenisuutta tai homogeenisuutta. Jakauman vinousarvo (skewness) ilmaisee muuttujan jakauman symmetrisyyden. Jakauma on symmetrinen, jos sen vasen ja oikea puoli ovat samanlaiset (keskiarvon kohdalta katsottuna). Normaalijakauma on symmetrinen, ja sen vinous saa arvon nolla. Muuttujan 'Opintomenestys1' jakauma on siis melko lailla symmetrinen, koska sen vinous (-0,06) poikkeaa hyvin vähän arvosta nolla. Etumerkki tarkoittaa sitä, mihin suuntaan jakauma on kallellaan (positiivisesti tai negatiivisesti). Positiivinen arvo tarkoittaa sitä, että jakauman oikea häntä on pitkä eli jakauma on kallellaan vasempaan. Negatiivisen arvon kohdalla tilanne on päinvastainen. Jos vinous on suuri, voidaan ajatella, että otoksen ilmoitettu keskiarvo on jossakin määrin erheellinen (jonkin ääriarvon vääristämä). Mitä suurempi vinousluku on, sitä heikommin keskiarvo edustaa todellista keskimääräistä arvoa, joten sellaisissa tilanteissa kannattaa laskea myös mediaanin arvo, koska mediaani ei ole herkkä tällaisille ”poikkeaville” arvoille. Täten aritmeettinen keskiarvo ei sovellu kovin vinojen jakaumien kuvaukseen eli tapauksiin, joissa on mukana toisista erittäin paljon poikkeavia arvoja.

Jakaumaa voidaan kuvata myös sen huipukkuuden (kurtosis) avulla. Tässäkin vertailukohtana on normaalijakauma, sillä sen huipukkuus saa arvon nolla. Jos muuttujan huipukkuus on negatiivinen, kyseisen muuttujan jakauma on normaalijakaumaa ”lättänämpi”. Positiivinen arvo tarkoittaa vuorostaan sitä, että jakauma on normaalijakaumaa ”terävämpi”. Lukion opintomenestysmuuttujan huipukkuudeksi on tullut arvo -1,07. Siten tämän muuttujan jakauma on normaalijakaumaa litistyneempi.

Taulukkoon on lisätty myös muuttujien normaalisuutta testaavan tilastollisen testin merkitsevyystasot. Testinä on ollut Kolmogorov-Smirnov Lillieforsin korjauksella. Näistä paljastuu käytettävän esimerkkiaineiston yksi suuri ongelma. Tiukasti testin arvoja tulkittaessa vain yhden muuttujan eli ’Kieliaineiden ka’ jakaumaa voidaan pitää normaalijakaumaa noudattavana (muista kuitenkin edellä näihin normaalisuusjakaumatesteihin liittyvä yliherkkyysongelma, ks. tietolaari 3, s. 103). Näin ei periaatteessa voitaisi käyttää myöhemmin esiteltäviä keskiarvotestejä, kuten t-testejä ja varianssianalyysia, muuttujia analysoitaessa, vaan tiukasti ottaen meidän tulisi käyttää näiden niin sanottuja epäparametrisia vastineita. Esimerkin vuoksi seuraavilla sivuilla analyysit tullaan tekemään kuitenkin pääasiallisesti keskiarvotesteillä. Samalla kuvataan lyhyesti näitä vastaavat epäparametristen analyysien käyttöä. Kuten jatkossa huomataan, parametriset ja epäparametriset testit antavat yleensä hyvin samansuuntaiset tulokset (p-arvot), jopa siinä määrin samansuuntaiset, että oikeassakaan analyysissa keskiarvotestien käytölle ei näin tutkijan näkökulmasta katsoen ole estettä, jos tutkimusaineiston otoskoko on riittävän suuri. Näin voidaan sanoa, esimerkiksi niin sanotun todennäköisyyslaskennan keskeisen raja-arvolauseen perusteella: koska keskiarvotesteissä p-arvo lasketaan testisuureen jakauman avulla ja testisuureen jakauma lähestyy normaalijakaumaa otoskoon kasvaessa, vaikka muuttujan jakaumaan liittyvä oletus ei täysin täyttyisikään.

SPSS: Keskiarvojen ja muiden jakauman tunnuslukujen laskeminen

Numeerisiin muuttujiin liittyvän taulukon 1 (s. 106) luvut on koottu kahdesta eri analyysista. Ensimmäisessä ajossa on saatu muuttujien jakaumia kuvaavia tunnuslukuja ja toisessa on testattu muuttujien normaalijakaumaoletuksen paikkansapitävyyttä. Ensimmäisen ajon saat toteutettua seuraavasti:

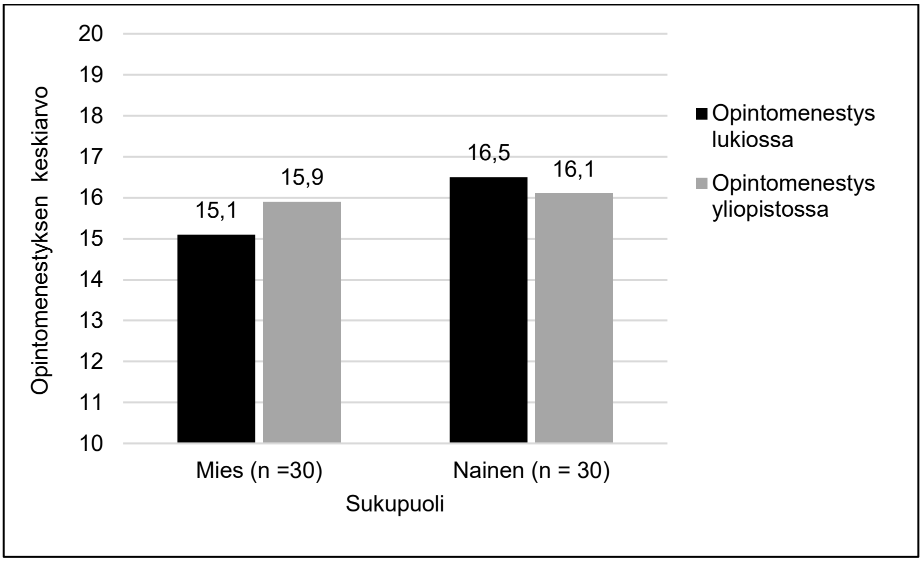
1. Valitse ylhäältä valikkoriviltä *Analyze → Descriptive Statistics → Descriptives*
   - siirrä valitsemasi (määrälliset) muuttujat muuttujaluettelon oikealla puolella olevalla nuolipainikkeella kohtaan *Variable(s)*
2. *Options* -painikkeen takaa pääset valitsemaan analyysissä tulostettavat tunnusluvut
   - oletusvalintojen lisäksi ruksaa kohdat *Kurtosis* (jakauman huipukkuutta kuvaava tunnusluku) ja *Skewness* (jakauman vinoutta kuvaava tunnusluku), hyväksy valinnat klikkaamalla *Continue*-painiketta
3. Toteuta tunnuslukujen laskeminen yllä määrittelemilläsi valinnoilla painamalla *OK* -painiketta ja saat tunnuslukutaulukot tulosteikkunaan

Itse asiassa edellä lasketut tunnusluvut voit saada samassa analyysissa muuttujien normaalijakaumatestien kanssa. Tämä toinen tapa tuottaa kuitenkin hyvin pitkän tulostelistauksen, joten edellä esitetty tapa saattaa olla käyttökelpoisempi silloin, kun et halua normaalijakaumatestauksia ja haluat muutenkin tiivistetymmän tulosteen. Toinen vaihtoehto toteutettaisiin seuraavasti:

1. Valitse ylhäältä valikkoriviltä *Analyze → Descriptive Statistics → Explore*
   - siirrä valitsemasi (numeeriset) muuttujat muuttujaluettelon oikealla puolella olevalla ylimmällä nuolipainikkeella kohtaan *Dependent List*
2. *Plots* -painikkeelta
   - varmista, että kohta *Normality plots with tests* on valittuna, hyväksy valinnat klikkaamalla *Continue*-painiketta
3. Toteuta tunnuslukujen laskeminen ja normaalisuustestaus yllä määrittelemilläsi valinnoilla painamalla *OK* -painiketta ja saat taulukot tulosteikkunaan

Normaalijakaumatestinä on *Kolmogorov-Smirnov* -testi, jossa merkitsevyystasoa laskettaessa on käytetty *Lillieforsin* korjausta. Pienemmillä otoksilla samaan taulukkoon tulostuva Shapiro-Wilk -testi on suositeltavampi. Tällä *Explore*-toiminnolla tunnusluvut ja testit ovat saatavissa myös ryhmittäin. Tällöin valitset ryhmittelevän tekijän, esimerkiksi sukupuolien, kohtaan *Factor List*, jolloin saat tulokset erikseen miehille ja naisille. *Factor List* -valintaruutuun voi valita myös esimerkiksi kaksi kategorista muuttujaa, jolloin tulokset saadaan näiden ryhmittelevien tekijöiden kaikissa erilaisissa ryhmäkombinaatioissa.

Keskiarvotaulukon avulla on siis mahdollista tulkita aineistoa, ennen kaikkea kuvata muuttujien yleispiirteitä. Taulukon avulla voidaan pohtia alustavasti myös jatkoanalyyseja. Keskiarvoja voidaan hyödyntää myös muuttujan graafiseen kuvaukseen. Esimerkkiaineistossa luontevin esimerkki tällaisesta kuviosta olisi opintomenestysmuuttujien vertailu, joko sellaisinaan tai esimerkiksi sukupuolittain tarkasteltuna, kuvio näyttäisi seuraavanlaiselta:



KUVIO 7. Naisten ja miesten opintomenestys lukiossa ja yliopistossa.

Kuviosta 7 ilmenee selkeästi sukupuolten välisen opintomenestyksen erilaisuus. Lyhyesti kuvattuna: lukiossa naisten opintomenestys (keskiarvo 16,5) on miehiä (15,1) parempi. Sen sijaan otoksemme miesten opintomenestys paranee, syystä tai toisesta, yliopistovaiheessa selkeästi, kun taas naisten opintomenestystaso jopa laskee lukioon verrattuna jonkin verran. Yliopistotasolla naisten ja miesten opintomenestystaso näyttäisi siis tasoittuvan lukioon verrattuna. Näin diagrammiesityksenkin avulla voi analysoida alustavasti omaa tutkimusaineistoaan, jolloin voi löytyä mielenkiintoisia uusia tulkintanäkökulmia ja miksei uusia tutkimusongelmiakin. Seuraavassa ohjeet siihen miten edellä oleva pylväskuvio tehdään SPSS:ssä.

SPSS: Pylväsdiagrammin piirtäminen (tässä tilanne, jossa yksi kategorinen ja kaksi numeerista muuttujaa)

1. Valitse ylhäältä valikkoriviltä *Graphs* → *Legacy Dialogs* → *Bar*
   - valitse esiin avautuvasta valikosta vuorostaan *Clustered*-kuvake ja kohdan *Data in Chart Area* vaihtoehdoista ruksaa *Summaries of separate variables*. Jatka *Define*-painikkeella eteenpäin.
   - valitse muuttujalistasta haluamasi kategorinen muuttuja ja siirrä se listan oikealla puolella olevalla nuolipainikkeella kohtaan *Category Axis*. Esimerkkikuviossamme tämä muuttuja on sukupuoli. Pylväät piirretään siis kategorisen muuttujan eri arvoille
   - kohtaan *Bars Represent* valitset sen määrällisen muuttujan tai ne muuttujat, joista haluat pylväät. Esimerkissämme on valittu kaksi opintomenestysmuuttujaa.

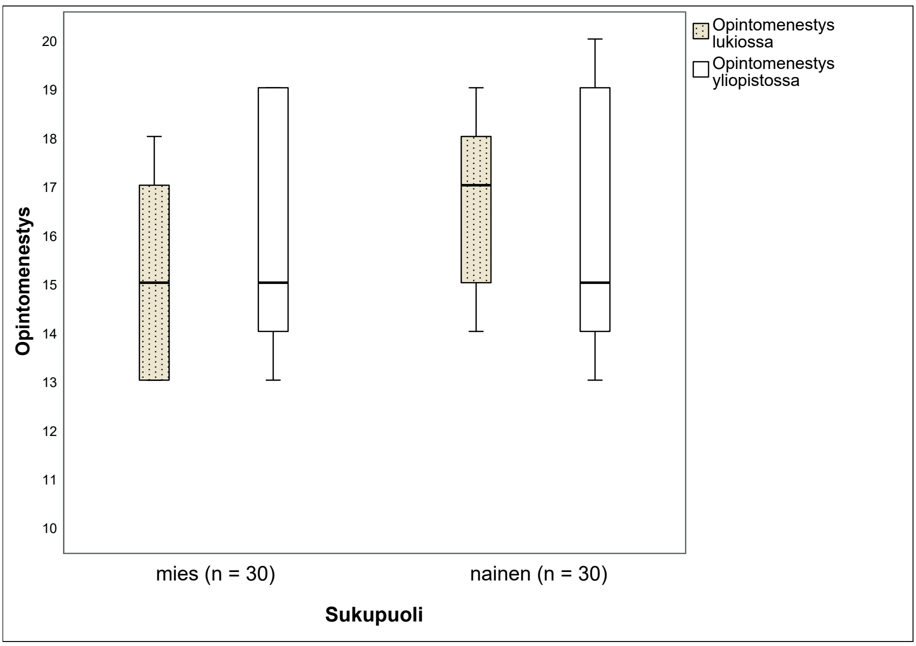
ohjetaulukko jatkuu...

- painikkeesta *Change Summary* määräät sen, mitä pylväät esittävät, esim. että kyseessä ovat keskiarvot (*Mean of values*)

2. Tarkista *Options*-painikkeesta, että mahdolliset puuttuvat tiedot eivät tule omaksi pylvääkseen, eli poista valintaruksi kohdasta *Display groups defined by missing values,* hyväksy valinnat klikkaamalla *Continue*-painiketta
3. Toteuta pylväsdiagrammin piirtäminen yllä määrittelemilläsi valinnoilla painamalla *OK* -painiketta ja saat kuvion tulosteikkunaan.

Halutessasi voit jatkaa kuvion muokkaamista tulosteikkunassa (ks. aiemmin olleet ohjeet sivulla 97).

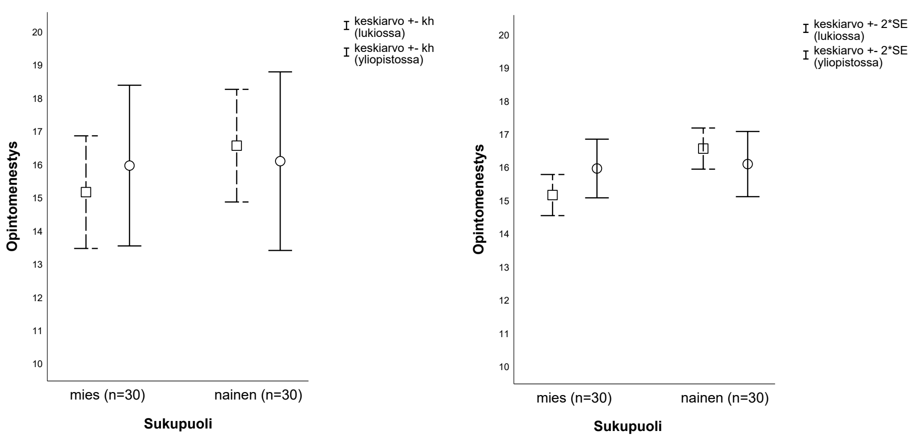
Tarkastellaan edellä esiteltyä kysymyksenasettelua vielä vaihtoehtoisilla graafisilla esitystavoilla. Näiden avulla kuvioon saadaan peruspylväskuviota enemmän informaatiota, kuten alla olevat laatikko-jana -esimerkit osoittavat. Sama asia kuin kuviossa 7 esitettiin, näyttää seuraavanlaiselta laatikko-jana -esityksenä:



KUVIO 8a. Naisten ja miesten opintomenestys lukiossa ja yliopistossa kuvattuna laatikko-jana -kuviolla.

Kuvio 8a:n perinteisessä laatikko-jana -esityksessä (Boxplot) kuvataan molemmille muuttujille samassa graafissa ryhmien mediaanit (laatikon sisällä oleva viiva), kvartiilivälit (laatikon ylä- ja alareunat) sekä vaihteluväli kun mukana on myös poikkeavammat havainnot (janat laatikon ylä- ja alapuolella). Katso tarkempi kuvaus esimerkiksi Grönroos (2011, 43–44). Näin mediaanilla kuvatun jakauman sijainnin lisäksi päästään samassa kuviossa havainnollistamaan myös havaintoarvojen hajontaa.

Perinteisen laatikko-jana -kuvion lisäksi jakauman sijaintia ja hajontaa voidaan kuvata myös muilla hieman samankaltaisilla graafeilla, joissa on käytetty muita keski- ja hajontalukuja. Alla olevassa kuviossa 8b on käytetty perinteisen laatikko-jana -esityksen modifikaatiota, jossa keskilukuna on keskiarvo ja hajontaa kuvataan keskihajonnalla tai keskiarvon keskivirheellä. Jos tätä verrataan aiemmin esitettyyn pylväsdiagrammiin (kuvio 7, s. 109), tässä saadaan keskiarvojen lisäksi esiin myös muuttujien arvojen hajontoihin liittyvää tietoa. Kuvio on siis pylväsdiagrammia informatiivisempi. Kuviossa pienet neliösymbolit kuvaavat keskiarvoja. Vasemmanpuoleisessa graafissa janat kuvaavat aluetta, joka on yhden keskihajonnan päässä keskiarvon molemmin puolin, jana on siis kokonaisuudessaan pituudeltaan 2 x kulloisenkin muuttujan keskihajonta. Kuvion tulkintaa voi helpottaa normaalijakauman ominaisuuksiin liittyvä tieto, että normaalijakautuneella muuttujalla tällainen vaihteluväli (ka ± 1 x kh) sisältää noin 70 % havaintoarvoista. Keskihajontoja tarkastelemalla nähdään, että opintomenestyksen hajonta yliopistossa on lukiota suurempaa sekä miesten että naisten kohdalla. Yliopistollisissa opinnoissa keskihajonta siis kasvaa molemmissa sukupuoliryhmissä, naisilla miehiä hivenen enemmänkin. Syy keskihajontojen kasvuun voi olla se, että opinnot yliopistossa ovat lukiota vaikeampia, ja se kasvattaa menestyseroja.



KUVIO 8b. Naisten ja miesten opintomenestys lukiossa ja yliopistossa kuvattuna error-bars -kuviolla (vasemman kuvion janat liittyvät keskihajontoihin, oikealla olevan kuvion janat puolestaan liittyvät keskiarvon keskivirheisiin).

Edellä piirretyssä kuviossa vasemmanpuoleisessa graafissa janat kuvasivat muuttujien keskihajontoja. Jos annetaan janojen kuvastaakin otoskeskiarvon estimointitarkkuuteen liittyviä niin sanottuja keskivirheitä (standard error, SE), niin saadaan ulkoisesti samannäköinen, mutta tulkinnaltaan aivan erilainen kuvio. Tässäkin kuviossa (kuvio 8b oikeanpuoleinen graafi) pienet neliösymbolit kuvaavat keskiarvoja, mutta janat ovatkin nyt keskivirheitä eli otoskeskiarvon keskihajontoja (mittaa keskiarvojen lukuarvoissa esiintyvää vaihtelua, mikäli vastaavan kokoinen otanta toteutettaisiin useita kertoja ja kustakin laskettaisiin keskiarvo). Otoksesta laskettuun tunnuslukuun liittyvä keskivirhe saadaan laskettua estimaattorin teoreettisten ominaisuuksien perusteella, hyödyntäen muuttujan keskihajontaa ja otoskokoa, eli tulkintaan tulee mukaan populaatio-otos -ajattelu ja luottamusvälien käyttöön liittyvä tarkastelu. Jos janojen pituus valitaan kuviossa esitetyllä tavalla (keskiarvo ± 2*SE), se tarkoittaa sitä, että noin 95 prosentin todennäköisyydellä todellinen keskiarvo eli populaatiokeskiarvo on janan rajaaman alueen sisällä. Janojen pituus on siis itse asiassa valittu siten, että se vastaa suunnilleen merkitsevyystasoa $p = 0,05$. Tämä vuorostaan tarkoittaa sitä, että janojen pituuden avulla pystytään alustavasti päättelemään ryhmien välisten erojen tilastollinen merkitsevyys. Tulkinnan lähtökohtana on janojen sijainti toisiinsa nähden: kun kaksi janaa (päällekkäin siirrettyinä) eivät kohtaa toisiaan, kyseisten kahden ryhmän välinen keskiarvoero on tilastollisesti merkitsevä (merkitsevyystasolla $p = 0,05$). Kummankin ryhmän populaatiokeskiarvo on siis jossakin oman janansa sisällä, ja jos janat eivät kohtaa toisiaan, voidaan suurella luotettavuudella päätellä, että populaatiokeskiarvot eivät ole samat, eli otoskeskiarvoissa havaittu ero on tilastollisesti merkitsevä.

Jos tarkastellaan miesten ja naisten välisiä opintomenestyseroja, voidaan kuviosta huomata kiinnostavia seikkoja. Ensinnäkin lukiotasolla naisten ja miesten välisen opintomenestyksen ero on tilastollisesti merkitsevä. Naisten opintomenestys on miehiä parempi. Sen sijaan yliopistotasolla sukupuolten välillä opintomenestysero ei ole enää tilastollisesti merkitsevä. Tämä selittyy sillä, että miesten opintomenestys kohoaa lukiosta yliopistoon siirryttäessä (tosin ei tilastollisesti merkitsevästi), kun taas naisten opintomenestys laskee yliopistoon tultaessa jonkin verran. Näin naisten ja miesten opintomenestys on yliopistossa lähes yhteneväinen. Esimerkin tilanteessa osa keskiarvojen vertailuista voisi liittyä toisistaan riippuvien otosten tarkasteluun, jolloin keskiarvo-erojen vertaamista ei niiltä osin voida tehdä yhtä suoraviivaisesti yllä esitetyllä tavalla (janojen päällekkäisyys/erillisyys). Tällainen tarkastelu voisi olla esimerkiksi, että onko miesten opintomenestyksessä eroa lukioon verrattuna. Kyseessä on toistetusti mitatut tiedot samoilta miesopiskelijoilta, ts. vertailtavat ryhmät eivät nyt ole toisistaan riippumattomia. Tällaisessa tilanteessa 95 %:n luottamusväli olisi muodostettava opintomenestyserojen (lukio miinus peruskoulu) keskiarvolle. Näin muodostettujen, arvioinnin luotettavuutta kuvaavien jonojen perusteella voitaisiin tulkita, josko arvo nolla (joka kuvaisi, että populaatiossa eroa ei esiinny) sijaitsee muodostuneiden janojen ulkopuolella. Tällöin tulkinta olisi, että miehillä menestys lukiossa on merkitsevästi erilaista, esimerkiksi parempaa, kuin peruskoulussa. Luottamusvälien käyttöä on käsitelty tarkemmin luvussa 1.6.1.3. ja 4.1.1.1. Alla olevassa ohjelaatikossa on kuvattu laatikko-jana -kuvion ja error-bar -kuvion SPSS-toteutus.

SPSS: Laatikko-jana -esityksen piirtäminen (useita muuttujia):

1. Valitse ylhäältä valikkoriviltä *Graphs → Legacy Dialogs → Boxplot*
   - esiin tulevista kuvakkeista valitse vuorostaan *Clustered*-kuvake, ja kohdan *Data in Chart Area* vaihtoehdoista ruksaa *Summaries of separate variables*
   - jatka *Define*-painikkeella eteenpäin
2. Valitse muuttujalistasta haluamasi kategorinen muuttuja, ja siirrä se listan oikealla puolella olevalla nuolipainikkeella kohtaan *Category Axis*. Pylväät piirretään siis kategorisen muuttujan eri arvoille
   - kohtaan *Boxes Represent* valitset sen muuttujan tai ne muuttujat, joista haluat laatikkojana -esityksen
3. *Options*-painike
   - tarkista, että mahdolliset puuttuvat tiedot eivät tule mukaan, eli poista valintaruksi kohdasta *Display groups defined by missing values,* hyväksy valinnat klikkaamalla *Continue*-painiketta
4. Toteuta pylväsdiagrammin piirtäminen yllä määrittelemilläsi valinnoilla painamalla *OK*-painiketta ja saat kuvion tulosteikkunaan

Halutessasi voit jatkaa kuvion muokkaamista tulosteikkunassa (ks. aiemmat ohjeet, s. 97).

SPSS:n laatikko-jana -esityksessä janojen päätepisteinä ovat pienin ja suurin arvo, laatikon yläreuna on yläkvartiili ja alareuna on alakvartiili; laatikon poikkiviiva on mediaani. Mikäli muuttujalla on joitain poikkeavan suuria tai pieniä havaintoarvoja, janojen päätepisteet asettuvat 1.5 * kvartiilivälin etäisyydelle laatikon ylä- ja alareunasta. Tällöin poikkeavat arvot näkyvät kuviossa erillisinä symboleina.

Alakvartiilin, mediaanin ja yläkvartiilin avulla tehty laatikko-jana -esitys ei kuitenkaan aina ole havainnollinen, vaan keskiarvojen avulla saadaan usein aikaan erottelevampi kuvio.

Tämän teet seuraavasti:

SPSS: Keskiarvoja ja arvojen vaihtelua kuvaavan error-bar -graafin muodostaminen:

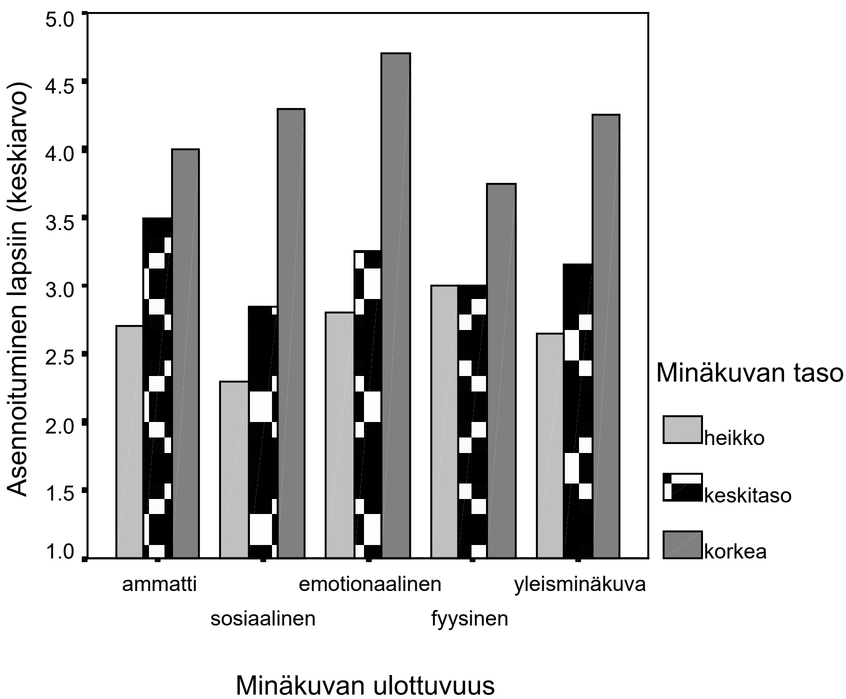
1. Valitse ylhäältä valikkoriviltä vaihtoehdot *Graphs → Legacy Dialogs → Error Bar*
   - esiin tulevista kuvakkeista valitse vuorostaan *Clustered*-kuvake, ja kohdan *Data in Chart Area* vaihtoehdoista ruksaa *Summaries of separate variables*
   - jatka *Define*-painikkeella eteenpäin

ohjetaulukko jatkuu...

2. Valitse muuttujalistasta haluamasi kategorinen muuttuja, ja siirrä se listan oikealla puolella olevalla nuolipainikkeella kohtaan *Category Axis*. Pylväät piirretään kategorisen muuttujan eri arvoille. Kohtaan *Variables* valitset sen muuttujan tai ne muuttujat, joista haluat error-bar -esityksen. *Bars Represent* -kohdasta pystyt valitsemaan sen, mitä kuvion janat esittävät: esimerkiksi muuttujan keskihajontaa (*Standard deviation*) tai keskiarvon keskivirhettä (*Standard error of mean*)
   - Jos valitset keskivirhevaihtoehdon, silloin myös kerroinkohta (*Multiplier*) on merkityksellinen; arvo 2 vastaa suunnilleen merkitsevyystasoa 0,05
   - kun siis käytät tätä vaihtoehtoa, kuviosta on tulkittavissa, ovatko kuvion esittämien keskiarvojen erot tilastollisesti merkitseviä (merkitsevyystasolla 0,05).
3. *Options*-painike
   - tarkista, että mahdolliset puuttuvat tiedot eivät tule mukaan, eli poista valintaruksi kohdasta *Display groups defined by missing values*.
4. Paina lopuksi *Continue*- ja *OK*-painikkeita, kuvio piirtyy tulosteikkunaan. Halutessasi voit jatkaa kuvion muokkaamista tulosteikkunassa (ks. aiemmat ohjeet, s 97).

Tämä SPSS:n *Error bar* -kuvio on siis pelkistetty laatikko-jana -esitys, jossa ovat vain pienet laatikot (keskiarvot) ja janat (esim. keskihajonnat tai keskivirheet). Pienet laatikot vastaavat varsinaisen laatikko-jana -esityksen poikkiviivaa (mediaani), joten *Error bar* -kuviossa itse asiassa ei ole lainkaan varsinaista (isoa) laatikkoa.

Laatikko-jana tai error-bar -kuvioista voidaan tehdä siis varsin pitkällekin meneviä johtopäätöksiä. Toki perinteisten diagrammiesitystenkin käyttö on sallittua. Esimerkiksi seuraava kuvio on melko informatiivinen ilman janojakin. Ajatellaan tilannetta, että on tutkittu lastenkotien lapsien ja hoitajien keskinäistä suhdetta. Eri testien perusteella on saatu mitatuksi hoitajien asennoitumista lapsiin. Koska tutkija ei tyydy pelkästään kuvailevaan tutkimusasetelmaan, on hän päättänyt tutkia myös hoitajien asennoitumisen taustalla olevia selittäviä tekijöitä. Näin hän voi analysoida esimerkiksi hoitajien minäkuvan yhteyttä heidän asennoitumispistemääriinsä. Minäkuva on mitattu mittarilla tai menetelmällä, josta ilmenevät hoitajien minäkuvan eri ulottuvuudet. Saamiensa tulosten perusteella tutkija olisi muodostanut seuraavanlaisen pylväskuvion:



KUVIO 9. Lastenkotien hoitajien minäkuvien yhteys heidän asennoitumiseensa lapsiin (kuviteltu esimerkki), N=75.

Myös tästä kuviosta voi päätellä melko paljon: voidaan todeta yleisesti deskriptiivisenä tuloksena sen, että mitä positiivisempi minäkuva hoitajilla on, sen positiivisempi suhde heillä on lapsiin. Kuvion avulla voitaisiin vertailla myös minäkuvan eri ulottuvuuksia toisiinsa, mutta tässä yhteydessä näitä ei tarkastella tämän tarkemmin. Tässä kuviossa esitetty Minäkuvamuuttuja on nominaaliasteikollinen, jolloin graafi voitaisiin hyvin muodostaa myös vaakapylväikkönä, jossa ulottuvuus-luokat järjestettäisiin esimerkiksi kuviossa tummimpina näkyvien pylväiden (korkea minäkuvan taso) mukaiseen järjestykseen. Tässä toisaalta mukana on järjestysasteikollinen luokitteleva muuttuja ryhmittelijänä, jolloin myös tällainen pystypylväikkö on perusteltu esitystapa. Toisinaan siis kuviotyypin valinnassa ja muotoilussa joudutaan käyttämään omaa harkintaa, eikä kuvaamisessa ole yhtä ainoata oikeata tapaa. Edellä olleiden graafisten esitysten esimerkkien tarkoituksena oli osoittaa erilaisia mahdollisuuksia muuttujien havaintoarvoja kuvaavien tunnuslukujen esittämiseen.

SPSS: Pylväsdiagrammin piirtäminen (kaksi kategorista ja yksi numeerinen muuttuja):

1. Valitse ylhäältä valikkoriviltä *Graphs* → *Legacy Dialogs* → *Bar*
   - valitse avautuvasta valintaruudusta kuvake *Clustered* ja kohdan *Data in Chart Area* vaihtoehdoista ruksaa *Summaries for groups of cases*, jatka *Define*-painikkeella seuraavaan vaiheeseen
2. Valitse kuvioissa käytettävät muuttujat
   - valitse muuttujalistasta haluamasi kategorinen muuttuja, ja siirrä se listan oikealla puolella olevalla nuolipainikkeella kohtaan *Category Axis,* pylväät piirretään siis kategorisen muuttujan eri arvoille
   - kohtaan *Define Clusters by* siirrät toisen kategorisen muuttujan. Pylväät esitetään tämän toisen muuttujan muodostamina ryhminä
   - vaihtoehtojen *Bars Represent* kohtaan *Variable* siirrät sen numeerisen muuttujan, josta haluat pylväät
   - painikkeesta *Change Summary* voit valita esim. sen, että pylväät kuvaavat keskiarvoja (*Mean of values*)
3. Tarkista *Options*-painikkeesta, että mahdolliset puuttuvat tiedot eivät tule omaksi pylvääkseen, eli poista valintaruksi kohdasta *Display groups defined by missing values,* hyväksy valinnat klikkaamalla *Continue*-painiketta
4. Toteuta pylväsdiagrammin piirtäminen yllä määrittelemilläsi valinnoilla painamalla *OK*-painiketta ja saat kuvion tulosteikkunaan

Halutessasi voit jatkaa kuvion muokkaamista tulosteikkunassa (ks. aiemmat ohjeet, sivuilta 97)

Tarkastellaan vielä sitä, miten SPSS-ohjelman avulla saadaan laskettua edellä esitettyjen tapaisia ryhmittäisiä keskiarvoja ja keskihajontoja. Verrataan esimerkkiaineiston kaikkien numeeristen muuttujien keskiarvoja ja keskihajontoja sukupuolen mukaan ryhmitettyinä taulukkomuodossa:

TAULUKKO 2. Esimerkkiaineiston numeeristen muuttujien keskiarvot ja keskihajonnat sukupuolittain.

		Sukupuoli	
		Miehet (n=30)	Naiset (n=30)
Verbaalitesti	Ka	24,33	26,60
	Kh	2,34	2,25
Järkeilytesti	Ka	35,80	34,13
	kh	1,83	1,38
Kieliaineiden ka	Ka	74,13	85,33
	Kh	8,65	8,72
Matematiikan numero	Ka	7,60	6,87
	Kh	1,16	0,90
Opintomenestys (lukio)	Ka	15,13	16,53
	Kh	1,70	1,70
Opintomenestys (yliopisto)	Ka	15,93	16,07
	Kh	2,42	2,69

Taulukosta on luettavissa selkeästi ja tarkasti eri muuttujien erot sukupuolittain. Tässä yhteydessä taulukkoa ei liene tarpeen selittää tämän tarkemmin. Meidän kannalta tärkeää on todeta tämänkin kaltaisen asetelman mahdollisuus. SPSS:ssä ryhmittäinen muuttujien taulukkomuotoinen kuvaus onnistuu esimerkiksi seuraavasti:

SPSS: Muuttujien keskiarvojen ja keskihajontojen laskeminen kategoristen muuttujien eri luokissa

1. Valitse ylhäältä valikkoriviltä *Analyze* → *Compare Means* → *Means*
   - siirrä valitsemasi numeeriset muuttujat muuttujaluettelon oikealla puolella olevalla ylemmällä nuolipainikkeella kohtaan *Dependent List*
   - haluamasi kategorisen muuttujan tai kategoriset muuttujat siirrät muuttujaluettelon oikealla puolella olevalla alemmalla nuolipainikkeella kohtaan *Independent List*
2. *Options* -painike
   - tässä voit valita, mitä tunnuslukuja haluat tulostetaulukkoon. Oletusvalintoina on keskiarvo, keskihajonta ja ryhmäkoko. Näiden lisäksi voit valita vasemmanpuoleisesta luettelosta myös muita tunnuslukuja, joita voit siirtää mukaan tarkasteluun keskellä olevalla nuolella, hyväksy valinnat klikkaamalla *Continue*-painiketta
3. Toteuta ryhmittäin laskettujen tunnuslukujen laskeminen yllä määrittelemilläsi valinnoilla painamalla *OK* -painiketta ja saat taulukot tulosteikkunaan

Taulukon dimensioita ja muita muotoiluja pääset muuttamaan ylävalikon valikoista *Format* ja *Pivot* (kunhan ensin olet kaksoisklikannut taulukkosi muokkaustilaan).

Huom: ylhäältä valikkoriviltä *Analyze* → *Tables* → *Custom Tables* valitsemalla saat tehtyä myös mm. ryhmittäisiä taulukointeja. Tämä toiminto on enemmän interaktiivinen, joten siinä kannattaa edetä kokeillen.

Sijaintiluvuista käsitellään lopuksi vielä hieman tarkemmin mediaania, jonka tulkintaa sivuttiin aiemmin laatikko-jana -kuvion yhteydessä (ks. kuvio 8a, s. 110). Mediaanin käyttö tulee kyseeseen ennen kaikkea silloin, kun otoksessa on toisistaan merkittävästi poikkeavia havaintoarvoja, arvojen jakauma poikkeaa muodoltaan normaalijakaumasta tai jos muuttujien asteikot eivät täytä välimatka-asteikon kriteereitä.

3.4 Mediaani

Mediaani (median, Md) on se luku, joka jakaa muuttujan suuruusjärjestyksessä olevat arvot kahteen osaan siten, että kumpaankin osaan tulee yhtä monta tapausta, jako on siis 50/50 prosenttia. Jos muuttujan arvojen määrä on pariton, silloin mediaani on keskimmäisin arvo. Jos arvoja on parillinen määrä, mediaani lasketaan kahden keskimmäisen arvon keskiarvona. Jos muuttujan suuruusjärjestyksessä olevien arvojen jako suoritetaan ala- (Q_1) tai yläkvartiilin (Q_3) avulla, alakvartiilin alapuolelle tai vastaavasti yläkvartiilin yläpuolelle tulee neljäsosa havaintoarvoista, eli jakosuhde alakvartiilin kohdalla on 25/75 prosenttia ja yläkvartiilin kohdalla 75/25 prosenttia. Ääriarvot eivät vaikuta mediaaniin tai kvartiileihin samalla tavoin kuin aritmeettiseen keskiarvoon. Keskiarvoon näiden vaikutus voi olla erittäin voimakas. Toinen ero on mitta-asteikkovaatimus; keskiarvolla se on välimatka- eli intervalliasteikko, kun taas mediaanille ja kvartiileille kelpaavat myös järjestysasteikolliset (ordinaaliasteikko) muuttujat. Mediaani ja kvartiilit ovat persentiilien (percentile) eli ”prosenttipisteiden” erikoistapauksia. Persentiili jakaa muuttujan suuruusjärjestyksessä olevat arvot kahtia minkä tahansa prosenttiosuuden perusteella. Esimerkiksi P_{10} tarkoittaisi sitä, että arvot on jaettu suhteessa 10/90 prosenttia, tätä kohtaa P_{10} kutsutaan myös jakauman ensimmäiseksi desiiliksi. Siten alakvartiilia voitaisiin merkitä myös P_{25}, mediaania P_{50} ja yläkvartiilia P_{75}.

Mediaaniesityksen haittapuolena on se, että monet yleisimmin käytetyt tilastoanalyysit perustuvat pitkälti keskiarvoihin, joten mediaaniesitys jää usein kovin irralliseksi muista käytetyistä testeistä. Yleisesti ottaen mediaanin käyttöä tulisi kuitenkin suosia nykyistä huomattavasti enemmän.

Lasketaan seuraavaksi mediaani ja kvartiilit esimerkin vuoksi samoista muuttujista kuin edellä keskiarvon yhteydessä. Saadaan seuraava taulukko:

TAULUKKO 3. Esimerkkiaineistomme numeeristen muuttujien mediaanit ja kvartiilit.

	Minimi (*Min*)	Alakvartiili(Q_1)	Mediaani (*Md*)	Yläkvartiili (Q_3)	Maksimi (*Max*)
Verbaalinen testi	20	24	26	27	30
Järkeilytesti	31	34	35	36	40
Kieliaineiden keskiarvo	59	72	80	87	97
Matematiikan numero	5	7	7	8	10
Opintomenestys 1	13	14	16	17	19
Opintomenestys 2	13	14	15	19	20

Esimerkissä muuttujat ovat kaikki määrällisiä, mutta mediaanin käyttö onnistuisi siis jo järjestysasteikollisten muuttujien kanssa. Taulukosta nähdään esimerkiksi se, että kieliaineiden keskiarvo -muuttujan mediaani on 80 (joka aineistossa merkitsee todistusarvosanaa 8). Tämän kummallekin puolelle sijoittuu 50 prosenttia tapauksista. Alimman eli kieliaineissa heikoimmin menestyneen neljänneksen arvot sijoittuvat välille 59–72, parhaan neljänneksen sijoittuessa pisteiden 87–97 väliin.

SPSS: Muuttujien mediaanin, ala- ja yläkvartiilin laskeminen

1. Valitse ylhäältä valikkoriviltä *Analyze → Descriptive Statistics → Frequencies*
   - siirrä valitsemasi muuttujat muuttujaluettelon oikealla puolella olevalla nuolipainikkeella kohtaan *Variable(s)*
   - poista ruksi kohdasta *Display frequency tables*, ellet samalla halua myös muuttujien frekvenssi- ja prosenttijakaumia
2. *Statistics* -painike
   - tässä pääset valitsemaan analyysiin mukaan tulevat tunnusluvut, kuten mediaanin (*Median*) sekä ala- ja yläkvartiilin (*Quartiles*). Yllä on otettu analyysiin mukaan myös muuttujan pienin (*Minimum*) ja suurin arvo (*Maximum*), hyväksy valinnat klikkaamalla *Continue*-painiketta
3. Toteuta näiden tunnuslukujen laskeminen yllä määrittelemilläsi valinnoilla painamalla *OK*-painiketta ja saat tunnuslukutaulukon tulosteikkunaan

Edellisessä alaluvussa hyödynnettiin esimerkinomaisesti ala- ja yläkvartiilia luokittelurajojen määrittelyssä. Näin muutettiin hieman epäortodoksisesti määrällinen numeerinen muuttuja kategoriseen muotoon. Arvojen luokittelu ja luokitellun muuttujan käyttö voi kuitenkin joissain tilanteissa olla perusteltua. Etenkin aloittelevalle tutkijalle luokiteltujen muuttujien käsittely ja tulkinta voi olla helpompaa ja selkeämpää verrattuna analysointiin vastaavia jatkuva-arvoisia muuttujia käyttäen. Samoin kuvaileva analyysi saattaa olla havainnollisemmin toteutettavissa luokitelluilla muuttujilla. Haittapuolena tässä luonnollisesti on informaation kadottaminen, kun toisiaan lähellä olevat arvot luokitellaan samaan luokkaan.

4. Kahden ryhmän keskiarvojen vertaaminen

Tässä luvussa keskitytään vertailutilanteeseen, jossa pyritään selvittämään, eroavatko kaksi ryhmää jonkin mitatun muuttujan suhteen toisistaan. Tällainen tutkimusasetelma on hyvin yleinen lähtökohta tutkimuksissa. Tutkija voi olla kiinnostunut esimerkiksi siitä, kuinka tytöt ja pojat menestyvät koulussa yleisesti tai eri aineissa, tai voidaan tutkia sitä, suoriutuvatko johonkin opetuskokeiluun osallistuneet (koeryhmä) kontrolliryhmään kuuluneita keskimäärin paremmin. Seuraavassa esitellään aluksi t-testi, jolla voidaan verrata kahden riippumattoman ryhmän keskiarvoja toisiinsa (luku 4.1.1). Toistettujen mittausten eli parittaisen t-testin (luku 4.1.2) avulla voidaan puolestaan verrata esimerkiksi saman ryhmän suoriutumiseroja alku- ja loppumittausten välillä tai, kuten esimerkkiaineistossa, sitä, miten opintomenetys eroaa lukio- ja yliopisto-opiskelun suhteen. T-testi on parametrinen testi, jossa käytön edellytyksenä on muun muassa normaalijakaumaoletuksen täyttyminen molemmissa testattavissa ryhmässä. Mikäli t-testin käytön edellytykset eivät ole voimassa, tulee valita kyseiseen tilanteeseen soveltuva epäparametrinen testi. Kahden ryhmän vertaamiseen voidaan käyttää epäparametrista järjestyslukuihin perustuvaa Mann-Whitneyn U-testiä (luku 4.2.1). Verrattaessa kahta riippuvaa otosta, esimerkiksi muutosta alkumittauksesta loppumittaukseen on Wilcoxonin testi epäparametrinen vastine toistettujen mittausten t-testille (luku 4.2.2).

4.1 T-Testi

Studentin t-testi on yksi käytetyimmistä menetelmistä kahden ryhmän välisten keskiarvoerojen arvioimiseksi. Testiä voidaan käyttää hyvinkin pienten aineistojen analysointiin, jos normaalijakaumaoletus on voimassa. Toisaalta, kuten aiemmin on jo esitetty, aineiston otoskoon kasvaessa normaalijakaumaoletuksen merkitys vähenee.

T-testin käyttö ja käytön edellytykset:

- asetelmassa vertaillaan numeerisen muuttujan keskiarvoja kategorisen muuttujan eri luokissa eli kyseessä on käytännössä kahden ryhmän keskiarvojen vertailemisesta; kategorisen muuttujan on oltava kaksiluokkainen, eli vertailtavia ryhmiä on oltava kaksi (jos vertailtavia ryhmiä on enemmän, menetelmänä voidaan käyttää varianssianalyysiä)
- numeerisen muuttujan mitta-asteikko on joko välimatka- tai suhdeasteikko (monesti sovelletaan myös Likert-asteikkoisille muuttujille)
- numeerisen muuttujan jakaumien tulee olla kummassakin ryhmässä normaalijakauman mukaisia
- numeerisen muuttujan populaatiovarianssien (hajontojen) tulee olla yhtä suuret eri ryhmissä (tämän voi testata Levenen testin avulla). Eri suurten varianssien tilanteissa t-testille tulostuu esimerkiksi SPSS-ohjelmassa vaihtoehtoinen, hieman eri tavalla laskettu t-testin versio.
- jos testin käytön edellytykset eivät ole voimassa, testin tilalla tulee käyttää esimerkiksi Mann-Whitneyn testiä (t-testin epäparametrinen vastine).

Studentin t-testi soveltuu siis kahden toisistaan riippumattoman ryhmän keskiarvojen eron testaukseen. T-testissä tarkasteltavan muuttujan tulee olla jatkuva (määrällinen). Likert-asteikkoa käytetään tässä tilanteessa kasvatustieteessä ja yhteiskuntatieteissä ja sen katsotaan täyttävän tämän kriteerin. Lisäksi t-testillä voidaan testata yhden ryhmän sisällä toistettavien mittaustulosten eron merkitsevyyttä (toistettujen mittausten t-testi).

Keskeisimmät tunnusluvut t-testissä ovat otoskeskiarvojen ja -keskihajontojen lisäksi t- ja p-arvot (t on testisuureen arvo ja p on havaittu merkitsevyystaso). Eri testien testisuureiden (esim. khiin neliö -testi, t-testi) arvot eivät ole keskenään vertailukelpoisia, mutta p-arvojen tulkinta on samanlainen. P-arvon avulla tehdään siis päätelmä siitä, onko esimerkiksi koe- ja kontrolliryhmien keskiarvojen välillä havaittu ero tilastollisesti merkitsevä. Nollahypoteesi olisi tässä tapauksessa H_0: ryhmien populaatiokeskiarvot ovat samat ja vaihtoehtoinen hypoteesi H_1: ryhmien populaatiokeskiarvot eroavat toisistaan. P-arvo kertoo sen riskin suuruuden, joka otetaan, mikäli nollahypoteesi jakaumien sijaintien samankaltaisuudesta hylätään. Kuten aiemmin on todettu, yleisimmin (perinteisen käytännön mukaisesti) tilastollisissa testeissä testin tulosta arvioidaan riskitasoa 0,05 käyttäen (5 %) siten, että p-arvon ollessa suuruudeltaan pienempi kuin 0,05 voitaisiin H_0-hypoteesi hylätä. (Ks. p-arvon tulkinnasta myös luku 1.6.1.1). Yleisemmin kasvatus- ja yhteiskuntatieteellisessä tutkimuksessa kyse on kuitenkin yksinkertaisesti kahden eri ryhmän keskiarvojen yhtäläisyyden tai eron tilastollisen merkitsevyyden tarkastelusta jonkin tietyn muuttujan suhteen. Jolloin on enemmänkin kyse siitä, voidaanko havaittu tulos yleistää riittävän luotettavasti tutkimuksen perusjoukkoon. Tämä ei kuitenkaan kerro mitään havaitun eron suuruudesta tai käytännön merkittävyydestä. Nykyisin yhä useammin korostetaankin sitä, ettei tuloksen tulkintaa pitäisikään jättää pelkän testin tuottaman p-arvon varaan, vaan testituloksen rinnalle kannattaisi tulkita ja raportoida myös luottamusväli- ja efektikoko-tarkastelut. Tämä on tärkeätä erityisesti suuria aineistoja analysoitaessa, jolloin tilastolliset testit tulevat herkästi merkitseviksi. Tämän jälkeen, kuten luvussa 1.6.2 on korostettu, näin muodostettua empiiristä tulkintaa tulee tarkastella vielä aiempaa tutkimustietoa ja asiaa koskevia teoreettisia tulkintoja vastaan. Oma empiirinen tulkinta tulee siis nostaa astetta teoreettisemmalle tasolle samalla kun tulkinta itsessään saa tätä kautta lisää luotettavuutta, vakuuttavuutta ja uskottavuutta.

4.1.1 Riippumattomien ryhmien t-testi

T-testi on kasvatustieteissä yleisesti käytetty ryhmien keskiarvojen erojen merkitsevyystason mittari. Kieltämättä se onkin useissa tapauksissa riittävä, mutta koska se perustuu keskiarvolaskelmiin, jotka eivät ota huomioon aineistossa mahdollisesti olevia poikkeavia havaintoja (outliers), se saattaa antaa erheellisiä tuloksia. Pienissä otoksissa ääriarvojen merkitys voi olla hyvinkin huomattava.

T-testissä (kuten keskiarvotesteissä ylipäätään) on periaatteessa kaksi tapaa määrittää vaihtoehtoinen hypoteesi. Toisessa tulkinnassa tehdään kaksisuuntainen vaihtoehtoinen hypoteesi (two-tailed), kun esimerkiksi sukupuolimuuttujien kohdalla ei voida ennakkoon esittää hypoteesia siitä, kumpi sukupuoli on tutkittavassa ilmiössä ”parempi” tai ”huonompi”. Toisessa tapauksessa lähdetään aikaisemman tiedon tai asiaa koskevien teorioiden perusteella yksisuuntaisesta vaihtoehtoisesta hypoteesista (one-tailed) eli meillä on jo ennakkoon tiedossa, kuinka tietyn muuttujan pitäisi vaikuttaa mitattavaan muuttujaan eli esimerkiksi, että tytöt ovat poikia parempia äidinkielessä. Tilasto-ohjelmissa on yleensä oletuksena kaksisuuntainen vaihtoehtoinen hypoteesi. (Yksisuuntaisen vaihtoehtoisen hypoteesin etu olisi se, että saatu p-arvo puolittuisi).

Esimerkkimatriisissa sukupuolimuuttuja on aineiston ainoa kaksiluokkainen muuttuja (tytöt ja pojat). Tutkija voisi nyt olla kiinnostunut esimerkiksi siitä, miten opintomenestys eroaa sukupuolen suhteen lukiossa. Seuraavassa esitellään tämä analyysi t-testin avulla (ks. taulukko 4):

TAULUKKO 4. Miesten ja naisten opintomenestys lukiossa.

Opintomenestys lukiossa	Miehet (n=30)	Naiset (n=30)	t-arvo	vapaus-asteet	p-arvo
Keskiarvo	15,13	16,53	-3,20	58	0,002
Keskihajonta	1,70	1,70			

Levenen testi: F = 0,04; p = 0,836

Aluksi voidaan todeta ryhmien varianssien olevan homogeeniset; t-testin käytön yksi kriteeri on, että ryhmien populaatiovarianssit ovat samankaltaiset. Tämän voidaan todentaa Levenen F-testin avulla. Levenen testituloksen voi liittää taulukoihin (ks. taulukko 4), joista lukija voi todeta muuttujien varianssien homogeenisuuskriteerin täyttymisen. Levenen testissä nollahypoteesina on jakaumien varianssien yhtäsuuruus. Kun testin p-arvo on 0,05 tai suurempi, variansseja voidaan pitää yhtäsuurina ja t-testin voi toteuttaa. Jos Levenen testin arvo on pienempi kuin 0,05, t-testi voidaan toteuttaa vaihtoehtoisella periaatteella (ns. Welchin testi), jonka tulos esimerkiksi SPSS-ohjelmassa saadaan automaattisesti t-testin tulosteen yhteyteen (t-testin tuloste-taulukossa alempi rivi, ks. alla olevat SPSS-ohjeet). Tarkasteltavat ryhmät ovat melko suuret (molempien n = 30), joka mahdollistaa t-testin käytön, vaikka ryhmien jakaumatietoja ei sen tarkemmin tarkasteltaisi (ns. keskeinen raja-arvolause). Tarvittaessa normaalijakaumaoletuksen täyttyminen voidaan tarkistaa vinous- ja huipukkuuslukujen avulla tai normaalisuuden testaukseen tarkoitetulla testillä (esim. Kolmogrov-Smirnovin testi, ks. tarkemmin tietolaari 3, normaalijakauma ja normaaliuden testaus, s. 105).

