



PRO GRADU -TUTKIELMA
Matematiikan ja tilastotieteen laitos
Tilastotiede
Helmikuu 2008

Heidi Kainulainen

Klusteroitunut aineisto kahden ryhmän vertailussa

Tampereen yliopisto

Matematiikan ja tilastotieteen laitos

KAINULAINEN, HEIDI: Klusteroitunut aineisto kahden ryhmän vertailussa

Pro gradu -tutkielma, 59 s., 17 liites.

Tilastotiede

Helmikuu 2008

Tiivistelmä

Havaintoaineisto voi koostua keskenään riippumattomista, homogeenisista osista, klustereista, joiden sisällä havainnot ovat keskenään riippuvia. Tässä tutkielmassa tarkastellaan klusteroitunutta aineistoa kahden ryhmän sijainnin vertailussa. Tunnetuimpia klassisia menetelmiä ovat kahden riippumattoman otoksen Studentin t-testi ja Wilcoxonin järjestyslukutesti. Klassinen tilastollinen päättely ei kuitenkaan huomioi klusterin sisäistä korrelaatiota ja antaa siten harhaanjohtavia tuloksia. Tutkielmassa esitellään uudet, klusteroidut versiot t-testistä sekä Wilcoxonin järjestyslukutestistä, joissa klusteroituneen, yksiulotteisen aineiston kahden ryhmän sijaintiongelmaa lähestytään uudella tavalla. Klusteroidut testit huomioivat klusterin sisäisen riippuvuuden estimaattorin varianssin estimoinnissa.

Klusteroituja ja klassisia menetelmiä vertaillaan simulointikokeissa testien voimakkuuksien avulla. Vertailu osoittaa, että klusteroituneen aineiston tilanteessa klusteroidut testit ovat valideja ja tehokkaampia kuin klassiset menetelmät. Tällöin perinteiset menetelmät osoittautuvat konservatiivisiksi tai ne antavat virheellisiä tuloksia eivätkä ole valideja testejä. Klusteroiduista testeistä Wilcoxonin testi on tehokkaampi kuin t-testi, jos jakauma on vino tai paksuhäntäinen. Lisäksi t-testi ei ole robusti vaan herkkä poikkeaville havainnoille.

Tutkimusaineisto koostuu Sosiaali- ja terveysalan tutkimus- ja kehittämiskeskuksen järjestämän valtakunnallisen Kouluterveyskyselyn pääkaupunkiseudun koulujen kahdeksas- ja yhdeksäsluokkalaisten oppilaiden vastauksista vuodelta 2006. Havaintoyksikköjä on 17078. Empiirisen aineiston klustereina pidetään kouluja. Vastemuuttujana on nuorten masentuneisuusoireet, joiden mittaamisessa on käytetty Raimo Raitasalon Suomeen soveltamaa Mielialakyselyä, Beck Depression Inventory -indikaattoria. Havaitaan, että masentuneisuus riippuu sukupuolesta, liikuntaharrastuksista, vanhempien työttömyydestä, seksuaalikokemuksista, suhteesta vanhempiin ja suhteesta opettajiin.

Asiasanat klusteri, klusterin sisäinen korrelaatio, Studentin t-testi, Wilcoxonin järjestyslukutesti

Sisältö

1	Johdanto	5
1.1	Taustaa	5
1.2	Tutkielman tavoitteet	6
1.3	Tutkielman rakenne	7
2	Tutkimusaineisto ja -ongelma	9
2.1	Kouluterveyskysely	9
2.2	Tutkimusongelma	10
2.3	Tutkimusaineisto	11
2.3.1	Vastemuuttuja: masentuneisuus	11
2.3.2	Ryhmittelevät muuttujat	12
3	Klassiset menetelmät kahden riippumattoman otoksen vertailussa	15
3.1	Studentin t-testi	15
3.2	Wilcoxonin järjestyslukutesti	17
3.3	Tasatulosten käsittely Wilcoxonin testissä	19
4	Klusteroitunut aineisto	21
4.1	Klusterin sisäinen riippuvuus	21
4.2	Klusterirakenteet	22
4.3	Klusteroitu t-testi	23
4.4	Klusteroitu Wilcoxonin testi	25
5	Menetelmien vertailu klusteroituneessa aineistossa	28
5.1	Tutkimusmenetelmät	28
5.2	Klusteroitujen mallien muodostaminen	28
5.3	Simuloinnit: rakennetyyppi <i>I</i>	30
5.3.1	Havainnot normaalijakaumasta	31
5.3.2	Havainnot t-jakaumasta	31
5.4	Simuloinnit: rakennetyyppi <i>II</i>	35
5.4.1	Havainnot normaalijakaumasta	36
5.4.2	Havainnot t-jakaumasta	36
6	Soveltaminen Kouluterveyskysely-aineistoon	43
6.1	Tutkimusmenetelmät	43
6.2	Osaotosten testausasetelma	44

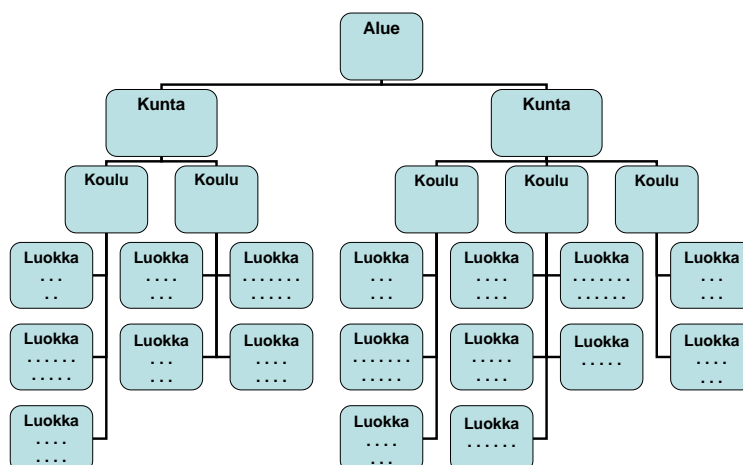
6.3	Analyyssien tulokset	44
6.3.1	Sukupuoli	47
6.3.2	Liikuntaharrastukset	47
6.3.3	Vanhempien työttömyys	49
6.3.4	Seksuaalikokemukset	51
6.3.5	Suhde vanhempiin	51
6.3.6	Suhde opettajiin	53
6.4	Analyyssien yhteenveto	55
7	Yhteenveto	56
	Lähdeluettelo	58
A	R-funktiot	60
A.1	t.test.cluster	60
A.2	wilcox.cluster	61
B	Indikaattorimuuttujien muodostaminen	62
C	Simulointien tulokset	64
C.1	Klusterirakennetyyppi I	64
C.2	Klusterirakennetyyppi II	69

1 Johdanto

1.1 Taustaa

Yleisterminä klusteri (ryväs) tarkoittaa joukkoa samanlaisia kohteita, jotka jollain tavoin pysyvät yhdessä tai hyötyvät keskinäisestä yhteydestään. Tässä tutkielmassa klusteroitumisella tarkoitetaan hierarkkista rakennetta, jossa tietyllä joukolla on jokin yhdistävä ja ryhmittelevä tekijä. Lisäksi ajatellaan, että tämä yhdistävä tekijä voi aiheuttaa samankaltaisuutta mitattaessa saman klusterin jäseniltä jotakin ominaisuutta tai asiaa. Tämä tarkoittaa sitä, että klusterin sisällä havainnot voivat olla korreloituneita keskenään, mutta samalla eri klusterien havaintojen ajatellaan olevan toisistaan riippumattomia. Käytännössä aineisto on siis luonnostaan klusteroitunut, eikä sitä klusteroida.

Ajatellaan esimerkiksi yhden koulun oppilaita, jotka voivat olla samankaltaisia keskenään, mutta erilaisia kuin toisen koulun oppilaat. Toisaalta saman koulun oppilaat ovat samankaltaisempia keskenään verrattuna oppilaisiin, jotka valitaan satunnaisesti kaikista Suomen kouluista. Hierarkkinen rakenne voi olla myös monitasoinen, jolloin klusteritasoja ja yhdistäviä tekijöitä on useita. Esimerkkinä monitasoisen hierarkkarakenteen klustereista voisi olla vaikka alue (lääni) → kunta → koulu → luokka, kuten kuviossa 1.1 on havainnollistettu.



Kuvio 1.1. Esimerkki monitasoisesta hierarkkarakenteesta, missä klustereina ovat alue, kunta, koulu ja luokka. (Luokka-laatikon sisällä piste kuvaa oppilasta.)

Klusteroinnin vaikutuksia perinteisten tilastollisten menetelmien validisuuteen on tutkittu paljon parin viimeisimmän vuosikymmenen aikana. Kuitenkin vasta viime vuosina tutkimustyö on keskittynyt myös epäparametrisiin (*non-parametric*) menetelmiin, kun puhutaan nimenomaan klusteroituneesta aineistosta. Saman klusterin havainnot ovat siis yleensä korreloituneita, mutta samalla eri klustereista tulevien havaintojen oletetaan olevan riippumattomia keskenään. Jos klusterien sisäistä riippuvuutta ei oteta huomioon tilastollisissa testeissä, testien p-arvot voivat olla liian pieniä tai suuria tai jotkut testit jopa epävalideja. Näin ollen analyysien tulokset voivat olla harhaanjohtavia. Tällöin edellisessä kappaleessa mainittu esimerkki koulun oppilaista tarkoittaisi sitä, että eri kouluilla ei olisi mitään roolia vaan kaikki oppilaat olisivat ikään kuin yhdestä koulusta.

Useimmiten klusteroitunutta aineistoa koskeva kirjallisuus käsittelee normaali- tai binomijakaumaa noudattavia populaatioita (esimerkiksi Rosner 1984; Liang & Zeger 1986). Tyypillisesti klusteroitunut aineisto analysoidaan käyttäen GEE-menetelmää (yleistetty estimointiyhtälö, *Generalized Estimating Equations*), joka olettaa vasteen jakauman normaalisuuden (Williamson, Datta & Satten 2003; Stoner & Leroux 2002).

Rosner, Glynn ja Lee (2003) esittivät epäparametrisen järjestyslukuihin perustuvan kahden otoksen testin klusteroituneelle aineistolle, kun klusterin jäsenet kuuluvat samaan ryhmään. Datta ja Satten (2005) kehittivät menetelmää edelleen, ja toisin kuin aikaisemmissa klusteroituneen aineiston järjestyslukutesteissä, saman klusterin jäsenet voivat kuulua kahteen (vrt. Wilcoxonin järjestyslukutesti) tai useampaan vertailtavaan ryhmään (vrt. Kruskal-Wallis testin). Lisäksi Dattan ja Sattenin testissä eri ryhmien havaintojen korrelaatiot klusterin sisällä voivat erota toisistaan.

Larocque (2003) laajensi moniulotteisen epäparametrisen yhden otoksen merkkitestin (affine invariant multivariate sign test) klusteroituneelle aineistolle. Nevalainen, Larocque ja Oja (2007a; 2007b) ovat myös lähestyneet vastaavaa moniulotteista yhden otoksen estimointiongelmää samansuuntaisesti, kuten myöskin Larocque, Nevalainen ja Oja (2007), jotka tarkastelivat yhden otoksen painotettua merkkitestiä.

1.2 Tutkielman tavoitteet

Tässä tutkielmassa esitellään kahden ryhmän sijainnin vertailuun uusi mallipohjainen lähestymistapa klusteroituneelle aineistolle. Käsittely rajataan yhteen klusteritasoon. Uudessa menetelmässä estimaattorin (ryhmäkeskiarvojen erotus) estimoitu varianssi huomioi klusterien sisäisen korrelaation (*intracluster correlation*). Tällöin analyysien tulosten oletetaan antavan oikeampia tuloksia verrattuna perinteisiin menetelmiin, jotka eivät huomioi klusteroitumista. Uudesta menetelmästä esitetään sekä parametrinen että epäparametrinen testi: niin sanottu klusteroitu versio Studentin t-testille ja Wilcoxonin järjestyslukutestille. Testisuureiden merkitsevyyksien laskemisessa käytetään normaaliap-

proksimaatiota asymptomatiikan perusteella ja esitetyt menetelmät on tarkoitettu suurille otoskoille, mikä tässä tarkoittaa suurta klusterien määrää.

Simulointikokeiden avulla vertaillaan eri menetelmiä ja tarkastellaan niiden antamien tulosten eroavuuksia kahden otoksen tilastollisissa testeissä: sekä perinteisessä Studentin t-testissä ja Wilcoxonin järjestyslukutestissä että uudessa, klusteroidussa t-testissä ja Wilcoxonin testissä. Tutkielman tavoitteena on tarkastella, kuinka klusterin sisäinen riippuvuus aineistossa vaikuttaa näihin testeihin. Simulointien avulla osoitetaan, miten menetelmät reagoivat klusterin sisäiseen korrelaatioon ja miten korrelaatio vaikuttaa tulosten luotettavuuteen. Työssä osoitetaan, että uudet testit huomioivat klusterin sisäisen riippuvuuden ja ovat valideja, kun taas perinteiset menetelmät ovat epävalideja ja antavat jopa vääriä tuloksia klusteroituneen aineiston tapauksessa.

Sovellusesimerkkinä käytettävä Kouluterveyskysely-aineisto on tyypillinen klusteroitunut aineisto. Klustereina pidetään kouluja, joiden oppilaiden vastauksissa voidaan olettaa jonkinasteista riippuvuutta. Esimerkiksi suurkaupungin koulun oppilaiden vastaukset ovat ehkä samansuuntaisia, mutta poikkeavat maaseutukunnan koulun oppilaiden vastauksista (esim. Helsinki vs. Kerimäki). Tutkielman yhtenä tavoitteena on analysoida tutkimusaineistoa klassisilla sekä tässä työssä esitetyillä uusilla menetelmillä. Koululaisten masentuneisuusoireita yritetään selittää heidän henkilökohtaisilla ja sosiaalisilla ominaisuuksillaan ja vertailut tehdään kahden ryhmän tilanteissa.

1.3 Tutkielman rakenne

Luvun 1 johdannon jälkeen luvussa 2 esitellään tutkimusaineistona oleva Kouluterveyskysely-aineisto ja tutkimusongelma, nuorten masentuneisuus. Luvussa 3 esitellään kahden populaation vertailussa käytettävät perinteiset menetelmät: parametrisen Studentin t-testin sekä epäparametrinen Wilcoxonin järjestyslukutesti.

Luvussa 4 tarkastellaan klusteroitumista ja esitetään klusteroituneelle aineistolle tyypilliset rakennetyypit sekä uusien, klusteroitujen menetelmien teoria ja notaatio. Klusteroidut versiot esitetään kahden otoksen Studentin t-testille sekä järjestyslukuihin perustuvalle Wilcoxonin testille.

Luvussa 5 vertaillaan eri tilastollisia menetelmiä simulointikokeiden avulla ja havainnollistetaan tuloksia kuvioilla. Sopivista jakaumista generoidaan klusteroituneita otoksia kahden ryhmän vertailuun ja otokset analysoidaan luvuissa 3 ja 4 esitetyillä parametrisilla ja epäparametrisilla testeillä. Generoiduista otoksista määritellään vertailtaville testeille voimakkuudet, joiden avulla testien tehokkuuksia vertaillaan klusteroituneessa aineistossa.

Luvussa 6 sovelletaan edellä esiteltyjä menetelmiä Kouluterveyskysely-aineistoon. Empiirisen osan vastemuuttujana on nuorten masentuneisuusoireet, joita selitetään sukupuoli, liikuntaharrastuksilla, vanhempien työttömyydellä, seksuaalikokemuksilla, suhteella vanhempiin sekä suhteella opettajiin. Lopuksi luvussa 7 on yhteenvedo ja pohdintaa aiheesta. Lähdeluettelo löytyy si-

vulta 58.

Esitetyille uusille klusteroiduille testeille on koodattu R-funktiot ja ne esitetään liitteessä A. Tutkimusaineistona olevan Kouluterveyskysely-aineiston indikaattorimuuttujien muodostaminen esitetään liitteessä B sekä simulointien numeeriset tulokset liitteessä C.

2 Tutkimusaineisto ja -ongelma

Tutkimusaineistona käytetään Sosiaali- ja terveysalan tutkimus- ja kehittämiskeskuksen (Stakes) järjestämän Kouluterveyskyselyn vastauksia. Tässä luvussa kerrotaan taustatietoja Kouluterveyskyselystä sekä esitellään tutkimusongelma, aineisto, vastemuuttuja ja kahden ryhmän vertailussa käytettävät ryhmittelymuuttujat.

2.1 Kouluterveyskysely

Kouluterveyskysely on vuosittain peruskoulujen 8. ja 9. luokilla sekä lukioiden 1. ja 2. vuosikursseilla järjestettävä valtakunnallinen kysely, joka on otettu käyttöön vuonna 1996. Opettajan ohjaama luokkakysely järjestetään kuntien kaikissa peruskouluissa sekä lukioissa. Kouluterveyskysely on toteutettu vuodesta 2001 alkaen kaksijaksoisena siten, että se järjestetään samoissa kunnissa joka toinen vuosi huhtikuun aikana. Kahden peräkkäisen vuoden aineisto kattaa siis koko maan peruskoulujen yläluokat ja lukiot. Parillisina vuosina kyselyyn osallistuvat Etelä-Suomen, Itä-Suomen ja Lapin läänien kunnat ja parittomina vuosina Länsi-Suomen ja Oulun läänien sekä Ahvenanmaan kunnat. Kyselyssä kootaan valtakunnallisesti ja yhtenäisesti vertailukelpoista tietoa 14–18-vuotiaiden elinoloista, koulukokemuksista, terveydestä ja terveystottumuksista. (Luopa, Räsänen, Jokela & Rimpelä 2005.)

Kouluterveyskyselyyn ovat osallistuneet lähes kaikki kunnat, joissa on peruskoulun yläluokkia ja lukioita. Vuosina 1996–2006 kyselyyn on osallistunut lähes 400 kuntaa ja vuosittain kyselyyn on vastannut 46000–90000 oppilasta (Luopa, Sinkkonen, Jokela, Puusniekka & Pietikäinen 2006). Vastausprosentti kyselyyn osallistuvissa kunnissa on ollut noin 70–80.

Kouluterveyskyselyn kysymykset jaotellaan seuraavien aihepiirien mukaan: koulutyö, koulukiusaaminen, terveys, mieliala, terveysopetus, seurustelu ja seksuaalikäyttäytyminen, tupakointi, päihteet, liikuntatottumukset ja nukkuminen, ruokailutottumukset, rikkeet ja rikokset sekä koti ja ystävät. Kysymyksiä vuoden 2006 lomakkeella on 101, alakohtineen 264, joista enemmistö on toistunut samanlaisena vuodesta toiseen. Lomake on saatavilla Internetissä: <http://info.stakes.fi/kouluterveyskysely/FI/perustiedot/lomakkeet.htm>.

2.2 Tutkimusongelma

Tutkimusongelmana käsitellään nuorten masentuneisuutta. Tässä työssä tarkastellaan mahdollisia syitä ja yhteyksiä nuorten masentuneisuusoireiden sekä sukupuolen tai elinolojen tai sosiaalisten olosuhteiden välillä.

Mielenterveyden häiriöt ilmaantuvat usein ensimmäisen kerran jo ihmisen nuoruusiän aikana. Tutkimustieto osoittaa, että 20–25 % nuorista kärsii tavanomaisesta kehityksestä poikkeavasta pitkäkestoisesta psyykkisestä oireilusta. Nuorten mielialat voivat vaihdella normaalistikin heidän nopean fyysisen, psyykkisen ja sosiaalisen kasvun seurauksena. Tällaiset yksittäiset tunne-elämän oireet, jotka eivät täytä diagnostisia masennustilan kriteereitä, ovatkin tavallisia nuorille. Sanalla masennus voidaan arkikielessä tarkoittaa ohimenevää tunnetilaa, mutta tässä yhteydessä masennuksella tai depressiolla tarkoitetaan pidempikestoista ja voimakasta masennusoireyhtymää kuten lääketieteellisessä kielenkäytössä. (Karlsson & Marttunen 2007.)

Nuorten masennuksen taustalla voi olla biologisia, psykologisia ja sosiaalisia tekijöitä ja tavallisesti on kyse usean tekijän yhteisvaikutuksesta. Usein masennustilojen yhteydessä puhutaan ns. stressi-haavoittuvuusmallista, jonka mukaan depressioon sairastuminen on yksilöllisen haavoittumisalttiuden sekä erilaisten stressitekijöiden yhteisvaikutusten tulosta. Yksilölliseen haavoittuvuuteen vaikuttavat sekä elämäntapahtumat että elinolot, ja stressitekijöillä tarkoitetaan fyysisistä tai psykososiaalista tasapainotilaa uhkaavia tekijöitä. Toisaalta fyysiset sairaudet altistavat masennusoireilulle, mutta mahdollinen kausaalisuus voi olla myös vastakkaissuuntaista. On nimittäin huomioitu, että psyykkisen terveyden lisäksi masentuneisuusoireista kärsivien nuorten somaattinen terveys (fyysiset oireet: esim. vatsakipu, päänsärky) on huonompi kuin muiden nuorten. Lisäksi tutkimuksien mukaan nuoruusikäisten tyttöjen depressiot ovat noin kaksi kertaa yleisempiä kuin poikien. Edelleen sosiaalisen tuen puute sekä erilaiset perheen psykososiaaliset kuormitustekijät, kuten vanhempien ero, pitkäaikaiset ristiriidat, sairaudet ja taloudelliset vaikeudet, vaikuttavat myös nuoriin sekä suoraan että vanhempien välityksellä ja lisäävät riskiä sairastua depressioon nuoruusiällä. (Karlsson & Marttunen 2007.)

Tutkimushypoteeseina tutkitaan onko sukupuolella, liikuntaharrastuksilla, vanhempien työttömyydellä, seksuaalikokemuksilla, suhteella vanhempiin tai suhteella opettajiin yhteyttä nuorten masentuneisuuteen. Kiinnostuksen kohteena olevat selittävät ryhmittelymuuttujat on valittu edellä mainittujen ja oletettujen masentuneisuuden syiden perusteella. Tarkastelut tehdään kahden ryhmän tilanteissa, ja ryhmittelymuuttujat ovat indikaattorimuuttujia, jotka muodostuvat Kouluterveyskyselyn kysymysryppäistä. Ryhmittelevät muuttujat esitellään tarkemmin seuraavassa alaluvussa sekä indikaattorimuuttujien muodostaminen liitteessä B.

2.3 Tutkimusaineisto

Tutkimusaineistona käytetään Kouluterveyskyselyn vastauksia vuodelta 2006. Aineisto rajataan pääkaupunkiseudun neljän anonyymien kunnan koulujen kahdeksas- ja yhdeksäsluokkalaisten vastauksiin yhdeltä vuosikursilta. Taulukossa 2.1 nähdään Kouluterveyskyselyyn osallistuneiden kuntien sekä vastanneiden oppilaiden lukumäärät. Vuonna 2006 kyselyyn osallistui 115 koulusta 17078 oppilasta ja vastausprosentti oli noin 75. Vastanneista 50.4 % on tyttöjä ja 49.6 % poikia.

Taulukko 2.1. Pääkaupunkiseudun kuntien Kouluterveyskyselyyn vastanneiden koulujen ja 8.–9.-luokkalaisten oppilaiden lukumäärät vuodelta 2006.

Kunta	Vastanneet	
	koulut	oppilaat
Kunta 1	27	4179
Kunta 2	64	8595
Kunta 3	22	3864
Kunta 4	2	440
Yhteensä	115	17078

Vastaukset kysymyksiin ovat kategorisia, joko luokittelu- tai järjestysasteikollisia. Mielipidekysymyksissä on yleensä neljä tai viisi vastausvaihtoehtoa. Tässä tutkielmassa on käytetty vain osaa kysymyksistä tutkimusongelmien mukaisesti. Aineiston klusterimuuttujana on koulu. Selitettävänä muuttujana eli vastemuuttujana on nuorten *masentuneisuus* ja selittävinä muuttujina *sukupuoli*, *liikuntaharrastukset*, *vanhempien työttömyys*, *seksuaalikokemukset*, *suhde vanhempiin* ja *suhde opettajiin*.

2.3.1 Vastemuuttuja: masentuneisuus

Vastemuuttuja on nuorten masentuneisuus, joka on kvantitatiivinen muuttuja. Masentuneisuuden mittaamisessa on käytetty Raimo Raitasaloon Suomeen soveltamaa indikaattoria: Beckin lyhyttä depressiokyselyä eli Mielialakyselyä (Raitasalo 1995, liite). Beck Depression Inventory (BDI) kehitettiin yli neljäkymmentä vuotta sitten mittaamaan depressiivisyyden syvyyttä (Beck, Ward, Mendelson, Mock & Erbaugh 1961). Alkuperäinen mittari oli 21-osiainen, mutta siitä on myöhemmin kehitetty lyhyempi 13-osiainen kysely (Beck & Beck 1972). Tästä lyhyemmästä versiosta on edelleen kehitetty Raitasaloon 13-osiainen Mielialakysely (RBDI) työkaluksi nuorten depressio-oireiden seulontaan esimerkiksi yleislääkärin praktiikassa.

Kouluterveyskyselyssä on vuodesta 1998 alkaen ollut mukana vain 12 RBDI-kysymystä masentuneisuuden mittaamiseen, koska itsensä vahingoittamista koskeva kysymys on luonteensa vuoksi Opetushallituksen määräyksestä jouduttu poistamaan. Vastemuuttuja on summamuuttuja näistä 12 RBDI-kysymyksestä, joiden vastaukset on pisteytetty, ja summamuuttujan maksimipistemäärä on 36. Indikaattorin muodostaminen esitetään tarkemmin liitteessä B.

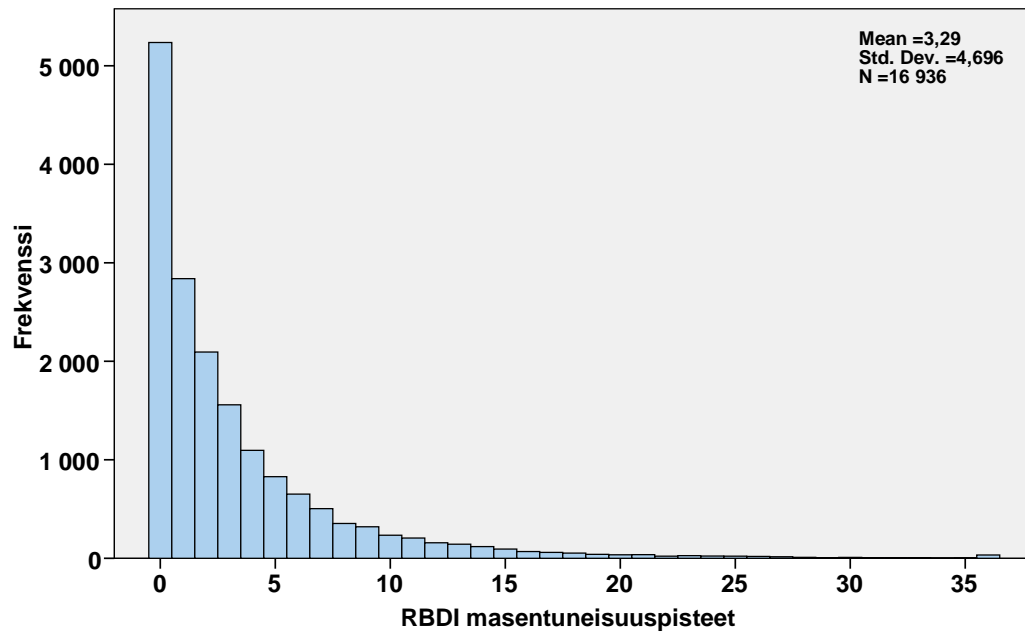
Kysymykset on tehty siten, että jokaisen kysymyksen ensimmäinen vaihtoehto kuvaa positiivista tunnetilaa tai ajatusta. Toinen vaihtoehto on neutraalimpi, mutta ei vielä kuvaa masennustilaan liittyviä mielialan piirteitä. Kolmannesta vaihtoehdosta alkaen väittämät kuvaavat depressio-oireita, joiden voimakkuus ja syvyys lisääntyvät kohti vaihtoehtoa viisi. Pisteiden perusteella ei voi tehdä diagnoosia masennustilasta, mutta niitä voidaan käyttää masentuneisuuden viitteelliseen arviointiin. Yhteenlasketut pisteet voidaan jakaa kuvaamaan masennusoireiden syvyyttä siten, että pisteet 0–4 viittaavat ei-depressiiviseen, 5–7 lievästi depressiiviseen, 8–15 keskivaikeasti depressiiviseen sekä pisteet 16–39 vakavasti depressiiviseen mielialaan. (Karlsson & Marttunen 2007.) Huomattavaa on, että edellä mainittu pisteluokittelu perustuu 13 kysymykseen.

Aineiston RBDI-pisteiden keskiarvo on 3.3, mediaani 2.0, keskihajonta 4.7 ja muuttuja saa arvoja välillä 0–36. RBDI-muuttujan havaintoja on yhteensä 16936 ja puuttuvia havaintoja 142. Nämä puuttuvan RBDI-havainnon sisältävät tilastoyksiköt on poistettu aineistosta. Kuten kuvioista 2.1 nähdään, kokonaisjakauma ei ole normaalin vaan huipukas ja oikealle vino. Kuviossa 2.2 puolestaan nähdään edellä mainitunlaisen masennusoireiden syvyyden mukaan luokitellun RBDI-pisteiden suhteelliset osuudet. Kaikista 16936 oppilaasta 76 %:lla ei ole depressiivisyyteen viittaavia oireita, 12 %:lla on lievästi, 10 %:lla keskivaikeasti ja 3 %:lla vakavasti depressiiviseen mielialaan viittaavia oireita.

