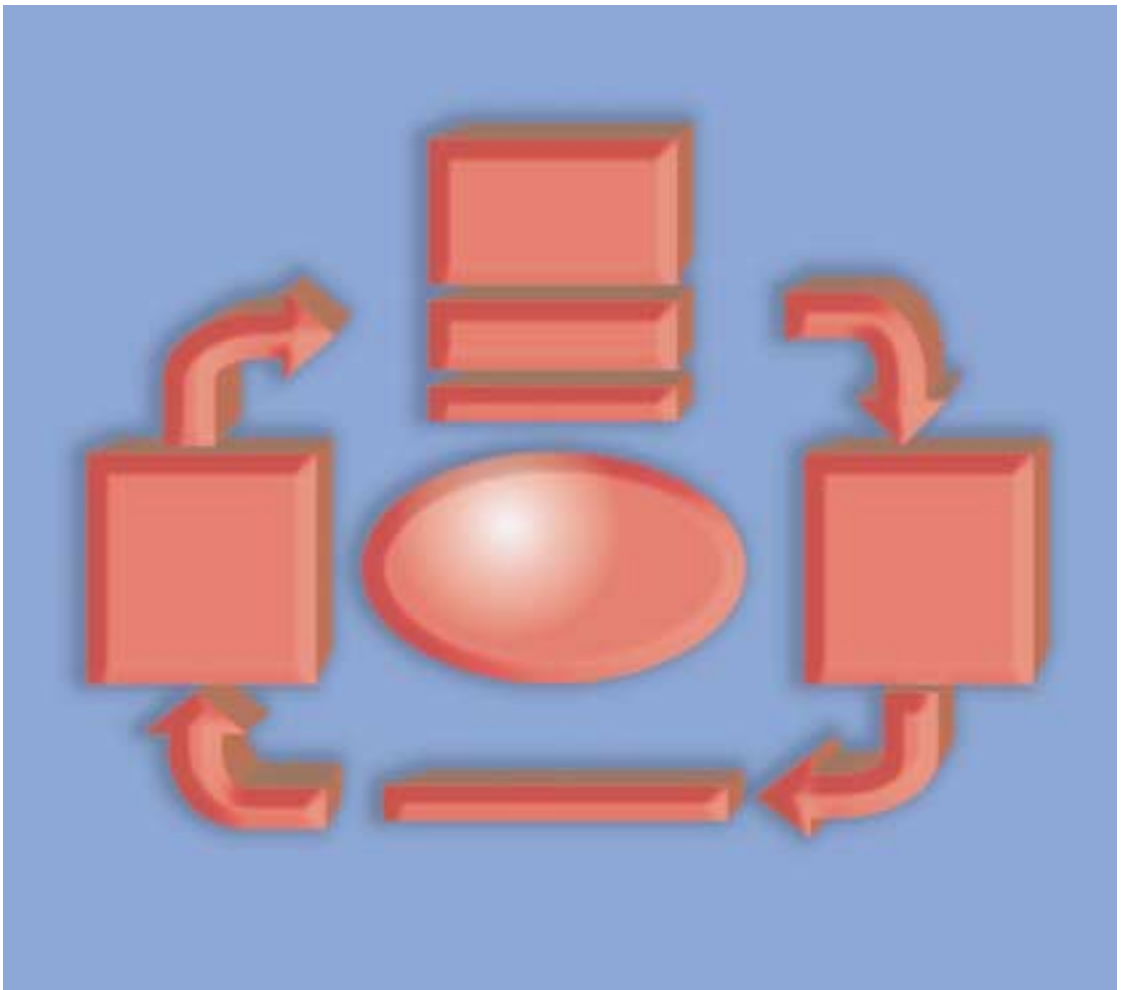


Laatua tilastoissa

2. uudistettu painos



Laatua tilastoissa

2. uudistettu painos

Tiedustelut – Förfrågningar – Inquiries:

*Johanna Laiho
Pentti Pietilä
Kari Djerf
(09) 17 341*

laatua.tilastoissa@tilastokeskus.fi

Kansikuva – Pärm bild – Cover graphics: Johanna Laiho

Taitto – Ombrytning – Layout: Hilikka Potila

© 2007 Tilastokeskus – Statistikcentralen – Statistics Finland

*Tietoja lainattaessa lähteenä on mainittava Tilastokeskus.
Uppgifterna får lånas med uppgivande av Statistikcentralen som källa.
Quoting is encouraged provided Statistics Finland is acknowledged as the source.*

*ISSN 0355–2063
= Käsikirjoja
ISBN 978–952–467–655–7*

Yliopistopaino, Helsinki 2007

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. Lisäksi tilastoammattilaisen tulee pystyä selkeästi ja ymmärrettävästi kertomaan käytetyistä menetelmistä sidosryhmille. Laatu tilastoissa -käsikirja pyrkii tukemaan tilastojen tuotannon kehittämistä ja vuoro-vaikutusta sidosryhmien kanssa. Lisäksi käsikirjaa tullaan hyödyntämään ensimmäisen painoksen tapaan uuden henkilöstön ja asiakkaiden kouluttamisessa.

Tilastojen tuotantoprosessi on toimintamme ydin. Tilastokeskuksessa on runsaasti tuotantoprosessiin sekä tilastojärjestelmien suunnitteluun ja toteuttamiseen liittyvää tietämystä ja osaamista. Käsillä olevan teoksen keskeinen tavoite on 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.

Käsikirjassa on neljä pääasiallista asiakokonaisuutta. Aluksi tarkastellaan yleisesti tilastoalan toimintakehikkoa, tilastotoimintaa säätelevää normistoa ja yleisiä laatuvaatimuksia. Toisessa luvussa perehdytään tilastotuotannon eri vaiheisiin edeten toiminnallisessa järjestyksessä tietojenkeruusta ja tietojen käsittelystä tilastojen, tunnuslukujen ja indikaattoreiden tuottamiseen. Tavoitteena on läpivalaista prosessin kokonaisuutta ja havainnollistaa, 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 tilastojen käytön edistämistä, julkaisemista ja asiakaspalvelua. Luvussa korostetaan tilastotoimen asiakaslähtöisyyden ja asiakaspalvelun jatkuvan kehittämisen tärkeyttä. Käsikirjan käytettävyyttä parantaa loppuun liitetty aakkosellinen asihakemisto.

Laatu tilastoissa -käsikirja on tarkoitettu kaikille tilastojärjestelmistä ja tilastoprosesseista kiinnostuneille, sekä tilastojen tuottajille että niiden käyttäjille. Laatu-käsikirja ei ole koskaan valmis. Käsillä oleva teos on Laatu tilastoissa -käsikirjan toinen uusittu painos. Lisäksi käsikirjasta on julkaistu verkkoversio, jota ylläpidetään ottaen huomioon tilastoalan jatkuvasti muuttuva ympäristö ja kehittyvät tietotarpeet.

Miltei kaikkien asiakokonaisuuksien tekstit on laadittu asiantuntijoiden muodostamissa pienryhmissä. Käsikirja perustuu aiempaan ensimmäiseen painokseen. Uusitun version kirjoittamiseen on osallistunut 22 Tilastokeskuksen asiantuntijaa ja 15 kommentoijaa. Heidän ohellaan kirjan valmistumiseen on vaikuttanut myös moni Tilastokeskuksen työntekijä toimittaen arvokasta lähdemateriaalia ja opastaen kirjoittajia avainhenkilöiden luo. Heille kaikille lämpimät kiitokset asiantuntevasta työstä ja käyttökelpoisista kommenteista. Erityisen suuret kiitokset ansaitsevat professori Carl-Erik Särndal ja professori Risto Lehtonen, joiden jatkuva kannustus ja työn eri vaiheissa esitetyt hyödylliset neuvot olivat ratkaisevan tärkeitä käsikirjan valmistumisen kannalta.

Päävastuun uusitun painoksen toimittamisesta on kantanut kehittäispäällikkö Johanna Laiho. Toimitustyön edistymistä ovat tukeneet toimitusneuvoston muut jäsenet: kehittäispäällikkö Kari Djerf, kehittäispäällikkö Markku Huttunen, professori Risto Lehtonen ja kehittäispäällikkö Pentti Pietilä. Heille kaikille esitän vilpittömät kiitokset asiantuntevasta työstä Laatussa -käsikirjan uusitun, 2. painoksen toteutuksesta.

Helsingissä, tammikuussa 2007

Heli Jeskanen-Sundström
pääjohtaja

Kirjoittajat

Faiz Alsuhail
Kari Djerf
Susanne Hellman-Ketola
Markku Huttunen
Riitta Kaisio
Janika Konnu
Vesa Kuusela
Johanna Laiho
Timo Laukkanen
Risto Lehtonen
Jussi Melkas

Sirkku Mertanen
Heli Mikkela
Perttu Muurimäki
Pekka Myrskylä
Antti Pasanen
Pentti Pietilä
Riitta Poukka
Anna-Leena Reinikainen
Antti Suoperä
Tuula Taivainen
Tuula Viitaharju

Sisällys

Alkusanat	3
1 Yleinen toimintakehikko	6
1.1 Tilastotoimen kehikko	6
1.2 Tilastojen laatukehikko	15
2 Tilastotutkimus ja -tuotanto	28
2.1 Tilastotutkimuksen vaiheet	28
2.2 Tietosuojat	34
2.3 Luokitukset, käsitteet ja määritelmät	41
2.4 Kehikot ja peittävyys	46
2.5 Hallinnolliset aineistot ja rekisterit	49
2.6 Otantamenetelmät	56
2.7 Tiedonkeruutavat	62
2.8 Tiedonkeruulomake: suunnittelu ja testaus	68
2.9 Vastauskato ja siihen vaikuttaminen	71
2.10 Tilastollinen editointi ja imputointi	73
2.11 Painokertoimet ja katovirheen oikaisu	79
2.12 Tilastollinen estimointi ja analyysi	83
2.13 Tilastotiedon esittäminen	88
2.14 Aikasarjat ja kausitasoitus	94
2.15 Indikaattorit ja indeksit	98
3 Dokumentointi ja arkistointi	105
3.1 Tilastotutkimuksen dokumentointi	105
3.2 Tilastotutkimuksen arkistointi	109
4 Tilastojen käytön edistäminen, julkaiseminen ja asiakaspalvelu	112
4.1 Tilastojen käytön edistäminen	112
4.2 Julkaiseminen	114
4.3 Laatu asiakaspalvelussa	118
Hakemisto	121
Käytetyt lyhenteet	126

1 Yleinen toimintakehikko

1.1 Tilastotoimen kehikko

1.1.1 Lainsäädäntö

Tilastolaki (280/2004) on valtion tilastotoimen yleislaki, jota sovelletaan koko valtion tilastotoimeen. Lain tarkoituksena on luoda edellytykset tehokkaalle ja yhtenäiselle tilastotoimelle, joka ottaa huomioon sekä kansalliset että kansainväliset tietotarpeet. Tilastolaki määrittää selkeän ja luottamusta herättävän toimintapolitiikan, jota sovelletaan tietojen antajiin ja käyttäjiin. Laki 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 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.

Kirjallisuus

EU:n lainsäädäntö ja kansainväliset toimintaohjeet

Neuvoston asetusta (EY) nro 322/97 yhteisön tilastoista.

http://forum.europa.eu.int/irc/dsis/bmethods/info/data/new/legislation/stat_law.html

Suomen lait ja asetukset sekä kansalliset toimintaohjeet

Tilastolaki (280/2004).

<http://www.finlex.fi/fi/laki/ajantasa/2004/20040280>

1.1.2 Ammattietiikka

Soveltamisala ja päämäärä

Tilastoalan ammattietiikka on keskeinen perusta tilastoviranomaisten toiminnalle sekä tilastolainsäädännölle ja sen soveltamiselle. 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 oppaansa. Euroopan neuvosto on myös hyväksynyt henkilötietoja koskevan tilastoalan tietosuojasuosituksen No. R (97) 18, joka tietosuojalainsäädännön lisäksi pohjautuu tilastoalan ammattieettisille periaatteille. Myös Euroopan yhteisöjen komissio on vuonna 2005 antanut suosituksen kansallisten ja yhteisön tilastoviranomaisten riippumattomuudesta, koskemattomuudesta ja vastuuvollisuudesta (KOM 217, 2005).

Periaatteet

Tilastontekijän arvopäämäärä on puolueettoman, monipuolisen ja luotettavan kuvan antaminen yhteiskunnasta tietojen kohteiden tietosuoja kunnioittaen. Toimintaympäristön jatkuva ja nopea muuttuminen aiheuttaa uudenlaisia eettisiä ongelmia. Tämän lisäksi vanhat eettiset ongelmat nousevat usein uudelleen keskusteluun. Tilastokeskuksen ammattieettisen oppaan tarkoituksena on helpottaa eettisten ongelmien ratkaisua (Tilastokeskus, 2006).

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

EU:n lainsäädäntö ja kansainväliset toimintaohjeet

Euroopan neuvoston henkilötietoja koskeva suositus. 1997. Council 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.

Euroopan Yhteisöjen Komissio. 2005. *Kansallisten ja yhteisön tilastoviranomaisten riippumattomuudesta, koskemattomuudesta ja vastuuvollisuudesta*. KOM (2005) 217.

http://eur-lex.europa.eu/LexUriServ/site/fi/com/2005/com2005_0217fi01.pdf

International Statistical Institute (ISI). 1985. *Declaration on Professional Ethics*.

<http://isi.cbs.nl/ethics.htm>

Suomen lait ja asetukset sekä kansalliset toimintaohjeet

Tilastolaki (280/2004).

<http://www.finlex.fi/fi/laki/ajantasa/2004/20040280>

Tilastokeskuksen toimintaohjeet

Tilastokeskus. 2006. *Tilastokeskuksen ammattieettinen opas*. Käsikirjoja 30. Helsinki.

http://www.stat.fi/org/etiikka/org_etiikka_etiikkaopas.html

1.1.3 Käytännäsäännöt

Soveltamisala ja päämäärä

Tilastojen laadun arvioinnin objektiivisuus ja vertailukelpoisuus edellyttävät laajaa yhteistyötä sekä yhteisesti hyväksytyjä ja sovellettuja laatukriteereitä. Euroopan tilastojärjestelmässä (*European Statistical System, ESS*) tilastoviranomaisten tuottaminen ja levittämien tilastojen tulee täyttää tiukat laatua ja luotettavuutta koskevat vaatimukset niin kansallisella kuin koko yhteisön tasolla. Tämän vuoksi Komissio on antanut Euroopan tilastojärjestelmälle Käytännäsäännöt (*Code of Practice*), jotka ohjaavat tilastojärjestelmän kehittymistä palvelemaan paremmin tietoyhteiskunnan tarpeita. Euroopan tilastojärjestelmän päämääränä on toimia tehokkaasti ja täyttää riippumattomuutta, koskemattomuutta ja vastuuvollisuutta koskevat vaatimukset.

Käytännäsäännöt ovat luonteeltaan itsesääntelevä väline, sillä ne ovat Euroopan tilastojen tärkeimpien tuottajien eli kansallisten tilastolaitosten valmistelemat ja vahvistamat. Käytännäsäännöt sisältävät 15 periaatetta, joita yhteisön ti-

lastojen tuottamisessa on noudatettava. Käytännessä on kaksi tarkoitusta:

- (i) parantaa luottamusta tilastoviranomaisiin ehdottamalla tiettyjä instituutioihin ja organisaatioon liittyviä järjestelyjä, sekä
 - (ii) vahvistaa tilastoviranomaisten tuottamien ja levittämien tilastojen laatua edistämällä sitä, että kaikki virallisia tilastoja Euroopassa tuottavat tahot noudattavat johdonmukaisesti parhaita kansainvälisiä tilastoperiaatteita, menetelmiä ja käytänteitä.
- (KOM 217, 2005).

Käytännessä perustuvat ESS:n laatujulistukseen (Statistical Programme Committee, 2001):

Euroopan tilastojärjestelmän missio:

"We provide the European Union and the world with high quality information on the economy and society at the European, national and regional levels and make the information available to everyone for decision making purposes, research and debate."

Euroopan tilastojärjestelmän visio:

"The ESS will be a world leader in statistical information services and the most important information provider for the European Union and its member states. Based on scientific principles and methods, the ESS will offer and continuously improve a programme of harmonised European statistics that constitutes an essential basis for democratic processes and progress in society."

ESS:n laatujulistus ohjaa Euroopan tilastojärjestelmää kehittämään kokonaislaadunhallintaa EFQM (*European Foundation for Quality Management*) -mallin mukaisesti (EFQM, 2003). Myös LEG (*Leadership Group*) -laatusuositukset korostavat EFQM -mallin soveltamista (LEG on Quality, 2000). LEG Implementation -työryhmä on ohjeistanut LEG-suositusten käyttöönottoa (*European Communities, 2002*).

Periaatteet

Käytännessä (Code of Practice) ohjaavat tilastojen tuottajia täyttämään Euroopan laatustandardit ja palvelemaan kansallisten tarpeiden lisäksi Euroopan unionin toimielinten, hallitusten, tutkimuslaitosten, elinkeinonharjoittajien ja suuren yleisön tarpeita. Tilastojen laatu määrittyy sen mukaan, missä määrin tilastot ovat relevantteja, tarkkoja ja luotettavia, ajantasaisia, yhtenäisiä ja vertailukelpoisia eri alueiden ja maiden kesken. Lisäksi tilastotietojen helppo saatavuus on oleellista tilastojen laatua arvioitaessa. Laadukkaan tilaston on täytettävä käyttäjien tarpeet kaikkien näiden kriteereiden mukaisesti.

Käytännessä periaatteet on ryhmitelty kolmeen pääkokonaisuuteen, joissa käsitellään institutionaalista ympäristöä, tilastoprosesseja ja tilastotietojen tuottamista (KOM 217, 2005):

I. Institutionaalinen ympäristö

Periaate 1: Ammatillinen riippumattomuus – Euroopan tilastojen uskottavuus varmistetaan tilastoviranomaisten ammatillisella riippumattomuudella politiikasta, sääntelystä tai hallinnosta vastaavista muista yksiköistä ja elimistä sekä yksityisen sektorin toimijoista.

Periaate 2: Tiedonkeruuvaltuudet – Tilastoviranomaisilla on oltavat selvät juridiset valtuudet kerätä Euroopan tilastotarpeisiin tarkoitettuja tietoja. Hallinto, yritykset ja kotitaloudet sekä suuri yleisö voidaan lailla määrätä tilastoviranomaisen pyynnöstä antamaan tietoja yhteisön tilastotarpeisiin.

Periaate 3: Resurssien riittävyys – Tilastoviranomaisten käytettävissä olevien resurssien on oltava sellaiset, että ne riittävät täyttämään Euroopan tilastoja koskevat vaatimukset.

Periaate 4: Laatuun sitoutuminen – Kaikki Euroopan tilastojärjestelmän jäsenet sitoutuvat työskentelemään ja tekemään yhteistyötä Euroopan tilastojärjestelmän laatujuhlituksen periaatteiden mukaisesti.

Periaate 5: Tilastosalaisuus – Tietojen toimittajien (kotitalouksien, yritysten, hallinnon sekä muiden vastaajien) yksityisyys, niiden tuottaman tiedon salassapito ja tiedon käyttäminen ainoastaan tilastotarkoituksiin on taattava ehdottomasti.

Periaate 6: Puolueettomuus ja objektiivisuus – Tilastoviranomaisten on tuotettava ja levitettävä Euroopan tilastoja tieteellistä riippumattomuutta noudattaen objektiivisella, ammattimaisella ja avoimella tavalla, joka kohtelee kaikkia käyttäjiä tasavertaisina.

II. Tilastoprosessit

Periaate 7: Pätevät menetelmät – Korkealaatuisten tilastojen tukena on oltava pätevät menetelmät. Tämä edellyttää riittäviä välineitä, menettelyjä ja asiantuntemusta.

Periaate 8: Asianmukaiset tilastolliset menettelyt – Korkealaatuisten tilastojen tukena käytetään asianmukaisia tilastollisia menettelyjä tiedonkeruusta tiedon validointiin.

Periaate 9: Kohtuullinen vastaustaakka – Vastaustaakan olisi oltava suhteessa käyttäjien tarpeisiin ja säilyttävä kohtuullisena. Tilastoviranomainen seuraa vastaustaakkaa ja asettaa tavoitteita sen vähentämiseksi.

Periaate 10: Kustannustehokkuus – Resursseja on käytettävä tehokkaasti.

III. Tilastotietojen tuottaminen

Periaate 11: Relevanssi – Yhteisön tilastojen on täytettävä käyttäjien tarpeet.

Periaate 12: Tarkkuus ja luotettavuus – Yhteisön tilastojen on kuvattava todellisuutta tarkasti ja luotettavasti.

Periaate 13: Ajantasaisuus ja oikea-aikaisuus – Yhteisön tilastot on levitettävä ajantasaisina ja oikea-aikaisina.

Periaate 14: Yhtenäisyys ja vertailukelpoisuus – Yhteisön tilastojen tulee olla sisäisesti ja ajallisesti yhtenäiset sekä alueiden ja maiden kesken vertailukelpoisia. Eri lähteistä peräisin olevia tietoja tulisi voida yhdistää ja käyttää yhteisesti.

Periaate 15: Saatavuus ja selkeys – Yhteisön tilastot olisi esitettävä selkeässä ja ymmärrettävässä muodossa, levitettävä sopivalla ja tarkoituksenmukaisella tavalla ja niiden sekä niitä tukevan metatiedon ja ohjeiden tulisi olla tasapuolisesti saatavilla.

Suosituksset

Laadukkaan tilastotuotannon takaamiseksi tilastontuottajat ovat vastuullisia seuraamaan säännöllisesti toiminnan ja tavoitteiden toteutumista Käytännesääntöjen mukaisesti. Tilastojen laadun varmistamiseksi tulee tilastotuotantoa kehittää jatkuvasti vastaamaan yhä tiukentuvia ja monimutkaisempia käyttäjien tietotarpeita.

Käytännesäännöissä on määritelty indikaattorit kullekin periaatteelle. Tilastontuottajan tulee pyrkiä noudattamaan niitä ja osallistua kehittämään toimintaympäristöään niiden mukaiseksi.

Tilastollisen laatukriteereiden täyttämiseksi tilastontuottajien on otettava huomioon seuraavat Käytännesäännöissä määritellyt indikaattorit (KOM 217, 2005):

1. Relevanssi:

- Menettelyt käyttäjien kuulemiseksi sekä nykyisten tilastojen relevanssin ja hyödyllisyyden seuraamiseksi käyttäjien kannalta ovat olemassa ja niiden avulla voidaan saada tietoa käyttäjien tulevista tarpeista ja prioriteeteista.
- Työohjelmassa otetaan ensisijaiset tarpeet huomioon ja vastataan niihin.
- Käyttäjien tyytyväisyydestä kerätään tietoa säännöllisesti.

2. Tarkkuus ja luotettavuus:

- Lähdetiedot, väliaikaiset tulokset ja tuotetut tilastotiedot arvioidaan ja validoidaan.
- Otantavirheet ja muut kuin otantavirheet mitataan ja dokumentoidaan järjestelmällisesti Euroopan tilastojärjestelmän laatukomponenttien mukaisesti.
- Tutkimukset ja korjausten analysointi kuuluvat rutineihin ja niistä saatua tietoa hyödynnetään sisäisesti tilastoprosesseissa.

3. Ajantasaisuus ja oikea-aikaisuus:

- Ajantasaisuus täyttää vaativimmat levitystä koskevat yhteisön ja kansainväliset standardit.
- Yhteisön tilastojen julkaisulle on asetettu päivittäinen vakioaika.
- Käyttäjien tarpeet otetaan Euroopan tilastojen julkaisemisen jaksollisuudessa mahdollisimman hyvin huomioon.
- Muutokset tilastojen levitysaikataulussa julkaistaan etukäteen, selitetään ja annetaan uusi julkaisupäivämäärä.
- Alustavia tuloksia hyväksyttävissä olevasta aggregaattilaadusta voidaan julkaista, jos se katsotaan hyödylliseksi.

4. Yhtenäisyys ja vertailukelpoisuus:

- Tilastot ovat sisäisesti yhtenäiset ja johdonmukaiset (esimerkiksi samojen aritmeettisten ja kirjanpitoyksiköiden noudattaminen).
- Tilastot ovat yhtenäisiä tai kohtuullisen ajan kuluessa yhteensovitettavia.

- Tilastotiedot kootaan yhteisten standardien perusteella eri tiedonkeruiden ja niissä käytettyjen lähteiden kattavuutta, määritelmiä, yksiköitä ja luokituksia noudattaen.
- Eri tiedonkeruista ja lähteistä kerättyjä tilastotietoja verrataan ja sovitaan yhteen.
- Tietoaineiston kansalliset rajat ylittävä vertailukelpoisuus varmistetaan vaihtamalla tietoja säännöllisesti Euroopan tilastojärjestelmän ja muiden tilastojärjestelmien välillä. Jäsenvaltiot ja Eurostat tekevät yhteistyössä menetelmä-tutkimuksia.

5. Saatavuus ja selkeys:

- Tilastojen esitysmuoto on sellainen, että se helpottaa tilastojen oikeaa tulkintaa ja merkityksellisten vertailujen tekemistä.
- Julkaisupalveluissa hyödynnetään nykyaikaista tieto- ja viestintäteknikkaa ja käytetään tarpeen mukaan perinteisiä paperijulkaisuja.
- Toteuttamiskelpoisia asiakkaalle räätälöityjä analyysejä tuotetaan ja julkaistaan.
- Mikrotietoihin pääsy voidaan sallia tutkimustarkoituksiin. Pääsyyn sovelletaan tiukkoja protokollia.
- Metatieto dokumentoidaan standardoitujen metatietojärjestelmien mukaisesti.
- Käyttäjille tiedotetaan tilastoprosesseissa käytetyistä menetelmistä ja tuotettujen tilastotietojen laadusta Euroopan tilastojärjestelmän laatukriteerin mukaisesti.

Kirjallisuus

EU:n laindäädäntö ja kansainväliset toimintaohjeet

Euroopan Yhteisöjen Komissio. 2005. *Kansallisten ja yhteisön tilastoviranomaisten riippumattomuudesta, koskemattomuudesta ja vastuuvollisuudesta*. KOM (2005) 217.

http://eur-lex.europa.eu/LexUriServ/site/fi/com/2005/com2005_0217fi01.pdf

Leadership Group (LEG) on Quality. 2000. *LEG on Quality Recommendations*. Eurostat.

Statistical Programme Committee. 2001. *ESS Quality Declaration*.

Muu kirjallisuus

European Foundation for Quality Management (EFQM). 2003. *The EFQM Excellence Model*. Brussels. (Suomen kielinen käännös. Laatukeskus. Excellence Finland).

European Communities. 2002. *Quality in the european statistical system – The way forward*. Eurostat & European Commission. Luxembourg: Office for Official Publications of the European Communities.

Lisätietoja

Eurostat. 2006. The Eurostat Quality site.

<http://ec.europa.eu/eurostat/quality>

1.1.4 Virallinen tilasto

Soveltamisala ja päämäärä

Virallinen tilasto on tilastollinen järjestelmä, johon on koottu keskeinen ja laadukas osa yhteiskuntatilastoja, jotka tuotetaan säännöllisesti tai riittävän usein ja ovat kansallisesti edustavia. Saatavilla olevien tilastojen tulee käytännössä heijastaa käyttäjien tarpeita ja prioriteetteja sekä noudattaa kansainvälisiä tilastojen laadinnalle asetettuja standardeja tai suosituksia. Useissa maissa virallisilla tilastoilla on yleisluontoinen määrittely; tilastojen tulee palvella eri käyttäjäryhmiä: kansalaisia, julkista hallintoa, elinkeinoelämää ja tutkimusta.

Useimmiten virallisen tilaston kokonaisuus määritellään yhteiskunnan keskeisten ilmiö- tai aihealueiden kautta. Aihealuejako voi perustua YK:n kansainvälisten tilastosarjojen pääjakoon: väestö- ja sosiaalitalastot, taloustilastot ja muut tilastot. Näiden alla tilastot ryhmitellään alasarjoihin, aakkoselliseen ryhmittelyyn tai muun vakiintuneen tavan mukaan. Sarjaryhmittelyt on pyritty suunnittelemaan siten, että ne säilyvät suhteellisen vakiintuneina, vaikka sarjojen sisällä tilastotuotanto voi uudistua paljonkin.

Periaatteet

Virallisen tilaston tuottamiseksi tarvitaan "luotettava tiedollinen toimija". Yleensä virallisen tilaston laadinta on viranomaistehtävä, joka on joko määrätty yhden erityisesti tätä tarkoitusta varten perustetun viranomaisen tehtäväksi tai hajautettu eri hallinnonalojen viranomaisille. Virallista tilastoa julkaisevat useimmiten valtion viranomaiset keskeisenä koordinoivana organisaationaan kansallinen tilastovirasto.

Virallisen tilaston laadintaperiaatteena ovat YK:n Tilastokomission 1994 hyväksymät ammattieettiset periaatteet (United Nations Statistical Commission, 1994; ks. myös 1.1.2):

1. Tilastot ovat välttämätön osa demokraattisen yhteiskunnan informaatioperustaa. Tilastoviranomaisten tulee laatia ja saattaa käyttöön viralliset tilastot puolueettomuutta noudattaen.
2. Ammattimaisuus, tieteelliset periaatteet ja ammattietiikka ovat toiminnan perustana.
3. Tilastojen lähteet, laadintamenetelmät ja prosessit perustuvat tieteellisiin standardeihin ja menetelmiin.
4. Tilastojen virheellisiä tulkintoja ja väärinkäyttöä tulee ehkäistä.
5. Tietolähteet valitaan ottaen huomioon laatu, ajantasaisuus, kustannukset ja vastaajien rasitus.
6. Tilastotoimessa on yksikkötietoja koskeva tiukka tietosuoja ja yksikkötietoja käytetään vain tilastollisiin tarkoituksiin.
7. Tilastotoiminta koskeva lainsäädäntö ja muut toimintaperiaatteet ovat julkisia.
8. Kansallisen tilastotoimen koordinointi on välttämätöntä tehokkuuden ja johdonmukaisuuden turvaamiseksi.
9. Kansainvälisesti hyväksytyjen käsitteiden, luokitusten ja menetelmien käyttöä edistetään.
10. Kansainvälinen ja monenkeskinen tilastoyhteistyö auttaa kaikkien maiden tilastotoimen kehittämistä.

Euroopan Yhteisöjen Komission Käytännösäännöt sisältävät itsesääntelevänä välineenä kansallisten ja yhteisön tilastoviranomaisten riippumattomuutta koskevat standardit. Käytännösäännöissä riippumattomuuden periaatetta kehitetään edelleen ottamalla käyttöön ammatillisen riippumattomuuden käsite, jonka yhteydessä erityisesti viitataan myös tilastojen levittämiseen ja saatavuuteen. Ammatillisen riippumattomuuden periaatetta kuvataan seuraavasti:

”Euroopan tilastojen uskottavuus varmistetaan tilastoviranomaisten ammatillisella riippumattomuudella politiikasta, sääntelystä tai hallinnosta vastaavista muista yksiköistä ja elimistä sekä yksityisen sektorin toimijoista. (KOM 217, 2005).”

Indikaattoreita, joilla tämän periaatteen noudattamista seurataan ovat:

- (i) riippumattomuuden osa-alueiden kirjaaminen lakiin,
- (ii) tilastoviranomaisten johtajien asema ja toiminnalliset vastuualueet, mukaan luettuina tilastollisista menetelmistä, standardeista ja menettelyistä tehtävät päätökset,
- (iii) julkaistavien tilastojen sisällöstä ja aikataulusta tehtävät päätökset,
- (iv) tilastoja koskevien työohjelmien julkaiseminen.

Viralliset tilastot kuvaavat taloudellista ja sosiaalista tilannetta sekä muutoksia yhteiskunnan eri sektoreilla. Tilastojen pohjalta tehdään usein kalliita ja kauaskantoisia päätöksiä. Virallisen tilaston luotettavuus on sen takia hyvin oleellista. Tilastojen laatu ja objektiivisuus ovat luottamuksen perusta.

Suosituks

Laadun parantamistoimenpiteet, yleisesti hyväksytyjen laatustandardien ja välineiden käyttö, laatuselvitykset, systemaattiset laatuarviointit ja myös tilaston riittävä dokumentointi sekä sen avoimuus ja läpinäkyvyys ovat osoituksia laatuun sitoutumisesta. Nämä ovat ammattiosaamisen tunnusmerkkejä, ja niistä riippuu pitkälti käyttäjien luottamus virallisia tilastoja tuottaviin organisaatioihin ja niiden palvelutuotteisiin.

- Virallisen tilaston tulee täyttää tietyt laatuvaatimukset, ja tilaston tuottajan olla uskottava, jotta asiakas voi luottaa siihen.
- Tilastontuottajan vastuulla on luotettavien tilastojen tuottamisen lisäksi niiden tehokas jakelu sekä laadun arviointi ja raportointi.
- Laadunraportoinnin tulee noudattaa yleisesti hyväksytyjä laatuindikaattoreita. Euroopan tilastojärjestelmän tilastojen laatuosioita ovat: i) relevanssi, ii) tarkkuus ja luotettavuus, iii) ajantasaisuus ja oikea-aikaisuus, iv) yhtenäisyys ja vertailukelpoisuus, v) saatavuus ja selkeys.
- Käyttäjien luottamuksen aikaansaamiseksi virallisen tilaston tekijöiden ja johdon on osoitettava hyvin voimakkaasti sisäisessä toimintakulttuurissaan sitoutuminen laatuun.

Kirjallisuus

- Euroopan Yhteisöjen Komissio. 2005. *Kansallisten ja yhteisön tilastoviranomaisten riippumattomuudesta, koskemattomuudesta ja vastuuvollisuudesta*. KOM (2005) 217.
- United Nations Statistics Division. 1994. *Fundamental Principles of Official Statistics*. Supplement No. 9. Report of the Statistical Commission on its Special Session. New York 11.–14.4.1994. E/CN.3./1994/18, E/1994/29.

Lisätietoja

- SVT-neuvottelukunta. 2006. *Suomen virallisen tilaston perus- ja laatukriteerit*. (TK-41-1227-06).
- United Nations. 1982. *Directory of International Statistics*. Vol. 1. Statistical Papers. Ser. M. No. 56. Rev 1. New York.
- United Nations Statistics Division. 2001. *Handbook of Statistical Organization. The Operation and Organization of a Statistical Agency*. 3rd ed. ST/ESA/STAT/SER.F/88 (Final Draft).
<http://unstats.un.org/unsd/methods/statorg/stathbook.htm>

1.2 Tilastojen laatukehikko

1.2.1 Kokonaislaadunhallinta

Soveltamisala ja päämäärä

Laatu on perinteisesti Tilastokeskuksessa merkinnyt tilastojen laatua, eli lähinnä tuotelaatua. Tuotelaatunäkemyksen täydennykseksi on määritelty Tilastokeskuksen strateginen kokonaislaadunhallinta (*total quality management, TQM*). 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 tuotantjärjestelmä, 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 suorituskyvylle (EFQM, 2003):

- *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 erottua muista

- *Kumppanuuksien kehittäminen*; painopisteen valinta ja kokonaisuuksien hallintaa vaatii luotettavia kumppanuuksia toimittaja- ja asiakasyhteistyössä
- *Yhteiskunnallinen vastuu*; esimerkillinen yritysuskonallisuus.

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 käänntö parhaiksi tilastometodologisiksi ratkaisuuksi, niiden käytäntöön vienti luomalla tarkoituksenmukaisia rakenteita ja järjestelmiä varmistamaan kustannustehokas keuruu, 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 tähtäävää toimintaa. Se voi olla kokonaan järjestelmää uusivaa, uudelleen suunnittelevaa (korjaavaa tai jatkuvan parantamisen toimintaa).

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

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.

Laadunhallintajärjestelmä pyrkii systematisoimaan ihmisten yhteistyötä ohjaavat toimintatavat kaikkien edun mukaisiksi ja edistämään organisaation asia-

kas- ja sidosryhmäkeskeisten ajattelu- ja toimintamallien parantamista. Järjestelmä edistää viestintää asiantuntijalta toiselle, mikä on olennaista oman organisaation, asiakkaiden ja muiden keskeisten sidosryhmien kannalta.

Laadunhallintajärjestelmällä luodaan edellytykset rakentaa ja vakiinnuttaa organisaatioon suunnitelmallisuutta, hyväksi koettuja toimintamalleja, seurantaa ja jatkuvaa parannusta. Se vähentää viestintäkatkoksia organisaation hierarkiasuunnassa ja horisontaalisessa funktioiden välisessä asiakassuunnassa sekä parantaa keskusteluyhteyttä ja hiljaisen tiedon levittämistä organisaation eri osapuolien välillä.

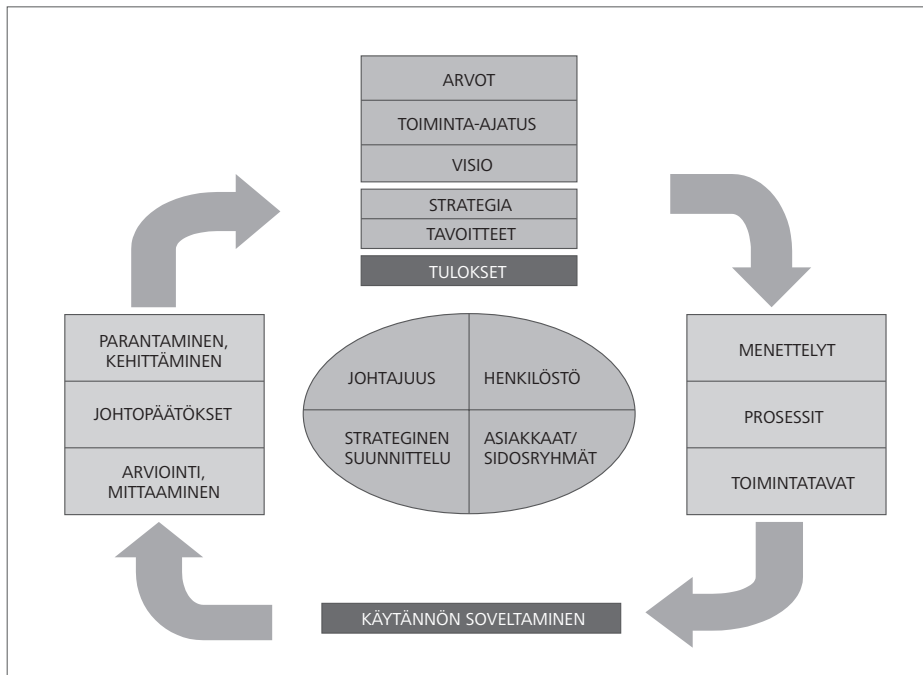
Hyvä laadunhallintajärjestelmä auttaa tunnistamaan, mitä pitää tehdä (oikeita asioita) ja antaa kokemuksen kautta hioutuneita vinkkejä siitä, miten kannattaa tehdä (oikeat asiat oikein). Lisäksi se tuottaa arvokasta palautetietoa, jonka perusteella suorituskyky voidaan edelleen hioa.

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

Organisaatiotaso

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

Kuvio 1.1 Kokonaislaadun hallinta



Tuotantoydinprosessitaso

Tuotantoydinprosessitasolla kehitetään ja ylläpidetään yhtenäistä, kustannustehokasta ja sujuvaa tuotantoprosessia ja siihen liittyvää logistiikkaa yhtenä ydinprosessina. Proessin 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 ja käsikirjat.

Tilastoprosessitaso

Tilastotutkimukset ovat tuotantoydinprosessin alaprosesseja. Tilastoprosessitasolla tutkitaan, toimiiko yksittäisen tilastotutkimuksen tuotantoprosessi optimaalisella tavalla. Yksittäisten prosessien parannukset eivät kuitenkaan pelkäänsä riitä. 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 toiminnan (laadun) hallintajärjestelmä, jossa hyvät laatuperiaatteet on integroitu tilastontuotantojärjestelmiin.
- Merkittävät ja koko organisaatiota koskevat laatutavoitteet sisällytetään strategiin tavoitteisiin, jolloin ne tulevat osaksi toiminnan suunnittelua ja seuranta.
- Keskeisenä kokonaislaadunhallinnan sovellusvälineenä ja mittaristona käytetään tasapainotettua tulokorttia.
- ISO 9000:2000 standardisarja luo puitteet laatuajattelun kehittämiseksi yhteensopivana laatuajattelu- 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 laatuvälineiden hankintaa.

Kirjallisuus

European Foundation for Quality Management (EFQM). 2003. *The EFQM Excellence Model*. Brussels. (Suomenkielinen käännös. Laatuokeskus. Excellence Finland)

Lisätietoja

Anttila, J. & Vakkuri, J. 2001. *Hyvä Parempi Paras*. Tampere: Sonera Oyj / Tammer Paino Oy.

Anttila, J. & Vakkuri, J. 2001. *Luovan johtajan ISO 9000*. Tampere: Sonera Oyj / Tammer Paino Oy.

1.2.2 Tilastojen laadun arviointi

Soveltamisala ja päämäärä

Laadun arviointi tilastotuotannossa perustuu pääosin kansainvälisiin suosituksiin ja sitoumuksiin tilastotoimen laadusta ja sen kehittämisestä. Erityisesti kansainväliset organisaatiot kuten YK, OECD, IMF ja Eurostat ovat asettaneet tavoitteita virallisen tilaston ja kansallisen tilastotoimen laadulle.

Laadun mittaamiseen on sovellettavissa vaihtoehtoisia työkaluja arvioimaan organisaatiota, tilastotuotantoa tai tuotettuja tilastoja eri näkökulmista. Laatuarviointeja voidaan tehdä koko organisaation tasolla, keskittyen yksittäisten tilastojen kehittämiseen tai asiakasnäkökulmaan. Pääosa työkaluista on organisaation johdon ja asiantuntijoiden käytettävissä, mutta osa edellyttää ulkopuolisen arvioinnin hyödyntämistä.

Laadunarviointia hyödynnetään suoraan tilastotuotannon johtamisen tukena ja kehittämistoiminnan tarpeiden paikantamiseen ja tuloksellisuuden arviointiin. Laadunarvioinnin lähtökohdaksi on toiminnan kehittäminen, avoimuus ja yhtenäisten menettelyjen hyödyntäminen. Usean arviointimenetelmän yhteiskäytöllä saadaan erilaisia leikkauspintoja sekä organisaatiosta että tilastotuotannosta. Eri arviointien tulosten keskinäinen vertailu ja analysointi antaa mahdollisuuden arvioida tulosten luotettavuutta ja yhdenmukaisuutta sekä myös näkemysten samankaltaisuutta tai poikkeamia. Laatuarviointien vertailussa havaittu tulosten yhdenmukaisuus tai erillaisuus mahdollistaa yksittäistä arviointia syvällisemmän analyysin organisaation eri osapuolten käsityksistä. Usean laatu työkalun yhteiskäyttö voi antaa kriittisen tärkeää tietoa tilastotuotannon laadun kokonaisvaltaiselle kehittämiselle.

Periaatteet

Useat organisaatiot pyrkivät jatkuvaan kehittämiseen ja tulosparannukseen. Tässä onnistuakseen organisaatio tarvitsee toimivan rakenteen ja resurssien lisäksi jatkuvaa kehitystä tukevaa organisaatiokulttuuria. Tämä käsittää hyvin motivoituneen henkilökunnan, muutoksen johtamista, läpinäkyvyyttä, tehokasta viestintää ja henkilöstön kannustamista parantamaan työympäristöä ja käytäntöjä. Nykyiset ja tulevat saavutukset, kehittämisideat, dialogi ja henkilöstön vahva motivointi kiihdyttävät edelleen organisaation kehittymistä. (Laiho ja Nimmergut, 2004).

Periaatteessa kaikki kehittämistyö perustuu organisaation laatu kulttuurille. Sitoutuminen ja uskallus tuoda uusia kehittämisideoita perustuvat luottamukselle, kunnioitukselle ja uusien ideoiden vastaanotolle. Tilastotuotannon läpivalaisu ja itsearviointityökalut kuten tilastotuotannosta vastaavien DESAP (*Development of Self-Assessment Program*) ja johdon EFQM (*European Foundation for Quality Management*) itsearviointit pyrkivät tuomaan esiin tällaisten kehittämisideoiden syntyä. Ne myös tukevat kehittämistyötä tukevaa keskustelua organisaatiossa. (Laiho ja Nimmergut, 2004).

Laadun arviointia tehdään useimmiten suhteessa standardiin, suosituksiin tai tavoitteisiin. Arviointi voidaan tehdä itse, sisäisesti tai ulkopuolisen voimin tai niiden yhdistelmänä. Arviointi voi perustua sekä laadullisiin että määrällisiin mittareihin ja kuvauksiin. Euroopan kansallisissa tilastovirastoissa ja Tilastokes-

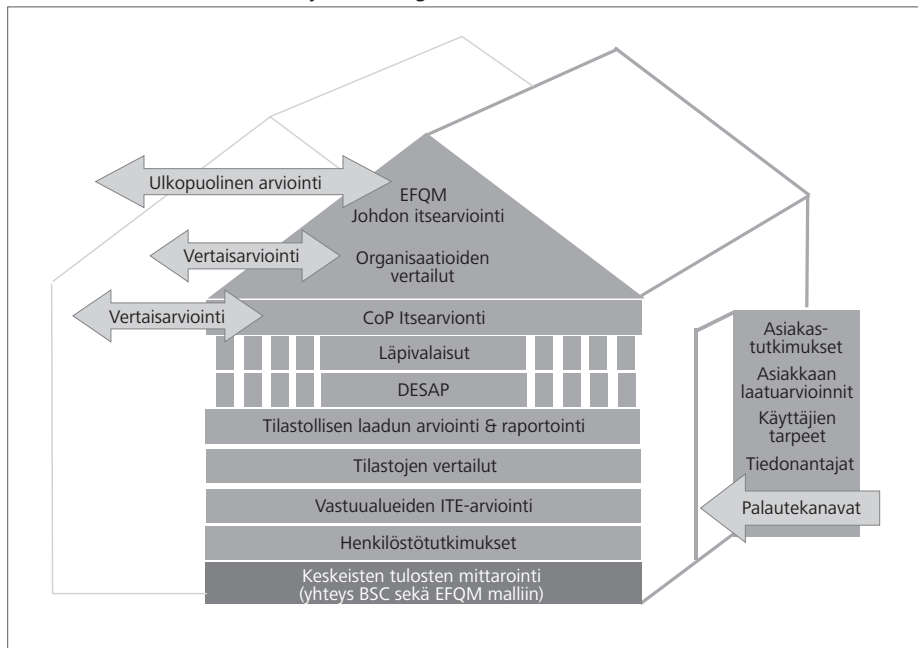
kuksessa käytettyjä tilastotuotantoa tukevia laadunarviointimenetelmiä on esitetty kuviossa 1.2.

