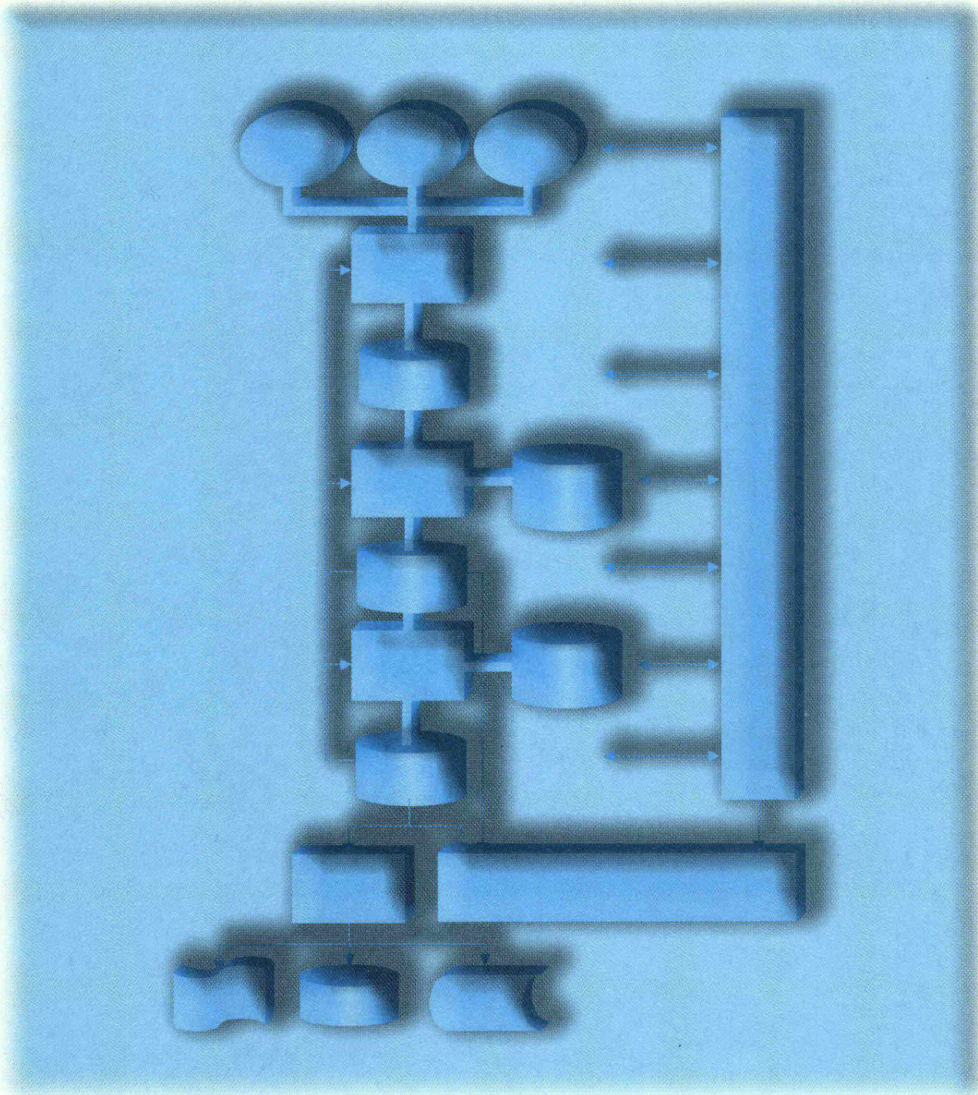


Laatua tilastoissa



Laatua tilastoissa

Tiedustelut:

*Johanna Laiho
Pentti Pietilä
Kari Djerf
(09) 17 341
laatua.tilastoissa@tilastokeskus.fi*

Kansikuva: Hilikka Lehikoinen

Taitto: Marita Potila

© 2002 Tilastokeskus

Tietoja lainattaessa lähteenä on mainittava Tilastokeskus.

*ISSN 0355-2063
= Käsikirjoja
ISBN 952-467-038-0*

Hakapaino Oy, Helsinki 2002

ALKUSANAT

Tilastotieto on monimutkaisen ja monivaiheisen prosessin lopputuote. Tilastojärjestelmien metodologia on kokoelma joko teoriaan tai käytäntöön pohjautuvia menettelytapoja. Niistä tilastoammattilaisten on tehtävä tarkoituksenmukaisia valintoja ja sovelluksia.

Tilastojen tuotantoprosessi on toimintamme ydin. Tilastokeskuksessa on runsaasti tuotantoprosessiin sekä tilastojärjestelmien suunnitteluun ja toteuttamiseen liittyvää tietämystä ja osaamista. Käsillä olevan teoksen keskeinen tavoite onkin koota olemassa oleva tietämys ja kehitetyt hyvät käytännöt yhteiseksi organisaation osaamispääomaksi. Tämän pääoman yhteinen hyödyntäminen ja tuotantoprosessien harmonisointi koko organisaatiossa on kulmakivi tuottavuuden nostamiselle ja tilastojen laadun jatkuvalla kehittämiselle.

Tässä käsikirjassa on neljä pääasiallista asiakokonaisuutta. Aluksi tarkastellaan yleisesti tilastoalan toimintakehikkoa, tilastotointia säätelevää normistoa ja yleisiä laatuvaatimuksia. Toisessa luvussa perehdytään tilastotuotannon ja -tutkimuksen prosessiin ja käsitellään prosessin eri vaiheita edeten toiminnallisessa järjestyksessä tietojenkeruusta ja tietojen käsittelystä tilastojen, tunnuslukujen ja indikaattoreiden tuotantoon. Tavoitteena on läpivalaista prosessin kokonaisuutta ja havainnollistaa samalla sitä, miten työprosessin eri vaiheet nivoutuvat toisiinsa ja täydentävät toinen toisiaan. Tilastotutkimuksen dokumentoinnin tärkeyttä korostaen se on nostettu erilliseksi kokonaisuudeksi kolmannessa luvussa. Dokumentoinnin tulee kuulua jokaisen tilastoasiantuntijan työtapoihin ja saumattomasti jokaiseen työvaiheeseen. Neljännessä luvussa tarkastellaan tilastotiedon julkistamista, levittämistä ja asiakaspalvelua. Luvussa korostetaan tilastotoimen asiakaslähtöisyyden ja asiakaspalvelun jatkuvan kehittämisen tärkeyttä. Käsikirjan käytettävyyttä parantaa loppuun liitetty aakkosellinen asiahakemisto.

Laatua tilastoissa -käsikirja on tarkoitettu kaikille tilastojärjestelmistä ja tilastoprosesseista kiinnostuneille, sekä tilastojen tuottajille että niiden käyttäjille. Laatukäsikirja ei ole koskaan valmis. Käsillä oleva teos on Laatua tilastoissa -käsikirjan ensimmäinen painettu versio. Tekeillä olevaa sähköistä versiota tullaan päivittämään ja kehittämään ottaen huomioon tilastoalan jatkuvasti muuttuva ympäristö ja kehittyvät tietotarpeet.

Miltei kaikkien asiakokonaisuuksien tekstit on laadittu pienryhmissä, joille on nimetty vastuhenkilö. Joihinkin ryhmiin on kutsuttu myös ulkopuolisia kommentaattoreita. Käsikirjan kirjoittamiseen on osallistunut kaiken kaikkiaan 45 kirjoittajaa. Heidän ohella kirjan valmistumiseen on vaikuttanut myös moni Tilastokeskuksen työntekijä toimittaen arvokasta lähdemateriaalia ja opastaen kirjoittajia avainhenkilöiden luo. Lisäksi lukuisa joukko asiantuntijoita on antanut käsikirjan laatimisen eri vaiheissa arvokkaita kommentteja. Heille kaikille lämpimät kiitokset asiantuntevasta työstä ja käyttökelpoisista kommentteista.

Erityisen suuret kiitokset ansaitsee professori Carl-Erik Särndal, jonka jatkuva kannustus ja työn eri vaiheissa esitetyt hyödylliset kommentit olivat ratkaisevan tärkeitä käsikirjan valmistumisen kannalta.

Käsikirjan toimitusprosessi kesti yli kaksi vuotta. Päävastuun teoksen toimittamisesta ovat kantaneet erikoistutkijat Johanna Laiho ja Leena Hietaniemi. Toimitustyön ja koko projektin edistymistä ovat tukeneet toimitusneuvoston muut jäsenet, professori Risto Lehtonen, kehittämisspäällikkö Pentti Pietilä ja kehittämisspäällikkö Kari Djerf. Heille kaikille esitän vilpittömät kiitokset asiantuntevasta ja pitkäjänteisestä työstä Laatusa tilastoissa -käsikirjahankkeen läpiviemiseksi ja laajan asiantuntijajoukon innostamiseksi tilastojen laadun kaikinpuoliseen kehittämiseen.

Helsingissä, huhtikuussa 2002

Heli Jeskanen-Sundström
pääjohtaja
Tilastokeskus

Kirjoittajat

Timo Alanko	Marko Nokkala
Jaana Andelin	Veijo Notkola
Petteri Baer	Pauli Ollila
Nicola Brun	Hannele Orjala
Timo Byckling	Maari Paasilinna
Kari Djerf	Kirsi Palteisto
Petri Godenhjelm	Rami Peltola
Riitta Harala	Pentti Pietilä
Tuula Hausmann	Kirsti Pohjanpää
Leena Hietaniemi	Riitta Poukka
Taina Koivula	Marjo Pyy-Martikainen
Päivi Koivuniemi	Anna-Leena Reinikainen
Timo Koskimäki	Tuomas Rothovius
Riitta Koskinen	Kaija Ruotsalainen
Vesa Kuusela	Antti Suoperä
Jaakko Laakso	Marja Tammilehto-Luode
Seppo Laaksonen	Ismo Teikari
Johanna Laiho	Ari Tyrkkö
Hannu Lehtimäki	Virpi Viertola
Risto Lehtonen	Kai Vikki
Markku Lindqvist	Vesa Virtanen
Kari Molnar	Paavo Väisänen
Pekka Myrskylä	

SISÄLLYS

Alkusanat	3
Sisällys	5
1 Yleinen toimintakehikko	6
1.1 Tilastotoimen normisto	6
1.1.1 Lainsäädäntö	6
1.1.2 Ammattietiikka	6
1.1.3 Tietosuoja	8
1.2 Tilastojen laatu	12
1.2.1 Kokonaislaadunhallinta (TQM)	12
1.2.2 Tilastojen laadun arviointi	15
1.2.3 Suomen Virallinen Tilasto (SVT)	18
1.2.4 Suomen Virallisen Tilaston laatukriteerit	20
2 Tilastotutkimus ja -tuotanto	24
2.1 Tilastotutkimuksen eli surveyn määrittely	24
2.2 Luokitukset, käsitteet ja määritelmät sekä harmonisointi	30
2.3 Kehikot ja peittävyys	34
2.4 Hallinnolliset rekisterit ja tilastorekisterit	37
2.5 Otantamenetelmät	45
2.6 Tiedonkeruutavat henkilö pohjaisessa ja yrityspohjaisessa tiedonkeruussa	48
2.7 Tiedonkeruulomakkeiden suunnittelu ja testaus	54
2.8 Vastauskato ja siihen vaikuttaminen	57
2.9 Tilastollinen editointi ja imputointi	60
2.10 Painokertoimen muodostus ja katovirheen oikaisu	64
2.11 Tilastollinen estimointi ja analyysi	67
2.12 Tilastotiedon esittäminen	72
2.13 Aikasarjat ja kausitasoitus	78
2.14 Indikaattorit ja indeksit	82
3 Dokumentointi	89
3.1 Tilastotutkimuksen dokumentointi	89
3.2 Arkistointi	92
4 Julkistaminen, jakelu ja asiakaspalvelu	96
4.1 Julkistaminen	96
4.2 Tilastotiedon levittäminen	99
4.3 Tilastotuotteet	101
4.3.1 SVT:n Tilastojulkaisut ja tuoteselostesuositus	101
4.3.2 Sähköiset tuotteet	105
4.3.3 Rääätälöidyt tuotteet ja palvelut	108
4.4 Laatu asiakaspalvelussa	110
Hakemisto	114

1 YLEINEN TOIMINTAKEHIKKO

1.1 TILASTOTOIMEN NORMISTO

1.1.1 Lainsäädäntö

Tilastolaki (62/1994) on valtion tilastotoimen yleislaki, jota sovelletaan koko valtion tilastotoimeen. Lain tarkoituksena on luoda edellytykset tehokkaalle ja yhteisölliselle tilastotoimelle, joka ottaa huomioon niin kansalliset kuin kansainväliset tietotarpeet. Tilastolaki määrittää selkeän ja luottamusta herättävän toimintapolitiikan, jota sovelletaan tietojen antajiin ja käyttäjiin. Tilastolaki sisältää säännöksiä tilastojen laatimisen eri vaiheista ja kiinnittää erityistä huomiota tietojen keruun rationalisointiin, tiedonantajasuhteisiin, tilastoeettisten periaatteiden toteuttamiseen ja tietosuojaan.

Tilastolain mukaan tilastojen laadinnassa on ensisijaisesti hyödynnettävä muussa yhteydessä kerättyjä tietoja kuten hallinnollisia aineistoja. Tietojen antajilta voidaan suoraan pyytää vain ne tiedot, joita ei voi saada muulla tavoin. Lain mukainen tiedonantovelvollisuus koskee laissa yksilöidyssä laajuudessa valtion viranomaisia, julkisia palvelujen tuottajia, elinkeinon-, liikkeen- ja ammatinharjoittajia, valtion ja kuntien liikelaitoksia, kuntia ja kuntayhtymiä sekä voittoa tavoittelemattomia yhteisöjä. Uudesta ja muuttuneesta tietojen keruusta on neuvoteltava tiedonantovelvollisten edustajien kanssa hyvissä ajoin ennen tiedonkeruun aloittamista. Tiedonantovelvollisille (lukuun ottamatta valtion viranomaisia) on myös annettava tilastomuotoinen palautetieto.

Tilastolaki edellyttää myös hyvän tilastotavan ja tilastoalalla yleensä sovellettavien kansainvälisten suositusten ja menettelytapojen noudattamista. Näillä säännöksillä viitataan tilastoeettisten periaatteiden noudattamiseen.

Tilastolakiin sisältyy pykälä tilastojen laadusta. Sen mukaan tilastojen on oltava mahdollisimman luotettavia ja niiden on annettava oikea kuva yhteiskuntaoloista ja niiden kehityksestä. Tilastot on julkistettava mahdollisimman pian niiden valmistumisen jälkeen. Tietojen vertailukelpoisuuden varmistamiseksi tilastoja laadittaessa on mahdollisuuksien mukaan käytettävä yhdenmukaisia käsitteitä, määritelmiä ja luokituksia.

Euroopan yhteisön tilasto-ohjelman mukaan laadittaviin tilastoihin sovelletaan neuvoston asetusta (EY) No 322/97 yhteisön tilastoista. Kansallisen tilastolain ja yhteisön tilastoasetuksen keskeiset periaatteet ovat samansuuntaisia.

1.1.2 Ammattietiikka

Soveltamisala ja päämäärä

Tilastoalan ammattietiikka on keskeinen perusta tilastoviranomaisten toiminnalle. Se on tilastolainsäädännön ja sen soveltamisen perusta. Se vaikuttaa myös tilastoviranomaisten toimintaperiaatteisiin ja -tapoihin.

Tilastoalan ammattieettisiä periaatteita on koottu Kansainvälisen Tilastoinstituutin (ISI) vuonna 1985 hyväksymään ammattietiikkajulistukseen, jonka pohjalta Tilastokeskus on laatinut oman ammattieettisen oppaan. Euroopan neuvosto on myös hyväksynyt henkilötietoja koskevan tilastoalan tietosuojasuosituksen No. R (97) 18, joka tietosuojalainsäädännön lisäksi pohjautuu tilastoalan ammattieettisille periaatteille.

Periaatteet

Tilastokeskuksen ammattieettisen oppaan mukaan tilastontekijän arvopäämäärä on puolueettoman, monipuolisen ja luotettavan kuvan antaminen yhteiskunnasta tietojen kohteiden tietosuoja kunnioittaen (Tilastokeskus, 1993). Toimintaympäristön jatkuva ja nopea muuttuminen aiheuttaa uudenlaisia eettisiä ongelmia ja vanhat eettiset ongelmat nousevat usein uudelleen keskusteluun. Ammattieettisen oppaan tarkoituksena on helpottaa eettisten ongelmien ratkaisua.

Suosituks

Ammattieettisen oppaan mukaan Tilastokeskus ja sen työntekijät soveltavat seuraavia ammattieettisiä periaatteita:

- **Puolueettomuus:**
Tilastot laaditaan noudattaen objektiivisuutta, tieteellisyyttä ja tasapuolisuutta. Tilastontekijät päättävät itse tilaston laadinnassa käytettävistä tekniikoista, menetelmistä ja määritelmistä. Näissä asioissa he eivät suosi mitään yhteiskunnallista ryhmää, eivätkä toimi vallassa olevien vaatimuksesta ammattieettisten periaatteiden vastaisesti.
- **Luotettavuus:**
Tilastot kuvaavat mahdollisimman paikkansapitävästi ja täsmällisesti niitä olosuhteita ja asioita, joita niillä halutaan kuvata.
- **Avoimuus:**
Yleinen yhteiskuntatilasto julkistetaan tehokkaasti. Tilastojen perustietojen antajille kerrotaan tietojen antoperuste, käyttötarkoitus, tärkeimmät tilaston laatimisessa sovellettavat menettelytavat sekä tietojen luovutusperusteet.
- **Tietosuoja:**
Tietoja käsitellessään ja säilyttäessään Tilastokeskuksen työntekijä menettelee aina niin, ettei kenenkään yksityisyyttä tai liike- ja ammattisalaisuutta vaaranneta.
- **Relevanssi (asiaankuuluvuus):**
Tilastot kuvaavat niitä alueita, joista yhteiskunta todella tarvitsee tietoja. Tilastojen tarjonta noudattaa kysyntää ja sen muutoksia. Tilastojen laadinnassa tarpeettomia tietoja ei kerätä eikä tarpeettomia tilastoja laadita.
- **Tehokkuus:**
Virallinen tilasto laaditaan taloudellisesti ja tehokkaasti, myös perustietojen antajien vastaustaakkaa minimoiden.

Vastaavat periaatteet sisältyvät kaikkien EU -maiden tilastovirastoja koskevana Euroopan yhteisön tilastoasetukseen.

Kirjallisuus

Tilastokeskus. 1993. *Toimi oikein tilastoalalla. Tilastokeskuksen ammattieettinen opas*. Käsikirjoja 30.

Tilastokeskus. 2002. *Tilastokeskuksen ammattieettinen opas*. (Uudistetun painoksen luonnos).

Tilastolaki (L 62/1994, muut. 21.4.1995/734, 18.12.1998/1039 ja L 1293/2001).

1.1.3 Tietosuoja

Sovelletamisala ja päämäärä

Tietosuojan tarkoituksena on varmistaa mm., ettei tarpeettomia tietoja kerätä ja että salassapidettävien tietojen paljastuminen ulkopuolisille estetään tietojen käsittelyn kaikissa vaiheissa. Tilastotoimen tietosuojaperiaatteilla pyritään varmistamaan kattavien ja luotettavien tietojen saaminen ylläpitämällä luottamuksellisia suhteita tiedonantajiin. Tietosuojavaatimukset rajaavat sen, mitä tietoja voidaan kerätä ja miten tietoja voidaan käsitellä sekä missä muodossa ne voidaan julkaista. Tietosuoja vaikuttaa siten tilastojen laatimisessa sovellettaviin menettelytapoihin ja julkaistavien tilastojen sisältöön. Tilastotoimessa tietosuoja on toiminnan keskeinen peruseriaate ja osa tilastoviranomaisten ulkoista imagoa.

Keskeiset säännökset ja periaatteet

Tilastotointa koskevat tietosuojasäännökset sisältyvät tilastolakiin (62/1994), neuvoston asetukseen (EY) 322/97 yhteisön tilastoista sekä henkilötietolakiin (523/1999). Lisäksi Tilastokeskus on antanut useita tietosuojaa koskevia sisäisiä ohjeita (esimerkiksi Ohjeet taulukkomuotoisten yritystietojen (TK-00-888-00) ja henkilötietojen (TK-00-314-02) suojaamiselle).

Tilastotarkoituksiin saadut tiedot ovat pääsääntöisesti salassapidettäviä. Julkisia ovat valtion ja kunnallisten viranomaisten toimintaa ja julkisten palvelujen tuottamista kuvaavat muun lainsäädännön mukaan julkiset tiedot sekä yritys- ja toimipaikkarekisteriin sisältyvät tiedot.

Tilastotarkoituksiin saatuja salassapidettäviä tietoja voidaan luovuttaa ainoastaan tieteelliseen tutkimukseen tai tilastollisiin selvityksiin pääsääntöisesti sellaisessa muodossa, ettei henkilöitä voida suoraan tai välillisesti tunnistaa. Muiden tilastoyksiköiden kuin henkilöiden tunnistetietoja ei saa myöskään luovuttaa. Tiedot iästä, sukupuolesta, koulutuksesta ja ammatista voidaan poikkeuksellisesti luovuttaa tieteelliseen tutkimukseen ja tilastollisiin selvityksiin tunnistetietoineen. Tietojen luovuttaminen käytettäväksi hallinnollisessa päätöksenteossa tai muussa vastaavassa asian käsittelyssä on erityisesti kielletty. Tietojen luovuttamisessa sovellettuihin menettelytapoihin ja periaatteisiin sovelletaan pääjohtajan

vahvistamaa ohjetta käyttölupien myöntämisestä Tilastokeskuksen perusaineistoon.

Tietosuojavaatimusten huomioon ottaminen edellyttää koko tietojen käsittelyprosessin huolellista suunnittelua etukäteen tietojen keruusta niiden julkaisemiseen. Tiedonantajilta voidaan kerätä vain ko. ilmiöalueen kuvaamisen kannalta välttämättömät tiedot. Otantaa ja muita tietojen keruuta keventäviä menetelmiä on käytettävä aina, kun se on mahdollista. Tietojen käsittely on toteutettava niin, ettei kenenkään yksityisyyttä tai liike- ja ammatillisaisuutta vaaranneta. Tilastoja laativan viranomaisen on huolehdittava siitä, että tiedot on asianmukaisesti suojattu luvaton käsittelyä ja paljastumista vastaan.

Tiedot voidaan julkaista ainoastaan sellaisessa muodossa, ettei tietojen kohdetta koskevia salassapidettäviä tietoja paljastu. Salassapidettäviä tietoja voidaan poikkeustapauksessa julkaista asianomaisen tietojen kohteen suostumuksella. Salassapidettävien tietojen paljastuminen tietoja julkaistaessa voidaan estää käyttämällä soveltuvia tilastollisia tietosuojamenetelmiä (Hänninen, 1997). Menetelmää valittaessa tulisi huomiota kiinnittää tietojen laatuun (mm. tietojen arkaluonteisuus) sekä erilaisten menetelmien vaikutukseen tilastojen sisältöön.

Suosituks

Tilastolain peruseriaate on, että kaikki tilastotarkoituksiin saadut tiedot ovat salassapidettäviä lukuun ottamatta edellä mainittuja poikkeuksia.

Tarkasteltaessa sitä, mitä tietoja voidaan luovuttaa viraston ulkopuolelle, on ensimmäiseksi erotettava toisistaan taulukkomuotoiset tiedot ja yksikötason tiedot (mikrotiedot).

Suosituksia mikrotietojen kohdalla

Yksikötason tietoja voidaan luovuttaa ainoastaan tieteelliseen tutkimukseen tai tilastollisiin selvityksiin ja luovuttaminen tapahtuu aina käyttöluvamenettelyn kautta. Käyttöluva voidaan hakemuksesta myöntää joko yksittäiselle tutkijalle tai tutkimuslaitokselle tai vastaavalle. Hakemuksessa on yksilöitävä kaikki aineiston käsittelyyn osallistavat ja lupa annetaan määräajaksi. Käyttöluva myönnetään aina yksilöityyn tutkimustarkoitukseen.

Luovutettava mikroaineisto tulee olla henkilöitä koskevien tietojen osalta sellainen, että yksittäiset henkilöt eivät siitä ole tunnistettavissa. Tämä tarkoittaa sitä, että sen lisäksi että henkilöiden tunnistetiedot on aineistosta poistettu, aineistosta poistetaan myös välillisen tunnistamisen mahdollistavat muuttujat tai aineistoa muuten käsitellään siten, että tunnistaminen ei enää ole mahdollista. Kokonaisaineistoja ei luovuteta, vaan ainoastaan otosaineistoja. Kokonaisaineistolla tarkoitetaan esimerkiksi kaikkia tietyn tutkinnon suorittaneita tai kaikkia tietyllä alueella asuvia.

Tunnistamisen estämiseksi voidaan käyttää erilaisia menetelmiä, joista yksinkertaisimpia ovat luokitustarkkuuden karkeistaminen, todellisten arvojen korvaaminen luokkakeskiarvolla ja tunnistamista helpottavien muuttujien poistaminen. Mikroaineistojen osalta kattavien, joka tilanteessa soveltuvien tietosuojajohjeiden

antaminen on mahdotonta, koska tarvittavat toimenpiteet riippuvat aina monesta eri osatekijästä kuten aineiston koosta, muuttujien määrästä ja luokitustarkkuudesta. Myös muuttujien yhdistelmät saattavat aiheuttaa tunnistamisongelman. Esimerkiksi henkilön asuinkuntatieto ei sinänsä riitä tunnistamiseen, mutta kun asuinkuntatietoon lisätään ammattitieto, henkilön tunnistaminen ainakin harvinaisempien ammattien osalta on jo mahdollista. Siten tietosuojan varmistaminen mikroaineistojen luovutuksissa on joka kerta tarkasteltava tapauskohtaisesti.

Yrityksiä koskevat tiedot

Muiden kuin oman yksikön yksikkötason tietojen käyttö on myös Tilastokeskuksen sisällä luvanvaraista. Tietojen käyttö lupaa anotaan yksikön johtajalta sisäisellä käyttöluvahakemuslomakkeella.

Suosituksia taulukkomuotoisen tiedon kohdalla

Tilastotaulukot voivat olla frekvenssitaulukoita ja määrätaulukoita (*tables of magnitudes*). Frekvenssitaulukot kertovat kuhunkin taulukon soluun kuuluvien tapausten määrän ja määrätaulukoissa esitetään numeerisia arvoja, kuten esimerkiksi tulojen keskiarvoja tai tulosummia markoissa. Tietosuojatoimenpiteet eri tyyppisten taulukoiden kohdalla voivat vaihdella.

Taulukon (tai sen solun) arkaluonteisuuden määrittelemiseen on periaatteessa olemassa kaksi kriteeriä: toinen on tapausten määrä solussa ja toinen solun tapausten niin sanottu dominanssi.

Henkilötietoihin perustuvat tilastotaulukot tulisi laatia niin, että taulukon solut eivät sisällä tarpeettoman pieniä arvoja. Taulukkoaineistojen sisältämät henkilöihin, yksittäisiin perheisiin tai asuntokuntiin kohdistuvat tietosuojariskit on aina arvioitava ennen tietojen julkaisemista ja/tai luovuttamista.

Taulukon muuttujien arkaluonteisuus määrittelee ensisijaisesti tarvittavien tietosuojatoimenpiteiden vahvuuden. Taulukon tietojen suojaustarpeen arvioimiseen vaikuttavat muuttujien arkaluonteisuuden lisäksi mm. tapausten määrä taulukon eri soluissa, tarkasteltavan perusjoukon suuruus, muuttujien määrä ja tilastointialueen koko ja sijaintitarkkuus.

Ykköstapaukset sinänsä eivät välttämättä vaaranna tietosuojaa. Esimerkiksi tieto siitä, että postinumeroalueella on yksi koneinsinööri ei sinänsä ole arkaluonteinen, mutta lisättäessä taulukkaan pääasiallinen toiminta-muuttuja tilanne saattaa muuttua (postinumeroalueella on yksi koneinsinööri, joka on työtön).

Tiedon suojaaminen voidaan tehdä solukohtaisesti tai luokkakohtaisesti (esimerkiksi rivi kokonaisuudessaan lukuun ottamatta luokassa olevien tapausten määrää). Suojaus voidaan myös tehdä joko arvojen peittämisellä tai luokitusta muuttamalla. Käytettävissä on myös erilaisia tietosuoaohjelmia, joita on testattu (mm. ARGUS-ohjelma) tai kehitetty Tilastokeskuksessa (pienalue-tilastojen tietosuojamenetelmät). Taulukkomuotoisten henkilöitä koskevien tietojen julkaisemisen osalta Tilastokeskuksella on yleisohjeet (TK-00-314-02) ja yritystietojen osalta ohjeet (TK-00-888-00), jotka helpottanevat tapauskohtaista tarkastelua.

Joskus taulukkomuotoinen tieto lähenee mikroaineistoa esimerkiksi muodostettaessa moniulotteinen taulukko tarkoilla luokituksilla. Tällöin on tapauskohtai-

sesti harkittava, onko aineistoa tarkasteltava yksikkötason aineistona, jonka luovuttamiseen tarvitaan käyttöluupa.

Muut suositukset

Jokaiselle Tilastokeskuksen keräämälle aineistolle on määriteltävä suojausluokka sekä suorassa että epäsuorassa tiedonkeruussa. Suojausluokan määrääminen on pakollinen arkistonmuodostumissäännön (AMS) yhteydessä.

Kirjallisuus

Lait

Henkilötietolaki (L 523/1999)

Laki viranomaisten toiminnan julkisuudesta (L 621/1999) ja asetus viranomaisten toiminnan julkisuudesta ja hyvästä tiedonhallintatavasta (L 1030/1999).

www.om.fi

Tilastolaki (L 62/1994, muut. 21.4.1995/734, 18.12.1998/1039 ja L 1293/2001).

Väestölaskentalaki (L 154/1938) myöhempine muutoksineen.

Kansainväliset asetukset ja ohjeet

Euroopan neuvoston henkilötietoja koskeva suositus. 1997. *Recommendation No.R (97) 18 and Explanatory Memorandum of the Committee of Ministers to Member States Concerning the Protection of Personal Data Collected and Processed for Statistical Purposes*. (Adopted by the Committee of Ministers on 30 September 1997 at the 602nd meeting of the Ministers' Deputies).

[www.coe.fr/dataprotection/rec/r\(97\)18e.htm](http://www.coe.fr/dataprotection/rec/r(97)18e.htm).

Euroopan parlamentin ja neuvoston direktiivi yksilöiden suojasta henkilötietojen käsitteilyssä 95/46/EY. ("EU:n tietosuojadirektiivi").

ISI - International Statistical Institute. 1985. *Declaration on Professional Ethics*. Adopted: August 1985. www.cbs.nl/isi/ethics.htm

Komission päätös (97/281/EY) Eurostatin asemasta yhteisön tilastojen tuottamisessa.

Neuvoston asetus (EY) n:o 322/97 yhteisön tilastoista.

Neuvoston asetus (EY) n:o 1588/90 salassapidettävien tilastotietojen luovuttamisesta Euroopan yhteisöjen tilastotoimistolle.

YK:n Tilastokomission hyväksymät virallisen tilaston peruseriaaatteet.1994. E/CN.3./1994/18, E/1994/29.

Tilastokeskuksen ohjeet ja muut kansalliset ohjeet

Harala, R., Orjala, H. & Öörni, S. 1996. *Henkilötilastot-yksikön asiakasrahoitteisia erityisselvityksiä koskevia eettisiä ja tietosuojaratkaisuja*. Muistio 28.5.1996.

Hänninen, M. 1997. *Tilastolliset tietosuojamenetelmät ja niiden käyttö*. Katsauksia 1997/3. Tilastokeskus. Helsinki: Yliopistopaino.

- Salassa pidettävien tietojen ja asiakirjojen turvaluokittelu.* JHS 147. Juhta 28.11.2000.
www.intermin.fi/juhta/suosituksset/jhs147.doc
- Tietoaaineiston luokitus ja käsittely.* 1995. Muistio 31.5.1995.
www.vn.fi/vm/suomi/muuta/vahti/tilurap2.htm
- Tilastokeskus. 1993. *Tilastokeskuksen ammattieettinen opas.* Toimi oikein tilastotalalla. Käsikirjoja 30. Helsinki.
- Tilastokeskus. 1994. *Ohje käyttö lupien myöntämisestä Tilastokeskuksen perusaaineistoon.* TK-00-854-94. 20.5.1994.
- Tilastokeskus. 1996. *Tietojen suojaus Tilastokeskuksessa.* Tietosuojaohje. TK-00-1603-96. 17.9.1996).
- Tilastokeskus. 2000. *Henkilötietolain soveltamisohje.* TK-00-578-00. 15.5.2000.
- Tilastokeskus. 2000. *Julkisuuslain mukaisen Tietojärjestelmäselosteen laadintaohje Tilastokeskuksessa.* TK-00-696-00. 17.5.2000.
- Tilastokeskus. 2000. *Ohjeet taulukkomuotoisen yritystiedon suojaamiselle.* TK-00-888-00, 21.6.2000.
- Tilastokeskus. 2000. *Tilastolain (62/1994) soveltamisohje Tilastokeskuksessa.* TK-00-579-00. 15.5.2000.
- Tilastokeskus. 2002. *Ohjeet taulukkomuotoisten henkilötietojen suojaamiselle.* TK-00-314-02. 1.3.2002.
- Tilastokeskus. 2002. *Tilastokeskuksen ammattieettinen opas.* (Luonnos).
- Valtiovarainministeriön päätös eräiden tilastotietojen salassapitämisestä ennen niiden julkistamista (1994). 27.1.1994/36.

1.2 TILASTOJEN LAATU

1.2.1 Kokonaislaadunhallinta (TQM)

Soveltamisala ja päämäärä

Laatu on perinteisesti Tilastokeskuksessa merkinnyt tilastojen laatua eli lähinnä tuotelaatua, jota käsitellään tarkemmin luvussa 4.3.1 SVT:n Tilastojulkaisut ja tuoteselostesuositus. Tuotelaatunäkemyksen täydennykseksi on määritelty Tilastokeskuksen strateginen kokonaislaadunhallinta. Sillä pyritään organisaation suorituskyvyn erinomaisuuteen ja sen eräänä keskeisenä tavoitteena on saavuttaa laatupalkintotason määrittelemä laatu. Laatupalkintokehikko on samalla moderni määritelmä kokonaislaadunhallinnalle.

Kokonaislaadunhallinnassa tarkastelun kohde ei ole enää pelkästään yksittäisen tuotteen, sen tuottamisprosessin tai aktiviteetin laatu, vaan koko tilastojen tuotantojärjestelmä, muut ydinprosessit ja edelleen koko johtamisjärjestelmä, henkilöstö, kumppanuudet ja resurssit. Näiden asioiden hallinnalla voidaan luoda laatua korostava toimintajärjestelmä ja parantaa myös lopputuotteiden laatua.

Seuraavat tunnuspiirteet luovat perustan hyvälle suorituskyvyllle:

- **Tuloshakuisuus**; odotuksiin ja tavoitteisiin vastaaminen, aikaansaannokset ratkaisevia
- **Asiakassuuntautuneisuus**; asiakastarpeet muovaavat tuotteita, palveluja ja toimintatapoja
- **Johtajuus ja toiminnan päämäärätietoisuus**; johdon henkilökohtainen esimerkkitoiminta, suunnan näyttäminen ja työympäristöstä ja ilmapiiristä huolehtiminen
- **Prosesseihin ja tosiasioihin perustuva johtaminen**; työnkulkujen selvittäminen, kehittäminen ja ohjaus sekä luotettavat ja kattavat tiedot
- **Henkilöstön kehittäminen ja osallistuminen**; organisaation tavoitteiden ja henkilöstön intressien kytkentä
- **Jatkuva oppiminen, parantaminen ja innovatiivisuus**; kyky kehittää uutta ja muista erottuminen
- **Kumppanuuksien kehittäminen**; painopisteen valinta ja kokonaisuuksien hallinta vaatii luotettavia kumppanuuksia toimittaja- ja asiakasyhteistyössä
- **Yhteiskunnallinen vastuu**; esimerkillinen yrityskansalaisuus.

Periaatteet

Laadunhallintajärjestelmä

Laadunhallintajärjestelmä on yhteisesti sovittu viitekehys organisaation toiminnalle, joka hahmottaa ulkoista ja sisäistä toimintaympäristöä ja organisaation omaa roolia. Laadunhallintajärjestelmän tavoitteena on luoda parhaiden käytäntöjen systemaattinen haku-, siirto-, ylläpito- ja kehittämismenettely omien ja asiakkaiden etujen turvaamiseksi. Se pyrkii myös saattamaan halutut toimintamallit läpinäkyviksi ja levittää niitä koko organisaation tietoisuuteen.

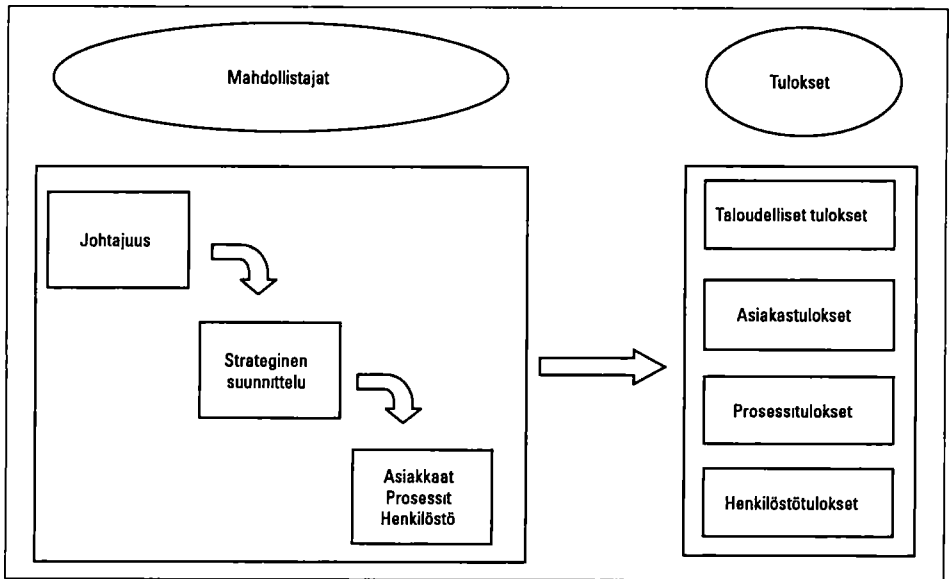
Lyhyesti sanoen laadunhallintajärjestelmä voidaan rakentaa seuraavasti:

- kehitä hyvä toimintatapa
- dokumentoi se
- toimi sen mukaan
- paranna sitä (Moisio ja Ritola, 2000).

Laadunhallintajärjestelmä pyrkii systematisoimaan ihmisten yhteistyötä ohjaavat toimintatavat kaikkien edun mukaisiksi ja edistää organisaation asiakas- ja sidosryhmäkeskeisten ajattelu- ja toimintamallien parantamista. Se edistää viestintää asiantuntijalta toiselle, mikä on olennaista oman organisaation, asiakkaiden ja muiden keskeisten sidosryhmien kannalta.

Toisin sanoen laadunhallintajärjestelmällä luodaan edellytykset rakentaa ja vakiinnuttaa organisaatioon suunnitelmallisuutta, hyväksi koettuja toimintamalleja, seurantaa ja jatkuvaa parannusta. Se pienentää organisaation niin kutsuttua aukkotautia eli katkoksia organisaation hierarkiasuunnassa ja horisontaalisessa funkti-

Kaavio 1.1 Kokonaislaadun hallintajärjestelmä



oiden välisessä asiakassuunnassa ja parantaa keskusteluyhteyttä ja hiljaisen tiedon levittämistä organisaation eri osapuolien välillä.

Pelkistettynä hyvä laadunhallintajärjestelmä auttaa tunnistamaan mitä pitää tehdä (oikeita asioita) ja antaa kokemuksen kautta hioutuneita vinkkejä siitä, miten kannatta tehdä (oikeat asiat oikein) ja tuottaa arvokasta palautetietoa, josta suorituskykyä voidaan edelleen hioa.

Laadunhallintajärjestelmää/laatua voidaan tarkastella organisaation tasolla, ydinprosessien tasolla tai yksittäisten prosessien tasolla seuraavasti:

Organisaatiotaso

Organisaatiotasolla haetaan laajoja, strategisia ja koko organisaatioon vaikuttavia parannushankkeita ja toteutetaan niitä. Koko organisaation tasoisen laadunhallinnan perusta on laatupalkintomalli ja sen avulla tehtävä arviointi vahvuuksista ja parantamisalueista. Tasapainoinen tulokortti (*balanced scorecard*, BSC) on strategisen suunnittelun ja sen toteutuksen sekä viestinnän väline. Niiden yhdistelmää kuvaa kokonaislaadun hallintajärjestelmä (kaavio 1.1).

Tuotantoydinprosessitaso

Tuotantoydinprosessitasolla kehitetään ja ylläpidetään yhtenäistä, kustannustehokasta ja sujuvaa tuotantoprosessia ja siihen liittyvää logistiikkaa yhtenä ydinprosessina. Prosessin tunnistaminen, määrittely, mallintaminen, dokumentointi, arviointi, mittarointi, laadunvarmistus ja parantaminen ovat tässä työssä oleellisia tekijöitä. Operationaalisina tekijöinä ydinprosessiin vaikuttavat muun muassa tilastometodit, välineet, aineistot, ohjelmistot, metatiedot, manuaalit ja käsikirjat.

Tilastoprosessitaso

Tilastotutkimukset ovat tuotantoydinprosessin alaprosesseja. Tilastoprosessitasolla tutkitaan toimiiko yksittäisen tilastotutkimuksen tuotantoprosessi optimaalisella tavalla. Yksittäisten prosessien parannukset eivät kuitenkaan riitä, sillä huippulaatuorganisaatioksi pääseminen edellyttää myös yllä mainittujen ylempien ja rinnakkaisten tasojen kehittämistä.

Suosituks

- Kokonaislaadunhallinnan soveltamisen kulmakivenä on välttää mahdollisimman pitkälle oman ja erillisen laatumaailman syntyminen. Pyrkimyksenä on luoda integroitu toiminnan (laadun) hallintajärjestelmä, jossa hyvät laatuperiaatteet on integroitu tilastontuotantjärjestelmiin.
- Merkittävät ja koko organisaatiota koskettavat laatutavoitteet sisällytetään strategisiin tavoitteisiin, jolloin ne tulevat osaksi toiminnan suunnittelua ja seuranta.
- Keskeisenä kokonaislaadunhallinnan sovellusvälineenä ja mittaristona käytetään tasapainotettua tuloskorttia.
- Käytännössä ISO 9000:2000 standardisarja luo puitteet laatu järjestelmän kehittämiseksi yhteensopivina laatu palkinto- ja BSC -kehikkojen kanssa.
- Kokonaislaatuajattelussa organisaation toiminta mallitetaan ydinprosessien ja niiden alaprosessien avulla.
- Hyvin motivoitunut ja osaava henkilöstö ymmärtää ja hyväksyy organisaation strategiset tavoitteet, mikä parantaa oleellisesti suorituskykyä.
- Muutoksen toteuttaminen edellyttää suuntaa, maaperää ja välineitä. Johto antaa suunnan ja edesauttaa sopivien (laatu)välineiden hankintaa.

Kirjallisuus

- European Foundation for Quality Management. 1999. *The EFQM Excellence Model*. Suomenkielinen käännös. Laatu keskus.
- Moisio, J., Ritola, O. 2000. *ISO 9000: 2000 ja menestyksen avaimet – Vinkejä pohdiskelijoille*. Suomen Standardisoimisliitto SFS. Gummeruksen kirjapaino Oy.
- Tilastokeskus. 1987. *Tilastojen laadun kuvaaminen – Ohjeita tuoteselosteiden laatimiseksi*. Käsikirjoja 23. Tilastokeskus: Helsinki.
- Suomen standardisoimisliitto SFS. 2001. *Sata usein esitettyä kysymystä uudesta ISO 9000 sarjasta*. Gummeruksen kirjapaino Oy.

1.2.2 Tilastojen laadun arviointi

Soveltamisala ja päämäärä

Laatua on ISO 9000 määritelmän mukaan kaikki asiakkaan haluamat tuotteen tai palvelun ominaisuudet. Laadunhallinnalla tarkoitetaan niitä organisaation toimenpiteitä, joilla se varmistetaan. Laadukkaiden tuotteiden ja palvelujen valmistaminen ja toimittaminen muodostuu kolmesta osasta:

- **Laadun suunnittelu**
Asiakastarpeiden määrittely, niiden käänö parhaiksi tilastometodologisiksi ratkaisuisi, niiden käytäntöön vienti luomalla tarkoituksenmukaisia rakenteita ja järjestelmiä varmistamaan kustannustehokas keruu, käsittely ja jakelu määrättyllä tarkkuustasolla.
- **Laadun varmistus/tarkistus**
Tuotteen ja tuotannon seuranta käymällä läpi lopputuotteiden laatua suhteessa asiakastarpeisiin ja suunnitelmien mukaisuuteen. Varmistukseen liittyy tuotantoprosessin yksittäisten prosessien, aktiviteettien ja tehtävien läpikäynti, jolla varmistetaan niiden asianmukainen toteutus.
- **Laadun parantaminen**
Havaittujen laatuongelmien poistamiseen tähtävää toimintaa. Se voi olla kokonaan järjestelmää uusivaa, uudelleen suunnittelevaa (korjaavaa tai jatkuvan parantamisen toimintaa).

Laatua joudutaan arvioimaan jatkuvasti, useista näkökulmista ja eri tasoilta ja se voidaan tehdä jollain seuraavista tavoista tai niiden yhdistelmällä:

- **Toteutuskeskeisyys**
 - suunnitelmien mukainen valmistus
 - tehokkuus ja tuottavuus
 - virheettömyys/tarkkuus
- **Tuotokeskeisyys**
 - tuotteen ominaisuudet määrittelevät sen laadun
 - tuotteen suunnittelu korostuu
- **Palvelukeskeisyys**
 - saatavuus ja käytön helppous
- **Asiakaskeskeisyys**
 - kyky täyttää asiakkaan tarpeet, toiveet ja odotukset tai tilaus
- **Arvokeskeisyys**
 - hinta/laatu-suhde
 - hyöty/arvo asiakkaan toiminnalle
 - parasta mitä rahalla saa
- **Markkinakeskeisyys**
 - tuotteen tai palvelun suhteellinen arvo verrattuna vastaaviin tuotteisiin
- **Yhteiskunta- ja ympäristökeskeisyys**
 - yhteiskunnallinen vaikuttavuus
 - ympäristövaikutukset
 - elinkaari.

Laadun arviointia tehdään useimmiten suhteessa tavoitteisiin tai laatustandardiin. Arviointi voidaan tehdä itse, sisäisesti tai ulkopuolisen voimin tai niiden yhdistelmänä. Arviointi voi perustua joko laadullisiin tai määrällisiin mittareihin tai niiden yhdistelmään.

Arviointi perustuu useimmiten kirjalliseen arvioitavan toiminnon menettelykuvaukseen, jota käyttäen selvitetään ovatko toiminnot suunnitelmien mukaiset, toteutetaanko niitä tehokkaasti ja ovatko ne tavoitteiden kannalta tarkoituksenmukaisia. Arvioinnissa selvitetään myös parannustoimenpiteiden tila. Myös menettelytapojen hyvyttä suhteessa parhaisiin käytettävissä oleviin voidaan arvioida.

Asiaa voidaan myös tarkastella tutkimalla tilastoaineiston eri virhelähteitä. Myös riskien vakavuuden ja niiden todennäköisyyden tunnistamisella voidaan selvittää laatua. Saatavalla informaatiolla voivat tilastojen tekijät parantaa laatua ja myös tietoja käyttävät asiakkaat voivat tehdä asianmukaisia tulkintoja tilastoista.

Periaatteet

Tilastojen laadun arvioinnilla tarkoitetaan tällä hetkellä lähinnä lopputuotteen arviointia eli tuotekeskeistä laatua. Tilastojen käyttäjillä tulee olla mahdollisuus arvioida mitkä tekijät aineistossa rajoittavat tilaston laatua. Harvat käyttäjät voivat itse arvioida tuotetun tilaston laatua. Tilastovirastolla on velvollisuus tehdä tarpeelliset laatuarvioinnit ja jakaa tehtyjen arviointien tulokset käyttäjille ajantasaisesti ja helppokäyttöisessä muodossa.

Tilaston laadun arvioinnit hyödyttävät myös tilastovirastoa. Jos virheet voidaan paikallistaa tiettyyn tilastosysteemin vaiheeseen, voidaan korjaavat toimenpiteet tehdä seuraavaan tai vastaavanlaiseen tilastosysteemiin.

Tilaston laadun arvioinnin ajantasaisuus on yhtä tärkeätä kuin tilaston ajantasaisuus. Laatuarvioinnin tulee olla käytettävissä samanaikaisesti tilaston kanssa.

Suosituks

- Minimi arviointi on Suomen Virallisen Tilaston (SVT:n) tuoteselosteen määrittelemä kuvaus. (ks. 4.3.1 SVT:n Tilastojulkaisut ja tuoteselostesuositus).
- Varmistetaan, että tiedon laadun arviointi sisältyy tilastotutkimuksen rakentamisen suunnitteluun, koska tarvittava tieto joudutaan keräämään tuotantoprosessin yhteydessä.
- Tilastotutkimuksen tai tuotteen laadun arvioinnissa on aina harkittava tilaston käyttöä ja käyttäjiä, virhepotentiaalia ja sen merkitystä tilaston käytössä, laadun mahdollista vaihtelua ajassa, laatuarvioinnin kustannuksia suhteessa kokonaiskustannuksiin, laadun parantamisen potentiaalia, tuottavuutta, arviointitulosten helppokäyttöisyyttä käyttäjille ja onko kysymyksessä kertaluonteinen vai jatkuva tilasto.
- Jos kvantitatiivinen laatuanalyysi on jostain syystä mahdoton, on tilasto varustettava asiantuntija-arvioinneilla tai subjektiivisella arvioinnilla.
- Kytetään tilaston sisäisiä ja ulkoisia käyttäjäryhmiä asettamaan laadun arvioinnille uusia tavoitteita. Heitä käytetään arvioinnissa konsultteina.

www.qualitas-fennica.fi/artikkelit/index.html#arviointi (Qualitas Fennican laatu-
arviointin artikkelikokoelma)

www.sfs.fi (Suomen Standardisoimisliitto ry)

1.2.3 Suomen Virallinen Tilasto (SVT)

Soveltamisala ja päämäärä

Virallisen tilaston tehtävänä on palvella yhteiskuntaa ja tuottaa tilastoinformaatiota, jota yhteiskunta tarvitsee toiminnassaan.

