



VAASAN YLIOPISTO

KAUKO MIKKONEN – MARIANNE HAAPALA

Aluetutkimuksen menetelmät

VAASAN YLIOPISTON JULKAISUJA

Opetusjulkaisu 55

Aluetiede 4

VAASA
2006

ALKUSANAT

Aluetutkimuksen menetelmät -opetusjulkaisulla on yhtä pitkä tausta kuin on aluetieteellä Vaasan yliopistossa. Kartografian perusteet ja tutkimusmenetelmät ovat kuuluneet talousmaantieteen opetussisältöön kauppakorkeakoulun alusta, vuodesta 1968 lähtien. Talousmaantiede muuttui aluetieteeksi 16.1.1998. Myös metodikurssien lukumäärä ja nimet ovat tutkintorakenteiden kehityksen myötä muuttuneet, kunnes yleisen suuntauksen mukaisesti päädyttiin lopulta yhteen metodikurssikokonaisuuteen.

Kurssien sisällöt ovat niin ikään aikojen saatossa jonkin verran vaihdelleet. Muutokset heijastelevat aluetutkimuksen tarpeiden ja välineiden kehitystä. Oman sävynsä ovat sisältöön tuoneet metodikursseista vuosien varrella vastanneet opettajat. Tämä opetusjulkaisu on Kauko Mikkosen ja Marianne Haapalan yhteisesti toimittama, vuonna 1997 julkaistun Kauko Mikkosen kirjoittaman ensimmäisen painoksen 2., uudistettu painos. Osittain uutta, Haapalan kirjoittamaa tekstiä on luvuissa 'Tiedon kerääminen', 'Alueyksikön valinta' ja 'Kartografian perusteet'. Muu teksti on pääsääntöisesti ennallaan. Vain käsitteistöä on ajantasaistettu, havaitut painovirheet korjattu ja joitain lähdeviitteitä lisätty.

Aluetutkimuksen menetelmät -kurssi harjoituksineen pyrkii antamaan menetelmälliset perusvalmiudet aluetutkimuksen tekijöille. Yksi kurssi ei kuitenkaan voi olla kaiken kattava. Tästä opetusjulkaisusta on jätetty pois sellaisia menetelmiä, joiden esittely sopii luontevimmin substanssikurssien yhteyteen. Kaikkien menetelmien osalta pätee sama totuus: kuuntelemalla ja lukemalla saa jonkinlaisen käsityksen asiasta, mutta vasta tekemällä oppii!

Vaasa, elokuussa 2006

Kauko Mikkonen

Marianne Haapala

Sisällys

ALKUSANAT	2
JOHDANTO	5
1. Käsitteitä	5
2. Tutkimusprosessi	6
3. Metodikurssin tarkoitus ja rakenne	8
TIEDON KERÄÄMINEN	10
1. Valmiit aineistot	10
2. Kysely ja haastattelu	16
3. Muu kenttätutkimus	23
ALUEYKSIKÖN VALINTA	26
1. Säännölliset matriisit	26
2. Epäsäännölliset alueverkot	27
3. Näkökohtia alueverkon valintaan	30
KARTOGRAFIAN PERUSTEET	31
1. Kartografian (kartan) historiaa	31
2. Suomen karttalaitos	35
2.1. Historiaa	35
2.2. Lehtijako	37
2.3. Karttatyyppejä	39
3. Diagrammit	42
3.1. Diagrammien jako	42
3.2. Diagrammien laatimisessa huomioon otettavaa	46
4. Kartogrammit eli teemakartat.....	49
4.1. Kvalitatiiviset kartogrammit	50
4.2. Kvantitatiiviset kartogrammit	52
4.3. Kartogrammien laatimisessa huomioon otettavaa	61
4.4. Tietokoneavusteinen tilastoteemakarttatuotanto	62

TILASTOLLISIA MITTAUS- JA ANALYYSIMENETELMIÄ	64
1. Mittaaminen ja mitta-asteikot	64
2. Luokittelu	66
2.1. Luokittelutyyppejä	66
2.2. Monimuuttujaisia luokittelumenetelmiä	70
3. Muuttujajoukon kuvaaminen	71
3.1. Tilastolliset tunnusluvut	71
3.2. Normaalijakauma	75
4. Tilastollisen testin valinta	76
5. Korrelaatio- ja regressioanalyysi	79
5.1. Korrelaatio	79
5.2. Järjestyskorrelaatiokerroin	82
5.3. Regressio	83
5.4. Multiregressio.....	87
6. Faktorianalyysi	89
7. Diskriminaatioanalyysi	93
8. Logistinen käyrä	96
ALUEELLISIA MITTAUS- JA ANALYYSIMENETELMIÄ	101
1. Alueelliset tunnusluvut	101
1.1. Sijaintiosamäärä	101
1.2. Alueellisen assosiaation mittaus	102
1.3. Alueelliset keskiluvut	103
1.4. Lorentz-käyrä	105
2. Gradienttianalyysi	107
3. Suodatinkartoitus	110
4. Trendipinta-analyysi	115
5. Potentiaalitekniikka	125
6. Minimimatkasumma	132
6.1. Kartografinen ratkaisu	132
6.2. P-mediaanianalyysi	134
7. Lähimmän naapurin menetelmä	136
8. Shift and share -menetelmä.....	138

JOHDANTO

1. Käsitteitä

Termit **metodologia** ja **metodi** muistuttavat toisiaan siinä määrin, että niitä yleisesti pidetään toistensa synonyymeinä.

- Metodologia on tieteenfilosofian käsite, joka liittyy tieteenharjoittamisen sääntöihin ja menettelytapoihin. Metodologiassa kysytään esimerkiksi: Miten tutkimusta harjoitetaan? Mitä tieteellinen selittäminen on?
- Metodit ovat aineiston käsittelyssä tarvittavia tilastollisia tai muita menetelmiä.

Metodologia ja metodit eivät ole toisistaan riippumattomia. Kuten professori Ilkka Niiniluoto (1999: 23) on asian ilmaissut: ”Kaikilla aloilla metodinen valmeus olennaisesti edellyttää myös riittävää *metodologis-filosofista yleissivistystä*”.

Tässä opetusjulkaisussa on kysymys aluetutkimuksen menetelmistä, ei metodologiasta. Maantieteen ja aluetieteen metodologis-filosofisia teemoja käsitellään erityisesti syventävien opintojen erikoiskurssilla Aluetieteen metodologia.

Tutkimukset voidaan lähestymistapansa perusteella jakaa kahteen erilaiseen tyyppiin, teoreettiseen tutkimukseen ja empiiriseen tutkimukseen.

Teoreettisen tutkimuksen tyypillisiä sisältöjä ovat:

- käsiteanalyysi (esim. käsitteiden alue, paikka, tila, raja pohdinta),
- itsestään selvinä pidettyjen lähtökohtien problematisointi,
- argumentointi, näkemysten perustelu ja niiden pätevyyden arviointi,
- teoreettinen mallintaminen ”kirjoituspöytätyönä”.

Empiirinen tutkimus perustuu tavalla tai toisella kerättyyn aineistoon reaali maailman ilmiöistä ja johtopäätösten tekemiseen kyseisen aineiston avulla. Empiiristä tutkimusta voidaan luonnehtia tutkimusprosessin avulla.

2. Tutkimusprosessi

TUTKIMUSAIHEEN JA -KOHTTEEN VALINTA

- tieteenalan sisäiset painotukset
- laitoksen tai ohjaajan tutkimussuuntaus, painopistealueet
- yhteiskunnan tai käytännön merkitys
- henkilökohtainen kiinnostus
- toimeksiantotutkimus

TUTKIMUKSEN VIITEKEHYS

- teoreettiset yleistyksiset, teorit, mallit
- oma etukäteismalli (*a priori model*)
- aikaisemmat tutkimukset

TUTKIMUSONGELMAN TÄSMENTÄMINEN

- rajaukset: alue, aika, asiasisältö
- tutkimusongelman osittaminen
- hypoteesit

TUTKIMUSAIKINEISTON HANKINTA

- valmiit aineistot
- kysely- ja haastatteluaineistot

AINEISTON KÄSITTELY

- graafiset esitykset, taulukot
- tilastolliset ja muut analyysit

JOHTOPÄÄTÖKSET

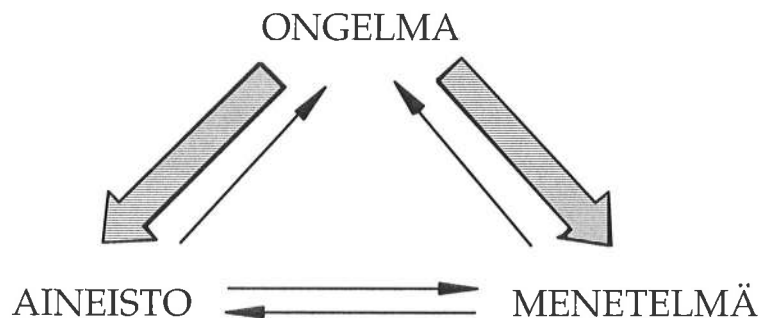
- tulosten suhteuttaminen hypoteeseihin (hypoteesien testaus), viitekehukseen, aikaisempiin tutkimuksiin
- tulosten tiivistäminen, jälkikäteismalli (*a posteriori model*)

RAPORTIN KIRJOITTAMINEN

JULKAISUTOIMENPITEET

Tutkimusongelma on tutkimuksen kriittinen yksityiskohta. Jos ei ole selkeästi muotoiltua ongelmaa ja siihen liittyviä hyviä kysymyksiä, ei voi syntyä kunnollista tieteellistä tutkimusta. Tutkimus on vastauksen tai vastausten etsimistä asetettuun tutkimusongelmaan. Vastauksen etsimiseen tarvitaan aineisto ja menetelmät, joilla aineistoa kuvataan ja käsitellään.

Korostettakoon, että tutkimusongelma määrää viime kädessä menetelmät eikä päinvastoin! Varsinkin niin sanotun kvantitatiivisen vallankumouksen yhteydessä, Suomessa 1960-luvun loppupuolella ja 1970-luvulla, syyllistyttiin myös aluetutkimuksessa tilastollisten menetelmien ylikorostamiseen. Tiettyjen muodikkaiden tilastomatematiikkaisten menetelmien käyttämisestä tuli itsetarkoitus. Tieteellistä kontribuutiota tavoiteltiin uuden menetelmän tai ohjelmiston soveltamisesta ennen kuin muut sen tekevät. Toissijaiseksi jäi sen seikan pohtiminen, millaisten ongelmien ratkaisemiseen mikin menetelmä sopii. Tästä tiedeyhteisöä vaivanneesta "lapsentaudista" on sittemmin palattu normaaliin etenemisjärjestykseen. **Tieteellinen tutkimus käynnistyy ongelman hahmottamisesta. Aineisto kootaan tutkimusongelman mukaan. Sen jälkeen tutkijan on kyettävä valitsemaan kaikista mahdollisista analyysi- ja kuvausmenetelmistä kyseisen ongelman selvittämiseen parhaiten soveltuvat analyysi- ja esitystavat.**



Lähde:

Niiniluoto, Ilkka (1999). *Johdatus tieteenfilosofiaan. Käsitteen- ja teorianmuodostus*. 2. painos. Helsinki: Kustannusosakeyhtiö Otava.

3. Metodikurssin tarkoitus ja rakenne

Metodikurssin ja tämän opetusjulkaisun tarkoituksena on selvittää,

- miten empiirisen aluetutkimuksen aineistoja kootaan,
- miten tietoja käsitellään,
- miten tietoja havainnollistetaan kartografian keinoin,
- miten tietoja analysoidaan tilastotieteen ja korologisen tekniikan menetelmin.

Opetusjulkaisu etenee – kuten tutkimusprosessi – tiedon keruusta aineiston kuvaukseen eli graafisten esitysten laatimiseen ja sitten tilastollisiin ja alueellisiin analyysimenetelmiin.

Empiiriset tutkimukset jaetaan karkeasti kahteen päätyyppiin niissä käytetyn tutkimusotteen perusteella, *kvantitatiiviseen* ja *kvalitatiiviseen* tutkimukseen. Näiden tutkimustapojen eroja on luonnehdittu seuraavan sivun taulukkoasetelmassa. Tässä opetusjulkaisussa analyysimenetelmien pääpaino on kvantitatiivisissa menetelmissä.

On luonnollista, että menetelmiä käsitellään asiayhteyksiinsä liittyen muillakin kursseilla. Esim. SWOT-analyysi on tyypillinen kvalitatiivinen tutkimusmenetelmä, joka sopii aluetieteen opetustarjonnassa luontevasti esiteltäväksi Aluekehittäminen -kurssin yhteydessä, strategisen suunnittelun menetelmänä. SWOT tulee englanninkielen sanoista *Strengths* (vahvuudet), *Weaknesses* (heikkoudet), *Opportunities* (mahdollisuudet), *Threats* (uhat). Liikenneverkkojen graafiteoreettiset mittaukset ja indeksit sisältyvät puolestaan Liikennejärjestelmät ja logistiikka -opintojaksoon, vaikutusalueiden rajausmenetelmät, kaupunkien roolitutkimuksen indeksit, CBD:n (*Central Business District*) rajausmenetelmät yms. kuuluvat Aluetalouden rakenteet -kurssin sisältöön. Poliittisessa maantieteessä käsitellään puolestaan sen tutkimusteemoihin sopivia menetelmiä jne.

Kvantitatiivisen ja kvalitatiivisen tutkimuksen eroja

Kvantitatiivinen	Kvalitatiivinen
Aineisto edustaa tilastollisesti perusjoukkoa	Aineisto edustaa tutkimuskohteen olennaisia piirteitä, on teoreettisesti merkittävä
Aineiston kerääminen, käsittely ja analyysi ovat toisistaan erottuvia vaiheita	Aineiston kerääminen, käsittely ja analyysi kietoutuvat tiiviimmin yhteen
Täsmällisesti rajattu aineisto, joka on esitettävissä lukuina ns. havaintomatriisin muodossa	Aineiston rajat "avoimet" Aineisto verbaalista tai kuvallista
Tyypilliset tutkimusaineistot: koe-, kysely- ja haastattelututkimukset, tilastot, sisällönanalyysi	Tyypilliset tutkimusaineistot: kenttähavainnointi, haastattelutekstit, erilaiset dokumentit, arkistoaineistot, kulttuurin tuotteet
Havaintomatriisin tilastollinen analyysi	Analyysin tavat vahvemmin aineistosidonnaisia ja vähemmän sääntöihin sidottavissa Laadullinen analyysi: diskussianalyysi, narratiivianalyysi jne.
Tilastollisen analyysin taitavuus ja onnistuneisuus	Aineiston kylläntyminen: uudet tapaukset eivät enää tuo uusia piirteitä Aineiston kattavuus: tulkinta ei perustu satunnaisiin poimintoihin
Mittauksen validiteetti ja reliabiliteetti Analyysin toistettavuus: yksiselitteiset luokittelu- ja tulkintasäännöt	Analyysin arvioitavuus: lukijan on mahdollista seurata tutkijan päättelyä
Teoriaa koetteleva	Teoriaa kehittävä

TIEDON KERÄÄMINEN

Tietoa voi kerätä monin eri menetelmin ja monista lähteistä. Tutkimusongelma ratkaisee ensisijaisesti, millaista tietoa on kerättävä. Mutta asiaan vaikuttaa myös se, millaisia tietoja on kohtuullisella vaivalla saatavissa. Joskus ongelman edellyttämien tietojen hankkiminen on niin vaivalloista tai kallista, että on tyydyttävä lähinnä sopiviin, korvaaviin aineistoihin.

Tietolähteet voidaan jakaa karkeasti kolmeen ryhmään: valmiit aineistot, kysely- ja haastatteluaineistot sekä muulla kenttätutkimuksella kerätyt aineistot.

1. Valmiit aineistot

1) Tilastot, niiden laatijat, tuottajat, tietokannat

Tilastot ovat aluetutkijoiden ehkä tärkeimpiä tietolähteitä. Suomessa on erittäin laaja, kansainvälisesti korkeatasoinen tilastotuotanto. Tilastojen tuottamisen päävastuu on Tilastokeskuksella. Tilastokeskuksen lisäksi Suomessa on myös laaja joukko muita tilastojen laatijoita ja julkaisijoita.

a) Tilastokeskus

Tilastokeskuksen toiminta-ajatuksena on yhdistää eri lähteistä saatavat tietoaaineistot tilastoiksi ja tietopalveluiksi yhteiskunnan tarpeisiin sekä edistää tilastojen käyttöä ja kehittää kansallista tilastointia. Standardituotteiden ja -palvelujen lisäksi tilastokeskukselta on saatavissa erilaisiin tarkoituksiin räätälöityjä tilastoaineistoja, selvityksiä ja tutkimuspalveluja. Tilastokeskus on hallinnollisesti valtiovarainministeriön alainen, mutta se vastaa itsenäisesti toiminnastaan ja palveluistaan. Tilastokeskuksen toiminta rahoitetaan pääasiassa valtion budjetista. Muita rahoituslähteitä ovat maksullinen palvelutoiminta ja muu ulkopuolinen rahoitus.

Tietoja Tilastokeskuksen tuottamista tuotteista ja palveluista löytyy muun muassa vuosittain päivitettävästä tuote- ja palvelukatalogista, Tilasto-oppaasta, Tilastokeskuksen kotisivuilta (www.stat.fi) ja Tilastotorilta, joka toimii samalla myös Tilastokeskuksen julkaisujen tilauspaikkana.

Esimerkkejä Tilastokeskuksen tarjoamista tietolähteistä:

Väestölaskennat: 1950, 1960, 1970, 1975, 1980, 1985, 1990, 1995, 2000, ...

Väestölaskentoja tehdään lähes kaikissa maailman maissa. Väestölaskennan tietosisältö määräytyy ensi sijassa aikaisempien laskentojen, olemassa olevan tietotarpeen sekä EU:n ja YK:n suositusten pohjalta. Väestölaskennassa tuotetaan väestön rakennetta, työssäkäyntiä, perhettä ja asumista kuvaavia perustietoja. Tietoja voidaan käyttää muun muassa yhteiskunnan rakenteen ja siinä tapahtuvien muutosten kuvaamisessa ja suunnittelussa. Vuodesta 1990 lähtien väestölaskenta on kerätty kokonaan rekisteripohjaisesti noin kolmestakymmenestä tilastollisesta rekisteristä ja hallinnollisesta aineistosta. Tietoja julkaistaan sekä maksullisina että veloituksettomina taulukkoina, teemakarttoina, julkaisuina ja tilastotietokantoina.

Keskeisiä tilastoja:

STV = Suomen tilastollinen vuosikirja (yleisteos)

SVT = Suomen virallinen tilasto

Tilastokatsaukset, erilliset tilastojulkaisut ja -julkistukset

Lehtiä:

Tieto&trendit. Monipuolista tietoa Suomen taloudesta ja yhteiskunnasta, aiemmin ilmestyneiden Tieto aika ja Economic Trends -lehtien yhdistelmä.

Kuntapuntari. Teemoittaisia asiantuntijakirjoituksia Suomen kunnista ja alueista.

Veloituksettomia verkkopalveluja:

Tilastotietokantoja:

StatFin-tilastopalvelu. Taulukkotietokantamuotoista Suomea koskevaa tilastotietoa eri aihealueilta.

Kuntien tunnusluvut. Kuntien talous- ja toimintatilaston tietoja ja niistä laskettuja tunnuslukuja.

Väestö. Alueittaisia väestörakennetilastoja.

Yritysten rakenteet -tilastopalvelu. Yritysten tilinpäätöstilastot sekä teollisuuden toimialoittaisia tietoja maakunnittain ja kunnittain luokiteltuna.

Muita veloitusettomia verkkopalveluja:

Maailma numeroina. Maakohtaisia rakennetietoja maailman valtioista.

Suomi lukuina. Keskeistä tilastotietoa Suomesta 25 eri aihealueelta, myös kansainvälistä vertailutietoa, ilmestyy myös taskutilastojulkaisuna.

Webstat-verkkotilastot. Osa Tieteen linkkitalo tiedonhaku- ja Tilastokirjaston tietopalvelua, sisältää valikoituja internetin kautta maksutta saatavilla olevia, eri maiden virallisten tilastontuottajien ja kansainvälisten järjestöjen tuottamia tilastotaulukoita ja -tietokantoja.

Maksullisia tilastotietokantoja:

Altika (aluetietokanta). Sisältää yli 700 000 yksittäistä tietosolua jokaisesta Suomen kunnasta useilta aihealueilta pitkinä aikasarjoina. Tilastotietoja voidaan tulostaa paitsi kunnittain myös kuntatiedoista summautuvilla aluejaoilla. Tietojen päivitys tapahtuu välittömästi uuden tilaston valmistuttua. ALTIKA on käytettävissä joko vuosittaiseen käyttöoikeusmaksuun perustuvana palveluna tai haetun tiedon määrään perustuvalla veloituksella.

Astika. Suhdanne- ja taloustilastoja sisältävä aikasarjatietokanta.

Tilastokatsaus. Sisältää noin 1200 keskeistä talous- ja suhdannetilastoaikasarjaa, päivitetään päivittäin.

KunTo. Kunnittaisia toimipaikkatilastoja toimialaluokituksen 5-numerotasolla.

Kaupunki- ja seutuindikaattorit. Kaupunkeja ja toiminnallisia kaupunkiseutuja kuvaavaa kunnittaista tilastoaineistoa aikasarjoina.

Maaseutuindikaattorit. Keskeisiä maaseutualueiden kehitystä kuvaavia tunnuslukuja erityyppisillä aluejaoilla.

Sijoittumispalvelu. Tietoja koulutuksesta, valmistuneiden työllistymisestä sekä alueellisesta sijoittumisesta.

Väestötilastopalvelu. Väestöä koskevaa monipuolista tilastotietoa kunnittain ja kunnan osa-alueittain (suur-, tilasto- ja pienalueittain).

Paikkatietopalveluja:

Paikkatiedot (GIS) ovat tietokonemuotoisia kartta- ja rekisteritietoja kohteista, joiden sijainti tunnetaan. Sijainti voidaan esittää esimerkiksi kohteen koordinaattitai osoitetietojen avulla. Paikkatiedon keskeisiä ominaisuuksia ovat sen riippumattomuus hallinnollisista rajoista ja sen antama mahdollisuus tuottaa hallinnollisia alueita tarkempaa sijaintitietoa. Tilastokeskuksen ohella myös useat

muut rekisterinpitäjät sekä yritykset ja organisaatiot keräävät ja ylläpitävät erilaisia paikkatietoaineistoja.

Tilastokeskuksen ruuduittaisia paikkatietotilastoja on koottu **Ruututietokanta**-tuotteeseen (ruutukoot 250 m x 250 m ja 1 km x 1 km). Tietokanta sisältää väestöä, koulutusta, työssäkäyntiä, rakennuksia ja asuntoja koskevia tietoja. Tietokantaa päivitetään vuosittain. Ruututietokannan lisäksi koordinaattipohjaista tietoa löytyy yritysrekisterin toimipaikkatiedoista. Ruutuaineistoa on saatavissa myös asiakkaan omia tarpeita varten räätälöityinä aineistoina.

Muita palveluja:

AlueOnline. Aluetietoa tarjoava palvelu, joka sisältää suoraan käyttäjän koneelle avautuvia valmiita teemakarttoja ja diagrammeja.

Aluekatsausjulkaisut. Aluetietoon erikoistuneita kunnittaista, seutukunnittaista, maakunnittaista ja lääneittäistä vertailutietoa sisältäviä vuosittain ilmestyviä julkaisuja.

Ohjelmistotuotteita:

SuomiCD. Alueellinen tietokanta, johon on koottu tietoa asukkaista, talouksista, yritystoimipaikoista, rakennuksista, palveluista, tuloista ja ostovoimasta sekä työpaikoista ja työmatkoista postinumeroalueittain ja kunnittain.

Kuntafakta. Kunnittaisia ja kuntatiedoista summattavissa olevia tietoja sisältävä tilasto-ohjelma, päivitetään kaksi kertaa vuodessa.

Euroavain. Perustiedot maailman kaikista valtioista sekä aluetietoa Euroopan maista aina maakuntatasolle asti, päivitetään vuosittain.

Rekisteripalveluja:

Yritystietopalvelu (tarjoaa tietoa Tilastokeskuksen ylläpitämästä yritysrekisteristä sekä julkisyhteisöjen rekisteristä muodostetusta palvelutietokannasta).

Yritysrekisteri on lakisääteisesti ylläpidetty tilastotoimen perusrekisteri, joka sisältää kaikki yritykset – myös yksityiset elinkeinonharjoittajat ja yhteisöt – jotka ovat työnantaja ja arvonlisäverovelvollisia. Tiedot kerätään Tilastokeskuksen omin yritys­kyselyin ja useista hallinnollisista tietolähteistä kuten Verohallinnon rekistereistä. Yritysrekisteristä on saatavilla mm. seuraavia tietoja: yritysten yhteystiedot, toimialat, henkilöstön ja liikevaihdon suuruusluokat, toiminnan

aloittamisaika, oikeudellinen muoto ja omistajatyyppe. Tietoja saa niin yritys- kuin toimipaikkatasolla. Tietojen ajantasaisuudessa esiintyy jossain määrin vaihtelua. Aloittamistiedot tulevat rekisteriin noin kuukauden ja lopettamistiedot 1–5 kuukauden viiveellä. Muut tiedot päivittyvät 1–10 kuukauden viiveellä. Yritys- ja toimipaikkatiedot ovat poimittavissa lukuisten valintakriteerien avulla. Tiedot luovutetaan korvausta vastaan joko tiedostona, valmiina osoitetarroina tai paperitulosteena. Yritysrekisterin tietoja julkaistaan mm. StatFin ja KunTo -tilastotietokannoissa sekä erillisinä tilastoina ja tilastojulkaisuuina (Yritysrekisterin vuositilasto, Aloittaneet ja lopettaneet yritykset, Suomen Yritykset).

Julkisyhteisöjen rekisterin palvelutietokannassa ovat mukana kaikki valtion virastot, kunnat ja kuntayhtymät sekä niiden toimipaikat. Julkisyhteisöjen rekisterin tietosisältö muistuttaa liikevaihtotietoa lukuun ottamatta yritysrekisterin tietosisältöä.

Oppilaitosrekisteri. Kaikkien oppilaitosten osoitteet sisältävä osoiterekisteri.

Tilastokeskus tarjoaa lisäksi suuren määrän erilaisia tutkimus- ja koulutuspalveluja (esim. verkkokoulu) sekä Tilastokirjaston tuottamia tiedonhaku- ym. palveluja. Tilastokeskuksen kotisivujen kautta on myös yhteys moniin kansallisten ja kansainvälisten virastojen, laitosten, yritysten ja yhteisöjen julkaisemiin tilastoaineistoihin.

Esimerkkejä muista tilastojen tuottajista ja julkaisijoista:

b) Suomen virallisen tilaston tuottajia:

Tilastoviranomaiset: Tilastokeskuksen lisäksi Tullihallitus, joka ylläpitää mm. ulkomaankaupan tilastotietokantaa ULTIKAA, STAKES, Maa- ja metsätalousministeriön tietopalvelukeskus Tike, joka tekee tilastoja maaseudusta sekä maa- ja elintarviketaloudesta sekä ylläpitää sähköistä maataloustilastojen tietopalvelua Matildaa.

Tilastoja laativia viranomaisia: Ilmailulaitos, Maanmittauslaitos, Merenkulku-
laitos, Ratahallintokeskus, Suomen ympäristökeskus, Tiehallinto, Työministeriö.

Muita tilastontuottajia: Kansaneläkelaitos, Suomen Pankki.

- c) Muita Suomen julkishallinnon tilastojen tuottajia:
Eri ministeriöt, Energiamarkkinavirasto, Kuluttajatutkimuskeskus, Matkailun edistämiskeskus MEK, Puolustusvoimat, Suomen kuntaliitto, TE-keskukset, Ulkomaalaisvirasto, Valtiokonttori, Valtion taloudellinen tutkimuskeskus VATT, Verohallitus, Väestörekisterikeskus.
- d) Kunnat, kuntainliitot, maakuntien liitot, kauppakamarit, yliopistot, tutkimuslaitokset jne.

Lähde:

Tilastokeskus (2006). *Tilastokeskuksen www-sivut*. Saatavissa 7.4.2006: <http://www.stat.fi/>.

2) Aikaisempien tutkimusten aineistot

Sekundäärianalyysi, kirjallisuus- ja tiedonhaut.

3) Organisaatioiden dokumentit

Vuosikertomukset, raportit ja katselmukset, tilinpäätökset, oikeuden päätökset, yritysten ja organisaatioiden www-sivut jne.

4) Henkilökohtaiset dokumentit

Elämäkerrat, päiväkirjat, kirjeet (esim. siirtolaisuustutkimus).
Lähdekritiikki eli aineiston luotettavuus ja edustavuus.

5) Kulttuurituotteet

Lehdet, elokuvat, mainokset, esitteet, www-sivut.

2. Kysely ja haastattelu

Usein tutkimusongelma on sellainen, että valmiit aineistot eivät tehtävään sovellu tai niitä ei ole käytettävissä. Tällöin tiedot on kerättävä kyselyn tai haastattelun avulla (ns. *survey* -tutkimus). Survey-tekniikkaa käytetään paljon eri tieteenaloilla, erilaisissa mielipidemittauksissa ja kaupallisissa tiedonkeruutehtävissä (gallupit George Gallupin mukaan). Kyselyn ja haastattelun ongelmana tänä päivänä on niiden runsaus, mikä ei voi olla vaikuttamatta käyttökelpoisten vastausten lukumäärään. Yritystiedustelujen haittapuolena on lisäksi vastausrasituksen kohdistuminen eri tavalla eri toimialoilla toimiviin ja erikokoisiin yrityksiin. Suurimmat yritykset (organisaatiot) joutuvat vastaajiksi huomattavasti useammin kuin pk-yritykset, joista yleensä poimitaan otos niiden suuremman lukumäärän vuoksi.

Kysely ja haastattelu ovat peräti yleisiä myös aluetieteellisissä tutkimustehtävissä. Esimerkiksi tietoja kuluttajakäyttäytymisestä (asiointien suuntautuminen, asiointifrekvenssit, myymälän valinnan perusteet), matkailijoiden motiiveista ja profiileista, yritysten sijaintipreferensseistä, logistiikasta, verkostoitumisen tilasta tai kansainvälistymisestä on vaikea saada ilman kyselyä tai haastattelua. Näistä kaikista teemoista on vuosien mittaan kertynyt runsaasti tutkimusesimerkkejä Vaasan yliopiston aluetieteen (ent. talousmaantiede) laitoksella. Mainittakoon kuriositeettina kansallisesti ja kansainvälisestikin ainoalaatuinen, laajoihin postikyselyihin perustuva entisen Vaasan läänin väestön asiointitutkimusten sarja vuosilta 1971, 1980, 1991 ja 2000.

Kyselyn ja haastattelun sekä niiden välimuotojen luonnehdinta:

- | | |
|--------------------|--|
| Kysely | Vastaaja täyttää esim. postitse, sähköpostitse tai kyselypisteestä saamansa lomakkeen ja palauttaa sen pyydetyllä tavalla ilman henkilökohtaista kontaktia kyselytutkimuksen tekijään. |
| Haastattelu | Haastatteliija esittää kysymykset suullisesti ja merkitsee vastaukset muistiin tai ottaa nauhalle. |
| Välimuotoja | Tutkija jakaa lomakkeet ja selittää tutkimuksen tarkoituksen ja kulun, vastaajat palauttavat lomakkeet esim. postitse tai sovitulla tavalla (ns. informoitu kysely). |

Aineisto voidaan koota myös niin, että lomakkeet postitetaan tai toimitetaan vastaajille muulla tavoin etukäteen, ja tutkija kerää vastaukset henkilökohtaisesti.

Kyselyn ja haastattelun välimuotona voidaan pitää myös sähköpostitse suoritettavia standardimuotoista kyselyä vapaa-
muotoisempia haastattelunomaisia kyselyjä.

Kyselylomake, kyselyn toteuttaminen ja jatkotoimenpiteet

Mikäli tutkimus perustuu kyselyyn, sen laatimista voidaan pitää yhtenä tärkeimmistä tutkimuksen tekoon liittyvistä työvaiheista. Kyselylomakkeessa esitetyillä kysymyksillä kerätään tutkimuksen empiirinen aineisto. Siksi kysymysten laatiminen on usein melko työläs ja aikaa vievä prosessi.

Yleisiä periaatteita kyselylomakkeen laatimisessa

Kyselylomakkeen kysymykset esitetään pääsääntöisesti standardoidussa muodossa. Avovastauksisia kysymyksiä käytetään mahdollisimman vähän. Kysymysten selkeä luokittelu helpottaa myös aineiston myöhempää käsittelyä.

Kysymysten luokittelun ja vastausvaihtoehtojen esittämisen kaksi pääsääntöä:

- * tyhjentyvyys = kaikkien tapausten sovittava johonkin luokkaan,
- * poissulkevuus = kukin tapaus saa sopia vain yhteen luokkaan.

Tyhjentyvyyden ongelma voidaan aina välttää kaatoluokalla: "jokin muu, mikä".
Esim. äidinkieli: 1 suomi, 2 ruotsi, 3 jokin muu, mikä _____.

Jos informaatiolähteet jaetaan seuraavasti:

- 1 joukkotiedotusvälineet
- 2 henkilökohtaiset kontaktit
- 3 lehdistö

Onko luokittelu poissulkeva? Ei, sillä lehdistö kuuluu joukkotiedotusvälineisiin, joten vastaaja voi rengastaa sekä 1:sen että 3:sen.

Kysymykset on laadittava niin, että vastaajat ymmärtävät ne samalla tavalla ja ennen kaikkea tutkijan tarkoittamalla tavalla. Vastaajilta ei voi odottaa eikä kysyä mahdottomia. Mikään vaihtoehto ei saa olla muita houkuttelevampi. Kysymyksen asettelun tulee olla neutraali.

Kysymystyyppejä

1) Täsmälliset tosiasiat eli faktat (esim. taustatiedot, usein luokitellaan)

Esimerkiksi ikä tai syntymävuosi, ammatti, koulutus, äidinkieli, sukupuoli

1 nainen 2 mies

2) Arvionvaraiset tosiasiat

Kuinka paljon, kuinka usein, missä/milloin viimeksi, tulotaso;

Voidaan tavoitella tarkkaa arviota jättämällä vastauskohta avoimeksi tai esittää valmiit vaihtoehtoluokat.

3) Asenne- ja mielipidekysymykset

Yleinen tapa on esittää väittämiä, ja vastaus on valittava esim. 5-portaiselta asteikolta (sanallinen muoto):

- 1 täysin samaa mieltä
- 2 jokseenkin samaa mieltä
- 3 ei samaa, ei eri mieltä
- 4 jokseenkin eri mieltä
- 5 täysin eri mieltä

Esimerkiksi yritystoiminnan sijaintitutkimuksissa esitetään lukuisa joukko sijaintitekijöitä, ja vastaajien on arvioitava, kuinka tärkeä kukin tekijä on yrityksen sijoittumisen kannalta. Vastausasteikko on tällöin esimerkiksi:

<i>Sijaintitekijä</i>	<i>Ei lainkaan merkitystä</i>				<i>Erittäin merkittävä</i>		<i>En osaa sanoa</i>
	1	2	3	4	5	6	

Vastausasteikko merkitään tärkeysjärjestyksessä siten, että merkittävin arvosana ilmaistaan korkeimmalla arvolla. Vastauksista lasketaan usein keskiarvoja ja suurempi arvo on helpompi mieltää tärkeämmäksi kuin matala arvo. Vastausten johdonmukaisuutta ja luotettavuutta voidaan kontrolloida esittämällä samasta asiasta kaksi tai useampia samansuuntaisia väitteitä.

4) Avoimet kysymykset

Avoimia kysymyksiä on yleensä helppo laatia, mutta helppous kostautuu poikkeuksetta aineiston purkuvaiheessa.

Kyselylomakkeen kokoaminen

Alkuun helpot kysymykset kuten taustatiedot. Taustakysymyksiksi valitaan tutkimusongelman kannalta oleelliset kysymykset. Joissain tapauksissa taustatiedot on tarkoituksenmukaista sijoittaa lomakkeen loppuun. Myös erityisen tärkeät kysymykset on hyvä sijoittaa kyselylomakkeen alkupuolelle. Loppuun varataan tilaa vapaamuotoisille huomautuksille, lisänäkökohdille ja kiitokselle. Vastaus- ja palautusohjeet sekä yhteystiedot annetaan joko kyselylomakkeessa tai sitä seuraavassa saatekirjeessä. Lomakkeen viimeistelyyn kannattaa kiinnittää huomiota. Viimeistely layout antaa vakuuttavamman kuvan tutkimuksesta ja parantaa vastausaktiiviteettia!

Lomakkeen pituus. Esimerkiksi 10 sivua on vastaajalle liikaa. Pituus riippuu kohdejoukosta, sen koosta ja laadusta, kysymysten vaikeusasteesta ja kyselytilanteesta. Suositeltavia maksimipituuksia: postikyselyssä 4 sivua (A3-arkki taitettuna), nopeassa haastattelutilanteessa 1 sivu.

Koekysely eli lomakkeen testaus on syytä suorittaa aina ennen lopullista käyttöä!

Kyselyn kohdejoukko ja kyselyn toteuttaminen

Kyselyä suunniteltaessa on päätettävä, tehdäänkö *kokonaistutkimus* vai *otantatutkimus*. Kokonaistutkimuksessa on kohteena tutkimuspopulaation kaikki jäsenet. Otantatutkimuksessa tutkimuskohteena on joko edustava otos tai näyte. Edustava otos valitaan tilastomatematiikan kriteereiden perusteella, ja siitä tehtävät johtopäätökset ulotetaan koskemaan koko perusjoukkoa. Edustavan otoksen

poimimiseen on kehitetty monia menetelmiä: tasavälinen poiminta sukunimen aakkosjärjestyksen mukaan, poiminta osoitetietojärjestelmästä, yritysrekisteristä tai vastaavista, tiettynä päivänä syntyneet jne.

Suoraan kohdejoukolle osoitetun kyselyn asemesta vastaajina voidaan joskus käyttää varsinaisen kohdejoukon sijaisedustajia. Esimerkiksi vaikutusalue-tiedusteluissa on väestön edustajina käytetty kansakoulujen johtajaopettajia ja kunnanvaltuuston jäseniä (Palomäki 1963, LTT 1967). Yrityksiä koskevissa kyselytutkimuksissa yrittäjien asemesta tietoja voidaan pyytää esim. elinkeino-asiamiehiltä. Näissä on kyse näytteestä eli valikoidusta kohdejoukosta.

Kyselyn vastaamiseen varataan aikaa noin 2 viikkoa. Vastaushävikki on pyrittävä saamaan mahdollisimman pieneksi, sillä hävikki voi aiheuttaa systemaattista vinoumaa tuloksiin. Hävikkiä pienennetään "karhunkierroksien" avulla. Tavallisesti suoritetaan kaksi karhunkierrosta, joista ensimmäinen voi olla pelkkä muistutus kyselyyn vastaamisesta. Tilanteesta riippuen karhunkierroksia voidaan hoitaa myös puhelimitse tai sähköpostitse. Karhunkierrosten vastausaika on lyhyempi kuin alkuperäisen kyselyn.

Tietojen koodaus, tallennus ja testaukset

Lomake on suunniteltava siten, että vastausten käsittelemistä eli koodattavaa ennen tallennusta on mahdollisimman vähän. Kysymyksistä ja täydentävistä tiedoista laaditaan yleensä **muuttujaluettelo**. Tietojen tallennus tapahtuu jonkin tietokantaohjelman avulla. Tällaisia ohjelmia ovat esimerkiksi *SPSS* ja *FileMaker (& SAS)*. Ohjelmaan laaditaan **tallennuspohja**, jolle tiedot syötetään. Nykyään tietokantaohjelmat ovat sellaisia, että jo niillä voidaan tehdä monenlaisia analyysejä. Huolellisesti tehty muuttujaluettelo helpottaa analyyysien tekemistä. Pitkälle standardoidut lomakkeet on nykyään mahdollista tallentaa sähköisesti suoraan tietokoneen muistiin.

Ensimmäisiä tulosteita ovat yleensä muuttujakohtaiset suorat jakaumat ja minimi-maksimi -arvot. Niiden perusteella on helppo havaita ja korjata tallennusvirheet, suunnitella luokitteluja, ristiintaulukointeja ja muita analyysejä.

Otoksen edustavuus. Otoksen edustavuutta testataan esim. vertaamalla otoksesta laskettuja jakaumia ja tunnuslukuja tilastoista saataviin, koko populaatiota kuvaava-

viin vastaaviin jakaumiin ja tunnuslukuihin. Otoksesta lasketut estimaatit on aina testattava, jotta voidaan päätellä, onko esimerkiksi kahden ryhmän välinen keskiarvojen ero satunnaisuudesta johtuva vai myös perusjoukon tasolla vallitseva. Testejä käsitellään lähemmin kurssin tilastomenetelmäosassa. Testaus kontrolloisen, ettei tehdä liian pienien erojen perusteella liian pitkälle vietyjä johtopäätöksiä.

Tärkeä otantatutkimukseen liittyvä käsite on **varmuusväli** (*confidence interval*).

Esimerkki:

Oletetaan, että meillä on 1000 henkilön suuruinen satunnaisotos. On saatu tulokseksi, että tiettyä puoluetta äänestäneitä oli 20 % vastanneista. Erilaisten matemaattisten operaatioiden jälkeen saadaan selville, että on vain yksi mahdollisuus tuhannesta, että etsitty perusjoukon todellinen tunnusluku olisi kauempana kuin 4,2 prosenttiyksikön päässä otoksesta saadusta luvusta 20 %. Tämä tarkoittaa sitä, että vain kerran (jos kysely tehtäisiin 1000 kertaa) saataisiin tulos, jossa puolueen kannatus olisi alle 15,8 % tai yli 24,2 %. Jos raja asetetaan lievemmin eli otetaan riski erehtyä viidellä kerralla sadasta, saadaan varmuusväliksi $\pm 1,3$ prosenttiyksikköä. Jos esimerkiksi kahden lyhyen ajan sisällä tehdyn otantatutkimuksen tuloksista halutaan tehdä päätelmiä jonkin asian muutoksesta, on aina laskettava varmuusväli, jotta tiedetään, onko muutos satunnainen vai tietyllä erehtymisriskillä todella tapahtunut.

Haastattelun lajit

1 Avoin haastattelu

- strukturoimaton, lähellä keskustelua, syvähaastattelu,
- asioita etukäteen vaikea jäsentää tai muotoilla,
- kun halutaan esille heikosti tiedostettuja tai arkaluontoisia asioita ja kun haastateltavia on vähän,
- käytetään esim. elämäkertatutkimuksissa,
- yleensä tarvitaan useita haastattelukertoja,
- materiaali jäsennetään jälkeinpäin.

2 Lomakehaastattelu

- strukturoitu, lomakkeen muoto on tärkeä,
- kysymyksillä tulee olla sama merkitys kaikille,
- strukturoitu lomakehaastattelu yleensä välttämätön nopeatempoisissa haastattelutilanteissa, esim. myymäläasiakkaiden haastattelu, haastattelu lossilla, laivassa, liikenneterminaalissa jne.
- haastattelutilanne määrää lomakkeen pituuden.

3 Teemahaastattelu

- avoimen ja lomakehaastattelun välimuoto,
- puolistrukturoitu haastattelu kohdennetaan tiettyihin ennalta, määrättyihin teemoihin, joita haastattelussa syvennetään,
- kysymyksiä ei muotoilla etukäteen tarkkaan.

Kyselyn ja haastattelun vertailua

<i>Kysely</i>	<i>Haastattelu</i>
– rajoitettu määrä selkeästi standardeoituja kysymyksiä	– kysymyksiä voi olla runsaammin (riippuu haastattelutilanteesta), avokysymykset yleisempiä
– kohdejoukko voi olla suurikin	– rajoitettu määrä haastateltavia
– taloudellisuus	– kalleus
– kohteiden tavoittaminen helpompaa	– haastateltavien tavoittaminen voi tuottaa vaikeuksia (puh tai tapaam.)
– kyselyn suorittajan henkilökohtainen vaikutus eliminoituu	– haastattelijä voi ohjata vastauksia
– ei vuorovaikutteinen kyselytilanteessa	– vuorovaikutus, ilmapiiri haastattelutilanteessa
– kato voi olla suuri eli puuttuvien vastausten ongelma	– kato yleensä pieni, ei puuttuvien vastausten ongelmaa
– vastaukset voivat olla ylimallaisia	– vastaukset luotettavampia ja laadukkaampia
– ei yleensä ylimääräistä tietoa	– mahdollista saada oheistietoja

3. Muu kenttätutkimus

1) Kenttämittaukset

Maantieteessä on perinteisesti kerätty varsin paljon tutkimusaineistoa kenttämittausten avulla. Tämä koskee erityisesti luonnonmaantiedettä. On menty maastoon ja tehty havaintoja maantieteellisistä ilmiöistä, esimerkiksi pinnanmuodosta, muinaisrannoista, on otettu näytteitä tai tehty mittauksia jääkauden jäljistä kallioperässä, maaperästä, vesistöistä, soista jne.

