Tutkimuksia Undersökningar Studies
Nro 104

Tilastojen laadun tutkiminen

HELSINKI 1984

Tilastokeskuksen laatupäivän esitelmät

Tilastokeskus
Statistikcentralen
Central Statistical Office of Finland

Tutkimuksia
Nro 104

Undersökningar

Studies

Tilastojen laadun tutkiminen

Tilastokeskuksen laatupäivän esitelmät

Tilastokeskus
Statistikcentralen
Central Statistical Office of Finland

HELSINKI 1984

ISSN 0355-2071
ISBN 951-46-7595-9

Helsinki 1984. Valtion painatuskeskus

ESIPUHE

Tilastokeskus järjesti 18.5.1983 teemapäivän aiheesta tilastojen laatu. Teemapäivän tavoitteena oli etsiä vastauksia kysymyksiin:

- mitä tarkoitetaan tilastojen laadulla
- kuinka laatua mitataan
- kuinka tilaston käyttäjille kerrotaan laadusta
- mitä meidän pitäisi tehdä tässä asiassa.

Tilaisuuteen osallistui Tilastokeskuksen johtoa, tilastojen vastuuhenkilöitä ja muiden tilastovirastojen edustajia yhteensä 58 henkilöä. Julkaisemalla teemapäivän esitelmät Tilastokeskus toivoo edistävänsä keskustelua tilastojen laadusta.

Helsingissä lokakuussa 1983

Olavi E. Niitamo

Erkki Pentinmäki

408400134V—12

SISÄLLYS

Sivu

Pääjohtaja Olavi E. Niitamo: Laatupäivän avaussanat 1

Toimistopäällikkö Erkki Pentinmäki: Tilaston laadusta 3

Suunnittelija Paavo Väisänen: Laadun tutkiminen Yhdysvaltojen ja Ruotsin tilastovirastojen suositusten nojalla 9

Yliaktuaari Carita Putkonen: Väestölaskennan laatututkimus; esimerkki laadun mittaamisesta 16

Suunnittelija Pasi Markelin: Laatuongelmat rekistereitä käytettäessä 28

Yliaktuaari Tuula Rissanen: Työvoimatutkimuksen laadun tutkiminen.... 38

Yliaktuaari Tuula Koskenkylä: Laadun tutkiminen yritystilastoissa ... 45

Suunnittelija Esa-Jukka Käär: Laatukysymykset kansantalouden tilinpidon näkökulmasta 64

Suunnittelija Pirjo Liewendahl: Yhteenveto yleiskeskustelusta 74

Pääjohtaja Olavi E. Niitamo

LAATUPÄIVÄN AVAUSSANAT

Yhteiskunnan toiminnan ja kehittämisen on perustuttava todenmukaiseen, riittävään ja luotettavaan tietoon yhteiskuntaoloista. Valtion tilastotoimen tehtävänä on tuottaa suuri osa näistä yhteiskunnan tarvitsemista tiedoista. Voimassa olevassa valtion tilastotoimen kehittämisohjelmassa on tietojen luotettavuus korostuneesti esillä. Tilastoja laativien valtion viranomaisten on seurattava paitsi tuottamiensa tilastotietojen tarpeellisuutta myös niiden luotettavuutta.

Tilastoneuvosto on pitänyt asiaa niin tärkeänä, että se asetti vuosi sitten erityisen jaoston selvittämään tilastojen laatu- ja jalostuskysymyksiä. Professori Fellmanin johdolla toiminut jaosto on saanut työnsä valmiiksi ja jaoston ehdotuksia käsitellään täällä tänään.

Asia on myös kansainvälisesti ajankohtainen. Euroopan talouskomission tilastoyhteistyönä on valmisteilla suunnitelma laatuselosteiden laatimiseksi. Asiaa käsitellään vielä tämän vuoden lopussa. Kuten tämän päivän esityksistä käy ilmi, on tällaisia suosituksia ainakin Ruotsin ja Yhdysvaltain tilastotoimella. Meiltä ne ovat vielä puuttuneet. Se ei toki merkitse sitä, että alue olisi meillä täysin laiminlyöty. Jonkin verran laadun tutkimista on harjoitettu ja siitä käyttäjillekin tietoa välitetty. Se on ollut kuitenkin riittämätöntä, sattumanvaraista ja epäsystemaattista.

Kun puhumme täällä tilastojen laadusta on syytä rajata keskustelun kohde. Me emme tarkastele tilastotoimen palvelukykyä suhteessa yhteiskunnan tietotarpeeseen. Jotkut voivat käyttää laatu-sanaa kuvaamaan yleensä palvelukykymme tehokkuutta ja pitää laatua huonona jos havaitaan tilastotuotannossa ja -palvelussa aukkoja. Joku voi todeta oikeutetusti puutteena esimerkiksi sellaisen seikan, että emme nykyisellään pysty tarjoamaan tietoa informaatioelinkeinon parissa työskentelevien lukumäärästä tai informaatioelinkeinon osuudesta kansan-

tuotteesta. Tämä on puute, mutta tällaista puutetta emme sisällytä tänään tarkastelun alaisena olevan laatukäsitteen piiriin. Laadulla tarkoitamme yksittäisen tilaston tai sen osatekijän ominaisuuksia.

Hyvät kuulijat! Tämä teemapäivä kuuluu osaltaan Tilastokeskuksen johdon koulutukseen. Johdon koulutuksen KTS-ohjelmassa on tuotevastuu nähty tärkeänä osana. Me kaikki tilastotoimen johtotehtävissä toimivat vastaamme tuotteittemme korkealaatuisuudesta. Olkoon tämä päivä virittämässä keskustelua laadusta ja vaikuttamassa suunnitelmallisen laaduntarkkailutyön aikaansaamiseen ja käyttäjien nykyistä parempaan informoimiseen laatutekijöistä.

Toivon virikkeistä teemapäivää kaikille osallistujille. Erityisesti toivotan tervetulleeksi muista virastoista paikalla olevat valtion tilastotoimessa työskentelevät kollegat.

Toimistopäällikkö Erkki Pentinmäki

TILASTON LAADUSTA

Viime vuosina on Tilastokeskuksen määräaikaisuunnitelmissa laadun kohottaminen ollut selvästi esillä. Laatu tai luotettavuus on mainittu suunnitelmissa rinnan muiden tavoiteltavien asioiden kanssa, joita ovat yleensä olleet nopeuttaminen ja jalostusasteen nostaminen. On jo merkkejä siitä että huolenpito laadusta on siirtymässä liturgian tasolta konkreettisiksi luotettavuutta koskeviksi tutkimuksiksi. Tilastotoimen kehitystä koskeva seurantaraportti tältä vuodelta mainitsee kolme esimerkkiä. Väestö- ja asuntolaskennan yhteydessä on tehty laaja luotettavuustutkimus. Työvoimatiedustelun luotettavuutta on tutkittu ja tuloksia käytetty hyväksi tilastojenkeruun uudistuksessa. Kolmantena esimerkkinä on tie- ja vesirakennushallituksen tavarankuljetustilaston yhteydessä tehty luotettavuusselvitys. Nämä eivät varmastikaan ole ainoita tilastojen laatua koskevia selvityksiä tilastotoimen piirissä, ovatpahan vain esimerkkejä huomattavista luotettavuustutkimuksista.

Olemme puhuneet tilaston laadusta, mutta tähän asti meiltä on puuttunut yhteisesti sovittu määritelmä siitä, mitä sillä tarkoitamme. Laatukäsitteemme on ollut hieman autereinen. Jokainen on liittänyt siihen omia mielikuviaan niistä ominaisuuksista, joiden avulla tilaston laatua tulisi arvioida. Tarkoittaahan sana laatu sinänsä ominaisuuksia, joita jollakin on tai jotka johonkin kuuluvat. Tilaston laatu on siten laajasti käsitettynä kaikkien niiden ominaisuuksien joukko, jotka tilastolla on. Tarvitsemme kuitenkin laatukäsitteen operationalisoimista.

Perinteen vaikutus laadun tutkimiseen

Tilastotyön ja organisaatiomme perinnetekijät vaikuttavat siihen miten laatua valvomme, tutkimme ja siitä tietoja annamme. Myönteisenä

piirteenä voimme panna merkille sen, että arvostus hyvää laatua kohtaan on tilastotyössä aivan luontaista. Se johtuu siitä, että työssämme vaikuttaa tutkijaperinne. Tilastotyö on hyvin lähellä tutkimustyötä. Varsinkin suunniteltaessa tilastollista kuvausta uudelta ongelma-alueelta on rajan tekeminen tilastotyön ja tutkimustyön välillä melko hankalaa. Tilastokeskuksen johdolla aina työnjohtoporrasta myöten on tutkijakoulutus. Toisaalta vastuun kantaminen hyvästä laadusta johtaa vääjäämättä tutkimuslaitostyyppiseen otteeseen. Tilastotoimi tarvitsee tuotantoa palvelevaa tutkimusta siinä kuin mikä muu tuotannonala hyvänsä. Laatua koskeva selvitystyö sekä tutkimus- ja jatkojalostusasiat liittyvät läheisesti yhteen. Tilastoneuvoston asettaman laatujaoston työ jäi siinä mielessä puolitiehen, että jaoston piti itse asiassa käsitellä tilastojen laatu- ja jalostuskysymyksiä, mutta laatutyön konkretisoimisessa oli niin paljon puuhaa että jalostus- ja tutkimusasiat jäivät vain maininnalle.

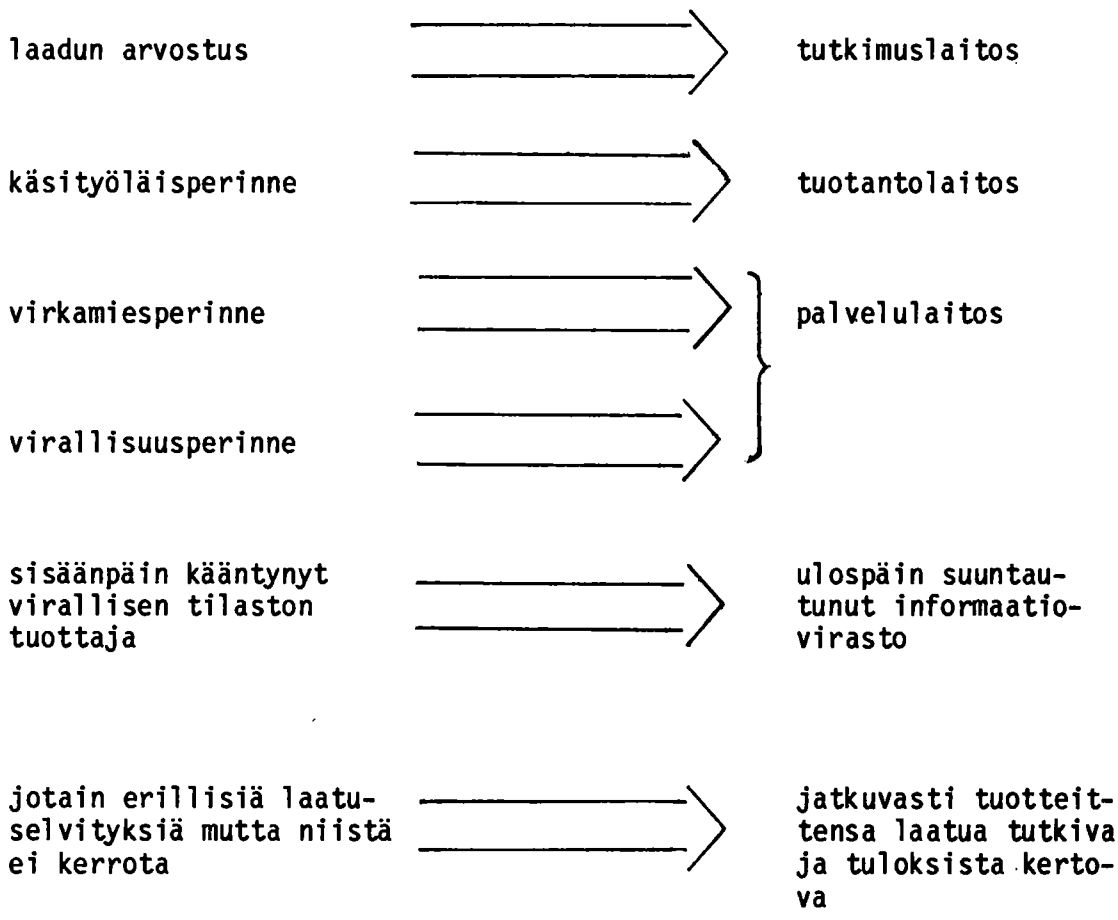
Organisaatiossamme on myös perinnetekijöitä, jotka saattavat olla haittaamassa ja hidastamassa laatutukimusten aloittamista. Tällaisista perinnetekijöistä mainitsisin tilastotyön alkuajoista juontuvan käsityöläisperinteen, asennepuolella vaikuttavan virkamiesperinteen ja ympäristöpuolella esiintulevan virallisuusperinteen. Käsityöläisperinteelle on ominaista henkilökohtainen vastuu tuotteesta ja sen laadusta. Valvonta on välitöntä ja korkea laatu pyritään takaamaan joka vaiheen huolellisella teolla ja henkilökohtaisella valvonnalla tuotannon kaikissa vaiheissa. Korkeaan laatuun on pyritty vanhastaan ihmistyötä säästämättä. Koko aineistoon kohdistuvat totaaliset tarkistukset ovat vanhaa perinnettä. Koko aineiston tarkistuskoodauksen ja tarkistustallennuksen kaltaisilla toimenpiteillä on toivottu saatavan aikaan mahdollisimman hyvä laatu. Tällaiseen käsityöperinteeseen sopii huonosti sana laaduntarkkailu. Sehän on jotain ulkokohtaista ja ikään kuin ammattityötä negatiivisesti arvostelevaa. Virkamies- ja virallisuusperinne ovat lyöneet myöskin oman leimansa toimintatapaamme. Kun tilastot ovat aina olleet lähellä hallintoa ja hallinollisten aineiston käyttöä, on syntynyt tietty virallisuusperinne, jonka takia on ollut ikään kuin oman virkatoimensa kyseenalaiseksi asettamista ryhtyä esittelemään tilastonsa virheitä ja virhe-

lähteitä. Suomen virallisen tilaston sarjassa on tuntunut vieraalta puhua virallisen tilaston virheistä. Epäilenpä, että tämäkään perinne ei ole kuollut vaan se asustaa virkamiessieluissamme ja on ollut estämässä laadun tutkimista ja virheistä kertomista. Kun julkaisen virhearviot, voivat panna vastuuseen virheistä ja vaikkapa lopettaa koko tilaston, saattaa virkamiesminämme alitajuisesti ajatella.

Tilastonkäyttäjätkään eivät ole aiemmin vaatineet tietoja tilaston laadusta. Joko ajattelevat, että virallisen tilaston on oltava luotettavaa ja ottavat totena tai korkeintaan ottavat esille yksittäisiä epäluotettavuuksia vaatimatta jatkuvasti selontekoa laadusta. Toisin sanoen tilastotoimen ympäristö elää myöskin virallisuusperinteen vallassa.

Ajan kuluessa on kuitenkin syntynyt uusia vaateita, jotka vaikuttavat näiden perinnetekijöiden vastaisesti ja vaativat meitä muuttamaan toimintatapojamme. Alla on kuviossa 1. kuvattu sitä, mihin suuntaan kehitys kulkee tilastotoimessa.

Kuvio 1. Kehityslinjat perinnetaustaan nähden



Olemme siis muuttumassa ja meidän on muututtava sisäänpäin kääntyneestä virallisen tilaston tuottajasta ulospäin suuntautuneen informaatioviraston suuntaan. Laatututkimuksiin ja laadusta kertomiseen tällä on se vaikutus, että meidän on työssämme otettava laadun tutkiminen omaksi osakseen tuotantoa ja kerrottava työmme laadusta myöskin ulospäin.

Laatukäsite

Kun tilaston laatua ryhdytään tutkimaan ja siitä kertomaan, on se ensiksikin määriteltävä. Laadusta on tullut standardoimiskohde. Tilastoneuvoston laatujaosto on raportissaan tutkaillut laadun käsitettä. Tilaston laadusta puhuttaessa on tahdottu rajautua yksittäisen tilaston ominaisuuksiin poissulkien arvioinnin tilastojärjestelmän tehokkuudesta tai palvelukyvystä. Työn laatu ja hinta ovat aina yhteydessä keskenään, niin myöskin tilastotyössä. Itse asiassa tuotteen laadun muodostavat ne ominaisuudet, joista käyttäjät ovat valmiita maksamaan. Laatu voi olla liian huonoa mutta se voi olla myöskin johonkin käyttötarkoitukseen nähden tarpeettoman hyvää ja näin ollen tarpeettoman kallista. Laatu on siten aina suhteutettava tiettyyn tarpeeseen tai käyttötarkoitukseen. Virallinen tilasto on erittäin ongelmallinen yleiskäyttökelpoisuutensa takia. Virallista tilastoa halutaan helposti virallisuutensa takia käyttää mihin käyttötarkoitukseen hyvänsä. Tämä johtaisi vääjäämättä ylikorkeaan laatuun ellei virallisen tilaston kohdallakin käyttökelpoisuutta rajata.

Laatu käsitteen operationalisointi on sidoksissa kolmeen tekijään:

1 laatua mitataan käänteisesti eli virheiden avulla
2 tavoitetaso on suhteellinen käsite ja riippuvainen käyttötarkoituksesta
3 mittaustulokset on esitettävä käyttäjille.

Laadun tutkimiseen tai kuvaamiseen on kaksi tapaa: tietyn virhemallin soveltaminen tai laadun jakaminen osatekijöihin. Laadun tutkimisen ja

laatutulosten esittämisen pitäisi olla meidän tilastoihmisten leipälaji, koska laatutukimus tapahtuu tilastollisten menetelmien avulla. Tilastolliset menetelmät tulivat teollisen laadun tutkimukseen 1920-luvulla. Klassinen teos on Walter A. Shewhartin Economic Control of Quality of Manifactured Product, joka perustuu Bell-tehtaissa tehtyihin kokeisiin ja tutkimuksiin. Sodan aikana laaduntutkimus sai lisävauhtia. Yhdysvaltoihin perustettiin erityinen Columbian yliopiston ryhmä kehittämään tieteellisiä menetelmiä. Eräs mainio lopputulos on Abraham Waldin sekvenssianalyysi. Laatua pitää tutkia ja siitä tulee puhua tilastollisten menetelmien ja tunnuslukujen avulla.

Laatuselosteet

Tietoa laadusta tarvitsee tilaston suunnittelija, tuottaja ja käyttäjä. Tämän takia laadusta täytyy kertoa erityisten laatuselosteiden avulla. Niitä suunniteltaessa on etenkin ajateltava käyttäjän tarvetta. Tilastoneuvoston laatujaosto suosittelee kahta laatuselosteen muotoa: varsinaista laatuselostetta ja teknistä raporttia. Laatuseloste riittää tavalliseen käyttöön kun taas tekninen raportti on varattu yksityiskohtaiseen erityiskäyttöön. Laatujaosto teki myöskin selvityksen siitä mitä tietoa käyttäjät tilaston laadusta haluavat. Yleistoteamus on, että käyttäjät painottavat enemmän laadun relevanssipuolta eli sisältöä ja oikea-aikaisuutta kuin kokonaisvirhepuolta. Käyttäjäryhmien välillä on kuitenkin eroja siinä miten mielenkiinto suuntautuu laadun eri osatekijöihin. Yliopistot tuntuivat painottavan enemmän kuin muut ryhmät myöskin laadun kokonaisvirhekomponenttia. Tässä esityksessä ei ole tarkoitus lähemmin tarkastella mitä laadun osatekijöitä laatuselosteessa tulisi käsitellä. Sen sijaan on paikallaan mainita mitkä ovat tärkeimmät laatuominaisuudet käyttäjäkyselyn perusteella. Eli mistä laadun osatekijöistä käyttäjät ovat kiinnostuneempia. Viisi tärkeintä ominaisuutta ovat seuraavat:

1 tilaston vertailtavuus ajassa
2 vertailtavuus muihin tilastoihin
3 käyttökelpoisuutta koskevat rajoitukset

4 käsitteiden määrittely

5 julkaisemattomien tulosten tallennusmuoto ja saatavuus.

Eräs käyttäjäkyselyssä esiin tullut piirre oli se, että mitä tärkeämmäksi joku tilasto koetaan, sitä enemmän siihen asetetaan lisävaateita eli ruokahalu kasvaa syödessä. Kun tilaston käyttäjä on tottunut saamaan tietyn tasoisen tuotteen, hän haluaa helposti entistä parempaa. Tässäkin näkyy tavoitetason suhteellisuus.