Yleistä, kansainväliseksi katsottavaa suositusta virallisen tilaston määrittelyksi ei ole olemassa. Saatavilla olevien tilastojen tulee käytännössä heijastella käyttäjien tarpeita ja prioriteetteja. Useissa maissa tilastoilla on yleisluontoinen määrittely; sen tulee palvella eri käyttäjäryhmiä: kansalaisia, julkista hallintoa, elinkeinoelämää ja tutkimusta.

Periaatteet

Useimmiten virallisen tilaston kokonaisuus voidaan määritellä yhteiskunnan keskeisten ilmiöalueiden/aihealueiden kautta. Aihealueet ryhmitellään sarjoihin. Sarjajako voi perustua YK:n kansainvälisten tilastosarjojen pääjakoon: väestö- ja sosiaalitilastot, taloustilastot ja muut tilastot sekä näiden alla alasarjoihin, aakkoselliseen ryhmittelyyn ja muuhun "vakiintuneeseen" tapaan. Mikäli tilastotuotanto kattaa hyvin eri ilmiöalueet, ovat sarjaryhmittelyt suhteellisen vakiintuneita, vaikka sarjojen sisällä tilastotuotanto voi uudistua paljonkin. Esimerkiksi Ruotsissa on 22 aihealuesarjaa, Iso-Britanniassa 16 ja Australiassa 17. YK:n ryhmityksessä on 20 alasarjaa.

Suomessa virallisen tilaston sarjajulkaisut on ryhmitelty 28 aihealueittaiseen sarjaan. Lisäksi on yleisjulkaisuja. Nämä periaatteissa kattavat kaikki yhteiskunnan ilmiöalueet.

Suosituks

Virallisen tilaston tuottamiseksi tarvitaan "luotettava tiedollinen toimija". Yleensä virallisen tilaston laadinta on viranomaistehtävä, joka on tilanteesta riippuen joko määrätty yhden erityisesti tätä tarkoitusta varten perustetun viranomaisen tehtäväksi tai on hajautettu eri hallinnonalojen viranomaisille.

Suomessa tilastolaki (62/1994) määrittelee kriteerit, joiden perusteella tilastotoimen piiriin luettavat viranomaiset on rajattu. Tilastotoimeen luettavia viranomaisia on nykyisin 18. Tämän lisäksi kattavuuskriteerin perusteella Suomen Virallisen Tilaston tuottajiksi luetaan Kela, ETK ja Suomen Pankki. Viime mainitut tuottavat omilta toimialueiltaan keskeisiä yhteiskuntatilastoja.

Virallisen tilaston tuottamisen peruskriteereitä ovat mm. puolueettomuus, riippumattomuus ja ammatillinen itsenäisyys. Nämä ovat tärkeitä, jotta tilastot voivat saavuttaa yleisen luottamuksen ja palvella yhtäläisesti eri käyttäjäryhmiä. Itsenäisyys ammatillisissa asioissa merkitsee mm. sitä, että tilastoviranomainen asettaa prioriteetit käyttäjälähtöisesti. Tilastojen laadinnassa käytetään kansainvälisiä standardeja, käsitteitä ja luokituksia sekä parhaita mahdollisia metodeja. Itsenäisyyteen kuuluu myös päättää tietojen julkistamistavasta ja -ajankohdasta.

Muita kriteereitä ovat mm. relevanssi, tiedonantajan tietosuojasta huolehtiminen ja toiminnan läpinäkyvyys. Tilastontuottajan tulee toimia kustannustehokkaasti sekä taloudellisuuden että tiedonantajien vastausrasituksen näkökulmasta. Yleisiin periaatteisiin kuuluu myös velvollisuus helpottaa tilastojen tulkintaa sekä reagoida tietojen virheelliseen tulkintaan.

Suomen Virallisen Tilaston (SVT) aihealueittaiset sarjat

1. Asuminen
2. Elinolot
3. Energia
4. Hinnat ja kustannukset
5. Julkinen talous
6. Kansantalous
7. Kauppa
8. Koulutus
9. Kulttuuri ja viestintä
10. Liikenne ja matkailu
11. Maa-, metsä- ja kalatalous
12. Oikeus
13. Palkat
14. Rahoitus
15. Rakentaminen
16. Sosiaaliturva
17. Teollisuus
18. Terveys
19. Tiede, teknologia ja tutkimus
20. Tulot ja kulutus
21. Työmarkkinat
22. Ulkomaankauppa
23. Vaalit
24. Vakuutus
25. Väestö
26. Väestölaskenta
27. Ympäristö ja luonnonvarat
28. Yritykset

- De Vries, W. 1998. How are we doing? Performance indicators for national statistical systems. *Netherlands Official Statistics*. Vol 13. 5-13.
- Office for National Statistics. 1996. *Guide to Official Statistics*. 1996 Edition. A Publication of the Government Service. London: HMSO.
- Statistical Office. 1982. Directory of International Statistics. Vol. 1, *Statistical Papers*. Ser. M. No. 56. Rev 1. New York: United Nations.
- Söder, L. 1998. *Ehdotus Suomen Virallisen Tilaston (SVT) laatukriteereiksi*. Muistio. Tilastokeskus.
- TES-kurssi: *The Role of Statistics in a Democracy*. 5.-6.3.1998. Luxemburg. (Kursiaineisto).
- United Nations Statistical Division. 1994. *Fundamental Principles of Official Statistics*. Supplement No. 9. Report of the Statistical Commission on its Special Session, held in New York 11-14 April 1994. E/CN.3/1994/18.

1.2.4 Suomen Virallisen Tilaston laatukriteerit

Soveltamisala

Suomen Virallisten Tilastojen laatukriteereiden määrittely on osa sitä työtä, jossa tavoitteena on koota virallisten tilastojen tuottajista ja näiden tuottamista tilastoista kokonaisuus sekä kehittää Suomen Virallisen Tilaston (SVT) tunnuksesta "tutkitun, virallisen tilaston" leima. Tarkoituksena on, että virallisen tilaston kokonaisuus erottuu muun tilastomuotoisen tiedon ja monien muiden yhteiskunnallisen tietopalvelun tuottajien joukosta. Virallisen tilaston tulee täyttää tietyt laatuvaatimukset ja tilaston tuottajan olla uskottava, jotta asiakas voi luottaa siihen.

Viralliset tilastot kuvaavat taloudellista ja sosiaalista tilannetta ja muutoksia yhteiskunnan eri sektoreilla ja niiden pohjalta tehdään usein kalliita ja kauaskantoisia päätöksiä. Luottamus näihin tilastoihin on sen takia hyvin oleellista. Virallisten tilastojen laatu ja niiden objektiivisuus on luottamuksen perusta.

Laadulle ei ole tarkkaa yksiselitteistä määritelmää. Tilastojen laatu on dynaaminen, moniulotteinen ja ristivaikutteinen kokonaisuus. Tilasto voi olla johonkin käyttötarkoitukseen korkealaatuinen, mutta toiseen tarkoitukseen sama laatu ei taas ole riittävä. Tilastojen laadun osatekijöinä käytetyt relevanssi, tarkkuus, ajantasaisuus, saatavuus, vertailukelpoisuus ja yhtenäisyys kuvaavat laatu-olottuuksia. Osalle näistä voidaan asettaa tarkat mitattavat raja-arvot laadulle, mutta useimmiten tilaston käyttäjät joutuvat tyytymään tilastontekijöiden tekemään laadun tai virhetekijöiden kuvailuun.

Käyttäjien luottamuksen aikaansaamiseksi virallisen tilaston tekijöiden ja johdon on osoitettava hyvin voimakkaasti sisäisessä toimintakulttuurissaan sitoutuminen laatuun. Laadun parantamistoimenpiteet, yleisesti hyväksytyjen laatustandardien ja välineiden käyttö, laatuselvitykset, systemaattiset laatu-arvioinnit ja myös virallisen tilaston riittävä dokumentointi sekä sen avoimuus ja läpinäkyvyys käyttäjille ovat osoituksia laatuun sitoutumisesta. Käyttäjille nämä ovat am-

mattiossaamisen tunnusmerkkejä ja tästä riippuu pitkälti käyttäjien luottamus organisaatioon ja sen palvelutuotteisiin. (Ks. 4.3.1 SVT:n Tilastojulkaisut ja tuoteselostesuositus).

Toimintaperiaatteet

SVT-laatukriteerit ovat:

1. tilastotietojen relevanssi (*relevance*)
2. tietojen oikeellisuus ja tarkkuus (*accuracy*)
3. julkaistujen tietojen ajantasaisuus (*timeliness*) ja oikea-aikaisuus (*promptness*), tai täsmällisyys tai luvattu aikaisuus
4. tietojen saatavuus ja läpinäkyvyys/selkeys (*accessibility and clarity*)
5. tilastojen vertailukelpoisuus (*comparability*)
6. selkeys ja eheys/yhtenäisyys (*consistency*)
7. dokumentointi (*documentation*).

Suosituksset

1. Relevanssi

Relevanssilla tarkoitetaan karkeasti käyttäjien tarpeiden mukaisen tiedon tuottamista. Tietojen käyttökelpoisuuden kannalta on tärkeää, että kerrotaan käytettävyyttä rajoittavat tekijät. Relevanssia täytyy jatkuvasti ylläpitää ja parantaa seuraamalla ympäristömuutoksia sekä käyttäjien odotusten ja vaatimusten muuttumisia.

2. Tietojen oikeellisuus

Tietojen oikeellisuudella tarkoitetaan sitä, että mitataan oikein ja luotettavasti sitä, mitä on tarkoitus mitata. Tilastoissa tulee selostaa erilaiset virhetyypit, jotka voivat liittyä tietojen käyttökelpoisuuteen. Virhetyypit ovat joko otantavirheitä tai kehikkoon, mittaukseen, prosessointiin, katoon tai mallintamiseen liittyviä virheitä.

3. Ajantasaisuus ja oikea-aikaisuus

Ajantasaisuudella ja oikea-aikaisuudella tarkoitetaan sitä, että tiedot ovat mahdollisimman tuoreita ja että ne ilmestyvät sovittuna aikana. Ajantasaisuus, oikeellisuus ja relevanssi ovat keskenään vuorovaikutuksessa ja niiden välillä on yleensä aina ristiriitaa. Tiedon tuoreuden mittarina käytetään yleensä sitä, kuinka monta viikkoa tilastointiajankohdasta menee siihen, kun tiedot ovat käytettävissä.

Yleinen tavoitetaso on ollut, että vuositilastojen tulisi ilmestyä seuraavan vuoden aikana, neljännesvuositilaston seuraavan vuosineljänneksen aikana ja kuukausitilaston seuraavan kuukauden aikana.

Useimmista vuositilastoista laaditaan myös ennakkotilastoja, jotka saattavat valmistua jopa puolivuotta lopullista tilastoa aikaisemmin. Niiden julkaisuajankohda määräytyy sen mukaan, miten luotettavana siihen mennessä käsiteltyä ai-

neistoa voidaan pitää lopullisen tilaston kannalta. Mittarina ennakkotietojen luotettavuudesta käytetään niiden eroavuutta lopullisista tiedoista.

Tietojen julkistamisajankohta ilmoitetaan etukäteen (ks. 4.1. Julkistaminen). Ajankohdan pitävyys on yksi korkean laadun merkki. Tilastojen julkistamisajoista ja -tavoista ilmoitetaan esimerkiksi julkistamiskalenterin kautta. Siinä kuukausitilastoista ilmoitetaan ilmestymispäivä, neljännesvuositilastoista ja vuositilastoista tietojen julkistamisviikko.

Joidenkin tilastojen julkistamista koskevat myös sisäpiirisäännökset (Tilastokeskus, 2000). Näillä tilastoilla on etukäteen ilmoitettu julkistamispäivä ja ajankohta, jota ennen vain erikseen nimetyt henkilöt ovat tekemisissä tilaston tietojen kanssa. Näin pyritään estämään tietojen spekulatiivinen käyttö.

4. Tietojen saatavuus ja läpinäkyvyys

Tilastotiedot ovat arvokkaampia, jos ne ovat helposti käyttäjien saatavissa, jos ne ovat valmiiksi heidän käyttötarkoituksiinsa soveltuviissa muodoissa ja jos tiedot ovat riittävästi dokumentoidut. Tietojen jakelu on ratkaiseva askel informaatioketjussa. Tietojen varastointikonttorina toimiminen ei riitä.

Tiedot tulee saattaa käyttäjien saataville heille soveliaassa muodossa. (i) Käyttäjillä tulee olla tieto, minkälaisia tietoja on saatavilla. (ii) Tietojen fyysinen saatavuus tulee tehdä mahdollisimman helpoksi. (iii) Käytetyistä käsitteistä ja metodeista pitää olla dokumentoitu selostus. (iv) Tilastoissa tulee olla myös analyysiä tuloksista. Kaksi viime mainittua helpottavat tietojen tulkittavuutta.

5. Vertailukelpoisuus

Hyvän tilaston ominaisuuksiin kuuluu myös tietojen vertailukelpoisuus yksiköiden kesken, yli ajan sekä muihin kotimaisiin ja kansainvälisiin tilastoihin. Tätä toteutetaan käyttämällä yleisiä tilastoluokituksia, jotka useimmiten on valmisteltu kansainvälisessä yhteistyössä. Muutokset esimerkiksi käsitteissä tulee dokumentoida ja vaikutukset arvioida. Samoin tulee menetellä yhteiskunnan yleisten, esimerkiksi lainsäädännön muutosten raportoinnissa.

6. Selkeys ja eheys

Eri tilastojen pitää olla keskenään loogisesti yhtenäisiä ja johdonmukaisia. Mikäli samasta aiheesta on useampia eri lähteitä, esimerkiksi hallinnollinen aineisto ja otantatutkimus, erot käsitteissä ja keruuprosessissa tulee kertoa. Eri lähteistä saaduissa tiedoissa tulee käyttää mielellään eri käsitteitä, jotta käyttäjää ei johdeta harhaan.

7. Dokumentointi

Dokumentoinnilla tarkoitetaan toiminnan kuvaamista. Siihen kuuluvat mm. käytetyt käsitteet, määritelmät, menetelmät sekä tilastotuotantoprosessi tietojärjestelmiseen ja tarvittavine työohjeineen. Dokumentoinnin tavoite on olla täydellinen, yksiselitteinen, monikäyttöinen ja ainakin Tilastokeskuksen sisällä julkinen kuvaus tilastosta ja sen tuotannosta. Kuvauksien on oltava ajantasaisia, hyvin organisoituja, helposti saatavilla, tiiviitä ja tarkkoja. Niiden avulla uuden henkilöstön

on voitava perehtyä tilastotuotannon pääpiirteisiin. Hyvä nyrkkisääntö dokumentoinnille on, että dokumentoidaan ne asiat, joiden dokumentoimattomuudesta aiheutuu oleellista haittaa. (Ks. 3. Dokumentointi).

Kustannukset

Kustannukset ovat enemmän rajoite kuin laadun komponentti. Tieto käytössä olevista resursseista antaa mahdollisuuden optimoida ne laadun suhteen. Kolme erilaista kustannuskomponenttia voidaan erottaa:

- kustannus tilastovirastolle
- kustannus tiedonantajalla
- kustannus/hyöty tietojen käyttäjälle.

Kirjallisuus

- Eurostat. 2000. *Assessment of the Quality in Statistics*. Definition of Quality in Statistics. Doc. Eurostat/A4/Quality/Definition.
- Eurostat. 2000. *Assessment of the Quality in Statistics. Standard Quality Report*. Doc. Eurostat/A4/Quality/Standard report.
- Söder, L. 1998. *Ehdotus Suomen Virallisen Tilaston (SVT) laatukriteereiksi*. Tilastokeskus. (Muistio).
- Tilastokeskus. 2000. *Tilastokeskuksen julkistamisohje*. Tietopalvelu/Tiedotus. (Muistio).

2 TILASTOTUTKIMUS JA -TUOTANTO

2.1 TILASTOTUTKIMUKSEN ELI SURVEYN MÄÄRITTELY

Soveltamisala ja päämäärä

Tilastotutkimukset jaotellaan perinteisesti aihealueittain. Kun tavoitteena on suunnitella, ylläpitää, tehostaa tai parantaa laadullisesti koko tilastotuotantoa ja/tai yksittäistä tilastoprosessia on hyödyllistä hahmottaa eri tilastotutkimusprosessien yhtäläisyydet ja erityispiirteet. Tässä luvussa tarkastellaan tilastotutkimusprosessien yleisluonnetta.

Tilastotutkimus rajataan tässä yhteydessä niihin tilastotutkimuksiin, jotka tuottavat luonteeltaan kuvailevaa tai analyttistä yhteiskunnallista tietoa yhteiskunnallisen päätöksenteon, tieteellisen tutkimuksen tai kansainvälisten vertailujen tarpeisiin. Niiden päämääränä on kuvata koko perusjoukon tilaa tarkasteltavan seikan suhteen. Yleensä perusjoukko koostuu henkilöistä, kotitalouksista, yrityksistä tai muista taloudellisista ja/tai sosiaalisista yksiköistä.

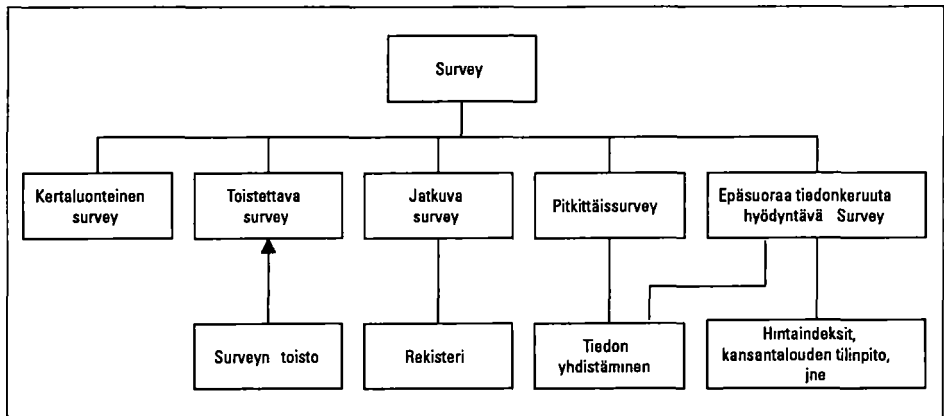
Kansainvälisessä kirjallisuudessa näihin tilastotutkimuksiin viitataan survey -termillä (*statistical survey*), joka käsittää laajasti kaiken empiirisyhteiskunnallisen tilastotutkimuksen ja -tuotannon. Tarkasti ottaen survey yhdistetään koko tavoiteperusjoukkoon ja siitä kerättyyn aineistoon (Marriott, 1990). Survey kattaa määritelmänsä mukaan:

- kokonaisaineistot (*census survey*), joissa tiedonkeruu kattaa kaikki perusjoukon yksiköt
 - otosaineistot (*sample survey*), joissa tietoa kerätään tavoiteperusjoukosta (tavallisesti satunnaisesti) poimitusta otoksesta
 - aineistot, jotka on muodostettu hallinnollisista rekistereistä (*administrative records*)
 - johdetut tilastolliset aineistot (*derived statistical activity*), joissa aineisto on esitimoitu, mallitettu tai muuten johdettu olemassa olevista tietovarastoista.
- (Statistics Canada, 1998).

Empiirisyhteiskunnallinen tilastotutkimus eli survey on siis yläkäsite, joka kattaa myös rekisteritietoihin perustuvat kokonaistutkimukset. Kokonaisaineistoon perustuvaan tilastotutkimukseen poimitaan jokainen tavoiteperusjoukon yksikkö. Otosperusteinen tilastotutkimus puolestaan poimii osan tavoiteperusjoukosta kuuluvista yksiköistä pyrkien edustavaan ja koko tavoiteperusjoukkoon yleistettävään tulokseen. Tilastotutkimukset kattavat suoran ja epäsuoran tiedonkeruun otoksesta tai koko perusjoukosta ja näiden yhdistelmät. Esimerkiksi Tilastokeskuksen otantatutkimuksissa hyödynnetään sekä rekisteritietoa että suoralla tiedonkeruulla saatavaa tietoa (esim. haastattelutietoa).

Tilastotutkimukset voidaan jakaa toistuviin ja kertaluontoisiin tutkimuksiin (ks. Kaavio 2.1). Toistuvat tilastotutkimukset tehdään tietyin määräajoin ja niiden perusteella voidaan laatia myös aikasarjoja ja ennusteita. Tilastotutkimukset voi-

Kaavio 2.1 Empiirisyhteiskunnallisen tilastotutkimuksen jaottelu (Sundgren, 2001)

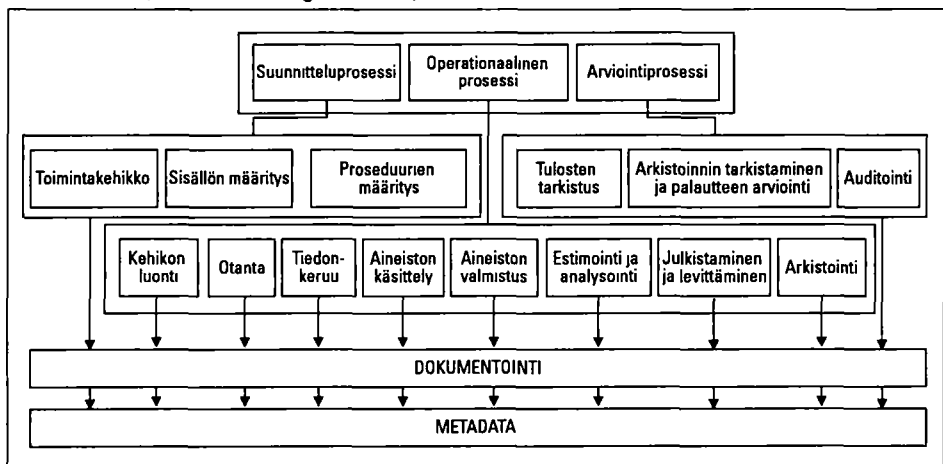


daan jaotella pysyviin, uusiin tai jo aiemmin tehdyn pohjalta kehitettäviin tutkimuksiin. Ensiksi mainitussa tapauksessa tilastotutkimus toistetaan jo monta kertaa käytössä olleilla kysymyksenasetteluilla. Aivan uuden tilastotutkimuksen suorittamisessa joudutaan kokonaisvaltaisesti konstruoimaan alusta alkaen koko tilastotutkimuksen kokonaisprosessi. Tiedonkeruulomakkeen testausvaiheessa voidaan apuna käyttää Tilastokeskuksen surveylaboratoriota ja/tai tehdä pilottitutkimus (ks. 2.7 Tiedonkeruulomakkeen suunnittelu ja testaus).

Periaatteet

Tilastotutkimus koostuu kolmesta pääprosessista: suunnittelusta, operationaalista prosessista ja arvioinnista (ks. kaavio 2.2).

Kaavio 2.2 Empiirisyhteiskunnallisen tilastotutkimuksen prosessikokonaisuus (muokaten Sundgren, 1999¹)



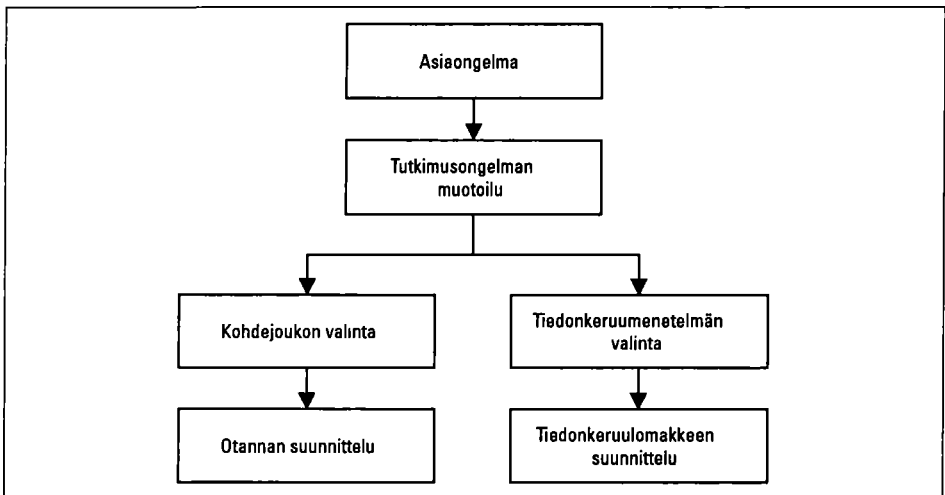
¹ Kaavioon on tehty neljä lisäystä verrattuna alkuperäiseen (Sundgren, 1999). Kaavioon on lisätty dokumentointi ja metatiedon karttuminen läpi tilastotutkimusprosessin. Tämän lisäksi suunnitteluvaiheeseen on lisätty 'toimintakehikko' ja arviointivaiheeseen 'auditointi'.

Suunnitteluprosessi

Empiirisyhteiskunnallisen tilastotutkimuksen suunnittelu voi pohjautua yhteiskunnallisen päätöksenteon, tieteellisen tutkimuksen tai tilastojulkaisujen tarpeisiin. Asiakaslähtöisen aineistotarpeen määrittämisessä suunnitteluvaiheessa yksilöidään asiakkaan/tutkijan/sidosryhmän tarve. Tilastolain asettamat toimintatavat, käsitteet, luokitukset ja standardit määrittävät tilastotutkimuksen toimintakehikon (ks. 1.1 Tilastotoimen normisto ja 2.2 Luokitukset, käsitteet ja määritelmät sekä harmonisointi). Tilastotutkimuksen tutkimusongelma tulee olla realistisesti toteutuskelpoinen ja tutkimussuunnitelmassa tulee yksilöidä tavoiteperusjoukko ja asiasisältö. Suunnitteluvaiheen keskeisiä asioita ovat tilastotutkimuksen sisällön määrittäminen, strategiset päätökset tiedonkeruumenetelmistä ja tiedonkeruun suunnittelu (ks. Kaavio 2.3). Suunnittelun alkuvaiheessa selvitetään myös käytettävissä olevat resurssit, kuten budjetti, henkilöstö ja muu välineistö; samalla määritetään tilastotutkimuksen prosessi ja sen aikataulu sekä varmistetaan että SVT:n mukaiset laatukriteerit ovat täytettävissä (ks. 1.2 Tilastojen laatu).

Aineiston tiedonkeruumenetelmän valinta on merkittävä strateginen ratkaisu tilastotutkimuksen suunnittelussa, jolla on suuri merkitys erityisesti kerättävän tiedon luonteelle, kysymysten asettelulle, kyselylomakkeen suunnittelulle, tilastotutkimuksen läpivientiajalle ja kustannuksille (ks. 2.7 Tiedonkeruulomakkeen suunnittelu ja testaus ja 2.6 Tiedonkeruutavat henkilö pohjaisessa ja yrityspohjaisessa tiedonkeruussa). Suunnitteluvaiheessa yksilöidään myös tietojenkäsittelymenetelmät sekä ennakoidaan lopullisen aineiston tuottamiseen tarvittavat editointi- ja imputointimenetelmät (ks. 2.9 Tilastollinen editointi ja imputointi). Yleisesti

Kaavio 2.3 Empiirisyhteiskunnallisen tilastotutkimuksen suunnitteluprosessi (muokaten Tilastokeskus, 1986²)



² Kaavio on osa julkaistua kaaviota (Tilastokeskus, 1986, 3). Alkuperäisestä kaaviosta on rajattu tähän yhteyteen vain tilastotutkimuksen suunnitteluun liittyvät kohdat seuraavin muutoksin: 'otantamenetelmän valinta' on muutettu 'otannan suunnitteluksi', mikä kuvaa tarkemmin tähän liittyvää suunnitteluprosessissa tehtävien päätösten sarjaa; 'tiedonkeruulomakkeen laatiminen' kohdasta on irrotettu suunnitteluvaiheessa tehtävä 'tiedonkeruulomakkeen suunnittelu'.

suunnitteluvaiheeseen kuuluu tärkeänä osana myös haastattelijoiden koulutus sekä kenttätöyön organisointi (ks. Tilastokeskus, 1999).

Otosperusteisen tilastotutkimuksen keskeistä tilastotieteellistä työvaihetta kutsutaan otannan suunnitteluksi, joka sisältää niin otanta-asetelman kuin otanta-aineiston estimointimenetelmien määrittämisen (ks. 2.5 Otantamenetelmät). Otannan suunnittelussa tulee arvioida käytettävissä olevien otantakehikkojen soveltuvuus ja niiden määrittämien kohdeperusjoukkojen vastaavuus otantatutkimuksen tavoiteperusjoukkoon (ks. 2.3 Kehikot ja peittävyys). Otoksen poimintamenetelmä, otoksen koko ja sisällymistodennäköisyydet tavoiteperusjoukon eri ryhmissä määritellään siten, että tilastotutkimuksen tuottamat estimaatit täyttävät niille asetetut luotettavuuskriteerit (ks. 2.10 Painokertoimien muodostus ja katovirheen oikaisu, 2.11 Tilastollinen estimointi ja analyysi). Eri otantamenetelmien vertailu on hyvin tärkeä työvaihe etsittäessä tutkittavan ongelman kannalta mahdollisimman täsmällisiin tuloksiin johtava otanta-asetelma (Pahkinen ja Lehtonen, 1989). Otanta-aineiston estimointimenetelmien määrittäminen kattaa katovirheen oikaisumenetelmät, estimaattorien johtamisen, estimaattien laskemisen ja niiden keskivirheiden kaavat.

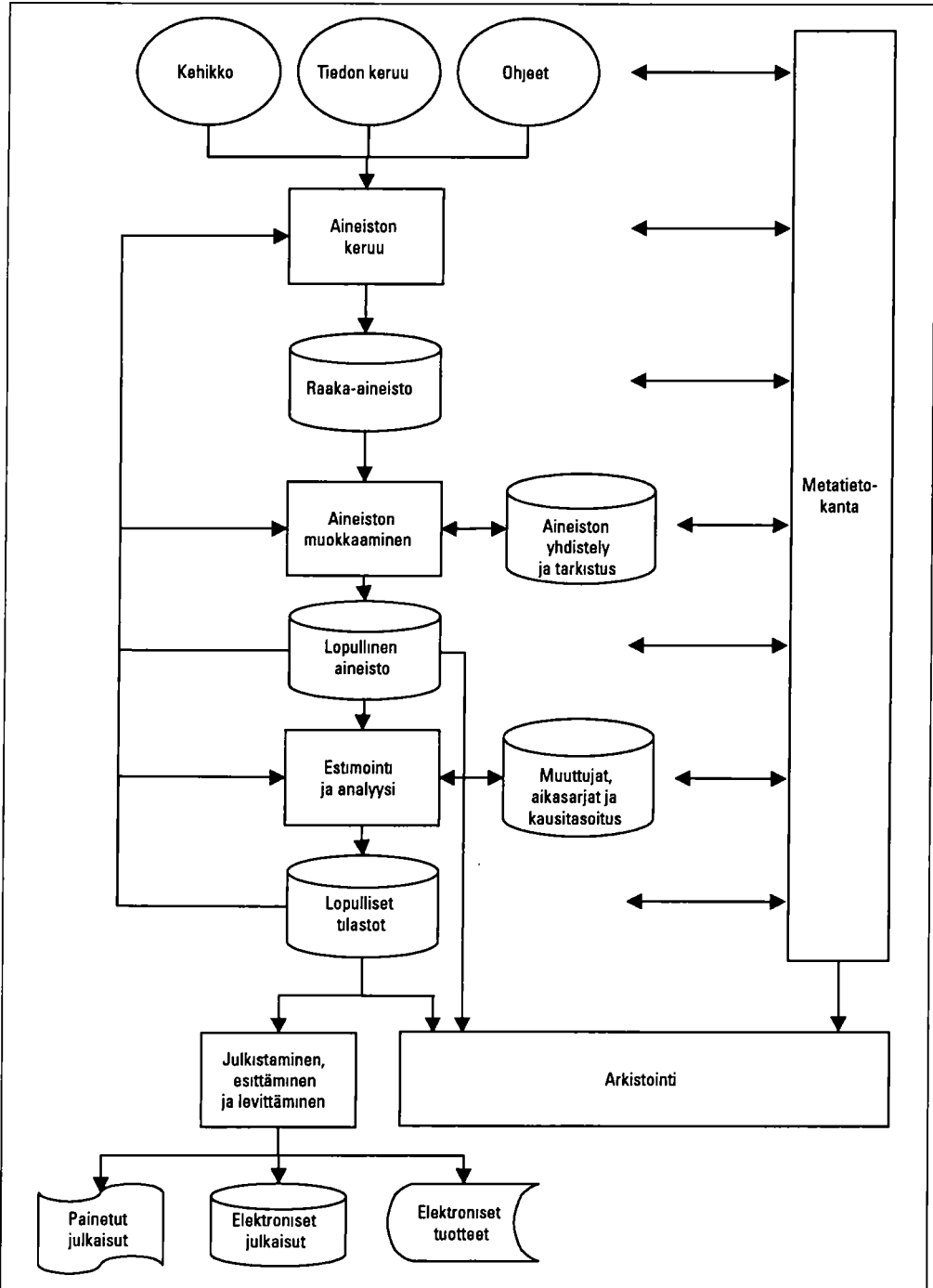
Suunnitteluprosessissa tulee pyrkiä ennakoimaan mahdollisesti esiintyvät ongelmat. Esimerkiksi haastattelututkimuksessa tulee varautua vastauskatoon ja sen vaatimiin korjausmenetelmiin. Tilastotutkimuksen suunnittelun dokumentointi nopeuttaa sen teknistä raportointia eli tuoteselosteiden, laatuselvitysten ja käyttäjän oppaiden laadintaa. Suunnittelun ensisijainen päämäärä on edesauttaa saavuttamaan mahdollisimman tarkka ja oikea tutkimustulos.

Operationaalinen prosessi

Tilastotutkimuksen operationaalinen vaihe koostuu seuraavista prosesseista (ks. Kaavio 2.4):

- otoskehikon konstruointi ja tarvittavan lisäinformaation yhdistäminen (ks. 2.3 Kehikot ja peittävyys)
- otoksen poiminta (ks. 2.5 Otantamenetelmät)
- kyselylomakkeen laadinta /ohjelmointi ja työohjeiden valmistaminen (ks. 2.7 Tiedonkeruulomakkeen suunnittelu ja testaus)
- tiedonkeruu ja sen koordinointi (ks. 2.6 Tiedonkeruutavat henkilö pohjaisessa ja yrityspohjaisessa tiedonkeruussa)
- aineiston muokkaus (tiedon koonti, koodaus, (loogisuus)tarkistus ja editointi) (ks. 2.9 Tilastollinen editointi ja imputointi)
- lopullisen aineiston tuottaminen (ks. 2.9 Tilastollinen editointi ja imputointi, 2.10 Painokertoimien muodostus ja katovirheen oikaisu, 2.11 Tilastollinen estimointi ja analyysi)
- estimointi ja analysointi (ks. 2.11 Tilastollinen estimointi ja analyysi)
- lopullisten tulosten julkistaminen (ks. 4.1 Julkistaminen)
- tilastotiedon levittäminen (ks. 4.2 Tilastotiedon levittäminen)
- arkistointi (ks. 3.2 Arkistointi).

Kaavio 2.4 Empiirisyhteiskunnallisen tilastotutkimuksen operationaalinen pääprosessi (muokaten Sundgren, 1999³)



³ Kaavioon (Sundgren, 1999, 3) on muutettu metatiedon yhdistämistä käsittelevä kohta siten, että metatietoa kerätään ja yhdistetään koko prosessin aikana. Lisäksi alkuperäiseen kaavioon on lisätty lopullisen aineiston, lopullisten tilastojen ja metatietojen "arkistointi".

Tiedonhallinta ja systeemityö nivoutuvat keskeisesti tilastotutkimuksen jokaiseen vaiheeseen. Tiedonhallintaa ja systeemityötä hyödynnetään systeemin suunnittelusta ja raaka-aineistosta lähtien lopullisen aineiston tuottamiseen ja arkistointiin.

Metatietoa syntyy jo tilastotutkimuksen suunnitteluprosessissa; operationaalissa prosessissa metatiedon määrä kumuloituu läpi koko prosessin. Metatiedolla tarkoitetaan tässä yhteydessä esimerkiksi käytettyjä määritelmiä, luokituksia, standardeja, otantakehikkoa, käytettyä lisäinformaatiota, kyselylomaketta, työohjeita, vastaajien ja haastattelijoiden kommentteja, tehtyjä editointeja ja muutujamuunnoksia.

Arviointiprosessi

Tilastotutkimuksen arviointivaiheessa tarkastellaan miten suunnitteluvaiheessa määritellyt lopputuotteet on tuotettu, julkaistu ja miten asetetut laatu kriteerit on saavutettu. Lisäksi tarkastetaan ja arvioidaan metatiedon dokumentointi; aineistojen, luokitusten ja käsitteiden arkistointi sekä asiakastutkimukset. (Sundgren, 1999).

Koko tilastotutkimuksen suoritusta ja laatu kriteerien täyttämistä voidaan tarkastella lisäksi erillisellä auditoinnilla, jossa tilastoammattilaiset ja puolueettomat asiantuntijat arvioivat yhdessä tilastotutkimuksen kokonaisprosessia ja pyrkivät selvittämään parannusmahdollisuuksia.

Suositukses

Empiirisyhteiskunnallisen tilastotutkimuksen korkea tasoinen laatu voidaan saavuttaa:

- resursoimalla jokainen tilastotutkimuksen vaihe oikein,
- panostamalla suunnitteluvaiheeseen,
- asiakastarpeen huolellisella analysoinnilla,
- oikein resursoidulla operationaalisella vaiheella, jolle on ominaista järjestelmällinen systeemityö,
- panostamalla yksittäisten toimintojen ja kokonaisuuden vuorovaikutukseen,
- dokumentoimalla yksittäiset työvaiheet siten, että ne ovat toistettavissa toisen tilastoammattilainen toimesta (ks. 3.1 Tilastotutkimuksen dokumentointi),
- hyväksymällä tilastotutkimus valmiiksi vasta kun se on arkistoitu (ks. 3.2 Arkistointi),
- arvioimalla tilastotutkimuksen tuotos,
- tekemällä yhteistyötä muiden tilastotutkimusten kanssa ja jakamalla hyviä käytäntöjä ja tietämystä.

Kirjallisuus

Marriott, F.H.C. 1990. *A Dictionary of Statistical Terms*. 5th ed. Burnt Mill: Longman Scientific & Technical.

- Pahkinen, E. & Lehtonen, R. 1989. *Otanta-asetelmat ja tilastollinen analyysi*. Helsinki: Gaudeamus.
- Statistics Canada. 1998. *Statistics Canada Quality Guidelines*. Ottawa: Statistics Canada.
- Sundgren, B. 1999. Information systems architecture for national and international statistical offices. Guidelines and recommendations. *Conference of European Statisticians Statistical Standards and Studies*. No.51. Geneva: United Nations.
- Sundgren, B. 2001. *Documentation and Quality in Official Statistics*. Q2001. 14-15.5.2001 Stockholm, Sweden. Conference CD. 1st ed. Statistics Sweden & Eurostat.
- Tilastokeskus. 1986. *Kysymyksiä ja vastauksia haastattelu- ja kyselytutkimuksen tarvisijoille*. Helsinki: Valtionpainatuskeskus.
- Tilastokeskus. 1999. *Tilastohaastattelijan opas*. Haastattelu- ja tutkimuspalvelut. Elokuu 1999.

2.2 LUOKITUKSET, KÄSITTEET JA MÄÄRITELMÄT SEKÄ HARMONISOINTI

Soveltamisala ja päämäärä

Tilasto tehdään yleensä suuresta joukosta yksittäisiä havaintoja. Johtopäätösten tekemiseksi empiirisiä havaintoja on järjestettävä, luokiteltava esimerkiksi samalaisuuden tai erilaisuuden perusteella. Luokituksen avulla havaintoaineisto voidaan järjestää tavalla, joka kuvaa tarkastelun kohteena olevaa ilmiötä johdonmukaisesti. Siten luokitus on olennainen osa tilastoa ja standardiluokitukset tilastoimen keskeisiä instrumentteja.

Käsitteet, määritelmät ja luokitukset yhdessä luovat osaltaan tilastoinnin keskeisen toimintakehikon. Käytetyt käsitteet perustuvat yleensä teoreettiseen tai tilastolliseen viitekehykseen ja niiden avulla rajataan aihetta, kuvattavia tilastoyksiköitä ja/tai perusjoukkoa, jota havainnot koskevat.

Tilastoissa käytettävät käsitteet ovat yleensä tilastollisia standardikäsitteitä, mutta ne voivat olla myös yleiskielen käsitteitä, jotka on määriteltävä. Tieteellisten käsitteiden käytössä hankaluuksia aiheutuu usein siitä, että yhdellä käsitteellä voidaan viitata moneen eri ilmiöön ja että käsitteellä tai sanalla voi olla monta merkitystä tai siitä, että kaikilla käsitteillä ei ole empiiristä vastinetta.

Pelkistetyimmillään luokitus koostuu nimetyistä luokista ja niiden tunnuksista (koodeista). Luokitukset koostuvat (i) luokkien määrittelystä (luokkien nimet, koodit ja usein myös luokan kuvaus eli sen tekstimuotoinen määrittely) sekä (ii) peruskäsitteiden ja luokitusperiaatteiden ja -kriteereiden esittelystä.

Luokitukset voidaan karkeasti ryhmitellä:

- väestöä, henkilöitä ja näiden elinoloja kuvaaviin luokituksiin (demografiset luokitukset),
- taloudellista toimintaa ja kansantaloutta kuvaaviin luokituksiin,
- alueluokituksiin ja ympäristötilastoluokituksiin,
- lisäksi on eräitä erityisluokituksia kuten tautiluokitus, tullitariffinimikkeistö ja rikosnimikkeistö.

Luokitteluperiaatteet

Aineisto, havainnot, on ryhmiteltävä eli luokiteltava niin, että laadittu tilasto tuottaa todellisuutta vastaavan, jäsenellyn ja ymmärrettävän kuvan reaali maailmasta – yhteiskunnan rakenteesta, sen tilasta ja siinä tapahtuvista muutoksista. Tilastossa käytettyjen luokitusten on oltava relevantteja keskeisten käyttäjien ja näiden käyttötarkoitusten näkökulmasta, koska tilastoilla on merkittävä rooli yhteiskunnallisessa tutkimuksessa ja päätöksenteossa.

Luokitus on aina systemaattinen: se luokittelee havainnot johdonmukaisesti käyttäen sovittuja kriteerejä. Luokitus on yleensä myös hierarkkinen niin, että ylin luokittelutaso (tavallisimmin puhutaan pääryhmätasosta tai 1-numerotasosta) kuvaa tilaston kohteena olevan ilmiön rakennetta, jota tarkennetaan siirryttäessä alemmille luokitustasoille.

Luokituksen luokkien tulee olla yksiselitteisiä, kattavia ja toisensa poissulkevia. Yksiselitteisyys tarkoittaa sitä, että havainnot voidaan luokitella määriteltyjen luokitusperiaatteiden ja kriteereiden perusteella selvästi tiettyyn luokkaan. Kattavuus merkitsee sitä, että kaikki havaintoaineiston tapaukset voidaan luokitella. Poissulkevuus toteutuu, kun luokkien määrittelyt on tehty niin, että tapauksia ei voi luokitella kahteen tai useampaan luokkaan.

Tarkemmin määrittelemättömien luokkien eli kaatoryhmien kuten "muualle luokittelemattomat..." tulisi käsittää suhteellisen vähän tapauksia verrattuna luokittelutason muihin luokkiin. Jos kaatoryhmän koko kasvaa huomattavaksi, luokitus ei enää toimi hyvin tilaston tai tutkimuksen aiheen kuvauksen välineenä.

Luokitukseen liittyvät käsitteet ja muuttujat on määriteltävä. Määritelmät voivat olla luonteeltaan yleisiä tai tietyn aineiston havaintojen käsittelyä kuvaavia operationaalisia määritelmiä.

Luokitusstandardit

Tilastolliset standardiluokitukset perustuvat lähes poikkeuksetta kansainvälisiin suosituksiin tai sopimuksiin. Osa Euroopan yhteisön jäsenmaiden käyttämistä luokituksista on vahvistettu säädöksinä. Tällaisia ovat muun muassa toimiala- ja tuoteluokitukset sekä uudistetun kansantalouden tilinpidon (ESA 95) perusluokitukset. Esimerkiksi väestöä ja sen elinoloja kuvaavat luokitukset perustuvat väestölaskentasuositukseen. Lisäksi tulonjakoa ja kulutusta kuvaavien luokitusten taustalla on kansantalouden tilinpidon kehikon lisäksi tulotilastosuositus.

Keskeiset tilastostandardit vahvistetaan Suomessa julkisen hallinnon suosituksina (JHS). Menettelyllä pyritään laajentamaan standardiluokitusten käyttöä kaikilla yhteiskunnan sektoreilla ja ennen muuta julkisen hallinnon toiminnassa. Yleisiä tilasto-

standardeja käytetään laajasti hallinnollisten viranomaisten ja yhä enemmän myös yksityisten yritysten ja organisaatioiden tietojärjestelmissä. Tällä seikalla on tilastotoimen kannalta suuri merkitys, koska hallinto, yritykset ja yhteisöt ovat merkittäviä tiedonantajia. Niiden tuottamia aineistoja käytetään hyväksi tilastojen perusaineistoina.

Suosituksia

Standardiluokituksia on käytettävä aina kun mahdollista. Tilastoja julkaistaessa on aina esiteltävä käytetyt luokitukset ja niihin liittyvät keskeiset käsitteet.

Kun käyttöön otetaan uusi luokitus, on pidettävä huolta siitä, että uuden luokituksen luokille ei anneta jo käytössä olevien kansallisten tai kansainvälisten standardiluokitusten nimikkeitä, jos määrittelyt ja käsitteet eroavat käytössä olevista.

Käyttäjää on informoitava aina tilastojulkaisun teksteissä tai muulla sopivalla tavalla, jos standardikäsitteitä ja luokituksia ei jostain syystä voida noudattaa puhtaasti. Poikkeaminen EU:n säädöksiin vahvistamista luokituksista on muodollisesti hyväksytettävä yhteisön tilastovirastossa, Eurostatissa.

Luokkia nimettäessä tulee pyrkiä välttämään kaatoryhmän (residuaaliluokan) nimikkeenä termiä ”Muut”, koska se toimii huonosti vapaassa tekstihaussa.

Jos havaintoaineiston koko ei mahdollista luokituksen soveltamista kokonaisuudessaan tai valitulla luokittelutasolla, joudutaan luokitusta karkeistamaan. Se tapahtuu aggregoimalla eli yhdistämällä luokkia samalla tasolla. Yleinen merkintätapa on antaa tällaiselle, edelleen erittelemättömälle luokalle 0-päätteinen tunnus. Vastaavasti kaatoryhmille annetaan usein 9-päätteinen tunnus.

Vertailukelpoisuus, harmonisointi

Käyttämällä kansainvälisiin suosituksiin tai sopimuksiin perustuvia luokituksia varmistetaan se, että eri maista tuotettu tilastoinformaatio on vertailukelpoista siitä huolimatta, että yhteiskunnat ja niiden rakenteet poikkeavat toisistaan. Kansainvälisistä luokituksista laaditaan yleisesti kansalliset versiot, jotka mahdollistavat kansallisten erityispiirteiden huomioon ottamisen sekä luokittelun tarkentamisen kansallisten olosuhteiden kannalta relevantteihin asioihin. Kansallisen version laatiminen tapahtuu tavallisimmin niin, että kansainväliseen standardiluokitukseen lisätään kansallinen detaljitaso luokituksen alimmaksi tasoksi. Näin meneteltäessä kansallinen luokituksen rakenne ja ylemmät luokitustasot seuraavat puhtaasti kansainvälistä luokitusta.

Dokumentointi

Luokitukset, käsitteet sekä määritelmät ja käsittelysäännöt on dokumentoitava huolellisesti ja sellaisessa muodossa, että tieto on Tilastokeskuksessa siirrettävissä helposti käyttäjille: luokitukset luokitustietokantaan ja käsitteet käsitetietotietokantaan. Otostutkimuksiin liittyvät käsittelysäännöt ja muut työohjeet talletetaan tilastoarkistoon siltä osin kuin niitä ei ole kuvattu koontitiedostossa.

Yksittäisiä aihealueittaisia suosituksia

Valtioiden ja maiden tunnuksset. Kansalaisuutta, syntymä- ja asuinmaata, muuttoliikettä, majoitusta, ulkomaankauppaa, yritysten kotimaata tilastoitaessa voi-

daan käyttää kansainvälistä standardia ISO 3166 valtioiden ja maiden tunnuksista. Tunnusjärjestelmään kuuluvat valtioiden ja maiden kolminumeroiset, kaksi-kirjaimiset sekä kolmikirjaimiset tunnukset. Ensisijaisesti suositellaan käytettävän kolminumeroista tunnusta. Tunnusjärjestelmään sisältyvät myös itsenäiset valtiot sekä epäitsenäiset alueet (maat). Valtiot ja maat voidaan ryhmitellä kuuteen maanosaan ja edelleen 19 suuralueeseen. Jako soveltuu käytettäväksi lähinnä väestötilastoissa. Valtioita ja maita voidaan ryhmitellä myös muilla tavoilla, esim. kauppapoliittisin perustein tai valtioliittoumien mukaan.