Maantieteen perinteisin väline kartta perustuu kenttämittauksiin. Aikanaan maastossa mitattiin olennaiset maastopisteet ennen kuin oli mahdollista piirtää kartta. Karttoitukseen on sittemmin tullut yhä kehittyneempiä välineitä ja menetelmiä. Nykyään ilmakuvat, satelliittikartoitus ja paikkatietoja sisältävä Maastotietokanta ovat karttojen laatimisen välttämättömiä apuvälineitä. Maaston kolmiulotteisia muotoja voidaan ilmakuvauksessa tulkita stereokuvauksen avulla. Kenttämittauksia voidaan tehdä myös kartasta. Monenlaisia asioita voidaan saada selville kartan perusteella. Esimerkkejä: asutuksen ryhmittäminen, maankäyttömuodot, teiden ja rautateiden kulku, vesistöjärjestelmät.

On syytä tähdentää, että tekniset edistysaskeleet ja nykyaikaiset apuvälineet ovat suuresti helpottaneet kenttämittausten tekemistä ”kirjoituspöytätyönä”, mutta ne eivät sittenkään kykene täysin korvaamaan omakohtaista havainnointia ”paikan päällä” esim. tutkittaessa maiseman moninaisia ilmentymiä.

2) Ulkopuolinen tarkkailu

Kenttämittausta on omalla tavallaan myös sellainen tilanne, jossa tutkija tekee havaintoja esimerkiksi jonkin kylän elinolosuhteista tarkkailemalla asioita paikan päällä. Havainnoitsija arvioi tutkimuskohdetta systemaattisella, jäsennellyllä tavalla etukäteen laaditun ohjelman mukaisesti ja kirjaa havainnot muistiin analysointia varten. Kohdetta ei irroteta ympäristöstään. Sen sijaan ulkopuolinen tarkkailija voi tehdä havaintoja kohteen ja ympäristön välisistä suhteista sekä vertailla esim. haastatteluissa ilmoitettua ja havaittua käyttäytymistä. Ulkopuolinen tarkkailu voi edeltää osallistuvaa havainnointia. Raja osallistumattoman ja osallistuvan havainnoinnin välillä on usein häilyvä.

3) Osallistuva havainnointi

Tutkija osallistuu ryhmän toimintaan sen yhtenä jäsenenä. Osallistumisen kesto ja syvyys vaihtelevat, mutta se on tietoista ja systemaattista. Tarkoituksena on aineiston keruu eri tilanteista. Joskus tutkijat ovat asuneet jopa vuosikausia tutkimuskohteidensa keskellä ja tehneet koko ajan havaintoja. Vuorovaikutus tapahtuu kohteiden ehdoilla. Tutkija vaikuttaa mahdollisimman vähän tapahtumiin.

4) Toimintatutkimus

Toimintatutkimuksessa (engl. *action research*) tutkijan ja tutkittavan välillä ei ole selvää rajaa. Tutkija osallistuu ja vaikuttaa kohdeyhteisön toimintaan ja tietoisuuteen ja pyrkii yhteisön jäsenten kanssa yhdessä ratkaisemaan ratkaistavaksi aiotut ongelmat.

Tutkimusesimerkki:

Koivuporras, Titta-Liisa (2006). Ihmiset ja pankki samaa maata? Yhteiskuntavastuu pankkisektorilla (käsikirjoitus). *Acta Wasaensia*.

Lähde:

Grönfors, Martti (1985). *Kvalitatiiviset kenttätutkimusmenetelmät*. 2. painos. Porvoo: WSOY.

HARJOITUS: **Kyselylomakkeen laatiminen** (aihe-esimerkkejä)

Aihe 1: Ravintolapalvelujen imago- ja kehittämistutkimus
Kohdeyrityksinä esim. VY:n ravintolapisteet

Kohderyhmä: Ravintoloiden asiakaskunta

Kyselyn avulla kerätään yrityksen tarpeisiin tietoja ravintolan (ravintoloiden) nykytilasta käyttäjien kannalta ja miten ravintoloiden ruokatarjontaa, palvelua, viihtyisyyttä yms. voitaisiin edelleen kehittää. Vastaajien taustatiedot ovat välttämättömiä aineiston analysoimiseksi.

Aihe 2: Aluetieteestä valmistuneiden sijoittuminen työelämään

Kohderyhmä: Aluetieteestä valmistuneet HTM ja KTM -opiskelijat

Kyselyn avulla pyritään selvittämään aluetieteestä valmistuneiden sijoittumista työelämään. Kuinka helppoa/vaikeaa työllistyminen on ollut, vastaavatko työtehtävät saatua koulutusta, miten koulutusta tulisi suunnata työllistymisen helpottamiseksi. Miten työllistyminen riippuu vastaajien tutkinnosta ja muista taustatekijöistä.

Aihe 3: Jonkin kuntakeskuksen palvelujen kehittäminen

Kohderyhmä: Otos jonkin kunnan väestöstä tai kyseisen kunnan kaupoissa ja palvelupisteissä asioivista asiakkaista.

Kyselyn avulla selvitetään, mikä on kuluttajien mielestä kyseisen kunnan palvelujen taso nykytilanteessa, miten palveluvarustus on viime vuosina kehittynyt ja mitä palveluja tai erikoisliikkeitä puuttuu tai tarvittaisiin lisää. Kyselyn avulla pyritään myös saamaan selville, missä määrin kunnassa asioivat tulisivat käyttämään mahdollisesti perustettavien erikoisliikkeiden palveluja.

Yleisiä ohjeita

- * Kyselylomakkeen johdannosta tulee käydä selville, miksi kysely suoritetaan.
- * Kyselylomakkeeseen vähintään 10 kysymystä.
- * Vähintään yksi avoin kysymys ja vähintään yksi asteikollinen valintakysymys.
- * Vähintään yksi kysymys täsmällisistä tosiasioista ja vähintään yksi kysymys asenteista ja mielipiteistä.

ALUEYKSIKÖN VALINTA

Monet ilmiöt näyttäytyvät erilaisina riippuen aluetasosta, jolla niitä tarkastellaan. Tutkijan on tämä tiedostettava ja valittava se alueyksiköiden muodostama verkko (*korologinen matriisi*) tai ne verkot, joilla tutkimuksen kohteena olevaa ilmiötä on tarkoituksenmukaista tarkastella. Periaatteessa vaihtoehtoja on runsaasti, todella vartenotettavia vaihtoehtoja niukemmin.

1. Säännölliset matriisit

a) Kolmioverkko

Kolmioverkko on enemmän tai vähemmän teoreettinen, koska ei ole olemassa valmiita tietoja tällä aluejaolla.

b) Nelikulmioverkko

Maastokartan 1:20 000 (korvautumassa uudella Peruskartalla 1:25 000) koordinaattiruudut 1 km x 1 km ja ruutujen neljännekset muodostavat säännöllisen matriisin. Esimerkiksi vuoden 1985 väestölaskentaan asti Suomen taajamat rajattiin peruskarttaruutujen neljännesten tarkkuudella manuaalisesti, sen jälkeen täysin rekisteripohjaisesti rakennusten koordinaattitietojen perusteella. Vuodesta 2000 lähtien taajamarajaustekniikka perustuu ns. ruutupuskurimenetelmään, koko Suomen yli muodostettuun yhtenäiskoordinaatistossa määriteltyyn 100 m x 100 m ruudukoon. Taajamittaisia tilastotietoja tuotetaan mm. väestörakenteesta, työvoimasta ja rakennuksista. Koordinaattiruuduittain (esim. 250 m x 250 m ja 1 km x 1 km) on saatavissa myös runsaasti erilaista Tilastokeskuksen tai muiden organisaatioiden ja yritysten tuottamaa paikkatietoa (GIS).

Alueverkkona voidaan käyttää myös maastokartta- tai peruskarttalehtien muodostamia erikokoisia matriiseja (esim. Juhani Hult väitöskirjassaan 1966) riippuen tarkastelun mittakaavasta. On tietenkin mahdollista tehdä kuhunkin tutkimustehtävään räätälöity, karttakoordinaatistosta riippumaton ruutumatriisi: esimerkiksi 1:400 000 kartalla 1 cm x 1 cm ruutu on luonnossa 4 km x 4 km; sama alueyksikkö 1:100 000 kartalla on 4 mm x 4 mm (esim. Palomäki & Mikkonen 1971; Vaasan läänin asiointitutkimukset 1971, 1980, 1991 ja 2000).

c) Kuusikulmioverkko

Kuusikulmioverkko on monikulmaisoin geometrinen kuvio (lähinnä ympyrää), jolla alue voidaan peittää aukottomasti eivätkä alueet mene päällekkäin (esim. Reijo Helle väitöskirjassaan 1965).

2. Epäsäännölliset alueverkot eli ns. administratiiviset korologiset matriisit

a) Kuntaverkko

Kunta on tilastoinnin perusyksikkö ja sen mukaisesti suurin osa tilastoista julkaistaan kunnittain. Kuntajako on moniin tarkoituksiin sopivankokoinen ja eniten käytetty alueverkko. Kuntajaosta saadaan yhdistelemällä muita suurempia alueverkkoja, esim. **seutukuntajako, maakuntajako, läänijako**.

Kunnat ovat hyvin erikokoisia pinta-alaltaan. Se voi aiheuttaa tulkintaongelmia kunnan pinta-alaan tai väkilukuun suhteutettuja ilmiöitä kuvattaessa ja analysoitaessa (esim. väentiheys, vähittäiskaupan myynti/asukasluku).

b) Pienalueet

Kuntatasoa tarkempaa tietoa voidaan tuottaa sekä koordinaatteihin perustuvilla aluejaoilla (säännöllinen ruudukko, osa-alueet) että hallinnollisista aluerajoista riippumattomilla aluejaoilla (postinumeroalueet, taajamat).

Rakennusten kotipaikkatunnukset (esim. kunta-, kylä-, kaupunginosa-, rakennus- ja huoneistonumero), lähiosoitteet (esim. osa-alue, postinumeroalue, äänestysalue) ja koordinaattitiedot muodostavat koko maan kattavan osoitetietojärjestelmän perustan. Kotikunta ja siellä oleva tarkka asuinpaikka tallennetaan maistraattien ja Väestörekisterikeskuksen jatkuvasti ylläpitämään lakisääteiseen, valtakunnalliseen väestötietojärjestelmään (entinen henkikirjoitus). Väestötietojärjestelmässä olevat henkilöt ovat liitettävissä rakennus- ja huoneistotunnuksen avulla rakennuksen keskipisteen koordinaatteihin ja sitä kautta yhteiskunnan muihin perusrekistereihin ja erilaisiin paikkatietojärjestelmiin. Väestötietojärjestelmä sisältää osoitetietojen lisäksi myös monenlaista muuta tietoa henkilöistä, rakennuksista, huoneistoista ja kiinteistöistä.

Väestötietojärjestelmässä olevien tietojen rekisteröinti perustuu kansalaisten ja viranomaisten lakisääteisiin ilmoituksiin. Osoitetiedoissa tapahtuvat muutokset tallennetaan väestötietojärjestelmään muuttoilmoitusten välityksellä. Muuttoilmoitus on tehtävä maistraatille aikaisintaan kuukausi ennen muuttopäivää ja viimeistään viikon sisällä muutosta (vakituinen muutto). Myös tilapäisestä asumisesta, joka kestää yli kolme kuukautta on tehtävä muuttoilmoitus. Muuttoilmoitus voidaan tehdä myös puhelimitse tai internetissä täytettävällä lomakkeella. Maistraatit tallentavat tiedot ja muutokset omalta toimialueeltaan.

Valtion paikallishallinto on 1.12.1996 alkaen järjestetty kihlakunnittain, joita on 90. Maistraatit toimivat valtionhallinnon rekisteriviranomaisina yhden tai useamman kihlakunnan alueella. Maistraatit vastaavat alueensa väestötietojärjestelmän lisäksi mm. nimenmuutoksista, siviilivihkimisistä, holhousasioista, kauppa- ja yhdistysrekisteristä, venerekisteristä sekä paikallisesta tietopalvelusta.

Väestötietojärjestelmä on maamme eniten käytetty perusrekisteri. Laajan ja ajantasaisen tietomäärän ansiosta väestötietojärjestelmä on laadukas lähde mitä moninaisimpiin tietotarpeisiin. Viranomais-, yritys- ja yhteisökäytön lisäksi väestötietojärjestelmä palvelee mm. tieteellisen tutkimusten tarpeita tarjoamalla esimerkiksi tilasto-, otanta- ja poimintapalveluja.

Suurin osa kunnista jakaa alueensa **osa-alueisiin**. Kunnan osa-alueet muodostuvat kunnan itsensä määrittelemistä toiminnallisista aluekokonaisuuksista, joita ne käyttävät oman aluesuunnittelunsa pohjana. Osa-aluejako on kunnasta riippuen 1–3-tasoinen hierarkkinen alueluokitus (pienalue, tilastoalue, suuralue). Tilastokeskus ylläpitää koko maan aluejakoja. Kunnilla on mahdollisuus tarkastaa osa-aluejakonsa vuosittain. Tietojen tuottaminen osa-alueittain edellyttää rakennusten koordinaattitietoja.

Kylä on kuntaa pienempi alueyksikkö, jonka rajat määrättiin virallisesti isonjaon yhteydessä (ensimmäiset Suomea koskevat isojakoasetukset vuodelta 1762). Talot merkittiin kylittäin maakirjaan ja myöhemmin maarekisteriin (=maakirjasta kehitetty kaavoittamattomien alueiden virallinen luettelo). Kyläjako on ollut perinteisesti henkikirjoituksen perusta. Nykyisessä väestötietojärjestelmässä kylä/kaupunginosanumero kirjataan osana kotipaikkatunnusta.

Kunta on jaettu **koulupiireihin** siten, etteivät koulumatkat peruskoulun ala-asteeseen yleensä ole 5 km pitempiä.

Vaalipiirijako on eräänlainen poliittinen aluejako, jolla tarkoitetaan vaaleilla täytettävien paikkojen alueellista jakautumista. Suomi on jaettu eduskuntavaaleja varten 15 vaalipiiriin, jotka on määritelty 2.10.1998 annetussa vaalilaissa (Säädös-kokoelma 714/1998). Uutta vaalipiirijakoa sovellettiin ensimmäisen kerran kevään 2003 eduskuntavaaleissa. Presidentinvaalissa ja europarlamenttivaaleissa maa on yhtenä vaalipiirinä. Kunnallisvaaleissa vaalipiirinä on kunta.

Väestötietojärjestelmä on äänioikeusrekisterin, ehdokkaiden tietojen tarkistuksen ja ehdokasrekisterin muodostuksen perusta. Ennakoäänestämisessä äänioikeus tarkistetaan äänioikeusrekisteristä. Äänioikeusrekisteristä valmistetaan vaalipäivän äänestyspaikoille vaaliluettelot niistä henkilöistä, jotka eivät ole äänestäneet ennakkoon. Kunta voi muodostua joko yhdestä tai useammasta **äänestysalueesta**. Äänestysaluejako tarkistetaan tarvittaessa vuosittain.

Postinumeroalueella tarkoitetaan postitoimipaikan toiminta-alueita, joka on yksilöity viiden merkin pituisella postinumerokoodilla. Koska postinumeroalueet ovat hallinnollisista aluejaoista riippumattomia, sama postinumeroalue voi jakaantua useammankin kuin yhden kunnan alueelle. Alueet päivitetään vuosittain.

c) Toiminnalliset aluejaot

Edellä mainittujen alueyksiköiden lisäksi on käytössä monia muitakin alueluokituksia, joiden perusteet on määritelty esimerkiksi väestön sijoittumisen, työssäkäynnin tai elinkeinon mukaan. Näiden ns. toiminnallisten alueiden määrittely perustuu yleensä aluetta koskeviin tilastoihin. Tällaisia pääasiassa kuntapohjaan perustuvia alueluokituksia ovat mm. tilastollinen kuntaryhmitys (kuntien jako kaupunkimaisiin, taajaan asuttuihin ja maaseutumaisiin kuntiin), työvoima- ja elinkeinokeskukset, tilastolliset työssäkäyntialueet ja sairaanhoitopiirit. Yksittäisten kuntien tyypittelyyn perustuvia luokituksia ovat esimerkiksi kalleus-, kantokyky-, kielisuhde- ja kuntamuotoluokitukset sekä EU:n tavoitealueet ja kansallisen kehitysalueen tukialueet.

Lähteet:

Maistraatit (2006). *Maistraattien www-sivut*. Saatavissa 1.6.2006: <http://www.maistraatti.fi/>.

Tilastokeskus (2006). *Tilastokeskuksen www-sivut*. Saatavissa 1.6.2006: <http://www.stat.fi/>.

Väestörekisterikeskus (2006). *Väestörekisterikeskuksen www-sivut*. Saatavissa 1.6.2006: <http://www.vaestorekisterikeskus.fi/>.

3. Näkökohtia alueverkon valintaan

Alueyksikön valinta ei tuota ongelmia, jos tieto on saatavilla vain jonkin suuraluejaon tarkkuudella (esim. itsemurhatilastot, tutkimus- ja tuotekehityspanokset). Siihen on tyytyminen. Tavallisempi tilanne on kuitenkin se, että tutkija joutuu valitsemaan. Valintaan vaikuttavat tarkastelun mittakaava (skaala esim. kunta – koko maa) ja haluttu yleistysaste. Mitä tarkempaan tulokseen halutaan päästä, sitä pienempää alueyksikkökoko on käytettävä. Toisaalta liian hienojakoiseen tarkasteluun mentäessä ongelmia voi muodostua 0-alueista (tilastollisissa laskutoimituksissa) ja tilastoinnin salassapitosäännöistä.

Jos tutkimuksen kohteena on koko Suomi, voi olla järkevää tarkastella ilmiötä kahdella aluejaolla, esimerkiksi lääneittäin tai maakunnittain ja seutukunnittain tai kunnittain. Lääni- tai maakuntatarkastelussa saadaan esiin ilmiön alueellisen vaihtelun yleispiirteet, mutta ne edustavat samalla eräänlaista keskiarvönäkemyttä. Suuraluekohtaiset säännönmukaisuudet eivät kerro mitään alueiden sisäisestä vaihtelusta. Monien ilmiöiden luonne on mosaiikkimainen, ja se paljastuu vasta yksityiskohtaisemmassa aluetarkastelussa.

Liian suurpiirteisestä, keskiarvoistavasta tarkastelusta voi olla jollekin alueelle käytännön haittaa. Esimerkiksi Pohjanmaan maakunnan työttömyystilanne näyttää maakuntien välisessä vertailussa yleensä hyvältä, mutta maakunnassa on myös korkeamman työttömyyden kuntia, jotka eivät maakuntakohtaisessa tarkastelussa tule esiin. Yleispiirteisessäkin tarkastelussa on siis tiedostettava alueen sisäisen vaihtelun mahdollisuus, mikä olisi selvyuden vuoksi myös tuotava esiin esim. kartografisesti ja tilastollisten mittalukujen avulla (keskihajonta ym.). Suurläänijako tuottaa luonnollisesti vielä karkeamman yleistyksen ilmiön alueellisesta vaihtelusta kuin maakuntajako.

Alueyksikön ja aluetietojen käyttöön liittyy myös ns. *ekologisen korrelaation* ongelma. Jos esimerkiksi teollisuustyöpaikkojen suuri osuus ja vasemmiston kannatus luonnehtivat tilastojen mukaan jotakin kuntaa, ei ole oikeutettua päätellä, että kaikki "duunarit" olisivat demareita tai vasemmistoliittolaisia. Osa on, osa ei ole. Toisin sanoen rakenteellisista muuttujista ei saa vetää liian pitkälle meneviä yksilötason johtopäätöksiä.

KARTOGRAFIAN PERUSTEET

Kartografia on kaikille tieteille yhteinen keino kuvata alueita ja ihmisen toiminnan tuloksia maapallolla. Kartografia käsittelee karttojen, kartogrammien ja diagrammien laatimista ja niiden käyttöä.

1. Kartografian (kartan) historiaa

I Kartografian varhaisvaiheet

- Alku ”hämärän peitossa”.
- Varhaisimmin kartoitettuja alueita olivat Babylonian ja Niilin laaksot.
- Vanhin tunnettu kartta Babylonian *Ga Suria* esittävä savitauluille tehty kartta noin vuodelta 2500 eKr.
- Noin vuodelta 500 eKr. tunnetaan yksi maailmankartta.
- Kreikkalaiset, mm. **Eratosthenes** (n. 200 v. ea) ja **Ptolemaios** (n. 100 jKr.), käyttivät jo asteverkkoa ja karttaprojektiota (kuvat 1.1. A ja B).

Ptolemaiosta pidetään länsimaisen kartografian ”perustajana”. Ptolemaioksen **Geographia**-teos sisälsi tarkan kuvauksen siitä, miten kartta tulee laatia (projektioppi), sekä yli 8 000 paikannimeä ja niiden pituus- ja leveyspiirit. Luultavasti hän esitti ne myös kartoilla, mutta karttoja ei ole löydetty. Hänen kirjoituksensa katosivat länsimaailmalta yli 1000 vuodeksi. Alkuperäisteokset tuhoutuivat Alexandrian kirjaston palossa vuonna 391. Teoksen kopioita kuitenkin säilyi arabialaisissa kirjastoissa, ja ne tulivat päivänvaloon vasta 1400-luvulla. Ptolemaioksen kirjoitusten perusteella rekonstruointiin hänen nimeään kantavat kartat, ja niillä oli syvä vaikutus Euroopan maantieteelliseen ja kartografiseen ajatteluun renessanssin aikana. Sitä ennen vallitsi kuitenkin pitkä ”pimeä aika”.

II *Pimeä aika* noin 200–1400 jKr.

- Kiinnostus suuntautui muihin tieteisiin.
- Roomalaiset laativat jonkinlaisia hallinnollisia karttoja ja tiekarttoja käytännön tarpeisiin.
- Tavallisimpia maailmankarttoja olivat ns. ympyräkartat (OT = *Orbis Terrarum*), joissa keskuksena oli jokin tärkeä paikkakunta (kuva 1.1. C).
- Välimeren alueelta laadittiin merikortteja, ns. portolaaneja, joihin oli merkitty rannikot, kaupungit ja kompassisuunnat (kuva 1.1. D).

III *Kehitys alkaa* noin 1400 –

- Kartografian kehityksen perustana oli mittaus- ja painotekniikan kehittyminen 1400–1700 luvuilla.
- Ptolemaioksen *Geographia*-kirjoitukset tulivat esiin ja käännettiin kreikasta latinaksi noin 1410-luvulla.
- Löytöretket avarsivat maailmankuvaa.
- **Mercator** ja hänen oikeakulmainen lieriöprojektiionsa (v. 1559).
- Kartografian monien käyttötarkoitusten tultua tunnetuksi myös valtion sisäisissä kontrollitehtävissä ja sodankäynnissä useat maat ”institutionalisoivat” kartantuotannon. Perustettiin karttojen valmistukseen ja maanmittaukseen erikoistuneita yksiköjä. Muun muassa Ruotsi–Suomessa 30-vuotisen sodan jälkeen 1600-luvun alkukymmenillä **Andreas Bureus** käynnisti sekä ns. geometriset (suurimittakaavaiset) että maantieteelliset (pienimittakaavaiset) kartoitus- ja maanmittaustyöt.
- Ranskan Akatemia perustettiin 1600-luvun lopulla. Sen toimesta suoritettiin mm. meridiaanin kaaren pituuden mittauksia uudella kolmiomittausmenetelmällä Perussa ja Torniossa (Maupertuis’in johtama retkikunta Lapissa 1736). Näillä mittauksilla todistettiin mm. maan litistyneisyys.
- Metrijärjestelmä otettiin käyttöön 1800-luvun alussa.
- Navigaatiokarttojen ja topografisten karttojen rinnalle alkoi ilmaantua temaattisia karttoja.

IV 1900-luvulta eteenpäin

Kartografinen tekniikka on kehittynyt enemmän kuin koskaan aikaisemmin.

Kehityksen syitä:

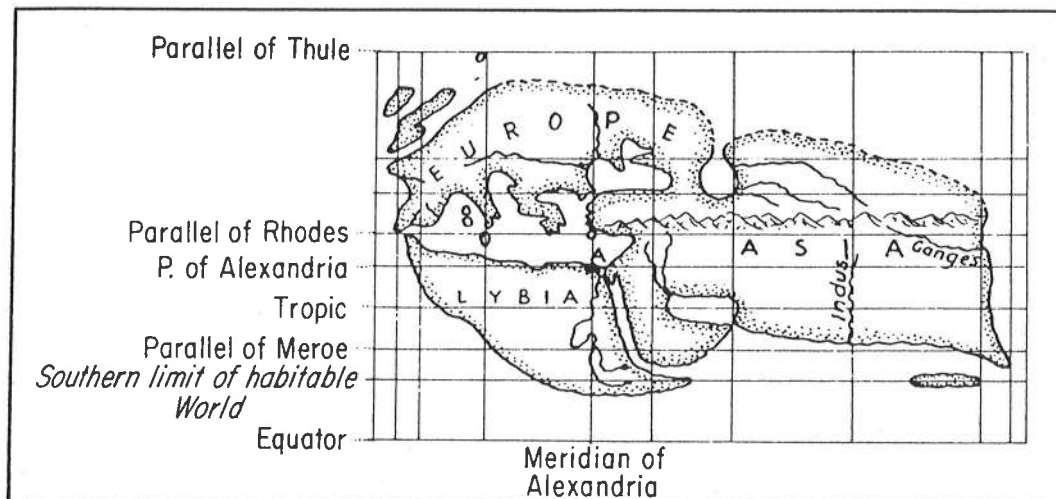
- 1) Kaksi maailmansotaa, tarvittiin karttoja sotilaallisiin tarkoituksiin.
- 2) Väestön nopea kasvu ja tiedon tarpeen lisääntyminen kaikkialla, karttoja mm. suunnittelun tarpeisiin.
- 3) Ilmakuvakartoitus, väärävarikuvat, ortokuvat
- 4) Stereokartoitus
- 5) Satelliittipaikannus (GPS)
- 6) Tietokonekartografian nopea kehittyminen; Maastotietokanta, digitaaliset karttatuotteet yms.

Kirjallisuutta:

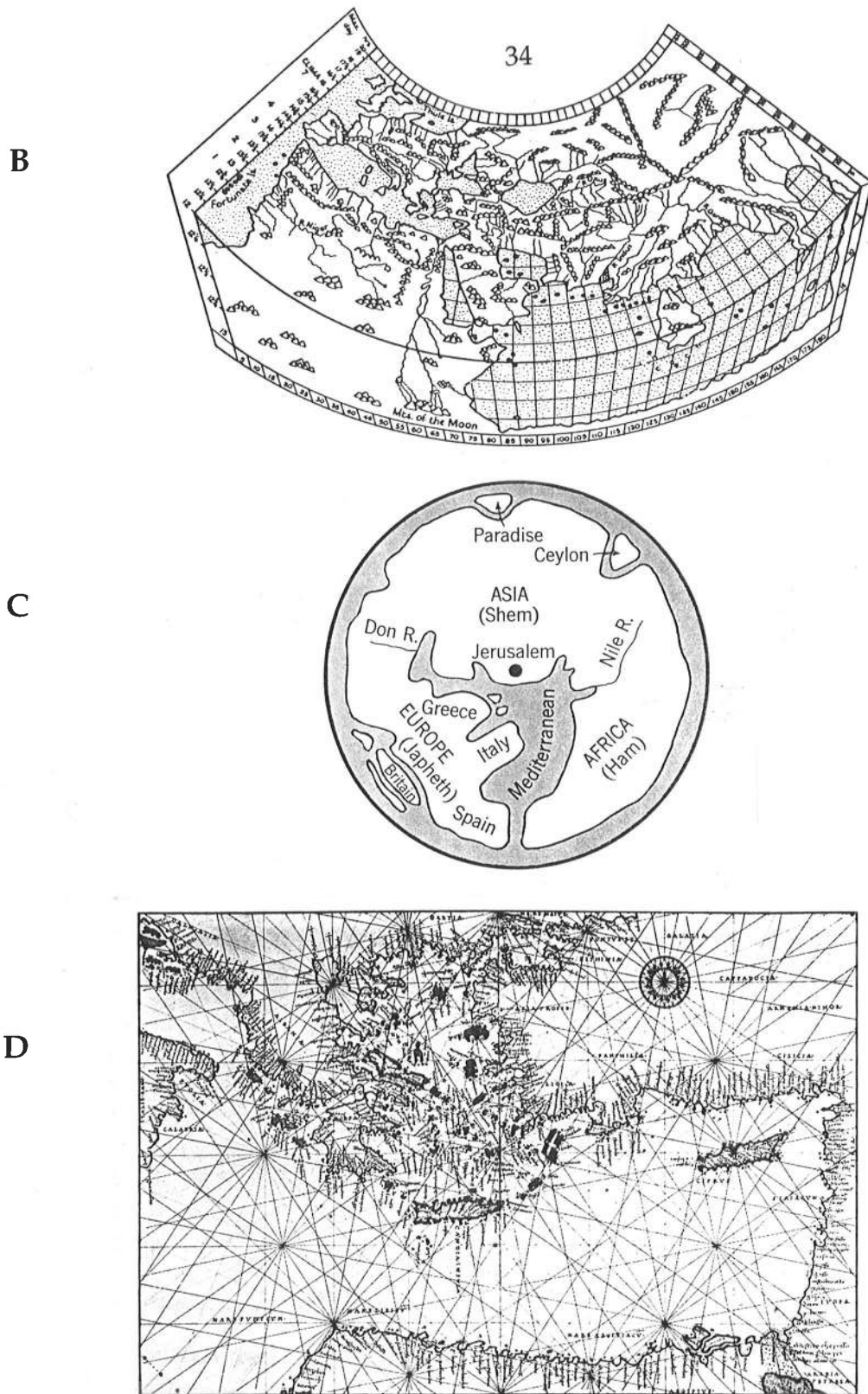
Robinson, A.H, R.D. Sale, J.L. Morrison & P.C. Muehrcke (1984). *Elements of Cartography*. 5th edition. New York etc.: John Wiley & Sons.

Abler, Adams & Gould (1977). *Spatial Organization. The Geographer's View of the World*. London: Prentice-Hall Inc. (s. 61–72).

A



Lähde: Abler, Adams & Gould 1977: 63.



Kuva 1.1. Maapallon klassillisen kartografian (kartan) kehittyminen.
 A: Eratostenesin (270–195 eKr.) maailmankartta (ed. sivu);
 B: Ptolemaioksen maailmankartta (noin 150 v. jKr.);
 C: OT-kartta (*Orbis Terrarum*);
 D: Portolaanikartta Välimeren alueelta (1400-luku).

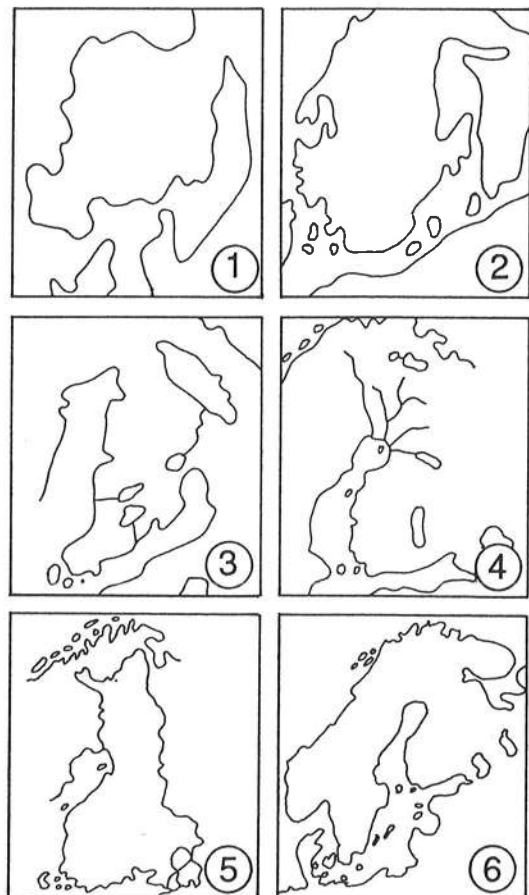
(Robinson ym. 1984; julkaistu John Wiley & Sons Inc.:n luvalla.)

2. Suomen karttalaitos

2.1. Historiaa

Suomi-neidon kehittyminen (kuva 2.1)

- 0) Vanhin Suomeakin koskeva kartta on **Al Idrisin** Maailmankartta vuodelta 1154 (ks. esim. Peltonen 1989: kuva 5). Kartassa Suomen piirteet ovat vielä epämääräiset.
- 1) **Claudius Clavius**; Skandinavian kuvaus vuodelta 1427. Suomi paikantuu kannaksena Skandinavian niemimaan ja "Venäjän" välille.
- 2) **Jacob Ziegler**; Scondia-teoksen tekstiosa ja Pohjoismaiden kartta vuodelta 1532 sisältävät suhteellisen yksityiskohtaisen kuvan Suomesta.
- 3) **Olaus Magnus**; Carta Marina vuodelta 1539 on runsaasti kopioitu karttakuva, joka sisältää historiallis-mystistä kuvitusta.
- 4) **Andreas Bureuksen** kartassa vuodelta 1626 Suomi esiintyy jo lähes nyky-muotoisena.
- 5) **C.W. Gyldenin** vuonna 1853 julkaisema Suomenmaan korkeuskartta 1:1 milj. on maailman ensimmäisiä korkeusvyöhykekarttoja.



Kuva 2.1.

Suomen ja Skandinavian karttakuvan kehittyminen:

- 1 C. Clavius v. 1427
- 2 J. Ziegler v. 1532
- 3 Olaus Magnus v. 1539
- 4 Andreas Bureus v. 1626
- 5 C.W. Gylden v. 1853
- 6 Nykyinen kartta

Muita edistysaskelia

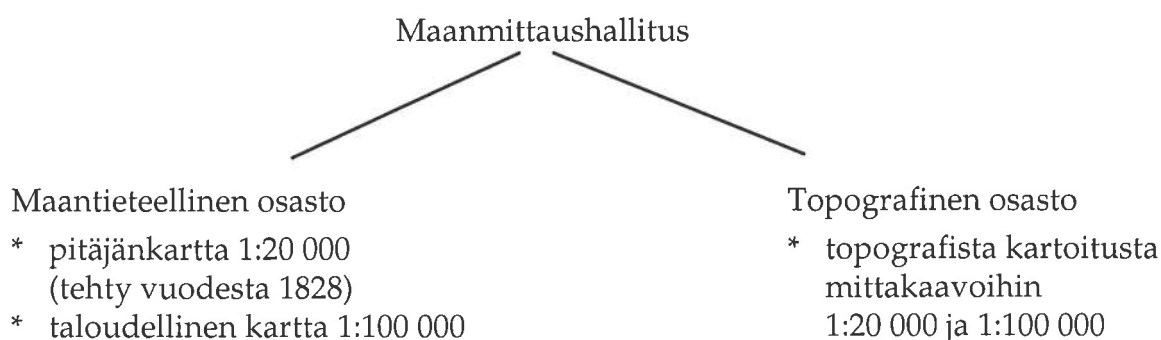
Maupertuis'in kolmiomittaukset Länsi-Lapissa 1736–37.

Päämaanmittauskonttorin perustaminen Turkuun 1812.

Suomen yleiskartta 1:400 000 (30 karttalehteä) valmistui 1873.

Venäjän armeijan topografikunta suoritti kartoituksia Suomessa 1870–1917.

Maamme itsenäistyttyä päävastuu topografisesta kartoituksesta siirtyi siviiliviranomaisille. Maanmittaushallitukseen perustettiin topografinen osasto 1919 siellä ennestään olleen maantieteellisen osaston rinnalle.



Ilmakuvaukseen perustuvien kartoitusmenetelmien käyttöönotto merkitsi kartoitustöiden käännteentekevää nopeutumista.

Vuonna 1940 tuli käyttöön – topografisten ja pitäjänkattojen tekstien ja muiden merkintöjen yhdenmukaistamista tutkivan toimikunnan aloitteesta – **Suomen karttojen yleislehtijako** ja siihen liittyvä karttalehtien numerointijärjestelmä.

Vuonna 1947 pitäjänkartta ja topografikartta yhdistettiin yhdeksi kartaksi = **peruskartta**, ja alkoi **Suomen peruskartoitus**, joka saatiin päätökseen vuonna 1977. Työ ei kuitenkaan päättynyt tähän, sillä karttoja on pidettävä jatkuvasti ajan tasalla ympäristössä tapahtuvien muutosten ja teknisen kehityksen edellyttämällä tavalla. Yksi esimerkki kehityksestä on karttojen tuotannossa käytettävien koordinaattijärjestelmien yhdenmukaistaminen ja uudistaminen.

2.2. Lehtijako

Yleislehtijako

Suomessa on käytetty vuodesta 1940 lähtien maastokarttojen tuotannossa Gaussin-Krügerin karttaprojektion mukaiseen suorakulmaiseen peruskoordinaatistoon vahvistettua **yleislehtijakoa** ja siihen liittyvää karttalehtien numerointijärjestelmää. Karttalehtijakoa täydennettiin 1960-luvulla koko Suomen yli ulottuvalla yhden projektiokaistan (keskimeridiaani 27°) sisältävällä yhtenäiskoordinaatistolla sekä pelastuspalvelulehdillä. Merikartoissa käytetään Mercatorin oikeakulmaista projektiota.

Peruskoordinaatistossa Suomi on jaettu neljään projektiokaistaan, joiden keskimeridiaanit ovat: 21° , 24° , 27° ja 30° itään Greenwichistä. Kukin projektiokaista muodostaa oman koordinaatistonsa, jonka origo on päiväntasaajan ja projektiokaistan keskimeridiaanin leikkauspisteessä. Projektiokaistan pohjoiskoordinaatin arvo ilmaisee pisteen etäisyyden päiväntasaajasta ja itäkoordinaatti etäisyyden keskimeridiaanista. Projektiokaistat on numeroitu 1, 2, 3 ja 4 lännestä itään. Lisäksi Suomen reuna-alueilla ovat käytössä kaistat 0 ja 5. Ensimmäinen numero lehtijaon numerointijärjestelmässä tarkoittaa juuri meridiaanikaistaa. Kukin kaista on jaettu puolestaan 10:een 120 km:n levyiseen osaan, jotka numeroidaan 0, 1, 2, 3, 4, 5, 6, 7, 8, 9. Toinen numero karttojen numerointijärjestelmässä tarkoittaa kyseistä osaa pohjois–eteläsuunnassa. Tämä kaksinumeroinen alue (esim. 20) sisältää 16 karttalehteä 1:100 000 (esim. 2043) jne. (Kuva 2.2; Suomen Kartasto 1984.)

Uusi lehtijako

Yli 60 vuotta käytössä ollutta yleislehtijakoa laajennuksineen on alettu pitää nykyajan karttatuotannossa paitsi ulkoasultaan melko epäyhtenäisenä myös teknisesti vanhentuneena. Tästä johtuen Maanmittauslaitos on siirtynyt käyttämään karttatuotannossaan eurooppalaisen ETRS89 koordinaattijärjestelmän mukaista EUREF-FIN (tarkemmin ETRS-TM35FIN) karttaprojektiota, joka perustuu kansainvälisesti tunnettuun UTM-projektioon (*Universal Transverse Mercator*). Projektiota käytetään yksikaistaisena keskimeridiaanin ollessa 27° itäistä pituutta. Keskimeridiaanin itäkoordinaatti on 8 500 000 m ja pohjoiskoordinaatin nollapiste päiväntasaajalla. Järjestelmä on aikaisempaa kartastokoordinaattijärjestelmää tarkempi ja sopii

yhteen yleismaailmallisten GPS-vastaanottimien WGS84-koordinaattien kanssa. Samalla myös karttalehtijako- ja numerointijärjestelmä muuttuvat (kuva 2.3).

Uudessa UTM-järjestelmän mukaisessa lehtijaossa Suomi on jaettu länsi-itäsuunnassa viiteen (numerot 2–6) ja etelä-pohjoissuunnassa kolmeentoista (tunnukset K–N, P, Q–X) osaan. Kukin näistä yhdestä kirjaintunnuksesta ja numerosta koostuvasta alueesta (esim. V3) sisältää 4 karttalehteä mittakaavassa 1:100 000 (esim. V31), 16 karttalehteä 1:50 000 (esim. V313) ja 64 karttalehteä 1:25 000 (esim. V3133). Vastaavalla tavalla yksi 1:100 000 karttalehti sisältää neljä 1:50 000 karttalehteä ja 1:50 000 karttalehti samoin neljä 1:25 000 karttalehteä.

Koordinaattijärjestelmän ja lehtijaon uudistuksen myötä aikaisemman lehtijaon mukaisten peruskarttojen, jotka on tunnettu vuodesta 1993 lähtien nimellä Maastokartta 1:20 000, sijaan Maanmittauslaitos on vuodesta 2006 lähtien alkanut painaa uutta karttasarjaa, nimeltään Peruskartta, mittakaavassa 1:25 000. Siinä yksi karttalehti kattaa 12 km x 24 km alueen entisen 10 km x 10 km sijaan. Uudistuksen seurauksena painettavien karttalehtien kokonaismäärä vähenee 3 800 lehdestä noin 1 500 karttalehteen. Koko maan käsittävän peruskartaston uudistus toteutetaan 2010-luvun alkupuolelle mennessä.

Vuoteen 2005 asti julkaistu Maastokartta 1:50 000 (= topografinen kartta) korvautuu puolestaan vähitellen uuden lehtijaon mukaisella 1:50 000 kartalla. Kukin karttalehdistä kattaa 24 km x 48 km kokoisen alueen (aiempi 30 km x 40 km). Koska mittakaava pysyy samana, karttasarjan lehtien lukumäärässä ei tapahdu suurta muutosta. Uusi Maastokartta 1:50 000 tuotetaan nopeutetulla aikataululla vuoteen 2008 mennessä lähinnä puolustusvoimien tarpeisiin olemassa olevan kartta-aineiston sekä tiestön ja hallintorajojen osalta ajantasaistetun Maastotietokannan tietoja käyttäen. Samanaikaisesti on kuitenkin aloitettu kokonaan Maastotietokannan pohjalta tuotettavan Maastokartan valmistus. Sen valmistumistavoite noudattelee peruskartoituksen valmistumistavoitetta.

Mittakaavoissa 1:10 000 sekä 1:20 000 tuotettavat karttalehdet Maanmittauslaitos jakaa pääasiassa tiedostoina ja tulosteina. Myöskään mittakaavoissa 1:100 000 ja 1:200 000 tuotettavia karttoja ei tällä hetkellä Maanmittauslaitoksen toimesta julkaista.

2.3. Karttatyyppejä

Maanmittauslaitoksen graafiset ja numeeriset maastokarttatuotteet perustuvat Maastotietokantaan, joka on tarkin valtakunnallinen maastoa kuvaava aineisto. Tietokannassa olevat tiestötiedot päivitetään vuosittain. Muut tiedot ajantasais-tetaan tietojen saatavuuden mukaan vaihtelevasti noin viiden vuoden väliajoin.

Tutustutaan erilaisiin karttatyyppeihin ja lehtijakoon Maanmittauslaitoksen esitteiden, www-sivujen ja aluetieteen karttakokoelman avulla.

Esimerkkejä erilaisista karttatyypeistä:

Peruskartta 1:25 000, Maastokarttaa 1:20 000 ei enää paineta, mutta molempia karttoja on markkinoilla rinnakkain, kunnes uudet kartat on saatu painettua.

Maastokartta 1:50 000, korvaa aikaisemman Maastokartan 1:50 000 (topografinen kartta).

Yleiskartta 1:500 000 (6-osainen).

GT-kartta 1:200 000, Autoilijan tiekartta 1:800 000.

Käyttäjän tarpeen mukaan räätälöidyt karttatulosteet.

Digitaaliset kartta-aineistot (rasteri- ja vektorimuotoiset aineistot, hyödyntäminen paikkatietojärjestelmissä).

Internet-verkossa olevat karttatuotteet esim. Maanmittauslaitoksen Karttapaikka (www.karttapaikka.fi).

Lähteet:

Maanmittauslaitos (2006). *Maanmittauslaitoksen www-sivut*. Saatavissa 8.5.2006.