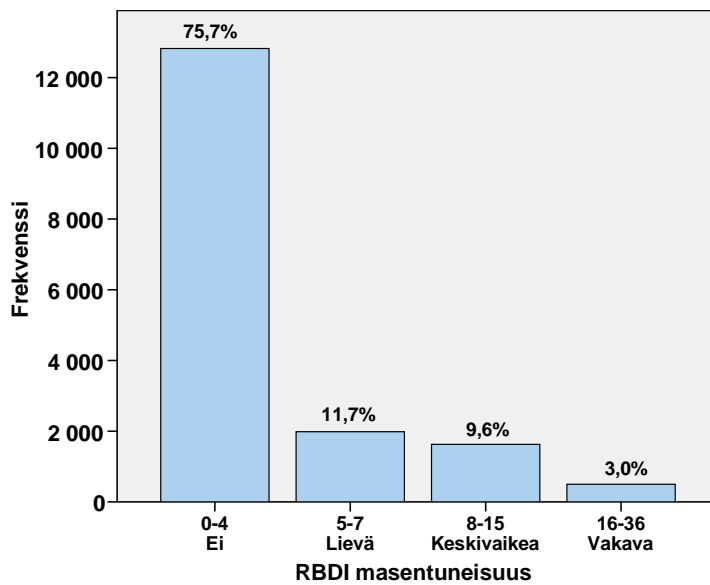
2.3.2 Ryhmittelevät muuttujat

Tutkimusongelmaa tarkastellaan kahden ryhmän vertailussa. Selittäjinä ja ryhmittelevinä muuttujina ovat sukupuoli, liikuntaharrastukset, vanhempien työttömyys, seksuaalikokemukset, suhde vanhempiin sekä suhde opettajiin. Kunkin ryhmän ominaisuus tulkitaan Kouluterveyskyselyn vastausten perusteella kyseessä olevan summamuuttujan pistemäärän mukaan. Ominaisuus on dikotomisoitu eli se saa arvon 0 tai 1, ja siis kullakin tilastoyksiköllä on jokin ominaisuus tai sitten ei. Lisäksi sukupuoliryhmässä tytöt saavat arvon 0 ja pojat arvon 1.

Muuttujien luokittelussa on osittain käytetty Stakesin määrittämiä indikaattoreita ja osa indikaattorimuuttujista on määriteltävä vastaavalla tavalla tätä tutkielmaa varten. Tarkemmin sanottuna indikaattorimuuttujat ovat sukupuoli, liikuntaa harvemmin kuin kerran viikossa, vähintään yksi vanhempi työttömänä vuoden aikana, on seksuaalikokemuksia, keskustelua vanhempien kanssa ei juuri koskaan sekä huono suhde ja kommunikointi oppilaan ja opettajien



Kuvio 2.1. RBDI-masentuneisuuspisteiden jakauma ($N = 16936$).



Kuvio 2.2. RBDI-pisteiden suhteelliset osuudet luokiteltuna masennusoireiden syvyyden mukaan ($N = 16936$).

välillä. Muuttujien muodostaminen esitetään siis liitteessä B.

Ryhmien sekä puuttuvien havaintojen suhteelliset osuudet nähdään taulukossa 2.2. Koska aineistossa on paljon havaintoyksikköjä ja vähän puuttuvia havaintoja ja puuttuminen näyttäisi olevan satunnaista, niin puuttuvien tietojen poistamisen ei katsota vaikuttavan tulosten tarkkuuteen tai vääristävän analyysien tuloksia. Puuttuvien havaintojen lukumäärät ryhmittäin ovat: sukupuoli 0, liikuntaharrastukset 169, vanhempien työttömyys 324, seksuaalikokemukset 696, suhde vanhempiin 265 ja suhde opettajiin 46. Puuttuvia tietoja sisältävät havaintoyksiköt on poistettu ryhmäkohtaisesti kyseessä olevan ryhmän analyyseistä. Kunkin ryhmän analyyseissä käytettävä otoskoko (N) sekä validit prosenttiosuudet nähdään myös taulukossa 2.2. Tyttöjen osuus on 50.5 % ja poikien osuus 49.5 %. Muissa ryhmittelymuuttujissa kyseessä oleva ominaisuus on 8–20 prosentilla vastaajista.

Taulukko 2.2. Ryhmittelymuuttujien (indikaattorimuuttujien) suhteelliset osuudet kaikille havainnoille ja puuttuvalle tiedolle, validit prosenttiosuudet sekä otoskoot (N).

Ryhmittelevä muuttuja	%osuus			Yht.	N
	Epätosi	Tosi	Puuttuva		
Sukupuoli (poika)	50.5	49.5	0.0	100.0	16936
	50.5	49.5		100.0	16936
Liikuntaa harvemmin kuin kerran viikossa	91.0	8.0	1.0	100.0	16936
	91.9	8.1		100.0	16767
Vähintään yksi vanhempi työttömänä vuoden aikana	78.9	19.1	1.9	100.0	16936
	80.5	19.5		100.0	16612
On seksuaalikokemuksia	76.6	19.3	4.1	100.0	16936
	79.8	20.2		100.0	16240
Keskustelua vanhempien kanssa ei juuri koskaan	87.6	10.8	1.6	100.0	16936
	89.0	11.0		100.0	16671
Huono suhde ja kommunikointi oppilaan ja opettajien välillä	80.2	19.5	0.3	100.0	16936
	80.4	19.6		100.0	16890

3 Klassiset menetelmät kahden riippumattoman otoksen vertailussa

Tässä tutkielmassa keskitytään tilastollisiin testeihin, joiden avulla voidaan selvittää kahden populaation välisiä eroja. Näitä eroja voidaan tutkia tarkastelemalla populaatioiden sijaintia ja usein se tapahtuu keskiarvojen avulla. Jos tutkittava muuttuja on vähintään välimatka-asteikollinen, voidaan populaatioiden sijaintia kuvata niiden keskiarvoilla. Koska keskiarvon käytöllä on kuitenkin rajoituksensa eikä sitä voida aina käyttää, tarvitaan myös muita testejä jakaumien vertailuun.

Testin valinnassa on siis otettava huomioon testin vaatimat edellytykset. Ranta, Rita ja Kouki (2005) korostavat, että koska parametriin testeihin usein liittyy voimakkaita oletuksia populaation jakauman muodolle, ei niiden käyttö ole aina mielekästä. Jos nämä oletukset eivät toteudu ja parametrisia testejä silti käytetään, tulokset ovat epäluotettavia. Toisin kuin parametrisilla menetelmillä yleensä, epäparametrisilla tai jakaumasta vapailla (*distribution free*) menetelmillä ei ole vaatimuksia populaation jakaumalle. (Tässä yhteydessä ei tehdä eroa epäparametristen ja jakaumasta vapaiden menetelmien välillä.) Epäparametrinen testi on yleensä kuitenkin parametrista tehottomampi tilanteissa, joissa normaalisuusoletukset ovat voimassa. Jos parametrinen testin oletukset eivät toteudu, niin silloin taas on perusteltua käyttää epäparametristä menetelmää, koska se on robustimpi kuin parametrinen versionsa.

Tässä luvussa esitellään klassiset kahden riippumattoman otoksen testit: parametrinen Studentin t-testi ja epäparametrinen Wilcoxonin järjestyslukutesti. Lisäksi tarkastellaan tasatulosten vaikutusta Wilcoxonin testissä.

3.1 Studentin t-testi

Alaluku perustuu pääpiirteissään Ranta et al. (2005, s. 187–192) teokseen, josta esimerkiksi voi tarkemmin lukea Studentin t-testistä, mutta tiedot löytyvät useimmista tilastotieteen perusteoksista. Studentin t-testi on yleisin ja yksi tehokkaimmista menetelmistä kahden riippumattoman otoksen keskiarvojen vertailuun. Testi edellyttää välimatka-asteikon tasoisen mittauksen lisäksi populaation jakauman normaalisuuden. Etenkin pienten otosten kohdalla jakauman normaalisuus on olennaista. Nollahypoteesi voidaan muotoilla monella eri tavalla, mutta tavallisimmin tarkastellaan keskiarvojen tai jakaumien yhtäsuuruutta. Kun tarkastelun kohteena on kahden keskiarvon vertailu (ryhmät A ja

B : esimerkiksi tytöt ja pojat tai käsittely (treatment) ja verrokki (control)), voidaan sama asia ilmaista myös yhden parametrin, keskiarvojen erotuksen $\mu_B - \mu_A$ (tai $\mu_A - \mu_B$), avulla.

Olko kaksi riippumatonta satunnaisotosta populaatioista A ja B siten, että ryhmän A satunnaisotos normaalijakaumasta $N(\mu_A, \sigma_A^2)$ on

$$X_1, X_2, \dots, X_{N_A}$$

ja ryhmän B satunnaisotos normaalijakaumasta $N(\mu_B, \sigma_B^2)$

$$X_{N_A+1}, X_{N_A+2}, \dots, X_{N_A+N_B},$$

missä μ_A ja μ_B ovat populaatioiden odotusarvoja, σ_A^2 ja σ_B^2 variansseja sekä N_A ja N_B otoskokoja. Nollahypoteesina t -testissä testataan, ovatko ryhmien A ja B keskiarvot yhtäsuuret ($H_0 : \mu_A = \mu_B$) ja keskiarvojen erotus -parametrin avulla kirjoitettuna siis

$$H_0 : \mu_B - \mu_A = 0.$$

Populaatioiden varianssit σ_A^2 ja σ_B^2 ovat tavallisesti tuntemattomia, jolloin ne on estimoitava otoksista. Jos kuitenkin oletetaan, että varianssit ovat yhtä suuria, saadaan yhteisen varianssin estimaattoriksi

$$s^2 = \frac{(N_A - 1)s_A^2 + (N_B - 1)s_B^2}{N_A + N_B - 2},$$

missä s_A^2 ja s_B^2 ovat otosvariensseja. Testisuure

$$(3.1) \quad t = \frac{\bar{X}_B - \bar{X}_A}{s \sqrt{\frac{1}{N_B} + \frac{1}{N_A}}}$$

noudattaa Studentin t -jakaumaa vapausastein $df = N_A + N_B - 2$, kun H_0 on tosi ja populaatiot ovat normaalijakautuneita eli $t \sim t(N_A + N_B - 2)$.

Jos ei voida olettaa populaatioiden varianssien yhtäsuuruutta, varianssit joudutaan estimoimaan kummallekin populaatiolle erikseen. Luonnollinen testisuure on tällöin

$$(3.2) \quad t = \frac{\bar{X}_B - \bar{X}_A}{\sqrt{\frac{s_B^2}{N_B} + \frac{s_A^2}{N_A}}},$$

joka suurilla otoskoilla on likimain $N(0, 1)$ -jakautunut, kun H_0 on tosi. Jos nyt varianssit ovat yhtäsuuret, niin $s_A^2 \approx s_B^2$ ja testisuureen (3.2) arvo on likimain sama kuin (3.1). Jos otoskoot ovat pieniä ja estimoidut varianssit poikkeavat paljon toisistaan, on perusteltua käyttää sopivaa epäparametrista testiä.

3.2 Wilcoxonin järjestysluketesti

Kahden riippumattoman otoksen epäparametrinen vaihtoehto, kun testataan tulevatko kahden ryhmän havainnot samasta jakaumasta, on Wilcoxonin järjestysluketesti (*Wilcoxon rank-sum test*). Testiä käytetään t-testin asemesta erityisesti silloin, kun jatkuvan muuttujan jakauma ei ole normaalin tai kun testattava muuttuja on järjestysasteikollinen. Testi on algebrallisesti sama (ekvivalentti) kuin Mannin–Whitneyn U-testi eli se antaa saman p-arvon. Testistä käytetäänkin joskus myös nimitystä Mannin–Whitneyn–Wilcoxonin testi. Tässä työssä testistä käytetään myös lyhyempää nimitystä Wilcoxonin testi, mutta testiä ei pidä sekoittaa Wilcoxonin parittaisten otosten testiin.

Wilcoxonin testillä ei siis ole jakaumaoletuksia, mutta otosten on oltava riippumattomia ja havaintojen vähintään järjestysasteikollisia. Wilcoxonin testi on niin sanottu järjestysluketesti, joka ei perustu varsinaisesti alkuperäiseen dataan vaan havaintojen järjestyslukuihin, jolloin populaation jakaumaoletuksia ei tarvita. Testi onkin yksi tehokkaimmista ja tunnetuimmista jakaumasta riippumattomista testeistä. Tämän kahden otoksen testin ovat alunperin kehittäneet Wilcoxon (1945) sekä Mann ja Whitney (1947).

Olkoot kaksi riippumatonta satunnaisotosta populaatioista A ja B siten, että ryhmän A satunnaisotos on

$$X_1, X_2, \dots, X_{N_A}$$

ja ryhmän B satunnaisotos

$$X_{N_A+1}, X_{N_A+2}, \dots, X_{N_A+N_B},$$

missä N_A on ryhmän A ja N_B ryhmän B otoskoko sekä $N_A + N_B = N$ on yhdistetyn havaintoaineiston otoskoko. Yksinkertaisuuden vuoksi, yleisyyttä menettämättä valitaan perusjoukosta B (esim. käsittely, treatment) poimittu otos jatkossa tehtävien vertailujen lähtökohdaksi. Tällöin yhdistetyn havaintoaineiston havainnot X_i , $i = 1, \dots, N_A, N_A + 1, \dots, N_A + N_B \equiv N$, asetetaan suuruusjärjestykseen ja korvataan järjestyslukuillaan R_i . Ryhmän A havaintojen järjestysluvut ovat siis

$$R_1, R_2, \dots, R_{N_A}$$

ja ryhmän B havaintojen järjestysluvut

$$R_{N_A+1}, R_{N_A+2}, \dots, R_N.$$

Kaikkien järjestyslukujen summa on $\sum_{i=1}^N R_i = 1 + 2 + \dots + N = \frac{1}{2}N(N + 1)$.

Testataan nollahypoteesia, jonka mukaan ryhmillä ei ole eroa ja siis

H_0 : Kaikki muuttujat X_i ovat samoin jakautuneita.

Nollahypoteesi voidaan muotoilla myös esimerkiksi seuraavasti:

H_0 : Käsittelyllä (treatment) ei ole vaikutusta.

Wilcoxonin testin testisuure

$$(3.3) \quad W_B = \sum_{i=N_A+1}^N R_i$$

on vertailuotoksen B järjestyslukujen summa, missä $N = N_A + N_B$. Jos vertailuotokseksi valitaan populaatiosta A poimittu otos, niin vastaava testisuure $W_A = \sum_{i=1}^{N_A} R_i$ antaa saman merkitsevyytason kuin W_B . (3.3):n minimiarvo on nyt selvästi $1 + 2 + \dots + N_B = \frac{1}{2}N_B(N_B + 1)$. Nollahypoteesi hylätään (ja uskotaan, että käsittelyllä on vaikutusta), kun

1. W_B on liian suuri ("suuri on huono"); $W_B \geq c$ ja täyttää ehdon $P_{H_0}(W_B \geq c) = \alpha$ tai
2. W_B on liian pieni ("pieni on huono"); $W_B \leq c$ ja täyttää ehdon $P_{H_0}(W_B \leq c) = \alpha$,

missä vakio c on kriittinen arvo ja α on valittu merkitsevyytaso (Lehmann 2006).

H_0 :n vallitessa todennäköisyys jokaiselle havaitulle joukolle (r_{N_A+1}, \dots, r_N) on

$$P_{H_0}(R_{N_A+1} = r_{N_A+1}, \dots, R_N = r_N) = \frac{N_A!}{N!}.$$

Olkoon $\#(w; N_B, N_A)$ kaikkien niiden järjestyslukupöjökköjen määrä järjestyslukuista $(1, \dots, N \equiv N_A + N_B)$, joiden summa ryhmälle B on w . Silloin

$$P_{H_0}(W_B = w) = \frac{\#(w; N_B, N_A)}{N!/N_A!}.$$

Wilcoxonin testin kriittiset arvot ja merkitsevyydet pienille otoskoille (kun $N_A, N_B \leq 10$) saadaan esimerkiksi niitä varten laadituista taulukoista. Lehmannin (2006, Table B, s. 414–416) taulukko perustuu Mann–Whitney-testisuureeseen

$$(3.4) \quad W_{XY} = W_B - \frac{1}{2}N_B(N_B + 1),$$

joka on ekvivalentti Wilcoxonin testisuureen (3.3) kanssa. Taulukossa on käytetty (3.4):n jakaumaa, koska nimittäin $1 + 2 + \dots + N_B = \frac{1}{2}N_B(N_B + 1)$ on (3.3):n minimiarvo, ja sitä käyttämällä taulukko saadaan kompaktiin muotoon. Taulukosta saadaan todennäköisyys $P_{H_0}(W_{XY} \leq a)$, missä a on testisuureeseen W_{XY} (3.4) perustuva kriittinen arvo. (Lehmann 2006.)

Otoskokojen kasvaessa testisuureen havaittu, likimääräinen merkitsevyytaso voidaan keskeisen raja-arvolauseen (*central limit theorem, CLT*) perusteella usein laskea normaalijakaumaan perustuen. Nyt nollahypoteesin vallitessa

$$(3.5) \quad E(W_B) = \frac{1}{2}N_B(N + 1)$$

ja

$$(3.6) \quad \text{Var}(W_B) = \frac{1}{12} N_A N_B (N + 1).$$

Normitettu Wilcoxonin testisuure on nyt

$$W = \frac{W_B - E(W_B)}{\sqrt{\text{Var}(W_B)}},$$

missä $W \sim N(0, 1)$ likimain, kun H_0 on tosi. (Lehmann 2006.)

3.3 Tasatulosten käsittely Wilcoxonin testissä

Tähän asti on oletettu, että havainnoissa ei esiinny tasatuloksia. Jos jakauma on jatkuva ja havainnot mitataan riittävän suurella tarkkuudella, niin tasatuloksia ei periaatteessa esiinny. Käytännössä mittauksissa tai muissa numeerisissa havainnoissa usein kuitenkin sattuu niin, että yhtä suuria havainnot on kaksi tai useampia. Lisäksi Wilcoxonin testi voidaan tehdä jo kategoriselle järjestysasteikolliselle muuttujalle, joten tasatulokset ovat sitenkin mahdollisia ja tuovat ongelmia testisuureen merkitsevyyden laskemiseen. Otosten sisäiset tasatulokset eivät vaikuta testisuureen arvoon, mutta otosten väliset tasatulokset vaikuttavat. Tasatulokset pienentävät yhdistetyn havaintoaineiston varianssia, mutta niillä ei ole vaikutusta normitettuun keskiarvoon. (Lehmann 2006; Ranta et al. 2005.)

Jos tasatuloksia ei huomioida testisuureen laskemisessa, saadaan liian suuria p-arvoja. Näin siksi, koska huomioidaessa tasatulokset estimoitu varianssi pienenee, ja sen seurauksena normitetun testisuureen arvo hieman kasvaa. Samalla kasvaa myös otosten erojen tilastollinen merkitsevyys. Tasatulosten aiheuttamat sidokset huomioidaan testisuureen laskemisessa siten, että käytetään sidoksissa olevien havaintojen *järjestyslukujen keskiarvoja*.

Kun tasatuloksia ei ole $R_i = \sum_{j=1}^N I(X_j \leq X_i)$. Tasatulosten sattuessa käytetään kaavaa

$$R_i^* = \sum_{j=1}^N \left[I(X_j < X_i) + \frac{1}{2} I(X_j = X_i) \right] + \frac{1}{2}.$$

Wilcoxonin testin testisuure keskiarvoisilla järjestysluvuilla (*Wilcoxon midrank-sum test*) on

$$(3.7) \quad W_B^* = \sum_{i=N_A+1}^N R_i^*,$$

missä $N = N_A + N_B$ ja R_i^* on siis sidoksissa olevien järjestyslukujen keskiarvo.

Koska (3.7):n jakauma riippuu otoskokojen N_A ja N_B lisäksi tasatulosten määrästä, jakauman taulukointi on monimutkaista. Siksi tarve approksimaation käytölle on ilmeinen. Tasatulokset eivät vaikuta odotusarvoon, vaan se on

vastaava kuin (3.5), siis

$$E(W_B^*) = \frac{1}{2}N_B(N+1).$$

Testisuureen W_B^* (3.7) tasatulokset huomioivan varianssin kaavassa on mukana korjaustermi, ja se on muotoa

$$(3.8) \quad \text{Var}(W_B^*) = \frac{1}{12}N_A N_B (N+1) - \frac{N_A N_B \sum_{i=1}^e (d_i^3 - d_i)}{12N(N-1)},$$

missä oletetaan, että kaikissa N :ssä ($= N_A + N_B$) havainnossa on e erisuuruista arvoa ja d_1 on pienimpien arvojen lukumäärä, d_2 toiseksi pienimpien, \dots , d_e suurimpien arvojen lukumäärä. Odotusarvon ja varianssin todistukset ovat luettavissa Lehmannin (2006, s. 332–333) teoksesta. Varianssin kaavan (3.8) ensimmäinen termi

$$\frac{1}{12}N_A N_B (N+1)$$

on yhtä kuin (3.6), ja toinen termi

$$\frac{N_A N_B \sum_{i=1}^e (d_i^3 - d_i)}{12N(N-1)}$$

korjaa tasatulosten aiheuttaman virheen. Siten erityisesti, kun tasatuloksia ei esiinny, $e = N$ ja jokainen $d_i = 1$, $i = 1, \dots, N$, jolloin korjaustermi on nolla ja $\text{Var}(W_B^*)$ (3.8) = $\text{Var}(W_B)$ (3.6). (Lehmann 2006.)

Standardoitu Wilcoxonin testisuure on nyt

$$(3.9) \quad W^* = \frac{W_B^* - E(W_B^*)}{\sqrt{\text{Var}(W_B^*)}}$$

ja $W^* \sim N(0, 1)$ likimain, kun H_0 on tosi. (3.7):lle on siis nollahypoteesin vallitessa voimassa asymptoottinen normaalisuus eli (3.9) noudattaa likimain standardoitua normaalijakaumaa Φ , kun $N_A, N_B \rightarrow \infty$ ja lisäksi kun

$$\frac{d_i}{N} \rightarrow \lambda_i, \quad i = 1, \dots, e, \quad \text{kun } N \rightarrow \infty, \quad 0 < \lambda_i < 1.$$

Indeksi i viittaa siis tässä e :hen mahdolliseen satunnaismuuttujan X_i arvoon. (Lehmann 2006.)

4 Klusteroitunut aineisto

Tässä luvussa taustoitetaan hieman klusteroitumista ja esitellään tyypilliset kahden ryhmän klusteroituneen aineiston populaatorakennemallit. Lisäksi lukijalle esitellään uudet menetelmät klusteroituneen aineiston kahden otoksen sijainnin vertailuun. Uudet, klusteroidut versiot esitetään Studentin t-testille ja Wilcoxonin järjestyslukutestille. Uusia testejä nimitetään myös lyhyemmin klusteroiduiksi testeiksi. Teoria perustuu yhteen klusteritasoon, ja lähtökohtana on klusteroitunut aineisto ja klusterin sisäinen riippuvuus.

4.1 Klusterin sisäinen riippuvuus

Hierarkkinen rakenne ja klusterien olemassaolo on tavallista monissa käytännön aineistoissa, ja havainnot klusterin sisällä ovat usein korreloituneita. Tässä työssä klusterien oletetaan olevan keskenään riippumattomia. Jos klusterien sisäinen riippuvuus jätetään huomioimatta, voidaan analyyseissä saada liian suuria p-arvoja ja siten tehdä vääriä johtopäätöksiä. Tämä tarkoittaisi sitä, että nollahypoteesi hyväksytään, vaikka vaihtoehtoinen hypoteesi olisi totta (tyypin II virhe). Toisaalta klusterin sisäisen korrelaation ignorointi voi johtaa jopa epävalidin tilastollisen testin käyttämiseen p-arvojen ollessa virheellisen pieniä.

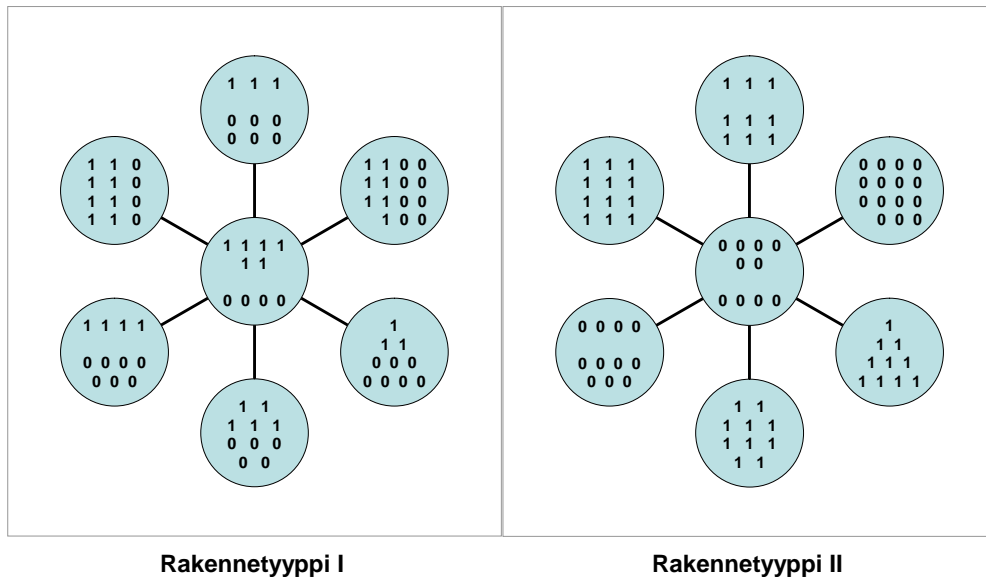
Koska klusteroituneessa aineistossa ajatellaan, että klusterin sisäinen samankaltaisuus johtuu jostain yhdistävästä tekijästä tai syistä, niin näitä syitä voisi ajatella ikään kuin sekoittavina tekijöinä. Sekoittavia tekijöitä pyritään kontrolloimaan jakamalla koko tutkimusaineisto pienempiin luontaisiin osiin, klustereihin, jotka ovat sekoittavien tekijöiden suhteen homogeenisempia kuin koko tutkimusaineisto.

Jos otetaan esimerkiksi tutkielmassa käytetyn empiirisen aineiston vaste-
muuttujana oleva koululaisten masentuneisuuspistemäärä, klusteroituminen tarkoittaisi sitä, että saman klusterin, eli tässä tapauksessa koulun, oppilaiden pistemäärät ovat riippuvia. Voi olla, että joissakin kouluissa on keskimääräistä korkeampia ja joissakin kouluissa taas keskimääräistä matalampia pistemääriä. Vastaavasti järjestyslukutestissä korkeita pistemääriä saavat koulut saavat luonnollisesti isoja järjestyslukuja ja matalia pistemääriä saavat koulut taas pieniä järjestyslukuja. Tutkitaan näiden vaikutuksia työn edetessä.

4.2 Klusterirakenteet

Lähtökohtaisesti tutkielman lähestymistavassa ajatellaan, että klusteroituneen aineiston populaatorakenteessa voidaan erotella kaksi rakennetyyppiä tai -mal-
lia. Lisäksi muistetaan, että tässä työssä käsitellään kahden otoksen, kahden ryhmän tapausta, mutta periaatteessa ryhmiä voisi olla useampiakin kuin kaksi.

Rakennetyypin *I* ryhmäjako ajatellaan siten, että jokaisessa klusterissa (esim. koulu) on molempien ryhmien (esim. tytöt ja pojat) jäseniä. Tyypis-
sä *II* puolestaan on kyse klustereista, joissa on vain toisen ryhmän jäseniä (esim. koulussa 1 on vain tyttöjä, koulussa 2 on vain poikia, jne.). On myös mahdollista, että rakennetyypit sekoittuvat eli osassa klustereista on molempien ryhmien jäseniä (tyttöjä sekä poikia) ja osassa taas pelkästään toisen ryhmän jäseniä (tyttöjä tai poikia). Tällöin voidaan ajatella, että kyseessä on tyyppin *I* erikoistapaus. Edellä mainittuja rakennetyyppejä havainnollistetaan kuviossa 4.1.



Kuvio 4.1. Klusteroituneen aineiston kahden ryhmän (0/1) populaatorakennetyypit. Tyyppi *I*: klustereissa on molempien ryhmien jäseniä (0 ja 1). Tyyppi *II*: klustereissa on vain toisen ryhmän jäseniä (0 tai 1).

4.3 Klusteroitu t-testi

Koostukoon havaintoaineisto

$$\mathbf{X} = (\mathbf{X}'_1, \mathbf{X}'_2, \dots, \mathbf{X}'_n)' = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \vdots \\ \mathbf{X}_n \end{pmatrix}$$

n klusterista \mathbf{X}_i ja klusterit satunnaismuuttujien X_{ij} havainnoista, siis

$$\mathbf{X}_i = (X_{i1}, X_{i2}, \dots, X_{im_i})', \quad i = 1, \dots, n; \quad j = 1, \dots, m_i,$$

missä n on klusterien lukumäärä ja m_i on i . klusterin havaintojen määrä. Olkoon otoskoko $N = \sum_{i=1}^n m_i$. Siis

$$\mathbf{X}_1 = \begin{pmatrix} X_{11} \\ X_{12} \\ \vdots \\ X_{1m_1} \end{pmatrix}, \quad \mathbf{X}_2 = \begin{pmatrix} X_{21} \\ X_{22} \\ \vdots \\ X_{2m_2} \end{pmatrix}, \dots, \quad \mathbf{X}_n = \begin{pmatrix} X_{n1} \\ X_{n2} \\ \vdots \\ X_{nm_n} \end{pmatrix}$$

ja $\mathbf{X} \in \mathbb{R}^{N \times 1}$. Klusterien määrän on oltava suurempi kuin yksi eli $1 < n \leq N$. Jos klusterointia ei ole, niin silloin klusterien koko on 1 eli jokainen havainto on ikään kuin omassa klusterissaan ja siis $n = N$.

Havainnot tulevat ryhmistä A ja B ja ryhmiä vastaavat otoskoot ovat N_A ja N_B sekä $N = N_A + N_B$. Olkoon G_{ij} ryhmän jäsenyys i . klusterin j . havainnolle ja

$$G_{ij} = \begin{cases} 0, & \text{jos } X_{ij} \text{ kuuluu ryhmään } A; \\ 1, & \text{jos } X_{ij} \text{ kuuluu ryhmään } B. \end{cases}$$

Nyt ryhmän B jäsenten lukumäärä i . klusterissa on $\sum_{j=1}^{m_i} G_{ij}$ ja ryhmän A jäsenten lukumäärä i . klusterissa $\sum_{j=1}^{m_i} (1 - G_{ij})$.