Laatujaoston suositukset

Tilastoneuvoston laatujaosto esitti seuraavat suositukset:

1 laaduntarkkailun tulee muodostua olennaiseksi osaksi tilastotyötä. Tämän takia suunnitelmissa on kiinnitettävä erillisesti huomiota laatuun osatekijöittäin. On laadittava teknisiä raportteja. Tuotantomenetelmien kehittämishankkeen yhteydessä olisi selvitettävä tarvittavat laadunvalvontamenetelmät

2 jokaiseen tilastoon tulee liittää riittävän yhdenmukainen laatuseloste

3 on tehtävä laatua koskevia tarvetiedusteluja.

Suunnittelija Paavo Väisänen

LAADUN TUTKIMINEN YHDYSVALTAIN JA RUOTSIN TILASTOVIRASTOJEN SUOSITUSTEN NOJALLA

Seuraavassa tarkastelen lähinnä Ruotsin tilastoviraston SCB:n laatumääritelmiä tilastotiedoille. SCB on omaa tilaston laatuselosteohjeistoaan tehdessään käyttänyt hyväksi mm. Yhdysvaltojen ja Kanadan tilastovirastojen laatuohjeistoja. SCB:n laatumalli on laajempi verrattuna esim. Yhdysvaltojen laatuohjeistoon (2). Tässä julkaisussa on esitetty yksityiskohtaisia ohjeita, miten tilastoon sisältyviä virheitä on analysoitava ja esitettävä. Samoin Kanadan tilastovirasto Statistics Canada on julkaissut oman ohjeistonsa (3). Virheiden analysointimalleja on edelleen kehitelty Survey Methodology sarjassa julkaistuissa artikkeleissa. Mielenkiintoista on, että kehittelyn pohjana on ollut mm. Kishin 1965 julkaisema virhemalli.

1 Mitä hyötyä laatuselosteista on ja miten tai mihin niitä voidaan käyttää?

- Kiinnitetään tilaston käyttäjien huomiota tekijöihin jotka rajoittavat tilaston käyttöä. Selostetaan tilastossa olevat virhelähteet.

- Antaa käyttäjälle paremmat lähtökohdat päätöksenteolle. Virheellisten päätösten riskiä voidaan vähentää.

- Käyttäjillä on paremmat mahdollisuudet määritellä vaatimuksensa tulevia tutkimuksia ajatellen.

2 Mitä on tilaston laatu? Miten tilaston laatu määritellään?

Laatu pitää aina suhteuttaa tilastotietojen käyttöön. Tilastollinen havaintoaineisto voi olla riittävä laatunsa

suhteen tiettyyn tarkoitukseen ja riittämätön johonkin toiseen tarkoitukseen. Tilastoissa voidaan ja pitäisi laatu määritellä ja selostaa jonkin virhemallin mukaan. Useita virhemalleja on kehitelty ja niiden kehittely näyttää koko ajan jatkuvan. Tunnetuin virhemalli lienee USA:n Bureau of Censuksen kehittämä. Siinä tilastoaineiston laatu määritellään sen eron perusteella, joka on saatujen tulosten ja niiden tulosten välillä, jotka olisi saatu ihanteellisissa tietojenkeruu- ja käsittelyolosuhteissa. Tämä määritelmä implikoi kokonaisvirheen käsitteen. Tilaston kokonaisvirhe lasketaan ja selostetaan jonkin virhemallin mukaisesti. Tämä määritelmä on yleisluontoinen eikä sinällään sovellu mihinkään tiettyyn virhemalliin.

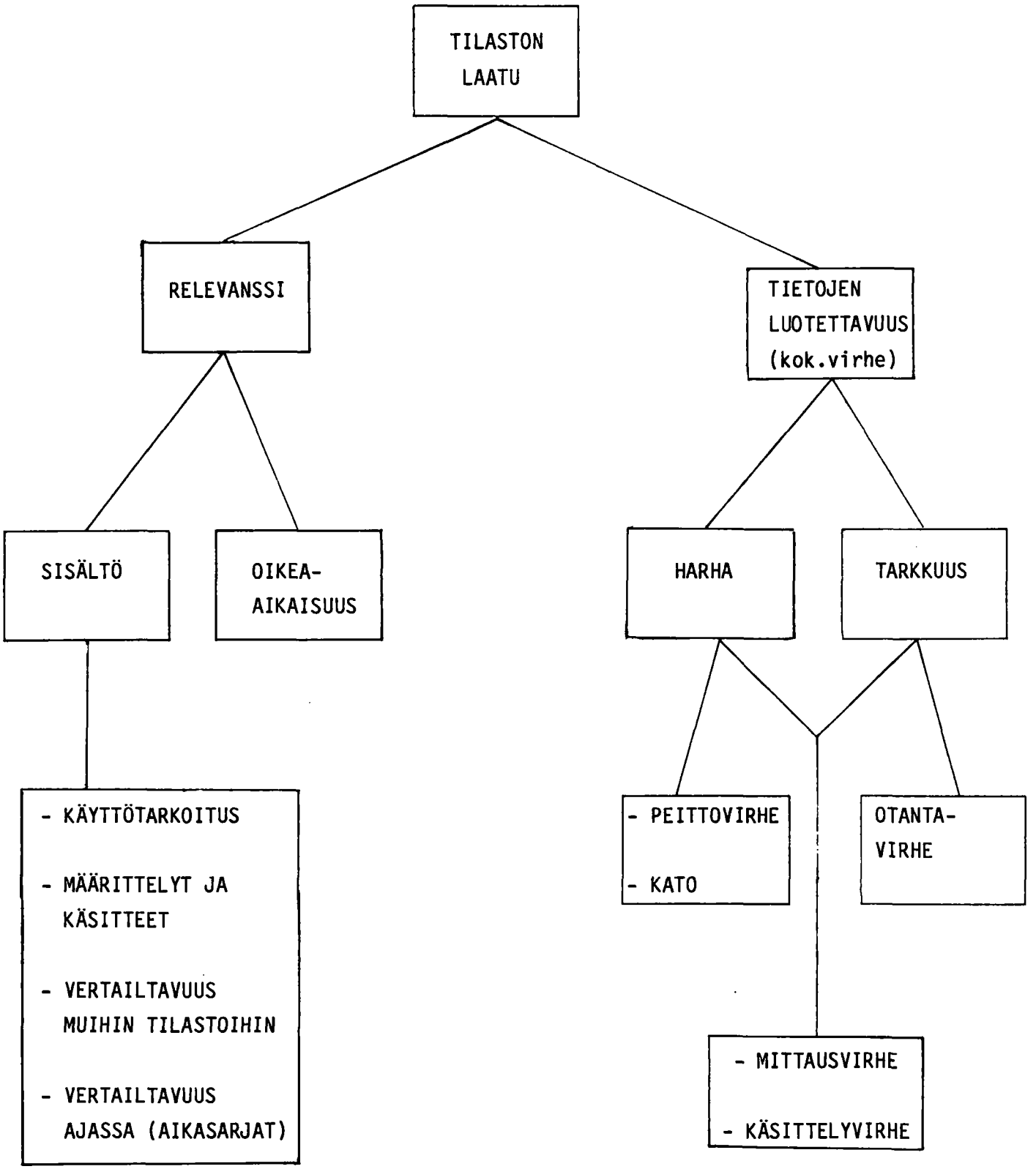
Sen sijaan laatu suhteutetaan tilastotietojen keskeisiin ominaisuuksiin, jotka ovat relevantteja ensi sijassa sellaisille tavanomaisille tilastollisille tunnusluvuille kuten kokonaissummat, keskiarvot ja prosenttiluvut. Soveltuvin osin mallia voidaan käyttää myös monimutkaisimmille tilastollisille tiedoille kuten esim. regressio- ja korrelaatiokertoimille.

SCB:ssä laatua on analysoitu seuraavalla mallilla (kuvio 2), jossa laatuominaisuudet jaetaan kahteen päätyyppiin, relevanssitekijöihin ja virhetyyppeihin.

3 Mitä laatuselosteen tulisi sisältää?

Kullakin tilaston käyttäjällä on tietty ongelma, jonka hän toivoo voivansa ratkaista tilastotietoja käyttämällä. Jokaisen tilastollisen selvityksen pitää siksi sisältää juuri sitä tietoa, informaatiota, jota tietojen käyttäjä tarvitsee, jotta tilasto olisi käyttökelpoinen hänen tarpeisiinsa. Seuraavat tilastotietojen relevanssia kuvaavat kohdat tulisi sisällyttää SCB:n mukaan tilastollisiin selvityksiin.

Kuvio 2. Tilastotietojen laatumalli SCB:n mukaan



- Tutkimuksen tarkoitus ja käyttömahdollisuudet
  Selvitetään, miksi tilasto on tehty ja mitä se kuvaa, jos se ei ole ilmeisen selvää. Jos on syytä olettaa, että tilastoa voidaan tulkita väärin tai käyttää virheellisesti, tästä pitää varoittaa käyttäjiä.

- Käsitteet, määritelmät ja luokitukset
  Tutkimusperusjoukko ja tutkimuskohde, -alkio, määritellään yksikäsitteisesti. Muuttujat ja luokitukset, jotka eivät ole itsestään selviä, pitää myös selostaa. Kysymyslomake suositellaan liitettäväksi mukaan varustettuna kommenteilla siitä, miten kysymykset on muotoiltu ja mitä vastausvaihtoehtoja on saatu.

- Vertailtavuus muihin tilastoihin
  Selostetaan muut samaa aihetta käsittelevät tilastot ja tutkimukset. Miten käytetyt määritelmät ja luokitukset poikkeavat. Mikä on eri virheiden vaikutus mahdollisiin eroihin tuloksissa.

- Oikea-aikaisuus
  Selvitetään mitä ajankohtaa tai ajanjaksoa tilasto kuvaa.

  Laatupäivänä tuotiin esille seuraavat näkökannat: Tilastotietojen pitäisi olla käytettävissä sellaisena ajankohtana, jolloin niitä tarvitaan. Tilastojen tulee valmistua riittävän nopeasti.

Tietojen luotettavuuden kannalta kokonaisvirheellä on oleellinen merkitys käyttäjille. Kokonaisvirhe voidaan selvittää joissakin tapauksissa luotettavuustutkimuksella.

Se on kuitekin melko vaikeaa ja kallista. Aiheellista on kuitenkin selvittää kunkin tutkimuksen kannalta oleelliset virhelähteet. Jos tilastotietojen kokonaisvirhettä ei voida estimoida, niistä pitäisi esittää vähintään edes karkea arvio.

- Peittovirheet
  Kun kehikkoperusjoukko ei vastaa tutkimusperusjoukkoa, syntyy peittovirhettä. Yli- ja alipeitto selostetaan jos se tunnetaan. Peittovirheen vaikutus tuloksiin lasketaan tai ellei se ole mahdollista, virheestä esitetään arviot.

- Kadon aiheuttama virhe
  Esitetään kadon suuruus tärkeimpien tutkimusosajoukkojen suhteen. Kadon jakauma esitetään kadon syiden mukaan luokiteltuna, jos ne ovat tiedossa. Lisäksi esitetään osittaiskato muuttujille, joilla kato on suuri. Kadon aiheuttaman virheen suuruus arvioidaan. Lisäksi ulee selittää mitä keinoja on käytetty kadon aiheuttaman virheen pienentämiseksi kuten esim. sijaistaminen, painottaminen tai mahdollinen katotutkimus.

- Mittausvirhe
  Jos tutkimuksen mittausvirheestä on saatu tietoja, ne tulisi liittää julkaistavien tilastotietojen yhteyteen. Ellei mittausvirhettä tunneta, esitetään mitä virhelähteitä sisältyy tutkimuksen tietojenkeruuvaiheeseen. Mahdollisuuksien mukaan estimoidaan virheen suuruus ja suunta.

- Käsittelyvirheet
  Aineiston koodauksen ja tallennuksen yhteydessä

syntyneet virheet selvitetään ja lasketaan niiden vaikutus tuloksiin.

- Otantavirhe
  Esitetään otannasta aiheutuva satunnaisvirhe keskivirheen tai luottamusvälin avulla.

4 Laatuselosteiden muotoilu ja sijoittaminen

Laatuselosteet tehdään tilastojen käyttäjiä varten. Ne pitää siten laatia aina käyttäjien näkökulmasta lähtien. Laatuselosteen tulee sen vuoksi olla helposti luettava. Liian laajoja laatuselosteita, jotka hämärtävät tilaston laadulliset ominaisuudet on ehdottomasti vältettävä. Samoin tilastotieteellisiä termejä, teknisiä käsitteitä, monimutkaisia matemaattisia kaavoja olisi vältettävä, jos se vain on mahdollista. Käytetty tilastotieteellinen terminologia on selvitettävä tavalla, joka on ymmärrettävää kenelle tahansa.

Laatuselosteet tulisi laatia samanlaisiksi jonkin standardin mukaan. Tällöin käyttäjät vähitellen oppivat lukemaan niitä ja selosteen käyttö tulee vaivattomaksi.

Käyttäjät lukevat tavallisesti suoraan taulukkoa, kun he tarvitsevat tilastotietoja. Tämä pätee yhtä hyvin julkaisuihin ja raakatauluihin kuin päätteellä käytettäviin tietokantoihin. Siksi pitäisi sellainen laatua koskeva informaatio, jolla on merkitystä tietyn taulukon käytön kannalta, sijoittaa sen loppuun. Tämä laatuseloste voisi olla melko lyhyt. Ellei lyhyttäkään selostetta saada mahtumaan taulukon loppuun, siihen on lisättävä viittaus, mistä yksityiskohtaiset kuvaukset laadusta ovat saatavissa. SCB:n mukaan koko tutkimusta koskeva laatuseloste pitäisi sijoittaa raportin alkuun esimerkiksi johdantoosaan.

Yksityiskohtainen laatuseloste, joka voi olla vaikeasti luettavaa tavalliselle tilaston käyttäjälle, suositellaan tehtäväksi erilliseksi liitteeksi tai raportiksi.

Kirjallisuutta

(1) Statistiska Centralbyrån, Riktlinjer för kvalitetsredovisning av statistik, 1979

(2) U.S. Bureau of the Census, Standard for discussion and presentation of errors in data, Technical Paper No 32, U.S. Government printing office, Washington D.C., 1974

(3) Statistics Canada, Policy of the Production and dissemination of Mesures of Data Quality and Methodology reports for Statistics Canada Surveys and their Outputs, 1978

(4) Kanadan tilastoviraston julkaisusarja Survey Methodology - Techniques d'enquete

(5) Kish L., Survey Sampling, John Wiley & Spms. N.Y., 1965, ss. 509-571

Yliaktuaari Carita Putkonen

VUODEN 1980 VÄESTÖ- JA ASUNTOLASKENNAN LUOTETTAVUUSTUTKIMUS

Vuoden 1980 väestö- ja asuntolaskennasta tehty laatututkimus on ensimmäinen Suomessa suoritettu väestölaskennan laatutukimus lukuunottamatta AET 75:stä tehtyä peittävyysselvitystä. Muissa maissa veästölaskentojen laatutukimuksia on suoritettu jo paljon aikaisemmin, esimerkiksi Ruotsissa vuodesta 1960, Norjassa vuodesta 1975 ja USA:ssa vuodesta 1950. Väestölaskentojen luotettavuuden mittaamisen menetelmät on kehitetty lähinnä USA:ssa. Sekä Ruotsissa, Norjassa että USA:ssa ja nyt myös Suomessa tehdyt väestölaskentojen laatututkimukset perustuvat Hansen-Hurwitzin keskineliöpoikkeamamalliin. Kaikissa näissä maissa tutkimukset on suoritettu periaatteessa samalla tavalla.

Ruotsin ja Norjan tutkimuksissa on keskitty taloudellisen toiminnan tietojen tutkimiseen. Vuonna 1970 tutkittiin lisäksi koulutustietoja. Ruotsissa on tutkittu myös jonkin talotyypin asuntotietoja. Otoskoot ovat Ruotsissa olleet yli 10 000. Ruotsissa on tästä syystä voitu tulostaa myös alueellisia tulosteita eräänlaisella suuraluejaolla.

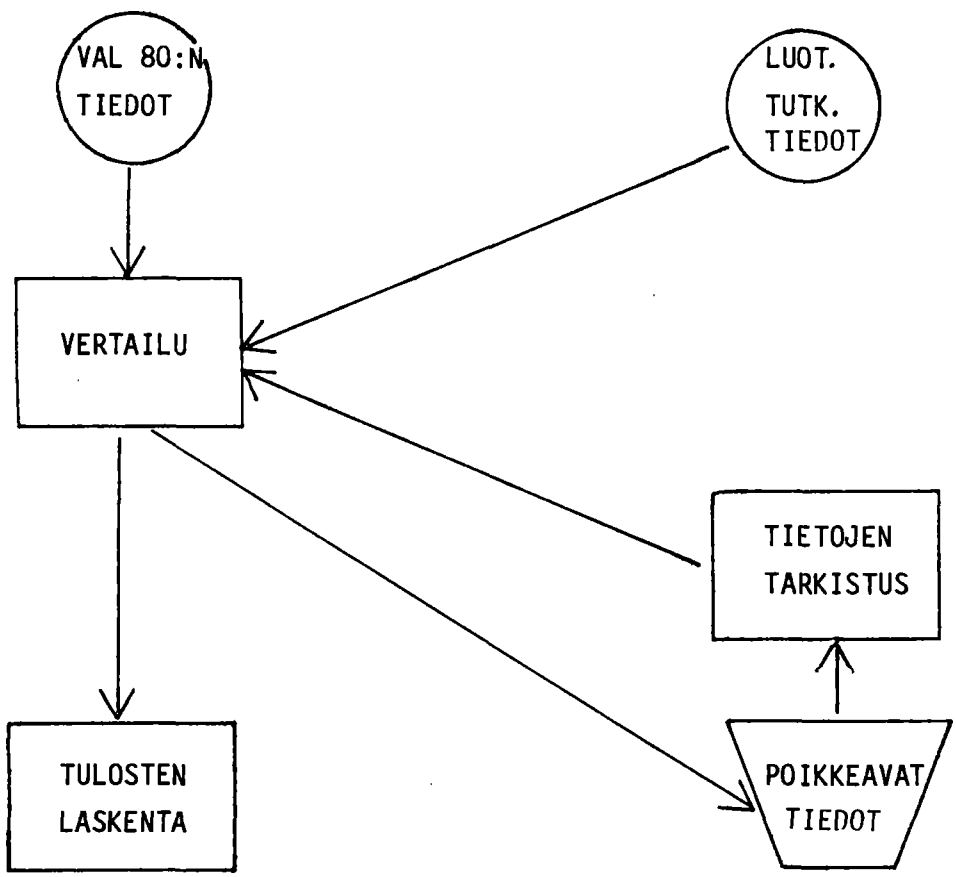
VAL 80:n laatutukimuksessa on tutkittu laadun komponenteista mittausvirheitä, käsittelyvirheitä ja peittovirheitä.

Mittaus- ja käsittelyvirheitä on tutkittu keräämällä tiedot uudestaan haastattelemalla, suorittamalla lomakkeilla uudelleen kooditus sekä määrittämällä rakennuksille uudet koordinaatit. Tavoitteena on ollut määrittää jokaiselle tutkittavan yksikön tiedolle oikeat arvot. Näitä oikeaksi katsottuja tietoja on verrattu VAL 80:n lopullisiin tietoihin ja laskettu virheellisyyttä kuvaavat luvut.

Haastattelututkimuksia suoritettiin kaksi: henkilö- ja asuntotietoja koskenut haastattelututkimus sekä rakennus- ja toimitilatietoja kos-

kenut haastattelututkimus. Henkilötietoja koskeneen haastattelututkimuksen otoskoko oli n. 4000, rakennustietoja koskeneen haastattelututkimuksen otoskoko oli n. 2000. Haastattelututkimusten työnkulku oli seuraava:

Kuvio 3. Haastattelututkimuksen työnkulku



VAL 80:n tiedot kerättiin ensin haastattelemalla. Haastattelutietojen oikeellisuuden parantamiseksi haastattelijat koulutettiin tehtäväänsä ja haastatteluohjeet laaditiin mahdollisimman tarkoiksi. Saaduille tiedoille suoritettiin validisuus- ja loogisuustarkistuksia. Tietoja verrattiin VAL 80:n tietoihin. Poikkeavat tiedot tulostettiin virhelistoille, joilla olleet haastattelutiedot tarkistettiin sekä koodi-

tusvirheiden että tietojen antajien aiheuttamien virheiden korjaamiseksi suorittamalla mm. puhelintiedusteluja. Tiedot pyrittiin tällä tavoin saamaan niin oikeiksi kuin inhimillisesti katsoen oli mahdollista. Tarkistettuja tietoja verrattiin VAL 80:n lopullisiin tietoihin ja laskettiin virheellisyyttä kuvaavat luvut: nettovirhe, bruttovirhe ja oikein luokiteltujen määrä.

Nettovirheellä tarkoitetaan tietyn muuttujaluokan VAL 80:n määrän ja luotettavuustutkimuksen määrän välistä eroa, joka voi olla joko positiivinen tai negatiivinen. Nettovirhe kuvaa luokkien kokonaismäärien oikeellisuutta VAL 80:ssä. Se kertoo tilaston käyttäjälle kuinka oikea tiettyyn luokkaan kuuluvien yksiköiden määrä on.