Organisaation johdon laadunarvioinneista laajalti käytetty ja standardoitu arviointimenettely on johdon EFQM-itsearviointi, joka noudattaa yhtenäistettyä rakennetta (EFQM, 2003). Sen tavoitteena on paikantaa organisaation vahvuuksia ja kehittämiskohteita. Organisaation johtotasolla voidaan tehdä myös vapaa-
muotoisempia organisaatioiden välisiä vertailuja. Niiden järjestäminen ja painopiste ovat yleisesti sovittavissa vertailtavien organisaatioiden kesken. Lisäksi Euroopan tilastovirastoissa on käynnistetty vuonna 2005 CoP-itsearviointit (Code of Practice) sekä vuonna 2006 Euroopan tilastovirastojen vertaisarviointit, jotka perustuvat Käytännesääntöihin (KOM 217, 2005). Niiden luonne on edellä mainittuja organisaatiovertailuja järjestelmällisempi ja ne noudattavat selkeää rakennetta perustuen Käytännesääntöjen sisältöön (ks. luku 1.1.3).

Tilastotuotannon laatuarviointityökaluiksi ymmärretään perinteisesti niin kutsutut laatuselosteet, jotka raportoivat asiakkaille tilastojen tiedon laadusta. Kokonaislaadun kannalta keskeisempiä laadunarviointimenetelyjä ovat kuitenkin itse laadunostatekijöihin pureutuvat arvioinnit, kuten tilastojen läpivalaisu (Eiderbrandt, 2004) ja tilastotuotantoon räätälöity DESAP-itsearviointi (Körner ja Nimmergut, 2003).

Läpivalaisun periaate perustuu kokonaislaadun hallintaan. Sen kohteena ovat yksittäisten tilastojen tuotanto. Siinä tunnustetaan, että kunkin tilaston parhaat asiantuntijat ovat kyseisessä tilastoyksikössä, mutta hyödynnetään muita tilastoalan asiantuntijoita sisäiseen vertaisarviointiin. Toiminnan tarkoituksena on hyödyntää tilastojen ulkopuolista näkökulmaa tilaston kehitysalueiden tunnistamisessa, niiden

Kuvio 1.2 Tilastotoimen laadun arviointimenetelmät (muokaten Laiho ja Nimmergut, 2004)

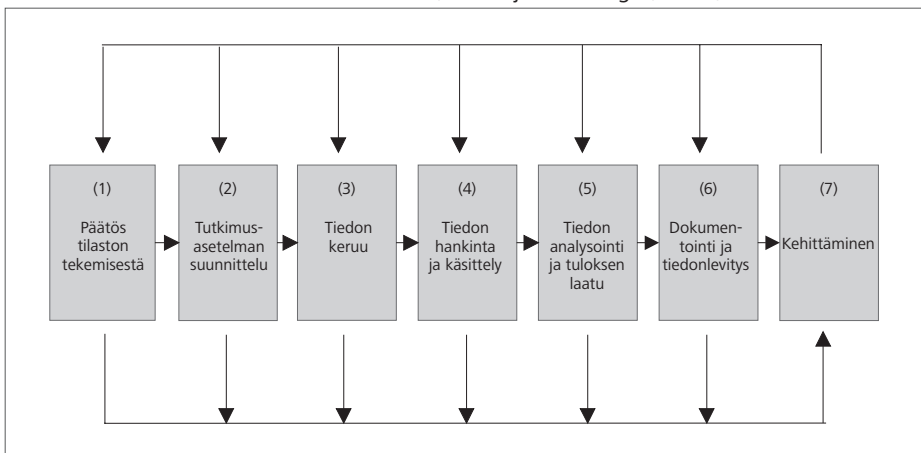


priorisoinnissa ja kehitystoimenpiteiden konkretisoinnissa laadullisten parannusten saavuttamiseksi tilastossa. Arviointi perustuu kirjalliseen arvioitavan toiminnon menettelykuvaukseen sekä itsearviointiin. Näiden avulla selvitetään ovatko toiminnot suunnitelmien mukaiset, toteutetaanko niitä tehokkaasti ja ovatko ne tavoitteiden kannalta tarkoituksenmukaisia. Arvioinnissa selvitetään nykyisten menettelytapojen hyvyttä suhteessa parhaisiin mahdollisiin käytänteisiin. Lisäksi läpivalaisut pyrkivät levittämään hyviä käytäntöjä läpi organisaation.

DESAP-itsearviointi (kuvio 1.3) on myös räätälöity tilastotuotannon tarpeisiin. Toisin kuin muiden asiantuntijoiden avustuksella tehtävä tilastojen läpivalaisu, DESAP tehdään kunkin tilaston vastuuhenkilöiden toimesta. Täten se voidaan toteuttaa riippumatta ulkoisista resursseista ja ajoittaa itsenäisesti ja sopivasti tilastotuotannon aikatauluun kyseisen tilaston laadun kehittämistyön tueksi. DESAP on tilastotuotantoon keskittyvä väline, joka on kehitetty nimenomaan Euroopan tilastojärjestelmään (ESS) kuuluvien tilastojen järjestelmällistä laadunarviointia varten. Täten se poikkeaa yleisistä organisaation kokonaislaadun itsearvioinneista. DESAP kattaa tilastojen laatuun vaikuttavat tuotannon osatekijät. Se on ESS:n laatuvaatimusten mukainen ja sisältää tilastollisen tiedon laatuun liittyvät pääkohdat. Tarkistuslista on siinä mielessä yleinen, että se soveltuu kaikkiin yksittäisiin tilastoihin, joihin kerätään mikroaineistoa – aihepiiristä tai erityisestä tutkimusmenetelmästä riippumatta. Se on tarkoitettu välineeksi tutkimuksesta vastaaville, tilastojen laadun arvioinnin ja parannustoimien suunnittelun avuksi. Vastaavia itsearviointeja on tehty aiemmin myös kansallisesta aloitteesta, esimerkiksi Tilastokeskuksessa 1980–1990 lukujen vaihteessa. DESAP on tuotettu harmonisoiduksi työkaluksi koko Euroopan tilastojärjestelmän kehittämistä ja koordinoimista varten.

Tilastotuotannossa keskeisellä sijalla on erityisesti tiedonantajilta ja asiakkailta saadun palautteen huomioiminen kehittämistyössä. Tiedonantajien vastausrasitteen minimointi, tiedonkeruiden koordinointi ja tiedonkeruunstrumenttien kehittäminen ovat jatkuvan kehittämisen alueita tilastotuotannossa. Tilastojen relevanssin, ymmärrettävyyden ja saatavuuden kannalta on tärkeitä ylläpitää ja kehittää dialogia asiakkaiden kanssa. Asiakasarviointit ja -palautteen analysointi ja aktiivinen kerääminen on keskeisiä laadunarvioinnin menettelyjä ja kehittämi-

Kuvio 1.3 DESAP-itsearviointin rakenne (Körner ja Nimmergut, 2003)



sen työkaluja. Tilastojen laadusta raportoidaan usein käyttäjille laatuselosteissa tai erillisissä menetelmäselvityksissä (luku 1.2.3). Tilastojen laatua voidaan arvioida myös niiden välisin vertailuin, jolloin voidaan puretua erityisesti tilastojen koherenssin ja vertailukelpoisuuden arviointiin.

Organisaation muita laadun arviointimenetelmiä ovat vastuualueiden ITE-arvioinnit (Holma, 2003) ja henkilöstötutkimukset, joiden tuloksia tulee hyödyntää yhdessä kaikkien edellä esitettyjen menettelyjen kanssa. Itsearviointi tarvitsee tuekseen muita arvioinnin muotoja; samanaikaisesti tulee käyttää myös muita arviointimenettelyjä: eri asiakasryhmiltä ja asianosaisilta hankittua palautetta, vertaisarviointeja, benchmarking -toimintaa sekä ulkopuolisen riippumattoman tahon tai tilaajan suorittamaan arviointia (Holma, 2003).

Suositukseset

- Tilastotuotannon laadunarvioinnissa on hyödynnettävä yhtenäisiä menettelyjä ja työkaluja yhtenäisesti läpi organisaation. Laatuarviointien tehokas ja yhtenäinen hyödyntäminen edellyttää jatkuvaa panostusta myös niiden koordinointiin ja ohjaukseen.
- Tilastotuotannon laadunarviointien tarkoituksenmukainen ajoitus tuotantotykliin nähden on kriittistä, jotta niistä saatava hyöty on riittävä ja niihin investoitu aika ei rasita tilastotuotantoa.
- Saatuja tuloksia on hyödynnettävä johtamisessa ja tilastotuotannon kehittämistyössä. Keskeisistä tuloksista ja päätetyistä toimenpiteistä tulee viestiä organisaation sisällä.
- Kaikessa laatuarvioinnissa keskeisenä edellytyksenä on avoimuus. Itsearvioinneissa on määriteltävä ja viestittävä selkeästi luottamuksellisuustasosta ja avoimuusperiaatteen noudattamisesta ennen arviointeja.
- Avoimuusperiaatteen toteutuminen laatuarvioinnissa edellyttää organisaatiokulttuurin jatkuvaa kehittämistä.
- Toiminnan laadun arvioinnissa tulee sisäisten arviointien lisäksi kerätä järjestelmällisesti tietoa tiedonantajilta, asiakkailta ja muilta sidosryhmiltä. Vastavasti tätä palautetta tulee analysoida suhteessa organisaation sisäisten arviointien tuloksiin.

Kirjallisuus

- Eiderbrandt, G. 2004. *Statistical Auditing at Statistics Sweden*. Q2004 conference proceedings. Mainz, Germany.
- European Foundation for Quality Management (EFQM). 2003. *The EFQM Excellence Model*. Brussels. (Suomenkielinen käännös. Laatukeskus. Excellence Finland).
- Holma, T. 2003. *ITE2 -opas uudistuneen itsearviointi- ja laadunhallintamenetelmän käyttöön*. Helsinki: Suomen Kuntaliitto.
- Körner, T. & Nimmergut, A. 2003. *DESAP. Development of a Self Assessment Programme*. Final report. Eurostat & Federal Statistical Office Germany – Institute for Research and Development in Federal Statistics.
- Laiho, J. & Nimmergut, A. 2004. *Using Self Assessments for Data Quality Management – DESAP Experiences*. Q2004 conference proceedings. Mainz, Germany.

Lisätietoja

Eurostat. 2006. The Eurostat Quality site.
<http://ec.europa.eu/eurostat/quality>

1.2.3 Suomen virallinen tilasto ja laatukriteerit

Soveltamisala ja päämäärä

Suomen Virallinen Tilasto (SVT) pyrkii antamaan kattavan kuvan suomalaisesta yhteiskunnasta ja sen tehtävä on palvella yhteiskuntaa tuottaen tilastoinformaatiota, jota yhteiskunta tarvitsee toiminnassaan. SVT-tilastot heijastavat sekä käyttäjien tarpeita että yhteiskunnan yleistä tietotarvetta. Tilastojen aihealueita on 26.

SVT-neuvottelukunta koordinoi virallisen tilaston järjestelmää ja ylläpitää luetteloja viralliseen tilastoon hyväksytyistä tilastoista ja julkaisuista. SVT-tuottajia ovat Tilastokeskus ja 19 muuta julkista viranomaista.

Periaatteet

Suomen virallinen tilasto (SVT) noudattaa YK:n virallisen tilaston periaatteita ja Euroopan Yhteisöjen Komission Käytännesääntöjä (ks. luku 1.1.3). Suomen virallisen tilaston neuvottelukunta on määritellyt viralliseen tilastoon otettujen tilastojen kriteerit.

Suomen viralliseen tilastoon kuuluvan yksittäisen tilaston tulee täyttää seuraavat kriteerit (SVT-neuvottelukunta, 2006):

SVT-peruskriteerit:

SVT-peruskriteeri 1: Tilaston on tuottanut virasto tai laitos, joka kuuluu Tilastokeskuksen ylläpitämään luetteloon SVT-tuottajista.

SVT-peruskriteeri 2: Julkaiseva organisaatio vastaa tilaston sisällöstä ja tietojen oikeellisuudesta taaten näin jatkuvuuden ko. tilaston tuottamiselle. SVT-tilastoa ei voida julkaista henkilön nimellä.

SVT-peruskriteeri 3: SVT-tilastoon liittyy SVT-neuvottelukunnan hyväksymä ajantasainen laatuseloste.

SVT-laatukriteerit:

1. *Relevanssi:* SVT-tilasto sisältää olennaisia ja valtakunnallisesti kattavia tietoja edustamastaan aihealueesta yhteiskunnan tarpeisiin.
2. *Tarkkuus ja luotettavuus:* SVT-tilasto kuvaa täsmällisesti ja luotettavasti tilastoitavaa ilmiötä, sen tilaa ja muutoksia. Tilastointimenetelmistä on esitettävä selkeä kuvaus sekä mahdolliset tilaston epävarmuustekijät on analysoitava ja raportoitava. SVT:ssa mahdollisesti havaittujen virheiden oikaisussa menetellään SVT-neuvottelukunnan hyväksymän suosituksen mukaisesti.
3. *Ajantasaisuus ja oikea-aikaisuus:* SVT-tilastojen tiedot ovat mahdollisimman ajantasaisia ja niiden julkistamisajankohdan pitää olla ennakoon tiedossa ja julkaistuna julkaisijan ylläpitämässä julkistamiskalenterissa, jonka menettelytavat noudattavat SVT-neuvottelukunnan ohjeita.

4. *Yhtenäisyys ja vertailukelpoisuus:* SVT-tilastot ovat keskenään loogisesti konsistentteja sekä mahdollisimman vertailukelpoisia yli ajan ja alueellisesti. Tilastoissa käytetään yleisiä ja vakiintuneita pääasiassa kansainvälisiä käsitteitä ja tilastoluokituksia. Tilaston yhtenäisyyteen ja vertailukelpoisuuteen vaikuttavat tekijät on dokumentoitava ja niistä on tilastossa kerrottava. Määritelmällisesti toisistaan poikkeavat tiedot, jotka kuvaavat samaa ilmiötä, tulee selkeyttää eri käsittein.
5. *Saatavuus ja selkeys:* SVT-tilastot on esitettävä selkeässä, läpinäkyvässä ja ymmärrettävässä muodossa, levitettävä sopivalla ja tarkoituksenmukaisella tavalla ja niiden sekä niitä tukevan metatiedon ja käyttäjille suunnattujen ohjeiden on oltava yleisesti saatavilla.

Laadun arviointi

SVT-tuottajien tulee säännöllisesti arvioida tilastojen laatua yllä esitetyn laatu-kriteerein. Lisäksi tulee arvioida ja seurata asiakkaiden tarpeita ja tyytyväisyyttä sekä tilastotuotantoprosessin tarkoituksenmukaisuutta.

Tilaston laatuseloste ja erillinen menetelmäseloste

Jokaiseen SVT-tilastoon tulee liittää laatuseloste, jossa arvioidaan tiiviisti tilaston laatu, luotettavuus ja sopivuus eri käyttötarkoituksiin. Laatuselosteen keskeinen tarkoitus on vastata, miten SVT-laatuksiteerit täyttyvät kyseisessä tilastossa.

Laatuselosteiden lisäksi kussakin SVT-tilastossa tulee arvioida sisäinen ja ulkoinen tarve erilliseen yksityiskohtaiseen menetelmäselosteeseen ja sen julkaisutiheyteen. Erityisen tärkeä menetelmäseloste on niissä tilastoissa, joiden laadinnassa on ollut merkittäviä menetelmämuutoksia.

Laatu- ja menetelmäselosteiden rakennetta ja eroja on käsitelty taulukossa 1.1. Menetelmäselosteiden rakenteen ja sisällön painotuksen ratkaisevat pitkälti yksittäisen tilaston erityispiirteet ja käyttäjäkunnan tarpeet.

Taulukko 1.1 Ehdotus laatuselosteiden ja menetelmäselosteiden rakenteesta

	Tilaston laatuseloste	Tilaston erillinen menetelmäseloste
1. Tilastotietojen relevanssi (relevance)	<p>1.1 Yhteenveto tilaston 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>1.4 Kuvataan menettelyt käyttäjien kuulemiseksi sekä nykyisen tilastojen relevanssin ja hyödyllisyyden seuraamiseksi.</p>	<p>1.1 Tuotetaan tarkka yhteenveto tilaston tietosisällöstä ja käyttötarkoituksesta. Kerrotaan, mitä ilmiötä kuvaamaan kyseinen tilasto 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 Kuvataan tarkemmalla tasolla menettelyt käyttäjien kuulemiseksi sekä nykyisen tilastojen relevanssin ja hyödyllisyyden seuraamiseksi.</p> <p>1.5 Arvioidaan miten tuotettujen tilastotietojen relevanssi vastaa sille asetettuja vaatimuksia, miten ilmiöalueen mahdolliset muutokset on huomioitu sen tilastoinnissa.</p>

Taulukko 1.1 jatkuu

	Tilaston laatuseloste	Tilaston erillinen menetelmäseloste
2. Tilastotutkimuksen menetelmäkuvaus	<p>2.1 Laatuselosteessa raportoidaan selkeästi tilaston tutkimusmenetelmä eli tilaston perusjoukko, käytetyt perusaineistot, tutkimusasetelma (kokonaistutkimus tai otantatutkimus), (otanta-asetelma), tiedonkeruutapa, estimointimenetelmät ja otantatutkimusten kohdalla painokertoimien käyttö.</p> <p>2.2 Otantavirheiden ja muiden kuin otantavirheiden vaikutus raportoidaan</p>	<p>2.1 Tilastotutkimuksen menetelmäkuvauksessa raportoidaan laatuselosteen kuvausta tarkemmin lopullisten tulosten vaatimat (otanta-,) tiedonkeruu-, editointi-, imputointi-, (painotus-) ja estimointimenetelmät</p> <p>2.2 Otantavirheet ja muut kuin otantavirheet mitataan ja niiden vaikutuksesta arvioidaan ja raportoidaan</p> <p>2.3 Lähdetiedot, väliaikaiset tulokset ja tuotetut tilastot arvioidaan ja validoidaan.</p> <p>2.4 Käytetyt menetelmät ja muutokset niissä perustellaan (myös muutosten vaikutukset aikasarjoille arvioidaan).</p> <p>2.5 Menetelmäkuvauksessa selvitetään tilastotuotannossa käytetyt tietolähteet (myös lisäinformaatiolle).</p> <p>2.6 Kartoitetaan koko tilaston tuotantoprosessi.</p>
3. Tietojen tarkkuus (accuracy) ja luotettavuus	<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>3.4 Otantatutkimuksissa raportoidaan poimittujen yksiköiden kato- tai vastausosuus</p>	<p>3.1 ja 3.2 kohdat laatuselosteesta käsitellään tarvittaessa tarkemmalla tasolla</p> <p>3.3 Esitetään ja tulkitaan estimaateille tilastoinnissa käytetyin keskeisin luokituksin tilastollisia tunnuslukuja, kuten otanta-asetelman huomioivat keskivirheet, keskineliövirheet (MSE) ja otanta-asetelman tehokkuutta estimoivat tunnusluvut (deff).</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>) - vastauskato (<i>non-response</i>). <p>3.5 Arvioidaan tavoiteperusjoukon ja kohdeperusjoukon vastaavuus sekä käytetyn kehikon laatu.</p> <p>3.6 Esitetään kvantitatiivisin indikaattorein vastauskadon, editoinnin ja imputoinnin vaikutus lopulliseen tilastoon.</p>

Taulukko 1.1 jatkuu

	Tilaston laatuseloste	Tilaston erillinen menetelmäseloste
4. Julkaistujen tietojen ajan-tasaisuus (timeliness) ja oikea-aikaisuus (promptness)	<p>4.1 Kerrotaan, mitä ajankohtaa tai -jaksoa julkaistu tilasto kuvaa.</p> <p>4.2a Kerrotaan myös, onko kyseessä ennakkotieto vai lopullinen tieto.</p> <p>4.2b Mikäli tiedot voivat muuttua tulevaisuudessa (esimerkiksi kausitasoituksen vuoksi) tulee muutosepävarmuus esittää selkeästi tiedon käyttäjille.</p> <p>4.2c Mikäli tilasto julkaistaan ensin ennakkollisena, kerrotaan, milloin lopulliset tiedot ovat saatavilla.</p>	<p>4.1 Kerrotaan, milloin tiettyä ajankoh-taa tai -jaksoa koskevat tiedot on julkaistu.</p> <p>4.2 Analysoidaan ennakkotiedon ja lopullisen tiedon mahdollisia poik-keamia.</p> <p>4.3 Tarkastellaan, onko tuotantoaikataulu SVT-suositusten mukaisella tavoitetasolla. Mikäli tilasto ei ilmesty SVT-suositusten mukaisessa tavoit-teajassa tai viivästyy sovitusta aika-tilusta, tulee kertoa tuotanto- ja julkaisuviipeeseen johtaneista syistä.</p> <p>4.4 Tarvittaessa tutkitaan miten ai-kasarjojen tiedot ovat muuttuneet ajassa (esimerkiksi kausitasoituksen vuoksi).</p>
5. Tilastojen yhtenäisyys ja vertailu-kelpoisuus (comparability)	<p>5.1 Kerrotaan tilaston ajallisesta ver-tailtavuudesta sekä vertailtavuu-desta muihin aineistoihin.</p> <p>5.2 Arvioidaan tilaston yhtenäisyyttä ja vertailukelpoisuutta muihin saman aihepiiriin tilastoihin.</p>	<p>5.1 Kerrotaan tilaston ajallisesta ver-tailtavuudesta sekä vertailtavuu-desta muihin aineistoihin laa-tuselostetta tarkemmalla tasolla.</p> <p>5.2 Arvioidaan tilaston yhtenäisyys ja vertailukelpoisuus muihin saman aihepiiriin tilastoihin laatuselostetta tarkemmalla tasolla. Tähän liittyen erot käsitteissä ja tiedonkeruu-prosessissa tulee kartoittaa ja niiden vaikutus tulee arvioida.</p> <p>5.3 Tarkastellaan tilaston yhtenäisyy-teen ja vertailukelpoisuuteen vai-kuttaneita muutoksia ja niiden merkittävyyttä esimerkiksi tilaston tuotantoprosessissa, käsitteissä ja luokituksissa.</p> <p>5.4 Tilastoissa, joiden tiedot muodos-tavat vertailukelpoisia aikasarjoja, ilmoitetaan saatavilla olevien ai-kasarjojen pituudet.</p>
6. Tietojen saatavuus ja selkeys (accessibility and clarity)	<p>6.1 Esitellään tilaston julkaisukanavat ja tietopalvelun yhteystiedot.</p> <p>6.2 Käyttäjää informoidaan tilasto-tuotannon menetelmistä ja tuotet-tujen tietojen laadusta Euroopan ti-lastojärjestelmän ja SVT laatukri-terien mukaisesti.</p> <p>6.3 Kuvataan tilastotutkimuksen meta-tiedon saatavuus ja tietolähde.</p>	<p>6.1 Esitellään tilaston julkaisukanavat (esimerkiksi sähköinen tai painettu-julkaisu, tietokanta ja/tai verkko-palvelu). Lisäksi kuvataan muut tietopalvelumahdollisuudet.</p> <p>6.2 Käyttäjää informoidaan tilasto-tuotannon menetelmistä ja tuotet-tujen tietojen laadusta Euroopan tilastojärjestelmän ja SVT laatu-kriteerien mukaisesti.</p> <p>6.3 Kuvataan tilastotutkimuksen meta-tieto ja viitataan tarkempiin meta-tietotietokantoihin.</p>
7. Dokumentointi (documentation)		<p>7.1 Kuvataan saatavissa olevat arkistoidut tiedostot.</p>

Suosituksset

- Suomen virallisen tilaston (SVT) julkaisuissa on noudatettava SVT-neuvottelukunnan hyväksymiä julkaisuohjeita ja SVT-julkaisuissa tulee olla SVT-standardin mukainen laatuseloste tai viittaus julkiseen verkko-osoitteeseen, jossa laatuseloste on luettavissa. Lisäksi on noudatettava luvussa 1.1.3 esitettyjä Euroopan Yhteisöjen Komission Euroopan tilastojärjestelmälle antamia Käytännönsääntöjä.
- Jokaisesta tilastosta tulee laatia laatuseloste.
- Sisäinen ja ulkoinen tarve erilliseen yksityiskohtaiseen menetelmäselosteseen, niiden sisältöön ja julkaisutiheyteen on erikseen arvioitava.
- Tilaston laatua tulee arvioida käyttämällä ensisijaisesti tilastolle asetettuja kansainvälisiä suosituksia laatuindikaattoreista tai Eurostatin määrittämiä laatuindikaattoreita.
- Laatuselosteessa tulee olla menetelmäkuvaus tilastotutkimuksesta. Tilastollista laatua kuvataan laatuselosteissa seuraavien SVT-laatuksiteerien mukaisesti:
 1. Relevanssi
 2. Tarkkuus ja luotettavuus
 3. Ajantasaisuus ja oikea-aikaisuus
 4. Yhtenäisyys ja vertailukelpoisuus
 5. Saatavuus ja selkeys

Kirjallisuus

EU:n laindäädäntö ja kansainväliset toimintaohjeet

Euroopan Yhteisöjen Komissio. 2005. *Kansallisten ja yhteisön tilastoviranomaisten riippumattomuudesta, koskemattomuudesta ja vastuuvollisuudesta*. KOM (2005) 217.

http://eur-lex.europa.eu/LexUriServ/site/fi/com/2005/com2005_0217fi01.pdf

United Nations Statistics Division. 1994. *Fundamental Principles of Official Statistics*. Supplement No. 9. Report of the Statistical Commission on its Special Session. New York 11.–14.4.1994. E/CN.3./1994/18, E/1994/29.

<http://unstats.un.org/unsd/statcom/doc94/e1994.htm>

Eurostat. 2005. *Standard Quality Indicators*. Doc/ESTAT/02/Quality/9/EN.

Kansalliset toimintaohjeet

SVT-neuvottelukunta. 2006. *Suomen virallisen tilaston perus- ja laatuksiteerit*. (TK-41-1227-06).

SVT-neuvottelukunta. 2004. *Suomen virallisen tilaston julkaisuohje*.

Tilastokeskuksen toimintaohjeet

Tilastokeskus. 2004. *Tilastokeskuksen julkistamisohje*. Tilastokeskus. Sisäinen muistio. Helsinki.

Lisätietoja

Eurostat. 2006. The Eurostat Quality site.

<http://ec.europa.eu/eurostat/quality>

2 Tilastotutkimus ja -tuotanto

2.1 Tilastotutkimuksen vaiheet

Soveltamisala ja päämäärä

Tässä luvussa tarkastellaan tilastotuotannon yleisluonnetta sekä niitä tekijöitä, jotka vaikuttavat keskeisesti tilastojen laatuun. Luvun tavoitteena on antaa lukijalle kokonaiskuva tilastotuotannon ja tilastollisen laadun yhteyksistä nojautuen laajalti käytettyihin ja vakiintuneisiin määritelmiin. Luvun 2 alaluvut heijastavat täten tilastollisen laadun osatekijöitä.

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 kattavasti koko perusjoukon tilaa tarkasteltavan ilmiön suhteen. Yleensä perusjoukko koostuu henkilöistä, kotitalouksista, yrityksistä tai muista taloudellisista ja/tai sosiaalisista yksiköistä. Tilastotutkimukset jaotellaan perinteisesti aihealueittain (ks. luku 1.1.4). Kun tavoitteena on suunnitella, ylläpitää, tehostaa tai parantaa laadullisesti koko tilastotuotantoa ja/tai yksittäistä tilastoa, on hyödyllistä hahmottaa eri tilastojen yhtäläisyydet ja erityispiirteet.

Kansainvälisessä kirjallisuudessa tilastotutkimuksiin viitataan survey-termillä (*statistical survey*), joka käsittää laajasti kaiken empiiris-yhteiskunnallisen tilastotutkimuksen ja -tuotannon. Tarkasti ottaen survey tarkoittaa tilastoja, jotka pyrkivät kattavasti kuvaamaan koko tavoiteperusjoukkoa (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 aineistoista ja/tai rekistereistä (*administrative records*)
- johdetut tilastolliset aineistot (*derived statistical activity*), joissa aineisto on es-timoitu, mallitettu tai muuten johdettu olemassa olevista tietovarastoista. (Statistics Canada, 1998).

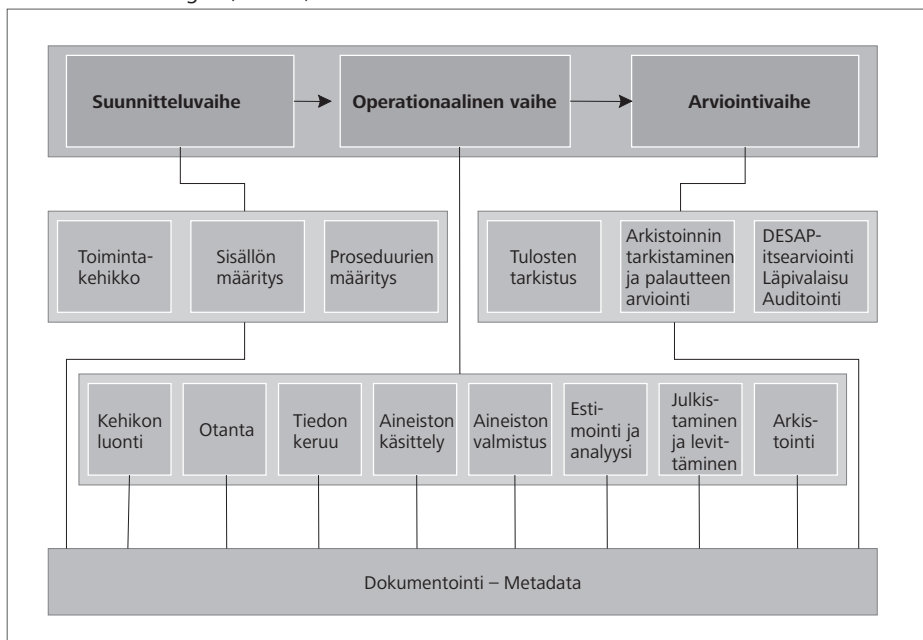
Tilastotutkimus eli survey on siis yläkäsite, joka kattaa myös hallinnollisiin aineistoihin tai rekisteritietoihin perustuvat kokonaistutkimukset. Kokonaisaineistoon perustuvaan tilastotutkimukseen poimitaan jokainen tavoiteperusjoukon yksikkö. Otosperusteisissa tilastotutkimuksissa puolestaan poimitaan osa tavoiteperusjoukkoon kuuluvista yksiköistä pyrittäessä edustavaan ja koko tavoiteperusjoukkoon yleistettävään tilastoon. Tilastotutkimukset kattavat suoran ja epä-suoran tiedonkeruun otoksesta tai koko perusjoukosta sekä näiden yhdistelmät.

Tilastotutkimukset voidaan jakaa toistuviin ja kertaluonteisiin tutkimuksiin. Toistuvat tutkimukset tehdään tietyin määräajoin eli niiden perusteella voidaan laatia myös aikasarjoja ja ennusteita. Tilastotutkimukset voidaan luokitella myös pysyviin, uusiin tai jo aiemmin tehdyn pohjalta kehitettäviin tutkimuksiin. Ensiksi mainitussa tapauksessa tilastotutkimus toistetaan jo käytössä olleella kysymyksenasettelulla. Aivan uuden tilastotutkimuksen suorittamisessa joudutaan suunnitella alusta alkaen tilastotutkimuksen tuotantovaiheet ja kehittää sen tarvitsemat instrumentit. Tiedonkeruulomakkeen testausvaiheessa voidaan apuna käyttää Tilastokeskuksen SurveyLaboratoriota sekä tehdä pilottitutkimuksia (ks. luku 2.8).

Periaatteet

Tilastotutkimuksen suunnittelussa ja tavoitteiden asettelussa tulee huomioida tilastotutkimus kokonaisuutena, joka koostuu kolmesta päävaiheesta: suunnittelusta, operationaalisesta vaiheesta ja arvioinnista (ks. kuvio 2.1).

Kuvio 2.1 Tilastotutkimuksen prosessikokonaisuus (muokaten United Nations, Sundgren, 1999¹)



1) Kuvioon on tehty neljä lisäystä verrattuna alkuperäiseen (United Nations, Sundgren, 1999). Kuvioon on lisätty dokumentointi ja metatiedon karttumisen läpi tilaston tuotannon. Tämän lisäksi suunnitteluvaiheeseen on lisätty 'toimintakehikko' ja arviointivaiheeseen DESAP-itsearviointi, läpivalaisu ja auditointi.

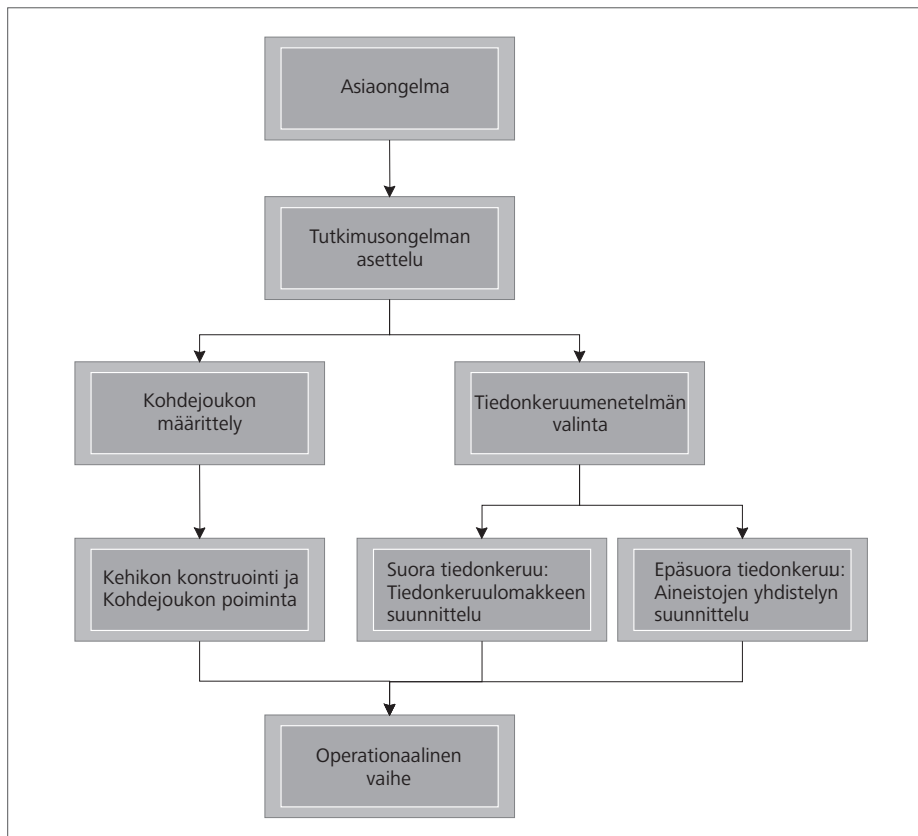
Tilastotutkimuksen suunnittelu

Tilastotutkimuksen suunnittelu voi pohjautua yhteiskunnallisen päätöksenteon, tieteellisen tutkimuksen tai tilastojulkaisujen tarpeisiin. Suunnitteluvaiheessa yksilöidään asiakkaan, tutkijan tai sidosryhmien tietotarve. Tilastolain (280/2004) vaatimukset sekä kansainväliset suositukset ja standardit määrittävät tilastotutki-

muksen toimintakehikon (ks. luvut 1.1). Tilastokeskuksen suunnittelu on esitetty kaaviona kuviossa 2.2. Tilastotutkimuksen ongelmanasettelu tulee olla realistisesti toteutuskelpoinen ja tutkimussuunnitelmassa tulee yksilöidä tavoiteperusjoukko sekä asiasisältö. Suunnitteluvaiheen keskeisiä asioita ovat tilastotutkimuksen sisällön määrittäminen, tiedonkeruun menetelmän valinta ja tiedonkeruun suunnittelu. Suunnittelun alkuvaiheessa selvitetään myös käytettävissä olevat resurssit, kuten budjetti, henkilöstö ja muu välineistö. Samalla määritetään tilastotutkimuksen työvaiheet, tietosuojaus ja aikataulu. Lisäksi pyritään varmistamaan, että tilastojen laatukriteerit ovat täytettävissä (ks. luvut 1.2).

Aineiston tiedonkeruumenetelmän valinta on merkittävä ratkaisu tilastotutkimuksen suunnittelussa. Tiedonkeruumenetelmä vaikuttaa erityisesti kerättävän tiedon luonteeseen, kysymysten asetteluun, kyselylomakkeen suunnitteluun, tilastotutkimuksen läpivientiaikaan ja kustannuksiin (ks. luvut 2.7 ja 2.8). Suunnitteluvaiheessa yksilöidään myös tietojenkäsittelymenetelmät sekä ennakoidaan lopullisen aineiston tuottamiseen tarvittavat tilastolliset menetelmät. Haastattelijatutkimuksissa suunnitteluvaiheeseen kuuluu tärkeänä osana myös haastattelijoiden koulutus sekä kenttätöön organisointi (Tilastokeskus, 1999). Lisäksi otosperusteisissa tilastotutkimuksissa suunnitteluvaiheeseen sisältyy otannan suunnittelu, jossa määritellään otanta-asetelma ja estimointimenetelmät (ks. luku 2.6).

Kuvio 2.2 Tilastotutkimuksen suunnittelu



Suunnittelun ensisijainen päämäärä on edesauttaa saavuttamaan mahdollisimman tarkka, täsmällinen ja ajantasainen tutkimustulos. Suunnitteluvaiheessa tulee pyrkiä ennakoimaan mahdollisesti esiintyvät ongelmat. Tilastotutkimuksissa tulee varautua puuttuvaan tietoon ja sen vaatimiin korjausmenetelmiin. Tilastotutkimuksen suunnittelun dokumentointi nopeuttaa laatuselosteiden ja käyttäjän oppaiden laadintaa ja julkaisemista.

Operationaalinen vaihe

Tilastotutkimuksen operationaalinen vaihe on esitetty työvaiheineen kuviossa 2.3). Nämä työvaiheet vaikuttavat oleellisesti lopputuloksen tilastolliseen laatuun:

- otoskehikon konstruointi ja tarvittavan lisäinformaation yhdistäminen (ks. luku 2.4)
- otoksen poiminta (ks. luku 2.6)
- tiedonkeruulomakkeen laadinta / ohjelmointi ja ohjeiden valmistaminen (ks. luku 2.8)
- tiedonkeruu ja sen koordinointi (ks. luku 2.7)
- aineiston muokkaus (tiedon koonti, koodaus, (loogisuus) tarkistus ja editointi) (ks. luku 2.10)
- lopullisen aineiston tuottaminen (ks. luvut 2.10, 2.11 ja 2.12)
- estimointi ja analysointi (ks. luku 2.12)
- lopullisten tulosten julkistaminen (ks. luku 4.2)
- arkistointi (ks. luku 3.2).

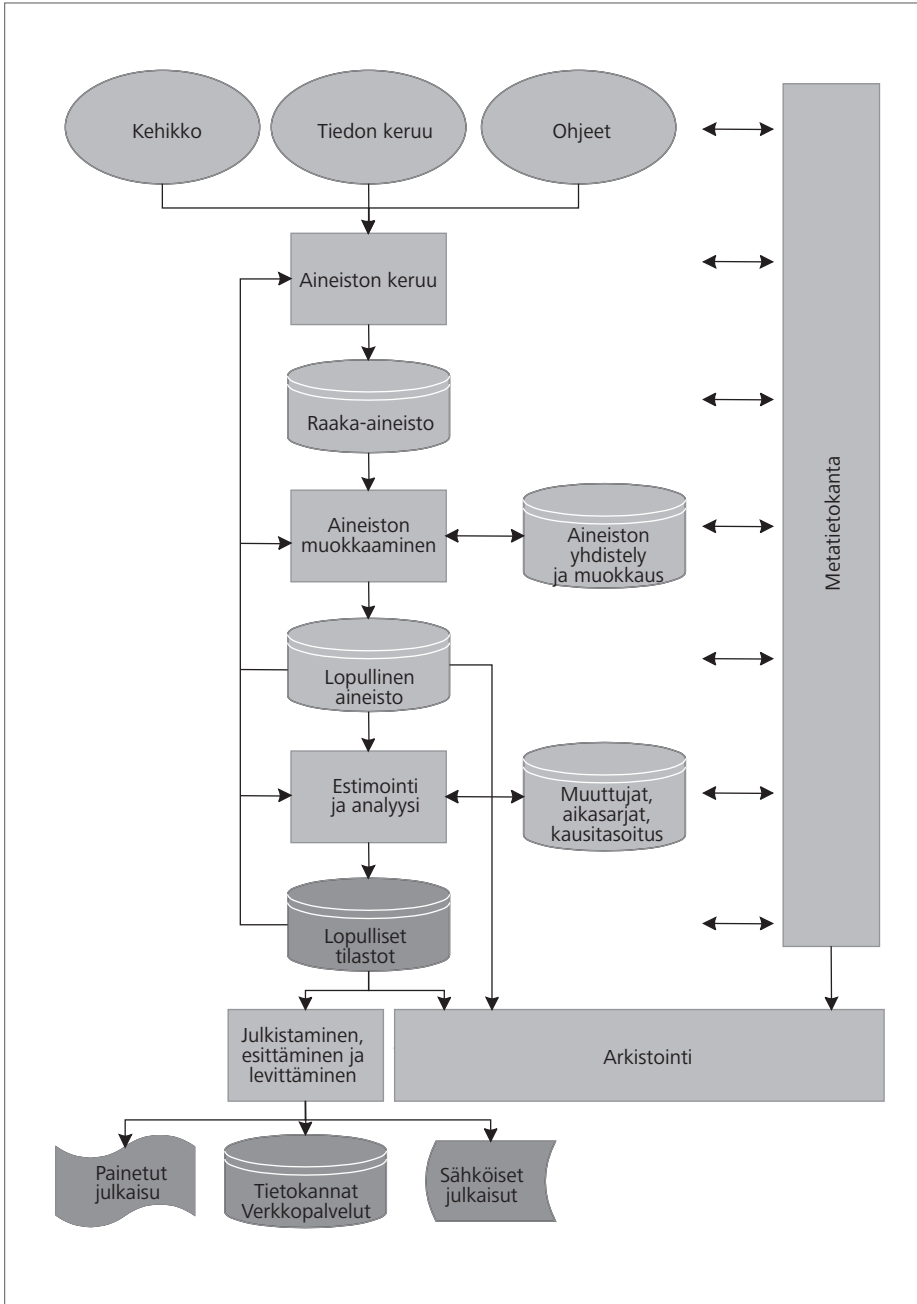
Tiedonhallinta ja systeemyö ovat tärkeä osa tilastotutkimuksen jokaista vaihetta. Kuviossa 2.3 on kuvattu miten tilastotutkimuksen dokumentointi alkaa ja metatietoa syntyy jo tilastotutkimuksen suunnitteluvaiheessa. Vastaavasti dokumentointi ja säilytetty metatieto karttavat läpi koko prosessin. Metatiedolla tarkoitetaan esimerkiksi käytettyjä määritelmiä, luokituksia, standardeja, otantakehikkoa, käytettyä lisäinformaatiota, kyselylomaketta, työohjeita, vastaajien ja haastattelijoiden kommentteja, tehtyjä editointeja ja muuttujamuunnoksia.

Tilastotutkimuksen arviointi

Tilastotutkimuksen arviointivaiheessa tarkastellaan, miten suunnitteluvaiheessa määritellyt lopputuotteet on tuotettu, julkaistu ja miten asetetut laatuksiteerit on saavutettu. Lisäksi tarkastetaan ja arvioidaan metatiedon dokumentointi: aineistojen, luokitusten ja käsitteiden arkistointi sekä asiakastutkimukset. (Muokaten United Nations, Sundgren, 1999).

Koko tilastotutkimuksen suoritusta ja laatuksiteerien täyttämistä voidaan tarkastella lisäksi tilastonvastuuhenkilöiden DESAP-itsearviointinilla, organisaation sisäisillä tilastotuotannon läpivalaisuilla sekä auditoinneilla, jossa tilastoammattilaiset ja puolueettomat asiantuntijat arvioivat tilastotutkimuksen kokonaisprosessia ja pyrkivät selvittämään parannusmahdollisuuksia (ks. 1.2.2).

Kuvio 2.3 Tilastotutkimuksen operationaaliset työvaiheet (muokaten, United Nations, Sundgren, 1999¹)



1) Kuvioon (United Nations, Sundgren, 1999) on muutettu metatiedon yhdistämistä käsittelevä kohta siten, että metatietoa kerätään ja yhdistetään koko prosessin aikana. Lisäksi alkuperäiseen kuvioon on lisätty lopullisen aineiston, lopullisten tilastojen ja metatietojen 'arkistointi'.

Suosituksset

Tilastotutkimuksen korkeatasoinen laatu voidaan saavuttaa:

- resursoimalla jokainen tilastotutkimuksen vaihe oikein,
- panostamalla suunnitteluvaiheeseen,
- asiakastarpeen huolellisella analysoinnilla,
- järjestelmällisellä systeemityöllä,
- panostamalla yksittäisten toimintojen ja kokonaisuuden vuorovaikutukseen,
- dokumentoimalla yksittäiset työvaiheet siten, että ne ovat toistettavissa toisen tilastoammattilaisen toimesta (ks. 3.1),
- hyväksymällä tilastotutkimus valmiiksi vasta, kun se on arkistoitu (ks. 3.2),
- arvioimalla tilastotutkimuksen tuotos,
- tekemällä yhteistyötä muiden tilastotutkimusten tekijöiden kanssa ja jakamalla hyviä käytäntöjä ja tietämystä.

Kirjallisuus

EU:n laindäädäntö ja kansainväliset toimintaohjeet

United Nations Statistical Commission & Economic Commission for Europe. 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.

http://www.unece.org/stats/documents/information_systems_architecture/1.e.pdf

Suomen lait ja asetukset sekä kansalliset toimintaohjeet

Tilastolaki (280/2004).

<http://www.finlex.fi/fi/laki/ajantasa/2004/20040280>

Tilastokeskuksen toimintaohjeet

Tilastokeskus. 1999. *Tilastohaastattelijan opas*. Tilastokeskus. Sisäinen muistio. Helsinki.

Muiden tilastovirastojen toimintaohjeet

Statistics Canada. 2003. *Statistics Canada Quality Guidelines*. Ottawa.