Merkitään

$$\Delta = \mu_B - \mu_A$$

sekä

$$\delta_{ij} = \begin{cases} -\frac{1}{N_A}(1 - G_{ij}), & \text{jos } X_{ij} \text{ kuuluu ryhmään } A; \\ \frac{1}{N_B}G_{ij}, & \text{jos } X_{ij} \text{ kuuluu ryhmään } B. \end{cases}$$

Siis

$$\delta_{ij} = \begin{cases} -\frac{1}{N_A}, & \text{jos } X_{ij} \text{ kuuluu ryhmään } A; \\ \frac{1}{N_B}, & \text{jos } X_{ij} \text{ kuuluu ryhmään } B. \end{cases}$$

Silloin ryhmäkeskiarvojen erotus on

$$(4.1a) \quad \bar{X}_B - \bar{X}_A = \sum_{i=1}^n \sum_{j=1}^{m_i} \delta_{ij} X_{ij}$$

$$(4.1b) \quad = \sum_{i=1}^n \sum_{j=1}^{m_i} \delta_{ij} (X_{ij} - \bar{X}).$$

Yhtälö (4.1a) voidaan todeta oikeaksi kirjoittamalla

$$\begin{aligned}\sum_{i=1}^n \sum_{j=1}^{m_i} \delta_{ij} X_{ij} &= \sum_{i=1}^n \sum_{j=1}^{m_i} \frac{1}{N_B} G_{ij} X_{ij} - \sum_{i=1}^n \sum_{j=1}^{m_i} \frac{1}{N_A} (1 - G_{ij}) X_{ij} \\ &= \frac{1}{N_B} \sum_{i=1}^n \sum_{j=1}^{m_i} X_{ij} G_{ij} - \frac{1}{N_A} \sum_{i=1}^n \sum_{j=1}^{m_i} X_{ij} (1 - G_{ij}) \\ &= \bar{X}_B - \bar{X}_A.\end{aligned}$$

Yhtälö (4.1b) puolestaan seuraa siitä, että

$$\begin{aligned}\sum_{i=1}^n \sum_{j=1}^{m_i} \delta_{ij} (X_{ij} - \bar{X}) &= \sum_{i=1}^n \sum_{j=1}^{m_i} \delta_{ij} X_{ij} - \bar{X} \underbrace{\sum_{i=1}^n \sum_{j=1}^{m_i} \delta_{ij}}_{=0} \\ &= \bar{X}_B - \bar{X}_A.\end{aligned}$$

Oletetaan, että

- klusterit $\mathbf{X}_1, \dots, \mathbf{X}_n$ ovat riippumattomia
- muuttujat $X_{ij} - \Delta G_{ij}$ ovat samoin jakautuneita
- kaksiulotteiset muuttujat $(X_{ij} - \Delta G_{ij}, X_{ij'} - \Delta G_{ij'})$ ovat samoin jakautuneita.

Testataan nollahypoteesia, jonka mukaan ryhmillä ei ole eroa ja siis

H_0 : Kaikki muuttujat X_{ij} ovat samoin jakautuneita.

Nollahypoteesi voidaan myös kirjoittaa: $H_0: \Delta = 0$. Oletusten mukaan muuttujien varianssit ja kovarianssit ovat yhtä suuria. Merkitään

$$\text{Var}(X_{ij}) = \sigma^2 \quad \text{ja} \quad \text{Cor}(X_{ij}, X_{ij'}) = \rho, \quad \forall i, j; j \neq j'.$$

Nollahypoteesin vallitessa

$$E(\bar{X}_B - \bar{X}_A) = 0$$

ja

$$(4.2) \quad \text{Var}(\bar{X}_B - \bar{X}_A) = \sum_{i=1}^n \sum_{j=1}^{m_i} \sum_{j'=1}^{m_i} \delta_{ij} \delta_{ij'} \text{Cov}(X_{ij}, X_{ij'}),$$

kun $1 < n \leq N$. Yhtälö (4.2) voidaan kirjoittaa myös:

$$\begin{aligned}
\text{Var}(\bar{X}_B - \bar{X}_A) &= \sum_{i=1}^n \left(\sigma^2 \sum_{j=j'} \delta_{ij}^2 + \rho \sigma^2 \sum_{j \neq j'} \delta_{ij} \delta_{ij'} \right) \\
&= \sigma^2 \left[\sum_{i=1}^n \sum_{j=1}^{m_i} \delta_{ij}^2 + \rho \sum_{i=1}^n \left(\sum_{j=1}^{m_i} \sum_{j'=1}^{m_i} \delta_{ij} \delta_{ij'} - \sum_{j=1}^{m_i} \delta_{ij}^2 \right) \right] \\
&= \sigma^2 \left[(1 - \rho) \underbrace{\sum_{i=1}^n \sum_{j=1}^{m_i} \delta_{ij}^2}_{=\frac{1}{N_A} + \frac{1}{N_B}} + \rho \sum_{i=1}^n \left(\sum_{j=1}^{m_i} \delta_{ij} \right)^2 \right] \\
&= \sigma^2 \left[(1 - \rho) \left(\frac{1}{N_A} + \frac{1}{N_B} \right) + \rho \sum_{i=1}^n \left(\sum_{j=1}^{m_i} \delta_{ij} \right)^2 \right].
\end{aligned}$$

Nähdään, että kun

$$\begin{aligned}
\rho \rightarrow 0 &\Rightarrow \text{Var}(\bar{X}_B - \bar{X}_A) \rightarrow \sigma^2 \left(\frac{1}{N_A} + \frac{1}{N_B} \right) \\
\rho \rightarrow 1 &\Rightarrow \text{Var}(\bar{X}_B - \bar{X}_A) \rightarrow \sigma^2 \sum_{i=1}^n \left(\sum_{j=1}^{m_i} \delta_{ij} \right)^2.
\end{aligned}$$

Edelleen, kun H_0 on tosi ja $1 < n \leq N$, ryhmäkeskiarvojen erotuksen varianssin luonnollinen, tarkentuva estimaattori on

$$(4.3a) \quad \widehat{\text{Var}}(\bar{X}_B - \bar{X}_A) = \sum_{i=1}^n \sum_{j=1}^{m_i} \sum_{j'=1}^{m_i} \delta_{ij} \delta_{ij'} (X_{ij} - \bar{X})(X_{ij'} - \bar{X})$$

$$(4.3b) \quad = \sum_{i=1}^n \left[\sum_{j=1}^{m_i} \delta_{ij} (X_{ij} - \bar{X}) \right]^2.$$

Kahden otoksen t-testisuure klusteroituneelle aineistolle, jossa klusterikoot $m_i \leq M$; $M \in \mathbb{Z}_+$, ovat rajoitettuja, on nyt

$$T^c = \frac{\bar{X}_B - \bar{X}_A}{\sqrt{\widehat{\text{Var}}(\bar{X}_B - \bar{X}_A)}},$$

joka suurilla klusterimäärillä, asympotoottisen normaalisuuden perusteella on likimain $N(0, 1)$ -jakautunut, kun H_0 on tosi. Tulos seuraa keskeisestä raja-arvolauseesta ja siitä, että erilaisia tapauksia m_i ; $\delta_{i1}, \dots, \delta_{im_i}$ on äärellinen määrä (Larocque, Haataja, Nevalainen & Oja 2008).

4.4 Klusteroitu Wilcoxonin testi

Kahden otoksen Wilcoxonin järjestyslukutestin klusteroitu versio on teorialtaan samankaltainen kuin edellä esitetty t-testin klusteroitu versio. Klusteroitu Wilcoxonin testi on epäparametrinen vastine klusteroidulle t-testille. Satunnaisotokset tulevat populaatioista A ja B , eikä otosten jakaumista tehdä

oletuksia. Wilcoxonin testiä varten yhdistetyn havaintoaineiston havainnot X_{ij} korvataan vastaavilla järjestyslukuillaan R_{ij} , $i = 1, \dots, n$; $j = 1, \dots, m_i$, missä n on klusterien määrä ja m_i on i . klusterin havaintojen määrä. Tasatulosten huomioimiseksi käytetään keskiarvoistettuja järjestyslukuja.

Havainnot tulevat ryhmistä A ja B , ja ryhmiä vastaavat otoskoot ovat N_A ja N_B sekä $N = N_A + N_B$. Lisäksi oletetaan, että $1 < n \leq N$. Olkoon G_{ij} ryhmän jäsenyys i . klusterin j . havainnolle ja

$$G_{ij} = \begin{cases} 0, & \text{jos } R_{ij} \text{ kuuluu ryhmään } A; \\ 1, & \text{jos } R_{ij} \text{ kuuluu ryhmään } B. \end{cases}$$

Ryhmän B jäsenten lukumäärä i . klusterissa on $\sum_{j=1}^{m_i} G_{ij}$ ja ryhmän A jäsenten lukumäärä i . klusterissa $\sum_{j=1}^{m_i} (1 - G_{ij})$.

Merkitään

$$\delta_{ij} = \begin{cases} -\frac{1}{N_A}(1 - G_{ij}), & \text{jos } R_{ij} \text{ kuuluu ryhmään } A; \\ \frac{1}{N_B}G_{ij}, & \text{jos } R_{ij} \text{ kuuluu ryhmään } B. \end{cases}$$

Siis

$$\delta_{ij} = \begin{cases} -\frac{1}{N_A}, & \text{jos } R_{ij} \text{ kuuluu ryhmään } A; \\ \frac{1}{N_B}, & \text{jos } R_{ij} \text{ kuuluu ryhmään } B. \end{cases}$$

Nyt ryhmäkeskiarvojen erotus on, vastaavasti kuin (4.1),

$$\bar{R}_B - \bar{R}_A = \sum_{i=1}^n \sum_{j=1}^{m_i} \delta_{ij} R_{ij} = \sum_{i=1}^n \sum_{j=1}^{m_i} \delta_{ij} (R_{ij} - \bar{R}).$$

Oletetaan, kuten t-testin tilanteessa, että

- klusterit $\mathbf{X}_1, \dots, \mathbf{X}_n$ ovat riippumattomia
- muuttujat $X_{ij} - \Delta G_{ij}$ ovat samoin jakautuneita
- kaksiuolotteiset muuttujat $(X_{ij} - \Delta G_{ij}, X_{ij'} - \Delta G_{ij'})$ ovat samoin jakautuneita.

Testataan nollahypoteesia, jonka mukaan ryhmillä ei ole eroa, ja siis

$$H_0: \text{Kaikki muuttujat } X_{ij} \text{ ovat samoin jakautuneita,}$$

jolloin myös R_{ij} :t ovat samoin jakautuneita. Nollahypoteesin vallitessa

$$E(\bar{R}_B - \bar{R}_A) = 0$$

ja varianssi, vastaavasti kuin (4.2),

$$\text{Var}(\bar{R}_B - \bar{R}_A) = \sum_{i=1}^n \sum_{j=1}^{m_i} \sum_{j'=1}^{m_i} \delta_{ij} \delta_{ij'} \text{Cov}(R_{ij}, R_{ij'}),$$

kun $1 < n \leq N$. Edelleen, vaikka eri klustereihin liittyvät järjestysluvut ovat riippuvia, ryhmäkeskiarvojen erotuksen varianssin luonnollinen, tarkentuva estimaattori on, vastaavasti kuin (4.3),

$$\begin{aligned}\widehat{\text{Var}}(\bar{R}_B - \bar{R}_A) &= \sum_{i=1}^n \sum_{j=1}^{m_i} \sum_{j'=1}^{m_i} \delta_{ij} \delta_{ij'} (R_{ij} - \bar{R})(R_{ij'} - \bar{R}) \\ &= \sum_{i=1}^n \left[\sum_{j=1}^{m_i} \delta_{ij} (R_{ij} - \bar{R}) \right]^2.\end{aligned}$$

Kahden otoksen Wilcoxonin testisuure klusteroituneelle aineistolle on

$$W^c = \frac{\bar{R}_B - \bar{R}_A}{\sqrt{\widehat{\text{Var}}(\bar{R}_B - \bar{R}_A)}},$$

joka on likimain $N(0, 1)$ -jakautunut, kun H_0 on tosi. Asymptoottinen normalisuus voidaan osoittaa kuten Wilcoxonin testin normalisuus riippumattomien otosten tapauksessa, kun erilaisia tapauksia m_i ; $\delta_{i1}, \dots, \delta_{im_i}$ on äärellinen määrä (Larocque et al. 2008).

5 Menetelmien vertailu klusteroituneessa aineistossa

Menetelmien ja testien vertailua varten tehdään simulointikokeita. Tässä luvussa ja tutkielman tuloksellisesti merkittävimmässä osassa esitetään simulointien avulla, kuinka kahden ryhmän vertailuun tarkoitetut perinteiset ja uudet, tässä työssä esitetyt menetelmät käyttäytyvät analysoitaessa klusteroitunutta aineistoa.

5.1 Tutkimusmenetelmät

Kokeellisia analyysejä varten generoidaan otoksia sopivista jakaumista ja lisäksi huomioidaan klusterin sisäinen korrelaatio. Generoiduilla otoksilla tehdään kahden riippumattoman otoksen klassinen Studentin t-testi ja Wilcoxonin järjestyslukutesti sekä klusteroitu t-testi ja Wilcoxonin testi. Menetelmiä vertaillaan testien voimakkuuksien avulla eli generoiduista otoksista määritellään testien todennäköisyydet hylätä epätosi nollahypoteesi. Kaikkien testien merkitsevyytasoksi valitaan simuloinneissa 0.05 (5 %). Puhuttaessa lyhyemmin *klassisista testeistä* tarkoitetaan niillä Studentin t-testiä ja Wilcoxonin testiä ja *klusteroituilla testeillä* tarkoitetaan uusia, tässä tutkielmassa esitettyjä t-testin ja Wilcoxonin testin klusteroituja versioita.

Testien vertailevissa analyyseissä käytetään tilastollista R-ohjelmistoa (versio 2.5.0, <http://www.r-project.org/>). Klusteroituneelle aineistolle edellä esitettyjen testisuureiden ja merkitsevyyksien laskemiseen on koodattu R-funktiot, koska markkinoilla olevissa ohjelmistoissa ei valmista testiä luonnollisestikaan ole. R-funktiot t-testin ja Wilcoxonin testin klusteroituille versioille esitetään liitteessä A.

Tutkimusaineistona käytettävästä Kouluterveyskysely-aineistosta tehdään analyysejä vastaavilla testeillä. Koska aineisto on havaintomäärältään hyvin suuri, siitä demonstroidaan myös pienempiä osaotoksia, jotta eri testien välisiä eroja olisi mahdollista saada esiin. Tähän empiiriseen osaan palataan tarkemmin luvussa 6.

5.2 Klusteroitujen mallien muodostaminen

Klusteroituneen datan simulointikokeissa käytetään alaluvussa 4.2 kuvailtuja rakennetyyppejä *I* ja *II*. Tyypin *I* klustereissa on molempien ryhmien jäseniä ja

tyypin II klustereissa vain toisen ryhmän jäseniä. Molempiin rakennetyyppeihin otetaan kolme jakaumaa: normaalijakauma sekä t-jakauma vapausasteilla (*df*) 3 ja 10. Yhteensä siis tarkastellaan kuutta eri mallia.

Merkitään

n = klusterien määrä,

m_i = havaintojen määrä i . klusterissa,

$$N = \sum_{i=1}^n m_i = \text{otoskoko},$$

m_{i_0} = ryhmän 0 jäsenten lukumäärä i . klusterissa,

m_{i_1} = ryhmän 1 jäsenten lukumäärä i . klusterissa,

G_{ij} = ryhmän jäsenyys (0 tai 1) i . klusterin j . havainnolle siten, että

$$m_{i_1} = \sum_{j=1}^{m_i} G_{ij} \quad \text{ja} \quad m_{i_0} = \sum_{j=1}^{m_i} (1 - G_{ij}), \quad i = 1, \dots, n,$$

$\Delta = \bar{X}_1 - \bar{X}_0 =$ ryhmäkeskiarvojen erotus yksilötasolla.

Havainnot simulointikokeisiin generoidaan mallin

$$(5.1) \quad X_{ij} = \Delta G_{ij} + \frac{\tau_i + \epsilon_{ij}}{\sqrt{\chi_k^2/k}}, \quad i = 1, \dots, n; \quad j = 1, \dots, m_i,$$

mukaisesti, missä τ_i on klusteriefektitermi ja ϵ_{ij} virhetermi. Oletetaan, että χ_k^2 noudattaa χ^2 -jakaumaa vapausastein k , $\tau_i \sim N(0, \tau^2)$ ja $\epsilon_{ij} \sim N(0, \sigma^2)$. Lisäksi oletetaan, että kaikki muuttujat τ_i , ϵ_{ij} ovat keskenään riippumattomia. Siis $X_{ij} - \Delta G_{ij} \sim t(k)$.

Nyt

$$(5.2) \quad \text{Cor}(X_{ij}, X_{ij'}) = \frac{\tau^2}{\tau^2 + \sigma^2}, \quad j \neq j'.$$

Jos $k \rightarrow \infty$, niin $(X_{ij} - \Delta G_{ij})$:n jakauma lähestyy normaalijakaumaa ja (5.1) voidaan kirjoittaa

$$(5.3) \quad X_{ij} \rightarrow \Delta G_{ij} + \tau_i + \epsilon_{ij}, \quad i = 1, \dots, n; \quad j = 1, \dots, m_i,$$

ja $X_{ij} - \Delta G_{ij} \sim N(\mu, \tau^2 + \sigma^2)$ likimain. Myös nyt

$$(5.4) \quad \text{Cor}(X_{ij}, X_{ij'}) = \frac{\tau^2}{\tau^2 + \sigma^2}, \quad j \neq j'.$$

Merkitään klusterin sisäistä korrelaatiota

$$(5.5) \quad \text{Cor}(X_{ij}, X_{ij'}) = \rho, \quad j \neq j'.$$

Yleisyyttä menettämättä kiinnitetään (laskennan helpottamiseksi) simuloinneissa $\tau^2 + \sigma^2 = 1$, jolloin (5.2), (5.4) ja (5.5) mukaan $\text{Cor}(X_{ij}, X_{ij'}) = \tau^2 = \rho$

ja edelleen $\sigma^2 = 1 - \rho$. Jos nyt $k \rightarrow \infty$ ja $\mu = 0$, niin (5.3) mukaan $X_{ij} - \Delta G_{ij}$ noudattaa likimain standardoitua normaalijakaumaa.

Simulointeja havainnollistavissa kuvioissa kukin piste kuvaa kyseessä olevan testin voimakkuutta. Yhden pisteen määrittämiseen on käytetty 1000 toistoa, joista lasketaan nollassa nollahypoteesin hylkäämiseen johtaneet testitulokset kriittisen pisteen mukaan ($p < 0.05$). Kaiken kaikkiaan simulointikokeita on tehty 1000:n toistoilla: $[3 \text{ jakaumaa} \times 8 \rho \times 11 \Delta \times 4 \text{ testiä}]_{\text{malli I}} + [(3 \text{ jakaumaa} \times 8 \rho \times 16 \Delta \times 4 \text{ testiä}) + (3 \text{ jakaumaa} \times 1 \rho \times 31 \Delta \times 4 \text{ testiä})]_{\text{malli II}} = 1056 + 1536 + 372 = 2964$. Simulointikokeiden tulokset taulukoituina esitetään liitteessä C. Tarkemmat tiedot ja simulointeihin liittyvät koodit annetaan pyydettäessä.

Havainnot generoidaan siis seuraavista jakaumista:

1. Mallit $\mathbf{M(I, norm)}$ ja $\mathbf{M(II, norm)}$ (rakennetyypit I ja II)

$$X_{ij} - \Delta G_{ij} = \tau_i + \epsilon_{ij} \sim N(0, 1)$$

2. Mallit $\mathbf{M(I, t(3))}$ ja $\mathbf{M(II, t(3))}$ (rakennetyypit I ja II)

$$X_{ij} - \Delta G_{ij} = \frac{\tau_i + \epsilon_{ij}}{\sqrt{\chi_3^2/3}} \sim t(3)$$

3. Mallit $\mathbf{M(I, t(10))}$ ja $\mathbf{M(II, t(10))}$ (rakennetyypit I ja II)

$$X_{ij} - \Delta G_{ij} = \frac{\tau_i + \epsilon_{ij}}{\sqrt{\chi_{10}^2/10}} \sim t(10).$$

5.3 Simuloinnit: rakennetyyppi I

Oletetaan ensin rakennetyypin I mukainen malli, joka valitaan niin, että klusterien koko on yhtä suuri. Poissuljetaan ryhmäkokojen välisen suhteen mahdollinen vaikutus, joten tarkastellaan tasapainoista ryhmäjakoja eli valitaan jokaiseen klusteriin yhtä monta ryhmien 0 ja 1 jäsentä. Valitaan otokseen klusterien määräksi 50 ja jokaiseen klusteriin 20 havaintoa, joista molempiin ryhmiin kuuluu 10 havaintoa. Otoksen koko on siis 1000. Esimerkiksi jokaisessa 50 koulussa on 20 oppilasta, joista puolet on tyttöjä ja puolet poikia. Tarkastellaan tapauksia, joissa odotusarvo on nollassa nollahypoteesin mukaisesti nolla ja klusterin sisäinen korrelaatio ρ saa arvoja väliltä $[0, 0.7]$ sekä ryhmäkeskiarvojen erotus Δ saa arvoja väliltä $[0, 0.2]$. Nyt

$$N = 1000, \quad n = 50, \quad m_1 = \dots = m_{50} = 20, \quad m_{i_0} = m_{i_1} = 10; \quad i = 1, \dots, 50$$

sekä

$$\begin{aligned} \rho &= 0.00, 0.10, 0.20, \dots, 0.70 \quad \text{ja} \\ \Delta &= 0.00, 0.02, 0.04, \dots, 0.20. \end{aligned}$$

5.3.1 Havainnot normaalijakaumasta

Generoidaan ensin klusteroituneita otoksia normaalijakaumasta $N(0, 1)$ mallin $\mathbf{M}(\mathbf{I}, \mathbf{norm})$ mukaisesti, jolloin $X_{ij} = \Delta G_{ij} + \tau_i + \epsilon_{ij}$. Kuviossa 5.1 havainnollistetaan tämän tasapainoisen normaalijakaumallisen mallin mukaisen simuloinnin tuloksia eri klusterin sisäisen korrelaation ρ (kuviossa *rho*) arvoilla. Kuviossahan pisteet kuvaavat testien voimakkuutta eri Δ :n (kuviossa *delta*) arvoilla. Kuvioista 5.1 nähdään, että kun klusterin sisäinen korrelaatio $\rho = 0$ eli periaatteessa aineisto ei ole klusteroitunut, niin kaikki testit antavat lähes saman tuloksen. Tällöin testin voimakkuus lisääntyy kaikilla neljällä testillä likimain yhtä paljon ryhmien eron Δ :n kasvaessa. Lisäksi kuvioista voidaan havaita, kuinka aineiston muuttuessa korreloituneeksi, ja nyt tarkoitetaan nimenomaan klusterin sisäistä korrelaatiota, klassisten testien ja klusteroitujen testien tulokset alkavat selkeästi erottua toisistaan. Etenkin silloin, kun korrelaatio on ≥ 0.40 , pienikin ero kahden ryhmän välillä on havaittavissa klusteroiduilla testeillä tehokkaammin kuin klassisilla testeillä.

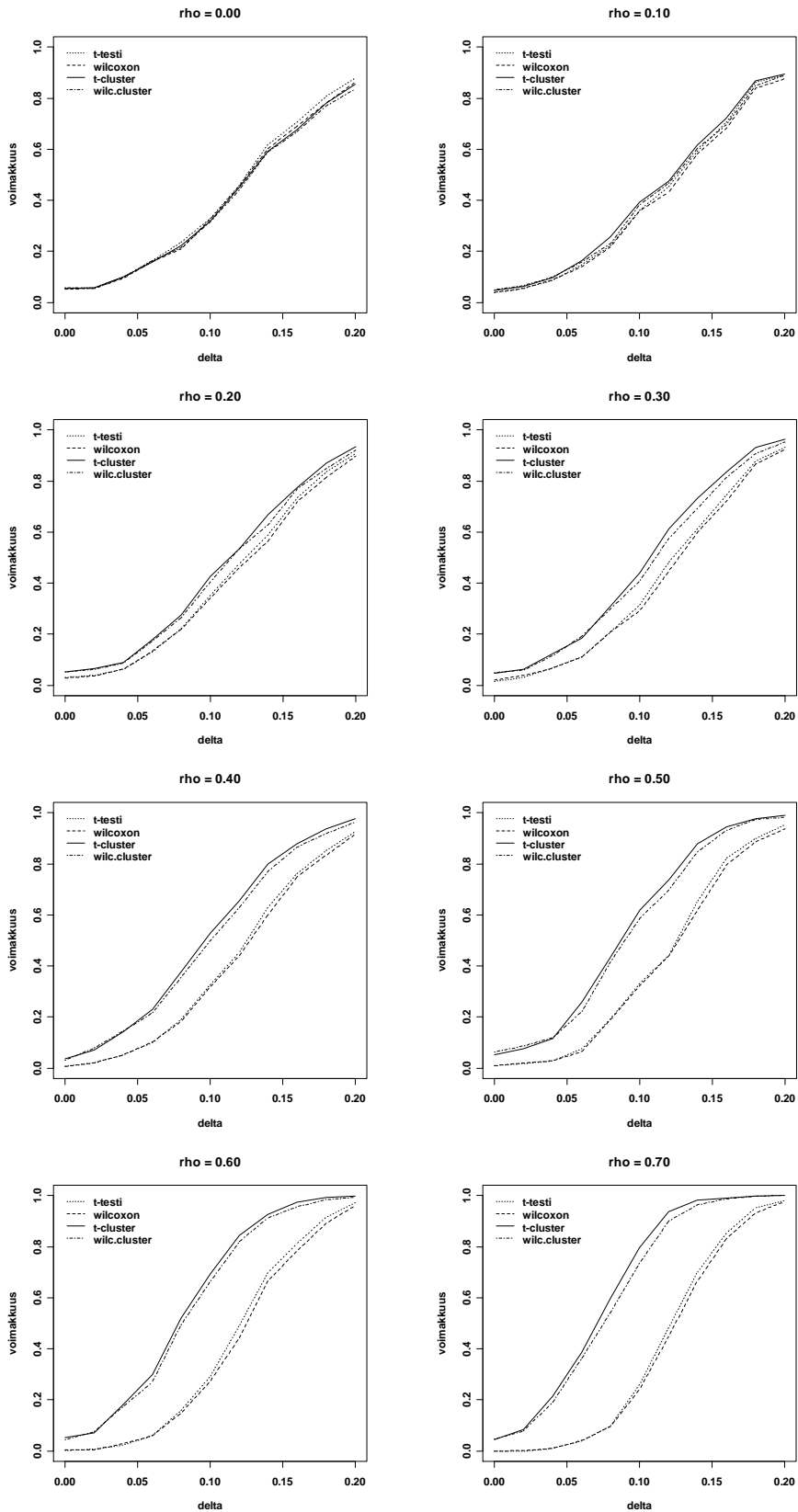
Molemmat klusteroidut testit havaitsevat siis poikkeamat nollahypoteesista huomattavasti tehokkaammin kuin klassiset testit. Esimerkiksi, kun $\rho = 0.50$ ja $\Delta = 0.10$, klassisen Studentin t-testin ja Wilcoxonin testin voimakkuudet ovat 0.330 ja 0.323, kun taas klusteroitu Wilcoxonin testi antaa voimakkuudeksi 0.587 ja klusteroitu t-testi peräti 0.619. Vastaavasti, kun $\rho = 0.70$ ja $\Delta = 0.10$, niin klassisen t-testin ja Wilcoxonin testin voimakkuudet jopa hieman laskevat (0.260 ja 0.241), mutta uusien klusteroitujen testien vastaavat arvot nousevat 0.735:een (klusteroitu Wilcoxonin testi) ja 0.797:ään (klusteroitu t-testi). Simulointikokeiden tulokset esitetään tarkemmin liitteessä C.1.

Jos tarkastellaan yleisesti testien voimakkuuksia tasapainoisessa normaalijakaumallisessa mallissa $\mathbf{M}(I, \mathbf{norm})$, niin klusterin sisäinen korrelaatio vaikuttaa siten, että klassiset testit ovat konservatiivisia ja epävalideja. Klassisten testien p-arvot ovat liian suuria, jolloin nollahypoteesi hylätään liian harvoin. Klusteroidut testit sitä vastoin ovat sitä voimakkaampia mitä suurempia sekä ρ että Δ on. Ne näyttävät havaitsevan poikkeamat nollahypoteesista varsin tehokkaasti. Jos puolestaan verrataan t-testiä ja Wilcoxonin testiä, sekä klassisia testejä keskenään että klusteroituja testejä keskenään, niin t-testit näyttävät olevan hieman tehokkaampia kuin vastaavat Wilcoxonin testit. Nythän oli kyseessä generoidut otokset normaalisesta jakaumasta, joten tulosten osoittama ”paremmuus” parametrusten t-testien hyväksi on luonnollista.