Alueluokitukset ja Ahvenanmaan erityisaseman huomioon ottaminen. Tilastoissa käytetään pääasiassa Euroopan unionin virallista NUTS -alueluokitusjärjestelmää (NUTS, Nomenclature des Unités Territoriales Statistiques, Eurostat 1995). Suomen NUTS-aluejaossa Manner-Suomi ja Ahvenanmaa muodostavat NUTS1-alueitaso, suuralueet NUTS2-, maakunnat NUTS3-, seutukunnat NUTS4- ja kunnat NUTS5-tason.

Maakunnat ja tilastollinen kuntaryhmitys ovat tilastoissa suositeltavat ensisijaiset alueluokitukset. Kuntatietoja (NUTS5) taulukoitaessa kunnat esitetään alueittain, yleensä maakunnittain, aakkosjärjestyksessä. Ahvenanmaa on aina maakunnista viimeisenä. Painetuissa julkaisuissa, joissa esitetään alueittaisia tilastolukuja, suositellaan otettavaksi käyttöön NUTS1 -aluejako (Manner-Suomi, Ahvenanmaa) väliotsikkona välisummineen tai ilman niitä.

Tilastollinen kuntaryhmitys. Kuntien kaupunkimaisuutta kuvattaessa käytetään tilastollista kuntaryhmitystä. Kuntamuotoluokituksen käytöstä luovuttiin vuonna 1997; sitä paitsi kuntamuoto ei ole ollut pitkään aikaan kaupunki/maaseutu -käsiteparin synonyymi.

Kotitalouskäsite. Suomessa kotitalouskäsitettä voidaan käyttää vain haastatellututkimuksissa ja postikyselyissä. Rekisteripohjaisissa väestölaskennoissa ja vuositilastoissa sekä muissa hallinnollisiin aineistoihin perustuvissa tilastoissa käytetään käsitettä asuntokunta.

Toimiala - ja tuoteluokitukset. Euroopan unionin jäsenvaltioiden toimialoittaiset tilastot pitää laatia käyttäen NACE Rev. 1:tä tai siitä johdettua kansallista luokitusta. Tilastovuodesta 2003 käyttöön otetaan em. luokituksen päivitetty versio, NACE 2002, ja siihen perustuva kansallinen toimialaluokitus, TOL 2002.

Tuoteluokitusta (CPA, Classification of Products by Activity) käytetään raportointiluokituksena toimitettaessa mm. tuotantoon, kulutukseen, ulkomaankauppaan ja jakeluun liittyviä tilastotietoja Eurostatille.

Teollisen tuotannon tilastoinnissa käytetään PRODCOM-luetteloa, joka kattaa tuotenimikkeet toimialaluokituksen pääluokille C, D ja E (mineraalien kaivaminen, teollisuus sekä sähkö-, kaasu- ja vesihuolto). Kansallisista tarpeista johtuen PRODCOM-luetteloa on täydennetty lisänimikkeillä, jotka joko kokonaan puuttuvat PRODCOM-luettelosta tai ovat 8-numeroisten PRODCOM-nimikkeiden alanimikkeitä. Kansalliset lisänimikkeet ovat 10-numeroisia.

Luokitustiedon lähteitä:

Tilastokeskuksessa luokituspalvelut -yksikkö ylläpitää luokituksia luokitustietokannassa, julkaisee luokituksia käsikirjoina ja välittää luokitusinformaatiota Internetissä. Keskeiset luokitukset on esitelty myös Luokitusopas -vihkosessa, jota voi jakaa asiakaskoulutuksen tilaisuuksissa ja esimerkiksi koordinoinnin tapaamisissa. Luokitusopas -vihkosta saa luokituspalveluista.

Luokitustietokanta sisältää tärkeimmät luokitusstandardit erikielisinä (suomi, ruotsi, englanti) ja -pituisine luokitusversioineen, luokitusten metatiedot, suurten luokitusten hakemistot (hakusanat) ja luokitusten väliset muunnos- avaimet. Luokitustietokanta toimii luokitustuotteiden kuten käsikirjojen tuotantoympäristönä ja palvelualustana. Luokitustietokanta on kaikkien käytettävissä, mutta vaatii ensin, että se asennetaan omaan työpisteeseen.

Luokituspalveluilla on oma sivusto Internetissä. Sivusto esittelee keskeiset Suomen tilastotoimissa käytettävät standardiluokitukset metatietoineen. Verko-osoite on: <http://tilastokeskus.fi/luokitukset/index.html>

Eurostatin käyttämät luokitusstandardit ja useat muut laajassa kansainvälisessä käytössä olevat luokitusstandardit on viety RAMON-luokitustietokantaan. Verko-osoite on: <http://europa.eu.int/comm/eurostat/ramon/>

Vastaavasti YK:n julkaisemat luokitusstandardit löytyvät YK:n tilastoviraston (UNSD) verkkosivuilta. Verkkosoite on <http://www.un.org/Depts/unsd/class/>

2.3 KEHIKOT JA PEITTÄVYYS

Soveltamisala ja päämäärä

Empiirisyhteiskunnallisen tilastotutkimuksen tavoitteet määräävät, mistä kohde-ryhmästä ja yksiköistä ollaan kiinnostuneita. Nämä yksiköt muodostavat perusjoukon. Lähes jokaisen tilastotutkimuksen päämääränä on yleistää tutkimuksen lopputulokset koko perusjoukkoon; tilastot ilmaistaan yleisimmin jakaumien, summien, keskiarvojen, desiilien ja suhdelukujen muodossa, jotka on laskettu koko perusjoukolle (Levy ja Lemeshow, 1991).

Tilastotutkimuksen tulosten yleistettävyys, relevanssi ja tarkkuus edellyttää oikeanmukaisten kehikon käytön tilastotutkimuksen tiedonkeruussa. Niin kokonaistutkimuksissa kuin otantatutkimuksissa kehikon peittävyys, täydellisyys, ajantasaisuus, tietosisältö ja tietojen tarkkuus ovat kriittisiä tekijöitä kehikon sopivuudesta tilastotutkimuksen käyttöön. Keskeistä on arvioida myös kehikon yksiköiden ja tilastoitavien yksiköiden välinen yhteys ja määritelmälliset erot ja yhteneväisyydet. Käytännön ongelmia saattaa esiintyä esimerkiksi kotitalouden tai perherakenteen identifioimisessa asuntokunta perusteisissa rekistereissä. (Kish, 1965).

Tässä kappaleessa määritetään empiirisyhteiskunnallisen tilastotutkimuksen käyttämien kehikoiden yleispiirteet, ja esitetään kehikoiden suhde tilastotutkimuksen tulosten yleistettävyteen ja laatuun. Seuraavassa kappaleessa (2.4 Hal-

linnolliset rekisterit ja tilastorekisterit) kuvataan tarkemmin Tilastokeskuksen yleisesti käyttämien eri kehikoiden luonne ja erityispiirteet.

Pääperiaatteet

Tilastotutkimuksen perusjoukko tulee rajata siten, että jokaisen yksikön kuuluvuus perusjoukkoon voidaan yksiselitteisesti määrittää. Seuraavaksi tilastotutkimuksen suunnittelussa tulee kartuttaa vaihtoehtoiset kehikot ja soveltumiskriteerit perusjoukon tutkimiseen niitä käyttäen. Kehikon soveltuvuuteen vaikuttaa myös tilastotutkimuksen luonne:

- Kokonaistutkimuksia tehtäessä kehikon ja siihen yhdistettävän lisäinformaation tulee vastata tilastotutkimuksen sisällöllisiä tarpeita.
- Otantatutkimusta tehtäessä on oleellista, että kehikko sisältää valitun otantamenetelmän tarvitsemää lisäinformaatiota esimerkiksi otoksen ositukseen.

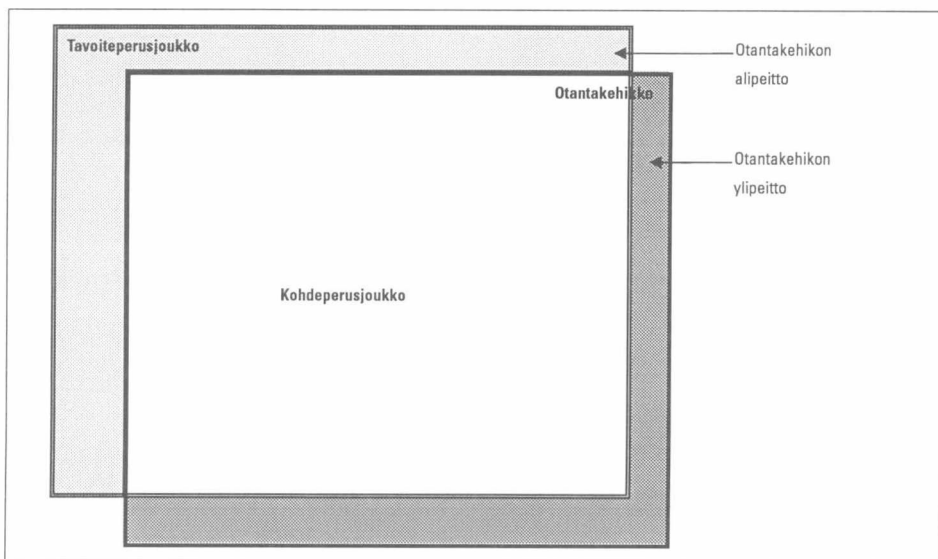
Koko perusjoukon tulisi olla tutkimuksen piirissä, mutta käytännössä on harvinaista, että kaikki perusjoukon alkiot olisivat saavutettavissa. Tämän vuoksi tavoiteperusjoukko ja kehikkoperusjoukko on erotettava toisistaan. Tavoiteperusjoukkoon kuuluvat ne alkiot, joista ollaan kiinnostuneita kun taas kehikkoperusjoukkoon kuuluvat ne yksiköt, jotka ovat saavutettavissa ja jotka kuuluvat otantakehikkoon. Yleisimmin otantakehikkoina käytetään hallinnollisia rekistereitä. (Ks. Kaavio 2.5).

Kehikon sisältämien tietojen tulee olla ajan tasalla. Rekistereitä käytettäessä osite- ja luokittelutietojen tulee olla päivitettyjä viimeisten mahdollisten tietojen mukaisiksi. Rekistereitä yhdisteltäessä otantakehikoksi tulee yhdistelytietojen kuten esimerkiksi henkilö- ja kotitaloustutkimusten kohdalla kohdehenkilöiden henkilötunnusten olla tarkistettuja. Rekistereihin sisältyviä tietoja tulee käyttää otoksen kohdentamiseen tutkimuksen kannalta tärkeisiin perusjoukon osajoukkoihin. Otoksen kohdentamisessa voidaan käyttää ositettua otantaa, kaksivaiheista otantaa tai otantaa vaihtelevin todennäköisyyksin.

Hallinnolliset rekisterit otantakehikkoina sisältävät myös muita alkioita kuin niitä jotka kuuluvat tavoiteperusjoukkoon. Näitä alkioita kutsutaan ylipeitoksi ja ne ovat yleensä identifioitavissa rekisteristä poimintavaiheessa. Hallinnollisissa rekistereissä voi olla ali- ja ylipeittoa. Otantakehikon ylipeitto voidaan mitata, mutta ali-peittoa on vaikeampi arvioida (Djerf, 2000). Alipeittoa varten tulee selvittää muut mahdolliset lähteet, joista kehikosta puuttuvat alkiot voidaan löytää.

Yrityksiin tai toimipaikkoihin kohdistuvissa tutkimuksissa perusjoukossa tapahtuvat muutokset ovat suurempia kuin henkilöihin ja kotitalouksiin kohdistuvissa tutkimuksissa. Otantakehikon ajantasaisuuteen on siten kiinnitettävä erityisesti huomiota. Yrityksiin suunnatuissa tutkimuksissa yrityksen koko otetaan yleensä huomioon otosta poimittaessa. Koko voidaan huomioida osituksella tai poimimalla yrityksiä suhteessa niiden kokoon. Suurimmat yritykset sijoitetaan usein samaan ositteeseen, josta kaikki alkiot tulevat mukaan otokseen. Päivitetty kehikkoperusjoukko tulisi muodostaa mikäli mahdollista estimointivaihetta varten, jolloin saadaan myös lisävalaistusta alipeitosta.

Kaavio 2.5 Tilastotutkimuksen tavoiteperusjoukko, kohdeperusjoukko, ja kehikoiden ali- ja ylipeitto



Kun tutkimuksen tavoiteperusjoukko on määritelty, tulee kohdeperusjoukon täsmällistä määrittystä varten tarkastella otantakehikkoa ja sen mahdollista ylipeittoa. Joissain tilastotutkimuksissa rajataan tutkimusyksiköitä kohdeperusjoukon ulkopuolelle:

- **maantieteellisin perustein:**
Joissain kotitalousväestön haastattelututkimuksissa saatetaan jättää tutkimuksen ulkopuolelle Ahvenanmaa ja ulkosaariston kuntia.
- **yksikkökohtaisin demograafisin perustein:**
Tilastotutkimus saattaa kohdentua henkilöihin tietyn demograafisen taustatekijän mukaan, esimerkiksi täysi-ikäiseen väestöön; vastaavasti yrityspohjainen tilastotutkimus voi keskittyä esimerkiksi tiettyjen toimialojen yrityksiin.
- **ryhmäkohtaisin perustein:**
Jos tilastotutkimuksen yksikkö on eri tyyppinen kuin kehikon, tulee tutkimusyksikkö määritellä yksiselitteisesti. Tällöin ne yksilöt, jotka selkeästi eivät ole tilastotutkimuksen yksikön määritelmän piirissä tulee poistaa kohdeperusjoukosta. Esimerkiksi laitosväestöön kuuluvat henkilöt pyritään poistamaan kotitalousväestöön kohdistuvasta tilastotutkimuksen otantakehikosta. (Rossi, Wright ja Anderson, 1983).

Suositukses

Ohjeita otantakehikon valintaan

- Selvitä mitä rekistereitä on käytettävissä otantakehikoiksi, mitä niiden käyttö maksaa ja mitä rekistereitä voidaan yhdistellä tutkimuksen kannalta parhaan kehikon luomiseksi.

- Henkilörekisteri soveltuu henkilöotosten poimimiseen. Henkilörekisteristä voidaan poimia myös kotitalousotokset. Tällöin tutkimukseen tulee se kotitalous, johon poimittu henkilö kuuluu. Henkilörekisteristä voidaan muokata myös asuntokunnat käsittävä rekisteri, josta voidaan poimia kotitalouksia, kun kotitaloutta approksimoidaan asuntokunnalla.
- Yritysootosten otantakehikkoina käytetään yritysrekisteriä tai siihen perustuvia tilastotuotantoon tehtyjä tiedostoja. Yritysrekisteri perustuu verohallinnon aineistoihin. Muutokset hallinnollisessa rekisterissä ovat hallinnollisia tapahtumia, jotka yritysrekisterissä on muutettu tilastollisiksi tapahtumiksi. Oikeudelliset yksiköt on tarvittaessa muutettu tilastollisiksi yksiköiksi.
- Maatilaotosten kehikkona käytetään maatilarekisteriä.
- Laske otoksen poiminnan jälkeen apumuuttujien arvot kehikosta ja otoksesta ja tarkista, että otoksen poiminta on tapahtunut oikein. Jos havaitset poikkeuksellisen suuria satunnaisvirheitä, niin poimi uusi otos ja toista tarkistukset. Kerää mahdollisimman paljon aputietoa ja tallenna ne myöhempää käyttöä varten.
- Arvioi rekisterin tilaa ja virheitä kerätyn aineiston avulla. Raportoi tietojenkeruuvaiheessa havaittu (i) ylipeitto, (ii) rekisteristä saatujen luokittelumuuttujien virheet ja (iii) arvioi alipeittoon kuuluvien määrät muita tietolähteitä käytäen. Seuraa haastattelihoilta tai muuten tulevaa palautetta osoitetietojen ajantasaisuudesta ja luokittelumuuttujien oikeellisuudesta.
- Raportoi kohdeperusjoukko, otantakehikko ja sen peittävyys tutkimustulosten yhteydessä.

Kirjallisuus

- Djerf, K. 2000. *Properties of Some Estimators under Unit Nonresponse*. Statistics Finland. Research Reports 231. Helsinki: Hakapaino Oy.
- Kish, L. 1965. *Survey Sampling*. New York: John Wiley & Sons.
- Levy, P. S. & Lemeshow, S. 1991. *Sampling of populations: methods and applications*. New York: John Wiley & Sons.
- Rossi, P.H., Wright, J.D., Anderson, A.B. 1983. *Handbook of Survey Research*. London: Academic Press.

2.4 HALLINNOLLISET REKISTERIT JA TILASTOREKISTERIT

Soveltamisala ja päämäärä

Virallisen tilastotuotannon kannalta hallinnollisella aineistolla tarkoitetaan mitä tahansa aineistoa, joka on ensisijaisesti kerätty muuta tarkoitusta kuin virallista tilastotuotantoa varten. Tämän määritelmän puitteissa myös muut tahot kuin viranomaiset voivat kerätä ja pitää hallinnollisia aineistoja. Tämä määritelmä kiinnittää huomion siihen, mikä virallisen tilastotuotannon kannalta on tärkein hallin-

nollisten aineistojen ominaisuus: koska tiedot on kerätty muuta tarkoitusta varten, niiden käyttäminen ei lisää tiedonantorasitusta. Kustannukset tilastojen tuottajalle ovat myös hyvin pienet verrattuina suoraan keruun. Nämä edut ovat niin merkittäviä, että ne yleensä painavat enemmän kuin hallinnollisten aineistojen käyttöön liittyvät haitat.

Hallinnolliset rekisterit ovat hallinnollisten aineistojen osajoukko. Jos hallinnollinen aineisto koostuu yksikötason tiedoista, sitä voi kutsua hallinnolliseksi rekisteriksi. Tämän määritelmän puitteissa voimme puhua rekistereistä, vaikka rekisteritiedot olisi tallennettu relaatiotietokantaan. Vaikka relaatiotietokannassa olevat tiedot eivät tietotekniikan mielessä muodosta rekisteriä, niin voimme sanoa, että tietokantaan sisältyy yksi tai useampi looginen rekisteri.

Tilastotuotannon kannalta hallinnolliset rekisterit ovat ylivoimaisesti tärkeimmät hallinnolliset aineistot. Tilastorekisterillä tarkoitetaan rekisteriä, jota ensisijaisesti ylläpidetään tilastotuotannon tarpeita varten. Valtaosassa tapauksia tilastolaki itse asiassa kieltää tilastorekisterin käyttöä mihinkään muuhun tarkoitukseen kuin tilastotuotantoa varten. Tärkeimmät tilastorekisterit perustuvat hallinnollisiin rekistereihin. Tilastorekisteriin on tällöin kerätty tilastotuotannon kannalta relevantteja muuttujia yhdestä tai useammasta hallinnollisesta rekisteristä.

Keskeinen käsite rekisterien käytössä on tunnusjärjestelmä, jolla tarkoitetaan yksikkötietoihin liitettävää tunnusta, jonka avulla ko. yksikkö voidaan yksikäsitteisesti tunnistaa. Tunnusjärjestelmää voidaan kutsua yleiseksi tunnusjärjestelmäksi, jos sitä käyttää useimmat ja keskeisimmät rekisterinpitäjät. Yleinen tunnusjärjestelmä, kuten henkilötunnus tai yritys- ja yhteisötunnus, mahdollistaa rekisterien yhdistelyn sekä suoran keruun aineistojen ja rekisteritietojen yhdistelyn.

Rekisteri

Tietorekisteri on yleensä konekielisessä muodossa oleva yksikötasoinen, jonkin ryhmän tai joukon kokonaisuudessaan käsittävä (totaalinen) tietoaaineisto, jota pidetään säännöllisesti ajan tasalla eli päivitetään muutostiedoilla.

Suurin osa rekistereistä on hallinnollisia eli viranomaisten toimintaa palvelevia: kuten esimerkiksi väestökirjanpito, verotusrekisteri, työeläkkeiden, työnhakijoiden, eläkeläisten ja varusmiesten rekisteröinti. Vain murto-osa rekistereistä on perustettu alunperin tilastollisiin tarkoituksiin. Näistä voidaan mainita yritys- ja toimipaikkarekisteri, siihen liittyvät valtion virastot ja toimintayksiköt kattava rekisteri, kuntien toimintayksikkörekisteri. Tutkintorekisteri kattaa väestön suorittamat perusasteen jälkeiset tutkinnot.

Keskeinen tietorekisteri on 1960-lopulla perustettu väestön keskusrekisteri, joka käsittää kaikki ne henkilöt, joilla on (tai on ollut) vakinaisen asuinpaikka Suomessa. Rekisteriä täydennetään päivittäin väestön muutostiedoilla eli syntyneillä, kuolleilla, maahan- ja maastamuuttaneilla, maan sisällä muuttaneilla, vihkimisillä, avioeroilla, adoptioilla ja nimenmuutoksilla. Väestön keskusrekisteri on kihlakuntien paikallisrekistereiden yhdistelmä ja sen päivitys hoidetaan kihlakuntien rekisteritoimistoissa.

Usein myös muita konekielisiä tiedostoja kutsutaan rekistereiksi, vaikka ne käytännössä ovat poikkileikkaustilanteesta kerättyjä tiedostoja, eikä niitä päivitetä muutostiedoilla. Esimerkkinä voidaan mainita niin kutsuttu yhteisvalintarekisteri, joka käsittää yhden vuoden aikana ammatillisiin oppilaitoksiin sekä lukioihin pyrkineet ja otetut henkilöt. Seuraavana vuotena tämä rekisteri koostuu suurimaksi osaksi eri henkilöistä, joten se oikeastaan on vain konekielinen tiedosto. Vastaavanlaisia tiedostoja ovat useat Tilastokeskuksen vuosiaineistot, vaikka niitä saatetaan kutsua rekistereiksi.

Rekisterin ja tiedoston välimuotona voidaan pitää myös 1960-luvun lopulla perustettua verotusrekisteriä (sitäkin kutsutaan aina rekisteriksi), vaikka se muodostuu kunkin vuoden verovelvollisista henkilöistä ja heidän yhden vuoden tuloja ja varallisuustiedoistaan. Tiedostoa päivitetään väestörekisterin muutoksilla, kuten kuolleilla, muuttaneilla ja perhetiedoilla, mutta verotustiedot kerätään joka vuosi uudelleen tulonsaajilta itseltään, heidän työnantajiltaan ja eläkelaitoksiltaan tms. Edellisen vuoden tietoja ei siten päivitetä. Osaa verotustiedoista säilytetään rekisterissä: varallisuustietoja käytetään mahdollisen omaisuusveron ja myyntivoiton määräämiseksi. Niitä ei tarvitse ilmoittaa uudelleen joka vuosi, ellei muutoksia ole tapahtunut.

Yritystilastojen kannalta tärkeimmät rekisterit ovat veroviranomaisten pitämät rekisterit. Ensimmäiset yritystoiminnan verotukseen tarkoitetut rekisterit kehittyivät 1970-luvulla. Tuolloin saatiin mm. liikevaihtoverovelvolliset rekisteröityä. 1980-luvulla verottaja uudisti laajalti yritysverotuksen tietojärjestelmiä, vei ne sähköisen muotoon ja otti käyttöön liike- ja yhteisötunnuksen (nykyinen yritys- ja yhteisötunnus). Näiden uudistusten jälkeen Tilastokeskus pystyi luomaan lähes kattavan yritysrekisterin, jossa yritysten aloitus- ja lopettamistietoja voitiin ylläpitää lähes ajantasaisesti. Yritysverotuksen rekisterien hyödyntäminen tilastotöissä keskittyi 1990-luvulle saakka yritysrekisteriin. 1990-luku oli suuren edistyksen aika hallinnollisten aineistojen hyödyntämisessä yritystilastoissa. Ensin tuli yritysten rakennetilastojen uudistus 1995 – 1996, jolloin elinkeinoverotusrekisteristä tehtiin keskeinen rakennetilastojen tietolähde. Vuonna 1998 käynnistettiin yritysten liikevaihtoa ja palkkasummaa koskevien kuukausitilastojen tuotanto. Nämä perustuvat verottajan maksuvalvontarekisteriin, joka sisältää arvonlisäverotusta ja rekisteröityjen työnantajien maksamia palkkoja koskevia tietoja.

Periaatteet

Rekisteriyksiköiden tunnukset

Rekisteriyksiköiden tunnustiedot identifioivat rekisteriaineistojen yksiköt. Tunnustietojen avulla eri rekistereiden yksiköt voidaan yhdistää toisiinsa: esimerkiksi henkilöt asuntoihin, asunnot rakennuksiin, rakennukset kiinteistöihin ja työlliset yrityksiin. Henkilörekistereissä tunnistamisvälineenä käytetään henkilötunnusta; rakennus- ja huoneistorekisterissä kiinteistö-, rakennus- ja huoneistotunnusta; yritysrekistereissä toimipaikka-, yritys - tai yhteisötunnusta ja ajoneuvorekisterissä auton rekisteritunnusta.

Henkilöille voidaan yhdistää henkilötunnuksen avulla myös tietoja muista rekistereistä kuten esimerkiksi työpaikan toimiala, työllisyys- ja/tai työttömyysjaksot, sairausjaksot, eläkejaksot, opiskelujaksot, suoritettut tutkinnot, tulot, varat, velat, toimeentulotukien saamiset, auton hallinta, sekä rakennusten, asuntojen ja kesämökkien omistukset.

Tärkeimmät tilastotuotannossa käytetyt rekisterit

- väestötietojärjestelmä, joka sisältää väestötiedot, rakennus- ja huoneistotiedot, toimilatiatiedot ja kesämökkitiedot (Väestörekisterikeskus VRK)
- verotuksen asiakasrekisteri (asiakastietokannasta), maksuvalvontarekisteri (sisältää kuukausittaiset arvonlisävero- ja työnantajasuoritustiedot), verotusrekisterit (mm. elinkeinoverotusrekisteri), tulo- ja varallisuustiedot, työnantajan vuosivalvontatiedot, liike- ja yhteisötunnusrekisteri (jota ollaan siirtämässä verohallituksen ja patentti- ja rekisterihallituksen yhteisesti ylläpitämään yritys ja yhteisötietojärjestelmään (YTJ)) (Verohallitus Veroh)
- kaupparekisteri (Patentti- ja rekisterihallitus)
- työeläkejärjestelmien työsuhdetiedot (Eläketurvakeskus ETK, Valtiokonttori, Kuntien Eläkevakuutus sekä muutamia muita pieniä tiedontuottajia)
- työnhakijarekisteri, työvoimapolitiittinen koulutus (Työministeriö)
- vanhuus-, työkyvyttömyys- ja työttömyyseläkeläiset, asumistukitiedot, sairaustiedot (Kansaneläkelaitos)
- varusmiespalveluksessa olevat (Pääesikunta)
- opiskelijarekisterit (Tilastokeskus TK ja muut)
- ajoneuvorekisteri (Ajoneuvorekisterikeskus ARK)
- toimeentulotukea saaneiden rekisteri (Sosiaali- ja terveysalan tutkimus- ja kehittämiskeskus Stakes)
- hoitoilmoitusrekisteri (Stakes)
- tutkintorekisteri (TK)
- yritys- ja toimipaikkarekisteri (ml. valtion ja kuntien toimintayksiköt) (TK).

Rekisteritietojen käyttö tilastotuotannossa

Tavallisesti rekistereitä käytetään tilastotuotannossa sekä (i) kantatilastoissa että (ii) muutostilastoissa. Esimerkiksi väkiluku ja väestön rakennetilastot tuotetaan rekisterin kantatiedoista. Väestömuutostilastot (syntyneet, kuolleet, maassa muuttaneet, maasta muuttaneet, maahan muuttaneet, vihityt ja eronneet) perustuvat rekisterin päivitystietoihin. Samoin rakennus- ja huoneistorekisteristä saadaan kantatiedot ja rakennustuotantotiedot perustuvat rekisterin ylläpitojärjestelmään. Säännöllistä kantatietojen ylläpitoa tapahtuu myös ajoneuvorekisterissä, tutkintorekisterissä, yritys- ja toimipaikkarekisterissä ja työnhakijarekisterissä, joista tuotetaan kantatilastoja ja muutostilastoja.

Verotus-, opiskelija-, ja toimeentulotukea saaneiden rekisterit ovat oikeastaan vuositiedostoja, joissa esimerkiksi kaikkien tuloa saaneiden tiedot on koottu kullakin vuodelta erikseen. Varsinaista ylläpitoa ei siten tapahdu eikä myöskään laadita verotuksen muutostilastoja. Samoin opiskelijarekisteri sisältää kunakin vuonna

oppilaitoksissa kirjoilla olevat. Muutostilastoksi voidaan ajatella joka vuosi julkaittavia tilastoja pyrkineistä, valituista ja aloittaneista ja tutkinnon suorittaneista.

Vuosittain laadittavat yritystoiminnan rakennetilastot laaditaan yhdistelemällä suoran tiedustelun kautta saatuihin tietoihin elinkeinoverotusrekisterin ja yritysrekisterin tietoja. Suora keruu kohdistetaan vain muutamalle prosentille yrityksiä.

Kuukausittain laadittavat liikevaihto- ja palkkasummaindikaattorit laaditaan yhdistelemällä suoran tiedustelun kautta saatuihin tietoihin arvolisävero- ja työnantajasuoritusrekisterin ja yritysrekisterin tietoja. Suora keruu kohdistetaan vain vajaan yhden prosentin yrityksistä.

Yritystoiminnan rakennetilastojen ja kuukausi-indikaattoreiden tiedon keruun arkkitehtuurin samankaltaisuudesta huolimatta metodologiset haasteet näissä kahdessa tilastossa ovat erilaisia. Rakennetilastoissa imputoinnilla on keskeinen asema metodologiassa jolla hallinnollisen aineiston puutteet korjataan siten, että ne kelpaavat tilastolliseen käyttöön. Kuukausi-indikaattoreissa paneelimenetelmät on keskeinen vastaus ajantasaisuuden ja alipeiton ongelmiin.

Suosituks

Rekisterien hyväksikäytön edellytykset laskennoissa ja muissa tutkimuksissa

Rekisterin on täytettävä seuraavat edellytykset, jotta sitä voidaan käyttää tilastotuotannossa. Kaikki yksittäiset edellytykset eivät aina ole ehdottoman välttämättömiä, vaan niitä tulee aina harkita muiden edellytysten ja vaihtoehtoisten keruumenetelmien kannalta.

Hallinnollisista aineistoista voisi yleisesti sanoa, että ne laatutekijöiden suhteen tuovat järjestään haasteita tilastojen tekijöille. On myös osoittautunut, että pitkäjänteisellä metodologisella työllä näihin haasteisiin pystytään vastaamaan.

Sisällölliset ja käsitteelliset edellytykset

Kattavuus: Rekisterin tulee kattaa koko kohdejoukkonsa, esimerkkinä maassa-asuva väestö, eläkeläiset, työttömät, ajoneuvot, verovelvolliset ja yritykset. (ks. 1.2.3 Suomen Virallisen Tilaston laatukriteerit). Myös sellaiset osajoukot kuten arvolisäverovelvolliset tai rekisteröidyt työnantajat voivat täyttää kattavuuden kriteerin, koska näiden osajoukkojen merkitys koko ilmiön kannalta tunnetaan. Arvolisäverovelvollisten yritysten liikevaihdon kehitys on siis tilastoinnin kannalta relevantti kohdejoukko.

Yksikkökohtaisuus: Rekisterin perustietojen tulee olla yksikkökohtaisia, henkilöitä, asuntokuntia, perheitä, asuntoja, toimitiloja, rakennuksia, ajoneuvoja, yrityksiä. Oman haasteensa tuo yritysten fuusiot ja diffuusiot. Niiden käsittely hallinnossa ei välttämättä vastaa tilastotoimen vaatimuksia.

Vastavuus käyttäjien tarpeisiin: Rekisteritietojen tulee olla relevantteja tilastollisen käytön kannalta ja niiden tulee kattaa esimerkiksi osa väestölaskennan tietosisällöstä ja vastata Tilastokeskuksen tilastollisia käsitteitä kuten: työllinen, työtön ja eläkeläinen. (ks. 1.2.3 Suomen Virallisen Tilaston laatukriteerit). Usein hallinnollisten aineistojen tietoihin on suhtauduttava kuten korvikemuuttujiin. Esi-

merkkejä: elinkeinoverotusilmoituksista saatavat investointeja ja poistoja koskevat tiedot eivät vastaa käsitteellisesti eikä luokituksen kannalta rakennetilastojen määritelmiä ja luokituksia. Tällöin tiedot on muunnettava joidenkin oletusten (teorian) pohjalta vastaamaan tilastotietojen tarpeita. Vastaavasti arvonlisäverotuksesta saatavat myynnin määrä koskevat tiedot sisältävät käyttöomaisuuden myyntejä, eivätkä siten vastaa käsitteellisesti liikevaihdon määritelmää.

Luotettavuus

Peittävyys: Rekisterin tulee kattaa kohdejoukkonsa mahdollisimman hyvin (ks. 2.3 Kehikot ja peittävyys). Ehdotonta peittävyysprosenttia ei voida määritellä, koska mikään systeemi ei ole täysin kattava ja useimmissa systeemeissä on myös ylipeittoa toisin sanoen yksiköitä, jotka eivät kuuluisi perusjoukkoon. Rakennusrekisterissä on tuhoutuneita rakennuksia. Väestölaskentojen kattavuus vaihtelee maittain jossain 95 ja sadan prosentin välillä. Jos laskennassa käytettyjen rekisterien peittävyys on tätä heikompi, on käyttö ongelmallista. Viime kädessä peittävyysvaatimukset esittävät tietojen käyttäjät, tulosten tulee olla käytettävyyden kannalta tyydyttäviä.

Tietojen luotettavuus: Yksittäisten tietojen tulee olla mahdollisimman luotettavia (ks. 1.2.3 Suomen Virallisen Tilaston laatukriteerit). Tämä koskee sekä tunnistietoja että ominaisuustietoja. Myös rekisterin päivityssysteemin tulee olla kattava. Esimerkiksi väestötilastosysteemi saa varsin luotettavasti tiedot syntyneistä, kuolleista, vihityistä eronneista, mutta muuttoilmoituksia laiminlyödään ja esimerkiksi ammattitieto päivitetään vain muuttoilmoituksen yhteydessä, perheellisiin rinnastettavat avoliittolaiset joudutaan tilastollisesti päättelemään ja asunnoissa tapahtuu merkittäviä muutoksia ilman rakennuslupia.

Kun yhdistetään esimerkiksi työpaikka-, ammatti- ja tulotiedot eri rekistereistä, ei voida aina olla varmoja että ammatti ja toimiala kuvaavat samaa työsuhdetta. Tällöin voi syntyä niin sanottu konsistenssiongelmaa.

Yleensä viranomaistyönä tehdyt ilmoitukset sisältävät tarkan ajankohdan. Henkilöiden omiin ilmoituksiin perustuvat muutostiedot ovat jossain määrin subjektiivisia: esimerkiksi avoliitosta voi yhdessä asuvilla olla erilainen käsitys, vähäistä työtä tekevä opiskelija ilmoittaa itsensä opiskelijaksi vaikka työllisen määritelmä täyttyy (palkattua työtä tehdään vähintään yksi tunti viikossa) ja työvoimakäsittelyn mukaan hänen ehdottomasti kuuluisi ilmoittaa itsensä iltasiivoojaksi. Rekisteritietoja käytettäessä vastaajien subjektiivisten käsitysten vaikutus pienee. Kaikki opiskelijat, joilla on vähäininkin työsuhde merkitään työllisiksi ja perheen viitehenkilöksi tulee aina suurituloisin.

Elinkeinoverotusta koskevat tiedot saattavat joiltakin yrityksiltä puuttua kokonaan ja useimmissa on jotain virheitä. Virheitten olemassaoloa voidaan testata tarkastamalla menevätkö tilit umpeen. Epäkelpojen tilien joukosta voidaan tämän jälkeen etsiä yleisiä virhetyyppejä, joita voi, sen jälkeen kun ne on ymmärretty, korjata koneellisesti. Tämä työ tuo tulosta, kun sitä tehdään sinnikkäästi ja pitkäjänteisesti. Vuoden 1995 yritysten tilinpäätöstiedoista saatiin vain noin 55 prosenttia menemään korjausten jälkeen umpeen. Vuoden 1999 tiedoista 90 prosenttia saatiin menemään umpeen. Niiltä osin kuin elinkeinoverotuksen tilinpä-

töstiedot puuttuvat tai niitä ei saada menemään umpeen, ne joudutaan imputoimaan proportionaalisesti liikevaihdon suhteen, koska liikevaihtotieto on yritysrekisteristä saatavissa kaikille yrityksille. Pitkäjänteisellä työllä imputointia voidaan parantaa huomioimalla mm. toimialan ja yrityksen suuruusluokan yhä sofistikoituneimmilla menetelmillä.

Ajantasaisuus

Tietojen tulee olla ajan tasalla (ks. 1.2.3 Suomen Virallisen Tilaston laatukriteerit). Nopeimmat rekisterit ovat käytettävissä muutaman viikon tai parin kuukauden kuluttua viiteajankohdasta (ajoneuvorekisteri, työnhakijarekisteri, väestön keskusrekisteri), hitaimpien valmistumiseen kuluu jopa vuosi (esimerkiksi verotusrekisteri). Ajantasaisuuteen vaikuttaa se kuinka kauan kuluu aikaa ennen kuin tapahtuma rekisteröidään ja kuinka usein tietoja rekisteriin päivitetään. Väestötietoja päivitetään joka päivä – tulotietoja ”päivitetään” kerran vuodessa, mutta jos muuttaja laiminlyö muuttoilmoituksen antamisen rekisterin tiedot eivät päivity. Rakennuksen valmistumisajankohta eli rakennuksen ottaminen rekisteriin on ongelmallinen käsite, rakennuksessa on saatettu jopa asua jo pitkän aikaa ennen kuin se loppukatselmuksessa todetaan ”valmistuneeksi rakennukseksi”.

Koska hallinnolliset aineistot määritelmällisesti kerätään muuta tarkoitusta kuin tilastotuotantoa varten, tilastoviranomainen ei voi suoraan vaikuttaa ajantasaisuuteen liittyviin ongelmiin. Tilastotuotannon on kehitettävä menetelmiä, joilla puutteet ajantasaisuudessa voidaan hallita. Näin on esimerkiksi yritysten kuukausittaista liikevaihtoa ja maksettua palkkasummaa koskevissa tilastoissa. Kun verottajan annetun kuukauden aineistot saapuvat ensimmäistä kertaa tilastoviranomaiselle, ne kattavat vain noin 70 % lopullisesta määrästä esimerkiksi arvonlisäveroilmoituksia. Yksittäisen yrityksen kohdalla on ongelmallista tietää, puuttuuko ilmoitus sen vuoksi, että se on myöhässä vaiko sen vuoksi, että yritys on lopettanut toimintansa. Tämän vuoksi annetun kuukauden tietoja pyydetään tilastotoimen käyttöön toistuvasti kuuden kuukauden ajan, kunnes ne ovat kattavia. Tämä ei kuitenkaan vastaa ajantasaisuuden haasteeseen. Kuukausi-indikaattoreissa paneelimenetelmä oli ensimmäinen vastaus tähän ajantasaisuuden haasteeseen. Pitkäjänteisellä metodologisella työllä on pystytty parantamaan edelleen näiden tilastojen laatua.

Tekniset edellytykset

Dokumentointi: Rekistereistä tulee olla riittävän hyvät ja yksityiskohtaiset kuvaukset, joiden perusteella käyttäjä saa kuvan rekisterin toiminnasta, rakenteesta ja tietosisällöstä. (Ks. 3. Dokumentointi).

Automaattinen tietojenkäsittely: Rekisterin tulee olla konekielisessä muodossa, jotta sitä voidaan käyttää hyväksi tilastotuotannossa. Kortistomuotoisia manuaalisesti käsiteltäviä rekistereitä voi olla esimerkiksi yritysten työntekijärekisterit ja yhdistysten jäsenluettelot. Niillä harvoin on tilastollista käyttöä.

Tunnukset: Rekisterin yksiköillä on tärkeätä olla yleisesti käytössä olevat tunnukset. Tunnetuin on henkilötunnus, joka sisältää oikean kirjoittamisen varmistavan tarkistusmerkin. Se on lähes muuttumaton, käyttäjien tuntema ja laajalti eri

rekistereihin levinnyt tunnus sekä täysin yksikäsitteinen. Osittain tunnusten puuttuminen voidaan korvata nimi- ja/tai osoitetunnistuksella (esimerkiksi yhdistettäessä niin sanottuja eläkejärjestelynumeroita LY-tunnuksiin).

Luokiteltavuus: Ominaisuustietojen tulee olla kooditettuina tai numeerisessa muodossa, jotta niitä voidaan luokitella. Jos tieto on selväkielinen, kuten esimerkiksi ammattinimike, se voidaan koodittaa automaattisen koodituksen avulla. Osa selväkielisistä nimikkeistä vaatii manuaalikäsittelyä. (Ks. 2.2 Luokitukset, käsitteet ja määritelmät sekä harmonisointi).

Hallinnolliset edellytykset

Rekisterin pitäjän tulee luovuttaa aineisto Tilastokeskukselle. Tilastolain mukaan Tilastokeskuksella on oikeus saada viranomaisilta ja useilta muilta yhteisöiltä niiden hallussa olevat tiedot tilaston laatimista varten. (Ks. 1.1.1 Lainsäädäntö).

Kustannukset

Rekisterin hyväksikäytön kustannukset eivät saa olla suuremmat kuin toisen vaihtoehdoisen keruutavan kustannukset. Rekisteripohjaisen väestölaskennan kustannukset ovat vain murto-osa lomakekeruun kustannuksista. Osa laskentatiedoista päätellään eri rekisteritietojen avulla (vaikka niitä ei suoraan missään rekisterissä esiintyisikään) kuten päätoiminta laskentaviikolla, perheet, asuntokunnat, perheen viitehenkilö, sosioekonominen asema, työpaikan ja asuinpaikan välinen etäisyys. Tilastokeskus korvaa rekisterin pitäjille yleensä vain tietojen hyväksikäytöstä aiheutuneet lisäkustannukset.

Kirjallisuus

- Alho, J. (toim). 1999. *Statistics, Registries and Science – Experiences from Finland*. Statistics Finland. Keuruu: Otava Book Printing.
- Elvers, E. 1999. Frame errors. Teoksessa: *Model Quality Report in Business Statistics Vol. 1. Theory and Methods for Quality Evaluation*. Assessment of the Quality in Statistics. Doc. Eurostat/Supcom 97/lot6/99/Final Report/Vol.1. Eurostat.
- Myrskylä, P. 1991. Census by Questionnaire - Census by Registers and Administrative Records: The Experience of Finland. *Journal of Official Statistics*. Vol. 7. No 4. 457-474.
- Myrskylä, P. 1996. Virallisten tilastojen tiedonkeruu. Teoksessa. *Tilastoista tiedoksi korkea-asteelle*. Niemi, H. & Tourunen, K. (toim). Tilastokeskus, Helsinki.
- Myrskylä, P. 1999. New statistics made available by the use of registers. *Statistical Journal of the United Nations ECE* 16. 165-180.
- Ragnarsøn, R. 2000. *Utilisation of Administrative Data in the Manufacturing Statistics*. The Norwegian Experience. Meeting of INSEE's and Nordic NSI's Business Statisticians, Helsinki 11-12 May 2000.
- Tilastokeskus. 2001. *Väestölaskenta 2000*. Käsikirjoja 35. Helsinki: Edita Oy. <http://2000.stat.fi/dialog/opas2000.pdf>

2.5 OTANTAMENETELMÄT

Soveltamisala ja päämäärä

Otantaa käytetään yhteiskunnallisissa tilastotutkimuksissa kokonaistutkimuksen asemesta kustannusten säästämiseksi. Oikein tehtynä otannalla saadaan käyttö-tarkoituksen kannalta tarkat estimaatit.

Edustavana otantamenetelmänä satunnaisotanta on 1900-luvulla syrjäyttänyt harkinnanvaraisen (*purposive*) otannan. Satunnaisotannan etuna on sen yleistettävyys perusjoukkoon sekä mahdollisuus laskea estimaattien luotettavuus. Yritysotoksissa käytetään vielä usein katkaisuoitosta, koska oikealle vinoissa populaation jakaumissa suurten yksiköiden vaikutus kokonaisestimaattiin on yleensä dominoiva. Katkaisuoitoksissa vain suurimmat yksiköt tutkitaan. Otantamenetelmän valintaan vaikuttavat muun muassa tutkimusasetelma, perusjoukon ominaisuudet ja käytettävissä olevat kehikot, tiedonkeruutapa ja kustannukset.

Otantamenetelmät

Satunnaisotanta

Otannan perusmenetelmä on yksinkertainen satunnaisotanta. Yksinkertaisessa satunnaisotannassa jokaisella perusjoukon alkiolla on poimintahetkellä yhtä suuri todennäköisyys tulla poimituksi otokseen. Helppo tapa yksinkertaisen satunnaisotoksen poimimiseksi on käyttää satunnaislukuja, joiden avulla tapahtuvassa poiminnassa numeroidaan perusjoukon alkiot ja otetaan otoskokoja vastaava määrä satunnaislukuja ykkösen ja perusjoukon koon väliltä. Tietokonetta käytettäessä toinen yksinkertainen tapa otoksen poimimiseksi on lajitella perusjoukko satunnaisjärjestykseen ja otokseen tulee otokseen mukainen määrä alkioita tiedoston alusta.

Otoksen kokoon vaikuttaa se miten tarkkoja tulosten halutaan olevan, kuinka pienistä perusjoukon osajoukoista halutaan tietoja eli taulukoitavien muuttujien luokittelusta, perusjoukon heterogeenisyydestä eli perusjoukon vaihtelusta. Luonnollisesti käytettävissä olevat taloudelliset resurssit vaikuttavat otoskoko. Yleensä tarkkuus paranee otoksen suuretessa, koska estimaattien keskivirheet pienenevät. Otoksen kasvaessa otantavirhe ei pienene otoksen koon suhteessa vaan hitaammin.

Ryväsotanta

Rypästä voidaan tutkia kaikki alkiot tai sen sisällä voidaan poimia otos tutkittavista alkiosta. Edellisessä tapauksessa on kyse yksiasteisesta ryväsotannasta ja jälkimmäisessä kaksiasteisesta ryväsotannasta. Ryväsotonnan esimerkkinä on henkilöiden työoloja koskeva tutkimus, jossa otokseen poimitaan yrityksiä, joiden työntekijöitä tutkitaan. Yritys muodostaa rypään ja tutkimusyksikkönä on työntekijä. Jos kaikki perusjoukkoon kuuluvat henkilöt tutkitaan niin kyseessä on yksiasteinen ryväsotanta. Jos yrityksen sisällä tutkimukseen poimitaan vain osa työntekijöistä niin kyseessä on kaksiasteinen ryväsotanta.

Ositettu otanta

Ositetussa otannassa perusjoukko jaetaan käytettävissä olevien muuttujien mukaan homogeenisiin osiin. Vinosti jakautuneille perusjoukoille ositus on välttämättömä yksinkertaista satunnaisotantaa käytettäessä. Perusjoukon ositus edellyttää, että kaikille perusjoukon yksiköille on käytettävissä tietoja, joiden perusteella ositus voidaan tehdä. Ositteisiin jaossa kukin perusjoukon yksikkö kuuluu vain yhteen ositteeseen. Eri ositteissa voidaan käyttää tarvittaessa erilaisia poimintamenetelmiä ja poimintasuhteita. Joskus on tarkoituksenmukaista tehdä tietyissä ositteissa kokonaistutkimus kun taas toisissa ositteissa käytetään otantaa. Väestön kohdistuvissa tutkimuksissa tavallisia ositusmuuttujia ovat esimerkiksi läänit sekä henkilön ikä ja sukupuoli.

Systemaattinen otanta

Systemaattisessa otannassa poimitaan yksi alkio tasavälein läpi koko otantakehikon. Kun perusjoukosta valitaan joka k :s alkio, niin poimintaväli on tällöin k . Poiminta aloitetaan arpomalla k :n ensimmäisen alkion joukosta ensimmäinen ja tämän jälkeen otokseen poimitaan joka k :s alkio.

Perusjoukko voi olla sopivassa järjestyksessä. Kun perusjoukko on satunnaisjärjestyksessä, niin saadaan yksinkertaista satunnaisotosta vastaava otos. Jos perusjoukossa esiintyy jaksollisuutta, otos saattaa muodostua samankaltaisista, muusta joukosta poikkeavista alkiosta, jolloin saadaan harhaiset tulokset. Joissain tapauksissa perusjoukon sopivalla järjestämisellä voidaan saada tarkempia tuloksia kuin käyttämällä satunnaisjärjestystä. Jos tutkitaan ilmiötä, jolla on alueellista riippuvuutta, ja tutkittavan kohteena on Suomen väestö ja otantakehikkona käytetään väestön keskusrekisterin hakemistoa, jossa henkilöt ovat kotipaikkatunnuksen mukaisessa järjestyksessä, saadaan alueellisesti oikein jakautunut otos, mikä tällöin on parempi kuin yksinkertainen satunnaisotos.

PPS-otanta (*Probabilities Proportional to Size*)

Jos otos poimitaan hyvin vinosta populaatiosta, jossa yksiköiden koko vaihtelee paljon, voidaan otoksen tarkkuutta parantaa antamalla kullekin yksikölle sisällymistodennäköisyys, joka on suhteessa yksikön kokoon. Tällöin voidaan käyttää palauttaen-tyyppistä tai palauttamatta-tyyppistä pps-otantaa.

Suositukses

Yksinkertainen satunnaisotanta on suositeltava perusmenetelmä monialaisissa empiirisyhteiskunnallisissa tilastotutkimuksissa, joissa otosta ei voi kohdentaa perusjoukon osajoukkoihin. Yksinkertaista satunnaisotantaa käytetään yleisesti henkilöihin ja kotitalouksiin kohdistuvissa kyselyissä.

Kaksivaiheista otantaa käyttämällä verorekisteristä saadaan kotitalouden sosioekonomista tilaa kuvaavia ositusmuuttujia ositusta varten kohdentamaan otosta haluttuihin väestöryhmiin.