<http://www.statcan.ca/english/freepub/12-539-XIE/free.htm>

Sundgren, B. 2001. *Documentation and Quality in Official Statistics*. Conference CD. Stockholm 14.–15.5.2001. Statistics Sweden & Eurostat.

Muu kirjallisuus

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

Dodge, Y., Cox, D., Commenges, D., Davison, A., Solomon, P., Wilson, S. 2003. *The Oxford Dictionary of Statistical Terms*. 6th ed. Oxford, New York: Oxford University Press.

2.2 Tietosuoja

Soveltamisala ja päämäärä

Tietosuojalla tarkoitetaan lakeja, määräyksiä, menetelmiä ja näiden soveltamista, joiden kaikkien tavoitteena on turvata tiedonantajien yksityisyys sekä liike- ja ammattisalaisuus. Tietosuoja otetaan huomioon jo tiedonkeruuvaiheessa ja sen suunnittelussa huolehtimalla, ettei tilaston laatimisen kannalta tarpeettomia tietoja edes kerätä. Tietoja käsitellään tietosuojaperiaatteita noudattaen eli salassapidettävien tietojen paljastuminen ulkopuolisille estetään tietojen käsittelyn kaikissa vaiheissa.

Tilastotoimessa tietosuoja on toiminnan keskeinen peruseriaate. Tietosuojaperiaatteiden tarkoituksena on ylläpitää luottamuksellisia suhteita tiedonantajiin. Näin varmistetaan kattavien ja luotettavien tietojen saaminen tilastojen laadintaa varten. Tietosuojaperiaatteet rajaavat sen, mitä tietoja voidaan kerätä, miten tietoja voidaan käsitellä sekä missä muodossa ne voidaan julkaista. Tietosuoja vaikuttaa siten tilastojen laatimisessa sovellettaviin menettelytapoihin ja julkaistavien tilastojen sisältöön.

Tilastotointa koskevat tietosuojasäännökset sisältyvät ensisijaisesti tilastolakiin (280/2004) ja neuvoston asetukseen (EY) 322/97 yhteisön tilastoista, mutta myös henkilötietolaki (523/1999) ja laki viranomaisen toiminnan julkisuudesta asetuksineen (621/1999 ja 1030/1999) sisältävät tietosuojaa koskevia säännöksiä. Lakien ja asetusten lisäksi tietosuojaan liittyviä ohjeita on antanut Valtionhallinnon tietoturvallisuuden johtoryhmä (VAHTI). VAHTI:n ohjeista tilastotointa koskevat lähinnä Valtionhallinnon tietoaineistojen käsittelyn tietoturvallisuusohje (Valtiovarainministeriö, 2000) sekä Arkaluonteiset kansainväliset tietoaineistot (Valtiovarainministeriö, 2002).

Tilastokeskuksessa on useita tietosuojaa koskevia sisäisiä ohjeita, kuten esimerkiksi tilastolain ja henkilötietolain soveltamisohjeet (Tilastokeskus, 2005a ja 2000a) ja ohjeet taulukkomuotoisten yritystietojen ja henkilötietojen suojaamiselle (Tilastokeskus, 2002a ja 2000b). Tietosuojaan liittyviä yleisiä toimintaperiaatteita antaa tietosuojaohje (Tilastokeskus, 2005b) sekä Tilastokeskuksen ammattieettinen opas (Tilastokeskus, 2006), joka pohjautuu EU:ssa, YK:ssa ja Kansainvälisessä Tilastoinstituutissa (ISI) hyväksytyille tilastoeettisille periaatteille. Näistä kansainvälisistä ohjeista erityisesti ISI:n ammattieettinen julistus (ISI, 1985), YK:n tilastokomission hyväksymät virallisen tilaston peruseriaatteet (YK, 1994) ja Euroopan yhteisöjen komission Käytännäsäännöt (KOM 217, 2005) ovat olleet tärkeimpinä lähtökohtina ammattieettistä opasta kirjoitettaessa.

Periaatteet

Tietosuoja vaatimusten huomioon ottaminen edellyttää tilastotutkimuksen huolellista suunnittelua etukäteen tietojen keruusta niiden julkaisemiseen. Tiedonantajilta tulee kerätä vain tutkimuksen kannalta välttämättömät tiedot. Otantaa ja muita tietojen keruuta keventäviä menetelmiä on käytettävä tiedonantajien vastausrasitteen minimoimiseksi aina, kun se on mahdollista. Kerättyjen

tietojen käsittely on toteutettava niin, ettei kenenkään yksityisyyttä tai liike- ja ammatillisalaisuutta vaaranneta. Tilastoja laativan viranomaisen on huolehdittava siitä, että tiedot on asianmukaisesti suojattu luvatonta käsittelyä ja paljastumista vastaan. Muiden kuin oman tilastoyksikön yksikkötason tietojen käyttö on myös Tilastokeskuksen sisällä luvanvaraista.

Tilastointitarkoituksiin saadut tiedot ovat pääsääntöisesti salassapidettäviä. Julkisia ovat valtion ja kunnallisten viranomaisten toimintaa ja julkisten palvelujen tuottamista kuvaavat tiedot sekä yritys- ja toimipaikkarekisteriin sisältyvät tiedot. Salassapidettäviä tietoja voidaan luovuttaa ainoastaan tieteellisiin tutkimuksiin tai tilastollisiin selvityksiin ja silloinkin pääsääntöisesti muodossa, josta yksiköitä ei voida suoraan tai välillisesti tunnistaa. Tiedot henkilöiden iästä, sukupuolesta, koulutuksesta ja ammatista voidaan poikkeuksellisesti luovuttaa tieteellisiin tutkimuksiin ja tilastollisiin selvityksiin tunnistetietoineen. Tunnistetietoineen tietoja voidaan luovuttaa myös toisille tilastolain määrittelemille tilastoviranomaisille niiden toimialaan kuuluvien tilastojen laatimista varten. Tietojen luovuttaminen käytettäväksi hallinnollisessa päätöksenteossa tai muussa vastaavassa asian käsittelyssä on erityisesti kielletty. Tietojen luovuttamisen menettelytavat ja periaatteet on kirjattu pääjohtajan vahvistamaan ohjeeseen käyttö lupien myöntämisestä Tilastokeskuksen yksikkötason aineistoon (Tilastokeskus, 2005c).

Euroopan Unionin jäsenyyden myötä tilastollisia aineistoja on luovutettava Komission päätöksellä (EY, Nro 281/1997) yhteisöjen tilastojen tuottajaksi asetetun Eurostatin käyttöön, mikäli yksittäisiä tilastoja koskevissa EY-säädöksissä niin edellytetään. Neuvoston asetukseen yhteisön tilastoista (EY, Nro 322/1997) sisältyy yleiset EU:n tilastotoimen periaatteet ja määräyksiä tietosuojan osalta. Lisäksi jäsenmaiden tilastotietojen luovuttamista Eurostatille säätelee Neuvoston asetus salassapidettävien tilastotietojen luovuttamisesta Euroopan yhteisöjen tilastotoimistolle (EY, Nro 1588/1990). Henkilötietojen suojaamisesta on määrätty sekä Euroopan parlamentin ja neuvoston direktiivillä (EY, Nro 46/1995) että Euroopan neuvoston henkilötietoja koskevalla suosituksella (EY, No. R (97) 18/30.9.1997). Lisäksi Käytännösäännöissä (KOM 217, 2005) annetaan muun muassa ehdoton suositus tilastosalaisuudesta. Komission asetuksessa yhteisön tilastoista annetun neuvoston asetuksen (EY) nro 322/1997 täytäntöönpanosta tieteellisiin tarkoituksiin käytettävien salassapidettävien tietojen käyttöoikeuden osalta (EY, Nro 831/2002) määritetään tilanteet ja menettelyt, joiden toteutuessa tutkijan käyttöön on mahdollista luovuttaa Eurostatin halussa olevia tietoja.

Julkaistavien tietojen on oltava sellaisessa muodossa, ettei tietojen kohde tai kohdetta koskevia salassapidettäviä tietoja paljastu. Näitä salassapidettäviä tietoja voidaan poikkeustapauksessa julkaista ainoastaan asianomaisen tietojen kohteen suostumuksella. Tätä mahdollisuutta sovelletaan pääasiassa vain suuryritysten tietoihin. Luovutettaessa tietoja tutkimuskäyttöön tai niitä julkaistaessa on aina tarkistettava salassapidettävien tietojen suojaamisen tarve. Aineisto on suojattava käyttämällä soveltuvia tilastollisia tietosuojamenetelmiä, mikäli yksikkötason tietoja on mahdollista paljastaa kohtuullisella vaivalla (Hänninen, 1997). Aineistoa suojatessa tulee kiinnittää huomiota tietojen laatuun, arkaluonteisuuteen sekä aineiston käyttötarkoitukseen sopivan menetelmän valintaan, koska tietosuojamenetelmät voivat vaikuttaa tilastojen sisältöön (Konnu, 2006).

Tilastolliset tietosuojamenetelmät

Salassapidettävien tietojen suojaamiseen on käytettävissä vaihtoehtoisia tietosuojamenetelmiä. Tietosuojamenetelmän valinnassa pitää ottaa huomioon aineiston tuleva käyttötarkoitus, koska menetelmien vaikutuksessa suojattavaan aineistoon voi olla suuriakin eroja. Kaikkien tietosuojamenetelmien tavoitteena on saada aineisto muokattua muotoon, josta yksikön tai sen ominaisuuksien paljastaminen on mahdotonta. Taulukkoaineistoille ja yksikkötason aineistoille on olemassa omat toisistaan eroavat tietosuojamenetelmänsä. Siksi tilastollisen tietosuojan menetelmiä on syytä esitellä erikseen mikroaineistoille ja taulukkoaineistoille. Seuraavassa on esimerkkejä Tilastokeskuksessa testatuista ja/tai käytössä olevista tietosuojamenetelmistä.

I. Taulukkomuotoisten aineistojen tietosuoja

Taulukkomuotoisten aineistojen (*tabular data*) muuttujien arkaluonteisuus määrittelee ensisijaisesti tarvittavien tietosuojatoimenpiteiden vahvuuden. Taulukon tietojen suojaustarpeen arvioimiseen vaikuttavat muuttujien arkaluonteisuuden lisäksi havaintojen määrä taulukon eri soluissa, tarkasteltavan perusjoukon suuruus, muuttujien määrä ja tilastointialueen koko ja luokittelun tarkkuus. Ennen taulukon julkaisemista on hyvä vielä kerran tarkistaa taulukosta ilmenevät tiedot. Varsinaisen taulukon lisäksi on aina muistettava tarkistaa kaikki julkaistavat taulukot, jotka ovat yhteydessä kyseiseen taulukkoon. Tilaston julkaisijan on varmistettava, ettei taulukoita vertaamalla ja tutkimalla voida paljastaa salassapidettävää tietoa.

Taulukkomuotoisten aineistojen tietosuojausta tarkastellessa tulee erottaa toisistaan frekvenssitaulukot ja määrätaulukot (*tables of magnitudes*). Frekvenssitaulukot kertovat kuhunkin taulukon soluun kuuluvien havaintojen määrän. Määrätaulukoissa esitetään puolestaan numeerisia arvoja, kuten esimerkiksi tulojen keskiarvoja tai tulosummia euroina. Määrätaulukoiden suojaamiseen kuuluu yleensä enemmän aikaa ja muita resursseja kuin frekvenssitaulukoiden suojaamiseen. Frekvenssitaulukoiden suojaamisessa käytetään yleensä ennalta määrättyä kynnsarvoa, joka kertoo sen lukumäärän, jota pienemmät solufrekvenssien arvot peitetään. Joskus peittämistä voidaan välttää yhdistämällä luokkia siten, että yhdistetyn luokan solufrekvenssit ylittävät kynnsarvon.

Määrätaulukoiden suojaamisessa tarvitaan tietoa taulukon arvojen lisäksi myös yksiköiden arvoista. Taulukon tai sen solun arkaluonteisuuden määrittelemiseen on periaatteessa olemassa kaksi kriteeriä: havaintojen määrä solussa ja solun havaintojen niin sanottu dominanssi. Mikäli tietyn solun arvoon vaikuttaa vain yksi tai muutama yksikkö, on solun arvo peitettävä. Tällöin on peittämisen lisäksi tarkistettava, ettei solun arvo ole laskettavissa taulukon marginaalisummiin kautta. Jos laskeminen on mahdollista, on peitettävä alkuperäisen solun lisäksi muitakin marginaalisummaan vaikuttavia soluja, eli käytettävä *komplementääristä peittämistä* (Duncan et al. 2001). Myös *dominanssisääntöä* eli (n,k)-sääntöä sovellettaessa tarvitaan tieto kaikista soluun vaikuttavista yksiköistä ja solun arvoon vaikuttavista luvuista. Esimerkiksi dominanssisäännön (2,80) tapauksessa solu täytyy suojata, jos kaksi suurinta soluun vaikuttavaa yksikköä muodosta yli 80 prosenttia solun arvosta.

Määrätaulukoiden tarkempien tietosuojamenetelmien tavoitteena on estää julkaistaviin lukuihin eniten vaikuttavien yksiköiden mahdollisuudet paljastaa toisten arvoja. Esimerkiksi alansa toiseksi suurin yritys voi halutessaan arvioida suurimman kilpailijansa liikevaihdon melko tarkasti, jos yrityksen arvio alan muiden yritysten liikevaihdon suuruudesta on kohdallaan. Tällainen paljastaminen voidaan estää julkaisemalla solujen tarkan arvon sijaan väli, jolle tarkka arvo kuuluu. Suojausta vaativien solujen havaitsemiseen voidaan käyttää ”*p%* -sääntöä” (Merola, 2005). Jos joku sellaisista yksiköistä, jonka tiedot vaikuttavat tarkasteltavan solun arvoon, pysyy päättelämään toisen, tarkasteltavan solun arvoon vaikuttavan, yksikön muuttujanarvon *p* prosentin tarkkuudella, on kyseinen solu *p%* -säännön mukaan suojattava. Tarkastellaan edellisen esimerkin suojausta 80%-säännön mukaan. Tällöin liikevaihdon sisältävä solu olisi suojattava, jos alan toiseksi suurin yritys saa arvion suurimman kilpailijansa liikevaihdosta 80 prosentin tarkkuudella.

Suojaamisen toteuttamiseksi voidaan käyttää apuna erilaisia tietosuojaohjelmia, kuten esimerkiksi Argus-ohjelma, joka on sisällytetty SuperCross -ohjelmaan, sekä pienaluetilastojen tietosuojamenetelmiä.

II. Yksikkötason aineistojen tietosuoja

Yksikkötason aineistoja (*microdata*) voidaan suojata aineistoa rajoittavin (*non-perturbative*) tai aineistoa muokkaavin (*perturbative*) menetelmin (Group Crises, 2004a ja 2004b). Aineistoa rajoittavat menetelmät ovat nimensä mukaisesti sellaisia, joiden suojaavuus perustuu salassapidettävien tietojen poistamiseen aineistosta. Tehokkain aineistoa rajoittava tapa suojata yksikköaineistoa paljastumiselta on julkaista kokonaisaineiston sijaan otosaineisto ilman tunnistetietoja. Kokonaisaineistolla tarkoitetaan esimerkiksi kaikkia tietyn tutkinnon suorittaneita tai kaikkia tietyllä alueella asuvia. Kaikki Tilastokeskuksesta tutkimuskäyttöön luovutettavat aineistot ovat otoksia kokonaisaineistosta.

Muita Tilastokeskuksessa jo käytössä olevia yksikkötason aineiston tietosuojamenetelmiä on esimerkiksi luokittelun *karkeistaminen* eli luokittelun muuttaminen laajempiin luokkiin (*global recoding*), suurimpien arvojen luokkien yhdistäminen (*top coding*) ja vastaavasti pienimpien arvojen luokkien yhdistäminen (*bottom coding*). Karkeistaminen auttaa harvinaisten arvojen suojaamisessa ilman aineiston käytettävyyden hankaloitumista. Esimerkiksi henkilöaineistoissa kaikki yli 80-vuotiaat voidaan luokitella samaan luokkaan, vaikka muu aineisto olisi luokiteltu viiden ikävuoden pituisiin luokkiin.

Kun aineiston muuttujanarvoista muodostuu ainutlaatuisia arvoyhdistelmiä, käytetään harvinaisten *arvojen peittämistä*. Esimerkiksi jos aineistossa on mukana vain yksi tietyn ammatin edustaja, voidaan hänen ammattitietonsa merkitä puuttuvaksi tiedoksi, jolloin tunnistamisen vaara yleensä poistuu. Tämän menetelmän heikkoutena on, että tunkeutuja eli paljastamista yrittävä henkilö voi käyttää puuttuvien tietojen merkintää apuna etsiessään helposti paljastettavia yksiköitä. Siksi peitetyt arvoja ei saa korvata muista puuttuvista arvoista eroavalla merkinnällä, vaan kaikille eri syistä puuttuville tiedoille on käytettävä samaa merkintää.

Aineistoa rajoittavien tietosuojamenetelmien avulla suojatuissa aineistoissa ei ole riskiä analyysien tulosten muuttumisesta muun kuin korkeintaan suojattavien arvojen osalta. Kuitenkin liian laajoihin luokkiin luokiteltu tai paljon puuttuvia tietoja sisältävä aineisto voi tuottaa hankaluutta aineiston käyttäjälle. Aineistoa

muokkaavia menetelmiä on syytä käyttää, jos aineiston käyttäjä tarvitsee erityisen tarkkoja tietoja. Nämä menetelmät voidaan jakaa edelleen systemaattisesti ja satunnaisesti muokkaaviin menetelmiin. Systemaattisten menetelmien aiheuttama muutos aineistossa on samanlainen jokaisella suojauskerralla, kun taas satunnaisen menetelmien tuloksena on joka suojauskerralla erilainen suojattu aineisto.

Systemaattisiin suojausmenetelmiin kuuluu esimerkiksi *mikroaggregointi* (Group Crises, 2004c), jossa aineiston yksiköt ryhmitellään samankaltaisten yksiköiden ryhmiin ja yksikön muuttujanarvot korvataan ryhmän keskiarvoilla. Ryhmittelyn toteuttamiseen on useita vaihtoehtoja ja ryhmäkoon voi antaa vaihdella tai sen voi pitää kiinteänä. Aineistoa satunnaisesti muokkaavista menetelmistä tyypillisin esimerkki on *kohinan lisääminen* muuttujanarvoon (Brand, 2002). Tällä tarkoitetaan satunnaisen arvon lisäämistä alkuperäiseen muuttujan saamaan arvoon. Kohinan lisääminen erilaisten muunnosten kautta tehostaa suojausvaikutusta. Toinen yksinkertainen esimerkki on tietyn muuttujan *arvojen vaihtaminen yksiköiden välillä* (Fienberg ja McIntyre, 2004) joko satunnaisesti tai arvojen järjestykseen perustuen. Kehittyneempiä satunnaisesti muokkaavia menetelmiä on esimerkiksi *PRAM -menetelmä* (Gouweleeuw, 1998), jonka suojaavuus perustuu muuttujan arvojen vaihtumiseen ennalta valitun todennäköisyysmatriisin mukaisesti.

Yksikköaineistojen osalta kattavien, jokaiseen tilanteeseen soveltuvien tietosuoaohjeiden antaminen on mahdotonta, koska tarvittavat toimenpiteet riippuvat aina monesta eri tekijästä, kuten aineiston koosta, muuttujien määrästä ja luokitustarkkuudesta. Myös mikroaineistojen suojaamisen apuna voidaan käyttää tietosuoaohjelmistoja, kuten esimerkiksi Argus -ohjelmaa.

Suosituks

Yleisperiaatteet

Tilastolain peruseriaate on, että kaikki tilastotarkoituksiin saadut tiedot ovat sallassapidettäviä. Tiedonkeruussa, käsittelyssä ja julkaisussa on otettava huomioon seuraavat perusohjeet:

- Vain tilaston laatimisen kannalta välttämättömät tiedot kerätään.
- Valmiiksi kerättyjä rekisteritietoja on hyödynnettävä mahdollisimman paljon.
- Tunnistetietoja kerätään ja käytetään vain silloin, kun se on välttämätöntä.
- Ennen aineiston luovuttamista aineistosta on poistettava tai karkeistettava myös välillisesti identifioivat muuttujat.
- Aineistojen tietosuojamenettelyt on aina dokumentoitava.

Yleisten tietojen käsittelyn periaatteiden lisäksi tietoja julkaistaessa tai luovutettaessa on otettava huomioon seuraavat tarkemmat ohjeet. Tarkasteltaessa, mitä tietoja voidaan luovuttaa, on erotettava toisistaan taulukkomuotoiset tiedot ja yksikötason tiedot.

Suosituksia taulukkomuotoisen tiedon kohdalla

- Taulukkomuotoisia tietoja julkaistaessa on noudatettava Tilastokeskuksen ohjetta taulukkomuotoisten henkilötietojen suojaamiselle (Tilastokeskus, 2002) tai Tilastokeskuksen ohjetta taulukkomuotoisen yritystiedon suojaamiselle (Tilastokeskus, 2000b).
- Henkilö- tai yritystietoihin perustuvat tilastotaulukot tulee laatia niin, että taulukon solut eivät sisällä tarpeettoman pieniä arvoja. Taulukkoaineistojen sisältämät henkilöihin, yksittäisiin perheisiin, asutokuntiin tai yrityksiin kohdistuvat tietosuojariskit on aina arvioitava ennen tietojen julkaisemista tai luovuttamista.
- Taulukkomuotoisen tiedon suojaamiseen tulee valita tarkoituksenmukainen menetelmä, jonka avulla taulukon hyödyllisyys säilyy suojauksen jälkeen mahdollisimman hyvänä.
- Taulukkomuotoinen tieto lähenee mikroaineistoa, esimerkiksi muodostettaessa moniulotteinen taulukko tarkoilla luokituksilla. Tällöin on tapauskohtaisesti harkittava, onko aineistoa tarkasteltava yksikkötason aineistona, jonka luovuttamiseen tarvitaan käyttöluupa.

Suosituksia yksikkötason tietojen kohdalla

- Yksikkötason tietoja luovutettaessa on noudatettava ohjetta käyttö lupien myöntämisestä Tilastokeskuksen yksikkötason aineistoon (Tilastokeskus, 2005c).
- Yksikkötason tietoja voidaan luovuttaa ainoastaan tieteellisiin tutkimuksiin tai tilastollisiin selvityksiin.
- Yksikkötason tietojen luovuttaminen tapahtuu aina käyttö lupamenettelyn kautta. Käyttö lupaa myönnetään yksilö ityyn tutkimustarkoitukseen määrä ajaksi.
- Kokonaisaineistoja ei luovuteta Tilastokeskuksen ulkopuolelle. Ainoastaan otosaineistoja voidaan luovuttaa muualle, mutta kokonaisaineistoja voidaan kuitenkin antaa tutkijoiden käytettäväksi Tilastokeskuksen tiloissa ja valvonnassa.
- Luovutettavan yksikkötason aineiston tulee olla sellainen, etteivät yksiköt ole siitä tunnistettavissa.
- Välillisen tunnistamisen estämiseksi on käytettävä tietosuojamenetelmiä. Yritysaineistoista ei yleensä voida sulkea pois välillisen tunnistamisen mahdollisuutta, joten ne voidaan antaa tutkijoiden käytettäväksi ainoastaan Tilastokeskuksen tiloissa ja valvonnassa.
- Yksikkötason aineistojen luovutuksissa on aina tarkastettava tietosuojaperiaatteiden toteutuminen.

Kirjallisuus

EU:n lainsäädäntö ja kansainväliset toimintaohjeet

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.

Euroopan parlamentin ja neuvoston direktiivi (EY) nro 46/95 yksilöiden suojasta henkilötietojen käsittelyssä.

<http://europa.eu.int/eur-lex/fin/index.html>

International Statistical Institute (ISI). 1985. *Declaration on Professional Ethics*.

<http://isi.cbs.nl/ethics.htm>

Komission asetus (EY) nro 831/2002 yhteisön tilastoista annetun neuvoston asetuksen (EY) N:o 322/97 täytäntöönpanosta tieteellisiin tarkoituksiin käytettävien salassapidettävien tietojen käyttöoikeuden osalta.

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

http://forum.europa.eu.int/irc/dsis/bmethods/info/data/new/legislation/role_of_eurostat.html

Euroopan Yhteisöjen Komissio. 2005. *Kansallisten ja yhteisön tilastoviranomaisten riippumattomuudesta, koskemattomuudesta ja vastuuvellisuudesta*. KOM (2005) 217.

http://eur-lex.europa.eu/LexUriServ/site/fi/com/2005/com2005_0217fi01.pdf

Neuvoston asetus (EY) nro 1588/90 salassapidettävien tilastotietojen luovuttamisesta Euroopan yhteisöjen tilastotoimistolle.

<http://forum.europa.eu.int/irc/dsis/bmethods/info/data/new/legislation/confidentiality.html>

Neuvoston asetus (EY) nro 322/97 yhteisön tilastoista.

http://forum.europa.eu.int/irc/dsis/bmethods/info/data/new/legislation/stat_law.html

Yhdistyneiden Kansakuntien tilastokomission hyväksymät virallisen tilaston peruseriaatteen. 1994

Suomen lait ja asetukset sekä kansalliset toimintaohjeet

Tilastolaki (280/2004).

<http://www.finlex.fi/fi/laki/ajantasa/2004/20040280>

Henkilötietolaki (L 523/1999).

Laki viranomaisten toiminnan julkisuudesta (L 621/1999).

Asetus viranomaisten toiminnan julkisuudesta ja hyvästä tiedonhallintatavasta (L 1030/1999).

Valtiovarainministeriö. 2000. *Valtionhallinnon tietoaineistojen käsittelyn tietoturvasuositukset*. Ohje VAHTI 2/2000.

Valtiovarainministeriö. 2002. *Arkaluonteiset kansainväliset tietoaineistot*. Ohje VAHTI 4/2002.

Tilastokeskuksen toimintaohjeet

Tilastokeskus. 2000a. *Henkilötietolain soveltamisohje*. (TK-00-578-00). Helsinki.

Tilastokeskus. 2000b. *Ohje taulukkomuotoisen yritystiedon suojaamiselle*. (TK-00-888-00). Helsinki.

Tilastokeskus. 2002. *Ohje taulukkomuotoisten henkilötietojen suojaamiselle*. (TK-00-314-02). Helsinki.

Tilastokeskus. 2005a. *Tilastolain (280/2004) soveltamisohje Tilastokeskuksessa*. (TK-00-198-05). Helsinki.

Tilastokeskus. 2005b. *Tietojen suojaaminen Tilastokeskuksessa. Tietosuojaohje*. (TK-00-334-05). Helsinki.

Tilastokeskus. 2005c. *Ohje käyttöluoppien myöntämisestä Tilastokeskuksen yksikötason aineistoon*. (TK-00-128-05). Helsinki.

Tilastokeskus. 2006. *Tilastokeskuksen ammattieettinen opas*. Käsikirjoja 30. Helsinki.

http://tilastokeskus.fi/org/etiikka/org_etiikka_etiikkaopas.html

Tilastolliset tietosuojamenetelmät

Brand, R. 2002. Microdata Protection through Noise Addition. Teoksessa: *Inference Control in Statistical Databases*, toim. Domingo-Ferrer, J. Berlin, Heidelberg: Springer. 97–116.

Group Crises 2004a. Non-Perturbative Methods for Microdata Privacy in Statistical Databases. *Research Report CRIREP-04-007*.

http://vneumann.etsu.es/publications/reports/nonperturbative_methods.pdf

Group Crises 2004b. Perturbative Masking for Microdata Privacy Protection in Statistical Databases. *Research Report CRIREP-04-008*.

http://vneumann.etsu.es/publications/reports/perturbative_methods.pdf

Group Crises, G. 2004c. Microaggregation for Privacy Protection in Statistical Databases. *Research Report CRIPREP-04-005*.

<http://vneumann.etsu.es/publications/reports/microaggregation.pdf>

Duncan, G.T., Fienberg, S.E., Krishnan, R., Padman, R. & Roehrig, S.F. 2001. Disclosure Limitation Methods and Information Loss for Tabular Data. Teoksessa: *Confidentiality, Disclosure, and Data Access: Theory and Practical Applications for Statistical Agencies*, toim. Doyle, P., Lane, J.I., Theeuwes, J.J.M. & Zayatz, L.V. Amsterdam: North-Holland. 135–166.

Fienberg, S.E. & McIntyre, J. 2004. Data Swapping: Variations on a Theme by Dalenius and Reiss. Teoksessa: *Privacy in Statistical Databases*, toim. Domingo-Ferrer, J. & Torra, V. Berlin, Heidelberg: Springer. 14–29.

- Gouweleeuw, J.M., Kooiman, P., Willenborg, L.C.R.J. & de Wolf, P.-P. 1998. Post Randomisation for Statistical Disclosure Control: Theory and Implementation. *Journal of Official Statistics*. Vol. 14. No. 4. 463–478.
- Hänninen, M. 1997. Tilastolliset tietosuojamenetelmät ja niiden käyttö. Tilastokeskus. *Katsauksia* 1997/3. Helsinki.
- Konnu, J. 2006. *Mikroaineistojen tilastolliset tietosuojamenetelmät henkilötalustoissa*. Tilastotieteen pro gradu -tutkielma. Jyväskylän yliopisto: Matematiikan ja tilastotieteen laitos.
- Merola, G.M. 2005. Safety Rules in Statistical Disclosure Control for Tabular Data. Teoksessa: *Monographs of official statistics, Work Session on Statistical Data Confidentiality (Geneva, 2005)*. Luxembourg: Office for Official Publications of the European Communities. 247–254.

2.3 Luokitukset, käsitteet ja määritelmät

Soveltamisala ja päämäärä

Tilasto tehdään yleensä suuresta joukosta yksittäisiä havaintoja. Johtopäätösten tekemiseksi empiirisiä havaintoja on järjestettävä, luokiteltava esimerkiksi samanlaisuuden tai erilaisuuden perusteella. Luokituksen avulla havaintoaineisto voidaan järjestää tavalla, joka kuvaa tarkastelun kohteena olevaa ilmiötä johdonmukaisesti. Siten luokitus on olennainen osa tilastoja ja standardiluokitukset tilastotoimen 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). Yleensä luokituksen kuvaus on kuitenkin tätä laajempi ja koostuu (i) luokkien määrittelystä (luokkien nimet, koodit ja usein myös luokan kuvaus, eli sen tekstimuotoinen määrittely) sekä monimutkaisten ja laajojen luokitusten tapauksessa (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.

Periaatteet

Luokitteluperiaatteet

Aineisto, havainnot, on ryhmiteltävä eli luokiteltava niin, että laadittu tilasto tuottaa todellisuutta vastaavan, jäsennellyn 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 siten, 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 määriteltujen luokitusperiaatteiden ja kriteereiden perusteella selvästi tiettyyn luokkaan. Kattavuus merkitsee sitä, että kaikki havaintoaineiston tapaukset voidaan luokitella. Poissulkevuus toteutuu, kun luokkien määrittelyt on tehty niin, ettei tapauksia voida sijoittaa 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. Kaatoryhmästä käytetään usein myös termiä "residuaaliluokka".

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, jotka velvoittavat EU-jäsenmaita käyttämään kyseisiä luokituksia tilastotoimessa tai toimitettaessa tietoja yhteisön tilastovirastolle, Eurostatille. Tällaisia säädöksiä vahvistettuja luokituksia ovat muun muassa toimiala- ja tuoteluokitukset sekä uudistetun kansantalouden tilinpidon (ESA 95) perusluokitukset. Sen sijaan väestöä ja elinoloja kuvaavat luokitukset perustuvat väestölaskentasuosituksiin. Tulonjakoa ja kulutusta kuvaavien luokitusten taustalla on kansantalouden tilinpidon kehikon lisäksi tulotilastosuositus.

Tilastokeskus ylläpitää luokitustietokantaa, johon on viety kansallisen tilastotoimen käyttämät luokitusstandardit. Tämä tietokanta on tarkoitettu vain viraston sisäiseen käyttöön. Tilastokeskuksen Internet-palvelussa on tietoa ulkoisille käyttäjille tilastojen luokituksista, verkko-osoite: <http://tilastokeskus.fi/luokitukset/index.html>

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

Yleisiä tilastostandardeja 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.

Suosituksset

- 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 säädöksin vahvistamista luokituksista on muodollisesti hyväksyttävä 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 kansainväliin standardiluokitukseen lisätään kansallinen detaljitaso luokituksen alimmaksi tasoksi. Näin meneteltäessä kansallisen luokituksen rakenne ja ylemmät luokitustasot seuraavat puhtaasti kansainvälistä luokitusta.

Yksittäisiä aihealueittaisia suosituksia

I. Valtioiden ja maiden tunnukset

Kansalaisuutta, syntymä- ja asuinmaata, muuttoliikettä, majoitusta, ulkomaankauppaa ja yritysten kotimaata tilastoitaessa voidaan käyttää kansainvälistä standardia ISO 3166 valtioiden ja maiden tunnuksista. Tunnusjärjestelmään kuuluvat valtioiden ja maiden kolminumeroiset, kaksikirjaimiset sekä kolmikirjaimiset tunnukset. Tilastojärjestelmissä suositellaan käytettäväksi ensisijaisesti kolminumeroista tunnusta, joka on YK:n tilastotoimiston antama. Kaksikirjaiminen tunnus on tarkoitettu yleiskäyttöön. Kolmikirjaimista tunnusta voidaan käyttää, kun halutaan helpottaa alueen tunnistamista ja lisätä havainnollisuutta. Kirjaintunnukset perustuvat lähinnä valtioiden ja maiden englanninkielisiin nimiin.

Tunnusjärjestelmään sisältyvät myös itsenäiset valtiot sekä epäitsenäiset alueet (maat). Valtioiden ja maan ryhmitely kuuluu maanosan ja edelleen 19 suuralueeseen. Jako soveltuu käytettäväksi lähinnä väestötilastoissa. Valtioita ja maita voidaan ryhmitellä myös muilla tavoilla, esimerkiksi kauppapoliittisin perustein tai valtioliittoumien mukaan.

II. Alueluokitukset ja Ahvenanmaan erityisaseman huomioon ottaminen

Tilastoissa käytetään pääasiassa Euroopan unionin virallista NUTS-alueluokitusjärjestelmää (NUTS, *Nomenclature des Unités Territoriales Statistiques*, joka on määriteltä Euroopan parlamentin ja neuvoston asetuksessa 1059/2003). Suomen NUTS-aluejaossa Manner-Suomi ja Ahvenanmaa muodostavat NUTS1-alueen, suuralueet NUTS2-, maakunnat NUTS3-, seutukunnat NUTS4- ja kunnat NUTS5-tason.

Maakunnat ja tilastollinen kuntaryhmitely 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 myös NUTS1-aluejako (Manner-Suomi, Ahvenanmaa) väliotsikkoina, välisummineen tai ilman niitä.

III. Tilastollinen kuntaryhmitely

Kuntien kaupunkimaisuutta kuvattaessa käytetään tilastollista kuntaryhmitystä, jonka mukaan kunnat luokitellaan kaupunkimaisiin, taajaan asuttuihin ja maa-seutumaisiin kuntiin. Aiemmin käytössä olleesta kuntamuotoluokituksesta ”kaupungit/maalaiskunnat” luovuttiin vuonna 1997, koska se ei ollut pitkään aikaan toiminut kaupunki/maaseutu -käsiteparin synonyyminä ja, koska vastaava jako poistettiin myös lainsäädännöstä. (Laki tuntee vain kunnat, jotka voivat itse päättää, kutsuvatko kuntaansa kaupungiksi vai ei.)

IV. Kotitalouskäsite

Suomessa kotitalouskäsitettä voidaan käyttää vain haastattelututkimuksissa ja postikyselyissä. Rekisteripohjaisissa väestölaskennoissa ja vuositilastoissa sekä muissa hallinnollisiin aineistoihin perustuvissa tilastoissa käytetään käsitettä asutokunta.

V. Toimiala- ja tuoteluokitukset

Euroopan unionin jäsenvaltioiden toimialoittaiset tilastot pitää laatia käyttäen NACEa tai siitä johdettua kansallista luokitusta. Tilastovuodesta 2003 käytössä on ollut NACE 2002 (NACE Rev.1.1), ja siihen perustuva kansallinen toimialaluokitus TOL 2002.

Tuoteluokitusta (*Classification of Products by Activity, CPA*) käytetään raportointiluokituksena toimitettaessa esimerkiksi 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.

Dokumentointi

Luokitukset, käsitteet sekä määritelmät ja käsittelysäännöt on dokumentoitava huolellisesti: luokitukset luokitustietokantaan ja käsitteet käsitetietokantaan.

Otostutkimuksiin liittyvät käsittelysäännöt ja muut työhöjeet talletetaan tilastoarkistoon, mikäli niitä ei ole kuvattu koontitiedostossa. Näin menetellen tilastoaineistot ja niihin liittyvät luokitukset, käsitteet ja muut metatiedot ovat käytettävissä esimerkiksi myöhemmin tapahtuvaa tutkimusta varten.

Kirjallisuus

Luokitustiedon lähteitä

Tilastokeskuksessa Luokitus- ja metatietopalvelut -vastuualue ylläpitää luokituksia luokitustietokannassa, julkaisee luokituskäsi kirjoja ja välittää luokitusinformaatiota intrassa ja Internetissä. Keskeiset luokitukset on esitelty myös Luokitusopas-vihkosessa, jota voi jakaa asiakaskoulutuksen tilaisuuksissa ja esimerkiksi koordinoinnin tapaamisissa. Luokitusopas-vihkosta saa luokituspalveluista.

Luokitustietokanta sisältää luokitusstandardit erikielisinä (suomi, ruotsi, englanti) ja -pituisine luokitusversioineen, luokitusten metatiedot, suurten luokitusten hakemistot (hakusanat) ja eri ajankohdilta olevien luokitusten väliset muunnosavaimet. Luokitustietokanta toimii luokitustuotteiden, kuten käsikirjojen, tuotantoympäristönä ja palvelualustana. Luokitustietokanta on kaikkien Tilastokeskuksen työntekijöiden käytettävissä.

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

Eurostatin käyttämät luokitusstandardit ja linkit useisiin laajassa kansainvälisessä käytössä oleviin luokitusstandardeihin on viety RAMON-luokitustietokantaan. Verkko-osoite on <http://euro-pa.eu.int/comm/eurostat/ramon/>

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

2.4 Kehikot ja peittävyys

Soveltamisala ja päämäärä

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

Tilastotutkimuksen tulosten yleistettävyys, relevanssi ja tarkkuus edellyttävät oikeanmukaisen kehikon käyttöä tiedonkeruussa. Sekä kokonais- että otantatutkimuksissa kehikon peittävyys, täydellisyys, ajantasaisuus, tietosisältö ja tietojen tarkkuus ovat kriittisiä tekijöitä kehikon sopivuudelle 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 tilastotutkimuksen käyttämien kehikoiden yleispiirteet ja esitetään kehikoiden suhde tilastotutkimuksen tulosten yleistettyteen ja laatuun. Seuraavassa kappaleessa (2.5 Hallinnolliset aineistot ja rekisterit) kuvataan tarkemmin Tilastokeskuksen yleisesti käyttämien kehikoiden luonne ja erityispiirteet.

Periaatteet

Tilastotutkimuksen perusjoukko tulee rajata siten, että jokaisen yksikön kuulumus perusjoukkoon voidaan yksiselitteisesti määrittää. Seuraavaksi tilastotutkimuksen suunnittelussa tulee selvittää vaihtoehtoiset kehikot ja niiden soveltuvuus. 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.

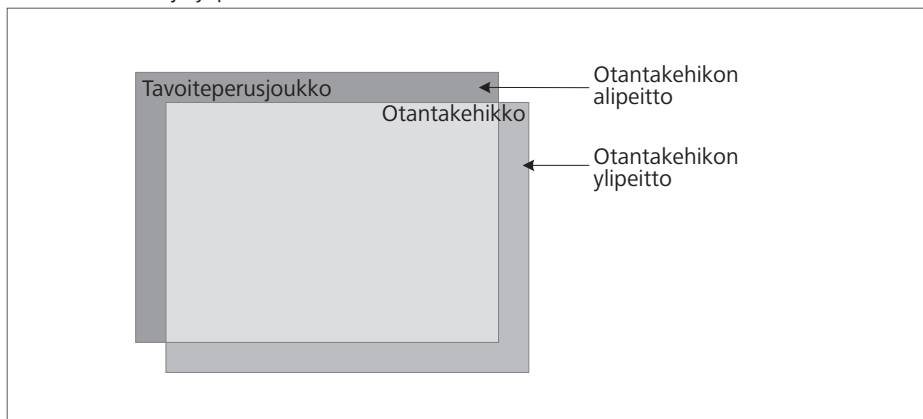
Kehikon sisältämien tietojen tulee olla ajan tasalla. Rekistereitä käytettäessä osite- ja luokittelutietojen tulee olla päivitettyjä viimeisten mahdollisten tietojen mukaisiksi. Rekistereitä yhdisteltäessä otantakehikoksi tulee yhdistelytietojen, esimerkiksi henkilö- ja kotitaloustutkimuksissa, kohdehenkilöiden henkilötun-

nusten 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 (kuvio 2.4). Otantakehikon ylipeitto voidaan mitata, mutta alipeittoa on vaikeampi arvioida (Djerf, 2000). Alipeittoa varten 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.

Kuvio 2.4 Tilastotutkimuksen tavoiteperusjoukko, kohdeperusjoukko, ja kehikoiden ali- ja ylipeitto



Kun tutkimuksen tavoiteperusjoukko on määritelty, tulee kohdeperusjoukon täsmällistä määrittelyä 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 demografisin perustein:* Tilastotutkimus saattaa kohdentaa henkilöihin tietyn demografisen taustatekijän mukaan, esimerkiksi täysi-ikäi-

seen väestöön; vastaavasti yritysperustainen tilastotutkimus voi keskittyä esimerkiksi tiettyjen toimialojen yrityksiin.

- *Ryhmäkohtaisin perustein:* Jos tilastotutkimuksen yksikkö on erityyppinen kuin kehikon, tulee tutkimusyksikkö määritellä yksiselitteisesti. Tällöin ne yksilöt, jotka selkeästi eivät ole tilastotutkimuksen yksikön määritelmän piirissä, tulee poistaa kohdeperusjoukosta. Esimerkiksi laitospöytäkirjoon kuuluvat henkilöt pyritään poistamaan kotitalousväestöön kohdistuvasta tilastotutkimuksen otantakehikosta. (Rossi, Wright ja Anderson, 1983).

Suositukseset

- Selvitetään mitä rekistereitä on käytettävissä kehikoiksi, niiden käytön kustannukset 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.
- Lasketaan otoksen poiminnan jälkeen apumuuttujien arvot kehikosta ja otoksesta ja tarkistetaan, että otoksen poiminta on tapahtunut oikein. Jos havaitaan poikkeuksellisen suuria satunnaisvirheitä, niin poimitaan uusi otos ja toistetaan tarkistukset. Kannattaa kerätä mahdollisimman paljon aputietoa ja tallentaa ne myöhempiä käyttöä varten.
- Arvioidaan rekisterin tilaa ja virheitä kerätyn aineiston avulla. Raportoidaan tietojenkeruuvaiheessa havaittu ylipeitto, rekisteristä saatujen luokittelumuuttujien virheet ja arvioidaan alipeittoon kuuluvien määrät muita tietolähteitä käyttäen. Seurataan haastattelijoilta tai muualta tulevaa palautetta osoitetietojen ajantasaisuudesta ja luokittelumuuttujien oikeellisuudesta.
- Tutkimustulosten julkistamisen yhteydessä raportoidaan kohdeperusjoukko, otantakehikko ja sen peittävyys.

Kirjallisuus

- Djerf, K. 2000. Properties of Some Estimators under Unit Nonresponse. Statistics Finland. *Research Reports 231*. Helsinki.
- 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.

2.5 Hallinnolliset aineistot ja rekisterit

Soveltamisala ja päämäärä

Virallisen tilastotuotannon kannalta hallinnollisella aineistolla tarkoitetaan mitä tahansa aineistoa, joka on ensisijaisesti kerätty muuta tarkoitusta kuin virallista tilastotuotantoa varten. Määritelmän puitteissa myös muut tahot kuin viranomaiset voivat kerätä ja pitää hallinnollisia aineistoja. Tämä määritelmä kiinnittää huomion virallisen tilastotuotannon kannalta tärkeimpään hallinnollisten aineistojen ominaisuuteen: koska tiedot on kerätty muuta tarkoitusta varten, niiden käyttäminen ei lisää tiedonantajien vastausrasitusta. Kustannukset tilastojen tuottajalle ovat myös hyvin pienet verrattuina suoraan keruuseen. Nämä edut ovat yleensä merkittävämpiä kuin hallinnollisten aineistojen käyttöön liittyvät haitat. Myös tilastolaki (280/2004) korostaa vastausrasitteen minimoimista, mikäli tilastoinnin tarvitsema tieto on saatavilla hallinnollisista aineistoista.

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

Tilastorekisterillä tarkoitetaan rekisteriä, jota ylläpidetään ensisijaisesti tilastotuotannon tarpeita varten. Valtaosassa tapauksia tilastolaki itse asiassa kieltää tilastorekisterin käytön mihinkään muuhun tarkoitukseen kuin tilastotuotantoon. 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, joiden avulla yksiköt voidaan yksikäsitteisesti tunnistaa. Tunnusjärjestelmää voidaan kutsua yleiseksi tunnusjärjestelmäksi, jos sitä käyttävät useimmat ja keskeisimmät rekisterinpitäjät. Yleinen tunnusjärjestelmä, kuten henkilötunnus tai yritys- ja yhteisötunnus, mahdollistaa rekisterien yhdistelyn, suoran tiedonkeruun sekä aineistojen ja rekisteritietojen yhdistelyn.

Rekisterit

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

Suurin osa rekistereistä on hallinnollisia eli viranomaisten toimintaa palvelevia, kuten väestötietojärjestelmä, verotietokanta sekä 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 sekä konsernirekisteri. Lisäksi tutkintorekisteri kattaa väestön suorittamat perusasteen jälkeiset tutkinnot ja opiskelijarekisteri kattaa perusasteen jälkeisessä koulutuksessa olevat opiskelijat.