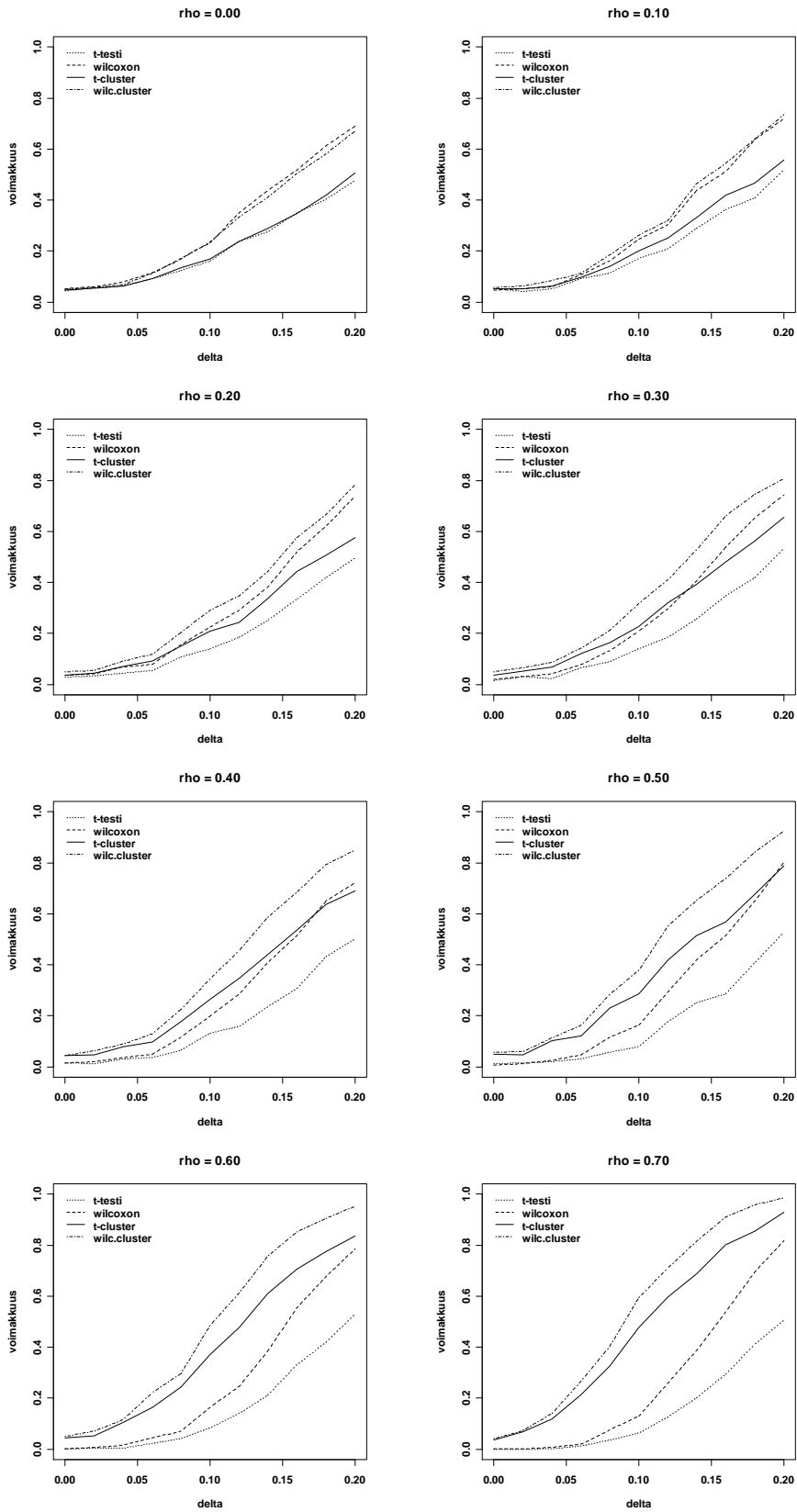
5.3.2 Havainnot t-jakaumasta

Generoidaan sitten havaintoja t-jakaumasta. Vapausasteiksi malliin $\mathbf{M}(\mathbf{I}, \mathbf{t}(3))$ valittiin 3, jolloin $X_{ij} = \Delta G_{ij} + (\tau_i + \epsilon_{ij}) / \sqrt{\chi_3^2/3}$. Malliin $\mathbf{M}(\mathbf{I}, \mathbf{t}(10))$ valittiin $df = 10$, jolloin $X_{ij} = \Delta G_{ij} + (\tau_i + \epsilon_{ij}) / \sqrt{\chi_{10}^2/10}$. Kuvioissa 5.2 ja 5.3 havainnollistetaan näiden tasapainoisten t-jakaumallisten mallien mukaisten simulointien tuloksia.

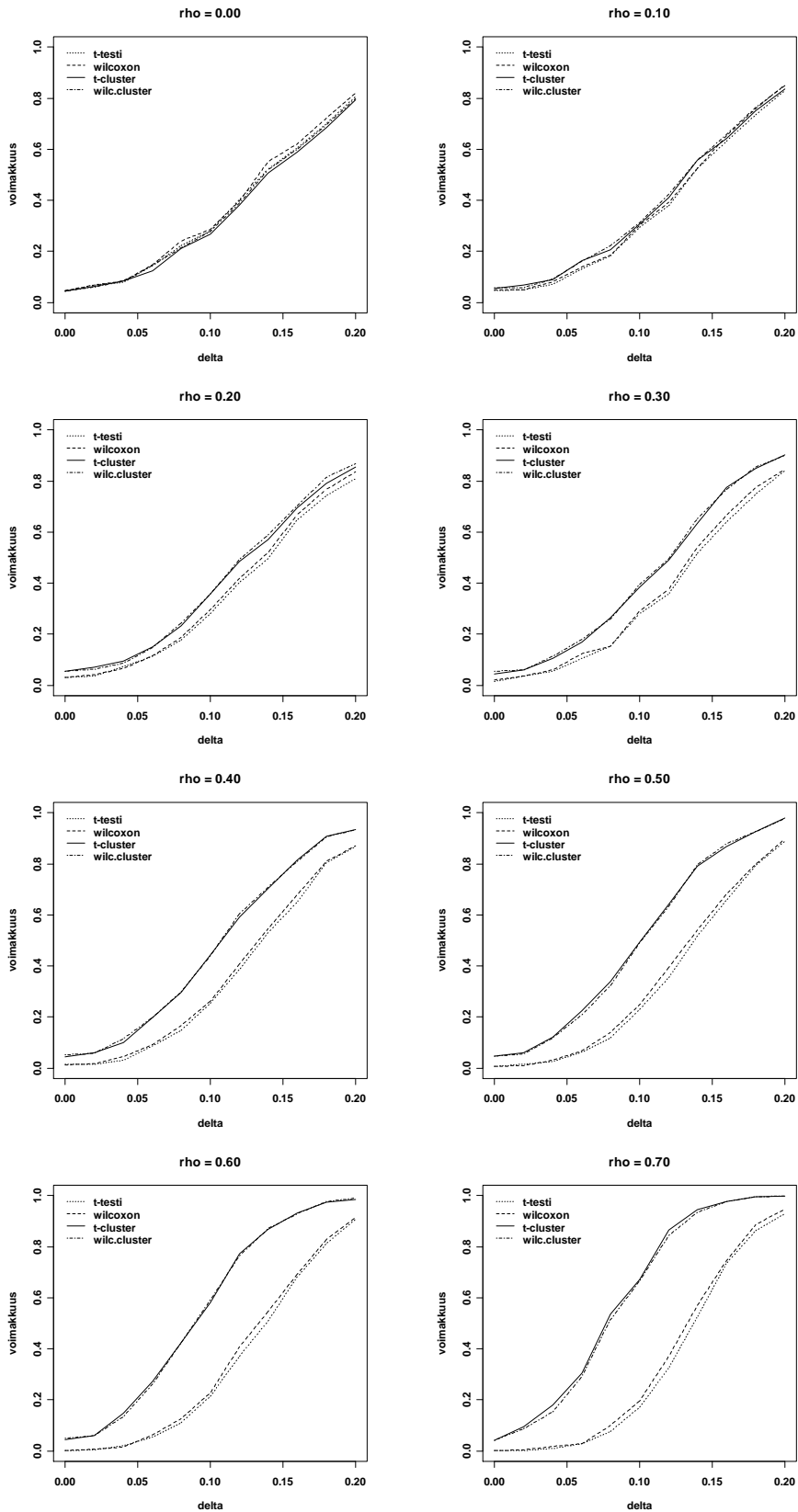
Tarkasteltaessa mallia $\mathbf{M}(I, \mathbf{t}(3))$ kuvioista 5.2 nähdään, että kun kluste-



Kuvio 5.1. Simulointikoe tasapainoiselle normaalijakaumalliselle mallille $M(I, \text{norm})$. Testien voimakkuuksien vertailu ρ :n ja Δ :n eri arvoilla, kun klustereita on 50 ja jokaisessa klusterissa 10 molempien ryhmien jäsentä ($N = 1000$).



Kuvio 5.2. Simulointikoe tasapainoiselle t-jakaumalliselle mallille $M(I, t(3))$. Testien voimakkuuksien vertailu ρ :n ja Δ :n eri arvoilla, kun klustereita on 50 ja jokaisessa klusterissa 10 molempien ryhmien jäsentä ($N = 1000$).



Kuvio 5.3. Simulointikoe tasapainoiselle t-jakaumalliselle mallille $M(I, t(10))$. Testien voimakkuuksien vertailu ρ :n ja Δ :n eri arvoilla, kun klustereita on 50 ja jokaisessa klusterissa 10 molempien ryhmien jäsentä ($N = 1000$).

rointia ei ole ($\rho = 0$), niin Δ :n kasvaessa epäparametriset testit näyttäisivät olevan jonkin verran tehokkaampia kuin parametriset testit. Klusterin sisäisen korrelaation kasvattaminen tuo testien erot jälleen esiin. Vielä pienillä ρ :n arvoilla 0.10 ja 0.20 epäparametristen testien voimakkuusarvot ovat suurempia kuin parametristen. Samalla uudet, klusteroidut testit ovat hieman tehokkaampia klassisiin vastaaviin testeihin verrattuna. Kun ρ :ta kasvatetaan yli 0.40:n, niin jälleen molemmat klusteroidut testit havaitsevat nollahypoteesin poikkeamat herkemmin kuin klassiset testit. Jos lisäksi verrataan epäparametrisia ja parametrisia testejä, niin sekä klusteroitujen että klassisten testien epäparametrinen testi on selkeästi tehokkaampi kuin parametrinen vastineensa. Erot tehokkuudessa korostuvat epäparametristen testien hyväksi, koska nyt on kyseessä ns. paksuhäntäinen (*heavy-tailed*) jakauma.

Jos tarkastellaan mallia $M(I, t(10))$, kuviosta 5.3 havaitaan, että parametristen ja epäparametristen testien välillä ei juurikaan ole eroa. Korrelaation kasvaessa, uudet testit ovat nytkin voimakkaampia kuin klassiset testit. Itse asiassa, jos tuloksia verrataan normaalijakaumallisen mallin $M(I, \text{norm})$ tuloksiin (kuvio 5.1), huomataan kuvioiden olevan melko yhteneviä. Erona voisi mainita, että epäparametriset testit ovat tässä t-jakauman mallissa odotetusti hieman, tosin vain minimaalisesti, tehokkaampia kuin parametriset testit, kun verrataan klassisia testejä keskenään ja klusteroituja testejä vastaavasti keskenään. Normaalijakaumatapauksessahan tulos oli päinvastainen.

Tarkasteltaessa jälleen yleisesti testien voimakkuuksia tasapainoisissa t-jakaumallisissa malleissa $M(I, t(3))$ ja $M(I, t(10))$ tulokset ovat samansuuntaisia kuin normaaliosessa mallissa $M(I, \text{norm})$. Aineiston ollessa klusterin sisäisesti korreloitunut klassiset testit hylkäävät nollahypoteesin liian harvoin eivätkä ole siis valideja. Klusteroitujen testien kohdalla ρ :n kasvattaminen puolestaan lisää testien tehokkuutta. Lisäksi paksuhäntäinen jakauma ($t(3)$) tuo testeihin eroja, jolloin Wilcoxonin testit ovat voimakkaampia kuin niitä vastaavat t-testit. Tarkemmat numeeriset tulokset simulointikokeista esitetään siis liitteessä C.1.

5.4 Simuloinnit: rakennetyyppi II

Oletetaan sitten alaluvun 4.2, rakennetyypin II mukainen malli, jossa klusterien koko on yhtä suuri, mutta puolessa klustereista on ainoastaan ryhmän 0 jäseniä ja puolessa ainoastaan ryhmän 1 jäseniä. Kutsutaan tätä mallia epätasapainoiseksi malliksi. Valitaan klusterien määräksi edelleen 50 ja jokaiseen klusteriin 20 havaintoa, jotka kaikki kuuluvat joko ryhmään 0 tai ryhmään 1. Otoskoko on siis myös nyt 1000. Esimerkiksi jokaisessa 50 koulussa on 20 oppilasta, ja kouluista 25 on tyttökouluja ja loput 25 poikakouluja. Tarkastellaan tapauksia, joissa odotusarvo on nollahypoteesin mukaisesti nolla ja klusterin sisäinen korrelaatio ρ saa arvoja väliltä $[0, 0.7]$ sekä ryhmäkeskiarvojen erotus Δ saa arvoja väliltä $[0, 0.3]$. Nyt

$$N = 1000, \quad n = 50, \quad m_1 = \dots = m_{50} = 20, \quad m_{1_1} = \dots = m_{25_1} = 0, \\ m_{26_1} = \dots = m_{50_1} = 20, \quad m_{1_0} = \dots = m_{25_0} = 20, \quad m_{26_0} = \dots = m_{50_0} = 0$$

sekä

$$\rho = 0.00, 0.02, 0.04, 0.06, 0.08, 0.10, 0.20, \dots, 0.70 \quad \text{ja} \\ \Delta = 0.00, 0.02, 0.04, \dots, 0.30.$$

5.4.1 Havainnot normaalijakaumasta

Generoidaan ensin havaintoja normaalijakaumasta $N(0, 1)$ mallin $\mathbf{M}(\mathbf{II}, \text{norm})$ mukaisesti. Aineisto on siis klusterin sisäisesti korreloitunut ja $X_{ij} = \Delta G_{ij} + \tau_i + \epsilon_{ij}$. Kuviossa 5.4 havainnollistetaan tämän epätasapainoisen mallin simuloinnin tuloksia eri ρ :n ja Δ :n arvoilla.

Kuvio 5.4 osoittaa samoin kuin aiemmin mallissa $\mathbf{M}(I, \text{norm})$ (kuvio 5.1), että kun klusterin sisäinen korrelaatio $\rho = 0$, kaikki neljä testiä antavat likimain saman tuloksen. Pienikin ρ :n kasvattaminen ($\rho \in [0.02, 0.06]$) saa aikaan sen, että nyt puolestaan klassiset testit ensin näyttäisivät havaitsevan nollahypoteesin poikkeamat tehokkaammin kuin klusteroidut testit. Kuitenkin samalla huomataan, että klassisten testien p-arvot ovat virheellisen pieniä. Havaintona voidaan todeta, että vaikka ryhmien välistä eroa ei ole ja aineistossa on klusterin sisäistä korrelaatiota ($\Delta = 0$ ja $\rho > 0$), niin klassiset testit hylkäävät nollahypoteesin liian usein. Esimerkiksi, kun $\Delta = 0$ ja $\rho = 0.10$, hylkäämisen todennäköisyys on klassisella Studentin t-testillä 0.243 ja Wilcoxonin testillä 0.236. Tai edelleen, kun $\Delta = 0$ ja $\rho = 0.70$, niin vastaavat hylkäämisen todennäköisyydet ovat peräti 0.610 ja 0.589 (liite C.2), kun ne valitulla riskitasolla ”saisivat” olla vain noin 0.05. Klassiset testit eivät siis ole lainkaan validea. Tässä mallissa klusterin sisäinen korrelaatio näyttäisi vahvasti aiheuttavan epäluotettavuutta klassisten testien tuloksiin.

Kuviosta 5.4 nähdään edelleen, että klusteroidut testit menettävät tehokkuuttaan ρ :n kasvaessa. Tällaisessa epätasapainoisen mallin tapauksessa, jossa yhdessä klusterissa on siis pelkästään toisen ryhmän edustajia, kuten esimerkiksi aiemmin mainitut poikakoulut ja tyttökoulut, uudet testit eivät pysty havaitsemaan nollahypoteesin poikkeamia yhtä tehokkaasti kuin tasapainoisessa mallissa (kuvio 5.1). Tarkastellaan vielä lähemmin tilannetta, kun $\rho \rightarrow 1$ ja Δ saa arvoja väliltä $[0, 1.5]$. Kuviosta 5.5 nähdään, että kun $\rho = 0.99$, testien voimakkuuden kasvu Δ :n kasvaessa on maltillisempaa kuin vastaavassa tasapainoisessa mallissa. Klusteroidut testit havaitsevat ryhmäkeskiarvojen välisen eron vasta suuremmilla Δ :n arvoilla. Parametrisella ja epäparametrisella testillä ei ole suurta eroa, mutta klusteroitu t-testi on aavistuksen verran tehokkaampi kuin vastaava Wilcoxonin testi.

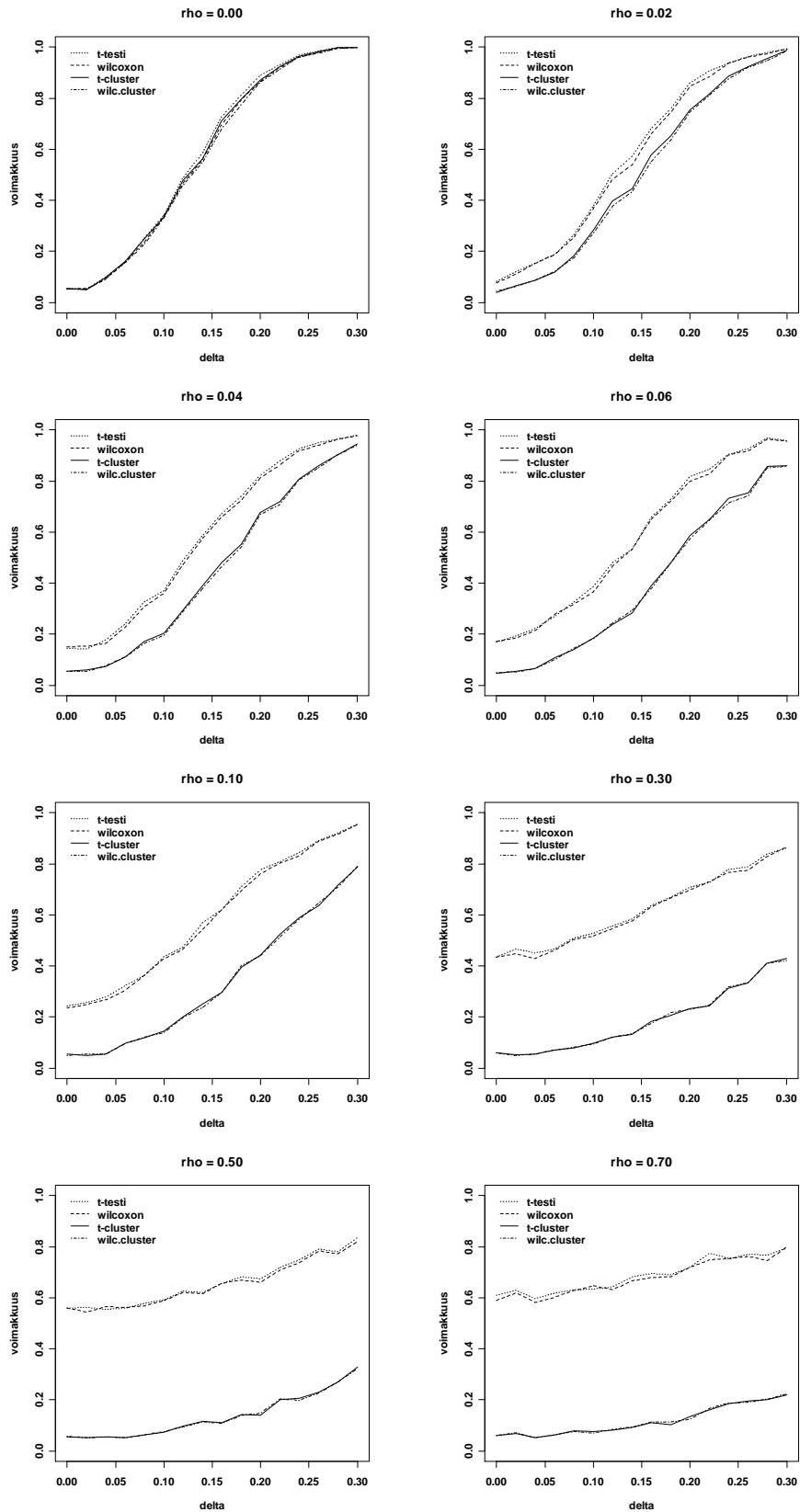
5.4.2 Havainnot t-jakaumasta

Generoidaan tähän malliin myös klusterin sisäisesti korreloituneita otoksia t-jakaumasta. Mallissa $\mathbf{M}(\mathbf{II}, \mathbf{t}(3))$ vapausasteiksi valittiin 3, jolloin $X_{ij} =$

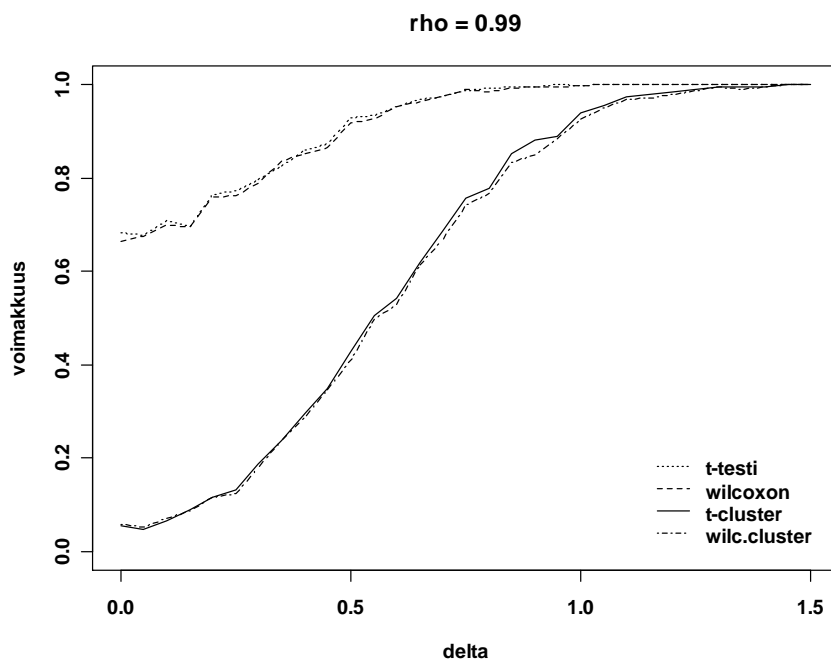
$\Delta G_{ij} + (\tau_i + \epsilon_{ij}) / \sqrt{\chi_3^2/3}$ ja mallissa $\mathbf{M}(\mathbf{II}, \mathbf{t}(\mathbf{10}))$ valittiin $df = 10$, jolloin $X_{ij} = \Delta G_{ij} + (\tau_i + \epsilon_{ij}) / \sqrt{\chi_{10}^2/10}$. Simuloinnin tuloksia näihin epätasapainoisiin malleihin havainnollistetaan kuvioissa 5.6 ja 5.7.

Kuviosta 5.6 nähdään mallin $\mathbf{M}(\mathbf{II}, \mathbf{t}(3))$ mukainen tilanne. Kun klusterin sisäistä korrelaatiota ei ole tai se on hyvin pieni ($\rho \in [0.00, 0.02]$), molemmat epäparametriset testit ovat tehokkaampia kuin parametriset testit. Edelleen ρ :n kasvattaminen saa aikaan saman tuloksen kuin normaalijakaumatilanteessa eli klassiset testit eivät ole valideja. Klassiset testit tekevät hylkäämisvirheen eli hylkäävät nollahypoteesin silloinkin, kun se on tosi. Lisäksi nähdään parametristen testien herkkyys jakauman muodolle (paksuhäntäinen jakauma), jolloin parametrinen testi on tehottomampi kuin vastaava epäparametrinen testi. Kuviosta 5.7 nähdään, että mallissa $\mathbf{M}(\mathbf{II}, \mathbf{t}(10))$ testit käyttäytyvät jokseenkin samankaltaisesti kuin mallissa $\mathbf{M}(\mathbf{II}, \mathbf{t}(3))$, mutta samalla aletaan lähestyä mallin $\mathbf{M}(\mathbf{II}, \text{norm})$ mukaista normaalijakaumatilannetta (kuvio 5.4).

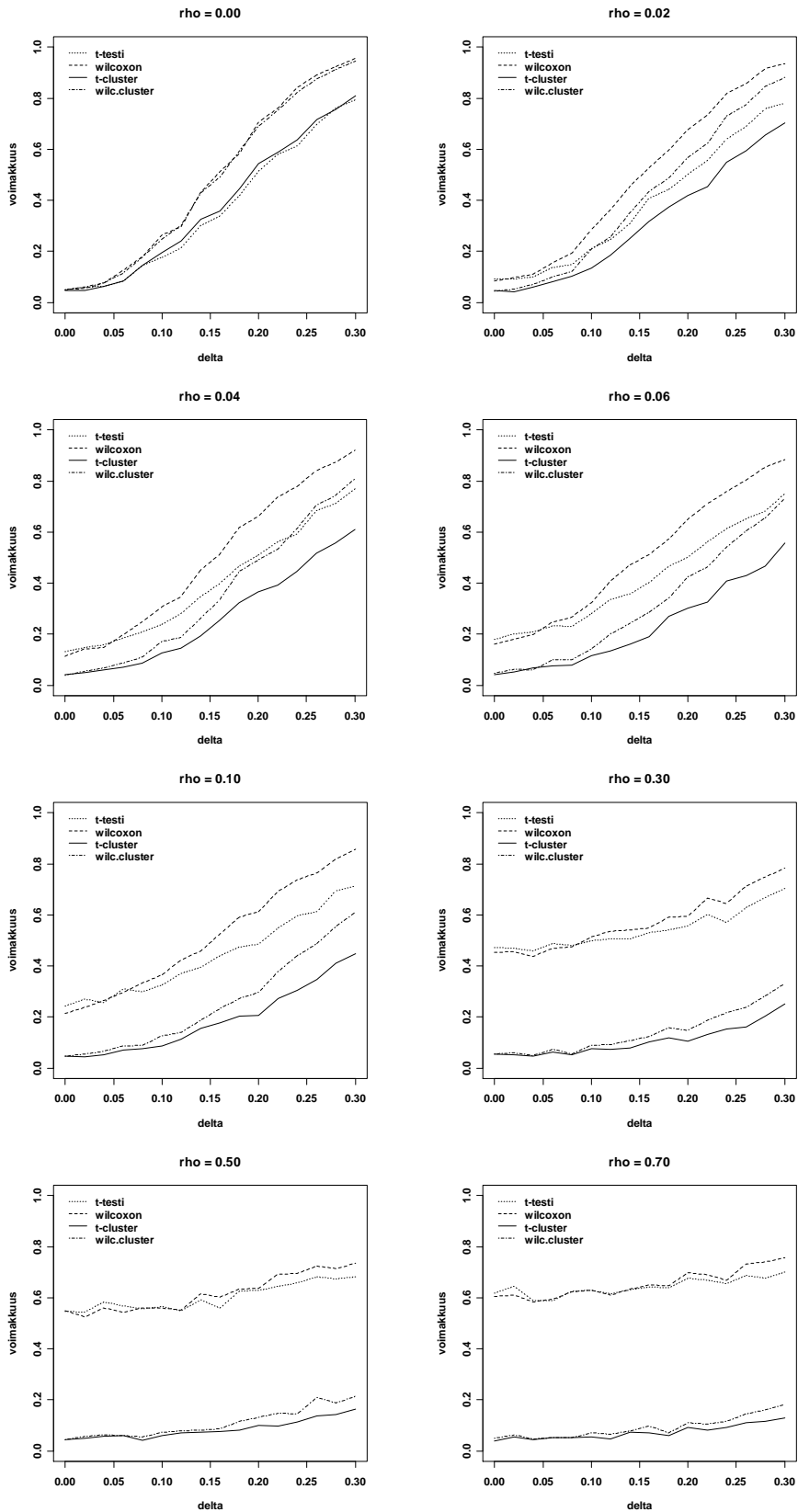
Myös epätasapainoisissa t-jakaumallisissa malleissa uudet klusteroidut testit menettävät tehokkuuttaan ρ :n kasvaessa, eivätkä pysty vielä Δ :n arvoilla $[0, 0.3]$ tehokkaasti havaitsemaan nollahypoteesin poikkeamia. Tarkastellaan siis vielä tilannetta, kun $\rho \rightarrow 1$ sekä Δ saa arvoja väliltä $[0, 1.5]$. Kuviosta 5.8 nähdään mallin $\mathbf{M}(\mathbf{II}, \mathbf{t}(3))$ simulointitulokset sekä kuviosta 5.9 mallin $\mathbf{M}(\mathbf{II}, \mathbf{t}(10))$ tulokset, kun $\rho = 0.99$. Tulokset ovat samansuuntaisia kuin epätasapainoisessa normaalijakaumatilanteessa (kuvio 5.5) eli klusteroitujen testien voimakkuudet kasvavat Δ :n arvon kasvaessa. Lisäksi $\mathbf{t}(3)$ -jakauman mallissa Wilcoxonin testin klusteroitu versio on jakauman muodon vuoksi odotetusti tehokkaampi kuin klusteroitu t-testi. Tarkemmat numeeriset tulokset mallien simulointikokeista esitetään liitteessä C.2.



