

SURVO 76-uutisia

Helsingin yliopiston tilastotieteen laitoksella kehitetty SURVO 76-järjestelmä on tällä hetkellä käytössä Wang 2200-laitteistoilla seuraavissa korkeakouluissa, tutkimuslaitoksissa ja liikeyrityksissä:

	<u>laitteistot</u>
Helsingin yliopisto, Tilastotieteen laitos	2 VP
ALKO, Taloudellinen tutk. ja suunn.yksikkö	1 VP + 1 T
Työterveyslaitos, Helsinki	2 VP + 1 T
CIBA-GEIGY AG, Basel, Sveitsi	4 VP
Siegfried AG, Zofingen, Sveitsi	1 VP
Sotakorkeakoulu, Helsinki	1 T
Postipankki, Helsinki	1 T
Freie Universität, Berlin	1 T
Joensuun KK, tilastotieteen laitos	1 VP
Valio, Helsinki	1 MVP
Valtiovarainministeriö, kansantal.osasto	1 VP
Työväen taloudellinen tutkimuslaitos, Helsinki	1 T
Työterveyslaitos, Tampere	1 VP
Lundin yliopisto, tilastotieteen laitos	1 VP
Vaasan korkeakoulu	1 VP
Alkoholipoliittinen tutkimuslaitos	1 VP
Elinkeinoelämän tutkimuslaitos (ETLA), Helsinki	1 VP

Tilattuna SURVO 76 on mm. seuraaviin paikkoihin:

Hoffmann-La Roche, Basel, Sveitsi

Boehringer, Ingelheim, L-Saksa

Päätellen lukuisista tiedusteluista on ilmeistä, että lähiaikoina SURVO 76-käyttäjien joukko tulee melkoisesti laajenemaan. On myös odotettavissa, että käyttäjien piirissä syntyy yhteistoimintaa ja aloitteita, jotka auttavat järjestelmän kehittämisessä. Olemme saaneetkin huomattavaa tukea jo tähän mennessä ennenkaikkea ALKO:lta ja Työterveyslaitokselta.

SURVO 76:n kehittämisestä ja ylläpidosta vastaa jatkuvasti Seppo Mustosen johdolla toimiva työryhmä. Järjestelmä toimitetaan suoraan Helsingin yliopiston tilastotieteen laitokselta valtionhallinnon alaisiin laitoksiin Suomessa ja ulkomaisiin korkeakouluihin.

SURVO 76:n myynti yksityisiin tutkimuslaitoksiin ja liikeyrityksiin hoidetaan Wang-laitteistojen maahantuoja Findip Oy:n kautta. Työryhmällämme on tähän liittyvä kehityssopimus, jonka puitteissa maahantuoja tulee ylläpitämään SURVO 76-kehitystyöhön tarvittavaa laitteistoa.

Opetustoiminta

Laitoksellamme toimii jatkuvasti viikottain kokoontuva tilastollisen tietojenkäsittelyn seminaari lähinnä jatko- ja täydennyskoulutusmielessä. Seminaari on avoin kaikille asiasta kiinnostuneille. Osanottajia on viime aikoina ollut noin 20. Tämän seminaarin yhteydessä on esitelty myös SURVO 76:een liittyvää tutkimus- ja suunnittelutoimintaa. Tänä syksynä ohjelma on ollut:

16. 9		Seminaarin järjestäminen,
	S.Mustonen	Kuulumisia COMPSTAT-kokouksesta,
23. 9	S.Mustonen	SURVO 76 EDITOR (yleistä),
30. 9	S.Mustonen	SURVO 76 EDITOR (moniulotteiset taulukot),
8.10	J.Puranen	Tilastotieteen opetuksessa käytettävän havaintomateriaalin tuottamisesta SURVO 76:ssa,
14.10	O.Hämäläinen	Cross-impact-analyysi,
21.10	J.Tuomikoski	Kaksi lähestymistapaa digressio-ongelmaan: vertailu simuloinnilla,
28.10	I.Mellin	Moniyhtälö-regressiomallien estimointi SURVO 76:ssa,
4.11	S.Mustonen	SURVO 76-EDITORin laajennuksesta numeeristen lausekkeiden käsittelyyn,
11.11	R.Luukkonen	Moniulotteisten taulukoiden käsittely SURVO 76:ssa,
18.11	"Workshop"	Tienpinnan liukkauden ennustamisesta,
25.11	"Workshop"	Lyijylle altistuneitten seurantatutkimuksesta,
1.12	J.Lappi	Optimointimenetelmistä monihuippuisten kohdefunktioiden yhteydessä,
9.12	L.Tenkanen	MATRI-ketjuista SURVO 76:ssa.

