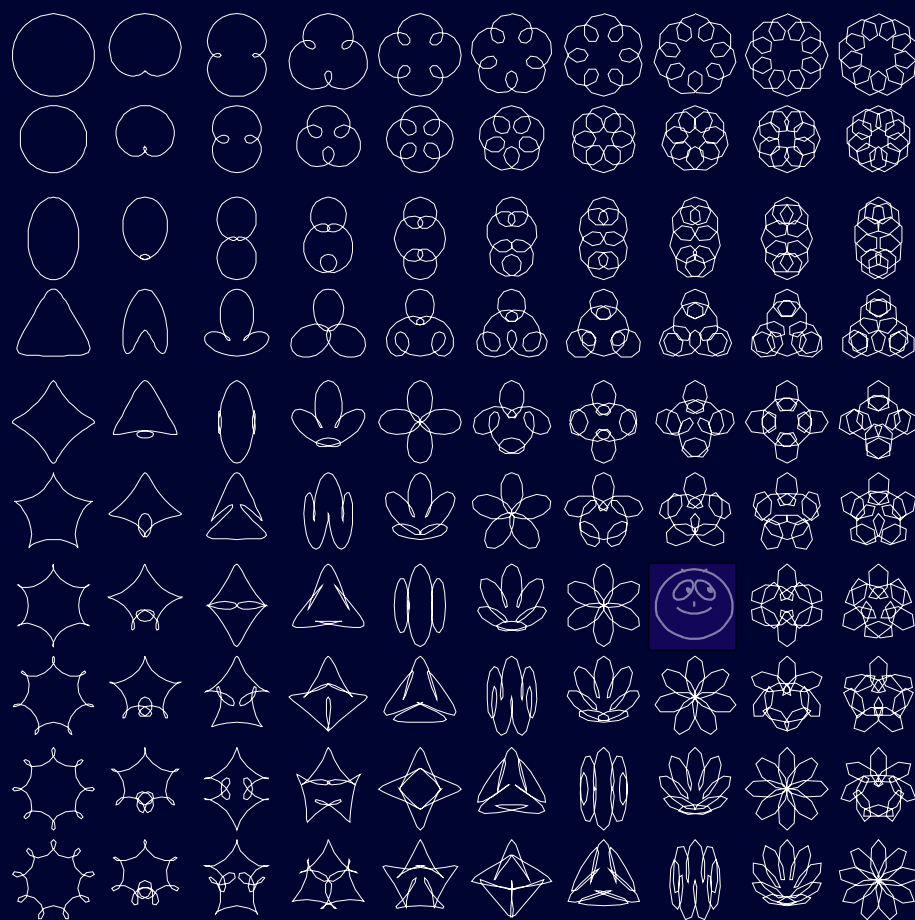


# Survo ja minä



Seppo Mustonen

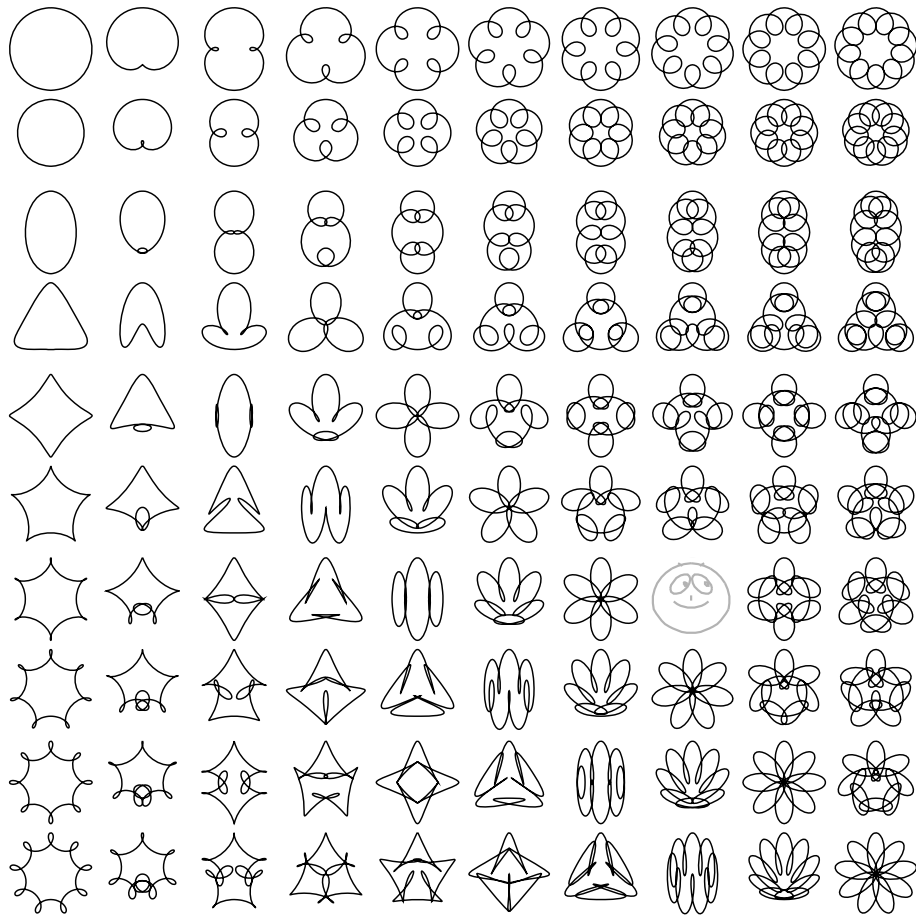


**Survo ja minä**





# Survo ja minä



**Seppo Mustonen**

© Seppo Mustonen 1996

Kustantaja Survo Systems Oy



Kirjassa kuvatut toiminnot ovat voimassa SURVO 84C -ohjelmiston versioissa 5.20- . Ne edellyttävät lisäksi MULTI2-levykkeellä olevia Survon laajennuksia.

Kirja on suunniteltu ja kirjoitettu SURVO 84C -työnä. Paino-originaali on tehty suoraan Survon PRINT-operaatiolla käyttäen PostScript-tulostinta, jonka tarkkuus on 1200 pistettä tuumalla. Teos on painettu Gummerus Kirjapaino Oy:ssä 1996.

ISBN 951-96634-2-8







## Sisällys

|                           |     |
|---------------------------|-----|
| Pientä laskentaa          | 9   |
| Yleisperiaatteita         | 17  |
| Töiden hallinta           | 29  |
| Alusta pitäen             | 47  |
| Nappulat pelikentällä     | 59  |
| 25. heinäkuuta 1996       | 73  |
| Tilastollista käsittelyä  | 89  |
| Keskustelu                | 111 |
| Sattuman pyyntiä          | 117 |
| Monen ulottuvilla         | 139 |
| Kuvien takaa              | 173 |
| Epäpuhdasta matematiikkaa | 193 |
| Makein juttu              | 205 |
| Laiskana ja tyhmänä       | 215 |

Tämän kirjan synnystä on kulunut aikaa jo 13 vuotta.  
Tällä välin on ehtinytkin tapahtua yhtä ja toista.  
Survossa rikottiin ahtaat DOS-ympäristön rajat (v.1998)  
ja siirryttiin Windows-ympäristöön (v.2000).  
Uusin versio SURVO MM on silti täysin yhteensopiva  
aikaisempien Survojen kanssa eli kirjan sisältö vastaa  
käyttökuvausten ja -esimerkkien osalta myös nykytilannetta.

Tässä kirjassa esittämistäni epäilyistä huolimatta  
Survon ja Windowsin yhteistyö on onnistunut myönteisellä  
tavalla, jota en todellakaan osannut etukäteen odottaa.  
Vaikka SURVO MM sisältää lukuisia uusia ominaisuuksia  
ja toimintatapoja, kirja toimii nähdäkseni edelleen  
kelvollisena johdatuksena Survon käyttöön.

Lukijan kannattaa myös tutustua Survon verkkosivuihin  
**[www.survo.fi](http://www.survo.fi)**  
saattaakseen tietonsa ajan tasalle.

Heinäkuussa 2009  
Seppo Mustonen



## Alkusanat

Ihmisen ja tietokoneen välinen kanssakäyminen on kiinnostanut minua lähes 40 vuoden ajan. Se tarjoaa haasteita, joita emme ole ennen kohdanneet. Eipä ole mitään itsestään selvää vastausta siihen, miten koneen kanssa pääsee hyviin väleihin. Olen mielestäni onnellisessa asemassa saatettuani kehitellä ja kokeilla omia ajatuksiani jo pitkään - etenkin kun ottaa huomioon alan lyhyen historian.

Puuhailuni tietojenkäsittelyn parissa ovat keskittyneet Survo-nimisen ohjelmiston kehittämiseen ja alunperin pelkästään tilastollisen tutkimuksen tarpeisiin. Olen kuitenkin aina halunnut kokeilla tietokoneen mahdollisuuksia myös oman alan ulkopuolella. Itse asiassa sellaiset asiat, joissa kone saadaan käyttäytymään jotenkin "ihmisiksi", houkuttelevat minua. Musiikin ystävänä olen varhain (1960-luvun alkupuolella) tehnyt kokeita, joissa tietokone "säveltää" ja myös "esittää" sävellyksensä. Myöhemmin 1970- ja 1980-lukujen taitteessa laadin erään ensimmäisistä nuotinpainatusohjelmista. Tuon kokeen ansiosta sain ratkaisevan herätteen Survon nykyisen käyttötavan muotoutumiselle.

Survo ei ole enää pelkästään "tilastopaketti". Se on yleinen käyttöympäristö, jossa käsitellään sekä tekstiä että numeerista tietoa. Tavoitteenani on ollut luoda vapaa temmellyskenttä erilaisten ajatusten kehittelylle silloin kun ajatukset ilmaistaan tekstinä, kuvina, lukuina, taulukkoina, kaavoina jne. En pakota ihmisiä kuljettamaan töitään nykyohjelmien tyyliin ahtaita putkistoja pitkin eli valmiiksi rakennettujen valikkojen kautta. Suosin luovaa käyttäjää, joka tarvittaessa raivaa omat polkunsa ja peltonsa tutkimattoman tiedon ryteikköön.

Kannan huonoa omaatuntoa siitä, että Survon todellinen olemus on toistaiseksi jäänyt vieraaksi useimmille tietojenkäsittelystä kiinnostuneille. Survo saataan vieläkin erilaisuutensa vuoksi leimata vaikeaksi ja vain asiantuntijoille tarkoitetuksi ohjelmistoksi. Tämä ei kerta kaikkiaan pidä paikkaansa. Survoa käyttävät Suomessa hyvin erilaiset ihmiset ja erilaisiin tarkoituksiin. Menestyksellisen käytön ehtona ei ole koulutustaso vaan halu päästä työssään hyviin tuloksiin eli "intohimo kvaliteettiin".

Tämä kirja kannesta kanteen ja viimeistä piirtoa myöten on tuotettu Survolla painovalmiiseen muotoon. Kirjan tarkoituksena on ennen muuta osoittaa Survon luonteenomaiset piirteet ja sen käytön mahdollisuudet erilaisten esimerkkien kautta. Heti alusta pitäen yritän näyttää, miten Survossa monet yksinker-

taisetkin tehtävät on helpompi hoitaa kuin tavanomaisilla ohjelmilla. Tämä ei ole varsinaisesti mikään oppikirja eikä kaikenkattava läpileikkaus eri toiminnoista, vaan aiheet otin omien mieltymysteni mukaisesti. Uskon kuitenkin sen auttavan alkuun niitä, jotka Survoon mielivät tutustua. Se antaa toivoakseni uutta tietoa ja näkemystä myös vanhoille Survon käyttäjille. Ennen kaikkea haluan osoittaa, miten Survo venyy käyttäjänsä taitojen kehittyessä yhä vaativampiin suorituksiin. Myös toivon merkillepantavan, miten monenlaatuisiin tehtäviin se yltää. Useat survoilijat eivät juuri muita ohjelmia kaipaakaan.

Olen tarkoituksella valinnut kevyehkön tyylilajin. Esitän jopa sangen omalaa-tuisia sovelluksia senkin uhalla, että lukija ajattelee Survolla vain leikittävän. Näin olen halunnut saada eloisuutta ja laveutta Survon periaatteiden tajuami-seksi.

Koska kirjan jotkut jaksot vaativat eriasteista perehtyneisyyttä matematiikkaan, tilastotieteeseen tai tietojenkäsittelyyn, kerron kulloisenkin kohdan vai-keusasteesta reunamerkinnein.

- Puhdas valkoinen ruutu vasemmalla sanoo, ettei mitään erikoistietoja edellytetä.
- Musta pylväs ruudun vasemmassa laidassa osoittaa, että jonkinasteisesta ma-tematiikan osaamisesta on hyötyä.
- Pylväs keskellä kertoo tilastotieteen osallisuudesta asian ymmärtämisen kan-nalta.
- Oikealla oleva pylväs viestii tietojenkäsittelytaitojen ja Survon tuntemuksen hyödyllisyydestä.
- Kun merkintöjä yhdistellään, niin esim. tämän rivin alussa oleva ruutu kielii siitä, että matematiikkaa tulisi hallita hieman, tilastotiedettä ei lainkaan, mutta tietojenkäsittelyä melkoisesti.

Siis mitä "mustosemmalta" tienviitta näyttää, sitä vähemmän tarvitsee mate-maattispitoisia tarkasteluja pelkäävän lukijan ottaa siitä itseensä, ettei asiaa ymmärrä. Lukija hyppiköön harkintansa mukaan yli synkimpien kohtien. Suo-sittelen kuitenkin rohkeutta näidenkin jaksujen silmäilyyn.

Olen moneen otteeseen keskustellut *Marjut Schreckin* ja *Kimmo Vehkalahden* kanssa kirjan aihepiiristä ja siihen liittyvistä kysymyksistä. Ilman heidän myötä-eloan tämä tarina olisi ehkä jäänyt syntymättä. Kirjan käsikirjoitukseen sen eri vaiheissa ovat myös tutustuneet *Kalevi Kantele*, *Jouko Manninen*, *Olli Mustonen*, *Anna-Riitta Niskanen*, *Martti Puohiniemi*, *Lauri Tarkkonen*, *Osmo A. Wiio* ja vaimoni *Marja-Liisa*. Kiitän heistä jokaista kannustavista virik-keistä ja huomautuksista.

Hituniemessä syyskuussa 1996

Seppo Mustonen

Seppo.Mustonen@Helsinki.Fi



## Pientä laskentaa

- Käy jotenkin yli ymmärrykseni, että lisääntyvästä "helppokäyttöisyydestä" huolimatta on edelleen tilanteita, joissa yksinkertaiset tehtävät saattavat olla yllättävän hankalia yleisesti käytetyille ohjelmille.

Olen usein kysynyt, miten eri käyttöympäristöissä toimitaan, kun esim. kesken kirjeen kirjoittamisen tarvitsee tehdä joitain yksinkertaisia laskutoimituksia, joiden tulokset pitäisi liittää tekstiin ja sitten jatkaa kirjoittamista.

Ajatelkaamme vaikka tunnontarkkaa raportoijaa selostamassa:

Kiinassa on tällä hetkellä arviolta 612461200 mies- ja 586041300 naispuolista kansalaista. Väestömäärä on siis \_

Miten siis toimit käyttäessäsi mieliteksturiasi saadaksesi summan lasketuksi ja kirjoitetuksi suoraan lauseen jatkeeksi?

Vastaukset vaihtelevat. Jotkut sanovat, ettei heillä koskaan ole mitään tarvetta laskea kirjoittaessaan. Tuo tuntuu minusta uskomattomalta esim. insinöörin suusta kuultuna. Toiset kertovat pitävänsä tasku- tai pöytälaskinta koneensa vierellä ja tekevät sillä laskutoimituksia ainakin silloin tällöin. Erään kyselyn mukaan peräti 80 prosenttia PC:n käyttäjistä toimii näin. Siitäkö johtuu, että kuulemani mukaan markkinoilta löytyy nykyisin näppäimistöjä, joihin on istutettu kiinteästi tavallinen nelilaskin! Minusta se vaikuttaa yhtä jälkijättöiseltä kuin auton varustaminen aisoilla hevosvetoa varten. Eräät taas siirtävät esim. Windows-ympäristössä numeeriset tiedot leikepöydän kautta laskinohjelmalle ja tulokset sieltä samaa kautta takaisin - tosi kömpelöä!

Survo-ympäristössä tällaiset pikkutehtävät on jo vuoden 1982 alusta lähtien helpoiten tehty ns. kosketuslaskennalla. Esimerkissäni toimitaan niin, että

```
11 1 SURVO TOUCH MODE Mon Jun 24 17:26:59 1996 TOUCH 100 100 0
1 *
2 *
3 *Väestötietoja vuodelta 1994:
4 *
5 *Kiinassa on tällä hetkellä arviolta 612461200 mies- ja 586041300 nais-
6 *puolista kansalaista. Väestömäärä on siis 1198502500_
7 *
8 *
9 *
10 *
11 *
12 *
13 *
14 *
15 *
16 *
17 *
18 *
19 *
20 *
21 *
22 *
23 *
+612461200+586041300
```

siirryn ensin kosketustilaan (Touch mode) napilla F3, koskettelen lukuja halutuilla laskutoimitusnapeilla (+ - \* / jne.) ja sijoitan tuloksen tekstin perään osoittamalla kohdistimella sopivaa paikkaa ja painamalla nappia =. Lopuksi palaan tavalliseen kirjoitustilaan napilla ENTER. Tästä voi välittömästi jatkaa kirjoittamista. Kokenut Survon käyttäjä tekee tämän noin 7 sekunnissa eikä hänen tarvitse pelätä kirjoitusvirheitä.

Jos samaa yrittää taskulaskimella, jolloin luvut joutuu numero numerolta napputtelemaan laskimeen ja sitten tuloksen vastaavasti lauseen jatkeeksi, aikaa kuluu yli 30 sekuntia ja koko ajan saa pinnistellä virheitä välttääkseen. Huolellinen laskija jopa tarkistaa tuloksen, jolloin aika kasvaa jo minuutin paikkeille.

En usko, että Windows-virtuoosikaan selviää tehtävästä juuri alle puolessa minuutissa, vaikka paitsi teksturin myös järjestelmän oman laskurin ikkuna on valmiina esillä ja vaikka hän käyttää pikanäppäimiä.

Itse kokeilin tätä Windows 95:ssä kirjoittaen tekstiä "Notepad"illa ja laskien "Calculator"illa. Aikaa kului yli puoli minuuttia. Työskentely oli paljon epähavainnollisempaa ja sekä sormille että ajatuksille rasittavampaa kuin Survossa.

On sinänsä jotenkin surkuhupaisaa, että Windows-ympäristössä numeerisiin laskutoimituksiin yhä tyrkytetään taskulaskinta (jopa "tieteellistä" sellaista). Ymmärrän kyllä hyvän tarkoituksen; ihmiset ovat oppineet tavanomaiset laskinten ominaisuudet ja tätä kokemusta voi hyödyntää.

Erityismoitinnan ansaitsee Windowsin "tieteislaskimen tilastollinen boksi" (Statistical box), johon voi syöttää helposti lukuja ja laskea niistä joitain

perustunnuslukuja. Tämä boksi on kuitenkin tiedot nielaiseva musta aukko ja nimeksi sopisi paremmin "Statistical black box", sillä syötettyä lukujonoa ei pysty siirtämään esim. leikepöytään eikä sitä siis voi käyttää hyväksi millään muulla tavalla. Tällainen tuhlauk on vastoin hyvän tiedonhallinnan sääntöjä. Jokaisen kerran koneeseen syötetyn tiedon - esim. tässä tuon lukujonon - on käyttäjän halutessa säilyttävä niin, että sen voi palauttaa aina takaisin työhön muokkausta ja uutta käsittelyä varten.

Lienee paikallaan lyhyesti todeta, miten Survossa tuollaisen "tilastoboksin" tehtävät hoidetaan ilman em. haittoja. Katsokaamme siis tilannetta, jossa tekstiä kirjoitettaessa sekaan on ilmaantunut lukuja, joista pitäisi laskea tilastollisia tunnuslukuja:

```

11 1 SURVO 84C EDITOR Mon Jun 24 13:50:00 1996 C:\KIRJA\ 100 100 0
1 *
2 *Viime englannin kieliopin kokeessa annoin oppilailleni seuraavat
3 *numerot:
4 *
5 *
6 *Pekka 7
7 *Marjut 8
8 *Liisa 10
9 *Seppo 4
10 *Kimmo 6
11 *Tapio 9
12 *Bill 9
13 *Raimo 6
14 *Tuula 7
15 *Risto 8
16 *Juha 7
17 *Pentti 5
18 *Minna 8
19 *Ulla 5_
20 *
21 *
22 *
23 *

```

Tässä on näkymä Survon toimituskentästä, joka vastaa täysin teksturien kirjoitustilaa ja johon opettaja on kirjoittamassa kokeen tuloksista. Näkymä vastaa koko kuvaruutua. Moniajoympäristössä se on oma ikkunansa. Ylinnä ikkunassa on sinipohjainen palkki. Siinä näkyy mm. päiväys ja kellonaika. Lisäksi kerrotaan, että tällä hetkellä työskennellään KIRJA-nimisessä hakemis-tossa C-levyllä. Lopussa ilmoitetaan vielä toimituskentän nykyinen koko eli 100 riviä ja 100 saraketta.

Toimituskentästä on näkyvillä 23 ensimmäistä riviä. Opettajan tähän mennessä kirjoittamat tiedot päättyvät riville 19, jonka perässä kohdistin vilkkuu juuri nyt.

Tilastollisiin yhteenvetoihin Survossa on monia keinoja. Eräs helpoimmista on STAT-komento. Sen käyttö - kuten kaikkien tilastoaineistojen käsittelyyn tarkoitettujen - edellyttää havaintotaulukon nimeämistä DATA-määritelmällä.

```

11 1 SURVO 84C EDITOR Mon Jun 24 13:50:48 1996 C:\KIRJA\ 100 100 0
1 *
2 *Viime englannin kieliopin kokeessa annoin oppilailleni seuraavat
3 *numerot:
4 *DATA KOE
5 *Nimi nro
6 *Pekka 7
7 *Marjut 8
8 *Liisa 10
9 *Seppo 4
10 *Kimmo 6
11 *Tapio 9
12 *Bill 9
13 *Raimo 6
14 *Tuula 7
15 *Risto 8
16 *Juha 7
17 *Pentti 5
18 *Minna 8
19 *Ulla 5
20 *
21 *STAT KOE,22_
22 *
23 *

```

Opettaja on nyt kirjoittanut DATA-määritelmän riville 4 antaen aineistonsa nimeksi KOE, jolla siihen viitataan Survon eri toiminnoissa. DATA-rivin alapuolelle hän on lisännyt aineiston sarakkeiden nimet (Nimi nro), joihin myös saatetaan myöhemmin viitata. Opettaja on edelleen kirjoittanut riville 21 STAT-komennon, jossa mainitaan käsiteltävä aineisto (KOE) ja joka kertoo myös sen toimituskentän rivin, josta tulosten pitäisi alkaa. Tulokset saadaan, kun kohdistin on STAT-rivillä 21 (missä kohtaa tahansa) ja painetaan Survon aktivointinappia ESC.

```

11 1 SURVO 84C EDITOR Mon Jun 24 13:50:49 1996 C:\KIRJA\ 100 100 0
17 *Pentti 5
18 *Minna 8
19 *Ulla 5
20 *
21 *STAT KOE,22
22 *Basic statistics: KOE N=14
23 *Variable: Nimi
24 *Cannot be classified
25 *
26 *Variable: nro
27 *min=4 in obs.#4 (Seppo)
28 *max=10 in obs.#3 (Liisa)
29 *mean=7.071429 stddev=1.730464 skewness=-0.109062 kurtosis=-1.010608
30 *lower_Q=6 median=7 upper_Q=8
31 *nro f %
32 * 4 1 7.1 *
33 * 5 2 14.3 **
34 * 6 2 14.3 **
35 * 7 3 21.4 ***
36 * 8 3 21.4 ***
37 * 9 2 14.3 **
38 * 10 1 7.1 *
39 *

```

STAT-komennon aikaansaannokset kirjoittuvat valittuun paikkaan. Nyt opettaja on kelannut toimituskentän ikkunaa sen verran alaspäin, että näkee ne tässä kokonaan. Nimi-tiedolle STAT ei osaa tehdä mitään (siitä ilmoitus rivillä 24).

Kokeen tuloksesta (nro) syntyy sen sijaan laajempi yhteenveto riveille 26-38. Näemme esim. että kokeen keskiarvo on ollut noin 7.1 ja keskihajonta 1.7 . On olennaista, että tulostekstit ovat tekstiä siinä kuin opettajan itse kirjoittamat sanat ja lauseet. Niinpä hän voi helposti muokata tuloksia tekstinkäsittelyn keinoin päätyen tällä kohtaa vaikkapa lyhyeen yhteenvetoon:

```

59 1 SURVO 84C EDITOR Mon Jun 24 13:51:55 1996 C:\KIRJA\ 100 100 0
1 *
2 *Viime englannin kieliopin kokeessa annoin oppilailleni seuraavat
3 *numerot:
4 *DATA KOE
5 *Nimi nro
6 *Pekka 7
7 *Marjut 8
8 *Liisa 10
9 *Seppo 4
10 *Kimmo 6
11 *Tapio 9
12 *Bill 9
13 *Raimo 6
14 *Tuula 7
15 *Risto 8
16 *Juha 7
17 *Pentti 5
18 *Minna 8
19 *Ulla 5
20 *
21 *mean=7.071429 stddev=1.730464 skewness=-0.109062 kurtosis=-1.010608
22 *Keskiarvo oli 7.1 eli hieman huonompi kuin viime kerralla._
23 *

```

Kosketuslaskennan ohella Survossa on mukana ns. editoriaalinen laskenta, jossa myös suoraan, mutta aivan toiseen tyyliin, tehdään numeerisia toimituksia. Editoriaalinen laskenta on Survossa vielä varhaisempaa perua kuin kosketuslaskenta.

Jos haluan laskea tekstin sekaan kirjoitetun numeerisen lausekkeen arvon esim.  $(12.5+18)/2=15.25$ , niin tämä tehdään Survossa juuri kuten edellä näkyy. Kirjoitan siis  $(12.5+18)/2=_$  ja kursorin vilkkuessa merkin = perässä painan Survon yleistä aktivointinäppäintä ESC. Tällöin Survon toimitinohjelma, joka on Survo-istunnon aikana jatkuvasti läsnä, "tajuaa", että haluan laskea numeerisen lausekkeen, ja kutsuu automaattisesti Survoon kuuluvan editoriaalisen laskennan ohjelmamodulin käyttöön. Tuo ohjelma tutkii lausekkeen, laskee sen arvon ja kirjoittaa tuloksen (15.25). Kaikki tämä syntyy silmänräpäyksessä käyttäjän tarvitsematta tietää mitään siitä, mitä pinnan alla tapahtuu. Tulos on lausekkeen perässä aivan kuin olisin sen itse kirjoittanut ja saatan jatkaa työskentelyä esim. pyyhkimällä pois lausekkeen jättäen vain tuloksen näkyville.

Vastaavanlaisissa tehtävissä Windowsissa laskettavan lausekkeen joutuu siirtämään teksturista leikepöydän kautta laskimelle ja tuloksen taas takaisin teksturiin; se tietää aikaa ja vaivaa. Survossa riitti yksi ainoa ESC-napin painallus. Jos lausekkeessa esiintyy tavanomaisia funktioita, kuten neliöjuuri



(sqrt), niitä ei auta syöttää Windowsin laskimelle, sillä esim. sqrt(49)= tuottaa pelkän nollan. Jotta lausekkeen saisi tuolla laskimella lasketuksi, se on annettava muodossa 49@= , missä @ on laskimen neliöjuurinapin vastine. Lausekkeita, joissa on kirjainmerkinnöillä annettuja vakiota, on turha edes yrittää Windowsin laskimella.

Näytteeksi Survon editoriaalisien laskennan yleisestä luonteesta sopinevat seuraavat esimerkit:

```
37 1 SURVO 84C EDITOR Mon Jun 24 19:53:03 1996 C:\KIRJA\ 40 100 0
1 *
2 *Kolmion pinta-alan laskeminen:
3 *Olkoon kolmion sivujen pituudet a, b ja c. Tällöin kolmion pinta-ala on
4 *   Ala=sqrt(p*(p-a)*(p-b)*(p-c))   (ns. Heronin kaava)
5 *missä
6 *   p=(a+b+c)/2
7 *on siis kolmion piirin puolikas.
8 *
9 *Olkoon a=3, b=4 ja c=6. Tällöin Ala=_
10 *
```

Kyseessä on asetelma, jollaista Survossa sanotaan laskentakaavioksi. Tässä kerrotaan sangen vapaamuotoisesti mutta kaavojen osalta tarkasti, miten kolmion pinta-ala lasketaan. Kun siis aktivoin ESC-napilla rivin 9 juuri merkin = perästä, toimituskenttään ilmestyy välittömästi tulos

```
37 1 SURVO 84C EDITOR Mon Jun 24 19:53:03 1996 C:\KIRJA\ 40 100 0
1 *
2 *Kolmion pinta-alan laskeminen:
3 *Olkoon kolmion sivujen pituudet a, b ja c. Tällöin kolmion pinta-ala on
4 *   Ala=sqrt(p*(p-a)*(p-b)*(p-c))   (ns. Heronin kaava)
5 *missä
6 *   p=(a+b+c)/2
7 *on siis kolmion piirin puolikas.
8 *
9 *Olkoon a=3, b=4 ja c=6. Tällöin Ala=5.3326822519254
10 *
```

mistä sopii jatkaa vaikka poistamalla turhat desimaalit tai aktivoimalla uudelleen toisilla lähtöarvoilla. On erityisesti pantava merkille, ettei Survolla ole etukäteen mitään ohjelmaa tai muuta tietoa kolmion pinta-alan laskemiseksi. Nuo tiedot saadaan juuri riveiltä 4 ja 6 löytyvistä merkinnöistä. Survo toimii tässä eräänlaisena agenttina, joka omatoimisesti hakee tarvittavat tiedot ja suorittaa tehtävän niiden mukaisesti.

Toisena näytteenä ajateltakoon lainanottajaa, joka tulee maksamaan lainansa takaisin kuukausittain tasaerinä. Seuraavassa kaaviossa näytetään, miten lainanottaja voi tarkastella kuukausimaksun muodostumista eri laina-ajoilla ja korkoprosenteilla. Kaaviossa määritellään ensin oma kkmaksu-funktio Survon kirjastofunktion FIN.PMT avulla ja sitten on tehty muutamia koelaskelmia.

```

19 1 SURVO 84C EDITOR Mon Jun 24 19:58:22 1996 C:\KIRJA\ 40 100 0
11 * ACCURACY=2
12 *Lainan koron ja lyhennyksen suuruus kuukausittain lasketaan
13 *funktiolla
14 * kkmaksu(s,p,n):=FIN.PMT(s,p/12,12*n)
15 *missä
16 * s on lainan määrä,
17 * p vuotuinen korkoprosentti ja
18 * n lainaika vuosina.
19 *
20 *Siis esim. kun s=450000,
21 *kkmaksu(s,8,10)=5460
22 *kkmaksu(s,8,15)=4300
23 *kkmaksu(s,7.5,15)=4172
24 *

```

Kannattaa tehdä vertailuja. Jokainen voi tutkiskella, miten muissa ympäristöissä lausekkeita lasketaan. Kysynpä, montako painallusta tarvitset, että saat suosikkiteksturissasi koneen aidosti laskemaan, mitä on 2+2 ja vieläpä kirjoittamaan tuloksen 4 niin, että se on heti käytettävissä ja muokattavissa? Survossa noita painalluksia, kuten edellä kuvasin, tarvitaan 5 (eli  $2 + 2 = \text{ESC}$ ). Näistäkin 3 (eli  $2 + 2$ ) liittyy itse lausekkeen kuvaamiseen ja vain 2 (eli = ESC) itse toimituksen aktivointiin.

Joku saattaa ihmetellä, onko edellisen kaltaisten, pienten toimintojen tehokkaalla suorittamisella mitään olennaista merkitystä todellista tietoa käsiteltäessä. Vastaan, että onpa vain! Survon kaltaisessa yhdennetyssä järjestelmässä kaikki toiminnot rakentuvat hierarkkisesti eli laajemmat toimenpiteet ovat monien pienten peräkkäistä yhdistelyä. Se, mitä pienissä asioissa voitetaan, kertautuu laajemmissa yhteyksissä.

Survon suunnittelussa ja kehittämisessä on ollut keskeistä miettiä jokaisessa tilanteessa, eikö kyseistä työvaihetta voisi tehdä vielä yksinkertaisemmin (esim. pienemmällä määrällä kirjoittamista ja erikoisnäppäinten painalluksia). Jos vastaus on myöntävä, on pyritty löytämään se yksinkertaisempi tapa.

Tämänkaltaisessa tehokkuusajattelussa piilee vaara, että työskentely muuttuu kikkailuksi, joka edellyttää mitä konstikkaimpien näppäinyhdistelyjen muistamista. Survossa tämä vaara on vältetty eri tavoin. Ensinnäkin, Survon käyttäjä tulee toimeen varsin pienellä määrällä näppäintoimintoja, jotka ovat joko yhden tai kahden näppäimen takana. Toiseksi, Survon oma voimakas makrokieli tekee helpoksi toimintojen automatisoinnin käyttäjän haluamalle asteelle. Kolmanneksi, Survon kehittyneet työpöytätoiminnot mahdollistavat "yhden kosketuksen pelin" siirryttäessä työvaiheesta toiseen tai sovelluksesta toiseen. Kaikki tämä tietää, ettei käyttäjän tarvitse olla mikään muistihirviö eikä pikakirjoittaja hallitakseen töitään.





## Yleisperiaatteita

- Survon toiminnot perustuvat sen omaan editoriaaliseen käyttöliittymään. Tälle liittymälle on ominaista, että perustila, jossa työskentelen, on monin tavoin samanlainen kuin tavallisessa teksturissa. Saatoin siis kirjoittaa kuvaruudussa näkyvään toimituskenttään (edit field) tekstiä kuten kirjoituskoneella. Toimituskenttä poikkeaa kuitenkin tavanomaisen teksturin työtilasta siinä suhteessa, että tekstin sekaan voin kirjoittaa vaikkapa numeerisia lausekkeita (kuten nähtiin edellisessä luvussa). Aktivoimalla niitä ESC-näppäimellä saan suoraan lausekkeiden arvot tekstin sisään tai jatkeeksi.

Tekstin yhteyteen voin myös kirjoittaa kommentoja ja niitä täsmentäviä tietoja, jotka samoin aktivoidaan ESC-näppäimellä (esim. STAT-komento edellisessä luvussa). Kommentojen antamat tulokset ilmestyvät toimituskenttään haluamaani paikkaan jälleen tekstinä. Saatoin siis välittömästi muokata tuloksia siinä kuin omaa kirjoitustani. Lisäksi voin siirtyä milloin tahansa kirjoitustilasta vaikkapa kosketuslaskentaan, missä näkymä kentästä on entinen, mutta toiminta poikkeaa tavanomaisesta esim. laskutoimitusnäppäinten kohdalla.

Tässä mainitut menettelyt ovat vain esimerkkejä niistä mahdollisuuksista, joita Survo tarjoaa käyttäjälleen. On pyritty löytämään järkevä työnjako koneen ja ihmisen välillä. Käyttäjä ilmaisee itseään kirjoittamalla. Aina kun edessä on tilanne, jossa hän arvelee koneen olevan etevämpi - kuten esim. laskutoimituksissa - apu on lähellä.

Survossa ei käytetä lainkaan valikkoja siinä mielessä kuin ne ovat alinomaan esillä esim. Windowsissa. Valikkopohjainen työskentely olisi tietenkin opetteluaiheessa helpontuntuista, mutta kun jotain jo osaa, alkaa kaivata oikoteitä.

Kuten myöhemmin ilmenee, valikkoja ei suinkaan ole kokonaan hylätty Survossa. Mm. töiden yleinen hallinta ja käytön opastus tapahtuvat valikkojen ohjaamana. Yleensä luotetaan kuitenkin suoriin yhteyksiin käyttäjän ja hänen työnsä välillä.

## Ikkunat levällään

Siitä että käyttäjän ja koneen toiminnot yhtyvät samassa tilassa on paljon hyötyä. Tässä Survo poikkeaa ratkaisevasti muista käyttöympäristöistä, joille on ominaista asioiden karsinoiminen erillisiin ikkunoihin. Yleisesti rehottava ikkunointivillitys tuo mieleeni vanhan hölmöläistarinan, jossa kerrotaan mitä sitten tapahtui, kun huomattiin, että juuri rakennettu talo oli pimeä eikä valon kantaminen säkillä pirttiin ottanut onnistuakseen:

### VALOA PIRTTIIN

Sattui parahiksi paikalle Matti, joka huviksensa joskus pistäytyi Hölmölän kylään sen touhuja katselemaan.

Saatuaan kuulla, mistä oli kysymys, hän lupasi hoitaa pirtin valaisun nokkelammin. ...

Matti otti kirveensä ja nakutti sillä seinään pienen ikkuna-aukon. Hölmöläiset ihastelivat noin yksinkertaista, mutta tehokasta keksintöä.

Matin mentyä he päättivät hankkia vielä enemmän erinomaista valoa ja hakkasivat tuvasta koko seinän pois. Pirtti oli nyt täpötäynnä valoa.

- Mutta miksi se ei voisi olla vieläkin valoisampi, aivan huikaiseva, hölmöläiset intoilivat touhukkaina.

Ja niin he hakkasivat vielä toisenkin seinän. Kolmatta eivät ehtineet hakata, sillä tupa äkkiarvaamatta romahti kasaan...

(Eero Salola: Hölmölän kylä, Weilin+Göös, 1966)

Ikkunoinnissa ei siis pitäisi mennä liiallisuuksiin, mutta jo järkevissäkin puitteissa saatetaan joutua hankaluuksiin. Monille ympäristöille on tyypillistä, että esim. tekstinkäsittely ja taulukkolaskenta tapahtuvat kumpikin omassa maisemassaan. Tällöin siirtyminen näiden maisemien välillä on kitkaista.

## Toimituskentän synty

Oli onnekasta, etten saadessani ajatuksen editoriaalisesta työtavasta voinut tietää mitään taulukkolaskennasta. Tuolloin 70- ja 80-lukujen vaihteessa ensimmäinen taulukkolaskentaohjelma, Visicalc, oli vasta kehitteillä. Esikuvanani olivat senaikaiset tekstieditorit, erityisesti *Hans-Peter Sehmin* Suomessa Wang 2200-pientietokoneelle laatima, tekstinkäsittelyyn tarkoitettu ohjelma. Työskentelin silloin samalla pientietokoneella ja olin rakentanut sen vahvalla Basic-tulkilla SURVO 76-ohjelmistoa, jolle oli tunnusomaista vuorovaikutteinen, keskusteleva käyttötapa. Sehm oli laatinut editorinsa koneen mikrokoodilla, johon taas itse en ollut koskaan perehtynyt. Hänen ohjelmansa peruskäyttöliittymä oli kuitenkin ulkonaisena mallina omalle ratkaisulleni. Nyt ainoa jäljelle jäänyt yhteinen piirre Sehmin ja oman editorini välillä on rivien alussa oleva tähti (\*), josta Survoon on tullut moninaisia merkityksiä omaava kontrollisarake. Toinen lähtökohta oli tarve tehdä hyvä editori nuotinpainatus-

1 SURVO 76 EDITOR (C)1979 S.Mustonen ( 70x 75 )

```

1 *SAVE QUAD
2 *PRINT 4,61
3 *
4 *M(4/1/0)X(800/400)Quatrilles"
5 *M(2/1/0)X(0/200)Moderato con anima"
6 *M(2/1/0)X(1670/200)Olli Mustonen (1979)"
7 *G.T(3/4).V90.Q.4.L.75.F/4..
8 *YCO.X*M(1.8/1.2/15)*,-Hb1/8,Q.9,XC1,<Eb2,H1>.XC1.*
9 *F/4.,Q.8.YA3.X(1/4);
10 *-<Hb1/16.XC1.YA3.X*M(1.5/1/0) = ca.108",Eb2>.<Ab,Db3,AE,E>.X
11 *UBO.-<Hb1/8.XC1.C2>.XC1.,+<H1.XC1.A>.XC1.,<Hb,Ab>*
12 *G/4.,YCO.X+1.-CN2/8.XC2.<Ct.F#>.X+1.,XC2.,R#
13 *-<A2/8.XC2.YGO.X-1.GM2>.XC2..
14 *<F.Eb/16.XC2.Db>.XC2.,+<H1.XC2.A.XC2.,G,F>.X-1.*
15 *UBO.<E/8.XC1.F#>.XC1.,<Ft.XC1.Eb>.XC1.,<Et,D>*
16 *<C#8/8,Z.,C>.Z.,<F#8.Z.,Ct>.Z.,<H0.Z.,F#1>.Z.*
17 *U100.K.2.<E1/8:AO,C#E2>.YCO.X*M(1.8/1.2/15)*#
18 *K9.-<Eb2/16.D.F#8>.<Ft/16.XC1.Eb.XC1.,Ab1.Hb>.R#
19 *UBO.,+<F#8.F/16.XC1.G#>.XC1..
20 *<Ft/8.F/16.XC1.Gt>.XC1.,<E/8.D#1/16.YC.XC1.C#>.YH0.XC1.*
21 *U90.,+<Ct1/8:G.E2>.YCO.X*M#
22 *K9.L.75.-<F#2/16.Ft.A/B>.<G#8/16.XC1.H.XC1.,Ct.F>*
23 *-<E2.YF.XC1.C# YA.XC1.,Eb,D>.<C#8.D#8.C.H1>.V80.,+<Ct2.H1.F#8.G#>.R

```

## Quatrilles

Moderato con anima  $\text{♩} \cdot \text{ca.108}$  Olli Mustonen (1979)

Tämä on pienennetty kopio ikivanhasta kalvosta, jolla aikoinaan havainnollistin nuottieditorin toimintaa. Kuva on samalla tässä kirjassa ainoa, jota ei ole tehty suoraan nykyisellä Survolla. Sävellyksen kolmas tahti on toimituskentässä kuvattu alleviivattuna koodijonon rivillä 11.

Nuotinpainatusohjelmaa suunniteltaessa kiintoisinta oli jäljitellä nuotinpainatuksen estetiikkaa kuvastavia sääntöjä matemaattisesti. Esim. kaaren piirtämisessä lähtökohtana oli syklodi eli ympyrän kehällä sijaitsevan pisteen ura ympyrän vieressä. Säättämällä kaaren korkeutta lisäparametrilla jopa kolmella eri arvolla sain mielestäni kauniin muodon ja sopivan viivanlevyden vaihtelun.

ohjelmaani varten. Tuon ohjelman tein harrastusmielessä auttaakseni silloin 12-vuotiaista poikaamme Ollia nuottien puhtaaksipiirtämisessä. Siinä yhteydessä tuli mieleeni, että samalle työtavalle saattaisi olla sijaa tilastollisen aineiston hallinnassa tekstinmuokkauksen ohessa. Näin nuotinpainatusohjelman jatkokehittely sai jäädä ryhtyessäni innokkaasti kokeilemaan editoriaalista käyttötappaa ja etsien sen pelisääntöjä.

Survossa ei ole juopaa tekstin ja taulukkojen käsittelyn välillä hyvin yksinkertaisesta syystä; kummassakin tapauksessa toimitaan niin, että pienin käsiteltävä yksikkö on toimituskentässä näkyvä merkki. Taulukkolaskennassahan pienin yksikkö on suurempi eli solu, jossa voi olla esim. moninumeroinen luku, sana, jopa kaava. Survossa ei tällaista solun käsitettä lainkaan tunneta, vaan ohjelmalla tulee olla kyky hahmottaa mm. taulukot tekstin joukosta ja siihen se tarvitsee vain vähäistä opastusta.

## Lajittelua

Tässä on eräs Survon toimittimen (editorin) varhaisimpia sovelluksia, toimituskenttään kirjoitetun taulukon lajittelu:

```

13 1 SURVO 84C EDITOR Fri Jun 28 18:46:33 1996 C:\KIRJA\ 100 100 0
1 *SAVE KULUTUS
2 *
3 *Nautintoaineiden kulutuslukuja/asukas vuodelta 1979
4 *
5 * Maa Kahvi Tee Olut Viini Viina
6 * Suomi 12.5 0.15 54.7 7.6 2.7
7 * Ruotsi 12.9 0.30 58.3 7.9 2.9
8 * Norja 9.4 0.19 43.5 3.1 1.8
9 * Tanska 11.8 0.41 113.9 10.4 1.7
10 * Englanti 1.8 3.49 113.7 5.1 1.4
11 * Irlanti 0.2 3.73 124.5 3.8 1.9
12 * Hollanti 9.2 0.58 75.5 9.7 2.7
13 * Sveitsi 9.1 0.25 73.5 44.9 2.1
14 * Ranska 5.2 0.10 44.5 104.3 2.5
15 * Italia 3.6 0.06 13.6 106.6 2.0
16 * Espanja 2.5 0.03 43.6 73.2 2.7
17 * Portugali 2.2 0.03 27.5 89.3 0.9
18 * 11111
19 *SORT 6,17,18_
20 *
21 *
22 *
23 *

```

Taulukko on joko kirjoitettu itse tai poimittu toimituskenttään esim. tekstitiedostosta. Sitä on saatettu muotoilla Survon tarjoamin keinoin niin, että sarakkeet on saatu suoristetuksi. Tällaisen taulukon voi järjestää minkä tahansa sarakkeen mukaan Survon SORT-komennolla.

Aikomuksena on lajitella maat oluen kulutuksen perusteella. Itse komennon saa kirjoittaa mille vapaalle riville tahansa. Tässä se on sijoitettu riville 19.

SORT-komennon perään kirjoitetaan pilkkujen tai välilyöntien erottamina ensimmäisen lajiteltavan rivin numero (6), viimeisen rivin numero (17) ja ns. maskirivin numero (18). Maskirivi osoittaa sen sarakkeen, jonka suhteesta lajitellaan. Tässä numeroin kirjoitettu maski (11111) tarkoittaa numeerista lajittelua.

Tämän jälkeen, kohdistimen ollessa SORT-rivillä 19, painetaan aktivointinappia ESC ja Survo järjestää taulukon rivit muotoon:

```

13 1 SURVO 84C EDITOR Fri Jun 28 18:46:33 1996 C:\KIRJA\ 100 100 0
1 *SAVE KULUTUS
2 *
3 *Nautintoaineiden kulutuslukuja/asukas vuodelta 1979
4 *
5 * Maa Kahvi Tee Olut Viini Viina
6 * Italia 3.6 0.06 13.6 106.6 2.0
7 * Portugali 2.2 0.03 27.5 89.3 0.9
8 * Norja 9.4 0.19 43.5 3.1 1.8
9 * Espanja 2.5 0.03 43.6 73.2 2.7
10 * Ranska 5.2 0.10 44.5 104.3 2.5
11 * Suomi 12.5 0.15 54.7 7.6 2.7
12 * Ruotsi 12.9 0.30 58.3 7.9 2.9
13 * Sveitsi 9.1 0.25 73.5 44.9 2.1
14 * Hollanti 9.2 0.58 75.5 9.7 2.7
15 * Englanti 1.8 3.49 113.7 5.1 1.4
16 * Tanska 11.8 0.41 113.9 10.4 1.7
17 * Irlanti 0.2 3.73 124.5 3.8 1.9
18 * 11111
19 *SORT 6,17,18_
20 *
21 *
22 *
23 *

```

## Palapeli mutta ei palikkatesti

Nykyisillä laitteilla toimitus on salamannopea siitä huolimatta, että todellisuudessa Survon toimitin ei itse tiedä lajittelusta mitään vaan tarvitsee siihen toisen ohjelman apua. Yleensäkin, kun Survossa aktivoidaan jokin komento, toimitin vain tutkii komentosan ja kutsuu väliaikaisesti avukseen tätä sanaan vastaavan erillisen ohjelman eli tässä tapauksessa Survon SORT-ohjelmamodulin. Tämä ohjelma setvii komennon parametrit ja toimii niiden mukaisesti, siis lajittelee rivit ja kirjoittaa ne takaisin toimituskenttään uudessa järjestyksessä. Tehtyään tämän SORT-ohjelma väistyy ja itse voin jatkaa työskentelyä toimittimella.

Jotta mitä erilaisimmat toimenpiteet saadaan tehdyksi toimituskentästä ohjattuina, Survo jakautuu satoihin ohjelmamoduleihin, joista toimitinohjelma esim. käynnistetyn komennon perusteella kutsuu sopivan. Ohjelmamodulien koot vaihtelevat 20 kilotavusta pariin sataan; yleensä ne ovat alle satakiloisia. Siirtyminen modulista toiseen on vikkellä. Käyttäjä ei siirtymisiä juuri huo-



maa eikä hänen tarvitse niistä välittääkään. Survon tulisi vaikuttaa saumattomalta kokonaisuudelta. Näin tällä palapelimaisella järjestelmän arkkitehtuurilla vältetään tarpeettomat odotusajat, jotka ovat ainakin vielä turhauttavan pitkiä esim. Windows-ympäristössä. Survo mm. käynnistyy parissa sekunnissa kun taas Windows vaatii ainakin puoli minuuttia.

Lainatakseni *Lauri Tarkkosen* vertausta, Survon käyttö on hengeltään rinnastettavissa Lego-palikoilla leikkimiseen. Voin yhdistellä palikoita vapaasti oman mielikuvituksen kannustamana eikä minun ole pakko toimia minkään palikkasarjan mallikuvan mukaisesti. Siis "auton" asemasta saan yrittää rakentaa vaikka "lentokonetta".

## Lajitteluesimerkin paluu

Viime tilanteessa SORT-komento on jäänyt edelleen riville 19. Voin pyyhkiä sen pois tai käyttää sitä uudelleen. Jos vaikkapa tahtoisinkin rivit päinvastaiseen järjestykseen eli oluen kulutuksen mukaan suurimmasta pienimpään, riittää, että lisään komentosanan SORT eteen merkin - ja aktivoin uudelleen:

```

13 1 SURVO 84C EDITOR Fri Jun 28 18:47:12 1996 C:\KIRJA\ 100 100 0
1 *SAVE KULUTUS
2 *
3 *Nautintoaineiden kulutuslukuja/asukas vuodelta 1979
4 *
5 * Maa Kahvi Tee Olut Viini Viina
6 * Irlanti 0.2 3.73 124.5 3.8 1.9
7 * Tanska 11.8 0.41 113.9 10.4 1.7
8 * Englanti 1.8 3.49 113.7 5.1 1.4
9 * Hollanti 9.2 0.58 75.5 9.7 2.7
10 * Sveitsi 9.1 0.25 73.5 44.9 2.1
11 * Ruotsi 12.9 0.30 58.3 7.9 2.9
12 * Suomi 12.5 0.15 54.7 7.6 2.7
13 * Ranska 5.2 0.10 44.5 104.3 2.5
14 * Espanja 2.5 0.03 43.6 73.2 2.7
15 * Norja 9.4 0.19 43.5 3.1 1.8
16 * Portugali 2.2 0.03 27.5 89.3 0.9
17 * Italia 3.6 0.06 13.6 106.6 2.0
18 * 11111
19 *-SORT 6,17,18
20 *
21 *
22 *
23 *

```

Jos taas haluan asettaa maat taulukossa aakkosjärjestykseen, riittää korvata nykyinen maski rivillä 18 pelkkien A-kirjaimien jonolla sarakkeen Maa kohdalla ja aktivoida SORT jälleen. Kannattaa miettiä, miten teet ylläolevat asiat mieliteksturillasi.

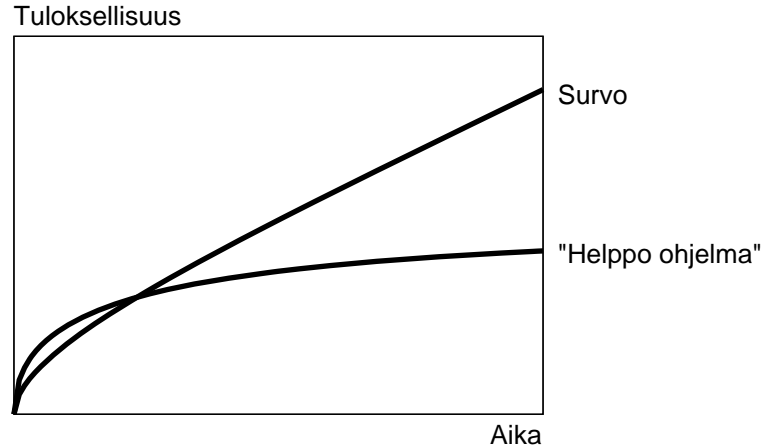
## Wysiwyg vai Wywiwyg

Esim. MS-Wordissa (ver. 7) taulukkoja voi lajitella. Se edellyttää kuitenkin liikkumista teksti- ja taulukkotilan välillä, mikä mielestäni on melko tahmeaa. Wordissa erillinen taulukkotila on välttämätön, jotta ohjelma pysyisi selvillä taulukon rakenteesta ja rajoista. Paljolti nuo hahmotusvaikeudet johtuvat Wysiwyg- (What you see is what you get) periaatteen noudattamisesta. Kun siis tekstin tulisi jo kuvaruudussa näkyä lopullisessa painoasussa ja kun mielellään käytetään vaihtuvanlevyisiä kirjaimistoja, taulukot menevät teksti-tilassa winksin wonksin.

Survossa - leikillisesti todeten - vallalla on Wywiwyg-periaate eli What you **want** is what you get. En pidä mielekkäänä sitä, että kaiken pitäisi koko ajan näkyä ruudulla lopullisessa painoasussa vaan ainoastaan tarvittaessa. Survon käyttäjän tulee "soittaessaan" olla "nuotinlukutaitoinen" kun esim. Windowssissa "tapailaan korvakuulolta". "Nuotinlukutaito" avaa valtavan paljon enemmän mahdollisuuksia itsensä toteuttamiseen myös tiedonkäsittelyssä.

Mielivertaukseni on kauan ollut myös viulunsoittotaito. Kuvittele, että jousisoittimia ei olisi ollutkaan olemassa, vaan joku vasta nyt lähtisi sellaista kehittämään. Sen perusteella miten ohjelmien käytön vaativuutta yleensä arvioidaan, olisi turha toivoa viululle hyväksyntää, koska siinä on niin hankala "käyttöliittymä"; ei mitään merkintöjä tai nauhoja otelaudassa kuten kitarassa puhumattakaan valmiista koskettimistosta. Jos kaikesta huolimatta joku sinnikäs muusikko saataisiin suostumaan moisen soittimen omaehtoiseen opetteluun, kestäisi ainakin kuukausia, ennenkuin soitto olisi muuta kuin korvia ja hermoja riipivää. Survossa eivät kuitenkaan vaativimmatkaan soittelut vedä vaikeudessaan vertoja esim. Bachin Chaconnen esittämiseen - sen tiedän kokemuksesta. Hyvän käyttöympäristön tulee kuitenkin olemukseltaan olla enemmän viulun kuin esim. automatiikalla tehostettujen sähköurkujen kaltainen. Viululla ylletään rikkaampaan musikaalisuuteen ja häikäisevämpään virtuositeettiin kuin sähköisillä ja mekaanisilla "soittokoneilla".

Vaikka käytänkin tuollaisia vertauksia, ei pidä kuvitella, että Survon opettelu olisi jollain tapaa erityisen hankalaa - se on vain erilaista. Helppokäyttöisyys, jolla esim. graafista käyttöliittymää perustellaan, on sitä pääasiassa vain aloittelevalle käyttäjälle. Mainoslauseilla tyyliin "Vain 5 pientä minuuttia ja olet mestari..." ihmiset on tuuditettu uskoon, että kaiken pitäisi olla mukavaa ilman opetteluun tuskaa. Kun ajatellaan, miten käyttötaito kehittyy ajan myötä, *Lauri Tarkkonen* on hahmotellut Survon ja "helppokäyttöisten" ohjelmien eron oppimiskäyrinä seuraavasti:



Helpon ohjelman oppiminen sujuu aluksi nopeasti, mutta myös melko pian saavutetaan tilanne, jossa ei ole enää juuri mitään uutta opittavaa. Helppouden varjopuolena on useimmiten mahdollisuuksien rajallisuus. Survon omaksuminen - ainakin joissain asioissa - saattaa olla aluksi hitaampaa, mutta taitojen karttuessa tuloksellisuudelle ei ole mitään muuta rajaa kuin käyttäjän oma mielikuvitus ja oppimiskyky.

### Toimituskentän talletus

Milloin tahansa toimituskenttä on tallettavissa omaksi tiedostokseen. Tätä varten on tyypillistä (vaikkei pakollista) kirjoittaa kentän ensimmäisen rivin alkuun SAVE-komento, jonka perässä mainitaan tiedoston nimi (lajitteluesimerkissä KULUTUS). Siis, jos siirrän kohdistimen tälle riville ja painan ESC-nappia, Survo tallettaa toimituskentän. Tiedosto sijoittuu hakemistoon, jonka polkunimi näkyy toimituskentän otsikkorivin loppupäässä eli tässä tapauksessa C-levyn hakemistoon KIRJA.

Vastaavasti talletetut toimituskentät otetaan uudelleen esiin LOAD-komennolla, jonka parametrina on kentän nimi. Niitä katsellaan SHOW-komennolla, jolloin nykyiseen työhön voi poimia esim. osia aikaisemmasta toimituskentästä. Survossa on helppo luoda uusia työhakemistoja ja siirtyä niihin toisista hakemistoista. Tästä puhutaan seuraavassa luvussa.

Pane merkille, että Survossa kaikki osatekijät, siis käsiteltävät tiedot, niitä ohjaavat komennot täydennyksineen ja komentojen antamat tulokset esitetään samassa tilassa, toimituskentässä. Ei myöskään ole eroa sillä, onko käsittelyn kohteena sanallinen vai numeerinen tieto. Asiat tulkitaan halutulla tavalla toiminnan luonteen mukaisesti.

## Hiiri työttömänä?

Monet pitävät Survoa vanhentuneena, kun se on "pelkkä" DOS-ohjelma ja kun siinä ei hiirellä ole mitään virkaa. Tämä pitää vain osaksi paikkansa. Survon itsenäisessä käytössä hiirtä ei tosiaankaan tarvita. Survon editoriaalisessa käyttöliittymässä näet nopeutettu, nuolinäppäimin ohjattu kohdistin on itse asiassa hiirtä näppärämpi osoitinväline. Koneiden normaaliasetuksissa kohdistimen liikuttaminen on hidasta, mutta se saadaan kulkemaan liukkaasti esim. DOS-komennolla

```
>MODE CON RATE=32 DELAY=1 ,
```

jonka Survon käyttäjät sijoittavat yleensä valmiiksi käynnistystiedostoon.

Kun korostetaan graafisen käyttöliittymän erinomaisuutta, mielessä vertailukohteena ovat todella vanhanaikaiset teksti- ja komentopohjaiset ratkaisut. Vaikka Survossakin esiintyy komentoja ja toimitaan pääosin tekstinäytössä, toiminta on luonteeltaan vähintäänkin semigraafista (puoligraafista) ja täysin verrattavissa hiirellä ohjailtuun käyttötapaan. Survossa voi haluttaessa peittää peruskäyttöliittymän ja korvata valikkojenvaraisella, kuten on tehty esim. Survon perusaloituksessa.

Survo sopeutuu lisäksi mainiosti esim. Windows 95:n alaisuuteen, jolloin hiirtä rakastavat saavat mielikilleen töitä ihan tarpeeksi. Voin pitää hallussani vaikkapa useita Survoja eri ikkunoissa samanaikaisesti ja siirtää tietoja ikkunasta toiseen. Tosin tämä tapahtuu kätevimmin Survon omien keinojen kautta; ei siihen leikepöytä tarvita.

Toisaalta Survossa ei ole mitään esteitä hiiren käyttöönotolle. Sekä *Juha Puranen* että *Jouko Manninen* ovat rakentaneet mm. tilastotieteen opetusta varten Survon päälle toimintoja, joita ohjataan hiirellä.

## Toimituskentän rajat

Toimituskenttä on kooltaan rajattu tila, koska sitä vastaa tietty alue koneen nopeassa keskusmuistissa. Maksimikoko on noin 64000 merkkiä eli siis esim. 600 riviä, 100 saraketta kullakin. Tämä saattaa hämmentää niitä, jotka ovat tottuneet käsittelemään mielivaltaisen pitkiä aineistoja ja tekstijonoja yhtenä kokonaisuutena.

Survossa tuo rajoitus on kuitenkin näennäinen. Survossa suositetaan sellaista järkevää käytäntöä, jossa esim. laajat numeroaineistot talletetaan omiin tie-

dostoihinsa pakatussa muodossa ja jossa tekstit luonnollisella tavalla, esim. lukuihin tai kappaleisiin jakautuen, sijoitetaan useisiin toimituskenttiin. Tällainen hajautus ei kuitenkaan estä näkemästä ja hallitsemasta tietoa kokonaisuutena. Esim. mielivaltainen jono eri toimituskenttiin kuuluvia osia on helposti määriteltävissä yhdeksi Survon LIST-rakenteeksi. Laajoja havaintoaineistoja taas voi katsella ja muokata Survon FILE-komennoin ikäänkuin kaikki tiedot olisivat samassa toimituskentässä.

Itse toimin tavallisimmin oletustilassa, joka on 100 riviä ja 100 saraketta. Kirjoitustöissä saatan käyttää hieman pitempiä kenttiä kuten tämän tekstin hallinnassa 100-400-rivisiä.

## **Vielä periaatteista**

Nykyistä Survoa ei ole luotu hetkessä eikä sen työskentelymuotoja - ei edes sisäisiä rakenteita - ole päätetty kerralla. Survo on noin 30-vuotisen historian-aikana elänyt erilaisia vaiheita. Uudelleen olen aloittanut kolme kertaa, viimeksi vuonna 1985, jolloin nykyinen C-kielinen laitos lähti liikkeelle. Valikojen rajoitusten vuoksi valitsin työskentelyn perustaksi editoriaalinen käyttötavan. Se oli vähitellen ilmaantunut keskustelevan Survon kylkeen jo vuodesta 1979 lähtien.

Nykyurvossa pidetään tiukasti kiinni siitä periaatteesta, ettei uudistusten myötä mitään aikaisempaa toimintamuotoa hylätä. Uusia luodaan ja vanhoja uudistetaan, mutta perinteitä kunnioittaen. Käytännössä tämä tarkoittaa, että kaikki aikaisemmalla Survolla tehdyt työt tulee voida toistaa sellaisenaan uusimmalla versiolla. Koska Survossa tärkeimmät ja laajimmat hankkeet toteutetaan Survon omalla makrokielellä ns. sukroina, yhteensopivuus on alati tarkkailun kohteena. Juuri sukroina tehdyt asiantuntija- ja opetussovellukset koetlevat järjestelmän toimivuutta erittäin rankasti.

Tiukka yhteensopivuuden vaaliminen ei ole estänyt tuomasta mukaan viimeisen kymmenen vuoden aikana - koneiden nopeutuessa - työtapoja, joita ei olisi voinut kuvitellakaan alussa. Sukrokielen kehitys on tästä oiva esimerkki.

Survon eteneminen on sillä tapaa liukuvaa, että uusia versioita ilmestyy parin viikon - kuukauden välein. Tällä vältetään ne ikävät yllätykset, joita suurin hyppäyksin tapahtuvat ja rakenteisiin kajoavat uudistukset aiheuttavat useiden muiden ohjelmien vakiintuneille käyttäjille. Saatan joskus "ajaa sisään" lisäyksiä ja parannuksia, jotka ovat mukana uusimmissa versioissa, mutta joista ei laajemmin kerrota ennen kuin urhoolliset koekäyttäjät ovat niitä tarpeeksi testanneet.

Survon modulaarisuus auttaa paljon järjestelmän ylläpidossa. Koska jokainen ohjelmamoduli on itsenäinen ohjelma, sitä voi rauhassa parannella ilman, että tarvitsisi huolehtia Survon muista toiminnoista.

Survohan itsenäisenä käyttöympäristönä sallii jopa sen, että parantelen tai laajennan sitä itseään käytön aikana. Kirjoitan siis Survon toimituskenttään C-koodia, jonka käänän Survon alaiseksi ohjelmamoduliksi. Välittömästi - poistumatta Survosta - voin myös lähteä soveltamaan uutta ohjelmaa omana komentonaan, omalla nimellään. Tämä ilo suodaan käyttäjille, jotka hankkivat tarvittavan C-kääntäjän (Microsoftin C, versio 6) ja Survon omat ohjelmakirjastot (vapaasti kopioitavissa). Se on kuitenkin harvojen herkkua. Survon käyttäjän ei toki tarvitse tietää C-kielestä eikä muistakaan tämänluonteisista asioista yhtään mitään.

Tuo joustava kehitystekniikka selittää osaltaan, miksi Survo on voinut laajeta nopeasti sängen rajoitetuin resurssein. Modulaarisuus takaa senkin, ettei järjestelmä haukkaa levyiltä eikä keskusmuistista liikaa tilaa. Survo toimii jatkuvasti jopa IBM XT-tasoisessa koneessa, kunhan levytilaa on vapaana 20 megatavua. Niinpä olen nähnyt sen pyörivän vain 300 grammaa painavassa, taskulaskimen kokoisessa laitteessa. Kuitenkin, mitä tehokkaampi laitteisto, sitä enemmän siitä on hyötyä survoillessakin.





## Töiden hallinta

- Monitaitoinen ja -tavoitteinen käyttäjä ansaitsee mukavat keinot töittensä hallintaan. Survossa vaihtoehtoisia tapoja on useita. Kun seuraa survoilijoiden työskentelyä, tuon vaihtelevuuden totisesti huomaa. Mm. omassa tilastollisen tietojenkäsittelyn seminaarissani ja Survon käyttäjäyhdistyksen tilaisuuksissa näistä kysymyksistä on monta kertaa keskusteltu. On ilmennyt, ettei töiden järjestelyyn voi olla mitään yksiselitteisesti parasta tapaa. Valinnat riippuvat suuresti käyttäjän tehtävien luonteesta ja hänen tavoistaan mieltää rakenteita. Silti on seikkoja, joista ei tarvitse kiistellä. Aloittelija jää helposti tilaan, jossa tiedot liikaa kasautuvat samoihin hakemistoihin. On tärkeää osata jäsentää työkohteensa niin, että mikä tahansa niistä löytyy välittömästi vuosienkin jälkeen. Tässäkään ei tule mennä liiallisuuksiin eli jotkut saattavat päätyä turhan moniportaisiin hakemistorakenteisiin. Töiden järjestelyssä ei voi pelkästään luottaa käyttöjärjestelmän tai minkään sovellusohjelman apuun. Survon alkuvaiheessa on kuitenkin uusi kohta (Töiden yleinen hallinta), jossa annetaan ohjeita käyttäjille näissä kysymyksissä. Tulee muistaa, että avaimet on silti pidettävä omassa käsissä.

Mieleni tekee kertoa, miten nykyisin olen järjestänyt "työpöytäni" Survossa niin, että löydän helposti kaikki aikaisemmat jutut, aineistot ja asiakokonaisuudet.

Työpöytämaainen toiminta on tullut Survoon jo silloin, kun editoriaalinen työtapa syntyi. Survossa toimituskentät ovat luonnollisia "työpöytiä". Niihin voi asetella tarvitsemansa linkit kunkin tehtävän eri osiin esim. selityksin varustettuina LOAD-komentoina haluamallaan tavalla. Asiat on järjestettävissä niin, ettei mitään tiedostonimiä, polkunimistä puhumattakaan, tarvitse muistaa ulkoa. Pitäisi riittää, että nimet on kertaalleen valittu ja kirjoitettu. Sen jälkeen tulee saavuttaa nimetyt kohteet ilman turhia miettimisiä ja selailuja.

Omien töitteni hallinnan perustana ovat automaattisesti ylläpidettävät valikot. Kun käynnistän Survon koneessani, kuvaruutuun ilmaantuu seuraavanlainen valikko; tämä oli tilanne 7. heinäkuuta 1996:



```

1 1 1 SURVO 84C EDITOR Sun Jul 07 09:34:08 1996 C:\KIRJA\ 100 100 0
  Työvalikko C:\E\D\SURVO.JOB
  0. S1 Survon päivitys C:\S\ 6 6 1 1 /
  1. INDEX Kirjeet, lausunnot etc. C:\ARK\ 5 1 1 1 /
  2. P Puhelinnrot, tietoliik. C:\ARK\ 2 6 1 1 /
  3. INDEX Arkistoituja levykkeitä C:\Z\ 4 1 1 1 /
  4. OPETUS Kurssit ja seminaarit C:\ARK\ 3 1 1 1 /
  5. ASENNUK SURVO 84C-levykkeet C:\PR\ 1 1 1 1 /
  6. TILAUS Tilauslomakkeet C:\SS\TILAUS\ 7 1 1 1 /
  7. TILIT7 Tilinpito 1995-96 C:\SS\ 1 1 1 1 /
  8. SURVO5 Etiketit C:\ETIK\ 24 1 24 1 /
  9. Jobfile: C:\M\MONIM.JOB / Monimuuttujamenetelmät, hyperteksti .
  A. Jobfile: C:\E\D\SM.JOB / Tekeillä olevat työt .
  B. Sucro: ESITTELY 12.10.1995 C:\E\TUT\ +
  C. KOPIO Varakopioiden teko C:\S\ 4 1 1 1 /
  D. RAMSUR2 Survo to Ramdrive C:\E\D\ 4 1 1 1 /
  E. INDEX Missing observations C:\MISS\ 4 1 1 1 /
  F. INDEX Survo ja minä C:\KIRJA\ 10 1 1 1 /
  G. INDEX Windows-kokeet C:\WINKOE\ 4 1 1 1 /
  H. INDEX GENREG-parannus C:\GENREG\ 4 1 1 1 /
  Valitse työ ao. napilla. Lisää työ napilla +. ENTER=Paluu ?=HELP
  S. Survon alkuvalikko M. Päätyövalikko

```

Tästä valikosta siirryn haluttuun kohteeseen yhdellä napin painalluksella. Siis esim. painamalla nappia F (kohde kolmanneksi viimeisellä rivillä) pääsen käsittelemään tätä kirjoitustani (Survo ja minä) koskevia asioita.

Vaikka paperilla musta, valkoinen ja harmaan eri sävyt eivät ilmaise kunnolla valikon todellisia väritehosteita, näkymä ei varmasti ole yhtä korea kuin vaikkapa Windowsin ikonein somistettu kuvaruutu, mutta mielestäni on parempi toimia "kerakkein kuin kuvakkein". Toisin sanoen ytimekkäät sanalliset kohteiden selitykset ovat lukutaitoisille otollisempia kuin pelkät ikonit.

Ei ole suuremmin aihetta lämmittää kiistaa eri käyttöliittymien eduista ja haitoista. Silti haluan palauttaa mieleen sen, mitä *Rudolf Arnheim* sanoi jo vuonna 1935:

**"Asioiden kuvaamisessa on yleistyksen johdettava yksityistapauksista; on valittava, verrattava, ajateltava. Mutta kun viestintä voi tapahtua sormella osoittamalla, suu hiljenee, kirjoittava käsi pysähtyy ja mieli kutistuu."**

Tämä lainaus on otettu Neil Postmanin kirjasta *Lyhenevä lapsuus* (WSOY 1982, suomentaja Ilkka Rekiarvo) sivulta 82. Arnheimin varoitus tarkoitti silloin läpimurtoaan tekevää näkötietoa, mutta sama koskee ihmisen ja tietokoneen välistä kanssakäymistä.

Survon työvalikossa kohteina esiintyy paitsi toimituskenttiä, myös toisia työvalikkoja (Jobfile: kohteet 9,A) ja Survon makroja eli sukroja (Sucro: kohde B). Tässä kohde B aloittaa automaattisen Survon viitosversion esittelyn, joka sukrotekniikalla tekee ja näyttää asiat kuvaruudulle ilmestyvinä värikkäinä ja grafiikalla höystettyinä "kalvoina".

Kunkin kohteen kuvaukseen on varattu 8+24 merkkiä eli DOS-käyttöjärjestelmän 8 merkin nimirajoitus on näin häivytetty. 24 lisäkirjainta antaa keinot kohteen riittävän luettavaan kuvailuun ja on enemmän kuin mitä ikoni kertoo. Uskallan myös väittää, että "yhden kosketuksen pelin" mukainen valinta on nopeampaa kuin kohdistimen kuljettaminen hiirellä tiettyyn paikkaan ruudulla ja kaksoisnäpätys.

Valikon oikeassa reunassa paljastetaan mm. kohteiden levy- ja polkutunnukset. Niistä käyttäjän ei tarvitse välttämättä tietää mitään. Tällaisia valikkoja luodaan ja muutetaan Survossa helposti. Esim. uuden työn lisääminen valikkoon tapahtuu painalluksella + ja työn poistaminen napilla - . Nappi ? kertoo tarkemmin työvalikkojen hallinnasta, esim. siitä, miten rivit voidaan värittää toisistaan eroaviksi. Milloin tahansa nykyinen työvalikko saadaan uudelleen esiin joko napilla F11 tai näppäinyhdistelmällä Ctrl-A.

Kun työ lisätään valikkoon, edellytetään, että sitä vastaava levyhakemisto on perustettu ja sinne on luotu toimituskenttä, joka toimii esim. työn tarkempana hakemistona. Tuo hakemistokenttä talletetaan ensimmäiselle riville sijoitetulla SAVE-komennolla, johon on liitetty sopiva selitys, esim.

SAVE INDEX / Survo ja minä

Tämän jälkeen kohdistin asetetaan (tuossa INDEX-kentässä) siihen kohtaan, mistä työ halutaan aloittaa. Sitten otetaan esiin nykyinen työvalikko (napilla F11 tai Ctrl-A) ja painetaan nappia +. Tällöin työ ilmaantuu näillä selityksillä varustettuna uutena kohteena valikon loppuun.

On tärkeää, että Survon käyttäjät osaavat sijoittaa eri työnsä sopiviin lokeroihin (levyhakemistoihin, "kansioihin"). Alkavalle käyttäjälle helpoin tapa on mennä Survon alkuvalikkoon (työhakemistosta napilla S) ja siirtyä sieltä kohtaan "Töiden yleinen hallinta" ja sieltä edelleen kohtaan "Työhakemiston perustaminen".

Survon oveluutta kuvaneeekin, että koko työvalikkotouhu on ohjelmoitu Survon omalla makrokielellä V-nimisenä sukrona. Tämä sukro käyttää hyväkseen lukuisia Survon C-kielellä ohjelmoituja peruskomentoja. Koska sukrojen ohjelmointikeinot ovat jokaisen käyttäjän ulottuvilla, esim. V-sukro on näin avoin käyttäjän omille parannoille. Maininnan ansainnee sekin, että työvali-

kot tallentuvat levyille tavallisina tekstitiedostoina. Niitä pystyy siis muokkamaan myös tekstinkäsittelyn keinoin.

## Päiväkirja

Palaan työvalikkooni ja katselen eräitä kohteita. Painamalla nappia 0 (Survon päivitys) saan esille Survon kehitystä koskevia tietoja:

```

6 6 SURVO 84C EDITOR Sun Jul 07 16:35:50 1996 C:\S\ 400 100 0
1 *SAVE S1 / Survon päivitys
2 *
3 * Tässä toimituskentässä on Survon tietoja Survon muutoksista ja
4 * laajennuksista.
5 *
6 *FIND _
7 *.....
8 *July 7, 1994
9 *STATL
10 *Usage: STATL <SURVO_data>,<output_line>
11 *works as STAT but allows frequencies > 65535.
12 *(\E1\AN !STATL.C, STAT2L.C,..., STAT6L.C STATL.BAT)
13 *
14 *July 8, 1994
15 *SHOW C:\E\D\RAMSURVO.EDT
16 *Moving vital parts of Survo to Ramdrive
17 *
18 *July 9, 1994
19 *Keys alt-Right, alt-Left, alt-Up, alt-Down (4 or 8 steps)
20 *
21 *July 11, 1994
22 *Touch mode: END key, touching fl.point numbers (123e-10)
23 *

```

Tämä on siis ohjelmatekninen päiväkirja, jota aloin pitää vuoden 1994 heinäkuussa. Sitä aikaisemmin olin pannut vastaavia asioita ylös "vahakantiseen" vihkoon. Tähän toimituskenttään yritän nyt muistaa merkitä kaikki olennaiset Survoon tekemäni laajennukset, muutokset, parannukset ja korjaukset.

Etsin tästä listasta asioita useimmiten yksinkertaisesti selaamalla sivu sivulta tai valmiiksi kirjoitetulla FIND-komennolla (rivi 6), jolle olen tullut suoraan työvalikon ohjaamana. Olen viime viikkoina työskennellyt mm. tekstitietokantojen käsittelyyn liittyvien Survon ohjelmien kehittämiseksi ja laatinut LIST COUNT -toiminnon, jolla lasketaan laajoista tekstimassoista samanaikaisesti useiden eri sanojen, sanontojen jne. lukuisuutta. Jos haluan löytää kohdan, jossa LIST COUNT mainitaan ensimmäisen kerran, jatkan suoraan FIND-riviä kirjoittamalla esim.

```
FIND LIST C_
```

ja painan aktivointinappia ESC:

```

1 1 SURVO 84C EDITOR Sun Jul 07 16:50:02 1996 C:\S\ 400 100 0
344 *LIST COUNT (new module)
345 */LCDEMO shows how to use LIST count.
346 *char *struprf(char *s) -> SM.LIB (SURVO.LIB)
347 *char *strnuprf(s,len)
348 * as strupr() but valid also for äöå...
349 *isletterf(char c) 1=upper case 2=lower case 0=not a letter
350 *
351 *July 4 - 5, 1996
352 *GENREG many times faster in large data.
353 *Instructions for logistic regression (LOGIST?)
354 *
355 *July 6, 1996
356 *Activating help texts of optional modules from the inquiry system
357 *select: !CORRMV:MULTI2? (example)
358 *
359 *July 7, 1996
360 *Macro [FOREBACK(x,y)] in PS.DEV. [REVERSE] now same as [FOREBACK(1,0)]
361 *
362 *
363 *
364 *
365 *
366 *

```

Saan tässä tapauksessa näkyviin toimituskentästä kaikkein tuoreimmat merkinnät. Nuo muistiinpanot on tarkoitettu vain itselleni eikä niiden tarkkaa sisältöä ole tässä mieltä lähteä selostamaan. Lisättyäni jotain päiväkirjaan talletan sen takaisin tiedostoonsa painamalla muutaman kerran HOME-nappia, joka palauttaa kohdistimen kentän ensimmäiselle riville. Siellä on valmis SAVE-komento, jonka aktivoin ESC-napilla ja lisäys on säilössä.

F7-näppäimellä merkitään nykyinen paikka toimituskentässä niin, että samaan kohtaan pääsee palaamaan esim. SAVE-komennon aktivoinnin jälkeen painamalla nappia F7 uudelleen. Tuon SAVE-komennessa poikkeamisen voi välttää kokonaan käyttämällä SAVENOW-sukroa, jonka käynnistys kannattaa määritellä omana näppäinyhdistelmänään.

## Muistitko tallettaa?

Huomaa, ettei Survossa poistuttaessa jostain työstä kysytä käyttäjältä: "Muistitko tallettaa?" tai muuta sellaista. Jos unohdan tallettaa kentän kirjoitettuani siihen lisätietoja, menetän nuo lisäykset. Unohtamisia sattuu joskus, mutta olenpa seuraavalla kerralla tarkempi. Miksi Survo ei siis huolehdi automaattisesti toimituskenttien talletuksesta?

Itse asiassa sen saa siitä huolehtimaan tarvittaessa (antamalla systeemi-parametrille "autosave" arvon, joka kertoo minuuteissa automaattisen talletuksen väliajan). En kuitenkaan pidä erityisemmin siitä, että ohjelmat - sinänsä hyvää tarkoittaen - liikaa holhoavat käyttäjiään. Minulle kuten monille muillekin on tärkeämpää se, että saatan keskittyä työhöni ohjelman häiritsemättä "turhilla kyselyillä".

Sama periaate koskee tietojen varmistamista Survossa yleensäkin. Käyttäjä on itse velvollinen pitämään varakopioita kaikista tärkeistä tiedostoistaan. Koke-maton käyttäjä saa tähän opastusta siirtymällä Survon alkuvalikosta kohtaan "Töiden yleinen hallinta" ja sieltä edelleen kohtaan "Työn kopiointi levyk-keelle".

Pelkät levykevarmistukset eivät riitä paljon työskenteleville. Itse talletan päi-väkohtaiset lisäykset ja muutokset kyllä levykkeelle, mutta kerran tai kahdesti viikossa kopioin kaikki muuttuneet tiedot koneeni kovalevyiltä irtokovalevyille. Tätä varten työhakemistossani on kohta C (Varakopioiden teko). Erillisiä ko-valevyjä on käytössäni kaksi ja pidän niitä eri paikoissa.

On siis totuttava ajatukseen, ettei henkilökohtaisella koneella työskenneltäessä ole mitään automaattista tapaa huolehtia tietoturvallisuudesta.

## Kirjeet ja lausunnot

Hoidan pienet, satunnaiset kirjalliset työni etupäässä työvalikkoni kohdasta 1, joka tuo näkyville toisenlaisen valikon:

| 1 1 SURVO 84C EDITOR Sun Jul 07 17:59:08 1996 |  | C:\ARK\ 250 80 0 |                |
|---|--|------------------|----------------|
| 1   | *SAVE INDEX / Kirjeet, lausunnot etc.                      |                  |                |
| 2   | *  |                  |                |
| 3   | *INDEX 5 / EDT=LOAD SORT=-TIME GROUPING=EDT ONLY=1 TYPES=0 |                  |                |
| 4   | *  |                  |                |
| 5   | *LOAD INDEX / Kirjeet, lausunnot etc.                      | 20412            | 07.07.96 17:53 |
| 6   | *LOAD P / Puhelinnrot, tietoliik.                          | 10605            | 05.07.96 13:48 |
| 7   | *LOAD JUHAP / Juha Partanen                                | 2025             | 05.07.96 13:34 |
| 8   | *LOAD SÄHKÖ / Hämeen Sähkö                                 | 4212             | 27.06.96 11:00 |
| 9   | *LOAD ADDRESS / +TEL & FAX                                 | 3266             | 18.06.96 10:39 |
| 10  | *LOAD PATCH / EDQ-paikkaus                                 | 5454             | 13.06.96 09:56 |
| 11  | *LOAD SURVO1 / Kimmon esittelyteksti                       | 2025             | 12.06.96 11:00 |
| 12  | *LOAD TIEDESCI / Tieteen suurhakemisto                     | 16362            | 07.06.96 12:05 |
| 13  | *LOAD CORRTEST / Tiedote                                   | 10302            | 04.06.96 13:18 |
| 14  | *LOAD TARTU / README.DOC                                   | 10302            | 01.06.96 14:04 |
| 15  | *LOAD SURVOE / Engl.esite (1 sivu)                         | 27472            | 31.05.96 18:38 |
| 16  | *LOAD STATUS / Computational statistics                    | 6565             | 30.05.96 14:41 |
| 17  | *LOAD SUOMTYÖ / Suom.työn liitto                           | 3645             | 29.05.96 10:47 |
| 18  | *LOAD HUNT / Teaching Statistics                           | 12322            | 07.05.96 18:53 |
| 19  | *LOAD LSADE / Leena Sadeniemi                              | 2025             | 26.04.96 08:49 |
| 20  | *LOAD EESTI4 / Tarton peruutus                             | 20604            | 24.04.96 08:40 |
| 21  | *LOAD NAEVE12 /  | 15554            | 24.04.96 08:31 |
| 22  | *LOAD EESTI3 / Tarton lyhennelmä2                          | 20705            | 24.03.96 16:16 |
| 23  | *LOAD EESTI2 / Tarton lyhennelmä                           | 20402            | 24.03.96 15:54 |

Tässä on listattuna osa viimeaikaista kirjeenvaihtoaani. Saan vanhat tekstini esiin viemällä kohdistimen ao. riville ja aktivoimalla ESC-napilla. Tässä ei ole valmista FIND-komentoa hakuja varten, mutta voin sellaisen kirjoittaa vaikka riville 4. Itse asiassa haku onnistuu myös Survon hakunapilla alt-F5. Kun painan sitä kerran, Survo-ruudun alimmalle riville ilmestyy teksti

Search for word: \_

Kirjoittamalla merkki merkiltä etsittävää sanaa Survo liikuttaa tekstiä ikkunnassaan niin, että kohdistin osoittaa joka vaiheessa ensimmäistä kohtaa, johon tuntomerkit täsmäävät. Jos ensimmäinen vaihtoehto ei kelpaa, ESC-napilla saa heti seuraavan. Päätän haun napilla ENTER.