Rekisteristä poimittaessa käytännöllinen poimintatapa on systemaattinen otanta. Tällöin otantakehikon lajittelulla saadaan otokseen likimain sama jakauma

kuin kehikossa. Tällä on merkitystä erityisesti haastattelututkimuksissa, jossa kotipaikkatunnuksen mukainen lajittelu tuottaa alueellisesti väestömäärien mukaan jakautuneen otoksen ja kun haastattelijat on sijoitettu väestömääritlään yhtä suuriin alueisiin, niin saadaan jaettua otos tasaisesti haastattelijolle.

Johtuen yritysajakauman vinoudesta yksinkertainen satunnaisotanta ei ole sopiva menettely yritysotoksissa. Niissä yleisimmin käytetty menetelmä on koon mukaan ositettu otanta. Päälinjana on esiositus toimialan ja yrityskoon mukaan. Lisäksi käytetään menetelmiä, joissa sisältymistodennäköisyys määräytyy suoraan yksikön koon perusteella. Suurimmat yksiköt yritysotoksissa yleensä määrätään ositteisiin, jotka sisältyvät otokseen todennäköisyydellä yksi.

Koordinoi säännöllisesti toistuvissa tutkimuksissa peräkkäiset otokset, ettei niissä ole samoja henkilöitä tai kotitalouksia vastausrasitteen vähentämiseksi. Tällä on merkitystä paneelitutkimuksissa, jossa on sama henkilö tai kotitalous on mukana useita kertoja. Koordinointi on tehtävä niin, ettei siitä aiheudu harhaa poimittuun otokseen. Jos siihen ei päästä, niin on käytettävä sopivia oikaisumenetelmiä.

Yritystutkimuksissa ollaan usein kiinnostuneita perusjoukossa tapahtuvista muutoksista, jolloin on välttämätöntä saada riittävä määrä samoja yksiköitä perättäisiin otoksiin. Jottei rasitettaisi liikaa samoja yksiköitä, tulee mahdollisuuksien mukaan kierrättää yksiköitä. Koska yrityspopulaatio on suhteellisen pieni verrattuna henkilöväestöön, sama yksikkö saattaa joutua useampaan perättäiseen tai samanaikaiseen tiedusteluun. Tämän vuoksi kaikkia yhdestä kehikosta poimittuja otoksia pitää koordinoida siten, ettei yhden vastaajan rasite nouse suuremmaksi kuin muille yksiköille kohdistuva vastausrasite.

Kirjallisuus

- Cochran, W.G. 1977. *Sampling Techniques*. 3rd ed. New York: John Wiley & Sons.
- Foreman, E.K. 1991. *Survey Sampling Principles*. New York: Marcel Dekker, Inc.
- Kish, L. 1965. *Survey Sampling*. New York: John Wiley & Sons.
- Lehtonen, R. & Pahkinen, E. 1996. *Practical Methods for Design and Analysis of Complex Surveys*. Revised Edition. Chichester: John Wiley & Sons.
- Levy, P.S. & Lemeshow, S. 1991. *Sampling of Populations: Methods and Applications*. New York: John Wiley & Sons.
- Pahkinen, E. & Lehtonen, R. 1989. *Otanta-asetelmat ja tilastollinen analyysi*. Helsinki: Gaudeamus.
- Sukhatme, P.V. & Sukhatme, B.V. 1970. *Sampling Theory of Surveys with Applications*. Ames, IA: Iowa State University Press.
- Särndal, C.-E., Swensson, B. & Wretman, J. 1992. *Model Assisted Survey Sampling*. New York: Springer-Verlag.
- Teikari, I. 2001. *Controlling the distribution of response burden in longitudinal and cross section business surveys*. Statistics Finland. Research Reports 232. Helsinki: Hakapaino Oy.

2.6 TIEDONKERUUTAVAT HENKILÖPOHJAISESSA JA YRITYSPOHJAISESSA TIEDONKERUUSSA

Soveltamisala ja päämäärä

Tässä luvussa tarkastellaan erilaisten tiedonkeruutapojen ominaisuuksia ja niiden valintaan vaikuttavia tekijöitä. Tiedonkeruutavalla tarkoitetaan sitä kokonaisuutta, missä määritetään tiedonantajien valinta ja miten heiltä kerätään tieto. Tiedonkeruuhankkeita kutsutaan jatkossa tutkimuksiksi. Yrityksille ja luonnollisille henkilöille kohdistettavat tutkimukset eroavat hieman toisistaan johtuen ensisijaisesti vain kohdejoukon luonteesta. Eroavuudet eivät kuitenkaan ole suuria ja siksi tämä tarkastelu koskee ensisijaisesti henkilö pohjaisia tutkimuksia. Lähtökohtaisesti tässä tarkastellaan otostutkimuksia, ei siis kokonaistutkimuksia.

Tiedonkeruutapa vaikuttaa moneen erilaiseen tekijään. Tärkeimpiä näistä ovat sellaiset, kuten tutkimusasetelma, tutkimuskohteiden valinta eli otanta, todennäköinen vastausosuuden (tai kadon) suuruus ja tiedonkeruun kustannukset. Tiedonkeruutapa vaikuttaa myös tutkimuksen aikatauluun ja tulosten laatuun. Vaatimukset vaikuttavat toinen toisiinsa myös monella muulla tavalla. Eri tyyppiset tiedonkeruutavat mahdollistavat varsin erilaisen otoskoon ja tietojen yleistettävyyden eri väestöryhmiin puolestaan riippuu siitä, miten suurelta joukolta tutkittavia tietoja on kerätty.

Tiedonkeruutavan valinta ei ole muista tekijöistä irrallinen toimenpide, vaan se vaikuttaa koko tilastotutkimuksen prosessiin. Esimerkiksi keruutapa vaikuttaa tutkimuslomakkeen suunnitteluun ja toteuttamiseen, kerättävän tiedon määrään ja laatuun sekä tutkimuksen kustannuksiin ja aikatauluun.

Tutkimustyytit

Tutkimustyyppien jaottelussa yleensä huomioidaan kolme tekijää: tutkimusasetelma, tietojen keruutapa ja tietojen kirjaamisen menetelmä.

Kenties merkittävin tutkimusten jaottelu on pitkittäistutkimuksiin ja poikittaistutkimuksiin. Pitkittäistutkimuksille on ominaista, että samoilta tutkimusyksiköiltä kerätään tietoa useamman kerran tietyin väliajoin, mutta poikittaistutkimuksissa tietoa kerätään vain kerran. Pitkittäistutkimuksissa tiedonkeruu voidaan toteuttaa joka kerralla samalla tavalla tai eri kerroilla eri tavoilla. Tietentyyppejä pitkittäistutkimuksia kutsutaan paneelitutkimuksiksi. Työvoimatutkimus on Tilastokeskuksen tunnetuin henkilö pohjainen pitkittäistutkimus ja kuluttajainterventojen keruu tunnetuin yrityspohjainen pitkittäistutkimus.

Lisäksi tutkimukset jaotellaan joko kokeellisiksi tai epäkokeellisiksi eli havaintotutkimuksiksi. Yhteiskuntaa koskevissa tutkimuksissa kokeelliset asetelmat ovat harvinaisia. Ne ovat yleensä paneelitutkimuksia, joissa seurantakertojen välissä on jonkinlainen interventio.

On tärkeää erottaa toisaalta tietojen keruutapa eli se, kerätäänkö tieto haastatteleamalla vai vastaavatko tutkittavat itse, ja toisaalta menetelmä, jolla kerättävä tieto kirjataan eli käytetäänkö paperilla olevaa keruualustaa vai sähköistä tiedon-

keruulinstrumenttia. Tiedonkeruutapoja ovat kyselyt, joissa vastaaja itse täyttää lomakkeen ja haastattelut, joissa haastatteliija esittää kysymykset ja merkitsee vastaukset. Haastatteluissa erotellaan vielä puhelin- ja käyntihaastattelut. Sekä kyselyissä että haastatteluissa tiedot voidaan tallentaa joko paperilomakkeelle tai jollekin sähköiselle tietovälineelle (ks. Taulukko 2.1).

Taulukko 2.1 Tutkimustyytit tiedonkeruutavan ja instrumentin mukaan

	Paperilomake	Tietokonelomake
Kysely	postikysely	Internet-kysely, Tietokoneavusteinen lomakekysely (CASI)
Haastattelu	haastattelulomake	Tietokoneavusteinen henkilöhaastattelu (CAPI), Tietokoneavusteinen puhelinhaastattelu (CATI)

Toimintaperiaate

Lähtökohtaisesti tutkimuksen tavoitteiden tulisi ohjata tiedonkeruutavan valintaa eikä päinvastoin eli tiedonkeruutavan ei pitäisi ohjata tutkimuksen tavoitteenasettelua. Kriteereinä tulisi olla seuraamusvaikutukset, jotka ovat erilaiset eri menetelmillä.

Lähtökohtana tutkimusta suunniteltaessa tulisi olla arvio siitä, miten suuri otos tai vastaajien joukon ylipäättään tulisi olla. Yleisperiaatteena on, että suuremman otoksen tulokset ovat luotettavampia. Toinen merkittävä asia on yhdeltä vastaajikokolta kerättävien tietojen määrä ja laatu. Esimerkiksi, pitkäkestoista haastattelua ei voi tehdä puhelimitse. Vastaavasti henkilökohtainen haastattelu on ainoa mahdollinen tiedonkeruutapa, mikäli vastaajille on näytettävä joitain vastaamiseen liittyviä asioita, kuten esimerkiksi visuaalisia skaaloja.

Usein tavoiteaikataulu eli se, milloin tiedot on oltava käytettävissä, on keskeinen suunnittelun tekijä. Jos tavoiteaika on tiedossa, siitä voidaan laskea taaksepäin eri vaiheiden tavoiteajat. On muistettava, että tutkimukset vaativat usein yhteistyötä monien eri tahojen kanssa. Siksi aikataulu useimmiten vaatii sopimista ja myös sovittamista.

Aikataulun suunnittelu

Aikataulussa on otettava huomioon kaikki tilastotutkimuksen työvaiheet, joita on erilaisissa tutkimuksissa erilainen määrä ja ne painottuvat eri tavoilla. Esimerkiksi sähköisten tiedonkeruulinstrumenttien suunnitteluun on varattava enemmän aikaa kuin paperilomakkeiden suunnitteluun ja painatukseen. Toisaalta hyvin suunnitellun sähköisen tiedonkeruun tuloksena tiedot ovat nopeammin keruuvaiheen jälkeen käytettävissä, koska tallennusta ei tarvita ja tiedot on jo tarkistettu automaattisesti sähköiseen lomakkeeseen ohjelmoitujen tarkistusten ansiosta.

Tutkimuksen aikatauluun vaikuttavat myös muiden yksiköiden aikataulut. Erityisesti haastattelututkimuksissa kenttähaastatteluyksikön ja cati-keskuksen

työtilanne vaikuttaa olennaisesti tiedonkeruun ajankohtaan ja keston. Lisäksi tiedonkeruun laadun kannalta on usein tärkeää, että haastattelijat koulutetaan ja että työohjeet on tehty huolellisesti. Myös muiden resurssien saatavuus (esim. esitarkastus ja tallennus) vaikuttaa aikatauluun.

Tiedonkeruuprosessin keston vaikuttaa keruutapa. Kestoon vaikuttavat monet erilaiset tekijät, mutta yleisesti ottaen tietokoneavusteinen puhelinhaastattelu on käyntihaastattelua ja lomakekyselyä nopeampi tiedonkeruutapa. Lomakekyselyn kesto saattaa joskus olla vaikea ennakoida ja kaikissa tapauksissa se riippuu siitä kuinka monta kertaa vastaamattomilta karhutaan vastauksia.

Vastaamattomuus eli kato

Usein tutkimuksen kesto ja kato ovat keskenään kääntäen riippuvia. Mitä enemmän aikaa ja resursseja käytetään sitä pienemmäksi vastaamattomuus jää. Hyväksytyin kadon suuruus on aina kompromissi. Yhtenä perusteena mainitaan usein, että alhaisista kustannuksista, suuresta vastausosuudesta ja nopeasta aikataulusta vain kaksi voi toteutua samassa tutkimuksessa.

Kato on tyypillisesti erilainen eri tiedonkeruutavoilla. Suurin kato on lomakekyselyissä. Tästä syystä nykyisin pyritään käyttämään haastatteluja. Internet-kyseilyissä kato ja kehikon alipeitto ovat erittäin suuria ongelmia.

Katoon vaikuttavat myös monet muut tekijät. Esimerkiksi kysymysten määrä (so. haastattelun kesto) ja lomakkeen monimutkaisuus lisäävät katoa. Samoin tutkimuksen ajoittuminen esimerkiksi loma-aikaan lisää katoa. Lisäksi tutkimuksen aihepiiri vaikuttaa katoon. Esimerkiksi terveyttä koskevissa tutkimuksissa on tavallista muita tutkimuksia alhaisempi kato. (Katso myös luku 2.8 Vastauskato ja siihen vaikuttaminen).

Kerättävän tiedon määrä ja laatu

Kerättävän tiedon määrä, tai volyyymi, saadaan, kun tutkimusyksiköiden määrällä kerrotaan yhdeltä yksiköltä kerättävän tiedon määrällä. Molemmat tekijät vaikuttavat tutkimuksen keston. Usein kerättävän tiedon määrää tarkasteltaessa keskitytään vain yhdeltä yksiköltä kerättävän tiedon määrään.

Kerättävän tiedon määrä rajaa mahdollisten keruutapojen joukkoa. Hyvänä puhelinhaastattelun pituutena pidetään korkeintaan 30 minuuttia. Käyntihaastattelu sen sijaan voi kestää huomattavasti kauemmin kuin puhelinhaastattelu. Toisaalta hyviä kokemuksia on saatu varsin pitkistä - mutta vastaajan kannalta hyvin motivoitavista - puhelinhaastatteluista. Lomakekyselyissä kato kasvaa nopeasti kysymysten määrän kasvaessa.

Haastatteluissa haastattelijä täyttää lomakkeen vastaajan puolesta ja kyselyissä vastaaja lukee kysymykset (ja ohjeet) ja vastaa ymmärtämänsä perusteella. Haastattelun etuna on se, että haastattelijä tietää, mitä kysytään ja kuinka vastaukset on kirjattava ja samalla huolehtii, että kaikkiin kysymyksiin tulee hyväksyttävä vastaus. Koska kysymyslomake on hänelle tuttu, pystyy hän lisäksi siirtymään kysymyksestä toiseen sujuvasti ja nopeasti. Kyselyissä on, lomakkeesta riippuen, selkeästi enemmän virheitä.

Keskitetyn puhelinhaastattelukeskuksesta tehdyn tiedonkeruun etuna on erityisesti se, että haastattelujen kulkua ja tutkimuksen laatua voidaan seurata koko tutkimusprosessin ajan. Samoin haastattelijoiden kouluttaminen ja työnohjaus sekä lisätietojen jakaminen ovat yksinkertaisempaa toteuttaa kuin hajautetussa haastattelusysteemissä.

Eri menetelmien etuja ja haittoja

Pitkittäistutkimukset ovat luotettavampia ja tarkempia kuvaamaan muutosta – tai yhteiskunnan dynamiikkaa – kuin toistetut poikittaistutkimukset. Lisäksi pitkittäistutkimuksilla voidaan saavuttaa säästöjä, koska siinä käytetään samaa otosta, samoja tiedonkeruulinstrumentteja ja käytännön rutiinit tulevat tehokkaammiksi. Haittapuolena joissain tilanteissa on niin sanottu paneeliefekti, mikä tarkoittaa sitä, että vastaajat tiedostavat kuuluvansa paneeliin ja muuttavat käyttäytymistään tai vastauksiaan.

Käyntihaastattelututkimuksissa kenttätyövaihe ja sen organisointi vaatii enemmän resursseja ja aiheuttaa enemmän kustannuksia kuin puhelinhaastattelu- tai lomakekyselyissä. Toisaalta haastattelijan läsnäolon kautta on haastattelu-/vastaustilanne haastattelijan toimesta hyvin kontrolloitavissa ja käyntihaastattelu tarjoaa hyvät mahdollisuudet selkeyttää kysymyksiä ja vastauksia. Kalleinta tiedonkeruu on käyntihaastattelujen avulla tehtynä.

Puhelinhaastattelussa ei ole mahdollista käyttää vastaamiseen apuna kuvia tai visuaalisia skaaloja - toisin kuin käyntihaastattelussa ja lomakekyselyssä. Sitä vastoin käyntihaastattelussa haastattelijan vaikutus on selvästi suurempi kuin puhelinhaastattelussa. Lisäksi kotiin tuleva haastattelija vaatii vastaajalta enemmän järjestelyjä kuin puhelinhaastattelu.

Lomakekyselyn etuna ovat pienet kustannukset. Myös sensitiivisiin kysymyksiin on helpompi vastata paperilomakkeella kuin puhelin- tai käyntihaastattelussa. Usean eri vastausvaihtoehdon (>6) erottaminen toisistaan edellyttää niiden visualisointia, mikä on mahdollista lomakekyselyssä ja käyntihaastattelussa.

Tietokoneavusteinen tiedonkeruu tarjoaa runsaasti uusia mahdollisuuksia paperilomakkeeseen verrattuna. Se mahdollistaa erilaisten kysymys- ja vastauskombinaatioiden käytön. Tietokoneavusteinen tiedonkeruussa erilaiset tiedonkeruulomakkeen reititykset on helppo toteuttaa. Lisäksi tiedonkeruulomakkeeseen voi liittää jo valmiiksi tutkittavaa yksikköä koskevia esimerkiksi rekisteritietoja. Tietokoneavusteinen tiedonkeruu mahdollistaa myös vastausten koodauksen jo haastattelutilanteessa, esimerkiksi ammattiluokitus. Tietokoneavusteisuus vähentää tallennusvirheiden määrää ja nopeuttaa aineiston valmistumisprosessia (koska haastattelija (tai vastaaja) tallentaa tiedot suoraan tietokoneelle haastattelutilanteessa ja erilaiset oikeellisuus- ja loogisuustarkistukset on ohjelmoitu osaksi tiedonkeruulomaketta). Lähes kaikki Tilastokeskuksen puhelin- ja käyntihaastattelut tehdään nykyään tietokoneavusteisesti.

Viimeaikoina on alettua tietoa kerätä sähköisellä itse täytettävällä lomakkeella. Sähköinen tiedonkeruulomake voi olla joko niin sanottu internet-lomake tai esimerkiksi sähköpostin liitetiedostona lähetettävä sähköinen lomake. Koko väestöl-

le tarkoitettujen sähköisen tiedonkeruun suurimpana ongelmana on otantakehikko; kaikilla ei ole mahdollisuutta käyttää internet-yhteyksiä.

Sensitiivisissä ja aroissa kysymyksissä saattaa itse täytettävä lomake tuoda luotettavampia vastauksia kuin haastattelu ja samalla vähentää vastauksetta.

Se että vastaaja kirjaa itse vastauksensa mahdollista myös henkilökohtaisissa haastattelussa joko osassa tai kaikkien kysymysten osalta. Perusteluna on se, että haastattelijan läsnäolo vaikuttaa aina vastauksiin, tosin vaikutusmekanismeja ja vaikutuksen suuruudesta ei ole kaikilta osin tietoa.

Suosituksset

Sähköiset itse täytettävät tiedonkeruunstrumentit soveltuvat tällä hetkellä vain yrityksille tehtäviin kyselyihin ja niissäkin on mietittävä, pystyvätkö kaikki yritykset vastaamaan tällaisiin kyselyihin.

Tilastokeskus edellyttää, että väestöhaastattelussa vastausprosentit kyetään laskemaan todellisesta tutkimuksen kohdejoukosta. Esimerkiksi käyntihaastattelussa kohteista kyetään tavoittamaan yli 90 prosenttia vastaajista ja puhelinhaastattelussa puhelinnumerot löydetään lähes 90 prosentille otokseen kuuluvista. Lopullinen vastausprosentti on käyntihaastattelussa yleensä yli 80 prosenttia ja

Taulukko 2.2 Yleisimmin käytettyjen tiedonkeruun menetelmien tietokoneavusteiset västineet (De Leeuw ja Nicholls, 1996)

Menetelmä	Tietokoneavusteinen muoto	Englanninkielinen västine
Käyntihaastattelu	Tietokoneavusteinen henkilöhaastattelu	Computer Assisted Personal Interviewing (CAPI)
Puhelinhaastattelu	Tietokoneavusteinen puhelinhaastattelu	Computer Assisted Telephone Interviewing (CATI)
Itsetäytettävä lomakekysely	Tietokoneavusteinen lomakekysely	Computer Assisted Self Interviewing (CASI)
Postikysely	Esim. Internet-kysely tai levykekysely	Disk by Mail (DBM), Electronic Mail Survey (EMS)
Paneeli- tai päiväkirjatutkimus	Tietokoneavusteinen jatkuva paneeli-tutkimus tai sähköinen päiväkirja	Computer Assisted Panel Research (CAPAR), Teleinterview, Electronic diaries
Osallistuva havainnointi	Esim. Tietokoneavusteinen videokoodaus	Computer Assisted Self Interviewing with Interviewer Present (CASI of CASIIP) Question Text on Screen: Visual (CASI-V) Text on Screen and on Audio (CASI-A)
Rekisteritutkimus	Rekistereiden tietokoneavusteinen yhdistäminen	

puhelinhaastattelussa lähellä 80 prosenttia. Tavanomaisessa postikyselyssä jää-
dään selvästi alle 70 prosentin.

Tilastokeskuksessa on kaksi haastatteluorganisaatiota (kenttä- ja puhelinhaas-
tattelijat), jotka käytännössä toteuttavat haastattelut. Tutkimuksen ajankohta on
aina sovittava näiden organisaatioiden kanssa.

Kirjallisuus

- Ahola, A. 1993. *Tiedonkeruumenetelmä: konteksti : haastatteluvastaukset*. Keskus-
telumuistioita. 4/1993. Tilastokeskus.
- Canoune, H. L. & Leyhe, E. W. 1985. Human Versus Computer Interviewing.
Journal of Personality Assessment. Vol. 49. 103-106.
- Catlin, G. & Ingram, S. 1988. The Effects of CATI on Costs and Data Quality: A
Comparison of CATI and Paper Methods in Centralized Interviewing. Teok-
sessa: R. M. Groves, P. P. Biemer, L. E. Lyberg, J. T. Massey, W. L. Nicholls II
& J. Waksberg (Eds.) *Telephone Survey Methodology*. New York: Wiley.
- Couper, M. P., Hansen, S. E. & Sadovsky, S. 1997. Evaluating Interviewer Use of
CAPI Technology. Teoksessa: L. E. Lyberg, P. Biemer, M. Collins, C. Dippo,
N. Schwarz, & D. Trewin (Eds.) *Survey Measurement and Process Quality*.
New York: Wiley.
- De Leeuw, E. D. & Collins, M. 1997. Data Collection Method and Data Quality:
An Overview. Teoksessa: L. E. Lyberg, P. Biemer, M. Collins, C. Dippo, N.
Schwarz, & D. Trewin (Eds.) *Survey Measurement and Process Quality*. New
York: Wiley.
- De Leeuw, E., Nicholls, W. II. 1996. Technological Innovations in Data Collec-
tion: Acceptance, Data Quality and Costs. *Sociological Research Online*. Vol. 1.
No. 4. <http://www.socresonline.org.uk/socresonline/1/4/leeuw.html>
- De Leeuw, E.D. 1993. *Data quality in mail, telephone and face-to-face surveys*.
Amsterdam: TT-Publikaties.
- Dillman, D. A. 2000. *Mail and Internet Surveys: The Total Design Method*. New
York: Wiley.
- Groves, R. M. 1989. *Survey Errors and Survey Costs*. New York: John Wiley &
Sons.
- Kasprzyk, D., Duncan, G. J. & Kalton, G. 1989. *Panel Surveys*. New York: Wiley.
- Martin, J. & Manners, T. 1995. Computer Assisted Personal Interviewing in Sur-
vey Research. Teoksessa: R. M. Lee (Ed.) *Information Technology for the Social
Scientist*. London: UCL Press.
- Nicholls, W. L. II, Baker, R. P., & Martin, J. 1997. The Effect of New Data Collec-
tion Technologies on Survey Data Quality. Teoksessa: L. Lyberg, P. Biemer,
M. Collins, C. Dippo, N. Schwarz, & D. Trewin (Eds.) *Survey Measurement
and Process Quality*. New York: Wiley.

2.7 TIEDONKERUULOMAKKEEN SUUNNITTELU JA TESTAUS

Soveltamisala ja päämäärä

Kerättävä tieto kirjataan aina jonkinlaiselle standardoidulle pohjalle, josta käytetään yleisnimeä lomake tai kaavake. Haluttu tieto osoitetaan kysymyksellä lomakkeissa ja kaavakkeissa usein yksinkertaisesti 'tiedon nimellä', esimerkiksi 'Liikevaihto'.

Oheiset toimintaperiaatteita ja tiedonkeruulomakkeen suunnittelua koskevat ohjeistukset on tehty ensisijaisesti henkilö pohjaisen tiedonkeruun näkökulmasta. Suurelta osin ne soveltuvat myös yrityskyselyihin. Yrityksille ja luonnollisille henkilöille tehdyt lomakkeet eivät suunnitteluperiaatteiltaan suuresti poikkea toisistaan. Erot johtuvat pääasiassa tutkimusten aiheista ja vastaajista. Yrityskyselyissä tiedustellaan pääasiassa tosiasioita ja käsitteiden voidaan olettaa olevan tuttuja vastaajille, kun taas luonnollisille henkilöille tehtävissä tutkimuksissa tiedustellaan hyvin monenlaisia asioita ja yleisimpien käsitteiden tuntemuksesta ei voida olla varmoja. Yrityskyselyjen vastaajien voidaan olettaa olevan paremmin perillä kysyttävistä asioista. Näistä syistä johtuen yrityskyselyissä käytetään usein kaavakkeita, joissa on erittäin vähän vastausta ohjaavia osia.

Laadinnan näkökulmasta tiedonkeruulomakkeet, instrumentit¹, jaotellaan kahden tekijän mukaan: toisaalta sen mukaan, kuka tiedot merkitsee lomakkeelle ja toisaalta sen mukaan, onko instrumentti paperilla vai sähköinen versio. Kaikki näin muodostuneet neljä muotoa asettavat erilaiset vaatimukset lomakkeen suunnittelulle (ks. Taulukko 2.1 s.49).

Tässä yhteydessä merkittävin on jako itse täytettäviin ja haastattelulomakkeisiin.

Toimintaperiaate

Kyselylomake tehdään vastaajia ajatellen ja haastattelulomake haastattelijoina ja vastaajia ajatellen. Keskeinen periaate kaikkien kysymysten suunnittelussa on, että vastaajan

- on ymmärrettävä, mitä häneltä kysytään,
- voidaan olettaa tietävän vastauksen kysymykseen ja
- on ymmärrettävä, miten vastaus annetaan.

Haastattelijoina aiheutuvan vaihtelun vähentämiseksi haastatteluissa käytetään niin sanottua standardoitua menetelmää, jolla pyritään siihen, että kaikki haastattelijat esittävät kysymykset tarkalleen samalla lailla. Tästä syystä haastatte-

1 Usein puhutaan instrumenteista, koska lomake assosioituu yleensä paperilomakkeeksi ja kuitenkin suuri osa lomakkeista nykyisin on sähköisiä.

lulomakkeissa käytetään esitysstandardeja, joilla ohjataan haastattelijan toimintaa ja kysymysten esittämistä. Pitkissä ja monimutkaisissa haastatteluissa on haastattelijat yleensä koulutettava ennen tutkimuksen alkua. Tämä saattaa pidentää aikataulua huomattavasti.

Tutkija vastaa aina itse tiedonkeruulomakkeensa toimivuudesta. Ohjeita ja apua lomakkeen suunnitteluun saa esimerkiksi haastatteluyksiköstä ja surveylaboratoriosta. Myös muissa tilastoyksiköissä on kokemusta tiedonkeruunstrumenttien suunnittelusta.

Tutkimusetiikkaan kuuluu, että lomakkeen alussa tai saatekirjeessä kerrotaan, miksi se tehdään, kuka rahoittaa tutkimusta ja mistä otos on poimittu. Lomakkeen alussa olisi myös hyvä olla tarkistus siitä, että vastaamassa on oikea henkilö. Kaikista Tilastonkeskuksen haastattelututkimuksista lähetetään vastaajalle etukäteen ilmoitus, jossa kerrotaan odotettavissa olevasta yhteydenotosta ja tutkimuksen sisällöstä.

Tilastokeskuksen toimintaperiaatteisiin kuuluu, että haastateltavien vastaus-
rasite tulee pitää mahdollisimman alhaisena. Tilastolaki edellyttää, että:

Tietoja hankittaessa tilastojen laatimista varten tulee ensi sijassa käyttää hyväksi muussa yhteydessä kerättyjä tietoja. (Tilastolaki 4 §)

Tilastoja laativan viranomaisen on huolehdittava siitä, että tiedonantajilta pyydetään vain tilastojen laatimisen kannalta välttämättömät tiedot. (Tilastolaki 6 §)

Tiedot on kerättävä taloudellisesti ja niin, että siitä aiheutuu tietojen antajille mahdollisimman vähän haittaa ja kustannuksia. (Tilastolaki 6 §)

(Ks. 1.1 Tilastotoimen normisto)

Suosituksset

Tiedonkeruulomakkeen suunnittelu

- Kysymysten laadinnasta sekä haastattelututkimuksiin että kyselytutkimuksiin on kirjoitettu useita oppikirjoja, joihin on hyvä tutustua ennen kuin aloittaa lomakkeen laadinnan, jos asia ei ole tuttu entuudestaan.
- Tilastokeskuksen tutkimuksissa käytetään vakiintuneita käsitelmärittelyjä, luokituksia ja kysymysmuotoiluja. Näistä asioista on kerrottu 'Tilastohaastattelijan oppaassa' ja 'Haastattelulomakkeiden suunnittelustandardit' monisteessa.
- Yleensä kannattaa pyrkiä käyttämään aikaisemmin testattuja ja hyväksi todettuja kysymyksiä ja kysymyspatteristoja. Tutkimusten vertailu edellyttää, että käytetään samoja kysymyksiä.
- Sähköinen ja paperilomake poikkeavat huomattavasti toisistaan siinä suhteessa, miten lomake laaditaan. Sähköinen lomake pitää aina ohjelmoida ja siksi pitää selvittää, kuka vastaa lomakkeen ohjelmoinnista ja kuinka kauan se vie aikaa.
- Jos tutkimus tehdään useammalla kielellä, varataan riittävästi aikaa ja resursseja lomakkeen, työohjeiden sekä muun oheismateriaalin kääntämiseen.
- Itse täytettävässä lomakkeessa on yksiselitteisesti kerrottava, kuka lomakkeen täyttää. Lomakkeen alkuun tai sen saatteeksi on kirjattava selkeät ohjeet siitä, kuinka lomake täytetään ja minne se palautetaan.

- Paperilomakkeesta on tehtävä helppo ja mukava täyttää. Lomakkeesta ei saa tehdä liian ahdasta. Paperilomakkeen on oltava myös tallennuksen kannalta selkeä, sujuva ja looginen. Tallennusta helpottaa, kun vastaukset on sijoitettu samalle sarakkeelle, esimerkiksi marginaaliin tai omaan sarakkeeseen.
- Kaikkiin saatekirjeisiin tulee merkitä selvästi erottuvana tutkimuksen nimi, lisätietoja antavien henkilöiden nimet ja suorat puhelinnumerot, jotta puhelin-tiedustelut voidaan ohjata nopeasti oikeaan osoitteeseen. Puhelinvaihteeseen tulee myös lähettää meneillään olevien kyselyjen saatekirjeet.

Tiedonkeruulomakkeen testaus

- Lomaketta voidaan testata monella eri tavalla. Lomakkeen suunnittelija tai tutkija voi tehdä itse koehaastattelun ja arvioida lomaketta yhdessä esimerkiksi kollegojen kanssa. Erityisesti sähköisen lomakkeen testaaminen on vaativa ja aikaa vievä tehtävä.
- Pilottitutkimus kannattaa tehdä aina, kun se vain on mahdollista. Pilottitutkimuksen etuna on, että sen avulla voidaan testata koko tiedonkeruuprosessin toimivuus eri vaiheissa.
- On suositeltavaa konsultoida tai käyttää Tilastokeskuksen surveylaboratorion palveluja ryhtyessäsi suunnittelemaan uuden tai uudistettavan lomakkeen testaamista.
- Ryhmäkeskusteluilla saa ennakkotietoja tutkittavasta aiheesta sekä kuva niistä käsitteistä ja ilmaisuista, joita ihmiset itse eri asioista käyttävät. Ryhmäkeskustelut nauhoitetaan.
- Kognitiiviset haastattelut ovat yksilöhaastatteluja, joissa tutkitaan vastaamisprosessia. Tarkastelun kohteena on erityisesti ymmärtämiseen, muistamiseen ja vastauksen tuottamisen liittyvät seikat.
- Asiantuntijapaneeli käsittelee järjestelmällisesti lomakeluonnoksen kysymyksiä soveltamalla tietoa ihmisen kognitiivisista prosesseista.
- Jaetulla otoksella (*split panel test*) voidaan arvioida kysymysten sanamuotojen, kysymysten järjestyksen sekä aineistonkeruumenetelmien vaikutusta vastauksiin. Puolittamistestissä testataan kahta tai useampaa vaihtoehtoista versioita kokeellisessa asetelmassa.

Kirjallisuus

- Ahola, A. 1992. *Haastattelukysymysten ymmärtämisestä*. Helsinki: Tilastokeskus. (Keskustelumuitioita 1/92).
- Biemer, P. P., Groves, R. M., Lyberg L. E., Mathiowetz, N. A. & Sudman, S. (eds.) 1991. *Measurement errors in surveys*. New York: John Wiley & Sons.
- Converse, J. M. & Presser, S. 1986. Survey Questions. *Handcrafting the Standardized Questionnaire*. Quantitative Applications in the Social Sciences No. 63. Newbury Park: SAGE Publications Ltd.
- Fowler, F. J. Jr. 1993. *Survey Research Methods*. 2nd ed. Newbury Park: SAGE Publications Ltd.

- Groves, R. M., Biemer P. P., Lyberg L. E., Massey J. T., Nicholls II W. L. & Waksberg J (eds.) 1988. *Telephone Survey Methodology*. New York: John Wiley & Sons.
- Holstein, J. A. & Gubrium, J. F. 1995. *The Active Interview*. Thousand Oaks, CA: SAGE Publications Ltd.
- Lyberg, L., Biemer, P., Collins, M., de Leeuw, E., Dippo, C., Schwarz, N. & Trewin, D. (eds.) 1997. *Survey Measurement and Process Quality*. New York: John Wiley & Sons.
- Paananen, S., Juntto, A., Sauli, H. (toim.) 1998. Faktajuttu. *Tilastollisen sosiaalitutkimuksen käytännöt*. Tampere: Vastapaino.
- Sudman, S., Bradburn, N.M. & Schwarz, N. 1996. Thinking about answers. The application of cognitive processes to survey methodology. San Francisco: Jossey-Bass Publishers.
- Tanur, J. M. (ed.) 1992. *Questions about questions. Inquiries into the cognitive bases of surveys*. New York: Russel Sage Foundation.
- Tilastokeskus. 1980. *Tilastohaastattelijan opas*. Toinen laitos. Käsikirjoja 7. Helsinki: Tilastokeskus.
- Tilastokeskus. 1999. *Tilastohaastattelijan opas*. Haastattelu- ja tutkimuspalvelut. (Muistio).
- Tilastokeskus. 2001. *Haastattelulomakkeiden suunnittelustandardit*. Elinolot/Lomakestandardit-työryhmä. (Muistio).

2.8 VASTAUSKATO JA SIIHEN VAIKUTTAMINEN

Soveltamisala ja päämäärä

Tiedonkeruun eri vaiheissa havaintojen määrä usein pienenee erilaisista syistä. Rekistereitä ynnä muista epäsuoraa tiedonkeruuta hyödyntävissä tutkimuksissa hävikki on pieni ongelma, kun taas erilaiset kehikko- ja mittausvirheet saattavat muodostua huomattaviksi. Sen sijaan suoran tiedonkeruun tapauksessa hävikki, jota kutsutaan vastauskadoksi (ja jäljempänä kadoksi), voi olla suuri ongelma. Perinteisesti vastauskatoa (*non-response*) esiintyy vapaaehtoisuuteen perustuvissa kysely- ja haastattelututkimuksissa, mutta edes lakisääteisissä tiedonkeruissa, kuten kokonaislaskennoissa siltä ei voi täysin välttyä, vaikka tiedonkeruun toteutus tapahtuisi aivan optimaalisesti.

Vastauskato jaetaan kahteen pääryhmään: eräkatoon (*item non-response*) ja yksikkökatoon (*unit non-response*). Eräkadolla tarkoitetaan sellaista vastausta, jossa tutkimusyksikkö antaa osan tiedoista hyväksyttävästi, mutta joko ei osaa tai ei tahdo vastata taikka antaa sellaisen vastauksen, joka myöhemmissä aineiston tarkistuksissa joudutaan hylkäämään. Yksikkökadon tapauksessa kaikki havaintoyksikköä koskevat tutkimustiedot puuttuvat tai joudutaan hylkäämään.

Kadolla on vaikutusta tutkimukseen tuloksiin: tutkimusjoukon pienenemisellä tavoitteeseen eli perusjoukkoon nähden on useimmiten harmillisia vaikutuksia.

Mikäli vastanneet ja kato ovat sekä tausta- että tutkimusmuuttujien suhteen samoin jakautuneita, otosvarianssi suurenee kadon vaikutuksesta. Myös kokonais- tutkimuksiin syntyy otosvarianssia tällä tavoin. Useimmiten vastaajat ja katoon jääneet yksiköt poikkeavat toisistaan, mikä aiheuttaa tutkimuksen tuloksiin virhettä, pahimmassa tapauksessa harhaa.

Periaatteet

Tärkeimpänä periaatteena on luonnollisesti tuottaa mahdollisimman laadukkaita ja harhattomia tilastoja sekä tutkimustuloksia, joiden kato-osuus on alhainen ja havaintoaineiston varianssi pieni. Niiden saavuttaminen edellyttää tutkimusten tiedonkeruun huolellista suunnittelua, tutkimusvälineiden testausta ja kenttätöiden sekä tietojen jatkokäsittelyn varmistamista. Jokaiseen vaiheeseen liittyy monia rajoituksia, kuten käytettävissä olevat määrärahat, aika ja henkilökunta. Erityinen huomio kohdistuu tiedonkeruun kenttätöiden järjestelyihin, koska ne muodostavat rajapinnan tutkimuksen kohteena olevan yksikön ("vastaajan") ja tiedonkerääjän välille.

Yksikkökadon syyt jaetaan tavallisesti kolmeen pääryhmään: tavoittamatta jääneisiin, kieltäytyneisiin ja muihin syihin. Periaatteena on yleensä pyrkiä maksimoimaan koko vastausosuutta, jolloin tavoittamatta jääneiden ja muiden syiden osuus yritetään pitää mahdollisimman pienenä. Vapaaehtoisissa tutkimuksissa kieltäytyminen on kohteena olevan yksikön valinta, joten kieltäytymispäätöksiin ei voida juuri vaikuttaa; tosin haastattelijoita voidaan kouluttaa kieltäytyneiden suostuttelussa. Sen sijaan lakisääteisissä ym. ei-vapaaehtoisissa tutkimuksissa erityisesti kieltäytyneiden osuutta pyritään pitämään mahdollisimman vähäisenä. Usein juuri kieltäytyneiden ryhmä aiheuttaa suurimmat riskit aineistojen vinoutumiseen ja harhaisuuteen.

Kenttätöiden jälkeen aineiston sisältämän vinouden ja virheen sekä mahdollisen harhan arvioimiseksi ja sen vähentämiseksi (eli tulosten "oikaisemiseksi") on tärkeää, että saatua havaintoaineistoa verrataan huolellisesti tutkimuksen perusjoukon kanssa. Tiedonkeruusta ja saadusta havaintoaineistosta laaditaan laatukuvaukset, jotka sisältävät myös mahdollisimman yksityiskohtaisen katoanalyysin, sen syistä ja vaikutuksista tuloksiin. Kadon aiheuttamaa harhaa pyritään korjaamaan tilastollisin menetelmin (ks. 2.10 Painokertoimien muodostus ja katovirheen oikaisu).

Suosituks

- Tutkimuksen perusjoukko ja mahdolliset poimintakehikot tutkitaan huolellisesti ja verrataan tutkimuksen tavoitteena olevana kohdeperusjoukkoon.
- Kenttätöitä ja siinä käytettävät menetelmät ovat asianmukaiset ja oikein suunnattu tutkimuksen kohdeperusjoukolle (kuten työn ajoitus, materiaalit, ohjeet, tiedonkeruuvälineet, mahdolliset tekniset ja muut apuvälineet ovat mahdollisimman huolellisesti valittuja ja pyrkivät pienentämään vastausrasitusta).
- Keruuvälineet ja menetelmät testataan huolellisesti ennen varsinaista tiedonkeruuta.

- Valittu otantamenetelmä tuottaa harhattomat tulokset perusjoukolle ja sen tärkeimmille osajoukoille.
- Tavoittamatta jääneitä tapauksia pyritään jäljittämään, joissain tapauksissa myös kieltäytyneitä pyritään suostuttelemaan vastaajiksi. Pakollisissa tutkimuksissa kieltäytymiset pyritään minimoimaan.
- Vinoja perusjoukkoja tutkittaessa (kuten yrityksiin suunnatuissa tutkimuksissa) pyritään tutkittavan ilmiön kannalta tärkeimmiltä yksiköiltä saamaan vastaus.
- Vastaajien ja kadon rakenne tutkitaan ja aineisto oikaistaan kadon vaikutuksista niin hyvin kuin mahdollista esimerkiksi painottamalla tai paikkaamalla (ks. 2.9 Tilastollinen editointi ja imputointi; 2.10 Painokertoimien muodostus ja katovirheen oikaisu).
- Kadon määrä ja rakenne raportoidaan ja aineistonkäyttäjiä informoidaan sen mahdollisista tuloksista vinouttavista vaikutuksista. Jatkuviissa tutkimuksissa kadon suuruutta seurataan, jotta kenttätöissä tai yhteiskunnallisessa ilmapiirissä tapahtuvat muutokset voitaisiin ottaa huomioon.

Kirjallisuus

- Groves, R.M., Couper, M.P., 1998. *Nonresponse in Household Interview Surveys*. New York: Wiley.
- Groves, R.M., Dillman, D.A., Eltinge, J.L., and Little, R.J.A. (eds.) 2002. *Survey Nonresponse*. New York: Wiley.
- Incomplete Data in Sample Surveys*. 1983. Vol. 1 Report and Case Studies (eds. W. Madow, H. Nisselsohn and I. Olkin); Vol. 2 Theory and Bibliographies (eds. W. Madow, I. Olkin and D. Rubin); Vol. 3 Proceedings of the Symposium (eds. W. Madow and I. Olkin). New York: Academic Press.
- Japac, L., Ahtiainen, A., Hörngren, J., Lindén, H., Lyberg, L., Nilsson, P. 1997. *Minska bortfallet*. Statistiska centralbyrån. Örebro: SCB-tryck.
- Kalton, G. 1983. *Compensating for Missing Survey Data*. The University of Michigan, Institute for Social Research, Survey Research Center, Research Report Series.
- Laaksonen, S. (ed.) 1996. *International Perspectives on Nonresponse*. Statistics Finland. Research Reports No. 219. Helsinki: Edita.
- Laaksonen, S. 1988. *Katovirheen korjaus kotitalousaineistossa*. Tutkimuksia 147. Helsinki: Tilastokeskus.
- Lessler, J. and Kalsbeek, W. (eds.). 1992. *Nonsampling Error in Surveys*. New York: Wiley.
- Lundström, S. and Särndal, C-E. 2001. *Estimation in the presence of Nonresponse and Frame Imperfections*. Stockholm: SCB.
- Särndal, C-E., Swensson, B. and Wretman, J. 1992. *Model Assisted Survey Sampling*. New York: Springer.

2.9 TILASTOLLINEN EDITOINTI JA IMPUTOINTI

Tilastollinen editointi

Soveltamisala ja päämäärä

Tilastollinen editointi tarkoittaa toimintoja, joilla tilastoaineiston havaintoarvot tarkistetaan ja korjataan mahdollisimman oikeiksi koskien yksittäisiä arvoja ja arvojen keskinäistä loogisuutta ja yhteensopivuutta. Editointi tulisi toteuttaa ensin kerätyllä aineistotasolla toisin sanoen perus- eli mikrotasolla (mikro-editointi) ja lisäksi myös karkeammilla tasoilla (makro-editointi). Lopuksi on syytä on pyrkiä eri tasojen keskinäiseen yhteensopivuuteen.

Editointia vaativia toimenpiteitä edellyttävät:

- Havaintojen puuttuminen, jolloin tapauksen tunnistamisen jälkeen jatketaan imputoinnilla.
- Kelvottomat vastaukset kaikkiin kysymyksiin tai osaan niistä.
- Vastausten ja muiden tietojen epäloogisuudet.

Tilastollista editointia tarvitaan kaikissa tiedonkeruun vaiheissa, alkaen keruun suunnittelusta, jatkuen tallentamiseen ja analysointiin. Aineiston lisäeditointi tulee tarpeelliseksi jopa vuosia keruun päättymisen jälkeenkin, kun poikkileikkausaineistoja yhdistetään muiden aineistojen kanssa eli muodostettaessa pitkittäisaineistoja ja muita yhdistelmiä. Kaikissa tiedusteluissa on tällaisia ongelmia ja pienikin määrä voi aiheuttaa tuloksiin merkittäviä vääristymiä, mikäli ei huolella käydä läpi koko aineistoa.

Periaatteet

Editoinnilla on kolmitahoinen tehtävä eli se (i) tuottaa pohjan tilastotutkimusten aineistojen parantamiseksi tulevaisuudessa, (ii) tuottaa tietoa tilastotutkimusten aineistojen laadusta, ja (iii) siistii aineiston (Granquist, 1994). Kaksi ensiksi mainittua editoinnin perustetta ovat tärkeitä, koska keruun virheiden korjaaminen on hankalaa. Lisäksi editointi luo pohjaa tuleville vähemmän puutteita sisältäville aineistoille.

Kaikissa näissä yhteyksissä on aineiston käsittelijän kyettävä erittelemään virheet ja puutteet niiden kokonaismerkityksen kannalta eli kaikki virheet eivät ole samanarvoisia. Vastaavasti eniten työvoimaa ja tutkimusta on panostettava merkittävimpien ongelmatekijöiden erittelemiseen ja ehkäisyyn. Pienet virheet voidaan lisäksi usein hoitaa automaattisilla ratkaisuilla, joiden laajamittaiseen käyttöön ja ohjelmointiin tulee pyrkiä.

Suosituks

- *Tunnistimien eli identifioijien* oikeellisuus on ratkaisevan tärkeä, jos aineistoa aiotaan myöhemmin yhdistää muihin aineistoihin. Mikäli tilastoyksiköiden aineistoja yhdistetään pitkittäin, niin vain poikkileikkauslinkkaukseen soveltuva

tunnistin ei riitä. Aineistoon on siten syytä suunnitella myös oma pitkittäistunnistinjärjestelmä.

- *Arvoaluetarkistus* on välttämätön koko aineiston tasolla ja tulisi tehdä mahdollisuuksien mukaan myös erillisesti sopivissa osajoukoissa. Edelleen olisi rakennettava hyvä järjestelmä hierarkkisten tietojen arvoaluetarkistukseen ja korjaukseen.
- *Erikoishavaintoarvojen* läpikäynti on selkeätä jatkoa arvoaluetarkistukselle. Hyväksyttävälläkin alueella sijaitsevat havaintoarvot voivat olla ongelmallisia, jos niiden vaikutus tuloksiin on liian suuri. Tämä koskee ääriarvoja tai poikkeavia arvoja, joilla tarkoitetaan datamassasta selvästi poikkeavia arvoja. Ääriarvo on joko (i) täysin virheellinen, tai (ii) se on perustunut väärään tulintaan tai (iii) se on aivan oikea.
- Erotta toisistaan *puuttuva, nollatieto ja mahdoton tieto*.
- Eri tietojen tulee olla yhteensopivia samalla kun yksittäiset tiedot ovat korrekkeja. Ristikkäistarkastuksilla eri osa-alueita tutkitaan mahdollisesti myös ulkopuolisten datojen kanssa.
- Pyri mahdollisimman pitkälle automaattisiin ratkaisuihin siten, että ensin havainnoidaan virheet, toiseksi tietyt kriteerit omaavat virheet korjataan automaattisesti ja kolmanneksi loput virhetilanteet tulevat manuaaliseen käsitteilyyn.
- Monimuuttuja-analyysieihin perustuvat tarkistukset on tehtävä aidon tutkimusasetelmaan sopivan mallin avulla eli normaalin analyysin yhteydessä.
- Editointiprosessi tulee dokumentoida sekä kirjallisesti sisältäen teoreettiset ja käytännölliset perustelut sekä atk-ohjelman avulla automaatiota käytettäessä.

Imputointi

Soveltamisala ja päämäärä

Imputointi tarkoittaa puuttuvan tiedon paikkausta korvikearvolla, jonka tulee olla mahdollisimman oikea. Tavallisesti imputointia käytetään eräkadon korjaamiseen, mutta sitä voidaan soveltaa yksikkövastaustasolla. Laajamittaisissa ja automatisoiduissa imputointitilanteissa käytetään termiä massaimputointi.

Periaatteet

Imputointimenetelmiä on hyvin monenlaisia (ks. Kalton ja Kasprzyk, 1986). Ne voidaan esittää kolmen pääryhmän puitteissa (Laaksonen, 2000):

(i) *Looginen imputointi* tapahtuu tilastollisen editoinnin yhteydessä. Tässä on olemassa looginen eli yksikäsitteisesti kuvattavissa oleva yhteys tietyn tai tiettyjen havaintoarvojen osalta tunnettujen muuttujien tai toisten havaintoarvojen osalta tuntemattomien muuttujien välillä. Menetelmää käytetään usein olettaen tämän yhteyden olevan, vaikkei varmaa osoitusta yhteydestä olisikaan.