Keskeinen tietorekisteri on 1960-lopulla perustettu väestön keskusrekisteri (nykyään väestötietojärjestelmä), joka käsittää kaikki ne henkilöt, joilla on tai on ollut vakinainen asuinpaikka Suomessa. Väestötietojärjestelmässä on mukana myös niin sanottuja historiatietoja eli maastamuuttaneet ja kuolleet. Tietojärjestelmää täydennetään päivittäin väestön muutostiedoilla, eli syntyneillä, kuolleilla, maahan- ja maastamuuttaneilla, maan sisällä muuttaneilla, vihkimisillä, avioeroilla, adoptioilla ja nimenmuutoksilla. Väestötietojärjestelmä on kihlakuntien paikallisrekistereiden yhdistelmä, ja sen päivitys hoidetaan kihlakuntien rekisteritoimistoissa. Vuodesta 1980 lähtien väestötietojärjestelmä laajeni kattamaan myös rakennukset, asunnot, toimitilat ja vapaa-ajan asuinrakennukset. Rakennuksista ovat kattavasti mukana kaikki asuinrakennukset sekä sellaiset rakennukset, joissa on työpaikkoja. Tunnustietojen avulla henkilöt yhdistetään asuntoihin ja yritykset toimitiloihin.

Usein myös muita tietokantoja 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 suurimmaksi osaksi eri henkilöistä, joten se on pikemmin tiedosto kuin varsinainen rekisteri. Vastaavanlaisia tiedostoja ovat useat Tilastokeskuksen vuosiaineistot, vaikka niitä saatetaan kutsua rekistereiksi.

Rekisterin ja tiedoston välimuotona voidaan pitää myös 1960-luvun lopulla perustettua verotusrekisteriä (nykyään verotietokanta). Verotietokanta tulkitaan myös rekisteriksi, vaikka se muodostuu kunkin vuoden verovelvollisista henkilöistä ja heidän yhden vuoden tulo- ja varallisuustiedoistaan. Tiedostoa päivitetään väestötietojärjestelmän muutoksilla, kuten kuolleilla, muuttaneilla ja perhetiedoilla, mutta verotustiedot kerätään joka vuosi uudelleen tulonsaajilta, heidän työnantajiltaan ja eläkelaitoksiltaan. Edellisen vuoden tietoja ei siten päivitetä. Osaa verotustiedoista säilytetään tietokannassa: varallisuustietoja käytetään mahdollisen omaisuusveron ja myyntivoiton määrittämiseksi. Niitä ei tarvitse ilmoittaa uudelleen joka vuosi, ellei muutoksia ole tapahtunut.

Yritystilastojen kannalta tärkeimmät rekisterit ovat veroviranomaisten pitämät rekisterit. Ensimmäiset yritystoiminnan verotukseen tarkoitetut rekisterit kehitettiin 1970-luvulla. Tuolloin saatiin muun muassa rekisteröityä kaikki liikevaihtoverovelvolliset. 1980-luvulla yritysverotuksen tietojärjestelmiä uudistettiin: ne talletettiin elektroniseen muotoon ja liike- ja yhteisötunnukset (nykyinen yritys- ja yhteisötunnus) otettiin käyttöön. Näiden uudistusten jälkeen Tilastokeskus pystyi luomaan lähes kattavan yritysrekisterin, jossa yritysten aloitus- ja lopettamistietoja voidaan ylläpitää lähes ajantasaisesti.

Yritysverotuksen rekisterien hyödyntäminen tilastotoimessa keskittyi 1990-luvulle saakka yritysrekisteriin. 1990-luvulla hallinnollisten aineistojen hyödyntäminen lisääntyi yritystilastoissa huomattavasti. Yritysten rakennetilastot uudistettiin 1995–1996, jolloin elinkeinoverotusrekisteristä tehtiin keskeinen rakennetilastojen tietolähde. Vuonna 1998 käynnistettiin yritysten liikevaihtoa ja palkkasummia koskevien kuukausitilastojen tuotanto. Nämä perustuvat verottajan arvonlisäverotuksen tietokannan -osarekisteriin, joka sisältää tietoja arvonlisäverotuksesta ja rekisteröityjen työnantajien maksamista palkoista.

Periaatteet

Yksiköitä identifioivat tunnukset

Tunnustiedot identifioivat hallinnollisten aineistojen ja rekisterien yksiköt. Tunnustietojen avulla eri tietojärjestelmien 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. Tunnetuin näistä 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 osoitetunnustuksella (esimerkiksi yhdistettäessä niin sanottuja eläkejärjestelynumeroita Y-tunnuksiin).

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

Tärkeimmät tilastotuotannossa käytetyt hallinnolliset aineistot ja rekisterit

Väestö- ja henkilötiedot:

- väestötietojärjestelmä, joka sisältää väestötiedot, rakennus- ja huoneistotiedot, toimitilatiedot ja kesämökkitiedot (Väestörekisterikeskus)
- vanhuus-, työkyvyttömyys- ja työttömyyseläkeläiset, asumistukitiedot, sairaustiedot (Kansaneläkelaitos)
- toimeentulotukea saaneiden rekisteri (Sosiaali- ja terveystieteiden tutkimus- ja kehittämiskeskus, Stakes)
- hoitoilmoitusrekisteri (Stakes)
- varusmiespalveluksessa olevat (Pääesikunta)
- työeläkejärjestelmien työsuhdetiedot (Eläketurvakeskus, Valtiokonttori, Kuntien eläkevakuutus sekä muutamia muita pieniä tiedontuottajia)
- työnhakijarekisteri, työvoimapolitiittinen koulutus (työministeriö)
- opiskelijarekisteri (Tilastokeskus)
- tutkintorekisteri (Tilastokeskus)

Yritystiedot:

- verotuksen asiakasrekisteri (asiakastietokannasta), arvonlisäverotuksen tietokanta (sisältää kuukausittaiset arvonlisävero- ja työnantajasuoritustiedot), elinkeinoveroaineisto, verotusrekisterit, tulo- ja varallisuustiedot, työnantajan vuosivalvontatiedot
- yritys- ja yhteisötietojärjestelmä, joka on verohallituksen ja patentti- ja rekisterihallituksen yhteisesti ylläpitämä tietojärjestelmä (YTJ)
- kaupparekisteri (Patentti- ja rekisterihallitus)

- yritys- ja toimipaikkarekisteri (ml. valtion ja kuntien toimintayksiköt) (Tilastokeskus)
- maatala- ja puutarhayritysrekisterit (Maa- ja metsätalousministeriön tietopalvelukeskus)

Muut rekisterit:

- ajoneuvorekisteri (Ajoneuvohallintokeskus AKE)

Hallinnollisten aineistojen ja rekisterien käyttö tilastotuotannossa

Tavallisesti rekistereitä käytetään tilastotuotannossa sekä kantatilastoissa että muutostilastoissa. Esimerkiksi väkiluku ja väestön rakennetilastot tuotetaan rekisterin kantatiedoista. Väestönmuutostilastot (syntyneet, kuolleet, maassamuuttaneet, maastamuuttaneet, maahanmuuttaneet, vihityt ja eronneet) perustuvat rekisterin päivytystietoihin. Samoin rakennus- ja huoneistorekisteristä saatavat kantatiedot ja rakennustuotantotiedot perustuvat rekisterin ylläpitojärjestelmään. Kantatietoja ylläpidetään säännöllisesti myös ajoneuvo-, tutkinto-, yritys- ja toimipaikka- ja työnhakijarekisterissä, joista tuotetaan kantatilastoja ja muutostilastoja.

Hallinnollisten aineistojen ja rekisterien käyttö tilastotuotannossa pienentää sekä tilastoinnin kustannuksia että tiedonantajien vastausrasitetta, mikä on yksi keskeisistä tilastolain (280/2004) periaatteista. Esimerkiksi rekisteripohjaisen väestölaskennan kustannukset ovat vain murto-osa suoran tiedonkeruun kustannuksista. Osa väestölaskennan tiedoista päätellään eri rekisteritietojen avulla, vaikka niitä ei suoraan missään rekisterissä esiintyisikään, kuten pääasiainen toiminta laskentaviikolla, perheet, asuntokunnat, perheen viitehenkilö, sosio-ekonominen asema sekä työpaikan ja asuinpaikan välinen etäisyys.

Useat tilastot tuottavat niin kutsuttuja kantatilastoja esimerkiksi alueiden väestöstä, asunnoista, rakennuksista, yrityskannasta ja tuottavat näiden lisäksi muutostietoja. Tällaisia muutostietoja ovat esimerkiksi vuosittain syntyneet, kuolleet, maassamuuttaneet, vihityt, eronneet, maahan- ja maastamuuttaneet, valmistuneet rakennukset ja asunnot, rekisteröidyt ajoneuvot, aloittaneet ja lopettaneet yritykset. Kun kantatiedot ovat rekisterissä niin tilastojen tuottamiseksi tarvitsee käsitellä vain niitä yksiköitä, joille on tapahtunut muutoksia. Esimerkiksi rakenusrekisterissä käsitellään uudet rakennukset, mutta muille rakennuksille ei tarvitse tehdä toimenpiteitä. Osa rekistereistä kerää uuttakin tietoa lähes kaikille yksiköille vuosittain, esimerkiksi verotuksen rekisterit, yritys- ja toimipaikkarekisteri, työeläkerrekisterit. Tällaiset rekisterit eivät eroa paljoa vuositiedostoista. Osa rekistereistä onkin oikeastaan vuositiedostoja, koska kaikki rekisterin yksiköt ja yksiköiden ominaisuustiedot kerätään joka vuosi uudelleen. Tällaisia rekisteritietoja ovat esimerkiksi asumis- ja opintotukea saaneet, toimeentulotukea saaneet sekä hoitoilmoitusrekisteri, joka koostuu vuoden aikana laitoshoitossa olleista potilaista. Myös opiskelijarekisteri sisältää kunakin vuonna oppilaitoksissa kirjoilla olevat. Opiskelijarekisteristä tehtäviä muutostilastoja ovat pyrkineet, valitut ja aloittaneet sekä tutkinnon suorittaneet.

Yleensä viranomaistyönä tehdyt ilmoitukset sisältävät tarkan ajankohdan. Henkilöiden omiin ilmoituksiin perustuvat muutostiedot ovat jossain määrin subjektiivisia: esimerkiksi avoliitosta voi yhdessä asuvilla olla erilainen käsitys;

työtä tekevä opiskelija ilmoittaa itsensä opiskelijaksi, vaikka työllisen määritelmä täyttyy (palkattua työtä tehdään vähintään yksi tunti viikossa, ja työvoimakäsittteen mukaan hänen tulisi ilmoittaa itsensä iltasiivoojaksi). Rekisteritietoja käytettäessä vastaajien subjektiivisten käsitysten vaikutus pienenee. Esimerkiksi kaikki päätoimiset opiskelijat, joilla on työsuhde, luokitellaan työllisiksi, ja kotitalouden viitehenkilöksi määritetään yleensä sen suurituloisin jäsen.

Tietojen oikeellisuutta, puuttuvuutta ja loogisuutta tulee tarkastella myös hallinnollisissa aineistoissa ja rekistereissä. Tämä on haasteellista tilastontuottajalle erityisesti silloin, kun tietojärjestelmiä ylläpitää eri organisaatio. Tämän vuoksi aineistojen tarkistuksissa tulee tehokkaasti hyödyntää käytettävissä olevia tietolähteitä sekä systemaattisesti kehitettyjä hierarkisia päättelysääntöjä. Esimerkiksi elinkeinoverotusta koskevat tiedot saattavat joiltakin yrityksiltä puuttua kokonaan tai ne voivat olla virheellisiä. Tietojen oikeellisuutta tulee testata tarkastamalla yritysten tilien umpeutumista. Epäkelpojen tilien joukosta voidaan tämän jälkeen etsiä yleisiä virhetyyppejä, jotka voidaan korjata ohjelmoitavien loogisuus- ja päättelysääntöjen avulla (ks. luku 2.10). Niiltä osin kuin elinkeinoverotuksen tilinpäätöstiedot puuttuvat tai ne eivät umpeudu, ne joudutaan imputoimaan suhteessa yrityksen liikevaihtoon. Kaikkien yritysten liikevaihtotieto on saatavissa yritysrekisteristä. Imputointia voidaan parantaa huomioimalla esimerkiksi toimialan ja yrityksen suuruusluokka entistä kehittyneemmällä menetelmällä. Toisaalta kun yhdistetään työpaikka-, ammatti- ja tulotiedot eri rekistereistä, ei voida aina olla varmoja, että ammatti ja toimiala kuvaavat samaa työsuhdetta. Tällöin voi syntyä niin sanottu konsistenssiongelma.

Nopeimmat hallinnolliset aineistot ja rekisterit ovat käytettävissä muutaman viikon tai parin kuukauden kuluttua viiteajankohdasta (ajoneuvorekisteri, työnhakijarekisteri, väestötietojärjestelmä ja verottajan maksuvalvonta-aineisto), kun taas hitaimpien valmistumiseen kuluu jopa vuosi (esimerkiksi verotustietokanta). Ajantasaisuuteen vaikuttaa se, kuinka kauan kuluu aikaa ennen kuin tapahtuma rekisteröidään ja se, kuinka usein tietoja päivitetään rekisteriin. Väestötietoja päivitetään joka päivä, tulotietoja kerran vuodessa. Mikäli muuttaja laiminlyö muuttoilmoituksen antamisen, rekisterin tiedot eivät päivity. Rakennuksen valmistumisajankohta, eli rakennuksen ottaminen rekisteriin, on myös ongelmallinen käsite: rakennuksessa on saatettu jopa asua jo kauan, ennen kuin se loppukatselmuksessa todetaan valmistuneeksi rakennukseksi.

Ajantasaisuus vaikuttaa myös tietojen kattavuuteen. Esimerkiksi maksuvalvonta-aineistojen saapussa ensimmäistä kertaa tilastoviranomaiselle kattavuus on noin 80 prosenttia yritysten liikevaihdon ja 75 prosenttia säännöllisten työntekijänsä maksaman palkkasumman osalta lopullisesta määrästä. Yksittäisen yrityksen kohdalla on ongelmallista tietää, puuttuuko ilmoitus tiedonantoviipeen vai yrityksen toiminnan loppumisen vuoksi. 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.

Vuosittain laadittavat yritystoiminnan rakennetilastot laaditaan yhdistämällä suoran tiedustelun kautta saatuihin tietoihin elinkeinoverotuksen ja yritysrekisterin tietoja, missä suora tiedonkeruu kohdistetaan vain muutamalle prosentille yrityksiä.

Kuukausittain laadittavat liikevaihto- ja palkkasummaindikaattorit puolestaan laaditaan käyttämällä suoran tiedustelun kautta saatujen kuukausihavaintojen lisäksi arvonlisävero- ja työnantajasuoritusrekisterin tietoja. Suora tiedonkeruu kohdistetaan tällöin vain vajaalle yhdelle prosentille yrityksistä. Yritystoiminnan rakennetilastojen ja kuukausi-indikaattoreiden tiedonkeruun samankaltaisuudesta huolimatta metodologiset haasteet näissä tilastoissa ovat erilaisia. Rakennetilastoissa imputoinnilla on keskeinen metodologinen asema: hallinnollisen aineiston puutteet korjataan siten, että ne kelpaavat tilastollisen käyttöön. Kuukausi-indikaattoreissa käytetään panee-limenetelmiä ratkaisemaan ajantasaisuuden ja alipeiton ongelmia. Pitkäjänteisellä metodologisella työllä on pystytty parantamaan edelleen näiden tilastojen laatua.

Suosituks

Hallinnollisten aineistojen ja rekisterien hyväksikäytön edellytykset

Hallinnollisten aineistojen ja rekisterin on täytettävä tietyt edellytykset, jotta niitä voidaan käyttää tilastotuotannossa. Kaikki yksittäiset edellytykset eivät aina ole ehdottoman välttämättömiä, vaan niiden tarpeellisuutta tulee harkita muiden edellytysten ja vaihtoehtoisten tiedonkeruumenetelmien kannalta. Hallinnolliset aineistot sisältävät laadullisia haasteita tilastojen tekijöille. Mutta pitkäjänteisellä metodologisella kehittämistyöllä sekä tietojärjestelmien ylläpitäjien tiiviillä yhteistyöllä näihin haasteisiin pystytään vastaamaan.

Sisällölliset ja käsitteelliset edellytykset

- *Kattavuus*: Rekisterin tulee kattaa koko kohdejoukkonsa (ks. luku 1.2.3.).
- *Yksikkökohtaisuus*: Rekisterin perustietojen tulee olla yksikkökohtaisia. Oman haasteensa tuovat yritysten fuusiot ja diffuusiot. Niiden hallinnollinen käsittely ei välttämättä vastaa tilastotoimen vaatimuksia.
- *Vastaavuus käyttäjien tarpeisiin*: Tilastotuotannossa hyödynnettävien rekisterien tulee olla relevantteja tilastollisen käytön kannalta. Rekisteritiedoissa esiintyvien käsitteiden on oltava riittävän yhtenäisiä tilastotuotannossa käytettyjen käsitteiden kanssa (ks. luku 1.2.3). Usein hallinnollisten aineistojen tietoihin on suhtauduttava kuten korvikemuuttujiin. Tällöin tiedot on muunnettava teoriapohjalta vastaamaan tilastotietojen tarpeita.

Luotettavuus

- *Peittävyys*: Rekisterin tulee kattaa kohdejoukkonsa mahdollisimman hyvin (ks. luku 2.4). 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. Väestölaskentojen kattavuus vaihtelee maittain 95–100 prosentin välillä. Jos laskennassa käytettyjen rekisterien peittävyys on tätä heikompi, on käyttö ongelmallista.
- *Tietojen luotettavuus*: Yksittäisten tietojen tulee olla mahdollisimman luotettavia (ks. luku 1.2.3). Tämä koskee sekä tunnus- että ominaisuustietoja.
- *Tietojen oikeellisuus*: Rekisterin päivityssysteemin tulee olla kattava. Esimerkiksi väestötilastosysteemi saa varsin luotettavasti tiedot syntyneistä, kuolleista, vihityistä ja eronneista, mutta muuttoilmoituksia laiminlyödään. Muita epätarkkuuksia aiheutuu siitä, että ammattitieto päivitetään vain muuttoil-

moituksen yhteydessä, ja että asunnoissa tehdään merkittäviä muutoksia ilman rakennuslupia. Myös avioliittoon rinnastettavat avoliitot joudutaan tilastollisesti päättelemään.

Ajantasaisuus

- Tietojen tulee olla ajan tasalla (ks. luku 1.2.3). Ajantasaisuuteen vaikuttaa se, kuinka kauan kuluu aikaa ennen kuin tapahtuma rekisteröidään ja se, kuinka usein tietoja päivitetään rekisteriin.
- Tilastotuotannossa on kehitettävä yhteistyötä eri tilastoviranomaisten välille ja menetelmiä, joilla puutteet ajantasaisuudessa voidaan hallita. Koska hallinnolliset aineistot määritelmällisesti kerätään muuta tarkoitusta kuin tilastotuotantoa varten, tilastoviranomainen ei voi suoraan vaikuttaa ajantasaisuuteen liittyviin ongelmiin.

Tekniset edellytykset

- *Dokumentointi*: Hallinnollisista aineistoista tulee olla riittävän hyvät ja yksityiskohtaiset kuvaukset, joiden perusteella käyttäjä saa kuvan aineiston toiminnasta, rakenteesta ja tietosisällöstä (ks. luvut 3.1 ja 3.2).
- *Tietojenkäsittely*: Hallinnollisen aineiston tulee olla elektronisessa muodossa, jotta sitä voidaan käyttää hyväksi tilastotuotannossa. Kortistomuotoisia manuaalisesti käsiteltäviä aineistoja voivat olla esimerkiksi yritysten työntekijärekisterit ja yhdistysten jäsenluettelot. Niillä on harvoin tilastollista käyttöä.
- *Tunnukset*: Hallinnollisten aineistojen yksiköillä on tärkeää olla yleisesti käytössä olevat tunnukset.
- *Luokiteltavuus*: Ominaisuustietojen tulee olla kooditettuina tai numeerisessa muodossa, jotta niitä voidaan luokitella. Jos tieto on selväkielinen, kuten esimerkiksi ammattinimike, se voidaan koodittaa automaattisesti. Osa selväkielistäkin nimikkeistä vaatii manuaalista käsittelyä (ks. luku 2.3).

Hallinnolliset edellytykset

Hallinnollisen aineiston ja/tai rekisterin ylläpitäjän tulee luovuttaa aineisto Tilastokeskukselle. Tilastolain (280/2004) mukaan Tilastokeskuksella on oikeus saada viranomaisilta ja useilta muilta yhteisöiltä niiden hallussa olevat tiedot tilaston laatimista varten (ks. luku 1.1.1).

Kustannukset

Tilastolain mukaan hankittaessa tietoja tilastojen laatimista varten tulee ensi sijassa käyttää hyväksi julkishallinnon tehtävien hoitamisessa kertyneitä sekä elinkeinon- ja ammatinharjoittajien, yhteisöjen ja säätiöiden tavanomaisen toiminnan seurauksena syntyneitä tietoja. Hallinnollisten aineistojen hyväksikäytön kustannukset tilastotuotannolle ei toisaalta tule olla suuremmat kuin toisen vaihtoehdoisen tiedonkeruutavan kustannukset. Tilastokeskus korvaa hallinnollisten aineistojen pitäjille yleensä vain tietojen hyväksikäytöstä aiheutuneet lisäkustannukset.

Kirjallisuus

Suomen lait ja asetukset sekä kansalliset toimintaohjeet

Tilastolaki (280/2004).

<http://www.finlex.fi/fi/laki/ajantasa/2004/20040280>

Lisätietoja

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. Luxembourg.

<http://amrads.jrc.it/WPs%20pages/Quality/Documents/eqpbvol1.pdf>

Harala, R. & Tammilehto-Luode, M. 1999. GIS and Register-based Population Census. Teoksessa: Alho, J. (toim.). *Statistics, Registries and Science. Experiences from Finland*. Tilastokeskus. 55–73.

Ilmakunnas, P., Laaksonen, S. & Maliranta, M. 1999. Enterprise Demography and Job Flows. Teoksessa: Alho J. (toim.). *Statistics, Registries and Science. Experiences from Finland*. Tilastokeskus. 73–89.

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.

<http://www.jos.nu>

Myrskylä, P. 1996. Virallisten tilastojen tiedonkeruu. Teoksessa: Niemi, H. & Tourunen, K. (toim.). *Tilastoista tiedoksi korkea-asteelle*. Tilastokeskus. Helsinki.

Myrskylä, P. 1999. New statistics made available by the use of registers. *Statistical Journal of the United Nations ECE*. Vol. 16. No. 2–3. 165–180.

Myrskylä, P. (toim.) 2004. *Use of Registers and Administrative Data Sources for Statistical Purposes – Best Practices of Statistics Finland*. Handbooks 45. Helsinki: Statistics Finland.

Ragnarson, R. 2000. *Utilisation of Administrative Data in the Manufacturing Statistics. The Norwegian Experience*. Meeting of INSEE's and Nordic NSI's Business Statisticians. Helsinki 11–12.5.2000.

Tilastokeskus. 2001. *Väestölaskenta 2000*. Käsikirjoja 35. Helsinki: Tilastokeskus.

2.6 Otantamenetelmät

Soveltamisala ja päämäärä

Otantatutkimuksia käytetään yhteiskunnallisissa tilastotutkimuksissa kokonaisuutkimuksen asemesta useista syistä. Koska kaikkia yksiköitä ei ole välttämätöntä tutkia, kustannuksia voidaan säästää, tutkimuksen kohteena olevasta ilmiöstä voidaan kerätä hyvin yksityiskohtaisia tietoja ja tulokset voidaan tuottaa suhteellisen nopeasti. Niitä käytetään myös täydentämään rekistereistä ja muista hallinnollisista lähteistä saatavaa tietoa. Oikein tehtynä otannalla saadaan käyttötarkoituksen kannalta riittävän tarkat estimaatit ja samalla voidaan pienentää yhteiskunnallista vastausrasitetta. (Kish, 1965; Särndal et al., 1992)

Nykyaikaiset otantamenetelmät perustuvat satunnaistamiseen ja niitä kutsutaan todennäköisyysotannan menetelmiksi. Otantamenetelmien etuna on yleislettävyys perusjoukkoon sekä mahdollisuus laskea estimaattien luotettavuutta kuvaavia tunnuslukuja. Otantamenetelmän valintaan vaikuttavat muun muassa tutkittava ilmiö, tutkimusasetelma, perusjoukon ominaisuudet ja käytettävissä olevat otantakehikot, tiedonkeruutapa ja kustannukset.

Periaatteet

Otantamenetelmiä voidaan jaotella usealla tavalla. Ensimmäinen jaottelu perustuu siihen, poimitaanko suoraan tutkittavia alkioita vai niiden yhdistelmiä, joita kutsutaan rypäiksi. Toisena valintakriteerinä on otantakehikosta tai muusta lähteestä saatavan lisäinformaation hyödyntäminen otanta-asetelmassa. Valinta voi perustua myös kustannusten ja tilastollisen tarkkuuden välillä tehtävään vertailuun.

Alkiotason otantamenetelmät ovat yleensä tilastotieteellisesti tehokkaita ja suhteellisen helposti toteutettavissa varsinkin rekisteripohjaisista otantakehikoista. Ryväotannon menetelmät eivät ole yleensä yhtä tehokkaita tilastolliselta kannalta, mutta usein kustannuksiltaan pienempiä kuin alkiotason menetelmät. Seuraavaksi esitetään tyypillisiä otantamenetelmiä, joita voidaan soveltaa niin alkioiden kuin rypäiden poiminnassa. Sen jälkeen käsitellään kehikoihin liittyviä kysymyksiä: osittamista ja ryvästystä.

Yksinkertainen satunnaisotanta

Otannon perusmenetelmä on yksinkertainen satunnaisotanta (*simple random sampling, SRS*), jossa jokaisella perusjoukon yksiköllä eli alkiolla on poimintahetkellä yhtä suuri todennäköisyys tulla poimituksi otokseen. Otanta perustuu kullekin yksikölle liitettyyn satunnaislukuun ja mitään lisäinformaatiota ei käytetä hyödyksi. (Lohr, 2003). Yksinkertaista satunnaisotantaa käytetään henkilöihin ja kotitalouksiin kohdistuvissa kysely- ja haastattelututkimuksissa. Yksinkertainen satunnaisotanta on myös vertailuasetelma eri otanta-asetelmien tilastollisen tehokkuuden arvioinnissa. (Kish, 1965).

Systemaattinen otanta

Systemaattisessa otannassa (*systematic sampling, SYS*) otosalkiot poimitaan tasavälein läpi koko otantakehikon. Menetelmä on kehitetty otospoimintaan kortistosta, mutta soveltuu hyvin myös tietokoneella tehtäviin poimintoihin.

Systemaattista otantaa käytettäessä ainoa lisäinformaatio sisältyy perusjoukon lajittelujärjestykseen. Kun perusjoukko on satunnaisjärjestyksessä, saadaan yksinkertaista satunnaisotantaa vastaava otos. Kehikko voidaan myös lajitella jonkin apumuuttujan suhteen, jolloin voidaan saada tarkempia tuloksia kuin käyttämällä satunnaisjärjestystä. Tällaista menettelyä kutsutaan implisiittiseksi ositukseksi. Kun esimerkiksi tutkitaan alueellisesti riippuvaista ilmiötä Suomen väestössä, saadaan alueellisesti oikein jakautunut otos käyttämällä otantakehikona väestötietojärjestelmän hakemistoa lajiteltuna tarkan osoitetiedon sisältävän kotipaikkatunnuksen mukaan.

Jos perusjoukossa esiintyy jaksollisuutta, systemaattinen otos saattaa muodostua samankaltaisista ja muusta joukosta poikkeavista alkioista. Tällöin menetelmä voi tuottaa perusjoukkoon nähden vinon otoksen, joka voi edelleen aiheuttaa harhaisia tuloksia.

PPS-otanta

Jos otos poimitaan vinosta populaatiosta, jossa yksiköiden koko vaihtelee paljon, voidaan otoksen tarkkuutta parantaa antamalla kullekin yksikölle sisällymisdennäköisyys, joka on suhteessa yksikön kokoon. Tällöin voidaan käyttää

PPS-otantaa (*probabilities proportional to size*). (Särndal et al., 1992; Lehtonen ja Pahkinen, 2004).

PPS-otanta edellyttää, että otoskehikossa kaikilla alkioilla on jokin kokoa mittaava muuttuja, jonka arvojen perusteella poimintatodennäköisyys määräytyy. Suurimpien alkoiden kokoa mittaavan muuttujan arvo voi joskus tulla niin suureksi, että alkio valitaan otokseen todennäköisyydellä yksi. Tällaiset alkiot valitaan suoraan ja kehikon muiden alkoiden poimintatodennäköisyydet pitää laskea operaation jälkeen uudelleen.

PPS-otokset ovat tilastotieteellisesti hyvin tehokkaita. Niiden haittapuolena on se, että ne eivät useinkaan sovellu muiden kuin poimintakriteerinä käytettyjen tai niiden kanssa vahvasti korreloivien tutkimusmuuttujien analysointiin.

Otoskehikon jakaminen osajoukkoihin

Otoskehikko voidaan jakaa sopivaa lisäinformaatiota käyttäen toisensa poissulkeviin osajoukkoihin kahdesta syystä. Jos osajoukkoihin jakaminen tehdään niin, että kukin osajoukko toimii itsenäisenä kehikkona otospoiminnassa, puhutaan osittamisesta. Jos taas osajoukot ovat ”ylemmän tason yksiköitä”, joista varsinainen otos poimitaan, mutta jonka alkioita varsinaisesti tutkitaan, kyse on ryväotannasta.

Ryväsotanta

Joskus on tarkoituksenmukaista hyödyntää perusjoukon ryvästymistä otosta poimittaessa. Rypäällä tarkoitetaan perusjoukon luonnollista osajoukkoa, kuten kotaloutta, toimipaikkaa, koulua tai vastaavaa yksikköä, joka sisältää tutkittavat alkiot. Poimituista rypäistä voidaan tutkia kaikki alkiot tai niiden sisältä voidaan poimia edelleen otos tutkittavista alkioista. Edellisessä tapauksessa on kyse yksiasteisesta (*one-stage*) ryväotannasta (*cluster sampling*) ja jälkimmäisessä kaksiasteisesta ryväotannasta. Esimerkkinä ryväotannasta on henkilöiden työoloja koskeva tutkimus, jossa otokseen poimitaan yrityksiä, joiden työntekijöitä tutkitaan. Yritys muodostaa rypään, ja tutkimusyksikkönä on työntekijä. Jos kaikki poimituihin yrityksiin kuuluvat henkilöt tutkitaan, on kyseessä yksiasteinen ryväotanta. Jos poimitujen yritysten sisällä tutkimukseen poimitaan vain osa kyseisten yritysten työntekijöistä, on kyseessä kaksiasteinen ryväotanta.

Ryväsotanta pienentää tiedonkeruun kustannuksia, koska samasta poimintayksiköstä voidaan kerätä usean alkion tiedot. Sen haittapuolena on usein otosvarianssin kasvaminen suhteessa yksinkertaiseen satunnaisotantaan. Tehokkuuden heikkeneminen on seurausta rypäeseen kuuluvien alkoiden samankaltaisuudesta eli sisäkorrelaatiosta. Ryväotos ei siten välttämättä sisällä yhtä paljon vaihtelua kuin yksinkertaisella satunnaisotannalla poimittu otos.

Ositettu otanta

Ositetussa otannassa (*stratified sampling*) perusjoukko jaetaan käytettävissä olevien apumuuttujien mukaan homogeenisiin osiin eli ositteisiin. Vinosti jakautuneissa perusjoukoissa ositus on suositeltavaa, jotta myös suurimmista yksiköistä saataisiin täsmällisiä tietoja – vaihtoehtoisesti voidaan käyttää PPS-otantaa. Perusjoukon ositus edellyttää, että kaikista perusjoukon yksiköistä on käytettävissä lisäinformaatiota, jonka perusteella ositus voidaan tehdä. Ositettaessa perusjoukkoa kukin 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. Väestöön kohdistuvissa tutkimuksissa tavallisia ositusmuuttujia ovat esimerkiksi asuinalue, henkilön ikä ja sukupuoli.

Ositekohtaiset otoskoot joudutaan kiintiöimään erikseen. Mikäli ositteiden otoskoot ovat yhtä suuret, puhutaan tasakiintiöinnistä. Se soveltuu erityisesti ositteiden välisten vertailujen tekemiseen. Jos ositekohtaiset otoskoot ovat suorassa suhteessa niiden osuuteen koko perusjoukosta, on kyse suhteellisesta kiintiöinnistä. Tämä menetelmä tehostaa hieman tavanomaista satunnaisotantaa. Voimakkaasti vinojen perusjoukkojen tapauksessa voidaan hyödyntää joko optimaalista kiintiöintiä tai potenssiikiintiöintiä. Tällöin otos kohdistuu hyvin vahvasti suurimpiin (tai muuten tutkimuksen kannalta tärkeisiin) yksiköihin (Lehtonen ja Pahkinen, 2004). Kumpikin menetelmä on tilastotieteellisesti hyvin tehokas. Ääritapauksissa kaikkein suurimmat yksiköt valitaan todennäköisyydellä yksi, jolloin niitä kutsutaan itse-edustaviksi ositteiksi. Ilman niiden sisältämää informaatiota tulosten estimointi kävisi epäluotettavaksi.

Monimutkaiset otanta-asetelmat

Monimutkaiset otanta-asetelmat (*complex sampling designs*) koostuvat eri menetelmien yhdistämisestä. Tyypillisesti kyseessä on moniasteinen ositettu ryväotanta, jossa poiminta kohdistuu laajempiin alue- tai muihin luokitteleviin yksiköihin ja jossa hyödynnetään PPS-otantaa. Varsinaisten tutkimusyksiköiden poiminta tapahtuu miltei poikkeuksetta yksinkertaista satunnaisotantaa tai systemaattista otantaa hyödyntäen. Yhdistämällä ositus ja ryvästäminen ja optimoimalla eri otosvaiheista saatavilla oleva informaatio voidaan päästä suhteellisen kustannustehokkaihin otanta-asetelmiin.

Rekisteripohjaisten perusjoukkojen tapauksessa voidaan hyödyntää myös kaksi- tai useampivaiheista otantaa (*multi-phase sampling*). (Särndal et al., 1992). Silloin poimitaan aluksi suuri perusotos (master-otos), johon yhdistetään muista lähteistä lisäinformaatiota. Tätä muodostettua perusotosta käytetään kehikkona varsinaisen tutkimusotoksen poimimiseen. Monivaiheisia otoksia käytämällä voidaan hyödyntää tehokkaasti lisäinformaatiota, kohdistaa tutkimus tarkasti sekä kontrolloida eri tekijöiden vaikutusta estimointiin.

Kiintiöpoimintamenetelmät

Kiintiöpoiminta tarkoittaa menetelmiä, jotka eivät ole varsinaisia todennäköisyysotannon menetelmiä. Niitä sovelletaan usein markkinatutkimuksissa hahmottaessa ennalta määrätty määrä tietyntyyppisiä vastaajia. Tällöin kiintiötä täytettäessä poimitaan vastaajia kuhunkin kiintiöön kunnes on saatu ennalta asetettu määrä kunkin kiintiöidyn ominaisuuden mukaisia vastaajia. Kiintiöpoiminta on altis valikoitumisharhalle, koska yksiköiden tavoitettavuus ja osallistumispiä-tös voi olla yhteydessä mielenkiintoon tutkimuksen aihepiiriä kohtaan.

Yritysoannoissa sovelletaan melko usein katkaisuoatantaa (*cut-off sampling*). Katkaisuoatoksissa määritellään kynnsarvo, useimmiten alaraja, jota pienemmät alkiot poistetaan kehikosta eivätkä ne siten kuulu otannon piiriin. Katkaisuoatanta on kuitenkin ongelmallinen, koska kynnsarvo aiheuttaa aina alipeittoa tuloksiin. Vanhentuneen kehikon tapauksessa perusjoukon muutokset voivat olla

huomattavia, jolloin katkaisuoitanta aiheuttaa selkeästi harhaisia tuloksia. Katkaisuoitoksia käytettäessä tulisi selkeästi kuvata menetelmän vaikutus tilaston kattavuuteen ja yleistettävyyteen sekä perustella kynnsarvon valinta.

Otoskoon määrittäminen

Tarvittavan otoksen koon määrittämisessä tulee ottaa huomioon lopullisille estimaateille asetetut tarkkuusvaatimukset, julkistettaville tiedoille käytettävät perusjoukon osajoukkojen luokitustasot ja perusjoukon heterogeenisuus. Myös käytettävissä olevat taloudelliset resurssit vaikuttavat otoskokoon. Yleensä tarkkuus paranee otoksen suuretessa, koska estimaattien keskivirheet pienenevät. Otoksen kasvaessa otantavirhe ei kuitenkaan pienene otoksen koon suhteessa, vaan hitaammin.

Luotettavuus

Otosten luotettavuutta mitataan ensisijaisesti tilastollisten laatuindikaattorien avulla. Otosvarianssi kuvaa ensisijaisesti otantamenetelmästä ja valitusta otoskoosta aiheutuvaa piste-estimaattien satunnaisvaihtelua. Otosvarianssi tulee aina laskea asianmukaisen, käytetylle otantamenetelmälle johdetun, otosvarianssin laskentakaavan avulla. Otosvarianssin avulla saadaan joukko yleisesti käytettyjä tilastollisen laadun tunnuslukuja, kuten keskivirhe, variaatiokerroin ja luottamusväli eli virhemarginaali. Luotettavuusmittojen avulla voidaan vetää johtopäätöksiä eri päätöstilanteissa sekä päättää myös tulosten mahdollisista julkistamiskriteereistä.

Puhtaasti tilastollisten luotettavuusindikaattorien ohella on syytä kiinnittää huomiota myös muihin otantatutkimuksissa ilmeneviin virhelähteisiin. Niitä ovat muiden muassa kehikoissa esiintyvät peittovirheet, vastauskato sekä mahdolliset mittaus- ja käsittelyvirheet. Laatuindikaattorien esittämistä käsitellään muun muassa Suomen virallisen tilaston laatuksiteerien yhteydessä luvussa 1.2.3.

Otosten koordinointi

Säännöllisesti toistuvissa tutkimuksissa on suotavaa koordinoita peräkkäisiä otoksia. Positiivinen koordinointi merkitsee sitä, että otoksiin pyritään saamaan mahdollisimman paljon samoja yksiköitä. Tämä toteutuu erilaisissa paneelitutkimuksissa, joissa muutoksen estimointi tapahtuu luotettavimmin juuri samoista alkioista. Negatiivisessa koordinoinnissa pyritään välttämään samojen alkioiden valintaa vastausrasitteen vähentämiseksi. Koordinointi on tehtävä niin, ettei siitä aiheudu harhaa poimittuun otokseen. Jos tämä tavoite ei toteudu, on estimoinnissa käytettävä sopivia oikaisumenetelmiä.

Yritystutkimuksissa ollaan usein kiinnostuneita perusjoukossa tapahtuvista muutoksista, jolloin on välttämätöntä saada riittävä määrä samoja yrityksiä perättäisiin otoksiin. Jottei tiedonantajia rasitettaisi liikaa, 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 tulee koordinoita siten, ettei yhden vastaajan rasite nousisi suuremmaksi kuin muiden. Koordinointia ei kuitenkaan voida ulottaa suurimpiin alkioihin, jotka poimitaan aina kaikkiin otoksiin peittävyuden turvaamiseksi. (Teikari, 2001).

Suosituksset

- Käytettävien otantamenetelmien tulee perustua todennäköisyysotannan menetelmiin. Yksinkertainen satunnaisotanta tai systemaattinen otanta on suositeltava perusmenetelmä monia kiinnostuksen kohteita sisältävissä tilastotutkimuksissa, joissa otosta ei voi kohdentaa perusjoukon osajoukkoihin.
- Rekisteristä poimittaessa käytännöllinen poimintatapa on systemaattinen otanta. Tällöin otantakehikon lajittelulla saadaan otokseen likimain sama jakauma kuin kehikossa.
- Ositettua otantaa on syytä käyttää haluttaessa varmistua tutkimuksen kannalta tärkeiden osajoukkojen sisällymisestä poimittavaan otokseen.
- Yritysjakauman vinoudesta johtuen yksinkertainen satunnaisotanta ei ole sopiva menettely yritysotoksissa. Niissä yleisimmin käytetty menetelmä on koon mukaan ositettu otanta. Päälinjana on ositus toimialan ja yrityskoon mukaan. Lisäksi käytetään menetelmiä, joissa sisällymistodennäköisyys määräytyy suoraan yksikön koon perusteella. Suurimmat yksiköt yritysotoksissa määrätään yleensä ositteisiin, jotka sisältyvät otokseen todennäköisyydellä yksi.
- Monimutkaisten otanta-asetelmien käytössä on punnittava tarkasti menetelmän tuottamat hyödyt suhteessa haittoihin ja työläyteen. Kaksi- tai useampi-vaiheiset otantamenetelmät ovat tehokkaita sekä kustannusten että tilastollisen tarkkuuden suhteen. Sen sijaan moniasteiset ryväsotantamenetelmät voivat olla tilastotieteellisesti hyvin tehottomia.
- Säännöllisesti toistuvissa tutkimuksissa peräkkäiset otokset on koordinoitava vastausrasitteen vähentämiseksi. Tällä on merkitystä erityisesti paneeli- ja yritystutkimuksissa. Henkilöitä ja kotitalouksia koskeissa otantatutkimuksissa tutkimuksiin osallistuneet yksiköt vapautetaan määrääjäksi.
- Otannasta johtuvaa virhettä ja satunnaisvaihtelua on mitattava, analysoitava ja dokumentoitava järjestelmällisesti. Keskeisten estimaattien julkistamisen yhteydessä on raportoitava niiden keskivirheet, otanta-asetelman tehokkuutta estimoivat tunnusluvut (*deff*), vastaus- tai kato-osuus ja muut soveltuvat laatuindikaattorit.
- Otosaineistoihin on liitettävä otospainot. Painottamista käsitellään tarkemmin luvussa 2.11. Otosaineistoihin perustuvaa tilastollista estimointia ja analyysia käsitellään luvussa 2.12.

Dokumentointi

Raportoinnissa keskeistä on perusjoukon määrittely, tehdyt rajaukset ja niiden perustelut, otoskehikon sekä valitun otanta-asetelman tarkka kuvaus. Laatuselosteissa on selvitettävä tärkeimpien estimaattoreiden kaavat, aineistoon si-

sältyvät vaihtelulähteet ja virheet, erityisesti katovirhe, ja niistä aiheutuneet korjaustarpeet painotuksessa. Laatu- ja menetelmäselosteiden menetelmäkuvauksiin tulee sisältyä perusasetelman mukaisten estimaattoreiden otosvarianssien lausekkeet sekä vastauskadon ja muiden mahdollisten virhelähteiden vaikutus niihin, jotta keskivirheet ja muut keskeiset laatuindikaattorit voidaan laskea mahdollisimman tarkasti.

Kirjallisuus

- Kish, L. 1965. *Survey Sampling*. New York: John Wiley & Sons.
- Lehtonen, R. & Pahkinen, E. 2004. *Practical Methods for Design and Analysis of Complex Surveys*. 2nd ed. Chichester: John Wiley & Sons.
- Lohr, S. 1999. *Sampling: Design and analysis*. Pacific Grove: Duxbury 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.

Lisätietoja

- Cochran, W. G. 1977. *Sampling Techniques*. 3rd ed. New York: John Wiley & Sons.
- Groves, Robert M., Fowler, F.J., Couper, M.P., Lepkowski, J.M., Singer, E., & Tourangeau, R. 2004. *Survey Methodology*. New York: John Wiley & Sons.
- Pahkinen, E. & Lehtonen, R. 1989. *Otanta-asetelmat ja tilastollinen analyysi*. Helsinki: Gaudeamus.

2.7 Tiedonkeruutavat

Soveltamisala ja päämäärä

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

Tiedonkeruutavan valinta ei ole muista tekijöistä irrallinen toimenpide, vaan se on yhteydessä koko tilastotutkimuksen prosessiin, kuten tutkimusasetelmaan, tutkimuskohteiden valintaan eli otantaan, kadon suuruuteen ja kustannuksiin. Tiedonkeruutapa vaikuttaa myös tutkimuksen aikatauluun ja tulosten laatuun sekä tutkimuslomakkeen suunnitteluun ja laadintaan. Eri tiedonkeruutavat mahdollistavat varsin erilaisen otoskoon. Tietojen yleistettävyyden perusjoukon osajoukkoihin puolestaan riippuu siitä, miten suurelta joukolta tutkittavia tietoja on kerätty (ks. luku 2.6).

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

Tutkimukset voidaan jakaa pitkittäistutkimuksiin ja poikittaistutkimuksiin. Pitkittäistutkimuksille on ominaista, että samoilta tutkimusyksiköiltä kerätään tietoa useamman kerran tietyin väliajoin. Sitä vastoin poikittaistutkimuksissa tietoa kerätään vain kerran. Pitkittäistutkimuksissa tiedonkeruu voidaan toteuttaa jokaisella kerralla samalla tavalla tai eri kerroilla eri tavoilla. Tietyntyyppisiä pitkittäistutkimuksia kutsutaan paneelitutkimuksiksi. Työvoimatutkimus on Tilastokeskuksen tunnetuin henkilöpohjainen pitkittäistutkimus ja kuluttajahinnojen keruu tunnetuin yrityspohjainen pitkittäistutkimus.

On tärkeää erottaa toisalta tietojen keruutapa ja toisalta menetelmä, jolla kerättävä tieto kirjataan. Tietojenkeruutapa määrittää kerätäänkö tieto haastatteleamalla vai vastaavtko tutkittavat itse. Tietojen keruumenetelmä puolestaan määrittää miten kerättävä tieto kirjataan, eli käytetäänkö paperilomakkeita vai sähköistä tiedonkeruulinstrumenttia.

Tiedonkeruutavat voidaan jakaa kyselyihin, joissa vastaaja itse täyttää lomakkeen, ja haastatteluihin, joissa haastattelija esittää kysymykset ja merkitsee vastaukset. Haastatteluissa erotellaan vielä puhelin- ja käyntihaastattelut. Sekä kyselyissä että haastatteluissa tiedot voidaan taltioida joko paperilomakkeelle tai jollekin sähköiselle tietovälineelle (ks. taulukko 2.1).

