Parametriset ja epäparametriset menetelmät vaihtovuoroasetelmassa

Auli Partanen

Tilastotieteen pro gradu -tutkielma

3. syyskuuta 2000

Jyväskylän yliopisto

Tilastotieteen laitos

Tilastotoimen menetelmien maisteriohjelma
(http://www.stat.jyu.fi/) (Päivitetty 16.5.1997)

Tilastotoimen menetelmien maisteriohjelman tavoite on kouluttaa opiskelija tilastotiedon keruun, jatkojalostuksen ja käytön asiantuntijaksi nykyaikaisessa tilastojärjestelmäympäristössä, jossa datat ovat survey-, koeasetelma- tai rekisteriperusteisia. Koulutus järjestetään yhteistyönä Jyväskylän yliopiston tilastotieteen laitoksen kanssa. Tilastotieteelliseltä kannalta kyse on survey-menetelmiin, biostatistiikkaan ja ekonometriaan erikoistuneiden tilastoasiantuntijoiden koulutuksesta. Ohjelman kesto päätoimisella opiskelijalla on kaksi lukuvuotta.

Tilastotoimen ytimenä ovat tilastotieteen, erityisesti tilastotoimen teorian ja tietojenkäsittelyn kurssit. Ohjelmaan voidaan sisällyttää myös koulutustavoitteita tukevia opintojaksoja lähitieteistä (esim. talous-, yhteiskunta- ja viestintätieteet) sekä yritystoimintaan perehdyttäviä kursseja. Tarjottava opetus koostuu osittain tilastotieteen laitoksen opetusohjelmasta, erityisesti suunnitelluista tilastotoimen teorian kursseista ja ostopalveluina hankituista muista erikoiskursseista. Opettajista pääosa on kotimaasta. Vierailijoina on myös ulkomaalaisia asiantuntijoita, joten opetus on osin englanninkielistä.

Ohjelman tärkeä osa on yliopiston ulkopuolisessa yhteistyötoimipaikassa suoritettu pro gradu -tutkielma ja siihen liittyvä harjoittelu. Yhteistyötoimipaikat ovat virallisesta tilastotoimesta, suuryrityksistä tai tutkimuslaitoksista. Ne osallistuvat osaltaan harjoittelun kustannuksiin. Perusajatus on, että pro gradu -tutkielma tai osa siitä toisi tutkimustulostensa osalta lisäarvoa yhteistyötoimipaikalle. Tätä tukenee se, että harjoittelu jaetaan kahteen osaan, joista ensimmäinen eli orientoiva osa on ensimmäisen opintovuoden lopussa. Sen jälkeen opiskelija palaa yliopisto-opiskeluihin yhden lukukauden ajaksi syventääkseen tietojaan siinä tilastotieteen osa-alueessa, joka on tarpeen aiotussa pro gradu -tutkielmassa. Maisteriohjelman viimeinen lukukausi muodostaa yhtenäisen tutkimusjakson, jonka aikana tehdään pro gradu -tutkielma.

Master's Programme in Statistical Systems
(http://www.stat.jyu.fi/) (Updated 16.5.1997)

The main traget of the Program is to educate high qualified professionals in the collection analysis, managing and dissemination of large data sets. The Program has been built mainly on the regular curriculum of the Department of Statistics in the University of Jyväskylä. In point of view of statistical sciences, the Program in Statistical Systems concentrate in the knowledge of survey methodology, biometry and econometry. Full-time students can graduate in two Academic years. The core courses cover the advanced theory in mathematical statistics, of statical systems expecially the survey methodology and some courses in information technology. Still the program is flexible that a moderate amount of different kinds of studies in the neighboring sciences can be included as studies in economics, social sciences and communication. Most of the instructors come from Finland but visiting professors from outboard are eventual using English.

An important part of the Program crows from the cooperational research work created between the Department of Statistics and the research units located outside the University of Jyväskylä. Those units are research and development departments in the bodies of official statistics, big business firms and research institutes. They share the costs of practices. Basic idea is that the MS.c. Thesis wil be written from the topics given by the cooperational research units and so the results could be of some contributed value to them. As an appropriate mode of policy may be here thus that the time of practice skills for the Thesis. This is proceeded in 6 final months of the Program, in collaboration with the same cooperating research unit.

Sisällysluettelo

1 VAIHTOVUOROASETELMA 3

1.1 KÄSITTEITÄ 3
1.2 I × J -PERUSASETELMA 5
1.3 KORKEAMMAN ASTEEN VAIHTOVUOROASETELMA 7
1.4 ESIMERKKIAINEISTON ESITTELY 9

2 LINEAARINEN SEKAMALLI 11

2.1. YLEINEN MUOTO 12
2.2 VAIHTOVUOROASETELMAAN LIITTYVÄ MALLI 13
2.3 MALLI OTSONIKOKEELLE 15
2.4 MALLIN ESITTÄMINEN MATRIISIMUODOSSA 16
2.5 KOVARIANSSIRAKENTEESTA 19

3 PARAMETRIEN ESTIMOINTI JA TESTAUS 22

3.1 MALLITTAMINEN *GLM*-PROSEDUURILLA 22
3.2 MALLITTAMINEN *MIXED*-PROSEDUURILLA 23
3.3 KOVARIANSSIRAKENTEEN SOPIVUUDEN TUTKIMINEN 26
3.3.1 Kovarianssirakenteita 26
3.3.2 Ohjelmointi *proc mixed*-proseduurilla 27
3.3.3 Informaatiokriteerit 28
3.4 KOVARIANSSIRAKENTEEN KIRJOITTAMINEN MATRIISIMUOTOON 29
3.5 ESTIMOINTI 30
3.6 TULOKSET JA NIIDEN TULKINTA 32
ESTIMOINNIN EROISTA 32

4 KONTRASTIMUUTTUJIEN KÄYTTÖ ESTIMOINNISSA JA TESTAUKSESSA 36

4.1 MÄÄRITELMÄ JA OMINAISUUKSIA 36
4.2 KONTRASTIMUUTTUJIEN VALINTA 39

5 ESTIMOINTI JA TESTAUS KÄYTTÄEN KONTRASTIMUUTTUJIA 45

5.1 KAHDEN OTOKSEN T-TESTI LUOTTAMUSVÄLEINEEN 45
5.2 KAHDEN OTOKSEN MANNIN-WHITNEYN TESTI LUOTTAMUSVÄLEINEEN 48

6 LOPUKSI 53

LÄHDELUETTELO 55

LIITTEET 57

1 Vaihtovuoroasetelma

Luvussa 1 käydään läpi vaihtovuorokokeeseen liittyviä käsitteitä sekä muutamia erilaisia vaihtovuoroasetelmia ja niiden ominaispiirteitä. Lisäksi kappaleessa 1.4 esitellään tutkielmassa käytetty havaintoaineisto.

Vaihtovuoroasetelma (*cross-over design*) on yleinen kahden tai useamman käsittelyn vertailuun kehitetty koeasetelma lääketieteellisissä tutkimuksissa. Tätä asetelmaa käytetään myös usein eläinlääketutkimuksissa, eläinten ruokintakoetutkimuksissa sekä liikuntatieteellisissä ja psykologisissa tutkimuksissa (Jones ja Kenward, 1989).

1.1 Käsitteitä

Tässä kappaleessa selitetään vaihtovuorokokeeseen liittyviä keskeisiä käsitteitä ja käydään läpi vaihtovuoroasetelman yleisiä ominaisuuksia. Käsitteen suomenkielisen nimen perässä suluissa on vastaava englannin kielinen termi ja näiden jälkeen on selitetty, mitä käsite tarkoittaa.

Vaihtovuoroasetelma eroaa muista koeasetelmista erityisesti siinä, että eri käsittelyjen mittaustulokset ovat samasta koeyksilöstä mitattuja. Vaihtovuorokokeessa jokaisella käsittelyllä samasta koeyksilöstä saatuja mittaustuloksia voidaan verrata keskenään eli kukin koeyksilö on itsensä kontrolli (Chinchilli, 1996). Tämä pätee oletuksella, että tutkimusolosuhteet ovat samat kaikilla jaksoilla eli jakson vaikutus on koko ajan sama. Samojen koeyksilöiden käytöstä seuraa, että vaihtovuorokokeessa koeyksilöitä tarvitaan vähemmän verrattuna muihin koeasetelmiin ja siten tarvittavien koeyksilöiden pieni määrä tekee asetelmasta sopivan lääketieteellisiin tutkimuksiin. Estimoidun käsittelyvaikutuksen varianssi on muotoa, jossa yksilön sisäisellä vaihtelulla (*within*) on suurempi merkitys kuin koeyksilöiden välisellä (*between*) vaihtelulla ja nimenomaan tätä yksilön sisäistä vaihtelua hyödynnetään käsittelyvaikutuksen testaamisessa.

Tässä tutkielmassa perusasetelma (taulukko 1.1) on $I \times J$ -koeasetelma (*design*), jossa on I ryhmää (käsittelyjärjestystä, *sequence*) ja J jaksoa (*period*). Esimerkiksi puhuttaessa 2×3 –koeasetelmasta tarkoitetaan asetelmaa, jossa on 2 ryhmää ja 3 käsittelyjaksoa.

Perusasetelmaa voidaan laajentaa (taulukko 1.1) lisäämällä totuttautumisjakso (*run-in period*), jonka aikana koeyksilöt totutetaan kokeeseen liittyviin rutiineihin, joita ovat esimerkiksi lääkkeen säännöllinen nauttiminen ja mittaukset. Rutiinien sujuvuudella voidaan vähentää harhaa, joka voi aiheutua rutiinien hallitsemisen eroista ensimmäisellä ja myöhemmillä jaksoilla. Totuttautumisjaksoa käytetään toisaalta usein myös seulontajaksona (*screening period*), jonka aikana valitaan koeyksilöt ja varmistetaan yksilöiden sopivuus kokeeseen. Lisäksi totuttautumisjakson aikana koeyksilöt voidaan valmistella kokeeseen sopiviksi eli esimerkiksi pitää yksilöt paastolla ja ilman lääkitystä. Totuttautumisjakson aikana ennen käsittelyä ja käsittelyjaksojen välillä tehdään perustasomittaukset (*baseline measurement*). Näiden avulla voidaan tarkastella tarkemmin viivästevaikutusta sekä käsittelyvaikutusta viivästevaikutuksen tapauksessa (Hinkka, 1993). Käsittelyjaksojen välillä on usein puhdistumisjakso (*wash-out period*), jonka aikana edellisen käsittelyn vaikutuksen annetaan poistua ja koeyksilöiden palata kokeen alussa olleelle tasolleen.

Taulukko 1.1 Perusasetelma vs. laajennettu perusasetelma

	Perusasetelma			Laajennettu perusasetelma			
Käsittely-järjestys	Jakso 1	Jakso 2		Totuttautumis-jakso	Jakso 1	Puhdistumis-jakso	Jakso 2
Käsittely-järjestys 1	B	A	vs.	Perustaso-mittaus	B	Perustaso-mittaus	A
Käsittely-järjestys 2	A	B		Perustaso-mittaus	A	Perustaso-mittaus	B

$I \times J$ -koeasetelmassa kiinnostuksen kohteena ovat suora käsittelyvaikutus, viivästevaikutus ja jakson vaikutus. Viivästevaikutus (*carry-over effect*) tarkoittaa sitä, että aikaisempi käsittely vaikuttaa yhä aloitettaessa seuraavaa käsittelyä. Esimerkiksi koeyksilön saadessa ensimmäisellä jaksolla kipulääkettä ja toisella lumelääkettä

(*placebo*) viivästevaikutuksella tarkoitetaan sitä, että ensimmäisen käsittelyn vaikutus ei ole poistunut ennen toista käsittelyä ja tämän takia kivut ovat pienemmät myös toisella jaksolla. Viivästevaikutus jaetaan ensimmäisen ja useamman asteen vaikutuksiin. Ensimmäisen asteen viivästevaikutus tarkoittaa edellisen käsittelyn vaikutusta vielä käsittelyä seuraavalla jaksolla. Yleisemmin g:nnen asteen viivästevaikutus tarkoittaa aiemmin annetun käsittelyn vaikutusta käsittelyä seuraavien g jakson aikana.

Suora käsittelyvaikutus tarkoittaa sitä välitöntä vaikutusta, joka käsittelyllä on mitattavaan muuttujaan. Jakson vaikutuksella (*period effect*) tarkoitetaan sitä, että jaksolla itsellään on jokin vaikutus mitattavaan muuttujaan. Jotta jakson vaikutus voitaisiin erottaa käsittelyjen vaikutuksista, ryhmät saavat käsittelyt eri järjestyksessä, esimerkiksi järjestyksissä AB ja BA. Ryhmän vaikutuksella (*sequence effect*) tarkoitetaan sitä, että käsittelyjärjestyksillä on eroa. Tämän vaikutuksen testaamisella ei yleensä ole merkitystä, koska sitä testaamalla voidaan mitata satunnaistamisen (*randomisation*) onnistumista ja yleensä se oletetaan onnistuneeksi.

1.2 I × J -perusasetelma

Perusasetelmassa kukin käsittelyjärjestys i ($i = 1, \ldots, I$) sisältää T käsittelyä satunnaisessa, yleensä jakson suhteen tasapainotetussa järjestyksessä ja koe kestää J jakson ajan (taulukot 1.2 ja 1.3).

Taulukko 1.2 $I \times J$ –perusasetelma

$I \times J$		
Käsittely-järjestys	Ryhmän koko	Jakso 1 … J
1	n_1	Järjestys 1
.	.	. .
.	.	. .
.	.	. .
I	n_I	Järjestys I

Taulukko 1.3 Havainnot y_{ijk} (i = käsittelyjärjestys, j = jakso ja k = yksilö) $I \times J$ – perusasetelmassa

$I \times J$				
Käsittely-järjestys	Yksilöt	Jakso 1	. . .	J
1	1 . . n_1	y_{111} . . y_{11n_1}	. . .	y_{1J1} . . y_{1Jn_1}
. . .	. . .	. . .		. . .
I	1 . . n_I	y_{I11} . . y_{I1n_I}	. . .	y_{IJ1} . . y_{IJn_I}

Yksinkertaisin vaihtovuoroasetelma on 2×2 –koeasetelma, jossa on siis 2 käsittelyjärjestystä ja 2 jaksoa. 2×2–perusasetelma toteutetaan siten, että koeyksilöt jaetaan satunnaisesti kahteen ryhmään, käsittelyjärjestykset 1 ja 2. Koeyksilöille annetaan kaksi eri käsittelyä, käsittelyt A ja B satunnaisesti valitussa käsittelyjärjestyksessä. Käsittelyt voivat olla esimerkiksi 10 mg tutkimuslääkettä ja 20 mg tutkimuslääkettä. Ryhmien käsittelyjärjestykset eroavat ja tällöin esimerkiksi koeyksilöt, joilla on käsittelyjärjestys 1, saavat ensimmäisellä jaksolla käsittelyä B ja toisella jaksolla käsittelyä A. Vastaavasti koeyksilöt, joilla on käsittelyjärjestys 2, saavat käsittelyt järjestyksessä A ja B.

Olosuhteet, joissa 2×2 –perusasetelmaa käytetään, tulee olla asetelman käytölle sopivat. Puhdistusjakson tulee olla niin pitkä, ettei viivästevaikutusta ole, jotta 2×2 – perusasetelma olisi käsittelyn vaikutuksen tutkimiseen sopiva. Oletuksena 2×2

–perusasetelman käytölle on tutkittavan sairauden tilan pysyminen vakaana eli koeyksilön tilassa ei saa tapahtua sellaista muutosta, että alkutilanne olisi erilainen eri jaksoilla, (Patel, 1996). Näiden olosuhteiden vallitessa myös 2×2 –perusasetelma antaa oikeita ja tehokkaita estimaatteja suoralle käsittelyvaikutukselle.

1.3 Korkeamman asteen vaihtovuoroasetelma

Chown ja Liun (1992) mukaan 2×2 –perusasetelma ei sovellu koeasetelmaksi, jos kokeessa on viivästevaikutusta, koska tällöin estimaatit ovat harhaisia. Koeasetelma ei myöskään tuota riippumattomia sisäisen vaihtelun estimaatteja. Jos viivästevaikutus aiheuttaa ongelmia, niin tällöin käytetään korkeamman asteen vaihtovuorokokeita (*higher order cross-over design*). Korkeamman asteen vaihtovuoroasetelma tarkoittaa (Chow ja Liu, 1992) vaihtovuoroasetelmaa, jossa käsittelyjärjestysten ja/tai jaksojen määrä on suurempi kuin käsittelyjen määrä eli suurempi kuin 2. Yleisemmin käytettyjä koeasetelmia kahta käsittelyä vertaavissa kokeissa ovat asetelmat 2×3 , 4×2 tai 4×3 .

Jonesin ja Kenwardin (1989) mukaan kirjallisuudessa asetelman hyvyyden mittarina on käytetty parametrin estimaattorin varianssia siten, että sellainen asetelma on paras, johon liittyvällä estimaattorilla on pienin varianssi. Cheng ja Wu (1980), Laska, Meisner ja Kushner (1983), Laska ja Meisner (1985) ja Matthews (1987) ovat Jonesin ja Kenwardin mukaan osoittaneet, että käytettäessä tasaista kovarianssirakennetta (*compound symmetry, CS*) taulukossa 1.4 esitetyt vaihtovuoroasetelmat ovat parhaimmat kahden käsittelyn vertaamiseen sopivat asetelmat.

Taulukko 1.4 Parhaimmat kahden käsittelyn (T=2) vertailuun liittyvät korkeamman asteen vaihtovuoroasetelmat

Asetelma 1 (4×2)

Käsittely-järjestys	Jakso	
	I	II
1	A	A
2	B	B
3	A	B
4	B	A

Asetelma 4 (2×5)

Käsittely-järjestys	Jakso				
	I	II	III	IV	V
1	A	B	B	A	A
2	B	A	A	B	B

Asetelma 2 (2×3)

Käsittely-järjestys	Jakso		
	I	II	III
1	A	B	B
2	B	A	A

Asetelma 5 (4×6)

Käsittely-järjestys	Jak so					
	I	II	III	IV	V	VI
1	A	B	B	A	A	B
2	B	A	A	B	B	A
3	A	A	B	B	B	A
4	B	B	A	A	A	B

Asetelma 3 (4×4)

Käsittely-järjestys	Jakso			
	I	II	III	IV
1	A	A	B	B
2	B	B	A	A
3	A	B	B	A
4	B	A	A	B

Lähde: Jones ja Kenward (1989)

1.4 Esimerkkiaineiston esittely

Seuraavaksi perehdytään tarkemmin asetelmaan, joka on tässä tutkielmassa käytetyn esimerkkiaineiston mukainen. Asetelman 2 (taulukko 1.4) mukaisessa 2×3 -koeasetelmassa on 2 käsittelyjärjestystä (1 ja 2) sekä 3 jaksoa (I, II ja III). Erilaisia käsittelyjä voi olla kaksi (käsittelyt A ja B). Asetelmassa 2 voivat ryhmät saada käsittelyt eri järjestyksissä, esimerkiksi järjestyksissä ABA ja BAB tai ABB ja BAA.

Jos jokaista käsittelyä seuraa jokainen toinen käsittely ainakin kerran (myös sama käsittely) kutsutaan 2×3 –asetelmaa tasapainotetuksi. Näin on käsittelyjärjestyksellä ABB ja BAA. Chinchillin (1996) mukaan verrattaessa 2×3 –asetelmaa 2×2 –perusasetelmaan edellinen on parempi, koska siinä voidaan estimoida suora käsittelyvaikutus myös viivästevaikutuksen tapauksessa. Lisäksi voidaan estimoida koeyksilön sisäinen vaihtelu, koska kukin koeyksilö saa toisen käsittelyistä kaksi kertaa.

Aineisto, jolla teoriaosiota sovelletaan käytäntöön, on Hafnerin, Kochin ja Canadan (1988) artikkelista. Kokeessa tutkittiin nuorilla ja aikuisilla hiirillä bentobarbituraatin poistumista aineenvaihdunnasta käsiteltäessä hiiret 0.3 yksikön annoksella otsonia 3 tunnin ja 45 minuutin ajan. Koeasetelmaksi valittiin kolmijaksoinen vaihtovuoroasetelma (2×3). Tällöin saman voimakkuuden saavuttamiseksi tarvittiin vähemmän koe-eläimiä kuin olisi tarvittu rinnakkaisryhmien asetelmassa (*parallel groups design*). Kymmenen nuorta hiirtä ja kymmenen aikuista hiirtä jaettiin satunnaisesti käsittelyjärjestyksiin 1 ja 2. Satunnaistamissuunnitelman mukaan siis käsittelyt molemmilla järjestyksillä sai viisi nuorta ja viisi aikuista hiirtä. Kokeen suorittamisessa oli kaksi käsittelyohjetta:

Hiiren annetaan hengittää otsonia ensimmäisenä päivänä, ilmaa kahdeksantena päivänä ja otsonia 15:ntenä päivänä. Käsittelyjärjestystä merkitään *O:A:O*.

Hiiren annetaan hengittää ilmaa ensimmäisenä päivänä, otsonia kahdeksantena päivänä ja ilmaa 15:ntenä päivänä. Käsittelyjärjestystä merkitään *A:O:A*.