Kirjeluettelo on jälleen aivan tavallista tekstiä ja saatan muokata sitä Survon tarjoamin keinoin. Tässä tapauksessa riittää yleensä, että päivitän listan tarvittaessa valmiilla INDEX-komennolla, joka täsmennyksineen on rivillä 3. Siis aktivoimalla INDEX-komennon saan listan tuoreimmillaan ja siten, että uusimmat asiat ovat alussa. Lista ei siis itse lisätä mitään, vaan INDEX huolehtii siitä, että hakemiston kaikki tiedostot tulevat mukaan.

INDEX kuuluu olennaisena osana *Kimmo Vehkalahden* laatimiin Survon työpöytäohjelmiin. Sillä on monenlaisia vaihtoehtoisia muotoja, joiden valinta riippuu käyttäjän mieltymyksistä. Saan tietoja sen käyttäytymisestä koska tahansa aktivoimalla pelkän komentosanan muodossa INDEX? . Se herättää Survon oman neuvontajärjestelmän, joka antaa tietoja kyselyn kohteesta väliaikaisessa ikkunassa. Tämä on Survossa yleinen tapa ottaa selville, mitä eri komennot tekevät. Aiheeseen palataan.

## Puhelimitse

Työvalikkoni kohdasta 2 (siis painamalla nappia 2) saan eteeni puhelinluettelon (näytän vain aivan alkua):

```

6 1 SURVO 84C EDITOR Sat Oct 12 10:02:26 1996 C:\ARK\ 100 100 0
1 *SAVE P / Puhelinrot, tietoliik.
2 *FIND _
3 *CD C:\VT / Tietoliikenne F2 {act}
4 *->VAXTERM
5 *CD C:\ARK
6 *
7 *PHONE Koti /
8 *PHONE Yliopisto /
9 *PHONE Elina&Teppo&Nappu /
10 *PHONE Olli&Raija /
11 *PHONE 09-1918871 / Yliopisto, tilastotieteen laitos
12 *PHONE 03-2156111 / Tampereen yliopisto
13 *PHONE 09-4521951 / ComPetit Consulting Oy

```

Kun koneeni on yhdistettynä puhelinverkkoon, vien kohdistimen sille riville, jossa mainittuun paikkaan aion soittaa ja aktivoin ao. PHONE-komennon. Näytön alariville ilmestyy vahvistukseksi soiton kohde ja kehoitus nostaa luuri. PHONE ohjelma pyörittää numerot ja jos soittooni vastataan, keskustelu voi alkaa.

Olen kohdannut ihmisiä, joiden mielestä Survon kaltaiset ohjelmistot ovat liian laajoja - eivät he tarvitse kuin pienen osan tarjolla olevista toiminnoista. Tähän olen vastannut, etten minäkään tarvitse Helsingin koko puhelinluetteloa. Yksi sivu riittäisi vallan hyvin, kun siinä olisivat ne oikeat numerot. Nyt toiveeni on toteutunut, heidän ei.

Myös PHONE on tyypillinen Survon työpöytäohjelma. Sen saa helposti toimimaan valitsemallaan näppäinyhdistelmällä jopa tekstin sekaan kirjoitetusta puhelinnumerosta. Usein tehdessäni työtä, jonka yhteydessä joudun pitämään yhteyttä toisiin henkilöihin, kirjoitan heidän numeronsa tekeillä olevan kirjoituksen alkuun. Tällöin saatan soittaa suoraan edes siirtymättä puhelinluetlooni.

PHONE sallii sen, että numeron asemasta mainitaan vain kohteen nimi. Tällöin edellytetään, että nimeä vastaava tieto on kerrottu jossain muualla samassa toimituskentässä (esim. oma numeroni yliopistolla Yliopisto=09-1918866).

On siis helppo soitella tietokoneen välityksellä, kun ei tarvitse muistaa numeroita. Aavistelen tähän vaivattomuuteen liittyvän kuitenkin tiettyjä vaaroja. Vanhastaan tutut numerot saattavat vähitellen pudota pois mielestä ja mitäs teet silloin kun pitäisi soittaa, eikä oma kone olekaan lähellä eikä puhelinluettelokaan ole saatavilla? Tämän vuoksi olen pyrkinyt säilyttämään tuntuman tärkeimpiin numeroihin ja soitan usein myös aivan vanhanaikaiseen tapaan.

Yleensäkin kannattaa pohtia, mitkä ovat sellaisia toimintoja, joita ei pidä uskoa pelkästään koneensa varaan.

## Tietoliikenne

Puhelinluetteloni alussa (rivit 3-5) on kolmen komennon yhdistelmä. Siirtämällä kohdistimen riville 3 ja antamalla ns. jatkuvan aktiivoinnin eli painamalla näppäimiä F2 ja ESC saan aktivoituksi tuon komentojonon. Ensimmäinen komento

```
CD C:\VT
```

vain siirtää Survon toiminnan kohdistumaan hakemistoon VT, johon olen sijoittanut *Hannu Kulokarin* VaxTerm-tietoliikenneohjelman. Toinen komento

```
>VAXTERM
```

käynnistää VaxTermin, joka valtaa näytön ja tästedes toimin täysin sen ohjelman ehdoilla. Saatan siis ottaa tietoliikenneyhteyksiä, lukea ja kirjoittaa säh-

köpostia sekä poimia viestejä tai muuta tietoa omaan koneeseeni tekstitiedostoiksi. Kun lopetan VaxTermin käytön, käynnistyy kolmas komento

CD C:\ARK

joka palauttaa Survon takaisin lähtöhakemistoon. Voin siis ottaa vaikka jonkin sähköpostista kaivamani viestin Survon toimituskenttään ja ryhtyä miettimään ja kirjoittamaan viestiin vastausta. Sen tehtyäni käynnistän uudelleen tuon kolmen komennon sarjan ja lähetän vastaukseni. Näin en haaska turhaa aikaa tietoliikenneyhteyksiin.

Tämä esimerkki osoittaa myös sen, että Survosta voi koska tahansa poiketa ulkopuolisiin ohjelmiin. Ulkopuoliset ohjelmat käynnistetään omilla nimillään, joiden eteen on asetettu merkki >.

### Valikosta valikkoon

Päätyövalikossani on kaksi kohdetta (9 ja A), joissa viitataan toiseen työvalikkoon (Jobfile). Painamalla nappia 9 nykyinen valikko vaihtuu uudeksi:

```

1 1 1 SURVO 84C EDITOR Sun Jul 07 09:34:08 1996 C:\KIRJA\ 100 100 0
-----
Monimuuttujamenetelmät 1996/SM C:\M\MONIM.JOB
-----
0. Jobfile: C:\E\D\SURVO.JOB / Takaisin päähakemistoon .
1. Jobfile: C:\M\MON\MON.JOB / Esipuhe, grafiikka .
2. Jobfile: C:\M\MON\M.JOB / Multinormaalijakauma .
3. Jobfile: C:\M\MEN\M.JOB / Pääkomponentti- ja faktorianalyysi .
4. Jobfile: C:\M\MEN2\M1.JOB / Kanoniset korr. ja erotteluanalyysi .
5. Jobfile: C:\M\MEN2\M2.JOB / Moniul.skaalaus, korrespondenssianal. .
6. DIM Kuutiot ja pallot C:\M\DIM\ 13 1 1 1 /
7. SING Matriisihajotelmat C:\M\SING\ 13 1 1 1 /
8. INDEX MULTI2 diskette C:\M\LEVYKE\ 1 1 1 1 /
9. EX-HIL1 HIL-esimerkki C:\M\HIL\ 1 1 1 1 /
A. EX-HIL!1 HIL-esimerkki (jatk.) C:\M\HIL!\ 1 1 1 1 /
B. INDEX Harjoitustehtäviä C:\M\H\ 9 1 1 1 /
C. INDEX Esimerkkikaaviot C:\M\EX\ 7 1 1 1 /
D. Jobfile: C:\M\TENTIT.JOB / Aikaisemmat Survo-tentit .
E. INDEX MULTI2 diskette C:\M\LEVYKE\ 1 1 1 1 /
-----
Valitse työ ao. napilla. Lisää työ napilla +. ENTER=Paluu ?=HELP
-----
S. Survon alkuvalikko M. Päätyövalikko

```

Tässä on ovi Survolle ominaiseen hypertekstimaailmaan, jossa käsittelemme tilastollisia monimuuttujamenetelmiä koskevaa kurssimateriaaliani. Valikon alussa (kohdat 1,2,3,4,5) mainitut uudet työvalikot viitoittavat tien kirjani "Tilastolliset monimuuttujamenetelmät" (1995) pääjaksoihin. Loput kohteista koskevat kurssimateriaalin liitteitä, esimerkkejä ja tenttitehtäviä sekä erillistä levykettä (MULTI2), jolle on kerätty monimuuttujamenetelmiä koskevia Survon laajennuksia.



Tästä työvalikosta lähtien teen hyvin monenlaisia asioita. Kun painan nappia 4, näkyville tulee uusi valikko, jossa kerrotaan kanonisista korrelaatioista ja erotteluanalyysistä:

```

1 1 1 SURVO 84C EDITOR Sun Jul 07 09:34:08 1996 C:\KIRJA\ 100 100 0
  Kanoniset korr. ja erotteluanalyysi C:\M\MEN\M1.JOB

0. Jobfile: C:\E\D\SURVO.JOB / Takaisin päähakemistoon .
1. Jobfile: C:\M\MONIM.JOB / Monimuuttujamenetelmien päähakemisto .
2. CA1 Kanoniset korrelaatiot C:\M\MEN2\ 12 1 1 1 /
3. CA2 Estimointi C:\M\MEN2\ 9 1 1 1 /
4. CA3 Informaatioteor.tulkinta C:\M\MEN2\ 9 1 1 1 /
5. DI1 Erotteluanalyysi C:\M\MEN2\ 6 1 1 1 /
6. DI3 Luokitteluongelma C:\M\MEN2\ 9 1 1 1 /
7. DI4 Hämmönnästä C:\M\MEN2\ 8 1 1 1 /
8. CL1 Ryhmittelyanalyysi C:\M\MEN2\ 9 1 1 1 /
9. CL2 Esimerkkejä C:\M\MEN2\ 8 1 1 1 /
A. MEN2 Painatus C:\M\MEN2\ 13 1 1 1 +
B. MEN2 Painatus (Eroitteluanal.) C:\M\MEN2\ 35 1 27 1 +

Valitse työ ao. napilla. Lisää työ napilla +. ENTER=Paluu ?=HELP

S. Survon alkuvalikko M. Päättyövalikko

```

Tulostaakseni kirjan erotteluanalyysiin liittyvät sivut valitsen nyt kohdan "Painatus (Eroitteluanal.)" napilla B ja esiin putkahtaa Survolle ominainen PRINT- eli tulostuskaavio:

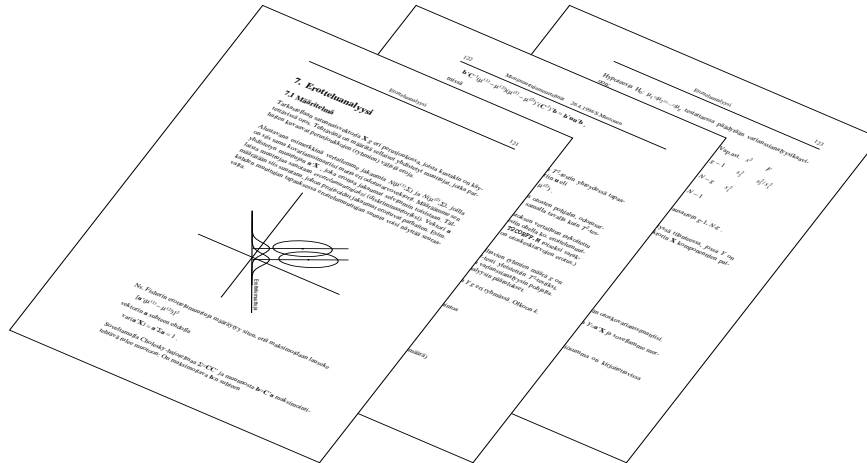
```

1 1 SURVO 84C EDITOR Mon Jul 08 16:21:04 1996 C:\KIRJA\ 150 80 0
27 - [PSAVE] [M] [Times(9)]
28 * Erotteluanalyysi ###
29 - [PRESTORE]
30 *
31 - [PSAVE] [M] [Times(9)]
32 *### Monimuuttujamenetelmät 26.4.1994/S.Mustonen
33 - [PRESTORE]
34 *
35 *PRINT CUR+1,E
36 - include PS.DEV
37 - include SMSTYLE.DV2
38 - include MATH2.DV2
39 - [page_number(121)] [M]
40 - replace
41 - header_lines 27,30,31,34
42 - [GCHAR_OFF]
43 - chapter A in DI1
44 - chapter A in DI2
45 - chapter A in DI3
46 - chapter A in DI4
47 - chapter A in DI5
48 - chapter A in DI6
49 - chapter A in DI7

```

Kohdistin asettuu suoraan PRINT-komennon kohdalle (rivi 35), jolloin aktiivisella sen napilla ESC saan tulostetuksi erotteluanalyysistä kertovan jakson kirjastani PostScript-kirjoittimella lopullisessa painoasussaan.

Tällä kohtaa en yritä kuvailla PRINT-kaavion rakennetta. Pyydän kuitenkin panemaan merkille, että matka päätyövalikostani tämän painatuksen käynnistykseen on vain neljän napin painalluksen (9,4,B ja ESC) päässä. Ainoa ulkoa muistettava painallus on ESC. Kaikki muut selviävät valikoista.



Täsmälleen painettua ulkoasua vastaavassa muodossa näytän kurssin aikana vastaavat asiat tietokoneeni ja siihen kytketyn opetusheijastimen välityksellä isokokoisina luokkahuoneen seinällä. Teen siis opetuskalvoni tosiaikaisesti. Tätä varten noudan ensin ko. tekstin toimituskenttään. Esim. erotteluanalyysin alku löytyy edellisestä työvalikosta napilla 5:

```

1 1 SURVO 84C EDITOR Mon Jul 08 20:32:36 1996 C:\M\MEN2\ 150 100 0
1 *SAVE DI1 / Erotteluanalyysi
2 *LOAD DI2
3 *TRIM3 63,P3
4 *DEF A,6,END
5 (
6 *7. Erotteluanalyysi
7 *
8 *7.1 Määritelmä
9 *Tarkastellaan satunnaisvektoria X g eri perusjoukossa, joista kustakin o
10 *tettävissä otos. Tehtävänä on määrätä sellaiset yhdistetyt muuttujat, jo
11 *haiten kuvaavat perusjoukkojen (ryhmien) välisiä eroja.
12 *
13 *Alustavana esimerkkinä vertailemme jakaumia N(m(1),S) ja N(m(2),S), joil
14 *on siis sama kovarianssimatriisi mutta eri odotusarvovektorit. Määräämme
15 *yhdistetyn muuttujan a'X, joka erottaa jakaumat selvimmän toisistaan. T
16 *laista muuttujaa sanotaan erottelumuuttujaksi (diskriminaattoriksi). Vek
17 *määrätään siis suuntana, johon projisoidut jakaumat erottuvat parhaiten.
18 *kahden muuttujan tapauksessa erottelumuuttujan suunta voisi näyttää seur
19 )valta.
20 *
21 % 800
22 - picture D2KUVA1.PS
23 - picture D2KUVA2.PS,*-150,*+558,1,1,299.70433243798

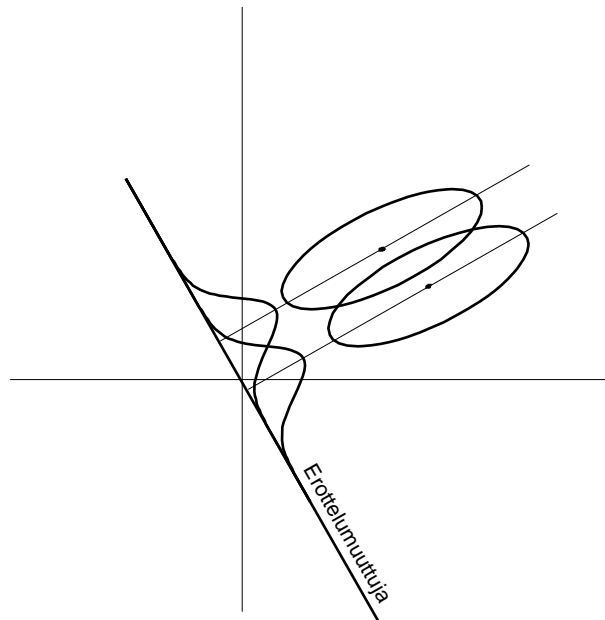
```

Et voi olla huomaamatta, että kentässä oleva teksti ei kaikin osin näytä samalta kuin sen painettu muoto. Tässä on kysymys Survolle ominaisesta "nuotinnuksesta", jossa esim. tekstilajin vaihdokset ja matemaattiset kaavat ja merkit ilmaistaan eri värein ns. *varjomerkkien* avulla. Riveillä 21-23 asetellaan valmiit kuvatiedostot paikoilleen.

Painamalla näppäinyhdistelmää F2 M M (tämä tarkoittaa M-nimisen sukron pikavalintaa) saan kuvaruudun alareunaan kehoitteen miten toimia. Osoitan sen mukaan ensin kohdistimella ensimmäistä näytettävää riviä ja painan nappia M. Sitten osoitan vastaavasti viimeistä näytettävää riviä ja painan vielä kerran nappia M. Sukro M aktivoi tällöin Survon PRINT-ohjelman siten, että valitut rivit tulostetaan PostScript-tiedostoksi. Tämän jälkeen M-sukro käynnistää ulkopuolisen Ghostscript-ohjelman, joka tulostaa PostScript-tiedoston koneen näyttöön ja sitä kautta valitut tekstit, kaavat ja kuvat myös heijastuvat seinälle.

Kun maalaan M-napilla vaikkapa rivit 13-23, kuvaruutuun ja opetusheijastimen kautta seinälle ilmestyy:

Alustavana esimerkkinä vertailemme jakaumia  $N(\mu^{(1)}, \Sigma)$  ja  $N(\mu^{(2)}, \Sigma)$ , joilla on siis sama kovarianssimatriisi mutta eri odotusarvovektorit. Määäämmen yhdistetyn muuttujan  $\mathbf{a}'\mathbf{X}$ , joka erottaa jakaumat selvimmin toisistaan. Tällaista muuttujaa sanotaan *erottelumuuttujaksi* (diskriminaattoriksi). Vektori  $\mathbf{a}$  määrätään siis suuntana, johon projisoidut jakaumat erottuvat parhaiten. Esim. kahden muuttujan tapauksessa erottelumuuttujan suunta voisi näyttää seuraavalta.



Ohjailemalla kohdistinta ja painelemalla samaa M-näppäintä toistuvasti saan ketterästi uusia otteita kirjastani kaikkien näkyville.

Kirjaan sisältyviä esimerkkejä näytän lisäksi sellaisinaan toimituskentistä. Toistan esimerkkilaskelmia ja -analyyseja sekä alkuperäisessä muodossa että kurssilla käydyin keskustelun perusteella muutettuina; improvisoinnille ja "Entäs jos?"-kysymyksille on aina tilaa.

## Tekeillä olevat työt

Palautan mieleen tämänhetkisen päätyövalikkoni. Saan sen aina esiin siirtymällä tarvittaessa ensin nykyiseen työvalikkoon (F11 tai Ctrl-A) ja painamalla sitten nappia M:

```

1 1 1 SURVO 84C EDITOR Tue Jul 09 09:15:20 1996 C:\KIRJA\ 100 100 0

```

| Työvalikko  |  | C:\E\D\SURVO.JOB |           |
|-------------|--|------------------|-----------|
| 0. S1       | Survon päivitys                                      | C:\S\            | 6 6 1 1   |
| 1. INDEX    | Kirjeet, lausunnot etc.                              | C:\ARK\          | 5 1 1 1   |
| 2. P        | Puhelinnrot, tietoliik.                              | C:\ARK\          | 2 6 1 1   |
| 3. INDEX    | Arkistoituja levykkeitä                              | C:\Z\            | 4 1 1 1   |
| 4. OPETUS   | Kurssit ja seminaarit                                | C:\ARK\          | 3 1 1 1   |
| 5. ASENNUK  | SURVO 84C-levykkeet                                  | C:\PR\           | 1 1 1 1   |
| 6. TILAUS   | Tilauslomakkeet                                      | C:\SS\TILAUS\    | 7 1 1 1   |
| 7. TILIT7   | Tilinpito 1995-96                                    | C:\SS\           | 1 1 1 1   |
| 8. SURVO5   | Etiketit   | C:\ETIK\         | 24 1 24 1 |
| 9. Jobfile: | C:\M\MONIM.JOB / Monimuuttujamenetelmät, hyperteksti |                  | .         |
| A. Jobfile: | C:\E\D\SM.JOB / Tekeillä olevat työt                 |                  | .         |
| B. Sucro:   | ESITTELY 12.10.1995                                  | C:\E\TUT\        | +         |
| C. KOPIO    | Varakopioiden teko                                   | C:\S\            | 4 1 1 1   |
| D. RAMSUR2  | Survo to Ramdrive                                    | C:\E\D\          | 4 1 1 1   |
| E. INDEX    | Missing observations                                 | C:\MISS\         | 4 1 1 1   |
| F. INDEX    | Survo ja minä  | C:\KIRJA\        | 10 1 1 1  |
| G. INDEX    | Windows-kokeet                                       | C:\WINKOE\       | 4 1 1 1   |
| H. INDEX    | GENREG-parannus                                      | C:\GENREG\       | 4 1 1 1   |

Valitse työ ao. napilla. Lisää työ napilla +. ENTER=Paluu ?=HELP  
S. Survon alkuvalikko M. Päätyövalikko

Kaikkein ajankohtaisimpia työn alla olevia asioita (mm. tutkimusaiheita, kirjoitustöitä ja Survon kehittämishankkeita) pidän tilapäisesti tässä päävalikossa. Nyt sellaisia ovat kohdat E,F,G ja H. Näistä F (Survo ja minä) sekä G (Windows-kokeet) koskevat juuri tätä kirjoitustani.

Muut tekeillä olevat tai toimenpiteitä vaativat kohteet olen sijoittanut omaan työvalikkoonsa "Tekeillä olevat työt", joka nousee esille napilla A:

```

1 1 1 SURVO 84C EDITOR Tue Jul 09 10:01:59 1996 C:\KIRJA\ 100 100 0
-----
Tekeillä olevat työt C:\E\D\SM.JOB
-----
0. Jobfile: C:\E\D\SURVO.JOB / Takaisin päähakemistoon .
1. Jobfile: C:\E\D\LIST.JOB / LIST-operaatiot .
2. Jobfile: C:\E\D\VANHAT.JOB / Aikaisempia töitä .
3. INDEX MULTVAR operation C:\MVAR\ 2 1 1 1 /
4. INDEX Mvar paper C:\MVARTXT\ 2 1 1 1 /
5. INDEX Survo ja minä C:\KIRJA\ 10 1 1 1 /
6. INDEX External MAT operations C:\MAT#\ 10 1 1 1 /
7. INDEX Nuoret 1995 C:\TUOMO\ 4 1 1 1 /
8. INDEX Vakiotermi C:\SIMO\ 4 1 1 1 /
9. COUNT LIST COUNT C:\LIST\ 6 1 1 1 /
A. CORRTEST C:\CORRTEST\ 7 1 1 1 /
-----
Valitse työ ao. napilla. Lisää työ napilla +. ENTER=Paluu ?=HELP
-----
S. Survon alkuvalikko M. Päättyövalikko

```

Täällä on osittain samoja aiheita, mutta myös sellaisia, jotka odottavat toimenpiteitäni. Laajahkosta LIST-operaatioiden ohjelmointityöstä on jopa oma valikkonsa, joka tulee näkyville napilla 1.

Kohteet 3 ja 4 (MULTVAR operation ja Mvar paper) edustavat viimeaikaista tutkimustyötäni tilastollisten monimuuttujamenetelmien alueella. Vaikka kerroin tästä aiheesta myöhemmin, on paikallaan todeta, että oma työni Survon äärellä ei ole ainoastaan Survon itsensä kehittämistä ja ohjelmointia. Survo on itselleni myös laboratorio, jossa kokeilen ja testaan tilastotieteellisiä tutkimusideoita. Tilastotieteen alueella tulee vastaan ongelmia, joita on hyvin vaikea ellei jopa mahdotonta selvittää pelkästään teoreettisin keinoin. Niissä tarvitaan laskennallista tukea mm. työläiden simulointikokeiden muodossa. Mikä parasta, tällaisten kokeiden tuloksista voi joskus pilkistää esiin löydöksiä, jotka vievät tutkimusta myös teoriatasolla eteenpäin.

Nämä simulointiin liittyvät vaativat laskennalliset tarpeet ovat osaltaan vaikuttaneet siihen, miten olen Survon kehitystä suunnannut. Olen tähdännyt siihen, että Survo muodostaisi tukevan alustan mainitunlaisten kokeiden suunnittelulle ja toteutukselle. Survoon on tullut tätä kautta käytännön sovelluksiin tarkoitettuja menetelmiä, jotka perustuvat aineistokohtaiseen simulointiin. Tällaiset menetelmät ovat "robustimpia" kuin klassiset, koska tutkimusaineiston tilastolliseen käyttäytymiseen riittävät väljemmät oletukset.

Tässä vaiheessa en kuitenkaan aio pitempään esitellä "vaatekaappini" sisältöä. Joihinkin aiheisiin palaan siis vielä myöhemmissä luvuissa.

## Survo nopeimmillaan

- ☐ Päättyövalikossa on kohta "Survo to Ramdrive", jota vastaava toimituskenttä löytyy napilla D:

```
1 1 SURVO 84C EDITOR Tue Jul 09 10:08:05 1996 C:\KIRJA\ 100 80 0
1 *SAVE RAMSUR2 / Survo to Ramdrive
2 *
3 *
4 *->COPY C:\E D:\
5 *->MD D:\SYS
6 *->COPY C:\E\SYS D:\SYS
7 *->MD D:\D
8 *->MD D:\FI
9 *->COPY C:\E\FI D:\FI
10 *->MD D:\TAB
11 *->COPY C:\E\tab D:\TAB
12 *->MD D:\P
13 *->COPY C:\E\P\P.EXE D:\P
14 *->MD D:\G
15 *->COPY C:\E\G D:\G
16 *->MD D:\F
17 *->COPY C:\E\F D:\F
18 *->MD D:\TBL
19 *->COPY C:\E\TBL D:\TBL
20 *->MD D:\TUT
21 *->COPY C:\E\TUT D:\TUT
22 *->MD D:\M
23 *->COPY C:\E\M D:\M
```

Kenttä sisältää jonon käyttöjärjestelmän komentoja, joiden tarkoitus on kopioida Survo miltei kokonaan D-levylle. Tämä on koneessani ns. "Ramdrive" eli osa koneen nopeasta keskusmuistista on varattu levytilan kaltaiseksi alueeksi. Kun käynnistän koneeni, voin valita mm. Windows 95:n tai pelkän DOS:in. Jälkimmäisessä tapauksessa käytössäni on ruhtinaalliset lähes 20 megatavua vapaata keskusmuistia, josta suurin osa on silloin nimetty D-levyksi. Koneeni keskusmuisti käsittää kaiken kaikkiaan 24 megatavua. Tämä sattuu olemaan tarkalleen 1000-kertainen tila verrattuna siihen Wang 2200-pientietokoneeseen, jolla aloitin vuorovaikutteisen Survon kehittelyn vuonna 1975.

Kun aloitan jatkuvan aktivoinnin (napeilla F2 ja ESC) riviltä 4, Survon keskeiset ohjelmamodulit kopioituvat alle puolessa minuutissa tuolle nopealle levylle.

Tämän menettelyn varsinainen hyöty ja tarkoitus on luoda erityisen suurta nopeutta edellyttäville tehtäville edulliset puitteet. Kun Survon nyt käynnistää D-levyltä, se toimii mm. vaativissa tilastollisissa simulointikokeissa jopa noin kaksinkertaisella nopeudella. Tuollaisissa kokeissa monet Survon ohjelma-modulit tulevat vaihtelevasti käyttöön toiminnan aikana. Silloin levyn nopeudella on todella merkitystä.

Menettely säästää myös koneen kovalevyä. Kun laajoissakin simulointitutkimuksissa kaikki pyörii pelkästään koneen keskusmuistissa ja ajo saattaa kestää vaikkapa yli yön, levymuistin ja näytön voi sammuttaa. Ajon päättymisen jälkeen tulokset siirretään sitten parempaan talteen.

Tämä esimerkki osoittaa myös, kuinka välittömästi käyttöjärjestelmän toiminnot ovat saatavilla suoraan Survosta.

## Survo vs. Windows

- Tiedonhallintaa aidosti vuorovaikutteisessa ja eloisassa ympäristössä on vaikea saada kuvatuksi paperilla. Tuntuu siltä, että monet tärkeät piirteet latistuvat tai vaikuttavat todellista mutkikkaammilta. Nämä asiat pitäisi siis nähdä ja kokea luonnossa.

Tästä huolimatta toivon, että sellainenkin lukija, joka ei ole nykyiseen Survoon tutustunut, on edellä kuvattujen esimerkkien kautta saanut jonkinlaisen otteen Survon olemuksesta.

Survo viihtyy mainiosti esim. Windows 95 -ympäristössä. On joissain tilanteissa varmasti etua siitä, että samanaikaisesti pystyn käyttämään useita Survoja samanaikaisesti, vaikka liiallisuuksia tulee kaihtaa. Joka tapauksessa tuntuu houkuttavalta, että voin vaikka piirtää kuvia yhdellä Survolla ja laskeskel-la ja kirjoittaa tekstiä toisella - jopa samanaikaisesti.

Tavantakaa kohtaan kysymyksen: "Milloin Survosta tulee aito Windows-versio?". Tämänhetkinen vastaukseni kuuluu: "Tuskin koskaan!". Nykyisen DOS-version ehdottomana valttina on, että se toimii kaikissa järjestelmissä, joissa on ns. DOS-boksi, eli paitsi eri Windowseissa myös esim. OS/2:ssa, Linuxissa, jopa Macissa. Tätä eivät tee näiden käyttöjärjestelmien omat ohjelmat keskenään. DOS on sitä paitsi mainettaan parempi. Vaikka sen tiettyjä puutteellisuuksia ja rajoituksia en halua kieltää, niitä voi *kiertää!* Olen aina työskennellyt varsin ahtaissa rajoissa ja opetellut rakentamaan "purjelaivaa pulloon". Tästä kerron myöhemmin kuvatessani Survon ohjelmointitekniikkaa.

Uskon jo tähän mennessä tehneeni selväksi, että Survon editoriaalista käyttötappaa ja töiden hallintaa ei voi suoraan rinnastaa vanhanaikaisiin komentokielisiin. Vapaus järjestellä tiedot ja niitä ohjaavat komentokaaviot sekä valikot mielensä mukaan erottaa Survon edukseen tavanomaisista komentokielistä, joissa käskyjen esittäminen tietystä järjestyksessä oli ehdotonta. Kaikkien Survoa paremmin tuntevien yhteinen vakaumus on, että tämä omaleimaisuus väistämättä menetettäisiin, jos Survosta yritettäisiin tehdä esim. "aito" Windows-versio. Survo ja Windows ovat rakenteittensa puolesta kuin tuli ja vesi;

niitä toisiinsa liittämällä yhdistyisivät pikemmin kummankin huonot kuin hyvät puolet.

Niin usein olen kohdannut ATK-asiantuntijoita, jotka tuntuvat tietävän Survosta enemmän kuin minä, vaikka eivät ole itse käyttäneet Survoa. He eivät ole edes yrittäneet päivittää tietojaan eikä käsityksiään, vaan perustavat väitteensä esim. "Survon vaikeakäyttöisyydestä" vanhentuneisiin kuulopuheisiin.

Haluamatta syyllystyä vastaavaan virheeseen, olen seurannut mitä alalla tapahtuu ja erityisesti viime aikoina toiminut myös aidoilla Windows 95-ohjelmilla (MS-Office-paketti ja SPSS-tilasto-ohjelma). Sen perusteella, mitä olen Survoon kohdistuneitten ennakkoluulojen muodossa saanut kokea, lukija ymmärtänee, että jossain suhteessa saatan tulkita Windows-ohjelmia kuin "piru raamattua". Puhun kuitenkin oman, elävän kokemuksen pohjalta.

Olen jo edellä esittänyt kysymyksiä tyyliin "Miten tekisit tämän mieli-ohjelmallasi?" ja tulen esittämään vastaavia jatkossakin. Ajateltakoon nyt pelkästään töiden hallintaa, mikä luonnostaan on kaiken muun perusta.

Tyypillisissä graafisissa käyttöliittymissä aikaa tuhlaantuu kaikenlaisiin oheistoimintoihin kuten jatkuvaan työpöydän järjestelyyn - ikkunoiden avaamiseen, siirtelyyn, venyttelyyn ja sulkemiseen - varsinaisen tekemisen kustannuksella. Vastaavasti työn aikana ja siitä poistuttaessa joudun vahvistamaan moneen kertaan jo tekemäni valinnat. Ergonomisesti siirtyminen edestakaisin hiireltä näppäimistöle ei ole erityisen miellyttävää - "*Salmonella typhimurium*" uhkaa. Windows 95:ssä on myös harmillista, että kulkiessani hiiren avulla valikosta valikkoon yksi ainoa lipsahdus palauttaa minut armotta lähtöruutuun. Ymmärrän, että lopulliset valinnat on saatettavissa oman kuvakkeensa alle. Aktiiviselle käyttäjälle niitä kuitenkin kertyy tuhattomasti, jolloin oikean kuvakkeen etsiminen ja hahmottaminen "ikonostaasista" vie taas aikaa. Miksi Windows 95:n käynnistykseen kuluu koneellani (90 Mhz Pentium, 24 MB) noin 40 sekuntia ja kovalevy narisee sen aikana niin kuin sitä tahallaan simputtaisiin?

Toisaalta Windows 95:n tarjoamat näkymät kuvaruudulla ovat lumoavia verrattuna Survon toimituskenttien tai työvalikkojen arkiseen olemukseen. Tästä kateellisena en kuitenkaan voi olla ajattelematta: "Moni kakku päältä kaunis, mutta ... "

Vielä en malta olla puuttumatta myyttiin Survon vaikeudesta. Koska Survossa pelataan osittain erilaisilla käsitteillä kuin muissa käyttöympäristöissä, tämä aiheuttaa ymmärtämättömyyttä. Survossa ei puhuta "oliopohjaisesta" työskentelystä eikä "OLE2-piirteistä", koska niitä ei sananmukaisesti siellä kerta kaikkiaan ole! Kuitenkin käytännössä hyvin pitkälle vastaavat toimintatavat ovat jo yli kymmenen vuoden ajan kuuluneet survoilijoiden arkeen.



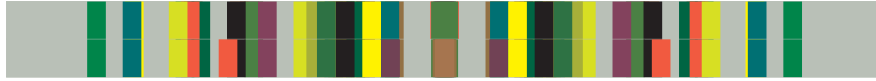
Jokainen aidosti erilainen työskentelypiiri synnyttää sille ominaisen pienois-kulttuurin, jonka oppiminen vaatii aikansa. Kukaan ei parin viikon turistimatkan jälkeen voi väittää omaksuneensa kiinalaista kulttuuria ja sen ajattelutapoja. Vaikkei samantasoisista asioista olekaan kysymys, myös eri ATK-ympäristöt vaativat enemmän kuin muutaman tunnin esittelyn tai parin päivän kurssin, ennenkuin paljastavat edes osan totuuttaan.

Kun pohditaan käytön helppoutta, se ei tarkoita, niin kuin usein uskotellaan, että käytön tulisi aina olla helppoa. Enin mihin päästään on, että helpot asiat pysyvät helppoina, mutta käsitteellisesti vaikeat jäävät edelleen vaikeiksi. Vain tiedon määrän ja hallinnan kannalta tai laskennallisesti raskaita ongelmia ATK voi tuntuvasti helpottaa. Siitä, että helpot asiat tehdään ATK:n myötä vaikeiksi, on surullisen paljon esimerkkejä.

Einstein sanoi aikoinaan: "Things must be done as simple as possible but no simpler" eli "Asiat on tehtävä niin yksinkertaisiksi kuin mahdollista mutta ei yhtään yksinkertaisemmiksi." En tiedä, mikä antoi aiheen toteamukseen, mutta arvata saattaa, että hän oli lopullisesti kyllästynyt lehtimiesten uteluun: "Eikö suhteellisuusteoriaa voisi selittää yksinkertaisemmin?"

Anttonio Pytinki oli Uuden Suomen palstoilla 1940-60-luvuilla vaikuttaneen mestaripakinoitsija Ollin sankareita. Pytinki näyttäytyi keksijänerona, jonka saavutuksiin kuului mm. "Einsteinin suhteellisuusteoria lapsille". Avoimeksi jää, oliko Olli tietoinen Einsteinin lausumasta.

Silti "suhteellisuusteorioita lapsille" keksitään jatkuvasti, mikäli koneiden ja ohjelmien mainontaan on uskominen.



## Alusta pitäen

- Edellisissä luvuissa olen tarkastellut Survon ominaisuuksia enimmältään kokeneen käyttäjän silmin. Kun Survo on vasta asennettu koneeseen ja sitä on tarkoitus käyttää ensimmäisen kerran, tilanne näyttää toiselta; "vaatekaappi" on tyhjä.

En nytkään ota tehtäväkseni toimia kaiken selittävänä oppaana aloittelevalle. Haluan kuitenkin kertoa, miten Survoon pääsee käsiksi ja miten sen kanssa mielestäni parhaiten oppii tulemaan toimeen.

Kaikessa vähänkin vaativammassa tietojenkäsittelyssä vallitsee raaka "luonnonlaki", jonka mukaan hyvät päätökset ja valinnat luovat edellytykset hyvälle ratkaisuille jatkossakin, kun taas tyhmät päätökset johtavat uusiin tyhmiin valintoihin. Käyttäjäraukalle, joka kerran on eksynyt "oikealta" tieltä, ongelmat vain kasautuvat, mikä on niin väärin, niin väärin...

On vaikea keksiä yleisiä sääntöjä sille, miten tietojärjestelmien kanssa menestyään. Periaate "pieni on kaunista" on kuitenkin hyvä pitää mielessä. Aivan liian usein ohjelmat antavat työkohteiden, esim. tekstidokumenttien, kasvaa mittoihin, joissa niiden hallitseminen koituu hitaaksi ja kankeaksi.

Survossa kasvun rajat tulevat äkkiä vastaan, mikäli suurta tilastoaineistoa yritetään itsepintaisesti pitää toimituskentässä eikä ymmärretä siirtää sitä ajoissa havaintotiedostoksi, jolloin rajoitukset poistuvat.

Monet istuvat sitkeästi kiinni totutuissa työtavoissaan eivätkä hievhadakaan, vaikka heille selittäisi, että löytyyhän niitä suurempiakin keinoja. Alituinen kiire on hyvä syy olla muuttamatta mitään, kun asiat jotenkuten sujuvat nykyiselläänkin.

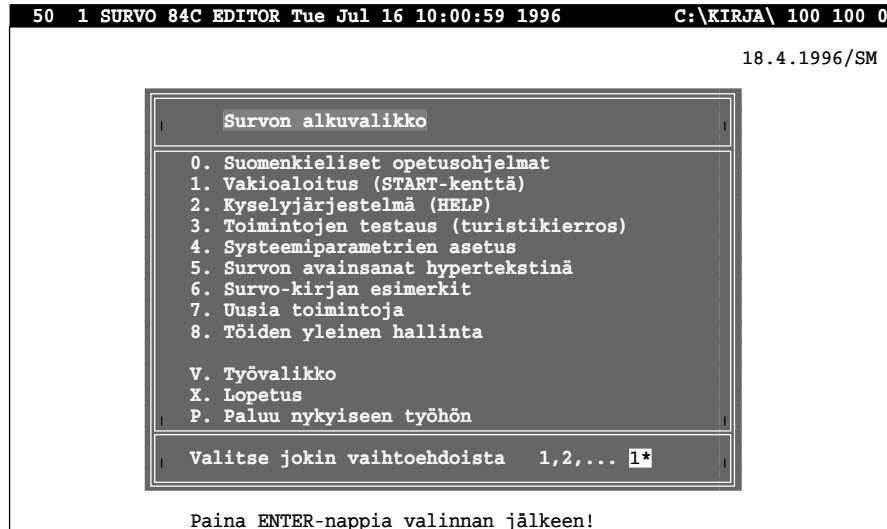
Vaikka on uskaliaista yleistää, tuntuu siltä, että Survon omaksuminen on helppointa niille, joilla ei ole pinttynyttä vakaumusta. Kun ei tiedä "paremmasta", Survon oppiminen on otollista. Johonkin toiseen järjestelmään hyvin perehty-

nyt saattaa aluksi nähdä Survonkin lähinnä vain tuon toisen järjestelmän ehdoilla; hänen on vaikea irrottautua tottumuksistaan ja hyväksyä Survon pelisääntöjä.

## Alkuvalikosta opetteluun

On monia tapoja Survon käynnistämiseen. Ensimmäisissä versioissa aloitus johti pelkkään tyhjään toimituskenttään. Siitä piti sitten omin toimin lähteä eteenpäin kirjoittamalla ja aktivoimalla sopivia komentoja. Tähän tuskin kukaan survoilija enää tyytyy. Hän luo itselleen sopivan lähtöasetelman esim. samantapaisen työvalikon muodossa, jollaista esittelin edellisessä luvussa tai ainakin hänellä on alkutilassa sopivilla ohjeilla ja komennoilla täytetty toimituskenttä.

Kenenkään ei tarvitse aloittaa tyhjästä. Kun Survo käynnistetään pelkällä nimellään, eteen ilmaantuu alkuvalikko:



Alkuvalikosta pääsee tutustumaan Survoon eri tavoin. Aivan aluksi kannattaa valita kohta **3. Toimintojen testaus**, koska se käy läpi useita keskeisiä toimintoja ja näin tutkii, että esim. juuri koneeseen asennettu Survo toimii siinä moitteettomasti. Samalla se pyörittää turistikierroksen, jolla saa pikaisen katsauksen Survon peruskäyttöön. Tämä näytematka kuten monet muutkin alkuvalikosta lähtevät toiminnat perustuu Survon sukrotekniikkaan. Se on siis kooste Survon makrokielellä tehdyistä ohjelmista. Kierroksen aikana, kuten aina sukron ollessa vauhdissa, eräänlainen aavekäyttäjä toimii koneella ja näyttää tarvittaessa - jopa yksittäisiä näppäinpainalluksia myöten - mitä ja miten tehdään. Käyttö on sikäli aivan aitoa eikä lavastettua, että voit koska tahansa

katkaista tämältyyppisen Survon opetusohjelman (napilla .) ja jatkaa saavutetusta tilanteesta omin avuin ja omaan suuntaasi. Jos näin teet, pääset kahdella painalluksella takaisin työvalikon kautta alkuvalikkoon.

Turistikierros kestää nykyisillä koneilla kymmenisen minuuttia ellei toimintaa kiihdytä matkan aikana (napilla +) tai hidasta (napilla -). Kun haluaa nähdä opetusohjelman aikana yksittäiset näppäinpainallukset, on painettava HELP-nappia (F1), jolloin ruudun oikeaan alakulmaan ilmaantuu aina seuraavaksi painettavan napin nimi. Jos et paina mitään nappia, Survo etenee itse seuraavaan painallukseen 10 sekunnin välein. Jos painat väärää nappia, kuulet merkkiään ja joudut yrittämään uudelleen. Niille, jotka näin askeltaen haluavat seurata tapahtumia, mutta eivät vaivaudu painelemaan "oikeita" nappeja, kannattaa vihjaista, että nuolinappi "vasemmalle" on tässä joka paikan näppäin. Painamalla siis nuolinappia toistuvasti, käytössäsi on "yhden napin editori", jolla etenet sutjakkaasti. Tästä vaiheittaisesta läpikäynnistä takaisin automaattiseen palataan yksinkertaisesti HELP-napin uudella painalluksella.

Huomattavasti mittavampi tutustumistilaisuus on edessä, kun valitsee kohdan **0. Suomenkieliset opetusohjelmat**. Se jakautuu asteittaisesti aiheen mukaan lukuisiin alakohtiin, joita vastaavat omat valikkonsa. Painamalla pelkkää ENTER-nappia - aina kun pyydetään - saat läpikäydyksi kaikki aiheet noin 10 tunnissa. Tämän saatat tietenkin tehdä hyvin pienissäkin jaksoissa. Nämä opetusohjelmat eivät laajuudestaan huolimatta kata läheskään kaikkea, mitä Survosta löytyy. Olen tehnyt tämän sukkokokoelman viisi vuotta sitten (v.1991) Survon etupäässä koulukäyttöön tarkoitettua, supistettua SURVOS-versiota varten. Tämä on vaikuttanut kokoelman luonteeseen, sillä olen halunnut sen avulla tukea paitsi Survon myös mm. koulumatematiikan opetusta ja antaa näin malleja opettajille.

Omasta mielestäni eräs hauskimista suomenkieliseen opetussarjaan kuuluvista jaksoista kertoo matemaatikosta nimeltä Cole. Tämä esimerkki löytyy, kun alkuvalikosta siis valitaan kohta 0 ja sen jälkeen

7. Matemaattiset toiminnat

ja sen jälkeen

2. Editoriaalinen laskenta

ja vielä sen jälkeen

A. Mielenkiintoinen esitelmä

eli aiheeseen päästään kolmen alivalikon kautta.

☐ Kertomus alkaa kuvaruudulla seuraavasti:

```

1 1 SURVO 84C EDITOR Tue Jul 16 16:30:54 1996 C:\E\OPETUS\ 100 100 0
1 *
2 *   Mielenkiintoinen esitelmä
3 *
4 *   Amerikan matemaattinen yhdistys järjesti New Yorkissa lokakuussa 1903
5 *kokouksen, jonka ohjelmassa mm. F.N. Cole oli luvannut pitää esitelmän
6 *otsikolla "Suurten lukujen tekijöihin jakamisesta". Kun puheenjohtaja
7 *pyysi häntä aloittamaan esityksensä, hyvin vähäpuheiseksi tunnettu Mr.
8 *Cole asteli taululle ja - sanomatta mitään - ryhtyi huolellisesti
9 *kirjoittamaan liidulla seuraavaa laskutoimitusta:
10 *
11 *   Huom! Mr. Cole tekee tässä useita laskutoimituksia suurilla luvuilla.
12 *       Olemme laatineet niiden jäljittelyä varten omat sukronsa,
13 *       joista kerrotaan erikseen, jos painat nappia F1:HELP.
14 * .....
15 * _
16 *
17 *
18 *
19 *
20 *
21 *
22 *
23 *
      Jatka painamalla ENTER (tai F1:HELP, jos haluat opastusta)

```

Painan nyt nappia ENTER, jolloin saamme nähdä, mitä Mr. Cole puuhaa:

```

15 1 SURVO 84C EDITOR Tue Jul 16 16:30:54 1996 C:\E\OPETUS\ 100 100 0
11 *
12 *           256      =2^8
13 *           256 x
14 *           1536
15 *           1280
16 *           512
17 *           65536   =2^16
18 *           65536 x
19 *           393216
20 *           196608
21 *           327680
22 *           327680
23 *           393216
24 *           4294967296 =2^32
25 *           4294967296 x
26 *           25769803776
27 *           38654705664
28 *           8589934592
29 *           30064771072
30 *           25769803776
31 *           38654705664
32 *
      Mr. Cole laskemassa taululla...

```

Mahdotonta on aivan tarkalleen tietää, miltä laskut todella näyttivät, mutta kakkosen korkeisiin potensseihin niillä kurkoteltiin ...

```

40 1 SURVO 84C EDITOR Tue Jul 16 16:30:54 1996 C:\E\OPETUS\ 100 100 0
32 *      17179869184
33 *      38654705664
34 *      8589934592
35 *      17179869184
36 *      18446744073709551616 =2^64
37 *              8 x =2^3
38 *      147573952589676412928 =2^67
39 *              1 -
40 *      147573952589676412927 =2^67 - 1 = M67
41 * .....
42 *      Kuulijat hämmästyivät hieman, sillä Mr. Cole oli laskenut ns.
43 *Mersennen luvun M67, jota Marin Mersenne jo vuonna 1644 oli väittänyt
44 *alkuluvuksi. Tämän seikan yhdistyksen jäsenet tiesivät hyvin.
45 *      Sanomatta edelleenkään mitään, Mr. Cole siirtyi taulun puhtaan
46 *osan eteen ja jatkoi...
47 * .....
48 *
49 *      761838257287 =A
50 *      193707721 x =B
51 *      761838257287
52 *      1523676514574
53 *      5332867801009
54 *      5332867801009
55 *      5332867801009
56 *      2285514771861
57 *      6856544315583
58 *      761838257287
59 *      147573952589676412927 = A x B = M67
60 * .....
61 *      Tulos täsmäsi aikaisempaan. M67 ei siis ollutkaan alkuluku!
62 *      Ensimmäisen ja toistaiseksi ainoan kerran yhdistyksen historian
63 *aikana yleisö osoitti innostuneesti esitykselle suosiotaan.
64 *      Mr. Cole palasi hiljaa paikalleen. Kukaan ei kysynyt mitään.
65 *
66 *      E.T.Bell, joka kertoo tapauksen kirjassaan Mathematics: Queen and
67 *Servant of Science (London 1952), tiedusteli esitelmän pitäjältä
68 *myöhemmin, kauanko niiden tekijöiden löytäminen kesti.
69 *Cole vastasi: "Three years of Sundays"._
70 *

```

Siitä huolimatta, että esityksen elävyys kuihtuu paperilla, toivon lukijan aisti-  
van, kuinka luontevasti tällainen kerronta sopii Survon tyyliin. Mitä tulee tari-  
nan varsinaiseen opetukseen, on ihailtavaa, mitä ihmiset pystyivät tekemään ja  
jaksoivat tehdä ilman nykyisiä laskentamahdollisuuksia. Herra Cole lienee ol-  
lut todellinen "sunnuntaimatemaatikko" ja utterasti puurtanut ongelmansa  
parissa. Esimerkki ei kuitenkaan edusta matemaattista tutkimusta tyypillisim-  
millään.

Saadakseni tarinan toimimaan Survossa tahdoin laskennan näkyvän taululla  
aidosti sellaisena kuin kertolaskuja tehdään käsin laskien. Tämän vuoksi laa-  
din joukon sukroja, jotka pystyvät peruslaskutoimituksiin suurilla kokonais-  
luvuilla näkyvässä muodossa. Siis, joka kerran kun esimerkki käydään läpi  
Survossa, myös laskut toistetaan aidosti. "Kynä" liikkuu ja kirjoittaa sukro-  
ohjelman "kursorikoreografiaa" noudattaen.

Tämäntapainen kädestä pitäen opettaminen ei suinkaan ole ominaista kaikille  
opetussarjan jaksoille. Tehtävän luonne määrää, kuinka tarkkaan yksityiskohdat  
ilmaistaan.

## Neuvontaa

- Sekä alkavalle että myös kokeneelle survoilijalle laajin tuki löytyy Survoon sisältyvästä kyselyjärjestelmästä. Neuvontaan pääsee suoraan alkuvalikosta (kohta 2). Useimmiten apua haetaan, kun ollaan jo työn keskellä jossain toimituskentässä. Tällöin viimeksi käynnistetystä tai yritetystä komennosta saa tietoja pelkällä HELP-napin (F1) painalluksella. Jos taas haluaa kysellä itse kirjoittamallaan hakusanalla, kirjoitetaan ko. hakusana kysymysmerkillä varustettuna tyhjän rivin alkuun (esim. PRINT?) ja kohdistimen ollessa tällä rivillä painetaan aktivointinäppäintä ESC. Tällöin kyselysanana alapuolelle avautuu tilapäinen opastusikkuna esim. seuraavaan tapaan:

```

7 1 SURVO 84C EDITOR Wed Jul 17 16:45:40 1996 C:\KIRJA\ 300 100 0
17 *Tällöin kyselysanana alapuolelle avautuu tilapäinen opastusikkuna
18 *esim. seuraavaan tapaan:
19 *
20 *PRINT?_
PRINT L1,L2
prints the lines L1-L2 of the current edit field on the printer.
Line numbers and the control column are not printed.
Lines are usually printed as they appear in the edit field and various
display modes (shadow characters) correspond to certain printing
modes of the printer.
Lines with a '-' in the control are not printed, but hold various
control information for printing. For example, a line of the form
    11 - chapter A in FILE3
specifies a chapter A from the edit file FILE3 to be printed.
Control lines of various type enable printing of multipage
reports with line numbers and headers. On the PostScript and Canon
LBP-8 printers also pictures made with the PLOT operation can be
automatically included.

    1 = PostScript printers
    2 = Canon LBP-8 A2 and II
    3 = Control information in the PRINT operation
Select 1-3 or press ENTER!

```

Kyselysanalla PRINT kerrotaan Survon tulostuskomennosta. Tästä ikkunasta mm. selviää, että esim. toimituskentän rivit 11-50 voi tulostaa kirjoittimella aktiivomalla komennon

```
PRINT 11,50
```

Kyselyikkunassa kerrotaan loppuriveillä, mistä saa lisävinkkejä. Tässäkin toimii "yhden kosketuksen peli" eli esimerkissämme napilla 1 kerrotaan PostScript-tulostuksesta ja napilla 3 saa tarkempia tietoja erilaisista tulostuksen ohjauskeinoista, kuten kirjasinlajien, rivivälin, palstan leveyden ja sivun pituuden asetuksista.

Survon kyselyjärjestelmässä voi seikkailla loppumattomiin ja näin pikaisesti tutustua järjestelmän ominaisuuksiin. On myös mahdollista peruuttaa edelli-

seen ikkunaan (BACKSPACE-napilla jopa useaan kertaan) ja poimia tekstiä ikkunoista toimituskenttään (napilla +). Neuvonnan aikana palataan alkuun painamalla HELP-nappia F1. Sieltä avautuu parilla askeleella Survon päätoimialojen valikko:

```

7 1 SURVO 84C EDITOR Wed Jul 17 16:46:12 1996 C:\KIRJA\ 300 100 0
17 *Tällöin kyselysanon alapuolelle avautuu tilapäinen opastusikkuna
18 *esim. seuraavaan tapaan:
19 *
20 *PRINT?
SURVO 84C EDITOR is the basic tool for all operations in SURVO 84C.
A = General information on SURVO 84C
B = Edit field
C = Keys and their functions
D = SURVO 84C operations
E = Control operations
F = Text processing
G = Operations for tables
H = File management (SURVO 84C data files)
I = Statistical data analysis
J = Mathematical operations
K = Plotting
L = Numerical conversions
N = New functions (since writing of the Survo book 1992)
Select A-N or press ENTER!

```

Luettelosta lähtien ja välillä siihen palaten voi lukea neuvontatietoja kuin käyttöopasta.

Kyselystä poistutaan painamalla ENTER-nappia. Tällöin yleensä viimeksi katsellusta aiheesta (esim. PRINT-komennosta) ilmestyy kyselyrivin täydennykseksi malli, joka osoittaa ko. komennon yleisen rakenteen. Mallin mukaan käyttäjän on helppo kirjoittaa oma komentonsa.

Kirjoitettuihin oppaisiin verrattuna kyselyjärjestelmän ehdoton etu on siinä, että se on aina ajan tasalla. Kyselytekstit ovat kaikki Survon toimituskentissä ja niitä päivitetään heti, kun itse ohjelmistosta on jotain uutta kerrottavaa. Neuvontapalvelu kuuluu olennaisena osana Survon jokapäiväiseen käyttöön. En itsekään edes yritä muistaa harvinaisempien toimintojen täsmällisiä muotoja, vaikka olen niistä valtaosan aikoinaan sommitellut. Kyselyjärjestelmästä saan aina nopeasti poimittua esiin tarvitsemani tiedot.

Kyselyn saattaa aloittaa myös rivin keskellä olevasta sanasta - tässä vaikkapa tuo PRINT - ja kohdistimen koskettaessa ko. sanaa painetaan nappeja F2 ja F1 peräkkäin. Periaatteessa mikä tahansa sana on näin aktivoitavissa. Siten toimituskentän jokainen piste on "kuuma". Samaten luettaessa tietoa kyselyikkunoissa kohdistimella voi osoittaa niissäkin kaikkia sanoja ja kysyä jopa pelkällä ESC-napilla. Eri asia on, tietääkö kyselyjärjestelmä sanasta mitään. Jos se ei tiedä, näkyviin tulee lähisukuisten sanojen luettelo, josta voi valita mieleisensä.



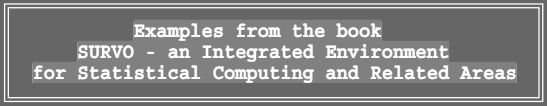
Kyselyjärjestelmää on mukava ylläpitää Survon omilla keinoilla. Saattaapa sen rinnalle tehdä eri aihealueita koskevia, uusia neuvontakokonaisuuksia. Kimmo Vehkalahti on rakentanut vuonna 1989 kirjoittamani tekstin pohjalta sellaisen, joka avustaa Survon C-kielistä ohjelmointia. *Armo Heikkilä* on vuorostaan suomentanut huomattavan osan kyselyteksteistä ja järjestänyt ne uudella tavalla suomenkieliseksi neuvontajärjestelmäksi. Kyselyjärjestelmästä toiseen siirrytään mm. QPATH-komennolla, josta saa tietoja vastaavalla kyselyllä.

## Matkimista

Kaikki edellä mainitut oppimista avustavat keinot tarjoavat valmiita esimerkkejä. Näitä mukailten on usein helpompi rakentaa Survossa omia ratkaisuja kuin lähtemällä puhtaalta pöydältä. Vaikkei aina heti ymmärtäisikään kaikkia mallina käytetyn esimerkin yksityiskohtia, matkiminen kannattaa. Suosittelen lämpimästi kokeilunomaista lähestymistä, jossa muutellaan mallia vähitellen ja jopa arvausten kautta haluttuun suuntaan. Kun itse lähden tekemään uutta työtä Survolla, ensimmäinen ajatus on muistella, olenko tehnyt aiemmin jotain vastaavaa. Jos tällainen pohjaksi kelpaava työ löytyy, alustan uuden työn tämän aikaisemman työn toimituskentillä ja muilla tiedoilla. Tästä muuttamalla ja kehittelemällä pääsen yleensä nopeasti liikkeelle.

Kirjoittamani Survon englanninkielinen käyttöopas (v.1992) sisältää myös runsaasti jopa sangen vaativia sovellustilanteita. Miltei jokainen näistä esimerkeistä on toistettavissa Survolla, sillä alkuvalikossa on kohta **6. Survo-kirjan esimerkit**. Kun valitsen tämän, saan eteeni toimituskentässä olevan luettelon:

```

1 1 SURVO 84C EDITOR Wed Jul 17 20:31:33 1996 C:\E\EX\ 300 80 0
1 /SURVO-RETURN / Return to the main menu
2
3
4 
5 S. Mustonen
6 22.12.1992
7
8
9 /EX-HELP / General information
10
11 2. An Overview of Survo
12 /EX-P8 / Population of Finland by age group (pie chart)
13 -/EX-P9 / Compound interest (curves)
14 */EX-P11 / Inverse matrix of AA'
15
16 3. Text processing
17 /EX-P24 / INSERT command
18 /EX-P26 / CHANGE command (for lines)
19 /EX-P27 / CLEAR command
20
21 4. Table management
22 /EX-P36 / FORM command
23 /EX-P37 / TAB key

```

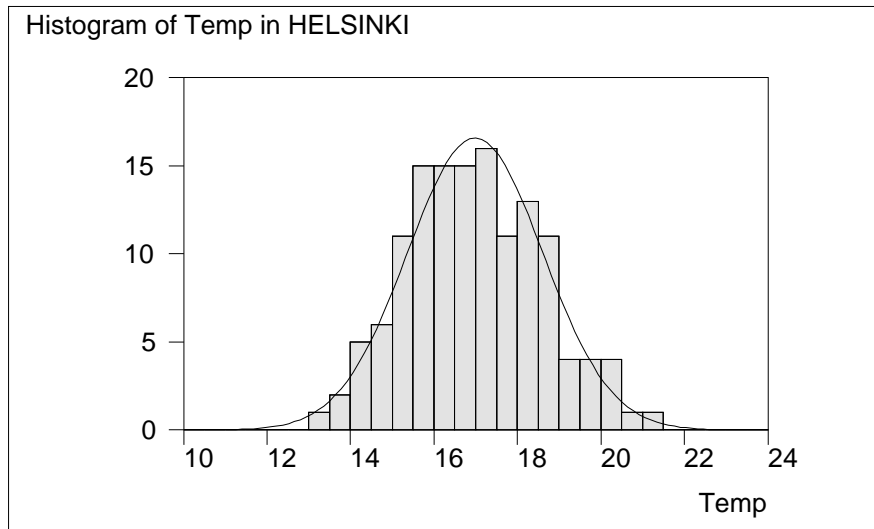
Luettelo käsittää lähes 300 esimerkkiä, jotka esiintyvät Survo-kirjassa. Kun selailen kirjaa ja haluan ottaa mallikseni jonkin kirjan esimerkeistä (vaikkapa Helsingin lämpötilaa kuvaavan histogrammin) katson vain kirjan sivunumeron (se on 151) ja etsin tätä vastaavan kohdan luettelosta esim. hakunapilla alt-F5.

```

1 1 SURVO 84C EDITOR Wed Jul 17 20:31:33 1996 C:\E\EX\ 300 80 0
104 */EX-P150 / TAB operation, Data BINORM
105 /EX-P151 / HISTO operation, Data HELSINKI
106 */EX-P154 / HISTO operation, Binomial distribution
107 /EX-P157 / HISTO operation, Beta distribution
108 */EX-P160 / HISTO operation, Mixture of normal distributions
109 /EX-P163 / COMPARE operation, Farm and town boys
110 /EX-P165B / COMPARE operation, Sucro /COMPARE
111 /EX-P166 / COMPARE operation, Pairwise comparison
112 /EX-P166B / COMPARE operation, Rank correlations
113 */EX-P167 / COMPARE operation, Testing for normality
114 /EX-P169 / LINREG operation
115 /EX-P170 / LINREG operation, Stepwise regression
116 */EX-P172 / REGDIAG operation
117 */EX-P176 / SMOOTH operation for semiparametric smoothing (Polyreg)
118 */EX-P177 / SMOOTH operation for semiparametric smoothing (Finland)
119 */EX-P178 / ESTIMATE operation, Simple example
120 */EX-P181 / DER operation
121 */EX-P183 / ESTIMATE operation, Initial estimates
122 */EX-P184 / ESTIMATE operation, Constants in the model
123 */EX-P186 / ESTIMATE operation, Weighting of observations
124 */EX-P188 / ESTIMATE operation, Estimation criteria
125 */EX-P189 / ESTIMATE operation, Residuals and predicted values
126 */EX-P190 / ESTIMATE operation, Maximum likelihood estimates

```

Aktivoimalla esimerkkiä koskevan rivin (tässä 105) pääsen valmiin sukron ohjaimena seuraamaan, kuinka ko. histogrammi piirretään Survon avulla.



```

23 1 SURVO 84C EDITOR Sun Apr 12 08:14:52 1992 D:\P2\STAT\ 100 100 0
12 *
13 *HISTO HELSINKI,Temp,17
14 *Temp=10(0.5)24 FIT=NORMAL
15 *XSCALE=10(2)24 YSCALE=0(5)20 SIZE=1160,700 DEVICE=PS
16 *
17 *Frequency distribution of Temp in HELSINKI: N=135
18 *
19 *Class midpoint f % Sum % e e f X^2
20 * <=13.00 0 0.0 0 0.0 1.0
21 * 13.25 1 0.7 1 0.7 1.2
22 * 13.75 2 1.5 3 2.2 2.3
23 * 14.25 5 3.7 8 5.9 4.0 8.5 8 0.0
24 * 14.75 6 4.4 14 10.4 6.5 6.5 6 0.0
25 * 15.25 11 8.1 25 18.5 9.4 9.4 11 0.3
26 * 15.75 15 11.1 40 29.6 12.4 12.4 15 0.6
27 * 16.25 15 11.1 55 40.7 14.9 14.9 15 0.0
28 * 16.75 15 11.1 70 51.9 16.3 16.3 15 0.1
29 * 17.25 16 11.9 86 63.7 16.3 16.3 16 0.0
30 * 17.75 11 8.1 97 71.9 14.8 14.8 11 1.0
31 * 18.25 13 9.6 110 81.5 12.2 12.2 13 0.0
32 * 18.75 11 8.1 121 89.6 9.2 9.2 11 0.3
33 * 19.25 4 3.0 125 92.6 6.3 6.3 4 0.8
34 * 19.75 4 3.0 129 95.6 3.9
35 * 20.25 4 3.0 133 98.5 2.2
36 * 20.75 1 0.7 134 99.3 1.2
37 * 21.25 1 0.7 135 100.0 0.5
38 * >21.50 0 0.0 135 100.0 0.4 8.3 10 0.4
39 *Mean=16.98704 Std.dev.=1.623701
40 *Fitted by NORMAL(16.987,2.6364) distribution
41 *Chi-square=3.596 df=9 P=0.9359
42 *

```

Edellä oleva kuva on piirretty Survon HISTO-operaatiolla, joka näkyy tarvittavine täydennyksineen yllä toimituskentässä riveillä 13-15. HISTO:n antamat numeeriset tulokset ovat riveillä 17-41, jotka tässä havainnollisuuden vuoksi esitetään harmaalla pohjalla.

Lauri Tarkkonen on sanonut Survon olevan "julkaisuohjelma, jolla voi tehdä faktorianalyysia". Jos koneessa on Survo ja siihen on kytketty PostScript-kirjoitin, jokainen voi välittömästi todentaa tämän käytännössä valitsemalla em. luettelosta esimerkit P337 ja P221. Edellinen tulostaa painovalmiina Survo-kirjan luvun 9 "Printing of reports" ja jälkimmäinen tekee faktorianalyysia automaattisesti sukron /FACTOR avulla.

Vaikka minkä tahansa Survo-kirjan esimerkin saattaa toistaa näin ilman vaihua, on eri asia ymmärtää, mitä oikein tapahtuu. Minulle on kerrottu, että eräässä Helsingin yliopistossa parisen vuotta sitten pidetyssä tiedotustilaisuudessa, jossa Survokin oli esillä, muuan lääketiedettä edustava tutkija oli todennut: "Survossa on vain se vika, että pitää ymmärtää, mitä on tekemässä!" Huomautus lienee luettava hyväntahtoisen leikinlaskun tiliin. Silti raaka on se totuus, että "Einsteinin suhteellisuusteorioita" ei ole lapsille tarjona. Katsokaamme siis, millaista ymmärrystä edellä kuvattu histogrammin piirto oikein edellyttää.

- Tarkastelun kohteena on ollut Helsingin heinäkuista keskilämpötilaa kuvaava aineisto vuosilta 1844-1978 eli siis 135 havainnon sarja. Tämä aineisto on tal-

letettuna Survon havaintotiedostoksi HELSINKI ja Temp on lämpötilaa kuvaava muuttuja tuossa tiedostossa. Histogrammi on luokitetun, yhden muuttujan graafinen esitys, jossa tapausten luokittaista määrää kuvaavat erikokoiset pylväät. Jotta Survon HISTO-komento pystyy kuvan piirtämään, se tarvitsee parametreikseen käsiteltävän aineiston (HELSINKI) ja siitä piirrettävän muuttujan (Temp). Tämä näkyy edellä esitettyssä piirroskaaviossa rivillä 13. Lisäksi siinä viitataan ensimmäiseen tulosriviin (17), mutta tämä ei ole pakollista silloin, kun ei välitetä mistään numeerisista tuloksista. Tutkittavan muuttujan Temp luokitustapa ja -tarkkuus kuvataan täsmennyksellä Temp=10 (0.5) 24, mikä sanoo, että lämpötila-arvot luokitellaan puolen asteen välein arvosta 10 arvoon 24.

Siinä oli kaikki, mikä on pakko ilmoittaa, jotta HISTO voisi hoitaa tehtävänsä, ja tuon tiedon saan esiin heti Survon neuvonnasta aktivoimalla kyselyn HISTO?. Jos ei tiedä edes kyselysanaa, aina voi lähteä liikkeelle pelkällä HELP- napilla (F1), valita "Statistical data analysis" (Tilastolliset menetelmät) ja sieltä edelleen "Frequency distributions, histograms".

Erilaisilla lisätäsmennyksillä parannetaan kuvan ulkonäköä ja laajennetaan numeerista tulostusta. Tässä tapauksessa täsmennys FIT=Normal sanoo, että halutaan tutkia samalla, miten hyvin luokitettu aineisto noudattaa normaalijakaumaa. Tulos näkyy pylväskuvan päälle piirrettynä Gaussin kellokäyränä, joka on mahdollisimman hyvin sovitettu aineistoa myötäileväksi. Yhteensopivuutta mitataan  $\chi^2$ -testin avulla. Sen antama tulos on rivillä 41. Erityisesti kannattaa kiinnittää huomiota testin  $P$ -arvoon 0.9359, joka kertoo, että mikäli aineisto on todella otos normaalijakaumasta, vain  $100 \cdot (1 - 0.9359) = 6.41$  tapauksessa sadasta yhteensopivuus olisi parempi kuin nyt havaittu. Tämä osoittaa, että yhteensopivuus on tilastollisesti katsoen hyvä; ei ole mitään syytä epäillä, etteikö Helsingin lämpötilan heinäkuinen keskiarvo noudattaisi normaalijakaumaa.

Rivillä 15 on kerrottu yksityiskohtaisesti, millaisin asteikoin (XSCALE, YSCALE), missä koossa (SIZE) ja millä laitteella (DEVICE) kuva piirretään. Tässä kooksi on valittu 116x70 mm (huom. yksikkönä 0.1 mm eli riittävän tarkka kaikissa tilanteissa) ja tulostuslaitteeksi PostScript-kirjoitin (PS).

- Tässä Survon sovelluksessa tulee "ymmärtää", mitä tarkoittavat käsitteet histogrammi, normaalijakauma ja  $\chi^2$ -yhteensopivuustesti. Ilman näitä tietoja ei koko tehtävää ole mielekäästä yrittääkään, vaikka matkimalla saisikin sen mekaanisesti suoritetuksi. On valitettavaa, että tilastollisia menetelmiä aika usein sovelletaan "vailla täyttä ymmärrystä". Tähän on osittain syynä "helppokäyttöiset" ohjelmat, joissa valikko-ohjatusti voi mukavaan tyyliin syöttää mitä tahansa tietoa mille tahansa tilastolliselle toiminnalle. Siis: "garbage in - garbage out" eli "roskaa sisään - roskaa ulos".

Tyhmyksien tekeminen onnistuu vallan mainiosti Survollakin, koska mitään järjestelmää ei voi jalostaa niin pitkälle, että se sallisi vain järkevät toimenpiteet. Vielä viitisen vuotta sitten puhuttiin paljon ns. "asiantuntijajärjestelmistä" myös tilastotieteen tutkijoiden piirissä. Kaunis ajatus oli, että "viisas" ohjelma siinä opastaa "tyhmää" käyttäjää, jonka ei tarvitse tietää kovin paljon tilastollisista menetelmistä. Tällaisten ratkaisujen kehittäminen on onnistunut kuitenkin vain hyvin rajatuissa tehtävissä. Epäilin jo silloin ja epäilen yhä tuota lähestymistapaa. Mielestäni on tässä vaiheessa parempi tyytyä luomaan ratkaisuja, jotka palvelevat käyttäjää monipuolisesti ja joissa vallitsee tuo luonnollinen suhde: "Helpot asiat tehdään helposti, vaikeat vaativat enemmän tietoa."

Yleensäkin vaatimus, että järjestelmä kieltäytyisi toteuttamasta mitään järjestömiä tehtäviä, kuten olen joskus kuullut esitettävän, on kerta kaikkiaan mahdoton. Jos sellaiseen pyrittäisiin, se rajoittaisi samalla liikaa uutta luovia ongelmanratkaisutapoja. Useinhan edistystä on se, että osataan yhdistellä vanhoja menettelyjä uudella tavalla. Tällaiseen Survo on omiaan. Toisaalta lempein keinoin tulisi usuttaa järkeviin valintoihin.

Kun puheet "asiantuntijajärjestelmistä" olivat kuumimmillaan, ryhdyin tarkoituksellisesti puhumaan Survosta "asiantuntijan järjestelmänä" eli englanninkielisen termin *Expert System* asemasta käytin luonnehdintaa "*System for the experts*". Tämän tein senkin uhalla, että se lisäisi vaikutelmaa Survosta **vain** asiantuntijoiden käyttöön tarkoitettuna käyttöympäristönä. Todellisuudessa Survon käyttäjät vaihtelevat koulutuksensa tason ja määrän suhteen tavattomasti. Se vaikuttaa myös heidän käyttötottumuksiinsa. Luultavimmin suurin osa survoilijoista harrastaa esim. tilastollista laskentaa vain nimeksi. Heillä monesti tekstinkäsittely, kuvien piirtäminen, taulukkolaskenta ja julkaisujen tuottaminen ovat keskeisiä toimintoja. Juuri tällaisissa tehtävissä käyttäjät ovat kehittyneet erittäin taitaviksi, vaikka peruskoulutus olisi vaatimatonta. On siis täysin vastuutonta sanoa, kuten olen joskus kuullut väitettävän, että Survo olisi muka "vain ohjelmoijan työväline".

Niillä, jotka alunperin lähtevät Survoa soveltamaan johonkin ennalta määrättyyn, rajoitettuun ongelmaan, syntyy usein kiinnostus katsella myös ympärille. Ennenpitkää he huomaavat, että voivat tehdä paljon toisenlaisiakin asioita. Survo venyy näin käyttäjänsä vaatimusten yltyessä yhä haastavampiin tehtäviin.

Viihtyvyyttä Survon seurassa lisää sekin, että työ dokumentoi itsensä. Työhön kuuluvat toimituskentät kannattaa tallettaa tarvittaessa pienin vihjein varustettuina. Tällöin vuosienkin päästä voi palata takaisin ja toistaa työn nykyhetkeen päivitettyä ilman, että joutuisi uudelleen opettelemaan kaikki yksityiskohdat.



## Nappulat pelikentällä

- Toimituskenttä on Survossa kaikkien "olioiden" kohtausta paikka. Siihen kirjoitetaan tekstiä, taulukoita, komentoja ja komentokaavioita. Siihen ilmaantuvat myös luettavaksi tarkoitetut tulokset. Kaikki asiat kietoutuvat yhteen samaan tilaan päinvastoin kuin ikkunoiduissa järjestelmissä, joissa kullekin tietotyypille on varattuna lokeronsa.

Survoilijan näkemyksistä ja tehtävien luonteesta riippuu, miltä toimituskenttä näyttää. Ehkä ainoa yhteinen piirre on talletuskomennon SAVE sijoittaminen ensimmäiselle riville, mutta sekään ei ole pakollista. Usein kenttä ainakin ulkopuolisen silmin voi vaikuttaa sekalaisten merkintöjen ja muistiinpanojen kokoelmalta. Jokaiselle kehittyä kokemuksen myötä oma persoonallinen esitystyyliä, mikä auttaa välittömästi hahmottamaan kentän sisältöä. Itse nautin tuosta vapaudesta asetella tietoja mieleni ja tavoitteitteni mukaan.

Erityyppisten tietojen kohtaaminen toimituskentässä antaa Survolle verrattoman edun. Kaikkeen pääsee suoraan käsiksi eikä tarvitse katsella voimattomana "ikkunalasin" takaa, mikä on usein tunnusomaista mm. Windows-työskentelylle.

Pelkässä tekstinkäsittelyssä Survo ei kuitenkaan juuri poikkea tavanomaisista tekstureista. Survossa on pyritty säilyttämään tuntuma siihen, mitä ennen tehtiin kirjoituskoneilla. Esim. rivinsiirtonäppäin (ENTER) siirtää aina kohdistimen seuraavan rivin alkuun ja siinä kaikki. Esim. Word-teksturissa ENTER-näppäin painallus rivin keskeltä siirtää loppurivin uuden rivin aluksi. Se on mielestäni hämmentävää, koska näin poiketaan kirjoituskoneen toiminnasta. Pelkässä tekstinkäsittelyssä on tietenkin mahdollista olla välittämättä siitä, miten sanat riveille alunperin asettuvat. Survon toimittimella on kuitenkin tekstureihin verrattuna siinä määrin lisävelvoitteita, että liika automatiikka tekstin hallinnassa olisi pahasta. Toimituskenttä on yleisempi pelitila, jossa nappuloiden liikutteluun tarvitaan osittain toisia sääntöjä.

Kuten jo huomautin aikaisemmin, on pidettävä huolta siitä, että merkit ovat ryhdikkäästi myös pystysuorassa suunnassa, jolloin ei tarvita mitään erillistä tilaa taulukkojen muotoiluun eikä käsittelyyn. Tavallisissa tekstoreissa taulukkojen kanssa on paljon hankalampi tulla toimeen.

## Näppituntumalla

- Monissa, vaativaan käyttöön tarkoitetuissa editoreissa (esim. emacs) toimenpiteet hoidetaan yleensä erilaisin näppäinyhdistelmin, joita voi olla satamäärin. Survossa näppäinyhdistelmiä on vain kymmenittäin, sillä varsinkin vaativia tekstinmuokkauksia varten on omat komentonsa. Komennotkin voi, jos niin tahtoo, piilottaa valitsemansa näppäinyhdistelmän taakse.

Aloitteleva survoilija tulee toimeen alle 20 erikoisnäppäimellä, joista esitän oman suosikkilistani. Tärkeintä on tietää, että Survon kaikki näppäintoiminnot saa esille kyselyllä KEYS? ja erityisesti funktionapeista tulee tietoa kyselyllä F?.

Kun laskuista jätetään pois tavalliset kirjoituskonenäppäimet, siis kirjaimet, numerot, välimerkit ja rivinsiirto (ENTER), listani 15 kärjessä on seuraava:

|                    |  |
|--------------------|--|
| 1. F1: HELP        | Kyselyjen aloitus  |
| 2. F8: EXIT        | Poistuminen Survosta   |
| 3. ESC             | Komentojen aktivointi  |
| 4. Nuolinapit      | Siirtyminen askeltaen paikasta toiseen                         |
| 5. Ins tai F9      | Siirtyminen päällekirjoitustilasta lisäystilaan ja päinvastoin |
| 6. Del tai F10     | Merkin poisto  |
| 7. Alt-F9          | Uuden tyhjän rivin lisäys                                      |
| 8. Alt-F10         | Nykyisen rivin poisto  |
| 9. Ctrl-END        | Nykyisen rivin tyhjennys                                       |
| 10. END            | Kohdistin rivin loppuun  |
| 11. HOME           | Paluu asteittain rivin alkuun - kentän alkuun                  |
| 12. Alt-F5         | Sanan haku kentästä merkki merkiltä                            |
| 13. Alt-F3         | Rivin kopiointi  |
| 14. F3: TOUCH      | Siirtyminen kosketuslaskentaan                                 |
| 15. Ctrl-A tai F11 | Työvalikko esiin   |

Survosta poistuminen (F8) on siksi kakkostilalla, että itseäni on harmittanut monien muiden (DOS-)ohjelmien tapa olla piittaamatta siitä, miten niistä pääsee pois. Ikäänkuin haluttaisiin itsepintaisesti käyttäjän "viihtyvän" ohjelman parissa loputtomiin. Survossa jo alkuvalikko sisältää kohdan **X. Lopetus**.

F2 on Survossa yleinen etunappi (PREFIX), joka usein muuttaa tilapäisesti seuraavan napin tehtävää. Esim. F2 ENTER ei viekään kohdistinta rivin alkuun vaan määrää kohdistimen paluusarakkeen seuraavissa ENTER-painalluksissa. Tästä on hyötyä, kun kirjoitetaan jotain esim. taulukon keskellä olevaan sarakkeeseen.

Koska tämä ei ole varsinainen käyttöopas, turhaa on jatkaa tarkemmilla luetteiloilla ja kuvauksilla. Jotkin näppäinyhdistelmät ansainnevat silti maininnan.

Kun tekstin tai taulukkojen osia kopioidaan paikasta toiseen, se tapahtuu kätevimmin napilla alt-F4 (BLOCK). Painettaessa ensimmäisen kerran Survoikkunan alariville ilmestyvät ohjeet, miten menetellä. Itse asiassa koko homma tehdään kohdistinta liikuttamalla (nuolinapeilla) ja painamalla toistuvasti samaa BLOCK-nappia kopioitavan lohkon vastakkaisissa kulmissa ja lopulta uuden paikan kohdalla. Kopioitava lohko tallentuu menettelyn aikana myös väliaikaiseen tiedostoon, joka toimii samaan tapaan kuin leikepöytä Windowsissa. Esim. viimeksi määritellyn lohkon saa kopioitua myöhemmin vaikka kuinka monta kertaa uusiin paikkoihin painamalla BLOCK-nappia surutta neljästi! Jos moniajoympäristössä on useita Survoja vedetty yhtä aikaa esille, näin voi kopioida tietoja hyvin nopeasti myös Survosta toiseen. Lauseen osia siirrellään vastaavasti napilla alt-F2 (WORDS).