Oikein luokiteltujen määrällä ja osuudella tarkoitetaan VAL 80:ssä ja luotettavuustutkimuksessa samalla tavalla luokiteltujen määrää ja osuutta VAL 80:n tiedoista. Oikein luokiteltujen määrä ja osuus kertoo kuinka suuri osa yksikkötason tiedoista on oikein VAL 80:ssä. Lisäksi oikein luokiteltujen määrä ja osuus kertovat tilastojen tuottajille kuinka hyvin työssä on onnistuttu.

Muuttujaluokan bruttovirheellä tarkoitetaan luokan suhteen virheellisesti luokiteltujen tapausten kokonaismäärää ts. virheellisesti luokasta poisjääneiden ja virheellisesti luokkaan mukaan laskettujen summaa.

Bruttovirhe ja oikein luokiteltujen määrä kuvaavat yksikkökohtaisten tietojen oikeellisuutta, nettovirhe sen sijaan tietyn luokan kokonaismäärän oikeellisuutta.

Taulukko 1. Esimerkki virheitä kuvaavien lukujen laskemisesta

Luotettavuus-tutkimuksen määrät	VAL 80:n määrät 1	2	Yhteensä
1	20	5	25
2	10	25	35
Yhteensä	30	30	60
Nettovirhe	5	-5	
Oikein luokiteltuja	20	25	45
Bruttovirhe	15	15	30

Mikä näistä luvuista on tilaston käyttäjille käyttökelpoisin laadun mitta riippuu siitä, miten hän VAL:n tietoja käyttää. Jos käytetään ainoastaan tiettyyn muuttujaluokkaan kuuluvien tapausten yhteismäärää, nettovirhe kertoo tämän tiedon luotettavuuden. Jos taasen käytetään jotakin osapopulaatiota koskevia tietoja esim. alueellisia tietoja, nettovirhe ei ole riittävä luotettavuuden mitta. Nettovirheitä laskettaessa virheelliset tapaukset voivat kumota toisiaan siten, että nettovirhe on pieni, vaikka virheellisten tapausten määrä on suuri. Tästä syystä nettovirhe voi vaihdella eri osajoukoissa. VAL:n alueittaisten tietojen luotettavuuden arvioimisessa käyttökelpoisempia laadun mittareita ovat oikein luokiteltujen määrä ja osuus sekä bruttovirhe. Jos esim. oikein luokiteltujen osuus on 80 %, joka 5:s yksikkö on luokiteltu ko. muuttujan tai muuttujaluokan suhteen väärin. Tällöin osapopulaatiotuloksiin on suhtauduttava varauksellisesti. Väärin luokitellut tapaukset voivat koskea juuri kyseessä olevaan osapopulaatioon sisältyviä tapauksia. Oikein luokiteltujen osuus kertoo lisäksi missä muuttujan luokissa tulosten luotettavuus on erityisen varmaa tai epävarmaa.

Peittävyyttä on VAL 80:n osalta tutkittu samaan tapaan kuin AET 75:n peittävyysselvityksessä. VAL 80:n perusjoukkoja on verrattu muihin tietoaineistoihin. Henkilöiden peittävyyttä on tutkittu selvittämällä ns. vastaamatta jättäneiden rakennetta. VAL 80:n henkilöiden perusjoukko saatiin VRK:sta, joten varsinaista ali- tai ylipeittoa ei henkilöiden osalta VAL 80:ssä ole verrattuna VRK:n väkilukuun.

Seuraavaksi esittelen joitakin lukuja VAL 80:n tietojen luotettavuudesta.

Ammatissa toimivien määrä on yksi tärkeimmistä väestölaskennoista saatavista tiedoista. Sen nettovirhe oli n. -5 %, ts. ammatissa toimivien määrä oli VAL 80:ssä 5 % liian pieni. Tästä määrästä lähes puolet oli tapauksia, joissa pääasiallinen toiminta oli VAL 80:ssä tuntematon. Tietojen antajien aiheuttamia ja käsittelyvirheitä oli tämän tiedon osalta siten hyvin vähän.

Pääasialliselta toiminnaltaan tuntemattomia oli VAL 80:ssä yhteensä lähes 90 000, mikä on n. 2 % 15 vuotta täyttäneestä väestöstä. Näistä n. 40 000 oli henkilöitä, joista ei saatu lomakkeilla tietoja ja jotka tästä syystä lisättiin VAL 80:n henkilöiden perusjoukkoon VRK:n henkilörekisteristä. Näiden ns. vastaamatta jättäneiden määrää olisi voitu jonkin verran pienentää lisäämällä lomakkeiden karhuntaa. Tämä olisi vaatinut lisää resursseja eikä tästä syystä ollut VAL 80:ssä mahdollista. Vastaamatta jättäneitä oli VAL 80:ssä puolet vähemmän kuin AET 75:ssä, jossa vastaamatta jättäneitä yli 15-vuotiaita henkilöitä oli n. 95 000. Vastaamatta jättäneiden rakenne oli VAL 80:ssä ja AET 75:ssä samanlainen. Molemmissa laskennoissa muita useammin vastaamatta jättivät 20-34-vuotiaat miehet. Ilmeisesti tällaisia henkilöitä on vaikeaa tavoittaa lomakelaskennoissa.

Kuten taulukosta 2 ilmenee, ammatissa toimivien nettovirheet ovat olleet Ruotsissa ja Norjassa lähes yhtä suuret kuin VAL 80:ssä. Ilmeisesti lomakelaskennoissa päädytään tämän tasoiseen tulokseen.

Taulukko 2. Ammatissa toimivien suhteelliset nettovirheet (%:na väestölaskentojen määristä)

		%
VAL 80		- 5,1
Norja	1970	- 5,8
Ruotsi:	1970	- 4,6
	1975	- 4,1

Olen valinnut ammatissa toimivuuden suhteellisen nettovirheen esimerkiksi paitsi siitä syystä, että käsite on keskeinen väestölaskennassa myös siitä syystä, että tulos 5 % vastaa suhteellisten nettovirheiden keskitasoa. Yli 10 %:n nettovirheet olivat VAL 80:ssä harvinaisia. Esimerkiksi luokan asuinrakennukset suhteellinen nettovirhe oli alle 1 %.

Taulukossa 3 on esitetty ammattiluokkien maatalous- ja teollinen työ suhteelliset nettovirheet luokituksen 1-numerotasolla. Maataloustyön suuri suhteellinen nettovirhe ei aiheudu niinkään siitä, että henkilöt olisi luokiteltu väärään ammattiluokkaan, vaan siitä, että maataloustyötä tehneet eivät ilmoittaneet VAL 80:n lomakkeella tehneensä maataloustöitä. Luokan nettovirheestä yli puolet muodostui tapauksista, joissa henkilö ei ollut VAL 80:ssä ammatissa toimiva. Maataloustyötä tehneet jättivät suhteellisesti muita useammin ilmoittamatta työkuukautensa VAL 80:ssä. Osaltaan tämä varmasti aiheutuu siitä, että käsite maatalouden normaalityöajasta on epäselvä ja siitä, että vastaajille on epäselvää mikä työ katsotaan maataloustyöksi. Tulosta voitaisiin varmasti parantaa pelkästään käsitteitä selventämällä ja saattamalla ne yleiseen tietoon. Samanlaiseen tulokseen on päädytty myös Ruotsissa ja Norjassa.

Taulukko 3. Maatalous- ja teollista työtä tehneiden suhtteelliset nettovirheet (%:na väestölaskentojen määristä)

		Maataloustyö %	Teollinen työ %
VAL 80		- 14,2	- 2,7
Norja	1970	- 18,3	- 0,9
Ruotsi:	1970	- 13,0	- 3,0
	1970	- 15,9	- 1,5

Oikein luokiteltujen osuudet vaihtelivat VAL 80:ssä tiedoittain melko paljon. Suurimmat oikein luokiteltujen osuudet olivat suppeita luokituksia koskevissa tiedoissa kuten rakennusten käyttö (96 %) ja asunnon hallintamuoto (94 %). Pienimmät oikein luokiteltujen osuudet ts, eniten virheitä oli tarkkaa vastausta vaatineissa tiedoissa kuten rakentamivuosi (80 %) ja huoneitten lukumäärä (82 %).

Taulukossa 4 on esitetty ammatin ja elinkeinon mukaan oikein luokiteltujen osuudet luokitusten 1-numerotasolla VAL 80:ssä ja Ruotsin ja Norjan väestölaskennoissa. Tulokset ovat jälleen lähes samat. Menetelmiä kehittämällä voidaan laskentojen tulosta parantaa. Ruotsin vuoden 1975 väestölaskennassa kooditus suoritettiin näyttöpäätteelle, mikä vähensi huomattavasti kooditusvirheitä. Se näkyy myös taulukossa esitetyissä tuloksissa, tosin ei niin selvästi kuin kooditusvirheiden määriä verrattaessa.

Taulukko 4. Ammatin ja elinkeinon mukaan oikein luokiteltujen osuudet (%:na väestölaskentojen ammatissa toimivien määristä)

		Ammatti %	Elinkeino %
VAL 80		90,7	90,0
Norja	1970	84,6	84,5
Ruotsi:	1970	89,7	91,6
	1975	92,0	94,4

Tietojen virhelähteet voidaan jakaa kahteen ryhmään: tietojen antajien aiheuttamiin virheisiin ja käsittelyvirheisiin.

Tietojen antajien aiheuttamia virheitä on VAL 80:ssä sekä saatujen vastausten perusteella koditetuissa tiedoissa että vastaajan suoraan lomakkeelle merkitsemissä tiedoissa.

Koditettaviin tietoihin vastaajat aiheuttivat virheitä ennen kaikkea liian epätarkoilla vastauksillaan. Varsinkin ammattitiedon koditusta vaikeuttivat VAL 80:ssä liian epätarkat vastaukset lomakkeilla. Henkilöille ei voitu määrittää luokituksen kriteerien mukaista oikeaa ammattiluokkaa kun vastauksena oli esim. merkonomi, ekonomi tai työnjohtaja. Liian epätarkkoja vastauksia ei varmaankaan ole ainoastaan VAL 80:n lomakkeilla, vaan kaikissa kyselylomakkeilla kerätyissä aineistoissa.

Eniten vastaajan aiheuttamia virheitä oli kuitenkin suoraan lomakkeelle merkityissä tarkkaa vastausta vaatineissa kysymyksissä. Tällaisia olivat kaikki tarkkaa määrää ja aikaa kuten pinta-alaa ja työssäolokuukausia koskeneet kysymykset. Usein vastaukset olivat noin vastauksia, esim. pinta-ala pyöristettynä lähempään kymmeneen neliömetriin. Kun esim. henkilön pääasiallinen toiminta pääteltiin työssäolo- ja työttömyyskuukausien perusteella, saattoi henkilö sijoittua väärään luokkaan pelkästään yhden kuukauden virheen takia. Oikein luokiteltujen osuudet olivat kaikista tutkituista tiedoista kaikkein pienimmät juuri tarkkaa vastausta vaatineissa tiedoissa.

VAL 80:ssä koditus suoritettiin hajautetusti kuntiin perustetuissa paikallistoimistoissa. Paikallistoimistojen esimiehille annettiin kahden päivän koulutus tehtäviinsä. Esimiesten tehtävänä oli kouluttaa alaisensa.

Saatujen tulosten mukaan suppeitten luokitusten käyttö onnistui paikallistoimistoissa hyvin. Esim. rakennuksen pääasiallisen käytön mukaan virheellisesti koditettujen osuus oli ainoastaan 4 %. Sen

sijaan laajojen luokitusten kuten ammatti- ja toimialaluokituksen käyttö vaatii huomattavaa ammattitaitoa ja kokemusta sekä riittävää koulutusta. Ammattitiedoissa oli koditusvirheitä 3-numerotasolla 14 %.

VAL 80:ssä käytettiin työpaikkatietojen koodituksessa apuna toimipaikkasysteemiä. Henkilölomakkeelle kooditettiin pääsääntöisesti työpaikkaa vastaava toimipaikkatunnus, jolloin työpaikkaa koskevat tiedot saatiin toimipaikkasysteemistä. Toimipaikkatunnuksissa oli koditusvirheitä ainoastaan 7 %, kun suoraan lomakkeille koodítetuissa toimialatiedoissa virheitä oli 11 %. Toimipaikkasysteemin kaltaiset aputiedostot vähentävät saatujen tulosten mukaan huomattavasti koditusvirheitä.

Luotettavuustutkimuksen yhteydessä tuli ilmi hyvin erilaisia ja eri tasoisia tietoja väestölaskennan tietojen oikeellisuudesta ja niihin vaikuttaneista tekijöistä. Esimerkkinä tutkimuksessa ilmi tulleista laatuun vaikuttaneista tekijöistä voidaan mainita kysymysten muotoilu, joka on tärkeä saatujen vastausten oikeellisuuteen vaikuttava tekijä.

VAL 80:n kysymysten muotoilussa havaittiin kolmen laatuisia puutteellisuuksia.

Ensinnäkin samassa kysmyksessä ei pitäisi kysyä useampia asioita. VAL 80:ssä kysyttiin rakentamisvuotta ja perusparannusvuotta samalla kysymyksellä. Kysymys oli muotoa: Mikä on rakennuksen rakentamisvuosi? Ohjeissa kysymystä täsmennettiin siten, että jos rakennuksessa oli suoritettu uudesti rakentamiseen verrattavissa olevia perusparannustöitä, pyydettiin ilmoittamaan perusparannusvuosi. Uudesti rakentamiseen verrattavissa olevia perusparannustöitä ei määritelty lomakkeella. Tuloksena oli, että 20 % rakentamisvuositiedoista poikkesi luokitellusta aineistossa luotettavuustutkimuksen tiedoista. Kaikista verratuista rakentamisvuositiedoista ainoastaan 2/3 oli täsmälleen samoja. Suuri osa poikkeavista tapauksista koski rakennuksia, joissa

joko VAL 80:ssä oli ilmoitettu alkuperäinen rakentamisvuosi eikä perusparannusvuotta, tai rakentamisvuotena oli ilmoitettu jokin korjausvuosi, vaikka korjaus ei ollut rinnastettavissa uudesti rakentamiseen. Saadut vastaukset eivät olleet mitenkään oikaistavissa eikä niitä voitu mitenkään tarkistaa. Kysymys sotki vastaajia.

Jos kysymys sisältää epämääräisiä käsitteitä, ne pitäisi ohjeissa selkeästi määritellä. VAL 80:n rakennuslomakkeella kysyttiin esim. mikä on rakennuksen pääasiallinen lämmitystapa, mutta ei määritelty millä perusteella pääasiallinen lämmitystapa valitaan. Onko esim. vanha omakotitalo, joka on alunperin rakennettu uunilämmitteiseksi, mutta myöhemmin varustettu myös sähkölämmityksellä, pääasialliselta lämmitystavaltaan uunilämmitteinen? Molempia lämmitystapoja voidaan käyttää samanaikaisesti. Ratkaisevatko käyttöaika, rakenteelliset seikat vai käyttökustannukset? Samoin kysymykseen onko osa-aikatyö vähintään puolet alan normaalityöajasta, voi joillakin aloilla olla vaikea vastata. Mikä on esim. maatalouden normaalityöaika? Syy siihen, ettei maatalouden normaalia työaikaa määritelty VAL:n lomakkeella, ei tietenkään ole lomakkeiden suunnittelijoiden, käsitteeltä puuttuvat tarkat määrittelyt.

Myös pelkästään vastausvaihtoehtojen asettelu lomakkeelle saattaa aiheuttaa virheitä. Jostain syystä kyllä-ei vastausvaihtoehtoihin vastaajat merkitsevät mielellään ensimmäisenä olevan vastausvaihtoehdon. VAL 80:n lomakkeilla ei vastaus oli ensimmäisenä. Tästä syystä esim. kysymykseen onko rakennus liitetty sähköverkkoon vastaajat olivat joissakin tapauksissa merkinneet ei ole, vaikka oli ilmeistä että rakennus oli liitetty sähköverkkoon. Vastauksia jouduttiin korjaamaan koneellisesti jälkeenpäin. Samoin kysymykseen oliko osa-aikatyö vähintään puolet alan normaalityöajasta, oli joissakin tapauksissa vastattu ensimmäisen vaihtoehdon mukaisesti ei, vaikka haastatteluissa saatujen tietojen mukaan henkilö työskenteli selvästi vähintään puolet alan normaalityöajasta.

Laatututkimuksen suorittamisella samanaikaisesti tilaston laadinnan kanssa on myös välittömiä vaikutuksia tilaston laatuun. Esim. tiedostot tarkistetaan vertailuajoissa jo ennen tilaston varsinaista tulostusta. Tässä säästetään ennen kaikkea aikaa. Samoin tilaston perusjoukot tulee tarkistettua tutkimuksen suorittamisen yhteydessä. Esim. VAL 80:n henkilöiden perusjoukko tuli tarkistetuksi hyvin varhaisessa vaiheessa. Tällöin havaittiin, että VRK:sta saatu henkilöiden perusjoukko ei vastannutkaan 1.11.1980 tilannetta vaan siitä puuttuivat 1.11-31.12.80 välisenä aikana kuolleet henkilöt. Perusjoukko saatiin korjatuksi ennenkuin tilastojen tulostaminen oli aloitettu. Kaikki tiedostoissa olleet virheet eivät näin pienellä otoksella tietenkään tule ilmi, ainoastaan yleisimmät virheet.

Laatututkimus vaikuttaa varmasti myös tilastoa laativaan henkilöstöön. Kun työntekijät tietävät että heidän työnsä laatuun kiinnitetään huomiota ja sitä tutkitaan, seurauksena on työn laadun paraneminen.

Laatutukimuksessa suunniteltaessa otoskoot olisi ratkaistava suhteessa tilaston tietojen käyttöön. Nyt suoritetussa tutkimuksessa otoskoot olivat niin pienet, ettei minkäänlainen alueellinen tulostus ollut mahdollista. Kuitenkin väestölaskennan tietoja käytetään ennen kaikkea alueittain, tärkeimpiä käyttäjiä ovat mm. kunnat. Väestölaskennan suorittamista lomakelaskentana, totaalitutkimuksena, perustellaan juuri alueellisten tietojen tarpeella. Luotettavuustutkimuksen tietojen perusteella ei alueellisten tulosten luotettavuudesta voida päätellä muuta kuin että alueellisia eroja ilmeisesti on.

Koska VAL 80:n luotettavuustutkimus oli ensimmäinen laatuaan Tilastokeskuksessa, haluttiin tutkia mahdollisimman monen tiedon luotettavuutta. Laatutukimuksen suorittamisen kannalta olisi kuitenkin parempi pitäytyä vain kaikkein keskeisimpien tietojen tutkimiseen, esim. 4-5 tietoon. Näin saadaan haastatteluissa kerättyjen tietojen oikeellisuutta parannetuksi. Haastattelutilanne on erilainen silloin, kun kysyttävänä on 4 kysymystä eikä 40. Saatujen vastausten tarkkuus lisääntyy kysymysten vähentyessä. Myös tietojen tarkistus on huomat-

tavasti helpompaa, kun voidaan keskittyä vain muutamaan oleelliseen tietoon. Samalla tietojen tarkistajat ja koodittajat voivat keskittyä yhden tiedon koodittamiseen ja tarkistamiseen, jolloin heidän työnsä taso paranee.

Tietojen rajaaminen vähentää kustannuksia ja säästää huomattavasti aikaa. Laatututkimuksen suunnittelussa olisi ratkaistava mitkä ovat ne keskeiset tiedot, joiden laatua pitäisi tutkia ja suhteuttaa tutkimuksen otoskoko tilaston käyttäjien tarpeisiin.

ATK-suunnittelussa on tietysti lähdettävä yleissuunnitelmasta kuten muissakin tutkimuksissa. Tämän tyyppiunen tutkimus asettaa tiedostojen rakenteelle kuitenkin erityisvaatimuksia. Tutkimuksen edetessä on voitava nopeasti ja joustavasti tuottaa lisätulosteita. Tutkimuksen tulostusta ei voida suunnitella tarkasti etukäteen. Kun esim. havaitaan jossakin ryhmässä paljon virheitä, nämä tapaukset pitäisi voida tutkia nopeasti edelleen joidenkin lisämuuttujien mukaan.

VAL 80 poikkeaa aikaisemmista laskennoista sikäli, että sen tietojen laadun taso tiedetään. Samoin tiedetään missä oli ongelmia ja missä onnistuttiin. Enää ei tarvitse arvailla miten oikeita VAL 80:n tiedot ovat.

Suunnittelija Pasi Markelin

LAATUONGELMAT REKISTEREITÄ KÄYTETTÄESSÄ

Rekisterillä tarkoitetaan seuraavassa sellaista järjestettyä tietoaineistoa, jonka yksiköt (esim. henkilöt, rakennukset) ovat yksikäsitteisesti tunnistettavissa jonkin tunnuksen (esim. henkilötunnuksen) avulla ja jonka yksiköihin liitettyjä ominaisuustietoja ylläpidetään "jatkuvasti" (vähintään kerran vuodessa). Lisäksi seuraavassa rajoitutaan vain atk-pohjaisiin rekistereihin.

