KALA- JA RIISTARAPORTTEJA nro 256

Kati Kekäläinen

Vastauskato ja otantayksikköongelma vapaa-ajankalastuskyselyissä

Helsinki 2002

RIISTAN- JA KALANTUTKIMUS

KUVAILULEHTI

Julkaisija	Julkaisuaika
Riista- ja kalatalouden tutkimuslaitos	Kesäkuu 2002

Tekijä(t)

Kati Kekäläinen

Julkaisun nimi

Vastauskato ja otantayksikköongelma vapaa-ajankalastuskyselyissä

Julkaisun laji	Toimeksiantaja	Toimeksiantopäätöksen määrä
Tutkimusraportti		

Projektin nimi ja numero

Tiivistelmä

Tässä raportissa käsitellään vastauskadosta aiheutuvia ongelmia ja niiden ratkaisemista vapaa-ajankalastuskyselyissä. Lisäksi tutkitaan myös otantayksikköongelmaa. Raportti pohjautuu Jyväskylän yliopistossa 9.4. 2002 julkaistuun pro gradu-tutkimukseen: Hierarkkinen jälkiositus, kalibrointi ja katopainotus surveytutkimuksessa: sovellus vapaa-ajankalastuskyselyyn.

Raportissa käsitellään kato-ongelman ratkaisua sekä painotus- että imputointimenetelmillä. Lisäksi etsitään ratkaisuja tilanteisiin, joissa edellisten menetelmien vaatimaa lisätietoa ei ole saatavissa rekistereistä. Tällaisia tilanteita varten esitellään hierarkkinen jälkiosltus, joka voidaan muodostaa vastaamattomista poimitun jälkikyselyn perusteella.

Esimerkkiaineistona raportissa käytetään RKTL:n Vapaa-ajankalastus 1998 -aineistoa, johon esiteltyjä menetelmiä sovelletaan ja näitä estimaatteja vertaillaan alkuperäisiin. Hierarkkisen jälkiosituksen havaitaan antavan pienempiä estimaatteja saalismäärälle ja kalastuspäiville kuin alkuperäisessä aineistossa käytetty kalibrointipainotus.

Asiasanat

vastauskato, otantayksikköongelma, imputointi, kalibrointipainotus, jälkikysely, hierarkkinen jälkiositus

Sarjan nimi ja numero	ISBN	ISSN
Kala- ja riistaraportteja 256	951-776-373-5	1238-3325

Sivumäärä	Kieli	Hinta	Luottamuksellisuus
55 s.	suomi		

Jakelu	Kustantaja
Riista- ja kalatalouden tutkimuslaitos Pukinmäenaukio 4, PL 6 00721 Helsinki Puh. 0205 7511 Faksi 0205 751 201	Riista- ja kalatalouden tutkimuslaitos PL 6 00721 Helsinki Puh. 0205 7511 Faksi 0205 751 201

Sisältö

1 Johdanto

Vapaa-ajankalastusta harrasti Suomessa vuonna 2000 noin kaksi miljoonaa henkilöä. Kotitalouksia, joissa joku jäsenistä on kalastanut, oli puolestaan noin 1,1 miljoonaa. Vuonna 2000 vapaa-ajankalastajien yhteissaaliin määräksi arvioitiin yli 4 milj. kg. Tämän saalismäärän arvo on ammattikalastajien keskihinnoilla arvioituna lähes 50 milj. euroa. Vapaa-ajankalastus Suomessa on siis hyvin laaja ilmiö sekä harrastajien määrän että saaliin arvon mukaan tarkasteltuna. Siihen liittyy paitsi huomattavia taloudellisia arvoja myös virkistyksellisiä ja elämyksellisiä arvoja.

1.1 Vapaa-ajankalastuskyselyjen tavoitteet

Riista- ja kalatalouden tutkimuslaitos tutkii vapaa-ajankalastusta useantyyppisillä tutkimuksilla. Vapaa-ajankalastajien määrää, saaliista ja kalastukseen käytettyä aikaa halutaan selvittää myös alueittain, pyydyksittäin ja lajeittain. Vapaa-ajankalastajien profiilitutkimuksessa kerätään tietoja mm. kalastajien ikä- ja sukupuolijakaumasta sekä tarkastellaan erikseen tiettyjen erityisryhmien, kuten nuorten, kalastamista. Vesialueiden omistajien profiilitutkimuksessa puolestaan tutkitaan kalavesien omistajia sekä näiden ja eri organisaatioiden toimintaa kalavesien käyttämisessä ja hoidossa. Tavoitteena on tuottaa tietoa, jonka avulla päätoksentekojärjestelmää voitaisiin kehittää vastaamaan entistä paremmin eri intressiryhmien tarpeita. Myös vapaa-ajankalastajien kalastuksessaan kokemia ongelmia pyritään selvittämään, jotta olemassa oleviin epäkohtiin voitaisiin puuttua.

Kalastuksen jakautuminen eri vesialueille on tärkeä tutkimuksen kohde, sillä sen perusteella jaetaan valtion keräämiä kalastusvaroja takaisin kalastusalueille käytettäväksi kalavesien hoitoon. Myös kalastusmatkailua tutkitaan, jotta sitä pystyttäisiin kehittämään ja laajentamaan.

Ilmiönä vapaa-ajankalastus on monimutkainen ja vaikeasti ennustettavissa, sillä se ei kovinkaan voimakkaasti riippuvainen rekistereistä saatavista taustatiedoista, kuten iästä, sukupuolesta, sosioekonomisesta taustasta tai asuinpaikasta. Alueelliset erot ovat suuria, mutta kalastamisaktiivisuuteen vaikuttaa paitsi karkea aluejako myös kuntatyyppi, asuinkunnan koko, lähivesistöjen laatu jne. Siten myöskään aluetiedolla ei pystytä ennustamaan kalastamista tarkasti. Lisäksi vapaa-ajankalastajista suuri osa on ns. satunnaiska-

lastajia. Näitä ovat lähinnä kesämökeillään kalastavat ihmiset, jotka saattavat kalastaa hyvinkin satunnaisesti esimerkiksi lomasäidensä mukaan. Nämä satunnaisonkijat eivät välttämättä kalasta edes joka vuosi ja hankaloittavat siten kalastajien määrän arviointia.

Vapaa-ajankalastuskyselyt poimitaan yleensä väestörekistereistä ja suoritetaan postikyselyinä. Tietoa vapaa-ajankalastajien määristä saadaan myös luparekistereistä. Näistä ei kuitenkaan saada riittävän kattavaa tietoa, sillä onkimiseen tai pilkkimiseen ei tarvita lainkaan lupia eivätkä myöskään alle 18-vuotiaat tarvitse kalastuslupia. Luonnollisesti kalastusluparekistereistä ei myöskään saada tietoa henkilöistä, jotka kalastavat luvatta.

Myös aikaisemmista tutkimuksista saadaan aputietoa. RKTL:llä on pitkät perinteet kalastuskyselyiden tekemisessä ja esimerkiksi väestörekisteripohjaisia vapaa-ajankalaskyselyitä tehdään säännöllisesti joka toinen vuosi. Lisäksi mm. Tilastokeskuksen työvoimatiedustelujen yhteydessä on selvitetty kalastaneiden määriä.

1.2 Otanta väestörekisteristä

Väestörekisteri tarjoaa kattavan poimintakehikon surveytutkimuksille, joissa ollaan kiinnostuneita henkilöistä tai kotitalouksista. Myös vapaa-ajankalastuskyselyiden yhteydessä otos poimitaan yleensä väestörekisteristä, sillä sopivaa kehikkoa, jolla kalastaneet saataisiin seulotuksi muusta väestöstä, ei ole olemassa. Kalastusluparekistereihin perustuvat kehikot eivät ole tarpeeksi kattavia, sillä esimerkiksi pelkästään onkimista tai pilkkimistä harrastavat sekä lapset ja nuoret jäävät näiden rekisterien ulkopuolelle.

Västörekisteri on hyvä poimintakehikko myös siksi, että se tarjoaa käyttöön erilaisia apumuuttujia, joilla voidaan mahdollisesti tehostaa estimointia. Tällaisia aputietoja ovat mm. henkilön ikä, sukupuoli, ammatti sekä osoitetiedot. Vapaa-ajankalastuskyselyjen kohdalla nämä taustatiedot eivät kuitenkaan yleensä korreloi riittävän voimakkaasti tutkimusmuuttujien kanssa, jotta niillä pystyttäisiin tehostamaan estimointia.

Kalastuskyselyissä käytetään usein ositettua otantaa, missä ositteet muodostetaan maantieteellisin perustein. Tällä pyritään siihen, että otos on mahdollisimman kattava, eli kaikilta kiinnostavilta alueilta saadaan havaintoja otokseen. Käytettäessä väestörekisteriä poimintakehikkona maantieteellis-

ten ositteiden muodostaminen on helppoa, sillä väestörekisteri sisältää tiedot henkilöiden asuinpaikasta.

Myös kaupallisia rekistereitä voidaan käyttää poimintakehikkoina. Tällaisia palveluja on mm. Tilastokeskuksella, Suomen Gallupilla ja Atkos Oy:llä. Näistä rekistereistä voi löytyä monipuolisempaa taustatietoa kuin väestörekisteristä.

Otantayksikköongelmalla tarkoitetaan tapauksia, joissa poimintayksikkö ja tilastoyksikkö eivät täysin vastaa toisiaan. Poimintayksiköllä tarkoitetaan poimintakehikon muodostavia yksilöitä, joista otos poimitaan. Tilastoyksikkö puolestaan on tutkimuksen kiinnostuksen kohde, jota koskevia tunnuslukuja halutaan estimoida. Otantayksikköongelmaan päädytään silloin, kun kattavaa rekisteriä tai muuta poimintakehikkona käytettävää listaa tilastoyksiköistä ei ole saatavissa, vaan poiminta joudutaan suorittamaan kehikosta, jossa listatut poimintayksiköt poikkeavat tilastoyksiköistä. Poiminta- ja otantayksikön välinen ero voi aiheuttaa harhaa estimointeihin, joten se täytyy aina ottaa huomioon. Harhan korjaamiseen voidaan käyttää sopivaa painotusta.

Vapaa-ajankalastuskyselyissä tuloksia halutaan yleensä estimoida kotitaloutta kohti. Tämä johtuu mm. siitä, että näin saadaan selvitetyksi myös alaikäisten lasten ja nuorten kalastamista, mikä ei muuten onnistuisi, sillä lapsiin ei voi kohdistaa postikyselyä. Väestörekisteristä saadaan tieto siitä, mihin asuntokuntaan henkilö kuuluu. Kotitalouden ja asuntokunnan määritelmät poikkevat kuitenkin toisistaan. Tämä ero on selitetty tarkemmin luvussa 2.1. Näin ollen myös vapaa-ajankalastuskyselyissä törmätään otantayksikköongelmaan. Tätä ongelmaa voidaan korjata käyttämällä kalibrointipainotusta, joka on tarkemmin kuvattu luvussa 7.

Otantamenetelmänä väestörekisteripohjaisissa surveytutkimuksissa käytetään systemaattista otantaa, joka voidaan suorittaa myös ositettuna. Estimoinnit palautuvat tällöin yksinkertaisen satunnaisotannan tapaukseen tai, käytettäessä ositettua otantaa, estimoinnit suoritetaan ositteiden sisällä kuten yksinkertaisen satunnaisotannan tapauksessa. Näistä ositteittain estimoiduista tunnusluvuista lasketaan sitten sopivasti painottamalla yhdistetty estimaatti koko aineistolle.

Väestörekisteripohjaisissa kyselyissä on mahdollista käyttää myös ryväsotantaa tai PPS-otantaa (poiminta otosalkion koon mukaan). Näiden menetelmien käyttö edellyttää luonnollisesti sitä, että ryväsmuuttuja tai kokoa mittaava

muuttuja tiedetään kaikilta perusjoukon alkioilta. Rypäinä voivat olla esimerkiksi asuntokunnat. Kokomuuttujia, joita koskeva tieto voidaan saada väestörekisteristä, ovat mm. asuntokunnan tai kunnan koko.

1.3 Otoksesta laskettujen tietojen yleistäminen perusjoukon tasolle

Jotta otoksesta lasketut estimaatit voidaan yleistää koko perusjoukkoon, täytyy havaintoja painottaa sopivasti estimaattien laskemisessa. Jos aineisto on täydellinen, riittää käyttää ns. perusotospainoja, jotka saadaan otokseen sisältymistodennäköisyyden käänteislukuna. Täydellisen aineiston tilannetta ei käytännössä juuri koskaan esiinny, joten peruspainotusta joudutaan yleensä korjaamaan uudelleenpainotuksella.

Painot muodostetaan siten, että ne summautuvat koko perusjoukon lukumäärään. Joissakin tapauksissa voi olla kuitenkin tarpeen analysoida otosta sellaisenaan, eikä perusjoukon tasolla. Tällaiseen tilanteeseen päädytään esimerkiksi silloin kun analyyseissa joudutaan käyttämään ohjelmistoja, jotka eivät ole asetelmaperusteisia. Tällöin painot on skaalattava siten, että ne summautuvat otoskooksi.

Sopivalla painotuksella voidaan korjata sekä vastauskadon että otantayksikköongelman aiheuttamia vaikutuksia. Luvussa 7 on käsitelty tarkemmin kadon korjaamiseen suunniteltuja painotusmenetelmiä sekä kalibrointipainotusta, joka sopii sekä kadon että otantayksikköongelman oikaisemiseen.

1.4 Raportin sisältö

Tämä raportti käsittelee vastauskadon vaikutuksia ja niiden korjaamista kalastuskyselyissä. Raportti pohjautuu Jyväskylän yliopistossa 9.4.2002 julkaistuun pro gradu -työhön *Hierarkkinen jälkiositus, kalibrointi ja katopainotus surveytutkimuksessa: sovellus vapaa-ajankalastuskyselyyn* (Kekäläinen 2002). Sekä pro gradu -tutkimus että tämä raportti ovat osa tilastotoimen maisteriohjelmaan kuuluvaa tutkimusharjoitteluani, jonka suoritin yhteistyössä Riista- ja kalatalouden tutkimuslaitoksen kanssa.

Raportin toisessa luvussa esitellään esimerkkitapauksena käsiteltävä RKTL:n Vapaa-ajankalastus 1998 -aineisto sekä aineiston poiminta. Kolmannessa lu-

vussa käsitellään vastauskatoa otantatutkimuksissa. Lisäksi tutkitaan vastauskadon mallittamista. Hierarkkinen jälkiositus on eräs keino kadon korjaamiseen ja sitä esitellään tarkemmin raportin neljännessä luvussa. Tämän menetelmän soveltamista Vapaa-ajankalastus 1998 -aineistoon kuvataan puolestaan luvussa 5. Viidennessa luvussa tehdään myös muita empiirisiä tarkasteluja aineistolle. Esimerkiksi koetetaan arvioida vastaamattomien kalastuskäyttäytymistä kontaktiryhmien avulla. Puuttuvien havaintojen ongelmaa tutkitaan luvussa 6 ja seitsemäs luku puolestaan käsittelee painotusmenetelmiä kadon korjaamisessa.

2 Aineistokuvaus

Tutkimusaineisto koostuu Riista- ja kalatalouden tutkimuslaitoksen Vapaa-ajankalastus 1998 -kyselyn tuloksista (Vapaa-ajankalastus 1998). Riista- ja kalatalouden tutkimuslaitos suorittaa vapaa-ajankalastusta koskevan kyselytutkimuksen joka toinen vuosi. Tämä vuoden 1998 aikana tapahtunutta kalastamista koskeva kysely suoritettiin vuoden 1999 alussa. Kyselyn tavoitteena on selvittää mm. kalastaneiden henkilöiden ja kotitalouksien määriä, kalastuspäiviä, saalismääriä, eri pyydysten käyttöä ja kalastuksen jakautumista eri vesialueiden, pyydysten ja kalalajien välillä vuonna 1998. Kyselyssä ei oteta huomioon ulkomaalaisten kalastamista Suomessa tai suomalaisten kalastamista ulkomailla.

2.1 Aineiston poiminta

Otos poimittiin väestörekisterikeskuksen väestötietojärjestelmästä ja otoskoko oli 4000 asuntokuntaa. Asuntokunta määritellään siten, että sen muodostavat samassa asuinhuoneistossa vakinaisesti asuvat henkilöt. Asuntokunta voi siis koostua yhdestä tai useammasta kotitaloudesta, sillä kotitalous puolestaan määritellään mm. rahojen yhteisen käytön suhteen.

Otantamenetelmänä käytettiin systemaattista ositettua otantaa. Ositteet muodostettiin maantieteellisin perustein. Poimintaa varten henkilöt luetteloitiin väestötietojärjestelmässä asuinpaikan mukaan siten, että samaan asuntokuntaan kuuluvat henkilöt ovat listassa peräkkäin. Tästä luettelosta poimittiin tasavälein 4000 henkilöä ja heidän edustamansa asuntokunta tuli mukaan otokseen. Koska poimintaväli oli riittävän pitkä, kaikki henkilöt kuuluivat eri asuntokuntiin. Poiminta kohdennettiin 18-74 -vuotiaisiin, eli jos valituksi tullut henkilö oli alle 18- tai yli 74-vuotias, asuntokunta jätettiin pois otoksesta ja seuraava ikäehdon täyttänyt henkilö ja hänen asuntokuntansa valittiin otokseen.