Survon ehdottomia valtteja tekstinkäsittelyssä on näppäinyhdistelmä F2 J (JATKA). Se on "kaikkien makrojen äiti", sillä jos juuri nyt kirjoittaisin

```
F2 J on siis "k_
```

ja painan nappeja F2 ja J, Survon toimitin osaa jatkaa lausetta kirjoittamalla

```
F2 J on siis "kaikkien
```

Samalla ikkunan alle ilmestyy ohjeita. Kun yksinkertaisesti painan välilyöntinäppäintä toistuvasti, saan pidennettyä tekstin muotoon

```
F2 J on siis "kaikkien makrojen äiti",_
```

ENTER-nappi lopettaa toimittimelta tämän jatkamisvietin ja saatan itse jälleen kirjoittaa eteenpäin. Kaikki tapahtuu sillä yksinkertaisella periaatteella, että kun keskeneräisen sanan perään painetaan F2 J, Survo etsii nykyisestä toimituskentästä ensimmäisen kohdan, johon sanan alku täsmää ja jatkaa sitten heti matkimalla tuota mallia. Tässä tapauksessa olin kirjoittanut mainitun tekstin jo kerran ja niin Survo apinoi tuota sanontaa.

"Jatkonappi" toimii myös silloin, kun malli on vasta nykyisen kohdan alapuolella näkymättömissä. Tästä on Survon opetussarjassa esimerkki, johon päästään valintojen

4. Tekstinkäsittelyn alkeita

ja

9. Tietojen haku toimituskentästä

kautta:



```

6 1 SURVO 84C EDITOR Mon Jul 22 15:02:47 1996 C:\E\OPETUS\ 600 100 0
17 *
18 *Kirjoittaessamme tekstiä, jossa toistuvasti esiintyy pitkiä nimiä ja
19 *termejä, voimme säästää työtä näppäinyhdistelmän PREFIX J avulla.
20 * Voimme poimia tällä tekniikalla jopa kokonaisia rivejä:
21 *
22 * Tampere HÄM 165453 523.70 15.00
23 *
24 C Kunta Lää Väestö Ala Äyri
25 A Alahärmä VAA 5358 348.23 17.00
26 * Alalampi VAA 8658 715.95 18.00
27 * Alastaro TUR 3686 260.92 15.00
28 * Alavieska OUL 3036 250.44 17.00
29 * Alavus VAA 10359 813.96 17.50
30 * Anjalankoski KYM 20307 721.60 16.00
31 * Anttola MIK 1792 256.16 16.50
32 * Artlampi UUS 2039 180.70 16.50
33 * Asikkala HÄM 8249 566.60 15.50
34 * Askainen TUR 825 58.30 16.50
35 * Askola UUS 3930 209.40 16.00
36 * Aura TUR 2445 95.05 15.00
37 * Brändö AHV 551 99.67 14.00
38 * Dragsfjärd TUR 4670 243.20 15.50
39 * Eckerö AHV 705 107.74 14.50
Next word by the space bar. Cancel by DELETE. Accept by ENTER!

```

Toimituskenttään on poimittu (FILE LOAD-komennolla tiedostosta KUN-NAT) tietoja Suomen kunnista. Kunnat ovat luetteloituna aakkosjärjestyksessä. Saadakseni esiin Tampereen tiedot, voin tietenkin etsiskellä niitä hakunapilla alt-F5. Jatkonappi F2 J antaa toisen mahdollisuuden. Olen kirjoittanut (yllä riville 22) sanan alun

Tamp\_

ja painanut nappeja F2 ja J. Tällöin ensivaiheessa nimi täydentyy oikein ja siten painamalla välilyöntinäppäintä toistuvasti saan koko Tampere-rivin kopioiduksi riville 22. Siis ei ole lainkaan tarvetta kelata toimituskenttää eteenpäin tiedon löytämiseksi.

Kun aikoinaan 1980-luvun alussa sain kuulla huomautuksia tyyliin: "Survo pystyy vai alkeelliseen tekstinkäsittelyyn!", tämä kaiveli mieltäni sen verran, että laadin joitakin toimintoja, joita tuskin vieläkään löytyy muista tekstoreista. Äskeinen jatkonappi on yksi tällainen kehittyneempi keino. Toinen on ns. *matomoodi* tai *koirankävelytytys*, joka avautuu kosketuslaskennassa (nappi F3) napin alt-F2 (WORDS) takaa. Jälleen ikkunan alalaitaan ilmestyvät tarpeelliset ohjeet. Tarkoituksena on ensin määritellä nuolinapeilla maalaamalla mielivaltainen jono toimituskenttään kirjoitettuja merkkejä, sitten "kuljettaa" tätä merkkijonoa ympäri kenttää ja lopulta kopioida se halutuissa "asennoissa" ja "kiemuroissa" eri paikkoihin. Tästä en yritäkään antaa paperilla esimerkkiä, koska ajatus väistämättä haalistuu. Kannattaa kokeilla luonnossa.

## Komentelua

Varsinainen tekstin ja taulukoiden muokkaus tapahtuu komennoilla, joista tärkeimpiä ovat:

|                |   |
|----------------|---|
| TRIM           | Tekstin tahtaminen  |
| FORM           | Taulukon muotoilu   |
| SORT           | Taulukon lajittelu  |
| REPLACE        | Sanojen ja merkkijonojen korvaaminen toisilla   |
| SHOW           | Toisten toimituskenttien ja tekstitiedostojen katselu sekä tietojen poiminta nykyiseen kenttään |
| COPY jne.      | Tietojen kopiointi  |
| INSERT, DELETE | Taulukkojen ja tekstinosien siirto sivuttaissuunnassa   |
| CLEAR, SCRATCH | Tietojen poispyyhkiminen  |
| SAVE           | Toimituskentän talletus tiedostoon  |
| LOAD           | Toimituskentän lukeminen tiedostosta  |
| SAVEP, LOADP   | Tekstitiedostojen talletus ja lukeminen   |
| PRINT          | Tulostus kirjoittimella tai tiedostoon  |

Kuten aikaisempi lajitteluesimerkki luvussa "Yleisperiaatteita" osoitti, komennot kirjoitetaan vapaasti tekstin sekaan, kuitenkin omille riveilleen. Tästä koituu suuri etu, sillä komennot kaikkine lisineen ovat pelkkää tekstiä tekstin joukossa. Komennolla on "suuri hetkensä" vasta silloin, kun se aktivoidaan ESC-napilla kohdistimen osoittaessa komentoriviä. Komennon jälkeen koko komentoteksti palaa takaisin muiden kuolevaisten joukkoon.

Joillekin näyttää olevan vastenmielistä, että "pakonapilla" ESC on tuo tärkeä käynnistysrooli; miksi siihen ei käytetä ENTER-nappia? Tähän ei tarvitse enää vastata, mikäli lukija on ymmärtänyt kantani ENTER-napin tehtävästä. ESC-napista ei tule napista! On valitettavaa, että PC:n vakionäppäimistöstä puuttuu varsinainen suoritusnäppäin (EXECUTE), jota käytin aktivointiin Survon aikaisemmissa versioissa Wang-pientietokoneella ja Wang-mikrolla. Kun katselee nykyistä näppäimistöä ja ajattelee Survolle ominaista käyttötapaa, on vaikea keksiä aktivointiin ESC-nappia parempaa vaihtoehtoa. Se erotuu muista hyvin eikä sillä Survossa olisi muuten mitään tekemistä.

Käyttäjä päättää itse, mitkä komennot hän jättää käytön jälkeen paikoilleen. Jos komentoja jää esim. tulostettavan tekstin keskelle, ne on luonnollisesti sieltä poistettava. Monissa töissä on mukava kirjoittaa toimituskentän alkuun tai käsiteltävien taulukkojen lähelle vakiokomentoja. Niitä sellaisenaan tai muokattuna sovelletaan aina uudelleen tarvittaessa. Tällaiset komentojen muodostamat työkaaviot olivat Survossa suosittuja vielä joitakin vuosia sitten. Uskon tälläkin hetkellä monien survoilijoiden rakentavan ja soveltavan niitä omissa tehtävissään. Nykyisin vastaavat asiat toteutetaan Survon makrokielen avulla usein näppärämmin.

Haluan kuitenkin tässä esitellä yhden esimerkin, joka edustaa komentokaavio-tekniikkaa suppeudestaan huolimatta monipuolisesti. Kuvitelkaamme tilannetta, jossa kunnittain ja äänestysalueittain on kerätty tietoja äänestysvilkkaudesta niin, että yhden kunnan tulokset ovat omassa tekstitiedostossaan (tässä VETINEN.TXT) seuraavassa muodossa:

```
Vetinen:
Kalliolahti 256 210
Likovesi 310 258
Löhölampi 111 68
Sumppusuo 220 175
Raivolähde 159 130
Intojoki 401 344
Kahluupuro 93 82
```

Äänestysalueittain on siis laskettuna äänioikeutettujen ja äänestäneiden lukumäärät. Jokaisesta kunnasta tulisi nyt tulostaa siisti taulukko, jossa äänestysalueet järjestetään äänestysaktiivisuuden mukaan seuraavaan tapaan:

| Vetinen:<br>Alue | Äänioikeutettuja | Äänestäneitä | Äänestysprosentti |
|------------------|------------------|--------------|-------------------|
| Kahluupuro       | 93               | 82           | 88.2              |
| Intojoki         | 401              | 344          | 85.8              |
| Likovesi         | 310              | 258          | 83.2              |
| Kalliolahti      | 256              | 210          | 82.0              |
| Raivolähde       | 159              | 130          | 81.8              |
| Sumppusuo        | 220              | 175          | 79.5              |
| Löhölampi        | 111              | 68           | 61.3              |

Jos tehtävä olisi ainutkertainen eikä sitä tarvitsisi toistaa uusilla aineistoilla, ei olisi mitään vaikeutta poimia tietoja toimituskenttään (esim. SHOW- tai LOADP-komennolla), sitten muotoilla ne siistimmäksi taulukoksi (FORM), laskea äänestysprosentit (esim. kosketuslaskennalla) ja lopulta lajitella (SORT) sekä tulostaa (PRINT) lopullinen taulukko.

Tilanne on toinen, kun tietää, että samat temput pitää välittömästi tai joskus myöhemmin tehdä kymmenille tai sadoille vastaaville aineistoille, joissa äänestysalueiden lukumäärä vaihtelee. Tavalla tai toisella tehtävä tulee automatisoida tai ohjelmoida niin, että seuraavilla kerroilla ei tarvitse yksityiskohtaisesti muistaa ja toistaa samoja työvaiheita. Mietipä, ennenkuin tarkastelet Survon ratkaisua, miten hoitaisit työn suosikkiohjelmillasi niin, että tehtävä on toistettavissa automaattisesti.

- Pitäen tuon toistotarpeen mielessä olen sommitellut seuraavanlaisen työkaavion valmiiksi toimituskenttään ÄÄNESTYS:

```

7 1 SURVO 84C EDITOR Mon Jul 22 19:25:11 1996 C:\KIRJA\ 100 100 0
1 *SAVE AANESTYS / Lajittelu äänestys#:n mukaan
2 *
3 * AAAAAAAAAAAAA 11111 11111
4 * 11111
5 *LOADP _____,A / SKIP=TX
6 *FORM B,END,3
7 *VAR Äänestysprosentti=100*Äänestäneitä/Äänioikeutettuja TO ALUE
8 *-SORT B,END,4
9 *PRINT A,END
10 *
11 *DATA ALUE,B,END,X,Y
12 Y AAAAAAAAAAAAA 11111 11111 111.1
13 A
14 T *5 ( ) ( ) ( )
15 X Alue Äänioikeutettuja Äänestäneitä Äänestysprosentti
16 B
17 *
18 *
19 *
20 *
21 *
22 *
23 *

```

Tässä kaaviossa aktivoitavia komentoja on viisi ja ne sijaitsevat peräkkäisillä riveillä 5-9. Komennot voi suorittaa jopa yhdellä ns. jatkuvalla aktivoinnilla (painetaan ensimmäisen kohdalla nappia F2 ennen aktivointia ESC-napilla), jolloin lopullinen taulukko syntyy ja tulostuu kertarysäyksellä. Nyt tarkastelen tapahtumien kulkua kuitenkin vaiheittain komento komennolta ja selitän samalla, mitä kulloinkin tapahtuu.

Kirjoittamalla tekstitiedoston nimen VETINEN.TXT riville 5 LOADP-komennon perään ja aktivoimalla sen, ko. vaalipiirin tiedot ilmaantuvat toimituskenttään muodossa:

```

1 1 SURVO 84C EDITOR Mon Jul 22 19:25:11 1996 C:\KIRJA\ 100 100 0
1 *SAVE AANESTYS / Lajittelu äänestys#:n mukaan
2 *
3 * AAAAAAAAAAAAA 11111 11111
4 * 11111
5 *LOADP VETINEN.TXT,A / SKIP=TX
6 *FORM B,END,3
7 *VAR Äänestysprosentti=100*Äänestäneitä/Äänioikeutettuja TO ALUE
8 *-SORT B,END,4
9 *PRINT A,END
10 *
11 *DATA ALUE,B,END,X,Y
12 Y AAAAAAAAAAAAA 11111 11111 111.1
13 AVetinen:
14 T *5 ( ) ( ) ( )
15 X Alue Äänioikeutettuja Äänestäneitä Äänestysprosentti
16 BKalliolahti 256 210
17 *Likovesi 310 258
18 *Löhölampi 111 68
19 *Sumppusuo 220 175
20 *Raivolähde 159 130
21 *Intojoki 401 344
22 *Kahluupuro 93 82
23 *

```

Selkeyden vuoksi esitän aktivoidun komennon tummalla ja sen aikaansaamat tulokset tai muutokset harmaalla pohjalla. Komento LOADP poimii siis tekstitiedoston VETINEN.TXT rivit toimituskenttään niin, että ensimmäinen niistä tulee riville A (eli 13). Tuo rivitunnus (numerokin kelpaisi) on LOADP-komennon toisena parametrina. Rivien edessä olevaa kontrollisaraketta, jolla alunperin on pelkkiä tähtiä, käytetään tässä esimerkissä hyväksi monella tavalla.

Kontrollisarake on sikäli kirjoitustilan ulkopuolella, että tekstiä kirjoitettaessa sille ei tapahdu mitään. Jos tätä saraketta kuitenkin haluaa käyttää, sinne pääsee yksinkertaisesti pitämällä nuolinappia "vasemmalle" jonkin aikaa alhaalla. Viitattaessa riveihin on varsinkin yleisissä kaavioissa turvallista käyttää kontrollisarakeeseen merkittyjä tunnuksia, koska mm. rivejä lisättäessä tai poistettaessa yläpuolelta rivinumerot muuttuvat mutta tunnuksat säilyvät.

Tarkkaan ottaen tekstitiedoston rivit eivät sijoitu peräkkäisille riveille, vaan rivit 14 ja 15 (T ja X), jotka ovat jo alkuperäisessä kaaviossa ja joita tarvitaan myöhemmin, on jätetty koskemattomiksi. Tämän on saanut aikaan täsmennys SKIP=TX, joka on kirjoitettu tässä tapauksessa komentoriville 5; se voisi olla muuallakin. Täsmennys tarkoittaa juuri sitä, että mahdolliset kirjaimin T ja X merkityt rivit tulee ohittaa, kun LOADP latao tekstiä toimituskenttään.

Nyt tulee laittaa taulukon sarakkeet oikeille paikoilleen, oikein muotoiltuina. Sen tekee rivillä 6 oleva FORM-komento:

```

7 1 SURVO 84C EDITOR Mon Jul 22 19:25:11 1996 C:\KIRJA\ 100 100 0
1 *SAVE AANESTYS / Lajittelu äänestys%:n mukaan
2 *
3 * AAAAAAAAAAAAA 11111 11111 11111
4 *
5 *LOADP VETINEN.TXT,A / SKIP=TX
6 *FORM B,END,3
7 *VAR Aänestysprosentti=100*Äänestäneitä/Äänioikeutettuja TO ALUE
8 *-SORT B,END,4
9 *PRINT A,END
10 *
11 *DATA ALUE,B,END,X,Y
12 Y AAAAAAAAAAAAA 11111 11111 111.1
13 AVetinen:
14 T *5 ( ) ( ) ( )
15 X Alue Äänioikeutettuja Äänestäneitä Äänestysprosentti
16 B Kalliolahti 256 210
17 * Likovesi 310 258
18 * Löhölampi 111 68
19 * Sumppusuo 220 175
20 * Raivolähde 159 130
21 * Intojoki 401 344
22 * Kahluupuro 93 82
23 *

```

FORM-komento on kohdistettu riviltä B (16) niin pitkälle kuin rivejä riittää (END). Lisäksi FORM-komennossa viitataan maskiriviin 3, jolla sijaitsee muotoilumalli eli osoitetaan sarakkeiden paikat ja esitysmuodot.

Taulukosta puuttuu äänestysprosenttia kuvaava sarake. Se on Survossa laskettavissa useilla eri keinoilla, mm. kosketuslaskennalla ja erityisellä C%-komennolla. Yleisin ja toistuviin sovelluksiin sopivin tapa on määritellä taulukko Survon havaintotaulukoksi DATA-määritelmällä, joka tässä on rivillä 11 ja käyttää VAR-komentoa, joka riviltä 7 aktivoidaankin seuraavaksi:

```

7 1 SURVO 84C EDITOR Mon Jul 22 19:25:11 1996 C:\KIRJA\ 100 100 0
1 *SAVE AANESTYS / Lajittelu äänestys%n mukaan
2 *
3 * AAAAAAAAAAAA 11111 11111
4 * 11111
5 *LOADP VETINEN.TXT,A / SKIP=TX
6 *FORM B,END,3
7 *VAR Äänestysprosentti=100*Äänestäneitä/Äänioikeutettuja TO ALUE
8 *-SORT B,END,4
9 *PRINT A,END
10 *
11 *DATA ALUE,B,END,X,Y
12 Y AAAAAAAAAAAA 11111 11111 111.1
13 AVetinen:
14 T *5 ( ) ( ) ( )
15 X Alue Äänioikeutettuja Äänestäneitä Äänestysprosentti
16 B Kalliolahti 256 210 82.0
17 * Likovesi 310 258 83.2
18 * Löhölampi 111 68 61.3
19 * Sumppusuo 220 175 79.5
20 * Raivolähde 159 130 81.8
21 * Intojoki 401 344 85.8
22 * Kahluupuro 93 82 88.2
23 *

```

DATA-määritelmässä rivillä 11 (jota ei siis koskaan aktivoida) kerrotaan taulukon nimi (ALUE), ensimmäinen varsinainen taulukkorivi (B), viimeinen rivi (END), sarakkeiden nimirivi (X) ja sarakkeiden maskirivi (Y). Näiden tietojen avulla esim. Survon laskenta-, piirto- ja tilastolliset komennot pystyvät hahmottamaan käsittelyn kohteen täydellisesti.

VAR-komennossa (rivillä 7) annetaan laskettavan sarakkeen (Äänestysprosentti) laskukaava normaalein matemaattisin merkinnöin ja ilmoitetaan, mihin aineistoon sitä tulee soveltaa (TO ALUE).

Enää on jäljellä taulukon rivien lajittelu äänestysprosentin mukaan ja tulostus paperille. Lajittelun tekee -SORT-komento, koska alueet halutaan laskevaan järjestykseen. Se lajittelee rivit B:stä eteenpäin laskevaan järjestykseen maskirivin 4 osoittaman sarakkeen mukaan.

Kun lopuksi aktivoidaan PRINT-komento rivillä 9, rivit A:sta eteenpäin tulostetaan kirjoittimella. Tässä tapauksessa on käytössä PostScript-kirjoitin, jolloin ilman mitään lisätäsmennyksiä tulostuksessa käytetään 12 pisteen Times-kirjasinlajeja. Koska kyseessä on vaihtelevan levyinen merkistö, taulukkojen sarakkeet saadaan pidetyksi suorina "tabulaattoririvillä" T (14), jolla esim. numeeristen sarakkeiden paikat osoitetaan sulkumerkeillä ja merkkien leveys-

kerroin annetaan muodossa \*5. Tämä tarkoittaa, että sarakeveveys yhtä merkkiä kohti on 5 pistettä.

```

12 1 SURVO 84C EDITOR Mon Jul 22 19:25:11 1996 C:\KIRJA\ 100 100 0
1 *SAVE AANESTYS / Lajittelu äänestys%:n mukaan
2 *
3 * AAAAAAAAAAAAA 11111 11111 11111
4 *
5 *LOADP VETINEN.TXT,A / SKIP=TX
6 *FORM B,END,3
7 *VAR Äänestysprosentti=100*Äänestäneitä/Äänioikeutettuja TO ALUE
8 *-SORT B,END,4
9 *PRINT A,END_
10 *
11 *DATA ALUE,B,END,X,Y
12 Y AAAAAAAAAAAAA 11111 11111 111.1
13 AVetinen:
14 T *5 ( ) ( ) ( )
15 X Alue Äänioikeutettuja Äänestäneitä Äänestysprosentti
16 B Kahluupuro 93 82 88.2
17 * Intojoki 401 344 85.8
18 * Likovesi 310 258 83.2
19 * Kalliolahti 256 210 82.0
20 * Raivolähte 159 130 81.8
21 * Sumpusuo 220 175 79.5
22 * Löhölampi 111 68 61.3
23 *

```

On selvää, että juuri esiteltyä kaaviota voi käyttää kuinka monta kertaa tahansa ottamalla LOAD-komennolla aina uudelleen ÄÄNESTYS-kentän puhtaana esiin, kirjoittamalla riville 5 uuden tiedostonimen ja aktivoimalla (jatkuvasti) komennot tältä samalta riviltä lähtien. Samoin on hyvä varustaa kaavio sanallisilla selityksillä niin, että sellainenkin, joka ei tunne Survon käyttötapoja hyvin, pystyy sitä soveltamaan.

Ensinäkemältä kuvattu kokonaisuus voi vaikuttaa hiukan mutkikkaalta; onhan siinä tullut esitellyksi samalla kertaa monia piirteitä, joista ei aikaisemmin ole tehty selkoa. Todellisuudessa työkaavioiden rakentelu on selkeää, kun tuntee tarvittavat rakenneosat. Yleensä kaavio syntyy vaiheittain yrityksen, erehdyksen ja lopulta onnistumisen kautta.

## Paitsiosäännöt

- Monet survoilijat mieltävät Survon käytön eräänlaiseksi peliksi, koska on mielenkiintoista sommitella ja rakentaa käteviä ratkaisuja vaikkapa edellisenkaltaisten työkaavioiden muodossa. Pelissä on jo kuvattujen piirteiden lisäksi muitakin sääntöjä, jotka koskevat pelikentällä liikkumista.

Editoriaalisessa laskennassa Survolta edellytetään kykyä hahmottaa ja löytää itsenäisesti laskentaan tarvittavia aineksia. Samoin järjestelmän tulee tunnistaa havaintotaulukkoja toimituskentästä. Jotta hahmottaminen onnistuisi, vaaditaan tiettyjä sopimuksia ja sääntöjä myös sen varalta, etteivät samaan kenttään

sijoitetut lähisukuiset kaaviot häiritsisi toisiaan. Tarvitaan siis eräänlaisia paitiosääntöjä.

Tarkat säännöt ovat tarpeen senkin vuoksi, että monet sovellukset koostuvat osakaavioista, jotka mukavuussyistä kannattaa asetella samaan toimituskenttään. Näitä osakaavioita säätelevät tiedot, joita Survossa kutsutaan täsmennyksiksi (specifications). Ne ovat eri kaavioissa osittain yhteisiä ja osittain yksityisiä. Kun jokin komento aktivoidaan, ei saa syntyä tulkintaongelmaa siitä, mitkä täsmennyksistä nyt ovat voimassa.

- Näytteeksi paitiosääntöjen olemuksesta otan Survo-kirjan esimerkin sivulta 319. Siinä tarkastellaan Helsingin heinäkuun keskilämpötilaa ja sademäärää 135 peräkkäiseltä vuodelta 1844-1978. Kun aikaisemmin piirrettiin histogrammi lämpötilasta ja verrattiin otosjakamaa normaalijakaumaan, käytettiin  $\chi^2$ -testiä eli numeerista kriteeriä. Samaa asiaa tarkkaillaan nyt pelkkien kuvien avulla piirtämällä sekä lämpötilan että sademäärän jakaumien kertymäfunktiot ns. todennäköisyyspaperille. Silloin aidon normaalisti jakautuneen otoksen pitäisi kuvautua likimäärin suoralle viivalle. Epänormaalin otoksen kuvaan tulee selvää käyryyttä. Silmämääräisesti on huomattavasti helpompi aistia suoraviivaisuutta kuin sitä, yhtyykö histogrammi Gaussin kellokäyrään kyllin hyvin.

```

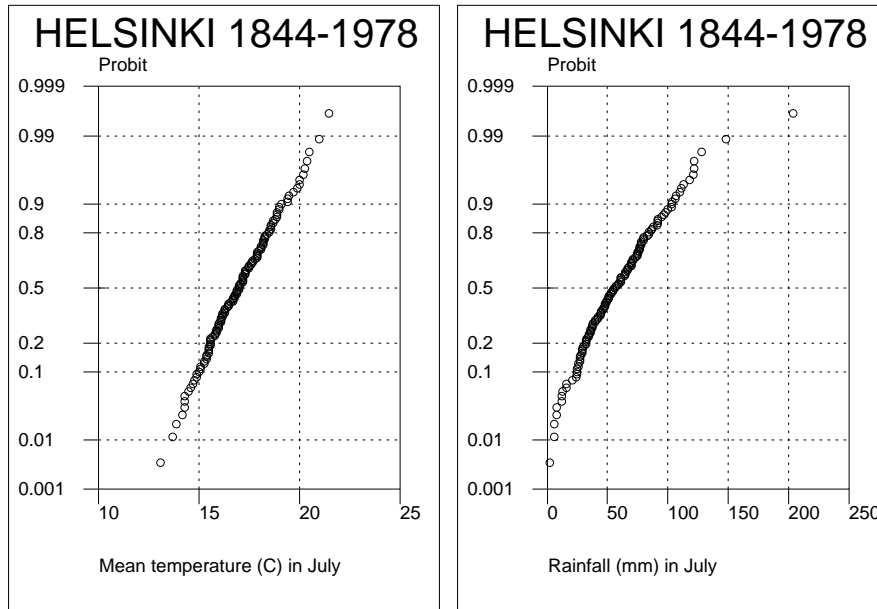
25 1 SURVO 84C EDITOR Wed Jul 24 11:16:26 1996 C:\E\EX\ 100 100 0
1 *
2 * *GLOBAL*
3 * SIZE=570,800 XDIV=120,400,50 GRID=[line type(2)],XY TICK=0,1
4 * YSCALE=*probit(y),0.001,0.01,0.1,0.2,0.5,0.8,0.9,0.99,0.999
5 * HEADER=[Swiss(15)],HELSINKI_1844-1978 PEN=[Swiss(7)]
6 * YLABEL=Probit XLABEL=
7 * DEVICE=PS
8 *
9 *FILE SORT HELSINKI BY Temp TO HELSORT
10 *PLOT HELSORT,Temp,PROBIT / DEVICE=PS,HEL1.PS
11 *TEXTS=T1 T1=Mean_temperature_(C)_in_July,120,50
12 *
13 *FILE SORT HELSINKI BY Rain TO HELSORT
14 *PLOT HELSORT,Rain,PROBIT / DEVICE=PS,HEL2.PS
15 *TEXTS=T1 T1=Rainfall_(mm)_in_July,120,50
16 *XSCALE=0(50)250
17 *

```

Kuvista (seuraavalla sivulla) näkee, että keskilämpötilaa vastaava pistekuvio on melko tarkkaan suoralla mutta sademäärän pisteet kaartuvat eli sademäärän jakauman normalisuudesta syntyy epäilystä.

Kumpaakin kuvaa varten on oma työkaavio. Ensimmäinen on toimituskentän riveillä 9-11 ja toinen riveillä 13-16. Kummassakin kaaviossa aineisto lajitellaan ensin tutkittavan muuttujan mukaan nousevaan järjestykseen FILE SORT-komennolla ja kuva piirretään sen jälkeen PLOT-komennolla. Selvyiden vuoksi kaikki aktivoitavat komennot on tässä esitetty tummalla pohjalla. Käytännössä ei ole väristä väliä eikä välistä väriä.





Lisäksi kentän alussa on yhteisiä ominaisuuksia kuvaava alue (rivit 1-7), joka on merkitty tunnussanalla \*GLOBAL\* (rivi 2).

Kaavioiden erottamiseksi toisistaan käytetään pisteiden muodostamia rivejä (tässä 8,12 ja 17). Itse asiassa erottimeksi riittää

\*.....

siis merkki \* , jonka perässä on vähintään 10 pistettä. Käytännössä tuo \* on yleensä kontrollisarakeessa valmiina ja pisteitä asetellaan rivin täydeltä, jolloin Survon toimitin näyttää pisterivin kaksoisjuovana. Näin osakaaviot erottuvat toisistaan selvästi.

Kun jossakin osakaaviossa aktivoidaan komento, Survo etsii tarvittavat täsmennystiedot ensisijaisesti tuosta omasta ympäristöstä. Jollei jotain tietoa, kuten esim. kuvan kokoa osoittavaa SIZE-täsmennystä, löydy omasta osakaaviosta, sitä haetaan alussa olevasta "yhteismaasta" eli avainsanalla \*GLOBAL\* merkitystä alueesta. Sieltä SIZE tässä tapauksessa löytyykin riviltä 3. Jos täsmennystietoa ei esiinny lainkaan, käytetään kuvan koolle oletusarvoja, jotka on valittu kuvatyyppin ja piirtävän laitteen ominaisuuksien perusteella.

Kyseessä on siis asteittainen etsintä suosituimmuusjärjestyksessä:

1. Komentorivi
2. Oma osakenttä
3. \*GLOBAL\*-kenttä
4. Oletusvalinta

Jos sama täsmennystieto (siis esim. SIZE) esiintyy useasti samassa osakentässä, valinta kohdistuu ensimmäiseen ylhäältä alas ja vasemmalta oikealle lukien. Tästä säännöstä ainoa poikkeus on itse komentorivi, sillä komennon perään (kauttaviivalla erotettuna) kirjoitettu täsmennys on aina ensisijainen. Komentorivin ensisijaisuus ei ollut säännöissä mukana aivan alusta. Muistan, että Armo Heikkilä aikoinaan aivan perustellusti ehdotti tätä lisäystä. Se antaa mahdollisuuden kirjoittaa peräkkäisiä komentoja niin, että kullakin on muutama oma täsmennys tarvitsematta erottaa komentoja omiin karsinoihinsa. Kesti kuitenkin jonkin aikaa ennenkuin uskalsin sen lisäyksen tehdä, koska pelkäsin muutoksen aiheuttavan häiriöitä sekä siihen asti tehdyissä ohjelmissa että käyttäjien sovelluksissa. Näin osittain kävikin, mutta olen silti tyytyväinen, että tähän päästiin ja kiitollinen Armo Heikkilälle hyvästä aloitteesta.

Äskeisessä esimerkissä kuvilla on hyvin paljon yhteisiä piirteitä, jotka on sijoitettu \*GLOBAL\*-alueeseen. Sinne on jäänyt riville 7 täsmennys DEVICE=PS, mikä piirtokomennon PLOT yhteydessä tarkoittaa PostScript-kirjoittimen käyttöä. Tätä on sovellettu kuvia kehiteltäessä. Nyt on kuitenkin kummassakin osakaaviossa omat DEVICE-ohjaukset (rivit 50 ja 54), joilla kuvat ohjataan PostScript-kielisinä omiin tiedostoihinsa (HEL1.PS ja HEL2.PS). Nämä omat DEVICE-täsmennykset ovat näin etusijalla edellä mainittujen paitsiosääntöjen mukaan.

Enempää ei ole tarkoitus selitellä tämän kuvaparin tekoa. Suosittelen kuitenkin yksityiskohtien tutkiskelua. Luulisin, että lukija helposti arvaa, miten eri täsmennykset ovat vaikuttaneet lopputulokseen. Tämä on luonnollinen tapo oppia itse tekemään Survolla kuvia. Aikoinaan, ennenkuin Survosta nyky muodossaan oli mitään käyttöoppaita, julkaisin eräänlaisen tyyppikuvasarjan. Siinä esitettiin piirrosten ohessa kaaviot, joilla kuvat olivat syntyneet. Jo pelkästään tältä pohjalta monet loivat piirroksia omiin tarkoituksiinsa.

- Työkaavioiden ja paitsiosääntöjen hallitseminen vaatii oman totuttelunsa. Joskus sattuu niin, että unohtaa asettaa rajaviivan osakaavioiden väliin. Jos kenttä on sattunut vierimään sen verran alaspäin, ettei edellistä kaaviota enää näy, saattaa ihmetellä, mikseivät jotkut nykyisen kaavion asetuksista toimi. Survopa lukee ne rajaviivan puuttuessa tuosta aikaisemmasta kaaviosta. Jokin hinta on siitä maksettava, että saa asetella nappuloita hyvin vapaasti pelilaudalle.

Kaavioita laadittaessa työ dokumentoi itse itsensä. Selityksiä saa tietenkin lisätä kaavioitten ohkeen mielin määrin. Niille, jotka katsovat, että asioiden muis-

tiinpanosta on liikaa vaivaa, tapanani on huomauttaa, että "luova laiskuus on tietojenkäsittelyssä hyve". Tällä tarkoitan sitä, että kannattaa miettiä, miten tehtävästä selviää kaikkein helpoimmalla.

Esim. Windows-ympäristössä monet yksinkertaiset työt saattaa tehdä ensimmäisellä kerralla helpontuntuisesti, mutta miten on sen seuraavan ja sitä seuraavan kerran laita. Monilla ohjelmilla jokainen kerta on yhtä työläs, sillä valinnat ikäänkuin haihtuvat ilmaan. Aikaisemmista yrityksistä on ehkä vain sen verran hyötyä, että muistaa paremmin, miten tehtävä oikein suoritettiin. Ainutkertaiset asiat ovat kuitenkin harvinaisia, toistuvat tavallisia.

Survossa kaikki mitä tehdään myös säilyy kirjoitetussa muodossa toimituskentissä, kun muistaa niitä tallettaa. Äskeinen piirrostehtävä on myös hyvä esimerkki siitä, miten "laiska" käyttäjä hyödyntää aikaisempia omia tai toisten tekemisiä. Samoilla kaavioilla on helppo piirtää mikä tahansa aineisto todennäköisyyspaperille. Kaaviot palvelevat käyttäjää itsenäisten ohjelmien tapaan.



## 25. heinäkuuta 1996

- On kulunut suunnilleen kuukausi siitä, kun ryhdyin tekemään tätä kirjaa. Tapani mukaan olen työskennellyt normaalin rasiituksen ja ylläpidon vaikeasti arvioitavassa välitilassa. Kesän alku oli suosiollinen, koska oli poikkeuksellisen viileää. Viime päivät ovat kuitenkin ylittäneet oman hellerajan, 20 astetta. Niskani on kipeytynyt liiasta koneen äärellä istumisesta, vaikka vietänkin melko säännöllisesti kesälomaa klo 19 ja 21 välillä soutelemalla Haltianselälä, mikäli sää sen sallii.

*Tapio Nummenmaa* soitti aamulla varmistaakseen, miten uusi vaillinaisista aineistoista korrelaatioita laskeva Survon ohjelmamoduli CORRMV asennetaan. Tämä oli viimeisiä ohjelmointikohteita, jonka kanssa olin puuhailut ennen kirjan aloittamista.

Sähköpostista löytyi *Ene-Margit Tiitin* Tartosta lähettämä viesti, jossa hän kutsui Kimmoa ja minua pitämään esitelmää Survosta tilastotieteen opetuksen välineenä kansainvälisen tilastojärjestön (ISI) seuraavassa maailmankokouksessa Istanbulissa. Soitin Kimmolle ja päätimme odottaa tarkempia tietoja, ennenkuin vastaamme. Itse en pidä kongressimatkailusta, mutta Kimmo sanoi olevansa valmis lähtemään.

Eilen olin juuri pannut postiin oikoluettavaksi lähetetyn artikkelini "A measure for variability in multivariate normal distribution", joka tulee ilmestymään ilmeisesti vielä tämän vuoden puolella "Computational Statistics & Data Analysis"-lehdessä. 14-sivuinen artikkeli koskee tutkimustani, joka muodosti puolet viime kesän urakastani. Olin jo vuosien ajan silloin tällöin pohdiskellut ongelmaa, miten mitata vaihtelua yhdellä mittaluvulla varianssin tapaan, kun tilastollinen aineisto on moniulotteinen. Yleisesti tunnetuilla mitoilla kun on selvät heikkoutensa. Kesän alussa viime vuonna sain ajatuksen, jonka saattelemana olen päätenyt omaan ehdotukseeni. Arvelen, että mittani tulee herättämään ristiriitaisia tunteita tilastotieteen tutkijoiden piirissä; hyvä jos syntyy keskustelua. Koko tuon tutkimustyön aikana Survosta on ollut korvaamatonta apua. Saatan palata tähän aiheeseen vielä myöhemmin.

Jouduin kuitenkin viettämään aikaa tylsän oikoluvun parissa ennakoitua enemmän. Olin lähettänyt lopullisen käsikirjoituksen täysin painovalmiina, Survolla tuotettuna PostScript-tiedostona. Tuntien ko. lehden typografian saatoon näet jäljitellä lopullista painoasua omasta mielestäni varsin tyylikkäästi. Kustantaja North-Holland on kuitenkin nyt hieman muuttanut lehden ulkomuotoa. Niinpä oikoluettavakseni toimitettu ladelmä olikin tehty osittain pelkän tekstitiedoston perusteella, jolloin varsinkin kaavojen suhteen piti olla tarkkana. Vedoksesta löytyikin yllättävän monta painovirhettä ja tuntui kiusalliselta joutua uudelleen tarkastamaan jo aiemmin puhdistettua tekstiä.

Kuluneen kuukauden aikana kirjoitustyön lomaan ilmaantui muitakin puuhia. Vanha ystäväni *Juha Partanen* oli pyytännyt ottamaan yhteyttä logistista regressioanalyysia koskevassa ongelmassa. Hän oli utelias tietämään, miten sitä voisi harrastaa Survossa suurehkoilla aineistoilla. Muistan hyvin, kuinka joskus vuoden 1962 kesällä vietimme yhdessä aikaa Suomen Kaapelitehtaan elektronikkaosastolla laskien jopa yökausia Juhan silloisen tutkimuksen analyyseja Elliott 803-tietokoneella. Kesä oli helteinen. Jouduimme hankkimaan erilaisia puhalluslaitteita - jopa tukankuivaajia - viilentämään ylikuumenneita, suuren komeron kokoisia "sähköaivoja".

Logistista regressioanalyysia pystyy Survossa tekemään yleistettyjä lineaarisia malleja varten laatimallani GENREG-ohjelmalla. Laskenta hidastuu huomattavasti silloin, kun koko aineistoa ei pystytä pitämään kerralla koneen keskusmuistissa, sillä iteratiivisissa menetelmissä ohjelman täytyy kelata aineisto läpi monia kertoja. Olin tehnyt tuon ohjelman alunperin jo vuoden 1988 alkupuolella pienille aineistoille ja parannellut sitä vuoden 1990 lopulla nostaakseni sen kapasiteettia.

Koska Juhan aineistoilla ja malleilla ajoajat olisivat kasvaneet kohtuuttomiksi, lupasin tehdä asialle jotain, jos vain osaisin. Etsin parempaa ratkaisua parin päivän ajan kesä-heinäkuun vaihteessa ja lopulta yllättäen huomasin, että pienellä muistinhallintaa koskevalla muutoksella saatoin ilokseni ja samalla häpeäkseni todeta, että ohjelma alkoi toimia suurilla aineistoilla noin 40 (!) kertaa nopeammin. Jokaiselle survoilijalle, joka kaipaa lisää tehoa GENREG:iin, lähetän mielelläni tämän parannetun version.

Samaan vaiheeseen lomittui myös ensikokeiluni aitojen Windows 95-ohjelmien kanssa. Vertasin Survon tilastollisia toimintoja SPSS-ohjelmiston uuden version (7.0) vastaaviin. SPSS:ssä on logistista regressiota varten oma ohjelmansa joka toimii noin kaksi kertaa nopeammin kuin GENREG. Tämä johtuu paljolti siitä, että ohjelman erikoistuessa vain logistiseen regressioon, laskennassa pystytään hyödyntämään analyysitavan erityispiirteitä.

Hämmästyin kuitenkin aika lailla, kun havaitsin, että SPSS kuluttaa pelkän korrelaatiomatriisin laskemiseen 50 muuttujan, 1000 havainnon aineistolla koneessani yli 25 sekuntia. Survo selviää samasta tehtävästä alle 4 sekunnissa. Matriisin tulostus PostScript-kirjoittimella ottaa yli kymmenkertaisen ajan Survoon verrattuna. Viesti on selvä: ohjelmat todella kadottavat tehoaan paitsuessaan Windows-ympäristössä.

SPSS:n viimeisestä versiosta näyttää syntyneen muutenkin vilkasta keskustelua Internetissä (comp.soft-sys.spss). Ns. Output Navigator eli tulosten uudellinen hallinta- ja esitystapa on syrjäyttänyt aikaisemman, jonka varaan SPSS:n käyttäjät ovat rakentaneet omat sovelluksensa. Tästä monet ovat oikeutetusti huolissaan ja jopa vihaisia. He eivät enää kykene siirtämään kaikkia tulostaulukkoja vaivatta teksturille edelleen muokattavaksi ja raportoitavaksi. SPSS joutunee viipymättä palauttamaan aikaisemman tulostustavan uuden rinnalle.

## Tekstin synty

- Eniten olen viettänyt aikaa kuukauden aikana tämän tekstin äärellä. Esipuheen olen päättänyt laatia vasta, kun näen, mihin tässä päädytään. Muuten olen tehnyt työtä siinä järjestyksessä, jonka uskon olevan lopullinen eli tähän päivään mennessä noin 60 sivua. Tehollista peliaikaa on kertynyt parhaina päivinä 7-8 tuntia, mutta välillä en muiden toimien ja lyhyiden kesälomamatkojen takia ole ehtinyt kirjoittaa riviäkään.

Yritän nyt kuvata, miten tämä teksti syntyy Survon avulla - miten sekalaiset mietteet vähitellen muotoutuvat painoasuun. Vuosien varrella olen erilaisten kirjoitustehtävien yhteydessä urautunut tiettyihin työtapoihin, joista pidän. Niinpä tekstin tekninen suoltaminen ja viimeistely on ongelmatonta, mutta mitä sanoa ja miten, se on ikuinen päänvaiva. Kun vielä varsinaisten tosiasioiden ohella joutuu ottamaan kantaa ja tekemään vertailuja, eteneminen tuppaa takkuuntumaan.

Aikaisemmin kirjoitustyö saattoi keskeytyä myös siihen, että välillä piti ryhtyä ohjelmoimaan. Muistan, miten kirjoittaessani ensimmäistä nykyisen Survon englanninkielistä käyttöopasta vuonna 1987 ja kuvatessani jotain Survon toimintoa takerruin silloin tällöin keskelle lausetta havaittuani, ettei lause pidä paikkaansa, ellen hiukan kehittele ohjelmia. Niinpä kirjoitustyö saattoi keskeytyä järjestelmän parantelun vuoksi päiviksi, jopa viikoiksi. Sitten vasta pääsin hyvällä omalla tunnolla kirjoittamaan lauseen lopun.

Tätä tekstiä tehdessäni tuollaisia keskeytyksiä ei enää ole onneksi juuri ilmaantunut. Vain yhden teknisen parannuksen olen joutunut tekemään tämän työn vuoksi. Saadakseni kuvaruutua jäljittelevät näkymät värien puuttuessa

vivahteikkaammiksi, lisäksi makron FOREBACK(x,y) Survon PostScript-ajuriin. Näin itse tekstin sävyä x ja sen taustan sävyä y voi jatkuvasti säätää harmaan eri astein mustasta valkoiseen.

Kuvaan nyt työn vaiheita antaakseni samalla vaikutelman siitä, miten yleensäkin laajoja tekstimassoja hallitaan Survossa. Koko aineiston, siis toimituskentät ja muut tiedostot olen sijoittanut hakemistoon C:\KIRJA. Työn keskuksena on INDEX-kenttä, jonka löydän tällä hetkellä suoraan päätyövalikostani napilla F.

```

1 1 SURVO 84C EDITOR Fri Jul 26 10:22:20 1996 C:\KIRJA\ 100 100 0
1 *SAVE INDEX / Survo ja minä
2 *>PKZIP A:KIRJA /U C:\KIRJA\*.*
3 *>PKUNZIP A:KIRJA /V
4 *Suomenkielinen Survo-kirja
5 *
6 *INDEX / EDT=LOAD SORT=-DATE
7 *CD C:\KIRJA
8 *CD C:\
9 *
10 *LOAD T13.EDT / 25. heinäkuuta 1996 2 20402 26.07.96 08:15
11 *LOAD T12.EDT / 25. heinäkuuta 1996 15554 25.07.96 08:34
12 *LOAD PAINATUS.EDT / 17415 25.07.96 15:21
13 *LOAD INDEX.EDT / Survo ja minä 10302 25.07.96 13:40
14 *LOAD T11.EDT / Nappulat pelikentällä 3 42218 25.07.96 11:26
15 *LOAD T9.EDT / Nappulat pelikentällä 21715 25.07.96 11:20
16 */GS-SHOW K.PS / 19282 24.07.96 16:15
17 *LOAD T10.EDT / Nappulat pelikentällä 2 46056 24.07.96 11:27
18 *LOAD T8.EDT / Alusta alkaen 2 48581 24.07.96 11:14
19 */GS-SHOW HEL2.PS / 22480 24.07.96 11:09
20 */GS-SHOW HEL1.PS / 22200 24.07.96 11:09
21 *LOAD T1.EDT / Alkua... 30704 24.07.96 10:27
22 *SHOW VETINEN.TXT / 142 23.07.96 14:01
23 *LOAD ÄÄNESTYS.EDT / Lajittelu äänestys%:n mukaan 10302 23.07.96 10:35

```

Tässä toimituskentässä tiedot ovat rivin 6 INDEX-komennolla listattuna tuoreimmasta vanhimpaan. Pääsen siis suoraan aktivoimaan rivillä 10 olevan LOAD-komennon, joka tuo eteeni juuri tällä hetkellä työn alla olevan tekstin, siis tämän, jota parhaillaan luet.

Listassa on nyt tarkalleen 50 viittausta, joista 12 koskee varsinaisia tekstikenttiä ja loput esimerkkejä, PostScript-kuvia jne. Aktivoimalla rivin 2, saan Survon ulkopuolisella PKZIP-ohjelmalla talletetuksi kaikki KIRJA-hakemiston (muuttuneet) tiedostot tiivistettyinä levykkeelle, jolloin vähän väliä saatan varmistaa, että varakopiot ovat ajan tasalla. Tähänastinen aines (noin 60 sivua valmista tekstiä kuvineen ja oheistiedostoineen) vie "zipattuna" tilaa alle 200 kilotavua, mikä lienee varsin vähän verrattuna monien muiden ohjelmien tilantarpeeseen.

Tapanani on antaa tekstikentille hyvin lyhyitä nimiä (kuten tässä T1, T2, ..., T13), koska niiden selitykset ovat luetteloissa selvästi näkyvillä. Tällainen järjestelmällisyys ei ole ehdotonta, mutta siitä on hyötyä esim. silloin, kun ha-

luaa selailta ja muokata tekstimassaa yhtenä jatkuvana kokonaisuutena Survon LIST-operaatioilla.

Tekstiä toimituskenttään kirjoitettaessa ei kannata aluksi mitenkään välittää muotoiluun liittyvistä seikoista. Tyyliini kuuluu kerätä kentän alkuun irrallisia avainsanoja ja ilmauksia ikäänkuin muistilistaksi siitä, mitä tässä yhteydessä pitäisi kertoa. Tätä voisi joku sanoa jopa jäsentelyksi. Kirjoittaessani annan tekstin rönsyillä vapaasti. Muuttelen lauserakenteita ja asioiden esitysjärjestyksestä. Kun kappaleen tai parin mittainen jakso alkaa näyttää sisällöltään kelloiselta, taitan sen lopulliseen painoasuun.

Tekstin muokkaukseen olen tehnyt sukroja (siis Survon makroja). Nämä pikkuapulaiset palvelevat vain omaa työtäni. Esim. kappaleen taitto tapahtuu viemällä kohdistin kappaleita edeltävälle tyhjälle riville ja painamalla 3 näppäimen yhdistelmää F2 N N . Näin käynnistetään ilman näkyvää komentoa sukro, jonka todellinen nimi on #N eli sen voisi aktivoida myös komennolla /#N, Yhdistelmä F2 N N (vain 3 painallusta, joista kaksi viimeistä samoja) on kuitenkin kätevämpi vaihtoehto.

- Vaikka sukroista kerron kunnolla vasta myöhemmin, kannattaa hetkeksi kurkistaa niiden näppäinten taakse. Saan sukron #N listatuksi toimituskenttään TUTLOAD-komennolla:

```
11 *TUTLOAD #N
12 *{tempo -1}TRIM3 63,P3{act}{home}{erase}{tempo +1}{end}
```

Sukron #N listaus mahtuu yllä riville 12 ja sen toiminta on seuraava:

|            |   |
|------------|---|
| {tempo -1} | Valitaan suurin mahdollinen nopeus.                 |
| TRIM 63,P3 | Kirjoitetaan tämä muotoilukomento.                  |
| {act}      | Aktivoidaan muotoilukomento.                        |
| {home}     | Palataan rivin alkuun.                              |
| {erase}    | Tyhjennetään rivi.                                  |
| {tempo +1} | Palataan normaalinopeuteen.                         |
| {end}      | Lopetetaan ja palataan tavalliseen kirjoitustilaan. |

Sukron tehtävänä on vain kirjoittaa tekstikappaleen muotoilukomento, aktivoida se ja pyyhkiä omat jälkensä. Kun on valittu suurin mahdollinen nopeus, nykyisillä koneilla silmä ei juuri ehdi havaita komennon pikavierailua. Vain sen tulos, haluttuun leveyteen (63 yksikköä) tavuttaen taitettu tekstikappale on tästä todisteena.

Toinen sukroapulainen, josta mainitsin jo aikaisemmin monimuuttujamenetelmä-kurssin yhteydessä, on sukro M, joka pika-aktivoinnilla F2 M M antaa mahdollisuuden valita valmiiksi taitetun tekstin osan. Sen jälkeen se näyttää



ulkopuolisen Ghostscript-ohjelman avulla kuvaruudulla suurennetussa koossa tekstin lopullisen painoasun. Näin käytössäni on oiva esikatselukeino.

- Tekstin oikolukuun on tarjolla *Kimmo Koskenniemen* nerokkaan OIKO-ohjelman ensimmäinen versio, johon olen laatinut Survolle ominaisen käyttöliittymän. Kun komento OIO aktivoidaan millä tahansa tyhjällä rivillä, OIKO-ohjelma tarkistaa kaikki alla olevat rivit. Se pysähtyy jokaisen väärinkirjoitetun tai oudon sanan kohdalla, näyttää ko. rivin ympäristöä ja antaa mahdollisuuden välittömään korjaukseen. Tämän vanhimman OIKO-ohjelman oma sanavarasto lienee hyvin suppea nykyiseen verrattuna, mutta saatan kasvattaa sitä jokaisen oikoluvun yhteydessä painamalla nappia + sellaisten sanojen kohdalla, joita OIKO ei hyväksy, mutta jotka itse kelpuutan. Tällaisia sanoja ovat mm. Survo ja monet tilastolliset termit sekä muut ammattisanat.

Olen kokemusteni myötä tullut siihen johtopäätökseen, että riittää mainiosti, että oikolukuohjelma tuntee vain suomenkielen perussanaston. Kun se pysähtyy jokaisen harvinaisemman sanan kohdalla, kannattaa edes hetken miettiä, onko tarvis käyttää kovin sofistikoituja foneemikomplekseja. Sitä paitsi mitä enemmän sanoja oikoluku osaa, sitä suurempi on todennäköisyys, että kirjoitusvirhe menee läpi, koska sana sattuu olemaan jonkin toisen sanan oikein kirjoitettu muoto. Pianovirheitä ei siis välttämättä keksitä. Tunnen hykerryttävää sääliä niitä ihmispoloja kohtaan, joiden tekstoreissa "kirjoitusvirheet korjataan automaattisesti".

Survossa on myös toinen tekstin laatua punnitseva komento LUE. Se ei kuulu Survon peruskalustoon mutta on vapaasti saatavilla. Komennon LUE tarkoituksena on nimensä mukaisesti lukea alapuolellaan olevat toimituskentän rivit (tai annettu tekstitiedosto) ja mitata paitsi tekstin luettavuutta myös sen lukemiseen kuluvaa aikaa.

LUE:n antama luettavuusarvio perustuu *Osmo A. Wiion* tutkimuksiin. Tekstin teoreettista vaikeusastetta mitataan vertaamalla sitä 300 lehtiartikkelin kielen tilastollisiin ominaisuuksiin. Lehtityypit ovat: lastenlehdet, sanomalehdet ja tiedelehdet. Tuloksena saatava vaikeusindeksi vaihtelee tyypillisesti nollasta sataan. Se lasketaan lineaarisen regressiomallin antamana ennusteena.

Kun Osmo Wiio aikoinaan ystävällisesti antoi käyttööni tutkimuksensa tulokset ja ryhdyin laatimaan LUE-ohjelmaa Survoon, sain ajatuksen laajentaa sen toimintaa myös lukuajan mittaamiseen. Jos joku haluaa todella mitata lukemiseen kuluvia aikoja, hän voi ensin "opettaa" ohjelmalle oman lukutyyhinsä. Tämä tapahtuu yksinkertaisesti lukemalla omaa tekstiänsä ohjeiden mukaan, jotka saa aktivoimalla komennon LUE ? . Tällöin ohjelma näyttää kirjoitusta yleensä muutaman virkkeen paloina väliaikaisessa ikkunassa. Lukija hokee virkkeet ja painaa ENTER-nappia kunkin jakson perään. Ohjelma muodostaa näin käsityksen lukijan tyylistä ja tallettaa tiedot myöhempiä lukukertoja var-

ten. Lukunopeutta kuvataan dynaamisella regressiomallilla, jossa vaihtelua selittävät virkkeiden, sanojen ja tavujen lukumäärät. LUE-ohjelma muuntaa myös luvut ja yleisesti tunnetut lyhenteet lausuttuun asuun, jotta tulos paremmin vastaisi todellisuutta. Lukija pystyy koeluennan aikana näkemään, milloin regressiomallin parametrit vakiintuvat riittävästi. Silloin hän keskeyttää toiminnan ja saattaa alkaa käyttää ohjelmaa tosimitoiminnassa.

Kun rakensin tuon ohjelman, pidin sitä enemmän hauskana kokeilukohteena dynaamiselle regressiotekniikalle kuin vakavasti otettavana lukunopeuden mittarina. Sain kuitenkin myöhemmin kuulla, että se on osoittautunut hyödylliseksi käytännössä. Mm. erään suuren liikelaitoksen tutkijat kertoivat, että he saavat usein tehtäväksi valmistella esim. 7 minuutin katsauksen taloudellisesta tilanteesta. Tuntiessaan katsauksen esittäjän puhetyylin, he tekevät juuri sopivan kokoisen puheen mittatilauksena.

Tarkastin koko tämän jakson "25. heinäkuuta 1996" tekstiosan luettavuuden ja lukuajan LUE-komennolla. Wiion luettavuusindeksiksi tuli 50 eli LUE ilmoitti tekstin olevan "sanomalehtikielen rajoissa". Kun "Seppo\_ääneti" käy tekstin läpi, arvioitu lukemiseen kuluva aika on noin 13 minuuttia. Ääneen lukeminen tavallisella puherytmillä veisi kuitenkin jo noin 43 minuuttia ja kuvittelemani "Paasikivi" kuluttaisi aikaa jopa 68 minuuttia.

## Tekstin painatus

- Siitä voi olla monta mieltä, miten aikaisin tulee päättää lopullisesta painoasusta, siis sivukoosta, rivivälistä, palstanleveydestä, kirjasinlajeista jne. Itse pidän ajatuksesta, että nämä asiat on mietitty suurin piirtein valmiiksi työtä aloitettaessa, vaikka Survossa on helppo myöhemmin vaihtaa painoasua. Tässä työssä päädyin heti alusta samaan tyyliin, jota olen soveltanut edellisissä laajemmissa kirjallisissa töissä. Pidän suhteellisen tiiviistä tekstinasettelusta, koska tällöin yhdelle sivulle mahtuu riittävästi asiaa.

Tulostan tekstin luku luvulta, kunkin omalla PRINT-komennollaan. Nämä komennot olen koonnut PAINATUS-nimiseen toimituskenttään. Tämän luvun painatus on tapahtunut oheisella kaaviolla.

Aktivoimalla rivin 160 /GS-PRINT-komennon tämän luvun sivut tulostuvat väliaikaiseksi PostScript-tiedostoksi K.PS levyllä D: (Ramdrive). Tämän komennon asemasta rivillä 160 voisi olla myös yksinkertaisemmin

```
160 *PRINT CUR+1,E
```

jolloin teksti tulisi ulos suoraan kirjoittimesta. Jottei tuhlaisi paperia, on kuitenkin hyvä saada käsitys siitä, miten teksti asettuu sivuille. /GS-PRINT-

komento (joka siis on itseasiassa sukro) paitsi tulostaa tekstin tiedostoon, kutsuu myös Ghostscript-ohjelman näyttämään sivut pienoiskoossa kuvaruudussa, jolloin näen etukäteen, kannattaako ne siirtää myös paperille. Jos näyttää hyvältä, aktivoin rivillä 168 olevan (DOS:in) COPY-komennon, joka siirtää tiedoston K.PS kirjoittimelle eli sivut painetaan paperille. Jos näyttää huonolta, palaan takaisin tekstiin, teen korjauksia ja toistan äskeisen menettelyn.

```

28 1 SURVO 84C EDITOR Sat Jul 27 16:52:34 1996 C:\KIRJA\ 200 80 0
150 *
151 - [PSAVE] [M] [Times(9)]
152 * 25. heinäkuuta 1996 #####
153 - [PRESTORE]
154 *
155 - [PSAVE] [M] [Times(9)]
156 *##### Survo ja minä 25.7.1996/S.Mustonen
157 - [PRESTORE]
158 *
159 *
160 */GS-PRINT CUR+1,E TO D:K.PS_
161 - include SMKIRJA.DV2
162 - [page_number(60)] [M]
163 - header_lines 151,154,155,158
164 - chapter A in T12
165 - chapter A in T13
166 E
167 *
168 *->COPY D:K.PS LPT1
169 *
170 *
171 *
172 *

```

Tulostuskaaviossa PRINT nimellisesti käsittelee komennossa ilmoitetut rivit rivistä CUR+1 (eli kohdistinrivistä 1 eteenpäin, siis 161) riviin E (eli 166). Yksinkertaisissa tapauksissa näillä riveillä on pelkästään tekstiä, joka tulostuu oletusvalintojen mukaan. Oletusvalinnat määrää kirjoitinajuri, tässä tapauksessa tiedosto PS.DEV (Survon yleinen PostScript-ajuri). Ajuri on tekstitiedosto, joka yhdistää Survolle ominaiset tulostusta ohjaavat koodisanat PostScript-kieleen. Survon PRINT-ohjelma käy aina ensin läpi tämän ajurin ja näin valmistee kirjoittimen ottamaan vastaan lisäohjaukset ja varsinaisen tekstin.

Ensimmäinen käsiteltävä rivi 161 viittaa tekstitiedostoon SMKIRJA.DV2. Se määrittelee tarkasti tämän kirjan tulostustyylin ja täydentää näin PostScript-ajuria.

Rivillä 162 kerrotaan ensimmäisen sivun sivunumero ja [M] on tyyli tiedostossa määritelty koodisana. Se valitsee tämän kirjan "leipätekstin", 10.5 pisteen Timesin, 13 pisteen rivivälin ja 55 pican (116 mm), palstanleveyden.

Yksinkertaistaen, tämän kirjan varsinainen tekstiosa voitaisiin tulostaa PRINT-käskyllä

```
*PRINT CUR+1,E
- [Times(10.5)] [line_spacing(13)] [trim(55)]
* Rivi 1
* Rivi 2
*
* jne.
*
E Viimeinen rivi
```

ilman mitään muita lisiä. Tässä merkillä - osoitettua riviä ei tulosteta, vaan se ohjaa tulostusta.

Kirjan oloiseen tulostukseen kuuluu kuitenkin monia yksityiskohtia, joista täytyy pitää huolta. Rivi 163 (header\_lines) koskee sivujen otsikoita ja kertoo, että rivit 151-154 kirjoitetaan aina parittomien sivujen alkuun ja rivit 155-158 parillisten. Koska näille otsikkoriveille on valittu pienempi kirjaskoko [Times(9)], koodisana [PSAVE] pitää huolta siitä, että sivunvaihdon yhteydessä oleva tyyli (tavallisesti leipäteksti) pannaan talteen ja [PRESTORE] palauttaa sivuotsikon jälkeen tuon tyylin takaisin. Merkein ##### osoitetaan sivunumeron paikka.

Vasta rivit 164 ja 165 viittaavat varsinaiseen tekstiin niin, että esim. rivi 165 määrää toimituskentässä T13 olevan kappaleen (chapter) A otettavaksi mukaan tulostukseen. Kun katselee kentän T13 alkua,

```
1 1 SURVO 84C EDITOR Sun Jul 28 09:56:02 1996 C:\KIRJA\ 400 100 0
1 *SAVE T13 / 25. heinäkuuta 1996 2
2 *LOAD T14
3 *
4 *
5 *DEF A,6,END
6 (
7 *Tekstin synty
8 *
9 *Eniten olen viettänyt aikaa kuukauden aikana tämän tekstin äärellä. Esip
10 *olen päättänyt laatia vasta, kun näen, mihin tässä päädytään. Muuten ole
11 )nyt työtä siinä järjestyksessä, jonka uskon olevan lopullinen eli tähän
12 *mennessä noin 60 sivua. Tehollista peliaikaa on kertynyt parhaina päivin
13 *tuntia, mutta välillä en muiden toimien ja lyhyiden kesälomamatkojen tak
14 *ehtinyt kirjoittaa riviäkään.
15 *
16 *Yritän nyt kuvata, miten tämä teksti syntyy Survon avulla - miten sekala
17 *mietteet vähitellen muotoutuvat painoasuun. Vuosien varrella olen erila
18 *kirjoitustehtävien yhteydessä urautunut tiettyihin työtapoihin, joista p
19 *Niinpä tekstin tekninen suoltaminen ja viimeistely on ongelmaton, mutt
20 *mitä sanoa ja miten, se on ikuinen päänvaiva. Kun vielä varsinaisten tos
21 *den ohella joutuu ottamaan kantaa ja tekemään vertailuja, eteneminen tup
22 *takuuntumaan.
23 *
```

näkee, että kyseinen kappale A nimetään rivin 5 DEF-määritelmällä tekstin-osaksi riviltä 6 niin pitkälle kuin kirjoitusta kentässä riittää (END). Survo taittaa rivit sivuiksi automaattisesti. Etteivät sivunvaihdot sattuisi keskelle

taulukkoja tai heti otsikkorivin perään, on tarjolla erilaisia keinoja taiton hienosäätöön. Esim. sulkumerkit kontrollisarakkeessa riveillä 6 ja 11 ohjaavat kyseisen välin tulostumaan samalle sivulle. PRINT-ohjelman tulee tarkkailla näitä säätöjä. Siis havaitessaan alkusulun kontrollimerkinä, PRINT etsii ennen tuon rivin tulostusta seuraavan loppusulun ja mittaa tarvittavan tilan. Jos suluihin merkityt rivit eivät näytä mahtuvan nykyiselle sivulle, siirrytään seuraavalle sivulle ennen painatusta.

Esimerkistä näkyy, että tulostuksen muotoa säädellään Survossa PRINT-listassa riveillä, joilla kontrollisarakkeessa on merkki -. Tulostusta ohjaavat sanat merkitään hakasulkuihin ja ne on määritelty kirjoittimen ohjaustiedostoissa eli tässä tapauksessa ajureissa PS.DEV ja SMKIRJA.DV2. Jokainen ohjaussana on nimettyä erikseen esim. omalla `define`-rivillään. Tällaisia määrittelyrivejä voi olla myös kirjoitettavan tekstin seassa (omilla `-define`-riveillä) ja samat sanat voi määritellä uudelleen kuinka monta kertaa tahansa. Määrittelyissä saa käyttää sekä kirjoittimen omaa ohjauskieltä (eli tässä PostScript) että aikaisemmin nimettyjä hakasuluissa olevia sanoja. Näin kerrostaen ilmaistaan yhä yleisempiä ja laajempia ohjaustietoja.

Survon tulostuksia ohjaavat [hakasulkusanat] ovat pitkälle samoja tulostavasta laitteesta riippumatta. PostScript on nykyisin paras vaihtoehto Survon käyttäjälle. PostScript-kieltä edes hieman osaava voi itse tehdä lisäohjauksia. Näin kaikki tämän yleisen sivunkuvauskielen keinot ovat hyödynnettävissä. Tätä kielitaitoa ei kuitenkaan edellytetä survoilijoilta normaalisovelluksissa.

PostScript-tuen liitin Survoon vuosien 1987 ja 1988 vaihteessa. Olin siihen asti pitänyt Canonin laserkirjoittimia etusijalla, sillä ne olivat ensimmäisiä, joilla saattoi tulostaa kätevästi yhtäaikaa sekä tekstiä että kuvia. PostScriptia olin harkinnut jo jonkin aikaa, mutta Lauri Tarkkonen sai minut lopullisesti innostumaan; antoipa oman kirjoittimensakin lainaksi joululoman ajaksi, jotta tekisin Survoon PostScript-ajurin.

PostScript-kielen rakenne on siinä määrin tavallisista ohjelmointikielistä poikkeava, että tehokkuudestaan huolimatta sen harrastaminen ei ole mikään erityinen nautinto. Jouduin kuitenkin muutaman viikon aikana siihen sen verran perehtymään, että tuo ajuri syntyi. Paljon on sittemmin unohtunut oppimastani, mutta tarvittaessa pystyn kyllä tekemään paikkauksia ja laajennuksia, kun hetken tutkailen vanhaa koodiani. Jokainen pääsee tutustumaan siihen katselemalla ajuria PS.DEV, sillä se on aivan tavallinen tekstitiedosto. Kuten sanoin, survoilijoiden ei kuitenkaan tarvitse vaivautua ymmärtämään tästä kielestä sanaakaan. Tarvittavat PostScript-ilmaisut ovat valmiiksi käännettyinä Survon tulostuskielelle.

En lähde tässä enempää kuvailemaan painatuksen yksityiskohtia, sillä niitä on edistyneemmissä sovelluksissa todella paljon. Survossa ei aloittelijan kannata

liikaa vaivata päätään kaikilla erityispiirteillä. Kirjeet, lausunnot ja muut tavanomaiset tekstit saa siististi paperille hyvin vähillä tiedoilla. Aitojen julkaisujen teossa voi aluksi käyttää hyväksi toisten tekemiä tyylytiedostoja. Tietoja mm. alaviitteiden käytöstä, kuvien liittämistä ja hakemistojen automaattisesta laatumisesta löytyy sekä kyselyn PRINT? kautta että Survon käyttöoppaasta.

## Varjomaailmasta

Tulostusten yhteydessä on erityistä syytä kiinnittää huomiota Survon editoriaalisen käyttötavan yhteen omaleimaisimpaan piirteeseen, varjomerkkien käyttöön. Ettei tarvitsisi pilata tekstiä sen keskelle sijoitetuilla koodausmerkinnoilla, Survossa on ollut alusta alkaen mahdollista liittää jokaiseen toimituskentän riviin varjorivi, joka toimii rivin taustana ja ilmaisee itseään kuvaruudussa merkkien värityksinä. Tekstiin saattaa siis Survossa kohdentaa erilaisia lisäyksiä rivikohtaisesti kontrollimerkeillä ja merkkikohtaisesti varjomerkeillä. Kummallakin keinolla on keskeinen tehtävä mm. tulostuksen säätelyssä.

Varjomerkkien tuottamat värit selkeyttävät näkymiä kuvaruudulla. Niillä tuotetaan silmäniloa valikkoihin ja otsikkoihin. Kun jotkin yksityiskohdat ovat kesken työssäni, saatan merkitä ko. paikat vilkkuvilla sanoilla (varjo 5), jolloin palatessani työhön uudelleen, en voi olla huomaamatta työstämistä kaipaavia kohteita.

Varjomerkit ovat hyödyllisiä myös suurten tekstitietokantojen käsittelyssä (Survon LIST-operaatiot). Tällöin eri värejä käytetään tietojen joustavaan koodaukseen. Väryksestä on hyötyä jopa ohjelmoinnissa. Itse kirjoitan Survon makrokielellä ohjelmia (sukroja) toimituskenttään. Tällöin saatan sopivalta värityksellä korostaa vaikkapa viimeksi muutettuja kohtia, jolloin ne pilkistävät selvästi esiin koodin kätkeistä.

Kohdistimen osoittaman rivin varjorivi ilmaantuu näkyville tilapäisesti rivin alapuolelle painamalla nappeja F2 ja S:

Kun näin on tehty seuraavassa kaaviossa rivillä 16, rivi 17 on korvautunut tilapäisesti rivin 16 varjorivillä. Itse värit eivät erotu tässä mustavalkoisessa kuvassa, mutta kuvaruudulla sana "sekaisin" on valkeaa tekstiä sinisellä pohjalla, ensimmäinen "kirjaimia" vihreää valkoisella ja toinen "kirjaimia" punaista valkoisella. Varjorivi poistuu näkyviltä painamalla uudelleen nappeja F2 ja S, mutta värit säilyvät osoittaen, että ko. riviin liittyy jotain erikoista.

```

1 1 SURVO 84C EDITOR Sun Jul 28 11:16:25 1996 C:\KIRJA\ 23 100 0
1 *SAVE VARJOT / Varjomerkit
2 *
3 *PRINT 4,END
4 - define [M] [Times(12)]
5 - define [I] [Courier(12)]
6 - shadow m: [M]
7 - shadow i: [I]
8 *
9 - [M]
10 *Näillä riveillä on tekstiä
11 *sekä Times- että Courier-kirjasinlajeilla.
12 - [I]
13 *Tämä on Courier-tekstiä.
14 - [M]
15 *Tämä on Times-tekstiä.
16 *Tässä on sekaisin Courier-kirjaimia ja Times-kirjaimia.
Shadow: 44444444 i 66666666m 11111111
18 *
19 *
20 *
21 *
22 *
23 *

```

Rivin 3 PRINT-komennon aktivointi tuottaa seuraavaa jälkeä kirjoittimella:

```

Näillä riveillä on tekstiä
sekä Times- että Courier-kirjasinlajeilla.
Tämä on Courier-tekstiä.
Tämä on Times-tekstiä.
Tässä on sekaisin Courier-kirjaimia ja Times-kirjaimia.

```

Vertaamalla tulostusnäytettä PRINT-kaavioon, jolla se tehtiin, näet rivin 16 varjomerkkien vaikutuksen. Varjo 4 alleviivaa näin merkityt kohdat, varjo 6 muuttaa tekstilajin *kursiiviksi* ja varjo 1 **lihavoit**. Määrätyt numeromerkit varjoina vastaavat kaikilla tulostuslaitteilla samoja tehosteita niin paljon kuin mahdollista. Näin olen päättänyt tehdessäni ajureita eri kirjoittimille. Survoilija oppii nopeasti muistamaan silloin jo värityksen perusteella, millaisesta tehostuksesta on kysymys.

PRINT-listan alussa on neljä ohjausriviä (merkki - kontrollisarakeessa). Kahdella ensimmäisellä määritellään koodisanalla [M] 12 pisteen Times ja koodisanalla [I] 12 pisteen Courier. Kahdella seuraavalla - shadow-rivillä määritellään varjomerkki m tarkoittamaan koodisanaa [M] ja varjomerkki i koodisanaa [I]. Tekstilajia saatetaan siten vaihtaa sekä rivien välillä ohjausriveillä (rivit 9, 12 ja 14) että rivin sisällä varjomerkeillä m ja i (kuten on tehty rivillä 16).

Tässä tulostusesimerkissä oli vain neljä määrittelyriviä. Käytännössä mieleistä tyyliä luotaessa niitä syntyy enemmän. Tällöin määrittelyrivit talletetaan usein

omaa tyylietiedostoon esim. OMA.DV2 ja ne tulevat voimaan panemalla PRINT-listan alkuun ohjausrivin - include OMA.DV2.

Kun tyyli on keskitetysti määriteltynä omassa tekstitiedostossaan, sitä on helppo muuttaa koska tahansa poimimalla tyylietiedot toimituskenttään, tekeillä parannukset ja tallettamalla takaisin tiedostoon.

## Selailua

Pitkät tekstit jaetaan Survossa useihin toimituskenttiin. Kenttien koot vaihtelevat riippuen tilanteesta. Uusi luku aloitetaan uudesta kentästä. Pitkät luvut on helppo jakaa sopiviin osiin. Tämän kirjan tähän asti kirjoitettu alkuosa koostuu 6 luvusta, jotka on jaettu 13 toimituskenttään T2,T3,...,T14.

Kentät on ketjutettu siten, että kunkin kentän toisella rivillä on valmis LOAD-komento seuraavalle. Jo tämä tekee koko tekstin selailun kenttien yli suhteellisen vaivattomaksi. Survo tarjoaa tekstimassan hallintaan myös yleisempiä ja tehokkaampia keinoja LIST-komentojen muodossa.

Esimerkiksi tähänastista materiaalia voi katsella yhtenä kokonaisuutena LIST SHOW-komennolla ja samalla etsiä kiinnostavia kohtia hakusanojen avulla.

```

18  1 SURVO 84C EDITOR Mon Jul 29 20:25:44 1996          C:\KIRJA\ 100 100 0
1  *
2  *LIST SMLISTA: A T2-14 END
3  *
4  *LIST SHOW SMLISTA
1  LIST SMLISTA: A in T2 (lines 6 - 374)
6  *
7  *Pientä laskentaa
8  *
9  - [mtk(0,0,0)] [M]
10 *Käy jotenkin yli ymmärrykseni, että lisääntyvästä "helppokäyttöisyydestä
11 *huolimatta on edelleen tilanteita, joissa yksinkertaiset tehtävät saatta
12 *yllättävän hankalia yleisesti käytetyille ohjelmille.
13 *
14 *Olen usein kysynyt, miten eri käyttöympäristöissä toimitaan, kun esim. k
15 *kirjeen kirjoittamisen tarvitsee tehdä joitain yksinkertaisia laskutoimi
16 *joiden tulokset pitäisi liittää tekstiin ja sitten jatkaa kirjoittamista
17 *
18 *Ajatelkaamme vaikka tunnontarkkaa raportoijaa kirjoittamassa:
19 *
20 - [U]
21 * Kiinassa on tällä hetkellä arviolta 612461200 mies- ja 586041300 nais-
22 * puolista kansalaista. Väestömäärä on siis
23 - [M]
To stop, press F8!      F1=HELP

```

Useista toimituskentistä tai niiden kappaleista (chapter) muodostuva tekstikonaisuus, jota Survossa kutsutaan listaksi, nimetään toimituskentässä LIST-määritelmällä, jollainen esimerkissämme on kirjoitettuna riville 2. Siinä määritellään lista, joka koostuu A-nimisistä kappaleista toimituskentissä T2-14 eli T2,T3, ..., T14. Kappaleet tulee määritellä DEF-riveillä ao. toimituskentissä,



mikä tässä tapauksessa on jo tehty tulostusta varten. Listalle on annettu nimi SMLISTA ja se sitoo yhteen mainitut kappaleet eli koko tähänastisen tekstin yhdeksi saumattomaksi koosteeksi.

Tätä koostetta selataan LIST SHOW-komennolla (rivi 4). Sen aktivointi avaa komentorivin alapuolelle tilapäisen ikkunan, johon ilmestyy aluksi listan alkupää. Nyt voi mm. nuolinapeilla selata listaa ylhäältä alas ja päinvastoin. Toimituskenttien rajat näkyvät värillisinä otsikkoriveinä, mutta muuten lista on täysin yhtenäinen. Selauksessa lista käyttäytyy ikäänkuin teksti olisi yhdessä pitkässä toimituskentässä eikä listan pituudelle ole mitään muuta ylärajaa kuin koneen levytila.

LIST SHOW on omiaan tekstinhauissa. Painetaan vain hakunappia Alt-F5, jolloin ikkunan alaosaan tulee kehotus antaa hakusana. Hakusanaa kirjoitettaessa LIST SHOW etsii jatkuvasti merkki merkiltä ensimmäistä kohtaa, johon tuntomerkki sopivat. Painamalla ESC-nappia löytyy seuraava esiintymiskohta. Onpa mahdollista laskea pikaisesti kaikkien esiintymiskohtien lukumääräkin.

LIST SHOW on kätevä pitkien tekstien korjailussa. Silloin sitä kannattaa käyttää /LIST-sukron välityksellä. Korjattava sana tai ilmaisu etsitään ensin hakunapilla listasta. Sitten LIST SHOW keskeytetään tilapäisesti napilla F8 (mistä on valmis kehotus alarivillä). Silloin se toimituskenttä, jonka kohdalle pysähdyttiin, tulee normaaliin tapaan näkyville. Kaikessa rauhassa tehdään tarvittavat korjaukset, talletetaan kenttä ja palataan listassa entiseen paikkaan komennolla /LIST \*, jonka voi määritellä makrokksi.

Toisena esimerkkinä LIST-operaatioiden soveltamisesta tekstien tarkistukseen kannattaa vielä mainita LIST COUNT-komento, jolla lasketaan listasta erilaisten merkkien, kirjainten, numeroiden, sanojen ja yleisten merkkijonojen esiintymiskertoja.

Etsittävien merkkijonojen luettelo annetaan ennen komennon LIST COUNT aktivointia PHRASES-rivin alapuolella. Oheisessa kaaviossa on laskettu luettelossa Survosanat kerrottujen ilmaisujen lukumäärät, jotka listan läpikäynnin jälkeen ilmestyvät rivien jatkeeksi edellä esitettyyn tapaan (tässä havainnollisuuden vuoksi harmaalla taustalla).

Joukossa voi olla erilaisia yleisilmaisuja, joista nyt on valittu vain rivien (#lines), sanojen (#words) ja merkkien (#characters) lukumäärät. Tarkistuksen vuoksi olen halunnut katsoa, että esim. alku- ja loppusulkujen lukumäärät täsmäävät. Koska PostScript-sanassa c-kirjain tahtoo minulta usein tippua pois, olen tutkiskellut tätäkin.

```
29 1 SURVO 84C EDITOR Tue Jul 30 11:22:37 1996 C:\KIRJA\ 40 80 0
1 *
2 *LIST SMLISTA: A T2-14 END
3 *
4 */LIST SMLISTA
5 *
6 *LIST COUNT SMLISTA, Survosanat_
7 *PHRASES Survosanat
8 *#lines // 3527
9 *#words // 20106
10 *#characters // 172802
11 *( // 288
12 *) // 288
13 *[ // 205
14 *] // 205
15 *Survo // 407
16 *voi // 166
17 *saattaa // 13
18 *PostScript // 27
19 *PostS // 27
20 *Suomenkieliset opetusohjelmat // 2
21 *
22 *
23 *
```

Jos olet jaksanut lukea tähän asti kaiken huolella, olet äskeisen tilaston mukaan nähnyt yli 170000 kirjainta, numeroa ja erikoismerkkiä. Mukana ovat myös sanavälit, sillä ilman niitä lukemisestatuleetosihankalaa. Itse olen naputellut merkeistä ehkä reilusti yli puolet. Loput olen poiminut valmiista esimerkkikaavioista tai Survo on niitä tuottanut toimintojensa tuloksina.





## Tilastollista käsittelyä

- On yhä tavallista, että Survo mielletään "tilasto-ohjelmaksi". Ei siis tiedetä, että kokonaisuuteen suhtauttaen varsinainen tilastollinen laskenta ja tilastolliset analyysit ovat jo kauan sitten jääneet vähemmistöön. Survo puhtaana tilasto-ohjelmiana on auttamatta historiaa. Oikeudenmukaisinta on luonnehtia Survoa yleiseksi käyttöympäristöksi, jossa erityistä huolta kannetaan tilastollisista tehtävistä.

Tilastollisia analyysimenetelmiä on tavattoman paljon. Yksikään edes ns. suurista tilasto-ohjelmista (esim. SAS ja S-Plus) ei kata kaikkea, vaan paljon tekemistä riittää edelleen erikoisohjelmille. Vaikka Survossa tilastomenetelmien valikoima ei vedä vertoja laajimmille tilasto-ohjelmille, se ei ole vakava puute. Tilastollisten menetelmien suhteen tarve jakautuu hyvin epätasaisesti. Soveltavassa tilastollisessa tutkimuksessa kysyntä painottuu selkeästi perusmenetelmiin, jotka ovat mukana kaikissa tilasto-ohjelmissa, myös Survossa.

Koska eri ohjelmistoissa tarjonta tilastotieteen erikoisalojen suhteen vaihtelee, tasaisesti parasta vaihtoehtoa ei voi olla olemassakaan. Survon vahvimmat alueet perusmenetelmien ulkopuolelta löytyvät käsittääkseni tilastollisesta grafiikasta, tilastollisista monimuuttujamenetelmistä ja tilastollisten mallien simuloinnista. Nämä alat kiinnostavat eniten itseäni ja edelleenkin haluan kehittää Survoa antaumuksella näillä suunnilla.

Keskeinen kysymys tilasto-ohjelmia vertailtaessa ei ole, mitä niillä voi tehdä, vaan miten ne sen tekevät. Tapaan sanoa, että tilastotiede ei ole yksinäänä tässä maailmassa. Tilastollista tutkimusta harrastavalle ei siis riitä, että hän saa haluamansa laskelmat ja analyysit tehdyksi. Yhä tärkeämmäksi vaatimukseksi on tietojenkäsittelyn kehittyessä noussut se, että tutkimustyöhön olennaisesti kuuluvat oheistoiminnot sujuvat kitkatta ja nivoutuvat ongelmanratkaisuun.

Tällaisia oheistoimintoja ovat mm. tutkimusaineistojen kerääminen, talletus, muokkaus, tutkimukseen liittyvä yleinen kirjanpito, varainhallinta, taulukko-

jen laatiminen, kuvien piirtäminen, muu kuin tilastollinen laskenta, tutkimus-selostuksen kirjoittaminen ja sen viimeistely painokuntoon.