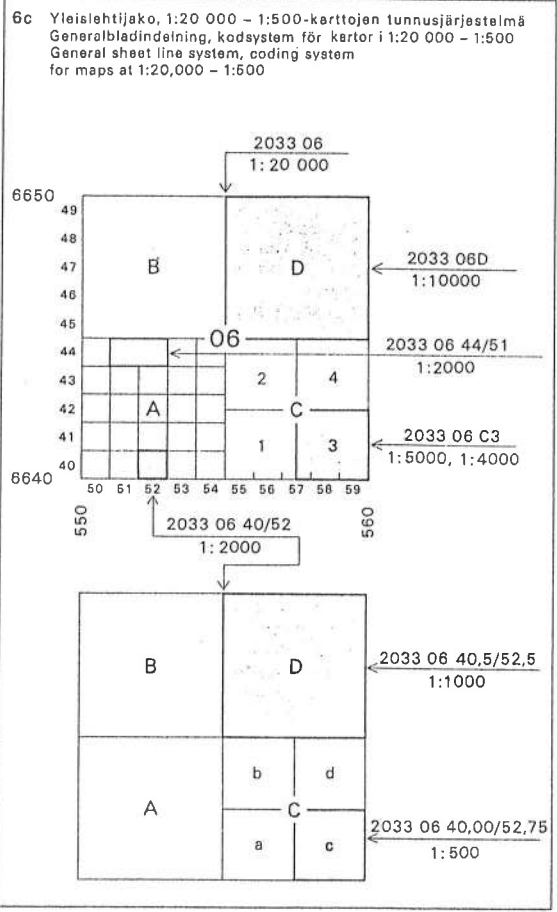
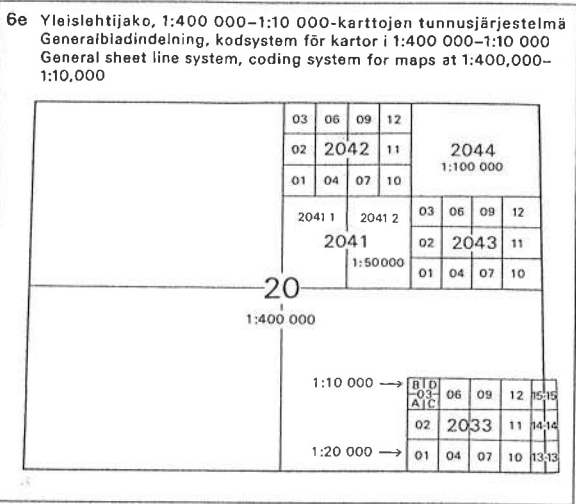
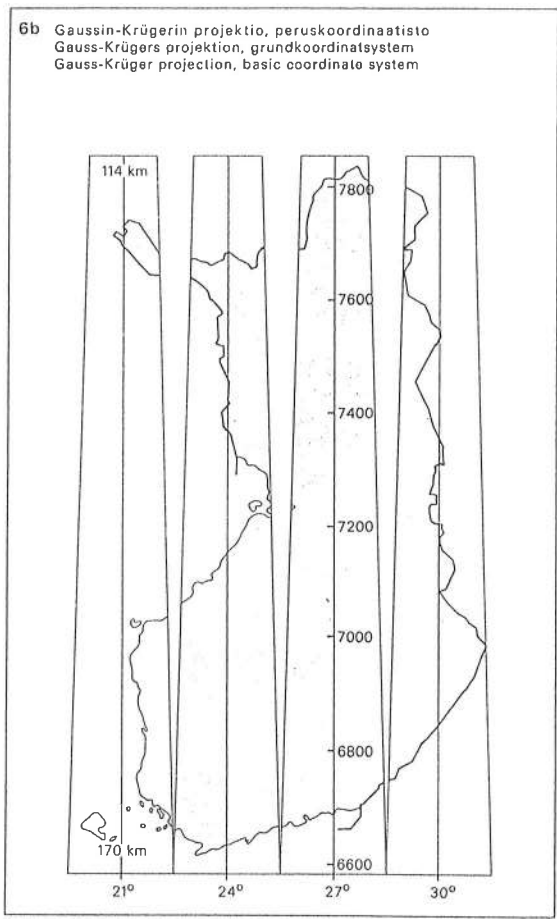
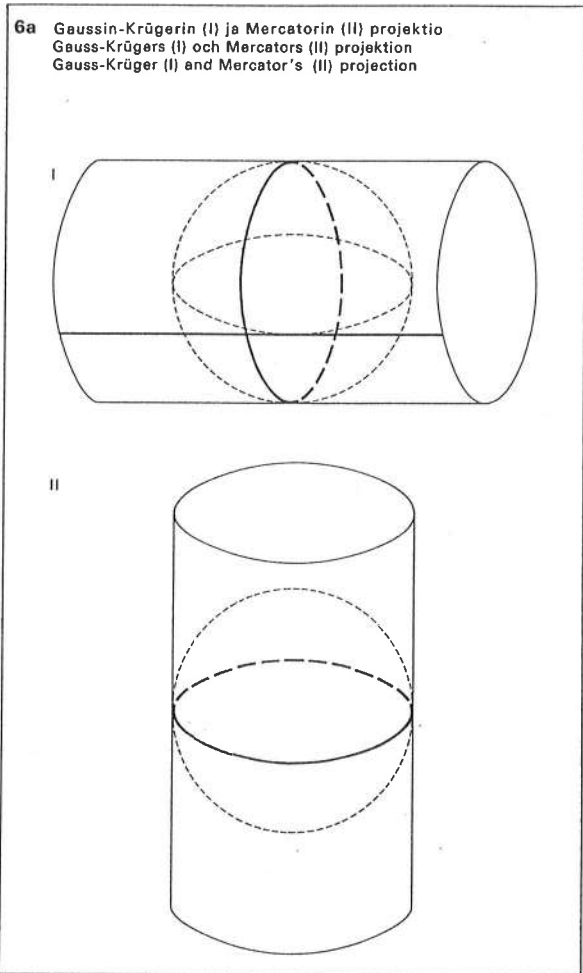
Maanmittauslaitoksen esitteet.

Maantieteen perusmenetelmät (1993). *Helsingin yliopiston maantieteen laitoksen opetusmonisteita* 1. 6. painos. Helsinki.

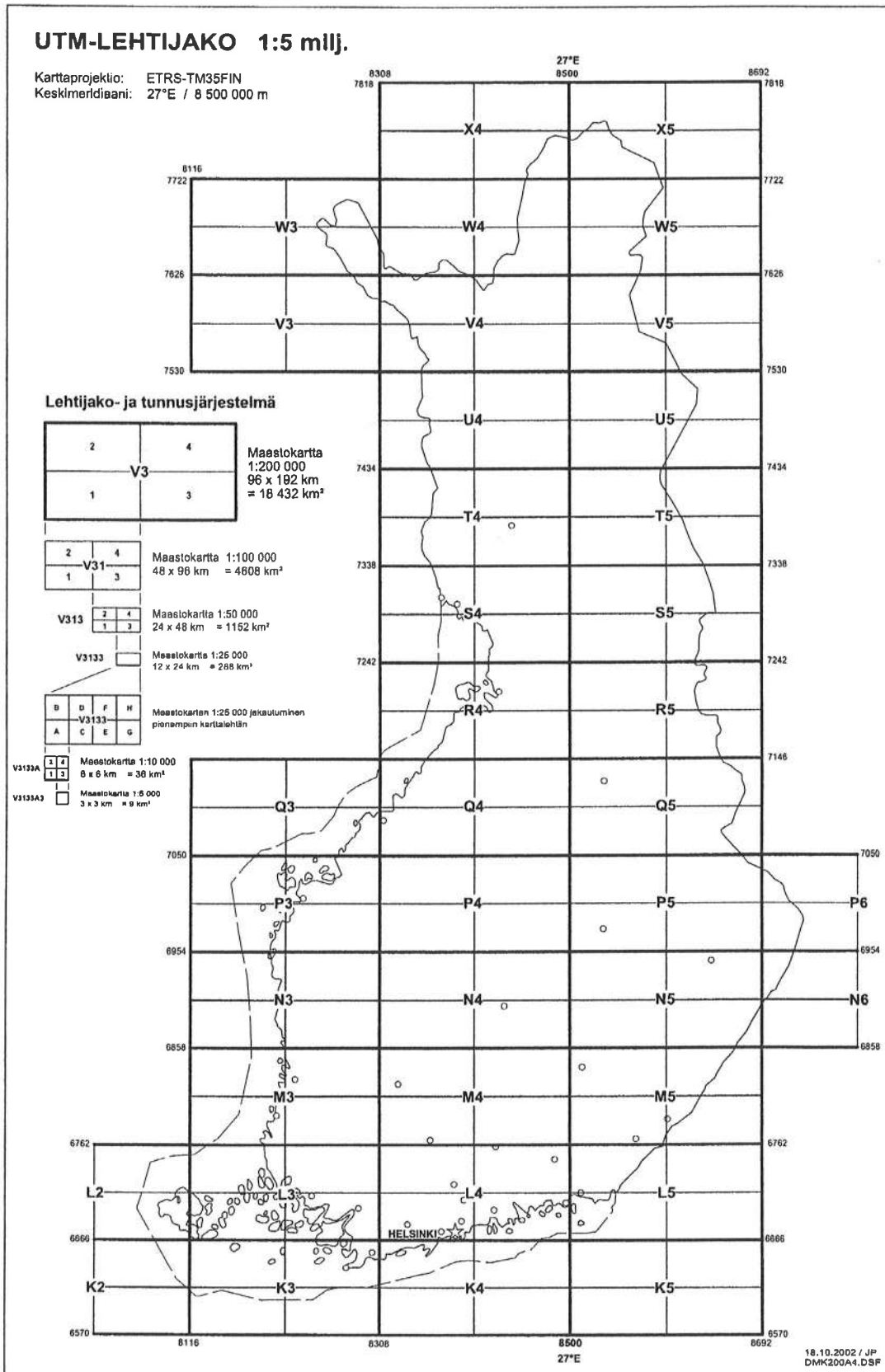
Peltonen, Arvo (1989). *Peruskartografia. Kartografian harjoitustyökurssin moniste*. Toinen painos. *Helsingin yliopiston maantieteen laitoksen opetusmonisteita* 32.

Suomen Kartasto (1984). *Vihko 112: Suomen kartoitus*. Helsinki: Maanmittaushallitus ja Suomen Maantieteellinen Seura.

Tietoa Maasta (2005). *Maanmittauslaitoksen asiakaslehti* 4/2005.



Kuva 2.2.
Suomen karttalaitoksen yleislehtijako.
(Suomen Kartasto 1984;
© Maanmittauslaitos Lupanro 49/Myy/06.)



Kuva 2.3. UTM-lehtijako (© Maanmittauslaitos Lupanro 49/Myy/06).

3. Diagrammit

Diagrammien pääkäyttötarkoituksena on mitattujen tai laskettujen, useimmiten taulukkomuodossa esitettyjen suureiden visualisointi ilman alueellista ulottuvuutta. Yleisesti tunnettu tosiasia, "kuva kertoo enemmän kuin tuhat sanaa", koskee myös diagrammeja mutta vain sillä edellytyksellä, että tilastokuvio on hyvin suunniteltu ja virheettömästi laadittu. Onnistunut tilastokuvio esittää tietoja selkeästi, riittävän tarkasti ja yksinkertaisesti. Huonosti suunniteltu tai toteutettu kuvio voi sotkea katsojaa enemmän kuin välittää tietoa.

Kaikissa tapauksissa kuvio ei välttämättä ole soveliain esitystapa. Erityistä tarkkuutta vaativien ja joskus myös hyvin yksinkertaisten taulukoiden tietosisältö saattaa olla sellainen, että sitä on mahdotonta esittää onnistuneesti kuviona. Esitystavan valintaan vaikuttaa tietosisällön ohella mm. se, kenelle tieto esitetään ja missä yhteydessä se esitetään.

3.1. Diagrammien jako

a) Koordinaatiston mukaan

1 Suorakulmainen koordinaatisto

Kaksiulotteinen, kolmiulotteinen (voi vielä esittää graafisesti).

Jatkuva-arvoinen aritmeettinen (tasavälinen) asteikko, epäjatkuvaa asteikko.

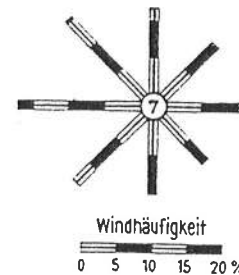
Puolilogaritmiasteikko, jossa toinen akseli (yleensä y-akseli) on logaritminen ja toinen akseli (x-akseli) aritmeettinen.

Kokologaritmiasteikko, jossa molemmat akselit ovat logaritmisia.

Logaritmisessa asteikossa ei ole nollaviivaa eikä negatiivisten lukujen esittäminen ole mahdollista.

2 Suuntaiskoordinaatisto tai polaarikoordinaatisto (kuva 3.1. A)

Esim. viennin volyyymi vuoden aikana, vuorokausilämpötilan vaihtelu; tuuliruuus (viereisen kuvion esimerkissä tuulensuuntien jakauma pylväsdiagrammina).



3 Kolmiodiagrammi

Esim. elinkeinorakenteen kolmijako (kuva 3.1. B).

b) Kuvattavien elementtien lukumäärän mukaan

1 Yksinkertaiset

Kuvattavana on vain yksi asia kerrallaan.

2 Yhdistetyt

Kaksi tai useampia elementtejä samassa diagrammissa.

Viivojen, rasterien tulee erottua selvästi toisistaan.

c) Käytettyjen merkkien mukaan

1 Pistediagrammit

Yhden mitatun ilmiön pisteellä ilmaistu frekvenssi tai määrä.

Kahden mitatun ilmiön x- ja y-akselistoon piirretty pistejoukko = hajontakuvio; yleinen esitystapa korrelaatio- ja regressiotarkastelussa.

2 Viivadiagrammit

Murtoviiva tai interpoloitu, liukuvan keskiarvon menetelmällä tasoitettu tai estimoitu käyräviiva.

Korostaa kehityssuuntaa ja vaihtelua, soveltuu aikasarjojen esittämiseen.

3 Kuviodiagrammit eli piktogrammit

1) Yksiulotteiset eli yksidimensionaaliset piktogrammit

Määrän ilmaisee pylvään, suorakulmion korkeus tai pituus.

Pylväät, suorakulmiot voivat olla pystysuorassa, vaakasuorassa tai polaarikoordinaatistossa.

Pystypylväskuvio korostaa määriä ja määrien vaihtelua; soveltuu aikasarjojen esittämiseen; vaihtoehtoinen viivakuvion kanssa.

Vaakapylväskuvio soveltuu epäjatkuvien luokkien määrän esittämiseen ja vertailuun. Pysty- ja vaakapylväskuviot eivät ole keskenään vaihtoehtoisia esitystapoja.

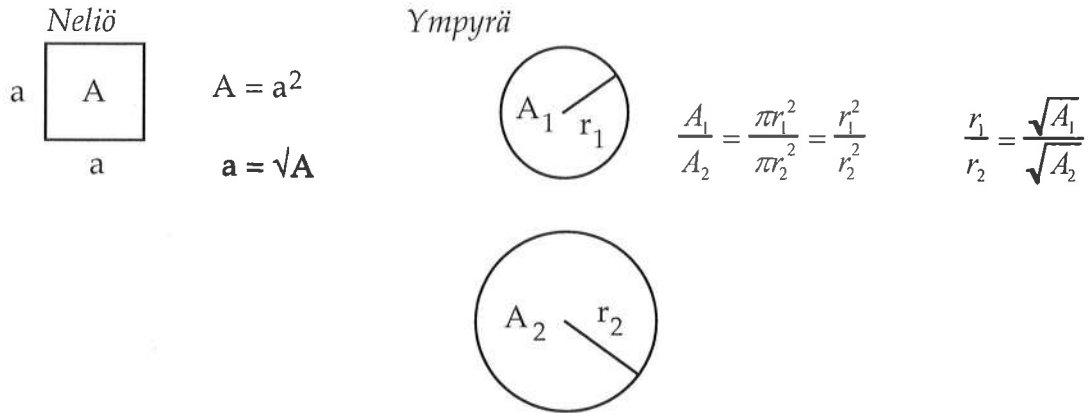
Pylväät, suorakulmiot voivat olla jakamattomia tai jaettuja. Jaetuissa pylväissä tärkein asia esitetään alimpana (pystysuorissa pylväissä) tai vasemmalla (vaakasuorissa). Luokkien selitykset näkyviin.

Kuviodiagrammi on yksinkertainen, havainnollinen esitystapa. Asioiden vertailtavuus on helppoa.

2) Kaksiulotteiset eli kaksidimensionaaliset piktogrammit

Määrän ilmaisijana kuvion pinta-ala.

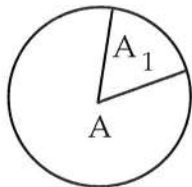
Särmän tai säteen laskeminen:



Ympyrädiagrammien mittakaava on aina muistettava merkitä näkyviin. Nykyaikaiset kartanpiirto-ohjelmat (esim. MapInfo) tekevät sen automaattisesti.

Jaettu ympyrä (piirakkakuviio) kuvaa jonkin kokonaisuuden jakautumista osiin; määrän ja/tai osuuden kokonaisuudesta osoittaa ympyräsektori. Informaatio välittyy sektoreiden pinta-alojen ja niiden suhteiden välityksellä. Koska piirakkakuviossa ei ole varsinaista asteikkoa, se on edellä mainituista kuviotyypeistä epätarkin. Mikäli ympyrän ala on vakioitu, sen edustama absoluuttinen määrä on syytä merkitä näkyviin samoin sektorien edustamat osuudet prosentteina tai absoluuttisina lukuina.

Sektorikulman laskeminen:



a) *absoluuttisista luvuista*

$$\frac{A_1}{A} = \frac{\alpha}{360^\circ} \Rightarrow \alpha = \frac{A_1}{A} \cdot 360^\circ$$

b) *prosenttiosuuksista*

$$\frac{A_1}{A} = \frac{\alpha}{100} \Rightarrow \alpha = \frac{A_1}{A} \cdot 100$$

A = kokonaismäärä,

A₁ = osa tai osuus kokonaismäärästä

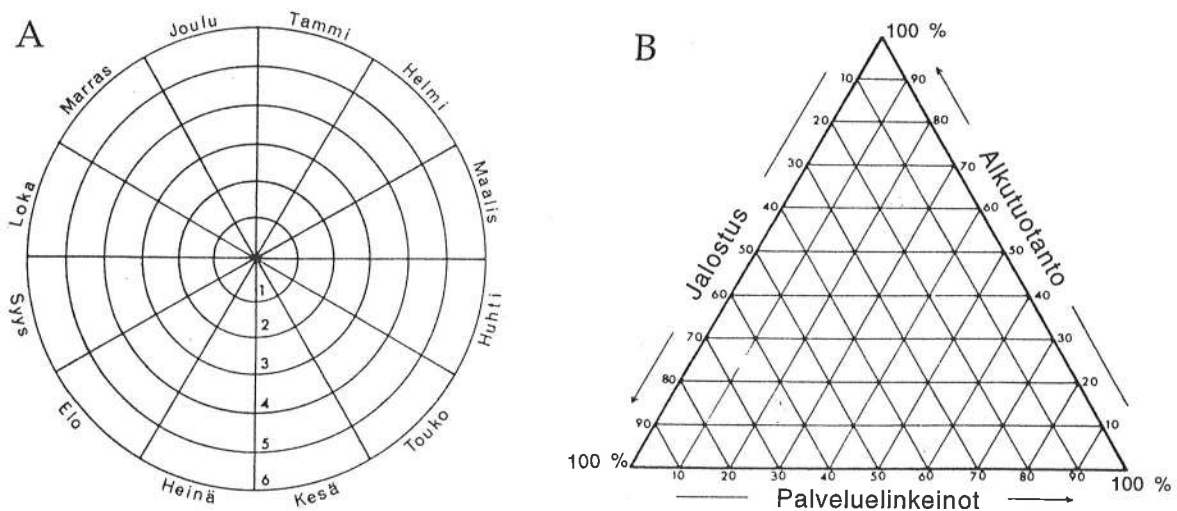
Sektoreiden sijoittelussa on käytössä kaksi eri periaatetta. Vanhemman käytännön mukaan sektorijako aloitetaan klo 12:sta myötöpäivään suuruusjärjestyksessä. Uudemman tyylin mukaan sektorit alkavat klo 3:sta ja kiertävät vastapäivään suuruusjärjestyksessä. Jos ympyröitä on useampia, osatekijöiden keskinäinen järjestys on aina sama. Sektoreita ei saa olla liikaa, enimmillään noin 6. Sektoreiden nimiöt on tarkoituksenmukaista sijoittaa lohkojen viereen, josta ne on helppo hahmottaa suhteessa kokonaisuuteen. Grafiikkaohjelmia käytettäessä piirakkaesitystä ei pidä kallistaa, sillä se vääristää sektoreiden pinta-alasuhteet!

3) Kolmiulotteiset eli kolmidimensionaaliset piktogrammit

Määrän ilmaisijana kuvion (tavallisesti kuutio tai pallo) tilavuus.

Särmän (s) tai säteen (r) laskeminen:

$$r \text{ tai } s = \sqrt[3]{V}, \text{ jossa } V \text{ on kuvattavan asian volyyymi}$$



Kuva 3.1. Polaarikoordinaatisto (A) ja kolmiodiagrammi (B).

3.2. Diagrammien laatimisessa huomioon otettavaa

1. Kuviotyypin valinta

Eri tilastokuviot soveltuvat eri tilanteisiin. Kuviotyypin valinta on tehtävä ensisijaisesti esitettävän asian luonteen perusteella. Siitä, millaiseen esitystilanteeseen kukin kuviotyyppi sopii, on kerrottu lyhyesti kunkin kuviotyypin esittelyn yhteydessä.

2. Asteikon valinta ja merkitseminen suorakulmaisessa koordinaatistossa

Havaintojen ääriarvot ratkaisevat asteikon valinnan. Asteikkoa muuttamalla voidaan saada erilainen mielikuva asiasta. Asteikot on valittava siten, että tuloksena on tasapainoinen kuva.

Perussääntö on, että kaikki asteikot ovat aritmeettisia (tasavälisiä).

Perinteinen suositus: 0-kohta näkyviin (paitsi aika-akselilla ja indeksikuvioissa). Nykyisin tässä suhteessa ei olla enää yhtä ehdottomia.

Logaritminen asteikko (puoli tai koko) valitaan, jos

- a) aineiston jakauma on kovin vino eli paljon pieniä arvoja ja muutama iso arvo,
- b) käyräviivainen jakauma halutaan esim. regressiosuoran estimointia varten "oikaista" suoraviivaiseen muotoon,
- c) halutaan kuvata suhteellista muutosta (esim. taloudelliset ilmiöt).

Asteikko merkitään akseliin lyhyin ulkonevin asteikkoviivoin, ja niiden tulee kattaa koko se alue, josta kuviossa on tietoja. Asteikkoviivojen viereen merkitään asteikon lukuarvot tai nimiöt. Asteikkonumeroinnissa suositellaan kymmenellä (viidellä) jaollisia lukuja. Desimaalilukuja pyritään välttämään.

On merkittävä näkyviin, mitä asteikolla mitataan (akselin otsikko).

Aika-asteikko merkitään yleensä vaaka-akselille ja määräasteikko pystyakselille.

Vaakapylväskuvioissa luokitteleva ominaisuus esitetään pystyakselilla ja jatkuva ulottuvuus (esim. määrä) vaaka-asteikolla.

3. Koordinaatti- eli hilaviivat

Hilaviivojen tarkoituksena on helpottaa kuvion eri osien vertailua. Niitä pitäisi piirtää mahdollisimman vähän mutta kuitenkin sen verran, että kuvion tulkinta on riittävän tarkkaa. Hilaviivat eivät saa erottua liikaa kuvioista.

4. 2–3-ulotteiset kuviot

Valitaan 2-ulotteinen kuvio, ellei aineiston suuri hajonta toisin vaadi. 3-ulotteinen kuvio (esim. pylväs ja ympyrädiagrammi) on erittäin vaikea tulkita oikein.

5. Yleisperiaatteena esityksen selvyys

Kuvion nimiöinnin on oltava selkeä ja riittävän yksityiskohtainen. Epäonnistuneet tekstit saattavat jopa kokonaan vääristää kuvion sanoman. Lyhenteiden käyttöä tulisi välttää. Diagrammin ja kuvatekstin avulla on voitava tulkita asia ilman muuhun tekstiin turvautumista.

Diagrammin havainnollisuutta voidaan parantaa lisäämällä siihen ilmiön vaihtelua selittäviä tekijöitä (esim. aikasarjassa). Samaan kuvioon ei saa ahtaa kuitenkaan liikaa asiaa!

Kuvioiden rastereita valittaessa on kiinnitettävä huomiota siihen, että ne erottuvat selkeästi toisistaan, myös paperille tulostamisen jälkeen. Rasterien aiheuttaman visuaalisen harhan (kallistumat, väreilyt) syntymistä kuvioon on vältettävä.

Vältettävä konstailua kuvioilla, vaikka nykyiset piirrosohjelmat antaisivat siihen mahdollisuuden. Grafiikkaohjelmat eivät vapauta niiden käyttäjää vastuusta.

Kuvion on kerrottava totuus aineistosta!

Kuvion yhdeksän hyvää

E. Tufte (1983) on esittänyt hyvälle tilastokuvioille yhdeksän vaatimusta. Hänen mukaansa hyvä tilastokuvio:

1. **Esittää tiedot kattavasti.** Tämä on tilastokuvion tärkein tehtävä. Muut tehtävät on sopeutettava tähän.
2. **Houkuttelee katsojan ajattelemaan asiaa** eikä menetelmiä, graafista suunnittelua, tuottamistekniikkaa tai jotain muuta vastaavaa. On aina muistettava, että kuvio tehdään *asian* takia eikä tekijän taitojen esittelemiseksi.
3. **Ei vääristä tietoihin sisältyvää sanomaa.** Visuaalinen vertailtavuus on kuvion tärkeimpiä ominaisuuksia. *Valekertoimella* mitataan visuaalisen vertailtavuuden puuttumista. Valekerroin mittaa, kuinka paljon kuvio liioittelee (tai vähättelee) lukujen suhteita.

4. **Esittää paljon lukuja pienessä tilassa** eli kuviossa pyritään maksimoimaan tietotiheys.
5. **Tiivistää suureen tietojen joukkoon sisältyvän sanoman.** Tilastokuvion ehkä tärkein ominaisuus on, että sen avulla on mahdollista tuoda esiin kaoottisessa massassa oleva säännönmukaisuus.
6. **Rohkaisee silmää tekemään vertailuja kuvien eri osien välillä.** Kuviossa pitäisi olla riittävästi apuneuvoja vertailujen tekemiseksi ja koko kuvion rakenteen palvella vertailua.
7. **Tuo esiin tietoihin liittyvän asian monella tasolla.** Kuvion tulisi antaa laaja yleiskuva asiasta ja samalla sen pitäisi olla hienojakoinen esitys yksityiskohdista.
8. **Palvelee kohtuullisen selvästi tiettyä päämäärää:** kuvio voi kuvailla, tutkia, taulukoida tai myös koristaa muuta aineistoa. On huomattava, että koristeidenakin oleva tilastokuvio mielletään lähes aina tietoa välittäväksi.
9. **Muodostaa kiinteän kokonaisuuden muun taulukoissa ja tekstissä olevan aineiston kuvauksen kanssa.** Tiedonvälityksessä kuviolla, taulukolla ja tekstillä on oma tehtävänsä. Kuitenkin kuviona viestitetty sanoma säilyy muistissa pisimpään.

Lähteet:

Tilastokeskus (2006). Verkkokoulu. Tilastografiikka. Saatavissa 15.6.2006: <http://www.stat.fi/tup/verkkokoulu/index.html>.

Tufte, E.R. (1983). *The Visual Display of Quantitative Information*. USA: Graphics Press (lainaus Vesa Kuuselan kirjoituksesta, *Tietoaika* 6/1994, s. 22).

HARJOITUS: Diagrammien laatiminen esim. Excel -ohjelmaa käyttäen

1. Laaditaan viivadiagrammi (yhdistetty) esim. kahden kaupungin väkiluvun kehityksestä.
2. Pylväsdiagrammi kyseisten kaupunkien väkiluvun kasvuprosenteista kymmenvuotisjaksoin.
3. Ympyräsektoriesitys erikseen sovittavasta teemasta, esim. yritysten liikevaihto markkina-alueittain, tuoteryhmittäin jne.

4. Kartogrammit eli teemakartat

Kartogrammien eli teemakarttojen avulla kuvataan kartan aiheeseen eli teemaan liittyviä paikannettavia laatu- tai määrätietoja sijoitettuna karttapohjalle. Teemakartat voidaan jakaa maastoteemakarttoihin (esim. kasvillisuuskartta, maaperäkartta jne.) ja tilastoteemakarttoihin, jotka kuvaavat ilmiöiden paikannettavia tilastotietoja kartografisin symbolein esitettyinä.

Kartogrammit voidaan jakaa erilaisiin tyyppeihin "ristiintaulukoimalla" ilmiön ominaisuus ja havaintojen mittaustarkkuus (kuva 4.1).

	POINT	LINE	AREA
NOMINAL	<ul style="list-style-type: none"> • Town ⌵ Mine † Church BM X Bench Mark 	<ul style="list-style-type: none"> River Road Graticule Boundary 	<ul style="list-style-type: none"> Swamp Desert Forest Census Regions
ORDINAL	<ul style="list-style-type: none"> Large Medium Small 	(Roads) <ul style="list-style-type: none"> Interstate U.S. numbered State County 	Industrial regions <ul style="list-style-type: none"> Major Minor Smoke pollution
INTERVAL - RATIO	REPETITION <ul style="list-style-type: none"> Each dot represents 75 persons GRADUATED <ul style="list-style-type: none"> One-dimensional <ul style="list-style-type: none"> Bars Two-dimensional <ul style="list-style-type: none"> Circles, squares, triangles, etc. 	REPETITION <ul style="list-style-type: none"> Isarithms GRADUATED <ul style="list-style-type: none"> Hachures Flowlines 	Density <ul style="list-style-type: none"> 40 30 20 10 0 Elevation <ul style="list-style-type: none"> 5000 2000 0

Kuva 4.1. Kartografisten symbolien järjestelmä

(Robinson ym. 1984: fig. 7.1; julkaistu John Wiley & Sons Inc.:n luvalla).

4.1. Kvalitatiiviset kartogrammit

Kvalitatiiviset kartogrammit kuvaavat kohteen sijainnin karttapohjalla maantieteellisesti oikeaan paikkaan sijoitettuna ilman kvantitatiivista merkitystä.

1 Ei-kvantitatiiviset yksikkömerkit karttapohjalla

Muodoltaan, kooltaan tai väriltään erilaisin merkein osoitetaan ne pisteet, joissa kuvattava ilmiö sijaitsee.

Merkit voivat olla geometrisia kuvioita, symbolisia kuvioita, kirjaimia, esitettävän kohteen kuvia jne. (valinta tehtävä ajatuksella).

Merkkien lukumäärä voi kertoa esiintymistiheydestä eli määrästä (maantieteellinen jakautuma).

2 Viivakartogrammit

Tyypillisiä viivakartogrammeja:

- maantie-, rautatie-, vesitiekartat (perustuvat ns. nominaalidataan)
- hallintoaluekartat, vaikutusaluekartat

Teiden ja vaikutusalueiden luokittelu (ns. ordinaaliasteikko).

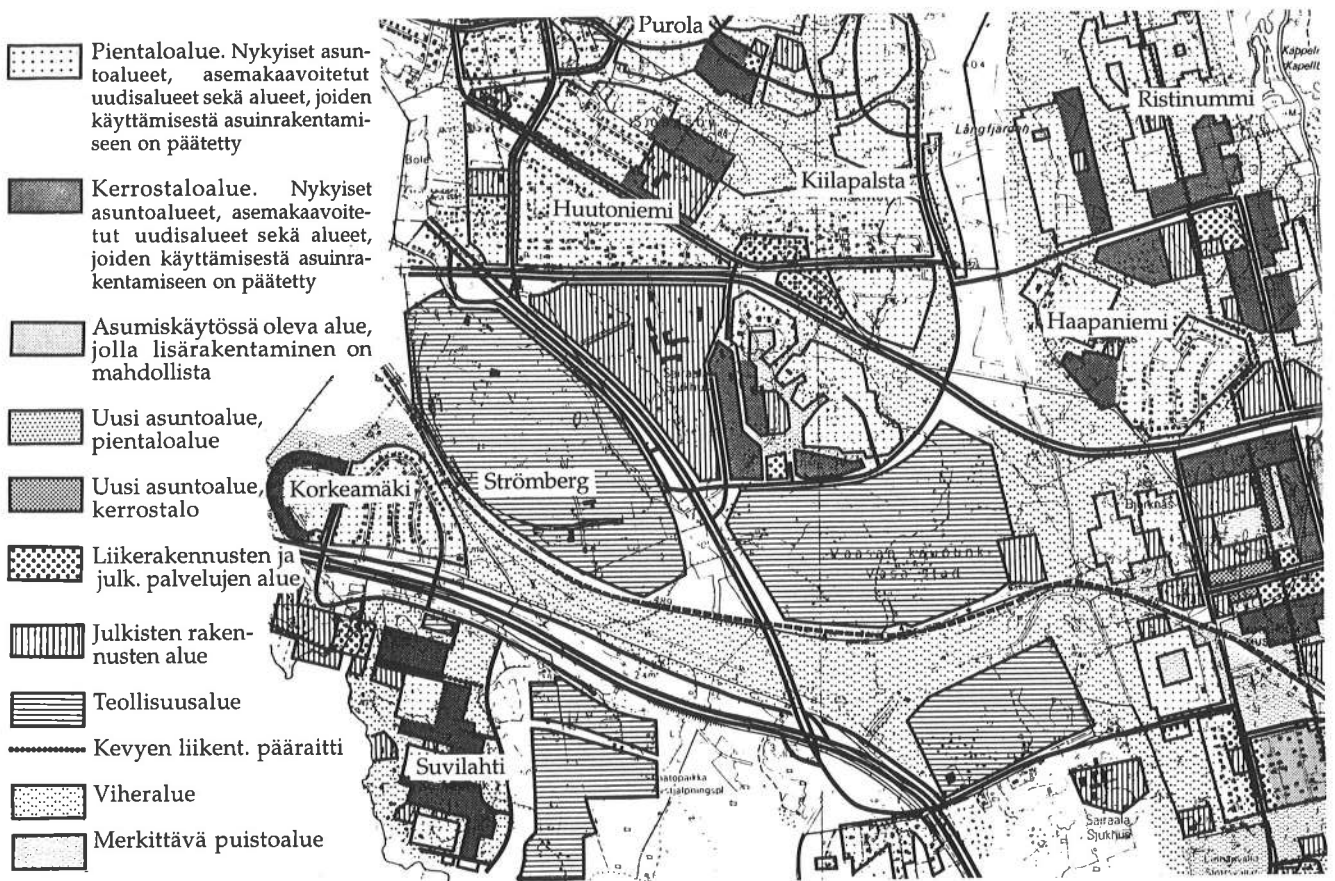
3 Pinta-alatekniikka eli ns. korokromaattinen tekniikka

Kuvataan yhden tai useamman elementin alueellista esiintymistä kartalla:

- esim. Englannin kivihiilikentät (absoluuttiset alueet, tarkasti rajattu), USA:n puuvillan viljelyalueet (relatiiviset alueet, ei tarkkoja rajoja);
- maankäyttö- ja maanluokituskartat (kuva 4.2), kasvillisuustyyppikartta;
- erilaisiin tilastolukuihin perustuva alueluokituskartta (esim. kuntien luokittelu kaupunkimaisiin, taajaan asuttuihin ja maaseutumaisiin kuntiin).

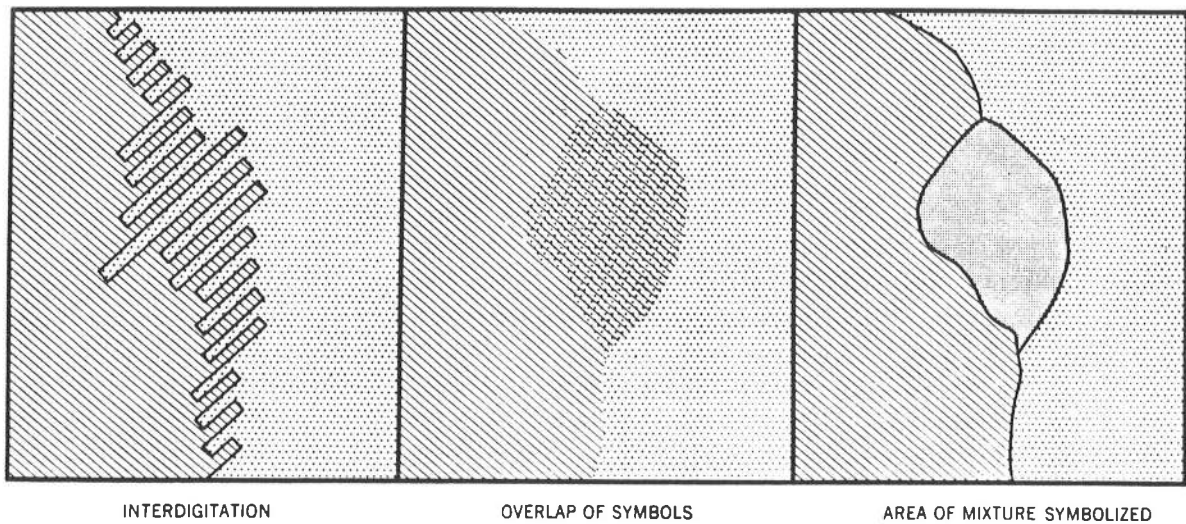
Pinta-alasymbolin (-lien) valinta kuvattavan ilmiön mukaan.

Päällekkäisten levinneisyyksien esittäminen (ks. kuva 4.3).



Kuva 4.2. Esimerkki kvalitatiivisesta pinta-alakartasta: Ote Vaasan yleiskaavasta, suositusmallin maankäyttö.

(Lähde: Yleiskaava, II osaraportti. Vaasan kaupunkisuunnitteluvirasto 1980.)



Kuva 4.3. Esimerkkejä päällekkäisten levinneisyyksien esittämistavoista

(Robinson ym. 1984: fig. 14.6; julkaistu John Wiley & Sons Inc.:n luvalla).

4.2. Kvantitatiiviset kartogrammit

1 Pistekartat (ns. absoluuttinen esitystapa)

Pisteen **yksikköarvo** riippuu esitettävän asian runsaudesta ja kartan mittakaavasta. Vaihtoehdot:

- 1) kukin piste = esim. 100 henkeä,
- 2) kukin piste = esim. 1 % tai 1 ‰ kokonaisväkiluvusta

Pisteen **koko** riippuu yksikköarvosta ja kartan mittakaavasta (kuva 4.4). Pisteiden

sijoittaminen:

- 1) tasaisesti kullekin alueelle (tilastollinen menetelmä),
- 2) mahdollisimman tarkoin oikeille paikoille (maantieteellinen menetelmä). (Kuva 4.5.)

Pisteen sijoittaminen oikeaan kohtaan kartalla ei ole nykyisillä karttaohjelmilla vielä täysin mahdollista tai vaatii erityistaitoja.

2 Kuviodiagrammit eli piktogrammit karttapohjalla (absol. esitystapa)

Karttapohjalle oikeille paikoilleen, alueilleen sijoitetut ympyrät (jaetut, jakamattomat), pylväät, neliöt, pallot ym. symbolit, jotka esittävät alle jäävän alueen arvoja (samat laatimissäännöt kuin diagrammeissa).

Tilastoalueet yleensä hallinnollisia.

Kuvioiden mittakaava valitaan aineiston hajonnan ja kartan mittakaavan perusteella (kuva 4.6). Mittakaava merkittävä aina näkyviin.

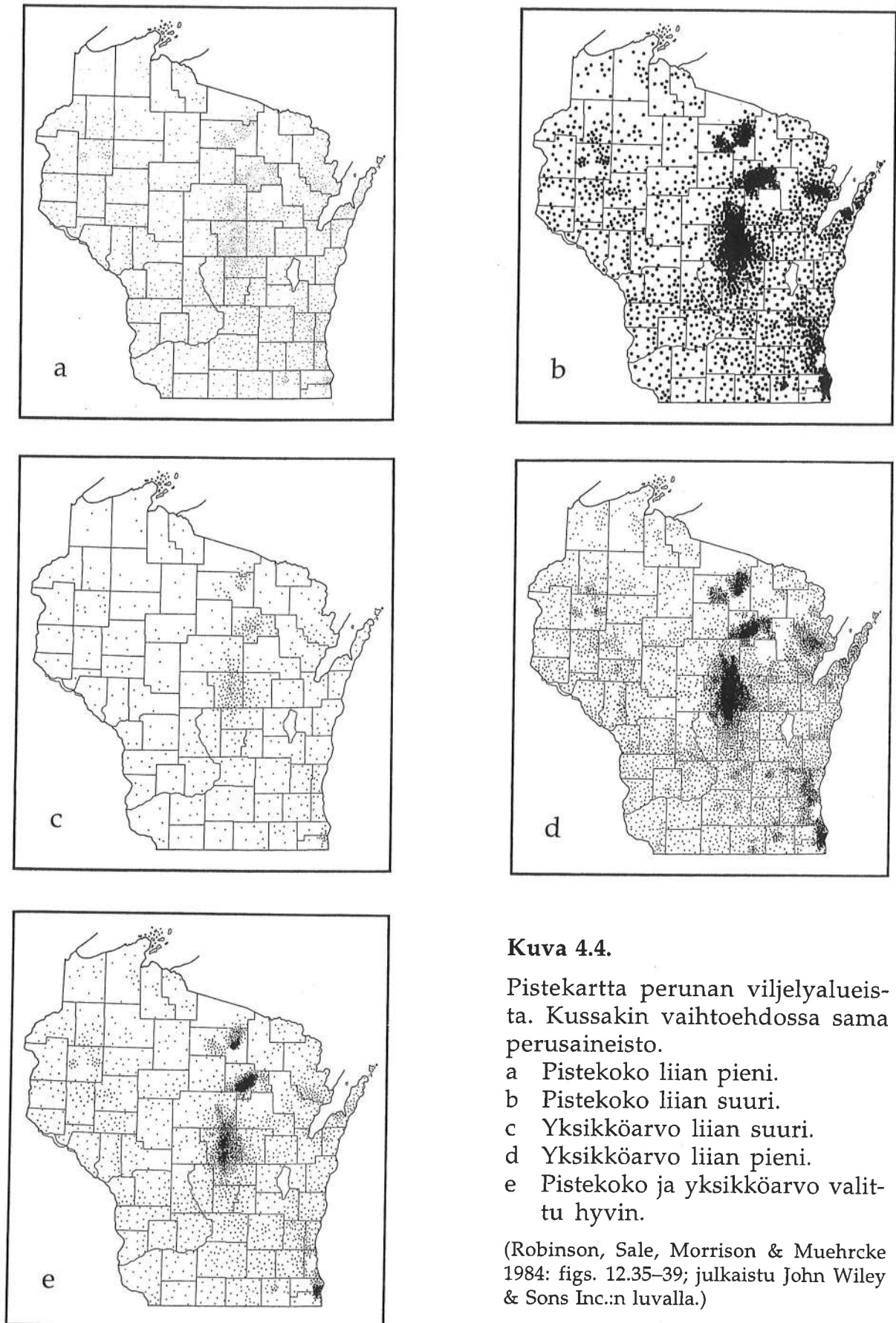
Aineiston suuri hajonta voi aiheuttaa esitysteknisiä ongelmia. Jos alue on pieni tai pienehkö (esim. Hki, Uudenmaan maakunta), mutta volyymit ja siten myös symbolit ovat suuria, ne peittävät alleen ison alueen kartasta.

Päällekkäisten ympyröiden piirtäminen:

- 1) Ympyräkehät kokonaisina näkyviin, jos ympyrät ovat avoimia tai ne on rasteroitu vaalealla rasterilla (kuva 4.7. A).
- 2) 'Sirppitekniikka' tarvittaessa. Leikkaamisen yleisperiaate: pienin päällekkäisistä ympyröistä kokonaisena, isoimmista leikataan mutta niin, että säde ja mahdolliset sektorit ovat vielä nähtävissä (kuva 4.7. B).

Nykyisissä karttaohjelmissa päällekkäisyyksien piirtäminen on pitkälle ennalta määritelty.

Alueiden välinen vertailu ja alueellisen trendin löytäminen ei ole niin helppoa kuin esim. koropleettikartoissa.



Kuva 4.4.

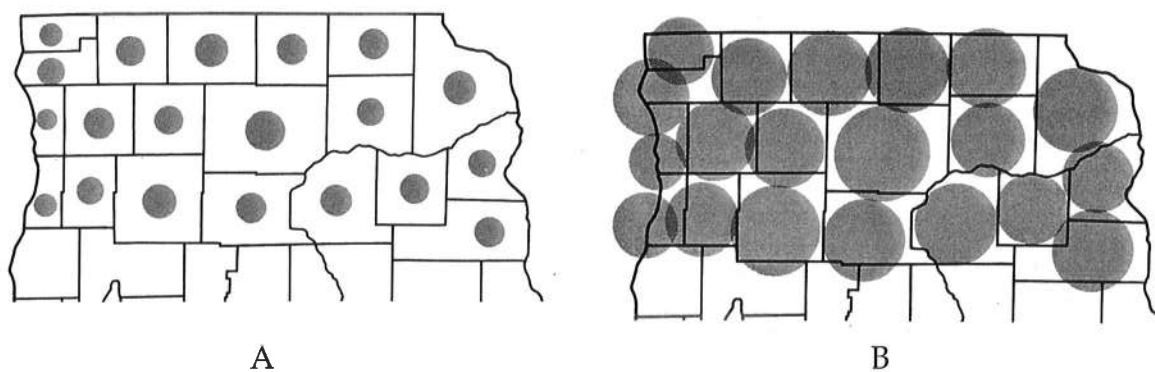
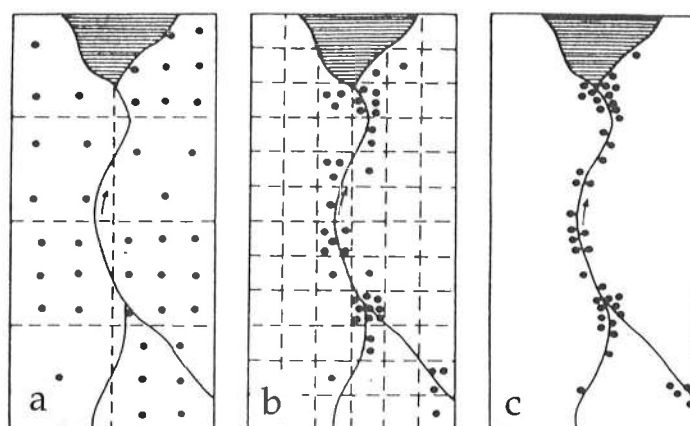
Pistekartta perunan viljelyalueista. Kussakin vaihtoehdossa sama perusaineisto.

