

sl

Tartu 1993

2



EESTI STATISTIKASELTS

Teabevihik



Tartu 1993



2

EESTI STATISTIKASELTS

Teabevihik

SISUKORD

Saateks 2

I

"REGISTRID EESTI STATISTIKA- JA INFORMAATIKASÜSTEEMIS",
VILJANDI, 2.–3. MÄRTS 1993. A.

E.-M. Tiit, H. Vigla, M. Kajandu: Eesti statistika- ja infopoliitika
päevaülesanded 4

H. Vigla: Ettevõttereegistri vajadusest statistikas 7

K. Ääremaa, R. Ääremaa: Statistika ja informaatika arengusuunad ... 9

II

"STATISTIKA TARKVARA EESTIS", TARTU, 20.–21. APRILL 1993. A.

E.-M. Tiit: Millist statistika tarkvara on kasutatud ja kasutatakse
Tartu Ülikoolis 17

E. Käärik, M. Yähi: STATGRAPHICS ja tema kasutamine õppetöös .. 23

J. Roosare: Territoriaalsete andmete analüüsi paketid ja
statistika osa nendes 27

E. Päma: Noorpullide kasvukiiruse aretusväärtuste hindamine 36

U. Oja: NSDstat ja selle kasutamise võimalused 39

A.-M. Parring, J. Vilismäe: Lühiülevaade paketest SYSDAT 45

L.-M. Tooding: Seosemudelite kasutamine ühiskonna- ja
majandusteadustes 50

A. Kiviste: Metsandusandmete töötlemise kogemustest 57

Varia 59

SAATEKS

Käesolev trükkis on Eesti Statistika seltsi teine väljaanne, mis jätkab eelmise teabevihiku traditsiooni — anda lugejale ülevaade seltsi tegevusest järgneva perioodi jooksul. Käesoleva vihiku üks ülesandeid on anda teavet erinevate statistikavaldkondade funktsioneerimisest, teadlaste, praktikute ja õppurite kokkusaamistest ühistel seminaridel. Teabevihik on mõeldud seeriväljaandena, kust seltsi liikmed ja statistikahuvilised leiavad informatsiooni seltsi tegevuse kohta ja väljavõtteid üritustel ettekantud olulistest ettekannetest. Et teabevihik võiks muutuda ka mõttevahetuspaigaks Eesti statistika ja selle arengu käsitluse kohta, siis on käesolevasse vihikusse lisatud materjale ka Riigi Statistikaameti väljaantavast infolehest.

Esimese tegevusaasta jooksul toimus neli seminari.

- 15. XII 1992. a. Tallinnas "Eesti rahvuslik arvepidamine" — korraldajaks Aarne Tihemets;
- 2.–3. III 1993. a. Viljandis "Registrid Eesti info- ja statistikapoliitikas" — korraldajaks Mihkel Servinski;
- 20.–21. IV 1993. a. Tartus "Statistika tarkvara" — korraldajateks Kuldev Ääremaa ja Ülo Randaru;
- 2.–3. XI 1993. a. Tartus "Majandus- ja inseneristatistikast" — korraldajateks Villem Tamm ja Ülo Randaru.

Seminaridel osalesid seltsi liikmed, teadlased, väliskülalised, tudengid jt. statistikahuvilised. Esimesel tegevusaastal võeti ESS vastu Rahvusvahelise Statistika instituudi (ISI) kollektiivliikmeks; personaalselt kuulub ISI liikmete hulka seltsi president prof. Ene-Margit Tiit.

ESS juhatus avaldab kõigi oma liikmete nimel tänu Informaatikafondile, kes on toetanud ESS-i materiaalselt seltsi ürituste läbiviimisel.

ESS ootab kõikidelt huvitatutelt ettepanekuid edaspidiste kavade ja ürituste kohta aadressil: Tartu, EE-2400, Liivi 2, ESS; telefonid 43 06 41 ja 43 53 91.

Kutsume kõiki koostööle ja ootame sooviavaldusi uute liikmete vastuvõtmiseks.

Ülo Randaru
ESS-i juhatuse liige



EESTI STATISTIKASEELTS

Esti StatistikaSeltsi seminar

Registrid Eesti statistika- ja infosüsteemis

Viljandi

2.märts - 3.märts 1993.

Viljandi StatistikaBüroo, Vabaduse plats 4.

2. märts.

I. ISTUNG.

10.00 - 11.00. Registreerimine.

11.00. Seminari avamine.

11.05. E.Tiit (TÜ). Statistilised andmed ja statistika tänapäeva Eestis Eesti StatistikaSeltsist vaadates.

11.45. T.Vapper (TÜ). Millal saame objektiivse statistika? Kas probleem on informatsioonis või organisatsioonis?

12.30. Läbirääkimised.

13.00 - 14.00. Lõunavaheaeg.

II. ISTUNG

14.00. M.Kajandu (Riigi Arvutuskeskus). Riiklik registripoliitika I

14.45. R.Vende, I.Odrats (Riigi Arvutuskeskus). Riiklik registripoliitika II.

15.30. H.Vigla (ESA). Ettevõttereegistri vajadus statistikas.

16.15. P.Küüts (ESA Peaarvutuskeskus). Ettevalmistustöödest riikliku rahvastikuregistri asutamiseks.

17.00. Läbirääkimised. Tarkvara demonstratsioon.

III. ISTUNG

18.00. ESS üldkoosolek.

Informatsioon hetkeolukorrast ja lähematest üritustest.

Uute liikmete vastuvõtt. Liikmekaartide kätteandmine.

3.märts.

IV ISTUNG

9.00. Tarkvara demonstratsioon.

10.00. K.Ääremaa, R.Ääremaa (TÜ). Statistika ja informaatika arengusuunad.

11.00. S.Rogova (ESA Peaarvutuskeskus). Eesti ettevõttereegister vabariiklikus infosüsteemis.

11.45. P.Tooming (Tartu StatistikaBüroo) Digitaalkaart ja registrid.

12.30. M.Servinski (Viljandi StatistikaBüroo). Registrip probleem Viljandist nähtuna.

13.00. Läbirääkimised.

13.45. Seminari lõpetamine.

Eesti StatistikaSelts, Liivi 2, Tartu, Estonia EE2400

Tel: 01434-30641 Fax: 01434-35440 E-mail: ess@ulam.ut.ee

EESTI STATISTIKA- JA INFOPOLIITIKA PÄEVAÜLESANDED

Viljandis toimus 2. ja 3. märtsil Eesti Statistikeseltsi seminar "Registrid Eesti statistika- ja informaatikasüsteemis". Osa võttis ligi 50 huvilist, nende seas suur osa neid inimesi, kelle igapäevatöö on seotud registritega.

Kuigi esindatud olid üpriski erinevad amet- ja meeskonnad, jõuti ühissuukohtadele selles, missugused on peamised probleemid tänase Eesti statistika- ja infopoliitika väljaarendamisel.

Selle mõttega, et areng viib meid varem või hiljem niisugusesse ühiskonna arengustaadiumisse, mida on hakatud nimetama infoühiskonnaks, oleme juba harjunud, kuid küllap mõistab iga inimene seda asjaolu erinevalt.

Infoühiskonda iseloomustab eriline vahekord teabe ja inimese vahel, teabe mõju inimesele ja ühiskonnale.

Väga oluline on teabe, s.h. eriti statistilise teabe tunnetuslik väärtus igale kodanikule. On loomulik, et iga inimene tahab end identifitseerida oma kodukoha, kodumaa, rahvaga. See identifitseerimine on võimalik teabe pinnal — tundes ja tunnetades ennast, oma kodukohta, oma ajalugu, oma rahvast, oma riiki.

Mida paremini on inimene informeeritud, seda sõltumatum ja enesekindlam ta on. Teadmatuses hoiatud inimene ja rahvas on manipuleeritav — see tõde oli hästi selge totalitarismi (ka natsismil!) ideoloogidele, kes igasuguse informatsiooni, eriti aga statistika salastamise abil kõigiti takistasid impeeriumi alluvaid ennast identifitseerimast oma rahva, rahvuse ja ajalooliselt kujunenud riigiga.

Tänapäevalgi teeb suurt kahju poliitik või kirjamees, kes kas asjatundmatusest või poliitiliste (politikaanlike) eesmärkide nimel kinnitab "statistikaline tuginedes", et eesti rahvas on hävingu või väljasuremise ohus. Veel ei ole iga eestlane nii hästi informeeritud, et olla immuunne selliste ütluste suhtes — kuid selleni on tarvis jõuda, et rahvas saaks olla sõltumatu ja enesekindel.

Info, kitsamalt statistilise teabe väärtuse ühiskonnale määrab suuresti selle pragmaatiline aspekt.

Statistilisi ja operatiivandmeid on tarvis igapäevaseks elukorralduseks ja juhtimisteggevuseks (nt. selleks, et korraldada sotsiaalabi vallas, peab valdavalitsus teadma abivajajate hulka ja nende seisundit).

Statistilist teavet on vaja ka ühiskonnas toimuvate protsesside ettenägemiseks, riigi ja selle kõigi haldusüksuste majandusliku seisundi pidevaks jälgimiseks. Ainult statistilise info põhjal on võimalik hinnata Eesti majandust turumajanduse huvidest lähtuvalt, teha põhjendatud järeldusi ja strateegilisi otsustusi.

Statistilist teavet on vaja ka infovahetuseks rahvusvahelisel tasemel, selleks, et olla mõõdetav ja võrreldav teiste (eriti Euroopa) riikidega, olla üks nende hulgas.

Vajalikul tasemel statistilise teabe saamiseks on tarvis:

— muuta Eesti statistika ja arvepidamine vastavaks rahvusvahelistele standarditele ja ÜRO poolt välja töötatud rahvamajanduse arvepidamise süsteemile (SNA);

— luua olulisimat statistilist teavet (n-ö. üldkogumi tasemel) sisaldavad omavahel kooskõlastatud registrid — rahvastiku-, ettevõtte-, kinnisvararegister jmt.;

— võtta ulatuslikult kasutusele statistiliselt korrektselt planeeritud valikvaatlused kui märksa odavamad ja usaldusväärsemad, sest nende täpsuse hindamiseks on võimalik kasutada statistikaaparatuuri.

Oluline on aga teadvustada tõsiasja, et *infokäsituse ning -levi tase ja kultuur määravad suurel määral ühiskonna arengu taseme, tema küpsusastme, samuti ka ühiskonna arengupotentsiaali*. Ilma vajalikul tasemel infoteenitusest pole võimalik ei riigikaitse ega võitlus kuritegevusega, ei mõistlik tööjõu- ega sotsiaalpoliitika. Info- ja statistikapoliitika vaegarengu taha ähvardab kinni jääda nii riigi kodanikkonna määratlus kui ka erastamise praktiline kulg.

Viljandis toimunud seminaril jõudsid statistika- ja infoala töötajad ühistele järeldustele nende määravate kitsaskohtade osas, mis praegusel hetkel on takistuseks statistika- ja infoalase tegevuse arengule Eestis.

Peaaegu kõik esinejad märkisid esimese raskusena *statistika- ja infoalase tegevuse juriidilise aluse puudumist*.

On tarvis tervet rida seadusi — registriseadust, andmekaitseseadust, uut versiooni statistikaseadusest jne. Need seadused peavad ühelt poolt kaitsma isikut (on ju üldlevinud hirm, et häid registreid võidakse hakata kasutama isikute vastu. Demokraatlikes riikides on kõik niisugused võimalused andmekaitse seadustega välistatud). Samuti peaksid need seadused kindlustama vajalike andmete laekumise, reguleerima andmete valdamise, haldamise, edastamise ja publitseerimise korra, andmete (andmetöötuste) hinna määramise reeglid jmt.

Mitmeski osas on tegelik elu seadusandlusest ette jõudnud — rahvastikuregistri ettevalmistustöö on jõudnud juba üsna kaugele, registriseadus puudub. Ka ettevõtteregister on teatavas mõttes oma seadusest välja kasvanud.

Tundub, et sedalaadi seadused on üpris kõvaks pähkliks Eesti klassikalise haridusega juristidele, kelle senistest õppekavadest info- ja statistikavaldkond on eemale jäänud. Siiski ootavad statistikud-informaatikud seaduste paketi kokkupanemisel hea tahtega juristide koostööd ja tuge.

Teisena nimetati rahapuudust. Eestis on statistika- ja infoalase tegevuse finantseerimine vähemalt suurusjärgu võrra väiksem kui naabermaades, kusjuures meil on töömaht märksa suurem — siin tuleb süsteeme (registreid, andmepanku, klassifikaatoreid) rajada, mitte üksnes normaalses töökorras hoida, teenindada ja edasi arendada, nagu arenenumates naaberriikides.

Rahapoliitika osas polnudki pretensioonid niivõrd suunatud finantseerimise mahu, kuivõrd selle sihipärasuse pihta. Oleks loomulik, et riikliku statistika- ja infopoliitika keskused peaksid töötama riiklikul finantseerimisel ja täitma riiklikke tellimusi. Oleks loomulik, et projekte finantseeritakse kuni tellimustöö valmimiseni ja tulemused ka realselt käiku lähevad. Praeguse olukorra iseloomustamiseks sobis väljend — linn, kus on väga palju pooleliolevaid maju — raha ei jätku nende lõpetamiseks ja kasutuselevõtmiseks, kuigi selline ehitamine on kalleim võimalik.

Kõneldi ka organisatsioonilistest puudustest, mõistes selle all riikliku info- ja statistikapoliitika üldist koordineerimist, mis praegu tõsiselt soovida jätab.

Olukorras, kus paljusid asju on tarvis väga kiiresti teha ning ei ole fikseeritud ülesannete prioriteete, pole võimalik asju teha otstarbekalt ja arukalt, õiges järjekorras (seadus—tellimus—projekt—realisatsioon—katsetamine—juurutamine). Alatihti toimib Parkinsoni seadus — enne ühe töö alustamist pidanuks olema lõpetatud teine, kuid osutub, et seda teist pole veel alustatudki. Mõningaid ülesandeid lahendatakse paralleelselt, koordineerimatult. Viimane puudutab ka välisabi kasutamist.

Üldine arvamus oli, et Eestis jätkub intellektuaalset potentsiaali ja reaalseid töötegijaid kõigi püstitatud ülesannete lahendamiseks, vältides sealjuures paljusid naabermaades tehtud vigu ning jõudes seega positiivsete tulemusteni neist märksa kiiremini.

Suuri lootusi pandi Riigikantselei juures äsja loodud Riigi Infosüsteemide osakonnale, mis peaks hakkama tegelema Eesti infopoliitika kujundamise ja arendamise, teadvustamise ja elluviimisega.

Samasuguste ülesannete täitmisele statistilise infotöötuse osas peaks asuma Riiklik Statistika Nõukogu, mille loomine tõsisis päevakorda käesoleva aasta alguses, kuid mille osas samuti otsustavad sammud on veel astumata.

Ometi jääb aktuaalseks ülesanne — teadvustada ühiskonna juhtivatele jõududele, poliitikutele, Riigikogu liikmetele, maksumaksjatele, et teabe ja statistika sihipärasele arendamisele kulutatud investeeringud ei ole luksus, need tasuvad end kõige kiiremini ja kõige kindlamini, ning põhjendamatu kokkuhoid selles valdkonnas tähendab paratamatult arengu aeglustumist, mahajäämust kõige elutähtsamates riigielu valdkondades.

Ene-Margit Tiit, Tartu Ülikooli matemaatilise statistika professor.

Helina Vigla, Eesti Statistikaameti peadirektori asetäitja.

Marje Kajandu, Riigi Arvutuskeskuse asejuhataja.

21. III 1993

ETTEVÖTTEREGISTRI VAJALIKKUSEST STATISTIKAS

Helina Vigla

Eesti Statistikaamet

Statistika eesmärgiks on üldsuse ja kõiqi tasandite juhtimisorganite varustamine adekvaatse ja õigeaegse informatsiooniga, mis võimaldab:

- kõiqi haldusüksuste majandusliku seisundi pidevat jälgimist;
- hinnata Eesti rahvamajanduse seisundit lähtudes turumajanduse huvidest;
- võrrelda Eesti majanduslikku olukorda teiste riikidega.

Seoses ümberkorraldustega majanduses tekib objektiivne vajadus statistika ja arvepidamise muutmiseks vastavalt rahvusvahelistele standarditele ja ÜRO poolt väljatöötatud rahvamajanduse arvepidamise süsteemile (SNA — *The System of National Accounts*).

Praeguse majandusstatistika taustal ei saa rahvamajanduse arvepidamist koostada nii, et see kajastaks kogu majandust. Peapõhjus on selles, et praegune majandusstatistika ei hõlma kogu majandust. 1992. a. laekus aruandeid 70–80% tegutsevatelt ettevõtetelt. Kogu majanduse kohta andmete kogumine on väga kulukas ja osa ettevõtteid jääb alati kogumist välja. Seepärast on otstarbekas kasutada väljavõttelise vaatluse meetodikat. Koondandmete kvaliteet väljavõttelise vaatluse meetodikat kasutades on rahvusvahelisi kogemusi arvestades parem kui kõiqi ettevõtete koormamisel aruandlusega.

Väljavõtteline vaatlus on odav, valikust tingitud viga on alati võimalik hinnata ja sellise vaatluse tulemusel saame kätte olulised seosed. Väljavõtteline vaatlus põhineb üldkogumil, millest valitakse valim kindlate tunnuste alusel, valim mõõdetakse ja järeldused tehakse üldkogumi kohta. Üldkogumi määramise ja valimi moodustamise aluseks on ettevõttereister. Ettevõttereistris on registreeritud üle 50 000 objekti, millest üle 44 000 on juriidilised isikud ja üle 6000 on tegevuskohad. Tegevuskoha ehk tegevuspaiga all mõistame juriidilise isiku teisel aadressil paiknevat objekti või samal aadressil paiknevat ja teise tegevusalaga tegelevat objekti.

Ettevõttereister peaks sisaldama kõiki juriidilisi isikuid. Tegelikult ei sisalda register kõiki tegutsevaid ettevõtteid, aga sisaldab lisaks tegutsevalle ettevõtetele veel mittetegutsevaid juriidilisi isikuid. Ettevõttereistrit korrigeeritakse kõiqi kättesaadavate infoallkate abil. See ei õnnestu kunagi sajaprotsendiliselt. Täiendavat infot saadakse sotsiaalfondi, maksuameti ja tolliameti kaudu. Ka statistiline aruandlus, mis laekub ettevõtetele, on üks infoallikas.

Kasutades ettevõttereistri informatsiooni, saab andmeid koguda tegevuspaikade lõikes tegevuspaikadelt, tegevuspaikade lõikes ettevõtetele ja kol-

mas võimalus on koguda andmeid ettevõtete lõikes ettevõtetelt. Igaüks neist kolmest variandist on seotud eri nõuetega ja annab erineva võimaluse statistika avaldamiseks.