Kun tarkastellaan taulukossa 4 esitettyä tulosta, voidaan aluksi todeta miesten opintomenestyksen keskiarvon olevan lukiossa 15,1 ja naisten vastaavan arvon 16,5. Naisten opintomenestys näyttäisi siis keskiarvon perusteella olevan miesten menestystä parempi. Taulukosta nähdään lisäksi, että testin t-arvoksi tulee -3,20 ja p-arvoksi 0,002. (Miinusmerkki t-arvon edessä johtuu siitä, että t-arvoa laskettaessa jälkimmäisen ryhmän keskiarvo on ollut suurempi). Näin voidaan todeta, että sukupuolten välillä ilmenevä ero opintomenetyksessä on tilastollisesti merkitsevä (eli $p < 0,05$). Testin tulos raportoidaan yleensä seuraavassa muodossa: $t(58) = -3,20$; $p = 0,002$ (58 = vapausasteet eli degrees of freedom, df, ks. sivu 41). T-testissä vapausasteiden määrä on N-2, koska molemmissa ryhmissä aineistosta on ryhmävertailua varten laskettu ryhmäkeskiarvot. Tällöin vapaita eli riippumattomia havaintoja on aineistossa jäljellä enää N-2 kappaletta. Naisten opintomenestys on siis miehiä parempi lukiossa, mutta entä yliopistotasolla? Tätä voidaan tutkia tekemällä uuden t-testin yliopisto-opintomenestystä kuvaavalla muuttujalla. Lisäksi voidaan tutkia ryhmien sisäisen opintomenestyksen muutoksen tilastollista merkitsevyyttä siirryttäessä lukiosta yliopistoon toistettujen mittausten t-testillä, jota tarkastellaan seuraavassa alaluvussa.

SPSS: Riippumattomien ryhmien t-testi

1. Valitse ylhäältä valikkoriviltä *Analyze → Compare Means → Independent-Samples T Test*
   - siirrä valitsemasi numeerinen muuttuja (tai useita, jos haluat monta t-testiä samalla kertaa) muuttujaluettelon oikealla puolella olevalla ylemmällä nuolipainikkeella kohtaan *Test Variable(s)*
   - siirrä vastaavalla tavalla valitsemasi kategorinen muuttuja alemmalla nuolipainikkeella kohtaan *Grouping Variable* ja paina *Define Groups* -painiketta. Kirjoita kenttiin *Group 1* ja *Group 2* niiden kahden ryhmän aineistoon koodatut numeroarvot, joita haluat vertailla toisiinsa, hyväksy valinnat klikkaamalla *Continue*-painiketta

2. Toteuta t-testi painamalla *OK*-painiketta ja saat testiin liittyvät tulokset tulosteikkunaan

Huom: t-testin yhteyteen, samaan taulukkoon, tulostuu automaattisesti testi varianssien yhtäsuuruusolettamuksen voimassaololle (*Levene's test*). Jos olettamus ei pidä paikkaansa, testitaulukossa on mukana myös sellainen versio t-testistä, jossa varianssien yhtäsuuruusolettamusta ei tarvita (*Equal variances not assumed*).

Mikäli t-testin käyttöedellytykset eivät täyty, tulee t-testin sijaan käyttää sopivaa epäparametrista testiä. Riippumattomien ryhmien t-testin epäparametrinen vastine on Mann-Whitneyn U-testi, jota käsitellään luvussa 4.2.1 (s. 135). Kun vertailtavia luokkia on enemmän kuin kaksi tai kun vertailuasetelma on moniulotteisempi, on syytä käyttää varianssianalyysia ja monivertailutestejä. Monivertailutesteillä pystytään vähentämään virhemahdollisuuksia. Jos vertailtavia ryhmiä on paljon, olisi tehtävä paljon erillisiä t-testejä. Kun tehdään kyseiset testit esimerkiksi 5 prosentin erehtymisriskillä (p-arvolla 0,05), on oletettavissa, että 20 analyysin joukossa on keskimäärin yksi tilastollisesti merkitsevä ero, joka perustuu pelkästään sattumaan. Monivertailutestit pyrkivät vähentämään tätä virhemahdollisuutta, jolloin merkitsevyystaso kyetään pitämään vaaditulla tasolla.

4.1.1.1 T-testin tulkinnan vankistamista efektikoko- ja luottamusvälitarkastelun avulla

Edellä luvussa 1.6.1 on tarkasteltu yleisluontoisesti efektikoko- ja luottamusvälien käytön perusajatusta ja niiden perusteita tilastollisen päättelyn yhteydessä. Näiden käyttö perinteisen p-arvotulkinnan rinnalla on lisääntynyt viime vuosina merkittävästi. Efektikoko- ja luottamusvälitarkastelut tarjoavat siis perinteiselle p-arvo testitulostarkastelulle vaihtoehtoisia tai tätä täydentäviä menettelytapoja. Lisäksi tarkastelun lopuksi esitellään ryhmäerojen Bootstrap-estimoinnin soveltamista tämän asetelman suhteen. Vaikka tässä kirjassa nämä lisätarkastelut esitellään tarkemmin vain tässä yhteydessä, eli kahden ryhmän vertailuun liittyen, näitä analyysejä on mahdollista toteuttaa vastaavaan tapaan myös muiden kirjassa esitettyjen analyysimenetelmien kohdalla.

Efektikoko ja sen laskeminen ja arviointi kahden ryhmän vertailussa

Edellä käsitelty riippumattomien ryhmien t-testi antoi tulokseksi, että miesten ja naisten välillä havaittiin tilastollisesti merkitsevä ero lukion opintomenestyksessä. Edellä tuloksen tulkinta perustui ryhmien keskiarvojen tulkintaan ja t-testiin liittyvän p-arvon vertaamiseen yleisesti sovittuun raja-arvoon (yleensä merkitsevyystasona 0,05). Tulos siis kertoo vain sen, että miesten ja naisten keskiarvojen välillä on populaatiossa tilastollisesti merkitsevä ero. Tulos ei kuitenkaan kerro, miten suuri tai merkittävä tuo havaittu ero on. Tällaista eron suuruutta/voimakkuutta/kokoa voidaan arvioida tilanteeseen soveltuvaa efektikoko-mittaa käyttäen. Varsinkin laajoissa aineistoissa ongelmaksi voi muodostua se, että pienikin ero voi tilastollisessa testissä osoittautua erittäin merkitseväksi. Efektikoon laskeminen voi tällaisissa tapauksissa osoittaa, että testillä merkitseväksi havaittu ero itse asiassa onkin varsin mitätön (ks. luku 1.6.1.2).

Esimerkkiaineistossa miesten opintomenestys (ka = 15,1, kh = 1,7) oli t-testin perusteella tilastollisesti merkitsevästi matalampi kuin naisilla (ka = 16,5, kh = 1,7). Keskiarvojen perusteella ero ei kuitenkaan vaikuttaisi kovinkaan suurelta. Myös keskihajonnat ovat molemmissa ryhmissä melko pienet eli havaintoarvoissa hajontaa on melko vähän. Tältäkään osalta sukupuoliryhmien välillä ei siis ollut eroa. Voidaanko havaittua sukupuolten eroa opintomenestyksessä pitää suurena vai ei? Efektikoon arviointiin kehitetyt eri menetelmät antavat tutkijoille tämän päättelyyn oman apunsa. Kahden ryhmän vertailussa tällaista keskiarvojen ryhmäeroon liittyvää efektin kokoa (miten voimakas vaikutus eli efekti sukupuolella on opintomenestykseen) mitataan tavallisimmin *Cohenin d* -suureen avulla. Tätä efektikoon mittaa ei jostain syystä edelleenkään saada toteutettua suoraan SPSS:ssä, mutta se on kohtuullisen helppo laskea käsin tai sen saa tuotettua jollakin netistä löytyvällä laskurilla. Tällaisia voit helposti löytää netin hakukoneella vaikkapa hakusanalla 'effect size calculator'. *Cohenin d* saadaan laskettua myös laskimella käyttäen laskukaavaa d = (ka1 – ka2) / kh, mikäli ryhmien keskihajonnat ovat yhtäsuuret. Jos keskihajonnat ovat erisuuruiset, nimittäjässä oleva keskihajonta (kh) on muodostettava erillisen laskukaavan mukaan (ks. esimerkiksi Field 2018, 115). Esimerkissä efektikoon d arvoksi tulisi d = (16,5–15,1)/1,7 = 0,82. Seuraavassa tulosteessa erään nettilaskurin tuottama tulos:

	Group 1	Group 2
Mean	15.1	16.5
Standard Deviation	1.7	1.7
Effect Size d_{Cohen}	0.824	
Effect Size *Glass' Δ*	0.824	
Common Language Effect Size *CLES*	0.72	

N (Total number of observations in both groups)	30
Confidence Coefficient	95% ▾
Confidence Interval for d_{Cohen}	0.078 - 1.569

URL: https://www.psychometrica.de/effect_size.html

Miten saatua arvoa sitten voitaisiin tulkita? Lähtökohta on se, että mitä suurempi lukuarvo, sitä voimakkaampi efekti, eli suurempi ero keskiarvoissa. J. Cohen (1988) on esittänyt d-suureen tulkintaan suuntaa-antaviksi raja-arvoiksi 0,2 – 0,5 – 0,8 siten, että näissä rajakohdissa tekijän voimakkuus voitaisiin tulkita vastaavassa järjestyksessä pieni – keskisuuri – suuri (ks. asetelma 5, s. 49). Esimerkissä d = 0,82, joka siis ylittää rajakohdan 0,8 eli efektikoko olisi "suuri" tai "voimakas". Saatujen tulosten mukaan nyt voitaisiin raportoida, että miesten ja naisten ero oli tilastollisesti merkitsevää ja efektikooltaan suurta, t(58) = -3,20, p = 0,002, d = 0,82.

Luottamusvälit ja niiden käyttö ja tulkinta kahden ryhmän keskiarvojen vertailussa

Kuten edellä on todettu jo useaan kertaan, p-arvon tulkintaan perustuva päättely ei kerro myöskään mitään siitä, miten tarkka otokseen perustuva päättely on suhteessa populaatioon. Toisin sanoen pelkällä p-arvoon perustuvalla testillä ei saada tietoa siitä, miten tarkasti tämän pohjalta voidaan arvioida ryhmien keskiarvoja sekä keskiarvojen eroa. Luottamusvälitarkastelu tarjoaa tähän yhden merkittävän lisätarkastelunäkökulman (ks. tarkemmin luku 1.6.1.3). Luottamusvälin pituus antaa tietoa otokseen perustuvan arvioinnin eli estimoinnin tarkkuudesta. Yleensä tämä tarkastelu tehdään 95 %:n luotettavuustasolla. Tältä pohjalta määritetään arviointiin liittyvä virhemarginaali, joka lisätään ja vähennetään saadusta piste-estimaatista. Näin saadaan aikaan estimoinnin vaihteluväli, jonka käytännön tulkinta on kuvattu alla olevan esimerkin yhteydessä. Tässä käsiteltävässä esimerkissä luottamusvälit muodostetaan ryhmäkeskiarvoille, mutta aivan vastaavalla periaatteella luottamusvälejä saadaan toteutettua myös keskiarvojen erolle tai vaikkapa prosenttiosuuksille, keskihajonnalle, myöhemmin käsiteltäville korrelaatio- ja regressiokertoimille ja niin edelleen. Edellä esitettyyn t-testiasetelmaan lisätään nyt siis 95 %:n luottamusvälit ryhmäkeskiarvoille sekä myös ryhmäkeskiarvojen erolle, SPSS-tulosteen näyttävät nyt tällaisilta:

Opintomenestys lukiossa

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean Lower Bound	95% Confidence Interval for Mean Upper Bound
mies	30	15,13	1,70	,31	14,50	15,77
nainen	30	16,53	1,70	,31	15,90	17,17
Total	60	15,83	1,82	,24	15,36	16,30

		95% Confidence Interval of the Difference Lower	95% Confidence Interval of the Difference Upper
Opintomenestys lukiossa	Equal variances assumed	-2,28	-,52
	Equal variances not assumed	-2,28	-,52

Taulukoissa sarakkeilla *'95 % Confidence Interval'* nähdään tarkasteltavaan suureeseen liittyvän 95 %:n luottamusvälin ala- ja ylärajat sarakkeilla *lower* ja *upper*. Ylemmässä taulukossa ala- ja ylärajat liittyvät miesten ja naisten keskiarvojen arviointiin ja alemmassa taulukossa ryhmien välisen eron arviointiin. Esimerkiksi miehillä opintomenestys on keskimäärin tasolla 15,1 (miesten opintomenestyksen populaatiokeskiarvon piste-estimaatti). Koska muuttujan skaala oli 10–20, tämä kuvaa yleisen opintomenestyksen suhteen keskitasoista oppimistasoa. Keskiarvon 95 %:n luottamusväli on tulosteen tietojen mukaan [14,5, 15,8].

Luottamusvälin lukuarvot kuvaavat ensinnäkin sitä, että jos vastaavanlaista otokseen perustuvaa arviointia toteutettaisiin esimerkiksi 100 kertaa, näistä 95 kertaa todellinen populaation keskiarvo asettuisi välille 14,5–15,80 (ks. esim. van de Schoot ym. 2013). Toisin sanoen vain 5 %:ssa tapauksista arvo asettuisi kyseisen välin ulkopuolelle. Nyt siis saadaan yksittäistä pistearvoa täydentävä tieto siitä, miten tarkkana voidaan pitää saatua otoskeskiarvoa. Naisilla luottamusväli tulkittaisiin samaan tapaan. Lisäksi alemmassa taulukossa on vielä muodostettu 95 %:n luottamusväli ryhmien keskiarvoerolle, joka olisi taulukossa olevien ala- ja ylärajojen mukaan väli [-2.28, -0,52]. Etumerkit tässä ovat negatiiviset vain koska erotus on satuttu laskemaan ka(miehet) miinus ka(naiset). Väli voitaisiin yhtä hyvin raportoida vastaavin positiivisin lukuarvoin. Välin pituus kuvaa tarkkuutta, jolla ryhmäkeskiarvojen ero kyetään otoksesta laskemaan. Toinen tärkeä tulkinta tässä olisi, että koska luottamusväli ei sisällä arvoa 0 (nolla tarkoittaisi, että ryhmien välillä ei ole eroa, eli ryhmien keskiarvot ovat samat), ryhmien välillä on eroa 95 %:n luotettavuudella. Jälkimmäisessä tulkinnassa on suora analogia tilastolliseen testaamiseen. Saatujen tulosten mukaan nyt voitaisiin raportoida, että miehillä ja naisilla ero lukion opintomenestyksessä oli tilastollisesti merkitsevää ja efektikooltaan suurta, $t(58) = -3{,}20$, $p = 0{,}002$, eron 95 % CI: [0,52, 2,28], $d = 0{,}82$.

SPSS: Luottamusvälien muodostaminen kahden ryhmän keskiarvojen vertailussa

Vertailtaville osaryhmille saat muodostettua luottamusvälit esimerkiksi seuraavasti:

1. Valitse ylhäältä valikkoriviltä *Analyze → Descriptive Statistics → Explore*
   - siirrä tutkittava numeerinen muuttujasi muuttujaluettelon oikealla puolella olevalla nuolipainikkeella kohtaan *Dependent List*
   - siirrä ryhmittelevä muuttuja keskellä olevalla nuolipainikkeella kohtaan *Factor List*
2. *Statistics* -painikkeen takaa pääset valitsemaan, millä luotettavuudella luottamusväli muodostetaan. Oletuksena on yleisimmin käytetty arvo 95 %, mutta tässä pääset muuttamaan arvoa halutessasi (esim. 90 % tai 99 %). Hyväksy valinta klikkaamalla painiketta *Continue*
3. Toteuta osaryhmille muodostetun tunnuslukutaulukon laskeminen painamalla *OK*. Tunnuslukutaulukon yläosassa on nyt myös ryhmien keskiarvoihin liittyvien luottamusvälien ylä- ja alarajat.

ohjetaulukko jatkuu...

Huom. Yllä olevan esimerkin taulukko on toteutettu SPSS-proseduurilla *Analyze → Compare Means → One-Way ANOVA*, jossa keskiarvoihin liittyvät 95 %:n luottamusvälit saadaan painikkeella *Options*, jossa klikataan valinta *Descriptives*

Vertailtavien ryhmien keskiarvo-erolle saat muodostettua luottamusvälin t-testin toteutuksen yhteydessä:

1. Valitse ylhäältä valikkoriviltä *Analyze → Compare Means → Independent-Samples T Test*
   - siirrä valitsemasi numeerinen muuttuja (tai useita, jos haluat tarkastella montaa muuttujaa samalla kertaa) muuttujaluettelon oikealla puolella olevalla ylemmällä nuolipainikkeella kohtaan *Test Variable(s)*
   - siirrä vastaavalla tavalla valitsemasi kategorinen muuttuja alemmalla nuolipainikkeella kohtaan *Grouping Variable* ja paina *Define Groups* -painiketta. Kirjoita kenttiin *Group 1* ja *Group 2* niiden kahden ryhmän numeroarvot, joita haluat vertailla toisiinsa
   - hyväksy valinnat klikkaamalla *Continue*-painiketta
2. Varmista *Option*-painikkeen takaa kohdassa *Confidence Interval Percentage*, että tulosteeseen saadaan haluttu, esimerkiksi 95 %:n luottamusväli. Hyväksy valintasi klikkaamalla *Continue*-painiketta
3. Toteuta analyysi painamalla *OK*. Ryhmien keskiarvojen ero sekä eron luottamusväli tulostuvat t-testin taulukon oikean puoleiseen osaan.

Bootstrap-estimointi ja sen käyttö ja tulkinta kahden ryhmän keskiarvojen vertailussa

Edellä esitellyt luottamusväli-estimointi ja luottamusvälin pituuden määräytyminen perustuivat käytetyn estimaattorin, eli keskiarvon, tilastomatemaattisiin ominaisuuksiin. Tällöin otoksen poimintaa ei tarvitse lähteä toistamaan lukuisia kertoja, jotta nähtäisiin, miten paljon eri otosten tuottamat keskiarvot vaihtelevat otoksesta toiseen. Sen sijaan keskiarvo-estimaattorin ominaisuuksista on tiedossa, että kun ryhmäkoot ovat riittävän suuria (ryhmillä n on vähintään 20, mieluusti yli 30) ja havaintoarvot noudattavat normaalijakaumaa, luottamusväliin liittyvä virhemarginaali saadaan laskettua suoraan otoksen arvoista, käyttäen keskivirhettä (SPSS-tulosteessa oleva Standard Error, s.e.) ja tilanteeseen soveltuvaa todennäköisyysjakaumaa (keskiarvojen yhteydessä t-jakaumaa). Käytännössä hyvä ja monesti riittävän tarkka arvio 95 %:n luottamusväliin liittyväksi virhemarginaaliksi on 2*s.e., jolloin otoksesta lasketun keskiarvon luottamusvälin ala- ja ylärajat asettuvat lukuarvoihin ka ± 2*s.e. On kuitenkin huomattava, että edellä oleva tieto pätee vain, kun ryhmäkokoihin ja normaalijakautuneisuuteen liittyvät oletukset ovat kunnossa. Mikäli nämä oletukset eivät toteudu tai niiden toteutuminen epäilyttää, luottamusvälejä voidaan muodostaa myös Bootstrap-estimointia käyttäen (ks. enemmän luku 1.6.1.4). Alla olevissa taulukoissa on edellisiin esimerkkeihin liittyvät miesten ja naisten keskiarvoerojen bootstrap-estimoinnit:

Report

Opintomenestys lukiossa

Sukupuoli		Statistic	Bootstrap [a] Bias	Std. Error	95% Confidence Interval Lower	Upper
mies	Mean	15,133	,001	,290	**14,543**	**15,719**
	N	30				
	Std. Deviation	1,697				
nainen	Mean	16,533	,015	,319	**15,929**	**17,179**
	N	30				
	Std. Deviation	1,697				
Total	Mean	15,833	,006	,227	15,400	16,317
	N	60				
	Std. Deviation	1,824				

a. Unless otherwise noted, bootstrap results are based on 1000 bootstrap samples

Bootstrap for Independent Samples Test

		Mean Difference	Bootstrap [a] Bias	Std. Error	Sig. (2-tailed)	95% Confidence Interval Lower	Upper
omene1	Equal variances assumed	-1,40	,00	,43	,004	-2,30	-,59
	Equal variances not assumed	-1,40	,00	,43	,005	-2,30	-,59

a. Unless otherwise noted, bootstrap results are based on 1000 bootstrap samples

Ylemmässä taulukossa on osaryhmien keskiarvoille muodostetut 95 % Bootstrap-luottamusvälit. Esimerkiksi miehillä 95 % otoksissa keskiarvo on jotain väliltä [14,5, 15,7]. Naisilla keskiarvon 95 % luottamusväli olisi [15,9, 17,2]. Koska saadut luottamusvälit eivät asetu päällekkäin, voitaisiin päätellä, että miesten ja naisten keskiarvoissa on merkitsevää eroa. Nyt on huomattava, että tämä arviointi ei vaadi esimerkiksi oletuksia normaalijakaumasta tai arvojen vaihtelun yhtäsuuruudesta. Jos verrataan esimerkiksi miesten keskiarvoon liittyviä luottamusvälejä, aiemmin edellä perinteisellä tavalla laskettu luottamusväli on hyvin samanlainen kuin tässä muodostettu bootstrap-luottamusväli. Tämä kertoo siitä, että perinteisen tavan taustaoletukset ovat riittävän hyvin voimassa. Jälkimmäinen taulukko sisältää bootstrap-estimoinnin ryhmien keskiarvojen erolle. Siinäkin saatu 1000 uusintaotokseen perustuva luottamusväli [-2,30, -0,59] on hyvin samankaltainen kuin aiemmin muodostettu 95 %:n luottamusväli [-2,28, -0,52]. Mikäli perinteinen ja bootstrap-periaatteen luottamusväli eroavat selvästi toisistaan, on mahdollista, että perinteisen tarkastelun oletukset eivät toteudu. Tällöin olisikin luotettavampaa käyttää tuloksissa bootstrap-ratkaisua.

SPSS: Bootstrap-estimointi kahden ryhmän keskiarvojen vertailussa

Vertailtaville osaryhmille saat muodostettua bootstrap-luottamusvälit esimerkiksi seuraavasti:

1. Valitse ylhäältä valikkoriviltä *Analyze → Compare Means → Means*
   - siirrä tutkittava numeerinen muuttujasi muuttujaluettelon oikealla puolella olevalla nuolipainikkeella kohtaan *Dependent List*
   - siirrä ryhmittelevä muuttujasi keskellä olevalla nuolipainikkeella kohtaan *Layer 1 of 1*
2. *Bootstrap*-painikkeen takaa pääset muodostamaan luottamusvälit. Tyypillisen tulosteen saat seuraavasti:
   - valitse yläosan valinta *'Perform Bootstrapping'*, sekä varmistamalla, että sen alla olevassa ruudussa muodostettavien uusintaotosten määrä on vähintään 1000
   - valintaikkunan keskellä pääset antamaan luottamusväliin liittyvän luotettavuustason, yleisimmin arvona käytetään 95 %
   - hyväksy valinnat painamalla *Continue* -painiketta
3. Toteuta ryhmittäisten tunnuslukujen laskenta painamalla *OK* ja saat nyt tulosteeseen myös 95 %:n bootstrap-luottamusvälit (mm. ryhmien keskiarvoille)

Vertailtavien ryhmien keskiarvoerolle saat laskettua bootstrap-luottamusvälin t-testin toteutuksen yhteydessä:

1. Valitse ylhäältä valikkoriviltä *Analyze → Compare Means → Independent-Samples T Test*
   - siirrä valitsemasi numeerinen muuttuja (tai useita, jos haluat tarkastella montaa muuttujaa samalla kertaa) muuttujaluettelon oikealla puolella olevalla ylemmällä nuolipainikkeella kohtaan *Test Variable(s)*
   - siirrä vastaavalla tavalla valitsemasi kategorinen muuttuja alemmalla nuolipainikkeella kohtaan *Grouping Variable* ja paina *Define Groups* -painiketta. Kirjoita kenttiin *Group 1* ja *Group 2* niiden kahden ryhmän, joita haluat vertailla toisiinsa (aineistossa käytetyt numeroarvot)
   - hyväksy valinnat klikkaamalla *Continue*-painiketta
2. *Bootstrap*-painikkeen takaa pääset muodostamaan bootstrap-estimoinnin. Tyypillisen tulosteen saat seuraavasti:
   - valitse yläosan valinta *'Perform Bootstrapping'*, sekä varmistamalla, että sen alla olevassa ruudussa muodostettavien uusintaotosten määrä on vähintään 1000
   - valintaikkunan keskellä pääset antamaan luottamusväliin liittyvän luotettavuustason, yleisimmin arvona käytetään 95 %
   - hyväksy valinnat painamalla *Continue*-painiketta
3. Toteuta analyysi painamalla *OK*
   - Bootstrap-tulokset löydät tulosteen viimeisestä taulukosta *Bootstrap for Independent Sample*
   - tässä taulukossa mm. on oikeassa reunassa keskiarvoeron 95 % luottamusvälin ala- ja ylärajat

HUOM: Bootstrap-analyysiä on mahdollista toteuttaa aivan samaan tapaan myös monissa muissa SPSS:n menetelmävalikoissa. Toisin sanoen aina, jos käytettävän menetelmän valikkoruudussa löytyy Bootstrap-painike, perinteiseen analyysiin on lisättävissä myös bootstrap-tarkastelu yllä olevilla valinnoilla.

Edellä on esitelty samaan esimerkkiin liittyen useita eri menetelmiä, joita voidaan käyttää ja soveltaa kahden ryhmän keskiarvojen vertailuissa. Käytännössä ei välttämättä ole tarkoituksenmukaista, että tällaisia osin vaihtoehtoisia, osin toisiaan täydentäviä menetelmiä pyrittäisiin aina kaikkia käyttämään ja raportoimaan. Näiden käyttöä voit kuitenkin harkita tutkimuskysymystesi, aineistosi ja muuttujiesi ominaisuuksien mukaan. Useissa viimeaikaisissa tutkimusraportointiin liittyvissä ohjeistuksissa ja tieteellisten julkaisusarjojen kirjoittajaohjeissa kuitenkin monesti suositetaan, että saatua tulosta ei pitäisi esittää pelkästään yhden tulosta kuvaavan arvioinnin avulla (ks. esimerkiksi APA 2020; JARS website, https:/apastyle.apa.org/jars). Yleisohjeena voisi olla, että tunnuslukutietojen ja tilastollisen testituloksen lisäksi raportoinnissa olisi hyvä olla näkyvissä myös tutkittavaan tilanteeseen soveltuvat efektikokoa kuvaava suure ja 95 % luottamusväli. Näidenkin käytön yhteydessä tulee muistaa, että nämä ovat p-arvotestien lailla vain apuvälineitä tutkijan tulkintaa ajatelleen, viimekätinen tulkinnan tarkoituksenmukaisuuden pohtiminen jää aina tutkijalle ja laajemmin tutkijayhteisölle.

4.1.2 Toistettujen mittausten t-testi (parittainen t-testi)

Seuraavassa tarkastellaan toistettujen mittausten t-testiä, joka soveltuu esimerkiksi tilanteisiin, joissa jollekin ryhmälle (tai ryhmille) on tehty opetuskokeilun yhteydessä alkumittaus opetustuokion alussa ja loppumittaus sen jälkeen. Toistettujen mittausten t-testissä vertaillaan koeryhmän alkutilanteen keskiarvoa loppumittauksessa saatuun keskiarvoon ja arvioidaan keskiarvojen erotuksen tilastollista merkitsevyyttä. Siten voidaan päätellä, onko esimerkiksi opetuskokeiluun liittyvällä opetusmenetelmällä halutun suuntaisia vaikutuksia.

Esimerkkihavaintoaineistosta voidaan tutkia toistettujen mittausten t-testillä esimerkiksi sitä, muuttuuko opintomenestys miesten ja naisten ryhmän sisällä ja onko tapahtunut muutos tilastollisesti merkitsevä. Toteutetaan ensin erikseen sekä miesten että naisten kohdalla omat toistettujen mittausten t-testit. Näistä voidaan koota vaikkapa seuraavanlaisen taulukon:

TAULUKKO 5. Miesten ja naisten opintomenestyksen muutokset lukiosta yliopistoon.

Omene1 = Opintomenestys lukiossa Omene2 = Opintomenestys yliopistossa	**Miehet (n = 30)**	**Naiset (n = 30)**
keskiarvo (Omene_1)	15,13	16,53
keskiarvo (Omene_2)	15,93	16,07
keskihajonta (Omene_1)	1,70	1,70
keskihajonta (Omene_2)	2,42	2,69
keskiarvojen erotus	-0,80	0,47
keskiarvojen erotuksen keskihajonta	1,97	2,10
t-arvo	-2,22	1,22
vapausasteet (*df*)	29	29
p-arvo	0,034	0,233

Taulukosta voidaan nähdä, että miesten ryhmässä lukioaikainen opintomenestys (keskiarvo 15,1) on noussut arvoon 15,9. Muutos on tilastollisesti merkitsevä ($p < 0,05$), ja tulos voitaisiin raportoida muodossa $t(29) = -2,22; p = 0,034$. On kuitenkin tärkeää huomata, että muutos opintomenestyksessä on käytännössä suhteellisen vähäistä, kun sitä tarkastellaan suhteessa muuttujissa käytettyyn mitta-skaalaan (10–20). Naisten ryhmässä kehitys on ollut jonkin verran laskevaa. Muutos on naistenkin kohdalla vähäistä, mitä osoittaa myös analyysin tilastollisesti ei-merkitsevä testitulos ($p > 0,05$). Taulukosta nähdään myös sen, että vaikka miesten opintomenestys on parantunut ja naisten laskenut, siitä huolimatta naisten keskimääräinen opintomenestys yliopistossa on edelleen miehiä jonkin verran korkeampi. Taulukon tulosten perusteella ei kuitenkaan voida sanoa, onko muutos erilaista naisilla ja miehillä. Tämä edellyttäisi toistettujen mittausten varianssianalyysin käyttöä, jota käsitellään seuraavassa pääluvussa.

SPSS: Riippuvien ryhmien (toistettujen mittausten) t-testi

1 Valitse ylhäältä valikkoriviltä *Analyze → Compare Means → Paired-Samples T Test*
- siirrä yhtä aikaa kaksi valitsemaasi numeerista muuttujaa muuttujaluettelon oikealla puolella olevalla ylemmällä nuolipainikkeella kohtaan *Paired Variables*

2. Toteuta riippuvien ryhmien t-testi painamalla *OK*-painiketta, ja tulokset tulevat tulosteikkunaan

Jos haluaisit suorittaa yllä olevan t-testauksen jollekin kiinnostuksen kohteena olevalle osaryhmälle (esim. vain tytöt), silloin poimitaan analyysiin mukaan sukupuoli-muuttujaa käyttäen vain ryhmä sukupuoli = 2 (tytöt oli syötetty aineistoon koodilla 2)

Poiminnan tekisit seuraavasti:

1. Valitse ylhäältä valikkoriviltä vaihtoehdot *Data → Select Cases*
   - valitse esiin tuleva vaihtoehto *If condition is satisfied*, paina *If*-painiketta.
   - siirrä valitsemasi kategorinen muuttuja (tässä tapauksessa sukupuoli-muuttuja) muuttujaluettelon oikealla puolella olevalla nuolipainikkeella sen vieressä olevaan tyhjään tilaan. Paina tilan alapuolella olevan laskinnäppäimistön yhtäsuuruuspainiketta (=) ja numeroa kaksi (2), jolloin aineistosta ”suodatetaan” mukaan sukupuoli-muuttujan arvo kaksi eli tytöt
   - hyväksy valinnat klikkaamalla *Continue*-painiketta

2. *OK*-painikkeella suoritat poiminnan, ja sen merkiksi SPSS:n alapalkkiin tulee teksti *Filter On*. Lisäksi pois jääneet tapaukset näkyvät havaintoaineistossa yliviivatuin rivinumeroin. Muista poistaa poiminta, kun et sitä enää tarvitse. Poistamisen teet näin. Valitse ylhäältä valikkoriviltä vaihtoehdot *Data → Select Cases, v*alitse kohdan *Select* vaihtoehto *All Cases* ja paina *OK*-painiketta.

ohjetaulukko jatkuu...

Mikäli haluaisit suorittaa yllä olevan t-testauksen kaikille kiinnostuksen kohteena oleville eri osaryhmille kerralla (esim. erikseen tytöille ja pojille), käytä ennen t-testiä toimintoa, jolla aineisto jaetaan erillisiin osaryhmiin:

1. Valitse ylhäältä valikkoriviltä vaihtoehdot *Data → Split File*
   - valitse esiin tuleva vaihtoehto *Organize output by groups*
   - siirrä valitsemasi kategorinen muuttuja (tässä tapauksessa sukupuoli-muuttuja) muuttujaluettelon oikealla puolella olevalla nuolipainikkeella sen vieressä olevaan tyhjään tilaan *Group based on*
2. *OK*-painikkeella suoritat aineiston jaon osaryhmiin, ja sen merkiksi SPSS:n alapalkkiin tulee teksti *Split File On*. Muista poistaa aineiston jako, kun et enää sitä tarvitse. Poistamisen teet näin. Valitse ylhäältä valikkoriviltä vaihtoehdot *Data → Split File, v*alitse kohdan *Select* vaihtoehto *Analyze all cases, do not create groups* ja paina *OK*-painiketta. Havaintomatriisin alareunan teksti *Split File On* poistuu.

T-testin käytölle edellytyksenä on, että alkumittauksen, loppumittauksen ja muuttujien välisten erotusten arvot ovat jakautuneet normaalijakauman mukaisesti. Normaalisuusoletusta selvitetään edellisen esimerkin tavoin visuaalisesti (histogrammit), tunnusluvuin (vinous ja huipukkuus) ja tilastollisen normaalisuustestin avulla. Otoskoon myös tulisi olla riittävän suuri. Mikäli t-testin kriteerit eivät täyty, tämän testin tilalla voidaan käyttää epäparametrista vastinetta, Wilcoxonin testiä (ks. tarkemmin kriteerit ja epäparametrinen testi luku 4.2.2). Tässäkin pätee se, jota edellä on jo useasti korostettu: tulosten tulkintaa ei pitäisi perustaa ainoastaan testin tuottaman p-arvon varaan, vaan testituloksen rinnalle kannattaisi toteuttaa ja raportoida mahdollisuuksien mukaan myös luottamusväli- ja efektikoon tarkastelua (ks. luku 1.6.1.4). Näitä ei käsitellä tässä yhteydessä, mutta kiinnostunut lukija löytää varsin helposti kirjallisuudesta tai verkosta etsien ohjeita tähän soveltuvista lisäanalyyseistä.

4.2 Kahden ryhmän erojen vertaileminen epäparametrisilla testeillä

Mikäli käyttöedellytykset eivät ole voimassa t-testiä ajatellen (esim. muuttuja ei ole jakautunut normaalijakauman tapaan), voidaan käyttää testin epäparametrista vastinetta. Riippumattomien ryhmien vertailussa voidaan käyttää Mann-Whitneyn U-testiä ja toistettujen mittausten tilanteessa Wilcoxonin testiä (ks. yksityiskohtaisempi esitys epäparametrisista menetelmistä esimerkiksi Metsämuuronen 2009, luku 9). Epäparametrisiin testeihin tutustuminen on siinäkin mielessä tarkoituksenmukaista, että keskiarvotestit edellyttävät minimissään välimatka-asteikon tasoisia mittauksia. Likert-asteikon muuttujat ovat välimaastossa, joten niiden kohdalla kannattaa kokeilla myös epäparametristen testien käyttöä ja katsoa, tapahtuuko p-arvoissa parametrisiin keskiarvotesteihin verrattuna suuria muutoksia. Jos ero on tulkinnan kannalta ratkaiseva, silloin on varminta tehdä p-arvoon perustuvat päätelmät epäparametrisen testin antaman p-arvon avulla. Lisäksi voidaan ja kannattaakin soveltaa ja raportoida p-arvon tulkinnan rinnalle myös täydentäviä analyysejä, kuten luottamusvälien ja efektikokosuureiden tulkintaa (ks. esimerkit t-testin yhteydessä, s. 122–131).

4.2.1 Mann-Whitneyn U-testi

Mann-Whitneyn U-testi soveltuu kahden ryhmän välisten erojen vertailuun. U-testissä edellytetään vertailtavien ryhmien olevan riippumattomia toisistaan (ei siis sovellu toistettujen mittausten tilanteeseen) ja sitä, että muuttujien mittaustaso on vähintään järjestysasteikollinen. Vertailtavissa ryhmissä jakaumien ei tarvitse olla normaalijakauman mukaisia, mutta jakaumien muodon olisi molemmilla ryhmillä kuitenkin hyvä olla samankaltainen. (Ks. tarkempi kuvaus testin periaatteista esimerkiksi Metsämuuronen 2009, 1102.)

Edellä olleesta tutkimusasetelmasta, jossa verrattiin naisten sekä miesten opintomenestystä lukiossa, SPSS antaa seuraavanlaiset tulosteet:

Ranks

	SUKUPUOL	N	Mean Rank	Sum of Ranks
omene1	mies	30	23,97	719,00
	nainen	30	37,03	1111,00
	Total	60		

Test Statistics[a]

	omene1
Mann-Whitney U	254,000
Wilcoxon W	719,000
Z	-2,937
Asymp. Sig. (2-tailed)	,003

a. Grouping Variable: SUKUPUOL

Mann-Whitneyn U-testi järjestää muuttujan arvot suuruusjärjestykseen ja laskee näin saadut järjestysluvut sukupuolittain yhteen (Sum of Ranks) ja jakaa saadun summan kummankin ryhmän koolla, jolloin saadaan järjestyslukujen keskiarvot (Mean Rank) ryhmittäin. Miesten keskimääräinen sijaluku on 24 ja naisten 37. Naiset ovat siis kohtalaisen selkeästi opintomenestysmuuttujan keskikohdan paremmalla puolella ja miehet vastaavasti heikommalla. (Keskikohdan voidaan päätellä olevan 30,5 eli järjestyslukujen 1 ja 60 keskiarvon). Testiosa (Test Statistics) osoittaa sukupuolten välisen eron olevan tilastollisesti merkitsevä, sillä testiin liittyvä p-arvo on 0,003. Eron voimakkuus kannattaa laskea sopivaa efektikoon suuretta käyttäen. U-testin yhteydessä tällainen on z-testisuureen avulla muodostettava r, joka saadaan laskemalla r = z/neliöjuuri(N). Näin laskien r:n arvoksi saataisiin r = 0,38, jolloin efektikoko voidaan tulkita keskivahvaksi (ks. enemmän esim. Field 2018, 295). Kokonaisuudessaan tilastolliseen päättelyyn liittyvät tulokset raportoitaisiin seuraavasti: U = 254, z = -2,94; p = 0,003; r = 0,38.

U-testistä saadaan siis aivan saman tulkinnan kuin t-testilläkin. Mann-Whitneyn U-testin p-arvo on hieman suurempi kuin t-testissä. Tällä ei ole kuitenkaan tulkinnan kannalta merkitystä. Molemmat arvot sijoittuvat välille 0,01–0,001, jolloin pysytään loppujen lopuksi samalla riskitasolla. Siten t-testin käyttö ei tässä tapauksessa tuottanut harhaista tulosta ja siihen liittyvä keskiarvojen käyttö tulkinnassa ja raportoinnissa on helpompaa kuin edellä esitettyjen keskimääräisten järjestyslukujen tai ainakin ihmiset mieltävät yleisesti helpommin sen, kuinka merkittävästä erosta vertailtavien ryhmien välillä on kysymys. Tutkimusten tulosten tulkinnan kannalta tämä ero on vähintään yhtä tärkeä tulkita kuin P-arvoon perustuva tulkinta (ks. luku 1.6.1). Silti mikään ei estä tutkijaa käyttämästä tässäkin tapauksessa epäparametrista analyysivaihtoehtoa.

SPSS: Mann-Whitneyn U-testi

1. Valitse ylhäältä valikkoriviltä *Analyze → Nonparametric Tests → Legacy Dialogs → 2 Independent Samples*
   - siirrä valitsemasi numeerinen muuttuja (tai useita, jos haluat monta U-testiä samalla kertaa) muuttujaluettelon oikealla puolella olevalla ylemmällä nuolipainikkeella kohtaan *Test Variable List.*
   - siirrä vastaavalla tavalla valitsemasi kategorinen muuttuja alemmalla nuolipainikkeella kohtaan *Grouping Variable*
   - tarkista, että kohdassa *Test Type* on valittuna vaihtoehto *Mann-Whitney U* ja paina *Define Groups*-painiketta. Kirjoita avautuviin kenttiin *Group 1* ja *Group 2* niiden kahden ryhmän numeroarvot, joita haluat vertailla testillä toisiinsa
   - hyväksy valinnat klikkaamalla *Continue*-painiketta
2. Toteuta Mann-Whitneyn U-testi painamalla *OK*-painiketta ja saat testitulokset tulosteikkunaan

HUOM. Mann-Whitneyn U-testin voi toteuttaa myös valikossa *Analyze → Nonparametric Tests → Independent Samples*. Tällä toiminnolla muuttujien ja testin valinta tehdään avautuvilta välilehdiltä. Välilehdellä *Fields* pääset valitsemaan numeerisen muuttujasi siirtämällä sen keskellä olevalla nuolella oikealla olevaan tilaan *Test Fields*. Vastaavasti siirrät kategorisen muuttujasi alemmalla nuolella kohtaan Groups. Toiminto toteuttaa muuttujatietojen perusteella automaattisesti oikean testin, kun painat alareunan *Run*-painiketta. Voit myös itse käydä valitsemassa Mann-Whitneyn U-testin ja muita lisävalintoja siirtymällä välilehdelle *Settings,* jossa voit valita vaihtoehdon *Customize tests* ja sen jälkeen testin klikkaamalla kohdan *Mann Whitney U (2 samples)*. Alareunan *Run* -painikkeella toteutat valitun testin. Huomaa, että testin toteutus edellyttää, että aineiston muuttujien mitta-asteikot on määritelty oikein havaintomatriisin alareunan *Variable View* -näkymässä (*Measure: Nominal/ Ordinal/ Scale*). Huomaa myös, että perinteiseen toteutustapaan verrattuna tässä toiminnossa on mahdollista saada erilaisia lisätulosteita, mm. ryhmämediaanien eron luottamusväli ja osaryhmien jakaumakuviot.

4.2.2 Wilcoxonin testi

Toistettujen mittausten t-testin epäparametrinen vastine on Wilcoxonin-testi. Tällä testillä voidaan selvittää kahden riippuvan muuttujan jakaumien välistä eroa (ks. tarkemmin esimerkiksi Field 2018, 297–305). Muuttujien pitää olla riippuvia, eli samat henkilöt ovat vastanneet kahteen eri kysymykseen. Seuraavassa analysoidaan sitä, pysyykö lukion ja yliopisto-opiskelun opintomenestys keskimäärin ennallaan vai onko siinä havaittavissa muutosta. Tarkastelu kannattaisi toki aloittaa koko aineistoa koskien, mutta tässä muutosta on tutkittu erikseen naisten ja miesten ryhmissä. SPSS-ohjelmasta saamme seuraavanlaisen tulosteen (tulosteeseen on otettu malliksi vain miehiä koskeva tuloste, n = 30):

Ranks

		N	Mean Rank	Sum of Ranks
Opintomenestys lukiossa - Opintomenestys yliopistossa	Negative Ranks	18[a]	16,39	295,00
	Positive Ranks	10[b]	11,10	111,00
	Ties	2[c]		
	Total	30		

a. Opintomenestys lukiossa < Opintomenestys yliopistossa
b. Opintomenestys lukiossa > Opintomenestys yliopistossa
c. Opintomenestys lukiossa = Opintomenestys yliopistossa

Test Statistics[b]

	Opintomenestys lukiossa - Opintomenestys yliopistossa
Z	-2,133[a]
Asymp. Sig. (2-tailed)	,033

a. Based on positive ranks.
b. Wilcoxon Signed Ranks Test

Wilcoxonin testi laskee ensin jokaisen havaintoyksikön kohdalla tarkastelun kohteena olevan muuttujaparin absoluuttisen erotuksen (erotuksen etumerkkiä ei oteta huomioon). Jos erotus on nolla eli kyseessä on niin sanottu ”sidos” (*Tie*), näitä tapauksia ei oteta mukaan. Sen jälkeen nämä absoluuttiset erotukset järjestetään suuruusjärjestykseen unohtamatta kuitenkaan sitä, oliko erotus alun perin positiivinen vai negatiivinen. Kun erotuksen alkuperä on tiedossa, saadut järjestysluvut voidaan laskea yhteen erotustyypeittäin (*Sum of Ranks*). Kun summat jaetaan kummankin tyyppisten erotusten lukumäärillä, saadaan järjestyslukujen keskiarvot (*Mean Rank*) erotustyypeittäin. Tämän jälkeen lasketaan testisuureen arvo ja siihen liittyvä p-arvo. Testiosa (*Test Statistics*) osoittaa, että miesten ryhmässä opintomenestyksen muutos on tilastollisesti merkitsevä ($p < 0{,}05$). Tämänkin testin yhteydessä on hyvä laskea lisäksi tilanteeseen soveltuva efektikokoa mittaava suure. U-testin tapaan tällainen on suure r, joka saadaan nyt lausekkeesta r = z/neliöjuuri(2*N). Tässä arvoksi tulisi r = 0,28, jonka mukaan havaittu muutos on voimakkuudeltaan vähäistä (ks. tulkinta Field 2018, 303–304). Kokonaisuudessaan tilastolliseen päättelyyn liittyvät tulokset raportoitaisiin seuraavasti. Testin tulos voidaan raportoida tulosteen Z-testisuuretta käyttäen: Z = -2,13; p = 0,033; r = 0,28.

Kuten edelläkin, analyysi olisi voitu suorittaa myös toistettujen mittausten t-testiä käyttäen, mikäli testille asetetut kriteerit täyttyvät. Tässäkin kohden t-testin etuna on se, että analyysi perustuu kaikille tuttuun keskiarvoon, joten lukion ja yliopisto-opiskelun opintomenestyksen todellinen tasoero on helpommin tulkittavissa. Wilcoxonin testistä tämä tulkinta on jo paljon mutkikkaampi esittää, eikä tätä siinä muodossa voikaan tulkita kuin keskiarvoeroja suoraan vertailemalla. Toki analyysin tulosteesta voidaan todeta, kuinka monella otokseen kuuluvista opintomenestys on laskenut tai pysynyt samalla tasolla ja kuinka monella opintomenestys on parantunut (tätähän ei suoraan taas t-testitulosteesta saa). Wilxoconin testin etu on kuitenkin se, että se soveltuu jo järjestysasteikollisille muuttujille, eikä sen käyttö edellytä parametrisille testeille tyypillisiä jakaumiin liittyviä oletuksia. Testiä siis kannattaa käyttää, jos keskiarvotestin kriteerit eivät selkeästi täyty.

Yleensäkin voidaan todeta, että parametrisia testejä kannattaa käyttää aina, kun se on mahdollista niiden helpon tulkittavuuden ja paremman voimakkuuden vuoksi (voimakkaampi testi kykenee havaitsemaan populaatiossa esiintyvän efektin jo pienemmästä otoksesta). Silloin, jos esimerkiksi t-testin käyttöedellytykset eivät täyty muuttujissa, asiaa pitää tietysti harkita, ja vähintäänkin tarkistaa näissä tapauksissa t-testin tulos sitä vastaavalla epäparametrisella testillä.

SPSS: Wilcoxonin testi

1. Valitse ylhäältä valikkoriviltä *Analyze → Nonparametric Tests → Legacy Dialogs → 2 Related Samples*
   - siirrä yhtä aikaa (voit käyttää Ctrl-painiketta) kaksi valitsemaasi numeerista muuttujaa muuttujaluettelon oikealla puolella olevalla ylemmällä nuolipainikkeella kohtaan *Test Pairs*
   - ruksita kohtaan *Test Type* valinnaksi *Wilcoxon*
2. Toteuta Wilcoxonin testi *OK*-painikkeella ja saat testitulokset tunnuslukuineen tulosteikkunaan

Jos haluaisit suorittaa yllä olevan Wilcoxonin testin jollekin kiinnostuksen kohteena olevalle osaryhmälle (esim. vain tytöt), silloin ennen yllä olevaa analyysia mukaan poimitaan sukupuoli-muuttujaa käyttäen vain ryhmä sukupuoli = 2 (tytöt oli syötetty aineistoon koodilla 2).

ohjetaulukko jatkuu...

Poiminnan tekisit seuraavasti:

1. Valitse ylhäältä valikkoriviltä vaihtoehdot *Data → Select Cases*
   - valitse esiin tuleva vaihtoehto *If condition is satisfied*, ja paina *If*-painiketta
   - siirrä valitsemasi kategorinen muuttuja (tässä tapauksessa sukupuolimuuttuja) muuttujaluettelon oikealla puolella olevalla nuolipainikkeella sen vieressä olevaan tyhjään tilaan. Paina tilan alapuolella olevan laskinnäppäimistön yhtäsuuruuspainiketta (=) ja numeroa kaksi (2), jolloin aineistosta ”suodatetaan” mukaan sukupuolimuuttujan arvot kaksi eli tytöt
   - hyväksy valinnat klikkaamalla *Continue*-painiketta
2. *OK*-painikkeella suoritat poiminnan, ja sen merkiksi SPSS:n alapalkkiin tulee teksti *Filter On*. Lisäksi pois jääneet tapaukset näkyvät yliviivatuin rivinumeroin havaintoaineistossa. Muista poistaa poiminta, kun et sitä enää tarvitse. Poistamisen teet näin. Valitse ylhäältä valikkoriviltä vaihtoehdot *Data → Select Cases, v*alitse kohdan *Select* vaihtoehto *All Cases* ja paina *OK*-painiketta.

Mikäli haluaisit suorittaa ylimpänä olevan t-testauksen kaikille kiinnostuksen kohteena oleville eri osaryhmille kerralla (esim. erikseen tytöille ja pojille), käytä ennen yllä olevaa Wilcoxonin testiä toimintoa, jolla aineisto jaetaan erillisiin osaryhmiin:

1. Valitse ylhäältä valikkoriviltä vaihtoehdot *Data → Split File*
   - valitse esiin tuleva vaihtoehto *Organize output by groups*
   - siirrä valitsemasi kategorinen muuttuja (tässä tapauksessa sukupuoli-muuttuja) muuttujaluettelon oikealla puolella olevalla nuolipainikkeella sen vieressä olevaan tyhjään tilaan *Group based on*
2. *OK*-painikkeella suoritat aineiston jaon osaryhmiin, ja sen merkiksi SPSS:n alapalkkiin tulee teksti *Split File On*. Muista poistaa aineiston jako, kun et enää sitä tarvitse. Poistamisen teet näin. Valitse ylhäältä valikkoriviltä vaihtoehdot *Data → Split File, v*alitse kohdan *Select* vaihtoehto *Analyze all cases, do not create groups* ja paina *OK*-painiketta (alareunan teksti *Split File On* poistuu)

Huom. Wilkoxonin testin voi toteuttaa myös valikossa *Analyze → Nonparametric Tests → Related Samples*. Tällä toiminnolla muuttujien ja testin valinta tehdään avautuvilta välilehdiltä. Välilehdellä *Fields* pääset valitsemaan kaksi numeerista muuttujaasi siirtämällä ne keskellä olevalla nuolella oikealla olevaan tilaan *Test Fields*. Välilehdellä *Settings* pääset valitsemaan testiksi Wilcoxonin testin klikkaamalla oikealla puoliskolla olevan kohdan *Wilcoxon matched-pair signed rank (two samples)*. Alareunan *Run* -painikkeella toteutat valitun testin. Huomaa, että testin toteutus edellyttää, että aineiston muuttujien mitta-asteikot on määritelty oikein havaintomatriisin *Variable View* -näkymässä (*Measure: Nominal/ Ordinal/ Scale*). Huomaat myös, että testitulostuksen muoto on hieman erilainen kuin perinteisellä tavalla toteutetussa testissä, josta oli esimerkki edellisellä sivulla.

Seuraavaksi siirrytään tarkastelemaan varianssianalyysia, joka mahdollistaa t-testejä monipuolisemmat tutkimusasetelmat (voidaan verrata esimerkiksi kolmen tai useamman ryhmän keskiarvoeroja keskenään tai käyttää samassa analyysissä kahta tai useampaa kategorista muuttujaa).

5. Usean ryhmän tai mittauskerran keskiarvojen vertaaminen

Usean ryhmän ja/tai mittauskerran keskiarvojen vertaamiseen voidaan käyttää varianssianalyysiä eli ANOVAa (analysis of variance, ANOVA). Varianssianalyysi soveltuu tutkimuskysymyksiin, joissa ollaan kiinnostuneita siitä, miten eri kategoriset muuttujat (tekijät) vaikuttavat analysoitavaan määrälliseen muuttujaan (vaste) eli miten ne selittävät vastemuuttujan vaihtelua. Käytännössä on siis kyse eri ryhmien ryhmäkeskiarvojen vertailusta. Yksisuuntainen varianssianalyysi soveltuu tilanteisiin, joissa vertailtavat ryhmät muodostuvat yhden kategorisen tekijän tasojen mukaisesti. Useampisuuntaisessa varianssianalyysissä mukana on useita ryhmitteleviä tekijöitä, joiden erilaisten tasokombinaatioiden keskiarvoja vertaillaan keskenään. Erityisesti kaksisuuntainen varianssianalyysi on varsin käyttökelpoinen analyysi esimerkiksi monissa kasvatustieteellisissä tutkimusasetelmissa. Esimerkiksi kaksisuuntaisessa analyysissä olisi tutkittavan määrällisen muuttujan lisäksi mukana kaksi kategorista muuttujaa eli tekijää, joissa molemmissa voisi olla vaikkapa kolme tasoa eli ryhmävaihtoehtoa. Tällainen kaksisuuntainen asetelma voitaisiin esittää muodossa 3x3 ANOVA. Useampisuuntainen varianssianalyysi on yksi monimuuttujamenetelmistä, jotka antavat meille mahdollisuuden analysoida aineistoa moniulotteisemmin kuin useat erillisinä analysoidut kahden muuttujan asetelmat. Monimuuttujamenetelmillä pyritään saamaan aineistosta esille tieteellisesti mielenkiintoisia selityksiä tutkimuksen aineistossa olevien muuttujien esiintyvien systemaattisten erojen pohjalta (Lehtonen & Pahkinen 2004, 257).