Aineisto koottiin postikyselyllä. Ensimmäiseen kyselyyn vastaamatta jättäneet saivat uuden kehotuksen, ja tähän kehotukseen vastaamattomat saivat vielä uuden kyselylomakkeen.

Vastausprosentiksi tuli 65, eli 2582 asuntokuntaa palautti kyselyn. Palautetuista lomakkeista jouduttiin kuitenkin hylkäämään vielä 96, sillä näistä ei

voitu päätellä, oliko kyseinen kotitalous kalastanut vai ei. Lopullisen aineiston kooksi tuli siten 2486 asuntokuntaa.

Poimintayksikkönä oli siis asuntokunta, mutta tilastoyksikkönä käytettiin kotitaloutta. Tästä erosta syntynyttä harhaa korjattiin kalibrointipainotuksella, jossa käytettiin reunajakaumina Tilastokeskuksen kotitalouskyselystä saatua kotitalouksien kokoluokkajakaumaa sekä väestörekisteristä saatuja naisten ja miesten ikäjakaumia lääneittäin.

2.2 Ositteiden muodostaminen

Ositteita on yhteensä kahdeksan ja ne muodostettiin asuinkunnan sijainnin (pääkaupunkiseutu, muu Etelä-Suomi, Länsi-Suomi, Itä-Suomi, Oulun lääni, Lappi ja Ahvenanmaa), kuntatyypin (kaupunkimainen, taajaan asuttu, maaseutumainen) sekä kunnan merellisyyden (saaristokunta, sisämaassa sijaitseva kunta ja rannikkokunta) mukaan. Seuraavalla sivulla olevassa taulukossa on esitetty ositteiden muodostuminen, koko ja otoskoot ositteittain.

Ensimmäinen osite koostuu Etelä- ja Länsi-Suomen lääneihin kuuluvista rannikkokunnista, jotka on luokiteltu kaupunkimaisiksi tai taajaan asutuiksi. Lisäksi tähän ositteeseen lasketaan mukaan pääkaupunkiseutu sekä Kaskisen kunta, vaikka se kuuluisikin oikeastaan saaristokuntiin. Toiseen ositteeseen kuuluvat Oulun ja Lapin läänien kaupunkimaiset ja taajaan asutut rannikkokunnat. Kolmas osite koostuu Etelä- ja Länsi-Suomen läänien maaseutumaisista rannikkokunnista. Lisäksi tähän ositteeseen on siirretty Luodon kunta. Neljäs osite sisältää Oulun ja Lapin läänien maaseutumaiset rannikkokunnat. Neljänteen ositteeseen on myös siirretty Hailuodon kunta. Viidennen ositteen muodostavat saaristokunnat. Saaristokuntien ositteesta on kuitenkin siirretty muihin ositteisiin Kaskinen, Luoto ja Hailuoto, jolloin tähän ositteeseen jäävät Ahvenanmaan kunnat ja Lounais-Suomen saaristo. Kuudes osite sisältää Etelä- ja Länsi-Suomen läänien kaupunkimaiset tai taajaan asutut sisämaan kunnat. Seitsemänteen ositteeseen kuuluvat Etelä- ja Länsi-Suomen läänien maaseutumaiset sisämaan kunnat sekä Itä-Suomen, Oulun ja Lapin läänien kaupunkimaiset ja taajaan asutut sisämaan kunnat. Kahdeksas osite koostuu puolestaan Itä-Suomen, Oulun ja Lapin läänien maaseutumaisista sisämaan kunnista

Taulukko 1: Ositteiden muodostaminen

osite	mistä alueista koostuu	henkilöiden määrä/osite	otoskoko (asuntokuntia/osite)
1	Etelä- ja Länsi-Suomen läänien kaupunkimaiset ja taajaan asutut rannikkokunnat (+ pääkaupunkiseutu & Kaskinen)	1199845	750
2	Oulun ja Lapin läänien kaupunkimaiset ja taajaan asutut rannikkokunnat	172769	250
3	Etelä- ja Länsi-Suomen läänien maaseutumaiset rannikkokunnat (+ Luoto)	104971	250
4	Oulun ja Lapin läänien maaseutumaiset rannikkokunnat (+ Hailuoto)	12816	250
5	Saaristokunnat (lukuunottamatta Kaskista, Luotoa ja Hailuotoa)	33641	500
6	Etelä- ja Länsi-Suomen läänien kaupunkimaiset ja taajaan asutut sisämaan kunnat	1069660	500
7	Etelä- ja Länsi-Suomen läänien maaseutumaiset sisämaan kunnat sekä Itä-Suomen, Oulun ja lapin läänien kaupunkimaiset ja taajaan asutut sisämaan kunnat	722743	1000
8	Itä-Suomen, Oulun ja Lapin läänien maaseutumaiset sisämaan kunnat	326328	500
	yhteensä	3642773	4000

2.3 Kyselylomake ja aineiston muuttujat

Otokseen valituksi tulleille postitse lähetetty kyselylomake on liitteessä 1. Liite 2 puolestaan sisältää listan aineistoon kuuluvista muuttujista ja näiden muuttujien selitykset.

Kyselylomakkeessa tiedusteltiin ensin kotitalouden jäsenien lukumäärää sekä ikä- ja sukupuolirakennetta. Toisessa kysymyksessä kysyttiin, oliko kukaan kotitaloudesta kalastanut vuoden 1998 aikana. Vaihtoehtoina olivat 1='kyllä, ja sai saalista', 2='kyllä, mutta kukaan ei saanut saalista', 3='ei, mutta on kalastanut tai ravustanut aikaisemmin' ja 4='ei ole kalastanut eikä ravustanut koskaan'.

Kolmannessa kysymyksessä selvitettiin kotitalouden kalastaneiden henkilöiden yhteislukumäärä sekä lukumäärät iän ja sukupuolen mukaan luokiteltuna. Myös kyselylomakkeen neljäs kohta koski kotitalouden kalastusaktiivisuutta. Kyselylomakkeeseen pyydettiin merkitsemään, kuinka moni kotitalouden jäsenistä kuului mihinkin seuraavista luokista: 1='ei kalastanut lainkaan', 2='osallistui kalastamiseen ainoastaan soutamalla tai ohjaamalla venettä', 3='kalastus oli yksi harrastus muiden joukossa', 4='kalastus oli tärkein tai lähes tärkein harrastus' ja 5='kalastus oli tärkein tai lähes tärkein harrastus ja myös kalastuskilpailuihin osallistuttiin'.

Lomakkeen viides kysymys koski kalastusmatkailua; siinä kysyttiin kotitalouden kalastusmatkailuun osallistuneiden henkilöiden lukumäärät ja kalastuspäivien määrät sekä lisäksi muuhun kalastukseen osallistuneiden henkilöiden määrät ja kalastuspäivät. Tämän kysymyksen vastauksia ei ole kuitenkaan raportoitu tuloksia analysoitaessa.

Kuudennessa kohdassa tiedusteltiin kalastusta eri pyydyksillä ja eri vesialueilla. Pyydykset oli jaettu kahdeksaan luokkaan (verkko, katiska/merta/rysä, pilkkivapa, onki, heittovapa, perhovapa, vetouistin ja muut pyydykset). Vesialueet oli jaettu ensin kahteen ryhmään: sisävesialue ja merialue. Lisäksi sisävesialue jakautui vielä viiteen osa-alueeseen: Etelä-Suomi, Länsi-Suomi, Itä-Suomi, Oulun lääni ja Lappi. Merialue puolestaan koostui neljästä osa-alueesta: Suomenlahti, Saaristomeri ja Ahvenanmaa, Selkämeri ja Merenkurk-

ku sekä Perämeri. Tässä kohdassa selvitettiin, onko kotitalous kalastanut kullakin alueista ja pyydyksistä sekä tiettyä pyydystä käyttäneiden ja tietyllä alueella kalastaneiden henkilöiden määriä.

Seuraavassa kysymyksessä puolestaan selvitettiin kalastuspäivien määrää kullakin vesialueella ja pyydyksellä. Pyydysten ja alueitten jako oli sama kuin kuudennessa kohdassa.

Lomakkeen kahdeksannessa kohdassa kysyttiin kalastusvuorokausia jäältä kalastuksessa ja avovesikalastuksessa. Näitä tietoja ei kuitenkaan raportoitu.

Kyselylomakkeen viimeisessä kohdassa pyydettiin täyttämään taulukko saadusta saaliista lajeittain ja pyydyksittäin. Lisäksi kunkin lajin kohdalla kysyttiin lajin tärkeintä kalastusaluetta.

Puuttuva tieto aineistossa (osittaisvastauskato) on merkitty imputointimuuttujilla, joita on yhteensä kuusi (*impla*, *implb*, *imp*3, *imp*4, *impsa* ja *impkhka*). Nämä muuttujat saavat arvon nolla, jos kyseessä on aito havainto ja arvon yksi, jos kyseinen havainto on imputoitu. (Poikkeuksena puuttuvaan saalistietoon liittyvä muuttuja *impsa*, joka voi saada myös arvon kaksi, jos kyseessä on osittain puuttuva tieto.) Myös nämä muuttujat on esitelty tarkemmin liitteessä 2.

3 Vastaamatta jättäneiden huomioonottaminen otantatutkimuksissa

3.1 Teoreettiset tarkastelut

Vastauskadon vaikutukset estimointituloksiin riippuvat kahdesta asiasta: kadon suuruudesta sekä valikoituneisuudesta. Kadon valikoituneisuudella tarkoitetaan sitä, että vastanneet ja vastaamattomat poikkeavat systemaattisesti toisistaan tutkimusmuuttujien suhteen. Vastauskato pienentää aina analysoitavan aineiston kokoa, jolloin jo tästä syystä estimaattien tarkkuus heikkenee. Tarkkuuden heikkeneminen on sitä vakavampi ongelma, mitä suurempaa vaje aineistossa on. Tarkkuuden heikkenemistä pystytään kuitenkin arvioimaan, sillä puuttuvien havaintojen osuus tiedetään.

Toinen vastauskadon aiheuttama ongelma on estimaattien harhaisuuden lisääntyminen. Tämä on myös hankalammin korjattava ongelma kuin estimaattien epätarkkuus, sillä harhan suuruusluokkaa ja toisinaan myös sitä, mihin suuntaan harhaisia estimaatit ovat, on hankala selvittää. Harha aiheutuu vastaamattomien ja vastanneiden välillä olevista systemaattisista eroista. Mitä suurempia erot näiden kahden ryhmän välillä ovat, sitä harhaisempia estimaatteja saadaan. Harhan suuruusluokkaan vaikuttaa lisäksi myös puuttuvien havaintojen määrä. On kuitenkin huomattava, että otoskoon lisääminen ei välttämättä korjaa harhaa.

3.2 Vastauskadon mallittaminen taustatiedoilla kalastuskyselyssä

Vastauskadon mallittamiseen tarvitaan sopivia taustamuuttujia, jotka ovat saatavilla myös vastaamattomista. Näiden muuttujien tulee luonnollisesti olla sellaisia, että vastaamisaste riippuu voimakkaasti apumuuttujista, jolloin aineisto voidaan esimerkiksi jälkiosittaa apumuuttujien mukaan ja muodostaa aineistoon homogeenisen vastausryhmän malli (Response Homogeneity Group -malli eli RHG-malli) (Särndal, Svensson & Wretman 1992), jossa oletetaan, että vastaamisaste on vakio näissä apumuuttujan tai apumuuttujien mukaan jaetuissa jälkiositteissa. Tällöin vastauskato voidaan ottaa huomioon painotuksessa kertomalla alkuperäinen painokerroin katokorjaustermillä, joka on ositteittain havaitun vastaamistodennäköisyyden käänteisluku.

Jos apumuuttujat puolestaan ovat jatkuvia, voidaan niiden ja vastaamisasteen välille sovittaa esimerkiksi regressiomalli.

Apumuuttujat saadaan usein tutkimuksen ulkopuolisesta lähteestä kuten rekistereistä. Muuttujista ei myöskään välttämättä tarvitse olla tiedossa muuta kuin reunajakaumat, jolloin voidaan käyttää kalibrointipainotusta, jota on kuvailtu tarkemmin luvussa 7.2. Tällöin voidaan myös yhdistää eri lähteistä saatua taustatietoa, sillä ristiintaulukon solufrekvenssejä ei tarvita.

3.3 Vastauskadon mallittaminen uusintakyselyllä kalastuskyselyssä

Aina sopivia apumuuttujia ei kuitenkaan ole saatavissa. Vastaamismekanismi voi esimerkiksi olla sen verran monimutkainen, etteivät rekistereistä saatavat apumuuttujat pysty selittämään sitä. Tällöin vastaamattomista täytyy yrittää hankkia tietoa muulla tavoin. Yksi keino tähän tiedonhankintaan on poimia alaotos vastaamattomista ja suorittaa heille uusi kysely. Tässä uudessa kyselyssä pyritään nyt saamaan vastaamattomista sellaista tietoa, mikä on olennaisinta tutkimuksessa ja mistä olisi hyötyä myös muiden tutkimusmuuttujien estimoinnissa.

Kysely täytyy suunnitella huolella, jotta vastaamisaste saataisiin pysymään mahdollisimman korkeana. Kysymyksien määrä kannattaa pitää mahdollisimman pienenä ja kysymysten tulisi olla helposti vastattavia. Helpoiten tällaisessa kyselyssä voidaan selvittää kvalitatiivista tietoa, esimerkiksi onko kotitaloudessa kalastettu lainkaan, jos on, niin millä pyydyksellä, jne. Sen sijaan kvantitatiivisen tiedon kuten saalismäärien ja kalastuspäivien selvittäminen on hankalampaa jo senkin takia, että tällainen tieto on vaikeasti muistettavaa ja arvioitavaa. Varsinkin jos puhutaan pitkästä ajanjaksosta, kuten vuodesta, ja kysely suoritetaan kauan kalastamisen jälkeen, muistivirheen osuus voi kasvaa todella suureksi tai sitten haastateltava kieltäytyy koko kyselystä vedoten muistamis- ja arviointivaikeuksiin.

Tällaiset vaikeasti vastattavat kysymykset voivat vielä lisätä kadon aiheuttamaa harhaa, sillä innokkaimmat ja ahkerimmat kalastajat todennäköisesti pystyvät arvioimaan saaliinsa ja kalastuspäiviensä määrän paremmin kuin vain satunnaisesti kalastavat. Tällöin vähän kalastavat kieltäytyvät myös helpommin kyselystä ja vastaamattomien ja vastanneiden välille syntyy systemaattisia eroja.

Kyselyssä selvitettävän kvalitatiivisen tiedon pitää kuitenkin korreloida voimakkaasti kvantitatiivisten tutkimusmuuttujien kanssa, jotta tästä lisätiedosta olisi hyötyä tutkimusmuuttujien estimoinnissa. Ainakin sen selvittäminen, onko kotitaloudessa kalastettu lainkaan ja jos on, niin millä pyydyksillä, näyttää olevan selvästi yhteydessä saalismääriin ja kalastuspäiviin. Luvussa 5.2 on kerrottu tarkemmin vuoden 2000 toteutetusta jälkikyselystä.

Vastaamattomista saadun kvalitatiivisen tiedon avulla voidaan aineistoon rakentaa jälkiositus. Jos muuttujien vaihtelu on jälkiositteiden sisällä pienempää kuin niiden välillä, saadaan tehokkaampia estimaattoreita. Jälkiositetusta aineistosta estimaatit keskiarvoille ja kokonaismäärille voidaan laskea estimoimalla halutut tunnusluvut ensin jälkiositteittain ja summaamalla ositteittain lasketut estimaatit sopivasti painottamalla.

Jälkiosituksen avulla voidaan myös muodostaa vastauskatoa korjaavat painot. Jos jälkiositus voidaan muodostaa siten, että osituksessa käytetty muuttuja korreloi vastaamisasteen kanssa, voidaan ositteittain estimoitujen vastaamisasteiden käänteisluvuista muodostaa katoa korjaavat painokertoimet. Nämä painot eivät kuitenkaan korjaa valikoituneesta vastauskadosta aiheutuvaa harhaa, joten ne eivät sellaisenaan yleensä riitä kalastuskyselyjen katoongelmien korjaamiseen. Seuraavassa luvussa on selitetty tarkemmin jälkiositusmenetelmän käyttöä.

4 Hierarkkisen jälkiosituksen soveltaminen vastauskadon paikkaamiseen

Jos sopivaa taustatietoa ei ole saatavilla, ei kadon korjaamiseen voida käyttää rekisteritiedon avulla tehtävään mallintamiseen perustuvia tekniikoita, vaan estimoinnin tehostamiseksi joudutaan etsimään muita keinoja. Yksi mahdollisuus vähentää vastauskadon aiheuttamaa harhaa on jälkiositus. Tämä tarkoittaa aineiston jakamista otannan jälkeen jonkin sopivan otokseen sisältyvän muuttujan mukaan mahdollisimman homogeenisiin osiin siten, että myös tutkittavan tulosmuuttujan vaihtelu on pienempää näiden ositteiden sisällä kuin niiden välillä.

Jälkiositetusta aineistosta voidaan laskea haluttuja suureita, kuten keskiarvoja ja variansseja, estimoimalla halutut tunnusluvut ensin jälkiositteiden sisällä ja sitten yhdistämällä näistä estimaatit koko aineistoa koskeville tunnusluvuille. Luvussa 4.3 on esitetty keskiarvon, varianssin ja kovarianssin kaavat, joilla pystytään arvioimaan kyseisiä suureita koko aineistossa jälkiositteista laskettujen tunnuslukujen avulla.