Seminaari tulee jatkumaan keväällä 1981 samankaltaisissa merkeissä. Syksyn aikana Ilkka Mellin on pitänyt SURVO 76-peruskurssin harjoitukseen noin 10 osanottajalle.

SURVO 76:n käyttötavoista

Järjestelmää kehitettäessä päätavoitteena on ollut ja tulee olemaan mahdollisimman kätevän ja monipuolisen tuen antaminen tilastolliseen tutkimustyöhön ja tilastotieteen opetukseen. Tällä hetkellä SURVO 76 on opetustyössä käytössä mm.

1. oppimateriaalin valmistamiseen (luentomonisteet, harjoitustehtävät, graafiset esitykset, taulukot)
2. havaintovälineenä (lentojen ja seminaarien yhteydessä),
3. vaativiin harjoitustöihin,
4. tilastollisen tietojenkäsittelyn opetukseen,
5. tilastollisten menetelmien opetukseen.

Kuitenkin puhtailla tutkimukseen liittyvillä tarpeilla on ollut vielä suurempi merkitys SURVO 76-järjestelmää kehitettäessä. Olemme lisäksi ryhtyneet laajentamaan sitä ulkoisen mielenkiinnon kasvaessa myös sellaisten toimintojen suuntaan, jotka eivät kuuluneet alkuperäisiin suunnitelmiin. Tällä hetkellä tätä ratkaisua voi luonnehtia siten, että se antaa tilastollista tutkimusta ja suunnittelua suorittavalle asiantuntijalle tuen paitsi normaaliin havaintomateriaalin tilastolliseen esittämiseen, muokkaamiseen ja analyysiin myös lopullisen tutkimusraportin laadintaan. Se pyrkii kattamaan kaikki tilastollisen tutkimuksenteon työvaiheet yhtenäisellä tavalla. Käyttö on yleensä keskustelunomaista, mutta SURVO 76:ssa sovelletaan myös muita vuorovaikutteisia työskentelymuotoja.

SURVO 76-ohjelmiston laajuus on tällä hetkellä noin 60 tilastollista osasysteemiä ja modulia, jotka yhdessä käsittävät noin 1 miljoonan tavun määrän ohjelmatekstiä kirjoitettuna laajennetulla BASIC-kielellä. Muodollisesti SURVO 76 on yksi ainoa BASIC-ohjelma, jonka moduläärinen rakenne mahdollistaa jatkuvan laajentamisen ja ylläpidon helpos-
ti.

SURVO 76 ja tekstinkäsittely

Uusimmista laajennuksista on erityisen hyödylliseksi on osoittautunut tekstinkäsittelytoimintojen yhdistäminen tilastolliseen ja muuhun laskentaan. Tätä toimintatapaa on kuvattu mm. raportissa:

Mustonen, S. (1980), SURVO 76 Editor, a new tool for interactive statistical computing, text and data management, Research Report No.19, Dept. of Statistics, University of Helsinki

SURVO 76 Editorin avulla on mahdollista suorittaa kuvaputken välityksellä hallittavassa editointikentässä mm. seuraavanlaisia toimenpiteitä:

- tekstin ja numeeristen aineistojen kirjoittaminen, editointi ja talletus
- kirjoitetun tekstin taittaminen haluttuun leveyteen,
- numeeristen taulukkojen muotoilu ja numeeriset muunnokset,
- numeeristen taulukkojen talletus SURVO 76-tiedostoiksi,
- tulostaulukkojen haku editointikenttään SURVO 76-tiedostoista,
- editointikentän rivien numeerinen ja aakkoslaajittelu,
- tilastollisten analyysien toteuttaminen editointikentässä olevilla havaintoaineistoilla ja taulukoilla,
- editointikentän ja -tiedostojen valittujen osien tulostus paperille, (esim. nyt näkyvillä oleva teksti on kirjoitettu ja työstetty tällä ohjelmalla. Lopullinen taitto ja tulostus paperille on tapahtunut yhdellä ainoalla PRINT-operaatiolla.)

Yksinkertaiset toimitukset hoidetaan suoraan ohjauspöydän funktio-napeilla. Vaativampiin suorituksiin käytetään editointioperaatioita, jotka kirjoitetaan editointikenttään ja niitä voidaan käsitellä samaan tapaan kuin tavallista tekstiä. Samoin koko kenttä voidaan tallettaa operaatioineen levyille ja ottaa myöhemmin käyttöön uusia toimenpiteitä varten.

Tekstinkäsittelytoimintojen yhdistäminen tilastolliseen tietojenkäsittelyyn lisää monin tavoin varsinkin henkilökohtaisen ATK:n hyödyntämismahdollisuuksia.