Yleensä ajatellaan, että kuhunkin oheistoimintoon valitaan siihen erikoistunut ohjelma. Esim. tutkimusraportti tehdään tilasto-ohjelman ulkopuolisella teksturilla, kuvat grafiikkaohjelmalla ja lopullinen selostus jollakin julkaisuohjelmalla. Tällöin luotetaan käyttöympäristön palvelukykyyn ja eri ohjelmien yhteistyöhön. Esim. Windows-ympäristössä tämä on luonnollinen tapa toimia. Se kuitenkin edellyttää perehtymistä erityyppisten ohjelmien käyttötapoihin ja tehokkaita laitteita. Näillä ehdoilla työskentelyn pitäisi onnistua. Kun seuraa Windowsista käytävää keskustelua, huomaa kuitenkin, ettei kaikki aina mene-kään niin joustavasti kuin mitä luvataan. Silmiin pistää, miten joutavanpäiväisistä asioista saattaa kehittyä vakavaa pohdintaa ja miten mitättömät ongelmat kiusaavat ihmisiä. Tämmöisestä malliksi voi viitata vaikkapa SPSS:n uusimman version tulostusongelmiin ja kyselyyn (sfnet.atk.ms-windows toukokuussa 1996) "...miten saisi Word 6.0-7.0:n oikoluvun hyväksymään esim. sanan kuorma-auto...".

Tekee mieleni sanoa, että Survossa asiat ovat toisin. Nykyistä Survoa lähdin heti alusta viemään itseriittoiseen suuntaan. Tahdoin koota tilastollista tutkimusta tukevat toiminnat saman katon alle yhdessä tilastollisen analyysin ja laskennan kanssa. Näin Survossa on päästy pitemmälle toimintojen yhdentämisessä kuin esim. Windowsissa erityisesti osaavien ja vaativien käyttäjien kannalta.

Ensimmäisiä laajempia ohjelmointikohteita oli PRINT-operaatio, sillä halusin nopeasti tuottaa selostuksia järjestelmän kehityksestä. Siihen, että erilaiset oheistoiminnot ovat olennainen osa Survoa, on osaltaan vaikuttanut työympäristöni. Ainakaan Helsingin yliopiston laitoksilla ei tutkijoilla ja opettajilla ole liiemmin käytettävissä sihteerityövoimaa. Itse olen vuosikymmenet vastannut omien juttujeni puhtaaksikirjoittamisesta ja viimeistelystä painoasuun. Niinpä päätin tehdä Survosta oman sihteerini.

Hiukan kärjistäen voisi väittää, että olen rakentanut Survoa lähinnä omia tutkimus- ja opetustarpeitani varten. Jos siitä on hyötyä muillekin, olen tyytyväinen. Totta puhuen, Survoon on tullut lisätyksi lukuisia ominaisuuksia myös toimeliiden käyttäjien aloitteesta. Olen aina yrittänyt aistia, mitä käyttäjät kaipaavat ja jollakin aikataululla toteuttaa toiveet.

Tietenkään askartelu tilastotieteestä sivussa olevien toimintojen parissa ei kuulu "toimenkuvaani". Jo silloin, kun Survo todella oli pelkkä tilasto-ohjelma, saatettiin kysyä, milloin taas palaan tekemään tutkimustyötä. Musertavimman arvostelun esitti eräs kirjapainoalan edustaja tieteellisen julkaisu-toiminnan tulevaisuutta koskevassa keskustelutilaisuudessa. Kertoessani omassa alustuksessani Survon mahdollisuuksista julkaisujen teossa hän närkästynenä

ihmetteli, miten "professori voi tuhlata aikaansa tehtävään, joka kuuluu vain alan ammattilaisille".

Ottaen oppia tästä lausumasta, keskityn nyt ehdottomasti kertomaan vain tilastollisesta tietojenkäsittelystä Survolla. Koska tämä ei ole sen enempää tilastotieteen kuin Survonkaan oppikirja, en aio selittää vastaan tulevia tilastollisia käsitteitä ja menetelmiä. Tästä huolimatta uskon, että lukija saa yleiskuvan tilastollisen tietojenkäsittelyn luonteesta. Rajatussa tilassa pystyn luomaan vain karkean läpileikkauksen Survon mahdollisuuksista tällä alueella.

## Tiedon esittäminen ja hallinta

Tilastollista ja muuta systemaattista tietoa kuvataan yleensä ns. havaintomatriisina, jonka sarakkeina ovat muuttujat ja riveinä havainnot. Havaintoaineisto on siis periaatteessa aina taulukon muotoinen, mutta sen ilmiasu saattaa Survossa vaihdella riippuen muuttujien ja havaintojen määrästä ja käsillä olevista tehtävistä. Tämä on poikkeuksellista, sillä monet tilastolliset ohjelmistot tukeutuvat vain yhteen aineiston esitystapaan, joka tyypillisesti on ajateltu suurten tietomäärien kannalta. Tilastotiede on kuitenkin usein parhaimmillaan pienten aineistojen käsittelyssä.

Suppeat aineistot kirjoitetaan Survossa suoraan toimituskenttään joko yksinkertaisena peräkkäisten lukujen jonona tai varsinkin usean muuttujan tapauksessa taulukkona. Näitäkään ei sananmukaisesti tarvitse aina "kirjoittaa". Jos tieto on jo valmiiksi tallennettuna esim. tekstitiedostoon, se on helppo poimia toimituskenttään ja muotoilla Survon edellyttämään asuun. Kannattaa muistaa se sääntö, että kunnon tiedonhallinnassa mitään tietoa ei pidä joutua kirjoittamaan uudelleen, kun se kerran on järjestelmään syötetty.

## Datalistat

- ☐ Luvussa "Pientä laskentaa" englannin opettaja olisi voinut - oppilaitten nimistä välittämättä - tehdä laskelmansa myös toisella tavalla:

```

15  1 SURVO 84C EDITOR Mon Aug 05 12:26:14 1996      C:\KIRJA\ 100 100 0
 1  *
 2  *Viime englannin kieliopin kokeessa annoin oppilailleni seuraavat
 3  *numerot:
 4  *DATA KOE: 7 8 10 4 6 9 9 6 7 8 7 5 8 5 END
 5  *
 6  *CORR KOE,CUR+1_
 7  *KOE  N=14
 8  *Variable  Mean      Std.dev.
 9  *KOE      7.071429   1.730464
10  *

```

Hän on kirjoittanut riville 4 koenumeroita ja antanut DATA-määritelmällä aineiston nimeksi KOE. Keskiarvo ja muut tunnusluvut laskettiin aikaisemmin STAT-komennolla. Nyt on käytetty CORR-komentoa, joka yleisesti laskee muuttujien keskiarvot, hajonnat ja korrelaatiokertoimet. Koska kyseessä on yksi ainoa muuttuja, mitään korrelaatioita ei voi laskea. Keskiarvo (Mean) ja hajonta (Std.dev.) ovat tietenkin samat kuin aiemmin STAT:illa saadut.

Hieman isompana data- eli havaintolistana katseltakoon "nopanheiton" tuloksia:

```

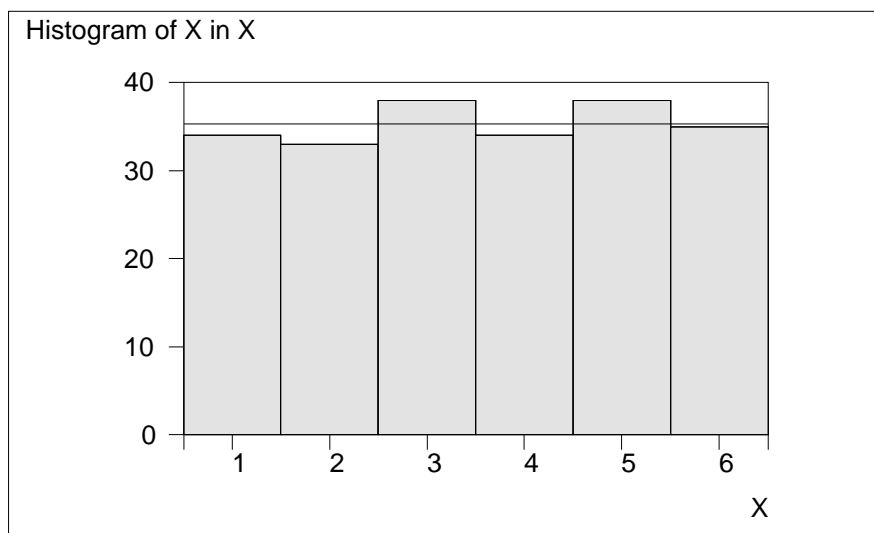
17 1 SURVO 84C EDITOR Mon Aug 05 12:37:48 1996 C:\KIRJA\ 100 100 0
12 *
13 *DATA X: 4 6 3 5 2 1 3 2 6 1 5 4 3 5 4 1 3 2 6 5 6 3 5 2 4 1 1 5 4 6 2 4
14 *3 2 6 5 4 6 2 4 1 3 1 6 3 5 2 1 3 6 5 4 3 5 4 1 3 6 6 3 5 2 2 1 4 6 3 2
15 *3 2 1 5 4 5 2 3 6 1 4 6 5 2 1 2 6 1 5 3 4 1 2 6 5 3 5 2 1 4 3 2 4 3 6 5
16 *4 6 5 4 1 3 1 4 6 3 2 1 3 1 5 4 5 1 2 5 6 5 4 1 2 1 5 6 3 3 6 4 6 2 4 1
17 *3 6 6 3 5 2 2 6 1 5 4 1 2 5 6 3 2 1 1 5 4 3 4 3 1 5 4 6 2 4 3 6 1 4 6 5
18 *2 6 1 5 4 3 4 3 2 3 5 1 2 5 4 3 4 2 6 5 4 6 5 3 6 3 5 2 1 3 2 6 1 5 3 5
19 *END
20 *
21 *GHISTO X,X,CUR+2 / X=0.5(1)6.5 FIT=Uniform(0.5,6.5)
22 * XSCALE=0.5:,1(1)6,6.5:
23 *Frequency distribution of X in X: N=212
24 *
25 *Class midpoint f % Sum % e e f X^2
26 * 1.0 34 16.0 34 16.0 35.3 35.3 34 0.1
27 * 2.0 33 15.6 67 31.6 35.3 35.3 33 0.2
28 * 3.0 38 17.9 105 49.5 35.3 35.3 38 0.2
29 * 4.0 34 16.0 139 65.6 35.3 35.3 34 0.1
30 * 5.0 38 17.9 177 83.5 35.3 35.3 38 0.2
31 * 6.0 35 16.5 212 100.0 35.3 35.3 35 0.0
32 *Mean=3.537736 Std.dev.=1.694467
33 *Fitted by UNIFORM(0.5,6.5) distribution
34 *Chi-square=0.660 df=5 P=0.9851

```

Tunnustan heti, etten tätä varten ole heittänyt noppaa kertaakaan enkä ole edes kirjoittanut tätä 212 heiton tulossarjaa toimituskenttään. Se on Survolla tehty.

Aineisto näkyy riveillä 13-19 ja sille on annettu nimi X, mikä samalla on aineiston ainoan muuttujan nimi. Niinpä rivin 21 GHISTO-komennossa ensimmäinen X viittaa aineistoon ja toinen X muuttujaan. GHISTO piirtää kuvaruutuun histogrammin ja kirjoittaa tarvittaessa numeeriset tulokset toimituskenttään (rivit 23-34). Ohjauksella FIT=Uniform(0.5,6.5) aineistoon on sovittu tasainen jakauma, koska kunkin pisteluvun todennäköisyys on sama harhattomalla nopalla. Yhteensopivuus tämän näkemyksen kanssa on sekä kuvasta katsoen että  $\chi^2$ -testin perusteella (rivi 34) hyvä eli vain  $100 \cdot (1 - 0.9851) = 1.49$  prosentissa vastaavia kokeita otos käyttäytyisi yhtä tasaisesti.

Näin hyvä tulos on tietenkin mahdollinen, mutta kokenut tilastotieteilijä saattaa haistaa "palaneen käryä". Tulos vaikuttaa liian siistiltä, vaikka onkin mahdollinen oikealla nopalla.



- Tässä tapauksessa, kuten jo tunnustin, en ole heittänyt noppaa, vaan sanoin Survon tehneen tämän koesarjan. Survolla on helppo matkia aidon nopan käyttäytymistä, kuten myöhemmin näytän. Tässä tapauksessa koesarja on kuitenkin kaikkea muuta kuin satunnainen. Se on saatu jakamalla lukua 7 suuremmat alkuluvut (jaottomat luvut) 11,13,17,19,23,29,31,37,... luvulla 7 ja ottamalla talteen jakojäännökset eli 4,6,3,5,2,1,3,2,... (Tämän olen helposti toteuttanut sukrolla, josta kerron aikanaan.)

Alkulukujen epäsäännöllisestä esiintymisestä johtuen saattaisi kuvitella, että nuo jakojäännökset käyttäytyisivät nopanheiton tapaan satunnaisesti. Näin ei ole asian laita. Jos vaikkapa lisätään "havaintojen" määrä 6x800:ksi, jakojäännösten frekvenssit ovat

|             |     |     |     |     |     |     |
|-------------|-----|-----|-----|-----|-----|-----|
| Jakojäännös | 1   | 2   | 3   | 4   | 5   | 6   |
| Frekvenssi  | 798 | 801 | 804 | 798 | 798 | 801 |
| Poikkeama   | -2  | +1  | +4  | -2  | -2  | +1  |

eli poikkeamat ovat aivan liian pieniä aitoon satunnaisuuteen verrattuna. Tässä tapauksessa  $\chi^2$ -testi asettaa satunnaisuuden kyseenalaiseksi 99.999-prosenttisesti. Itselleni oli yllätys, että nuo jakojäännökset ovat tilastollisesti katsoen yllättävän säännöllisiä. Uskoisin, että lukuteoreetikot pystyvät selittämään, mistä se johtuu.

Kun Survolla matkii niin hyvin kuin osaa aitoa sattumaa, yhteenveto tyypillisestä heittosarjasta voisi olla



|            |     |     |     |     |     |     |
|------------|-----|-----|-----|-----|-----|-----|
| Pisteluku  | 1   | 2   | 3   | 4   | 5   | 6   |
| Frekvenssi | 773 | 800 | 768 | 829 | 794 | 836 |
| Poikkeama  | -27 | 0   | -32 | +29 | -6  | +36 |

eli poikkeamat ihannetuloksesta tulevat selvästi suuremmiksi. Paradoksaalista kyllä, tulos, jossa kaikki poikkeamat ovat nolliä, on silti se kaikkein todennäköisin, mutta tuonkin tapahtuman todennäköisyys on häviävän pieni eli noin 0.0000000014 . Lotossa täysosuman - 7 oikein 39 numerosta - saa lähes 50-kertaisella todennäköisyydellä.

Nämä todennäköisyydet on laskettu Survossa seuraavasti:

```

1 1 SURVO 84C EDITOR Mon Aug 05 16:09:47 1996 C:\KIRJA\ 100 100 0
28 *
29 *Noppaa heitetään 4800 kertaa , jolloin odotettu frekvenssi kullekin
30 *silmäluvulle on f=800.
31 *Todennäköisyys sille, että kutakin silmälukua saadaan juuri 800 kpl., on
32 *
33 *P1=exp (fact.l(m*f) -m*fact.l(f) -m*f*log(m) )
34 *
35 *missä m=6 ja fact.l(n) on Survon kirjastofunktio luvun n kertoman
36 *luonnolliselle logaritmillem.
37 *Tällöin P1=0.00000000136659
38 *
39 *Lotossa täysosuman todennäköisyys on vuorostaan
40 *
41 *P2=1/C(39,7)
42 *
43 *missä C(n,m) tarkoittaa binomikerrointa.
44 *Siis P2=0.00000006501554
45 *ja P2/P1=47.575144286705
46 *

```

Ihmisen on yleensä vaikea hahmottaa ja matkia satunnaisuutta. Tavallisesti käy juuri niin, ettei liian siistiä koesarjaa osata epäillä. Sattuma on omaa todellisuuttamme ihmeellisempi.

### Datataulukot

- Vaikka myös usean muuttujan aineistoja voi esittää Survossa havaintolistoina, taulukkomuoto on niille tavallisesti suotuisampi. Taulukkoja kirjoitettaessa sarakkeet tulevat automaattisesti suoristetuksi käyttämällä T-kirjaimella merkitty sarkainriviä, joissa sarakkeiden paikat osoitetaan myös T-kirjaimilla. Sarakkeesta seuraavaan liikutaan sarkainapilla. Taulukkoja muotoillaan tarvittaessa mm. FORM-komennolla.

Tässä on hieman vanhentunutta pohjoismaista väestötilastoa:

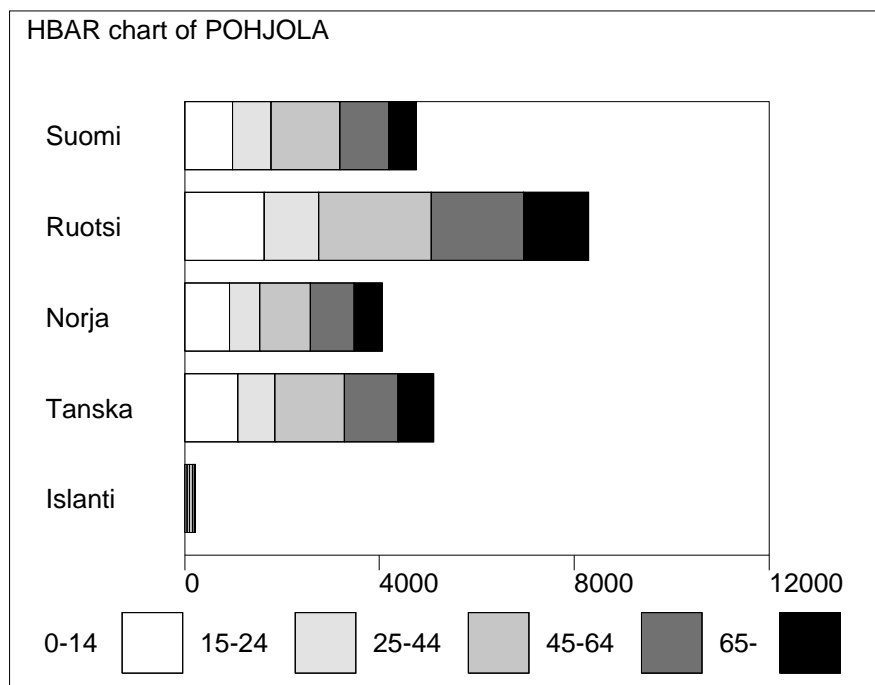
```

14 1 SURVO 84C EDITOR Mon Aug 05 16:36:39 1996 C:\KIRJA\ 100 100 0
60 *
61 * Pohjoismaat: Asukasluku v.1980 tuhansina ikäluokittain
62 *
63 *DATA POHJOLA
64 * Maa      0-14 15-24 25-44 45-64 65-
65 * Suomi   990  780 1420 1015 555
66 * Ruotsi 1641 1117 2316 1883 1345
67 * Norja   925  617 1041  895 588
68 * Tanska 1102  753 1433 1099 724
69 * Islanti  61   43   59   39  22
70 *
71 *GLOT POHJOLA
72 *

```

Nyt DATA-määritelmällä (rivi 63) on nimetty havaintotaulukko POHJOLA. Ellei muuta ole mainittu, Survo olettaa taulukon alkavan heti seuraavalta riviltä niin, että ensin ovat sarakeotsikot (muuttujien nimet rivillä 64) ja sen perässä itse aineisto ensimmäiseen tyhjiin riviin asti (eli rivit 65-69).

Tavallinen pylväskuvio (ikäryhmien suuruudet maittain) syntyy kuvaruutuun, kun aktivoi komennon GLOT POHJOLA (tässä rivillä 71).

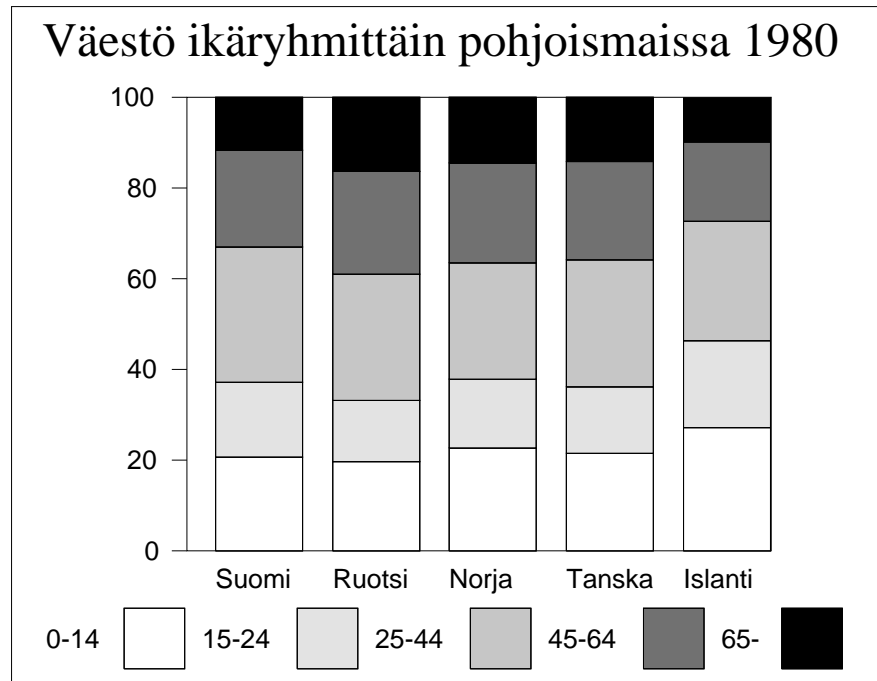


Survolla on toimissaan tietty ennakkokäsitys, mitä tehdä, jos annetussa tehtävässä ei ole täsmällisiä ohjeita. Tämantapaisessa piirroksessa Survo valitsee vaakasuoran pylväikön ja värittää eri osat (ikäryhmät) toisistaan poikkeavin

värein - mustavalkotulostuksessa harmaan eri astein. Kaikkiin yksityiskohtiin pääsee vaikuttamaan.

Koska Islanti on kovin pieni maa, sen tiedot jäävät melkein näkymättömiin. Tilanne paranee, jos siirrytään prosentuaaliseen esitykseen. Käännetään samalla pylväät pystyyn. Tämä saadaan aikaan lisäämällä kaavioon täsmennys `TYPE=%VBAR` (Vertical BARs in percentages). Samalla kuvan oletusotsikko on muutettu `HEADER`-täsmennyksellä.

```
27 1 SURVO 84C EDITOR Mon Aug 05 16:36:39 1996 C:\KIRJA\ 100 100 0
60 *
61 * Pohjoismaat: Asukasluku v.1980 tuhansina ikäluokittain
62 *
63 *DATA POHJOLA
64 * Maa      0-14 15-24 25-44 45-64 65-
65 * Suomi   990   780  1420  1015  555
66 * Ruotsi  1641  1117  2316  1883  1345
67 * Norja   925   617  1041   895  588
68 * Tanska  1102   753  1433  1099  724
69 * Islanti  61    43    59    39   22
70 *
71 *GPLOT POHJOLA / TYPE=%VBAR_
72 *HEADER=[Times(18)],Väestö_ikäryhmittäin_pohjoismaissa_1980
```



Tästä mm. havaitaan, että Suomessa ja Islannissa 1980-luvun alussa vanhimpien ikäluokkien osuus on selvästi pienempi kuin muissa pohjoismaissa.

Erilaisia `TYPE`-täsmennyksellä valittavia pylväs- ja sektorikuvatyyppejä Survossa on toistakymmentä ja niistä löytyy tietoa kyselyllä `TYPE?`. Eri tavoin

kuvia kerrostamalla yhdistelmiä syntyy vielä lisää. Kuvan kokoon, muotoon, otsikoihin, lisäteksteihin jne. voi vaikuttaa monin tavoin.

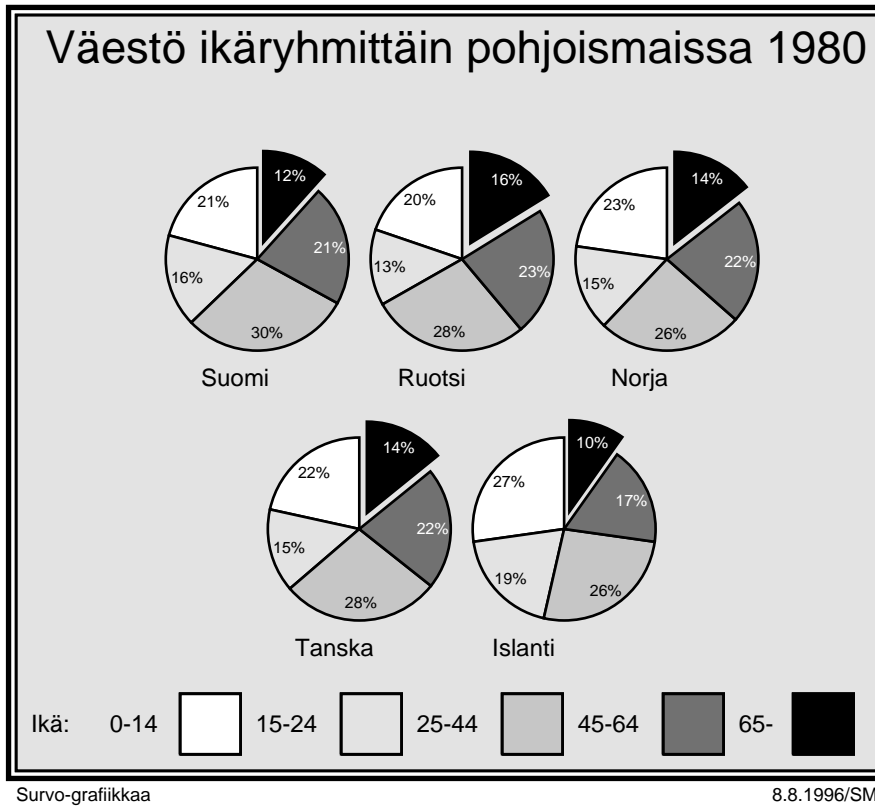
- Jalostamalla piirroskaavio vaikkapa muotoon

```

13 1 SURVO 84C EDITOR Thu Aug 08 10:59:02 1996 C:\KIRJA\ 150 100 0
82 *
83 *PLOT POHJOLA / TYPE=%PIE DEVICE=PS,POHJOLA3.PS
84 *HEADER=[Swiss(17)],Väestö ikäryhmittäin pohjoismaissa 1980
85 *SIZE=1140,1000 PEN=[Swiss(9)][autom_color(0)] LINETYPE=[line_width(1)]
86 *SHADING=0,1,2,5,9P (varjostukset, P tarkoittaa sektorin ulosvetoa)
87 *ANGLE=90 (lähtökulma ensimmäiselle sektorille)
88 *VALUES=[Swiss(6)][autom_color(0.6)],##%,8
89 *XDIV=1,5,1 YDIV=1,5,1
90 *LEGEND=Ikä:
91 *FRAMES=F1,F2 F1=0,0,1140,1000,1
92 * F2=[line_width(2)],-10,-10,1160,1020
93 *TEXTS=T1,T2 T1=[Swiss(7)],Survo-grafiikka,0,-40
94 * T2=[Swiss(7)],8.8.1996/SM,1000,-40
95 *

```

pohjoismaista syntyy sektorikuva



Kukaan ei rakenna tällaista kaaviota suoraan, vaan se muotoutuu vaiheittain kokeillen ja lisäillen erilaisia täydennyksiä. Minulla ei ole mitään sitä vastaan,

että käytetään mielikuvitusta, rohkeita värejä ja sävyjä, kunhan kuvan tietosisältö ei hämähärry.

- Ns. "business-grafiikassa" on kuitenkin valitettavasti päässyt vallalle innostus tehdä pylväs- ja sektorikuvista kolmiulotteisen näköisiä, vaikka tämmöisellä temppuilla vain heikennetään kuvan kertomaa. Kun vielä piirakkakuvat pannaan kallelleen, saatetaan niitä käyttää jopa tietoiseen valehteluun, koska etummaisat sektorit (esim. omat tuotteet verrattuna muihin) saadaan näyttämään todellisuutta mahtavammilta.

Ilmeisesti "yleisön pyynnöstä" jotkin ohjelmat tyrkyttävät harhaanjohtavia esitystapoja. Esim. Excel antaa valittavaksi 10 erilaista pylväs- ja sektorikuva-tyyppiä, joista kokonaista 6 on "kolmiulotteista". Tämä osoittaa ohjelman tekijältä täydellistä vastuuttomuutta ja alistumista markkinoiden ehdoille. Tämä ei ole vain oma kantani, vaan asiasta tilastotieteen tutkijat ovat yksimielisiä.

On opetuksen kannalta joskus ihan suotavaakin sisällyttää tilastolliseen ohjelmistoon huonojakin menetelmiä, jotta pääsee tekemään vertailuja. Silti tuollaisia "kolmiulotteisia" kuvia en sentään suostu Survoon liittämään. Aidosti moniulotteisten tilastoaineistojen kuvallinen esittäminen on sen sijaan erittäin haastava tehtävä, johon vielä palaan.

## **Painavia aivoituksia**

- Hiukan laajempaan näytteenä datataulukoiden hallinnasta käy nisäkkäiden aivojen painoa koskeva tutkimusaineisto. Tämä esimerkki löytyy Survon suomenkielisistä opetusohjelmista valitsemalla

### **A. Tilastolliset menetelmät**

#### **9. Lineaarinen regressioanalyysi**

#### **4. Nisäkkäiden aivojen painot.**

Tämä opetusjakso esittelee aineiston seuraavan kaavion osoittamassa muodossa.

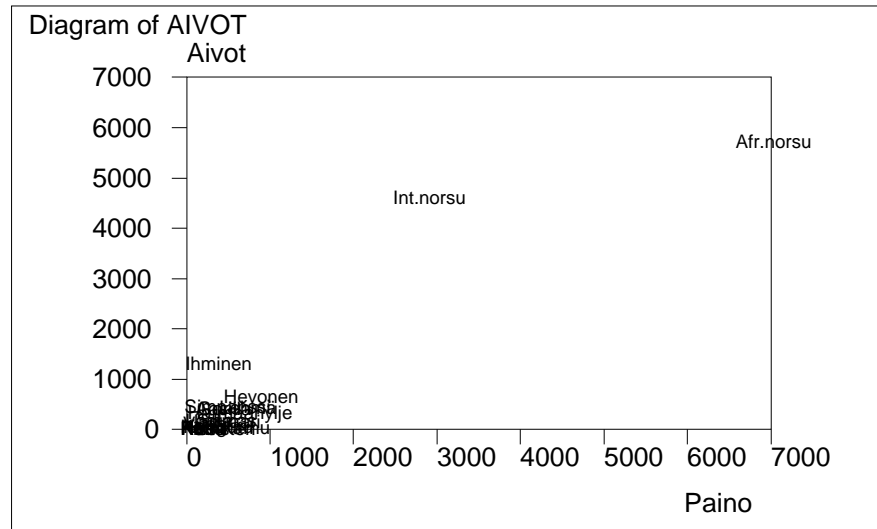
Havaintotaulukossa AIVOT nisäkkäiden painot on annettu kiloina ja aivojen painot grammoina. Koska tähän taulukkoon on tarkoitus muodostaa laskemalla uusia sarakkeita, DATA-määritelmässä (rivi 12) viitataan paitsi varsinaisen taulukon ensimmäiseen ja viimeiseen riviin (A,B) myös otsikkoriviin N ja maskiriviin M. Tällainen DATA-määritelmä voisi olla muuallakin kuin juuri taulukon yläpuolella.

On mielenkiintoista tutkia aivojen painon riippuvuutta koko kehon painosta. Kannattaa aloittaa piirtämällä riippuvuutta kuvaava korrelaatiodiagrammi, joka syntyy komennolla GPLOT:

```

1 1 SURVO 84C EDITOR Tue Aug 06 08:07:08 1996 C:\KIRJA\ 200 100 0
1 *
2 *   Nisäkkäiden aivojen painot
3 *
4 *   Tutkimme aineistoa, johon on koottu eräiden nisäkkäiden keski-
5 *määräinen paino kiloina ja aivojen keskipaino grammoina. Tarkoitus
6 *on selvittää, millainen on aivojen painon riippuvuus koko painosta
7 *ja millaiset ovat lajikohtaiset erot.
8 *   Otamme tässä esille vain osan aineistosta, jonka T.Allison ja
9 *D.V.Cicchetti keräsivät 1970-luvulla tutkiessaan varsinaisesti
10 *nisäkkäiden nukkumista.
11 *
12 *DATA AIVOT,A,B,N,M
13 N Laji      Paino      Aivot
14 M AAAAAAAAAA 1111.111  1111.1
15 A Afr.norsu  6654      5712
16 * Int.norsu  2547      4603
17 * Hevonen   521       655
18 * Lehmä     465       423
19 * Gorilla   207       406
20 * Sika      192       180
21 * Aasi     187.1     419
22 * Jaguaari  100       157
23 * Harmaahylje 85       325
24 * Ihminen   62       1320
25 * Lammas    55.5     175
26 * Simpanssi 52.160   440
27 * Kenguru   35       56
28 * Vuohi    27.66    115
29 * Pesukarhu 4.288    39.2
30 * Kettu    4.235    50.4
31 * Kissa    3.3     25.6
32 * Rotta    0.28    1.9
33 * Hamsteri 0.12    1
34 B Hiiri    0.023   0.4
35 *
37 *GPLOT AIVOT,Paino,Aivot_ / POINT=Laji
38 *

```



Kuvasta ei saa paljoakaan selvää, sillä suurikokoiset norsut tallovat muut alle. Ohjauksella POINT=Laji pelkän pisteen asemasta kuvaan tulevat nisäkäiden nimet. Silloin alkukirjaimen keskikohta asettuu siihen kohtaan kartalla, jonka X-akselin määrää paino ja Y-akselin aivojen paino.

Äkkijyrkkien vaihtelujen tasoittamiseksi ja ottaen tutkittavan ilmiön luonteen huomioon kannattaa yrittää logaritmista muunnosta kummassakin muuttujassa ennen piirtämistä. Survossa toki voi piirtää suoraankin epälineaarisille asteikoille, mutta jatkoa ajatellen tehdään ensin nuo muunnokset VAR-komennolla.

```

23 1 SURVO 84C EDITOR Tue Aug 06 08:30:01 1996 C:\KIRJA\ 200 100 0
11 *
12 *DATA AIVOT,A,B,N,M
13 N Laji Paino Aivot LogP LogA
14 M AAAAAAAAAA 1111.111 1111.1 11.111 11.111
15 A Afr.norsu 6654 5712 8.803 8.650
16 * Int.norsu 2547 4603 7.843 8.434
17 * Hevonen 521 655 6.256 6.485
18 * Lehmä 465 423 6.142 6.047
19 * Gorilla 207 406 5.333 6.006
20 * Sika 192 180 5.257 5.193
21 * Aasi 187.1 419 5.232 6.038
22 * Jaguaari 100 157 4.605 5.056
23 * Harmaahylje 85 325 4.443 5.784
24 * Ihminen 62 1320 4.127 7.185
25 * Lammas 55.5 175 4.016 5.165
26 * Simpanssi 52.160 440 3.954 6.087
27 * Kenguru 35 56 3.555 4.025
28 * Vuohi 27.66 115 3.320 4.745
29 * Pesukarhu 4.288 39.2 1.456 3.669
30 * Kettu 4.235 50.4 1.443 3.920
31 * Kissa 3.3 25.6 1.194 3.243
32 * Rotta 0.28 1.9 -1.273 0.642
33 * Hamsteri 0.12 1 -2.120 0.000
34 B Hiiri 0.023 0.4 -3.772 -0.916
35 *
36 *GPLOT AIVOT,Paino,Aivot / POINT=Laji
37 *
38 *VAR LogP,LogA TO AIVOT
39 *LogP=log(Paino) LogA=log(Aivot)
40 *

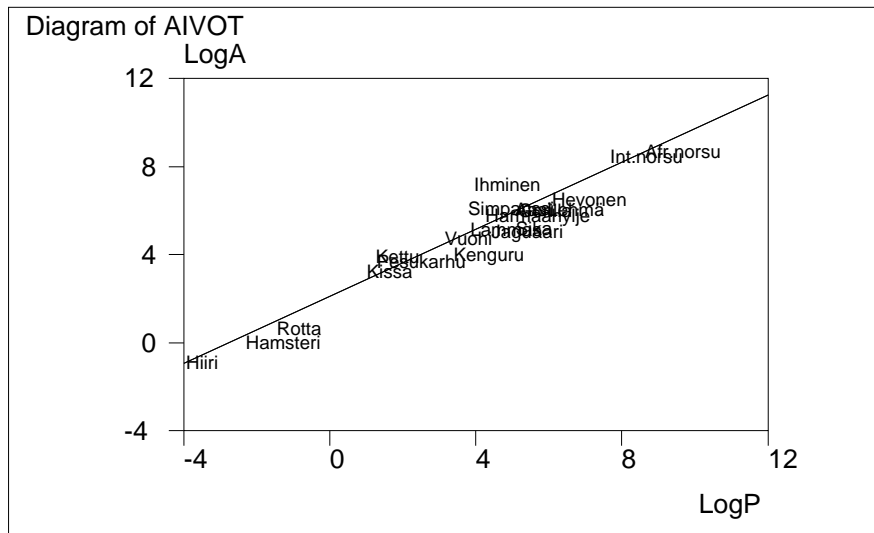
```

Uusien muuttujien nimet ja maskit on ensin kirjoitettu riveille 13 ja 14. Itse VAR-komennossa luetellaan laskettavat muuttujat ja niitä kuvaavat lausekkeet saa kirjoittaa minne tahansa. Tässä ne on asetettu riville 39. Kun siis rivin 38 VAR-komento aktivoidaan, kenttään ilmestyvät uudet sarakkeet (yllä merkitynä harmaalla pohjalla).

Kun nyt rivin 36 GPLOT-komento editoidaan muotoon

```
36 *GPLOT AIVOT,LogP,LogA / POINT=Laji TREND=0
```

ja aktivoidaan, saadaan selkeämpi kuva:



Riippuvuus näyttää siis muunnetuilla muuttujilla olevan melko suoraviivaista. Kuvan määrittelyyn on sen vuoksi lisätty täsmennys `TREND=0`, joka piirtää lineaarisen trendin. Jo nyt nähdään, että ihminen poikkeaa muista havainnoista. Sitä vastaava piste on selvimmin trendin yläpuolella eli ihmisen aivot ovat painoltaan suhteettoman suuret muihin nisäkkäisiin verrattuna.

Tuota riippuvuutta voi tutkia tarkemmin lineaarisella regressioanalyysillä, mikä tapahtuu Survossa esim. seuraavalla `LINREG`-komennolla:

```

16 1 SURVO 84C EDITOR Tue Aug 06 08:35:44 1996 C:\KIRJA\ 200 100 0
40 *
41 *LINREG AIVOT,42 / VARS=LogP(X),LogA(Y),Jäännös(R) RESULTS=0
42 *Linear regression analysis: Data AIVOT, Regressand LogA N=20
43 *Variable Reqr.coeff. Std.dev. t beta
44 *LogP 0.762731 0.049807 15.31 0.964
45 *constant 2.110434 0.233268 9.047
46 *Variance of regressand LogA=6.428643253 df=19
47 *Residual variance=0.483719603 df=18
48 *R=0.9637 R^2=0.9287
49 *DW=1.8264
50 *

```

Rivin 41 `LINREG`-komento laskee lineaarisen mallin tulokset riviltä 42 eteenpäin (riville 49). Malli on määritelty tässä tapauksessa `VARS`-täsmennyksellä, joka sattuu olemaan komentorivillä. Se määrittelee muuttujan `LogP` selittäjäksi (`X`), muuttujan `LogA` selitettäväksi (`Y`) ja muuttujan `Jäännös` (kts. seuraavaa kaaviota) mallin residuaaleja kuvaavaksi muuttujaksi (`R`). Tuloksista havaitsee, että lineaarinen regressiomalli toimii suhteellisen hyvin. Aivojen (logaritmissen) painon vaihtelusta koko painon logaritmi selittää lähes 93%.



```

1 1 SURVO 84C EDITOR Tue Aug 06 08:35:40 1996 C:\KIRJA\ 200 100 0
11 *
12 *DATA AIVOT,A,B,N,M
13 N Laji Paino Aivot LogP LogA Jäännös
14 M AAAAAAAAAA 1111.111 1111.1 11.111 11.111 11.111
15 A Afr.norsu 6654 5712 8.803 8.650 -0.175
16 * Int.norsu 2547 4603 7.843 8.434 0.341
17 * Hevonen 521 655 6.256 6.485 -0.397
18 * Lehmä 465 423 6.142 6.047 -0.748
19 * Gorilla 207 406 5.333 6.006 -0.172
20 * Sika 192 180 5.257 5.193 -0.927
21 * Aasi 187.1 419 5.232 6.038 -0.063
22 * Jaguaari 100 157 4.605 5.056 -0.567
23 * Harmaahylje 85 325 4.443 5.784 0.285
24 * Ihminen 62 1320 4.127 7.185 1.927
25 * Lammas 55.5 175 4.016 5.165 -0.009
26 * Simpanssi 52.160 440 3.954 6.087 0.961
27 * Kenguru 35 56 3.555 4.025 -0.797
28 * Vuohi 27.66 115 3.320 4.745 0.102
29 * Pesukarhu 4.288 39.2 1.456 3.669 0.448
30 * Kettu 4.235 50.4 1.443 3.920 0.709
31 * Kissa 3.3 25.6 1.194 3.243 0.222
32 * Rotta 0.28 1.9 -1.273 0.642 -0.497
33 * Hamsteri 0.12 1 -2.120 0.000 -0.493
34 B Hiiri 0.023 0.4 -3.772 -0.916 -0.149
35 *

```

Malliin nähden "ylipainoiset" aivot ovat ennen muuta ihmisellä ja jossain määrin myös simpanssilla sekä ehkä ketulla. Selvin negatiivinen jäännös on siällä, mutta se johtunee enemmän raskaasta ruumiinrakenteesta kuin aivojen kehittymättömyydestä. Sama koskee lehmää. Tuloksia saattaa vielä tarkentaa, mutta tämä riittääköön nyt.

Tämä aineisto oli vielä suhteellisen ihanteellinen esitettäväksi ja käsiteltäväksi suoraan toimituskentässä. Huomaa, miten helposti siihen lisättiin uusia muutujia. Kaikki tiedot, siis aineisto, sitä käsittelevät komennot ja niiden antamat tulokset sekä käyttäjän mahdollisesti väliin kirjoittamat kommentit ovat kootusti ja kätevästi samassa tilassa.

Selvästi suurempiakin aineistoja on mahdollista pitää toimituskentässä, mutta omasta mielestäni jo äskeistä hiemankin laajempi havaintotaulukko kannattaa yleensä siirtää ajoissa havaintotiedostoon. Se on hyvin yksinkertaista. Esim. AIVOT-aineisto kopioidaan Survon datatiedostoksi AIVOT2 komennolla

```
FILE COPY AIVOT TO AIVOT2 .
```

## Datatiedostot

- ☐ Survon pääasiallista ruokaa numeerisen tiedon käsittelyssä ovat havaintotiedostot, joiden kokoa ei rajoita juuri muu kuin käytettävissä oleva levytila. Suurimmat tiedostot, joita tiedän survoilijoiden luoneen, ovat käsittäneet pari tuhatta muuttujaa ja kymmeniä tuhansia havaintoja. Simulointikokeissa olen itse tarvinnut tiedostoja, joissa on miljoona havaintoyksikköä.

Vaikka muuttujia voi tiedostossa olla hyvin paljon, kerralla tilastolliseen analyysiin niitä saa Survossa mahtumaan tyypillisesti korkeintaan 90. Rajoitus johtuu DOS-käyttöjärjestelmän muistinhallinnasta, mutta se tulee käytännössä harvoin vastaan. Tavallisissa selitysmalleissa, esim. regressioanalyysissa, muuttujien määrä nousee harvoin useisiin kymmeneen. Faktoriaanalyysissa jotkut saattavat kaivata suurempiakin muuttujamääriä. Piirtämällä kuvia, tutkiskelemalla perusjakaumia ja tekemällä järkeviä osa-analyysseja muuttujien määrä karisee faktoriaanalyysissäkin pian rajan 90 alapuolelle. Useimmissa toiminnoissa havaintoyksiköiden määrälle ei käytännössä ole mitään ylärajaa.

Tietojen siirto joko tekstitiedostosta tai toimituskentästä Survon datatiedostoon on vaivatonta. Edellisestä varten on FILE SAVE-komento ja tätä vielä yleisempänä Kimmo Vehkalahden DATAFIND-komento. Toimituskentässä olevan aineiston kopiointi tiedostoon onnistuu FILE COPY-komennolla.

AIVOT-aineistoa käsiteltäessä edellisessä jaksossa tämä jo tehtiinkin. Datatiedostoa AIVOT2 katsellaan tilapäisessä ikkunassa FILE SHOW-komennolla.

```

17 1 SURVO 84C EDITOR Wed Aug 07 08:54:02 1996 C:\KIRJA\ 200 100 0
51 *
52 *FILE COPY AIVOT TO AIVOT2
53 *FILE SHOW AIVOT2_
54 *

```

Kun rivin 53 FILE SHOW-komento aktivoidaan, tiedoston AIVOT2 sisältö tulee näkyville:

```

1 1 SURVO 84C EDITOR Wed Aug 07 08:54:02 1996 C:\KIRJA\ 200 100 0
File AIVOT2 N=20

```

|               |          |        | Ihminen |        | Aivot   |  |
|---------------|----------|--------|---------|--------|---------|--|
| 3 Laji        | Paino    | Aivot  | LogP    | LogA   | Jäännös |  |
| 1 Afr.norsu   | 6654.000 | 5712.0 | 8.803   | 8.650  | -0.175  |  |
| 2 Ent.norsu   | 2547.000 | 4603.0 | 7.843   | 8.434  | 0.341   |  |
| 3 Hevonen     | 521.000  | 655.0  | 6.256   | 6.485  | -0.397  |  |
| 4 Lehmä       | 465.000  | 423.0  | 6.142   | 6.047  | -0.748  |  |
| 5 Gorilla     | 207.000  | 406.0  | 5.333   | 6.006  | -0.172  |  |
| 6 Sika        | 192.000  | 180.0  | 5.257   | 5.193  | -0.927  |  |
| 7 Aasi        | 187.100  | 419.0  | 5.232   | 6.038  | -0.063  |  |
| 8 Jaguaari    | 100.000  | 157.0  | 4.605   | 5.056  | -0.567  |  |
| 9 Harmaahylje | 85.000   | 325.0  | 4.443   | 5.784  | 0.285   |  |
| 10 Ihminen    | 62.000   | 1320.0 | 4.127   | 7.185  | 1.927   |  |
| 11 Lammas     | 55.500   | 175.0  | 4.016   | 5.165  | -0.009  |  |
| 12 Simpanssi  | 52.160   | 440.0  | 3.954   | 6.087  | 0.961   |  |
| 13 Kenguru    | 35.000   | 56.0   | 3.555   | 4.025  | -0.797  |  |
| 14 Vuohi      | 27.660   | 115.0  | 3.320   | 4.745  | 0.102   |  |
| 15 Pesukarhu  | 4.288    | 39.2   | 1.456   | 3.669  | 0.448   |  |
| 16 Kettu      | 4.235    | 50.4   | 1.443   | 3.920  | 0.709   |  |
| 17 Kissa      | 3.300    | 25.6   | 1.194   | 3.243  | 0.222   |  |
| 18 Rotta      | 0.280    | 1.9    | -1.273  | 0.642  | -0.497  |  |
| 19 Hamsteri   | 0.120    | 1.0    | -2.120  | 0.000  | -0.493  |  |
| 20 Hiiri      | 0.023    | 0.4    | -3.772  | -0.916 | -0.149  |  |
| 21            |          |        |         |        |         |  |

```

To stop, press EXIT! (F1-HELP)

```

Survo pysyy tässä näyttömuodossa, kunnes painetaan poistumisnappia F8 (EXIT). Kohdistinta liikutellaan aivan kuin kuljettaisiin toimituskentässä. Nyt vain liikkumtila on suurempi eli koko aineiston kattava. Näkymää voi siis vierittää mm. nuolinapeilla kaikkiin suuntiin.

Tässä kohdistin on viety ihmisen aivojen painon kohdalle. Siitä on vahvistus kuvaruudun toisella rivillä, jonka loppupäässä näkyy havainnon tunnus (Ihminen) kohdistimen osoittama muuttuja (Aivot) ja muuttujan arvo (1320.0). Kun ensimmäinen muuttuja on sanallista muotoa, sitä käytetään havainnon tunnukseksi. Tuo tieto säilyy aina näkyvillä. Siis liikuttaessa miten tahansa, pysytään selvillä missä päin aineistoa ollaan.

Kun siis toimin FILE SHOW-tilassa, tunne on sama kuin liikkuisin rajattomassa toimituskentässä. Tässä tilassa voin korjailla tietoja ja kirjoittaa niitä lisää aivan kuten toimituskenttään. Samoin on mahdollista hakea tietoja määrätyn muuttujan arvojen mukaan siirtymällä ao. sarakkeeseen ja painamalla hakunappia Alt-F5. Tietoja eri näppäintoinnoista saan painamalla nappia F1, jolloin siirrytään väliaikaisesti Survon neuvontajärjestelmään.

- Survon datatiedostoilla on tiukka rakenne, joka tiivistää tiedot yleensä pienempään tilaan kuin tekstitiedostoissa. Kunkin datatiedoston rakennetiedot saadaan listatuksi toimituskenttään komennolla FILE STATUS:

```

19  1 SURVO 84C EDITOR Wed Aug 07 19:51:32 1996      C:\KIRJA\ 200 100 0
51 * .....
52 *FILE COPY AIVOT TO AIVOT2
53 *FILE SHOW AIVOT2
54 *FILE STATUS AIVOT2
55 * Copy of data matrix AIVOT
56 *FIELDS: (active)
57 *   1 SA   11 Laji
58 *   2 N1   8 Paino   (####.###)
59 *   3 N1   4 Aivot   (####.#)
60 *   4 N1   4 LoqP   (##.###)
61 *   5 N1   4 LoqA   (##.###)
62 *   6 N1   4 Jäännös (##.###)
63 *END
64 *SURVO 84C data file AIVOT2: record=63 bytes, M1=11 L=64 M=6 N=20
65 *

```

Tässä ovat mainitun komennon antamat tulokset riveillä 55-64. Riviltä 57 alkaen muuttujaluettelossa kerrotaan muuttujien tyypit ja nimet. Esim. 1. muuttuja (Laji) on merkkijono (S) ja sille on varattu tilaa 11 tavua (merkkiä). 2. muuttuja (Paino) on vuorostaan numeerinen (N) ja sille varattu tila, 8 tavua, osoittaa, että arvot talletetaan ns. kaksoistarkkuudella. Tämä tietää 15-16 merkitsevää numeroa. Muut kaikki ovat 4 tavun numeerisia muuttujia eli talletettuina 6-7 merkitsevällä numerolla. Survo tuntee lisäksi 1 ja 2 tavun tarkkuudet ja näitä käytetään kokonaislukuarvoisille muuttujille. Näitä ei tässä aineistossa esiinny lainkaan.

Numeerisista muuttujista on nimen perässä kuvaus sopivasta painoasusta, esim. (##.###). Näitä Survo hyödynsi jo äskeisessä FILE SHOW-näytössä.

Aivan viimeinen rivi (64) kertoo tiedoston yleisrakenteesta. Yhtä havaintoyksikköä varten on varattu 63 tavua korkeintaan 11 muuttujalle (M1=11), mutta näistä on käytössä vasta 6 (M=6). Muuttujan nimeä varten selityksineen on varattu 64 tavua (L=64), mistä 8 ensimmäistä tarkoittaa muuttujan "kutsumanimeä". Havaintojen lukumäärä on 20 (N=20).

Kun aineisto kopioidaan uuteen tiedostoon (FILE COPY), tilaa varataan muuttujille lisää hieman yli 40%. Tästä johtuu, että esim. suurin mahdollinen muuttujamäärä tiedostossa AIVOT2 on 11, vaikka vain 6 on käytössä. Lisätilaa tarvitaan, sillä muunnokset ja useat Survon tilasto-operaatiot synnyttävät uusia muuttujia. Jos muuttujatila käy ahtaaksi, on hyvin helppo kopioida tiedosto uudelleen, jolloin taas syntyy 40%:n lisäys. Rajojen tiukkuus seuraa siitä, että datatiedostot ovat Survossa suorasaantitiedostoja. Tämä tarkoittaa mm. sitä, että jos tietää muuttujan ja havainnon numeron, ohjelma pystyy suoraan laskemaan kyseisen tiedon paikan tiedoston sisällä, jolloin se löytyy nopeasti ilman peräkkäishakua tai vastaavaa.

Huomaa erityisesti, että kaikki rakennetiedot Survo on päättänyt automaattisesti silloin, kun toimituskentässä ollut havaintotaulukko AIVOT kopioitiin datatiedostoksi AIVOT2 komennolla FILE COPY. Näin tapahtuu myös, kun tekstitiedostoja siirretään Survoon datatiedostoiksi FILE SAVE-komennolla. Survo tutkii mainittujen komentojen yhteydessä aineiston rakenteen, tunnistaa sanalliset (merkkijonomuotoiset) ja numeeriset muuttujat sekä päättelee sopivan tarkkuuden (1, 2, 4 tai 8) ja ulkoisen esityksen kullekin muuttujalle.

Jos käyttäjä ei tyydy Survon valintoihin, hän voi määritellä oman rakenteensa (FILE CREATE) ja kopioida sitten tiedot sen mukaisesti.

## Suomen kunnat

- Tietojen hallinta- ja esittämistapoja kannattaa vielä tarkastella hieman laajemman aineiston avulla. Survo-esittelyjen peruskalustoon on kauan kuulunut Suomen kuntia kuvaava havaintotiedosto KUNNAT. Tiedot ovat vuosilta 1978-80 eli jo auttamattoman vanhentuneita. Tämä aineisto kummittelee kuitenkin niin monessa Survoon liittyvässä dokumentissa, esim. suomenkielisessä opeussarjassa, että ajantasaistamisesta olisi enemmän haittaa kuin hyötyä. Tilastolliselta kannalta aineisto on jatkuvasti säilyttänyt mielenkiintonsa.

Tiedoston rakenne on seuraava:

```

19 1 SURVO 84C EDITOR Thu Aug 08 12:40:20 1996 C:\KIRJA\ 200 100 0
1 *
2 *FILE STATUS KUNNAT
3 * Suomen kunnat aakkosjärjestyksessä
4 * Tiedot ovat pääosin vuosilta 1978-80. 5.2.84/SM
5 * SORT:Kunta
6 *FIELDS: (active)
7 * 1 SA- 16 Kunta Kunnan nimi
8 * 2 SA- 3 Lääni [UUS,TUR,AHV,HÄM,KYM,MIK,KAR,KUO,KES,VAA,OUL,LAP]
9 * 3 NA- 4 Väestö Arvioitu maassa asuva väestö 1.1.1980 (#####)
10 * 4 NA- 4 Synt. Elävänä syntyneet v.1978 (####)
11 * 5 NA- 8 Ala Maapinta-ala km^2 1.1.1979 (#####.##)
12 * 6 SA- 1 Maamet Maa- ja metsätaloudessa toimivien osuus (10%)
13 * 7 SA- 1 Teoll Teollisuudessa toimivien osuus (10%)
14 * 8 SA- 1 Palvelu Ammatissa ja palveluelinkeinoissa toim.osuus (10%)
15 * 9 NA- 4 Asuin Valmistuneet asuinhuoneistot v.1978 (####)
16 * 10 NA- 4 Äyri Veroäyri hinta v.1979 (##.##)
17 * 11 NA- 2 Tulotaso Veroäyriä asukasta kohti v.1979 (#####)
18 *END
19 *SURVO 84C data file KUNNAT: record=128 bytes, M1=30 L=64 M=11 N=464
20 *

```

FILE STATUS-komento on antanut riveille 3-19 rakennetiedot. Muuttujien yhteyteen on liitetty sanallisia selityksiä. Muuttujan Lääni kohdalla on hakusuluissa annettu sallitut arvot (lääniä nimien lyhenteet). Numeerisille muuttujille on ilmoitettu sopivat painoasut. Niille voisi aaltosuluissa antaa myös pienimmän ja suurimman sallitun arvon. Tällaiset vihjeet vähentävät tallennusvirheitä. Kun tiedoston sisältöä katsellaan ja täydennetään FILE SHOW-komennolla, alkunäkymä on:

```

1 1 SURVO 84C EDITOR Wed Aug 08 12:41:12 1996 C:\KIRJA\ 200 100 0
File KUNNAT N=464
Eckerö Ayri 14.50
3 Kunta Lää Väestö Synt Ala TP Asui Äyri Tulot
1 Alahärmä VAA 5358 78 348.23422 6417.0012711
2 Alajärvi VAA 8658 129 715.95333 10618.00 9099
3 Alastaro TUR 3686 39 260.92522 5515.0011583
4 Alavieska OUL 3036 34 250.44512 3517.00 9231
5 Alavus VAA 10359 120 813.96333 10617.5010743
6 Anjalankoski KYM 20307 229 721.60143 12516.0015325
7 Anttola MIK 1792 11 256.16512 1516.5011015
8 Artjärvi UUS 2039 14 180.70611 116.5010308
9 Asikkala HÄM 8249 93 566.60333 8415.5013347
10 Askainen TUR 825 10 58.30512 4616.5011415
11 Askola UUS 3930 64 209.40333 4116.0012473
12 Aura TUR 2445 31 95.05333 2315.0012574
13 Brändö AHV 551 3 99.67404 814.0012265
14 Dragsfjärd TUR 4670 48 243.20063 4015.5015652
15 Eckerö AHV 705 5 107.74215 1314.5013753
16 Elimäki KYM 8230 106 385.31333 8216.0014009
17 Eno KAR 8814 104 981.80233 9016.0011768
18 Enonkoski MIK 2078 16 318.59512 1116.00 9682
19 Enontekiö LAP 2259 29 8123.53215 3117.0010401
20 Espoo UUS1335562063 308.87027196414.5024665
21 Eura TUR 9621 113 426.35152 6815.5014990
To stop, press EXIT! (F1=HELP)

```

Hakunappi Alt-F5 on erityisen hyödyllinen suurien aineistojen selauksessa. Kun kohdistimen siirtää esim. nimisarakeen alkuun ja valitsee hakusanaksi \*järvi, saa helposti selville, että järvipäätteisiä kuntia oli 1980-luvun alussa

peräti 31 kpl. Voipa harrastaa hakua jopa äänen avulla niin, että esim. erisuuruiset arvot antavat erikorkuisia ääniä. Kun haluaa lepuuttaa silmiään tai etsiä aineistosta jaksollisuuksia, äänen vaihtelun avulla saattaa löytää kiinnostavia yksityiskohtia ja säännönmukaisuuksia.

- Uusia muuttujia tehdään VAR-komennolla niin kuin havaintotaulukoissakin. Syntyneisyys 1000 asukasta kohti (SYNT) lasketaan näin:

```
37 1 SURVO 84C EDITOR Thu Aug 08 13:44:13 1996 C:\KIRJA\ 200 100 0
22 *
23 *VAR SYNT=1000*Synt./Väestö TO KUNNAT_
24 *
```

Sekä FILE STATUS- että FILE SHOW-komennoilla näkee, että uusi muuttuja on ilmaantunut entisten jatkeeksi. Aineiston hallinnassa on kätevä käyttää myös FILE ACTIVATE-toimintoa. Se käynnistyy jopa pelkällä Alt-F6 -napilla. Tällöin Survon ikkunan valtaa tilapäisesti seuraava näkymä:

```
1 1 SURVO 84C EDITOR Thu Aug 08 13:45:11 1996 C:\KIRJA\ 200 100
Data file KUNNAT M= 12 Col= 1 M(active)= 12 M(protected)= 0
1 S A- Kunta Kunnan nimi
2 S A-N Lääni [UUS,TUR,AHV,HÄM,KYM,MIK,KAR,KUO,KES,VAA,OUL,LAP]
3 4 A-F Väestö Arvioitu maassa asuva väestö 1.1.1980 (#####)
4 4 A-F Synt. Elävänä syntyneet v.1978 (####)
5 8 A-R Ala Maapinta-ala km^2 1.1.1979 (#####.##)
6 S A-I Maamet Maa- ja metsätaloudessa toimivien osuus (10%)
7 S A-I Teoll Teollisuudessa toimivien osuus (10%)
8 S A-I Palvelu Ammatissa ja palveluelinkeinoissa toim.osuus (10%)
9 4 A-F Asuin Valmistuneet asuinhuoneistot v.1978 (####)
10 4 A-I Äyri Veroäyrin hinta v.1979 (##.##)
11 2 A-R Tulotaso Veroäyriä asukasta kohti v.1979 (#####)
12 4 A- SYNT ~1000*Synt./Väestö
Denote passive variables by - and active by A,X,Y etc. Stop by EXIT!
```

Uusi muuttuja SYNT on aikaisempien jatkeena vieläpä automaattisesti varustettuna lausekkeella, jolla se juuri laskettiin. Tässä muuttujien aktivointitilassa kohdistin liikkuu vain ns. maskisarakkeiden alueella (tässä tummalla pohjalla). Ensimmäinen sarake (nyt täytettynä pelkillä A-kirjaimilla) on yleinen aktivointisarake. Toisen avulla muuttujan arvot voi suojata tahattomilta muutoksilta (kirjaimella P). Kolmas sarake kertoo muuttujan mitta-asteikon laadun. Esim. N tarkoittaa nominaaliasteikkoa, R suhdeasteikkoa ja F frekvenssiä. Asteikkomerkitöjä ei ole pakko antaa, mutta niistä on hyötyä, sillä useat Survon tilastolliset ohjelmat tarkkailevat asteikkomerkitöjä. Jos vaikka yrittää laskea keskiarvoja nominaaliasteikollisesta muuttujasta, mikä on mieletöntä, Survo joko varoittaa tästä tai jopa kieltäytyy tekemästä.

Maskisarakkeita saa helposti käyttöön useampiakin ja niiden avulla osoitetaan väliaikaisesti tai pysyvästi aineistoon liitettynä erilaisia muuttujavalintoja. Useimmiten riittää käyttää valintoihin ensimmäistä maskisaraketta.

```

1 1 SURVO 84C EDITOR Thu Aug 08 13:45:11 1996 C:\KIRJA\ 200 100
Data file KUNNAT M= 12 Col= 1 M(active)= 12 M(protected)= 0
1 S --- Kunta Kunnan nimi
2 S --N Lääni [UUS, TUR, AHV, HÄM, KYM, MIK, KAR, KUO, KES, VAA, OUL, LAP]
3 4 --F Väestö Arvioitu maassa asuva väestö 1.1.1980 (#####)
4 4 --F Synt. Elävänä syntyneet v.1978 (####)
5 8 --R Ala Maapinta-ala km^2 1.1.1979 (#####.##)
6 S X-I Maamet Maa- ja metsätaloudessa toimivien osuus (10%)
7 S --I Teoll Teollisuudessa toimivien osuus (10%)
8 S --I Palvelu Ammatissa ja palveluelinkeinoissa toim.osuus (10%)
9 4 --F Asuin Valmistuneet asuinhuoneistot v.1978 (####)
10 4 X-I Äyri Veroäyrin hinta v.1979 (##.###)
11 2 --R Tulotaso Veroäyrejä asukasta kohti v.1979 (#####)
12 4 Y- SYNT ~1000*Synt./Väestö

Denote passive variables by - and active by A,X,Y etc. Stop by EXIT!

```

Laskeakseni regressiomallin, jossa selitettävänä muuttujana on uusi SYNT ja selittäjinä Maamet ja Äyri, olen merkinnyt selitettävän muuttujan Y:llä ja selittäjät X:llä. Muut muuttujat on merkitty passiivisiksi merkillä -.

- Tällöin lineaarisen regressioanalyysin tulokset saadaan yksinkertaisesti aktivoimalla LINREG-komento, joka viittaa aineistoon ja ensimmäiseen tulosriviin:

```

22 1 SURVO 84C EDITOR Thu Aug 08 14:37:51 1996 C:\KIRJA\ 200 100 0
44 *
45 *LINREG KUNNAT, CUR+1
46 *Means, std.devs and correlations of KUNNAT N=464
47 *Variable Mean Std.dev.
48 *Maamet 2.956897 1.847968
49 *Äyri 16.22511 0.975172
50 *SYNT 12.02225 3.459748
51 *Correlations:
52 * Maamet Äyri SYNT
53 * Maamet 1.0000 0.3088 -0.4378
54 * Äyri 0.3088 1.0000 0.1016
55 * SYNT -0.4378 0.1016 1.0000
56 *
57 *Linear regression analysis: Data KUNNAT, Regressand SYNT N=464
58 *Variable Reqr.coeff. Std.dev. t beta
59 *Maamet -0.970845 0.079205 -12.26 -0.519
60 *Äyri 0.928446 0.150095 6.186 0.262
61 *constant -0.171188 2.377533 -0.072
62 *Variance of regressand SYNT=11.96985915 df=463
63 *Residual variance=8.973323663 df=461
64 *R=0.5036 R^2=0.2536
65 *

```

Vaikka regressiokertoimet ovatkin merkitseviä, mallin selitysaste on keho (vain 25%), mutta koskapa ihmiskunnan sikiäminen on vallan järkiperaistä ollut.

Monesti havaintojen suhteen ehdolliset tarkastelut ovat hyödyllisiä. Ehdollistaminen tapahtuu Survon toiminnoissa esim. seuraavasti:

```

22 1 SURVO 84C EDITOR Thu Aug 08 14:52:25 1996 C:\KIRJA\ 200 100 0
67 *
68 *CASES=Lääni:UUS,TUR,KYM IND=Väestö,20000,500000
69 *VARS=Kunta,Väestö,Äyri,SYNT
70 *FILE LOAD KUNNAT,71
71 *DATA KUNNAT*,A,B,C
72 C Kunta Väestö Äyri SYNT
73 A Anjalankoski 20307 16.00 11.277
74 * Espoo 133556 14.50 15.447
75 * Helsinki 483057 15.00 12.112
76 * Hyvinkää 37204 15.00 14.165
77 * Imatra 36510 16.00 12.709
78 * Järvenpää 23129 15.50 16.170
79 * Kerava 23300 16.00 16.738
80 * Kotka 61035 16.00 11.715
81 * Kouvola 30906 15.00 14.593
82 * Kuusankoski 22504 15.00 11.998
83 * Lappeenranta 53610 16.50 12.479
84 * Nurmijärvi 21875 13.75 14.720
85 * Pori 79261 17.00 14.698
86 * Rauma 30593 16.00 17.651
87 * Turku 164081 15.00 14.200
88 * Tuusula 22036 14.75 14.794
89 B Vantaa 129669 15.00 16.866
90 *

```

Tässä on poimittu aineistosta ne kunnat, jotka kuuluvat johonkin lääneistä UUS,TUR ja KYM (CASES-täsmennys rivillä 68) ja joissa väkiluku ylittää arvon 20000 (IND samalla rivillä). Käytettävissä olisi vielä yleisempi SELECT-täsmennys, jonka avulla kaikki loogiset ehdot ovat kuvattavissa, mutta niin pitkälle ei tässä ole tarvetta mennä.

Poiminnasta vastaa FILE LOAD-komento, jonka tuloksena toimituskenttään on ilmaantunut havaintotaulukko KUNNAT\* (määriteltynä rivillä 71). Tätä taulukkoa voisi nyt lähteä käsittelemään omana itsenäisenä aineistonaan ja vaikkapa lajitella sen muuttujan SYNT mukaiseen laskevaan järjestykseen. Saman olisin voinut saada aikaan suoraankin lajittelemalla ensin koko KUNNAT-tiedoston, mikä tapahtuu komennolla

```
FILE SORT KUNNAT BY -SYNT TO KUNNAT2 .
```

Tällöin uudelleen järjestetty tiedosto olisi KUNNAT2. Tiedostojen lajittelussa on mahdollista käyttää asteittain useita lajitteluavaimia ja ottaa uuteen tiedostoon talteen havainnoista vain "parhaat päältä".

Kuten jo alussa totesin, tässä yhteydessä on ollut tyytyminen vain karkeaan läpileikkaukseen siitä, miten aineistoja Survossa hallitaan ja käsitellään. Jatkossa tulee vielä ilmi joitain muita aiheeseen liittyviä seikkoja.







## Keskustelu

- S.M.** Olen tässä parin päivän aikana jonkin verran lueskellut käsikirjoitustasi tai sen alkuosaa - kirjahan ei näytä vielä valmiilta. Suoraan sanoen, mihin oikein pyrit? Voitko noin vain uskotella, että Survo olisi parempi kuin yleisesti hyväksytyt *kansainväliset* ohjelmat?

**s.m.** Joo, täytyy myöntää, että tuntuu vähän taas siltä kuin olisin tässä altavastajan asemassa. Siihen olen totisesti saanut tottua, että Survoa usein hakataan kuin vierasta sikaa. Olenkin sanonut, että Survoa moni inhoaa Suomessa, koska se on suomalainen ohjelma ja ulkomailla sitä taas vierastetaan siitä syystä, että se on suomalainen ohjelma.

Mitä kirjoituksissani olen puhunut Survon eduista, vetää tuskin vertoja monien ohjelmien kaupallisten edustajien kehuskeluille. He puhuvat usein vastoin parempaa tietoa omaan pussiinsa, koska eiväthän he itse vastaa noista ohjelmista. Minulta tutkijana ja ohjelmani tekijänä sitä vastoin edellytetään ehdotonta rehellisyyttä. Niinpä väitteeni Survon suotuisista ominaisuuksista perustuvat asioihin, jotka on näytettävissä toteen. Olenhan usein kysellyt tuon tekstin varrella, miten tekisit vastaavat temput omalla mielihjelmallasi ja ...

**S.M.** Anteeksi, että keskeytän. Sinun on turha yrittää toistella argumentteja, jotka olen jo lukenut. Ne eivät kerta kaikkiaan vakuuta minua. En tosin ole koskaan Survoa käyttänyt, mutta niin usein kuulen puhuttavan sen vaikeudesta.

**s.m.** Niin, minkä Survon? Myönnän, että Survo on ollut "vaikea" joskus menneisyydessä, kuten ohjelmat yleensäkin, mutta se on luonut nahkansa aika monta kertaa vuosikymmenten aikana. Survolla on oma alansa nähden tavallista pitempi historiansa, mikä lienee raskas. Olisi kai aina välillä pitänyt vaihtaa nimeä ja puhua "uudesta, ennennäkemättömästä ratkaisusta". Ehdoton pyrkimyksemme on viime vuosina ollut helpottaa aloituskynnystä ja siinä olemme onnistuneet.

Tietoa Survon "vaikeudesta" levittävät kaikei sellaiset, jotka ammoin muinaisuudessa ovat Survoa kokeilleet eivätkä tiedä, missä nykyisin kuljetaan. Jou-

kossa on valitettavasti niitä, joilla asenteet ovat kerta kaikkiaan lukossa tai joiden kyky tajuta omilla aivoillaan on rajoittunut. He seuraavat vain sitä, mitä alan vaikuttajat sanovat. Näitä vaikuttajia löytyy erityisesti ATK:n näennäisosaajien joukosta - jopa eri organisaatioiden ATK-tukihenkilöistä.

**S.M.** Etkö sitten ole saanut noita vaikuttajia huomaamaan Survon "erinomaisuutta"?

**s.m.** Aina kun on luonteva tilaisuus, yritän sitä heille selittää. Kokemukset eivät vain ole rohkaisevia. Pari vuotta sitten asentaessani Survon verkkoversiota eräässä oppilaitoksessa tarvitsin verkosta vastaavan henkilön apua, jotta pääsin verkkoon sisälle. Hän seurasi asennusta vierestä ja kerroin hänelle innostuneesti, mistä on kysymys. Tapani mukaan koekäytin Survoa valmiilla sukroilla nähdäkseni, että ohjelma toimii kunnolla. Kuviteltuani, että Survo olisi tehnyt jonkinmoisen vaikutuksen, sain kuulla musiikillisen vertauksen. Hän sanoi jotain siihen tapaan, että eläähän niitä vielä sellaisia vanhempia ihmisiä, joille "Tapsa Rautavaaran Isoisän olkihattu on käypää musaa."

Varsinkin verkkoympäristöissä toimivat survoilijat kertovat usein, että teknisten ongelmien ilmaantuessa tukihenkilöt syyttävät aina ensin Survoa - aiheetta. Onpa paikkoja, jonne Survoa ei ole saanut asentaa edes omaan koneeseen mm. virustartuntavaaran takia. Joissain organisaatioissa on taas tehty virallinen päätös: "Käytämme vain Windows-ohjelmia!"

**S.M.** No, mutta eikös Survo olekin auttamattoman vanhentunut ohjelma, kun siitä ei ole edes Windows-versiota?

**s.m.** Totta, jos vanhanaikaisuus tarkoittaa pysyttelemistä ulkona hetken virtauksista. Kuten on tapana sanoa, graafiset käyttöliittymät ovat tulleet jäädäkseen ja monilla tehtäväalueilla ne voittavat entiset. Tämä ei kuitenkaan tarkoita sitä, että hiirellä ja valikkojen kautta **kaikki** työt tulisivat mukavammiksi. Editoriaalinen käyttöliittymä on Survon toimialaan kuuluvissa tehtävissä joustavampi. Mitä vaativammasta sovelluksesta on kysymys, sitä paremmin se pääsee oikeuksiinsa.

Minua on kiinnostanut lähes 40 vuoden ajan ihmisen ja tietokoneen välinen kanssakäyminen. Siihen sisältyy haasteita, joita emme aikaisemmin ole kohdanneet. Ihmiset ovat usein orjuuttaneet toisiaan ja tai muuten häikäilemättä käyttäneet hyväkseen. Tietokone on kuitenkin ihmisen orja ilman että kenenkään tarvitsisi kärsiä tunnonvaivoja. Tätä taustaa vasten on omituista, että tietojenkäsittely on saanut muotoja, joissa roolit vaihtuvat ja ihminen saattaa joutua koneen orjaksi.

**S.M.** Miten ihmeessä niin?

**s.m.** Tarkoitan tällä kaikkea sitä hössötystä ja riippuvuutta, johon nykyiset "tietoyhteiskunnan valtaapitävät" ovat meitä johdattamassa. Tietotekniikka on

edennyntä niin huimasti viime vuosikymmeninä, että periaatteellisia vaihtoehtoja ihmisen ja koneen väliseen kommunikointiin on runsaasti. Teknisen kehityksen nopeudesta johtuu, ettei esiin tulevien vaihtoehtojen välille pääse syntymään aitoa luonnonvalintaa. Vain kaupallinen kilpailu on rajua. Kehitys näet määräytyy pääoman ja alan vaikuttajapiirien ehdoilla ja mieltymyksillä. Myös luonnollisen valinnan ehtona olevalla sattumalla on siinä näppinsä pelissä - mutta tapahtumat vierivät liian nopeasti, jotta luonnollisen valinnan muut ehdot täytyisivät.

**S.M.** Etkö siis sallisi enää vapaata kilpailua ATK-alalla?

**s.m.** Turha kysymys, sillä kukapa sitä olisi estämään. On silti joskus vaikea välttää turhautumista. Jo monta vuotta sitten minulla oli tilaisuus esitellä Survoa merkittävän amerikkalaisen ohjelmistotalon johtajalle, joka teki pikavie-railun Helsinkiin. Jo kättelyssä - ennenkuin olin päässyt puhumaan Survosta sanaakaan - hän ilmoitti, ettei merkitse mitään, mikä on ohjelmatuotteen laatu, vaan se, miten hyvin sitä markkinoidaan. Noin tunnin kestäneen esittelyn aikana sain hänet oireellisesti tajuamaan, että Survossa saattaisi olla jotain kiinnostavaakin. Hän ilmoitti lähtiessään olevansa valmis järjestämään USA:ssa tapaamisen heidän asiantuntijoittensa kanssa. Kehotti varautumaan siihen, että 15 minuutissa minun on saatava heidät vakuuttamaan Survon mahdollisuuksista. Ehdotettua matkaa en koskaan tehnyt. Kyseinen ohjelmistotalo meni konkurssiin muutaman vuoden kuluttua.

Vastapainoksi ynseille suhtautumisille on aina ollut ilo kohdata ihmisiä, jotka hyvin nopeasti äkkäävät, mistä Survossa on kysymys. Näitä olen kohdannut sekä Suomessa että ulkomailla. Mieltäni on erityisesti lämmittänyt Osmo Soininvaaran lausuma sen jälkeen, kun olin ensimmäistä kertaa esitelty edito-riaalista käyttötappaa tilastotieteen laitoksen tutkijoille ja opettajille. Hän totesi viileästi: "Jos olisit elänyt keskiajalla, ne olisivat polttaneet sinut noitana!"

En ole yksin, kun väitän, että mainonnalla suosituiksi tyrkytetyt työtavat ovat monin tavoin epäonnistuneita. Ennen muuta niiden tuottamat hyödyt kustannuksiin nähden ovat liian pienet. Ne myös rajoittavat ihmisten valinnanvapautta eivätkä suosi omintakeisuutta eikä luovuutta.

Olen ollut sikäli onnellisessa asemassa, että olen voinut riittävän ajoissa hahmotella itselleni mieluisia työtapoja. Kun jo 1970-luvun puolivälissä pääsin kokeilemaan pientietokoneella vuorovaikutteista työskentelyä ja rakensin ensimmäistä "keskustelevaa" Survoa, minulle tultiin sanomaan, että "tuo on vain leikkiä tietokoneella". Väitettiin jopa, että tieteellisissä sovelluksissa vuorovaikutteisuus on "turhaa ylellisyyttä" eikä sitä tulla koskaan tarvitsemaan.

Koetin puolestani selittää, että vaikkei interaktiivisuuden aika vielä ollutkaan koittanut, olisi kyllin varhain opittava, miten koneen kanssa seurustellaan. Sitä taitoa ei kukaan keksi hetkessä.

Olen siis varsin pitkään päässyt kokeilemaan ja toteuttamaan ideoitani. Edelleen olen vapaa jatkamaan omalla tavallani riippumatta siitä, miltä trendit näyttävät.

**S.M.** Sikäli kun sain jotain selvää tuosta vuodatuksestasi, syytät siis johtavia alan vaikuttajia siitä, ettei kehitys ole mennyt oikeaan suuntaan - eli sinun suuntaasi. Vaikka aina löytyy jotain arvosteltavaa, eiköhän sentään valtaosa ihmisistä ole tyytyväisiä nykytilanteeseen. Eikö ole nähtävissä, että tietojenkäsittelyyn on syntymässä selvä yhtenäiskulttuuri? Tiedon valtatietä me kaikki kohta kuljemme.

**s.m.** Vaikka en ole sokea tehokkuusajattelulle, en ymmärrä, miksi pitäisi pyrkiä siihen, että olisi vain yksi tapa ajatella ja työskennellä. Kun luonnon monimuotoisuuden vaalimisesta ja harvinaisten eliölajien säilyttämisestä pidetään huolta, eikö tämän pitäisi koskea myös henkisiä asioita. Itse näkisin tietojenkäsittelyn tulevaisuuden mieluummin monimuotoisena. Tulee olemaan jatkuvasti tiettyjä valtakulttuureja massoille, mutta tilaa myös vähemmistökuultuureille.

Musiikissa, joka on lähellä sydäntäni, näin on tapahtunut. Kun sielläkin kaupallisuus on saanut ylliotteen luonnollisesta kehityksestä, pinnalla rehoittaa pop-musiikki lieveilmiöineen. Jää kuitenkin ihmettelemään, miksi ihmiset tyytyvät noin vähään. Esim. klassisen musiikin ystävät muodostavat valitettavasti yhä pienenevän vähemmistön.

On tyypillistä, että helpoimmin ihmisiin tehoavat "suuret" ilmiöt. Määrä korvaa laadun. Jopa kielemme on iskostunut pysyvästi "suurenmoisia" ilmaisuja. Klassisessa musiikissakin "suuret" teokset, ennen muuta oopperat, nauttivat suurinta suosiota. "Suurten" tenorien yhteiset konsertit ovat suuria mediatapahtumia. Kuitenkin pienikin voi olla "suurta". Domenico Scarlatti (1685-1757), joka syntyi samana vuonna kuin Bach ja Händel, jäi aikanaan monien muiden säveltäjien varjoon. Domenico Scarlatti ei luonut suurimuotoisia teoksia. Hänen tuotantonsa pääosa koostuu yli 500 lyhyestä yksiosaisesta cembalosonaatista. Tämän muodoltaan suppean mutta sisällöltään uskomattoman rikkaan tuotannon arvostus on noussut vasta viime vuosikymmeninä. Niinpä musiikin ystävät asettavat hänet nykyisin "suurten" mestareiden rinnalle. Monia hänen oman aikansa "suuria" säveltäjiä muistaa enää tuskin kukaan.

**S.M.** Hei, maltapa nyt! Tajuan jotenkuten innostuksesi musiikkiin, mutta eihän noita asioita mitenkään sovi rinnastaa puhtaasti tiedollisiin ilmiöihin?

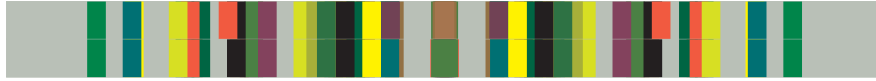
**s.m.** Miksi ei! Tunteehan tieteenkin historia neroja, joita oma aika ei koskaan ymmärtänyt ja joiden saavutukset on noteerattu vasta paljon myöhemmin. On mielestäni syytä uskoa, että tälläkin hetkellä keskuudessamme toimii tutkijoita, joiden saavutuksia vasta jälkipolvet arvostavat. "Tietoyhteiskuntamme" tulvii tietoa, josta vain murto-osa pääsee perille. Kun maailman kaikkina aikoina eläneistä tutkijoista yli puolet elää juuri tällä hetkellä, voi paremminkin epäillä tilanteen pahentuneen kuin parantuneen.

Haluamatta millään tavalla leimautua "väärin ymmärretyksi neroksi" minulla on sen verran itsetuntoa, että väitän tiedemaailmankin arvostusten olevan jälkeen jääneitä ja "paperinmakuisia". Nyt keskustellaan laajasti yliopistojen tu-

losvastuusta ja tieteellisen työn arvioinnista. Tilastotieteen edustajana joudun kohtaamaan mittaamisen ylivoimaiset vaikeudet moniulotteisissa ilmiöissä. Silloin tiedän, että on vastuutonta väittää, että jopa yksilötasolla joillain numeerisilla osoittimilla voitaisiin mitata tutkijan saavutuksia. Varsinkin puhuttaessa vapaasta perustutkimuksesta kaikki se työ, mitä maailman tutkijat tällä hetkellä tekevät, tarvitaan, että tuolta maaperältä aina silloin tällöin nousee jotain uutta, merkitsevää, ainutlaatuista. Mitä nousee ja milloin, sitä kukaan ei pysty ennakoimaan saaticka välittömästi mittaamaan.

En henkilökohtaiselta kannalta niin paljon piittaa siitä, miten työtäni arvostetaan. Minua ahdistaa enemmän huoli, etten ole kyennyt tekemään tiettäväksi, miten moni ihminen, tutkija, asiantuntija, suunnittelija, opettaja, toimistotyön tekijä, opiskelija saattaisi hyötyä Survon antamista mahdollisuuksista. Tästä syystä päätin kirjoittaa tämän Sepostuksen.