Luvussa 5.3 on esitelty empiirisiä tarkasteluja jälkiosituksen käytölle Vapaa-ajankalastus 1998 -aineistossa. Näissä kokeiluissa aineisto on ensin jälkiositettu kalastaneisiin ja kalastamattomiin. Lisäksi kalastaneet on vielä jaettu pienempiin ositteisiin pyydysten käytön mukaan. Empiirisissä tarkasteluissa nyt sovellettu jälkiositus on vain yksi mahdollisuus osittaa aineistoa ja myös muita jälkiosituskriteerejä kannattaa pohtia. Muita mahdollisia kriteerejä kalastaneiden jakamiseen pienempiin ositteisiin voisivat olla esimerkiksi kalastuspäivien tai saaliin määrä (vähän ja paljon kalastaneet omiksi ryhmikseen).

Tässä raportissa esitetyissä empiirisissä tarkasteluissa on laskujen ja merkintöjen yksinkertaistamiseksi käsitelty vain hierarkkiseen ositukseen johtavia binäärisiä jakoja. Vastaavasti kaikki esitetyt estimointikaavat yleistyvät kuitenkin myös jälkiosituksiin, joissa aineisto jaetaan kolmeen tai useampaan luokkaan.

4.1 Populaation ositusmalli

Halutaan estimoida vapaa-ajankalastusaineiston tietyn muuttujan kokonaismäärä tai keskiarvo (esimerkiksi saaliin määrä tai kalastuspäivien lukumäärä). Merkitään tätä kiinnostuksen kohteena olevaa muuttujaa y:llä.

Havaitaan, että aineistossa voidaan ajatella olevan taustalla hierarkkinen malli, sillä y:n arvot riippuvat toisesta muuttujasta k, joka kertoo, onko kotitalous kalastanut kuluneen vuoden aikana vai ei. Lisäksi, jos aineistossa esiintyy vastauskatoa, myös sillä on vaikutusta tutkittavan muuttujan y estimointiin, sillä vastaamattomista ei ole saatavilla tietoa tutkittavasta muuttujasta y, eikä edes siitä, onko kotitalous kalastanut vai ei. Merkitään jatkossa tätä vastauskatomuuttujaa i:llä.

Siis

$$k = \begin{cases} 0 & \text{, kun kotitalous ei kalastanut} \\ 1 & \text{, kun kotitalous kalasti} \end{cases}$$

ja

$$i = \begin{cases} 0 & \text{, kun kotitalous ei vastannut kyselyyn} \\ 1 & \text{, kun kotitalous vastasi kyselyyn.} \end{cases}$$

Populaatio voidaan tällöin luokitella muuttujien k ja i avulla ristiintaulukkoon ja konstruoida jakauma seuraavasti.

	$i=0$	$i=1$	
$k=0$	π_{00}	π_{01}	$\pi_{0.}$
$k=1$	π_{10}	π_{11}	$\pi_{1.}$
	$\pi_{.0}$	$\pi_{.1}$	

Tällöin siis muuttujan i reunajakaumista $\pi_{.0}$ ja $\pi_{.1}$ nähdään vastaamattomien ja vastanneiden osuudet koko aineistossa ja muuttujan k reunajakaumat $\pi_{0.}$ ja $\pi_{1.}$ kertovat kalastaneiden ja kalastamattomien osuudet aineistossa, jossa vastauskatoa ei esiinny.

Ristiintaulukon frekvensseistä π_{01} ja π_{11} saadaan estimoiduksi aineistosta, mutta kalastaneiden ja kalastamattomien osuudet vastaamattomien joukossa ovat tuntemattomia. Myös osuudet vastaamattomilla pitäisi kuitenkin pystyä estimoimaan, jotta päästäisiin käsiksi "täydelliseen aineistoon", joka edustaa riittävän hyvin tavoiteperusjoukkoa.

Jotta vastaamattomuus olisi harmitonta (epäinformatiivista) eli se ei riippuisi lainkaan kalastamisesta, vaaditaan, että vastanneiden ja vastaamattomien suhde kalastaneilla ja kalastamattomilla kotitalouksilla olisivat samat,

ts. $\frac{\pi_{01}}{\pi_{00}} = \frac{\pi_{11}}{\pi_{10}}$. Tällöin siis ristitulosuhde

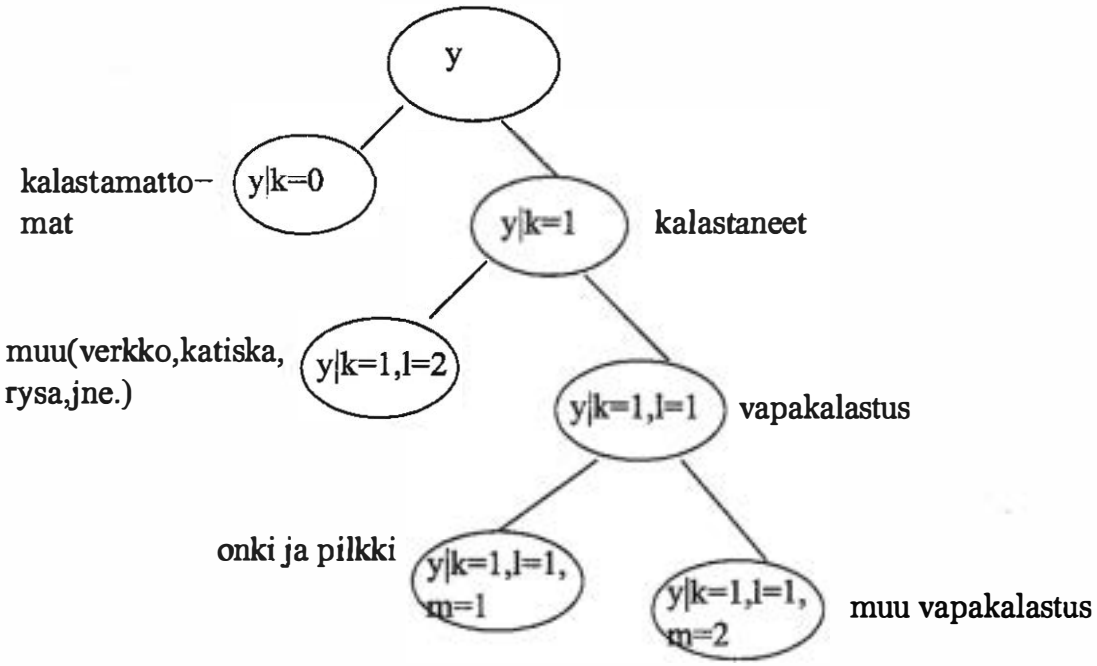
$$\Omega = \frac{\frac{\pi_{01}}{\pi_{00}}}{\frac{\pi_{11}}{\pi_{10}}} = \frac{\pi_{01}\pi_{10}}{\pi_{11}\pi_{00}} = 1$$

ja yhtäpitävästi $\log \Omega = 0$. Tämä vastaa siis riippumattomuusmallia muuttujien i ja k kontingenssitaulussa.

Aineistosta pystytään estimoimaan kalastaneiden ja kalastamattomien osuudet kyselyyn vastanneiden osalta (π_{01} ja π_{11}). Vastaavat osuudet vastaamatta jättäneiden kotitalouksien osalta eivät puolestaan ole tiedossa, mutta ne voidaan estimoida esimerkiksi poimimalla otos kyselyyn vastaamatta jättäneistä ja suorittamalla otokseen valituksi tulleille puhelimitse jälkikysely, jossa kysytään, onko kotitalous kalastanut vai ei. Tämän puhelinhaastattelun tulosten perusteella saadaan estimoiduksi kaikki ristiintaulukon solufrekvenssit ja pystytään sovittamaan aineistoon loglineaarinen malli.

4.2 Hierarkkinen ositus

Binääristä puuta hyväksi käyttäen aineisto voidaan jaotella hierarkkisesti pienempiin osiin jakamalla se ensin kahtia, minkä jälkeen nämä ositteet voidaan edelleen jakaa kahtia jne. Vapaa-ajankalastusaineistoon tätä hierarkkisen osituksen ideaa voidaan soveltaa esimerkiksi jakamalla aineisto ensin kahtia kalastamisen mukaan (kalastaneet ja kalastamattomat kotitaloudet omiin jälkiositteisiinsa). Tämän jälkeen kalastaneet voidaan vielä jakaa kahtia esimerkiksi käytettyjen pyydysten suhteen vaikkapa pelkästään vapakalastusta harrastaneisiin ja myös muilla pyydyksillä kalastaneisiin. Kolmas taso hierarkkiaan saadaan, kun vapakalastus jaetaan vielä kahteen osaan: onki- ja pilkkikalastukseen sekä muuhun vapakalastukseen. Tätä esimerkkijakoa on havainnollistettu kuvassa 1.



Kuva 1: Aineiston strukturointi pyydysten käytön mukaan

Otamme siis käyttöön uudet binääriset muuttujat l ja m seuraavasti:

$$l = \begin{cases} 1 & \text{, kun on harrastettu vain vapakalastusta} \\ 2 & \text{, kun on käytetty (myös) muita pyydyksiä} \end{cases}$$

ja

$$m = \begin{cases} 1 & \text{, kun on kalastettu vain ongella tai pilkillä} \\ 2 & \text{, kun on harrastettu (myös) muuta vapakalastusta.} \end{cases}$$

Lisäksi merkitään p:llä todennäköisyyttä $\mathrm{P}(k = 1)$, q:lla todennäköisyyttä $\mathrm{P}(l = 1)$ ja r:llä todennäköisyyttä $\mathrm{P}(m = 1)$.

Perusajatuksena on odotus, että hierarkkinen jako luo homogeenisia osajoukkoja. Homogeenisuudesta seuraa, että niissä oleva puuttuva tieto on imputoitavissa tavanomaisilla menetelmillä kuten keskiarvoimputoinnilla. Sen sijaan jakosuhteet on saatavissa ulkopuolisesta lähteestä kuten jälkikyselystä. On huomattava, että tässä nojaudutaan apumuuttujainformaatioon vain jakosuhteiden osalta.

4.3 Keskiarvo, varianssi ja kovarianssi binäärisen jaon tilanteessa

Oletetaan aluksi, että aineisto on jaettu vain kalastaneisiin ja kalastamattomiin. Oletetaan lisäksi, että aineistossa ei esiinny vastauskatoa. Tällöin tutkittavan muuttujan y (esimerkiksi saaliin määrä, kalastuspäivät, jne.) vaihtelu muodostuu muuttujan oman vaihtelun lisäksi siitä, kalastiko kotitalous vai ei.

Jos oletetaan, että vastauskatoa ei ole (käytössä olisi täydellinen aineisto), voidaan muuttujan y odotusarvo ja varianssi laskea seuraavasti:

$$\mathrm{E}y = \mathrm{E}_k\mathrm{E}(y|k) = p\mathrm{E}(y|k=1) + (1-p)\mathrm{E}(y|k=0).$$

Tässä $\mathrm{E}(y|k=0) = 0$, joten lauseke sievenee muotoon

$$\mathrm{E}y = p\mathrm{E}(y|k=1).$$

Koko aineiston odotusarvolle saadaan siis estimaatti kertomalle kalastaneiden keskisaalis tai keskimääräiset kalastuspäivät kalastaneiden osuudella. Odotusarvo kalastaneiden joukossa voidaan estimoida otoksesta kuten myös kalastaneiden osuus. Kalastaneiden osuudelle saadaan kuitenkin todennäköisesti parempi arvio selvittämällä kalastaneiden osuus vastaamatta jättäneiden joukossa esimerkiksi jälkikyselyllä ja arvioimalla kalastaneiden osuutta koko aineistossa ottamalla huomioon vastauskadon suuruus ja vastaamattomien joukossa kalastaneiden määrä.

Vastaavasti

$$\begin{aligned}
\mathrm{Var}(y) &= \mathrm{E}_k\mathrm{Var}(y|k) + \mathrm{Var}_k\mathrm{E}(y|k) \\
&= \mathrm{E}(\mathrm{E}(y^2|k) - (\mathrm{E}(y|k))^2) + \mathrm{E}(\mathrm{E}(y|k))^2 - (\mathrm{E}(y|k))^2 \\
&= p\mathrm{E}(y^2|k=1) - p(\mathrm{E}(y|k=1))^2 + \mathrm{E}(\mathrm{E}(y|k))^2 - \mathrm{E}(\mathrm{E}(y|k))^2 \\
&= p\mathrm{Var}(y|k=1) + p(1-p)\mathrm{E}(y|k=1)^2,
\end{aligned}$$

missä summan ensimmäinen termi tulee siis muuttujan y vaihtelusta kalastaneiden joukossa ja toinen termi kalastaneiden osuuden $p = \mathrm{P}(k=1)$ vaihtelusta.

Varianssi koko aineistossa saadaan siis kertomalla varianssi kalastaneiden joukossa kalastaneiden osuudella ja lisäämällä tähän termi, joka saadaan

kertomalla kalastaneiden osuus, kalastamattomien osuus ja kalastaneiden odotusarvon neliö keskenään. Kuten edellä odotusarvoa laskettaessa ehdollinen odotusarvo ja varianssi kalastaneiden joukossa voidaan estimoida aineistosta ja kalastaneiden osuus saadaan arvioiduksi esimerkiksi jälkikyselyn perusteella.

Jos merkitsemme kahta tulosmuuttujaa (esimerkiksi saalis ja kalastuspäivät) x:llä ja y:llä, voidaan myös näiden kahden muuttujan välinen kovarianssi kirjoittaa muodossa, jossa kovarianssi koko aineistossa on jaettu termeiksi. Nämä termit voidaan estimoida kalastaneiden ja kalastamattomien osajoukoissa.

Koska kovarianssi voidaan esittää muodossa

$$\begin{aligned}\mathrm{Cov}(x,y) &= \mathrm{E}[(x-\mathrm{E}x)(y-\mathrm{E}y)] \\ &= \mathrm{E}_k\mathrm{E}[(x-\mathrm{E}x)(y-\mathrm{E}y)|k] \\ &= \mathrm{E}_k\mathrm{Cov}(x,y) - \mathrm{E}_k[\mathrm{E}(x|k)\mathrm{E}(y|k)|k] + \mathrm{E}_k(\mathrm{E}x\mathrm{E}y)|k) \\ &= \mathrm{E}_k\mathrm{Cov}(x,y|k) - \mathrm{Cov}_k[\mathrm{E}(x|k),\mathrm{E}(y|k)],\end{aligned}$$

saadaan muuttujien x ja y välinen kovarianssi lopulta muotoon

$$\begin{aligned}\mathrm{Cov}(x,y) &= p\mathrm{Cov}(x,y|k=1) + (1-p)\mathrm{Cov}(x,y|k=0) \\ &\quad - p(\mathrm{E}(x|k=1) - \mathrm{E}x)(\mathrm{E}(y|k=1) - \mathrm{E}y) \\ &\quad - (1-p)(\mathrm{E}(x|k=0) - \mathrm{E}x)(\mathrm{E}(y|k=0) - \mathrm{E}y) \\ &= p\mathrm{Cov}(x,y|k=1) - p[\mathrm{E}(x|k=1) - \mathrm{E}x][\mathrm{E}(y|k=1) - \mathrm{E}y].\end{aligned}$$

Muuttujan y ehdollinen odotusarvo ja varianssi sekä muuttujien x ja y välinen ehdollinen kovarianssi $\mathrm{E}(y|k=1)$, $\mathrm{Var}(y|k=1)$ ja $\mathrm{Cov}(x,y|k=1)$ voidaan estimoida otoksesta. Estimaatit näille saadaan laskemalla kyseiset tunnusluvut kalastaneiden osalta otokseen kuuluvilta kotitalouksilta. Ongelmana on kuitenkin kalastaneiden osuuden $\mathrm{P}(k=1)$ arviointi. Kalastaneiden osuus on nimittäin todennäköisesti pienempi vastaamattomilla kuin vastanneilla, joten jos todennäköisyyttä $\mathrm{P}(k=1)$ estimoidaan aineistosta, saadaan luultavasti liian suuri arvio tälle osuudelle.

Edellä esitettyjä keskiarvo-, varianssi- ja kovarianssikaavoja käyttäen saadaan kyseiset tunnusluvut lasketuksi myös monimutkaisemman hierarkkisen

mallin tapauksessa. Lasketaan vain ensin halutut tunnusluvut pienimmissä ositteissa ja yleistetään kaavoja rekursiivisesti aina ylemmälle ja ylemmälle tasolle, kunnes saadaan koko aineistoa koskevat tunnusluvut.

4.4 Estimaattorin keskivirheen estimointi

Jälkiositettujen estimaattien varianssien estimointi on kuitenkin vaikeaa, jos ositteisiin kuulumistodennäköisyyksiä ei voida olettaa kiinteiksi. Lisäksi näiden varianssien estimoinnissa täytyy ottaa myös huomioon ositteittain laskettujen estimaattien kovarianssit, koska ositteittain laskettuja estimaatteja ei yleensä voida olettaa riippumattomiksi.