Muissa kahdessa menetelmäryhmässä puuttuvan arvon korvike eli imputoitu havaintoarvo lainataan joltakin, jota voidaan kutsua luovuttajaksi (*donor*). Imputointimenetelmän pääryhmät määräytyvät luovuttajan tyyppin mukaan:

(ii) *Malliluovuttaja-imputointimenetelmässä (model-donor imputation)* luovuttajana on malli, joka voi olla yksinkertainen tai monimuuttujainen/tasoinen. Malliluovuttajamenetelmä voi antaa sekä havaittuja että havaitsemattomia arvoja.

(iii) *Vastaajaluovuttaja-imputointimenetelmässä (real-donor imputation)* luovuttajana on aito vastaaja. Arvo on siis varmasti sellainen joka on aidossa tilanteessa mahdollinen.

Käytännön tilanteessa voi olla hyvä soveltaa useita menetelmiä, jopa yhtä muuttujaa imputoitaessa, koska menetelmä (ii) voi sopia hyvin sellaisille alueille joihin menetelmällä (iii) ei löydy oikeita havaintoja.

Imputoinnin pääryhmien sisältä voidaan löytää myös alajako: deterministinen vs. stokastinen. Viimeksi mainittuun sisältyy satunnaislementti. Kaikkia imputointimenetelmiä (i), (ii) ja (iii) voidaan soveltaa sekä stokastisesti että deterministisesti.

Imputointimallit voidaan edellisen lisäksi jakaa kahteen ryhmään. Ensinnäkin sellaisiin joissa puuttuva arvo korvataan yhdellä (*single imputation*) tai usealla korvikearvolla (*multiple imputation*) (Rubin, 1987).

Moni-imputointia puoltaa piste-estimointiin verrattuna se, että yhdellä arvolla korvaaminen voi olla liian riskialtista tilanteessa jossa on huomattavaa epävarmuutta. Eniten moni-imputointia käytetään arvioitaessa imputoinnista johtuvaa lisävirhettä eli imputointivarianssia. *Imputointi aiheuttaa aina tiettyä lisäepätarkkuutta tuloksiin, joka pitää pyrkiä estimoimaan.* Tämä ei kuitenkaan edellytä moni-imputointia, myös muita menetelmiä on olemassa (ks. Shao, 1997; Rao ja Shao, 1992; Särndal, 1996). Malliluovuttaja -menetelmät ovat ilman stokastista elementtiä hankalia varianssiestimoinnin ja jakauman hyvyyden kannalta.

Suositukses

Ei ole ilman muuta selvää, missä tilanteissa imputointia kannattaa käyttää. Alkuperäisen muuttujan lisäksi on tarkoituksenmukaista tehdä toinen muuttuja, johon puuttuvat tiedot imputoidaan ja merkataan yksiselitteisesti (liputtaminen). Näin on mahdollisuus käyttää sekä alkuperäisiä että imputoituja arvoja ja verrata tuloksia keskenään.

Imputoitu tulos on helpommin saatavissa luotettavaksi, jos:

- Taustalla on hyvä apumuuttuja ja sen yhteys tulosmuuttujaan on hyvin mallitettu.
- Estimoitava parametri on yksinkertainen. Yksinkertaisimpia ovat kokonaisummat ja keskiarvot. Jakaumien kanssa on usein ongelmia. Erityisen hankalia ovat monimuuttujaiset tilanteet eli imputoituihin arvoihin perustuvia tuloksia käytetään jatkossa yhteyksien tutkimisessa. Tällöin imputointi voi vaikuttaa väärällä tavalla lopputulokseen, erityisesti jos samoja muuttujia on käytetty sekä imputointimallissa että analyysimallissa.

- Ota huomioon, että imputointi on erityisen hankala muuttujille, jotka koskevat asenteita ja mielipiteitä. Sen sijaan faktamuuttujille kuten tuloille imputointi on usein onnistunut paljon paremmin (Laaksonen, 1991).
- Imputointi on kokonaisuudessaan prosessi:
 - (i) rakenna imputointimalli, jonka pohjalta ratkaistaan imputointimenetelmä (useita menetelmiä kokeillen ja testaten),
 - (ii) suorita itse imputointi,
 - (iii) tutki imputointimallin spesifiointia monesta näkökulmasta etsien ratkaisua joka on parhaiten ennustava (*predictable*),
 - (iv) diagnostisoi imputointitulosta sekä graafisesti että taulukoimalla sekä vertaamalla tuloksia benchmarking-tietojen kanssa; diagnostisointiin liittyy vastaajaluovuttaja-menetelmää käytettäessä tarkastelu siitä kuinka paljon eri luovuttajia on käytetty,
 - (v) laske imputointivarianssi,
 - (vi) merkitse imputoidut tiedot tulevia käyttäjiä varten,
 - (vii) dokumentoi koko prosessi,
 - (viii) laadi imputoidun aineiston käytettävyysselostus eli mihin jatkokäyttöön se soveltuu ja kuinka hyvin.

Kirjallisuus

- Granquist, L. 1994. Macro Editing. *A Review of Some Methods for Rationalizing the Editing of Survey Data*. Statistical Data Editing. Volume No 1. Methods and Techniques. UN/ECE, 1994.
- Kalton, G. & Kasprzyk, D. 1986. The Treatment of Missing Survey Data. *Survey Methodology*, Vol. 12, 1-16.
- Laaksonen, S. 1991. Adjustment for Non-response in Two-year Panel Data: Applications to Problems of Household Income Distribution. *The Statistician*. 40, 153-168.
- Laaksonen, S. 1996. Statistical Methodology and the Principle of Subsidiarity: A Challenging Optimization Task for Eurostat and European Statistical Institutes. *Proceedings the 1st International Conference on Methodological Issues in Official Statistics*. Stockholm: Statistics Sweden.
- Laaksonen, S. 2000. Regression-Based Nearest Neighbour Hot Decking. *Computational Statistics*. Vol. 15. 65-71.
- Little, R. & Rubin, D. 1987. *Statistical Analysis with Missing Data*. New York: John Wiley & Sons.
- Rao, J.N.K. & Shao, J. 1992. Jack-Knife Variance Estimation With Survey Data Under Hot Deck Imputation. *Biometrika*. Vol. 79. No. 4. 811-822.
- Rubin, D. 1987. *Multiple Imputation for Nonresponse in Surveys*. New York: John Wiley & Sons.
- Rubin, D. 1996. Multiple Imputation After 18+ Years. *Journal of the American Statistical Association*. Vol.91. No. 434. 473-520.
- Schafer, J.L. 1997. *Analysis of Incomplete Multivariate Data*. London: Chapman & Hall.

- Shao, J. 1997. *Variance Estimation for Imputed Survey Data With Non-Negligible Sampling Fractions*. Invited Paper for the 51st Session of the ISI, Istanbul.
- Särndal, C.-E. 1996. For a Better Understanding of Imputation. Teoksessa: Laaksonen, S. (ed.). *International Perspectives on Non-response*. Statistics Finland. Research Reports No. 219. Helsinki: Edita.7-22.
- West S.A., Kratzke, D-T. & Robertson, K.W. 1996. Alternative Imputation Procedures for Item-Non-Response from New Establishments in the Universe. *ASA Proceedings of the Section in Survey Research Methods*.

2.10 PAINOKERTOIMIEN MUODOSTUS JA KATOVIRHEEN OIKAISU

Soveltamisala ja päämäärä

Useimmissa otantatutkimuksissa havaintoihin kuuluu liittää painokerroin, joka sisältää kuhunkin havaintoon kohdentuvan informaation otoksen poiminnassa käytetystä otanta-asetelmasta sekä joissain tapauksissa myös lisäinformaatioon perustuvista estimoinnin tehostus- ja/tai kadonkorjausmenetelmistä. Painokertoimen käytöllä on tarkoitus huolehtia siitä, että estimointimenetelmän tuottamat parametriestimaatit olisivat mahdollisimman virheettömiä.

Painokertoimien perusmuoto on asetelmapaino, joka useimpien otanta-asetelmien tapauksessa muodostuu kunkin alkion sisällyttämistodennäköisyyden käänteislukuna. Kunkin poimitun alkion todennäköisyys tulla valituksi on sen tähden kyettävä johtamaan. Asetelmapaino ei useimmissa käytännön tilanteissa riitä, koska otoksen poiminnan ja tiedonkeruun jälkeen aineisto saatetaan havaita vinoksi kehikkovirheiden, otantavirheen, kadon tai mittausvirheiden vuoksi. Asetelmapainoja joudutaan muokkaamaan erilaisten mallioletusten perusteella ja näin johdettua painotusta kutsutaan uudelleenpainotukseksi.

Ilman painokerrointa voidaan estimoida vain, jos asetelma on itsepainottuva (esim. yksinkertainen satunnaisotanta tai satunnaistetusta perusjoukosta poimitu systemaattinen otos) ja kadon rakenne voidaan syystä olettaa sellaiseksi, ettei sitä tarvitse oikaista. Tässäkin tapauksessa estimoitavana parametrinä tulee olla muu kuin kokonaismäärä (esimerkiksi keskiarvo, korrelaatiokerroin tai kahden keskiarvon suhde). Jos yksikin näistä ehdoista jää toteutumatta, on painokertoimen käyttö välttämätöntä asianmukaisen estimoinnin toteuttamiseksi.

Toimintaperiaate

Painokerroin on yleensä korottava, eli laskettaessa kokonaismääräestimaatteja tulokset ovat tarkastelussa perusjoukon tasolla. Tällöin painokertoimen antama lukema voidaan tulkita niin, että otokseen valittu havainto edustaa painonsa osoittamaa määrää havaintoja perusjoukossa. Painokertoimen muodostusta varten on

selvitettävä käytettävän kokonaismäärän estimaattorin rakenne. Estimaattorikaavasta on löydettävissä painokertoimen muoto.

Joskus korottavaa vaikutusta ei haluta painoon mukaan, jolloin painon arvot vaihtelevat ykkösen molemmin puolin ja painojen summaksi tulee hyväksytyin havaintoaineiston koko. Tällöin puhutaan analyysipainosta. Analyysipainon käyttöä puolletaan tutkimustraditiolla, tulostussyillä tai sillä, ettei sovellettavalle analyysimenetelmälle ole toistaiseksi johdettu otanta-asetelman huomioon otettavaa versiota. Korottavan vaikutuksen poistaminen painokertoimesta edellyttää sen suhteuttamista yksinkertaisen satunnaisotannan korottavaan painokertoimeen, jossa kato on otettu huomioon. Joissain tutkimustilanteissa painokertoimen muokkauksen avulla pyritään muodostamaan vakioituja jakaumia tai tunnuslukuja.

Uudelleenpainotus on yleisesti käytössä oleva menetelmä, jolla alkuperäisiä asetelmapainoja muokataan käyttämällä hyödyksi lisäinformaatiota joko perusjoukosta, otoksesta tai kummastakin (Oh ja Scheuren, 1983). Yksinkertaisimmat uudelleen-painotusmenetelmät ovat jälkiositus sekä suhdetehostus. Jälkiosituksessa otos painotetaan perusjoukon tunnettujen jakaumatietojen mukaan, esimerkiksi henkilö pohjaisissa tutkimuksissa demograafisten tietojen, kuten ikä - ja sukupuoliryhmien sekä asuinalueen mukaan (Särndal et al., 1992). Vastaavasti suhdetehostuksessa hyödynnetään jatkuvien muuttujien summatietoja, kuten yrityksiä koskevista tutkimuksista yrityksen liikevaihdon määrää suhteessa koko toimialan liikevaihdon summaan. Monimutkaisemmat uudelleenpainotusmenetelmät perustuvat joko eksplisiittisten tai implisiittisten tilastollisten mallien käyttöön (Rosenbaum, 1987; Ekholm ja Laaksonen, 1991).

Eräs malliin perustuva operaatio on valittujen havaintoyksiköiden vastaamattomuudesta johtuvan katovirheen oikaisu eli adjustointi. Painokertoimen käyttämättä jättäminen tällaisissa tapauksissa voi tuottaa harhaisia tuloksia ja johtaa virheellisiin johtopäätöksiin. Katovirheen oikaisu voidaan yksinkertaisimmillaan toteuttaa korvaamalla alkuperäisen otoksen koko vastanneiden määrällä. Suositeltavampaa on oikaista kadon vaikutusta pohjautuen tutkittavan ilmiön kannalta relevanttiin informaatioon perusjoukosta. Tällöin oletetaan otoksen ulkopuolisen tiedon vähentävää katovirhettä. Joissakin tapauksissa myös skaalaaminen perusjoukon (rekisterin) tiedossa oleviin tunnuslukuihin on tarkoituksenmukaista.

Suositukses

Painokerroin koostuu yleensä perusjoukkoon, otokseen sekä vastanneisiin liittyvistä lukumäärätiedoista, jotka ovat usein jostain perusjoukon ryhmistä, yleensä ositteista. Jos ryhmittelyä on tehty, on kunkin havainnon aina identifioitava johonkin ryhmään. Kyseisten ryhmien perusjoukkotasoisien tietojen täytyy olla käytettävissä painoja laskettaessa. Jos poiminnassa on ollut mukana erisuuria todennäköisyyksiä, täytyy näiden todennäköisyyksien olla muodostetut havaintokohtaisesti, jotta niitä voidaan käyttää painonmuodostuksessa. Malliavusteista estimointia varten tehdään yleensä alustavia estimaatteja, jotta mahdollisimman virheetön menetelmä kyetään löytämään. Painokerrointen muodostamisen onnis-

tuminen voidaan todentaa ensinnäkin laskemalla havaintoaineistossa painokertoimien summa. Tällöin summan pitäisi olla sama kuin perusjoukon koko. Vastavasti osituksen ja/tai jälkiosituksen kuhunkin ositteeseen saatavien painojen summien pitäisi täsmätä ositteiden kokojen kanssa. Kalibroinnissa apuna käytettävien perusjoukon apumuuttujien reunajakaumien sekä määräestimaattien pitäisi myös olla yhtä aineistosta laskettujen vastaavien painotettujen lukujen kanssa (Deville ja Särndal, 1992; Lundström, 1997).

Eri otanta-asetelmien mukaisen asetelmapainojen kaavat on muodostettu tilanteeseen, jossa yksikkökatoa ei ole. Toimenpiteenä tilanteessa voi olla alkupe- räisen otoksen koon korvaaminen vastanneiden määrällä. Tällöin siis oletetaan ha- vainnon empiiriseksi vastaustodennäköisyydeksi vastanneiden määrän osuus otok- sesta.

Katovirheen oikaisu eli adjustointi perustuu usein jälkiositusta muistuttavaan käytäntöön, jossa rekisteristä löytyvän taustamuuttujan tai -muuttujien luokitte- lulla saatavia jakauma- tai ristiintaulukko-frekvenssejä käytetään osituksen tapaan painokertoimen muodostamisessa (Smith, 1990; Särndal et al., 1992; Deville, Särndal, ja Sautory, 1993). Teoriamielessä pidemmälle menevässä menetelmässä eli painojen kalibroinnissa käytetään useamman muuttujan jakaumatietoa, mutta solufrekvenssitietojen sijasta käytettävissä on pelkästään reunajakaumat taikka jat- kuvien muuttujien kokonaismäärät (Deville ja Särndal, 1992; Lundström, 1997). Useimmiten käytetään perusjoukon tai sen ositteiden tasoista tietoa, joskus myös otostason tietoa. Painojen kalibrointi on suhteellisen monimutkainen mallinta- mismenetelmä, jota varten on olemassa omat ohjelmansa (esimerkkeinä CALMAR (Sautory, 1993) ja CLAN97 (Andersson ja Nordberg, 1998).

Painokertoimien jakaumaa on syytä tarkastella jo otanta-asetelman muodos- tusvaiheessa ja ennakoida mahdollista katoa eri ositteissa. Hyvin suuret painot muuten pienien painojen joukossa voivat vaikuttaa estimaattoreiden varianssei- hin odottamattoman voimakkaasti. Jos ositus on voimakkaan painonvaihtelun syynä, on hyvä karkeistaa tältä osin ositusta (ositteiden luhistaminen) tai lisätä on- gelmaositteisiin otoskokoa. Jälkiosituksen tapauksessa jaon karkeistus on ainoa keino. Kalibrointia käytettäessä annettujen reunajakaumajakojen karkeistaminen ja määräestimaattiehtojen vähentäminen tulevat tältä osin kysymykseen.

Dokumentointi

Raportoinnissa keskeistä on otanta-asetelman, estimointimenetelmän ja uudel- leenpainotuksen tarkka kuvaus. Laatuselosteissa on selvitettävä tärkeimpien esti- maattoreiden rakenne, aineistoon sisältyvät virheet, erityisesti katovirhe, ja niistä aiheutuneet korjaustarpeet painotuksessa. Tuoteselosteiden ja/tai laatuselvitys- ten menetelmäkuvauksiin tulee sisältyä painokertoimen muodostamisperiaat- teet sekä kuvaus lopullisen painotuksen tuottamista tärkeimmistä parametriesti- maateista (ks. 4.3.1. SVT:n Tilastojulkaisut ja tuoteselostesuositus). Luonnolli- sesti kuvauksissa on selvitettävä painotuksen aiheuttamat muutokset perusasetel- masta johdettuihin otosvarianssien lausekkeisiin, jotta keskivirhe ym. laatuindi- kaattorit voidaan laskea mahdollisimman virheettömästi.

- Andersson, C. & Nordberg, L. 1998. *A User's Guide to CLAN*. Statistiska centralbyrån. Örebro: SCB-tryck.
- Bethlehem, J.G. & Keller, W.J. 1987. Linear Weighting of Sample Survey Data. *Journal of Official Statistics*. Vol. 3. No. 2. 141-153.
- Deville, J.-C. & Särndal C.-E. 1992. Calibration Estimators in Survey Sampling. *Journal of the American Statistical Association*. Vol. 87. No. 418. 376-382.
- Deville, J.-C., Särndal, C.-E. & Sautory, O. 1993. Generalized Raking Procedures in Survey Sampling. *Journal of American Statistical Association*. Vol. 88. No. 423. 1013-1020.
- Djerf, K. 2000. *Properties of Some Estimators under Unit Nonresponse*. Statistics Finland. Research Reports 231. Helsinki: Hakapaino Oy.
- Ekholm, A. & Laaksonen, S. 1991. Weighting via Response Modelling in the Finnish Household Budget Survey. *Journal of Official Statistics*. Vol. 7. No. 3. 325-337.
- Kalton, G. & Maligalig, D.S. 1991. *A Comparison of Methods of Weighting Adjustment for Nonresponse*. (With discussion). Bureau of the Census. Proceedings of the Annual Research Conference. 409-447.
- Laaksonen, S. 1988. *Katovirheen korjaus kotitalousaineistossa*. Tilastokeskus. Tutkimuksia 147. Helsinki: Valtion painatuskeskus.
- Lundström, S. 1997. *Calibration as a Standard Method for Treatment of Nonresponse*. Doctoral Dissertation. University of Stockholm. Department of Statistics. Stockholm: Akademityck AB.
- Oh, J.L. & Scheuren, F. 1983. Weighting Adjustment for Unit Nonresponse. Teoksessa: *Incomplete Data in Sample Surveys*, toim. Madow, W.G., Olkin, I. & Rubin, D.B. Vol 2. 143-184. New York: Academic Press.
- Rosenbaum, P.R. 1987. Model-Based Direct Adjustment. *Journal of the American Statistical Association*. Vol. 82. No. 398. 387-394.
- Sautory, O. 1993. La macro CALMAR. Redressement d'un échantillon par calage sur marges. I.N.S.E.E., Série des documents de travail n° F 9310. Paris: I.N.S.E.E.
- Smith, T.M.F. 1990. Post-stratification. *The Statistician*. 40. 315-323.
- Särndal, C.-E., Swensson, B. & Wretman, J. 1992. *Model Assisted Survey Sampling*. New York: Springer-Verlag.

2.11 TILASTOLLINEN ESTIMOINTI JA ANALYYSI

Soveltamisala ja päämäärä

Kuvailevissa tilastotutkimuksissa kiinnostuksen kohteena ovat parametrit ovat yleensä kokonaismääriä, keskiarvoja ja osuuksia koko perusjoukon tasolla tai perusjoukon osajoukoissa. Analyyttispainotteisissa tutkimuksissa kiinnostuksen kohteena ovat ilmiöiden väliset yhteydet ja riippuvuudet. Parametrit liittyvät tilastollisiin malleihin, kuten lineaarisiin malleihin, ja ovat esimerkiksi korrelaatiokertoimia tai regressiokertoimia. Molemmissa tutkimustyypeissä tärkeä työvaihe on estimoida

tuntemattomat parametrit mahdollisimman luotettavasti. Analyyttisiin tutkimuksiin liittyy usein kuvailevia tutkimuksia laajempi tulosten tulkinnallinen osuus.

Estimoinnin avulla saadaan tuntemattomille parametreille numeeriset arviot käytävissä olevan otos- tai muun havaintoaineiston perusteella. Estimoinnissa tarvittava laskentamenetelmää tai -algoritmia kutsutaan estimaattoriksi, ja sen avulla saatua numeerista arvoa estimaatiksi. Estimaatin luotettavuutta arvioidaan laskemalla käytetyn estimaattorin keskivirhe tai variaatiokerroin. Keskivirhe-estimaatin avulla voidaan laskea parametrille luottamusväli eli virhemarginaali (esimerkiksi 95 % luottamusväli). Luottamusvälit ja variaatiokertoimet ovat tilaston luotettavuuden arviointiin liittyviä keskeisiä tunnuslukuja, joita tulee myös raportoida käyttäjiä varten.

Tilastollinen analyysi viittaa työvaiheeseen, jossa havaintoaineistoa analysoidaan tilastollisilla analyysimenetelmillä, kuten regressio- tai varianssianalysilla. Analyysin yhteydessä joudutaan estimoimaan käytettäviin analyysimenetelmiin liittyvien tilastollisten mallien parametrit, esimerkiksi regressiokertoimet, ja niiden keskivirheet kuvaamaan estimaattien luotettavuutta. Tilastollinen analyysi liittyy aineiston jatkojalostukseen ja viittaa datan muuntamiseen tiedoksi.

Tilastollista analyysia voidaan käyttää myös tilastotutkimusprosessin sisäisen laadun seuraamiseen. Tällöin analyysin tavoitteena on tuottaa parannuksia identifioimalla virhelähteitä, arvioimalla niiden merkitystä tietojen luotettavuuden kannalta ja ehdottamalla toimenpiteitä virheiden korjaamiseksi tai eliminoimiseksi.

Tässä luvussa tarkastellaan estimaattoreita ja niiden laatuun liittyviä ominaisuuksia sekä tilastolliseen analyysiin liittyviä laatuäkökohtia.

Periaatteet

Tärkeimmät estimaattoreiden luotettavuuteen liittyvät teoreettiset ominaisuudet ovat harhattomuus, tarkkuus ja täsmällisyys. Estimaattori on harhaton, jos kaikista mahdollisista otoksista laskettujen estimaattien keskiarvo on sama kuin tuntemattoman parametrin arvo. Tällöin estimaattorin odotusarvo on yhtä kuin parametrin arvo. Tätä lievempi ominaisuus, joka on käytännössä useammin voimassa, on estimaattorin tarkentuvuus. Estimaattori on tarkentuva, jos odotusarvo lähennee parametrin arvoa, kun otoksen kokoa kasvatetaan, ja yhtyy parametrin arvoon, kun otoskoko on sama kuin perusjoukon koko. Tarkkuus viittaa estimaattien vaihteluun odotusarvonsa ympärillä. Mitä pienempi vaihtelu eli varianssi on, sitä tarkempi on estimaattori. Täsmällinen estimaattori on sekä harhaton (tai ainakin tarkentuva) että tarkka. Tilastokeskuksessa pyritään käyttämään mahdollisimman täsmällisiä estimaattoreita, esimerkiksi Horvitz-Thompson-tyyppisiä estimaattoreita, yleistettyjä regressioestimaattoreita tai kalibrointiestimaattoreita.

Erityisesti estimaattoreiden varianssiominaisuuksiin voidaan vaikuttaa sisällyttämällä käytävissä olevaa, otoksen ulkopuolista lisäinformaatiota otanta-asetelmaan tai estimointiasetelmaan. Lisäinformaatiota saadaan eri tilasto- ja rekisterilähteistä. Esimerkkejä lisäinformaation sisällyttämisestä otanta-asetelmaan ovat ositettu otanta ja pps-otannon (*Probabilities Proportional to Size*) eri versiot. Regressioestimointi ja suhde-estimointi ovat esimerkkejä lisäinformaation sisällyttämisestä estimointiasetelmaan. Estimointi paranee, jos lisäinformaatio korreloi voi-

makkaasti tulosmuuttujan kanssa. Tämä näkyy estimaattorin varianssiestimaatin pienemisenä eli tehokkuuden paranemisenä. Tehokkuutta mitataan asetelmakertoimella, joka saadaan laskemalla käytettyyn otanta- ja estimointiasetelmaan liittyvän varianssiestimaatin suhde varianssiestimaattiin, joka perustuu vastaavan kokoiseen yksinkertaiseen satunnaisotantaan.

Lisäinformaatiota voidaan sisällyttää estimointiasetelmaan myös vastauskadon aiheuttaman harhan oikaisemiseksi. Esimerkiksi otantatutkimuksen havaintoaineistossa voi olla yksikkökatoa, jonka aiheuttaman harhan oikaisemista varten aineistoa painotetaan uudelleen. Silloin alkuperäisen (tavallisesti Horvitz-Thompson) estimaattorin painorakennetta muokataan tai käytetään jälkiositukseen tai kalibrointiin perustuvaa estimaattoria. Uudelleenpainottamisen huomioon ottavan estimaattorin harhan tulee olla mahdollisimman pieni ja varianssin pienempi kuin korvattavassa estimaattorissa.

Pieniä osajoukkoja tarkasteltaessa perusestimattorit eivät usein toimi riittävän täsmällisesti, jolloin joudutaan käyttämään erityisiä pienalue-estimaattoreita. Erittäin malliperusteiset pienalue-estimaattorit voivat olla otanta-asetelman suhteen harhaisia, mutta niin pieniä varianssiltaan, että niiden käyttäminen on perusteltua. Vastaavasti analyttisessä tutkimuksessa regressiomallin sovittaminen voi makkaasti korreloivassa aineistossa saattaa tuottaa yllättäviä (harhaisia) tuloksia. Tässäkin tapauksessa voidaan harhaton pienimmän neliösumman estimaattori joutua korvaamaan harhaisella, mutta tarkentuvalla harjaestimaattorilla.

Suosituksukset

Yleinen menettelyohje on valita estimointi- ja analyysimenetelmät siten, että havaintoaineiston ominaisuudet tukevat niitä mahdollisimman hyvin. Siten alkuperäisen otanta- tai koeasetelman huomioon ottaminen estimoinnin ja analyysin yhteydessä on välttämätöntä luotettavien tulosten saamiseksi. Tämä voidaan toteuttaa käyttämällä asetelmaan reagoivia asetelmaperusteisia menetelmiä.

Otosaineistot

Alkuperäinen otosasetelma on säilytettävä estimointiasetelmaa luotaessa. Siten alkuperäiset tiedot osituksesta, ryvästyksestä, moniasteisuudesta ja -vaiheisuudesta samoin kuin tiedot todennäköisyydellä yksi poimituista alkioista on pidettävä ennallaan. Jos otosalkiot ovat siirtyneet poiminnan jälkeen asetelman suhteen tilasta toiseen, esimerkiksi yritystutkimuksessa yrityksen ositteena oleva toimiala on muuttunut, tulee muuttuneet tiedot estimoida tutkimuksen tilanteen mukaisina osajoukkoina rikkomatta ositusrakennetta.

Otoksen ulkopuolista informaatiota on syytä hyödyntää aina kun sitä on luotettavasti saatavissa. Lisäinformaatiota voidaan käyttää mm. uudelleenpainotuksessa, esimerkiksi erilaisilla kalibrointiestimaattoreilla (ml. jälkiositus, suhde-estimointi tai regressioestimointi). Uudelleenpainotuksen avulla otosjakaumat saadaan lähemmäs perusjoukon jakaumia.

Piste-estimaattien luotettavuutta arvioidaan keskivirheillä tai niistä johdetuilla variaatiokertoimilla (suhteellinen keskivirhe siten, että alkuperäinen otosasetelma ja siitä johdettu estimointiasetelma otetaan mahdollisimman tarkasti huomi-

oon. Monimutkaisten otanta-asetelmien tapauksessa sekä tilanteissa, joissa estimoidaan epälineaarisen estimaattorin (esimerkiksi regressiokerroin) varianssia, estimaattorin otosvarianssia ei yleensä voida johtaa analyttisesti, jolloin joudutaan turvautumaan approksimaatioihin. Menetelmää valittaessa on sovellettava asetelman suhteen niin oikeaa approksimaatiota kuin mahdollista (ei kuitenkaan sellaista approksimaatiota, joka tuottaa liian pienen otosvarianssin). Tilasto-ohjelmissa tavallisin approksimointimenetelmä on linearisointimenetelmä; muita ovat otoksen uudiskäyttökäytännöt, kuten jackknife, tasapainotetut puoliotokset tai bootstrap. Menetelmää valittaessa on syytä neuvotella menetelmäasiantuntijan kanssa.

Ellei erityistä syytä ilmene, on piste-estimaattien ja niiden otosvarianssien laskemisessa käytettävä jotain vakiintunutta ja asetelmaan sopivaa valmisohjelmistoa (kuten CLAN97, Stata, SUDAAN tai WesVar).

Vastauskato on pyrittävä ottamaan huomioon estimoinnissa. Yksikkökadon vaikutusta pyritään useimmiten pienentämään uudelleenpainotusmenetelmillä, mm. kalibrointiestimaattorin avulla. Kadon otosvarianssia kasvattava vaikutus on pyrittävä ottamaan huomioon soveltamalla esimerkiksi kaksivaiheisen otannan varianssiestimaattoria tai malliperusteista muunnettua varianssiestimaattoria. Eräkadon tapauksessa puuttuvia havaintoarvoja useimmiten paikataan satunnaisesti valittujen alkiodien vastaavilla arvoilla (imputointi). Paikatut arvot tulee merkitä indikaattoreilla ja varianssiestimaattori johtaa siten, että se sisältää paikkaamisesta syntyvän varianssikomponentin.

Otosaineistoissa on usein niin kutsuttuja poikkeavia havaintoja, joiden arvoalueet ylittävät selvästi muiden havaintojen arvot. Jos havainnot ovat oikeita, ne pidetään ennallaan. Tällöin piste-estimoinnissa voidaan soveltaa mm. robusteja menetelmiä, kuten leikattuja keskiarvoja tai kvanttileja. Jos poikkeavuus aiheutuu painotuksesta, on painojen vaihtelua syytä rajoittaa.

Asetelmaperusteiseen tilastolliseen analyysiin soveltuvat esimerkiksi SUDAAN-ohjelmisto sekä eräät SAS- ja Stata -ohjelmistojen proseduurit. Menetelmävalikoima kattaa mm. lineaaristen mallien sovelluksia (regressio-, varianssi- ja kovarianssianalyysi), log-lineaariset ja logistiset mallit sekä elinaikamallien soveltamisen. Analysoitavassa aineistossa on oltava mukana painomuuttuja sekä otanta-asetelman identifiointiin liittyvät muuttujat (ositus, ryvästys). Monimutkaiseen otanta-asetelmaan perustuvaa aineistoa on mahdollista analysoida myös monitasomallien avulla (esimerkiksi MLwiN -ohjelmisto).

Muut aineistot

Poikkileikkausaineistojen käsittelyssä on usein sekä paljon havaintoja että paljon muuttujia. On luontevaa aloittaa aineiston tutkiminen perinpohjaisella kuvailevalla analyysillä, josta saadaan selville mm. muuttujien jakaumat ja niiden väliset yksinkertaiset riippuvuudet. Kuvailevan analyysin perusteella on helpompaa suunnitella aineiston tutkimisessa ja mahdollisessa mallittamisessa käytettävät muuttujien muunnokset (skaalaus, transformaatiot, muuttujien yhdistäminen esim. pääkomponenteiksi), jotta tilastollisten mallien oletukset olisivat tutkimusaineiston suhteen riittävästi voimassa.

Aikasarja -aineistoissa on niin ikään syytä tutkia sarjojen välisiä riippuvuuksia, jotta mallittaminen voitaisiin tehdä harhattomasti ja tarkentuvasti. Ongelmana on usein aikasarjojen välinen korreloituneisuus eli multikollineaarisuus ja sovitettujen mallien jäännösten autokorreloituneisuus. Menetelmiä valittaessa on pyrittävä poistamaan tällaiset haitat mahdollisimman tehokkaasti. Joissain tapauksissa voidaan joutua soveltamaan menetelmiä, jotka ovat tutkimusasetelman suhteen hiukan harhaisia, mutta tarkentuvia – esimerkkinä harjaestimaattorit.

Dokumentointi

Hyvään tilastointitapaan kuuluu selvittää koko havaintoaineiston käsittelyyn liittyvä tuotantoketju. Estimoinnin osalta se merkitsee mm. perustelua piste-estimaattorin ja siitä johdetun varianssiestimaattorin valinnalle, koeasetelmien ja sovelletun menetelmän, esimerkiksi mallin kuvaamista. Otosaineistoista on tuotettava tiedot keskeisten tulosmuuttujien piste-estimaateista ja niiden luotettavuudesta tärkeimmissä osajoukoissa. Luotettavuus merkitsee sekä arvioita piste-estimaatteihin mahdollisesti liittyvästä harhasta että täsmällisyydestä. Joissain tapauksissa kaikkea luotettavuuden arvioinnin kannalta välttämätöntä informaatiota ei kyetä sisällyttämään raportointiin, mutta silloinkin tiedot on syytä tuottaa mahdollisia myöhempiä tarpeita varten.

Kirjallisuus

- Cochran, W. G. 1977. *Sampling Techniques*. 3rd ed. New York: John Wiley & Sons.
- Diggle, P.J., Liang, K.-Y. & Zeger, S.L. 1994. *Analysis of Longitudinal Data*. Oxford: Oxford University Press.
- Djerf, K. 2000. *Properties of Some Estimators under Unit Nonresponse*. Statistics Finland. Research Reports 231. Helsinki: Hakapaino Oy.
- Goldstein, H. 1995. *Multilevel Statistical Models*. New York: John Wiley & Sons.
- Kasprzyk et al. 1989. *Panel Surveys*. New York: John Wiley & Sons.
- Kish, L. 1965. *Survey Sampling*. New York: John Wiley & Sons.
- Lehtonen, R. & Djerf, K. (eds.) 2001. *Lecture Notes on Estimation for Population Domains and Small Areas*. Reviews 2001/5. Helsinki: Statistics Finland.
- Lehtonen, R. & Pahkinen, E. 1996. *Practical Methods for Design and Analysis of Complex Surveys*. Revised Edition. Chichester: John Wiley & Sons.
- Lohr, S. 1999. *Sampling: Design and Analysis*. Pacific Grove: Duxbury.
- McCullagh, P. & Nelder, J.A. 1989. *Generalized Linear Models*. London: Chapman and Hall.
- Nummenmaa, T. Konttinen, R., Kuusinen, J. & Leskinen, E. 1997. *Tutkimusaineiston analyysi*. Porvoo: WSOY.
- Skinner, C.J., Holt, D. & Smith, T.M.F. 1989. *Analysis of Complex Surveys*. Chichester: John Wiley & Sons.
- Särndal, C.-E., Swensson, B. & Wretman, J. 1992. *Model Assisted Survey Sampling*. New York: Springer-Verlag.
- Wolter, K. 1985. *Introduction to Variance Estimation*. New York: Springer-Verlag.

2.12 TILASTOTIEDON ESITTÄMINEN

Soveltamisala ja päämäärä

Tilastotiedon esittäminen, julkistaminen ja jakelu on arkistoinnin kanssa viimeinen vaihe tilastotutkimusten operationaalisisessa prosessissa. Tämä on huomattavan merkittävä vaihe, koska siinä määritetään miten aikaisempien vaiheiden panos saadaan hyödylliseen ja laajaan käyttöön. Tämän luvun aiheena on perinteiset esittämistavat, vaikka tietojen esittäminen liittyy myös rekistereihin. Perinteisiä esittämistapoja käytetään niin paperille painettuna kuin sähköisissä esityksissäkin (esimerkiksi Internet).

Tilastotieto – oikeammin määrällinen tieto – voidaan esittää kolmella tavalla: tekstin, taulukoiden tai kuvioiden avulla. Taulukko on ylivoimaisesti tavallisin esitystapa. Tilastokuvioita on alettu käyttää taulukon ohella yhä enemmän, koska oikein tehtynä kuvio välittää asian havainnollisemmin ja tehokkaammin, ja antaa usein aiheesta paljon syvällisemmän käsityksen kuin taulukko. Toisaalta on muistettava, että kuvion tarkkuus on aina rajallista, mutta taulukossa tarkkuutta voidaan lisätä tarpeen mukaan. Kuvio on nopean ja havainnollisen tiedonvälityksen väline, kun taas taulukon käyttö hidasta ja epähavainnollista.

Kuvioiden, taulukon ja tekstin roolit menevät hieman toistensa kanssa päällekkäin, koska usein sama asia voidaan esittää kaikilla tavoilla. Tekijän on siis mietittävä esitystapojen roolit. Kuviolla kannattaa tuoda esiin merkittävimmät tulokset ja ne tulokset, joiden viestittäminen on tärkeintä. Lisäksi joidenkin ilmiöiden luonteen kuvaaminen on mahdollista vain graafisesti. Kuvio on ylivoimaisesti paras säännönmukaisuuksien ja riippuvuuksien esittämisessä. Teksti on yleensä subjektiivisempi esitys kuin taulukko tai kuvio ja siinä voidaan tietoa eritellä laajemmin sekä liittää se laajempaan yhteyteen. Pelkästään taulukon lukujen toistaminen tekstinä ei ole suositeltavaa.

Tavoitteena pitää olla, että sekä kuvio että taulukko ovat itsenäisiä kokonaisuuksia siten, että ne sellaisenaan välittävät oleellisen tiedon eivätkä edellytä muualta haettavaa tietoa.

Taulukko

Taulumuotoisista esityksistä erotetaan kaksi perusmuotoa: asetelma ja taulukko. Asetelma on yleensä tekstin joukossa oleva pienikokoinen taulukko, joka eroaa varsinaisesta taulukosta siinä, että se on vapaamuotoisempi. Tilastotaulukko puolestaan on suhteellisen tarkkaan formuloitu esitys, mutta myös siinä on valittavana erilaisia ratkaisuja, joiden avulla taulukon luettavuutta voidaan parantaa.

Taulukon soluissa voi olla monenlaista tietoa, useimmiten lukuja. Taulukon tulisi sisältää mahdollisimman homogeenisiä alkioita, ideaalisessa tapauksessa yhden muuttujan samalla tavalla mitattuja arvoja. Asioiden saattaminen helposti vertailtavaan muotoon on yksi taulukon tavoitteista. Jos vertailu ei ole järkevää asioiden luonteen vuoksi, myös taulukon mielekkyys voi olla kyseenalainen.

Taulukon osat

Taulukon otsikko on varsinaisen taulukon ulkopuolella ja vastaa kysymyksiin Mitä? Missä? Milloin? Taulukon nimen tulee olla kuvaava, pelkistetty ja ytimekäs. Jos taulukon kaikki alkiot esittävät samaa asiaa, myös muuttuja on hyvä esittää otsikossa.

Sarakeotsikot voivat olla hierarkkisia siten, että ylemmän tason otsikolla sidotaan yhteen alemman tason otsikoita. Alimman tason sarakeotsikoissa ilmoitetaan (yleensä) sarakkeissa käytetty muuttuja ja mittayksikkö, jos ne pysyvät samana kaikissa sarakkeen soluissa. Sarakeotsikoiden kokonaisuutta kutsutaan joskus taulukkopääksi.

Riviotsikot voivat olla pidempiä kuin sarakeotsikot. Riviotsikot voivat myös olla hierarkkisia siten, että vasemmalla olevilla otsikolla sidotaan yhteen oikealla olevia otsikoita. Oikean puolimaisissa otsikoissa ilmoitetaan (yleensä) rivillä esitetty muuttuja ja mittayksikkö, jos se pysyy samana koko rivillä.

Tämän lisäksi taulukoissa voi käyttää apuviivoja helpottamaan taulukon lukua.

Suosituksia taulukoille

- Lukuja ei pidä esittää liian monella merkitsevällä numerolla, ja pyöristys on aina suoritettava matemaattisesti oikein.
- Rivien ja sarakkeiden keskiarvot/summat esitetään silloin, kun se on loogista ja helpottaa vertaamista.
- Taulukoissa esitettävien rivien ja sarakkeiden järjestyksen on oltava looginen.
- Sijoita vertailtavat asiat päällekkäin, koska taulukossa lukuja on helpompi verrata sarakkeittain kuin riveittäin.
- Taulukon visuaalisen asettelun tulee olla selkeä.
- Taulukkoja varten on laadittu joukko vakiintuneita merkintöjä erikoistilanteita (esim. puuttuva tieto) varten. Malli löytyy esimerkiksi Suomen tilastollisesta vuosikirjasta.

Tilastokuvio

Taulukosta poiketen tilastokuvion tärkein tehtävä on tiedon välittäminen visuaalisesti. Kuvioden tekemisessä on vaarana tahaton vääristely, ellei tunne tekniikkaa. Kuvion laatijan vastuulla on, että kuvio antaa kuvion lukijalle oikean käsityksen asioiden todellisesta luonteesta eli välittää oikeaa tietoa.

Tilastokuvio on taulukkoa parempi esittäessä aineiston rakenteellisia ominaisuuksia, tiivistettäessä suuria tietomääriä, osoitettaessa asioiden keskinäistä yhteyttä, välittämässä ajatuksia ja johtopäätöksiä tai esimerkiksi luomassa asiayhteyden haluttu tunnelma.

Kuvion osat

- Kuvion otsikon tulee vastata kysymyksiin Mitä? Missä? Milloin? Joissain tilanteissa niin sanottu tarinaotsikko on hyvä ratkaisu. Joissain tilanteissa (esimer-

- kiksi tutkimusraportissa) otsikon voi korvata kuviotekstillä. Tosin otsikko ja kuvioteksti eivät sulje toisiaan pois.
- Koordinaattiakselit on nimettävä riittävällä tarkkuudella ja ilmoitettava niissä käytetty muuttuja ja mittayksikkö.
 - Koordinaattiakselien asteikkomerkinnot tulee olla selkeitä ja riittävän suurella kirjasinkoolla tehtyjä.
 - Kaikki kuvioelementit (viivat, osat, sektorit jne.) on nimettävä.
 - Kuvion tulkintaa ja johtopäätöksiä helpottavia apuvälineitä tulee käyttää riittävästi, mutta ei häiritsevän paljon.
 - Mahdolliset poikkeavuudet on tuotava esiin ja selvítettävä.

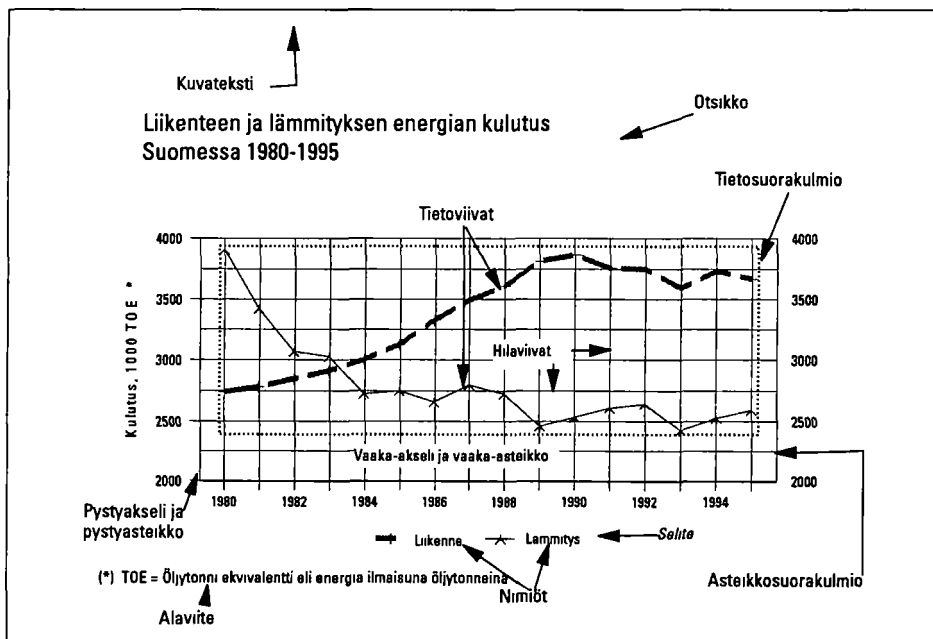
Yleisimmin käytettyjä kuviotyyppejä ovat viivakuviot, pylväskuviot, piirakkakuviot ja teemakartat. Monilla näistä kuviotyypeistä on lukuisia alatyyppejä. Lisäksi on joukko harvinaisempia, mutta erittäin informatiivisia kuviotyyppejä.

Viivakuviot ja pystyylväskuvio

Viivakuviolle ja pystyylväskuvio on toistensa vaihtoehtoja. Niille on yhteistä, että vaaka-akselilla on jatkuva-arvoinen ominaisuus, yleensä juokseva aika, ja että ne soveltuvat parhaiten aikasarjojen esittämiseen. Vähimmäisvaatimus vaaka-akselin muuttujalle on, että asteikolla on järjestysominaisuus.

Viivakuviot korostaa vaihtelua ja kehityssuuntaa eli trendiä. Pylväskuvio korostaa määrää ja määrän muutoksia. Tästä johtuen viivakuviossa y-akselin asteikon katkaisemisella on vain vähän merkitystä kuvion sanoman kannalta, mutta pystyylväskuviossa y-akselin katkaisu turmelee visuaalisen tiedonvälityksen.

Kaavio 2.6 Tilastokuvion osat ja termistö



Vaakapylväät

Vaakapylväät soveltuvat luokkiin tai ryhmiin liittyvien määrien kuvaamiseen. Pystyakselilla olevan ominaisuuden ei tarvitse olla jatkuva-arvoinen. Vaakapylväskuvio ei ole vaihtoehto pystypylväskuviolle.

Piirakkakuviot

Piirakkakuviossa ympyrä jaetaan sektoreihin ja sektoreiden pinta-ala välittää määrällisen tiedon. Tämä on otettava huomioon silloin, kun piirakoita sijoitetaan vierekkäin ja piirakoiden koko on suhteessa esitettävien asioiden "kokoon" (ympyrän pinta-ala kasvaa suhteessa säteen neliöön).

Piirakkakuviolla voidaan välittää vain suhteellista jakaumaa (prosenttijakaumaa) koskevaa tietoa ja se välittää tietoa huomattavasti epätarkemmin kuin viiva- tai pylväskuvio. Piirakkakuviota ei pidä käyttää, jos haluaa välittää tieto tarkasti. Rakenteellisen epätarkkuutensa vuoksi piirakkakuvioiden yhteydessä on perusteltua käyttää myös lukuarvoja, jotka kertovat prosenttijakauman.

Tilastolliset teemakartat

Tilastolliset teemakartat ovat karttapohjalle tehtyjä tilastokuvioita ja niiden tarkoituksena on kertoa asian tai asioiden maantieteellinen jakauma. Teemakarttoja on useita tyyppisiä, mutta tunnetuin lienee koropleettikartta, jossa maantieteelliset alueet varjostetaan eri tummuusasteilla osoittamaan yhden asian suhteellista jakaumaa.

Näyttävyyden lisääminen kuvioihin

Kuviot herättävät usein tarpeen tehdä koristeellinen tai koristeltu esitys. Tyypillisimmillään kuvioon lisätään keinotekoinen kolmas ulottuvuus tai käytetään hahmokuviota. Värien harkitsematon käyttö kuuluu myös tähän luokkaan. Tilastojulkaisun koristelemiseksi on kuitenkin vaikea keksiä mitään perustetta.

Hahmokuviot ovat esityksiä, joissa pylväiden sijasta käytetään symbolisia hahmoja, esimerkiksi ihmisen kuvia. Näitä käytetään yleensä populaäreissa esityksissä niiden mielenkiintoa herättävän ja dramaattisen vaikutuksen vuoksi.

Symbolihahmokuva on niin vaikeasti hallittava kuviotyyppi, että sen käyttö pitäisi jättää kokonaan. Tilastografikan periaatteiden mukaan oikein tehtyjä symbolikuvioita on lähes mahdoton tehdä.

Kuvioiden näennäistä kolmiulotteisuutta suositaan luultavasti samasta syystä kuin symbolihahmokuviotakin eli sillä saadaan vaatimattomastakin asiasta vaikuttava esitys. Vain harvoissa tapauksissa kuvioon tuotettu lisäulottuvuus kantaa informaatiota. Yleensä kyseessä on esityksen koristelu, joka lisäksi haittaa tiedon välittymistä eli useimmiten kolmiulotteisuus on kuvioroinaa.

Kolmiulotteisuuden käyttö pitäisi rajoittaa vain niihin tapauksiin, jolloin kolmas ulottuvuus todella kantaa informaatiota.

Suosituksia tilastokuvioille

Kuvioiden laadinta

Parhaille tilastokuvioille on ominaista muun muassa, että ne:

- Välittävät tiedot visuaalisesti ja toimivat ilman niiden yhteyteen liitettyjä lukuja.
- Välittävät suuren määrän tietoa pienessä tilassa. Parhaimmat tilastokuviot tiivistävät suuren tietojoukon ja toisaalta parista luvusta ei kannata kuviota tehdä.
- Eivät vääristä tietoihin liittyvää sanomaa. Tilastokuvioissa määrää (tai osuutta) kuvaava tieto esitetään kuviona ja eri määriä esittävien kuvioiden suhde on oltava sama kuin määrien suhde. Siis kaksi kertaa suurempaa määrää kuvataan kaksi kertaa pidemmällä pylväällä. Tällöin valekertoimen arvo on 1.00.
- Esittävät tiedot liittäen ne asiayhteyteensä. Riittävällä otsikoinnilla ja muulla oheisinformaatiolla esitys saadaan liittymään todellisuuteen.
- Antavat katsojalle monitasoisen oivalluksen. Parhaissa tilastokuvioissa on monta tasoa: niistä näkee jotakin silmäyksellä, lähempi tutustuminen antaa jotakin uutta tietoa ja tutkimalla kuviota löytää vielä jotakin.
- Saavat katsojan kiinnostumaan kuvion sanomasta, ei kuviosta sinänsä. On muistettava, että katsoja todennäköisesti on kiinnostuneempi asiasta kuin esitystekniikasta.