## Sattuman pyyntiä

- Tilastollisissa näytelmissä sattumalle on varattuna konnan rooli. Tilastollisten menetelmien, näytelmän sankareiden, tehtävä on estää tuon roiston tihutyöt, jotta totuus valkenee. Se ei milloinkaan täysin onnistu. Sattuman aseena on sen yllätyksellisyys - täsmälleen ei kukaan tiedä, missä muodossa se näyttäytyy. Silti seuratessa sattuman toimia kyllin kauan saattaa havaita, että sen käyttäytyminen noudattaa omia sääntöjään, jotka tulee tunnistaa. Sanotaan, että sattuma on sokea. Siksi sen voi panna koetteelle hännämällä sitä toistuvasti eli toimimalla kyllin järjestelmällisesti. Näin on toivoa saada se kiinni itse teosta. Sankarit saattavat myös viekkaasti varjostaa sattumaa värväämällä omia sattumiaan. Nämä panevat roistot ahtaalle.

Tuskin tuo peitekieli riittävästi valaisee tilastollista tutkimustoimintaa, joka jännittävimmillään tosiaankin muistuttaa salapoliisinäytelmää. Tutkimus perustuu malliajatteluun. Tilastolliset mallit eroavat täsmällisistä luonnonlaeista siinä, että niissä jätetään pelivaraa satunnaistekijöille. Malleihin liitetään ns. satunnaiskomponentteja. Jos malli on yksinkertainen ja satunnaistekijän käyttäytyminen on tuttua tyyppiä, satunnaisuuden vaikutukset tutkittavaan ilmiöön on laskettavissa suorilla kaavoilla. Mutkikkaammissa tilanteissa joudutaan turvautumaan simulointikokeisiin. Järjestetään laskennallisin keinoin toistuvasti "näytelmiä", joissa seurataan tarkkaan, miten sattuma vaikuttaa.

Jo aikaisemmin alkulukuihin perustuvaa "noppaa" heitettäessä totesin, kuinka vaikea ihmisen on luonnostaan hahmottaa satunnaisuutta. Tilastotieteen opeuksessa on siksi syytä eri tavoin yrittää esitellä sattuman käyttäytymismuotoja. Myös tutkimustyössä sattumalle löytyy paljon tekemistä mm. simulointikokeiden yhteydessä.

On ehkä hämmästyttävää, että sattumaa saattaa menestyksellisesti matkia tietokoneella, vaikka koneen toiminnassa ei saisi olla mitään sijaa epämääräisyyksille. Aitoa sattumaa - liekö sitä lainkaan olemassa - on kuitenkin mahdollista matkia hassuntuntuksilla laskennallisilla keinoilla. Tätä on harrastettu jo silloin, kun ensimmäiset koneet tulivat käyttöön.



## Onnistuuko kertomalla?

- ☐ *John von Neumann* (1903 - 1957), joka on tämän vuosisadan huomattavimpia matemaatikkoja, osallistui ensimmäisen tietokoneen (EDSAC vuonna 1949) suunnitteluun. Häneltä lienee peräisin seuraavanlainen tapa luoda satunnaisen kaltaisia lukuja:

```

13 1 SURVO 84C EDITOR Sat Aug 10 12:11:29 1996 C:\KIRJA\ 300 100 0
1 *
2 *          473299      Valitaan kaksi 6-numeroista lukua
3 *          210956 x   ja kerrotaan ne keskenään.
4 * 99845263844      Tuloksen keskeltä erotetaan 6 numeron jono.
5 *          210956      Otetaan toinen alkuperäisistä luvuista
6 *          845263 x   ja kerrotaan se tuolla uudella 6-numeroisella luvulla.
7 * 178513301428      Jälleen erotetaan tulosta 6 numeron jono keskeltä
8 *          845263      ja jatketaan kertomista
9 *          313301 x   käyttäen aina kahta viimeksi syntynyttä lukua...
10 * 264821743163
11 *

```

Olen tässä jäljitellyt von Neumannin tekniikkaa Survon kosketuslaskennalla eli käsin olen kirjoittanut vain kaksi ensimmäistä lukua (rivit 2 ja 3) sekä oikealle sijoitetut selitykset.

"Ohjelmoidakseni" tämän laskutavan ja saadakseni laajemman näytteen rakensin aluksi pienen laskentakaavion:

```

11 1 SURVO 84C EDITOR Sat Aug 10 12:11:59 1996 C:\KIRJA\ 300 100 0
12 * .....
13 *mod(x,y) :=int(y*(x/y-int(x/y))+0.5)   jakojäännös, kun x jaetaan y:llä
14 *
15 *VAR U TO K / IND=ORDER,3,11
16 *V=U[-1]*U[-2] U=mod(int(V/1000),1000000)
17 *
18 *DATA K,A,A+10,N,M
19 M 1111111
20 N U
21 A 473299
22 * 210956
23 * 845263
24 * 313301
25 * 821743
26 * 452903
27 * 169869
28 * 934179
29 * 688052
30 * 763729
31 * 485265
32 *

```

Tässä lukuja lasketaan muuttujan  $U$  arvoina rivin 15 VAR-komennolla aineistoon  $K$ , joka alkaa riviltä  $A$  (21). Siemeniksi olen valmiiksi sijoittanut kaksi ensimmäistä lukua. IND-täsmennys pitää huolta siitä, että laskenta alkaa vasta kolmannelta luvusta. Varsinainen laskusääntö ilmaistaan rivillä 16. Ensin las-

ketaan luku  $V$  kahden edellisen  $U$ -luvun tulona (esim.  $U[-1]$  tarkoittaa nykyistä  $U$ :ta edeltävää lukua aineistossa). Sitten tuon 6-numeroisen luvun poiminta tulon keskeltä tapahtuu jakamalla  $V$  ensin 1000:lla, ottamalla tästä luvusta kokonaisosa (int) sekä lopuksi saadun luvun jakojäännös, kun jaetaan miljoonalla. Koska VAR-ohjelma ei tunne jakojäännösfunktiota mod(), se on määritelty ns. väliaikaisena funktiona rivillä 13 (joskin kömpelösti).

Havaitaan, että näin todella syntyy samoja lukuja kuin aivan alussa kosketuslaskennalla, mutta toimituskentässä ei koskaan pitkälle pötkitä. Sen vuoksi olen uudistanut asetelman niin, että lukuja tehdään uuteen Survon havaintotiedostoon MULT:

```

16 1 SURVO 84C EDITOR Sat Aug 10 12:12:33 1996 C:\KIRJA\ 300 100 0
32 *
33 *FILE CREATE MULT,12,2
34 *FIELDS:
35 *1 N 8 U          (#####)
36 *2 N 4 X          (#.#####)
37 *END
38 *
39 *DATA H
40 *U          X
41 *434411 0.434411
42 *263457 0.263457
43 *
44 *FILE COPY H TO MULT
45 *
46 *FILE INIT MULT 24998
47 *
48 *mod(x,y) :=int(y*(x/y-int(x/y))+0.5)
49 *
50 *VAR U,X TO MULT- / IND=ORDER,3,25000
51 *V=U[-1]*U[-2]    U=mod(int(V/1000),1000000)
52 *X=U/1000000
53 *

```

Ensimmäinen on perustettu tiedosto MULT rivin 33 FILE CREATE-komennolla. Siinä määritellään kaksi muuttujaa  $U$  ja  $X$ , joista ensimmäinen vastaa aikaisempaa  $U$ :ta ja toinen tulee olemaan  $U$  jaettuna miljoonalla eli desimaaliluku välillä  $(0,1)$ .

Tiedosto MULT alustetaan taas kahdella ensimmäisellä havainnolla kopioidulla siihen pikkuaineisto H (rivit 39-42) rivin 44 FILE COPY-komennolla. Tiedostoa MULT jatketaan sitten 24998 tyhjällä havainnolla, jotta saataisiin kaiken kaikkiaan 25000 havaintoyksikköä (FILE INIT rivillä 46). Lopulta kaikki havainnot lasketaan täsmälleen samanlaisella VAR-komennolla kuin mitä käytettiin edellä.

Miksi laskin juuri 25000 lukua? Siksi, että tiesin aikaisemmasta kokemuksesta, että tämä menettely surkastuu tavallisesti hyvin nopeasti niin, että se ryhtyy tuottamaan pelkkiä nollia. Tämä nähdään katselemalla aineistoa FILE SHOW-komennolla ja etsimällä ensimmäistä nollaa.

```

3 1 SURVO 84C EDITOR Sat Aug 10 12:19:47 1996 C:\KIRJA\ 300 100
File MULT N=25000 U 211874
1 U X
24714 2118740.211874
24715 3095790.309579
24716 5917410.591741
24717 1905870.190587
24718 7781410.778141
24719 3035580.303558
24720 2109250.210925
24721 279710.027971
24722 8997830.899783
24723 1678300.167830
24724 105800.010580
24725 7756410.775641
24726 2062810.206281
24727 10.000001
24728 2060.000206
24729 00.000000
24730 00.000000
24731 00.000000
24732 00.000000
24733 00.000000
24734 00.000000
To stop, press EXIT! (F1=HELP)

```

Niin ikävästi käy, että tarina päättyy havaintoon nro 24728 ja sen jälkeen on turha jatkaa. Tulo  $775641 * 206281 = 160000001121$  (laskin sen juuri editoriaalisesti) "sattuu" sisältämään keskellään niin paljon nollia, että se tuottaa pelkän ykkösen. Aloitekykyinen lukija keksii ehkä nopeasti, miten lukujonon pelastaisi tuolta syöveriltä, mutta turha on yrittää. Tätä kautta ei satunnaisuutta kannata jäljittää.

Olen laatinut Survoon erityisen testiohjelman RNDTEST satunnaislukujen laadun arviointiin. Se tutkii, miten hyvin luvut ovat nollan ja ykkösen välillä tasaisesti jakautuneita. Tätä varten tiedostoon MULT laskettiin myös muuttuja X. On kehitetty monenlaisia koettimia lukujen satunnaisuudelle. Kymmenkunta erilaista tarkistinta on koottu yhteen RNDTEST-ohjelmassa. Tutkitaan, että jaettaessa väli (0,1) esim. 10 samanmittaiseen osaan, luvut osuvat näihin väleihin kyllin tasaisesti - mutta eivät liian tasaisesti, kuten tapahtui alkulukukokeessa. Tarkastellaan myös toistoksia eli mm. kuinka pitkiä nousevia ja laskevia jaksoja lukujonossa esiintyy. Täysin satunnaisista luvuista tiedetään, miten tuollaiset toistokset jakautuvat.

Kun Survon RNDTEST-komennolla käy läpi kaikki 24728 arvoa, ei synny mitään mainittavaa epäilystä siitä, etteivätkö juuri nuo luvut voisi olla satunnaisia. Vakavaan satunnaisuuden arviointiin tarvitaan kuitenkin huomattavasti pitempiä koesarjoja. Koska tällä kertolaskumenetelmällä on paha tapa surkastua ennen pitkää, se ei kelpaa tosikäyttöön.

## Luonnollista vai keinotekoista

Joissakin varhaisissa tietokoneissa oli erityinen satunnaislukugeneraattori, joka säteilyilmiön avulla loihdi satunnaisbittejä koneen rekistereihin. Vaikka uskoisi tällaisen menetelmän toimivuuteen, on todellisissa tutkimustilanteissa kuitenkin pakko jatkuvasti tarkkailla tuotoksen laatua. Se hidastaa toimintaa liikaa, sillä useissa sovelluksissa satunnaislukuja tarvitaan aivan tuhattomasti. Niinpä jo pitkään on tyydytty ns. *pseudosatunnaislukugeneraattoreihin*, jotka jäljittelevät sattumaa yksinkertaisilla laskutoimituksilla. Koska laskeminen tapahtuu kokonaisluvuilla ja on täysin tarkkaa, menettelyn etuna on, että samat luvut voidaan laskea uudelleen kuinka monta kertaa tahansa. Näin ko. generaattorin yleinen laatu ja mikä tahansa sen tuottamista lukujonoista saatetaan testata etukäteen. Näin myös taataan, että esim. satunnaistamiseen perustuvat tietokonesimuloinnit ovat toisten tutkijoiden toistettavissa ja tarkistettavissa. Mitään ei siis jätetä sattuman varaan!?

Ne, jotka eivät usko tuollaiseen valesatunnaisuuteen, käyttäkööt edelleen vanhoja hyväksi koettuja arpomisvälineitä, vaikkapa noppaa tai lanttia. Aina voi kuitenkin epäillä myös kolikon luotettavuutta. Äsken mainitsemani von Neumann on osoittanut, että vaikka raha olisikin harhainen niin, että se antaa kruunan todennäköisyydellä  $P$  ja klaavan todennäköisyydellä  $1 - P$  eikä  $P$ :tä edes tunneta, sitä on mahdollista käyttää reilun rahan kaltaisesti, jolla  $P = 1/2$ . Heittoja vain tarvitaan vaihteleva määrä, keskimäärin korkeintaan  $1/[P(1 - P)]$  kappaletta. Jätänpä lukijan kiusaksi päätellä, miten se onnistuu? Vihjeeksi annettakoon, että heittoja sopii tarkastella pareittain.

Nopat, kolikot eivätkä valmiina saatavat satunnaislukutaulukot kuitenkaan riitä vaativampiin sovelluksiin. On pakko käyttää laskennallisia keinoja. Valitettavasti on käynyt niin, että vuosikymmenten ajan esim. ohjelmointikielien yhteydessä tarjotut valmiit satunnaislukuohjelmat ovat usein osoittautuneet kelvottomiksi. Vasta viimeisten kymmenen vuoden aikana on asiaan osattu kiinnittää riittävästi huomiota.

Suosituimmat generaattorit perustuvat ns. kongruenssimenetelmään eli niissä luku  $U_n$  lasketaan edellisestä luvusta  $U_{n-1}$  muotoa

$$U_n = aU_{n-1} + c \pmod{m}$$

olevalla kaavalla, missä  $\text{mod } m$  tarkoittaa jakojäännöstä luvulla  $m$  jaettaessa. Generaattorin laatu riippuu ratkaisevasti siitä, miten vakiot  $a$ ,  $c$  ja  $m$  on valittu. Usein tässä on tehty kohtalokkaita erehdyksiä. Nykyisin on päädytty siihen, että  $c:n$  on parasta olla yksinkertaisesti nolla.

Vähimmäisvaatimukset täyttää esim. generaattori

$$U_n = 16807 U_{n-1} \pmod{2^{31} - 1}.$$

Tämä jaettuna modulillaan on Survossa on saatavilla funktiona urand(N). Tässä N on siemenluku, josta lähdetään liikkeelle. Tämä generaattori antaa täyden jakson eli  $2^{31} - 1$  eli noin  $2 \times 10^9$  erilaista lukua ennen kuin alkaa toistaa itseään. Kehittyneimmät generaattorit ovat usean generaattorin sekoituksia. Tällaisista Survossa on käytössä eräs ns. yhdistetty Tausworth-generaattori funktiona rand(N). Jakson pituus on peräti noin  $10^{18}$ . Kysely RAND? kertoo Survossa näistä asioista enemmän.

## Tasaista sattumaa

Edellä mainitut generaattorit tuottavat aina saman lukujonon, kun lähdetään liikkeelle annetusta siemenluvusta. Vaihtamalla siemenlukua saadaan aikaan erilaisia jonoja. Opetustilanteissa on mukava käyttää kolmatta generaattoria rnd(), joka perustuu Microsoftin C-ohjelmakirjastossa olevaan rand-funktioon. Sen takana on tietty kongruenssimenetelmä. Olen ohjelmoinut Survon rnd-generaattorin niin, että käyttämällä sitä argumentilla 0 siemenluku otetaan koneen sisäisestä kellosta. Silloin joka kerta ilmaantuu eri lukuja. Kun tätä juuri kirjoittaessa aktivoin lausekkeen rnd(0) toistuvasti, saan rnd(0)=0.1292 rnd(0)=0.4418 rnd(0)=0.6166 rnd(0)=0.9507 eli sama lauseke tuottaa eri arvoja. Jos joskus myöhemmin yritän samaa, on melkoista sattumaa, että täsmälleen nuo luvut putkahtaisivat esiin.

Survoilija voi aina tehdä itselleen arpomisvälineitä. Kirjoittamalla - kuten juuri teen - Noppa=int(6\*rnd(0)+1) käytössäni on noppa, jota heitetään aktiivimalla sitä tekstin keskellä toistuvasti eli

Noppa=5

Noppa=3

Noppa=1

Riittää kirjoittaa vain kerran Noppa=\_ ja kohdistimen pysyessä samalla paikalla painella uudestaan ja uudestaan ESC-nappia.

Jos tarvitsen kahta noppaa ja suoraan myös niiden pistelukujen summan, painelen yksinkertaisesti ESC-nappia toistuvasti kaaviossa:

```

8  1 SURVO 84C EDITOR Sun Aug 11 11:12:43 1996 C:\KIRJA\ 100 100 0
1  *
2  *Kahden nopan heittoa:
3  *Noppa1=int(6*rnd(0)+1) Noppa2=int(6*rnd(0)+1) Summa=Noppa1+Noppa2
4  *
5  *Noppa1.=4 Noppa2.=6
6  *Summa.=10 pistemäärien summa
7  *

```

Nyt on kyseessä ns. yhteisaktivointi eli olen merkinnyt aktivoitavat lausekkeet niin, että perässä on pelkän =-merkin asemasta .= . Tällöin aktivoidessani ESC-napilla yhdenkin näistä lausekkeista saan päivitetyksi kaikki muutkin eli tässä tapauksessa kohdistimen ollessa =-merkin perässä rivillä 6, saan aina uudelleen paitsi summan myös kummankin nopan pistemäärät toimituskenttään.

Perustamalla havaintotiedoston NOPAT ja aktivoimalla tässä samassa toimituskentässä komennon

```
VAR Noppa1,Noppa2,Summa TO NOPAT
```

tiedostoon kerääntyisi haluamani määrä kaksoisnopan tuloksia.

Kertoessani aikaisemmin alkulukunopan yhteydessä, millaisia ovat tyypilliset heittosarjat, mainitsemani esimerkki perustui seuraavaan kaavioon:

```
21 1 SURVO 84C EDITOR Sun Aug 11 11:26:49 1996 C:\KIRJA\ 100 100 0
1 *
2 *FILE CREATE NOPPA,1,1
3 *FIELDS:
4 *1 N 1 X
5 *END
6 *
7 *FILE INIT NOPPA,4800
8 *
9 *VAR X=int(1+6*rand(19964)) TO NOPPA
10 *
11 *GHISTO NOPPA,X,CUR+2 / X=0.5(1)6.5 FIT=Uniform(0.5,6.5)
12 * XSCALE=0.5:,1(1)6,6.5:
13 *Frequency distribution of X in NOPPA: N=4800
14 *
15 *Class midpoint f % Sum % e e f X^2
16 * 1.0 773 16.1 773 16.1 800.0 800.0 773 0.9
17 * 2.0 800 16.7 1573 32.8 800.0 800.0 800 0.0
18 * 3.0 768 16.0 2341 48.8 800.0 800.0 768 1.3
19 * 4.0 829 17.3 3170 66.0 800.0 800.0 829 1.1
20 * 5.0 794 16.5 3964 82.6 800.0 800.0 794 0.0
21 * 6.0 836 17.4 4800 100.0 800.0 800.0 836 1.6
22 *Mean=3.537292 Std.dev.=1.709978
23 *Fitted by UNIFORM(0.5,6.5) distribution
24 *Chi-square=4.908 df=5 P=0.4273
25 *
```

Koska  $\chi^2$ -testin  $P$ -arvo (rivillä 24) on lähellä puolikasta, tulosta voi pitää tyypillisenä eli eri pistelukujen frekvenssit (sarake f) kertovat, mikä on ominaista pitkän heittokokeen satunnaisuudelle.

## Normaalia sattumaa

Useimmiten satunnaisuustarkasteluissa kummittelee normaalijakauma, jossa vaihtelua kuvaa ns. Gaussin kellokäyrä. Normaalijakauman käyttökelpoisuus perustuu vankkaan "luonnonlakiin". Silloin, kun jotain ilmiötä kuvaava mittaluku koostuu useista yhteen kasautuvista pikkusatunnaisuuksista, sen jakauma alkaa muistuttaa normaalijakaumaa. Tämä tapahtuu mitä vaihtelevimmissa tilanteissa ja on sitä selvempää, mitä enemmän noita satunnaistekijöitä on ja mitä riippumattomampia ne ovat toisistaan. Näin ei kuitenkaan aina käy eli tulee olla tarkka, milloin normaalin sattuman tunnustaa omakseen.

Niinpä heitettäessä noppaa riittävän monta (vaikkapa 12) kertaa ja tarkkailtaessa pistelukujen summaa tuo summa alkaa olla jo hyvin normaalijakautuneen näköinen. Seuraava koe sen osoittaa.

```

21  1 SURVO 84C EDITOR Sun Aug 11 13:40:43 1996          C:\KIRJA\ 100 100 0
 7  *
 8  *N1=int(6*rand(1996)+1)
 9  *N2=int(6*rand(1996)+1)
10  *N3=int(6*rand(1996)+1)
11  *N4=int(6*rand(1996)+1)
12  *N5=int(6*rand(1996)+1)
13  *N6=int(6*rand(1996)+1)
14  *N7=int(6*rand(1996)+1)
15  *N8=int(6*rand(1996)+1)
16  *N9=int(6*rand(1996)+1)
17  *N10=int(6*rand(1996)+1)
18  *N11=int(6*rand(1996)+1)
19  *N12=int(6*rand(1996)+1)
20  *
21  *Summa=N1+N2+N3+N4+N5+N6+N7+N8+N9+N10+N11+N12
22  *
23  *FILE CREATE NOPAT12,1,1
24  *12 nopanheiton pistemäärä generaattorilla rand(1996)
25  *FIELDS:
26  *1 N 1 Summa
27  *END
28  *
29  *FILE INIT NOPAT12,1000
30  *
31  *VAR Summa TO NOPAT12
32  *
```

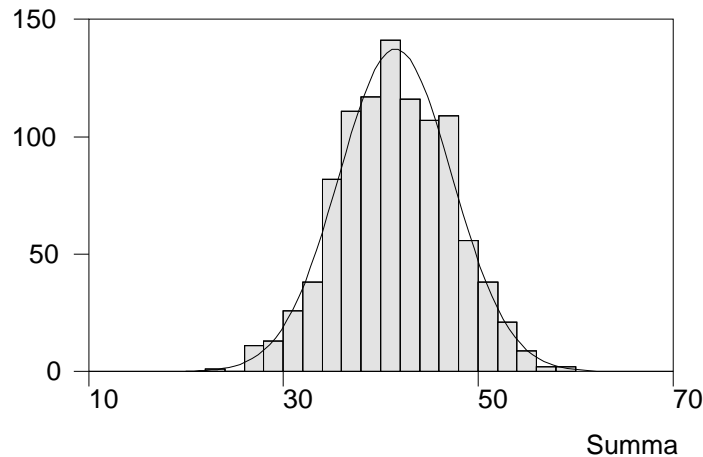
Riveillä 8-19 on määritelty 12 toisistaan riippumatonta noppaa ja rivillä 21 niiden pistelukujen summa. Tarvittava tiedosto on luotu (rivit 23-27) ja alustettu tuhanta tulosta varten (rivi 29). Noppaa on sitten heitetty 1000 kertaa aktivoimalla rivin 31 VAR-komento.

```

27 1 SURVO 84C EDITOR Sun Aug 11 13:45:30 1996 C:\KIRJA\ 100 100 0
32 *
33 *CHISTO NOPAT12, Summa, CUR+1 / Summa=20(2)60 FIT=Normal
34 *Frequency distribution of Summa in NOPAT12: N=1000
35 *
36 *Class midpoint f % Sum % e e f X^2
37 * 21 0 0.0 0 0.0 0.3
38 * 23 1 0.1 1 0.1 0.9
39 * 25 0 0.0 1 0.1 2.5
40 * 27 11 1.1 12 1.2 6.3 10.0 12 0.4
41 * 29 13 1.3 25 2.5 13.8 13.8 13 0.0
42 * 31 26 2.6 51 5.1 27.1 27.1 26 0.0
43 * 33 38 3.8 89 8.9 47.2 47.2 38 1.8
44 * 35 82 8.2 171 17.1 73.3 73.3 82 1.0
45 * 37 111 11.1 282 28.2 101.2 101.2 111 1.0
46 * 39 117 11.7 399 39.9 124.3 124.3 117 0.4
47 * 41 141 14.1 540 54.0 135.9 135.9 141 0.2
48 * 43 116 11.6 656 65.6 132.2 132.2 116 2.0
49 * 45 107 10.7 763 76.3 114.4 114.4 107 0.5
50 * 47 109 10.9 872 87.2 88.1 88.1 109 5.0
51 * 49 56 5.6 928 92.8 60.4 60.4 56 0.3
52 * 51 38 3.8 966 96.6 36.8 36.8 38 0.0
53 * 53 21 2.1 987 98.7 20.0 20.0 21 0.1
54 * 55 9 0.9 996 99.6 9.6 9.6 9 0.0
55 * 57 2 0.2 998 99.8 4.1
56 * 59 2 0.2 1000 100.0 1.6 5.7 4 0.5
57 *Mean=41.52600 Std.dev.=5.808212
58 *Fitted by NORMAL(41.526,33.735) distribution
59 *Chi-square=13.29 df=13 P=0.4257
60 *

```

Histogram of Summa in NOPAT12



Normaalijakauman sopivuutta on tarkkailtu piirtämällä histogrammi ja sovitamalla siihen normaalijakauma (rivi 33). Rivillä 57 luokitetusta aineistosta laskettu keskiarvo ja hajonta ovat uskottavia, sillä teoreettiset vastaavat tunnusluvut ovat tässä tapauksessa 42 ja  $\sqrt{35}=5.916$ . Mitään epäilystä jakauman normaalisuudesta ei synny, sillä  $\chi^2$ -testin  $P$ -arvo on 0.43 - ei siitäkään huolimatta, että summan jakauman vaihteluväli on tiukka [12,72] eikä äärettömyyksiin levittäytyvä kuten aidossa normaalijakaumassa. Tosin hajonnan ollessa noin 6, vaihtelua riittää  $(42-12)/6=5$  hajonnan mittaa kumpaankin suun-



taan. Näin poikkeavat arvot ovat normaalijakaumassakin erittäin harvinaisia - todennäköisyydeltään luokkaa  $6 \times 10^{-7}$ .

- ▣ Normaalijakauman simulointiin ei toki tarvita näin raskasta menettelyä. Yleinen tapa sellaisen jakauman generointiin, jonka kertymäfunktio on  $F(x)$ , on laskea lukuja  $F^{-1}(U)$ , missä  $U$  jakautuu tasaisesti välillä  $(0,1)$ . Eli soveltamalla kertymäfunktion käänteisfunktioita tasaisesti jakautuneeseen satunnaislukuun luodaan jakauman  $F$  mukaisia satunnaisarvoja. Tämän kertymäfunktion laskeminen hyvin tarkasti normaalijakaumalla ei ole kovin nopeaa. Sen vuoksi normaalijakaumaa todella raskaissa simulointikokeissa generoidaan toisella tavalla. Pieniin kokeiluihin Survossa riittää likiarvoinen probit-funktio eli esim. `probit (rnd (0))` antaa koneen kelloon sidottuja, standardoitua normaalijakaumaa likimain noudattavia lukuja. Kiitos lankeaa *Timo Patovaaralle* siitä, että hän aikoinaan ohjelmoi Survon C-kirjastoon tärkeimpien tilastollisten jakaumien perusfunktiot. Niiden kautta Survossa saatetaan luoda eri jakaumien mukaista tuotantoa.

Jos halutaan matkia sellaisen ihmispopulaation pituuden jakaumaa, jossa keskiarvo on 175 cm ja hajonta 6 cm olettaen, että pituus noudattaa normaalijakaumaa, lausekkeen

```
int (6*probit (rnd (0)) +175.5)=177
```

toistuva aktivointi toimituskentässä antaa jatkuvasti uusia senteiksi pyöristettyjä satunnaisarvoja tuosta jakaumasta.

Ottamalla myös paino mukaan (keskiarvolla 72 kg ja hajonnalla 5 kg) ja olettamalla, että pituuden ja painon välillä on suhteellisen korkea korrelaatiokerroin 0.82, saadaan aikaan jo kiintoisampi koetilanne. 30 havainnon otos ehdot täyttävästä 2-ulotteisesta normaalijakaumasta kehittyä kaaviolla:

```
25 1 SURVO 84C EDITOR Sun Aug 11 16:27:48 1996 C:\KIRJA\ 100 100 0
1 *
2 * Pituuden ja painon arvonta:
3 * keskiarvo hajonta
4 * Pituus m1=175 cm s1=6
5 * Paino m2=72 kg s2=5
6 * Pituuden ja painon korrelaatiokerroin r=0.82
7 *
8 * V1=probit (rnd(0)) V2=probit (rnd(0))
9 * Pituus=int (s1*V1+m1+0.5) int () ottaa lausekkeen kokonaisuosan
10 * Paino=int (s2*(r*V1+sqrt (1-r*r)*V2)+m2+0.5)
11 *
12 *VAR Pituus,Paino TO OTOS_
13 *DATA OTOS,A,A+29,N,M
14 M 111 111
15 N Pituus Paino
16 A
```

Muuttujat luodaan kahden  $(0,1)$ -normaalisen muuttujan  $V_1$  ja  $V_2$  avulla. Pituus lasketaan tässä aivan kuin edellä käyttäen muuttujaa  $V_1$ . Jos paino las-

kettaiisiin vastaavalla tavalla muuttujasta  $V_2$ , pituus ja paino jäisivät toisistaan riippumattomiksi. Oikeanlaisen riippuvuuden takaamiseksi painoa muodostettaessa (rivi 10) muuttujia  $V_1$  ja  $V_2$  sekoitetaan sopivassa suhteessa. Suhteen määrää korrelaatiokerroin  $r$ . Jos  $r = 0$ , paino riippuisi vain muuttujasta  $V_2$ . Jos taas  $r = 1$ , paino riippuisi vain muuttujasta  $V_1$  ja olisi siten puhtaan lineaarisesti riippuvainen pituudesta. Nyt valittiin  $r = 0.82$  ja rivin 12 VAR-komento antaa esim. seuraavanlaisen otoksen:

```

25 1 SURVO 84C EDITOR Sun Aug 11 16:28:11 1996 C:\KIRJA\ 100 100 0
11 *
12 *VAR Pituus,Paino TO OTOS / Toistuva aktivointi antaa uusia otoksia!
13 *DATA OTOS,A,A+29,N,M
14 M 111 111
15 N Pituus Paino
16 A 173 68
17 * 173 69
18 * 178 77
19 * 181 77
20 * 183 73
21 * 171 67
22 * 173 69
23 * 169 63
24 * 176 71
25 * 168 68
26 * 175 69
27 * 174 69
28 * 171 70
29 * 170 67
30 * 179 74
31 * 183 80
32 * 175 71
33 * 183 81
34 * 175 73
35 * 159 62
36 * 174 70
37 * 161 64
38 * 173 72
39 * 181 74
40 * 181 75
41 * 177 74
42 * 177 75
43 * 178 75
44 * 169 73
45 * 173 70
46 *
47 *CORR OTOS,CUR+1
48 *Means, std.devs and correlations of OTOS N=30
49 *Variable Mean Std.dev.
50 *Pituus 174.4333 5.829079
51 *Paino 71.33333 4.551115
52 *Correlations:
53 * Pituus Paino
54 * Pituus 1.0000 0.8626
55 * Paino 0.8626 1.0000
56 *

```

Lopuksi on tarkastettu, että otoksesta lasketut tunnusluvut täsmäävät kyllin hyvin odotettuihin arvoihin. Koska tässä aineistoa on luotu perimmältään  $\text{rnd}(0)$ -generaattorilla (rivi 8 kentän alussa), uusi VAR-aktivointi tuottaisi aina uuden otoksen. Tallentamalla peräkkäisistä näytteistä CORR-komennon tuloksia voitaisiin tutkia opetusmielessä esim. korrelaatiokertoimen tilastollista käyttäytymistä.

## Satunnaistetut testit

- Viittauksessa salapoliisinäytelmään totesin, että sattumaa voi käyttää aseena haitallista sattumaa vastaan. Näin tapahtuu tilastollisessa tutkimustyössä mm. silloin kun viljelykokeissa erilaiset käsittely- ja lannoitustavat arvotaan koealoittain. Simulointikokeilla, joissa satunnaisuuden olemusta jäljitellään keinotekoisesti, on ehkä nykyisin vielä suurempi merkitys.

Tilastollisessa päättelyssä käytetään testisuureita, joiden avulla pyritään toteamaan, missä määrin tutkittavaa ilmiötä koskevat hypoteesit pitävät paikkansa. Useimmiten päättely perustuu ilmiötä kuvaaviin otoksiin. Jos esim. halutaan tutkia, onko kahdessa eri ihmispopulaatiossa pituuden keskiarvo sama, verrataan näistä populaatioista umpimähkään poimituista otoksista laskettuja pituuskeskiarvoja toisiinsa. Jos pituuden jakauma voidaan olettaa edes lähes normaaliksi, tavanomainen keino on soveltaa ns. *t*-testiä.

Tämän testin periaatteen ymmärtäminen on helppoa, mutta miksi se lasketaan niin kuin se lasketaan ja miksi se noudattaa nollahypoteesin pätiessä (eli todellisten keskiarvojen ollessa samat) ns. *t*-jakaumaa? Kysymykseen vastaaminen vaatii melko perusteellisia tilastotieteen opintoja. Useat tutkijat tekevät näitä testejä mekaanisesti ymmärtämättä niiden edellytyksiä.

Kehittynyt tietotekniikka tuo ulottuville periaatteessa yksinkertaisempia keinoja, vaikka ne laskennallisesti saattavat olla huikeasti raskaampia. Voimme nykyisin soveltaa ns. satunnaistettuja testejä, joiden periaatteet, laskutavan ja toimivuuden ymmärtää hyvin vähillä tilastotieteen esitiedoilla.

Survoilija löytää tästä aiheesta lisätietoja suomenkielisestä opetussarjasta valinnalla

### A. Tilastolliset menetelmät

#### 7. Tilastolliset testit, pienten otosten vertailu.

Esitän tässä lyhennelmän ko. opetusjaksoista. Tarkoituksena ei ole selittää otosten vertailuun tarkoitettujen testien teoriaa. Kuvaan vain lyhyesti simulointiin perustuvaa riskitasojen määrittämistä. Tämä on tilastollisissa yleisohjelmissa toistaiseksi harvinainen piirre. Se ansaitsisi paljon suuremman huomion.

Kysymys on toisaalta testin satunnaistamisesta, jonka periaatteet tilastotieteen englantilainen uranuurtaja *R.A.Fisher* (1890 - 1962) esitti jo 1930-luvulla ja toisaalta simuloinnin käytöstä riskitason laskemisessa, joka muuten olisi kohuttoman vaivalloista. Satunnaistettujen testien merkitys on siis kyllä tunnet-

tu kauan, mutta niitä ei ole pystytty käyttämään laskennallisen työn suuruuden vuoksi. Tätä käyttöä koskevaa rajoitusta ei kuitenkaan enää pitäisi olla.

Tarkastelen seuraavassa riippumattomien otosten vertailua keskiarvojen perusteella ja näytän, miten mutkattomasti satunnaistamisperiaate toimii siinä tilanteessa. Teoria on paljon yksinkertaisempi kuin esim. tavanomaisella  $t$ -testillä, jota käytettäessä joudutaan lisäksi tekemään rajoittava normaaliajakau-  
maolettamus.

- ☐ Valaisen asiaa em. opetusjaksosta lainatulla pienellä esimerkillä. Vertailtavat aineistot on valittu aluksi tahallaan niin pieniksi, että on helppo laskea myös tarkat tulokset. Survossa pienten otosten vertailut tapahtuvat COMPARE-komennolla. Opetussarjassa on kuitenkin lisäksi käytettävissä VERTAA-niminen sukro, joka antaa helpot keinot vertailujen tekoon, vaikkei käyttäjällä olisi juuri mitään tietoa tilastollisesta testaamisesta. VERTAA esittää käyttäjälle muutamia perusolettamuksia koskevia kysymyksiä, soveltaa sitten kommentoa COMPARE, suomentaa tulokset ja kertoo lopuksi johtopäätökset sanallisesti.

```

1 1 SURVO 84C EDITOR Wed Aug 14 09:46:32 1996 C:\KIRJA\ 100 100 0
1 *
2 *Tarkastelun kohteena olkoot otokset
3 *DATA X: 5 6 7 END
4 *DATA Y: 5.5 6.5 8 9 END
5 *joista X-otoksessa on 3 havaintoa ja Y-otoksessa 4.
6 *Nollahypoteesimme on "Otokset ovat samasta perusjoukosta" ja
7 *vastahypoteesimme on yksisuuntainen "Y-perusjoukossa keskiarvo on
8 *suurempi kuin X-perusjoukossa".
9 *
10 *Vertailusukro kertoo tästä seuraavaa:
11 */VERTAA X,Y
12 *Olettamukset:
13 * Otokset X ja Y ovat riippumattomia.
14 * Mittaukset intervalliasteikollisia.
15 * Otokset eivät normaali jakaumasta.
16 *Nollahypoteesi: Kummankin otoksen havainnot saatu samasta jakaumasta.
17 *Vastahypoteesi:
18 * Otoksen Y jakauma poikkeaa ylöspäin (1-suuntainen testi).
19 *-----
20 *COMPARE X,Y,CUR+1 / SIMUMAX=0
21 *Riippumattomat otokset X Y
22 *Otoskoko 3 4
23 *Keskiarvo 6.000000 7.250000
24 *Hajonta 1.000000 1.554563
25 *t-testi: t=-1.203 df=5 (P=0.1414 1-suuntainen testi)
26 *Järjestyslukusumma (R) 9 19
27 *Mann-Whitney-testi (U) 3 9
28 *(P=0.1444 1-suuntainen testi, normaali jakauma-approksimaatio)
29 *Pienten otoskokojen johdosta testien kriittiset tasot myös simuloimalla:
30 *COMPARE X,Y,CUR+1 / SIMUMAX=5000
31 *Simuloidut testien kriittiset tasot:
32 * K.arvo R tai U
33 *Kriittinen taso 0.16940 0.19420 N=5000
34 *Keskivirhe 0.00530 0.00559
35 *-----
36 *Päätelmät:
37 * Johtopäätökset perustuvat keskiarvon satunnaistettuun testiin.
38 * Yksisuuntaisen testin kriittinen taso on 0.16940 .
39 * Hypoteesi "Otokset samasta jakaumasta" jää voimaan.
40 *_

```

Esimerkissä VERTAA-sukron kirjoittama teksti on lihavoituna ja sen käyttämän COMPARE-komennon tulokset harmaalla pohjalla. VERTAA esittää käyttäjälle aluksi eräitä kysymyksiä siitä, mitä etukäteen aineistoista tulee olettaa. Olettamuksista VERTAA on kirjoittanut tiivistelmän riveille 12-18. Se käynnistää COMPARE-komennon jopa kahdella tavalla tässä tapauksessa (rivit 20 ja 30) ja kirjoittaa lopulliset päätelmät (rivit 36-39).

VERTAA siis perustaa päätelmänsä keskiarvon satunnaistetun testin simuloi- tuun kriittiseen tasoon 0.16940. Miten tähän on tultu?

Asetamme testattavien aineistojen havainnot yhteen jonoon suuruusjärjestyk- sessä:

**5 5.5 6 6.5 7 8 9** (X-havainnot lihavoituna)

Jos nollahypoteesi pätee eli havainnot ovat samasta perusjoukosta, niin satun- naistamisperiaatteella toimittaessa ajatellaan, että X-havainnot on saatu poi- mimalla satunnaisesti 3 havaintoa näiden 7 havainnon joukosta. Nyt on saatu X-otos 5, 6, 7, mutta yhtä hyvin se olisi voinut olla esim. 5.5, 6, 9 .

Mahdollisia, keskenään yhtä todennäköisiä 3 havainnon otoksia 7 havainnosta on binomikertoimen  $C(7,3)=35$  osoittama määrä. Kun vastahypoteesina oli "Y-perusjoukon keskiarvo on suurempi kuin X-perusjoukon", niin on kiinnos- tavaa katsoa, moniko näistä 35 mahdollisesta X-otoksesta on sellainen, että sen keskiarvo on korkeintaan todellisuudessa havaittu  $(5+6+7)/3=6$  .

Koska otoskoko on vakio (3), riittää laskea havaintosummaa ja etsiä ne otok- set, joissa havaintojen summa on korkeintaan  $5+6+7=18$  . Näitä ovat:

| Otos      | Havaintosumma      |
|-----------|--------------------|
| 5 5.5 6   | 16.5               |
| 5 5.5 6.5 | 17                 |
| 5 5.5 7   | 17.5               |
| 5 6 6.5   | 17.5               |
| 5 6 7     | 18 (havaittu otos) |
| 5.5 6 6.5 | 18                 |

eli yhteensä 6 kappaletta 35:stä.

Siis todennäköisyys, että X-otokseksi saataisiin nollahypoteesin kannalta huo- nompi tai yhtä huono vaihtoehto kuin mitä on havaittu, on  $6/35=0.1714$  eli tä- mä on keskiarvon satunnaistetun testin todellinen kriittinen taso. Edellä saim- me simuloimalla sille arvion 0.16940, joka on kohtuullisen lähellä oikeaa ar- voa.

Simulointi tapahtui arpomalla 3 havainnon "koeotos" noista 7 havainnosta  $N=5000$  kertaa ja laskemalla niiden otosten suhteellinen frekvenssi, joiden havaintosumma on korkeintaan havaittu summa 18. Koska näin laskettu suhteellinen frekvenssi oli 0.16940, voimme todeta takaperin, että "kriittisiä koeotoksia" oli  $0.16940 \cdot 5000 = 847$  kpl.

Miksi uhrata näin paljon turhaa laskutyötä tulokseen, joka ei ole täysin tarkka ja joka saataisiin suoraan juuri tehdyllä lyhyellä päättelyllä?

Syy simuloinnin käyttöön on, että se toimii melko suurillakin otoskoilla. Suora kombinatorinen päättely sitä vastoin kangistuu.

Jos esim. molemmissa otoksissa olisi 20 havaintoa, kriittisen tason 0.05 alitaisi systemaattisessa vertailussa

$$0.05 \cdot C(40, 20) = 6892326441$$

eli lähes 7 miljardia "koeotosta". Nämä on tietenkin mahdollista käydä läpi tietokoneella, mutta se olisi **yli miljoona kertaa** raskaampaa kuin tehdä vain 5000 arvottua "koeotosta".

Näytän nyt, miten juuri 20+20 havainnon testauksessa 5000 "koeotoksen" simulointi riittää erittäin hyvin antamaan arvion todellisesta kriittisestä tasosta. Käytän suoraan COMPARE-operaatiota:

```

15 1 SURVO 84C EDITOR Wed Aug 14 11:15:35 1996 C:\KIRJA\ 100 100 0
41 *
42 *DATA A: 2 4 4 3 4 2 5 5 4 2 3 3 2 1 4 3 4 5 3 2 END
43 *DATA B: 3 3 3 5 5 3 4 5 3 3 5 5 5 4 2 5 5 2 4 4 END
44 *
45 *SIMUMAX=5000 (simulointikertojen määrä)
46 *COMPARE A,B,47
47 *Independent samples          A          B
48 *Sample size                  20          20
49 *Mean                         3.250000    3.900000
50 *Standard deviation          1.164158    1.071153
51 *Student's t=-1.838 df=38 (P=0.0370 one-sided test)
52 *Sum of ranks (R)             348          472
53 *Mann-Whitney (U)            138          262
54 *(P=0.0468 one-sided Mann-Whitney, normal approximation)
55 *Critical levels by simulation:
56 *      Mean      R or U
57 *Critical level 0.04580 0.04280 N=5000
58 *Standard error 0.00296 0.00286
59 *

```

Saimme tällä kertaa keskiarvon satunnaistetun testin kriittiseksi tasoksi simuloinnilla arvon  $P=0.04580$ , kun simulointikokeita tehtiin 5000 kappaletta. Tämän arvion keskivirhe (Standard error) on  $s=0.00296$ , joka kuvaa arvion tarkkuutta. Voimme esim. päätellä, että todellinen riskitaso on 95%:n varmuudella välillä, jonka alaraja on  $N.G(P, s^2, 0.025) = 0.0400$  ja yläraja vuorostaan  $N.G(P, s^2, 0.975) = 0.0516$ . Tässä N.G-funktio laskee Survossa normaalijakauman kertymäfunktion käänteisfunktion ja päättely perustuu siihen, että simuloituiden  $P$ -arvot noudattavat hyvin tarkkaan normaalijakaumaa. Jos joku ihmet-

telee arvion väljyyttä, on muistettava, että tarkan arvon laskemiseen menisi miljoona kertaa enemmän aikaa.

Toisaalta nykyisillä koneilla arviota pystyy nopeasti parantamaan, vaikkei sillä tässä esimerkissä liene mitään käytännön merkitystä. Esim. miljoona simulointia onnistuu hieman yli minuutissa ja tällöin päädytään arvoon  $P=0.04962$  keskivirheellä  $s=0.00022$ , jolloin  $P$ -arvon 95%:n luottamusväli kapenee ja on enää (0.0492, 0.0501). Täydellinen läpikäynti olisi vieläkin 5000 kertaa hitaampaa ja kestäisi yli sata tuntia.

Simuloinnin avulla saamme siis Fisherin hienon satunnaistamisperiaatteen toimimaan käytännössä. Keskiarvon satunnaistettu testi ei edellytä normaalisuutta tai muuta vastaavaa jakaumaoletusta. Havaintojen on vain oltava riippumattomia ja vähintään intervalliasteikollisia.

Jos havaintoarvot on mitattu vain järjestysasteikolla, on tyydyttävä Mann-Whitney-testiin, joka lasketaan aivan samalla tavalla käyttäen järjestysnumeroita alkuperäisten havaintoarvojen sijasta. COMPARE-operaatio laskee ja tulostaa rinnakkain molempien testien simulointitulokset. Ne ovat yleensä lähellä toisiaan, joskus jopa aivan samat.

Olen viime vuosien kuluessa liittänyt Survoon yhä useampia satunnaistamisperiaatteen varaan rakennettuja testejä. On mm. mahdollista tutkia frekvenssitaulukoissa muuttujien riippumattomuutta erilaisin reunajakaumiin liittyvin oletuksin TABTEST-komennolla ja korrelaatiokertoimien merkitsevyyttä jakaumasta riippumatta CORRTTEST-komennolla. Satunnaistettuja tekniikkoja on yleensä sovellettu vain pieniin aineistoihin, joissa ne yleensä parhaiten pääsevät oikeuksiinsa. Nyt näyttää siltä, että koneiden nopeuduttua vastaavia menettelyjä saatetaan soveltaa suurehkoilla aineistoilla ja jopa monimuuttujamenetelmissä.

### **Kevyttä ketjutusta**

- Edellä silmäiltiin sattuman olemusta numeerisesti ilmaistavissa tilanteissa. Survoon sisältyy muuan toiminto, joka pelaa sattuman kanssa myös mm. sanallisilla aineistoilla. Olen ohjelmoinut sen MARKOV-nimisenä jo vuonna 1988 eli 8 vuotta sitten. Se on ollut mukana siitä lähtien kaikissa Survon versioissa, mutta en ole kehdannut mainita tästä edes neuvontajärjestelmässä. Ensimmäisen kerran taisin kertoa julkisemmin tilastollisen tietojenkäsittelyn seminaarissa viime keväänä puhuessani Survon harvinaisista toiminnoista. Arvelen siis, etteivät useimmat survoilijat tiedä siitä mitään, vaikka voisivat koska tahansa sen käynnistää. Ehkä ei ole niin väliksikään, etteivät tiedä, sillä MARKOV ei uskoakseni kelpaa mihinkään vakaviin tehtäviin. Sillä lienee

merkitystä lähinnä opetuksessa, kun havainnollistetaan ns. Markovin ketjujen käyttäytymistä - niin ja sillä saattaa hieman hauskuttaa...

Nämä ketjut ovat saaneet nimensä tunnetun venäläisen matemaatikon *Andrei Markovin* (1856 - 1922) mukaan ja niiden teorialla on todennäköisyyslaskennassa suuri merkitys. Yleisesti tarkastellaan systeemiä, joka voi kerrallaan olla jossakin tiloista 1,2, ...,  $n$ . Systeemi saattaa siirtyä tilasta toiseen aina määrävällein annetulla todennäköisyydellä, joka riippuu sekä lähtötilasta että uudesta tilasta.

Seurataanpa erään mielikuvituksellisen olion käyttäytymistä "Elon laskuopin" mukaan. Oliolla olkoon kolme tilaa: se joko 1. nukkuu, 2. on hereillä ja liikkuu tai 3. syö. Seurataan sen elämää minuutin välein. Oletetaan esim., että nukkuessaan (tila 1) se herää seuraavan minuutin aikana todennäköisyydellä  $1/20$  (tilaan 2), mutta pysyy sikeästi edelleen unitilassa todennäköisyydellä  $19/20=0.95$ . Herättyään se alkaa liikkua ja etsii ruokaa. Ruokaa löytyy minuutin aikana (eli olio siirtyy tilaan 3) todennäköisyydellä  $1/3$ . Muuten se jatkaa liikkumistaan. Päästyään lihapatojen äärelle se jatkaa syömistään (tilassa 3) aina minuutin ajan todennäköisyydellä  $4/5$ , mutta vaipuu joko kylläisenä tai ruoan loputtua uneen todennäköisyydellä  $1/5$  herätäkseen taas joskus...

- ▣ Seuraava kaavio näyttää, kuinka Survolla matkitaan tuon olion käyttäytymistä 1000 minuutin aikana. Tässä vaiheessa ei vielä tarvita MARKOV-komentoa.

```

33 1 SURVO 84C EDITOR Thu Aug 15 09:25:10 1996 C:\KIRJA\ 150 100 0
1 *
2 *Olion käyttäytymistodennäköisyydet:
3 *      Nukkuu   Liikkuu   Syö
4 *1. Nukkuu      19/20    1/20    0
5 *2. Liikkuu      0        2/3    1/3
6 *3. Syö         1/5      0      4/5
7 *
8 *FILE CREATE OLIO,1,1
9 *
10 *FIELDS:
11 *1 N 1 T      1=Nukkuu, 2=Liikkuu, 3=Syö
12 *END
13 *
14 *FILE INIT OLIO,1000
15 *.....
16 *S=1996148
17 *VAR T=if (ORDER=1) then (1) else (T1)
18 *   T1=if (T[-1]=1) then (T11) else (T2)
19 *   T11=if (rand(S)<19/20) then (1) else (2)
20 *   T2=if (T[-1]=2) then (T22) else (T33)
21 *   T22=if (rand(S)<2/3) then (2) else (3)
22 *   T33=if (rand(S)<4/5) then (3) else (1)
23 *

```

Luodaan siis tiedosto OLIO, jossa tilaa varataan vain yhdelle muuttujalle T. Se tulee saamaan arvoja 1,2,3. Varaudutaan seuraamaan olion elämää 1000 minuutin ajan (rivi 14). Olion tarinan kertoo VAR-komento, jonka käyttäytymiseen vaikuttavat rivit 17-22. Oletetaan, että olio nukkuu alussa (ORDER=1)



eli tällöin  $T=1$ . Muulloin noudatetaan sääntöä T1 (rivi 18). Tällöin, jos edellisen minuutin aikana on nukuttu ( $T[-1]=1$ ), siirrytään sääntöön T11, muuten sääntöön T2 jne. Voit itse tarkistaa, että rivien 17-22 säännöt säätelevät olion eloa juuri edellä kuvaamallani tavalla. Kun siis VAR-komento aktivoidaan ESC-napilla, 1000 minuuttia olion kohtaloita tallentuu tiedostoon OLIO.

- Tapahtumat otetaan näkyville toimituskenttään yksinkertaisimmin FILE LOAD-komennolla:

```
21 1 SURVO 84C EDITOR Thu Aug 15 09:39:17 1996 C:\KIRJA\ 150 100 0
23 *
24 *FILE LOAD OLIO,CUR+1 / FORMAT=LIST
25 *DATA T: 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 3 3 3 1 1 1 1 1 1
26 *1 1 1 1 1 1 1 1 1 2 3 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
27 *2 2 2 3 3 3 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
28 *1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 3 3 1 1 1 1 1 1 2 2 2
29 *3 3 3 3 3 3 3 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
30 *1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 2 2 2 3 3 3
31 *3 3 1 2 2 3 3 3 3 3 3 3 3 1 1 1 1 1 1 1 1 2 2 3 3 3 3 1 1 1 1 1 1 1
32 *1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 3 3 3 3 3 3 3 3 1 1 1 1 1 1 1 1 1 1
33 *1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
34 *1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 2 2 3 1 1 1 1
35 *1 1 1 1 2 2 2 2 2 2 2 3 3 3 3 3 3 1 1 1 1 1 1 1 1 1 1 1 1 1 2 3 3 3
36 *3 3 3 3 3 3 3 3 3 3 3 3 3 3 1 1 2 2 2 2 2 3 3 1 1 1 1 1 1 2 3 3 3 3 3
37 *3 3 3 3 1 1 1 2 2 3 3 3 1 1 1 1 2 2 2 3 3 3 3 3 1 1 1 1 1 1 1 1 1 1 1
38 *2 3 3 3 3 3 3 3 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2
39 *2 3 1 1 2 3 3 1 1 2 2 2 2 2 2 2 2 2 3 3 3 1 1 1 1 1 1 1 2 3 3 3 3 3
40 *3 3 3 3 3 3 3 3 3 3 1 1 1 1 2 2 2 2 2 2 2 2 2 3 3 1 1 1 1 1 1 2 2 3 3
41 *3 3 3 1 1 1 2 2 2 3 3 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
42 *2 2 2 3 3 3 3 3 3 1 1 1 1 1 1 2 2 3 3 3 3 1 1 1 1 1 1 1 1 1 1 1 1 1
43 *1 2 2 2 3 3 3 1 1 1 1 1 2 2 2 2 2 2 2 2 3 3 3 3 3 3 3 3 3 3 3 1 1 1
44 *1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 2 3 3 3 3 1 1 1 1 1 1 1 1
45 *1 1 1 1 1 2 3 3 3 3 3 1 2 2 2 2 3 3 3 3 3 1 1 1 1 1 1 1 1 1 1 1 1 1 1
46 *1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
47 *1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 2 2 2 3 3 3 3
48 *3 3 3 3 3 3 3 3 1 1 1 1 1 1 1 1 1 1 1 1 2 3 3 1 1 1 1 1 1 1 1 1 1 1 1
49 *1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
50 *2 3 3 3 3 3 3 3 3 1 1 1 2 2 2 2 3 3 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
51 *1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 2 2 3 3 3 3 1 1 1 1 1 1 1 1
52 *1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 END
53 *
54 *STAT OLIO,CUR+1
55 *Basic statistics: OLIO N=1000
56 *Variable: T 1=Nukkuu, 2=Liikkuu, 3=Syö
57 *min=1 in obs.#1
58 *max=3 in obs.#24
59 *mean=1.503 stddev=0.796634 skewness=1.130649 kurtosis=-0.475777
60 *autocorrelation=0.8296
61 *lower_Q=1 median=1 upper_Q=2
62 *T f % *=16 obs.
63 * 1 689 68.9 *****
64 * 2 119 11.9 *****
65 * 3 192 19.2 *****
66 *
```

Elämäkerrasta ilmenee, että olio pysyy yleensä minuuttikaupalla unessa herätäkseen ruokaa etsimään ja syödäkseen lyhyemmän aikaa. Tätä saattoi ounnastella jo etukäteen. Herää kysymys, kuinka kauan suhteellisesti ajastaan olio viettää eri tiloissa. Summittaisen vastauksen on antanut STAT-komento.

Olion elo on nukkumista noin 70-, liikumista noin 10- ja syömistä noin 20-prosenttisesti. On näytettävissä, että teoreettisten todennäköisyyksien suh-

teet ovat tarkalleen 20:3:5, kun olion elämää seurataan äärettömän pitkään. Näinhän ei luonnossa tapahdu, vaan halutessamme tehdä todennäköisyysmal- lista realistisemman, siihen tulisi mm. liittää tila 4. "Kuolee". Siihen siirtymi- nen tulisi sitä todennäköisemmäksi, mitä kauemmin olio on ollut ilman ruo- kaa. Kaikenlaiset parannukset olisivatkin mahdollisia "oliokielellä", jonka Survossa esim. ehdolliset VAR-kaaviot tarjoavat, mutta sellaisiin en tässä ryhdy.

Nyt on sen sijaan aika vihdoin näyttää, mitä tuo MARKOV-komento tekee. Pyyhitään ensin olion elämäkerrasta (rivit 25-52) alusta ja lopusta kaikki sa- nalliset ilmaisut pois niin, että jäljelle jäävät pelkät tilanumerot. Kun MAR- KOV kohdistetaan näihin riveihin, se alkaa matkia rivien sisältöä hiukan eri- koisella tavalla, jonka informaatioteorian kehittäjä *Claude Shannon* (1916 - ) aikoinaan esitti Markovin ketjujen simulointiin.

```

16 1 SURVO 84C EDITOR Thu Aug 15 11:10:36 1996 C:\KIRJA\ 400 100 0
66 *
67 *MARKOV 25,52,68 / DEGREE=2
68 *1 1 1 1 1 2 2 3 1 1 2 2 2 2 2 2 2 2 2 2 2 3 3 3 3 3 3 1 1 1 1 1 1
69 *2 3 3 3 3 3 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
70 *1 1 1 1 1 1 1 1 2 3 3 3 3 3 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
71 *1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 3 3 3 3 3 1 1 1 1 1 1 1 1 1 1
72 *1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 3 3 3 3 3 1 1 1 1
73 *1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 3 3 3
74 *3 3 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
75 *1 1 1 1 2 3 3 3 3 3 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
76 *1 1 1 1 1 1 1 1 1 1 1 2 3 3 3 3 3 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
77 *1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 3 3 3 3 3 1 1 1 1 1 1
78 *1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 3 3 3 3 1
79 *1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
80 *2 3 3 3 3 3 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
81 *1 1 1 1 1 1 1 1 2 3 3 3 3 3 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
82 *1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 3 3 3 3 3 3 3 3 3 3 3 1 1 1 1 1 1
83 *1 1 1 2 2 2 3 3 3 3 1 1 2 2 2 3 3 3 3 3 3 3 1 1 1 1 1 1 1 1 1 1
84 *1 1 1 1 1 1 1 1 2 2 2 2 3 3 3 3 3 3 1 1 1 1 1 1 1 2 2 2 2 2 3 3 3
85 *3 3 3 3 3 3 1 1 1 1 1 2 3 3 3 3 3 3 1 1 1 1 1 1 1 1 1 1 1 1 1 1
86 *1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 3 3 1 1 1 1 1 1 1 1 2 2 3 3 3 3 3
87 *3 3 3 3 3 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
88 *1 2 2 2 3 3 3 3 3 3 3 3 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

```

MARKOV-ohjelma ei tiedä mitään siirtymätodennäköisyyksistä vaan matkii alkuperäisen lukujonon käyttäytymistä hyvin yksinkertaisella periaatteella. Riviltä 68 eteenpäin näkyy MARKOVin tuotosta, mikä silmämäärin on samana- näköistä kuin alkuperäinen, mutta ei suinkaan identtistä. MARKOV ottaa kaikki lähteaineiston merkit tosissaan - siis myös välilyönnit, jotka erottavat numeroita. Niinpä tässä tapauksessa kaikki kahden merkin (DEGREE=2) yh- delmät tulkitaan lähtötiloiksi. Nämä ovat 1S, 2S, 3S, S1, S2, S3, missä S:llä olen selvyuden vuoksi merkinnyt välilyöntiä.

Matkiminen alkaa katsomalla, mitkä ovat alkuperäisessä aineistossa kaksi en- simmäistä merkkiä. Ne ovat 1 ja välilyönti. Alkutila on siis 1S. Shannonin idea oli, että tästä eteenpäin lähdetään etsimään seuraavaa kohtaa, jossa esiin- tyy 1S esiintyy ja katsotaan sitten, mitä seuraa ja otetaan se uudeksi lähtöti- laksi. MARKOV poikkeaa käytökseltään hieman tästä. Se harppaa tuosta läh-

tökohtasta umpimähkään eteenpäin korkeintaan 100 merkkiä ja lähtee etsimään lähtötilan mukaista merkkijonoa vasta siitä alkaen. Näin vältetään tämäntapaisissa aineistoissa liiallista jäljittelyä.

Jos siis ensimmäisen harppauksen jälkeisen ensimmäisen 1S-yhdistelmän jälkeen olisi merkki 2, uusi lähtötila olisi S2. Siitä harpattaisiin jälleen eteenpäin etsimään merkkijonoa S2 ja katsottaisiin sen perässä oleva merkki jne.

MARKOV tulostaa näkyville saavutetut tilat ja lähdeaineiston loppuessa kierretään aina uudelleen alkuun. Näin jäljiteltä aineistoa syntyy loputtomiin ja voidaan osoittaa, että menettely vastaa siirtymätodennäköisyyksillä arpomista.

- Mitä tällä voitetaan? Näin menetellen ei tarvitse laskea tai arvioida etukäteen siirtymätodennäköisyyksiä. Se olisikin työlästä moniasteisilla Markovin ketjuilla. Seuraavia esimerkkejä ei kukaan jaksaisi rakentaa ilman Shannonin menetelmää.

```
18 1 SURVO 84C EDITOR Thu Aug 15 13:39:06 1996 C:\KIRJA\ 200 100 0
26 *Aune Helmi Karoliina Sirkka Sirja Hilma Veera Lotta Vappu Onerva Kielo
27 *Elisa Aliisa Lauha Mari Unelma Meri Lemmikki Irja Kaino Anni Pirita
28 *Sinikka Kaisu Venla Jonna Alma Amanda Milla Hilla Marjukka Henni Vieno
29 *Aura Eine Hilikka Eija Liisa Ulpu Irma Mirkka Impi Kirsi Heli Valpuri
30 *Linnea Henna Floora Virva Anu Maire Kristiina Virpi Sylvia Silva Lilja
31 *Raisa Ruut Alli Kaarina Emmi Milja Solja Tuomi Raija Heljä Aili Lea
32 *Inari Saima Ritva Liina Henriikka Mira Soila Pirjo Auri Johanna Suvi
33 *Päivi Inkeri Saara Auli Maili Armi Alina Emilia Raila Sointu Loviisa
34 *Sylvi Sirkku Helvi
35 *
36 *MARKOV 2,34,CUR+1 / DEGREE=2
37 *Helikka Sata Maija Helma Jonja Vieli Lota Tuulma Maija Soin Hanjaari Ina
38 *Siri Maro Eira Oilda Alma Maara Velmi Tea Vanni Sisterhikko Tiitta
39 *Reijukka Estiina Soha Mertta Maara Lilja Mari Maritvi Maara One Katju
40 *Auri Sora Marja Lovia Maakki Sinea Marmetta Leni Raija Lota Saija Elisa
41 *Jonja Marjo Sivi Leenli Jeervi Marjo Iira Marja Raila Rita Silda
42 *Tuijukki Estiina Salvi Airvokki Irmu Päiviitta Veen Ilmelli Marjo Iira
43 *Pirkki Pukka Maimma Henna Leni Ailma Varka Tuijukki Raija Soin Hanjaari
44 *Ella Emi Tyylla Elmu Sina Suvikka Helmikka Sanelja Villiisa Iina Anne
45 *Helvi Nienna Marju Alma Maara Velmi Tea Vanni Sisterhikko Tiitta
46 *Reijukka Estiina Soha Lauiirsi Leenli Jeli Ulpu Heleoha Elvi Hilli Siri
47 *Maro Eira Oilda Tuijukki Estiina Salvi Airvokki Irmu Päiviitta Veen
48 *Ilmelli Marjo Impi Sanelja Helmikku Helja Kasa Inketa Elvi Rauha Mili
```

Olen aikoinaan kopioinut nimipäiväkalenterista kaikkien tyttöjen nimet Survoon. Ne on tässä siirretty toimituskentän riveille 2-34 umpimähkäisessä epäjärjestyksessä. Elokuun 15. päivänä 1996 MARKOV tuotti tämän luettelon innostamana riviltä 37 alkavan uuden luettelon. Joukossa on tuttuja nimiä mutta enemmistö lienee "uusia". Jos tuon MARKOV-komennon aktivoi toistamiseen, luettelosta tulee aina erilainen. Tosin aika lailla samoja nimiä löytyy kaikista yriyksistä.

Lopuksi seuraa suppea näyte neliasteisen MARKOVin luomasta uudesta Kalevalasta:

Mielelleni keralla riihen  
 yhteheni virittelemahan  
 Yhteheni virittelemahan,  
 Saa kera saattelin  
 sommeloltamme käärin,  
 Vetoset ajoilla,  
 Katoi Louhi luoksi,  
 Risukoilla,  
 Metisillä mättähilleni  
 keralla riihen yhyttyämme;  
 Hampahiltamahani,  
 Puien leikkosehen yhymme,  
 Sovitteleni hajoovat,  
 Minun laihemmalta ahjon  
 Ilmarisen jousevassa,  
 Vilu mulle kuulun  
 kuulun kultaisien,  
 Ve'in kanssa noien luottehia.  
 Lajivirttä vanhan virret,

Pohjan laulamahani,  
 Puien leikkosehen lomahan,  
 Tienohesta,  
 Tietä mielen,  
 Veli runoja,  
 Kerittämiä  
 Vyöltä virteni minä kerältä,  
 Puhe'et putolevi  
 Lähteäni lauloaksemme iloksi,  
 Risukoilla.  
 Ne minun tekevi,  
 Kanervoin yhymme,  
 Sovitteleni sanoja  
 Eessänsä värttinätä,  
 Päästä kalkuttelevalan hyviä,  
 Kerän pistin  
 kanssa noien luottehia.  
 Lajivirttä vanhan viritän,  
 Aivoni lausehia:

Pyydän anteeksi kansanrunouden ystäviltä sitä, että tämmöistä puppua Survon Sampo sylkee solkenaan. Innoitus irtoaa, kuten lukija jo varmaan huomasi, Kalevalan ensimmäisen runon 102 säkeestä.





## Monen ulottuvilla

- Otsikko on harhaanjohtava, mutta sitä ovat usein ne asiat, joita nyt aion kuvata. Puhun moniulotteisuudesta tilastollisissa tutkimuksissa. Mainitsin jo aikaisemmin järjettömän innostuksen piirtää "kolmiulotteisia" pylviä ja piirakoita silloin, kun se on täysin tarpeetonta ja jopa hämäävää. Aidosti moniulotteisissa tilastollisissa ongelmissa pyrkimys on päinvastainen. Kun muuttujia voi esiintyä tutkimuksessa kymmenittäin, ilmiötä periaatteessa katsellaan hyvin moniulotteisessa kuvitteellisessa avaruudessa. Tällöin on toivottavaa, että ulotteisuusluku supistuisi joillain keinoin mahdollisimman pieneksi, jotta tilanne hahmotettaisiin selkeästi.

Ihmisillä on taipumus uskotella itselleen, että kaksi- tai kolmiulotteisesta maailmasta syntyneet mielikuvat ja ominaisuudet yleistyvät sellaisinaan korkeampiin dimensioihin siirryttäessä. Kuvaamme moniulotteisia asioita tasossa tai avaruudessa malleilla, joiden uskomme pätevän täysin yleisesti.

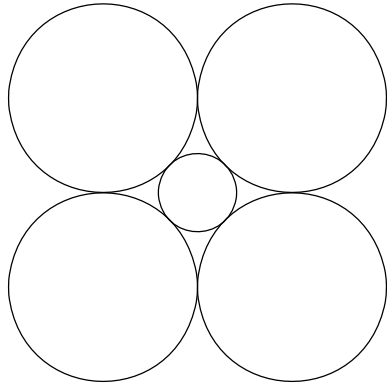
Itse asiassa jo hyppäys 2-ulotteisuudesta 3-ulotteisuuteen tuottaa yllätyksiä. Ainakaan minä en pysty suoraan aistimaan, että yleisissä peräkkäisissä koordinaatiston kierroissa vain kaksiulotteisessa tapauksessa ei ole väliä sillä missä järjestyksessä kierrot tehdään. "Avaruusmatkailu" on myös sikäli vaarallisempaa, että satunnaisliikkeessä, jossa kuljetaan aikayksikössä aina yksi askel umpimähkään jonkin koordinaattiakselin suuntaan, takaisinpaluu onnistuu 3-ulotteisessa avaruudessa vain todennäköisyydellä 0.35, kun vastaava todennäköisyys kahdessa ulottuvuudessa on lohdullisesti 1. Tosin paluu aikaan lähtökohtaan kummassakin tapauksessa kasvaa keskimäärin äärettömäksi. Nämä elämän tosiseikat on kerrottu mm. *William Fellerin* kirjassa "An Introduction to Probability Theory and Its Applications" (1967).

### Pallot moniulotteisina

Pallo on tunnetusti esine, jota ei voi sen enempää kaataa kuin nostaa pystyyn. Jos eläisimme 2-ulotteisessa maailmassa ja tajuaisimme vain sen olemuksen, uskoisimme samaa ympyrästä. Tiedämme kuitenkin, että vannetta pyöritettäessä se hyvin herkästi kaatuu. Näin ollen pallonkin pitäisi kaatua 4-ulottei-

seen avaruuteen. Jos siihen pystyisimme, rikkoisimme fysikaalisen todellisuutemme rajat.

- Palloihin liittyy seuraava paradoksin kaltainen esimerkki, jonka kuulin *A.N. Shiryaevilta* ja jota hänen kertomansa mukaan jo *A.N. Kolmogorov* (todennäköisyyslaskennan venäläinen uranuurtaja) oli käyttänyt luennoillaan. Tarkastellaan neljää yksikkösäteistä ympyrää, jotka sivuavat toisiaan oheisen kuvan mukaisesti ja joiden keskelle on piirretty kaikkia neljää sivuava (paljon pienempi) ympyrä.



On helppo todeta, että pienen ympyrän säde on  $\sqrt{2} - 1 \approx 0.4142$ . Kun vastaava rakennelma toteutetaan kolmiulotteisesti asettamalla 8 yksikkösäteistä palloa kasaan, jää keskelle tilaa pienemmälle pallolle, joka sivuaa kaikkia kahdeksaa. Tämän pallon säteeksi saadaan  $\sqrt{3} - 1 \approx 0.7321$ . Yleisesti on mahdollista päätellä, että vastaavassa  $n$ -ulotteisessa rakennelmassa pienen  $n$ -ulotteisen pallon säteeksi saataisiin  $\sqrt{n} - 1$ . Siis esim. kun  $n=4$ , "pieni" pallo olisi samankokoinen kuin sitä ympäröivät ja arvolla  $n=9$  sen säde olisi kaksinkertainen. Dimension  $n$  kasvaessa keskellä olevan pallon säde kasvaisi rajatta. Missä on vika?

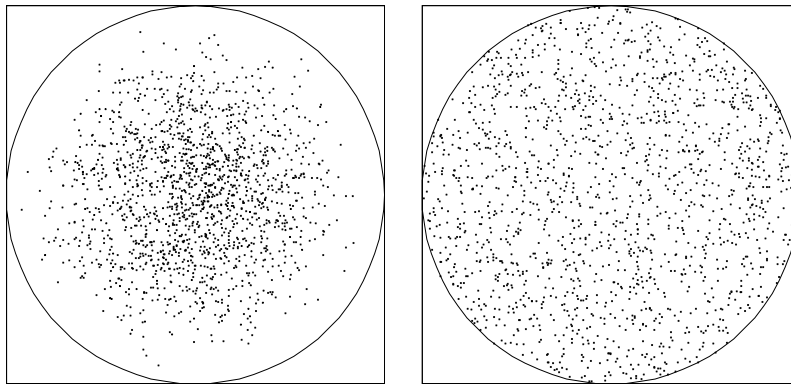
- Vikaa ei ole missään. Käsitksemme  $n$ -ulotteisen pallon olemuksesta vain on harhaanjohtava. Se on paljon "hennompi olento" kuin mitä saatamme kuvitella 2- ja 3-ulotteisten mielikuvien varassa. Ajatelkaamme  $n$ -ulotteista kuutiota, jonka sivun pituus on 2 ja sen sisälle asetettua mahdollisimman suurta  $n$ -ulotteista palloa, jonka säde on siis 1. Pallon suurin ulottuvuus on sen halkaisija  $=2$ , mutta kuution suurin ulottuvuus on sen päälävistäjä, jonka pituus on  $2\sqrt{n}$  (saadaan soveltamalla Pythagoraan lausetta  $n-1$  kertaa). Pallo täyttää kuution hyvin vajaasti ja näin jää toisiaan sivuavien pallojen väliin tilaa paljon enemmän kuin saatamme etukäteen kuvitella.

Kun  $n=2k$ ,  $n$ -ulotteisen,  $R$ -säteisen pallon mitta (tilavuus) on  $\pi^k R^{2k}/k!$  ja sen suhde ympäröivän kuution mittaan  $(2R)^{2k}$  lähestyy pikaisesti nollaa, kuten nähdään taulukosta

|     |         |
|-----|---------|
| $n$ | suhde   |
| 2   | 0.78540 |
| 4   | 0.30843 |
| 6   | 0.08075 |
| 8   | 0.01585 |
| 10  | 0.00249 |
| 12  | 0.00033 |

- ☐ Seuraava kuvapari näyttää tilastollisesti, miten pallon sisältö kasautuu sitä enemmän keskipisteen läheisyyteen, mitä suuremmaksi ulotteisuus  $n$  kasvaa. Olen arponut tasajakaumaperiaatteella 8-ulotteisen kuution sisältä 100000 pistettä ja valinnut ne, jotka ovat myös tämän kuution sisään asetetun 8-ulotteisen pallon sisällä. Näitä pisteitä löytyi 1635 kappaletta (odotettu arvo olisi taulukon mukaan 1585 keskivirheellä 39.5) eli uskottava määrä.

Jos nyt projisoin tämän 8-ulotteista palloa hahmottavan pistekuvion tasoon, saamme seuraavanlaisen 2-ulotteisen jakauman. Vertailun vuoksi, viereinen, oikeanpuolinen kuva näyttää neliön sisälle piirrettyyn ympyrään tasajakauman mukaisesti arvoit 1635 pistettä.



8-ulotteisen pallon projektiossa pisteet eivät enää jakaudu tasaisesti ympyrälle, vaan ne keskittyvät voimakkaasti kohti keskustaa. Tämäkin osoittaa, kuinka hintelä on moniulotteinen pallo verrattuna sitä ympäröivään moniulotteiseen kuution. Kun tutkitaan tasaista jakaumaa  $n$ -ulotteisessa pallossa, voidaan osoittaa, että näiden projektiopisteiden yhteisjakauman tiheysfunktio on tässä tapauksessa  $f(x,y)=\text{vakio}\cdot(1-x^2-y^2)^3$ . Tämä tarkastelu löytyy tilastollisten monimuuttujamenetelmien kirjastani (1995).

- ☐ Selvitän vielä lyhyesti, miten tuon 8-ulotteisen pallon pisteistö ja sen projektiokuva tehdään Survolla.



```

16 1 SURVO 84C EDITOR Fri Aug 16 17:01:25 1996 C:\KIRJA\ 50 100 0
1 *
2 *FILE CREATE PALLO8,33,9
3 *FIELDS:
4 *1 N 4 X1
5 *2 N 4 X2
6 *3 N 4 X3
7 *4 N 4 X4
8 *5 N 4 X5
9 *6 N 4 X6
10 *7 N 4 X7
11 *8 N 4 X8
12 *9 N 1 P
13 *END
14 *
15 *FILE INIT PALLO8,100000
16 *VAR X1,X2,X3,X4,X5,X6,X7,X8 TO PALLO8
17 *X1=2*rnd(1)-1 X2=2*rnd(1)-1 X3=2*rnd(1)-1 X4=2*rnd(1)-1 X5=2*rnd(1)-1
18 *X6=2*rnd(1)-1 X7=2*rnd(1)-1 X8=2*rnd(1)-1
19 *
20 *VAR P TO PALLO8
21 *P=if (X1*X1+X2*X2+X3*X3+X4*X4+X5*X5+X6*X6+X7*X7+X8*X8<1) then (1) else (0)
22 *
23 *

```

Alunperin luodaan tiedosto PALLO8 100000 pistettä varten, jotka rivin 16 VAR-komennolla arvotaan tasajakaumaperiaatteella 8-ulotteisesta kuutiosta. Kuution keskipiste on origo ja sen sivun pituus 2. Tämän jälkeen erotetaan ne pisteet, jotka ovat vastaavan pallon sisällä rivin 20 VAR-komennolla. Siinä apumuuttuja P saa arvon 1, jos arvottu piste on myös yksikkösaiteisen 8-ulotteisen pallon sisällä. Ulkopuolisilla pisteillä P saa arvon nolla.

```

19 1 SURVO 84C EDITOR Fri Aug 16 17:02:31 1996 C:\KIRJA\ 50 100 0
22 *
23 *GPLOT X(T)=cos(T),Y(T)=sin(T) / T=0,2*pi,pi/20 pi=3.141592653589793
24 *SCALE=-1,1 HEADER= FRAME=3 XDIV=0,1,0 YDIV=0,1,0
25 *SIZE=479,479 MODE=VGA OUTFILE=A
26 *
27 *GPLOT PALLO8,X1,X2 / IND=P
28 *SCALE=-1,1 HEADER= FRAME=3 XDIV=0,1,0 YDIV=0,1,0
29 *SIZE=479,479 MODE=VGA INFILE=A
30 *

```

Kuva syntyy (tässä kuvaruutuun) kahdessa vaiheessa. Ensin piirretään yksikköympyrä rivien 23-25 GPLOT-kaaviolla. Tulos tallentuu täsmennyksellä OUTFILE=A kuvatiedostoksi A.SPX. Jälkimmäinen GPLOT-kaavio piirtää 8-ulotteisen pisteparven projektion X1-X2-tasoon edellisen kuvan päälle (INFILE=A). Vain ne pisteet piirretään, joilla P=1 (täsmennys IND=P).

## Latteita kuvia

- Tilastollisen aineiston graafisen esittämisen ongelmat korostuvat moniulotteisissa aineistoissa, sillä esim. monikymmenulotteisen pisteparven litistäminen tasoon tarkkuudesta tinkimättä on täysi mahdottomuus. Kolmiulotteisuus esim. stereokuvapareina tai kuvaruudulla pyörivinä ns. spin-kuvina ei anna juuri mitään lisähyötyä näissä tilanteissa. Parasta on tunnustaa tosiasiat ja esittää se, mikä esitettävissä on, tasossa.

Tilastollisten monimuuttujamenetelmien eräänä tarkoituksena on tuottaa usean muuttujan aineistoista vähäulotteisia esityksiä esim. karsimalla tutkittavan ilmiön kannalta tarpeetonta satunnaisuutta. Tällöin menetelmien tuloksia tarkasteltaessa graafiset keinot tulevat paremmin ulottuvillemme.

Sopii kysyä, onko moniulotteisen ilmiön graafisessa esittämisessä mitään mieltä, koska itse ilmiöllä on harvoin suoraa suhdetta fysikaaliseen, näkyvään todellisuuteen. Kaikki kuvalliset keinot ovat tällöin täysin sopimuksenvaraisia. On kuitenkin kiistatonta, että ihmisen on jopa huonostikin suunnitellusta kuvallisesta esityksestä helpompi nähdä asioiden välisiä yhteyksiä kuin katselemalla pelkkää lukujen muodostamaa havaintotaulukkoa. Kuvien hahmottamisessa ihminen on vielä ylivoimainen tehokkaimpiinkin robotteihin verrattuna.

Miltei kaikkien kuvallisten keinojen perustana ovat tavanomaiset kaksiulotteiset, suorakulmaiset koordinaattiesitykset, joissa havainnot näkyvät pisteinä tai pisteen laajennuksina. Laajennuksella tarkoitetaan sitä, että "pisteet" voivat olla erikokoisia, -muotoisia ja -värisiä. Niiden ympärille voi kasautua myös eri muuttujista riippuvaa tietoa erimittaisilla ja -suuntaisilla janoilla tai käyränpätkillä kuvattuina. Siis erilaisilla pisteen liitännäisillä saadaan kuvaan jollain tavoin mukaan hyvinkin monen muuttujan osuus.

On ehdotettu myös hyvinkin erikoistuneita kuvaustapoja. Näistä eksoottisimpia ovat ns. Chernoffin naamat. Niissä muuttujat asetetaan vastaamaan kasvojen eri piirteitä. Menetelmän viehätys piilee siinä, että tukeudutaan suoraan ihmisen opittuun kykyyn tunnistaa lähimmäisensä naamataulusta.

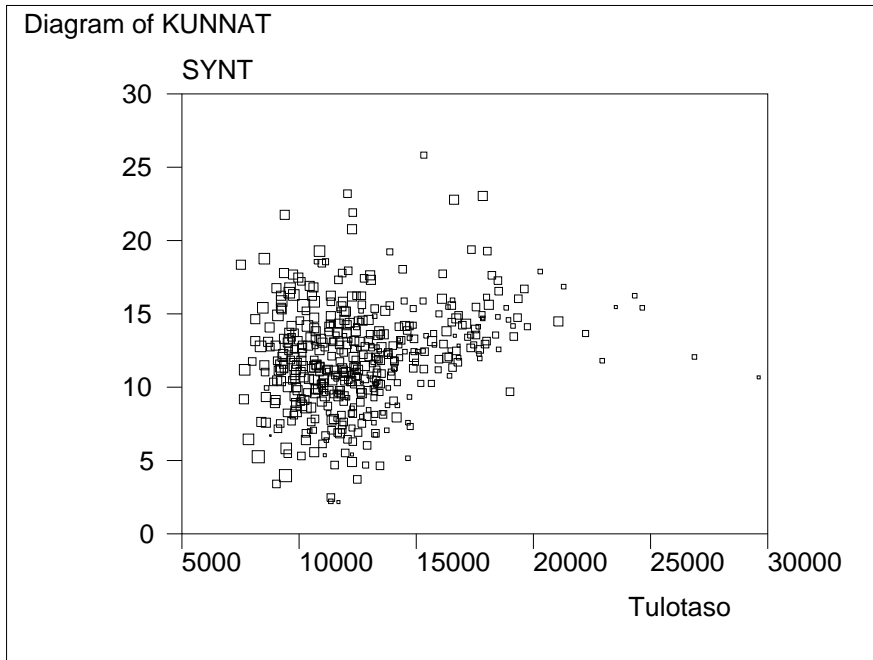
## Hajontakuvien yleistyksiset

- Kahden muuttujan hajontakuvissa, joita myös kutsutaan korrelaatiodiagrammoiksi, tarkastellaan ko. muuttujien keskinäisiä riippuvuuksia. Kutakin havaintoa vastaa kaksiulotteisessa koordinaatistossa piste, jonka asema x-akselin suunnassa määräytyy ensimmäisen muuttujan arvon ja y-akselin suunnassa toisen muuttujan arvon mukaan.