Tarkastellaan merkintöjen yksinkertaistamiseksi kahteen osaan jaettua aineistoa, jossa p on todennäköisyys kuulua ensimmäiseen jälkiositteeseen. Tulokset voidaan suoraviivaisesti yleistää tapaukseen, jossa jälkiositteita on useampia tai jälkiositteisiin jako on tehty hierarkkisesti. Oletetaan ensin, että ositteisiin kuulumistodennäköisyydet voidaan olettaa kiinteiksi, jolloin niistä ei tule lisää vaihtelua estimaatin vaihteluun. Tällöin muuttujan y keskiarvon varianssi voidaan jakaa ositteittain laskettujen varianssien painotetuksi summaksi, jossa painoina käytetään ositteisiin kuulumistodennäköisyyksien neliöitä:

$$\widehat{\mathrm{Var}}(\overline{y}) = p^2\mathrm{Var}(\overline{y}_1) + (1-p)^2\mathrm{Var}(\overline{y}_2) + p(1-p)\mathrm{Cov}(\overline{y}_1, \overline{y}_2).$$

Tässä $\overline{y}_1$ on siis keskiarvoestimaattori ensimmäisessä jälkiositteessa ja $\overline{y}_2$ toisessa. Kovarianssitermin laskeminen on yleensä hankalaa. Useissa tapauksissa estimaatin varianssin suuruusluokalle saadaan kuitenkin riittävän hyvä arvio ottamalla mukaan pelkät varianssitermit ja jättämällä kovarianssit pois.

Kun myös p oletetaan satunnaismuuttujaksi, ei varianssille saada tarkkaa lauseketta, vaan se joudutaan approksimoimaan. Tähän voidaan käyttää Taylorin sarjakehitelmään perustuvaa delta-menetelmää (Wolter 1985). Merkitään keskiarvoa $g(p, \overline{y}_1, \overline{y}_2) = \overline{y} = p\overline{y}_1 + (1-p)\overline{y}_2$. Tällöin $\frac{\partial g}{\partial p} = \overline{y}_1 - \overline{y}_2$, $\frac{\partial g}{\partial \overline{y}_1} = p$ ja $\frac{\partial g}{\partial \overline{y}_2} = 1-p$. Delta-menetelmän mukaan funktion g varianssi voidaan kirjoittaa nyt muodossa

$$\mathrm{Var}(g(p, \overline{y}_1, \overline{y}_2)) \approx [Dg(p, \overline{y}_1, \overline{y}_2)]^T \mathrm{Cov}(p, \overline{y}_1, \overline{y}_2)[Dg(p, \overline{y}_1, \overline{y}_2)].$$

Keskiarvoestimaattorin varianssille saadaan siis seuraava estimaattori:

$$\hat{\mathrm{Var}}(\overline{y}) = p^2\mathrm{Var}(\overline{y}_1) + (1-p)^2\mathrm{Var}(\overline{y}_2) + \overline{y}_1^2\mathrm{Var}(p) + \overline{y}_2^2\mathrm{Var}(p) - 2\overline{y}_1\overline{y}_2\mathrm{Var}(p) + p(1-p)\mathrm{Cov}(\overline{y}_1, \overline{y}_2).$$

Koska p:n ja y:n estimaatit tulevat riippumattomista lähteistä (esimerkiksi y:n ositteittaiset keskiarvot otoksesta ja p jälkikyselystä), voidaan p ja y olettaa riippumattomiksi, joten näiden välisiä kovarianssitermejä ei tarvita.

5 Empiirisiä tarkasteluja

5.1 Vastaamattomien arviointi kontaktiryhmien avulla ja RHG-mallin soveltaminen aineistoon

Kun tarkastellaan eri kontaktiryhmiä (monennellako kontaktilla kotitalous on vastannut kyselyyn), havaitaan sekä kalastaneiden osuuden että keskimääräisten kalastuspäivien ja saalismäärän välillä eroja. Ensimmäisellä kontaktilla vastanneiden joukossa kalastaneita on suhteellisesti enemmän kuin toisella ja kolmannella kontaktilla vastanneissa. Kalastaneiden osuus putoaa vielä hieman siirryttäessä toisesta kontaktiryhmästä kolmanteen. Myös kalastuspäivien ja saalismäärän suhteen on nähtävissä samansuuntaisia eroja ryhmien välillä. Seuraavaan taulukkoon on koottu joitakin tunnuslukuja estimoituna erikseen eri kontaktiryhmissä.

Taulukko 2: Kontaktiryhmittäin laskettuja tunnuslukuja

	1. kontakti-ryhmä	2. kontakti-ryhmä	3. kontakti-ryhmä
vastaamisaste	36.6 %	12.1 %	32.1 %
kalastaneiden kotitalouksien osuus	61 %	50 %	46 %
kalastaneiden kotitalouksien keskisaalis (kg)	65.65	41.34	42.08
kalastaneiden kotitalouksien keskimääräiset kalastuspäivät	45.19	35.86	30.53

Tässä vapaa-ajankalastusaineistossa RHG-mallia sovellettiin siten, että ensimmäisellä ja toisella kontaktilla kyselyyn vastanneet kotitaloudet muodostivat oman vastausaktiivisuusryhmänsä, jonka sisällä vastaamistodennäköisyys oletettiin vakioksi. Toinen kontaktiryhmä sai postissa vain vastauskehoituksen ja monet tämän kehoituksen saaneista olivat ehtineet jo heittää kyselylomakkeen pois, joten heillä ei ollut mahdollisuutta vastata. Tämä selittää osaltaan toisen ryhmän muita pienempää vastaamisastetta. Näin ollen ensimmäisen ja toisen ryhmän yhdistäminen oli vastausaktiivisuusryhmien vertailukelpoisuuden vuoksi järkevää.

Toinen vastusaktiivisuusryhmä muodostui kolmannella kontaktilla vastanneista ja vastaamattomista. Vastaamattomien oletettiin siis olevan samanlaisia kuin kolmanteen kontaktiryhmään kuuluvat. Tämän mallin käytön epävarmuutta lisää kuitenkin se, että vastauskato kolmannessa kontaktiryhmässä on suurta (vastanneita on vain n. 32 %), ja on syytä epäillä, että siitä lasketut estimaatit eivät välttämättä vastaa kovinkaan hyvin vastaamattomia.

Seuraavassa taulukossa on esitetty joitakin aineistosta laskettuja tunnuslukuja sekä alkuperäistä käytettyä painotusta käyttäen että RHG-mallia soveltaen. RHG-mallin antamat tulokset ovat selvästi alkuperäisiä pienempiä niin kalastaneiden osuuden kuin kalastuspäivien ja saalismäärän keskiarvojenkin osalta.

Taulukko 3: Tunnuslukuja estimoituna kalibrointipainotusta käyttäen ja RHG-mallia soveltaen

	aineistossa käytetty yhdistettyä kalibrointi-katopainotusta	estimaatit laskettu soveltaen RHG-mallia
kalastaneiden kotitalouksien osuus (%)	46	44
kalastaneiden kotitalouksien keskisaalis (kg)	45	39
kalastaneiden kotitalouksien keskimääräiset kalastuspäivät	38	34

5.2 Aliotosmenetelmän soveltaminen Vapaa-ajankalastus 2000 -aineistoon

Vuoden 2000 Vapaa-ajankalastuskyselyssä kokeiltiin jälkikyselytekniikkaa yhdelle ositteelle. Tämän ositteen vastaamattomille suoritetussa jälkiotannassa poimittiin 300 kotitalouden otos n. 400 vastaamattoman joukosta. Näistä 300 otokseen valituksi tulleesta löydettiin puhelinnumerot 277 kotitaloudelle, joista 233 vastasi puhelintiedusteluun. Hyväksyttyjä vastauksia kysymykseen, kalastettiinko kotitaloudessa kyseisen vuoden aikana tuli 217. Vastanneista kotitalouksista 106 (48.8 %) oli kalastaneita ja 111 (51.2 %) kalastamattomia.

Alkuperäinen otoskoko tutkittavassa ositteessa oli 1250, joista vastanneita oli 838. Kalastaneiden osuus ensimmäisessä kontaktiryhmässä oli 58.2 %, toisessa 58 % ja kolmannessa 48.6 %.

Kalastaneiden osuus vastaamattomista poimitussa aliotoksessa on siis lähes sama kuin kolmannen kontrolliryhmän vastaava osuus. Myös vuoden 1998 kyselyssä kolmannesta kontaktiryhmästä laskettu kalastaneiden osuus (46 %) on hyvin lähellä näitä. Kalastaneiden osuus vastaamattomien joukossa on siis todennäköisesti lähellä vastaavaa osuutta kolmannessa kontaktiryhmässä, eli RHG-mallin antamat estimointitulokset ovat luultavasti lähellä oikeita.

5.3 Jälkiositusmenetelmän soveltaminen Vapaa-ajankalastus 1998 -aineistoon

Jälkiositusmenetelmää sovelletaan tähän tutkimukseen siten, että aineisto jaettiin ensin kalastaneisiin ja kalastamattomiin. Tämän jälkeen kalastaneet jaettiin pyydysten käytön mukaan ensin kahteen ryhmään: vain vapakalastusta harrastaneisiin ja muihin, ja sen jälkeen vapakalastaneet jaettiin vielä pelkästään onkineisiin ja pilkkineisiin sekä muihin. Tämä jako on kuvattu luvussa 4.2 ja käytämme jatkossa samoja siellä esiteltyjä merkintöjä.

Tällainen jako perustuu nyt siihen, että koska aineistossa ei ole sopivia estimoinnissa hyödynnettävää lisätietoa sisältäviä muuttujia, aineiston ositus rakennetaan jonkin sellaisen muuttujan avulla, joka voidaan selvittää aliotosmenetelmällä myös vastaamattomista. Sekä kalastaminen että pyydysten käyttö ovat molemmat asioita, jotka on helppo selvittää jälkikäteen yksinkertaisilla kyllä/ei-kysymyksillä vastaamattomista poimitusta aliotoksesta. Sen sijaan tutkittavat muuttujat, kuten saalismäärät tai kalastuspäivät ovat vaikeita selvittää vastaamattomilta, koska ne ovat kvantitatiivista tietoa. Esimerkiksi jos aliotos poimitaan paljon alkuperäisen otoksen poimimisen jälkeen, otokseen valituksitulleilla on varmasti myös vaikeuksia muistaa saalismääriä tai kalastuspäiviä tarkasti. Niiden arvioiminen ei ole muutenkaan helppoa, vaikkei kalastustapahtumasta edes olisi kulunut kovin kauaa. Näin ollen vastaamattomista pyritään saamaan sellaista kvalitatiivista tietoa, jonka avulla näiden kvantitatiivisten muuttujien estimointia voitaisiin parantaa.

Kalastaneiden osuus vaikuttaa selvästi kalastuspäivien ja saalismäärän estimointiin, sillä kalastamattomilla ei ole saalista eikä kalastuspäiviä. Myös pyydyksellä näyttää olevan selvää yhteyttä kalastuspäiviin ja saaliiseen.

Seuraavissa taulukoissa on aineistosta estimoidut kalastaneiden kotitalouksien kalastuspäivien ja saalismäärän keskiarvot, varianssit ja variaatiokertoimet pyydysten käytön mukaan jaetuissa ryhmissä. Kaksi ensimmäistä taulukkoa esittävät edellä mainitut tunnusluvut vapakalastusta harrastaneille ja muita pyydyksiä käyttäneille. Kahdessa seuraavassa taulukossa puolestaan on samat tunnusluvut laskettuna erikseen vain onkineille ja pilkkineille sekä myös muuta vapakalastusta harrastaneille.

Taulukko 4: Kalastuspäivien keskiarvo, varianssi ja variaatiokerroin eri pyydysryhmissä, l=1 vain vapakalastusta harrastaneilla ja l=2 myös muita pyydyksiä käyttäneillä

y=kalastuspäivät				
	N	E(y|k=1,l)	Var(y|k=1,l)	c.v.(y|k=1,l)
l=1	545	15.91	1542.52	2.47
l=2	833	56.79	5989.02	1.36

Kalastuspäivien määrä on paljon pienempi vain vapakalastusta harrastaneilla kuin kotitalouksilla, joissa on käytetty muita pyydyksiä (esim. verkko, katiska tai rysä). Pyydysten käytön suhteen kalastuspäiväksi lasketaan pyydysten tarkastuspäivät. Kalastuspäivät kuitenkin myös vaihtelevat enemmän ensimmäisessä ryhmässä. Pelkästään vapakalastusta harrastaneita oli 545 kotitaloutta ja muita pyydyksiä käyttäneitä 833.

Taulukko 5: Saalismäärän keskiarvo, varianssi ja variaatiokerroin eri pyydysryhmissä, l=1 vain vapakalastusta harrastaneilla ja l=2 myös muita pyydyksiä käyttäneillä

y=saalismäärä				
	N	E(y|k=1,l)	Var(y|k=1,l)	c.v.(y|k=1,l)
l=1	545	18.05	11022.08	5.82
l=2	833	82.90	27270.73	1.99

Saalismäärä jää myös selvästi pienemmäksi vapakalastusryhmässä. Saalismäärän vaihtelu on kuitenkin selvästi suurempaa tässä ensimmäisessä ryhmässä.

Taulukko 6: Kalastuspäivien keskiarvo, varianssi ja variaatiokerroin eri pyydysryhmissä, m=1 vain onkineilla ja pilkkineillä ja m=2 myös muuta vapakalastusta harrastaneilla

y=kalastuspäivät				
	N	E(y|k=1,l=1,m)	Var(y|k=1,l=1,m)	c.v.(y|k=1,l=1,m)
m=1	235	9.79	2235.72	4.83
m=2	310	20.55	972.49	1.52

Kun vapakalastusta harrastaneet jaetaan vielä kahteen ryhmään, joista ensimmäiseen kuuluvat vain ongella ja pilkillä kalastaneet ja toiseen (myös) muuta vapakalastusta (heittovapa, vetouistin tai perhovapa) harrastaneet, on ensimmäiseen ryhmään kuuluvilla kalastuspäivien keskiarvo selvästi pienempi, mutta vaihtelu on toista ryhmää suurempaa.

Taulukko 7: Saalismäärän keskiarvo, varianssi ja variaatiokerroin eri pyydysryhmissä, m=1 vain onkineilla ja pilkkineillä ja m=2 myös muuta vapakalastusta harrastaneilla

y=saalismäärä				
	N	E(y|k=1,l=1,m)	Var(y|k=1,l=1,m)	c.v.(y|k=1,l=1,m)
m=1	235	14.36	21150.00	10.12
m=2	310	20.85	3369.89	2.78

Samoin saalismäärän suhteen ryhmiä verratessa ensimmäisessä ryhmässä saalista saadaan vähemmän, mutta myös vaihtelu on selvästi suurempaa kuin toisessa ryhmässä.

Sovelletaan seuraavaksi luvussa 4.3 kehitettyä menetelmää saalismäärän ja kalastuspäivien keskiarvon estimointiin, eli arvioidaan näitä keskiarvoja koko aineistossa jälkiositteittain laskettujen estimaattien avulla. Ehdolliset odotusarvot pystytään estimoimaan jälkiositteiden sisällä, mutta ongelmana on osuuksien p, q ja r estimointi. Näillekin voidaan toki laskea estimaatit aineistosta, mutta kuten aiemmin on todettu, voi varsinkin kalastaneiden osuuden p estimointi otoksesta antaa melko harhaisia tuloksia, sillä vastanneiden ja vastaamattomien välillä kalastusaktiivisuuden voidaan olettaa poikkeavan

toisistaan selvästi. Seuraavissa taulukoissa on esitetty estimaatit kalastuspäivien ja saalismäärän keskiarvolle muutamilla eri p:n arvoilla. Osuuksiksi r ja q on näissä tarkasteluissa valittu otoksesta lasketut estimaatit 0.40 ja 0.43.

Taulukko 8: Kalastuspäivien keskiarvoja estimoituna eri p:n arvoilla

	y=kalastuspäivät
p	E(y)
0.46	18.60
0.49	19.82
0.52	21.03
0.53	21.43
0.55	22.24

Taulukko 9: Saalismäärän keskiarvoja estimoituna eri p:n arvoilla

	y=saalismäärä
p	E(y)
0.46	26.20
0.49	27.91
0.52	29.62
0.53	30.19
0.55	31.33

Kalastaneiden osuuden arvioinnilla on siis suuri vaikutus tutkittavien muuttujien odotusarvojen estimoinnissa, joten tämä osuus täytyisi saada mahdollisimman hyvin estimoiduksi. Jos kalastaneiden osuus vastaamattomien joukossa saadaan selvitetyksi esimerkiksi jälkikyselyn avulla, saadaan vastanneiden ja vastaamattomien osuudet huomioon ottamalla kalastaneiden osuudelle koko aineistossa hyvä arvio. Seuraavassa taulukossa on esitetty aineistosta estimoidut kalastaneiden osuudet, vain vapakalastusta harrastaneiden osuudet sekä vain onkineiden ja pilkkineiden osuudet lääneittäin sekä koko maan tasolla. Nämä estimaatit on laskettu olettaen vastaamattomat samanlaisiksi kuin kolmannella kontaktilla vastanneet.