VL 90 -projektin tarkastelutapa rekisterien tilastollisen hyväksikäytön edellytyksistä

VL 90 -projektin työn keskeisenä tavoitteena on ollut luoda edellytykset kokonaan rekisteripohjaiselle tilastotuotannolle tulevaisuudessa väestö- ja asuntolaskentatyyppisten tietojen osalta. Tätä varten projekti on tutkinut mahdollisimman perusteellisesti noin 40 olemassa olevan tai rakenteilla olevan rekisterin tai muun hallinnollisen tietoaineiston hyväksikäyttöedellytyksiä tulevaisuudessa lähinnä vuosia 1985 ja 1990 ajatellen [1]. Tässä työssä käytettiin "kaavaa", jonka mukaan kaikkia rekistereitä tarkasteltiin. Näistä edellytyksistä varsin monet liittyvät päivän teemaan eli tietojen laatuongelmiin, erityisesti 3 ensimmäistä kohtaa. Tilastoneuvoston laatujaoston raportin kannalta voitaisiin sanoa, että ensimmäinen ja kolmas edellytys liittyvät laadun relevanssipuoleen, kun taas toinen kohta liittyy kokonaisvirheeseen eli luotettavuuteen.

---

1) VL 90 -edellytysprojekti: rekisteriraportti, TK Raportti n:o 1982:3, 26.1.1982

Rekistereiden hyväksikäytön edellytykset laskennoissa
(VL 90:n käyttämä tarkastelutapa)

1. Sisällölliset ja käsitteelliset edellytykset
   - kattavuus
   - yksikkökohtaisuus
   - vastaavuus käyttäjien tarpeeseen
   (- YK:n suositukset)
   - viiteajankohta/-jakso

2. "Luotettavuus" (= kokonaisvirhe)
   - peittävyys
   - tietojen luotettavuus (mittaus/käsittelyvirhe)

3. Ajantasaisuus
4. Tekniset edellytykset
   - kuvaukset
   - ATK
   - tunnukset
   - luokiteltavuus

5. Hallinnolliset edellytykset
6. Kustannukset
7. Muut
   - stabiilisuus
   - tietosuoja

(VL 90 - edellytysprojekti: Rekisteriraportti, 26.1.1982)

Yleisesti voi sanoa, että sisällöllisten ja käsitteellisten edellytysten selvittäminen oli varsin työläs tehtävä. "Kattavuudella" tarkoitetaan sitä, että rekisterin yksiköiden tulee kattaa koko maa, väestö, tms. totaalinen perusjoukko, koska väestölaskennoissa ei juuri voida käyttää jotain osapopulaatiota - esim. yhtä kuntaa - kuvaavia rekistereitä.

"Yksikkökohtaisuudella" tarkoitetaan lyhyesti sitä, että rekisterin perustietojen on oltava yksikkökohtaisia, muuten ei tietojen yhdistely laskennoissa yleensä onnistu, verrattuna agregaattitason rekistereihin (esim. tilastotietokannat).

Rekisterin tietojen tulee luonnollisesti vastata käyttäjien tarvetta. Rekisterin tietojen tulisi vastata yleisesti käytettyjä standardeja ja luokituksia. Lisäksi edellytyksenä pidettiin, että rekisterin tiedot ovat ajallisesti vertailukelpoisia aikaisempiin laskentoihin verrattuna. Tämä on varsin selvästi käyttäjien ilmaisema keskeien vaatimus. Voitaneen sanoa, että YK:n suositukset heijastavat melko pitkälle käyttäjien tarpeita ja kansainvälistä vertailukelpoisuutta, jonka vuoksi niitäkin verrattiin rekistereiden tietosisältöön.

Laskentojen kannalta keskeisenä vaatimuksena on, että tuotettavat tiedot kuvaavat samaa ajanjaksoa tai ajankohtaa. Yleisesti ottaen tämä edellytys ei tuottanut suurempia vaikeuksia, mutta eräänä ongelmallisena esimerkkinä voidaan mainita asuntohallituksen asumistukirekisteri, jonka eri tiedoilla on erilainen viiteajankohta: eri tietoja pidetään eri lailla ajan tasalla, osa tiedoista päivitetään vain kerran vuodessa, kun taas eräitä tietoja päivitetään jopa kaksi kertaa kuukaudessa.

Rekistereiden luotettavuuden eli kokonaisvirheen selvittäminen sen sijaan oli todella hankalaa, koska **käytännöllisesti katsoen yhdestäkään valtakunnallisesta hallinnollisesta aineistosta ei ole tehty systemaattisia laatuselosteita tai luotettavuustutkimuksia;** pari poikkeusta kuitenkin on, näitä tarkastellaan jäljempänä. Luotettavuuden tarkastelu jaettiin kahteen komponentiin: peittävyyteen ja itse tietojen luotettavuuteen eli mittaus- ja käsittelyvirheisiin. Parhaimmillaankin arviot rekisterien luotettavuudesta olivat sentapaisia kuin "rekisteristä saattaa puuttua alle 5 % yksiköitä" tai "Mitä tulee Kelan rekisteritietojen luotettavuuteen voidaan sääntönä pitää, että henkilö- ja etuustiedot ovat oikeita lukuunottamatta puhtaasti tilastotarkoituksiin kerättyjä tietoja, joissa saattaa esiintyä virheitä".

Ajantasaisuuden suhteen asetettiin yleinen vaatimus, että rekisterin tulee olla Tilastokeskuksen käytettävissä viimeistään vuoden kuluttua viiteajankohdasta. Esimerkkeinä "hitaimmasta päästä" voidaan mainita Verohallituksen verotusrekisteri ja maatilahallituksen maatilarekisteri, jotka molemmat valmistuvat vasta noin vuoden kuluttua viiteajankohdastaan.

Teknisiä edellytyksiä ei tässä lähdetä laajemmin tarkastelemaan.[1)]

"Luokiteltavuus"-edellytyksellä tarkoitetaan sitä, että rekisterin tietojen tulisi olla valmiiksi luokiteltuja (ts. kooditettuja), jotta tietoja voidaan tietokoneella helposti käsitellä. Konkreettisena esimerkkinä tästä voidaan mainita väestön keskusrekisterin ammattitieto, joka on rekisterissä selväkielinen tieto, eikä sitä näin ollen ole voitu käyttää tilastotarkoituksiin. (Tämä ongelma pyritään kuitenkin ratkaisemaan tulevaisuudessa ns. automaattisen kooditusjärjestelmän avulla.)

"Hallinnollisilla edellytyksillä" tarkoitetaan yksinkertaisesti sitä, että aineisto tulee voida luovuttaa Tilastokeskuksen käyttöön yksikkötasolla. Suurimpana "kantona kaskessa" on tässä ollut Eläketurvakeskuksen laajat aineistot, joita Tilastokeskus ei toistaiseksi ole voinut käyttää tilastollisiin tarkoituksiin, koska virasto ei ole lain tarkoittama valtion virasto (joilta Tilastokeskuksella on oikeus saada yksikkökohtaisia tietoja).

"Kustannus"-edellytyksellä tarkoitetaan sitä, että rekisterin käytön kustannukset eivät saa olla valtiolle suuremmat kuin jonkin muun vaihtoehtoisen tietojenkeruuvaihtoehdon käyttäminen. Yleisesti ottaen rekisterien käyttö verrattuna esim. postitiedusteluun onkin ratkaisevasti halvempaa ainakin Tilastokeskukselle.

---

1) Tunnus-edellytyksestä mainittakoon, että VAL 80:ssä suunniteltiin maatilahallituksen (MTH) maatilarekisterin hyväksikäyttöä ennen vuotta 1980, saimme kuulla, että rekisteristä puuttuu tunnukset - viljelijän henkilötunnukset - noin 10 %:lta perusjoukosta. Tämän pohjalta TK esittikin neuvotteluissa vaatimuksen rekisterin hyväksikäytölle VAL 80:ssä, että yksiköiden tunnukset tulisi täydentää rekisteriin ennen laskentahetkeä. MTH ryhtyikin täydentämään rekisteriänsä ja laskentahetkellä henkilötunnuksia puuttui enää 1-2 %. Näin tuon rekisterin laatutasoa saatiin tilastolliselta kannalta parannetuksi oleellisesti.

Sen sijaan rekisterin "stabiilisuus" ajassa saattaa useinkin olla ongelmallinen, kun järjestemään voidaan tehdä jostakin hallinnollisesta syystä muutoksia esim. vuosittain, puhumattakaan siitä mahdollisuudesta, että koko rekisteri lakkautetaan jonkin viranomaisen päätöksen perusteella.

Rekisteritietojen relevanssi

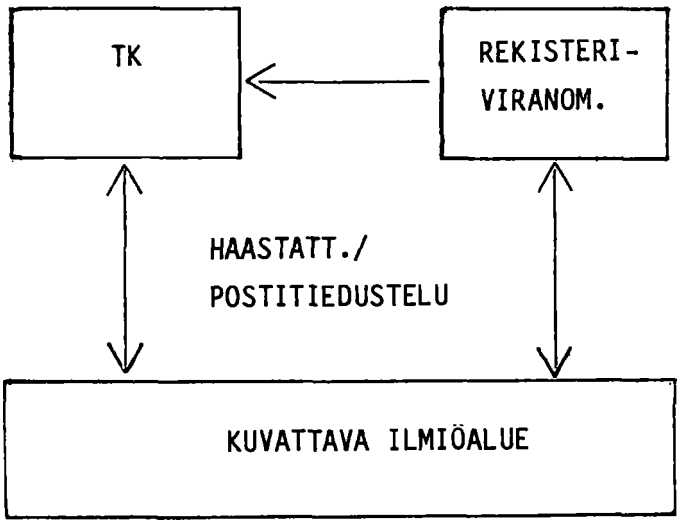
Yleisesti ottaen voidaan todeta, että rekisteritietojen relevanssi on tilastollisen käytön kannalta hankala asia. Usein rekisterit on rakennettu "räätälin työnä" johonkin erityiseen hallinnolliseen tarkoitukseen, eikä yleisiä tilastollisia vaatimuksia ole otettu huomioon. Luonnollisesti eräitä poikkeuksiakin on. **Tulevaisuutta ajatellen saattaa hyvinkin käydä niin, että kun hallinnollisten tietoaineistojen hyväksikäyttö lisääntyy hyvää vauhtia, niin ao. tilastojen relevanssi suhteessa kuvattavaan realimaailmaan heikkenee nykyisestään.** Tilastoilla ei ehkä aina kuvata todellisuutta - niinkuin luonnollisesti pitäisi - vaan rekisterien omaa suljettua maailmaa. Myös kansainvälisesti tällainen kehitys saattaa olla todella dramaattista, kun kansainvälinen vertailukelpoisuus tilastoissa väistämättä näyttää heikkenevän rekisterien käytön myötä.

Relevanssin eräs osatekijä on "ajantasaisuus" tai tässä tapauksessa rekisterien käytön nopeus. Olen poiminut seuraavat luvut Olli Janhusen tekemästä tuotantomenetelmien nykyjärjestelmäselvityksestä [1]. Siinä selvitettiin Tilastokeskuksen keskeisten systeemien ajoitusta tietojenkeruun, tiedostojen luonnin ja tuotteiden valmistumisen osalta. Systeemit on taulukossa jaettu kolmeen ryhmään tietojen keruutyypin perusteella. Rekistereiden hyväksikäyttöä kuvaa lähinnä ensimmäinen rivi/tietojenkeruutyyppi: "koneellinen".

Taulukko 5. Rekisterien käytön nopeus
(Nykyjärjestelmäselvitys)

	"Viive" viikossa			
Tietojen keruutyyppi	Keruu	Tiedostot	1. tuote	Viim.tuote
Koneellinen	31	41	54	59
Postitiedustelu	12	19	23	44
Haastatt.	9	20	53	57

1) Nykyjärjestelmäselvitys, yleiskuvaus, TK Raportti n:o 1983:5, Olli Janhunen, 28.3.1983



Taulukosta nähdään, että **koneellinen tietojenkeruu on tarkasteltavissa systeemeissä ajoitukseltaan hitain verrattuna muihin tietojen keruutyyppeihin.** Tähän on ainakin kaksi syytä: ko. koneellista tietojen keruuta käyttävät systeemit ovat yksikkömäärältään suuria systeemeitä ja toisaalta se, että Tilastokeskus joutuu odottamaan ao. viranomaisten pitämien rekisterien valmistumista ennen kuin se voi aloittaa oman tuotantoprosessinsa. Lisäksi on muistettava, että rekisterien hyväksikäytölle tilastotuotannossa ei käytännössä ole juuri vaihtoehtoja: ei voida ajatella, että Tilastokeskus ryhtyisi keräämään esim. verotustietoja postitiedustelulla tai haastattelemalla verovelvollisia.

Rekisterien kokonaisvirhe

Valtakunnallisten rekisterien luotettavuudesta on erittäin vähän systemaattista tietoa luotettavuustutkimusten tai laatuselosteiden muodossa. Ainoat tuntemani esimerkit ovat

- TK:n tutkintorekisteri ja
- VRK:n pitämä väestön keskusrekisteri (vain eräät tiedot).

Molemmille on yhteistä se, että luotettavuustutkimukset ovat Tilastokeskuksen suorittamia. Muut tilastovirastot ja rekisteriviranomaiset eivät siis tietääkseni ole omien rekistereidensä luotettavuutta selvittäneet. **TK:n tutkintorekisterin** luotettavuutta - peittovirhettä ja tutkintotietojen luotettavuutta - on tutkittu haastattelemalla sekä vuonna 1977 koskien vuotta 1975 (kaksi eri haastattelua) että VAL 80:n yhteydessä 1980/81. Keskeiset tulokset näkyvät taulukosta 6.

Taulukko 6. Tutkintorekisterin luotettavuus

	v. 1975 %	1980 %
(Brutto)alipeitto	6	2
(Brutto)ylipeitto	0,8	0,6
Virheellisiä	1,3	3,4

VAL 80:ssä TR:ää täydennettiin n. 60 000 tutkinnolla

1.11.1980 TR:stä puuttui n. 78 000 henkilön tutkintotiedot, joka on n. 5 % TR:n perusjoukosta.

Huomataan, että rekisterin peittävyys on parantunut melkoisesti vuodesta 1975: tähän on selityksenä se, että VAL 80:n yhteydessä rekisteriä täydennettiin noin 60 000 tutkinnolla. Sen sijaa yksittäisten tutkintotietojen oikeellisuutta ei ole tiettävästi (jälkikäteen) korjattu, joten virheellisten tietojen osuus näyttää jonkin verran kasvaneen. **Väestön keskusrekisterin äidinkieli-tieto** on ehkä ääripään esimerkki tiedon virheettömyydestä. Tätä selvitettiin VAL 80:n luotettavuustutkimuksessa haastattelemalla n. 3 600 henkilöä. Taulukosta 7 huomataan, että virheiden määrä on äärimmäisen pieni (jos niitä yleensä esiintyy, kun otetaan huomioon otoksesta aiheutuva satunnaisvirhe). Tiedon hyvään tasoon on luonnollinen selitys: se tarkistetaan joka vuosi henkikirjoituksen yhteydessä.

Taulukko 7. VKR:n äidinkieli (VAL 80:ssä)

	VAL 80 (VKR:stä)	Luotettavuus-tutkimus
Suomi	94,0	93,9
Ruotsi	5,9	6,0
Muu	0,1	0,1
Yht.	100,0	100,0

(Haastateltuja 3 620)

Kolmantena esimerkkinä on **väestön keskusrekisteriin (VKR) sisältyvä tieto rakennusten koordinaateista.** (Karttakoordinaatit kuvaavat rakennusten tarkkaa sijaintia ± 10 metrin tarkkuudella pohjois-etelä ja itä-länsi -suunassa.[1]) VAL 80 -luotettavuustutkimuksessa selvitettiin myös koordinaattitietojen luotettavuutta, koska niiden avulla laaditaan mm. taajama- ja kunnan osa-aluekohtaiset tiedot.

Taulukko 8. Rakennuksen koordinaatit VAL 80:SSÄ (VRK:stä)

- Alipeitto n. 2,1 % 20 000 rakennusta (VAL 80:ssä täydennettiin n. 50 000 rakennuksen koordinaatit

  Kunnittainen vaihtelu 0,0 - 36,4 %, Ahvenanmaalla yli 10 % puuttuu

- Virheelliset koordinaatit n. 19,8 %:lla (yli 30 m:n virhe) 185 000 rakennusta

  - kaupungeissa 13,4 %
  - muissa kunnissa 21,3 %

---

1) Ks. LIITE 1 rakennuksen koordinaattien määrittelystä

Selvityksen perusteella (Taulukko 8) voidaan arvioida, että rekisteristä puuttui 1.11.1980 n. 2,1 %:lta rakennuksista koordinaatit, mikä merkitsee noin 20 000 rakennusta. Kuitenkin kunnittainen vaihtelu oli huomattavan suurta: 0,0 - 36,4 %:iin, Ahvenanmaalla puuttui keskimäärin yli 10 %. Virheellisiä koordinaatteja oli rekisterissä vastaavasti n. 19,8 % (virheellisiksi katsottiin yli 30:n virhe). Tämä merkitsee, että rekisterissä on n. 185 000 virheellistä rakennusten koordinaattia. Kaupunkimuodon mukaan erot ovat melkoisia:

- kaupungeissa n. 13,4 % virheellisiä
- muissa kunnissa n. 21,3 % virheellisiä

Lopuksi esitetään yhteenveto tutkituista rekistereistä ja niiden tietojen virheistä: Taulukko 9. VAL 80 -luotettavuustutkimuksessa selvitettiin keskeisten rakennus-, asunto- ja toimitilatietojen virheitä, jotka olivat syntyneet lomaketietojen keruussa. Nämä tiedot siirrettiin sitten Väestörekisterikeskuksen juuri perustamaan rakennus- ja huoneistorekisteriin. Tulokset kuvaavat siis rekisterin tietojen virheellisyyttä sen perustamishetkellä eli 1.11.1980, sen jälkeen syntyneistä virheitä ei ole ainakaan vielä tehty luotettavuustutkimuksia.

Edellä sanotun perusteella on välttämätöntä, että TK yhdessä keskeisten rekisteriviranomaisten kanssa ryhtyi mahdollisimman nopeasti laatimaan laatuselosteita valtakunnallisten rekistereiden luotettavuudesta.

Taulukko 9. Systemaattisia luotettavuustutkimuksia on tehty seuraavista rekistereistä/tiedoista

	Peittävyys	Mittausvirhe	Käsittelyvirhe
TK/Tutkintorekisteri	x	x	(x)
VRK/Väestön keskusrekisteri			
- Äidinkieli		x	(x)
- Rakennusten koordinaatit		x	(x)
VAL 80 -lomakkeilla kerätyt tiedot 1.11.1980:			
- Asutut asunnot	x		
- Asuttu/tyhjä asunto		x	
- Vuokranantaja		x	x
- Huoneistotyyppi		x	
- Pinta-ala		x	
- Hallintaperuste		x	
- Rakennusten käyttötarkoitus		x	x
- Valmistumisvuosi		x	
- Lämmitysaine ja -tapa		x	
- Kerrosluku		x	
- Kerrosala		x	
- Omistaja			x
- Toimitilojen käyttötarkoitus		x	x
- Pinta-ala		x	
- Pinta-alan käyttö		x	
- Vuokranantaja			x

Yliaktuaari Tuula Rissanen

TYÖVOIMATIEDUSTELUN LUOTETTAVUUDEN TUTKIMINEN

Yleistä

Työvoimatiedustelu oli otokseen pohjautuva postikyselytutkimus vuoden 1982 loppuun asti. Vuoden 1983 alusta alkaen se muutettiin haastattelututkimukseksi. Uutta tutkimusta kutsutaan nimellä työvoimatutkimus. Tässä esityksessä käsitellään postikyselyn luotettavuusongelmia.

Työvoimatiedustelun otokseen on poimittu 36 000 henkilöä vuodessa. Perusjoukkona on 15-74-vuotias väestö. Koska työvomatiedustelussa tutkittavien muuttujien muutosten estimointia pidetään tärkeänä, otos suunniteltiin vähitellen vaihtuvaksi. Otos jaettiin 3 000 henkilön ryhmiin, joille tiedustelu toistettiin kuusi kertaa kolmen kuukauden välein. Tiedustelun toistoa nimetään rotaatioksi ja otosryhmiä tästä johtuen rotaatioryhmiksi. Joka kuukausi otokseen tuli mukaan yksi uusi rotaatioryhmä. Yhden kuukauden otoksessa oli kuusi rotaatioryhmää, jotka ovat olleet tiedustelussa aiemmin 0-5 kertaa. Kolmen peräkkäisen kuukauden otokset ovat erilliset ja peräkkäisten vuosineljännesten otoksissa 5/6 on samoja henkilöitä.