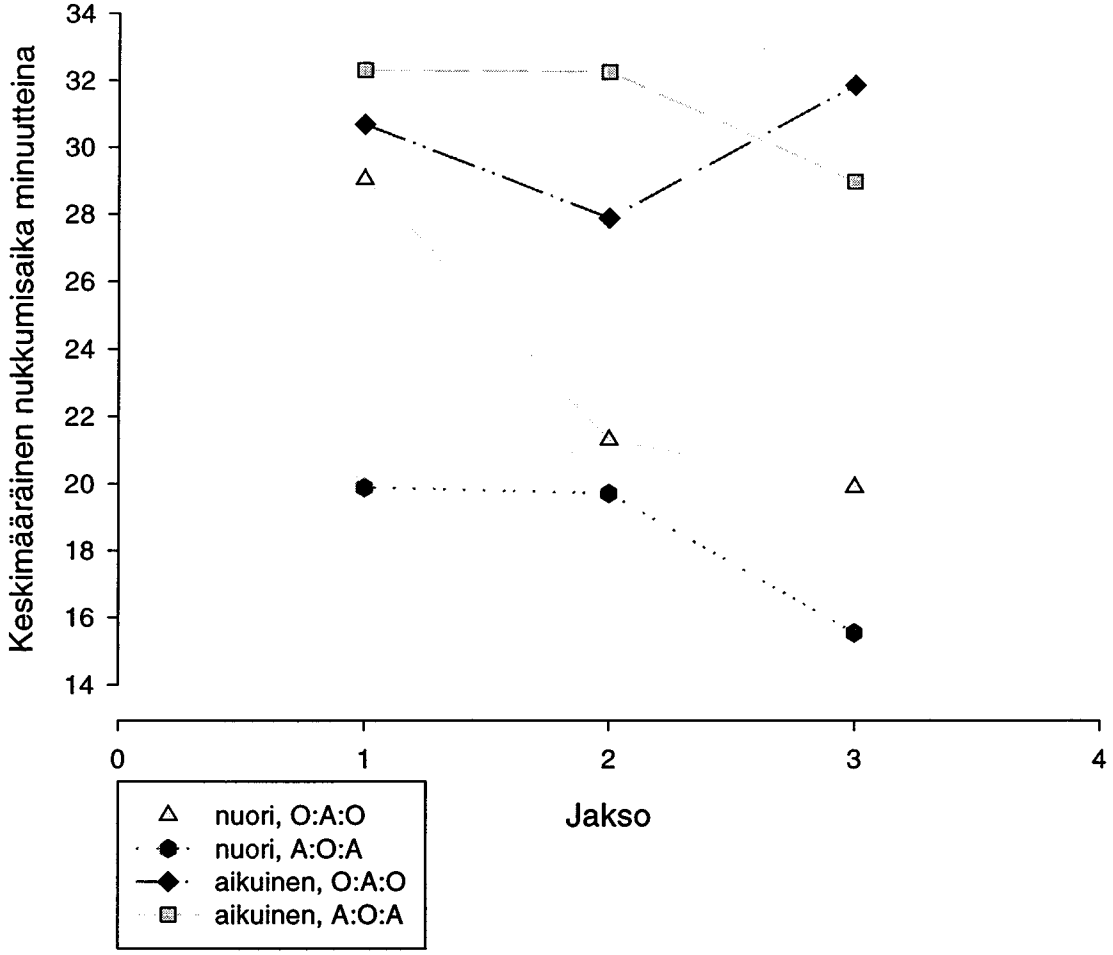
Mahdollisten viivästevaikutusten poistamiseksi päivien yksi ja kahdeksan jälkeen oli viikon puhdistumisjakso. Bentobarbituraatti annettiin lihakseen pistettynä välittömästi ilma- tai otsonikäsittelyn jälkeen ja hiirten painon vaihtelu otettiin huomioon annostuksessa (50 mg/kg). Tulos, josta oltiin kiinnostuneita, oli kunkin hiiren nukkumisaika minuutteina bentobarbituraattipistoksen jälkeen.

Taulukko 1.5 Hiirten nukkumisajat minuutteina bentobarbituraattipistoksen jälkeen. Liitteessä aineisto on esitetty SAS-ohjelmiston vaatimassa muodossa.

Ikä	Käsittely-järjestys*	Jakso I	II	III
Nuori	*O:A:O*	38.1	23.8	16.4
Nuori	*O:A:O*	19.7	16.6	15.3
Nuori	*O:A:O*	39.0	25.3	30.2
Nuori	*O:A:O*	31.7	19.1	17.6
Nuori	*O:A:O*	16.8	21.8	20.1
Nuori	*A:O:A*	22.3	21.8	15.1
Nuori	*A:O:A*	24.0	19.5	15.1
Nuori	*A:O:A*	17.8	17.6	21.9
Nuori	*A:O:A*	15.5	20.0	10.1
Aikuinen	*O:A:O*	30.0	31.2	37.5
Aikuinen	*O:A:O*	29.4	34.4	35.3
Aikuinen	*O:A:O*	45.8	27.5	30.9
Aikuinen	*O:A:O*	25.7	21.7	23.7
Aikuinen	*O:A:O*	22.6	24.7	32.0
Aikuinen	*A:O:A*	33.5	22.6	28.6
Aikuinen	*A:O:A*	36.1	38.0	32.9
Aikuinen	*A:O:A*	32.1	42.4	30.5
Aikuinen	*A:O:A*	32.9	35.1	27.9
Aikuinen	*A:O:A*	27.0	23.2	25.1

* Yksi nuori hiiri järjestyksellä *A:O:A* kuoli kokeen alkuvaiheessa.

Kuva 1.1.
Keskimääräiset nukkumisajat bentobarbituraattipistoksen jälkeen jaksottain ikä- ja käsittelyryhmittäin



Kuvasta 1.1 nähdään, että otsonikäsitellyt hiiret näyttäisivät nukkuvan kauemmin kuin ilmakäsittelyn saaneet hiiret. Kuvassa esitetyistä keskiarvoista voidaan päätellä myös se, että aikuisten hiirten nukkumisajat ovat molemmilla käsittelyillä suuremmat.

2 Lineaarinen sekamalli

Havaintoaineisto $I \times J$-perusasetelmalle voidaan tavallisesti esittää vektori- ja matriisimuodossa seuraavasti:

$$(\underset{n\times 1}{\mathbf{y}}\ \underset{n\times p}{\mathbf{X}}\ \underset{n\times q}{\mathbf{Z}}) = \begin{bmatrix} y_1 & x_{11} & \cdots & x_{1p} & z_{11} & \cdots & z_{1q} \\ y_2 & x_{21} & & x_{2p} & z_{21} & & z_{2q} \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ y_n & x_{n1} & \cdots & x_{np} & z_{n1} & \cdots & z_{nq} \end{bmatrix} = \begin{bmatrix} y_1 & \mathbf{x}_1^T & \mathbf{z}_1^T \\ y_2 & \mathbf{x}_2^T & \mathbf{z}_2^T \\ \vdots & \vdots & \vdots \\ y_n & \mathbf{x}_n^T & \mathbf{z}_n^T \end{bmatrix},$$

missä n on $N \times J$ eli mittausten kokonaislukumäärä ($N = \sum_{i=1}^{I} n_i$). Jatkossa käytetään merkintää n yksinkertaisuuden vuoksi.

Vektori $\mathbf{y}$ on $n \times 1$ -vektori ja sisältää tutkittavan vastemuuttujan havaitut arvot. Matriisi $\mathbf{X}$ on $n \times p$ -matriisi (*design matrix*), joka liittää kiinteät selittäjät vasteeseen $\mathbf{y}$. $\mathbf{Z}$ on $n \times q$ -matriisi, joka liittää yksilökohtaiset selittäjämuuttujat vasteeseen $\mathbf{y}$. Erityispiirteenä on, että matriisiin $\mathbf{Z}$ liittyvien selittäjämuuttujien vaikutuksen ajatellaan olevan satunnaista eli yksilöillä on oma taustastaan riippuva yksilöllinen vastetasonsa. Luvussa 2.4 matriiseja $\mathbf{X}$ ja $\mathbf{Z}$ on havainnollistettu esimerkkiaineiston avulla.

2.1. Yleinen muoto

Vaihtovuorokokeeseen liittyvä yleinen malli kirjoitetaan yleensä muodossa

$$\mathbf{y} = \mathbf{X}\alpha + \mathbf{Z}\mathbf{b} + \mathbf{e}, \tag{1}$$

missä $\mathbf{b} \sim \mathrm{N}_q(0, \Sigma_b)$, $\mathbf{e} \sim N_n(0, \Sigma_e)$ ja α sisältää kiinnostuksen kohteena olevat kiinteät selittäjät. Siis vektori $\mathbf{y}$ noudattaa n-ulotteista multinormaalijakaumaa $\mathrm{N}_n(\mu, \Sigma)$. Koska $E(\mathbf{b}) = 0$ ja $E(\mathbf{e}) = 0$ saadaan odotusarvoksi $\mathrm{E}(y) = \mu = \mathbf{X}\alpha$ ja koska $Cov(\mathbf{b}) = \Sigma_b$ ja $Cov(\mathbf{e}) = \Sigma_e$, saadaan kovarianssimatriisiksi

$$\mathbf{Cov(y)} = \Sigma = \mathbf{Z}\Sigma_b\mathbf{Z}^{\mathrm{T}} + \Sigma_e. \tag{2}$$

Jatkossa tarkastellaan kahta mallia, jotka kovarianssirakenteineen ovat:

Kiinteä yksilövaikutuksen malli (*GLM* ja *Mixed*, kiinteä):

$$\mathbf{y} = \mathbf{X}\alpha + \mathbf{e}$$

ja (3)

$$\Sigma = \Sigma_e = \sigma_e^2\mathbf{I}$$

Satunnaisen yksilövaikutuksen malli (*Mixed*, satunnainen):

$$\mathbf{y} = \mathbf{X}\alpha + \mathbf{Zb} + \mathbf{e}$$

ja (4)

$$\Sigma = \Sigma_{\mathbf{b}} + \Sigma_{\mathbf{e}} = \sigma_s^2 \mathbf{ZZ}^{\mathrm{T}} + \sigma_e^2 \mathbf{I}$$

Malli (3) on yleinen lineaarinen malli (*general linear model, GLM*), joka on kaavassa (1) esitetyn sekamallin (*mixed model*) erikoistapaus. Tällöin oletetaan $\mathbf{Z} = 0$ ja $\Sigma = \Sigma_e = \sigma_e^2 \mathrm{I}$, missä $\mathbf{I}$ on identtinen matriisi, jonka diagonaalilla on ykkösiä ja muualla nollia. Käytännössä tämä tarkoittaa sitä, että koeyksilö sovitetaan kiinteänä tekijänä.

Malli (4) tunnetaan kirjallisuudessa nimellä lineaarinen sekamalli (*general linear mixed model*). Sekamallilla koeyksilö voidaan sovittaa myös satunnaisena ja tarkastella muitakin kuin tasaista kovarianssirakennetta (katso luku 2.5). Satunnaistekijänä voidaan käsitellä koeyksilöiden ohella myös tutkimuskeskuksia.

Malleissa (3) ja (4) selitettävän muuttujan arvot ovat vektorissa $\mathbf{y}$ ja termi $\mathbf{X}\alpha$ edustaa mallin kiinteää osaa (*fixed effect*), joka tässä sovelluksessa sisältää jaksosta, käsittelystä, mahdollisesta viivästevaikutuksesta ja käsittelyjärjestyksestä malliin tulevan vaikutuksen. Parametrivektori α siis sisältää yleensä kiinnostuksen kohteena olevat kiinteät parametrit. Lisäksi mallissa (4) olevassa termissä $\mathbf{Zb}$ on mallin satunnainen osa (*random effect*), joka nyt sisältää yksilöstä tulevan satunnaisvaihtelun.

2.2 Vaihtovuoroasetelmaan liittyvä malli

Yleensä vaihtovuoroasetelma on järjestetty siten, että yksilöt on jaettu satunnaisesti I käsittelyjärjestykseen, joista jokaisessa on K yksilöä ($n_i = K$ kaikilla i:n arvoilla). Mittauksia tehdään J jakson aikana, joten havaittuja arvoja on $I \cdot J \cdot K$ kappaletta. Kokeessa halutaan verrata T käsittelyä ja koeyksilöiden käsittely toteutetaan eri ryhmissä erilaisissa järjestyksissä.

Lineaarinen sekamalli käsittelyjärjestyksen i koeyksilön k vasteelle Y_{ijk} jakson j aikana on

$$Y_{ijk} = \mu + G_i + S_{ik} + P_j + F_{t(i,j)} + C_{t(i,j-1)} + e_{ijk}, \qquad (5)$$

missä μ = kokonaiskeskiarvo,

G_i = käsittelyjärjestyksen (*sequence*) i kiinteä vaikutus

S_{ik} = käsittelyjärjestyksen i koeyksilön k edustamasta perusjoukosta tuleva "satunnaistaso"

P_j = jakson j vaikutus

F_t = käsittelyn t vaikutus, missä $t(i, j)$ on käsittely käsittelyjärjestyksellä i jaksolla j

$C_{t(i,j-1)}$ = käsittelyn t viivästevaikutus, $C_{t(i,0)} = 0$

e_{ijk} = vastaus- ja mittausvirheestä käsittelyjärjestyksellä i yksilölle k jaksolla j aiheutuva virhemuuttuja.

Yleensä oletetaan, että yksilöt on poimittu satunnaisesti perusjoukosta, jossa yksilöä kuvaava satunnaistaso on normaalisti jakautunut odotusarvolla 0 ja varianssilla σ_S^2 eli $S_{ik} \sim N(0, \sigma_S^2)$ ja satunnaisvaihtelu on normaalisti jakautunut odotusarvolla 0 ja varianssilla σ_e^2 eli $e_{ijk} \sim N(0, \sigma_e^2)$. Oletetaan myös, että satunnaistaso S_{ik} ja vastaus- ja mittausvirheestä aiheutuva satunnaisvaihtelu e_{ijk} ovat toisistaan riippumattomia. Kun käytetään näitä oletuksia, mittaustuloksiin liittyvä kovarianssirakenne on niin sanottu tasainen kovarianssirakenne.

Summarajoitus ja nurkkasolurajoitus

Identifioituvuuden vuoksi parametreille joudutaan yleensä asettamaan summarajoitus tai nurkkasolurajoitus. Summarajoituksessa parametrisointi tehdään siten, että jokaisen selittäjän eri tasoihin liittyvien parametrien summa on nolla, mikä esimerkiksi jakson vaikutukselle tarkoittaa sitä, että eri jaksojen vaikutusten summa on nolla eli

$$\sum_{j=1}^{J} P_j = 0.$$

Summarajoitusta käytettäessä selittäjien tasoille laskettuja estimaatteja tulkitaan siten, että selittäjän tasoon liittyvä estimaatti kuvaa poikkeavuutta keskimääräisestä tasosta selittäjän eri luokissa. Nurkkasolurajoituksessa oletetaan jokin selittäjän taso nollaksi eli esimerkiksi ensimmäisen jakson vaikutus on nolla eli $P_1 = 0$. Yhtä selittäjää koskevassa nurkkasolurajoituksessa estimoitavana on $p-1$ estimaattia selittäjien tasojen mittaustuloksille. Nurkkasolurajoitusta käytettäessä parametreja tulkitaan siten, että

nollaksi asetettu parametrin taso on vertailuluokka. *GLM*-proseduuri olettaa nurkkasolurajoituksen siten, että parametrin viimeinen taso on vertailutaso, kun taas *Mixed*-proseduurissa tällaisia oletuksia ei eksplisiittisesti tehdä. Tällöin esimerkiksi sovitettaessa mallia (3) esimerkkiaineistolle nurkkasolurajoituksena oletetaan, että ikäryhmä nuori on vertailuluokka, jakso 1 on vertailuluokka, ilmakäsittely on vertailuluokka ja ilman viivästevaikutus on vertailuluokka. Käytettäessä *GLM*-proseduuria aineisto järjestetään siten, että halutut vertailuluokat ovat muuttujan viimeisenä tasona.

2.3 Malli otsonikokeelle

Pro gradu –tutkielmassa esimerkkiaineistona käytetyn 2×3 –koeasetelman malli on esitetty seuraavassa kaavassa. Malli on tekijöiden (Canada, Hafner ja Koch) artikkelissa esitetyn mukainen.

$$Y_{ijk} = \mu + A_h + S_{ik} + P_j + F_{t(i,j)} + C_{t(i,j-1)} + e_{ijk}, \qquad (6)$$

missä symbolit ovat muuten samat kuin kaavassa (5), mutta

A_h = ikäryhmän h vaikutus, missä 1 = nuori, 2 = aikuinen,

$F_{t(i,j)}$ = käsittelyn $t(i, j)$ vaikutus, missä 1 = ilma ja 2 = otsoni

ja indeksit ovat

i = 1 (nuori, käsittelyjärjestys *O:A:O*) ... 4 (aikuinen, käsittelyjärjestys *A:O:A*)

j = 1 (jakso 1) ... 3 (jakso 3)

k = 1 ... n_i.

Oletukset ovat samat kuin kaavan (5) jälkeen kappaleessa 2.2 esitetyt. Viivästevaikutuksen parametri otettiin malliin mukaan, vaikka käsittelyjen välillä olikin viikon puhdistumisjakso. Toisen asteen viivästevaikutus ei ole mallissa mukana. Kaavassa (5) olleen käsittelyjärjestyksen sijaan malliin otettiin termi ikäryhmä, jolla tässä koeasetelmassa on vaikutusta.

Kaavan (6) mukaiset mallit on taulukoissa 2.1 ja 2.2 selvyyden vuoksi purettu molempien ikäryhmien kummallekin käsittelyjärjestykselle omiksi malleikseen.

Taulukko 2.1 Vasteet ikäryhmittäin, käsittelyittäin ja jaksoittain, $k = 1,\ldots,K$

Ikäryhmä	Käsittely-järjestys	Jakso I	Jakso II	Jakso III
Nuori	*O:A:O*	Y_{11k}	Y_{12k}	Y_{13k}
	A:O:A	Y_{21k}	Y_{22k}	Y_{23k}
Aikuinen	*O:A:O*	Y_{31k}	Y_{32k}	Y_{33k}
	A:O:A	Y_{41k}	Y_{42k}	Y_{43k}

Taulukko 2.2 Tilastolliset mallit taulukon 2.1 vasteille

Ikä-ryhmä	Käsittely-järjestys	Jakso I	Jakso II	Jakso III
Nuori	*O:A:O*	$\mu + A_1 + S_{1k} + P_1 + F_2 + e_{11k}$	$\mu + A_1 + S_{1k} + P_2 + F_1 + C_2 + e_{12k}$	$\mu + A_1 + S_{1k} + P_3 + F_2 + C_1 + e_{13k}$
	A:O:A	$\mu + A_1 + S_{2k} + P_1 + F_1 + e_{21k}$	$\mu + A_1 + S_{2k} + P_2 + F_2 + C_1 + e_{22k}$	$\mu + A_1 + S_{2k} + P_3 + F_1 + C_2 + e_{23k}$
Aik.	*O:A:O*	$\mu + A_2 + S_{3k} + P_1 + F_2 + e_{31k}$	$\mu + A_2 + S_{3k} + P_2 + F_1 + C_2 + e_{32k}$	$\mu + A_2 + S_{3k} + P_3 + F_2 + C_1 + e_{33k}$
	A:O:A	$\mu + A_2 + S_{4k} + P_1 + F_1 + e_{41k}$	$\mu + A_2 + S_{4k} + P_2 + F_2 + C_1 + e_{42k}$	$\mu + A_2 + S_{4k} + P_3 + F_1 + C_2 + e_{43k}$

2.4 Mallin esittäminen matriisimuodossa

Seuraavaksi esitetään esimerkkiaineistoon liittyvä tilastollinen malli $\mathbf{Y} = \mathbf{X}\alpha + \mathbf{Zb} + \mathbf{e}$ luvussa 2.1 esitetyssä yleisessä muodossa. Havaitut vasteen arvot muodostavat siis n-ulotteisen eli (19×3)-vektorin $\mathbf{Y}$. Vektori α sisältää kaikki estimoitavat vaihtovuoroasetelmaan liittyvät parametrit ja esimerkkiaineiston tilanteessa

$$\underset{10\times 1}{\alpha} = \begin{bmatrix} \mu \\ A_1 \\ A_2 \\ P_1 \\ P_2 \\ P_3 \\ F_1 \\ F_2 \\ C_1 \\ C_2 \end{bmatrix}$$

Esimerkkiaineistossa kiinteiden vaikutuksiin liittyvä asetelmamatriisi **X** on $n \times p$ eli $(19\times 3)\times 10$-matriisi, jonka rivit vastaavat 19×3 mittausta. Sarakkeet puolestaan vastaavat vektorin α mukaisesti parametreja, joilla on kiinteät vaikutukset. Matriisi **X** kahden ensimmäisen ja viimeisen koeyksilön, yhteensä yhdeksän mittauksen, osalta on

$$\underset{57\times 10}{\mathbf{X}} = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 \end{bmatrix}$$

Matriisissa **X** kolmella ensimmäisellä rivillä ovat ensimmäisen koeyksilön ensimmäiseen, toiseen ja kolmanteen jaksoon liittyvien kiinteiden selittäjien arvot. Seuraavilla kolmella rivillä ovat toisen koeyksilön ensimmäiseen, toiseen ja kolmanteen jaksoon liittyvien selittäjien arvot ja viimeisillä kolmella rivillä ovat 19:nnen koeyksilön ensimmäiseen, toiseen ja kolmanteen jaksoon liittyvien selittäjien arvot.

Mallin termi **Z** esitettynä matriisimuodossa on 57×19 -matriisi, jonka rivit vastaavat erillisiä mittauksia (mittaukset kolmella jaksolla 19 koeyksilöstä) ja sarakkeet koeyksilöitä (19 hiirtä). Esimerkkiaineiston tilanteessa **Z**-matriisin kuusi ensimmäistä ja

kolme viimeistä riviä kahden ensimmäisen ja viimeisen koeyksilön mittauksia vastaten ovat

$$\underset{57\times19}{\mathbf{Z}} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 1 & 0 & 0 & \cdots & 0 \\ 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & & & & \vdots \\ 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & \cdots & 0 & 1 \end{bmatrix}$$

Vektori **b** antaa koeyksilöihin liittyvät satunnaistasot kullekin mittaukselle ja esimerkkiaineiston tapauksessa **b** on

$$\underset{19\times1}{\mathbf{b}} = \begin{bmatrix} s_{11} \\ s_{12} \\ s_{13} \\ s_{14} \\ s_{15} \\ s_{21} \\ s_{22} \\ s_{23} \\ s_{24} \\ \vdots \end{bmatrix}.$$

Vektorissa **b** viisi ensimmäistä solua sisältävät aineiston viiden ensimmäisen koeyksilön satunnaistasot eli niiden nuorten koeyksilöiden satunnaistasot, joiden käsittelyjärjestys on *O:A:O*. Seuraavat neljä solua sisältävät satunnaistasot koeyksilöille viidestä yhdeksään eli koeyksilöille, jotka ovat aikuisia ja joiden käsittelyjärjestys on *A:O:A*.