Taulukko 10: Osuudet p, q ja r lääneittäin sekä koko Suomessa aineistosta estimoituna käyttäen kalibrointipainotusta.

	kalastaneiden osuus p	vapakalastaneiden osuus q	onkineiden ja pilkkineiden osuus r
Etelä-Suomi	0.42	0.45	0.53
Länsi-Suomi	0.43	0.52	0.46
Itä-Suomi	0.50	0.41	0.60
Oulu	0.54	0.34	0.36
Lappi	0.63	0.36	0.37
Ahvenanmaa	0.45	0.41	0.04
Koko maa	0.45	0.45	0.49

Seuraaviin taulukoihin on koottu saalismäärän ja kalastuspäivien keskiarvoestimaatit laskettuna sekä kalibrointipainotusta käyttäen että jälkiositusta hyödyntäen. Jälkiositetut tulokset on laskettu käyttäen edellä esitetyllä tavalla estimoituja osuuksia. Kalastaneiden osuus p, vain vapakalastaneiden osuus q ja vain onkineiden ja pilkkineiden osuus r estimoitiin alkuperäisestä otoksesta olettaen, että vastaamattomat käyttäytyvät samoin kuin kolmas kontaktiryhmä. Samaa painotusmenetelmää käyttäen estimoitiin myös keskiarvot jälkiositteiden sisällä. Estimaatit on laskettu sekä lääneittäin että koko maan tasolla.

Taulukko 11: Kalastaneiden kotitalouksien keskisaaliit lääneittäin sekä koko Suomessa laskettuna kalibrointipainotusta käyttäen sekä jälkiositusta hyödyntäen.

	kalibrointi-painotuksella lasketut estimaatit	jälkiositetut estimaatit
	keskiarvo	keskiarvo
Etelä-Suomi	33.87	32.56
Länsi-Suomi	49.93	49.18
Itä-Suomi	57.74	51.75
Oulu	49.56	49.55
Lappi	57.88	57.01
Ahvenanmaa	71.47	58.65
Koko maa	44.97	43.62

Eri menetelmillä saatuja tuloksia verrattaessa havaitaan, että jälkiositus antaa saaliin keskiarvolle jonkin verran pienempiä estimaatteja kuin alkuperäinen kalibrointipainotus. Erot näiden estimaattien välillä eivät kuitenkaan ole kovin suuria. Suurimmat erot esiintyvät Itä-Suomen läänin ja Ahvenanmaan kohdalla. Koko maan keskisaaliissa ero on kolmen prosentin luokkaa.

Taulukko 12: Kalastaneiden kotitalouksien keskimääräiset kalastuspäivät lääneittäin sekä koko Suomessa laskettuna kalibrointipainotusta käyttäen sekä jälkiositusta hyödyntäen.

	kalibrointi-painotuksella lasketut estimaatit	jälkiositetut estimaatit
	keskiarvo	keskiarvo
Etelä-Suomi	32.92	31.78
Länsi-Suomi	34.71	34.26
Itä-Suomi	51.41	45.69
Oulu	53.27	53.51
Lappi	47.74	46.80
Ahvenanmaa	19.13	17.05
Koko maa	38.30	37.22

Kalastuspäivien kohdalla tulokset ovat samansuuntaisia kuin saalismäärälläkin. Jälkiositetut estimaatit ovat jälleen hieman alkuperäisiä pienempiä, mutta erot ovat pieniä. Suurin ero on jälleen Itä-Suomen läänin kohdalla.

Seuraavissa kahdessa taulukossa on esitetty vastaavilla tavoilla estimoidut kokonaismäärien estimaatit. Tulokset ovat samansuuntaisia kuin keskiarvojenkin estimoinnissa. Jälkiositus antaa siis hiukan pienempiä arvoja myös kokonaismäärille.

Taulukko 13: Kokonaissaaliit lääneittäin sekä koko Suomessa laskettuna kalibrointipainotusta käyttäen sekä jälkiositusta hyödyntäen.

	kalibrointi-painotuksella lasketut estimaatit	jälkiositetut estimaatit
	kokonais-määrä (milj. kg)	kokonais-määrä (milj. kg)
Etelä-Suomi	14.505	13.942
Länsi-Suomi	17.739	17.472
Itä-Suomi	8.177	7.329
Oulu	4.654	4.653
Lappi	2.658	2.618
Ahvenanmaa	0.422	0.346
Koko maa	48.154	46.703

Taulukko 14: Kalastuspäivien kokonaismäärät lääneittäin sekä koko Suomessa laskettuna kalibrointipainotusta käyttäen sekä jälkiositusta hyödyntäen.

	kalibrointi-painotuksella lasketut estimaatit	jälkiositetut estimaatit
	kokonais-määrä (milj. pv.)	kokonais-määrä (milj. pv.)
Etelä-Suomi	14.095	13.607
Länsi-Suomi	12.329	12.171
Itä-Suomi	7.280	6.471
Oulu	5.002	5.026
Lappi	2.192	2.149
Ahvenanmaa	0.113	0.101
Koko maa	41.011	39.852

Jälkiositettujen estimaattien keskihajonnat ovat vain arvioita todellisen keskihajonnan suuruusluokalle, sillä keskihajonnan tarkan arvon laskeminen ei onnistu jälkiositetuille estimaateille. Arviot on laskettu soveltaen delta-menetelmää (Wolter 1985), eli varianssi on laskettu painotettuna summana jälkiositekohtaisista variansseista ja kovarianssitermit on jätetty summasta pois. Lisäksi summaan on lisätty osuuksien vaihtelua kuvaavat termit, eli osuusestimaattien varianssit kerrottuna osittaittain estimoitujen keskiarvojen neliöillä. Koska kovarianssitermejä ei ole otettu huomioon, ovat nämä arviot todellisia keskihajontoja pienempiä, mutta ne ovat kuitenkin riittävän tarkkoja arvioita keskihajontojen suuruusluokalle. Tämän voi todeta esimerkiksi tekemällä bootstrap-menetelmällä simulointikokeita (Efron & Tibshirani 1998).

Nämä arviot jälkiosituksella saatujen estimaattoreiden keskihajonnoille olivat miltei kaikissa lääneissä jonkin verran kalibrointiestimaattoreiden keskihajontoja pienempiä, kuten jälkiosituksen tapauksessa kuuluukin olla. Tämä aiheutuu luonnollisesti siitä, että vaihtelu on pienempää jälkiositteiden sisällä kuin välillä. Estimaattoreiden hyvyyden arvioinnissa on nyt kuitenkin otettava huomioon se, että jälkiositettujen estimaattoreiden harha voi olla hyvinkin suuri, sillä harhan määrään vaikuttaa se, kuinka luotettavasti osuudet p, q ja r on pystytty estimoimaan.

Taulukko 15: Saalismäärän keskiarvot ja keskiarvon keskivirheet lääneittäin sekä koko Suomessa laskettuna kalibrointipainotusta käyttäen sekä jälkiositusta hyödyntäen.

	kalibrointi-painotuksella lasketut estimaatit		jälkiositetut estimaatit	
	keskiarvo	s.e.	keskiarvo	arvio s.e.:lle
Etelä-Suomi	33.87	5.12	32.56	2.54
Länsi-Suomi	49.93	5.96	49.18	3.26
Itä-Suomi	57.74	3.76	51.75	2.20
Oulu	49.56	3.10	49.55	2.22
Lappi	57.88	5.69	57.01	7.62
Ahvenanmaa	71.47	15.42	58.65	6.40
Koko maa	44.97	2.90	43.62	1.65

Taulukko 16: Kalastuspäivien keskiarvot ja keskiarvon keskivirheet lääneittäin sekä koko Suomessa laskettuna kalibrointipainotusta käyttäen sekä jälkiositusta hyödyntäen.

	kalibrointi-painotuksella lasketut estimaatit		jälkiositetut estimaatit	
	keskiarvo	s.e.	keskiarvo	arvio s.e.:lle
Etelä-Suomi	32.92	2.89	31.78	1.78
Länsi-Suomi	34.71	2.64	34.26	1.77
Itä-Suomi	51.46	3.00	45.69	1.84
Oulu	53.27	4.27	53.51	2.85
Lappi	47.74	3.54	46.80	6.09
Ahvenanmaa	19.13	1.99	17.05	1.17
Koko maa	38.30	1.57	37.22	1.11

5.4 Suositukset vastauskadon huomioonottamiseen

Jos vastauskadon mallittamiseen sopivaa lisätietoa on saatavilla rekistereistä, tämä tieto kannattaa ehdottomasti käyttää hyväksi vastauskadon korjaamisessa. Apumuuttujia voidaan käyttää joko vastausaktiivisuusryhmien (response homogeneity groups) muodostamiseen tai muuhun kadon mallittamiseen (esimerkiksi apumuuttujien ollessa jatkuvia regressiomallin sovittamiseen apumuuttujien ja vastaamistodennäköisyyden välille). Taustatiedon laadusta riippuu, mitä menetelmää mallittamisessa kannattaa suosia. Jos apumuuttujat ovat jatkuvia ja ennustavat vastaamisastetta riittävän hyvin, antaa regressiomallinnus hyviä tuloksia. Jälkiosittamalla voidaan kuitenkin päästä lähes yhtä hyviin tuloksiin. Jälkiosittamista voidaan tehdä vastauskatoa mallittavien muuttujien mukaan, jolloin saadaan RHG-mallin mukaiset ositteet, joissa vastaamistodennäköisyydet voidaan olettaa vakioiksi. Toinen vaihtoehto jälkiosituksen käytölle on valita osittamiseen muuttuja, joka korreloi voimakkaasti tulosmuuttujan kanssa ja imputoida tulosmuuttujan puuttuvat arvot näiden jälkisoitteiden sisältä otetuilla havaituilla arvoilla. Jos apumuut-

tujista ei puolestaan ole tiedossa kuin reunajakaumat, voidaan käyttää kalibrointipainotusta.

Kaikki edellä mainitut menetelmät perustuvat kuitenkin siihen, että sopivia apumuuttujia saadaan käyttöön. Jos tällaisia apumuuttujia ei ole saatavilla, tai vastausaktiivisuuteen vaikuttaa monta muuttujaa, joista vain osasta voidaan saada tietoa vastaamattomien osalta, mallittamisesta ei ole apua vastauskadon huomioonottamiseen. Tällaisessa tilanteessa ainoa keino saada vastaamattomista lisätietoa on jälkikysely.

Jälkikysely toimii hyvin, jos se on suunniteltu huolellisesti. Kyselyn suunnittelussa on kiinnitettävä huomiota sekä vastaamisasteen korkeana pysymiseen että oikean tiedon hankintaan. Kysymykset on siis valittava siten, että niillä saadaan juuri sellaista tietoa, jonka avulla estimointi todella tehostuu. Jälkikyselyllä hankittujen tietojen on siis korreloitava voimakkaasti tutkittavan muuttujan kanssa. Vastaavasti, jos jälkikyselyllä halutaan vain mallittaa vastaamisaktiivisuutta, muuttujien on korreloitava voimakkaasti vastaamisasteen kanssa.

Jälkikyselyn järjestäminen lisää aina tutkimuksen kustannuksia. Jos tutkittavissa muuttujissa ei kuitenkaan tapahdu kovin nopeaa arvojen muuttumista, voidaan vuosittain järjestettävissä kyselyissä käyttää saman jälkikyselyn tuloksia useampana vuonna peräkkäin.

Jos vastaamisaste jälkikyselyssä saadaan pysymään suurena, saadaan jälkikyselyllä luotettavaa tietoa vastaamattomista. Jos saatu lisätieto on voimakkaasti yhteydessä tutkittavien muuttujien kanssa, voidaan aineisto jälkiosittaa tämän muuttujan mukaan. Jälkiosituksen käyttö tarkentaa estimointeja. Myös estimaattoreiden harha pysyy pienenä, jos jälkiositteisiin kuuluvien osuuksille saadaan luotettavat arviot jälkikyselystä.

6 Puuttuvien havaintojen ongelma

Puuttuvien havaintojen ongelmasta eli osittais- tai eräkadosta puhutaan silloin, kun tutkimusyksiköltä saadaan kerätyksi halutut tiedot vain osittain. Esimerkiksi henkilö palauttaa postitse lähetetyn kyselylomakkeen, mutta ei ole vastannut kaikkiin kysymyksiin tai kieltäytyy puhelinhaastattelussa vastaamasta joihinkin kysymyksiin, mutta suostuu kyllä vastaamaan muihin.

Osittaiskato vaikuttaa estimaatteihin samoin kuin kokonaiskatokin, eli aiheuttaa estimointeihin epätarkkuutta ja harhaa, mutta vain niiden muuttujien osalta, joita kato koskee. Harhan määrä riippuu luonnollisesti siitä, kuinka paljon vastanneet ja vastaamattomat poikkeavat toisistaan juuri näiden kyseessä olevien muuttujien suhteen.

Myös osittaiskatoa voi torjua jo ennalta kyselylomakkeen suunnittelulla. Mitä selkeämmiksi ja helpommin vastattaviksi kysymykset saadaan, sitä pienemmäksi osittaiskato näiden kysymysten kohdalla muodostuu. Silti vastaamisaktiivisuuteen vaikuttavat myös asiat, joihin tutkija ei voi vaikuttaa, kuten esimerkiksi kysymysten aihepiiri ja tilanne, jossa lomake täytetään tai puhelinhaastatteluun vastataan.

Osittaiskadon korjaamiseen ei juurikaan käytetä kokonaiskadon korjaamisessa yleisiä painotusmenetelmiä. Sen sijaan useimmiten osittaiskatoa paikataan imputoimalla.

6.1 Imputointi

Imputoinnilla tarkoitetaan puuttuvien arvojen korvaamista joillakin sopivilla arvoilla. Näitä imputoituja arvoja voidaan pitää puuttuvien arvojen ennusteina. Usein imputointi perustuu jonkun lisäinformaation käyttöön.

Korvaamalla kaikki puuttuvat havainnot saadaan aineisto, jota voidaan sitten analysoida kuten tavoitteena olevaa täydellistä aineistoa. Imputoidut arvot on kuitenkin aina merkittävä siten, että aineistosta voidaan jälkeenpäin tunnistaa, mitkä havainnoista ovat oikeasti havaittuja ja mitkä imputoituja. Imputoitua aineistoa käsitellessä täytyy aina pitää mielessä, ettei se ole täydellinen havaittu aineisto. Imputoinnin käyttö mm. aina heikentää estimointien tarkkuutta.

Imputointia voidaan käyttää sekä kokonais- että osittaiskadon paikkaamiseen. Molemmissa tapauksissa on kuitenkin muistettava, että imputointi lisää

usein harhaa ja vaikeuttaa otantavirheen arviointia. Myös imputointimenetelmän valintaa kannattaa miettiä tarkoin. Tähän valintaan vaikuttaa mm. se, mitä lisätietoa on käytettävissä sekä se, mitä aineistosta halutaan estimoida.

Parhaiten imputointi toimii yleensä tilanteissa, joissa taustalla on jokin sopiva apumuuttuja ja tämän apumuuttujan yhteys tulosmuuttujaan on hyvin mallitettu, sekä estimoitavat parametrit ovat yksinkertaisia (kuten keskiarvoja tai kokonaismääriä). Mm. jakaumien estimoiminen on usein hankalaa, varsinkin yksiarvoimputoinnin tapauksessa. Yleisesti ottaen imputoimalla saadaan parempia tuloksia, kun imputoitava muuttuja on jokin faktamuuttuja eikä mielipidettä tai asennetta mittaava muuttuja. Lisätietoa imputoinnista ja imputointimenetelmistä löytyy useista lähteistä (Laaksonen 1991, Laaksonen 1992, Little & Rubin 1987, Rubin 1987, Särndal, Swensson & Wretman 1992).

Hot deck

Hot deck -menetelmille on yhteistä se, että puuttuvan havainnon imputoinnissa käytetään jonkin muun vastaajan arvoa. Kyseinen arvo voidaan ottaa satunnaisesti joltain vastanneelta joko koko aineistosta tai jonkin samankaltaisista havainnoista koostuvan solun sisältä. Aineiston soluihin jakaminen tehdään jonkin sopivan taustamuuttujan arvojen homogeenisuuden mukaan.

Hot deck -menetelmien etuna on se, että puuttuvan arvon tilalle tulee todellinen aineistossa esiintyvä arvo. Soluittain tehtävässä hot deck -imputoinnissa on usein ongelmana oikean solukoon määrääminen. Imputoinnin tarkkuus paranee solukoon pienentyessä, mutta liian pienistä soluista voi olla mahdotonta löytää sopivaa korvaavaa arvoa (Laaksonen 1991, Laaksonen 1992, Little & Rubin 1987, Särndal, Swensson & Wretman 1992).

Cold deck -imputointi

Cold deck -imputoinnissa käytetään muusta lähteestä saatua arvoa, esimerkiksi aiemmasta tutkimuksesta tai jostain muusta historiallisesta lähteestä löydettyä sopivaa arvoa.

Cold deck -imputointi toimii hyvin, jos aikaisempi tutkimus antaa tarkasteltavan muuttujan osalta hyvin samanlaista tietoa kuin nykytilanne. Toisaalta

imputointi voi antaa hyvinkin vääristyneitä arvoja, jos aiemmasta tutkimuksesta on esimerkiksi kulunut kauan aikaa tai muuttujien arvot ovat muuten muuttuneet ratkaisevasti (Little & Rubin 1987, Särndal, Swensson & Wretman 1992).

Keskiarvoimputointi

Keskiarvoimputoinnilla tarkoitetaan sitä, että korvataan puuttuva arvo havaittujen arvojen keskiarvolla. Tämä menetelmä luonnollisestikin aiheuttaa arvojen keskittymistä, joten vaihtelu pienenee ja varianssi- ja kovarianssiestimaateista tulee harhaisia (liian pieniä) ja luottamusväleistä liian kapeita. Keskiarvoimputointi sopiikin vain tilanteisiin, joissa halutaan estimoida keskiarvoja tai kokonaismääriä, ei jakaumia.