Työvoimatiedustelussa tutkittavia muuttujia ovat mm. työikäisen väestön toiminnan laatu, työvoimaosuus, työttömyysaste ja työpanos. Estimaatit lasketaan ositetun otannan avulla ja kato sijaistetaan oletuksella, että se on samoin jakaantunut kuin vastanneet.

Työvoimatiedustelun luotettavuusongelmat

Työvoimatiedustelun kokonaisvirheen suurimmat osatekijät ovat kato ja mittausvirheet. Tässä esityksessä keskitytään kato-ongelman vaikutusten selvittämiseen. Katoon vaikuttavana tekijänä on tarkasteltu rotaatiota postikyselytutkimuksessa. Muut työvoimatiedustelun virhetekijät ovat helpommin hallittavissa.

Kehikkovirheet ovat suhteellisen pieniä tarkkojen väestörekisterien ansiosta. Satunnaisvaihtelu heikentää erityisesti kuukausitulosten tarkkuutta, mutta pidemmällä ajanjaksolla vaikutus on pienempi. Erityisongelma on ollut työvoimatiedustelujen vastausten kopiointi, jos vastauksia joudutaan kopioimaan, tulisi kopioitavista poistaa satunnaisesti katoa vastaava osuus.

Vastausten laadusta

Työvoimatiedustelun vastausten laatua on seurattu lomakkeilla ilmenevien puutteiden ja ristiriitaisuuksien avulla. Mittausvirhetutkimusta ei ole tehty erillisenä, se olisi vaatinut haastattelumäärärahan. Näin ollen havaitut puutteet kuvastavat vain osaa mittausvirheistä. Keskimäärin 20 % lomakkeista on täytetty puutteellisesti. Merkinnän yhteydessä voidaan täydentää 15 % lomakkeista päättelemällä vastauksia. Puutteellisia vastauksia on työllisillä henkilöillä allaolevia määriä:

Taulukko 10.

	Vastaus puuttuu	Vastaus täydennetty tulkinnanvaraisesti	Tuntematon
	%	%	%
Toiminnan laatu	10,3	1,9	-
Ammatti	6,1	2,7	1,0
Ammattiasema	13,5	4,0	1,2
Ajankäyttö	32,2	2,9	3,0
Työtunnit	6,4	1,5	4,3
Normaali työaika	9,2	5,4	3,5

Alkuperäisessä aineistossa on kohtuuttoman paljon puutteellisia vastauksia, joita joudutaan täydentämään. Osa tiedoista voidaan täydentää lomakkeen muista osista (esim. lisätiedoista), mutta myös tulkinnanvaraisesti täydennettyjä vastauksia on paljon. Tuntemattomien määrä jää näinkin meneteltäessä suureksi.

Kadon vaikutus tuloksiin

Katoprosentti on keskimäärin 30 %. Kato on vino iän, sukupuolen, läänien ja tutkittavien ominaisuuksien suhteen, lisäksi vinous muuttuu vuodenaikojen ja suhdannevaihtelujen mukaan. Rotaatio vaikuttaa kadon määrään siten, että kato kasvaa tiedustelukertojen määrän kasvaessa.

Rotaatioryhmäharha

Kun sama mittaus toistetaan samoille henkilöille kuusi kertaa uusintatiedusteluissa kohdehenkilöiden vastaamiskäyttäytyminen muuttuu. Kadon määrän lisääntyminen uusintatiedustelussa aiheutunee vastaamisväsymyksestä, toiminnan muutosten "salaamisesta" tai toiminnan pysyvyyden "salaamisesta". Tästä johtuen eri rotaatioryhmissä lasketut estimaatit poikkeavat merkittävästi toisistaan. Ilmiötä kutsutaan rotaatioryhmäharhaksi. Rotaation vaikutusta voidaan tutkia laskemalla estimaatit erikseen eri rotaatioryhmissä ja koko otoksessa. Ryhmien välisiä eroja voidaan havainnollistaa ns. rotaatioryhmäindeksillä:

$$\frac{\text{Ryhmän estimaatti}}{\text{Koko otoksen estimaatti}} \cdot 100$$

Tällöin indeksiluku 100 ilmaisee, että ko. ryhmä tuottaa saman estimaatin kuin koko otos, suurempi luku kuin 100 ilmaisee, kuinka paljon enemmän ominaisuutta on ryhmässä kuin koko otoksessa, ja vastaavasti

pienempi luku vähemmän. Luvun tulisi luotettavuuden kannalta olla mahdollisimman sama eri rotaatioryhmissä.

Taulukko 11. Indeksiluvut vuosina 1977-1981

	Tiedustelukerta						
	1	2	3	4	5	6	1-6
Työlliset	104,1	101,2	100,0	98,1	97,2	97,5	100,0
Työttömät	106,1	102,4	101,0	98,3	98,2	95,9	100,0
Työvoiman ulkopuoliset	93,5	98,1	100,0	102,9	104,2	104,1	100,0

Indeksiluvuista nähdään esimerkiksi, että työllisiä on 1. kertaa tiedusteltavassa ryhmässä n. 4,1 % enemmän kuin koko otoksessa. 3. tiedustelun jälkeen indeksien suunta muuttuu.

Taulukko 12. Työllisten rotaatioryhmäindeksi eri vuosineljänneksinä

	Tiedustelukerta						
	1	2	3	4	5	6	1-6
1. nelj.	104,9	100,3	100,8	98,0	97,4	97,7	100,0
2. nelj.	104,7	101,7	99,7	99,4	96,4	97,5	100,0
3. nelj.	106,2	101,4	100,3	97,3	98,3	95,7	100,0
4. nelj.	102,1	103,5	100,6	97,8	97,4	98,3	100,0

3. ja 4. neljänneksen indeksiarvot poikkeavat selvästi keskimääräisistä arvoista, mikä on merkki siitä, että vastauskäyttäytyminen vaikuttaa kausivaihtelutuloksiin.

Vastaamiskäyttäytymisestä

Noin 16 % otoshenkilöistä ei vastaa yhteenkään tiedusteluun, noin 50 % vastaa kaikkiin tiedusteluihin ja loput vain osaan tiedusteluista.

Taulukko 13. Yhden ajankohdan tiedustelun keskimääräinen jakauma vastaamiskertojen lukumäärän mukaan vuosina 1977-1981

	Vastaamiskertojen lukumäärä (%)						
	1	2	3	4	5	6	1-6
Vastanneet	1,7	2,9	4,4	8,4	15,1	67,5	100,0

Kerran vastanneista on tiedustelussa alle 2 % ja kuusi kertaa vastanneista peräti 68 %. Kerran vastanneista 62 % vastaa 1. tiedusteluun ja siirtyy sen jälkeen katoon.

Esimerkkinä eri määrän vastauksia antaneiden henkilöiden ominaisuuksien eroista verrataan 1-2 kertaa vastanneiden ja koko otoksen työvoimaosuuksia (työvoimaosuus = työvoima/työikäinen väestö).

Taulukko 14.

Vastaamiskertoja \ Vuosi	1977	1978	1979	1980	1981
1-2	71,9	72,9	73,8	75,0	71,9
1-6	59,9	60,0	60,3	61,2	62,5

(Huom. yllä esitetyt työvoimaosuudet ovat noin 3 % alhaisemmat kuin julkaisuissa, tämä johtuu siitä, että ylläolevat luvut on laskettu otosfrekvensseistä.)

Vuosina 1977-79 työvoimaosuus kasvoi koko otoksessa 0,4 %, mutta 1-2 kertaa vastanneilla 3 %. Vuosina 1980-81 koko otoksen työvoimaosuus kasvoi 1,3 %, mutta 1-2 kertaa vastanneiden luku väheni 3 %. Otoksen rakenne on siis muuttunut eri vuosina, mikä vaikuttaa muutostietoihin.

Loppupäätelmiä

Rotaation vaikutus postikyselyn vastaamiskäyttäytymiseen on kohtuuttoman suuri. Uusintatiedustelut aiheuttavat jopa kymmenen prosentin eroja verrattuna ensimmäisiin tiedusteluihin. Myös vastausten laatu on heikko postikyselyssä.

Edellä kuvattujen tekijöiden vaikutus kokonaisvirheeseen voidaan arvioida työvoimatiedustelun uudistuksen rinnakkaistulosten perusteella. Haastattelun ja postikyselyn erot näkyvät alla olevista luvuista.

Taulukko 15. Rinnakkaisvaihe, 4. neljännes 1982

Toiminnan laatu	haast. - posti (1 000)	%
Työlliset	104	4,6
Työttömät	-3	-2,0
Työvoiman ulkopuoliset	101	4,2

Työvoiman määrä kasvoi haastattelussa 100 000 henkilöllä. Työttömien määrän laskuun pidetään syynä postitiedustelun mittausvirheitä. Vaikka kadossa oli työttömiä enemmän kuin vastanneiden joukossa, osa työttömistä ei täytä kaikkia työttömän kriteereitä. Tässä tapauksessa mittausvirhe ja kadon aiheuttama harha kumosivat toisiaan osittain.

Verrattaessa rinnakkaisvaiheen harha-arvioita ja rotaatioryhmäindeksejä, todetaan, että työllisten estimaatti 1. tiedustelun ryhmässä on ollut 4,1 % koko otosta suurempi, eli varsin lähellä rinnakkaisvaiheen arviota (+ 4,6 %). Postitiedustelun tulokset olisivat siis olleet huomattavasti luotettavammat, jos tiedustelukertoja olisi ollut vähemmän.

Työvoimatiedustelun aikasarjat korjataan vastaamaan uudistetun tutkimuksen lukuja. Harhojen systemaattisen luonteen vuoksi lukujen tasoa

voidaan korjata rinnakkaistulosten perusteella, mutta muutostietojen parantamiseksi joudutaan harkitsemaan korjausta myös erilliselvitysten ja muista lähteistä saatujen tietojen perusteella.

Yliaktuaari Tuula Koskenkylä

LAATUONGELMISTA YRITYSTILASTOISSA

Yritystilastot

Yritystilastojen tarkoituksena on tuottaa säännöllisin väliajoin luotettavaa ja yksityiskohtaista perustietoa sekä yritysten taloudellisista toiminnoista että yritysten ja muiden talousyksikköjen välisistä suhteista. Yritystilastot perustuvat keskeisiltä osin yritysten laskentatoimesta ennenkaikkea kirjanpidosta ja tilinpäätöksestä saataviin tietoihin. Tärkeimmät muuttujaryhmät, joista halutaan tietoja, ovat tuloslaskelma, tase, vaihto-omaisuuden erittely, käyttöomaisuuden ja muiden pitkävaikutteisten menojen erittely, liikevaihdon tai myyntituottojen erittely sekä toimialakohtaiset erityistiedot.

Yritystilastoissa mielenkiinto on kohdistunut ensisijaisesti näiden ilmiöalueella havaittavien muuttujien kokonaismääräestimaattien tuottamiseen. Yritystilastotiedot kerätään otantatutkimuksen ja osittain myös kokonaistutkimuksen avulla postitse tätä tarkoitusta varten suunnitelluilla tiedustelulomakkeilla. Ajanjakso, jolta tiedot halutaan, on yrityksen tilikausi.

Yritystilastojen perusjoukot

Yritystilastojen kohdeperusjoukko eli se perusjoukko, josta ollaan varsinaisesti kiinnostuneita, on määritelty siten, että sen muodostavat yritykset, joilla on ollut toimintaa tiedustelua koskevana vuonna ja jotka kuuluvat institutionaalisen sektoriluokituksen luokkiin:

1. Yhteisöyritykset ja yhteisömäiset yritykset (lukuunottamatta valtion liikelaitoksia ja kuntien liikelaitoksia) ja

5.1 Elinkeinonharjoittajien kotitaloudet (lukuunottamatta ammatin ja maatilatalouden harjoittajia).

Sektoriluokituksella rajatulta alueelta perusjoukkoon kuuluvat funktionaalisen luokituksen eli toiminnan laadun perusteella toimialaluokituksen toimialat 2, 3, 4, 5, 6, ja osin 7 kolminumerotasolla. Kohdeperusjoukon alkeisyksikköjä ovat näin ollen yritykset, jotka kuuluvat e.m. toimialoille ja institutionaalisella sektoriluokituksella rajatulle alueelle.

Yritystilastojen otosperusjoukko ja otoskehikko on muodostettu useista erillisistä rekistereistä ja muista tietolähteistä. Useimmilla yritystilastotoimialoilla otosperusjoukkona on ollut Tilastokeskuksen ylläpitämä yritysrekisteri.

Ositettu otanta yritystilastoissa

Yritystilastoissa on käytössä kaikilla osajoukoilla ositettu otanta. Perusteluina tälle on se, että eri toimialat eroavat toisistaan rakenteensa, kokonsa ja toimintansa laadun suhteen. Tällöin on tarpeellista käsitellä eri toimialoja tai ainakin toiminna laadun suhteen samanlaisia toimialoja tarvittaessa omana perusjoukkonaan. Koska tarkastellaan yrityksiä ja niiden taloudellisia toimintoja, on näiden toimintojen suuruus tai laajuus markkamääräisesti mitattuna riippuvainen myöskin yrityksen koosta. Yrityksen kokoa kuvaavista muuttujista on osittamismuuttujaksi valittu henkilökunnan lukumäärä. Toisaalta on valittu osittamismuuttujaksi myös toimialaluokitus kolminumerotasolla.

Taulukko 16. Yritystilastotoimialojen osittaminen käyttäen yrityksen koon mittana henkilökunnan lukumäärää

Osite	Henkilökunnan lukumäärä
1	0 - 4
2	5 - 19
3	20 - 49
4	50 - 99
5	100 -

Toimialoilla, joissa ei ole käytettävissä henkilökuntatietoa perusjoukon osalta, on ositerajat arvioitu henkilökuntarajoja vastaaviksi.

Yritystilastotiedustelussa tiedusteltavien yritysten kokonaismäärä on noin 4200 yritystä. Tämä kokonaismäärä muodostuu yrityksistä, jotka poimitaan otannan avulla ja yrityksistä, jotka kuuluvat kokonaistutkimukseen. Otoskokoa määrättäessä jokainen toimiala on periaatteessa oma perusjoukkonsa ts. jokaisen toimialan otoskoko määrätään muista toimialoista riippumattomasti asettamalla tietyt tarkkuusvaatimukset estimoitavalle kokonaismäärätiedolle.

Yritystilastoissa tuotettavien kokonaismäärätietojen estimointi on kaikilla toimialoilla yhdenmukainen. Estimointimenetelmänä on erillinen suhde-estimointi. Erillisessä suhde-estimoinnissa lasketaan ensin jokaisessa ositteessa halutulle muuttujalle kokonaismäärän estimaatti. Koko toimialan tai haluttaessa koko osajoukon esim. teollisuuden kokonaismäärän estimaatti on tämän jälkeen näiden osite-estimaattien summa.

Virhelähteistä otantatutkimuksissa

Edellä esitetyissä yritystilastojen menetelmällisissä tarkasteluissa oletetaan, että on käytetty jonkinlaista todennäköisyysotantaa ja jokaisella havaintoyksiköllä on oikea arvo. Tällöin estimaatin virhe aiheutuu pelkästään satunnaisesta otantavaihtelusta, joka on seurauksena siitä, että koko perusjoukon N:n yksikön sijasta mitataan vain osa perusjoukon yksiköistä eli n yksikköä. Käytännön otantatutkimuksissa, joissa mitataan useita muuttujia ja näihin mittauksiin liittyy usein vaikeita ongelmia, ovat edellä esitetyt olettamukset kaukana todellisuudesta. Satunnaisvaihtelun lisäksi saattaa otantatutkimuksissa olla seuraavia virhelähteitä:

1 Ei onnistuta mittaamaan joitakin otokseen valittuja yksikköjä (kato).

2 Yksiköiden mittauksissa sattuneet mittausvirheet,

3. Virheet, jotka syntyvät tulosten koodauksessa, taulukoinnissa tai julkaisemisessa (käsittelyvirheet) ja

4. Virheet, jotka syntyvät, kun poiketaan todennäköisyysotannasta.

Ruotsin SCB:n laaduntarkastelumalli

Kun tarkastellaan yritystilastojen laatua tai laatuongelmia eivät edellä esitetyt virhelähteet kata vielä kokonaan sitä, vaan käsite laatu sisältää myös tilaston tietyn määritellyn käyttötarpeen tai käytön tarkastelun. Tällöin voidaan nojautua esim. Ruotsin SCB:n laatukäsitteeseen, jossa laatu on hajoitettu osatekijöihin, joiden merkitys on sidoksissa paitsi käyttötarpeeseen myös käytettävissä olevaan tuotantomentelmään ja siihen liittyviin resursseihin.

Laaduntarkastelumallissa laatu on jaettu tilastotiedon relevanssitekijöihin ja toisaalta virhelähteisiin (kokonaisvirhe). Relevanssitekijät voidaan jakaa edelleen tiedon sisältöominaisuuksiin ja oikeaaikaisuuteen. Tietojen kokonaisvirhe voidaan jakaa harhaan ja tarkkuuteen.

Kun puhutaan yritystilastojen laadusta, siihen vaikuttavista tekijöistä, on huomioitava, että tilastossa ei ole tehty minkäänlaisia laatuselvityksiä tai laatututkimuksia, vaan seuraava laatutarkastelu on lähtöisin niistä ongelmista, jotka tilastossamme on koettu vaikeiksi.

Yritystilastojen näkökulmasta keskeisimmät laatuongelmat löytyvät luotettavuustekijöistä, joita seuraavassa tarkastellaan osatekijöittäin. Pohdiskelu relevanssitekijöistä, joita on yritystilastoissa yritetty kartoittaa myöskin tilastojen käyttäjille kohdistetun tarvetiedustelun avulla, jätetään seuraavassa suorittamatta.

Virhelähteet yritystilastoissa

Merkityksellisin näistä kokonaisvirheen komponenteista on peittovirhe. Peittovirhettä syntyy tilanteessa, jossa otoskehikko ei kata yksikäsitteisesti koko tarkasteltavaa ilmiötä (kohdeperusjoukkoa) tutkimusajankohtana. Yritystilastojen otoskehikko on rajoitetumpi kuin kohdeperusjoukko. Otoskehikkona käytetyt reksiterit ja muut tietoaineistot ovat keskimäärin 1 - 4 vuotta aikaisemmalta ajankohdalta kuin kohdeperusjoukko. Yritystilastojen kannalta keskeisin rekisteri yritysrekisteri on saattanut olla pahimmassa tilanteessa 4 vuotta vanha. Tällä aikavälillä perusjoukko on muuttunut. Perusjoukossa ovat jotkut yritykset fuusioituneet, lopettaneet toimintansa tai uudet yritykset ovat aloittaneet toimintansa. Osa perusjoukon yrityksistä on voinut siirtyä myös toimialalta toiselle, kohdeperusjoukon ulkopuolelle tai muuttua merkitsevästi kooltaan (koko mitattuna henkilökunnan määrällä).

Vertaamalla kahden yritysrekisterivuoden yritystietoja saadaan jonkinlainen käsitys, mitä ovat erot, kun esim. vuoden 1981 yritystilastotiedustelu lähtee YR 78:sta poimitulla otoksella.

Peittävyyseroja kohde- ja otosperusjoukon välille aiheutuu lisäksi siitä, että yritysrekisteriin kuuluvat vain yritykset, joilla on liikevaihtoveronalaista toimintaa. Otosperusjoukon ulkopuolelle jäävät tällöin kohdeperuskoukosta ne yritykset, joilla ei ole liikevaihtoveronalaista toimintaa. Tällaista toimintaa on mm. puhdas agentuuritoiminta. Toisaalta kaikki liikevaihtoveronalainen toiminta ei kuulu kohdeperusjoukkoon.

Taulukko 17. Yritystilastojen otoskehikko muodostettuna vuosien 1978 ja 1980 yritysrekistereistä.

Toimiala	Yritysten lukumäärä 1978	1980	Muutos %	Uudet yritykset 1979 - 80
29	23	60	160	21
30	14	13	- 7	-
31	1 135	1 122	- 1	119
32	1 511	1 622	10	547
33	2 018	2 376	18	764
34	1 176	1 346	15	320
35	572	596	4	129
36	514	523	2	94
37	88	81	- 8	7
38	3 311	3 978	20	1 232
39	340	379	12	141
41	79	77	- 3	2
51	2 836	3 777	33	930
60	52	54	4	1
61	4 727	5 601	19	1 409
62	22 811	23 633	4	3 616
63	3 715	4 083	10	608
Yht.	44 922	49 361	10	9 940

Yritystilastoille, joita tehdään otantatutkimuksena aikaerosta syntyvän peittävyyseron merkitys on:

1. Otoskehikossa on melkoista yli- ja alipeittävyyttä (lopettaneet yritykset/uudet yritykset, joita ei korjata mitenkään estimaateissa).

2. Yritystilastojen otantamentelmässä tärkeiden ositepainojen sijasta meillä onkin näiden estimaatteja. Perusjoukon keskiarvoilla on myös taipumusta muuttua aineistoissa, jotka ovat vinoja jakaumaltaan. Samoin varianssi muuttuu ajan mukana tämä tyyppisissä aineistoissa.