Tähän kuvaustapaan voi lisätä tietoa muista muuttujista laajentamalla eri tavoin "pisteen" ulkoista muotoa. Survon grafiikassa tämä käy helpoiten käyttämällä POINT-täsmennystä yleisimmässä muodossaan, kuten tapahtuu seuraavassa esimerkissä. Tällöin otetaan mukaan kolmas muuttuja, joka vaikuttaa pistettä vastaavan symbolin kokoon. LINE-täsmennys laajennuksineen tuottaa vielä monipuolisempia tehostuksia.

Seuraavassa diagrammassa on piirretty vastakkain Suomen kunnista (Survon esimerkkiaineisto KUNNAT) muuttujat Tulotaso ja SYNT (syntyneisyys 1000 asukasta kohti) siten, että kuntaa vastaavan neliömäisen "pisteen" sivun pituus on verrannollinen muuttujaan Äyriero=Äyri-12. Muunnos veroäyri-muuttujassa tehdään, jotta erot todella näkyisivät kuvassa.

```
27 1 SURVO 84C EDITOR Thu Jul 21 17:53:39 1994 C:\M\MON\ 100 100 0
1 *
2 *VAR SYNT=1000*Synt./Väestö TO KUNNAT
3 *VAR Äyriero=Äyri-12 TO KUNNAT
4 *
5 *GPLOT KUNNAT,Tulotaso,SYNT
6 *POINT=5,10,Äyriero,8
7 *
```



Kuvasta ilmenee paitsi tulotason ja syntyneisyyden riippuvuus myös se, että veroäyri on odotetusti alhaisen tulotason kunnissa suurimmillaan ja korkean tulotason kunnissa pienimmillään.

Rivin 6 POINT-täsmennys määrää, mitä kussakin tapauksessa tulee pisteen paikalle. Ensimmäinen parametri 5 valitsee symboliksi avoimen neliön. Toinen parametri 10 ilmoittaa neliön peruskoon ja kolmas parametri (Äyriero)

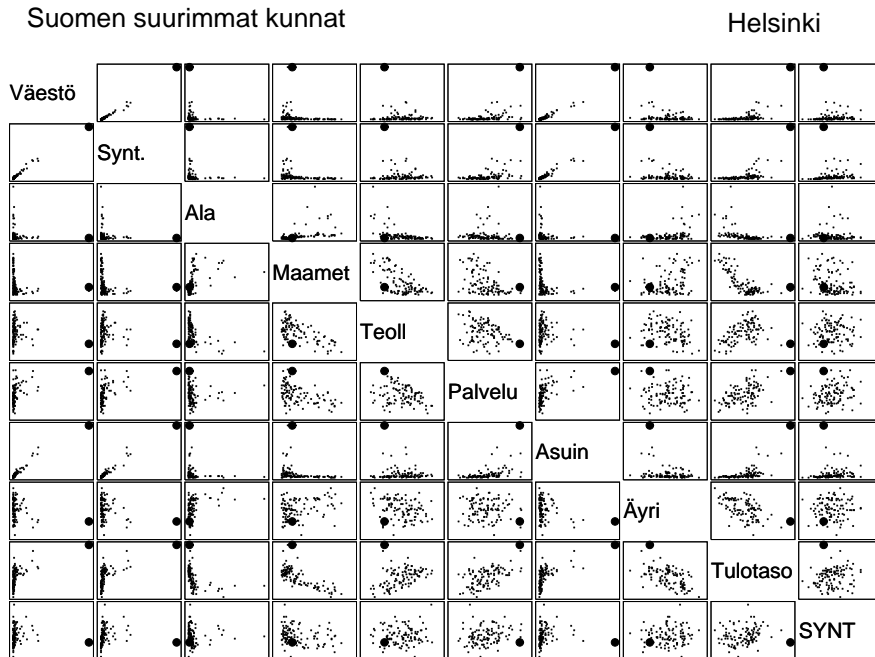
kokoon vaikuttavan muuttujan. Peruskokoa käytetään, kun  $\ddot{A}yriero$  on viimeisen parametrin (8) suuruinen. Yleisesti neliön koko (sivun pituus) on suhteessa muuttujan  $\ddot{A}yriero$  arvoon.

Survoilijat voivat vaatimusten kasvaessa kehittää itse paljon monimuotoisempiakin esityksiä havaintopisteille, jotta vielä useamman muuttujan vaikutus paistaisi kuvasta.

**Hajontakuvat koosteena**

- Hajontakuvamatriisilla (Draftsman’s display) tarkoitetaan kuvakoostetta, joka asetelultaan vastaa esim. korrelaatiomatriisia, mutta jonka "alkioina" ovat asianomaisten muuttujien korrelaatiodiagrammat. Englanninkielinen nimitys juontaa alkunsa teknisten laitteiden projektiopiirrosteoppiin. Tutkittavan aineiston kaikkien mahdollisten kaksiulotteisten hajontakuvien samanaikainen esittäminen antaa melko hyvän kokonaisnäkemyksen riippuvuuksien luonteesta. Se ei kuitenkaan voi tuottaa täydellistä kuvaa aineiston kokonaisvaihtelusta, koska minkäänulotteiset reunajakaumat eivät määrittele yhteisjakaumaa yksikäsitteisesti. Tästä huolimatta hajontakuvamatriisin piirtäminen on oivallinen keino tutustua uuden aineiston käyttäytymiseen ja auttaa esim. sopivien muuttujamuunnosten löytämisessä.

Suomen suurimpien kuntien (asukasluku yli 10000) 10 valitusta muuttujasta tehty hajontakuvamatriisi näyttää seuraavalta. Kussakin korrelaatiodiagrammassa Helsinki erottuu suurempana pisteenä.



Tällainen kuva syntyy Survon avulla vähimmillään PLOT- (kuvaruutuun GPlot-) komennolla, joka on varustettu täsmennyksellä TYPE=DRAFTS. Hyvin vähäluokkaisten muuttujien osalta on hyötyä JITTER-täsmennyksestä, joka täristää muuten päällekkäin tulevat pisteet "oikean" paikan ympärille satunnaisesti pisteparveksi. Ilman täristystä diskreettien muuttujien hajontakuvat surkastuvat usein mielenkiinnottomiksi hilapisteistöiksi eikä riippuvuuden luonteesta saa kunnon käsitystä. Täristys on hyvä esimerkki satunnaisuuden tahallisesta tuottamisesta. Mikään täristetty piste ei sijaitse täsmälleen "oikeassa" paikassa, mutta kuva silti kertoo totuuden näin paremmin.

- Tässä tapauksessa, koska yksi havainto halutaan erottaa muiden joukosta ja seurata sen asemaa kussakin osakuvassa erikseen, kuva rakennetaan kahdessa vaiheessa. Ensinnäkin piirretään koko aineisto (rivit 2-6) tallettaen sekä kuva (OUTFILE-täsmennys) että automaattisesti valitut piirrosasteikot (OUTSCALE-täsmennys). Tämän päälle saadaan sopeutettu kuva toisesta aineistosta (tai kuten tässä yhdestä havainnosta) eri värillä tai toisentyyppisinä pisteinä merkittynä muuten vastaavalla kaaviolla (rivit 8-14) käyttäen kohdistukseen INFILE- ja INSCALE-täsmennyksiä:

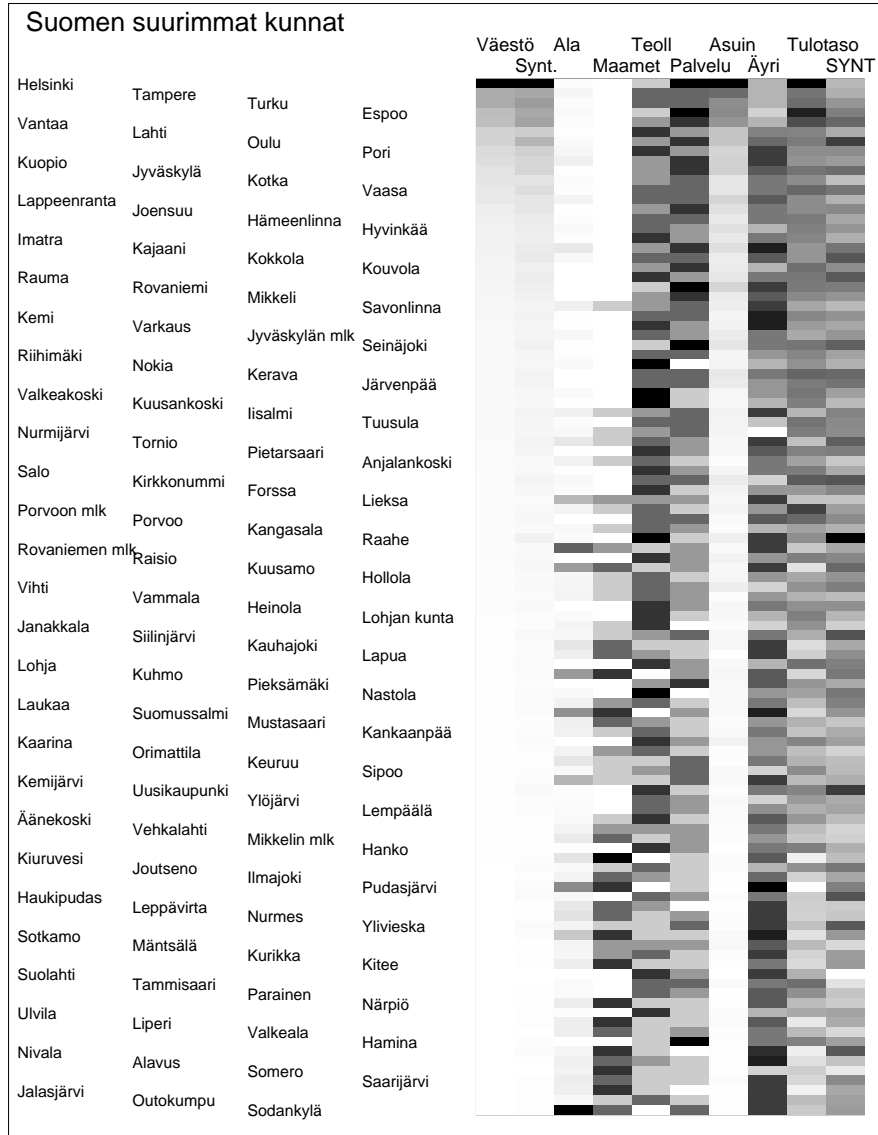
```

13 1 SURVO 84C EDITOR Sat Jul 23 15:45:38 1994 C:\M\MON\ 100 100 0
1 *
2 *Koko aineiston piirto (skaalausten valinta ja täristys):
3 *GPlot KUNNAT / TYPE=DRAFTS OUTSCALE=SKAALAT.TXT JITTER=30
4 *IND=Väestö,10000,500000
5 *XDIV=0,1,0 YDIV=0,10,1 HEADER=Suomen suurimmat_kunnat
6 *MASK=--AAAAAAAAA MODE=VGA OUTFILE=A
7 *.....
8 *Yhden havainnon lisäys isommalla merkinnällä:
9 *GPlot KUNNAT_ / TYPE=DRAFTS INSCALE=SKAALAT.TXT
10 *
11 *XDIV=0,1,0 YDIV=0,10,1 HEADER=
12 *MASK=--AAAAAAAAA MODE=VGA INFILE=A POINT=[RED],0,3 TEXTS=Kunta
13 *CASES=Kunta:Helsinki Helsinki voidaan vaihtaa
14 *Kunta=Helsinki,500,450 mihin tahansa muuhun kuntaan.
15 *

```

### Havintomatriisi rasterikuvana

- Toinen tapa yleiskuvan saamiseksi on piirtää koko havintomatriisi matriisi-diagrammana siten, että havaintoarvojen paikalla ovat niiden suuruuksia vastaavat tummuusasteeltaan vaihtelevat viivat tai laatikot. Esim. muuttujakohdistaan säädetään, miten tummuusaste muuttuu mustasta valkoiseen arvon kasvaessa tai päinvastoin.



Esimerkkinä esiintyvät jälleen Suomen suurimpia kuntia kuvaavat 10 muuttujaa. Kunnat on järjestetty asukasluvun mukaan suurimmasta pienimpään, jolloin kuvasta voi päätellä helpommin, mitkä tiedot korreloivat hyvin asukasluvun kanssa. Tässä esitystavassa kannattaa kiinnittää huomiota poikkeaviin tapauksiin, jotka näkyvät silmiinpistävinä "spektriviivoina".

□ Kuva on saatu aikaan seuraavasti:

```

13 1 SURVO 84C EDITOR Sat Jul 23 17:40:34 1994 C:\M\MON\ 100 100 0
15 * .....
16 *IND=Väestö,10000,500000
17 *FILE SORT KUNNAT BY -Väestö TO KUNNAT2
18 * .....
19 *MASK=A-AAAAAAAAA--
20 *HEADER=Suomen suurimmat kunnat
21 *PLOT KUNNAT2 / TYPE=MATRIX SCREEN=NEG DEVICE=PS,KUNNAT4.PS
22 *SIZE=1164,1500 XDIV=620,514,30 YDIV=30,1370,100
23 *ROWLABELS=[Swiss(6)],1,4,10 COLUMNLABELS=[Swiss(7)],1,2
24 *

```

Tiedoston KUNNAT mukaan otettavat havainnot on ensin lajiteltu väkiluvun mukaan laskevaan järjestykseen havaintotiedostoksi KUNNAT2 (rivit 16-17). Kuvan piirto tapahtuu riveillä 19-23 olevalla PLOT-kaaviolla, jossa kuvatyypin määrää täsmennys TYPE=MATRIX. Täsmennys SCREEN=NEG tarkoittaa, että muuttujanarvon vähetessä myös tummuusaste vähenee. Täsmennysten ROWLABELS ja COLUMNLABELS avulla rivi- ja sarakeotsikot saadaan lomittumaan niin, etteivät ne ahtaudu päällekkäin.

### Andrews-käyrät

- Kokonaan toisenlaisen näkökulman moniulotteisen aineiston graafiseen tarkasteluun tarjoaa *D.F.Andrewsin* (1972) esittämä Fourier-käyräteknikka. Kutakin  $p$  muuttujan  $\mathbf{X}=(X_1, X_2, \dots, X_p)$  havaintoa vastaa funktion

$$f_{\mathbf{X}}(t) = X_1/\sqrt{2} + X_2 \sin(t) + X_3 \cos(t) + X_4 \sin(2t) + X_5 \cos(2t) + X_6 \sin(3t) + \dots$$

kuvaaja välillä  $-\pi < t < \pi$ . Kun havainnot esitetään samassa koordinaatistossa, toisiaan muistuttavia havaintoja edustavat luonnollisesti toisiaan muistuttavat käyrät. Käyrien etäisyys toisistaan vastaa jopa tarkkaan havaintojen euklidista etäisyyttä  $p$ -ulotteisessa avaruudessa siinä mielessä, että havainnoille  $\mathbf{X}$  ja  $\mathbf{Y}$  pätee

$$\frac{1}{\pi} \int_{-\pi}^{\pi} [f_{\mathbf{X}}(t) - f_{\mathbf{Y}}(t)]^2 dt = \|\mathbf{X} - \mathbf{Y}\|^2 = (X_1 - Y_1)^2 + (X_2 - Y_2)^2 + \dots + (X_p - Y_p)^2.$$

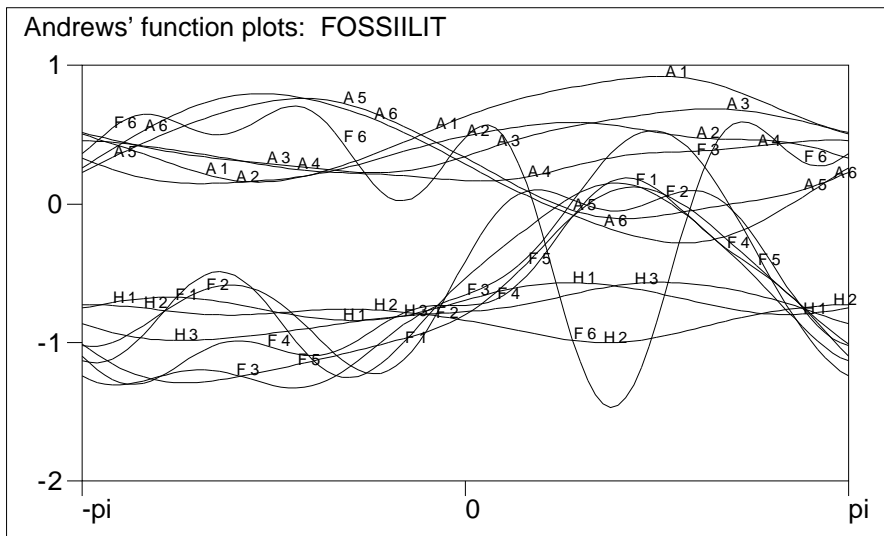
Andrews käytti eräänä esimerkkinään ihmis- ja apinalajien sekä fossiilien leualuista tehdyistä mittauksista koottua aineistoa. Alkuperäiset 8 muuttujaa on seuraavassa havaintotaulukossa korvattu erotteluanalyysin antamilla erottelumuuttujilla, jolloin eri lajien ja rotujen poikkeamat näkyvät muuttujissa  $X_1$ - $X_8$  voimakkuusjärjestyksessä. Muuttujat kannattaa asettaa tärkeysjärjestykseen, koska niiden vaikutukset itse käyrissä ilmenevät sitä paremmin mitä alhaisemmasta taajuudesta on kysymys. Erityisesti ensimmäinen muuttuja ( $X_1$ ) määrää yksinkertaisesti, millä perustasolla havaintoa vastaava käyrä kulkee.

Aineiston kolme ensimmäistä havaintoa vastaavat nykyisiä ihmisrotuja (länsi-afrikkalainen, britti, australialainen), 6 seuraavaa tunnettuja apinalajeja ja loput 6 muinaisia löydöksiä. Mielenkiintoista on tarkastella viimeistä (*Proconsul Africanus*), jota ainakin joskus on pidetty apinoiden ja ihmisten välisenä "puuttavana renkaana".

```

1 1 SURVO 84C EDITOR Sun Jul 24 14:24:08 1994 C:\M\MON\ 150 100 0
1 *
2 *DATA FOSSIILIT
3 * Laji X1 X2 X3 X4 X5 X6 X7 X8 Tunnus
4 * Westafr -8.09 0.49 0.18 0.75 -0.06 -0.04 0.04 0.03 H1
5 * British -9.37 -0.68 -0.44 -0.37 0.37 0.02 -0.01 0.05 H2
6 * Austral -8.87 1.44 0.36 -0.34 -0.29 -0.02 -0.01 -0.05 H3
7 * Gorilla1 6.28 2.89 0.43 -0.03 0.10 -0.14 0.07 0.08 A1
8 * Gorilla2 4.28 1.52 0.71 -0.06 0.25 0.15 -0.07 -0.10 A2
9 * Orang1 5.11 1.61 -0.72 0.04 -0.17 0.13 0.03 0.05 A3
10 * Orang2 3.60 0.28 -1.05 0.01 -0.03 -0.11 -0.11 -0.08 A4
11 * Chimpan1 3.46 -3.37 0.33 -0.32 -0.19 -0.04 0.09 0.09 A5
12 * Chimpan2 3.05 -4.21 0.17 0.28 0.04 0.02 -0.06 -0.06 A6
13 * Pith.Pek -6.73 3.63 1.14 2.11 -1.90 0.24 1.23 -0.55 F1
14 * Pith.P2 -5.90 3.95 0.89 1.58 -1.56 1.10 1.53 0.58 F2
15 * Par.Robu -7.56 6.34 1.66 0.10 -2.23 -1.01 0.68 -0.23 F3
16 * Par.Cras -7.79 4.33 1.42 0.01 -1.80 -0.25 0.04 -0.87 F4
17 * Megantro -8.23 5.03 1.13 -0.02 -1.41 -0.13 -0.28 -0.13 F5
18 * Proc.Afr 1.86 -4.28 -2.14 -1.73 2.06 1.80 2.61 2.48 F6
19 *_
    
```

Tämän aineiston Andrews-käyrät



☐ syntyvät Survon piirroskaaviolla:

```

14 1 SURVO 84C EDITOR Sun Jul 24 14:27:16 1994 C:\M\MON\ 150 100 0
19 *
20 *GPLOT FOSSIILIT / TYPE=ANDREWS LABEL=[Small],Tunnus
21 * YSCALE=-2 (1)1
22 *
23 *VARIABLES: A B Term
24 *X1 0 1 1/sqrt(2)
25 *X2 0 1 sin(t)
26 *X3 0 1 cos(t)
27 *X4 0 1 sin(2*t)
28 *X5 0 1 cos(2*t)
29 *X6 0 1 sin(3*t)
30 *X7 0 1 cos(3*t)
31 *X8 0 1 sin(4*t)
32 *END of plotting specifications
33 *
    
```



PLOT-komenossa täsmennys TYPE=ANDREWS synnyttää Andrews-käyrät. Se edellyttää erityistä VARIABLES-luetteloa, joka on tässä riveillä 23-32. Jos luettelo puuttuu, kun GPLOT käynnistetään, Survo tuottaa valmiin mallin toimituskenttään. Luettelossa kerrotaan muuttujat  $X$  (tärkeysjärjestyksessä). Joista on lupa skaalata muotoon  $(X-A)/B$  antamalla parametrit  $A$  ja  $B$ . Tässä tapauksessa on  $A=0$  ja  $B=1$  kaikilla muuttujilla eli muuttujanarvoja käytetään sellaisenaan. Eri havaintoja vastaavien käyrien tunnistamiseksi annetaan LABEL-täsmennys. Se ilmoittaa muuttujan, jonka arvoilla jokainen käyristä merkitään sopivasti porrastetuin välein. Havaintotaulukon viimeisenä sarakkeena on muuttuja Tunnus tätä tarkoitusta varten.

- Ensimmäisen erottelumuuttujan  $X_1$  mukaisesti Andrews-käyrät jakautuvat kahteen ryhmään, joista ylemmässä ovat apinat, alemmassa ihmisrodut ja useimmat fossiilit. Mystinen Proconsul Africanus (F6) kulkee etupäässä apinoitten puolella poiketen kerran ihmisten seuraan.

### Chernoffin naamat

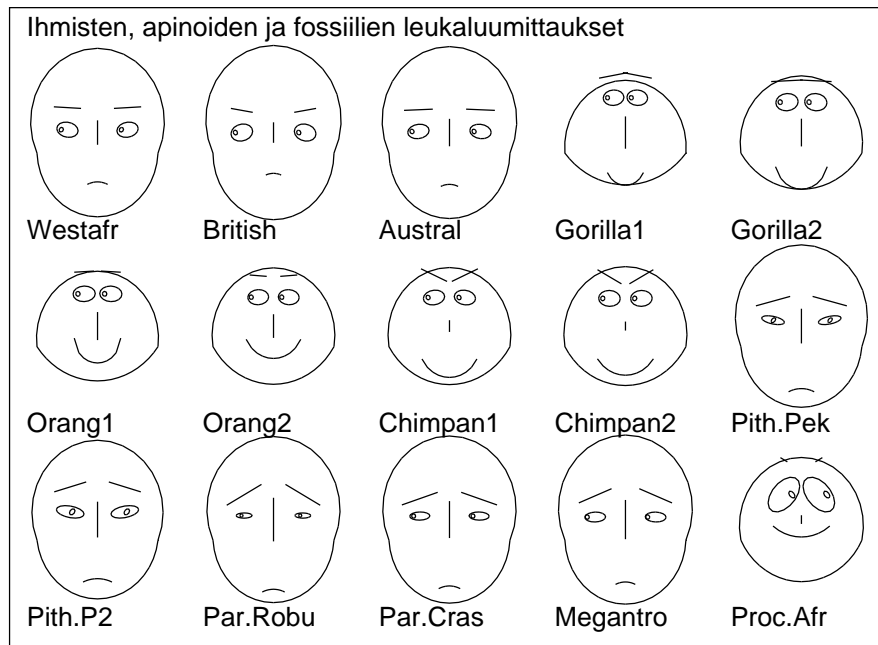
- Tässä piirrostavassa, jonka *H. Chernoff* on esittänyt vuonna 1973, muuttujat asetetaan vastaamaan karkeasti piirrettyjen kasvojen eri piirteitä. Survossa on seurattu tarkasti Chernoffin alkuperäistä ehdotusta, jossa valittavia piirteitä oli kaikkiaan 18. Jos aktivoidaan PLOT-komento varustettuna pelkällä TYPE=FACES täsmennyksellä, toimituskenttään kopioituu mallikaavio, jota muokkaamalla soveltaja liittää muuttujat ja kasvojen piirteet toisiinsa. Tämän mallin keskeinen osa on VARIABLES-luettelo, jonka olen jo valmiiksi täyttänyt fossiiliaineiston mukaisesti.

| 1 1 SURVO 84C EDITOR Sun Aug 18 11:28:21 1996 |   |                                 |        | C:\M\MON\ 120 100 0 |                               |
|---|---|---------------------------------|--------|---------------------|-------------------------------|
| 20  | * | _                               |        |                     |                               |
| 21  | * | VARIABLES:                      | xmin   | xmax                | Features                      |
| 22  | * | -                               | *      | **                  | Radius to corner of face OP   |
| 23  | * | -                               | *      | **                  | Angle of OP to horizontal     |
| 24  | * | D1                              | 6.28** | -9.37*              | Vertical size of face OU      |
| 25  | * | -                               | *      | **                  | Eccentricity of upper face    |
| 26  | * | -                               | *      | **                  | Eccentricity of lower face    |
| 27  | * | D2                              | -4.28* | 6.34**              | Length of nose                |
| 28  | * | D3                              | -2.14* | 1.66**              | Vertical position of mouth    |
| 29  | * | D1                              | -9.37* | 6.28**              | Curvature of mouth 1/R        |
| 30  | * | D1                              | -9.37* | 6.28**              | Width of mouth                |
| 31  | * | D1                              | -9.37* | 6.28**              | Vertical position of eyes     |
| 32  | * | D1                              | 6.28** | -9.37*              | Separation of eyes            |
| 33  | * | D4                              | -1.73* | 2.11**              | Slant of eyes                 |
| 34  | * | D5                              | -2.23* | 2.06**              | Eccentricity of eyes          |
| 35  | * | D6                              | -1.01* | 1.8**               | Size of eyes                  |
| 36  | * | D7                              | -0.28* | 2.61**              | Position of pupils            |
| 37  | * | D8                              | -0.87* | 2.48**              | Vertical position of eyebrows |
| 38  | * | D2                              | -4.28* | 6.34**              | Slant of eyebrows             |
| 39  | * | D3                              | -2.14* | 1.66**              | Size of eyebrows              |
| 40  | * | *END of plotting specifications |        |                     |                               |
| 41  | * |                                 |        |                     |                               |

Taulukon viimeisinä sarakkeina ovat kasvojen piirteiden selitykset (Features) ja niiden luonnolliset minimi- ja maksimiarvot. Soveltajan tehtävä on päivittää

kolme ensimmäistä saraketta, joissa nimetään eri piirteisiin vaikuttavat muuttajat ja niiden minimi- ja maksimiarvot. Tällöin muuttujien arvot kuvautuvat piirteiksi lineaarisella muunnoksella, joka asettaa minimin minimejä vastaan ja maksimit maksimeja vastaan. Muuttujien minimejä ja maksimeja ei tarvitse erikseen laskea aineistosta, vaan minimin paikalla on valmiiksi merkintä \* ja maksimin paikalla \*\*. Kuvauksen voi kääntää vastakkaiseen suuntaan asettamalla minimin paikalle \*\* ja maksimin \*. Erityisesti kun muuttujia on vähemmän kuin naaman piirteitä, tärkeinä pidettyjä muuttujia kannattaa käyttää useasti. Naaman piirteen voi vakioda (minimin ja maksimin keskiväliin) panemalla muuttujan paikalle merkinnän - .

- Fossiiliaineistoa piirrettäessä on houkutus yrittää valita vastaavuudet siten, että ihmisistä ja apinoista tulee jossain määrin itsensä näköisiä. On kuitenkin kohtuutonta kuvitella, että näin saataisiin fossiilit myös näyttämään "oikeilta". Havaitsemme vain, että Chernoffin naamoina useimmat fossiilit ovat enemmän ihmisen kuin apinan kaltaisia ja että Proconsul Africanus on tässä seurassa tosi outo ilmestys.



Survon Chernoff-ohjelmassa kuvaruudulla on myös mahdollista käyttää värejä ja esim. "maalata" kasvot ja silmämunat. Naamakuvien todellinen hyöty käytännön sovelluksissa on kuitenkin jäänyt vähäiseksi alkuperäisen idean hauskuudesta huolimatta.

### Vaiheittaista ryhmittelyä

- Eräissä monimuuttujamenetelmissä tulosten esittäminen kuvallisessa muodossa kuuluu menetelmän ominaispiirteisiin. Havaintojen hierarkkinen ryhmittely on tässä suhteessa otollinen.

Yleisesti ryhmittelyanalyysi kohdistuu tilastollisiin aineistoihin, jotka ovat useasta eri perusjoukosta saatujen otosten (ryhmien) yhdistelmiä. Tarkoituksena on paljastaa oikea ryhmien lukumäärä ja luokitella havainnot näihin ryhmiin.

Useimmat ryhmittelymenetelmät ovat luonteeltaan heuristisia; niiltä puuttuu selkeä teoreettinen tausta. Suosittuja ovat juuri hierarkkiset menetelmät. Tällöin esim. aluksi jokainen havainto muodostaa oman ryhmänsä ja etsitään ne havainnot, jotka ovat kaikkein läheisimpiä ja yhdistetään ne kahden havainnon ryhmäksi. Tämän jälkeen uusitaan sama menettely, jolloin syntyy toinen kahden havainnon ryhmä tai jokin havainto yhtyy ensimmäiseen kahden havainnon ryhmään. Tätä menettelyä toistetaan jatkuvasti, jolloin joka kerralla ryhmien lukumäärä vähenee yhdellä. Kun näin jatketaan, lopulta kaikki havainnot kasautuvat yhdeksi ryhmäksi. Tarkoitus on kuitenkin keskeyttää menettely sellaiseen vaiheeseen, jossa esim. ryhmittelyn hyvyttä kuvaavassa kriteerissä tapahtuu selvä muutos. Hierarkkinen ryhmittely voi tapahtua myös toisinpäin lähtemällä jakamaan kaikkien havaintojen muodostamaa ryhmää vaiheittain pienempiin. Menetelmästä on lisäksi monenlaisia muunnelmia. On mm. eri tapoja mitata etäisyyksiä havaintoyksiköiden ja ryhmien välillä.

Survossa hierarkkisesta ryhmittelystä vastaa uusi HCLUSTER-ohjelma, jonka *Fredrik Åberg* on laatinut osana opinnäytettään. Esimerkkinä sen toiminnasta tarkastelen pientä aineistoa, jonka Helsingin Sanomat keräsi keväällä 1996 EU-maiden Brysselin edustustoilta. Tässä kyselyssä maat ottivat kantaa 15 EU:n lähitulevaisuutta koskevaan kysymykseen.

Vastaukset olen koodannut tiedostoon EUMAAT 5-luokkaisina siten, että

- 5 = Voimakas kannatus,
- 4 = Kannattaa,
- 3 = Kiinnostunut,
- 2 = Ei hyväksy,
- 1 = Vastustaa,

Kysymykset olivat

```

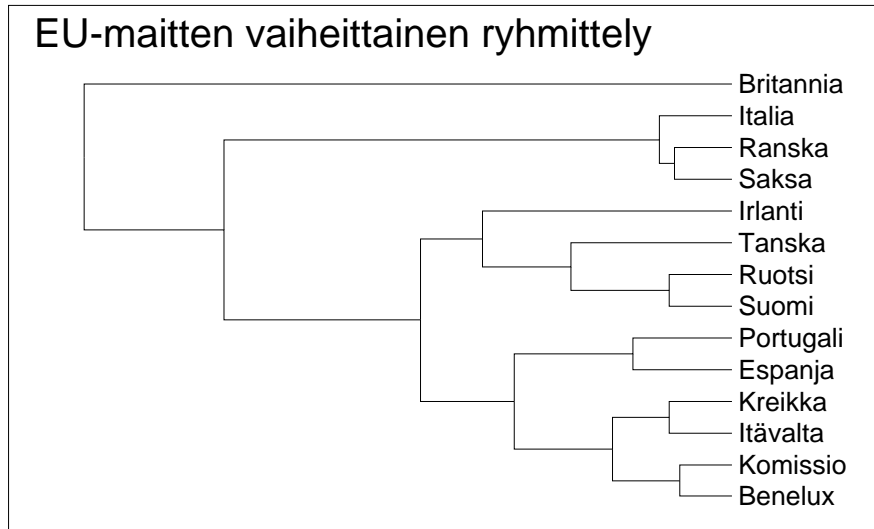
19 1 SURVO 84C EDITOR Sat Aug 17 13:52:30 1996 C:\KIRJA\ 240 100 0
1 *
2 *FILE STATUS EUMAAT
3 *EU-maitten kannat keväällä 1996
4 *5=Kannattaa - 1=Vastustaa
5 *FIELDS: (active)
6 * 1 SA 12 Maa
7 * 2 NA 1 TiivisEU Pyrittävä yhä tiiviimpään integraatioon (#)
8 * 3 NA 1 ItäEU Laajennuttava itä- ja keski-Eurooppaan (#)
9 * 4 NA 1 Äänet Äänimääriä muutettava ministerineuvostossa (#)
10 * 5 NA 1 Puhjoht Puheenjohtajuus-järjestelmää syytä muuttaa (#)
11 * 6 NA 1 Komissio Komission kokoa ratkaisevasti rajoitettava (#)
12 * 7 NA 1 Parlvalt Parlamentille lisää valtaa (#)
13 * 8 NA 1 Yhtpäät Yhteispäätösmenettelyä lisättävä (#)
14 * 9 NA 1 Määräen Määräenemmistöpäätösten osuutta lisättävä (#)
15 * 10 NA 1 Työttöm Työttömyyden torjunta perussopimukseen (#)
16 * 11 NA 1 Päätulko Määräenemmistöpäätöksiä lisättävä ulkopoliitikassa
17 * 12 NA 1 Ulkomin Ulkoministeriä tarvitaan (#)
18 * 13 NA 1 WEU EU ja WEU yhdistettävä (#)
19 * 14 NA 1 Siirtol Siirtolaisuusasiat unionin toimivaltaan (#)
20 * 15 NA 1 Poliisi Poliisiasiat unionin toimivaltaan (#)
21 * 16 NA 1 Schengen Schengenin sopimusta laajennettava (#)
22 *END
23 *SURVO 84C data file EUMAAT: record=53 bytes, M1=24 L=64 M=16 N=14
    
```

ja eri edustustojen kannanotot:

|           | TiivisEU | Äänet   | Komissio | Yhtpäät | Työttöm  | Ulkomin | Siirtol | Schengen |   |   |   |   |   |   |   |
|-----------|----------|---------|----------|---------|----------|---------|---------|----------|---|---|---|---|---|---|---|
|           | ItäEU    | Puhjoht | Parlvalt | Määräen | Päätulko | WEU     | Poliisi |          |   |   |   |   |   |   |   |
| Saksa     | 5        | 5       | 4        | 4       | 5        | 5       | 5       | 5        | 2 | 5 | 3 | 5 | 5 | 3 | 5 |
| Ranska    | 5        | 5       | 5        | 3       | 5        | 5       | 5       | 5        | 1 | 5 | 5 | 4 | 4 | 4 | 2 |
| Britannia | 1        | 5       | 4        | 3       | 4        | 1       | 2       | 1        | 1 | 1 | 1 | 1 | 1 | 1 | 2 |
| Italia    | 5        | 4       | 5        | 3       | 4        | 4       | 4       | 4        | 2 | 4 | 3 | 4 | 4 | 3 | 3 |
| Espanja   | 4        | 3       | 5        | 3       | 2        | 2       | 3       | 3        | 4 | 4 | 3 | 4 | 2 | 2 | 2 |
| Suomi     | 2        | 5       | 1        | 1       | 1        | 3       | 3       | 4        | 5 | 3 | 1 | 2 | 2 | 2 | 4 |
| Ruotsi    | 3        | 5       | 1        | 2       | 1        | 2       | 3       | 3        | 5 | 3 | 3 | 2 | 4 | 2 | 4 |
| Itävalta  | 5        | 5       | 2        | 3       | 3        | 4       | 4       | 4        | 5 | 4 | 3 | 2 | 4 | 3 | 4 |
| Tanska    | 2        | 5       | 3        | 3       | 5        | 3       | 3       | 4        | 5 | 2 | 3 | 3 | 2 | 2 | 3 |
| Irlanti   | 5        | 5       | 3        | 1       | 1        | 2       | 3       | 5        | 3 | 2 | 3 | 1 | 4 | 4 | 4 |
| Benelux   | 5        | 5       | 3        | 4       | 1        | 4       | 5       | 4        | 4 | 5 | 2 | 5 | 5 | 3 | 5 |
| Portugali | 4        | 3       | 3        | 3       | 1        | 2       | 3       | 4        | 4 | 4 | 2 | 4 | 4 | 4 | 4 |
| Kreikka   | 5        | 5       | 2        | 2       | 2        | 5       | 5       | 4        | 5 | 3 | 2 | 4 | 4 | 2 | 3 |
| Komissio  | 5        | 5       | 4        | 2       | 2        | 4       | 5       | 4        | 5 | 5 | 1 | 5 | 4 | 2 | 5 |

Itse asiassa en ollut huomannut tätä juttua lehdestä, vaan Väinö Lignell, maamme markkinatutkimuksen pioneereja ja muinainen tilastotieteen assistentti, ilmaantui laitokselle lehtileike kädessään. Lehdessä tuo taulukko oli kuvattu käyttäen numeroiden asemasta erilaisia kuvakkeita, jolloin hieman paremmin saattoi nähdä, mitkä maat ovat samoilla linjoilla, mitkä taas eivät. Väinö Lignellillä oli kuitenkin vahva aavistus, että yhteenvedon voisi esittää iskeväminkin kuvallisessa muodossa. Olipa ajatellut mm. Chernoffin naamoja tähän tarkoitukseen.

Naamakuvat saattaisivatkin kertoa jotain, mutta itselleni oli heti selvää, että tässä kannattaa kokeilla hierarkkista ryhmittelyä. Käyttämällä siis Survon HCLUSTER-ohjelmaa syntyi muutaman kokeilun kautta seuraavanlainen esitys:



Päädyin ns. complete linkage-ryhmittelyyn, jossa ryhmien etäisyydeksi tulkitaan niiden kaukaisimpien jäsenten etäisyys. Kuvassa vaaka-akseli - oikealta vasemmalle - kertoo ryhmien välisen etäisyyden ja ryhmittelypuu sen, mitkä maat ja missä vaiheessa liittyvät yhteen. Esim. Suomi ja Ruotsi muodostavat heti oman ryhmänsä, johon liittyy pian Tanska ja myöhemmin Irlanti. Britannia osoittautui kaikissa ratkaisuisissa muista poikkeavimmaksi kannanotoissaan. Mielestäni näin syntyy selkeä ja uskottava käsitys maitten välisistä suhteista EU-kysymyksissä.

```

18 1 SURVO 84C EDITOR Sat Aug 17 15:46:56 1996 C:\KIRJA\ 240 100 0
25 *
26 *CLUSTER EUMAAT,CUR+1 / METHOD=COMPLETE
27 * Hierarchical cluster analysis of observations in EUMAAT N=14
28 * Method : Complete linkage Distance : Squared Euclidian
29 *
30 * Activate line below to plot the tree on CRT.
31 * GPLOT #TREE#,X,Y_
32 * Change YSCALE and XSCALE below to plot subtrees.
33 * YSCALE=1,14 POINT=Label FRAME=0 XLABEL=
34 * XSCALE=-125/1,0 PEN=[Swiss(10)] MODE=VGA YLABEL=
35 * Change LINE spec. to 1 below for 'triangular' tree.
36 * LINE=3 XDIV=1,65,10 YDIV=2,20,1
37 * HEADER=[Swiss(15)],EU-maitten vaiheittainen ryhmittely
38 *
39 * Stage Cluster 1 Cluster 2 Distance
40 * 1 Benelux Komissio 10.00000
41 * 2 Saksa Ranska 11.00000
42 * 3 Itävalta Kreikka 12.00000
43 * 4 Suomi Ruotsi 12.00000
44 * 5 GROUP 2 Italia 14.00000
45 * 6 Espanja Portugali 19.00000
46 * 7 GROUP 3 GROUP 1 23.00000
47 * 8 GROUP 4 Tanska 31.00000
48 * 9 GROUP 6 GROUP 7 42.00000
49 * 10 GROUP 8 Irlanti 48.00000
50 * 11 GROUP 9 GROUP 10 60.00000
51 * 12 GROUP 5 GROUP 11 98.00000
52 * 13 GROUP 12 Britannia 125.00000
53 *

```

Kuva on tehty kahdessa vaiheessa.

Aktivoinnalla HCLUSTER-komento (rivi 26) saadaan tulokset riveille 27-52. Aluksi ilmoitetaan valittu menetelmä (rivit 27-28), sitten HCLUSTER kirjoittaa kuvallista esitystä varten valmiin GLOT-kaavion (rivit 30-37) ja lopuksi numeeriset tulokset (rivit 39-52), joista näkyy ryhmien kehittyminen ja etäisyydet.

Tämän jälkeen ei muuta tarvita kuin aktivoida tulosten joukossa oleva GLOT-komento (rivi 31). Se piirtää aineiston #TREE# muuttujat X ja Y vastakkain yhdistäen LINE-täsmennyksellä sopivasti peräkkäisiä pisteitä. Näin saadaan vailla vaivaa aikaan edellä nähty ryhmittelypuu. Aineisto #TREE# on EUMAAT-aineistosta ryhmittelytulosten perusteella johdettu keinotekoinen Survon datatiedosto, jonka siis Fredrik Åbergin ohjelma ovelasti luo sivutuotteenaan. Kuvakaaviossa olen muuttanut eräitä otsikkotietoja jälkikäteen.

Piirtämistä varten HCLUSTER on siis muodostanut datatiedoston #TREE#, jossa ryhmittelypuun koordinaattien (X,Y) lisäksi on tarvituissa paikoissa maitten nimet (Label). Kun tiedostoa katselee FILE SHOW-komennolla, sen alku näyttää seuraavalta:

```

1 1 SURVO 84C EDITOR Sat Aug 17 15:49:03 1996 C:\KIRJA\ 240 100
File #TREE# N=78 X 0.000
1 X Y Label
1 0.000 1.000 Benelux
2 -10.000 1.500
3 -10.000
4 0.000 2.000 Komissio
5 -10.000 1.500
6 -10.000
7 0.000 3.000 Itävalta
8 -12.000 3.500
9 -12.000
10 0.000 4.000 Kreikka
11 -12.000 3.500
12 -12.000
13 -12.000 3.500
14 -23.000 2.500
15 -23.000
16 -10.000 1.500
17 -23.000 2.500
18 -23.000
19 0.000 5.000 Espanja
20 -19.000 5.500
21 -19.000
To stop, press EXIT! (F1=HELP)

```

On Survolle tunnusomaista, että eri toiminnot hyödyntävät toisiaan. Tässä tapauksessa kuvan piirtämistä ei ole tarvinnut ohjelmoida C-kielellä vaan HCLUSTER synnyttää kuvakaavion ja keinotekoisena aineiston niin, että päästään soveltamaan valmiita GLOT-ohjelmaa.

### Säveltäjät kartalla

- Moniulotteisella skaalauksella tarkoitetaan menettelyä, jolla usean havainnon välisiä eroja koskevien etäisyys-, samanlaisuus- tai erilaisuustietojen perusteella havaintoja vastaavat pisteet yritetään sijoittaa "kartalle" eli tavallisesti 1- , 2- tai korkeintaan 3-ulotteiseen avaruuteen siten, että näiden pisteiden keskinäiset etäisyydet vastaavat annettuja etäisyydestietoja.

Tämäntyyllisellä menettelyllä on pitkät perinteet psykologisissa tutkimuksissa. Myös erilaisissa tuotevertailuissa se lienee suosittu. Sovelluksissa on tavoitteena löytää mielekkäitä rakenteita jopa täysin abstrakteista ilmiöistä. Lähtökohtana on esim. yhden tai useamman henkilön subjektiiviset arviot havaintokohteiden keskinäisistä samanlaisuuksista, erilaisuuksista tai etäisyyksistä.

Ongelmia syntyy miltei aina siitä, etteivät ilmoitetut havaintokohteiden etäisyydet välttämättä ole yhteensopivia, vaan niissä saattaa esiintyä keskinäisiä ristiriitaisuuksia. Skaalausmenettelyn tulee tasoittaa nämä ristiriitaisuudet mahdollisimman taitavasti. Samalla on pidettävä huolta, ettei lopullinen esitys jää liian moniulotteiseksi, koska silloin menetettäisiin tuloksen havainnollisuus.

Pyysin *Olli Mustosta* valitsemaan 10 huomattavaa säveltäjää musiikin eri aikakausilta ja vertaamaan näitä täysin intuitiivisesti heidän koko tuotantonsa ja tyyliensä pohjalta. Sovimme, että hän käyttää asteikkoa 0 - 100 siten, että mitä enemmän hän katsoo säveltäjien eroavan toisistaan, sitä suuremman pistemäärän hän antaa. Noin puolen tunnin harkinnan jälkeen hän esitti seuraavan etäisyystaulukon, joka on valmiiksi muotoiltu Survon matriisiksi:

```

29 1 SURVO 84C EDITOR Sun Jun 12 09:07:20 1994 D:\M\MEN2\ 200 100 0
1 *
2 *MATRIX MUS
3 *///      Bach Hayd Moza Beet Schu Brah Sibe Debu Bart Sost
4 *Bach      0  50  30  20  40  40  40  50  30  30
5 *Haydn     50  0  10  15  30  70  90  50  80  40
6 *Mozart    30  10  0  20  25  40  70  50  80  50
7 *Beethven  20  15  20  0  10  20  25  80  60  40
8 *Schubert  40  30  25  10  0  15  60  50  70  60
9 *Brahms    40  70  40  20  15  0  20  70  70  70
10 *Sibelius 40  90  70  25  60  20  0  35  35  20
11 *Debussy  50  50  50  80  50  70  35  0  15  40
12 *Bartok   30  80  80  60  70  70  35  15  0  20
13 *Sostakov 30  40  50  40  60  70  20  40  20  0
14 *
15 *MAT SAVE MUS
16 *MAT MUST=MUS' / Pitempien riviotsikoiden
17 *MAT CLABELS FROM MUST TO MUS_ / kopiointi sarakeotsikoiksi
18 *

```

Säveltäjät esiintyvät taulukossa suurin piirtein aikajärjestyksessä. O.M. käytti asteikkoa 5 yksikön välein, koska hän katsoi, ettei ole edellytyksiä tarkempaan arviointiin. Suurin etäisyys 90 esiintyy Sibeliuksen ja Haydnin välillä.

Useissa monimuuttujamenetelmissä syntyy tai esiintyy luonnostaan tietoja, jotka on parasta kuvata Survossa ns. matriisitiedostoina. Matriisitiedostoja käsitellään paitsi näitä menetelmiä vastaavilla Survon komennoilla myös Survon matriisitulkilla. Matriisitulkilla jatketaan usein laskentaa siitä eteenpäin, mihin perusmenetelmät päättyvät. Kerron tästä tulkista lisää myöhemmin.

Tässä esimerkissä matriisitulkkia on vaivattu jo sen verran, että etäisyysmatriisi on talletettu matriisitiedostoksi MUS (rivillä 15) ja lisäksi otsikkotietoja on hieman muokkailtu (rivit 16-17).

- ☐ Survo sisältää kaksi eri menetelmää moniulotteiseen skaalaukseen. Parempi ja yleisempi niistä on pienimmän neliösumman skaalaus. Tässä tyydyn esittelemään vain yksinkertaisempaa eli ns. klassista skaalausta. Se tapahtuu sukrokomennolla /CSCAL, joka puolestaan soveltaa matriisitulkkia automaattisesti kaikissa laskelmissa.

```

22 1 SURVO 84C EDITOR Sun Jun 12 09:23:02 1994 D:\M\MEN2\ 200 100 0
18 *
19 */CSCAL MUS,2
20 *Classical multidimensional scaling for MUS:
21 *MAT LOAD CSCAL.M,END+2 / Scale values (2 dimensions)
22 *MAT LOAD CSEIGEN.M,END+2 / Eigenvalues
23 *MAT LOAD CSCENT.M,END+2 / Eigenvalues (percentages)
24 *MAT LOAD CSDIST.M,END+2 / Reproduced distances
25 *GPLOT CSCAL.M,DIM1,DIM2 / POINT=[SMALL],CASE
26 *LSCAL MUS,CSCAL.M,END+2 / Least Squares Scaling
27 *Distance matrix MUS is not Euclidean!
28 *
29 *MAT CSCENT=CSCENT.M' / *CSCENT-Eigenvalues (in percentages)'
30 *MAT LOAD CSCENT,CUR+1_
31 *MATRIX CSCENT
32 *Eigenvalues (in percentages)'
33 */// Per cent Cumulat.
34 *DIM1 35.090 35.090
35 *DIM2 22.720 57.809
36 *DIM3 10.765 68.574
37 *DIM4 3.536 72.110
38 *DIM5 2.416 74.526
39 *DIM6 -0.000 74.526
40 *DIM7 -3.112 77.637
41 *DIM8 -4.260 81.897
42 *DIM9 -4.405 86.302
43 *DIM10 -13.698 100.000
44 *

```

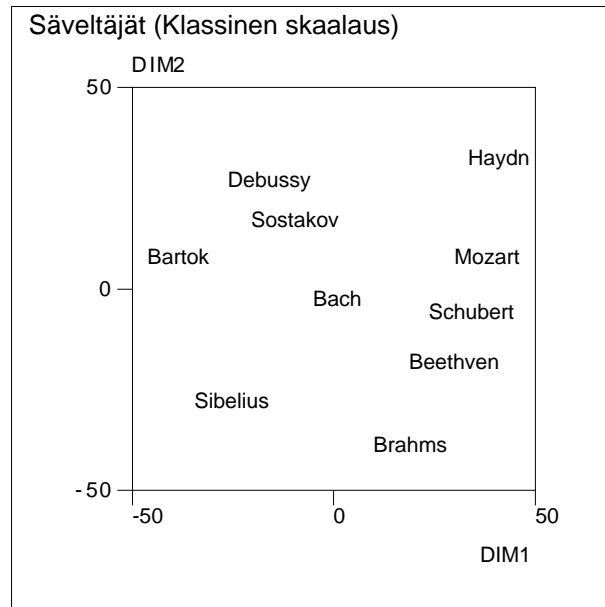
Rivin 19 /CSCAL-komento kohdistuu matriisiin MUS ja siinä ilmoitetaan myös, monikoulotteisena tuloksia halutaan tarkastella. Tulokset on saatu riveille 20-27 hyvin keskitetysti. Mukana ei suoraan ole mitään numeerisia tai kuvallisia esityksiä. Toimituskentässä ilmoitetut tulokset ovat valmiita komentoja ja selityksiä, joiden avulla saatan onkia näkyville haluamiani asioita. Kaikki tarpeellinen on jo siis laskettu. Riippuu tutkijasta, mitä hän halua tarkkailla lähemmin.



Tässä tapauksessa on otettu esiin riveille 34-43 muunnetun etäisyysmatriisin ominaisarvot prosentuaalisesti. Ne paljastavat alkutietojen johdonmukaisuuden ja kertovat myös, miten vahvoja ulottuvuuksia aineistosta löytyy. Aidon etäisyysmatriisin pitäisi olla sellainen, ettei negatiivisia ominaisarvoja esiintyisi lainkaan. Silloin annetut etäisyydet määrittelevät selkeän pistekuvion ja etäisyysmatriisia sanotaan euklidiseksi.

Kuten ominaisarvojen käyttäytymisestä nyt näkyy, etäisyysmatriisi sellaisenaan ei ole euklidinen. Kaksi ensimmäistä ulottuvuutta kuitenkin selittävät lähes 60% koko vaihtelusta ja 77% "positiivisesta" vaihtelusta. Kohtalaiset negatiiviset ominaisarvot viittaavat siihen, ettei arviointi voinut olla aivan ristiriidatonta. Sopivalla etäisyyksien epälineaarilla muunnoksella (esim. ottamalla neliöjuuri) matriisi tulisi lähemmäksi euklidista, mutta tällaisiin toimiin ei tässä tapauksessa ryhdytty.

- Kun aktivoidaan rivin 25 valmis GLOT-komento, kuvaruutuun ilmaantuu mielenkiintoisin tulos eli kaksiulotteinen esitys siitä, miten säveltäjät sijoittuvat toisiinsa nähden kartalla, jonka määrää kaksi ensimmäistä ulottuvuutta.



Ensimmäinen ulottuvuus oikealta vasemmalle vastaa muuten varsin hyvin aikaa vain sillä huomattavalla poikkeamalla, että "ajaton" Bach asettuu keskelle. Toinen dimensio on tulkittavissa ylhäältä alaspäin siirtymisenä "kevyestä raskaaseen" musiikkiin. Niinpä Wieniläisklassikot (Haydn, Mozart, Schubert ja Beethoven) muodostavat johdonmukaisen ketjun ja saavat jatkokseen vielä Brahmsin, joka Sibeliuksen kera sijoittuu "raskaimpaan sarjaan". Moderneimmat säveltäjät (Debussy, Shostakovits ja Bartok) muodostavat oman ryhmänsä ja on täysin ymmärrettävää, että näistä Shostakovits on lähinnä Bachia. Sibe-

lius jää omaan yksinäisyyteensä. On ilmeistä, ettei muusikoilla ole yhtenäistä käsitystä säveltäjien suhteista. Jokainen musiikin ystävä kokeilkoon, miltä näyttää hänen säveltäjämäailmansa.

## Muuttujakuvauksia

- Varsinaisissa monimuuttujamenetelmissä ei saata kokonaan välttyä numero-taulukoiden katselulta. Aina kuitenkin tietoa yritetään tiivistää niin pieneen tilaan kuin mahdollista. Tämä pyrkimys kuultaa esiin nimestä Survo. Raakaa tietoa survotaan niin, että tiedon "mehu" siitä tihkuu. Toisaalta nimen lähtökohtana on ollut "Survey analysis", sillä Survon alkutaipaleella haastattelututkimukset olivat tärkeitä kohteita.

Tiedon survonta monimuuttujamenetelmissä on useimmiten uusien, alkuperäisistä muuttujista yhdistämällä saatavien muuttujien rakentelua. Nämä ovat melkein poikkeuksetta muuttujien painotettuja summia eli lineaarisia yhdistelmiä, kuten on tapana sanoa. Johtava tavoite uusien muuttujien muodostettaessa on löytää selkeämmin tulkittavia tuloksia kuin mitä alkuperäiset muuttujat sellaisenaan kertovat.

Otan tässä esiin vain kaksi menetelmää, erotteluanalyysin ja faktorianalyysin. Niitäkin tarkkailen pelkästään yhden aineiston analyysin kautta. Aineisto liittyy *Tuomo Martikaisen* tutkimuksiin nuorten suhtautumisesta politiikkaan ja yhteiskuntaan. Hänen johdolla on kerätty puhelinhaastatteluina Helsingistä ja sen lähikunnista tietoja 18-30-vuotiailta suomalaisilta. Aineistossa on 705 havaintoa vuodelta 1988 ja 600 havaintoa vuodelta 1996. Tuomo Martikainen on ystävällisesti antanut aineistot käyttööni ja esittelen niiden avulla em. menetelmien soveltamista Survossa.

Tarkastelen tässä ainoastaan 37 muuttujaa. Näistä 36 on tyypillisiä asenne-muuttujia, jotka olen koodannut uudelleen asteikolle 1 - 5 siten, että 5 tarkoittaa yhtymistä väitteeseen ja 1 vastakkaista mielipidettä. Lisänä on vielä vastaajan sukupuoli (1=nainen, 2=mies).

- Kummastakin aineistosta on aluksi laskettu näiden muuttujien keskiarvot, hajonnat ja korrelaatiot. Muuttujat on tätä ennen jo valittu asettamalla tiedostoissa NUORET88 ja NUORET95 vain 37 muuttujaa aktiivisiksi.

```

14 1 SURVO 84C EDITOR Mon Aug 19 14:55:44 1996 C:\KIRJA\ 400 100 0
1 *
2 *CORR NUORET88
3 *MAT R88=CORR.M / *R88~R (NUORET88) S37*37
4 *MAT M88=MSN.M / *M88~MSN (NUORET88) 37*3
5 *
6 *CORR NUORET95
7 *MAT R95=CORR.M / *R95~R (NUORET95) S37*37
8 *MAT M95=MSN.M / *M95~MSN (NUORET95) 37*3
9 *

```

Kyseiset tunnusluvut lasketaan CORR-komennoilla (rivit 2 ja 6). Tulokset, korrelaatiomatriisi CORR.M sekä keskiarvojen ja hajontojen matriisi MSN.M kopioidaan uusiin matriisitiedostoihin.

- ☐ Poimimalla tuloksia toimituskenttään (tässä on käytetty lisäksi STATMSF-komentoa, joka laskee ns. suorat jakaumat) on helppo luoda tiivistelmä taulukon muodossa. Taulukosta näkyy muuttujittain nimet, keskiarvot (M), hajonnat (S) ja prosentuaaliset jakaumat. Kunkin muuttujan ylempi rivi koskee vuotta 1995 ja alempi vuotta 1988. Taulukon oikeassa reunassa ovat selitykset, jotka olen liittänyt tiedostoihin muuttujien nimien jatkeeksi. Tällöin selitykset saa vaivatta mukaan tulostaulukoihin.

|          | M    | S    | 1  | 2  | 3  | 4  | 5  |   |
|----------|------|------|----|----|----|----|----|---|
| INTEREST | 2.99 | 1.19 | 8  | 43 | 0  | 43 | 7  | Kiinnostunut politiikasta                     |
|          | 2.83 | 1.23 | 10 | 47 | 0  | 35 | 8  |   |
| TRUST    | 2.79 | 1.12 | 9  | 48 | 1  | 42 | 1  | Luottamus nuorten ongelmien ratkaisuun        |
|          | 2.52 | 1.16 | 17 | 48 | 1  | 32 | 2  |   |
| STRENGTH | 3.07 | 1.29 | 14 | 30 | 1  | 47 | 9  | Vakiintunut oma poliittinen käsitys           |
|          | 3.22 | 1.30 | 11 | 28 | 0  | 47 | 13 |   |
| STABLE   | 2.57 | 1.37 | 29 | 32 | 0  | 32 | 7  | Vakiintunut puoluekanta                       |
|          | 2.69 | 1.33 | 19 | 41 | 0  | 30 | 9  |   |
| ONLYVOTE | 3.56 | 1.24 | 6  | 24 | 1  | 46 | 23 | Äänestäminen ainoa keino vaikuttaa            |
|          | 2.96 | 1.25 | 10 | 40 | 2  | 38 | 9  |   |
| COMPLIC  | 2.95 | 1.48 | 24 | 24 | 1  | 36 | 16 | En ymmärrä politiikkaa                        |
|          | 3.07 | 1.30 | 12 | 33 | 3  | 40 | 12 |   |
| MUCHSAME | 2.86 | 1.53 | 26 | 27 | 0  | 28 | 19 | Asiat jatkuvat hallituspuolueista riippumatta |
|          | 3.03 | 1.23 | 10 | 37 | 4  | 41 | 9  |   |
| NOPEOPLE | 3.37 | 1.31 | 8  | 28 | 2  | 42 | 20 | Puolueet kiinnostuneita vain äänistä          |
|          | 3.33 | 1.12 | 2  | 33 | 5  | 48 | 12 |   |
| VOTEINT  | 4.83 | 0.47 | 0  | 1  | 1  | 12 | 86 | Aion käyttää äänioikeuttani                   |
|          | 4.46 | 1.02 | 5  | 3  | 0  | 26 | 66 |   |
| POLSYST  | 2.92 | 1.11 | 6  | 47 | 0  | 45 | 2  | Poliittinen järjestelmä toimii                |
|          | 3.02 | 1.06 | 4  | 41 | 5  | 48 | 2  |   |
| PSCIENCE | 3.82 | 1.79 | 28 | 0  | 4  | 0  | 69 | Ylpeä tieteellisistä saavutuksista            |
|          | 3.55 | 1.80 | 30 | 0  | 12 | 0  | 58 |   |
| PSPORT   | 4.16 | 1.61 | 20 | 0  | 1  | 0  | 79 | Ylpeä urheilusaavutuksista                    |
|          | 3.66 | 1.82 | 30 | 0  | 6  | 0  | 64 |   |
| PPOLSYST | 2.26 | 1.79 | 66 | 0  | 6  | 0  | 29 | Ylpeä eduskunnasta ja pol.järjestelmästä      |
|          | 2.41 | 1.76 | 58 | 0  | 14 | 0  | 28 |   |
| PARTS    | 3.67 | 1.85 | 31 | 0  | 4  | 0  | 65 | Ylpeä taiteellisista saavutuksista            |
|          | 3.97 | 1.64 | 21 | 0  | 9  | 0  | 70 |   |
| PECONOMY | 2.65 | 1.91 | 56 | 0  | 6  | 0  | 38 | Ylpeä taloudellisista saavutuksista           |
|          | 2.81 | 1.84 | 47 | 0  | 15 | 0  | 38 |   |
| PWELFARE | 4.48 | 1.33 | 12 | 0  | 1  | 0  | 87 | Ylpeä sosiaaliturvasta ja terv.huollosta      |
|          | 4.30 | 1.44 | 15 | 0  | 6  | 0  | 80 |   |

|          |      |      |       |          |  |
|----------|------|------|-------|----------|--|
| LEASH    | 3.77 | 1.05 | 2 20  | 1 56 22  | Poliitikot puolueensa "liekanuorassa"        |
|          | 4.37 | 1.42 | 14 0  | 3 0 83   |  |
| FAKES    | 3.80 | 1.14 | 3 20  | 0 49 28  | Poliitikot eivät anna oikeaa kuvaa itsestään |
|          | 4.05 | 1.68 | 23 0  | 2 0 75   |  |
| PLAYERS  | 3.21 | 1.33 | 9 34  | 1 38 18  | Poliitikot "pelaavat varman päälle"          |
|          | 3.30 | 1.95 | 41 0  | 3 0 56   |  |
| WASTEMON | 3.34 | 1.36 | 11 27 | 2 38 23  | Poliitikot tuhlailevat                       |
|          | 3.54 | 1.86 | 33 0  | 6 0 60   |  |
| PROTECT  | 3.29 | 1.30 | 10 26 | 5 42 17  | Poliitikot suojelevat toisiaan               |
|          | 3.07 | 1.88 | 42 0  | 12 0 46  |  |
| RECKLESS | 2.55 | 1.36 | 26 37 | 2 26 9   | Poliitikot häikäilemättömiä ja julk.kipeitä  |
|          | 2.37 | 1.86 | 64 0  | 4 0 32   |  |
| DEVELAID | 3.24 | 1.50 | 18 23 | 3 30 26  | Kehitysapua on lisättävä                     |
|          | 3.81 | 1.24 | 6 17  | 3 39 35  |  |
| REFUGEE  | 2.75 | 1.52 | 31 24 | 2 27 16  | Otettava vastaan enemmän pakolaisia          |
|          | 3.27 | 1.51 | 18 23 | 3 28 28  |  |
| EXPLOIT  | 2.98 | 1.41 | 18 30 | 4 33 16  | Suromistajat rikastuvat työl.kustannuksella  |
|          | 3.35 | 1.22 | 5 30  | 6 42 17  |  |
| FREERIDE | 2.44 | 1.36 | 32 32 | 2 27 7   | Useimmat tuensaajat eivät sitä ansaitse      |
|          | 2.56 | 1.16 | 13 53 | 5 22 7   |  |
| OWNFAULT | 2.66 | 1.44 | 30 27 | 1 33 10  | Työttömyyden syyt ihmisessä itsessään        |
|          | 3.17 | 1.27 | 9 33  | 5 40 14  |  |
| RICHPRIV | 2.92 | 1.44 | 23 25 | 3 35 14  | Rikkaille eri lait kuin köyhille             |
|          | 3.05 | 1.22 | 8 38  | 6 38 10  |  |
| DEPEND   | 3.44 | 1.29 | 13 16 | 2 53 17  | Hyvinvointivaltiossa ei huolehdi itsestä     |
|          | 3.17 | 1.15 | 4 37  | 6 44 9   |  |
| OWNFEET  | 3.40 | 1.34 | 11 24 | 1 43 22  | Sos.avustukset vähentävät ihmisten omatoim.  |
|          | 2.80 | 1.19 | 10 46 | 4 33 7   |  |
| CONFLICT | 2.41 | 1.26 | 27 41 | 3 24 6   | Yhteistyö yrityksissä mahdotonta             |
|          | 2.52 | 1.07 | 10 58 | 6 22 4   |  |
| RESPECT  | 2.83 | 1.37 | 21 31 | 3 35 10  | Nuoret eivät kunnioita perinteisiä arvoja    |
|          | 2.56 | 1.04 | 9 55  | 8 25 2   |  |
| PUNISH   | 3.55 | 1.30 | 10 19 | 3 44 25  | Lainrikkojille ankarampia tuomioita          |
|          | 3.06 | 1.17 | 7 35  | 13 35 10 |  |
| DEATHPEN | 2.54 | 1.71 | 49 9  | 3 16 23  | Joihinkin rikoksiin kuolema sopivin ratkaisu |
|          | 2.22 | 1.35 | 40 31 | 3 17 9   |  |
| AUTHOBEY | 3.65 | 1.37 | 10 19 | 1 37 34  | Koulujen opetettava lapsia tottelevaisuuteen |
|          | 3.11 | 1.21 | 10 31 | 5 47 8   |  |
| LAWOBEY  | 3.05 | 1.45 | 18 29 | 1 33 19  | Lakia aina noudatettava (vaikka väärää)      |
|          | 2.55 | 1.12 | 14 50 | 6 27 3   |  |
| SEX      | 1.49 | 0.50 | 52 49 | 0 0 0    | Miespuolisuus (Nainen=1 Mies=2)              |
|          | 1.49 | 0.50 | 51 49 | 0 0 0    |  |

Olisi luonnollista verrata keskiarvoja toisiinsa ja luoda käsitystä siitä, miten nuorten asenteet ovat muuttuneet vuodesta 1988 vuoteen 1995. Parittaisten vertailujen asemasta, mihin liittyy omat vaaransa, olen soveltanut erotteluanalyysia.

### Erotteluanalyysi

- Tämä menetelmä etsii kahden vertailtavan ryhmän tapauksessa sen muuttujajohdistelmän, joka selvimmin tekee eroa ryhmien välille. Survossa erotteluanalyysi lasketaan äskeisten tulosten perusteella (huom. korrelaatiot tulevat mukaan olennaisena tekijänä) esim. /DISCRI-nimisellä sukrolla yksinkertaisesti näin:

```

8 1 SURVO 84C EDITOR Sun Aug 18 15:00:55 1996 C:\KIRJA\ 400 100 0
10 * .....
11 *CORR=R88,R95
12 *MSN=M88,M95
13 */DISCRI
14 * Eig.val. % Can.corr Chi^2 df P
15 * 1 0.611523 100.00 0.616010 612.9370 37 0.9999
16 *
17 *MAT LOAD DISCR.L.M,END+2 / Discriminant coefficients
18 *MAT LOAD DISCRXR.M,END+2 / Correlations variables/discriminators
19 *Correlations, means and standard deviations of discriminators
20 *for each of the 2 groups are saved in matrix files corresponding
21 *to CORR and MSN files with their names preceded by letter 'D'.
22 *Discriminant scores are computed by
23 *LINCO <data_file>,DISCR.L.M(D1,D2,...)
24 *

```

Rivillä 13 oleva /DISCRI-komento käyttää hyväkseen vertailtavien ryhmien CORR- ja MSN-matriiseja, jotka on tässä täsmennetty riveillä 11 ja 12. Jälleen suoraan toimituskenttään tulostuu vain vähän tietoja (rivit 14-23). Laajimmat tulostaulukot /DISCRI tallettaa omiin matriisitiedostoihinsa, joista kiinnostavimmat käyttäjä ottaa esiin toimituskenttään valmiilla MAT LOAD-komennoilla.

- Kahden vertailtavan ryhmän tapauksessa ryhmiä erottelevia uusia muuttujia voi syntyä vain yksi. Sitä koskevat perustulokset rivillä 15 osoittavat, että ryhmät eroavat erittäin selvästi toisistaan tilastollisesti.

Olen aktivoinut rivin 18, jolloin saan toimituskenttään tämän erottelumuuttujan ja alkuperäisten väliset korrelaatiokertoimet. Olen liittänyt jälleen muuttujien nimien selitykset mukaan ja sitten järjestänyt muuttujat korrelaatioiden suuruuden mukaan SORT-komennolla.

```

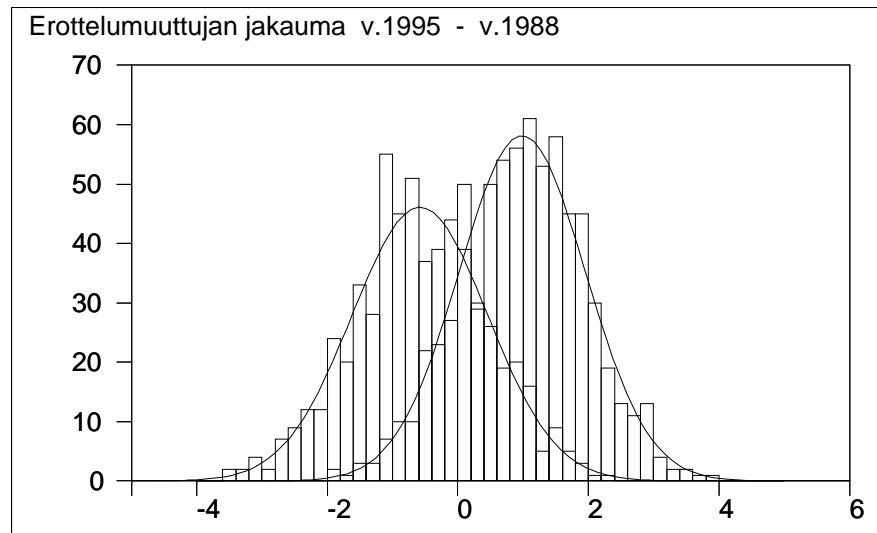
ONLYVOTE -0.37547 Äänestäminen ainoa keino vaikuttaa
OWNFEET -0.37459 Sos.avustukset vähentävät ihmisten omatoim.
VOTEINT -0.36441 Aion käyttää äänioikeuttani
AUTHOBEY -0.33108 Koulujen opetettava lapsia tottelevaisuuteen
PUNISH -0.31812 Lainrikkokojille ankarampia tuomioita
LAWOBEY -0.30919 Lakia aina noudatettava (vaikka väärää)
PSPORT -0.23255 Ylpeä urheilusaavutuksista
TRUST -0.18526 Luottamus nuorten ongelmien ratkaisuun
RESPECT -0.17831 Nuoret eivät kunnioita perinteisiä arvoja
DEPEND -0.17786 Hyvinvointivaltiossa ei huolehdi itsestä
DEATHPEN -0.16862 Joihinkin rikoksiin kuolema sopivin ratkaisu
PSCIENCE -0.12169 Ylpeä tieteellisistä saavutuksista
PWELFARE -0.10697 Ylpeä sosiaaliturvasta ja terv.huollosta
PROTECT -0.10431 Poliitikot suojelevat toisiaan
INTEREST -0.10352 Kiinnostunut politiikasta
RECKLESS -0.08851 Poliitikot häikäilemättömiä ja julk.kipeitä
NOPEOPLE -0.02862 Puolueet kiinnostuneita vain äänistä
SEX 0.00936 Miespuolisuus (Nainen=1 Mies=2)
PLAYERS 0.03910 Poliitikot "pelaavat varman päälle"
PPOLSYST 0.06832 Ylpeä eduskunnasta ja pol.järjestelmästä
PECONOMY 0.06941 Ylpeä taloudellisista saavutuksista
COMPLIC 0.07077 En ymmärrä politiikkaa
FREERIDE 0.07357 Useimmat tuensaaajat eivät sitä ansaitse

```

|          |         |   |
|----------|---------|---|
| STABLE   | 0.07384 | Vakiintunut puoluekanta                       |
| POLYSYST | 0.07440 | Poliittinen järjestelmä toimii                |
| CONFLICT | 0.07504 | Yhteistyö yrityksissä mahdotonta              |
| RICHPRIV | 0.07965 | Rikkaille eri lait kuin köyhille              |
| STRENGTH | 0.09416 | Vakiintunut oma poliittinen käsitys           |
| WASTEMON | 0.09434 | Poliitikot tuhlailevat                        |
| MUCHSAME | 0.09913 | Asiat jatkuvat hallituspuolueista riippumatta |
| PARTS    | 0.13649 | Ylpeä taiteellisista saavutuksista            |
| FAKES    | 0.13991 | Poliitikot eivät anna oikeaa kuvaa itsestään  |
| EXPLOIT  | 0.22550 | Suromistajat rikastuvat työl.kustannuksella   |
| REFUGEE  | 0.27454 | Otettava vastaan enemmän pakolaisia           |
| OWNFAULT | 0.30402 | Työttömyyden syyt ihmisessä itsessään         |
| DEVELAID | 0.33353 | Kehitysapua on lisättävä                      |
| LEASH    | 0.37379 | Poliitikot puolueensa "liekanuorassa"         |

Listan ääripäät ovat mielenkiintoisia. Alussa ovat ne asenteet, jotka ovat ominaisempia vuoden 1995 nuorille ja loppupäässä ne, jotka kuvaavat vuoden 1988 nuoria. Selvästi näkyy, että asenteet vähäosaisia mm. työttömiä ja pakolaisia kohtaan ovat nykynuorilla koventuneet. Myös halu käyttää äänioikeutta on lisääntynyt.

Erot näkyvät selvästi myös piirtämällä erottelumuuttujan frekvenssijakaumat samaan kuvaan, missä vuoden 1995 nuoret sijoittuvat vasemmalle:



Nämä ryhmät erottuvat yli 70%:n varmuudella, sillä kun kaikki havainnot luokitellaan Survon CLASSI-komennolla erottelumuuttujan perusteella, virheluokitusten määrä jää kummassakin suunnassa noin 23 prosenttiin.

## Faktorianalyysi

- Tarkastelen asennemuuttujien välisiä suhteita faktorianalyysin avulla. Perustieto sisältyy 37 x 37 -korrelaatiomatriisiin, josta riippuvuuksien ulosmittaus olisi välittömästikin mahdollista. Faktorianalyysi kuitenkin tiivistää tietoa huomattavan tehokkaasti.

Analyysin luonne on muihin monimuuttujamenetelmiin verrattuna poikkeuksellinen. Kun mm. erotteluanalyysissa etsitään suoraan hyviä muuttujien lineaarikombinaatioita, faktorianalyysissa lähdetään liikkeelle siitä ajatuksesta, että tutkittavan ilmiön taustalla on joitain yleisiä, ilmiötä hallitsevia tekijöitä (faktoreita), joita ei kyetä välittömästi havaitsemaan eikä mittaamaan. Ne vain näyttäytyvät osatekijöinä havaittavissa tiedoissa. Kunkin muuttujan vaihtelun ajatellaan jakautuvan kahteen osaan, yhteisvaihteluun ja ominaisvaihteluun. Yhteisvaihtelua kuvataan piilossa lymyävien faktorien avulla. Ominaisvaihtelua katsotaan kohinaksi, mitä ei tarvitse ottaa lukuun.

Selvittääksemme jotain tuosta piilorakenteesta noilla 37 asennemuuttujalla teen Survolla faktorianalyysin vuoden 1995 aineistolla. Olen esitarkasteluilla päätenyt siihen, että faktoreita löytyy viisi.

```
18 1 SURVO 84C EDITOR Mon Aug 19 17:28:44 1996 C:\KIRJA\ 400 100 0
1 *
2 *CORR NUORI95
3 *FACTA CORR.M,5
4 *ROTATE FACT.M,5,6_
5 *
```

Aluksi lasketaan korrelaatiomatriisi CORR.M (rivi 2). Seuraavaksi etsitään faktoriratkaisu suurimman uskottavuuden menetelmällä korrelaatiomatriisista 5 faktorilla. Tuloksena saadaan 37 muuttujan ja 5 faktorin välisten korrelaatiokertoimien muodostama faktorimatriisi FACT.M (rivi 3).

- Faktorianalyysin toinen erikoisuus on siinä, ettei tulos ole yksikäsitteinen, vaan siihen jää ns. rotaatiomahdollisuus. Alkuperäistä ratkaisua FACT.M voi kiertää faktorien virittämässä koordinaatistossa (tässä siis 5-ulotteisessa) niin, että tulos näyttäytyy tulkinnan kannalta edullisessa muodossa. Faktorirotaatioita varten on eri mahdollisuuksia. Niistä perinteisin on graafinen rotaatio, jota tietokonekaudella on käytetty vähän, sillä sitä on ollut hankala automatisoida. Tilalle ovat tulleet jo 1950-luvulla ns. analyttiset menetelmät, joissa silmän korvaa usean muuttujan funktion optimointitehtävä.

Ohjelmoin graafisen rotaatoratkaisun Survoon 1970-luvun lopulla. Se tehdään kuvaruudussa vaiheittain, kuten perinteeseen kuuluu. Kussakin vaiheessa tutkija näkee faktoriavaruuden jonkin 2-ulotteisen projektion ja voi näppäinten

avulla osoittaa haluamansa kiertokulman. Käymällä läpi - jopa toistuvasti - kaikki faktoriyhdistelmät syntyy lopullinen rotaattoriratkaisu.

Rotaatio tapahtuu Survossa ROTATE-komennolla. Jos siihen liitetään täsmennys ROTATION=Graphical, siirrytään graafiseen rotaatioon ja avaimet ovat tutkijalla. Muilla ROTATION-täsmennyksillä valitaan jokin analyttisistä menetelmistä. Jos ei mainita mitään, kuten tässä tapauksessa, käytetään ns. Varimax-menetelmää, joka on kaikkein tavallisin.

En ole ottanut näkyville ratkaisua alkuperäisessä muodossa vaan muokannut sitä Survolle ominaisin keinoin luettavampaan asuun. Lopullinen faktorimatriisi näyttää tältä:

|            | F1           | F2          | F3          | F4          | F5           | $h^2$ |   |
|------------|--------------|-------------|-------------|-------------|--------------|-------|---|
| PLAYERS    | <b>0.63</b>  | 0.08        | -0.08       | -0.04       | -0.10        | 0.42  | Poliitikot "pelaavat varman päälle"           |
| NOPEOPLE   | <b>0.62</b>  | 0.11        | -0.16       | -0.13       | -0.19        | 0.48  | Puolueet kiinnostuneita vain äänistä          |
| WASTEMON   | <b>0.58</b>  | -0.01       | -0.05       | -0.03       | 0.19         | 0.37  | Poliitikot tuhlailevat                        |
| POLSYST    | <b>-0.53</b> | -0.02       | 0.10        | -0.04       | 0.11         | 0.31  | Poliittinen järjestelmä toimii                |
| RECKLESS   | <b>0.53</b>  | 0.01        | -0.01       | -0.03       | 0.06         | 0.29  | Poliitikot häikäilemättömiä ja julk.kipeitä   |
| RICHPRIV   | <b>0.48</b>  | -0.02       | 0.02        | 0.05        | 0.11         | 0.25  | Rikkaille eri lait kuin köyhille              |
| FAKES      | <b>0.46</b>  | 0.05        | -0.12       | -0.08       | 0.04         | 0.24  | Poliitikot eivät anna oikeaa kuvaa itsestään  |
| EXPLOIT    | <b>0.46</b>  | -0.06       | -0.03       | 0.01        | 0.12         | 0.23  | Suromistajat rikastuvat työlkustannuksella    |
| TRUST      | <b>-0.45</b> | -0.00       | 0.15        | 0.02        | 0.13         | 0.25  | Luottamus nuorten ongelmien ratkaisuun        |
| PROTECT    | <b>0.43</b>  | -0.06       | 0.08        | -0.05       | -0.01        | 0.20  | Poliitikot suojelevat toisiaan                |
| CONFLICT   | <b>0.42</b>  | 0.05        | -0.12       | 0.09        | 0.13         | 0.22  | Yhteistyö yrityksissä mahdotonta              |
| MUCHSAME   | <b>0.40</b>  | 0.12        | -0.31       | -0.05       | 0.02         | 0.27  | Asiat jatkuvat hallituspuolueista riippumatta |
| LEASH      | <b>0.40</b>  | -0.05       | -0.06       | -0.14       | -0.19        | 0.22  | Poliitikot puolueensa "liekanuorassa"         |
| PPOLSYST   | <b>-0.34</b> | -0.12       | 0.13        | 0.10        | 0.05         | 0.16  | Ylpeä eduskunnasta ja pol.järjestelmästä      |
| OWNFEET    | 0.03         | <b>0.71</b> | -0.01       | -0.17       | 0.15         | 0.55  | Sos.avustukset vähentävät ihmisten omatoim.   |
| DEPEND     | 0.02         | <b>0.58</b> | 0.07        | -0.04       | 0.04         | 0.34  | Hyvinvointivaltiossa ei huolehdi itsestä      |
| OWNFAULT   | -0.09        | <b>0.55</b> | -0.11       | -0.08       | 0.16         | 0.35  | Työttömyyden syyt ihmisessä itsessään         |
| FREERIDE   | 0.09         | <b>0.44</b> | -0.04       | -0.08       | 0.27         | 0.28  | Useimmat tuensaaajat eivät sitä ansaitse      |
| STRENGTH   | -0.07        | 0.07        | <b>0.76</b> | -0.02       | -0.10        | 0.60  | Vakiintunut oma poliittinen käsitys           |
| STABLE     | -0.11        | 0.03        | <b>0.76</b> | 0.02        | 0.02         | 0.59  | Vakiintunut puoluekanta                       |
| INTEREST   | -0.21        | -0.08       | <b>0.42</b> | -0.04       | -0.15        | 0.25  | Kiinnostunut politiikasta                     |
| VOTEINT    | -0.25        | -0.06       | 0.24        | 0.03        | 0.16         | 0.15  | Aion käyttää äänioikeuttani                   |
| DEVELAID   | 0.01         | -0.06       | -0.05       | <b>0.70</b> | -0.04        | 0.50  | Kehitysapua on lisättävä                      |
| REFUGEE    | -0.18        | -0.04       | 0.07        | <b>0.68</b> | <b>-0.33</b> | 0.61  | Otettava vastaan enemmän pakolaisia           |
| SEX        | -0.09        | 0.16        | 0.12        | -0.27       | -0.18        | 0.15  | Miespuolisuus (Nainen=1 Mies=2)               |
| DEATHPEN   | 0.14         | 0.13        | 0.01        | -0.27       | 0.27         | 0.18  | Joihinkin rikoksiin kuolema sopivin ratkaisu  |
| PUNISH     | 0.22         | 0.17        | -0.01       | -0.15       | <b>0.38</b>  | 0.24  | Lainrikkojille ankarampia tuomioita           |
| COMPLIC    | <b>0.33</b>  | -0.02       | -0.21       | 0.13        | <b>0.38</b>  | 0.31  | En ymmärrä politiikkaa                        |
| AUTHOBEGY  | 0.04         | 0.18        | -0.01       | -0.10       | <b>0.34</b>  | 0.16  | Koulujen opetettava lapsia tottelevaisuuteen  |
| PSPORT     | -0.05        | 0.06        | -0.04       | 0.06        | <b>0.34</b>  | 0.13  | Ylpeä urheilusaavutuksista                    |
| RESPECT    | 0.10         | 0.26        | 0.03        | -0.01       | <b>0.34</b>  | 0.19  | Nuoret eivät kunnioita perinteisiä arvoja     |
| ONLYVOTE   | -0.04        | -0.02       | 0.03        | -0.03       | 0.26         | 0.07  | Äänestäminen ainoa keino vaikuttaa            |
| LAWOBEGY   | -0.10        | 0.12        | -0.09       | -0.04       | 0.22         | 0.08  | Lakia aina noudatettava (vaikka väärää)       |
| PWELFARE   | -0.07        | -0.18       | -0.04       | 0.12        | 0.21         | 0.10  | Ylpeä sosiaaliturvasta ja terv.huollosta      |
| PECONOMY   | -0.13        | 0.05        | -0.01       | 0.04        | 0.20         | 0.06  | Ylpeä taloudellisista saavutuksista           |
| PSCIENCE   | -0.18        | -0.00       | -0.02       | 0.03        | 0.10         | 0.04  | Ylpeä tieteellisistä saavutuksista            |
| PARTS      | -0.05        | -0.14       | 0.04        | 0.26        | 0.07         | 0.10  | Ylpeä taiteellisista saavutuksista            |
| neliösumma | 3.77         | 1.65        | 1.71        | 1.36        | 1.44         | 9.94  |   |



Matriisi sisältää 37 havaitun muuttujan ja 5 faktorin väliset 185 korrelaatio-kerrointa eli latausta. Tämä on jo sellaisenaan mukavampaa katsottavaa kuin korrelaatiomatriisin  $37 \cdot 36/2 = 666$  (pedon luku!) erilaista korrelaatiokerrointa. Koska faktorimatriisissa varsinkin hyvän rotaatoratkaisun kautta kuvastuu tietty rakenteen yksinkertaisuus, suurin osa näistä latauksista on käytännöllisesti katsoen nolliä. Kun otetaan vain edellä tehostetusti merkityt lataukset huomioon, rakenteen kuvaamiseen tarvitaan alle 40 lukua.