Imputoinnissa voidaan käyttää kaikkien havaittujen arvojen kokonaiskeskiarvoa, mutta parempaan tulokseen päädytään jakamalla aineisto jonkin sopivan apumuuttujan mukaan mahdollisimman homogeenisiin soluihin ja korvaamalla puuttuvat arvot oman solunsa keskiarvoilla. Vaikka tulokset ovatkin parempia kuin kokonaiskeskiarvomenetelmällä, varianssi- ja kovarianssiestimaatit ovat edelleen liian pieniä (Little & Rubin 1987, Särndal, Swensson & Wretman 1992).

Mediaani-imputointi

Puuttuva arvo korvataan nyt havaittujen arvojen mediaanilla. Myös mediaani-imputointi voidaan tehdä koko aineiston mediaanilla tai soluittain.

Mediaani-imputoinnilla on sama haittapuoli kuin keskiarvoimputoinnillakin, eli myös tässä tapauksessa muuttujan vaihtelu tulee aliestimoiduksi. Hyvänä puolena on se, että mediaani ei ole herkkä poikkeaville havainnoille.

Moodi-imputointi

Korvataan puuttuva arvo moodilla eli yleisimmällä havaintoarvolla. Imputoinnissa voidaan käyttää koko aineiston moodia tai soluittain havaittuja moodeja. Myös moodi-imputointi aliestimoi muuttujan varianssia.

Malli-imputointi

Malli-imputointi perustuu konstruoitavaan tilastolliseen malliin, jolla voidaan ennustaa puuttuvat arvot. Yleinen menettelytapa on käyttää mallina jotain regressiomallia. Yleensä selittäjinä mallissa käytetään havaittuja muuttujien arvoja, jotka voivat olla joko tutkittavan muuttujan arvoja tai jonkin sopivan apumuuttujan havaittuja arvoja.

Imputointia voidaan vielä parantaa lisäämällä malliin jokin satunnaisosa, jolloin myös muuttujan vaihtelu saadaan paremmin otetuksi huomioon. Tällöin puhutaan stokastisesta regressioimputoinnista.

Malli-imputoinnin toimivuus riippuu siitä, kuinka sopiva malli aineistoon löydetään. Tähän puolestaan vaikuttaa mm. kadon suuruus ja se, onko aineistossa mallintamiseen sopivia apumuuttujia (Laaksonen 1991, Laaksonen 1992, Little & Rubin 1987, Särndal, Swensson & Wretman 1992).

Korvausmenetelmä (substitution)

Korvausmenetelmä ei oikeastaan ole varsinainen imputointimenetelmä. Siinä puuttuva havainto korvataan ottamalla arvo joltain otokseen kuulumattomalta yksiköltä eli tavallaan poimitaan otokseen uusi vastaaja, jolta kyseinen arvo saadaan. Näin saatua otosta ei tietenkään voida käsitellä kuten täydellistä otosta, koska vastanneiden ja vastaamattomien välillä muuttujan arvot voivat vaihdella systemaattisesti. Otosta analysoitaessa korvattuja arvoja täytyykin muistaa käsitellä kuten imputoituja arvoja.

Korvausmenetelmän huonona puolena on se, että tulokset voivat olla pahasti vääristyneitä, jos vastanneiden ja vastaamattomien välillä on suurta vaihtelua. Tämä menetelmä soveltuukin vain tilanteisiin, joissa tiedetään, etteivät vastanneet ja vastaamattomat poikkea systemaattisesti toisistaan tutkimusmuuttujien suhteen (Little & Rubin 1987).

Yhdistelmämenetelmät (composite methods)

Yhdistelmämenetelmillä tarkoitetaan imputointimenetelmiä, jotka yhdistävät ideoita kahdesta tai useammasta imputointimenetelmästä. Esimerkiksi voidaan käyttää regressiomenetelmää tuottamaan ennustettuja keskiarvoja sekä

lisätä näihin residuaalit, jotka on valittu hot deck -menetelmällä havaittujen jäännösten joukosta. Yhdistämällä residuaalit keskiarvoihin saadaan lopulliset imputoidut arvot (Little & Rubin 1987).

Moni-imputointi

Moni-imputoinnilla tarkoitetaan puuttuvan havainnon korvaamista useammalla kuin yhdellä arvolla. Näin saadaan useita täydellisiä aineistoja, joista jokaisesta voidaan estimoida halutut tunnusluvut erikseen ja lopuksi laskea näistä yhdistetty estimaatti.

Moni-imputoinnin etuna on se, että nyt muuttujan hajontaa ja jakaumaa pystytään arvioimaan paremmin kuin yksiarvoimputoinnin yhteydessä. Yksiarvoimputoinnissahan muuttujan vaihtelua on hankala arvioida, sillä jos puuttuva arvo imputoidaan jonkin muun yksikön vastaavalla arvolla, se voi saada hyvinkin poikkeuksellisen arvon, jos muuttujan vaihtelu on suurta. Keskiarvoimputointi taas aina kutistaa muuttujan varianssia, jolloin vaihtelu tulee aliestimoiduksi (Laaksonen 1991, Laaksonen 1992, Little & Rubin 1987, Rubin 1987, Särndal, Swensson & Wretman 1992).

K:n lähimmän naapurin menetelmä (k-NN)

Sekä kokonais- että osittaiskadon paikkaamiseen voidaan käyttää myös k:n lähimmän naapurin menetelmää. Tässä menetelmässä valitaan ensin sopivat taustamuuttujat, joiden avulla voidaan määritellä otosyksikköjen väliset etäisyydet. Etäisyysmitan määrittelyssä käytettävien apumuuttujien tulee olla sellaisia, että niistä on saatavilla tietoa myös vastaamattomien osalta. Sen jälkeen valitaan k havaittua yksikköä, jotka sijaitsevat "lähimpänä" puuttuvaa yksikköä. Näistä k:sta lähimmästä naapurista voidaan sitten esimerkiksi valita satunnaisesti yksi, jolta otetaan arvo puuttuvan havainnon paikalle. Toinen vaihtoehto on laskea kaikista k:sta havainnosta jollain sopivalla tapaa painotettu keskiarvo, ja korvata puuttuva havainto sillä.

Korvattaessa havainto vain yhdellä arvolla tuloksiin voi aiheutua suurtakin vaihtelua, jos k:n lähimmän naapurin joukossa on paljon vaihtelua. Painotetulla keskiarvolla korjaaminen puolestaan aiheuttaa varianssiestimaattorille liian pieniä arvoja, varsinkin jos havaittujen arvojen vaihtelu on todellisuudessa suurta.

Yksi vaihtoehto on käyttää moni-imputointia, eli valita tietty määrä havaittuja arvoja k:n lähimmän naapurin joukosta, jolloin saadaan useita täydellisiä aineistoja. Jokaisesta aineistosta voidaan estimoida erikseen haluttuja tunnuslukuja ja sitten laskea näistä jokin yhdistetty estimaatti. Moni-imputoinnin etuna on se, että muuttujan jakaumaa ja varianssia pystytään nyt arvioimaan paremmin kuin yhdellä arvolla imputoitaessa.

6.2 Jälkiosituksen käyttö puuttuvan tiedon imputoinnissa

Jälkiositusta käyttäen voidaan puuttuvan tiedon imputointia tehostaa, jos jälkiositus on tehokas, eli ryhmät ovat sisäisesti homogeenisia. Tällöin imputoinnit voidaan suorittaa kunkin jälkiositteen sisällä, jolloin aineiston vaihtelu tulee paremmin huomioiduksi kuin käyttämällä esimerkiksi keskiarvoimputointia koko aineiston keskiarvolla.

Parhaiten aineiston vaihtelun säilyttäminen onnistuu, jos käytetään moniarvoimputointia, eli korvataan puuttuvat havainnot useammalla havaitulla arvolla ja valitaan lopulliseksi korvaavaksi arvoksi joko satunnaisesti jokin näistä tai lasketaan jokin yhdistetty estimaatti havaituista arvoista (Särndal, Swensson & Wretman 1992). Moniarvoimputointi edellyttää kuitenkin, että jälkiositteiden on oltava riittävän suuria, jotta imputointia varten tarvittavia havaittuja arvoja on tarpeeksi.

6.3 Kvalitatiivinen tieto (puhelinhaastattelupaikkausta)

Jotta tiedettäisiin, kuinka paljon puuttuvia havaintoja on, täytyy saada selville kalastaneiden osuus vastaamattomien joukossa. Ne vastaamattomat, jotka eivät kalastaneet lainkaan, voidaan siis tässä jättää huomiotta, sillä kalastamattomat eivät tarjoa mitään informaatiota tutkimuksessa. Kalastamattomien osuus vastaamattomien joukossa voidaan selvittää esimerkiksi tekemällä vastaamattomille jälkikysely (esimerkiksi puhelinhaastatteluna), jossa selvitetään vain se, kuinka moni vastaamattomista on kalastanut. Näin saadaan siis selville imputoitavien havaintojen määrä. Jos puhelinhaastattelussa päästään pieneen vastauskatoon, saadaan kalastamattomien osuudelle lähes harhaton arvio.

Kalastaneiden osuutta vastaamattomien joukossa voidaan myös arvioida kolmannen kontaktiryhmän perusteella. Tämän arvion varianssi pysyy melko pienenä, mutta arvio voi olla harhainen.

6.4 Kvantitatiivisen tiedon imputointi

Ongelmana kalastuspäivien ja saalismäärän estimoinnissa on mm. se, että näiden muuttujien vaihtelu on suurta. Tämän vuoksi puuttuvien havaintojen paikkaus imputoimalla pitäisi suorittaa mahdollisimman homogeenisten ryhmien sisällä, jottei vaihtelua aliestimoitaisi. Aineiston jälkiosittamisesta voisi olla apua tämän ongelman ratkaisemiseen. Aineisto siis jaetaan jonkin apumuuttujan mukaan ositteisiin, joiden sisällä vaihtelu on koko aineiston vaihtelua pienempää, ja imputointi suoritetaan kunkin ositteen sisällä. Näin imputoinnissa menetetään huomattavasti vähemmän informaatiota vaihtelusta, kuin jos imputoinnit tehtäisiin koko aineistossa.

Kun tarvittava kvalitatiivinen tieto (esim. kalastaneiden osuus tai tiettyyn pyydysryhmään kuuluvien osuus kalastaneista) on saatu selville ja tarvittavien imputointien määrä on selvitetty, tarvitaan enää kvantitatiivinen tieto tutkittavasta muuttujasta (esimerkiksi keskiarvo, jos käytetään keskiarvoimputointia tai useita havaittuja muuttujan arvoja, jos käytetään moniimputointia). Tämä kvantitatiivinen tieto voidaan nyt estimoida erikseen kunkin jälkiositteen sisällä. Imputoinnit saadaan luonnollisesti sitä tarkemmiksi, mitä pienempiä ja homogeenisempia jälkiositteet ovat. Kuitenkin moniarvoimputointia käytettäessä jälkiositteiden pitää olla sen verran suuria, että ositteiden sisällä on riittävä määrä havaittuja muuttujan arvoja, joita voidaan käyttää imputoinneissa.

6.5 Herkkyysanalyysi

Herkkyysanalyysilla voidaan tutkia, kuinka hyvin imputointimenetelmien antamat korjatut aineistot vastaavat todellista tilannetta. Herkkyysanalyysia voidaan suorittaa esimerkiksi siten, että poistetaan havaitusta täydellisestä aineistosta tietyt havainnot ja korvataan nämä jonkin imputointimenetelmän antamilla ennusteilla. Nyt tätä imputoitua aineistoa voidaan verrata alkuperäiseen ja tutkia, miten hyvin imputoidut havainnot vastaavat alkuperäisiä sekä millainen vaikutus imputoinnilla on estimoituihin tuloksiin. Luonnolli-

sesti näin voidaan myös vertailla eri imputointimenetelmiä toisiinsa luomalla useita imputoituja aineistoja eri imputointitekniikoita käyttäen. Herkkyysanalyysilla voidaan tutkia myös sitä, kuinka suurta puuttuneisuus saa olla ennen kuin se vaikuttaa tuloksiin merkittävästi.

6.6 Puuttuvien havaintojen ongelma kalastuskyselyaineistossa

Vapaa-ajankalastus 1998 -aineistossa osittaiskadon määrä on melko pieni, joten puuttuvat havainnot eivät ole mikään suuri ongelma. Yleisesti ottaen puuttuvia havaintoja on eniten kysymyksissä, jotka käsittelevät kvantitatiivista tietoa, kuten saaliin määriä tai kalastuspäivien määriä. Näiden tietojen paikkaamiseen on käytetty pääasiassa hot deck -imputointia. Joitakin puuttuvia tietoja on myös pystytty päättelemään vastattujen kysymysten perusteella.

Yksi tapa parantaa imputointien tarkkuutta on käyttää jälkiositusta hyväksi imputointien suorittamisessa. Jos aineisto pystytään jälkiosittamaan jonkin sellaisen muuttujan mukaan, joka korreloi voimakkaasti tutkittavan muuttujan kanssa, voidaan puuttuvien arvojen imputointi tehdä jälkiositteiden sisällä. Tällöin imputoidut arvot saadaan todennäköisesti lähemmäs todellisia arvoja kuin tekemällä imputointeja koko aineistossa. Puuttuvat arvot voidaan korvata joko jälkiositteen keskiarvolla, satunnaisesti ositteesta valitulla havaitulla arvolla tai käyttämällä moniarvoimputointia. Jos jälkiositteet ovat riittävän suuria, moniarvoimputointi on paras ratkaisu, sillä tällöin pystytään arvioimaan myös muuttujan vaihtelua.

7 Otantayksikköongelman korjaaminen painotuksella

Jos poimintayksikkö ja tilastoyksikkö poikkeavat toisistaan, täytyy tämä ero korjata käyttämällä estimoinnissa sopivaa painotusta. Kalibrointimenetelmän antavat painokertoimet sopivat hyvin tällaisen eroavaisuuden korjaamiseen, jos käyttöön saadaan sopivat reunajakaumat kalibrointia varten. Kalibrointipainotuksella voidaan lisäksi myös oikaista vastauskadon aiheuttamaa harhaa.

7.1 Otospainojen laskeminen

Otantatutkimuksissa havainnoille täytyy aina laskea otospainot. Yksinkertaisimmassa tilanteessa, jossa poimintayksikön ja tilastoyksikön välillä ei ole eroa, aineistossa ei ole katoa, eikä ole muitakaan syitä painottaa havaintoja jollain erityisellä tavalla, otospainot saadaan alkion otokseensisältymistodennäköisyyden käänteislukuna. Tämän painokertoimen tehtävänä on siis vain yleistää otoksesta lasketut estimaatit koko perusjoukon tasolle.

Tällaista täydellistä otanta-asetelmaa ei kuitenkaan käytännössä pystytä juuri koskaan rakentamaan, joten painotusta joudutaan monimutkaistamaan. Sopivalla painotuksella voidaan korjata poimintayksikön ja tilastoyksikön välisiä eroja ja vastauskadosta aiheutuvaa harhaa. Joskus myös tutkija voi haluta painottaa otosalkioita jollain tietyllä tavalla, esimerkiksi taloustutkimuksissa voidaan haluta suuremmille yrityksille enenmmän painoarvoa, jolloin yrityksiä voidaan painottaa esimerkiksi liikevaihdollaan.

Katoa korjaavaa painotusta varten tarvitaan jokin apumuuttuja, jolla on vaikutusta vastaamiseen. Aineisto jaetaan tämän muuttujan avulla ositteisiin, joiden sisällä vastaamistodennäköisyydet ovat homogeenisia. Sen jälkeen voidaan havainnoille laskea painokertoimet, jotka ottavat huomioon osittaittain havaitun vastaamistodennäköisyyden.

7.2 Kalibrointi

Jos poimintayksikkö ja tilastoyksikkö eroavat toisistaan, täytyy tämän eron aiheuttama harha korjata kalibroimalla. Kalibroinnin tavoitteena on muo-

dostaa painokertoimet, jotka ovat mahdollisimman lähellä alkuperäisiä painoja (esimerkiksi perusotospainoja) siten, että otoksesta lasketut reunajakaumat saadaan mahdollisimman oikeiksi. Lisäksi kalibroinnissa voidaan myös määrätä painoille tietyt rajoitteet, jolloin vältytään mm. regressiomenetelmän tapauksessa usein ongelmia aiheuttavilta liian suurilta tai negatiivisilta painokertoimilta.

Kalibroinnissa tarvitaan sopivia apumuuttujia, joista tiedetään reunajakaumat koko perusjoukossa. Nämä apumuuttujat saadaan usein otantatutkimuksen ulkopuolisesta lähteestä, esimerkiksi rekisteristä (Deville & Särndal 1992).

Merkitään populaatiota U:lla siten, että $U = \{1, ..., u, ..., N\}$. Lisäksi merkitään s:llä kokoa n olevaa otosta, joka on poimittu populaatiosta U käyttäen annettua otanta-asetelmaa. Oletetaan, että kiinnostuksen kohteena on muuttujan y kokonaismäärä $t_y = \sum_U y_u$ ja estimoinnissa hyödynnettävä lisäinformaatio on muuttujissa x_j. Merkitään vielä alkion u otokseensisältymistodennäköisyyttä π_u:lla. Otokseen kuuluvaa alkiota u kohti on siis tiedossa vektori $(y_u, \mathbf{x}_u)$, missä $\mathbf{x}_u = (x_{u1}, ..., x_{uj}, ..., x_{uJ})'$. Lisäksi oletetaan että apumuuttujista x on tiedossa koko populaatiosta lasketut kokonaismäärät $\mathbf{t}_x = \sum_U \mathbf{x}_u$.