Sen lisäksi, että se parantaa tutkimusselostusten ja muiden raporttien laatimista, myös tilastollisen tietojenkäsittelyn normaaleja toimintoja voidaan ottaa käyttöön uudella tavalla. Useat tilastolliset operaatiot voidaan jo nyt toteuttaa suoraan editointikentän muodostamassa tilassa editointioperaatioiden kaltaisesti. Tällä hetkellä editointioperaatioilla on mahdollista tehdä suoraan mm. lineaarista regressioanalyysia, varianssianalyysia ja estimoida log-lineaarisia malleja moniulotteisten frekvenssitaulukoiden pohjalta. Tukena näille analyyseille ovat moniulotteisten taulukkojen muokkaukseen suunnitellut erikoisoperaatiot. Näitä toimintoja kuvataan yksityiskohtaisesti em. tutkimusraportissa.

Myös listojen muokkaus ja lajittelu on mahdollista. Esim. viimekevään Maarianhaminan kongressin osanottajien rekisteröinti ja luetteloiden laadinta tapahtui suoraan Editorilla ilman mitään erityisohjelmia.

Numeeristen lausekkeitten käsittely

Viimeisin laajennus, joka on tapahtunut em. raportin ilmestymisen jälkeen, suo mahdollisuuden toteuttaa aritmeettisia ja käyttäjän antamien "kaavakokoelmien" mukaisia laskelmia suoraan editointikentässä.

Yksinkertaisin käyttötapa on tällöin se, että kirjoitetaan mihin tahansa paikkaan editointikentässä (tekstin ja numeeristen taulukkojen sekaan) laskettavaksi aiottu lauseke:

Esim.1a

	1	SURVO 76 EDITOR	(C)1979 S.Mustonen	(100x100)
1	*			
2	*			
3	*			
4	*	Lukujen 11,13 ja 17 geometrinen keskiarvo on		
5	*	$(11*13*17)^{(1/3)}=_$		
6	*			
7	*			
8	*			

Kun nyt painetaan nappia CONTINUE, jolla aina aktivoidaan editointioperaatiot, Editor analysoi aluksi kursorin osittaman rivin ja = kursorin edessä saa aikaan edessä olevan aritmeettisen lausekkeen tulokinnan ja tarvittavat laskentatoimenpiteet, joiden jälkeen kuvaputkelle ilmestyy tulos:

Esim.1b

	1	SURVO 76 EDITOR	(C)1979 S.Mustonen	(100x100)
1	*			
2	*			
3	*			
4	*	Lukujen 11,13 ja 17 geometrinen keskiarvo on		
5	*	$(11*13*17)^{(1/3)}=\underline{13.4460581862}$		
6	*			
7	*			
8	*			

Käyttäjä voi määritellä editointikentässä myös muuttujia ja antaa erilaisille lausekkeille nimiä. Näin voi laatia vaikkapa seuraavanlaisia laskentaohjeita:

Esim.2a

```

1  SURVO 76 EDITOR      (C)1979 S.Mustonen      (100x100)
44  *
45  *      Yleisen normaali-jakauman kertymäfunktio:
46  *
47  *      Kun  $Z > 0$ , normeeratun normaali-jakauman
48  *      kertymäfunktio  $F(Z)$  on likimäärin
49  *       $F = 1 - f * (b_1 * t + b_2 * t^2 + b_3 * t^3 + b_4 * t^4 + b_5 * t^5)$ ,
50  *      missä
51  *       $f = (1 / \sqrt{2 * \pi}) * \exp(-Z^2 / 2)$  (tiheysfunktio) ja
52  *       $t = 1 / (1 + r * Z)$ ,  $r = 0.2316419$ 
53  *       $b_1 = .31938153$ ,       $b_2 = -.356563782$ 
54  *       $b_3 = 1.781477937$ ,    $b_4 = -1.821255978$ 
55  *       $b_5 = 1.330274429$ ,
56  *       $\pi = 3.14159265359$  .
57  *
58  *       $Z = (X - my) / \sigma$ 
59  *       $my = 0$ ,  $\sigma = 1$ ,
60  *
61  *       $X = 3$     $F =$  _
62  *
```

Ylläoleva editointikenttään kirjoitettu teksti sisältää kaikki normaali-jakauman kertymäfunktion numeeriseen laskemiseen tarvittavat tiedot. Se on siis eräänlainen ohjelma, vaikkei käyttäjä sitä ehkä koe ohjelmaksi vaan pikemmin luonnolliseksi tavaksi kertoa, miten ko. arvoja lasketaan.