Useampisuuntaisella analyysillä voidaan siis tutkia useiden tekijöiden samanaikaista vaikutusta tutkittavaan ilmiöön. Voidaan esimerkiksi analysoida tietyn opetusmenetelmän ja eri opetustyylien vaikutuksia oppimissaavutuksiin tai analysoida esimerkiksi sukupuolen, sosiaaliluokan tai kotiympäristön vaikutuksia tutkittavaan ilmiöön. Tulevissa esimerkeissä ollaan kiinnostuneita ensinnäkin siitä, kuinka kolmiluokkaisena mitattu oppilaiden viriketausta on yhteydessä oppilaiden opintomenestykseen lukiossa (yksisuuntainen varianssianalyysi). Tämän jälkeen lisätään asetelmaan vielä toiseksi kategoriseksi tekijäksi sukupuolen (kaksisuuntainen varianssianalyysi). Näin voidaan tutkia ja tulkita opintomenestykseen vaikuttavia tekijöitä monipuolisemmin, esimerkiksi sitä, vaikuttaako erilainen viriketaustan laatu eri tavalla vaikkapa poikiin kuin tyttöihin. Luvun lopuksi esitetään vielä esimerkki toistettujen mittausten varianssianalyysista vastaavaan tapaan kuin tehtiin aiemmin t-testinkin kohdalla.

Tiiviisti varianssianalyysin perusteista:

- verrataan numeerisen muuttujan kahta tai useampaa ryhmäkeskiarvoa
- tekijöitä voi olla useita; jos on vain yksi tekijä ja yksi mittauskerta, kyseessä on yksisuuntainen varianssianalyysi; kaksisuuntaisessa varianssianalyysissa on kaksi tekijää tai yksi tekijä ja useampi mittauskerta jne.
- numeerista muuttujaa, jonka keskiarvoista ollaan kiinnostuneita, sanotaan vasteeksi[4] (tarkoittaa tässä yhteydessä selitettävää muuttujaa)
- muuttujaa, jonka eri luokissa numeerisen muuttujan keskiarvoja tutkitaan, sanotaan tekijäksi, ja tekijän arvoja sanotaan tasoiksi (tekijä tarkoittaa lähinnä selittävää muuttujaa)
- testaus perustuu F-testiin, kirjain F viittaa englantilaiseen tilastotieteilijään ja geneetikkoon R.A. Fisheriin (1890–1962), joka kehitti voimakkaasti kokeellista tutkimusta ja sitä kautta varianssianalyysia
- useamman kuin yhden tekijän varianssianalyysissä keskeisenä tuloksena saadaan analysoitavien tekijöiden yhdysvaikutukset (ks. s. 144) yksittäisten tekijöiden päävaikutusten lisäksi
- nollahypoteesina on ryhmäkeskiarvojen yhtäsuuruus
- varianssianalyysin tuloksista voidaan päätellä, ovatko eri ryhmien keskiarvoissa havaitut erot tilastollisesti merkitseviä

---

4 Vasteella viitataan tilastotieteellisessä kirjallisuudessa muuttujaan, jonka arvot selviävät vasta havaintojen teon kautta. Tekijäksi taas kutsutaan muuttujaa, jonka arvot ovat tiedossa ennen havaintojen tekoa. Katso enemmän vaste-tekijä -käsitteistä esim. Helenius (1989, 22–24). Kasvatustieteellisessä ja yhteiskuntatieteellisessä kirjallisuudessa vakiintuneet selitettävä-selittävä tai riippuva-riippumaton -käsiteparit saattavat johdattaa meidät ajattelemaan liian suoraviivaisesti analyysien tuloksia kausaaliluontoisiksi, joita tilastollisen analyysin tulokset harvoin ovat.

Käytön edellytyksiä:

- vasteen on oltava vähintään välimatka-asteikollinen, ja sen jakauman pitäisi olla tekijän tai tekijöiden eri luokissa normaalijakauman mukainen
- vasteen populaatiovarianssien (populaatiokeskihajontojen) tulisi olla yhtä suuret tekijän tai tekijöiden eri luokissa
- tekijöiden välillä ei saa olla liian voimakasta yhteyttä tai eivät saa edustaa loogisesti samaa asiaa
- tekijöiden asteikkotason on oltava kategorinen

5.1 Yleistä varianssianalyysistä

Varianssianalyysiä voidaan siis käyttää analysoitaessa yhden tai useamman tekijän samanaikaista vaikutusta vasteeseen. Kullakin tekijällä on useita eri tasoja eli eri luokkia/ryhmiä kuvaavia arvoja. Esimerkkiaineistossa tekijöitä voisivat olla sukupuoli, jolla on kaksi tasoa (miehet/naiset; nykyään myös kolmantena luokkana esimerkiksi muunsukupuolinen), vanhempien peruskoulutus, jolla on kolme tasoa (kansakoulu/peruskoulu/lukio) ja viriketausta, jolla myös on kolme tasoa (matala/keskinkertainen/korkea). Vaikka tekijöiden mittaustasoille ei varianssianalyysissa aseteta ehtoja, on selvää, että liian monitasoiset tekijät aiheuttavat hankaluuksia, mikäli aineisto on pieni. Käytännössä tekijät ovat lähes aina kategorisia muuttujia. Vastemuuttujan tulee olla numeerinen ja jakaumaltaan sellainen, että keskiarvo sopii kuvaamaan sen jakaumaa (normaalijakaumavaade). Varianssianalyysi kuuluu yleisen lineaarisen mallin analyysimenetelmiin samoin kuin kovarianssi- ja regressioanalyysitkin. Tekijöiden mitta-asteikoista riippuen, se on lähellä muita saman ”analyysiperheen” menetelmiä. Kategorisilla tekijöillä kyse on siis varianssianalyysistä. Jos taas joku tai jotkut tekijöistä ovat mitta-asteikoltaan määrällisiä, niin analyysi toteutettaisiin kovarianssianalyysina. Jos taas kaikki selittävät tekijät ovat määrällisiä, niin analyysissa käytettäisiin regressioanalyysia (ks. luku 7).

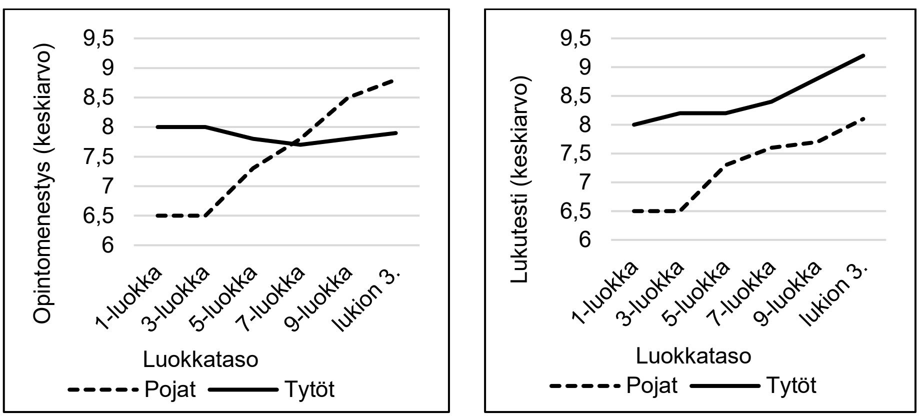
Varianssianalyysiä voidaan pitää t-testin laajennuksena kahdessakin mielessä: ensinnäkin varianssianalyysilla voidaan tutkia usean tekijän samanaikaista vaikutusta yhteen muuttujaan, vasteeseen, ja toiseksi kullakin tekijällä voi olla enemmän kuin kaksi tasoa. T-testihän sopii ryhmäkeskiarvojen vertaamiseen ainoastaan siinä tapauksessa, että verrattavia ryhmiä on vain kaksi. Varianssianalyysi tuo lisää mahdollisuuksia myös toistettujen mittausten tutkimusaineistojen analyysiin. Menetelmä soveltuu myös sen tarkasteluun, onko kehitys erilaista eri tekijän tasoilla (eri ryhmissä).

Varianssianalyysi, kuten muutkin monimuuttujamenetelmät, mahdollistaa siis usean samanaikaisen tekijän vaikutuksen mittaamisen. Näin päästään lähemmäksi todellisuutta; onhan niin, että todellisuudessa esimerkiksi yleistä opiskelumenestystä ei voida redusoida yhteen taustatekijään, johon esimerkiksi tulomomenttikorrelaatiossa tai t-testissä joudutaan tyytymään. Mikäli varianssianalyysin tulos on se, että jonkin tekijän vaikutus vasteeseen on tilastollisesti merkitsevä (eli vasteen keskiarvot poikkeavat toisistaan tilastollisesti merkitsevästi tämän tekijän eri tasoilla), ei kyyetä tarkalleen sanomaan, mitkä kyseisen tekijän tasoerot aiheuttavat tuon tilastollisen merkitsevyyden. Jos saataisiin esimerkiksi aineistosta sellaisen tuloksen, että viriketausta vaikuttaa tilastollisesti merkitsevästi opintomenestykseen (eli että opintomenestysmuuttujan keskiarvot poikkeavat toisistaan tilastollisesti merkitsevästi eri viriketaustaluokissa matala/keskinkertainen/korkea), ei vielä tämän perusteella voisi sanoa, mitkä eri luokkien väliset keskiarvoerot aiheuttavat tuon tilastollisen merkitsevyyden. Jatkoanalyyseinä onkin lisäksi mahdollista toteuttaa yksityiskohtaisempia ryhmäkohtaisia vertailuja niin sanottujen post-hoc -testien (ryhmien parittaiset vertailut) ja ennalta suunniteltujen niin sanottujen kontrastien avulla, mistä saatu tilastollinen merkitsevyys johtuu. Varianssianalyysin jatkoanalyysinä ei tule käyttää lukuisia erillisiä tavallisia t-testejä, sillä näin testaukseen liittyvää merkitsevyystasoa ei saada ilman jotakin korjausmenettelyä kontrolloitua yleisesti sovitulle tasolle (esim. 5 %). Erilliset t-testit eivät ota huomioon vasteen kokonaisvaihtelua, johon parivertailut suhteutetaan. Ilman jatkoanalyyseja varianssianalyysin tulos jää melko yleiselle tasolle.

Vaikka tilasto-ohjelmissa onkin mahdollista tutkia varianssianalyysillä usean eri tekijän samanaikaista vaikutusta vasteeseen kannattaa tutkimushypoteesit monesti muodostaa siten, että analyysissä rajoitutaan tarkastelemaan korkeintaan kolmen tekijän yhtäaikaista vaikutusta vasteeseen. Jos esimerkkiaineiston mukaisesti meillä on edellä mainitut kolme tekijää (sukupuoli, vanhempien peruskoulutus ja viriketausta), joiden vaikutusta opintomenestykseen halutaan tutkia, joudutaan asetelmaan, jossa verrataan 2×3×3 = 18 eri ryhmäkeskiarvoa toisiinsa! Jos havaintoja on 60, jää kuhunkin soluun keskimäärin 60/18 eli noin kolme tapausta. Kun käytännössä havainnot eivät koskaan jakaudu tasaisesti tekijöiden eri tasoihin, päädytään helposti tilanteeseen, jossa johonkin tasoyhdistelmään ei tule yhtään tapausta. Tutkimusasetelmat, joissa tekijöitä olisi yli neljä, eivät esimerkiksi Kerlingerin (1981) mukaan ole kovin käyttökelpoisia. Jos on tarvetta tutkia useamman muuttujan vaikutusta tutkittavaan tekijään, on monesti käytännöllisempää käyttää varianssianalyysin sijaan muita monimuuttujamenetelmiä, kuten regressioanalyysia tai muita monimuuttujamallinnukseen sopivia menetelmiä (mm. rakenneyhtälömallitukseen liittyvät polkumallit jne.)

Varianssianalyysissa tutkitaan siis sekä kunkin tekijän vaikutusta erikseen että eri tekijöiden yhdysvaikutusta vasteeseen. Kahden tekijän yhdysvaikutuksella tarkoitetaan toisen tekijän vaikutusta vasteeseen ensimmäisen tekijän eri luokissa. Jos esimerkkiaineistossa sukupuolen ja viriketaustan välinen yhdysvaikutus opintomenestykseen olisi tilastollisesti merkitsevä, tarkoittaisi se sitä, että miehillä ja naisilla viriketausta vaikuttaisi eri tavalla opintomenestykseen. Mikäli varianssianalyysin tulos on se, että yhdysvaikutus on tilastollisesti merkitsevä, on selvitettävä aina, esimerkiksi ryhmäkeskiarvojen avulla, mistä tuo merkitsevyys johtuu. Järvikoski & von Wright (1970, 10) antavat yhdysvaikutuksesta hyvän esimerkin: "...Yhdysvaikutuksen käsitettä selvittää seuraava esimerkki: Edellä mainitussa lukutestissä oli luokitusperusteina mm. sukupuoli ja luokka-aste. Jos havaitaan, että tytöt ovat keskimäärin parempia alaluokilla, pojat yläluokilla, voidaan puhua sukupuolen ja luokka-asteen yhdysvaikutuksesta." Toisin sanoen, tyttöjen ja poikien suoritusten väliset erot muuttuvat luokka-asteen funktiona. Asian selkeyttämiseksi tarkastellaan yhdysvaikutuskäsitettä myös graafisesti.

Seuraavassa kuviossa 10 kuvion vasemmanpuoleisessa osakuviossa on esitetty ensinnäkin tilanne, jossa tekijöillä on yhdysvaikutusta (Järvikosken & Wrightin esimerkin mukainen tilanne), ja toiseksi, kuvion oikeanpuoleisessa osakuviossa, tilanne, jossa sukupuolella ja luokalla ei ole yhdysvaikutusta.



KUVIO 10. Esimerkki tilanteesta, jossa tekijämuuttujien sukupuoli ja luokkataso välillä on ja ei ole yhdysvaikutusta (vasemmanpuolisessa asetelmassa muuttujien välillä on yhdysvaikutusta, oikeanpuolisessa tätä ei ole).

Kuviossa on siis tilanne, jossa tekijöillä on yhdysvaikutusta. Tyttöjen opintomenestys on alaluokilla poikien menestystä parempi. Koulunkäynnin myötä poikien menestyminen opinnoissa kuitenkin paranee; sen sijaan tyttöjen saavuttama opintomenestys pysyy melko vakiona riippumatta luokka-asteesta. Tyttöjen ja poikien kehitysurat eivät siis ole samanlaiset. Tilanteessa, jossa sukupuolella ja luokalla ei ole yhdysvaikutusta (kuvio 10, oikeanpuolinen kuvio) lukutestin vasteeseen, kuvio näyttää aivan erilaiselta kuin vasemmanpuolinen kuvio. Tyttöjen ja poikien lukutestipistemäärät nousevat koulunkäynnin edistyessä suunnilleen samalla tavalla (käyrät muodoltaan samanlaisia ja samansuuntaisia). Ero tyttöjen ja poikien saavutustasossa ei muutu olennaisesti heidän koulunkäyntinsä edistymisen myötä, eli tytöt saavuttavat lukutesteissä poikia korkeampia pistemääriä kaikilla luokka-asteilla. Näin asetelma pysyy jatkuvasti samansuuntaisena. Voidaan todeta sekä sukupuolen että luokka-asteen vaikuttavan positiivisesti lukutestissä menestymiseen.

SPSS: Viivadiagrammin piirtäminen (kaksi kategorista muuttujaa ja yksi numeerinen):

1. Valitse ylhäältä valikkoriviltä Graphs → Legacy Dialogs → Line
   - valitse esiin tulevista kuvakkeista *Multiple*-kuvake ja kohdan *Data in Chart Area* vaihtoehdoista ruksaa *Summaries for groups of cases*. Jatka *Define*-painikkeella seuraavaan vaiheeseen
   - valitse muuttujalistasta haluamasi kategorinen muuttuja, ja siirrä se listan oikealla puolella olevalla nuolipainikkeella kohtaan *Category Axis*. Tämän muuttujan arvot tulevat x-akselin arvoiksi
   - kohtaan *Define* Lines *by* valitset toisen kategorisen muuttujan; tämän muuttujan arvot määräävät kuvioon muodostuvien janaviivojen lukumäärän.
   - valitse lopuksi numeerinen muuttuja seuraavasti: aktivoi klikkaamalla kohdan *Lines Represent* vaihtoehto *Others statistic*. Valitse vasemmalla olevasta muuttujalistasta haluamasi numeerinen muuttuja, ja siirrä se listan oikealla puolella olevalla nuolipainikkeella kohtaan *Variable*. Kohdasta *Change Statistic* valitset sen, mitä numeerisen muuttujan arvoja kuvaavat viivat esittävät, esimerkiksi vaikkapa keskiarvoja (*Mean of values*).
2. *Options*-painike
   - tarkista, että mahdolliset puuttuvat tiedot eivät tule mukaan kuvioon omaksi luokakseen, eli ettei kohdassa *Display groups defined by missing values* ole valintaruksia
   - hyväksy valinnat klikkaamalla *Continue*-painiketta
3. Toteuta viivadiagrammi painamalla *OK*-painiketta, ja saat graafin tulosteikkunaan

Halutessasi voit jatkaa kuvion muokkaamista tulosteikkunassa (ks. aiempi ohje, 97).

5.2 Yksisuuntainen varianssianalyysi

Seuraavassa tehdään yksisuuntainen eli yhden tekijän varianssianalyysi. Otetaan tekijäksi viriketausta ja vasteeksi ensimmäinen opintomenestysmuuttuja. Tavoitteena on siis selvittää, eroaako opintomenestys viriketaustamuuttujan kolmessa eri viriketaustakategoriassa. Nollahypoteesina on keskiarvojen yhtäsuuruus, eli opintomenestys on samanlaista kaikissa viriketaustaluokissa. Mikäli viriketaustalla ei ole vaikutusta opintomenestykseen, eivät keskiarvot poikkea tilastollisesti merkitsevästi toisistaan. Koska valitussa tekijässä, viriketaustassa, on peräti kolme eri tasoa (ryhmät matala/keskinkertainen/korkea), ei sen vaikutuksen analysoiminen onnistu t-testillä, vaan on käytettävä varianssianalyysiä.

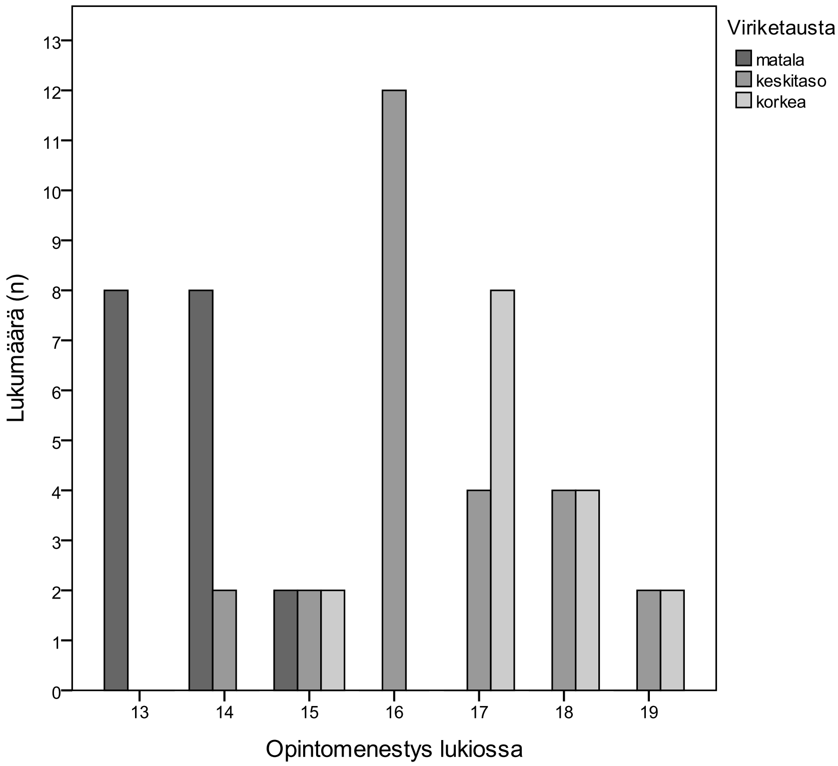
Ennen varsinaista analyysia käydään läpi varianssianalyysin käytön edellytysten voimassaolo sekä tutustutaan tutkimuskohteeseen laskemalla vastemuuttujan keskiarvot tekijän eri tasoilla:

TAULUKKO 6. Opintomenestys lukiossa -muuttujan keskiarvot luottamusväleineen ja keskihajonnat viriketaustan mukaan.

Opintomenestys lukiossa

				95% Confidence Interval for Mean	
	N	Mean	Std. Deviation	Lower Bound	Upper Bound
matala	18	13,67	,69	13,33	14,01
keskitaso	26	16,46	1,30	15,94	16,99
korkea	16	17,25	1,13	16,65	17,85
Total	60	15,83	1,82	15,36	16,30

Yllä oleva taulukko sisältää ryhmäkoot, keskiarvot, keskihajonnat ja keskiarvojen luottamusvälit (ala- ja ylärajat). Kuten keskiarvoista huomataan, opintomenestys näyttäisi olevan sitä parempi, mitä korkeampi opiskelijan viriketausta on. Luottamusvälien avulla voidaan tulkita, että esimerkiksi matalan viriketausta opiskelijoiden keskimääräinen opintomenestys sijoittuu välille [13,33, 14,01], joka on kokonaisuudessaan matalampi kuin esimerkiksi keskitason viriketaustan omaavien opiskelijoiden (ks. tarkemmin luottamusvälit ja niiden tulkinta, s. 49–51, 127–129). Ryhmien keskihajonnat vaihtelevat siten, että selvästi pienintä vaihtelu opintomenestyksessä on matalan viriketaustan ryhmässä. Katsotaan vielä, miltä opintomenestysmuuttujan jakaumat näyttävät graafisesti tarkasteltuna eri viriketaustaryhmissä:



KUVIO 11. Opintomenestys lukiossa viriketaustan mukaan.

Ryhmäkohtaisista jakaumista näkyy selvästi matalan viriketaustaryhmän pieni varianssi: peräti 16/18 matalaryhmäläisistä jää alle 15 opintomenestyspistemäärän, ja arvot painottuvat pistemääriin 13 ja 14. Kuviosta voi myös nähdä ryhmien väliset keskiarvoerot, sillä matalan viriketaustan jakauma painottuu kuviossa vasempaan reunaan, keskitaso keskelle ja korkea viriketausta oikean puoleiseen osaan.

Ryhmien keskiarvojen välisiä eroja testataan ANOVA:ssa F-testiä käyttäen. Testi on herkkä niin sanotuille haja-arvoille, toisin sanoen jos jossain ryhmässä on havaintoarvo, joka poikkeaa paljon muista ryhmän arvoista jompaankumpaan suuntaan ja tätä kautta suurentaa hajontaa ja muuttaa keskiarvoa, saattaa varianssianalyysin tulos olla varsin epäluotettava (jo yksi tai kaksi tällaista haja-arvoa alle 10 havainnon ryhmässä voi nostaa F-testin arvon kohtuuttoman suureksi ja sitä kautta saada jonkin tekijän vaikutuksen näyttämään virheellisesti tilastollisesti merkitsevältä). Yllä olevista jakaumista nähdään, ettei tällaisia mahdollisesti haitallisia haja-arvoja näyttäisi olevan. Kuvio antaa myös hyvän esimerkin siitä, miksi usein tutkimusasetelmaa tai yleensä käsiteltäviä asioita kannattaa tarkastella aluksi graafisesti: se selkeyttää usein tutkijalle tutkimusasetelmaa. Kaikkea ei toki kannata laittaa lopulliseen raporttiin, tässä suhteessa tarvitaan harkintaa.

Nyt kun meillä on tuntuma tutkimuskohteeseen, siirrytään käsittelemään tarkemmin varianssianalyysin käytön edellytysten voimassaoloa. Levenen testin p-arvo populaatiovarianssien yhtäsuuruudelle on 0,106 eli $p \geq 0{,}05$, joten erot ryhmien otosvariansseissa eivät ole niin merkittäviä, etteikö voitaisi pitää myös ryhmien populaatiovariansseja yhtä suurina. Viriketaustaryhmät ovat selvästi myös toisistaan riippumattomat, sekin varianssianalyysin edellytyksistä on siis voimassa. Varianssianalyysissä käytettävä F-testi toimii varsin vakaasti, vaikka normaalisuusoletus ei olisikaan täysin voimassa, kunhan ryhmäkoot ovat riittävän suuret (n on vähintään 20 kaikilla tekijän tasoilla eli ryhmissä, ks. esimerkiksi Hair ym. 2010). Esimerkkiaineistossa ryhmien koot ovat 18, 26 ja 16, joten voidaan varovasti tältäkin pohjalta päätellä, että normaalijakaumaoletus on voimassa. Tarvittaessa normaalijakaumaoletuksen täyttyminen voidaan tarkistaa graafisesti histogrammin avulla, jakauman muotoa kuvaavien tunnuslukujen avulla ja normaalisuuden testaukseen tarkoitetulla sopivalla testillä, esimerkiksi Shapiro-Wilkin tai Kolmogrov-Smirnovin testeillä (ks. tarkemmin tietolaari 3, s. 103, normaalijakauma ja normaalijakauman testaus). Näin voidaan päätellä, että viriketaustan vaikutusta opintomenestykseen voidaan tutkia varianssianalyysin avulla. Mikäli oletuskriteerit eivät olisi täyttyneet, analyysin olisi voinut suorittaa vastaavalla epäparametrisella Kruskal-Wallisin testillä (ks. luku 5.5, s. 162).

Varianssianalyysin testiosan SPSS-tuloste näyttää esimerkissämme seuraavalta:

ANOVA

Opintomenestys lukiossa

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	126,872	2	63,436	52,055	,000
Within Groups	69,462	57	1,219		
Total	196,333	59			

ANOVA-taulukossa neliösummat (Sum of Squares) kuvaavat ryhmien välistä (Between Groups) ja ryhmien sisäistä (Within Groups) vaihtelua; näiden kahden summa on kokonaisvaihtelu (Total). Kun ryhmävaihtelut jaetaan vapausasteillaan (df), saadaan esiin keskimääräinen vaihtelu eli keskineliösummat (Mean Square). Keskineliösummat jaetaan keskenään, jolloin saadaan testisuureen F-arvo. Lopuksi testisuureen arvon ja vapausastelukujen avulla lasketaan testiin liittyvä p-arvo (Sig. = Significance). Varianssianalyysin idea on siis se, että mitä suurempi on ryhmien välinen vaihtelu suhteessa ryhmien sisäiseen vaihteluun, sitä suuremmaksi tulee testisuureen F-arvo ja sitä pienemmäksi siihen liittyvä p-arvo. Viriketaustan vaikutus opintomenestykseen lukiossa on siis tilastollisesti merkitsevä, $p = 0{,}000$ ($p < 0{,}001$). Yllä olevaan taulukkoon ei tulostunut efektikokoon suureita (eta^2, eta_p^2 tai omega), mutta niitä saadaan laskettua käyttämällä sopivaa SPSS:n varianssianalyysiin liittyvää toteutustapaa (ks. alla oleva SPSS-ohjelaatikko). Tässä tapauksessa efektikokoon arvoksi saataisiin $eta^2 = 0{,}65$, jolloin viriketaustan vaikutus opintomenestykseen voidaan tulkita voimakkuudeltaan suureksi, kun tulkinnassa käytetään raja-arvoja 0,01 pieni – 0,06 keskisuuri – 0,14 suuri (ks. asetelma 5, s.49; Ellis 2010, 41). Tilastolliseen päättelyyn liittyvä tulos raportoidaan yleensä seuraavasti: $F(2, 57) = 52{,}06$; $p < 0{,}001$; $eta^2 = 0{,}65$ (arvot 2 ja 57 = vapausasteet, eli degrees of freedom df. Yksisuuntaisessa varianssianalyysissä vapausasteet muodostuvat ensiksikin ryhmien lukumäärän mukaan, jossa huomioidaan kokonaiskeskiarvon laskeminen (tässä $df_1 = 3\text{-}1 = 2$), sekä toiseksi havaintojen lukumäärän mukaan, kun huomioidaan ryhmäkeskiarvojen laskeminen (tässä $df_2 = N\text{-}3 = 57$).

Keskiarvojen avulla tulkittuna ANOVA:n tulos kertoo tässä sen, että ainakin jonkin keskiarvoparin erotus (ainakin suurimman ja pienimmän keskiarvon erotus) on tilastollisesti merkitsevä. Mutta voihan tilastollisesti merkitseviä eroja olla muidenkin keskiarvoparien joukossa, joten meitä kiinnostaa edelleen, mitkä viriketaustaryhmien keskiarvot pareittain verrattuina sitten poikkeavat toisistaan. Jotta saataisiin tarkempi kuva siitä, miten viriketaustaltaan erilaisten opiskelijoiden opintomenestys eroaa toisistaan ja eroavatko eri ryhmien keskiarvot toisistaan tilastollisesti merkitsevästi, meidän on tehtävä varsinaisen varianssianalyysin lisäksi myös parittaiset vertailut post hoc -testein. Näitä monivertailutestejä on monenlaisia, tässä käytetään yleisimpiä eli Tukeyn ja Bonferronin testejä. SPSS antaa seuraavanlaisen tulosteen:

Multiple Comparisons

Dependent Variable:Opintomenestys lukiossa

	(I) Viriketausta	(J) Viriketausta	Mean Difference (I-J)	Std. Error	Sig.
Tukey HSD	matala	keskitaso	-2,79*	,34	,000
		korkea	-3,58*	,38	,000
	keskitaso	matala	2,79*	,34	,000
		korkea	-,79	,35	,072
	korkea	matala	3,58*	,38	,000
		keskitaso	,79	,35	,072
Bonferroni	matala	keskitaso	-2,79*	,34	,000
		korkea	-3,58*	,38	,000
	keskitaso	matala	2,79*	,34	,000
		korkea	-,79	,35	,085
	korkea	matala	3,58*	,38	,000
		keskitaso	,79	,35	,085

*. The mean difference is significant at the 0.05 level.

Matalaviriketaustaisten opintomenestyspistemäärän keskiarvo poikkeaa tilastollisesti merkitsevästi sekä keski- että korkeaviriketaustaisten keskiarvoista (p-arvo molemmissa vertailuissa on 0,000 eli p < 0,001). Sen sijaan keski- ja korkeaviriketaustaiset eivät tämän aineiston perusteella poikkeaisi toisistaan lukioaikaisen opintomenestyksen suhteen (Tukey antaa p-arvon 0,072 ja Bonferroni arvon 0,085; siten p ≥ 0,05). Tarkempaa esittelyä ja vertailua eri post hoc -testien ominaisuuksista löydät esimerkiksi Andy Fieldin kirjasta (2018, 549–551). Seuraavassa on ohjeet analyysin suorittamiseksi SPSS-ohjelmassa.

SPSS: Yksisuuntainen varianssianalyysi

1. Löydät analyysiin valikkoriviltä *Analyze: Compare Means → One-Way ANOVA*
2. Siirrä valitsemasi numeerinen muuttuja muuttujaluettelon oikealla puolella olevalla ylemmällä nuolipainikkeella kohtaan *Dependent List*

- kyse on siis vastemuuttujan (selitettävän tai riippuvan muuttujan) valinnasta
- jos haluat tehdä samalla kertaa useita varianssianalyyseja, voit valita tähän muitakin muuttujia

3. Siirrä sitten valitsemasi kategorinen muuttuja alemmalla nuolinäppäimellä kohtaan *Factor*
   - kyse on siis selittävästä tekijästä eli riippumattomasta/selittävästä muuttujasta

Lisävalinnat:

4. *Options*-painikkeen alta
   - valitaan *Descriptives*
   analyysi tulostaa tällöin tekijämuuttujan ryhmissä tutkittavan vastemuuttujan keskiarvot, keskihajonnat ja luottamusvälit. Tämä on tärkeä valinta, koska näin tulosteisiin tulee varsinaisen varianssianalyysin lisäksi vertailtavien ryhmien

ohjetaulukko jatkuu...

keskiarvot yms. tärkeät tunnusluvut näkyviin, tältä pohjalta voit tehdä yleisen ryhmissä esiintyvien erojen tai yhtäläisyyksien tulkinnan
  - varianssianalyysin käyttöoletuksiin kuului varianssien yhtäsuuruusoletus, joten valitse myös kohta *Homogeneity-of-variance* tämä tuottaa tulosteeseen lisätaulukon, joka sisältää testin sille, että onko arvojen vaihtelu likimain samanlaista kaikissa vertailtavissa ryhmissä.
  - lisäksi samalla kannattaa ruksittaa valinnoista myös *Brown-Forsythe-testi* tai *Welchin testi* testi, jonka tulosta kannattaa käyttää ANOVA:n normaalin F-testin sijaan, jos vertailtavissa ryhmissä varianssien yhtäsuuruus ei ole voimassa
- hyväksy valinnat klikkaamalla *Continue*-painiketta

5. *Post Hoc* -painikkeen alta
   - jos valitsemassasi tekijämuuttajassa (kategorinen muuttuja) on useampia ryhmiä, joiden keskiarvoja haluat vertailla, valitse *Equal Variances Assumed* -testeistä esimerkiksi *Tukey* (jos ryhmien varianssien yhtäsuuruusolettamus pitää paikkansa; jos ei pidä, valitse testiksi *Equal Variances Not Assumed* -osiosta esimerkiksi *Dunnett's T3*
   - hyväksy valinnat *Continue*-painikkeella
6. Toteuta yksisuuntainen varianssianalyysi yllä määrittelemilläsi valinnoilla *OK*-painiketta painamalla ja saat analyysien tulosteet tulosteikkunaan.

Huom: Ryhmäkeskiarvojen luottamusvälit tulostuvat tunnuslukutaulukkoon, joka saadaan yllä olevassa kohdassa 4 kuvatulla tavalla (*Option* -painikkeen takana olevalla valinnalla *Descriptives*)

Huom: Efektikoon saat laskettua toiminnossa *Analyze → Compare Means → Means* tai *Analyze → General Linear Model → Univariate*, käyttäen painiketta *Options*, josta rastitaan kohta *Estimates of Effect Size.*

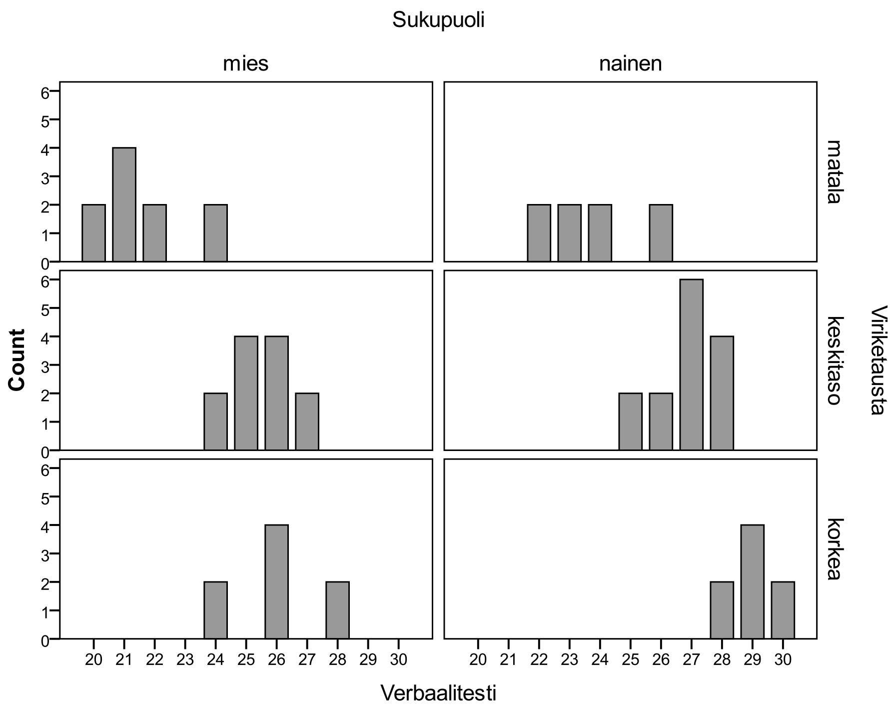
5.3 Kaksisuuntainen varianssianalyysi

Seuraavaksi tarkastellaan sitä, miten sukupuoli ja viriketausta vaikuttavat samanaikaisesti verbaalisuustestin tulokseen. Kuten yksisuuntaista varianssinanalyysiä koskevan esimerkin kohdalla, havaintoaineistoa tarkastellaan ensin yleisluontoisesti, sen jälkeen tarkastellaan varianssianalyysin soveltuvuutta kyseessä olevan tutkimusongelman ratkaisemiseen. Kun vastaus on tähän myönteinen, tehdään itse analyysiin mahdollisesti tarvittavat täsmennykset. Tutkitaan siis aluksi ryhmäkeskiarvoja ja keskihajontoja. Koska sukupuolimuuttujalla on kaksi tasoa (miehet/naiset) ja viriketaustamuuttujalla kolme (matala /keskitaso/korkea), ryhmiä on kaikkiaan kuusi kappaletta:

TAULUKKO 7. Verbaalisuustestimuuttujan tunnuslukuja viriketaustan ja sukupuolen mukaan.

	Verbaalisuustesti		
Sukupuoli/Viriketausta	**keskiarvo**	**keskihajonta**	**n**
miehet/matala	21,60	1,43	10
miehet/keskitaso	25,50	1,00	12
miehet/korkea	26,00	1,51	8
naiset/matala	23,75	1,58	8
naiset/keskitaso	26,86	1,03	14
naiset/korkea	29,00	0,76	8
kaikki ryhmät	25,47	2,55	60

Keskiarvoista voi nähdä, miten sekä miehillä että naisilla verbaalisuustestipistemäärän keskiarvo nousee siirryttäessä korkeampaan viriketaustaluokkaan. Keskiarvoille olisi hyvä tulostaa taulukkoon näkyviin myös estimoinnin tarkkuutta ilmaisevat 95 %:n luottamusvälit, jotka tästä puuttuvat. Esimerkiksi ylimpänä olevalle ryhmälle keskiarvoon liittyvä luottamusväli olisi: 95 % CI: [20,6 , 22,6]. Tämän perusteella estimointi on varsin tarkkaa, sillä keskiarvon estimointiin liittyy vain noin yhden yksikön (testipisteen) mittainen virhemarginaali. Keskiarvoja silmämääräisesti vertailtaessa nähdään, että miehillä suurin nousu on siirryttäessä matalasta viriketasosta keskinkertaiseen, kun taas naisilla nousu tapahtuu tasaisemmin. Tarkasteltaessa naisia ja miehiä erikseen nähdään, että naisten keskiarvot ovat kussakin viriketaustamuuttujan luokassa vastaavia miesten keskiarvoja korkeampia. Solujen varianssit tai tässä niiden neliöjuuret eli keskihajonnat, eivät poikkea paljonkaan toisistaan. Tarkastellaan seuraavaksi vasteen jakaumia eri soluissa. Tässä graafinen tarkastelu on jälleen paikallaan:



KUVIO 12. Verbaalisuustestin taso sukupuolittain ja viriketaustoittain.

Kuviossa visualisoituvat selvästi samat seikat, jotka havaittiin jo edellä solujen tunnuslukujen tarkastelun yhteydessä: mitä korkeampi viriketaso on, sitä lähemmäksi kuvan oikeaa reunaa havainnot siirtyvät. Näin käy sekä miehillä että naisilla. F-testin, johon varianssianalyysin tulos perustuu, kannalta vaarallisia haja-arvoja ei ryhmissä näytä olevan (haja-arvot voivat muuttaa ryhmäkeskiarvoa ja -hajontaa dramaattisesti ja johtaa vääriin päätelmiin, kuten yksisuuntaisen varianssianalyysin kohdalla todettiin).

Seuraavaksi arvioidaan, miten kaksisuuntainen varianssianalyysi sopii menetelmäksi tähän tutkimusongelmaan. Tutkitaan ensin ryhmien populaatiovarianssien yhtäsuuruutta. Levenen testi antaa p-arvon 0,341 eli $p \geq 0{,}05$, joten kyseisiä variansseja voidaan pitää riittävän yhtä suurina. Koska ryhmien koot ovat pienet ($n = 8–14$), myös se, noudattaako vaste normaalijakaumaa, on syytä tarkistaa. Vasteen normaalijakautuneisuus pitää tarkistaa kussakin ryhmässä erikseen; tulokset on koottu seuraavaan taulukkoon:

TAULUKKO 8. Verbaalisuustestimuuttujan normaalijakautuneisuuden testaus.

Sukupuoli/ Viriketausta	Kolmogorov-Smirnov -testin p-arvo	Shapiro-Wilk -testin p-arvo
miehet/matala	0,049	0,054
miehet/keskitaso	0,200	0,187
miehet/korkea	0,150	0,093
naiset/matala	0,200	0,175
naiset/keskitaso	0,007	0,025
naiset/korkea	0,150	0,093

Kuten aiemmin luvussa 3.3 oli esillä, Shapiro-Wilkin testiä käytetään yleensä pienten aineistojen ($n < 50$) ja Kolmogorov-Smirnovin testiä suurempien aineistojen ($n \geq 50$) normaalisuuden testaamiseen. Normaalijakaumaoletus on testien perusteella riittävän hyvin voimassa, sillä molempien testien p-arvot ovat pääsääntöisesti suurempia kuin 0,05. Vain ryhmä naiset/keskitaso on ongelmallinen, mutta siinäkin Shapiro-Wilk on melko lähellä raja-arvoa 0,05. Niinpä ei tarvitse olla kovin huolissaan tästä yhdestä poikkeuksesta. Myös kolmas varianssianalyysin käytön perusedellytys, ryhmien riippumattomuus, on voimassa, sillä kukin vastaajista on vain yhdessä ryhmässä.

Kun varianssianalyysin suoritetaan, saadaan päätulokseksi seuraavanlainen SPSS-tuloste:

Tests of Between-Subjects Effects

Dependent Variable: Verbaalisuustestin pistemäärä

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	302,32[a]	5	60,46	40,50	,000	,789
Intercept	37029,03	1	37029,0	24804,1	,000	,998
vir_t	217,77	2	108,89	72,94	,000	,730
SUKUPUOL	67,24	1	67,24	45,04	,000	,455
vir_t * SUKUPUOL	6,75	2	3,38	2,26	,114	,077
Error	80,61	54	1,49			
Total	39296,00	60				
Corrected Total	382,93	59				

a. R Squared = ,789 (Adjusted R Squared = ,770)

Taulukon otsikossa esiintyy sana *Effect* eli vaikutus. Sarakkeessa Sig. on mallissa olevien vaikutusten p-arvot ja viimeisellä sarakkeessa efektikoot eta_p^2-suureen arvoilla. Efektikoon arvot voidaan tulkita prosenttilukuina, paljonko kyseinen vaikutus selittää tutkittavan muuttujan vaihtelusta (ks. Ellis 2010, 41). Kaksisuuntaisessa varianssianalyysissä ehkä keskeisin tulos on yhdysvaikutuksen tilastolliseen merkitsevyyteen liittyvä arvo, joka nähdään taulukossa rivillä VIRIKET*SUKUPUOL. Tässä se ei ole tilastollisesti merkitsevä, koska $p = 0{,}114$ eli $p \geq 0{,}05$. Yhdysvaikutuksen efektikoko on 8 prosenttia, eta_p^2 arvon ollessa 0,08. Tutkimusraportissa tilastolliseen päättelyyn liittyvät tulokset merkitään yleensä APA-ohjeistuksen suosittamassa muodossa, jossa keskeiset testiin liittyvät suureet ja lukuarvot (vapausasteluvut df, testisuure F ja p-arvo) sekä efektikoko ovat näkyvissä, eli tässä nämä tiedot raportoitaisiin tekstiin sanallisen tulkinnan yhteyteen muodossa $F(2, 54) = 2{,}26$; $p = 0{,}114$; $eta_p^2 = 0{,}08$. Riviltä SUKUPUOL nähdään, että sukupuolen vaikutus verbaalisuustestin tulokseen on tilastollisesti merkitsevä, sillä p-arvo on 0,000 eli $p < 0{,}001$. Vastaavasti nähtäisiin, että myös viriketausta (rivi VIRIKET) on vaikutukseltaan merkitsevä. Näillä niin sanotuilla päävaikutuksilla efektikokoja voidaan pitää suurina. Tulokset tarkoittavat sitä, että vaikka sukupuoli ja viriketausta vaikuttavatkin verbaalisuustestin tulokseen, niin viriketaustan vaikutus on samanlaista kummallakin sukupuolella ja että sukupuolen vaikutus on samanlaista kaikissa kolmessa viriketaustaryhmässä.

Kuten yksisuuntaisenkin varianssianalyysin kohdalla, analyysia jatketaan vielä sen selville saamiseksi, miten eri ryhmät poikkeavat toisistaan. On siis monivertailutestin vuoro. Koska viriketausta osoittautui merkitseväksi selittäjäksi ja siinä eri tasoja oli 3, toteutetaan näiden kolmen ryhmän parittaiset vertailut post hoc -testein. Koska ryhmiä on kolme, parittaisia vertailuja muodostuu kolme kappaletta. Toteutetaan tässä vertailut Tukeyn testein, joiden tulokset nähdään alla olevasta taulukosta:

TAULUKKO 9. Tukeyn monivertailutestit p-arvoineen viriketason eri tasokombinaatioilla.

Multiple Comparisons

Dependent Variable: Verbaalisuustestin pistemäärä
Tukey HSD

(I) Viriketausta	(J) Viriketausta	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Lower Bound	95% Confidence Upper Bound
matala	keskitaso	-3,68*	,50	,000	-4,87	-2,48
	korkea	-4,94*	,56	,000	-6,28	-3,61
keskitaso	matala	3,68*	,50	,000	2,48	4,87
	korkea	-1,27*	,51	,043	-2,51	-,03
korkea	matala	4,94*	,56	,000	3,61	6,28
	keskitaso	1,27*	,51	,043	,03	2,51

Taulukossa kolmannessa sarakkeessa *Mean Difference* nähdään vertailtavien ryhmien keskiarvojen eron suuruus ja sarakkeessa *Sig.* on tuttuun tapaan nähtävissä eron tilastollinen merkitsevyys. Kuten sarakkeesta nähdään, ovat kaikki ryhmien väliset erot tilastollisesti merkitseviä ($p < 0,05$).

Toinen tapa vertailla eri ryhmäkeskiarvoja toisiinsa on käyttää hyväkseen kontrasteja. Kontrastien avulla voidaan määrätä, millaisia vertailuja halutaan tehdä. Periaatteena on, että kullekin tekijän tasolle annetaan kontrastikerroin, joka ilmaisee sen aseman kontrastitarkastelussa. Jos kontrastikerroin on nolla, tarkoittaa se sitä, että tätä tasoa ei haluta mukaan vertailuun. Verrattaville tasoille annetaan kontrastikertoimiksi vastaluvut, ja kaikkien tasojen kontrastikerrointen summan tulee olla nolla. Jos siis halutaan verrata viriketaustamuuttujan tasojen ”matala” ja ”korkea” ryhmäkeskiarvoja toisiinsa ja halutaan jättää tarkastelusta pois keskiryhmän, olisivat kontrastikertoimet -1, 0 ja 1. Kontrasteja voi käyttää myös luokkayhdistelmien vertaamiseen. Jos vaikka halutaan verrata luokkien ”matala” ja ”keskinkertainen” yhdistettyä ryhmäkeskiarvoa ryhmän ”korkea” keskiarvoon, asettaisimme kontrasteiksi -1, -1 ja 2. Itse määritettävien kontrastien lisäksi tilasto-ohjelmissa on myös valittavissa valmiiksi määriteltyjä, tyypillisimmin käytettyjä kontrastivertailuja. Tällaisia ovat esimerkiksi SPSS-ohjelman kontrastit *simple* tai *repeated*, joista ensiksi mainitussa ensimmäinen tekijän taso (eli vertailtava ryhmä) asetettaisiin vertailuryhmäksi, johon muita tasoja verrataan, ja jälkimmäisessä vertailut kohdistettaisiin aina vierekkäisiin ryhmiin (käytetyssä esimerkissämme vertailut ”matala” vs. ”keskinkertainen” ja ”keskinkertainen” vs. ”korkea”).

Katsotaan lopuksi, miten kaksisuuntainen varianssianalyysi toteutetaan SPSS-ohjelmalla.

SPSS: Kaksisuuntainen varianssianalyysi

1. Löydät analyysin valikkoriviltä *Analyze: General Linear Model → Univariate*
2. Siirrä valitsemasi numeerinen muuttuja (vaste, riippuva/selitettävä muuttuja) muuttujaluetteloon *Dependent Variable* oikealla puolella olevan ylemmän nuolipainikkeen avulla
3. Siirrä valitsemasi kaksi kategorista muuttujaa (tekijää, riippumatonta muuttujaa) kohtaan *Fixed Factor(s)* oikealla puolella olevan toiseksi ylimmän nuolinäppäimen avulla

Lisää valinnat:

4. *Options*-painikkeen alta
   - valitaan *Descriptives*
     analyysi tulostaa tällöin tekijämuuttujan ryhmien keskiarvot, keskihajonnat ja luottamusvälit
     tärkeä valinta, koska näin tulosteisiin tulee näkyviin varsinaisen varianssianalyysin lisäksi vertailtavien ryhmien keskiarvot yms. tärkeät tekijät, tältä pohjalta voit tehdä yleisen ryhmissä esiintyvien erojen tai yhtäläisyyksien tulkinnan; pelkkä varianssianalyysien pohjalta todettu erojen tilastollinen merkitsevyystaso ei yksinään riitä tuloksen tulkinnaksi
   - valitse kohta *Homogeneity tests* testaa tekijämuuttujien eri ryhmien jakaumat vastemuuttujan suhteen (on siis varianssien yhtäsuuruus-testi)
   - ruksaa lopuksi myös *Estimates of effect size*, jolla saadaan mallin tulosteeseen merkitsevyystestien taulukkoon myös efektikoon sarakkeelleen
   - hyväksy valinnat klikkaamalla *Continue*-painiketta
5. *Post Hoc* -painikkeelta
   - jos teet monivertailutestejä, valitse ne täältä. Valittavissa on mm. yksisuuntaisessa varianssianalyysissä esillä olleet *Tukey* ja *Dunnett* testit, joita tässäkin voi käyttää ja jotka myös tässä yhteydessä toteutetaan erikseen analyysissä oleville ryhmitteleville tekijöille.
   - jos haluat verrata ryhmäkeskiarvoja erityisesti tutkimushypoteeseihin perustuvien vertailujen mukaisesti (ns. kontrastit, planned contrasts), on tällaisia mahdollista määritellä erillisen *Contrasts* -painikkeen takana
   - hyväksy valinnat klikkaamalla *Continue*-painiketta
6. *Plots*-painikkeelta
   - jos haluat tulosteeseen mukaan myös tekijöiden yhdysvaikutuksia kuvaavan viivadiagrammin (tällainen on monesti hyödyllinen tulosten tulkinnassa), vie täällä *Horizontal Axis* -kohtaan toinen analyysissä olevista kategorisista muuttujista (tekijä/riippumaton/selittävätekijä; kirjan esimerkissä tämä olisi viriketaustamuuttuja) ja *Separate Lines* -kohtaan toinen kategorisista tekijöistä eli esimerkissämme sukupuolimuuttuja
   - lisää yllä määritelty kuvio *Add*-painikkeella ja sitten koko valinta *Continue*-painikkeella
7. Toteuta kaksisuuntainen varianssianalyysi yllä määrittelemilläsi valinnoilla *OK*-painiketta painamalla ja saat analyysin tulosteet tulosteikkunaan.

Huom: Jos haluat laskea osaryhmien keskiarvoille myös väliestimaatit (eli luottamusvälit), ne saat toteutettua siten, että ensiksi jaat aineistosi toisen tekijän mukaan osaryhmiin toiminnolla *Data → Split File*, jonka jälkeen teet tunnuslukutaulukon osaryhmittäin toiminnolla *Analyze → Descriptive Statistics → Explore*. Siellä tutkittava vastemuuttuja siirretään yläosan kohtaan *Dependent List* ja toinen jäljellä oleva kategorinen tekijä keskimmäiseen laatikkoon *Factor List*. *Ok*-painike tuottaa tulosteen, jossa avautuvan taulukon yläosassa on osaryhmille keskiarvon luottamusvälin ala- ja yläraja.

5.4 Toistettujen mittausten varianssianalyysi

Edellä esitetyn varianssianalyysin käytön eräs edellytys oli se, että vertailtavat ryhmät ovat riippumattomat. Mikäli meillä on tilanne, jossa halutaan verrata toisiinsa kahta toisistaan riippuvaa ryhmää, päädytään toistettujen mittausten varianssianalyysiin (Repeated Measures). Tällainen tilanne syntyy esimerkiksi silloin, kun mitataan samaa muuttujaa kahtena (tai useampana) ajankohtana ja käytetään samoja koehenkilöitä. Jos ajankohtia on vain kaksi eikä muiden tekijöiden vaikutusta haluta tutkia samanaikaisesti, voidaan käyttää myös riippuvien otosten t-testiä, mutta jos ajankohtia on useampia kuin kaksi tai halutaan tutkia samanaikaisesti jonkin ulkoisen tekijän vaikutusta vasteeseen, tarvitaan toistettujen mittausten varianssianalyysia.

Tarkastellaan toistettujen mittausten t-testin yhteydessä esitettyä esimerkkiä, jossa tutkittiin opintomenestyksen muutosta siirryttäessä lukiosta yliopistoon erikseen miehillä ja naisilla. Tuossa esimerkissä datan muodosti kaksi samanlaista opintomenestysmuuttujaa, jotka oli mitattu samoilta koehenkilöiltä kahtena eri ajankohtana. Lisäksi haluttiin tutkia opintomenestyksessä tapahtunutta muutosta erikseen miehillä ja naisilla. Käytettäessä t-testiä analyysi tehtiin erikseen miehille ja erikseen naisille, joten sellaista analyysia ei pystytty tekemään, jossa olisi tarkasteltu samanaikaisesti sukupuolen ja opintomenestyksen muutosta. Toistettujen mittausten varianssianalyysin avulla tämä on mahdollista. Ongelmaa on hyvä lähestyä taas ryhmäkeskiarvojen avulla:

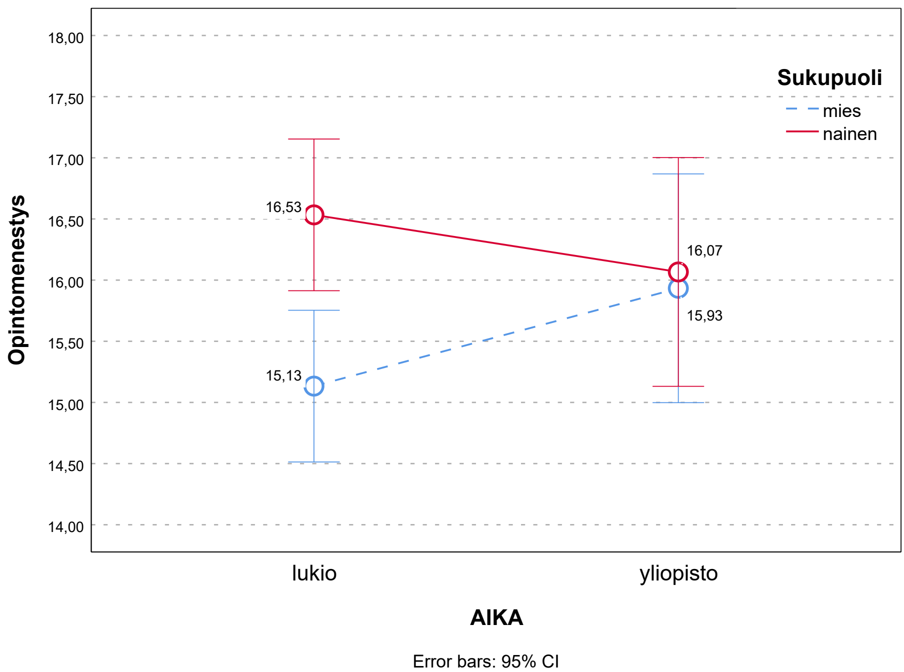
TAULUKKO 10. Ryhmien tunnuslukuja eri ajankohtina.