Jos aineisto on täydellinen ja otantayksikön ja poimintayksikön välillä ei ole eroa, voidaan kokonaismäärä t_y estimoida käyttämällä Horvitz-Thompson-estimaattoria $\hat{t}_{y\pi} = \sum_s \frac{y_u}{\pi_u}$ (Särndal, Swensson & Wretman 1992). Kalibrointipainotettu estimaatti t_y:lle muodostetaan siten, että uudet painokertoimet w_u ovat mahdollisimman lähellä alkuperäisiä otantapainoja $d_u = \frac{1}{\pi_u}$. Jos halutaan lisäksi korjata vastauskatoa, paino d_u saa muodon $d_u = \frac{1}{\pi_u \theta_u}$, missä θ_u on (estimoitu) vastaamistodennäköisyys. Lisäksi kalibrointipainon muodostamisessa halutaan, että apumuuttujien $\mathbf{x}$ otoksesta lasketut painotetut reunajakaumat täsmäävät etukäteen tiedettyihin perusjoukon reunajakaumiin siten, että

$$\sum_s w_u \mathbf{x}_u = \sum_U \mathbf{x}_u.$$

Edellisessä yhtälössä siis $\mathbf{x}_u$ sisältää apumuuttujien $\mathbf{x}$ arvot u:nnella tilastoyksiköllä, ja vasemmanpuoleinen summaus tapahtuu otoksen yli, kun taas oikeanpuoleinen koko perusjoukon yli. Tämän yhtälön toteutuminen takaa sen, että painotetut lisämuuttujien totaaliestimaatit täsmäävät tunnettujen perusjoukon kokonaismäärien kanssa. Toisin sanoen tällöin kalibrointipainot antavat täydelliset estimaatit apumuuttujan kokonaismäärien estimointiin sovellettuna. Näin voidaan siis varmistaa estimaattoreiden tarkkuus. Jos

apumuuttuja korreloi voimakkaasti tutkimusmuuttujan kanssa ja painotus toimii hyvin apumuuttujan estimoinnissa, voidaan olettaa, että painokertoimet antavat järkeviä tuloksia myös tutkimusmuuttujan estimointiin sovellettuna.

Painojen välisen etäisyyden määrittelyyn on olemassa esitetty vaihtoehtoisia etäisyysfunktioita (Deville, Särndal & Sautory 1993, Deville & Särndal 1992). Myös etäisyysfunktion minimoinnin suorittamiseen voidaan käyttää useita menetelmiä. Yleisimmin käytössä on yleistetty pienimmän neliösumman menetelmä (Generalized Least Squares, GLS).

Etäisyysfunktiolle G on olemassa tiettyjä vaatimuksia, jotka sen pitää toteuttaa: G:n pitää olla positiivinen, kahdesti derivoituva ja konveksi funktio. Lisäksi seuraavat ehdot täytyy olla voimassa: $G(1) = G'(1) = 0$ ja $G''(1) = 1$. Minimoitava lauseke on tällöin

$$\sum_s d_u G(\frac{w_u}{d_u}) - \lambda'(\sum_s w_u \mathbf{x}_u - \sum_U \mathbf{x}_u),$$

missä vektori λ on Lagrangen kerroin ja etäisyysfunktio $\sum_s G(\frac{w_u}{d_u})$ mittaa d_u:n ja w_u:n etäisyyttä otoksessa. Kun tämä lauseke derivoidaan w_u:n suhteen, päädytään yhtälöön

$$g(\frac{w_u}{d_u}) - \mathbf{x}_u'\lambda = 0.$$

Funktio g on nyt siis G:n derivaatta. Kun w_u ratkaistaan tästä, lopuksi saadaan

$$w_u = d_u F(\mathbf{x}_u'\lambda),$$

missä $F(t) = g^{-1}(t)$. Painokertoimien laskemiseksi täytyy vielä määritellä kertoimen λ arvo. Tämä tapahtuu ratkaisemalla yhtälö

$$\sum_s d_u F(\mathbf{x}_u'\lambda)\mathbf{x}_u = \sum_U \mathbf{x}_u.$$

Tämä ratkaistaan yleensä käyttämällä soveltuvaa numeerista menetelmää.

Yksinkertaisimmillaan minimoitavaksi etäisyysfuktioksi voidaan valita seuraava funktio

$$\mathrm{E}_p\left[\sum_s \frac{(w_u - d_u)^2}{d_u q_u}\right],$$

missä $E_p(.)$ tarkoittaa odotusarvoa käytetyn otanta-asetelman suhteen ja q_u on jokin positiivinen painokerroin, jonka ei tarvitse riippua d_u:sta. Erityisesti SAS-makro CLAN97 laskee kalibrointipainot käyttäen tätä etäisyysfunktiota. Tätä menetelmää kutsutaan myös lineaariseksi menetelmäksi, koska tämä G:n valinta johtaa lineaariseen funktioon $F(t) = g^{-1}(t)$.

Kun ylläoleva etäisyysfuktio minimoidaan w_u:n suhteen, saadaan kalibrointipainoiksi

$$w_u = d_u(1 + q_u \mathbf{x}_u \lambda),$$

missä vektori λ määrätään siten, että yhtälö $\Sigma_s w_u \mathbf{x}_u = \Sigma_U \mathbf{x}_u$ toteutuu. Näin saadaan λ:lle ratkaisu

$$\lambda = \mathbf{T}_s^{-1}(\mathbf{t}_x - \hat{\mathbf{t}}_{x\pi}),$$

missä

$$\mathbf{T}_s = \sum_s d_u q_u \mathbf{x}_u \mathbf{x}_u'$$

ja $\hat{\mathbf{t}}_{x\pi} = \sum_s d_u x_u$ on t_x:n Horvitz-Thompson-estimaattori. Tavallisesti painokerroin $q_u = 1$, jolloin kalibrointipaino w_u sievenee muotoon

$$w_u = d_u(1 + \mathbf{x}_u \lambda)$$

ja

$$\mathbf{T}_s = \sum_s d_u \mathbf{x}_u \mathbf{x}_u'.$$

Muuttujan y kokonaismäärän t_y kalibrointiestimaattori voidaan siis kirjoittaa lausekkeena

$$\hat{t}_{ycal} = \sum_s w_u y_u = \hat{t}_{y\pi} + (\mathbf{t}_x - \hat{\mathbf{t}}_{x\pi})' \hat{\mathbf{B}}_s,$$

missä

$$\hat{\mathbf{B}}_s = \mathbf{T}_s^{-1} \sum_s d_u \mathbf{x}_u y_u.$$

Tästä yhtälöstä nähdään, että kun painojen välisen etäisyyden minimointi suoritetaan käyttäen lineaarista menetelmää, kalibrointiestimaattori on sama kuin yleistetty regressioestimaattori kokonaismäärälle t_y.

Kalibroitujen estimaattoreiden varianssien estimoinnissa lasketaan ensin jäännökset $e_u = y_u - \mathbf{x}_u \hat{\mathbf{B}}_s$. Varianssiestimaattori voidaan sen jälkeen kirjoittaa muodossa

$$\hat{\mathrm{Var}}(\hat{t}_{yw}) = \sum_{u \in s} \sum_{v \in s} \frac{\Delta_{uv}}{\pi_{uv}} (w_u e_u)(w_v e_v),$$

missä $\Delta_{uv} = \pi_{uv} - \pi_u \pi_v$, (Deville, Särndal & Sautory 1993).

Kalibrointiin liittyvä laskenta suoritetaan käyttäen jotain kyseiseen tarkoitukseen kehitettyä tietokoneohjelmaa. Tällaisia ovat esimerkiksi Ruotsin tilastoviraston kehittämä SAS-makro CLAN97 ja Ranskan tilastoviraston kehittämä SAS-makro CALMAR.

Kalibroinnin etuna on mm. se, että siinä voidaan käyttää yhtä aikaa hyödyksi useita perusjoukkotason muuttujia. Muuttujat voivat tulla myös eri lähteistä, sillä oletukseksi riittää, että niiden reunajakaumat tunnetaan. Koska ristiinluokittelua ei tarvita, eivät myöskään ristiinluokittelulle tyypilliset liian pienet solukoot aiheuta ongelmia. Kalibroinnin voi toki myös suorittaa käyttämällä solufrekvenssejä, jos sellaiset ovat tiedossa. Tällöin puhutaan täydellisestä jälkiosittamisesta, kun taas kalibrointia marginaalijakaumilla kutsutaan vastaavasti epätäydelliseksi jälkiosittamiseksi. Pelkillä reunajakaumilla kalibrointi voi kuitenkin tietyissä tapauksissa osoittautua lähes yhtä tehokkaaksi kuin solufrekvenssien käyttö.

Lopullista painotusmenetelmää valittaessa suositellaan yhdistelemään kumulatiivisesti eri menetelmiä, kuitenkin siten, että viimeiseksi suoritetaan kalibrointi, jolla saadaan reunajakaumat oikeiksi.

Kalibrointi on tilastotieteellisesti tehokas menetelmä, sillä jos apumuuttujat korreloivat voimakkaasti vasteen kanssa, otosvarianssi pysyy pienenä. Kalibrointi on myös approksimatiivisesti harhaton menetelmä ja tarkentuva kaikille estimaateille. Kalibrointimenetelmän huonona puolena on otosvarianssien estimoinnin vaikeus. Lisäksi menetelmän käyttöä rajoittaa se, että sopivaa taustatietoa ei aina ole saatavissa.

7.3 Painotus Vapaa-ajankalastus 1998 -aineiston analysoinnissa

Kalastuskyselyaineistoissa esiintyy yleensä ainakin jossain määrin vastauskatoa. Tämä kato on myös usein valikoitunutta, eli vastanneet ja vastaamattomat poikkeavat toisistaan tutkimusmuuttujien suhteen. Vapaa-ajankalastus 1998 -aineistossa kokonaiskadon suuruus on yli 30 % otoskoosta. Tällöin vastauskato täytyy ehdottomasti ottaa huomioon painokertoimia muodostettaessa. Myös toinen painotusta vaativa ongelma eli otantayksikköongelma esiintyy tässä kyselyssä; poimintayksikkönä käytettiin asuntokuntaa, mutta

tilastoyksikkönä kotitaloutta. Tämä ero täytyy myös korjata sopivalla painotuksella.

Vapaa-ajankalastus 1998 -aineistossa käytetty painotus koostuu kolmesta osasta: otokseensisältymistodennäköisyyden käänteisluvusta, ositteittain havaitun vastaamistodennäköisyyden käänteisluvusta sekä kalibrointipainosta. Ositteittain havaitun poimintatodennäköisyyden pitäisi nyt korjata vastauskadon vaikutuksia aineistossa. Tämä painokerroin on kuitenkin puutteellinen, sillä oletus siitä, että vastauskato olisi ositteittain vakio, ei näytä kovin hyvin pitävän paikkaansa. Tämä painokerroin ei myöskään ota huomioon vastauskadon valikoituneisuutta. Kalibrointipainon tehtävänä on puolestaan korjata poiminta- ja tilastoyksikön välistä eroa. Tässä aineistossa kalibroinnissa käytettiin reunajakaumina tilastokeskuksen kotitalouskyselystä saatua kotitalouksien kokoluokkajakaumaa sekä väestörekisteristä saatuja miesten ja naisten ikäjakaumia lääneittäin. Kalibrointipainotus korjaa myös vastauskatoa, sillä vastaamisaktiivisuus näyttää riippuvan kotitalouden koosta sekä vastaajan iästä ja sukupuolesta.

Kalastuskyselyissä ongelmia tuottaa usein se, ettei vastaamisaktiivisuutta hyvin mallittavaa taustatietoa ole saatavissa vastaamattomista, jolloin katoa korjaava painotus jää melko tehottomaksi tai sitä ei voida käyttää lainkaan. Jälkikyselyn suorittaminen vastaamattomille on kuitenkin yksi ratkaisu saada vastaamattomista tietoa, jonka avulla aineisto esimerkiksi pystytään jälkiosittamaan vastaamistodennäköisyyden mukaan homogeenisiin osittеisiin, ja näille voidaan havaittuja vastaamisasteita käyttäen laskea osittеittaiset vastauskatoa korjaavat painot.

8 Diskussio

8.1 Vastaamattajättäneiden ongelman ratkaisu

Vastauskadon korjaamiseen voidaan käyttää sekä imputointeihin että painottamiseen pohjautuvia menetelmiä. Imputoinnit soveltuvat paremmin osittaiskadon paikkaamiseen, kun taas painotusmenetelmiä kannattaa käyttää kokonaiskadon korjaamisessa. Molemmissa menetelmissä tarvitaan lisätietoa, joka pitäisi saada koko populaatiosta ja jonka pitäisi korreloida voimakkaasti tutkittavien muuttujien kanssa. Sopivaa tietoa on kalastuskyselyiden tapauksessa usein hankala saada rekistereistä, joten tiedon hankintaan täytyy löytää muita tapoja.

8.2 Lisäinformaation hankkiminen aineistoa täydentämällä

Jos sopivaa rekisteritietoa vastaamattomista ei ole saatavilla, on jälkikysely ainoa keino hankkia vastaamattomista lisätietoa. Jälkyselyllä voidaan pyrkiä hankkimaan lisätietoa vastauskadon mallittamiseen, imputointien tehostamiseen tai tutkimusmuuttujien mallittamiseen apumuuttujien avulla.

Jälkikyselyllä saadaan hyviä tuloksia, jos kyselyn vastaamisaste saadaan pidettyä korkeana ja kyselyssä saadaan tietoa sellaisista muuttujista, jotka korreloivat voimakkaasti joko vastaustodennäköisyyden tai tutkittavien muuttujien kanssa. Tietysti parasta olisi, jos jälkikyselyllä pystytään selvittämään suoraan tutkittavien muuttujien arvoja vastaamattomilla. Yleensä kalastuskyselyissä tämä ei kuitenkaan toimi, sillä tutkittavat muuttujat ovat usein kvantitatiivisia, kuten saalismääriä. Tällaista tietoa jälkikyselyllä on hankala kerätä, sillä esimerkiksi saalismäärien arvioiminen pitkältä ajanjaksolta on aktiivisillekin kalastajille melko vaikeaa. Yleensä vastaamattomissa on suurin osa satunnaisia kalastajia, jotka jättävät helposti vastaamatta koko kyselyyn, jos heiltä kysytään vaikeasti muistettavaa tai arvioitavaa tietoa. Lisäksi ne, jotka suostuvat vastaamaan saattavat arvioida saaliinsa niin huonosti, että tulosten luotettavuus jää heikoksi.

Jälkikyselyllä kannattaakin siis hankkia kvalitatiivista tietoa, koska tällöin vastaamisaste saadaan todennäköisesti pidettyä suhteellisen hyvänä. Tämän kvalitatiivisen tiedon avulla aineisto voidaan jälkiosittaa.

Jälkiositus voidaan muodostaa imputointien tehostamiseksi. Tällöin pyritään siihen, että tutkimusmuuttujan vaihtelu ositteiden sisällä on mahdollisimman pientä. Jos tällainen ositus saadaan muodostetuksi, voidaan sitä käyttää myös siten, että estimaatit keskiarvoille ja totaaleille lasketaan jälkiositteisiin kuuluvien osuuksilla painotettuina summina ositteittain lasketuista estimaateista. Tällöin estimaattorin varianssia saadaan yleensä pienennettyä.

Jälkiositus voidaan muodostaa myös sellaisen muuttujan mukaan, joka vaikuttaa voimakkaasti vastaamisaktiivisuuteen. Tällöin vastaamisasteelle voidaan laskea ositteittaiset estimaatit ja käyttää näitä katoa korjaavien painokertoimien muodostamisessa.

8.3 Yhteenveto empiirisistä kokemuksista

Luvussa 5 tehtiin erilaisia empiirisiä tarkasteluja vastauskadon tutkimiseksi. Kontaktiryhmiä verratessa havaittiin, että kalastaneiden osuus sekä keskimääräiset kalastuspäivät ja saalismäärät pienenevät siirryttäessä ensimmäisestä kontaktiryhmästä toiseen ja saavat kolmannessa kontaktiryhmässä vielä jonkin verran toista ryhmää pienempiä arvoja.

Aineistoon kokeiltiin RHG-mallia, jossa oletettiin, että vastaamattomien kalastuskäyttäytyminen muistuttaisi kolmannen kontaktiryhmän käyttäytymistä. Tällä mallilla saadaan jonkin verran pienempiä tuloksia kalastuspäivien ja saaliin määrälle kuin alkuperäistä painotusta käytettäessä.

Vuoden 2000 kyselyyn sovelletussa jälkikyselyssä havaitaan, että RHG-mallin oletus vastaamattomien ja kolmannen kontaktiryhmän homogeenisuudesta näyttäisi pitävän melko hyvin paikkaansa. RHG-mallin antamia tuloksia voidaan siis pitää luotettavina.