- a Pistekoko liian pieni.
- b Pistekoko liian suuri.
- c Yksikköarvo liian suuri.
- d Yksikköarvo liian pieni.
- e Pistekoko ja yksikköarvo valittu hyvin.

(Robinson, Sale, Morrison & Muehrcke 1984: figs. 12.35–39; julkaistu John Wiley & Sons Inc.:n luvalla.)

Kuva 4.5.

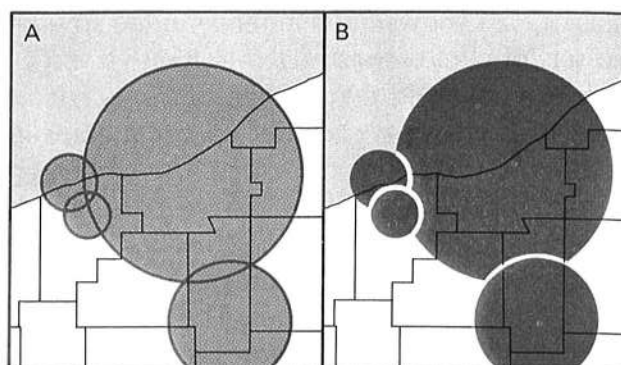
Pisteiden sijoittaminen karttapohjalle tilastollisen menetelmän (a ja b) ja maantieteellisen menetelmän mukaan (c) (Havukkala 1969: kuva 43).

**Kuva 4.6.**

Ympyräkartogrammi: viljanviljelyala maakunnittain. Kuvassa A ympyrän sädeyksikkö on liian pieni, kuvassa B sädeyksikkö on liian suuri; molemmissa esityksissä sama data (Robinson ym. 1984: figs 12.26 ja 12.27).

Kuva 4.7.

Päällekkäisten ympyröiden piirtäminen, kaksi vaihtoehtoista tapaa (Robinson ym. 1984: fig. 12.25; kuvat 4.6 ja 4.7 julkaistu John Wiley & Sons Inc.:in luvalla).



3 Virta- eli nauhakartat

Esitetään maantieteellisen alueen sisällä tapahtuvaa liikettä (esim. muuttoliike).

Viivan/nuolen/nauhan paksuus kuvaa virtausmäärää. Virran suunnan tulee käydä ilmi, ellei virta kuvaa edestakaista kokonaisliikennettä.

Havainnollinen kuvaamaan muutoksen suuntaa, ei niinkään suuruutta (viivan pituus vaikuttaa mielikuvaan).

Asteikko näkyviin: sanallinen, graafinen tai eri tummuusasteet.

Reittiverkko voi olla todellinen tai topologinen (ks. luku 8).

4 Isaritmikartat tai isopleettikartat

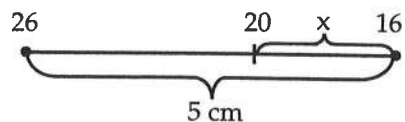
Isaritmi eli samanarvokäyrä yhdistää kartalla toisiinsa ne pisteet, joilla on sama arvo esitettävän asian suhteen.

Termejä, joita käytetään asiayhteyden mukaan samanarvokäyristä: isotermit (lämpötilan yhteydessä), isohyeetit (sademäärien), isobaarit (ilmanpaineen), isobaasit (maankohoamisen), isobaatit (syvyyssuhteiden), isohypsit (korkeuskäyrien), isokronit (saman ajan), isodapanit (samojen kuljetuskustannusten yhteydessä). Termiä isaritm*i* suositellaan yleistermiksi samanarvokäyristä.

Valeisaritmikartoissa yhdistetään sellaiset pisteet, joilla on sama keskimääräinen tms. laskettu arvo.

Laadittaessa tärkeää:

- riittävän tiheä havaintoverkko, joka muodostuu säännöllisistä tai epäsäännöllisistä alueyksiköistä tai havaintopisteistä,
- pisteiden sijoittaminen alueille (tilastollinen tai maantiet. menettely),
- luokkarajojen valinta: aritmeettinen asteikko, esim. 10, 20, 30, ... ;
geometrinen asteikko, esim. 2, 4, 8, 16, ...
- interpolointi, esim. 20-samanarvokäyrä:
- vyöhykkeiden rasterointi
tai värisävyjen valinta.



$$\frac{x}{5} = \frac{20-16}{26-16} = \frac{4}{10} \quad x = 5 \times 4/10 = 2 \text{ cm}$$

Isaritmikarttojen vyöhykkeet eivät risteä eikä vyöhykkeiden järjestys muutu eli kahden vyöhykkeen välissä oleva vyöhyke piirretään ja rasteroidaan, vaikka tähän vyöhykkeeseen kuuluvia havaintoarvoja ei olisikaan (kuva 4.8. C).

5 Koropleettikartat (ns. relatiivinen esitystapa)

Eniten käytetty tilastoteemakarttatyyppi/kvantitatiivinen alueluokituskartta.

Koropleettikartoissa (*choros* = pinta, *plethos* = määrä) on kyseessä aina alueelta laskettu kahden elementin suhde: esim. asukkaita maaneliökilometriä kohti, metsää prosentteina maa-alasta, teollisuusväestöä prosentteina työllisestä työvoimasta jne. Suhdeluvut ja niiden perusteella laaditut karttaesitykset keskiarvoistavat kuvattavan asian kullakin alueella.

Koropleettikartta (kuva 4.8. A) laaditaan aina annettua aluejakoa, tavallisimmin hallinnollisia alueita käyttäen.

Laadittaessa ratkaistava:

- luokkarajojen valinta: aineiston jakauma määrää, luokkarajojen valinnalla suuri vaikutus lopputulokseen; esim. tasaväliset luokkavälit, sama määrä havaintoja, luokittelu keskihajonnan/keskiarvon perusteella, aritmeettiset ja geometriset asteikot, luonnolliset luokkarajat, oma keinotekoinen jaottelu;
- useimmissa tapauksissa 5–6 luokkaa on maksimi;
- luokkien rasterointi tavoitteen ja asian mukaan (saman värin eri sävyt);
- luokkarajojen esittäminen (1–4,9 5–9,9...);
- alueyksiköiden rajojen piirtäminen: 1) kaikki rajat piirretään, 2) vain rajat, joissa luokka muuttuu, 3) ei lainkaan rajoja näkyviin;
- koropleettikartan havaintoharhat: pinta-alaharha, alueharha, rajaamisharha.

6 Ruutukartat

Ruutukartassa kuvattava alue on jaettu vakiokokoiisiin, hallinnollisista aluejaoista riippumattomiin ruutuihin. Ruutujen koko voidaan määritellä esitettävän asian mukaan (esim. 250 m x 250 m, 500 m x 500m, 1 km x 1 km). Ruutukartat mahdollistavat siten hyvinkin tarkan tiedon esittämisen karttapohjalla. Ruutukartoilla esitetään yleensä absoluuttisia määriä kuvaavia arvoja. Relatiivistenkin arvojen esittäminen on mahdollista kuten koropleettikartalla, mutta tällöin saatetaan törmätä ruutuaineiston esittämisessä yleisiin ongelmiin eli tietosuojan ja satunnaisvaihtelun aiheuttamaan ruuduttaisten tietojen ylikorostumiseen. Ruutukartan laatimisessa (kuvattavien luokkien, rasterien valinta) noudatetaan samoja periaatteita kuin koropleettikartan laatimisessa.

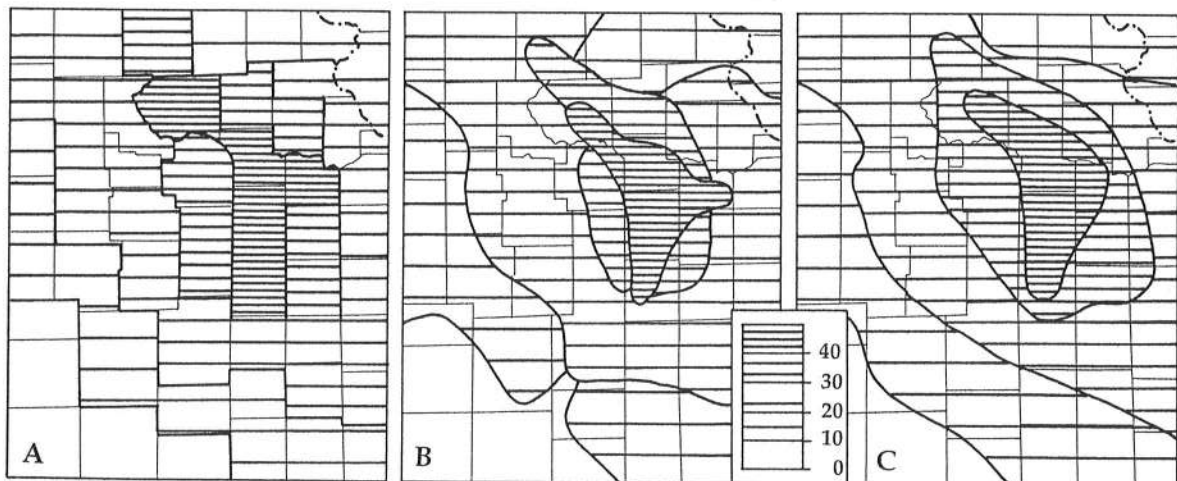
7 Dasymmetriset kartat

Sama rasteripinta tai värisävy kattaa samat lukuarvot tai samaan luokkaan kuuluvat havaintopisteet. Dasymmetrisen kartan tietyn rasteri- tai väripinnan saavat alueet määräytyvät aineiston perusteella! Alueet eivät siis ole etukäteen annettuja kuten koropleettikartoissa. Vyöhykekartoista poiketen dasymmetrisissä kartoissa naapurialueiden rajat voivat ristetä, eikä naapurialueiden tarvitse kuulua perättäisiin luokkiin (kuva 4.8. B).

Dasymmetristä esitystapaa käytetään mm. erilaisissa tiheys- ja tehokkuuskartoissa.

8 Yhdistelmätemakartat

Edellä käsiteltyjä esitystapoja käytetään usein yhtä aikaa samoissa kartoissa. Esim. piste- ja ympyrätekniiikan yhdistelmä on sopiva esitystapa väestökarttoja laadittaessa. Maaseutualueiden asutus esitetään pistein ja taajamien väkiluvut ympyröillä. Yleistä on, että isopleetti-, koropleetti- tai dasymmetrisen esitystavan "päälle" sijoitetaan pistemäisiä tai viivaelementtejä. Niiden tehtävänä on lisätä informaatiota tai selittää alkuperäisen ilmiön alueellista vaihtelua.



Kuva 4.8. Kolme kvantitatiivista pinta-alaesitystä: A. koropleettikartta, B. dasymmetrinen kartta, C. isopleettikartta
(Robinson ym. 1984: fig. 14.8; julkaistu John Wiley & Sons Inc.:n luvalla).

Kirjallisuutta:

Havukkala, Jaakko (1969). Kartografia. Helsinki: KY:n Kirja- ja paperikauppa.

Monkhouse, F. J. and H. R. Wilkinson (1973). *Maps and Diagrams*. London: Methuen & Co Ltd.

Robinson, A. H., R. D. Sale, J. L. Morrison & P. C. Muehrcke (1984). *Elements of Cartography*. 5th edition. New York etc.: John Wiley & Sons, Inc.

Tilastokeskus (2006). Verkkokoulu. Tilastoteemakartat. Saatavissa 15.6.2006: <http://www.stat.fi/tup/verkkokoulu/index.html>.

9 Topologinen esitystapa

Topologia juontaa juurensa 1700-luvulle, mutta sitä on kehitetty erityisesti 1800-luvulla. Topologia eroaa euclidisestä geometriasta siinä, että sille ei ole olennaista ilmiöiden etäisyys ja oikea muoto vaan järjestys ja naapuruus. Esimerkiksi maanalaisten ja linja-autoyhtiöiden reittiverkot pysähdyspaikkoineen esitetään tavallisesti topologisten karttojen avulla.

Topologian käsitteistöön kuuluvat termit *transformaatio* ja *ekvivalenttiset kuviot*. Kun euclidisessä geometriassa venytetty neliö ja eläinhahmo (ks. kuva 4.9) eivät enää ole ekvivalentteja alkuperäisen neliön kanssa tai keskenään, topologiassa ne ovat!

Topologisen kartan konstruointi

- a) Tietyn alueen erisuuruiset osa-alueet (esim. kunnat, valtiot) piirretään yhtä suuriksi säilyttäen alkuperäiset naapuruudet.
- b) Osa-alueiden pinta-alat suhteutetaan kuvattavan ilmiön määrään, esim. väkilukuihin, tulojen määrään, BKT:hen tms. (kuva 4.10).

Kritiikkiä

- topologiset transformaatiot voidaan piirtää monella eri tavalla; kaksi transformoijaa tuskin päätyy samaan lopputulokseen, vaikka tehtävä on sama,
- topologinen kartta on epätavallinen, useimmille vieras,
- topologinen kartta on työläs laatia,
- tapauksiin, joissa oikea mittakaava ja suunta ovat oleellisia (esim. merikortit), topologinen esitystapa ei sovi lainkaan.

Etuja

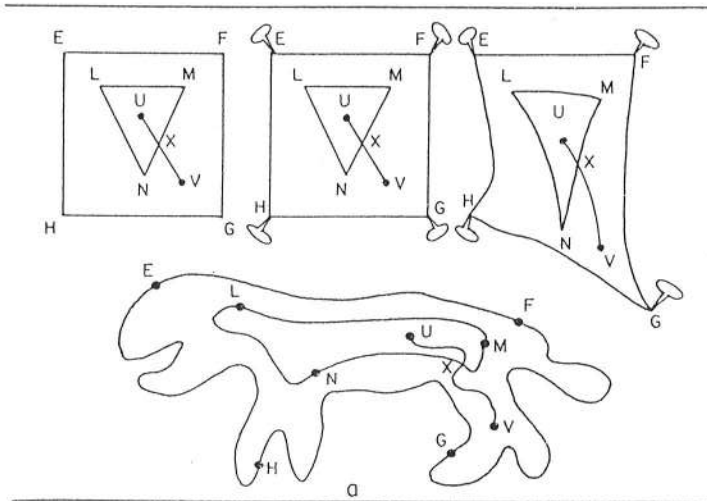
- topologinen esitystapa on joustava, euclidinen jäykkä,
- huomiota herättävät uudet faktat ja suhteet saadaan topologisilla kartoilla paremmin esiin kuin topografisilla – pelkistämällä oleellinen esiin,
- tietynlaisia jakaumia esitettäessä (värein tai rasterein) topologiset kartat antavat oikeamman kuvan asiasta kuin tavanomaiselle karttapohjalle laaditut esitykset.

Esim. maailmanlaajuisella tulotasokartalla Australia ja Belgia esiintyisivät samassa korkeassa tulotasoluokassa. Maiden väkiluvut ja ansaittujen tulojen määrät ovat suurin piirtein yhtä suuret. Kuitenkin tavanomaisella karttapohjalla Australian rasterointi olisi noin 300 kertaa laaja-alaisempi kuin Belgian johtuen maiden pinta-alaeroista. Tulo on jakautunut ihmisten kesken eikä maa-alan mukaan. Oikea esitystapa tässä esimerkissä olisi lähinnä: a) ympyrät suhteutettuina väkilukuihin ja väritettyinä tai rasteroituina tulotason mukaan tai b) topologinen esitystapa.

Yleensä topologisissa kartoissa muodon vääristyminen kasvaa käsiteltävien osaluokkien lukumäärän myötä sekä jakautuman hajonnan kasvun tai ääriarvojen eron kasvun myötä.

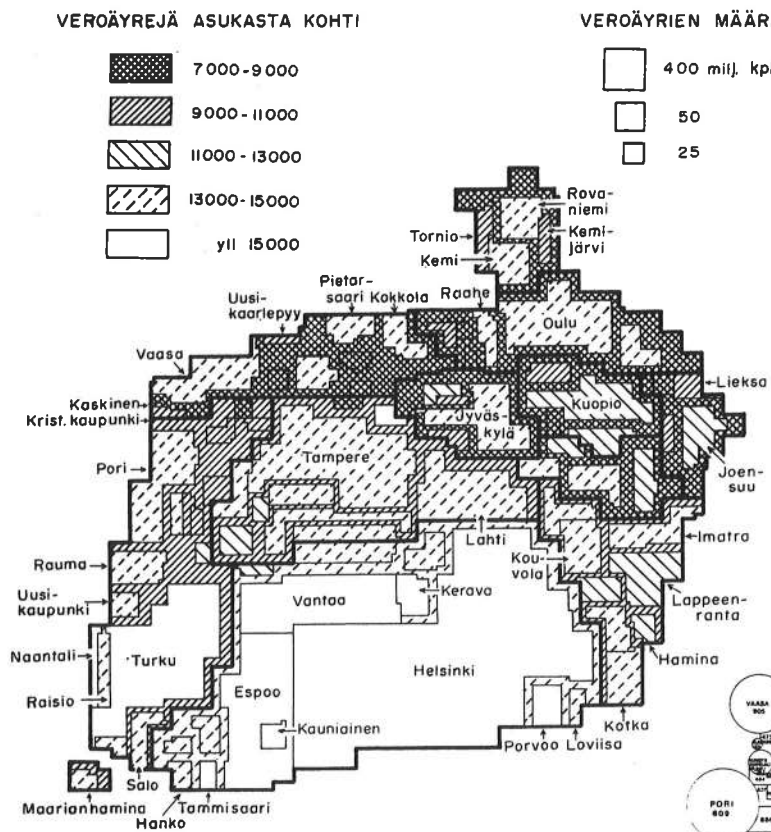
Lähteet:

- Cole & King (1968). *Quantitative Geography*. London etc.: John Wiley & Sons (s. 85–91).
Härö, Elina S. (1989). Suomen kuntien väestö pinta-aladiagrammina. *Terra* 101:2, 173–180.
Mikkonen, K. (1978). Tulojen al... *VKKK:n julkaisuja. Tutkimuksia* No 53 (s. 19–21).



Kuva 4.9.

Transformaatio (Cole & King 1968: fig. 2.28 a).

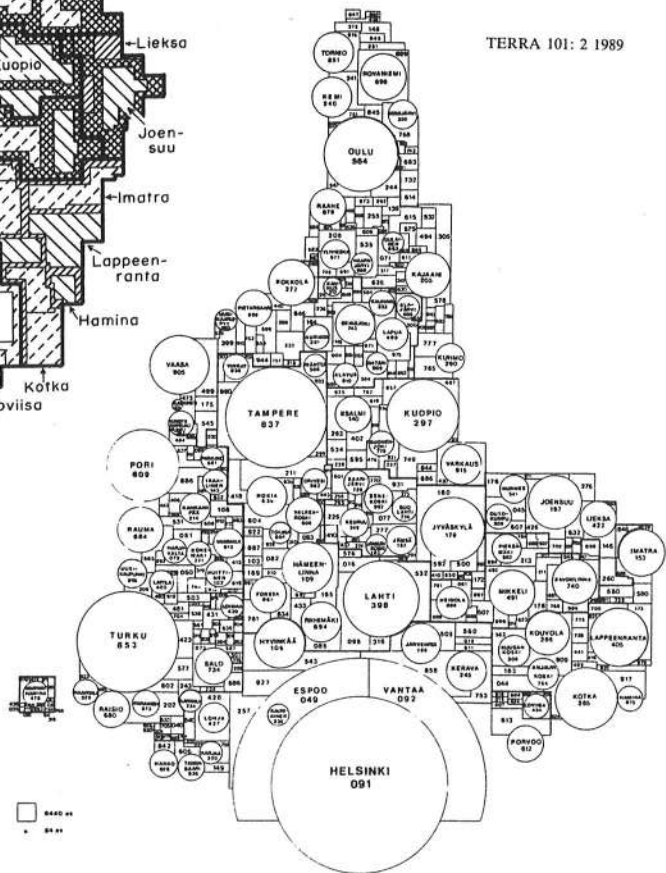


Kuva 4.10 a.

Tulojen alueellinen jakautuminen Suomessa 1975 (Mikkonen 1978: 20).

Kuva 4.10.

Kaksi topologista kuvaa Suomesta.



Kuva 4.10 b.

Kuntien väestö 1985 pinta-ala-diagrammina (Härö 1989: kuva 1).

4.3. Kartogrammien laatimisessa huomioon otettavaa

Teemakarttoja laadittaessa esitettävä asia on pääasia, karttapohja on taustalla. Kartat laaditaan yleensä lopullista esityskokoaan suurempaan mittakaavaan. Nykyaikaisia teemakarttaohjelmia käytettäessä valintoja tehdään ohjelman tarjoamien vaihtoehtojen pohjalta. Karttojen viivaelementit, symbolit, rasterit tekstin koko jne. on pyrittävä valitsemaan siten, että ne näkyvät myös pienennöksessä.

Esityksessä pyritään selkeyteen ja samalla taloudelliseen tilan käyttöön. On varottava sijoittamasta liikaa informaatiota samalle karttapohjalle. Aineksen harkitsematon lisääminen voi heikentää kartan havainnollisuutta ja tulkittavuutta.

Kuvan nimiöinnin ja selitysten on oltava riittävän yksityiskohtaisia, jotta kuvan sanoma selviää ilman muuhun tekstiin turvautumista.

Alkuperäiseen karttaan piirretään **janamittakaava**, joka pienenee kartan mukana. Alkuperäisen kartan suhdelukumittakaava ei päde pienennöksen jälkeen!

Pääsääntö: Kuvan tulee välittää lukijalle mahdollisimman oikeaa tietoa mahdollisimman selkeässä ja havainnollisessa muodossa. Diagrammiohjelmien tavoin myöskään karttaohjelmat eivät vapauta tekijää vastuusta (ks. luku 3.2.).

4.4. Tietokoneavusteinen tilastoteemakarttatuotanto

Maantieteellistä tietoa on vuosikymmenien ajan esitetty kartoilla perinteisten "käsin piirrettyjen" karttojen avulla. Teemakarttojen tuottamiseen soveltuvat tietokoneohjelmat ovat olleet varsin kalliita ja monimutkaisia kokonaisuuksia, joiden hallinta on vaatinut käyttäjältään pitkäkestoista paneutumista ja erikoisosaamista. Tietojenkäsittelykapasiteetin halpenemisen, mikrotietokoneiden kapasiteetin kohoamisen, uusien käyttäjäystävällisempien PC- ja internet -ympäristössä toimivien nk. Desktop Mapping -ohjelmien lisääntymisen sekä paikkatietotekniikan (GIS) kehittymisen myötä tietokoneavusteisen karttatuotannon mahdollisuudet ovat nopeasti avautumassa yhä useammalle käyttäjälle. Kehitys on johtanut siihen, että tietokoneavusteisesti valmistettavat teemakartat ovat korvaamassa perinteiset käsin piirrettävät kartat.

Teemakartta voidaan laatia tietokoneelle tallennettujen tietojen avulla nopeammin ja helpommin kuin perinteisillä manuaalisilla menetelmillä. Tiedot kytketään karttapohjaan sijaintitiedon (paikkatiedon) perusteella, jonka avulla tietoja voidaan analysoida, vertailla ja yhdistellä. Kartan esitystapaa, esimerkiksi karttatyyppejä, symbolien, viivojen, värien ja tekstin kokoa ja muotoa voidaan helposti vaihdella ja siten luoda erilaisia karttaesityksiä. Nykyiset karttaohjelmat sisältävät lisäksi useita monipuolisia maantieteelliseen analyysiin soveltuvia työkaluja, joista mainittakoon esimerkkeinä erilaiset aineistojen luokittelutavat, reittien optimointi ja sijainnin suunnitteluratkaisut.

Karttaohjelmilla ja niihin liittyvillä paikkatietojärjestelmillä on jo varsin paljon käyttäjiä Suomessa sekä julkisella että yksityisellä sektorilla. Järjestelmiä hyödynnetään apuvälineenä mm. kartoituksessa ja maankäytössä, kaavoituksessa, tilastoinnissa, liikenteen ja kuljetusten suunnittelussa ja valvonnassa, markkinatutkimuksissa jne. Paikkatietotekniikan koulutus on lisääntynyt merkittävästi viime vuosina ja laajentunut yliopistoista ja korkeakouluista ammattikorkeakouluihin, ammattioppilaitoksiin ja jopa lukioihin. Vaasan yliopiston aluetieteen laitoksella on käytössä MapInfo -ohjelma. Muita Suomessa yleisesti käytössä olevia paikkatieto-ohjelmia ovat mm. ArcGIS -tuoteperheeseen kuuluvat tuotteet (esim. ArcInfo, ArcView).

Kirjallisuusesimerkki: Löytönen Markku, Tuuli Toivonen, Ilta-Kanerva Kankaanrinta (2003). *Globus GIS. Paikkatietojärjestelmä*. Porvoo: WSOY.

**HARJOITUS: Koropleetti- ja isopleettikartan laatiminen, alueelliset keski-
luvut ja Lorentz-käyrä**

Lähtöaineisto voi olla esim. väestöpistekartta, jonka päälle asetetaan ruutumatriisi 10 cm x 10 cm, ruudun koko 1 cm x 1 cm.

Alustavat työvaiheet:

- 1 Ruutumatriisin ruutuihin lasketaan väestöpisteiden lukumäärät
- 2 Toinen ruutumatriisi, jossa väkiluku/ruutu
- 3 Kolmas ruutumatriisi, jossa väentiheys/ruutu

Tehtävät:

1. Laaditaan korologisen väestömatriisin perusteella väestön levinneisyyttä kuvaava koropleettikartta ja isopleettikartta.
2. Laaditaan väestömatriisin perusteella väestön painopiste ja mediaanipiste sekä haetaan modaalipiste(et).
3. Laaditaan väestömatriisin perusteella Lorentzin käyrä ja lasketaan väestön agglomeraatioaste.

Kohdat 2 ja 3 suoritetaan metodikurssin osassa Alueellisia mittaus- ja analyysimenetelmiä (luvut 1.3 ja 1.4).

HARJOITUS: Tilastoteemakarttojen piirtäminen MapInfo-ohjelmalla

Tehtävät:

1. Eriyryppisten tilastoteemakarttojen piirtämisen harjoittelu MapInfo-karttaohjelmalla.
2. Tilastotietojen syöttö ja koropleettikartan laatiminen tallennettujen tietojen pohjalta esim. tulotason alueellisesta vaihtelusta tarkastelun kohteena olevalla alueella.

TILASTOLLISIA MITTAUS- JA ANALYYSIMENETELMIÄ

1. Mittaaminen ja mitta-asteikot

Tutkimuskohteen ominaisuuksia mitataan ja analysoidaan erilaisten muuttujien avulla. Muuttujien (*variables*) valinta riippuu tutkimustehtävästä ja käytettävissä olevista tietolähteistä. Muuttujat voivat olla laatua tai määrää mittaavia.

Mitta-asteikot, joilla muuttujien kuvaamia asioita mitataan, voidaan jakaa seuraaviin tyyppeihin:

1 Nominaali- eli laatueroasteikko (*nominal scale*)

Luokitellaan kohteita laadullisten ominaisuuksien perusteella, mutta luokkia ei voida asettaa mihinkään järjestykseen eli ei voida sanoa, että jokin tutkimusyksikkö olisi "suurempi kuin" tai "pienempi kuin" jokin toinen tutkimusyksikkö.

Esim. ihmisten luokittelu sukupuolen mukaan: miehet/naiset.

2 Ordinaali- eli järjestysasteikko (*ordinal scale*)

Järjestysasteikolla kohteet luokitellaan ja luokat voidaan lisäksi asettaa tiettyyn järjestykseen eli voidaan tehdä "suurempi kuin" ja "pienempi kuin" tyyppisiä vertailuja. Luokkia osoittavat luvut ovat vain luokkien symboleja. Niiden perusteella ei siis voida sanoa, että jokin tutkimusyksikkö olisi esimerkiksi kaksi kertaa suurempi tai pienempi kuin joku toinen.

Esim. tieluokat, sotilasarvot jne.

3 Intervalli- eli välimatka-asteikko (*interval scale*)

Välimatka-asteikolla mitatulle numerosarjalle on ominaista, että numerot eivät ilmoita vain kohteiden järjestyksen, vaan ne määrittelevät myös niiden erotuksen.

Tyypillinen välimatka-asteikko on lämpömittari: kahden lämpötilan välinen ero asteissa mitattuna on suurempi, yhtä suuri tai pienempi kuin kahden muun lämpötilan erotus. On järkevää sanoa, että mittarin laskiessa nolasta kymmeneen pakkasasteeseen pakkanen on kiristynyt enemmän kuin mittarin laskiessa nolasta viiteen pakkasasteeseen. Silti ei ole oikein sanoa, että lämpötila -10°C olisi "kaksi kertaa niin kylmä" kuin lämpötila -5°C . Syynä on se, että lämpömittarin nolapiste ei ole absoluuttinen, vaan sopimuksenvarainen ($t = 0^{\circ}\text{C}$ on veden jäätymispiste).

Toinen esim. on pH-asteikko, jossa $\text{pH} = 7$ merkitsee neutraalia happamuus-astetta.

4 Suhdelukuasteikko (*ratio scale*)

Asteikkotyyppi, joka täyttää kaikki intervalliasteikon ominaisuudet, mutta lisäksi asteikolla on absoluuttinen nolapiste, jossa mitattava ominaisuus häviää. Suhdelukuasteikolla mitattaessa voidaan tehdä "kaksi kertaa suurempi kuin" tyyppisiä vertailuja.

Esimerkkejä: ikä, lasten lukumäärä, maatalouden työpaikkojen osuus kunnan työpaikoista jne.

* * *

Mitta-asteikolla on ratkaiseva merkitys sille, millaisia analyysimenetelmiä ja tilastollisia testejä aineistoon voidaan soveltaa. Jos mittauksessa on käytetty esim. laatueroasteikkoa, ei tulosten käsittelyssä voi käyttää kaikkein tehokkaimpia menetelmiä. Laatueroasteikolla mitatuista muuttujista ei esimerkiksi ole mahdollista laskea keskiarvoa tai keskihajontaa. Monet kehittyneet tilastomatemattiset tekniikat perustuvat juuri keskiarvoihin ja keskihajontoihin. Suhdelukuasteikko edustaa korkeimman asteista mittausta ja mahdollistaa vaativimmatkin analyysimenetelmät ja tilastolliset testit. Jos mittaukset on tehty suhdelukuasteikolla, ei kannata tyytyä kaikkein yksinkertaisimpiin menetelmiin. Suhdelukuasteikolla mitatuista luvuista päästään esim. luokittelemalla karkeampaan tarkasteluun mutta ei päinvastoin. Tilastollisen testin valinnasta lähemmin luvussa 4.

2. Luokittelu

Luokittelun tarkoitus voi olla kaksinainen:

- 1 Luokittelu voi olla suuren datamäärän alustavaa järjestämistä ennen datojen analysointia.
- 2 Luokittelu voi edustaa datojen tai tutkimuskohteen analysoinnin lopputulosta.

2.1. Luokittelutyyppejä

Luokittelun kompleksisuus riippuu paitsi tutkimuskohteen luonteesta myös luokittelun perustana olevien muuttujien lukumäärästä, luonteesta ja aineiston jakaumasta kullakin muuttujalla. Seuraavassa esitellään maantieteellisissä tutkimuksissa esiin tulevia luokittelutapoja.

1. *Mielivaltainen luokittelu*

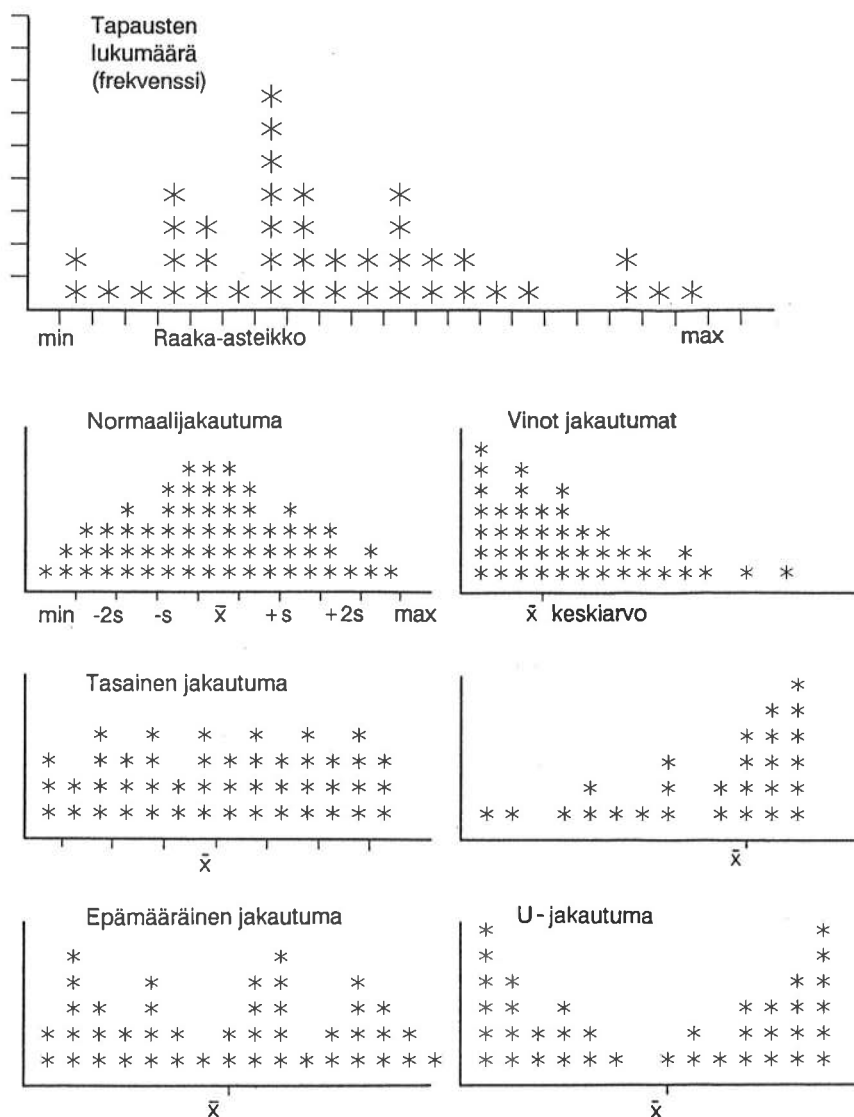
Voidaan esimerkiksi määritellä, että jyrkkä rinne on yli 25 astetta ja loiva rinne on alle 5 astetta. Tässä esimerkissä ilmiötä (rinteen kaltevuus) mitataan suhteellisella asteikolla. Mielivaltaista luokittelua voidaan käyttää myös, kun aineisto on jakautunut rypäisiin, joiden väliin jää epäsäännöllisiä tyhjiä aukkoja.

2. *Epäjatkuvan jakauman luokittelu*

Jos luokiteltavana ovat erilaiset ja toisistaan erotettavissa olevat kohteet, silloin luokitteluratkaisu sisältyy jo mittaustapaan (luokitteleva tai järjestävä mittaustapa). Kukin kohde sijoittuu yhteen ja vain yhteen käytettävissä olevista luokista. Esimerkiksi mies/nainen.

3. *Jatkuvan jakauman luokittelu*

Jatkuvien jakaumien yhteydessä (intervalli- ja suhdemittaus) yleisin asetelma lienee se, että aineisto on jakautunut lukusuoralle ilman teräviä ja itsestään selviä luokkarajojen paikkoja. Aineiston jakauman graafinen kuvaus auttaa luokkarajojen valinnassa (ks. kuva 2.1).



Kuva 2.1. Aineiston jakauman tutkiminen ennen luokittelua (Peltonen 1989: kuva 41).

Luokkarajojen valinta on tärkeä vaihe ilmiön analysoimisessa. Aineiston jakaumasta riippuen sopivat luokkarajat voivat noudattaa esimerkiksi aritmeettista jonoa, geometrista jonoa, näiden yhdistelmää tai muuta loogista luokitteluperiaatetta. Luokittelun apuna voidaan käyttää keskiarvoa ja luokkavälin mittana keskihajontaa. Kun mittaustulokset asetetaan suuruusjärjestykseen, luokkarajojen paikat voidaan löytää peräkkäisten suhdelukujen menetelmällä. Tämä menetelmä tuottaa yleensä epäsäännöllisiä luokkia.

4. *Kaksiulotteinen epäjatkuvien jakaumien luokittelu*

Kyseessä on erinomaisen tavallinen kahden luokitellun muuttujan ristiintaulukointitilanne ($m \times n$ -taulukko). Ristiintaulukointi auttaa näkemään muuttujien välisiä yhteyksiä. Ristiintaulukointi voi useissa tutkimustehtävissä olla ainoa ja riittävä tutkimusmenetelmä!

5. *Kaksiulotteinen jatkuvien jakaumien luokittelu*

Luokittelu kahden ominaisuuden perusteella. Luokittelu voi tapahtua diagrammin hajontakuvion avulla silmämääräisesti, määrittelemällä luokkarajat kummallakin ulottuvuudella kuten kohdassa 3 tai tilastollisia menetelmiä käyttäen (ks. kohta 6).

6. *Kompleksinen luokittelu*

Aineiston tai tutkimuskohteen luokittelu useiden ominaisuuksien perusteella. Kolmen ulottuvuuden tapauksessa luokittelua voidaan vielä havainnollistaa kolmiodiagrammin avulla, esim. taajamien luokittelu elinkeinojen 3-jaon mukaan. Useampien muuttujien tapauksessa lopulliseen luokittelutulokseen voidaan edetä eri tavoin riippuen tutkimustehtävästä ja aineistosta:

- 1) Luokittelu analyttisten karttojen synteessin perusteella (esim. Vaasan läänin maaseudun suuralueet; J.G. Granön maisemamaantieteelliset alueet).
- 2) Monimuuttujaisen datajoukon luokittelu tilastotieteen monimuuttujamenetelmien, esim. klusterianalyysin (lähemmin luvussa 2.2) ja faktorianalyysin (luku 6) avulla.

7. *Funktionaalinen luokittelu*

Edellyttää tavallisesti aineiston monipuolista analysointia ennen kuin luokittelu voidaan tehdä. Esimerkiksi kaupunkien luokittelu niiden palvelusisällön perusteella (keskusluokitukset) tai yleensä taloudellisten toimintojen perusteella (rooli-tutkimus): kauppakaupungit, satamakaupungit, teollisuuskaupungit, yliopisto-kaupungit.

8. *Ikäluokittelu*

Kohteiden ryhmittely ikäluokkiin edellyttää edeltävää tietoa kohteista. Esimerkiksi kaupungit tai kaupunginosat voidaan ryhmitellä ikäluokittelun perusteella. Luonnonmaantieteen puolella tunnetuin esimerkki on W.M. Davisin esittämä pinnanmuotojen jako nuoreen, kypsään ja vanhaan.

9. *Geneettinen luokittelu*

Kohteiden luokittelu niiden syntyvän ja kehityspiirteiden perusteella. Yleinen luokittelutapa luonnonmaantieteessä (esim. jokityypit).

* * *

Luokittelutyyppinä 8 ja 9 ei voida pitää varsinaisesti tilastollisina luokittelumenetelminä.

Useimmissa tutkimustehtävissä luokittelu voidaan suorittaa useilla vaihtoehtoisilla tavoilla.

Lähteet:

Cole, John P. & Cuchlaine A.M. King (1968). *Quantitative Geography*. London etc.: John Wiley & Sons, Ltd. (s. 574–583).

Peltonen, Arvo (1989). Peruskartografia. Kartografian harjoitustyökurssin moniste. Toinen painos. *Helsingin yliopiston maantieteen laitoksen opetusmonisteita* 32 (kuva 41).

2.2. Monimuuttujaisia luokittelumenetelmiä

1. Klusterianalyysi

Klusterianalyysi on yleisnimi erilaisille menetelmille, joiden avulla aineisto ryhmitellään ryhmiin yhden tai useamman muuttujan suhteen. Ryhmien lukumäärä voidaan antaa etukäteen. Ryhmittelykriteereitä on kymmeniä, ellei satoja. Eräs yleisimmistä on "varianssikriteeri", jossa ryhmien sisäinen vaihtelu pyritään saamaan mahdollisimman pieneksi ja ryhmien välinen vaihtelu mahdollisimman suureksi.

Tietyn tyyppisessä klusterianalyysissä voidaan edetä myös siten, että asia- tai ilmiöjoukko, esim. kunnat, ryhmitellään mittaustulosten (pistemäärien) perusteella askel askeleelta pienempään määrään ryhmiä lähtien liikkeelle eniten toisiaan muistuttavien yksiköiden (kuntien) yhdistämisestä. Kussakin tilanteessa yhdistetään (linkitetään) lähimpänä toisiaan olevat pisteet: joko yksikkö toisen yksikön kanssa, yksikkö useamman yksikön muodostaman ryhmän kanssa tai kaksi ryhmää keskenään. Näin edetään haluttuun luokkien lukumäärään asti tai prosessi loppuun, jolloin kaikki yksiköt ovat yhdistyneet samaan luokaan. Ryhmiä muodostettaessa käytetään yleensä yksiköiden (kuntien) alkuperäisiä muuttujakohtaisia pistemääriä. Laskentatekniikat saattavat kuitenkin yksityiskohdissaan poiketa toisistaan riippuen klusterointiohjelmasta. Tällaisesta klusterianalyysistä käytetään myös nimitystä *linkage-analyysi*.

Esimerkit: Kytömäki & Seppälä 1976: kuvat 14 ja 15,
Mikkonen & Luoma 1996: fig. 4.

Lähteitä:

Everitt, Brian (1980). *Cluster Analysis*. 2nd ed. London.

Aldenderfer & Blashfield (1985). *Cluster Analysis*. Beverly Hill Calif.: Sage.

Mikkonen, Kauko & Martti Luoma (1996). The regional possibilities in the process of external integration of the Finnish economy. *Fennia* 174:1, 97-111.

2. Faktorianalyysi

Faktorianalyysiä käsitellään muiden monimuuttujaisien analyysimenetelmien yhteydessä myöhemmin luvussa 6.

3. Muuttujajoukon kuvaaminen

Joukko eli populaatio. Reaalimaailman tilannetta edustava täydellinen data- (tai yksilöiden) joukko = perusjoukko. Joukon jakaumaa kuvaavia lukuja, ns. **joukon parametrejä**, ovat esim.: keskiarvo, mediaani, moodi, varianssi, keskihajonta. Joukon parametreja merkitään tavallisesti kreikkalaisin kirjaimin (β , μ jne.).

Otos (sample) = populaation osajoukko. Otoksen jakaumaa kuvaavia lukuja kutsutaan **tilastollisiksi tunnusluvuiksi**. Tunnuslukuja merkitään tavallisesti latinalaisin kirjaimin (a , b , \bar{X} , s jne.) tai hatulla varustetuin kreikkalaisin kirjaimin.

3.1. Tilastolliset tunnusluvut

Aritmeettinen keskiarvo

Merkitään: μ populaatio, \bar{X} otos

Luokittelematon aineisto

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_N}{N} = \frac{\sum_{i=1}^N X_i}{N}$$

Luokiteltu aineisto

$$\bar{X} = \frac{\sum_{i=1}^k n_i \bar{X}_i}{\sum_{i=1}^k n_i}$$

\bar{X}_i = luokan i luokkakeskiarvo,

n_i = tapausten lkm luokassa i ,

k = luokkien lkm.

Painotettu aritmeettinen keskiarvo

Jos arvot X_1, X_2, \dots, X_k esiintyvät f_1, f_2, \dots, f_k kertaa, niiden aritmeettinen keskiarvo on

$$\bar{X} = \frac{f_1 X_1 + f_2 X_2 + \dots + f_k X_k}{f_1 + f_2 + \dots + f_k} = \frac{\sum_{j=1}^K f_j X_j}{\sum_{j=1}^K f_j} = \frac{\sum fX}{\sum f} = \frac{\sum fX}{N}$$

jossa $N = \sum f$ on kokonaisfrekvenssi eli tapausten kokonaismäärä.

Mediaani (Md)

Md jakaa aineiston kahteen yhtä suureen osaan.

Luokittelematon aineisto:

- Md = suuruusjärjestykseen järjestetyn aineiston keskimäinen arvo, jos pariton määrä havaintoja,
= kahden keskimäisen keskiarvo, jos parillinen määrä havaintoja
- Md jakaa histogrammin pinta-alaltaan kahteen yhtä suureen osaan.