Sukupuoli	Aika (toisto)	keskiarvo	keskihajonta	n
miehet	opintomenestys lukiossa	15,13	1,70	30
miehet	opintomenestys yliopistossa	15,93	2,42	30
naiset	opintomenestys lukiossa	16,53	1,70	30
naiset	opintomenestys yliopistossa	16,07	2,69	30

Keskiarvoille olisi hyvä tulostaa tunnuslukutietoihin näkyviin myös estimoinnin tarkkuutta ilmaisevat 95 %:n luottamusvälit, jotka yllä olevasta tästä taulukosta puuttuvat. Esimerkiksi ylimpänä olevalle ryhmälle (miehet & lukiomenestys) keskiarvoon liittyvä luottamusväli olisi: 95 % CI: [14,51, 15,75]. Tämän perusteella estimointi on varsin tarkkaa, sillä keskiarvon estimointiin liittyy vain noin 0,6 yksikön (menestysindeksin) mittainen virhemarginaali keskiarvon ympärille suuntaan ja toiseen.

Sukupuolimuuttuja saa arvot mies ja nainen. Muuttuja on kategorinen. Toistetusti mitatusta muuttujasta ollaan yllä käyttäneet myös nimitystä aikamuuttuja. Sen arvot ovat opintomenestys lukiossa (1. mittaus) ja opintomenestys yliopistossa (2. mittaus). Molemmat opintomenestysmuuttujat ovat määrällisiä. Toistettujen mittausten varianssianalyysissamme on siis yksi kategorinen muuttuja ja kaksi määrällistä muuttujaa. Täten asetelma eroaa esimerkiksi kaksisuuntaisesta varianssianalyysista, jossa on kaksi kategorista muuttujaa ja yksi määrällinen muuttuja.

Yllä olevasta taulukosta nähdään, että miesten opintomenestyksen taso on alempana kuin naisten. Havaitaan kuitenkin, että miehillä opintomenestyksen taso nousee siirryttäessä lukiosta yliopistoon, kun taas naisilla käy päinvastoin. Miesten ja naisten opintomenestyksen taso näyttäisikin yliopistossa olevan jo lähes samaa tasoa. Ryhmien keskiarvoja kuvaava graafinen esitys valaisee asiaa:



KUVIO 13. Miesten ja naisten opintomenestyksen kehitys lukiosta yliopistoon.

Miesten ja naisten opintomenestystä ja sen muutosta kuvaavat suorat näyttävät todellakin kovin erilaisilta: muutos opintomenestyksessä on miehillä parempaan ja naisilla heikompaan suuntaan. Ennen testauksen suorittamista varmistetaan, että testin suorittamiseen liittyvät olettamukset pitävät paikkansa. Ryhmät ovat melko suuret (molempien n = 30), joten voidaan ajatella, että normaalijakaumaan liittyvä oletus on kunnossa. Tarvittaessa normaalijakaumaoletuksen täyttyminen voidaan tarkistaa normaalisuuden testaukseen tarkoitetulla testillä (esim. Kolmogrov-Smirnovin testi, ks. tarkemmin tietolaari 3, s. 103, normaalijakauma ja normaalisuuden testaus). Varianssien yhtäsuuruusolettamuksen testaamiseen (itse asiassa ryhmien varianssi-kovarianssi-matriisin yhtäsuuruuden testaamiseen) toistettujen mittausten tilanteessa voidaan käyttää Boxin M -testiä. Sen p-arvoksi saadaan 0,948, eli myös varianssien yhtäsuuruusolettamus pitää paikkansa. Tulosteeseen saadaan myös varianssien yhtäsuuruustestit, joissa myös p-arvot ylittävät merkitsevyystason 0,05 (Levenen-testit, joissa testien tulokset tulkitaan aiempien esimerkkien tapaan).

Box's Test of Equality of Covariance Matrices

Box's M	,377
F	,121
df1	3
df2	605520,000
Sig.	,948

Tests the null hypothesis that the observed covariance matrices of the dependent variables are equal across groups.

Levene's Test of Equality of Error Variances

	F	df1	df2	Sig.
Opintomenestys (lukio)	,043	1	58	,837
Opintomenestys (yliopisto)	1,391	1	58	,243

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

Toistettujen mittausten varianssianalyysin tulokset testauksen osalta näyttävät seuraavilta (SPSS-tuloste):

Tests of Between-Subjects Effects

Measure: MEASURE_1
Transformed Variable: Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Intercept	30400,83	1	30400,83	4133,91	,000	,986
SUKUPUOL	17,63	1	17,63	2,40	,127	,040
Error	426,53	58	7,35			

Mauchly's Test of Sphericity

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon Greenhouse-Geisser	Huynh-Feldt	Lower-bound
AIKA	1,000	,000	0	.	1,000	1,000	1,000

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

Tests of Within-Subjects Effects

Measure: Opintomenestys

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
AIKA	Sphericity Assumed	,833	1	,833	,402	,528	,007
	Greenhouse-Geisser	,833	1,000	,833	,402	,528	,007
	Huynh-Feldt	,833	1,000	,833	,402	,528	,007
	Lower-bound	,833	1,000	,833	,402	,528	,007
AIKA * SUKUPUOL	Sphericity Assumed	12,033	1	12,033	5,810	,019	,091
	Greenhouse-Geisser	12,033	1,000	12,033	5,810	,019	,091
	Huynh-Feldt	12,033	1,000	12,033	5,810	,019	,091
	Lower-bound	12,033	1,000	12,033	5,810	,019	,091
Error(AIKA)	Sphericity Assumed	120,133	58	2,071			
	Greenhouse-Geisser	120,133	58,000	2,071			
	Huynh-Feldt	120,133	58,000	2,071			
	Lower-bound	120,133	58,000	2,071			

Ylemmässä taulussa rivillä SUKUPUOL on testattu sukupuolen vaikutusta. Sukupuolten välinen ero ei ole tilastollisesti merkitsevä (eli $p \geq 0{,}05$). Myös efektikoolla eta_p^2 mitattuna sukupuolen vaikutus opintomenestykseen on efektikoolla mitattuna pientä. Tilastolliseen päättelyyn liittyvä tulos raportoitaisiin kyseisellä rivillä olevien arvojen perusteella muodossa $F(1, 58) = 2{,}40$; $p = 0{,}127$; $eta_p^2 = 0{,}04$. Sivun 157 taulukosta voidaan laskea miesten keskiarvoksi 15,5 ja naisten 16,3. Nämä ja niihin liittyvät 95 % luottamusvälit saadaan myös SPSS-tulosteisiin otsikolla *Estimated Marginal Means*. Vaikka siis naisten keskiarvo on korkeampi, ero ei kuitenkaan ole niin suuri, että eroa voitaisiin pitää tilastollisesti merkitsevä.

Alemmassa taulussa on testattu ajan päävaikutusta sekä ajan ja sukupuolen yhdysvaikutusta eli interaktioita. Within-Subjects -taulusta voidaan käyttää ns. univariaatteja testejä (rivit Sphericity Assumed), mikäli Mauchlyn sfäärisyystesti menee läpi (sfäärisyysoletus voidaan hyväksyä, jos testin p-arvo $\geq 0{,}05$). Sfäärisyydellä tarkoitetaan sitä, että mittauskertojen välisissä riippuvuuksissa ei ole merkitsevää vaihtelua seurantajakson aikana. Nyt kun mittauskertoja oli vain kaksi, kaikki eri testiversiot antavat samat tulokset, mutta jos mittauskertoja olisi useampia, testitulokset tulisi valita Mauchlyn sfäärisyystestin perusteella. Vaihtoehtoisissa testeissä on tehty korjauksia vapausastelukuihin, yleisimmin näistä käytetään Greenhouse-Geisser -korjausta.

Rivillä AIKA on testattu aikamuuttujan vaikutusta, eli sitä, onko lukiomenestyksen ja yliopistomenestyksen välinen keskiarvoero tilastollisesti merkitsevä. Sivun 157 taulukosta voidaan laskea opintomenestys lukiossa -muuttujan keskiarvoksi 15,8 ja opintomenestys yliopistossa -muuttujan keskiarvoksi 16,0. Nämä ja niihin liittyvät 95 % luottamusvälit saadaan myös SPSS-tulosteisiin otsikolla *Estimated Marginal Means*. Tämäkään ero ei ole tilastollisesti merkitsevä, sillä $F(1, 58) = 0{,}402$; $p = 0{,}528$ eli $p \geq 0{,}05$). Myös efektikoko on hyvin pieni $eta_p^2 = 0{,}007$.

Alemman taulukon rivillä AIKA*SUKUPUOL on testattu aikamuuttujan ja sukupuolimuuttujan yhdysvaikutusta. Tähän kannattaa kiinnittää erityistä huomiota, koska pyrkimyksenähän on selvittää nimenomaan sitä, millaiseen tulokseen päädytään, jos molempia muuttujia tarkastellaan samanaikaisesti. Yhdysvaikutus on tilastollisesti merkitsevä ja efektikooltaan kohtalainen [$F(1, 58) = 5{,}81$; $p = 0{,}019$; $eta_p^2 = 0{,}091$]. Tämä tulos on helppo ymmärtää kuviota 13 katsomalla: miehillä ja naisilla muutos opintomenestyksessä on erilaista siirryttäessä lukiosta yliopistoon. Miesten opintomenestys paranee ja naisten heikkenee.

SPSS: Toistettujen mittausten varianssianalyysi

1. Löydät analyysin valikkoriviltä *Analyze: General Linear Model → Repeated Measures*
2. *Within-Subject Factor Name* -kohdassa
   - määrittelet mittauskerrat haluamallasi tavalla (voit nimetä esim. mittauskerta, aika, toisto tms.)
   - jos kyse on kahdesta mittauskerrasta, kirjoita kohtaan *Number of Levels* numero 2
   - a*dd*-painike lisää nimeämäsi muuttujan mukaan analyysiin
3. *Define*-painike vie sinut seuraavaan vaiheeseen, jossa valitset analyysiin mukaan tulevat muuttujasi
   - valitse muuttujalistasta varsinaiset toistettuihin mittauksiin liittyvät muuttujat (esimerkissämme 1. mittaus ja 2. mittaus) ja siirrät ne kohtaan *Within-Subjects Variables* tämän kohdan vasemmalla puolella olevalla nuolipainikkeella.
   - vastaavalla tavalla siirrät kategorisen muuttujan (esim. sukupuolimuuttuja) kohtaan *Between-SubjectsFactor(s)* tämän kohdan vasemmalla puolella olevalla nuolipainikkeella
   - hyväksy valinnat klikkaamalla *Continue-painiketta*
4. *Options*-painike
   - valitaan kohta *Descriptive statistics*
   analyysi tulostaa tällöin tekijämuuttujan ryhmien keskiarvot, keskihajonnat ja luottamusvälit; tärkeä valinta, koska näin tulosteisiin tulee näkyviin varsinaisen varianssianalyysin lisäksi vertailtavien ryhmien keskiarvot yms. tärkeät tekijät; tältä pohjalta voit tehdä yleisen ryhmissä esiintyvien erojen tai yhtäläisyyksien

ohjetaulukko jatkuu...

tulkinnan; pelkkä varianssianalyysien pohjalta todettu erojen tilastollinen merkitsevyystaso ei yksinään riitä tuloksen tulkinnaksi
- valitse myös *Homogeneity Tests* testaa tekijämuuttujien eri ryhmien jakaumat vastemuuttujan suhteen (on siis varianssien yhtäsuuruus -testi)
- hyväksy valinnat klikkaamalla *Continue*-painiketta

5. *Plots*-valikko
   - jos haluat tulosteeseen mukaan myös tekijöiden yhdysvaikutuksia kuvaavan viivadiagrammin (tällainen on monesti hyödyllinen tulosten tulkinnassa), vie täällä Horizontal Axis -kohtaan eli kuvion vaaka-akselille nimeämäsi mittauskertamuuttuja ja Separate Lines -kohtaan mahdollinen kategorinen tekijä eli esimerkissämme sukupuolimuuttuja
   - lisää yllä määritelty kuvio *Add*-painikkeella ja sitten koko valinta *Continue*-painikkeella

6. Toteuta toistettujen mittausten varianssianalyysi yllä määrittelemilläsi valinnoilla *OK* -painiketta painamalla, ja saat analyysin tulosteet tulosteikkunaan.

5.5 Epäparametrinen Kruskal-Wallisin testi

Mikäli varianssianalyysin kriteerit eivät täyty, voidaan käyttää sen epäparametrista vastinetta, Kruskal-Wallisin testiä. Se soveltuu kolmen tai useamman ryhmän välisten erojen vertailuun. Kruskal-Wallisin testissä edellytetään vertailtavien ryhmien olevan riippumattomia toisistaan ja sitä, että muuttujien mittaustaso on vähintään järjestysasteikollinen. Kruskal-Wallisin testi perustuu järjestyslukujen käyttöön; siinä vertaillaan keskimääräisiä järjestyslukuja ja niiden perusteella jakaumien muotoa. Tässä testin periaatteesta esitetään pääpiirteet, katso tarkempi kuvaus esimerkiksi Metsämuurosen (2009, 1115–1120) teoksesta.

Edellä olleesta tutkimusasetelmasta, jossa selvitettiin, eroaako opintomenestys kolmen eri viriketaustaluokan suhteen, SPSS antaa seuraavanlaisen tulosteen:

Ranks

	Viriketausta	N	Mean Rank
Opintomenestys lukiossa	matala	18	10,39
	keskitaso	26	35,88
	korkea	16	44,38
	Total	60	

Test Statistics[a,b]

	Opintomenestys lukiossa
Chi-Square	37,428
df	2
Asymp. Sig.	,000

a. Kruskal Wallis Test
b. Grouping Variable: Viriketausta

Kruskal-Wallisin testi järjestää muuttujan arvot suuruusjärjestykseen ja laskee näin saadut järjestysluvut viriketaustoittain yhteen (Sum of Ranks) ja jakaa saadun summan kummankin ryhmän koolla, jolloin saadaan järjestyslukujen keskiarvot (Mean Rank) ryhmittäin. Alimman viriketaustan ryhmässä keskimääräinen sijaluku on 10,39, keskitason 35,88 ja ylimmän viriketaustan ryhmässä 44,38. Tämä tarkoittaisi sitä, että matalan viriketaustan omaavien opiskelijoiden opintomenestys lukiossa oli heikoin ja korkean viriketason omaavien puolestaan paras. Tuloksen eli ryhmäerojen raportoinnissa käytetään yleensä Mean Rank -lukujen sijaan ryhmämediaaneja. Niitä ei SPSS:ssä saada suoraan Kruskal-Wallis -testin tulosteesta, vaan ne on tulostettava erikseen tunnuslukujen taulukointeihin liittyvillä proseduureilla. Alemmasta taulukosta nähdään, että tulos oli tilastollisesti merkitsevä ($p=0{,}000$). Ryhmäerojen voimakkuus kannattaa laskea sopivaa efektikokoon suuretta käyttäen. Kruskall-Wallis -testin yhteydessä tällainen on z-testisuureen avulla muodostettava tunnusluku r, joka saadaan laskemalla r = z/neliöjuuri(N). Tässä tarvittava suure z ei sisälly yllä olevaan tulosteeseen, vaan se jouduttaisiin etsimään standardoidun normaalijakauman kertymäfunktion taulukoidusta arvoista (tai vaikkapa excel-funktiolla) siten, että z-arvo rajaisi jakaumasta Kruskall-Wallis -testin p-arvoa vastaavan osan. Toinen, ehkä hyödyllisempikin tapa arvioida efektikokoa, olisi kokonaisefektin sijaan toteuttaa ensin vertailtavien osaryhmien parittaiset vertailut Mann-Whitneyn U-testin avulla ja arvioida niistä erikseen kullekin parittaiselle vertailulle efektikoot. Tällaista efektikoon arviointia U-testin yhteydessä on käsitelty edellä luvussa 4.2.1. Mikäli tulos kuitenkin raportoitaisiin merkitsevyystestin ja kokonaisefektin koon avulla, niin tilastolliseen päättelyyn liittyvät numeeriset tulokset raportoitaisiin muodossa: $H(2) = 37{,}43$; $p < 0{,}001$; $r = 0{,}75$ (ks. p-arvon tulkinnasta tarkemmin luku 1.6.1). Tätä toki olisi tärkeätä, kuten aina tilastollisten tulosten raportoinnissa tulisi tehdä, täydentää ja pohtia sisällöllisellä käytännön tulkinnalla, jossa saatua tulosta avattaisiin käytännössä ja tuloksen käytännön merkittävyyttä arvioitaisiin.

SPSS: Kruskal-Wallisin testi

1. Löydät analyysin ylhäältä valikkoriviltä Analyze → Nonparametric Tests → Legacy Dialogs → k Independent Samples
   - siirrä valitsemasi numeerinen muuttuja (tai useita, jos haluat monta testiä samalla kertaa) muuttujaluettelon oikealla puolella olevalla ylemmällä nuolipainikkeella kohtaan *Test Variable List*
   - siirrä vastaavalla tavalla valitsemasi kategorinen muuttuja alemmalla nuolipainikkeella kohtaan *Grouping Variable*
   - varmista, että *Test Type* -kohdassa on valittuna vaihtoehto *Kruskall-Wallis H* ja paina *Define Range* -painiketta. Kirjoita kenttiin *Minimum* ja *Maximum* ryhmämuuttujan numeroarvot (kategorisen muuttujan pienin ja suurin kategoria)
   - hyväksy valinnat klikkaamalla *Continue*-painiketta
2. Toteuta nyt Kruskal-Wallis -analyysi painamalla *OK*-painiketta ja saat analyysien tulosteet tulosteikkunaan
3. Tässä Kruskal-Wallis -analyysin toteutuksessa ei saada suoraan eri ryhmien välille parivertailuja eli post hoc -testejä. Tarvittaessa parittaiset vertailut on tehtävä erillisillä U-testeillä kaikille eri ryhmäkombinaatioille. Mann-Whitneyn U-testin toteutus on kuvattu aiemmin luvussa 4.2.1 (s. 135–136). Lopuksi U-testien tuottamille p-arvoille on vielä tehtävä ns. bonferroni-korjaus, jossa kaikki p-arvot vielä kerrotaan vertailujen lukumäärällä (ks. esim. Nummenmaa 2011, 207)

Huom. Toinen tapa toteuttaa Kruskal-Wallis -testi on käyttää valikkoa *Analyze → Nonparametric Tests → Independent Samples*. Tällä toiminnolla muuttujien ja testin valinta tehdään valikossa avautuvilta välilehdiltä. Välilehdellä *Fields* pääset valitsemaan numeerisen vastemuuttujasi siirtämällä sen keskellä olevalla nuolella oikealla olevaan tilaan *Test Fields*. Vastaavasti siirrät kategorisen muuttujasi alemmalla nuolella kohtaan *Groups*. Toiminto toteuttaa muuttujatietojen perusteella automaattisesti oikean testin, kun painat alareunan *Run*-painiketta. Voit myös itse käydä valitsemassa Kruskall-Wallisin testin ja halutessasi myös muita lisävalintoja siirtymällä välilehdelle *Settings.* Siellä voit valita vaihtoehdon *Customize tests* ja sen jälkeen testin klikkaamalla valikon oikealta puoliskolta kohdan *Kruskall-Wallis 1-way anova (k samples)*. Samalla kannattaa valita myös testin alla oleva valinta *Multiple comparisons: all pairwise*, jolloin saat tulosten yhteyteen myös taulukon ryhmien välisistä parivertauluista. Alareunan *Run*-painikkeella toteutat valitun testin. Huomaa, että testin toteutus edellyttää, että aineiston muuttujien mitta-asteikot on määritelty oikein havaintomatriisin alareunan *Variable View* -näkymässä (*Measure: Nominal/ Ordinal/ Scale*). Tällä toteutustavalla tulostaulukoiden muoto poikkeaa yllä kuvatun perinteisen toteutustavan taulukoista. Esimerkiksi parivertaillut saa näkyviin vasta, kun kaksoisklikkaat tulosikkunassa taulukon muokkaustilaan. Proseduurin tulostustapa on jokseenkin erikoinen!

6. Tutkittavien muuttujien välisen yhteyden eli riippuvuuden analysointi

Tilastollisessa tutkimuksessa ollaan yleisesti kiinnostuneita eri tekijöiden (muuttujien) välisistä yhteyksistä, jotka vastaavat kullekin tutkimukselle asetettuihin kysymyksiin. Tästähän yhteiskuntatieteellisissä ja kasvatustieteellisissä tutkimuksissa usein on juuri kyse. Tällaista tietoa tarvitaan paitsi tutkittavan ilmiön tieteellisen tulkitsemisen tueksi, myös yhteiskunnallisen päätöksenteon perustaksi. Voidaan olla kiinnostuneita esimerkiksi siitä, mitkä tekijät selittävät eri sukupuolten koululle antamia merkityseroja tai vaikka siitä, selittääkö se, mitä lukiota ylioppilaskokelaat ovat käyneet, kuinka he ovat menestyneet ylioppilaskirjoituksissa. Vastaavia esimerkkejä tutkimuskysymyksistä voisi listata loputtomasti. Tässä luvussa käsitellään tähän liittyviä perusanalyyseja, kuten ristiintaulukointia ja korrelaatioanalyysia. Ensiksi mainittu soveltuu luokitteluasteikollisten muuttujien välisten yhteyksien analysointiin. Lisäksi, kuten aluksi esiteltävä esimerkki osoittaa, ristintaulukoinnilla saatuja frekvenssi- ja prosenttiosuusjakaumia käytetään myös aineiston yleiseen kuvailemiseen, jota on käsitelty jo aiemmissa luvuissa. Tämän jälkeen tarkastellaan korrelaatioanalyysia, joka soveltuu ennen kaikkea kahden määrällisen muuttujan välisen yhteyden eli riippuvuuden voimakkuuden yleiseen tarkasteluun. Korrelaatiotarkastelun yleisperiaate on tärkeä ymmärtää senkin vuoksi, koska monet monimuuttujamenetelmät perustuvat muuttujien välisiin korrelaatiorakenteisiin. Seuraavissa luvuissa tarkastelu ja laajennetaan tilanteisiin, joissa analyyseissä huomioidaan samanaikaisesti useampia muuttujia ja niiden välisiä yhteyksiä.

6.1 Ristiintaulukointi

Ristiintaulukointi on yksi käytetyimmistä ja yksinkertaisimmista menetelmistä, ja ehkä juuri siksi se on myös vähän aliarvostettu kasvatustieteellisissä ja yhteiskuntatieteellisissä tutkimuksissa. Kuitenkin sen yksinkertaisuus tekee siitä selkeän, tehokkaan ja siten myös suositeltavan analyysimenetelmän, jonka avulla on helppo hahmottaa tutkimusaineiston muuttujien suhteita, niiden luonnetta ja jatkoanalysoinnin tarpeita. Ristiintaulukointi on tarkoitettu kategoristen muuttujien analysointiin ja ristiinluokitteluun. Sukupuoli-, ikä-, kotitausta-, erilaiset asenne- ja preferenssimuuttujat ovat tyypillisiä ristiintaulukoitavia muuttujia. Ristiintaulukoinnissa aineisto esitetään frekvensseinä ja prosenttiosuuksina. Kaiken kaikkiaan ristiintaulukointi, yhdistettynä esimerkiksi riippuvuuden merkitsevyyttä testaavaan khiin neliö -testiin sekä riippuvuuden voimakkuutta mittaavan Cramerin V suureeseen, sopii monen tutkimusongelman käsittelyyn.

Tyypillisessä ristiintaulukkoanalyysissä tarkastelussa on vain kaksi muuttujaa, joiden välistä yhteyttä halutaan selvittää. Analyysiä voidaan laajentaa myös niin, että lisäksi siinä huomioidaan myös kolmas tekijä (muuttuja). Tällaista niin sanottua elaboraatio-menettelyä hyödyntäen nähdään, miten kahden muuttujan välinen yhteys ilmenee kolmannen muuttujan eri luokissa (ks. taulukko 11, s. 180). Vastaavaan tapaan analyysiin voisi ottaa mukaan tarkasteluun useampienkin muuttujien jakaumien yhteyksiä, joskin silloin tulostaulukot alkavat muodostua hankalasti hahmotettaviksi ja vaikeiksi tulkita. Tällaisessa useamman kategorisen muuttujan yhteyksien analyysissa kannattaakin siirtyä mieluummin log-lineaaristen mallien käyttöön. Tätä menetelmää ei käsitellä kuitenkaan tässä kirjassa, mutta asiasta kiinnostunut lukija voi tutustua aiheeseen muista lähteistä (mm. Gröönroos 2011, 163–171). Ristiintaulukoinnin ja log-lineaaristen analyysien vertailua on käsitelty myös esimerkiksi Toivosen (2005) artikkelissa.

Kerlingerin (1981, 160–161) mukaan ei ole mitään yleisesti hyväksyttyjä kriteereitä ristiintaulukoinnin laatimiseksi. Koska kyseessä on ristiinositus (cross partitions), on käytettävien kategorioiden kuitenkin nojauduttava sivulla 74 esitettyihin luokkien muodostamiskriteereihin, eli käytettävien kategorioiden tulee perustua tutkimusasetelmaan ja esitettyihin hypoteeseihin, niiden tulee olla toisistaan riippumattomia ja toisensa poissulkevia sekä niiden tulee olla tyhjentäviä. Lisäksi käytettävien luokitusten tulee muodostaa yhtenäinen kokonaisuus ja luokittelun tulee olla johdettu yhdestä luokitteluperiaatteesta. Vastaavasti tulkittaessa ristiintaulukoiden näyttämiä suhteita tai khiin neliö -testin tuloksia tulee muistaa, ettei esimerkiksi kahden muuttujan välillä havaittavia riippuvuussuhteita tule tulkita kausaalisiksi, eli syysuhdepäätelmiä tulee varoa, jos ei ole esittää laajempaa evidenssiä tulkinnoilleen

(tällöin tutkimusasetelma on usein hypoteesien testausmalli); paremminkin saadut suhteet ilmaisevat jotakin vertailtavien muuttujien välisen suhteen luonteesta.

6.1.1 Khiin neliö -testi ja eräitä yhteyden voimakkuutta kuvaavia suureita

Khiin neliö -testi (χ^2, Pearsonin khiin neliö -testi) on jakauman (frekvenssien) yhteensopivuustesti (saatu frekvenssijakauma vs. odotettu jakauma). Lisäksi testiä käytetään riippumattomuustestinä etsittäessä vastausta kysymykseen, ovatko tarkasteltavat muuttujat toisistaan riippuvia, eli onko muuttujien välillä yhteyttä vai ei.

Testissä tutkitaan ristiintaulukon havaittujen solufrekvenssien ja odotettujen solufrekvenssien eli teoreettisten solufrekvenssien välisen eron tilastollista merkitsevyyttä. Teoreettiset solufrekvenssit ovat ne frekvenssit, jotka saataisiin, jos muuttujien jakaumat olisivat samat tai jos muuttujat olisivat riippumattomia toisistaan. Mitä suurempi havaittujen ja teoreettisten solufrekvenssien erotus on, sitä suuremmaksi tulee testisuureen χ^2 arvo ja sitä pienemmäksi siihen liittyvä p-arvo. Testin tulosteisiin liitetään toisinaan mukaan myös niin sanottu vapausasteluku (df = degrees of freedom), joka liittyy taulukon solujen määrään. Jos ristiintaulukon toisella muuttujalla on esimerkiksi kaksi mahdollista arvoa ja toisella muuttujalla kolme, kyseisen taulun koko on 2*3 = 6 solua. Tällaisen taulukon vapausasteluku ei kuitenkaan ole kuusi, vaan kaksi (2–1)*(3–1), sillä tauluun voidaan sijoittaa vain kaksi frekvenssiä vapaasti (muuttujien reunajakaumien ollessa tunnettuja).

Lähtöoletuksena khiin neliö -testin käytölle sen lisäksi, että otos on satunnaisotos, on se, etteivät taulukon solujen odotusarvot eli teoreettiset frekvenssit saisi olla liian pieniä. Esimerkiksi Siegel (1956, 46–47, 110, 178–179) antaa seuraavanlaiset kriteerit testin käytölle:

- muuttujat voivat olla nominaali(luokittelu)asteikkoisia
- korkeintaan 20 % teoreettisista solufrekvensseistä saa olla alle 5 ja jokaisen teoreettisen solufrekvenssin on oltava suurempi kuin 1

Mikäli khiin neliö -testin käyttöön liittyvät oletukset eivät ole voimassa, vaihtoehtoinen tapa toteuttaa riippumattomuustesti on käyttää Fisherin tarkkaa testiä (Fisher’s exact test). Tässä merkitsevyysarvo saadaan asymptoottisten todennäköisyysjakaumien käytön sijaan tutkimalla kaikki mahdolliset ristiintaulukon vastaavat solukombinaatiot. Proseduuri on laskennallisesti hyvin raskas, mutta nykyisillä tietokoneilla tulos saadaan yleensä muutamassa sekunnissa. SPSS tulostaa myös tämän testin p-arvoineen automaattisesti khiin neliö -testin yhteyteen, jos ristiintaulukon yksikin teoreettinen eli odotettu solufrekvenssi on alle viiden. Tarkan testin arvo tulostuu samaan taulukkoon kuin varsinainen khiin neliö -testi ja tuloksena saatava p-arvo (sig.) tulkitaan muiden testien tapaan. Käytännössä lopputulos on monesti molemmilla testeillä hyvin samankaltainen. Fisherin tarkan testin tulos raportoidaan tekstiin siten, että siitä esitetään p-arvo, jonka jälkeen teksti *Fisherin tarkka testi*.

Seuraavassa luvussa 6.1.2. käydään monipuolisesti läpi esimerkkejä χ^2-testin käytöstä, joten tässä riittänee lyhyt tulkintaesimerkki. Voidaan esimerkiksi tutkia, onko viriketaustalla ja sukupuolella yhteyttä keskenään. Nollahypoteesi on nyt muotoa H_0: viriketaustalla ja sukupuolimuuttujalla ei ole yhteyttä keskenään ja vastaavasti vaihtoehtoisen hypoteesin H_1: viriketaustalla ja sukupuolella on yhteyttä keskenään. Tähän liittyvät ajot on esitetty taulukossa 10, josta nähdään, että khiin arvoksi tulee $\chi^2 = 0{,}38$ ja siihen liittyväksi p:n arvoksi $p = 0{,}830$. Tämän perusteella voidaan todeta, ettei sukupuoli- ja viriketaustamuuttujilla ole tilastollisesti merkitsevää yhteyttä (joka olisikin käytännössä ollut aika outoa). Edellä esitettyjen hypoteesilausekkeiden avulla esitettynä voidaan todeta nollahypoteesin saaneen vahvistusta, joten se hyväksyttäisiin. Testituloksen numeeriset arvot raportoitaisiin tutkimusraporttiin sanallisen tulkinnan perään muodossa $\chi^2(2) = 0{,}38$, $p = 0{,}830$.

Tilasto-ohjelmat antavat p-arvon hyvinkin tarkkaan. Esityksissä riittävän tarkkuuden antaa kolmen desimaalin tarkkuus. Nykyisin suositeltava käytäntö on raportoida tarkka ohjelman tuottama p-arvo, eikä siis viittausta yleisiin merkitsevyystasojen rajoihin, jotka on kuvattu sivulla 42. Mikäli kuitenkin p-arvo on hyvin pieni niin, että ohjelma tulostaa sen muodossa $p = 0{,}000$, tämä on hyvä esittää tekstissä muodossa $p < 0{,}001$ tai pyöristyssääntöihin liittyvässä muodossa $p < 0{,}0005$. Tällä ilmaistaan sitä, että p-arvo on matala mutta ei kuitenkaan nolla. Taulukon avulla ja tekstin merkitsevyystasoviittausten avulla lukija voi päätellä, etteivät sukupuolten viriketaustat eroa toisistaan tilastollisesti merkitsevästi.

Khiin neliö -testin avulla voidaan päätellä muuttujien välisen yhteyden tilastollista merkitsevyyttä, mutta yhteyden voimakkuudesta tämäkään testi ei kerro. Sen vuoksi ristiintaulukoinnin ja khiin neliö -testin yhteyteen on hyvä laskea myös voimakkuutta mittaavia tunnuslukuja, toisin sanoen efektikokoa mittaavia suureita. SPSS:ssä näitä ovat esimerkiksi Cramérin V, kontingenssikerroin C sekä fi (phi) -kerroin. Cohenin Kappa -kerrointa käytetään kahden eri arvioitsijan suorittaman luokittelun yksimielisyyden mittaamiseen. Katso näiden tunnuslukujen tulkinnan raja-arvoista asetelma 5:stä (s. 49). Seuraavissa ristiintaulukkoesimerkeissä on käytetty norjalaisen Harald Cramérin kehittämää V-kerrointa. Cramérin V-arvot sijoittuvat välille 0,00–1,00 (kyseiset arvot mukaan lukien). Jos arvo on nolla, tarkoittaa se, ettei kategoristen muuttujien välillä ole lainkaan yhteyttä, kun taas arvo yksi viittaa muuttujien täydelliseen yhteyteen. Käyttäytymistieteessä yleisiä (likimääräisiä) kriteeriarvoja Cramérin V:n kohdalla ovat

V-arvo 0,10 tai yli, muuttujien välillä on yhteyttä, joka on voimakkuudeltaan heikkoa

V-arvo 0,30 tai yli, yhteys voimakkuudeltaan kohtalainen

V-arvo 0,50 tai yli, yhteys voimakkuudeltaan voimakas (Cohen 1988).

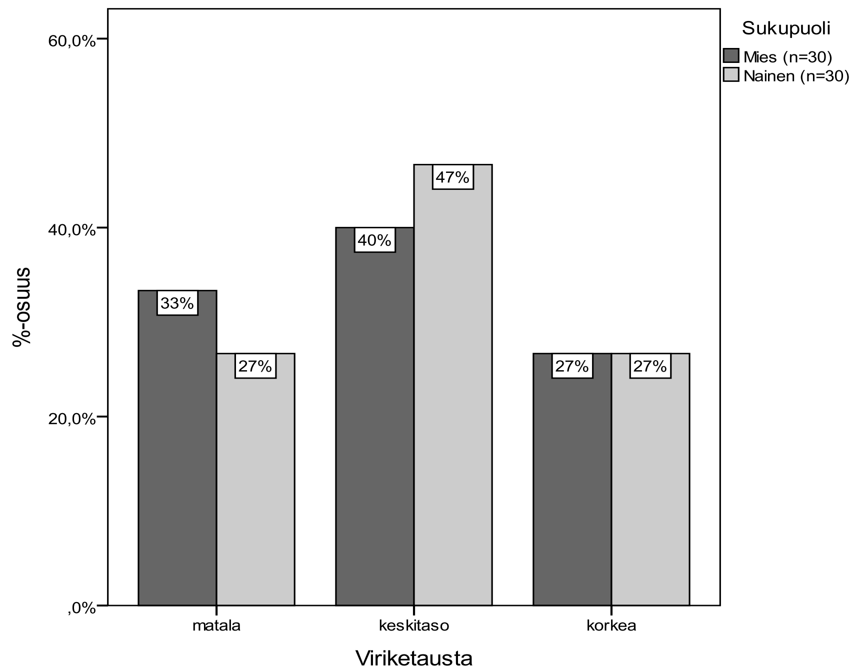
Arvot, jotka jäävät alle 0,10 kuvaavat olematonta tai hyvin heikkoa yhteyttä. Näin tunnusluvun avulla voidaan täydentää khiin neliö -testin antamaa informaatiota. Kaiken kaikkiaan khiin neliö -testi ja Cramérin V ovat useassa yhteydessä käyttökelpoiset. Yhteiskuntatieteissä ja kasvatustieteissä niitä käytetäänkin runsaasti. Toisaalta on muistettava, että khiin neliö -testi testaa ristiintaulukointia kokonaisuudessaan. Mikäli halutaan tarkastella tarkemmin taulukon jotakin yksityiskohtaa, silloin voidaan analysoida soluihin liittyviä jäännösarvoja ja riveittäin toteutettavia z-testejä. Näitä käsitellään tarkemmin sivulla 177 olevassa esimerkissä.

6.1.2 Ristiintaulukointiesimerkkejä

Ristiintaulukoinnissa on yleistä, että muuttujien arvot luokitellaan. Kategoristen muuttujien kohdalla asia on yksinkertainen, näin esimerkiksi sukupuolen tai asuinpaikan suhteen. Esimerkkihavaintomatriisissa ei ole kovin monta puhtaasti kategorista muuttujaa, joita voitaisiin tai olisi mielekästä tutkia keskenään ristiintaulukoinnilla, jos emme luokittele jatkuvia muuttujia siten kuin edellä esitettiin. Käytännössä tällainen muuttujien luokittelu ei olisi välttämätöntä.

Tarkastellaan aluksi esimerkin vuoksi miesten ja naisten viriketaustoja ja niiden eroja, joiden tarkastelu ristiintaulukoinnin avulla käytännössä on melko turhanaikaista, mutta tehdään se nyt tämän menetelmän periaatteen esittämiseksi. Toki tämän avulla saadaan lisää syvyyttä esimerkiksi sukupuolten koulumenestyksen mahdollisten erojen taustatekijöihin. Yleisesti voidaan todeta ensinnä se, että nuoret naiset näyttävät menestyvän varsinkin lukio-opinnoissaan paremmin kuin nuoret miehet. Mutta, olisikohan koulumenestyksen taustalla kuitenkin muitakin vaikuttavia tekijöitä kun sukupuoli? Tutkijan on tärkeä esittää itselleen koko tutkimusprosessin aikana tällaisia lisäkysymyksiä, kuten kirjan alussa tätä jo painotettiinkin. Ei ole kaukaa haettu ajatus, että lukiolaisten kotien viriketaustaerot ovat yhteydessä myös heidän opintomenestykseensä koulussa. Laajemmin voitaisiin peräänkuuluttaa muun muassa sukupuolten kasvatukseen ja harrastuksiin liittyvien erojen yhteyttä heidän lukioaikaiseen opintomenestykseen. Toki sukupuolten erilaiset biologisetkin tekijät voivat vaikuttaa siihen ja moni muukin tekijä. Seuraavassa esitetään vähän pohjaa tämän kaltaisille pohdinnoille (muistetaan kuitenkin, että aineistoesimerkin pohjalta ei voida tehdä reaalimaailmaa koskevia empiirisiä yleistyksiä).

Pylväsdiagrammilla tarkasteltuna sukupuolten viriketausta näyttää seuraavalta:



KUVIO 14. Miesten ja naisten viriketaustat.

Kuten kuviosta nähdään, graafiset esitykset eri tutkimuskysymysten suhteen ovat usein hyvin informatiivisia. Kuviosta voidaan nähdä sen, että naisten ja miesten viriketausta on osin yhtenevä (korkea viriketausta on 27 prosentilla sekä naisista että miehistä). Kahden muun arvon (matala ja keskitaso) suhteen on jonkin verran poikkeamia. Tosin poikkeamat ovat melko pieniä, joten kokonaisuudessa muuttujien välinen yhteys ei liene tilastollisesti merkitsevä. Tutkitaan tätä vähän lähemmin samalla, kun tarkastellaan muuttujien jakaumia ristiintaulukoinnin avulla.

TAULUKKO 11. Miesten ja naisten viriketaustat.

			Viriketausta			
			matala	**keskitaso**	**korkea**	**Yhteensä**
Sukupuoli	mies	n	10	12	8	30
		%	33,3	40,0	26,7	100,0
	nainen	n	8	14	8	30
		%	26,7	46,7	26,7	100,0
Yhteensä		n	18	26	16	60
		%	30,0	43,3	26,7	100,0

χ^2 = 0,38; df = 2; p = 0,830, Cramérin V = 0,08

Yllä oleva taulukko on lähes SPSS:n alkuperäisen tulosteen mukainen (taulukkomallina SPSS:n '*Academic*'), vain englanninkieliset termit on korvattu suomenkielisillä.

Esitettäköön tässä yhteydessä muutama taulukon laatimisen perusohje:

1. Taulukon alalaitaan alaviitteeksi on lisätty khiin neliö -testiin liittyvät arvot sekä Cramérin V -suure. Khiin neliö -testissä teoreettisten solufrekvenssien (Expected Count) tulisi olla riittävän suuria (kirjallisuudessa useimmin esiintyy kriteerinä, että teoreettisista solufrekvensseistä saa olla viittä pienempiä enintään 20 %). Siten itselle kannattaa tulostaa myös nämä, mutta lopulliseen raporttiin niitä ei useinkaan laiteta. Tätä kriteeriä, joka on khiin neliö -testin p-arvon tarkoituksenmukaisen tulkinnan ehto, on usein vaikea täyttää pienissä aineistoissa. Tämä ongelma voidaan usein välttää sillä, että vähennetään muuttujien kategorioita luokkia yhdistelemällä, tehdään esimerkiksi kolmiluokkaisesta kaksiluokkainen ja niin edelleen. Esimerkkiaineistosta voitaisiin viriketaustamuuttuja muuttaa kaksijakoiseksi arvoinaan matala ja korkea. Nyt tätä muutosta ei ole tarvinnut tehdä, koska teoreettiset solufrekvenssit (joita emme ole esittäneet yllä olevassa taulukossa) olivat taulukon kaikissa soluissa yli viiden. Jos solujen arvot jäävät kovin pieniksi tai meillä on ylipäätään pieni otos, tulisi arvioida myös sitä, olisiko aineisto tulkittavissa jollakin muulla tavalla kuin tilastollisesti, esimerkiksi kvalitatiivisin metodein. Käytännössä

joudutaan usein sellaisen tilanteen eteen (vaikka tutkimusasetelmaa pohdittaessa tällaisetkin asiat tulisi ottaa huomioon), ettei kaikkia testien kriteerejä pystytä täyttämään, mutta silti on tarkoituksenmukaista käyttää joitakin testejä. Tällöin tulisi tulkintoja tehtäessä olla varovainen ja tulkintaongelmasta pitäisi tehdä lukijakin tietoiseksi.

2. Taulukon riviväli kannattaa pitää kohtuullisen pienenä, mutta luettavana. Usein on tarkoituksenmukaista pienentää taulukon kirjasinkokoakin (fonttia). Näin saadaan sijoitettua laajemmankin taulukon yhdelle sivulle. Näin saadaan taulukko myös ulkoasultaan paremmaksi. Samalla taulukko erottuu tekstistä paremmin, kun kirjasinkoko on erilainen.

3. Nimetään taulukko. Nimen tulisi olla mahdollisimman selkeä, ytimekäs ja informatiivinen, vältä monen rivin otsikoita.

4. Taulukoissa esiintyy usein rivi-, sarake- tai kokonaisprosentteja, joiden tulisi määritelmän mukaan summautua täsmälleen lukuun 100 prosenttia. Jos katsotaan esimerkkitaulukkoa, nähdään, että naisia koskevien riviprosenttien summaksi tuleekin 100,1. Kyse ei ole mistään laskuvirheestä, vaan lukuja ei kerta kaikkiaan aina saada summautumaan tasan 100 prosenttiin. Suositeltava menettely tällaisissa tilanteissa on seuraava. Käytä yhteensä-kohdassa aina lukua 100 %, ei siis lukua 99,9 % tai 100,1 %. Jotta summaksi tulisi täsmälleen 100 %, voit tehdä tahallisen pyöristysvirheen sen summaan kuuluvan arvon kohdalla, johon pyöristysvirheen vaikutus on suhteellisesti katsoen pienin. Esimerkkiaineistossa voisit naisten kohdalla pienentää arvon 46,7 % arvoon 46,6 prosenttia.

Miten taulukkoa sitten tulisi tulkita:

1. 1. Yleensäkin ristiintaulukointia tulkittaessa tulee muistaa, ettei ole syytä mennä syysuhde-selityksiin, varsinkaan jos tulkinnan tueksi ei ole esittää muuta empiiristä evidenssiä tai teoriaa. Tällainen selittäminen ei siis ole paikallaan esimerkkitilanteessa, jossa kumpaakaan muuttujaa ei voi ajatella selitettäväksi muuttujaksi. Taulukosta ilmenevä suhde ilmaisee sukupuolen ja viriketaustan välistä yhteyttä.

2. Ristiintaulukointia tulkittaessa käytetään yleisesti prosenttiosuuksia, koska usein eri ryhmien otoskoot eivät ole yhtä suuria. Siten frek-

venssien vertailu ei ole tarkoituksenmukaista (tulkinta vääristyisi). Prosenttiosuuksien käyttökelpoisuus perustuu siihen, että aineistojen jakaumia käsitellään kunkin ryhmän suhteellisina osuuksina. Esimerkiksi miehistä noin 33 prosentilla on matala, 40 prosentilla keskitasoinen ja noin 27 prosentilla korkea viriketausta. Näitä prosenttiosuuksia voidaan verrata muiden ryhmien, tässä tapauksessa naisten, vastaaviin lukuihin. Vertailu osoittaa, ettei sukupuolten jakaumissa viriketaustamuuttujan suhteen ole kovinkaan suuria eroja (suurin solukohtainen ero on noin seitsemän prosenttiyksikköä).

3. Taulukkoon on lisätty khiin neliö -testi ja Cramerin V-kertoimen arvo (näitä tarkastellaan tarkemmin luvussa 6.1.1). Näiden perusteella voidaan tehdä päätelmiä muuttujien välisistä suhteista ja Cramerin V-kertoimen avulla myös yhteyden voimakkuudesta. Khiin neliö -testin perusteella voidaan päätellä ensinnäkin, ettei sukupuolen ja viriketaustamuuttujan välillä ole tilastollisesti merkitsevää yhteyttä (p = 0,830). Tämä on tärkeä tieto esimerkiksi jatkoanalyysejä varten: jos jatkossa todetaan esimerkiksi sukupuolen ja opintomenestyksen välillä olevan yhteyttä, voidaan tämän analyysin perusteella päätellä, etteivät ainakaan naisten ja miesten viriketaustaerot ole merkittävästi vaikuttamassa tähän tulokseen. Lisäksi riippuvuuden voimakkuuden tulkinnassa voidaan käyttää suuretta Cramerin V, joka kuvaa efektikokoa. Tässä sen lukuarvon perusteella voidaan tulkita, että asioiden välinen yhteys on voimakkuudeltaan varsin heikkoa, lähes olematonta: lukuarvo V = 0,08 asettuu suureen vaihteluvälillä [0,1] aivan alkupäähän, joka tarkoittaa heikkoa yhteyttä. V:n lukuarvoille on kirjallisuudessa esitetty raja-arvoiksi 0,10 - 0,30 - 0,50: heikko - kohtalainen - voimakas yhteys. Toisin sanoen vasta V = 0,30 suuruiset arvot voidaan tulkita kuvaavan kohtalaista riippuvuutta ja samoin vasta kun V = 0,50 riippuvuus voidaan tulkita voimakkaaksi. (Ks. Cohen 1988.) Suureen V tulkintaa esiteltiin tarkemmin sivulla 169.

Kuten yllä olevasta esimerkistä nähdään, ristiintaulukointi on kätevä ja selkeä analyysimenetelmä, varsinkin kun siihen voidaan yhdistää lisäanalyyseinä khiin neliö -testi ja efektikokoa mittaava Cramerin V -kerroin. Seuraavassa esimerkissä katsotaan sitä, miten ristiintaulukointi testeineen saadaan toteutettua SPSS:ssä.

SPSS: Kahden kategorisen muuttujan ristiintaulukointi

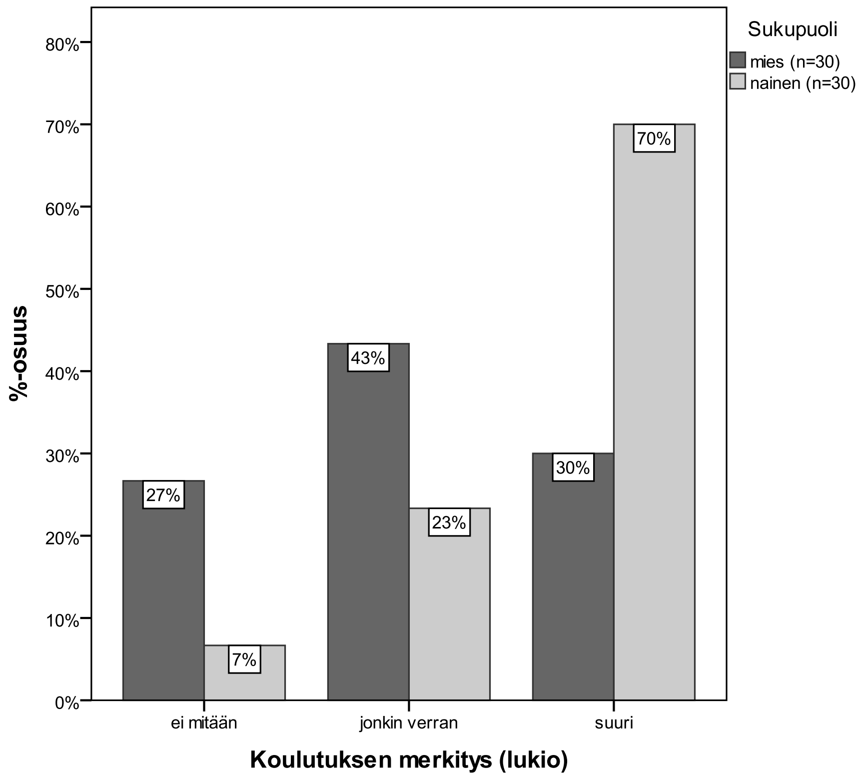
1. Löydät analyysin valikkoriviltä *Analyze: Descriptive Statistics → Crosstabs*.
2. Valitaan aineiston muuttujat, joille ristiintaulukointi halutaan muodostaa
   - siirrä se valitsemasi kategorinen muuttuja, jonka arvot haluat ristiintaulukossa riveille, muuttujaluettelon oikealla puolella olevalla nuolipainikkeella kohtaan *Row(s)*. Esimerkissämme valitsimme rivimuuttujaksi sukupuolen, jolloin miesten ja naisten tiedot asettuvat omille riveilleen vastaavalla tavalla siirrät kohtaan *Column(s)* sen kategorisen muuttujan, jonka arvot haluat ristiintaulukossa tulevan sarakkeille. Esimerkissämme sarakemuuttujana on viriketausta
3. *Statistics*-painike
   - valitaan tulosteisiin riippumattomuustesti χ^2-riippumattomuustesti (*Chi-square*) ja yhteyden riippuvuuden voimakkuutta mittaava Cramérin V:n (*Phi and Cramér´s V*) mikäli epäilet, että χ^2-testin oletukset eivät toteudu (taulukossasi on liian pieniä odotettuja frekvenssejä), voit lisäksi valita erilliseltä *Exact* -painikkeelta kohdan *Exact*, jolla saat tulosteisiisi myös Fisherin tarkan testin, jota siis voidaan käyttää silloin kun χ^2-testin oletukset eivät toteudu
   - hyväksy valinnat klikkaamalla *Continue -painiketta*
4. *Cells*-painike
   - tässä voit määritellä sen, mitä erilaisia lisätietoja pelkkien frekvenssien lisäksi ristiintaulukon soluihin tulostuu mukaan
   - jos esimerkiksi haluat, että taulukon prosenttiluvut summautuvat riveittäin arvoon 100 %, silloin valitset *Percentages*-kohdan vaihtoehdon *Row*
   - jos taas haluat tulostukseen mukaan myös teoreettiset eli odotetut solufrekvenssit, valitse *Count*-kohdan *Expected*-vaihtoehto
   - lisäksi taulukkoon voit tarvittaessa tulostaa myös ns. standardoidut jäännökset ja riveittäin lasketut z-testit, joiden perusteella voidaan tulkita mahdollisen riippuvuuden luonnetta (katso myöhemmin tuleva esimerkki sivulta 177). Nämä standardoidut jäännökset ja rivikohtaiset z-testit saat *Cells*-painikkeen kohdissa *Residual: Adjusted standardized* ja *z-test* -kohdan molemmat valinnat
   - hyväksy valinnat klikkaamalla *Continue-painiketta*
5. Toteuta ristiintaulukointi nyt *OK*-painiketta painamalla ja saat analyysien tulosteet tulosteikkunaan

Huom. Ristiintaulukossa oleville prosenttiosuuksille on mahdollista muodostaa luottamusvälit toiminnolla Analyze → Tables → Custom Tables.

Ennen kuin tarkastellaan tarkemmin seuraavaa ristiintaulukointiesimerkkiä, todetaan vielä kaksi tämän menetelmän keskeistä yleiskäyttöohjetta. Ensinnäkin prosenttiosuudet lasketaan selittävän muuttujan mukaan, jos analyysissa sellainen on tulkittavissa (periaatteessa meidän ensimmäisessä esimerkissä ei voida määritellä kumpaakaan toisen selittäväksi muuttujaksi). Yleisesti ottaen tämä on hyvä pitää mielessä. Lisäksi prosenttiosuudet lasketaan otoksen edustavuuden mukaisesti. Edellä olleessa esimerkissä oltiin kiinnostuneita sukupuolten viriketaustojen jakaumasta, joten prosentit laskettiin sukupuolimuuttujan mukaan.