Taulukko 2.1 Tutkimustyyppitiedonkeruutavan ja instrumentin mukaan

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

Periaatteet

Tutkimuksen tavoitteiden tulisi ohjata tiedonkeruutavan valintaa eikä päinvastoin. Valintakriteereinä tulisi olla erilaisten menetelmien seuraamusvaikutukset. Tutkimuksen suunnittelun lähtökohtana tulee olla lopullisten estimaattien tarkkuusvaatimukset ja arvio otoksen tai vastaajajoukon suuruudesta. Yleisperiaatteena on, että suuremman otoksen tulokset ovat luotettavampia. Toinen merkittävä tekijä on vastausyksiköiltä kerättävien tietojen määrä ja laatu. Esimerkiksi pitkäkestoista haastattelua ei voida 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 tietojen on oltava käytettävissä, on keskeinen suunnittelun tekijä. Jos tavoiteaika on tiedossa, voidaan laskea eri vaiheiden tavoiteajat. On muistettava, että tutkimukset vaativat usein yhteistyötä muiden eri tahojen kanssa. Siksi aikataulu useimmiten vaatii sopimista ja myös sovittamista.

Aikataulun suunnittelu

Aikataulussa on otettava huomioon kaikki tilastotutkimuksen työvaiheet, joiden määrä ja painotukset vaihtelevat tutkimuksittain. Esimerkiksi sähköisten tiedonkeruulinstrumenttien suunnitteluun on varattava enemmän aikaa kuin paperilomakkeiden suunnitteluun ja painatukseen. Toisaalta hyvin suunnitellun sähköisen tiedonkeruun tuloksena tiedot ovat keruuvaiheen jälkeen nopeammin 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 tärkeää, että haastattelijat koulutetaan ja että työohjeet on tehty huolellisesti. Myös muiden resurssien saatavuus (esimerkiksi esitarkastus ja tallennus) vaikuttaa aikatauluun.

Tiedonkeruuprosessin keston vaikuttaa keruutapa. Yleisesti ottaen tietokoneavusteinen puhelinhaastattelu on käyntihaastattelua ja lomakekyselyä nopeampi tiedonkeruutapa. Lomakekyselyn kesto saattaa joskus olla vaikea ennakoita. Kaikissa keruutavoissa koko tiedonkeruun kenttätyn keston vaikuttaa se, kuinka monta kertaa vastaamattomilta karhutaan vastauksia.

Vastaamattomuus eli kato

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

Tiedonkeruumenetelmä vaikuttaa kadon kokoon: suurin kato on itsetäytettävissä lomakekyselyissä. Tästä syystä nykyisin pyritään käyttämään haastatteluja. Internet-kyselyissä kato ja kehikon alipeitto ovat erittäin suuria ongelmia.

Katoon vaikuttavat myös monet muut tekijät (ks. myös luku 2.9). Esimerkiksi kysymysten määrä (so. haastattelun kesto), lomakkeen monimutkaisuus ja tutkimuksen ajoittuminen loma-aikaan lisäävät katoa. Lisäksi tutkimuksen aihepiiri vaikuttaa katoon. Esimerkiksi terveyttä koskevissa tutkimuksissa on tavallisesti muita tutkimuksia alhaisempi kato.

Kerättävän tiedon määrä ja laatu

Kerättävän tiedon määrä tai volyyymi saadaan, kun tutkimusyksiköiden määrä kerrotaan yhdeltä yksiköltä kerättävän tiedon määrällä. Molemmat tekijät vaikuttavat tutkimuksen keston. Varsin paperilomaketta käytettäessä kerättävän tiedon määrä vaikuttaa suoraan tehtävien tarkistusten määrään ja tallennettavan tiedon kokonaismäärään ja sitä kautta tutkimuksen keston.

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 motivoivista puhelinhaastatteluista. Lomakekyselyissä kato kasvaa nopeasti kysymysten määrän kasvaessa.

Haastatteluissa haastattelija täyttää lomakkeen vastaajan puolesta. Kyselyissä vastaaja lukee kysymykset sekä ohjeet ja vastaa ymmärtämänsä perusteella. Haastattelun etuna on se, että haastattelija 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 työn-ohjaus sekä lisätietojen jakaminen ovat yksinkertaisempia toteuttaa kuin haautetussa haastattelusteemissä.

Eri menetelmien etuja ja haittoja

Pitkittäistutkimukset ovat luotettavampia ja tarkempia kuvaamaan muutosta – tai yhteiskunnan dynamiikkaa – kuin toistetut poikittaistutkimukset. Lisäksi pitkittäistutkimuksilla voidaan saavuttaa säästöjä, koska niissä käytetään samaa otosta, samoja tiedonkeruustruktuureja 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-/vastaus-tilanne hyvin kontrolloitavissa, sillä 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 lisää vastausrasitetta, koska se 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 (taulukossa 2.2 on lueteltu menetelmien tietokoneavusteiset vastineet). Se mahdollistaa erilaisten kysymys- ja vastauskombinaatioiden käytön. Tietokoneavusteisessa tiedonkeruussa erilaiset tiedonkeruulomakkeen reititykset on helppo toteuttaa. Lisäksi tiedonkeruulomakkeeseen voi liittää jo valmiiksi tutkittavaa yksikköä koskevia tietoja, esimerkiksi rekisteritietoja. Tietokoneavusteinen tiedonkeruu mahdollistaa myös vastausten koodauksen jo haastattelutilanteessa, esimerkiksi ammattiluokituksen mukaan. Tietokoneavusteisuus vähentää tallennusvirheiden määrää ja nopeuttaa aineiston valmistumisprosessia, koska haastattelija tai vastaaja tallentaa tiedot suoraan tietokoneelle haastattelutilanteessa. Lisäksi erilaiset oikeellisuus- ja loogisuustarkistukset on oh-

jelmoitu osaksi tiedonkeruulomaketta. Lähes kaikki Tilastokeskuksen puhelin- ja käyntihaastattelut tehdään nykyään tietokoneavusteisesti.

Viime aikoina on alettu kerätä tietoa sähköisellä itse täytettävällä lomakkeella. Sähköinen tiedonkeruulomake voi olla Internet-lomake tai esimerkiksi sähköpostin liitetiedostona lähetettävä sähköinen lomake. Koko väestölle tarkoitettun sähköisen tiedonkeruun suurimpana ongelmana on otantakehikko; kaikilla ei ole mahdollisuutta käyttää Internet-yhteyksiä.

Sensitiivisissä kysymyksissä saattaa itsetäytettävä lomake tuoda luotettavampia vastauksia kuin haastattelu ja samalla vähentää vastaukskatoa.

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

Sähköiset itsetäytettävät tiedonkeruulinstrumentit soveltuvat tällä hetkellä vain yrityksille tehtäviin kyselyihin. On kuitenkin mietittävä, pystyvätkö kaikki yritykset vastaamaan sähköisiin kyselyihin. Henkilö- tai kotitalouspohjaisissa tutkimuksissa internet-kyselyä voidaan käyttää täydentävänä tiedonkeruumuotona.

Taulukko 2.2 Yleisimmin käytettyjen tiedonkeruumenetelmien tietokoneavusteiset vastineet (de Leeuw ja Nicholls, 1996)

Menetelmä	Tietokoneavusteinen muoto	Englanninkielinen vastine
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	Esimerkiksi internet-kysely tai levykekysely	Web-interview CAWI, Disk by Mail (DBM), Electronic Mail Survey (EMS)
Paneeli- tai päiväkirjatutkimus	Tietokoneavusteinen jatkuva paneelitutkimus tai sähköinen päiväkirja	Computer Assisted Panel Research (CAPAR), Teleinterview, Electronic diaries
Osallistuva havainnointi	Esimerkiksi tietokoneavusteinen videokoodaus	Computer Assisted Self Interviewing with Interviewer Present (CASIIP) Question Text on Screen: Visual (CASI-V) Text on Screen and on Audio (CASI-A)
Rekisteritutkimus	Rekistereiden tietokoneavusteinen yhdistäminen	

Huom: Web-interview CAWI on lisätty tähän taulukkoon. Se ei sisälly alkuperäiseen taulukkoon.

Suositukseset

- Tilastokeskus edellyttää, että väestöhaastatteluissa vastausosuudet kyetään laskemaan todellisesta tutkimuksen kohdejoukosta.
- Käyntihaastatteluissa kohteista kyetään tavoittamaan yli 90 prosenttia vastaajista ja puhelinhaastatteluissa puhelinnumerot löydetään myös yli 90 prosentille otokseen kuuluvista. Lopullinen vastausosuus on käyntihaastattelussa yleensä yli 80

prosenttia ja puhelinhaastattelussa lähellä 80 prosenttia. Tavanomaisessa postikyselyssä jäädyään selvästi alle 70 prosentin. Ruotsin tilastovirasto on koonnut laajan ohjeiston kadon pienentämiseksi (ks. Japac et al., 2000).

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

Kirjallisuus

de Leeuw, E. D., Nicholls, W. L. II. 1996. Technological Innovations in Data Collection: Acceptance, Data Quality and Costs. *Sociological Research Online*. Vol. 1. No. 4.

<http://www.socresonline.org.uk/1/4/leeuw.html>

Japac, L., Ahtiainen, A., Hörngren, J., Lindén, H., Lyberg, L. & Nilson, P. 2000. *Minska bortfallet*. Statistiska centralbyrån, Örebro.

Lisätietoja:

Ahola, A. 1993. *Tiedonkeruumenetelmä: konteksti: haastatteluhaustaukset*. Tilastokeskus. Keskustelu-
muistoita 4/1993. Helsinki.

Biemer, P. & Lyberg, L. 2003. *Introduction to Survey Quality*. New York: John Wiley & Sons.

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. Teoksessa: Groves, R. M., Biemer, P. P., Lyberg, L. E., Massey, J. T., Nicholls, W. L. II & Waksberg, J. (toim.). *Telephone Survey Methodology*. New York: John Wiley & Sons. 437–450.

Couper, M. P., Hansen, S. E. & Sadovsky, S. 1997. Evaluating Interviewer Use of CAPI Technology. Teoksessa: Lyberg, L. E., Biemer, P., Collins, M., Dippo, C., Schwarz, N. & Trewin, D. (toim.). *Survey Measurement and Process Quality*. New York: John Wiley & Sons. 267–287.

de Leeuw, E. D. 1993. Data quality in mail, telephone and face-to-face surveys. Amsterdam: TT-Publikaties.

de Leeuw, E. D. & Collins, M. 1997. Data Collection Method and Data Quality: An Overview. Teoksessa: Lyberg, L. E., Biemer, P., Collins, M., Dippo, C., Schwarz, N. & Trewin, D. (toim.). *Survey Measurement and Process Quality*. New York: John Wiley & Sons. 199–221.

Dillman, D. A. 2000. *Mail and Internet Surveys: The Total Design Method*. New York: John Wiley & Sons. (Second edition, 2006)

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: John Wiley & Sons.

Lynn, P., Clarke, P., Martin, J. & Sturgis, P. 2002. The Effects of Extended Interviewer Efforts on Nonresponse Bias. Teoksessa: Groves, R. M., Dillman, D. A., Eltinge, J. L. & Little, R. J. A. (toim.). *Survey Nonresponse*. New York: John Wiley & Sons. 135–149.

Martin, J. & Manners, T. 1995. Computer Assisted Personal Interviewing in Survey Research. Teoksessa: Lee, R. M. (toim.). *Information Technology for the Social Scientist*. London: UCL Press.

Nicholls, W. L. II, Baker, R. P. & Martin, J. 1997. The Effect of New Data Collection Technologies on Survey Data Quality. Teoksessa: Lyberg, L., Biemer, P., Collins, M., Dippo, C., Schwarz, N. & Trewin D. (toim.). *Survey Measurement and Process Quality*. New York: John Wiley & Sons.

Vehovar, V., Batagelj, Z., Lozar Manfreda, K. & Zaletel, M. 2002. Nonresponse in Web Surveys. Teoksessa: Groves, R. M., Dillman, D. A., Eltinge, J. L. & Little, R. J. A. (toim.). *Survey Nonresponse*. New York: John Wiley & Sons. 229–243.

2.8 Tiedonkeruulomake: suunnittelu ja testaus

Soveltamisala ja päämäärä

Kerättävä tieto kirjataan aina 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 lähinnä 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 eikä yleisimpien käsitteiden tuntemuksesta voida olla varmoja. Yrityskyselyjen vastaajien voidaan olettaa tietävän enemmän kysyttävistä asioista. Näistä syistä johtuen yrityskyselyissä käytetään usein kaavakkeita, joissa on vain vähän vastausta ohjaavia osia.

Laadinnan näkökulmasta tiedonkeruulomakkeet eli instrumentit, jaotellaan kahden tekijän mukaan: toisaalta sen mukaan, kuka tiedot merkitsee lomakkeelle ja toisaalta sen mukaan, onko instrumentti painettu vai sähköinen. Kaikki näin muodostuneet neljä muotoa asettavat erilaiset vaatimukset lomakkeen suunnittelulle. Usein puhutaan instrumenteista tiedonkeruulomakkeen sijasta, koska lomake assosioituu yleensä paperilomakkeeksi ja kuitenkin suuri osa lomakkeista nykyisin on sähköisiä. (Ks. taulukko 2.1)

Tässä yhteydessä merkittävien tiedonkeruulomakkeiden jaottelu on jako itseytettäviin ja haastattelulomakkeisiin.

Periaatteet

Kyselylomake laaditaan vastaajia ajatellen ja haastattelulomake haastatteli joita ja vastaajia ajatellen. Keskeiset periaatteet kaikkien kysymysten suunnittelussa ovat, että vastaajan

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

Haastatteli joista aiheutuvan vaihtelun vähentämiseksi haastatteluissa käytetään niin sanottua standardoitua menetelmää, jolla pyritään siihen, että kaikki haastattelijat esittävät kysymykset tarkalleen samalla lailla. Tästä syystä haastattelulomakkeissa 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 Tilastokeskuksen Tutkimus- ja haastattelupalvelut -yksiköstä ja SurveyLaboratoriosta. Myös muissa tilastoyksiköissä on kokemusta tiedonkeruuinstrumenttien 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 Tilastokeskuksen haastattelututkimuksista lähetetään vastaajalle etukäteen ilmoitus, jossa kerrotaan odotettavissa olevasta yhteydenotosta ja tutkimuksen sisällöstä.

Tilastokeskuksen toimintaperiaatteisiin kuuluu, että haastateltavien vastausrasite tulee pitää mahdollisimman alhaisena. Tilastolaki edellyttää, että (Ks. luku 1.1.1)

Hankittaessa tietoja tilastojen laatimista varten tulee ensi sijassa käyttää hyväksi julkishallinnon tehtävien hoitamisessa kertyneitä sekä elinkeinon- ja ammatinharjoittajien, yhteisöjen ja säätiöiden tavanomaisen toiminnan seurauksena syntyneitä tietoja. (Tilastolaki 4 §)

Tilastoja laativan viranomaisen on huolehdittava siitä, että tiedonantajilta pyydetään vain tilastojen laatimisen kannalta välttämättömät tiedot. (Tilastolaki 5 §)

Tiedot on kerättävä taloudellisesti ja niin, että siitä aiheutuu tiedonantajille mahdollisimman vähän haittaa ja kustannuksia. (Tilastolaki 4 §)

Suosituksset

Tiedonkeruulomakkeen suunnittelu

- Kysymysten laadinnasta sekä haastattelututkimuksiin että kyselytutkimuksiin on kirjoitettu useita oppikirjoja, joihin on hyvä tutustua, ennen kuin aloittaa lomakkeen laadinnan.
- Tilastokeskuksen tutkimuksissa käytetään vakiintuneita käsitelmäärittelyjä, 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.
- Itsetäytettävässä lomakkeessa on yksiselitteisesti kerrottava, kenen se tulee 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, jos vastaukset on sijoitettu samalle sarakkeelle, esimerkiksi marginaaliin tai omaan sarakkeeseen.
- Kaikkiin saatekirjeisiin tulee merkitä selvästi erottuvana tutkimuksen nimi sekä lisätietoja antavien henkilöiden nimet ja suorat puhelinnumerot, jotta puhelintiedustelut voidaan ohjata nopeasti oikealle henkilölle. Puhelinvaihteen 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 kollegojen kanssa. Erityisesti sähköisen lomakkeen testaaminen on vaativa ja aikaa vievä tehtävä.
- Pilottitutkimus kannattaa tehdä aina, kun se on mahdollista. Pilottitutkimuksen etuna on, että sen avulla voidaan testata koko tiedonkeruuprosessin toimivuus eri vaiheissa.
- On suositeltavaa konsultoida tai käyttää Tilastokeskuksen SurveyLaboratorion palveluja uuden tai uudistettavan lomakkeen suunnittelussa ja testaamisessa.
- Ryhmäkeskusteluilla saa ennakkotietoja tutkittavasta aiheesta sekä kuvan niistä käsitteistä ja ilmaisuista, joita ihmiset eri asioista käyttävät. Ryhmäkeskustelut nauhoitetaan.
- Kognitiiviset haastattelut ovat yksilöhaastatteluja, joissa tutkitaan vastaamisprosessia. Tarkastelun kohteena ovat 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

Suomen lait ja asetukset sekä kansalliset toimintaohjeet

Tilastolaki (280/2004).

<http://www.finlex.fi/fi/laki/ajantasa/2004/20040280>

Tilastokeskuksen toimintaohjeet

Tilastokeskus. 1980. *Tilastohaastattelijan opas*. Tilastokeskus. Käsikirjoja 7. Helsinki.

Tilastokeskus. 1999. *Tilastohaastattelijan opas*. Tilastokeskus. Sisäinen muistio. Helsinki.

Tilastokeskus. 2001. *Haastattelulomakkeiden suunnittelustandardit*. Tilastokeskus. Sisäinen muistio. Helsinki.

Lisätietoja

Ahola, A. 1992. *Haastattelukysymysten ymmärtämisestä*. Tilastokeskus. Keskustelumuiistioita. 1/92. Helsinki.

Biemer, P. P., Groves, R. M., Lyberg L. E., Mathiowetz, N. A. & Sudman, S. (toim.). 1991. *Measurement errors in surveys*. New York: John Wiley & Sons.

Biemer, P., Lyberg L. 2003. *An Introduction to Survey Quality*. 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, W. L. II & Waksberg, J. (toim.). 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.
- Lehto, A.-M. (1998). Laatu Surveytutkimukseen. Teoksessa: Paananen, S., Juntto, A. & Sauli, H. (toim.). *Faktajuttu*. Tilastollisen sosiaalitutkimuksen käytännöt. Tampere: Vastapaino. 207–233.
- Lyberg, L., Biemer, P., Collins, M., de Leeuw, E. D., Dippo, C., Schwarz, N. & Trewin, D. (toim.). 1997. *Survey Measurement and Process Quality*. New York: John Wiley & Sons.
- Presser, S., Rothgeb, J. M., Couper, M. P., Lessler, J. T., Martin, E., Martin, J., & Singer, E. (toim.). 2004. *Methods for testing and evaluating survey questionnaires*. New Jersey: John Wiley & Sons.
- 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. (toim.). 1992. *Questions about questions. Inquiries into the cognitive bases of surveys*. New York: Russel Sage Foundation.

2.9 Vastauskato ja siihen vaikuttaminen

Soveltamisala ja päämäärä

Tiedonkeruun eri vaiheissa havaintojen määrä usein pienenee erilaisista syistä. Epäsuoraa tiedonkeruuta, kuten rekistereitä, hyödyntävissä tutkimuksissa puuttuva tieto on pieni ongelma, kun taas erilaiset kehikko- ja mittausvirheet saattavat muodostua huomattaviksi. Sen sijaan suoran tiedonkeruun tapauksessa vastauskato, voi olla suuri ongelma. Perinteisesti vastauskatoa (*non-response*) esiintyy vapaaehtoisuuteen perustuvissa kysely- ja haastattelututkimuksissa. Mutta vastauskadolta ei voi täysin välttyä lakisääteisissä tiedonkeruissa.

Vastauskato jaetaan perinteisesti kahteen pääryhmään: eräkatoon (*item non-response*) ja yksikkökatoon (*unit non-response*) (Groves ja Couper, 1998). Eräkadolla tarkoitetaan sellaista vastausta, jossa tutkimusyksikkö antaa vain osan tiedoista hyväksyttävästi tai antaa sellaisen vastauksen, joka myöhemmissä aineiston tarkistuksissa joudutaan hylkäämään. Yksikkökadon tapauksessa kaikki havaintoyksikköä koskevat tutkimustiedot puuttuvat tai joudutaan hylkäämään.

Tutkimusjoukon pienenemisellä tavoitteeseen eli perusjoukkoon nähden on useimmiten harmillisia vaikutuksia tilastotutkimuksen tuloksiin (Lessler ja Kalsbeek, 1992; Särndal, Swensson ja Wretman, 1992; Särndal ja Lundström, 2005). Mikäli vastanneet ja kato ovat sekä tausta- että tutkimusmuuttujien suhteen samoin jakautuneita, otosvarianssi suurenee kadon vaikutuksesta. Myös kokonaistutkimuksiin 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. Eriytyinen 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 tiedonantajien kieltäytymispäätöksiin ei voida juuri vaikuttaa; tosin haastattelijoita ja muuta tiedonkeruuseen osallistuvaa henkilökuntaa voidaan kouluttaa kieltäytyneiden suostuttelussa. Sen sijaan lakisääteisissä ja muissa ei-vapaaehtoisissa tutkimuksissa erityisesti kieltäytyneiden osuutta pyritään pitämään mahdollisimman vähäisenä. Usein juuri kieltäytyneiden ryhmä aiheuttaa suurimmat riskit aineistojen vinoutumiseen ja harhaisuuteen.

Tiedonkeruun jälkeen aineiston sisältämän vinouden ja virheen sekä mahdollisen harhan arvioimiseksi ja sen vähentämiseksi (eli tulosten oikaistamiseksi) on tärkeää, että saatua havaintoaineistoa verrataan huolellisesti tutkimuksen perusjoukon kanssa. Tilastotutkimusten laatuselosteissa ja menetelmäraporteissa informoidaan tiedonkeruusta, saadusta havaintoaineistosta ja vastausosuuksista. Laatuselosteita perusteellisemmat menetelmäraportit sisältävät myös mahdollisimman yksityiskohtaisen katoanalyysin: kadon syistä ja vaikutuksista tuloksiin. Kadon aiheuttamaa harhaa pyritään korjaamaan tilastollisin menetelmin (ks. luku 2.11).

Suosituks

- Tutkimuksen perusjoukko ja mahdolliset poimintakehikot on tutkittava huolellisesti ja verrattava tutkimuksen tavoiteperusjoukkoon.
- Kenttätyöt ja käytettävät menetelmät ovat asianmukaiset ja oikein suunnatut tutkimuksen kohdeperusjoukolle (kuten työn ajoitus, materiaalit, ohjeet, tiedonkeruuvälineet, mahdolliset tekniset ja muut apuvälineet ovat mahdollisimman huolellisesti valittuja ja pyrkivät pienentämään vastausrasitusta).
- Tiedonkeruuvälineet ja menetelmät on testattava huolellisesti ennen varsinaista tiedonkeruuta.
- Valitun otantamenetelmän on tuotettava harhattomat tulokset perusjoukolle ja sen tärkeimmille osajoukoille.
- Tavoittamatta jääneitä tapauksia on pyrittävä jäljittämään. Joissain tapauksissa myös kieltäytyneitä tulee suostutella vastaajiksi. Pakollisissa tutkimuksissa kieltäytymiset on minimoitava.
- Vinoja perusjoukkoja tutkittaessa (kuten yrityksiin suunnatuissa tutkimuksissa) on pyrittävä saamaan vastaus tutkittavan ilmiön kannalta tärkeimmiltä yksiköiltä.
- Vastaajien ja kadon rakenne on tutkittava ja aineisto oikaistava kadon vaikutuksista niin hyvin kuin mahdollista, esimerkiksi painottamalla tai imputoimalla (ks. luvut 2.10 ja 2.11).

- Katovirhe on mitattava, dokumentoitava ja raportoitava. Katovirheen mitauksessa tulee käyttää yleisesti hyväksytyjä menetelmiä kato- ja/tai vastausosuuden määrittämiseksi.
- Kadon määrä ja rakenne on raportoitava ja aineiston käyttäjiä informoitava sen mahdollisista vaikutuksista tilastolliseen laatuun ja tutkimustulosten luotettavuuteen. Toistuvissa tilastotutkimuksissa kadon suuruutta on seurattava, jotta kenttätöissä tai yhteiskunnallisessa ilmapiirissä tapahtuvat muutokset voidaan huomioida.

Kirjallisuus

- Groves, R. M. & Couper, M. P. 1998. *Nonresponse in Household Interview Surveys*. New York: John Wiley & Sons.
- Lessler, J. & Kalsbeek, W. (toim.). 1992. *Nonsampling Error in Surveys*. New York: John Wiley & Sons.
- Särndal, C.-E., Swensson, B. & Wretman, J. 1992. *Model Assisted Survey Sampling*. New York: Springer-Verlag.
- Särndal, C.-E. & Lundström, S. 2005. *Estimation in Surveys with Nonresponse*. Chichester: John Wiley & Sons.

Lisätietoja

- Groves, R. M., Dillman, D. A., Eltinge, J. L. & Little, R. J. A. (toim.). 2002. *Survey Nonresponse*. New York: John Wiley & Sons.
- Groves, R. M., Fowler, F.J., Couper, M. P., Lepkowski, J. M., Singer, E., and Tourangeau, R. 2004. *Survey Methodology*. New York: John Wiley & Sons.
- Japac, L., Ahtiainen, A., Hörngren, J., Lindén, H., Lyberg, L. & Nilsson, P. 1997. *Minska bortfallet*. Örebro: Statistiska centralbyrån.
- Laaksonen, S. 1988. *Katovirheen korjaus kotitalousaineistossa*. Tilastokeskus. Tutkimuksia 147. Helsinki.
- Laaksonen, S. (toim.). 1996. *International Perspectives on Nonresponse*. Statistics Finland. Research Reports No. 219. Helsinki.

2.10 Tilastollinen editointi ja imputointi

Soveltamisala ja päämäärä

Tilastollista editointia ja imputointia hyödynnetään virheellisten havaintojen havaitsemiseen ja korjaamiseen aineistoissa. Näitä menetelmiä sovelletaan yleensä aineiston keräämisen jälkeen, mutta elektronisessa tiedonkeruussa ne voidaan toteuttaa myös automaattisesti.

Editoinnissa ja imputoinnissa käytetyt menetelmät ja sovellukset kehittyvät koko ajan. Tällä hetkellä käytännössä ja kirjallisuudessa esiintyy vielä epä johdonmukaisuuksia terminologiassa, minkä vuoksi seuraavassa esitetään tässä luvussa käytetyt määritelmät. Terminologian yhdenmukaistaminen on kuitenkin edellytys menetelmien merkitykselliselle ja tehokkaalle kehittämiselle ja systemaattiselle soveltamiselle. Eurostat (2005) määrittelee editoinnin ja imputoinnin seuraavasti:

- Tilastollisella editoinnilla tunnistetaan puuttuvia, virheellisiä tai epäyhdenmukaisia havaintoja aineistossa.
- Imputointia käytetään puuttuvien, virheellisten ja epäyhdenmukaisten havaintojen korjaamiseen. Imputointi toteutetaan korvaamalla virheelliset tai puuttuvat havainnot uskottavilla, sisäisesti yhdenmukaisilla arvoilla.

Myös YK:n tilastollinen komissio ja Euroopan talouskomissio (UN/ECE, 2000a) ovat tuottaneet kattavan dokumentin editointi ja imputointi -käsitteistä, periaatteista, tekniikoista ja menetelmistä. Molemmissa yllä mainituissa lähteissä aineistojen editointi on kiinteästi yhteydessä imputointiin. Editointia hyödynnetään enimmäkseen tunnistamaan virheellisiä tai epäjohdonmukaisia havaintoja. Imputoinnilla vastaavasti paikataan puuttuvia, virheellisiä tai epäjohdonmukaisia havaintoja.

Tilastollinen editointi (*statistical editing*) tarkoittaa niitä toimintoja, joilla tilastoaineiston havaintoarvot tarkistetaan. Tämä käsittää yksittäisten arvojen, arvojen keskinäisen loogisuuden ja yhteensopivuuden tarkistukset. Editointi tulee tehdä sekä kerätyn aineiston perus- eli mikrotasolla (*mikro-editointi*) että karkeammalla tasolla (*makro-editointi*). Tällöin tulee pyrkiä aineiston eri tasojen keskinäiseen yhteensopivuuteen. Lopuksi koko aineisto täytyy tarkistaa huolella, jotta aineiston estimaatteihin ja niiden jakaumiin ei aiheudu merkittäviä vääristymiä.

Imputointi (*imputation*) tarkoittaa puuttuvien tai virheellisten tietojen paikkaamista korvikearvoilla, joiden tulee olla mahdollisimman oikeita. Imputointimenetelmät vaihtelevat merkittävästi aineiston tyyppin, laajuuden ja puuttuvien havaintojen mukaan.

Käytännössä editointia ja imputointia hyödynnetään aineiston käsittelyssä rinnakkain. Tilastollista editointia tarvitaan tarkistamaan ja identifioimaan löytyykö sopimattomia tai loogisesti epäyhdenmukaisia arvoja. Tämän jälkeen imputointia käytetään paikkaamaan nämä tunnistetut arvot. Tilastollista editointia tehdään useissa tilastotuotannon vaiheissa, alkaen tiedonkeruun suunnittelusta aineiston muodostamiseen ja analysointiin. Aineiston lisäeditointi voi tulla tarpeelliseksi jopa vuosia keruun päättymisen jälkeen, esimerkiksi muodostettaessa pitkittäisaineistoja.

Periaatteet

Editoinnilla ja imputoinnilla on kolmitahoinen tehtävä: aineiston siistiminen, informaation tuottaminen aineiston laadusta ja tulevien tilastotutkimusten aineistojen laadun parantaminen. Laadun arvioimiseksi editoinnissa ja imputoinnissa havaittujen virheiden ja puutteiden kokonaisvaikutus ja tärkeys on analysoitava. Aineiston suurten ongelmien ehkäisemiseen ja analysoimiseen on panostettava riittävästi resursseja. Sen sijaan pienempiin puutteisiin tulee soveltaa automaattisia editointi- ja imputointisovelluksia.

Tilastollinen editointi

Editointisääntöjen suunnittelu on avainasemassa tilastollisessa editoinnissa. Sääntöjen suunnittelussa on huomioitava editoinnin ja imputoinnin vaikutus aineiston jakaumiin. Editointisääntöjä ei tule suunnitella siten, että ne maksimoivat

havaittujen virheiden määrän. Editointisääntöjä määritettäessä on huomioitava muun muassa seuraavat vaatimukset (UN/ECE, 1997):

- *Kattavuus*: Kaikki mahdolliset virheellisten arvojen identifiointisäännöt on määritettävä.
- *Oikeellisuus*: Jokaisen editointisäännön on perustuttava aineiston, teorian ja ilmiöalueen tuntemukseen.
- *Selkeä esitys ja dokumentointi*: Editointisäännöt on dokumentoitava siten, että ne voidaan jäljittää ja niiden vaikutusta arvioida.

Tilastotutkimuksen aineisto on editoitava eli tarkistettava aina ennen tulosten julkistamista. Tällä tavalla varmistetaan, että tieto on tarkkaa, täsmällistä ja yhdenmukaista. Editoinnin aiheuttaman laadun muutoksen mittaaminen perustuu lopullisen ja raaka-aineiston vertaamiseen. *Editointiaste* kuvaa tehdyn kontrollin suhteellista laajuutta. Se mitataan virheelliseksi havaittujen osuudella kaikista havainnoista. *Editointisuhde* mittaa virheellisiksi havaittujen arvojen suhteellista vaikutusta. Se muodostetaan virheellisiksi havaittujen arvojen summan ja kaikkien havaintojen summan suhteena.

Editointiastetta ja -suhdetta voidaan hyödyntää aineiston muokkauksen tehokkuustarkasteluun. Pieni editointisuhde viittaa siihen, että virheelliset havainnot eivät muodosta merkittävää osaa kaikista havainnoista. Tämä yhdistettynä suureen editointiasteeseen voi merkitä virheellisten havaintojen etsimisen ylisurssointia (UN/ECE, 2000b). Tällöin resurssien uudelleen allokoinnilla voidaan saavuttaa sekä laadullisia että tehokkuusparannuksia.

Editoinnissa on pyrittävä varmistamaan, että estimaattien harha minimoidaan käyttäen tehokkaasti resursseja hyväksi, välttäen ylieditointia. Editointiin liittyy neljä yleisesti hyväksyttyä vaihetta: alkueditointi, perinteinen editointi, valitseva editointia ja makro-editointi, joilla on luonnollinen järjestys. Jokaisen editointivaiheen jälkeen yleensä seuraa imputointiaskel korjaamaan havaitut virheet. Editointivaiheet voivat olla myös limittäisiä niiden erilaisten vahvuuksien ja vaihtelevan paljastuskyvyn vuoksi (Scarrott, 2006). Yksi käytetyimmistä editointimenetelmistä on Fellegi ja Holtin menetelmä (Fellegi ja Holt, 1976), jonka sovellus on käytettävissä useimmissa editointiohjelmistoissa (esimerkiksi BANFF).

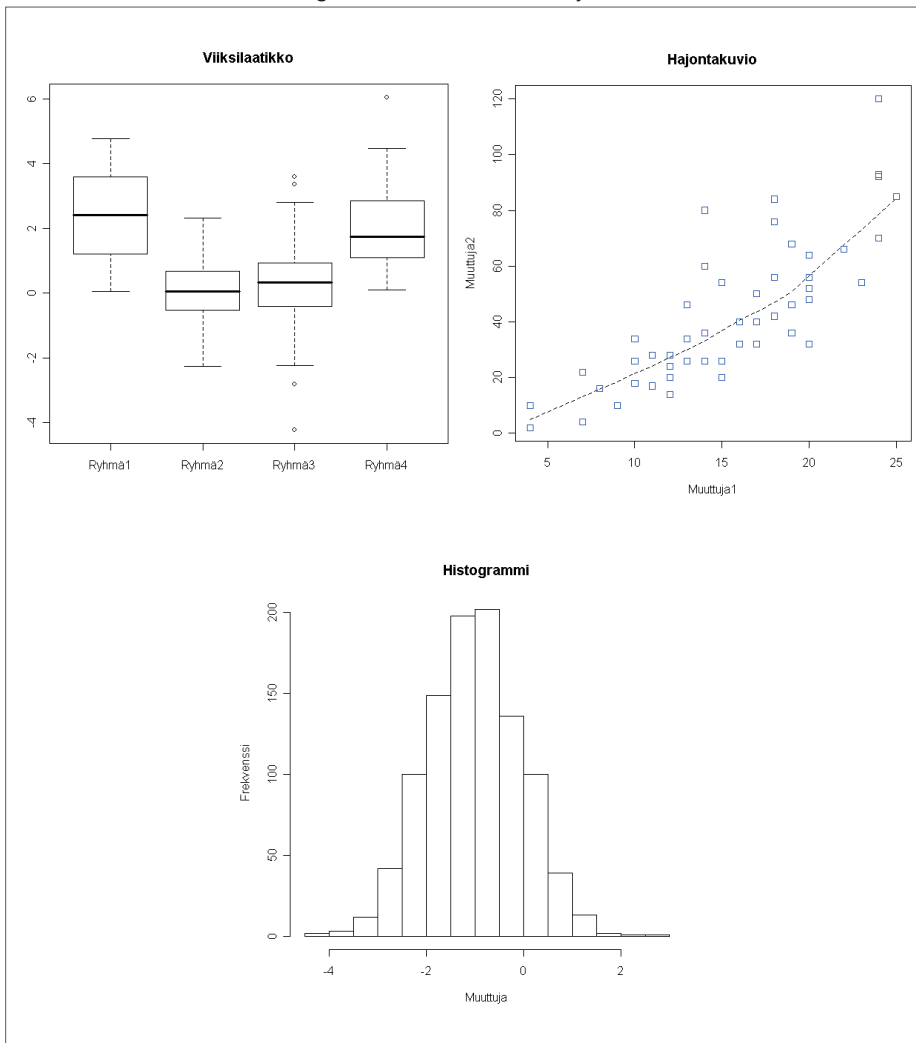
Graafinen editointi

Graafisessa editoinnissa hyödynnetään ihmisen visuaalista kykyä havaita poikkeavuuksia, säännöllisyyksiä tai yhteyksiä aineistossa, joiden havainnointi olisi tehostomampaa käytettäessä analyyttisiä keinoja. Graafinen editointi tukee tavallisia editointimenetelmiä, jotka eivät yleensä pysty osoittamaan poikkeavien havaintojen vaikutusta estimaatteihin (UN/ECE, 1997).

Alustavasta aineiston tutkimisesta graafisin keinoin voi olla suurta hyötyä suunnitellessa editointi- ja imputointisääntöjä. Hyödyntäen esimerkiksi viiksiläätkoita, hajontakuvioita, regressiosuoria ja -suhteita, voidaan tutkia ääriarvoja, trendiä ja muita ominaisuuksia. Kolme tyypillistä graafisen tutkimisen tapaa on esitetty kuviossa 2.5:

- Laatikkokuviot (*boxplot*) esittävät aineiston jakaumia eri kategorisissa ryhmissä. Kuvaajista havaitaan aineiston arvojen vaihtelu eri ryhmissä sekä identifioidaan ääriarvot.
- Hajontakuviot (*scatter plot*) paljastavat muuttujien välisiä yhteyksiä. Regressio suorasta poikkeavat havainnot on syytä tutkia huolella poikkeavien arvojen tunnistamiseksi.
- Histogrammit näyttävät graafisesti aineiston jakauman muodon, hajonnan ja ääriarvojen esiintymisen. Kuvaajasta havaitaan aineiston vinous sekä muista selvästi poikkeavat havainnot.

Kuvio 2.5 Kolme esimerkkiä graafisen visualisoinnin käytöstä editoinnissa



Imputointi

Imputoinnilla korvataan muuttujien puuttuvat tai virheelliseksi havaitut arvot hyväksyttävillä arvoilla. Imputoinnilla pyritään tehokkaasti tuottamaan samat tulokset, mitä täydellinen aineisto tuottaisi (Chambers, 2002). Kalton ja Kasprzyk (1986) ovat esittäneet yleiskuvauksen eri imputointimenetelmistä. Ne voidaan jakaa kolmeen pääryhmään: loogiseen imputointiin, malliluovuttaja-imputointiin ja vastaajaluovuttajaimputointiin (Laaksonen, 2000):

- Looginen imputointi tapahtuu tilastollisen editoinnin yhteydessä. Imputoidut arvot konstruoidaan asiantuntijan avulla tai parhaan mahdollisen päätelyn pohjalta, käyttäen hyväksi vahvoja apumuuttujia tai aikaisempia arvoja.
- Malliluovuttaja-imputointimenetelmässä (*model-donor imputation*) imputoidut arvot konstruoidaan yksinkertaisen tai monimuuttujaisen tilastollisen mallin tuottamina sovitteina.
- Vastaajaluovuttaja-imputointimenetelmässä (*real-donor imputation*) imputoidut arvot saadaan samankaltaisilta aidoilta vastaajilta. Arvo on siis varmasti sellainen, joka on aidossa tilanteessa mahdollinen.

Kahdessa viimeisessä menetelmäryhmässä puuttuvan arvon korvike eli imputoitu havaintoarvo lainataan joltakin, jota voidaan kutsua luovuttajaksi (*donor*). Imputointimenetelmän pääryhmät määräytyvät luovuttajan tyyppin mukaan. Käytännön tilanteessa on suositeltavaa verrata useita menetelmiä, jopa yhtä muuttujaa imputoitaessa, koska esimerkiksi malliluovuttaja-imputointi voi sopia hyvin sellaisiin tilanteisiin, joihin vastaajaluovuttaja-imputoinnilla ei löydy oikeita havaintoja.

Kaikkia imputointimenetelmiä voidaan soveltaa sekä deterministisesti että stokastisesti. Viimeksi mainittuun sisältyy satunnaislementti. Imputointimallit voidaan edellisen lisäksi jakaa kahteen ryhmään. Ensimmäisessä ryhmässä puuttuva arvo korvataan yhdellä arvolla (*single imputation*) ja toisessa ryhmässä usealla korvikearvolla (*multiple imputation*) (Rubin, 1987). Tilastokeskuksen tilastotuotannossa ei tällä hetkellä hyödynnetä moni-imputointimenetelmiä.

Imputointi aiheuttaa tuloksiin aina tiettyä lisäepätarkkuutta, joka pitää pyrkiä estimoimaan. Moni-imputointia voidaan käyttää lisävarianssin estimoimiseen (Rubin, 1987; Rubin, 1996; Schafer, 1997). Muita soveltuvia menetelmiä imputointivarianssin estimoimiseen ovat esittäneet (esimerkiksi Rao ja Shao, 1992; Särndal, 1996; Lee, Rancourt ja Särndal, 2002; sekä Särndal ja Lundström, 2005).

Imputointiprosessin monitorointiin on käytettävissä laatuindikaattoreita, joihin tulee tulkita varoen. *Imputointiaste* lasketaan imputoitujen havaintojen osuutena kaikista havainnoista. *Imputointisuhde* kuvaa imputoitujen arvojen kontribuutiota lopulliseen estimaattiin. Tämä suhdeluku lasketaan imputoitujen arvojen summan osuutena kaikkien arvojen summasta (Eurostat, 2005).

Suositukseset

- Eräkadosta aiheutuvaa puuttuvaa tietoa tulee tarpeen mukaan imputoida kaikissa tilastoaineistoissa.

- Yksikkökadon aiheuttamat puuttuvat tiedot oikaistaan aineiston uudelleen painottamisella kotitalous ja henkilöaineistoissa (ks. luku 2.11). Muissa tilastoaineistoissa yksikkökatoa voidaan oikaista myös imputoimalla.
- Tunnistimien eli identifioijien oikeellisuus on ratkaisevan tärkeä, jos aineistoa käytetään painottamiseen, estimointiin tai yhdistetään muihin aineistoihin. Mikäli tilastoaineistoja yhdistetään pitkittäin, niin vain poikkileikkauslinkkaukseen soveltuva tunnistin ei riitä. Pitkittäisaineistoihin on siten suunniteltava myös erillinen pitkittäisaineiston tunnistinjärjestelmä.
- Automatisoituja editointi- ja imputointisovelluksia on hyödynnettävä niin paljon kuin mahdollista.
- Editointi- ja imputointiprosessit tulee dokumentoida tarkasti. Dokumentoinnissa tulee ilmetä teoreettiset ja käytännölliset perustelut. Lisäksi alkuperäinen raaka-aineisto ja automatisoitu editointi- ja imputointisyntaksi täytyy dokumentoida ja arkistoida.
- Editoinnin ja imputoinnin vaikutusten arviointiin käytetään laatuindikaattoreina editointiastetta ja -suhdetta, sekä imputointiastetta ja -suhdetta (Eurostat, 2005; UN/ECE, 2000b). Laatuindikaattorit on tuotettava ja tarkistettava. Niiden tulkinnessa tulee kuitenkin noudattaa varovaisuutta.

Tilastollinen editointi

- Arvoaluetarkistus on välttämätön koko aineiston tasolla ja tulee tehdä mahdollisuuksien mukaan myös erillisesti tarkoituksenmukaisissa osajoukoissa. Esiohjelmoituja tarkastussysteemejä täytyy soveltaa käytettäessä tietokoneavusteista tiedonkeruujärjestelmää (kuten CAPI tai CATI, ks. luku 2.7).
- Kaikkien tunnistettujen ääriarvojen tai poikkeavien arvojen oikeellisuus on tarkastettava huolellisesti.
- Puuttuva, nollatieto ja mahdoton tieto tulee erottaa toisistaan.
- Yksittäisien tietojen ja osajoukkojen tulee olla johdonmukaisia. Ristikkäistarkastuksilla eri osa-alueita tutkitaan mahdollisesti myös ulkopuolisten aineistojen kanssa.

Imputointi

- Aineistoon sopiva imputointimalli täytyy konstruoida kokeilemalla ja testaamalla useita menetelmiä.
- Imputoidut arvot tulee merkitä indikaattorimuuttujalla, jotta imputoinnit voidaan erottaa alkuperäisistä havainnoista.
- Imputointitulokset täytyy tutkia sekä graafisesti että taulukoimalla. Malliluovuttaja-menetelmää käytettäessä on imputoitujen arvojen jakaumaa tutkittava mahdollisten virheellisten tai ääriarvojen varalta. Vastaajaluovuttaja-menetelmää käytettäessä on tarkasteltava sitä, onko eri luovuttajia on käytetty riittävästi.
- Imputoinnista aiheutuva lisävarianssi tulee pyrkiä estimoimaan.

Kirjallisuus

Chambers, R. 2002. *Evaluation Criteria for Statistical Editing and Imputation*. National Statistics Methodology Series 28.

Eurostat, 2005. *European statistics code of practice, Glossary of quality terms*.

Fellegi, I.P. and Holt, D. 1976. A Systematic Approach to Automatic Edit and Imputation. *Journal of the American Statistical Association*. Vol. 71. 17–35.

- Kalton, G. & Kasprzyk, D. 1986. The Treatment of Missing Survey Data. *Survey Methodology*. Vol. 12. No. 1. 1–16.
- Laaksonen, S. 2000. Regression-based Nearest Neighbour Hot Decking. *Computational Statistics*. Vol. 15. No.1. 65–71.
- Lee, H., Rancourt, E. & Särndal, C.-E. 2002. Variance Estimation from Survey Data under Single Imputation. Teoksessa: Groves, R. M., Dillman, D. A., Eltinge, J. L. & Little, R. J. A. (toim.). *Survey Nonresponse*. New York: John Wiley & Sons. 315–329.
- Little, R. & Rubin, D. 1987. *Statistical Analysis with Missing Data*. New York: John Wiley & Sons.
- Rao, J. N. K. & Shao, J. 1992. Jackknife 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.
- Scarrott, C. 2006. *Feasibility study: A review of selective editing*. Mathematics and statistics department, University of Canterbury.
- Schafer, J. L. 1997. *Analysis of Incomplete Multivariate Data*. London: Chapman & Hall.
- Särndal, C.-E. 1996. For a Better Understanding of Imputation. Teoksessa: Laaksonen, S. (toim.). *International Perspectives on Non-response*. Statistics Finland. Research Reports 219. Helsinki. 7–22.
- Särndal, C.-E. & Lundström, S. 2005. *Estimation in Surveys with Nonresponse*. Chichester: John Wiley & Sons.
- United Nations statistical commission and economic commission for Europe (UN/ECE). 1997. *Statistical data editing: Methods and techniques*. United Nations, New York and Geneva.
- United Nations statistical commission and economic commission for Europe (UN/ECE). 2000a. *Glossary of terms on statistical data editing*. United Nations, Geneva.
- United Nations statistical commission and economic commission for Europe (UN/ECE). 2000b. *Evaluating efficiency of statistical data editing: General framework*. United Nations, Geneva.