3. Yritystilastojen joillakin toimialoilla (esim. rakennustoiminta) yhteismitallisten luokitusten muodostaminen ei ole mahdollista.

4. Yritysrekisterin aikaviive siitä johtuvine peittävyyseroineen on johtanut yritystilastoissa käytäntöön, että otos on jouduttu poimimaan eri kehikosta ja tietojen estimointi suoritetaan eri kehikon avulla. Näin ollen tilastollisia laskelmia estimaattien tarkuudesta ei voida esittää.

Kato

Kato on yksi otantatutkimuksen vaikeista ongelmista silloin kun se alkaa olla yli 10 %:n luokkaa. Kadolla tarkoitetaan sitä, että epäonnistutaan joidenkin otokseen valittujen yksikköjen mittaamisessa kokonaan tai osittain. Yritystilastoissa kato muodostuu joukoista:

1. Yrityksiä ei ole onnistuttu mittaamaan mitenkään minkään muuttujan osalta.

2. Yritys on onnistuttu mittaamaan vain joidenkin muuttujien osalta, mutta näidenkään osalta ei riittävän tarkasti.

3. Katolukuihin sisältyy lopettaneita yrityksiä.

Taulukko 18. Yritystilastojen bruttokato vuoden 1981 tiedustelussa toimialoittain

Toimiala	Otoskoko lukumäärä	Yrityksiä julk.lkm.	Bruttokato lukumäärä	Bruttokato/ otoskoko %
29	16	15	1	6,3
30	14	12	2	14,3
31	181	141	40	22,1
32	182	140	42	23,1
33	152	104	48	31,6
34	142	136	6	4,2
35	93	75	18	19,4
36	48	37	11	22,9
37	39	26	13	33,3
38	216	186	30	13,9
39	52	43	9	17,3
41	47	46	1	2,1
51	410	309	101	24,6
52	129	108	21	16,3
60	27	25	2	7,4
61	539	467	72	13,4
62	1 032	800	232	22,5
63	33	22	11	33,3
7	623	529	94	15,1
Yht.	3 975	3 221	754	19,1

Edellä olevassa taulukossa on esitetty mittaamatta jääneiden otosyksiköiden lukumäärät toimialoittain vuonna 1981. Vuonna 1981 bruttokato on ollut yritystilastoissa 19 %:n luokkaa. Sen sijaan toimialoittain on ollut melkoisia vaihteluita aina 2,1 %:sta 33,3 %:iin.

Jos verrataan eri vuosina Tilastokeskuksessa mittaamatta jääneiden yksiköiden perosenttiosuutta otoskoosta, on sitä merkitsevästi saatu pienennetyksi.

Taulukko 19. Yritystilastotiedustelussa vuosina 1974 - 76 mittaamatta jääneiden otosyksiköiden osuus otoskoosta toimialoittain.

Vuosi	Mittaamatta jääneiden yksiköiden %-osuus otoskoosta		
Toimiala	1974	1975	1976
51	30	22	23
52	22	20	25
29	40	20	20
31	22	15	14
32	39	23	22
33	48	30	39
34	26	13	12
35	34	21	26
36	37	16	34
37	29	29	23
38	38	23	26
39	59	28	16
30	8	0	0
41	16	0	0
7	20	14	11
60	26	0	9
61	30	29	31
62	37	25	31
Yht.	33	22	26

Taulukossa on esitetty vuosilta 1974, 1975 ja 1976 mittaamatta jääneiden otosyksiköiden %-osuudet otoskoosta toimialoittain. Kuviosta voidaan havaita, että vuoden 1974 tiedustelussa, jolloin yritystilastot tuotettiin lähes nykyisessä muodossaan ensimmäistä kertaa, kato on ollut noin kolmannes, joillakin toimialoilla lähes puolet tiedustelun otoskoosta.

Taulukko 20. Mittaamatta jääneet yritykset teollisuuden osajoukossa yrityksen koon mukaan vuonna 1981.

Suuruusluokkaosite	Otoskoko lukumäärä	Yrityksiä julkaisussa	Bruttokato lukumäärä	Bruttok./ otoskoko %
1	116	53	63	54,3
2	193	128	65	33,8
3	149	120	29	19,5
4	152	147	5	3,3
5	577	513	64	11,1

Tarkastelemalla yllä olevaa taulukkoa, jossa on esitetty mittaamatta jääneet yritykset teollisuuden toimaloilla yrityksen koon mukaan vuonna 1981, voidaan havaita, että yritystilastoissa mittaamatta jääneiden ongelma on suurin pienten yritysten kohdalla. Osite 1:ssä, joka on alle 5 hengen yritysten osite, on kato ollut 54,3 % otoskoosta. Osite 5:n yritykset kuuluvat totaalitutkimukseen. Siten 11,1 % koko perusjoukosta tässä ositteessa jää mittaamatta. Kadon suuruus selittyy osittain sillä, että tiedustelun tietosisältö näille isoille yrityksille on laajempi.

Mittaamatta jääneiden yksiköiden vaikutus estimaatteihin ei ole suuri, jos mittaamatta jääneet yksiköt ovat täysin samanlaisia kuin mitatut otosyksiköt. Vertaamalla yritystilastojen mittaamatta jääneitä ja mitattuja yksiköitä voidaan huomata, että mittaamatta jääneet yksiköt ovat keskimäärin pienempiä ja huonommin menestyviä yrityksiä kuin mitatut. Alla olevassa taulukossa on verrattu joillakin suurilla toimialoilla vuoden 1974 ja vuoden 1977 tiedusteluissa mittaamatta jääneitä ja mitattuja yrityksiä keskenään.

Taulukko 21. Mitattujen yritysten liikevaihdon keskiarvo jaettuna mittaamatta jääneiden yritysten liikevaihdon keskiarvolla.

Toimiala	vuosi 1974	vuosi 1977
311	2,40	2,10
331	2,24	3,42
332	6,10	1,15
381	1,53	3,25
384	6,47	5,27
512	1,88	4,85
616	2,26	5,98
617	0,80	3,28
620	2,40	3,96
621	1,25	1,76
622	0,26	4,36
627	2,11	2,57

Vertailu on tehty otoskehikosta saatavalla liikevaihtotiedolla osite 5:n yrityksille. Mitattujen yritysten liikevaihdon keskiarvo on jaettu mittaamatta jääneiden yritysten liikevaihdon keskiarvolla.

Taulukon luvuista huomataan, että kumpanakin vuonna mitattujen yritysten liikevaihdon keskiarvo on huomattavasti suurempi kuin mittaamatta jääneiden yritysten liikevaihdon keskiarvo. Näin ollen yritykset näissä kahdessa joukossa eivät ole samanlaisia ainakaan liikevaihdolla mitattuna.

Yritystilastoissa ei ole tehty mitään erityisiä katotutkimuksia tai katoselvityksiä ts. on lähdetty olettamuksesta, että vastaamatta jättäneet ovat samanlaisia kuin vastanneet yritykset.

Otantavirhe

Otoksen perusteella lasketuissa parametrien arvoissa esiintyy aina nk. otantavirhettä, joka johtuu siitä, että perusjoukkoa koskevat "johtopäätökset" tehdään perusjoukon osajoukon perusteella. Otantatutkimuksen etuja on, että tutkimustuloksiin sisältyvän otantavirheen yläraja on mahdollista määrätä etukäteen halutun suuruiseksi. Tämä suurin sallittu virhe puolestaan sanelee otoskoon. Otantamentelmiin sisältyy kuitenkin riski, että virhe on sallittua suurempi.

Vuoden 1982 yritystilastojen tiedustelun otoskokoa määrättäessä on tämä suurin sallittu virhe (virheen yläraja) ilmaistu %-määräisenä poikkeamana toimialoittaisen kokonaismäärän estimaatille. Todennäköisyydeksi, että virhe ylitetään, on otettu 0,05. Virheen ylärajaksi on määrätty keskimäärin 9 % poikkeama toimialan liikevaihtomuuttujan kokonaismääräestimaatille. Tämä määrittelytapa merkitsee huomattavasti suurempaa virheen ylärajaa pienten yritysten kokonaismääräestimaatille. Yritystilastoissa ei ole laskettu mitään keskivirheitä tai luottamusvälejä millekään muuttujalle. Otantavirheen suuruutta voidaan kuvata vain oheisen taulukon liikevaihtomuuttujan otantavirheen ylärajamäärittelyn avulla.

Taulukko 22. Valittu otantavirheen yläraja otoskokoa määrättäessa ilmaistuna %:eina kokonaismäärän estimaatille (liikevaihto muuttuja)

TOL [1]	Pienet yritykset	Koko toimiala	TOL	Pienet yritykset	Koko toimiala
290	14,9	9	384	53,4	9
311	45,0	9	385	12,8	9
312	26,0	9	390	16,7	9
321	31,3	9	410	47,8	9
322	32,6	9	512	14,9	9
323	14,1	9	612	15,9	8
324	34,8	9	613	10,0	10,0
331	33,1	9	614	18,5	9,0
332	15,1	9	615	27,8	6,0
342	29,2	9	616	26,2	9,0
351	61,3	9	617	13,5	9,0
352	54,5	9	618	19,3	17,0
355	11,9	9	621	23,7	8,0
356	17,1	9	622	10,6	8,0
361	9,0	9	624	12,6	10,5
362	23,7	9	625	12,1	11,0
369	25,9	9	626	12,0	12,0
37	49,6	9	627	12,9	11,0
381	16,9	9	628	10,0	10,0
382	39,6	9	629	14,1	12,0
383	66,8	9			

1) Pieni yritys = yritys, jonka henkilökunnan määrä < 100 henkilöä.

Yritystilastoissa mielenkiinto kohdistuu useisiin muuttujiin, joille estimoidaan perusjoukon kokonaismäärätietoja. Kuitenkin otoskoko määrätään asettamalla tarkkuusvaatimukset vain yhdelle muuttujalle, liikevaihdolle.

Käsittelyvirheet

Käsittelyvirheet ovat merkitykseltään vähäisimmät yritystilastojen virhelähteinä. Niitä pyritään pienentämään useampikertaisin tarkistus- ja korjausmenettelyin. Käsittelyvirheitä saattaa kuitenkin esiintyä yritystilastoissa hyvin rajoitetussa laajuudessa esim. yritystä kuvaavien yleistietojen kohdalla, joihin on vaikea rakentaa koneellisia loogisuus-, identiteetti- tms. tarkistuksia.

Mittausvirheet

Paljon merkityksellisimpiä kuin käsittelyvirheet ovat kuitenkin mittausvirheet ja mittaamiseen liittyvät ongelmat yritystilastoissa.

Mittausvirheillä tarkoitetaan virheitä, jotka liittyvät muuttujien arvojen mittaamiseen. Syyt mittausvirheille voivat löytyä esim. puutteellisista ohjeista, liian yksityiskohtaisista muuttujien määrittelystä, tai käsitteistä ja määritelmistä, jotka ovat tuntemattomia vastaajille.

Jos tarkastellaan yritystilastoissa eri laatutekijöiden painottamista, on näistä keskeisin ollut muuttujien arvojen mittaaminen. Tämä syystä, että tilastojen yrityksiä koskeva mitattavien muuttujien määrä (tietosisältö) on suuri.

Mitattavat muuttujat perustuvat keskeisiltä osin yritysten laskentatoimesta, ennen kaikkea kirjanpidosta ja tilinpäätöksestä saataviin tietoihin. Yrityksen kirjanpitoa ja tilinpäätöstä taasen säätelee joukko lakeja, joista keskeisimmät ovat kirjanpitolaki ja -asetus. Tämä lainsäädäntö määrittelee kuitenkin tavallaan vain yleiset minimivaatimukset, jotka yrityksen tulisi täyttää ulosannettavasta informaatiosta. Nämä minimivaatimuksetkaan eivät ole yksiselitteisiä, vaan niissä on varsin paljon yrityskohtaista soveltamis- ja harkintavaltaa.

Kun yritystilastoissa halutaan mm. kansantalouden tilinpidon takia useat tilinpäätöstä koskevat tiedot huomattavasti tarkemmin eriteltyinä kuin virallinen tilinpäätös edellyttää ja kun lisäksi samat tiedot pyritään keräämään kaikilta yritystilastotoimialojen tilastoyksiköiltä samansisältöisinä, ei näiden tietojen mittaaminen ole ongelmatonta.

Mittaamisongelmat tai -vaikeudet ja virheet mittaamisessa perustuvat osaltaan seuraaviin taustatekijöihin:

1. Kirjanpitolaki sellaisenaan ei kovin hyvin sovellu kaikille toimialoille,

2. yritystilastojen yhdenmukainen tietosisältö kaikille toimialoille eli yksi yleislomake, jolla tiedot kerätään kaikilta yrityksiltä,

3. yritykset käyttävät samoja nimiä eri asioista,

4. yritystilastotiedustelun muuttujat, jotka yritys pystyy selvittämään vain käymällä läpi kirjanpitoaan "kuittitasolla",

5. yritystilaston tietosisältö t.s. mitattavien muuttujien määrä on liian suuri.

Kohdan 1 taustatekijöille esimerkkinä on rakennustoiminta kirjavine kirjanpitokäytäntöineen. Esimerkiksi rakennustoiminnan yrityksen gryndaamien talojen osakkeiden myynti eli arvopaperikauppa. Rakennustoiminnassa on 3 tapaa käsitellä arvopaperikauppaa tuloslaskelmassa:

1. Erikseen liikevaihto rakennustoiminnasta ja arvopaperikaupasta,

2. osa yrityksistä piilottaa arvopaperikaupan liikevaihtoon muuten vain,

3. osa ilmoittaa vain nettokaupan (nettokauppaa tai myyntivoittoa on ns. odotusajan lisähinta eli rakennuksen valmistumisen ja myymishetken välisenä aikana mahdollisesti perittävä lisähinta).

Tämä käytäntö heijastuu yritystilastotiedusteluun ja siinä liikevaihtomuuttujaan. Liikevaihto ei ole osin tästä syystä rakennustoiminnan eri yritysten välillä yhteismitallinen.

Kohdan 2 yritystilastojen yhdenmukainen tietosisältö kaikille toimialoille eli yksi yleislomake, jolla tiedot kerätään kaikilta yrityksiltä, ongelmana on eri toimialoille ominaiset kirjaustavat, joita on vaikea sovittaa yhdelle yleislomakkeelle. Esim. liikenteen toimialoille tyypilliset korjaus- ja huoltokulut eivät sisälly yleislomakkeelle, mistä seuraa, että yritys ilmoittaa ne aineet ja tarvikkeet tai muut liikekulut kohdassa yritystilastojen yleislomakkeella. Toimialalle tyypillisten kirjaustapojen sovittamiseen yhdelle yleislomakkeelle on myöskin hyvin keskeinen useiden muuttujien osalta rakennustoiminnan toimialoilla.

Kohdan 3 esimerkkinä eli miten yritykset käyttävät samaa nimeä eri asioista, on yrityksen palkkakulut.

Palkkojen esittämisessä tuloslaskelmassa yritykset ovat käyttäneet mm. seuraavia vaihtoehtoja:

A. Palkkoihin kirjataan vain työajalta maksetut palkat; muut palkkojen luontoiset erät luetaan sosiaalikuluihin,

B. Palkkojen kuluryhmässä esitetään kaikki veronalaiset palkat ja korvaukset.

A) Tapauksessa sosiaalikuluihin katsotaan mm. sairausajalta maksetut palkat, tapaturma-ajan palkat, lomapalkat, lomaltapaluurahat, koulutusajan palkat, palkat lasten sairausajalta, arkipyhäkorvaukset ja muut vapaapäiväkorvaukset.

Kohdan 4, yritystilastotiedustelun muuttujat, jotka yritys pystyy selvittämään vain käymällä läpi kirjanpitoaan "kuittitasolla". Tällaisia ovat muuttujat, joiden tarve on lähtöisin pelkästään jostakin yhteiskunnallisesta tietotarpeista eikä niillä ole mielenkiintoa yritysten päätöksenteossa, esimerkiksi julkiset maksut tai tukipalkkiot yms. muuttujat.

Yrityksen saamat tukipalkkiot menojen kattamiseksi voivat olla mm. seuraavia:

- Yrityksen tuotantotoiminnan käynnistysavustus
- Yrityksen kehittämisavustus (markkinointi, tutkimus, tuotekehitys)
- Nuorten työllistämistuki (1000 mk/henkilö/kk)
- Nuorten työelämään perehdyttämistuki
- Korkeakoulu- ja opistotasoista koulutusta saaneiden työllistämistuki
- Viennin edistämismäärärahat
- Tuotekehitysavustukset (avustukset uusien tuotteiden ja tuotantomenetelmien kehittämiseen)
- Kuljetusten alueellinen tukeminen (VR ja kuorma-autokuljetukset)

Yritys kirjaa ne suoraan menojen vähennyksinä ao. tileille, jolloin riippuen tukipalkkion luonteesta palkat, välilliset palkkamenot, muut liikekulut, ostomenot tai myynnin oikaisuerät ovat pienemmät. Toinen tapa yrityksellä on kirjata ne muihin tuottoihin, jolloin muut tuotot ovat tukipalkkioiden määrää suuremmat. Kolmas tapa käsitellä tukipalkkioita on, että yrityksellä ovat kaikki tukipalkkiot yhdessä ts. valtiolta ja kunnilta saadut tukipalkkiot hinnan alentamiseksi ja tukipalkkiot menojen kattamiseksi. Näin ollen tukipalkkiot muuttujan mittausvirhe ei ole yksinomaan muuttujassa tukipalkkiot vaan se vaikuttaa useaan mitattavaan muuttujaan. Vuodelta 1981 yritystilastojen tukipalkkiot estimaatti on esim. teollisuudessa vain noin 1/2 teollisuudelle myönnettyjen tukipalkkioiden määrästä.

Viidentenä kohtana mittaamisongelmissa on, että yritystilastojen tietosisältö ts. mitattavien muuttujien määrä on liian suuri. Mitattavien tietojen määrä isoilta yrityksiltä on n. 327 tietoa ja 246 tietoa pieniltä yrityksiltä. Yleensä hyvin runsaan tietomäärän kerääminen aiheuttaa tietojen antajissa väsymystä ja siten kerätyn aineiston laatu huononee.

Edellä käydyissä tilanteissa muuttujien arvojen mittaaminen oikein on vaikeaa tai mahdotonta. Miten suuria mittausvirheitä yritystilastoissa mitattavien muuttujien arvoihin sisältyy ei ole mahdollista kvantitatiivisesti määritellä.

Estimaattorin harha

Yritystilastoissa on käytössä kaikilla toimialoilla estimointimenetelmänä erillinen suhde-estimointimenetelmä. Suhde-estimaatti on harhainen paitsi tilanteessa, kun perusjoukon kaikilla alkioilla on voimassa $E(y \mid x) = \beta x$ eli y:n regressiosuora x:n suhteen kulkee origon kautta. (x on apumuuttuja, y on tutkittava muuttuja). Tällöin suhde-estimaatti on erillisellä ja yhdistetyllä ositetun otannan suhde-estimaattorilla harhaton otoskoosta riippumatta.

Harha on kuitenkin vähäinen suurissa otoksissa. Edellä mainituilla ositetun otannan suhde-estimaattoreilla harha on erilainen riippuen otoksen koosta ja ositteiden lukumäärästä. Jos otoskoko ositteessa on suhteellisen suuri (50 - 100), niin harha on vähäinen kummallakin estimaattorilla useimmissa perusjoukoissa. Kun ositteiden ototskoot ovat pieniä ja lisäksi ositteiden lukumäärä on suuri, niin käytettäessä erillistä suhde-estimaattoria, harha kokonaismäärän estimaatissa ei ehkä ole merkityksetön.

Jos harha on lisäksi samanmerkkinen kaikissa ositteissa, saattaa tapahtua että harha koko perusjoukon kokonaismäärän estimaatissa on karkeasti ositteiden lukumäärä kertaa yhden ositteen kokonaismääräestimaatin harha.

Taulukko 23. Teollisuuden yritystilaston vuoden 1982 tiedustelun otos

TOL	Suuruusluokkaosite 1	2	3	4	5	yht.
290	5	5	5	2	4	19
300	-	-	1	-	12	13
311	5	9	13	8	70	103
312	5	6	8	-	5	23
313	5	3	-	2	5	13
314	-	-	-	-	2	5
321	5	5	6	10	35	59
322	5	4	6	7	67	87
323	6	9	11	15	3	44
324	5	3	5	5	20	36
331	5	5	5	5	50	68
332	8	22	13	10	16	69
341	5	3	3	3	20	32
342	5	9	10	9	74	106
351	5	3	4	3	8	21
352	5	3	4	3	19	32
353	-	-	-	-	1	1
354	3	3	1	1	1	9
355	5	3	2	1	1	10
356	5	12	7	9	12	45
361	10	7	1	1	-	19
362	5	4	4	3	6	20
369	5	12	6	4	12	38
371	5	3	7	3	5	21
372	5	3	3	1	4	14
381	10	18	24	13	46	111
382	5	3	4	3	49	62
383	5	3	4	3	27	40
384	5	3	3	4	30	43
385	5	5	7	5	2	22
390	7	13	8	12	9	49
410	1	3	3	7	19	33
Yht.	155	184	179	152	637	1 267

Yritystilastoissa teollisuuden toimialat on tyyppiesimerkki, jossa ei missään tilanteessa pitäisi käyttää erillistä suhde-estimointia. Teollisuuden toimialoilla (edellä esitetty taulukko) ositteiden lukumäärä on noin 160 ja otoskoot muutaman yrityksen luokkaa. Tällöin jäänee harkittavaksi, mitä luokkaa käytetyllä menetelmällä harha todella on.