2.5 Kovarianssirakenteesta

Tutkielmassa käytetty yleistetty lineaarinen sekamalli on muotoa $\mathbf{y} = \mathbf{X}\alpha + \mathbf{Z}\mathbf{b} + \mathbf{e}$. Malli on esitetty kaavassa (4) ja matriisimuotoisena kappaleessa 2.4. Kaavassa yksilön satunnaisvaihtelu S_{ik} sisältyy termiin **b** ja jäännösvaihtelu e_{ijk} sisältyy termiin **e**. Vaihtelun S_{ik} oletetaan olevan riippumatonta ja noudattavan normaalijakaumaa keskiarvolla 0 ja varianssilla σ_S^2 eli $S_{ik} \sim N(0, \sigma_S^2)$. Oletetaan myös, että virhemuuttujat e_{ijk} ovat riippumattomia ja normaalisti jakautuneita odotusarvolla 0 ja varianssilla σ_e^2 eli $e_{ijk} \sim N(0, \sigma_e^2)$. Tästä saadaan mallin (3) tapauksessa kovarianssimatriisi $\mathbf{Cov(y)} = \Sigma_e = \sigma_e^2\mathbf{I}$ ja mallille (4) kovarianssimatriisi $\mathbf{Cov(y)} = \Sigma = \Sigma_b + \Sigma_e = \sigma_S^2\mathbf{Z}\mathbf{Z}^T + \sigma_e^2\mathbf{I}$. Siis mallille (4) esimerkkiaineiston tapauksessa rakenne on tasainen kovarianssirakenne (*Compound symmetry, CS*). Esimerkkiaineiston kovarianssirakenteeksi saadaan mallin (4) mukaisesti (katso luku 2.4)

$$\mathbf{Cov(y)} = \Sigma_b + \Sigma_e = \sigma_s^2\mathbf{Z}\mathbf{Z}^T + \sigma_e^2\mathbf{I}$$

$$= \sigma_s^2 \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 1 & 0 & 0 & \cdots & 0 \\ 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & \cdots & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & \cdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & \cdots \\ \vdots & & & \vdots & & \vdots & & & \vdots & \end{bmatrix} + \sigma_e^2 \begin{bmatrix} 1 & 0 & 0 & \cdots \\ 0 & 1 & 0 & \\ 0 & 0 & 1 & \\ \vdots & & & \ddots \end{bmatrix}$$

$$= \sigma_s^2 \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 & \cdots \\ 1 & 1 & 1 & 0 & 0 & 0 & \cdots \\ 1 & 1 & 1 & 0 & 0 & 0 & \cdots \\ 0 & 0 & 0 & 1 & 1 & 1 & \cdots \\ 0 & 0 & 0 & 1 & 1 & 1 & \cdots \\ 0 & 0 & 0 & 1 & 1 & 1 & \cdots \\ \vdots & & & & & \vdots & \ddots \end{bmatrix} + \sigma_e^2 \begin{bmatrix} 1 & 0 & 0 & \cdots \\ 0 & 1 & 0 & \\ 0 & 0 & 1 & \\ \vdots & & & \ddots \end{bmatrix}.$$

Kuten matriisitulosta $\mathbf{ZZ}^{\mathrm{T}}$ nähdään, kovarianssimatriisi on lohkodiagonaalinen ja yksi lohko vastaa yhden koeyksilön mittaustuloksia. Kunkin lohkon eli koeyksilön kovarianssimatriisin diagonaalilla ovat varianssit $\sigma_s^2+\sigma_e^2$ ja muualla ovat kovarianssit σ_s^2 eli kovarianssimatriisi käsittelyjärjestyksen i koeyksilön k mittaustuloksille on

$$\mathbf{Cov}(\mathbf{y}_{\mathbf{ik}})=\mathbf{R}=\begin{bmatrix} \sigma_s^2+\sigma_e^2 & \sigma_s^2 & \cdots & \sigma_s^2 \\ \sigma_s^2 & \ddots & \sigma_s^2 & \vdots \\ \vdots & \sigma_s^2 & \ddots & \sigma_s^2 \\ \sigma_s^2 & \cdots & \sigma_s^2 & \sigma_s^2+\sigma_e^2 \end{bmatrix} \tag{7}$$

Kukin lohko on symmetrinen $J\times J$-matriisi eli analysoitavassa havaintoaineistossa 3×3 -matriisi. Tasainen kovarianssirakenne (*CS*) on siis rakenne, jonka vallitessa saman koeyksilön eri mittaustuloksiin liittyvät varianssit $(Var(y_{ijk})=\sigma_s^2+\sigma_e^2)$ samoin kuin saman koeyksilön eri mittaustulosten väliset kovarianssit $(Cov(y_{ijk},y_{ij'k})=\sigma_s^2)$ ovat keskenään samanlaisia.

Edellä kuvatussa tasaisessa kovarianssirakenteessa lohkodiagonaalielementti on kuten kaavassa (7). Lohkodiagonaalielementti esitetään joskus myös muodossa $\mathbf{R}=(\mathbf{J}\rho+\mathbf{I}(1-\rho)\sigma_e^2)$, missä koeyksilöiden välinen korrelaatio on ρ eli

$$\rho=\frac{\sigma_s^2}{\sigma_s^2+\sigma_e^2}$$

ja $\mathbf{J}$ on ykkösiä täynnä oleva matriisi.

Käytettäessä sekamallia (kaava 4) ei tarvitse rajoittua tasaiseen kovarianssirakenteeseen, vaan on mahdollista tarkastella muitakin kovarianssirakenteita. Merkitään edelleen

$$\Sigma=\mathbf{Z}\Sigma_b\mathbf{Z}^T+\Sigma_e,$$

tällöin $\Sigma=diag(\mathbf{R})$.

Varianssi-kovarianssimatriisi $\mathbf{Cov(y)}$ mallille $\mathbf{y}=\mathbf{X}\alpha+\mathbf{Zb}+\mathbf{e}$ voidaan esittää

$$
Cov\begin{bmatrix} y_{11} \\ \vdots \\ y_{1n_1} \\ y_{21} \\ \vdots \\ y_{2n_2} \\ \vdots \\ y_{I1} \\ \vdots \\ y_{In_I} \end{bmatrix} = \begin{bmatrix} \mathbf{R} & \cdots & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots & & \vdots & \vdots \\ 0 & \cdots & \mathbf{R} & 0 & \cdots & 0 & \vdots \\ 0 & \cdots & 0 & \mathbf{R} & & \vdots & 0 \\ \vdots & \cdots & \vdots & 0 & \ddots & 0 & \vdots \\ \vdots & \cdots & \vdots & \vdots & & \mathbf{R} & 0 \\ 0 & \cdots & 0 & 0 & & 0 & \mathbf{R} \end{bmatrix}.
$$

Sekamallien (esimerkiksi SAS-ohjelmiston proseduuri *Proc Mixed*) avulla on mahdollista valita lohkodiagonaalielementille **R** muitakin kovarianssirakenteita kuin tasainen kovarianssirakenne (katso luku 3.3).

Normaalisuuden testaaminen

Tässä kappaleessa käydään lyhyesti läpi normaalisuusoletuksen testaaminen, mutta esimerkkiaineiston oletusten voimassaoloa ei voida testata pienen otoskoon vuoksi. Normaalisuusoletuksen testaustapa riippuu siitä, mihin kovarianssirakenteeseen päädytään, mutta tasaisen kovarianssirakenteen tilanteessa normaalisuus testataan tässä kappaleessa esitetyllä tavalla. Enemmän oletusten testaamiseen sopivista testeistä on teoksessa Encyclopedia of Statistical Sciences, jonka ovat toimittaneet Kotz, Johnson ja Read (painokset 1982 ja 1988).

Tutkielmassa olevia parametrisia menetelmiä sovelletaan oletuksella, että kovarianssirakenne on tasainen (*CS*). Tällöin siis yksilön satunnaistasoa kuvaava parametri S_{ik} ja satunnaisvaihtelua kuvaava parametri e_{ijk} ovat toisistaan riippumattomia ja parametrien S_{ik} ja e_{ijk} oletetaan olevan normaalijakautuneita odotusarvolla 0 ja vakiovariansseilla σ_s^2 ja σ_e^2 eli $S_{ik} \sim N(0,\sigma_s^2)$ ja $e_{ijk} \sim N(0,\sigma_e^2)$. Tutkittaessa normaalisuusoletusta on molemmat jäännökset testattava erikseen.

Jos oletukset ovat voimassa, voidaan parametrisia testejä pitää luotettavina ja oikeita tuloksia antavina. Vastaavasti tuloksia ei voida pitää luotettavina, jos oletukset eivät ole voimassa.

Yhden parametrin jäännösten normaalijakaumaoletusta voidaan tarkastella esimerkiksi Shapiron-Wilksin testillä (Royston, 1982). *SAS*-ohjelmisto käyttää kahdenlaista normaalisuustestiä riippuen otoskoosta: pienille otoksille ($n \leq 2000$) käytetään Shapiron-Wilksin testiä ja suurille otoksille ($n > 2000$) ohjelmistossa käytetään Kolmogorov-Smirnov -testiä. Esimerkkiaineiston tapauksessa normaalisuustestit eivät anna luotettavia tuloksia pienen otoskoon takia. Multinormaalisuutta voidaan testata esimerkiksi Mardian multinormaalisuustestillä (Mardia, 1985).

3 Parametrien estimointi ja testaus

Luvussa 3 tarkastellaan otsonikokeen aineiston mallittamista *SAS*-ohjelmiston version 6.12 *glm*- ja *mixed*-proseduurien avulla. Luvun alussa on lyhyesti kerrottu proseduurien ominaisuuksista ja tarkasteltu parametrisiin testeihin liittyviä oletuksia. Tämän jälkeen käsitellään sovitetulle mallille sopivan kovarianssirakenteen valintaa ja luvun lopussa on mallin parametrien estimointia parametrisilla menetelmillä ja tulkintaa.

3.1 Mallittaminen *glm*-proseduurilla

SAS-ohjelmiston *GLM*-proseduuri käyttää tavallista pienimmän neliösumman menetelmää (*ordinary least squares, OLS*) normaalijakaumavastemallin sovittamiseen. Proseduuria voidaan käyttää regressio-, varianssianalyysi-, kovarianssianalyysi-, moniulotteisen varianssianalyysi- eli MANOVA- (*multivariate analysis of variance*) ja osittaiskorrelaatio-ongelmien (*partial correlation*) tarkasteluun. Selittäjät voivat olla luokiteltuja tai jatkuvia; luokitellut ilmoitetaan `class`-lauseessa ja muut selittäjät oletetaan jatkuviksi. *GLM*-proseduurin sovittama malli (kaava 3) on $\mathbf{y} = \mathbf{X}\alpha + \mathbf{e}$ ja kovarianssirakenne on $\mathbf{Cov(y)} = \Sigma_e = \sigma_e^2\mathbf{I}$.

Estimaattien laskemisessa *GLM*-proseduuri käyttää kappaleessa 2.2 esiteltyä nurkkasolurajoitusta siten, että viimeinen selittäjien tasoista asetetaan vertailutasoksi. Tällöin estimaatit ja testit on laskettu asetettu nurkkasolurajoitus huomioiden.

Tässä tutkielmassa käytetty perusmalli *GLM*-proseduurilla on siis malli (3). Koska yksilö on mallissa kiinteänä tekijänä, joudutaan poikkeamaan oletuksesta, että yksilöt on satunnaisesti poimittu perusjoukosta ja yksilötasoa kuvaavat parametrit ovat riippumattomia ja noudattavat normaalijakaumaa odotusarvolla 0 ja varianssilla σ_s^2. Näiden oletusten ollessa voimassa voidaan estimoida yksilöiden välisiä (*between*) ja yksilöiden sisäisiä (*within*) tekijöitä sekä näiden välistä yhdysvaikutusta (*interaction*), kuten myöhemmin sovitettaessa yksilö satunnaisena voidaan tehdä.

Hiiret-aineistoa (katso liitteestä havaintomatriisi koe2.sd2) estimoitiin ensin ohjelmalla, joka on mallin (3) mukainen:

```
Title 'glm, fixed effects';
proc glm data=koe2;
        class ika trt yksilo jakso C2;
        model sl_time=ika yksilo(ika) trt jakso C2;
        estimate 'ilma vs. otsoni'    trt    1 -1;
        estimate 'nuori vs. aikuinen' ika    1 -1;
        estimate 'jakso1 vs. jakso2'  jakso  1  0 -1;
        estimate 'jakso1 vs. jakso3'  jakso  0  1 -1;
        estimate 'jakso2 vs. jakso3'  jakso -1  1  0;
        estimate 'carry1 vs. carry2'  C2    -1  1;
        contrast 'ilma vs. otsoni'    trt    1 -1;
        contrast 'nuori vs. aikuinen' ika    1 -1;
        contrast 'jakso1 vs. jakso2'  jakso  1  0 -1;
        contrast 'jakso1 vs. jakso3'  jakso  0  1 -1;
        contrast 'jakso2 vs. jakso3'  jakso -1  1  0;
        contrast 'carry1 vs. carry2'  C2    -1  1;
run;
```

Seuraavaksi mallista poistettiin viivästevaikutuksen parametri ja ohjelma ajettiin uudelleen. Molempien ohjelmien antamat estimaatit, testisuureet ja p-arvot ovat taulukossa 3.1 ja ohjelmat sekä aineistomatriisit ovat täydellisinä liitteessä.

3.2 Mallittaminen *mixed*-proseduurilla

SAS-ohjelmiston *proc mixed*-proseduuri on laajennus *proc glm*-proseduurista. *Mixed* ja *glm* -proseduurien ominaisuudet vastaavat osittain toisiaan ja suurin ero niiden välillä

on se, että *mixed* perustuu sekamallien teoriaan. Teoriataustaltaan *mixed* perustuu siis uskottavuuspäättelyyn, jossa maksimoidaan logaritminen uskottavuusfunktio $l(\Psi)$ (Brown ja Kempton, 1994). *Mixed* käyttää rajoitettua suurimman uskottavuuden menetelmää (*Restricted/Residual Maximum Likelihood* eli *REML*) tai suurimman uskottavuuden menetelmää (*Maximum Likelihood* eli *ML*).

Sekamalleja käsiteltiin aiemmin luvussa 2. Kiinteänä selittäjänä voidaan pitää esimerkiksi lääkkeen annostasoa, jos asetelmassa on käytetty vain joitain ennalta määrättyjä lääkeannoksia. Tällöin tietyllä annoskoolla saatuja tuloksia ei voida yleistää kaikkia saman lääkkeen annoskokoja koskeviksi. Satunnaisista selittäjistä esimerkkinä olkoon koeyksilö, jos kokeesta yhdellä tai useammalla otoksella saatuja tuloksia käytetään koko perusjoukkoa koskevina yleistyksinä. *Mixed*-proseduuria käytettäessä täytyy tietää kumpaan osaan selittäjä kuuluu, sillä kiinteän osan selittäjät annetaan `model`-lauseessa ja satunnaiset `random`-lauseessa. Jos `random`-lausetta ei käytetä, malli pelkistyy *glm*-proseduurin käyttämäksi kiinteiden tekijöiden malliksi.

Mixed parametrisoi mallin siten, että jokainen parametrin taso estimoidaan itsenäisesti. Proseduuri ei siis estimoi esimerkiksi käsittelyjen välistä eroa suoraan, vaan se estimoi parametrin tason saaman arvon ja tarvittaessa yksilökohtainen kokonaisestimaatti voidaan laskea parametrin tasojen estimaattien summana. Estimaatit kaikille tasoille saadaan lisäämällä `model`-lauseeseen komento `/solution`. Haluttaessa testata hypoteeseja käytetään `Contrast`-lausetta. `Estimate`-lause eroaa `Contrast`-lauseesta siten, että `Estimate`-lauseessa kontrastien käyttö on rajoitetumpaa.

Mixed-proseduurilla voidaan analysoida monenlaisia aineistoja. Tutkimuksen koeasetelmana voi olla esimerkiksi satunnaistettujen lohkojen koe, hierarkinen asetelma tai toistettujen mittausten asetelma.

Seuraavana perusproseduuri:

```
proc mixed data=koe;
class rokote potilas;
model kuume=rokote / solution;
random potilas;
run;
```

`Class`-lauseessa ovat luokitellut muuttujat, jotka voivat olla joko numeerisia tai merkkimuodossa. `Model`-lauseessa selitettävä muuttuja on kuume ja selittäjinä ovat rokote ja potilas. Koska potilaan ajatellaan kuuluvan mallin satunnaiseen osaan, sitä ei ole laitettu `model`-lauseeseen vaan satunnaisen selittäjän sisältävään `Random`-lauseeseen. Potilas on luokiteltu muuttuja, jonka luokkiin (potilaisiin) liittyvät tasot ovat otos perusjoukosta. Kokeen perusteella siis oletetaan, että kaikilla ihmisillä rokotteen aiheuttaman sivuvaikutuksen esiintyminen on samanlaista kuin kokeessa olleilla potilailla. `Repeated`-lausetta käytetään, kun halutaan estimoida koeyksilöiden sisäistä (*within*) vaihtelua eli vaihtelua saman koeyksilön eri mittaustulosten välillä ja laskea mallin kovarianssimatriisi. `Repeated`-lauseessa olevien muuttujien pitää olla luokiteltuja.

Mixed-proseduurin hyödyllisimpiin ominaisuuksiin kuuluvat `lsmeans`, `estimate` ja `contrast`-lauseet, joilla voidaan estimoida keskiarvoja (`lsmeans`) ja testata hypoteeseja (`contrast` ja `estimate`). `Lsmeans`-, `estimate`- ja `contrast`-lauseet ovat

```
lsmeans rokote/pdiff;
estimate 'rokote1 vs rokote2' rokote 1 -1 0 0;
contrast 'rokote1 vs rokote2' rokote 1 -1 0 0;
```

`Estimate`-lauseessa tutkittava ero on esimerkin tapauksessa rokotteiden 1 ja 2 vaikutusten ero. Kontrastien avulla tutkittavan nollahypoteesin mukaan rokotuksen 1 ja rokotuksen 2 vaikutuksilla ei ole eroa.

Mixed-proseduurin `Repeated`-lauseen avulla voidaan etsiä aineistoon sopiva kovarianssirakenne, jonka teoriaa on enemmän luvussa 2.5. Analysoitavaan aineistoon sovitetulle mallille oikean kovarianssirakenteen käyttäminen on tärkeää sen vuoksi, että oikealla kovarianssirakenteella hypoteesien testaus ja lasketut luottamusvälit ovat luotettavampia. *Mixed* laskee *REML*- ja *ML*-menetelmillä varianssiparametrien estimaatit. *Mixed*-proseduuri käyttää estimoinnissa oletusmenetelmänä rajoitettua suurimman uskottavuuden menetelmää (*REML*) ja tätä menetelmää on käytetty myös tässä tutkielmassa.

3.3 Kovarianssirakenteen sopivuuden tutkiminen

Yleensä sopivia kovarianssirakenteita ovat tasainen (yhtenäinen ja symmetrinen) *CS*-rakenne (*Compound Symmetry*), *rajoittamaton* (vapaa) *UN*-rakenne (*Unstructured*) ja *astetta 1 oleva autoregressiivinen AR(1)*-rakenne (*Autoregressive order one*). Tutkielmassa on lisäksi sovitettu mallia *CSH*-rakenteella (*Compound Symmetry Heterogeneous*). Näiden kovarianssirakenteiden lisäksi sopivan kovarianssirakenteen valinnan yhteydessä mallia sovitettiin muutamalla harvemmin sopivalla kovarianssirakenteella (SAS/STAT, Mixed Procedure, p.273-274).

3.3.1 Kovarianssirakenteita

CS on yleisemmin käytetty kovarianssirakenne. Siinä on kaksi tuntematonta parametria, toinen homogeenisen varianssin parametri ($\sigma_s^2 + \sigma_e^2$) ja toinen kovarianssin parametri (σ_s^2) (katso kaava 7). Matriisi Σ on lohkodiagonaalinen siten, että kussakin lohkossa ovat yhtä koeyksilöä koskevat varianssit ja kovarianssit.

UN-rakenne tarkoittaa sitä, että rajoittavaa kovarianssirakennetta ei ole. Kaikilla koeyksilöillä on sama lohkodiagonaalielementti $\mathbf{R}$, joka voidaan määritellä. Tällöin jokaisella mittauksella on oma varianssi ja kovarianssi. Käytettäessä sopivimman kovarianssirakenteen etsimisessä loguskottavuuden arvoja (`-2 Res Log Likelihood`), sovitetun kovarianssirakenteen saamaa loguskottavuutta verrataan rajoittamattoman kovarianssirakenteen (*UN*) loguskottavuuteen (katso kappale 3.3.3). Tässä aineistossa *UN*-rakennetta käytettäessä on kuusi tuntematonta kovarianssirakenteeseen liittyvää parametria, jotka ovat eri jaksoihin liittyvät varianssit sekä jaksojen väliset kovarianssit.

AR(1) on sopiva kovarianssirakenne silloin, kun korrelaatio vähenee mittausten välisen ajan lisääntyessä eli tällöin aika-akselilla lähellä toisiaan lähellä olevilla mittauksilla on suurempi korrelaatio kuin kauempana olevilla. *AR(1)*-rakenne määrittelee kovarianssin muodossa $\sigma^2\rho^w$, missä w on mittausten välisten ajanjaksojen lukumäärä ja σ^2 on havainnon varianssi eli $Var(y_{ijk}) = \sigma^2$. Parametri ρ tarkoittaa saman koeyksilön aika-akselin vierekkäisten havaintojen välistä korrelaatiota eli ensimmäisen ja toisen mittauksen välinen korrelaatio on ρ ja esimerkiksi ensimmäisen ja neljännen

mittauksen välinen korrelaatio on ρ^3. *AR(1)*-rakenteeseen liittyy kaksi tuntematonta parametria.

CSH on kovarianssirakenteen *CS* suora yleistys kuten *ARH(1)* on kovarianssirakenteelle *AR(1)*. *CSH*-rakenteella on vakiokorrelaatiot, mutta heterogeeniset varianssit. Tässä aineistossa *CSH*-rakenteella on neljä $(J+1)$ tuntematonta parametria.

3.3.2 Ohjelmointi *proc mixed*-proseduurilla

Vertailun vuoksi ensin on esitetty ohjelma, jossa ei määrätä mitään kovarianssirakennetta. Tälöin *mixed*-proseduuri olettaa kovarianssirakenteen olevan *compound symmetry* (CS) eli tasainen kovarianssirakenne.

```
Title ' Mixed - random ';
proc mixed data=koe2 order=internal;
        class ika trt yksilo jakso C2;
        model sl_time=ika trt jakso C2;
        random yksilo;
run;
```

Lause `order=internal` määrittelee järjestyksen, johon luokiteltujen muuttujien tasot järjestetään. Vaihtoehto `internal` määrittelee, että kunkin muuttujan tasot järjestetään siten, että pienin tasoa kuvaava numero tai aakkosten ensimmäinen tasoa kuvaava merkki (taso aikuinen ennen tasoa nuori) on ensimmäisenä.

Sopivimman kovarianssirakenteen löytämiseksi mallia ajettiin seuraavalla ohjelmalla tutkittavan kovarianssirakenteen ollessa `type= CS, UN, CSH` tai `AR(1)`, jolloin mallitetaan kappaleessa 2.5 esitettyä lohkodiagonaalielementtiä **R**. Tasaisella kovarianssirakenteella (*CS*) saadaan täsmälleen sama malli ja kovarianssirakenne kuin edellä olleella ohjelmalla.

```
CS, UN, CSH, AR(1):

Title ' Mixed: kovarianssirakenne=CS';
Proc mixed data=koe2;
       Class ika trt yksilo jakso C2;
       Model sl_time=ika trt jakso C2 / ddfm=satterth;
       Repeated jakso / subject=yksilo(ika) type= CS r rcorr; Run;
```

Tulokset esitetään kappaleessa 3.3.3 taulukossa 3.1.

3.3.3 Informaatiokriteerit

Akaiken (*AIC*) ja Schwarz-Bayesin (*SBC*) informaatiokriteerien avulla voidaan valita aineistoon sopivin kovarianssirakenne. *AIC*- (*Akaike's Information Criterion*) ja *SBC*- (*Schwarz Bayesian Criterion*) eli *BIC*- (*Bayesian Information Criterion*) arvot ovat loguskottavuuden arvoja, joihin on lisätty estimoitujen parametrien määrästä riippuva rankaisutermi. Olkoon Ψ tuntemattomien kovarianssiparametrien vektori estimointimenetelmän ollessa REML-menetelmä. Lisäksi olkoon $l(\Psi)$ logaritminen uskottavuusfunktio.