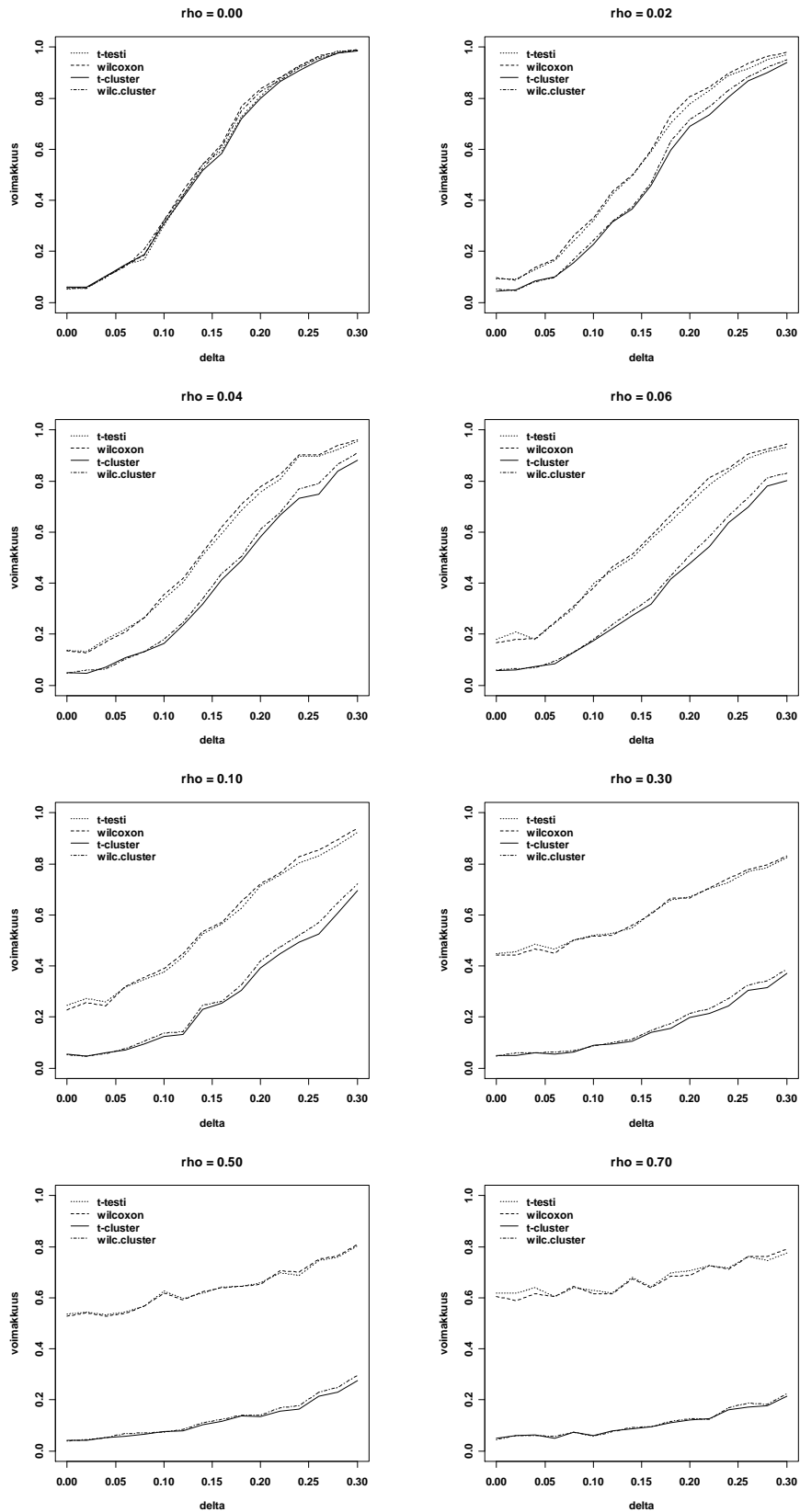
Kuvio 5.4. Simulointikoe epätasapainoiselle normaalijakaumalliselle mallille $M(II, \text{norm})$. Testien voimakkuuksien vertailu ρ :n ja Δ :n eri arvoilla, kun klustereita on 50, ja 25 klusterissa on pelkästään 20 ryhmän 0 jäsentä ja 25 klusterissa pelkästään 20 ryhmän 1 jäsentä ($N = 1000$).



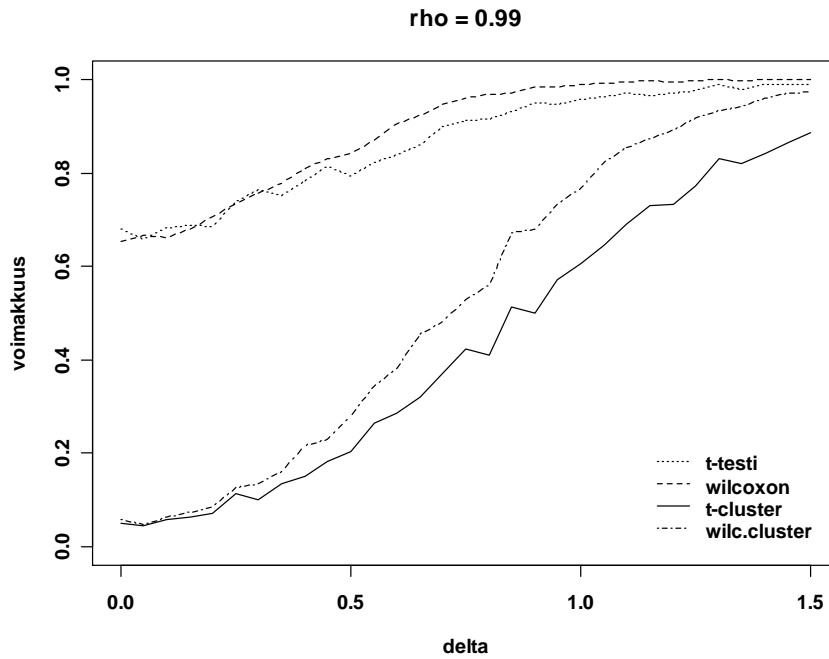
Kuvio 5.5. Simulointikoe epätasapainoiselle normaalijakaumalliselle mallille $M(II, \text{norm})$. Testien voimakkuuksien vertailu Δ :n eri arvoilla, kun $\rho = 0.99$. Klustereita on 50, ja 25 klusterissa on pelkästään 20 ryhmän 0 jäsentä ja 25 klusterissa pelkästään 20 ryhmän 1 jäsentä ($N = 1000$).



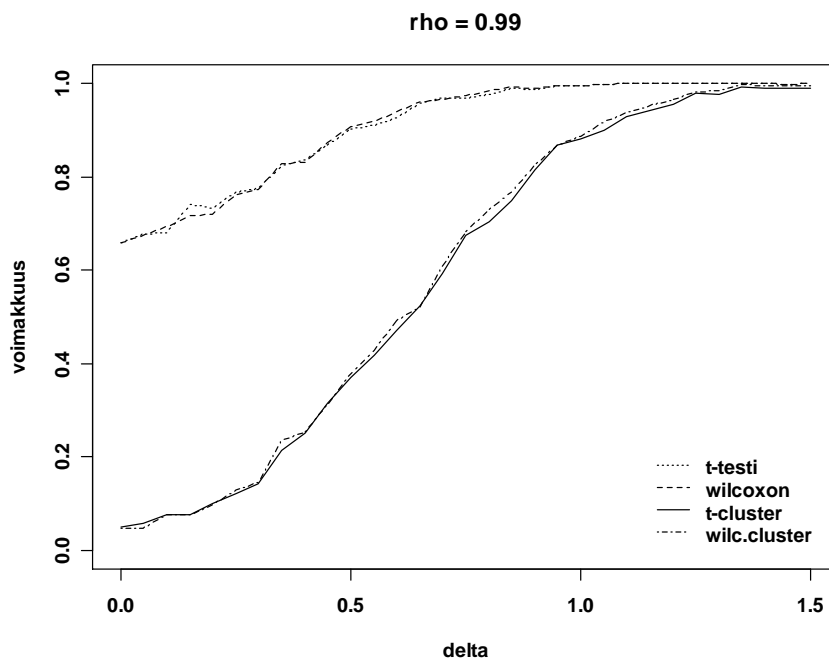
Kuvio 5.6. Simulointikoe epätasapainoiselle t -jakaumalliselle mallille $M(II, t(3))$. Testien voimakkuuksien vertailu ρ :n ja Δ :n eri arvoilla, kun klustereita on 50, ja 25 klusterissa on pelkästään 20 ryhmän 0 jäsentä ja 25 klusterissa pelkästään 20 ryhmän 1 jäsentä ($N = 1000$).



Kuvio 5.7. Simulointikoe epätasapainoiselle t-jakaumalliselle mallille $M(II, t(10))$. Testien voimakkuuksien vertailu ρ :n ja Δ :n eri arvoilla, kun klustereita on 50, ja 25 klusterissa on pelkästään 20 ryhmän 0 jäsentä ja 25 klusterissa pelkästään 20 ryhmän 1 jäsentä ($N = 1000$).



Kuvio 5.8. Simulointikoe epätasapainoiselle t-jakaumalliselle mallille $M(II, t(3))$. Testien voimakkuuksien vertailu Δ :n eri arvoilla, kun $\rho = 0.99$. Klustereita on 50, ja 25 klusterissa on pelkästään 20 ryhmän 0 jäsentä ja 25 klusterissa pelkästään 20 ryhmän 1 jäsentä ($N = 1000$).



Kuvio 5.9. Simulointikoe epätasapainoiselle t-jakaumalliselle mallille $M(II, t(10))$. Testien voimakkuuksien vertailu Δ :n eri arvoilla, kun $\rho = 0.99$. Klustereita on 50, ja 25 klusterissa on pelkästään 20 ryhmän 0 jäsentä ja 25 klusterissa pelkästään 20 ryhmän 1 jäsentä ($N = 1000$).

6 Soveltaminen

Kouluterveyskysely-aineistoon

Tutkimusaineistona käytetään Stakesin Kouluterveyskysely-aineistoa, joka on tyypillinen klusteroitunut aineisto, jossa kouluja voidaan pitää klustereina. Aineisto on rajattu vuoden 2006 pääkaupunkiseudun neljän anonyymien kunnan koulujen 8.–9.-luokkalaisten vastauksiin. Vastemuuttujana on nuorten masentuneisuuden oireita mittaava RBDI-masentuneisuuspistemäärä, joka on Raitasalon Mielialakyselyyn perustuva indikaattorimuuttuja (pykälä 2.3.1). Tässä luvussa esitellään analyysien tulokset kahden ryhmän vertailuissa, joissa ryhmittelevinä muuttujina ovat sukupuoli, liikuntaharrastukset, vanhempien työttömyys, seksuaalikokemukset, suhde vanhempiin sekä suhde opettajiin.

6.1 Tutkimusmenetelmät

Empiirisen aineiston analyysissä vertaillaan toisaalta klassisten ja klusteroitujen testien tuloksia keskenään, toisaalta parametristen ja epäparametristen testien tuloksia keskenään. Testeinä käytetään kahden riippumattoman otoksen Studentin t-testiä ja Wilcoxonin järjestyslukutestiä sekä näiden testien klusteroituja versioita. Jakaumien vinouden vuoksi keskiarvo ja parametriset testit eivät ole sopivimpia jakaumien sijainnin vertailuun, mutta tässä niitä käytetään, koska eri testien tuloksia halutaan vertailla samoin kuin simulointikokeissa.

Analyysien tekemiseen käytetään R-ohjelmistoa (versio 2.5.0) ja klusteroituille testeille R:lle koodattuja funktioita: `t.test.cluster` ja `wilcox.cluster` (liite A). Koko aineiston analyysissä klassiset testit toteutetaan myös näillä klusteroiduilla funktioilla, jolloin kukin havainto kuuluu omaan klusteriin eli klusterien koko on 1. Näin testisuureiden arvot ovat samassa mittakaavassa ja vertailukelpoisia keskenään. Analyysien riskitasoksi valitaan 5 %. Datan käsittelyssä käytetään myös SPSS 15.0 for Windows -ohjelmistoa.

Kouluterveyskysely-aineistossa on 16936 vastemuuttujan havainnon sisältävää tilastoyksikköä. Aineiston suuri koko jo sinänsä kasvattaa testisuureiden arvoja, jolloin saadaan käytännössä aina pieniä p-arvoja, vaikka ryhmien välillä ei olisikaan suurta eroa. Lisätarkasteluna aineistosta demonstroidaan pienempiä otoksia, jolloin testisuureet eivät kasva pelkästään aineiston koon vuoksi. Näille osaotoksille tehdään vastaavat analyysit kuin koko aineistolle. Tällöin koulut, joiden oppilaiden vastauksia käytetään analyysihin, valitaan satunnaisesti kuhunkin osaotokseen, joita poimitaan 1000 kertaa palauttaen.

Analyyysien lähtökohta:

- Klusteri: koulu
- Vastemuuttuja: RBDI-masentuneisuuspisteet
- Kahden ryhmän (0/Epätosi ja 1/Tosi) vertailu, ryhmät:
 - Sukupuoli (poika)
 - Liikuntaa harvemmin kuin kerran viikossa
 - Vähintään yksi vanhempi työttömänä vuoden aikana
 - On seksuaalikokemuksia
 - Keskustelua vanhempien kanssa ei juuri koskaan
 - Huono suhde ja kommunikointi oppilaan ja opettajien välillä.

6.2 Osaotosten testausasetelma

Osaotosten analyyseissä mukaan otetaan koulut, joissa on vähintään 20 oppilasta. Viidessä koulussa on alle 20 oppilasta, joten aineistoon otetaan mukaan 110 koulua. Koulut valitaan kaksiasteisella ryväsotannalla, jossa valituista kouluista poimitaan vielä satunnaisesti vain osa mukaan otettaviksi havainnoiksi.

Demonstroidaan siis osaotoksia, joihin valitaan 110 koulusta yksinkertaisella satunnaisotannalla (YSO) palauttamatta 20 koulua eli klusteria, joista jokaisesta valitaan edelleen 20 oppilasta YSO:lla. Siten otokseen tulee yhteensä 400 oppilasta. Ryhmien kokoa tai niiden välistä suhdetta ei määritellä, vaan satunnaisotannalla saadaan likimain samat todennäköisyydet tulla valituiksi ryhmistä 0 ja 1 kuin koko aineiston ryhmien osuudet kussakin klusterissa. Osaotos sisältää siis:

- otoskoko 400 ($N = 400$), jossa
- 20 klusteria ($n = 20$), joissa
- 20 oppilasta kussakin klusterissa ($m_i = 20$; $i = 1, \dots, 20$) ja oppilaat kuuluvat ryhmään 0 tai 1.

Testien p-arvot ja otosten ryhmien osuudet palautetaan 1000 toistosta.

6.3 Analyyysien tulokset

Tarkastellaan vastemuuttuja RBDI-masentuneisuuspisteiden jakaumia ensin ryhmittäin. Havaitut RBDI-pisteiden arvot ovat välillä 0–36, ja taulukosta 6.1 nähdään ryhmittäiset tunnusluvut. Suurimmat erot ryhmien keskiarvojen välillä ovat indikaattorimuuttujilla suhde vanhempiin (4.9 pistettä), liikuntaharrastukset (3.4 pistettä) ja suhde opettajiin (2.6 pistettä). Indikaattorimuuttujien sukupuoli, vanhempien työttömyys ja seksuaalikokemukset ryhmien keskiarvojen erot ovat 1.5–1.9 pistettä. Vastemuuttujan ryhmittäiset jakaumat ovat hui-

Taulukko 6.1. Ryhmittäiset RBDI-masentuneisuuspistemäärän tunnusluvut pääkaupunkiseudun 8.–9.-luokkalaisille oppilaille. (Ka=keskiarvo, Md=mediaani, s=otoshajonta)

Ryhmittelevä muuttuja		RBDI-pisteet		
		Ka	Md	s
Sukupuoli ($N = 16936$)	Tyttö	4.0	2.0	4.9
	Poika	2.5	1.0	4.3
Liikuntaa harvemmin kuin kerran viikossa ($N = 16767$)	Epätosi	3.0	2.0	4.2
	Tosi	6.4	4.0	7.6
Vähintään yksi vanhempi työttömänä vuoden aikana ($N = 16612$)	Epätosi	3.0	1.0	4.3
	Tosi	4.6	3.0	5.8
On seksuaalikokemuksia ($N = 16240$)	Epätosi	2.9	1.0	4.2
	Tosi	4.8	3.0	6.3
Keskustelua vanhempien kanssa ei juuri koskaan ($N = 16671$)	Epätosi	2.8	1.0	3.9
	Tosi	7.7	5.0	7.6
Huono suhde ja kommunikointi oppilaan ja opettajien välillä ($N = 16890$)	Epätosi	2.8	1.0	3.9
	Tosi	5.4	3.0	6.6

pukkaita ja oikealle vinoja. Kuviossa 6.1 nähdään masennusoireiden syvyyden mukaan luokitellun RBDI-muuttujan ryhmittäiset frekvenssit.

Koulujen kokonaisvaihtelu muodostuu koulujen välisestä ja koulujen sisäisestä vaihtelusta

$$\sum_i \sum_j (X_{ij} - \bar{X}_{..})^2 = \sum_i \sum_j (\bar{X}_{i.} - \bar{X}_{..})^2 + \sum_i \sum_j (X_{ij} - \bar{X}_{i.})^2.$$

Koulujen välisen vaihtelun tarkentuva estimaattori on

$$\hat{\tau}^2 = \frac{1}{N} \sum_{i=1}^n \sum_{j=1}^{m_i} (\bar{X}_{i.} - \bar{X}_{..})^2$$

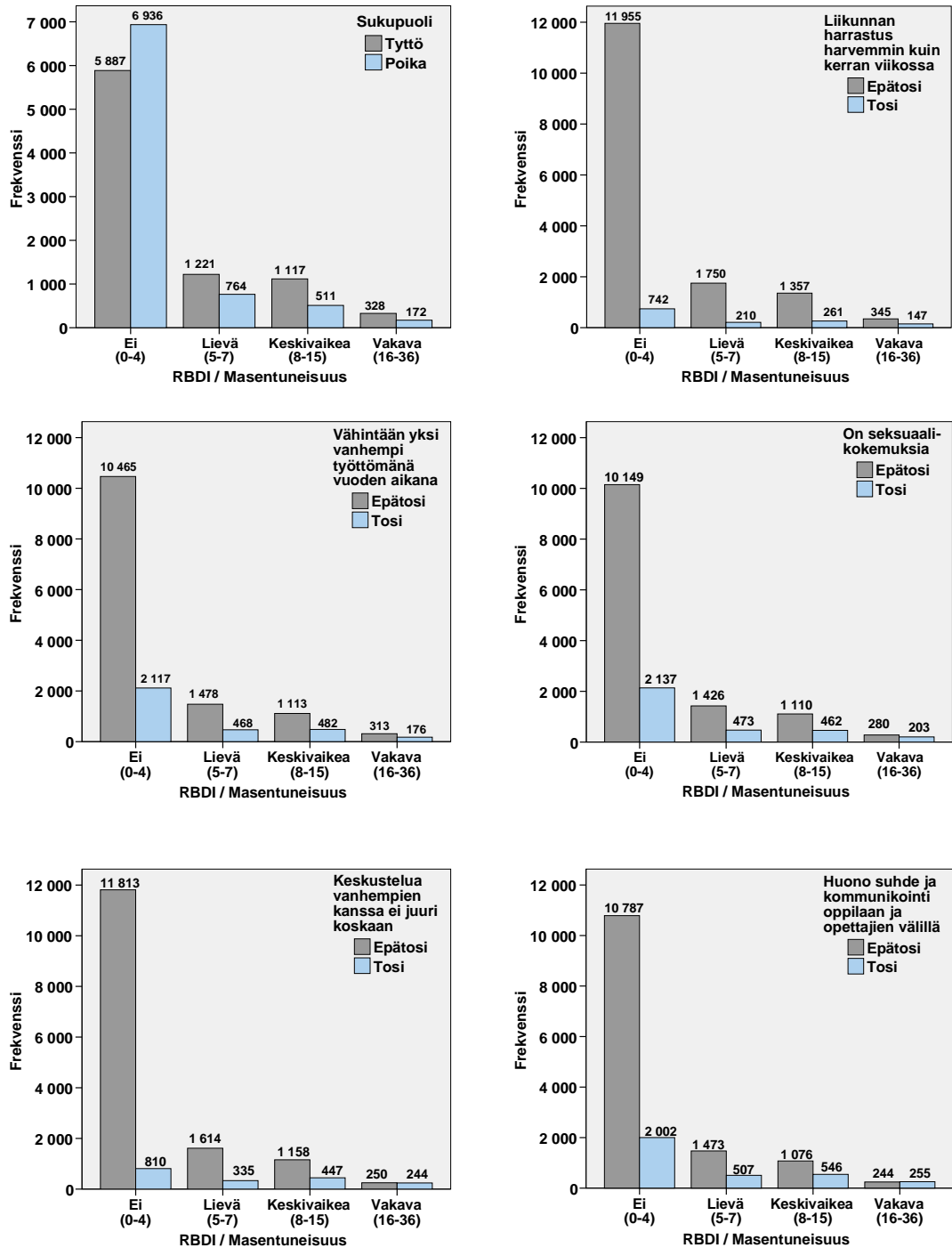
ja koulujen sisäisen vaihtelun tarkentuva estimaattori

$$\hat{\sigma}^2 = \frac{1}{N} \sum_{i=1}^n \sum_{j=1}^{m_i} (X_{ij} - \bar{X}_{i.})^2.$$

Aineistosta lasketut estimaatit ovat: $\hat{\tau}^2 = 0.317$ ja $\hat{\sigma}^2 = 21.736$, joten kokonaisvaihtelun estimaatti on 22.053. Klusterin sisäisen korrelaatiokertoimen

$$\hat{\rho} = \frac{\hat{\tau}^2}{\hat{\tau}^2 + \hat{\sigma}^2}$$

estimaatiksi saadaan 0.014. Vastemuuttujan klusterin sisäinen riippuvuus on siis hyvin heikko.



Kuvio 6.1. Pylväskuviot ryhmittäisistä masennusoireiden syvyyden mukaan luokitellun RBDI-muuttujan jakaumista.

6.3.1 Sukupuoli

Kun ryhmittelevänä muuttujana on sukupuoli, tyttöjen RBDI-pisteiden keskiarvo on 1.5 pistettä suurempi kuin poikien. Koko aineistosta ($N = 16936$) tehtyjen analyysien testisuureet ja p-arvot ovat:

Testi	Testisuure	p-arvo
klusteroitu Wilcoxonin testi	W^C -8.74	2.3e-18
klusteroitu t-testi	T^C -8.65	5.1e-18
klassinen Wilcoxonin testi	W -27.22	4.0e-163
klassinen t-testi	t -20.65	9.4e-95

Kaikkien testien p-arvot ovat < 0.001 eli tyttöjen ja poikien RBDI-pisteet eroavat toisistaan tilastollisesti erittäin merkitsevästi.

Tarkastellaan sitten tilannetta osaotosten avulla ja valitaan satunnaisesti 20 koulua, joista valitaan satunnaisesti 20 oppilasta, joten otoskoko on 400. Tyttöjen osuus otoksissa on keskimäärin 49.9 % ja poikien osuus 50.1 %. Koko aineiston vastaavat osuudet ovat 50.5 % ja 49.5 % (taulukko 2.2). Kuviossa 6.2 nähdään 1000 osaotoksen p-arvojen jakaumat kaikilla neljällä testillä. Nollahypoteesin hylkäämistodennäköisyys 5 %:n riskitasolla on klusteroidulla Wilcoxonin testillä 0.96, klusteroidulla t-testillä 0.77, Wilcoxonin testillä 0.98 ja t-testillä 0.83.

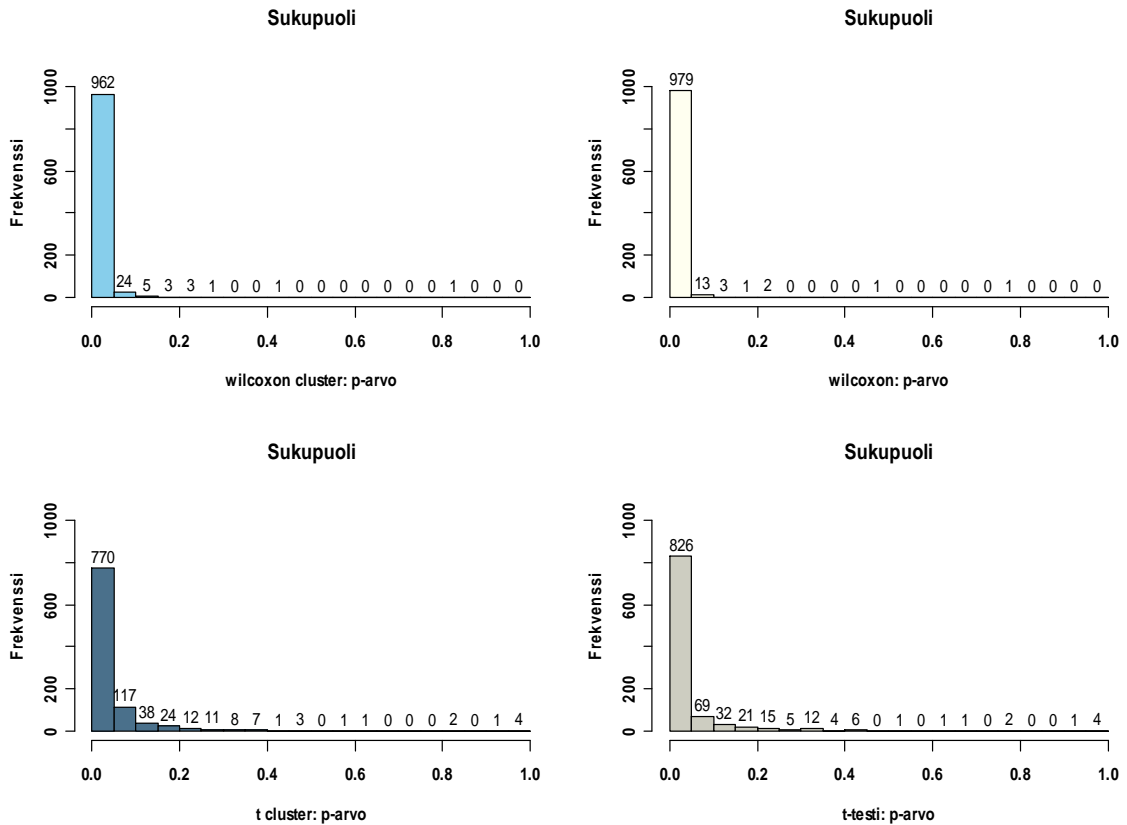
Jos verrataan erikseen klusteroituja (W^C vs. T^C) ja klassisia testejä (W vs. t), niin molemmissa tapauksissa epäparametrinen testi hylkää nollahypoteesin useammin kuin parametrinen vastaava testi. Jos taas verrataan erikseen epäparametrisia (W^C vs. W) ja parametrisia testejä (T^C vs. t), niin molemmissa tapauksissa klassinen testi hylkää nollahypoteesin useammin kuin klusteroitu vastaava testi.

6.3.2 Liikuntaharrastukset

Kun ryhmittelevänä muuttujana on liikuntaharrastukset, liikuntaa harvemmin kuin kerran viikossa harrastavilla (ryhmä 1) oppilailla on keskimäärin 3.4 pistettä suuremmat RBDI-pisteet kuin niillä, jotka harrastavat liikuntaa vähintään kerran viikossa (ryhmä 0). Koko aineistosta ($N = 16767$) tehtyjen analyysien testisuureet ja p-arvot ovat:

Testi	Testisuure	p-arvo
klusteroitu Wilcoxonin testi	W^C 8.34	$< 1e-16$
klusteroitu t-testi	T^C 7.80	6.0e-15
klassinen Wilcoxonin testi	W 17.02	$< 1e-16$
klassinen t-testi	t 15.00	$< 1e-16$

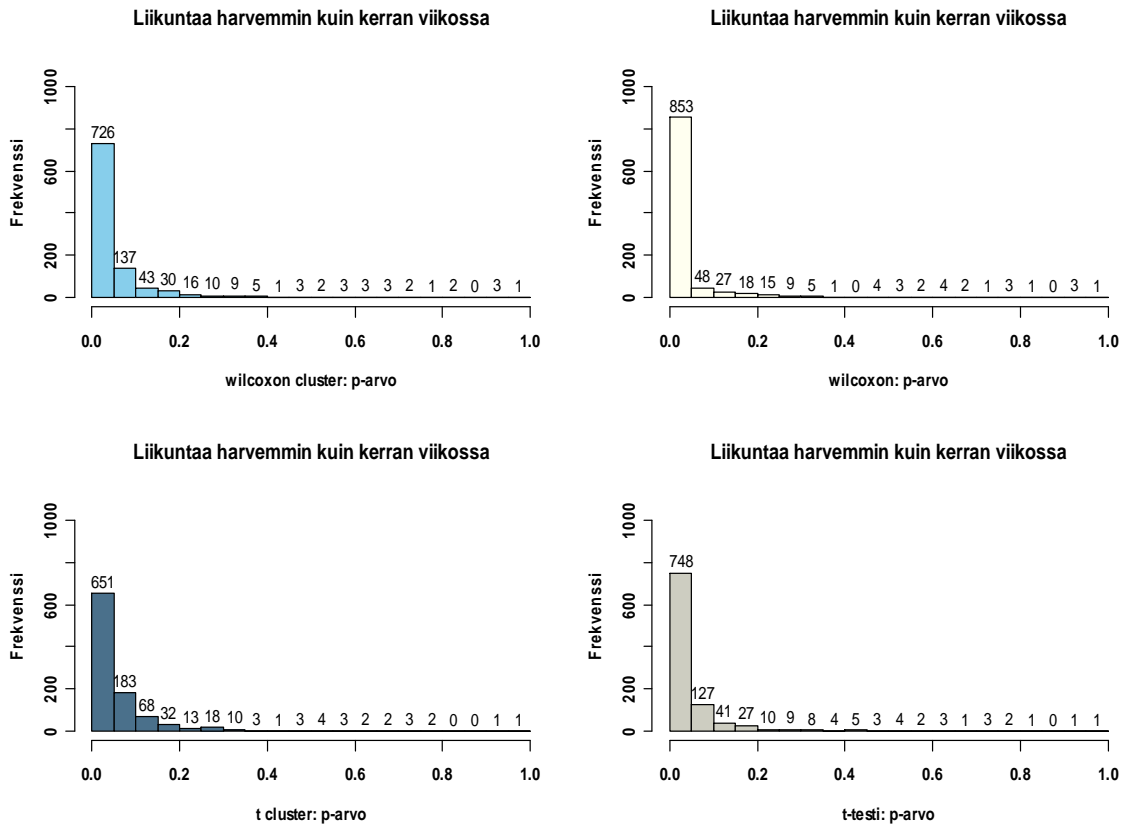
Kaikkien testien mukaan ryhmät eroavat toisistaan (p-arvot < 0.001) tilastollisesti erittäin merkitsevästi eli liikuntaharrastuksilla näyttäisi olevan vaikutusta nuorten masentuneisuusoireisiin, joita on mitattu RBDI-mittarilla.



Kuvio 6.2. 1000 osaotoksen p-arvojen jakaumat, kun vastemuuttujana on RBDI-masentuneisuuspisteet ja ryhmittelevänä muuttujana sukupuoli. Otoksissa on 20 koulua, joissa 20 oppilasta ($N = 400$).