Laatutekijöiden arviointia

Edellä on käyty läpi Ruotsin SCB:n laaduntarkastelumallin pohjalta yritystilastojen kokonaisvirheeseen vaikuttavia tekijöitä. Ei ole mahdollista esittää, mitä luokkaa kokonaisvirhe on ehkä yritystilastoissa. Tämän selvittäminen vaatisi todella laajaa virhetekijöiden selvittämistä ja tutkimista. Sitäpaitsi nämä tekijät ovat sidoksissa toisiinsa ja niiden vaikutus kohdistuu eri tavoin eri muuttujiin.

Joskaan yritystilastoissa ei ole tehty laatuun liittyviä luotettavuustutkimuksia, on toki laadun hyväksi tehty jotain. Keskeisimmät näistä toimenpiteistä ovat olleet muuttujien mittausvirheiden, kadon, osittaiskadon pienentäminen ja relevanssitekijöiden puolelta tilastojen nopeuttaminen merkitsevästi.

Tekijöitä, joille yritystilastojen laatuun voidaan vaikuttaa on useita lähtien menetelmällisten puutteiden tai ongelmien ratkaisemisesta aina tietosisällön ongelmien selvittämiseen, jollakin tasolla myöskin hallinnollisten rekistereiden hyväksikäyttöön.

Mutta aivan erityisesti laatuun vaikuttavana tekijänä on huomioitava, että yritystilastot tehdään otantatutkimuksena. Otantatutkimuksen lähtökohtana on, että tutkimusajankohdalta (tai ainakin lähes) pitää olla käytettävissä rekisteri, otoskehikko, joka kattaa yksikäsitteisesti tarkasteltavan ilmiön, kohdeperusjoukon.

Suunnittelija Esa-Jukka Käär

LAATUKYSYMYKSET KANSANTALOUDEN TILINPIDON NÄKÖKULMASTA

Johdanto

Kuten tunnettua kansantalouden tilinpito on tilastojärjestelmä, jonka avulla pyritään kuvaamaan kansakunnan taloudellista toimintaa yhtenäisessä tilikehikossa. Perustilastoista ja niihin verrattavista tietoaineistoista kansantalouden tilinpito poikkeaa siinä, että suoraan "kentältä" kerättävät tiedot muodostavat vain hyvin pienen osan tilinpidon laadintaa varten tarvittavasta informaatiosta. Pääasiallisina tietolähteinä tilinpidon laadinnassa käytetään nimittäin juuri perustilastoja, suurten poikkileikkaustutkimusten tietoja, erilaisia hallinnollisia tietoaineistoja ja muita niihin verrattavissa olevia tietoja. Tässä mielessä kansantalouden tilinpitoa voidaankin kutsua nimellä "tilastojen tilasto".

Tätä taustaa vasten on selvää, että perustilastojen yms. laadulla on suora vaikutus myös kansantalouden tilinpidon laatuun. Koska tänään on jo muissa yhteyksissä käsitelty perustilastoja ja niihin verrattavien aineistojen laatutekijöitä varsin perusteellisesti, ei tässä yhteydessä enää puututa siihen puoleen, vaan pyritään keskittymään kansantalouden tilinpidon sisäisten laatutekijöiden tarkasteluun: mitä laadulla ymmärretään kansantalouden tilinpidossa, miten sitä voitaisiin mitata ja miten laatua voitaisiin parantaa.

Relevanssitekijät

Kansantalouden tilinpitoa on useasti kritisoitu mm. siitä, että BKT ei mittaisi oikealla tavalla hyvinvointia tai että järjestelmästä puuttuu elementtejä, joiden tulisi siihen sisältyä (esimerkiksi taloudellisen kasvun aiheuttama ympäristön saastuminen ja palkattoman kotityön oleminen kuvauksen ulkopuolella). Tässä yhteydessä ei kuitenkaan liene tarkoituksenmukaista pohtia edellä mainitun tyyppisten

seikkojen vaikutusta tilinpidon laatuun, vaan kansantalouden tilinpidon käsitteellinen sisältö otetaan annettuna eli sellaisena kuin se tällä hetkellä voimassa olevassa YK:n suosituksessa on esitetty, jota suositusta siis voidaan pitää tavoitekehikkona.

Suomen kansantalouden tilinpito uudistettiin vuonna 1979 ja noudattaa tällä hetkellä varsin tarkasti viimeisintä YK:n suositusta, joten tältä osin tilinpidon laatua voitaneen pitää varsin hyvänä.

Vertailtavuus muihin (tilastoihin) mainitaan laatumallissa eräänä relevanssitekijänä. Tilinpidon näkökulmasta olen tällä ymmärtänyt tarkoitettavan käsitteellistä vertailukelpoisuutta muiden maiden kansantalouden tilinpitotietoihin. Tämän laatukomponentin arvioimiseen Suomen kansantalouden tilinpidon osalta ei oikeastaan ole muuta kriteeriä *kuin mainitun YK:n suosituksen noudattamisen* tarkkuus, josta jo edellä oli puhe.

Ajallinen vertailtavuus on aina ollut ja on kansantalouden tilinpidon laadinnassa eräs johtava periaate. Niinpä tilinpidon uudistuksen yhteydessä laadittiin uusilla tasoilla ja uusia käsitteitä, määritelmiä ja menetelmiä noudattaen aikasarjat taaksepäin aina vuoteen 1960 asti eli taannehtivat aikasarjat viidentoista vuoden ajalta. Vastaavasti laadintakehikon muuttumattomuudella pyritään varmistamaan tällä hetkellä ja tästä eteenpäin laadittavan tilinpidon vertailtavuus aikaisempiin vuosiin.

Edellä sanottu ei kuitenkaan merkitse sitä, että kansantalouden tilinpito olisi staattinen, yhteiskunnan muuttumiseen reagoimaton järjestelmä. Kansantalouden tilinpidon laadinnassa pyritään jatkuvasti ottamaan huomioon uusien ilmiöiden ja kuvauskohteiden syntyminen varsinaisen kehikon pysyessä silti kuitenkin ennallaan. Tässä työssä tilinpidon käyttäjillä on varsin suuret mahdollisuudet vaikuttaa tilinpidon laatuun positiivisesti saattamalla tietämyksensä ja näkemyksensä tilinpidon laatijoiden tietoon. Tätä tarkoitusta varten tärkeimpiin käyttäjiin onkin olemassa kiinteästi organisoidut yhteydet, jolla on pyritty varmistamaan käyttäjien tietotarpeiden ja toisaalta palautteiden kulkeutuminen tilinpidon laatijoille.

Kansantalouden tilinpidon ehkä tärkeimpänä käyttäjänä voitaneen pitää Valtiovarainministeriötä, jonka vaatimuksesta tilinpidon laadinnan aikataulu on varsin kiinteästi kytketty valtion tulo- ja menoarvion sekä ministeriön ennusteiden laadinnan rytmiin. Tässä mielessä kansantalouden tilinpidon tietojen voitaneen sanoa täyttävän oikea-aikaisuuden laatukriteerin.

Kritiikkiä on sen sijaan kuulunut tilinpidon lopullisten tietojen valmistumisen nopeudesta vai pitäisikö sanoa hitaudesta. Tässä suhteessa pallo on kuitenkin pitkälti perustilastoilla, joista eräiden valmistumisnopeus ei anna mahdollisuuksia tilinpidon lopullisten tietojen tuottamiseen nykyistä nopeammassa aikataulussa.

Kansantalouden tilinpidon ensimmäiset ennakkotiedot vuodesta t julkaistaan kuitenkin jo heti vuoden päättymisen jälkeen tammi-helmikuussa hieman suppeammalla tietosisällöllä kuin lopulliset luvut. Seuraavana kesänä eli runsaat puoli vuotta tilastovuoden päättymisen jälkeen julkaistava seuraava ennakollinen kansantulotilasto vuodesta t tuotetaankin jo lopullisten lukujen laajalla tietosisällöllä. Tällä aikataululla Suomen kansantalouden tilinpito on eräs maailman nopeimmista.

Huoltotaseen tilastollinen ero

Kansantalouden tilinpitohan on luonteeltaan kahdenkertaista kirjanpitoa, jossa jokainen taloustoimi esiintyy kaksi kertaa, kerran menona ja kerran tulona. Teoriassa tili-identiteettien toteutuminen merkitsee myös järjestelmän "umpeenmenevyyttä". Käytännössä täydellistä umpeenmenevyyttä ei yleensä kuitenkaan saavuteta, vaan esimerkiksi kansantalouden kokonaiskysynnän ja -tarjonnan estimaatit poikkeavat toisistaan huoltotaseen tilastollisen eron verran. Tähän eroon vaikuttavat mm. eri perusaineistojen toisistaan poikkeavat taloustoimien kirjausajankohdat ja kirjausperusteet, epätäydellisesti kuvatut kansantalouden toiminnot, perusaineistoissa mahdollisesti olevat puutteet jne.

Tilinpidon laadintavaiheessa huoltotaseen tilastollisella erolla on tärkeä tehtävä tilinpidon sisäisen tasapainottomuuden paljastajana ja siten se toimiii jo tilinpidon laadinnan kuluessa eräänlaisena "laadunvalvojana". Huoltotaseen tilastollisen eron suuruudelle on kuitenkin varsin vaikea määrittää mitään ehdotonta, vielä hyväksyttävissä olevaa ylärajaa, jonka ylittyminen aiheuttaisi laadintavaiheessa automaattisesti jonkinlaisen "virheenmetsästyksen". Suomen kansantalouden tilinpidossa huoltotaseen tilastollisen eron suhteellinen suuruus on 1970-luvulla ollut keskimäärin 0,3 % kokonaiskysynnästä vaihdellen 0,1 % ja 0,9 % välillä. Myöskään mitään selvää trendiä ei tilastollisen eron kehityksessä 1970-luvulla ole havaittavissa. Perusvuodesta (tällä hetkellä vuosi 1975) etääntymisen on kuitenkin havaittu jossain määrin vaikeuttavan järjestelmän umpeenmenoa ja hieman lisäävän tilastollista eroa, mitä heijastelevat ennakollisten vuosien 1981 ja 1982 luvut taulukossa 24. Tähän ongelmakenttään liittyy tilinpidon perusvuoden vaihtaminen ja tasotarkistukset, joista lisää hieman jäljempänä.

Taulukko 24. Huoltotaseen tilastollinen ero, osuus kokonaiskysynnästä

	Suomi		Australia	Kanada	Ruotsi
Vuosi	Mmk	%		%	
1970	- 265,6	- 0,5	- 0,0	- 0,3	1,7
1971	- 83,0	- 0,1	0,5	- 0,8	0,1
1972	- 53,3	- 0.0	1,2	- 0,1	1,4
1973	763,8	0,8	2,6	0,0	1,7
1974	296,8	0,2	1,7	0,3	- 0,5
1975	306,5	0,2	0,6	0,1	0,1
1976	351,1	0,2	0,4	0,1	0,4
1977	- 1 466,8	- 0,9	0,7	0,3	0,0
1978	131,0	0,1	0,0	0,1	0,2
1979	- 491,9	- 0,2	- 0,6	0,0	0,2
1980	- 823,5	- 0,3	0,5	0,1	0,5
1981*	1 052,7	0,4	-	-	0,4
1982*	1 267,4	0,4	-	-	
Keskimäärin 1970-1980		0,3	0,8	0,2	0,6

Tilastollisen eron kansainvälistä vertailua vaikeuttaa se, että esimerkiksi OECD:n kansantulotilastossa kyseinen luku esiintyy vain muutaman maan tiedoissa. Muiden maiden kohdalla se sisältynee varastojen lisäys-komponenttiin. Australian, Kanadan ja Ruotsin tiedoista huoltotaseen tilastollinen ero on kuitenkin saatavissa. Suhteutettuna kokonaiskysyntään sen suuruus oli näissä maissa 1970-luvulla keskimäärin vastaavasti 0,8, 0,2 ja 0,6 prosenttia. Näihin verrattuna Suomen kansantalouden tilinpidosta laskettu vastaava suhdeluku näyttää varsin "kilpailukykyiseltä".

Vertailu hallinnollisiin aineistoihin

Tilinpidon joidenkin aggregaattien luotettavuutta voidaan arvioida vertaamalla niitä hallinnollisista aineistoista saataviin vastaaviin tietoihin. Tällainen vertailu voidaan tehdä tilinpidon toimialoittaisen palkkasumman ja verotusaineiston palkkasumman välillä. Nämä kaksi palkkasummaestimaattia ovat toisistaan riippumattomia, koska tilinpidon estimaatti syntyy useiden perustilastojen palkkatietojen pohjalta ja verotusaineiston tieto taas perustuu verovelvollisten ilmoituksiin.

Tilinpidossa verotusaineiston tieto on katsottu luotettavammaksi, jolloin sen ja toimialoittaisen palkkasumman erotusta käsitellään korjauseränä. Sen suuruus on ollut 1970-luvulla keskimäärin 0,7 % verotuksen palkkasummasta vaihdellen lopullisissa luvuissa nollasta lähes yhteen prosenttiin. Yhtä vuotta lukuunottamatta verotusaineiston tieto on ollut suurempi, minkä voitaneen tulkita viittaavan jonkinlaisen peittävyysongelman olemassa oloon perustilastojen puolella. Viime vuosina ero on lisäksi ollut hienoisessa kasvusuunnassa.

Taulukko 25. Verotusaineiston palkkasumman ja tilinpidon toimialoittaisen palkkasumman ero

Vuosi	Mmk	%
1975	291,5	0,6
1976	416,9	0,7
1977	292,1	0,5
1978	- 58,1	- 0,0
1979	164,5	0,2
1980	793,5	0,9
1981	1 265,9	1,3

Ennakolliset vs. lopulliset tiedot

Edellä todettiin lopullisiin lukuihin sisältyvän eräitä puutteita, joiden vaikutuksesta tilinpidossa esiintyy mm. huoltotaseen tilastollinen ero. Kun seuraavassa käsitellään tilinpidon ennakollisten lukujen suhdetta lopullisiin tietoihin, pidetään tässä tarkastelussa vertailun mielekkyyden vuoksi lopullisia tietoja kuitenkin "virheettöminä".

Kansantalouden tilinpidon lopulliset tiedot vuodesta t laaditaan vuoden t+2 kesällä. Tätä ennen kyseisen vuoden lukuja on estimoitu tilinpidossa kolmesti ennakollisten tietojen tuottamisen muodossa. Eräänä laadun mittarina kansantalouden tilinpidosa voidaankin pitää ennakkotietojen poikkeamien suuruutta lopullisista tiedoista tilinpidon eri komponenteissa. Ennakkotietojen laadinnassa korostuu tilinpidon perustilastoista poikkeava luonne varsin voimakkaasti, sillä mitä aikaisemman vaiheen ennakkotilasto on kyseessä sitä vähemmän informaatiota talouden kehityksestä on käytettävissä ja sitä enemmän tulos riippuu laatijoiden asiantuntijapanoksesta.

Seuraavassa on pyritty tutkimaan ennakkotietojen ja lopullisten tietojen välistä eroa eräissä huoltotaseaggregaateissa laskemalla kullekin poikkeamien voimakkuutta mittaava tunnusluku kaavasta[1]

$$P = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (E_i - L)^2}$$

missä P on ennustevirheen voimakkuutta mittaava tunnusluku, ns. RMS-virhe (root mean square prediction error)

n on havaintojen eli ennakollisten laskentakierrosten lukumäärä

E_i on ennakkotieto laskentakierroksella i

L on lopullinen tieto

Tarkastelun kohteeksi on valittu huoltotaseaggregaattien volyymin muutokset ennakollisissa ja lopullisissa tiedoissa[2]. Koska nykyisen kansantalouden tilinpidon ensimmäiset ennakkotiedot koskevat vuotta 1978, voidaan näitä vertailuja tästä aineistosta tehdä vain muutamalta vuodelta. Tämän vuoksi onkin rinnalle otettu lisäksi vanhan tilinpidon pohjalta lasketut vastaavat luvut vuosilta 1970-1976. Uuden ja vanhan tilinpidon pohjalta lasketut keskimääräiset poikkeamat on esitetty taulukossa 26. Vastaavat vuosittaiset luvut ovat taulukossa 27.

---

1 Theil, Henri: Applied Economic Forecasting, Amsterdam 1971

2 Jos tarkasteltavat muuttujat ovat prosenttiyksikköjä, on poikkeamien voimakkuutta mittaava tunnusluku P myös samaa laatua

Taulukko 26. Ennakollisten ja lopullisten volyymin muutosten poikkeamat huoltotaseen komponenteissa keskimäärin 1970-luvulla, prosenttiyksikköä

	Vanha tilinpito 1970-1976	Uusi tilinpito 1978-1980
Bruttokansantuote	0,8	0,8
Tuonti	1,3	0,6
Vienti	1,5	0,5
Yksityinen kulutus	0,8	0,9
Julkinen kulutus	0,9	0,9
Yksityiset investoinnit	2,9	1,6
Julkiset investoinnit	2,9	2,8

Huoltotaseen eristä vaikeimmin ennakoitavissa olivat 1970-luvulla investoinnit, joiden volyymin muutosten poikkeama ennakollisten ja lopullisten tietojen välillä oli tuona aikana keskimäärin 2,5 prosenttiyksikköä. Vastaava poikkeama BKT:n osalta oli 0,8 prosenttiyksikköä.

Verrattaessa ennakkolukujen osuvuutta vanhan ja uuden tilinpidon välillä näyttävät viennin, tuonnin ja myös yksityisten investointien ennakkotiedot osuneen uudessa tilinpidossa paremmin paikalleen. Muissa huoltotaseen erissä ei sen sijaan ole havaittavissa olennaisia eroja.

Kun eri huoltotase-erien poikkeamalukuja tarkastellaan vuositasolla taulukossa 27. havaitaan, että poikkeamat ovat eri vuosina varsin eri suuruisia. Siten esimerkiksi vuoden 1972 bruttokansantuotteen volyymin kasvu poikkesi ennakollisissa kansantulotilastoissa keskimäärin 1,4 prosenttiyksikköä lopullisesta tiedosta kun vastaava poikkeama vuoden 1976 osalta oli vain 0,1 prosenttiyksikköä. Kaiken kaikkiaan vuoden 1972 huoltotase-erien ennakoinnissa onnistuttiin varsin huonosti. Kokonaisuutena huomattavasti parempaan tulokseen päästiin sen sijaan esimerkiksi vuoden 1979 ennakkoluvuissa.

Taulukko 27. Ennakollisten ja lopullisten volyymin muutosten poikkeamat, prosenttiyksikköä

Huoltotaseen erä	Vanha tilinpito							Uusi tilinpito		
	1970	-71	-72	-73	-74	-75	-76	-78	-79	-80
Bruttokansantuote	1,0	0,9	1,4	0,9	0,7	0,8	0,1	0,7	0,8	0,9
Tuonti	1,6	1,6	1,5	0,1	1,2	1,9	1,2	0,6	0,8	0,2
Vienti	0,4	0,6	1,3	2,4	2,4	1,7	1,5	0,1	0,4	0,3
Yksityinen kulutus	0,6	0,5	2,4	0,6	0,5	0,2	1,0	2,2	0,4	0,3
Julkinen kulutus	0,6	0,1	1,6	1,2	0,5	1,4	1,1	1,4	0,3	1,0
Yksityiset investoinnit	3,7	3,2	2,3	1,4	2,5	2,7	4,4	1,4	0,6	2,3
Julkiset investoinnit	1,8	1,7	6,7	2,8	2,3	2,7	2,4	0,9	2,0	5,6

Laadun parantamismahdollisuuksia

Edellä on jo viitattu perusvuodesta etääntymisen vaikeuttavan järjestelmän "umpeenmenoa" ja lisäävän tilastollista eroa. Tällaisia havaintoja on tehty sekä nykyisestä tilinpidosta sen varsin lyhyenä "elinaikana" että vanhasta tilinpidosta. Nyttemmin onkin päädytty siihen, että tilinpidon perusvuotta muutetaan tiheämmin kuin tähän asti, esimerkiksi viiden vuoden välein. Tässä yhteydessä korjataan tilinpidon tasoissa olevat puutteet ja kehitetään laadintamenetelmiä siten, että ne vastaisivat paremmin tämän hetken todellisuutta. Tällä keinoin pyritään saamaan kansantalouden rakenteellinen kuvaus ajan tasalle, jolloin myös hintaindeksien perusvuoden vaihtaminen voi nojautua oikeampaan rakenteeseen. Edellä mainituilla keinoilla pyritään vaikuttamaan lähinnä huoltotaseen tilastollisen eron pienentämiseen.