Selleks, et koguda andmeid tegevuspaikadelt, peab olema täiuslik tegevuspaikade register, mis meil praegu puudub. Teine variant — koguda andmeid tegevuspaikade lõikes ettevõtetelt, ei nõua tegevuspaikade registrit, vaid annab hoopis andmeid selle registri koostamiseks. Selle tööga on alustatud, sest 1992. a. aasta aruandluses küsiti ettevõtetelt mõnesid andmeid ka tegevuspaikade lõikes. Suurem osa andmetest aga kogutakse 1993. aastal ettevõtetelt ettevõtte kui terviku kohta.

Ettevõttereistri baasil määratakse statistiline kogum majandusanalüüsiks ehk nn. ettevõttereistri statistiline profiil, mis on tarbimiskõlblik siis, kui ta haarab kõiki suurettevõtteid ning suurt osa väikeettevõtteid. Kõrvalejäänud väikeettevõtete osatähtsus ei tohi mõjutada majandusanalüüsi.

Statistiline profiil kontrollitakse täielikult, määratakse kindlaks ettevõtte põhitegevusala, omandivorm, ettevõtlusvorm.

Kõige tähtsam reegel on, et kõigis statistikaharudes kasutatakse ühist kogumit, igale ettevõttele rakendatakse seda tegevusala ja teisi tunnuseid, mis talle kuuluvad ettevõttereistri statistilise profiili kohaselt. Kui mingil põhjusel tegevusala kood ettevõttereistri statistilises profiilis ei ole kooskõlas tegeliku tegevusalaga, saab seda muuta, aga ainult juhul, kui muutmiseks tehakse keskne otsus, mis kajastub kõigis statistikaharudes.

Tuginedes ettevõttereistri andmetele ja lähtudes registri statistilisest profiilist, koostatakse valimid statistiliste küsitluste läbiviimiseks. Tänapäeval ei ole võimalik korraldada majandusalaseid statistilisi vaatlusi ilma ettevõttereistriga.

STATISTIKA JA INFORMAATIKA ARENGUSUUNAD

Kuldev Ääremaa, Ruth Ääremaa

Tartu Ülikool, Matemaatilise Statistika Instituut

1. Statistika ja informaatika

Riigi arengutaseme üheks näitajaks on rahvusvahelistele standarditele vastava kirjeldava statistika olemasolu. Kuid see näitaja jääb tühiseks, kui riik siseselt ei oska kasutada statistilist informatsiooni oma arengu määramiseks. Teatud rajajooneks riigi arengus on tema üleminek tööstushiskonnast infoühiskonda. Kaasaegne statistika saab tugineda vaid kaasaegsele informaatikale.

Ei ole võimalik tõmmata ranget piiri statistika ja informaatika vahele ja ka käesolevas ettekandes me ei püüa seda teha. Mõne sõnaga selgitaks siiski nende kahe teadusharu olemust ja kokkupuuteid.

Statistika on massiliselt esinevate nähtuste vallas seaduspärasusi käsitlev teadus. Statistika esmane etapp on andmehõive, s.t. küsimus sellest, milliseid andmeid koguda, kuidas saavutada andmete esinduslikkus, ja muud sellega seotud probleemid. Seejuures ei tegele statistika mitte niivõrd infohõive tehniliste, kui just sisuliste aspektidega. Andmehõivele järgnevad töötusetapid, kus andmeid töödeldakse statistikale omaste meetoditega.

Informaatika on valdkond, mis hõlmab info kogumist, edastamist ja töötlemist ning selle tarbeks meetodite ja tehniliste vahendite loomist ja kasutamist.

Toodud kahe definitsiooni järgi võiks statistika sisalduda informaatikas, kuid nagu eelpool öeldud, ei ole praegu oluline mitte piiri tõmbamine, kui just tõmbamata jätmine. Teatud probleemide sarnasus selgitab ka seda, miks riigi juhtimiseks vajalike andmete hõive ja töötlemise probleemidega tegelevad nii statistikud kui ka informaatikud ning omavahelisi ebakõlasid ei teki seni, kuni kumbki pool ei soovi monopoliseerida selgelt kahe valdkonna ühiselt lahendamisse kuuluvaid probleeme. Selgelt eralduvaid osi on olemas, näiteks jääb informaatika pärusosaks sidekanalite läbilaskevõime, koodide optimeerimine jne.

Vaadeldes statistika vaatepunktist, peab informaatika looma baasi kaas- aegseks andmehõiveks ja statistiliste meetodite kasutamiseks kaasaja tasemel.

2. Registri mõiste

Viimasel ajal on väga laia kasutuse leidnud sõna "register". Kõikjal luuakse registreid: hoonete, kinnisvara, sõidukite, passide, invaliidide, looduskaitsese jne. jne. Oma semantiliselt tähenduselt on register kindlal kujul korras- tatud nimestik või loend. Oma tähenduse järgi ei seo register endaga mitte

mingit andmehõivesüsteemi ega struktuuri, ning selles mõttes on enamik viimasel ajal loodud registreid tööpoolest vaid registrid.

Registrite loomine ja kasutuselevõtt iseloomustab seisundit, milles me praegu oleme. Kuid registritest liigne vaimustumine kätkeb endas ka ohte. Võtame näiteks juba laialt tuntuks saanud sõna "rahvastikuregister". Väga paljudes kohtades mõistetakse seda kui isikuandmeid sisaldavat arvutustehnika vahenditega töödeldavat andmekogumit, mis sisaldab riigi kõigi elanike, samuti mujal elavate Eesti riigi kodanike isikuandmeid. Kui võrrelda seda varajasemast ajast pärineva mõistega — vabariiklik aadresslaud, siis erinevus on ainult infotöötuses kasutatavates tehnilistes vahendites. Kaasaegsed vahendid infotöötuses on loomulikult suur samm edasi informajanduses, kuid mitte veel see samm, mis viib meid üle infoühiskonna lävepaku.

Arvutustehnikal põhinevate üksikregistrite loomine on viimasel paaril kolmel aastal saanud igapäevaseks nähtuseks. Teatud piirides on see mõistetav ja teatud piirides ka vajalik, kui kõige kiirem andmete korrasdamise ja süstematiseerimise viis. Keeruliseks läheb aga olukord siis, kui registri andmehõive on väga lai ning andmete aktualiseerimine nõuab hästi funktsioneerivat süsteemi. Sellised registrid püüavad muutuda omaette administratiivüksusteks, mis tegutsevad oma põhikirja alusel ja millele info andmise kohustus pannakse peale seadusandlikus korras. Infotöötlemine muutub tegutsemisvaldkonnaks iseeneses ja teatud mõttes kaugeneb ühiskonna juhtimisest.

Omades hästi palju registreid, on loomulik, et meil on olemas ka automatiseeritud registrite register ja valitsuse määrus registrite registri loomise kohta.

Ei maksa arvata, et siia maani on arvepidamine puudunud ja alles nüüd, seoses arvutite kasutuselevõttuga on ometi võimalus luua täpne arvestus. Tegelikult on meil enamik registreid (v.a. näiteks repressseeritute register, endiste omanike register jms.) alati olemas olnud ja sageli veel paremas korras kui praegu. Ainult et mitte arvutis, vaid nõõrraamatutena, perfokaartkataloogidena jne. Kõik oleks korras, kui arvutite tulekuga oleks kaasnenu registreerimise lihtsustumine ja suurem kord, kuid seda ei ole mitte alati märgata. Põhipõhjuseks siin tuleb pidada seda, et me oleme alles selles arengustaadiumis, kus automatiseerime üksikelemente, terviklike infosüsteemide loomise aeg on veel ees.

3. Registrite ajastu

Üksikregistrite ajajärk oli (on) meie arengus tööstusühiskonnast infoühiskonda loomulik nähtus. Selle tekkimine oli paratamatus ja teatud võimenduse sellele andis küllaltki järsk üleminek kaasaegse infotehnoloogia kasutamisse. Objektivsete põhjustena võib esile tuua kolme:

- 1) teadmiste ja kogemuste pagas — kõik meie juhtivad infotöötuskeskused tegelesid agaralt automatiseeritud juhtimissüsteemide projekteerimisega ülimalt tsentraliseeritud ühiskonnas;

2) tehniline baas — EC tüüpi suurarvutid olid pikka aega arvutustehnika vallas valitsevaks; nende kasutamine orienteeris ka mõtlemismallid rohkem tsentraalsete kui hajutatud andmebaaside poole;

3) vajadus kiiresti kohaneda turumajanduse (isemajanduse) tingimustega, leida uued finantseerimisallikad ja püsitöö tulevikuks; on ju üksikuna tegutseva registri loomine võrratult lihtsam kui infotöötlusvõrgu loomine.

Eriti tuntav on praegu püüe luua mitmesuguseid keskregistreid, mõtlemata sellele, kuidas nad praktikas funktsioneerima hakkavad ja kas infohõive probleemid lihtsustuvad või muutuvad hoopistükkis keerulisemaks. Sageli tundub, et siin ei ole tegemist infopoliitilise küsimusega, vaid poliitilise — kes valdab informatsiooni, see valitseb. Ehitades üles tallinnakeskset Eestit, on loomulik koondada sinna ka kogu informatsioon.

Loodame, et üksikregistrite ajastu saab siiski otsa ning infotöötlusprobleemidele hakkame lähenema infovõrgu tasandil, kus kindla valdkonna probleeme lahendatakse terviklikult selle valdkonna suhtes, arvestades ka infovõrkude omavahelist infovahetust. Sellele peavad kaasa aitama kindla suunitlusega riiklikud sihtprogrammid, nagu seda on praegu näiteks Eesti Informaatika Nõukogu kureeritav riigiturvüsteemide sihtprogramm. Ilmselt saab ja tuleb ka praeguseks loodud registrid liita vastava valdkonna terviklikesse infosüsteemidesse, kuid see võib osutada juba administratiivselt raskeks (tegutseb ju iga riiklik register oma põhimääruse järgi ja igasse struktuuriüksusesse on programmeeritud enese alalhoiuinstinkt), rääkimata juba infooloogilisest sobivusest ühistööks.

4. Ait üles või ülalt alla?

Eeldame, et korralikult projekteeritud infosüsteem sisaldab nii lokaalseid registreid kui ka tsentraalset registrit ja andmete paigutus (asukoht) on optimeeritud kasutamise vajadusest, kasutamissageduset ja sidekanalite maksumusest lähtudes. Teeme veel täiendava eelduse, et kogu infosüsteemi üheaegne käikulaskmine on raskendatud kas majanduslikel kaalutlustel või siis selletõttu, et ei ole võimalik katkestada korraga kogu teatud valdkonna infotöötlust uue süsteemi installaerimise ajaks. Nendel tingimustel kerkib üles küsimus: kust alustada, kas lokaalsetest registritest või tsentraalsest. *Pro et contra* argumente saab tuua nii ühe kui teise variandi kasuks, kindlasti sõltub see aga otseselt süsteemi olemusest, nii et allpool toodud analüüs ei pretendeeri täielikkusele ja õigupoolest ei võimalda seda ka selle arutluse maht.

Oletame, et (lihtsustatud käsitluses) on keskregister lokaalsete registrite summa. Keskregistri selline ülesehitamine on vajalik sel juhul, kui registris sisalduvate objektide kohta käivate kõikide üksikute teadmine on oluline keskregistri andmeid tarbivates asutustes. See on küll vastuolus üldise seaduspäraga, et mida kõrgema võimutasandiga on tegemist, seda üldistatumat informatsiooni ta vajab, kuid see selleks. Alustades sel juhul infosüsteemi käivitamist keskregistrist, loome võimaluse lokaalregistrite käivitamiseks info põhjal, mis on kooskõlastatud eri regioonide vahel. Tehniliselt tekib siin aga küllaltki palju ebaotstarbekust. Nimelt, kui eeldame, et

info kogumine toimub lokaaltasandil ja ta ei ole arvutipärasel infokandjal, siis ei ole otstarbekas kasvõi juba tööhõive seisukohalt koondada informatsiooni sisestamistööd keskusesse. Kui aga info on juba arvutis, siis on sisuliselt lokaalregister käivitatud ja küsimus võib olla vaid lokaalregistri töö korrigeerimises. Erandi moodustaksid siin olukorrad, kus keskregister luuakse lokaaltasandist sõltumatu informatsiooni põhjal ja vastava ala info-töötlus käivitub alles pärast keskusest andmete saamist. Selle skeemi alla kuulub idee käivitada rahvastikuregister rahvaloendusandmete põhjal. Siin on aga mitmed vastunäidustused, sest praegu kasutavad lokaaltasandid palju täpsemat informatsiooni kui seda on rahvaloendusel suusõnaliselt isiku poolt tehtud ütlused. (Küll on aga võimalik vastupidine: rahvaloendusel kattuvad andmed võtta rahvastikuregistrist ja sellega vähendada loendamisel tehtavat tööd.)

Muidugi ei tohi siin unustada teist poolt — kui keskregistri käivitamine on riigi seisukohalt niivõrd oluline, et see korvab muud kahjud, siis ülaltoodud argumente ei tarvitse arvestada.

Üldiselt ei pruugi keskregister olla lokaalregistrite summa, vaid informatsioon sinna kogutakse kasutaja üldistustaseme vajadust arvestades. Sellisel juhul võime valmis keskregistri põhjal käivitada vaid lokaalregistri teatud skeemi, mis on aluseks lokaaltaseme infovajadust rahuldava registri käivitamisele. Infohõivega ja info ettevalmistamisega seotud probleemid on aga samad, mis eespool vaadeldud juhul.

Ait-üles ja ülalt-alla probleemid kaovad ära, kui loome infovõrgu, kus kõik andmed on tsentraalses baasis ja lokaaltaseme tarbijal on sellesse baasi otsejuurdepääs. Selle kõige lihtsama mudeli oluliseks puuduseks on aga infokanalite kasutamise maksumus. Kui kohalik omavalitsus peab maksuma kinni need kulud, mida tema ametnik kulutab igapäevatööks vajaliku info hankimiseks keskbaasist, siis tekivad paratamatult "põrandaalused" kohalikud infopangad.

5. Kaasnev või iseseisev infosüsteem

Infotöötlusprojekti loomisel on tarvis analüüsida, millisel määral nõuab loodav infotöötlussüsteem spetsiaalsete infotöötlusinstituutide loomist ja millisel määral laekub informatsioon võrku vastava ala töötajate igapäevase tegevuse tulemusena. Arvame, et vaid mõningatel erijuhtudel on otstarbekas luua iseseisev infosüsteem, mis ei kopeeri eksisteerivaid haldusstruktuure. Tegelikult on elu juba enamuse asju paika pannud: ainukese puhtakujulise iseseisva andmetöötlusvõrguna on senini eksisteerinud Eesti Statistika ja ilmselt ei vähene tema roll ka tulevikus. Enamike teiste valdkondade korral paistab aga eelistatum olevat kaasnev infosüsteem, mis ei nõua täiendavate töökohtade loomist, tema eesmärgiks on parandada infotöötlust kaasaegse infotehnoloogia baasil.

Rõhutame siinjuures, et meie väited ei ole absoluutsed ja infotöötluse kvalitatiivne erinevus võib tingida ka teatud muudatusi arusaamades.

Vaatleme allpool mõningaid näiteid kaasnevatest ja iseseisvatest infovõrkudest.

A. Autoregister.

Kuni arvutitel baseeruva registri loomiseni oli autode kartoteek liiklusmiilitsa alluvuses ja seda loeti vähemalt rahva arvatës normaalseks. Mõne aja eest tekkisid iseseisvad registrikeskused (autode registreerimise infovõrk), mille põhiülesandeks on info salvestamine (auto registreerimine) ja info väljastamine (näiteks politsei vajadusteks).

Kas antud tingimustel selline lahutamine on õige? Kui uskuda, et autoregistri loomisel viidi läbi igakülgne analüüs, siis ei ole praegu meil ilma konkreetsete andmeteta mingit põhjust selles kahelda.

B. Sotsiaalabi infosüsteem.

Arvutustehnikal põhinevat sotsiaalabi süsteemi praegu veel ei eksisteeri, küll on aga mitmetes valdades (ja maakondades) loodud korrektsed kartoteegid, mida kasutatakse igapäevatöös. Raske oleks ette kujutada, et püstitaksime ülesande luua sotsiaalabi vajajate keskregister või spetsiaalne infovõrk, kuhu sotsiaaltöötajad peaksid sisestama andmed kõikide infovajajate kohta. Loodav sotsiaalabi infosüsteem peab olema abivahendiks sotsiaaltöötajatele ja igale järgmisele juhtimistasandile peab jõudma vaid teatud statistilise üldistuse läbiteinud info.

C. Rahvastikuarvestuse infosüsteem (rahvastikuregister).

Rahvastikuarvestuse infosüsteem on üks keerulisemaid ja tähtsamaid Eesti infoaastikul. Tema andmehõive on väga lai, andmed tekivad väga mitmesuguste toimingute tulemusena (sünnid, surmad, abiellumine, kodakondsuse andmine, passi väljastamine jne.) ning nende kogumaht on suur. Olulise tähtsusega on siin vajadus infosüsteemi järele vahetus haldus- ja majandustegevuses. Ilma laitmatult funktsioneeriva rahvastikuarvestuse süsteemita ei saa olla kaasaegset politseisüsteemi, meditsiiniteenistuse süsteemi, kinnisvararegistrit jne. jne. Rahvastikuarvestuse süsteemi asendamine keskregistriga on primitiivne lahendus, mis jätab enamiku probleemidest ikkagi õhku rippuma. Kaasaegsest rahvastiku arvestamise süsteemist võime rääkida alles siis, kui kogu probleemide komplektile lähenetakse kui ühtsele tervikule, vaadeldes info kogumist, edastamist, arhiveerimist, levitamist ja teisi toiminguid süsteemselt.

6. Vabariiklik infopoliitika

Riigi infopoliitika kujundamiseks ja elluviimiseks loodi Eesti Vabariigi valitsuse määrusega 31.07.90. a. Eesti Informaatikafond (tegelik töö algas 1991. aasta algusest), mille tööd juhib Informaatiikanõukogu, olles pädev tegema strateegilisi otsuseid infopoliitika valdkonnas. Valitud poliitika elluviimiseks on kaks võimalust — kas finantseerida (mittefinantseerida) konkreetseid projekte või teha valitsusele ettepanekuid informaatikat puudutavate määruste vastuvõtmiseks.