Informaatiokriteerit on määritelty (Smith, 1988):

$AIC = l(\hat{\Psi}) - z$, missä z on estimoitavien varianssiparametrien lukumäärä

$SBC = l(\hat{\Psi}) - \frac{1}{2} z \log n$, missä n on havaintojen lukumäärä (otoskoko).

Kriteerejä tulkitaan siten, että suurimman arvon tuottava kovarianssirakenne on paras.

Kovarianssirakenteen valinnassa voidaan käyttää myös maksimoidun uskottavuusfunktion (*ML*) logaritmin arvoja ($l(\hat{\Psi})$ tai $-2l(\hat{\Psi})$). Tällöin erilaisilla kovarianssirakenteilla saatuja loguskottavuuden arvoja verrataan rajoittamattomalla (*UN*) kovarianssirakenteella saatuun arvoon siten, että loguskottavuuden arvojen ($-2l(\hat{\Psi})$) erotus on χ^2 -jakautunut vapausasteilla, jotka saadaan tutkittavalla ja rajoittamattomalla kovarianssirakenteella sovitetun mallin parametrien erotuksena. Tällöin nollahypoteesina on, että vertailtava kovarianssirakenne sopii yhtä hyvin kuin rajoittamaton kovarianssirakenne. Michael Kenwardin (*The Analysis of Repeated Measurements and Longitudinal Data* –kurssi) mukaan AIC- ja SBC-kriteerit suosittelevat liian yksinkertaista kovarianssirakennetta ja hänen mukaansa kannattaa käyttää loguskottavuuden arvoja tehtäessä päätelmiä kovarianssirakenteen sopivuudesta. Tässä tutkielmassa sopivin kovarianssirakenne on kuitenkin valittu informaatiokriteerien perusteella.

Mallin sovittamisen tuottamat loguskottavuuden arvot, AIC- ja SBC-informaatiokriteerien arvot ovat taulukossa 3.1. Täydelliset ohjelmat ovat liitteenä.

Taulukko 3.1 Logaritmisen uskottavuuden, *AIC*- ja *SBC*-kriteerien arvot sovitettaessa mallia *UN*-, *CS*-, *CSH*- ja *AR(1)* -kovarianssirakenteilla.

	SBC	*AIC*	$-2l(\hat{\Psi})$	parametrien lkm
UN	−178.902	−173.107	334.2139	6
CS	−174.266	−172.334	340.6685	2
CSH	−175.435	−171.571	335.1418	4
AR(1)	−175.084	−173.152	342.3042	2

AIC-kriteerin mukaan sovitetulla mallille paras kovarianssirakenne oli *CSH*. Melko lähellä *CSH*:n saamaa *AIC*-arvoa oli myös kovarianssirakenteen *CS* saama *AIC*-arvo. *SBC*:n mukaan sopivin kovarianssirakenne oli *CS* ja muista rakenteista *AR(1)*:n saama arvo oli lähinnä rakenteen *CS* arvoa. *AIC*- ja *SBC*-kriteerien perusteella ja myös rakenteen tulkittavuuden vuoksi tässä tutkielmassa käytetylle mallille kovarianssirakenteeksi valittiin tasainen kovarianssirakenne (*CS*). Toinen vaihtoehto kovarianssirakenteeksi näiden kriteerien perusteella olisi *AR(1)*, mutta aikasarjatyyppinen kovarianssirakenne ei tulkinnallisesti ole sopiva tähän malliin. Havaintojen keskinäisen riippuvuuden ei voida ajatella muuttuvan sen mukaan, mitä pitempi mittausten välinen aika on. Esimerkiksi ei ole mitään syytä olettaa, että jaksojen 1 ja 2 korrelaatio olisi suurempi kuin jaksojen 1 ja 3 mittausten välinen korrelaatio.

3.4 Kovarianssirakenteen kirjoittaminen matriisimuotoon

Kovarianssimatriisi kullekin lohkolle (koeyksilölle) saadaan esimerkiksi tulostuksen kohdasta `R Matrix for YKSILO(IKA) 10 1`. Tässä ensimmäinen numero viittaa muuttujan YKSILO kymmenenteen tasoon eli tässä kymmenenteen koeyksilöön ja toinen numero muuttujan IKA luokkaan 1 (katso liitteestä koe2.sd2).

Koko aineiston kovarianssimatriisi on muotoa (katso kappale 2.5)

$$\Sigma = \begin{bmatrix} \mathbf{R} & 0 & \cdots & 0 \\ 0 & \mathbf{R} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \mathbf{R} \end{bmatrix},$$

missä kovarianssimatriisi **R** on sama kaikilla koeyksilöillä

$$\mathbf{R} = \begin{bmatrix} 38.382 & 12.863 & 12.863 \\ 12.863 & 38.382 & 12.863 \\ 12.863 & 12.863 & 38.382 \end{bmatrix},$$

missä riveinä ja sarakkeina ovat jaksot j. Tulostuksesta kohdasta `Covariance Parameter Estimates (REML)` koeyksilön kahden mittauksen välisen kovarianssin estimaatiksi $\hat{\sigma}_s^2$ saatiin 12.863 ja mittauksen yksilötason varianssin estimaatiksi $\hat{\sigma}_s^2 + \hat{\sigma}_e^2$ saatiin 38.382. Näistä laskemalla saatiin yksilön virhevaihtelun estimaatti $\hat{\sigma}_e^2$=38.382-12.863=25.519.

Korrelaatio kahden satunnaismuuttujan, merkitään W ja Z, välillä on $r_{w,z} = Cov(w,z) / (Var(w)Var(z))^{1/2}$. Tätä kaavaa soveltamalla saatiin saman koeyksilön kahden mittauksen välinen korrelaatio, joka tässä aineistossa oli $r = 12.863 / (38.382 \cdot 38.382)^{1/2} = 0.335$. Korrelaatiot ovat tulostuksen kohdassa `R Correlation Matrix for YKSILO(IKA) 10 1` ja ne ovat oletusten perusteella samat kaikille koeyksilöille. Siis kahden mittauskerran välinen korrelaatio oli 0.335 mittausajankohdasta riippumatta, eli mittausten välillä ei juurikaan ollut korrelaatiota.

3.5 Estimointi

Estimaatit ja niiden luottamusvälit sekä testisuureet ja p-arvot on laskettu seuraavilla ohjelmilla

```
Title 'glm, fixed effects';
proc glm data=koe2;
        class ika trt yksilo jakso C2;
        model sl_time=ika yksilo(ika) trt jakso C2;
        estimate 'ilma vs. otsoni'    trt    1 -1;
        estimate 'nuori vs. aikuinen' ika    1 -1;
        estimate 'jakso1 vs. jakso2'  jakso  1  0 -1;
        estimate 'jakso1 vs. jakso3'  jakso  0  1 -1;
        estimate 'jakso2 vs. jakso3'  jakso -1  1  0;
        estimate 'carry1 vs. carry2'  C2    -1  1;
        contrast 'ilma vs. otsoni'    trt    1 -1;
        contrast 'nuori vs. aikuinen' ika    1 -1;
        contrast 'jakso1 vs. jakso2'  jakso  1  0 -1;
        contrast 'jakso1 vs. jakso3'  jakso  0  1 -1;
```

```
        contrast 'jakso2 vs. jakso3'  jakso -1  1  0;
        contrast 'carry1 vs. carry2'  C2    -1  1; run;

Title 'mixed, fixed effects';
proc mixed data=koe2;
        class ika trt yksilo jakso C2;
        model sl_time=ika yksilo(ika) trt jakso C2 / solution
        ddfm=satterth;
        estimate 'ilma vs. otsoni'    trt    1 -1    / cl;
        estimate 'nuori vs. aikuinen' ika    1 -1    / cl;
        estimate 'jakso1 vs. jakso2'  jakso  1  0 -1 / cl;
        estimate 'jakso1 vs. jakso3'  jakso  0  1 -1 / cl;
        estimate 'jakso2 vs. jakso3'  jakso -1  1  0 / cl;
        estimate 'carry1 vs. carry2'  C2    -1  1    / cl;
        contrast 'ilma vs. otsoni'    trt    1 -1;
        contrast 'nuori vs. aikuinen' ika    1 -1;
        contrast 'jakso1 vs. jakso2'  jakso  1  0 -1;
        contrast 'jakso1 vs. jakso3'  jakso  0  1 -1;
        contrast 'jakso2 vs. jakso3'  jakso -1  1  0;
        contrast 'carry1 vs. carry2'  C2    -1  1; run;


Title 'mixed,random subject';
proc mixed data=koe2;
        class ika trt yksilo jakso C2;
        model sl_time=ika trt jakso C2 / solution ddfm=satterth;
        random yksilo(ika);
        estimate 'ilma vs. otsoni'    trt    1 -1    / cl;
        estimate 'nuori vs. aikuinen' ika    1 -1    / cl;
        estimate 'jakso1 vs. jakso2'  jakso  1  0 -1 / cl;
        estimate 'jakso1 vs. jakso3'  jakso  0  1 -1 / cl;
        estimate 'jakso2 vs. jakso3'  jakso -1  1  0 / cl;
        estimate 'carry1 vs. carry2'  C2    -1  1    / cl;
        contrast 'ilma vs. otsoni'    trt    1 -1;
        contrast 'nuori vs. aikuinen' ika    1 -1;
        contrast 'jakso1 vs. jakso2'  jakso  1  0 -1;
        contrast 'jakso1 vs. jakso3'  jakso  0  1 -1;
        contrast 'jakso2 vs. jakso3'  jakso -1  1  0;
        contrast 'carry1 vs. carry2'  C2    -1  1; run;
```

Ensin ohjelma ajettiin *glm*-proseduurilla, jossa kaikki malliin (malli 3) mukaan otetut tekijät ovat kiinteitä. Toisena ohjelmana oli *mixed*-proseduuri, jolla sovitetulla mallilla (3) myös kaikki tekijät ovat kiinteitä. Vertailun vuoksi kolmannessa ohjelmassa on mukana satunnainen tekijä (malli 4) ja muuten se on vastaava ohjelma kuin ensimmäisenä ajettu *mixed*-proseduurin ohjelma. `Random`-lauseeseen on laitettu muuttuja yksilo, koska muuttujan yksilo vaikutuksen eli koeyksilön vaihtelun oletetaan olevan satunnaista (*random*). Tällä lauseella voidaan malliin sisällyttää yksilön satunnaisvaihtelu ja estimoida yksilötasolla mittausten välinen kovarianssi σ_S^2 . Lisäksi vastaavilla ohjelmilla ajettiin mallit, joista oli poistettu viivästevaikutuksen parametri.

Estimaatit ja niiden 95 %:n luottamusvälit, keskivirheet ja p-arvot ovat taulukossa 3.2 ja täydelliset ohjelmat ja aineistomatriisit ovat liitteenä. Luvussa 4 siirrytään estimoimaan ja testaamaan kontrasteilla ja tällöin tilanne oletetaan mallin (3) mukaiseksi.

3.6 Tulokset ja niiden tulkinta

Taulukossa 3.2 ovat estimaatit, keskivirheet, p-arvot ja 95 %:n luottamusvälit. Ensin ovat tulokset, jotka saatiin sovittamalla mallia (3) (Canada, Hafner ja Koch, 1988) ja mallia (4) (Gough, 1988), joissa oli mukana viivästevaikutuksen parametri. Näiden mallien sovittamisesta saatujen tulosten jälkeen ovat suluissa tulokset, jotka saatiin sovittamalla kutakin mallia ilman viivästevaikutuksen parametria.

Estimoinnin eroista

Käytettäessä pienimmän neliösumman menetelmän (pns-menetelmä) tilalla rajoitettua suurimman uskottavuuden menetelmää eli REML-menetelmää, voidaan estimoinnissa käyttää koeyksilöiden välistä (*between subject*) tietoa (Gough, 1989). *GLM*-proseduuri ja *Mixed*-proseduuri (kiinteät tekijät, malli 3) käyttävät estimoinnissa pns-menetelmää ja *Mixed*-proseduuri (kiinteät tekijät ja satunnainen tekijä, malli 4) käyttää estimoinnissa myös REML-menetelmää (katso kappale 3.2). REML-menetelmää käytettäessä estimoidaan kiinteät selittäjät yleistetyllä pienimmän neliösumman menetelmällä (*GLS, generalized least squares*) ja vain satunnaisosa estimoidaan rajoitetulla suurimman uskottavuuden menetelmällä. Gough on estimoinut REML-menetelmällä tässä tutkielmassa käytettyä esimerkkiaineistoa ja päätynyt samoihin tuloksiin kuin taulukossa 3.2 esitetään.

Vaihtovuoroasetelmat ovat usein epätasapainoisia eli jaksoja on vähemmän kuin käsittelyjä tai aineistossa on puuttuvia havaintoja. Koska esimerkkiaineiston asetelma ei ole tasapainoinen, REML pystyy tehokkaampaan estimointiin. Kun yksilö määritellään satunnaiseksi, REML-menetelmä pystyy tehokkaampaan estimointiin, koska se käyttää myös koeyksilöiden välisen tiedon (*between-subject*) eli kaiken mahdollisen tiedon. Määriteltäessä yksilö kiinteäksi REML-menetelmä palautuu pienimmän neliösumman menetelmäksi.

Taulukko 3.2 Estimaatit, keskivirheet, p-arvot ja 95 %:n luottamusvälit proseduureilla *GLM* ja *Mixed*. Suluissa olevat arvot ovat mallille, josta on poistettu viivästevaikutuksen termi. Parametri-sarakkeessa on ensin selittäjän nimi aineistossa ja sitten estimoitu ero.

	Malli (3)					**Malli (4)**		
	GLM		***MIXED***			***MIXED***		
Parametri	Estimaatti se	p-arvo	Estimaatti se	[CI]	p-arvo	Estimaatti se	[CI]	p-arvo
Ikä : Aikuinen-nuori	9.53 (9.53) 1.344 (1.324)	0.0001 (0.0001)	9.53 (9.53) 2.849 (1.324)	[6.803, 12.265] ([6.843, 12.218])	0.0001 (0.0001)	9.54 (9.53) 2.125 (2.120)	[5.059, 14.026] ([5.060, 14.007])	0.0003 (0.0003)
Käsittely: Otsoni-ilma	2.68 (2.53) 2.849 (1.404)	0.3533 (0.0803)	2.68 (2.53) 2.849 (1.404)	[-3.108, 8.469] ([-0.321, 5.379])	0.3533 (0.803)	3.14 (2.67) 2.595 (1.370)	[-2.085, 8.372] ([-0.108, 5.438])	0.2323 (0.0591)
Jakso: Jakso2-jakso1	-2.79 (-2.69) 2.295 (1.620)	0.2322 (0.1054)	-2.79 (-2.69) 2.295 (1.620)	[-7.456, 1.872] ([-5.983, 0.597])	0.2322 (0.1054)	-3.01 (-2.68) 2.224 (1.620)	[-7.513, 1.492] ([-5.974, 0.602])	0.1838 (0.1061)
Jakso: Jakso3-jakso1	-3.98 (-3.88) 2.264 (1.619)	0.0877 (0.0219)	-3.98 (-3.88) 2.264 (1.619)	[-8.581, 0.621] ([-7.171, -0.598])	0.0877 (0.0219)	-4.20 (-3.88) 2.192 (1.618)	[-8.637, 0.238] ([-7.169, -0.600])	0.0629 (0.0218)
Jakso: Jakso3-jakso2	-1.18 (-1.19) 1.645 (1.620)	0.4749 (0.4673)	-1.19 (-1.19) 1.645 (1.620)	[-4.531, 2.154] ([-4.481, 2.010])	0.4749 (0.4673)	-1.19 (-1.19) 1.641 (1.620)	[-4.523, 2.146] ([-4.486, 2.090])	0.4740 (0.4643)
Viivästevaikutus: Otsoni-ilma	0.20 3.289	0.9513	0.20 3.289	[-6.4822, 6.8866]	0.9513	0.67 3.072	[-5.532, 6.863]	0.8296

Tulosten tulkinta

Tulokset malleille (3) ja (4) ilman viivästevaikutuksen parametria on esitetty siten, että ensin ovat malleilla (3) ja (4) saadut tulokset ja sitten vastaavasti suluissa ilman viivästevaikutuksen parametria saadut tulokset.

Ensimmäisenä tutkitaan, onko viivästevaikutus merkitsevää. Nollahypoteesille $H_0: C_2 - C_1 = 0$ eli otsonin ja ilman viivästevaikutuksella ei ole eroa, *glm* ja *mixed*

(malli 3) antavat saman F-testisuureen 0.00 (0.05) p-arvolla 0.9513 (0.8296). Käsittelyjen viivästevaikutukset eivät tässä tapauksessa näyttäisi eroavan ja nollahypoteesi jää voimaan merkitsevyystasolla 0.05. Mallien (3) ja (4) testisuureet eroavat hieman, mutta tässä ero ei vaikuta lopputulokseen. Otsonin ja ilman viivästevaikutusten eron estimaatti on 0.20 (0.67) minuuttia ilman viivästevaikutuksen ollessa vertailutasona mallin (3) tapauksessa. Tulosten perusteella voidaan jatkossa käyttää mallia, josta viivästevaikutuksen termi on poistettu.

Mallista, jossa viivästevaikutuksen parametri ei ole mukana, saadaan F-testisuure 3.25 (3.78) testattaessa nollahypoteesia $H_0: F_2 - F_1 = 0$ eli tutkittaessa onko otsonikäsittelyllä vaikutusta ilmakäsittelyn ollessa vertailutasona mallin (3) tapauksessa. Tällöin p-arvo on 0.0803 (0.0591) eli nollahypoteesia ei voida hylätä merkitsevyystasolla 0.05. Mallilla (4) saatu tulos on hyvin lähellä merkitsevää tulosta ja tämän perusteella voidaan olettaa otsoni- ja ilmakäsittelyn vaikutusten erotuksen eroavan jonkin verran nollasta. Käsittelyjen välisen eron estimaatiksi saadaan 2.53 (2.67) minuuttia 95 %:n luottamusvälin ollessa [-0.321, 5.379] ([-0.108, 5.438]).

Testattaessa parametrin ika vaikutusta saatiin F-testisuureeksi 51.83 (20.22) p-arvolla 0.0001 (0.0003). Nollahypoteesina oli, etteivät aikuisten ikäryhmän nukkumisajat eroa vertailutasona olleesta nuorten hiirien ikäryhmän nukkumisajoista mallin (3) tapauksessa. Nollahypoteesi hylätään merkitsevyystasolla 0.05 eli nukkumisaikojen erotus ikäryhmittäin eroaa nollasta. Nukkumisajan eron estimaatiksi ikäryhmittäin saatiin 9.53 (9.53) minuuttia vertailutason ollessa nuorten ryhmä mallilla (3) ja 95 %:n luottamusväliksi saatiin [6.843, 12.218] ([5.060, 14.007]).

Mallia 3 sovitettaessa jaksoista vertailutasoksi asetettiin jakso 1. Tutkittaessa jaksojen 1 ja 2 eroa (mallissa 3) F-testisuureen arvoksi saatiin 2.76 (2.75) p-arvon ollessa 0.1054 (0.1061) eli jaksot 1 ja 2 eivät eroa toisistaan tilastollisesti merkitsevästi. Estimaatiksi saatiin –2.69 (-2.69) 95 %:n luottamusvälin ollessa [-5.983, 0.597] ([-5.974, 0.602]). Jaksojen 1 ja 3 eron F-testisuureeksi saatiin 5.76 (5.76) p-arvon ollessa 0.0219 (0.0218). Nollahypoteesi, etteivät jakson 3 nukkumisajat eroa vertailutason jakson 1 nukkumisajoista mallissa (3), hylätään merkitsevyystasolla 0.05 estimaatin 95 %:n luottamusvälin ollessa [-7.171, -0.598] ([-7.169, -0.600]). Testattaessa jaksojen 2 ja 3

välistä eroa F-testisuureeksi saatiin 0.54 (0.55) p-arvolla 0.4673 (0.4643). Tämän testin perusteella jaksojen 2 ja 3 nukkumisajat eivät eroa toisistaan tilastollisesti merkitsevästi merkitsevyystason ollessa 0.05. Estimaatiksi saatiin -1.19 (-1.19) 95 %:n luottamusvälin ollessa [-4.481, 2.010] ([-4.486, 5.32]).

Yksilökohtaiset estimaatit

Seuraavaksi lasketaan esimerkkinä yksilöiden 3 ja 17 nukkumisajan kokonaisestimaatit jaksolla 3 ja jaksolla 1 käyttäen mallilla (3) hyväksi *proc mixed*in tulostuksen kohtaa `Solution for Fixed Effect`. Tulokset tarkastettiin ratkaisemalla *SAS*-ohjelmistolla yksilökohtaiset ennusteet, jotka saatiin lisäämällä ohjelmiin model-lauseen optio `/predicted`. Mallia (4) sovittamalla saadut estimaatit on otettu suoraan predicted-tulostuksesta.

Yksilö 3 on ikäryhmässä nuori ja sille annettiin jaksolla 3 otsonikäsittely. Estimaatin laskemiseksi käytetään tuloksia, jotka saatiin sovittamalla mallia ilman viivästevaikutuksen parametria. Tämän perusteella kokonaisestimaatiksi kiinteiden tekijöiden mallilla saadaan (vakio+ika+trt+yksilo(ika)+jakso)

$16.5495 + 0 + 2.5289 + 15.4570 - 3.8842 \approx 30.7$.

Vastaavasti mallilla (4), jossa on mukana satunnainen tekijä, yksilölle 3 kokonaisestimaatiksi saadaan tulostuksesta 26.8.

Havaittu arvo on 30.2 eli tässä mallin (3) mukainen yksilökohtainen estimaatti on lähes oikea ja mallin (4) mukainen estimaatti eroaa melko paljon oikeasta arvosta.

Yksilö 17 on ikäryhmässä aikuinen ja sille annettiin jaksolla 1 ilmakäsittely. Kokonaisestimaatti lasketaan samalla tavalla kuin yksilölle 3 laskettiin ja estimaatiksi saadaan kiinteiden tekijöiden mallilla (3)

$16.5495 + 9.9000 + 0 + 9.9000 + 0 = 36.3$.

Vastaavasti mallilla (4) yksilölle 17 saadaan tulostuksesta kokonaisestimaatiksi 34.4.

Havaittu arvo on 32.1 eli tässä satunnaisen tekijän avulla saaduista estimaateista laskettu kokonaisestimaatti on melko lähellä todellista mittaustulosta.

4 Kontrastimuuttujien käyttö estimoinnissa ja testauksessa

Tässä tutkielmassa kontrastimuuttujalla tarkoitetaan yksilön mittaustulosten $y_{i1k}, y_{i2k}, \ldots, y_{iJk}$ lineaarikombinaatiota $c_1 y_{i1k} + \ldots + c_J y_{iJk}$ käyttäen kertoimia $c_1, \ldots, c_J$, joiden summa on 0. Kontrastimuuttujia tutkitaan mallilla (3). Käytettäessä parametrien estimoinnissa ja testauksessa kontrastimuuttujia, kokonaiskeskiarvoja mallissa kuvaava termi (μ) ja koeyksilön satunnaistason parametri (S_{ik}) kumoutuvat lausekkeista pois (taulukko 2.2).