Viivakuviossa

- Vaaka-akselilla on oltava ominaisuus, jolla on jatkuva mitta-asteikko, esim. aika.
- Määräasteikon eli pystyakselin asteikon ei ole välttämättä alettava nolasta.
- x- ja y-akselin asteikon suhde, niin sanottu aspektisuhde, vaikuttaa ratkaisevasti kuvion ulkoasuun ja suhde on määrättävä niin, että tasaista kasvua kuvataan 45° kulmassa kulkevalla suoralla.
- Sekä y- että x-akselin asteikkojen on oltava tasavälisiä, vain logaritmiset asteikot ovat poikkeus.
- Kuviossa ei saa olla liikaa tietoviivoja ja kunkin viivan on erotuttava selkeästi muista viivoista.

Pystyylväskuvioissa

- Vaaka-akselilla on oltava ominaisuus, jolla on jatkuva mitta-asteikko, esim. aika.
- Pystyakselin asteikkoa (määräasteikkoa) ei saa katkaista, vaan sen on aina alettava nolasta.
- Sekä pysty- että vaaka-akselin asteikkojen on oltava tasavälisiä.
- Pylväiden väli on 25 % - 50 % pylväiden leveydestä.
- Pylväsryhmissä on korkeintaan kolme pylvästä rinnakkain.
- x- ja y-akselin asteikkojen suhde, aspektisuhde, vaikuttaa kuvion ulkoasuun.

Vaakapylväskuvioissa

- x-akselin asteikkoa ei saa katkaista. Sen sijaan ylipitkiä pylväitä voi katkaista yläpäästä.
- Pylväsryhmissä on korkeintaan kolme pylvästä rinnakkain.
- Pylväiden väli on 25 % - 50 % pylväiden leveydestä.
- Pylväät on sijoitettava suuruusjärjestykseen, yleensä pisin ylimmäksi.

Piirakkakuviossa

- Sektoreita saa olla korkeintaan kuusi.
- Asiat järjestetään siten, että sektorit ovat piirakassa suuruusjärjestyksessä, suurimmasta pienempään.
- Sektorit lähtevät joko klo 12:sta myötäpäivään tai yleisemmin klo 3:sta vastaapäivään.
- Rinnakkain piirrettyjen piirakoiden pinta-ala voi välittää tietoa, mutta (jopa saman kokoisten) piirakoiden vertailu on vaikeata.

Tunne yleisösi ja mediasii

Kuviota tehtäessä on pidettävä mielessä, ketkä kuviota lukevat ja otettava huomioon yleisön vastaanottokyvyn erilaisuus. Esimerkiksi laajalevikkiseen sanomalehteen ja suppealle asiantuntijapiirille ei kannata tehdä kuviota samalla periaatteella. Myös esitystapa (televisio, dia, esityskalvo, lehti, Internet-sivu, tms.) on otettava huomioon, koska jokainen media toistaa esityksen eri tavoin. Myös esitysfoorumeilla on merkitystä.

Tunne taitosi ja rajoituksesi

Pahiten epäonnistuneet esitykset ovat luultavasti syntyneet siksi, että kuvion tekijä on sortunut yleiseen uskomukseen "kuka tahansa osaa tehdä tilastokuvion tilastografiikan perusteisiin tutustumatta".

Älä tyydy grafiikkaohjelmien esitykseen

Grafiikkaohjelmia on monen tasoisia. Valitettavasti yksikään niistä ei ole täydellinen. On suuri kiusaus tyytyä siihen, mitä grafiikkaohjelmat antavat myöten, vaikka tulos ei olisi kovin hyvä. Myös toinen äärimmäisyys, ohjelmistojen monista mahdollisuuksista innostuminen saattaa tuottaa epäonnistuneen lopputuloksen.

Älä tyydy grafiikkaohjelmien oletusarvoihin

Tilastografiikkaa varten on erikoisohjelmia, joilla pystyy tekemään tavallisimmat kuviot. Ohjelmistojen oletusasetukset on tehty lähinnä myyntitarkoituksiin. Usein syntyykin huono tulos, jos tyydytään ohjelmistojen oletusarvoihin. Kokeilemalla eri vaihtoehtoja löytää useimmiten paremman ratkaisun. On hyvä huomata, että ohjelmistojen tekijöistä vain harvat ovat perehtyneet tilastografiikkaan ja ohjelmistoja kehitetään pitkälle markkinointia silmällä pitäen.

Käytä harkiten kolmatta ulottuvuutta

Kolmiulotteinen esitystapa on perusteltu vain siinä tapauksessa, että kolmanteen ulottuvuuteen sisältyy tietoa, jonka esittäminen muulla tavoin on vaikeata.

Tasapainota tekstin taulukoiden ja kuvioiden osuus esityksessä

Kullakin tiedonvälitystavalla on oma roolinsa ja tehtävänsä. Kuvioilla kannattaa esittää keskeiset ja mielenkiintoiset asiat sekä sellaiset asiat, jotka vaativat vakuuttelua. Taulukko on hyvä suurten tietomäärien esittämiseen.

Kirjallisuus

- Kuusela, V. 1994. Hyvin laadittu tilastokuvaio. Teoksessa: Melkas J. (ed.) *Tilastokeskuksen ulkoisen tiedottamisen käsikirja*. Helsinki: Tilastokeskus.
- Kuusela, V. 2000. *Tilastografikan perusteet*. Helsinki: Tilastokeskus ja Oy Edita Ab.
- Tufte, E.R. 1983. *Visual Display of Quantitative Information*. Cheshire, CT: Graphics Press.
- Wainer, H. 1997. *Visual Revelations: Graphical tales of fate and deception from Napoleon Bonaparte to Ross Perot*. New York: Copernicus.

2.13 AIKASARJAT JA KAUSITASOITUS

Soveltaisala ja päämäärä

Aikasarja on joukko ajan suhteen järjestettyjä havaintoja muuttujasta, jotka jokin satunnaisprosessi on tuottanut. Havaittu aikasarja on satunnaisprosessin eräs reaalisaaio. Aikasarja-analyysi pyrkii päättämään havaittavan aikasarjan perusteella, millainen satunnaisprosessi on luonteeltaan. Aikasarja-analyysissä kuvataan satunnaisprosessin käyttäytymistä matemaattisesti, estimoidaan malliin liittyvät parametrit, tutkitaan sarjan jaksollisuutta sekä laaditaan mallin avulla sarjan ennusteita. Varsin paljon käytetty menetelmä on myös aikasarjan kausitasoitus. Siinä aikasarjasta poistetaan vuoden sisäinen eli kausivaihtelu, jotta sarjan yleisestä kehityksestä saadaan parempi kuva. Laatukontrolliin liittyvissä menetelmissä taas sarjan analysoinnin perusteella pyritään vaikuttamaan itse sarjan tuottaneeseen prosessiin.

Kausivaihtelulla tarkoitetaan aikasarjan kehitykseen liittyvää toistuvaa, säännöllistä vuoden sisäistä vaihtelua, jonka aiheuttaa vuodenaikojen vaihtelusta johdettavat ilmiöt sekä muut vuosirytmien liittyvät ilmiöt kuten lomien, juhlahäät ja kulutustottumukset. Kausitasoituksella tarkoitetaan kausivaihtelun estimoimista ja sen vaikutuksen eliminoimista aikasarjasta. Kausitasoitettu sarjasta on helpompi havaita sarjan yleiseen kehitykseen ja suhdannevaihteluihin liittyviä ilmiöitä. Kausitasoitus tekee myös eri kuukaudet keskenään vertailukelpoisiksi ja siitä havaitsee paremmin sarjan kriittiset kohdat kuten käännepeitteet. Kausikomponentin ja

tasoitettun sarjan näkeminen saattaa myös helpottaa sarjan mallittamista. Joskus saatetaan olla kiinnostuneita myös itse kausikomponentista ja sen ominaisuuksista.

Yleisesti käytetyt perinteiset kausitasoitusmenetelmät eivät perustu mihinkään hyvin määriteltyyn malliin, vaan ovat käytännön kautta syntyneitä laskuoperaatioita. Kausitasoitus turmelee myös sarjojen sisäisen ja eri sarjojen välisen riippuvuusrakenteen, mikä tekee kausitasoitettun sarjan mallittamisen hankalaksi.

Toimintaperiaate

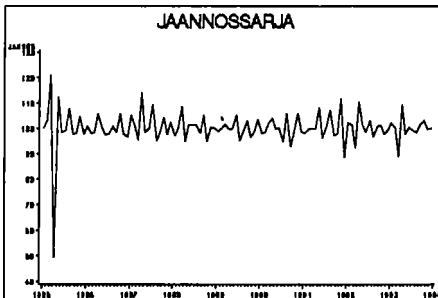
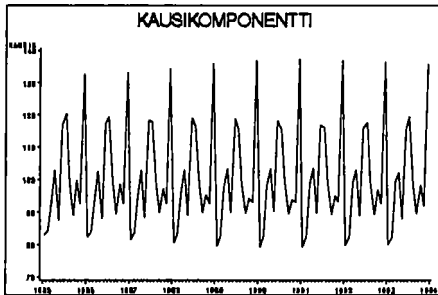
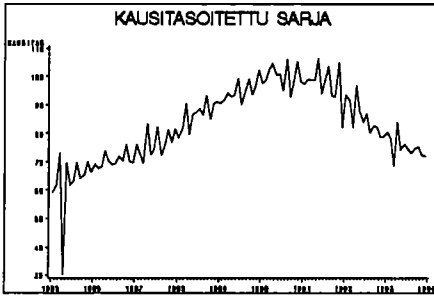
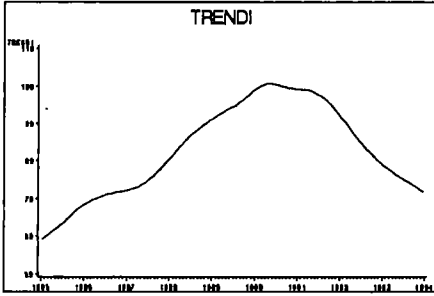
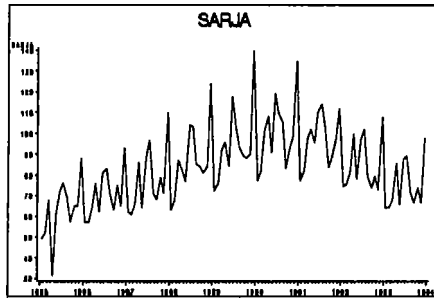
Perinteisesti on ajateltu, että taloudelliset ja muut yhteiskunnalliset aikasarjat koostuvat eri komponenteista (ks. Kaavio 2.7).

1. **Trendi** kuvaa sarjan pitkään ajanjaksoon liittyvää lähinnä rakenteellisista syistä johtuvaa vaihtelua.
2. **Suhdannevaihtelu** eli **sykli** kuvaa useamman vuoden mittaista lähes jaksollista taloudellisista syistä johtuvaa vaihtelua.
3. **Kausivaihtelu** kuvaa lähinnä vuodenaikojen vaihtelusta johtuvaa vuosittain toistuvaa säännöllistä vaihtelua.
4. **Epäsäännöllinen vaihtelu** kuvaa sitä sarjan ennalta arvaamatonta vaihtelua, jota ei voi sijoittaa kolmeen edellä olevaan komponenttiin. Epäsäännölliseen vaihteluun kuuluu myös muutamat erikoistapaukset kuten esimerkiksi lakon aiheuttamat poikkeuksellisen pienet tai suuret arvot.

Usein aikasarjasta ei ole riittävästi havaintoja, jotta trendi ja suhdannevaihtelu voitaisiin erottaa toisistaan ja niitä käsitellä yhdessä. Tällöin puhutaan lyhyesti vain trendistä. Kausitasoitettussa sarjassa on siis kausikomponentti estimoitu ja se on poistettu sarjasta. Kausitasoitettu sarja koostuu tällöin sarjan trendistä ja epäsäännöllisestä vaihtelusta. TK:ssa on viime vuosina tasoitetuista sarjoista poistettu myös epäsäännöllinen vaihtelu ja julkaistu myös pelkkä trendi.

Ennen sarjan varsinaista kausitasoitusta on sarjaan tehtävä tarvittavat työpäivä- ja juhlapyhäkorjaukset sekä mahdolliset lakoista tai muista erikoistapauksista johtuvat poikkeavien havaintojen korjaukset. Perinteiset kausitasoitusmenetelmät trendin ja kausikomponentin estimoimiseksi ovat liukuvan keskiarvon menetelmiä. Tunnetuin näistä on Yhdysvaltain tilastovirastossa kehitetty **X11-menetelmä**. Liukuvan keskiarvon menetelmien ongelma on se, että sarjan alku- ja loppupäätä tasoitettaessa joudutaan käyttämään eri suodinta kuin tasoitettaessa sarjan keskellä olevia havaintoja. Tämä aiheuttaa sen, että kun alkuperäiseen sarjaan lisätään havainto ja sarja tasoitetaan, niin myös aiemmin tasoitettujen arvojen muuttuvat tietyllä ajanjaksolla. Parannuksen tähän tuo Kanadan tilastovirastossa kehitetty **X11ARIMA-menetelmä**. Siinä tasoitettavasta sarjasta tehdään ARIMA-malli ja mallin avulla estimoidaan sarjan alku- ja loppupäähän uusia havaintoja ja näin saatuja havaintoja käytetään hyväksi tasoitettaessa varsinaisen sarjan alku- ja loppupäätä. Tällä menetelmällä voidaan aiemmin tasoitettujen havaintojen muuttumista pienentää. X11ARIMA on tällä hetkellä Tilastokeskuksessa käytössä oleva kausitasoitusmenetelmä.

Kaavio 2.7 Aikasarjat ja niiden komponentit



Toinen olennaisesti erilainen tapa suorittaa kausitasoitus on käyttää hyväksi aikasarjojen rakennemallien lähestymistapaa. Rakennemalleissa kausikomponentti mallitetaan yhdessä sarjan muiden komponenttien kanssa, joten sitä voidaan pitää mallipohjaisena kausitasoitusmenetelmänä. Mallin parametrit estimoidaan Kalmanin suotimen avulla. Rakennemallien avulla tasoitettaessa aiemmin tasoitettujen havaintojen muuttuminen on pienempää kuin liukuvien keskiarvojen menetelmillä tasoitettaessa, mutta heikkoutena on runsas parametrien estimointi. EUROSTAT:in standardimenetelmä kausitasoituksessa on TRAMO/SEATS ("Signal Extraction in Arima Time Series"), joka perustuu signaalin estimointiin sarjasta.

Suosituksset

- Ennen sarjan kausitasoitusta tulee testata onko sarjassa kausivaihtelua; liukuvan keskiarvon menetelmät voivat muussa tapauksessa antaa harhaanjohtavia tuloksia.
- Tasoitettaessa aikasarjoja X11ARIMA-menetelmällä SAS- tai SPSS-ohjelmistolla, tulee aina käyttää ARIMA-optiota. Jos sopivaa ARIMA-mallia ei löydy vakiomallien joukosta, niin ohjelmisto käyttää tavallista X11-menetelmää.
- Ennen kausitasoitusta tulee aikasarjaan tehdä tarvittavat työpäivä-, juhlapyhä- ja lakkokorjaukset. Näiden vaikutus lopputulokseen saattaa olla suurempi kuin käytettävän menetelmän.
- Trendin ja kausitasoitettun sarjan loppupään arvot saattavat revisioitua huomattavasti, kun alkuperäiseen sarjaan lisätään uusi havainto ja sarja tasoitetaan. Niitä ei tule käyttää laskennallisessa mielessä arvioitaessa sarjan viimeaikaista kehitystä.
- Aggregaattisarjat koostuvat komponenttisarjoista ja niiden kausitasoituksessa voidaan käyttää joko suoraa tai epäsuoraa menetelmää. Epäsuorassa menetelmässä aggregaattisarja muodostuu kausitasoitetuista komponenttisarjoista. Tällöin on kausitasoitettava ainoastaan ne komponenttisarjat, joissa on havaittavissa selvää kausivaihtelua, ja jätettävä loput komponenttisarjat tasoittamatta. Suorassa menetelmässä kausitasoitetaan aggregaattisarja ja additiivisuuden säilyttämiseksi komponenttisarjoihin viedään tasoituksessa syntynyt ero.

Kirjallisuus

- Dagum, E. 1980. *The X11ARIMA Seasonal Adjustment Method*. Ottawa: Statistics Canada.
- Dagum, E. 1988. *The X11ARIMA/88 Seasonal Adjustment Method- Foundations and User's Manual*. Ottawa: Statistics Canada.
- Den Butter F.A. & Fase M.M. 1991. *Seasonal Adjustment as a Practical Problem*. Amsterdam: North-Holland.
- Gomez, V. & Maravall A. 1996. *Programs TRAMO and SEATS. Instructions for the User*. Banco de Espana - Servicio de Estudios.
- Harvey, A. C. 1989. *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge: Cambridge University Press.

- Nyblom, J. & Ritola, V. 1996. *Esikorjaukset kausitasoitettavissa vähittäiskaupan aikasarjoissa*. Helsinki: Tilastokeskus. (Muistio).
- Nyblom, J. 1994. Aikasarjojen rakennemalleista. Julkaisussa: *Suhdannekäännne ja taloudelliset aikasarjat*. (Toim.) Pahkinen, E., Ritola, V. Tilastokeskus. Tutkimuksia 210. Helsinki: Hakapaino Oy.
- Nyblom, J. 1998. *Taustaa TRAMO/SEATS-kausitasoitusohjelmalle*. Helsinki: Tilastokeskus. (Muistio).
- Ritola, V. 1994. Tukku- ja vähittäiskaupan aikasarjat. Julkaisussa: *Suhdannekäännne ja taloudelliset aikasarjat*. (Toim.) Pahkinen, E., Ritola, V. Tilastokeskus. Tutkimuksia 210. Helsinki: Hakapaino Oy.
- SAS-Institute. *SAS/ETS Software: Applications Guide 1*. SAS Institute, Inc.
- SAS-Institute. *SAS/ETS User's Guide*. SAS Institute, Inc.
- Shiskin, J., Young, A. H. & Musgrave, J. C. 1967. *The X-11 Variant of the Census Method II Seasonal Adjustment program*. Technical Paper No. 15. Washington, DC: U.S. Department of Commerce.

2.14 INDIKAATTORIT JA INDEKSIT

Soveltamisala ja päämäärä

Mittaamalla esimerkiksi samaa taloudellista ilmiötä ajassa saadaan tilastoaineisto, joita kutsutaan aikasarjaksi. Tyypillisiä aikasarjoja ovat esimerkiksi hyödykkeiden hinnat ja määrät, työntekijöiden palkat ja tehdyt työtuntien määrät, arvopapereiden hinnat ja määrät. Aikasarja on pelkistetyksi jono lukuja, jotka saadaan mittamalla täsmällisesti kontrolloitua samaa ilmiötä ajassa. Aikasarjat kuvaavat siis tämän ilmiön kehitystä ajassa.

Ongelmalliseksi aikasarjojen hyödyntämisen tekee se, että jo hyvin rajatuissa ja suppeissa päätöstilanteissa tutkittavaa ilmiöaluetta voidaan esittää monilla erilaisilla ei-yhteismitallisilla aikasarjoilla. 'Aikasarjoja on olemassa yksinkertaisesti liian monta, jotta niiden sisältämän informaation voisi kerrallaan hallita. Niitä on tarve yhdistää siten, että menetetään mahdollisimman vähän informaatiota' (Törnquist, 1974 s.21). Yleensä taloustoimija ei ole kiinnostunut yksittäisistä aikasarjoista, vaan niiden tilastollisesta yleistyksestä. Tätä menettelyä, joissa useiden aikasarjojen tilastollista informaatiota "tiivistetään" harvoiksi tunnusluvuiksi, kutsutaan aikasarjojen aggregoinniksi. Yksinkertaisin esimerkki aikasarjojen aggregoinnista saadaan laskemalla aikasarjahavainnoista kullekin ajankohdalle aritmeettinen keskiarvo. Hinta-aggregoinnissa aritmeettisiä keskiarvoja kutsutaan tapauksesta riippuen joko yksikköhinnoiniksi tai yksikköarvoiksi.

Yksikköhintojen tai -arvojen mittaamisessa oleelliseksi kysymykseksi muodostuu se, minkälaisella suhteellisella osuudella tai suhteellisella painotuksella kukin aikasarjahavainno vaikuttaa kokonaisuuteen. Mikäli suhteellinen painotus muuttuu jatkuvasti, aggregoidut aikasarjahavainnot eivät ole vertailukelpoisia ajassa. Hintojen ja palkkojen aggregoinnissa tämä tilanne realisoituu, kun kulutus- tai

työvoimarakenteet muuttuvat ajassa. Ongelma voidaan ratkaista indeksiteorian avulla.

Indeksiteoria on taloustieteen osa-alue, jossa ollaan kiinnostuneita samanlaa-
tuisten ja mittayksiköistä vapaiden suhdelukujen aggregoinnista. Indeksilaskenta
ratkaisee laadullisesti erilaisten hyödykkeiden hinta- ja määräaika-
sarjojen aggregoinnin (normeeratut summat) tiivistämällä informaation indeksisarjoihin, jotka
mittaavat hyödykeryhmän hintojen ja/tai määrien suhteellista muuttumista (so.
hintojen ja/tai määrien kehitystä) ajassa.

Esimerkiksi kulutusrakenteen muuttuminen yli ajan ei synnytä laadullisia eroja
indeksisarjoihin, koska indeksikaavasta riippuen, vertailuun vaikuttavat kulutusra-
kenteet vakioidaan perus- ja/tai vertailuajankohdan mukaisiksi. Indeksilaskennal-
la pyritään vastaamaan seuraavaan kysymykseen: Paljonko enemmän/vähemmän
perusajankohdan (vertailuajankohdan) hyödykkeistä on maksettava vertailuajan-
kohdalla (perusajankohdalla) perus- ja/tai vertailuajankohdan kulutustottumuk-
silla.

Tyypilliset indeksit ovat yleensä joko hintasuhteiden tai määräsuh-
teiden painotettuja keskiarvoja – painorakenne määritellään siten, että ne edustavat joko pe-
rus- ja/tai vertailuajankohdan tilannetta. Tämä indeksilaskennan metodi tarkoittaa
esimerkiksi hinnanmuutoksen arvioinnissa sitä, että vaikka hinnat ja määrät
muuttuvat samanaikaisesti, niin määrien muutoksen ei sallita 'sokea' tavoitteena
olevan hinnanmuutoksen mittaamista.

Indeksiteoria jakaantuu pelkistetysti kuvailevaan (deskriptiiviseen) ja analyyt-
tiseen (talusteoreettiseen) tutkimussuuntaukseen. Analyyttinen tutkimus perus-
tuu lähtökohtaan, jossa taloustoimijoiden (kuten kuluttajien ja yritysten) olet-
taan käyttäytyvän rationaalisesti matemaattisen käyttäytymismallin mukaan. Ana-
lyyttisessä tutkimussuuntauksessa ongelmaksi muodostuu, että taloustoimijat voi-
vat muuttaa käyttäytymistä yli ajan esimerkiksi sosioekonomisen aseman tai vaik-
kapa makutottumusten muutoksen seurauksena. Usein myös taloudelliset reali-
teetit, kuten kulutusmarkkinoille tulevat uudet ja sieltä poistuvat hyödykkeet
(esimerkkeinä tietokoneet, kännykät, kirjoituskoneet ja höyrykoneet) on vaikeaa
liittää taloustoimijoiden käyttäytymismalleihin. Vastaavanlaiseen tilanteeseen
päädytään uusien ja häviävien työpanosten osalta. Talusteoriaan perustuvat inde-
ksit ovat teoriasidonnaisia ja usein matemaattisesti vaativia. Lähestymistavan
perusajatuksista ovat kirjoittaneet esimerkiksi Vartia (1976), Samuelson & Swa-
my (1974), Allen (1975), Diewert (1974, 1975).

Deskriptiivinen tutkimussuuntaus hyödyntää indeksilaskennassa tilastoaineis-
ton tyypillisimpiä tunnuslukuja (esimerkiksi keskiarvo, geometrinen keskiarvo,
keskiverto ja varianssi). Se ei edellytä teoreettisen viitekehyksen määrittelemistä
taloustoimijoiden joukolle. Pelkistetysti deskriptiivisessä lähestymistavassa indek-
si määritellään siten, että se mittaa esimerkiksi hinta- tai palkkatasojen muutosta
tai yksittäisten hintojen tai palkkojen keskimääräistä muutosta. Deskriptiivisestä
lähestymistavasta ovat kirjoittaneet Vartia (1976, 1983, 1995), Vartia Y. & Vartia
P. (1984).

Periaatteet

Indeksit ovat suhteellisen muutoksen indikaattoreita, jotka kertovat paljonko esimerkiksi määrätyn hyödykkeen tai hyödykeryhmän (hyödykekori) hinnat ja/tai määrät ovat suhteellisesti muuttuneet perusajankohdasta vertailuajankohtaan nähden. Indeksilaskennassa lähtökohdan muodostaa homogeeniset hyödykkeet ja näihin hyödykkeisiin liittyvät otannalla selvitettyt yksittäiset hinta-, määrä ja arvosuhteet. Homogeenisen hyödykkeen (Emmental juusto) tai homogeenisista hyödykkeistä muodostuvan hyödykeryhmän (esim. juustot) hinta- tai volyyymi-indeksit ovat yleensä yksittäisistä hinta- tai määräsuhteista laskettuja painotettuja keskiarvoja – painot määrätellään yksittäiselle hinta- tai määräsuhteelle joko perus- ja/tai vertailuajankohdan hinta- tai määrä tiedolla.

Yksikköarvoindeksi

Yksikköarvoindeksi mittaa keskimääräisten hintojen muutosta ajassa. Yksikköarvoindeksi esittää esimerkiksi kahden rakenteellisesti erilaisen hyödykekorin hintasuhteen perus- ja vertailuajankohdan välillä. Koska perus- ja vertailuajankohtien hyödykekorit ovat rakenteellisesti erilaisia, on mahdotonta sanoa mikä osa keskimääräisten hintojen muutoksesta on todellista hinnanmuutosta ja mikä osa hinnanmuutoksesta on kulutusrakenteiden muuttumisen aiheuttamaa. Yksikköarvoindeksi on täsmällinen hinnanmuutoksen indikaattori ainoastaan, jos perus- ja vertailuajankohdan kaikki hinta- ja/tai määräsuhteet eivät ole korreloituneita keskenään (Vartia, 1995, s.6).

Yksikköarvoindeksi (so. keskiarvojen vertailu) ei ole hyvä hintojen suhteellisen muutoksen mittari edes silloin, kun perus- ja vertailuajankohtien keskiarvot on mitattu tarkasti. Toisin sanoen, vaikka perus- ja vertailuajankohdina arvioitujen keskiarvojen keskivirheet olisivat nollia (keskiarvo vastaa perusjoukon keskiarvoa), se ei tee keskiarvojen vertailusta hyvää indeksiä, koska perus- ja vertailuajankohtien hyödykekorit ovat edelleen rakenteellisesti erilaisia. Yksikköarvoindeksin virhe syntyy siihen sisältyvän implisiittisen volyyymi-indeksin heikkouksista. Eri-laatuisten tai rakenteellisesti heterogeenisten määrien summa ei kuvaa oikein volyyymikehitystä. Asiallinen vertailu edellyttää aina hyödykekorien rakenteellista samanlaatuisuutta.

Laspeyresin, Paaschen, Edgeworthin ja Fisherin indeksit

Tyypillisesti sekä hinnat ja/tai määrät muuttuvat samanaikaisesti perus- ja vertailuajankohdilla. Tällöin luonnollisesti vertailtavat hyödykekorit ovat rakenteellisesti erilaisia, joten esimerkiksi keskimääräisten hintojen suorassa vertailussa emme voi tietää sitä, mikä osa hinnanmuutoksesta on todellista ja mikä rakenteellisen muutoksen aikaansaamaa. Indeksilaskennassa ongelma ratkaistaan konstruoimalla hypoteettinen vertailutilanne, jossa vertailtavat hyödykekorit määritellään rakenteellisesti samanlaisiksi. Hypoteettisia vertailutilanteita voidaan konstruoida ainakin seuraavat: Valitaan lähtökohdaksi perusajankohdan kulutus rakenne ja konstruoidaan vertailuajankohdan hypoteettinen kulutusmeno perusajankohdan hyödykekorille vertailuajankohdan hinnoilla. Tätä hinnanmuutoksen mittaa kutsu-

taan Laspeyresin hintaindeksiksi ja se kertoo perus- ja vertailuajankohtien välisen suhteellisen hinnanmuutoksen ehdolla, että kulutusrakenne kiinnitetään vastaamaan perusajankohdan kulutusrakennetta. Laspeyresin indeksi ei salli hintojen muutoksen aiheuttamaa substituutiovaikutusta, joten se voi joko yli- tai aliarvioida ”todellista hinnanmuutosta”. Indeksien tarkkuus suhteessa ”todelliseen hinnanmuutokseen” riippuu siitä, miten kulutuksen volyyymi ja hinnat ovat muuttuneet perusajankohdasta vertailuajankohtaan nähden. Mikäli hinnat ja volyymin muutos ovat korreloimattomia, ’indeksiharha’ häviää.

Toisessa yleisesti käytetyssä indeksissä, jota kutsutaan Paaschen indeksiksi, kulutusrakenne vakioidaan vertailuajankohdan kulutuksen mukaiseksi. Indeksillä kerrotaan kuinka paljon enemmän/vähemmän vertailuajankohdan kulutuskori olisi maksanut perusajankohdan hinnoilla vertailuajankohdan hintoihin nähden. Paaschen hintaindeksin ongelmat ovat analogisia Laspeyresin indeksin kanssa – se ei kykene huomioimaan hinnanmuutoksen aiheuttamaa kulutuksen uudelleenallokaatiota, joten se voi joko yli- tai aliarvioida ”todellista” hinnanmuutosta.

Laspeyresin ja Paaschen indeksin perusongelmaksi jää se, että niiden kulutusrakenne saadaan joko perus- tai vertailuajankohdalta. Koska Fisherin indeksi arvioidaan Laspeyresin ja Paaschen geometrisena keskiarvona, se on näitä ”tarkempi” hinnanmuutoksen mitta. Se huomioi kulutusrakenteessa tapahtuneet muutokset, joten se eliminoi kulutusrakenteen muuttumisen synnyttämää substituutioharhaa hinnanmuutoksen mittaamisessa. Fisherin hintaindeksin ongelma on se, että sillä ei ole selvää kulutuskoriturkintaa.

Toinen Laspeyresin ja Paaschen indeksien ”keskiarvona” arvioitu indeksi on Edgeworthin indeksi, jossa indeksin painot määrätään uuden ja vanhan tuotekorin keskiarvona. Koska indeksi huomioi kulutusrakenteiden muuttumisen, se on Laspeyresin ja Paaschen nähden tarkempi hinnanmuutoksen mittari.

Muita yleisiä hinnan- ja määränmuutoksen mittareita ovat Törnqvist, Vartia I ja Sato-Vartia indeksit. Näissä indekseissä hinnat, määrät ja arvot määritellään liukuvasti ajassa ja ne ovat jatkuvasti muuttuvia ja indeksit ovat näiden argumenttien suhteen ajan funktioita. Indeksit huomioivat kulutustottumuksissa tapahtuneet muutokset, joten ne ovat kiinteäpainoisia indeksejä (Laspeyres ja Paasche) tarkempia.

Indeksin konstruointistrategian suositukset (Vartia, 1976)

Ensimmäiseksi ratkaistaan indeksin käyttötarkoitus määrittelemällä:

- vertailtavien ’hyödykkeiden’ yleisluonne,
- taloustoimijat, joiden näkökulmasta suhteellista muutosta mitataan ja
- aikaperiodit tai aikajänne, joille suhteellinen muutos arvioidaan.

Toiseksi ratkaistaan indeksilaskennan teknisiä ongelmia määrittelemällä

- vertailtavien hyödykkeiden luokitteluratkaisu,
- näihin hyödykkeisiin liittyvän hinta-aineiston keruumetodi ja
- asianmukainen painorakenne.

Kolmas päätäntä liittyy indeksin laskentamenetelmiin, jossa valitaan

- indeksikaava ja
- indeksisarjojen konstruointistrategia.

Indeksit voidaan konstruoida joko kanta- tai ketjuindekseinä. Kantaindekseissä perusajankohta pidetään kiinteänä ja vertailuajankohta muuttuu liukuvasti ajassa. Ketjuindekseissä puolestaan perus- ja vertailuajankohdat ovat esimerkiksi peräkkäisiä ajankohtia (tai esim. peräkkäisten vuosien vastaavia neljänneksiä/kuukausia) ja ne muuttuvat liukuvasti ajassa.

Neljäs indeksien konstruointistrategian ongelma-alue liittyy

- hyödykkeiden laadulliseen muuttumiseen ja
- uusien ja häviävien hyödykkeiden ongelmaan indeksilaskennassa.

Uusien ja häviävien hyödykkeiden ongelmaa ratkaistaessa keskeiseksi kysymykseksi muodostuu, kuinka pitkä aikajänne perus- ja vertailuajankohtien välille sallitaan. Mitä pitempi aikajänne on, sitä varmemmin "vanhat" hyödykkeet substituoidaan uusilla. Tästä seuraa luonnollisesti, että indeksisarjat eivät ole ajassa vertailukelpoisia. Uudet ja häviävät hyödykkeet ovat enemmän kanta- kuin ketjuindeksien ongelmana. Kantaindeksien ongelmia voidaan vähentää lyhentämällä perus- ja vertailuajankohtien aikajännettä.

Laadunmuutos ja indeksilaskenta

Kuluttajahintaindeksissä hyödykkeiden laadullista muuttumista voidaan kontrolloida otannalla. Tällöin hinnanmuutosta mitataan vain samanlaatuisesta hyödykkeestä perus- ja vertailuajankohtien välillä. Esimerkiksi asunto-osakkeiden, tonttien, omakotitalojen ja palkkojen tapauksessa otantamenetelmällä ei yksistään voida ratkaista laadunmuutoksen synnyttämää indeksilaskennan ongelmaa. Aina-kin seuraavat seikat komplisoivat indeksilaskentaa:

- Tilastoyksiköt (esimerkiksi eri asunnot tai eri työntekijät) voivat vaihdella osittain tai kokonaan perus- ja vertailuajankohtien välillä.
- Mittauksen kohteissa saattaa tapahtua rakenteellisia muutoksia perus- ja vertailuajankohtien välillä.
- Tilastoyksiköiden ominaisuuksien arvostuksissa saattaa tapahtua muutoksia perus- ja vertailuajankohtien välillä.

Jos tilastoyksiköt eivät pysy muuttumattomina perus - ja vertailuajankohtien välillä, on mahdollista, että vertailuryhmään perus- ja vertailuajankohdalla sijoittuvat tilastoyksiköt ovat laadullisilta ominaisuuksiltaan erilaiset. Hinnanmuutoksen arvioinnissa ei voida olla varmoja siitä, mikä osa hinnanmuutoksesta on todellista (laatupuhdistettu hinnanmuutos) ja mikä laadullisen muutoksen aikaansaamaa.

Toinen kohta liittyy luokitellun tilastoaineiston rakenteen (esim. suhteellisesti tarkasteltuna) muuttumiseen ajassa. Esimerkiksi ammattinimikkeiden, sukupuolen, iän ja koulutustason mukaan luokiteltu työvoima voi muuttua työ-

panososuuksien suhteen perus- ja vertailuajankohdan välillä. Vastaavasti asuntokaupan suhteellinen painopiste voi muuttua alueellisesti perus- ja vertailuajankohdten välillä.

Kohdissa yksi, kaksi ja kolme esitettyjen indeksilaskennan ongelmien ratkaisumetodina käytetään hedonisia indeksimenetelmiä. Näiden menetelmien tavoitteena on eliminoida kohdissa yksi, kaksi ja kolme esiin tuodut laadullisten muutosten vaikutukset hinnanmuutoksen mittaamisessa. Menettely johtaa tilanteeseen, jossa todellinen hinnanmuutos (Log-muutoksella mitattuna) jaetaan osatekijöilleen seuraavasti: Todellinen hinnanmuutos \equiv laaturuhdistettu hinnanmuutos + laadullisten ominaisuuksien muuttumisesta syntyvä hinnanmuutos.

Metodisesti todellisen hinnanmuutoksen jakaminen osatekijöilleen voidaan toteuttaa karkeasti luokitellen kahdenlaisilla hedonisilla laadunvakiointimenetelmillä: (i) Griliches -tyyppisellä hedonisella menetelmällä (Griliches, 1971), jossa keskitytään kohdissa yksi ja kaksi esitettyjen indeksilaskennan ongelmien eliminoimiseen. Tyypillisiä Griliches -menetelmällä toteutettuja indeksejä ovat asunto-osakkeiden hintaindeksi ja jälleenmyytyjen autojen hintaindeksi. (ii) Toisessa hedonisessa laadunvakiointimenetelmässä, jota kutsutaan hedoniseksi imputoinniksi (Koev, 1997) indeksiongelmat ratkaistaan kahdessa vaiheessa. Ensimmäisessä vaiheessa arvioidaan tilastoyksiköille (esimerkiksi myydylle kiinteistöille tai ammattiryhmän työntekijöille) laaturakioituidut hinta-arviot ja toisessa vaiheessa näistä hinta-arvioista konstruoidaan laaturuhdistetut hintasuhteet samaan tapaan kuin tyypillisessä indeksilaskennassa. Hedonista imputointia on sovellettu metalliteollisuuden palkkojen muutoksen mittaamiseen, omakotitalojen ja tonttien indeksisovellukseen, laaturakioituu työ hintaindeksiin ja sektoreiden välisten palkkaerojen estimointiin (Valpas, Kunpas ja Kirpas -tutkimukset).

Molemmille hedonisille menetelmille on tyypillistä hintojen määräytymisen mallintaminen. Hintojen ajatellaan määräytyvät ehdollisina odotusarvoina ehdolla tilastoyksiköiden ominaisuudet. Griliches -tyyppisessä menetelmässä ominaisuuksien hintavaikutukset ovat estimointiperiodeilla vakioisia, joten siinä ei huomioida mahdollisia makutottumusten muutoksia (ominaisuuksien markkina-arvostusten muutoksia) ajassa. Hintamallit ovat tyypillisesti (vakiokerroin) kovarianssianalyysi-malleja, joissa estimointijakson ajankohdat eritellään toisistaan aikaosoitinmuuttujilla. Aikaosoitinmuuttujien arvioidut kertoimet kertovat suoraan hintojen suhteellisen muutoksen perusajankohtaan nähden. Menetelmässä ratkaistaan laadunmuutoksen ja indeksilaskennan ongelmat samanaikaisesti. Arvioidut hinnanmuutokset ovat proportionaalisia kahden ajankohdan välillä kaikille tilastoyksiköille.

Hedonisessa imputoinnissa laaturakiointi ja indeksilaskenta toteutetaan menetelmällisesti kahdessa vaiheessa. Ensimmäisessä vaiheessa perus- ja vertailuajankohdten hintamallit spesifioidaan funktiomuodon ja muuttujien suhteen samantavaksi. Perus- ja vertailuajankohdten hintamallit estimoidaan toisistaan riippumattomasti, joten ominaisuusmuuttujien kertoimia ei sidota niitä estimoidessa. Tämän jälkeen laadunvakiointi suoritetaan seuraavasti. Arvioimme perus- ja vertailuajankohdten hintamallien avulla, paljonko perusajankohdan hyödykkeet maksaisivat ominaisuuksineen ja 'rakenteineen' (esimerkiksi alue-, hyödyke- tai työpanosuokitus) vertailuajankohdalla vertailuajankohdan ominaisuuksien ja raken-

teellisten tekijöiden markkina-arvostuksilla? Vertailtavien hyödykkeiden ominaisuudet ja luokittelun rakenteelliset tekijät vastaavat perusajankohdan tilannetta, joten perus- ja vertailuajankohtien hyödykkeet ovat laadullisesti samanlaisia. Hinnanmuutos voidaan arvioida näin perusajankohdan hintojen ja laatuvarioitujen vertailuajankohdan hintojen välillä esimerkiksi tyypillisen Laspeyresin kaavan avulla. Vastaavalla menettelyllä voimme arvioida, paljonko vertailuajankohdan hyödykkeet maksaisivat ominaisuuksineen ja 'rakenteineen' (esim. alueluokitus) perusajankohdalla perusajankohdan ominaisuuksien ja rakenteellisten tekijöiden markkina-arvostuksilla? Vertailtavien hyödykkeiden ominaisuudet ja luokittelun rakenteelliset tekijät vastaavat tällöin vertailuajankohdan tilannetta, joten vertailuajankohdan määrätiedoilla hinnanmuutos voidaan arvioida esimerkiksi Paaschen kaavan avulla. Hedoninen imputointi ei rajoita indeksikaavan valintaa – kaikkia tyypillisiä indeksikaavoja voidaan soveltaa hinnanmuutoksen arvioinnissa.

Laatuvakioidulle, samoin kuin mille tahansa indeksille pätee, että ne ovat ehdollisia tilastoja ehdolla indeksilaskennassa kontrolloidut tekijät. Laatuvakioiduille indekseille on tavanomaista, että indeksiarviot muuttuvat, jos hintamallin selittäjät vaihdetaan toisiksi. Siksi on huomioitava relevantit hintojen muodostumista selittävät tekijät, joiden valinta voidaan testata yleisesti hyväksytyillä tilastollisilla menetelmillä. Jos indeksilaskennassa ei hyödynnetä todellisuudessa hintoihin vaikuttavia tekijöitä, indeksisarjojen konstruointi ja niiden vertailtavuus on epämääräistä.

Kirjallisuus

- Allen, R. G. D. 1975. *Index numbers in theory and practise*. London.
- Diewert, W. E. 1974. Intertemporal consumer theory and the demand for durables. *Econometrica*. Vol. 42. 97-516.
- Diewert, W. E. 1975. Ideal log-change index numbers and consistency in aggregation. DP 75-12. University of British Columbia. Department of Economics.
- Griliches, Z. (ed.) 1971. *Price indexes and quality change*. Cambridge: Harvard University Press.
- Koev, E. 1997. *Constructing a Hedonic Wage index: Pilot Study for the Finnish Metal Industry*. (Moniste, saatavilla mm. Tilastokeskuksen Hinnat ja Palkatyksiköstä).
- Samuelson, P. A. & Swamy, S. 1974. Invariant economic index numbers and canonical duality: survey and synthesis. *American Economic Review*. Vol. 64. 566-593.
- Törnqvist, L. 1974. *Aikasarjojen konstruointi*. (Toim.) Kautio-Toikka, A. Helsinki: Gaudeamus.
- Vartia, Y. 1976. *Relative changes and index numbers*. Helsinki.
- Vartia, Y. 1983. Aikasarjojen konstruointi ja indeksiluvut. *Helsingin yliopiston tilastotieteen laitoksen opetusmonisteita no. 3*. Helsinki.
- Vartia, Y. 1995. *Quality and price of labour*. DP no. 366. Helsingin yliopisto.
- Vartia, Y. & Vartia, P. 1984. Descriptive index number theory and the Bank of Finland currency index. *Scandinavian Journal of Economics*. Vol. 86. 352-364.

3. DOKUMENTOINTI

3.1 TILASTOTUTKIMUKSEN DOKUMENTOINTI

Soveltamisala ja päämäärä

Dokumentoinnilla tarkoitetaan tilastollisen toiminnan kuvaamista. Tilastolliseen toimintaan kuuluvat käytetyt käsitteet, määritelmät, menetelmät ja itse tilastotuotantoprosessi tietojärjestelmineen. Lisäksi on dokumentoitava myös selvitykset tiedon laadusta ja vertailtavuudesta. Tilastollinen dokumentointi voidaan jakaa tilastojen käyttäjille ja tuottajille tarkoitettuun dokumentointiin.

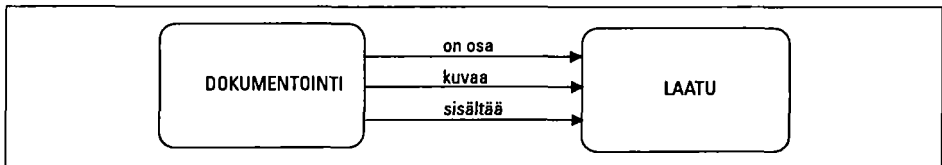
Tilastojen käyttäjien on dokumentoinnin avulla:

- saatava selville mitä tilastoja ja tilastollista dataa (tiedostoja) on olemassa
- löydettävä tarvitsemansa tilastot ja data
- pystyttävä tulkitsemaan ja analysoimaan tilastoja
- pystyttävä käsittelemään dataa.

Tilastojen tuottajille tarkoitettu dokumentointi auttaa tilastontekijää:

- tekemään tilastoja luotettavasti ja tehokkaasti
- parantamaan tilaston laatua
- ylläpitämään ja kehittämään työtapoja, tuotantoprosesseja ja tietojärjestelmiä
- keräämään tietoa muita vastaavia prosesseja varten
- perehdyttämään ja kouluttamaan uusia työntekijöitä.

Kaavio 3.1 Tilastotutkimuksen dokumentoinnin ja kokonaislaadun yhteys (Sundgren, 2001)



Dokumentointi voidaan jakaa tilastollisten prosessien ja tilastotuotteiden dokumentointiin. Prosessiorientoitunut dokumentointi kuvaa miten tuote valmistetaan; tilaston käyttäjä tarvitsee puolestaan tuoteselosteen tilastosta ja sen laadusta.

Periaatteet

Dokumentoinnin tavoite on olla täydellinen, yksiselitteinen ja monikäyttöinen kuvaus tilastosta ja sen tuotannosta. Kuvauksien olisi oltava ajantasaisia, hyvin organisoituja, helposti saavavilla, tiiviitä ja tarkkoja. Myös yhteenvedo tuloksista sisältyy dokumentointiin.

Kaikki tehty dokumentointi tulisi tallettaa koko viraston yhteisiin tietokantoihin (Koonti, Luokitus, Käsite, Systemirekisteri ja AMS), joista voidaan tuottaa

dokumentteja eri tarkoituksiin kuten jakelutietokantoihin (StatFin), arkistoon ja muihin sähköisiin välineisiin sekä paperijulkaisuihin.

Suosituks

Tarjottavan dokumentoinnin laajuus ja tarkkuus riippuu käyttäjän tarpeesta. Tilastokeskuksen metatietokannat ohjaavat kuvausten tekemistä ohjeiden, kuvausmallien ja jo rakenteensa avulla. Tilastosta ja sen tuotantoprosessista tulee pyrkiä kuvaamaan seuraavat tiedot SVT-laatukriteerien mukaisesti:

1. Tilastotietojen relevanssi

Käyttötarkoitus

- mitä tarkoitusta varten tilasto on tehty,
- historia.

Määritelmät

- tutkimuskohde,
- tietojen kerääjä ja tiedonantaja,
- käsitteet ja niiden määritelmät,
- käytetyt luokitukset.

2. Tietojen oikeellisuus ja tarkkuus

Menetelmät

- tiedonkeruumenetelmä,
- käytetyt otanta- ja estimointimenetelmät.

Aineiston tilastollinen laatu

- tavoiteperusjoukon ja kohdeperusjoukon vastaavuus,
- otantatutkimuksien kato ja katovirheen oikaisumenetelmät,
- mittaus-, prosessointi- ja otantavirheet.

3. Julkaistujen tietojen ajantasaisuus ja oikea-aikaisuus

- tiedonkeruu-aika,
- tiedon julkaisuaika,
- täsmällisyys tai luvattu aikaisuus.

4. Tilastotietojen saatavuus ja tilastoprosessin läpinäkyvyys /selkeys

Tietojen jakelu

- tiedonjakelukanavat,
- tiedon fyysinen saatavuus,
- käyttäjien tarvitsemien muotojen kattavuus,
- kuvaukset saatavilla olevasta tiedosta.

Tietojen tulkittavuus

- käytetyistä käsitteiden ja menetelmien kuvausten saatavuus,
- tulosten analysointiaste.

5. Tilastojen vertailukelpoisuus

- samaa aihepiiriä koskeviin muihin kansallisiin ja kansainvälisiin tilastoihin,
- aiemmin kerättyihin tilastoihin ja aikasarjoihin,
- suosituksiin ja standardeihin.

6. Selkeys ja eheys/yhtenäisyys

- yhtenäisyys/johdonmukaisuus muihin saman aihepiirin tilastoihin,
- erot käsitteissä ja keruuprosessissa tulee raportoida.

7. Dokumentointi

Tietojärjestelmä

- arkkitehtuuri, tietomalli, toiminta, lähdekoodi ja testaus,
- käsittelysäännöt,
- virheet ja muutokset.

Tuotantoprosessi

- prosessin eri työvaiheet.

Tiedostot

- tiedoston tekninen- ja yleiskuvaus,
- kenttien tekninen- ja yleiskuvaus,
- päättelysäännöt.

Tiedon hallintaan liittyvät metatiedot

- Tilaston nimi,
- Tilastollinen aihealue,
- Vastuuhenkilö,
- Tiedonantovelvollisuus,
- Saatavuus,
- Salassapitomääräykset ja käyttörajoitukset,
- Säilytys ja arkistointi,
- EU-regulaatiot,
- Käytetyt resurssit,
- Muut tietolähteet ja lisäinformaatio.

Kirjallisuus

- Sundgren, B. 2001. *Documentation and Quality in Official Statistics*. Q2001. 14-15.5.2001 Stockholm, Sweden. Conference CD. 1st ed. Statistics Sweden & Eurostat.
- Statistics Canada. 1998. *Statistics Canada Quality Guidelines*. Ottawa: Statistics Canada.
- Tilastokeskus. 1985. *Tilastojen laadun kuvaaminen*. Käsikirjoja 23. Helsinki: Tilastokeskus.
- von Rothstein, I. & Sundgren, B. 1996. *Sweden's Statistical Databases: Detailed Descriptions of the Metadata in the Macrodatabase*. Statistics Sweden. 1996.