Kahjuks on aga asjal veel üks külg, mis võib sageli saada määravaks, ja nimelt vajaduste—võimaluste vahekord. Eesti vajadused informaatikaalaks tööks on niivõrd suured, et see pisku, mis on Informaatikafondist

jagamiseks, ei saagi luua korralikku alust ühtse infotööstusstrateegia väljaarendamiseks.

Käesolevaga me ei sea endale ülesandeks hinnata Informaatikafondi tegevust ja alljärgnev Informaatikafondi kaudu finantseeritavate tööde lühiloetelu annab teatud läbilõike seisust kus me oleme. Tööde täielik nimekiri on Informaatika fondis koostatud ja fondi põhikirjas sätestatud avalikustamise põhimõttele ka sealt kättesaadav. Siintoodud loetelu iseloomustab 1990.–1992. aastatel finantseeritud töid vaid põgusalt, liigitades neid töö nimetuses kasutatud mõistete järgi ja arvestades, et üht tüüpi objektidele orienteeritud infosüsteem on enamasti ikkagi vaid register.

Registrid ja andmebaasid: sundvõrandatud eluhoonete andmebaas, eesti filmide andmebaas, vabariiklik ehitusregister, hariduse andmebaas, metallikaitse infosüsteem ja andmepank, majandusinfo klassifikaatorite andmepank, kohalikud hooneregistrid, ökoloogia bibliograafia andmebaas, firmanimede teabesüsteem, eesti trükiste andmebaas, välisperioodika andmebaas ja kataloog, ettevõttere register, demograafia andmepank, eesti teederegister, kaubalitsentside andmebaas ja kontrollisüsteem, õigusaktide register, endiste omanike ja vara register, tootmisliitsentside andmepank, rahvastikuregister, registrite andmebaas, andmebaas "Maksuhäired", represseeritute register, keskne hooneregister, välisraamatute register, terminoloogia andmebaas.

Haldusinfosüsteemid: valla omavalitsuse tüüpinfosüsteem, Keila linna infosüsteem, Tartu linna infosüsteem.

Ametkonnasisesed Infosüsteemid: tolli infosüsteem, kõrgkoolide infosüsteem, Eesti Valitsuse infosüsteem, Kaubandus-tööstuskoja infosüsteem, tööturu infosüsteem, välisministeeriumi infosüsteem, riigi haldusinfosüsteemi kontseptsioon.

Temaatilised Infosüsteemid: konverentsi tüüpinfosüsteem, valimistulemuste analüüsi infosüsteem, ehitusinvesteeringute infosüsteem, eesti ehituse infosüsteem.

Toodud nimekirja võib analüüsida mitmest aspektist, kuid praegu piirduksime vaid mõningate aspektide esiletoomisega:

- 1) killustatus, selge suunitluse puudumine;
- 2) väga suur registrite ja teabesüsteemide osakaal;
- 3) haldusinfosüsteemide arendamise mahajäämus;
- 4) Informaatikanõukogu on finantseerinud esitatud taotlusi, mitte neid tekitanud, suunamaks üldist infopoliitikat.

Pöördudes veelkord Informaatikafondi juurde tagasi, peaks määratlema täpsed tegutsemisvaldkonnad ja tulevikku suunatud strateegilise liini. Ametkonnasiseste infosüsteemide loomine tuleb jätta täielikult ametkonna sisemiselt finantseeritavaks tegevuseks, Informaatikanõukogu võib olla eksperdik, et hinnata projektide tähtsust või osatähtsust.



EESTI STATISTIKASELTS

Seminari 'Statistika tarkvara Eestis'
programm

20. aprill 1993.

Tartu Statistikabüroo, Sõpruse pst 4, saal.

I. istung

- 10.00 - 10.10. Seminari avamine.
10.10 - 10.30. E.Tiit - Missugust statistika tarkvara me kasutame õppe- ja teadustöös?
10.30 - 11.30. P.Nõykki - Matemaatikapakett Mathematica (inglise keeles).
11.30 - 12.00. T.Kinkar - SAS ja selle võimalused.
12.00 - 12.15. M.Viil - SAS GRAPH.
12.15 - 12.30. Vaheaeg
12.30 - 13.00. J.Pelt - Aegridade analüüsi originaalpakett ISDA.
13.00 - 13.30. J.Roosaare - Territoriaalsete andmete analüüsi paketid ja statistika osa nendes.
13.30 - 15.00. Lõunavaheaeg ja demonstratsioonid.
I arvuti - MATHEMATICA (P.Nõykki),
II arvuti - IDRISI (J.Roosaare),
III arvuti - ISDA (J.Pelt).

II. istung

- 15.00 - 15.30. R.Veetõusme - Statistika tarkvara rakendamine ESAs. Olevik ja perspektiivid.
15.30 - 16.00. A.Kaarna - Programmipakett Blaise ankeetide kirjutamiseks.
16.00 - 16.30. T.Mõls - Andmebaasisüsteemide kasutamisest statistikaandmete ettevalmistamisel.
16.30 - 17.00. E.Käärrik, M.Vähi - STATGRAPHICS ja tema kasutamine õppetöös.
17.00 - 17.15. R.Oks - Aegridade analüüsi võimalused süsteemis STATGRAPHICS.
17.15 - 18.30. Demonstratsioonid.
I arvuti - SAS (S.Koskel, M.Viil),
II arvuti - Blaise (A.Kaarna),
III arvuti - STATGRAPHICS (M.Vähi, M.Thetloff),

Eesti Statistika Selts, Liivi 2, Tartu, Estonia EE2400

Tel: 01434-30641 Fax: 01434-35440 E-mail: ess@ulam.ut.ee



EESTI STATISTIKASELTS

21. aprill 1993.

III. istung

- 9.00 - 10.00. J.Puranen - SURVO ja selle kasutamine õppetõõs (inglise keeles).
10.00 - 10.30. L.M.Tooding - Seosemudelite kasutamine Uhis-konna- ja majandusteadustes.
10.30 - 11.00. U.Oja - NSDstat ja selle koolis kasutamise võimalused.
11.00 - 11.15. Vaheaeg.
11.15 - 11.30. I.Traat - Valimiteooria õppepakett ja selle kasutamise kogemus.
11.30 - 12.00. A.M.Parring, J.Vilismäe - Ülevaade paketi SYSTAT võimalustest.
12.00 - 12.15. S.Dudkina - Kindlustusmatemaatika pakett (vene keeles).

12.15 - 14.00. Lõuna ja demonstratsioonid.

- I arvuti - SURVO (J.Puranen),
II arvuti - NSDstat (U.Oja),
III arvuti - SYSTAT (A.M.Parring, J.Vilismäe).

IV. istung

- 14.00 - 14.30. T.Mõls - Lineaarsete mudelite rakendamine bioloogiaülesannetes (SAS GLM).
14.30 - 15.00. E.Pärna, M.Uba - SAS kasutamise kogemus ELVIS.
15.00 - 15.30. A.Kiviste - Metsandusandmete analüüsi kogemustest.
15.30 - 16.00. B.Kubo - Arvutivõrgu kasutamisest statistilise andmestiku analüüsimisel.
16.00 - 16.15. U.Randaru - Tarkvarakonvrentsi muljeid Saksa-maal.
16.15 - 16.30. Seminari lõpetamine.
16.30 - 18.30. Demonstratsioonid.

- I arvuti - SPSS (E.Käärrik),
II arvuti - valimiteooria pakett (I.Traat),
III arvuti - vaba ekraan.

Eesti Statistikeselts, Liivi 2, Tartu, Estonia EE2400

Tel: 01434-30641 Fax: 01434-35440 E-mail: ess@ulam.ut.ee

MILLIST STATISTIKA TARKVARA ON KASUTATUD JA KASUTATAKSE TARTU ÜLIKOO LIS?

Ene-Margit Tiit

Tartu Ülikool, Matemaatilise Statistika Instituut

Statistika tarkvara kasutamise algus Tartu Ülikoolis langeb ühte Tartu Ülikooli arvutuskeskuse rajamisega (1959), mil siin hakkas tööle esimese põlvkonna suurarvuti *Ural*.

Nagu paljudes muudeski arvutuskeskustes kogu maailmas, olid esimeste loodud programmide seas keskväärtuste ja korrelatsioonikordajate arvutamise, ühe- ja kahemõõtmeliste jaotustabelite koostamise jmt. statistika-programmid.

Arenedes kaunis iseseisvalt, eraldatuna Läänest ja vältides sõltuvust Idast, läbis Tartu Ülikooli statistikute kollektiiv ja tema poolt kasutatav statistika tarkvara (mõningase hiinemisega) kõik need arenguetapid, mida nimetatud valdkonnad on läbi teinud kogu maailma praktikas.

1. *Statistika üksikprogrammide loomise periood.* Niipea, kui TÜ Arvutuskeskuse suurarvuti tööle hakkas, ilmusid kohe ka tellijad, kes seni olid suuremahulisi, kuid lihtsaid statistikaarvutusi teinud käsitsi (mehhaaniliste arvutitega). Need olid kõigepealt arstiteadlased ja sotsioloogid, ning nende huvides hakati peagi standardseid statistikameetodite programme edasi arendama vastavalt *andmeanalüüsi* nõuetele.

Nn. ankeeditöötlusprogrammis (S. Veldre, S. Koskel, L. Vöhandu), mis võimaldas leida kahemõõtmelisi (tingimatuid ja ka tinglikke) jaotustabeleid koos jaotusparameetritega (s.h. mitmed seosekordajad), hakati arvestama tühikute, s.o. puuduvate vastuste olemasolu. Võeti kasutusele tunnuste tüpologia (E. Tiit, T. Veldre).

Populaarseks meetodiks paljutunnuseliste andmete analüüsimisel sai *faktoranalüüs*, mida L. Vöhandu tutvustas fakultatiivloengute seerias. Kuuekümnendate aastate algusest peale on TÜ arvutuskeskuses töötanud mitu versiooni faktoranalüüsi programme (T. Veldre, K. Pragi jt.), mis realiseerisid erinevaid hindamis- ja pööramise meetodeid, individuaalsete faktorkaalude arvutamist jmt. Huvitav oli ka Ahmavaara *transformatsioonanalüüsi* programm (T. Veldre), mis võimaldas erinevaid faktoranalüüse omavahel võrrelda.

2. *Originaalse statistikasüsteemi loomine.* Alates kuuekümnendate aastate keskpaigast hakkas Tartu statistikutele selgeks saama, et andmete edukaks statistiliseks analüüsimiseks on tarvis üksikprogrammidele üle minna statistikaprogrammide paketele või süsteemile, mis võimaldaks erinevaid meetodeid vabalt kombineerida. Niisuguste süsteemide loomise ja kasutamise kohta Läänes (eriti BMDP ja selle eelkäijad) jõudis pärale teave ka

statistikakirjanduse vahendusel. Tihenes side Moskvaga, kus üleliidulistel seminaridel statistikasüsteemide vajalikkusest palju räägiti, ometi reaalse-
te tulemusteni jõudmata.

Eesti esimeseks diskreetse andmestiku töötlemise süsteemiks sai L. Võhandu juhtimisel loodud SODI (TPI ja Eesti Raadio), mis käivitus kuuekümnendate aastate keskel ja leidis väga ulatuslikku rakendamist ankeetide töötlemisel.

Mõned aastad hiljem valmis TÜ originaalse andmetöötussüsteemi esimene versioon (T. Veldre, L.- M. Tooding). See andmetöötussüsteem (mida järkjärgult täiendati ning mis hiljem sai nimeks "Stella") teenis TÜ kõigi teaduskondade andmetöötusvajadusi enam kui kahekümne aasta jooksul.

Paralleelselt programmisüsteemi kirjutamisega toimus ka teoreetiline töö: mõtestati lahti statistilise andmeanalüüsi eesmärk ja eripära klassikalise matemaatilise statistikaga võrreldes, töötati välja kasutatav mõistetesüsteem, rakendatavad printsiibid (E. Tiit, L.- M. Tooding). Programmipakett "Stella" (valminud L.- M. Toodingu juhatusel ja E. Tiidu teoreetilisel juhendamisel) sisaldas põhilisi klassikalisi statistikaprotseduure — ühe- ja kahemõõtmelisi analüüsi (tabelid, jaotusparameetrid, hüpoteeside kontroll statistilise, monotoonse, regressioon- ja korrelatsioonse olemasolu kohta — J. Viilismäe), korrelatsioon- ja regressioonanalüüsi (S. Koskel), faktoranalüüsi koos ligikaudsete usalduspiiride arvutamisega (I. Traat), dispersioonanalüüsi juhusliku ja determineeritud mudeli alusel (s.h. mudelite kirjeldamine küllaltki üldise keele abil — T. Kelder). Klasteranalüüsi alamsüsteem (R. Ääremaa) sisaldas üsna palju võimalusi, s.h. tunnuste rühmitamist ja mitnesuguse kattuvusastmega rühmade moodustamist. Mitmeses regressioonanalüüsis oli realiseeritud ka võimalus leida kõik statistiliselt samaväärsed optimaalsed mudelid (T. Kelder).

Tartu Ülikoolis loodud tarkvarasüsteemis oli algusest peale rakendatud üpris paindlik tühikutega opereerimise protseduur, mis võimaldas maksimaalselt säästa vaatlusmaterjali. Samuti toimis protseduuride lubatavuse kontroll tunnuse tüübi järgi ning oli võimalik automaatselt välja eraldada jämedad vead. Tunnuseteisendused olid võimalikud loogiliste ja aritmeetiliste operatsioonide alusel, samuti oli realiseeritud lai valik teisendusprotseduure andmestike jaoks (alamhulgad tingimuste järgi, ühendamised jm.).

TÜ statistikasüsteemi loojad publitseerisid oma tulemusi pidevalt kahes Arvutuskeskuse väljaandes — eestikeelses laiale tarbijaskonnale mõeldud seerias "Programme kõigile" ning teadustööde kogumikus "Arvutuskeskuse tööd". Esimesest kogumikust oli statistikale pühendatud 8 vihku, kokku ligi 800 lehekülge, ning tänu sellele oli "Stella" üks paremini kirjeldatud ja dokumenteeritud statistikapakette terves omaaegses Nõukogude Liidus.

Teises trükisteseerias avaldati niihästi programmide teaduslikke kirjeldusi kui ka originaalseid meetodeid käsitlevaid teadusartikleid statistikateadlaste, õppejõudude, programmeerijate ja aspirantide sulest. Kajastamist leidsid ka mõned üliõpilastööd. Statistikaemalisi kogumikke ilmus paarikümne aasta jooksul tosinajagu.

3. *Programmipakettide adapteerimine ja testimine.* Programmide loomine kohapeal oli paratamatu seni, kuni N. Liidus toodetav riistvara ei ühildunud Läänes toodetuga — nii oli see suurarvutite *Ural* ja *Minsk* puhul.

Tõsi, oli olemas võimalus võtta üle idakolleegide loodud tarkvara, kuid siin oli takistuseks põhimõtteliselt erinev lähenemine: programmide kui valmisproduktide turu täieliku puudumise tõttu oli valdav enamus programme toodetud kohalikuks kasutamiseks ja neil puudus vajalik dokumentatsioon kasutuselevõtmiseks mujal. Lisandusid veel tehnilised probleemid seoses arvutite erinevustega, seadmete kapriissusega jn. Seetõttu — hoolimata küllalt headest töösuhetest Vilniuse, Minski, Moskva, Novosibirski jt. kolleegidega, oli nende poolt loodud tarkvara kasutamine Tartus pigem erand kui reegel.

Olukord muutus IBM-arvutitega ühilduva nn. ühtsusseeria (EC) suurarvutite kasutuselevõtmise järel, mil avanes põhimõtteline võimalus kasutada Läänes toodetud standardset tarkvara.

Kuivõrd litsentseeritud tarkvara ostmise valuuta eest oli mõeldamatu, tuli kasutada mitmesuguseid muid teid, neist käepäraseim oli programmipakettide nn. adapteerimine lepingute alusel.

Mitmesugusel teel (näiteks India kaudu, kontrollimata andmetel vahetuskaubana riistvara vastu) saabus Nõukogude Liitu ebamäärase päritoluga tarkvara pakette, milles oli suuremal või vähemal määral kasutatud Läänes valmistatud standardse komertstarkvara (enamasti vananenud versioonide) mooduleid. Valdavalt oli niisugune tarkvara nõrgalt dokumenteeritud ja vähe töökindel. Suhteliselt soodsaks lepingutööks oli taolise paketi läbiurimine, testimine ning sellele venekeelsele tarbijajuhendi koostamine. Niisugusel teel omandas Tartu Ülikooli Arvutuskeskus paketi SSP, hiljem ka paketi nimega GSARS, mille adapteeritud variandile anti nimeks SAI-SI, ning mis osutus sisaldavat SPSS 1977. aasta versiooni. Arusaadavalt kirjutati nendele pakettidele kohe ka eestikeelne tarbijajuhend.

Samal viisil adapteeriti ja kommenteeriti Valgevene Matemaatika Instituudis Minskis paketti BMDP, mis sai üldkasutatavaks N. Liidus, ning mis ka Tartus kasutusele võeti. Läbirääkimised käisid ka SAS-süsteemi adaptatsioonist Tartus, meie kätte sai selle süsteemi suurarvuti 1985. aasta versioonist ka üks N. Liidus paljundatud koopia, kuid lõplikult otsustati see prestiižikas ja majanduslikult tasuv töö siiski ühele Moskva firmale anda.

Seega muutus 1980. aastatel ka N. Liidus üsna mõttetuks uue tarkvara loomine standardsete statistikaprotseduuride tarvis, ning selle asemel muutus aktuaalseks tarkvara testimise ülesanne, mis saigi lähiaastate teadustöö üheks oluliseks suunaks Tartus.

Loodi originaalne testimise meetodika nn. täpsete valimite meetodil (E. Tiit), mis aitas hästi ka dokumenteerimata protseduuride sisu mõistatada. Teiseks tegevusvaldkonnaks oli tarkvara testimine reaalsete andmestike baasil. Testimisülesannete lahendamiseks loodi vastav tarkvara testimispakettide näol.

4. *Suurarvutite kasutusaja tipp — 1980. aastad.* Kokkuvõttes võib kinnitada, et 1980. aastate keskpaigaks saavutati Tartu Ülikooli Arvutuskeskuses statistika tarkvara osas märkimisväärne tase — praktiliselt oli võimalik kasutada pakette BMDP, SAS ja SPSS (kuigi mitte uusimaid versioone). Töötas originaalsüsteem "Stella", mis, olgugi kodukootud, vastas oma põhimõtelt standardsete statistikapakettide nõuetele, kuid oli eestikeelne ja korralike juhendmaterjalidega varustatud. Selletõttu eelistaski enamus TÜ Arvutuskeskuse arvukatest klientidest nimelt seda paketti. Eesti- ja venekeelne juhendmaterjal oli olemas ka SPSS (SAISI) jaoks, mida kasutati peamiselt õppetöös.