4.1 Määritelmä ja ominaisuuksia

Oletetaan, että vektorit **c** ja **y** ovat $J-$ulotteisia sarakevektoreita ja **y** sisältää tietyn yksilön tutkittavan vasteen arvot eri jaksoilla. Oletetaan, että kerroinvektorin **c** tekijöiden summa on nolla eli

$$\sum_{j=1}^{J} c_j = c_1 + c_2 + \ldots + c_J = 0 .$$

Tällöin kontrastimuuttuja u on vektoreiden **c** ja **y** pistetulo eli

$$u = \underset{1\times J}{\mathbf{c}^T} \underset{J\times 1}{\mathbf{y}} = (c_1 \quad c_2 \quad \cdots \quad c_J)(y_1 \quad y_2 \quad \cdots \quad y_J)^T = \sum_{j=1}^{J} c_j y_j = c_1 y_1 + c_2 y_2 + \ldots + c_J y_J .$$

Ideana on tällöin etsiä oma kerroinvektori **c** erikseen kullekin tutkittavalle parametrille ja muodostaa näistä kerroinmatriisi **C**.

Esimerkiksi havaintoaineistossa hiiret on kolme jaksoa ($j = 1, 2, 3$), jolloin vektori **c** on 3×1 -sarakevektori eli $\mathbf{c}^T = (c_1\ c_2\ c_3)$. Tällöin tutkittavan vasteen arvot sisältävä vektori **y** on 3×1 -sarakevektori eli $\mathbf{y} = (y_1\ y_2\ y_3)^T$. Kontrastimuuttuja u on vektorien $\mathbf{c}^T$ ja **y** pistetulo eli

$$\underset{1\times 3}{\mathbf{c}^T} \underset{3\times 1}{\mathbf{y}} = (c_1 \quad c_2 \quad c_3)(y_1 \quad y_2 \quad y_3)^T = \sum_{j=1}^{3} c_j y_j = c_1 y_1 + c_2 y_2 + c_3 y_3 ,$$

missä kertoimet toteuttavat ehdon $\sum_{j=0}^{3} c_j = c_1 + c_2 + c_3 = 0$.

Nyt oletetaan, että $\mathbf{y}$ noudattaa normaalijakaumaa odotusarvolla μ ja tasaisella kovarianssirakenteella Σ eli

$$\mathbf{y} \sim N(\mu, \Sigma)\text{ ,missä } \Sigma = \sigma_e^2\mathbf{I} + \sigma_s^2\mathbf{J} = \Sigma_e + \Sigma_b \tag{5}$$

Kun $\mathbf{y}$ kerrotaan kontrastivektorin transpoosilla $\mathbf{c}^T$, on kovarianssirakenteen koeyksilön satunnaistasoon liittyvä termi nolla, koska $\mathbf{c}^T\mathbf{J} = 0$. Tällöin kontrastimuuttuja eli $\mathbf{c}^T\mathbf{y}$ noudattaa normaalijakaumaa

$$N(\mathbf{c}^T\mu, \sigma_e^2\mathbf{c}^T\mathbf{c}) = N(\mathbf{c}^T\mu, \sigma_e^2\mathbf{I}) \tag{6}$$

Kaavassa (6) oleva kovarianssirakennetta koskeva tulos on vastaava kuin aiemmin mallille (3) saatu. Siis kontrastimuuttujan $\mathbf{c}^T\mathbf{y}_{ik}$ jakauma ei riipu kokonaiskeskiarvosta μ eikä koeyksilön satunnaistasosta S_{ik} (katso kaava 7).

Esimerkiksi valitaan kontrasti $c = (-1 \quad 1 \quad 0)$ eli tutkittava kontrastimuuttuja on $\mathbf{u}_{ik} = \mathbf{c}^T\mathbf{y}_{ik} = \mathbf{y}_{i2k} - \mathbf{y}_{i1k}$. Jos ei oleteta summarajoitusta eikä nurkkasolurajoitusta, saadaan kontrastimuuttujan $\mathbf{u}_{ik}$ odotusarvot (katso taulukko 2.2)

$$\begin{cases} u_{1k} = (P_2 - P_1) + (F_2 - F_1) + (e_{12k} - e_{11k}) \\ u_{2k} = (P_2 - P_1) - (F_2 - F_1) + C_2 + (e_{22k} - e_{21k}) \\ u_{3k} = (P_2 - P_1) + (F_2 - F_1) + (e_{32k} - e_{31k}) \\ u_{4k} = (P_2 - P_1) - (F_2 - F_1) + C_2 + (e_{42k} - e_{41k}) \end{cases} \tag{7}.$$

ja $\qquad Var(u_{ik}) = 2\sigma_e^2.$

Yksilöllisten kontrastimuuttujien kovarianssit

Ryhmän i yksilön k kontrastimuuttujan $p\,(p = 1, \ldots, P)$, missä P on tutkittavien vaikutusparametrien määrä, varianssi on $Var(u_{ipk}) = \sigma_e^2(c_{p1}^2 + c_{p2}^2 + \ldots + c_{pJ}^2) = \sigma_e^2\mathbf{c}_p^T\mathbf{c}_p$ ja kontrastimuuttujien u_{ipk} ja u_{iqk} väliset kovarianssit ovat $Cov(u_{ipk}, u_{iqk}) = \sigma_e^2(\mathbf{c}_{p1}\mathbf{c}_{q1} + \mathbf{c}_{p2}\mathbf{c}_{q2} + \ldots + \mathbf{c}_{pJ}\mathbf{c}_{qJ}) = \sigma_e^2\mathbf{c}_p^T\mathbf{c}_q$, joten yksilöllisen kontrastimuuttujan u_{ik} kovarianssimatriisi on

$$\mathbf{Cov}(\mathbf{u}_{ik}) = \sigma_e^2 \begin{bmatrix} c_1^T c_1 & c_1^T c_2 & c_1^T c_3 & \cdots & c_1^T c_p \\ c_1^T c_2 & c_2^T c_2 & c_2^T c_3 & \cdots & c_2^T c_p \\ c_1^T c_3 & c_2^T c_3 & c_3^T c_3 & & \vdots \\ \vdots & \vdots & \vdots & \ddots & \\ c_1^T c_p & c_2^T c_p & c_3^T c_p & \cdots & c_p^T c_p \end{bmatrix} = \sigma_e^2 \mathbf{C}\mathbf{C}^T, \text{kun } i = 1, \ldots, I.$$

Esimerkkiaineiston kovarianssimuuttujan $\mathbf{u}_{1k}$ kovarianssimatriisi on

$$Var(\mathbf{u}_{1k}) = \sigma_e^2 \begin{bmatrix} c_1^T c_1 & c_1^T c_2 \\ c_1^T c_2 & c_2^T c_2 \end{bmatrix} = \sigma_e^2 \begin{bmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \end{bmatrix} \begin{bmatrix} c_{11} & c_{21} \\ c_{12} & c_{22} \\ c_{13} & c_{23} \end{bmatrix} = \sigma_e^2 \mathbf{C}\mathbf{C}^T.$$

Ortogonaaliset kontrastit

Määritellään kontrastimuuttujat $\mathbf{u}_1$ ja $\mathbf{u}_2$ seuraavasti

$$\mathbf{u}_1 = \mathbf{c}_1^T \mathbf{y} = \sum_{j=1}^{J} c_{1j} y_j$$

ja

$$\mathbf{u}_2 = \mathbf{c}_2^T \mathbf{y} = \sum_{j=1}^{J} c_{2j} y_j$$

Kontrastimuuttujat $\mathbf{u}_1$ ja $\mathbf{u}_2$ ovat ortogonaalisia, jos kerroinvektorien elementtien pistetulo on nolla eli

$$\sum_{j=1}^{J} c_{1j} c_{2j} = \mathbf{c}_1^T \mathbf{c}_2 = 0.$$

Tällöin kaksiulotteinen kontrastimuuttuja noudattaa kaksiulotteista normaalijakaumaa eli

$$\begin{pmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{c}_1^T \mathbf{y} \\ \mathbf{c}_2^T \mathbf{y} \end{pmatrix} \sim N_2\left(\begin{pmatrix} \mathbf{c}_1^T \boldsymbol{\mu} \\ \mathbf{c}_2^T \boldsymbol{\mu} \end{pmatrix}, \sigma^2 \begin{pmatrix} \mathbf{c}_1^T \mathbf{c}_1 & 0 \\ 0 & \mathbf{c}_2^T \mathbf{c}_2 \end{pmatrix} \right).$$

Kuten kovarianssimatriisista nähdään, ortogonaaliset kontrastimuuttujat ovat riippumattomia. Ortogonaalisten kontrastimuuttujien avulla voidaan siis konstruoida eri parametreille tilastollisesti riippumattomia estimaatteja ja testejä.

4.2 Kontrastimuuttujien valinta

Ehdot

Kun tutkittavia vaikutusparametreja on P, jokaisella ryhmän $i\,(i=1,\ldots,I)$ yksilölle $k\,(k=1,\ldots,K)$ muodostetaan P-ulotteinen kontrastimuuttuja $\mathbf{u}_{ik}=\mathbf{C}\mathbf{y}_{ik}$, missä siis

$$\underset{P\times 1}{u_{ik}} = \begin{bmatrix} u_{i1k} \\ u_{i2k} \\ \vdots \\ u_{iPk} \end{bmatrix}, \quad \underset{P\times J}{C} = \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1J} \\ c_{21} & c_{22} & \cdots & c_{2J} \\ \vdots & \vdots & & \vdots \\ c_{P1} & c_{P2} & \cdots & c_{PJ} \end{bmatrix} = \begin{bmatrix} c_1^T \\ c_2^T \\ \vdots \\ c_P^T \end{bmatrix} \; ja \; \underset{J\times 1}{y_{ik}} = \begin{bmatrix} y_{i1k} \\ y_{i2k} \\ \vdots \\ y_{iJk} \end{bmatrix}.$$

Kerroinmatriisissa $\mathbf{C}$ sarakkeita vastaavat jaksot siten, että esimerkiksi ensimmäisessä sarakkeessa ovat ensimmäistä jaksoa vastaavat kertoimet parametreille $(1,\ldots,P)$ ja toisessa sarakkeessa ovat toista jaksoa vastaavat kertoimet. Matriisin riveillä puolestaan ovat kutakin kiinnostavaa parametria vastaavat kertoimet. Matriisin $\mathbf{C}$ rivielementit siis täyttävät ehdon $\sum_{j=1}^{J} c_{pj}=0$, kaikilla $p=1,\ldots,P$ eli jokaisen parametrin kertoimien $c_{pj}\,(j=1,\ldots,J)$ summa on nolla. Kuten aikaisemminkin vektori $\mathbf{y}_{ik}$ sisältää ryhmän i yksilöön k liittyvät jaksoittaiset mittaustulokset.

Esimerkkiaineistolle ryhmän 1 tapauksessa (ikä nuori, käsittelyjärjestys O:A:O) yksilöllinen kontrastivektori $\mathbf{u}_{1k}$, kerroinmatriisi $\mathbf{C}$ ja havaintovektori $\mathbf{y}_{1k}$ ovat

$$\underset{2\times 1}{\mathbf{u}_{1k}} = \begin{bmatrix} u_{11k} \\ u_{12k} \end{bmatrix}, \quad \underset{2\times 3}{\mathbf{C}} = \begin{bmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \end{bmatrix} = \begin{bmatrix} \mathbf{c}_1^{\mathrm{T}} \\ \mathbf{c}_2^{\mathrm{T}} \end{bmatrix} \text{ja } \underset{3\times 1}{\mathbf{y}_{ik}} = \begin{bmatrix} y_{11k} \\ y_{12k} \\ y_{13k} \end{bmatrix}.$$

Matriisin C optimaalinen valinta

Matriisi $\mathbf{C}$ valitaan siten, että ryhmien tai käsittelyjen eroja kuvaavat parametrit saadaan ryhmien kontrastimuuttujien odotusarvojen erotuksina ja kontrastimuuttujien erotuksen varianssi on mahdollisimman pieni.

Tällöin kontrastimuuttujien odotusarvojen erotus on $E(\mathbf{c}_p^T y_{1k}) - E(\mathbf{c}_p^T y_{2k}) = \Theta$, missä Θ on tutkittava parametri ja varianssi on kontrastimuuttujien erotuksen varianssi $Var(\mathbf{c}_p^T y_{1k} - \mathbf{c}_p^T y_{2k}) = 2Var(\mathbf{c}_p^T y_{1k}) = \sigma_e^2 \mathbf{c}_p^T \mathbf{c}_p$, koska samavarianssisuusoletuksen (*CS*, tasainen kovarianssirakenne, katso kappale 2.5) mukaan oletetaan, että $Var(y_{1k}) = Var(y_{2k}) = \ldots = Var(y_{lk}) = \sigma_e^2$.

Kerroinmatriisin C tekijöiden ratkaiseminen

Kerroinmatriisin **C** tekijöiden laskemisessa on kaksi periaatetta: valitaan sellainen matriisi, että yksilötason termit kumoutuvat kontrasteja laskettaessa ja kontrastien avulla laskettavien estimaattien varianssit tulevat mahdollisimman pieniksi. Estimaatin varianssien ollessa mahdollisimman pieniä, käytössä on optimaalinen kerroinmatriisi **C**. Pienimmän varianssin tilanteessa saadaan parhaimmat mahdolliset estimaatit ja kapeimmat parametrien luottamusvälit ja täten löydetään tehokkain menetelmä parametrien estimoimiseksi.

Esimerkkiaineiston avulla tarkastellaan yksilöllistä kontrastivektoria ja optimaalisen kerroinmatriisin **C** ratkaisemista. Lisäksi tutkitaan parametrien estimaattien odotusarvoja ja varianssia.

Odotusarvot

Luvussa 2.2 esitettiin nurkkasolurajoitus kiinnostavien parametrien tasoille siten, että ikäryhmä nuori on vertailuluokka ($A_1 = 0$), jakso 1 ($P_1 = 0$), ilmakäsittely ($F_1 = 0$) ja ilman viivästevaikutus ($C_1 = 0$) ovat vertailuluokkia. Tällöin estimoitaviksi jäävät parametrit A_2, P_2, P_3, F_2 ja C_2 eli ikäryhmän aikuinen vaikutus, jaksojen 2 ja 3 vaikutus, otsonikäsittelyn vaikutus ja otsonin viivästevaikutus. Nurkkasolurajoituksella saadut odotusarvot ovat taulukossa 4.1 (vrt. taulukko 2.2).

Taulukko 4.1 Odotusarvot ikäryhmittäin, käsittelyjärjestyksittäin ja jaksoittain

Ikäryhmä	Käsittelyjärjestys	Jakso		
		I	II	III
Nuori	*O:A:O*	$\mu + F_2$	$\mu + P_2 + C_2$	$\mu + P_3 + F_2$
	A:O:A	μ	$\mu + P_2 + F_2$	$\mu + P_3 + C_2$
Aikuinen	*O:A:O*	$\mu + A_2 + F_2$	$\mu + P_2 + A_2 + C_2$	$\mu + P_3 + F_2 + A_2$
	A:O:A	$\mu + A_2$	$\mu + P_2 + F_2 + A_2$	$\mu + P_3 + A_2 + C_2$

Kuten taulukosta 4.1 nähdään kontrastimuuttujien jakaumat riippuvat ainoastaan parametreista P_2, P_3, C_2 ja F_2 ja neljän ryhmän tilanne pelkistyy kahden ryhmän tilanteeksi. Jos on syytä olettaa, että jakson tai käsittelyn vaikutus on erilaista eri ikäryhmissä, parametreja täytyy testata ja estimoida ikäryhmittäin. Tässä oletetaan, että jakson ja käsittelyn vaikutus sekä viivästevaikutus on samanlaista riippumatta ikäryhmästä.

Seuraavaksi lasketaan kerroinmatriisi **C** vaakarivi kerrallaan etsimällä parametrien estimaatit tapauksessa, jossa molemmat parametrit F_2 ja C_2 ovat mukana mallissa.

Kontrastimuuttuja parametrille F_2

Nyt on siis etsittävä ne kontrastimuuttujat, joiden odotusarvot ryhmissä 1 ja 2 poikkeavat F_2:n verran. Niiden löytämiseksi muodostetaan kyseinen odotusarvojen erotus, joka on

$$\begin{aligned} & c_{11}F_2 + c_{12}P_2 + c_{12}C_2 + c_{13}P_3 + c_{13}F_2 - c_{12}P_2 - c_{12}F_2 - c_{13}P_3 - c_{13}C_2 \\ & = (c_{11} + c_{13} - c_{12})F_2 + (c_{12} - c_{13})C_2 \end{aligned}$$

Tästä saadaan ehdot parametrin F_2 kontrastimuuttujien kertoimille: Kertoimien summan on oltava 0, parametrin F_2 kerroin on 1 ja parametrin C_2 kerroin on 0 yllä olevassa yhtälössä. Matriisimuodossa esitettyinä ehdot ovat

$$\begin{pmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \\ 0 & 1 & -1 \end{pmatrix}\begin{pmatrix} c_{11} \\ c_{12} \\ c_{13} \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}.$$

Ehdoista saadaan kolmen yhtälön yhtälöryhmä, joka ratkaisemalla saadaan kerroinmatriisin **C** ylimmän rivin tekijät. Siis

$$\begin{cases} c_{11} + c_{12} + c_{13} = 0 \\ c_{11} + c_{13} - c_{12} = 1 \\ c_{12} - c_{13} = 0 \end{cases} \Rightarrow \begin{cases} c_{11} = 1 \\ c_{12} = c_{13} = -\frac{1}{2} \end{cases}$$

Varianssiksi parametrille F_2 saadaan $Var(\mathbf{c}^{\mathrm{T}}\mathbf{y}_{\mathrm{ik}}) = \frac{3}{2}\sigma_e^2$.

Kontrastimuuttuja parametrille C_2

Vastaavasti kuin etsittäessä käsittelyvaikutukseen F_2 liittyviä kertoimia, tarkastellaan ryhmien 1 ja 2 odotusarvojen erotusta

$$(c_{21} + c_{23} - c_{22})F_2 + (c_{22} - c_{23})C_2$$

Koska erotuksen tulee olla C_2, saadaan ehdot matriisimuodossa

$$\begin{pmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \\ 0 & 1 & -1 \end{pmatrix}\begin{pmatrix} c_{21} \\ c_{22} \\ c_{23} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

Ehdoista saadaan kolmen yhtälön yhtälöryhmä, jotka ratkaisemalla saadaan kerroinmatriisin **C** toisen rivin tekijät.

$$\begin{cases} c_{21} + c_{22} + c_{23} = 0 \\ c_{21} + c_{23} - c_{22} = 0 \\ c_{22} - c_{23} = 1 \end{cases} \Rightarrow \begin{cases} c_{21} = -1 \\ c_{22} = 0 \\ c_{23} = 1 \end{cases}$$

Yhdistämällä lasketut ensimmäisen ja toisen rivin tekijät saadaan kerroinmatriisi **C**. Siis jos mallissa on mukana viivästevaikutusta kuvaava parametri, niin ainoa mahdollinen kerroinmatriisi on

$$C = \begin{pmatrix} 1 & -\frac{1}{2} & -\frac{1}{2} \\ -1 & 0 & 1 \end{pmatrix}.$$

Optimaalinen kontrastimuuttuja

Kappaleen 4.2 lopussa esitetään periaate sille, miten optimaalinen kontrastimuuttuja määritellään ja lasketaan esimerkkiaineistolle. Lisäksi kappaleessa 5.1 tehdään t-testi ja kappaleessa 5.2 Mannin-Whitneyn -testi kontrastimuuttujalle.

Optimaalinen kontrastimuuttuja otsonin vaikutukselle F_2

Aikaisempien tulosten perusteella voidaan olettaa, ettei otsonilla ole viivästevaikutusta ja siten saadaan ratkaistavaksi kerroinmatriisia laskettaessa kolme tekijää kahden yhtälön avulla. Tällöin joudutaan laskemaan optimaalinen eli pienimmän varianssin antava kerroinmatriisi.

Halutaan testata nollahypoteesia $H_0: F_2 = 0$ ja vastahypoteesia $H_1: F_2 \neq 0$. Ensin lasketaan kerroinmatriisi $\mathbf{C}$, jolla kertomalla havaintovektorit saadaan mahdollisimman pienivarianssinen estimaatti.

Otsonin vaikutuksen (F_2) estimaatin varianssin suuruuteen vaikuttavat kerroinmatriisin $\mathbf{C}$ ensimmäisen (ja ainoan) rivin tekijöiden suuruudet eli estimaatin odotusarvoksi saadaan

$$\begin{aligned} & c_1F_2 + c_2P_2 + c_3P_3 + c_3F_2 - c_2P_2 - c_2F_2 - c_3P_3 \\ & = (c_1 + c_3 - c_2)F_2 \end{aligned}$$

Tästä saadaan parametrin F_2 kontrastitekijöiden yhtälö asettamalla ehdoiksi, että kontrastimatriisin rivin tekijöiden summa on 0 ja parametrin F_2 kontrastitekijöiden summa on 1. Matriisimuodossa esitettynä ehdot ovat

$$\begin{pmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

Ehdoista saadaan kahden yhtälön yhtälöryhmä, jolla on ääretön määrä ratkaisuja

$$\begin{cases} c_1 + c_2 + c_3 = 0 \\ c_1 + c_3 - c_2 = 1 \end{cases} \tag{8}$$

Minimoimalla kontrastimuuttujien varianssi $Var(\mathbf{c}^{\mathrm{T}} y_{ik}) = c_1^2\sigma_e^2 + c_2^2\sigma_e^2 + c_3^2\sigma_e^2$ saadaan optimaaliset kerroinvektorin tekijät.

Ensin huomataan tarkastelemalla yhtälöiden erotusta, että välttämättä

$$c_2 = -\tfrac{1}{2}.$$

Yhtälöiden summa antaa ehdon

$$c_1 = \tfrac{1}{2} - c_3.$$

Siis kerroinvektori on muotoa

$$\mathbf{c} = (\tfrac{1}{2} - c \quad -\tfrac{1}{2} \quad c),$$

jolloin

$$Var(\mathbf{c}^{\mathrm{T}} \mathbf{y}_{\mathbf{ik}}) = \left[(\tfrac{1}{2} - c)^2 + (\tfrac{1}{2})^2 + c^2\right]\sigma_e^2$$

minimoituu, kun $c = \frac{1}{4}$. Minimivarianssi on siis $\frac{3}{8}\sigma_e^2$ ja optimaalinen valinta on siten

$$\begin{cases} c_1 = \frac{1}{4} \\ c_2 = -\frac{1}{2}. \\ c_3 = \frac{1}{4} \end{cases}$$

Huomataan, että optimaalisen kontrastimuuttujan varianssi on vain $\frac{1}{4}$ aiemmin viivästevaikutuksen tapauksessa lasketusta parametrin F_2 kontrastimuuttujan kertoimien varianssista.

