

# Yleistäminen vaikeaa sekä tieteessä että mediassa

■ Seppo Laaksonen

**Tilastotieteessä yleistäminen tarkoittaa johtopäätösten tekemistä otoksesta tai muusta suppeammasta joukosta isompaan kokonaisuuteen, perusjoukkoon. Media ja muut julkiset toimijat voivat vastaavalla tavalla tehdä yleistyksiä. Aina on vaarana niin sanottu yleistysvirhe. Tarkastelen näitä kysymyksiä esimerkkien, etupäässä galluppien ja Pisa-tulosten valossa.**

Tieteellisen tiedon ja termien käyttö on medialle hankalaa, koska sen tulee saada sanomansa perille tavallisellekin lukijalle. Sen vuoksi yksi vaihtoehto on jättää julkaisematta kirjoitukset joissa pyritään sanomaan asiat niin oikein kuin mahdollista. Näin on uskoakseni usein käynyt. Toinen ääri vaihtoehto on tehdä tieteellisistä tuloksista pinnallisia analyyskejä, jopa huvittavia. Kolmas linja on esittää ihan oikeita tuloksia mutta jättää sivummalle tulosten tarkkuus. Tällöin pienetkin erot voivat nostaa suuria otsikoita.

Tilastollisen tiedon käsittely on medialle erityisen hankalaa. Tarkastelen tilannetta erityisesti Pisan eli OECD:n koulusaavutustutkimuksen uusimman uutisoinnin näkökulmasta, mutta esitän muitakin esimerkkejä. Esimerkkini ovat pääosin maamme päälehdessä, *Helsingin Sanomista*. Lehdellä on myös hyviä käytäntöjä, jollainen on kertoa galluppien ja muiden tulosten laadusta jotakin. Niinpä 19.12.2013 sen Fakta-laatikossa sanotaan: ”TNS Gallup selvitti Helsingin Sanomien toimeksiannosta puolueiden kannatusta edustakuntavaaleissa. Puhelinhaastatteluja oli 25.11.–14.12.2013 yhteensä 2 429 kappaletta. Vastaajat edustavat Suomen äänestysikäistä väestöä Ahvenanmaata lukuun ottamatta. 33 prosenttia vastanneista ei osannut tai halunnut sanoa, mitä puoluetta he äänestäisivät. Tähän ryhmään kuuluvat myös ne, jotka

eivät aio äänestää. Tulosten virhemarginaali on vajaat 2 prosenttiyksikköä suuntaansa suurimpien puolueiden kohdalla.”

Seloste ei toki kerro mitään siitä, kuinka hyvin vastaajat edustavat äänestysikäistä väestöä. Ilmeisesti yli 80-vuotiaat yhä puuttuvat aineistosta, kuten vuonna 2011, jolloin sain aineistot käyttöni (Laaksonen 2011). Virhemarginaali kerrotaan mutta ilmeisesti tilan sääntämisen takia ei sitä, miten paljon pienempi se on pienemmällä puolueilla. Tämä seloste on mainio siksi, että lukijat kiinnittäisivät tulosten epävarmuuteen jotain huomiota. Eri asia on sitten se, kuinka vähän media itse siitä välittää. Muutaman kymmenyksenkin ero voidaan uutisoinnissa katsoa merkittäväksi, vaikkei ero olisi tilastollisesti merkitsevä.

Tilastollinen merkitsevyys ei toki ole yksikäsitteinen. Se riippuu myös siitä, kuinka suuren riskin tulkitsija on valmis ottamaan johtopäätöksenä epäonnistumiselle. Median virhemarginaalissa riski on 5 prosenttia. Isommankin riskin toki voi ottaa, jolloin on mahdollista pitää pienempiä eroja olennaisina. Kaiken kaikkiaan tässä on kyse siitä, miten otoksen tai muun osa-aineiston tieto yleistetään tavoiteperusjoukkoon, mikä puoluekannatusmittauksissa tarkoittaa vaaleissa äänestäviä, jos vaalit pidettäisiin kyselyn tekoaikana. Pisa-tutkimuksessa yleistäminen taas tapahtuu 15 vuotta täyttäneiden ikäluokkaan, joka käy koulua. Suomessa tämän ikäluokan koko vuoden 2012 tutkimuksessa on hieman yli 60 000.

Huonompi käytäntö *Helsingin Sanomissa* koskee muun muassa yhä enemmän käyttöön tulleiden internet-paneelien hyväksi käyttöä. Tämä kyselytapahan on helpohko, jos paneeli on luotu, mutta menetelmä ei takaa laatua. *Hel-*

*singin Sanomat* uutisoi 22.12.2013 ”Kehno työrauha syy Pisa-tappioon.” Tässäkin uutisessa on Fakta-laatikko, jossa kerrotaan 18–74-vuotiaita vastaajia olevan 1 079 ja virhemarginaalin kolme prosenttiyksikköä suuntaansa. Lisäksi todetaan käytetyn TNS Gallupin internet-paneelin. Tämä on periaatteessa hyvä asia, mutta harva lukija tietää, mitä se tarkoittaa. Itse olen äskettäin kirjoittanut paneelista kahdellakin julkisella foorumilla (Laaksonen 2013a ja b). Niitä ei lehden lukijan voi edellyttää tutkineen. Siksi tarkempi tieto olisi paikallaan. Sellaista suosittelee STT-Lehtikuva, jonka tyylikirjassa sanotaan (<http://www.stt.fi/content/tyylikirja>):