Ristiintaulukointi sopii erityisen hyvin sellaisten kysymysten ja muuttujien analysointiin, joissa on kaksi tai kolme kategorista luokkaa. Useinhan Likert-asteikoissa, joita käytetään paljon kasvatustieteellisessä tutkimuksessa, on viisiasteikkoinen skaala ja tämä muuttuja käsitellään usein välimatka-asteikkoisena, vaikkei se sitä todellisuudessa olekaan. Myös viisiasteikkoisen skaalamittarin antamaa dataa voidaan tutkia ristiintaulukoinnin avulla.

Seuraavaksi kysytään sitä, miten miesten ja naisten näkemykset koulutuksen merkityksestä heidän tulevaisuutensa kannalta lukiossa ja yliopistossa eroavat toisistaan. Lisäksi kysytään sitä, minkälainen kehitystendenssi tässä on nähtävissä, eli muuttuvatko asenteet lukiosta yliopistoon siirryttäessä. Tarkastelu on hyvä aloittaa ensin kuvion avulla, koska kuviosta on yleensä helppo hahmottaa tutkittavan kysymyksen perusjakaumat ja vastaavat:



KUVIO 15a. Naisten ja miesten näkemykset koulutuksen merkityksestä heidän tulevaisuudelleen lukiotasolla.

Yleisesti tulkiten kuvion perusteella vaikuttaisi siltä, että koulutukseen suhtautumista ja sen merkityksen näkemistä oman tulevaisuuden kannalta pystytään selittämään sukupuolimuuttujan avulla. Kyseisten muuttujien välillä on siis varsin selvä yhteys. Kuviosta nähdään yksityiskohtaisesti myös sen, että lukioiässä olevat miehet eivät anna koulutukselle niin paljon painoa kuin naispuoliset opiskelijat. Naisista 70 prosenttia on valinnut vastausvaihtoehdon 'suuri' kysymykseen "minkälainen merkitys lukioaikaisella koulutuksella on oman tulevaisuuden kannalta". Vastaavasti vain 30 prosenttia miehistä näkee opiskelun merkityksen suurena tässä vaiheessa. Samanaikaisesti miehistä 27 prosenttia ei näe koulutuksella olevan mitään merkitystä, kun taas otoksen naisopiskelijoista vain 7 prosenttia on tätä mieltä lukiovaiheessa. Jos vielä suoritetaan tilastollinen testaus, khiin neliö -testi antaa tuloksen $\chi^2 = 10{,}20$; df = 2; p = 0,006. Yhteys on siis tilastollisesti merkitsevä ($p < 0{,}05$). Jos vielä lasketaan yhteyden voimakkuutta eli efektikokoa kuvaava tunnusluku, niin saadaan arvoksi Cramérin V = 0,41. Edellä esillä olleiden tulkintaohjeiden perusteella V-arvon perusteella yhteys on voimakkuudeltaan kohtalaista (V ylittää arvon 0,30).

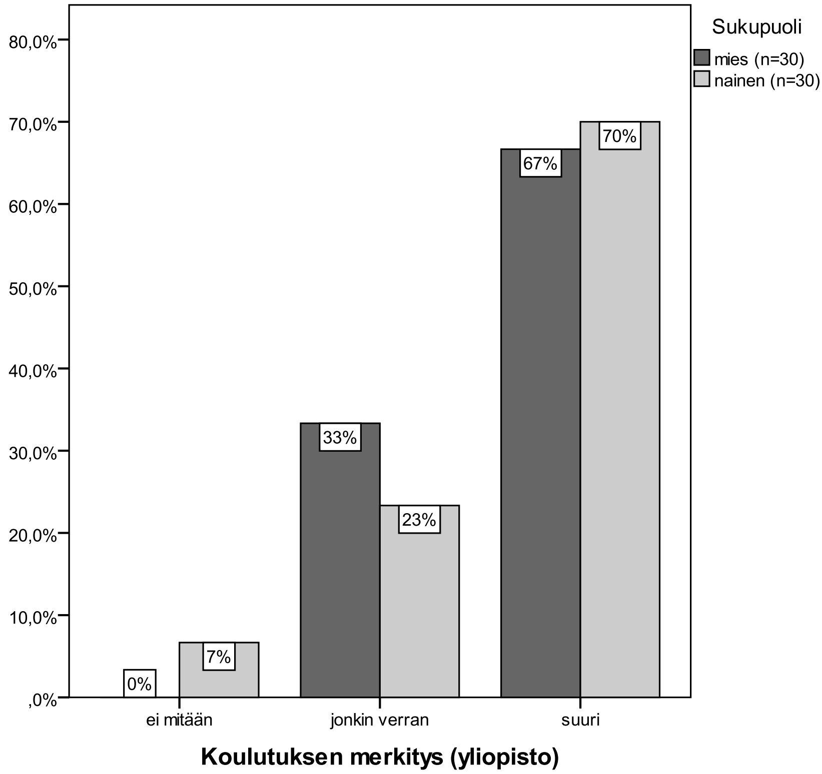
Khiin neliö -testi siis ilmaisi yleisesti, että asioiden välillä on yhteyttä. Tätä testitulosta voidaan vielä tarkentaa laskemalla ristiintaulukkoon niin sanotut jäännösarvot, joiden avulla nähdään missä ristiintaulukon kohdissa erityisesti on poikkeamaa riippumattomuustilanteesta eli tilanteesta, jossa sukupuolten koulutuksen merkitys -jakaumat ovat keskenään identtiset. Toisin sanoen, mitkä taulukon soluista voimakkaimmin vaikuttavat merkitsevään tulokseen. Esimerkkitapauksessa tällä voitaisiin selvittää, missä vastausvaihtoehdoissa sukupuoli näyttäisi erityisesti vaikuttavan vastaamiseen. Jäännösarvoina käytetään yleensä standardoituja suureita, jolloin niiden tulkinta perustuu normaalijakauman ominaisuuksiin, missä itseisarvoltaan lukuarvoa kaksi suuremmat jäännökset voidaan tulkita merkitseviksi. Solukohtaista vertailua voidaan jäännöstermien lisäksi vielä testata erillisiä z-testejä käyttäen ikään kuin post hoc -vertailujen tapaan (kuten aiemmin varianssianalyysin yhteydessä oli esillä). Esimerkin kuvion 14a mukaisessa ristiintaulukossa havaitut frekvenssit, odotetut frekvenssit sekä SPSS:n tuottamat standardoidut jäännökset (adjusted standardized residuals) näyttäisivät seuraavilta:

			Sukupuoli		
			mies	nainen	Total
Koulutuksen merkitys (lukio)	ei mitään	Havaittu frekvenssi	8_a	2_b	10
		Teor. frekvenssi	5,0	5,0	10,0
		% sukupuoliluokissa	26,7%	6,7%	16,7%
		Jäännös (adj. stand.)	2,1	-2,1	
	jonkin verran	Havaittu frekvenssi	13_a	7_a	20
		Teor. frekvenssi	10,0	10,0	20,0
		% sukupuoliluokissa	43,3%	23,3%	33,3%
		Jäännös (adj. stand.)	1,6	-1,6	
	suuri	Havaittu frekvenssi	9_a	21_b	30
		Teor. frekvenssi	15,0	15,0	30,0
		% sukupuoliluokissa	30,0%	70,0%	50,0%
		Jäännös (adj. stand.)	-3,1	3,1	
Total		Havaittu frekvenssi	30	30	60
		Teor. frekvenssi	30,0	30,0	60,0
		% sukupuoliluokissa	100,0%	100,0%	100,0%

Each subscript letter denotes a subset of Sukupuoli categories whose column proportions do not differ significantly from each other at the ,05 level.

Taulukon prosenttiluvut vastaavat kuvion 15a pylväitä. Havaittu frekvenssi kuvaa aineistossa esiintyvää lukumäärää ja teoreettinen frekvenssi kertoo, mikä frekvenssin tulisi olla, mikäli asioilla ei olisi yhteyttä. Standardoitu jäännösarvo liittyy näiden frekvenssien erotukseen. Vastausvaihtoehdoissa 'ei mitään' ja 'suuri' standardoidut jäännökset ylittävät itseisarvoillaan arvon 2, toisin sanoen niissä erityisesti sukupuoli näyttäisi merkitsevästi vaikuttavan vastaamiseen. Taulukkoon tulostuu myös riveillä olevien prosenttilukujen eroa vertaavien z-testien tulokset. Testien tulokset on tiivistetty riveittäin alaindekseinä oleviin kirjainsymboleihin siten, että jos rivin havaittujen frekvenssien alaindeksinä ovat eri kirjaimet (a, b, jne.), kyseiset frekvenssit poikkeavat merkitsevästi toisistaan ($p < 0,05$). 'Jonkin verran' -vastauksissa sukupuolella ei näyttäisi olevan merkitsevää yhteyttä vastaamiseen, mutta muissa vastausluokissa miesten ja naisten prosenttiarvot poikkeavat toisistaan merkitsevästi. (ks. SPSS-ohje standardoitujen jäännösten ja z-testien tuottamisesta sivulla 174).

Edellä käydyssä esimerkissä tutkittiin koulutuksen merkityksen ja sukupuolen välistä yhteyttä lukiossa. Tarkastellaan seuraavaksi kuvion avulla samaa asiaa yliopistovaiheessa:



KUVIO 15b. Naisten ja miesten näkemykset koulutuksen merkityksestä heidän tulevaisuudelleen yliopistotasolla.

Kuviosta nähdään ensinnäkin se, että sukupuolten väliset erot ovat tasoittuneet niin paljon, että erot eivät ole enää tilastollisesti merkitseviä: $\chi^2 = 2{,}55$; df = 2; p = 0,279. (Saatuun p-arvoon liittyy tosin se varaus, että teoreettisista solufrekvensseistä 33 prosenttia on alle viiden). Kuviosta voidaan nähdä myös sen, että yliopistovaiheessa ei enää juuri kukaan näe koulutuksen asemaa merkityksettömänä oman tulevaisuutensa kannalta. Koulutuksen arvostus nousee siis koulutuksen funktiona. Jos lisäksi verrataan tätä kuviota edelliseen, huomataan, että etenkin miesten koulutuksen arvostus lisääntyy merkittävästi lukioaikaisiin näkemyksiin nähden. Lukiovaiheessa 27 prosenttia miehistä ei nähnyt koulutuksella olevan mitään merkitystä tulevaisuutensa kannalta, nyt yliopistovaiheessa näin arvioivia ei enää ole. Sama tendenssi, mutta voimakkaampana, on nähtävissä arvon 'suuri' kohdalla. Lukiossa miesopiskelijoista 30 prosenttia näki koulutuksella olevan suuren vaikutuksen tulevaisuutensa kannalta, mutta yliopistossa peräti 67 prosenttia miehistä kokee koulutuksen merkityksen erittäin suureksi. Asennemuutos on siis tapahtunut erityisesti miesten kohdalla, tai itse asiassa, jos verrataan naisten vastauksia lukion ja yliopiston osalta, nähdään, että naisten prosenttiosuudet ovat pysyneet täysin muuttumattomina. Siten vain miesten koulutukselle antama merkitys muuttuu suhteessa koulutusuralla etenemiseen. Edellä oleva synnyttää heti jatkokysymyksen: entä, jos tarkastelu viedään yksilötasolle, onko silloin naisten kohdalla tapahtunut muutoksia? Voihan olla, että siellä on ollut muutoksia, mutta ne ovat olleet vastakkaisia, jolloin ne ovat kumonneet toisensa. Tähän saadaan vastaus, jos teemme kolmisuuntaisen ristiintaulukoinnin. Samalla saadaan näkyviin edellisten kahden kuvion perustana olevat prosenttiluvut.

SPSS: Kolmen kategorisen muuttujan ristiintaulukointi

1. Löydät analyysin valikkoriviltä *Analyze: Descriptive Statistics → Crosstabs*.
   - siirrä se valitsemasi kategorinen muuttuja, jonka arvot haluat ristiintaulukossa riveille, muuttujaluettelon oikealla puolella olevalla ylimmällä nuolipainikkeella kohtaan *Row(s)*. Esimerkissämme rivimuuttujana on nyt koulutuksen merkitys lukiossa
   - vastaavalla tavalla siirrä keskimmäisellä nuolipainikkeella kohtaan *Column(s)* sen kategorisen muuttujan, jonka arvot haluat ristiintaulukossa sarakkeisiin. Sarakemuuttujana esimerkissämme on koulutuksen merkitys yliopistossa
   - analyysin kolmas kategorinen muuttuja on nyt sukupuoli, siirrä tämä alimmalla nuolipainikkeella kohtaan *Layer 1 of 1.* Tämä tarkoittaa sitä, että koulutuksen merkitykseen lukiossa ja yliopistossa liittyvä ristiintaulukointi toteutetaan nyt erikseen sukupuolittain
   - hyväksy valinnat klikkaamalla *Continue*-painiketta
2. *Cells-valikossa*
   - määrittelet mm. sen, mitkä prosenttiluvut taulukkoihin tulevat mukaan (ks. edellinen SPSS-ohje, s. 174). Nyt haluamme esiin kokonaisprosentit, joten valitsemme vaihtoehdon *Total.*
   - hyväksy valinnat klikkaamalla *Continue*-painiketta
3. Toteuta ristiintaulukointi nyt *OK*-painiketta painamalla ja saat analyysien tulosteet tulosteikkunaan.

Kolmen kategorisen muuttujan ristiintaulukointiajo tuottaa seuraavanlaisen taulukon (olemme muokanneet SPSS-tulostetta vain hieman):

TAULUKKO 11. Naisten ja miesten koulutukselle antama arvostus oman tulevaisuutensa suhteen lukio- ja yliopistovaiheessa.

Sukupuoli				Koulutuksen merkitys (yliopisto) ei mitään	jonkin verran	suuri	Yhteensä
mies	Koulutuksen merkitys (lukio)	ei mitään	n		3	5	8
			% kaikista		10,0%	16,7%	26,7%
		jonkin verran	n		6	7	13
			% kaikista		20,0%	23,3%	43,3%
		suuri	n		1	8	9
			% kaikista		3,3%	26,7%	30,0%
	Yhteensä		n		10	20	30
			% kaikista		33,3%	66,7%	100,0%
nainen	Koulutuksen merkitys (lukio)	ei mitään	n			2	2
			% kaikista			6,7%	6,7%
		jonkin verran	n		4	3	7
			% kaikista		13,3%	10,0%	23,3%
		suuri	n	2	3	16	21
			% kaikista	6,7%	10,0%	53,3%	70,0%
	Yhteensä		n	2	7	21	30
			% kaikista	6,7%	23,3%	70,0%	100,0%

Taulukossa kohdissa *Yhteensä* olevat arvoon 100,0 prosenttia summautuvat prosenttiluvut tai vastaavasti lukuun 30 summautuvat frekvenssit muodostavat ns. reunajakaumat. Niitä on neljä kappaletta: 1) minkälaisen merkityksen miehet antavat koulutukselle lukiossa, 2) minkälaisen merkityksen naiset antavat koulutukselle lukiossa, 3) minkälaisen merkityksen miehet antavat koulutukselle yliopistossa ja 4) minkälaisen merkityksen naiset antavat koulutukselle yliopistossa. Reunajakaumissa ei siis vielä ole tarkasteltu lukiotasoa ja yliopistotasoa samanaikaisesti. Kahdessa edellisessä kuviossa ovat siis nämä reunajakaumat, eli lukiotason kuviota piirrettäessä ei ole käytetty hyödyksi yliopistotasoon liittyvää informaatiota ja päinvastoin; niissä on siis tarkasteltu vain kahta muuttujaa kerrallaan. Tämä puuttuva informaatio tulee näkyviin vasta, kun tarkastellaan kaikkia kolmea muuttujaa yhtäaikaisesti.

Sukupuolimuuttujalla on kaksi eri arvoa ja molemmilla koulutuksen merkitys -muuttujilla kolme, joten kolmiulotteiseen ristiintaulukkoon tulevien solujen määrä on 2*3*3 eli 18. Taulukossa on varjostettu nämä 18 solua. Niissä on puuttuva informaatio, eli niistä nähdään, miten aineiston 60 vastaajaa jakautuvat eri soluihin. Jos tarkastellaan samanaikaisesti kaikkia kolmea muuttujaa, nähdään muun muassa se, että muutoksia on sittenkin tapahtunut naisten kohdalla. Itse asiassa nähdään, että naisista vain kahdella kolmasosalla on asenne koulutuksen merkitykseen säilynyt samanlaisena. Lukioaikainen ja yliopistoaikainen asenne on pysynyt siis samana 20 naisopiskelijalla (4+16) eli 66,6 %:lla (13,3 % + 53,3 %). Se, että nämä muutokset eivät näy reunajakaumissa, johtuu siitä, että muutokset ovat olleet vastakkaisia, esimerkiksi kaksi naista on siirtynyt arvosta ’ei mitään’ arvoon ’suuri’, mutta samanaikaisesti kaksi naista on myös siirtynyt toiseen suuntaan eli arvosta ’suuri’ arvoon ’ei mitään’.

Miesten kohdalla tapahtunut muutos on kuitenkin näinkin tarkastellen ollut naisia suurempi, sillä miehistä vain 46,7 prosentilla asenne on säilynyt muuttumattomana. Kyseinen prosenttiluku on siis lukujen 20,0 prosenttia ja 26,7 prosenttia summa, ja näitä lukuja vastaavat frekvenssit vuorostaan ovat olleet 6 ja 8.

Kaiken kaikkiaan edellä esitetyt tarkastelut osoittavat sen, että kolmen muuttujan samanaikainen ristiintaulukointi voi johtaa eri tulokseen kuin pelkkä kaksiulotteisten ristiintaulukointien tarkastelu. Lohdullista oli kuitenkin se, että kaksiulotteisten taulujen avulla (kuvioina esitettyinä) päästiin oikeille jäljille! Tarkoituksena olikin esittää esimerkinomaisesti, kuinka ristiintaulukointia käyttäen voidaan analysoida moniulotteisiakin tutkimuskysymyksiä. Toki tähän tarkoitukseen on kehitetty muitakin menetelmiä. Esimerkiksi log-lineaariset mallit soveltuvat useiden kategoristen muuttujien samanaikaiseen tarkasteluun. Tässä kirjassa ei käsitellä näitä malleja, mutta jos tätä haluaa kokeilla, miten tällainen mallinnus onnistuu, SPSS:stä löytyy nämä mallit reittiä *Analyze → Loglinear* pitkin. Kovin montaa muuttujaa ei log-lineaarisissa malleissakaan ole järkevää ottaa samanaikaisen tarkastelun kohteeksi, koska tarvittavan havaintoaineiston koko kasvaa nopeasti todella suureksi. Esimerkiksi jos haluttaisiin tarkastella samanaikaisesti viittä kategorista muuttujaa, joista kullakin on kolme eri arvoa, tällaiseen taulukkoon tulisi jo pelkkiä solujakin 3×3×3×3×3=243 kappaletta. Log-lineaaristen mallien tarkastelussa aineiston koon tulee siis olla mieluummin satoja kuin kymmeniä. (Ks. tarkemmin esim. Gröönroos 2011, 157–169; Högmander ym. 2006, 56–69).

Seuraavassa annetaan vielä lyhyt esimerkki numeerisen (eli ei siis alun perin kategorisen) muuttujan analysoimisesta ristiintaulukoinnilla. Tällaisten määrällisten muuttujien kohdalla ristiintaulukoinnin käyttäminen on aina vähän kyseenalaista, koska tähän liittyvä numeerisen tiedon luokittelu luonnollisesti hävittää jossain määrin muuttujan sisältämää informaatiota. Joskus kuitenkin tutkimusasetelma tai joku muu aineiston käsittelyyn liittyvä syy voi antaa tälle perusteet. Esimerkiksi kasvatustieteessä voi hyvinkin tietyissä tutkimusasetelmissa olla tarkoituksenmukaista jakaa tutkimuksessa mukana olevat matemaattisilta taidoiltaan ”heikkojen”, ”keskinkertaisten” ja ”taitavien” ryhmiin, joita käytetään sitten hyödyksi jatkoanalyysejä tehdessä. Tällainen uudelleen ryhmittely mahdollistaa esimerkiksi ristiintaulukoinnin käytön aineiston analyysissa. Myös jotkut tilastolliset analyysit vaativat analysoitavien muuttujien tai ainakin yhden niistä olevan tällaisia kategorisia muuttujia.

Ajatellaan siis, että ollaan kiinnostuneita sukupuolen ja kieliaineissa pärjäämisen suhteesta ja tämä halutaan jostain syystä analysoida ristiintaulukoinnin avulla. Edellä on ryhmitelty kieliainemuuttuja kolmeen ryhmään (ks. luku 2.2): kieliaineissa välttävästi pärjääviin (5,9–7,1), hyviin tai keskitasoisiin (7,2–8,7) ja kiitettävästi kieliaineissa menestyneisiin (8,8–10,0). SPSS tulostaa seuraavanlaisen taulukon (taulukkoon suomennettu englanninkieliset termit).

TAULUKKO 12. Kieliaineiden keskiarvo sukupuolen mukaan.

			Kieliaineiden keskiarvo (luokiteltuna)			
			välttävä	hyvä	kiitettävä	Yhteensä
Sukupuoli	mies	n	10	20	0	30
		%	33,3%	66,7%	,0%	100,0%
	nainen	n	2	14	14	30
		%	6,7%	46,7%	46,7%	100,0%
Yhteensä		n	12	34	14	60
		%	20,0%	56,7%	23,3%	100,0%

$\chi 2$ = 20,44; df = 2; p = 0,000; Cramérin V = 0,58

Taulukosta voidaan nähdä selkeästi, kuinka sukupuolella ja kieliaineiden välillä on yhteyttä lukiovaiheessa. Naisista noin 47 prosentilla kieliaineiden keskiarvo on kiitettävä (8,8–10,0), kun miehillä osuus on nolla. Vastaavasti miehistä peräti 33 prosentilla on kieliaineiden keskiarvo välttävää tasoa (4,0–7,1), kun naisista vain noin 7 prosentilla. Testauskin puhuu selkeää kieltä. Khiin neliö -testin (χ^2) p-arvon avulla (p<0,001) voidaan todeta sukupuolen ja kieliaineissa menestymisen yhteyden olevan tilastollisesti erittäin merkitsevä. Tästä voidaan päätellä, että sukupuolella on merkitystä kieliaineissa menestymisen suhteen. Näin nollahypoteesi, jossa todetaan, ettei sukupuolella ja kieliaineiden keskiarvolla ole yhteyttä keskenään, hylätään. Tilastolliset suureet raportoitaisiin tutkimusraportissa selkeän sanallisen tulkinnan perään APA-tyylin mukaisesti muodossa $\chi^2(2) = 20,44; p < 0,001; V = 0,58$.

Ristiintaulukointia on tässä tarkasteltu varsin pitkään ja perusteellisesti, koska se on eräs luokittelevien muuttujien perusanalyysimenetelmä, jonka käyttömahdollisuudet ja tulkinnat on tärkeätä ymmärtää ja hallita. Seuraavaksi siirrytään korrelaation tarkasteluun, joka sekin on yksi tilastollisen analyysin perusmenetelmiä

6.2 Korrelaatio

Tutkittavien asioiden välisten yhteyksien eli riippuvuuksien tarkastelu liittyy hyvin tyypillisesti määrällisen ja tilastollisen tutkimuksen tutkimuskysymyksiin. Aihetta käsiteltiin jo edellisessä luvussa tilanteissa, joissa riippuvuusanalyysiä toteutettiin kategorisille muuttujille. Tällöin analyysimenetelmänä sovellettiin ristiintaulukointia tilastollisine testeineen ja tunnuslukuineen. Tilanteissa, joissa tutkittavat muuttujat voidaan tulkita mitta-asteikoltaan numeerisiksi (vähintään välimatka-asteikon muuttujat, usein käytännössä myös likert-asteikolliset mielipidemuuttujat), riippuvuustarkastelua voidaan tehdä korrelaatioanalyysin avulla. Tällöin lasketaan kyseiseen tilanteeseen soveltuvia korrelaatiokertoimia, joiden perusteella voidaan arvioida riippuvuuden luonnetta ja voimakkuutta. Ja aivan kuten oli ristiintaulukoinnissakin, myös korrelaation tilastollinen merkitsevyys on testattavissa tilastollisin testein. Yleisimmin käytetty korrelaatiokerroin on lineaariseen eli suoraviivaiseen riippuvuuteen liittyvä Pearsonin tulomomenttikorrelaatio. Tätä käsitellään seuraavaksi heti kappaleen aluksi. Lisäksi tutustutaan järjestyslukuihin perustuvaan Spearmanin järjestyskorrelaatiokertoimen käyttöön sekä eräisiin vaihtoehtoisiin kertoimiin ja niiden tulkintaan.

Yleistä:

- korrelaatiokerroin mittaa kahden vähintään järjestysasteikollisen muuttujan välistä yhteyttä
- muuttujat voivat olla mittaskaalaltaan erilaisia, esim. muuttujat, joissa toisessa asteikko 4-10 ja toisessa 0-20
- korrelaatioanalyysi edellyttää, että muuttujien välinen yhteys on lineaarinen eli suoraviivainen. Graafisten kuvioiden käyttö analyysien yhteydessä (ks. kuvio 16, s. 187) helpottaa havaitsemaan muuttujien välisen riippuvuuden luonnetta ja voimakkuutta
- korrelaatio on positiivista, jos muuttujan X suuriin arvoihin liittyvät muuttujan Y suuret arvot ja jos muuttujan X pieniin arvoihin liittyvät muuttujan Y pienet arvot
- korrelaatio on negatiivista, jos muuttujan X suuriin arvoihin liittyvät muuttujan Y pienet arvot ja jos muuttujan X pieniin arvoihin liittyvät muuttujan Y suuret arvot
- korrelaatiokertoimien arvot vaihtelevat välillä [-1, +1] miinus ykkösen kuvatessa täydellistä negatiivista korrelaatiota ja plus ykkösen täydellistä positiivista korrelaatiota. Arvo nolla tarkoittaa, että muuttujat eivät korreloi lainkaan keskenään

- korrelaatiota voidaan testata tilastollisin merkitsevyystestin; nollahypoteesina on se, että *populaatiossa muuttujien välillä ei ole korrelaatiota* ja vaihtoehtoinen kaksisuuntainen hypoteesi on muotoa *populaatiossa muuttujien välillä on korrelaatiota*
- osittaiskorrelaatiolla (ks. luku 6.2.3) voidaan huomioida eli *vakioida* jonkun kolmannen muuttujan vaikutus kahden muuttujan välisen havaittuun yhteyteen. Näin voidaan kohdentaa muuttujien välisen korrelaatiotarkastelun täsmällisemmin juuri tutkittavien muuttujien välille
- korrelaatio ei kerro, kumpi muuttujista on syy, kumpi seuraus; tosin sanoen korrelaatio ei ilmaise kausaliteettia.

Korrelaatio kuvaa siis kahden muuttujan välistä yhteyttä, tämän yhteyden suuntaa (positiivinen, samansuuntainen vai negatiivinen, käänteinen yhteys) ja yhteyden voimakkuutta. Korrelaatio ilmaisee, millaista on kahden eri muuttujan samanaikainen vaihtelu eli minkälainen on näiden muuttujien yhteisjakauma vai onko sitä lainkaan. Esimerkkiaineistosta saattaisi tutkijaa kiinnostaa esimerkiksi kieliaineiden pistemäärän ja verbaalitestin pisteiden välinen korrelaatio. Todennäköisesti meitä kiinnostaisi myös tieto siitä, mitkä muuttujat korreloivat opintomenestyksen kanssa. Käytetty esimerkkiaineisto on sellainen, että todennäköisesti kaikki opintomenestystä mittaavat muuttujat korreloivat positiivisesti yleiseen opintomenestyksen kanssa, toiset voimakkaammin ja toiset heikommin. Esimerkkinä opintomenestyksen kanssa negatiivisesti korreloivasta muuttujasta voisi olla luvattomien poissaolojen lukumäärä: mitä enemmän oppilaalla on luvattomia poissaoloja koulusta, sitä heikompi on hänen opintomenestyksensä (tai mitä vähemmän poissaoloja, sitä parempi opintomenestys). Korrelaation tulkinnan keskeiset kriteerit voidaan tiivistää vielä seuraavasti:

1. Korrelaatio sopii vain lineaaristen yhteyksien kuvaamiseen. Käyräviivaisen yhteyden kyseessä ollessa korrelaatio ei anna oikeata kuvaa muuttujien välisestä suhteesta. Korrelaatiokertoimen arvo tulee matalaksi, vaikka todellisuudessa muuttujien välillä on riippuvuussuhde. Tällöin korrelaatiota voi koettaa muodostaa aineistoon paloittain.

2. Tapauksessa, jossa toinen muuttujista on kategorinen, esimerkiksi dikotominen eli muuttujalla on vain kaksi arvoa (esimerkiksi 0 ja 1), korrelaatiokerroin saattaa tuottaa käyttökelpoista informaatiota. Tällaises-

sa tilanteissa kannattaa kuitenkin käyttää muita, muuttujien mitta-asteikoille paremmin soveltuvia menetelmiä (esim. t-testiä).

3. Joskus aineiston alaryhmien sisällä voi olla vahvakin korrelatiivinen suhde, joka kuitenkin jää piiloon koko aineistoa käsiteltäessä. Tällöin tulisi korrelaatiot laskea alaryhmittäin.

4. Korrelaatiokerroin ilmaisee muuttujien välisen yhteyden hyvin ylimalkaisesti, eikä sen avulla kyetä tekemään tarkempia analyyseja siitä, miten yhteys muodostuu. Muilla analyysimenetelmillä (esimerkiksi keskiarvotestit ja ristiintaulukointi) voidaan tarkentaa korrelaation antamaa kuvaa.

Näihin voidaan lisätä vielä yhden kriteerin:

5. Erityisesti pienissä aineistoissa muista selvästi poikkeavat arvot voivat vaikuttaa voimakkaasti korrelaatiokertoimen arvoon. Siksi esimerkiksi sirontakuvion avulla kannattaa aina tarkistaa, ettei tällaisia ääriarvoja ole. Jos niitä on, silloin voi kokeilla niiden tilapäistä poistamista ja katsoa, kuinka paljon korrelaatiokertoimen arvo muuttuu, ja ottaa nämä muutokset huomioon tulkintoja tehtäessä.

6.2.1 Pearsonin korrelaatiokerroin

Pearsonin (tulomomentti)korrelaatiokerroin (r) on käytetyin kahden muuttujan korrelaatiota kuvaavista korrelaatiokertoimista. Pearsonin korrelaatiokerroin edellyttää, että muuttujat ovat vähintään välimatka-asteikkoisia.

Yleistä:

- kuvaa kahden muuttujan välisen lineaarisen eli suoraviivaisen riippuvuuden suuntaa ja voimakkuutta
- riippuvuussuhde voi olla positiivinen tai negatiivinen
- korrelaatiokertoimen arvo nolla tarkoittaa täydellistä lineaarista riippumattomuutta, ts. muuttujien välillä ei ole lainkaan lineaarista yhteyttä
- edellyttää vähintään välimatka-asteikkoisia muuttujia
- otosjakauman tulisi olla normaalijakauman mukainen

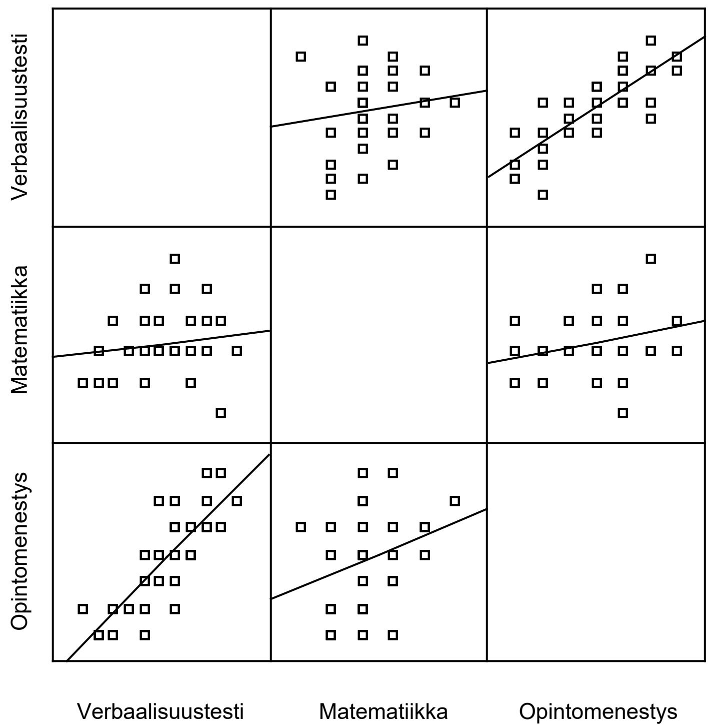
- kuten moniin muihinkin tilastollisiin tunnuslukuihin (keskiarvo, keskihajonta, jne), vaikuttavat korrelaatioonkin haitallisesti tutkimusaineistossa esiintyvät haja-arvot
- korrelaatiokertoimen arvo voidaan tulkita myös efektikoon mittana lineaarisen riippuvuuden voimakkuudelle
- korrelaatiokertoimen tulkintaa voidaan täydentää väliestimointia eli luottamusvälitarkastelua käyttäen
- useat monimuuttujamenetelmät perustuvat Pearsonin tulomomenttikertoimista muodostettuun korrelaatiomatriisiin

Korrelaatiokertoimen tilastollista merkitsevyyttä testattaessa testin nollahypoteesi on muotoa ”tarkasteltavien muuttujien välillä ei ole korrelaatiota populaatiossa”, eli H_0: $\rho = 0$. Mitä suurempi on korrelaatiokertoimen otoksesta laskettu arvo, sitä merkitsevämpi on testin tulos. Tilastolliseen merkitsevyyteen, p-arvoon, vaikuttaa monien muiden testien tapaan kuitenkin havaitun arvon lisäksi myös otoskoko. Mitä pienempi otos on, sitä suurempi korrelaatiokertoimen tulee olla ennen kuin sen voidaan katsoa osoittavan kahden muuttujan välisen riippuvuuden olevan tilastollisesti merkitsevää eikä sattumasta johtuvaa. Korrelaatiokerroin voidaan mieltää myös riippuvuuteen liittyvän efektikoon mittana, jonka tulkintaan löytyy eri lähteissä hieman erilaisia raja-arvoja (ks. esimerkiksi Ellis 2010, 41; Nummenmaa 2011, 290). Suuntaa antavina kriteereinä voidaan pitää seuraavia: $r \geq 0{,}7$ riippuvuuden voi tulkita voimakkaaksi; $0{,}3 < r < 0{,}7$ riippuvuuden voi tulkita kohtalaiseksi tai merkittäväksi, alle 0,3 olevan riippuvuuden voi katsoa heikoksi tai olemattomaksi (tosin jos otos on suuri, esimerkiksi $N > 50$, niin jonkin verran alle 0,3 oleva kerroinkin voidaan jo tulkita kohtalaiseksi). Cohen (1988) on esittänyt heikon – kohtalaisen – voimakkaan korrelaation raja-arvoiksi niinkin matalia raja-arvoja kuin 0,10 – 0,30 – 0,50. Vastaavia raja-arvoja voidaan käyttää myös negatiivisen eli käänteisen korrelaation tapauksissa. (Ks. asetelma 5, s. 49).

Korrelaatiokerroin (Pearsonin korrelaatio) kuvaa lineaarista yhteyttä, tulkinnassa tulee olla varovainen. Yhteys voi esimerkiksi olla käyräviivainen, eli korrelaatio voi olla nolla ($r = 0$) vaikka muuttujien välinen yhteys olisi täydellistä, vaikkei lineaarista. Havaintoyksiköiden sijoittuminen tulisi aina tutkia esimerkiksi sirontakuvion eli korrelaatiodiagrammin avulla. Lisäksi muuttujien suhteen luonnetta voidaan tarkastella monien

analyysimenetelmienkin avulla, esimerkiksi regressioanalyysissa niin sanottuja jäännöksiä tutkimalla. Näin vältytään muuttujien välistä yhteyttä koskevilta virhepäätelmiltä.

Jos haluaa nopeasti selvittää pääpiirteissään (pienehkön) muuttujajoukon muuttujien välisten yhteyksien luonteen, voi piirtää seuraavanlaisen kuvion, jossa samanaikaisesti on useita sirontakuvioita. Tarkastelun kohteena ovat esimerkkiaineiston verbaalisuustestin pistemäärän, matematiikan numeron ja lukioaikaisen opintomenestyksen väliset yhteydet:



KUVIO 16. Verbaalisuustestiä, matematiikan numeroa ja lukion opintomenestystä kuvaavien muuttujien yhteydet sirontakuvioiden avulla esitettyinä.

Sirontakuviot osoittavat, että yhteyksien joukossa ei näyttäisi olevan käyräviivaisia yhteyksiä. Pienillä neliöillä kuvatut havaintoarvot keskittyvät (enemmän tai vähemmän) muuttujien välistä lineaarista yhteyttä kuvaavan niin sanotun regressiosuoran lähettyville. Parhaiten suora näyttää kuvaavan verbaalisuustestin ja opintomenestyksen välistä yhteyttä, eli näiden välinen lineaarinen yhteys on kaikkein voimakkain. Kaikki saadut korrelaatiokertoimet ovat positiivisia, eli pääsuunta on se, että toisen muuttujan suuriin arvoihin liittyvät toisen muuttujan suuret arvot (tai päinvastoin: pieniin liittyvät pienet arvot). Sirontakuvioissa positiivinen yhteys näkyi siten, että regressiosuorat ovat nousevia (negatiivista korrelaatiota kuvaisi laskeva suora). Korrelaatioiden ja sirontakuvioiden avulla voidaan siis jo saada varsin hyvä käsitys eri muuttujien välisistä yhteyksistä, joten näiden avulla tutkimuksen teossa päästään hyvään vauhtiin.

SPSS: Sirontakuvion piirtäminen (useita määrällisiä muuttujia):

Jos haluat kuvioiden avulla nopean yleissilmäyksen siitä, miten aineistosi määrälliset (numeeriset) muuttujat korreloivat keskenään, kokeile seuraavaa tapaa.

1. Valitse ylhäältä valikkoriviltä *Graphs* → *Legacy Dialogs* → *Scatter*.
   - esiin tulevista kuvakkeista valitse *Matrix*-kuvake, ja jatka *Define*-painikkeella eteenpäin
   - valitse muuttujalistasta haluamasi numeeriset muuttujat ja siirrä ne listan oikealla puolella olevalla ylimmällä nuolipainikkeella kohtaan *Matrix Variables*
2. *Options*-painike
   - tarkista, että mahdolliset puuttuvat tiedot eivät tule mukaan kuvioon, eli poista valintaruksi kohdasta *Display groups defined by missing values*.
   - hyväksy valinnat klikkaamalla *Continue* -painiketta
3. Paina lopuksi *OK*-painiketta, niin kuvio piirtyy tulosteikkunaan

Jatka tulosteikkunassa olevan kuvion muokkaamista kaksoisnapsauttamalla sitä, jolloin saat kuvion muokkaustilaan (SPSS Chart Editor).

4. Lisää kuvioon muuttujien välistä lineaarista yhteyttä kuvaava regressiosuora valikkorivin vaihtoehtojen *Chart* ja *Options* avulla
   - ruksaa *Fit Line* -kohdan vaihtoehto *Total* ja napsauta painiketta *Fit Options*. Varmista, että kohdassa *Fit Method* on valittuna vaihtoehto *Linear regression*.
5. Kuvion muokkaustilasta pääset takaisin valitsemalla ylhäältä valikkoriviltä vaihtoehdot *File* ja *Close*.

Tutkitaan esimerkin muuttujien välisiä yhteyksiä myös korrelaatiokertoimien ja niiden tilastollisen testauksen avulla:

		Verbaalisuustestin pistemäärä	Matematiikan numero	Opintomenestys lukiossa
Verbaalisuustestin pistemäärä	Pearsonin korrelaatio	1		
	p-arvo			
Matematiikan numero	Pearsonin korrelaatio	,143	1	
	p-arvo	,277		
Opintomenestys lukiossa	Pearsonin korrelaatio	,812	,291	1
	p-arvo	,000	,024	

Taulukossa esiintyvät korrelaatiokertoimen arvot 1 liittyvät siihen, että muuttujan korrelaatio itsensä kanssa on luonnollisesti 1,00. Varsinaiset kiinnostuksen kohteena olevat muuttujien väliset korrelaatiot sijaitsevat yllä olevassa taulukossa tämän ykkösiä sisältävän päälävistäjän alapuolella. Verbaalisuustestin ja matematiikan numeron välinen korrelaatio on heikohko, $r = 0{,}14$; yhteys ei myöskään ole tilastollisesti merkitsevä, sillä $p = 0{,}277$. Verbaalisuustesti ja opintomenestys korreloivat voimakkaasti, $r = 0{,}81$; ja yhteys on tilastollisestikin erittäin merkitsevä. Tähän liittyvä korrelaatiokertoimen 95 % luottamusväli on 0,74–0,88. Tämän voi tulkita niin, että todellinen korrelaatio sisältyy näiden arvojen väliin 95 % varmuudella. Luottamusväliestimointi on SPSS-ohjelmassa toteutettavissa bootstrap-menettelyä hyödyntäen (ks. luku 1.6.1.4 ja sivun 129–131 esimerkki). Yleisesti korrelaatiotaulukon perusteella voidaan päätellä, että verbaalinen menestyminen on yhteydessä yleiseen opinnoissa menestymiseen. Leikkimielisesti voidaankin kysyä, voisiko opintomenestystä mitata pelkästään verbaalisuustestillä. Toisaalta opintomenestykseen näyttää vaikuttavan myös matematiikan numero, sillä näiden välinen korrelaatio on kohtalaista, $r = 0{,}29$; yhteys on tilastollisesti merkitsevä, $p = 0{,}024$ eli $p < 0{,}05$. Lisäksi tulosta voitaisiin täydentää edellä olevaan tapaan luottamusvälitarkastelulla.

SPSS: Korrelaatiokertoimen (Pearson) laskeminen

1. Valitse ylhäältä valikkoriviltä vaihtoehdot *Analyze → Correlate → Bivariate*.
   - siirrä valitsemasi numeeriset muuttujat muuttujaluettelon oikealla puolella olevalla nuolipainikkeella kohtaan *Variables*
   - varmista, että kohdan *Correlation Coefficients* vaihtoehto *Pearson* on valittuna
2. Toteuta korrelaatiotarkastelu *OK*-painikkeella ja saat analyysin tulokset tulosteikkunaan.

Huom: korrelaatiokertoimien bootstrap-luottamusvälit saat tulostettua mukaan korrelaatiotaulukkoon valitsemalla Bootstrap-painikkeen takaa kohdan *Perform Bootstrapping*.

6.2.2 Järjestyskorrelaatiokertoimet

Järjestyskorrelaatiokertoimen avulla tutkitaan kahden muuttujan välistä yhteyttä. Järjestyskorrelaation laskeminen tulee ajankohtaiseksi, kun muuttujat ovat järjestysasteikollisia ja/tai ollaan kiinnostuneita esimerkiksi siitä, miten otoksessa matematiikan numeron mukaan meneystyneet (järjestetty paremmuusjärjestykseen) ovat sijoittuneet esimerkiksi verbaalisessa testissä eli onko järjestys kummassakin samankaltainen vai ei. Järjestyskorrelaatiokertoimiin ei liity normaalijakaumaoletuksia, joten niitä kannattaa käyttää myös silloin, kun esimerkiksi Pearsonin korrelaatiokertoimeen liittyvä normaalijakaumaolettamus ei pidä paikkaansa. Edellisessä luvussa käsitelty Pearsonin korrelaatio liittyi lineaarisen eli suoraviivaisen yhteyden tutkimiseen, järjestyskorrelaatio sen sijaan liittyy monotonisen riippuvuuden voimakkuuden tutkimiseen. Tällainen riippuvuus voi olla myös epälineaarista eli käyräviivaista.

Järjestyskorrelaatio lasketaan useimmiten Spearmanin järjestyskorrelaatiokertoimen (r_s) avulla. Se on ajatustavaltaan hyvin samankaltainen kuin Pearsonin korrelaatiokerroin, paitsi että Spearman käyttää kerrointa laskiessaan muuttujien alkuperäisten arvojen sijasta niiden avulla muodostettuja järjestyslukuja. Jos järjestykset ovat täysin samoja eri muuttujissa, tulee kertoimen arvoksi plus yksi; miinus yksi kuvaa taas sitä, että järjestys on mitatuissa muuttujissa täysin käänteinen.

Jos lasketaan Spearmanin järjestyskorrelaatiokertoimen esimerkkiaineiston muuttujien 'verbaalisuustestin pistemäärä' ja 'matematiikan numero' välille, saadaan arvon $r_s = 0{,}14$ ($p = 0{,}293$). Kerroin ei siis ole tilastollisesti merkitsevä; toisin sanoen kerroin ei merkitsevästi poikkea nollasta. Tulosta voitaisiin täydentää edellä olevaan tapaan luottamusvälitarkastelulla (ks. 189). Pearsonin korrelaatiokertoimen arvo näillä samoilla muuttujilla oli hyvin samankaltainen; kolmen desimaalin tarkkuudella ilmaistuna arvo oli 0,143.

Toinen kohtalaisen usein esiintyvä järjestyskorrelaatiokerroin on Kendallin järjestyskorrelaatiokerroin (Kendall's tau-b,τ). Se eroaa tulkinnaltaan Spearmanin ja Pearsonin korrelaatiokertoimista. Kendallin järjestyskorrelaatiokerroin perustuu todennäköisyyksiin. Jos todennäköisyydestä, että kahden muuttujan arvojen järjestys on sama, vähennetään todennäköisyys, että arvojen järjestys on eri, päädytään kyseiseen kertoimeen. Tätä kerrointa voi siis käyttää silloin, kun kysymyksenasettelu on sellainen, että nimenomaan muuttujien arvojen järjestyksen samuus (tai erilaisuus) on kiinnostuksen kohteena. Tämänkin kertoimen arvojen vaihteluväli on [-1,+1]. Jos muuttujien arvojen järjestys on täysin sama, kerroin saa arvon plus yksi (=todennäköisyyksien 1 ja 0 erotus). Jos taas järjestys on täysin päinvastainen eli siis täysin erilainen, kertoimen arvoksi tulee miinus yksi (=todennäköisyyksien 0 ja 1 erotus).

Jos lasketaan Kendallin järjestyskorrelaatiokertoimen τ esimerkkiaineiston muuttujien 'verbaalisuustestin pistemäärä' ja 'matematiikan numero' välille, saadaan heikkoa yhteyttä kuvaavan arvon $\tau = 0{,}11$ ($p = 0{,}277$). Kerroin ei siis ole tilastollisesti merkitsevä.

SPSS: Järjestyskorrelaatiokertoimen laskeminen

1. Valitse ylhäältä valikkoriviltä vaihtoehdot *Analyze → Correlate → Bivariate*.
   - siirrä valitsemasi vähintään järjestysasteikolliset muuttujat muuttujaluettelon oikealla puolella olevalla nuolipainikkeella kohtaan *Variables*
   - varmista, että kohdan *Correlation Coefficients* vaihtoehto *Spearman* on valittuna
   - voit valita myös vaihtoehdon *Kendall's tau-b*, sillä myös se on järjestyskorrelaatiokerroin
2. Toteuta korrelaatiotarkastelu *OK*-painikkeella ja saat analyysin tulokset tulosteikkunaan.

Huom: korrelaatiokertoimien bootstrap-luottamusvälit saat tulostettua mukaan korrelaatiotaulukkoon valitsemalla Bootstrap-painikkeen takaa kohdan *Perform Bootstrapping*.

6.2.3 Osittaiskorrelaatio

Kun halutaan tarkastella kahden muuttujan välistä riippuvuutta niin että muiden, yhden tai useamman, muuttujan vaikutus tähän suhteeseen on vakioitu eli poistettu, tämän tarkastelu on mahdollista osittaiskorrelaation (partial correlation) avulla. Tässä siis halutaan eliminoida kolmannen tai useamman tekijän mahdollinen (lineaarinen) vaikutus tutkittavien kahden muuttujan välisen yhteyden ilmenemiseen. Kyse on siis näennäisyhteyksien, joita helposti tulee, kun tarkastellaan vain kahden muuttujan välistä riippuvuutta, esille nostamisesta. Tutkimuksenhan tavoitteena on näiden sijaan löytää todellisia muuttujien välisiä yhteyksiä. Se, että kahden muuttujan välillä havaitaan yhteys, ei vielä todista sitä, että havaittu yhteys on todellinen. Voihan olla niinkin, että joku kolmas tekijä selittää joko kokonaan tai osittain havaitun kahden tekijän välisen riippuvuuden. Tuskinpa esimerkiksi opintomenestystä kouluissa voi selittää vain yhdellä taustatekijällä.

Ajatuksena siis osittaiskorrelaatioanalyysissa, kuten monimuuttuja-analyyseissa yleensäkin, on se, ettei kahden tekijän tai ilmiön välinen tilastollinen yhteys ole riittävä osoittamaan sitä, että näillä olisi todellista yhteyttä toisiinsa: ”todellisen yhteyden toteamiseksi on empiirisesti kyettävä osoittamaan, ettei jokin kolmas, kahteen alkuperäiseen muuttujan yhteydessä oleva tekijä synnytä tätä yhteyttä” (Jokivuori & Hietala 2007, 15). Jokivuori ja Hietala (emt.) toteavatkin tähän liittyen sattuvasti sen, että tulkinta ilman havaittujen yhteyksien monipuolista ja systemaattista analyysia (elaboraatiota) voi pahimmillaan nojata tutkijan tieteellisiin ennakkoluuloihin ja omasta tulkinnasta voi tulla näin vain ”kokoelma tieteellisiä ennakkoluuloja.” Heidän mukaansa kvantitatiivinen tutkimus ilman elaboraatiota ei oikeastaan ole tutkimusta lainkaan. (Jokivuori & Hietala 2007, 15–19.) Elaboraatio-käsitteellä viitataan määrällisissä tutkimuksissa siihen, kun tutkijat pyrkivät esimerkiksi vakioimalla jonkun kolmannen tekijän vaikutusta varmistumaan siitä, ettei kahden muuttujan tutkittavien välinen yhteys johdu tästä kolmannesta tekijästä tai toisinpäin, kuinka eri tutkijakombinaatiot yhdessä vaikuttavat tutkittavaan muuttujaan (ks. esim. Toivonen 1999, 194–205).

Osittaiskorrelaatio perustuu Pearsonin korrelaatiokertoimeen, joten se sopii ainoastaan numeerisille muuttujille. Oikeastaan osittaiskorrelaatiot liittyvät tilanteisiin, joissa on jokin selitettävä muuttuja ja joukko selittäviä muuttujia. Tämän vuoksi osittaiskorrelaatiota käytetäänkin aika harvoin, regressioanalyysi tarjoaa usein tutkijalla tähän paremmat tarkasteluvälineet (vrt. regressioanalyysi luku 7). Kuitenkin osittaiskorrelaatioon perustuvakin tarkastelu on käypä analyysimenetelmä etenkin niissä asetelmissa, jossa halutaan vakioida kahden muuttujan välisen yhteyden tarkastelussa yhden tai useamman muuttujan vaikutus tähän yhteyteen. Esimerkiksi käyttämämme aineistoon liittyen, voisimme tarkastella sitä, kuinka merkittävästi oppilaiden verbaalisuustestin ja lukiossa menestymisen välinen korrelaatio muuttuisi, jos tässä testiin lisättäisiin mukaan kolmanneksi tekijäksi kieliaineiden keskiarvo (tämän vaikutus siis vakioitaisiin), joka myös korreloi voimakkaasti opintomenetykseen. Osittaiskorrelaation arvon muodostuminen perustuu muuttujien välisten yhteisvaihteluiden tarkasteluun suhteessa kokonaisvaihteluun. Periaate on lähellä regressioanalyysin selitysasteen käsitettä (paljonko muuttujan vaihtelusta kyetään selittämään toisen tai usean muun muuttujan avulla). Osittaiskorrelaatiossa muuttujien välisestä yhteisvaihtelusta huomioidaan vain tarkasteltavien muuttujien itsenäinen (unique) yhteisvaihtelu suhteessa kokonaisvaihteluun, josta on poistettu vakioitavien muuttujien osuus. Näin saadun selitysosuuden neliöjuuri tuottaa osittaiskorrelaation arvon. Periaate tässäkin on sama kuin regressioanalyysissä, jossa selitysaste muodostuu korrelaatiokertoimen neliönä (ks. yhden selittäjän regressioanalyysi luvussa 7.1, s. 197).

Tarkastellaan lyhyesti, miten edellä esitetyn kysymyksen suhteen todellisuudessa käy. Otetaan esiin tarkastelun kohteena olevaksi muuttujaksi opintomenestys lukiossa (selitettävä muuttuja). Aikaisemmin korrelaatiotarkasteluissa (s. 189) havaittiin, että verbaalisuustestin pistemäärä korreloi voimakkaasti opintomenestykseen lukiossa ($r = 0{,}81$). Tämän perustella voidaan siis päätellä, että verbaalisuustestin pistemäärällä ja opintomenestyksellä on hyvinkin vahva positiivinen yhteys. Toisaalta myös kieliaineiden keskiarvo korreloi voimakkaasti kyseiseen opintomenestykseen ($r = 0{,}83$). Tämäkin muuttuja olisi siis hyvä selittäjä opintomenestykselle. Koska myös nämä kaksi selittäjää korreloivat keskenään voimakkaasti (r = 0,91), on näiden kahden muuttujan yhteisvaihtelu syytä huomioida korrelaatioanalyysissä. Täsmennetään siis perinteistä korrelaatioanalyysiä osittaiskorrelaatiokertoimien avulla.