Tuleviin perusvuoden vaihtamisiin liittyy kuten vuoden 1979 uudistukseenkin itsestään selvänä toimenpiteetnä riittävän pitkien, uusilla tasoilla olevien taannehtivien aikasarjojen tuottaminen.

Taso- ja rakennekorjausten ohella tilipidon ennakkotietojen osuvuuden lisäämistä voidaan pitää keskeisenä osatavoitteena pyrittäessä tilinpidon laadun parantamiseen. Mahdollisimman luotettavan kuvan saaminen kansantalouden kehityksestä jo tilinpidon ennakkotietojen pohjalta merkitsee useille käyttäjille tilinpidon keskeistä antia. Tilinpitoon kohdistuvat käyttäjien arvostelut käsittelevätkin useimmiten juuri ennakkotietojen luotettavuutta.

Tämän tavoitteen saavuttamista voisivat palvella mm. uusien, paremmin ennakoivien indikaattoreiden kehittäminen, ennakollisten perustilastojen luotetavuuden lisääminen sekä panostaminen tilinpidon laatijoiden ammattitaitoon.

Tilinpidon käyttökelpoisuutta ja sitä kautta laatua nostavat myös tilinpidon toimialoittaisia ja sektoreittaisia laadintamenetelmiä käsittelevät kuvaukset, jotka kattavat jo lähes kaikki tilinpidon laskentakokonaisuudet. Näiden kuvausten pohjalta käyttäjillä on mahdollisuus esimerkiksi arvioida tietojen sopivuutta omiin tarkoituksiinsa ja niiden vertailukelpoisuutta muihin tietolähteisiin.

Suunnittelija Pirjo Liewendahl

YHTEENVETO YLEISKESKUSTELUISTA

Tilastokeskuksen toimintaa tulisi suunnata jatkuvasti tuotteidensa laatua tutkivaksi laitokseksi ja laitokseksi, joka myös kertoo tuotteidensa laadusta. Laatuselosteita pitäisi ruveta tekemään mahdollisimman pian. Laatu ja laatuseloste on kuitenkin pidettävä selvästi erillään toisistaan, koska laatuselosteen olemassaolo on vain yksi laadun ulottuvuus. Laatuseloste on eräänlainen tuoteseloste, sisältöseloste tai käyttöohje, jossa on käyttäjän näkökulma otettu huomioon. Tilaston laadun analysoinnin pohjana ja laatuselosteen mallina voitaisiin pitää Ruotsin tilastokeskuksen, SCB:n laatumallia. Laatuseloste voisi myös sisältää laatujaoston raportissa esillä olevat viisi kohtaa:

- tilaston vertailtavuus ajassa
- vertailtavuus muihin tilastoihin
- käyttökelpoisuutta koskevat rajoitukset
- käsitteiden määrittely
- julkaisemattomien tulosten tallennus ja saatavuus

Laatuselosteet palvelevat tilastojen käyttäjiä. Ne eivät kuitenkaan sellaisenaan riitä käyttäjien tarpeisiin. Tarvitaankin kaksi erilaista laatuselostetta: suuremmalle käyttäjäryhmälle tarkoitettu selväkielinen, ymmärrettävän tuoteselosteen tyyppinen sekä erikoisryhmille ns. tekninen raportti, jossa esitetään yksityiskohtaisia tietoja tilaston "virheistä", tilaston laadinnassa käytetyistä menetelmistä, laskukaavat jne.

Keskustelussa kysyttiin myös sitä, kuka tarvitsee laatuselostetta. Eikö tilaston käyttäjä voi luottaa Tilastokeskuksen tietoihin ilman tällaista selostetta? Tieteellinen tutkimustyö on tietysti poikkeus. Tutkija haluaa tietää, miten tilasto tai analyysi on tehty. Olisikin ensisijassa keskityttävä laadun parantamiseen ennemmin kuin selosteiden tekemiseen. Tavoitteena pitää olla, että ilmiöitä pyritään kuvaamaan mahdollisimman oikein.

Oikea laatu on ilmiön riittävän tarkka ja riittävän oikea kuvaus. Tiedon ei tarvitse olla tarkempaa kuin käyttäjä tarvitsee. Laatukeskustelussa on otettava huomioon myös tilastotuotannon määrä sekä mistä saadaan voimavarat laadun parantamiseen. On selvitettävä, tuotetaanko nyt turhia tilastoja tai turhia tietoja. Turhaa tuotantoa on karsittava, ja vapautuvilla voimavaroilla voidaan parantaa tärkeimpien tilastojen laatua.

Laatuongelma nähtiin myös paljolti systeemien ongelmana. Laatuselosteet ja tekniset raportit voitaisiin rakentaa niin, että systeemit kuvataan niissä hyvin.

Erilaisten tilastojen laatuongelmat ovat myös erilaisia. Kansantalouden tilinpidon järjestelmän (SNA:n) kannalta perustilastojen laadulla on merkitystä. Rekistereillä on omat laatuongelmansa. Viraston ulkopuolella tuotettaviin rekistereihinkin kertyy vuosittain aina jonkin verran uusia virheitä. Myös rekistereistä olisi tehtävä luotettavuustutkimuksia.

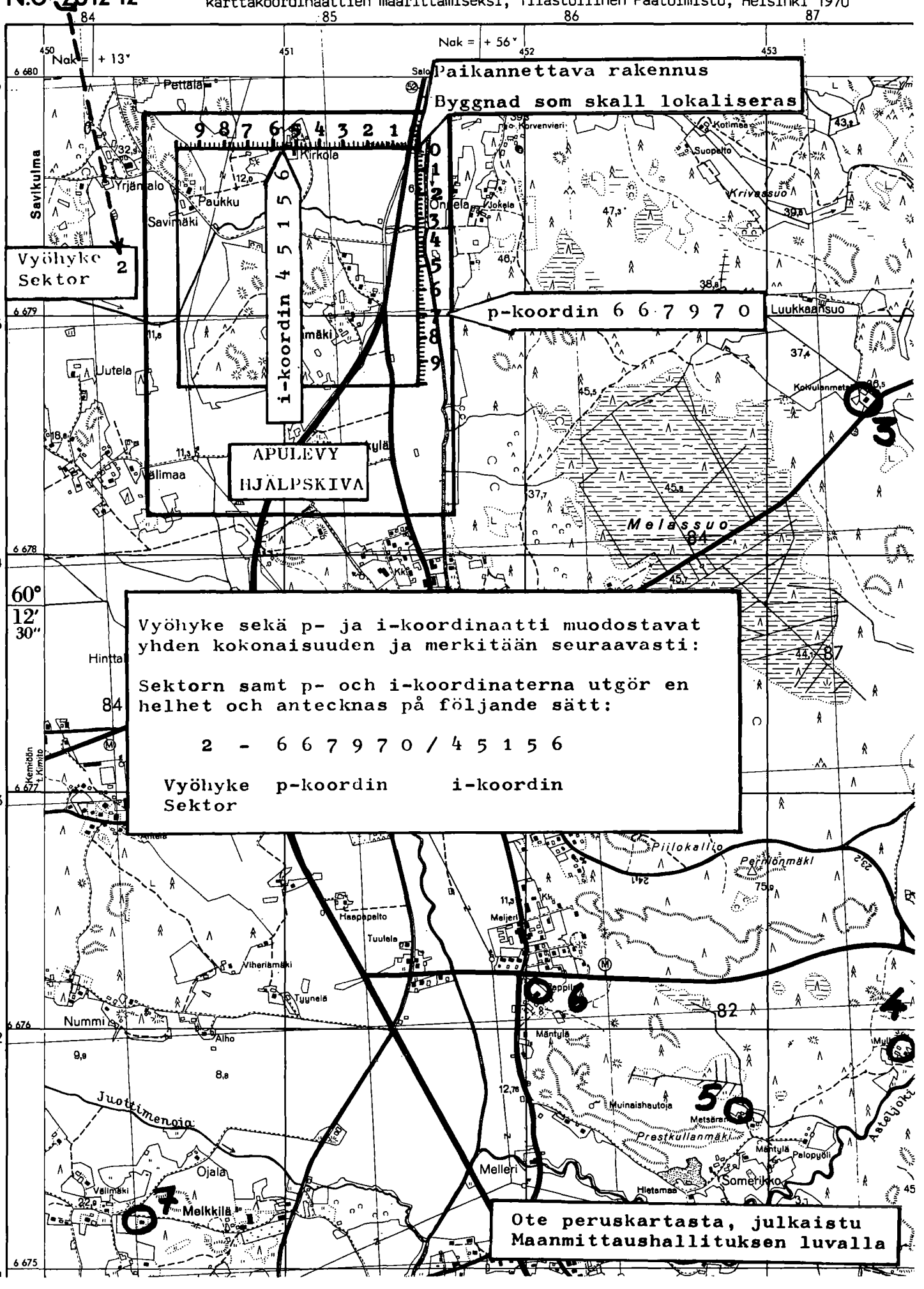
Yksi konkreettinen jatkotoimenpide-ehdotus on laatuselosteohjeiston tuottaminen työryhmässä, joka ottaa huomioon eri tilastojen tarpeet ja laatii yhtenäisen ohjeiston laatuselostetta varten tai vaihtoehtoisia laatuselostemalleja. Jatkossa on myös pidettävä yhteyttä tilastojen käyttäjiin ja tietojen antajiin. On otettava huomioon, mitä he laatuselosteilta odottavat. Vähitellen käyttäjät oppivat vaatimaan tietoja tilastojen käytettävyydestä ja rajoituksista.

Jonkinlaisia ohjeita tai normeja voitaisiin pohtia myös sille, milloin tilastossa on liikaa virheitä ja miten silloinkin pitäisi menetellä.

Peruskartta 1:20000 Grundkarta

N:o 2012 12

Väestökirjanpidossa, väestölaskennassa ja talonrakennustilastossa sovellettavat ohjeet karttakoordinaattien määrittämiseksi, Tilastollinen Päätoimisto, Helsinki 1970



TILASTOKESKUS

TUTKIMUKSIA

Tilastokeskus on julkaissut Tutkimuksia v. 1966 alkaen, v. 1980 lähtien ovat ilmestyneet seuraavat:

55. *Carita Putkonen,* Tupakkatuotteiden kulutus, Tupakkatuotteiden kokonaiskulutus Suomessa vuosina 1960–1978. Tammikuu 1980. 23 s.

56. *Pentti Tuominen,* Tuottajahintaiset indeksit, Menetelmät ja käytäntö, Teollisuuden tuottajahintaindeksi 1975=100, Kotimarkkinoiden perushintaindeksi 1975=100, Vientihintaindeksi 1975=100, Tuontihintaindeksi 1975=100. Maaliskuu 1980. 124 s.

56. *Pentti Tuominen,* Producer Price Indices, Methods and Practice, Producer Price Index for Manufactured Products 1975=100, Basic Price Index for Domestic Supply 1975=100, Export Price Index 1975=100, Import Price Index 1975=100. August 1980. 125 p.

57. *Risto Kolari,* Kuolleisuus, Kohorttikuolleisuus Suomessa v:sta 1851 lähtien. Maaliskuu 1980. 94 s.

58. *Vihavainen – Valppu – Suokko – Björk,* Kansantalouden tilinpito, Pääomakanta vuosina 1965–1977. Huhtikuu 1980. 101 s.

59. *Pekka Mäkelä,* Maatalouden yritystilastojärjestelmän kehittäminen. Toukokuu 1980. 35 s.

60. Tuotantomenetelmien kehittäminen, Tilastokeskuksen pitkän tähtäyksen suunnitelman erillisselvitys. Joulukuu 1980. 38 s.

61. *Pekka Mäkelä – Raimo Nurminen,* Kansantalouden tilinpito, Maa-, metsä- ja kalatalous sekä metsästys kansantalouden tilinpidossa. Syyskuu 1980. 129 s.

62. *Uotila – Leppä – Katajala,* Kansantalouden tilinpito, Neljännesvuosittainen kansantalouden tilinpito. Marraskuu 1980. 58 s.

63. *Heikki Sourama – Olli Saariaho,* Kansantalouden tilinpito, Rakenne, määritelmät ja luokitukset. Joulukuu 1980. 87 s.

64. *Hannu Pääkkönen,* Työttömyyden keston mittaaminen Suomessa. Maaliskuu 1981. 35 s.

65. *Iiris Niemi – Salme Kiiski – Mirja Liikkanen,* Suomalaisten ajankäyttö 1979. Maaliskuu 1981. 143 s

65. *Iiris Niemi – Salme Kiiski – Mirja Liikkanen,* Use of time in Finland 1979. September 1981. 143 p.

66. *Sinikka Askelo,* Opettajien työmäärätutkimus 1979, Peruskoulu, Lukio, Harjoittelukoulut, Iltakoulut. Huhtikuu 1981. 234 s.

66. *Sinikka Askelo,* Lärarnas arbetsmängd 1979, Grundskolan, Gymnasiet, Övningsskolorna, Aftonläroverken. April 1982. 226 s.

67. *Ari Tyrkkö,* Palkansaajien ansio- ja tulotason kehityksen mittaaminen. Toukokuu 1981. 74 s.

68. *Olavi E. Niitamo,* Tilastollinen tietohuolto 1980-luvulla. Elokuu 1981. 40 s.

69. *Juha Pietilä – Aku Alanen,* Valtion tulot ja menot lääneittäin 1978. Elokuu 1981, 88 s.

69. *Juha Pietilä – Aku Alanen,* Statens inkomster och utgifter länsvis, December 1981. 86 s.

70. Rakennuskustannusindeksit 1980=100. Elokuu 1981, 59 s.

71. Tilastokeskuksen pitkän tähtäyksen runkosuunnitelma. Syyskuu 1981. 49 s.

71. *Niitamo, O.E. – Laihonen, A. – Tiihonen, P.,* Long-Term Planning in the Central Statistical Office of Finland, November 1981. 70 p.

72. *Raili Broas,* Kansantalouden tilinpito, Yritykset kansantalouden tilinpidossa. Huhtikuu 1982. 62 s.

73. *Veli-Jukka Leppänen – Henry Takala,* Kansantalouden tilinpito, Rakennustoiminta kansantalouden tilinpidossa. Huhtikuu 1982. 112 s.

74. *Heli Jeskanen-Sundström,* Kansantalouden tilinpito, Teollinen toiminta kansantalouden tilinpidossa. Huhtikuu 1982. 55 s.

75. Tie- ja maarakennuskustannusindeksit 1980=100. Toukokuu 1982. 15 s.

76. *Matti Parkkinen,* Teollisuustuotannon volyymi-indeksin laskenta. Kesäkuu 1982. 80 s.

77. *Kari Ritvanen,* Kansantalouden tilinpito, Voittoa tavoittelemattomat yhteisöt kansantalouden tilinpidossa, Kesäkuu 1982. 70 s.

78. *Reino Hjerppe – Jouko Kaartinen,* Markkinaton tuotanto, hyvinvointi ja kansantalouden tilinpito. Elokuu 1982. 71 s.

79. *Pertti Marjomaa,* Väestökehitys ja kotitalouksien kulutusmenot. Elokuu 1982. 60 s.

80. *Lättilä – Heiskanen – Komulainen – Niskanen – Siren,* Tapaturmat ja väkivalta 1980, Haastattelututkimus tapaturmien ja väkivallan yleisyydestä sekä terveyshaitoista. Syyskuu 1982. 277 s.

80. *Lättilä – Heiskanen – Komulainen – Niskanen – Siren,* Accidents and Violence 1980, A Survey of the Incidence of Accidents, Criminal Violence and the Resulting Injuries. April 1983. 177 p.

81. *Wolfgang Lutz,* Demographic Transition and Socio-economic Development in Finland 1871 – 1978 – a Multivariate Analysis. November 1982. 81 p.

82. *Mauri Nieminen,* Muuttoliike, Maassamuuton syyt 1977 – 78. Marraskuu 1982. 90 s.

83. Suomen kansantalouden tilinpitoa sata vuotta. Joulukuu 1982. 93 s.

84. *Risto Kolari,* Korkeakouluopiskelu, Korkeakoulututkinnot ja opintojen keskeyttäminen 1966 – 1979. Joulukuu 1982. 100 s.

85. *Eeva Hamunen,* Kansantalouden tilinpito, Liikenne kansantalouden tilinpidossa. Joulukuu 1982. 79 s.

86. *Jarmo Hyrkkö – Ilkka Lehtinen,* Nettohintaindeksi 1977=100. Tammikuu 1983. 37 s.

87. *Veli-Matti Lehtonen – Timo Puustinen – Pentti Tuominen.* Vuoden 1980 kuntien kalleustutkimus. Maaliskuu 1983. 176 s.

87. *Veli-Matti Lehtonen – Timo Puustinen – Pentti Tuominen.* Dyrortsundersökningen 1980. Mars 1983. 194 s.

88. *Markku Tahvanainen – Markku Lindqvist,* Talonrakennustoiminnan volyymi-indeksi 1980=100. Huhtikuu 1983. 79 s.

89. *Olavi E. Niitamo,* Johdonmukaisuus ja ristiriita. Huhtikuu 1983. 54 s.

90. Future Trends of Data Management, *Olavi E. Niitamo,* Data Management in the 1980s. *Olli Janhunen,* Integration of Data, Applications and the Mass Media. May 1983. 23 p.

91. *Iiris Niemi,* The 1979 Time Use Study Method. May 1983. 27 p.

92. *Heikki Havén – Risto Syvänperä,* Aikuiskoulutukseen osallistuminen 1980. Toukokuu 1983. 100 s.

93. *Lättilä – Heiskanen,* Piilorikollisuus 1980, Haastattelututkimus henkilöön kohdistuneesta ilmitulleesta rikollisuudesta ja piilorikollisuudesta. Toukokuu 1983. 99 s.

93. *Lättilä – Heiskanen,* Hidden Criminality 1980, A Survey on Reported and Unreported Offences Against the Person. December 1983. 92p.

94. *Aila Repo,* Väestön tutkinto- ja koulutusrakenne-ennuste 1981 – 1995. Toukokuu 1983. 154 s.

95. *Hilkka Vihavainen – Seppo Varjonen,* Kansantalouden tilinpito, Investoinnit kansantalouden tilinpidossa. Toukokuu 1983. 75 s.

96. *Tuomo Niskanen,* Rikosvahingot 1980, Otantatutkimus poliisin tietoon tulleista omaisuus- ja väkivaltarikoksista. Kesäkuu 1983. 99 s.

96. *Tuomo Niskanen,* Crime Damages 1980, A Study of Property Offences and Violent Offences Reported to the Police. February 1984. 88 p.

97. *Risto Kolari,* Ammatillinen liikkuvuus Suomessa 1970/1975. Heinäkuu 1983. 127 s.

98. *Mauri Nieminen,* Suomen ruotsinkielinen väestö vuosina 1980–2020. Heinäkuu 1983. 51 s.

99. *Carita Putkonen,* Luotettavuustutkimus Väestö- ja asuntolaskenta 1980 osa I, Selvitys mittaus- ja käsittelyvirheistä. Marraskuu 1983. 215 s.

100. *Carita Putkonen,* Luotettavuustutkimus Väestö- ja asuntolaskenta 1980 osa II, Peittävyysselvitys.

101. *Rudolf Andorka – Istvdn Harcsa – Iiris Niemi,* Use of Time in Hungary and in Finland, Comparison of results of time budget surveys by the Central Statistical Offices of Finland and Hungary. November 1983. 62 p.

102. Air Quality Statistics: methods and principles Finland. Hungary. Sweden. December 1983. 103 p.

103. *Esa-Jukka Käär – Veli-Jukka Leppänen,* Kansantalouden tilinpito, Aluetilinpito 1976 ja 1978: menetelmät ja tulosten analyysi. Joulukuu 1983. 106 s.

104. Tilastojen laadun tutkiminen, Tilastokeskuksen laatupäivän esitelmät. Helmikuu 1984. 75 s.

Tätä julkaisua myy

VALTION PAINATUSKESKUS

Postimyynti

PL 516
00101 HELSINKI 10
Vaihde (90) 539 011

Kirjakaupat Helsingissä

Annankatu 44
(Et. Rautatiekadun kulma)
Vaihde (90) 17 341

Eteläesplanadi 4
Puh. (90) 662 801

Denna publikation säljes av

STATENS TRYCKERICENTRAL

Postförsäljning

PB 516
00101 HELSINGFORS 10
Växel (90) 539 011

Bokhandlar i Helsingfors

Annegatan 44
(i hörnet av S. Järnv.g.)
Växel (90) 17 341

Södra esplanaden 4
Tel. (90) 662 801

This publication can be obtained from

GOVERNMENT PRINTING CENTRE

Mail-order business

P.O.Box 516
SF-00101 HELSINKI 10
Phone (90) 539 011

Bookshops in Helsinki

Annankatu 44
Phone (90) 17 341

Eteläesplanadi 4
Phone (90) 662 801

ISSN 0355-2071
ISBN 951-46-7595-9