Tarkastellaan tilannetta osaotosten avulla ja valitaan satunnaisesti 20 koulua, joista mukaan valitaan satunnaisesti 20 oppilasta. Ryhmän 1 osuus otoksissa on keskimäärin 8.7 % ja ryhmän 0 osuus 91.3 %. Koko aineiston vastaavat osuudet ovat 8.1 % ja 91.9 % (taulukko 2.2). Kuviossa 6.3 nähdään 1000 osaotoksen p-arvojen jakaumat kaikilla neljällä testillä. Nollahypoteesin hylkäämistodennäköisyys valitulla 5 %:n riskitasolla on klusteroidulla Wilcoxonin testillä 0.73, klusteroidulla t-testillä 0.65, Wilcoxonin testillä 0.85 ja t-testillä 0.75.

Jos verrataan erikseen klusteroituja (W^C vs. T^C) ja klassisia testejä (W vs. t), niin samoin nyt molemmissa tapauksissa epäparametrinen testi hylkää nollahypoteesin useammin kuin parametrinen vastaava testi. Jos taas verrataan erikseen epäparametrisia (W^C vs. W) ja parametrisia testejä (T^C vs. t), niin myös nyt molemmissa tapauksissa klassinen testi hylkää nollahypoteesin useammin kuin klusteroitu vastaava testi.



Kuvio 6.3. 1000 osaotoksen p-arvojen jakaumat, kun vastemuuttujana on RBDI-masentuneisuus pisteet ja ryhmittelevänä muuttujana liikuntaharrastukset. Otoksissa on 20 koulua, joissa 20 oppilasta ($N = 400$).

6.3.3 Vanhempien työttömyys

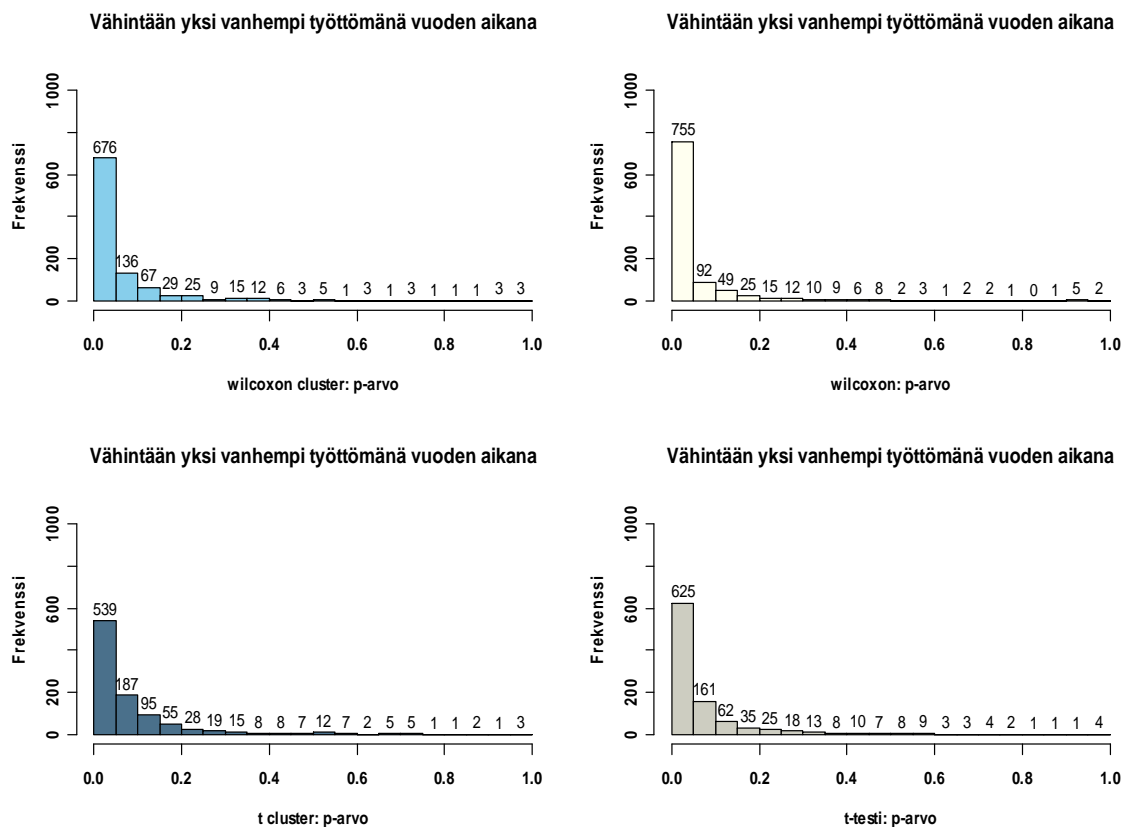
Kun ryhmittelevänä muuttujana on vanhempien työttömyys, niillä oppilailla, joiden vanhemmista ainakin toinen on ollut työttömänä viimeksi kuluneen vuoden aikana (ryhmä 1), on keskimäärin 1.6 pistettä suuremmat RBDI-pisteet kuin niillä, joiden vanhemmat eivät ole olleet työttömänä (ryhmä 0). Koko aineistosta ($N = 16612$) tehtyjen analyysien testisuureet ja p-arvot ovat:

Testi	Testisuure	p-arvo
klusteroitu Wilcoxonin testi	W^C 8.24	2.2e-16
klusteroitu t-testi	T^C 7.76	8.7e-15
klassinen Wilcoxonin testi	W 16.58	< 1e-16
klassinen t-testi	t 14.46	< 1e-16

Kaikkien testien tulokset ovat tilastollisesti erittäin merkitseviä (p-arvot < 0.001), ja siis vanhempien työttömyys näyttäisi vaikuttavan nuorten masennusoireisiin.

Tarkastellaan vielä tilannetta osaotosten avulla ja valitaan satunnaisesti 20 koulua, joista mukaan valitaan satunnaisesti 20 oppilasta. Ryhmän 1 osuus otoksissa on keskimäärin 21.4 % ja ryhmän 0 osuus 78.6 %. Koko aineiston vastaavat osuudet ovat 19.5 % ja 80.5 % (taulukko 2.2). Kuviossa 6.4 nähdään 1000 osaotoksen p-arvojen jakaumat kaikilla neljällä testillä. Nollahypoteesin hylkäämistodennäköisyys 5 %:n riskitasolla on klusteroidulla Wilcoxonin testillä 0.68, klusteroidulla t-testillä 0.54, Wilcoxonin testillä 0.76 ja t-testillä 0.63.

Jos verrataan erikseen klusteroituja (W^C vs. T^C) ja klassisia testejä (W vs. t), niin jälleen molemmissa tapauksissa epäparametrinen testi hylkää nollahypoteesin hieman useammin kuin parametrinen vastaava testi. Jos taas verrataan erikseen epäparametrisia (W^C vs. W) ja parametrisia testejä (T^C vs. t), niin myöskin nyt molemmissa tapauksissa klassinen testi hylkää nollahypoteesin useammin kuin klusteroitu testi.



Kuvio 6.4. 1000 osaotoksen p-arvojen jakaumat, kun vastemuuttujana on RBDI-masentuneisuus pisteet ja ryhmittelevänä muuttujana vanhempien työttömyys. Otoksissa on 20 koulua, joissa 20 oppilasta ($N = 400$).

6.3.4 Seksuaalikokemukset

Kun ryhmittelevänä muuttujana on seksuaalikokemukset, niillä oppilailla, joilla on seksuaalikokemuksia (ryhmä 1), on keskimäärin 1.9 pistettä suuremmat RBDI-pisteet kuin niillä, joilla ei ole seksuaalikokemuksia (ryhmä 0). Koko aineistosta ($N = 16240$) tehtyjen analyysien testisuureet ja p-arvot ovat:

Testi	Testisuure	p-arvo
klusteroitu Wilcoxonin testi	W^C 8.08	6.7e-16
klusteroitu t-testi	T^C 7.84	4.7e-15
klassinen Wilcoxonin testi	W 16.50	< 1e-16
klassinen t-testi	t 15.46	< 1e-16

Kaikkien testien tulokset ovat tilastollisesti erittäin merkitseviä (p-arvot < 0.001), joten nuorten seksuaalikokemuksilla näyttäisi olevan yhteys masennusoireisiin.

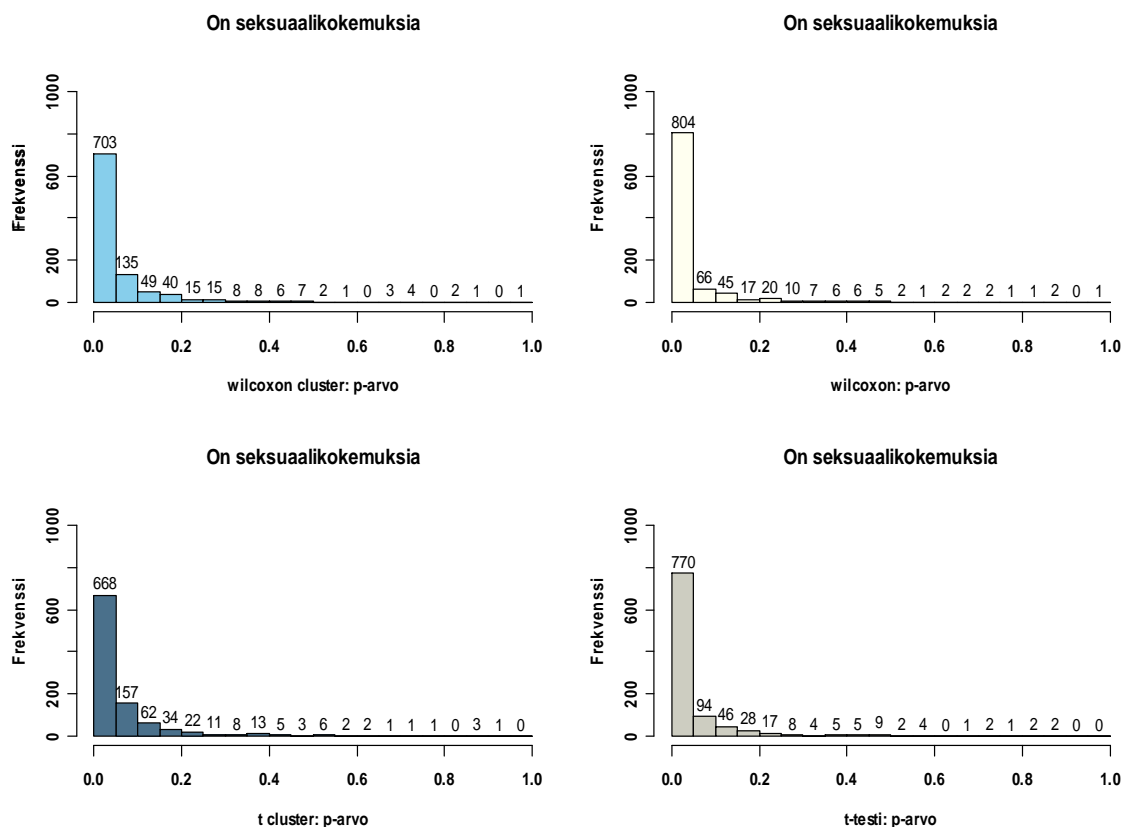
Tarkastellaan tilannetta myös osaotosten avulla ja valitaan satunnaisesti 20 koulua, joista mukaan valitaan satunnaisesti 20 oppilasta. Ryhmän 1 osuus otoksissa on keskimäärin 22.2 % ja ryhmän 0 osuus 77.8 %. Koko aineiston vastaavat osuudet ovat 20.2 % ja 79.8 % (taulukko 2.2). Kuviossa 6.5 nähdään 1000 osaotoksen p-arvojen jakaumat kaikilla neljällä testillä. Nollahypoteesin hylkäämistodennäköisyys valitulla 5 %:n riskitasolla on klusteroidulla Wilcoxonin testillä 0.70, klusteroidulla t-testillä 0.67, Wilcoxonin testillä 0.80 ja t-testillä 0.77.

Jos verrataan erikseen klusteroituja (W^C vs. T^C) ja klassisia testejä (W vs. t), niin nytkin molemmissa tapauksissa epäparametrinen testi hylkää nollahypoteesin hieman useammin kuin parametrinen testi. Jos taas verrataan erikseen epäparametrisia (W^C vs. W) ja parametrisia testejä (T^C vs. t), niin myös nyt molemmissa tapauksissa klassinen testi hylkää nollahypoteesin useammin kuin klusteroitu vastaava testi.

6.3.5 Suhde vanhempiin

Kun ryhmittelevänä muuttujana on suhde vanhempiin, niillä oppilailla, jotka eivät juuri koskaan pysty keskustelemaan omista asioistaan vanhempiensa kanssa (ryhmä 1), on keskimäärin 4.9 pistettä suuremmat RBDI-pisteet kuin niillä, jotka pystyvät keskustelemaan asioistaan vanhempiensa kanssa (ryhmä 0). Koko aineistosta ($N = 16671$) tehtyjen analyysien testisuureet ja p-arvot ovat:

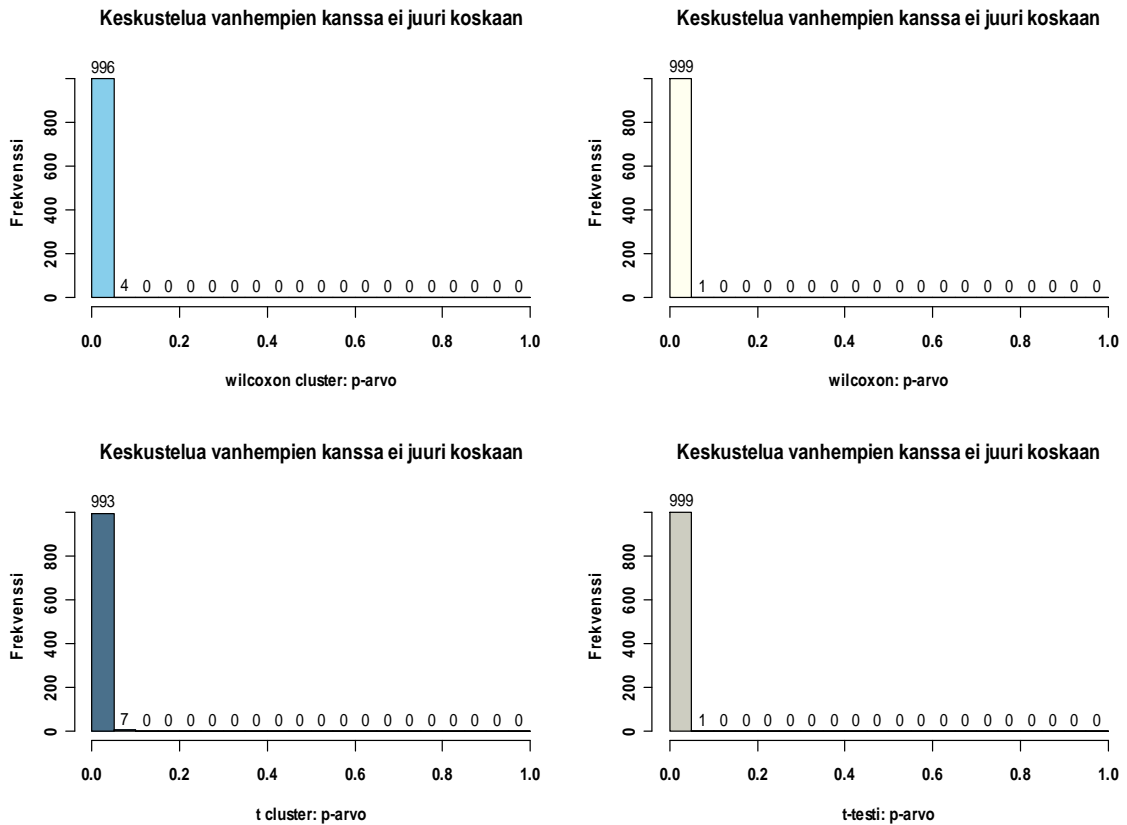
Testi	Testisuure	p-arvo
klusteroitu Wilcoxonin testi	W^C 9.18	< 1e-16
klusteroitu t-testi	T^C 9.02	< 1e-16
klassinen Wilcoxonin testi	W 29.20	< 1e-16
klassinen t-testi	t 23.72	< 1e-16



Kuvio 6.5. 1000 osaotoksen p-arvojen jakaumat, kun vastemuuttujana on RBDI-masentuneisuuspisteet ja ryhmittelevänä muuttujana seksuaalikoemukset. Otoksissa on 20 koulua, joissa 20 oppilasta ($N = 400$).

Kaikkien testien p-arvot ovat < 0.001 , joten huonon keskusteluyhteyden vanhempiin omaavilla oppilailla on tilastollisesti erittäin merkitsevästi suuremmat masentuneisuusoirepisteet kuin niillä, jotka pystyvät keskustelemaan vanhempiensa kanssa.

Tarkastellaan tilannetta osaotosten avulla ja valitaan satunnaisesti 20 koulua, joista mukaan valitaan satunnaisesti 20 oppilasta. Ryhmän 1 osuus otoksissa on keskimäärin 11.7 % ja ryhmän 0 osuus 88.3 %. Koko aineiston vastaavat osuudet ovat 11.0 % ja 89.0 % (taulukko 2.2). Kuviossa 6.6 nähdään 1000 osaotoksen p-arvojen jakaumat kaikilla neljällä testillä. Nollahypoteesin hylkäämistodennäköisyys valitulla 5 %:n riskitasolla on klusteroidulla t-testillä 0.99 ja muilla testeillä noin 1.00. Eli lähes jokaisessa 1000 osaotoksessa kaikkien testien p-arvot ovat < 0.05 .



Kuvio 6.6. 1000 osaotoksen p-arvojen jakaumat, kun vastemuuttujana on RBDI-masentuneisuuspisteet ja ryhmittelevänä muuttujana suhde vanhempiin. Otoksissa on 20 koulua, joissa 20 oppilasta ($N = 400$).

6.3.6 Suhde opettajiin

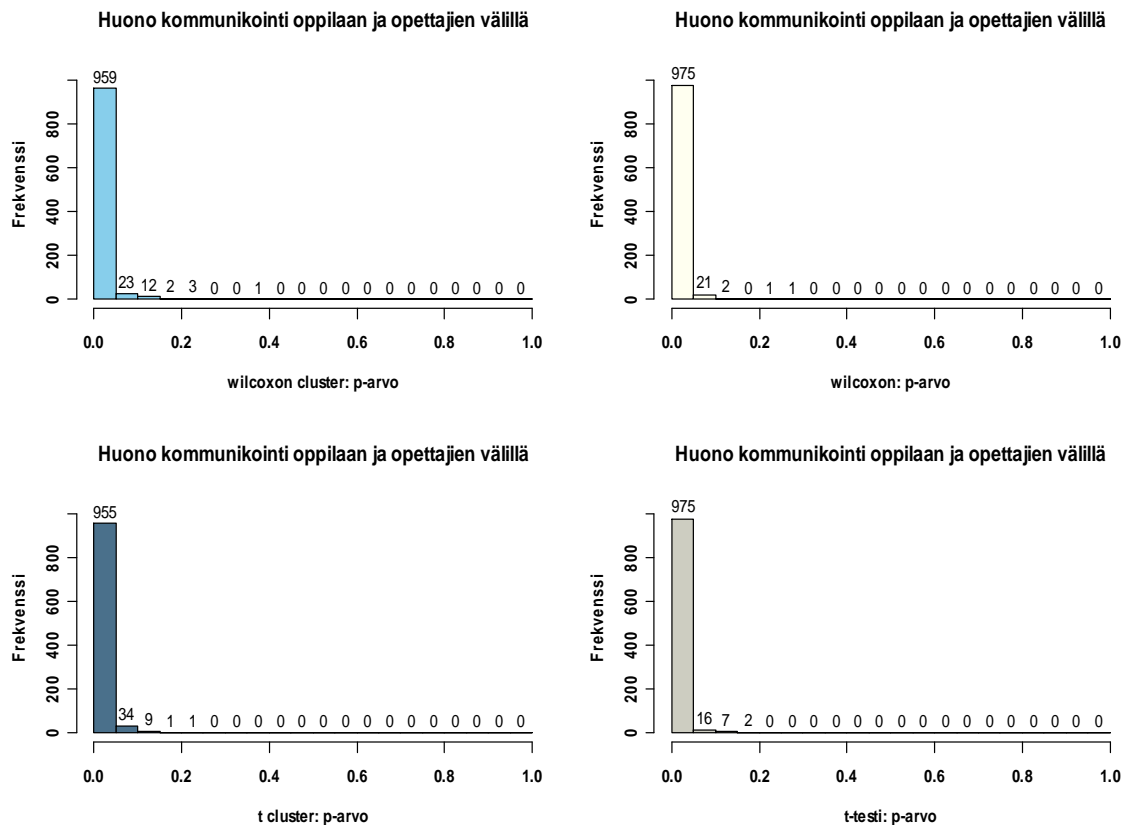
Kun ryhmittelevänä muuttujana on suhde opettajiin, niillä oppilailla, jotka kokevat, etteivät opettajat ole riittävän rohkaisevia, oikeudenmukaisia ja kiinnostuneita heistä tai heidän mielipiteistään (ryhmä 1), on keskimäärin 2.6 pistettä suuremmat RBDI-pisteet kuin niillä oppilailla, jotka kokevat opettajilta rohkaisevaa, oikeudenmukaista ja mielipiteitä huomioivaa suhtautumista (ryhmä 0). Koko aineistosta ($N = 16890$) tehtyjen analyysien testisuureet ja p-arvot ovat:

Testi	Testisuure	p-arvo
klusteroitu Wilcoxonin testi	W^C 8.69	$< 1e-16$
klusteroitu t-testi	T^C 8.54	$< 1e-16$
klassinen Wilcoxonin testi	W 23.47	$< 1e-16$
klassinen t-testi	t 21.00	$< 1e-16$

Kaikkien testien tulokset ovat tilastollisesti erittäin merkitseviä (p-arvot < 0.001), joten nuorten masentuneisuusoireisiin näyttäisi vaikuttavan myös se, kokevatko oppilaat omaavansa hyvän suhteen opettajiinsa vai eivät.

Tarkastellaan tilannetta osaotosten avulla ja valitaan satunnaisesti 20 koulua, joista mukaan valitaan satunnaisesti 20 oppilasta. Ryhmän 1 osuus otoksissa on keskimäärin 19.3 % ja ryhmän 0 osuus 80.7 %. Koko aineiston vastaavat osuudet ovat 19.6 % ja 80.4 % (taulukko 2.2). Kuviossa 6.7 nähdään 1000 otoksen p-arvojen jakaumat kaikilla neljällä testillä. Nollahypoteesin hylkäämistodennäköisyys 5 %:n riskitasolla on molemmilla klusteroiduilla testeillä 0.96 ja molemmilla klassisilla testeillä 0.98.

Jos verrataan erikseen klusteroituja (W^C vs. T^C) ja klassisia testejä (W vs. t), niin klusteroidut testit hylkäävät nollahypoteesin likimain yhtä usein sekä myös klassiset testit keskenään yhtä usein. Jos taas verrataan erikseen epäparametrisia (W^C vs. W) ja parametrisia testejä (T^C vs. t), niin jälleen nyt molemmissa tapauksissa klassinen testi hylkää nollahypoteesin hieman useammin kuin klusteroitu vastaava testi.



Kuvio 6.7. 1000 osaotoksen p-arvojen jakaumat, kun vastemuuttujana on RBDI-masentuneisuuspisteet ja ryhmittelevänä muuttujana suhde opettajiin. Otoksissa on 20 koulua, joissa 20 oppilasta ($N = 400$).

6.4 Analyysien yhteenveto

Koko aineistosta tehtyjen analyysien perusteella kaikkien ryhmien (sukupuoli, liikuntaharrastukset, vanhempien työttömyys, seksuaalikokemukset, suhde vanhempiin, suhde opettajiin) erot nuorten masentuneisuuspisteissä ovat tilastollisesti merkitseviä. Kaikkien ryhmittelymuuttujien tilanteissa pienin riskitaso, jolla nollahypoteesi voidaan hylätä, on pienempi kuin 0.001.

Osaotosten perusteella kaikkien ryhmittelymuuttujien tulokset ovat samansuuntaisia. Verrattaessa sekä klusteroituja testejä (klusteroitu Wilcoxon vs. klusteroitu t-testi) että klassisia testejä (klassinen Wilcoxon vs. klassinen t-testi) keskenään molemmissa tapauksissa epäparametrinen testi hylkää nollahypoteesin useammin kuin parametrinen vastineensa. Siis epäparametristen testien p-arvot ovat pienempiä kuin parametristen. Kyseessä on vinot jakaumat, jolloin epäparametristen testien voidaankin olettaa olevan voimakkaampia ja luotettavampia kuin parametristen testien. Kun verrataan sekä epäparametrisia (klusteroitu Wilcoxon vs. klassinen Wilcoxon) että parametrisia (klusteroitu t-testi vs. klassinen t-testi) testejä keskenään, klassiset testit puolestaan antavat pienempiä p-arvoja ja hylkäävät nollahypoteesin useammin kuin klusteroidut vastaavat testit.

Osaotosten ryhmittelymuuttujien tulokset vastaavat kutakuinkin luvun 5 simulointikokeiden tuloksia: 1) ryhmittelymuuttuja sukupuoli vastaa tasapainoisen (klusterirakennetyyppi *I*), ei-normaalisen mallin tilannetta sekä 2) muut viisi ryhmittelymuuttujaa: liikuntaharrastukset, vanhempien työttömyys, seksuaalikokemukset, suhde vanhempiin ja suhde opettajiin, lähinnä epätasapainoisen (klusterirakennetyyppi *II*), ei-normaalisen mallin tilannetta. Ensin mainitussa tapauksessa, korrelaation ollessa lähellä nollaa, simulointikokeiden epäparametriset testit antoivat pienempiä p-arvoja kuin parametriset testit (kuvio 5.2) kuten nytkin ryhmittelymuuttuja sukupuolen kohdalla. Jälkimmäisessä tapauksessa klassiset testit antoivat simuloinneissa virheellisen pieniä p-arvoja (kuvio 5.6). Nytkin siis klassiset testit antavat pienempiä p-arvoja kuin klusteroidut testit, mutta syy saattaa olla siinä, että klassiset testit ovat myös nyt epävalideja ja antavat kenties virheellisiä tuloksia. On kuitenkin huomioitavaa, että nyt kouluissa on molempien ryhmien jäseniä eikä kyseessä ole täysin epätasapainoinen klusterirakenne. Ryhmien väliset suhteet (siis sukupuolta luukuunottamatta) ovat välillä $\frac{1}{12} - \frac{1}{5}$.

Aineistosta estimoitu klusterin sisäinen korrelaatio on vain 0.014, mutta kuten aiemmin simulointien epätasapainoisissa malleissa huomattiin, klassiset testit antoivat virheellisen pieniä p-arvoja jo korrelaation arvolla 0.02, vaikka ryhmien välillä ei ollut eroa. Nuorten masentuneisuusoireet eivät siis riipu kouluista. Sitä vastoin sukupuolella, liikuntaharrastuksilla, vanhempien työttömyydellä, seksuaalikokemuksilla, suhteella vanhempiin ja suhteella opettajiin näyttäisi olevan vaikutusta nuorten masentuneisuuteen. Näin alaluvussa 2.2 mainitut tutkimushypoteesit saavat aineiston perusteella tukea riippuvuuskien olemassaoloon.

7 Yhteenveto

Tutkielmassa tarkastellaan klusteroitumisen ja klusterin sisäisen riippuvuuden vaikutuksia tilastollisissa analyysissä. Käsittelyssä rajoitutaan yhteen klusteritasoon kahden ryhmän vertailussa. Klusterit jakavat aineiston riippumattomiin osiin, mutta saman klusterin havainnot ovat riippuvia keskenään. Perinteiset menetelmät käsittelevät aineistoa yhtenä kokonaisuutena eivätkä näin ollen huomioi klusterien sisäistä riippuvuutta. Työssä esitellään klassisten menetelmien lisäksi mallipohjainen teoria yksiulotteisen aineiston kahden ryhmän sijainnin vertailuun. Esitettyjen menetelmien perusideana on klusterin sisäisen korrelaation huomioiminen varianssin estimoinnissa.

Menetelmien vertailut simulointikokeissa osoittavat, että tasapainoisessa tilanteessa, kun klusterien ryhmäkoot ovat yhtä suuret, uudet klusteroidut menetelmät ovat tehokkaampia kuin klassiset menetelmät aineiston ollessa klusterin sisäisesti korreloitunut. Klassisten testien p-arvot ovat liian suuria, koska varianssiestimaatit ovat ylöspäin harhaisia, ja testit hylkäävät nollahypoteesin liian harvoin. Tällöin parametrusten ja epäparametrusten testien erot, kun verrataan toisaalta klassisia testejä keskenään, toisaalta klusteroituja testejä keskenään, kääntyvät normaalijakaumatilanteessa parametrusten testien hyväksi ja paksuhäntäisen jakauman tilanteessa epäparametrusten testien hyväksi.

Simulointikokeiden epätasapainoisessa tilanteessa, kun klusterissa on vain toisen ryhmän jäseniä, klusterin sisäinen riippuvuus vaikuttaa siten, että klassisten menetelmien varianssiestimaatit ovat puolestaan alaspäin harhaisia. Klassisten testien p-arvot ovat siten liian pieniä, joten nollahypoteesi hylätään liian usein. Klassiset testit osoittautuvat epävalideiksi ja antavat jopa virheellisiä tuloksia. Lisäksi tällöin klusteroidut testit menettävät tehokkuuttaan, mutta ovat kuitenkin valideja. Parametrisilla ja epäparametrisilla menetelmillä, kun verrataan klassisia testejä keskenään ja klusteroituja testejä keskenään, ei epätasapainoisen mallin normaali- ja $t(10)$ -jakauman tilanteissa ole merkittävää eroa, mutta $t(3)$ -jakaumallisen mallin tilanteessa Wilcoxonin testit ovat voimakkaampia kuin vastaavat t -testit.