Suurarvutite ebasõbralikkuse tõttu (mis eriti omame oli Nõukogude toodetele) toimus kogu töö dispetšersüsteemis — tarbija ei olnud üldjuhul otsekontaktis arvutiga, vaid esitas tellimuse suuliselt vahendajale, kelleks oli statistiku väljaõppe ja kogemusega matemaatik-programmeerija. Reeglina konsulteeris iga tellijat ka kõrgema kvalifikatsiooniga statistikateadlane (õppejõud). See tagas tulemuste suhteliselt kõrge kvaliteedi ja usaldusväärsuse, kuid paratamatult vähendas mõnevõrra lahendusvariantide mitmekesisust.

Sellele perioodile on iseloomulik statistikameetodite üha ulatuslikum kasutuselevõtmine Tartu Ülikooli teadustegevuses: Aasta jooksul Tartu Ülikooli Arvutuskeskuses analüüsitud statistikaandmetike arv (dissertatsioonid, plaanilised ja lepingulised uurimistööd, diplomitööd) ulatus kaheksakümnendate aastate keskpaiku üle saja.

Suurarvutite osas oli Arvutuskeskus ülikoolis monopoolses seisundis, kuid teistes allüksustes (Majandusteaduskond, keemia osakond jm.) leidis mõningaid miniarvuteid, mis esialgu suutsid vaid üsna väikesi statistikaülesandeid lahendada.

5. *Personaalarvutid ja nende tarkvara.* Asjatundjate hinnangute kohaselt oli riistvara areng N. Liidus 1980. aastatel Läänest üle kümne aasta maha jäänud, kusjuures mahajäämus aina süvenes.

Ometi õnnestus Tartu ülikoolil tänu matemaatikute-arvutiteadlaste initsiatiivile ja mõningatele otsesidemetele Läänega ilma eriti suure hiliinemiseta siseneda personaalarvutite ajastusse. Juba 1980. aastate alguses hakati matemaatikute õppetöös kasutama personaalarvuteid Apple II ja ka edasi toimus areng kaunis kiiresti.

Personal arvutid jõudsid peagi Tartu Ülikooli teistegi teaduskondade laboritesse ja auditoriumidesse, ning sellega seoses suurenes üldine huvi arvutiteaduse, eriti aga andmetöötluse vastu.

Statistika tarkvara probleem kerkis üles uue teravusega. Oli selge, et nii kiire arengu perioodil ei ole mõeldavgi ise asuda standardse tarkvara loomisega, millega maailmas tegelevad suured, hästi varustatud kollektiivid. Paratamatult tuli esialgu hakata kasutama standardpakettide piraatkopieid.

Suhteliselt väikese leviku osaliseks sai esimesena kättesaadavaks osutunud STATPACK, mis oma ideoloogialt sarnanes veel suurarvuti pakettide-

ga. Seevastu järgmine, STATGRAPHICS, mille eeliseks oli personaalarvuti võimaluste (graafika, dialoog, inimsõbralikkus, läbimõeldud vaikimisi režiim) ja leplikkus arvutite tehniliste võimaluste suhtes, muutus väga populaarseks, hoolimata tema tagasihoidlikest võimalustest eriti suurte andmetike analüüsimisel. Paketi STATGRAPHICS kohta on ilmunud ka põhjalik eestikeelne tarbijajuhend.

Mõningast vilja on kandnud ka pingutused legaalse statistika tarkvara hankimiseks personaalarvutile.

Esimene katse — sõlmida leping SPSS firmaga (Moskvas paikneva Dialog'i vahendusel) jooksis siiski tühja. Paketist SPSS on Tartus praegu kasutusel vaid piraatkoopiad.

Originaalpaketist SURVO kinkis selle autor, Helsinki Ülikooli professor S. Mustonen, Tartu Ülikoolile kaks koopiat (1990 ja 1991). SURVO kohta on olemas niihästi originaalne tarbijajuhend (paar eksemplari) kui ka matemaatikatudengite lühike eestindus.

Peale läbirääkimisi ja konsultatsioone õnnestus leida vorm ka SAS-firmaga suhtlemiseks statistikute loodud eraaktiaseltsi "Resta" vahendusel. See selts sai SAS paketi litsentseeritud koopia ning täieliku dokumentatsiooni omanikuks. Eesti keeles on seni ilmunud kaks vihku SAS-juhendeid.

Saanud rahvusvahelisest statistikaajakirjandusest teada aktiivsest statistikaelust Eestis, legaliseeris ka BMDP firma (kirja teel) selle tarkvara kasutamise õiguse meil.

Firmapoolse kingitusena on kasutatav matemaatikapakett MATHEMATICA.

Kindlustusfirma "Swiss Re" kingitusena on õppetöös kasutatav kindlustusmatemaatika pakett.

Lisaks sellele on meil olemas mõningaid nn. professionaalseid pakette, sealhulgas TÜ kasvandiku Jaan Pelti kirjutatud aegridade analüüsi pakett ISDA.

Õppetöös kasutatakse valimiteooria paketti, mille I. Traat vahetult selle autorilt sai.

Kahjuks ei õnnestu meil siiski õppetöös praegu toime tulla ainult litsentseeritud, autorite poolt kingitud või aegumise tõttu praktiliselt litsentsivaba tarkvaraga (süüa hulka kuuluvad STATGRAPHICS-i versioonid 2.1–3.0, mida seni õppetöös kõige rohkem on rakendatud). Piraatkoopiade kujul kasutatakse pakette SYSTAT (tõsi, suhteliselt vana versiooni) ja SPSS, mis hetkel kasutamisiintensiivsuse poolest järgnevad STATGRAPHICS'ile.

Täiesti puudub Tartu Ülikoolis praegu statistika tarkvara suuremate arvutite — niihästi humanitaarabina saadud VAX'ile, samuti ka IBM suurarvutile. Standardse statistika tarkvara kasutamise eelduseks VAX'il on ka vajaliku operatsioonisüsteemi UNIX soetamine.

Hakata praegu IBM suurarvutil kasutama kaheksakümnendatel aastatel koduste vahenditega loodud paketti "Stella" on küll tehniliselt võimalik, kuid õpetamise ja teaduse arengu seisul kohalt täiesti mõeldamatu samm.

Arvestade aga suurarvutite ja tööjaamade jaoks loodud tarkvara kõrget hinda ning litsentseerimistingimuste rangust, ei ole nende jaoks vajaliku tarkvara hankimine statistikute enda jõududega mõeldav. Küll on võimalik juba loodud sidemeid kasutada soodsamate ostuvõimaluste leidmiseks.

Tartu statistikute pingutuste üks olulisi suundi on legaalse statistika tarkvara hankimine väljaõppe kaasaegsel tasemel hoidmise eesmärgil. Samuti peame oluliseks tarkvara eestikeelse juhendmaterjali koostamist, mis peaks aitama neid tarbijaid, kes iseseisvalt lahendavad personaalarvutite juures statistikameetodite abil oma andmeanalüüsi ülesandeid.

STATGRAPHICS JA TEMA KASUTAMINE ÕPETÖÖS

Ene Käärik, Mare Vähi

Tartu Ülikool, Matemaatilise Statistika Instituut

Esimene STATGRAPHICS'i versioon 1.0 valmis 1985. a. arvutile IBM XT. Järgnesid versioonid 2.1 — 1986. a., 2.6 — 1987. a. ja 3.0 — 1988. a. STATGRAPHICS (SG) on erinevalt mitmest teisest tuntud statistilise andmetöötuse süsteemist loodud otse personaalarvutile. Kõigil versioonidel on ühesugune päis ja ühesugune põhimenüü. Madalama tasemega versioonis sisestatud andmed sobivad kõrgema taseme versiooni (mitte vastupidil). SG edasiarendamisel on muudetud eeskätt töötamine süsteemiga käepärasemaks: on kasutusele võetud funktsionaalklahvid ja lisaaknad, saab analüüsida mitut andmestikku korraga, tellimuspaneelid on muudetud standardseks jms.

Süsteem SG ei ole eriti ressursinõudlik, ta töötab rahuldavalt (siiski vaid rahuldavalt) ka tagasihoidliku mälumahu ja kiirusega arvutil. Kohustuslikud eeldused süsteemi keskkonnale on:

- 1) IBM tüüpi või sellega ühilduv personaalarvuti (XT, AT);
- 2) operatiivmälu vähemalt 512 KB;
- 3) standardne klaviatuur;
- 4) operatsioonisüsteemi DOS versioon 2.0 või hilisem.

Süsteem tugineb oluliselt kasutaja ja arvuti pideva vastastikuse suhtlemise võimalusele. See teeb süsteemi kasutajasõbralikuks. Teda on võimalik kasutada ka üsna väheste eelteadmistega nii arvutite kui ka statistika alal. Süsteemil on väga head graafilised võimalused, millele viitab ka tema nimi STATGRAPHICS. Praktiliselt iga analüüsi saab illustreerida graafikuga, kusjuures väljundid on lihtsalt redigeeritavad, mis teeb nende kasutamise aruannetes ja teaduslikes artiklites eriti mugavaks. Arvestades süsteemi käsitlemise lihtsust ja tema vähenõudlikkust ressursside suhtes, on süsteem igati sobiv kasutamiseks õppetöös matemaatilise statistika või andmeanalüüsi kursuste praktilises rakendamises. Süsteem on kergesti kasutatav ja paigaldatav arvutiklassides. Üliõpilased omandavad SG süsteemi käsitlemise üsna lühikese aja jooksul, ei teki probleeme arvutiga, sest süsteem ei nõua praktiliselt eelteadmisi arvuti käsitluses, suheldes kasutajaga küllalt lihtsate ja arusaadavate dialoogide abil.

Süsteemi SG käivitamine toimub võtmesõna STATGRAF trükkimisega. Süsteemi käivitamine ilma kõvakettata arvutil võtab aega kuni 80 sekundit. Süsteem on tööks valmis, kui ekraanile ilmub põhimenüü:

STATGRAPHICS Statistical Graphics System

DATA MANAGEMENT AND SYST. UTIL.

- A. Data Management
- B. System Environment
- C. Report Wr. and Graphics Rep
- D. Plotter Interface

PLOTTING AND DESCRIPTIVE STAT.

- E. Plotting Functions
- F. Descriptive Methods
- G. Estimation and Testing
- H. Distribution Functions
- I. Exploratory Data Analysis

ANOVA AND REGRESSION ANALYSIS

- J. Analysis of Variance
- K. Regression Analysis

TIME SERIES PROCEDURES

- L. Forecasting
- M. Quality Control
- N. Smoothing
- O. Time Series Anal.

ADVANCED PROCEDURES

- P. Categ. Data Anal.
 - Q. Multivariate Methods
 - R. Nonparametric Methods
 - S. Sampling
 - T. Experimental Design
- ### MATHEM. AND USER PROC.
- U. Mathematical Funct.
 - V. Supplementary Oper.

SG süsteem on üles ehitatud hierarhiliste menüüde abil. Menüüs liikumine toimub nooleklahvidega ja valik fikseeritakse ENTER klahviga. Menüü on 2-astmeline, igale põhimenüü punktile vastab oma alammenüü. Alammenüü igale jaotusele vastab protseduur, mis on kas

- andmesisestusprotseduur (A),
- juhtimisprotseduur (B, C, D, V),
- analüüsiprotseduur (kõik ülejäänud).

Keskne koht süsteemis kuulub statistilise analüüsi protseduuridele. Siin on esindatud väga rikkalik valik matemaatilise statistika meetodeid. Põhimenüüs on analüüsi meetodid jaotatud nelja alarühma:

- *Plotting and Descriptive Statistics* (trükkimine ja kirjeldavad meetodid),
- *Anova and Regression Analysis* (dispersioon- ja regressioonanalüüs),
- *Time Series Procedures* (aegridade analüüs),
- *Advanced Procedures* (nn. edasiarenenud meetodid).

Esimene alarühm hõlmab põhiliselt võimalusi üksiktunnuse analüüsimiseks, mille tulemuseks on mitmesugused tabelid ja graafikud. Üksiktunnuse analüüsi meetodid sisaldavad tunnuse empiirilise jaotuse kirjeldamise, jaotusparameetrite ja mitmesuguste karakteristikute hindamise (punkt hinnangu mõttes), jaotusparameetrite vahemikhindamise ning statistiliste hüpoteeside kontrollimise jaotusparameetrite kohta, empiirilise jaotuse võrdlemise teoreetilisega ja statistiliste hüpoteeside kontrolli jaotuste kohta.

Kahe järgmise alarühma nimetused ütlevad juba ise, milliseid meetodeid nad sisaldavad: teises alajaotuses on dispersioon- ja regressioonanalüüs ja kolmandas aegridade analüüs.

Neljanda alarühma moodustavad kvalitatiivsete andmete analüüsi, mitme-mõõtmelise statistika ja mitteparameetrilise statistika meetodid. Väga suur

meetodite hulk on koondatud mitmemõõtmeliste meetodite alla. Siin on kovariatsioon- ja korrelatsioonanalüüs, faktoranalüüs, klasteranalüüs ja diskriminantanalüüs.

Iga protseduur algab dialoogi vormis tellimuse esitamisega. Tellimus koosneb kohustuslikust osast, kuhu kuuluvad analüüsitavaate tunnuste nimed (identifikaatorid) ja koodid, ning täiendavast informatsioonist, mille kasutamine muudab lahenduse efektiivsemaks. Ülesannete juures kasutatavate parameetrite väärtustamiseks pakub SG ise teatavaid "vaikimisi" väärtusi, mis enamasti osutuvad sobivaiks. Kui tellimus ei vasta kõigile vajalikele nõuetele, ei võta SG süsteem seda vastu, vaid väljastab veateate ja annab näpunäiteid selle parandamiseks. Vea korral antakse märku ka helisignaaliga. Tüüpilisemad vead võivad olla järgmised:

- tunnuse nimes viga (näiteks unustatakse, et nimi oli suurte tähtedega; tunnuse nimes eristab süsteem suur- ja väiketähti),
- tunnuse tüüp vale (näiteks ei luba antud protseduur sümboltunnust),
- tunnuste pikkused pole võrdsed.

SG süsteemi analüüsi tulemused esitatakse kas tabelina või graafikuna ekraanil. Osa tulemusi väljastatakse automaatselt, osa kujutab endast nn. lisainfot, mida saadakse tavaliselt lisainfo menüüst. Ekraanile väljastatud tulemusi (nii tabeleid kui graafikuid) on võimalik redigeerida: kirjutada juurde lisainfot, muuta graafikutel nii suurst, värve, pindasid, punktide kujutusi jne. Ekraanipildi modifitseerimiseks on SG süsteemis väga palju erinevaid võimalusi. Ekraanile väljastatud tulemused võib salvestada faili või vahetult välja trükkida.

Kui tellimus on täidetud täies ulatuses on võimalik sama protseduuri uuesti tellida või lõpetada töö sellega. Kordustellimuse korral on soovitatav pööruda allmenüüsse ja sealst protseduur uuesti käivitada.

SG süsteemiga töötamise hõlbustamiseks on ekraanil 3 alumist rida reserveeritud süsteemi hetkeseisu kirjeldamiseks nn. seisundirida. Esimesel neist kolmest antakse juhtnööre, mida parajasti tuleb teha. Teisel real on funktionaalklahvide seletus. SG süsteemis on defineeritud arvuti klaviatuuri abil terve rida funktsioone, mille tähendus püsib üldiselt kõigis protseduurides ühesugune. Mõned olulisemad neist:

F1 — *Help* — abistav süsteem. Ekraanile ilmub abiinfo parajasti teoksil oleva protseduuri kohta.

F6 — *Go* — salvestamine ja käsu täitmine. See on SG süsteemi tähtsaim klahv. Kui tellimus on antud, käivitatakse protseduur F6 abil.

F7 — tunnuste loetelu aken, mis hõlbustab tunnuste valikut protseduuri jaoks.

F8 — üleminek käskude režiimi, mille abil saab pooleliolevast protseduurist ajutiselt lahkuda, täita muid operatsioone (näiteks kasutada arvutit taskuarvutina) ja jätkata pärast seda tööd pooleli jäänud kohast.

F10 — *Esc* — *Quit* — menüü hierarhias samm tagasi. See on ka ühtlasi töö lõpetamise klahv: töö lõpetamiseks vajutatakse korduvalt klahvile, kuni

ilmub küsimus "Kas tahate lõpetada? N/Y". Lõpetamine toimub ainult vastusega Y, mistahes teise vastuse loeb süsteem eitavaks ja ekraanile ilmub tagasi põhimenüü.

Hetkeseisu kirjelduse alumisel (kolmandal) real on mitmesugune lisainfo süsteemi tegevuse kohta (s.h. kuupäev, kellaaeg, versiooni nr.).

Andmestik SG süsteemis kujutab endast kindlal viisil korrastatud mõõtmistulemuste kogumit koos andmekirjeldusega. SG muutuja võib olla arvtunnus, mittearvuline tunnus või m -mõõtmeline maatriks. Muutuja koosneb kahest osast: väärtusest ja andmekirjeldusest. Andmekirjeldus sisaldab nime, tüüpi, järgu, laiuse ja pikkuse (objektide arv). SG andmestik (fail) on ühest või paljudest SG muutujatest koosnev statistiliste andmete kogum, mida iseloomustab tema nimi ja temasse kuuluvate muutujate loetelu. Andmestiku sisestamine toimub alajaotuse A. *Data Management* (andmekäsitus) allmenüüs 2. *File Operations*. Üldiselt ei ole SG andmestikku kuuluvate muutujate valikul mingeid piiranguid. Ühes andmestikus võivad olla eri tüüpi, erineva pikkusega ning erinevatel objektidel mõõdetud andmed. Tuleb aga silmas pidada, et ühes protseduuris võib kasutada ainult võrdse pikkusega muutujaid. See nõue on täidetud, kui andmestikuks on objekt—tunnus maatriks. SG protseduurid ei nõua, et neis kasutata- vad tunnused oleksid ühest andmestikust. Korraga saab kasutada samas kataloogis paiknevaid andmestikke. Objektide ja tunnuste arv ühes andmestikus on piiratud SG andmestikele eraldatava mälu mahuga (64 K). Kuna üheaegselt saab töödelda erinevaid andmestikke, ei ole see nõue kitsendav. SG andmestik on teatavas mõttes ühilduv ka teiste süsteemidega s.t. ta ei ole küll vahetult teiste süsteemide poolt kasutatav, aga on loodud vahendid andmestike importimiseks ja eksportimiseks. Üle saab kanda ASCII, Lotuse, d'Base III ja DIF faile, mille kaudu on võimalik andmete vahetus väga mitmete süsteemidega.