3.2 ARKISTOINTI

Soveltamisala ja päämäärä

Arkistoinnin tavoitteena on turvata aineiston säilyminen ja uudelleenkäyttö yhteiskunnallista tutkimusta ja tietopalvelua varten.

Tilastokeskuksen toiminta-ajatuksessa todetaan, että Tilastokeskus toimii kansakunnan muistina. Tilastoaineistojen arkistointi on tämän tietopääoman kartuttamista. Aineistojen hyödyntämisen ja oikean käytön kannalta on tärkeää, että arkistoidut aineistot kuvataan laadukkaasti.

Nykyisin Tilastokeskuksessa kaikki konekieliset aineistot arkistoidaan keskuskoneelle standardoidussa muodossa eli koontitiedostomuodossa. Tällä taataan se, että aineistot ovat tulevaisuudessa palautettavissa käyttäjien tarpeisiin, vaikka alkuperäiset tallennusohjelmistot ja -laitteistot vanhenisivat.

Toimintaperiaate

Tilastokeskuksessa arkistoinnin tärkein työväline on Tilastokeskuksen Koonti-tietokanta. Koontitiedostojen kuvausjärjestelmässä kuvataan arkistoitavasta aineistosta sen tiedosto- ja kenttäkuvaukset eli metatiedot. Koonti-tietokanta kokoa yhteen myös muut Tilastokeskuksen metajärjestelmät: Luokitustietokannan ja Käsitietokannan, joita hyödynnetään edelleen koontitiedostojen kenttäkuvauksissa sekä Systeemirekisterin (SYSRAK) ja Arkistonmuodostussuunnitelman (AMS), joita puolestaan hyödynnetään tiedostonkuvauksissa (Kaavio 3.2).

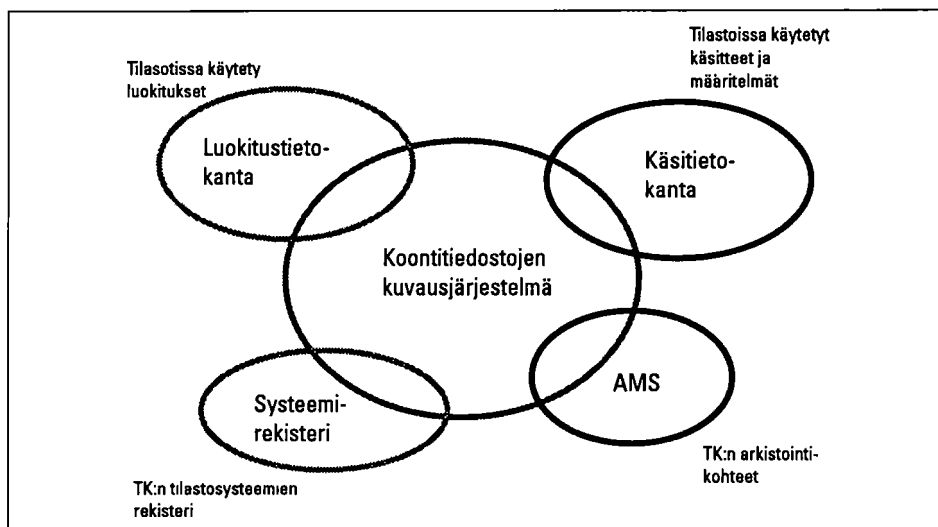
Periaatteen mukaisesti kuvaus tehdään aina vain yhteen paikkaan. Jos esimerkiksi arkistoitavassa aineistossa käytetään Luokitustietokannassa ennestään kuvattua luokitusta, ei kyseen omaista luokitusta tule kuvata erikseen, vaan se linkitetään koontitiedoston kenttäkuvaukseen luokitusliitoksen avulla. Vastaavasti tehdään myös käsiteliitos. Arkistoitavan aineiston AMS-liitos puolestaan kertoo sen, mihin arkistointikohteeseen koontitiedosto kuuluu ja systeemitunnus sen, mihin systeemiin se liittyy.

Arkistoinnissa kannattaa hyödyntää mahdollisimman paljon Tilastokeskuksen metajärjestelmiä, koska se helpottaa ja nopeuttaa joskus paljon aikaa vievää kuvaustyötä. Arkistoinnin ABC-oppaaseen on koottu ohjeet arkistoinnista ja tarvittavat sovellukset, niiden käyttö ja tuki sekä tietoa käyttäjätunnuksista ja sovellusten asentamisesta.

Arkistointitapahtuman päävaiheet ovat seuraavat:

1. Tarkista AMS-kuvauksen arkistointikohteen tiedot. Pienet muutokset voit tehdä päivittämällä kuvausta. Jos on kyse oleellisista muutoksista (esim. muutos kokonaisaineistosta otosaineistoksi), tee uusi arkistointikohde.
2. Kuvaa AMS-kuvauksen mukaiset koontitiedostot koontitiedostojen kuvausjärjestelmään (Koonti-sovellus). AMS-kuvauksessa ei luetella koontitiedostojen nimiä vaan ne aineistot, jotka arkistoidaan. Arkistointikohteeseen voi liittyä useita koontitiedostaja yksikön päätöksen mukaisesti. Koontitiedostoja ni-

Kaavio 3.2 Tilastokeskuksen metajärjestelmä



mettässä muista käyttää AMS:ssa määriteltyä systeemitunnusta, jotta tiedostot löytyvät jatkossakin.

3. Arkistoi koontitiedostot keskuskoneelle (= tee keskuskoneen koontitiedosto ja tee arkistointipyyntö). Kun arkistointi on tehty, Koonti-sovelluksessa oleviin koontitiedostoihin tulee Säilytys -kohtaan merkintä arkistoitu.
4. Täytä AMS :n luovutuslomake ja lähetä se mahdollisen paperiaineiston kanssa tilastoarkistoon. Jos arkistointikohteeseen liittyy useita koontitiedostoja, jotka valmistuvat eri aikaan, niin täytä luovutuslomake vasta, kun kaikki koontitiedostot on arkistoitu.

Arkistointiprosessin eri vaiheet on kuvattu kaavioon 3.3.

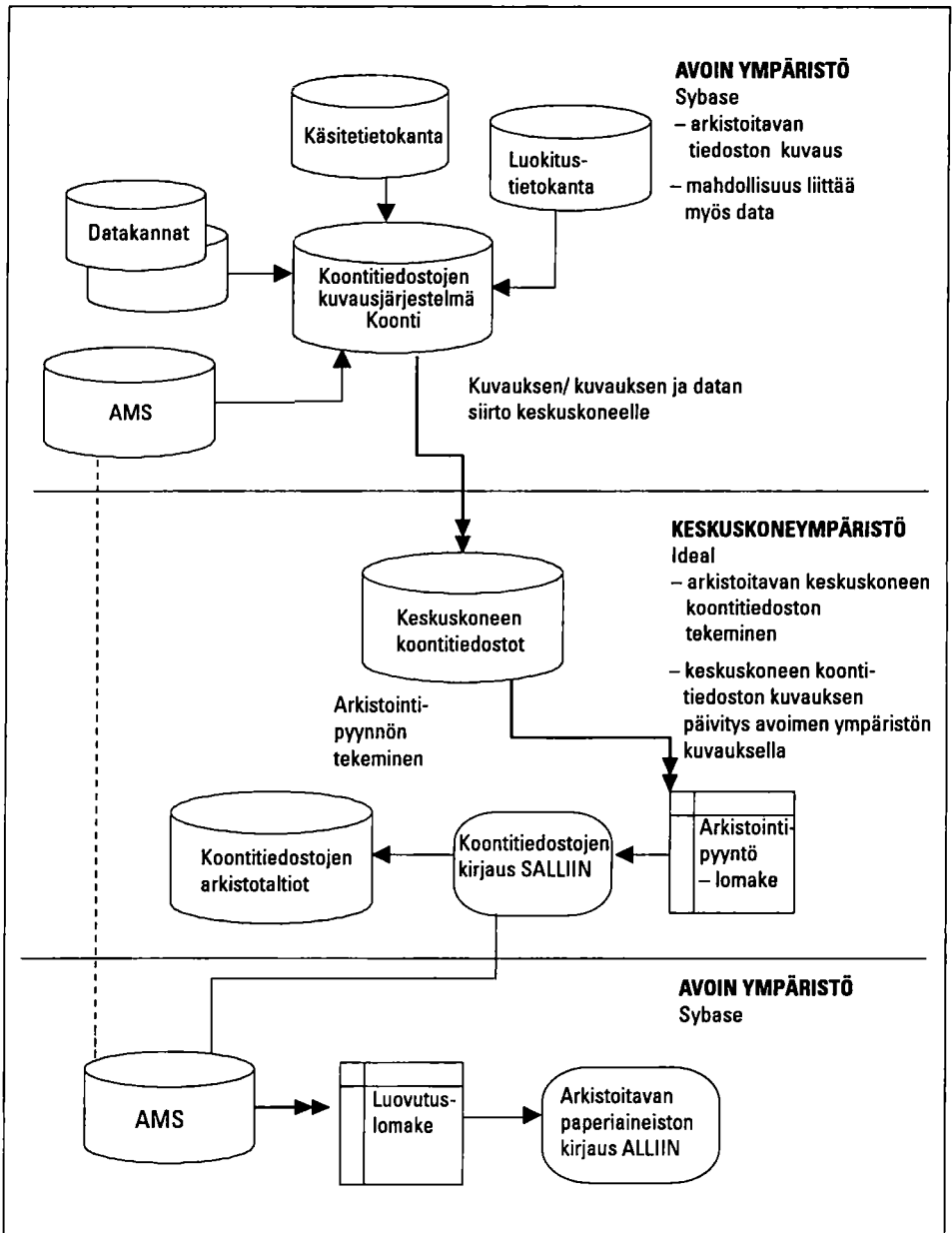
Suosituks

Kun Koonti-sovellusta käytetään arkistointiin, koontitiedoston kuvaus on tehtävä huolellisesti, jotta aineistojen hyödyntäminen ja oikea jatkokäyttö tulisi turvatuksi. Aineistojen jatkokäytön osalta keskeistä on, että metatiedot (kuvaus) ovat laadukkaita.

Aineistojen kuvauksen on vastattava kuvattavaa datatiedostoa. Ennen arkistointikuvauksen tekoa tulee selvittää arkistoitavan datatiedoston tiedot (minkä niminen se on, mikä versio arkistoidaan, missä se sijaitsee, kenellä siihen on oikeudet, mitä ja minkälaisia muuttujia siinä on jne). Tämän lisäksi tulee perehtyä arkistoinninoheisiin (ks. Koivuniemi, 2001).

Kuvauksen tekeminen on aina yhteistyön tulos: hyvä kuvaus pitää sisällään ymmärryksen tilaston tuotantoprosessista ja niistä keskeisistä päätöksistä, joita kulloinkin on tehty, jotta kyseinen aineisto on saatu aikaiseksi. Tämä ymmärrys on hyvä välittää myös jatkokäyttäjän tietoon.

Kaavio 3.3 Arkistointiprosessin vaiheet



Tiedostonkuvaus on hyvä tehdä heti, kun tietosisältö on valmis. Tämän ajoitus, riippuu tuotantoprosessin luonteesta. Kun kuvaus tehdään riittävän ajoissa samaa kuvausta voi käyttää myös muihin, arkistointia edeltäviin, tarpeisiin, esimerkiksi verkkopalveluissa StatFin -taulukoiden metakuvauksiin.

Kun olet tehnyt kuvauksen, tarkistuta se vielä koontitiedoston hoitajalla. Hän merkitsee kuvauksen hyväksytyksi. Vain hyväksytyn kuvauksen voi arkistoida.

Kirjallisuus

- Jouhki, S. 1999. *Käsitietokanta*. Tilastokeskus. (Muistio).
- Koivuniemi, P. & Toikka, A. 2001. SAS-tiedostojen arkistointiohje. SAS-tiedostojen arkistointiohje. Tilastokeskus.
- Koivuniemi, P. 2001. *Arkistoinnin ABC - Perustietoja arkistoinnista ja metatietojen kuvaamisesta*. Tilastokeskus. (Muistio).
- Poukka, R. 2001. *Tilastoaineistojen arkistoinnin yleisperiaatteet*. Arkistoprojekti. Tilastokeskus.
- Räikkönen, T. 2001. *AMS-sovelluksen käyttäjän ohje*. AMS-projekti. Tilastokeskus.
- Saijets, M. & Toivonen, P. 1995. *Keskuskoneen koontitiedostojen kuvausjärjestelmä*. Versio 1. Helsinki: Tilastokeskus.
- Saijets, M. & Toivonen, P. 1995. *Keskuskoneen koontitiedostojärjestelmän käyttöohje*. Versio 2, toinen korjattu painos. Helsinki: Tilastokeskus.
- Syvänperä, R. & Jouhki, S. 2000. *Käsitietokannan käyttö*. Tilastokeskus. Manuaali.
- Tilastokeskus. 2001. *Käsitietokanta- tarkennettu työsuunnitelma*. Käsitietokantaprojekti.
- Toikka, A. 2001. *SAS-tiedoston vienti koontitiedostoksi SASKOTI3-makrolla*. Tilastokeskus. (Ohje).

4 JULKISTAMINEN, JAKELU JA ASIAKASPALVELU

4.1 JULKISTAMINEN

Tavoite ja päämäärä

Kaikki Tilastokeskuksen valmistuvat tilastot tulee julkistaa. Julkistamisella välitetään tieto tilaston, tuotteen tai palvelun valmistumisesta ja saataville tulosta sekä kerrotaan keskeisten tilastojen uutisarvoinen sisältö laajempaan julkisuuteen. Julkistamistoiminnallaan Tilastokeskus pyrkii saattamaan tuottamansa tiedon viivytyksettä käyttöön. Samalla Tilastokeskus hoitaa myös yleistä, valtion viranomaisille määrättyä tiedotusvelvollisuutta.

Tilastokeskuksen julkistamistoiminnan välineitä ovat julkistamiskalenterit, lehdistötiedotteet ja -tilaisuudet sekä julkaisemansa lehdet.

Julkistamiskalentereillaan Tilastokeskus pyrkii lisäämään toimintansa ennakoitavuutta niin, että tiedon tarvitsijat voivat omassa toiminnassaan ottaa huomioon uuden tiedon ilmestymisen. Sisäisesti julkistamiskalenterit parantavat Tilastokeskuksen prosessien toimivuutta ja asiakastoimintoja niin tietopalvelussa, markkinoinnissa kuin tiedotuksessakin. Näin eri toiminnoista vastaavat voivat suunnitella toimintaansa ennakolta julkistamiseen nähden sekä toimia yhteisten käytäntöjen mukaisesti.

Periaatteet

Julkistamisen yleisperiaatteena on välittää aina julkisuuteen tieto uuden tiedon, toisin sanoen tilaston tai tutkimuksen valmistumisesta. Uuden tiedon ensijulkistaminen tapahtuu Tilastokeskuksen julkistamiskalentereissa. Julkistamiskalenterien toimittamisesta ja jakelusta vastaa Tilastokeskuksen viestintäyksikkö.

Vuoden alussa ilmestyvässä Tilastojen julkistamiskalenterissa esitetään Tilastokeskuksen tilastojen julkistamisajankohdat päivämäärän tai viikon tarkkuudella. Sisäpiirisääntöjen alaisten tilastojen ja joidenkin muiden suhdannetilastojen julkistamispäivämäärät ilmoitetaan päivän tarkkuudella. Viestintäyksikkö kerää julkistamistiedot kalenteriin tilastoyksiköistä kyselyllä. Kalenterissa kerrotaan myös Suomen Pankin, Tullihallituksen, työministeriön, maa- ja metsätalousministeriön tietopalvelukeskuksen, Valtiokonttorin ja Metsäntutkimuslaitoksen tilastojen julkistamistietoja.

Tilastokeskuksen aina perjantaisin ilmestyvässä viikkokalenterissa tarkennetaan julkistamispäivät kaikille seuraavalla viikolla ilmestyville Tilastokeskuksen tilastoille, julkaisuille ja muille tuotteille. Viikkokalenterissa kerrotaan tulevan viikon lehdistötiedotteet julkistamispäivän ja -kellonajan tarkkuudella. Sekä Tilastojen julkistamiskalenteri että viikkokalenteri jaetaan tiedotusvälineille sekä julkaistaan Tilastokeskuksen Internet-sivuilla. Viikoittain kuvataan otsikko- ja jul-

kaisusarjatietoja laajemmin myös niiden uusien tilastojen ja tuotteiden sisältö, joista ei laadita lehdistötiedotetta.

Lehdistötiedotteessa kerrotaan uuden tiedon uutisarvoisin sisältö, kuvataan uusia tuotteita ja palveluita tai uutisoidaan tilaisuuksia ja tapahtumia. Tiedotteet välitetään sähköpostina tai faksina tiedotusvälineille ja muille asiakkaille sekä julkistetaan viraston Internet-sivuilla, jolloin niihin voi liittää oheisaineistoa esimerkiksi StatFin -palvelusta.

Tilastokeskuksella on käytössään poikkeuksellisen laaja kosketuspinta suomalaisen julkisuuteen, mikä antaa erinomaiset mahdollisuudet myös Tilastokeskuksen tuotteiden markkinointiin. Siksi tuotteiden ja palveluiden tiedotus- ja markkinointisuunnittelu pitäisi tehdä rinnakkain. Joukkoviestinten luottamuksen säilyttämiseksi uutis- ja markkinointiviestintä on pidettävä julkistamisessa toisistaan erillään, mutta julkisuudessa esillä ollut tilastouutinen on hyvä perusta markkinointiviestinnälle. (Tilastokeskus, 2000).

Julkistamisohjeessa (Tilastokeskus, 2000) on määritelty tarkasti Tilastokeskuksen yleiset tiedotustoiminnan periaatteet, jotka muodostavat myös tilastojen julkistamiskäytäntöjen laadun arvioinnin ulottuvuudet. Tietoja julkistettaessa tulee kiinnittää huomiota seuraaviin asioihin (ks. myös luku 1.1.2 Ammattietiikka):

– **Tietojen luotettavuus:**

Tiedot ovat oikeita ja niiden luotettavuustaso on esitetty (ks. 4.3.1 SVT:n Tilastojulkaisut ja tuoteselostesuositus).

– **Tasapuolisuus:**

Tiedot julkistetaan ennalta ilmoitettuna ajankohtana ja ne julkistetaan niin, että kenellä tahansa on mahdollisuus saada ne käyttöönsä mahdollisimman pian julkistamishetken jälkeen.

– **Ajantasaisuus:**

Tiedot julkistetaan niin nopeasti kuin mahdollista.

– **Havainnollisuus:**

Tiedot välitetään havainnollisessa muodossa, yleisölle annetaan riittävät mahdollisuudet myös omien johtopäätösten tekoon, tiedon käyttäjien tarpeet otetaan huomioon (ks. 2.11 Tilastollinen estimointi ja analyysi, 2.12 Tilastotiedon esittäminen, ja 4.2 Tilastotiedon levittäminen).

– **Puolueettomuus:**

Leimaamista ja kannanottoja yhteiskunnallisiin kiistakysymyksiin vältetään, mutta tietoa tuotetaan myös kiistanalaisista asioista.

– **Tulkinta:**

Tietoja tulkitaan ja analysoidaan kuvaamalla eri ilmiöiden mittasuhteita (onko suuri/pieni verrattuna aiempaan kehitykseen tms.), kertomalla mistä ne johtuvat ja mitä niistä seuraa. Yksittäisen tilaston tiedot liitetään ja rinnastetaan mahdollisuuksien mukaan muihin samaan asiaan liittyviin tilastotietoihin ja muuhun mahdollisesti relevanttiin informaatioon.

– **Ajankohtaisuus:**

Tietoja julkistettaessa ne kytketään ajankohtaiseen yhteiskunnalliseen keskus-

teluun ja teemoihin, ajankohtaisista asioista tuotetaan tilastotietoa julkisuu-
teen oma-aloitteisesti.

– **Avoimuus:**

Luotettavaa tilastotietoa ei salata. Tällä periaatteella sovitetaan yhteen julkisen
hallinnon avoimuusperiaatetta ja edellä kohdassa kolme mainittua tasapuo-
lisuusperiaatetta.

– **Ohjaus:**

Tiedon käyttäjää pyritään opastamaan tiedon järkevään hankintaan ja luotetta-
van tiedon lähteille; kun julkistetaan muiden kokoamaa tilastotietoa tuodaan
tiedon tuottaja selkeästi esiin.

– **Myönteisyys:**

Tiedotuksessa annetaan todenmukainen myönteinen kuva Tilastokeskuksesta
ja sen palveluista sekä pyritään lisäämään niiden tunnettavuutta.

Suosituks

Uutta tietoa julkistettaessa keskeistä on ennakkointi, luotettavuus ja virheettö-
myys. Huolellinen julkistamisen ennakkosuunnittelu vähentää virhemahdolli-
suutta sekä uuden tiedon tuotannossa, itse tuotteessa että julkistamisessa. Sen si-
jaan ennakkosuunnittelu lisää tuotantoprosessin sujuvuutta ja vähentää työkus-
tannuksia. Tietojen virheettömyys ja julkistamisaikataulun noudattaminen vah-
vistavat Tilastokeskuksen yhteisökuvaa luotettavana tiedon tuottajana.

Hyvän asiakaspalvelun takaamiseksi on syytä pyrkiä siihen, että lehdistötiedote
ja siihen liittyvä julkaisu tai muu tuote ilmestyvät samaan aikaan. Lisäksi suositel-
laan, että viikkokalenterissa tiedotetaan myös niistä hankkeista, joissa Tilastokes-
kus on mukana yhtenä tuottajana. Tällaisia ovat lähinnä erilaiset yhteistutkimuk-
set ja -julkaisut ministeriöiden ja tutkimuslaitosten kanssa. Jo tutkimussopimuksia
tehtäessä olisi syytä käsitellä ja sopia tutkimuksen tiedottamisesta ja Tilastokes-
kuksen mahdollisesta roolista siinä.

Tilastokeskuksen tiedon tuotannossa ja julkistamisessa mukana olevien on pe-
rehdyttävä Tilastokeskuksen julkistamisohjeeseen. Ohjeessa annetaan yksityis-
kohtaisia neuvoja julkistamiskalentereista sekä lehdistötiedotteista, -tilaisuuksista
ja -seurannasta. Ohjeisiin tutustuminen ja niiden noudattaminen opastaa käytän-
nön työssä, yhdenmukaistaa julkistamisprosessia, parantaa yksiköiden välistä yh-
teistyötä ja turvaa Tilastokeskuksen julkistamistoiminnan häiriötöntä sujumista.

Sisäpiirisääntöjen alaisiksi määriteltyjen tilastojen (kokonaistuotannon kuu-
kausikuvaaja, työvoimatutkimuksen kuukausitiedot, teollisuustuotannon volyy-
mi-indeksi, kuluttajahintaindeksi, tuottajahintaindeksit) julkistamisessa on lisäksi
noudatettava niitä koskevia määräyksiä ja ohjeita (Valtionvarainministeriön pää-
tös eräiden tilastotietojen salassapitamisesta ennen niiden julkistamista).

- Melkas J. 1994. *Tilastokeskuksen ulkoisen tiedottamisen käsikirja*. Helsinki: Tilastokeskus.
- Tilastokeskus. 2001. *Ohjeita TK:n julkaisupalvelujen asiakkaille. Tilastokeskuksen julkaisuperiaatteet*. TP/Julkaisupalvelut. (Muistio).
- Tilastokeskus. 2000. *Tilastokeskuksen julkistamisohje*. Tietopalvelu/Tiedotus. (Muistio).
- Valtioneuvoston kanslia. 1994. *Valtionhallinnon tiedotusohje*. Valtioneuvoston tiedotusyksikkö. Helsinki: Painatuskeskus.
- Valtiovarainministeriön päätös eräiden tilastotietojen salassapitamisesta ennen niiden julkistamista. 27.1.1994/36.
- Eurostat. 1998. *Statistics and quality*. Sigma. 3/1998.

4.2 TILASTOTIEDON LEVITTÄMINEN

Soveltamisala ja päämäärä

Tilastotiedon levittäminen käsittää tehtävänä niin veloituksettoman tiedon kuin maksullisten tilastopalvelujen ja -tuotteiden tunnetuksi tekemisen, asiakkaiden tietotarpeiden kartoittamisen sekä asiakkaiden johdattamisen tarvittavan tiedon oikealle lähteelle. Asiakkaiden ja laajemminkin tiedon tarvitsijoiden tietotarpeiden kartoittamisella on suuri merkitys sekä tilastotiedon keruun suunnitteluvaiheessa että lopputuotteiden käytön edistämiseen liittyvässä tarjonta- ja markkinointityössä. Näin toimien laadukkailla tilastopalveluilla on mahdollista edistää julkisen hallinnon ja yritysten tietoon ja tosiasioihin perustuvaa päätöksentekoa ja toimintaedellytyksiä, luoda edellytyksiä yhteiskuntatutkimukselle ja auttaa kansalaisia muodostamaan luotettavan käsityksen yhteiskunnasta.

Tilastotietoa ja tilastopalveluja koskevan monipuolisen viestinnän tavoitteena on palvella jo olemassa olevia asiakkaita sekä myös tiedottaa olemassa olevasta tiedosta, jotta myös potentiaaliset asiakkaat löytäisivät tarvitsemansa tiedon tai tiedostaisivat tietotarpeensa. Tilastotuottaja-organisaation tulee jalostaa, yhdistää tietoa ja tiedottaa tuottamastaan tiedosta siten, että käyttäjille turvataan relevantin ja ajantasaisen tiedon helppo löytäminen sekä se, että tieto on helposti siirrettävissä asiakkaalle käyttökelpoisessa muodossa. Vastaavasti tilastotiedon viestittämisessä tulee huomioida tietoyhteiskunnan tiedon tulva ja aktiivisesti markkinoida kohdennetusti uusinta tilastotietoa.

Toimintaperiaate

Tuotetun tilaston saatavuudesta tulee tiedottaa mahdollisemmin kattavasti. Ihanetilanteessa tilastollisen palvelutoiminnan tehokasta toteutumista edistetään tietoarkkitehtuurilla, jonka puitteissa kaikki Tilastokeskuksessa valmistettu aggregoitu tilastotieto jatkuvasti koostetaan yhtenäiseen, monipuoliseen ja helpokäyttöi-

seen palvelutietokantaan, joka varsinaisen tilastoinformaation lisäksi sisältää myös kaikki tarvittavat meta- ja luokitustiedot ja josta tiedon irti ottaminen voi tapahtua helppokäyttöisen hakumoottorin avulla.

Nykytilanteessa tilaston saatavuudesta tiedotetaan eri kanavia hyödyntäen. Maksuttomasta tiedosta tiedotetaan mm. lehdistötiedottein, Tilasto-oppaassa sekä StatFin -tietokannan kautta. Maksullisesta tilastosta tiedotetaan, paitsi kohdennetulla suoramarkkinoinnilla, myös lukuisten foorumien avulla: Tilastotori, Tilastokeskuksen palvelukatalogi, Tilasto-opas, asiakaslehti Stat.Fi sekä mahdollisesti myös lehdistötiedotteilla.

Ulkomaailmalle tarkoitettun viestinnän lisäksi on tärkeää muistaa tiedottaa tilastoista myös Tilastokeskuksen sisällä. Toistaiseksi parhaita välineitä sisäiseen tiedotukseen ovat Tiimifoorumit sekä Tilastokeskuksen intranet-sivut. Muutaman kerran vuodessa järjestettävillä tuotetuntemusaamuilla esitellään Tilastokeskuksen uusia tuotteita.

Suosituksukset

- Tilastotoimen tulee pyrkiä kasvattamaan jatkuvasti tilastotuotteiden käytettävyyttä, tunnettavuutta ja käyttöastetta.
- Tilastotuotannon ja -tutkimuksen tulee tunnistaa niiden eri asiakasryhmät.
- Asiakasryhmiä tulee analysoida ja kuunnella; tilastotuotteet tulee kehittää siten, että ne vastaavat eri kohderyhmien tietotarpeita.
- Asiakaspalautteen huomioiminen on tärkeä osa toiminnan kehittämistä.
- Asiakkaat tulee johdattaa tarvitsemiensa tietojen lähteille.
- Tiedon tulvasta erotutaan oikeille kohde- ja käyttäjäryhmille kohdennetulla aktiivisella markkinoinnilla.
- Tilastokeskuksen aktiivinen vuorovaikutus asiakasryhmien kanssa mahdollistaa uusien potentiaalisten tilastotuotteiden kehittämisen.

Kirjallisuus

- Helle, M., Lautanala M., Salonen R., Turunen P. (2000). *Sähköisten tietopalvelujen segmentointi Tilastokeskuksessa*. JOKO Executive Education. Raportti 8.2.2000. Tilastokeskus.
- Pöllänen J. (1999). *Yksilömarkkinointi – Oppivan asiakassuhteen rakentaminen*. Helsinki: Kauppakaari.
- Pöllänen J. (2001). *Tietomarkkinoiden johtaminen*. Kurssimoniste. (Tilastokeskuksen tilaama Markkinointi-instituutin kurssi).
- Storbacka K., Blomqvist R., Dahl J., Haeger T. (1999). *Asiakkuuden arvon lähteillä*. Juva: WSOY.
- Storbacka K., Kaario K., Sivula P. (2000). *Arvoa strategisista asiakkuuksista*. Helsinki: Kauppakaari.
- Tilastokeskus. 1996. *Tilastokeskuksen asiakashallintaprojekti*. Keskusteluaineisto ja 1. väliraportti. Raportti 11.12.1996.
- Tilastokeskus. 2001. *Tilastokeskuksen asiakassuhteiden kehittäminen*. Tilastokeskuksen asiakassuhteiden kehittämistyöryhmä. Raportti 14.12.2001.

4.3 TILASTOTUOTTEET

4.3.1 SVT:n Tilastojulkaisut ja tuoteselostesuositus

Soveltamisala ja päämäärä

Suomen Virallinen Tilasto (SVT) on kattava ja objektiivinen tilastojulkaisusarja Suomen yhteiskunnasta. SVT-kriteerit täyttävät tilastot ja julkaisut kuuluvat SVT-sarjassa ilmestyviin tilastoihin, jos:

- julkaisun on tuottanut virasto tai laitos, joka kuuluu Tilastokeskuksen ylläpitämään luetteloon SVT-tuottajista, tai
- julkaiseva organisaatio vastaa julkaisun sisällöstä ja oikeellisuudesta taaten näin jatkuvuuden tilaston tuottamiselle samansisältöisenä. SVT-julkaisua ei voida julkaista henkilön nimellä.

SVT Tilastojulkaisu voidaan julkaista perinteisenä sarjajulkaisuna SVT sarjassa sarjajaon mukaisesti tai elektronisesti SVT:n tuottajien internet-sivuilla (ks. 1.2.3 Suomen Virallinen Tilasto (SVT)). Kaikkien SVT-sarjassa julkaistavien tilastojulkaisujen tulee täyttää Tilastokeskuksessa hyväksytyt SVT-laatukriteerit ja niissä tulee olla SVT-standardin mukainen tuoteseloste (ks. 1.2.4 Suomen Virallisen Tilaston laatukriteerit).

Tilastotiedon tuottajan vastuulla on tiedon oikeaoppinen tulkinta ja käyttö. Tämän varmistamiseksi on tiedon käyttäjiä opastettava riittävällä ohjeistuksella. Tämän vuoksi SVT-tilastojulkaisuihin on sisällytettävä tuoteselostus tilastotuotteen laadusta aina tilastotiedon julkaisemisen yhteydessä.

Tilastojen käyttäjillä tulee olla mahdollisuus arvioida mitkä tekijät aineistossa rajoittavat tilaston laatua. Harvat käyttäjät voivat itse arvioida tuotetun tilaston laatua. Tilastovirastolla on velvollisuus tehdä tarpeelliset laatuarviointit ja jakaa tehtyjen arviointien tulokset käyttäjille ajantasaisesti ja helppokäyttöisessä muodossa.

Julkisen ja helposti saatavilla olevan tuoteselosteen tekeminen ja ylläpitäminen ovat tilaston tuottajan keskeisiä tehtäviä. Tuoteselosteen antama informaatio lisää tilaston arvoa loppukäyttäjälle. Seuraavat seikat ovat tuoteselosteiden keskeisiä toiminnallisia tavoitteita:

- Auttaa tieteellisen tutkimuksen ja asianmukaisten tulkintojen teossa.
- On takuu käyttäjälle aineiston korkeasta laadusta.
- Voi kasvattaa aineiston käyttäjien määrää ja käyttöastetta.
- Mahdollistaa laadun parantamisen jatkuvissa ja toistettavissa tutkimuksissa.
- Lisää Suomen Virallisen Tilaston arvostusta ja tunnettavuutta.

Tuoteselosteen laajuutta tulee harkita julkaisukanava- ja tuotekohtaisesti. Seuraavassa on eritelty suositukset tuoteselosteiden vähimmäisvaatimuksiksi yleisesti käytetyille julkaisukanaville sekä erillisten, syvällisempien laatuselvitysten suosi-

teltava sisältö. Yleisten suositusten lisäksi on kunkin tilaston kohdalla tarvittaessa selvitettävä sen laadullisia erityispiirteitä.

Toimintaperiaate

Sarjajulkaisujen tuoteseloste ja erillinen laatuselvitys

Jokaiseen SVT-sarjajulkaisuun tulee liittää tuoteseloste. Tuoteselosteessa tulee pyrkiä tiiviiseen arviointiin tilastotutkimuksen laadusta, luotettavuudesta ja sopevuudesta eri käyttötarkoituksiin. Tuoteselosteen keskeinen tarkoitus on vastata, miten SVT-laatukriteerit täyttyvät kyseisessä SVT tilastossa.

Tuoteselosteiden lisäksi kunkin SVT julkaisusarjan kohdalla tulee arvioida sisäinen ja ulkoinen tarve erilliseen yksityiskohtaiseen laatuselvitykseen ja niiden julkaisu-tiheyteen. Tilastollinen laatuselvitys tulee tehdä jokaisesta SVT tilastosta, jonka asiakaskunta sitä tarvitsee. Erityisen tärkeä se on niiden tilastojen kohdalla, joilla on laaja aineistojen käyttäjäkunta.

Tuoteselosteen ja laatuselvityksen rakennetta ja eroja on käsitelty taulukossa 4.1. Laatuselvityksen sisällön painotuksen ratkaisee pitkälti yksittäisen tilaston erityispiirteet ja sen käyttäjäkunnan tarpeet.

Lehdistötiedotteen tuoteseloste

Lehdistötiedotteen tulee sisältää hyvin lyhyt ja tiivis tuoteseloste, josta tulee ilmetä kyseisen tilaston kannalta keskeiset laatuseloste. Esimerkiksi, tulee kertoa perustuuko tilasto kokonaistutkimukseen vai otantaan. Lisäksi tulee mainita mihin ajankohtaan tiedot perustuvat. Lehdistötiedotteessa kerrotaan myös onko kyseessä ennakkotieto vai lopullinen tieto. Tilastojen luotettavuudesta on myös mainittava lyhyesti, samoin keskeisistä käsitteistä ja määritelmistä.

Ennakkotiedotteen tuoteseloste

Tilastotietojen luotettavuus, tiedonlähde ja tiedonkeruutapa on kuvattava ennakkotietoreportoinnissa tarkemmin kuin lehdistötiedotteessa. Ennakkotietojen kohdalla tulee korostaa tiedon luonnetta ja sitä että tiedot voivat vielä muuttua. Jos ennako- ja lopullisen tiedon välisen eron tiedetään olevan systemaattista tulee tämä selvittää tiedonkäyttäjille. Lisäksi kerrotaan, milloin lopulliset tiedot ovat saatavilla.

Internet-julkaisujen tuoteseloste

Internet-sivuilla tietokannoissa ja html-sivujen muodossa julkaistujen taulukoiden yhteydessä tulee esittää lyhyt laatu kuvaus (vrt. lehdistötiedote), linkki tuoteselosteeseen ja viittaus asianomaiseen SVT-julkaisuun. Jos tilasto julkaistaan vain SVT-verkkopalvelussa/sähköisesti, on siitä myös tällöin laadittava sarjajulkaisun tuoteseloste vaatimukset täyttävä tuoteseloste, johon tulee viitata lähdeviittauksin ja/tai linkein.

Taulukko 4.1 Ehdotus standardoidun tuoteselosteen ja laatuselvityksen sisällysluetteloksi

	SVT Sarjajulkaisun tuoteseloste	SVT-tilaston erillinen laatuselvitys
1. Tilastotietojen relevanssi (relevance)	<p>1.1 Yhteenveto tuotteen tietosisällöstä ja käyttötarkoituksesta.</p> <p>1.2 Esitellään tilaston ymmärtämisen kannalta olennaiset käsitteet, käytetyt luokitukset, tutkimuskohde, tietojen kerääjä ja tiedonantajat.</p> <p>1.3 Viitataan mahdollisiin lakeihin, asetuksiin ja suosituksiin, joihin tilasto perustuu.</p>	<p>1.1 Tuotetaan tarkka yhteenveto tuotteen tietosisällöstä ja käyttötarkoituksesta. Kerrotaan, mitä ilmiötä kuvaamaan kyseinen tilastotuote on laadittu ja sen historia.</p> <p>1.2 Esitellään tilaston ymmärtämisen kannalta olennaiset käsitteet, luokitukset, tutkimuskohde, tietojen kerääjä ja tiedonantajat.</p> <p>1.3 Viitataan mahdollisiin lakeihin, asetuksiin ja suosituksiin, joihin tilasto perustuu.</p> <p>1.4 Arvioidaan miten tuotettujen tilastotietojen relevanssi vastaa sille asetettuja vaatimuksia, miten ilmiöalueen mahdolliset muutokset on huomioitu sen tilastoinnissa.</p>
2. Tilastotutkimuksen menetelmäkuvaus	<p>2.1 Tuoteselosteessa raportoidaan selkeästi tilaston tutkimusmenetelmä eli tilaston perusjoukko, käytetyt perusaineistot, tutkimusasetelma (kokonaistutkimus tai otantatutkimus), (otanta-asetelma), tiedonkeruutapa, estimointimenetelmät ja otantatutkimusten kohdalla painokertoimien käyttö.</p>	<p>2.1 Tilastotutkimuksen menetelmäkuvauksessa raportoidaan tuoteselosteen kuvausta tarkemmin lopullisten tulosten vaatimat (otanta-,) tiedonkeruu-, editointi-, imputointi-, (painutus-) ja estimointimenetelmät.</p> <p>2.2 Käytetyt menetelmät ja muutokset niissä perustellaan (myös muutosten vaikutukset aikasarjoille arvioidaan).</p> <p>2.3 Menetelmäkuvauksessa selvitetään tilastotuotannossa käytetyt tietolähteet (myös lisäinformaatiolle).</p> <p>2.4 Kartoitetaan koko tilastotutkimuksen prosessi.</p>
3. Tietojen oikeellisuus ja tarkkuus (accuracy)	<p>3.1 Osoitetaan, että tilasto mittaa sitä ilmiötä, mitä on tarkoitus mitata.</p> <p>3.2 Kerrotaan seikat, jotka voivat vaikuttaa tilaston luotettavuuteen. Myös keskeisimmät epävarmuustekijät eli mahdolliset virhelähteet (esimerkiksi kato tai alipeitto) on mainittava.</p> <p>3.3 Tarvittaessa taulukoidaan estimaateille pääluokitusten mukaan tilastollisia tunnuslukuja (esimerkiksi otantatutkimuksissa otanta-asetelman huomioivat keskivirheet, keskineliövirheet (MSE) ja otanta-asetelman tehokkuutta estimoivat tunnusluvut (deff)).</p>	<p>3.1 ja 3.2 kohdat tuoteselosteesta käsitellään tarvittaessa tarkemmalla tasolla.</p> <p>3.3 Arvioidaan tavoiteperusjoukon ja kohdeperusjoukon vastaavuus sekä käytetyn kehikon laatu.</p> <p>3.4 Keskeisimmät epävarmuustekijät eli mahdolliset virhelähteet on kuvattava ja arvioitava niiden vaikutusta julkaistuille estimaateille:</p> <ul style="list-style-type: none"> – otanta (<i>sampling errors</i>) – muut virhelähteet (<i>non sampling errors</i>): <ul style="list-style-type: none"> – kehikko (<i>frame</i>) – mittaus (<i>measurement</i>) – prosessointi (<i>processing</i>) – kato (<i>non-response</i>). <p>3.5 Taulukoidaan estimaateille tilastoinnissa käytetyin keskeisin luokituksin tilastollisia tunnuslukuja, kuten otanta-asetelman huomioivat keskivirheet, keskineliövirheet (MSE) ja otanta-asetelman tehokkuutta estimoivat tunnusluvut (deff).</p> <p>3.6 Tulkitaan kohdassa 3.5 tuotetut taulukot.</p>

Taulukko 4.1 Jatkuu

	SVT Sarjajulkaisun tuoteseloste	SVT-tilaston erillinen laatuselvitys
4. Julkaistujen tietojen ajantasaisuus (timeliness) ja oikea-aikaisuus (promptness)	<p>4.1 Kerrotaan, mitä ajankohtaa tai -jaksoa julkaistu tilasto kuvaa.</p> <p>4.2 Kerrotaan myös, onko kyseessä ennakkotieto vai lopullinen tieto.</p> <p>4.3 Mikäli tiedot voivat muuttua tulevaisuudessa (esimerkiksi kausitasoituksen vuoksi) tulee muutosepävarmuus esittää selkeästi tiedon käyttäjille.</p> <p>4.4 Mikäli tilasto julkaistaan ensin ennakkollisena, kerrotaan, milloin lopulliset tiedot ovat saatavilla.</p>	<p>4.1 Kerrotaan, milloin tiettyä ajankohtaa tai -jaksoa koskevat tiedot on julkaistu.</p> <p>4.2 Lisäksi tarkastellaan, onko tuotantoaikataulu SVT-suositusten mukaisella tavoitetasolla. Mikäli tilasto ei ilmesty SVT-suositusten mukaisessa tavoiteajassa tai viivästyy sovitusta aikataulusta, tulee kertoa tuotanto- ja julkaisuviiveeseen johtaneista syistä.</p> <p>4.3 Tarvittaessa tutkitaan miten aikasarjojen tiedot ovat muuttuneet ajassa (esimerkiksi kausitasoituksen vuoksi).</p>
5. Tietojen saatavuus ja läpinäkyvyys/selkeys (accessibility and clarity)	<p>5.1 Esitellään tilaston julkaisukanavat ja tietopalvelun yhteystiedot.</p>	<p>5.1 Esitellään tilaston julkaisukanavat (esim. paperijulkaisu, Astika, Altika, StatFin). Lisäksi kuvataan muut tietopalvelumahdollisuudet.</p> <p>5.2 Tilastoissa, joiden tiedot muodostavat vertailukelpoisia aikasarjoja, ilmoitetaan saatavilla olevien aikasarjojen pituudet.</p>
6. Tilastojen vertailukelpoisuus (comparability)	<p>6.1 Kerrotaan tilaston ajallisesta vertailtavuudesta sekä vertailtavuudesta muihin aineistoihin.</p>	<p>6.1 Kerrotaan tilaston ajallisesta vertailtavuudesta sekä vertailtavuudesta muihin aineistoihin tuoteselostetta tarkemmalla tasolla.</p> <p>6.2 Tarkastellaan vertailtavuuteen vaikuttaneita muutoksia ja niiden merkittävyyttä esimerkiksi tilaston tuotantoprosessissa, käsitteissä ja luokituksissa.</p>
7. Selkeys ja eheys/yhtenäisyys (consistency)	<p>7.1 Arvioidaan tilaston yhtenäisyys/johdonmukaisuus muihin saman aihepiiriin tilastoihin.</p>	<p>7.1 Tarvittaessa arvioidaan tilaston yhtenäisyys/johdonmukaisuus muihin saman aihepiiriin tilastoihin tuoteselostetta tarkemmalla tasolla. Tähän liittyen erot käsitteissä ja tiedonkeruuprosessissa tulee kartoittaa ja niiden vaikutus tulee arvioida.</p>
8. Dokumentointi (documentation)		<p>8.1 Kuvataan tilastotutkimuksen metatieto.</p> <p>8.2 Viitataan tarkempiin metatietotietokantoihin ja kuvataan saatavissa olevat arkistoidut tiedostot.</p>

Suosituksset

- Jokaisesta tilastosta tulee laatia tuoteseloste.
- Arvioidaan sisäinen ja ulkoinen tarve erilliseen yksityiskohtaiseen laatuselvitykseen ja niiden julkaisutiheyteen.
- Tuoteselosteen ja laatuselvityksen keskeinen tarkoitus on vastata ja selvittää, miten SVT-laatukriteerit täyttyvät kyseisessä SVT tilastossa.

Kirjallisuus

- Eurostat. 2000. *Assessment of the Quality in Statistics. Definition of Quality in Statistics*. Doc. Eurostat/A4/Quality/Definition.
- Eurostat. 2000. *Assessment of the Quality in Statistics. Standard Quality Report*. Doc. Eurostat/A4/Quality/Standard report.
- Office of Management and Budget. 2001. *Measuring and Reporting Sources of Error in Surveys*. Statistical Policy Working Paper 31. Subcommittee on Measuring and Reporting the Quality of Survey Data. Statistical Policy Office. Office of Information and Regulatory Affairs. Office of Management and Budget.
- Statistiska Centralbyrån. 2001. *Kvalitetsbegrepp och riktlinjer för kvalitetsdeklaration av officiell statistik*. Meddelanden i samordningsfrågor för Sveriges officiella statistik. Örebro: SCB-Tryck.
- Tilastokeskus. 1987. *Tilastojen laadun kuvaaminen – Ohjeita tuoteselosteiden laatimiseksi*. Käsikirjoja 23. Helsinki: Tilastokeskus.

4.3.2 Sähköiset tuotteet

Soveltamisala ja päämäärä

Tilastokeskuksen tuotantoprosessit ovat suuntautumassa entistä enemmän kohti sähköisiä jakelukanavia. Sähköisessä jakelussa on eräitä perinteisestä julkaisemisesta poikkeavia piirteitä, jotka on otettava huomioon, jotta sähköisen palvelun asettamat vaatimukset tulevat täytetyiksi.

Sähköisen tuotteen tai palvelun laatua ilmentäviä erityispiirteitä ovat:

- ajantasaisuus, jatkuva saatavuus ja jatkuva päivitettävyys (verkkojakelu)
- pyrkimys päätelaite- ja selainriippumattomuuteen (verkkojakelu)
- tietojen hyvä edelleen käytettävyys asiakkaan omissa tietojärjestelmissä sekä
- tietojen yhteensopivuus (harmonisointi) muiden tilastojen kanssa.

Toimintaperiaate

Sähköiset tuotteet ja palvelut voidaan jakaa laajalle asiakaskunnalle tehtäviin vakiotuotteisiin ja toisaalta sisällöltään, toimitustavoiltaan tai tiedostomuodoiltaan asiakaskohtaisesti räätälöityihin palveluihin.

Lisäksi tuotteet voidaan erotella jakelumedian perusteella. Online -verkkojaketu on sähköisen tilastopalvelun ensisijainen ja kasvava jakelumedia. Offline jakelua puolestaan edustavat CD-ROM -tuotteet. Levykkeitä käytetään tietojen toimittamiseen enää satunnaisissa ja asiakkaan kanssa erikseen sovituisissa tapauksissa.

Käsitteenä sähköinen jakelu on yhä enemmän muodostumassa synonyymiksi verkkojaketulle, jossa erilaisia tiedon toimitustapoja tuottajalta asiakkaalle on käytössä useita.

Perinteisin online -palvelu on tietokannan interaktiivinen itsepalvelukäyttö verkossa. Siinä asiakas saa käyttöönsä maksullisen tai maksuttoman tietokannan, jonka sisältökuvauksia selaillemalla ja niistä valitsemalla hän itsepalveluna tuottaa haluamansa taulukot. Online -itsepalvelutietokantoja edustavat StatFin -palvelu ja aikasarjatietokanta ASTIKA.

Online -palvelu voi myös koostua valmiista www-sivustoista, joissa on taulukoita, grafiikkaa ja tekstimuotoista tietoa. Tyypillisesti näillä sivuilla olevat taulukot ovat myös poimittavissa käyttäjän omiin sovelluksiin taulukkolaskentaohjelmille sopivassa tiedostomuodossa. Tällaisia www-sivustoja voi olla sekä maksullisina että ilmaispalveluina. Maksullisissa www-palveluissa on käytössä useille asiakkaille samansisältöisinä tarjottavia vakiopalveluja kuten Tilastokatsaus sekä asiakaskohtaisesti räätälöidyllä sisällöllä ja esitystavalla toteutettuja sivustoja kuten Valtioneuvoston ja eduskunnan käyttöön tehty Senaattori-palvelu.

Online -palvelun eräs erikoismuoto on CD-ROM -tyyppisten tuotteiden jakelu verkon kautta. Tuotetta ei tällöin käytetä verkon yli, vaan asennuslevyketä vastaava sisältö on asiakkaan noudettavissa verkon kautta. Tuotteen ostaja saa käyttäjätunnuksen ja salasanan, joilla voi ladata tuotteen www-palvelusta omalle koneelleen. Verkosta ladattavia tuotteita ovat mm. Kuntafakta ja Euroavain. Esimerkiksi Kuntafaktan jakelusta vuonna 1999 noin 90 % meni verkon kautta, ja tilaajat saivat tuotteen käyttöönsä yleensä jo samana päivänä kun tilaus saapui Tilastokeskukseen.

Sähköinen jakelu voi myös perustua sähköpostiin tai internet-pohjaiseen tiedostosiirtoon (FTP). Sähköposti sopii parhaiten vain sellaiseen jakeluun, jossa nopeus ja luotettavuus eivät ole ensisijaisen tärkeitä.