”SUHTAUDU VARAUKSELLESTI NETTIPANEELIEN JA HARKITSE JULKAISUA ERIKSEEN. KERRO LUKIJALLE MENETELMÄN EPÄVARMUUSTEKIJÄT. Nettipaneeliin voi ilmoittautua itse netissä, mutta siihen saatetaan kutsua väkeä myös esimerkiksi puhelimitse ja postitse. Lopullinen vastaajajoukko edustaa usein karkeasti koko väestön ikä-, sukupuoli- ja kotipaikkajakaumaa, mutta otos ei välttämättä vastaa mielipiteiden jakautumaltaan koko väestöstä. Syynä on se, että mukaan saattaa valikoitua keskimääräistä aktiivisempia ihmisiä. Lisäksi otoksesta voi jäädä pois ihmisiä, jotka eivät käytä internetiä ainakaan aktiivisesti.”

Oletan, että STT-Lehtikuvan asiakkaat noudattavat tyylikirjan ohjeistusta.

Kummassakin *Helsingin Sanomien* Fakta-laatikossa tuodaan rutiininomaisesti esille tulosten epävarmuus. Eri asia on sitten, kuinka vakavissaan lehden toimittajat siihen suhtautuvat. Monet esimerkit viittaavat siihen, että mitään muuta ei tarvita. Juttu itse voidaan kirjoittaa siitä välittämättä, tavoitteena riittävä ”vetävyys”. Pelkona lienee, ettei kukaan lue juttua, ellei mukana ole liioittelua.

Pisa 2012 antaa tästä hyviä esimerkkejä. Esimerkiksi Suomen sijoitukseksi matematiikassa uutisoidaan 12:s, vaikka julkisen tiedon mukaan tilastollisesti samalla tasolla on useita maita, ja Suomen sijoitus objektiivisesti määriteltynä on välillä 11–17.

Toinen asia näissä mittauksissa on se, että osaaminen uutisoitiin matematiikkana, luonnontieteenä ja lukemisena, vaikka kaikissa kyse on lukutaidosta (*literacy*), joka tarkoittaa oppilaalle annettujen käytännönläheisten tehtävien ymmärtämistä ja oikein tulkitsemista. Tämä

näkökulma oli mediassa hyvin vähäisessä roolissa. Itse olen käyttänyt pitkään opetuksessa Pisaa esimerkeissä, ja matematiikan sijasta termi on ”matemaattis-tilastollinen lukutaito”. Termi sinänsä ei ole tärkeä mutta ansaitsisi asianmukaisen selosteen.

Pisa-uutisoinnissa minua häiritsi myös se, että hyvin vähän kerrotaan tuloksia sen laajasta kyselymateriaalista, joka mielestäni tarjoaisi arvokkaampaa tietoa koulujen kehittämiseen kuin pelkät osaamistulokset. Olisi esimerkiksi hyvä löytää syyt sille, miksi Suomi sijoittuu koulussaolon onnellisuudessa aivan viimeisten joukkoon. Naapuri Viro on samalla tasolla.

*Helsingin Sanomat* teki 8.12.2013 myös laajan ja monilta osin ansiokkaan jutun poikien ja tyttöjen osaamiseroista: ”Ensimmäistä kertaa Pisa-tutkimusten historiassa Suomen tytöt olivat kaikissa testeissä parempia kuin pojat.” Näin voidaan ilman virhemarginaalia toki todeta, mutta jos se otetaan huomioon, ei tyttöjen ja poikien keskiarvojen ero matemaattis-tilastollisessa lukutaidossa ole lähimainkaan merkitsevä. Edellisessäkään Pisassa vuonna 2009 ero ei ollut merkitsevä, vaikka keskiarvo oli niukasti toisin päin.

Tytöt ja pojat ovat samalla tasolla matematiikassa valtaosassa maita. Australia ja Sveitsi ovat niukasti poikkeuksia, poikien ollessa parempia. Sen sijaan tilanne on ollut kaikkien kolmen viimeaikaisen Pisa-tutkimuksen osalta sama kirjoitetun tekstin ymmärtämisessä ja tulkitsemisessä. Tyttöjen taso on huomattavasti parempi. Suomi on tältä osin ollut kärjessä, mutta nyt Viro nousi sen edelle. Tähän eroon olisi voitu havahtua jo monta vuotta sitten.

Otoksen ymmärtäminen ja sen pohjalta yleistäminen ei ole helppoa. Pisa-esimerkeissäni toimittajat tekivät ”yleistysvirheen”. Tätä ihan hyvää termiä käyttää kolumnisti Matti Apunen (HS 31.12.2013) väittäessään arkkipiispa Kari Mäkisen syyllystyneen siihen lausuessaan: ”Meidän pitäisi toivoa tiedostusvälineille suhteellisuudentajua, oikeudenmukaisuutta ja totuudellisuutta.” Apusen mielestä Mäkinen tekee yleistysvirheen, koska arvostelee tiedostusvälineitä noiden puutteesta. Apusen perustelua en

ymmärrä, koska lause ei sisällä tällaista väitettä.