TERRITORIAALSETE ANDMETE ANALÜÜSI PAKETID JA STATISTIKA OSA NENDES

Jüri Roosaare

Tartu Ülikool, Geograafia Instituut

Sissejuhatus

Kui vaatlusaluste objektide või nähtuste kohta käivate andmete oluliseks atribuudiks on *asend* (paiknemine mingis koordinaatide süsteemis), siis on geograafias tavaks rääkida territoriaalsetest andmetest, ehkki vaatluse all ei tarvitse olla andmed just mingi territooriumi kohta. Tavaliselt on tegemist vähemalt ühega kolmest juhust:

- 1) vajalikke algandmeid soovitakse hankida mitmesugustelt (geograafiliselt) kaartidelt,
- 2) analüüsil tuleb arvestada objektide või tunnuste topoloogilisi suhteid,
- 3) parema näitlikustamise huvides tahetakse tulemused esitada kartograafiliselt.

Et seda tüüpi ülesannete jaoks vajaliku arvutitarkvara kujunemisel on kesksel kohal olnud arvutigraafika probleemid, siis on territoriaalsete andmete analüüsi programmid algselt olnud seotud põhiliselt kas kompuuterkartograafia, kosmosefotode töötuse või geograafiliste informatsioonisüsteemidega. Kartogrammid kui (3) lihtsaim viis on leidnud tee ka mitmesse statistikapaketti.

Käesolev artikkel ei püüa vastata küsimusele, kas andmete territoriaalsus lisab ka teooria tasandil midagi spetsiifilist statistikale — ehkki see võiks huvi pakkuda —, vaid piirdub kasutatava lühikirjeldusega põhiliselt ühe paketi — IDRISI¹ — näitel.

Andmemudel

Tavapärane andmeanalüüs lähtub m objekti ja n tunnuse $m \times n$ maatriksist $\|a_{ij}\|$. Territoriaalsete andmete puhul on enim kasutusel:

— vektormudel, mis võimaldab areaalide, joonte, sõlmede ja punktide kui horopleetide kaardi² elementide topoloogiliste seoste täielikku kirjeldamist ja sobib hästi andmebaaside loomiseks, ning

— rastermudel, mis katab territooriumi korrapäraste elementide võrgustikuga ja sobib hästi andmetöötuseks (Roosaare, 1992).

Piirdume siinkohal viimase vaatlemisega.

Ristkülikukujulist ala, mille nurkade koordinaadid on (x_{min}, y_{min}) , (x_{max}, y_{min}) , (x_{max}, y_{max}) , (x_{min}, y_{max}) , kirjeldaks rastermudeli puhul kolme-mõõtmeline andmemassiiv

$$A(k, l, n),$$

kus rastrielemendiks (territoriaalse informatsiooni kandjaks) oleks riskülik küljepikkustega $c = (x_{max} - x_{min})/k$ ja $d = (y_{max} - y_{min})/l$; kokku oleks $m = k \times l$ objekti.

Tavaliselt:

— esitatakse andmemassiiv n erineva maatriksina $\|a_{uv}\|$, $u = 1, \dots, k$ ja $v = 1, \dots, l$; iga j -ndas maatriks (neid nimetatakse *kaardikihtideks*) kirjeldab j -nda tunnuse territoriaalset paiknemist;

— levinumaks andmetöötlusliigiks on nn. kaardialgebra, mille puhul kaardikihtide vahel defineeritakse algebralised või loogilised seosed ning uus (huvipakkuv) kaardikiht esitatakse kui lähtekihtide funktsioon; sisuliselt rakendatakse seda funktsiooni vastavatele rastrielementidele $k \times l$ korda;

— c ja d valitakse nii väikesed, et antud ülesande raamides oleksid kõik rastrielemendid vaadeldavad homogeensetena; seetõttu võivad k ja l olla väga suured, n on tavaliselt väike;

— töötuse käigus kasutatakse täisarvulisi koordinaate (u, v) .

Näiteks IDRISI puhul $k < 32\ 000$, $l < 32\ 000$ ja enamike operatsioonide puhul $n = 2$, s.t. kaardikihte töödeldakse paarikaupa. Kaardialgebral on kasutada 14 erinevat operaatorit: ühe kihi skalaar teisendused (konstandiga liitmine, lahutamine, korrutamine, jagamine, astendamine), samad tehted kahe kihi vahel (on kasutatavad ka loogilisteks teheteks) ja funktsioonid $b = \min(a_1, a_2)$, $b = \max(a_1, a_2)$, $b = (a_1 - a_2)/(a_1 + a_2)$, $b = a_1$ kui $a_2 > 0$, muidu $b = 0$. Igale kaardikihile vastab failipaar, millest üks (laiendiga .IMG) sisaldab üksnes (baidilise, täisarvulise või reaalarvulise) maatriksi a_{uv} (ASCII, binaar- või pakitud formaadis) ja teine (laiendiga .DOC) nii vajalikud muutujad (x_{min} , x_{max} , y_{min} , y_{max} , k , l , a_{min} , a_{max}), andmed formaadi, täpsuse, mõõtühikute jms. kohta kui ka tiitlid ja legendi.

Kuigi IDRISI on orienteeritud eelkõige rasterpiltide töötuseks, võib vajadusel ilmselt suurema vaevata eeltöödeldud andmemudeli kujule teisendada ka alamhulki muudest andmekogumitest ning rakendada seejärel seda tarkvara muude ülesannete lahendamiseks.

Geostatistika

Nagu paljudes teistes valdkondades, nii on ka geograafias kujunenud välja oma enamkasutatavate statistiliste võtete kogum ning nendega tegelevate professionaalide ring(id), kelle organiseeritud nõupidamised ja eelistatud ajakirjad³ seda kogumit kujundavad. Territoriaalsus pole suutnud siiski olla piisavalt integreeriv kategooria geograafia paljudele eriharudele. Nii on viimaste aastakümnete jooksul statistiliste rakendustega tegeldud alul matemaatiliste meetodite ja hiljem geograafiliste informatsioonisüsteemide lipu all. Mõistmise laiuusest võib ülevaate saada Rahvusvahelise Geograafia-uniooni Matemaatiliste meetodite töögrupi korraldatud ekspertide küsitluse

järgi (tabel 1), milles segaduste vältimiseks on teemad esitatud originaalkeeles, sest kindel eestikeelne terminoloogia pole veel välja kujunenud.

Meetodeid kokku võtva kirjanduse seast võiks viidata N. Cressie (1991) raamatule.

Üheks statistikaga piirnevaks küsimusteringiks territoriaalsete andmete analüüsil on ruumilise interpolatsiooni meetodid (RIPM), millel näitena peatume veidi lähemalt.

Sageli me eeldame, et omadused muutuvad territooriumil järk-järguliselt ja meil on olemas andmed teatud vaatluspunktide kohta. Tekib vajadus anda mingi hinnang uuritava nähtuse väärtusele nendes punktides, kus vaatlusi pole tehtud, kasutades vaatluspunktides saadud väärtusi. RIPM aluseks on intuiitivne arusaam ruumilisest autokorrelatsioonist: punktid, mis asuvad üksteisele lähedamal, on samasamad. Niisugune seaduspärasus ei pruugi kehtida kaugeltki igasuguste kauguste puhul, seepärast on kesksiks küsimuseks, kuipalju punkte ja missuguste kaaludega interpoleerimisel arvesse võtta. Teiseks eeldatakse, et omaduste ruumilisel muutumisel on olemas mingisugune seaduspära, olgu siis determineeritud või stohhastiline.

Nende omaduste konkretiseerimist nimetatakse interpolatsioonimudeli valikuks, millest tulenevalt on võimalik kasutada üht või teist RIPM — subjektiivsetest kuni enam või vähem rangete (ja erineva keerukusega) matemaatiliste meetoditeni.

Küllalt sageli lähtutakse punkti x väärtuse Z selgitamisel järgmisest mudelist:

$$Z(x) = m(x) + \varepsilon'(x) + \varepsilon'', \quad (1)$$

kus

$m(x)$ — determineeritud, seaduspärane komponent, mis kirjeldab Z muutumise "struktuurset" osa või trendi;

$\varepsilon'(x)$ — stohhastiline, lokaalselt muutuv jääkkomponent [$Z(x) - m(x)$], mille varieeruvus oleneb kaugusest ja

ε'' — kaugustest sõltumatu, normaaljaotusega (keskmisega $E(x)=0$; dispersiooniga σ^2) jääkmüra.

P. Burrough (1987) järgi võib erinevad RIPM klassifitseerida järgmiselt:

• diskreetsete muutuste piiritlemise meetodid [*discrete techniques*]:

— üksuste silma järgi piiritlemine [*'Eyeball' method*],

— piirjoonte leidmise algoritm [*edge-finding algorithm*],

— lähima naabri [*proximal*] meetodid — Thiesseni (Voronoi, Dirichlet) polügoonid,

• sujuvaid muutusi eeldavad meetodid:

— globaalsed meetodid [*global techniques*]

— trendpinna analüüs [*trend surface analysis*],

Tabel 1

**EKSPERTIDE HINNANGUD TERRITORIAALSE STATISTIKA ERINEVATE
MEETODITE KASULIKKUSEST JA PERSPEKTIIVSUSEST**

MEETOD	1984*	N**
Theory of Spatial Statistics	3.52***	84
Geocoding and Geosampling	3.30	73
Classification	2.60	84
Interactive Cartography	3.90	67
Computer Mapping	3.64	81
Trend Surface Analysis	2.37	82
Spatial Autocorrelation	3.17	90
Spatial Equality Measures	2.84	70
Spatial Time Series	3.49	86
Directional Data Analysis	2.92	53
Topology	2.70	71
Shape Analysis	2.22	69
Spatial Choice Models	3.78	80
Spatial Interaction Models	3.38	89
Simulation Models	3.87	82
Computational Process Models	3.82	51
Point Pattern Analysis	2.44	79
Categorical Data Analysis	3.59	82
Fractals	3.39	64
Geo. Information Systems	4.02	86
Spatial Forecasting Models	3.74	84

* 91 vastajat 17 riigist; ** meetodit tundvaid vastajaid; *** meetodit tundvate ekspertide kasulikkuse hinnangute kaalutud keskmine (1 — kõige vähem, ..., 5 — kõige rohkem kasulik). (Gaile, 1990 andmete alusel.)

- Fourier read [*Fourier series*];
- lokaalsed interpolaatorid [*local fitting techniques*]
- B-splainid [*B-splines*],
- libisevad (kaalutud) keskmised [*weighted moving averages*],
- optimaalse interpolatsiooni meetodid e. kraiging [*kriging*].

Kraiging (nimetus tuleneb meetodi ühe autori, Lõuna-Aafrika kaevandusinseneri D. G. Krige nimest) lähtub eelpooltoodud kauguse mudelist (1) ja võimaldab määrata interpoleerimisel kasutatavate punktide kaalud nii, et interpoleeritud väärtus $Z(x_0)$ oleks nihketa hinnang ja tema dispersiooni hinnang σ^2 oleks väikseim vaatluspunktide ja kaalude lineaarkombinatsioon. Kauguse mõju (ja "akna" suuruse) määramiseks leitakse empiiriliste andmete (vaatluspunktide väärtuste) alusel semivariogramm $\gamma(h)$: $Z(x)$ ja $Z(x+h)$ ruutkeskmise hälbe sõltuvus kaugusest h . Lihtsaim kõvera $\gamma(h)$ kirjeldamiseks on lineaarne mudel:

$$\gamma(h) = \begin{cases} c_0 + bh & \text{for } 0 < h < a \\ c_0 + c_1 & \text{for } h > a, \end{cases}$$

kus:

c_0 — kirjeldab kaugusest sõltumatut müra [*'nugget' variance*] ϵ^2 ;

c_1 — kirjeldab graafiku platoo-osa, kus alates väärtusest $h=a$ variatsioon ei muutu ja seetõttu a on sobiv akna suuruseks.

Sõltuvalt $Z(x_0)$ arvutusvalemitest (need määratakse võrrandisüsteemi lahendamiseks) räägitakse punkt-kraigingust [*point or simple kriging*], blokk-kraigingust [*block kriging*], mis silub "väljaulatavaid" punkte paremini, ja universaalsest kraigingust [*universal kriging*], mis võimaldab arvestada lookaalseid trende.

Kraiging on hea tehnika, mis tugineb statistika teooriale ja nõuab olulisi arvutiresse, eriti kaartide (pindade) puhul. Ta võimaldab määrata vea suuruse erinevate akna kujude ja suuruste korral. Lisaks arvutusmahukusele võivad meetodi võimalusi piirama hakata ka vaatluspunktide vähesus ning võimalikest ruumiomadustest või nende muutustest tingitud andmete mittestatsionaarsus (eri suundade või eri ruumiosade semivariogrammid tulevad erinevad).

Kraiging on keskne tehnika ka spetsiaalses geostatistika pakettis GEOEAS4.

IDRISI

1987. a. käivitunud *IDRISI Project*⁵ on kasumivaba organisatsioon, mis seadis ülesandeks luua odav professionaalse tasemega geograafilise informatsiooni analüüsisüsteem. Praeguseks on Clarki Ülikooli juures olev paarikümnest inimesest koosnev IDRISI meeskond tihedates koostööside-metes selliste organisatsioonidega nagu UNITAR⁶ ja UNEP/GRID⁷, loodud tarkvarapakett on kasutuses enam kui sajast riigist, olles akadeemilise

kallakuga rastersüsteemidest kõige laiemal levikuga. Versioon 4.0 koosneb 120 iseseisvast EXE-moodulist (kokku 4 Mb kõvakettal), mis töötavad MSDOS keskkonnas, nõuavad vähemalt 512 Kb vaba mälu ja soovitatavalt VGA monitori. Reaalseks tööks on optimaalne siiski 4Mb RAM, 386/387 protsessor ja 8514/A monitor.

IDRISI sellel versioonil on kaasas ka küllaltki primitiivne menüüdesüsteem ja võimalik on moodulitega töötada mitte eriti tarbijasõbralikus dialoogrežiimis. Koos paketi lisatud näidisandmestiku (6 Mb) ja 16 harjutusülesandega (IDRISI, 1992a) loob see hea võimaluse paketi kasutamiseks mitmesuguses õppetöös, mida kogu maailmas ka laialdaselt tehakse.

Professionaalsel andmetöötlejal on aga põhiliseks kasutusviisiks IDRISI moodulite sidumine oma (suvalises keeles) rakendusprogrammidega, sest andmeformaadid, paketi pisike keskkonnafail ja välisseadmete draiverid on kasutajale avatud ning enamik mooduleist töötab ka käsurea parameetritega.

IDRISI ise andmebaase ei toeta, kuid on kergesti ühendatav mõne d'Base-tüüpi andmeohjesüsteemiga, kasutades liidesena IDRISI nn. atribuutide väärtusfaile. Küllaltki laialdased on andmete impordi/eksporti võimalused. Mitmesugustel kartograafilise informatsiooni sisend/väljund võimalustel, mis IDRISI-l olemas on (kaasa arvatud raster—vektor ja vektor—raster teisendused), siinkohal ei peatuta.

Paketi kommertsversiooni hind oli 1993. a. kevadel \$ 560 (akadeemilised ja valitsusasutused \$ 280, statsionaarsed üliõpilased \$ 140). Lisaks saatekulud. Pakutakse ka paketi mitmesuguseid keskkonnakaitse-alaseid rakendusi koos vastavate õpperaamatutega (á \$ 75).

Saamaks ülevaadet paketi statistilistest võimalustest, on tabelis 2 esitatud selle valdkonnaga seostuvate moodulite lühikirjeldused (terminoloogiliste segaduste vältimiseks originaalkeeles). Nii lühikeset kirjeldusest piisab muidugi vaid üldise ülevaate saamiseks.

Kokkuvõttes võime nentida, et suur osa statistikaga seotud mooduleid IDRISI-s on orienteeritud eeskätt satelliitfotode töötuseks. Moodulid kajastavad arvestataval määral seda, missugused statistika valdkonnad territoriaalsel andmeanalüüsil kasutamist leiavad. Ülesannete spetsiifika näiteks võib olla peakomponentide meetod (moodul PCA), mis lubab ainult kuni 12 tunnust objektide väga suure hulga juures.

IDRISI VERSION 4.0 (1992B) MOODULID,
MIS ON SEOTUD STATISTIKAGA

EXE-moodul	Kirjeldus
AUTOCORR	first-lag autocorrelation (using Moran's I); Rook's Case & King's Case
CENTER	mean center (weighted or not) of a point distribution and standard radius
CLUSTER	unsupervised classification of a (color composite) image
CONFUSE	evaluation of confusion matrices in classification accuracy assessment
CROSSTAB	image crosstabulation and crosscorrelation
EXTRACT	extraction of a summary of data (min, max, mean...) found within defined regions
FILTER	convolves (strictly correlates) an image with a digital filter (6+user-def.)
HISTO	histograms of image file values; proportional and cumulative frequencies along with simple statistics
INTERCON	interpolation of a surface from a set of contours by linear interpolation
INTERPOL	interpolation of a surface from point data using either a weighted-distance or potential surface model
MAKESIG	creation of signatures from training sites (for supervised classification)
MAXLIKE	a maximum likelihood classifier
MINDIST	a minimum distance to mean classifier
PATTERN	evaluation of pattern measures of relative richness, diversity, dominance etc.

PCA	principal components analysis (standardized and unstandardized)
PCLASS	evaluation of probability that pixels exceed by a specified threshold
PIPED	a parallelepiped classifier
QUADRAT	density, variance and variance-mean ratio of quadrat cell counts
RANDOM	creation of random images according to rectilinear or normal distribution
REGRESS	regression analysis on images or attribute values files
RESAMPLE	determines the data values for a rectangular grid by interpolation (some methods) of the values in a different (and possibly warped) grid
SAMPLE	creation of systematic, random and stratified random point sampling schemes
SIGCOMP	a signature comparison utility
STRETCH	linear, linear with saturation and histogram equalization contrast stretches
TREND	calculation of the best fit linear, quadratic or cubic trend surface to a set of irregular cell control points by least-squares procedures

Viidatud kirjandus

- Burrough, P. A. (1987). Principles of Geographical Information Systems for Land Resources Assessment. — Oxford: Clarendon Press. — 193 p.
- Cressie, N. (1991). Statistics for Spatial Data. — J. Wiley & Sons. — 900 p.
- Gaile, L. G. (1990). Whither Spatial Statistics?// Professional Geographer, 42(1), p. 95–100.
- IDRISI. Version 4.0. (1992a). User's Guide. 178 p.
- IDRISI. Version 4.0. (1992b). Technical Reference. 213 p.
- Roosaare, J. (1992). Geograafilised informatsioonisüsteemid. Sissejuhatav loeng // Elva sügiskooli materjale, 26.–28. november 1991. — Tartu, lk. 5–24.