Lisätietoja

- Laaksonen, S. 1991. Adjustment for Non-response in Two-year Panel Data: Applications to Problems of Household Income Distribution. *The Statistician*. Vol. 40. No. 2. 153–168.
- Laaksonen, S. 1996. *Statistical Methodology and the Principle of Subsidiarity: A Challenging Optimization Task for Eurostat and European Statistical Institutes*. Proceedings of the 1st International Conference on Methodological Issues in Official Statistics. Stockholm: Statistics Sweden.
- Longford, N.T. 2005. *Missing Data and Small-Area Estimation: Modern Analytical Equipment for the Survey Statistician*. New York: Springer-Verlag.
- Statistics Canada. 2006. Banff. Functional Description of the Banff System for Edit and Imputation. Generalised System Method Section & Business Survey Methods Division.

2.11 Painokertoimet ja katovirheen oikaisu

Soveltamisala ja päämäärä

Useimmissa otantatutkimuksissa havaintoihin kuuluu liittyy painokerroin, jonka käytöllä pyritään siihen, että estimointimenetelmän tuottamat parametriestimaatit olisivat mahdollisimman virheettömiä. Painokerroin eli otospaino sisältää kuhunkin havaintoon kohdentuvan informaation otoksen poiminnassa käytettyä otanta-asetelmasta. Joissain tapauksissa painokertoimet sisältävät tietoa myös lisäinformaatioon perustuvista estimoinnin tehostus- ja/tai kadonoikaisu menetelmistä, joita selitetään tarkemmin periaatteet-osiossa.

Asetelmapaino on yksikertaisin esimerkki painokertoimista. Asetelmapaino, kuvaa otosasetelmaa ja otoskokoa suhteessa perusjoukkoon. Asetelmapaino ei

useimmissa käytännön tilanteissa riittä, koska otoksen poiminnan ja tiedonkeruun jälkeen aineisto saatetaan havaita vinoksi kehikkovirheiden, otantavirheen, vastauskadon tai mittausvirheiden vuoksi. Asetelmapainoja joudutaan muokkaamaan erilaisten mallioletusten perusteella, ja näin johdettua painotusta kutsutaan uudelleenpainotukseksi.

Periaatteet

Asetelmapaino

Asetelmapaino on painokerroin, joka johdetaan otosasetelmasta. Useimpien otanta-asetelmien tapauksessa se muodostuu kunkin alkion sisällymistodennäköisyyden käänteislukuna. Kunkin poimitun alkion todennäköisyys tulla valituksi on sen tähden kyettävä johtamaan.

Otospaino on perusmuodossaan korottava eli laskettaessa kokonaismäärien estimaatteja tulokset ovat perusjoukon tasolla. Tällöin painokertoimen antama lukema voidaan tulkita niin, että otokseen valittu havainto edustaa painonsa osoittamaa määrää havaintoja perusjoukossa. Painokertoimen muodostusta varten on selvitetty käytettävän kokonaismäärän laskentakaavan eli estimaattorin rakenne.

Ilman painokerrointa voidaan estimoida vain, jos asetelma on itsepainottuva (esimerkiksi yksinkertainen satunnaisotanta tai satunnaistetusta perusjoukosta poimittu 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.

Joskus korottavaa vaikutusta ei haluta painoon mukaan, jolloin painon arvot vaihtelevat ykkösen molemmiin puoliin ja painojen summaksi tulee hyväksytyyn havaintoaineiston koko. Tällöin puhutaan analyysipainosta. Analyysipainon käyttöä puolletaan tutkimustraditiolla, tulostussyllä tai sillä, ettei sovellettavalle analyysimenetelmälle ole toistaiseksi johdettu otanta-asetelman huomioon ottavaa 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

Uudelleenpainotus on yleisesti käytössä oleva menetelmä, jolla alkuperäisiä asetelmapainoja muokataan käyttämällä hyödyksi lisäinformaatiota joko perusjoukosta, otoksesta tai kummastakin. Yksinkertaisimmat uudelleenpainotusmenetelmät ovat jälkiositus sekä suhdetehostus. Jälkiosituksessa otos painotetaan perusjoukon tunnettujen jakaumatietojen mukaan, esimerkiksi henkilöperustaisissa tutkimuksissa demografisten tietojen, kuten ikä- ja sukupuoliryhmien sekä asuinalueen mukaan (Särndal et al., 1992). Vastaavasti suhdetehostuksessa hyödynnetään jatkuviin muuttujien summatietoja, kuten yrityksiä koskeissa tutkimuksissa yrityksen liikevaihdon määrää suhteessa koko toimialan liikevaihdon summaan. Suhdetehostus edellyttää, että käytetyn lisäinformaation ja tutkimusmuuttujien välillä on suoraviivainen positiivinen suhde. (Särndal et al., 1992).

Monimutkaisemmat uudelleenpainotusmenetelmät perustuvat joko eksplisiittisten tai implisiittisten tilastollisten mallien käyttöön. Niissä asetelmapainoa muokataan mallittamalla sitä erilaisen lisäinformaation avulla. Tyypillisesti käytetään joko lineaarista regressiomallia tai logit-mallia. Eräs yleisimmistä menetelmistä on painojen kalibrointi (Deville ja Särndal, 1992). Siinä asetelmapainoja muokataan mallin avulla mahdollisimman vähän ehdolla, että perusjoukon jakaumatiedot saadaan oikeiksi. Näin menetellen pyritään sekä säilyttämään alkuperäisen otosasetelman ominaisuudet että korjaamaan erilaisia perusjoukosta tiedossa olevia taustamuuttujien jakaumia mahdollisimman virheettömiksi.

Uudelleenpainotus lisää miltei aina painokerrointen vaihtelua, joten sen vaikutusta esimerkiksi keskivirheisiin ja muihin laatuindikaattoreihin on syytä arvioida. Erityisen suureksi painojen vaihtelun merkitys voi kasvaa, jos alkuperäinen otos on poimittu PPS-otannalla tai hyvin voimakkaalla osituksella kokoa mittaavan muuttujan suhteen. Tarvittaessa tulee pyrkiä rajoittamaan uudelleenpainotuksesta syntyvää lisävaihtelua.

Katovirheen oikaisu

Eräs malliin perustuva operaatio on valittujen havaintoyksiköiden vastaamattomuudesta johtuvan katovirheen oikaisu eli adjustointi. Vastauskato saattaa olla suuri ja sen huomiotta jättäminen tuottaisi alipeittäviä estimaatteja kokonaisu-määrille. Lisäksi vastauskato ei juuri koskaan ole täysin satunnaista, vaan sen ja väestöjakaumien tai tutkimusmuuttujien välillä voidaan arvioida olevan riippuvuutta. Asetelmapainokertoimen käyttäminen voi tällaisissa tapauksissa tuottaa harhaisia tuloksia ja johtaa virheellisiin johtopäätöksiin.

Katovirheen oikaisu voidaan yksinkertaisimmillaan toteuttaa korvaamalla alkuperäisen otoksen koko vastanneiden määrällä, toisin sanoen otospainoja inflatoidaan kato-osuuden verran. Suositeltavampaa on oikaista kadon vaikutusta pohjautuen tutkittavan ilmiön kannalta relevanttiin informaatioon perusjoukosta sopivaa uudelleenpainotusmenetelmää hyväksikäyttäen. Tällöin oletetaan otoksen ulkopuolisen tiedon vähentävän katovirhettä edellyttäen, että otos kyettään jakamaan vastaustodennäköisyyksien suhteen mahdollisimman homogeenisiin osajoukkoihin. Joissakin tapauksissa voi olla tarkoituksenmukaista skaalata otospainoja perusjoukon (rekisterin) tiedossa oleviin tunnuslukuihin. (Rosenbaum, 1987; Ekholm ja Laaksonen, 1991; Särndal ja Lundström, 2005)

Painokertoimien kalibrointi

Katovirheen oikaisu perustuu usein jälkiositusta muistuttavaan käytäntöön, jossa lisäinformaation avulla saatavia jakaumia tai ristiintaulukoituja tietoja käytetään osituksen tapaan painokertoimen muodostamisessa (Smith, 1991; Särndal et al., 1992; Deville, Särndal ja Sautory, 1993). Teoreettisesti vaativammassa menetelmässä eli painojen kalibroinnissa käytetään useamman muuttujan jakaumatietoa, mutta solufrekvenssitietojen sijasta käytettävissä on pelkästään reunajakaumat taikka jatkuvien muuttujien kokonaisu-määrät (Deville ja Särndal, 1992; Lundström, 1997). Useimmiten käytetään perusjoukon tai sen ositteiden tasoista tietoa, joskus myös otostason tietoa. Painojen kalibrointiin soveltuvat CALMAR (Sautory, 2003) ja CLAN97 (Andersson ja Nordberg, 1998) ohjelmistot.

Suosituksset

- Painokerroin koostuu yleensä perusjoukkoon, otokseen sekä vastanneisiin liitetyistä lukumäärätiedoista, jotka on usein laskettu jostain perusjoukon ryhmistä, yleensä ositteista. Jos ryhmittelyä on tehty, on kunkin havainnon aina identifioitava johonkin ryhmään. Kyseisten ryhmien perusjoukkotasosten 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 onnistuminen tulee todentaa ensinnäkin laskemalla havaintoaineistossa painokertoimien summa. Tällöin summan pitäisi olla sama kuin perusjoukon koko.
- 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 yhdenmukaisia aineistosta laskettujen vastaavien painotettujen lukujen kanssa (Deville ja Särndal, 1992; Särndal ja Lundström, 2005).
- Eri otanta-asetelmien mukaisten asetelmapainojen kaavat on muodostettu tilanteeseen, jossa yksikkökatoa ei ole. Toimenpiteenä tilanteessa voi olla alkuperäisen otoksen koon korvaaminen vastanneiden määrällä. Tällöin siis oletetaan havainnon empiiriseksi vastaustodennäköisyydeksi vastanneiden määrän osuus otoksesta.
- Painokertoimien jakaumaa on tarkasteltava jo otanta-asetelman muodostusvaiheessa ja ennakoitava mahdollista katoa eri ositteissa. Hyvin suuret painot muuten pienien painojen joukossa voivat vaikuttaa estimaattoreiden variansseihin odottamattoman voimakkaasti.
- Jos ositus aiheuttaa voimakasta painonvaihtelua, on suositeltavaa karkeistaa ositusta (ositteiden luhistaminen) tai lisätä ongelmaositteisiin otoskokoja. Jälkiosituksen tapauksessa ositteiden karkeistus on ainoa keino. Kalibrointia käytettäessä annettujen reunajakaumajakojen karkeistaminen ja määräestimaattietojen vähentäminen johtavat usein toivottuun lopputulokseen.

Dokumentointi

Raportoinnissa keskeistä on otanta-asetelman, estimointimenetelmän ja uudelleenpainotuksen tarkka kuvaus. Laatuselosteissa on selvitettävä tärkeimpien estimaattoreiden rakenne, aineistoon sisältyvät virheet, erityisesti katovirhe, ja niistä aiheutuneet korjaustarpeet painotuksessa. Laatu- ja menetelmäselosteiden menetelmäkuvauksiin tulee sisältyä painokertoimen muodostamisperiaatteet sekä kuvaus lopullisen painotuksen tuottamista tärkeimmistä parametriestimaateista (ks. luku 1.2.3). Kuvauksissa on selvitettävä painotuksen aiheuttamat muutokset perusasetelmasta johdettuihin otosvariانسien lausekkeisiin, jotta keskivirhe ja muut keskeiset laatuindikaattorit voidaan laskea mahdollisimman virheettömästi.

Kirjallisuus

- Andersson, C. & Nordberg, L. 1998. *A User's Guide to CLAN*. Örebro: Statistiska centralbyrå.
- 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.
- 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.
http://www.jos.nu/Contents/jos_online.asp
- Lundström, S. 1997. *Calibration as a Standard Method for Treatment of Nonresponse*. Doctoral Dissertation. University of Stockholm. Department of Statistics. Stockholm.
- Oh, J. L. & Scheuren, F. 1983. Weighting Adjustment for Unit Nonresponse. Teoksessa: Madow, W. G., Olkin, I. & Rubin, D. B. (toim.). *Incomplete Data in Sample Surveys. Theory and bibliographies*. Vol 2. New York: Academic Press. 143–184.
- Rosenbaum, P. R. 1987. Model-Based Direct Adjustment. *Journal of the American Statistical Association*. Vol. 82. No. 398. 387–394.
- Sautory, O. 2003. *CALMAR2: A new version of the CALMAR calibration adjustment program*. Proceedings of Statistics Canada's Symposium 2003. <http://www.statcan.ca/english/freepub/11-522-XIE/2003001/session13/sautory.pdf>
- Smith, T. M. F. 1991. Post-stratification. *The Statistician*. Vol. 40. No. 3. 315–323.
- Särndal, C.-E., & Lundström, S. 2005. *Estimation in Surveys with Nonresponse*. Chichester: John Wiley & Sons.
- Särndal, C.-E., Swensson, B. & Wretman, J. 1992. *Model Assisted Survey Sampling*. New York: Springer-Verlag.

Lisätietoja

- Bethlehem, J. G. & Keller, W. J. 1987. Linear Weighting of Sample Survey Data. *Journal of Official Statistics*. Vol. 3. No. 2. 141–153.
http://www.jos.nu/Contents/jos_online.asp
- Djerf, K. 2000. *Properties of Some Estimators under Unit Nonresponse*. Statistics Finland. Research Reports 231. Helsinki.
- Kalton, G., & Maligalig, D. S. 1991. *A Comparison of Methods of Weighting Adjustment for Nonresponse. (With discussion)*. The U.S. Bureau of the Census. Proceedings of the Annual Research Conference. 409–447.
- Laaksonen, S. 1988. *Katovirheen korjaus kotitalousaineistossa*. Tutkimuksia 147. Helsinki: Tilastokeskus.
- Sautory, O. & Le Guennec, J. 2005. *La macro CALMAR2. Redressement d'un échantillon par calage sur marges*. I.N.S.E.E. Paris: I.N.S.E.E.

2.12 Tilastollinen estimointi ja analyysi

Soveltamisala ja päämäärä

Tässä luvussa käsitellään kuvailevia ja analyyttispainotteisia tutkimuksia. Kuvailuvissa tilastotutkimuksissa kiinnostuksen kohteena ovat yleensä kokonaismäärät, keskiarvot, osuudet ja vastaavat parametrit koko perusjoukon tasolla tai perusjoukon osajoukoissa. Analyyttispainotteisissa tutkimuksissa sovitetaan usein tilastollisia malleja, kuten lineaarisia regressiomalleja, ja tavoitteena on ilmiöiden välisten yhteyksien ja riippuvuuksien selvittäminen. Molemmista tutkimustyypeissä tärkeä työvaihe on laskea tarvittavat tunnusluvut mahdollisimman luotet-

tavasti. Analyytisissä tutkimuksissa tuloksia tulkitaan usein laajemmin kuin kuvailevissa tutkimuksissa. Tilastovirastoissa tyypillisiä ovat kuvailevat tilastotutkimukset ja niihin liittyvät laskennat, joskin myös erilaisten tilastollisten mallien käyttö on laajentumassa (esimerkiksi erilaisten regressioperusteisten menetelmien yleistymisen myötä).

Tässä luvussa tarkastellaan estimaattoreita ja niiden laatuun liittyviä ominaisuuksia sekä tilastolliseen analyysiin liittyviä laatuäkökohtia. Tilastollisessa estimoinnissa tarkoituksena on saada kohteena oleville tuntemattomille parametreille numeeriset arviot. Estimointiin käytetään otosaineistoja tai hallinnollisista aineistoista ja/tai rekistereistä johdettuja aineistoja. Otosaineistojen ja rekisteriaineistojen yhdistelmien käyttö on myös yleistymässä. Estimoinnissa tarvittavaa laskentamenetelmää tai -algoritmia kutsutaan estimaattoriksi, ja sen avulla saatuja numeerisia arvoja estimaateiksi. Esimerkkinä estimaateista ovat tilaston kokonaismääriä kuvaavat piste-estimaatit.

Käyttäjille tulee raportoida tilaston luotettavuuden arviointiin liittyvistä keskeisistä tekijöistä. Estimaatin luotettavuutta arvioidaan laskemalla käytetyn estimaattorin keskivirhe tai variaatiokerroin. Keskivirheen avulla voidaan laskea luottamusväli eli virhemarginaali (usein 95 % luottamusväli). Esimerkiksi työvoimatutkimuksessa estimoidaan työllisten ja työttömien kokonaismäärät ja työttömyysaste sekä vastaavat keskivirheet ja virhemarginaalit.

Tilastollinen analyysi liittyy aineiston jatkojalostukseen. Esimerkiksi työvoimatutkimuksessa voidaan tutkia soveltuvan tilastollisen mallin avulla työttömänä olemisen yhteyttä ikään, sukupuoleen ja ammattiin. Tilastollista analyysia voidaan käyttää myös tilaston tuotantoprosessin sisäisen laadun arviointiin. Tällöin analyysin tavoitteena on tuottaa prosessiin parannuksia identifioimalla virhelähteitä, arvioimalla niiden merkitystä tietojen luotettavuuden kannalta ja ehdottamalla toimenpiteitä virheiden korjaamiseksi tai eliminoimiseksi.

Yleinen menettelyohje on valita estimointi- ja analyysimenetelmät siten, että havaintoaineiston ominaisuudet tukevat niitä mahdollisimman hyvin. Siten alkuperäisen tutkimus- tai otanta-asetelman piirteiden huomioon ottaminen estimoinnin ja analyysin yhteydessä on välttämätöntä luotettavien tulosten saamiseksi.

Periaatteet

Tärkeimmät estimaattoreiden luotettavuuteen liittyvät tilastolliset ominaisuudet ovat harhattomuus, tarkentuvuus, tarkkuus ja täsmällisyys (Cochran, 1977; Lehtonen ja Pahkinen, 2004; Lohr, 1999). Estimaattori on harhaton (*unbiased*), jos estimaattorin odotusarvo on yhtä kuin parametrin arvo. Tätä lievempi ominaisuus, joka on käytännössä useammin voimassa, on estimaattorin tarkentuvuus (*consistency*). Estimaattori on tarkentuva, jos otoksen kokoa kasvatettaessa odotusarvo lähenee parametrin arvoa ja yhtyy parametrin arvoon, kun otoskoko on sama kuin perusjoukon koko. Täsmällisyys (*precision*) viittaa estimaattien vaihteluun odotusarvonsa ympärillä. Mitä pienempi vaihtelu eli varianssi on, sitä tarkempi on estimaattori. Tätä ominaisuutta kutsutaan myös tehokkuudeksi. Tarkka (*accurate*) estimaattori on sekä harhaton (tai ainakin tarkentuva) että täsmällinen. Tilastokeskuksessa pyritään käyttämään estimoinnissa asetelmaperusteisia estimaattoreita, jotka ovat harhattomia tai ainakin tarkentuvia. Tällaisia ovat

ominaisuuksiltaan esimerkiksi Horvitz-Thompson-tyyppiset estimaattorit, kalibroitimestimaattorit ja malliavusteiset estimaattorit, kuten yleistetyt regressioestimaattorit (Särndal et al., 1992).

Estimaattoreiden tehokkuutta voidaan parantaa hyödyntämällä lisäinformaatiota otannassa tai estimoinnissa. Lisäinformaatiota saadaan eri tilasto- ja rekisterilähteistä. Malliavusteiset menetelmät regressioestimointi ja suhde-estimointi sekä kalibroitimestimaattorit ovat esimerkkejä lisäinformaation hyödyntämisestä estimoinnissa (Särndal et al., 1992; Lehtonen ja Pahkinen, 2004). Tehokkuus paranee, jos lisäinformaatio korreloi voimakkaasti tulosmuuttujan kanssa. Tämä näkyy estimaattorin varianssiestimaatin pienenemisenä. Tehokkuutta mitataan asetelmakertoimella, joka saadaan laskemalla käytettyyn otanta- ja estimointiasetelmaan liittyvän varianssiestimaatin suhde varianssiestimaattiin, joka perustuu vastaavankokoiseen yksinkertaiseen satunnaisotantaan (Kish, 1965, ks. myös luku 2.6). Tehokkuus on parantunut, kun asetelmakerroin on pienempi kuin yksi.

Lisäinformaatiota voidaan hyödyntää 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 kalibroiintiin perustuvaa estimaattoria (Djerf, 2000). Otokseltaan suuria perusjoukon osajoukkoja tarkasteltaessa Horvitz-Thompson-tyyppiset estimaattorit ja kalibroitimestimaattorit tuottavat yleensä riittävän tarkkoja tuloksia (Lehtonen ja Djerf, 2001). Otokseltaan pienten osajoukkojen tapauksissa nämä estimaattorit eivät välttämättä ole riittäviä, jolloin joudutaan käyttämään erityisiä pienalue-estimaattoreita (Rao, 2003; Longford, 2005). Erityisesti malliperusteiset pienalue-estimaattorit voivat olla otanta-asetelman suhteen harhaisia, mutta tuottavat niin pienen varianssin, että niiden käyttäminen on perusteltua.

Paneeliasetelmiin, kuten rotaatiopaneeleihin ja koordinoituihin otoksiin liittyviä erityispiirteitä on mahdollista käyttää esimerkiksi ajallisen muutoksen estimoinnin parantamiseksi (Kasprzyk et al., 1989; Diggle et al., 2002). Myös näissä yhteyksissä voidaan estimoinnissa hyödyntää lisäinformaatiota. Tämä koskee myös paneeliasetelmaa hyödyntäviä otostutkimuksia, joissa esiintyy erityistä vastauskatoa, paneeliatritiota (*panel attrition*), joka tapahtuu siirryttäessä paneelin vaiheesta seuraavaan.

Myös analyttisissä tutkimuksissa käytetään tilastollisten mallien parametreille harhattomia tai tarkentuvia estimaattoreita. Mutkikkaisiin otanta-asetelmiin (ks. luku 2.6) perustuvien aineistojen analyysissä harhattomuus mallin suhteen ei yleensä ole riittävää, vaan lisäksi tarvittava ominaisuus on harhattomuus (tai tarkentuvuus) otanta-asetelman suhteen (Chambers ja Skinner, 2003).

Suositukseset

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

teisia menetelmiä. Niiden avulla voidaan estimoinnin ja analyysin yhteydessä ottaa huomioon esimerkiksi aineiston painotus sekä otanta-asetelman ominaisuudet kuten ositus ja ryvästyminen.

Otosaineistot

- Käsiteltävässä otosaineistossa on oltava estimointia ja analyysia varten riittävät otanta-asetelmaa koskevat tiedot (esimerkiksi muuttujat asetelmapainolle ja analyysipainolle sekä tarvittaessa ositukselle ja ryvästykseksi).
- Alkuperäinen otosasetelma on säilytettävä muuttumattomana estimointiasetelmaa luotaessa. Siten alkuperäiset tiedot osituksesta, ryvästyksestä, moniasteisuudesta ja -vaiheisuudesta, samoin kuin tiedot todennäköisyydellä yksi poimituista alkioista, on pidettävä aineistossa ennallaan. Nämä tiedot on myös dokumentoitava aineiston käyttäjiä varten. Aineistoa tulee aina estimoida tutkimuksen tilanteen mukaisina osajoukkoina rikkomatta ositusrakennetta.
- Otoksen ulkopuolista lisäinformaatiota on syytä hyödyntää estimoinnissa aina, kun sitä on luotettavasti saatavissa. Käyttämällä lisäinformaatiota estimoinnissa saadaan otosjakaumat lähemmäksi perusjoukon jakaumia sekä usein tehokkuushyötyjä. Menetelmää valittaessa on hyödyllistä neuvotella menetelmäasiantuntijan kanssa.
- Piste-estimaattien luotettavuutta arvioidaan keskivirheillä tai variaatiokertoimilla (suhteellinen keskivirhe) siten, että alkuperäinen otosasetelma ja siitä johdettu estimointiasetelma otetaan mahdollisimman tarkasti huomioon. On hyödyllistä estimoida asetelmakertoimia aina kun se on mahdollista.
- Estimaattoreiden otosvariansseja ei yleensä voida johtaa analyttisesti monimutkaisten otanta-asetelmien tapauksessa sekä tilanteissa, joissa estimoidaan epälineaarisen estimaattorin (esimerkiksi regressiokerroin) varianssia. Tällöin joudutaan turvautumaan likimääräisiin menetelmiin eli approksimaatioihin. Tilasto-ohjelmissa tavallisin approksimointimenetelmä on linearisointimenetelmä; muita ovat otoksen uudiskäyttökäytännöt, kuten jackknife, tasapainotetut puoliotokset tai bootstrap (Wolter, 1985; Lohr, 1999; Lehtonen ja Pahkinen, 2004). Approksimointimenetelmää valittaessa on syytä neuvotella menetelmäasiantuntijan kanssa.
- Otosaineistoissa on usein niin kutsuttuja poikkeavia havaintoja, joiden arvot poikkeavat selvästi muiden havaintojen arvoista. Jos havainnot ovat oikeita, ne pidetään ennallaan. Tällöin piste-estimoinnissa voidaan soveltaa keskiarvojen sijasta robusteja menetelmiä, kuten mediaaneja ja leikattuja keskiarvoja (*trimmed means*). Jos poikkeavuus aiheutuu painotuksesta, on painojen vaihtelua syytä rajoittaa.
- Vastauskato on pyrittävä ottamaan huomioon estimoinnissa. Yksikkökadon vaikutusta pyritään usein pienentämään esimerkiksi uudelleenpainotusmenetelmillä, esimerkiksi kalibroitimestimaattorin avulla (ks. luku 2.11). Eräkadon tapauksessa, ts. kun hyväksytyissä havainnoissa on puuttuvia tietoja joidenkin muuttujien osalta, puuttuvia havaintoarvoja imputoidaan soveltuvin menetelmin (ks. luku 2.10). Estimaattoreiden otosvarianssien laskennassa on pyrittävä ottamaan huomioon myös yksikkö- ja eräkadon oikaisun aiheuttama lisävariaatio.

- Asetelmaperusteiseen tilastolliseen analyysiin soveltuvat esimerkiksi SUDAAN-ohjelmisto sekä SAS-ohjelmiston proseduurit SURVEYREG ja SURVEYLOGISTIC ja eräät Stata-ohjelmiston proseduurit. Analyysipainojen (ks. luku 2.6) käyttö tilastollisen analyysin yhteydessä on suositeltavaa.

Rekisteriaineistot ja aikasarja-aineistot

- Rekisteriperusteisten kokonaisaineistojen tilastolliseen käsittelyyn liittyviä seikkoja on selvitetty luvussa 2.5 (Hallinnolliset rekisterit ja tilastorekisterit) (ks. lisäksi Wallgren ja Wallgren, 2007).
- Aikasarja-aineistoihin liittyviä estimointi- ja analyysiasioita on selvitetty luvussa 2.14 (Aikasarjat ja kausitasoitus).

Dokumentointi

Hyvään tilastointitapaan kuuluu dokumentoida koko havaintoaineiston käsittelyyn liittyvä tuotantoketju. Estimoinnin ja analyysin osalta tämä merkitsee käytettyjen estimaattoreiden, mallien, estimointimenetelmien ja ohjelmistojen kuvaamista.

Hyvä tilastointitapa edellyttää riittävää raportointia julkaistujen estimaattien luotettavuudesta käyttäjille. Mahdollisimman kattava analyysi tilastojen ja tulosten luotettavuudesta on dokumentoitava, myös mahdollisia myöhempiä tietotarpeita varten. Käyttäjille suunnatut laatu- ja menetelmäselosteet ovat tähän tarkoitukseen hyvin soveltuva julkaisukanava.

Keskeisten tulosmuuttujien piste-estimaateista ja niiden luotettavuudesta on raportoitava koko aineistossa ja sen tärkeimmissä osajoukoissa. Piste-estimaattien, keskvirhe-estimaattien ja luottamusvälien ohella asetelmakertoimien raportointi on suositeltavaa. Otosaineistojen kohdalla myös on arvioitava piste-estimaatteihin mahdollisesti liittyvää harhaa ja analysoitava estimaattien täsmällisyyttä.

Kirjallisuus

Tilastollinen estimointi

- Cochran, W. G. 1977. *Sampling Techniques*. 3rd ed. New York: John Wiley & Sons.
- Djerf, K. 2000. *Properties of Some Estimators under Unit Nonresponse*. Statistics Finland. Research Reports 231. Helsinki.
- Kasprzyk, D., Duncan, G., Kalton, G. & Singh, M. P. (toim.). 1989. *Panel Surveys*. New York: John Wiley & Sons.
- Kish, L. 1965. *Survey Sampling*. New York: John Wiley & Sons.
- Lehtonen, R. & Djerf, K. (toim.) 2001. *Lecture Notes on Estimation for Population Domains and Small Areas*. Statistics Finland. Reviews 2001/5. Helsinki.
- Lehtonen, R. & Pahkinen, E. 2004. *Practical Methods for Design and Analysis of Complex Surveys*. 2nd Edition. Chichester: John Wiley & Sons.
- Lohr, S. 1999. *Sampling: Design and Analysis*. Pacific Grove: Duxbury.
- Longford, N.T. 2005. *Missing Data and Small-Area Estimation: Modern Analytical Equipment for the Survey Statistician*. New York: Springer-Verlag.
- Rao, J.N.K. 2003. *Small Area Estimation*. Hoboken: John Wiley & Sons.
- Särndal, C.-E., Swensson, B. & Wretman, J. 1992. *Model Assisted Survey Sampling*. New York: Springer-Verlag.
- Wallgren, A. & Wallgren, B. 2007. *Register-based Statistics: Administrative Data for Statistical Purposes*. Chichester: John Wiley & Sons.
- Wolter, K. 1985. *Introduction to Variance Estimation*. New York: Springer-Verlag.

Tilastollinen analyysi

- Chambers, R.L. & Skinner, C.J. (toim.) 2003. *Analysis of Survey Data*. Chichester: John Wiley & Sons.
- Diggle, P. J., Heagerty, P., Liang, K.-Y., Zeger, S. L. (2002). *Analysis of Longitudinal Data. 2nd Edition*. Oxford: Oxford University Press.
- Lehtonen, R. & Pahkinen, E. 2004. *Practical Methods for Design and Analysis of Complex Surveys*. 2nd Edition. Chichester: John Wiley & Sons.
- Lohr, S. 1999. *Sampling: Design and Analysis*. Pacific Grove: Duxbury Press.

Lisätietoja

- Goldstein, H. 2003. *Multilevel Statistical Models*. 3rd Edition. New York: John Wiley & Sons.

2.13 Tilastotiedon esittäminen

Soveltamisala ja päämäärä

Tilastotiedon esittäminen ja julkistaminen ovat arkistoinnin kanssa viimeisiä vaiheita tilastotutkimusten operationaalisessa 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 painetuissa kuin sähköisissä esityksissä (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 on hidasta ja epähavainnollista.

Kuvioiden, taulukon ja tekstin roolit menevät hieman 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.

Periaatteet

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 taulukkopääksi.

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

Tämän lisäksi taulukoissa voi käyttää apuviivoja helpottamaan taulukon lukua.

Tilastokuvio

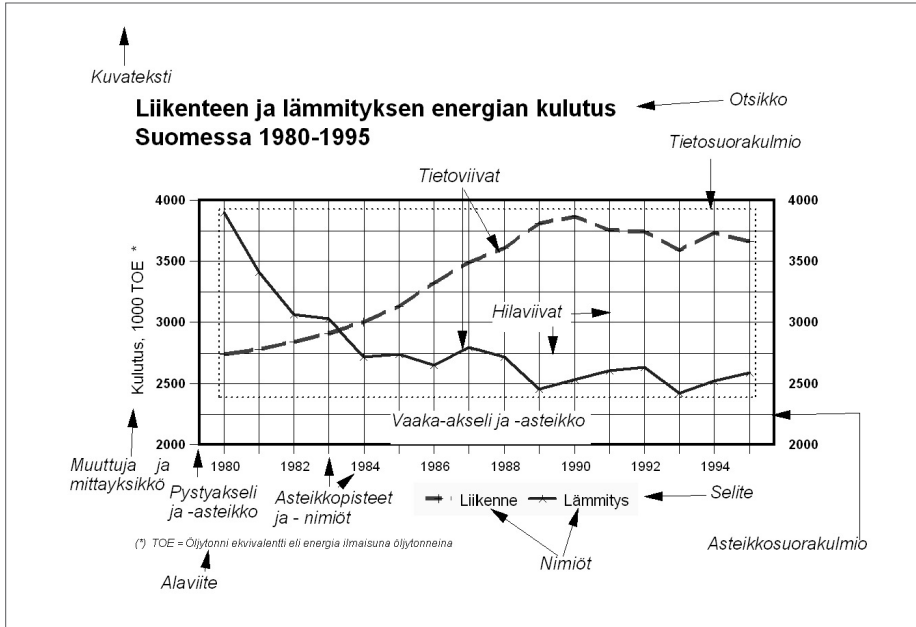
Toisin kuin taulukon, tilastokuvion tärkein tehtävä on tiedon välittäminen visuaalisesti. Kuviodien tekemisessä on vaarana tahaton vääristely, ellei tunneta tekniikkaa. Kuvion laatijan vastuulla on, että kuvio antaa lukijalleen oikean käsityksen asioiden todellisesta luonteesta, eli välittää oikeaa tietoa.

Tilastokuvio esittää aineiston rakenteellisia ominaisuuksia, tiivistää suuria tietomääriä, osoittaa asioiden keskinäistä kytkentää, välittää ajatuksia ja johtopäätöksiä sekä luo asiayhteyden haluttua tunnelmaa paremmin kuin taulukko.

Kuvion osat

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

Kuvio 2.6 Tilastokuvion osat ja termistö



- Koordinaattiakselien asteikkomerkintöjen tulee olla selkeitä ja riittävän suurella kirjasinkoolla tehtyjä.
- Kaikki kuvioelementit (esimerkiksi viivat, osat, sektorit) on nimettävä.
- Kuvion tulkintaa ja johtopäätöksiä helpottavia apuvälineitä tulee käyttää riittävästi, mutta ei häiritsevän paljon.
- Mahdolliset poikkeavuudet on tuotava esiin ja selvitettävä.

Kuviotyypit

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 ovat 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.

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

Piirakkakuviossa ympyrä jaetaan sektoreihin ja sektoreiden pinta-ala välittää määrällisen tiedon. Tämä on otettava huomioon silloin, kun piirakoita sijoite-

taan 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. Piirakkakuvioita ei pidä käyttää, jos halutaan välittää tietoa tarkasti. Rakenteellisen epätarkkuutensa vuoksi piirakkakuvion yhteydessä on perusteltua käyttää myös lukuarvoja, jotka kertovat prosenttijakauman.

Tilastolliset teemakartat ovat karttapohjalle tehtyjä tilastokuvioita. 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 koristella esitystä. 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 populistisissa esityksissä niiden mielenkiintoa herättävän ja dramaattisen vaikutuksen vuoksi.

Symbolihahmokuva on niin vaikeasti hallittava kuvioityyppi, ettei sitä tulisi käyttää ollenkaan. Tilastografiikan periaatteiden mukaan oikein tehtyjä symbolikuvioita on lähes mahdoton tehdä.

Kuvioiden näennäistä kolmiulotteisuutta suositaan luultavasti samasta syystä kuin symbolihahmokuviotakin, eli sillä saadaan vaatimatonkin asia esitettyä vaikuttavasti. Vain harvoissa tapauksissa kuvioon tuotettu lisäulottuvuus kantaa informaatiota. Yleensä kyseessä on esityksen koristelu, joka lisäksi haittaa tiedon välittymistä. Useimmiten kolmiulotteisuus on kuvioroinaa.

Kolmiulotteisuuden käyttö pitäisi rajoittaa vain niihin tapauksiin, joihin kolmas ulottuvuus todella antaa lisäinformaatiota.

Suosituksset

1. Taulukot

- Lukuja ei pidä esittää liian monella merkitsevällä numerolla, ja pyöristys on aina suoritettava matemaattisesti oikein.
- Sijoita vertailtavat asiat päällekkäin, koska taulukossa lukuja on helpompi verrata sarakkeittain kuin riveittäin.
- Esitä rivien ja sarakkeiden keskiarvot/summat silloin, kun se on loogista ja helpottaa vertaamista.
- Taulukoissa esitettävien rivien ja sarakkeiden järjestyksen on oltava looginen. Joskus paras tulos saavutetaan järjestämällä ensimmäisen sarakkeen ja rivin luvut suuruusjärjestykseen.
- Valittavat asiat sijoitetaan päällekkäin, koska taulukossa lukuja on helpompi verrata sarakkeittain kuin riveittäin.
- Taulukon visuaalisen asettelun tulee olla selkeä.

- Taulukkoja varten on laadittu joukko vakiintuneita merkintöjä erikoistilanteita kuten puuttuvaa tietoa varten. Malleja löytyy esimerkiksi Suomen tilastollisesta vuosikirjasta.

II. Tilastokuviot

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 tietomäärän ja toisaalta, vain muutamasta luvusta ei kannata tehdä kuviota.
- 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 yhä tarkemmin kuviota löytää vielä syvällisempää tietoa.
- Saavat katsojan kiinnostumaan kuvion sanomasta, ei kuviosta sinänsä. On muistettava, että katsoja todennäköisesti on kiinnostuneempi asiasta kuin esitystekniikasta.

Viivakuviassa

- Vaaka-akselilla on oltava ominaisuus, jolla on jatkuva mitta-asteikko, esimerkiksi aika.
- Määräasteikon, eli pystyakselin asteikon, ei ole välttämättä alettava nollassa.
- 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, esimerkiksi aika.
- Pystyakselin asteikkoa (määräasteikkoa) ei saa katkaista, vaan sen on aina alettava nollassa.
- 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 pylviäitä 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 vasta-päivään.
- 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 esitysfoorumilla 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 olisikaan 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 lähinnä markkinoinnin tarkoituksiin.

Käytä harkiten kolmatta ulottuvuutta

Kolmiulotteinen esitystapa on perusteltua vain siinä tapauksessa, että kolmanneen 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 vaikuuttelua. Taulukko on hyvä suurten tietomäärien esittämiseen.

Kirjallisuus

Tilastokeskuksen toimintaohjeet

Kuusela, V. 1994. Hyvin laadittu tilastokuvio. Teoksessa: Melkas, J. (toim.). *Tilastokeskuksen ulkoisen tiedottamisen käsikirja*. Helsinki: Tilastokeskus. 23–24.

Lisätietoja

Kuusela, V. 2000. *Tilastografiikan perusteet*. Helsinki: Tilastokeskus & Oy Edita Ab.

Tilastokeskus. 2006. *Suomen tilastollinen vuosikirja 2006*. 101. vuosikerta. Suomen virallinen tilasto (SVT). Tilastokeskus.

Tufte, E. R. 1983. *Visual Display of Quantitative Information*. Cheshire: Graphics Press.

Wainer, H. 1997. *Visual Revelations: Graphical tales of fate and deception from Napoleon Bonaparte to Ross Perot*. New York: Copernicus.

2.14 Aikasarjat ja kausitasoitus

Soveltamisala ja päämäärä

Monia taloudellisia ja yhteiskunnallisia ilmiöitä mitataan säännöllisin väliajoin. Säännöllisesti ilmestyviä talouden tunnuslukuja ovat esimerkiksi työttömyysprosentti ja asuntojen hinnat. Niissä kiinnostavaa ei ole vain muuttujan kunakin ajanhetkenä saama arvo, vaan myös muutos verrattuna edellisiin havaintoihin.

Aikasarja määritellään joukkona ajan suhteen järjestettyjä havaintoja. Työttömyyttä kuvaava aikasarja saadaan, kun eri ajankohtina lasketut työttömyysluvut asetetaan aikajärjestykseen. Tällöin on tärkeää, että työttömyyttä on kaikkina ajankohtina mitattu samalla tavalla, jotta havaintojen vertailu aika-akselilla olisi mielekäästä.

Tarkastelemalla työttömyyttä kuvaavaa aikasarjaa, saadaan kuva työttömyyden kehityksen trendistä, eli pitkän aikavälin muutoksen suunnasta. Samalla voidaan huomata, että työttömyysluvut vaihtelevat vuodenaikojen mukaan. Tätä kausityöttömyydestä johtuvaa aaltomaista liikehdintää kutsutaan kausivaihteluksi.

Kausitasoituksessa (*seasonal adjustment*) aikasarjasta poistetaan vuoden sisäinen kausivaihtelu, jotta sarjan yleisestä kehityksestä saadaan parempi kuva (Gomez ja Maravall, 1996). Kausivaihtelulla tarkoitetaan aikasarjan toistuvaa, vuoden sisäistä vaihtelua. Kausivaihtelu johtuu vuodenaajoista, sekä muista vuosirytmiiin liittyvistä ilmiöistä, kuten lomista, juhlapyhistä ja kulutustottumuksista. Siksi kausivaihtelu on jokseenkin säännöllistä.

Kausitasoitettujen sarjan yleistä kehitystä ja suhdannevaihteluiden käyttäytymistä on helppo tulkita. Kausitasoitus tekee myös eri kuukaudet keskenään vertailukelpoisiksi (Kokkinen ja Alshail, 2005; Bell ja Hillmer, 1984). Tilastokes-

kuksessa käytetään kausitasoitusta muun muassa teollisuustuotannon volyyminindeksiin, työvoimakustannusindeksiin sekä kansantalouden tilinpidon aikasarjoissa.

Kausitasoitusmenetelmiä on monenlaisia. Niin sanotut ad-hoc suotimet perustuvat pitkälle liukuvan keskiarvon menetelmiin. X-12-ARIMA on esimerkki ad-hoc suotimesta (Findley et al., 1998). TRAMO/SEATS on puolestaan mallipohjainen kausitasoitusmenetelmä (Gomez ja Maravall, 1996). TRAMO/SEATS:in ja X-12-ARIMA:n parhaiden puolien yhdistäminen on ajankohtainen tutkimuskysymys (Findley, 2005).

Periaatteet

Kausitasoitus perustuu periaatteeseen, jonka mukaan aikasarja voidaan esittää oheisten komponenttien summana.

1. Trendi (*trend*) kuvaa sarjan pitkän aikavälin vaihtelua.
2. Suhdannevaihtelu (*cycle*) kuvaa aikasarjan vaihtelua, joka johtuu esimerkiksi taloussuhdanteista.
3. Kausivaihtelu (*seasonal variation*) on vuoden sisäistä vaihtelua, joka toistuu vuodesta toiseen miltei säännöllisenä.
4. Epäsäännöllistä vaihtelua (*irregular variation*) ei voida ennustaa eikä sijoittaa edellä mainittuihin kolmeen komponenttiin. Epäsäännölliseen vaihteluun kuuluvat myös muutamat erikoistapaukset, kuten lakon aiheuttamat poikkeuksellisen pienet tai suuret arvot.

Trendi ja sykli ovat niin lähellä toisiaan, että niiden erottaminen toisistaan yksikäsitteisellä ja selkeällä tavalla on hankalaa. Tämän vuoksi niitä käsitellään yhdessä ja puhutaan lyhyesti vain trendistä (Kokkinen ja Alsuhail, 2005; Kaiser ja Maravall, 2001). Kausitasoitettu sarjasta on siis poistettu kausikomponentti. Tällöin kausitasoitettu sarja koostuu vain trendistä ja epäsäännöllisestä vaihtelusta. Tilastokeskus esittää julkaisuissaan kausitasoitettua aikasarjan lisäksi myös trendin. Tilastokeskuksen julkaisemissa aikasarjoissa siirryttiin vuonna 2005 käyttämään mallipohjaista TRAMO/SEATS-menetelmää.

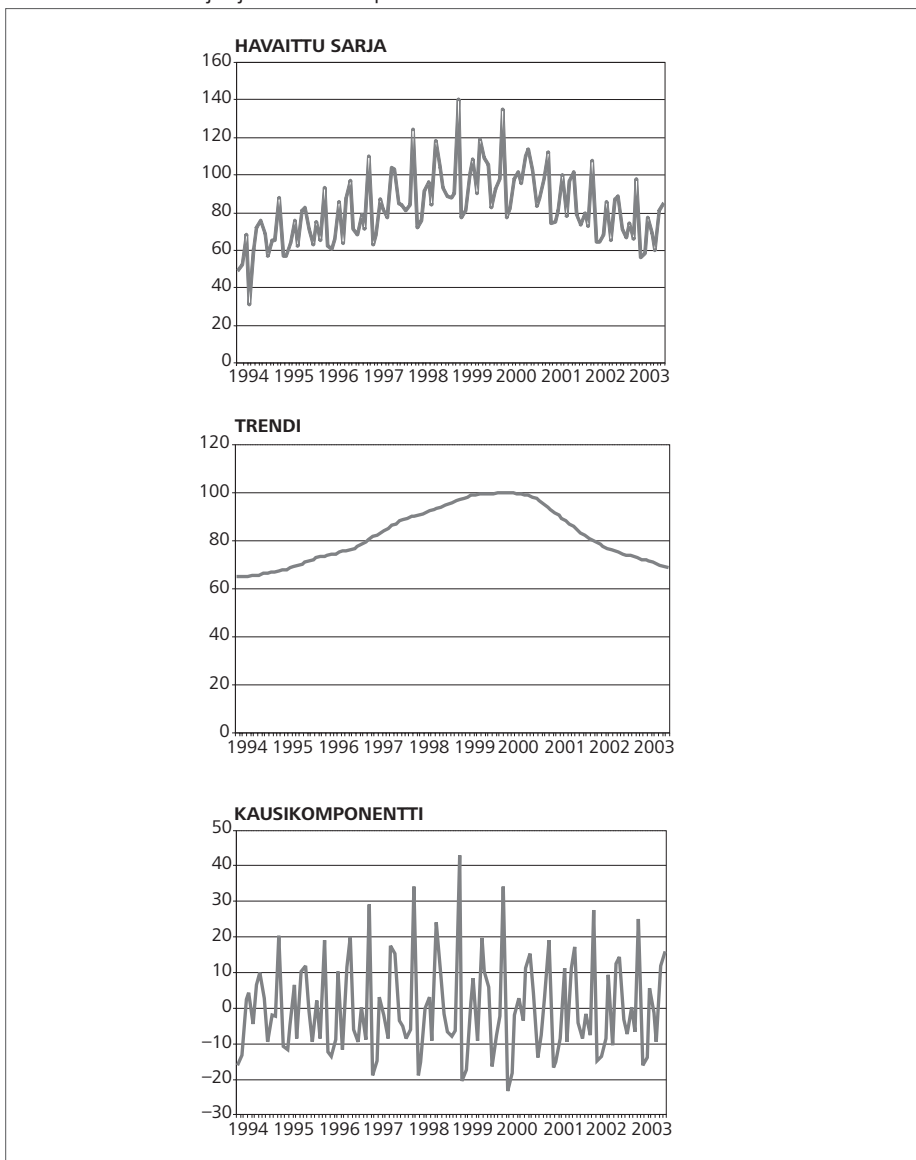
Ennen sarjan varsinaista kausitasoitusta TRAMO/SEATS-menetelmällä on sarjaan tehtävä tarvittavat työpäivä- ja juhlapyhäkorjaukset sekä mahdolliset lakoista tai muista erikoistapauksista johtuvat äärihavaintojen korjaukset. Tämän jälkeen esipuhdistetulle aineistolle sovitetaan aikasarjamalli (*time series model*). Tällä tarkoitetaan havaintojen aika-akselilla esiintyvän riippuvuuden kuvaamista matemaattisen yhtälön avulla. Mallin sovitus koostuu oikean mallin valinnasta ja sen parametrien estimoinnista. Tavoitteena on kuvata havaintojen riippuvuus tarkasti, mutta mahdollisimman yksinkertaisella tavalla. Saadun mallin avulla voidaan esimerkiksi tutkia aikasarjan jaksollisuutta ja laatia ennusteita (Hamilton, 1994).