Kun lasketaan muuttujien 'opintomenestys lukiossa' ja 'verbaalisuustestin pistemäärä' välinen osittaiskorrelaatio, kun kieliaineiden keskiarvon vaikutus on eliminoitu eli vakioitu, niin saadaan korrelaation arvoksi $r_p = 0{,}26$ ($p = 0{,}045$). Yhteys on vieläkin tilastollisesti merkitsevää tasoa, mutta yhteys ei ole enää lainkaan niin voimakasta kuin se alun perin näytti peruskorrelaatioanalyysin pohjalta. Liikkeellehän lähdettiin arvosta r = 0,81. Sama pätee myös toiseen selittäjään. Jos lasketaan muuttujien 'opintomenestys lukiossa' ja 'kieliaineiden keskiarvo' välinen osittaiskorrelaatio, kun verbaalisuustestin pistemäärän vaikutus on vakioitu, saadaan arvoksi $r_p = 0{,}36$ ($p = 0{,}005$). Yhteys on tilastollisesti merkitsevä, mutta tämäkään yhteys ei ole niin voimakas, miltä se alun perin näytti (r = 0,83). Sen, että korrelaatiot alenivat näin paljon, aiheutti selittäjien välinen voimakas korrelaatio (r = 0,91). Sääntö on siis se, että mitä pienempi on selittäjien välinen korrelaatio, sitä lähempänä osittaiskorrelaatiot ja tavalliset korrelaatiot ovat toisiaan. Osittaiskorrelaatio tulosta voidaan myös täydentää edellä olevaan tapaan luottamusvälitarkastelulla (ks. 189).

Edellä tarkasteltu esimerkki osoittaa konkreettisesti sen, joka käy järkeen jo arkiajattelunkin perusteella, että kahden tekijän väliseen riippuvuuteen tai yhteyteen vaikuttaa yleisesti muitakin tekijöitä. Joten yleensä kahden muuttujan riippuvuuden tarkastelua onkin syytä jatkaa juuri osittaiskorrelaation tai regressioanalyysin tarjoamilla lisäanalyyseillä.

SPSS: Osittaiskorrelaatiokertoimen laskeminen

1. Valitse ylhäältä valikkoriviltä vaihtoehdot *Analyze → Correlate → Partial*
   - siirrä ne valitsemasi numeeriset muuttujat, joiden välisestä korrelaatiosta olet ensisijaisesti kiinnostunut, muuttujaluettelon oikealla puolella olevalla ylemmällä nuolipainikkeella kohtaan *Variables*
   - siirrä se numeerinen muuttuja, jonka suhteen haluat vakioida edellä valitsemiesi muuttujien korrelaation, toiseksi ylimmällä nuolipainikkeella kohtaan *Controlling for*
2. Toteuta osittaiskorrelaatiotarkastelu *OK*-painikkeella ja saat analyysin tulokset tulosteikkunaan.

Huom: korrelaatiokertoimien bootstrap-luottamusvälit saat tulostettua mukaan korrelaatiotaulukkoon valitsemalla Bootstrap-painikkeen takaa kohdan *Perform Bootstrapping*.

7. Muuttujien välisten yhteyksien mallintaminen regressioanalyysilla

Edellisen luvun korrelaatiotarkasteluissa oltiin kiinnostuneita siitä, onko määrällisten muuttujien välillä lineaarista eli suoraviivaista yhteyttä ja miten voimakasta se on. Korrelaatioanalyysiä on luontevaa jatkaa siten, että selvitetään lisäksi, millaista asioiden välillä oleva suoraviivainen yhteys voisi olla. Regressioanalyysin avulla on mahdollista muodostaa ja ratkaista lineaarista yhteyttä kuvaava tilastollinen malli, jonka avulla saadaan tulkittua yksityiskohtaisemmin, miten tarkasteltavat muuttujat ovat yhteydessä keskenään. Tällainen regressiomalli voidaan muodostaa kahden tai myös useamman muuttujan välisten yhteyksien tutkimiseen. Lisäksi mallissa eli regressioanalyysissa muuttujat voidaan asettaa tutkimuskysymysten mukaisesti selitettävien ja selittävien tekijöiden rooleihin, jolloin analyysin tuottamat tulokset täsmentävät ja tarkentavat selvästi korrelaatioanalyysia.

Yleistä:

- regressioanalyysia käytetään tutkittaessa, miten numeerisen selittävän muuttujan vaihtelu (X)
- johtaa vaihteluun numeerisessa selitettävässä muuttujassa (Y)
- selittäviä (independent) muuttujia voi olla useampia kuin yksi, selitettäviä (dependent) muuttujia vain yksi (selitettäviä muuttujia toki voi regressioanalyysissa olla useitakin, mutta tässä kirjassa tällaista asetelmaa ei käsitellä)
- analyysiin mukaan otettavissa muuttujissa pitää esiintyä varianssia, arvojen vaihtelua
- muuttujien välinen yhteys on lineaarinen (oletuskriteeri), ja se on kuvattavissa esimerkiksi yhden selittäjän regressiomallissa yhtälön $Y = a + bX$ avulla, jossa a (selittävän muuttujaan liittyvä vakiotekijä) on kyseisen suoran ja y-akselin leikkauspiste ja b on regressiosuoran kulmakerroin.
- regressiosuoralla kuvataan selittävän muuttujan/selittävien muuttujien yhteyden voimakkuutta ja suuntaa

- yhteys voi olla: positiivinen, negatiivinen tai sitä ei ole lainkaan
- regressiokertoimella kuvataan voimakkuuden suuruutta (mitä suurempi tämä on, sitä voimakkaampi on havaittu yhteys muuttujien välillä selitettävään muuttujaan); kertoimen etumerkki (negatiivinen, positiivinen) kertoo suoran eli yhteyden suunnan
- regressiomalli on sitä parempi (selitysvoimaisempi), mitä lähemmäksi yksittäiset havaintoarvot sijoittuvat saatua regressiosuoraa
- analyysin mukaan valittujen muuttujien tulee olla tutkimuskysymyksen suhteen tarkoituksenmukaisia ja myös teoreettisesti perusteltuja
- kertoimet a ja b pyritään valitsemaan siten, että regressiosuoran avulla saadut kuvitteelliset Y:n arvot ovat mahdollisimman lähellä havaittuja Y:n arvoja (ns. pienimmän neliösumman menetelmä)
- käytetään usein tutkittaessa, mitkä muuttujat selittävät parhaiten selitettävää muuttujaa; selitysaste kertoo sen, kuinka hyvin selittämisessä on onnistuttu; sopii myös ennustamiseen

Käytön edellytykset:

- Tyypillisesti kaikki muuttujat välimatka- tai suhdeasteikkoisia. Erikoistapauksina kategoristen muuttujien regressioanalyysit, joissa mallissa voi olla kategorisia selittäjiä niin sanottuina dummy-muuttujina (arvot 0 tai 1), ja logistinen regressioanalyysi, jossa selitettävä muuttuja on kategorinen
- selitettävä muuttuja Y aina satunnaismuuttuja (ts. sen vaihtelu ei ole ennakoitavissa), selittävät muuttujat joko satunnaisia tai kiinteitä
- regressiosuoran antamien Y:n arvojen ja havaittujen Y:n arvojen välisten erojen eli jäännösten tulee olla normaalisti jakautuneita
- regressioanalyysi soveltuu muuttujien välisten yhteyksien kuvailuun, tilastolliseen päättelyyn ja kausaalisten yhteyksien tulkintaan. Jälkimmäisessä tapauksessa tosin on syytä olla tarkkana syy-seuraus -suhteiden tulkinnassa.
- selittäviä muuttujia ei saa olla liikaa (karkeana nyrkkisääntönä voisi pitää sitä, että havaintoja on 10–20 kertaa enemmän kuin selittäviä muuttujia)
- selittävät muuttujat eivät saa korreloida keskenään voimakkaasti (tämä ns. multikollineaarisuus on usein ongelma sovellettaessa regressioanalyysia kasvatustieteellisessä tutkimuksessa)

Regressioanalyysi soveltuu useiden erityyppisten tutkimuskysymysten tai -ongelmien ratkaisemiseen. Kasvatustieteellisessä tutkimuksessa regressioanalyysin käyttö tulee ajankohtaiseksi, kun tutkimme esimerkiksi sitä, mitkä tekijät vaikuttavat oppilaan tulevaan opintomenestykseen. Selittäviä muuttujia olisivat silloin esimerkiksi todistuksen arvosanat, oppilaan luonteenpiirteet, kotitausta yms. Lisäksi voidaan arvioida esimerkiksi eri opetusmenetelmien keskinäistä paremmuutta suhteessa asetettuun oppimistavoitteeseen.

Metsämuuronen (2009, 709–780) ja Nummenmaa (2011, 309–329) ovat esittäneet regressioanalyysin keskeisiä piirteitä siinä määrin kattavasti, ettei tässä yhteydessä ole tarpeen käsitellä menetelmän laskentatekniikkaa. Lisäksi englanninkielisistä monimuuttujamenetelmiä käsittelevissä kirjoista löytyy yksityiskohtaiset ja kattavat kuvaukset regressioanalyysin taustoista, toteutuksista ja tulkinnoista (ks. esim. Hair 2010; Tabachnic 2019). Tässä riittää, kun korostetaan sitä, että regressioanalyysin perusajatuksena on tarkastella selittävien muuttujien avulla selitettävän muuttujan vaihtelua ja muodostaa näistä mahdollisimman hyvä malli tämän suhteen.

Tilasto-ohjelmissa regressioanalyysin voi yleensä toteuttaa joko tavalliseen tapaan (ks. luku 7.1) tai askeltavasti (*stepwise,* ks. luku 7.2). Askeltavassa regressioanalyysissa ohjelmalle annetaan joukko muuttujia selittäviksi muuttujiksi, joista ohjelma valitsee askel askeleelta parhaat selittäjät. Analyysi loppuu, kun tilastollisesti merkitseviä selittäjiä ei enää löydy. Askeltavan regressioanalyysin käytön pääpiirteet voidaan kiteyttää vaikkapa seuraavasti: analyysissa pyritään malliin, jossa mallin selitysosuus pyritään maksimoimaan ja selitettävien muuttujien määrä minimoimaan. Askeltavan regressioanalyysin varjopuolena on sen tietynlainen mekaanisuus. Tutkimusta tehdessä onkin mietittävä, onko olemassa sellaisia tekijöitä, jotka sisällöllisin perustein on syytä ottaa malliin mukaan, vaikkei niiden selityskyky olisikaan tilastollisesti merkitsevä. Toisaalta voi miettiä myös sitä, kannattaako malliin ottaa sellaista tilastollisesti merkitsevää selittäjää, jonka tulkinta ontuu tai tuntuu keinotekoiselta.

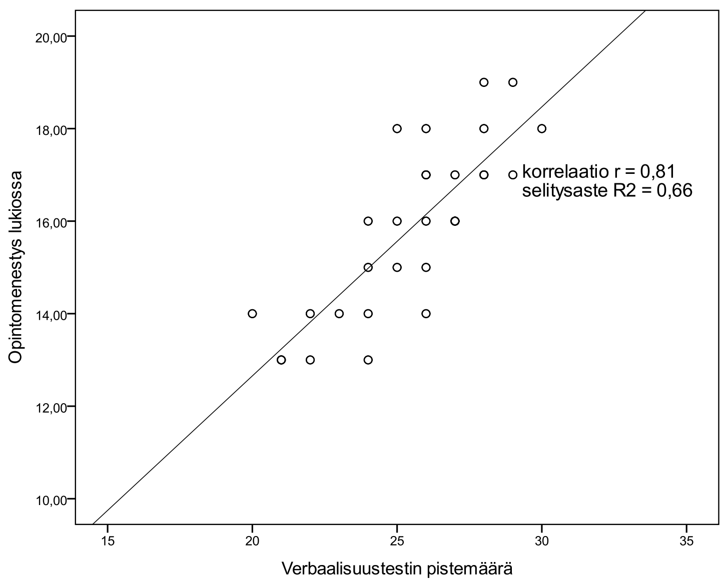
Tavallisessa regressioanalyysissa lähdetään siitä, että tutkijalla on jokin olettamus (teoria, aikaisemmat tutkimukset) joidenkin määrättyjen muuttujien keskeisestä selitysosuudesta selitettävään muuttujaan nähden. Tutkija haluaa testata lähtökohtansa oikeellisuutta. Siten tutkijan asettamat olettamukset joko saavat vahvistusta tai tulevat kumotuiksi. Tässä mielessä analyysin käyttö edellyttää tutkijan tietoista otetta ja hyvää tutkimusalueen tuntemusta.

7.1 Tavallinen regressioanalyysi

Tavallisessa regressioanalyysissa kaikki selittävät muuttujat otetaan mukaan malliin yhdellä kertaa, eikä niitä lisätä tai poisteta missään vaiheessa. Toteutetaan tällainen tyypillinen regressioanalyysi esimerkkiaineiston muuttujien avulla. Ajatuksena on tutkia sitä, miten verbaalisuustestin pistemäärä ja matematiikan arvosana selittävät opintomenestystä lukiossa. Kaikki muuttujat ovat numeerisia, joten siltä osin regressioanalyysi sopii analyysimenetelmäksi. Katsotaan seuraavaksi regressioanalyysin muiden edellytysten voimassaoloa.

Edellä on tutkittu esimerkkiaineiston muuttujien välisiä yhteyksiä korrelaatioanalyysien yhteydessä (luku 8.1), ja siellä todettiin, että verbaalisuustestin pistemäärän ja matematiikan arvosanan välinen korrelaatio r = 0,14 oli heikko. Koska selittävien muuttujien välinen yhteys on näin vähäinen, multikollineaarisuus ei ole ongelma.

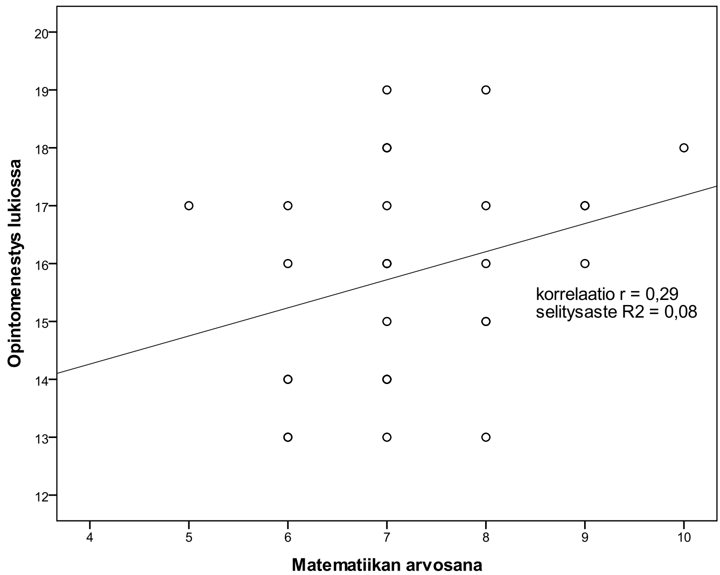
Koska regressioanalyysi tutkii muuttujien välisiä lineaarisia yhteyksiä ja lisäksi muuttujien ääriarvoilla saattaa olla ratkaiseva merkitys analyysin tuloksiin, tarkistetaan seuraavaksi visuaalisella tarkastelulla kuvioiden avulla yhteyden lineaarisuus sekä se, että aineistossa ei esiinny selvästi poikkeavia arvoja. Kuvailevana graafisena tarkasteluna voidaan käyttää hajontakuviota (ts. sirontakuvio tai korrelaatiodiagrammi) sekä siihen lisättyä lineaarista yhteyttä kuvaavaa regressiosuoraa:



KUVIO 17. Verbaalisuustestin ja opintomenestyksen yhteys regressiosuoran ja sirontakuvion avulla tarkasteltuna.

Kuvion havaintopisteissä ei ole havaittavissa mitään käyräviivaista muotoa, joten lineaarisuusehto täyttyy. Lisäksi havaintopisteet sijoittuvat tasaisesti regressiosuoran ympäristöön, joten myöskään haitallisia ääriarvoja ei muuttujissa vaikuttaisi olevan. Kuviosta näkyy selvästi muuttujien välinen voimakas positiivinen korrelaatio. Kuvioon on lisätty myös korrelaatiota ilmaiseva luku ja sen neliö eli selitysaste. Selitysasteesta nähdään se, että aineistossa kyetään selittämään verbaalisuustestin pistemäärän avulla 66 % lukion opintomenestysmuuttujan vaihtelusta!

Tutkitaan vielä vastaavasti matematiikan numeron ja opintomenestyksen välistä yhteyttä sirontakuvion avulla:



KUVIO 18. Matematiikan arvosanan ja opintomenestyksen yhteys regressiosuoran ja sirontakuvion avulla tarkasteltuna.

Sirontakuvionkaan perusteella havaintopisteissä ei ole havaittavissa mitään erityistä käyräviivaista muotoa, joten regressioanalyysin lineaarisuusehdon voidaan katsoa tältäkin pohjalta täyttyvän. Kun vielä havaintopisteet sijoittuvat kohtalaisen tasaisesti (vaikka ovatkin varsin levällään) regressiosuoran ympäristöön, voidaan todeta, ettei tarkasteltavissa muuttujissa ole merkittävästi muista poikkeavia havaintoarvoja. Se, että arvot asettuvat varsin levälleen, aiheutuu siitä, että muuttujien välinen yhteys ei ole kovin vahva. Näin mallin selitysvoima ja ennustearvo eivät ole kovin hyviä.

Mallin selitysasteen R^2 arvosta nähdään, että aineistossa matematiikan arvosanan avulla pystytään selittämään vain kahdeksan prosenttia lukion opintomenestyksestä. Selitysaste, jonka laskenta perustuu mallissa mukana olevien muuttujien korrelaation (tai yhteiskorrelaation) neliöön, ilmaisee kuviossa 18 siis sitä, kuinka suuri osa lukion opintomenestyksestä selittyy tilastollisessa tarkastelussa opiskelijoiden matematiikan taidoista (matematiikan arvosanasta). Kahdeksan prosentin selitysosuus ei ole erityisen suuri. Yleisesti tutkija joutuu tässäkin kohden tekemään itse päätöksen siitä, mikä selitysosuus kulloinkin olisi riittävä mallin hyväksymistä ajatellen. Selitysosuuden suuruuden laskemiseen vaikuttaa paljon se, kuinka monisyistä asiaa ollaan selittämässä. Yleisesti yhteiskunta- ja käyttäytymistieteissä tämä selitysosuus jää aika matalalle, etenkin korrelatiivisissa ja yhteyksiä kartoittavissa tutkimuksissa harvoin tämä nousee yli 20 prosentin. Esimerkiksi laboratoriokokeissa sen sijaan päästään hyvinkin paljon suurempiin selitysosuuksiin. Se, että selitysaste jää monesti suhteellisen matalaksi, johtuu siitä, ettei useinkaan analyysiin saada mukaan kaikkia tutkittavaan muuttujaan vaikuttavia tekijöitä. Tutkijan onkin aina hyvä kysyä itseltään, mitä saatu selitysaste oikeastaan tarkoittaa kunkin analyysin kohdalla. (Ketokivi 2015, 149–150.)

SPSS: Sirontakuvion ja regressiosuoran piirtäminen (kaksi numeerista muuttujaa):

1. Valitse ylhäältä valikkoriviltä vaihtoehdot *Graphs* → *Legacy Dialogs* → *Scatter*
   - esiin tulevista kuvakkeista valitse kuvake *Simple Scatter* ja jatka *Define*-painikkeella eteenpäin
   - valitse muuttujalistasta haluamasi selittävä muuttuja, ja siirrä se listan oikealla puolella olevalla nuolipainikkeella kohtaan *X Axis*
   - selitettävän muuttujan siirrät vastaavalla tavalla kohtaan *Y Axis*
2. *Options*-painike
   - tarkista, että mahdolliset puuttuvat tiedot eivät tule mukaan kuvioon, eli poista valintaruksi kohdasta *Display groups defined by missing values,*
   - hyväksy valinnat klikkaamalla *Continue*-painiketta
3. Paina lopuksi *OK*-painiketta, kuvio piirtyy tulosteikkunaan

Jatka tulosteikkunassa olevan kuvion muokkaamista kaksoisnapsauttamalla kuviota, jolloin saat kuvion muokkaustilaan (Chart Editor)

4. Lisää kuvioon muuttujien välistä lineaarista yhteyttä kuvaava regressiosuora valikkorivin vaihtoehdolla *Elements* ja *Fit Line at TotalTotal*
   - varmista, että valittuna on vaihtoehto *Linear* (tarvittaessa voi valita jonkun toisenkin muodon)
   - korrelaatiokerroin r ja regressiomallin selitysaste R^2 *Linear* tulostuvat automaattisesti kuvioon regressiosuoran yhteyteen
5. Kuvion muokkaustilasta pääset takaisin valitsemalla ylhäältä valikkoriviltä vaihtoehdot *File* ja *Close*.

Yksi regressioanalyysin käytön edellytys on myös se, että mallin niin sanotut jäännökset noudattavat normaalijakaumaa. Tämä voidaan kuitenkin todeta vasta analyysin toteutuksen yhteydessä, joten seuraavaksi voidaan tehdä varsinainen regressioanalyysi. Sen SPSS-tulosteet näyttävät seuraavilta:

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,831[a]	,691	,680	1,03179

a. Predictors: (Constant), Matematiikan arvosana, Verbaalisuustestin pistemäärä
b. Dependent Variable: Opintomenestys lukiossa

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	135,652	2	67,826	63,711	,000[a]
	Residual	60,682	57	1,065		
	Total	196,333	59			

a. Predictors: (Constant), Matematiikan arvosana, Verbaalisuustestin pistemäärä
b. Dependent Variable: Opintomenestys lukiossa

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients			95,0% Confidence Interval for B	
		B	Std. Error	Beta	t	Sig.	Lower Bound	Upper Bound
1	(Constant)	-,67	1,52		-,44	,662	-3,72	2,38
	Verbaalisuustestin pistemäärä	,56	,05	,79	10,57	,000	,46	,67
	Matematiikan arvosana	,30	,12	,18	2,41	,019	,05	,55

a. Dependent Variable: Opintomenestys lukiossa

Tulkitaan tulostetta: *Model Summary* -kohdasta nähdään, että mallin kaksi selittävää muuttujaa selittävät yhdessä 69 % 'opintomenestys lukiossa' -muuttujan vaihtelusta. *ANOVA*-taulussa on testattu mallin tilastollista merkitsevyyttä varianssianalyysista tutulla F-testillä. Malli on tilastollisesti erittäin merkitsevä, F-arvon 63,71 ja vapausasteiden (df) avulla laskettu p-arvo p = 0,000. Ei siis ole epäilystäkään siitä, etteivätkö kaksi muuttujaa pystyisi selittämään lukion opintomenestystä. Tutkimusraportissa testitulos on tapana raportoida muodossa F(2, 57) = 63,71; p < 0,001.

Coefficients-taulukkoon on koottu muuttujakohtaiset tulokset. Regressiokerrointen tilastollista merkitsevyyttä on testattu t-testien avulla[5]. Nähdään,

[5] Regressiokertomiin liittyvä t-testi ei ole sama kuin aiemmin esillä ollut ryhmäkeskiarvoihin liittyvä studentin t-testi. Testin vapausasteen määräytyvät otoskoon ja selittävien muuttujien lukumäärän perusteella df = N-k-1. Tällöin esimerkiksi matematiikan arvosanan tilastolliseen merkitsevyyteen liittyvä tulos voidaan tarvittaessa raportoida Coefficients-taulukon tietojen avulla muodossa: Matematiikan arvosana oli merkitsevä selittäjä lukion opintomenestykselle (t(57) = 2,41; p = 0,019). Tässä siis vapausasteet df = 60-2-1 = 57.

että molemmat muuttujat ovat tilastollisesti merkitseviä selittäjiä (p-arvot ovat alle 0,05). Sarakkeessa B ovat regressiosuoran yhtälön kertoimet. Regressiomalli on siis muotoa (kertoimet on pyöristetty kahteen desimaaliin):

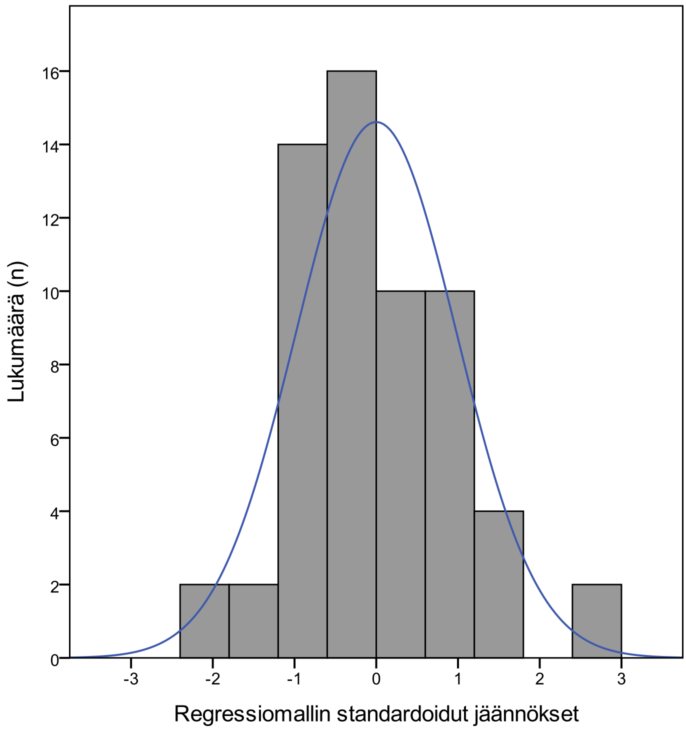
$$y = -0{,}67 + 0{,}56x_1 + 0{,}30x_2$$

eli jos esimerkiksi oppilaan verbaalisuustestin (x_1) tulos on 24 ja saman oppilaan matematiikan arvosana (x_2) on 8, tämän regressiosuoran yhtälön avulla voidaan laskea ennustearvon kyseisen oppilaan opintomenestykselle lukiossa (y). Se olisi y = -0,67 + 0,56 * 24 + 0,30 * 8 = 15,2, eli tämän mukaan kyseisellä opiskelijalla opintomenestys tulisi ennusteen mukaan lukiossa olemaan todennäköisesti keskitasoa (opintomenetyksen kokonaispistevaihteluvälihän on aineistosta 10–20 pistettä). Regressiomallia voidaan käyttää ennustamiseen yleisemminkin, eli jos tiedetään kenen tahansa oppilaan verbaalisuustestin pistemäärä ja matematiikan arvosana, voidaan niiden perusteella laskea regressiomallista ennuste hänen opintomenestykselleen lukiossa.

Taulukossa oikean puoleisimmissa sarakkeissa on regressiokertoimiin liittyvät 95 %:n luottamusvälit, joiden avulla voidaan tulkita estimoinnin tarkkuutta. Jos luottamusväli on kapea, niin saatu regressiokerroin on kyetty estimoimaan tarkasti, ja toisin päin leveä väli kertoo arvioinnin epätarkkuudesta. Esimerkiksi verbaalisuustestiin liittyvä regressiokerroin oli 0,56 ja tälle laskettu 95 %:n luottamusväli on 0,46–0,67. Toisin sanoen arviointiin liittyy 0,11 yksikön virhemarginaali. Virhemarginaali vaikuttaa melko pieneltä, mutta käytännössä tutkija joutuu arvioimaan virhemarginaalin suuruusluokan merkitystä kulloisenkin tilanteen ja tarkasteltavan aiheen mukaan. Mitä pienempi virhemarginaali, sitä luotettavampia tulkintoja voidaan regressiokertoimen perusteella tehdä.

Coefficients-taulukossa on myös sarake *Beta*. Siinä ovat standardoidut regressiokertoimet. Standardointi mahdollistaa regressiokertoimien vertailun. Koska verbaalisuustestin pistemäärään liittyvä kerroin 0,79 on selkeästi suurempi kuin matematiikan arvosanaan liittyvä kerroin 0,18, voidaan päätellä, että verbaalisuustesti on näistä kahdesta selittäjästä vaikutukseltaan voimakkaampi. Näitä kertoimia voidaan käyttää ja tulkita myös efektikoon mittoina: mitä suurempi arvo, sitä suurempi efektikoko.

Vielä on jäljellä yksi tarkistus. Kuten muistetaan, regressioanalyysin käytön edellytys on myös se, että jäännökset (residuals) noudattavat normaalijakaumaa. Jokaiselle havainnolle voidaan laskea jäännösarvo, joka mittaa kyseisen yksittäisen havaintoarvon etäisyyttä regressioanalyysillä tuotetusta ennusteesta (eli mallin mukaisesta arvosta). Mikäli havainto on lähellä mallin tuottamaa ennustetta, jäännösarvo on silloin pieni. Pääjoukosta ja lineaarisesta muodosta poikkeavilla havaintoarvoilla jäännösarvo muodostuu suureksi. Näiden jäännösten eli residuaalien tulisi jakautua normaalijakauman tapaan. Graafisesti jäännöksiä voidaan kuvat histogrammin avulla. Tässäkin kohden on järkevää käyttää standardoituja arvoja (standardized residuals).



KUVIO 19. Regressiomallin standardoitujen jäännösten jakauma suhteessa normaalijakaumaan.

Silmämääräisesti jakauma (pylväikkö) näyttää noudattelevan varsin mukavasti normaalijakauman muotoa (graafiin piirretty normaalijakaumaa kuvaava käyrä). Varminta on tehdä vielä testaus, Kolmogorov-Smirnov -testi (Lillieforsin korjauksella) antaa arvon $p \geq 0{,}200$. Näiden tarkastelujen perusteella voitaisiin siis tulkita, että normaalijakaumaoletus on voimassa.

SPSS: Tavallinen regressioanalyysi

1. Valitse ylhäältä valikkoriviltä *Analyze* → *Regression* → *Linear*
   - varmista, että kohdassa *Method* on valittuna vaihtoehto *Enter*
   - siirrä se valitsemasi numeerinen muuttuja, jota haluat selittää, muuttujaluettelon oikealla puolella olevalla ylemmällä nuolipainikkeella kohtaan *Dependent*
   - ne numeeriset muuttujat, joilla haluat selittää edellä valitsemaasi muuttujaa, siirrät toiseksi ylimmällä nuolinäppäimellä kohtaan *Independent(s)*
2. *Plots*-painike
   - ruksaa kohdan *Standardized Residual Plots* vaihtoehto *Histogram*. Haluamme siis standardoituja jäännöksiä esittävän kuvion,
   - hyväksy valinnat *Continue*-painikkeella
3. Statistics-painike
   - ruksaa avautuvasta valikosta kohta *Confidence intervals*,
   - hyväksy valinnat *Continue*-painikkeella
4. *Save*-painike
   - jos haluat myös testata jäännösten normaalijakautuneisuutta, valitse *Residuals*-kohdan vaihtoehto *Standardized*
   - nyt ohjelma tallentaa analyysin yhteydessä standardoidut jäännökset havaintomatriisin loppuun uudeksi muuttujaksi ja halutessasi voit testata tämän muuttujan normaalijakautuneisuutta luvussa 3 esitetyllä tavalla (s. 101–108),
   - hyväksy valinnat *Continue*-painikkeella
5. Toteuta regressioanalyysi *OK*-painikkeella ja saat analyysin tulokset tulosteikkunaan.

7.2 Askeltava regressioanalyysi

Kasvatus- ja yhteiskuntatieteellisessä tutkimuksessa voi tulla kyseeseen myös tilanne, ettei etukäteen kyetä päättelemään mikä selittävien muuttujin kombinaation pitäisi selittää tutkittavaa ilmiötä. Näin on esimerkiksi silloin, kun kyseessä vähän tutkittua aihepiiriä kartoittavasta tutkimuksesta. Tällöin tutkimuksella pyritään löytämään selitettävää muuttujaa parhaiten selittävät muuttujat ja jättää pois regressiomallista sellaiset selittäjät, joilla on heikko selityskyky tämän suhteen. Tavallisen regressioanalyysin sijasta tällöin tutkijalla on käytössä niin sanottu askeltavaa (stepwise) regressioanalyysi.

Askeltavan regressioanalyysin ensimmäisessä askeleessa mukaan malliin otetaan se selittäjä, jonka korrelaatio selitettävän muuttujan kanssa on korkein; seuraavassa askeleessa mukaan tulee se selittäjä, jonka tuoma selityslisä jäljellä olevista selittäjistä on korkein, ja niin edelleen. Selittäjien lisääminen muuttaa mallissa mukana olevia selitysosuuksia, joten jossain vaiheessa voi käydä niinkin, että mallissa jo mukana ollut selittäjä tippuu sieltä pois. Analyysi päättyy, kun annettujen kriteerien puitteissa yhtään selittäjää ei voida lisätä tai poistaa.

Lisäyksen ja poiston kriteerinä käytetään selittävän muuttujan tilastollista merkitsevyyttä, kun kriteerinä on F-testin p-arvo. SPSS:ssä oletusarvot ovat seuraavat: muuttuja otetaan malliin, jos $p < 0{,}05$; muuttuja poistetaan mallista, jos $p > 0{,}100$. Näitä oletusarvoja voi muuttaa. F-arvoja voisi käyttää myös suoraan, mutta niitä ei kannata suosia, sillä niiden tulkinta ei ole niin selkeää kuin p-arvojen.

Tehdään seuraavaksi askeltava regressioanalyysi. Valitaan esimerkkiaineistosta selitettäväksi muuttujaksi opintomenestys lukiossa ja mahdollisiksi selittäjiksi verbaalisuustestin pistemäärä, matematiikan arvosana, järkeilytestin pistemäärä ja kieliaineiden keskiarvo. Vilkaistaan vielä ennen analyysia kyseisten muuttujien välisiä korrelaatioita, sillä niitä tarkastelemalla voi aavistella, mitä tuleman pitää:

	Opintomenestys lukiossa	Verbaalisuustestin pistemäärä	Matematiikan arvosana	Järkeilytestin pistemäärä	Kieliaineiden keskiarvo
Opintomenestys lukiossa	1,00				
Verbaalisuustestin pistemäärä	,81	1,00			
Matematiikan arvosana	,29	,14	1,00		
Järkeilytestin pistemäärä	,32	,25	,74	1,00	
Kieliaineiden keskiarvo	,83	,91	-,02	,08	1,00

Selittävien muuttujien korrelaatiot selitettävään muuttujaan näkyvät varjostetussa sarakkeessa. Kieliaineiden keskiarvo on suurin, joten askeltavassa regressioanalyysissa se tulee ilmeisesti ensimmäisenä valituksi malliin mukaan.

Selittäjien joukossa huomio kiinnittyy korkeaan korrelaatioon $r = 0{,}91$. Tällaisissa tilanteissa piilee multikollineaarisuuden vaara, sillä selittäjien joukossa ei saisi olla muuttujia, jotka ovat likipitäen samoja. ”Selitysasteeksi” muutettuna 0,91 on 0,83, joten aivan samoja muuttujia verbaalisuustestin pistemäärä ja kieliaineiden keskiarvo eivät kuitenkaan ole.

Askeltavalla regressioanalyysi tuottaa seuraavanlaisen SPSS-tulosteen:

Model Summary[c]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,826[a]	,682	,676	1,038
2	,880[b]	,774	,766	,882

a. Predictors: (Constant), Kieliaineiden keskiarvo
b. Predictors: (Constant), Kieliaineiden keskiarvo, Matematiikan arvosana
c. Dependent Variable: Opintomenestys lukiossa

ANOVA[c]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	133,845	1	133,845	124,232	,000[a]
	Residual	62,488	58	1,077		
	Total	196,333	59			
2	Regression	152,004	2	76,002	97,725	,000[b]
	Residual	44,329	57	,778		
	Total	196,333	59			

a. Predictors: (Constant), Kieliaineiden keskiarvo
b. Predictors: (Constant), Kieliaineiden keskiarvo, Matematiikan arvosana
c. Dependent Variable: Opintomenestys lukiossa

Coefficients[a]

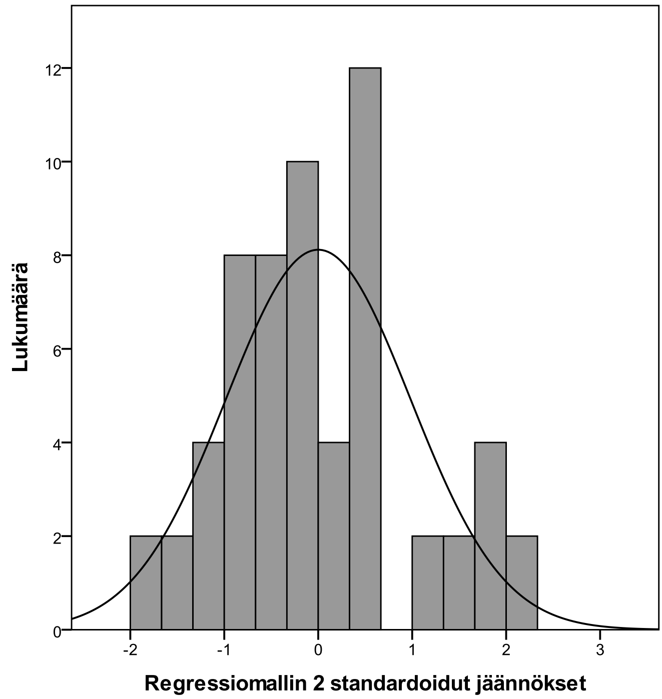
Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	4,174	1,055		3,958	,000
	Kieliaineiden keskiarvo	,146	,013	,826	11,146	,000
2	(Constant)	,443	1,183		,374	,710
	Kieliaineiden keskiarvo	,147	,011	,830	13,192	,000
	Matematiikan arvosana	,507	,105	,304	4,832	,000

a. Dependent Variable: Opintomenestys lukiossa

Kohdassa *Model Summary* on analyysin yhteenveto. Askeleella 1 (Model 1) regressiomalliin on ensimmäisenä tullut mukaan kieliaineiden keskiarvo. Se on nostanut selitysosuuden nollasta arvoon 68 %. Seuraavassa vaiheessa (Model 2) mukaan on päässyt matematiikan arvosana ja selitysosuus on kasvanut lukuun 77 %. Nämä kaksi muuttujaa selittävät siis yhdessä yli kolme neljännestä opintomenestyksestä lukiossa. Muita selittäjiä ei mukaan tulekaan, sillä jäljelle jääneiden muuttujien mukanaan tuoma selitysosuuden lisäys ei ole ollut tarpeeksi hyvä. Tarkasteltujen korrelaatioiden perusteella tämä on ymmärrettävää. Verbaalisuustestin pistemäärä jäi pois, koska se korreloi voimakkaasti malliin mukaan tulleeseen kieliaineiden keskiarvoon. Järkeilytestin pistemäärä taas jäi pois sen tähden, että se korreloi voimakkaasti mukaan päässeeseen matematiikan arvosanaan.

ANOVA-taulukosta nähdään, että kumpikin malli on tilastollisesti erittäin merkitsevä ($p < 0,001$). Kohdassa *Coefficients* vuorostaan on testattu erikseen kunkin malleihin tulleen selittäjän tilastollista merkitsevyyttä. Molempiin muuttujiin liittyvät regressiokertoimet ovat tilastollisesti erittäin merkitseviä ($p < 0,001$). Jos vielä tutkitaan selittäjien keskinäistä paremmuutta, kieliaineiden keskiarvon standardoitu regressiokerroin on 0,83. Se on selkeästi korkeampi kuin matematiikan arvosanan kerroin 0,30. Lisäksi kertoimille voi halutessaan muodostaa luottamusvälit edellä kuvatulla tavalla (ks. s. 200–201). Nämä tulostuisivat taulukon oikeaan reunaan lisäsarakkeina.

Lopuksi voidaan jälleen tutkia mallin standardoituja jäännöksiä, joiden avulla voidaan selvittää regressioanalyysin käyttöön liittyviä oletuksia (ks myös aiheen käsittelyä sivulla 201–202). Jäännösten normaalijakaumaoletuksen paikkansapitävyyttä selvitetään aluksi histogrammi-kuvion avulla.



KUVIO 20. Askeltavalla regressioanalyysilla saadun lopullisen mallin standardoitujen jäännösten jakauma suhteessa normaalijakaumaan.

Jäännöstermien jakaumassa on arvon yksi tuntumassa ikävän näköinen aukko, mutta jos tehdään normaalisuustestaus, Kolmogorov-Smirnov -testi (Lillieforsin korjauksella) antaa p-arvoksi p = 0,067. Normaalijakaumaolettamus voidaan siis kuitenkin hyväksyä. On myös hyvä huomata, että graafissa histogrammin muotoon saattaa vaikuttaa myös luokittelun tiheys vaaka-akselilla. Tässäkin voisi SPSS:n muodostaman oletusluokittelun sijaan koettaa histogrammin toteutusta erilaisilla luokkien lukumäärillä. On mahdollista, että erilaisella luokittelulla jakauman todellinen muoto havainnollistuisi selkeämmin. Kuvion perusteella jäännöksissä ei myöskään ole selvästi poikkeavia jäännösarvoja, vaan arvot sijaitsevat käytännössä välillä ±2.

SPSS: Askeltava regressioanalyysi

1. Valitse ylhäältä valikkoriviltä *Analyze → Regression → Linear*.
   - siirrä se valitsemasi numeerinen muuttuja, jota haluat selittää, muuttujaluettelon oikealla puolella olevalla ylemmällä nuolipainikkeella kohtaan *Dependent*
   - ne numeeriset muuttujat, joilla haluat selittää edellä valitsemaasi muuttujaa, siirrät toiseksi ylimmällä nuolinäppäimellä kohtaan *Independent(s),*
   - valitse seuraavaksi kohtaan *Method* vaihtoehdoksi *Stepwise*
2. Statistics-painike
   - ruksaa avautuvasta valikosta kohta *Confidence intervals*,
   - hyväksy valinnat *Continue*-painikkeella
3. *Plots*-painike
   - ruksaa kohdan *Standardized Residual Plots* vaihtoehto *Histogram*. Haluamme siis standardoituja jäännöksiä esittävän kuvion,
   - hyväksy valinnat *Continue*-painikkeella
4. *Save*-painike
   - jos haluat myös testata jäännösten normaalijakautuneisuutta, paina vielä *Save*-painiketta ja valitse sieltä *Residuals*-kohdan vaihtoehto *Standardized*
   - ohjelma tallentaa analyysin yhteydessä standardoidut jäännökset havaintomatriisin loppuun uudeksi muuttujaksi
   - halutessasi voit testata tämän muuttujan normaalijakautuneisuutta luvussa 3 esittämällämme tavalla (s. 101–108)
   - hyväksy valinnat *Continue*-painikkeella
5. Toteuta askeltava regressioanalyysi *OK*-painikkeella ja saat analyysin tulokset tulosteikkunaan.

7.3 Logistinen regressioanalyysi

Logistinen regressioanalyysi soveltuu tilanteisiin, joissa selitettävä muuttuja Y on kategorinen. Tyypillinen tilanne on sellainen, jossa muuttuja sisältää kaksi luokkaa, jotka on koodattu aineistoon arvoilla 0 ja 1. Tällöin kyse on kaksiluokkaisesta eli binäärisestä logistisesta regressioanalyysistä. Usein arvolla yksi kuvataan tutkimuksen kannalta kiinnostavaa vaihtoehtoa. Esimerkiksi jos tutkittaisiin lukion keskeyttämiseen vaikuttavia tekijöitä, kaksiluokkainen eli dikotominen selitettävä muuttuja voisi olla vaikkapa lukion keskeyttäminen (0 = ei, 1 = kyllä) tai tutkittaessa oppilaan eritysopetuksen tarpeeseen vaikuttavia tekijöitä voisi vastemuuttuja olla osallistunut erityisopetukseen (0 = ei, 1 = kyllä). Selitettävän muuttujan luokkia voi olla myös useita, jolloin puhutaan multinomiaalisesta logistisesta regressioanalyysistä (ks. esimerkiksi Jokivuori & Hietala 2007, 56–77; Gröönroos 2011, 237–240). Selittävät muuttujat x voivat olla jatkuva-arvoisia numeerisia muuttujia tai kaksi- tai useampiluokkaisia kategorisia muuttujia. Erona tavalliseen regressioanalyysiin on se, että nyt malliin pyritään löytämään muuttujia, joilla kyettäisiin selittämään luokkaan 1 kuulumisen todennäköisyyttä. Saatujen tulosten avulla voidaan arvioida jonkin tapahtuman toteutumisen riskiä ja siihen vaikuttavia tekijöitä; lisäksi mallin perusteella voidaan ennustaa esimerkiksi oppilaan kuulumista selitettävän muuttujan luokkaan 1 (tai vastaavasti vertailuluokkaan 0).

Logistisessa regressiomallissa sovelletaan niin sanottuja logistista funktiota, joka mahdollistaa todennäköisyyksien arvioinnin tavallisen regressionanalyysin keinoin. Merkitään kirjaimella Z regressiomallin yhtälöä $Z = a + b_1{*}x_1 + b_2{*}x_2 + \ldots + b_k{*}x_k$, jossa muuttujat $x_1, \ldots, x_k$ ovat malliin valittuja selittäviä tekijöitä ja kertoimet $b_1, b_2, \ldots, b_k$ aineiston perusteella estimoituja regressiokertoimia. Ns. logistisen funktion avulla voidaan tapahtuman *'y saa arvon 1'* todennäköisyys laskea lausekkeesta

$$P(y = 1) = \frac{1}{1 + e^{-Z}}$$

Lausekkeessa esiintyvä *e* on vakio, Neperin luku (e = 2,718). Oletuksena mallissa on, että mikäli tämä todennäköisyys ylittää henkilön (tai muun tutkittavan tilastoyksikön) havaintoarvoilla arvon 0,50, malli luokittelee henkilön luokkaan 1 (esim. hyvä tai erinomainen). Vastaavasti alle 0,50 todennäköisyyksillä henkilö luokittuu luokkaan 0 (esim. heikko tai kohtalainen).

Seuraavassa esimerkissä käydään läpi logistisen regressioanalyysin vaiheet, tulostus ja tulkinnat. Esimerkin analyysi voitaisiin toteuttaa myös askeltavana analyysinä, jossa toteutus ja tulosteet noudattaisivat pitkälti samoja vaiheita kuin aiemmin esitellyssä askeltavassa regressioanalyysissä (luku 7.2., sivu 203). Esimerkkiaineistossa ei ole valmiina sopivaa logistiseen regressioanalyysiin soveltuvaa kaksiluokkaista muuttujaa, mutta tämä voidaan muodostaa esimerkkiä varten uusi kaksiluokkainen opintomenestysmuuttuja OMENE1_LK: opintomenestys lukiossa luokiteltuna (0 = heikko tai kohtalainen, 1 = hyvä tai erinomainen). Näin selitettävänä muuttujana on dikotominen opintomenestysmuuttuja ja selittäjinä käytetään samoja muuttujia kuin edellisen luvun regressioanalyysin esimerkissä. Tavoitteena analyysissä on siis selvittää se, miten muuttujilla matematiikan arvosana ja kieliaineiden keskiarvo kyetään selittämään opintomenetysluokkaan 1 (hyvä tai erinomainen) kuuluminen.

Binääriseen logistiseen regressioanalyysiin liittyvä keskeinen SPSS-tulostus näyttää seuraavalta:

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	44,950	2	,000
	Block	44,950	2	,000
	Model	44,950	2	,000

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	35,812[a]	,527	,713

a. Estimation terminated at iteration number 7 because parameter estimates changed by less than ,001.

Classification Table[a]

	Observed		Predicted: opintomenestys lukiossa (luokiteltuna): heikko tai kohtalainen	Predicted: opintomenestys lukiossa (luokiteltuna): hyvä tai erinomainen	Predicted: Percentage Correct
Step 1	opintomenestys lukiossa (luokiteltuna)	heikko tai kohtalainen	30	6	83,3
		hyvä tai erinomainen	2	22	91,7
	Overall Percentage				86,7

a. The cut value is ,500

Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	2,265	8	,972

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B) Lower	95% C.I.for EXP(B) Upper
Step 1[a]	kieliaineiden keskiarvo	,359	,103	12,256	1	,000	1,432	1,171	1,752
	matematiikan arvosana	1,253	,530	5,579	1	,018	3,500	1,238	9,897
	Constant	-39,138	11,483	11,616	1	,001	,000		

a. Variable(s) entered on step 1: kieliaineiden keskiarvo, matematiikan arvosana

Alkuosan taulukoissa on mallin sopivuutta eli hyvyyttä kuvaavia tuloksia. Ensimmäisessä taulukossa khiin neliö -testi kertoo, poikkeaako malli merkitsevästi mallista, jossa ei olisi lainkaan selittäjiä (ns. tyhjä malli). Merkitsevä tulos ilmaisee, että malli on ainakin tyhjää parempi, $p < 0{,}001$. Tämä olisi tulkittavissa niin, että mallissa on ainakin jokin tai joitain tekijöitä, jolla/joilla hyvää lukion opintomenestystä kyetään selittämään. Analyysi toteutettiin ilman askellusta (koska nyt asetettiin kerralla halutut selittäjät analyysiin), joten kaikki taulukon testit tuottavat identtisen tuloksen. Seuraavassa taulukossa *Model Summary* on tunnuslukuja mallin sopivuudelle. Suureet *Cox & Snell* ja *Negelkerke* vastaavat tulkinnaltaan tavallisen regressioanalyysin selitysastetta. Kyseinen logistinen regressiomalli selittää noin 71 prosenttia muuttujan 'luokiteltu opintomenestys lukiossa' vaihtelusta (Negelkerke R^2), joten mallin selitysaste on varsin hyvä. *Classification Table* -taulukosta nähdään, miten hyvin malli kykenee selittävien muuttujien avulla luokittelemaan tilastoyksiköt (opiskelijat) oikeisiin opintomenestysluokkiin. Opiskelijoita, jotka kuuluvat luokkaan 0 = heikko tai kohtalainen oli yhteensä 36, ja näistä 30 eli 83,3 prosenttia malli luokittelee oikein. Vastaavasti luokkaan 1 = hyvä tai erinomainen kuuluvista opiskelijoista (n = 24) lähes 92 prosenttia luokitellaan oikein. Kaikkiaan malli tuottaa oikean luokituksen 86,7 prosentille havainnoista. Mallin ja aineiston välistä yhteensopivuutta voidaan testata myös Hosmer-Lemeshow -testillä, jossa pieni p-arvo liittyy merkitsevään poikkeavuuteen aineiston ja mallin välillä, ts. heikkoon yhteensopivuuteen. Taulukossa *Hosmer and Lemeshow* p-arvoksi on saatu p = 0,972, jolloin testi ilmentää mallin hyvää yhteensopivuutta.

Taulukko *Variables in the Equation* on tulosten kannalta kiinnostavin. Kuten tavallisessa regressioanalyysissä, se kertoo mitkä mallin selittäjistä ovat merkitseviä ja millaista on muuttujien vaikutus hyvää opintomenestystä selitettäessä. Taulukossa kullekin muuttujalle on tulostunut mallin estimoitu regressiokerroin (sarake *B*), kertoimien keskivirheet (sarake *S.E.*) sekä Waldin testi kertoimien tilastolliselle merkitsevyydelle (sarakkeet *Wald*, *df* ja *Sig.*). Molemmat malliin valitut *x*-muuttujat ovat tilastollisesti merkitseviä selittäjiä (p-arvot ovat alle 0,05: p < 0,001 ja p = 0,018). Regressiokertoimen etumerkki ilmaisee vaikutuksen suuntaa: negatiivinen kerroin pienentää opintomenestysluokkaan 1 kuulumisen riskiä ja positiivinen kerroin lisää sitä. Kertoimien avulla on laskettavissa henkilön todennäköisyys kuulua luokkaan 1 (1 = hyvä tai erinomainen opintomenestys). Esimerkiksi henkilölle, jonka kieliaineiden keskiarvo on 80 ja matematiikan arvosana on 7, saadaan todennäköisyydeksi kuulua opintomenestysluokkaan 1:

$$P(y=1) = \frac{1}{1+e^{-(-39{,}138+0{,}359*80+1{,}253*7)}} = 0{,}16$$

Malli luokittaisi kyseisen opiskelijan luokkaan 0 (0 = heikko tai keskinkertainen opintomenestys), sillä todennäköisyys jää alle arvon 0,50. Todennäköisyysarvon lisäksi tapahtuman y = 1 esiintymisen yleisyyttä voidaan ilmaista myös vedonlyönnistä peräisin olevalla riskisuureella Veto eli Odds, joka saadaan todennäköisyyksien osamääränä siten, että Odds = todennäköisyys kuulua luokkaan 1/todennäköisyys, että ei kuulu luokkaan 1. Esimerkin arvoilla:

$$Odds = \frac{0{,}16}{1 - 0{,}16} = 0{,}19$$

Tästä päästään erääseen logistisen regressioanalyysin tulkinnan kannalta tärkeään suureeseen Odds Ratio (OR). Tunnuslukua on suomennettu muotoon riskisuhde tai vetosuhde (ks. esimerkiksi Töttö 2012, 113–117). Usein analyysin tulosta tulkitaan juuri OR-arvojen avulla, ja yleensä ne myös raportoidaan tutkimusraporttiin. OR-luvuilla saadaan ilmaistua kunkin selittävän muuttujan vaikutuksen voimakkuus logistisessa regressioanalyysissä. Jatketaan edellä ollutta esimerkkiä niin, että olkoon jollakin toisella opiskelijalla kieliaineiden keskiarvo myös 80 ja matematiikan arvosana yhtä numeroa korkeampi 8. Hänelle saataisiin vastaavasti laskien Odds:

$$Odds = \frac{0{,}40}{1 - 0{,}40} = 0{,}67$$

Odds Ratio (*OR*) olisi näiden saatujen *Odds*-lukujen osamäärä:

$$OR = \frac{0{,}67}{0{,}19} = 3{,}50$$

Luku voidaan tulkita siten, että yhden yksikön lisäys muuttujassa x_2 = matematiikan arvosana lisää mahdollisuutta (Odds) kuulua parempaan opintomenestysluokkaan 3,50-kertaiseksi. SPSS-tulosteissa tämä arvo nähdään viimeisen taulukon sarakkeessa *Exp(B)*. Taulukossa lukuarvon viereen tulostuu vielä suureen luotettavuutta kuvaava 95 prosentin luottamusväli, jonka mukaan esimerkin OR on 95 prosentin varmuudella jotakin väliltä 1,24–9,90. Myös toinen mallin selittäjä, kieliaineiden keskiarvo, on tilastollisesti merkitsevä ($p < 0{,}001$), OR = 1,43 ja 95 %:n luottamusväli [1,17 , 1,75].

Esimerkissä molemmat selittäjät eli x-muuttujat olivat määrällisiä. Mallissa voisi käyttää myös kategorisia kaksi- tai useampiluokkaisia muuttujia, ja silloinkin tulosten tulkinta sujuisi pääpiirteissään edellä kuvatulla tavalla. Kaksiluokkaisella selittäjällä toinen luokista asetetaan vertailuryhmäksi, jolloin muuttujalle estimoitu OR kuvaa toisen ryhmän riskiä verrattuna tähän vertailu- eli referenssiluokkaan. Mikäli esimerkiksi mallin selittäviin muuttujiin lisättäisiin kategorinen muuttuja sukupuoli (0= miehet, 1= naiset) ja mallin tulostukseen sukupuolimuuttujalle saataisiin OR = 1,61, tarkoittaisi se, että naisilla on 1,61 kertainen riski (Odds) päätyä hyvän opintomenestyksen ryhmään verrattuna miehiin.