Märkused

- ¹ IDRISI Source Code © 1987–1992 J. Ronald Eastman; IDRISI Production © 1987–1992 Clark University Graduate School of Geography.
- ² Horopleetide (areaalide, levikualade) kaart vaatab kõiki punkte antud areaali sees ühesugustena ja muutusi nende piiridel diskreetsetena.
- ³ Siinkohal võiks nimetada Ohio State University Pressi poolt välja antavat *Geographical Analysis. An International Journal of Theoretical Geography*, ISSN 0016-7363.
- ⁴ GEOEAS — GEOstatistical Environmental Assessment Software on USA EPA Environmental Systems Laboratory, Las Vegas ja Stanfordini Ülikooli Applied Earth Sciences Department poolt 1988. a. loodud nn. PUBLIC DOMAIN tarkvara (seega vabalt kopeeritav). Meil kasutada olev v. 1.2.1 (1990–1991) nõuab 640 K RAM ja umbes 3 megabaiti kõval kettal.
- ⁵ Address: The IDRISI Project, Graduate School of Geography, Clark University, 950 Main Street, Worcester, Massachusetts, 01610-1477, U.S.A. FAX: (508) 793-8842. E-mail: idrisi@vax.clarku.edu.
- ⁶ UNITAR — United Nations Institute for Training and Research.
- ⁷ UNEP/GRID — United Nations Environment Programme Global Resource Information Database.

NOORPULLIDE KASVUKIIRUSE ARETUSVÄÄRTUSE HINDAMINE

Elli Pärna

Eesti Loomakasvatuse ja Veterinaaria Instituut

Pullikasvanduste pullikute omajõudluse hindamine ainult fenotüübiliste andmete alusel on põhjendatud aretustöö teatud etapini. Selline fenotüübiline tunnus, nagu massi-iive, mida kasutatakse kasvanduse pullikute omajõudluse hindamiseks, sisaldab aga tema väärtust mõjutavaid erinevaid süstemaatilisi faktoreid. Statistilises mõttes on "puhas" noorpulli fenotüübiline mõju õigemini jääkmõju, mis jääb järele pärast seda, kui nimetatud süstemaatilised mõjud on eemaldatud. See tähendab, et pullikute võrdlemine on korrektne alles pärast selliste mõjude (nn. fikseeritud mõjud) kõrvaldamist.

Töös uuriti vähimruutude meetodil Tori kasvanduse noorpullide kasvukiiruse fikseeritud mõjudena isa sünniaastat (grupeerituna 3 rühma vastavalt sünniaastale ajavahemikus 1969–1976 1. rühm, 1977–1981 2. rühm, 1982–1986 3. rühm) noorpulli sünniaastat (grupeerituna 8 rühma alates detsembrist 1982 kuni novembrini 1983 1. rühm, detsembrist 1983 kuni novembrini 1984 2. rühm jne.) aastaaega sünnimomendil (grupeerituna 2 rühma vastavalt detsembrist maini 1. rühm ning juunist novembrini 2. rühm), kasvandusse saabumise vanust (päevades) ning majandit. Kasvukiiruse fikseeritud mõjudena olid statistiliselt olulised pulliku isa sünniaasta, pulliku sünniaasta ja aastaaeg pulliku sünnimomendil ($p < 0.001$). Kasvukiirus ei olenenud vasika kasvandusse saabumise vanusest ega majandist, kust ta pärines, samuti kasvanduse-eelsest massi-iibest ($p > 0.05$). Andmed Tori noorpullide kasvu ja arengu kohta vanuses 6–12 kuud ajavahemikus 1984–1990 on esitatud tabelis 1. Juhul, kui vasikal puudus 6 ja/või 12 kuu kaal, prognoositi need regressioonanalüüsi teel olemasolevate andmete põhjal (determinatsioonikordaja $r^2 = 0.81$, $p < 0.001$).

Fikseeritud mõjudele alternatiivne tüüp on juhuslikud mõjud, mis tähendab seda, et mõjud, mis realiseeruvad vaadeldavas kogumis, on käsitletavad kui juhuslikud mõjud populatsiooni mõjude hulgas. Tori kasvanduse noorpullide isade mõju (juhuslik mõju) oli nii massile kui ka kasvukiirusele statistiliselt oluline ($p < 0.001$).

Kasvukiiruse aretusväärtuse ehk T -indeksi arvutamisel kasutati Rootsis kasutusel olnud meetodikat, mis põhines eakaaslastega võrdlemisel (T. Henningsson, 1985) ning alljärgnevat valemit:

$$T = h^2(R - 100) + 100,$$

kus h^2 on kasvukiiruse päritavuskoeffitsient ja R on suhteline kasvukiirus, s.t. noorpulli kasvukiiruse protsent eakaaslaste suhtes.

Tabel 1

TORI KASVANDUSE NOORPULLIDE ISELOOMUSTUS

Tunnus	n	Keskmine	Standardhävle	Standardviga	Min	Max
Algvanus, päevades	379	154.3	64.83	3.33	29	327
Lõppvanus, päevades	342	513.9	71.0	3.84	257	765
Mass 6 kuud, kg	297	209.5	27.4	1.6	140	290
Mass 12 kuud, kg	375	413.4	38.6	2.0	305	526
Massi-ivve, g/ööp.-s	294	1140	163	0.00	520	1580

T -indeksile väärtusele ei avalda mõju pulliku sünniaasta, aastaeg sünnimomendil, kasvanduse eelse perioodi ööpäevane massi-ivve, majand, kust vasikas pärines ning kasvandusse saabumise vanus ($p > 0.05$). Ainsa mõjuna ei õnnestunud kõrvaldada pulliku isa sünniaasta mõju kasvukiiruse aretusväärtusele ($p < 0.05$). Tori kasvanduse pullikute T -indeksi väärtuse keskmine oli 99.96, dispersioon 34.54, standardhävle 5.9, standardviga 0.3, minimaalväärtus 75.6 ning maksimaalväärtus 116.5. Seega võime öelda, et kasvukiiruse aretusväärtuse väljendamine T -indeksina annab hea tulemuse parima kasvukiiruse aretusväärtusega pullikute väljaselgitamisel, sest tema väärtus ei olene keskkonna tingimustest.

Eakaaslastega võrdlemise meetodil aretusväärtuse leidmise üks oluline puudus aga säilib, nimelt ei võta nimetatud meetod arvesse aastate jooksul tekkivaid geneetilisi erinevusi, samuti informatsiooni poolõdede kohta. Seetõttu jätkati käesolevat uurimist aretusväärtuse hindamise meetodi täiustamiseks, püüdes suhteliselt vähest informatsiooni ära kasutada võimalikult efektiivselt.

Aretuskarja geneetilise parandamise peamiseks vahendiks on seleksioon. Seleksioon tähendab seda, et loomad, kellel on tulevaste põlvkondade mõjutamisel soovitud suunas parimad väljavaated, on reproduktsioonil eelistatud. Seleksiooni jaoks tuleb soovitavad või paremad loomad potentsiaalsete aretusloomade hulgast leida. Varasematel aegadel tehti seda küllaltki subjektiivselt, s.t. valituks osutusid loomad, kellel oli hea omajõudlus või kelle kohta oli teada, et nad parandavad hea jõudluse oma järglastele. Tegelikult on selleks, mida me vajame, potentsiaalsete kandidaatide aretusväärtuse prognoosimine. Teisisõnu tähendab see, et tuleb hinnata potentsiaalsete kandjdaatide aditiivne geneetiline väärtus või vähemalt olema võimeline järjestama neid vastavalt kriteeriumile, mis on otseselt proportsionaalne aretusväärtusega (Ö. Danell, 1990).

Antud töö järgmiseks etapiks oli kasutada kasvanduse pullide kasvukiiruse aretusväärtuse hindamiseks mudelid, kus nii fikseeritud kui ka juhuslikke mõjusid vaadeldakse samaaegselt. Selliseid mudelid nimetatakse segamudeliteks, ning nad on praktilises aretustöös laialdaselt kasutusel. Isade mõju (juhuslik mõju) uurimise eesmärgiks on prognoosida pullide mõju järglastele. Selliseid juhuslike mõjude hinnanguid nimetatakse parimaks lineaarseks nihketa prognoosiks (BLUP, *Best Linear Unbiased Predictor*). "Parim" ja "lineaarne" tähendavad seda, et hindajate veadispersioonid on minimaalsed ning nad on vaatluste lineaarfunktsioonid. Veelgi enam, prognoosijad on kasutatud meetodiga garanteeritud olema nihketa, s.t ilma süstemaatiliste vigadeta.

Pulli genotüübi aditiivne geneetiline mõju pole indiviidil endal vaadeldav. Genotüübi aditiivne mõju on genotüüpi lülitatud geenide keskmine mõju pärast seda, kui nad on transformeerunud järgmisesse generatsiooni ja kombineerunud teiste geenidega. See on täpselt see, mida me tahame prognoosida, eeldades et pull selekteeritakse ja säilitatakse selleks, et ta saaks oma geenid üle anda järgmisele põlvkonnale. Tõeline aretusväärtus on indiviidi tegelik, kuid mittemõõdetav aditiivne geneetiline mõju. Tõelise aretusväärtuse prognoosi nimetatakse hinnatud aretusväärtuseks ehk indeksiks, tõelise aretusväärtuse hinnangu loomiseks on kaalutud mitut informatsiooniallikat. Et aretusväärtuse leidmisel on tähtis võimalikult täpselt teada tunnuse päritavuskoefitsienti, kasutatakse kasvukiiruse päritavuskoefitsiendi leidmiseks erinevaid meetodeid, püüdes leida neist objektivseima.

Tori kasvanduse pullikute kasvukiiruse aretusväärtuse prognoosimisel kasutati ainult dispersioone ja kovariatsioone kui teadaolevaid ning hinnati geneetilisi mõjusid aretusväärtusele samaaegselt teiste mõjudega ühtses statistilise hindamise protseduuris. Tavapärast tehakse seda Hendersoni segamudeli võrrandite lahendamise teel. Protseuur koosneb kolmest astmest, millest käesolevaks momendiks käib töö teise astme lahendamiseks:

- (1) Kirjeldati andmeid, mida kasutati statistilises mudelis koos andmeid mõjutavate süstemaatiliste faktorite fikseeritud mõjudega ning vaadeldavate pullikute suhtes juhuslike mõjudega (s.t. isade mõjudega).
- (2) Järgnevalt tekitatakse ning lahendatakse segamudel samaaegselt fikseeritud ja juhusliku mõju suhtes.
- (3) Seejärel arvutatakse lahendi alusel kasvukiiruse aretusväärtuse prognoos iga pulliku jaoks, mis võtab arvesse ka geneetilisi mõjusid.

Kasutatud kirjandus

- Henningsson, T. Performance testing for Beef Production Traits in Swedish Dual Purpose and Beef Cattle. Rapport 64, 1985.
- Danell, Ö. Basic quantitative genetic theory. Swedish University of Agricultural Sciences, 1990, Uppsala.

NSDstat JA SELLE KASUTAMISE VÕIMALUSED

Urmas Oja

Tartu Ülikool, Perelabor

Antud kirjutise eesmärgiks on tutvustada pisut lähemalt sellist statistikapaketti, mis käesoleval ajal on Eestis üpris tundmatu. Ma arvan ja loodan, et peagi võidab ta endale poolehoidjaid. Jutt tuleb programmist NSDstat. Pakett on väljatöötatud Norra Sotsiaalteaduste Andmetalituses (*Norsk satsunnsvitenskapelig datatjeneste*) ning sealt siis ka programmi nimi.

Idee

NSDstat loodi algselt Norra koolisüsteemi jaoks. Seega on programm eelkõige lihtsa andmeanalüüsi ja andmeesituse tarvis. Kuid siiski võib lisada, et peale hariduse kasutatakse NSDstat-i laialdaselt ka muudes eluvaldkondades, kus on vajadus andmete statistilise analüüsi ja nende esituse järele. Näiteks võiks tuua meditsiini, omavalitsusorganid, aga samuti ka mõned teadusliikud uurimisasutused.

NSDstat-i kasutamise eesmärk on tunduvalt avaram kui teistel minule teadaolevatel statistikapakettidel. Tavapärasele oma andmete analüüsile ja esitusele on lisandunud võimalus analüüsida NSD-s varem ettevalmistatud andmebaase. Loomulikult eeldab võõra andmebaasi analüüs programmit ka mingite uute võimaluste olemasolu. Ja need on NSDstat-il ka olemas. Kui enamikes statistika programmides on sisestatava dokumentatsiooni hulk piiratud lühikeste muutujate ja väärtuste nimedega, siis NSDstat annab võimaluse määrata andmebaasi igale muutujale tervikliku vabateksti dokumendi.

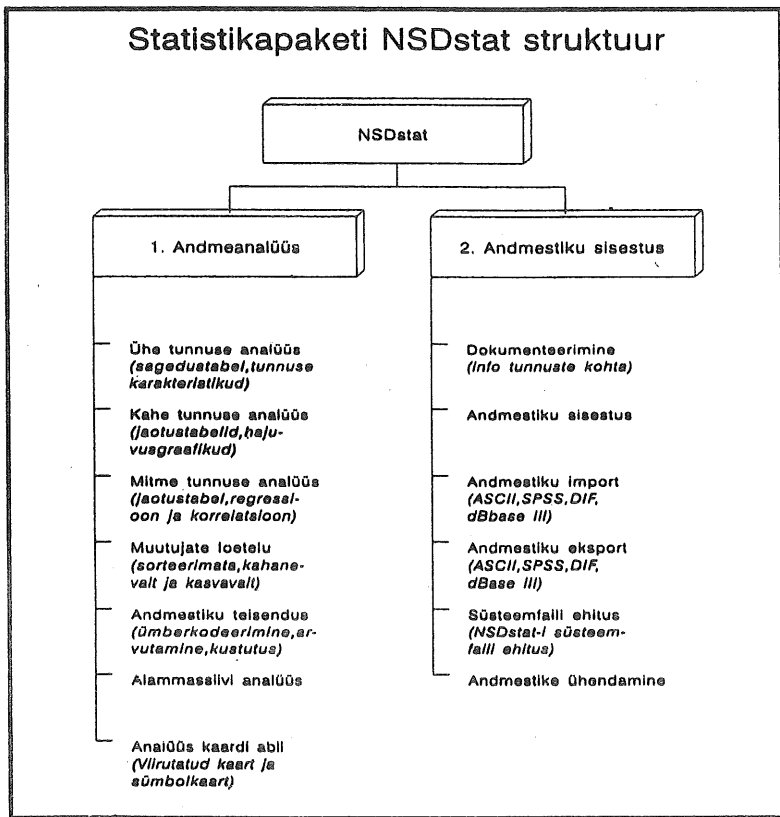
NSDstat-i põhiliseks ideeks on pakkuda keskkooliõpilastele ja üliõpilastele võimalust harjutada ennast sotsiaalteadlase rolliga analüüsides juba spetsiaalselt õppeotstarbeks kokkupandud korralikku andmestikku (on valminud paketid "Noorus ja Euroopa", "Demograafia ja tootmisstruktuur" ning "Poliitiline geograafia"). Loomulikult on programmis säilinud ka andmete sisestamise ja lisamise võimalus, mida eriti innukalt kasutavad just keskkooliõpilased.

Nõuded riistvarale

Programmi NSDstat suurus koos kõigi pakutavate ekraanidraiveritega on alla 950 kilobaidi. Paketi miinimumnõuded riistvarale on vägagi tagasihoidlikud: IBM-iga ühilduv personalarvuti, vähemalt 256 K RAM, kaks 360 K kettadraiverit ja CGA, EGA, VGA või Hercules graafikakaart.

NSDstat-i struktuur ja võimalused

NSDstat koosneb kahest moodulist: andmestiku sisestusest ja andmestiku analüüsist. Lähema ülevaate saamiseks programmi võimalustest vaata ka joonist 1.



Joonis 1. Programmi NSDstat struktuur.

NSDstat-i sisestusmoodul. Programmi andmestiku ehitamiseks on vaja kas andmed sisestada endal või siis importida ASCII, d'Base III, SPSS või DIF-failina (andmestiku ekspordi võimalused NSDstat-ist on analoogsed). Kui on tegemist klaviatuurilt sisestamise või ASCII faili importimisega, siis tuleb andmestik eelnevalt dokumenteerida — luua vastav tekstifail infoga muutujate kohta (eelkõige iga muutuja eeldatavad miinimum- ja maksimumväärtused; kas muutuja on numbriline või sümbolmuutuja; sümbol-

muutuja väärtused; soovi korral on võimalus ka sisestada iga muutuja kohta selgitavat lisainformatsiooni). Andmestiku dokumentatsiooni saab valmistada nii NSDstat-i programsete vahenditega, kui ka ükskõik millise tekstiredaktori võimalusi kasutades ASCII tekstina, arvestades muidugi vastavat NSDstat-i poolt nõutavat süntaksit. Mina isiklikult kaldun pooldama viimast võimalust just oma blokioperatsioonide võimaluste tõttu.

Kõik sisestamise võimalused lõpevad NSDstat-i süsteemfaili ehitusega.

NSDstat-i analüüsimoodul. Andmeanalüüsiks pakub NSDstat laia meetodite ja tehnikate valikut (joonis 1). Kuid enne analüüsimooduli lähemat tutvustust annan ülevaate funktsionaalklahvide abil sooritavatest operatsioonidest, mis iseloomustavad hästi NSDstat-i kasutamise mõningaid võimalusi.

F1. Võimaldab saada analüüsi või sisestuse igal etapil konkreetset abi. Lisaks nõuannetele, mis puudutavad puhtalt programmiga töötamisega seotud tegevusi, on võimalus kohe saada infot väljaarvutatud statistiliste karakteristikute kohta — s.o. mis on mis ja kuidas vastavaid tulemusi tõlgendada. See on teave, mis õigupoolest kogusummana annab tervikliku sotsiaalteaduste andmeanalüüsi lihtsa käsiraamatu (joonis 2).

KESKMIINE		Graafika	
		GEST	% OLEMASOLEVAITES
Ariteetiline keskmine on suuruste summa jagatud nende suuruste arvuga. Ta iseloomustab kogumi elementide paiknemist.	- Kesk	2.48	42.48
	- Mood	2.91	22.91
Keskmiist saab kasutada vaid numbrilise muutuja korral.	- Mediaan	23.87	23.87
	00	7.16	7.16
	30	2.39	2.39
	18	0.72	0.72
	3	0.48	0.48
SUM =		419	100.00
			100.00

*** Kaasaarvatud: 419 *** Puuduvaad: 0 *** Kokku: 419 ***

Joonis 2. Näide NSDstat-i statistikaalase abi kohta.