Ryhtymättä mihinkään perusteelliseen pohdintaan, totean vain, että faktoreille sopinevat esim. seuraavat tulkinnat:

- F1 Poliitikkovastaisuus
- F2 Hyvinvointivaltion kritiikki
- F3 Poliittinen sitoutuneisuus
- F4 "Pehmeät arvot"
- F5 Kovaluontoisuus

Jos samanlainen analyysi tehdään vuoden 1988 aineistolle, saadaan yllättävän yhtäpitävä faktorirakenne. Tätä ei yleensä voida todeta suoraan faktorimatriiseja vertailemalla, sillä rotaatio voi toisessa tapauksessa viedä ratkaisua hiegan erilaiselta näyttävään muotoon. Faktoritulosten vertailuun tarvitaankin omia menetelyjä, joista varsinkin Suomessa on käytetty *Yrjö Ahmavaaran* alunperin luomaa transformaatioanalyysia. Olen itse osallistunut tämän menetelmän jatkokehitykseen ja viime vuosina tutkinut mahdollisuuksia tulosten tilastolliseen vertailuun satunnaistamisen avulla.

## Vaihtelun vuoksi

- Monimuuttujamenetelmien teoriassa tavallisen yksiulotteisen normaalijakauman yleistys, multinormaalijakauma näyttölee keskeistä osaa. Vaikka hyvin tiedetään, että useimmat aineistot eivät tarkalleen ole otoksia tuosta jakaumasta, menetelmät ovat sen verran karkeita, etteivät poikkeamat juuri häiritse.

$p$ -ulotteinen multinormaalijakauma syntyy yksinkertaisesti muodostamalla useista yksiulotteisista, riippumattomista normaalimuuttujista erilailta painotettuja summia  $p$  kappaletta. Tällaisesta konstruktiiivisesta määritelmästä johdetaan kaikki multinormaalijakauman ominaisuudet. On mm. helppo todeta se miellyttävä seikka, että muuttujien välillä ei voi vallita muita kuin lineaarisia riippuvuuksia. Siis korrelaatiokertoimet kertovat multinormaalijakaumassa riippuvuuksista kaiken, mikä on kertomisen arvoista. Kun tunnetaan korrelaatiomatriisin lisäksi muuttujien keskiarvot ja hajonnat, multinormaalijakauma on täysin määrätty.

Keskiarvot asettavat jakauman keskipisteen paikalleen  $p$ -ulotteisessa avaruudessa eikä niillä ole mitään merkitystä vaihtelun kannalta yhden otoksen ta-

pauksessa. Vaihtelua kuvaavat yksittäisten muuttujien hajonnat (tai niiden neiliöt, varianssit) ja korrelaatiomatriisi yhdessä. Kaikki vaihtelua koskeva tieto kootaan tavallisesti  $p \times p$  -kovarianssimatriisiksi, jonka lävistäjällä ovat varianssit. Muut alkioit, kovarianssit, saadaan kertomalla korrelaatiokertoimet aomuttujien hajonnoilla. Kovarianssimatriisia merkitään yleisesti kreikkalaisella kirjaimella  $\Sigma$ .

Kun  $p = 1$  eli palataan tavalliseen normaalijakaumaan, muuttujan varianssi vastaa kovarianssimatriisia ja mittaa hyväksyttävällä tavalla tilastollista vaihtelua. Aidosti moniulotteisissa tilanteissa sellaisen mittaluvun löytäminen, joka vastaisi varianssia on osoittautunut hankalaksi. Kovarianssimatriisi sen tekee, mutta se ei ole pelkkä luku vaan peräti  $p(p + 1)/2$  erilaisen luvun kokoelma eli siis oikea "vaihtelun vuoksi".

- ▣ Kysymys vaihtelun puristamisesta tai sanoisinko survomisesta yhteen lukuun on askarruttanut minua jo kauan. Yleisesti tähän tarkoitukseen on tiedossa ainoastaan kaksi mittalukua. Näistä toinen on niisanottu *kokonaisvarianssi*, joka on kovarianssimatriisin jälki eli lävistäjäalkioiden, varianssien summa. Se voidaan lausua myös kovarianssimatriisin ominaisarvojen  $\lambda_1, \lambda_2, \dots, \lambda_p$  summana

$$\text{tr } \Sigma = \lambda_1 + \lambda_2 + \dots + \lambda_p.$$

Toinen mittaluku on kovarianssimatriisin determinantti eli ominaisarvojen tulo

$$|\Sigma| = \lambda_1 \lambda_2 \dots \lambda_p,$$

jota sanotaan *yleistetyksi varianssiksi*.

Molemmilla mitoilla on vakavia heikkouksia, jotka yleisesti tunnetaan, mutta joista ei näy liiemmin keskustellun tilastotieteen piirissä. Kokonaisvarianssi ei ota lainkaan korrelaatioita huomioon. Yleistetty varianssi on päinvastoin yliherkkä korrelaatioille, sillä se surkastuu nolnaan, jos muuttujien välillä on voimakkaita lineaarisia riippuvuuksia.

- ▣ Seuraava 3-ulotteinen esimerkki paljastaa nämä puutteet. Olkoon kovarianssimatriisi muotoa

$$\Sigma = \begin{bmatrix} 1 & \rho & \rho \\ \rho & 1 & \rho \\ \rho & \rho & 1 \end{bmatrix}$$

missä muuttujien yhteinen korrelaatiokerroin  $\rho$  on hyvin korkea, sanokaamme  $\rho = 0.999$ . Tässä tapauksessa kokonaisvarianssi on itsepintaisesti 3, vaikka

muuttujien ollessa miltei täydellisesti toistensa kopioita järki sanoo, että vaihtelua siinä esiintyy himpun verran enemmän kuin 1, mikä on kunkin muuttujan varianssi.

Yleistetty varianssi puolestaan antaa arvon  $(1 + 2\rho)(1 - \rho)^2 = 2.998 \times 10^{-6} \approx 0$ , mikä vaikuttaa omituiselta. Yleensäkin käy niin, että mikäli tiettyyn muuttujajoukkoon, jolla yleistetty varianssi on siisti positiivinen luku, lisätään vielä yksi muuttuja, joka sattuu olemaan voimakkaasti lineaarisesti riippuvainen aikaisemmista, yleistetty varianssi tippuu nolnaan.

- ▣ Koska jotain parempaa mielestäni voisi löytyä, kulutin kesällä 1995 melkoisesti aikaa ja Survoa erilaisten vaihtoehtoisten ideoiden kehittelyyn. Monien vaiheiden jälkeen päädyin ehdottamaan mittalukua  $M\text{var}(\Sigma)$ , jonka lauseke on

$$M\text{var}(\Sigma) = \max \sum_{i=1}^p \sigma_{i,12\dots i-1}^2$$

missä  $\sigma_{i,12\dots i-1}^2$  tarkoittaa  $i$ :n muuttujan jäännösvariانسsia, kun aikaisemmat muuttujat vakioidaan ja maksimia etsitään muuttujien kaikkien permutaatioiden yli.

Jos tuo maksimointi hetkeksi unohdetaan, takana on ajatus, että jäännösvariانسseja lasketaan yhteen niin, että ensimmäisen muuttujan variانسsi otetaan sellaisenaan, sitten toisen muuttujan varianssin se osa, jota ei saada selitetyksi ensimmäisellä muuttujalla, sitten kolmannen muuttujan varianssin se osa, jota ei saada selitetyksi kahdella ensimmäisellä muuttujalla, jne.

Tästä sukeutuu kuitenkin ongelmia. Tuon summan arvo vaihtelee riippuen siitä, missä järjestyksessä muuttujat luetaan. Toistaiseksi paras vastaukseni on: Valitaan järjestyksistä se, joka antaa suurimman arvon.

Kun katselee tätä ehdotusta yksinkertaisten tapausten valossa, huomaa mitan käyttäytyvän järkevästi. Yhden muuttujan tapauksessa mitta yhtyy varianssiin. Mittayksikkönä on näin 0,1-normaalisen muuttujan variانسsi. Kun  $p = 2$ , mitalle tulee lauseke

$$M\text{var}(\Sigma) = \sigma_1^2 + (1 - \rho^2)\sigma_2^2,$$

missä  $\sigma_1$  ja  $\sigma_2$  ovat muuttujien hajonnat ja  $\rho$  on korrelaatiokerroin. Muuttujien järjestys valitaan niin, että  $\sigma_1 \geq \sigma_2$ . Esimerkiksi tapauksessa  $\sigma_1 = \sigma_2 = 1$  on yksinkertaisesti  $M\text{var}(\Sigma) = 2 - \rho^2$  ja tästä tulee 2, kun  $\rho = 0$  ja 1, kun  $|\rho| = 1$ .

Tuossa 3 muuttujan tilanteessa, jossa kaikki korrelaatiot olivat samoja eli  $\rho = 0.999$ , mitta saa luontevan arvon

$$\text{Mvar}(\Sigma) = 3 - \rho^2(3 + \rho)/(1 + \rho) \approx 1.0035.$$

$p$ -ulotteisessa tilanteessa pitää paikkansa

$$\text{Mvar}(\Sigma) = \sigma_1^2 + \sigma_2^2 + \dots + \sigma_p^2 = \text{tr } \Sigma,$$

kun muuttujat ovat korreloimattomia eli mitta yhtyy kokonaisvarianssiin ja täysin yleisesti pätevät rajat

$$\max(\sigma_1^2, \sigma_2^2, \dots, \sigma_p^2) \leq \text{Mvar}(\Sigma) \leq \sigma_1^2 + \sigma_2^2 + \dots + \sigma_p^2,$$

missä alaraja saavutetaan, kun kaikki korrelaatiot ovat  $\pm 1$ . Jos varianssit ovat ykkösiä,  $\Sigma$  on korrelaatiomatriisi ja

$$1 \leq \text{Mvar}(\Sigma) \leq p.$$

### Alkuperäinen ajatus

- ▣ Määritelmä, jonka edellä esitin Mvar-mitalle, ei ollut ensimmäisenä mielessäni. Koko juttu lähti liikkeelle halustani esittää multinormaalijakauma alussa mainitsemani konstruktiivisen määritelmän mukaan mahdollisimman taloudellisesti.

Tällöin katsoin, että lopulliset muuttujat  $X_1, X_2, \dots, X_p$  tulee luoda  $(0,1)$ -normaalisista, riippumattomista muuttujista  $U_1, U_2, \dots, U_p$  muodossa

$$\begin{aligned} X_1 &= c_{11}U_1 \\ X_2 &= c_{21}U_1 + c_{22}U_2 \\ &\dots \\ X_p &= c_{p1}U_1 + c_{p2}U_2 + \dots + c_{pp}U_p. \end{aligned}$$

Keskiarvot olen "unohtanut", sillä niillä ei ole vaihtelun mittaamisen kannalta mitään vaikutusta.  $X$ -muuttujat syntyvät siis astenevasti niin, että joka vaiheessa tulee yksi uusi  $U$ -muuttuja mukaan. Jos halutaan tehdä  $X$ :t taloudellisesti eli mahdollisimman pienellä määrällä "satunnaisenergiaa" - hyvin epätihteellinen termi - katsoin, että niiden kertoimien, joilla kukin  $U$  astuu kuvaan mukaan, tulee olla itseisarvoltaan suurempia kuin saman  $U$ :n myöhemmät (alla olevat) kertoimet. Kertoimien muodostamassa alakolmiomatriisissa  $\mathbf{C}$  päälävistäjän, jolla on luvut  $c_{11}, c_{22}, \dots, c_{pp}$ , tulee siis dominoida ja tällöin vain näillä kertoimilla on merkitystä vaihtelun määrää mitatessa.

Mielessäni oli erilaisia tapoja muodostaa tuo mitta noista lävistäjäkalkioista. Useimmat johtivat ristiriitaisiin tuloksiin. Ainoa, joka kesti, oli ottaa niiden neliosuma

$$c_{11}^2 + \dots + c_{pp}^2$$

ja maksimoida se hakemalla muuttujien paras järjestys.

Kesti häpeäkseni pari viikkoa ennenkuin huomasin, että kyseinen mitta syntyy suoraan kovarianssimatriisin Cholesky-hajotelmasta

$$\Sigma = \mathbf{C}\mathbf{C}^T,$$

sillä  $X$ :t voidaan luoda juuri tämän  $\mathbf{C}$ :n avulla ja niiden lävistjäalkioiden ne-liöt ovat suoraan Mvar-mitan määritelmässä mainittuja jäännösvariansseja. Sain siten tilastollisen tulkinnan alkuperäiselle heuristiselle lähestymistavalle.

### Laskennallinen puoli

- ▣ Cholesky-hajotelman avulla mitan arvo on askeltaen suhteellisen nopeasti las-kettavissa yksittäiselle järjestykselle. Kun maksimia etsitään, joudutaan käy-mään kuitenkin läpi  $p!=1\cdot2\cdot3\cdot\dots\cdot(p-1)p$  vaihtoehtoja, mikä on hirvuinen määrä, kun esim.  $p = 20$ . Silloisella PC:lläni tuolla muuttujamäärällä aikaa ratkaisun löytämiseen olisi kulunut lähes 400 miljoonaa vuotta!

Oli pakko löytää lyhyempi tapa. Riittävän hyväksi käytännössä on osoittautu-nut askeltava menettely. Siinä valitaan kerrallaan jokainen muuttujista ensim-mäiseksi. Valitulle muuttujalle etsitään muista kumppani, jonka jäännösvar-ianssi on suurin, kun sitä selitetään ensimmäisellä muuttujalla. Kolmanneksi valitaan se jäljellejäävistä, jonka jäännösvarianssi on suurin, kun sitä selite-tään kahdella aikaisemmalla jne.

Näin menetellen ei saada aina parasta ratkaisua, mutta simulointikokein olen selvittänyt, että askeltava tulos on riittävän lähellä oikeata käytännössä. Nyt ei tulosta 20 muuttujankaan tapauksessa tarvitse odottaa satoja miljoonia vuosia, vaan se selviää muutamassa sekunnissa. Tämä on huikempia nopeuden li-säyksiä, mitä olen koskaan kohdannut eteeni tulleissa laskentaongelmissa. Kannattaa siis joskus hieman harkita, millä tavalla laskee.

Mitan arvo lasketaan Survossa MULTVAR-komennolla. Esim. vuoden 1995 nuorten 37 muuttujan korrelaatiomatriisi antaa tuloksen

```

18 1 SURVO 84C EDITOR Wed Aug 21 10:03:39 1996 C:\KIRJA\ 400 100 0
95 * .....
96 *MULTVAR R88,CUR+1
97 Mvar[R88]=30.17549 (Total variability in a 37*37 matrix)
98 *MAT LOAD COVVAR.M,END+2 / Optimally permuted covariance matrix
99 *

```

Korrelaatiot siis pienentävät vaihtelua  $100\cdot(1-30.17549/37)=18.4$  prosentilla. Eräänlainen keskimääräinen korrelaatiokerroin voidaan laskea ottamalla ver-

tailukohteeksi korrelaatiomatriisi, jossa kaikki korrelaatiot ovat samat ja jossa Mvar-arvo on sama. Tämä vertailu antaa keskimääräiseksi korrelaatioksi noin 0.24. Tulos on hyvin samanlainen vuoden 1988 nuorilla.

Jo edellä sanotusta käynee ilmi, miten Survo on ollut avuksi monella tapaa ongelman pohdinnassa. Mvar-mitan muoto on matemaattisessa katsannossa lievästi sanoen epämukava vaikka periaatteessa yksinkertainen. Siksi olen joutunut eräiden kiinnostavien tapausten yhteydessä tyytymään laskennallisiin kokeisiin ja yrittänyt löytää niiden kautta säännönmukaisuuksia.

Kun tarkastellaan hyvin yksinkertaista erikoistapausta, missä  $X$ :t ovat  $U$ -muuttujien kasautuvia summia eli

$$\begin{aligned} X_1 &= U_1 \\ X_2 &= U_1 + U_2 \\ &\dots \\ X_p &= U_1 + U_2 + \dots + U_p \end{aligned}$$

olen Survon avulla "keksinyt" ja vahvistanut aina arvoon  $p = 90$  asti, että tässä tapauksessa mitan arvo on

$$\text{Mvar}(\Sigma) = p[\log_2 p/4 + 1 - \varepsilon(p)]$$

missä  $0 \leq \varepsilon(p) < 0.01$ . Erityisesti  $\varepsilon(p) \equiv 0$ , kun  $p$  on kakkosen potenssi. Minulla ei ole harmainta aavistustakaan, miten todistaisin tuloksen yleisesti. Tässäkin tapauksessa yleistetyn varianssin käy huonosti. Se on aina 1 dimensioista  $p$  riippumatta, vaikka vaihtelu kiistatta lisääntyy  $p$ :n kasvaessa.

### Periaatteellinen ongelma

- ▣ Mitalla Mvar on se tärkeä ominaisuus, joka puuttuu yleistetyltä varianssilta, että lisättäessä uusi muuttuja entisten jatkeeksi mitta ei ainakaan vähene arvostaan. Se kasvaa aidosti silloin, kun sitä sopii odottaakin.

Ongelmana vain on, että Mvar-mittaa ei voi lausua pelkästään kovarianssimatriisin ominaisarvojen avulla. Se ei ole invariantti muuttujien ortogonaalisessa muunnoksessa. Jos kuitenkin pidetään kiinni äsken mainitusta monotonisuusehdosta, olen voinut helposti osoittaa, ettei löydy mitään mitta, joka monotonisuuden lisäksi olisi invariantti edellä mainitussa mielessä.

Tämän olen kertonut (ilmeisesti vielä tänä vuonna ilmestyvässä) artikkelissani Computational Statistics & Data Analysis -lehdessä. Ehdotukseni tulee mitä ilmeisimmin herättämään kysymyksiä ja ehkä vastustustakin. Katson kuitenkin avanneeni keskustelun pitkään uinuneesta tilastollisesta ongelmasta. Odotan kiinnostuneena, keksiikö joku paremman vastauksen ja millainen se on.





## Kuvien takaa

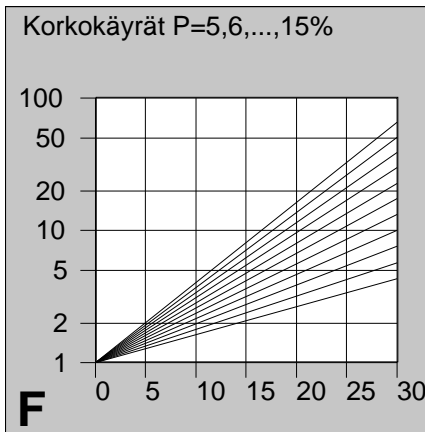
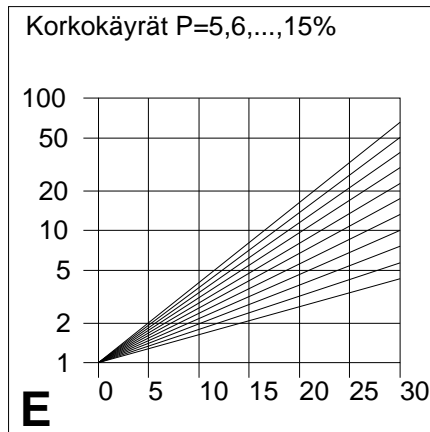
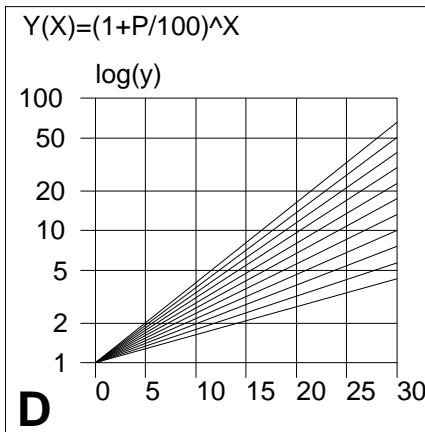
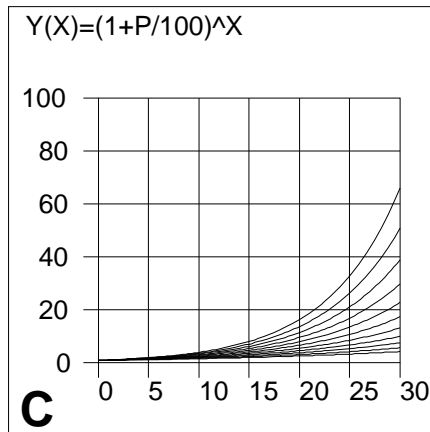
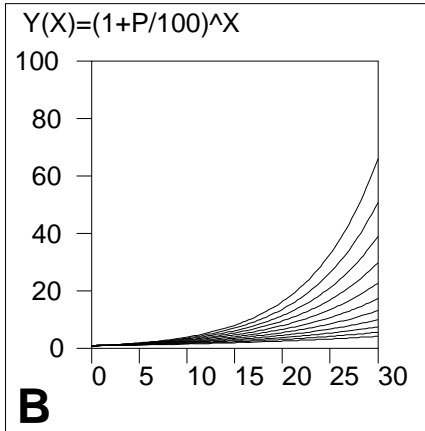
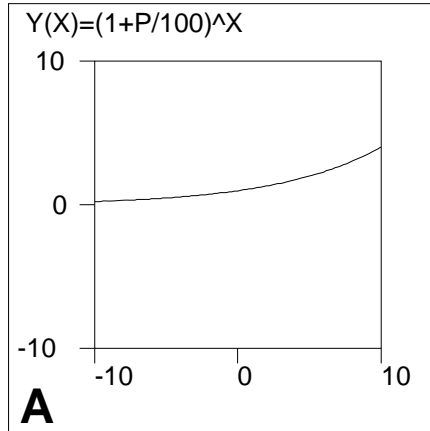
- Sanotaan, että "yksi kuva kertoo enemmän kuin tuhat sanaa". Rohkenen väittää, että Survo vuorostaan kertoo enemmän kuin tuhat kuvaa. Survoon kuuluu oma kuvakielensä, jonka perustana ovat pelkistetyt kuvatyypit. Kuvakieli sallii eri kuvatyypien vapaan yhdistelyn, värityksen ja koristelun. Näin syntyvät hyvinkin monimutkaisilta näyttävät graafiset esitykset. Kun katsoo vain lopullista tulosta, moni pelkää, että tuon täytyy olla vaikeata. On syytä kuitenkin aina uskotella itselleen, ettei ole olemassakaan mitään monimutkaisia asioita. Ne jotka sellaisilta näyttävät, ovat vain yksinkertaisten asioiden yhdistelmiä. Taito on siinä, että kääntää asiat toisinpäin ja löytää ne yksinkertaiset osat, joista haluttu kuva tulee koostumaan. Tämän taidon oppimisessa auttavat sekä valmiit mallit että kokeilemisen halu eli harhailu yrityksen ja erehdyksen mutkaisella polulla.

Tämän kirjan aikaisemmissa luvuissa on esitetty kuvia Survon keinoin. Useat noista kuvista on tehty Survon GPLOT-komennolla kuvaruutuun. Ne on tosin siirretty tähän paperille vastaavanlaisista PLOT-kaavioista PostScript-kirjoittimella. Kuvat viettävät Survossa eräänlaista kaksoiselämää. Tavallisesti kuvaa suunniteltaessa kannattaa käyttää GPLOT-komentoa, sillä piirrokset tulevat paljon nopeammin kuvaruutuun kuin paperille. Vasta lopullisemmassa muotoilussa otetaan PostScript käyttöön. Sitäkin voi jäljitellä kuvaruudussa ulkopuolisen Ghostscript-ohjelman avulla ja säästää kuluja erityisesti arvokkailla väritulostimilla.

Nyt tulen käsittelemään ainoastaan PostScript-grafiikkaa. PostScript on yleinen sivunkuvauskieli ja olen luonut sille Survoon liittymän, joka sallii kaikki temput, mitä PostScript itse osaa. Liittämällä PLOT-kaavioon täsmennyksen `DEVICE=PS` Survo ymmärtää ottaa tuon liittymän käyttöön.

Vaikka monenlaisia kuvia onkin jo tehty, vielä ei ole kuitenkaan piirretty käyriä ja käyräparvia, joita tarvitaan varsinkin tutkimussovelluksissa. Otan nyt korkokäyristä esimerkin, joka valaisee sitä, miten kuva kehittyy lopulliseen muotoonsa vaiheittain.





Muodonmuutokset näkyvät sarjakuvasta, jonka kuusi ruutua olen merkinnyt kirjaimin A - F.

- Lähtökohtana piirtämiselle käy tieto, että mikäli vuotuinen korkokanta on  $P$  prosenttia, niin 1 markka kasvaa  $X$  vuodessa summaksi, joka on

$$Y(X) = (1 + P/100)^X.$$

Arvolla  $P = 15$  kuvaa on ensiksi yritetty piirtää (ruutu A) PLOT-kaaviolla

```
22 1 SURVO 84C EDITOR Sat Aug 24 17:48:29 1996 C:\KIRJA\ 100 100 0
1 *
2 *P=15 SIZE=570,570 DEVICE=PS
3 *PLOT Y(X)=(1+P/100)^X
4 *
```

Aktivoidussa PLOT-komennessa kaava on ilmaistu sellaisenaan. Eksponentti on vain osoitettu merkin ^ avulla. Kuvan kooksi on määritelty 57 x 57 mm eli SIZE-täsmennyksessä yksikkönä on 0.1 mm. Koko on valittu niin, että kaksi tällaista kuvaa sopii sivulle vierekkäin.

Tulos on surkea. Koska muita tietoja ei ole annettu, Survo olettaa, että käyrä piirretään koordinaatistoon, jossa kumpikin kuva-alaan mahtuva akseli kulkee arvosta -10 arvoon 10. Tämä ikkuna on nyt liian suuri eli sitä pitää kaventaa antamalla sopivat XSCALE- ja YSCALE-täsmennykset. Samalla siirrytään yhdestä käyrästä käyräparveen, jossa korkokäyrät piirretään arvoilla  $P = 5, 6, \dots, 15$ . Lisäysten jälkeen kaavio on

```
22 1 SURVO 84C EDITOR Sat Aug 24 17:50:12 1996 C:\KIRJA\ 100 100 0
1 *
2 *P=5,15,1 SIZE=570,570 DEVICE=PS
3 *PLOT Y(X)=(1+P/100)^X
4 *XSCALE=0(5)30 siis 0:sta 5:n välein 30:een
5 *YSCALE=0(20)100
6 *
```

ja se synnyttää kuvan B. Piirroksen lisätään nyt ruudukko täsmennyksellä GRID=XY, jolloin kuva tulee muotoon C. Siitä huolimatta lukematarkeus varsinkin alle 20 vuoden ajalta on tässä esityksessä kovin heikko. Ratkaisevaa on siirtyä Y-akselilla logaritmiseen asteikkoon muuttamalla YSCALE-täsmennys muotoon

YSCALE=\*log(y) , 1, 2, 5, 10, 20, 50, 100,

jolloin saadaan kuva D. Eksponentiaalinen kasvu taittuu näin lineaariseksi ja luettavuus paranee olennaisesti. Seuraavaksi kohennetaan piirroksen tekstejä täsmennyksin

HEADER=Korkokäyrät\_P=5;6;...;15% YLABEL=

eli alkuperäinen otsikko (käyrän yhtälö) korvataan sisältöä kuvaavalla. Samalla poistetaan logaritmistä asteikkoa osoittava merkintä Y-akselin päästä paneamalla YLABEL tyhjäksi (kuva E).

Lopulta kaaviota somistetaan vielä harmaalla reunuksella maalaamalla ensin koko kuva-alue harmaaksi ja palauttamalla sitten varsinainen piirrosalue valkoiseksi. Tämä tapahtuu FRAMES-täsmennyksillä ja lopullinen kuva F on siis tullut PLOT-kaaviosta

```

22 1 SURVO 84C EDITOR Sat Aug 24 17:52:56 1996 C:\KIRJA\ 100 100 0
1 *
2 *P=5,15,1 SIZE=570,570 DEVICE=PS
3 *PLOT Y(X)=(1+P/100)^X
4 *XSCALE=0(5)30
5 *YSCALE=*log(y),1,2,5,10,20,50,100
6 *GRID=XY XDIV=120,400,50 YDIV=100,350,120
7 *HEADER=Korkokäyrät P=5;6;...;15% YLABEL=
8 *FRAMES=F1,F2 F1=0,0,570,570,2 F2=120,100,400,350,0
9 *

```

FRAMES-täsmennyksellä nimetään kaksi suorakaidetta F1 ja F2. Kumpikin määritellään omalla opasteellaan antamalla vasemman alakulman koordinaatit, leveys, korkeus ja harmauden aste, jolloin 0 tarkoittaa valkoista. Kuvissa C - F on lisäksi käytetty XDIV- ja YDIV-täsmennyksiä (rivi 6). Ne määrittelevät tarkasti, miten koko kuva-ala jakautuu suhteellisesti reuna-alueisiin ja itse piirrosalueeseen. Yksityiskohtaista tietoa löytyy Survosta aloittamalla kyselysanalla PLOT?.

Lopullisesta kuvasta mm. selviää, että pääoman kaksinkertaistumiseen kuluu 5%:n korolla noin 15 vuotta, mutta 15%:n kiskurikorolla vain 5 vuotta.

## Olympiarenkaat

- Survo ei kilpaile "vapaan käden" piirrosohjelmien kanssa, koska esim. käyrien piirtoa ohjataan yhtälöillä tai Survon havaintoaineistojen avulla. Jouko Manninen on tosin rakentanut taidokkaasti Survon päälle omia hiirellä ohjattavia graafisia toimintoja. Survon kuvakieli järjestelmällisyydessään johtaa usein vaativissa tehtävissä helpommin hallittaviin ja muunneltaviin ratkaisuihin kuin mihin tavanomaiset piirrosohjelmat yltävät.

Varmasti moni piirtää olympiarenkaat omalla mielihajelmallaan helpommin kuin miten ne nyt sommitellaan. Survon edut tulevat näkyville vasta mutkikkaammissa tilanteissa, joista kerron tämän jälkeen. Teen nyt olympiarenkaat vain näyttääkseni pienoiskoossa, miten Survossa toimitaan.

- Ensimmäinen täytyy osata piirtää ympyröitä. Tämän vuoksi näytän piirroskaavion, joka tuottaa Kolmogorovin paradoksin yhteydessä esitetyn, viiden toisiaan sivuavan ympyrän yhdistelmän.

```

39 1 SURVO 84C EDITOR Sun Aug 25 10:04:38 1996 C:\KIRJA\ 100 100 0
1 *
2 *PLOT X(T)=X0+R*cos(T),Y(T)=Y0+R*sin(T)
3 *
4 *DATA YMPYRÄT
5 *X0 Y0 R
6 * 0 0 0.41421 sqrt(2)-1=0.4142135623731
7 * 1 1 1
8 * -1 1 1
9 * -1 -1 1
10 * 1 -1 1
11 *
12 *DEVICE=PS
13 *X0=DATA:YMPYRÄT,X0 Y0=DATA:YMPYRÄT,Y0 R=DATA:YMPYRÄT,R
14 *T=0,2*pi,pi/40 pi=3.1416
15 *SCALE=-2,0,2 SIZE=700,700 DEVICE=PS,YMP5.PS
16 *

```

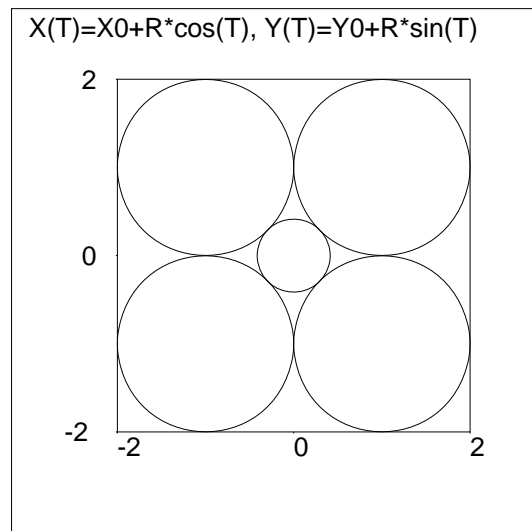
Ympyrät kannattaa piirtää aina parametriesityksessä

$$x(t) = x_0 + r \cos(t),$$

$$y(t) = y_0 + r \sin(t),$$

missä  $r$  on säde,  $(x_0, y_0)$  keskipiste ja  $t$  vaihekulma, joka kulkiessaan välin  $[0, 2\pi]$  piirtää koko ympyränkaaren.

Kaaviossa ko. parametriesitys ilmoitetaan sellaisenaan itse PLOT-komennon yhteydessä. Parametrit  $x_0$ ,  $y_0$  ja  $r$ , jotka vaihtelevat, on annettu havaintotaulukkona YMPYRÄT (rivit 4-10) ja ne sidotaan käyrän parametreihin rivin 13 täsmennyksillä. Jokainen havainto tuottaa oman ympyränsä. Rivillä 14 määrätään vaihekulma  $t$  kulkemaan väli  $[0, 2\pi]$  askelin  $\pi/40$  eli itse asiassa Survo piirtää tässä ympyröiden asemasta säännöllisiä 80-kulmioita. Saa olla tarkkanäköinen, että ne ympyrästä erottaa. Näin syntyy piirros



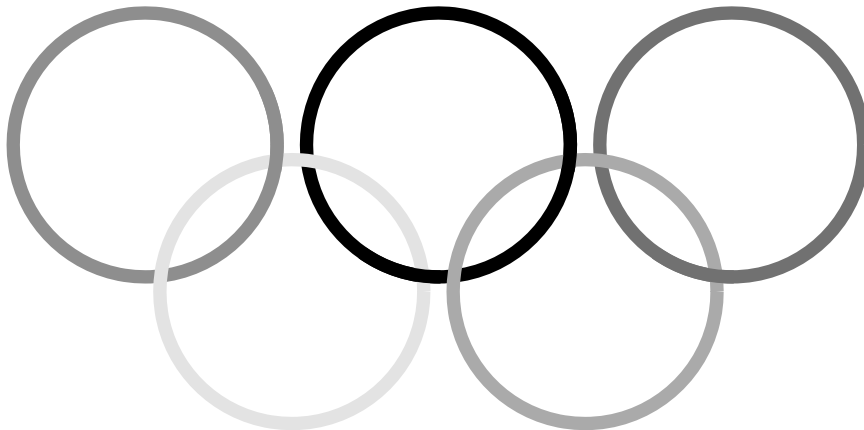
Täsmennyksillä HEADER= ja FRAME=0 saataisiin kaikki turhat apuviivat ja tekstit pyyhityksi niin, että pelkät ympyrät vain jäävät. Siirryn nyt kuitenkin suoraan olympiarengaskaavioon. Se on äskeisen välitön yleistys.

```

45 1 SURVO 84C EDITOR Sun Aug 25 10:28:47 1996 C:\KIRJA\ 100 100 0
18 *
19 *Olympiarengaat
20 *DATA OLYMPIA
21 * A B C D E Selitys
22 * -2 0 4 1 0 Sininen A: Keskipisteen X-koord.
23 * 0 0 9 1 0 Musta B: Keskipisteen Y-koord.
24 * 2 0 5 1 0 Punainen C: Väri
25 * -1 -1 1 1 0 Keltainen D: Kaarenpituus
26 * 1 -1 3 1 0 Vihreä E: Kaaren lähtökulma
27 * -2 0 4 0.1 -0.2 Sininen yli keltaisen (paikkaus)
28 * 0 0 9 0.1 4.1 Musta yli keltaisen (paikkaus)
29 * 0 0 9 0.1 -0.2 Musta yli vihreän (paikkaus)
30 * 2 0 5 0.1 4.1 Punainen yli vihreän (paikkaus)
31 *
32 *PLOT X(T)=A+R*COS(D*T+E), Y(T)=B+R*SIN(D*T+E)
33 *DEVICE=PS
34 *R=0.90
35 *T=[line_width(5)],0,2*pi,pi/20 pi=3.1416
36 *A=DATA:OLYMPIA,A B=DATA:OLYMPIA,B D=DATA:OLYMPIA,D E=DATA:OLYMPIA,E
37 *XSCALE=-3,3 YSCALE=-2,1 SIZE=1160,580
38 *COLOR_CHANGE=C C=DATA:OLYMPIA,C FRAME=0 XDIV=0,1,0 YDIV=0,1,0 HEADER=
39 *

```

Kaavio pitää huolta siitä, että kyseessä ovat todelliset renkaat, joiden paksuus ilmaistaan [line\_width(5)]-ohjauksella (rivi 35) ja jotka todella linkittyvät oikein. Tämä hoidetaan pienillä ympyränkaaren paikkauksilla (viimeiset havainnot). Esim. laitimmekin vasemmalla kulkee viereisen ylitse ja alitse. Täsmennys COLOR\_CHANGE pitää huolta värin muutoksista. Tässä värin asemasta on käytetty harmaan eri asteita. Värikuvassa värit määriteltäisiin negatiivisin indekseihin esim. niin, että täsmennyksellä [FILL-1]=0,0,1,0 arvo -1 C-sarakkeessa tarkoittaisi puhdasta keltaista. Survossa on PostScript-kielen mukaisesti käytössä CMYK-väritysjärjestelmä eli värikomponentteja on 4 (Cyan, Magenta, Yellow, Black). Ne voivat vaihdella jatkuvasti nollasta ykköseen, jolloin hyvillä laitteilla väriyhdistelmiä syntyy erittäin paljon.



Muistan, että *Simo Puntanen* aikoinaan ehdotti mahdollisuutta antaa havaintotiedoston arvojen vaikuttaa käyrien parametriin. Hän tarvitsi tätä ominaisuutta tilastollisissa piirroksissa eri otoksista laskettujen tunnuslukujen vaihtelun graafiseen esittämiseen. Olen hyvin kiitollinen hänen ideastaan, sillä se on rikastuttanut Survon kuvakieltä monissa muissakin yhteyksissä.

## Kuvia kerroksittain

- ▣ Survossa luonnollinen tapa kehittää monimuotoisia piirroksia on koota ne erillisistä kuvista. Useimmiten kuvia yksinkertaisesti kerrostetaan asettamalla ne lopullisessa julkitulossa päällekkäin. Tällöin osakuvat piirretään aluksi omiin tiedostoihinsa. Itse tiedostoja saatetaan yhdistää ja luoda niistä EPS-standardin mukaisia. Näitä kuvatiedostoja myös muut julkaisuohjelmat pystyvät painamaan.

Seuraavassa kuvassa on piirretty kaksi Survolla tehtyä kuvaa toistensa päälle. Kuvassa on 48 havaintopisteen hajontakuva survoilijoille tutun kymmenotteluaineiston urheilijoiden pituuden ja painon riippuvuudesta. Muuttujien korrelaatiokerroin on noin 0.85. Pohjalle on kuitenkin asetettu korrelaatiokertoimen herkkyyttä poikkeaville havainnoille kuvaava käyrästä. Se on piirretty rasterikuvana, josta ilmenevät muotoa  $z = f(x,y)$  olevaa funktiota vastaavan pinnan vaihtelut korkeuskäyrästä. Tässä tapauksessa  $x$  = pituus,  $y$  = paino ja funktion arvo  $z$  kuvaa kuinka paljon korrelaatiokertoimen arvo muuttuu, jos saadaan uusi havainto  $x,y$ .

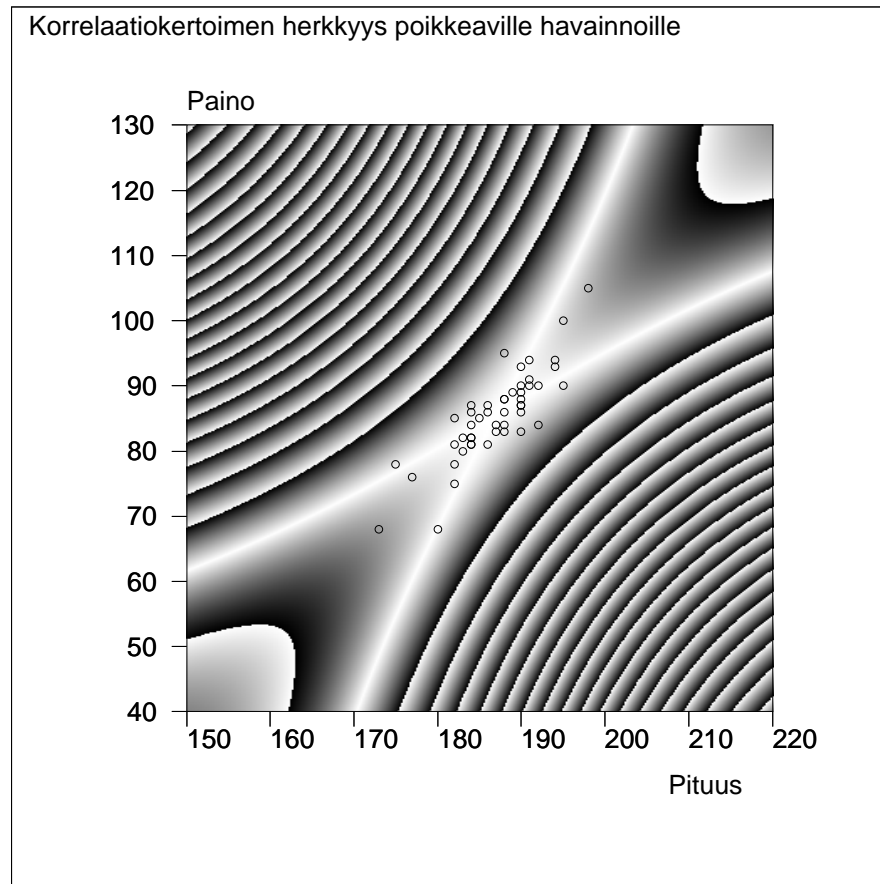
En rasita lukijaa piirroskaavioilla, sillä ne löytyvät mm. englanninkielisen Survo-kirjani (1992) sivulta 189. Herkkyyssäyrästä kuvataan siten, että eri tummuusarvot vastaavat eri korkeustasoja. Korrelaatiokertoimen 0.05-kokoista muutosta vastaa kuvassa aina yksi vyöhyke valkoisesta mustaan. Näin valkea, hieman käyristynyt ristikko kuvan keskellä on alue, jonne osunut uusi havainto ei muuttaisi juuri lainkaan korrelaatiokertoimen arvoa. Oikealle alas ja vasemmalle ylös syntyy vyöhykkeitä, jotka osoittavat kertoimen alenemista. Voi se jopa kasvaakin, kun mennään esim. tarpeeksi oikealle ylös. Siis laihe-liini, jonka pituus on 190 cm ja paino vain 40 kg alentaishi korrelaatiokerrointa 6 vyöhykkeen verran eli arvoon  $0.85 - 6 \times 0.05 = 0.55$ .

Yhdistetty kuva on painettu PRINT-listassa olevilla ohjausriveillä

```
% 1200  
- picture HERKKYYS.PS  
- picture KYMMEN.PS
```

missä ensimmäinen (% kontrollisarakeessa) tekee tilaa 12 cm korkuiselle kuralle ja kaksi muuta riviä liittävät kuvat tekstiin päällekkäin. Jos niitä halut-

taisiin asetella tästä poikkeavasti, tiedostonimien perään kirjoitetaan koordinaatit, mahdolliset kutistus- ja venytyskertoimet ja jopa kallistuskulma.



Minulla oli esitelmä tilastollisen tietojenkäsittelyn COMPSTAT-kokouksessa Neuchâtelissa Sveitsissä vuonna 1992. Tämän joka toinen vuosi järjestettävän kokouksen käytäntöön kuuluu, että kaikki esitelmät julkaistaan kirjoina edeltäpäin, jolloin osanottajat saavat kirjat käteensä kokoukseen saapuessaan. Yllätyksekseni kokouksen järjestäjät olivat ottaneet esitelmästäni tämän kuvan keskiosan kirjojensa kansikuvaksi. Iloani hämmensi vain hieman, että olivat pahukset kääntäneet sitä "taiteellisista syistä" 90 astetta.

## Ääniä paperilla

- Usein minulta kysytään, mikä yhdistää musiikkia ja matematiikkaa. Tavallisesti vastaan, ettei mikään, sillä tunnen yhtä lailla musikaalisia ja epämusikaalisia matemaatikkoja. Vastaavasti muusikkojen joukosta löytyy sekä matematiikasta kiinnostuneita että sitä kammoavia. Saattaahan sitä kuitenkin ajatella, että yhdistäviä tekijöitäkin on. Matematiikka tieteenalana ja musiikki taiteessa ovat puhtaimmalla muodollaan kumpikin abstrakteja eivätkä vaadi sanallisia selityksiä. Matematiikalla on myös ollut jo antiikista lähtien vankka sijansa musiikin teoriassa.

Otan esille pari musiikkiakustiikkaa koskevaa aihetta ja yritän valaista niitä Survon avulla. Mm. *Jouko Tolonen* on käsitellyt väitöskirjassaan (1969) mollisoinnun ongelmaa, joka on askarruttanut sekä matemaatikkoja että musiikin tutkijoita kautta vuosikymmenten. Jo Pythagoras havaitsi, miten hyvin sointuvat intervallit vastaavat yksinkertaisia soittimen kielenpituuksien suhteita. Niinpä oktaaville on leimallista suhde 1:2, kvintille 2:3, kvartille 3:4, suurelle terssille 4:5, pienelle terssille 5:6 jne. Sama koskee kääntäen sävelten värähdyslukuja. Näillä harmonisilla, sopusointuisilla suhteilla oli oma merkityksensä sille, että kreikkalaiset arvostivat kokonaislukuja ja niiden suhteita myös musiikin ulkopuolella. Niinpä havainto, ettei neliön lävistäjän ja sivun suhdetta  $\sqrt{2}$  voi esittää kokonaislukujen suhteena, kauhistutti ja sai heidät nimeämään moiset epäsiikiöt irrationaaliluvuiksi.

Lukusuhteet kertautuvat kolmisoinnuissa. Duurisoinnussa värähdyslukujen suhteet ovat 4:5:6, siis alla on suuri ja päällä pieni terssi. Mollisoinnussa käy huonosti, sillä suhteet ovat 10:12:15, koska nyt suuri ja pieni terssi ovat vaihtaneet paikkojaan. Mollisoinnun ongelma on lyhykäisestään siinä, miksi tonaalisessa musiikissa se nousee ainakin miltei tasaveroisesti duurisoinnun rinnalle, vaikka sitä hallitsevat selvästi suuremmat kokonaisluvut kuin duurisointua. Tähän on kehitetty lukuisia teorioita. Oma selitykseni näkyy seuraavasta kuvasta.

Kuvan yläosassa ovat sointujen puhtaat muodot hyvin tiheään puristettuina oskillogrammeina, jotta soinnun olemus tulisi kylliksi esille. Alaosassa näkyvät vastaavien sointujen tasavireiset muodot, joissa varsinkin terssien epäpuhtaudet aikaansaavat epäsäännöllisyyttä. Tämä vastaa mielestäni paremmin sitä todellisuutta, missä muusikot ovat eläneet, sillä jo ennen tasavireisyyden yleistymistä musisointi saattoi olla epäpuhdasta syystä jos toisesta.

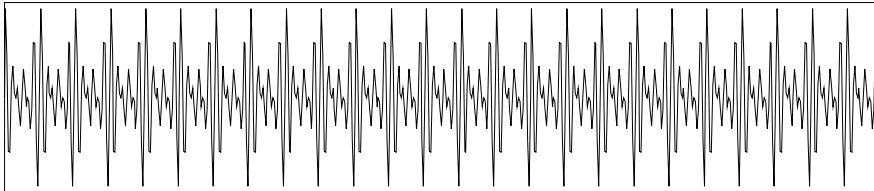
Vaikka siis puhdas duurisointu on kuvassa huomattavasti selkeämpi kuin puhdas mollisointu, tämä ei koske enää sointujen tasavireisiä muotoja, jotka ovat miltei yhtä sameat. Varsinkin duurisointu menettää enkelimäisen puhtautensa.



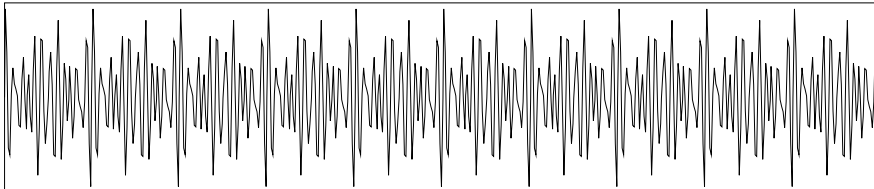
Pohtiessaan mollisoinnun ongelmaa teoreetikot eivät ehkä ole ottaneet huomioon sitä tosiasiaa, että käytännön musisoinnissa esiintyy harvoin täydellistä sointupuhtautta. Kun siis käytetään esim. tasavireistä viritystä, ongelma kaatuu myös kuuluvilta.

Kiinnostuin 1960-luvun lopulla vanhoista viritysjärjestelmistä, joita aikoinaan käytettiin kiinteäasteisilla soittimilla kuten cembalolla. Esim. keskisävelvirityksessä 2/3 tersseistä on periaatteessa puhtaita. Halusin näyttää pianovirituksen mestarille, *Eemeli Paavoselle*, joka pitkään vastasi konserttiflyygeleiden huollosta Suomessa, miltä vanhat viritykset kuulostavat. Lempeän suloiset terssit ja sekstit aiheuttivat tyrmistyksen ja hän tempaisi minulta välittömästi viritysavaimen pannakseen ne paikalleen eli tasavireiseen suuntaan. Toisaalta hän pian oivalsi, mistä on kysymys - lahjoittipa minulle muutaman vuoden kuluttua kirjan, jossa jopa pianon viritykseen suositeltiin vaihtelevia viritystapoja.

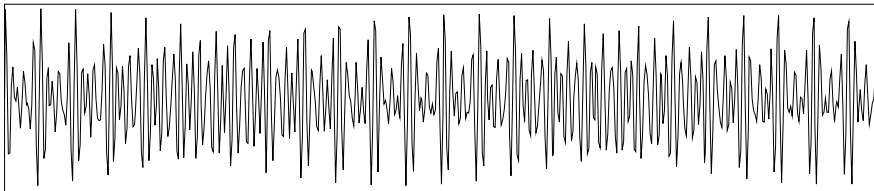
Duurikolmisointu (puhdas)



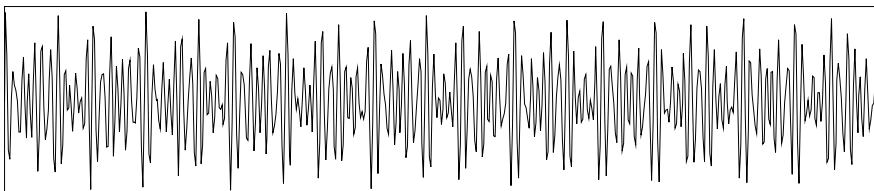
Mollikolmisointu (puhdas)



Duurikolmisointu (tasavireinen)



Mollikolmisointu (tasavireinen)



Alla on näyte siitä, miten puhdas duurisointu piirrettiin.

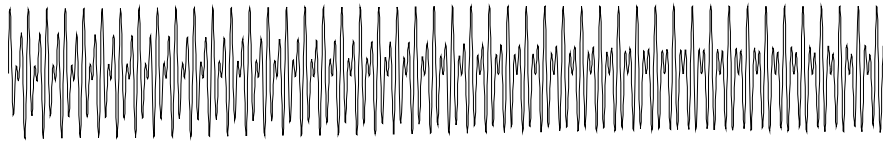
```

40 1 SURVO 84C EDITOR Sun Aug 25 17:13:22 1996 C:\KIRJA\ 100 100 0
1 *
2 *HEADER=Duurikolmisointu_(puhdas)
3 *SIZE=1160,310
4 *XSCALE=0:_,10*pi: YSCALE=-3:,3: pi=3.14159265
5 *X=0,10*pi,pi/60 XDIV=0,1160,0 YDIV=0,250,60
6 *PLOT Y(X)=sin(20*X)+sin(25*X)+sin(30*X)
7 *DEVICE=PS,S1.PS
8 *

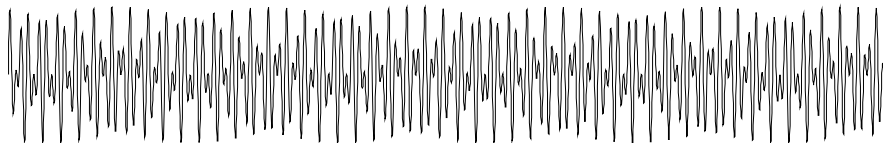
```

Liikaa ei kannata luottaa sointujen ja niiden graafisten esitysten vastaavuuteen. Tasavireiset kvintit ovat kauttaaltaan melko puhtaita. Keskisävelvirityksessä käy siten, että yksi 12 kvintistä muuttuu ulvovan epäpuhtaaksi susi-intervalliksi. Kun tasavireisen kvintin ja tuon suden esittää käyrinä allekkain, silmä ei aisti eroa samalla tavalla epämiellyttäväksi kuin korva. Suden ulvonnan vilkas huojunta vaikuttaa kuvassa jopa kauniilta.

Tasavireinen kvintti



Susikvintti



- ☐ *Hermann von Helmholtz* (1821 - 1894), huomattava fysiologi ja fyysikko on kirjoittanut urauurtavan teoksen akustiikasta, jossa hän mm. esittää kuvallisesti, miten dissonoiviksi eri intervallit priimistä oktaaviin koetaan. Hänen lainehtiva käyränsä perustuu viululla tehtyihin ääninäytteisiin. Olen viritysongelmia pohtiessani päätenyt yksinkertaiseen intervallien dissonoivuutta kuvaavaan laskennalliseen mittaan, jonka perusteella saan aikaan melko samanlaisen käyrän.

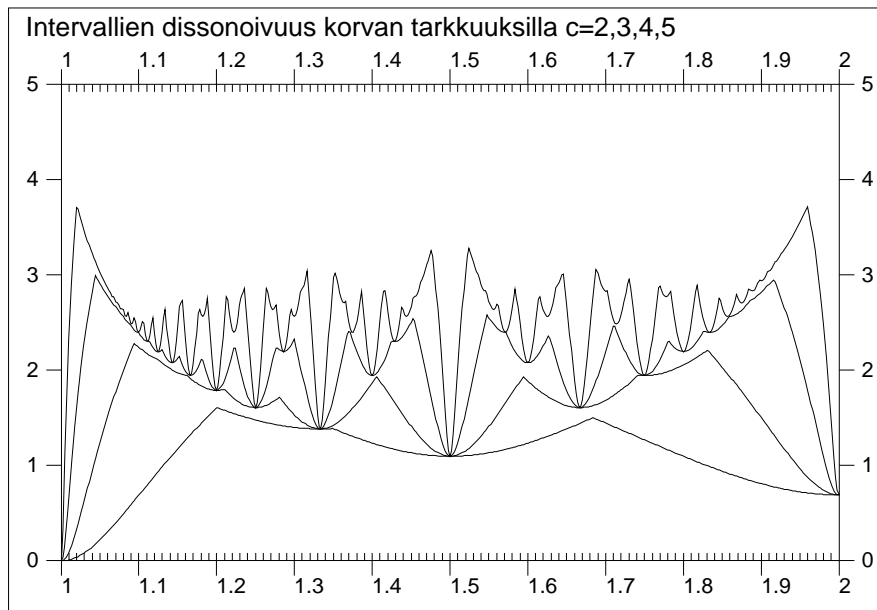
Kun tarkastellaan intervallia, jossa värähdyslukujen suhde on  $x$ , musikaalinen ihminen tunnistaa sen aina jonkin puhtaan intervallin  $n:m$  puhtaaksi tai epäpuhtaaksi muodoksi. Korva suosii pieniä  $m$ - ja  $n$ -arvoja, mutta ei siedä suurta epäpuhtautta. Epäpuhtaus ilmenee huojuntana, jonka suhteellinen määrä on  $mx/n - 1$ . Tarkalle korvalle on ominaista, että se aistii mutkikkaampiakin intervaleja ja on herkkä epäpuhtaalle äänille. Intervallien tunnistusta voi sen perusteella mielestäni jäljitellä etsimällä ne  $m$ - ja  $n$ -arvot, jotka tekevät lausekkeen

$$n + 10^c (mx/n - 1)^2$$

mahdollisimman pieneksi. Tässä  $c$  on korvan tarkkuutta kuvastava parametri, arvoltaan jossain kahden ja viiden välillä. Kyseinen minimiarvo saadaan Survossa lasketuksi funktiona  $\text{diss}(c,x)$ . Esimerkiksi huipputarkalla korvalla  $c=5$  puhtaan suuren terssin dissonoivuusarvo on  $\text{diss}(5,5/4)=1.61$ , mutta terävä tasavireinen terssi saa selvästi korkeamman arvon  $\text{diss}(5,2^{4/12})=2.42$ .

Seuraavassa kuvassa intervallien dissonoivuus priimistä 1 oktaaviin 2 esitetään eri korvan tarkkuuksia vastaavina käyriä. Näissä käyrissä laaksokohdat vastaavat puhtaiksi koettuja intervaleja. Alin käyrä edustaa tyypillistä "jytäkoria", jolle riittää toonika-dominantti-suhteen erottaminen eli väli-intervalleista vain kvintti ja ehkä kvartti tunnustetaan. Ylin käyrä saattaa jo ylittää parhaankin korvan herkkyuden.

Käyrästäön piirtävä kaavio löytyy jälleen valmiina englanninkielisestä Survokirjasta (1992) sivulta 330. Koska  $\text{diss}$  on ns. kirjastofunktio Survossa, ensin sen arvot lasketaan tasavälisesti omaan tiedostoon ja käyrät piirretään ikäänkuin aikasarjoina tästä tiedostosta.



### Myöhästynyttä mainontaa

- Musiikkiin aivan toisella tavalla liittyvät seuraavat kaksi konserttimainosta. Kumpikin on painettu lopullisessa muodossaan Survon PRINT-kaaviolla A3-kokoon. Ensimmäinen on värillinen (vaaleata kullankeltaista tekstiä tummansinisellä pohjalla). Sen taustalla hohtaa utumainen käyrästä, jonka tarkoitus on keventää ilmettä.

# KONSERTTI

## Montessori-luokkien hyväksi

Temppeliaukion kirkossa tiistaina 17.11. klo 19.30

Ralf Gothoni, piano  
Maris Gothoni, piano  
Tuula Hanhinen, piano  
Sharon Jaari, viulu  
Anu Komi, laulu  
Piia Komi, sello  
Ritva-Liisa Korhonen, laulu  
Timo Korhonen, kitara  
Ipo Mansnerus, huilu  
Elina Mustonen, cembalo  
Ilkka Paananen, piano  
Martti Rousi, sello  
Anja Voipio, huilu

Ohjelmassa Mozartin, Haydnin, Chopinin, Scarlattin, Popperin, Villa Lobosin, Blochin, Schönbergin ja Tárregan musiikkia.

Konsertin tuotto käytetään kokonaisuudessaan Pakilan Montessori-luokkien perustamiseen.

Liput 100 mk 3.11. alkaen Lippupalvelusta ja tuntia ennen ovelta  
Ennakkotilauksia myös Anja Voipiolta puh. 743276

# Siikkien

temppeliaukion kirkko  
to 24.11.1994 klo 19.30

**mozart**

pianokvartetto g

**milhaud**

la création du monde

**schumann**

pianokvartetto Es

kaija saarikettu viulu

kreetta-maria kentala viulu

jouko mansnerus alttoviulu

tuija rantamäki sello

raija kerppo piano

liput 60/40 mk tuntia ennen ovelta

## Mielikuvia

- Tähän asti nähdyllä Survon piirroksilla on jokin selvä merkitys tai tulkinta. Minua on pitkään kiinnostanut myös täysin merkityksettömien kuvien synnyttäminen. Harrastusmielessä olen jopa tehnyt 13 kansikuvan sarjan Matemaattisten Aineiden Opettajien Liiton Dimensio-nimiseen lehteen vuosina 1990-91. Yhteistä kaikille näille kuville on ollut se, että ne ovat Survolla piirrettyjä käyräparvia. Jokainen syntyi vain yhden PLOT-komennon aktivoinnilla ja kyseinen kuvakaavio näytettiin aina saman lehden sisäsivuilla niin, että kuka tahansa voi Survolla piirtää kuvan uudelleen ja myös muunnella sitä. Vaikka sanon näitä kuvia merkityksettömiksi, niihin mielestäni sisältyy erilaisia opetuksia. Yksi on se, että hyvin yksinkertaisia funktiomuotoja lievästi sekoittaen syntyy hyvinkin vaihtelevia ja vaikeasti ennustettavia kuvakokonaisuuksia. Siis usein kokeillen ja arvailten olen päätenyt tulokseen, joka näyttää mielenkiintoiselta.

Toistan seuraavilla sivuilla näistä kansikuvista neljä kaavioineen. Paitsi vanhoista Dimensio-lehden numeroista myös Englanninkielisestä Survo-kirjasta (1992) löytyy lisää tämänlaatuisia esimerkkejä. Vaikka merkityksetöntä onkin merkityksetöntä selittää, selitän kuitenkin.

### Lajien synty

Kuva on Survo-kirjassa nimellä "Origin of species". Kuvan keskellä on origo tai solu, josta lajit kehittyvät eri suuntiin. Kyseessä on kaksiparametrinen käyräparvi, joka antaa hyvin vaihtelevia muotoja. Tämä kuva opettanee huomamaan, ettei kovin paljon informaatiota tarvita erilaisten luonnossa tavattavien alkeellisten kukkais- ja eliömuotojen kuvaamiseen. Niiden eloonherättäminen on jo vaikeampi tehtävä. Kuvan oikea alakulma on suurennettuna tämän kirjan kansikuvassa. Joukkoon on siellä eksynyt Chernoffin naamoista tuttu Proconsul Africanus, puuttuva rengas.

### Sukellus mustaan aukkoon

Kyseessä esim. konformikuvauksissa esiin tuleva kahden paraabeliparven yhdistelmä. Risteävät paraabelit kohtaavat toisensa tarkkaan suorakulmaisesti

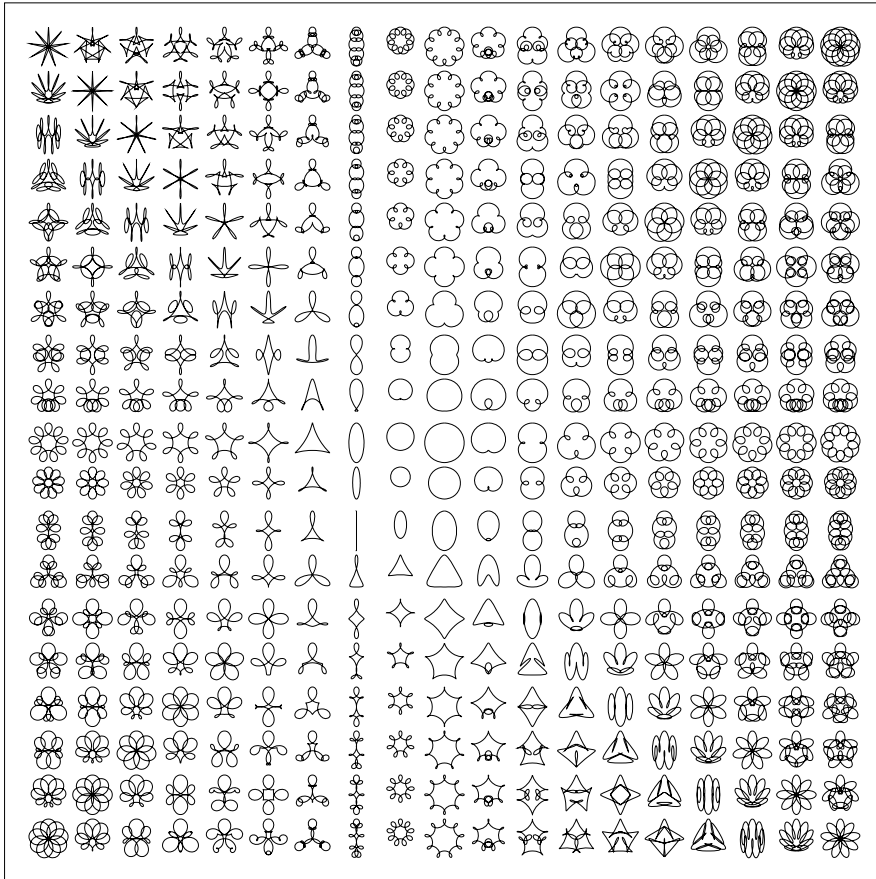
### Kuvakirjoitusta

Tämä on onneton yritys tavoitella jonkinlaista "kiinalaista" kirjoitustyyliä. Kun aikoinaan näytin kuvaa laitoksellamme kiinalaiselle opiskelijalle, hän vain hymyili kohteliaasti ja pudisti päätään.

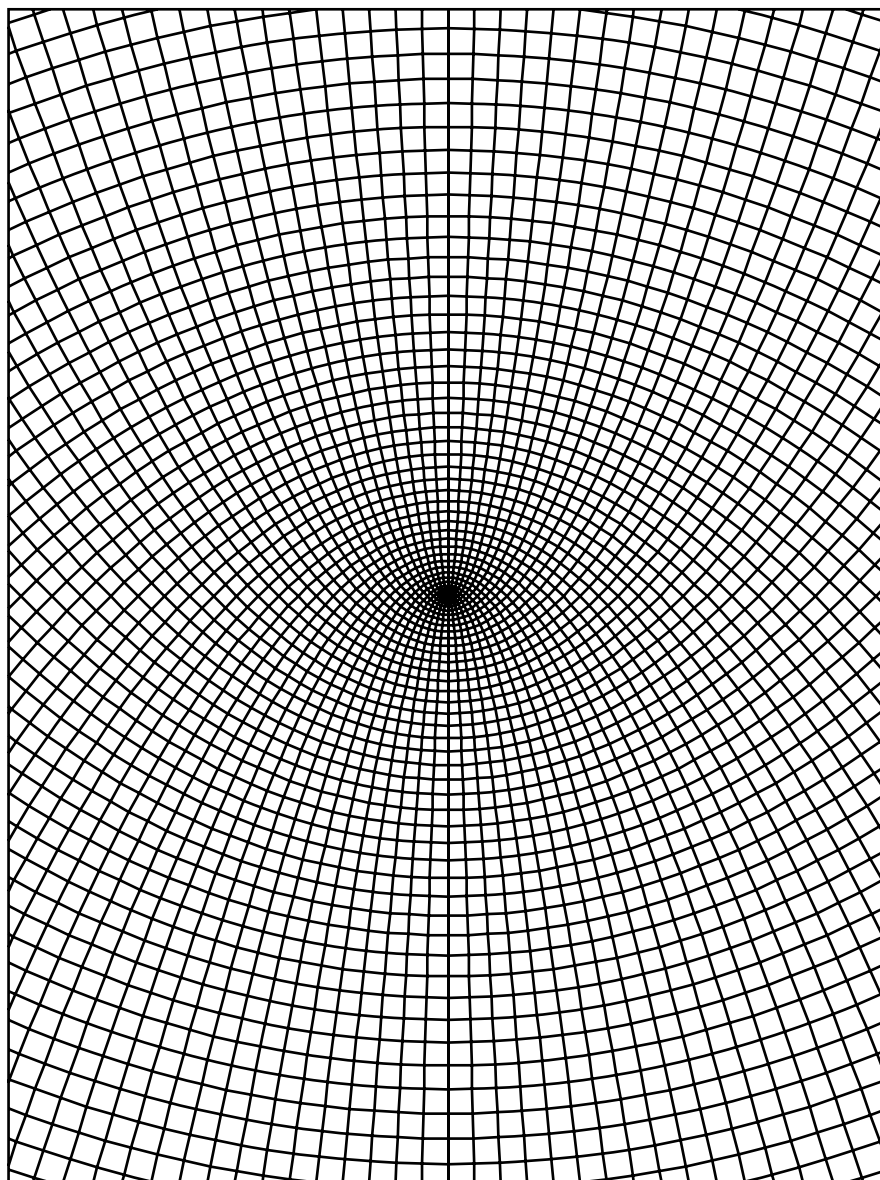
### Itämainen matto

Kuva on hauska kaksikäsitteinen. Päältä katsoen se näyttää monimuotoiselta verkolta tai matolta. Jos sitä tähyää viistosti vaikkapa vasemmasta alakulmasta oikeaan yläkulmaan ja pitää silmänsä lähellä kuvan tasoa, näkeekin lievästi kaartuvia vierekkäin olevia viivoja. Kyseessä onkin yhdellä jatkuvalla viivanvedolla punottu vyyhti.

```
56 1 SURVO 84C EDITOR Tue Aug 27 09:52:51 1996 C:\KIRJA\ 210 100 0
1 *
2 * LAJIEN SYNTY
3 *
4 *XDIV=0,1,0 YDIV=0,1,0 SIZE=1164,1164 HEADER= FRAME=3
5 *A=-8,10,1 B=-8,10,1 T=0,2*pi,pi/120 pi=3.14159265
6 *XSCALE=-9,11 YSCALE=-9,11 DEVICE=PS,SPECIES.PS
7 *
8 *PLOT X(T)=A+0.225*SIN(T)+0.139*SIN(A*T)+0.086*SIN(B*T),
9 * Y(T)=B+0.225*COS(T)+0.139*COS(A*T)+0.086*COS(B*T)
10 *
```



```
17 1 SURVO 84C EDITOR Tue Aug 27 10:02:35 1996 C:\KIRJA\ 210 100 0
12 *
13 * SUKELLUS MUSTAAN AUKKON
14 *
15 *PLOT X(T)=2*C*T,
16 * Y(T)=ind*(T*T-C*C)
17 *T=-5,5,0.1 C=0,5,0.1 ind=-1,1,2 SIZE=1164,1552 XDIV=0,1,0 YDIV=0,1,0
18 *XSCALE=-15,15 YSCALE=-20,20 DEVICE=PS,AUKKO.PS
19 *FRAME=3 LINETYPE=[line_width(1)] HEADER=
20 *
```





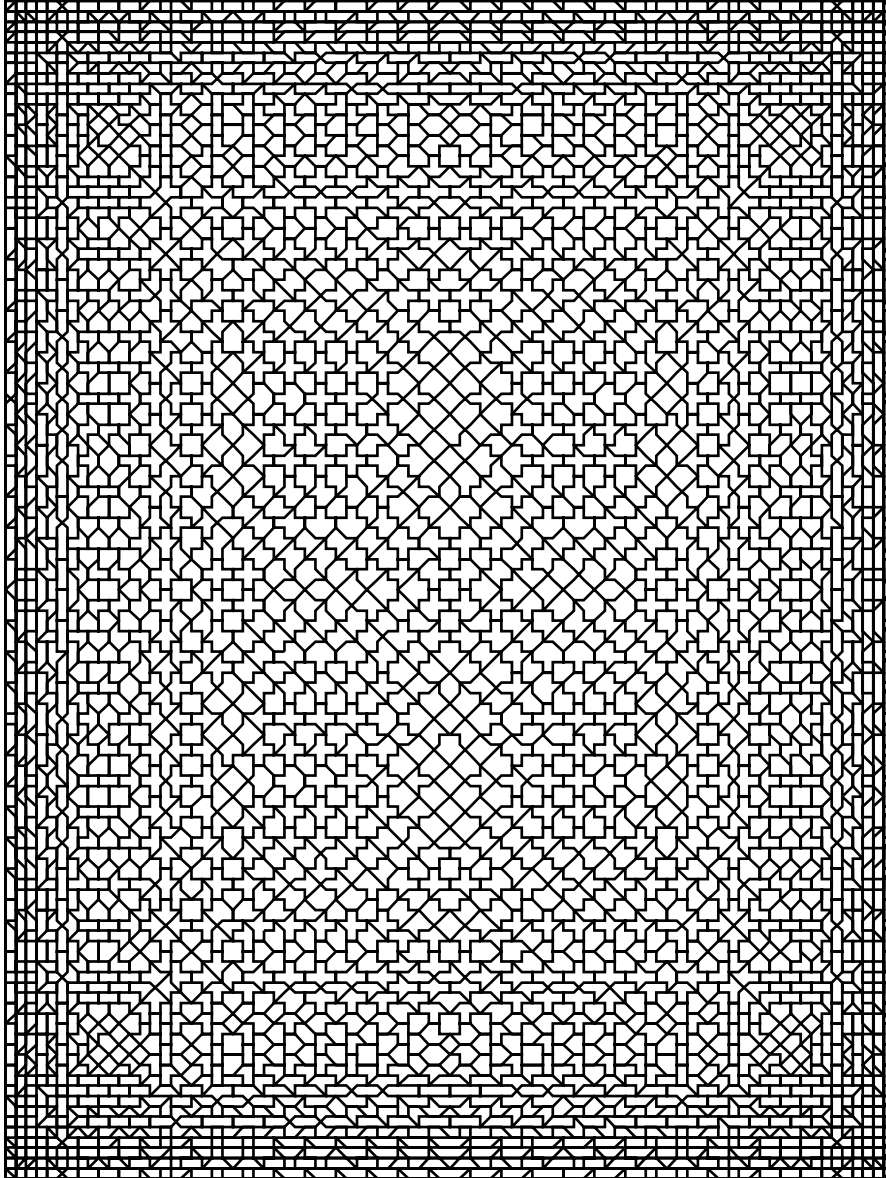
```

37 1 SURVO 84C EDITOR Tue Aug 27 10:14:55 1996 C:\KIRJA\ 210 100 0
22 *
23 * KUVAKIRJOITUSTA
24 *
25 *PLOT X(T)=M-0.5+R*cos(H1*(M+9)*T/5),
26 * Y(T)=N-0.5+R*sin(H2*(N+9)*T/5)
27 *M=1,3,1 N=1,4,1 h=4,10,2 H1=6*sin(3.7*h) H2=5*cos(5.8*h)
28 *R=0.25,0.32,0.01 XSCALE=0,4 YSCALE=0,4 DEVICE=PS,KUVAKIRJ.PS
29 *HEADER= FRAME=0 SIZE=1309,1300 XDIV=0,1,0 YDIV=0,1,0
30 *T=[line_width(2)],1.57,1.87,0.02
31 *

```



```
31 1 SURVO 84C EDITOR Tue Aug 27 10:27:00 1996 C:\KIRJA\ 210 100 0
33 *
34 * ITÄMAINEN MATTO
35 *
36 *PLOT x(t)=int(N*cos(M*t)+0.5),
37 * y(t)=int(M*sin(N*t)+0.5)
38 *M=57 N=43 t=[line width(1)],0,2*pi,pi/16384 pi=3.14159
39 *HEADER= FRAME=0 SCALE=-60,60 SIZE=1640,1640 XDIV=0,1,0 YDIV=0,1,0
40 *DEVICE=PS,MATTO.PS
41 *
```







## Epäpuhdasta matematiikkaa

- Survo ei puhtaasti matemaattisten toimintojensa laajuudessa ja syvällisyydessä vedä vertoja matemaatikoille tarkoitetuille erikoisohjelmille (esim. Mathematica ja Maple). Survon editoriaalinen käyttöliittymä tarjoaa kuitenkin sekä matematiikan opetukseen että kokeelliseen matematiikkaan keinoja, jotka eivät ole yhtä mukavasti saatavilla noissa erikoisohjelmissä.

Varsinaista symbolista matematiikkaa Survossa edustaa vain DER-komento, jolla muodostetaan funktion derivaattoja. Tämän ohjelmoin itse asiassa ensimmäisenä laajempaa C-kielisenä ohjelmani. Tarvitsin sitä epälineaarista regressioanalyysia tekevässä ESTIMATE-ohjelmassa.

Varsinaista iloa minulle tuottaa mahdollisuus numeerisilla kokeilla ja simuloinnilla vahvistaa tai kumota erilaisia eteen sattuvia teoreettisia oletuksia. Survo toimii näin tilastomatematiikan tutkimuslaboratoriona. Olen monesti näin saanut vihiä, mikä on esim. oikea muoto jollekin laskennallisesti tai kokeellisesti määritellylle funktiolle. Tästä mainitsin jo esimerkin moniulotteista vaihtelumittaa kuvatessani.

Tavallisesti yhdistelemällä editoriaalista ja kosketuslaskentaa, matriisitulkkia, muuttujamuunnoksia, simulointia, epälineaarista regressioanalyysia jne. pystyy nopeasti kokeilemaan erilaisia ideoita. Vaativammassa tilanteissa kannattaa rakentaa näistä aineksista sukroja.

## Kokonaislukujen ystäville

Lukumystiikka ja -symboliikka kiehtoo monia ihmisiä. Itse en katso lukeutuvani sellaisiin. Kaikki luvut voidaan selittää "mielenkiintoisiksi". Ennen oli työhuoneeni numero yliopistolla  $314$  eli noin  $100\pi$ . Nyt se on  $2025$  eli  $45^2$ , mikä on myös sillä tavalla kiintoisa, että jos jokaista numeroa kasvatetaan yhdellä eli siirrytään lukuun  $3136$ , niin se vuorostaan on  $56^2$ . Vanhempieni puhelinnumero oli aikoinaan Helsingissä  $9\ 36\ 25$  eli lukujen  $3,6,5$  neliöt peräk-

käin, mistä taas muodostuu vuoden päivien lukumäärä. Numerouudistuksen myötä eteen lisättiin 4 ja tämän numeron olen perinyt itselleni siis muodossa 49 36 25 eli lukujen 7,6,5 neliöt peräkkäin...

Matemaatikkojen keskuudessa on kauan kulkenut "todistus" sille, että itse asiassa kaikki positiiviset kokonaisluvut ovat mielenkiintoisia. Todistus on epäsuora. On ilmeistä, että esim. kokonaisluvut 1,2 ja 3 ovat mielenkiintoisia. Teemme vastaoletuksen, joka sanoo, että kaikki kokonaisluvut eivät olekaan mielenkiintoisia. Näiden joukosta valitaan pienin, mutta sepä vasta onkin mielenkiintoinen luku. Vastaoletus johtaa siis ristiriitaan eli kaikki kokonaisluvut ovat mielenkiintoisia.

Tämä juttu, jonka alkuperää en tunne, huvittaa matemaatikoita, koska siinä sovelletaan epäsuoraa todistustapaa hiukan yllätyksellisesti. Muita se tuskin naurattaa lainkaan.

Tämän vuosisadan tunnetuin kokonaislukujen ystävä lienee omaperäisen nerokas *Srinivasa Ramanujan* (1887 - 1920), joka vietti elämänsä alku- ja loppuosan Intiassa. Hän oli itseoppinut ja hänen suhtautumistaan matematiikkaan leimasi kaikki muodollisuudet ohittava syvä näkemyksellisyys. Tämä johtui osaltaan siitä, miten hän tutustui matematiikkaan. Hän sai käsiinsä 15-vuotiaana noin 5000 matematiikan peruskaavaa sisältävän kokoelman, jonka oli laatinut englantilainen *G.S.Carr* oppaaksi opiskelijoille. Vaikka kaavakokoelma etenee jotenkin johdonmukaisesti alkaen identiteetistä

$$a^2 - b^2 = (a - b)(a + b),$$

se ei sisällä juuri mitään todistuksia. Ilmeisesti tätä kirjaa selatessaan ja muistiinpanoja tehden Ramanujan kehitti omaleimaisen oivalluskäytön. Hänellä oli silmää nähdä suoraan tuloksia ilman hankalia todistuksia. Vaikka hänen muodollinen koulutuksensa jäi vähiin, hän herätti ennen pitkää Intian matemaatikopiireissä sen verran huomiota, että ystävät kehottivat häntä hakeutumaan Englantiin. Hän kirjoitti johtaville matemaatikoille kirjeitä, joissa esitteli (ilman todistuksia) omia tuloksiaan. Vasta kolmannen kirjeen saaja, *G.H.Hardy* oivalsi, mistä on kysymys ja kutsui hänet vuonna 1914 Cambridgeen. Ennen pitkää Hardy ja Ramanujan tekivät yhteistä tutkimustyötä, joka päättyi jo vuonna 1919, jolloin Ramanujan palasi tuberkuloosiin sairastuneena takaisin Intiaan. Ramanujanin jäämistö koostuu suurelta osalta muistikirjoista. Niissä on kosolti aineistoa, josta on riittänyt selvittämistä matemaatikoille näihin päiviin asti.