Kun nyt painetaan CONTINUE-nappia kursorin osoittaessa kohtaa $F =$ rivillä 61, Editor analysoi kaikki editointikentässä esiintyvät määrittelyt, kunnes saa F :lle numeerisen arvon ja synnyttää tässä tapauksessa tuloksen:

Esim.2b

```

1  SURVO 76 EDITOR      (C)1979 S.Mustonen      (100x100)
44  *
45  *      Yleisen normaali-jakauman kertymäfunktio:
46  *
47  *      Kun  $Z > 0$ , normeeratun normaali-jakauman
48  *      kertymäfunktio  $F(Z)$  on likimäärin
49  *       $F = 1 - f * (b_1 * t + b_2 * t^2 + b_3 * t^3 + b_4 * t^4 + b_5 * t^5)$ ,
50  *      missä
51  *       $f = (1 / \sqrt{2 * \pi}) * \exp(-Z^2 / 2)$  (tiheysfunktio) ja
52  *       $t = 1 / (1 + r * Z)$ ,  $r = 0.2316419$ 
53  *       $b_1 = .31938153$ ,       $b_2 = -.356563782$ 
54  *       $b_3 = 1.781477937$ ,    $b_4 = -1.821255978$ 
55  *       $b_5 = 1.330274429$ ,
56  *       $\pi = 3.14159265359$  .
57  *
58  *       $Z = (X - my) / \sigma$ 
59  *       $my = 0$ ,  $\sigma = 1$ ,
60  *
61  *       $X = 3$     $F = 0.9986500327$ 
62  *
```

Editoimalla kuvaputkella parametreja my , σ ja muuttujaa X on helppo laskea erilaisia kertymäfunktion arvoja. Yhtäläillä on mahdollista kysyä, mitä on $1 - F$ kirjoittamalla $1 - F =$ tai mitä on f kirjoitta-

malla $f=$.

Koska laskeminen ja kaikki muukin toiminta Editorilla tapahtuu editointikentässä, käyttäjän on helppo täydentää kentässä olevaa tietoa ja pyyhkiä tarpeettomia kohtia (esim. pyöristykset) ennen lopullista tulostusta tai talletusta.

Tarkoituksena on sisällyttää tähän ratkaisuun normaali- alkeis-funktioiden ohella erilaisia tilastollisia funktioita. Esim. binomijakaumaan liittyviä todennäköisyyksiä voidaan laskea jo nyt seuraavilla tavoilla:

Esim.3

	1	SURVO 76 EDITOR	(C)1979 S.Mustonen	(100x100)
24	*			
25	*			
26	*	Käytettävissä ovat binomikertoimet $C(n,m)$,		
27	*	binomijakauman $\text{bin}(n,p)$ pistetodennäköisyydet $\text{fbin}(n,p,x)$ ja		
28	*	kertymäfunktio $\text{Fbin}(n,p,x)$.		
29	*			
30	*	Siis binomitodennäköisyys, kun $n=10$, $p=0.5$ ja $x=5$		
31	*	voidaan laskea esim. muodossa		
32	*	$C(n,x)*p^x*(1-p)^{(n-x)}=0.24609375$		
33	*	tai yksinkertaisemmin suoraan muodossa $\text{fbin}(n,p,x)=0.24609375$		
34	*	Vastaava kertymäfunktion arvo on $\text{Fbin}(n,p,x)=0.623046875$		
35	*			
36	*	Funktioiden argumentit voivat vuorostaan olla funktioita		
37	*	tai lausekkeita; esim. $C(\text{sqr}(49),C(4,2)-1)=21$		
38	*			

Ylläolevan tekstin sisältämät laskutoimitukset ovat tapahtuneet aktiivoimalla CONTINUE-napilla lausekkeet riveillä 32,33,34 ja 37. Jos halutaan yksittäisten arvojen laskemisen sijasta taulukoida yhden tai useamman muuttujan funktioita, tähän on käytettävissä uusi COMP-operaatio, jonka yhteydessä määritellään tarvittavat argumentit ja lausekkeet. Apulausekkeita voi olla määriteltynä editointikentän kaikissa osissa.

SURVO 76-kirjallisuutta

- Mustonen, S. (1976), Tilastollinen tietojenkäsittely pientietokoneella, Helsingin yliopiston tilastotieteen laitoksen tutkimusraportti No.4
- Mustonen, S. ja Mellin, I. (1980), SURVO 76 program descriptions
- Mustonen, S. (1980), On interactive statistical data processing, 8th Nordic Conference on Mathematical Statistics, Mariehamn
- Mustonen, S. (1980), Interactive analysis in SURVO 76, Proceedings of Computational Statistics, Physica-Verlag, Wien
- Mustonen, S. (1980), SURVO 76 Editor, a new tool for interactive statistical computing, text and data management, Research report No.19, Dept. of Statistics, University of Helsinki