Kouluterveyskysely-aineiston kohdalla aineiston suuri koko aiheuttaa testi-suureiden arvojen ilmeistä kasvamista. Kaikkien ryhmien kohdalla saadaankin tilastollisesti erittäin merkitsevät erot kaikilla menetelmillä. Tutkimusaineistoa testataan lisäksi pienempien osaotosten avulla, ja vertailut menetelmien kesken antavat samansuuntaisia tuloksia kuin kokeelliset analyysit. Vastemuuttujan klusterin sisäinen korrelaatio on hyvin pieni, mutta jo simuloinneissa huomioitiin, että pienikin klusterin sisäinen korrelaatio vaikuttaa tuloksiin, ja klassisten

ja klusteroitujen testien välillä on eroja jo tällöin. Simulointikokeiden perusteella ja jakaumien muodon vuoksi sopivin, tehokkain ja validi analysointimenetelmä tutkimusaineiston analysointiin on järjestyslukuihin perustuva klusteroitu Wilcoxonin testi.

Tutkielmassa on osoitettu, että klusteroituneen aineiston tapauksessa klassinen Studentin t-testi ja Wilcoxonin järjestyslukutesti eivät ole valideja. Vaikka klassinen Wilcoxonin testi on aidosti jakaumasta vapaa menetelmä, niin tämä hyöty menetetään klusteroitumisessa. Klusteroitujen testien valinnan välillä ovat robustisuus- ja tehokkuussyyt. Klusteroitu t-testi on herkkä poikkeaville havainnoille, ja erot tehokkuudessa tulevat esiin, kun jakauma on vino tai paksuhäntäinen, jolloin Wilcoxonin testi on voimakkaampi kuin t-testi. Vaikka klusteroidussa t-testissä ei oleteta jakauman normalisuutta, niin testi vaatii 2. momentin. Sen sijaan Wilcoxonin testi on asymptoottisesti jakaumasta vapaa testi eikä vaadi momenteja.

Tutkielmaa työstäessä on noussut kysymyksiä useiden asioiden lisätarkasteluihin ja jatkotutkimuksiin. Työ on rajattu klusterirakennetyyppien ääritilanteisiin, mutta miten menetelmät käyttäytyvät erilaisilla ryhmien suhejaoilla tai erisuurilla klusterikoilla? Teorian eteenpäin vieminen monitasoisen hierarkiarakenteen tilanteessa, siis silloin kun klusteritasoja on useita, antaa myös aiheita lisätutkimuksiin kuten myös useamman kuin kahden ryhmän tilanne.

Lopuksi haluan esittää kiitokset ohjaajalleni akatemiaprofessori Hannu Ojalle, joka mahdollisti tämän mielenkiintoisen aihevalinnan ja perehdytti aiheeseen sekä tarjosi monia neuvoja ja oljenkorsia työn eri vaiheissa. Kiitokset myös Stakesin tutkimusprofessori Matti Rimpelälle, joka antoi tutkimusaineiston käyttööni.

Lähdeluettelo

- Beck, A. T. & Beck, R. W. (1972). "Screening depressed patients in family practice. A rapid technic", *Postgrad Med*, 52, 81–85.
- Beck, A. T., Ward, C. H., Mendelson, M., Mock, J. & Erbaugh, J. (1961). "An inventory for measuring depression", *Arch Gen Psychiatry*, 4, 561–571.
- Datta, S. & Satten, G. A. (2005). "Rank-Sum Tests for Clustered Data", *Journal of the American Statistical Association*, 100, 908–915.
- Karlsson, L. & Marttunen, M. (2007), "Nuorten depressio, Tietoa nuorten kanssa työskenteleville aikuisille", *Kansanterveyslaitoksen julkaisuja B10/2007*, Helsinki: Kansanterveyslaitos. http://www.ktl.fi/attachments/suomi/julkaisut/julkaisusarja_b/2007/2007b10.pdf.
- Larocque, D. (2003), "An affine-invariant multivariate sign test for cluster correlated data", *The Canadian Journal of Statistics*, 31, 437–455.
- Larocque, D., Nevalainen, J. & Oja, H. (2007), "A weighted multivariate sign test for cluster-correlated data", *Biometrika*, 94, 267–283.
- Larocque, D., Haataja, R., Nevalainen, J. & Oja, H. (2008), "Two Sample Tests for the Nonparametric Behrens-Fisher Problem with Clustered Data", unpublished manuscript.
- Lehmann, E. L. (2006), *Nonparametrics Statistical Methods Based on Ranks*. Revised 1st ed., New York: Springer.
- Liang, K.-Y., Zeger, S. L. (1986). "Longitudinal Data Analysis Using Generalized Linear Models", *Biometrika*, 73, 13–22.
- Luopa, P., Räsänen, M., Jokela, J. & Rimpelä, M. (2005). "Kouluterveyskyselyn valtakunnalliset tulokset vuosina 1999–2004", *Aiheita 12/2005*. Helsinki: Stakes. <http://info.stakes.fi/NR/rdonlyres/08B4923B-5D9A-44B1-9C7D-7002C0137E15/0/Aiheita122005.pdf>.
- Luopa, P., Sinkkonen, A., Jokela, J., Puusniekka, R. & Pietikäinen, M. (2006). "Kouluterveys 2006: Pääkaupunkiseudun raportti". Helsinki: Stakes. http://info.stakes.fi/NR/rdonlyres/E1FCCDE9-3AE7-4995-9BCA-81C94CF0BAEF/0/pkseutu2006_raportti.pdf.
- Mann, H. B., Whitney, D. R. (1947). "On a Test of Whether One of Two Random Variables is Stochastically Larger than the Other", *Annals of Mathematical Statistics*, 18, 50–60.
- Nevalainen, J., Larocque, D. & Oja, H. (2007a) "On the multivariate spatial median for clustered data", *The Canadian Journal of Statistics*, 35, 215–231.
- (2007b) "A weighted spatial median for clustered data", *Statistical Methods & Applications*, 15, 355–379.

- Raitasalo, R. (1995). *Elämänhallinta sosiaalipolitiikan tavoitteena*, Helsinki: Kansaneläkelaitos, Sosiaali- ja terveysturvan tutkimuksia 1.
- Ranta, E., Rita, H. & Kouki, J. (2005), *Biometria: tilastotiedettä Ekologeille*. 9. painos, Helsinki: Yliopistopaino.
- Rosner, B. (1984). "Multivariate Methods in Ophthalmology with Application to Other Paired-Data situations", *Biometrics*, 40, 1025–1035.
- Rosner, B., Glynn, R. J. & Lee, M.-L. T. (2003). "Incorporation of Clustering Effects for the Wilcoxon Rank Sum Test: A Large-Sample Approach", *Biometrics*, 59, 1089–1098.
- Stoner, J. A. & Leroux, B. G. (2002). "Analysis of clustered data: A combined estimating equations approach", *Biometrika*, 89, 567–578.
- Wilcoxon, F. (1945). "Individual Comparisons by Ranking Methods", *Biometrics*, 1, 80–83.
- Williamson, J. M., Datta, S. & Satten, G. A. (2003). "Marginal Analyses of Clustered Data When Cluster Size is Informative", *Biometrics*, 59, 36–42.

Liite A

R-funktiot

R-funktiot klusteroituneelle datalle: kahden otoksen *klusteroitu Studentin t-testi* ja *Wilcoxonin testi*

A.1 t.test.cluster

Funktion `t.test.cluster`(x , *group*, *clust*) parametreina vastemuuttujavektori x , ryhmittelymuuttujavektori *group* ja klusterinumerovektori *clust*.

Koodi:

```
t.test.cluster <- function(x, group, clust)
{
  if (sum(is.na(x))+sum(is.na(group))+sum(is.na(clust))==0)
  {
    N <- length(x)
    n_b <- length(group[group==1])
    n_a <- length(group[group==0])
    ncluster <- length(table(clust))
    delta <- array(dim=c(length(x)))
    deltaX <- array(dim=c(length(x)))
    delta[group==0] <- -1/(n_a)
    delta[group==1] <- 1/(n_b)
    meanDif <- sum(delta*x)
    clnr <- as.numeric(rownames(table(clust)))
    apu <- array(dim=c(ncluster))
    j <- 1
    for (i in clnr)
    {
      apu[j] <- (sum(delta[clust==i]*(x[clust==i]-mean(x))))^2
      j <- j+1
    }
    var.diff <- sum(apu)
    Tc1 <- meanDif / sqrt(var.diff)
    p <- 2*min(pnorm(Tc1),1-pnorm(Tc1))
    list(T_c1=Tc1, p.value=p)
  }
  else cat("Missing values! \n")
}
```

A.2 wilcox.cluster

Funktio `wilcox.cluster(x, group, clust)` parametreina vastemuuttujavektori x , ryhmittelymuuttujavektori $group$ ja klusterinumerovektori $clust$.

Koodi:

```
wilcox.cluster <- function(x, group, clust)
{
  if (sum(is.na(x))+sum(is.na(group))+sum(is.na(clust))==0)
  {
    rankx <- rank(x)
    N <- length(x)
    n_b <- length(group[group==1])
    n_a <- length(group[group==0])
    ncluster <- length(table(clust))
    delta <- array(dim=c(length(x)))
    deltaR <- array(dim=c(length(x)))
    delta[group==0] <- -1/(n_a)
    delta[group==1] <- 1/(n_b)
    meanDif <- sum(delta*rankx)
    clnr <- as.numeric(rownames(table(clust)))
    apu <- array(dim=c(ncluster))
    j <- 1
    for (i in clnr)
    {
      apu[j] <- (sum(delta[clust==i]*(rankx[clust==i]-mean(rankx))))^2
      j <- j+1
    }
    var.diff <- sum(apu)
    Wcl <- meanDif / sqrt(var.diff)
    p <- 2*min(pnorm(Wcl),1-pnorm(Wcl))
    list(W_cl=Wcl, p.value=p)
  }
  else cat("Missing values! \n")
}
```

Liite B

Indikaattorimuuttujien muodostaminen

Kouluterveyskyselyn muuttujien pisteytys ja indikaattoreiden muodostaminen. Kysymysten numerot viittaavat vuoden 2006 lomakkeeseen.

Masentuneisuus, (tieto saatavilla vuodesta 1997)

- *Lyhyt kuvaus:* Masentuneisuutta mitataan 12 kysymyksellä Raitosalon Mielialakyselystä (RBDI), joka perustuu Beckin depressio-mittariin. Muuttuja saa arvoja väliltä 0–36.
- *Valmis indikaattori; indikaattorin muodostus:*
”Seuraavat kysymykset käsittelevät mielialan erilaisia piirteitä. Vastaa kuhunkin kysymykseen siten, millaiseksi tunnet itsesi tänään. Valitse kustakin kysymyksestä vain yksi vaihtoehto. Joku kysymys voi tuntua Sinusta yllättävältä. Toivomme kuitenkin, että vastaat kaikkiin kysymyksiin.”
24. Minkälainen on mielialasi?
25. Miten suhtaudut tulevaisuuteen?
26. Miten katsot elämäsi sujuneen?
27. Miten tyytyväiseksi tai tyytymättömäksi tunnet itsesi?
28. Minkälaisena pidät itseäsi?
29. Onko Sinulla pettymyksen tunteita?
30. Miten suhtaudut vieraitten ihmisten tapaamiseen?
31. Miten koet päätösten tekemisen?
32. Minkälaisena pidät olemustasi ja ulkonäköäsi?
33. Minkälaista nukkumisesi on?
34. Tunnetko väsymystä ja uupumusta?
35. Minkälainen ruokahalusi on?
Pisteytys: vaihtoehdot 1,2 = 0; vaihtoehto 3 = 1; vaihtoehto 4 = 2; vaihtoehto 5 = 3. (1. vaihtoehto kuvaa positiivista tunnetilaa tai ajatusta, 2. vaihtoehto on neutraalimpi, mutta ei vielä kuvaa masennustilaan liittyviä mielialan piirteitä. 3. vaihtoehdosta alkaen väittämät kuvaavat depressio-oireita, joiden voimakkuus ja syvyys lisääntyvät kohti vaihtoehtoa 5.)

Sukupuoli

- Kaksiluokkainen, kysymys 1. Sukupuoli (Tyttö = 0, Poika = 1), (tieto saatavilla vuodesta 1996)

Liikuntaharrastukset

- **Liikunnan harrastusta harvemmin kuin kerran viikossa,** (tieto saatavilla vuodesta 1996)

- *Lyhyt kuvaus:* Oppilas harrastaa vapaa-aikana urheilua tai liikuntaa, joka kestää vähintään puoli tuntia kerrallaan, harvemmin kuin kerran viikossa.
- *Valmis indikaattori; indikaattorin muodostus:*
77. Kuinka usein harrastat urheilua tai liikuntaa vapaa-aikanasi vähintään puolen tunnin ajan? (Pisteytys: Useita kertoja päivässä, Noim kerran päivässä, 4–6 kertaa viikossa, 2–3 kertaa viikossa, Kerran viikossa = 0; Harvemmin, En lainkaan = 1. Ominaisuus toteutuu, jos vastaaja saa 1 pisteen.)

Vanhempien työttömyys

- **Vähintään yksi vanhempi työttömänä vuoden aikana**, (tieto saatavilla vuodesta 1996)
- *Lyhyt kuvaus:* Vähintään yksi vanhemmista on ollut työttömänä tai pakkolomalla viimeksi kuluneen vuoden aikana.
- *Valmis indikaattori; indikaattorin muodostus:*
98. Ovatko vanhempasi olleet viimeksi kuluneen vuoden aikana työttöminä tai pakkolomalla? (Pisteytys: Ei kumpikaan = 0; Toinen vanhemmistani, Molemmat vanhempani = 1. Ominaisuus toteutuu, jos vastaaja saa 1 pisteen.)

Seksuaalikokemukset

- **On seksuaalikokemuksia**, (tieto saatavilla vuodesta 1996)
- *Lyhyt kuvaus:* On ollut sukupuoliyhteydessä vähintään kerran.
- *Indikaattorin muodostus:*
47. Oletko ollut sukupuoliyhdyneenä? (Pisteytys: Ei = 0, Kyllä = 1. Ominaisuus toteutuu, jos vastaaja saa 1 pisteen.)

Suhde vanhempiin

- **Keskustelua vanhempien kanssa ei juuri koskaan**, (tieto saatavilla vuodesta 1996)
- *Lyhyt kuvaus:* Oppilas ei juuri koskaan pysty keskustelemaan omista asioistaan vanhempiensa kanssa.
- *Indikaattorin muodostus:*
96. Pystytkö keskustelemaan vanhempiesi kanssa omista asioistasi? (Pisteytys: En juuri koskaan = 1; Silloin tällöin, Melko usein, Usein = 0. Ominaisuus toteutuu, jos vastaaja saa 1 pisteen.)

Suhde opettajiin

- **Huono suhde ja kommunikointi oppilaan ja opettajien välillä**, (tieto saatavilla vuodesta 1998)
- *Lyhyt kuvaus:* Opettajat eivät ole riittävän rohkaisevia, oikeudenmukaisia ja kiinnostuneita oppilaista tai heidän mielipiteistään.
- *Indikaattorin muodostus:*
7.1. Opettajat rohkaisevat minua ilmaisemaan oman mielipiteeni oppitunnilla. 7.2. Opettajat ovat kiinnostuneita siitä, mitä minulle kuuluu. 7.4. Opettajat kohtelevat meitä oppilaita oikeudenmukaisesti. 7.7. Oppilaiden mielipiteet otetaan huomioon koulutyön kehittämisessä. (Pisteytys: Täysin samaa mieltä = 0; Samaa mieltä = 1; Eri mieltä = 2; Täysin eri mieltä = 3. Ominaisuus toteutuu, jos vastaaja saa 8 pistettä.)

Liite C

Simulointien tulokset

Taulukoissa: t = Studentin t-testi, w = Wilcoxonin testi, t-cl = t-testin klusteroitu versio, w-cl = Wilcoxonin testin klusteroitu versio.

C.1 Klusterirakennetyyppi I

$$\text{Malli } M(I, \text{norm}): X_{ij} - \Delta G_{ij} = \tau_i + \epsilon_{ij} \sim N(0, 1)$$

$df = \infty, \quad \rho = 0.00$					$df = \infty, \quad \rho = 0.10$				
Δ	voimakkuus				Δ	voimakkuus			
	t	w	t-cl	w-cl		t	w	t-cl	w-cl
0.00	0.054	0.053	0.055	0.057	0.00	0.038	0.038	0.048	0.050
0.02	0.057	0.055	0.058	0.056	0.02	0.054	0.056	0.062	0.066
0.04	0.101	0.094	0.099	0.098	0.04	0.086	0.089	0.097	0.100
0.06	0.165	0.162	0.158	0.163	0.06	0.147	0.139	0.165	0.158
0.08	0.235	0.208	0.221	0.213	0.08	0.223	0.218	0.256	0.231
0.10	0.329	0.327	0.318	0.314	0.10	0.361	0.358	0.392	0.382
0.12	0.459	0.457	0.450	0.441	0.12	0.454	0.431	0.476	0.466
0.14	0.618	0.601	0.593	0.589	0.14	0.591	0.580	0.615	0.602
0.16	0.707	0.689	0.678	0.670	0.16	0.707	0.681	0.723	0.695
0.18	0.807	0.781	0.781	0.769	0.18	0.862	0.839	0.868	0.850
0.20	0.879	0.862	0.855	0.837	0.20	0.890	0.876	0.894	0.888

Malli M(I , norm): $X_{ij} - \Delta G_{ij} = \tau_i + \epsilon_{ij} \sim N(0, 1)$

$df = \infty, \quad \rho = 0.20$				
Δ	voimakkuus			
	t	w	t-cl	w-cl
0.00	0.032	0.027	0.051	0.053
0.02	0.040	0.037	0.066	0.062
0.04	0.063	0.063	0.090	0.086
0.06	0.132	0.134	0.180	0.175
0.08	0.221	0.219	0.275	0.265
0.10	0.352	0.341	0.427	0.406
0.12	0.477	0.461	0.535	0.535
0.14	0.589	0.564	0.668	0.630
0.16	0.736	0.720	0.775	0.770
0.18	0.836	0.815	0.870	0.849
0.20	0.907	0.897	0.935	0.922

$df = \infty, \quad \rho = 0.30$				
Δ	voimakkuus			
	t	w	t-cl	w-cl
0.00	0.014	0.020	0.047	0.050
0.02	0.031	0.038	0.062	0.060
0.04	0.067	0.068	0.123	0.115
0.06	0.112	0.112	0.184	0.194
0.08	0.210	0.209	0.310	0.299
0.10	0.315	0.292	0.439	0.408
0.12	0.482	0.448	0.613	0.577
0.14	0.613	0.599	0.732	0.693
0.16	0.745	0.723	0.837	0.816
0.18	0.878	0.867	0.933	0.907
0.20	0.933	0.925	0.963	0.953

$df = \infty, \quad \rho = 0.40$				
Δ	voimakkuus			
	t	w	t-cl	w-cl
0.00	0.007	0.008	0.036	0.031
0.02	0.019	0.020	0.070	0.080
0.04	0.052	0.051	0.142	0.144
0.06	0.100	0.104	0.230	0.218
0.08	0.191	0.182	0.376	0.356
0.10	0.326	0.319	0.527	0.498
0.12	0.454	0.441	0.656	0.630
0.14	0.629	0.599	0.800	0.769
0.16	0.762	0.749	0.879	0.866
0.18	0.853	0.833	0.936	0.918
0.20	0.926	0.917	0.977	0.965

$df = \infty, \quad \rho = 0.50$				
Δ	voimakkuus			
	t	w	t-cl	w-cl
0.00	0.010	0.009	0.053	0.062
0.02	0.017	0.020	0.076	0.086
0.04	0.027	0.028	0.117	0.119
0.06	0.076	0.066	0.260	0.223
0.08	0.193	0.191	0.435	0.417
0.10	0.330	0.323	0.619	0.587
0.12	0.443	0.439	0.739	0.698
0.14	0.653	0.618	0.878	0.847
0.16	0.824	0.797	0.944	0.931
0.18	0.900	0.887	0.978	0.974
0.20	0.952	0.937	0.990	0.983

$df = \infty, \quad \rho = 0.60$				
Δ	voimakkuus			
	t	w	t-cl	w-cl
0.00	0.002	0.005	0.051	0.045
0.02	0.006	0.005	0.072	0.077
0.04	0.024	0.027	0.182	0.174
0.06	0.061	0.060	0.299	0.269
0.08	0.159	0.149	0.521	0.493
0.10	0.292	0.272	0.692	0.663
0.12	0.490	0.442	0.845	0.820
0.14	0.699	0.665	0.926	0.912
0.16	0.812	0.784	0.974	0.955
0.18	0.916	0.893	0.993	0.986
0.20	0.972	0.962	0.998	0.993

$df = \infty, \quad \rho = 0.70$				
Δ	voimakkuus			
	t	w	t-cl	w-cl
0.00	0.000	0.000	0.045	0.048
0.02	0.000	0.001	0.084	0.079
0.04	0.009	0.010	0.213	0.191
0.06	0.039	0.042	0.387	0.364
0.08	0.098	0.095	0.596	0.540
0.10	0.260	0.241	0.797	0.735
0.12	0.484	0.447	0.937	0.901
0.14	0.697	0.663	0.982	0.964
0.16	0.851	0.831	0.990	0.987
0.18	0.953	0.933	0.998	0.998
0.20	0.981	0.977	1.000	1.000

$$\text{Malli } M(I, t(3)): X_{ij} - \Delta G_{ij} = \frac{\tau_i + \epsilon_{ij}}{\sqrt{\chi_3^2/3}} \sim t(3)$$

$df = 3, \quad \rho = 0.00$				
Δ	voimakkuus			
	t	w	t-cl	w-cl
0.00	0.045	0.051	0.047	0.049
0.02	0.054	0.059	0.054	0.057
0.04	0.063	0.078	0.062	0.068
0.06	0.092	0.115	0.091	0.114
0.08	0.125	0.173	0.134	0.169
0.10	0.160	0.233	0.170	0.236
0.12	0.237	0.351	0.237	0.335
0.14	0.276	0.438	0.288	0.411
0.16	0.349	0.516	0.347	0.504
0.18	0.402	0.614	0.419	0.582
0.20	0.478	0.691	0.506	0.671

$df = 3, \quad \rho = 0.10$				
Δ	voimakkuus			
	t	w	t-cl	w-cl
0.00	0.051	0.046	0.053	0.058
0.02	0.041	0.053	0.052	0.064
0.04	0.051	0.059	0.063	0.084
0.06	0.091	0.109	0.097	0.113
0.08	0.113	0.160	0.140	0.185
0.10	0.172	0.245	0.201	0.263
0.12	0.209	0.305	0.252	0.321
0.14	0.289	0.438	0.330	0.464
0.16	0.363	0.511	0.419	0.545
0.18	0.409	0.638	0.467	0.639
0.20	0.520	0.720	0.557	0.734

$df = 3, \quad \rho = 0.20$				
Δ	voimakkuus			
	t	w	t-cl	w-cl
0.00	0.029	0.036	0.035	0.049
0.02	0.033	0.041	0.044	0.056
0.04	0.043	0.069	0.072	0.092
0.06	0.055	0.078	0.091	0.118
0.08	0.109	0.157	0.150	0.203
0.10	0.139	0.224	0.210	0.292
0.12	0.185	0.291	0.244	0.348
0.14	0.251	0.382	0.336	0.444
0.16	0.333	0.520	0.444	0.575
0.18	0.420	0.621	0.507	0.667
0.20	0.496	0.739	0.577	0.784

$df = 3, \quad \rho = 0.30$				
Δ	voimakkuus			
	t	w	t-cl	w-cl
0.00	0.015	0.019	0.036	0.049
0.02	0.030	0.032	0.051	0.065
0.04	0.022	0.041	0.068	0.086
0.06	0.066	0.079	0.122	0.143
0.08	0.089	0.133	0.165	0.212
0.10	0.141	0.210	0.228	0.319
0.12	0.185	0.296	0.321	0.410
0.14	0.258	0.405	0.393	0.527
0.16	0.347	0.539	0.480	0.660
0.18	0.420	0.653	0.563	0.747
0.20	0.532	0.744	0.656	0.807

$df = 3, \quad \rho = 0.40$				
Δ	voimakkuus			
	t	w	t-cl	w-cl
0.00	0.015	0.015	0.043	0.043
0.02	0.013	0.020	0.046	0.063
0.04	0.030	0.037	0.080	0.090
0.06	0.035	0.050	0.098	0.129
0.08	0.066	0.116	0.176	0.225
0.10	0.131	0.199	0.266	0.344
0.12	0.159	0.286	0.347	0.457
0.14	0.236	0.407	0.439	0.586
0.16	0.308	0.514	0.535	0.684
0.18	0.431	0.651	0.636	0.794
0.20	0.502	0.723	0.691	0.849

$df = 3, \quad \rho = 0.50$				
Δ	voimakkuus			
	t	w	t-cl	w-cl
0.00	0.013	0.008	0.050	0.057
0.02	0.014	0.013	0.048	0.060
0.04	0.019	0.026	0.103	0.114
0.06	0.030	0.046	0.122	0.163
0.08	0.058	0.116	0.229	0.284
0.10	0.079	0.163	0.286	0.379
0.12	0.176	0.295	0.420	0.553
0.14	0.251	0.418	0.514	0.654
0.16	0.287	0.515	0.568	0.737
0.18	0.407	0.651	0.678	0.842
0.20	0.529	0.803	0.789	0.923

$$\text{Malli } M(I, t(3)): X_{ij} - \Delta G_{ij} = \frac{\tau_i + \epsilon_{ij}}{\sqrt{\chi_3^2/3}} \sim t(3)$$

$df = 3, \quad \rho = 0.60$					$df = 3, \quad \rho = 0.70$				
Δ	voimakkuus				Δ	voimakkuus			
	t	w	t-cl	w-cl		t	w	t-cl	w-cl
0.00	0.002	0.002	0.045	0.050	0.00	0.000	0.001	0.036	0.042
0.02	0.005	0.007	0.053	0.071	0.02	0.000	0.003	0.067	0.074
0.04	0.004	0.014	0.106	0.117	0.04	0.002	0.007	0.118	0.140
0.06	0.023	0.045	0.163	0.222	0.06	0.013	0.020	0.215	0.268
0.08	0.041	0.071	0.244	0.297	0.08	0.036	0.075	0.325	0.403
0.10	0.083	0.165	0.370	0.486	0.10	0.063	0.128	0.478	0.595
0.12	0.140	0.245	0.477	0.614	0.12	0.127	0.259	0.596	0.711
0.14	0.211	0.384	0.610	0.757	0.14	0.200	0.387	0.687	0.816
0.16	0.330	0.554	0.705	0.853	0.16	0.295	0.538	0.802	0.911
0.18	0.420	0.678	0.775	0.904	0.18	0.412	0.696	0.856	0.959
0.20	0.531	0.786	0.837	0.952	0.20	0.506	0.818	0.928	0.985

$$\text{Malli } M(I, t(10)): X_{ij} - \Delta G_{ij} = \frac{\tau_i + \epsilon_{ij}}{\sqrt{\chi_{10}^2/10}} \sim t(10)$$

$df = 10, \quad \rho = 0.00$					$df = 10, \quad \rho = 0.10$				
Δ	voimakkuus				Δ	voimakkuus			
	t	w	t-cl	w-cl		t	w	t-cl	w-cl
0.00	0.045	0.046	0.044	0.043	0.00	0.046	0.049	0.056	0.058
0.02	0.069	0.068	0.064	0.060	0.02	0.049	0.049	0.067	0.058
0.04	0.079	0.084	0.083	0.083	0.04	0.072	0.082	0.089	0.093
0.06	0.145	0.149	0.124	0.146	0.06	0.133	0.140	0.165	0.162
0.08	0.222	0.241	0.212	0.215	0.08	0.183	0.186	0.206	0.222
0.10	0.282	0.286	0.267	0.277	0.10	0.294	0.303	0.307	0.312
0.12	0.404	0.399	0.382	0.390	0.12	0.379	0.392	0.410	0.424
0.14	0.523	0.551	0.506	0.519	0.14	0.525	0.525	0.558	0.557
0.16	0.604	0.620	0.589	0.600	0.16	0.632	0.654	0.642	0.659
0.18	0.700	0.721	0.685	0.695	0.18	0.735	0.759	0.750	0.764
0.20	0.807	0.819	0.793	0.799	0.20	0.832	0.852	0.836	0.850

$$\text{Malli } M(I, t(10)): X_{ij} - \Delta G_{ij} = \frac{\tau_i + \epsilon_{ij}}{\sqrt{\chi_{10}^2/10}} \sim t(10)$$

$df = 10, \quad \rho = 0.20$				
Δ	voimakkuus			
	t	w	t-cl	w-cl
0.00	0.031	0.032	0.055	0.054
0.02	0.037	0.042	0.071	0.064
0.04	0.073	0.065	0.096	0.088
0.06	0.113	0.115	0.150	0.149
0.08	0.178	0.187	0.234	0.244
0.10	0.280	0.296	0.358	0.359
0.12	0.402	0.420	0.486	0.494
0.14	0.496	0.520	0.571	0.588
0.16	0.648	0.670	0.696	0.704
0.18	0.743	0.768	0.791	0.814
0.20	0.810	0.835	0.855	0.869

$df = 10, \quad \rho = 0.30$				
Δ	voimakkuus			
	t	w	t-cl	w-cl
0.00	0.016	0.019	0.045	0.056
0.02	0.035	0.037	0.059	0.061
0.04	0.055	0.059	0.106	0.114
0.06	0.105	0.125	0.168	0.179
0.08	0.154	0.152	0.266	0.259
0.10	0.282	0.291	0.384	0.396
0.12	0.358	0.377	0.490	0.497
0.14	0.518	0.541	0.634	0.652
0.16	0.640	0.665	0.774	0.766
0.18	0.748	0.776	0.850	0.855
0.20	0.842	0.848	0.902	0.901

$df = 10, \quad \rho = 0.40$				
Δ	voimakkuus			
	t	w	t-cl	w-cl
0.00	0.015	0.012	0.043	0.052
0.02	0.014	0.017	0.059	0.057
0.04	0.031	0.043	0.100	0.115
0.06	0.086	0.091	0.198	0.198
0.08	0.147	0.166	0.296	0.300
0.10	0.253	0.262	0.443	0.440
0.12	0.386	0.409	0.593	0.604
0.14	0.531	0.547	0.704	0.709
0.16	0.649	0.680	0.816	0.809
0.18	0.804	0.810	0.908	0.906
0.20	0.868	0.872	0.935	0.935