TRAMO/SEATS -menetelmässä esipuhdistettuun aineistoon sovitetaan lineaarinen ARIMA-malli (Box ja Jenkins, 1970; Box, Jenkins ja Reinsel, 1994), johon trendin ja kausikomponentin estimointi perustuu (Gomez ja Maravall, 1996). Mikäli aineisto omaa esimerkiksi epälineaarisia piirteitä, voi suoritettu esipuhdistus olla tulkinnaltaan epäintuitiivinen. Toisinaan ilmenee tapauksia,

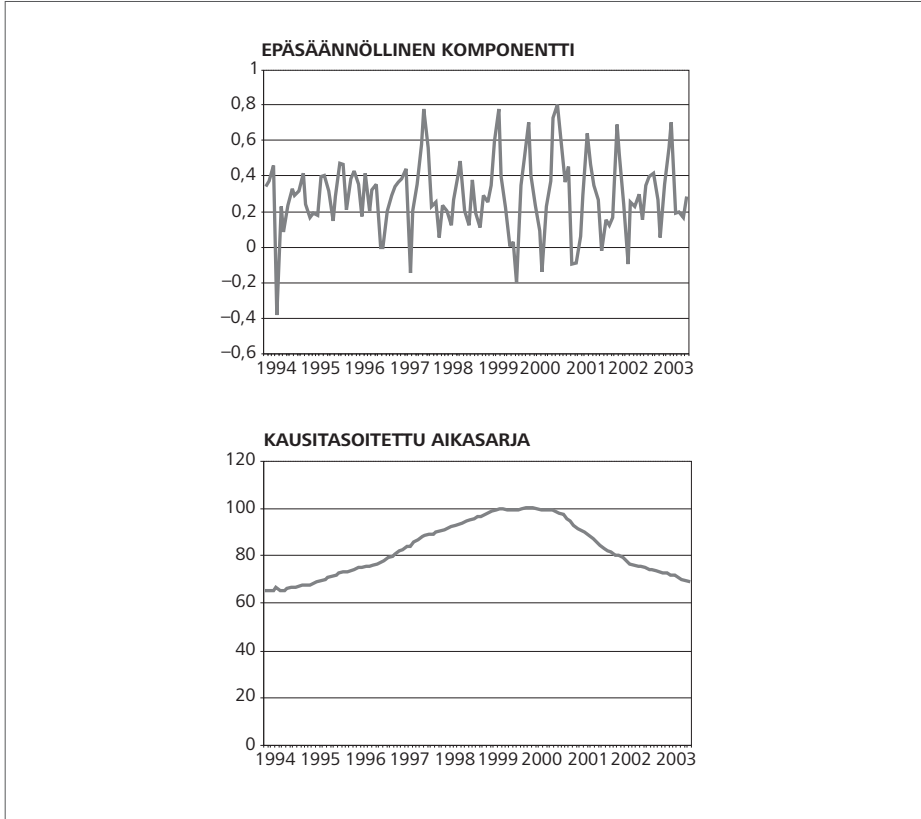
joissa visuaalisesti miellyttävä kausitasoitus ei ole diagnostisessa mielessä onnistunut ja päinvastoin (Alsuhail, 2005). Tällöin valinta tulee tehdä tilaston ensisijaista käyttötarkoitusta silmälläpitäen.

Kuviossa 2.7 havainnollistetaan aikasarjan jakautumista eri komponentteihin. Ensimmäinen esittää havaittua aikasarjaa, joka koostuu siis trendin, kausikomponentin ja epäsäännöllisen vaihtelun summasta. Trendi kuvaa aikasarjan pitkän aikavälin kehitystä ja kausikomponentti on miltei säännöllinen vuodesta toiseen. Kausitasoitettu sarja sisältää trendin lisäksi epäsäännöllisen jäännössarjan, minkä vuoksi se on trendiä sahalaitaisempi.

Kuvio 2.7 Aikasarjat ja niiden komponentit



Kuvio 2.7 jatkuu



Suosituksset

- Tilastokeskuksen julkaisemissa aikasarjoissa tulee käyttää TRAMO/SEATS-menetelmää. Kausitasoituksessa käytetään Demetra-ohjelmistoa.
- Aikasarjan esipuhdistuksessa tulee huomioida pyhäpäivien sijoittuminen ja käyttää aina Suomen arkipyhät huomioivaa kalenteria.
- Trendin ja kausitasoitettun aikasarjan tuoreimmat arvot saattavat muuttua, kun aikasarjaa päivitetään uusilla havainnoilla. Tämä johtuu siitä, että aikasarjan loppupäälle estimoidut trendi- ja kausikomponentti perustuvat osittain ennustettuihin aikasarjan arvoihin. Tämä epävarmuus tulee tuoda ilmi tilastojulkaisuissa.
- Mikäli tutkittavan tilastosuureen määritelmässä tai mittaamistavassa on tapahtunut muutos, jonka vuoksi aikasarjan havainnot eivät ole ajallisesti vertailukelpoisia, tulee siitä ilmoittaa erikseen.
- Kausitasoituksen onnistumista tulee arvioida sekä visuaalisesti että diagnostisten testien avulla.
- Tilastokeskuksen julkaisuissa ei tule esittää kuvioita, joiden lukuarvoja ei löydy julkaisun taulukoista.

Kirjallisuus

- Alsu hail, F. 2005. *Kausivaihtelun poistaminen taloudellisista aikasarjoista*. Tilastokeskus. Sisäinen muistio. Helsinki.
- Box, G. E. P. & Jenkins, G. M. 1970. *Time Series Analysis Forecasting and Control*. San Francisco: Holden-Day.
- Box, G. E. P., Jenkins, G. M. & Reinsel, G. C. 1994. *Time series analysis: forecasting and control*. 3rd edition. New Jersey: Prentice-Hall.
- Bell, W. R. & Hillmer, S. C. 1984. Issues Involved With the Seasonal Adjustment of Economic Time Series. *Journal of Business & Economic Statistics*. Vol. 2. No. 4. 98–127.
- Findley, D. F., Monsel, B. C., Bell, W. R., Otto, M. C. & Chen, B. 1998. New Capabilities and Methods of X-12-ARIMA Seasonal Adjustment Program. *Journal of Business & Economic Statistics*. Vol. 16. No. 2. 127–152.
- Findley, D. F. 2005. Some Recent Developments and Directions in Seasonal Adjustment. *Journal of Official Statistics*. Vol. 21. No. 2. 343–365.
- Gomez, V. & Maravall, A. 1996. *Programs TRAMO and SEATS. Instructions for the User*. Banco de Espana – Servicio de Estudios.
- Hamilton, J. D. 1994. *Time Series Analysis*. Princeton: Princeton University Press.
- Kaiser, R. & Maravall, A. 2000. *Notes on Time Series Analysis, ARIMA Models and Signal Extraction*. Working Papers 0012. Banco de Espana. (Research report available from Internet).
- Kokkinen, A. & Alsu hail, F. 2005. ARIMA-mallipohjaisesta kausitasoituksesta. *Kansantaloudellinen aikakauskirja*. Vol. 101. No. 4. 469–483.

Lisätietoja

Statistical Office of the European Communities. 2002. *Final Report. Task Force on Seasonal Adjustment of Quarterly National Accounts*. European Commission, Eurostat, European Central Bank.

2.15 Indikaattorit ja indeksit

Soveltamisala ja päämäärä

Mittaamalla esimerkiksi samaa taloudellista ilmiötä ajassa saadaan tilastoaineisto, jota kutsutaan aikasarjaksi. Tyypillisiä aikasarjoja ovat esimerkiksi hyödykkeiden hinnat ja määrät, työntekijöiden palkat ja tehdyt työtuntien määrät sekä arvopapereiden hinnat ja määrät. Aikasarja on pelkistetyksi jono lukuja, jotka saadaan mittaamalla täsmällisesti kontrolloitua samaa ilmiötä ajassa. Aikasarjat kuvaavat siis tämän ilmiön kehitystä ajassa (ks. luku 2.14).

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' (Törnqvist, 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 aikasarjahavainto vaikuttaa kokonaisuuteen. Mikäli suhteellinen painotus muuttuu jatkuvasti, aggregoidut aikasarjahavainnot eivät ole ajallisesti vertailukelpoisia. 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 samanlaatuisten ja mittayksiköistä vapaiden suhdelukujen aggregoinnista. Indeksilaskenta ratkaisee laadullisesti erilaisten hyödykkeiden hinta- ja määräaikasarjojen 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 kulutusrakenteet vakioidaan perus- ja/tai vertailuajankohdan mukaisiksi. Indeksilaskennalla pyritään vastaamaan seuraavaan kysymykseen: Paljonko enemmän tai vähemmän perusajankohdan (vertailuajankohdan) hyödykkeistä on maksettava vertailuajankohdalla (perusajankohdalla) perus- ja/tai vertailuajankohdan kulutustottumuksilla.

Tyypilliset indeksit ovat yleensä joko hintasuhteiden tai määräsuhteiden painotettuja keskiarvoja. Painorakenteet määritellään siten, että ne edustavat joko perus- ja/tai vertailuajankohdan tilannetta. Tämä indeksilaskennan metodi tarkoittaa esimerkiksi hinnanmuutoksen arvioinnissa sitä, että vaikka hinnat ja määrät muuttuvat samanaikaisesti, ei määrien muutoksen sallita sekoittaa tavoitteena olevan hinnanmuutoksen mittaamista.

Indeksiteoria jakaantuu pelkistetysti kuvailevaan (deskriptiiviseen) ja analyttiseen (talusteoreettiseen) tutkimussuuntaukseen. Analyttinen tutkimus perustuu lähtökohtaan, jossa taloustoimijoiden (kuten kuluttajien ja yritysten) oletetaan käyttäytyvän rationaalisesti matemaattisen käyttäytymismallin mukaan. Analyttisessä tutkimussuuntauksessa ongelmaksi muodostuu se, että taloustoimijat voivat muuttaa käyttäytymistään yli ajan – esimerkiksi sosioekonomisen aseman tai vaikkapa makutottumusten muutoksen seurauksena. Usein myös taloudelliset realiteetit, kuten kulutusmarkkinoille tulevat uudet ja sieltä poistuvat hyödykkeet (esimerkkeinä tietokoneet, matkapuhelimet, kirjoituskoneet ja höyrykoneet), on vaikeaa liittää taloustoimijoiden käyttäytymismalleihin. Vastaavanlaiseen tilanteeseen päädytään uusien ja häviävien työpanosten osalta. Talusteoriaan perustuvat indeksit ovat teoriasidonnaisia ja usein matemaattisesti vaativia. Lähestymistavan perusajatuksista ovat kirjoittaneet esimerkiksi Vartia (1976), Samuelson ja Swamy (1974), Allen (1975) ja Diewert (1974, 1975).

Deskriptiivinen tutkimussuuntaus hyödyntää indeksilaskennassa tilastoaineiston tyypillisimpiä tunnuslukuja (esimerkiksi keskiarvo, geometrinen keskiarvo, keskiverto ja varianssi). Se ei edellytä teoreettisen viitekehysten määrittämistä taloustoimijoiden joukolle. Pelkistetysti deskriptiivisessä lähestymistavassa indeksi määritellään siten, että se mittaa esimerkiksi hinta- tai palkkatasojen muutosta tai yksittäisten hintojen tai palkkojen keskimääräistä muutosta. Des-

kriptiivisistä lähestymistavasta ovat kirjoittaneet Vartia (1976, 1983, 1995) sekä Vartia ja Vartia (1984).

Periaatteet

Indeksit ovat suhteellisen muutoksen indikaattoreita, jotka kertovat paljonko esimerkiksi määrätyn hyödykkeen tai hyödykeryhmän (hyödykekori) hinnat ja/tai määrät ovat suhteellisesti muuttuneet perusajankohdasta vertailuajankohdastaan. Indeksilaskennassa lähtökohdan muodostavat homogeeniset hyödykkeet ja niihin hyödykkeisiin liittyvät otannalla selvitettyt yksittäiset hinta-, määrä- ja arvosuhteet. Homogeenisen hyödykkeen (Emmental-juusto) tai homogeenisista hyödykkeistä muodostuvan hyödykeryhmän (esimerkiksi 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 määrätiedolla.

Yksikköarvoindeksi

Yksikköarvoindeksi mittaa keskimääräisten hintojen muutosta ajassa. Yksikköarvoindeksi esittää esimerkiksi kahden rakenteellisesti erilaisen hyödykekorin hintasuhteen perus- ja vertailuajankohdan välillä. Koska perus- ja vertailuajankohdtien 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 vertailuajankohdtien keskiarvot on mitattu tarkasti. Toisin sanoen, vaikka perus- ja vertailuajankohdina arvioitujen keskiarvojen keskivirheet olisivat nolliä (keskiarvo vastaa perusjoukon keskiarvoa), se ei tee keskiarvojen vertailusta hyvää indeksiä, koska perus- ja vertailuajankohdtien hyödykekorit ovat edelleen rakenteellisesti erilaisia. Yksikköarvoindeksin virhe syntyy siihen sisältyvän implisiittisen volyyymi-indeksin heikkouksista. Erilaatuisten 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 että määrät muuttuvat samanaikaisesti perus- ja vertailuajankohdilla. Tällöin vertailtavat hyödykekorit ovat luonnollisesti rakenteellisesti erilaisia, joten esimerkiksi keskimääräisten hintojen suorassa vertailussa ei voida tietää sitä, mikä osa hinnanmuutoksesta on todellista ja mikä rakenteellisen muutoksen aikaansaamaa. Ongelman ratkaisemiseksi on indeksilaskennassa kehitetty erilaisia hinnan- ja määränmuutosten mittareita, kuten Laspeyresin, Paaschen, Edgeworthin ja Fisherin indeksit, joissa vertailtavat hyödykekorit määritellään rakenteellisesti samanlaisiksi. Laspeyresin hintaindeksiä käytettäessä valitaan lähtökohdaksi perusajankohdan kulusrakenne ja konstruoidaan ver-

tailuajankohdan hypoteettinen kulutusmeno perusajankohdan hyödykekorille vertailuajankohdan hinnoilla. Indeksi 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 volyymi ja hinnat ovat muuttuneet perusajankohdasta vertailuajankohtaan nähden. Mikäli hinnat ja volyymien muutokset ovat korreloimattomia, indeksiharha häviää.

Toisessa yleisesti käytetyssä indeksissä, jota kutsutaan Paaschen indeksiksi, kulutusrakenne vakioidaan vertailuajankohdan kulutuksen mukaiseksi. Indeksi kertoo, kuinka paljon enemmän tai 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 kulutuskoritulkintaa.

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 Laspeyresiin ja Paascheen 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. Indeksit ovat näiden argumenttien suhteen ajan funktioita. Indeksit huomioivat kulutustottumuksissa tapahtuneet muutokset, joten ne ovat kiinteäpainoisia indeksejä (Laspeyres ja Paasche) tarkempia.

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. Ainakin seuraavat seikat komplisoivat indeksilaskentaa

1. Tilastoyksiköt (esimerkiksi eri asunnot tai eri työntekijät) voivat vaihdella osittain tai kokonaan perus- ja vertailuajankohtien välillä.
2. Mittauksen kohteissa saattaa tapahtua rakenteellisia muutoksia perus- ja vertailuajankohtien välillä.
3. 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 (esimerkiksi suhteellisesti tarkasteltuna) muuttumiseen ajassa. Esimerkiksi ammattinimikkeiden, sukupuolen, iän ja koulutustason mukaan luokiteltu työvoima voi muuttua työpanososuuksien suhteen perus- ja vertailuajankohdan välillä. Vastaavasti asunokaupan suhteellinen painopiste voi muuttua alueellisesti perus- ja vertailuajankohtien välillä.

Edellä esitettyjen indeksilaskennan ongelmien ratkaisumetodina käytetään hedonisia indeksimenetelmiä. Näiden menetelmien tavoitteena on eliminoida esiin tuodut laadullisten muutosten vaikutukset hinnanmuutoksen mittaamisessa. Menettely johtaa tilanteeseen, jossa todellinen hinnanmuutos (Log-muutoksella mitattuna) jaetaan osatekijöilleen seuraavasti: Todellinen hinnanmuutos laatupuhdistettu hinnanmuutos + laadullisten ominaisuuksien muuttumisesta syntyvä hinnanmuutos.

Metodisesti todellisen hinnanmuutoksen jakaminen osatekijöilleen voidaan toteuttaa karkeasti luokitellen kahdenlaisilla hedonisilla laadunvakiointimenetelmillä: Griliches-tyyppisellä hedonisella menetelmällä (Griliches, 1971) tai hedonisella imputoinnilla (Koev, 1997). Griliches-tyyppisessä hedonisessa menetelmässä keskitytään edellä esitettyjen kohtien 1–2 indeksilaskennan ongelmien eliminointiin. Tyypillisiä Griliches-menetelmällä toteutettuja indeksejä ovat asunto-osakkeiden hintaindeksi ja jälleenmyytyjen autojen hintaindeksi. Hedonisessa imputoinnissa indeksiongelmat ratkaistaan kahdessa vaiheessa: Ensimmäisessä vaiheessa arvioidaan tilastoyksiköille (esimerkiksi myydyille kiinteistöille tai ammattiryhmän työntekijöille) laatuvaikoidut hinta-arviot, ja toisessa vaiheessa näistä hinta-arvioista konstruoidaan laatupuhdistetut hintasuhteet samaan tapaan kuin tyypillisessä indeksilaskennassa. Hedonista imputointia on sovellettu metalliteollisuuden palkkojen muutoksen mittaamiseen, omakotitalojen ja tonttien indeksisovellukseen, laatuvaikoidun työn hintaindeksiin ja sektoreiden välisten palkkaerojen estimointiin (Valpas, Kunpas ja Kirpas -tutkimukset).

Molemmille hedonisille menetelmille on tyypillistä hintojen määräytymisen mallintaminen. Hintojen ajatellaan määräytyvän ehdollisina odotusarvoina tilastoyksiköiden ominaisuuksille. Griliches -tyyppisessä menetelmässä ominaisuuksien hintavaikutukset ovat estimointiperiodeilla vakioisia, joten siinä ei huomioida mahdollisia makutottomusten muutoksia (ominaisuuksien markkina-arvostusten muutoksia) ajassa. Hintamallit ovat tyypillisesti (vakiokerroin) kovarianssianalyysimallia, 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 laatuvaikointi ja indeksilaskenta toteutetaan menetelmällisesti kahdessa vaiheessa. Ensimmäisessä vaiheessa perus- ja vertailuajankohtien hintamallit spesifioidaan funktiomuodon ja muuttujien suhteen samanlai-

siksi. Perus- ja vertailuajankohtien hintamallit estimoidaan toisistaan riippumattomasti, joten ominaisuusmuuttujien kertoimia ei sidota niitä estimoituessa.

Tämän jälkeen laadunvakiointi suoritetaan seuraavasti. Arvioidaan perus- ja vertailuajankohtien hintamallien avulla, paljonko perusajankohdan hyödykkeet maksaisivat ominaisuuksineen ja rakenteineen vertailuajankohtana (esimerkiksi alue-, hyödyke- tai työpanosluokitus) vertailuajankohdan ominaisuuksien ja rakenteellisten 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 vertailuajankohdan laatuvarakioitujen hintojen välillä, esimerkiksi tyypillisen Laspeyresin kaavan avulla.

Vastaavalla menettelyllä voidaan arvioida, paljonko vertailuajankohdan hyödykkeet maksaisivat ominaisuuksineen ja rakenteineen (esimerkiksi alueluokitus) perusajankohdalla perusajankohdan ominaisuuksien ja rakenteellisten tekijöiden markkina-arvostuksilla. Vertailtavien hyödykkeiden ominaisuudet ja luokittelun rakenteelliset tekijät vastaavat tällöin vertailuajankohdan tilannetta, joten vertailuajankohdan määrätiedoilla hinnannuutos voidaan arvioida esimerkiksi Paaschen kaavan avulla. Hedoninen imputointi ei rajoita indeksikaavan valintaa – kaikkia tyypillisiä indeksikaavoja voidaan soveltaa hinnannuutoksen arvioinnissa.

Laatuvakioituille, samoin kuin mille tahansa, indekseille pätee, että ne ovat ehdollisia tilastoja, joissa ehtoina ovat indeksilaskennassa kontrolloidut tekijät. Laatuvakioituille 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.

Suosituksset

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, joissa valitaan

- indeksikaava ja
- indeksisarjojen konstruointistrategia.

Indeksit voidaan konstruoida joko kanta- tai ketjuindekseinä. Kantaindekseissä perusajankohta pidetään kiinteänä ja vertailuajankohtaa muutetaan liukuvasti ajassa. Ketjuindekseissä puolestaan perus- ja vertailuajankohdat ovat esimerkiksi peräkkäisiä ajankohtia tai esimerkiksi peräkkäisten vuosien vastaavia neljänneksiä tai kuukausia ja ne muuttuvat liukuvasti ajassa.

Neljäs indeksien konstruointistrategian ongelma-alue liittyy hyödykkeiden laadulliseen muuttumiseen sekä 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.

Kirjallisuus

- Allen, R. G. D. 1975. *Index numbers in theory and practice*. London: Palgrave Macmillan.
- Diewert, W. E. 1974. Intertemporal consumer theory and the demand for durables. *Econometrica*. Vol. 42. No. 3. 497–516.
- Diewert, W. E. 1975. *Ideal log-change index numbers and consistency in aggregation*. Discussion Papers. No. 75–12. University of British Columbia. Department of Economics.
- Griliches, Z. (toim.) 1971. *Price indexes and quality change*. Cambridge: Harvard University Press.
- Koiv, E. 1997. *Constructing a Hedonic Wage index: Pilot Study for the Finnish Metal Industry*. Tilastokeskus. Helsinki.
- Samuelson, P. A. & Swamy, S. 1974. Invariant economic index numbers and canonical duality: survey and synthesis. *American Economic Review*. Vol. 64. No. 4. 566–593.
- Törnqvist, L. 1974. *Aikasarjojen konstruointi*. Kautio-Toikka, A. (toim.). Helsinki: Gaudeamus.
- Vartia, Y. 1976. *Relative changes and index numbers*. Research Institute of the Finnish Economy. Series A. Helsinki.
- Vartia, Y. 1983. *Aikasarjojen konstruointi ja indeksiluvut*. Helsingin yliopiston tilastotieteen laitoksen opetusmonisteita. No. 3. Helsinki.
- Vartia, Y. & Vartia, P. 1984. Descriptive index number theory and the Bank of Finland currency index. *Scandinavian Journal of Economics*. Vol. 86. No. 3. 352–364.
- Vartia, Y. 1995. *Quality and price of labour*. Discussion Papers. No. 366. University of Helsinki. Department of Economics. Helsinki.

3 Dokumentointi ja arkistointi

3.1 Tilastotutkimuksen dokumentointi

Sovellusala ja päämäärä

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

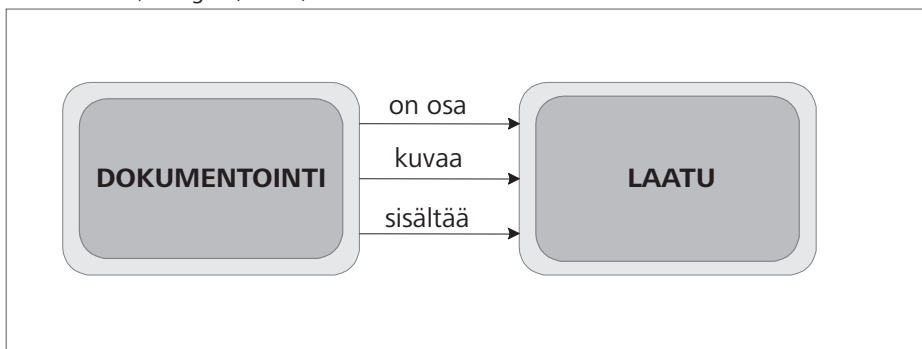
Tilastojen käyttäjien on dokumentoinnin avulla:

- 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.

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



Dokumentointi voidaan jakaa tilastollisten prosessien ja tilastotuotteiden dokumentointiin. Prosessorientoitunut dokumentointi kuvaa miten tuote valmistetaan; tilaston käyttäjä tarvitsee puolestaan laatuselosteen tilaston laadusta.

Periaatteet

Dokumentoinnilla tarkoitetaan toiminnan kuvaamista. Siihen kuuluvat 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 (Söder, 1998).

Dokumentoinnissa hyödynnetään koko viraston yhteisiä metatietokantoja: koonti-, luokitus- ja käsitetietokantoja sekä systeemirekisteriä ja arkistonmuodostussuunnitelmaa. Metatietokantojen käyttöliittymät on suunniteltu siten, että ne tukevat dokumentointia. Tietokantojen käyttöliittymien rakenne ohjaa käyttäjää metatietojen kirjaamisessa. Lisäksi käyttäjien tueksi on laadittu käyttöohjeita ja kuvausmalleja. Metatietokannoista voidaan tuottaa dokumentteja eri tarkoituksiin, esimerkiksi jakelutietokantoihin (StatFin), arkistoon ja muihin sähköisiin välineisiin sekä paperijulkaisuihin.

Näiden metatietojen dokumentoinnin lisäksi tulee kehittää yhteiset käytännöt muun muassa prosessikuvausten ja työohjeiden dokumentoinnille. Tilastotutkimuksen selkeä prosessikuvaus tukee tilaston tuotannon jatkuvuutta ja kehittämistä. Tarjottavan dokumentoinnin laajuus riippuu tilaston tuottajan, ylläpitäjän ja asiakkaan tarpeesta.

Suosituks

Dokumentoinnin on täytettävä kaikki toimeksiantajan tilastolle asettamat vaatimukset. Tilastojen ja niiden tuotantoprosessien yleiskuvauksen tulee noudattaa seuraavaa tarkistuslistaa, joka on jäsennetty 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 tarkkuus ja luotettavuus

Menetelmät

- tiedonkeruumenetelmä,
- käytetyt otanta- ja estimointimenetelmät.

Aineiston tilastollinen laatu

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

3. Julkaistujen tietojen ajantasaisuus ja oikea-aikaisuus

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

4. Tilastojen yhtenäisyys ja vertailukelpoisuus

Vertailukelpoisuus

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

Yhtenäisyys

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

5. Tilastotietojen saatavuus ja selkeys

Tietojen jakelu

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

Tietojen tulkittavuus

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

Dokumentointi

Tietojärjestelmä

- arkkitehtuuri, tietomalli, toiminta ja versionhallinta,
- lähdekoodi ja testaus,
- käsittely- ja tarkistussäännöt,
- virheiden ja muutosten hallinta.

Tuotantoprosessi

- prosessin eri työvaiheet, organisointi ja vastuu.

Tietovarastot

- tietovarastojen tekninen- ja yleiskuvaus,
- tietojen tekninen- ja yleiskuvaus,
- muodostussäännöt.

Tiedon hallintaan liittyvät metatiedot

- tilaston nimi,
- tilastollinen aihealue,
- vastuuhenkilö,
- tiedonantovelvollisuus,
- saatavuus,
- salassapitomääräykset ja käyttörajoitukset,
- säilytys, arkistointi ja versionhallinta,
- EU-regulaatiot,
- käytetyt resurssit,
- muut tietolähteet ja lisäinformaatio.

Kirjallisuus

Sundgren, B. 2001. *Documentation and Quality in Official Statistics*. The International Conference on Quality in Official Statistics. Stockholm 14.–15.5.2001. Proceedings. Statistics Sweden & Eurostat.

Söder, L. 1998. *Ehdotus Suomen Virallisen Tilaston (SVT) laatukriteereiksi*. Tilastokeskus. (Muistio).

Lisätietoja

Haikala, I., & Märijärvi, J. 2006. *Ohjelmistotuotanto*. Helsinki: Talentum.

Mertanen, S., Heinilä, A., & Hietanen, H. 2006. *Systeemityön tietopankki*. Tilastokeskuksen sisäinen ohjeistus intranetissa tietojärjestelmän rakentajalle.

Mertanen, S. (toim.) 2006. *Projektityön tietopankki*. Tilastokeskuksen sisäinen ohjeistus intranetissa projektipäälliköille, projektiryhmille, projektin ohjaus- ja johtoryhmille.

Statistics Canada. 2003. *Statistics Canada Quality Guidelines*. Ottawa. <http://www.statcan.ca/english/freepub/12-539-XIE/free.htm>

Suomen virallisen tilaston neuvottelukunta. 2006. *Suomen virallisen tilaston perus- ja laatukriteerit*. TK-41-1227-06.

von Rothstein, I. & Sundgren, B. 1996. *Sweden's Statistical Databases: Detailed Descriptions of the Metadata in the Macrodatabase*. Statistics Sweden.

3.2 Tilastotutkimuksen arkistointi

Sovellusala 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 myös kuvataan laadukkaasti.

Arkistotoimen tarkoituksena on turvata aineiston säilyminen uudelleenkäytön mahdollistavassa muodossa yhteiskunnallista tutkimusta ja tietopalvelua varten sekä edistää tietovarannon hoitoa hyvien asiakirjahallinnollisten ja yhtenäisten tietohallinnollisten periaatteiden mukaisesti.

Arkistolain (831/1994) mukaan arkistotoimen keskeisiä tehtäviä ovat asiakirjojen säilytysajoista ja -tavoista päättäminen sekä arkistonmuodostussuunnitelman (AMS) ylläpitäminen. Arkistolain mukaisia tehtäviä ovat:

- aineistojen arvonmääritys ja seulonta,
- arvokkaiden aineistojen säilyttäminen,
- säilytettävien aineistojen järjestäminen ja luettelointi sekä
- lainaaminen ja luovutus tutkimustarkoituksiin,
- tietopalvelu arkistoaineistoista,
- tarpeettomien aineistojen hävittäminen ja
- aineistojen suojaaminen poikkeusolojen varalta.

Arkistonmuodostussuunnitelma (AMS) toimii Tilastokeskuksen tilastoaineistojen karttumisohjeena. Siinä määritellään pysyvästi tai Tilastokeskuksen päätöksellä määrääjän säilytettävät aineistot. Kansallisarkisto hyväksyy arkistonmuodostussuunnitelman pysyvästi säilytettävien aineistojen osalta.

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

Periaatteet

Tilastokeskuksessa arkistoinnin tärkein työväline on Tilastokeskuksen Koonti-tietokanta. Koontitiedostojen kuvausjärjestelmässä kuvataan arkistoitavasta aineistosta sen tiedosto- ja kenttäkuvaukset eli metatiedot.

Myös muut Tilastokeskuksen metajärjestelmät: Luokitustietokanta ja Käsite-tietokanta, joita hyödynnetään edelleen koontitiedostojen kenttäkuvauksissa sekä Systeemirekisteri (SYSREK) ja Arkistonmuodostussuunnitelma (AMS), joita puolestaan hyödynnetään tiedostonkuvauksissa, on linkitetty koontitiedostoon.

Periaatteen mukaisesti kuvaus tehdään aina vain yhteen paikkaan. Jos esimerkiksi arkistoitavassa aineistossa käytetään Luokitustietokannassa kuvattua

luokitusta, ei kyseistä luokitusta tule kuvata Koonnissa erikseen, vaan se linkitetään koontitiedoston kenttäkuvaukseen luokitusliitoksen avulla. Vastaavasti tehdään myös käsiteliitos. Arkistoitavan aineiston tiedostokuvauksen AMS-liitos kertoo sen, mihin arkistointikohteeseen koontitiedosto kuuluu ja systeemitunnus sen, mihin systeemiin se liittyy.

Arkistoinnissa hyödynnetään mahdollisimman paljon Tilastokeskuksen metajärjestelmiä, koska tämä helpottaa ja nopeuttaa kuvaustyötä. Arkistoijan-ABC-oppaaseen (Taivainen, 2005b) on koottu ohjeet arkistoinnista ja tarvittavista sovelluksista, niiden käytöstä ja tuesta sekä tietoa käyttäjätunnuksista ja sovellusten asentamisesta.

Arkistointitapahtuman päävaiheet ovat:

Vaihe 1. Arkistointikohteen Arkistonmuodostussuunnitelma -kuvauksen (AMS) tiedot on tarkistettava. Pienet muutokset voidaan tehdä päivittämällä kuvausta. Jos on kyse oleellisista muutoksista (esimerkiksi muutos kokonaisaineistosta otosaineistoksi), tehdään uusi arkistointikohte.

Vaihe 2. AMS-kuvauksen mukaiset koontitiedostot on kuvattava koontitiedostojen kuvausjärjestelmään (Koonti-sovellus). AMS-kuvauksessa ei luetella koontitiedostojen nimiä, vaan ne aineistot, jotka arkistoidaan. Arkistointikohteeseen voi liittyä useita koontitiedostoja, joita nimettäessä tulee käyttää AMS:ssa määriteltyä systeemitunnusta, jotta tiedostot löytyvät jatkossakin.

Vaihe 3. Koontitiedostot arkistoidaan Unix-koneelle, jonka jälkeen Koonti-sovelluksessa oleviin koontitiedostoihin päivittyvä Säilytys-kohtaan merkintä arkistoituu.

Vaihe 4. Arkistoidusta koontitiedostosta täytetään AMS-sovelluksessa Luovutuslomake, johon liitetään kyseiset koontitiedostonimet ja muut luovutustiedot. Tämän jälkeen lähetetään luovutuslomake, sekä mahdollinen paperiaineisto tilastoarkistoon. Jos arkistointikohteeseen liittyy useita eri aikaan valmistuvia koontitiedostoja, luovutuslomake täytetään vasta, kun kaikki koontitiedostot on arkistoitu.

Suosituksia

- Arkistoitavan aineiston koontitiedoston kuvaus on aina tehtävä huolellisesti, jotta aineistojen hyödyntäminen ja oikea jatkokäyttö tulisi turvatuksi. Aineistojen jatkokäytön osalta keskeistä on, että metatiedot 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, onko luokitukset kuvattu luokituskantaan ja käsitteet käsittekantaan). Tämän lisäksi tulee perehtyä arkistoinnin ohjeisiin (Taivainen, 2005a; 2005b; 2005c).
- 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.

- 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 PX-Web- ja StatFin-taulukoiden metakuvauksiin.
- Kun kuvaus on valmis, se tarkistutetaan vielä koontitukihenkilöllä, joka merkitsee kuvauksen hyväksytyksi. Vain hyväksytyn koontitiedoston voi arkistoida.

Kirjallisuus

Suomen lait ja asetukset sekä kansalliset toimintaohjeet

Arkistolaki (831/1994).

Tilastokeskuksen toimintaohjeet

Taivainen, T. 2005a. AMS-sovelluksen käyttäjän ohje. Tilastokeskus. Sisäinen muistio. Helsinki.

Taivainen, T. 2005b. Arkistojen-ABC – Perustietoja arkistoinnista ja metatietojen kuvaamisesta. Tilastokeskus. Sisäinen muistio. Helsinki.

Taivainen, T. 2005c. Tilastoaineistojen arkistoinnin yleisohje. Tilastokeskus. Sisäinen muistio. Helsinki.

Lisätietoja

Saijets, M. & Toivonen, P. 1995a. Keskuskoneen koontitiedostojen kuvausjärjestelmä. Versio 1. Tilastokeskus. Sisäinen muistio. Helsinki.

Saijets, M. & Toivonen, P. 1995b. Keskuskoneen koontitiedostojärjestelmän käyttöohje. Versio 2. Toinen painos. Tilastokeskus. Sisäinen muistio. Helsinki.

Syvänperä, R. & Jouhki, S. 2000. Käsitietokannan käyttö. Tilastokeskus. Sisäinen muistio. Helsinki.

Taivainen, T. & Toikka, A. 2005. SAS-tiedostojen arkistointiohje+Koonti-makro. Tilastokeskus. Sisäinen muistio. Helsinki.

4 *Tilastojen käytön edistäminen, julkaiseminen ja asiakaspalvelu*

4.1 *Tilastojen käytön edistäminen*

Sovellusala ja päämäärä

Tilastojen käytön edistämiseen kuuluvat tilastotietojen sekä tilastotietoihin perustuvan tuote- ja palveluvalikoiman ja tuotteiden käytettävyyden kehittäminen, niiden tunnetuksi tekeminen sekä tilastojen luku- ja käyttötaidon edistäminen.

Tilastojen käytön edistämisen tavoitteena on varmistaa tilastojen laaja ja tehokas käyttö yhteiskunnassa. Tämän saavuttamiseksi tilastojen tulee olla helposti saatavilla ja käytettäviä. Tilastojen saatavuutta parannetaan yleisellä ja tuotetai tilastokohtaisella tiedottamisella ja markkinoinnilla sekä kehittämällä tuotteita ja palveluita helposti löydettäviksi ja saavutettaviksi. Tilastojen käytettävyyttä lisätään tekemällä tilastotuotteista ja -palveluista selkeitä, havainnollisia ja käyttäjien tarpeisiin soveltuvia. Tilastojen käyttöä edistetään myös tarjoamalla opastusta ja koulutusta tilastojen käyttämiseen ja lukemiseen.

Tilastojen käytön edistämässä, julkaisemisessa ja asiakaspalvelussa noudatetaan Tilastokeskuksen viestintästrategiassa (2003) määriteltyjä viestinnän tavoitteita ja periaatteita. Viestintästrategian mukaan viestinnässä ja asiakaspalvelussa tulee noudattaa seuraavia periaatteita:

- *Luotettavuus:* Tiedot ovat oikeita ja niiden luotettavuustaso on esitetty.
- *Puolueettomuus:* Tilastokeskus korostaa viestinnässään tiedontuotannon objektiivisuutta, riippumattomuutta ja tasapuolisuutta sekä noudattaa tilasto-eettisiä periaatteita.
- *Avoimuus ja tasapuolisuus:* Tiedot julkistetaan tasapuolisesti niin, että julkistamisajat ilmoitetaan etukäteen ja kaikille kiinnostuneilla on mahdollisuus saada tiedot yhtä aikaa julkistamishetkellä. Tilastokeskuksen henkilöstö esiintyy aktiivisesti oman alueensa asiantuntijana ja välittää tietoa tilastoista ja palveluista organisaation sisällä ja ulkopuolella.
- *Ymmärrettävyys ja havainnollisuus:* Tilastokeskus pyrkii kaikessa viestinnässään ymmärrettävyyteen ja havainnollisuuteen välttämällä liiallisia ammattitermejä ja analysoimalla tilaston antamaa kuvaa. Tiedot esitetään niin täydellisinä, että käyttäjällä on mahdollisuus tulkintaan ja omien johtopäätösten tekoon.

Tilastokeskuksen asiakaspalvelun periaatteet on muotoiltu palvelulupauksiksi (ks. luku 4.3).

Periaatteet

Tuotteiden ja palveluiden tunnettuutta ja saatavuutta edistetään lisäämällä tietoisuutta yleisestä tilastotarjonnasta sekä auttamalla tiedon käyttäjää hänen tarvitsemansa tiedon löytämisessä. Tiedon käyttäjiä informoidaan uusista tiedoista, tuotteista ja palveluista mm. esitteillä, tiedotteilla, asiakaslehdessä, Internet-palvelun uutis- ja ajankohtaispöytäkirjoilla sekä erilaisissa tilaisuuksissa, tapahtumissa ja messuilla. Tiedotteita laaditaan muun muassa uusista tilastotiedoista ja tutkimustuloksista, tilastojen menetelmämuutoksista, tuotteista ja palveluista sekä ajankohtaiseen keskusteluun, uutisiin tai muihin asioihin liittyvistä aiheista. Tilastokeskuksen tuotteista ja palveluista toimitetaan vuosittain painettu katalogi.

Tilastokeskuksen Internet-palvelussa julkaistaan yleistä tilastotarjontaa kuvaavaa tietoa ja luetteloita. Tietoa tilastoista ja niiden saatavuudesta on esimerkiksi tilastoluettelossa ja tilastojen kuvauksissa, Suomen virallisen tilaston julkaisuluettelossa, tilastojen julkistamiskalenterissa ja Tilastokeskuksen viikkokalenterissa. Verkossa julkaistut luettelot tukevat myös tietojen hakua, sillä niiden kautta voidaan edetä sähköisesti julkaistuun tilastotietoon. Internet-palvelussa on myös kuvaukset tilastoihin liittyvistä tuotteista ja palveluista.

Tilastokirjaston TILDA-luettelotietokanta sisältää tiedot kirjastossa saatavilla olevista aineistoista ja WebStat-verkkopalvelu tiedot luotettavimmista kansainvälisistä verkossa olevista tilastoaineistoista. Tilastokirjaston julkaisuvaihtokäytäntöjen, kirjastojärjestelmien ja kaukolainaverkoston avulla varmistetaan, että tilastojulkaisuja on saatavilla myös muissa kirjastoissa.

Tilastojen käytettävyyttä parannetaan menetelmä- ja tuotekehityksellä. Tuotekehityksen kannalta tiedon käyttäjiltä systemaattisesti kerätty palautetieto on erityisen tärkeää. Verkkopalveluiden tuotekehityksessä hyödynnetään myös käytettävyydestä. Julkaisujen käytettävyyttä parannetaan kehittämällä taulukoita, tilastografikkua sekä tekstin luettavuutta ja rakennetta. Tämä edellyttää järjestelmällistä ohjeistusta ja koulutusta julkaisujen tekijöille.

Tilastokeskus toimittaa julkistamia tilastoja ja tilastoaineistoja kansainvälisille organisaatioille kansainvälisten sopimusten ja säädösten mukaisesti. Kansainvälisten järjestöjen tiedustelut toimitetaan kirjaamon kautta, jonne jää tiedot toimitetuista tiedoista.

Tilastojen luku- ja käyttötaidon edistämiseksi Tilastokeskus tekee yhteistyötä muiden tilastontuottajien sekä tilastonkäyttäjien, esimerkiksi oppilaitosten ja joukkoviestinten kanssa. Tilastokeskus järjestää asiakaskoulutusta, jonka kurseilla käsitellään tilastojen laadintaa, tulkintaa ja käyttöä sekä tilastoihin perustuvia tuotteita ja palveluita. Asiakkaille järjestetään myös räätälöityä koulutusta. Lisäksi Tilastokeskus pitää yllä verkkokoulua, joka perehdyttää tilastojen tulkintaan ja käyttöön, tiedonhakuun ja tilastolliseen ajatteluun sekä johdattaa keskeisten tilastojen perusteisiin.

Suosituks

- Tilastotarjonnan tunnettuutta tulee lisätä viestinnällä ja markkinoinnilla.
- Tilastojen saatavuutta tulee parantaa kehittämällä verkkopalveluita ja tuemalla tilastojen näkyvyyttä kirjastoverkossa.

- Tilastotuotteiden ja -palveluiden käytettävyyttä tulee parantaa.
- Tiedon tarvitsijoita tulee opastaa heidän tarvitsemansa tiedon lähteille.
- Asiakkaiden tarpeita tulee selvittää ja tilastotuotteita ja -palveluita kehittää niiden pohjalta.
- Tilastojen luku- ja käyttötaitoa tulee edistää yhteistyössä muiden tilastontuottajien sekä tilastokäyttäjien kuten oppilaitosten ja joukkoviestinten kanssa.
- Tilastokeskus toteuttaa julkisten palvelujen laatustrategiaa (Suomen kunta-liitto, 1998) ja pyrkii saamaan tuottamilleen tilastotiedoille yhä laajemman ja monipuolisemman käytön.

Kirjallisuus

Suomen Kuntaliitto. 1998. Julkisten palvelujen laatustrategia. Julkisten palvelujen laatustrategia-projekti. Helsinki.

Tilastokeskus. 2003. Tilastokeskuksen viestintästrategia.

4.2 *Julkaiseminen*

Sovellusala ja päämäärä

Tilastoalan eettiset periaatteet korostavat tilastoviranomaisen yhteiskunnallista vastuuta. Luotettavat ja riittävän kattavat tilastot ovat asioiden tehokkaan hoitamisen perusta demokraattisessa yhteiskunnassa. Kansalaiset tarvitsevat tilastotietoja muodostaakseen realistisen käsityksen elinoloistaan ja yhteiskunnasta, valvoakseen etujaan ja osallistuakseen päätöksentekoon. Myös julkinen hallinto, EU, elinkeinoelämä ja tutkimustoiminta tarvitsevat luotettavia ja puolueettomia tilastotietoja.

Tilastokeskus suunnittelee ja kehittää tuotteita ja palveluita sekä kansallisiin tietotarpeisiin että Euroopan tilastojärjestelmän tarpeisiin. Tilastokeskuksen päätuote, yleiset yhteiskuntatilastot, ovat kaikkien saatavilla maksutta Internetissä. Tilastokeskus toimittaa sopimusten ja EU-lainsäädännön edellyttämät tilastotiedot EU:lle ja kansainvälisille järjestöille maksutta. Tilastokeskus palvelee tilastotiedon käyttäjiä myös monipuolisilla tietopalveluilla, joista osa tehdään asiakkaan tilauksesta maksua vastaan. Julkaisemisen tavoitteena on saattaa Tilastokeskuksen tuottama tilastotieto viivytyksettä tiedon käyttäjien saataville.