5 Estimointi ja testaus käyttäen kontrastimuuttujia

5.1 Kahden otoksen t-testi luottamusväleineen

Tässä kappaleessa lasketaan tavanomaisin parametrisin menetelmin kontrastimuuttujia käyttäen testisuureet, p-arvot ja luottamusvälit parametreille F_2 ja C_2. Kappaleessa 5.2 lasketaan vastaavalla epäparametrisella menetelmällä testisuureet, p-arvot ja luottamusvälit.

Parametrinen menetelmä

Parametrisena menetelmänä käytetään kahden otoksen t-testiä (esimerkiksi Jones ja Kenward). Parametrien estimaatit lasketaan eri ryhmien yksilöllisten kontrastimuuttujien keskiarvojen $\bar{\mathbf{u}}_{1\cdot}$ ja $\bar{\mathbf{u}}_{2\cdot}$ erotuksina seuraavasti

$$\hat{\Theta} = \begin{bmatrix} \hat{F}_2 \\ \hat{C}_2 \end{bmatrix} = \bar{\mathbf{u}}_{1\cdot} - \bar{\mathbf{u}}_{2\cdot} = \begin{bmatrix} \bar{\mathbf{u}}_{11\cdot} \\ \bar{\mathbf{u}}_{12\cdot} \end{bmatrix} - \begin{bmatrix} \bar{\mathbf{u}}_{21\cdot} \\ \bar{\mathbf{u}}_{22\cdot} \end{bmatrix}.$$

Kontrastimuuttujien keskiarvojen $\bar{\mathbf{u}}_{1\cdot}$ ja $\bar{\mathbf{u}}_{2\cdot}$ avulla voidaan tutkia, onko ryhmien välillä eroa. Testisuure, p-arvo, estimaatit ja luottamusvälit on laskettu SAS-ohjelmiston Version 8 *PROC TTEST*-proseduurilla.

Hypoteesit

$H_0: F_2 = 0, H_1: F_2 \neq 0$

Halutaan tutkia eroavatko ryhmien 1 ja 2 havainnot toisistaan. Havainnot kerrotaan kappaleessa 4.2 lasketulla kerroinmatriisilla, jonka avulla saaduilla kontrastimuuttujilla voidaan testata käsittely- ja viivästevaikutusta. Aiemmin on tehty nurkkasolurajoitus (kappale 2.2), jossa ilman käsittely- ja viivästevaikutus asetetaan vertailutasoiksi eli oletetaan, ettei niillä ole vaikutusta. Tämän perusteella voidaan nolla- ja vastahypoteeseiksi asettaa $H_0: C_2 = 0$, $H_1: C_2 \neq 0$ ja $H_0: F_2 = 0, H_1: F_2 \neq 0$ eli tällöin nollahypoteesina on, ettei otsonilla ole viivästevaikutusta ja vastahypoteesina on, että otsonilla on viivästevaikutusta sekä vastaavasti otsonikäsittelyllä ei ole vaikutusta ja otsonikäsittelyllä on vaikutusta.

Otsonikäsittelyn viivästevaikutuksen testaus, estimointi ja luottamusvälit

Kappaleessa 4.2 lasketun kerroinmatriisin **C** avulla saadaan viivästevaikutukselle estimaattori

$$\hat{C}_2 = -\overline{Y}_{11\cdot} + \overline{Y}_{13\cdot} + \overline{Y}_{21\cdot} - \overline{Y}_{23\cdot}.$$

Estimaatin estimoitu varianssi on

$$V\hat{a}r(\hat{C}_2) = s^2(\tfrac{1}{n_1} + \tfrac{1}{n_2}),$$

missä

$$s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

ja s_1^2 ja s_2^2 ovat kontrastimuuttujien otosvarianssit ryhmittäin.

Viivästevaikutuksen estimaattorin $\hat{C}_2$ $(1-2\alpha)\cdot 100\%$:n luottamusväli on $\hat{C}_2 \pm t(1-\alpha/2, (n_1+n_2-2))\sqrt{V\hat{a}r(\hat{C}_2)}$, missä $t(1-\alpha/2, n_1+n_2-2)$ on kriittinen arvo Studentin t-jakaumasta (n_1+n_2-2) vapausasteella.

Kahden otoksen t-testisuure viivästevaikutukselle on

$$t = \frac{\hat{C}_2}{\sqrt{V\hat{a}r(\hat{C}_2)}} = \frac{(\overline{x}_1 - \overline{x}_2)}{\sqrt{s^2(1/n_1 + 1/n_2)}} \sim t(\alpha, (n_1+n_2-2)),$$

missä $s^2 = \dfrac{[(n_1-1)s_1^2 + (n_2-1)s_2^2]}{n_1+n_2-2}$ ja s_1^2, s_2^2 ovat ryhmien 1 ja 2 varianssit.

Testisuure, p-arvo, estimaatit ja luottamusvälit ovat taulukossa 5.1.

Otsonikäsittelyn vaikutuksen testaus, estimointi ja luottamusvälit

Kappaleessa 4.2 lasketun kerroinmatriisin **C** avulla saadaan otsonikäsittelylle estimaattori

$$\hat{F}_2 = \overline{Y}_{11\cdot} - \tfrac{1}{2}\overline{Y}_{12\cdot} - \tfrac{1}{2}\overline{Y}_{13\cdot} - \overline{Y}_{21\cdot} + \tfrac{1}{2}\overline{Y}_{22\cdot} + \tfrac{1}{2}\overline{Y}_{23\cdot}.$$

Estimaatin estimoitu varianssi saadaan vastaavasti kuin viivästevaikutuksen estimaatille

$$V\hat{a}r(\hat{F}_2) = s^2(\tfrac{1}{n_1} + \tfrac{1}{n_2}).$$

Käsittelyvaikutuksen $\hat{F}_2$ $(1-2\alpha)\cdot 100\%$:n luottamusväli on

$$\hat{F}_2 \pm t(1-\alpha/2, (n_1+n_2-2))\sqrt{V\hat{a}r(\hat{F}_2)},$$

missä $t(1-\alpha/2, n_1+n_2-2)$ on kriittinen arvo Studentin t-jakaumasta (n_1+n_2-2) vapausasteella.

Kahden otoksen t-testisuure käsittelyvaikutukselle on

$$t = \frac{\hat{F}_2}{\sqrt{V\hat{a}r(\hat{F}_2)}} = \frac{(\overline{x}_1 - \overline{x}_2)}{\sqrt{s^2(1/n_1 + 1/n_2)}} \sim t(\alpha, (n_1+n_2-2)),$$

missä $s^2 = \dfrac{[(n_1-1)s_1^2 + (n_2-1)s_2^2]}{n_1+n_2-2}$ ja s_1^2, s_2^2 ovat ryhmien 1 ja 2 varianssit.

Testisuure, p-arvo, estimaatit ja luottamusvälit ovat taulukossa 5.1 ja ohjelmat ovat liitteessä.

Taulukko 5.1 Viivästevaikutuksen ja otsonin vaikutuksen testisuure, p-arvo, estimaatit ja 95 %:n luottamusvälit

	hypoteesit	testisuure	p-arvo	estimaatit	Luottamusväli
viivästevaikutus	$H_0: C_2 = 0$	$t = -0.06 \sim t(17)$, kun H_0 on tosi	0.9562	−0.202	(−8.135, 7.731)
otsonin vaikutus	$H_0: F_2 = 0$	$t = 0.82 \sim t(17)$, kun H_0 on tosi	0.4285	2.6806	(−4.505, 9.8657)

Huomataan, että 5 %:n merkitsevyystasolla otsonilla ei ole viivästevaikutusta (p=0.9562) vertailutason ollessa ilman viivästevaikutus ja otsonikäsittelyllä ei ole vaikutusta nukkumisaikaan (p=0.4285). Estimaatiksi käsittelyvaikutukselle saatiin 2.6806 minuuttia eli otsonikäsittelyn ja ilmakäsittelyn saaneen hiiren nukkumisajan

estimoitu ero on 2.6806 minuuttia siten, että otsonikäsittely pidentää nukkumisaikaa. Otsonin viivästevaikutuksen estimaatti on -0.202 minuuttia siten, että edellisen jakson otsonikäsittelyn estimoidaan lyhentävän nukkumisaikaa 0.202 minuuttia. 95 %:n luottamusvälit ovat käsittelyvaikutukselle $(-4.505, 9.8657)$ ja viivästevaikutukselle $(-8.135, 7.731)$. Tuloksista huomataan, että estimaatit ovat identtiset taulukossa 3.2. esitettyjen mallin (3) estimaattien kanssa. Estimaattien keskivirheet, t-testisuureet ja vapausasteet kuitenkin eroavat mallin (3) vastaavista, koska nyt estimaatit lasketaan käsittelyjärjestysten 1 (O:A:O) ja 2 (A:O:A) yksilöllisten kontrastimuuttujien keskiarvojen erotuksina (Chow ja Liu, sivut 263-272).

Otsonikäsittelyn vaikutuksen estimointi ja testaus optimaalisella kontrastimuuttujalla ja t-testillä

Otsonikäsittelyn vaikutus estimoitiin ja testattiin samalla tavalla kuin edellisessä kappaleessa. Tulokset ovat taulukossa 5.2 ja ohjelma on liitteessä.

Taulukko 5.2 Käsittelyvaikutuksen hypoteesit, testisuure, p-arvo, estimaatit ja 95 %:n luottamusvälit

	Hypoteesit	testisuure	p-arvo	estimaatit	luottamusväli
Otsonin vaikutus	$H_0: F_2 = 0$	$t = 2.10 \sim t(17)$, kun H_0 on tosi	0.0531	2.5289	$(0.0269, 5.0309)$

Kun mallista poistettiin viivästevaikutuksen parametri, saatiin p-arvoksi testattaessa nollahypoteesia 0.0531. Estimaatiksi otsonikäsittelyn vaikutukselle nukkumisaikaan saatiin 2.5289 eli otsonikäsittelyn estimoidaan pidentävän nukkumisaikaa 2.5 minuuttia verrattuna ilmakäsittelyn vaikutukseen. Otsonikäsitellyn hiiren nukkumisaika ei eroa tilastollisesti merkitsevästi ilmakäsitellyn hiiren nukkumisajasta.

5.2 Kahden otoksen Mannin-Whitneyn testi luottamusväleineen

Jos oletus aineiston normaalijakautuneisuudesta ei ole voimassa, vaihtoehtona kappaleessa 5.1 esitellylle parametriselle t-testille ovat epäparametriset menetelmät. Kappaleessa 5.2 epäparametrisista testeistä esitellään Mannin-Whitneyn testi eli Mannin-Whitneyn-Wilcoxonin testi (WMW-testi), jonka asymptoottinen tehokkuus

t-testiin verrattuna normaalijakaumavasteen tapauksessa on $3/\pi \approx 95.5\%$ otoskoon kasvaessa.

Hypoteesit

WMW-testillä testataan hypoteesia, jonka mukaan havainnot ovat riippumattomia ja noudattavat samaa jakaumaa (ryhmien välillä ei ole eroa).

Kahta ryhmää (otoskoot n_1 ja n_2) vertailtaessa testisuure konstruoidaan siten, että koko aineisto järjestetään suuruusjärjestykseen, jolloin testisuure R on toiseen ryhmään liittyvien järjestyslukujen summa.

Jos aineistossa ei ole sidoksia eli kaikki havainnot ovat erisuuria, nollahypoteesin vallitessa testisuureen R jakauma ei riipu tuntemattomasta yhteisestä jakaumasta (*distribution-free*) ja

$$E(R) = \frac{n_2(n_1 + n_2 + 1)}{2} \text{ ja } Var(R) = \frac{n_1 n_2 (n_1 + n_2 + 1)}{12}.$$

Suurille otoksille nollahypoteesin vallitessa pätee likimain

$$z = \frac{R - \frac{n_2(n_1 + n_2 + 1)}{2}}{\sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}} \sim N(0,1). \tag{7}$$

Suuret testisuureen z^2 arvot johtavat nollahypoteesin hylkäämiseen. Kriittisiä arvoja normitetun normaalijakauman tiheysfunktiolle on taulukoissa (esimerkiksi Ranta, Rita ja Kouki) ja tilastollisissa ohjelmistoissa sisään asennettuna, esimerkiksi *SAS*-ohjelmiston *probnorm*.

WMW-testi havaintoaineistolle

Esimerkkiaineiston tapauksessa muodostetaan ensin edellä esitellyt kontrastimuuttujat ja sovelletaan WMW-testiä. Tässä ryhmässä 1 ovat ikäryhmien nuori ja aikuinen havainnot, joiden käsittelyjärjestys on *O:A:O* . Ryhmässä 2 ovat ikäryhmien nuori ja aikuinen havainnot, joiden käsittelyjärjestys on *A:O:A*. Merkitään ryhmiä i (i = 1,2).

Ensin täytyy tutkia, onko otsonilla viivästevaikutusta. Tällöin aiemmin asetettu nurkkasolurajoitus huomioiden hypoteesit ovat

$$H_0: C_2 = 0 \text{ ja } H_1: C_2 \neq 0$$

eli nollahypoteesina on, ettei otsonilla ole viivästevaikutusta ja vastahypoteesina on, että otsonilla on viivästevaikutusta vertailutason ollessa ilman viivästevaikutus. Testisuure, p-arvo, estimaatti ja 95 %:n luottamusväli on esitetty taulukossa 5.3 ja ohjelmat ovat liitteessä.

Seuraavaksi voidaan tutkia, onko otsonikäsittelyllä vaikutusta. Tällöin hypoteesit ovat

$$H_0: F_2 = 0 \text{ ja } H_1: F_2 \neq 0.$$

Testisuure, p-arvo, estimaatti ja 95 %:n luottamusväli on esitetty taulukossa 5.3 ja ohjelmat ovat liitteessä.

Hodges-Lehmannin estimaattori ja luottamusvälit

Jakaumasta riippumaton luottamusväli voidaan WMW-testin yhteydessä laskea *HL (Hodgesin-Lehmannin)*-estimaattorin $\hat{\Delta}$ avulla (Randles ja Wolfe, 1979 ja Chow ja Liu, 1992).

Hodges-Lehmannin estimaattori

Kunkin parametrin piste-estimaatti on ryhmien $(i = 1,2)$ vastaavien kontrastimuuttujien erotusten $D_{k_1 k_2} = \mathbf{c}^{\mathrm{T}}\mathbf{y}_{1\mathbf{k}_1} - \mathbf{c}^{\mathrm{T}}\mathbf{y}_{2\mathbf{k}_2}$ mediaani. Tälläisiä erotuksia on $n_1 \cdot n_2$ kappaletta. Olkoot järjestetyt erotukset $D(1) \leq D(2) \leq \ldots \leq D(n_1 n_2)$.

Siis HL-estimaattori on

$$\hat{\Delta} = \begin{cases} \frac{1}{2}\left[D\left(\frac{n_1 n_2}{2}\right) + D\left(\frac{n_1 n_2}{2} + 1\right)\right], \text{ jos } n_1 n_2 \text{ on parillinen.} \\ D\left(\frac{n_1 n_2 - 1}{2} + 1\right), \text{ jos } n_1 n_2 \text{ on pariton.} \end{cases}$$

Luottamusvälit

Järjestämällä parittaiset erotukset $D_{k_1 k_2}$ suuruusjärjestykseen

$$D(1) \leq D(2) \leq \cdots \leq D(2)^* \leq D(1)^*,$$

missä $D(k)$ on k:nneksi pienin ja $D(k)^*$ on k:nneksi suurin erotus, ja tarkastellaan välejä

$$[D(k), D(k)^*], \text{missä } k = 1, \ldots, n_1 n_2.$$

Todennäköisyys, että väli $[D(k), D(k)^*]$ sisältää oikean parametrin arvon on (Spiegel,1996, Hahn ja Meeker, 1991)

$$P_k = 1 - 2P(R \leq k - 1),$$

kun R on järjestyslukujen summista pienempi ja P on nollahypoteesin määräämä todennäköisyysjakauma.

Pienillä otoksilla tarkat kertymäfunktion arvot on taulukoitu (esimerkiksi Ranta, Rita ja Kouki) ja isoilla otoksilla voidaan käyttää normaalijakauma-approksimaatiota

$$P_k = 1 - 2\Phi\left(\frac{k - n_1 n_2/2}{\sqrt{n_1 n_2 (n_1 + n_2 + 1)/12}}\right).$$

Tästä kaavasta voidaan ratkaista k, kun todennäköisyyden P_k oletetaan olevan lähellä haluttua merkitsevyystasoa kuten 0.95. Edellä olevan perusteella $(1-\alpha)100\%$:n luottamusväliksi saadaan

$$\hat{\Delta} \pm \Phi^{-1}(\alpha/2)\sqrt{n_1 n_2((n_1 + n_2) + 1)/12} = \hat{\Delta} \pm z_{\alpha/2}\sqrt{n_1 n_2((n_1 + n_2) + 1)/12}.$$

Kun k valitaan siten, että

$$\alpha/2 = \Phi\left(\frac{k - n_1 n_2/2}{\sqrt{n_1 n_2 (n_1 + n_2 + 1)/12}}\right).$$

Taulukko 5.3 Viiväste- ja käsittelyvaikutuksen hypoteesit, testisuure, p-arvot, HL-estimaattori ja 95 %:n luottamusvälit.

	Hypoteesit	testisuure	p-arvo	estimaatti	luottamusväli
Viivästevaikutus	$H_0: C_2 = 0$	$z = -0.1633$ $(R = 88)$	0.8702	−4.85	$(-12.75, 3.05)$
Otsonin vaikutus	$H_0: F_2 = 0$	$z = -0.3266$ $(R = 86)$	0.7440	1.975	$(-7.35, 11.30)$

Testisuureen arvot on laskettu käsin SAS-ohjelmiston version 8 tulostusta hyväksi käyttäen. SAS-ohjelmiston proseduuri *Proc npar1way* laskee WMW-testisuureen suurille otoksille ja käyttää p-arvoa laskiessaan normaaliaproksimaatiota. Tulostuksesta voidaan käyttää pienille otoksille testisuureita laskettaessa *Sum of Scores* -arvoja eli järjestyslukujen summia, jotka SAS-ohjelmisto laskee oikein kaikille otoksille. Tarkat p-arvot on laskettu *SAS*-ohjelmiston *probnorm*-funktiolla.

Tuloksena saadaan, että otsonilla ei näyttäisi olevan viivästevaikutusta merkitsevyystason ollessa 0.05 kaksisuuntaisessa testauksessa. Tämän perusteella voidaan testata mallia, josta on poistettu viivästevaikutuksen parametri. Testattaessa otsonin käsittelyvaikutusta merkitsevyystasolla 0.05 mallissa, jossa on viivästevaikutuksen parametri, otsonilla ei ole käsittelyvaikutusta.

Viivästevaikutuksen *HL*-estimaattoriksi saadaan −4.450 eli edellisellä jaksolla otsonikäsitellyn hiiren nukkumisajan estimoitu ero verrattuna edellisellä jaksolla ilmakäsittelyn saaneen hiiren nukkumisaikaan on −4.450 minuuttia.

Kun mallissa on mukana viivästevaikutuksen parametri, *HL*-estimaatti otsonikäsittelyn vaikutukselle nukkumisaikaan on 1.975 minuuttia ilmakäsitellyn saaneen hiiren nukkumisajan ollessa vertailutaso.

Viivästevaikutuksen estimaatin 95 %:n luottamusväliksi saatiin $(-7.35, 11.30)$. Luvun 5.2 alussa esitetyllä periaatteella saatiin muuttujan k arvoksi 69 ja tämän perusteella yläraja on järjestettyjen erotusten 69. havainto. Symmetrian nojalla alarajaksi saadaan estimaatin ja ylärajan erotus vähennettynä estimaatista.

Käsittelyvaikutuksen estimaatin 95 %:n luottamusväliksi saatiin $(-12.75, 3.05)$ vastaavasti kuin saatiin viivästevaikutuksen luottamusväli.

Otsonikäsittelyn estimointi ja testaus optimaalisen kontrastimuuttujan ja WMW-testin avulla

Otsonin käsittelyvaikutus estimoitiin ja testattiin samalla tavalla kuin mallille, jossa oli mukana viivästevaikutuksen parametri. Tulokset ovat taulukossa 5.4.

Taulukko 5.4 Käsittelyvaikutuksen hypoteesit, testisuure, p-arvo, HL-estimaatti ja 95 %:n luottamusvälit

	Hypoteesit	testisuure	p-arvo	estimaatti	luottamusväli
Otsonin vaikutus	$H_0: F_2 = 0$	$z = -2.2123$ $(R = 64)$	0.0269	−0.0875	(−3.400,3.225)

Kun mallista poistettiin viivästevaikutuksen parametri, saatiin testattaessa nollahypoteesia p-arvoksi 0.0269. Estimaatiksi otsonikäsittelyn vaikutukselle nukkumisaikaan saatiin −0.0875 eli otsonikäsittelyn estimoidaan lyhentävän nukkumisaikaa 0.09 minuuttia verrattuna ilmakäsittelyn vaikutukseen. Otsonikäsitellyn hiiren nukkumisaika eroaa tilastollisesti merkitsevästi ilmakäsitellyn hiiren nukkumisajasta.

6 Lopuksi

Tutkielman alkuperäisenä tarkoituksena oli esittää kontrastimuuttujiin ja epäparametrisiin menetelmiin liittyvää teoriaa sekä tämän teorian soveltamista käytäntöön. Aihetta työstettäessä tutkielmaan lisättiin myös muuta teoriaa eli tutkielmaan kirjoitettiin kovarianssirakenteista ja parametrisista menetelmistä erityisesti SAS-ohjelmiston käyttämästä näkökulmasta sekä teoriaa malleista, joissa on kiinteitä tekijöitä tai kiinteitä ja satunnaisia tekijöitä.

Tutkielmassa käytiin läpi lineaarisen sekamallin ominaisuuksia purkamalla malli matriisimuotoon asti, kuten tällä tasolla kirjoitettaessa on käytännöllistä. Teoriaosiossa käsiteltiin myös kovarianssirakenteita huolellisesti perusasioita myöten. Kovarianssirakenteista käytettiin mallien ja teorian yksinkertaisuuden vuoksi tasaista kovarianssirakennetta (*CS*), vaikka olemassa on monia muitakin kovarianssirakenteeksi

sopivia vaihtoehtoja, joita käytetään vähemmän. Tulevaisuudessa voisi vastaavan teorian läpikäynti johonkin muuhun kovarianssirakenteeseen perustuen olla hyödyllistä.