Luokiteltu aineisto:

$$Md = x_a + \frac{N/2 - N_{md-1}}{N_{md}} w$$

- x_a = mediaaniluokan alaraja,
- N_{md-1} = edellisen luokan summafrekvenssi,
- N_{md} = mediaaniluokan frekvenssi,
- w = luokkaväli,
- N = havaintojen kokonaisfrekvenssi.

Moodi

Tyyppiarvo = frekvenssiltään yleisin luku (luvut). Aina moodia ei ole.

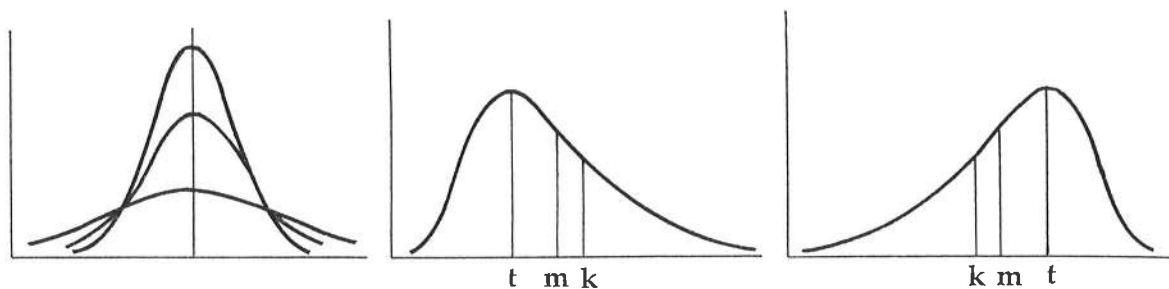
- unimodaalinen - yksi tyyppiarvo
- bimodaalinen - kaksi tyyppiarvoa
- multimodaalinen - useita tyyppiarvoja

Luokitellussa aineistossa moodin likiarvo on sen luokan keskus, jonka frekvenssi on suurin. Tarkemmassa määrittelyssä otetaan huomioon naapuriluokkien frekvenssit, jotka vetävät moodia jompaan kumpaan suuntaan. Moodin tarkka arvo saadaan tällöin kaavasta:

$$Moodi = x_a + \frac{a}{a+b} w$$

- x_a = moodiluokan alaraja,
- a = moodiluokan ja edellisen luokan frekvenssien erotus,
- b = moodiluokan ja seuraavan luokan frekvenssien erotus,
- w = luokkaväli.

Esimerkkejä (kuva 3.1).



a) Normaalijakauma erilaisilla hajonnoilla

b) Positiivinen vino jakauma

c) Negatiivinen vino jakauma

Kuva 3.1. Keskiarvo (k), mediaani (m) ja tyyppiarvo eli moodi (t) erilaisissa jakaumissa.

Geometrinen keskiarvo

$$G = \sqrt[N]{X_1 \cdot X_2 \cdots X_N}$$

Esimerkkejä geometrisen keskiarvon sovelluksista:

- 1) Woldenbergin konvergentin keskiarvon laskelmat keskusverkkotutkimuksissa.
- 2) Logistisen käyrän manuaalinen estimointi.

Keskipoikkeama (*mean deviation* eli *average deviation*)

$$M. D = \frac{\sum_{i=1}^N |X_i - \bar{X}|}{N}$$

Sovellus:
$$M. D = \frac{\sum |X_{tod} - X_{est}|}{N}$$

Varianssi eli keskihajonnan neliö

Merkitään: σ^2 populaatio
 s^2 otos

Populaatio

$$\sigma^2 = \frac{\sum_{i=1}^N (X_i - \mu)^2}{N}$$

N = havaintojen määrä

Otos

$$s^2 = \frac{\sum_{i=1}^N (X_i - \bar{X})^2}{N-1}$$

N-1 = vapausasteiden luku

Huom. Populaation ja otoksen käsittelyssä oleva ero: N tai N-1!

Keskihajonta eli standardipoikkeama (*standard deviation*)

- aineiston hajonnan mitta
- varianssin neliöjuuri

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{\sum (X_i - \mu)^2}{N}}$$

populaatio

$$s = \sqrt{s^2} = \sqrt{\frac{\sum (X_i - \bar{X})^2}{N-1}}$$

otos

Keskihajonnalla on tärkeä merkitys tilastollisissa luotettavuustesteissä. Sen suuruus ilmaisee aineiston homo- tai heterogeenisuuden: mitä pienempi s on, sitä homogeenisempi on aineisto.

Keskihajontaa voidaan käyttää luokittelussa luokkavälin mittana.

Vaihtelukerroin (*variation coefficient*)

Keskihajonnan ja keskiarvon osamäärä, joka yleensä ilmaistaan prosentteina eli lasketaan, montako prosenttia keskihajonta on keskiarvosta:

$$\frac{s}{\bar{X}} \times 100$$

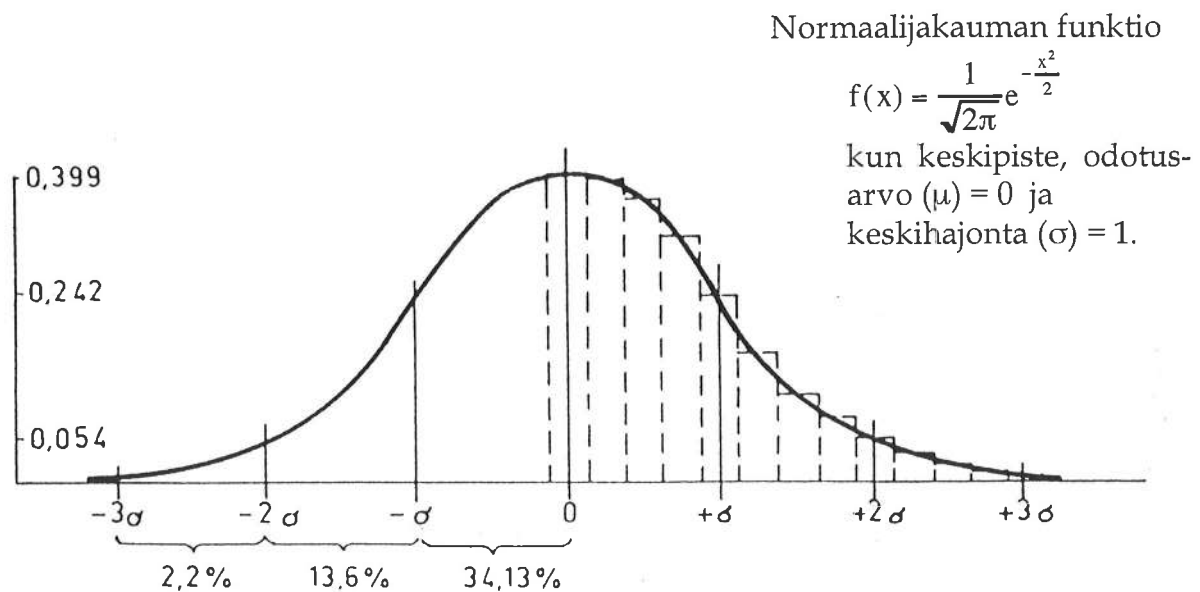
Mitä suurempi vaihtelukerroin on, sitä suurempi on tutkittavan ilmiön vaihtelu. Vaihtelukertoimen avulla voidaan tutkia ilmiön alueittaista vaihtelua, esimerkiksi työttömyyden maakunnittaista vaihtelua Suomessa.

3.2. Normaalijakauma

Normaalijakauma on jatkuva, hyppäyksetön jakauma, jonka arvot voidaan määrittellä mistä kohdasta tahansa. Käytännössä se voidaan ajatella muodostuneen äärettömän monista vierekkäisistä pylväistä (ks. kuva 3.2). Normaalijakauma on symmetrinen ja siitä seuraa, että keskiarvo, mediaani ja moodi ovat samassa pisteessä.

Normaalijakautunut aineisto keskittyy keskiarvon ympärille niin, että enintään keskihajonnan etäisyydelle siitä sopii molemmin puolin 34,13 % aineistosta eli yhteensä 68,26 %. Enintään kahden keskihajonnan etäisyydelle sopii vastaavasti yhteensä $2 \times 47,72 = 95,4$ %. Jos esim. miesten keskipituus on 175 cm ja pituuksien keskihajonta 5 cm, niin kaikkiaan 68,2 % miehistä sopii väliin 170–180 cm ja 95,4 % väliin 165–185 cm.

Käytännössä useimmat maantieteelliset ilmiöt eivät ole normaalijakautuneita vaan kaukana siitä. Jakautuma voidaan usein "normalisoida" eli saattaa lähemmäksi normaalijakautuman muotoa esim. logaritmeja käyttäen. Jos logaritmi-transformaatio normalisoi jakauman, puhutaan log-normaalista muuttujasta. Muita transformaatiomahdollisuuksia ovat neliöjuuri, kuutiojuuri, toiseen, kolmanteen jne. potenssiin korottaminen ym.



Kuva 3.2. Normaalijakauman graafinen kuvaus eli normaalikäyrä.

4. Tilastollisen testin valinta

Parametriset testit

Parametrisissä testeissä lasketaan ensin otoksesta tunnuslukuja ja testaamalla arvioidaan niiden luotettavuutta ja edustavuutta suhteessa perusjoukkoon. Parametriset testit vaativat eniten, mutta ne ovat myös tehokkaimpia. Kaikki parametriset testit ja parametriset tilastomenetelmät edellyttävät, että mitattavan ilmiön jakautuma on normaali tai sitä lähellä. Normaalisuuden tutkiminen laiminlyödään valitettavan usein. Normaalisuusehdon täytyminen jää oletuksen varaan, jos siihenkään.

Esimerkkejä tavallisimmista parametrisistä testeistä:

t-testi, varianssianalyysi, korrelaatio- ja regressioanalyysiin liittyvät testaukset.

Esim. t-testin edellytyksiä ovat:

- 1 havainnot ovat toisistaan riippumattomat,
- 2 peruspopulaatiot ovat normaalisti jakautuneet,
- 3 mittaus- ja havaintotarkkuus välimatka-asteikolla tai suhdeasteikolla.

Parametrittomat testit

Parametrittomat eli nonparametriset testit suoritetaan laskematta tunnuslukuja. Ne eivät edellytä normaalijakaumia. Ne soveltuvat luokitellun aineiston testaukseen. Parametrittomien testien teho on pienempi kuin vastaavien parametristen testien.

Esimerkkejä tavallisimmista parametrittomista testeistä:

Khin neliötesti, nelikenttätesti, binomitesti, Mann-Whitneyn U-testi (eräs t-testin vastine) jne.

Parametrittomista menetelmistä lähemmin, ks. esimerkiksi:

Metsämuuronen, Jari (2004). Pienten aineistojen analyysi. Parametrittomien menetelmien perusteet ihmistieteissä. *Metodologia-sarja* 9. Jyväskylä: Gummerus Kirjapaino Oy.

Esim. Khin neliötesti

Ristiintaulukoinneissa saadaan luokkiin jakautuneita havaintoja. Usein on tällöin joku ennakolta laadittu hypoteesi, jonka mukaan luokkiin jakautumisen odotetaan tapahtuvan. Se voi olla esim. jostakin aiemmasta tutkimuksesta saatu tulos tai teoreettinen jakauma. Yleensä empiirisesti saatu jakauma ei satu täsmälleen yhteen odotetun jakauman kanssa. Sen tähden on saatava vastaus kysymykseen, onko odotetun ja empiirisen jakauman ero suurempi kuin mitä sattumanvarainen vaihtelu voi aikaansaada. Tehtävään sopii *khin neliötesti*.

Khin neliötestillä voidaan esimerkiksi testata kahden tai useamman luokitellun empiirisen jakauman poikkeamista toisistaan. Verailupohjana ovat teoreettiset (odotetut) jakaumat.

χ^2 -testin kaava:

$$\chi^2 = \sum_{i=1}^k \frac{(n_i - n'_i)^2}{n'_i} = \sum \frac{(\text{havaitut} - \text{odotetut})^2}{\text{odotetut}}$$

jossa n_i = havaittu frekvenssi luokassa i ,
 n'_i = odotettu frekvenssi luokassa i ,
 N = kokonaisfrekvenssi,
 k = luokkien lukumäärä.

Odotusarvojen laskeminen 2 x 3 -taulukossa

	A	B	Σ
1	n'_{1A}	n'_{1B}	N_1
2	n'_{2A}	n'_{2B}	N_2
3	n'_{3A}	n'_{3B}	N_3
Σ	N_A	N_B	N

Esim.

$$n'_{1A} = \frac{N_1 \cdot N_A}{N}$$

Taulukkomalli khin neliötestiarvoa manuaalisesti laskettaessa

luokka i	n_i	n'_i	$d_i = n_i - n'_i$	d_i^2	$\frac{d_i^2}{n'_i}$
k					
					$\chi^2 = \sum \frac{d_i^2}{n'_i}$

Jos havaitut frekvenssit sopivat täsmälleen yhteen odotettujen kanssa, $\chi^2 = 0$. Mitä suurempi testiarvo on, sitä huonommin empiirinen jakautuma vastaa odotettua jakautumaa tai empiiriset jakautumat toisiaan ja sitä suuremmalla varmuudella voidaan hylätä nollahypoteesi frekvenssien yhtäsuuruudesta.

Khin neliötesti voi johtaa harhaan, jos frekvenssit ovat pienet. Vasama & Vartian (1979: 547) mukaan khin neliötestin luotettavuus käy kyseenalaiseksi, jos taulukon soluista yli 20 % saa odotusarvon, joka on pienempi kuin 5. Ongelmasta selviää tavallisesti luokkia yhdistelemällä. Esim. SPSS-ohjelmisto kyllä varoittaa automaattisesti luotettavuusongelmasta.

* * *

Usein numeerisia riskitasoja kuvataan seuraavilla symboleilla ja annetaan seuraavanlaisia sanallisia tulkintoja:

Riskitaso	Symboli	Tulkinta
10 %	o	suuntaa antava
5 %	*	melkein merkitsevä
1 %	**	merkitsevä
0,1 %	***	erittäin merkitsevä

Lähde:

Vasama, Pyry-Matti & Yrjö Vartia (1979). *Johdatus tilastotieteeseen*, Osa II. 3. painos. 341-725. Helsinki: Gaudeamus.

5. Korrelaatio- ja regressioanalyysi

5.1. Korrelaatio

Korrelaatio- ja regressiokäsitteet esitti englantilainen *Galton* 1800-luvun lopulla. Korrelaatioanalyysissä on kyse kahden muuttujan välisen riippuvuussuhteen mittaamisesta. Korrelaatiotesti vastaa vain kysymykseen, ovatko muuttujat keskenään korreloituneita. Korrelaatioanalyysissä ei ole riippuvaa muuttujaa eikä riippumatonta eli selitettävää ja selittävää muuttujaa. Kumpi tahansa voi riippua kummasta tahansa.

Merkitään: ρ (populaatio)

r (otos)

Tulomomenttikorrelaatiokerroin eli *Pearsonin* korrelaatiokerroin lasketaan:

$$r = \frac{\sum (X_i - \bar{X}) \cdot (Y_i - \bar{Y})}{\sqrt{[\sum (X_i - \bar{X})^2][\sum (Y_i - \bar{Y})^2]}} = \frac{N \sum XY - (\sum X)(\sum Y)}{\sqrt{[N \sum X^2 - (\sum X)^2][N \sum Y^2 - (\sum Y)^2]}}$$

$$= \frac{S_{XY}}{S_X S_Y} = \frac{\text{X : n ja Y : n kovarianssi eli yhteisvaihtelu}}{\text{X : n ja Y : n keskihajonnat}}$$

$$\text{Kovarianssi } S_{XY} = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{N - 1}$$

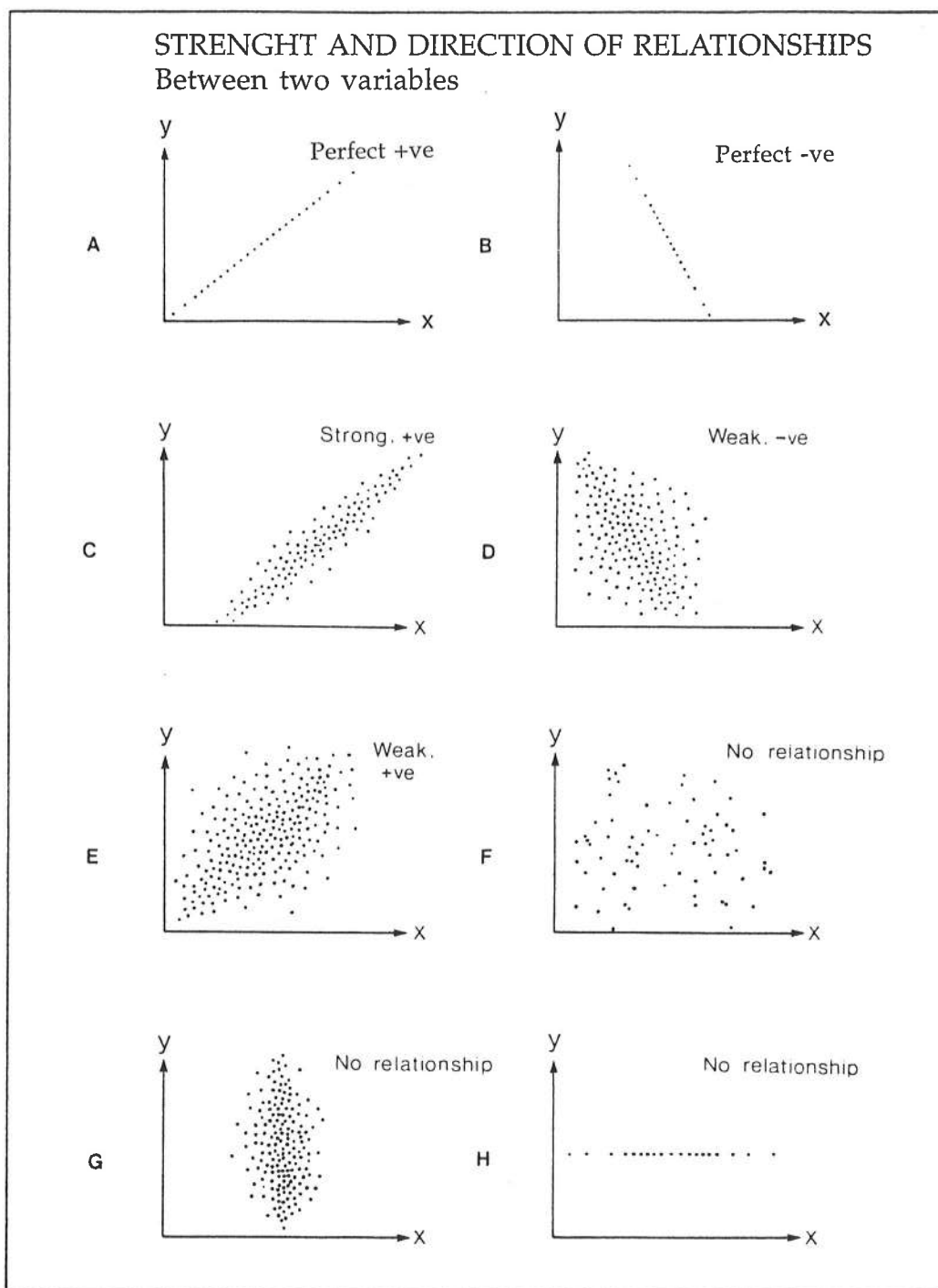
r:n ominaisuuksia:

arvot välillä +1 - -1 (ks. kuva 5.1)

$r = +1$ täydellinen positiivinen korrelaatio

$r = -1$ täydellinen negatiivinen korrelaatio

$r = 0$ ei korrelaatiota



Kuva 5.1. Kahden muuttujan x ja y välisen riippuvuussuhteen suunta ja voimakkuus, esimerkkejä.

Lähde: Matthews, John A. (1981). *Quantitative and Statistical Approaches to Geography*. Oxford etc.: Oxford Pergamon Press. (fig. 59)

Korrelaatiokertoimen merkitsevyyden testaus

Nollahypoteesi $H_0: \rho = 0$

Testataan t-testillä $t = \frac{r\sqrt{N-2}}{\sqrt{1-r^2}}$

joka noudattaa Student-jakaumaa vapausasteiden luvulla $N - 2$.

Laskettua testiarvoa verrataan asianomaisen vapausasteiden luvun kohdalla oleviin talukkoarvoihin (taulukko C).

Tulkinta: Riskitasolla, jolla laskettu testiarvo ylittää taulukkoarvon, 0-hypoteesi hylätään.

On laadittu myös valmis taulukko (taulukko H), jonka arvoihin korrelaatiokerrointa r vertaamalla voidaan suoraan päätellä sen merkitsevyys.

Z-muunnokset

a) Korrelaatiokertoimen poikkeaminen odotusarvosta

Nollahypoteesi $H_0: \rho = \rho_0 \neq 0$

r :t muunnetaan ensin Fisherin Z -muunnoksen avulla (saadaan taulukosta I), jolloin ne saadaan likimäärin normaalisti jakautuviksi muuttujiksi. Tämän jälkeen testaus tapahtuu seuraavalla testisuureella

$$t = \frac{Z - Z_0}{\sigma_Z} = (Z - Z_0)\sqrt{N-3} \quad \text{vapausasteiden luku} = \infty \text{ (t-testi)}$$

Testaus voidaan suorittaa myös normaalijakaumataulukkoa käyttäen, sillä testisuure on likimäärin normaalisti jakautunut.

$$Z = \frac{1}{2} \log_e \left(\frac{1+r}{1-r} \right) = 1.1513 \log_{10} \left(\frac{1+r}{1-r} \right) \text{ ja keskihajonta } \sigma_Z = \frac{1}{\sqrt{N-3}}$$

Z_0 on teorettinen arvo, jolla ei ole keskihajontaa.

b) Kahdesta eri aineistosta lasketun korrelaatiokertoimen eron merkitsevyyden testaus

Nollahypoteesi $H_0: \rho_1 = \rho_2 \neq 0$

Testataan testisuurella (normaalijakautunut)

$$t = \frac{Z_1 - Z_2}{\sqrt{\frac{1}{N_1 - 3} + \frac{1}{N_2 - 3}}} \quad \text{vapausasteiden luku} = \infty$$

Z:t saadaan taulukosta I. N_1 ja N_2 ovat havaintojen määrät aineistoista 1 ja 2.

5.2. Järjestyskorrelaatiokerroin

Sen sijaan, että käytettäisiin muuttujien täsmällisiä arvoja tai jos sellaisia ei ole, data voidaan järjestää esim. koon mukaiseen suuruusjärjestykseen, tärkeysjärjestykseen tai jollakin muulla perusteella käyttäen järjestysnumeroita: 1, 2, 3, . . . , n. Jos kaksi muuttujaa X ja Y on tällä tavalla järjestetty, niiden välinen järjestyskorrelaatiokerroin (*coefficient of rank correlation*) voidaan laskea kaavalla

$$r_{\text{rank}} = 1 - \frac{6 \sum D^2}{N(N^2 - 1)}$$

jossa D = X:n ja Y:n vastaavien järjestyslukujen erotukset,
 N = havaintoparien lukumäärä.

Kaava on nimeltään *Spearmanin* järjestyskorrelaatiokaava.

Esimerkki:

The following table shows how 10 students, arranged in alphabetical order, were ranked according to their achievements in both the laboratory and lecture portions of a biology course. Find the coefficient of rank correlation.

Laboratory	8	3	9	2	7	10	4	6	1	5
Lecture	9	5	10	1	8	7	3	4	2	6

Solution:

The difference of ranks D in laboratory and lecture for each student is given in the following table. Also given in the table are D^2 and $\sum D^2$.

Difference of ranks, D	-1	-2	-1	1	-1	3	1	2	-1	-1	
D^2	1	4	1	1	1	9	1	4	1	1	$\sum D^2 = 24$

Then

$$r_{\text{rank}} = 1 - \frac{6 \sum D^2}{N(N^2 - 1)} = 1 - \frac{6(24)}{10(10^2 - 1)} = .8545$$

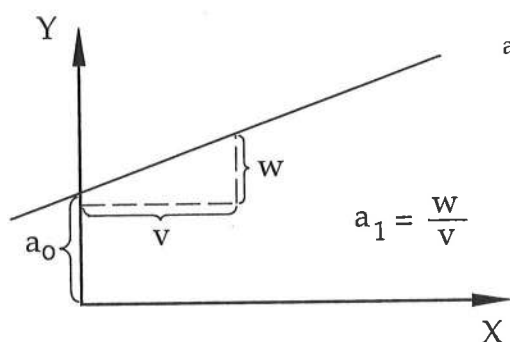
indicating that there is a marked relationship between achievements in laboratory and lecture.

5.3. Regressio (*simple regression*)

Regressioanalyysissä on riippuva eli selitettävä muuttuja (merk. tavall. Y), jonka riippuvuutta selittävästä eli riippumattomasta X -muuttujasta tutkitaan. Tutkijan on siis tehtävä oletamus siitä, kumpi on syy ja kumpi seuraus.

Lineaarinen regressio on kyseessä silloin, kun Y :n riippuvuutta X :stä voidaan kuvata suoralla viivalla. Regressiosuoraa kuvaava lineaarifunktio on muotoa:

$Y = a_0 + a_1X$, jossa a_0 = vakio, joka osoittaa, missä kohtaa suora leikkaa Y -akselin ($Y = a_0$, kun $X = 0$)



a_1 = regressiokerroin eli regressiosuoran kulmakerroin joka osoittaa regressiosuoran kaltevuuden ja tarkoittaa, että yhden yksikön muutosta X -akselin suunnassa vastaa a_1 yksikön muutos Y -akselin suunnassa.

Jos regressiokerroin on positiivinen, niin X :n kasvaessa Y kasvaa. Jos kerroin on negatiivinen, niin X :n kasvaessa Y pienenee.

Lineaarinen regressiosuora määritetään empiirisestä aineistosta ns. pienimmän neliösumman keinolla (*the least square method*). Tällöin minimoituu suorasta laskettujen poikkeamien neliöiden summa.

Lineaarifunktioiden parametrien estimointikaavat ovat:

$$a_0 = \frac{(\sum Y)(\sum X^2) - (\sum X)(\sum XY)}{N \sum X^2 - (\sum X)^2} \quad a_1 = \frac{N \sum XY - (\sum X)(\sum Y)}{N \sum X^2 - (\sum X)^2}$$

Regressiokerroimen ja korrelaatiokerroimen välillä on yhteys:

$$a_1 = \frac{r_{xy} \cdot s_y}{s_x}$$

Selitysaste R^2 (coefficient of determination)

Yleensä regressioanalyysin yhteydessä lasketaan myös selitysaste. Se ilmoittaa, montako prosenttia X selittää Y:n vaihtelusta.

$$R^2 (\%) = 100 r^2$$

Testaukset

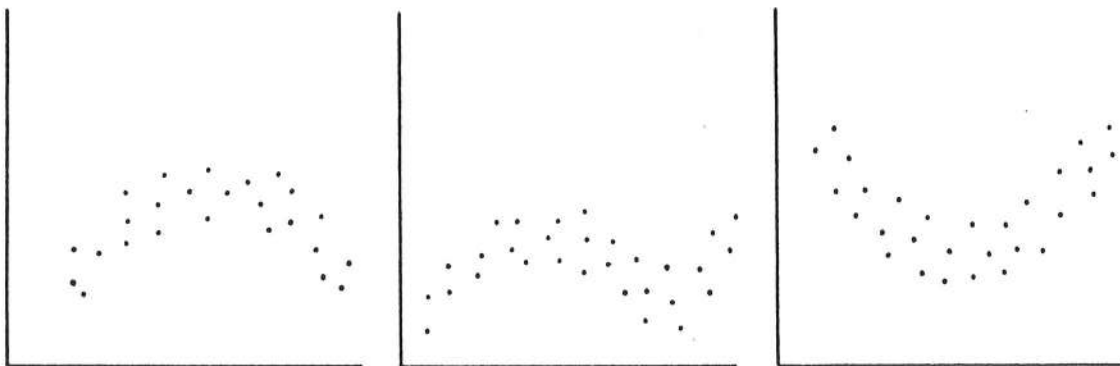
Regressiokertoimen merkitsevyys testataan *t*-testillä, $H_0: a_1 = 0$

Kahden regressiokertoimen eron merkitsevyys voidaan tiettyjen edellytysten vallitessa testata niin ikään *t*-testillä, $H_0: a_1 = a_1' \neq 0$.

T-testiarvon laskutoimitukset sivuutetaan tässä yhteydessä (ks. lähemmin esim. Mikkonen 1972: 61–62).

Käyräviivaiset regressiot

Käytännössä muuttujien väliset riippuvuussuhteet eivät suinkaan aina ole suora-
viivaisia. Jos nähdään – esim. korrelaatiodiagrammin avulla (kuva 5.2) – että toisen asteen käyrä, paraabeli, sopii kuvaamaan pistejoukon jakaumaa, paraabelin parametrit voidaan estimoida suoraan pienimmän neliösumman keinolla (*least square parabola*). Samoin voidaan menetellä kolmannen, neljännen, . . . , *n*:nnen asteen käyrän kanssa.



Kuva 5.2. Kolme käyräviivaista korrelaationsuhdetta.

Käyrien yhtälöt:

(1)	$Y = a_0 + a_1X$	Straight line
(2)	$Y = a_0 + a_1X + a_2X^2$	Parabola or Quadratic curve
(3)	$Y = a_0 + a_1X + a_2X^2 + a_3X^3$	Cubic curve
(4)	$Y = a_0 + a_1X + a_2X^2 + a_3X^3 + a_4X^4$	Quartic curve
(5)	$Y = a_0 + a_1X + a_2X^2 + \dots + a_nX^n$	n th degree curve

Vuorovaikutusmalli

Käyräviivaisesta riippuvuudesta on mallin parametrien estimoimiseksi usein välttämätöntä pyrkiä suoraviivaiseen riippuvuuteen. Tämä koskee esim. vuorovaikutusmallia. Monien inhimillisten toimintojen alueellinen systemaattisuus noudattaa tiettyä toiminnallista perusperiaatetta, ns. *allometrista lakia*, joka voidaan ilmaista yleisessä muodossa

$$(1) \quad Y = ae^{bf(X)}$$

Y on kuvattava ilmiö, jonka toiminnallista riippuvuutta X :stä tutkitaan. Vuorovaikutusmalleissa X on yleensä etäisyys johonkin mittauskeskipisteeseen, a ja b ovat empiirisen aineiston avulla estimoitavia parametreja, e on luonnollisen logaritmijärjestelmän kantaluku (kymmenkantaisessa järjestelmässä e :n paikalla on luku 10). Allometrisesta lainalaisuudesta ovat esimerkkejä *rank size rule* -sovellukset ja kaupunkien väkiluvun ja väentihyden väliset riippuvuudet. Asiointitiheydet keskuksista etäännyttäessä vähenevät samaa perusmallia noudattaen. Sitä noudattavat myös eksponentiaaliset kasvu-urat jne.

Kaavan 1 erilaiset muunnokset ryhmitellään $f(X)$:n avulla kahteen funktio-tyyppiin:

1. ryhmä

$$f(X) = X^m$$

m:n arvoina käytetty yleisesti
lukuja 0.5, 1 ja 2

2. ryhmä

$$f(X) = (\ln X)^m$$

m:n arvoina käytetty yleisesti
lukuja 1 ja 2

Kaavan 1 yleisimmän käytettyjä muotoja ovat siten:

$$(2) \quad Y = ae^{b\sqrt{X}} \quad (\text{ns. neliöjuuri-eksponenttimalli})$$

$$(3) \quad Y = ae^{bX}$$

$$(4) \quad Y = ae^{bX^2}$$

$$(5) \quad Y = ae^{b(\ln X)} \quad (\text{ns. Pareto-malli})$$

$$(6) \quad Y = ae^{b(\ln X)^2}$$

Lineaarimuunnoksiin päästään, kun yhtälöt logaritmoidaan:

$$(2') \quad \ln Y = \ln a + b\sqrt{X}$$

$$(5') \quad \ln Y = \ln a + b(\ln X)$$

$$(3') \quad \ln Y = \ln a + bX$$

$$(6') \quad \ln Y = \ln a + b(\ln X)^2$$

$$(4') \quad \ln Y = \ln a + bx^2$$

Tämän jälkeen parametrit estimoidaan empiirisestä aineistosta pienimmän neliösumman menetelmällä.

Huom! On kehitetty menetelmiä, joilla pistejoukon linearisoiva transformaattori (m:n arvo) voidaan määrätä tarkasti edellä kuvatun likimääräisen valintamenetelyn, esim. neliöjuuritransformaation asemesta.

Esimerkkejä:

Mikkonen, Kauko (1972). The gravitation field of Forssa. A gravitational analysis of the functional system. *Fennia* 115.

Mikkonen, Kauko (1983). Vuorovaikutusmallin etäisyysparametri ja aluerakenne. *Terra* 95: 2, 109-118.

5.4. Multiregressio

Multiregressioanalyysissä muuttujan Y vaihtelua selitetään kahdella tai useammalla selittävällä muuttujalla X_1, X_2, \dots, X_n .

Yhden selitettävän ja kahden selittävän muuttujan tapaus voidaan vielä kuvata graafisesti pintana 3-ulotteisessa avaruudessa.

Lineaarisen multiregressioyhtälön yleinen muoto on

$$Y = a_0 + a_1X_1 + a_2X_2 + \dots + a_nX_n + e,$$

jossa a_0 on vakio ja a_1, a_2, \dots, a_n ovat vastaavien muuttujien osittaisregressiokertoimia. Ne kuvaavat Y :n ja kunkin X :n keskinäistä riippuvuutta muiden tekijöiden pysyessä vakioina. Riippuvuuksia voidaan tällä tavoin kuvata myös graafisesti. e on regressiomallin selittämätön osa eli jäännösvarianssi.

Parametrien estimointi suoritetaan pienimmän neliösumman keinolla ja käytännössä esim. SPSS-ohjelmalla.

Tulostusvaihtoehdot

- 1) Yleisin tulostustapa on sellainen, jossa on etukäteen valitut selittävät tekijät ja malli estimoidaan kerralla.
- 2) Step wise- eli askeltava multiregressiotulostus. Siinä valitaan selittävät tekijät korrelaatiomatriisin perusteella selityskyvyn mukaisessa järjestyksessä: paras selittäjä ensin, seuraavana tulee tekijä, joka selittää jäännösvarianssista suurimman osan jne.
- 3) Kiinnitetty tulostusjärjestys. Selittävät tekijät tulostetaan tutkijan ennalta valitsemassa järjestyksessä.
- 4) Edellisten (2-3) yhdistelmä: esim. haluttu tekijä ensimmäisenä, muut selityskyvyn mukaisessa järjestyksessä.

Yhteiskorrelaatiokerroin ja selitysaste

Multiregressioanalyysin yhteydessä lasketaan myös yhteiskorrelaatiokerroin R^2 (*multiple regression coefficient*) ja selitysaste $100 R^2$ (*multiple coefficient of deter-*

mination), joka osoittaa, kuinka monta prosenttia selittävät tekijät yhdessä selittävät Y:n vaihtelusta.

Testaukset

Regressiokertoimien merkitsevyys testataan *t-testillä*. $H_0: a_i = 0$

Koko mallin merkitsevyys testataan *F-testillä*: $H_0: a_1 = a_2 = \dots a_n = 0$

Yleistä regressio- ja multiregressioanalyysin käytöstä

Analyysien tulkinta on selkein, jos selittävät tekijät ovat riippumattomia toisistaan. Tämä ei toteudu useinkaan käytännössä. Selittävien tekijöiden riippuvuutta voidaan kontrolloida korrelaatiomatriisin avulla ja poistaa tarpeeton tai tarpeettomat, keskenään voimakkaasti korreloivista muuttujista jo ennen regressioanalyysiä. Näin selkiytetään tulkintaa. Selittävien tekijöiden korreloitumista kutsutaan *multikollineaarisuudeksi*. Jos multikollineaarisuus on liian korkea, regressiokertoimien etumerkit voivat käyttäytyä epäloogisesti! Tekijöiden karsiminen voi tapahtua myös step wise -tulostuksesta testaamalla regressiokertoimien merkitsevyyttä *t-testillä* ja nojautumalla selityksasteen muutoksiin (jos tekijä ei lisää selityksastetta, ei syytä ottaa mukaan). Joskus taavoitteena on nimenomaan korkean korrelaation saavuttaminen (esim. ennustemalleissa). Tällöin selittävien tekijöiden keskinäiseen riippuvuuteen ei ole syytä liikaa tuijottaa.

Regressiomallien avulla voidaan

- 1) löytää ilmiön vaihtelua parhaiten selittävä(t) tekijä(t);
- 2) laskea empiiristen ja teoreettisten (mallin antamien) arvojen erotuksia, ns. *residuaaleja*, ja selittää niiden syitä;
- 3) poimia erityisen selvästi teoreettisista tuloksista poikkeavia havaintoja lähempään tarkasteluun;
- 4) optimoida ja ennustaa (esimerkkejä: Palomäki & Mikkonen 1971, Mikkonen 1975).

Special example:

Lineaarisen ja aaltoliikemallin yhdistäminen monimuuttujamalliksi

Mikkonen, Kauko (1992). The areal dynamics of the daily product retail trade. A case study from the market area of Vaasa, Finland. *Managerial and Decision Economics* 13:3, 255-266.

6. Faktorianalyysi

Faktorianalyysiä on sovellettu alunperin käyttäytymistieteissä, älykkyyteen liittyvissä tutkimuksissa, mutta varsin pian - tietokoneiden mahdollistettua suuren datamäärän käsittelyn, Suomessa 1960-luvulta lähtien - analyysitekniikkaa alettiin käyttää runsaasti myös aluetutkimuksen apuvälineenä.

Faktorianalyysi, siihen läheisesti liittyvä pääkomponenttianalyysi sekä seuraavassa luvussa käsiteltävä erotteluanalyysi ovat menetelmiä, joiden tuloksena useista muuttujista muodostetaan yhdistettyjä muuttujia. Ne poikkeavat korrelaatio- ja regressioanalyysistä sekä polku- ja kanonisesta analyysistä (ei käsitelty tässä opetusjulkaisussa), jotka kuuluvat selitysmallimenetelmien ryhmään.

Faktorianalyysin perusajatus

Faktorianalyysissä tutkimuskohteiden, esim. yksilöiden, yritysten, alueyksiköiden, ominaisuuksia kuvaavat muuttujat muunnetaan uusiksi muuttujiksi, faktoreiksi. Faktorit ovat aineiston perusulottuvuuksia. Ne identifioivat ne ominaisuudet, jotka ovat havaintoyksiköille yhteisiä ja ilmenevät muuttujien ja kyseisten ominaisuuksien (ulottuvuuksien) välisinä korrelaatioina. Esim. alueellisessa kehittyneisyystutkimuksessa tarkastelun kohteena voisivat olla Suomen kaikki kunnat (kuten ovat usein olleet), joista on mitattu vaikkapa 50 kehittyneisyyden astetta tavalla tai toisella kuvaavaa ominaisuutta. Nämä 50 ominaisuutta on faktorianalyysissä mahdollista tiivistää esim. kuntien ominaisuutta kuvaavaksi 4-6 perusulottuvuudeksi.

Faktorianalyysin toteutus

Faktorianalyysin **muuttujat** on valittava siten, että niillä on looginen yhteys tutkittavaan asiaan (ns. relevantit muuttujat). Muuttujien mittaamisessa joudutaan joskus (tietojen saatavuusongelman takia) parhaan vaihtoehdon asemesta tyytymään lähinnä sopivaan, alkuperäistä asiayhteyttä mahdollisimman hyvin edustavaan "korvike"ratkaisuun. Muuttujien mittaamisesta ovat voimassa parametristen testien yleiset mittaustarkkuusehdot.

Faktorianalyysin lähtökohtana on tarkastelussa mukana olevien muuttujien **korrelaatiomatriisi**. Sen laatiminen edellyttää täydellistä datajoukkoa siten, että jokaisesta havaintoyksiköstä on mittaustulos jokaisella muuttujalla.

Korrelaatiomatriisista tulostetaan **faktorimatriisi**. Menetelmänä käytetään ns. pääakseliratkaisua. Tulkintojen helpottamiseksi pääakselit yleensä rotatoidaan esim. *varimax*-menetelmällä. Rotaation avulla kukin muuttuja sijoittuu mahdollisimman lähelle yhtä faktoria (saa korkean latauksen tällä faktorilla) ja mahdollisimman etäälle muista faktoreista (saa matalia latauksia). Mikäli muuttuja korreloi voimakkaasti esim. kahden perusulottuvuuden kanssa, se saa korkeat lataukset näillä molemmilla faktoreilla – rotatoinnista huolimatta. Ulottuvuuksia, faktoreita, voidaan laskea korkeintaan yhtä monta kuin analyysissä on muuttujia. Toisaalta faktorianalyysin tavoitteena on tavallisesti suuren muuttujajoukon tiivistäminen muutamaankin faktoriin, jolloin maksimimäärän tulostamisessa ei ole mieltä.

Faktorimatriisin aikaansaaminen voi olla faktorianalyysin lopputulema, jonka perusteella tehdään johtopäätökset. Usein suurin mielenkiinto kohdistuu ensimmäisen faktorin sisältämään informaatioon. Tutkimustehtävästä ja faktoroinnin tuloksista riippuen faktorimatriisiin tulostetaan yleensä 4–5 faktoria tunnuslukuineen. Faktorianalyysiä voidaan myös jatkaa laskemalla tutkimuskohteiden (esim. alueyksiköiden) **faktoripistemäärät** (the *factor scores*). Se tapahtuu lausumalla halutut faktorit muuttujien avulla ja antamalla tutkimuskohteiden arvot muuttujille. Yleensä faktoripisteet standardoidaan siten, että niiden keskiarvoksi tulee 0 ja hajonnaksi 1. Tuloksia havainnollistetaan tarvittaessa diagrammesityksin: havaintoyksiköt sijoitetaan faktoripistemääriensä perusteella kaksiulotteiseen faktoridiagrammiin. Pistejoukkoa voidaan edelleen käsitellä tilastollisin menetelmin, esim. ryhmitellä klusterianalyysin avulla. Alueyksiköittäin lasketut faktoripistemäärät voidaan esittää karttapohjalla luokittelemalla piste-määrät ja rasteroimalla alueet sen mukaan, mihin luokkaan ne sijoittuvat.

Keskeisiä käsitteitä

1. **Faktorilataus** eli **painokerroin** on alkuperäisen muuttujan ja faktorin välinen korrelaatio. Lataus voi siis saada arvoja välillä $-1,00$ – $+1,00$. Latauksen neliö ilmoittaa muuttujan ja faktorin yhteisen varianssin suhteellisen osuuden. Jos

lataus on esim. 0,50, se merkitsee, että faktori selittää muuttujan varianssista osuuden 0,25 eli 25 %. Faktorilatauksien avulla voidaan valita tärkeimmät muuttujat (ns. kärkimuuttujat) luonnehtimaan faktorin edustamaa ulottuvuutta.

2. **Kommunaliteetti** on alkuperäisen muuttujan latausten neliöiden summa. Kommunaliteetti osoittaa, kuinka suuren osan kaikki faktorit yhteensä selittävät ko. muuttujan vaihtelusta.

3. **Ominaisarvo** (*eigenvalue*) on kunkin faktorin latausten neliöiden summa. Se ilmoittaa, kuinka suuren osan kukin faktori selittää kaikkien muuttujien vaihtelusta yhteensä. Ominaisarvoja voidaan käyttää apuna harkittaessa, montako faktoria otetaan lähempään tarkasteluun. Pienen ominaisarvon (esim. ykköistä pienemmät) faktorit jätetään tarkastelun ulkopuolelle.

4. **Kumulatiivinen ominaisarvo** on ominaisarvojen summa. Sen prosentuaalinen osuus alkuperäisten muuttujien kokonaismäärästä on faktorimallin "selitysaste".

Pääkomponenttianalyysi

Pääkomponenttianalyysi on lähellä faktorianalyysiä. Useissa aineistoissa pääkomponentit ja faktorianalyysin pääakseliratkaisun faktorit johtavat samanlaisiin tulkintoihin. Faktorianalyysi poikkeaa pääkomponenttianalyysistä lähinnä siinä, että faktorianalyysissä tarkastellaan muuttujien yhteisvaihtelua, pääkomponenttianalyysissä muuttujien kokonaisvaihtelun selittämistä. Pääkomponenttianalyysissä ei yleensä käytetä rotatointimenettelyä kuten faktorianalyysissä, sillä tavoitteena on kasata mahdollisimman suuri osa muuttujajoukon vaihtelusta ensimmäiselle pääkomponentille. Rotatointi helpottaa tulkintaa mutta heikentää samalla ensimmäisen pääkomponentin selityskykyä.

Huomautuksia

Muuttujien valinta on faktorianalyysin (kuten myös pääkomponenttianalyysin) kriittisin vaihe. Faktorit voivat heijastella vain alkuperäisten muuttujien sisältämää informaatiota. Faktorianalyysi voi olla vain niin hyvä kuin alkuperäinen data on. Faktorianalyysissä pätee kaksi tunnettua periaatetta:

- Gigo -periaate: jos panet roskaa sisään, saat roskaa ulos,
- Lilo -periaate: jos panet paljon dataa sisään, saat paljon dataa ulos.