Periaatteet

Julkaisemisella tarkoitetaan tilastotiedon jakelua yleisesti saataville muun muassa painetuissa ja sähköisissä julkaisuissa ja tilastotietokannoissa. Tilastokeskuksen julkaisemisen tavoitteet ja periaatteet noudattavat Tilastokeskuksen toimintastrategiaa (Tilastokeskus, 2003a), verkkopalveluiden kehittäislinjauksia (Tilastokeskus, 2003b) ja viestintäperiaatteita (4.1). Julkaisemisen periaatteita ovat

- *Tietojen luotettavuus*: tiedot ovat oikeita ja niiden luotettavuustaso on esitetty.
- *Tasapuolisuus*: tiedot julkistetaan ennalta ilmoitettuna ajankohtana niin, että ne ovat kaikkien käyttäjien saatavilla samanaikaisesti.
- *Ajantasaisuus*: tiedot julkistetaan niin pian kuin mahdollista niiden kuvaaman viiteajankohdan jälkeen.
- *Havainnollisuus*: tiedot välitetään havainnollisessa muodossa ottaen huomioon tiedon käyttäjien tarpeet. Tiedon käyttäjille annetaan riittävät mahdollisuudet omien johtopäätösten tekoon.
- *Puolueettomuus*: tuotettaessa tietoa yhteiskunnallisista kiistakysymyksistä välitetään leimaamista ja kannanottoja.
- *Tulkinta*: tietoja tulkitaan ja analysoidaan kuvaamalla eri ilmiöiden mittasuhteet sekä kertomalla muutosten ja ilmiöiden syistä ja seurauksista. Yksittäisen tilaston tiedot liitetään ja rinnastetaan mahdollisuuksien mukaan muihin samaan asiaan liittyviin tilastotietoihin ja muuhun mahdolliseen relevanttiin tietoon.
- *Ajankohtaisuus*: tietoja julkistettaessa ne kytketään ajankohtaiseen yhteiskunnalliseen keskusteluun ja teemoihin. Ajankohtaisista asioista tuotetaan tilastotietoa oma-aloitteisesti.
- *Avoimuus*: luotettavaa tilastotietoa ei salata.
- *Opastus*: käyttäjän tiedonhankintaa ja tietolähteiden etsintää tuetaan.

Julkistamisella tarkoitetaan tilastotiedon julkaisemista, joka tuo ensi kertaa julki jonkin tilaston tiettyä ajankohtaa kuvaavat keskeiset tiedot. Tilastolain periaatteiden mukaisesti valmistunut tilastotieto julkistetaan viivytyksettä ja samanaikaisesti kaikkien tiedonkäyttäjien saataville (*Tilastokeskus, 2004a*).

Uudet tilastotiedot julkistetaan tilastojulkistuksina kunkin tilaston kotisivuilla. Tiedot julkistetaan kielilain mukaan kaikille käyttäjille samanaikaisesti sekä suomeksi että ruotsiksi (*Tilastokeskus, 2004b*). Englanninkielinen tilastojulkistus tehdään kansainväliseltä kannalta tärkeistä tilastoista (esimerkiksi suhdanteet, kansantalouden tilinpito ja väestökehitys; *Tilastokeskus, 2006a* ja *2006b*). Julkistaminen sekä siihen liittyvä tiedon käyttäjien informoiminen ovat osa Tilastokeskukselle kuuluvaa yleistä, valtion viranomaisille määrättyä tiedotusvelvollisuutta.

Julkistaminen on suunnitelmallista. Tulevista julkistuksista kerrotaan kattavasti etukäteen vuotuisessa julkistamiskalenterissa ja tarkemmin viikoittaisessa uutiskirjeessä eli viikkokalenterissa (*Tilastokeskus, 2006c*). Kalentereilla Tilastokeskus lisää toimintansa ennakoitavuutta niin, että tiedonkäyttäjät voivat omassa toiminnassaan ottaa huomioon uuden tiedon ilmestymisen. Kalenterit tukevat prosessien toimivuutta ja asiakaspalvelua muun muassa tietopalvelussa, markkinoinnissa ja viestinnässä. Tilastojen suunnitelmien mukaista julkistamista seurataan säännöllisesti.

Tilastokeskus tuottaa ja julkistaa tietoja noin 200 eri tilastosta, jotka kattavat yhteiskunnan eri osa-alueet. Suurin osa tilastoista kuuluu Suomen viralliseen tilastoon (SVT). Tilastokeskuksen säännöllisessä tuotannossa olevista tilastoista tehdään noin 700 julkistusta vuodessa.

Tietojen salassapito ja luovutus ennen julkistusta

Julkistettavat tilastotiedot ovat salassa pidettäviä julkistushetkeen asti (*Tilastokeskus, 2008*). Erityisesti salassa pidettäviä ovat pääoma- ja rahoitusmarkkinoin-

hin vaikuttavat tilastot. Esimerkkejä salassa pidettävistä tilastoista ovat kokonaistuotannon kuukausikuvaaja, kansantalouden neljännesvuositilinpito ja kulluttajabarometri. Muidenkaan tilastojen tuloksista ei saa antaa tietoja etukäteen, vaan ne julkistetaan kaikille samanaikaisesti.

Tietoja voidaan erityistapauksissa luovuttaa säädösten tai sopimuksien perusteella esim. Eurostatille Eurooppatason aggregaattien tuottamista varten tai muille virallisen tilaston tuottajille näiden omien julkaisujen toimittamista varten. Tiedot toimitetaan aina embargo-periaatteella (tiedot on pidettävä salassa niiden viralliseen julkitukseen asti).

Painetut julkaisut

Tilastokeskus julkaisee painojulkaisuja sekä sarjoina että yksittäisjulkaisuuksina. Suomen virallisen tilaston painettujen tilastojen pääjulkaisusarja on nimeltään Suomen virallinen tilasto, jossa julkaisevat Tilastokeskuksen lisäksi myös muut virallisen tilaston tuottajat. Kussakin julkaisussa on ilmoitettu aihealue, jolle julkaisu sijoittuu. Suomen virallisen tilaston neuvottelukunta on julkaissut sarjan julkaisuohjeet (2005) ja ylläpitää luetteloa sarjaan kuuluvista julkaisuista.

Virallisen tilaston julkaisujen lisäksi Tilastokeskus julkaisee Tutkimuksia-, Katsauksia-, Käsikirjoja ja Sukupuolten tasa-arvo -sarjoja. Tutkimuksia-sarja on korkeatasoisten tieteellisten tutkimusten julkaisusarja. Sarjan julkaisut käyvät läpi tieteellisen etukäteisarvioinnin. Katsauksia-sarjassa julkaistaan tilasto- ja tutkimusraportteja, jotka eivät sovellu em. sarjoissa julkaistaviksi. Käsikirjoja-sarjassa julkaistaan keskeisiä tilastollisia käsikirjoja, luokituksia ja ohjeita. Sukupuolten tasa-arvo -sarjassa julkaistaan sukupuolten tasa-arvoon liittyviä tilasto- ja tutkimusraportteja.

Julkaisusarjojen lisäksi Tilastokeskus julkaisee yksittäisjulkaisuja eli monografi- oita sekä muita julkaisuja (aluekatsaukset, taskutilastot, lehdet). Yksittäisjulkaisut ovat erillisiä korkeatasoisia, kirjamaisia tilasto- ja tutkimusraportteja. Aluekatsaukset sisältävät kunnittaista tilastotietoa havainnollisessa muodossa. Taskutilastoina ilmestyy sekä yleisjulkaisu Suomi lukuina että aihealuekohtaisia julkaisuja. Tilastokeskuksen lehdet kertovat tilastoista havainnollisten artikkelien avulla.

Sähköiset julkaisut

Tilastokeskuksen säännöllisessä tuotannossa olevilla noin 200 tilastolla on kullakin oma pysyvä kotisivunsa Tilastokeskuksen verkkopalvelussa. Yleiset yhteiskuntatilastot ja niiden sisältämät tilastotiedot, jotka ovat kaikkien saatavilla Internetissä, julkaistaan ja myös säilytetään käyttäjien saatavilla tilastojen kotisivuilla (Tilastojulkistusarkisto <http://tilastokeskus.fi/til/arkisto/index.html>). Tilastojen kotisivujen kautta myös tietokannoissa julkaistut tilastotiedot ovat löydettävissä. Tilaston kotisivulta löytyy varsinaisen julkaistun tilastotiedon lisäksi tilastoon liittyvää syventävää ja taustoittavaa aineistoa, kuten tilaston kuvaus, laatu- ja menetelmäselosteita, käsitteitä ja määritelmiä sekä luokituksia.

Tiedonhaun helpottamiseksi tilastot ryhmitellään Suomen virallisen tilaston aihealuejaon mukaisesti (ks. luku 1.1.4). Ryhmittely, tilastojen asiansanoitus sekä tilastojen laajat tausta-aineistot auttavat käyttäjää jäsentämään asiakokonaisuuksia ja tutustumaan aiheeseen laajemmin.

Havaittujen virheiden korjausmenettelyt

Julkaistuissa tilastotiedoissa havaitut virheet on korjattava välittömästi Suomen virallisen tilaston neuvottelukunnan suositusten (SVT-neuvottelukunta, 2005) ja Tilastokeskuksen sisäisten ohjeiden (Tilastokeskus, 2007) mukaisesti. Ohjeet ja suositukset kattavat menettelytavat eri julkaisukanavissa. Julkaisuissa tai verkkopalveluissa esiintyvien virheiden korjaaminen on luottamuksen säilyttämisen kannalta ensiarvoisen tärkeää.

Korjausmenettely ei koske ennakkotietoina esitettyjä tietoja, joiden muuttumismahdollisuus kerrotaan niitä julkaistaessa. Ennakkotietojen suoranaiset virheet tulee korjata asianmukaisesti.

Suosituks

- Tilastojen julkistamisessa ja julkaisemisessa on noudatettava Suomen virallisen tilaston ja Tilastokeskuksen julkaisuohjeita, jotka koskevat muun muassa yhtenäisiä esitys- ja merkintätapoja, havaittujen virheiden korjauskäytäntöjä, julkistamiskalenterien käyttöä ja sisältöä (Tilastokeskus, 2003a; 2004a; 2004b; 2006c; 2007, ja 2008; SVT-neuvottelukunta, 2005).
- Tilaston tietoja saa Tilastokeskuksessa käsitellä ja niistä saa tietoja ennen julkistamista vain henkilö, joka osallistuu kyseisen tilaston laadintaan tai tarvitsee kyseisen tilaston tietoja omassa työssään ennen tietojen julkistamista (Tilastokeskus, 2008).
- Tilastojen luotettavuuden arvioimiseksi laadintamenetelmien on oltava julkisia ja yleisesti hyväksyttäviä. Julkaisujen ja käyttäjille suunnattujen ohjeiden tavoitteena on opastaa tilastojen oikeaan käyttöön ja tulkintaan. Tämän vuoksi tilastojulkaisuun on liitettävä laatuseloste, josta voi päätellä tietojen luotettavuuden ja mahdolliset epätarkkuusriskit (SVT-neuvottelukunta, 2006).

Kirjallisuus

Kansalliset toimintaohjeet

Suomen virallisen tilaston (SVT) neuvottelukunta. 2005. *Suomen virallisen tilaston julkaisuohje*.

Suomen virallisen tilaston (SVT) neuvottelukunta. 2006. *Suomen virallisen tilaston perus- ja laatukriteerit*.

Tilastokeskuksen toimintaohjeet

Tilastokeskus. 2003a. *Tilastokeskuksen toimintastrategia*. (TK-04-1225-03).

<http://tilastokeskus.fi/org/tilastokeskus/strategia.html>

Tilastokeskus 2003b. *Tilastokeskuksen verkkopalveluiden kehittämislinjaukset*. (TK-04-1225-03).

Tilastokeskus. 2004a. *Päätös tilastojen julkistamisajankohdasta Tilastokeskuksessa*. (TK-00-1576-04).

Tilastokeskus. 2004b. *Päätös kielilain huomioonottamisesta Tilastokeskuksessa*. (TK-00-1425-04).

Tilastokeskus. 2006a. *Päätös englannin kielilinjauksesta*. (TK-00-1100-06).

Tilastokeskus. 2006b. *Englannin kielen käyttö Tilastokeskuksen tuotteissa ja palveluissa*. Sisäinen ohje.

Tilastokeskus. 2006c. *Tilastojen ja tiedotteiden julkistaminen Tilastokeskuksessa – periaatteet ja käytännön ohjeet*. Sisäiset ohjeet Tilastokeskuksen intranetissa.

Tilastokeskus. 2007. *Ohje: Julkistetun tilastotiedon muuttuminen ja julkistamisen virhetilanteet*. TK-00-861-07.

Tilastokeskus. 2008. *Ohje tilastotietojen salassapidosta ja luovutuksesta ennen julkistamista*. (TK-00-23-08).

Lisätietoja

Kielilaki (423/2003).

Valtioneuvoston kanslia. 2002. *Valtionhallinnon viestintäsuositus*. Valtioneuvoston tiedotusyksikkö. Helsinki.

4.3 Laatu asiakaspalvelussa

Sovellusala ja päämäärä

Tilastokeskuksen perustehtäviin kuuluu tiedon saattaminen käyttöön mahdollisimman laajasti. Tavoitteena on, että Tilastokeskuksen tuottamat yleisen yhteiskuntatilaston keskeisimmät tilastotiedot ovat asiakkaiden käytettävissä maksuttomana itsepalveluna verkossa. Lisäksi Tilastokirjaston palvelut ovat kaikkien käytettävissä. Maksullisena palveluna tarjotaan esimerkiksi tutkimus- ja haastattelupalveluita. Tilastotietoja toimitetaan myös asiakkaan tilauksesta tehtävinä maksullisina toimeksiantoina.

Yhtenäiset toimintaperiaatteet ja käytännöt asiakaspalvelussa ovat tärkeitä, jotta asiakkaat saavat tarvitsemansa tiedot ja palvelun riippumatta siitä, minkä tulosityksikön kanssa he asioivat. Tämän vuoksi Tilastokeskus on määritellyt asiakaspalveluperiaatteet 2004 luodakseen yhtenäiset ja hyvät menettelytavat omalle asiakaspalvelulle.

Periaatteet

Asiakaspalveluperiaatteet

Asiakaspalvelu on keskeinen Tilastokeskuksen yrityskuvaan vaikuttava tekijä. Asiakkaiden tulee saada tarvitsemansa tiedot luotettavasti, ajallaan, ymmärrettävästi, käyttökelpoisessa muodossa, tietosuojan huomioiden ja oikein hinnoiteltuina. Asiakkaan tietotarpeisiin on kiinnitettävä huomiota ja toimitettava tiedot siinä muodossa, että he voivat käyttää tietoa työssään kustannustehokkaasti. Asiakaspalautteen kerääminen ja huomioon ottaminen toiminnassa on tärkeää palvelun kehittämisen ja laadun kannalta.

Jokaisen Tilastokeskuksen työntekijän tulee tuntee asiakaspalveluperiaatteet (Tilastokeskus, 2004). Niiden mukaan:

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

Opastus ja tietopalvelu

Jotta tiedontarvitsija saa tarvitsemansa tiedot ja osaa tulkita näitä tietoja oikein, Tilastokeskuksen tuotteista ja palveluista on tiedotettava, opastettava ja ohjeistettava. Tilastokeskuksen tuotteista ja palveluista sekä niiden hinnoista tehdään vuosittain katalogi. Lisäksi Tilastokeskuksen Internet-palvelussa on tietoa tuotteista, palveluista ja näihin liittyvistä metatiedoista sekä luokituksista, tiedonkeruista, laatuselosteista ja tilastotiedon saatavuudesta.

Tilastokeskuksessa asiakaspalvelu on yli yksikkörajojen tapahtuvaa yhteistyötä. Asiakkaan ottaessa yhteyttä puhelimitse, hänen puheluaan ei siirretä paikasta

toiseen. Sen sijaan yhteydenottotilanteessa selvitetään asiakkaan tiedon tarve ja yhteystiedot jatkotoimenpiteitä varten. Tämän jälkeen selvitetään organisaatiossa oikea asiantuntija, joka hoitaa toimeksiannon ja ottaa yhteyttä asiakkaaseen. Asiakkaiden kyselyihin ja palautteeseen vastataan mieluiten heti tai viimeistään kahden työpäivän aikana.

Asiakkaan toimeksianto käsitellään luottamuksellisesti. Lisäksi asiakastoimeksiannoissa huolehditaan, että tietopalvelussa noudatetaan Tilastokeskuksen lainsäädäntö- ja tietosuojavaatimuksia (ks. luvut 1.1.1 ja 2.2). Käyttöluvan alaisesta aineistosta tehdään käyttölupa-anomus, jonka tultua hyväksytyksi aineiston saa käyttöönsä (Tilastokeskus, 2005).

Keskeisimmät tilastotiedot ovat asiakkaiden käytettävissä maksuttomana itsepalveluna Tilastokeskuksen Internet-palvelussa Tilastot-sivustolla. Lyhytkestoiset opastus ja tietopalvelut – yleensä alle 15 minuutin työpanoksen vaatimat – tietopalvelut ovat maksuttomia. Muutoin erilliskustannukset kuten käytetty työaika veloitetaan.

Asiakaspalvelupisteissä on päivystys virka-aikana. Julkaisujen tilaukset voi tehdä ympärivuorokautisesti myös Tilastotori -verkkokaupan kautta. Tilastokirjaston tilat, laitteet ja kokoelmat ovat asiakkaiden käytettävissä virka-aikana. Tilastokeskuksen aluetoimistot palvelevat Turussa, Tampereella, Seinäjoella ja Oulussa.

Suosituks

- Asiakasta on kohdeltava ystävällisesti ja pyrittävä siihen, että asiakas saa yksilöllistä ja asiantuntevaa palvelua.
- Asiakkaan tulee saada tarvitsemansa tiedot sovittuna toimitusaikana. Viivästyksistä on tiedotettava välittömästi.
- Asiakkaan tilauksesta toimitettavat tuotteet ja palvelut ovat maksullisia. Palvelun hinnoittelussa on noudatettava voimassa olevaa hinnastoa. Hinnoittelu pohjaa valtion maksuperustelakiin (150/1992) ja -asetukseen (211/1992) sekä valtiovarainministeriön päätökseen Tilastokeskuksen suoritteista perittävistä maksuista (1208/2006).
- Tuotteisiin ja palveluihin on liitettävä laatuseloste. Muutoin asiakkaalle tulee toimittaa tiedot tilastoissa käytetyistä menetelmistä, käsitteistä ja määritelmistä. Maksullisissa toimeksiannoissa toimeksiantosopimukset ja toimitusedot on toimitettava asiakkaalle.
- Tuotteiden ja tilastojen osalta on pyrittävä virheettömään palveluun. Toimitettavat tiedot ja tuotteet on tarkistettava ennen niiden toimittamista asiakkaalle.
- Asiakaspalautetta kerätään keskitetysti Anoppi-asiakaspalautejärjestelmällä. Myös henkilöstö tallentaa saamansa palautteen. Reklamaatioihin on vastattava ystävällisesti ja nopeasti.
- Asiakkaiden tarpeita ja tyytyväisyyttä on arvioitava ja seurattava. Tilastokeskuksessa toteutetaan säännöllisesti asiakastytyväisyystutkimus. Lisäksi voidaan tehdä käyttäjätutkimuksia myös tuotteisiin ja palveluihin liittyen.
- Tilastokeskuksen työntekijöiden on sitouduttava noudattamaan Tilastokeskuksen asiakaspalveluperiaatteita.
- Asiakaspalveluperiaatteiden toteutumista seurataan tutkimuksilla ja muun palautteen avulla.

Dokumentointi

- Tietopalvelutoimeksiantoista tehdään toimeksianto- ja tietopalvelusopimukset (Tilastokeskus, 2002).
- Toimeksiantoja tehtäessä työvaiheet merkitään asiakashallintajärjestelmään.
- Asiakashallintajärjestelmään tallennetaan asiakasta koskevat toimenpiteet, kuten tehdyt tarjoukset ja sopimukset tietojen toimittamisesta.

Kirjallisuus

Suomen lait ja asetukset sekä kansalliset toimintaohjeet

Valtion maksuperusteasetus (211/1992).

Valtion maksuperustelaki (150/1992).

Valtiovarainministeriön asetus Tilastokeskuksen suoritteiden maksullisuudesta (1208/2006).

Tilastokeskuksen toimintaohjeet

Tilastokeskus. 2002. *Tietopalvelusopimuksen yleiset sopimusehdot.*

Tilastokeskus 2004. *Tilastokeskuksen asiakaspalveluperiaatteet.*

<http://www.stat.fi/org/palveluperiaatteet/index.html>

Tilastokeskus. 2005. *Ohje käyttöluupien myöntämisestä Tilastokeskuksen yksikötason aineistoon.* (TK-00-128-05).

Lisätietoja

Heino, J., Kaisio, R., Koskenkylä, T., Notkola, V., Orjala, H., Ranta, J., Rantanen, J., & Siljander, A. 2000. *Hinnoitteluopas.* Tilastokeskus. Helsinki. (TK-50-1453-00). (Uusi versio tulossa 2007.)

StatFin -tilastopalvelu. Tilastokeskus.

<http://statfin.tilastokeskus.fi>

Tilastokeskus.fi-verkkopalvelu. Tilastokeskus.

<http://tilastokeskus.fi/index.html>

Hakemisto

A

accessibility; 26
accuracy; 25
administrative records; 28
aggregointi; 98
Ahvenanmaa; 44, 47
aihealue; 13, 28, 44
aikasarjan komponentit; 95
aikasarjat; 26, 29, 87, 90, 94, 96, 98
ajankohtaisuus; 115
ajantasaisuus; 10, 11, 13, 14, 23, 26, 27, 46, 47, 48, 53, 54, 55, 107, 115
alipeitto; 25, 47, 48, 54, 64
alueluokitus; 41, 44
ammattillinen riippumattomuus; 10, 14
ammattieettiset periaatteet; 7, 13
ammattietiikka; 7, 13
analyysipaino; 80, 86, 87
ARGUS; 37
arkistointi; 26, 31, 33, 45, 88, 108, 109, 110
arkistolaki; 109
arviointi; 29, 31
arvoluetarkistus; 86
asetelmakerroin; 85
asetelmapaino; 79, 80, 81, 82
asetelmaperusteiset menetelmät; 86
asiakasarviointi; 21
asiakaskoulutus; 45, 113
asiakaspalaute; 21, 22, 113, 114, 119
asiakaspalvelu; 112, 118
asiakaspalveluperiaatteet; 118, 119
asiakastutkimukset; 11, 31, 119
asiakastyytyväisyys; 119
asiakkaan tietotarve; 15, 33
asuntokunta; 44, 46, 48
auditointi; 29, 31
avoimuus; 7, 14, 112, 115

B

balanced scorecard, BSC; 17
benchmarking; 22
bootstrap; 86
boxplot; 76
BSC; 18

C

CAPAR; 66
CAPI; 63
CASI; 63
CASIIP; 66
CASI-A; 66
CASI-V; 66

CATI; 63
CATI-keskus; 64
CAWI; 63, 66
census survey; 28
clarity; 26
cluster sampling; 58
Code of Practice; 8, 9
comparability; 26
complex sampling designs; 59
cut-off sampling; 59

D

DBM; 66
demografiset luokitukset; 41
derived statistical activity; 28
DESAP; 19, 20, 21, 29, 31
dokumentointi; 14, 18, 26, 29, 31, 33, 45, 55, 61, 75, 78, 82, 87, 105, 106

E

Edgeworthin indeksi; 100, 101
editointi; 25, 31, 73, 74, 78
editointiaste; 75, 78
editointisuhde; 75, 78
editointisäännöt; 74
EFQM; 9, 15, 19, 20
embargo-periaate; 116
EMS; 66
eräkatu; 71, 77, 86
esittäminen; 88
estimaattori; 80, 82, 84
estimointi; 30, 31, 47, 61, 79, 82, 84, 87
estimointimenetelmä; 107

F

Fisherin indeksi; 100, 101

G

graafinen editointi; 75

H

haastattelijapalaute; 48
haastattelijoiden koulutus; 30, 64, 65, 68, 72
haastattelu; 63
haastattelulomake; 63
haastatteluyksikkö; 64, 69
hajontakuvio; 76
hallinnollinen aineisto; 49
hallinnollinen rekisteri; 49
hallinnolliset aineistot; 6, 28
hallinnolliset rekisterit; 28, 46, 47, 48

harha; 57, 59, 60, 71, 72, 81, 85, 87
harhattomuus; 71, 72, 84
harjaestimaattori; 87
harmonisointi; 43
havainnollisuus; 88, 112, 115
hedoninen imputointi; 102
hedoniset menetelmät; 102
henkilöstön kehittäminen; 15
henkilötunnus; 47, 49, 51
histogrammi; 76
Horvitz-Thompson -estimaattori; 85
huoneistotunnus; 51

I

identifiointitunnukset; 55
identifiointitunnus; 38, 51, 78
imputointi; 25, 53, 73, 74, 77, 78, 86
imputointiaste; 77, 78
imputointisuhde; 77, 78
indeksilaskenta; 101
indeksin konstruointi; 103
indeksiteoria; 99
innovatiivisuus; 16
Internet-kysely; 63, 64, 66
Internet-palvelut; 45
irregular variation; 95
ISI; 7
ISO 3166; 44
ISO 9000; 16, 18
ITE-arviointi; 22
itsepainottuva; 80
itsetätettävä lomakekysely; 66

J

jackknife; 86
jakelukanavat; 107
jatkuva oppiminen; 16
johdetut tilastolliset aineistot; 28
johtajuus; 15, 18, 19
julkaiseminen; 11, 14, 34, 36, 37, 39, 112, 114, 117
julkisen hallinnon suositus, JHS; 42
julkistaminen; 14, 31, 88, 115, 117
jälkiositus; 80, 82, 85

K

kaatoryhmä; 42, 43
kadon adjustointi; 81
kaksivaiheinen otanta; 47
kalibrointi; 81, 82, 85
kalibrointiestimaattori; 85
Kansainvälisen tilastoinstituutti; 7
kansainväliset standardit; 44
kansainväliset suositukset; 6, 42
kansantalouden luokitukset; 41

kantatilasto; 52
katkaisuotanta; 59
kato; 25, 61, 64, 66, 71, 72, 73, 80, 81, 82, 85, 86
kato-osuus; 73
katovirhe; 62, 82
katovirheen oikaisu; 79, 81, 86, 107
kattavuus; 15, 34, 42, 54, 75
kausitasoitus; 26, 94, 95, 97
kausivaihtelu; 94, 95
kehikkoperusjoukko; 46, 47
kehikkovirhe; 25, 71, 80
kehittäminen; 15, 16, 18
kenttätyön organisointi; 30, 65, 72
keskivirhe; 25, 60, 62, 82, 84, 86
kiinteistötunnus; 51
kiintiöpoiminta; 59
kohdehenkilö; 46
kohdeperusjoukko; 25, 47, 48, 72, 107
kohderyhmä; 46
kokonaisuaineisto; 28, 39
kokonaisuadunhallinta; 15, 18
kokonaistutkimus; 25, 28, 46, 56, 59, 71
koontiedosto; 45, 109, 110, 111
kotitalous; 28, 44, 46, 47, 48, 57, 66, 78
kotitalouskäsite; 44
kotitalousotos; 48
kumppanuus; 15, 16
kustannukset; 13, 30, 48, 49, 52, 55, 56, 57, 58, 62, 64, 65, 119
kustannustehokkuus; 10
kuvion osat; 89
kuviotyypit; 90
kysely; 63
kyselylomakkeen ohjelmointi; 31
käsitteet; 6, 24, 26, 31, 41, 45, 68, 70
käyntihaastattelu; 63, 64, 65, 66
käytettävyyys; 114
käyttäjien informoiminen; 115
käyttäjien kuuleminen; 24
käyttäjien luottamus; 14
käyttäjien opastus; 105, 112, 113, 114, 115, 117, 118
käyttäjien palaute; 113, 114
käyttäjien tarpeet; 9, 10, 11, 13, 23, 24, 42, 54, 109, 114, 115
käyttäjien tyytyväisyys; 11, 119
käyttäjän opas; 31
käyttäjän opastaminen; 73, 84
käyttäjryhmät; 13
käyttäjätutkimukset; 119
käyttölupa; 35, 39, 119
käyttötarkoitus; 7, 24, 35, 36, 42, 56, 96, 103, 106
käytännėsäännöt; 8, 9, 11, 14, 20, 27, 34, 35

L

laadun arviointi; 14, 16, 17, 19, 20, 22, 24
laadun parannus; 105
laadun parantaminen; 14, 16
laadun suunnittelu; 16
laadun varmistus/tarkistus; 16
laadunhallinta; 15, 16
laadunhallintajärjestelmä; 16
laatikkokuvio; 76
laatuindikaattori; 81
laatuindikaattorit; 27, 60, 61, 62, 82
laatukriteerit; 8, 11, 12, 23, 27, 31
laatuseloste; 24, 27, 62, 82, 117, 118
laatuselvitys; 14, 31
laustandardit; 14, 16
laatuun sitoutuminen; 10, 14, 19
Laspeyresin indeksi; 100, 101
leikattu keskiarvo; 86
levykekysely; 66
linearisointimenetelmä; 86
lisäinformaatio; 25, 31, 46, 57, 58, 59, 79, 80,
81, 85, 86, 108
lomakkeen suunnittelu; 30, 68
lomakkeen testaus; 29, 72
loogisuustarkistukset; 31, 65
luokiteltavuus; 55
luokitteluperiaatteet; 42
luokitukset; 6, 24, 26, 31, 41, 45, 69, 118
Luokitus- ja metatietopalvelut; 45
luokitusopas; 45
luokitusstandardit; 42, 45
luokitustietokanta; 45
luotettavuus; 6, 7, 10, 11, 15, 23, 24, 25, 27, 34,
54, 56, 60, 63, 84, 86, 87, 105, 107, 112,
115
luottamuksellisuus; 34
luottamusväli; 84
läpinäkyvyys; 14, 16

M

makro-editointi; 74
Manner-Suomi; 44
master-otos; 59
menetelmäkuvaus; 25, 62, 82
menetelmäseloste; 24, 62, 82
metatieto; 10, 12, 18, 24, 26, 29, 31, 32, 45,
106, 108, 109, 110, 118
mikro-editointi; 74
mittausvirhe; 25, 71, 80, 107
monivaiheinen otos; 59
multi-phase sampling; 59
muutostiedot; 25, 49, 50, 52, 65, 100, 101, 103,
108, 115
muutostilasto; 52
muuttujamuunnos; 31
määritelmät; 6, 41, 45, 69, 106

N

NUTS; 44
NUTS-aluejako; 44

O

objektiivisuus; 8, 10, 14, 112
ohjaus; 15, 65
oikea-aikaisuus; 10, 11, 23, 26, 27, 107
oikeellisuus; 48, 54, 75
oikeellisuustarkistukset; 65
oikeudelliset yksiköt; 48
ongelmanasettelu; 30
opastus ks. myös käyttäjien opastus; 115
operationaalinen prosessi; 88
operationaalinen vaihe; 29, 31, 32
organisaatiotaso; 17
osallistuva havainnointi; 66
ositettu otanta; 47, 58, 59, 61, 81
ositus; 46, 57, 58, 59, 61, 82, 86
otanta-asetelma; 25, 30, 61, 79, 80, 81, 82, 85,
86
otantakehikko; 31, 46, 47, 48, 56, 57, 58, 61, 66
otantamenetelmä; 46, 56, 57, 60, 61, 72, 101,
107
otantatutkimus; 25, 28, 46, 56
otantavirhe; 11, 25, 60, 80, 107
otoksen kohdentaminen; 47
otoksen koko; 59, 60, 62, 79, 81, 82, 84, 85
otoksen poiminta; 31, 48, 79
otoksen tarkkuus; 57, 60
otosaineisto; 28, 39, 86
otoskehikon konstruointi; 31
otospaino; 61, 79, 80, 81
otosten koordinointi; 60
otosvarianssi; 60, 62, 71, 82, 86
otosyksikkö; 46

P

Paaschen indeksi; 100, 101
painetut julkaisut; 116
painokerroin; 79, 80, 81, 82
painokertoimien kalibrointi; 81
paneeliattribitio; 85
paneelitutkimus; 61
paperilomake; 63
parametristimaatit; 79, 82
parametrit; 80, 83
peittävyys; 46, 48, 54
perherakenne; 46
perusajankohta; 104
perusjoukko; 25, 28, 41, 46, 47, 56, 57, 58, 60,
61, 71, 72, 80, 81, 82
pienalue-estimaattori; 85
piirakkakuviot; 90, 93
pilottitutkimus; 29, 70

piste-estimaatti; 86, 87
pitkittäistutkimus; 63, 65
poikittaistutkimus; 63
paimintatodennäköisyys; 58
poissulkevuus; 42
postikysely; 63
PPS-otanta; 58, 59, 81
PRODCOM-luettelo; 45
promptness; 26
prosessijohdaminen; 15
prosessointivirhe; 25, 107
puhelinhaastattelu; 64, 65, 67
puolueettomuus; 7, 10, 13, 31, 112, 115
puuttuva tieto; 31, 47, 71, 92
pylväskuviot; 76, 90
pystypylväskuviot; 90, 92
pääryhmäso; 42

R

rakennustunnus; 51
regressioestimaattori; 85
regressioestimointi; 85
rekistereiden tietokoneavusteinen yhdistäminen;
66
rekisteri; 49
rekisteritutkimus; 66
relevanssi; 8, 10, 11, 14, 23, 24, 27, 46, 106
resurssien riittävyys; 10
resurssit; 15, 30, 33, 60, 64, 65, 69
riippumattomuus; 7, 8, 10, 14, 22, 112
rikosnimikkeistö; 41
ryväsotanta; 57, 58, 59, 61
ryvästys; 57, 58, 86

S

saatavuus; 10, 12, 14, 24, 26, 27, 64, 107, 108,
113, 118
salassapidettävät tiedot; 10, 34, 35, 36, 37, 38,
108, 115
sample survey; 28
SAS; 87
satunnaisotanta; 80, 85
scatterplot; 76
seasonal adjustment; 94
seasonal variation; 95
selkeys; 10, 12, 24, 26, 27, 107
simple random sampling, SRS; 57
sisältymistodennäköisyys; 57, 61, 80
standardiluoikitukset; 42
standardit; 18, 29, 31, 42
Stata; 87
statistical survey; 28
stratified sampling; 58
SUDAAN; 87
suhdannevaihtelu; 94, 95
suhde-estimointi; 85

suhdetehostus; 80
Suomen virallinen tilasto, SVT; 14, 23, 27, 115,
116
survey; 28
SurveyLaboratorio; 29, 69, 70
suunnittelu; 29
suunnitteluprosessi; 30
suunnitteluvaihe; 29
SVT-laatukriteerit; 23, 24, 27, 106
SVT-neuvottelukunta; 23, 117
SVT-peruskriteerit; 23
SVT-tuottajat; 14, 23
sykli; 95
systemityö; 31, 33
systemaattinen otanta; 57, 59, 61
sähköinen päiväkirja; 66
sähköiset julkaisut; 116

T

taloudellisen toiminnan luokitukset; 41
tarkentuvuus; 84
tarkkuus; 10, 11, 14, 16, 23, 25, 27, 46, 56, 57,
84, 88, 107
tasapuolisuus; 112, 115
tasavertaisuus; 10
taulukko; 34, 36, 38, 39, 88, 89, 91, 94
taulukon osat; 89
tautiluokitus; 41
tavoiteperusjoukko; 25, 28, 30, 46, 47, 72, 107
teemakartat; 90, 91
tehokkuus; 6, 7, 8, 16, 105
tiedonantovelvollisuus; 6, 108
tiedonhallinta; 31
tiedonkeruu; 107, 118
tiedonkeruulomake; 29, 65, 68, 69, 70
tiedonkeruulomakkeen laadinta; 31
tiedonkeruumenetelmä; 107
tiedonkeruumenetelmän valinta; 30
tiedonkeruutavat; 25, 62, 63
tiedonkeruuvaltuudet; 10
tietojenkäsittelymenetelmät; 30
tietojärjestelmä; 108
tietokoneavusteinen henkilöhaastattelu; 63
tietokoneavusteinen lomakekysely; 63
tietokoneavusteinen puhelinhaastattelu; 63, 64
tietokonelomake; 63
tietopalvelu; 114, 118
tietosisältö; 24, 46
tietosuoja; 6, 7, 13, 34, 37
tietosuojamenetelmä; 35, 36, 37, 39
tietovarastot; 108
tilastoarkisto; 45
tilastografiikka; 88, 89, 92
tilastojen käyttö; 10, 13, 16, 35, 109, 112, 113,
114
tilastojen laatu; 13

tilastojen lukutaito; 112, 113, 114
tilastojen läpivalaisu; 19, 20, 29, 31
tilastokuvio; 89, 92
tilastolaki; 6, 7, 29, 34, 35, 38, 49, 52, 55, 69, 115
tilastollinen kuntaryhmitys; 44
tilastolliset tietosuojamenetelmät; 36
tilastolliset yksiköt; 48
tilastorekisteri; 49
tilastosalaisuus; 10
tilastotoimen koordinointi; 13
tilastotutkimuksen toimintakehikko; 30
tilastotutkimus; 18, 25, 28, 29, 31, 46, 47, 56, 64
tilastoyhteistyö; 13
tilastoyksikkö; 41, 101, 102
timeliness; 26
toimialaluokitus; 45
toimintakehikko; 29, 41
toimipaikkatunnus; 51
total quality management, TQM; 15
TRAMO/SEATS; 95, 97
trendi; 90, 94, 95, 96, 97
trimmed means; 86
tulkinta; 12, 16, 90, 107, 112, 113, 115, 117
tullitariffinimikkeistö; 41
tuloshakuisuus; 15
tunnettuus; 113
tunnusjärjestelmä; 44, 49
tuotantoprosessi; 16, 26, 108
tuotantoydinprosessitaso; 18
tuotelaatu; 15
tuoteluokitukset; 45
tuoteluokitus; 42
tutkimusyksikkö; 46, 47, 48, 58, 63, 64, 71
työohjeet; 31, 45, 64, 69
täsmällisyys; 7, 84, 87, 98
täydellisyys; 46

U

uudelleenpainotus; 80

V

vaakapylväskuviot; 90, 93
variaatiokerroin; 60, 84, 86

vastaukato; 85
vastauskato; 78, 80, 81, 85, 107
vastauskato ks. myös kato; 66, 71, 86
vastausosuus; 73
vastausrasite; 8, 10, 13, 49, 52, 56, 60, 61, 65, 69, 72
versionhallinta; 108
vertailtavuus; 89
vertailuajankohta; 104
vertailukelpoisuus; 6, 8, 10, 11, 14, 24, 26, 27, 43, 94, 107
vertaisarviointi; 22
viestintästrategia; 112
viivakuvio; 92
viivakuviot; 90
virallinen tilasto; 13
virheettömyys; 62, 79, 82
virheiden korjaus; 117
virheiden oikaisu; 23
virhelähteet; 16, 25, 84
väestötietojärjestelmä; 57

X

X-12-ARIMA; 95

Y

ydinprosessit; 15, 17, 18
yhdistelytiedot; 46
yhteiskunnallinen vastuu; 16
yhtenäisyys; 6, 10, 11, 14, 18, 24, 27, 107
YKn kansainvälisten tilastosarjojen pääjako; 13
yksikköarvoindeksi; 100
yksikkökato; 71, 72, 78, 82, 85
yksikkökohtaisuus; 54
yksikkötason tiedot; 13, 35, 36, 37, 38, 39, 49, 65
yksinkertainen satunnaisotanta; 57, 58, 61, 80
yksiselitteisyys; 42, 46, 48, 69
yleistettävyys; 28, 46, 56, 62
ylipeitto; 47, 48
ymmärrettävyys; 112
ympäristötilastoluokitukset; 41
yritys- ja toimipaikkarekisteri; 35
yritys- ja yhteisötunnus; 49, 51
yritysrekisteri; 48

Käytetyt lyhenteet

AMS, Tilastokeskuksen arkistonmuodostussuunnitelma	EY, Euroopan yhteisö
BSC, Balanced scorecard (Tasapainoinen tuloskortti)	ISI, International Statistical Institute (Kansainvälinen tilastoinstituutti)
CALMAR, Calage sur marges (calibration on margins)	ISO, International Organization for Standardization
CAPI, Computer Assisted Personal Interview	JHS, julkisen hallinnon suositus
CASI, Computer Assisted Self Interview	MSE, Mean square error
CASI-A, Text on Screen and on Audio	NACE, Nomenclature Générale des Activités Economiques dans les Communautés Européennes (Statistical classification of economic activities in the European Community)
CASIC, Computer Assisted Survey Information Collection	NSI, National Statistical Institute (kansalliset tilastovirastot)
CASIIP, Computer Assisted Self Interviewing with Interviewer Present	NUTS, Nomenclature des Unités Territoriales Statistiques (Nomenclature of territorial units for statistics; Euroopan unionin virallinen alueluokitusjärjestelmää)
CASI-V, Question Text on Screen	PPS, Probabilities Proportional to Size
CATI, Computer Assisted Telephone Interview	PRODCOM, List of products of the European Community (Teollisen tuotannon tilastoinnissa käytetty luokitus)
CAWI, Computer Assisted Web Interview	RAMON, Eurostatin luokitustietokanta
CoP, Code of Practice (Käytännäsäännöt)	SRS, Simple random sampling
CPA, Classification of Products by Activity (Tuoteluokitus)	Stakes, National Research and Development Centre for Welfare and Health
DBM, Disk by Mail	SVT, Suomen virallinen tilasto
deff, design effect	SYS, Systematic sampling (Systemaattinen otanta)
DESAP, Development of Self-Assessment Program	TK, Tilastokeskus (Statistics Finland)
ECE, Economic Commission for Europe (Euroopan talouskomissio)	TQM, Total quality management
EFQM, European Foundation for Quality Management	UN, United Nations
ESA, European system of accounts (uudistetun kansantalouden tilinpidon ESA 95 perusluokitus)	UNSD, United Nations' Statistical Department
EU, European Union	YTJ, Yritys- ja yhteisötietojärjestelmä
Eurostat, The Statistical Office of the European Communities	

Kuvailulehti

Julkaisija
Tilastokeskus

Julkaisun päivämäärä
Tammikuu 2007

Tekijä: Johanna Laiho	Julkaisun laji Käsikirja
	Toimeksiantaja Tilastokeskus
	Toimielimen asettamispvm

Julkaisun nimi
Laatua tilastoissa 2. uudistettu painos

Julkaisun osat

Tiivistelmä

Tilastojen tuotantoprosessi on Tilastokeskuksen toiminnan ydin. Laatua tilastoissa -käsikirjan tavoite on parantaa tilastosysteemien suunnitteluun ja toteutumiseen 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 978-952-467-655-7
Kokonaissivumäärä 130	Kieli Suomi	Hinta 30 €	Luottamuksellisuus Julkinen
Jakaja Tilastokeskus Työpajakatu 13, Helsinki 00022 TILASTOKESKUS		Kustantaja Tilastokeskus	

KÄSIKIRJOJA
HANDBÖCKER
HANDBOOKS

Nro 1	Koulutusluokitus 2003 Utbildningsklassificering 2003		2004
	Koulutusluokitus 2000 Liite 1 ISCED 1997 Suomalainen soveltamisopas 2000 Koulutusluokituksen ja Unescon kansainvälisen koulutusluokituksen ISCED 1997 välinen koodiavain	12. uusittu laitos	2001
	Liite 2 Opetushallinnon koulutusala- ja koulutusasteluokitus Koulutusluokituksen ja opetushallinnon koulutusala-, opintoala- ja koulutusasteluokituksen välinen koodiavain		
	Liite 3 Koulutuskoodimuutokset 2000		
	Liite 3 Koulutuskoodimuutokset 2001 <i>Bilaga 3 Ändringar i utbildningskoderna 2001</i>		2002
	Liite 3 Koulutuskoodimuutokset 2002 <i>Bilaga 3 Ändringar i utbildningskoderna 2002</i>		2003
	Liite 3 Koulutuskoodimuutokset 2003 <i>Bilaga 3 Ändringar i utbildningskoderna 2003</i>		2004
Nr 1	<i>Utbildningsklassificering 2000</i>	8:e reviderade upplagan	2001
Nro 4	Toimialaluokitus TOL 2002 Pääjulkaisu		2002
	Toimialaluokitus TOL 2002 Liite 1 Hakemisto		2002
	Toimialaluokitus TOL 2002 <i>Näringsgrensindelningen TOL 2002</i> Standard Industrial Classification TOL 2002 Liite 2 Tiivistelmä <i>Bilaga 2 Sammandrag</i> Annex 2 Summary		2002
Nro 5	Sektoriluokitus 2000 <i>Sektorindelningen 2000</i> Classification of Sectors 2000		2000
Nro 6	Rahoitusvaadeluokitus 1996 Classification of financial assets and liabilities 1996		1995
Nro 10	Yhteisöjen tehtävuokituksset Julkisyhteisöjen ja voittoa tavoittelemattomien yhteisöjen tehtävuokituksset <i>Uppgiftsklassificeringar för sammanslutningar</i> <i>Uppgiftsklassificeringar för offentliga sammanslutningar</i> <i>och icke vinstsyftande sammanslutningar</i> Classifications of the functions of 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 2004 <i>Stater och länder 2004</i> Countries 2004		2004
Nro 14	Ammattiluokitus 2001 Pääjulkaisu		2001
	Ammattiluokitus 2001 Liite 1 Hakemisto, huhtikuu 2005		2005
	Ammattiluokitus 2001 <i>Yrkesklassificeringen 2001</i> Classification of Occupation 2001 Liite 3 Tiivistelmä <i>Bilaga 3 Sammandrag</i> 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 Kiinteä 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> <i>Riktlinjer för användning av åldersklassificeringar</i>		1986
Nro 27	Rikosnimikkeistö <i>Brottsnomenklatur</i> Crime nomenclature		1999
Nro 28	Kunnat ja kuntapohjaiset aluejaot 2007 <i>Kommuner och kommunbaserade indelningar 2007</i> <i>Municipalities and Regional Divisions Based on Municipalities 2007</i>		2007

Nro 30	Tilastokeskuksen ammattieettinen opas		2006
Nro 30b	Guidelines on Professional Ethics		2006
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		2005
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	uusittu laitos	2007
Nro 43b	Quality Guidelines for Official Statistics		2002
Nro 44	Yksilöllisen kulutuksen käyttötarkoituksen mukainen luokitus (COICOP)		2002
Nro 45	Use of Registers and Administrative Data Sources for Statistical Purposes		2004

Tiastojen tuotantoprosessi on Tilastokeskuksen toiminnan ydin. Laatua tilastoissa -käsikirjan tavoite on parantaa tilastosysteemien suunnitteluun ja toteuttamiseen liittyvään tietämyksen ja osaamisen käytettävyyttä ko-koamalla olemassa olevat periaatteet yhteiseksi osaamisräömäksi.

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

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

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

ISSN 0355-2063
= Käsikirjoja
ISBN 978-952-467-655-7
Tuotenumero 3174
CO