Parasta ajantasaisuutta vaativat toimitukset hoidetaan yleensä niin, että asiakkaan kanssa sovitun tietosisällön kantava ja asiakkaan määritysten mukaista tiedostomuotoa noudattava tietopaketti lähetetään ftp-siirtona suoraan asiakkaan omaan palvelimen vastaanottohakemistoon, johon asiakas on antanut Tilastokeskukselle lähetysoikeudet. Lähettävää ftp-jakelua (push-tekniikka tai shooting) käytetään yleensä sisäpiiritilastojen toimittamiseen heti julkaisuhetkellä kaupallisille jälleenjakelija-asiakkaille ja tietotoimistoille. Nopean ftp-jakelun lähetylähteenä Tilastokeskuksessa on pääsääntöisesti ASTIKA.

Jakelua tapahtuu myös noutoperiaatteella Tilastokeskuksen omalta ftp-palvelimelta. Palvelun tilanneet asiakkaat saavat käyttöönsä oman hakemiston (tunnus ja salasana) TK:n palvelimelta, josta käyvät sitten itsepalveluna noutamassa sinne toimitetut tiedostot. Tällaista noutavaa palvelua käytetään mm. väestötilastopakettien toimittamiseen kunnille. Tietopaketit toimitetaan palvelimelle joko ennal-

ta sovitun aikataulun mukaan tai Tilastokeskus ilmoittaa asiakkaalle aina erikseen milloin uusia tiedostoja on noudettavissa.

Verkkojakelun ajantasaisuusvaatimuksesta johtuvat järjestelyt

Tietoverkkojen kautta tapahtuva jakelu on nopeaa ja tehokasta. Usein tilastotiedot edellytetään toimitettaviksi asiakkaille välittömästi niiden julkistushetkellä. Tärkeimpiä ajantasaispalvelun asiakkaita ovat tietojen jälleenjakelijoina toimivat koti- ja ulkomaiset tietopankki-isännät sekä uutistoimistot. Nopean julkistushetkellä tapahtuvan toimituksen varmistamiseksi tiedot viedään riittävän ajoissa ennen julkistusta Tilastokeskuksen sisäiseen tietokantaan. Tarvittavat sähköiset tulosteet ehditään näin valmistella odottamaan julkistushetkellä tapahtuvaa asiakkaalle lähettämistä.

Niin sanotut herkat taloustilastot, joilla on ennalta ilmoitettu tarkka julkaisuajankohta, viedään asiakkaille toimitettavien tulosteiden tuottamiseksi Tilastokeskuksen sisäiseen suojattuun tietokantaan (sisäpiiri-ASTIKA). Tähän tietokantaan on käyttöoikeus vain kyseisen tilastojen sisäpiirivaltakirjan omaavilla työntekijöillä. Julkistamisajankohtaa odottavat sisäpiiritietoa sisältävät valmiit tulostetiedostot säilytetään niin ikään tässä järjestelmässä, johon on pääsy vain sisäpiirivaltuudet omaavilla henkilöillä.

Suosituks

Tietojen käytettävyys asiakkaan omassa tietojärjestelmässä

Sähköisessä tiedonjakelussa keskeinen ominaisuus on tiedon hyvä jatkokäsiteltävyys asiakkaan omassa tietojärjestelmissä. Jatkokäyttöarvo määräytyy paljolti sen mukaan miten yleiskäyttöisissä ja päätelaiteriippumattomissa tiedostomuodoissa tieto on saatavilla. Päätelaiteriippumattomuutta kehitetään XHTML- ja XML-tekniikkaa käyttämällä. Tieto voi myös olla Tilastokeskuksen toimesta saatettu (sopimusasiakkaalle) täysin valmiiksi sellaiseen muotoon, että asiakkaan tietojärjestelmä pystyy ottamaan sen vastaan sellaisenaan ilman asiakkaan omia jatkokäsittelytoimia.

Tietojen käytettävyys sähköisessä muodossa edellyttää myös sitä, että ennen uuden tilaston julkistamista sitä koskevaa metatietoa toimitetaan asiakkaille riittävästi etukäteen. Uuden tilaston syntyessä tai tilaston uudistuessa tulee sähköisen jakelun asiakkaille toimittaa ajoissa kuvaus uusista/muuttuneista tilaston laadintamenetelmistä sekä uudet/muuttuneet julkaisutason luokitukset tai taulukkomallit. Näin asiakas voi etukäteen alustaa oman tietojärjestelmänsä niin, että varsinaisen datan ilmestyessä sen jatkokäsittely (tai jälleenjakelua harjoittavan tietokantaisännän tapauksessa jatkopakelu), voi alkaa välittömästi tilaston ilmestyttyä. Uuden tai uudistetun tilaston varsinaisena julkistuspäivänä tulee julkaista uutena tietona vain varsinainen data sekä siitä tehdyt yhteenvedot ja tulkinnat. Kansainvälisiä asiakkaita varten tilaston uudistuneet tai muuttuneet metatiedot toimitetaan sähköisen jakelun tietokantoihin (esim. ASTIKA) jakelun valmistelua varten ajoissa myös englanninkielisinä.

Sähköisen jakelun, myös ilmaisen jakelun, tietojen käyttöoikeuksia (*copyright*) koskevat samat säännöt kuin perinteisempien jakelumedioiden kautta jaettavia tietoja. WWW-sivustoihin ja muihin sähköisen palvelun tuotteisiin tulee aina sisällyttää selkeä ja hyvin erottuva Copyright -merkintä, josta tulisi olla linkki sivulle, jossa käyttöoikeuksien yleiset periaatteet ja lähdeviitteiden käyttämisen säännöt käyvät seikkaperäisemmin ilmi.

Huom. Ilmaisestakaan internet-palvelusta haetun tiedon käyttöoikeus ei ilman erillistä lupaa koske muuta kuin tiedonhakua omaan käyttöön.

Yhteensopivuus tuotannossa on yhteensopivuutta myös tiedon käytössä

Tietojen käytettävyyttä edistetään myös yhdenmukaisilla esitystavoilla. Tilastoissa käytettyjen luokitusten harmonisointi on tilastojen sähköisessä jakelussa ehkä vieläkin tärkeämpää kuin painotuotteina jaeltavissa tiedoissa. Jopa samaa luokitusta käytettäessä voi toisistaan poikkeava luokkien esitysjärjestys heikentää tietojen käytettävyyttä tai aiheuttaa tarpeettomasti virheitä, kun eri tuotteista poimittuja tietoja mekaanisesti yhdistellään.

Yhtenäisyyttä voidaan parantaa parhaiten tuotantoprosessien ja tietoaarkkitehtuurin yhtenäistämällä. Yhteistä standardoituun metatietojärjestelmään kytettyä tuotantotietokantaa (Tilastokeskuksen sisäistä tietovarastoa) käyttäen yhdenmukaisuus tuotteissa toteutuu helpoimmin; suurelta osin jopa aivan itsestään.

Asiakkaan kannalta sähköisen tuotteiden mieltämistä edistää se, että sähköisten tuotteiden esitystavat (taulukointi) mahdollisuuksien mukaan vastaavat painettujen tuotteiden vastaavia. Tätä edistetään käyttämällä sekä perinteisen julkaisun että sähköisen tuotteen tuottamiseen samoja prosesseja niin, että perinteisen julkaisun ja sähköisen tuotteen tuotantoprosessit erkanevat mahdollisimman myöhäisessä vaiheessa toisistaan. Se myös takaa paremman tietojen virheettömyyden, kun samoja julkaistavia tietoja ei tarvitse tarkistaa kahdessa erillisissä jakelumieli- tai tuotekohtaisissa tai tuotekohtaisissa tuotantoputkessa, vaan tiedot ovat tarkistettuja ja virheettömiä jo silloin kun ne viedään jakelua ja julkaisutuotantoa palveleviin tietokantoihin, jotka sitten toimivat erilaisten sähköisten ja ei-sähköisten tuotteiden ja julkaisujen yhteisenä lähteenä.

4.3.3 Räätylöödyt tuotteet ja palvelut

Soveltamisala ja päämäärä

Tilastotietojen tarve lähtee asiakkaalta. Hänellä on mielessään joko tutkimus, selvitys- tai seurantarve, jonka tueksi asiakas tarvitsee tilastotietoja.

Toimintaperiaate

Neuvottelut

Vastuu tilastolukujen soveltuvuudesta asiakkaan/toimeksiantajan/tilaajan käyttöön on tilastoasiantuntijalla. Usein neuvottelut aloitetaan yleiskartoituksella, mitä lukuja Tilastokeskus tuottaa. Pääsääntöisesti yksi henkilö ottaa hankkeen/pro-

jektin/toimeksiannon vetovastuun Tilastokeskuksen puolesta asiakkaaseen päin. Hänen tehtäviinsä kuuluu tarvittavien tilastojen selvittäminen. Kunkin tilaston soveltuvuutta asiakkaan tarpeeseen voidaan vielä pohtia yhdessä vastuullisen tilastoasiantuntijan kanssa. Laajoissa hankkeissa on hyvä, että asiantuntijat ovat myös neuvotteluissa paikalla.

Yhdys henkilön tehtävänä on myös arvioida työmäärä ja sopivat tekijät hankkeelle. Mikäli toimeksianto on laaja, olisi hyvä, että työn tekijä olisi alusta saakka mukana neuvotteluissa. Näin varmistetaan toteutuksen onnistuminen ja vältetään väärinkäsitykset.

Neuvotteluissa voi tulla ilmi tietopalvelupyynnö, joka koskee niin sanottuja arkaluontoisia tietoja, kuten kuolemansyy. Tilastolain mukaan Tilastokeskus voi luovuttaa tieteellistä tutkimusta ja tilastollista selvitystä varten tiedot iästä, sukupuolesta, koulutuksesta ja ammatista edellyttäen, että tietojen saajalla on lain mukainen oikeus kerätä näitä tietoja.

Tarjous ja sopimus

Tarjousta varten yhdys henkilön on varmistuttava, että toivotut tilastot, aluejaot ja aikasarjat ovat saatavilla. Yhdys henkilön on myös pystyttävä arvioimaan hankkeen laajuus ja luonne. Kaikki räätälöinnit eivät johda maksulliseen toimeksiantoon, vaan asiakkaan pyynnö voi käynnistää tilaston kehittämishankkeen.

Tietosuoja koskevat säännökset on aina otettava huomioon toimeksiantojen yhteydessä. Samoin on noudatettava tilastoalan eettisiä periaatteita.

Yhdys henkilön antama tarjous asiakkaalle on sitova. Tarjouksessa on mainittava, mitä tehdään, aikataulu ja hinta. Tarjoukseen tehdyt muutokset asiakkaan puolelta tulkitaan tarjouspyynnöksi ja ne vaativat uutta tarjousta. Hyväksytty tarjous johtaa sopimukseen. Aineistojen säilytyksestä sovitaan erikseen ja siitä on olettava maininta sopimuksessa.

Työohjeet

Tilastohenkilö laatii työohjeet atk-henkilölle aina kirjallisesti. Ohjeita laadittaessa tilastohenkilö neuvottelee usein sekä eri ilmiöalueiden edustajien että atk-henkilön kanssa. Tilasto- ja atk-henkilön on syytä käydä ohjeet ja työn vaiheet yhdessä lävitse. Ohjeista tuli käydä ilmi kaikki toimeksiannon laatimisessa tarvittavat tiedot.

Tilastohenkilö selvittää mistä aineistoista toimeksianto tehdään ja atk-henkilö etsii ajan tasalla olevat atk-kuvaukset aineistoille.

Kaikki toimeksiantoon sidoksissa olevat henkilöt ovat vastuussa työn hoitamisesta aikataulussa ja kustannusarvion pitävyydestä.

Suosituks

Tarkistukset

Tarkistuksilla varmistetaan työn tuloksen oikeellisuus ennen kuin työ toimitetaan asiakkaalle. Tarkistamisvastuu kuuluu sekä tilasto- että atk-henkilölle. Työn eri

vaiheissa tehtävillä tarkistuksilla voidaan vähentää turhaa työtä ja tarpeettomia kustannuksia.

Tuoteseloste

Monitilastollisessa toimituksessa asiakkaalle on toimitettava riittävät tiedot käytetyistä lähteistä. Lähdeaineistosta on kerrottava ainakin nimi, mutta tarkempi kuvaus aineistosta on usein paikallaan. Tuoteselosteen voi tehdä joko erillisenä tai liittää tietojen yhteyteen.

Dokumentointi

Kaikki toimeksiantoon liittyvät merkitykselliset asiapaperit säilytetään toimeksiantoa varten perustettuun kansioon. Kirjaamoon toimitetuista asiakirjoista säilytetään kopiot. Tietotekniikkapalvelu arkistoi toimeksiannon yhteydessä syntyneet atk-dokumentit.

4.4 LAATU ASIAKASPALVELUSSA

Soveltamisala ja päämäärä

Tilastokeskuksen perustehtäviin kuuluu tiedon saattaminen käyttöön useita kanavia käyttäen: lehdistötiedotteet, tiedotustilaisuudet, julkaisut sekä elektroniset välineet mukaan lukien Internetin kautta saatavat palvelut. Tilastotietoja toimitetaan asiakkaan pyynnöstä tehtävänä toimeksiantona, jotka saapuvat puhelimitse, telefaksilla tai sähköpostilla. Lisäksi Tilastokirjaston palvelut ovat kaikkien käytävissä.

Asiakaspalvelun yhtenäiset toimintaperiaatteet ja käytännöt ovat tärkeitä, jotta asiakkaat saavat tarvitsemansa tiedot ja palvelun riippumatta siitä, minkä toimintayksikön kanssa he asioivat.

Periaatteet

Asiakaspalvelu on keskeisesti Tilastokeskuksen yrityskuvaan vaikuttava tekijä. Asiakaspalvelussa on huomioitava, että tiedontarvitsijat saavat tarvitsemansa tiedot luotettavasti, ajallaan, ymmärrettävästi, käyttökelpoisessa muodossa, tietosuojan huomioiden ja oikein hinnoiteltuna. Käyttäjän tietotarpeisiin on kiinnitettävä huomiota ja toimitettava tiedot siinä muodossa, että asiakkaat voivat käyttää tietoa työssään kustannustehokkaasti. Asiakaspalautteen kerääminen ja huomioon ottaminen toiminnassa on tärkeää palvelun laadun kannalta.

Tilastokeskuksen palveluperiaatteet

Palveluperiaatteilla luomme yhtenäiset ja hyvät menettelytavat omalle asiakaspalvelullemme. Jokaisen tilastokeskuksen työntekijän tulee tuntea periaatteet ja toimia niiden mukaisesti:

- olemme helposti tavoitettavissa,
- palvelemme monipuolisesti ja tehokkaasti,
- paneudumme asiakkaan tarpeisiin,
- toimimme nopeasti ja täsmällisesti,
- pidämme asiakkaamme ajan tasalla ja
- opimme jatkuvasti asiakkailtamme.

Opastus tilastotiedon käyttöön

Tilastokeskuksen tuotteista ja palveluista on tiedotettava riittävästi sekä opastettava, jotta tiedontarvitsija saa tarvitsemansa tiedot.

Internetissä pidetään yllä:

- Tilasto-opasta,
- Tilastotoria,
- kirjaston Tilda-kokoelmatietokantaa ja
- tietoa verkosta saatavista tilastolähteistä.

Palvelujen saatavuus

Asiakaspalvelupisteissä on päivystys virka-aikana 8.00 – 16.15 maanantaista perjantaihin. Julkaisujen myynti- ja indeksipuhelimet toimivat koko vuorokauden. Internet-palvelut ovat käytettävissä myös vuorokauden ympäri (Tilastotori, asiakaspalautteen antaminen, tietopalvelutoimeksiantojen lähettäminen sähköpostin kautta –asiakaspalvelupisteiden sähköpostiosoitteet tiedoksi käyttäjille). Tilastokirjaston tilat, laitteet ja kokoelmat ovat asiakkaiden käytettävissä virka-aikana. Tilastokeskuksen aluetoimistot palvelevat Seinäjoella, Tampereella, Turussa ja Oulussa.

Neuvonta ja yleistietopalvelu

Tilastokeskus neuvoo asiakkaita tilastotiedon käytössä. Tilastokeskuksesta saa puhelimitse tai sähköpostin kautta muutaman tilastotiedon maksutta. StatFin -palvelusta saatavat keskeiset perustilastot ovat ilmaisia itsepalveluna. Muutoin erilliskustannukset kuten käytetty työaika veloitetaan, mikäli työaika menee yli 15 min.

Tilastokeskuksessa asiakaspalvelu on yli yksikkörajojen tapahtuvaa yhteistyötä. Asiakasta ei siirretä paikasta toiseen, vaan asiakkaan ottaessa yhteyttä puhelimitse otetaan asiakkaan yhteystiedot ja tiedon tarve selville, selvitetään organisaatiossa oikea palvelupiste, joka toimeksiannon voi hoitaa ja otetaan yhteyttä asiakkaaseen tämän jälkeen.

Sähköinen viestintä luo odotuksia nopealle yhteydenotolle. Tilastokeskuksesta otetaan yhteyttä asiakkaaseen vuorokauden sisällä tai seuraavana työpäivänä sähköpostitiedusteluiden osalta.

Toimeksiannot

Asiakkaan toimeksianto käsitellään luottamuksellisesti. Lisäksi asiakastoimeksiannoissa huolehditaan, että tietopalvelussa noudatetaan TK:n lainsäädäntö- ja tietosuojavaatimuksia. Käyttöluvan alaisesta aineistosta tehdään käyttölupa-anomus, jonka tultua hyväksytyksi aineiston saa käyttöönsä.

Suosituks

- Asiakasta kohdellaan ystävällisesti ja pyritään siihen, että asiakas saa yksilöllistä ja asiantuntevaa palvelua.
- Asiakas saa tarvitsemansa tiedot sovittuna toimitusaikana. Viivästymisistä tiedotetaan välittömästi.
- Asiakkaan pyynnöstä toimitettavat tuotteet ja palvelut ovat maksullisia. Palvelun hinnoittelussa noudatetaan voimassa olevaa hinnastoa. Hinnoittelu pohjaa valtion maksuperustelakiin ja -asetukseen sekä valtiovarainministeriön päätökseen Tilastokeskuksen suoritteista perittävistä maksuista. Tilastokeskus toimii markkinoilla hintasopeutujana ei-johtajana.
- Tuotteisiin ja palveluihin sisältyy laatuseloste. Muutoin asiakkaalle toimitetaan tiedot tilastoissa käytetyistä menetelmistä, käsitteistä ja määrittelmistä. Maksullisissa toimeksiannoissa on toimeksiantosopimukset ja toimitusehdot toimitettava asiakkaalle.
- Tuotteiden ja tilastojen osalta pyritään virheettömään palveluun. Toimitettavat tiedot ja tuotteet tarkistetaan ennen niiden toimittamista asiakkaalle.
- Asiakaspalautetta kerätään keskitetysti Anoppi-asiakaspalautejärjestelmällä. Myös henkilöstö tallentaa saamansa palautteen. Reklamaatioihin vastataan ystävällisesti ja nopeasti.
- Asiakashallintajärjestelmään tallennetaan asiakasta koskevat toimenpiteet, kuten tehdyt tarjoukset ja sopimukset tietojen toimittamisesta.
- Tilastokeskuksessa toteutetaan vuosittain asiakastytyväisyystutkimus. Lisäksi voidaan tehdä asiakastytyväisyystutkimuksia myös tuotteisiin/palveluihin liittyen.
- Tilastokeskus toteuttaa julkisten palvelujen laatustrategiaa ja pyrkii saaman tietonsa yhä laajempaan ja monipuolisempaan käyttöön. Tilastokeskus sitoutuu noudattamaan julkistamia palveluperiaatteita, joiden avulla asiakkaat tietävät millaista palvelua he voivat edellyttää saavansa Tilastokeskuksen henkilöstöltä. Palveluperiaatteiden toteutumista seurataan tutkimuksilla ja muun palautteen avulla.

Dokumentointi

- Tietopalvelutoimeksiannoista tehdään toimeksianto/tietopalvelusopimukset.
- Toimeksiantoja tehtäessä työvaiheet kirjataan ja viedään asiakashallintajärjestelmään.
- Kansainvälisten järjestöjen tiedustelut toimitetaan kirjaamon kautta, jonne jää tiedot toimitetuista tiedoista.

Kirjallisuus

- Eurostat. 1998. *Quality Charter of the Data Shop Network*. Versio 2.
- Heino, J., Kaisio, R., Koskenkylä, T., Notkola, V., Orjala, H., Ranta, J., Rantanen, J., & Siljander, A. 2000. *Hinnoitteluopas*. Tilastokeskus.
- Huuskonen, J., Ijäs, N. & Lehtoranta, O. 1997. *Julkisten palvelujen laadun arviointi*. Helsinki: Tilastokeskus.
- Tilastokeskus. 1999. *HETI-työryhmän laatima opas*.
- Tilastokeskus. 1994. *Ohje käyttölupien myöntämisestä Tilastokeskuksen perusaineistoon*. TK-00-854-94. 20.5.1994.
- Tilastokeskus. 2002. *Tietopalvelusopimuksen yleiset sopimusehdot*.
- Tilastokeskus. *Tietopalvelusopimuslomake*.
- Valtion maksuperusteasetus 6.3.1992/211.
- Valtion maksuperustelaki 21.2.1992/150.
- Valtiovarainministeriön päätös Tilastokeskuksen suoritteiden maksullisuudesta 1.11.2001/946.

Hakemisto

A

accessibility; 21, 104
accuracy; 21, 103
administrative records; 24
aggregointi; 32, 82 - 83, 99
Ahvenanmaa; 33, 36
aihealue; 18 - 19, 24, 32, 91
aikasarjat; 24, 71, 74, 78 - 83, 91, 104, 106, 109
aineistotarpeen määrittäminen; 26
ajankohtaisuus; 97
ajantasaisuus; 17, 20 - 21, 34 - 35, 37, 39, 41, 43, 89 - 90, 97, 99, 101, 104 - 107
ajoneuvorekisteri; 39 - 40, 43
alipeitto; 35 - 37, 41, 50, 103
Altika; 104
alueluokitus; 31, 33, 88
ammattieettiset periaatteet; 7, 109
ammattietiikka; 6 - 7
AMS; 11, 89, 92 - 93
analysointi; 27, 60, 68, 70, 78, 89, 97, 100
analysointiaste; 90
analyysipaino; 65
Anoppi-asiakaspalautejärjestelmä; 112
ARGUS ; 10
ARIMA; 79, 81
arkistointi; 27, 29, 32, 91 - 94, 104, 110
arkistonmuodostumissääntö (AMS); 11
arkistonmuodostussuunnitelma (AMS); 92
arviointi; 22, 25, 29, 83, 86, 88
arvoaluetarkistus; 61, 70
arvoskeisyys; 16
asetelmakerroin; 69
asetelmapaino; 64 - 66
asetelmaperusteiset menetelmät; 69
asiaankuuluvuus; 7
asiakaskeskeisyys; 16
asiakaskoulutus; 34
asiakasneuvottelut; 108
asiakaspalvelupisteet; 111

asiakastutkimus; 29
asiakkaan tietotarve; 13, 29, 96, 99, 108, 111

ASTIKA; 104, 106 - 107
asuntokunta; 33 - 34, 37, 41, 44
auditointi; 25, 29
avoimuus; 7, 20, 98

B

balanced scorecard; 14
bootstrap; 70
BSC; 14 - 15

C

CALMAR; 66
CAPAR; 52
CAPI; 49, 52
CASI; 49, 52
CASI of CASIIP; 52
CASI-A; 52
CASI-V; 52
CATI; 49, 52
cati-keskus; 49
census survey; 24
CLAN97; 66, 70
clarity; 21, 104
comparability; 21, 104
consistency; 21, 104
copyright; 108

D

DBM; 52
demografiset luokitukset; 31
derived statistical activity; 24
documentation; 21, 104
dokumentoinnin laajuus; 90
dokumentoinnin tarkkuus; 90
dokumentointi; 14, 20 - 23, 25, 27, 29, 32, 43, 61, 63, 66, 71, 89 - 91, 104, 110, 112

donor imputation; 62

E

Edgeworthin indeksi; 84 - 85
editointi; 26 - 27, 29, 60 - 61, 103
eheys; 21 - 22, 91, 104
EMS; 52
eräkatu; 57, 61, 70
esittäminen; 72
estimaattori; 27, 65 - 66, 68 - 69
estimointi; 27, 35, 62, 64 - 66, 68, 71, 81, 87, 90

F

Fisherin indeksi; 84 - 85
frekvenssitaulukot; 10

H

haastattelijoiden koulutus; 27, 50 - 51, 55, 58
haastattelu; 49
haastattelulomake; 49
haastatteluysikkö; 49, 55
hallinnolliset aineistot; 6, 22, 24, 37 - 39, 41, 43
hallinnolliset rekisterit; 24, 35, 37 - 38
harha; 46 - 47, 58, 69, 71, 85
harhattomuus; 58 - 59, 68, 71
harjaestimaattori; 69, 71
harmonisointi; 32, 105, 108
havainnollisuus; 72, 97
hedoniset indeksimenetelmät; 87
helppokäyttöisyys; 17, 100 - 101
henkilöstön kehittäminen; 13, 22, 89
henkilötunnus; 35, 38 - 39, 43
huoneistotunnus; 39

I

imputointi; 26, 41, 43, 61 - 63, 70, 87 - 88, 103
indeksilaskenta; 86
innovatiivisuus; 13
Internet-kysely; 49, 52
ISI; 7
ISO 3166; 33

ISO 9000; 15

itsenäisyys; 19
itsepainottuva; 64
itsetäytettävä lomakekysely; 52

J

jackknife; 70
jakelu; 72
jatkuva oppiminen; 13
JHS; 31
johdetut tilastolliset aineistot; 24
johtajuus; 13
julkisen hallinnon suositus; 31
julkistaminen; 22, 27, 72, 96 - 97, 107
julkistamiskalenteri; 22, 96
jälkiositus; 65 - 66, 69

K

kaatoryhmä; 31 - 32
kadon adjustointi; 65 - 66
kaksivaiheinen otanta; 35, 46, 70
kalibrointi; 66, 69 - 70
kalibrointiestimaattori; 68 - 70
Kansainvälisen Tilastoinstituutti; 7
kansainväliset käsitteet; 19
kansainväliset luokitukset; 19, 32
kansainväliset standardit; 19, 32 - 33
kansainväliset suositukset; 6, 31 - 32
kansantalouden luokitukset; 31
kantatilasto; 40
89 - 91, 102 - 104, 112
katkaisuosot; 45
kato; 21, 27, 50, 52, 57 - 59, 64 - 66, 69 - 70, 90, 103
katovirhe; 27, 66
katovirheen oikaisu; 27, 65 - 66, 90
kattavuus; 8, 13, 18, 31, 38 - 39, 41 - 43, 90, 101
kaupparekisteri; 40
kausiasoitus; 78 - 79, 81, 104
kausivaihtelu; 78 - 79, 81
kehikkoperusjoukko; 35
kehikkovirhe; 21, 57, 64, 103
kehittäminen; 13, 15, 89, 100

kenttätöyön organisointi; 27, 51, 58
 keskivirhe; 27, 45, 66, 68 - 69, 84, 103
 kiinteistötunnus; 39
 kohdehenkilö; 35
 kohdeperusjoukko; 27, 36 - 37, 58, 90, 103
 kohderyhmä; 34, 100
 kokonaisaineisto; 9, 24, 92
 kokonaislaadunhallinta; 12, 15
 kokonaistutkimus; 24, 34 - 35, 45 - 46, 58, 102 - 103
 koontitiedosto; 32, 92 - 94
 kotitalous; 24, 33 - 34, 36, 46 - 47
 kotitalouskäsite; 33
 kotitalousotos; 37
 kovarianssianalyysi; 70, 87
 kumppanuus; 12 - 13
 kuntien toimintayksikkörekisteri; 38
 kustannukset; 17, 23, 26, 38, 44 - 45, 48, 50, 55, 98, 110 - 111
 kysely; 49
 kyselylomakkeen laadinta; 27
 kyselylomakkeen ohjelmointi; 27
 kyselylomakkeen työohjeet; 27
 käsitietokanta; 92
 käsitteet; 6, 22, 26, 29 - 30, 32, 41, 54, 56,
 käyntihaastattelu; 53
 käytettävyyttä; 21, 42, 63, 100, 105, 107 - 108
 käyttäjien tarpeet; 18, 21, 31, 41 - 42, 89 - 90, 92, 97, 102, 110
 käyttäjän opas; 27
 käyttäjän opastaminen; 43, 59, 63, 68, 89, 101
 käyttäjäryhmät; 17 - 19, 100
 käyttöaste; 100 - 101
 käyttölupa; 9 - 11, 112
 käyttörajoitukset; 91

L

laadun arviointi; 14, 16 - 17, 20, 97, 101 - 102
 laadun parantaminen; 16, 20, 101
 laadun suunnittelu; 16
 laadun varmistus/tarkistus; 16
 laadunhallinta; 12, 15
 laadunhallintajärjestelmä; 13
 laadunmuutos; 86
 laatuindikaattorit; 66
 laatuksiteerit; 29
 laatuselvitys; 20, 27, 66, 102 - 103
 laatustandardit; 16, 20
 Laspeyresin indeksi; 84 - 85
 lehdistötiedotteet; 96, 98, 100, 102, 110
 levykekysely; 52
 linearisointimenetelmä; 70
 lisäinformaatio; 27, 29, 35, 64 - 65, 68, 91, 103
 lomakkeen suunnittelu; 26, 54
 lomakkeen testaus; 25, 58
 looginen imputointi; 61
 loogisuustarkistukset; 27, 51, 60
 luokiteltavuus; 44
 luokitteluperiaatteet; 31
 luokitukset; 6, 10, 26, 29 - 32, 55, 90, 103 - 104, 107 - 108
 luokitusopas; 34
 luokituspalvelut; 34
 luokitusstandardit; 31, 34
 luokitustarkkuus; 9
 luokitustietokanta; 32, 34, 92
 luotettavuus; 6 - 8, 13, 21, 42, 45, 49, 68 - 69, 71, 97 - 98, 102 - 103, 106
 luotettavuuskriteerit; 27
 luottamuksellisuus; 8, 112
 luottamusväli; 68
 luvattu aikaisuus; 21, 90
 läpinäkyvyys; 13, 19 - 22, 90, 104

M

malliluovuttaja-imputointimenetelmä; 62
 Manner-Suomi; 33
 markkinakeskeisyys; 16
 massaimputointi; 61
 menetelmäkuvaus; 66, 103
 metatieto; 14, 25, 28 - 29, 34, 90 - 93, 100, 104, 107 - 108
 mittausvirhe; 57, 64, 90, 103

MLwiN; 70
model-donor imputation; 62
multiple imputation; 62
muutostieto; 38, 42
muutostilasto; 40
muuttujamuunnokset; 29
myönteisyys; 98
määrätaulukot; 10
määritelmät; 6, 22, 30, 32, 55, 89 - 90,
102, 112

N

neuvonta; 111
NUTS; 33
NUTS-aluejako; 33

O

ohjaus; 13, 51, 98
oikea-aikaisuus; 21, 90, 104
oikeellisuus; 21, 37, 60, 90, 101, 103, 109
oikeudelliset yksiköt; 37
operationaalinen prosessi; 25, 27 - 28, 72
opiskelijarekisterit; 40
organisaatiotasot; 14
osallistuva havainnointi; 52
ositettu otanta; 35, 46 - 47, 68 - 69
ositus; 35, 46, 66, 70
otannan suunnittelu; 27
otanta-asetelma; 27, 64 - 66, 68 - 70, 103
otantakehikko; 27, 29, 34 - 37, 45 - 46, 52
otantatutkimus; 22, 24, 27, 34 - 35, 90, 103
otantavirhe; 21, 64, 90
otoksen kohdentaminen; 35
otoksen koko; 27, 45, 48, 65 - 66, 68
otoksen poiminta; 27, 37, 45, 64
otoksen tarkkuus; 45 - 46
otosaineisto; 9, 24, 69 - 71, 92
otoskehikon konstruointi; 27
otosvariassi; 58, 66, 70
otosyksikkö; 34

P

Paaschen indeksi; 84 - 85
painokerroin; 64 - 66
palvelukatalogi; 100

palvelukeskeisyys; 16
paneelitutkimus; 41, 43, 47 - 48, 52
paperilomake; 49
parametriestimaattit; 64, 66
parametrit; 62, 64, 67 - 68, 78, 81
peittävyys; 34, 37, 42
perhe; 41, 44
perherakenne; 34
perhetiedot; 39
perusajankohta; 83 - 87
perusjoukko; 24, 30, 34 - 35, 42, 45 - 47,
57 - 59, 64 - 66, 103
pienalue-estimaattori; 69
piirakkakuviot; 74 - 75, 77
pilottitutkimus; 25, 56
piste-estimaatti; 69, 71
piste-estimaattori; 71
piste-estimointi; 62, 70
pitkittäistutkimus; 48, 51, 61
poikittaistutkimus; 48
poikkileikkausaineisto; 60, 70
poikkileikkaustilanne; 39
poimintaväli; 46
poissulkevuus; 31
postikysely; 49, 52
pps-otanta; 68
PRODCOM-luettelo; 33
promptness; 21, 104
prosessijohtaminen; 13
prosessointivirhe; 90, 103
puhelinhaastattelu; 52
puolueettomuus; 7, 19, 29, 97
puuttuva tieto; 35, 43, 57, 61 - 62, 73
pylväskuviot; 74
pystypylväskuviot; 74 - 76
päiväkirjatutkimus; 52
pääryhmätaso; 31

R

rakennustunnus; 39
real-donor imputation; 62
regressioanalyysi; 68, 70
regressioestimaattori; 68
regressioestimointi; 68 - 69

rekistereiden tietokoneavusteinen yhdistäminen; 52
rekisteritutkimus; 52
rekisteriyksiköiden tunnukset; 39
relevance; 21, 103
relevanssi; 7, 19 - 21, 34, 90, 99, 103
resurssit; 12, 23, 26, 29, 45, 50 - 51, 55, 91
riippumattomuus; 19
rikosnimikkeistö; 31
ryväotanta; 45, 69
ryvästys; 70

S

saatavuus; 16, 20 - 22, 50, 90 - 91, 104 - 105, 111
salassapidettävät tiedot; 8 - 9
salassapitomääräykset; 91
sample survey; 24
SAS; 70, 81
satunnaisotanta; 45 - 46, 65, 69
SEATS; 81
selkeys; 21 - 22, 90 - 91, 104
single imputation; 62
sisäinen tiedotus; 100
sisältymistodennäköisyys; 27, 46 - 47, 64
sisäpiirisäännökset; 22, 96, 106 - 107
SPSS; 81
standardiluokitukset; 31 - 32
standardit; 15, 19, 26, 29, 31, 54, 81, 91
Stat.Fi; 100
Stata; 70
StatFin; 90, 94, 97, 100, 104, 106, 111
statistical survey; 24
SUDAAN; 70
suhdannevaihtelu; 78 - 79
suhde-estimointi; 68 - 69
suhdetehostus; 65
suojausluokka; 11
Suomen Virallinen Tilasto (SVT); 18 - 20, 101
survey; 24
surveylaboratorio; 25, 55 - 56

suunnitteluprosessi; 25 - 26
SVT-julkaisut; 18, 101 - 103
SVT-kriteerit; 19, 101
SVT-laatuksiteerit; 20 - 21, 26, 90, 101 - 102, 105
SVT-sarjajako; 19, 101
SVT-standardi; 101
SVT-tuottajat; 18, 20
SYSRAK; 92
systemityö; 29
systemaattinen otanta; 46
sähköinen päiväkirja; 52

T

taloudellisen toiminnan luokitukset; 31
tarkentuvuus; 68 - 69, 71
tarkkuus; 16, 20 - 21, 34, 45, 68, 72, 85, 90, 103
tasapuolisuus; 97
taulukko; 9 - 10, 72 - 73, 78, 106
taulukkomallit; 107
tautiluokitus; 31
tavoiteperusjoukko; 24, 26 - 27, 35 - 36, 90, 103
teemakartat; 74 - 75
tehokkuus; 6 - 7, 16 - 17, 107
tiedonantorasitus; 38
tiedonantovelvollisuus; 6, 91
tiedonhallinta; 29
tiedonkeruulomake; 25, 51, 54 - 56
tiedonkeruumenetelmän valinta; 26
tiedonkeruutavat; 48 - 49, 102 - 103
tietojenkäsittelymenetelmät; 26
tietojärjestelmä; 91, 107
tietokoneavusteinen henkilöhaastattelu; 49, 52
tietokoneavusteinen lomakekysely; 49, 52
tietokoneavusteinen puhelinhaastattelu; 49 - 50, 52
tietokoneavusteinen videokoodaus; 52
tietokonelomake; 49
tietosisältö; 34, 41, 43, 94, 103, 106
tietosuoja; 6 - 10, 19, 110, 112

tilastoarkisto; 32, 93
tilastografiikka; 72 - 73, 76
tilastolaki; 6, 8, 18, 38, 44, 55, 109
tilastollinen kuntaryhmitys; 33
tilastolliset yksiköt; 37
Tilasto-opas; 100, 111
tilastoprosessitaso; 15
tilastotiedon levittäminen; 27, 99
Tilastotori; 100, 111
tilastotuotantoprosessi; 22
tilastotutkimuksen toimintakehikko; 26
tilastotutkimus; 15, 17, 24, 27, 34 - 36, 45, 49, 68, 102 - 103
tilastoyksikkö; 30
Tilda-kokoelmatietokantaa; 111
timeliness; 21, 104
toimeentulotukea saaneiden rekisteri; 40
toimeksiannot; 108, 110 - 112
toimialaluokitus; 33
toimintakehikko; 25, 30
toimipaikkarekisteri; 38
toimipaikkatunnus; 39
toteutuskeskeisyys; 16
TQM; 12
TRAMO; 81
trendi; 74, 79, 81
tulkinta; 17, 19, 74, 97, 101, 107
tulkittavuus; 22, 89 - 90
tullitariffinimikkeistö; 31
tuloshakuisuus; 13
tunnettavuus; 98, 100 - 101
tunnistetiedot; 8 - 9
tunnusjärjestelmä; 33, 38
tuotantoprosessi; 16 - 17, 89, 91, 93, 98, 104 - 105, 108
tuotantoydinprosessitaso; 14
tuotekeskeisyys; 16
tuotelaatu; 12
tuoteluokitukset; 33
tuoteluokitus; 31
tuoteseloste; 17, 27, 66, 101 - 103, 105, 110
tuottavuus; 16 - 17
tutkimusongelma; 26

tutkimusyksikkö; 34, 36, 45, 48, 50, 57
tutkintorekisteri; 38, 40
työnhakijarekisteri; 40, 43
työohjeet; 22, 27, 29, 32, 50, 55, 109
täsmällisyys; 7, 21, 27, 68, 71, 82, 84, 90
täydellisyys; 22, 34, 89

U

uudelleenpainotus; 65

V

vaakapylväskuviot; 75, 77
variaatiokerroin; 68 - 69
varianssianalyysi; 68, 70
vastaajaluovuttaja-imputointimenetelmä; 62 - 63
vastauskato ks. myös kato; 27, 52, 57, 70
vastausrasite; 19, 38, 47, 55, 58
verotusrekisteri; 39 - 40, 43
vertailtavuus; 72, 89
vertailuajankohta; 83 - 87
vertailukelpoisuus; 6, 20 - 22, 32, 78, 82, 91, 104
viitehenkilö; 42, 44
viivakuvio; 76
viivakuviot; 74
virheettömyys; 16, 64, 66, 98, 108, 112
virhelähteet; 17, 68, 103
väestön keskusrekisteri; 38, 43, 46
väestötietojärjestelmä; 40

W

WesVar; 70

X

X11ARIMA-menetelmä; 79, 81
X11-menetelmä; 79, 81
XHTLM-tekniikka; 107
XML-tekniikka; 107

Y

ydinprosessit; 12, 14 - 15
yhdistelytiedot; 35
yhteiskunnallinen vastuu; 13
yhteiskuntakeskeisyys; 16

yhteisötunnus; 38 - 39
yhteisötunnusrekisteri; 40
yhtenäisyys; 6, 14, 20 - 22, 91, 99, 104,
108, 110
YK:n kansainvälisten tilastosarjojen
pääjako; 18
yksikköarvoindeksi; 84
yksikkökato; 57 - 58, 66, 69 - 70
yksikkökohtaisuus; 41
yksikkötason tiedot; 9 - 11, 38
yksinkertainen satunnaisotanta; 45 - 47, 64
yksiselitteisyys; 22, 31, 35 - 36, 55, 89
yleistettävyys; 24, 34, 45, 48
yleistietopalvelu; 111
ylipeitto; 35 - 37, 42
ympäristötilastoluokitukset; 31
yritys- ja toimipaikkarekisteri; 8, 40
yritysrekisteri; 37 - 39, 41, 43
yritystunnus; 38 - 39

Ä

ääriarvot; 61

Kuvailulehti

Julkaisija

Tilastokeskus

Julkaisun päivämäärä

Toukokuu 2002

Tekijät (toimielimestä: toimielimen nimi, puheenjohtaja, sihteeri)

Johanna Laiho

Leena Hietaniemi

Julkaisun laji

Käsikirja

Toimeksiantaja

Tilastokeskus

Toimielimen asettamispvm

Julkaisun nimi

Laatua tilastoissa

Julkaisun osat

Tiivistelmä

Tilastojen tuotantoprosessi on Tilastokeskuksen toiminnan ydin. Laatua tilastoissa -käsikirjan tavoite on parantaa tilastosysteemien suunnitteluun ja toteuttamiseen liittyvän tietämyksen ja osaamisen käytettävyyttä kokoamalla olemassa olevat periaatteet yhteiseksi osaamispääomaksi.

Laatua tilastoissa -käsikirja on tarkoitettu kaikille tilastojärjestelmien toiminnasta kiinnostuneille, niiden käyttäjille ja tuottajille. Käsikirja kuvaa tilastoalan toimintakehikon Suomen oloissa ja käsittää tilastoalan lainsäädännön lisäksi hyväksi havaittuja käytäntöjä ja suosituksia.

Avainsanat (asiasanat)

Tilastojen laatu, tilastotutkimus, tilastotuotanto, tilastotoimi,

Muut tiedot

Sarjan nimi ja numero

Käsikirjoja 43

ISSN

0355-2063

ISBN

952-467-038-0

Kokonaissivumäärä

124

Kieli

Suomi

Hinta

30 €

Luottamuksellisuus

Julkinen

Jakaja

Tilastokeskus

Työpajakatu 13, Helsinki

00022 TILASTOKESKUS

Kustantaja

Tilastokeskus

Nro 1	Koulutusluokitus 2000 Liite 1, ISCED 1997 SUOMALAINEN SOVELTAMISOPAS 2000 Koulutusluokituksen ja Unescon kansainvälisen koulutusluokituksen ISCED 1997 välinen koodiavain Liite 2, OPETUSHALLINNON KOULUTUSALA- JA KOULUTUSASTELOUKITUS Koulutusluokituksen ja opetushallinnon koulutusala-, opintoala- ja koulutusasteluokituksen välinen koodiavain Liite 3, KOULUTUSKODIMUUTOKSET 2000 Liite 3, KOULUTUSKODIMUUTOKSET 2001 Bilaga 3, ÄNDRINGAR I UTBILDNINGSKODERNA 2001	12. uusittu laitos	2001
Nr 1	<i>Utbildningsklassificering 2000</i>	8:e reviderade upplagan	2001
Nro 4	Toimialaluokitus 1995 Liite 1, HAKEMISTO Liite 2, MUUNNOSAVAIN TOIMIALALUOKITUS 1988:STA Liite 3, TIIVISTELMÄ <i>Näringsgrensindelningen 1995</i> <i>Bilaga 3, SAMMANDRAG</i> Standard Industrial Classification 1995 Annex 3, SUMMARY	2. tarkistettu painos 3. tarkistettu painos 3. tarkistettu painos 2. korjattu painos	1999 1996 1995 1996
Nro 5	Sektoriluokitus 2000 <i>Sektorindelningen 2000</i> Classification of Sectors 2000 Institutionaalinen sektoriluokitus 1996 <i>Den institutionella sektorindelningen 1996</i> Classification of institutional sectors 1996		2000 1995
Nro 6	Rahoitusvaadeluokitus 1996 Classification of financial assets and liabilities 1996		1995
Nro 10	Yhteisöjen tehtäväloukitukset Julkisyhteisöjen ja voittoa tavoittelemattomien yhteisöjen tehtäväloukitukset <i>Uppgiftsklassificeringar för sammanslutningar</i> <i>Uppgiftsklassificeringar för offentliga sammanslutningar och icke vinstsyftande sammanslutningar</i> Classifications of the functions of revised edition government and non-profit institutions serving households	uusittu laitos förnyad upplaga revised edition	1986

Nro 11	Pääasiallisen toiminnan luokitus Pääasiallisen toimeentulolähteen luokitus <i>Klassificering av befolkningen efter huvudsaklig verksamhet</i> <i>Klassificering av befolkningen efter huvudsaklig inkomstkälla</i> Classification of the Population by Type of Activity Classification of the Population by Main Source of Livelihood		1980
Nro 12	Valtiot ja maat 2002 <i>Stater och länder 2002</i> Countries 2002		2002
Nro 14	Ammattiluokitus 2001 Pääjulkaisu		2001
	Ammattiluokitus 2001 Liite 1 Hakemisto		2001
	Ammattiluokitus 2001 Liite 3 Tiivistelmä <i>Yrkesklassificeringen 2001</i> <i>Bilaga 3 Sammandrag</i> Classification of Occupation 2001 Annex 3 Summary		2001
Nro 16	Rakennusluokitus 1994 <i>Byggnadsklassificering 1994</i> Classification of Buildings 1994		1994
Nro 17	Sosioekonomisen aseman luokitus 1989 Classification of Socio-economic Groups	uusittu laitos revised edition	1989
Nr 17	Sosioekonomisk indelning 1989	förnyad upplaga	1990
Nro 18	Demografiset ja sosiaaliset perusluokitukset Ikä, sukupuoli, siviilisääty, kieli, kansalaisuus, uskontokunta <i>Demografiska och sociala grundklassificeringar</i> <i>Ålder, kön, civilstånd, språk, nationalitet, trossamfund</i> Demographic and Social Basic Classifications Age, Sex, Marital Status, Language, Nationality, Religion		1983
Nro 20	Suomen ympäristötiedostot		1996
Nro 21	Aineellisten varojen luokitukset Klinterä pääoma, varastovarot, muut aineelliset varat <i>Klassificeringar av materiella tillgångar</i> <i>Fast kapital, lagerkapital, övriga materiella tillgångar</i> Classifications of Tangible Assets Fixed Capital, Inventories, Other Tangible Assets		1985
Nro 22	Ikäluokitukset Ohjeita ikäluokituksen käytöstä <i>Åldersklassificeringar</i> Riktlinjer för användning av åldersklassificeringar		1986
Nro 27	Rikosnimikkeistö <i>Brottsnomenklatur</i> Crime nomenclature		1999
Nro 28	Kunnat ja kuntapohjalset aluejaot 2002 <i>Kommuner och kommunbaserade indelningar 2002</i> Municipalities and Regional Divisions Based on Municipalities 2002		2002

Nro 30	Toimi oikein tilastoalalla Tilastokeskuksen ammattieettinen opas	1992
Nro 31	Tuottajahintaindeksit 1990=100 Indeksien käyttäjän käsikirja Producer Price Indices 1990=100 Handbook for Indices Users	1993
Nro 32	Maanrakennuskustannusindeksi 1990=100 Käyttäjän käsikirja	1993
Nro 33	Ansiotasoindeksi 1995=100 Käyttäjän käsikirja The Index of Wage and Salary Earnings 1995=100 Handbook for Users	1999
Nro 33	Ansiotasoindeksi 1990=100 Käyttäjän käsikirja The Index of Wage and Salary Earnings 1990=100 Handbook for Users	1994
Nro 34	Tuoteluokitus	1995
Nro 35	Väestölaskenta 2000 Käsikirja	2001
35b	Folkräkningen 2000 Handbok	2001
35c	Population Census 2000 Handbook	2001
Nro 36	Siviiliasian nimikkeistö <i>Nomenklatur för civilmål</i> Nomenclature of civil cases	2002
Nro 37	Jäteluokitusopas	1997
Nro 37b	Guide to Waste Classification	1999
Nro 38	Vuoden 1950 väestölaskennan otosaineiston käsikirja	1997
Nro 39	Kuluttajahintaindeksi 1995=100 Käyttäjän käsikirja <i>Consumer Price Index 1995=100</i> Handbook for Users	1998
Nro 40	Maankäyttöluokitus <i>Markanvändningsklassificering</i> Land Use Classification	2000
Nro 41	Julkisyhteisöjen tehtäväluokitus	2001
Nro 42	Rakennuskustannusindeksi 2000 =100 Käyttäjän käsikirja Building Cost Index 2000=100 User's Handbook	2001
Nro 43	Laatua tilastoissa	2002

Tilastojen tuotantoprosessi on Tilastokeskuksen toiminnan ydin. Laatu tilastoissa -käsikirjan tavoite on parantaa tilastosysteemien suunnitteluun ja toteuttamiseen liittyvään tietämyksen ja osaamisen käytettävyyttä koamalla olemassa olevat periaatteet yhteiseksi osaamispääomaksi.

Laatu tilastoissa -käsikirja on tarkoitettu kaikille tilastojärjestelmien toiminnasta kiinnostuneille, niiden käyttäjille ja tuottajille. Käsikirja kuvaa tilastoalan toimintakehikon Suomen oloissa ja käsittää tilastoalan lainsäädännön lisäksi hyväksi havaittuja käytäntöjä ja suosituksia.

Tilastokeskus, markkinointipalvelut
PL 4C
00022 TILASTOKESKUS
puh. (09) 1734 2011
faksi (09) 1734 2500
myynti.@tilastokeskus.fi
www.tilastokeskus.fi

Statistikcentralen, marknadsföringstjänster
PB 4C
00022 STATISTIKCENTRALEN
tfn (09) 1734 2011
fax (09) 1734 2500
myynti.@stat.fi
www.stat.fi

Statistics Finland, Marketing Services
P.O.Box 4C
FIN-00022 STATISTICS FINLAND
Tel. +358 9 1734 2011
Fax +358 9 1734 2500
myynti.@stat.fi
www.stat.fi

ISSN 0355-2063
= Käsikirjoja
ISBN 952-467-038-0
Tuotenro 9552
CO