Esimerkki:

Mikkonen, Kauko & Martti Luoma (1996). The regional possibilities in the process of external integration of the Finnish economy. *Fennia* 174:1, 97-111.

Lähteitä:

Short, J. R. (1991). *An introduction to urban geography*. London: Routledge & Kegan Paul, Ltd. (s. 146-152).

Smith, D. M. (1973). *The Geography of Social Well-being in the United States*. New York: McGraw-Hill.

7. Diskriminaatioanalyysi

Diskriminaatio- eli erotteluanalyysiä käytetään tutkittavan aineiston osajoukkojen erotteluun ja vertailuun eri ominaisuuksien perusteella.

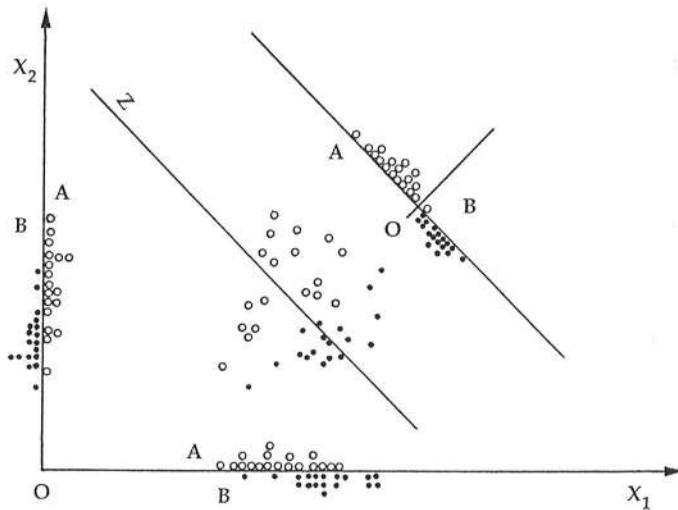
Erotteluanalyysillä voidaan

- 1) testata osajoukkojen välisiä eroavuuksia eri ominaisuuksien suhteen ja selvittää, mitkä ovat tärkeitä erottavia ominaisuuksia,
- 2) saada mitta osajoukkojen välisille eroavuuksille,
- 3) ratkaista, mihin osajoukkoon mikin alkio olisi parhaiten luettava ja vertailla alkuperäisen luokittelun ja erotteluanalyysin antaman "optimaalisen" luokittelun yhtenevyyttä,
- 4) ratkaista, mihin osajoukkoon tutkimuksen piiriin myöhemmin tulevat yksilöt kuuluvat.

Erotteluanalyysin perusajatus

Kahdesta (tai useammasta) perusjoukosta on tehty otos (= osajoukot, jotka voivat olla myös yhdestä perusjoukosta tehdyn otoksen alaryhmiä), ja yksilöistä on mitattu eräs ominaisuus x_1 . Vaikka tämän ominaisuuden keskiarvot poikkeavatkin jonkin verran toisistaan, ominaisuuden vaihtelurajat peittävät kuitenkin toisiaan niin paljon, että pelkästään tämän ominaisuuden perusteella ryhmien välille ei saada merkitsevää eroa eikä tuntematonta yksilöä pystyttäisi sijoittamaan kumpaankaan perusjoukkoon. Samoin voi olla jonkin toisen ominaisuuden x_2 laita (tai toisten ominaisuuksien).

Asiaa havainnollistaa kuva 7.1. Siinä osajoukon A jakauma muuttujien x_1 ja x_2 suhteen on esitetty avoimin ympyröin kuvion sisäpuolella ja osajoukon B jakauma pistein kuvion ulkopuolella. Kuvioista nähdään, ettei kummarkaan ominaisuuden perusteella osajoukkoja voida erottaa toisistaan.



Kuva 7.1. Diskriminaatioanalyysin havainnollistus.

Erottelufunktion muodostaminen

Erotteluanalyysissä muodostetaan uusi muuttuja z , joka on sekä x_1 :n että x_2 :n (ja ehkä vielä x_3 :n, x_4 :n jne.) funktio

$$z = a_1x_1 + a_2x_2 + \dots + a_mx_m$$

Jos z on lineaarinen kuten esimerkkikuviossa, sitä kuvaava suora kulkee keskiarvopisteen (\bar{x}_1, \bar{x}_2) kautta. Erottelufunktio, ns. **diskriminaattori** z konstruoidaan siten, että se erottelee tarkasteltavat ryhmät toisistaan, ts. painokertoimet a_1, a_2, \dots, a_m valitaan niin, että mahdollisimman moni toisen osajoukon yksilö saa suuria z -arvoja ja mahdollisimman moni toisen osajoukon yksilö pieniä z -arvoja. Tällöin z :n perusteella osajoukot voidaan erottaa paremmin kuin x_1 :n tai x_2 :n avulla (kuvassa 7.1 on osajoukkojen A ja B jakaumat piirretty z -suoran kanssa yhdensuuntaiselle "apu"suoralle, A yläpuolelle ja B alapuolelle).

Jos nyt tutkittavaksi tulee tuntematon uusi yksilö, mitataan siitä ominaisuudet x_1 ja x_2 , lasketaan niiden avulla z käyttäen kertoimia a_1 ja a_2 , ja tehdään z :n perusteella päätös kumpaan osajoukkoon uusi yksilö kuuluu.

Parhaan erottelijan lisäksi voidaan muodostaa muita toisistaan riippumattomia erottelijoita. Diskriminaattorien lukumäärä riippuu osajoukkojen ja muuttujien määrästä ollen pienempi luvuista: osajoukkojen lukumäärä - 1, muuttujien lukumäärä.

Erotteluanalyysin matematiikka pohjautuu yleiseen varianssianalyysiin, ts. pyrkimyksenä on maksimoida ryhmien välinen varianssi suhteessa ryhmien sisäiseen varianssiin.

Erotteluanalyysin välivaiheena lasketaan usein ns. *Mahalanobiksen etäisyysmitta* D^2 , jolla testataan, poikkeavatko muuttujittain lasketut ryhmien keskiarvot toisistaan. D^2 ilmoittaa ryhmien keskipisteiden keskinäisen etäisyyden keskihajonnan suuruusina yksikköinä.

$$D^2 = a_1 d_1 + a_2 d_2 + \dots + a_m d_m,$$

jossa

$$d_1 = \bar{x}_{A1} - \bar{x}_{B1}, d_2 = \bar{x}_{A2} - \bar{x}_{B2}, \dots$$

Etäisyysmitan arvoa voidaan käyttää kuten *khin neliötestiä*.

Erotteluanalyysiohjelmissa on yleensä mahdollista tulostaa muuttujat erottelukyvyn mukaisessa järjestyksessä. Järjestyksen määrittämisessä käytetään *Wilksin lambda*-suuretta siten, että kullakin askeleella sitä parannetaan mahdollisimman paljon. F-testillä testataan kussakin vaiheessa keskiarvojen yhtäsuuruus.

Esimerkki:

Mikkonen, Kauko (1977). Maantieteen jatko-opiskelijat ja heidän muodostamansa työpaikkakuvat. *Terra* 89:2, 53-68.

Lähde:

Mäkinen, Yrjö (1974). *Tilastotiedettä biologeille*. 3. painos, s. 191-195. Turku: Synapsi r.y.:n kurssi-moniste.

8. Logistinen käyrä

Erialaisten innovaatioiden eli uudisteiden omaksumisfrekvenssi noudattaa yleensä tiettyä säännönmukaista kehitysuraa. Alussa uudisteen omaksuvat muutamat harvat innovaattorit. Vähitellen omaksujien määrä aikayksikköä kohti lisääntyy (varhainen enemmistö), kunnes saavutetaan maksimitaso. Tämän jälkeen uusien omaksujien määrä alkaa laskea (myöhäinen enemmistö). Kyllästymistaso on saavutettu, kun viimeisetkin 'vitkastelijat' ovat uudisteen omaksuneet.

Omaksumisfrekvenssien kumulatiiviset arvot, summafrekvenssit, kasvavat aluksi kiihtyvällä nopeudella. Kasvu hidastuu käännepisteen (= omaksumistiheyden huippupiste) jälkeen ja loppuu kokonaan, kun lähestytään kyllästymistasoa (kuva 8.1).

Kuvattu summafrekvenssijakautuma on muodoltaan ns. **logistinen käyrä**, jonka kaava on

$$P_T = \frac{U}{1 + e^{a+bT}}$$

Muuttujat:

P_T = innovaation omaksuneiden kokonaismäärä tai suhteellinen osuus ajankohdassa T ,

T = aikatekijä.

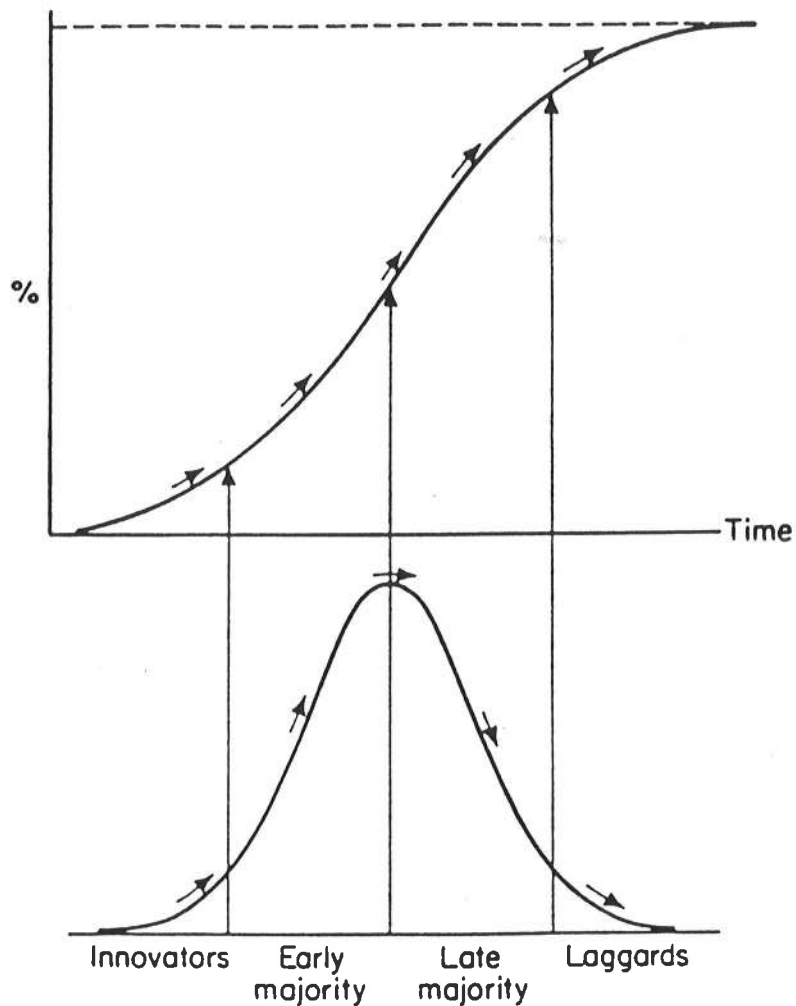
Estimoitavat parametrit:

U = yläraja eli ns. kyllästymistaso, mittaustavasta riippuen joko absoluuttinen lukumäärä tai 100 %,

b = vakio, joka ilmaisee sen nopeuden, jolla P muuttuu T :n muuttuessa,

a = p :n arvo, kun $T = 0$.

e on luonnollisen logaritmijärjestelmän kantaluku. Kymmenkantaisessa järjestelmässä e :n paikalla on luku 10.



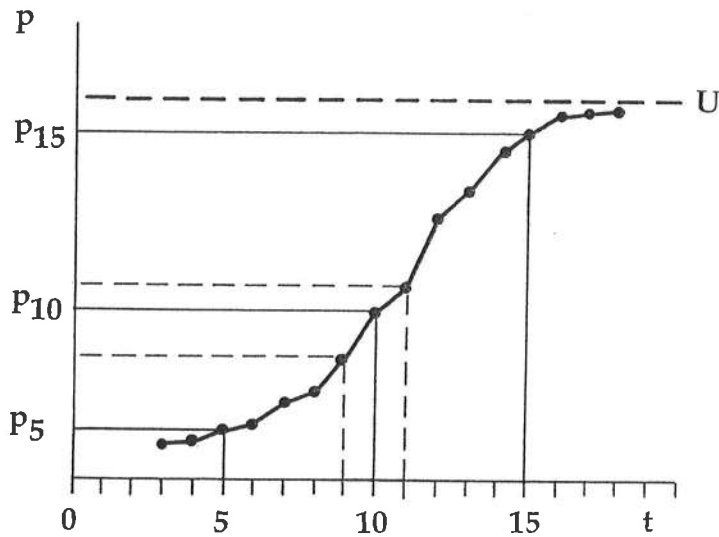
Kuva 8.1. Logistinen käyrä: innovaation omaksujien määrän kumuloituminen ajan funktiona.

Logistisen käyrän parametrien manuaalinen estimointi

Valitaan aika-akselilta kolme vuotta tasavälein, yksi havaintoaineiston alkupäästä, toinen keskeltä ja kolmas loppupäästä. Kuvan 8.2 esimerkissä vuodet ovat t_5 , t_{10} , t_{15} . Vastaavat p :n arvot ovat p_5 , p_{10} , p_{15} . Satunnaisen vaihtelun eliminoimiseksi kukin p on kolmen havainnon geometrinen keskiarvo, esim.

$$p_{10} = \sqrt[3]{p_9 \cdot p_{10} \cdot p_{11}}$$

N on vuosien lukumäärä t_5 :sta t_{10} :een = t_{10} :stä t_{15} :een.



Kuva 8.2.

Empiirinen esimerkki logistisesta käyrästä (ks. selostus tekstissä).

Logistisen käyrän parametrit voidaan nyt estimoida kolmella kaavalla:

$$U = \frac{2p_5p_{10}p_{15} - p_{10}^2(p_5 + p_{15})}{p_5p_{15} - p_{10}^2}$$

$$a = \ln \frac{U - p_5}{p_5} \quad \text{ja} \quad b = \frac{1}{N} \left[\ln \frac{p_5(U - p_{10})}{p_{10}(U - p_5)} \right]$$

b:n tulkinta: Mitä suurempi vakion arvo on, sitä jyrkempi on kuvaaja ja nopeampi uudisteen omaksumisprosessi.

Logistisen käyrän tilastollinen estimointi

$$P_T = \frac{U}{1 + e^{a+bT}}$$

$$P_T(1 + e^{a+bT}) = U$$

$$P_T + P_T e^{a+bT} = U$$

$$P_T e^{a+bT} = U - P_T$$

$$e^{a+bT} = \frac{U - P_T}{P_T}$$

: logaritmoidaan yhtälön molemmat puolet

$$\ln\left(\frac{U - P_T}{P_T}\right) = a + bT$$

: merkitään $Y_T = \ln\left(\frac{U - P_T}{P_T}\right)$

$$Y_T = a + bT$$

T:tä vastaavat Y:n arvot saadaan lasketuksi tekemällä oletuksia U:n tasosta.

On kehitetty myös hienostuneempi menetelmä, jossa U voidaan estimoida aineistosta.

Lähde:

Löytönen, Markku (1986). Logistisen funktion sovitushjelma LogFit. *Helsingin yliopiston maantieteen laitoksen julkaisuja B 27*.

Aineistoasetelma logistisen funktion estimoimiseksi:

T	Innovaation omaksujien	Innovaation omaks. kumulatiivinen määrä	Y_T
Ajanjakso	lkm	lkm %*	$Y_T = \ln\left(\frac{U - P_T}{P_T}\right)$
1	a_1	a_1	$\frac{a_1}{\sum a_T} \cdot 100$
2	a_2	$a_1 + a_2$.
3	a_3	$a_1 + a_2 + a_3$.
.	.	.	.
.	.	.	.
.	.	.	.
n	a_n	$\sum a_T$	100
	$\sum a_T$		

* Kumulatiivisia prosenttiarvoja lasakettaessa perusarvo voi olla myös omaksujien potentiaalinen maksimi (tiedossa oleva tai arvioitu), joka asiayhteydestä ja innovaation etenemisvaiheesta riippuen on yleensä suurempi kuin $\sum a_T$.

HARJOITUS: Tilastollisten menetelmien käyttö tutkimustehtävässä

Aihe

Tulotason alueellinen vaihtelu ja sitä selittävät tekijät, tutkimusalueena maakunta tai useamman maakunnan muodostama aluekokonaisuus.

Menetelmät

Tarkastelu tilastollisten tunnuslukujen sekä korrelaatio- ja regressioanalyysin avulla; tulosten havainnollistaminen graafisten esitysten avulla.

Muuttujat

- 1 y = kunnan tulotaso (veroäyryjä/asukas)
- 2 x_1 = etäisyys lähimpää D-tasolla toimivaan keskukseen
- 3 x_2 = tutkinnon suorittaneiden %-osuus väestöstä
- 4 x_3 = työllisen työvoiman %-osuus väestöstä
- 5 x_4 = maa- ja metsätalouden %-osuus työllisestä työvoimasta
- 6 x_5 = teollisuuden ja rakennustoiminnan %-osuus työll. työvoimasta
- 7 x_6 = palveluelinkeinojen %-osuus työllisestä työvoimasta

Tehtävät

1. Käsitellään aineistoa SPSS -ohjelmalla.
 - a) Tiedoston muodostaminen
 - b) Kuntien tulotasojen keskiarvo, mediaani ja hajonta
 - c) Korrelaatiomatriisi: muuttujat 1-7
 - d) Regressiomallit:

$$y = a_0 + a_1x_1$$

$$y = a_0 + a_1x_1 + a_2x_2 + a_3x_3 + a_4x_4$$
 (step wise -tulostus)
2. Graafiset esitykset
 - a) Tulotasokartta
 - b) Residuaalikartta: kunnan tulotaso - tulotasojen keskiarvo
 - c) Korrelaatiodiagrammit ja regressiosuorat: y/x_1 ja y/x_i (paras selittäjä)
3. Järjestyskorrelaatiokerroin: tulot / työllinen työvoima
4. Työselostus

Tulkitaan tuloksia loogisessa järjestyksessä.
Liitteeksi esitystä tukevat kuvat ja taulukot.

ALUEELLISIA MITTAUS- JA ANALYYSIMENETELMIÄ

1. Alueelliset tunnusluvut

1.1. Sijaintiosamäärä

Sijaintiosamäärä (*location quotient*) on yksinkertainen suhdelukumitta, jonka avulla voidaan yleispiirteisesti tutkia ja tulkita jonkin asian (esim. työpaikat, työttömyys, tuotannollinen toiminta) esiintymistä alueella (esim. kunnassa, seutukunnassa, maakunnassa) suhteessa koko maan vastaavaan tilanteeseen. Sijaintiosamäärä sopii myös muutoksen analysointiin.

$$LQ = \frac{\frac{S_i}{N_i}}{\frac{S}{N}} \quad \text{tai} \quad \frac{\frac{S_i}{S}}{\frac{N_i}{N}}$$

S_i = toimialan i työpaikkojen (tai vast.) määrä alueella,

S = työpaikkojen kokonaismäärä samalla alueella,

N_i = toimialan i työpaikkojen määrä koko maassa,

N = työpaikkojen kokonaismäärä maassa.

LQ voidaan esittää myös %-lukujen osamääränä. Tällöin %-luku $100 \times S_i/S$ kuvaa toimialan i %-osuutta alueen työpaikoista ja %-luku $100 \times N_i/N$ vastaavaa suhdetta koko maassa.

LQ:n tulkinta:

$LQ = 0$, alueella ei ole kyseisen toimialan työpaikkoja,

$LQ = 1$, alueella on toimialan i osuus sama kuin koko maassa,

$LQ > 1$, alueella on toimialan i osuus suurempi kuin koko maassa keskimäärin.

Esimerkki:

Rosenqvist, Olli (1997). *Maatalouden alueellinen erikoistuminen Suomessa*. Talusmaantieteen lisensiaatintutkimus. Vaasan yliopisto.

1.2. Alueellisen assosiaation mittaus

Tutkija voi joutua vertaamaan kahden alueellisen jakauman samankaltaisuutta eli assosiaatiota. Tehtävänä voi olla esim.

- kahden peräkkäisen ajanjakson väestön levinneisyyden vertailu tai
- erilaisten väestöelementtien vertailu samasta poikkileikkaustilanteesta.

Vertailuun voidaan käyttää ns. assosiaatiokerrointa D . Se lasketaan prosenttijakaumien avulla seuraavalla kaavalla

$$D = 1/2 \sum | X_i - Y_i | ,$$

jossa X_i = alueen i %-osuus koko X -populaatiosta,
 Y_i = alueen i %-osuus koko Y -populaatiosta.

Assosiaatiokerroin osoittaa, mikä prosenttiosuus toisesta populaatiosta olisi siirrettävä, jotta päästäisiin toisen populaation levinneisyyteen.

Esim. vertailtavana on valkoisten ja värillisten alueellinen jakautuminen Etelä-Afrikassa, työllisten ja työttömien jakautuminen maakunnittain tai lääneittäin, opiskelupaikkojen ja väestön alueelliset jakaumat Suomessa (Palomäki 1985, Swanljung 1991).

Teoreettinen esimerkki:

Alue	X-popul.	Y-popul.	Erotus
1	15 %	5 %	10 %
2	20	10	10
3	20	20	0
4	30	20	10
5	15	45	30
Yhteensä	100	100	60

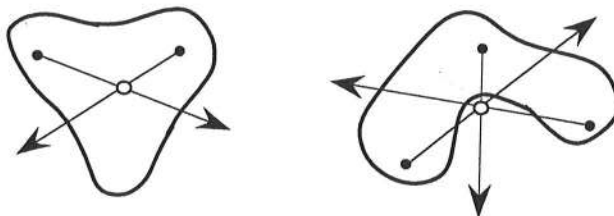
Yllä olevassa esimerkissä: $D = 1/2 \times 60 = 30 \%$.

1.3. Alueelliset keskiluvut

a) Alueen painopiste

Painopiste on se piste kappaleessa, josta ripustettuna kappale pysyy tasapainossa. Alueen painopiste määritetään luotilangan avulla ripustustekniikkaa käyttäen. Alueen painopiste voi tulla myös alueen ulkopuolelle.

Alueen painopiste on "tyhjän" alueen teoreettisesti edullisin piste.



b) Ilmiön levinneisyyden painopiste

Kyseessä voi olla esimerkiksi väestön, asumusten, yritysten tms. sijainnin alueellisen painopisteen määrittäminen. Painopisteeseen vaikuttavina tekijöinä ovat tapausten lukumäärät (painot) ruuduissa ja ruutujen keskipisteen etäisyydet (painotettavat) painopisteestä.

Painopiste määritetään seuraavasti:

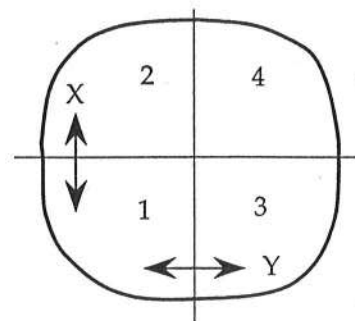
- 1) Valitaan ensin silmämääräinen painopiste jonkin koordinaattiruudun kulmaan.
- 2) Lasketaan x- ja y-koordinaattien korjaukset D_x ja D_y seuraavasti:

$$D_x = \frac{\sum P_{2+4} X_{2,4} - \sum P_{1+3} X_{1,3}}{P} \quad D_y = \frac{\sum P_{3+4} Y_{3,4} - \sum P_{1+2} Y_{1,2}}{P}$$

X = ruudun keskipisteen etäisyys Y-akselista,

Y = ruudun keskipisteen etäisyys X-akselista,

P = tapausten lukumäärä ruudussa.



Huomautuksia:

- ääriarvot vaikuttavat voimakkaasti painopisteen sijoittumiseen,
- kaikki muutokset perusjoukossa vaikuttavat painopisteen sijaintiin, minkä johdosta painopiste soveltuu hyvin alueellisten muutosten analysointiin (esim. USA:n, Suomen ym. väestön painopisteen siirtyminen).

c) Mediaanipiste

Mediaanipisteen kautta kulkeva sekä N-S- että W-E-suuntainen suora jakaa aineiston kahteen yhtä suureen osaan.

Määritetään seuraavasti:

- 1) Valitaan ensin silmämääräinen mediaanipiste jonkin koordinaattiruudun kulmaan.
- 2) Lasketaan x- ja y-koordinaattien korjaukset D_x ja D_y seuraavasti:

$$D_x = \frac{\sum P_{2+4} - \sum P_{1+3}}{2} \qquad D_y = \frac{\sum P_{3+4} - \sum P_{1+2}}{2}$$

Huomautuksia:

- mediaanipistettä määritettäessä etäisyyden vaikutus ajatellaan poistetuksi, aineiston jokaisella osalla on siis samanlainen paino,
- aineiston sisäiset muutokset vaikuttavat mediaanipisteeseen vain, jos jokin alkio ylittää mediaaniviivan,
- mediaaniakselit jakavat koko aineiston yhtä suuriin osiin, mutta eivät siten, että vierekkäisiin neljänneksiin tulisi yhtä monta alkiota,
- jos akselien kulmia muutetaan, mediaanipistekin muuttuu paikkaansa,
- mediaanipiste on vakaampi kuin painopiste, minkä johdosta se sopii painopistettä paremmin jonkin poikkileikkaustilanteen väestön levinneisyyskuvan analysointiin.

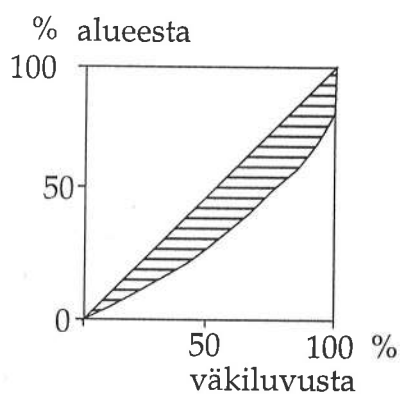
d) Modaalipiste

Modaalipisteessä tapausten frekvenssi on suurin. Sen avulla voidaan karkeasti vertailla väestön levinneisyyttä esim. eri valtioissa, maakunnissa ja jopa kunnissa (korologisen matriisin avulla).

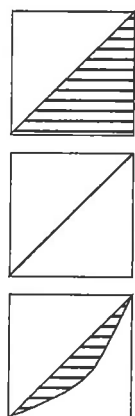
- 1) Unimodaaliset alueet
Alueita, joissa on vain yksi suuren konsentraation piste, esim. Englanti/Lontoo, Ranska/Pariisi, Suomi/Hki, Keski-Suomi/Jyväskylä.
- 2) Bi-modaaliset alueet
Alueita, joilla on kaksi suurin piirtein tasavertaista konsentraatiopistettä, esim. Australia/Sidney - Melbourne, Kymenlaakso/Kotka - Kouvola.
- 3) Tri-modaaliset alueet, esim. Kotka - Kouvola - Lappeenranta.

1.4. Lorentz-käyrä

Keksijänsä M.O. Lorentzin (1905) mukaan nimetty Lorentz-käyrä on kahden muuttujan suhteen laadittu kumulatiivinen summakäyrä, jonka avulla voidaan verrata eri alueiden väestön, tulojen tai jonkin muun ilmiön agglomeroitumisasteita. Lorentz-käyrä soveltuu mainiosti myös agglomeroitumisasteen muutoksen kuvaamiseen.



Agglomeraatioaste = viivoitetun alueen pinta-ala prosentteina neliön puolikkaan alasta



Agglomer.aste	Tulkinta
100 %	täydellinen keskittyminen (klusteroituminen) yhteen alueyksikköön
0 %	täysin tasainen jakauma
20-30 %	jonkin verran klusteroitunut

Lorentzin käyrän määrittäminen, kun lähtökohtana on väestöpistekartta

- 1) Ruutumatriisi (esim. 1 cm x 1 cm ruudukko), johon merkitään väestöpisteiden lukumäärä/ruutu.
- 2) Taulukko ruutumatriisin lukumäärätietojen perusteella. Taulukkoon merkitään väestöpisteiden lukumäärät ja vastaavat ruutujen lukumäärät väestöpisteiden lukumäärien mukaisessa suuruusjärjestyksessä sekä näiden perusteella pisteiden ja ruutujen kumulatiiviset summa- ja %-arvot.
- 3) Taulukon %-lukujen perusteella laaditaan Lorenzin käyrä ja lasketaan keskittyneisyysaste.

Esimerkki Lorentzin käyrän määrittämisestä (Hannerberg 1961)

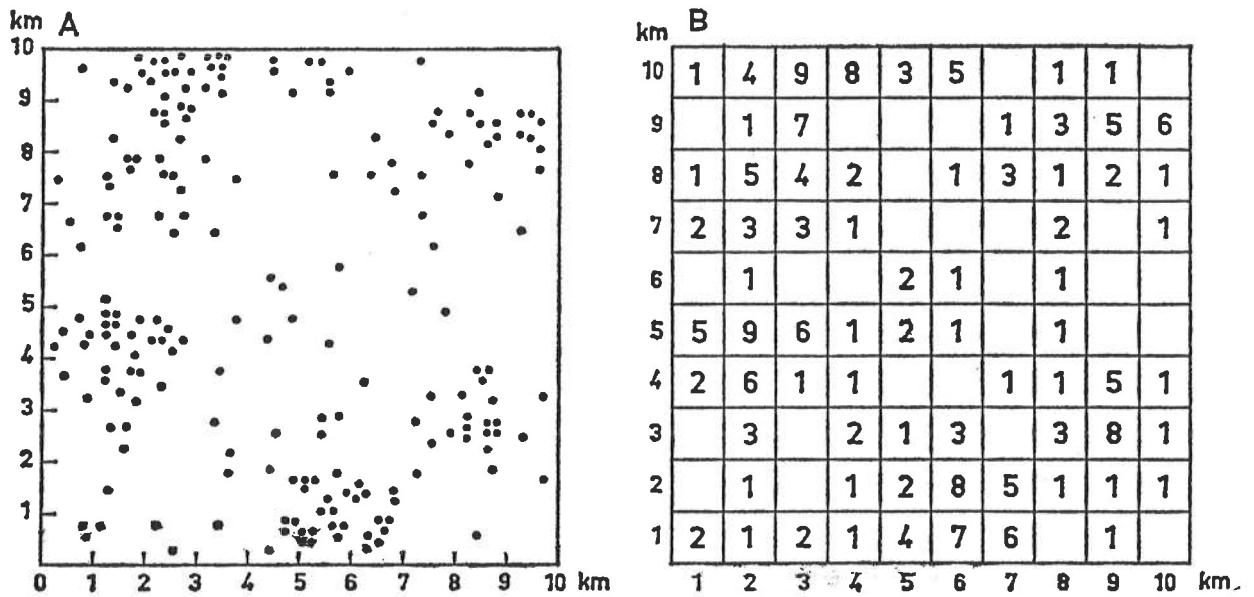
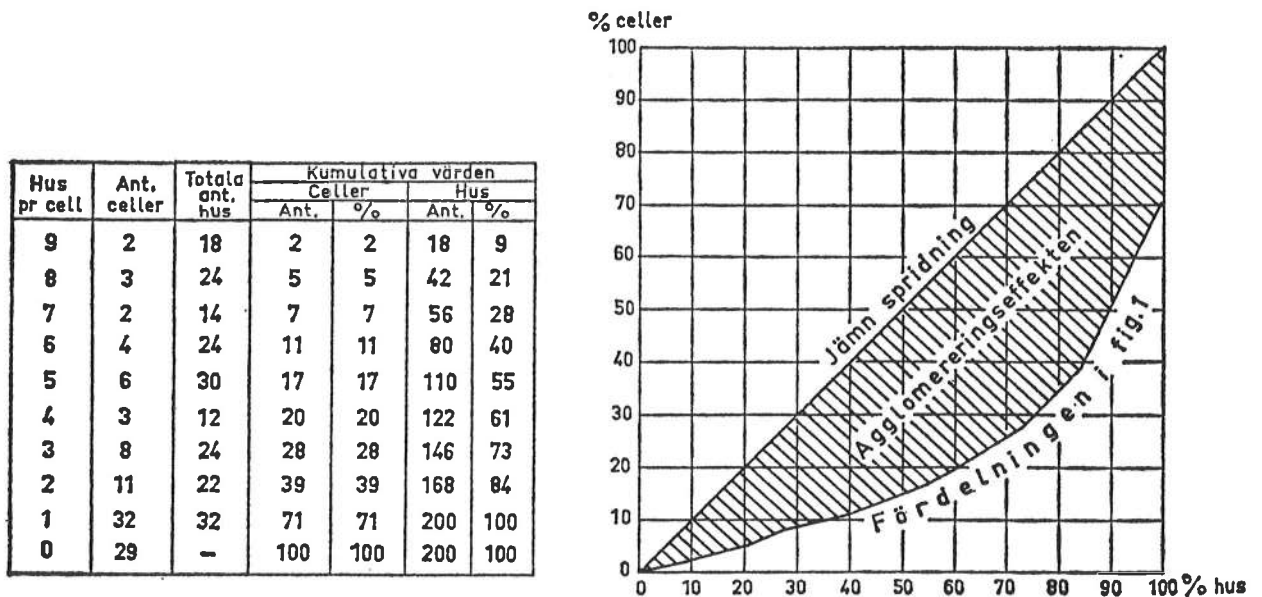


Fig. 1. T.v. Bostadshus inom ett 10 km x 10 km stort område. För att göra uträkningarna lätt överskådliga har antalet hus avrundats till 200. Bilden åsyftar närmast mellansvenska jordbruksområden med små slätter, sjöar och glesst bebyggda skogar. T.h. Fördelningen i fig. 1 A i form av en reguljär matris med 100 celler à 1 km².



Lorentz-diagram för bestämning av agglomereringsgraden hos den korologiska populationen i fig. 1 B. Agglomereringsgraden i figuren visar sig vid uträkning vara 0.59.

Lähteet:

Hannerberg, David (1961). *Att studera kulturgeografi*. Andra upplagen. Stockholm.

Lorentz, M.O. (1905). Methods of measuring the concentration of wealth. *Journal of American Statistical Association*, no. 70.

2. Gradienttianalyysi

Gradienttianalyysi on tutkimusmenetelmä, jossa tutkittavan asian alueellinen muuttuminen tietyistä mittauskeskipisteistä lähtien pyritään "vangitsemaan" kaksiulotteisen diagrammin muotoon. Tämä tapahtuu siten, että havainto-alueyksiköistä, esim. kunnista, muodostetaan kehä mittauskeskipisteen ympärille ja kullekin kehälle lasketaan ko. kehän havaintoarvojen perusteella keskimääräinen arvo. Mittauskeskipiste voi asiayhteydestä riippuen olla esim. väestöllinen keskipiste, maan arvon tai maan käytön intensiteetin huippua edustava piste tms. Etäisyysvyöhykkeet voidaan muodostaa myös säännöllisten ympyräkehien avulla, jolloin etäisyysluokat voivat olla tasavälisiä, geometrisen jonon tai jonkin muun järjestelmän mukaan asteittain suurenevia. Etäisyysvyöhykkeittäin lasketut arvot sijoitetaan tämän jälkeen diagrammiin, jossa analyysin kohteena oleva muuttuja on pystysuoralla akselilla ja etäisyys vaakasuoralla akselilla.

Gradienttianalyysin graafinen lopputulos saadaan yksinkertaisimmin yhdistämällä diagrammipisteet murtoviivalla. Johtopäätökset diagrammista voivat perustua yksinomaan visuaaliseen havainnointiin. Toinen, vaativampi tie on etsiä ja estimoida diagrammin pistejoukkoa mahdollisimman hyvin edustava funktio ja vetää johtopäätöksiä funktion kuvaajan ja parametrien avulla (estimointimenetelmistä, ks. Tilastollisia mittaus- ja analyysimenetelmiä, luku 5.3).

Gradienttianalyysin avulla voidaan tutkia esim. väentihedden vaihtelua keskuksittaisen etäisyyden ja ajan funktiona (esim. tutkimus Tukholmasta) tai tulotason vaihtelua suurten työpaikkakeskusten ympäristökunnissa.

Tutkimusesimerkki

Tutkimusesimerkissä (Mikkonen 1978: 61–64) gradienttianalyysillä havainnollistetaan maalaiskuntien tulotason muuttumista kaupunkitaso keskuksista etäännyttäessä. Tutkimusaineisto ryhmiteltiin etäisyysvyöhykkeisiin 5 km:n välein 30 km:iin asti ja siitä eteenpäin 10 km:n välein. Etäisyys kunnan pääkeskuksesta lähimpään kaupunkitaso keskukseseen määräsi kunnan sijoittumisen etäisyysluokkaan. Kussakin etäisyysluokassa laskettiin tulotasojen ja etäisyyksien keskiarvot, jotka määrivät pisteen paikan kaksiulotteisessa diagrammissa. Tällainen diagrammi laadittiin erikseen kunkin läänin aineistosta ja koko maan aineistoista. Tiedot ovat vuodelta 1975.

Diagrammisarja osoittaa ensinnäkin, että tulotaso laskee, kun etäisyys kaupunkitason keskukseen kasvaa ja toiseksi laskusuunnassa esiintyy porras koko maan ja useimpien läänien diagrammissa 20 km:n kohdalla tai välillä 10–20 km. Selvimmillään tämä porras on Hämeen, Kymen, Mikkelin, Keski-Suomen ja Oulun läänin diagrammeissa. Pohjois-Karjalan ja Kuopion läänin diagrammeissa mahdollinen porras ei ”ehdi” tulla näkyviin, koska ensimmäiset tulotasokeskiarvot ovat vasta 20 km:n paikkeilta. Lapin läänistä vastaavaa diagrammia ei voitu piirtää lainkaan, koska 18 maalaiskunnasta 12 sijoittui etäisyysluokkaan yli 60 km loppujen hajotessa välille 20–60 km yksi tai korkeintaan kaksi kuntaa etäisyysluokkaa kohti. Tulotasoportaan esiintyminen on yhteydessä kunnanrajan ylittävään työssäkäyntiliikenteeseen eli sukkulointiin. Useissa tutkimuksissa on yhtäpitävästi osoitettu, että ainakin Suomen suurimpien kaupunkien pendeliliikenteen määrissä tapahtuu jyrkkä lasku 15–20 km:n etäisyydellä kaupungin keskustasta.

Etäisyysluokasta 20–25 lähtien tulotaso laskee loivasti tai pysyy lähes vakiona (esim. Mikkelin, Pohjois-Karjalan, Kuopion ja Vaasan läänit sekä koko maan aineisto). Eräiden läänien diagrammeissa esiintyy paikallisiin erityisolosuhteisiin viittaavia pykäliä (esim. Hämeen, Keski-Suomen ja Oulun läänit).

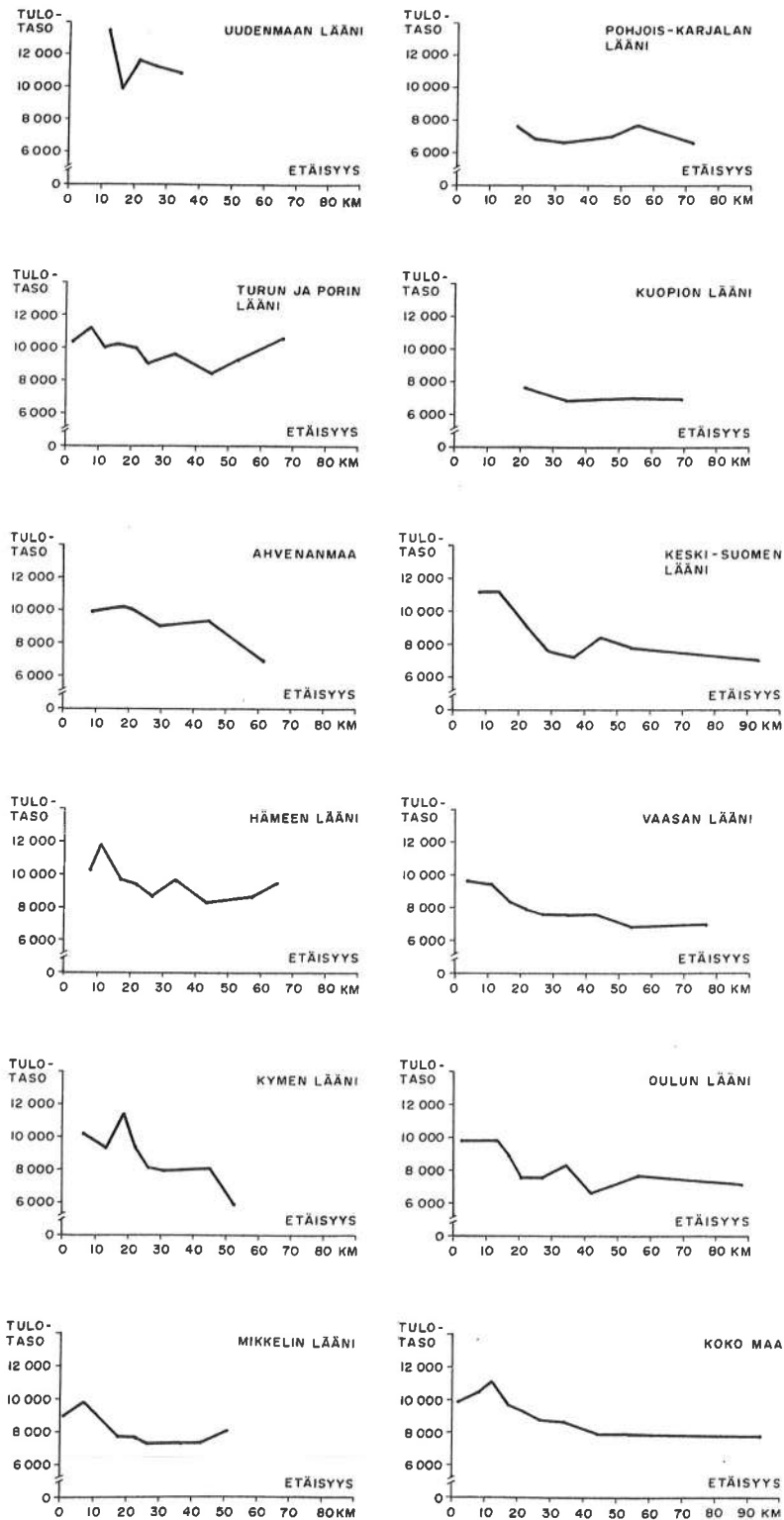
Riittävän etäälle (40–50 km) kaupunkitason keskuksesta päästäessä useiden läänien tulotasomurtoviiva osoittaa jälleen nousun merkkejä. Näin käy esim. Turun ja Porin, Hämeen, Mikkelin ja Pohjois-Karjalan lääneissä. Tällä noususuunnalla ei ole enää yhteyttä sukkulointiin alkuperäisissä kaupunkitason tutkimuksissa. Noususuunta onkin selitettävissä Reino Ajon käyttöön ottaman käsitteen *rajaseudun markkinat* avulla. Rajaseudun markkinoilla Ajo tarkoittaa liikennealueiden rajavyöhykkeen olosuhteita, joissa suuret etäisyydet pääkeskuksiin ja näiden välinen kilpailu muodostavat erikoisen edun ja tilaisuuden oman keskuksen kehittämiseen, ostovoiman keskittämiseen, työpaikkojen muodostamiseen ja sitä kautta ansiomahdollisuuksien lisääntymiseen (Ajo 1944: 283–285 ja 1946).

Tutkimusesimerkki:

Mikkonen, Kauko (1978). Tulojen alueellinen vaihtelu Suomessa. *Vaasan kauppa- ja korkeakoulun julkaisuja. Tutkimuksia* No 53 (s. 61–64).

Lähteet:

Ajo, Reino (1944). Tampereen liikennealue. *Kansantaloudellisia tutkimuksia* 13.
Ajo, Reino (1946). Liikennealueiden kehittyminen Suomessa. *Fennia* 69:3.



Kuva 2.1. Maalaiskuntien tulotason riippuminen etäisyydestä lähimpään kaupunkitason keskukseen kussakin läänissä ja koko maassa 1975 (Mikkonen 1978: 63).

3. Suodatinkartoitus

Suodatinkartoitus (*filtering mapping*) on menetelmä, jonka avulla voidaan yleistää kartalla olevia levinneisyystietoja.

Käyttötarkoitukset

1) Ilmiön levinneisyyden analyysi

Pyrkimyksenä on saada esiin

- a) mahdollinen alueellinen trendipinta ja/tai
- b) tutkia yksittäisten havaintojen poikkeamia trendipinnasta ja selittää poikkeamien syitä.

Levinneisyyden analyysissä on siis yleensä tavoitteena havaita ja erottaa tutkitavan ilmiön paikalliset (*local*) ja alueelliset (*regional*) piirteet alueella (*region*). Alueelliset piirteet edustavat ilmiön muuttumisen suurpiirteistä vaihtelua, eli ne kuvaavat muuttumisen suuntaa ja voimakkuutta (= muuttumisen alueellinen trendi). Paikalliset piirteet ovat ilmiön pienpiirteistä vaihtelua eli hajontaa trendipinnan suhteen.

2) Havaintoaineiston sattumanvaraisten heilahtelujen tasoittaminen

Tarpeellinen esim. silloin, kun aineisto muodostuu vastaajien subjektiivisista arvioista, jotka kohdistuvat kvantitatiivisen määrän ilmaisemiseen.