F2. Ekraanile ilmub analüüsitava andmestiku muutujate nimestik.

F3. Selle klahvivajutuse abil saame teada muutuja pikema nime ning vastavalt siis, kas info sümboolmuutuja võimalikest väärtutest või arvulise muutuja miinumumist-maksimumist.

F4. Analüüsitava andmestiku vastava muutuja kohta täiendav pikem informatsioon (joonis 3).

OECD: Esimene täiskasvanu annab perekonna lüdkoefitsendile juurde 1.0, iga järgmine täiskasvanu aga 0.7. Iga laps vanuses 14 - 18 aastat on arvesse võetud koefitsendiga 0.7. Iga laps vanuses 0 - 13 aastat on arvesse võetud koefitsendiga 0.5

Min	1.00
Max	5.10
Number	419
1. kvartil	1.70
Mediaan	2.40
3. kvartil	2.90

== Kaasaarvatud: 419 == Puuduva: 0 == Kokku: 419 ==

Joonis 3. Näide NSDstat-i vabateksti dokumentidst.

F5. Annab võimaluse programmi tööd lõpetamata operatsioonisüsteemi minekuks.

Ülejäänud funktsionaalklahvidest võiks nimetada veel:

F7. Tekstiliste või graafiliste analüüsiväljundite suunamine printerile väljatrükiks.

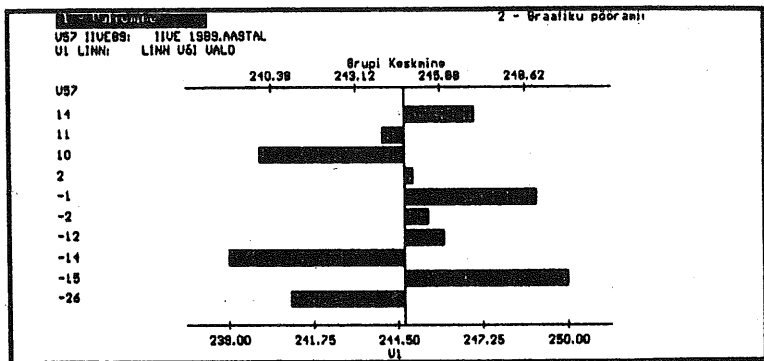
F8. Tekstiliste väljundite suunamine faili, mida saab hiljem töödelda tekstiredaktoriga (joonis 4):

v3 TEENIJAD v4 LAPSED							
	0.00	1.00	2.00	3.00	4.00	5.00	SUM
0.00	00.6	47.0	21.5	42.9	75.0	100.0	42.5
1.00	9.7	24.8	25.0	38.1	25.0	0.0	22.9
2.00	4.2	18.8	39.5	4.8	0.0	0.0	23.9
3.00	1.4	4.7	12.2	4.8	0.0	0.0	7.2
4.00	4.2	2.7	0.6	9.5	0.0	0.0	2.4
5.00	0.0	1.3	0.6	0.0	0.0	0.0	0.7
6.00	0.0	0.7	0.6	0.0	0.0	0.0	0.5
SUM	100.0	100.0	100.0	100.0	100.0	100.0	100.0
N=	72	149	172	21	4	1	419

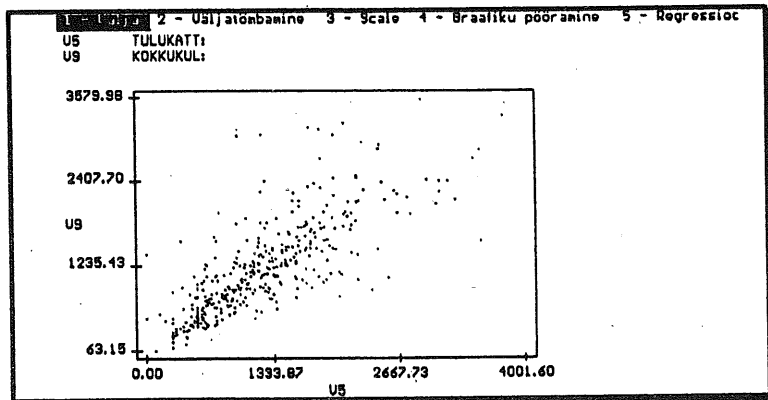
== Kaasaarvatud: 419 == Puuduva: 0 == Kokku: 419 ==

Joonis 4. Näide väljatrükist faili.

Andmeanalüüsi võimalustest paketi NSDstat-i abil annab hea ülevaate joonis 1. Lisaks programmi poolt väljaarvutatavatele statistilistele suurus-tele pakub NSDstat laialdase graafiliste esituste valiku. Hästikujundatud graafik on ju kaheldamatult kõigist statistilise informatsiooni analüüsi ja esitamise meetoditest lihtsaim ja samal ajal efektiivseim. NSDstat-is toimub ümberlülitamine erinevate graafiliste esituse võimaluste vahel kiiresti ja lihtsalt. (Vt. väikest graafiliste esituste valikut joonistel 5 ja 6.)

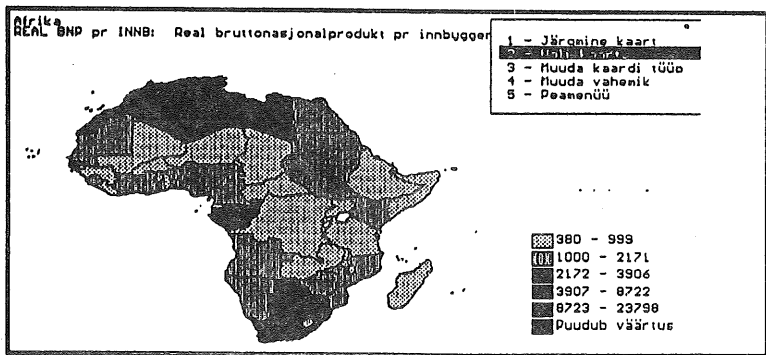


Joonis 5. Tulpdiagrammi näide.



Joonis 6. Hajusgraafiku näide.

Analüüsimooduli tutvustuse lõpuks jõuame suhteliselt harvaesineva and-
mestike analüüsi meetodi juurde. See on võimalus siduda mingi geo-



Joonis 7. NSDstat-i viirutatud kaardi näide.

graafilise üksuse andmed vastava administratiivse jaotusega kaardil ning saadud temaatilisi kaarte kasutada efektiivselt analüüsi käigus (joonis 7).

NSDstat-i olevik ja tulevik

NSDstat-i praegune variant on umbes seitsmeaastase arengu tulemus. NSDstat on praegu kõige laialdasemalt kasutatavam õppetarkvara Norra keskkoolides ja alates 1990. aastast levitatakse seda ka mitmetes teistes Euroopa maades. NSDstat-ist on peale norrakeelse versiooni olemas veel rootsi, soome, taani, saksa, inglise ja nüüd ka eestikeelne algversioon. On ju väga tähtis, et kobavale arvutikasutamisele ja ebakindlatele statistilistele teadmistele ei lisanduks veel statistika- ja arvutialase võõrkeele mittetundmine. Eestikeelse programmi edasine käekäik on momendil kahjuks takerdunud objektivsete ja subjektiivsete tegurite, s.o. siiski peamiselt rahaliste raskuste tõttu.

Rääkides NSDstat-i kasutamise õiguslikest aspektidest, siis tema suureks eeliseks teiste statistikapakettide ees on tema legaalse kasutamise võimalus. NSD on andnud Eesti Vabariigi mittekommertsstruktuuridele NSDstat-i tasuta kasutamise õiguse. Eestis lisanduvad vaid siin tõlkimise ja levitamise seotud kulutused, mis on eelduste kohaselt tühisel võrreldes programmi tegeliku väärtusega.

Lõpetuseks võib märkida, et mulle teadaolevalt on paketi loojad oma valitud tees, s.o. lihtne andmeanalüüs ja -esitus kindlad ning ei kavatse juurde tuua keerulisemaid analüüsi elemente, nagu näiteks faktor- või klasteranalüüsi. Keerulisema statistilise analüüsi jaoks on olemas ju tunduvalt mahukamad ja loomulikult ka tunduvalt kallimad programmid.

LÜHIÜLEVAADE PAKETIST SYSTAT

Anne-Mai Parring, Jüri Vilismäe

Tartu Ülikool, Matemaatilise Statistika Instituut

Statistikapakett SYSTAT on firma SYSTAT inc. poolt toodetud ja täiendatav tarkvarapakett. Lahendatavate ülesannete mahu ja realiseeritud statistika-meetodite hulga poolest on see keskmise suurusega pakett. Võimalik on töödelda tunduvalt mahukamaid andmestikke kui paketiiga STATGRAPHICS (silmas on peetud versiooni 3.0), ka realiseeritud meetodite hulk on mõnevõrra täielikum. Paketi SYSTAT võimalused on aga piiratumad kui paketil SAS.

Tartu Ülikoolis on paketti SYSTAT rakendatud nii õppetöös kui ka praktiliste ülesannete lahendamisel. Põhiliselt kasutatakse versiooni 4.1 (ilma graafikamoodulita SYGRAPH). Seda versiooni on silmas peetud ka järgnevas kirjelduses. Kõige uuem versioon kannab numbrit 5.2, töötab keskkonnas WINDOWS ja käskude asemel võib teda juhtida valikutega etteantud menüüdest.

Paketi kasutusjuhend on kirjutatud väga heas stiilis ning ilmekaid näiteid kasutades. Autoril, Leland Wilkinsonil, on oma kindel vaatepunkt mitmes statistika vaidlusaluses küsimuses ja seetõttu on juhendi lugemine niihästi õpetlik kui ka huvitav.

Pakett SYSTAT koosneb üksikutest moodulitest, mis töötavad üksteisest sõltumatult. Ta on kasutatav ka võrdlemisi tagasihoidlikult komplekteeritud arvutitel — korraga võib arvutis hoida ainult ühte-kahte konkreetse ülesande lahendamiseks vajalikku moodulit.

SYSTAT'i moodulid, abistav teave, andmestikud ja käsufailid on tavaliselt koondatud kataloogi SYSTAT. Paketi käivitamiseks tuleb sisestada tema nimi SYSTAT. Selle peale ilmub paketti kuuluvate moodulite loetelu, millest vajaliku valimiseks tuleb sisestada mooduli nimi või number.

Erinevate moodulite poolt teostatav tegevus on erinev. Andmete sisestamiseks on ette nähtud kaks moodulit EDIT ja DATA. Sisestada võib (klaviatuurilt või ASCII-failina) nii traditsioonilist objekt—tunnus-maatriksit, kuid ka korrelatsioonimaatriksit jms. ebatraditsioonilist andmestikku. Mõnede populaarsete andmetöötussüsteemide (Lotus, d'BASE jt.) faile võib importida originaalkujul, neid ASCII-failideks teisendamata.

Samades moodulites on realiseeritud väga rikkalikud võimalused andmete teisendamiseks ning uute tunnuste arvutamiseks. Uue tunnuse arvutamisel võib kasutada mitmesuguseid matemaatilisi funktsioone ja keerulisi aritmeetilisi avaldisi. Arvutuseeskirja kirjeldamine toimub BASIC-keelega väga lähedases keeles.

Moodulite EDIT või DATA töö tulemuseks on enamasti SYSTAT-fail. Mooduli DATA abil võib väljastada ka ASCII-faili või eksportida mõnda populaarsesse andmetöötussüsteemi (Lotus, d'BASE jt.) sisestatava faili.

Suurem osa ülejäänud moodulitest on ette nähtud andmete statistiliseks töötlemiseks. Moodul MACRO võimaldab kirjutada "makrosid" — programme uute algoritmide realiseerimiseks ning ebastandardseks töötuseks.

Mingi töötuse teostamiseks tuleb siseneda vajalikku moodulisse ja sisestada klaviatuurilt soovivat tegevust määrav käskude jada. Igas moodulis on kasutatavad käsud mõnevõrra erinevad, nende loetelu saab välja kutsuda käsuga help.

SYSTAT-käsud jagunevad kuumadeks ja külmadeks. Kuum käsk täidetakse kohe, külmad käsud täidetakse niipea, kui nende järel sisestada mistahes kuum käsk.

Vaatame näitena moodulis DATA täidetavat käskude jada, mille eesmärgiks on väljastada ekraanile failis ltest oleva andmestiku 10 esimese objekti andmed.

Soovist kasutada mingit SYSTAT-faili standardse laiendiga sys teatatakse käsuga use, mille järel kirjas faili nimi

use ltest

See on kuum käsk. Kohe pärast vajutust klahvile "Enter" käsk täidetakse. Ekraanile ilmub failis olevate tunnuste nimede loetelu. Paketis SYSTAT kehtivad nimede määramisel tavalised kitsendused: nimi on tähega algav täntede ja numbrite jada, mis ei ole pikem kui 8 sümbolit. Mittearvulise tunnuse nimi peab lõppema dollari sümboliga \$. Paketis SYSTAT ei erista suuri ja väikesi tähti.

Faili esimese 10 objekti andmete ekraanile saamiseks tuleb sisestada käsud:

casellst

repeat 10

Need on külmad käsud, mille käivitamiseks tuleb sisestada kuum käsk. Praegu sobib selleks kuum käsk

run

Tulemusena ilmub ekraanile soovitud väärtuste loetelu.

Käsus esinevast põhisonast võib reeglina kirjutada välja ainult kaks esimest tähte. Seega võime eelnevad käsud sisestada ka kujul:

ca

re 10

ru

SYSTAT suudab meeles pidada 5 viimasena sisestatud käsku. Neist mõne ekraanile kutsumiseks tuleb (korduvalt) vajutada võtmeklahvile "F9". Ekraanile kutsutud käsku võib parandada ja täiendada ning seejärel uuesti sisestada.

Saadud tulemused väljastatakse alati ekraanile. Soovi korral on tulemusi võimalik dubleerida ka trükkalil või salvestada nad väljundfaili (ASCII-fail). Trükkalit saab valida käsuga

output

väljundfaili nimi määratakse sama käsuga, näiteks

output tulem.out

Väljundi määramiseks ainult ekraanile kasutatakse käsku

output *

SYSTAT protokollib automaatselt ühe töökäigu ajal täidetud käsud. Käskude faili vaatamiseks tuleb pöörduda paketisisesse tekstiredaktori poole käsuga

fedt >

Selle käsu täitmisel tuuakse ekraanile käskude faili sisu. Ekraanil olevaid käske võib muuta ja muudetud faili saab salvestada uue nime all, näiteks **kasud1.cmd**. Kord salvestatud käskude faili võib uuesti käivitada käsuga **submit**, mille järel näidatakse ära käivitatava faili nimi. Näiteks

submit kasud1

Niisugune lihtne võimalus oma töötuseeskirja säilitamiseks ja selle korduvas kasutamiseks on paketi SYSTAT üks võluvamaid ja kaasaegsemaid jooni. See muudab paketi kasutamise mugavaks ja kiireks. Sama käsuga on võimalik käivitada ka kasutaja kirjutatud "makrosid".

Automaatselt protokollitud käskude fail kustutatakse süsteemist väljumisel. Nii moodulist kui ka süsteemist saab väljuda käsuga **quit**.

Kui käsu **quit** järel kirjutada mingi faili nimi, salvestatakse automaatselt protokollitud käsufaili vastava nimega faili.

SYSTAT säilitab automaatselt ka osa väljundist (viimased 256 rida). Väljundfaili vaatamiseks tuleb pöörduda paketi sisese tekstiredaktori poole käsuga

fedt *

Väljundfaili saab — nagu kõiki teisi faile — "lehitseda" klahvidega "PgUp" ja "PgDn".

Ülejäänud moodulite abil saab teostada andmete statistilist analüüsi. Kõikides moodulites võib kasutada puuduvate väärtustega andmestikke. Samuti on võimalik teostada töötlust rühmade kaupa. Rühmitava tunnuse nimi määratakse käsuga **by**. Rühmade kaupa töötlemiseks aga peavad andmed olema rühmitava tunnuse väärtuste järgi järjestatud. Veel on võimalik käsu **select** abil valida olemasolevast SYSTAT-failist etteantud tingimust täitev objektide hulk ja töödelda vaid neid objekte.

Kirjeldame lühidalt statistilise analüüsi mooduleid.

CLUSTER — moodul, mis on ette nähtud nii objektide kui ka tunnuste rühmitamiseks. Lähteandmestikuna on kasutatav nii objekt—tunnus-maatriks kui ka samasuste või kauguste maatriks.

Rühmitamiseks võib kasutada 6 hierarhilist meetodit või mittehierarhilist keskmete meetodit. Soovitava klastrite arvu juures võib klasteriseerimise peatada ning salvestada klasteriseerimise tulemuse.

CORR — moodul, mis on ette nähtud sümmeetrilise kolmnurkse seosekordajate, sarnasuste või kauguste maatriksi arvutamiseks. Maatriksi elementideks võib määrata mõne järgmistest suurustest:

- Pearsoni lineaarse korrelatsioonikordaja;
- kovariatsiooni,
- ristkorrutise;
- gamma-kordaja;
- Guttmani kordaja;
- Spearmani astakkorrelatsioonikordaja;
- Kendalli korrelatsioonikordaja;
- eukleidilise kauguse;
- dihhotoomiliste tunnuste seosekordaja.

Pearsoni korrelatsioonikordajate jaoks arvutatakse ka olulisustõenäosused ning Bartlett statistik kontrollimaks, kas korrelatsioonimaatriks erineb ühikmaatriksist.

FACTOR — moodul, mis on ette nähtud faktoranalüüsi teostamiseks. Faktorite leidmiseks on realiseeritud kaks meetodit — peakomponentide meetod ja iteratiivne peafaktorite meetod. Faktooreid võib pöörata kolme erineva meetodiga: *varimax*, *equamax*, *quartimax*. Nii faktorkaale, leitud faktooreid kui ka faktorite arvutamiseks kasutatud kordajaid võib salvestada. Lähteandmestikuna võib kasutada nii objekt—tunnus-maatriksit kui ka korrelatsioonimaatriksit (sel juhul pole muidugi faktorite väärtusi võimalik välja arvutada).

GRAPH — moodul, mis on ette nähtud andmete graafiliseks illustreerimiseks. Selle mooduli abil võib joonistada histogramme, hajuvusdiagramme, tüvi—leht-diagramme, karp—vurrud-diagramme, funktsioonide graafikuid. Saab konstrueerida tõenäosusdiagramme normaaljaotuse sobivuse kontrollimiseks. On ka valik parameetreid, mille abil saab graafikut kujundada.