Estimoinnissa ja testauksessa käytettiin ensin SAS-ohjelmiston sisältämiä *GLM*- ja *Mixed*-proseduureja, joista erityisesti *Mixed*-proseduuria tutkielman tekijä työssään tarvitsee. Näiden jälkeen siirryttiin kontrastimuuttujien määrittelyyn ja kontrastimuuttujien avulla estimointiin sekä parametrisella että epäparametrisella menetelmällä. Tulevaisuudessa kontrastimuuttujia ja niiden avulla estimointia kannattaisi miettiä myös monimutkaisempien ja epätasapainoisten asetelmien tapauksessa, koska tässä tutkielmassa rajoituttiin tasapainoiseen 2×3-koeasetelmaan. Jatkossa voisi miettiä myös perustasomittausten hyödyntämistä kontrastimuuttujien avulla estimoitaessa ja sitä, voiko kontrastimuuttujia käytettäessä ottaa jotenkin huomioon REML-menetelmän käyttämää koeyksilöiden mittaustulosten välistä tietoa.

Lähdeluettelo

Brown, E. B., Iyer, H. K. and Wang, C. M. (1997) *Tolerance intervals for assessing individual bioequivalence, Statistics in Medicine*, **16**, 803-820.

Brown, H. K. and Kempton, R. A. (1994) *The application of REML in clinical trials, Statistics in Medicine*, **13**, 1601-1617.

Cheng, C. S. and Wu, C. F. (1980) *Balanced repeated measurements designs. Annuals of Statistics*, **8**, 1272-83. Corrigendum (1983), **11**, 349.

Chinchilli, V. M. and Esinhart, J. D. (1996) *Design and analysis of intra-subject variability in cross-over experiments, Statistics in Medicine*, **15**, 1619-1634.

Chow, S. C. and Liu, J. P. (1992) *Design and analysis of bioavailability and bioequivalence studies.* Marc Dekker, Inc., New York.

Chow, S. C. and Liu, J. P. (1992) *On assessment of bioequivalence under a higher-order crossover design, Journal of biopharmaceutical statistics*, **2(2)**, 239-256.

Gough, K. (1989) *Letter to the editor. Statistics in Medicine*, **8**, 891-892.

Hafner, K. B., Canada, A. T. and Koch, G. G. (1988) *Some analysis strategies for three-period changeover designs with two treatments. Statistics in Medicine*, 7, 471-481.

Hahn, G. J. and Meeker, W. Q. (1991) *Statistical intervals.* John Wiley & Sons, Inc., New York.

Hinkka, S. (1993) Kaksiperiodiseen ristikkäiskaaviokokeeseen liittyvät ongelmat ja niiden ratkaiseminen alkumittauksia käyttäen. Analyysit ja niiden toteutus BMDP-ohjelmistolla. Pro gradu -tutkielma, Tilastotieteen laitos, Turun yliopisto.

Jones, B. and Kenward. M.G. (1989) *Design and analysis of cross-over trials.* Chapman-Hall, London.

Kotz, S., Johnson, N. L. and Read, C. B. (1982) and (1988) *Encyclopedia of Statistical Sciences,* New York, John Wiley & Sons.

Laska, E. and Meisner, M. (1985) *A variatonial approach to optimal two-treatment crossover designs: applications to carryover-effect models. Journal of the American Statistical Association*, **80**, 704-710.

Laska, E., Meisner, M. and Kushner, H. B. (1983) *Optimal crossover designs in the presence of carryover effects. Biometrics*, **39**, 1089-1091.

Leinonen, L. (1999) Tilastollisten peruskäsitteiden oppiminen peruskoulun yläasteella: tietokoneavusteisen opetusmenetelmän vaikutuksen tutkiminen. Lisensiaattityö. Matemaattisten tieteiden laitos, Oulun yliopisto.

Mardia, K. V. (1985) *Mardia's Test of Multinormality.* Teoksessa: *Encyclopedia of Statistical Sciences.* (Toim.) Kotz, S., Johnson, N. C. and Read, C. B. New York, John Wiley & Sons. Volume 5, p. 217-221.

Matthews, J., N., S. (1987) *Optimal crossover designs for the comparison of two treatments in the presence of carryover effects and autocorrelated errors. Biometrika*, **74**, 311-20.

Patel, H. L. and Tsai, K. T. (1996) *Robust procedures for analysing a two-period cross-over design with baseline measurements, Statistics in Medicine*, **15**, 117-126.

Ranta, E., Rita, H. and Kouki, J. (1989) *Biometria.* Yliopistopaino, Helsinki.

Royston, J. P. (1982) Shapiro-Wilk W Statistics, Teoksessa: *Encyclopedia of Statistical Sciences.* (Toim.) Kotz, S., Johnson, N. C. and Read, C. B. New York, John Wiley & Sons. Volume 8, p. 430-431.

SAS Institute Inc. (1996) SAS/STAT Software. Mixed Procedure. *Modelling covariance structure: Using the REPEATED statement in PROC MIXED*, 92-102.

SAS Institute Inc. (1996) SAS/STAT Software. Mixed Procedure. *Other covariance structure*, 273-274.
SAS Institute Inc. (1996) SAS/STAT Software. User's Guide. Volume 2, *GLM-VARCOMP*, 891-996.
Smith, A. F. M. (1988) Schwarz Criterion. Teoksessa: *Encyclopedia of Statistical Sciences.* (Toim.) Kotz, S., Johnson, N. C. and Read, C. B. New York, John Wiley & Sons. Volume 8, 289-290.
Tuominen, J., Hinkka, S. (1997) Analysis with PROC MIXED in SAS. Käsikirjasto, Orion Pharma, Turku.

Liitteet

Aineistot

Koe.sd2

Koe.sd2 -aineistoa käytettiin muokattaessa aineistosta *Anova3.sd2* -aineisto. Aineistossa selittäjien tasoja on merkitty seuraavasti

nuoret hiiret =	IKA (1)
aikuiset hiiret =	IKA (2)
jaksot1, 2 ja 3 =	JAKSO (1, 2, 3)
ilmakäsittely =	TRT (1)
otsonikäsittely =	TRT (2)
nukkumisaika minuutteina =	SL_TIME
ilman viivästevaikutus =	C2 (0)
otsonin viivästevaikutus =	C2 (1).

OBS	SEQ	YKSILO	IKA	JAKSO	TRT	SL_TIME	C2
1	121	6	1	1	1	22.3	0
2	121	7	1	1	1	24.0	0
3	121	8	1	1	1	17.8	0
4	121	9	1	1	1	15.5	0
5	121	6	1	2	2	21.8	0
6	121	7	1	2	2	19.5	0
7	121	8	1	2	2	17.6	0
8	121	9	1	2	2	20.0	0
9	121	6	1	3	1	15.1	1
10	121	7	1	3	1	15.1	1
11	121	8	1	3	1	21.9	1
12	121	9	1	3	1	10.1	1
13	121	15	2	1	1	33.5	0
14	121	16	2	1	1	36.1	0
15	121	17	2	1	1	32.1	0
16	121	18	2	1	1	32.9	0
17	121	19	2	1	1	27.0	0
18	121	15	2	2	2	22.6	0
19	121	16	2	2	2	38.0	0
20	121	17	2	2	2	42.4	0
21	121	18	2	2	2	35.1	0
22	121	19	2	2	2	23.2	0
23	121	15	2	3	1	28.6	1
24	121	16	2	3	1	32.9	1
25	121	17	2	3	1	30.5	1
26	121	18	2	3	1	27.9	1
27	121	19	2	3	1	25.1	1
28	212	1	1	1	2	38.1	0
29	212	2	1	1	2	19.7	0
30	212	3	1	1	2	39.0	0
31	212	4	1	1	2	31.7	0
32	212	5	1	1	2	16.8	0
33	212	1	1	2	1	23.8	1
34	212	2	1	2	1	16.6	1
35	212	3	1	2	1	25.3	1
36	212	4	1	2	1	19.1	1
37	212	5	1	2	1	21.8	1

38	212	1	1	3	2	16.4	0
39	212	2	1	3	2	15.3	0
40	212	3	1	3	2	30.2	0
41	212	4	1	3	2	17.6	0
42	212	5	1	3	2	20.1	0
43	212	10	2	1	2	30.0	0
44	212	11	2	1	2	29.4	0
45	212	12	2	1	2	45.8	0
46	212	13	2	1	2	25.7	0
47	212	14	2	1	2	22.6	0
48	212	10	2	2	1	31.2	1
49	212	11	2	2	1	34.4	1
50	212	12	2	2	1	27.5	1
51	212	13	2	2	1	21.7	1
52	212	14	2	2	1	24.7	1
53	212	10	2	3	2	37.5	0
54	212	11	2	3	2	35.3	0
55	212	12	2	3	2	30.9	0
56	212	13	2	3	2	23.7	0
57	212	14	2	3	2	32.0	0

Koe2.sd2

Koe2.sd2 -aineistoa käytettiin *GLM*- ja *Mixed*-proseduureilla ajetuissa ohjelmissa. Aineistossa on merkitty selittäjien tasoja seuraavasti

nuoret hiiret =	IKA (1)
aikuiset hiiret=	IKA (0)
jaksot 1, 2 ja 3=	JAKSO (4, 2, 3)
otsonikäsittely=	TRT (0)
ilmakäsittely=	TRT (1)
nukkumisaika minuutteina=	SL_TIME
ilman viivästevaikutus=	C2 (0)
otsonin viivästevaikutus=	C2 (1).

OBS	SEQ	YKSILO	IKA	JAKSO	TRT	SL_TIME	C2
1	212	1	1	4	0	38.1	0
2	212	1	1	2	1	23.8	1
3	212	1	1	3	0	16.4	0
4	212	2	1	4	0	19.7	0
5	212	2	1	2	1	16.6	1
6	212	2	1	3	0	15.3	0
7	212	3	1	4	0	39.0	0
8	212	3	1	2	1	25.3	1
9	212	3	1	3	0	30.2	0
10	212	4	1	4	0	31.7	0
11	212	4	1	2	1	19.1	1
12	212	4	1	3	0	17.6	0
13	212	5	1	4	0	16.8	0
14	212	5	1	2	1	21.8	1
15	212	5	1	3	0	20.1	0
16	121	6	1	4	1	22.3	0
17	121	6	1	2	0	21.8	0
18	121	6	1	3	1	15.1	1
19	121	7	1	4	1	24.0	0
20	121	7	1	2	0	19.5	0

21	121	7	1	3	1	15.1	1
22	121	8	1	4	1	17.8	0
23	121	8	1	2	0	17.6	0
24	121	8	1	3	1	21.9	1
25	121	9	1	4	1	15.5	0
26	121	9	1	2	0	20.0	0
27	121	9	1	3	1	10.1	1
28	212	10	0	4	0	30.0	0
29	212	10	0	2	1	31.2	1
30	212	10	0	3	0	37.5	0
31	212	11	0	4	0	29.4	0
32	212	11	0	2	1	34.4	1
33	212	11	0	3	0	35.3	0
34	212	12	0	4	0	45.8	0
35	212	12	0	2	1	27.5	1
36	212	12	0	3	0	30.9	0
37	212	13	0	4	0	25.7	0
38	212	13	0	2	1	21.7	1
39	212	13	0	3	0	23.7	0
40	212	14	0	4	0	22.6	0
41	212	14	0	2	1	24.7	1
42	212	14	0	3	0	32.0	0
43	121	15	0	4	1	33.5	0
44	121	15	0	2	0	22.6	0
45	121	15	0	3	1	28.6	1
46	121	16	0	4	1	36.1	0
47	121	16	0	2	0	38.0	0
48	121	16	0	3	1	32.9	1
49	121	17	0	4	1	32.1	0
50	121	17	0	2	0	42.4	0
51	121	17	0	3	1	30.5	1
52	121	18	0	4	1	32.9	0
53	121	18	0	2	0	35.1	0
54	121	18	0	3	1	27.9	1
55	121	19	0	4	1	27.0	0
56	121	19	0	2	0	23.2	0
57	121	19	0	3	1	25.1	1

Anova3.sd2

Anova3.sd2 -aineistoa käytettiin *ttest*- ja *npar1way* -proseduureilla analysoitaessa. Aineistossa selittäjien tasoja on merkitty seuraavasti

käsittelyjärjestys OAO (otsoni, ilma, otsoni) = SEQ (212)
käsittelyjärjestys AOA (ilma, otsoni, ilma) = SEQ (121)
JAKSO1-, JAKSO2- ja JAKSO3-sarakkeissa ovat hiiren nukkumisajat minuutteina jaksoittain.

OBS	RYHMA	SEQ	YKSILO	JAKSO1	JAKSO2	JAKSO3
1	1	212	1	38.1	23.8	16.4
2	1	212	2	19.7	16.6	15.3
3	1	212	3	39.0	25.3	30.2
4	1	212	4	31.7	19.1	17.6
5	1	212	5	16.8	21.8	20.1
6	1	212	10	30.0	31.2	37.5
7	1	212	11	29.4	34.4	35.3
8	1	212	12	45.8	27.5	30.9