SPSS: Logistinen regressioanalyysi, kun selitettävä muuttuja on kaksiluokkainen

1. Valitse ylhäältä valikkoriviltä *Analyze* → *Regression* → *Binary Logistic*.
   - siirrä valitsemasi kategorinen kaksiluokkainen muuttuja, jota haluat selittää, muuttujaluettelon oikealla puolella olevalla ylemmällä nuolipainikkeella kohtaan *Dependent*
   - ne muuttujat (numeeriset tai kategoriset), joita haluat käyttää mallissasi selittäjinä, siirrät toiseksi ylimmällä nuolinäppäimellä kohtaan *Covariates*
2. *Categorical*-painike
   - mikäli valituissa selittäjissäsi on kategorisia muuttujia, siirrä valintaikkunaan *Categorical Covariates*.
   - saman valintaikkunan alalaidassa olevalla valinnalla *Reference Category* voit asettaa haluamasi luokka-arvon vertailuryhmäksi, hyväksy valinnat *Continue*-painikkeella
3. *Options*-painikkeella voidaan valita lisätulosteita. Monesti kannattaa valita:
   - yhteensopivuustesti *Hosmer-Lemeshow Goodness-of-fit*
   - *OR*-vetosuhteille luottamusvälit *CI for exp(B),* hyväksy valinnat Continue-painikkeella
4. *Save*-painike
   - tällä saat tallennettua jäännöstermeihin ja poikkeaviin havaintoihin liittyviä suureita aineistoosi jatkotarkasteluja varten
   - standardoidut jäännösarvot tallentuvat aineistoon omaksi muuttujakseen valinnalla *Residuals: Standardized*, hyväksy valinnat *Continue*-painikkeella
5. Toteuta logistinen regressioanalyysi *OK*-painikkeella ja saat analyysin tulokset tulosteikkunaan.

8. Muuttujien ryhmittely faktori- tai pääkomponenttianalyysillä

Faktori- ja pääkomponenttianalyysiä käytetään yleisesti havaintoaineiston muuttujien määrän tiivistämiseen. Analyysien tavoitteena on muodostaa mittarissa käytetyistä muuttujista (monesti väittämiä tietyn ilmiön eri puolista) suurempia ja helpommin tulkittavia kokonaisuuksia. Jos meillä olisi kyselylomakkeessa 30 liikuntaan ja liikuntaharrastukseen liittyvää väittämää, voitaisiin näistä muodostaa pääkomponenttianalyysiä hyödyntäen esimerkiksi viisi uutta muuttujaa, jotka jäsentäisivät liikunnanmerkitystekijät mielekkäämmin eri osatekijöihin samalla, kun kokonaisuudesta tulisi helpommin tulkittava (ks. kuvio 21, s, 219). Lisäksi faktorianalyysin avulla voidaan selvittää faktorimallin rakennetta eli muuttujien välisiä suhteita ennalta määrättyjen hypoteesien pohjalta. Analyysia on käytetty paljon myös mittarien laadintaan, mittarien validointeihin ja niiden reliabiliteetin arvioimiseen. Jos faktorianalyysissa testataan ennalta asetettuja selkeitä hypoteeseja, kyse on konfirmatorisesta faktorianalyysista (ks. Leskinen 1987; Brown 2015); muussa tapauksessa faktorianalyysi on eksploratiivista (kartoittavaa). Tulevassa keskitytään eksploratiiviseen analyysiin, jossa esimerkkinä käydään läpi pääkomponenttianalyysiin liittyviä vaiheita ja tuloksia.

Pääkomponentti- (PCA) ja faktorianalyysi (FA) muistuttavat toisiaan, mutta menetelmien taustalla oleva filosofia ja tilastomatemaattinen toteutus poikkeavat toisistaan. Faktorianalyysin taustalla on malli, jossa muun muassa muuttujiin liittyvästä kokonaisvaihtelusta pyritään erottamaan mittausvirheestä johtuva vaihtelu. Lisäksi ajatellaan, että taustalla olevat latentit tekijät, eli faktorit ja niiden välinen rakenne eli yhteydet, muodostavat havaitut muuttujat. Tavoitteena on muodostaa faktorit, joilla saadaan riittävän hyvin selitettyä havaittujen muuttujien sisältämää yhteistä vaihtelua. Pääkomponenttianalyysissä tavoitteena on muodostaa matemaattinen ratkaisu, jossa joukolla pääkomponentteja (muuttujista muodostettuja lineaarikombinaatioita), kyettäisiin selittämään riittävän suuri osuus muuttujien sisältämästä kokonaisvaihtelusta. Käytännössä pääkomponenttianalyysiä ja exploratiivista faktorianalyysia voidaan soveltaa hyvin samanlaisiin tarkoituksiin. Sen sijaan konfirmatorisen faktorianalyysin käyttö eroaa näistä siinä, että analyysin toteutus edellyttää, että tutkijalla on etukäteen taustateoriaan tai aiempaan tutkimukseen perustuvaa tietoa faktorirakenteen luonteesta (ts. havaittujen, analyysiin mukaan tulevien muuttujien ryhmittymisestä faktoreille, faktorien lukumäärästä ja niiden välisistä riippuvuuksista). Faktorianalyysin ja pääkomponenttianalyysin käytön suhteen voisikin tältä pohjalta muotoilla seuraavanlaisen nyrkkisäännön:

1. käytä faktorianalyysia, jos tavoitteena on testata teoreettista mallia tai testata analyysissa mukana olevien muuttujien taustalla vaikuttavien latenttien tekijöiden teoreettista mallia

2. sovella pääkomponenttianalyysiä, jos haluat yksinkertaisesti tiivistää tai vähentää havaittujen muuttujien määrää muuttujajoukkoon, jotka muodostavat keskeisiä, toisistaan riippumattomia muuttujakokonaisuuksia. (Anglim 2020)

Se kumpaa menetelmää käytetään, riippuu siis siitä millaisia perusolettamuksia tutkijat asettavat analyysissa mukana olevien muuttujien suhteille. Käytännössä molemmat analyysit tuottavat samankaltaisia muuttujien uudelleen ryhmittelymalleja.

Yleisiä analyysin lähtökohtia ja periaatteita:

- monimuuttujamenetelmä
- perustuu Pearsonin (tulomomentti) korrelaatiokertoimista muodostettuun korrelaatiomatriisiin
- muuttujien tulee olla vähintään välimatka-asteikollisia
- varsinainen faktorianalyysi on tilastolliseen malliin perustuva monimuuttujamenetelmä, kun taas pääkomponenttianalyysi on kartoittavampi laskennallinen menetelmä, jossa muun muassa jakaumaoletukset eivät ole niin tiukkoja kuin faktorianalyysissa
- muuttujien tulee täyttää normaalijakaumaoletus (pääkomponenttianalyysissä tämä ei ole ehdoton vaatimus)
- muuttajissa mukana ei saisi olla suuresti poikkeavia arvoja
- vaatii suurehkon tai suuren aineiston. Yleisesti esitetään minimiksi noin 100 havaintoyksikkö, ja suositeltava aineiston koko olisi 200 tai tätä suurempi (riippuen myös analyysissä mukana olevien muuttujien määrästä). ks. yksityiskohtaisempi esitys suositeltavasta otoskoos-

ta ja muista menetelmän käytön perusoletuksista esim. Field (2018, 797–800).

8.1 Faktorianalyysin peruslähtökohtia

Eroavaisuuksistaan huolimatta faktori- ja pääkomponenttianalyysi kuuluvat läheisesti samaan malliperheeseen, jossa analyysien käytännön toteutus ja tulosteiden tulkinta ovat hyvin samankaltaisia. Tästä syystä menetelmien taustaa ja soveltamista esitellään tässä luvussa yhteisesti faktorianalyysinä. On kuitenkin huomattava, että tässä esiintyvät faktoreihin liittyvät suureet voitaisiin pääosin korvata myös pääkomponentteihin liittyvillä vastaavilla termeillä, kuten pääkomponenttianalyysin esimerkissä luvussa 8.2 huomataan.

Faktorianalyysin tavoitteena on siis tiivistää ja jäsentää analyysissa mukana olevat muuttujat uudelleen eli alkuperäiset muuttujat (esim. väittämät) pyritään korvaamaan pienemmällä määrällä uusia muuttujia, jotka kuitenkin säilyttävät mahdollisimman suuren osan alkuperäisten muuttujien yhteisestä vaihtelusta. Analyysi perustuu muuttujien välisiin korrelaatioihin, eli samaan uuteen muuttujaan valikoituvat alkuperäisistä muuttujista ne, joihin vastaajat ovat vastanneet samansuuntaisesti. Korrelaatioon liittyviä periaatteita on käsitelty aikaisemmin, luvussa 6. Faktorien lukumäärän määräämiseen ei ole olemassa mitään yksioikoista sääntöä. Tutkimuksellisesta näkökulmasta onnellisin tilanne olisi se, että käytettävien faktoreiden määrän pystyisi valitsemaan tutkittavaan ilmiöön liittyvien aikaisempien tutkimusten tai ilmiöön liittyvän teorian perusteella.

Faktorien lukumäärästä päättäminen on tärkeä kysymys analyysin ja tutkimuksenkin onnistumisen kannalta. On tärkeää, ensinnäkin, ettei muodosteta sellaisia faktoreita, jotka eivät liity keskeisinä tekijöinä tutkittavaan ilmiöön, tai ettei muodosteta sellaisia faktoreita, jotka eivät sisällä todellista vaihtelua ja joiden selitysosuus jää näin ollen mitättömäksi. Faktorien lukumäärän määrittämiseksi on joitakin yleisperiaatteita. Yleisin näistä on niin sanottu ominaisarvokriteeri (eigenvalue). Analyysissä kullekin faktorille voidaan laskea ominaisarvo, joka kuvaa määrää, jonka faktori pystyy selittämään muuttujien vaihtelusta. Suuri ominaisarvo kuvaa faktorin tärkeyttä. Ominaisarvokriteerin mukaan minimiarvoksi ominaisarvoille asetetaan arvo yksi: jos faktorin ominaisarvo ei ylitä tätä, se voidaan jättää faktorimallista pois. Tai toisin päin ilmaistuna, faktoriratkaisuun otetaan mukaan vain faktorit, joiden ominaisarvo on yli yksi. Ominaisarvojen summa on sama kuin analyysissa mukana olevien muuttujien lukumäärä. Kuhunkin faktoriin liittyvä ominaisarvo kertoo sen, kuinka hyvin kyseinen faktori on pystynyt uuttamaan havaittavien muuttujien varianssin itseensä. Jos esimerkiksi faktoriin liittyvä ominaisarvo on 3,1 ja havaittavia muuttujia on viisi kappaletta, faktori on onnistunut tehtävässään hyvin, sillä muille faktoreille on jäänyt uutettavaa vain 1,9. Jos ominaisarvo jaetaan havaittavien muuttujien lukumäärällä ja kerrotaan luvulla sata, saadaan selville se, kuinka monta prosenttia kukin faktori selittää havaittavien muuttujien vaihtelusta, eli mikä on kunkin faktorin selitysosuus prosentteina. Esimerkkifaktorin selitysosuus prosentteina on 62 % eli (3,1/5)*100. Kumulatiivinen ominaisarvo (faktoreiden ominaisarvojen summa) ilmentää sitä, kuinka paljon faktorit yhdessä selittävät mitatusta alueesta eli siis havaittujen muuttujien vaihtelusta. Jos tämäkin luku jaetaan havaittavien muuttujien lukumäärällä ja kerrotaan sadalla, saadaan selitysosuus prosentteina.

Palataan takaisin itse uusien faktorien määrään liittyvään ratkaisuun. Jos ominaisarvo ylittää raja-arvo yhden, tarkoittaa tämä pelkistetysti sanoen sitä, että kyseinen faktori on onnistunut perustehtävässään eli muuttujien lukumäärän pienentämisessä. Onnistuminen on sitä parempaa, mitä suurempi ominaisarvo on. Ominaisarvokriteeri yksi on usein riittävä, ja siksi monet tilasto-ohjelmistot käyttävät sitä oletuskriteerinään. Faktoreihin ei myöskään kannata ottaa mukaan sellaista komponentteja, joiden muuttujalataukset jäävät kauttaaltaan hyvin mataliksi. Faktorimäärien määrittelemiseen voidaan käyttää myös Cattellin scree-testiä (ks. esim. Child 1973, 43–44). Siinä piirretään suurimmasta pienimpään järjestettyjen ominaisarvojen mukainen viivadiagrammi, niin sanottu scree plot -kuvio, ja kuviossa mahdollisesti esiintyvän taitekohdan perusteella voidaan valita faktoreiden lukumäärä. Lisäksi yleisesti esitetään, ettei faktorimääräksi saisi valita enempää kuin 20 prosenttia analyysissa mukana olevasta muuttujamäärästä. Tärkeintä on se, että faktoriasetelmaan otetaan mukaan vain ne faktorit, joilla on riittävästi selityskykyä ja joiden tulkinta on mielekäs asetetun tutkimusongelman suhteen. Kun faktoreiden lukumäärä on selvillä, lasketaan jokaisen analyysissa mukana olevan muuttujan painoarvo eli lataus (loading) kunkin faktorin kohdalla. Lataukset kertovat sen, kuinka voimakkaasti muuttujat kuuluvat eli latautuvat eri faktoreille.

Faktoroinnin yhteydessä kannattaa yleensä suorittaa myös niin sanottu rotaatio. Koska faktorit voidaan ajatella koordinaatistoon, jossa kukin faktori on oma akselinsa, rotatointi tarkoittaa tämän akseliston kiertämistä tai kääntämistä. Rotaation avulla pyritään siihen, että kukin muuttuja latautuisi mahdollisimman selvästi vain yhdelle faktorille. Rotaation avulla päästään pääkomponenttiesityksen yksinkertaistamiseen. Rotaatio voidaan tehdä joko suorakulmaisen (varimax rotaatio) tai vinokulmaisen (Oblimin rotaatio) ratkaisun mukaisesti. Rotatoinnilla saatua ratkaisua selkeytetään niin, että faktoreille tulee sekä mahdollisimman korkeita, että mahdollisimman alhaisia latauksia. Näin faktorin sisällä olevien muuttujien latausten vaihtelu saadaan mahdollisimman suureksi. Suorakulmaisuus eli kohtisuoruus tarkoittaa sitä, että ratkaisussa faktorit ovat riippumattomia toisistaan, eli ne eivät korreloi. Suorakulmaisista rotaatioista yleisin on varimax-ratkaisu. Vinokulmainen tarkoittaa sitä, että faktoriakselit eivät ole kohtisuorassa toisiinsa nähden, jolloin faktorien sallitaan korreloida keskenään. Yleensä pyritään faktoreihin, jotka ovat riippumattomia toisistaan. Tällöin niiden tulkinta on selkeämpää.

Faktorianalyysin tulkinnan kannalta on esille nostettava ja kerrattava vielä muutamia keskeisiä käsitteitä: faktorilataukset (loadning), faktorien selitysasteet (% of Variance) ja kommunaliteetit. Nämä kaikki auttavat tutkijaa tekemään päätöksiä faktorianalyysin tarjoaman mallin tarkoituksenmukaisuudesta ja käyttökelpoisuudesta. Ensinnäkin muuttujan latausarvo ilmaisee kunkin muuttujan suhdetta (korrelaatiota) kyseiseen faktoriin. Latausarvosta voidaan laskea myös se, kuinka suuren osan faktori selittää tietyn muuttujan varianssista, eli jos jonkin muuttujan lataus on 0,90, selittää kyseinen faktori sen varianssista (vaihtelusta) 81 %. Samassa faktorissa lataukset voivat olla joko negatiivisia (latauksen etumerkki negatiivinen) tai positiivisia (etumerkki positiivinen). Jos näiden molempien arvot ovat korkeita, voidaan faktori tulkita kaksidimensioiseksi, jonka positiiviset ja negatiiviset ääripäät edustavat toisilleen vastakkaisia ulottuvuuksia (esim. tehtävä- ja välttämisorientoituneisuus).

Havaitulle muuttujalle voidaan laskea niin sanottu kommunaliteetti laskemalla yhteen kyseisen muuttujan eri faktoreilla saamien latausten (painokertoimien) neliöt. Kommunaliteetti kertoo sen, kuinka suuren osuuden asetelmassa olevat faktorit selittävät kyseisestä muuttujasta, eli kuinka paljon kyseisen muuttujan vaihtelusta faktorit ovat pystyneet uuttamaan itseensä. Kommunaliteetti voi saada arvoja nollasta yhteen. Tämä luku kerrottuna sadalla ilmaisee selitysosuuden prosenteissa.

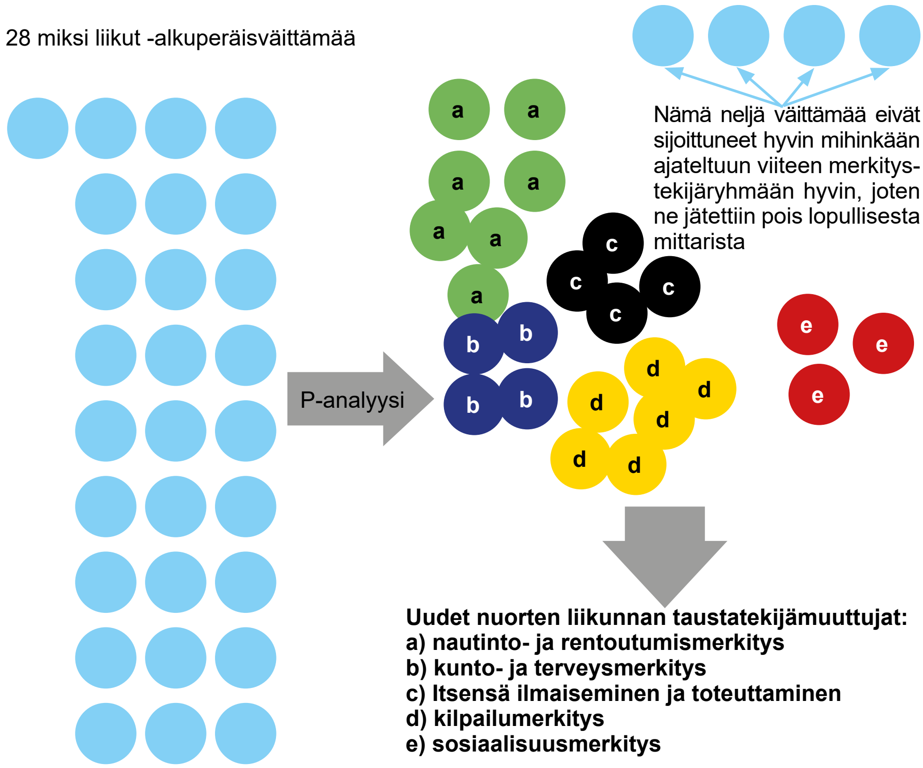
Saadun faktoriratkaisun tulkinta voi toisinaan olla haasteellista. Ei riitä, että ohjelmasta saadaan ulos jonkinlainen ratkaisu, vaan tämä tulee myös kyetä tulkitsemaan. Aluksi tulkinnassa kannattaa pohtia, miten looginen ja tarkoituksenmukainen ratkaisu on. Kyse on sekä ratkaisun sisällön tarkastelusta että saatujen uusien faktorien nimeämisestä. Faktorien nimeämisen avuksi voidaan antaa seuraavanaisia perusvinkkejä ja -kriteereitä:

- faktori nimetään sen muuttujan mukaan, jolla on faktorissa suurin lataus (ei aina suositeltavaa, pelkistää faktoria ehkä liiaksi)
- faktori nimetään niiden muuttujien mukaan, jotka ovat saaneet faktorissa suurimpia latauksia
- faktorit voidaan nimetä myös niiden prosessien avulla, joilla on ollut osuutta faktorien syntyyn
- edellisiä relevantimmaksi faktorin nimeäminen ja analysointi tulee, jos ne perustuvat aikaisempiin tutkimuksiin ja teoreettisiin lähtökohtiin. Näin päästään faktorin kuvausta pidemmälle, saadaan käyttöön teoreettisia käsitteitä, joiden avulla faktoria voidaan selittää (Sänkiaho 1974, 18–19).

Kaiken kaikkiaan faktoreiden tulkintaan ja siihen, miten saatuja faktoreita tutkimuksessa kyetään hyödyntämään, tulisi kiinnittää erityistä huomiota, koska faktorianalyysin käyttö saattaa muuten jäädä kovin mekaaniseksi. Esimerkiksi faktoreiden käytön rajoittaminen mittarien reliaabeliuden toteamiseen ei oikeastaan ole kovin relevanttia (tähän on muitakin testejä). Faktorianalyysi soveltuu erityisen hyvin aineistossa olevien eri ominaisuuksien tai ulottuvuuksien määrittelyyn, jonka perustalta meidän tulisi jatkaa aineiston varsinaista analyysia eikä lopettaa sitä tähän (ks. esim. Sänkiaho 1974; Eskola 1975, 266–270). Voidaan esimerkiksi luoda tältä pohjalta muuttujista kokoavia summamuuttujia (tätä tarkastellaan myöhemmin tarkemmin). Näin saatuja muuttujia voidaan sitten käyttää jatkoanalyyseissa hyödyksi. Jatkoanalyysia voidaan tehdä myös analyysin tuottamilla faktoripistemäärä-muuttujien (factor scores) avulla ja niin edelleen.

8.2 Pääkomponenttianalyysiesimerkki

Tässä kirjassa käsitellään yksityiskohtaisemmin faktorianalyysiryhmään kuuluvista menetelmistä pääkomponenttianalyysiä, joka on edelleen varsin yleisesti käytetty menetelmä kasvatustieteissä. Menetelmää käytetään erityisesti silloin, kun muuttujien määrää halutaan vähentää ilman erityistä oletusta muuttujien taustalla olevasta teoriasta (ks. s. 214 esitetyt valintaperusteet). Seuraavaan kuvioon on vielä tiivistetty graafisessa muodossa faktori- ja pääkomponenttianalyysin perusajatus eli suuren muuttujamäärän (väittämien) jäsentämisen ja tiivistämisen periaate:



KUVIO 21. Mittarin laadinta esimerkki pääkomponenttianalyysillä.

Kuvion pääkomponenttianalyysiesimerkissä, joka ei liity kirjan esimerkkiaineistoon, oli alkujaan 28 erilaista väittämää lasten ja nuorten oman liikunnan taustalla olevista merkitystekijöistä (motiivitekijöistä), siis tekijöistä, joiden vuoksi he harrastivat jossakin muodossa liikuntaa. Kaikkien yksittäisten väittämien käsittely olisi työlästä, jos ei ihan mahdotonta niin tutkijalle kuin lukijoillekin. Pääkomponentti- tai faktorianalyysi antavatkin toimivan ratkaisun jäsentää ja tiivistää näiden 28 väittämien muodostama patteristo harvempiin ja laaja-alaisempiin uusiin muuttujiin. Esimerkkitapauksessa tutkijat olivat aikanaan laatineet nämä 28 kysymystä niin, että näiden ajateltiin alkujaankin mittaavan viittä omalle liikunnalle annettavaa merkitysulottuvuutta aiempien tutkimusten ja tutkittavaan ilmiöön liittyvien teorioiden perusteella (ks. kuvio 21). Koska tutkijoilla oli tiedossaan nämä viisi laajempaa omalle liikunnalle annettava taustamerkitysluokkaa, käytettiin tutkimuksessa konfirmatorista faktorianalyysia (ei siis pääkomponenttianalyysilla, jonka jäsentely kuitenkin olisi samanlainen kuin kuviossa on esitetty) testatakseen sitä, kuinka heidän aineistossaan tämä viisiluokkainen mittariston todentuisi. Tässä vaiheessa tutkijat vielä muokkasivat mittaristoaan, koska neljä alkuperäiskysymyspatteristoon kuuluvaa väittämää eivät sopineet analyysin perusteella hyvin suunniteltuun mittaristorakenteeseen. Näin saatiin, kuten kuviosta 21 ilmenee, alkuperäisen 28 väittämän sijaan viisiulotteinen oman liikunnan taustamerkityksiä mittaava mittaristo, joka on alkuperäistä väittämäkokonaisuutta paljon helpommin käsitettävissä, tulkittavissa ja myös astetta yleisluontoisempi mittari. Sen, miten tutkijat hyödynsivät kyseistä uutta mittaria, voit lukea vuonna 2017 julkaistuista *Liikkumattomuuden jäljillä* -teoksesta (Vanttaja, Tähtinen, Zacheus & Koski 2017, 58–73.)

Palataan nyt kirjassa käytettyyn aineistoesimerkkiin ja suoritetaan esimerkkiaineistosta valituille muuttujille faktorianalyysiperheeseen kuuluva pääkomponenttianalyysi. Tämän aineiston kohdalla pääkomponenttianalyysiesimerkki on hiukan ongelmallinen, koska otoksen koko ja muuttujien lukumäärä ovat verrattain pienet. Lisäksi aineistossa ei ole kuin kaksi merkittävää ulottuvuutta, kielellisyys ja matemaattisuus-loogisuus, jotka voivat muodostaa relevantit faktorit. Kun myös muuttujia on vähän, faktoreiden selitysosuuksilla on taipumus maksimoitua, eli muuttujien varianssi tulee selitetyksi melkein kokonaan kahden faktorin avulla. Aineiston käyttö tässä esimerkkinä on kuitenkin puolusteltavissa sillä, että se on tullut tutuksi muiden analyysien yhteydessä. Näin lukijan on helpompi ymmärtää, mistä faktorianalyysissa oikein on kyse, ja hahmottaa sen pääperiaatteet.

Aineistossa on kuusi numeerista muuttujaa. Pääkomponenttianalyysiin niistä valittiin kuitenkin vain neljä, eli opintomenestysmuuttujat jätettiin tästä pois. Opintomenestysmuuttujat eivät oikein sovi suunniteltuun asetelmaan. Tarkoituksena on siis selvittää, löytyisikö aineistosta kaksi ulottuvuutta: kielellisyyden ja matemaattisuus-loogisuuden pääkomponentit. Valitaan analyysiin mukaan verbaalisuustestin pistemäärä, matematiikan arvosana, järkeilytestin pistemäärä ja kieliaineiden keskiarvo. Tulkoon vielä kerran mainituksi, että tässä muuttujien määrä on kovin pieni verrattuna tyypillisiin pääkomponentti- ja faktorianalyysien sovelluksiin. Aloitetaan analyysiin liittyvät tarkastelut katsomalla ensiksi mukaan valittujen muuttujien välisiä korrelaatioita. Näin saadaan seuraavanlainen korrelaatiomatriisi:

	Verbaalisuustestin pistemäärä	Matematiikan arvosana	Järkeilytestin pistemäärä	Kieliaineiden keskiarvo
Verbaalisuustestin pistemäärä	1,00			
Matematiikan arvosana	0,14	1,00		
Järkeilytestin pistemäärä	0,25	0,74	1,00	
Kieliaineiden keskiarvo	0,91	-0,02	0,08	1,00

Jo näiden neljän muuttujan korrelaatioista näkyy selvästi toisaalta kielellisyyteen, toisaalta matemaattisuuteen perustuva parittaisuus: verbaalisuustestin pistemäärä ja kieliaineiden keskiarvo liittyvät vahvan positiivisesti toisiinsa, samoin matematiikan arvosana ja järkeilytestin pistemäärä. Tässä esimerkissä korrelaatiot sattuivat olemaan positiivisia, mutta myös vahva negatiivinen korrelaatio olisi merkityksellinen. Nyt kun muuttujien väliset korrelaatiot on todettu, ja lisäksi tiedetään kysymystä koskevan teoreettisen tiedon perusteella, kuinka monta pääkomponenttia halutaan, joten seuraavaksi ajetaan varsinainen analyysi ja käydään läpi SPSS-tulosteita ja niiden tulkintaa.

Aluksi tulostuu kaksi korrelaatiomatriisiin liittyvää testiä, joiden avulla voidaan selvittää aineiston sopivuutta faktori- tai pääkomponenttianalyysiin.

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		,504
Bartlett's Test of Sphericity	Approx. Chi-Square	156,895
	df	6
	Sig.	,000

Kaiser-Meyer-Olkin -indeksin (KMO) kriteeriarvoksi on toisissa lähteissä esitetty 0,50–0,60 (Kaiser & Rise 1974): mikäli saatu lukuarvo on 0,50 tai alle, muuttujien välisiä korrelaatioita ei voida selittää muiden muuttujien avulla, eikä aineisto näin sovellu faktori- tai pääkomponenttianalyysiin. Kaiserin mukaan tällöin aineisto ei soveltuisi pääkomponenttianalyysiin (unacceptable). Toisaalta on hyvä muistaa, että tällaisia tiukkoja raja-arvoja ei nykyisin pidetä erityisen relevantteina. Silti tämän tapaisia kriteeriarvoja on hyvä pitää mielessä, kun arvioit oman aineistosi sopivuutta kyseiseen analyysiin. Esimerkissä KMO-arvo on 0,50. Arvo ei ole erityisen korkea mutta sen mukaan pääkomponenttianalyysiä voidaan ja kannattaa soveltaa. Bartlettin testissä (Bartlettin sfäärisyystesti) pieni p-arvo ilmaisee, että analyysissä olevien muuttujien välillä esiintyy nollasta poikkeavia korrelaatioita. Tässä on saatu erittäin merkitsevä testitulos (p = 0,000), mikä myös puoltaa pääkomponenttianalyysin käyttöä.

Total Variance Explained

Component	Initial Eigenvalues			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	2,07	51,83	51,83	1,91	47,85	47,85
2	1,60	39,90	91,74	1,76	43,88	91,74
3	,26	6,40	98,13			
4	,07	1,87	100,00			

Extraction Method: Principal Component Analysis.

Tässä taulussa on paljon asiaa. Edetään sarake kerrallaan, *Component*-sarakkeessa olevat luvut tarkoittavat kaikkia mahdollisia pääkomponentteja. Koska analyysiin otettiin mukaan neljä havaittavaa muuttujaa, pääkomponenttejakin voi enimmillään olla neljä. *Initial Eigenvalues Total* -sarakkeessa ovat kuhunkin komponenttiin liittyvät ominaisarvot. Ominaisarvojenkin summa on neljä. Komponentit on sijoitettu ominaisarvojen perusteella paremmuusjärjestykseen, joten nähdään, että kaksi ensimmäistä pääkomponenttia ovat ylivertaisia kahteen muuhun verrattuna. Tulkinnallisesti selkeimmin tämä näkyy seuraavassa sarakkeessa (% of Variance), jossa ovat kunkin komponentin selitysosuudet prosenttilukuina. Paras pääkomponentti pystyy selittämään 52 prosenttia havaittujen muuttujien vaihtelusta, toiseksi paras 40 prosenttia. Cumulative % -sarakkeesta nähdään suoraan, että näiden kahden yhteenlaskettu selitysosuus on 92 prosenttia. Selitysosuuden tulkinnasta esimerkiksi Jokivuori & Hietala (2007, 98) mainitsevat, että noin 50 prosentin kokonaisselitysastettakin voidaan pitää tyypillisenä ja varsin hyvänä tasona ihmistieteissä. Muille ei siis jää paljoakaan; kolmas komponentti tosin toisi selitysvoimaa lisää 6 %, mutta siihen liittyvä ominaisarvo on selvästi alle yhden, joten tästä näkökulmasta sitä ei kannattaisi erottaa omaksi pääkomponentikseen. Vaikka pääkomponenttien määräksi ei olisi etukäteen määritelty kahta, tuloksena olisi ollut siitä huolimatta kaksi pääkomponenttia. Ohjelman oletusarvona on, että vain ominaisarvoltaan yli yhden olevat komponentit tulevat omiksi pääkomponenteikseen (ns. ominaisarvokriteeri). Kun suoritetaan näiden kahden pääkomponentin muodostaman akseliston kierto (rotaatio), komponenttien keskinäiset suhteet muuttuvat hieman; kokonaisselitysosuus (92 %) ei tietenkään enää muutu. Rotaatiotulokset näkyvät otsakkeen *Rotation Sums of Squared Loadings* alla olevissa kolmessa sarakkeessa.

Tutustutaan seuraavaksi tulosteisiin, joissa ovat havaittujen muuttujien kummallakin pääkomponentilla saamat lataukset. Katsotaan ensiksi, millaiset ne olivat ennen rotaatiota (Component Matrix -tuloste):

Component Matrix

	Component	
	1	2
Verbaalitestin pistemäärä	,88	-,43
Kieliaineiden keskiarvo	,78	-,59
Matematiikan arvosana	,54	,76
Järkeilytestin pistemäärä	,63	,69

Extraction Method: Principal Component Analysis.

Ensimmäisellä pääkomponentilla suurin lataus on verbaalisuustestin pistemäärällä (0,88); tämä on siis ilmeisesti kielellisyyteen liittyvä komponentti.

Toisella pääkomponentilla suurin lataus (0,76) on matematiikan arvosanalla, eli kyseessä olisi siten matemaattisuus-loogisuus -komponentti. Muutkin lataukset ovat kuitenkin melko suuria, joten faktorirakenne ei tule riittävän selkeästi esiin. Siksi on suoritettu rotaatio. Rotaation jälkeen lataukset muuttuivat tällaisiksi (Rotated Component Matrix -tuloste):

Rotated Component Matrix

	Component	
	1	2
Verbaalitestin pistemäärä	**,97**	,15
Kieliaineiden keskiarvo	**,98**	-,03
Matematiikan arvosana	,00	**,93**
Järkeilytestin pistemäärä	,12	**,93**

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

Rotaation merkitys tulee hyvin esille, kun verrataan latauksia ennen ja jälkeen rotaation. Rotaatio on helpottanut huomattavasti latausten tulkintaa: nyt pääkomponenteille saadaan selkeä rakenne. Verbaalisuustestin pistemäärä (lataus 0,97) ja kieliaineiden keskiarvo (0,98) latautuvat ylivoimaisen vahvasti ensimmäiselle pääkomponentille, joten sen voi selkeästi nimetä kielellisyyskomponentiksi. Matematiikan arvosana (lataus 0,93) ja järkeilytestin pistemäärä (0,93) vuorostaan latautuvat lähestulkoon yhtä vahvasti toiselle komponentille; se on siis matemaattis-loogisuutta mittaava pääkomponentti. Tärkeimmät lataukset ovat yllä olevassa SPSS-tulosteessa korostettu tummemmalla.

Tutkitaan vielä kommunaliteetteja tarkastelemalla, miten hyvin kaksi pääkomponenttia pystyi uuttamaan itseensä analyysiin otetun neljän havaittavan muuttujan vaihtelun:

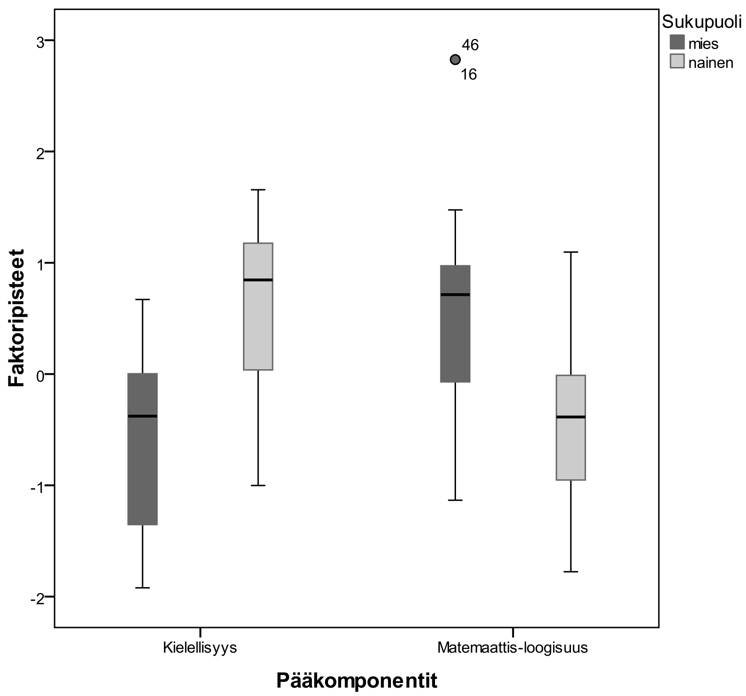
Communalities

	Initial	Extraction
Verbaalitestin pistemäärä	1,000	,962
Kieliaineiden keskiarvo	1,000	,963
Matematiikan arvosana	1,000	,874
Järkeilytestin pistemäärä	1,000	,871

Extraction Method: Principal Component Analysis.

Muodostetut pääkomponentit pystyvät selittämään erityisen hyvin kieliaineiden keskiarvoa (96,3 %) ja verbaalisuustestin pistemäärää (96,2 %); huono tilanne ei ole kahden muunkaan muuttujan kohdalla. Siten informaatiokato siirryttäessä neljästä muuttujasta kahteen ei ole ollut suuri. Oikeastaan voitaisiin puhua informaatiolisästä: alkuperäisiä neljää muuttujaa voidaan luonnollisestikin edelleen käyttää sellaisinaan, eivät ne ole kadonneet minnekään, mutta uutena on kuvaan tullut kaksi uutta ylemmällä abstraktiotasolla olevaa muuttujaa, joiden avulla pystytään viemään tutkimusta askeleen verran eteenpäin.

Pääkomponenttianalyysin avulla ollaan siis kyetty typistämään neljä alkuperäistä muuttujaa kahdeksi pääkomponentiksi, jotka silti selittävät noin 92 prosenttia alkuperäisten muuttujien vaihtelusta. Tässä esimerkkitapauksessa analyysin merkitys ei tunnu luonnollisestikaan kovin merkittävältä, mutta ajatellaanpa vaikka sellaista hyvin tyypillistä tilannetta, että meillä on 40 muuttujan joukko esimerkiksi kyselylomakkeella olevia väittämiä (ks. esim. kuvio 21, s. 219). Jos suuren joukon muuttujia kyetään tiivistämään esimerkiksi neljään ulottuvuuteen eli pääkomponenttiin tai faktoriin, saadaan asetelma, jota on huomattavasti helpompi hallita ja tulkita. Yhtenä esimerkkinä pääkomponenttianalyysin hyödyntämisestä jatkoanalyyseissä voisi olla pääkomponenttipisteiden käyttäminen ryhmävertailuissa. Voidaan tutkia esimerkiksi sitä, miten mies- ja naisvastaajat poikkeavat toisistaan saatujen pääkomponenttien suhteen. Asiaa voitaisiin valottaa esimerkiksi oheisella laatikko-jana -kuviolla (kuvio 22).



KUVIO 22. Pääkomponenttien erot miehillä ja naisilla.

Kuvion perusteella näyttäisi siltä, että kielellisyydessä tutkimuksessa mukana olleet naiset olivat miehiä etevämpiä ja vastaavasti toisen pääkomponentin ominaisuudessa eli matemaattis-loogisuudessa miehet saavat korkeampia arvoja. Erojen merkitsevyyttä voitaisiin tutkia vielä vaikkapa komponenttikohtaisilla t-testeillä. Kuvion yläosassa olevat kaksi havaintopistettä osoittavat tapausten numero 16 ja 46 poikkeavan merkittävästi ryhmänsä muista henkilöistä matemaattis-loogisuuden suhteen. Joissain analyysimenetelmissä tällaiset laatikko-jana -kuviossa erottuvat poikkeavat tapaukset tulee poistaa analyysistä, mikäli tälle on jokin perustelu.

SPSS: Faktori- ja pääkomponenttianalyysi

1. Valitse ylhäältä valikkoriviltä *Analyze → Data Reduction → Factor*.
   - siirrä valitsemasi numeeriset muuttujat muuttujaluettelon oikealla puolella olevalla ylemmällä nuolipainikkeella kohtaan *Variables*
2. *Descriptives*-panike: ruksataan KMO and Bartlett's test….
3. *Rotation*-painike
   - valitse kohtaan *Method* vaihtoehdoksi *Varimax.* Haluamme siis tehdä myös rotaation, jolla saatua ratkaisua on monesti selkeämpi tulkita,
   - hyväksy valinnat *Continue*-painikkeella
4. Painamalla *OK* saat pääkomponenttianalyysin "perustulostuksen" tulosteikkunaan.
5. *Extraction* -painike
   - mikäli haluaisit toteuttaa analyysin varsinaisena faktorianalyysinä, täällä kohdassa *Method* voisit vaihtaa *Principal components* -valinnan tilalle vaihtoehdon *Maximum Likelihood*, joka on yleisimmin käytetty faktorianalyysin estimointimenetelmä.
   - esimerkissämme tiesimme, että haluamme "sovittaa" aineistoomme kahden pääkomponentin ratkaisua. Tämän valinnan tekisit seuraavasti: valitse pääkomponenttianalyysi kohdassa *Method* valinnalla *Principal Components.* Valitse lisäksi vaihtoehto *Number of factors*, jonka vieressä olevaan tyhjään tilaan numero 2. Jos taas et oikein tiedä mikä olisi sopiva faktorimäärä, valitsemalla kohdan *Display* vaihtoehdon *Scree plot* saisit ohjelman piirtämään kuvion, joka auttaisi tässä päätöksenteossa,
   - hyväksy valinnat *Continue*-painikkeella
6. *Scores*-painike
   - ruksaamalla kohdan *Save as variables*, saat tallennettua uudet faktoreita tai pääkomponentteja edustavat muuttujat (ns. faktori- tai pääkomponenttipisteet) saat havaintomatriisiisi,
   - hyväksy valinnat *Continue*-painikkeella
7. *Options*-painike
   - valitse *Sorted by size*, jos haluat, että pääkomponenttitulosteeseen väittämät/muuttujat järjestyvätniiden latausten mukaiseen suuruusjärjestykseen
   - valitse *Suppress small coefficients, absolute value below.* Lukuarvoksi voit asettaa esimerkiksi 0.3, jolloin tätä pienemmät lataukset jäävät taulukossa tulostumatta. Tämä monesti selkeyttää lataustaulukon tulkintaa. Huomaa kuitenkin, että kaikki analyysissa mukana olevat muuttujat latautuvat kaikkiin pääkomponentteihin jossakin määrin, vaikka niitä ei nyt näy tulosteessa edellä tehdyn valinnan vuoksi.
8. Painamalla *OK* saat lisävalinnoilla täydennetyn tulostuksen tulosteikkunaan
   - uudet faktorimuuttujat ilmestyvät datamatriisin oikeaan reunaan
   - kuviota 22 vastaavan laatikko-jana -kuvion piirtämisestä on ohjeistus aineiston kuvailua käsittelevässä luvussa 3 (s. 113).

9. Havaintoyksiköiden ryhmittely klusterianalyysillä

Klusterianalyysi on yksi monimuuttujamenetelmistä. Analyysillä voidaan jakaa aineisto erilaisiin uusiin ryhmiin eli klustereihin. Yleensä klusterianalyysiä käytetään siten, että havaintoyksiköistä (esim. oppilaista) muodostetaan erilaisia ryhmiä muutaman tutkimuksessa olevan muuttujan perusteella. Havaintoyksiköiden ryhmittelyn lisäksi menetelmää on mahdollista käyttää myös aineistossa olevien muuttujien ryhmittelyyn. Tähän tarkoitukseen tosin tavallisimmin käytetään edellisessä luvussa esiteltyä faktori- tai pääkomponenttianalyysiä. Klusterianalyysissa ryhmien eli klusterien lukumäärä ei yleensä ole ennalta tiedossa, kuitenkin tutkijalla olisi hyvä olla joku idea näiden ryhmittymisestä ja ehkä määrästäkin, muuten ryhmittelystä ja tulkinnasta voi tulla hyvinkin sattumanvaraista. Menetelmässä käytetyt muuttujat ovat tyypillisesti määrällisiä, jotka on ennen analyysiä muunnettu yhtenäiselle mitta-asteikolle esimerkiksi standardoimalla. Ryhmät eli klusterit muodostetaan niin, että kuhunkin ryhmään kuuluvat havaintoyksiköt ovat ominaisuuksiltaan mahdollisimman samankaltaisia, mutta keskenään ryhmät eroavat toisistaan. Ryhmittelyyn vaikuttavat kaikki analyysissä mukana olevat muuttujat.

Analyysin yleisiä lähtökohtia:

- monimuuttujamenetelmä
- kyseessä on eksploratiivinen eli aineiston ominaisuuksia kartoittava menetelmä. Menetelmä ei perustu tilastolliseen malliin, kuten esimerkiksi faktori- tai regressioanalyysi, vaan tulos saadaan laskennallisten kriteerien ja jopa erilaisten kokeilujen kautta
- ei erityisiä jakaumaoletuksia, mutta poikkeavat havainnot (outliers) saattavat aiheuttaa tuloksiin epäselvyyttä. Lisäksi voimakkaat korrelaatiot analyysiin valittujen muuttujien välillä voivat aiheuttaa ongelmaa analyysin toteutuksessa (multikollineaarisuus).
- yleisimmin käytetään hierarkkista tai k-means -klusterianalyysiä. Näistä edellinen soveltuu erityisesti pienille aineistoille (max.

300–400, tällaisillakin esimerkiksi tulostukseen tulevan dendogrammi-kuvion tulkinta alkaa olla vaikeata) ja jälkimmäistä kannattaa käyttää isompien aineistojen yhteydessä
- menetelmällä saatua ryhmittelyä on mahdollista hyödyntää jatkoanalyyseissä kuten ristiintaulukoinneissa tai ryhmävertailuissa
- huom! Saadun ratkaisun arvioinnissa ja tulkinnassa on käytettävä harkintaa. Vaihtoehtoisia ryhmittelyjä voi olla useita, ja eri ryhmittelymenetelmät saattavat tuottaa hyvinkin erilaisia lopputuloksia. Ratkaisun hyvyyttä ei voida arvioida tilastollisilla testeillä vaan päätös perustuu tutkijan omaan arviointiin ja tutkimusaiheen sisällölliseen asiantuntemukseen.

9.1 Yleistä klusteri- eli ryhmittelyanalyyseistä

Edellä käsiteltiin faktorianalyysiä, joka periaatteessa mahdollistaisi myös havaintoyksiköiden ryhmittelyn, mutta tavallisempaa on käyttää ryhmittelyyn muita menetelmiä. Silloin kun aineistossa on jo valmiina tiedossa vastaajaryhmät ja halutaan löytää aineiston muuttujista tähän tunnettuun ryhmittelyyn vaikuttavia tekijöitä, voidaan käyttää erottelu- eli diskriminanttianalyysiä. Se perustuu regressio- ja faktorianalyysin tavoin tilastolliseen mallintamiseen, ja tuloksissa hyödynnetään todennäköisyysjakaumia, jolloin käytön edellytyksenä on moniulotteinen normaalijakauma. Yleisempi tilanne kuitenkin on sellainen, jossa ei tiedetä etukäteen, sisältääkö tutkimusaineisto samankaltaisista havainnoista koostuvia ryhmiä tai miten monta tällaisia ryhmiä mahdollisesti on. Ryhmien eli klusterien etsimiseen ja muodostamiseen käytetään ryhmittely- eli klusterianalyysiä.

Ryhmittelyanalyysin menetelmät ovat luonteeltaan kuvailevia ja aineistolähtöisiä, ja niissä pyritään aineistosta käsin ryhmittelemään aineiston havaintoja matemaattisin algoritmein. Tavoitteena on muodostaa sellaisia ryhmiä, jotka ovat keskenään mahdollisimman poikkeavia, mutta ryhmien sisällä havainnot ovat keskenään mahdollisimman samanlaisia. Ryhmittelyanalyysissä ei saada aikaan yhtä ”oikeaa” ratkaisua vaan tutkija joutuu itse päättämään vaihtoehtoisista ratkaisuista tulkinnan kannalta parhaan. (Ks. esim. Everitt ym. 2011.)

Yleisimmin käytettyjä ryhmittelyanalyysin menetelmiä ovat hierarkkinen klusterianalyysi tai K-means -klusterianalyysi. Lisäksi viime vuosina on yleistynyt tilastolliseen mallintamiseen perustuva latentti profiilianalyysi LPA, jollaista voidaan soveltaa rakenneyhtälömallituksen avulla niin sanottujen sekoittuneiden jakaumien mixture-mallina. Aivan samoin kuin klusterianalyyseissä, LPA:ssakin tavoitteena on löytää ennalta tuntemattomia aineistosta ominaisuuksiltaan erilaisia klustereita (profiiliryhmiä), joihin kuuluvat jäsenet ovat keskenään samankaltaisia. Latenttia profiilianalyysiä ei käsitellä tarkemmin tässä kirjassa, mutta aiheeseen voi tutustua esimerkiksi Mplus-ohjelmiston verkkomateriaaleista osoitteessa statmodel.com.

Hierarkkinen klusterianalyysi soveltuu parhaiten pienemmille aineistoille (väljästi määritellen monesti N on enintään ”kymmeniä”). Laajoillakin aineistoilla ryhmittely voidaan kyllä toteuttaa, mutta analyysin tulosta voi olla hankala tulkita, muun muassa puu-muotoiset graafiset esitykset alkavat isoissa otoksissa muodostua varsin epäselviksi. Toki päätulos, klusteriryhmiin kuuluminen, on selkeästi nähtävissä ja tallennettavissa esimerkiksi kuvailua ja jatkoanalyysejä varten. Menetelmässä luokitteluun vaikuttavina muuttujina – eli kriteerimuuttujina – voidaan käyttää sekä määrällisiä että kategorisia muuttujia. Ennen analyysin toteuttamista määrälliset muuttujat on hyvä standardoida[6] yhteiselle mitta-asteikolle, jolloin kaikki valitut muuttujat vaikuttavat ryhmittelyyn yhtä voimakkaasti. Ilman standardointia suuruusluokaltaan isot muuttujat painottuisivat ryhmittelyssä enemmän kuin pieniä arvoja sisältävät muuttujat.

Havaintojen ryhmittely etenee vaiheittain, alussa klustereita on yhtä monta kuin havaintoyksiköitä. Seuraavaksi keskenään lähimpänä toisiaan olevat kaksi havaintoyksikköä yhdistetään samaan klusteriin. Havaintojen läheisyyttä/etäisyyttä mitataan jollakin sopivalla etäisyysluvulla, joka esimerkiksi SPSS:ssä on nimeltään euklidisen etäisyyden neliö (oletusmenetelmä, joka tarvittaessa voidaan vaihtaa). Seuraavassa vaiheessa kolmas havaintoyksikkö joko liitetään edellisten muodostamaan ryhmään tai yhdistetään jonkin muun havaintoyksikön kanssa omaksi ryhmäkseen. Näin edetään askel kerrallaan, ja jokaisessa vaiheessa ryhmittelemätön yksikkö joko liitetään aiempiin klustereihin tai johonkin ryhmittelemättömään havaintoyksikköön, tai sitten kaksi jo muodostettua ryhmää yhdistetään suuremmaksi klusteriksi. Menetelmän hierarkkisuus tarkoittaa sitä, että kerran ryhmitelty havaintoyksikkö pysyy loppuun asti samassa ryhmässä. Viimeinen vaihe on se, että kaikki havain-

[6] Standardoinnissa muuttujalle tehdään muunnos, jonka jälkeen muuttuja-arvojen keskiarvo on 0 ja keskihajonta 1.

toyksiköt ovat yhdessä ja samassa klusterissa (tämä ei tietenkään ole kovin kiinnostava ratkaisu). Tutkijan on itse päätettävä, missä vaiheessa ryhmien lukumäärä olisi tulkinnan kannalta järkevä. Monesti ongelmana onkin löytää optimaalinen kohta lopettaa ryhmittely. Ratkaisua voi etsiä tarkastelemalla ryhmittymisen etenemistä kuvaavaa jääpuikkokuviota (Icicle plot) tai etäisyysarvoja ja ryhmittymistä kuvaavaa dendogrammia (Dendogram plot). Kun ryhmien lukumäärä on päätetty, niin lopuksi havaintomatriisiin kannattaa tallentaa uudeksi muuttujaksi tieto siitä, mihin klusteriin kukin havainto kuuluu. Näin saatua ryhmittelyä voitaisiin hyödyntää jatkoanalyyseissä. Esimerkkejä hierarkkisen klusterianalyysin toteutuksesta, SPSS-tulosteista ja niiden tulkinnasta löytyy lähteissä mainituista kirjoista (ks. mm. Metsämuuronen 2009, 877–884).

K-means -klusterianalyysi soveltuu erityisesti isommille aineistoille (N ”satoja”), joissa hierarkkisella analyysillä olisi yleensä vaikeaa saada tulkittavaa ja käyttökelpoista ratkaisua. Menetelmä on käyttökelpoinen myös silloin, kun muodostettavien ryhmien lukumäärä on etukäteen tiedossa. K-means -ryhmittelyssä kriteerimuuttujina käytettävien muuttujien pitää kaikkien olla määrällisiä ja ne on hyvä standardoida yhtenäiselle mitta-asteikolle. SPSS:ssä standardointi on tehtävä muuttujille jo ennen klusterianalyysin aloittamista, sillä K-means -proseduurissa ei ole valittavissa valmista optiota standardointiin. Menetelmässä tutkijan on aluksi itse päätettävä muodostettavien klusterien lukumäärä (k kappaletta), minkä jälkeen klustereille lasketaan alustavan ryhmityksen mukaiset klusterikeskukset (cluster centers). Ryhmittelyä pyritään parantamaan etäisyysmittojen pohjalta, minkä jälkeen klusterikeskusten arvot lasketaan uudestaan. Havaintojen ryhmittely etenee iteratiivisesti eli askeltaen, ja iterointia jatketaan, kunnes klusterikeskusten arvot eivät merkittävästi muutu. Tällöin tulkitaan, että ratkaisu on löytynyt, ja havaintoyksiköt kuuluvat kukin yhteen klusteriin. Käydään seuraavassa esimerkissä läpi K-means -ryhmittelyyn liittyvät vaiheet sekä keskeiset SPSS-tulosteet ja niiden tulkinta.

9.2 K-means -ryhmittelyanalyysiesimerkki

Toteutetaan klusterianalyysi (K-means Cluster), jossa halutaan ryhmitellä esimerkkiaineiston opiskelijat kolmeen ryhmään eli klusteriin heidän opintomenestyksensä mukaan. Klusterien lukumäärä voisi olla jokin muukin ja käytännössä analyysi toistettaisiin ryhmien eri lukumäärillä ja tulosten perusteella päätettäisiin, monenko ryhmän ratkaisu olisi järkevin.