3) Harvan pisteverkon tihentäminen (esim. Tuominen 1949) tai puuttuvien arvojen interpolointi sellaisiin korologisen matriisin ruutuihin, joista ei ole saatu vastausta, mutta joissa asuu potentiaalisia vastaajia. Tiheämpi havaintopisteverkko helpottaa isaritmikartan laatimista. Suodatinkartoitusta ei ole mielekästä kuitenkaan laskea alueen sellaisiin pisteisiin tai ruutuihin, joissa ilmiötä ei voi esiintyä.

Lähtötiedot

1) Pistekartan muodossa, jolloin

- a) pisteiden arvo voi olla mittaustuloksen mukaan vaihteleva,

b) pisteiden yksikköarvo voi olla aina sama (esim. 1 piste = 40 asukasta). Tällöin pistekartta ennen suodatinkartoitusta tavallisesti muunnetaan korologisen matriisin muotoon eli lähtötiedostona onkin pisteiden määrä/matriisin ruutu.

- 2) Korologisen matriisin muodossa, esim.:
 - a) Säännöllisiä monikulmiot kuten neliöt, 6-kulmiot jne. Tällöin havaintoarvot ajatellaan sijoitetuiksi alueiden keskipisteisiin.
 - b) Hallinnolliset alueet (koulupiirit, kunnat, seutukunnat, maakunnat jne.), jolloin havaintoarvot ajatellaan sijoitetuiksi alueiden väestöllisiin tai muihin paino- tai mediaanipisteisiin.

Menetelmät

1) Ympyrätasointusmenetelmä

Ympyrätasointusmenetelmää käytetään, kun mittaukset on esitetty pisteverkon avulla eikä niitä haluta tai viitsitä muuntaa korologisen matriisin muotoon. Menetelmä on joustavampi kuin ruutumatriisin käyttö, mutta se on toisaalta työläs, jos havaintopisteitä on paljon. Menetelmä on manuaalinen.

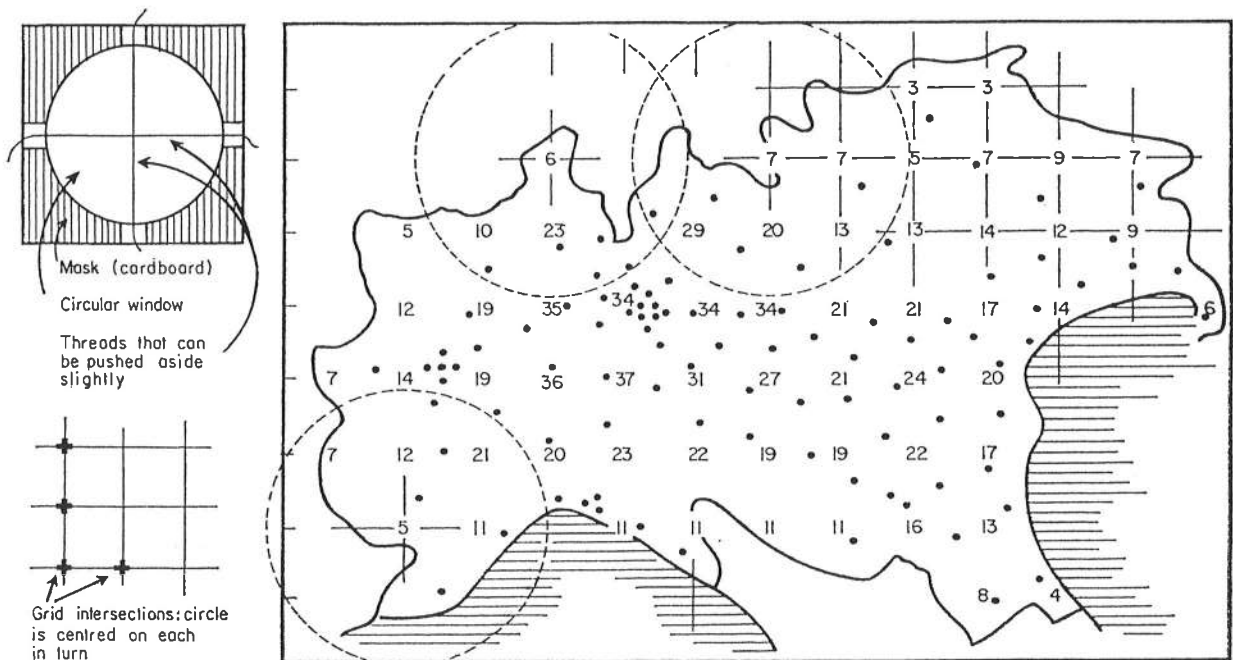
Vaiheet

- a) Laaditaan ruutuverkko, joka sijoitetaan havaintopisteverkon päälle. Ruutuverkon ruutujen koko vaikuttaa lopputulokseen (kuva 3.1).
- b) Tehdään ympyrämuotoinen maskipiirros läpinäkyvälle paperille. Ympyrän säteellä suuri vaikutus lopputulokseen:
 - pieni säde - pieni tasointus
 - suuri säde - voimakas tasointus
- c) Sijoitetaan ympyrämaskin keskipiste kuhunkin leikkauspisteeseen (ruudun kulmaan) ja lasketaan, montako havaintoa jää ympyrän sisään. Merkitään lukumäärä ko. leikkauspisteeseen. Tuloksena on tasointettu levinneisyyskuva.

Jos alkuperäisissä havaintopisteissä tutkittava ilmiö on mittauksissa saanut eri

arvoja, tasoitettuun esitykseen päästään laskemalla havaintoarvojen summat ympyrän sisällä ja jakamalla havaintojen lukumäärällä eli laskemalla ympyrän kussakin sijaintipisteessä havaintoarvojen keskiarvot ympyrän sisältä.

- d) Piirretään isaritmit uusien arvojen ja uuden pisteverkon perusteella. Näin esim. pistekartasta (jos pisteellä aina sama arvo) voidaan laatia suoraan väestön levinneisyyden isaritmiesitys, jonka yleistysastetta voidaan säädellä muuttamalla ympyrän sädettä ja/tai ruutuverkon tiheyttä.



Kuva 3.1. Esimerkki liukuvan keskiarvon ympyrätasointekniikasta (Cole & King 1969: 205).

2) Ruutumatriisitasoitus

Voidaan käyttää erilaisia alueellisen liukuvan keskiarvon menetelmiä, jotka ovat analogisia aikasarjojen liukuvan keskiarvon menetelmän kanssa. Tasoitettavia havaintoarvoja voidaan tarvittaessa painottaa. Laskelmat on kätevintä suorittaa tietokoneella, jolloin kukin ruutu paikannetaan koordinaattien avulla. Tieto-

koneen käyttö antaa tilaisuuden kokeiluihin. Tietokoneen ansiosta ruutumatriisi-tasointusmenetelmä on vähemmän työläs verrattuna ympyrätasointukseen.

Suodattimena käytetään valittua matriisin ruutujen lukumäärää. Ruutuihin osuneiden havaintojen avulla keskusruudulle lasketaan uusi arvo, joka on

$$\frac{\text{Havaintoarvojen summa}}{\text{Havaintojen lkm}}$$

Havaintojen lukumäärä voi vaihdella ruuduittain tai kussakin ruudussa voi olla vain yksi havaintoarvo tai ruudun keskiarvotieto. Havaintojen lkm on tällöin sama kuin suodattimen ruutujen lukumäärä.

Esimerkkejä:

$$\begin{array}{|c|c|} \hline 5 & 10 \\ \hline 15 & 30 \\ \hline \end{array} \quad \frac{60}{4} \qquad \begin{array}{c} 20 \\ 5 \quad 15 \quad 10 \\ 15 \end{array} \quad \frac{65}{5} \qquad \begin{array}{|c|c|c|} \hline 15 & 25 & 17 \\ \hline 16 & 30 & 10 \\ \hline 16 & 18 & 8 \\ \hline \end{array} \quad \frac{153}{9}$$

Tasointuksen voimakkuuteen voidaan vaikuttaa

- suodattimen koolla: mitä suurempi suodatin sitä voimakkampi tasointus,
- toistamalla tasointuksia riittävän monta kertaa,
- painottamalla keskusruudun ja sivuruutujen havaintoarvoja eri tavoin: mitä suurempi on keskusruudun painokerroin sitä pienempi tasointus.

Painotetun suodatinmenetelmän laskukaava:

$$\bar{Y} = \frac{m \cdot A_k + n \cdot A_s}{m \cdot N_k + n \cdot N_s}$$

- jossa
- A_k = havaintoarvojen summa keskusruudussa,
 - A_s = havaintoarvojen summa ympäröivissä ruuduissa,
 - N_k = havaintojen lkm keskusruudussa,
 - N_s = havaintojen lkm sivuruuduissa,
 - m = keskusruudun havaintojen painokerroin,
 - n = sivuruutujen havaintojen painokerroin.

Matriisin reunaruutujen muodostama ongelma. Reunaruutuihin ei voida laskea tasoitusta normaalilla tavalla. Vastaavan ongelman muodostavat aikasarjoissa sarjojen päät.

Menettelytavat:

- lasketaan jäljellä olevien ruutujen keskiarvo eli muutetaan suodattimen kokoa ja tasoituksen tehoa,
- jätetään pois,
- otetaan mukaan ylimääräinen reunaruuturivi, jos se on mahdollista.

1	2	3 a	3 b	3 c																																													
<table border="1" style="border-collapse: collapse; width: 100px; height: 100px;"> <tr><td></td><td>5</td><td>10</td></tr> <tr><td>12</td><td>20</td><td>6</td></tr> <tr><td>30</td><td>52</td><td>40</td></tr> </table>		5	10	12	20	6	30	52	40	<table border="1" style="border-collapse: collapse; width: 100px; height: 100px;"> <tr><td></td><td>5</td><td>16</td></tr> <tr><td>12</td><td>72</td><td></td></tr> <tr><td>30</td><td></td><td>40</td></tr> </table>		5	16	12	72		30		40	<table border="1" style="border-collapse: collapse; width: 100px; height: 100px;"> <tr><td></td><td></td><td></td></tr> <tr><td></td><td>36</td><td></td></tr> <tr><td></td><td></td><td></td></tr> </table>					36					<table border="1" style="border-collapse: collapse; width: 100px; height: 100px;"> <tr><td></td><td></td><td></td></tr> <tr><td></td><td>22</td><td></td></tr> <tr><td></td><td></td><td></td></tr> </table>					22					<table border="1" style="border-collapse: collapse; width: 100px; height: 100px;"> <tr><td></td><td></td><td></td></tr> <tr><td></td><td>25</td><td></td></tr> <tr><td></td><td></td><td></td></tr> </table>					25				
	5	10																																															
12	20	6																																															
30	52	40																																															
	5	16																																															
12	72																																																
30		40																																															
	36																																																
	22																																																
	25																																																

Kuva 3.2. Esimerkki painotetun liukuvan keskiarvon ruutumatriisitasoituksesta. 1. Asiointifrekvenssit ruuduittain. 2. Asiointifrekvenssien summa ruuduittain. 3. Asiointifrekvenssien keskiarvo keskusruudussa: a. aritmeettinen keskiarvo; b. painottamaton liukuva keskiarvo; c. painotettu liukuva keskiarvo, keskusruudun paino on 2, muiden ruutujen 1. (Mikkonen 1972: 44–45.)

Lähteet:

Cole, John P. & Cuchlaine A.M. King (1968). *Quantitative Geography*. London etc.: John Wiley & Sons, Ltd. (s. 201–207).

Mikkonen, Kauko (1972). The gravitation field of Forssa. A gravitational analysis of the functional system. *Fennia* 115.

4. Trendipinta-analyysi

Alueyksiköittäin, esim. kuntapohjalla tai säännöllisessä korologisessa matriisissa esitetyt mittaustulokset voidaan analysoida ikään kuin ne edustaisivat jatkuvan pinnan havaintoja. Menetelmänä käytetään *trend surface* -analyysiä. Suomenkielinen nimitys tälle menetelmälle on *trendipinta-analyysi*. Etuliite trendi viittaa siihen komponenttiin mittaustuloksia, joka aiheuttaa tasaista vaihtelua erotuksena paikallista vaihtelua synnyttävästä komponentista. Trendipinta-analyysin eräänä tavoitteena onkin konstruoida esiin mittaustulosten vaihtelun alueellinen trendi, jos sellainen on olemassa.

Aiemmin käsiteltyä suodatinkartoitusta voidaan pitää eräänä trendipinta-analyysin apuvälineenä. Yleisimmin trendipinta-analyysillä tarkoitetaan kuitenkin pinnan matemaattista mallintamista, joka tapahtuu monen muuttujan regressio-analyysin keinoin. Esim. tulotasotutkimukseen sovellettuna tämä tapahtuu siten, että tulotason vaihtelua selitetään havaintoyksiköiden eli kuntien sijainnilla, joka ilmaistaan x- ja y-koordinaattien avulla, ja regressioyhtälön parametrit estimoidaan normaaliin tapaan pienimmän neliösumman menetelmällä. Yksinkertaisinta regressiopintaa kutsutaan lineaariseksi eli ensimmäisen asteen pinnaksi. Pinnan kompleksisuus lisääntyy regressioyhtälön astelukua korotettaessa.

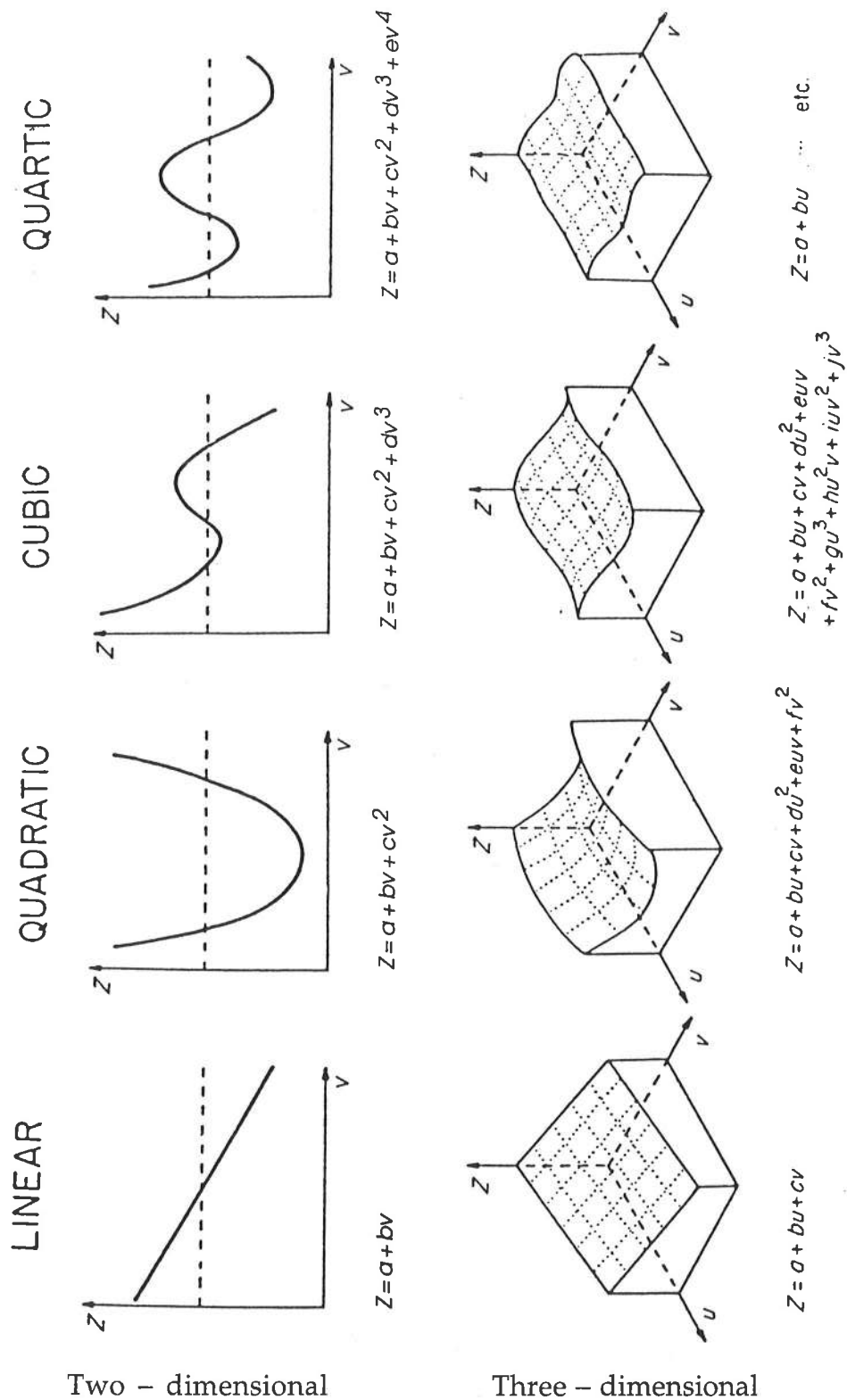
Ensimmäisen, toisen ja kolmannen asteen regressiopintojen perusmuodot ovat:

lineaari: $z = a_0 + a_1x + a_2y$

neliö: $z = a_0 + a_1x + a_2y + a_3x^2 + a_4xy + a_5y^2$

kuutio: $z = a_0 + a_1x + a_2y + a_3x^2 + a_4xy + a_5y^2$
 $+ a_6x^3 + a_7x^2y + a_8xy^2 + a_9y^3$

Neliö- ja sitä korkeamman asteen trendimallia estimoitaessa malliin eivät välttämättä tule mukaan kaikki alemman asteen 'trendikomponentit'.



Kuva 4.1. Eriasteisten trendipintojen graafinen havainnollistus (Wilson & Bennett 1985: fig. 5.1; julkaistu John Wiley & Sons Inc.:n luvalla).

Alkuperäisestä kuvasta poiketen muuttujan u tilalle on kaksidimensionaaliin diagrammeihin ja vastaaviin kaavoihin vaihdettu muuttuja v .

Kaikki selitettävän muuttujan arvot kolmiulotteisessa avaruudessa eivät suinkaan sijaitse estimoidulla trendipinnalla, vaikka pinnan "fittaus" eli sovittaminen datajoukkoon olisi onnistunut erinomaisesti. Useimmat empiirisistä arvoista ovat joko pinnan yläpuolella eli ovat suurempia kuin regressioyhtälöstä asianomaisilla selittävien muuttujien arvoilla saadut tulokset tai ovat pinnan alapuolella eli pienempiä kuin lasketut arvot. Usein trend surface -analyysin keskeisenä tavoitteena onkin poimia esiin alueellisesta trendipinnasta erityisen paljon poikkeavat havaintoyksiköt ja pyrkiä löytämään syitä näille poikkeamille. Poikkeamien eli residuaalien suuruuden luokittelumittana voidaan käyttää - kuten poikkeamia keskiarvosta tai mediaanista luokiteltaessa - havaintoarvojen keskihajontaa regressiopinnan suhteen (kuva 4.2).

Trend surface -analyysin yhteydessä lasketaan normaalin regressioanalyysin tapaan erilaisia tilastollisia tunnuslukuja mallin ominaisuuksien testaamiseksi. Mallin hyvyyttä mitataan yhteiskorrelaatiokertoimen (R) avulla muunnettuna selitysasteen muotoon ($100 R^2$). Se ilmaisee, kuinka monta prosenttia malli eli regressiopinta selittää havaintoarvojen kokonaisvarianssista. F-testillä testataan koko mallin merkitsevyys ja t-testillä kunkin regressiokertoimen merkitsevyys eli poikkeaminen 0:sta.

Tutkimusesimerkki

Tarkastelun kohteena on maalaiskuntien tulotason alueellinen vaihtelu Suomessa vuonna 1975. Olennaista esimerkissä on menetelmän esittely, ei poikkileikkausajankohta eikä käsitteen maalaiskunta poistuminen uuden kuntalain nimikkeistöstä.

Lähtökohtana trend surface -analyysitulosten käsittelylle tarkastellaan ensin maalaiskuntien tulotasojen poikkeamia maalaiskuntien valtakunnallisesta mediaanitulotasosta (kuva 4.3). Kolmiulotteisessa tilassa tätä tulotasoa edustava tasopinta on vaakasuorassa asennossa ja 8522 yksikön korkeudella 0-tasosta. Teollisuus-Suomen maalaiskunnat nousevat runsaslukuisesti ja selvästi (tulotaso suurempi kuin $md + 1 s$) valtakunnallisen md -tulotason yläpuolelle. Maan keskiosien poikki kulkee leveä vyöhyke, jossa tulotasot jäävät yleisesti tämän tason alapuolelle. Tässä vyöhykkeessä Vaasan läänin rannikolta löytyy muutamia positiivisen residuaalin maalaiskuntia (kaupunkien naapurikuntia), Pohjois-Karjalan läänissä ei ainuttakaan. Keski-Suomessa negatiivisten residuaalien luonnehtimasta yleiskuvasta poikkeavat näkyvimmin Jyväskylän ympäristön useat positiivisen residuaalin kunnat, Lapin läänissä ja Oulun läänin itäosissa laajat valkoiksi jääneet alueet (tulotaso välillä $md \pm 1/2 s$).

Alueelliset säännönmukaisuudet tulotasojen poikkeamisessa mediaanitulotasosta kertovat vakuuttavasti siitä, että tulotason vaihtelussa on tiettyä trendimäisyyttä. Se voidaan kvantifioida karkeasti ensimmäisen asteen lineaaripinnan avulla, jonka yhtälö on

$$(I) \quad z = 10658,7 - 15,43 x - 11,11 y .$$

x- ja y-muuttujien negatiiviset regressiokertoimet osoittavat, että tulotasoa kuvaava lineaaripinta laskee etelästä pohjoiseen ja lännestä itään. Tätä havainnollistavat myös kuva 4.7 a ja b. Kuvissa on esitetty esimerkinomaisesti vain yksi pohjois(N)-etelä(S)- ja yksi länsi(W)-itä(E)suuntainen profiili. Periaatteessa koko regressiotaso voitaisiin hahmottaa esim. peräkkäisten N-S-suuntaisten profiilien avulla. Ne olisivat keskenään samansuuntaisia, mutta eri korkeudella 0-tasosta (lännessä korkeammalla kuin idässä). Mallin I regressiokertoimien itseisarvojen suuruuden perusteella voidaan päätellä, että laskusuunta pohjoiseen on hieman jyrkempi kuin itään. Molemmat kertoimet poikkeavat erittäin merkitsevästi 0:sta ja koko malli on tässä tilanteessa luonnollisesti merkitsevä 0,1 %:n riskitasolla (taulukko 4.1). Mallin merkitsevyydestä huolimatta selitysaste ei nouse sen korkeammaksi kuin 20,4 %:iin. Myös keskihajonta regressiotason ympärillä on varsin suuri ($s = 1877$).

Kalteva regressiotaso asettuu ainakin Etelä-Suomessa empiiriseen havaintojoukkoon paremmin kuin vaakasuora mediaanitulotaso. Tämä näkyy kuvan 4.4 residuaalikartalla siten, että + -residuaaleja on nyt harvemmassa kuin kuvassa 4.3, ja joukossa esiintyy muutamia - -residuaalejakin (erityisesti Ahvenanmaan saaristossa sekä Turun ja Porin läänissä). Ne osoittavat, että kyseisen kunnan tulotaso on jäänyt vähintään 1/2 s:n päähän lineaaritasosta, sen alapuolelle. Tässä residuaalikartassa alkaa niin ikään hahmottua + -merkkien kasautuminen voittopuoleisesti kaupunkien ympäristökuntiin. Väli-Suomi tulee esiin negatiivisine residuaaleineen ja Jyväskylän ympäristö tästä säännöstä poikkeavine kuntineen kuten kuvassa 4.3. Pohjois-Suomen tulotasojen estimaatiksi lineaarinen regressiopinta sopii huonosti. Useimpien Lapin läänin maalais kuntien tulotaso on selvästi korkeampi (yli 1 s) kuin regressiomallin mukainen laskennallinen tulos.

Lineaarinen regressiopinta on siis vielä kovin karkea koko maan maalais kuntien tulotasojen alueellisen trendin osoittaja. Se ei voi samanaikaisesti ottaa huomioon Etelä-Suomen korkeiden tulotasojen trendiä, Väli-Suomen matalia arvoja sekä Lapin läänin keskimääräistä tasoa edustavia tulotasoja. Toisen asteen regressiopinta on

$$(II) \quad z = 10577,9 - 63,42 x + 31,03 y + 0,20 x^2 + 0,09 xy - 0,26 y^2 .$$

Se edustaa todellisuutta selvästi paremmin kuin lineaaripinta. Mallin selitysaste on 33 % eli lähes 13 %-yksikköä korkeampi kuin lineaarimallin (I) ja hajonta selvästi

pienempi ($s = 1728$, ks. taulukko 4.1). Malli on F-testin mukaan erittäin merkitsevä, mikä tarkoittaa sitä, että mallissa on ainakin yksi (vastakohtana 0-hypoteesille: ei yksikään) selittävä tekijä, jonka regressiokerroin poikkeaa 0:sta 0,1 %:n riskitasolla. Näitä ovat t-testin mukaan muuttujien x ja x^2 regressiokertoimet. Muuttujien y ja y^2 kertoimien poikkeaminen 0:sta on merkitsevää luokkaa.

Mallin II edustama pinta on satulapinta, jonka N-S-suuntaiset profiilit (paraabelit) aukeavat ylöspäin ja W-E-suuntaiset profiilit alaspäin. Tämä ilmenee mallista, tarkemmin sanottuna sen toisen asteen termien kertoimien etumerkeistä ja esimerkkiprofiileista kuvissa 4.7 a ja b. Profiilit edustavat mallista II saatuja yhtälöitä

$$z = 10991,9 - 53,73 x + 0,20 x^2, \text{ kun } y = 105 \quad (\text{kuva 4.7 a})$$

$$z = 6348,4 + 39,71 y - 0,26 y^2, \text{ kun } x = 94 \quad (\text{kuva 4.7 b})$$

Neliöpinta kuvaa keskeisiltä osin verraten hyvin tulotason vaihtelun alueellista trendiä. Eräissä kohdin se jättää kuitenkin toivomisen varaa. Esimerkiksi Perä-Pohjolassa neliöpinnan kuvaama jatkuvasti nouseva suunta (ks. kuva 4.7 a) on liian radikaali eikä vastaa todellisuutta. Itä-länsisuunnassa neliöpinnan alaspäin vievä suunta on nimenomaan Vaasan läänissä päinvastainen kuin tulotason vaihtelun todellinen trendi (ks. kuva 4.7 b). Viimeksi mainittu ristiriita tuottaa sen tuloksen, että aikaisempiin residuaalitarkasteluihin nähden poikkeuksellisen monet rannikon maalaiskunnat saavat nyt residuaalin, joka nähdään yhtenä (tulotaso korkeampi kuin $z + 1/2 s$) tai kahtena (tulotaso korkeampi kuin $z + 1 s$) + -merkinä kuvassa 4.5. Nämäkin puutteet korjautuvat pitkälti, kun regressiopinnan astelukua korotetaan vielä yhdellä eli siirrytään tarkastelemaan kuutiopintaa.

Kolmannen asteen regressiopinnan yhtälö on

$$(III) \quad z = 9658,8 - 49,03 x + 67,46 y + 0,77 x^2 - 1,57 xy - 0,20 y^2 - 0,002 x^3 \\ + 0,001 x^2 y + 0,008 xy^2 - 0,001 y^3 .$$

Malli selittää tulotason kokonaisvaihtelusta 38 %, keskihajonta on 1670 (taulukko 4.1). Mallin selityskyvyssä x^2 - ja x^3 -termit näyttävät ratkaisevinta osaa. Niiden kertoimet poikkeavat t-testin mukaan 0:sta erittäin merkitsevästi. Selvästi vaatimattomampi rooli jää W-E-suuntaista muutosta edustaville trendikomponenteille. Regressiomallin asteluvun korottaminen siis parantaa pinnan mukautumista havaintoarvoihin, mutta samalla mallin kompleksisuus ja siihen liittyvät tulkintaongelmat lisääntyvät. Niitä voidaan jonkin verran väistää hahmottamalla regressiopinnan kulkua jälleen yhden N-S-suuntaisen ja yhden W-E-suuntaisen profiilin avulla. Kuvien 4.7 a ja b esimerkkiprofiilien mallista III saadut kolmannen asteen yhtälöt ovat:

$$z = 12915,7 - 129,99 x + 0,91 x^2 - 0,002 x^3, \text{ kun } y = 105 \quad (\text{kuva 4.7 a})$$

$$z = 10352,7 - 68,09 y + 0,51 y^2 - 0,001 y^3, \text{ kun } x = 94 \quad (\text{kuva 4.7 b})$$

Regressiopinta edustaa hyvin Etelä-Suomen maalaiskuntien korkeita tulotasoarvoja. Näistä arvoista lähtien tulotasojen trendi on laskeva Väli-Suomeen asti, joita profiililinjalta edustavat Mikkelin ja Kuopion läänin havainnot. Profiilin laskennallinen minimi sijoittuu Kuopion läänin kohdalle ($x = 104$). Oulun läänissä trendipinta alkaa jälleen nousta ja saavuttaa laskennallisen maksimin Lapin läänin keskiosissa ($x = 225$) laskeakseen jälleen pohjoisinta Lappia kohti edettäessä. Länsi-itäsuunnassa trendipinta on aluksi laskeva, mikä on esim. Vaasan läänissä myös empiirinen todellisuus. Maan keskiosissa, Keski-Suomen ja Kuopion läänin alueella, laskusuunta hidastuu ja nopeutuu jälleen itään mentäessä. Profiililla ei ole maksimi- eikä minimiarvoa. Sen sijaan käännepiste on ja se sattuu Kuopion läänin kohdalle ($y = 121$, kun $z = 0$).

Kuviin 4.7 a ja b on piirretty näkyviin myös läänikohtaisista aineistoista lasketujen lineaarimallien profiilit niiden läänien osalta, joiden kautta profiililinjat kulkevat. Kuvista nähdään, että erityisesti etelä-pohjoissuunnassa kolmannen asteen kuvaaja on erittäin hyvä approksimaatio epäjatkuville läänikohtaisille kuvaajille. Länsi-itäsuunnassa kolmannen asteen kuvaaja edustaa hieman tasaisempaa tulotason vaihtelua kuin läänien kuvaajat, mutta suunta on yleensä sama. Eniten "hämminkiä" on Pohjois-Karjalan läänin kohdalla. Valtakunnallisen kuutiomallin mukaan tulotaso laskisi sen alueella itää kohti mentäessä, läänin aineistosta saadun lineaarimallin mukaan taas nousisi. Ristiriitaa lieventää se testituloksena, että Pohjois-Karjalan lineaarimallin x - ja y -termien kertoimet eivät ole tilastollisesti merkitseviä.

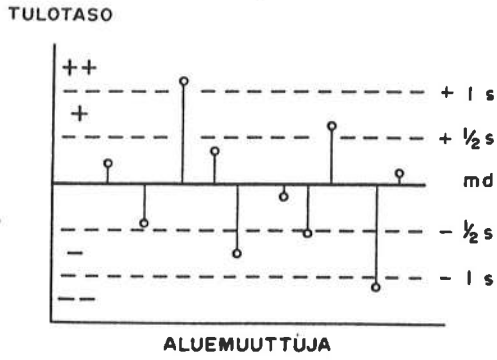
Residuaalikartalta 4 (kuva 4.6) nähdään maalaiskuntien tulotasojen vähintään 1/2 s:n (= 835) suuruiset poikkeamat kolmannen asteen regressiopinnasta. Kartta toistaa havainnollisesti sen tuloksen, että kaupunkien ympäristökunnissa tulotaso nousee useimmin alueellisen trendipinnan yläpuolelle. Läänien rajavyöhykkeille tai muuten syrjäisille alueille ovat puolestaan tyypillisiä trendipinnan alapuolelle jääneet tulotasot, esimerkkeinä Suomenselän alue sekä vyöhyke, johon kuuluu Turun ja Porin läänin itärajan kuntia sekä Uudenmaan läänin luoteisreunan ja Hämeen läänin länsirajan kuntia.

Tutkimusesimerkki:

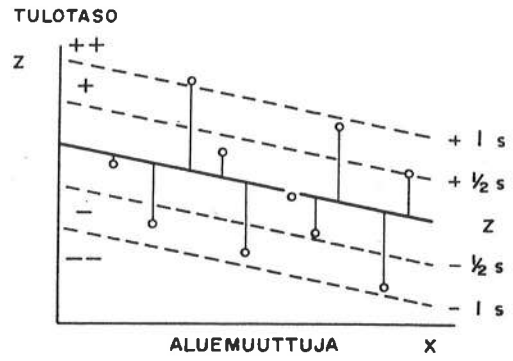
Mikkonen, Kauko (1978). Tulojen alueellinen vaihtelu Suomessa. *Vaasan kaupunkorakentamiskoulun julkaisuja. Tutkimuksia* No 53 (s. 30–40).

Lähde:

Wilson, A.G. & R.J. Bennett (1985). *Mathematical methods in human geography and planning*. New York: John Wiley & Sons.



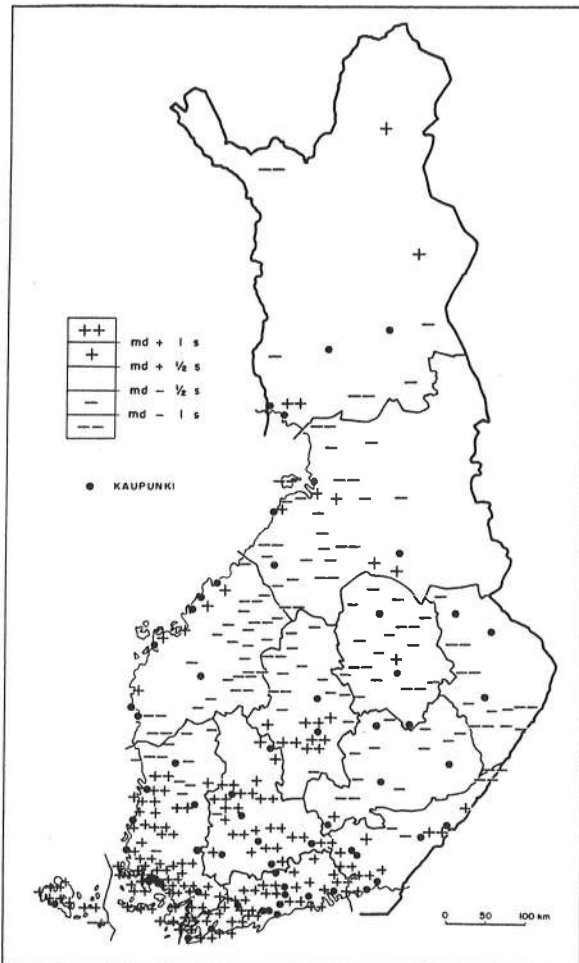
a) Havaintoarvojen poikkeaminen mediaanitulotasosta.



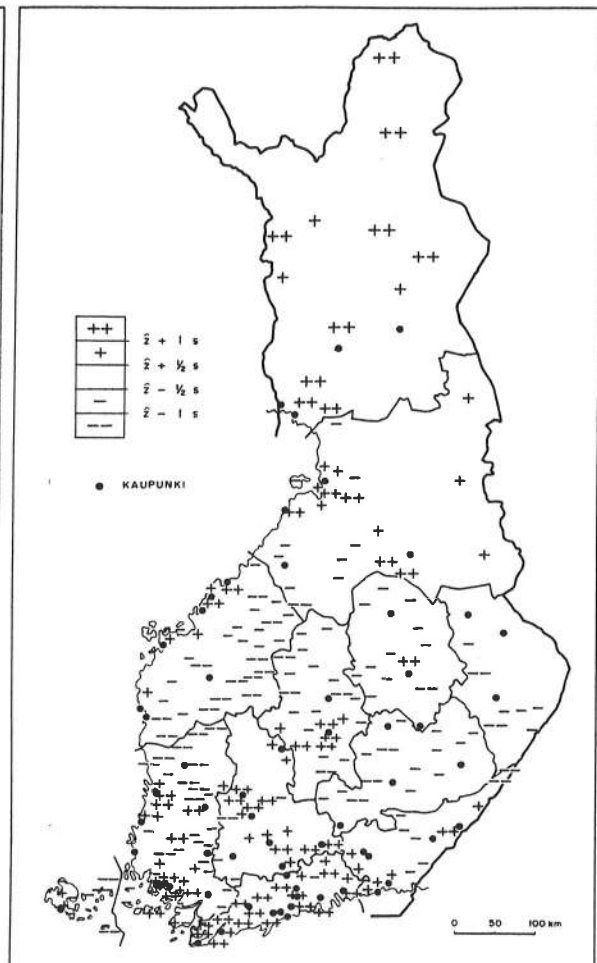
b) Havaintoarvojen poikkeaminen lineaaripinnasta $z = a_0 + a_1x + a_2y$.

Kuva 4.2. Residuaalianalyysin havainnollistus.

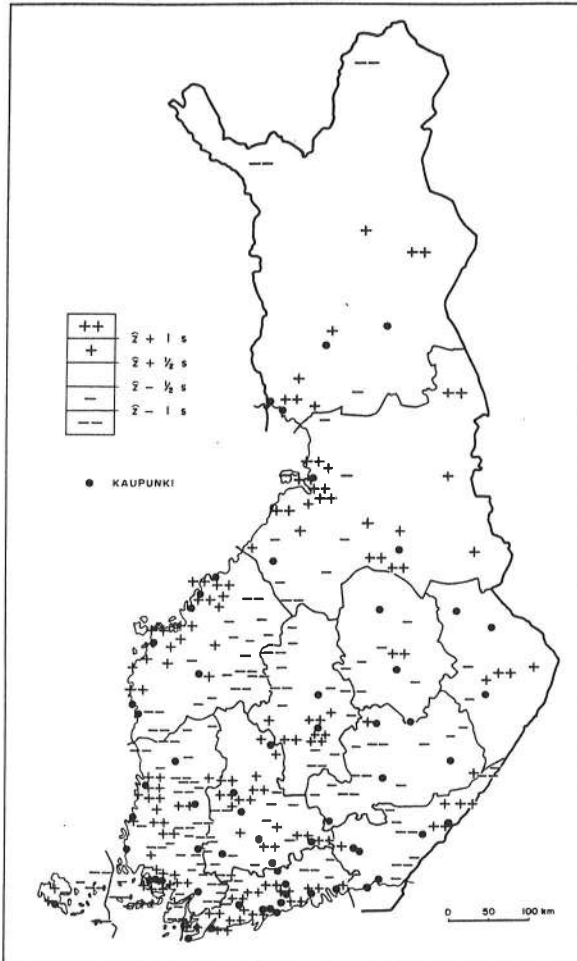
Profiili mallista, kun y-koordinaatti on vakioitu.



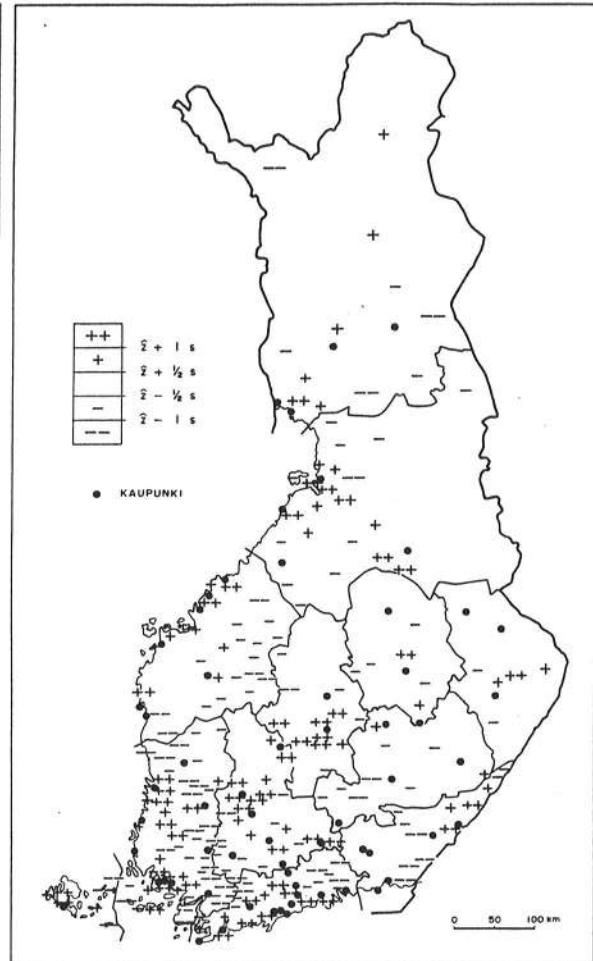
Kuva 4.3. Residuaalikartta 1: Maalaiskuntien tulotasojen poikkeaminen koko maan maalaiskuntien mediaanitulotasosta. $Md = 8522$, $s = 2098$.



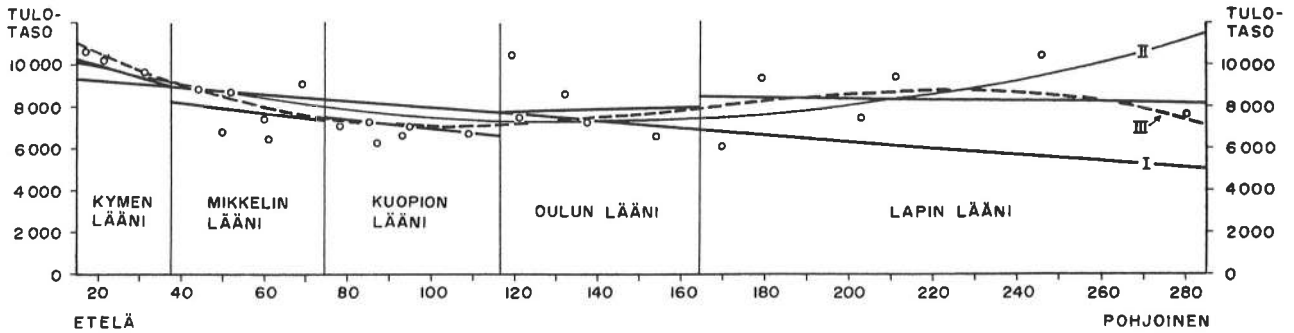
Kuva 4.4. Residuaalikartta 2: Tulotasojen poikkeaminen estimoidusta lineaaripinnasta I. \hat{z} = estimoidusta mallista laskettu tulotaso, $s = 1877$.



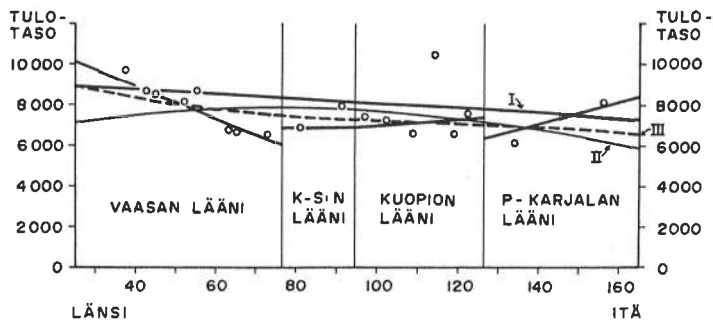
Kuva 4.5. Residuaalikartta 3:
Tulotasojen poikkeaminen
estimoidusta neliöpinnasta II.
 \hat{z} = estimoidusta mallista laskettu
tulotaso, $s = 1728$.



Kuva 4.6. Residuaalikartta 4:
Tulotasojen poikkeaminen
estimoidusta kuutiopinnasta III.
 \hat{z} = estimoidusta mallista laskettu
tulotaso, $s = 1670$.

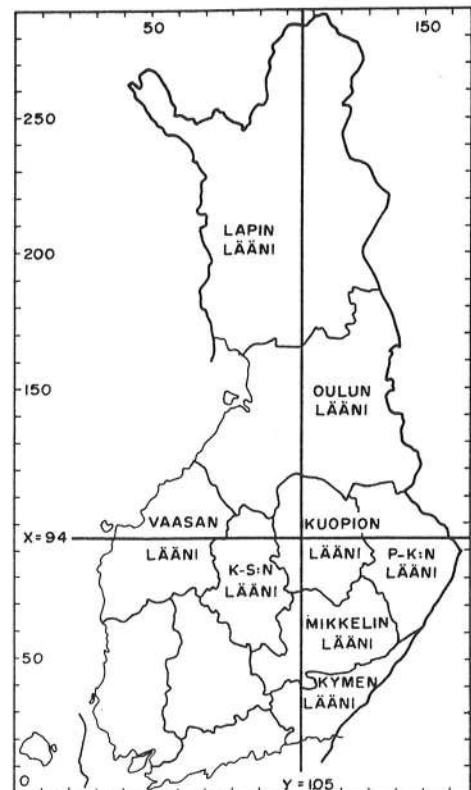


Kuva 4.7 a. Etelä-pohjoissuntaiset profiilit, $y = 105$.



Kuva 4.7 b. Länsi-itäsuuntaiset profiilit, $x = 94$.

Kuva 4.7 c. Trend surface -analyysissä käytetty koordinaatisto ja profiililinjat.