9	1	212	13	25.7	21.7	23.7
10	1	212	14	22.6	24.7	32.0
11	2	121	6	22.3	21.8	15.1
12	2	121	7	24.0	19.5	15.1
13	2	121	8	17.8	17.6	21.9
14	2	121	9	15.5	20.0	10.1
15	2	121	15	33.5	22.6	28.6
16	2	121	16	36.1	38.0	32.9
17	2	121	17	32.1	42.4	30.5
18	2	121	18	32.9	35.1	27.9
19	2	121	19	27.0	23.2	25.1

```
*DESCRIPTION:   Mallin sovittamista erilaisilla kovarianssirakenteilla
*NAME:          kov_rak2.sas
*               Analyysit
*AUTHOR:        Auli Partanen
*DATE:          1JUNE2000

libname data 'h:\gradu\last_version\sas\';

options nodate nocenter nonumber linesize=80;

data koe2; set data.koe2;
run;

Title 'mixed: kovarianssirakenne=UN';
proc mixed data=koe2;
        class ika trt yksilo jakso C2;
        model sl_time=ika trt jakso C2 / ddfm=satterth;
        repeated jakso / subject=yksilo(ika) type=UN  r rcorr; run;

Title 'mixed: kovarianssirakenne=CS';
proc mixed data=koe2;
        class ika trt yksilo jakso C2;
        model sl_time=ika trt jakso C2 / ddfm=satterth;
        repeated jakso / subject=yksilo(ika) type=CS  r rcorr; run;

Title 'mixed: kovarianssirakenne=CSH';
proc mixed data=koe2;
        class ika trt yksilo jakso C2;
        model sl_time=ika trt jakso C2 / ddfm=satterth;
        repeated jakso / subject=yksilo(ika) type=CSH  r rcorr; run;

Title 'mixed: kovarianssirakenne=AR(1)';
proc mixed data=koe2;
        class ika trt yksilo jakso C2;
        model sl_time=ika trt jakso C2 / ddfm=satterth;
        repeated jakso / subject=yksilo(ika) type=AR(1)  r rcorr; run;

Title 'mixed: kovarianssirakenne=ARH(1)';
proc mixed data=koe2;
        class ika trt yksilo jakso C2;
        model sl_time=ika trt jakso C2 / ddfm=satterth;
        repeated jakso / subject=yksilo(ika) type=ARH(1)  r rcorr; run;

Title 'mixed: kovarianssirakenne=FA(1)';
proc mixed data=koe2;
        class ika trt yksilo jakso C2;
        model sl_time=ika trt jakso C2 / ddfm=satterth;
        repeated jakso / subject=yksilo(ika) type=FA(1)  r rcorr; run;
```

```
Title 'mixed: kovarianssirakenne=FA1(1)';
proc mixed data=koe2;
        class ika trt yksilo jakso C2;
        model sl_time=ika trt jakso C2 / ddfm=satterth;
        repeated jakso / subject=yksilo(ika) type=FA1(1)  r rcorr; run; quit;



*DESCRIPTION: Analyysit ja estimaatit glm:llä, mixedillä (fixed ja
random)
*NAME:          mixed10.sas
*AUTHOR:        Auli Partanen
*DATE:          1JUNE2000

libname data 'h:\gradu\last_version\sas\';

options nodate nocenter nonumber linesize=80;

data koe2; set data.koe2;
run;

Title 'glm, fixed effects';
proc glm data=koe2;
        class ika trt yksilo jakso C2;
        model sl_time=ika yksilo(ika) trt jakso C2;
        estimate 'ilma vs. otsoni'    trt    1 -1;
        estimate 'nuori vs. aikuinen' ika    1 -1;
        estimate 'jakso1 vs. jakso2'  jakso  1  0 -1;
        estimate 'jakso1 vs. jakso3'  jakso  0  1 -1;
        estimate 'jakso2 vs. jakso3'  jakso -1  1  0;
        estimate 'carry1 vs. carry2'  C2    -1  1;
        contrast 'ilma vs. otsoni'    trt    1 -1;
        contrast 'nuori vs. aikuinen' ika    1 -1;
        contrast 'jakso1 vs. jakso2'  jakso  1  0 -1;
        contrast 'jakso1 vs. jakso3'  jakso  0  1 -1;
        contrast 'jakso2 vs. jakso3'  jakso -1  1  0;
        contrast 'carry1 vs. carry2'  C2    -1  1;
run;

Title 'mixed, fixed effects';
proc mixed data=koe2;
        class ika trt yksilo jakso C2;
        model sl_time=ika yksilo(ika) trt jakso C2 / solution ddfm=satterth;
        estimate 'ilma vs. otsoni'    trt    1 -1    / cl;
        estimate 'nuori vs. aikuinen' ika    1 -1    / cl;
        estimate 'jakso1 vs. jakso2'  jakso  1  0 -1 / cl;
        estimate 'jakso1 vs. jakso3'  jakso  0  1 -1 / cl;
        estimate 'jakso2 vs. jakso3'  jakso -1  1  0 / cl;
        estimate 'carry1 vs. carry2'  C2    -1  1    / cl;
        contrast 'ilma vs. otsoni'    trt    1 -1;
        contrast 'nuori vs. aikuinen' ika    1 -1;
        contrast 'jakso1 vs. jakso2'  jakso  1  0 -1;
        contrast 'jakso1 vs. jakso3'  jakso  0  1 -1;
        contrast 'jakso2 vs. jakso3'  jakso -1  1  0;
        contrast 'carry1 vs. carry2'  C2    -1  1;
run;

Title 'mixed,random subject';
proc mixed data=koe2;
        class ika trt yksilo jakso C2;
        model sl_time=ika trt jakso C2 / solution ddfm=satterth;
        random yksilo(ika);
```

```
        estimate 'ilma vs. otsoni'    trt    1 -1    / cl;
        estimate 'nuori vs. aikuinen' ika    1 -1    / cl;
        estimate 'jakso1 vs. jakso2'  jakso  1  0 -1 / cl;
        estimate 'jakso1 vs. jakso3'  jakso  0  1 -1 / cl;
        estimate 'jakso2 vs. jakso3'  jakso -1  1  0 / cl;
        estimate 'carry1 vs. carry2'  C2    -1  1    / cl;
        contrast 'ilma vs. otsoni'    trt    1 -1;
        contrast 'nuori vs. aikuinen' ika    1 -1;
        contrast 'jakso1 vs. jakso2'  jakso  1  0 -1;
        contrast 'jakso1 vs. jakso3'  jakso  0  1 -1;
        contrast 'jakso2 vs. jakso3'  jakso -1  1  0;
        contrast 'carry1 vs. carry2'  C2    -1  1;
run;


Title 'glm, fixed effects, ei viivästevaikutusta';
proc glm data=koe2;
        class ika trt yksilo jakso;
        model sl_time=ika yksilo(ika) trt jakso;
        estimate 'ilma vs. otsoni'    trt    1 -1;
        estimate 'nuori vs. aikuinen' ika    1 -1;
        estimate 'jakso1 vs. jakso2'  jakso  1  0 -1;
        estimate 'jakso1 vs. jakso3'  jakso  0  1 -1;
        estimate 'jakso2 vs. jakso3'  jakso -1  1  0;
        contrast 'ilma vs. otsoni'    trt    1 -1;
        contrast 'nuori vs. aikuinen' ika    1 -1;
        contrast 'jakso1 vs. jakso2'  jakso  1  0 -1;
        contrast 'jakso1 vs. jakso3'  jakso  0  1 -1;
        contrast 'jakso2 vs. jakso3'  jakso -1  1  0;
run;

Title 'mixed, fixed effects, ei viivästevaikutusta';
proc mixed data=koe2;
        class ika trt yksilo jakso;
        model sl_time=ika yksilo(ika) trt jakso / solution ddfm=satterth;
        estimate 'ilma vs. otsoni'    trt    1 -1    / cl;
        estimate 'nuori vs. aikuinen' ika    1 -1    / cl;
        estimate 'jakso1 vs. jakso2'  jakso  1  0 -1 / cl;
        estimate 'jakso1 vs. jakso3'  jakso  0  1 -1 / cl;
        estimate 'jakso2 vs. jakso3'  jakso -1  1  0 / cl;
        contrast 'ilma vs. otsoni'    trt    1 -1;
        contrast 'nuori vs. aikuinen' ika    1 -1;
        contrast 'jakso1 vs. jakso2'  jakso  1  0 -1;
        contrast 'jakso1 vs. jakso3'  jakso  0  1 -1;
        contrast 'jakso2 vs. jakso3'  jakso -1  1  0;
run;

Title 'mixed,random subject, ei viivästevaikutusta';
proc mixed data=koe2;
        class ika trt yksilo jakso;
        model sl_time=ika trt jakso / solution ddfm=satterth;
        random yksilo(ika);
        estimate 'ilma vs. otsoni'    trt    1 -1    / cl;
        estimate 'nuori vs. aikuinen' ika    1 -1    / cl;
        estimate 'jakso1 vs. jakso2'  jakso  1  0 -1 / cl;
        estimate 'jakso1 vs. jakso3'  jakso  0  1 -1 / cl;
        estimate 'jakso2 vs. jakso3'  jakso -1  1  0 / cl;
        contrast 'ilma vs. otsoni'    trt    1 -1;
        contrast 'nuori vs. aikuinen' ika    1 -1;
        contrast 'jakso1 vs. jakso2'  jakso  1  0 -1;
        contrast 'jakso1 vs. jakso3'  jakso  0  1 -1;
        contrast 'jakso2 vs. jakso3'  jakso -1  1  0; run; quit;
```

```
*DESCRIPTION:   t-testi, estimaatit ja luottamusvälit
*NAME:          t_test.sas
*AUTHOR:        Auli Partanen
*DATE:          1JUNE2000

options nodate nonumber nocenter linesize=132;
libname data 'h:\gradu\last_version\sas\';

**Aineiston muokkausta analysoitavaan muotoon;

data anova; set data.koe;
keep yksilo jakso seq ika sl_time C2;
proc sort; by sl_time yksilo seq;
run;

data anova1; set anova;
keep yksilo seq ryhma jakso sl_time C2;
if seq=212 then ryhma=1;
if seq=121 then ryhma=2;
ryhma=ryhma*1;
run;
proc sort; by ryhma seq yksilo jakso;
run;

proc transpose out=anova2;
var sl_time jakso;
by ryhma seq yksilo;
run;

data data.anova3; set anova2;
keep jakso1 jakso2 jakso3 yksilo seq ryhma;
if _name_='JAKSO' then delete;
jakso1=col1;
jakso2=col2;
jakso3=col3;
run;

**Käsittelyvaikutuksen kontrastimuuttuja;

data otsoni; set data.anova3;
*ryhmä1*;
if yksilo in (1,2,3,4,5,10,11,12,13,14) then jakso1=jakso1;
if yksilo in (1,2,3,4,5,10,11,12,13,14) then jakso2=jakso2*(-0.5);
if yksilo in (1,2,3,4,5,10,11,12,13,14) then jakso3=jakso3*(-0.5);
*ryhma2*;
if yksilo in (6,7,8,9,15,16,17,18,19) then jakso1=jakso1;
if yksilo in (6,7,8,9,15,16,17,18,19) then jakso2=jakso2*(-0.5);
if yksilo in (6,7,8,9,15,16,17,18,19) then jakso3=jakso3*(-0.5);

otsoni=jakso1+jakso2+jakso3;
run;
proc sort; by ryhma otsoni yksilo;
run;

Title 'Käsittelyvaikutuksen t-testi, estimaatti ja luottamusväli';

proc ttest data=otsoni;
class ryhma;
var otsoni;
run;
```

```
**Viivästevaikutuksen kontrastimuuttuja;

data carry; set data.anova3;
*ryhmä1*;
if yksilo in (1,2,3,4,5,10,11,12,13,14) then jakso1=jakso1*(-1);
if yksilo in (1,2,3,4,5,10,11,12,13,14) then jakso2=0;
if yksilo in (1,2,3,4,5,10,11,12,13,14) then jakso3=jakso3;
*ryhma2*;
if yksilo in (6,7,8,9,15,16,17,18,19) then jakso1=jakso1*(-1);
if yksilo in (6,7,8,9,15,16,17,18,19) then jakso2=0;
if yksilo in (6,7,8,9,15,16,17,18,19) then jakso3=jakso3;

carry=jakso1+jakso2+jakso3;
run;
proc sort; by ryhma carry yksilo;
run;

Title 'Viivästevaikutuksen t-testi, estimaatti ja luottamusväli';

proc ttest data=carry;
class ryhma;
var carry;
run;


*DESCRIPTION:WMW-testi
*NAME:          WMW.sas
*AUTHOR:        Auli Partanen
*DATE:          1JUNE2000

options nodate nonumber nocenter linesize=132 formchar='___';
libname data 'h:\gradu\last_version\sas\';

**Käsittelyvaikutuksen kontrastimuuttuja;

data otsoni; set data.anova3;
*ryhmä1*;
if yksilo in (1,2,3,4,5,10,11,12,13,14) then jakso1=jakso1;
if yksilo in (1,2,3,4,5,10,11,12,13,14) then jakso2=jakso2*(-0.5);
if yksilo in (1,2,3,4,5,10,11,12,13,14) then jakso3=jakso3*(-0.5);
*ryhma2*;
if yksilo in (6,7,8,9,15,16,17,18,19) then jakso1=jakso1;
if yksilo in (6,7,8,9,15,16,17,18,19) then jakso2=jakso2*(-0.5);
if yksilo in (6,7,8,9,15,16,17,18,19) then jakso3=jakso3*(-0.5);

otsoni=jakso1+jakso2+jakso3;
run;
proc sort; by ryhma otsoni yksilo;
run;

Title 'Käsittelyvaikutuksen WMW-testi';

proc npar1way data=otsoni
wilcoxon;
class ryhma;
var otsoni;
run;

**Viivästevaikutuksen kontrastimuuttuja;

data carry; set data.anova3;
```

```
*ryhmä1*;
if yksilo in (1,2,3,4,5,10,11,12,13,14) then jakso1=jakso1*(-1);
if yksilo in (1,2,3,4,5,10,11,12,13,14) then jakso2=0;
if yksilo in (1,2,3,4,5,10,11,12,13,14) then jakso3=jakso3;
*ryhma2*;
if yksilo in (6,7,8,9,15,16,17,18,19) then jakso1=jakso1*(-1);
if yksilo in (6,7,8,9,15,16,17,18,19) then jakso2=0;
if yksilo in (6,7,8,9,15,16,17,18,19) then jakso3=jakso3;

carry=jakso1+jakso2+jakso3;
run;
proc sort; by ryhma carry yksilo;
run;

Title 'Viivästevaikutuksen WMW-testi';

proc npar1way data=carry
wilcoxon;
class ryhma;
var carry;
run;


*DESCRIPTION:optimaalinen kontrasti:t-testi, estimaatit ja luottamusväli
*NAME:          op_ttest.sas
*AUTHOR:        Auli Partanen
*DATE:          1JUNE2000

options nodate nonumber nocenter linesize=132;
libname data 'h:\gradu\last_version\sas\';

**Käsittelyvaikutuksen optimaalinen kontrastimuuttuja;

data otsoni; set data.anova3;
*ryhmä1*;
if yksilo in (1,2,3,4,5,10,11,12,13,14) then jakso1=jakso1*(0.25);
if yksilo in (1,2,3,4,5,10,11,12,13,14) then jakso2=jakso2*(-0.5);
if yksilo in (1,2,3,4,5,10,11,12,13,14) then jakso3=jakso3*(0.25);
*ryhma2*;
if yksilo in (6,7,8,9,15,16,17,18,19) then jakso1=jakso1*(0.25);
if yksilo in (6,7,8,9,15,16,17,18,19) then jakso2=jakso2*(-0.5);
if yksilo in (6,7,8,9,15,16,17,18,19) then jakso3=jakso3*(0.25);

otsoni=jakso1+jakso2+jakso3;
run;
proc sort; by ryhma otsoni yksilo;
run;

Title 'Optimaalinen kontrastimuuttuja: käsittelyvaikutuksen t-testi, estimaatti ja
luottamusväli';

proc ttest data=otsoni;
class ryhma;
var otsoni;
run;


*DESCRIPTION:   Optimaalinen kontrastimuuttuja: WMW-testi
*NAME:          Opt_WMW.sas
*AUTHOR:        Auli Partanen
*DATE:          1JUNE2000

options nodate nonumber nocenter linesize=132 formchar='___';
```

```
libname data 'h:\gradu\last_version\sas\';

**Käsittelyvaikutuksen optimaalinen kontrastimuuttuja;

data otsoni; set data.anova3;
*ryhmä1*;
if yksilo in (1,2,3,4,5,10,11,12,13,14) then jakso1=jakso1*(0.25);
if yksilo in (1,2,3,4,5,10,11,12,13,14) then jakso2=jakso2*(-0.5);
if yksilo in (1,2,3,4,5,10,11,12,13,14) then jakso3=jakso3*(0.25);
*ryhma2*;
if yksilo in (6,7,8,9,15,16,17,18,19) then jakso1=jakso1*(0.25);
if yksilo in (6,7,8,9,15,16,17,18,19) then jakso2=jakso2*(-0.5);
if yksilo in (6,7,8,9,15,16,17,18,19) then jakso3=jakso3*(0.25);

otsoni=jakso1+jakso2+jakso3;
run;
proc sort; by ryhma otsoni yksilo;
run;

Title 'Optimaalinen kontrastimuuttuja: käsittelyvaikutuksen WMW-testi';

proc npar1way data=otsoni
wilcoxon;
class ryhma;
var otsoni;
run;

data w;
p_c2=          2*probnorm(-0.1633);
p_f2=          2*probnorm(-0.3266);
p_opt_f2=      2*probnorm(-2.2123);
run;

*DESCRIPTION:   Erotusten laskeminen Hodgess-Lehmannin estimaatteja varten
*NAME:          hl_estim4.sas
*AUTHOR:        Auli Partanen
*DATE:          8JUNE2000
*MODIFIED:

options nodate nonumber nocenter linesize=132;
libname data 'h:\gradu\last_version\sas\';

*käsittelyvaikutuksen kontrastimuuttuja*;

data otsoni; set data.anova3;
*ryhmä1*;
if yksilo in (1,2,3,4,5,10,11,12,13,14) then jakso11=jakso1;
if yksilo in (1,2,3,4,5,10,11,12,13,14) then jakso12=jakso2*(-0.5);
if yksilo in (1,2,3,4,5,10,11,12,13,14) then jakso13=jakso3*(-0.5);
*ryhma2*;
if yksilo in (6,7,8,9,15,16,17,18,19) then jakso11=jakso1;
if yksilo in (6,7,8,9,15,16,17,18,19) then jakso12=jakso2*(-0.5);
if yksilo in (6,7,8,9,15,16,17,18,19) then jakso13=jakso3*(-0.5);
kontmja1=jakso11+jakso12+jakso13;
if yksilo in (6,7,8,9,15,16,17,18,19) then kontmja2=kontmja1;
run;
proc sort; by ryhma yksilo;
run;

Title 'otsonikäsittelyn kontrastimuuttujat';
proc print data=otsoni;
var yksilo kontmja1 kontmja2; run;
```

```
*data otsonikäsittelyn ensimmäisen ja toisen ryhmän kontrastimuuttujien erotuksista;
*erotukset nimetty: kontrastimuuttuja, josta vähennetään+kontrastimuuttuja ryhmän
sisällä,
jonka arvo vähennetään;

%macro erotus(nimi1,nimi2,nimi3,nimi4,nimi5,nimi6,nimi7,nimi8,nimi9,nimi10,
              kont1,kont2,kont3,kont4,kont5,kont6,kont7,kont8,kont9,kont10);
data erF21_2; set otsoni;
if yksilo in (6,7,8,9,15,16,17,18,19) then k1_&nimi1=&kont1-kontmja2;
if yksilo in (6,7,8,9,15,16,17,18,19) then k1_&nimi2=&kont2-kontmja2;
if yksilo in (6,7,8,9,15,16,17,18,19) then k1_&nimi3=&kont3-kontmja2;
if yksilo in (6,7,8,9,15,16,17,18,19) then k1_&nimi4=&kont4-kontmja2;
if yksilo in (6,7,8,9,15,16,17,18,19) then k1_&nimi5=&kont5-kontmja2;
if yksilo in (6,7,8,9,15,16,17,18,19) then k1_&nimi6=&kont6-kontmja2;
if yksilo in (6,7,8,9,15,16,17,18,19) then k1_&nimi7=&kont7-kontmja2;
if yksilo in (6,7,8,9,15,16,17,18,19) then k1_&nimi8=&kont8-kontmja2;
if yksilo in (6,7,8,9,15,16,17,18,19) then k1_&nimi9=&kont9-kontmja2;
if yksilo in (6,7,8,9,15,16,17,18,19) then k1_&nimi10=&kont10-kontmja2;

%mend erotus;

*kont1-kont10-muuttujat saatu proc print-tulostuksesta;

%erotus(k21,k22,k23,k24,k25,k26,k27,k28,k29,k210,
18.0, 3.75, 11.25, 13.35, -4.15, -4.35, -5.45, 16.6, 3, -5.75);
run;

proc sort data=erF21_2;
by yksilo;
run;

proc transpose out=jarj;
var k1_k21 k1_k22 k1_k23 k1_k24 k1_k25 k1_k26 k1_k27 k1_k28 k1_k29 k1_k210 ;
by yksilo;
run;

data trt_diff; set jarj;
ero=col1;
if ero=. then delete;
keep yksilo ero;
run;

proc sort data=trt_diff;
by ero;
run;

proc print data=trt_diff;
var yksilo ero;
run;


*viivästevaikutuksen kontrastimuuttuja*;

data carry; set anova3;
*ryhmä1*;
if yksilo in (1,2,3,4,5,10,11,12,13,14) then jakso11=jakso1*(-1);
if yksilo in (1,2,3,4,5,10,11,12,13,14) then jakso12=0;
if yksilo in (1,2,3,4,5,10,11,12,13,14) then jakso13=jakso3;
*ryhma2*;
if yksilo in (6,7,8,9,15,16,17,18,19) then jakso11=jakso1*(-1);
if yksilo in (6,7,8,9,15,16,17,18,19) then jakso12=0;
```

```
if yksilo in (6,7,8,9,15,16,17,18,19) then jakso13=jakso3;
kontmja1=jakso11+jakso12+jakso13;
if yksilo in (6,7,8,9,15,16,17,18,19) then kontmja2=kontmja1;
run;
proc sort; by ryhma yksilo;
run;

Title 'otsonikäsittelyn viivästevaikutuksen kontrastimuuttujat';
proc print data=carry;
var yksilo kontmja1 kontmja2;
run;

*data otsonikäsittelyn viivästevaikutuksen ensimmäisen
ja toisen ryhmän kontrastimuuttujien erotuksista;
*erotukset nimetty: kontrastimuuttuja, josta vähennetään+kontrastimuuttuja ryhmän
sisällä,
jonka arvo vähennetään;

%macro erotus(nimi1,nimi2,nimi3,nimi4,nimi5,nimi6,nimi7,nimi8,nimi9,nimi10,
              kont1,kont2,kont3,kont4,kont5,kont6,kont7,kont8,kont9,kont10);
data erC21_2; set otsoni;
if yksilo in (6,7,8,9,15,16,17,18,19) then k1_&nimi1=&kont1-kontmja2;
if yksilo in (6,7,8,9,15,16,17,18,19) then k1_&nimi2=&kont2-kontmja2;
if yksilo in (6,7,8,9,15,16,17,18,19) then k1_&nimi3=&kont3-kontmja2;
if yksilo in (6,7,8,9,15,16,17,18,19) then k1_&nimi4=&kont4-kontmja2;
if yksilo in (6,7,8,9,15,16,17,18,19) then k1_&nimi5=&kont5-kontmja2;
if yksilo in (6,7,8,9,15,16,17,18,19) then k1_&nimi6=&kont6-kontmja2;
if yksilo in (6,7,8,9,15,16,17,18,19) then k1_&nimi7=&kont7-kontmja2;
if yksilo in (6,7,8,9,15,16,17,18,19) then k1_&nimi8=&kont8-kontmja2;
if yksilo in (6,7,8,9,15,16,17,18,19) then k1_&nimi9=&kont9-kontmja2;
if yksilo in (6,7,8,9,15,16,17,18,19) then k1_&nimi10=&kont10-kontmja2;

%mend erotus;

*kont1-kont10-muuttujat saatu proc print-tulostuksesta;

%erotus(k21,k22,k23,k24,k25,k26,k27,k28,k29,k210,
-21.7,-4.4,-8.8,-14.1,3.3,7.5,5.9,-14.9,-2.0,9.4);
run;

proc sort data=erC21_2;
by yksilo;
run;

proc transpose out=jarj;
var k1_k21 k1_k22 k1_k23 k1_k24 k1_k25 k1_k26 k1_k27 k1_k28 k1_k29 k1_k210 ;
by yksilo;
run;

data carry_di; set jarj;
ero=col1;
if ero=. then delete;
keep yksilo ero;
run;

proc sort data=carry_di;
by ero;
run;

proc print data=carry_di;
var yksilo ero;
run;
```

```
*optimaalinen käsittelyvaikutuksen kontrastimuuttuja*;

data op_otson; set anova3;
*ryhmä1*;
if yksilo in (1,2,3,4,5,10,11,12,13,14) then jakso11=jakso1*(0.25);
if yksilo in (1,2,3,4,5,10,11,12,13,14) then jakso12=jakso2*(-0.5);
if yksilo in (1,2,3,4,5,10,11,12,13,14) then jakso13=jakso3*(0.25);
*ryhma2*;
if yksilo in (6,7,8,9,15,16,17,18,19) then jakso11=jakso1*(0.25);
if yksilo in (6,7,8,9,15,16,17,18,19) then jakso12=jakso2*(-0.5);
if yksilo in (6,7,8,9,15,16,17,18,19) then jakso13=jakso3*(0.25);
kontmja1=jakso11+jakso12+jakso13;
if yksilo in (6,7,8,9,15,16,17,18,19) then kontmja2=kontmja1;
run;
proc sort; by ryhma yksilo;
run;

Title 'otsonin käsittelyvaikutuksen optimaaliset kontrastimuuttujat';
proc print data=op_otson;
var yksilo kontmja1 kontmja2;
run;

*data otsonikäsittelyn viivästevaikutuksen ensimmäisen
ja toisen ryhmän kontrastimuuttujien erotuksista;
*erotukset nimetty: kontrastimuuttuja, josta vähennetään+kontrastimuuttuja ryhmän
sisällä,
jonka arvo vähennetään;

%macro erotus(nimi1,nimi2,nimi3,nimi4,nimi5,nimi6,nimi7,nimi8,nimi9,nimi10,
             kont1,kont2,kont3,kont4,kont5,kont6,kont7,kont8,kont9,kont10);
data erOF21_2; set otsoni;
if yksilo in (6,7,8,9,15,16,17,18,19) then k1_&nimi1=&kont1-kontmja2;
if yksilo in (6,7,8,9,15,16,17,18,19) then k1_&nimi2=&kont2-kontmja2;
if yksilo in (6,7,8,9,15,16,17,18,19) then k1_&nimi3=&kont3-kontmja2;
if yksilo in (6,7,8,9,15,16,17,18,19) then k1_&nimi4=&kont4-kontmja2;
if yksilo in (6,7,8,9,15,16,17,18,19) then k1_&nimi5=&kont5-kontmja2;
if yksilo in (6,7,8,9,15,16,17,18,19) then k1_&nimi6=&kont6-kontmja2;
if yksilo in (6,7,8,9,15,16,17,18,19) then k1_&nimi7=&kont7-kontmja2;
if yksilo in (6,7,8,9,15,16,17,18,19) then k1_&nimi8=&kont8-kontmja2;
if yksilo in (6,7,8,9,15,16,17,18,19) then k1_&nimi9=&kont9-kontmja2;
if yksilo in (6,7,8,9,15,16,17,18,19) then k1_&nimi10=&kont10-kontmja2;

%mend erotus;

*kont1-kont10-muuttujat saatu proc print-tulostuksesta;

%erotus(k21,k22,k23,k24,k25,k26,k27,k28,k29,k210,
1.725,0.45,4.65,2.775,-1.675,1.275,-1.025,5.425,1.5,1.3);
run;

proc sort data=erOF21_2;
by yksilo;
run;

proc transpose out=jarj;
var k1_k21 k1_k22 k1_k23 k1_k24 k1_k25 k1_k26 k1_k27 k1_k28 k1_k29 k1_k210 ;
by yksilo;
run;

data otrt_di; set jarj;
```

```
ero=col1;
if ero=. then delete;
keep yksilo ero;
run;

proc sort data=otrt_di;
by ero;
run;

proc print data=otrt_di;
var yksilo ero;
run;
```

Tilastotoimen menetelmien maisteriohjelman pro gradu -tutkielma sarja

1. Salmikuukka, J. (1997)	Aikasarjojen perusrakennemalleista ja niiden soveltaminen Jyväskylän kaukolämmön kulutuksen analysointiin ja ennustamiseen. (76 s., 1 liite)	Jyväskylän Energia Oy, Jyväskylä
2. Yrjölä, T. (1997)	Lasten päivähoidon tuottavuusvertailu suurissa kaupungeissa DEA-menetelmällä. (72 s., 2 liitettä)	Jyväskylän kaupungin terveystoimi, Jyväskylä
3. Ainiala, N. (1997)	Helsingin osa-alueiden työvoimatilastojen estimointi pienaluetekniikalla valtakunnallisesta työvoimatutkimuksesta. (73 s., 3 liitettä)	Helsingin kaupungin tietokeskus, Helsinki
4. Puhakka, E. (1997)	Kiintiöpoiminnan tilastolliset ominaisuudet pk-yritysbarometritutkimuksessa. Sovelluksena 2/1996 aineisto (82 s., 2 liitettä) (salainen)	Tietoykkönen, Jyväskylä
5. Salonen, R. (1997)	Muutoksen ja tason estimointi rotatoivassa paneeliaineistossa eri estimaattoreiden avulla. Sovellus työvoimatutkimuksen aineistoon. (39 s, 4 liitettä)	Tilastokeskus, Helsinki
6. Kunttu, S. (1997)	Alueellisen teollisuustuotannon volyymi-indeksin estimointi Etelä-Pohjanmaalle. (103 s.)	Tilastokeskus, Seinäjoki
7. Koponen, M. (1997)	Epätäydelliseen otantakehikkoon perustuvan yritysaineiston estimointi. Käsiteanalyysiä ja soveluksia Tilastokeskuskuksen järjestäytymättömiä yrityksiä koskevaan palkkatiedusteluun. (56 s., 5 liitettä)	Tilastokeskus, Helsinki
8. Kainulainen, P. (1997)	Toimialatilastot sanomalehdessä: Monimuuttujaisen rekisteriaineiston havainnollistaminen ja tuotteistaminen (62 s., 6 liitettä)	Sanomalehtien liitto, Helsinki
9. Storfors, S. (1998)	Toistojen määrän arviointi kokeellisessa tutkimuksessa. (34 s., 3 liitettä)	Cultor Oy, Helsinki
10. Peltoniemi-Laajanen, L. (1998)	Pääteiden kehittämistarpeet: kuljettajahaastattelun monimuuttuja-analyysi (78 s., 8 liitettä)	VTT Yhdyskuntatekniikka, Helsinki
11. Hirvonen, M. (1999)	Vastauskadon käsittely yritysten innovaatiokyselyssä 1996 (81s., 9 liitettä)	Tilastokeskus, Helsinki
12. Nevalainen, J. (1999)	Välttöfunktion estimointi kliinisessä toistomittauskokeessa (56 s., 5 liitettä)	Leiras, Helsinki
13. Vierinkari, K. (1999)	Aaltojuotoskoneen parametrien optimointi Taguchi-menetelmällä.(92 s, salainen)	Essex Communication EMS Oy, Äänekoski
14. Pylvänäinen, I. (1999)	Kliinisen toistomittauskokeen analysointi toisistaan riippuvien binääri- ja ordinaali-vasteiden tapauksessa. (82 s., 3 liitettä)	Orion Oyj, Turku

15. Saarnio, K. (1999)	Työvoimaa kuvaavien tunnuslukujen estimointi pienaluetasolla kuukausittaisesta työvoimatutkimuksesta; hierarkkisten mallien sovellus. (107 s., 4 liitettä)	Tilastokeskus, Helsinki
16. Lamberg, P. (1999)	Pienalue-estimaatteja pitkäaikaissairastavuudelle seutukuntatasolla. (76 s., 4 liitettä)	Tilastokeskus, Helsinki
17. Piela, P, (1999)	Massaimputointi kaupan lyhyen aikavälin tilastotuotannossa. (120 s., 12 liitettä, salainen)	Tilastokeskus Helsinki
18. Tuominen, T. (1999)	Vastausrasite ja sen mittaaminen Tilastokeskuksen yrityssurveyssä 1998. (173 s)	Tilastokeskus, Helsinki
19. Pohjanjousi, P. (2000)	Kiinteiden ja satunnaisvaikutusten meta-analyysi (65 s.+ 29 s. liitteitä)	Orion Helsinki
20. Partanen, A. (2000)	Parametriset ja epäparametriset menetelmät vaihtovuoroasetelmassa (70 s.)	Orion, Turku