Internet-paneelin tulosten perusteella on mahdollista tehdä yleistyvirhe, mutta tuloksen hyvyteen on mahdotonta päästä kunnolla kiinni, koska aineisto ei ole kunnan otos. Sen sijaan Pisa-aineistot ovat hyvin tarkasti valvottuja ja otokset korkeatasoisia. On kiinnostavaa, että tuloksia voidaan epäillä muista syistä, erityisesti jos joistakin ryhmistä poimitaan suhteellisesti isompia otoksia. Suomen 2012 aineistossa haluttiin saada kahdesta ryhmästä selvästi isompi aineisto, ruotsinkielisistä ja maahan muuttaneista. Näin siksi, että olisi mahdollista laskea tarkempia tuloksia heistä.

En havainnut, että ruotsinkielisten yliedustus olisi herättänyt keskustelua, kuten maahan muuttaneiden yliedustus. Kansanedustaja Osmo Soininvaara otsikoikin bloginsa ([www.soininvaara.fi](http://www.soininvaara.fi)) 8.12.: ”Vaikuttiko näytteen valinta Pisa-katastrofiin?” Vaikka kirjoittajaa ja muita lukijoita valistettiin otosaineiston pohjalta tehtävän arvion eli estimoinnin periaatteista, lienee joillakin yhä tuo sama usko. Osallistuin itsekin samassa blogissa keskusteluun ja tein muutamia lisäanalyyskejä. Ne eivät puolla Soininvaara väitettä, mutta tosiasia on että Suomessa maahanmuuttajien lasten osaamistaso on melko matala. Syntyperäisten taso on joka tapauksessa laskeutunut. Eräs blogiosallistuja epäili, että maahanmuuttajien määrä huonontaa syntyperäisten tasoa. En löytänyt tähän näyttöä.

## Kirjallisuutta

Laaksonen, Seppo (2011). Gallupeilla politikoidaan. Elokuu. *Kanava*.

Laaksonen, Seppo (2013a). *Surveymetodiikka. Aineiston kokoamisesta puhdistamisen kautta analyysiin*. Ventus Publishing. 2nd Edition. 230 ss. <http://bookboon.com/fi/surveymetodiikka-ebook>

Laaksonen, Seppo (2013b). Nettikyselyt ovat nykyaikaa. *Kansantaloudellinen Aikakauskirja* 4, 541–548.

**Kirjoittaja on Helsingin yliopiston tilastotieteen professori.**

## Klaus Mäkelä – tekstit, teot ja elämä

Professori Klaus Mäkelä kuoli viime joulukuun 17. päivänä. Hänen muistokirjansa ilmestyy kesäkuun alussa.

Klaus Mäkelä oli kansainvälisesti arvostettu alkoholitutkija, tieteen kehittäjä, ajattelijana, yhteiskunnallisten uudistusten vaikuttajana, satojen intohimojen yhteiskuntatieteilijä ja suomen kielen salarakastaja.

Näistä aiheista kirjoittavat Risto Alapuro, Claes Andersson, Ilkka Arminen, Irmgard Eisenbach-Stangl, Lars D. Eriksson, Antti Eskola, Katarina Eskola, Elina Haavio-Mannila, Auli Hakulinen, Elina Hemminki, Risto Heiskala, Risto Jaakkola, Kalevi Kivistö, Risto Kolanen, Kimmo Leppo, Leila Melart, Margaretha Mickwitz, Pekka Peltola, Teuvo Peltoniemi, Kari Puro, Robin Room, Pia Rosenqvist, Tom Sandlund, Jussi Simpurra, Astrid Skretting, Marja-Leena Sorjonen, Kerstin Stenius, Pekka Sulkunen, Ilkka Taipale, Kaarlo Tuori, Ritva Turunen-Grönick ja Matti Virtanen. Kirjassa on myös Klaus Mäkelän omia tekstejä.

Kirja ilmestyy Sosiaalipoliittisen yhdistyksen tutkimusten sarjassa. Sen toimittavat Auli Hakulinen, Risto Jaakkola, Kerstin Stenius ja Matti Virtanen.

Alkoholitutkimussäätiö on myöntänyt apurahan kirjan julkaisemiseen.

Jos haluat tilata kirjan, maksa viimeistään toukokuun 10. päivänä 25 euroa tilille FI98 4055 3020 1099 89 (Aktia Bank, Kerstin Stenius) ja ilmoita nimesi ja postiosoitteesi lomakkeella, joka löytyy osoitteesta [www.nordicwelfare.org/klausmakela](http://www.nordicwelfare.org/klausmakela).

Yhteystiedot voi ilmoittaa myös osoitteella Kerstin Stenius, NVC, Topeliuksenkatu 41 a A, 00250 Helsinki. Tilauksesi postitetaan ilmoittamaasi osoitteeseen heti kirjan tultua painosta.