Hardy meni taksilla tapaamaan Ramanujania tämän ollessa toipilaana Lontoossa. Poistuessaan taksista Hardy pani merkille rekisterinumeron 1729 ja kertoi nähdessään ystävänsä, miten olikaan tylsä numero - toivottavasti ei huono enne. Tähän Ramanujan totesi välittömästi: "Ei, Hardy, se on hyvin

mielenkiintoinen luku - pienin, joka voidaan lausua kahdella eri tavalla kahden luvun kuutioiden summana" eli  $1729 = 12^3 + 1^3 = 10^3 + 9^3$ .

Päästyään jo ennen Englannin kautta mukaan matemaatikkojen toimintaan Intiassa hän julkaisi ystävänsä lehdessä ongelmapalstalla mm. tehtävän, jossa pyydettiin vain laskemaan lausekkeen

$$\sqrt{1 + 2\sqrt{1 + 3\sqrt{1 + \dots}}}$$

arvo. Kului puoli vuotta, eikä kukaan lähettänyt lehteen ratkaisua. Ramanujanin piti kertoa se itse. Oikea vastaus on 3.

Ylläolevat tiedot olen lainannut *Robert Kanigelin* Ramanujan-elämäkerrasta, joka ilmestyi vuonna 1991 nimellä "The Man Who Knew Infinity".

- Tarkastelen äskeistä tehtävää omasta näkökulmastani. Kun luin kirjaa ja näin tehtävän, yritin ratkaista sen. Vaikka vastaus mainitaan Kanigelin kirjassa, tein seuraavanlaisen numeerisen kokeen:

```
17 1 SURVO 84C EDITOR Tue Aug 27 14:07:05 1996 C:\KIRJA\ 23 80 0
1 *
2 *F(n) :=if (n=29) then (x) else (sqrt (1+n*F(n+1)))
3 *
4 *x=1      F(2)=2.999999948216
5 *x=1000000 F(2)=3.0000001612848
6 *x=30     F(2)=3.
7 *
```

Matemaatikko luultavimmin lähtee tutkimaan tehtävää havaitsemalla, että kysymys on arvolla  $n = 2$  funktiosta  $F(n)$ , jolle pätee rekursiokaava

$$F(n) = \sqrt{1 + nF(n+1)}, n = 2, 3, \dots$$

Edellä olevassa Survon kaaviossa olen määritellyt tämän funktion rivillä 2 ja merkinnyt  $F(29) = x$ . Jos nyt muuttujalle  $x$  annetaan erilaisia arvoja, niin lähes riippumatta valitusta arvosta  $F(2)$  hakeutuu lähelle kolmosta. Näyttää ilmeiseltä, että  $F(29) = 30$  ja yleisesti  $F(n) = n + 1$ .

Kun oikean vastauksen tietää, tämän jälkeen sen oikeaksi todistaminen on olennaisesti helpompaa. Niinpä, jos ajattelee numeerisia laskentamahdollisuuksia 1900-luvun alussa, tuskin kukaan pystyi saamaan sitä kautta selville, mitä arvoa tässä ollaan etsimässä. Vaikka tietääkin tuloksen, sen todistaminen oikeaksi saattaa silti tuottaa päänvaivaa. Useat, kuten minä, yrittävät varmaan käyttää hyväksi äskeistä rekursiota, mutta se ei johda nähdäkseni helposti tulokseen. Ratkaisu on lopulta ällistyttävän yksinkertainen ja perustuu siihen, että lähdetäänkin "epämatemaattisesti" liikkeelle vastauksesta 3 ja kehitellään sitä lausekkeen suuntaan. Siis

$$3 = \sqrt{9} = \sqrt{1 + 2\sqrt{16}} = \sqrt{1 + 2\sqrt{1 + 3\sqrt{25}}}$$

ja niin edespäin. Loppujen lopuksi kaikki oikeastaan viriää Carrin kirjan ensimmäisestä kaavasta erikoistapauksessa

$$n^2 - 1 = (n - 1)(n + 1) \text{ eli } n = \sqrt{1 + (n - 1)(n + 1)} .$$

## Matriisitulkki

- Taulukkolaskentaa sivuavat toiminnot jakautuvat Survossa usean erilaisen laskentamuodon kesken. Näitä toimintoja edustavat kosketuslaskenta ja editoriaalinen laskenta, jälkimmäinen erityisesti muuttujamuunnosten, mm. VAR-komennon kautta. Esim. tapa, jolla Pascalin kolmio muodostetaan kosketuslaskennalla toimituskenttään, on näppärä verrattuna tavanomaisen taulukkolaskennan tyyliin. Asia on kuvattu englanninkielisen Survo-kirjan (1992) sivuilla 73-74 ja löytyy myös valmiina kirjan esimerkkien valikosta Survoa käytettäessä. VAR-komento yhdistettynä datatiedostojen hallintaan puolestaan hoitaa raskaat taulukkojen laskentaan ja muokkaukseen kuuluvat tehtävät.

Matriisilaskentaa ei kai yleensä lueta mukaan taulukkolaskentaan, vaikka sen avulla saatetaan käsitellä myös tuon tehtäväalueen ongelmia paljon yleisempien kysymysten ohessa. Matriisilaskennalle on tilastotieteessä paljon kysyntää mm. lineaaristen mallien ja monimuuttujamenetelmien yhteydessä.

Matriisilaskenta tulee käyttäjän ulottuville Survon matriisitulkin kautta. Tulkkia ohjataan MAT-alkuisilla komennolla. Tämä on kolmas matriisitulkki, jota olen elämäni aikana ollut tekemässä. Ensimmäisen rakensin 1960-luvun alussa Elliott 803 -autokoodiin eli Basicin tapaiseen ohjelmointikieleen. Seuraava syntyi SURVO 76:een ja tämä viimeinen nykyiseen Survoon. Ne ovat jokainen käyttötavaltaan erilaisia eikä aikaisempia kannata tässä enempää muistella.

- Matriisitulkki käsittelee omiin tiedostoihinsa talletettuja matriiseja ja matriiseihin viitataan näillä tiedostonimillä. Matriiseja syntyy Survossa tallettamalla taulukkoja toimituskentästä, poimimalla osia datatiedostoista tai erilaisten Survon toimintojen tuloksina. Survon matriiseja saatetaan käyttää jopa datatiedostojen tapaan eli esim. faktorimatriisista saatetaan piirtää kuvia PLOT-komennolla. Matriiseilla on paitsi tiedostonimensä, niiden syntyhistoriaa kuvaava lauseke - esim.  $(A+B) * C$  - sekä rivi- ja sarakeotsikot. Kun matriiseja käsitellään MAT-komennolla, niiden lausekkeet ja otsikot muuttuvat toimenpiteiden mukaisesti. Näin käyttäjä pääsee seuraamaan, mitä matriiseille tapahtuu. Matriisit näkyvät toimituskentässä nimiensä ja lausekkeittensa kautta. Vasta tarvittaessa niitä poimitaan numeerisina taulukkoina toimituskenttään esim. MAT LOAD-komennolla.

Lineaarisen yhtälöryhmän ratkaiseminen osoittaa, miten matriisitulkki toimii.

```

17 1 SURVO 84C EDITOR Wed Aug 28 17:12:56 1996 C:\KIRJA\ 100 100 0
1 *
2 *Ratkaise yhtälöryhmä:
3 * 3x -4y +7z = 110
4 * x + y +2z = 55
5 * 8x -3y +2z = 77
6 *
7 *MATRIX YHTÄLÖT
8 */// x y z ratkaisu
9 *Yht1 3 -4 7 110
10 *Yht2 1 1 2 55
11 *Yht3 8 -3 2 77
12 *
13 *MAT SAVE YHTÄLÖT
14 *MAT A!=YHTÄLÖT(*,1:3) / Erotetaan 3 ensimmäistä saraketta matriisiksi A
15 *MAT B!=YHTÄLÖT(*,4) / Erotetaan viimeinen sarake matriisiksi B
16 *MAT C=INV(A) / *C=INV(A) 3*3
17 *MAT X=C*B / *X=INV(A)*B 3*1
18 *
19 *MAT LOAD X,CUR+1
20 *MATRIX X
21 *INV(A)*B
22 */// ratkaisu
23 *x 9.08257
24 *y 10.39450
25 *z 17.76147
26 *

```

Yhtälöryhmän kertoimet ja vakiotermit on koottu matriisiksi YHTÄLÖT riveille 7-11. Tämä  $3 \times 4$  -matriisi on varustettu sopivilla rivi- ja sarakeotsikoilla. Tässä muodossa se talletetaan rivin 13 MAT SAVE-komennolla matriisitiedostoksi YHTÄLÖT.MAT. Rivien 14 ja 15 komennot erottavat tästä matriisista kerroinmatriisin A ja vakiovektorin B. Yhtälöryhmä on tässä ratkaistu ensin hiukan kömpelästi kääntämällä matriisi A matriisiksi C ja kertomalla vektori B käänteismatriisilla. Matriisitulo eli ratkaisuvektori on lopuksi kirjoitettu toimituskenttään MAT LOAD-komennolla. Huomaa, miten Survon matriisitulkki kommentoi komentoja riveillä 16-17 ja tulostaessaan ratkaisun. Ratkaisuvektori X on varustettu laskentasuoritusta vastaavalla lausekkeella ja otsikot kertovat kaiken olennaisen. Tuloksesta todella näkee, miten se on saatu aikaan. Matriisitulkki paitsi laskee myös noudattaa eräänlaista lauseke- ja otsikkoalgebraa. Esim. matriisia transponoitaessa ja käännettäessä rivi- ja sarakeotsikot vaihtavat paikkoja.

Saman yhtälöryhmän ratkaisu käy kätevämminkin.

```

37 1 SURVO 84C EDITOR Wed Aug 28 18:40:40 1996 C:\KIRJA\ 100 100 0
26 *
27 *MAT SOLVE X FROM A*X=B
28 *MAT LOAD X,111.123456789012345,CUR+1
29 *MATRIX X
30 *(X from A*X=B)
31 */// ratkaisu
32 *x 9.082568807339454
33 *y 10.394495412844040
34 *z 17.761467889908260
35 *

```



Rivillä 27 on aktivoitu MAT SOLVE-komento, joka ratkaisee yhtälöryhmän suoraan ja edellistä tapaa tehokkaammin. MAT SOLVE pystyy ottamaan huomioon kerroinrakenteen ja valitsee ratkaisualgoritmin tämän mukaan. Esim. tapauksissa, joissa yhtälöitä on tuntemattomia muuttujia enemmän, etsitään pienimmän neliösumman ratkaisu.

Ratkaisu on nyt otettu esiin peräti 15 desimaalilla. Tästä on mahdollista terävöittää tulos täsmälliseksi käyttämällä Survon muunnostoimintoja.

```

40 1 SURVO 84C EDITOR Wed Aug 28 18:51:21 1996 C:\KIRJA\ 100 100 0
29 *MATRIX X
30 *(X from A*X=B)
31 */// ratkaisu
32 *x 9.082568807339454 (15:suhde) =990/109 (4.6117623608843e-015)
33 *y 10.394495412844040 (15:suhde) =1133/109 (4.0739980833315e-015)
34 *z 17.761467889908260 (15:suhde) =1936/109 (3.6498581934552e-015)
35 *

```

Ratkaisuna saadut desimaaliluvut on muunnettu mahdollisimman tarkkaan kahden kokonaisluvun suhteiksi aktivoimalla muotoa

<desimaaliluku> (<tarkkuus> :suhde) =\_

oleva lauseke. Survo käyttää tässä täsmälleen samaa periaatetta, jonka esittelin musikaalisen korvan intervallitarkkuutta tarkasteltaessa. Siis ylitarkka ( $c = 15$ ) korva kuuntelee desimaalilukuja ikäänkuin intervalleina. Rationaalilukuarvion perään Survo kirjoittaa vielä sulkuihin desimaaliluvun ja arvion välisen erotuksen. Erotukset ovat nyt niin pieniä, että saadut rationaaliluvut muodostavat todella yhtälöryhmän tarkan ratkaisun. Näin epäpuhdas tulos viritettiin puhtaaksi.

- Survo sisältää muutenkin runsaasti muunnoskeinoja, joista on hyötyä laskennallisissa puuhissa. Muuntaa voi mittayksiköstä toiseen mm. pituutta, nopeutta, pinta-alaa, tilavuutta, massaa, voimaa, työtä, tehoa, painetta, aikaa, kulmaa ja valuuttaa. Lisäksi kokonaislukuja voi esittää eri lukujärjestelmissä ja niitä voi jakaa alkutekijöihin. Muunnoskokoelmaa hallitaan Survossa yhden toimituskentän (MEASURES) avulla. Siellä on omalla rivillään kunkin mittayksikön nimi, laatu ja suhde perusyksikköön. Kokoelma on vapaasti laajennettavissa. Muunnosjärjestelmä tuntee lisäksi yleiset fysikaalisten mittojen etuliitteet kuten mega, giga, mikro ja nano.



|            | Nukkuu | Liikkuu | Syö   |
|------------|--------|---------|-------|
| 1. Nukkuu  | 19/20  | 1/20    | 0     |
| 2. Liikkuu | 0      | 2/3     | 1/3   |
| 3. Syö     | 1/5    | 0       | 4/5 , |

yhden aikayksikön päästä olio on edelleen tilassa 1 todennäköisyydellä  $19/20$  ja tilassa 2 todennäköisyydellä  $1/20$ . Kahden aikayksikön päästä eri tilojen todennäköisyydet tulevat olemaan  $(19/20)^2$ ,  $(19/20) \times (1/20) + (1/20) \times (2/3)$  ja  $(1/20) \times (1/3)$ .

Merkitsemällä  $\mathbf{P}$ :llä siirtymätodennäköisyysmatriisia ja  $\mathbf{x}$ :llä alkujakaumaa  $\mathbf{x} = [1 \ 0 \ 0]'$  on todettavissa, että yleisesti olion tilajakauma  $n$  aikayksikön kuluttua on  $(\mathbf{P}')^n \mathbf{x}$ .

```

8 1 SURVO 84C EDITOR Thu Aug 29 11:31:38 1996 C:\KIRJA\ 400 100 0
275 *
276 *MATRIX P
277 *///      Nukkuu  Liikkuu  Syö
278 *Nukkuu  19/20   1/20    0
279 *Liikkuu  0      2/3    1/3
280 *Syö     1/5    0      4/5
281 *
282 *MATRIX X
283 *///      Olio
284 *Nukkuu   1
285 *Liikkuu  0
286 *Syö     0
287 *
288 * MAT SAVE P
289 * MAT SAVE X
290 * MAT Q=P' / *Q~P' 3*3
291 S MAT X!=Q*X / *X~P'*X 3*1
292 * MAT LOAD X,1.123456789012345,CUR+3
293 * GOTO S
294 *
295 *MATRIX X
296 *///      Olio
297 *Nukkuu   0.714285714285712
298 *Liikkuu  0.107142857142857
299 *Syö     0.178571428571428
300 *

```

Oheisessa kaaviossa matriisit  $\mathbf{P}$  ja  $\mathbf{X}$  on ensin talletettu MAT SAVE-komennoilla ja  $\mathbf{P}$  transponoitu matriisiksi  $\mathbf{Q}$ . Sitten alkaa iteraatio, jossa  $\mathbf{X}$ :ää muutetaan kertomalla sitä toistuvasti matriisilla  $\mathbf{Q} = \mathbf{P}'$ . Joka askeleella  $\mathbf{X}$  tulostetaan ja palataan takaisin GOTO-komennolla riville S (291). Koko toimenpiteiden sarja rullaa automaattisesti, kun siirrytään vaikka jo rivillä 288 jatkuvaan aktivointiin näppäinyhdistelmällä F2 ESC.

Hieman yli sadan kierroksen jälkeen tilatodennäköisyydet vakiintuvat tässä näkyviin arvoihin. Niihin joudutaan, olipa alkujakauma mikä tahansa. Samalla tavalla kuin yhtälöryhmän ratkaisussa todetaan rationaalilukumuunnoksella, että tarkka rajajakauma on  $(20/28, 3/28, 5/28)$ .

Tehtävän voi myös selvittää ilman iterointia havaitsemalla, että ratkaisulle  $\mathbf{x}$  pätee yhtälö  $\mathbf{P}'\mathbf{x} = \mathbf{x}$  eli  $(\mathbf{P}' - \mathbf{I})\mathbf{x} = \mathbf{0}$ . Tämän homogeenisen yhtälöryhmän ratkaisu onnistuu suoraan matriisitulkilla.

## Funktiomuodon metsästys

- ▣ Palaan moniulotteista vaihtelua koskevaan Mvar-mittaani. Mainitsin, että saatoin Survon avulla löytää yleisen lausekkeen tapaukseen, jossa tarkastellaan riippumattomien 0,1-normaalisten muuttujien  $U_1, U_2, \dots, U_p$  kasautuvia summia

$$\begin{aligned} X_1 &= U_1 \\ X_2 &= U_1 + U_2 \\ &\dots \\ X_p &= U_1 + U_2 + \dots + U_p. \end{aligned}$$

Näytän nyt, miten se tapahtui.

```

16 1 SURVO 84C EDITOR Thu Aug 29 13:41:25 1996 C:\KIRJA\ 200 100 0
1 *
2 *p=16
3 *MAT C!=TRI(p,p,1)
4 *MAT S=MMT(C) / *S~C*C' S16*16
5 *
6 *SYSTEM accuracy=10
7 *MULTVAR S,CUR+1 / METHOD=1
8 *Mvar[S]=32.00000 (Total variability in a 16*16 matrix)
9 *MAT LOAD COVVAR.M,END+2 / Optimally permuted covariance matrix
10 *
11 *DATA MVAR,A,B,N,M
12 M 11 11111111111111
13 N p Mvar
14 A 1 1
15 * 2 2.50000000
16 * 3 4.16666667
17 * 4 6.00000000
18 * 5 7.86666667
19 * 6 9.83333333
20 * 7 11.88095238
21 * 8 14.00000000
22 * 9 16.08888889
23 * 10 18.23333333
24 * 11 20.42727273
25 * 12 22.66666667
26 * 13 24.94505495
27 * 14 27.26190476
28 * 15 29.61428571
29 * 16 32.00000000

```

Kullekin dimensiolle  $p$  (tässä  $p = 16$ ) on ensin muodostettu kovarianssimatriisi  $\mathbf{S}$  kertomalla yläkolmiomatriisi  $\mathbf{C}$  (rivi 3), jonka kaikki päälävistäjän yläpuolella olevat alkioit ovat ykkösiä, omalla transpoosillaan eli  $\mathbf{S} = \mathbf{C}'\mathbf{C}$ . Tulostarkkuus on nostettu 10 merkitsevään numeroon rivin 6 SYSTEM-komennolla ja mitan arvo on laskettu rivin 7 MULTVAR-komennolla.

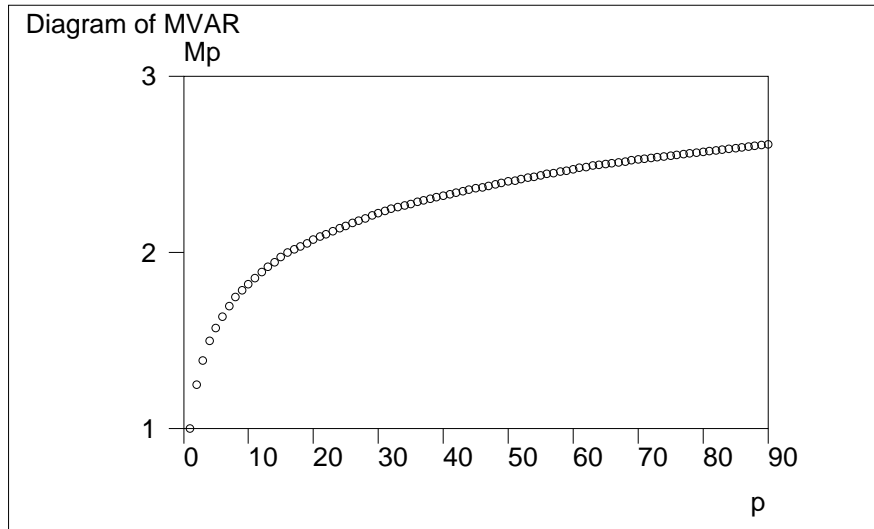
Eri  $p$ :n arvoilla saadut tulokset on siirretty valmiiksi pohjustettuun datataulukoon MVAR aina arvoon  $p = 90$  asti, mikä on matriisitulkissa neliömatriisin suurin ulotteisuus. Jo tässä vaiheessa ilmenee, että  $p$ :n ollessa muotoa  $2^k$ , missä  $k > 1$ , mitta näyttää saavan kokonaislukuarvoja.

Kun yrittää saada käsityksen vaihtelun kasvusta, kannattaa piirtää kuvia. Jos piirretään  $p$  ja Mvar vastakkain, riippuvuus on niin lievästi kaareutuva tarkkaillulla välillä, ettei siitä saa mitään osviittaa. Jos sen sijaan lasketaan vaihtelun arvo muuttujaa kohti eli uusi muuttuja  $Mp = Mvar/p$  ja piirretään tämä  $p$ :tä vastaan,

```

16 1 SURVO 84C EDITOR Thu Aug 29 14:05:13 1996 C:\KIRJA\ 200 100 0
100 * 87 226.4536501 2.60292
101 * 88 229.4181818 2.60702
102 * 89 232.3883416 2.61110
103 B 90 235.3641195 2.61516
104 *
105 *VAR Mp=Mvar/p TO MVAR
106 *GLOT MVAR,p,Mp / XSCALE=0(10)90 YSCALE=1(1)3
107 *

```



tottunut silmä näkee esiin piirtyvän pistejonon logaritmista käyrää muistuttavana. Jos siis ajatellaan mittaa Mvar ulotteisuuden  $p$  funktiona  $f(p)$ , sopivalta tuntuu esitys

$$f(p) = p(a \log_2 p + 1),$$

koska  $f(1) = 1$  ja kakkosen potensseilla saadaan siistejä arvoja. Tällaisen mallin sovittaminen aineistoon käy Survossa mukavimmin ESTIMATE-komennolla, koska siinä käyttäjä voi kirjoittaa estimoitavan mallin juuri haluamassaan muodossa.

```

23 1 SURVO 84C EDITOR Thu Aug 29 14:34:59 1996 C:\KIRJA\ 200 100 0
108 *
109 *MODEL M1
110 *Mvar=p*(a*log(p)/log(2)+1)
111 *
112 *ESTIMATE MVAR,M1,CUR+1
113 *Estimated parameters of model M1:
114 *a=0.249062 (3.92477E-005)
115 *n=90 rss=1.234893 R^2=1.00000 nf=4
116 *
117 *ESTIMATE MVAR,M1,CUR+1 / IND=I
118 *Estimated parameters of model M1:
119 *a=0.25 (1.50723E-017)
120 *n=6 rss=0.000000 R^2=1.00000 nf=4
121 *

```

Malli on nimetty (M1) ja kuvattu riveillä 109-110. Koska Survossa on suoraan saatavilla vain luonnollinen logaritmfunktio  $\log()$ , 2-kantaiset logaritmit muodostetaan suhteina  $\log(p)/\log(2)$ .

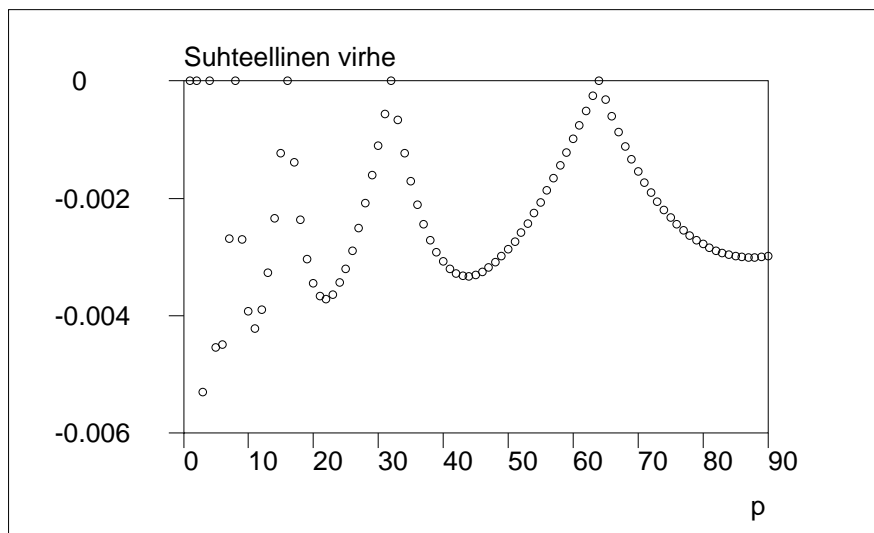
ESTIMATE-komennossa viitataan aineistoon (MVAR) ja estimoitavaan malliin (M1). Ellei toisin sanota, Survo käyttää pienimmän neliösumman keinoa. Ratkaisu tapahtuu tunnistamalla ensin mallin merkinnöistä, mitkä ovat aineiston muuttujia (tässä  $p$  ja Mvar) ja mitkä estimoitavia parametreja eli kaikki tuntemattomat suureet (tässä vain  $a$ ). Tämän jälkeen ESTIMATE muodostaa mallifunktion 1. ja 2. kertaluvun muodolliset derivaatat estimoitavien parametrien suhteen. Esim. siitä, että toiset derivaatat ovat nolliä, kuten tässä käy, ESTIMATE pystyy päättämään, että malli on lineaarinen parametrien suhteen. Muodollisten derivaattojen lausekkeet antavat huomattavasti tukea etenkin aidosti epälineaarisisa malleissa. Mallin ollessa lineaarinen riittää pyörittää Newtonin menetelmää yksi kierros, mikä vastaa tavallista normaaliyhtälöiden ratkaisua.

Ensimmäisessä yrityksessä kaikki 90 havaintoa ovat mukana ja kertoimelle  $a$  saadaan arvio, joka on hyvin lähellä neljäsosaa. Jos aineistoon liitetään apumuuttuja  $I$ , joka saa arvon 1, kun  $p$  on muotoa  $2^k$  ja on muuten 0, malli voidaan laskea uudelleen näin valikoiduilla havainnoilla. Tämä tapahtuu liittämällä ESTIMATE-kaavioon ehdollistava täsmennys  $IND=I$  (rivi 117). Nyt parametrille saadaan arvoksi tarkasti  $1/4$ .

Näin funktiolle  $f(p)$  silloin, kun  $p$  on kakkosen potenssi saadaan lauseke

$$f(p) = p(\log_2 p/4 + 1).$$

Laskemalla tämä lauseke kaikilla arvoilla ja piirtämällä suhteellinen virhe  $[Mvar - f(p)]/Mvar$   $p$ :tä vastaan



nähdään, että aikaisemmin esitetty arvio mitan lausekkeesta pitää paikkansa ainakin arvoon  $p = 90$  asti. On ilmeistä, mutta ei todistettu, että "helminauhojen" alaosat tästä eteenpäin roikkuvat korkeammalla ja korkeammalla eli suhteellinen virhe lähestyy nollaa.



## Makein juttu

- Survon kehittämisessä ja laajentamisessa yhä lisääntyvä osuus kuuluu sukroille. Sokerin kanssa niillä ei ole mitään tekemistä. Nimi juontuu lyhennyksenä "Survon makrosta" tai "supermakrosta". Alunperin puhuttiin "tutoreista", sillä ajatus lähti liikkeelle tarpeesta laatia Survon käyttäjille opetusohjelmia. Avainasemassa on tällöin ns. opetustila (tutoriaalinen moodi), johon siirryttäessä Survo uskollisesti tallettaa kaikki käyttäjän toimenpiteet - siis näppäinten painallukset - valittuun tiedostoon. Näin säilötty keskustelu voidaan välittömästi toistaa, mutta sitä on myös mahdollista muokata ja laajentaa. Tähän tarvitaan oma tulkkiohjelmansa, jolla levyille tallentunut sukro listataan lukukelpoisessa muodossa toimituskenttään muutoksia varten. Sen jälkeen sukro-ohjelma taas talletetaan takaisin tiedostoon. Sukro käynnistetään kirjoittamalla kauttaviiva ja sukron nimi (esim. /OPETUS) sekä painamalla ESC-nappia. Taajaan tarvittavien sukrojen varalta löytyy nopeampiakin aktivointitapoja.

Sukrojen toimiala on vuosien varrella levittäytynyt pelkistä opetusohjelmista moneenkin suuntaan. Yksinkertaisimmillaan sukrot muistuttavat tavanomaisia makroja. Laajimmillaan ne ovat toistuvia asiantuntijasovelluksia toteuttavia ohjelmia ja muodostavat monijäsenisiä perheitä. Jopa osa Survon systeemi-ohjelmista on sukroja. Esim. työvalikkoja hallitaan V-nimisen sukron avulla.

Survossa kehitys kulkeekin nykyisin sitä kautta, että uuden laajennustarpeen ilmaantuessa pohdin ensiksi, voiko tehtävän toteuttaa sukrona. Usein ainakin ohjelman prototyyppi kannattaa tehdä näin, jotta pääsee koettelemaan ratkaisun toimivuutta. Sukroja on näet paljon nopeampi laatia kuin varsinaisia C-ohjelmia.

Sukrojen etuna on myös se, että ne vievät C-kielisiin ohjelmamoduleihin verrattuna huomattavasti vähemmän tilaa. Esim. tyypillinen tilastollinen analyysiketju, joka koostuu osittain Survon valmiista komennoista, mahtuu muutamaa kilotavuun. Vastaava C-moduli saattaa tarvita 100 kilotavua. Tilaa säästetään, koska sukrokieltä tarvitaan yleensä vain liimaamaan valmiit osat yhteen. Näin saatetaan rakentaa vähällä vaivalla käyttäjän kannalta ystävällisem-



piä toimintoja, jotka sekä suomentavat että selittävät saatuja tuloksia. VER-TAA-sukro, jolla pienten satunnaisotosten testaaminen tuodaan maallikoidenkin ulottuville, kelvannee esimerkiksi.

Sukrokieli sisältää jopa laskennallisia toimintoja, mutta yleisiin ohjelmointikieliin sitä ei voi verrata, koska sukrotulkki vie tehosta oman veronsa. Tavallisesti työt jaetaan sopivasti sukron sisällä. Työn yleinen hallinta tapahtuu sukrokielellä, sillä toimintojen ehdollistamiseen ja käyttäjän suuntaiseen vuorovaikutukseen sukrokieli riittää mainiosti. Ainoastaan raskaat työvaiheet teetetään alihankintoina Survon laskennallisilla, tilastollisilla ja tietokantaohjelmilla. Tällöin sukro ei teholtaan ratkaisevasti eroa aidosta C-ohjelmasta.

Sukrojen teko poikkeaa myös siinä mielessä tavanomaisesta ohjelmoinnista, että sukrossa kuvataan pikkutarkasti jok'ikinen kohdistimen liikahtus. Samoin sokean ohjelman on hapuilemalla kyettävä tunnistamaan mm. tilastollisten toimintojen tuottamia tuloksia. Sen on löydettävä vaikkapa testin  $P$ -arvo ja päätettävä, mitä sen perusteella olisi sanottava. Siis eräänlainen aavekäyttäjä tekee töitä kohdistimen "koreografian" kiemuroita noudattaen.

Opetusohjelmissa on tärkeätä, että voi koko ajan seurata, missä mennään, ja vaikuttaa etenemisnopeuteen. Työkalusukroissa ja asiantuntijasovelluksissa, joissa yksityiskohtia ei tarvitse nähdä, toiminta on niin vilkasta, ettei katse pysy perässä. Tällöin sukron touhuilu saattaa vaikuttaa jopa turhan levottomalta. Siksi on mahdollista piilottaa sukron toiminta niiltä osin kuin sen katsoo aiheelliseksi. Joissakin graafisissa sovelluksissa sukro työskentelee jatkuvasti taustalla ja vain vaihtuvat kuvalliset ilmaisut osuvat katsojan silmään. Tällä tavoin koko Survolle ominainen käyttötapa saatetaan häivyttää näkyviltä ja luoda sovelluksia, joiden äärellä käyttäjän ei tarvitse edes tietää Survon olemassaolosta.

Sukroja on varsinaisiin ohjelmiin verrattuna helpompi ylläpitää ja huoltaa. Taitavat käyttäjät saattavat kehittää Survon mukana tulevista sukroista omia muunnoksiaan. Niinpä käyttäjien toivoessa jotain laajennusta Survoon vakio-vastaukseni on: Tee se itse sukrona!

Sukrokieli on käsitykseni mukaan ratkaisevasti nopeuttanut Survon kehitystyötä viime vuosina. Myös monet tottuneiden käyttäjien määräaikaisista töistä lepäävät sukrotekniikan varassa ja olisivat ehkä jääneet syntymättä ilman sitä.

Ahkerien survoilijoiden soisi tekevän sukroja sekä tilapäiseen että pysyvämpään käyttöönsä, koska näin he voittavat huomattavasti aikaa ja vaivaa. Hyvänä ohjeena pitäisin sitä, että jos lyhyen ajan sisällä huomaa tehneensä saman toimenpidesarjan viisi kertaa ja tietää, että se tulee pian taas uudelleen eteen, kannattaa keskeyttää arkitoimensa hetkeksi, miettiä, ja laatia sukro.

Tiedän, että useat pelkäävät sukrokielen opettelua, sillä onhan siinä kysymys eräänlaisesta ohjelmoinnista. Toisaalta rohkaisisin heitä yrittämään, sillä makealta tuntuu, kun saa koneen päivittämään viikottaisen raportin laskelmineen, kuvineen, taulukkoineen ja jopa sanallisine johtopäätöksineen yhdellä ESC-napin painalluksella.

Sukrojen teossa esiintyy kaksi koulukuntaa. Toiset kirjoittavat kaiken heti alusta pitäen sukrokielellä. Toiset taas laativat ohjelmarungon opetustilassa, ottavat sitten syntyneen nauhoituksen sukrokielisenä toimituskenttään ja jatkavat siitä muokkauksin ja lisäyksin. Kaikille sukroilijoille yhteistä on hyödyntää niin paljon kuin mahdollista aikaisempia ratkaisuja. Itse pyrin usein ottamaan jonkin vanhemman sukron listauksen kenttään ja lähden siitä sitten kehrittelemään uutta. Kesken koodauksen ja testauksen saatan kuitenkin tehdä pieniä lisäyksiä opetustilassa, jos niin tuntuu hyvältä.

## Murusia

- Sukrojen teosta löytyy esimerkkejä Survo-kirjasta ja Survoon sisältyvästä esimerkkivalikoimasta. Tässä esittelen vain muutamia näytteitä lyhyemmästä päästä.

```

9 1 SURVO 84C EDITOR Sun Sep 01 17:58:33 1996 C:\KIRJA\ 200 100 0
1 *
2 *TUTSAVE PÄIVÄYS
3 / Siirrytään nykyisen (komento)rivin alkuun ja tyhjennetään se:
4 *{init}{line start}{erase}{erase}
5 /
6 / Aktivoidaan apusukro PVM, joka kirjoittaa päiväyksen muodossa
7 / 23. syyskuuta 1996
8 */PVM{act}
9 /
10 / Lisätään vauhtia, siirrytään rivin alkuun ja lisäystilaan:
11 *{tempo -1}{line start}{ins}
12 /
13 / Kirjoitetaan lisäystilassa 5 väliä ja Helsingissä, :
14 *      Helsingissä, {}
15 /
16 / Poistutaan lisäystilasta ja palataan rivin alkuun:
17 *{ins}{line start}
18 /
19 / Palautetaan alkuperäinen nopeus ja lopetetaan:
20 *{tempo +1}{end}
21 *
22 */PÄIVÄYS-
23 *

```

Tässä on sukron PÄIVÄYS ohjelmakoodi kirjoitettuna riveille 3-20. Sukro talletetaan listan ylle kirjoitetulla TUTSAVE-komennolla. Jos nyt aktivoidaan komento /PÄIVÄYS rivillä 22, sukro tyhjentää tämän rivin ja kirjoittaa tilalle nykyisen päiväyksen muodossa

|    |   |                                |
|----|---|--------------------------------|
| 9  | 1 SURVO 84C EDITOR Sun Sep 01 17:58:33 1996 | C:\KIRJA\ 200 100 0            |
| 21 | *   |                                |
| 22 | *   | Helsingissä, 1. syyskuuta 1996 |
| 23 | *   |                                |

Sukron listauksessa kaikki kontrollimerkillä / osoitetut rivit ovat pelkkiä selityksiä eivätkä vaikuta sukron toimintaan. Selityksiä onkin hyvä antaa jonkin verran. Tässä tapauksessa niitä on poikkeuksellisen monta, jotta lukija saisi otteen koodin tarkoituksesta. Varsinaista sukrokoodia on vain kuudella rivillä. Tällaisen sukron (ilman selityksiä) totunut sukroilija loihtii opetustilassa alle minuutissa. Kokeilin sitä juuri.

Sukron PÄIVÄYS toiminnassa tulee panna merkille, että se aktivoi avukseen valmiin PVM-sukron. Sukrot voivat siis käyttää toisia sukroja aliohjelminaan. PVM osaa kirjoittaa itse päivästekstin. Sen jälkeen paikkakunnan lisääminen on pikkujuttu.

Sukrokoodissa kirjoitettava teksti näkyy sellaisenaan, mutta erikoisnappien painallukset aaltosuluissa olevina koodisanoina. Esim. {R} tarkoittaa rivinsiirtoa ja {act} aktivointia eli ESC-nappia.

Ilman selityksiä koko sukro-ohjelman listaus lyhenisi muotoon

```
{init}{line start}{erase}{erase}/PVM{act}
{tempo -1}{line start}{ins}
    Helsingissä, {ins}{line start}
{tempo +1}{end}
```

Jos joku pitää koodia vaikeaselkoisena, voin lohduttaa toteamalla, että alunperin aaltosulkukoodien paikalla oli vielä lyhyempiä merkintöjä. Niinpä kymmenisen vuotta sitten sama sukro olisi ilmaissut itsensä muodossa

```
2Gb2GB2e2e/PVM2D2GTa1@2GB2i
    Helsingissä, 2i2GB2GTa-1@2GT2S
```

Tämäntyyliseltä siansaksalta näyttävät vielä nykyäänkin joidenkin ohjelmien makrot.

Toiseksi esimerkiksi otan sukron S. Se on tärkeä apuväline varjomerkkien kirjoittamisessa. Kun sukro S käynnistetään, Survon toimitin muuttaa hiukan käyttäytymissääntöjään. Survo-ikkunan alariville ilmestyy kehotus

Enter shadow mode by -. Stop by #.

Edelleen on mahdollista kirjoittaa tekstiä tavanomaiseen tyyliin, mutta jos painetaan nappia -, kehotus muuttuu muotoon

Shadow typing. Resume edit mode by -. Stop by #.

Tällöin jokainen kirjoitettu merkki ei jäljennykään toimituskenttään kohdistimen osoittamaan paikkaan vaan varjomeriksi kyseiseen kohtaan. Näin varjomerkkejä pystyy asettelemaan paikoilleen suoraan naputtelemalla. Yksi painallus korvaa seitsemän erilaista ja tästä on totisesti ollut hyötyä tätäkin kirjaa tehdessä, koska mm. tekstilajin vaihdokset ja kreikkalaiset kirjaimet osoitetaan varjomerkeillä.

```

10 1 SURVO 84C EDITOR Mon Sep 02 10:35:06 1996 C:\KIRJA\ 200 100 0
23 *
24 *TUTLOAD S
25 *{tempo -1}{break off}
26 + 1: {message} Enter shadow mode by -. Stop: #@
27 + A: {get key W1}
28 - if W1 '=' SK then goto A
29 - if W1 '=' - then goto 2
30 - if W1 '=' # then goto End
31 *{write W1}{goto A}
32 + 2: {message} SHADOW typing. Resume edit mode by -. Stop: #@
33 + B: {get key W1}
34 - if W1 '=' SK then goto B
35 - if W1 '=' - then goto 1
36 - if W1 '=' # then goto End
37 *{d}{u}{pre}S{d}{write W1}{u}{pre}S{goto B}
38 + End: {message}@{break on}{tempo +1}{end}
39 *

```

Olen poistanut tilaa säästääkseni sukrosta S kaikki selitykset, jolloin myös koodi näkyy puhtaana. Sukroon sisältyy kaksi ohjelmakierrosta, ensimmäinen riveillä 26-31 ja toinen riveillä 32-37. Ensimmäisessä kierroksessa käyttäjän kirjoittamat merkit pelkästään kopioidaan. Rivien 28-30 ehtolauseet tarkkailevat erityisesti merkkien – ja # painalluksia, sillä näillä on nyt poikkeukselliset tehtävät. Merkin – painallus vie toiseen ohjelmakierrokseen, jossa rivin 37 koodimerkinpitävät huolta siitä, että kirjoitettu merkki asettuu varjoriville. Siis varjorivi avataan painalluksin F2 S eli {pre} vastaa nappia F2. {d} (down) siirtää kohdistimen varjoriville, {write W1} kirjoittaa viimeksi painetun merkin tälle paikalle, {u} (up) siirtää kohdistimen pykälän ylöspäin eli varsinaiselle riville ja lopulta {pre}S poistaa varjorivin näkyviltä. Talletettuna S-sukro mahtuu 296 tavuun.

- En yleensä käynnistä S-sukroa komennolla /S vaan painelen kolmea nappia (F2 M ja S) peräkkäin. Näistä viimeinen on juuri sukron nimi. Yksikirjaimisilla nimillä varustetut sukrot voi siis aktivoida näin koska tahansa keskellä tekstiä, sillä käynnistys ei jätä toimituskenttään mitään merkintöjä. Yksikirjaimisia sukroja on Survon vakiokokoonpanossa useitakin mitä erilaisimpiin avustustehtäviin. Yleiskäyttöiset sukrot sijaitsevat omassa hakemistossaan. Käyttäjä voi varata omille sukroilleen nimikkohakemistonsa. Samanimisten sukrojen tapauksessa ensisijaisesti tulee käyttöön nykyisessä työhakemistossa oleva, toissijaisesti käyttäjän nimikkohakemistossa oleva ja vasta sen jälkeen yleisessä sukrohakemistossa oleva. Koska monet kirjaimet on varattu hyviin tarkoituksiin, käyttäjä voi käynnistää muotoa #X nimeltään olevat sukrot

myös kolmen näppäimen yhdistelmällä. Se on F2 N X, missä X vastaa sukron nimeä. Olen leikilläni alkanut kutsua näitä apusukroja *mukroiksi* ja *nukroiksi*. Esim. N-nukro aktivoituna tämän tekstin kappaleen yllä tasaa rivit vastaamaan lopullista painoasua. Tekstin taitto sujuu näin vaivatta painelemalla tarpeen vaatiessa nappeja F2 N N. Vastaavasti mukro M näyttää maalaamani tekstin osan lopullisessa painoasussaan kuvaruudussa Ghostscript-ohjelman välityksellä.

- Kolmas murunen havainnollistaa, miltä sukro näyttää silloin, kun tekniikkaa sovelletaan aitoon numeeriseen laskentaan. Tämä sukro kirjoittaa peräkkäisiä alkulukuja toimituskentän täydeltä, ellei sitä keskeytä napilla . (piste), joka on yleinen lopetusnäppäin.

```

1 1 SURVO 84C EDITOR Mon Sep 02 12:54:15 1996 C:\KIRJA\ 200 100 0
39 *
40 *TUTSAVE ALUVUT
41 / def Wluku=W1 Wjakaja=W2 Wjäännös=W3 Wneliö=W4
42 *{tempo 1}{R}
43 *SCRATCH {act}{home}2 3{Wluku=1}
44 + A: {Wluku=Wluku+2}{Wjakaja=1}
45 + B: {Wjakaja=Wjakaja+2}{Wjäännös=Wluku%Wjakaja}
46 - if Wjäännös = 0 then goto A
47 *{Wneliö=Wjakaja*Wjakaja}
48 - if Wneliö < Wluku then goto B
49 * {write Wluku}{goto A}{end}
50 *
51 */ALUVUT
52 *2 3 5 7 11 13 17 19 23 29 31 37 41 43 47 53 59 61 67 71 73 79 83 89 97
53 *101 103 107 109 113 127 131 137 139 149 151 157 163 167 173 179 181 191
54 *193 197 199 211 223 227 229 233 239 241 251 257 263 269 271 277 281 283
55 *293 307 311 313 317 331 337 347 349 353 359 367 373 379 383 389 397 401
56 *409 419 421 431 433 439 443 449 457 461 463 467 479 487 491 499 503 509

```

Tästä huomaa, että sukrokieli näin käytettynä muistuttaa jossain määrin mm. Basicia. Erikoispiirteenä mainittakoon, että ohjelman muuttujat, joille voi antaa selväkieliset W-alkuiset nimet (tässä rivillä 41 määriteltyinä) ovat kaikki merkkijonomuotoisia tehtävän numeerisuudesta huolimatta. Muuttujien arvot jopa sijaitsevat yhtenä merkkijonona ns. sukromuistissa, jossa eri muistipaikoja erottavat @-merkit. Siitä seuraa, ettei sukrokielellä juuri kannata koodata mitään puhtaasti numeerisia tehtäviä. Sukro ALUVUT toimii sitä paitsi hyvin alkeellisella ja tehottomalla periaatteella. Opetustilanteissa hitaudesta ei kuitenkaan yleensä ole mitään haittaa.

Muuttamalla viimeinen koodirivi muotoon

```
{Wjäännös=Wluku%7}{write Wjäännös}{goto A}{end}
```

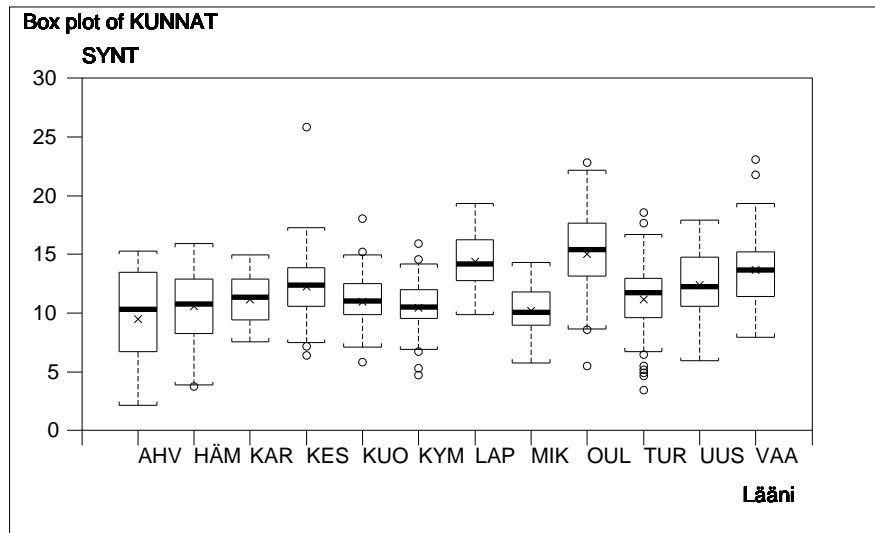
sukro ryhtyykin simuloimaan alkulukunoppaa, jota tarkastelin luvussa "Sattuman pyyntiä".

### Laatikoittain

- ☐ Näytteeksi siitä, miten monipuolisesti sukrot hyödyntävät Survon valmiita toimintoja, käy Kimmo Vehkalahden äskettäin ohjelmoima /BOXPLOT. Kun toimituskentässä aktivoidaan komento

/BOXPLOT KUNNAT,Lääni,SYNT,

kuvaruudulle ilmestyy syntyneisyyttä lääneittäin esittävä piirros.



Kuva kertoo, miten syntyneisyys 1000 asukasta kohti (SYNT) jakautui vuonna 1978 Suomen kunnissa, kun ne ryhmitellään lääneittäin. Kutakin lääniä kuvataan ns. Box & Whiskers- (laatikko & viikset) kaaviolla. Tämä on suosittu vaihtelun esitystapa tilastollisessa data-analyysissä. Laatikko näyttää tarkastellun muuttujan ylä- ja alakvartiilin eli alueen, jolla sijaitsee puolet läänin kunnista syntyneisyyden suhteen. Viikset ovat puolentoista kvartiilivälin mitaiset. Keskimäinen havainto (mediaani) on merkitty paksulla poikkiviivalla ja keskiarvo rastilla. Suuresti poikkeavat tapaukset näkyvät omina pisteinään. Kuva antaa siis yleiskäsityksen syntyneisyyden tilastollisesta käyttäytymisestä ja tarjoaa tilaisuuden läänien väliseen vertailuun. Esim. Oulun läänissä sekä syntyneisyyden taso että vaihtelu on ollut suurta useimpiin muihin lääneihin verrattuna.

Tällä kohtaa haluan kiinnittää enemmän huomiota siihen, miten BOXPLOT-sukro suorittaa tehtävänsä. Sukron listaus käsittää lähes 700 riviä ja se on 11.5 kilotavun laajuinen. Kimmo on sukroa laatiessaan soveltanut erittäin moni-

puolisesti Survon perustoimintoja. Sukro käyttää 25 erilaista Survon komentoa - joitakin moneen kertaan - ja näiden lisäksi sekä editoriaalista että kosketuslaskentaa. Se perustaa työnsä aikana erilaisia väliaikaisia tiedostoja. Lopullinen kuva rakentuu monen erillisen Survon PLOT-komennoilla tehdyn piirroksen koosteena.

Jos saman kuvatyypin ohjelmoi suoraan C-kielellä, siitä tulisi noin 200 kilotavun kokoinen. Samalla toiminta kyllä nopeutuisi kuvaruudussa omalla koneellani nykyisestä 15 sekunnista noin 5 sekuntiin. PostScript-puolella nopeuden lisäys jäisi kuitenkin täysin merkityksettömäksi.

Kuvaruutuun ilmestyneen piirroksen lisäksi BOXPLOT jättää jälkeensä toimituskentän, jossa kuvanteon ainesosat ovat valmiina. Niitä muokaten käyttäjä saattaa muotoilla tulosta mielensä mukaan. Tätä kautta piirroksen saa vaihatta myös siirrettyä PostScript-kirjoittimelle.

Kun komennon /BOXPLOT aktivoi ilman parametreja, se antaa kuvaruudussa seikkaperäiset tiedot erilaisista käyttötavoistaan. Tämä on ominaista kaikille sukroille, jotka on tehty yleiseen käyttöön.

- Nykyisin melkein kaikki vaativimmat Survon sovellukset tehdään sukroina. Ne ovat ikäänkuin laatikkoja, joihin temput kätevästi pakataan. Kookkaimpia sukrokokoelmia on suomenkielinen opetussarjani, joka koostuu noin sadasta sukroperheestä. Näistä useissa on jäsenenä toistakymmentä sukroa. Koko opetussarjan laajuus on yli puoli megatavua. Toinen huomattava kokoelma on Jouko Mannisen TILJAT, jonka hän on kehittänyt tilastotieteen opetukseen.

Erityisen maininnan ansaitsevat *Marjut Schreckin* kaupallis-hallinnollisiin ja toimistotehtäviin tarkoitettut sukrosovellutukset. Näistä mittavin on yleinen ja joustava kirjanpito-ohjelma, joka riittää ainakin pienten yritysten tarpeisiin.

Kaikki Survon perusvalikot ja työvalikot näyttäytyvät ja toimivat sukrojen ohjaamina. Näin koko järjestelmän ylläpito on kevyempää ja joustavampaa kuin perinteisin keinoin toteutettuna.

Varsinkin tilastollisten monimuuttujamenetelmien puolella olen tehnyt lukuisia toimintoja sukrotekniikan avulla. Yhden tällaisen sukron saattaa ohjelmoida tunnissa parissa ja virittää lopulliseen muotoonsa testauksen ja kokeilujen jälkeen muutamassa päivässä. Vastaavan C-kielisen modulin ohjelmointi ja testaus vie karkeasti viitisen kertaa enemmän aikaa. Se kysyy myös runsaasti enemmän ohjelmointitaitoa ja tuottaa usein harmillisiakin yllätyksiä.

## Kimmon oivallus

### ☐ 2.7.1996 klo 12.10-12.40

Tavatessamme tilastotieteen laitoksella Kimmo Vehkalahti kertoi, että *Peter Naeve* oli Tartossa kesäkuun alussa pidetyssä kokouksessa kaivannut sitä, että Survo-istunnon aikana voisi katsella eri ikkunassa esim. tutkimustilanteeseen liittyvää piirrosta. Tästä asiasta oli ollut jo aikaisemmin puhetta ja olimme pitäneet riittävänä, että esim. Windows- tai OS/2-ympäristössä työskennellään samanaikaisesti kahdella Survolla eri ikkunoissa. Yhdellä siis esim. kirjoitetaan tekstiä ja toisella piirretään kuvia. Vaikka tämä on aivan tyydyttävä ratkaisu, pohdiskelimme, olisiko silti grafiikka-ikkunan automaattinen ylläpito mahdollista.

Yht'äkkiä Kimmo totesi, että onhan se, kun pannaan toinen Survo jatkuvasti "nuuskimaan" ensimmäisen tekemisiä. Tästä oivalluksesta päädyimme yhdessä hahmottelemaan toteutuksen suuntaviivoja. Tulimme siihen tulokseen, että koko asian saa toteutetuksi kahdella sukrolla, joista toinen (isäntä MASTER) toimii ensimmäisessä Survossa ja toinen (orja SLAVE) toisessa Survossa. Orjasurvossa aktivoidaan SLAVE-sukro, joka panee tuon Survon odottamaan "viestiä" sovitun tiedoston ilmestymisen muodossa. Käyttäjä siirtyy tämän jälkeen isäntäsurvoon ja toimii sillä normaalisti - orja odottaa. Kun sitten isäntä haluaa orjan tekevän jotain puolestaan, kirjoitetaan tarvittava komento tai komentokaavio isäntäsurvossa niin, että se voitaisiin aktivoida siellä ESC-napilla normaaliin tapaan. Nyt painetaankin sopivaa vakionäppäinyhdistelmää (jollaiseksi seuraavana päivänä sovittiin ctrl-E). Se käynnistää MASTER-sukron. Tämä tallettaa nykyisen toimituskentän ja siihen liittyvät taustatiedot (siis systeemin tilan) kahtena tietynnimisenä tiedostona. Näistä toinen on se, jonka ilmaantumista orjasurvo on odottanut. Huomattuaan tämän orjasurvo tekeytyy näiden tietojen avulla isäntäsurvon "kopioksi" ja aktivoi komennon (komento-kaavion) "ESC-napilla". Esim. GPLOT-kaaviolla määritelty kuva ilmaantuu orjan ikkunaan ja isäntäsurvossa voi työskennellä tästä riippumatta samanaikaisesti eteenpäin.

Tämän aivoriihen lopussa sovimme, että Kimmo idean isänä toteuttaa tarvittavan ratkaisun.

### 2.7.1996 klo 18.37

Kimmo jätti minulle sähköpostiin viestin, joka sisälsi mainitut kaksi sukroa. Hän kertoi viestissään käyttäneensä puolisen tuntia niiden tekoon ja kokeiluun OS/2-ympäristössä.

### 3.7.1996 klo 12.45

Tulin lukeneeksi saapuneet viestit. Löysin Kimmon tekemät sukrot MASTER



ja SLAVE. Siirsin ne omaan koneeseeni ja viidessä minuutissa saatoin havaita, että ne toimivat myös Windows 95-ympäristössä.

Kimmon oivallus on tyypiesimerkki siitä, miten Survon keinoja avarretaan vähäisin elein. Periaatteessa kuka tahansa survoilija tai ainakin sukroilija olisi voinut keksiä saman jo monta vuotta sitten. Aikoinaan SURVO 76:ssa laitteistoon saattoi kuulua tavallisen näytön lisäksi erityinen graafinen näyttölaite, jolla piirrettiin kuvat. Nyt tällaisia "näyttölaitteita" saattaa siis Survolla työkennellessä pitää jopa useita omissa Survo-ikkunoissaan.



## Laiskana ja tyhmänä

- Laajan ohjelmiston rakentaminen edellyttää otsikossa mainittuja ominaisuuksia. Ohjelman tekijän tulee olla *laiska*, jotta hän - edes silloin kun viitsii ajatella - ihmettelisi koneen äärellä, eikä käsillä olevaa työtä voisi tehdä helpommin ja yksinkertaisemmin. Ihmettelystä seuraa, että eteen tulevia vaivoja säästääkseen tekijä päättää parantaa ohjelmaansa uskoen pystyvänsä suoriutumaan muutoksista nopeasti. Vain kyllin *tyhmä* ryhtyy moiseen puuhaan. Hän ei edes monien aikaisempienkaan kokemusten jälkeen ole oppinut ymmärtämään, että todellisuudessa ohjelman parantaminen kuluttaa noin kymmenen kertaa enemmän aikaa ja voimia kuin mitä etukäteen saattoi aavistaa.

Silloin kun tietokoneala otti ensi askeliaan, kehittyi nopeasti arvojärjestys eri osa-ammateille. Korkeimmalla tasolla keikkuivat "systeemisuunnittelijat", jotka vastasivat ATK-projektien yleisestä kehittämisestä ja hallinnasta. He näkivät ohjelmat lähinnä kulkukaavioina eivätkä tohtineet liata käsiään ohjelmoinnilla ja koodauksella. Jotkut johtoportaan henkilöt jopa kerskuivat sillä, etteivät koskaan olleet kirjoittaneet riviäkään ohjelmaa. "Ohjelmoijat" ja heidän alapuolellaan vielä "koodaajat" - nuo laiskat ja tyhmät - tekivät raa'an työn.

Itselleni ei koskaan ole merkinnyt mitään se, millä näistä tasoista työskenteleminen. Mielestäni ATK-alan johtotehtävissäkin toimivien pitäisi periaatteessa osata hommat, joita alaiset tekevät. Kuulun niihin, joille mahdollisimman kiinteä tuntuma esiin tuleviin ongelmiin on tärkeää. Tämän vuoksi mitään ponnisteluja, jos ne auttavat tavoitteen saavuttamisessa, ei tulisi kaihtaa.

Monet ovat kuvitelleet, että olisin teettänyt Survon ohjelmoinnin esim. opiskelijoiden opinnäytteinä. Todellisuudessa vastaan henkilökohtaisesti vieläkin suurimmasta osasta Survon nykyistä C-kielistä ohjelmakoodia. Opinnäytteinä on syntynyt vain muutama Survon osista. C-kielen tasolla *Markku Korhosen* panos erityisesti varianssianalyysin osalta on ollut arvokas. Kimmo Vehkalahti on vuorostaan parina viime vuotena osallistunut yhä voimakkaammin Survon kehittämiseen.

Survon ohjelmoinnin historiasta todettakoon sen verran, että ensimmäinen eli SURVO 66 (numerot viittaavat samalla vuosilukuihin) tehtiin Elliott 803-autokoodilla. SURVO 76 ja SURVO 84 toteutettiin tulkkaavalla Basic-kielellä, mikä tarjosi kohtalaisen suotuisan ympäristön vuorovaikutteiselle ohjelmoinnille. Tätä eivät alan asiantuntijat voineet olla taivastelematta, sillä tuota kieltä pidettiin vain harrastelijoille tarkoitettuna. Sitä paitsi sen arveltiin "pilaavan ohjelmoijan aivot". Niinpä pilaantunein aivoin lähdin vuonna 1985, tutustutuani Microsoftin uuteen C-kääntäjään, jälleen alusta liikkeelle. Tuo kääntäjä tarjosi ensimmäisenä riittävät keinot soveltaa itselleni ominaista välitöntä ohjelmointityyliä, johon juuri Basic-kielen ansiosta olin sortunut. Olin hyvin tietoinen Basic-kielen puutteista ja sen vuoksi etsinyt pitkään parempaa vaihtoehtoa. PC:t vain olivat aluksi niin verkkaisia ja kääntäjät kohmeisia, että pienenkin ohjelmapätkän kääntämiseen kului viitisen minuuttia. Se oli minulle liikaa. Olin tottunut toimimaan yrityksen ja erehdyksen kautta niin, että saatoin kehittää uutta ohjelmaa lyhyissä jaksoissa tietokoneen tukemana. Joka vaiheessa himotti päästä heti näkemään, miten jokin muutos vaikutti. En malttanut odottaa tuloksia viittä minuuttia vaan tovin korkeintaan. Kuten sanottu, tuo uusi kääntäjä yllättäen täytti odotukseni.

Ohjelmoituani aikaisemmin monilla eri tavoilla C-kielen omaksuminen ei tuottanut vaikeuksia. Ensimmäiset laajemmat kokeiluohjelmani kirjoitin vuoden 1985 kesällä parin viikon aikana lomalla paikassa, jossa saatoin käyttää vain kynää ja paperia. Nuo kokeilut koskivat funktion muodollista derivointia ja epälineaarista regressioanalyysia eli niistä kehittyivät uuden Survon ensimmäiset modulit DER ja ESTIMATE. Otin tehtävät itselleni haasteena. Jos saan nämä suhteellisen vaativat asiat toimimaan C-kielellä, selviän sitten muistakin.

Koska olin kirjoittanut vastaavat ohjelmat aiemmin Basic-kielellä, tuotti melkoista iloa - päästyäni loman jälkeen koneen äärelle ja saatuaani ohjelmat jotenkuten toimimaan - havaita, että nopeuden kasvu oli jopa 50-kertainen. Iloani tosin himmensi mm. kokemus, jossa ohjelmaan pujahtanut virhe - en tiedä vieläkö mikä - sai aikaan, että kone "kuoli". Kuvaputki musteni eikä näppäimistö toiminut. Kun kone vihdoin virkosi, se ei tuntenut enää omaa kokoonpanoaan, ei edes sitä, että siihen kuuluu kiintolevy. Tällaiset sattumukset johtuvat C-kielen vallattomuudesta. Se on kuin urheiluauto, jolla pääsee lujaa, mutta josta puuttuvat jarrut. Jyrkässä mutkassa lentää ulos ja saa kauan paikkailla, ennenkuin taas pääsee ajamaan.

Vähitellen opin varomaan mutkia. Ohjelmoinnissa se tarkoittaa paitsi ohjelmointikielen ominaisuuksien hallintaa vielä enemmän itsensä, eritoten omien heikkouksiensa tuntemista. Aina, kun törmään virhetilanteeseen, yritän muistella, onko aikaisemmin sattunut jotain vastaavaa ja jos on, mikä silloin oli onnettomuuden syy. Päätinpä jossain vaiheessa ryhtyä pitämään oikein kirjaa heikkouksistani. Laiskuuttani ja tyhmyyttäni yritys tyrehtyi alkuunsa.

Kun syksyllä 1985 rohkenin lopullisesti ryhtyä C-kielisen Survon kehittämiseen, kuvittelin tietäväni melko tarkkaan, että kuluu noin kolme vuotta, ennenkuin pääsen suunnilleen aikaisempien Survojen tasalle. Tässä suhteessa todellisuus ei heittänyt kymmenkertaisesti, vaan jälkikäteen arvioiden kuvitelmani vastasi suurin piirtein totuutta. Survo vuonna 1988 ei kuitenkaan vastannut aikaisempaa joka suhteessa. Useimmat toiminnot olivat toki jo mukana ja vain muutamat puuttuivat. Vastalahjaksi työn aikana siunaantui lukuisia uusia ajatuksia, jotka veivät kauemmaksi kuin ikinä olisin saattanut etukäteen aavistaa.

Alku näytti todella hitaalta, sillä ensin oli luotava sopivat työkalut. Oli myös selvitettävä ohjelmiston yleinen rakenne, etenkin se, miten edelleenkin saatan toimia pienten ohjelmamodulien varassa tarvitsematta koota niitä yhdeksi suureksi ohjelmajätiksi. Tämän puolisen vuotta kestäneen perustyön jälkeen eteneminen olikin ennakoitua ripeämpää. En ole koskaan raatanut niin pitkiä päiviä kuin viiden ensimmäisen vuoden aikana.

Useat kokeilukäyttäjät osoittelivat kilvan löytämiään virheitä. Niitä esiintyi ensimmäisinä vuosina tuhkatiheään, mutta onneksi tahti on harventunut ja saatan nukkua yöni rauhallisemmin. Aina kun tietooni tuodaan uusi Survosta löytynyt vakavahko virhetoiminto, mieleni täyttää lievä ahdistus; melkein kuin oma lapsi sairastaisi. Vaikka silloin hätäisesti yritän päästä virheen lähteelle, tiedän kokemuksesta, miten vaarallista on uskoa ensimmäiseen mieleen putkahtavaan korjauskeinoon. Siinä käy usein niin, että paikkaa virheen toisella virheellä. Sitten sitä vasta joutuukin pulaan.

Ohjelmoinnin jalo taito ei ole siinä, että välttäisi virheitä ja punoisi koko ajan puhdasta koodia. Tähän kykenee tuskin kukaan. Taito piilee enemmän siinä, että osaa metsästä virhettä ja korjata sen niin, että ohjelmaan jää suurella todennäköisyydellä yksi virhe vähemmän. Tuo oman heikkouksiensa tunnistaminen ja tunteminen auttaa paikkaustyössä melkoisesti.

Harvoin olen onnistunut "maalaamaan itseni nurkkaan" ja joutunut aloittamaan tekeillä olevan ohjelmanosan alusta. Tässä auttaa yksinkertaisesti kokemus ja vaisto. En juuri milloinkaan tee ohjelmia tarkkaan laadittujen piirustusten mukaan. Improvisoin. Pitäisi olla paljon viisaampi, jotta etukäteen näkisi eri vaihtoehdot. Jäykkä ennakkokartoitus pahimmillaan sulkee tien matkan varrella ilmaantuvilta oivalluksilta.

Alusta asti Survoa kehitellessäni olen ikäänkuin asunut talossa, jota vasta rakennan. Moni asia on rempallaan, mutta sitä suurempi syy on työllään lisätä viihtyvyyttään. Asuminen merkitsee myös tilaisuutta soveltaa Survoa itseään kehitystyössä. Tätä voisi sanoa myös aidoksi vuorovaikutteisuudeksi tekijän ja ohjelman välillä. Olen usein korostanut, että vuorovaikutteisen ohjelman luominen ja jatkuva ylläpito edellyttää vuorovaikutteisuutta kaikilla tasoilla.

Näistä näkyvin on luonnollisesti käyttäjän ja ohjelman välinen, mutta tärkeimmäksi kohoaa tekijän ja käyttäjien välinen vuorovaikutus. Toimeliaat käyttäjät ovat tietoisesti ja tietämättään auttaneet paljon Survon edistämisessä. Käyttäjien taholta tulleet virheilmoitukset, epäilyt, moitteet ja ehdotukset otan aina tuskastuneena ja kiitollisena vastaan.

Joku saattaa ihmetellä, miksi en innosta hehkuen ole heti valmis toteuttamaan hänen ainutlaatuista parannusehdotustaan. Minun on pidettävä huolta Survon rakenteellisesta eheydestä. Tarjottu palikka - olipa kuinka upea tahansa - ei välttämättä sovi sellaisenaan. Pitääkseni taloni kasassa paneudun mielelläni sellaisiin ehdotuksiin, joille löytyy perusteluita myös aivan toisilta suunnilta. Jos siis samalle parannukselle osoitetaan useita eri perusteluita, olen valmiimpi sen välittömään toteutukseen. Uskon tällöin, että taloni lujittuu. Silkkää itsepäisyyttäni olen varmasti torjunut oivallisiakin ehdotuksia. Ovathan ne sitten toteutuneet, kun muutaman vuoden päästä "keksin" ne itse.

Joskus hyvänkin uudistuksen tiellä on teknisiä hankaluuksia. Aiemmat ratkaisut, joita ei saa puretuksi, näyttävät estävän parannuksen. Ongelmasta selviää kuitenkin useimmiten tavallista ankarammalla pohdinnalla. Monenkirjajaan kudelmaan on pujotettava uusi punainen lanka entisiä kuvioita rikkomatta.

## **Kaunis yksinkertaisuus**

- Ohjelman tekijä viettää outoa kaksoiselämää. Hänen tulee nähdä samanaikaisesti työnsä sekä käyttäjän näkökulmasta että pinnan alta ohjelmoijan silmin. Käyttäjän ei tarvitse tietää välttämättä tuosta alamaailmasta mitään, mutta jos jotain tuntee, siitä saattaa hyötyä. Silloin käsittää paremmin tehdyt valinnat ja ratkaisut.

Hyppään suoraan Survon ohjelmoijan toiseen todellisuuteen ja näytän, miten Survoon kirjoitetaan uusia toimintoja C-kielillä. Havainnollistan tätä esimerkiksi avulla jopa niin, että kaikki työvaiheet, ohjelman kirjoittaminen, sen kääntäminen ja testaaminen uutena Survon komentona toteutetaan samassa toimituskentässä Survon ollessa jatkuvasti läsnä. Periaatteessa Survon kehittäjä saattaa kohentaa järjestelmää ja laatia siihen uusia C-kielisiä ohjelmia suoraan Survon keinoin poistumatta koskaan Survosta!

Tarkoituksena on luoda Survoon uusi komento PIENIN. Se etsii valitusta havaintoaineistosta tietyn muuttujan pienimmän arvon ja kirjoittaa vastauksen toimituskenttään komentorivin alapuolelle. Tätä ohjelmamodulia sellaisenaan tuskin kukaan kaipaa, sillä esim. Survon STAT-ohjelma tuottaa monen muun tiedon ohella myös muuttujien minimiarvot. Tila ei kuitenkaan salli todellisten modulien ohjelmoinnin esittelyä. Vakuutan silti, että jo tämä rehellinen

pienoiskuva riittää kertomaan jotain olennaista. En siis ole lakaissut roskiamaton alle.

```

20 1 SURVO 84C EDITOR Fri Sep 06 16:07:09 1996 C:\KIRJA\ 50 80 0
1 *
2 *CD C:\KIRJA
3 *>CL /c /AL !PIENIN.C
4 *>LINK !PIENIN,,NUL.MAP,SURVO,NUL.DEF /EXE /STACK:4000
5 *>COPY !PIENIN.EXE C:\E
6 *
7 *PIENIN KUNNAT,Väestö
8 *Muuttujan Väestö pienin arvo aineistossa KUNNAT on 127
9 *PIENIN AIVOT2,Aivot_
10 *Muuttujan Aivot pienin arvo aineistossa AIVOT2 on 0.4
11 *
12 *SAVEP !PIENIN.C

```

Ennen varsinaisen ohjelmakoodin näyttämistä osoitan, miten uusi ohjelma liitetään Survo-perheeseen. Oheisessa toimituskentän palasessa rivit 2-5 sisältävät komennot, joilla ohjelma käännetään ja talletetaan oikeaan paikkaan. Riveillä 7-10 on kaksi esimerkkiä osoittamassa, miten komento PIENIN toimii. Ensin on kuitenkin aktivoitu rivin 12 SAVEP-komento, joka tallettaa alapuolellaan sijaitsevan (tässä vielä näkymättömän) lähdekoodin tekstitiedostona !PIENIN.C . Tämän jälkeen tapahtuu lähdekoodin kääntäminen aktivoimalla peräkkäin - vaikkapa yhdellä jatkuvalla aktivoinnilla - rivit 2-5. Komento CD vain varmistaa, että syntyvät ohjelmatiedostot tallentuvat tämän kirjan nimikkohakemistoon. Komento >CL (rivi 3) kutsuu Microsoftin C-kääntäjän muuntamaan lähdekoodi konekielelle. Tarvittavat apuohjelmat liitetään koodiin rivin 4 >LINK-komennolla. Näin syntyy ohjelmatiedosto !PIENIN.EXE, joka rivin 5 >COPY-komennolla kopioidaan Survon ohjelmamodulien vakiohakemistoon.

Näiden vaiheiden jälkeen, jotka vievät koneellani muutaman sekunnin, uusi komento on valmis kokeiltavaksi. Tilanne muuttuu vain, jos kääntäjä löytää lähdekoodista muodollisia virheitä. Näistä ilmaantuu kuvaruutuun varoituksia. Tällöin on palattava koodin lähteille ja tehtävä korjauksia ennen uutta yritystä.

Ohjelmointikieliet poikkeavat luonnollisista kielistä tiukkuudessaan. Ne eivät tosikkoina ymmärrä mitään rivien välistä. Kun luonnolliset kielet laajenevat ja kehittyvät usein jopa huonokuuloisuuden ja väärinymmärtämisen kautta, ohjelmointikielten ilmaisuvoima lisääntyy vain uusien tarkkojen määrittelyjen välityksellä.

C-kielillä on eräitä huomattavia etuja moneen muuhun ohjelmointikielen verrattuna. Se on perusrakenteiltaan suhteellisen suppea, mutta sitä saattaa jatkuvasti parantaa omilla ilmaisulla. Sen suosimat tietorakenteet ovat suoraan yhteydessä todellisen tietokoneen perusominaisuuksiin eivätkä nojaa hahateluun jostain ihannekoneesta. Itse käytän päinvastoin kuin monet muut ohjel-

moijat C-kielen ominaisuuksia verraten säästeliäästi. Jos purkaisin C-kielen ja kokoaisin osat uudelleen yhteen, paljon jäisi "turhia" osia yli.

Sen sijaan olen noiden vähien perusrakenteiden avulla luonut runsaasti rakenteita ja aliohjelmia eli C-funktioita, jotka palvelevat Survon ohjelmointia. Tällä työkaluvarastolla katson tulevani hyvin toimeen enkä hyödy kielen esim. nimellä C++ kulkevista yleistyksistä. Nämä Survon työkalut laadin ja kokosin aliohjelmakirjastoiksi aivan alussa. Julkaisin ne alustavasti jo ensimmäisenä vuonna eikä niitä sen jälkeen ole tarvinnut kovin paljon lisätä tai kehittää. Työkalut ovat kustannuksitta kenen tahansa saatavissa. Niiden avulla muut voivat laajentaa ja kehittää Survoa siinä kuin minäkin. Tietenkin tarvitaan lisäksi Microsoftin C-kääntäjä (ver.6), joka tulee hankkia erikseen.

```

16 1 SURVO 84C EDITOR Fri Sep 06 16:32:29 1996 C:\KIRJA\ 50 80 0
11 *
12 *SAVEP !PIENIN.C_
13 *#include <stdio.h>
14 *#include "survo.h"
15 *#include "survoext.h"
16 *#include "survodat.h"
17 *
18 *SURVO_DATA d; int muuttuja; double pienin;
19 *void main(int argc, char *argv[])
20 * {
21 *     int i; long j; double x;
22 *     s_init(argv[1]);
23 *     i=data_open(word[1],&d); if (i<0) return;
24 *     muuttuja=varfind(&d,word[2]); if (i<0) return;
25 *     pienin=1e100;
26 *     for (j=d.l1; j<=d.l2; ++j)
27 *     {
28 *         data_load(&d,j,muuttuja,&x);
29 *         if (x<pienin) pienin=x;
30 *     }
31 *     sprintf(sbuf,"Muuttujan %s pienin arvo aineistossa %s on %g",
32 *             word[2],word[1],pienin);
33 *     edwrite(space,r1+r,1); edwrite(sbuf,r1+r,1);
34 * }
35 *

```

Koko lähdekoodi on kirjoitettu riveille 13-34. Olen siinä lihavoinnilla osoittanut kaikki ne ilmaisut, joita C-kieli ei sinällään sisällä vaan jotka ovat Survon ohjelmoinnin käyttöön laadittuja laajennuksia eli työkaluja.

En lähde yksityiskohtaisesti selittämään komennon PIENIN ohjelmakoodia. Tärkeämpää on kuvata, miten tämä lyhyt ohjelmanpalanen toimii Survon helmassa. Kun komento

PIENIN KUNNAT, Väestö

aktivoidaan ESC-napilla, Survon toimitin tarkkailee aktivoitua riviä ja erityisesti ensimmäistä sanaa (tässä PIENIN). Kun se ei tunne tätä sanaa, se etsii levyiltä !PIENIN.EXE-nimisen ohjelmätiedoston, hakee ohjelman keskusmuistin vapaana olevaan tilaan ja käynnistää sen ns. lapsiprosessina. Toimitin (emoprosessi) jää lepäämään ja odottaa lapsosen toiminnan päättymistä.

Emoa ja lasta yhdistää napanuorana vain yksi osoite koneen keskusmuistissa. Lapsi tarttuu tähän osoitteeseen `argv[1]` (ohjelmassa rivillä 22) Survon `s_init`-funktion avulla. Kyseinen osoite on itse asiassa viittaus laajaan osoiteluetteloon, joka tarkassa järjestyksessä kertoo kaiken Survon senhetkisestä tilasta, esim. toimituskentän sisällöstä, kohdistimen paikasta jne.

Koska lapsi näin "tietää kaiken" emostaan, Survon ohjelmoijalle on luotuna valmiiksi alusta, jolla on mukava toimia. Itse asiassa ohjelmoidessani uutta Survon modulia tunnen vain laajentavani yhtä suurenlaiseksi paisunutta ohjelmaa. On valtava etu, ettei moduleita ole mitenkään tarpeen yhdistää suuriksi ohjelmätiedostoiksi. Järjestelmää on tällöin vaivatonta huoltaa. Virheiden aiheuttamat säteilyvaikutukset vähenevät. Tehokkuudesta ei menetetä tippaakaan.

Tyypillisesti Survon ohjelmamodulit ovat 20-100 kilotavun laajuisia, suurimmatkin alle 200-kiloisia ja Survon toimitin `S.EXE` vain 85-kiloinen. Esim. Microsoft Excelin (ver.7) pääohjelma `EXCEL.EXE` on lähes 5000-kiloinen ja Wordin vastaava `WINWORD.EXE` lähes 4000-kiloinen. Survossa moduleita ja muita järjestelmätiedostoja onkin sitten muutamia satoja, jolloin tarvitaan yhteensä 16 megatavua koko nyky-survon tallellapitioon. Windows-ohjelmilla tilatarpeet ovat tässäkin suhteessa aivan toista luokkaa.

Palaan lähdekoodiin. Emon luovutettua valtuudet lapsi saa heti haltuunsa mm. komennon parametrit, siis esim. `KUNNAT` ja `Väestö`, joista ensimmäinen kulkee nimellä `word[1]` ja jälkimmäinen nimellä `word[2]`. Niinpä rivillä 23 se avaa Survon datan, olipa se mitä tyyppiä tahansa. On syytä panna merkille, että ohjelmoijan ei tarvitse tietää, onko data pelkkä havaintolista tai taulukko vai peräti datatiedosto, kuten `KUNNAT`. Hän saa datan olomuodosta riippumatta otteen aineistosta ja voi tunnistaa tutkittavan muuttujan (rivi 24).

Nyt päästään varsinaiseen asiaan eli etsimään valitun muuttujan pienintä arvoa. Tämä tapahtuu riveillä 25-30, joilla käydään koko aineisto läpi ko. muuttujan osalta ja yksinkertaisella jatkuvalla vertailulla löydetään tulos. Jäljelle jää tuloksen kirjoittaminen toimituskenttään (rivit 31-33).

Tähän päättyy lapsen työskentely. Emo herää itsestään horroksestaan ja jää odottamaan vuorostaan käyttäjän toimenpiteitä tämän nähtyä tuloksen.

On paikallaan huomauttaa, että vaikka `PIENIN` on täysin toimiva, aito Survon ohjelma, muutamia olennaisia vaiheita jätin lähdekoodista pois. Ohjelma ei mm. tarkista mahdollisia puuttuvia tietoja eikä siihen sisälly aineiston ehdollista käsittelyä niin, että vain tietynlaatuiset havainnot käsiteltäisiin. Toiminnan ehdollistaminen onnistuisi kuitenkin vaivatta lisäämällä ohjelmakoodiin kaksi vakioriviä. Puuttuvien arvojen testaamiseen riittää yksi.



Survon ohjelmamodulin laatiminen onkin rakenteellisilta osiltaan valmiiden palasten asettamista oikeaan järjestykseen. Veri punnitaan vasta kulloisenkin tehtävän erityispiirteiden, muistinhallinnan ja laskentamenetelmien yhteydessä. Tärkeintä ohjelmoijan kannalta on päästä keskittymään näihin seikkoihin. Järjestelmä huolehtii itsestään siitä, mistä saadaan alkutiedot ja miten tulokset tehdään käyttäjälle tietäväksi.

Kuten jo edellä totesin, virheenetsintä on tärkeä osa ohjelmointityötä. Tässäkin tyydyn hyvin yksinkertaisiin keinoihin. En koskaan käytä varsinaisia virheenetsintäohjelmia. Virheen ilmetessä sirottelen väliaikaisesti lähdekoodiin sopivia tulostuskäskyjä nähdäkseni, miten asiat ovat. Toistamalla menettelyä pääsen yleensä nopeasti ongelman ytimeen.

- Ilman edellä kuvattuja selkeitä rakenneperiaatteita olisi ollut tuiki mahdotonta viedä C-kielistä Survoa niin pitkälle kuin on vähän yli 10 vuoden aikana tapahtunut. Järjestelmä on täysin avoin. Juuri kuvattu esimerkki osoittaa, miten helposti Survoa voi laajentaa. Mainittakoon, että nykysurvon C-kielinen lähdekoodi vastaa merkeittäin laskettuna lähes kymmentä tällaista kirjaa.

Alati on toimittu alkeelliseksi tuomitun DOS-käyttöjärjestelmän ehdoilla. Tuntuu siltä, etteivät arvostelijat ole oikein ymmärtäneet DOS:in vaatimattoman kuoren alle kätkeytyviä voimakkaita ominaisuuksia.

Sitä taakkaa en ota harteilleni, että lähtisin pohtimaan, mistä sitten johtuu esim. Windows-ohjelmien pöhöttyneisyys. Sitä eivät selitä graafisen käyttöliittymän tai moniajon asettamat vaatimukset. Ymmärtääkseni kunnan ohjelmointiympäristössä tulisi vallita erinomaiset edellytykset sille, että ohjelmat yhteistointojensa kautta pikemmin kutistuisivat kuin laajenisivat. Ympäristön tulisi vaalia sekä ohjelmoijien että käyttäjien etuja. Näin ei vain käy. Mielestäni Windows lisäksi pakottaa ohjelmat omaan omituisen ahtaaseen muottiinsa. Ohjelmoijasta voi varomattomuuttaan tulla "Windows-torppari". Hän on sidottuna turpeeseen ja pysyy näin uskollisena isännälleen.

Tämän perusteella ei todellakaan ole syytä asettaa toiveita Survon "Windows-version" kehittämiseksi. Nykyinen Survo viihtyy ja kukoistaa mainiosti paitsi Windows 95:n, myös muidenkin moniajojärjestelmien, kuten OS/2 ja Linux, suojeluksessa. Se hyötyy niiden myönteisistä vaikutuksista juuri niin paljon kuin kohtuudella sopii odottaa. Tuleepa vallan hyvin toimeen omillaankin.

Päätin oman kirjoitusosuuteni sunnuntaina 2<sup>3</sup>. syyskuuta. Olen työskennellyt sen eteen Survolla noin 80 päivän ajan keskimäärin kahtena päivänä kolmesta. Arvioin noina päivinä tehneeni tätä työtä keskimäärin 7 tuntia. Tästä seuraa, että tehokasta peliaikaa on kertynyt noin 350 tuntia. Aikaa on mennyt paitsi kirjoittamiseen, erilaisten esimerkkien laadintaan ja hiomiseen. Paljon on karsiutunut matkalla pois. Vielä riittää aherrusta jonkin verran viimeistelyn vuoksi. Kirja menee painoon 23. syyskuuta 1996.