Kuva 4.7. Mallien I, II ja III kaksiulotteiset profiilit sekä läänikohtaisten lineaarimallien profiilit. Pienet ympyrät profiileissa kuvaavat profiililinjoilla tai niiden tuntumassa sijaitsevien kuntien tulotasoja.

Taulukko 4.1. Trend surface -analyysin tulokset koko maan aineistosta 1975 (laskelmissa mukana 392 maalaiskuntaa).

Muuttuja	Regressiokertoimet		
	Lin. malli	Neliömalli	Kuutiomalli
x	-15,42 ^{***}	-63,42 ^{***}	-49,03*
y	-11,11 ^{***}	31,03 ^{**}	67,46*
x ²		0,20 ^{***}	0,77 ^{***}
xy		0,09	-1,57 ^{**}
y ²		-0,26 ^{**}	-0,20
x ³			-0,002 ^{***}
x ² y			0,001
xy ²			0,008*
y ³			-0,001
Vakiotermin	10658,7	10577,9	9658,8
100 R ²	20,4	33,1	38,1
F	49,96 ^{***}	38,19 ^{***}	26,17 ^{***}
Vap.asteiden luku	2; 389	5; 386	9; 382
Hajonta	1877,2	1728,0	1670,2

Merkintöjen selitys:

Riskitaso	Merkintä	Tulkinta
10 %	o	suuntaa antava
5 %	*	melkein merkitsevä
1 %	**	merkitsevä
0,1 %	***	erittäin merkitsevä

5. Potentiaalitekniikka

1 Yleistä

Tietyn alueen kokonaispotentiaalilla, esim. vuorovaikutuksen kokonaismäärällä tarkoitetaan alueen vuorovaikutuksen summaa kaikkien muiden alueiden kanssa mukaan lukien oman alueen vuorovaikutus. Vuorovaikutus voi ilmetä ihmis-, hyödyke-, tavara-, rahavirtoina, palvelusten, ideoiden, informaation tms. liikkumisena alueiden välillä. Potentiaalitermin etuliitteessä kerrotaan yleensä suoraan tai ainakin vihjeellisesti, mistä vuorovaikutuksesta kulloinkin on kyse.

Oman alueen merkitys kokonaispotentiaalin muodostumisessa on luonnollisesti keskeinen. Muiden alueiden vaikutus on suoraan verrannollinen niiden voimakkuuteen tutkittavan asian suhteen ja kääntäen verrannollinen etäisyyteen eli vaikutus on sitä pienempi mitä kauempana ne sijaitsevat alkuperäisestä alueesta. Tämä lausuma voidaan pukea kaavaksi

$$V_i = \frac{A_1}{D_{i1}} + \frac{A_2}{D_{i2}} + \frac{A_3}{D_{i3}} + \dots + \frac{A_N}{D_{iN}} = \sum_{j=1}^N \frac{A_j}{D_{ij}},$$

jossa

V_i = alueiden j kokonaispotentiaali alueella i (alueen i oma potentiaali mukaan lukien),

A_j = alueen j voimakkuus, massatekijä,

D_{ij} = alueiden i ja j välimatka,

N = alueiden lukumäärä.

Yllä esitetty kaava voidaan saattaa sisällöltään monipuolisemmaksi esim. varustamalla etäisyystekijä D_{ij} sen merkityksen voimakkuutta säätelevällä eksponentilla (b) ja lisäämällä kaavan osoittajaan painokerroin (w_j), jolla voidaan korostaa alueiden tai kohderyhmien välisiä eroja. Näiden lisäysten jälkeen potentiaalimallin peruskaava on

$$V_i = \sum_{j=1}^N \frac{w_j A_j}{D_{ij}^b}$$

Sen kehittäminen yhä komplisoidumpaan, useampia parametreja sisältävään muotoon on edelleen mahdollista. Perusteellinen esitys potentiaaliajattelun teoreettisista perusteista ja mallin kehittäminen löytyy esim. Isardin teoksesta (1960: 494–510).

Potentiaalianalyysin eräs sovellusalue on käyttää alueyksittäin laskettuja potentiaaliarvoja jonkin ilmiön vaihtelun selittämiseen. Erinomainen esimerkki tämänkaltaisesta soveltamisesta on ruotsalaisten Tegsjön ja Obergin (1966) yritys selittää kananmunien hintatasoa Ruotsissa tarjonta- ja kysyntäpotentiaalien avulla. Väestöpotentiaaleja voitaisiin erittäin perustellusti käyttää myös keskusvoimakkuuksien selittäjinä. Toisaalta potentiaaliajatteluun nojaavaa logiikkaa on voitu soveltaa keskusvoimakkuuksien laskukaavojen johtamiseen. Niiden avulla on laskettu teoreettisia (optimaalisia) sentraliteetteja, joiden korrelaatio todellisten keskusvoimakkuuksien kanssa lähentelee arvoa 1 (Palomäki & Mikkonen 1971: 25–27; Mikkonen 1975: 38–40 ja 59; Mikkonen 2005).

Yleisimmin potentiaalilaskelmia ja niiden kartografisia tulostuksia lienee kuitenkin käytetty viitekehukseksi jonkin ilmiön tai toiminnon esiintymiselle (esim. markkinapotentiaalikäyrät/*manufacturing belt* USA:ssa), yritysten markkinoinnin alueellisen strategian suunnittelua varten (markkinapotentiaalikartat) tai yritystoiminnan sijaintipäätöksenteon tueksi (väestö- ja ostopotentiaalikartat). Tulo- ja väestöpotentiaalikartoilta sijaintipäätöstä harkitseva yritys voi välittömästi nähdä ne mahdolliset alueet, joilla yrityksen toimintaansa varten edellyttämä väestöpotentiaali tai ostovoima tulee katetuksi.

2 Potentiaalianalyysin toteutus

Potentiaalimallissa, erityisesti sen täydellisimmissä versioissa, esiintyy useita muuttujia ja parametreja (A , w , D ja b), joiden mittaaminen ja määrittely tai estimointi on välttämätöntä ennen kuin potentiaalilaskut voidaan suorittaa.

Massatekijä A . Analysoitavasta asiasta riippuu potentiaalimallin muuttujan A valinta. Esim. kysyntäpotentiaaleja laskettaessa muuttujana A käytetään tavallisesti väkilukua, ostopotentiaalilaskuissa A :ksi sopii hyvin tulojen määrä per

alueyksikkö, markkinapotentiaaleja kartoitettaessa A:na voidaan käyttää esim. vähittäiskaupan tai tukkukaupan myynnin volyyymiä.

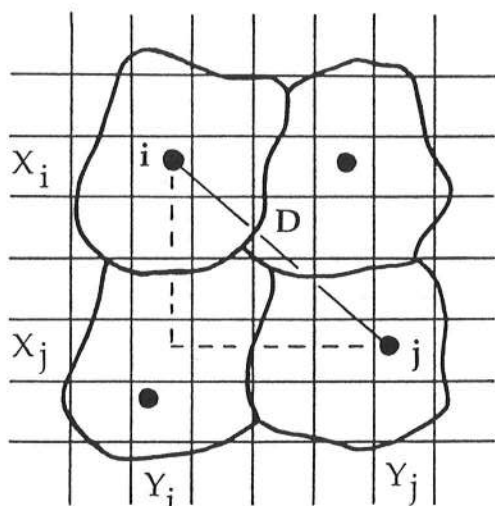
Esimerkkinä massatekijän painottamisesta esitetään usein taloudellisen aspektin kytkeminen väestöpotentiaaliin. Tämä tapahtuu siten, että kunkin alueen väkilukua painotetaan väestön tulotasoa osoittavalla mittaluvulla. Painokerroin w_A saadaan jakolaskulla Tulot/Asukasluvu. Potentiaalimallin osoittajaksi tulee siten

$$w_A = \frac{\text{Tulot}}{\text{Asukasluku}} \times \text{Asukasluku} = \text{Tulot}$$

Yllä oleva tarkoittaa sitä, että painottamalla väkilukua tulotasolla, saadaan aikaan tulopotentiaaliesitys, johon voidaan tietenkin edetä ilman kertolaskuvaihetta. Painokertoimen valinnassa on siis käytettävä harkintaa.

Etäisyystekijä D_{ij} . Etäisyyttä voidaan mitata monilla eri tavoilla. Fyysinen etäisyys on aina varteenotettava vaihtoehto. Mittaaminen voi tapahtua tällöin linnuntietä tai yleisimmin käytettyä reittiä (tavallisesti maantieyhteys) pitkin. Väestön liikkumista vastustava etäisyyskitka voidaan määrittää myös matkakulujen avulla. Ostosmatkojen yhteydessä matkaan käytetty aika on puolestaan havaittu erittäin edustavaksi etäisyysvastuksen mittariksi.

Etäisyyden mittaustapaa valittaessa on huomioon otettava tutkimustehtävän laajuus. Potentiaalialyysissä on mitattava etäisyydet jokaisesta pisteestä jokaiseen pisteeseen. Mitattavien etäisyyksien kokonaislukumäärä saadaan kaavasta $N(N-1)/2$, jossa N on havaintoyksiköiden lukumäärä. Suomessa oli 477 kuntaa vuoden 1975 alussa, joten etäisyysmatriisin muodostamiseksi potentiaalialyysiä varten tarvittiin 113 526 etäisyysmittausta. Tällaisen ja pienemmänkin työmäärän edessä on turvauduttava tietokoneen apuun ja linnuntie-etäisyyksiin. Tehtävä voidaan suorittaa helposti esim. havaintoyksiköiden (kuntien) koordinaatteja ja Pythagoraan lausetta hyväksi käyttäen (ks. kuva 5.1). Vaihtoehtoinen tapa on käyttää tehtävään pallotrigonometrian cosinilausesta: $\cos a = \cos b \times \cos c + \sin b \times \sin c \times \cos \alpha$, jossa a on mitattava välimatka, b ja c leveysasteet navalta mitattuna ja α on pituusasteiden erotus.



$$D_{ij} = (X_i - X_j)^2 + (Y_i - Y_j)^2$$

$$D_{ij} = \sqrt{(X_i - X_j)^2 + (Y_i - Y_j)^2}$$

Kuva 5.1. Kahden keskuksen i ja j välisen etäisyyden laskeminen.

Oman alueen huomioon ottaminen. Potentiaalianalyysissä joudutaan aina erikseen ratkaisemaan, miten oma alue otetaan laskelmissa huomioon, jaetaanko sisäisen vuorovaikutuksen määrää osoittava mittaluku jollakin etäisyydellä ja jos jaetaan niin millä. Käytäntö on ollut kirjavaa. Jakajana voi olla mm. alueen pinta-alaa vastaavan ympyrän säde, alueen suurimman halkaisijan puolikas tai neljännes, alueen keskietäisyys tai väestön sijainnilla painotettu etäisyys, puolet etäisyydestä lähimpään mittauskeskipisteeseen tai arvo 1. Tegsjö ja Oberg (1966) käyttivät jakajana sellaisen ympyrän sädettä ($\times 1,35$), joka maakunnan väestön mediaanipisteeseen sijoitettuna peittää puolet maakunnan väestöstä. Linnuntie-etäisyyden kertomisella 1,35:llä tutkijat pyrkivät lähestymään maakuntien välisiä todellisia maantie- tai rautatie-etäisyyksiä. On syytä todeta, että oman alueen etäisyystekijän valinnalla on lopputuloksen määräytymisessä varsin huomattava rooli.

Etäisyysseksponentti b. Etäisyysseksponentin valintaan liittyy läheisesti etäisyysseksponentin määrittämisongelma. Mitä suurempi etäisyysseksponentin arvo on, sitä voimakkaammin etäisyysseksponentin koetaan ehkäisevän vuorovaikutusta. Jos $b = 0$, etäisyydellä ei ole mitään merkitystä; jos $b = 1$, etäisyysseksponentin vaikutus lisääntyy suorassa suhteessa etäisyyden kasvuun; jos $b = 2$, etäisyysseksponentin vaikutus lisääntyy nopeammin kuin pelkkää fyysistä etäisyyttä käytettäessä jne. Mikäli etäisyyttä mitataan kuljetuskustannusten avulla, etäisyysseksponentin määrittäminen käy tarpeettomaksi, sillä kustannusten epälineaarinen kasvu sisältyy automaattisesti kuljetuskustannuskäyrään. Useissa empiirisissä tutkimuksissa on

voitu todeta, että etäisyyskseenponentti ei ole välttämättä kokonaisluku eikä vakio vaan liikkuu tietyllä vaihteluvälillä. Käytännössä potentiaalilaskelmia kuitenkin usein yksinkertaistetaan valitsemalla etäisyyskseenponentiksi jokin tasaluku, esim. arvo 1 tai 2.

Potentiaalipinnan graafinen esittäminen. Potentiaalialyysin tuloksia havainnollistetaan yleensä kartografisesti potentiaalipinnan avulla. Se saadaan esiin samanarvokäyrätekniikan avulla eli kartalla yhdistetään toisiinsa ne pisteet, joilla on sama potentiaaliarvo. Potentiaalikartan samanarvokäyrät voidaan piirtää absoluuttisten potentiaaliervojen avulla tai arvot voidaan ensin muuntaa laskemalla, kuinka monta prosenttia kukin potentiaaliervo on korkeimmasta arvosta. Korkein potentiaaliervo saa tällöin luvun 100 ja muut sen alle. Samanarvokäyrät piirretään tavallisesti 10 prosenttiyksikön välein. Tällainen 100:aan suhteutettu esitystapa sopii erityisen hyvin eri asioista laadittujen potentiaalipintojen tai samasta ilmiöistä eri ajankohdilta laadittujen potentiaalipintojen vertailuun.

Tutkimusesimerkki

Tutkimusesimerkissä laaditaan koko maan kattavat tulo- ja väestöpotentiaalikatrat. Potentiaalimallin A:na käytetään toisaalta kunnan veroäyrien kokonaismäärää vuoden 1975 tuloista ja toisaalta kunnan väkilukua 1.1.1975. Potentiaalialyysillä ei ole tässä sellaisia erityistavoitteita, että tulojen tai väkiluvun painottaminen esim. jollakin kuntaa luonnehtivalla muuttujalla olisi välttämättöä, ts. potentiaalimallin painokerroin $w = 1$. Kuntien pääkeskusten väliset etäisyydet mitattiin niiden sijaintikoordinaattien ja Pythagoraan kaavan avulla (kuva 5.1). Oman alueen massatekijän jakajaksi valittiin arvo 1. Myös etäisyyskseenponenttina käytetään arvoa 1. Se ei redukoisi muiden alueiden merkitystä eikä vahvista oman alueen osuutta niin suuresti kuin tekisi eksponentin arvo 2.

Edellä esitettyjen ratkaisujen jälkeen tutkimusesimerkissä käytetty potentiaalimalli on saatu muotoon

$$V_i = A_i + \sum_{j=1}^N \frac{A_j}{D_{ij}}, \quad j \neq i$$

Potentiaaliervojen avulla on konstruoitu vastaavat potentiaalikatrat (kuvat 5.2 ja 5.3). Samanarvokäyrät on piirretty karttoihin absoluuttisten potentiaaliervojen perusteella redukoimatta niitä prosenteiksi korkeimmasta arvosta.

Potentiaalikartoissa on huomattava määrä yhdenmukaisuutta. Molemmat potentiaalipinnat laskevat vähitellen Pohjois- ja Itä-Suomea kohti. Tästä yleistasosta nousevat voimakkaasti esiin suuret työpaikkakeskukset korkein tulo- ja väestöpotentiaaliarvoin.

Teollisuus-Suomi rajautuu tulopotentiaalikartalla suurin piirtein 500 milj. äyriä osoittavan samanarvokäyrän eteläpuolelle. Väestöpotentiaalikartalla vastaava rajaus tapahtuu 40 000 asukkaan käyrän kohdalla. Nämä käyrät sattuvat pääosin verraten hyvin yksin Teollisuus-Suomen ja vuonna 1975 voimassa olleen kehitysalueiden II-vyöhykkeen välisen rajan kanssa.

Teollisuus-Suomen ydinalueet näkyvät potentiaalikartoilla selvästi. Näitä ovat Helsingin ja Tampereen seutu sekä radanvarsiakseli näiden napojen välillä, haarautuma Lahteen, Kokemäenjokilaakso, jonka päässä Porin seutua edustava potentiaalihuippu sekä Turun ja Kotkan alueet. Muualla maassa korkeimpina huippuina potentiaalipinnasta työntyvät esiin Lappeenrannan, Mikkelin, Kuopion, Joensuun, Jyväskylän ja Vaasan seudut.

Pohjois-Suomessa Oulun potentiaaliarvot osoittautuvat ylivoimaisiksi muihin huippuihinkin varrattuna.

Tutkimusesimerkki:

Mikkonen, Kauko (1978). Tulojen alueellinen vaihtelu Suomessa. *Vaasan kauppakorkeakoulun julkaisuja. Tutkimuksia* No 53 (s. 71-78).

Muita lähteitä:

Isard, Walter (1960). *Methods of Regional Analysis: An Introduction to Regional Science*. Cambridge, Mass., and London: The M.I.T. Press.

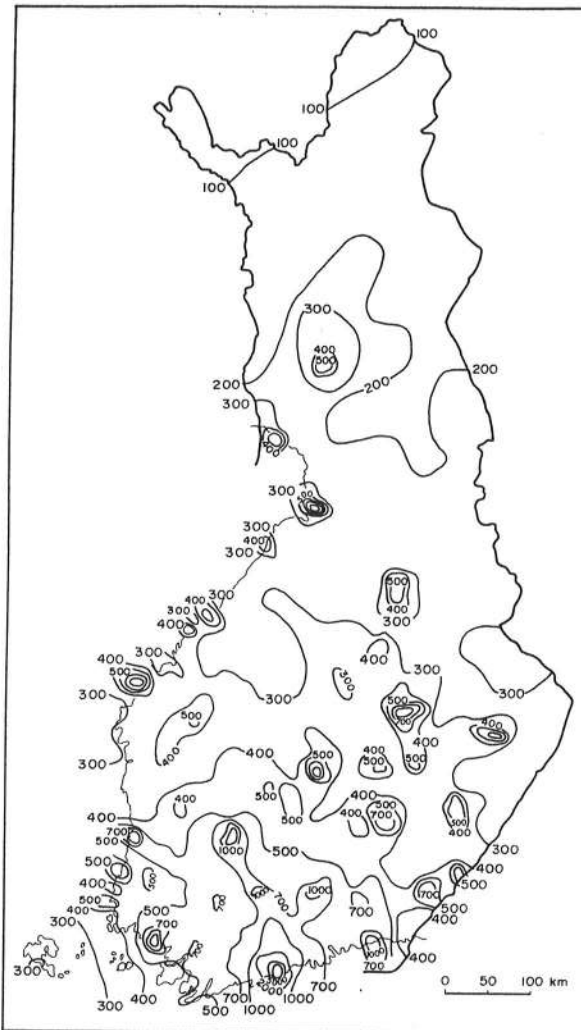
Mikkonen, Kauko (1975). Causal analysis of the system of central places and prediction of functional regional structure in the administrative province of Vaasa, Finland. *Fennia* 138.

Mikkonen, Kauko (2005). Toiminnallisen aluerakenteen ennustaminen - menneestä tulevaisuuteen. *Futura* 3-4/05, 91-103.

Palomäki, Mauri & Kauko Mikkonen (1971). Optimaalisen keskusverkon simulointi Suomeen. *Valtakunnansuunnittelutoimiston julkaisusarja A:25*.

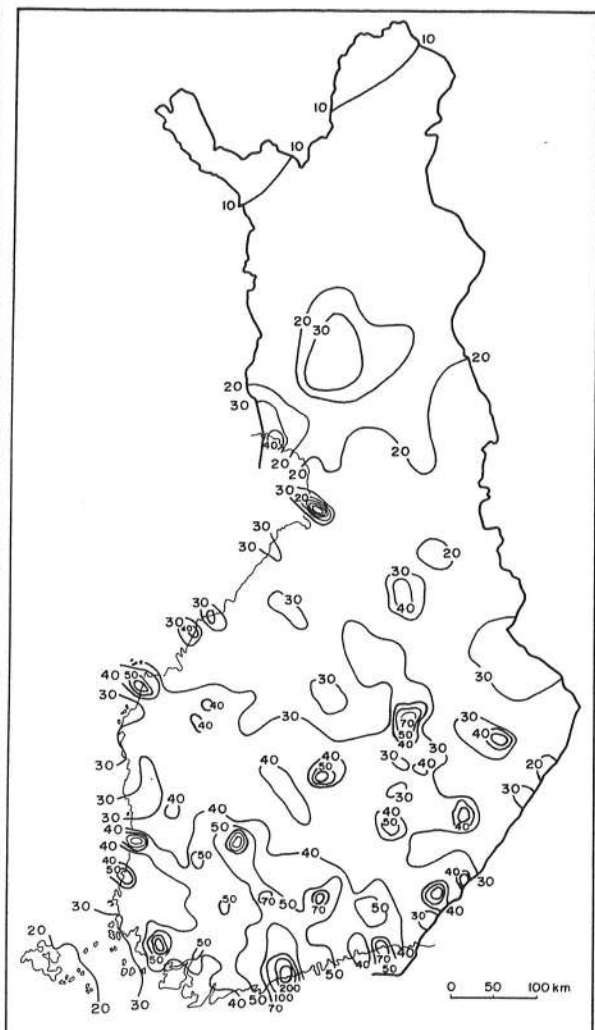
Tegsjö, Björn & Sture Oberg (1966). Concept of potential applied to price formation. *Geografiska Annaler*, Vol. 48, 51-58.

Yeates, Maurice (1968). *An Introduction to Quantitative Analysis in Economic Geography*. New York etc.: Mc Graw-Hill Book Company.



Kuva 5.2. Tulopotentiaalikartta 1975. Numerot samanarvokäyrillä osoittavat miljoonia äyriä.

(Mikkonen 1978: kuva 16.)



Kuva 5.3. Väestöpötkartta 1975. Numerot samanarvokäyrillä osoittavat 1000 henkilöitä.

(Mikkonen 1978: kuva 17.)

6. Minimimatkasumma

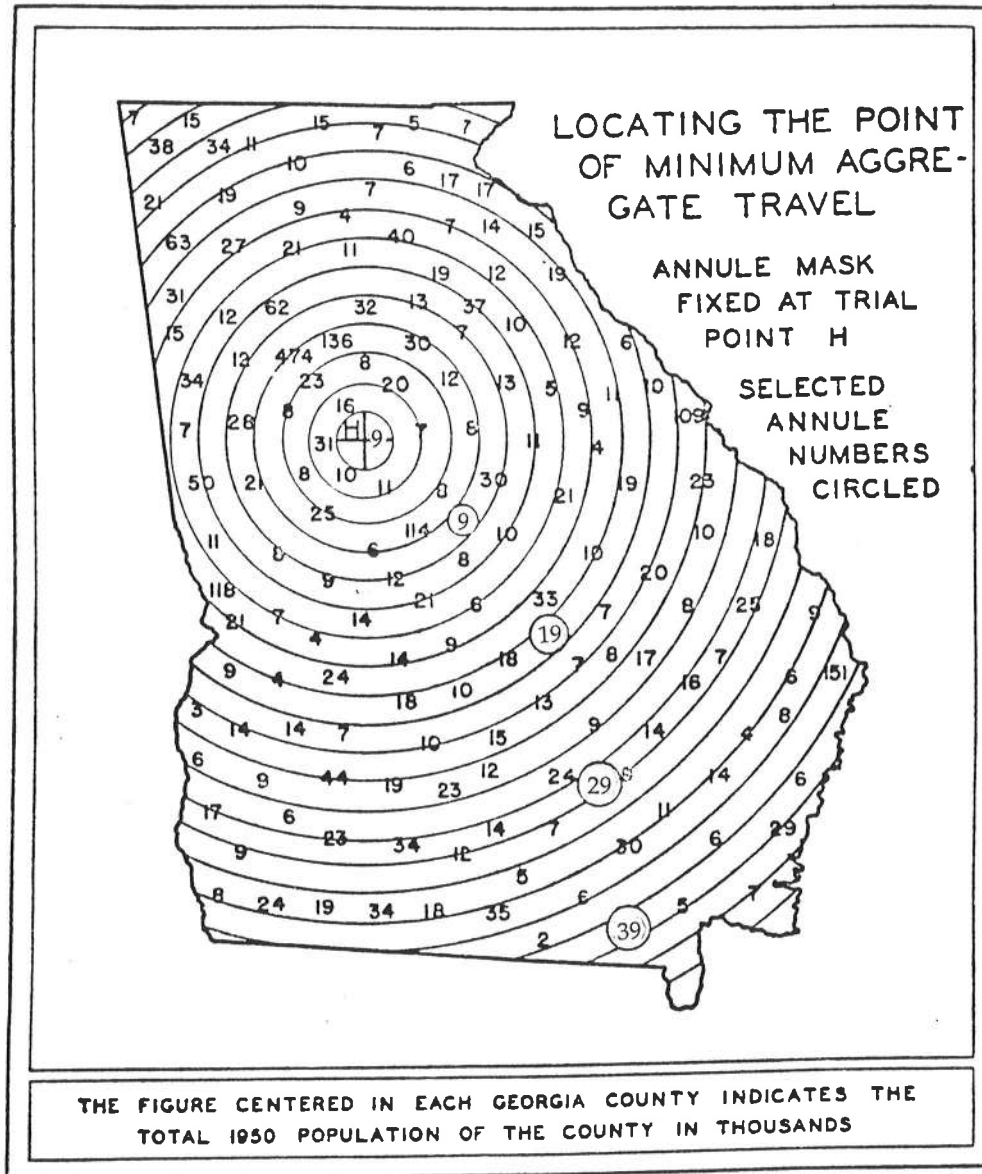
Minimimatkasumma liittyy yleensä alueellisen tehokkuuden (*spatial efficiency*) mittaamiseen tai optimointiin. Alueellinen tehokkuus on suurimmillaan, kun kuljetus- tai liikkumiskustannukset tai kokonaismatkatuotos on minimissään. Tyypillisiä kysymyksiä tässä yhteydessä ovat: onko palvelusten sijoittuminen tai aluejako (esim. koulupiirijako) väestön liikkumiskustannusten tai -tarpeen kannalta edullisin mahdollinen; millainen tulisi optimaalisen varastoverkon, palvelupiste- tai keskusverkon ja vastaavan alueverkon olla?

Tärkeitä kysymyksiä alue- ja yhdyskuntasuunnittelussa sekä logistiikassa!

6.1. Kartografinen ratkaisu

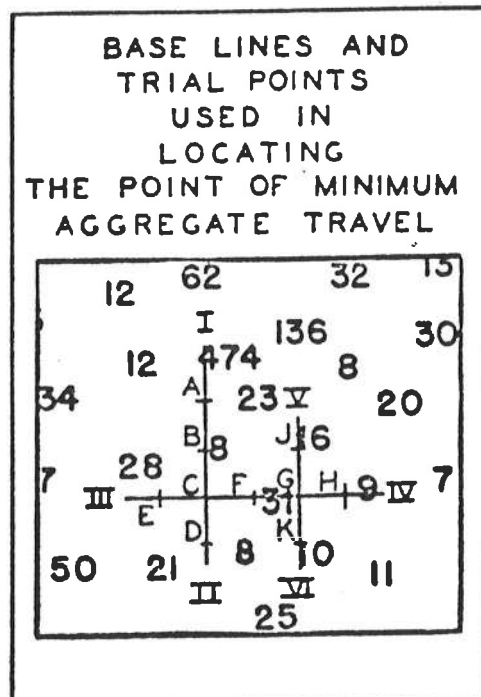
- 1) Tehdään läpinäkyvälle paperille sapluuna, jossa on samankeskisiä ympyröitä. Ympyröiden säde kasvaa aina yhtä paljon (kuva 6.1). Mitä pienempi säteen lisäys on, sitä tarkempi on tulos. Varustetaan ympyröiden kehät parittomilla numeroilla keskipisteestä lähtien.
- 2) Valitaan tutkimusalueelta silmämääräisesti arvioitu minimimatkapiste ja piirretään sen kautta perusviiva I-II (kuva 6.2). Tälle viivalle merkitään koepisteitä A, B, C, D, joiden välimatka on sama kuin säteen lisäys ympyröitä piirrettäessä.
- 3) Lasketaan koepisteiden matkasummat seuraavasti:
 - a) sijoitetaan kehäpiirroksen keskipiste koepisteeseen,
 - b) lasketaan jokaisen kehän havaintoarvot yhteen ja kerrotaan summa kehän järjestysnumerolla,
 - b) lasketaan näin saadut tulot yhteen ja tuloksena on tämän koepisteen matkasumma.
- 4) Siihen perusviivan pisteeseen, jossa laskettu matkasumma oli pienin, piirretään toinen perusviiva III-IV (kuva 6.2) kohtisuoraan ensimmäistä perusviivaa vastaan.
- 5) Lasketaan jälleen siihen merkittyihin koepisteisiin E, F, G, H matkasummat.
- 6) Koepisteeseen, josta saatiin pienin summa, piirretään kolmas perusviiva V-IV.

Iterointia jatketaan, kunnes on löydetty minimimatkasumman piste.



Kuva 6.1.
Minimimatkasummapisteen
etsimisessä käytetään ympyrä-
sapluunaa (esim. Georgiasta
USA:sta).

Kuva 6.2.
Perusviivat ja koepisteet
minimimatkasummapisteen
etsimisessä.



6.2. P-mediaanianalyysi

P-mediaanianalyysiksi kutsuttu menetelmä on minimimatkasummamenetelmän sovellus. P-mediaanianalyysin tehtäväkuvaus kuuluu verkkoteoreettisiin termein seuraavasti:

On valittava annetusta solmujoukosta p solmua siten, että muista solmuista lähimpään p -solmuun mitattujen painotettujen etäisyyksien summa tulee mahdollisimman pieneksi.

**Tulokseksi saatua solmujoukkoa kutsutaan verkon p -mediaaniksi.
Jos $p = 1$, tulokseksi saatua solmua kutsutaan verkon mediaaniksi.**

P-mediaanianalyysi tuottaa sivutuotteena myös aluejakoja!

Solmujen väliset etäisyydet. P-mediaanianalyysin lähtökohtana on etäisyysmatriisi, joka sisältää etäisyydet jokaisesta solmusta jokaiseen solmuun. Etäisyysmatriisia muodostettaessa pätee sama kuin potentiaalianalyysissä. Tehtävään loogisesti sopivimpia ovat matkustamiseen kuluva aika, matkakustannukset tai yleisimmin käytetty (tavallisesti lyhin) maantie-etäisyys. Niiden käytön esteenä on suuri työmäärä, jos etäisyysmittauksia tulee runsaasti. Tästä syystä käytetään yleensä linnuntie-etäisyyttä, koska se voidaan laskea tietokoneella Pythagoraan lauseen (tai cosinilauseen) avulla, kun solmujen koordinaatit (tai leveys- ja pituusasteet) on määritetty (ks. kuva 5.1).

Etäisyysmekijän eksponentti. Etäisyysmekijän avulla voidaan säädellä fyysisen etäisyysmekijän merkitystä lopputuloksessa:

korkea eksponentin arvo - etäisyysmekijä suuri
alhainen eksponentin arvo - etäisyysmekijä pieni

Yleisesti käytettyjä eksponentin arvoja ovat 0,5, 1 ja 2. Ykköstä pienemmällä eksponentin arvolla fyysinen etäisyys saadaan paremmin vastaamaan matkakustannusten hidastuvaa kasvua etäisyyden funktiona.

Etäisyyden painokerroin. Tehtävästä riippuen painokertoimeksi valitaan esim. väkiluku, eläkeläisten lukumäärä, ammattikoulutusiässä (15-19 v.) olevien määrä, tulopotentiaali, teollisuustyöpaikkojen määrä tms.

Laskentamenetelmä. P-mediaanianalyysin ratkaisemiseksi on kehitetty useita menetelmiä. Lineaariseen ohjelmointiin perustuva menetelmä kuuluu pätevimpään, mutta se vaatii runsaasti tietokoneaikaa. Käytännössä ehkä yleisimmin käytetty on nk. **korvausmenetelmä**, jossa ei tutkita kaikkia mahdollisia p-solmun valintamahdollisuuksia. Se ei anna kaikissa tapauksissa optimiratkaisua, mutta kohtuullisella työmäärällä (tietokoneajalla) joka tapauksessa hyvän lopputuloksen.

Minimimatkasummapisteen, -pisteiden ominaisuuksia

Minimimatkasummaratkaisussa ovat aineistosta lasketut säteen suuntaiset painotetut etäisyydet summaltaan pienimmät. Minimimatkasummapiste tai -pisteet edustavat näin teoreettisesti edullisinta ratkaisua olettaen, että yhteydet kuhunkin vaihtoehtoiseen pisteeseen ovat suoraviivaiset ja kohdejoukon kaikilla jäsenillä on yhtäläinen tarve ja mahdollisuudet liikkua. Käytännössä tällaista täydellisesti yhdistävää liikenneverkkoa ei ole olemassa. Tulosten tulkinnassa onkin tarpeen ottaa huomioon liikennevirtojen tosiasiallinen suuntautuminen, liikenne-esteet (esim. vesistöt), kielirajan mahdollinen vaikutus jne.

Minimimatkasummamenetelmän (manuaalisen tai p-mediaanianalyysin) avulla voidaan vertailla esim. kahden tai useamman pisteen, kirkonkylien, tienristeys-solmujen tms. edullisuutta matkatuotosten suhteen.

Esimerkkejä:

Hautamäki, Lauri (1972). Suomen keskusverkon optimointi. *Fennia* 112.

Mikkonen, K. & Riitta Koski (1973). Lapuan kaupunkihankkeen tarkoituksenmukaisuus. *Vaasan kaupparakennuslaitoksen julkaisuja. Tutkimuksia* No 20.

Mikkonen (1976). Vaasan läänin yhdyskuntarakenne. Kulttuuripalvelukset. *Vaasan kaupparakennuslaitoksen julkaisuja. Tutkimuksia* No 37.

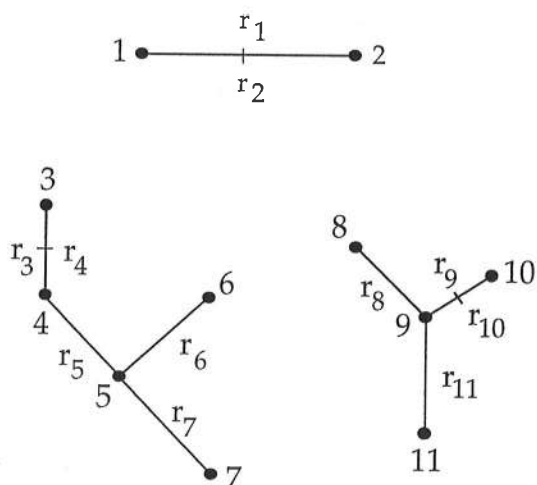
Nuutinen, Raakel (1977). Etelä-Pohjanmaan tuotesuunnitteluopiston sijoittaminen. Talousmaantieteen laudaturtutkielma. Vaasan kaupparakennuslaitoksen julkaisuja.

7. Lähimmän naapurin menetelmä

Lähimmän naapurin menetelmällä (*nearest neighbor analysis*) tutkitaan

- 1) jonkin ilmiön levinneisyyden keskittyneisyyden tai hajanaisuuden astetta tietyssä poikkileikkaustilanteessa tai eri poikkileikkaustilanteissa (esim. uudisteen leviämisen eri vaiheissa);
- 2) kahden ilmiön levinneisyyskuvien eroavuutta.

Lähimmän naapurin menetelmässä tarkastelun kohteena olevaa levinneisyyskuvaa, pisteverkkoa, verrataan satunnaisesti jakautuneeseen levinneisyyskuvaan. Menetelmässä nojaututaan mittauksiin siitä, kuinka tutkittavan ilmiön yksilöt sijaitsevat toisiinsa nähden jollakin alueella (kuva 7.1).



Pisteitä (esim. talouksia) 11.

Pisteen 1 lähin naapuri on piste 2 ja näiden välinen etäisyys on r_1 .

Pisteellä 2 on lähimpänä naapurinaan piste 1 ja etäisyys sinne on r_2 . Tässä on kysymys ns. *refleksiivisestä pisteparista*, ts. pisteet ovat toistensa lähimpiä naapureita. Pisteen 6 lähin naapuri on piste 5 ja etäisyys sinne r_6 jne.

Etäisyyksien lukumäärä on aina yhtä suuri kuin pisteiden lukumäärä.

Kuva 7.1. Esimerkki lähimmän naapurin periaatteesta.

Lähin-naapuritekniikan testisuure

$$R = \bar{r}_A / \bar{r}_E, \quad \text{jossa}$$

$\bar{r}_A = (\sum r) / N$ lähimpiin naapureihin mitattujen todellisten etäisyyksien keskiarvo,

$\bar{r}_E = 1 / (2 \cdot \sqrt{p})$ odostusarvo eli eniten todennäköinen etäisyyskeskiarvo lähimpään naapuriin,

$p = N / A$, jossa N = pisteiden lukumäärä, A = alueen maapinta-ala.

\bar{r}_E :n johtaminen:

Jos pisteet (esim. ruokakunnat) ovat täydellisesti agglomeroituneet, silloin $\bar{r}_A = 0$.

Jos pisteet ovat tasaisesti jakaantuneet pinnalle A, silloin \bar{r}_A saadaan yhtälöstä

$$\bar{r}_A = \frac{1.0746}{\sqrt{p}},$$

jossa p on pisteiden lukumäärä pinta-alayksikköä kohti eli N/A.

Jos N pistettä jaetaan sattumanvaraisesti pinnalle A rajattoman monta kertaa, todennäköisyys sille, että \bar{r} joutuu ääriarvojen $\bar{r} = 0$ ja $\bar{r} = 1.0746/\sqrt{p}$ puoleenväliin, on erittäin suuri.

Näin päästään odotusarvoon eli eniten todennäköiseen etäisyyskeskiarvoon

$$\bar{r}_E = \frac{0 + \frac{1.0746}{\sqrt{p}}}{2} \approx \frac{1}{2\sqrt{p}}$$

todennäköisyyskentässä, jonka oletetaan täyttävän sille asetetut yleiset ehdot.

R:n tulkinta	Skaala:	0 - 2.1491
	R = 0	täydellinen agglomeroituminen
	R = 2,1491	täysin tasainen levinneisyys
	R = 1,0746	sattumanvarainen

Esimerkkejä:

Mikkonen, Kauko (1978). The rotary movement. Diffusion of an innovation. *Fennia* 155, 1-21.

Nurminen, Eero (1981). Kanta-Hämeen ala-asteen kouluverkon alueellinen rakenne ja kehitys-prosessi. *Acta Universitatis Tamperensis Ser. A, Vol.129*.

Löytönen, Markku (1985). Spatial development of the post office network in the province of Mikkeli, Finland 1860-1980. *Fennia* 163, 1-112.

8. Shift and share -menetelmä

Shift and share -tekniikkaa käytettiin ensimmäisen kerran jo 1940-luvun alkupuolella, käyttäjänä Daniel B. Creamer (1943), mutta tekniikan popularisointi tapahtui vasta 1960-luvulla Perloffin ym. toimesta.

Shift and share -tekniikka on eräs yleisimmin käytetyistä aluetalouden analyysivälineistä. Suosion syytä: menetelmä on sopivan yksinkertainen, ja sitä voidaan soveltaa mitä moninaisimpiin taloudellisiin indikaattoreihin, esim. työpaikka-, toimipaikka-, kansantuote-, tulotilastoihin. Työllisyys (työpaikkakehitys) lienee kuitenkin yleisin sovellusalue ehkä siitä syystä, että työllisyystiedot ovat helposti saatavissa ja itse asia on kovin keskeinen yhteiskunta- ja aluepoliittisessa päätöksenteossa ja sen seurannassa. Shift and share -menetelmän ehkä tärkein ominaispiirre on se, että se näyttää tarjoavan systemaattisen keinon, jolla alueen sisäisestä kasvusta erotetaan koko maan kasvuosuus.

Kokonaissiirtymä

Alueittaisiin kokonaissiirtymiin, *total shift*, edetään seuraavien vaiheiden kautta (tarkastelun kohteena työpaikat):

1 Kullekin alueelle lasketaan työpaikkojen lukumäärän odotusarvo ajankohdassa t kertomalla alueen työpaikkojen lkm ajankohdassa 0 valtakunnallisella siirtymäkertoimella = työpaikkojen lkm ajankohdassa t / työpaikkojen lkm ajankohdassa 0.

2 Alueen työpaikkojen lukumäärän kokonaissiirtymä = työpaikkojen todellinen lkm ajankohdassa t - odotusarvo ajankohdassa t .

Mikäli siirtymä on merkittävästi positiivinen, työpaikkojen lukumäärän kehitys on ollut koko maan kehitystä suotuisampi. Miinusmerkkinen siirtymä merkitsee heikompaa kehitystä kuin maassa keskimäärin.

3 Erisuuruisten aineistojen vertailemiseksi kokonaissiirtymä ilmaistaan usein prosentteina ajankohdan 0 lähtöarvosta.

Kokonaissiirtymän laskukaava:

$$(1) \quad S = E_{rt} - (E_{nt} / E_{n0}) \times E_{r0},$$

jossa

$$\begin{aligned} t &= \text{pääteajankohta} & 0 &= \text{lähtöajankohta} \\ E_r &= \text{alueen työpaikkojen määrä} & E_n &= \text{koko maan työpaikkojen määrä} \end{aligned}$$

Differentiaallinen ja rakenteellinen siirtymä

Todellisen ja odotetun kasvun erotus, positiivinen tai negatiivinen kokonaisu-siirtymä, jaetaan kahteen komponenttiin, differentiaaliseen siirtymään ja raken-teelliseen siirtymään.

Differentiaallinen siirtymä (*differential shift*) osoittaa, missä määrin alueen toimialat kasvavat nopeammin tai hitaammin kuin vastaavat toimialat koko maassa. Sen tulkitaan heijastavan suhteellisia sijaintietoja, jotka selittävät toimialan erilaisia kasvunopeuksia eri alueilla. Differentiaallinen siirtymä S_d saadaan soveltamalla kaavaa (1) yksittäisiin toimialoihin tai toimialaryhmiin i ja näin saadut siirtymät lasketaan yhteen eli

$$(2) \quad S_d = \sum_i [E_{ri}^i - (E_{ni}^i / E_{no}^i) E_{ro}^i]$$

Rakenteellinen siirtymä (*structural shift*) heijastaa alueen toimialarakenteen vaikutusta työpaikkojen määrän kehitykseen. Jos alueen toimialarakenne on painottunut heikon kehityksen toimialoille koko maahan verrattuna, raken-teellisestä siirtymästä tulee negatiivinen ja päin vastoin. Kun kokonaisu-siirtymä ja differentiaallinen siirtymä tunnetaan, rakenteellinen siirtymä S_p saadaan yksin-kertaisimmin vähentämällä kokonaisu-siirtymästä differentiaallinen siirtymä eli

$$\begin{aligned} (3) \quad S_p &= S - S_d \\ &= E_{rt} - (E_{nt} / E_{no}) E_{ro} - \sum_i [E_{ri}^i - (E_{ni}^i / E_{no}^i) E_{ro}^i] \\ &= \sum_i [(E_{ni}^i / E_{no}^i) - (E_{nt} / E_{no})] E_{ro}^i \end{aligned}$$

Kritiikkiä

Shift and share -analyysi on paitsi suosittu myös ehkä yliarvostetuin aluetalouden analyysiväline. Se on vain kuvaileva menetelmä. Sen avulla ei voida selittää työllisyyden alueittaisia kasvueroja. Jollakin alueella toimialan kehitys voi olla positiivinen, vaikka alan kehitys koko maassa olisi negatiivinen ja päin vastoin. Menetelmä ei kerro mitään alueen kapasiteetista säilyttää kasvavia toimialoja tai siitä, miten vetää niitä puoleensa.

Positiivinen differentiaalinen siirtymä ei välttämättä johdu suhteellisesta sijaintiedusta. Se voi johtua jopa virheellisestä toimialaluokituksesta, tuotannon heterogeenisuudesta, yritysten tuotannonhaarojen uudelleenjärjestelyistä jne. Toivoa kuitenkin sopii, että suuressa aineistossa em. erikoispiirteet katoavat ja päästään yleisempiin johtopäätöksiin. Useissa tutkimuksissa on osoitettu, että differentiaaliset siirtymäkomponentit ovat ajassa muuttuvia. Toisaalta eräät tutkijat ovat havainneet, että kyseiset komponentit olisivat pysyviä joillain toimialoilla mutta epävakaita joillakin toisilla.

Kaikista varauksista huolimatta uusia shift and share -sovelluksia löytyy jatkuvasti arvostetuissa kansainvälisissä julkaisusarjoissa, mikä puhuu selkeää kieltä menetelmän käyttökelpoisuudesta. Kriittisyyttä toki tarvitaan tässä kuten kaikkia muitakin menetelmiä käytettäessä!

Lähde:

Richardson, Harry W. (1976). *Regional economics. Location theory, urban structure and regional change*. First published 1968, reprinted 1976. London: Weidenfeld and Nicolson (s. 342–346).

Esimerkki:

Mikkonen, Kauko (1996). Suomen teollisuuden toimipaikkarakenteen muutos taantumavuosina 1990–1993. *Vaasan yliopiston julkaisuja. Tutkimuksia* 204.