$df = 10, \quad \rho = 0.50$				
Δ	voimakkuus			
	t	w	t-cl	w-cl
0.00	0.006	0.006	0.046	0.046
0.02	0.014	0.010	0.061	0.056
0.04	0.025	0.030	0.122	0.119
0.06	0.064	0.067	0.225	0.208
0.08	0.118	0.139	0.339	0.323
0.10	0.229	0.249	0.493	0.490
0.12	0.355	0.394	0.641	0.633
0.14	0.518	0.541	0.792	0.796
0.16	0.658	0.683	0.867	0.880
0.18	0.793	0.800	0.927	0.926
0.20	0.889	0.898	0.979	0.978

$df = 10, \quad \rho = 0.60$				
Δ	voimakkuus			
	t	w	t-cl	w-cl
0.00	0.002	0.003	0.043	0.050
0.02	0.005	0.006	0.060	0.061
0.04	0.020	0.016	0.149	0.135
0.06	0.055	0.064	0.272	0.263
0.08	0.111	0.126	0.425	0.425
0.10	0.213	0.228	0.580	0.592
0.12	0.371	0.409	0.772	0.765
0.14	0.509	0.547	0.868	0.872
0.16	0.682	0.692	0.932	0.930
0.18	0.811	0.829	0.975	0.978
0.20	0.908	0.914	0.986	0.990

$df = 10, \quad \rho = 0.70$				
Δ	voimakkuus			
	t	w	t-cl	w-cl
0.00	0.002	0.002	0.042	0.041
0.02	0.002	0.005	0.095	0.087
0.04	0.010	0.018	0.181	0.153
0.06	0.027	0.028	0.306	0.292
0.08	0.077	0.100	0.536	0.514
0.10	0.173	0.195	0.671	0.666
0.12	0.325	0.370	0.865	0.844
0.14	0.526	0.570	0.945	0.934
0.16	0.734	0.745	0.978	0.977
0.18	0.864	0.886	0.995	0.995
0.20	0.929	0.947	0.999	0.999

C.2 Klusterirakennetyyppi II

Malli M(II, norm): $X_{ij} - \Delta G_{ij} = \tau_i + \epsilon_{ij} \sim N(0, 1)$

$df = \infty, \quad \rho = 0.00$					$df = \infty, \quad \rho = 0.02$				
Δ	voimakkuus				Δ	voimakkuus			
	t	w	t-cl	w-cl		t	w	t-cl	w-cl
0.00	0.054	0.051	0.054	0.053	0.00	0.081	0.077	0.038	0.043
0.02	0.052	0.056	0.050	0.052	0.02	0.120	0.111	0.066	0.063
0.04	0.099	0.092	0.097	0.091	0.04	0.154	0.153	0.088	0.086
0.06	0.156	0.158	0.161	0.155	0.06	0.188	0.187	0.119	0.121
0.08	0.251	0.235	0.248	0.228	0.08	0.265	0.255	0.183	0.175
0.10	0.342	0.338	0.335	0.331	0.10	0.379	0.369	0.283	0.274
0.12	0.487	0.467	0.478	0.459	0.12	0.504	0.484	0.397	0.376
0.14	0.586	0.560	0.562	0.549	0.14	0.572	0.539	0.445	0.436
0.16	0.728	0.695	0.710	0.679	0.16	0.682	0.661	0.578	0.552
0.18	0.811	0.794	0.795	0.776	0.18	0.759	0.745	0.654	0.636
0.20	0.888	0.866	0.872	0.863	0.20	0.860	0.847	0.755	0.746
0.22	0.932	0.921	0.923	0.913	0.22	0.907	0.884	0.817	0.811
0.24	0.968	0.962	0.963	0.960	0.24	0.937	0.936	0.887	0.877
0.26	0.985	0.980	0.982	0.976	0.26	0.965	0.960	0.924	0.922
0.28	0.997	0.996	0.997	0.995	0.28	0.980	0.974	0.957	0.947
0.30	0.998	0.998	0.998	0.998	0.30	0.993	0.993	0.988	0.986

Malli M(II , norm): $X_{ij} - \Delta G_{ij} = \tau_i + \epsilon_{ij} \sim N(0, 1)$

$df = \infty, \quad \rho = 0.04$				
Δ	voimakkuus			
	t	w	t-cl	w-cl
0.00	0.144	0.150	0.056	0.056
0.02	0.143	0.154	0.059	0.055
0.04	0.177	0.164	0.074	0.077
0.06	0.241	0.228	0.111	0.111
0.08	0.325	0.307	0.173	0.165
0.10	0.372	0.361	0.204	0.197
0.12	0.488	0.471	0.295	0.290
0.14	0.587	0.575	0.389	0.379
0.16	0.671	0.657	0.480	0.465
0.18	0.740	0.724	0.553	0.541
0.20	0.823	0.812	0.678	0.668
0.22	0.878	0.863	0.720	0.708
0.24	0.926	0.918	0.808	0.805
0.26	0.950	0.940	0.859	0.851
0.28	0.965	0.964	0.903	0.903
0.30	0.981	0.977	0.945	0.939

$df = \infty, \quad \rho = 0.06$				
Δ	voimakkuus			
	t	w	t-cl	w-cl
0.00	0.170	0.172	0.048	0.050
0.02	0.194	0.185	0.056	0.053
0.04	0.222	0.214	0.065	0.065
0.06	0.271	0.279	0.107	0.101
0.08	0.325	0.318	0.140	0.145
0.10	0.386	0.366	0.186	0.183
0.12	0.481	0.467	0.238	0.244
0.14	0.533	0.534	0.284	0.295
0.16	0.657	0.650	0.390	0.379
0.18	0.731	0.723	0.480	0.479
0.20	0.818	0.798	0.586	0.572
0.22	0.846	0.828	0.649	0.647
0.24	0.904	0.902	0.733	0.715
0.26	0.926	0.918	0.753	0.743
0.28	0.969	0.964	0.857	0.851
0.30	0.959	0.955	0.860	0.858

$df = \infty, \quad \rho = 0.10$				
Δ	voimakkuus			
	t	w	t-cl	w-cl
0.00	0.243	0.236	0.054	0.049
0.02	0.258	0.250	0.050	0.055
0.04	0.279	0.267	0.054	0.054
0.06	0.322	0.304	0.098	0.097
0.08	0.364	0.363	0.118	0.122
0.10	0.437	0.429	0.144	0.140
0.12	0.475	0.468	0.200	0.198
0.14	0.570	0.545	0.252	0.238
0.16	0.622	0.620	0.298	0.297
0.18	0.710	0.695	0.394	0.402
0.20	0.779	0.762	0.442	0.440
0.22	0.810	0.805	0.525	0.515
0.24	0.843	0.832	0.590	0.583
0.26	0.892	0.889	0.636	0.647
0.28	0.920	0.917	0.717	0.708
0.30	0.957	0.952	0.788	0.792

$df = \infty, \quad \rho = 0.30$				
Δ	voimakkuus			
	t	w	t-cl	w-cl
0.00	0.435	0.436	0.060	0.059
0.02	0.467	0.449	0.051	0.050
0.04	0.451	0.430	0.054	0.054
0.06	0.467	0.461	0.071	0.071
0.08	0.510	0.505	0.078	0.082
0.10	0.529	0.516	0.098	0.094
0.12	0.556	0.547	0.120	0.122
0.14	0.584	0.576	0.132	0.134
0.16	0.636	0.631	0.183	0.174
0.18	0.669	0.667	0.206	0.218
0.20	0.708	0.699	0.234	0.229
0.22	0.728	0.730	0.244	0.246
0.24	0.778	0.766	0.313	0.318
0.26	0.788	0.776	0.335	0.335
0.28	0.840	0.827	0.411	0.411
0.30	0.863	0.866	0.429	0.422

Malli M(II , norm): $X_{ij} - \Delta G_{ij} = \tau_i + \epsilon_{ij} \sim N(0, 1)$

$df = \infty, \quad \rho = 0.50$				
Δ	voimakkuus			
	t	w	t-cl	w-cl
0.00	0.559	0.559	0.054	0.058
0.02	0.563	0.544	0.053	0.053
0.04	0.555	0.564	0.056	0.056
0.06	0.560	0.563	0.052	0.051
0.08	0.578	0.568	0.062	0.064
0.10	0.592	0.588	0.074	0.077
0.12	0.625	0.620	0.097	0.095
0.14	0.622	0.616	0.115	0.114
0.16	0.656	0.658	0.112	0.109
0.18	0.681	0.669	0.143	0.141
0.20	0.673	0.662	0.140	0.148
0.22	0.719	0.708	0.201	0.203
0.24	0.749	0.739	0.206	0.198
0.26	0.790	0.783	0.230	0.227
0.28	0.781	0.773	0.270	0.269
0.30	0.836	0.821	0.329	0.324

$df = \infty, \quad \rho = 0.70$				
Δ	voimakkuus			
	t	w	t-cl	w-cl
0.00	0.610	0.589	0.061	0.060
0.02	0.630	0.619	0.067	0.070
0.04	0.596	0.580	0.052	0.052
0.06	0.619	0.602	0.064	0.062
0.08	0.632	0.628	0.078	0.075
0.10	0.634	0.647	0.075	0.071
0.12	0.641	0.632	0.081	0.083
0.14	0.683	0.667	0.093	0.096
0.16	0.695	0.680	0.112	0.113
0.18	0.691	0.682	0.103	0.114
0.20	0.718	0.718	0.135	0.125
0.22	0.776	0.749	0.162	0.167
0.24	0.755	0.755	0.184	0.189
0.26	0.770	0.763	0.195	0.190
0.28	0.766	0.745	0.200	0.204
0.30	0.796	0.798	0.220	0.226

$df = \infty, \quad \rho = 0.99$				
Δ	voimakkuus			
	t	w	t-cl	w-cl
0.00	0.682	0.664	0.055	0.057
0.05	0.677	0.676	0.046	0.053
0.10	0.709	0.698	0.065	0.071
0.15	0.696	0.695	0.090	0.087
0.20	0.766	0.760	0.115	0.115
0.25	0.774	0.762	0.132	0.123
0.30	0.797	0.790	0.186	0.179
0.35	0.827	0.837	0.238	0.237
0.40	0.861	0.852	0.296	0.288
0.45	0.874	0.866	0.350	0.346
0.50	0.929	0.919	0.428	0.410
0.55	0.935	0.928	0.506	0.497
0.60	0.954	0.952	0.542	0.530
0.65	0.968	0.964	0.618	0.613
0.70	0.974	0.975	0.685	0.667
0.75	0.987	0.991	0.757	0.742

$df = \infty, \quad \rho = 0.99$				
Δ	voimakkuus			
	t	w	t-cl	w-cl
0.80	0.992	0.986	0.778	0.768
0.85	0.996	0.992	0.852	0.833
0.90	0.996	0.996	0.882	0.849
0.95	1.000	0.997	0.891	0.885
1.00	0.999	0.999	0.939	0.927
1.05	1.000	1.000	0.955	0.950
1.10	1.000	1.000	0.975	0.968
1.15	1.000	1.000	0.981	0.973
1.20	1.000	1.000	0.984	0.980
1.25	1.000	1.000	0.991	0.989
1.30	1.000	1.000	0.997	0.995
1.35	1.000	1.000	0.997	0.991
1.40	1.000	1.000	0.997	0.996
1.45	1.000	1.000	1.000	1.000
1.50	1.000	1.000	1.000	1.000

$$\text{Malli } M(II, t(3)): X_{ij} - \Delta G_{ij} = \frac{\tau_i + \epsilon_{ij}}{\sqrt{\chi_3^2/3}} \sim t(3)$$

$df = 3, \quad \rho = 0.00$				
Δ	voimakkuus			
	t	w	t-cl	w-cl
0.00	0.049	0.049	0.047	0.050
0.02	0.061	0.056	0.048	0.060
0.04	0.062	0.077	0.064	0.077
0.06	0.085	0.126	0.083	0.117
0.08	0.145	0.179	0.146	0.178
0.10	0.178	0.264	0.195	0.249
0.12	0.215	0.296	0.242	0.301
0.14	0.301	0.432	0.327	0.430
0.16	0.340	0.511	0.357	0.491
0.18	0.418	0.584	0.446	0.595
0.20	0.515	0.707	0.544	0.691
0.22	0.581	0.761	0.590	0.753
0.24	0.614	0.842	0.636	0.824
0.26	0.698	0.893	0.716	0.875
0.28	0.761	0.923	0.757	0.913
0.30	0.793	0.955	0.809	0.946

$df = 3, \quad \rho = 0.02$				
Δ	voimakkuus			
	t	w	t-cl	w-cl
0.00	0.093	0.083	0.047	0.045
0.02	0.092	0.098	0.041	0.053
0.04	0.101	0.111	0.059	0.070
0.06	0.138	0.155	0.082	0.101
0.08	0.149	0.194	0.103	0.121
0.10	0.211	0.286	0.134	0.209
0.12	0.246	0.362	0.185	0.258
0.14	0.310	0.456	0.252	0.353
0.16	0.407	0.529	0.319	0.434
0.18	0.444	0.598	0.374	0.488
0.20	0.502	0.677	0.420	0.567
0.22	0.557	0.734	0.454	0.623
0.24	0.640	0.818	0.549	0.729
0.26	0.691	0.857	0.595	0.774
0.28	0.759	0.917	0.655	0.847
0.30	0.780	0.935	0.704	0.882

$df = 3, \quad \rho = 0.04$				
Δ	voimakkuus			
	t	w	t-cl	w-cl
0.00	0.131	0.113	0.042	0.040
0.02	0.147	0.142	0.050	0.055
0.04	0.158	0.148	0.061	0.068
0.06	0.186	0.198	0.072	0.086
0.08	0.209	0.248	0.087	0.112
0.10	0.239	0.308	0.126	0.171
0.12	0.280	0.346	0.144	0.189
0.14	0.348	0.450	0.194	0.262
0.16	0.398	0.512	0.254	0.334
0.18	0.466	0.618	0.322	0.446
0.20	0.508	0.662	0.366	0.490
0.22	0.562	0.737	0.393	0.533
0.24	0.592	0.778	0.445	0.614
0.26	0.686	0.842	0.516	0.707
0.28	0.710	0.873	0.558	0.744
0.30	0.771	0.922	0.611	0.809

$df = 3, \quad \rho = 0.06$				
Δ	voimakkuus			
	t	w	t-cl	w-cl
0.00	0.180	0.160	0.042	0.046
0.02	0.202	0.179	0.052	0.064
0.04	0.210	0.199	0.068	0.061
0.06	0.234	0.250	0.076	0.099
0.08	0.230	0.267	0.079	0.100
0.10	0.281	0.324	0.116	0.143
0.12	0.336	0.407	0.135	0.200
0.14	0.359	0.472	0.162	0.243
0.16	0.403	0.511	0.190	0.287
0.18	0.468	0.574	0.270	0.343
0.20	0.502	0.651	0.302	0.423
0.22	0.562	0.711	0.326	0.463
0.24	0.612	0.760	0.408	0.541
0.26	0.652	0.805	0.430	0.605
0.28	0.681	0.854	0.468	0.655
0.30	0.751	0.885	0.556	0.732

$$\text{Malli } M(II, t(3)): X_{ij} - \Delta G_{ij} = \frac{\tau_i + \epsilon_{ij}}{\sqrt{\chi_3^2/3}} \sim t(3)$$

$df = 3, \quad \rho = 0.10$				
Δ	voimakkuus			
	t	w	t-cl	w-cl
0.00	0.244	0.214	0.048	0.047
0.02	0.270	0.237	0.045	0.056
0.04	0.258	0.265	0.052	0.065
0.06	0.311	0.297	0.070	0.086
0.08	0.300	0.333	0.077	0.090
0.10	0.325	0.367	0.087	0.126
0.12	0.370	0.425	0.113	0.140
0.14	0.395	0.458	0.155	0.188
0.16	0.441	0.524	0.177	0.234
0.18	0.476	0.591	0.205	0.274
0.20	0.485	0.613	0.206	0.296
0.22	0.550	0.692	0.272	0.380
0.24	0.598	0.739	0.305	0.439
0.26	0.614	0.764	0.346	0.487
0.28	0.695	0.820	0.412	0.554
0.30	0.715	0.857	0.448	0.611

$df = 3, \quad \rho = 0.30$				
Δ	voimakkuus			
	t	w	t-cl	w-cl
0.00	0.473	0.454	0.055	0.056
0.02	0.469	0.455	0.053	0.059
0.04	0.459	0.437	0.047	0.050
0.06	0.487	0.470	0.063	0.074
0.08	0.479	0.474	0.051	0.055
0.10	0.499	0.515	0.076	0.090
0.12	0.506	0.536	0.074	0.091
0.14	0.507	0.542	0.079	0.108
0.16	0.530	0.549	0.102	0.124
0.18	0.541	0.592	0.119	0.158
0.20	0.557	0.594	0.106	0.147
0.22	0.601	0.666	0.132	0.188
0.24	0.570	0.646	0.154	0.218
0.26	0.628	0.713	0.162	0.237
0.28	0.670	0.749	0.204	0.284
0.30	0.704	0.782	0.252	0.332

$df = 3, \quad \rho = 0.50$				
Δ	voimakkuus			
	t	w	t-cl	w-cl
0.00	0.547	0.548	0.044	0.045
0.02	0.544	0.524	0.049	0.057
0.04	0.583	0.561	0.057	0.063
0.06	0.567	0.544	0.059	0.061
0.08	0.557	0.559	0.042	0.056
0.10	0.565	0.559	0.059	0.073
0.12	0.548	0.552	0.070	0.080
0.14	0.592	0.616	0.073	0.081
0.16	0.561	0.602	0.077	0.088
0.18	0.626	0.634	0.082	0.117
0.20	0.630	0.636	0.101	0.131
0.22	0.644	0.692	0.097	0.148
0.24	0.657	0.696	0.113	0.144
0.26	0.681	0.724	0.137	0.209
0.28	0.674	0.715	0.142	0.188
0.30	0.683	0.735	0.163	0.214

$df = 3, \quad \rho = 0.70$				
Δ	voimakkuus			
	t	w	t-cl	w-cl
0.00	0.618	0.604	0.040	0.050
0.02	0.645	0.611	0.054	0.063
0.04	0.588	0.584	0.045	0.046
0.06	0.590	0.595	0.051	0.051
0.08	0.625	0.624	0.051	0.051
0.10	0.628	0.630	0.056	0.071
0.12	0.616	0.610	0.046	0.066
0.14	0.631	0.633	0.074	0.079
0.16	0.643	0.649	0.070	0.097
0.18	0.639	0.648	0.060	0.071
0.20	0.677	0.699	0.091	0.112
0.22	0.670	0.691	0.081	0.105
0.24	0.655	0.670	0.091	0.117
0.26	0.687	0.732	0.110	0.144
0.28	0.677	0.741	0.115	0.161
0.30	0.700	0.756	0.130	0.183

$$\text{Malli } M(II, t(3)): X_{ij} - \Delta G_{ij} = \frac{\tau_i + \epsilon_{ij}}{\sqrt{\chi_3^2/3}} \sim t(3)$$

$df = 3, \quad \rho = 0.99$				
Δ	voimakkuus			
	t	w	t-cl	w-cl
0.00	0.681	0.655	0.049	0.058
0.05	0.658	0.666	0.045	0.047
0.10	0.683	0.663	0.056	0.063
0.15	0.688	0.679	0.062	0.073
0.20	0.686	0.708	0.070	0.085
0.25	0.738	0.736	0.112	0.125
0.30	0.766	0.757	0.101	0.135
0.35	0.751	0.779	0.135	0.160
0.40	0.783	0.810	0.151	0.216
0.45	0.816	0.832	0.182	0.230
0.50	0.794	0.841	0.202	0.280
0.55	0.824	0.871	0.263	0.344
0.60	0.840	0.906	0.284	0.381
0.65	0.860	0.923	0.319	0.455
0.70	0.901	0.948	0.370	0.481
0.75	0.914	0.961	0.424	0.528

$df = 3, \quad \rho = 0.99$				
Δ	voimakkuus			
	t	w	t-cl	w-cl
0.80	0.917	0.968	0.410	0.560
0.85	0.933	0.971	0.513	0.672
0.90	0.950	0.985	0.499	0.679
0.95	0.948	0.985	0.571	0.733
1.00	0.958	0.991	0.605	0.768
1.05	0.964	0.994	0.647	0.824
1.10	0.971	0.997	0.691	0.855
1.15	0.967	0.998	0.730	0.874
1.20	0.971	0.996	0.733	0.892
1.25	0.978	0.998	0.774	0.919
1.30	0.990	1.000	0.831	0.935
1.35	0.981	0.998	0.822	0.943
1.40	0.990	1.000	0.843	0.961
1.45	0.991	1.000	0.866	0.972
1.50	0.990	1.000	0.887	0.974

$$\text{Malli } M(II, t(10)): X_{ij} - \Delta G_{ij} = \frac{\tau_i + \epsilon_{ij}}{\sqrt{\chi_{10}^2/10}} \sim t(10)$$

$df = 10, \quad \rho = 0.00$				
Δ	voimakkuus			
	t	w	t-cl	w-cl
0.00	0.053	0.055	0.060	0.059
0.02	0.058	0.057	0.060	0.056
0.04	0.097	0.100	0.103	0.103
0.06	0.149	0.145	0.149	0.139
0.08	0.170	0.184	0.187	0.209
0.10	0.302	0.321	0.309	0.320
0.12	0.416	0.437	0.412	0.422
0.14	0.542	0.542	0.516	0.526
0.16	0.595	0.619	0.583	0.611
0.18	0.727	0.767	0.719	0.752
0.20	0.808	0.837	0.800	0.826
0.22	0.879	0.882	0.865	0.871
0.24	0.923	0.926	0.909	0.918
0.26	0.960	0.967	0.948	0.957
0.28	0.984	0.981	0.976	0.978
0.30	0.987	0.990	0.986	0.988

$df = 10, \quad \rho = 0.02$				
Δ	voimakkuus			
	t	w	t-cl	w-cl
0.00	0.092	0.097	0.043	0.051
0.02	0.091	0.088	0.049	0.048
0.04	0.128	0.138	0.083	0.081
0.06	0.164	0.170	0.101	0.098
0.08	0.242	0.262	0.155	0.169
0.10	0.320	0.332	0.228	0.244
0.12	0.425	0.434	0.316	0.317
0.14	0.497	0.499	0.365	0.375
0.16	0.592	0.597	0.458	0.469
0.18	0.704	0.733	0.596	0.632
0.20	0.777	0.806	0.690	0.717
0.22	0.832	0.844	0.735	0.766
0.24	0.889	0.897	0.805	0.831
0.26	0.917	0.936	0.867	0.883
0.28	0.951	0.963	0.901	0.921
0.30	0.973	0.979	0.940	0.950

$$\text{Malli } M(II, t(10)): X_{ij} - \Delta G_{ij} = \frac{\tau_i + \epsilon_{ij}}{\sqrt{\chi_{10}^2/10}} \sim t(10)$$

$df = 10, \quad \rho = 0.04$				
Δ	voimakkuus			
	t	w	t-cl	w-cl
0.00	0.137	0.134	0.049	0.048
0.02	0.132	0.127	0.046	0.059
0.04	0.179	0.169	0.071	0.062
0.06	0.219	0.209	0.107	0.104
0.08	0.266	0.264	0.132	0.131
0.10	0.340	0.356	0.164	0.180
0.12	0.402	0.419	0.236	0.247
0.14	0.509	0.523	0.319	0.338
0.16	0.595	0.621	0.413	0.438
0.18	0.685	0.709	0.488	0.504
0.20	0.757	0.777	0.580	0.609
0.22	0.807	0.825	0.665	0.677
0.24	0.898	0.902	0.733	0.771
0.26	0.896	0.903	0.748	0.792
0.28	0.924	0.939	0.839	0.866
0.30	0.956	0.961	0.881	0.911

$df = 10, \quad \rho = 0.06$				
Δ	voimakkuus			
	t	w	t-cl	w-cl
0.00	0.179	0.166	0.057	0.061
0.02	0.208	0.180	0.061	0.065
0.04	0.179	0.183	0.073	0.067
0.06	0.243	0.245	0.083	0.095
0.08	0.303	0.310	0.129	0.130
0.10	0.397	0.382	0.174	0.181
0.12	0.451	0.463	0.222	0.238
0.14	0.499	0.512	0.274	0.291
0.16	0.573	0.586	0.319	0.343
0.18	0.643	0.666	0.416	0.429
0.20	0.714	0.738	0.478	0.510
0.22	0.786	0.814	0.545	0.584
0.24	0.839	0.849	0.637	0.663
0.26	0.890	0.904	0.699	0.735
0.28	0.916	0.926	0.781	0.812
0.30	0.933	0.944	0.801	0.830

$df = 10, \quad \rho = 0.10$				
Δ	voimakkuus			
	t	w	t-cl	w-cl
0.00	0.247	0.228	0.054	0.053
0.02	0.272	0.256	0.047	0.048
0.04	0.259	0.244	0.060	0.058
0.06	0.317	0.321	0.070	0.077
0.08	0.346	0.356	0.094	0.106
0.10	0.376	0.389	0.124	0.137
0.12	0.434	0.447	0.133	0.143
0.14	0.524	0.537	0.229	0.246
0.16	0.565	0.570	0.255	0.262
0.18	0.625	0.653	0.306	0.325
0.20	0.713	0.722	0.393	0.419
0.22	0.756	0.765	0.448	0.476
0.24	0.803	0.829	0.494	0.521
0.26	0.832	0.855	0.525	0.570
0.28	0.874	0.895	0.608	0.648
0.30	0.924	0.940	0.695	0.721

$df = 10, \quad \rho = 0.30$				
Δ	voimakkuus			
	t	w	t-cl	w-cl
0.00	0.449	0.443	0.050	0.046
0.02	0.457	0.443	0.049	0.060
0.04	0.485	0.466	0.061	0.059
0.06	0.466	0.450	0.056	0.064
0.08	0.501	0.500	0.063	0.068
0.10	0.520	0.516	0.090	0.088
0.12	0.527	0.521	0.094	0.099
0.14	0.549	0.559	0.105	0.114
0.16	0.607	0.603	0.140	0.148
0.18	0.658	0.667	0.157	0.174
0.20	0.672	0.667	0.199	0.214
0.22	0.702	0.706	0.214	0.233
0.24	0.727	0.742	0.244	0.272
0.26	0.771	0.778	0.306	0.325
0.28	0.785	0.795	0.316	0.341
0.30	0.825	0.832	0.370	0.388

$$\text{Malli } M(II, t(10)): X_{ij} - \Delta G_{ij} = \frac{\tau_i + \epsilon_{ij}}{\sqrt{\chi_{10}^2/10}} \sim t(10)$$

$df = 10, \quad \rho = 0.50$				
Δ	voimakkuus			
	t	w	t-cl	w-cl
0.00	0.536	0.529	0.041	0.038
0.02	0.543	0.542	0.041	0.043
0.04	0.534	0.529	0.053	0.052
0.06	0.543	0.539	0.058	0.067
0.08	0.569	0.568	0.065	0.070
0.10	0.626	0.618	0.077	0.074
0.12	0.596	0.591	0.078	0.085
0.14	0.618	0.624	0.103	0.112
0.16	0.642	0.640	0.117	0.124
0.18	0.645	0.646	0.136	0.140
0.20	0.658	0.653	0.134	0.139
0.22	0.699	0.705	0.156	0.168
0.24	0.687	0.701	0.164	0.176
0.26	0.745	0.752	0.213	0.231
0.28	0.759	0.764	0.231	0.249
0.30	0.805	0.810	0.275	0.296

$df = 10, \quad \rho = 0.70$				
Δ	voimakkuus			
	t	w	t-cl	w-cl
0.00	0.619	0.606	0.049	0.045
0.02	0.618	0.589	0.059	0.060
0.04	0.639	0.616	0.062	0.061
0.06	0.604	0.604	0.049	0.058
0.08	0.640	0.644	0.073	0.074
0.10	0.628	0.616	0.059	0.058
0.12	0.617	0.615	0.078	0.077
0.14	0.679	0.675	0.087	0.091
0.16	0.641	0.640	0.095	0.095
0.18	0.698	0.686	0.110	0.117
0.20	0.705	0.687	0.122	0.126
0.22	0.725	0.726	0.127	0.124
0.24	0.717	0.710	0.162	0.170
0.26	0.763	0.763	0.173	0.188
0.28	0.747	0.762	0.176	0.183
0.30	0.774	0.790	0.214	0.226

$df = 10, \quad \rho = 0.99$				
Δ	voimakkuus			
	t	w	t-cl	w-cl
0.00	0.659	0.660	0.050	0.046
0.05	0.677	0.676	0.056	0.048
0.10	0.680	0.694	0.076	0.076
0.15	0.740	0.718	0.076	0.077
0.20	0.733	0.721	0.099	0.098
0.25	0.768	0.763	0.121	0.128
0.30	0.775	0.772	0.143	0.144
0.35	0.823	0.829	0.214	0.234
0.40	0.837	0.832	0.251	0.254
0.45	0.869	0.874	0.315	0.311
0.50	0.902	0.908	0.369	0.379
0.55	0.910	0.920	0.417	0.429
0.60	0.928	0.941	0.472	0.492
0.65	0.958	0.960	0.523	0.520
0.70	0.970	0.967	0.594	0.609
0.75	0.970	0.974	0.675	0.683

$df = 10, \quad \rho = 0.99$				
Δ	voimakkuus			
	t	w	t-cl	w-cl
0.80	0.978	0.986	0.703	0.731
0.85	0.990	0.993	0.749	0.769
0.90	0.989	0.990	0.816	0.825
0.95	0.996	0.997	0.869	0.868
1.00	0.996	0.997	0.881	0.886
1.05	0.998	0.998	0.900	0.920
1.10	1.000	1.000	0.929	0.937
1.15	1.000	1.000	0.943	0.952
1.20	1.000	1.000	0.956	0.967
1.25	1.000	1.000	0.980	0.982
1.30	1.000	1.000	0.978	0.984
1.35	1.000	1.000	0.994	0.998
1.40	1.000	1.000	0.991	0.995
1.45	0.999	0.999	0.991	0.995
1.50	1.000	1.000	0.990	0.996