Sovellettaessa jälkiositusta Vapaa-ajankalastus 1998 -aineistoon saadaan myös hieman alkuperäisiä tuloksia pienempiä estimaatteja kalastuspäiville ja saaliin määrälle. Tässä kokeillussa jälkiosituksessa osuuksiinkuuluvien osuudet jouduttiin estimoimaan aineistosta, koska vuoden 1998 kyselyn vastaamattomista ei ollut käytössä sopivaa lisätietoa. Osuuksien estimoinnissa tosin käytettiin hyväksi RHG-mallia, joka korjaa vastauskatoa. Jälkiosituksen käyttö pienentää myös estimaattoreiden vaihtelua.

8.4 Suositukset

Rekistereistä saadun lisätiedon käyttö vastauskadon mallittamisessa ja estimointien tehostamisessa on luonnollisesti helpoin ja kustannuksiltaan edullinen tapa parantaa estimointeja. Kalastuskyselyissä ongelmana on usein se, ettei sopivaa rekisteritietoa ole saatavilla.

Jos kato on suurta, vastaamattomien kalastuskäyttäytymistä täytyy kuitenkin jotenkin pystyä selvittämään. Tällöin jälkikysely tarjoaa menetelmän hankkia vastaamattomista sellaista lisätietoa, jonka avulla kadon vaikutuksia voidaan korjata. Jälkikysely parantaa kuitenkin estimointeja vain, jos se on hyvin toteutettu, eli vastauskato jälkikyselyssä on pientä ja kysymykset on valittu siten, että saatu aputieto on voimakkaasti yhteydessä tutkimusmuuttujiin.

Jälkikyselyn avulla toteutettu jälkiositus näyttää pienentävän muuttujien välistä vaihtelua ja korjaavan katoa. Myös harha pysyy pienenä, kun vastauskato jälkikyselyssä pidetään pienenä, eli kyselyn avulla voidaan luotettavasti estimoida ositteisiinkuulumisosuudet.

Jälkiosituksen avulla voidaan myös korjata osittaiskatoa tekemällä imputointeja ositteiden sisällä. Näin puuttuville arvoille saadaan todennäköisesti lähempänä todellista oleva arvo kuin imputoimalla koko aineiston sisällä, missä vaihtelu on paljon suurempaa.

Viitteet

Deville, J.-C. & Särndal, C.-E. : *Calibration estimators in survey sampling.* Journal of American Statistical Association. Vol. 87, pp. 376-381, 1992.

Deville, J.-C., Särndal, C.-E. & Sautory, O. : *Generalized raking procedures in survey sampling.* Journal of American Statistical Association. Vol. 88, pp. 1013-1020, 1993.

Efron, B. & Tibshirani, R.J. : *An introduction to the bootstrap* Chapman & Hall. 1998.

Kekäläinen, K. : *Hierarkkinen jälkiositus, kalibrointi ja katopainotus surveytutkimuksessa: sovellus vapaa-ajankalastuskyselyyn.* Pro gradu -tutkimus. Jyväskylän yliopisto. 2002.

Laaksonen, S. : *Comparative adjustments for missingness in short-term panels.* Tilastokeskus. Tutkimuksia 179. 1991.

Laaksonen, S. : *Handling household survey nonresponse data.* Suomen tilastoseura. Tilastotieteellisiä tutkimuksia 13. 1992.

Little, R.J.A. & Rubin, D.B. : *Statistical analysis of missing data.* John Wiley & Sons, Inc. 1987.

Rubin, D.B. : *Multiple imputation for nonresponse in surveys.* John Wiley & Sons, Inc. 1987.

Särndal, C.-E., Swensson, B. & Wretman, J. : *Model assisted survey sampling.* Springer-Verlag New York, Inc. 1992.

Vapaa-ajankalastus 1998. Riista- ja kalatalouden tutkimuslaitos. 2000.

Wolter, K.M. : *Introduction to variance estimation.* Springer-Verlag New York, Inc. 1985.

LIITE 1

RIISTAN- JA KALANTUTKIMUS

Riista- ja kalatalouden tutkimuslaitos
PL 6
00721 Helsinki

1. A. Kuinka monta henkilöä kuuluu kotitalouteenne?
Tutkimushetkellä kotitalouteen kuuluu yhteensä _____ henkilöä.

B. Merkitkää seuraavaan taulukkoon kotitaloutenne henkilöiden lukumäärä iän ja sukupuolen mukaan

	alle 10-vuotiaita	10-17 vuotiaita	18-24 vuotiaita	25-44 vuotiaita	45-64 vuotiaita	vähintään 65-vuotiaita
Naisia						
Miehiä						

2. Kalastiko tai ravustiko yksikään kotitaloutenne jäsenistä vuonna 1998? Merkitkää yksi rasti.
Kalastamiseksi katsotaan se, että on käyttänyt mitä tahansa pyyntimuotoa (esim. verkot, katiskat, onget, pilkkivavat yms.) edes yhdenkin kerran vuoden 1998 aikana. Henkilön katsotaan kalastaneen vaikka hän olisi vain soutanut tai ohjannut venettä toisen kalastaessa.

☐ 1 Kyllä, ja sai saalista.

☐ 2 Kyllä, mutta kukaan ei saanut saalista.

☐ 3 Ei, mutta on kalastanut tai ravustanut aikaisemmin.

☐ 4 Ei ole kalastanut eikä ravustanut koskaan.

3. Merkitkää seuraavaan taulukkoon kotitaloutenne **vuonna 1998 kalastaneiden (tai ravustaneiden) henkilöiden** lukumäärät

	alle 10-vuotiaita	10-17 vuotiaita	18-24 vuotiaita	25-44 vuotiaita	45-64 vuotiaita	vähintään 65-vuotiaita
Naisia						
Miehiä						

4. Kuinka monta kotitaloutenne jäsentä kuului kuhunkin seuraavista ryhmistä kalastuksensa (tai ravustuksensa) perusteella vuonna 1998? Merkitkää jokainen kotitalouden jäsen kuuluvaksi johonkin ryhmään, mutta vain yhteen ryhmään.

Lukumäärä

_____ 1 Ei kalastanut lainkaan.

_____ 2 Osallistui kalastamiseen ainoastaan soutamalla tai ohjaamalla venettä.

_____ 3 Kalastus oli yksi harrastus muiden joukossa.

_____ 4 Kalastus oli tärkein tai lähes tärkein harrastus.

_____ 5 Kalastus oli tärkein tai lähes tärkein harrastus ja myös kalastuskiloailuihin osallistuttiin.

5. **Arvioikaa kotitaloutenne kalastaneiden (tai ravustaneiden) henkilöiden lukumäärät ja heidän yhteenlasketut kalastuspäivien lukumääränsä kalastusmatkailussa ja muussa kalastuksessa Suomessa vuonna 1998.**
Kalastusmatkailulla tarkoitetaan matkaa, jolla myös kalastettiin ja kalastus tapahtui muualla kuin asuinkunnassa tai vapaa-ajanasuntonne läheisyydessä.

	Kalastusmatkailu 1	Muu kalastus 2
Kalastaneita henkilöitä 1		
Kalastuspäiviä 2		

---

6. **Kuinka moni kotitaloutenne jäsen kalasti (tai ravusti) seuraavilla pyydystyypeillä Suomessa eri alueilla vuonna 1998?**

Pyydys	Sisävesialue					Merialue			
	Etelä-Suomi 1	Länsi-Suomi 2	Itä-Suomi 3	Oulunlääni 4	Lappi 5	Suomenlahti 6	Saaristomeri ja Ahvenanmaa	Selkämeri ja Merenkurkku	Perämeri 9
Verkko 1									
Katiska, merta tai rysä 2									
Pilkkivapa 3									
Onki 4									
Heittovapa 5									
Perhovapa 6									
Vetouistin 7									
Muu pyydys, mikä? (esim. nuotta, syöttikoukku, pilkkiminä) 8									

---

7. **Arvioikaa myös kotitaloutenne kaikkien henkilöiden yhteenlasketut kalastuspäivien lukumäärät eri pyydystyypeillä.**
Kalastuspäivällä tarkoitetaan vapapyydysten osalta sitä, että yksi henkilö on kalastanut tietyllä vapapyydyksellä yhtenä päivänä. Verkko-, katiska-, merta- ja rysäpyydysten osalta kalastuspäivällä tarkoitetaan sitä, että henkilö on kokenut kyseisen tyyppisiä pyydyksiä yhtenä päivänä.

Pyydys	Sisävesialue					Merialue			
	Etelä-Suomi 1	Länsi-Suomi 2	Itä-Suomi 3	Oulunlääni 4	Lappi 5	Suomenlahti 6	Saaristomeri ja Ahvenanmaa	Selkämeri ja Merenkurkku	Perämeri 9
Verkko 1									
Katiska, merta tai rysä 2									
Pilkkivapa 3									
Onki 4									
Heittovapa 5									
Perhovapa 6									
Vetouistin 7									
Muu pyydys, mikä? (esim. nuotta, syöttikoukku, pilkkiminä) 8									

---

8. **Arvioikaa kotitaloutenne verkkokalastusvuorokausien lukumäärät jää- ja avovesikalastuksessa.**
Kalastusvuorokausi on yksi vuorokausi, jonka aikana teillä on ollut pyynnissä yksi tai useampia verkkoja.

Vuorokausia ☐ 1 **Jäältäkalastus**

Vuorokausia ☐ 2 **Avovesikalastus**

KOTITALOUDEN SAALIS 1998

Arvioikaa alla olevaan taulukkoon kotitaloutenne **Suomesta vuonna 1998** eri pyydyksillä saama saalis. Kalansaaliit ilmoitetaan kiloina, rapusaaliit kappaleina.

Arvioikaa saaliit perkaamattomana painona.

Jos olette kalastanut ja saanut saalista muiden kotitalouksien kanssa, ilmoittakaa **vain oman kotitaloutenne osuus**.

Ilmoittakaa myös jokaisen lajin tärkein kalastusalue (koodi kartasta) sekä saaliin jakautuminen (kg) **merestä**, järvestä ja joesta saatuun saaliiseen.

Pyydys	Kotitalouden saalis kiloina (kg)																			Saalis kpl.
	Ahven	Hauki	Särki	Lahna	Siika	Muikku	Made	Kuha	Taimen	Merilohi	Järvilohi	Kirjolohi	Silakka	Kilohaili	Turska	Kampela	Säyne	Kuore (Norssi)	Harjus (Harri)	Muu	Rapu / Tapa-rapu
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
Verkko 1																					
Katiska, merta tai rysä 2																					
Pilkkivapa 3																					
Onki 4																					
Heittovapa 5																					
Perhovapa 6																					
Vetouistin 7																					
Muu pyydys, mikä? (esim. nuotta, syöttikoukku, pitkäsiima) 8																					
Lajin tärkein kalastusalue (koodi kartasta) 9																					
Saalis merestä 10																					
Saalis järvestä 11																					
Saalis jo sta 12																					

KOTITALOUDEN SAALIIN KÄYTTÖ 1998

Arvioikaa alla olevaan taulukkoon kotitaloutenne 1998 saaman saaliin käyttö kalalajeittain.

Ilmoittakaa saaliit perkaamattomana painona.

Saaliin käyttömuodot	Saaliin käyttö kiloina (kg)																			Kpl.
	Ahven 1	Hauki 2	Särki 3	Lahna 4	Siika 5	Muikku 6	Made 7	Kuha 8	Taimen 9	Meritohi 10	Järvitohi 11	Kirjolohi 12	Sitakka 13	Kiichaili 14	Turska 15	Kampela 16	Säyne 17	Kuore (Norssi) 18	Harjus (Harri) 19	Muu 20	Rapu / Tapiarapu 21
Käytetty ihmisravinnoksi omassa taloudessa 1																				
Myyty tai annettu pois ihmisravinnoksi 2																				
Käytetty eläinten ruoaksi 3																				
Heitetty pois tai kompostoitu 4																				
Muu käyttö, mikä?_____ 5																				

Lisätietoja:

Kiitos vastauksestanne.

LIITE 2

MUUTTUJA	SELITYS
aht	poimintatodennäköisyydestä tuleva painokerroin = 1/piisi
akkoko	asuntokunnan koko
al1-al21	kalalajin tärkein kalastusalue (saa arvot 1-9)
hlkm	kotitalouteen kuuluvien henkilöiden määrä (kysymys 1b))
hryh	kotitalouden henkilöiden lkm (summana kysymyksen 4 vastauksista)
imp1a	kysymykseen 1a) liittyvä imput.mja (0, jos aito havainto; 1, jos imputoitu)
imp1b	kysymykseen 1b) liittyvä imput.mja (0, jos aito havainto; 1, jos imputoitu)
imp3	kysymykseen 3 liittyvä imput.mja (0, jos aito havainto; 1, jos imputoitu)
imp4	kysymykseen 4 liittyvä imput.mja (0, jos aito havainto; 1, jos imputoitu)
impkhka	kysymykseen 6 liittyvä imput.mja (0, jos aito havainto; 1, jos imputoitu)
kala1	kysymys nro 4, kohta 1
kala2	kysymys nro 4, kohta 2
kala3	kysymys nro 4, kohta 3
kala4	kysymys nro 4, kohta 4
kala5	kysymys nro 4, kohta 5
kalast1	1=(kalastus=1 tai kalastus=2), eli kotitaloudessa kalastettiin ja 0=(kalastus=3 tai kalastus=4) eli kotitaloudessa ei kalastettu
kalast2	1=((kalastus=1 tai kalastus=2) ja (kala3>0 tai kala4>0 tai kala5>0)), 2=((kalastus=1 tai kalastus=2) ja (kala3=0 ja kala4=0 ja kala5=0)), 3=(kalastus=3 tai kalastus=4)
kalastus	kysymys nro 2; saa arvot 1-4
khal1-khal9	kalastaneiden henkilöiden määrä/kotitalous vesialueittain
khkap1-khkap8	kalastaneiden henkilöiden määrä/kotitalous pyydyksittäin
khlkm	kotitalouden kalastaneiden henkilöiden määrä (kysymys 3)
khryh	kotitalouden kalastaneiden henkilöiden määrä (kysymys 4)
kktit	1, jos kotitaloudessa on kalastaneita, 0, jos ei ole
kmi1-kmi6	kotitalouden kalastaneiden miesten määrät ikäluokittain
kmiy	kalastaneet miehet yhteensä
kna1-kna6	kotitalouden kalastaneiden naisten määrät ikäluokittain
knay	kalastaneet naiset yhteensä
knm1-knm6	kalastaneet naiset ja miehet ikäluokittain
kont	1=(kontakti=1 tai 2), 0=(kontakti=3)
kontakti	montako kontaktia otettiin (saa arvot 1-3)
kotit	aina 1
kotitalo	montako hlöä kotitalouteen kuuluu (kysymys 1 a))
kthl4	kotitalouteen kuuluvien henkilöiden määrä, saa arvot 1-5 (jos >6->5)
laani	lääni, johon asuntokunta kuuluu
ls1-ls21	saaliin määrä/kotitalous lajeittain
lsyh	saaliin määrä yhteensä/kotitalous
makh	merialueella kalastaneiden hlöiden määrä/kotitalous
mh	henkilöiden määrä/osite
mi	18-74 v. /asuntokunta
mi1la1-mi6la6	miesten lukumäärä/kotitalous ikäluokittain ja lääneittäin
mi1-mi6	kotitalouden miesten määrät ikäluokittain
miy	miesten määrä/kotitalous
na1la1-na6la6	naisten lukumäärä/kotitalous ikäluokittain ja lääneittäin
na1-na6	kotitalouden naisten määrät ikäluokittain
nay	naisten määrä/kotitalous
nh	otoskoko (otokseen tulleiden asuntokuntien määrä/osite)
nm1-nm6	naiset ja miehet /kotitalous ikäluokittain
nrhg	RHG-ryhmien koot ositteittain

nro	indeksi
nshg	apumuuttuja; jos kont=1, nshg=vastanneiden määrä 1. tai 2. kontaktilla, jos kont=2, nshg=niiden kotitalouksien määrä, jotka eivät ole vastanneet 1. tai 2. kontaktilla
okal1-okal9	1, jos kotitalous on kalastanut ko. alueella; 0, jos ei ole
okkp1-okkp8	1, jos kotitalous on saanut saalista ko. pyydyksellä; 0, jos ei ole
ols1-ols21	1, jos kotitalous on saanut ko. lajia saaliiksi; 0, jos ei ole
olsyh	1, jos kotitalous on saanut saalista; 0, jos ei ole
onal	0=kotitalous ei kalast. lainkaan, 1=kotitalous kalasti sisävesillä,
	2=kalasti merialueella, 3=kalasti molemmilla
onma	1= kotitalous kalasti merialueella; 0, ei kalastanut
onsv	1=kotitalous kalasti sisävesillä; 0, ei kalastanut
otos	aina 1 (sama kuin kotit)
payp1-payp8	kalastuspäivät/kotitalous pyydyksittäin
plisi	asuinkunnan poimintatn. = (nh*mi)/mh
pop	ositekoko (otokseen tulleiden asuntokuntien määrä/osite)
posite	poimintaosite
pva1p1-pva9p8	kalastuspäivät/kotitalous alueittain ja pyydyksittäin
pva1-pva9	kalastuspäivät/kotitalous alueittain
resp	vastanneet/osite
sp	vastaajan sukupuoli (1=mies ,2=nainen)
sv	vastaajan syntymävuosi
svkh	sisävesillä kalastaneiden hlöiden määrä/kotitalous
vakal	1=vapaa-ajan kalastus, 2=ammattikalastus