Ryhmittelyssä käytetään kuutta aineistossa olevaa numeerista muuttujaa: verbaalitestin pisteet, järkeilytestin pisteet, kieliaineiden keskiarvo, matematiikan arvosana, opintomenestys lukiossa ja opintomenestys yliopistossa. Ennen analyysiä kaikki muuttujat on standardoitu, ja ryhmittelyssä käytetään näitä standardoituja muuttujia (alla olevissa SPSS-taulukoissa esimerkiksi merkintä Zscore: kieliaineiden keskiarvo tarkoittaa alkuperäistä kieliaineiden keskiarvomuuttujaa standardoidussa muodossa).

Aluksi SPSS-tulosteisiin tulee tiedot alustavista ryhmäkeskuksista ja iterointivaiheista. Nämä eivät ole tulkinnan kannalta erityisen kiinnostavia, joten ne on jätetty tästä pois. Alla olevassa taulukossa *Number of Cases in each Cluster* nähdään muodostuneiden klustereiden koko, tosin sanoen montako opiskelijaa kuhunkin klusteriin kuuluu.

Number of Cases in each Cluster

Cluster	1	14,00
	2	24,00
	3	22,00
Valid		60,00
Missing		,00

Tässä ryhmittely on tuottanut melko tasapainoisen tuloksen, pienimmässäkin ryhmässä (Cluster 1) on lähes neljännes havainnoista.

Seuraavassa taulussa *Final Cluster Centers* esitetään lopullisen klusteriratkaisun ryhmäkeskukset. Näiden perusteella voidaan tarkastella eri klusterien sisällöllistä tulkintaa. Esimerkiksi klusterin 2 sarakkeella olevat negatiiviset arvot voidaan tulkita niin, että klusterin 2 opiskelijat saavat muita ryhmiä matalampia arvoja lähes kaikissa kriteerimuuttujissa. Klusterien tulkintaan liittyviä seikkoja tarkastellaan enemmän hieman tuonnempana.

Final Cluster Centers

	Cluster		
	1	2	3
Zscore: verbaalitestin pisteet	,32	-,90	,78
Zscore: Järkeilytestin pisteet	1,12	-,40	-,28
Zscore: Kieliaineiden keskiarvo	,18	-,94	,91
Zscore: Matematiikan numero	1,22	-,29	-,46
Zscore: Opintomenestys lukiossa	,33	-,96	,84
Zscore: Opintomenestys yo:ssa	1,01	-,82	,25

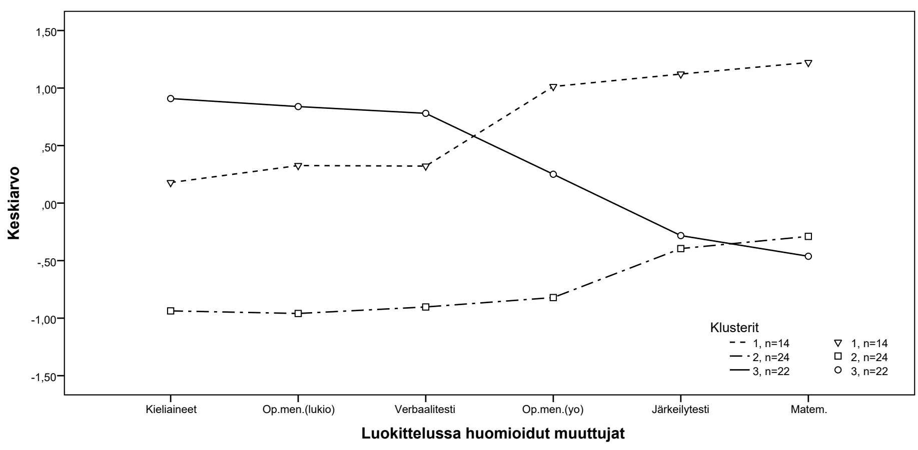
Taulukosta ANOVA nähdään, miten klusterien väliset erot ilmenevät eri muuttujilla. Toiset muuttujista vaikuttavat enemmän ryhmittelyyn, silloin F-suureen arvot ovat muita suurempia; samoin toisilla muuttujilla ei ole aivan yhtä suurta vaikutusta, jolloin F-arvot ovat matalampia. Varsinaiseen ryhmien väliseen erojen testaamiseen ei sig.-arvoja ole tarkoitettu käytettäväksi (huomaa SPSS:n tulostama huomautus taulukon alla), mutta F-suureen arvoista voidaan nähdä, missä muuttujissa klusterit selkeimmin eroavat toisistaan.

ANOVA

	Cluster		Error			
	Mean Square	df	Mean Square	df	F	Sig.
Zscore: Verbaalitestin pisteet	17,201	2	,432	57	39,862	,000
Zscore: Järkeilytestin pisteet	11,557	2	,630	57	18,358	,000
Zscore: Kieliaineiden keskiarvo	19,834	2	,339	57	58,480	,000
Zscore: Matematiikan numero	13,808	2	,551	57	25,078	,000
Zscore: Opintomenestys lukiossa	19,530	2	,350	57	55,830	,000
Zscore: Opintomenestys yo:ssa	15,965	2	,475	57	33,615	,000

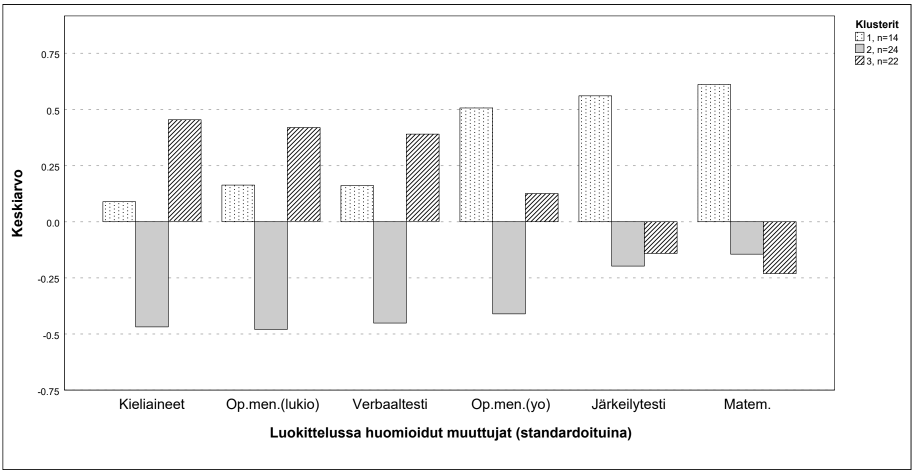
The F tests should be used only for descriptive purposes because the clusters have been chosen to maximize the differences among cases in different clusters. The observed significance levels are not corrected for this and thus cannot be interpreted as tests of the hypothesis that the cluster means are equal.

Tämän mukaan suurimpia eroja klusterien välillä on kieliaineiden keskiarvossa ja lukion opintomenestyksessä (suurimmat F-arvot). Vastaavasti järkeilytestin pistemäärässä klusterien väliset erot eivät ole yhtä selviä (pienin F-arvo). Muuttujien F-arvoja tutkimalla saadaankin alustavaa tuntumaa siitä, mitä ominaisuuksien klusterien identifioinnissa kannattaisi painottaa. Yksityiskohtaisempi käsitys klusterien ominaisuuksista saadaan tarkastelemalla niiden keskiarvoprofiileja. Profiililla tarkoitetaan tässä yhteydessä klusteroinnissa mukana olleiden muuttujien keskiarvojen havainnollistamista joko keskiarvotaulukon tai keskiarvoja kuvaavan viiva- tai pylväskuvion avulla. Kuviossa 23a nähdään esimerkistä muodostettujen kolmen klusterin profiilit viivakuviona.



KUVIO 23a. Keskiarvoprofiilit kolmen klusterin ratkaisulla viivadiagrammilla kuvattuna.

Yllä oleva viivadiagrammi (kuvio 23a) kuvaa havainnollisesti eri klusterien profiilimuotoja; toisaalta kuviosta voi saada erheellisen käsityksen, että vaaka-akselilla olevat eri tekijät olisivat jollain lailla jatkuva-arvoisesti yhteydessä toisiinsa. Toinen varsin yleisesti käytetty tapa onkin esittää klusterien keskiarvoprofiilit ryhmiteltyä pylväikköä käyttäen esimerkiksi alla olevan graafin tapaan (kuvio 23b):



KUVIO 23b. Keskiarvoprofiilit kolmen klusterin ratkaisulla pylväsdiagrammia käyttäen.

Kuvioissa pystyakseli kuvaa muuttujien standardoituja arvoja, jolloin arvo nolla tarkoittaa kunkin muuttujan keskimääräistä tasoa. Positiivinen lukuarvo tarkoittaa, että ominaisuutta on keskimääräistä enemmän, ja negatiivinen, että sitä on keskimääräistä vähemmän. Klusterin 2 (n = 24) profiili näyttäisi kulkevan pääosin alimpana ja kaikkien kuuden muuttujan suhteen negatiivisella alueella. Tätä vastaajaryhmää voitaisiin luonnehtia termillä ”taidoiltaan heikot”. Klusteri 1 oli pienin ryhmä (n = 14), ja heillä näyttäisi olevan kielelliset taidot ja lukiomenestys keskimääräistä korkeammalla tasolla ja parempaa kuin muilla ryhmillä. Ryhmä voitaisiin nimetä esimerkiksi verbaalisesti lahjakkaiden ryhmäksi ”verbaaliset opintomenestyjät”. Klusterin 3 (n = 22) opiskelijat taas näyttäisivät erottuvan erityisesti matemaattisuutta mittaavissa ominaisuuksissa, nimitettäköön heitä tässä ”matemaattis-loogiset opintomenestyjät”. Kuvion 23b pylväikkö voitaisiin muodostaa myös niin, että kuvio sisältäisi klustereittain kolme pylväsryhmää, jossa kussakin olisi kuusi pylvästä. On mahdollista, että tällaisesta asettelusta klusterien väliset erot havainnollistuisivat vielä selvemmin.

On huomattava, että ryhmien profiilien tulkinta saattaa joskus olla melko epäselvää ja ryhmiä erottelevia piirteitä voi olla vaikeata nimetä, jolloin tulkinta voi jäädä kovin subjektiiviselle tasolle. Kuten Kimmo Vehkalahti (2014, 151) toteaa, joskus ryhmittely voi olla lähinnä ”sokeata hapuilua aineiston seassa, jossa huonoimmillaan tuloksena on pelkkää sekasotkua”. Tällaisissa tapauksissa klusterianalyysin tulosta ei luonnollisestikaan kannatta käyttää.

Mihin klusterianalyysin tulosta sitten voitaisiin käyttää? Joskus riittää pelkästään deskriptiivinen eli kuvaileva tulos, jossa on pyritty saamaan käsitystä aineiston rakenteesta: millaisia osaryhmiä havaintoaineistossa on löydettävissä ja miten moneen klusteriin aineisto näyttäisi jakautuvan. Tai onko erillisiä klustereita ylipäänsä löydettävissä vai vaikuttaako siltä, että koko aineisto on ominaisuuksiltaan melko homogeenista? Lisäksi klusterianalyysillä saatua tulosta voidaan käyttää hyödyksi jatkotarkasteluissa.

Jos analyysillä löytyi selkeästi erottuvat, tulkinnaltaan järkevät osaryhmät, havaintomatriisiin kannattaa tallettaa uudeksi muuttujaksi klusteriin kuuluminen (SPSS: *Cluster membership*, joka saadaan tallennettua analyysin yhteydessä havaintomatriisin viimeiseksi sarakkeeksi). Tätä kategorista muuttujaa voidaan sen jälkeen käyttää kuten mitä tahansa aineiston ryhmittelevää muuttujaa. Esimerkissä voidaan jatkaa analysointia vaikkapa selvittämällä ryhmittelyn ja sukupuolen välistä yhteyttä ristiintaulukoimalla. Tai voidaan tutkia varianssianalyysillä klusterien välisiä keskiarvoeroja jossakin tutkimusongelman kannalta kiinnostavassa ominaisuudessa. Kannattaa kuitenkin muistaa, että jatkoanalyyseissä ei pidä käyttää klusterianalyysin pohjana jo käytettyjä muuttujia, sillä onhan selvää tai ainakin melko todennäköistä, että näissä muuttujissa klusterit eroavat toisistaan. Seuraava SPSS-taulukko kuvaa klusterianalyysin käyttöä jatkoanalyysissä, jossa ristiintaulukoinnilla on haluttu selvittää sukupuolen ja klusteriin kuulumisen yhteyttä.

TAULUKKO 13. Miesten ja naisten jakautuminen kolmelle klusterille.

			Klusteri			
			klusteri1: matem. taidot	klusteri2: heikot taidot	klusteri3: kielell. taidot	Yhteensä
Sukupuoli	miehet	lkm	12	16	2	30
		%	40,0%	53,3%	6,7%	100,0%
	naiset	lkm	2	8	20	30
		%	6,7%	26,7%	66,7%	100,0%
Yhteensä		lkm	14	24	22	60
		%	23,3%	40,0%	36,7%	100,0%

Miehistä valtaosa kuuluu klustereihin 1 ja 2 ja vain vajaa 7 % kuuluu klusteriin 3 (vahvat kielelliset taidot). Tutkimukseen osallistuneilla naisilla prosentuaalinen jakauma on hyvin erilainen: suurin osa naisista kuuluu klusteriin 3 ja klusterissa 1 on vain 7 prosenttia naisista. Näin siis sukupuolella ja klusteriin kuulumisella näyttäisi olevan yhteyttä. Yhteyden tilastollinen merkitsevyys nähdään khiin neliö -testillä taulukossa *Chi-Square Tests*. Sen mukaan yhteys on tilastollisesti erittäin merkitsevä ja efektikooltaan voimakas ($\chi^2(2) = 24{,}54$; $p < 0{,}001$; $V = 0{,}64$).

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	24,537[a]	2	,000
Likelihood Ratio	27,738	2	,000
Linear-by-Linear Association	22,069	1	,000
N of Valid Cases	60		

a. 0 cells (,0%) have expected count less than 5. The minimum expected count is 7,00.

Jatkoanalyyseissä voitaisiin edetä myös laadullisen analysoinnin suuntaan. Klusterianalyysin yhteydessä aineistoon saadaan tallennettua klusteriin kuulumisen lisäksi myös klusteriin kuulumisen tyypillisyyttä kuvaava mittaluku *distance from cluster center*. Tämä lukuarvo tallentuu havaintomatriisiin jokaiselle havaintoyksikölle, ja se tarkoittaa yksikön (henkilö, vastaaja, oppilas tms.) etäisyyttä oman klusterinsa keskipisteestä. Silloin kun arvo on hyvin pieni, lähellä nollaa, havaintoyksikkö on hyvin tyypillinen klusterinsa edustaja. Vastaavasti suuri etäisyysarvo tarkoittaa epätyypillistä klusterin jäsentä. Seuraavassa SPSS-taulukossa on listattu joka klusterista kaksi tyypillisintä henkilöä ja heidän profiilinsa.

Case Summaries[a]

id	etäisyys keskuksesta	klusteri	sp	kieli_ka	verb_t	omene1	matik	järk_t	omene2
9	,596	1	1	79	26	17	9	37	19
39	,596	1	1	79	26	17	9	37	19
7	,259	2	2	72	23	14	7	34	14
37	,259	2	2	72	23	14	7	34	14
24	1,105	3	2	86	27	16	7	34	15
54	1,105	3	2	86	27	16	7	34	15

a. Limited to first 6 cases.

Ensimmäistä klusteria edustavat parhaiten henkilöt id = 9 ja 39, klusterin 2 edustajina ovat henkilöt id = 7 ja 37, ja klusterin 3 tyypillisiä edustajia ovat id = 24 ja 54. Nyt tutkimukseen voisi yhdistää kvalitatiivista analyysiä, niin että keskityttäisiin tarkemmin näihin kuuteen tapaukseen, heidän taustoihinsa ja muihin ominaisuuksiin.

SPSS: Klusterianalyysi (K-means Cluster)

1. Standardoi ensin valitsemasi määrälliset muuttujat
   - valitse valikkoriviltä *Analyze → Descriptive Statistics → Descriptives*
   - siirrä valitsemasi muuttujat nuolipainikkeella kohtaan *Variables*, ja napsauta rasti ruutuun valintaikkunan alareunassa olevaan kohtaan *Save standardized values as variables*
   - paina *OK*, saat havaintomatriisin oikeaan reunaan uudet standardoidut muuttujat, joita voit käyttää klusterianalyysissä.
2. Aloita K-means-klusterianalyysi
   - valitse ylhäältä valikkoriviltä *Analyze → Classify → K-means Cluster*.
   - siirrä valitsemasi standardoidut numeeriset muuttujat muuttujaluettelon oikealla puolella olevalla ylemmällä nuolipainikkeella kohtaan *Variables*
   - anna muodostettavien klusterien lukumäärä numeroarvolla kenttään *Number of Clusters.*
3. *Options*-painike
   - kannattaa valita kohta *ANOVA table*, jolla saadaan tulostukseen varianssianalyysin taulukko, josta nähdään eri muuttujien vaikutus havaintojen ryhmittelyyn,
   - hyväksy valinnat *Continue*-painikkeella
4. *Save*-painike
   - tässä saat tallennettua aineistoosi kullekin vastaajalle tiedon siitä, mihin klusteriin hän kuuluu valitsemalla kohdan *Cluster membership* sekä miten kaukana hän on oman klusterinsa keskikohdasta (valinta *Distance from cluster center*), hyväksy valinnat *Continue*-painikkeella
5. Painamalla *OK* saat klusterianalyysin tulokset tulosteikkunaan.

HUOM. mikäli aineisto on kovin suuri, esim. tuhansia havaintoja, ratkaisun löytymistä voidaan helpottaa siten, että aluksi analyysi toteutetaan vain osalle aineistoa ja näin saadut ryhmäkeskukset tallennetaan omaksi erilliseksi SPSS-aineistoksi. Seuraavaksi analyysi tehdään koko aineistolle tallennettujen ryhmäkeskuksien avulla. Tällainen kaksivaiheinen toteutus saadaan tehtyä valintaikkunan kohdassa *Cluster Centers*.

HUOM: kuviot saat muodostettua valikosta *Graphs → Legacy Dialogs*, josta voit valita kuviotyypiksi joko viivakuvion *Line* tai pylväikön *Bar*.

10. Lopputulemaa

Kasvatus- ja yhteiskuntatieteellisessä määrällisessä, kvantitatiivisessa, tutkimuksessa ollaan kiinnostuneita ennen kaikkea eri tekijöiden (muuttujien) yhteyksistä, riippuvuuksista ja mallintamisesta. Ihmiset ja laajemmin eri instituutiot ja yhteiskuntapolitiikka tarvitsevat tällaista tietoa. Ilman tätä päätöksenteko olisi paljon nykyistä haasteellisempaa. Oikeastaan ilman survey-tutkimuksen ja tilastoaineiston tuottamaa tietoa emme voisi tarkastella laajempia yhteiskunnallisia kysymyksiä yleisellä tasolla lainkaan. Tämä koskee myös koululaitosta, esimerkiksi ilman määrällisen ja systemaattisen aineiston keräämistä ja analysointia sen arviointi, miten koulut eri puolella Suomea toimivat ja palveleeko järjestelmä koko maassa niin kuin sen pitäisi, olisi lähes mahdotonta selvittää. Tai ajatellaanpa vaikkapa sen selvittämistä, eroavatko eri alueiden nuorten koulutuspolut toisistaan, ja jos eroavat, mistä tämä mahdollisesti johtuu. Vastaavasti määrällinen ja tilastollinen tutkimus mahdollistaa eri maiden kouluissa tapahtuvan oppimisen vertaamisen. Tällaista laajaa tutkimusta edustavat muun muassa eri vuosina tehtävät PISA-tutkimukset. Määrällinen tutkimus palvelee tavoitteita erityisesti laajoissa makrotason tutkimuksissa, mutta kvantitatiivisella tutkimuksella voidaan tarkastella myös mikrotason ilmiöitä, esimerkiksi luokkahuoneiden tai koulujen ryhmädynamiikkaa tai opetukseen ja oppimiseen liittyviä eri ilmiöitä, prosesseja ja mekanismeja. Mikään ei sulje pois sitäkään mahdollisuutta, että määrällisillä aineistoilla pyritään kartoittamaan laadullisen tutkimuksen lailla tutkittaviin ilmiöihin liittyviä uusia piirteitä tai niihin liittyvien eri tekijöiden uusia yhteyksiä, vaikka tämä onkin vähän vierasta perinteiselle hypoteesi-tilastolliselle -tutkimusasetelmalle. Määrällisen aineiston avulla voidaan tarkastella myös tulevaa: esimerkiksi kirjassa esitetyssä regressioasetelmassa analysoitiin sitä, kuinka lukion menestys matemaattisissa tai äidinkielen aineissa ennakoi pärjäämistä mahdollisissa tulevissa yliopisto-opinnoissa.

Määrällisen tutkimuksen merkitystä käyttäytymis- ja yhteiskuntatutkimuksen kentässä ei voi kiistää. Tällä toteamuksella ei ole tarkoitus missään nimessä vähätellä laadullista tutkimusta, päinvastoin, molempia tarvitaan. Ja vielä jopa näitä lähestymistapoja yhdistävää tutkimusta. Mitä kattavammin voimme tutkia tutkittavaa ilmiötä tai kysymystä, sitä kokonaisvaltaisempia tulkintoja saamme tehtyä näistä ja näihin liittyvistä tekijöistä ja näiden välisistä yhteyksistä. Meille on tyypillistä yksinkertaistaa tai pelkistää tulkinnoissamme eri ilmiöitä ja asioita. Tämä koskee myös käyttäytymis- ja yhteiskuntatieteellistä tutkimusta. Mikko Ketokiven mukaan tämä johtuu ihmisten kognitiivisesta rajallisuudesta. Tieteenteon tehtävänä ei kuitenkaan ole hänen mukaansa pyrkiä selittämään ”monimutkaisia ilmiöitä yksinkertaisilla teorioilla” tai lievemmin ilmaistuna yksinkertaistavilla tulkinnoilla. Sen sijaan Ketokiven mukaan tieteen ja tutkimuksen tehtävänä on tehdä monimutkaisisista ilmiöistä ymmärrettäviä. (Ketokivi 2015, 55–58.) Tämä lähtökohta korostaa sitä, että tutkijan tulisi olla varovainen liian suoraviivaisten yhteyksien, ehdottomien ja kapea-alaisten selitysten suhteen. Toki tämän vaateen suhteen ei kannata olla liian kunnianhimoinen, ei varsinkaan ensimmäisiä opinnäytetöitään määrällistä aineistoa hyödyntäen tekevien. Monesti omaan aineistoon nojaava kuvailevakin tulkinta voi olla hyvin kiintoisa, varsinkin kun sen laittaa keskustelemaan aiemman asiaa tai kysymystä koskevan tutkimustiedon kanssa.

Määrällinen survey-tutkimus perustuu otantatutkimusasetelmaan, jossa tutkimusasetelman määrittämästä populaatiosta kerätyssä otoksessa havaittujen jakaumien erojen ja samanlaisuuksien oletetaan esiintyvän tietyllä riskitasolla kyseisessä populaatiossa. Periaatteessa tilastollinen päättely on perustaltaan siis aineistolähtöistä, induktiivista tulkintaa. Tässä teoksessa on pyritty avaamaan tilastollisten menetelmien perusmenetelmiä ja sitä, kuinka näiden analyysitulosteita tulkitaan. Lisäksi teoksessa on käyty yleispiirteittäin läpi tutkimusprosessin eri vaiheet, tutkimuksen suunnittelusta ja asemoimisesta aineiston keräämiseen ja tulkitsemiseen. Nämä kaikki vaiheet ovat keskeisiä onnistuneelle tutkimuksenteolle. Pääpaino teoksessa on kuitenkin siis tilastollisten menetelmien käyttöön liittyvissä kysymyksissä. Tämän vuoksi kirjassa vilisee erilaisia käsitteitä, menetelmien ja tulkinnan kriteereitä ja varoituksen sanoja, näitä kaikkia ei ole helppo ottaa kerralla haltuunsa. Ajan kanssa, lisäopiskelun ja ennen kaikkea eri menetelmiä käyttäen, ne saa kyllä haltuun. Yksi kirjan tekemisen päätavoite onkin ollut madaltaa eri menetelmien käyttöön, ja laajemmin tutkimuksentekoon, liittyviä pelkoja ja paineita. Jotta lukija saisi otteen paremmin eri menetelmien ja niiden tulosten analyysien perusteista, olemme esimerkiksi rajanneet eri menetelmien tulosteiden tarkastelun lähinnä näiden tilastolliseen tulkintaan. Rajauksen vuoksi näiden tulkinnassa on kyse ollut siis lähinnä tilastollisesta päättelystä, vaikka lähtökohtamme onkin se, että tulkinta ei saisi jäädä konkreettisissa tutkimuksissa tälle tasolle, vaan tilastollisten menetelmien saatuja suureita pitäisi aina tarkastella laajempaa tutkimuskysymystä koskevaa teoreettisia ja aiempien tutkimusten tulkintoja vastaan (ks. luku 1, erityisesti kuvio 2). Rajautumalla menetelmien tulosteiden tilastollisten suureiden tulkintaan ajatuksena on ollut se, että näin varsinkin niille, jotka ottavat vasta ensiaskeleitaan määrällisen aineiston maailmaan, on helpompi hahmottaa ja saada kiinni siitä, mistä eri menetelmien käytössä on kyse, miten niitä käytetään ja miten niiden tulosteita tulkitaan. Kuitenkin haluamme vielä tässäkin yhteydessä muistuttaa lukijaa siitä, kuinka pelkkä aineistopohjainen tilastollinen päättely jää helposti ”lattean empirismin” ja kuvailemisen tasolle. Kuten esimerkiksi Pertti Töttö (2012, 10) on osuvasti todennut, eivät analyysitulosten sisältämät luvut sinänsä ole vielä tutkimustuloksia, vaan tuloksiksi ne muuttuvat vasta siinä vaiheessa, kun nämä kertovat jotain tutkittavista asioista ja ilmiöistä. Pidetään kaikki tämä mielessämme!

Jos halutaan, meitä voi nimetä menetelmien käytön suhteen varsin liberaaleiksi, koska edustamme sellaista näkemystä, että tutkijan on aina viime kädessä päätettävä itse tapauskohtaisesti, miten hän tulkitsee eri menetelmille asetettuja joskus hyvinkin tiukkoja kriteereitä ja niissä havaittuja poikkeamia omassa aineistossa. Poikkeamat näiden suhteen todellisissa aineistoissa ovat enemmänkin sääntö kuin poikkeus. Yleisesti esitetyt menetelmien ja tulkintojen kriteerit ovat hyvä peruskartasto, mutta jos niitä tulkitsee liian tiukasti, saa takuuvarmasti päänsärkyä. Näidenkin suhteen, kuten analyysien tulosten tulkinnan kohdalla, hyvä nyrkkisääntö on se, että tulkitsee näitä oman tutkimusasetelmansa kokonaisuuden suhteen. On tärkeämpää saada aikaiseksi ymmärrettävä analyysi ja mielenkiintoinen tulos kuin se, että kaikki menetelmälliset kriteerit on täytetty pilkulleen (ks. esim. Jokivuori & Hietala 2007). Toki näihin poikkeamiin pitää suhtautua vakavasti, eikä kaikkia menetelmien käyttöön liittyviä ehtoja missään nimessä saa ohittaa turhina. Nämä ainakin tulee ottaa huomioon tulosten tulkinnassa ja ne tulee myös aina raportoida tutkimusraporteissa. Onneksi monet menetelmät ovat varsin ”vakaita” pienten (ja vähän suurempienkin) poikkeamien suhteen. Vastaavasti suurien aineistojen kohdalla esimerkiksi normaalijakaumavaatimuksen katsotaan täyttyvän yleisesti, vaikkei se analyysien mukaan tätä ihan täyttäisikään. Tämä sama visainen kysymys koskee siis myös analyysien tulosten tulkintakriteereitä, esimerkiksi sitä, miten tulkita saamamme tilastollisia merkitsevyys tai efektitasoon liittyviä arvoja. Turvaudummeko mekaanisesti annettuihin kriteeriarvoihin, vai onko meillä jokin syy tulkita näitä jossakin suhteessa toisin, on monesti kysymys, johon tutkija käytännön tutkimusaineistoa tulkitessaan törmää. Tästäkin selviää yleensä kunnialla, kun tuntee oman tutkimuskysymyksen alaa ja käyttämiensä menetelmien perustoja riittävästi, pohtii rauhassa ja perustelee tulkintaansa. Kaikki tämä vaatii tutkijalta luonnollisesti omaa panostusta, avointa, tietoista ja aktiivista suhdetta tekemiseensä. Muun muassa Wasserstein, Schrim & Lazar (2019) korostavat tähän liittyen osuvasti ”ajattelevan tutkijan” (thoughtful researcher) ja spesifimmin ”tilastollisesti ajattelevan tutkijan” keskeisyyttä, siis tutkijaa, jolla on tietoinen, avoin, kriittinen ja tietyssä mielessä nöyräkin suhtautuminen tutkimuksen tekoon.

Kvantitatiivisen aineiston tilastomenetelmien yksi vahvuus on se, että niillä voidaan analysoida samanaikaisesti monien tekijöiden yhteyttä tutkittavaan ilmiöön, myös niin, että näiden eri tekijöiden välisiä yhteyksien merkityksiä voidaan tulkita toistensa suhteen samalla kun niiden yhteyttä tarkastellaan selitettävien muuttujien/tekijöiden suhteen. Tällä päästään lähemmäksi todellisia ”vaikutusyhteyskimppuja” kuin esimerkiksi kahden muuttujan välisissä tarkasteluissa. Kyseeseen tulevat tällöin niin sanotut monimuuttajamenetelmät. Tässä teoksessa näidenkin suhteen on rajauduttu perusanalyysimenetelmiin (varianssi-, regressio- ja pääkomponenttianalyyseihin, lisäksi on avattu tarkemmin klusterianalyysia). Hienosyisempien tilastollisten mallinnusmenetelmien (esim. monitasomallien, polkuanalyysien tai yleisemmin sanottuna rakenneyhtälömallien, Structural Equation Models, SEM) esittely on jätetty muiden teosten tehtäväksi (ks. esim. Metsämuuronen 2009; Nummenmaa 2011; Kline 2016; Byrne 2011). Näitä ei ole otettu tähän perusteokseen, koska näiden käyttäminen ja tulkinta ovat jo paljon vaativampaa kuin edellä mainittujen perusmenetelmien käyttö. Kuitenkaan näiden tuomia lisäanalyysi- ja tulkintamahdollisuuksia ei kannata unohtaa, koska näiden avulla voidaan tulkita esimerkiksi kasvatukseen ja koulutukseen liittyviä ilmiöitä ja prosesseja paljon monipuolisemmin kuin muilla menetelmillä on mahdollista. Voidaan esimerkiksi tarkastella sitä, mitkä tekijät vaikuttavat oppimisen taustalla, testata tähän liittyviä teoreettisia konstruktioita ja niin edelleen. Tarpeen tullen näidenkin menetelmien tarjonta on siis käytössäsi, vaativuudestaan huolimatta monen tutkimusasetelman kannalta nämä antaisivat tärkeitä lisätyökaluja tutkijan käsiin. Esimerkiksi edellä mainittujen kirjojen avulla näidenkin käyttö ja tulkinta ovat otettavissa haltuun.

Kaiken kaikkiaan, yksi tämän teoksen johtoajatus, jota olemme teoksessa halunneet korostaa, on se, että tilastolliseen tutkimukseen sisältyy aina tulkintaa, kuten kaikkeen empiiriseen tutkimukseen yleensäkin. Tutkimukseen liittyvän tulkinnan vuoksi on erityisen tärkeää se, että tutkija tuntee mahdollisimman hyvin tutkimansa tutkimusalueen tutkimus- ja tulkintatraditioita ja on tarkkana (ja tietoinen) käyttämiensä tilastomenetelmienkin suhteen. Tämän vuoksi suosittelemme kirjan lukijoille tutustumista muidenkin vastaavien menetelmäkirjojen antiin. Voimme oman kokemuksemme perusteella sanoa, että niistä jokaisesta saa jotain arvokasta ja ajateltavaa sen ymmärtämiseksi, miksi tutkimusta yleensä tehdään, ja siihen, miten tämä kannattaisi (ja pitäisi) tehdä. Myös tutustumalla oman aiheen tutkimusraportteihin ja -artikkeleihin saa runsaasti vinkkejä siitä, mitkä menetelmät voisivat soveltua oman tutkimusasetelman kysymysten ja tutkimusaineiston analysointiin ja myös siihen, miten näiden tuloksia yleensä tulkitaan. Oman alan tutkimuskirjallisuuteen tutustuminen on tämänkin vuoksi tärkeää ja antoisaa. Mitä paremmin hallitsee menetelmiin liittyviä kysymyksiä, sitä enemmän saa itsekin irti tutkimuksen teosta ja näin esiin nousevista uusista tulkinnoista, ja joka tapauksessa tutkimuksen tuloksista tulee näin luotettavampia ja uskottavampia.

Kirjallisuus ja muut lähteet

Alkula, T., Pöntinen, S. & Ylöstalo, P. 2002. Sosiaalitutkimuksen kvantitatiiviset menetelmät. 4.p. Porvoo-Helsinki-Juva: WSOY.

Anglim, J. 2020. What are the differences between Factor Analysis and Principal Component Analysis? (Luettu 22.1.2020). Cross Validated. https://stats.stackexchange.com/questions/1576/what-are-the-differences-between-factor-analysis-and-principal-component-analysi

APA 2020. American Psychological Association. Publication Manual of the American Psychological Association (7th ed.). https://doi.org/10.1037/0000165-000.

Atjonen, P. 2008. Miksi kasvattaja ja kasvatustieteilijä tarvitsevat etiikkaa? Pauli Siljander & Ari Kivelä (toim.) Kasvatustieteen tila ja tutkimuskäytännöt. Paradigmat katosivat, mitä jäljellä? Kasvatusalan tutkimuksia 38. Helsinki: Suomen Kasvatustieteellisen seuran Kasvatusalan tutkimuksia, 113–142.

Bhattacherjee, A. 2012. Social Science Research: Principles, Methods, and Pratices. 2.p. (Luettu 11.12.2019) http://scholarcommons.usf.edu/cgi/viewcontent.cgi?article=1002&context=oa_textbooks.

Brown, T. A. 2015. Confirmatory Factor Analysis for Applied Research, 2.p., New York: Guilford Press.

Bulmer, M. G. 1979. Principles of Statistics. New York, Dover Publications.

Byrne, B. 2011. Structural Equation Modeling with Mplus. London: Routledge.

Child, D. 1973. The Essentials of Factor Analysis.3.p. London: Holt, Rinehart & Winston.

Cohen, J. 1988. Statistical power analysis for the behavioural sciences, 2.p. Hillsdale. NJ: Lawrence Erlbaum.

Cohen, L. & Manion, L. 2007.Research Methods in Education.London: Routledge.

Cumming, G. & Calin-Jageman, R. 2017. Introduction The New Statistics. Estimation, Open Science, & Beyond. New York and London: Routledge.

Ellis, P. D. 2010. The Essential Guide to Effect Sizes. Statistical Power, Meta-Analysis, and the Interpretation of Research Results. Cambridge: Cambridge University Press.

Eskola, A. 1975. Sosiologian tutkimusmenetelmät II. 2.p:n toinen muuttumaton lisäpainos. Porvoo-Helsinki: WSOY.

Everitt, B. S., Landau, S., Leese, M. & Stahl, D. 2011. Cluster Analysis, 5.p. Chichester: John Wiley & Sons.

Field, A. P. 2018. Discovering statistics using IBM SPSS Statistics. 5th ed. London: Sage.

Field, A. P. 2016. An Adventure in Statistics. The reality enigma. Lontoo: Sage Publications.

Fielding, J. & Gilbert, N. 2006. Understanding social statistics. 2.p. Lontoo: Sage Publications.

Freedman, D.A. 2010. Statistical Models and Causal Inference. A Dialogu with tje Social Sciences. Cambridge: Cambridge University Press.

Gill, J. 2002. Bayesian methods. A social and behavioral sciences approach. 2.p Boca Raton, Fla: Chapman & Hall/CRC.

Gröönroos 2011. Johdatus tilastotieteeseen. Kuvailu, mallit ja päättely. 4.p. Tampere: Finn Lectura.

Hair, J.F., Black, W.C., Babin, B.J. & Anderson, R.E. 2010. Multivariate data analysis. 7.p. New Jersey: Pearson.

Hakala, J. T. 2017. Tulevan maisterin graduopas. 3. uudistettu painos. Helsinki: Gaudeamus.

Hallamaa, J. & Lötjönen, S. 2002. Suomalainen tiedeyhteisö ja tutkimusetiikka. Teoksessa S. Karjalainen, V. Launis, R. Pelkonen & J. Pietarinen (toim.) Tutkijan eettiset valinnat. Helsinki: Gaudeamus, 372–383.

Helenius, H. 1989. Tilastollisten menetelmien perustiedot. Salo: Statcon.

Heinonen, V. 1989. Kasvatustieteen perusteet. Jyväskylä: Gummerus.

Heikkilä, T. 2014. Tilastollinen tutkimus. 9. uudistettu p. Helsinki: Edita.

Hirsjärvi, S., Remes, P. & Sajavaara, P. 2009. Tutki ja kirjoita. 15. uudistettu painos. Helsinki: Tammi.

Hirvonen, A. 2006. Eettisesti hyvä tutkimus. Teoksessa J. Hallamaa, V. Launis, S. Lötjönen & I. Sorvali (toim.) Etiikka Ihmistieteille. Helsinki: Suomalaisen Kirjallisuuden Seura, 31–49.

Huff, D. 1954. How to Lie with Statistics. New York: W. W. Norton & Company.

Högmander, H., Kankainen, A., Kärkkäinen S., Leskinen, E., Lyyra, A., Nissinen, K., Pahkinen, E. 2006. Tilastolliset analyysimenetelmät, osa II. 5. uudistettu painos. Jyväskylän yliopiston matematiikan ja tilastotieteen laitoksen opetusmonisteita,

Ikonen, R. 2016. Kvantitatiivinen tutkimus aikuiskasvatustieteessä. Aikuiskasvatus (2), 114–118.

Jokivuori, P. & Hietala, R. 2007. Määrällisiä tarinoita - monimuuttujamenetelmien käyttö ja tulkinta. Helsinki: WSOY.

Järvikoski, A. & von Wright. J.M. 1970. Varianssianalyysin alkeita. Turun yliopiston psykologian laitoksen opetusmonisteita 1.

Kaidesoja, T., Kankainen, T. & Ylikoski, P. (toim.) 2018. Syistä selityksiin. Kausaalisuus ja selittäminen yhteiskuntatieteissä. Helsinki: Gaudeamus.

Kaiser, H. F. & Rice, J. 1974. Little jiffy, Mark 4. Educational and Psychological Measurement 34 (1), 111-117.

Kakkuri-Knuutila, M-L. 1992. KAS. Kuvaus, argumentti ja selitys tutkimusraportissa. Helsingin kauppakorkeakoulun julkaisuja D 165.

Kakkuri-Knuutila, M-L. 2006. Kausaalisuhteet ja selittäminen tulkitsevassa tutkimuksessa. Teoksessa Rolin, K., Kakkuri-Knuutila, M-L. & Hentonen Elina. Helsinki: Gaudeamus, 54–87.

Kakkuri-Knuuttila, M-L. & Heinilahti, K. 2006. Mitä on tutkimus? Argumentaatio ja tieteenfilosofia. Helsinki: Gaudeamus.

Kari, J. & Huttunen, J. 1994. Johdatus kasvatuksen ongelmien tutkimiseen. 1.–4.p. Helsinki: Otava.

Karjalainen, L. 2000. Tilastomatematiikka. Mikkeli: Pii-Kirjat.

Karma, K. & Komulainen, E. 1990. Käyttäytymistieteiden tilastomenetelmien jatkokurssi. Helsinki: Yliopistopaino.

Karvanen, Juha, Luoma, Arto, Penttinen, Antti & Tikka, Santtu 2019 Bayes-tilastotiede 2

Jyväskylän yliopisto, Matematiikan ja tilastotieteen. Luettu 30.3.2020) http://users.jyu.fi/~santikka/bayes2/Bayes-tilastotiede_2.pdf

Kerlinger, F. N. 1981. Foundation of Behavioral Research. 2.p. Tokio: Holt-Saunders International Editions.

Ketokivi, M. 2015. Tilastollinen päättely ja tieteellinen argumentointi. 2. uud.p. Helsinki: Gaudeamus.

Kline, R. 2013. Beyond Significance Testing: Statistics Reform in the Behavioral Sciences (2nd ed.). American Psychological Association.

Kline, R. 2016. Principles and Practice of structural equation modeling. 4.p. New York: Guilford Press.

Koch, K-R. 2007. Introduction to Bayesian statistics. 2. ed. Berlin & New York: Springer.

Kuula, A. 2011. Tutkimusetiikka. Aineiston hankinta, käyttö ja säilytys. 2. uudistettu painos. Tampere: Vastapaino.

Kuusela, V. 2000. Tilastografiikan perusteet, Helsinki: Edita.

KvantiMOTV - Menetelmäopetuksen tietovaranto. havaintoaineiston painottaminen [verkkojulkaisu]. Tampere: Yhteiskuntatieteellinen tietoarkisto [ylläpitäjä ja tuottaja]. <https://www.fsd.tuni.fi/menetelmaopetus/>. (Viitattu 28.02.2020.)

Lehtonen, R. & Pahkinen, E. 2004. Practical Methods for Design and Analysis of Complex Surveys. 2. p. Chichester: John Wiley & Sons.

Leiman, M. & Toivonen, V-M. 1991. Kadonnutta reliabiliteettia etsimässä. Psykologia 26 (3), 180–186.

Leskinen, E. 1987. Faktorianalyysi. Konfirmatoristen faktorimallien teoria ja rakentaminen. Jyväskylän yliopiston tilastotieteen laitoksen julkaisuja 1.

Marascuilo, L.A. & McSweeney, M. 1977. Nonparametric and Distribution-Free Methods for the Social Sciences. Monterey: Brooks/Cole Publishing Company.

Metsämuuronen, J. 2009. Tutkimuksen tekemisen perusteet ihmistieteissä. 4. laitos, 1 p. Helsinki: International Methelp.

Nokelainen, P. 2010. Bayesilaisen tilastoanalyysin käyttömahdollisuudet ammattikasvatuksen tutkimuksessa. Ammattikasvatuksen aikakauskirja 12 (1), 34–46.

Nummenmaa L. 2005. Efektikoko psykologisessa tutkimuksessa. Psykologia 40 (5–6), 559–567.

Nummenmaa, L. 2011. Käyttäytymistieteiden tilastolliset menetelmät. 3.p. (uud. laitos). Helsinki: Tammi.

Nummenmaa, T., Konttinen, R., Kuusinen, J. & Leskinen, E. 1997. Tutkimusaineiston analyysi. Porvoo-Helsinki: WSOY.

Nurminen, M. 1997. Olemmeko p-arvon orjia? Duodecim 113 (4), 277–280.

Pahkinen, E. 2012. Kyselytutkimusten otantamenetelmät ja aineistoanalyysi. Jyväskylä: Julpu.

Patton, M. Q. 1990. Qualitative Evaluation and Research Methods.London: Sage Publishers.

Piispa, M. 2006. Kvantitatiivisen tutkimuksen eettiset lähtökohdat. Esimerkkinä naisiin kohdistuvan väkivallan kyselytutkimus. Teoksessa J. Hallamaa, V. Launis, S. Lötjönen & I. Sorvali (toim.) Etiikka Ihmistieteille. Helsinki: Suomalaisen Kirjallisuuden Seura, 141–160.

Pohjola, A. 2007. Eettisyyden haaste tutkimuksessa. Leena Viinamäki & Erkki Saari (toim.) Polkuja soveltavaan yhteiskuntatieteelliseen tutkimukseen. Helsinki: Tammi, 11–31.

Reynolds, H.T. 1977 Analysis of Nominal Data. Beverly Hills: Sage publications. (Series: Quantitative Application in the Social Sciences 07-007).

Ronkainen, S. 2008. Otanta, edustavuus ja kadon analyysi. Sähköä kyselyyn! Web-kysely tutkimuksessa ja tiedonkeruussa. Rovaniemi: Lapin yliopisto, 70–76

Ronkainen, S. & Karjalainen, A. 2008. Sähköä kyselyyn! Web-kysely tutkimuksessa ja tiedonkeruussa. Rovaniemi: Lapin yliopisto.

Rose, G. 1991. Deciphering Sociological Research.5.p. Houndmills: Macmillan.

van de Schoot, R., Kaplan, D., Denissen, J., Asendorpf, J.B., Neyer F.J., van Aken, M.A.G. 2014. A gentle introduction to bayesian analysis: applications to developmental research. Child Development 85 (3), 842-860.

Siegel, S. 1956. Nonparametric Statistics for the Behavioral Sciences. New York: McGraw-Hill Book Company, INC.

Siegel, S. & Castellan Jr, N. J. 1988. Nonparametric Statistics for The Behavioral Sciences. 2. p. New York, N.Y. : McGraw-Hill.

Sänkiaho, R. 1974. Temput ja kuinka ne tehdään. Monimuuttujamenetelmät kansan palvelijoina. Jyväskylän Yliopisto. Kasvatustieteiden tutkimuslaitoksen julkaisuja 220/1974.

Thalheimer, W., & Cook, S. 2002. How to calculate effect sizes from published research articles: A simplified methodology. Retrieved April 14, 2020 from http://work-learning.com/effect_sizes.htm.

Toivonen, T. 999. Empiirinen sosiaalitutkimus. Filosofia ja metodologia. Porvoo-Helsinki-Juva: WSOY.

Toivonen, T. 2005. Elaboraatio ja interaktio. Eli mitä log-lineaarisilla malleilla on hankala havaita. Pekka Räsänen, Anu-Hanna Anttila & Harri Melin (toim.) Tutkimus menetelmien pyörteissä. PS-kustannus. Jyväskylä.

Tuckman, B. W. 1988. Conducting Educational Research.3.p. San Diego: Harcourt Brace Jovanovich, Publishers.

Töttö, P. 2004. Syvällistä ja pinnallista. Teoria, empiria ja kausaalisuus sosiaalitutkimuksessa. Tampere: Vastapaino.

Töttö, P. 2012. Paljonko on paljon. Luvuilla argumentoinnista empiirisessä tutkimuksessa. Tampere: Vastapaino.

Töttö, P. 2018. Tilastolliset mallit ja mekanismeilla selittäminen. Teoksessa T. Kaidesoja, T. Kankainen & P. Ylikoski (toim.) 2018. Syistä selityksiin. Kausaalisuus ja selittäminen yhteiskuntatieteissä. Helsinki: Gaudeamus, 136–166.

Tilastokeskus: Tietoa tilastoista, käsitteet. https://www.stat.fi/meta/index.html (luettu 1.15.2020)

Tutkimuseettinen neuvottelukunta 2012. Hyvä tieteellinen käytäntö ja sen loukkausepäilyjen käsitteleminen Suomessa. Helsinki: Tutkimuseettisen neuvottelukunta. https://www.tenk.fi/sites/tenk.fi/files/HTK_ohje_2012.pdf (luettu 1.8.2019).

Varto, J. 2011. Miksi miettiä metodologioita? Teoksessa K. Holma & K. Mälkki (toim.) Tutkimusmatkalla. Metodologia, teoria ja filosofia kasvatustutkimuksessa. Helsinki: Gaudeamus, 13–24.

Vanttaja, M., Tähtinen, J., Zacheus, T. & Koski, P. 2017. Liikkumattomuuden jäljillä. Pitkittäistutkimus vähän liikuntaa harrastavien nuorten liikuntasuhteista ja liikunta-aktiivisuuden muutoksista. Helsinki: Nuorisotutkimusverkoston ja Nuorisotutkimusseura.

Vehkalahti, K. 2014. Kyselytutkimuksen mittarit ja menetelmät. Helsinki: Finn Lectura.

Vuorinen, R. 1982. Psykologisen tutkimuksen ja sovelluksen metodologiset perusteet. Turun yliopiston filosofian, psykologian ja menetelmätieteiden laitoksen psykologian opintomonisteita 2.

Wasserstein, R.L., Schrim, A.L. & Lazar, N.A. 2019. Moving to a World Beyond ”p<0.05”. The American Statistician 73:sup1, 1–19. DOI: 10.1080/00031305.2019.1583913.

Ylikoski, P. 2018. Selittäminen, ymmärtäminen ja kausaaliset mekanismit. Teoksessa T. Kaidesoja, T. Kankainen & P. Ylikoski (toim.) 2018. Syistä selityksiin. Kausaalisuus ja selittäminen yhteiskuntatieteissä. Helsinki: Gaudeamus, 20–54.

Asiahakemisto

O

P

R

V

W

Y

Z

Ä

Esimerkeissä käytetty havaintoaineisto

id	sp	vkou	vir_t	verb_t	järk_t	kieli_ka	matik	omene1	omene2	kmer_1	kmer_2
1	1	1	1	22	37	63	8	13	16	1	3
2	2	1	2	26	32	82	7	16	18	3	3
3	2	2	3	29	33	96	5	17	20	3	3
4	1	3	3	24	36	73	8	15	19	2	3
5	1	2	2	24	35	67	9	16	14	2	2
6	2	1	2	28	36	92	7	19	20	3	3
7	2	1	1	23	34	72	7	14	14	3	3
8	1	1	1	24	33	72	6	13	13	1	2
9	1	2	3	26	37	79	9	17	19	3	3
10	2	2	3	29	34	97	8	19	19	3	3
11	2	2	2	28	35	96	7	18	14	3	2
12	1	1	2	26	36	86	7	17	15	2	3
13	1	1	2	27	37	87	8	16	18	3	3
14	2	3	2	25	34	82	7	18	15	3	3
15	1	1	1	20	33	62	6	14	13	2	2
16	1	3	3	26	40	81	10	18	19	3	3
17	2	3	3	28	36	92	9	17	18	2	3
18	2	2	2	27	35	89	6	16	13	3	1
19	2	1	1	26	34	79	7	15	13	2	2
20	2	1	1	22	31	72	6	14	14	1	3
21	1	2	1	21	33	69	6	13	15	1	3
22	1	2	2	25	36	75	8	15	18	2	3
23	1	3	2	26	35	76	7	14	13	1	2
24	2	2	2	27	34	86	7	16	15	3	3
25	1	3	3	28	37	84	8	17	19	3	3
26	2	2	3	30	36	85	7	18	20	3	3
27	2	1	2	27	34	89	6	17	14	3	2
28	1	2	2	25	36	79	7	16	14	2	2
29	1	2	1	21	36	59	7	13	14	1	3
30	2	3	1	24	34	71	7	14	14	2	2
31	1	1	1	22	37	63	8	13	16	1	3
32	2	1	2	26	32	82	7	16	18	3	3
33	2	2	3	29	33	96	5	17	20	3	3
34	1	3	3	24	36	73	8	15	19	2	3
35	1	2	2	24	35	67	9	16	14	2	2
36	2	1	2	28	36	92	7	19	20	3	3
37	2	1	1	23	34	72	7	14	14	2	3
38	1	1	1	24	33	72	6	13	13	1	2
39	1	2	3	26	37	79	9	17	19	3	3
40	2	2	3	29	34	97	8	19	19	3	3
41	2	2	2	28	35	96	7	18	14	3	2
42	1	1	2	26	36	86	7	17	15	2	3
43	1	1	2	27	37	87	8	16	18	3	3
44	2	3	2	25	34	82	7	18	15	3	3
45	1	1	1	20	33	62	6	14	13	2	2
46	1	3	3	26	40	81	10	18	19	3	3
47	2	3	3	28	36	92	9	17	18	2	3
48	2	2	2	27	35	89	6	16	13	3	1
49	2	1	1	26	34	79	7	15	13	2	2
50	2	1	1	22	31	72	6	14	14	1	3
51	1	2	1	21	33	69	6	13	15	1	3
52	1	2	2	25	36	75	8	15	18	2	3
53	1	3	2	26	35	76	7	14	13	2	2
54	2	2	2	27	34	86	7	16	15	3	3
55	1	3	3	28	37	84	8	17	19	3	3
56	2	2	3	30	36	85	7	18	20	3	3
57	2	1	2	27	34	89	6	17	14	3	3
58	1	2	2	25	36	79	7	16	14	3	2
59	1	2	1	21	36	59	7	13	14	2	3
60	2	3	1	24	34	71	7	14	14	2	2

Aineisto on tallennettavissa kirjan verkkosivuilta www.utu.fi/tilastoanalyysiopas. Tarkempi kuvaus muuttujista löytyy kirjan luvusta 2.1 (s. 62)

Käsissäsi oleva teos on toinen uudistettu ja täydennetty painos vuoden 2011 painoksesta. Teoksen saaman positiivisen palautteen vuoksi olemme halunneet päivittää sen. Aiempaan teokseen verrattuna tähän versioon on lisätty erityisesti määrällisen tutkimuksen suunnitteluun, toteutukseen ja tulosten tulkinnan merkitystä korostavaa tarkastelua. Olemme myös lisänneet teokseen niin sanotun uuden tilastotieteen esittämien tulkintojen innoittamana erilaisia analyysimenetelmiä ja lisätarkasteluja, joiden avulla voidaan vahvistaa tutkimuksen otoksesta tutkimuksen populaatioon tehtävien tulkintojen luotettavuutta ja uskottavuutta. Pääpaino teoksessa on kuitenkin edelleen tilastollisten perusmenetelmien ja niihin liittyvien tulkintojen avaamisessa, kirja onkin edelleen suunnattu ennen kaikkea tilastollisten analyysien maailmaan ensiaskeleita ottaville, esimerkiksi kasvatustieteen ja yhteiskuntatieteen alan opiskelijoille. Tosin vuosien varrella olemme saaneet kuulla monilta vähän pidemmälle ehtineiltä jatko-opiskelijoilta ja tutkijoilta kirjan palvelleen heitäkin monessa suhteessa.

ISBN 978-951-29-8090-1 (PRINT)
ISBN 978-951-29-8091-8 (PDF)
ISSN 1237-7643
Painosalama Oy – Turku, 2020

Turun yliopisto,
Kasvatustieteiden tiedekunta,
Julkaisusarja C:22