MDS — moodul, mis on ette nähtud andmete mitmemõõtmeliseks skaalerimiseks. Mitmemõõtmeline skaalerimine on võimas vahend andmete näitlikustamiseks — etteantud kauguste (või sarnasuste) maatriksi põhjal paigutatakse punktid etteantud dimensiooniga ruumi (kui võimalik, siis kahemõõtmelisse) nii, et punktide vahelised kaugused (või sarnasused) oleks võimalikult lähedased esialgsetele.

MGLH — moodul, mis on ette nähtud mitmemõõtmelise lineaarse mudeli

$$Y = XB + E,$$

kus Y — sõltuvate tunnuste väärtuste $n \times k$ maatriks;

X — argumentide $n \times p$ -maatriks,

B — parameetrite pxk -maatriks,

E — juhuslike vigade $n \times k$ -maatriks,

parameetrite hindamiseks ja üldiste lineaarsete hüpoteeside:

$$ABC=D$$

kontrollimiseks parameetrite maatriksi B kohta. Selle mooduliga saab lähendada regressioonanalüüsi, dispersioonanalüüsi, mitmemõõtmelise dispersioonanalüüsi, diskriminantanalüüsi ja kanoonilise analüüsi ülesandeid, teostada profiili analüüsi. Moodulis on olemas rikkalik valik võimalusi vahe- ja abitulemuste salvestamiseks ning eraldi analüüsimiseks.

NONLIN — moodul, mis on ette nähtud mittelineaarse mudeli kordajate hindamiseks. Mudeli kuju etteandmisel võib kasutada avaldist, mis on lähedane **BASIC**-avaldisele. Parameetrite hindamiseks on realiseeritud kaks meetodit: Newtoni ja simpleks-meetod. Regressioonifunktsioone on võimalik määrata erinevates argumentide piirkondades erinevalt. Jääke ja hinnanguid saab salvestada ja omaette analüüsida.

NPAR — moodul, milles on realiseeritud rida mitteparameetrilisi meetodeid. Selle mooduli abil on võimalik kontrollida Kolmogorov—Smirnovi meetodiga normaaljaotuse sobivust, võrrelda kahte sõltuvat valimit (märgitest ja Wilcoxon'i test), mitut sõltuvat valimit (Friedmani test) või mitut sõltumatut valimit (Kruskal—Wallise ja Mann—Whitny. test, Kolmogorovi—Smirnovi test).

SERIES — moodul, mis on ette nähtud aegridade analüüsimiseks. Võimalik on valida mitu erineva aegrea mudelit: lineaarne ja mittelineaarne filtreerimine, Fourier' analüüs, sessoonne dekompositsioon, mittesessoonne ja sessoonne eksponentsiaalne silumine, Box—Jenkinsi analüüs mittesessoonsete ja sessoonsete mudelite jaoks. Samuti on võimalik aegrea graafiline illustreerimine ja kahe aegrea üheaegne analüüs.

STATS — moodul, mis on ette nähtud arvuliste tunnuste põhistatistikute arvutamiseks. Statistiku on võimalik arvutada ka teatava tunnuse väärtuste abil määratud rühmade jaoks. Samuti on võimalik tellida T -testi kahe rühma keskväärtuse võrdlemiseks ja ühefaktorilist dispersioonanalüüsi mitme rühma keskväärtuse võrdlemiseks.

TABLES — moodul, mis on ette nähtud ühe- ja mitmemõõtmeliste saagedustabelite koostamiseks. Võimaldab kontrollida tunnuste sõltumatust, arvutada Mantel—Haenzeli statistikut tunnuste vahelise seose uurimiseks, kus segavate tunnuste mõju on eemaldatud. Samuti saab konstrueerida \log -lineaarseid mudeleid ja analüüsida nende sobivust.

SEOSEMUDELITE KASUTAMINE ÜHISKONNA- JA MAJANDUSTEADUSTES

Liina-Mai Tooding

Tartu Ülikool, Matemaatilise Statistika Instituut

Käesolev artikkel ei ole pühendatud ühele konkreetsele programmile, vaid teatud mudelite klassile, mis on realiseeritud mitme programmipaketi koosseisus. Käsitlemisele tulevad koos tarkvaraülevaatega:

- 1) *log*-lineaarse analüüsi mudel tunnustevahelise statistilise seose kindlaksmääramiseks ja
- 2) lineaarsete struktuurivõrrandite mudel põhjusliku seose analüüsiks tunnuste vahel.

Tegemist on mitmemõõtmelise statistilise analüüsi meetoditega, mille kasutusala peetakse tavakohaselt üksnes süvateaduslikke uurimusi. Vaieldamatult on kõne alla tulevate mudelite korral tegemist kõrge matemaatilise abstraktsiooni astmega, kuid need mudelid ei ole sugugi elukauged.

Mõlema meetodi korral piirdume üksnes idee populaarse tutvustamisega, sest vastav matemaatiline tagapõhi ei mahu käesoleva väljaande raamidesse.

Uuritavaid tunnuseid tähistame edaspidi A, B, C, \dots , lisades vajaduse korral indekseid.

1. Log-lineaarne analüüs

Kahe tunnuse A ja B jaotustabel on hästi mõistetav konstruktsioon tunnuste ühisjaotuse ja tunnustevahelise seose iseloomustamiseks. Jaotustabeli elementideks on rea ja veeru ristumiskohale vastava väärtuspaari tõenäosus p_{ij} , $i = 1, 2, \dots, k$, $j = 1, 2, \dots, l$ (vt tabel 1). Väärtuspaari all mõistame kas üksikväärtuste või väärtusklasside paari, sest sageli on tegemist klassifitseeritud andmetega, seda niihästi pidevate väärtustega kui ka diskreetsete tunnuste korral. Kui tunnused on sõltumatud, siis tõenäosusteooria hästi tuntud tulemuse põhjal kehtib seos:

$$p_{ij} = p_i \cdot p_j, \quad (1)$$

kus p_i ja p_j on vastavalt vaadeldavate rea- ja veeruväärtuste serva- ehk marginaaltõenäosused. Statistiline sõltuvus on teatavasti jaotuste sõltuvus: teades indiviidi paiknemist tabelis ühe tunnuse väärtuse alusel, saame prognoosida tema paiknemist teise tunnuse seisukohalt. Sõltuvuse puududes ei sisalda tabeli siseosa täiendavat informatsiooni servajaotustega võrreldes.

Empiirilisel tasandil esitab tunnuse jaotust empiiriline jaotustabel, mille elementideks on tõenäosuse hinnangud n_{ij}/n , kus n_{ij} on vastava väärtuspaari sagedus valimis mahuga n . Marginaaltõenäosuste hinnanguteks on vastavalt arvud n_i/n ja n_j/n .

Jaotustabeliga analoogiline ehitus on sagedustabelil, milles paiknevad tõenäosuse p_{ij} asemel vastava väärtuspaari esinemissagedused N_{ij} . Sõltumatus korral saame seosest (1) tingimuse

$$n_{ij} = n (n_i/n) (n_j/n) = (1/n) n_i \cdot n_j, \quad (2)$$

Analoogia põhjal selle seosega võetakse *log*-lineaarses analüüsis populatsioonisageduse (ehk teoreetilise sageduse) N_{ij} esituseks kasutusele mudel

$$N_{ij} = T T_i^A T_j^B T_{ij}^{AB}. \quad (3)$$

milles ülemine indeks näitab, millisele tunnusele ja alumine indeks, millisele tunnuse väärtusele vastab mõjukomponent.

Tabel 1

B						
A	1	...	j	...	1	SUMMA
1	P_{11}	...	P_{1j}	...	P_{1l}	$P_{1.}$
...
i	P_{i1}	...	P_{ij}	...	P_{il}	$P_{i.}$
...
k	P_{k1}	...	P_{kj}	...	P_{kl}	$P_{k.}$
SUMMA	$P_{.1}$...	$P_{.j}$...	$P_{.l}$	1

Mudelisse (3) on seosega (2) võrreldes lisatud liige T_{ij}^{AB} , et arvesse võtta võimalikku tunnuste A ja B ühismõju.

Leides seose (3) mõlemast poolest logaritmi, saame seose

$$L_{ij} = L + L_i^A + L_j^B + L_{ij}^{AB}, \quad (4)$$

mis kujutab endast lineaarset avaldist parameetrite suhtes. Selle põhjal ongi valitud meetodi nimi — *log*-lineaarse analüüsi.

Mudel (4) on analoogiline dispersioonanalüüsi mudeliga ja seetõttu võiks *log*-lineaarset analüüsi nimetada ka sageduste dispersioonanalüüsiks.

Vastavalt sellele, millised on aineteeoreetilised hüpoteesid, valitakse mudeli konkreetne kuju. Oletades näiteks, et sagedustabelis peegelduva variatiivsuse ainsaks allikaks on tunnus A, tuleks uurida mudelit

$$N_{ij} = T T_i^A,$$

milles puuduvad tunnusele B vastavad mõjutegurid.

Kui ühismõju tegur $T_{ij}^{AB} = 1$ (või mudeli (4) kohaselt $L_{ij}^{AB} = 0$), siis on tegemist sõltumatuse mudeliga:

$$N_{ij} = T T_i^A T_j^B.$$

Log-lineaarse mudeli võib püstitada mis tahes arvu tunnuste omavahelise seose analüüsimiseks. Elemendiks, mille kohta püstitatakse avaldise (3) või (4) tüüpi mudel, on p tunnuse korral p -mõõtmelise ühisjaotuse element — vastava väärtuskombinatsiooni sagedus.

Kolme tunnuse juhul tuleks mudel kombineerida mõjudest, mis tulenevad tunnustest A , B ja C ning kõikvõimalikest koosmõjudest, mis tekivad tunnuste koosvaatlemisel kahe- või kolmekaupa: AB , AC , BC , ABC . Koosanalüüsivate tunnuste arvu suurenemisel kasvab võimalike mõjukomponentide hulk kiiresti. Reeglina on aga praktilistele ülesannetele omane see, et uurimishüpoteesid ei haara korraga kuigi suurt tunnuskomplekti, sageli mitte enam kui 5–8 tunnust, ja seetõttu on *log*-lineaarne mudel tõlgenduslikult veel jõukohane.

Mudeli (3) abil on võimalik prognoosida sagedusi ja mudeli (4) alusel sageduste logaritme. Komponentid T (või L) arvutatakse valimiandmete põhjal nii, et mudeli alusel saadud "teoreetiliste" ja valimi põhjal saadud empiiriliste sageduste erinevus oleks võimalikult väike. Kui see erinevus on mingi statistilise kriteeriumi alusel siiski lubatavast suurem, siis püütakse mudelit parandada, lisades sellesse veel täiendavaid mõjusid.

Mida suurem on mudelis kasutatavate parameetrite arv, seda parema prognoosi me saame. Rakenduslikust aspektist lähtudes tuleb aga eelistada vähem parameetreid sisaldavat mudelit, sest seda on lihtsam tõlgendada.

Sageduste "taastamine" teoreetilise mudeli alusel ei ole iseenesest *log*-lineaarse analüüsi eesmärgiks. Analüüsi eesmärgiks on välja selgitada, millised mõjukomponendid on vajalikud selleks, et saada küllalt häid prognoose. Kui vajalikuks osutub mitme tunnuse koosmõju tegur, siis võime järeldada, et nende tunnuste ühisjaotus kujuneb teatava stereotüübi ehk tunnustevahelise seose alusel. Nii osutubki *log*-lineaarse analüüsi oma olemuselt tunnustevahelise seose analüüsiks. Näiteks, kui mudelis (3) osutub oluliseks komponent T_{ij}^{AB} , siis on tunnused A ja B omavahel seotud.

Log-lineaarse analüüsi rakendamisel ei püstitata tunnuste statistilise iseloomu kohta praktiliselt mingeid kitsendusi, meetodit võib rakendada niihästi kvalitatiivsete kui ka kvantitatiivsete tunnuste analüüsiks. Ainsaks piiranguks on see, et tunnus peab olema agregeeritav võimalikult jämedale skaalale. Vastasel juhul — seda eriti väikese mahuga valimi korral — osutub jaotustabel hõredalt täidetuks ja järeldused võivad tulla ekslikud.

Rõhutame eriliselt, et statistiliselt korrektsete järelduste tegemiseks ei ole *log*-lineaarse analüüsi puhul vaja eeldada tunnuste tavapärasest normaal-sust, mis teeb selle meetodi sobivaks ka sotsiaalteaduste vallas, kus sageli

on tegemist hinnanguskaaladel mõõdetud diskreetsete tunnustega või tunnustega, mille puhul ei ole täidetud isegi jaotuse sümmeetrilisuse eeldus. Statistiliste otsustuste aluseks *log*-lineaarses analüüsis on hii-ruut-statistik. *Log*-lineaarse analüüsi põhjalikuma käsitluse leiab lugeja näiteks monograafiast [3].

Log-lineaarse analüüsi programmid on lülitatud enamusse statistikapaketitesse, olles mitnemõõtmelise jaotuse analüüsi loomulik jätk. Tabelis 2 on loetletud *log*-lineaarse analüüsi teostamiseks sobivad moodulid reas meil enamlevinud statistikaalastes programmpakettides (vt. [4]).

Tüüpiliselt sisaldavad *log*-lineaarse analüüsi programmid järgmisi võimalusi:

- 1) etteantud *log*-lineaarse mudeli komponentide olulisuse kontroll, et välja selgitada olulised mitmesed seosed,
- 2) mõjukomponentide arvilise suuruse hindamine, et leida parameetrite *T* empiirilised väärtused sageduste lahutuse mudelis,
- 3) mudeli konstrueerimine sammuviisiliselt kas lihtsat mudelit keerulisemaks muutes uute mõjukomponentide lisamise teel, kuni saame küllalt hea sageduste prognoosi, või vastupidi, keerulisemalt mudelit lihtsamatele üle minnes, senikaua kuni prognoos on veel küllalt hea.

Tabel 2

PAKETT	MOODUL, KÄSK
BMDP	programm 4F
SPSS	HILOGLINEAR, LOGLINEAR
SAS	CATMOD
SYSTAT	TABLES: MODEL
STATGRAPHICS	CATEGORICAL DATA ANALYSIS: LOG-LINEAR ANALYSIS

Lihtsa mudeli all mõteme seejuures võimalikult vähest arvu mõjukomponente sisaldavat mudelit.

Log-lineaarse analüüsi rakendusvõimalused ei ole praktiliselt millegagi piiratud ja ta sobib iga valdkonna ülesandeks, kus on probleemi võimalik formaliseerida tunnuste vastastikuse seose hüpoteesidena. Tulemuste seas on enamasti ka lisainfo selle kohta, millest seos tuleneb. Mõju suunda tunnustevahelises seoses saab siiski käsitleda ainult tõlgenduslikult, kuid mitte formaalselt, mudeli abil. *Log*-lineaarse analüüs ei anna informatsiooni seose põhjusliku iseloomu kohta, kuid just nimelt see on sageli praktiliste ülesannete eesmärgiks.

2. Põhjusliku seose analüüs

Mudel, mida vaatleme artikli teises pooles — lineaarsete struktuurivõrandite mudel —, ongi loodud suunaga seoste analüüsiks. Täpsemalt,

probleemi formaliseerides ja mudelit püstitades on võimalik fikseerida mõju oletatav suund ja viia analüüs läbi selle eelduse korral.

Käsitleme järgnevalt seda lähenemisviisi K. Jöreskogi ideede põhjal välja töötatud Lisrel-mudelite alusel (vt. [2]). Lineaarsete struktuurivõrandite mudeli idee mõistmiseks meenutame kaht hästi tuntud mitmemõõtmelise statistilise analüüsi meetodit — regressioonanalüüsi ja faktoranalüüsi.

Regressioonanalüüs on teatavasti mõeldud teatava tunnusvektori, nn. sõltuva ehk funktsioontunnuse Y prognoosimiseks sõltumatute tunnuste vektori ehk argumenttunnuse X kaudu:

$$Y = G X + h, \quad (5)$$

kus G on regressioonikordajate maatriks ja h on prognoosimudeli viga. Ülesandeks on hinnata valimi põhjal maatriksi G elemente.

Faktoranalüüsi idee tuleneb põhieeldusest, et tunnuste omavahelise seose allikaks on nende tunnuste seos teatavate "kolmandate", varjatud tunnustega. Nii võiksime mõõdetud tunnusvektorit x esitada nende "taga" olevate ühiskomponentide vektori X kaudu:

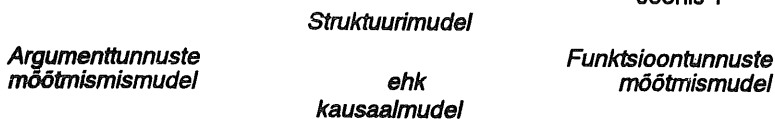
$$x = L X + e, \quad (6)$$

kus e on sellise esituse viga.

Lineaarsete struktuurivõrandite mudelis ühendatakse need kaks ideed. Tunnused jagunevad mõõdetud tunnusteks ja latentseteks, varjatud tunnusteks. Põhjusliku seose mudel ehk kausaalmodell püstitatakse latentsete tunnuste kohta, lugedes üht osa neist argumenttunnusteks (eksogeenseteks tunnusteks) ja teist osa funktsioontunnusteks (endogeenseteks tunnusteks). Seos funktsioon- ja argumenttunnuste vahel moodustab struktuurivõrandi. Seos mõõdetud tunnuste ja latentsete tunnuste vahel moodustab latentsete tunnuste mõõtmismudeli. Mõõtmismudelid formuleeritakse eraldi funktsioontunnuste ja argumenttunnuste jaoks. Kõik seosed on lineaarsed käsitatavate tunnuste suhtes.

Struktuurivõrandite mudel koosneb niisiis kolmest osast:

Joonis 1



Esitame mudelid matemaatilisel kujul. Olgu X latentne argumenttunnuste vektor ja x sellele vastav mõõdetud tunnuste vektor. Analoogiliselt olgu Y latentne funktsioontunnuste vektor ja y sellele vastav mõõdetud tunnuste

vektor. Struktuurivõrandite mudel avaldub siis kujul

$$\begin{aligned} Y &= B Y + G X + h && \text{kausaalmudel} \\ x &= L^X X + d && \text{argumenttunnuste mõõtmismudel} \\ y &= L^Y Y + e, && \text{funktsioontunnuste mõõtmismudel,} \end{aligned} \quad (7)$$

kus B , G , L^X ja L^Y on kordajate maatriksid ja h , d ning e on vigade vektorid. Eeldatakse vigade mittekorreleeritust omavahel.

Lineaarsete struktuurivõrandite süsteemi (7) lahendamiseks tuleb leida kordajate maatriksid, hinnates neid statistiliselt valimi põhjal. Kordaja väärtus näitab vastavate tunnuste vahelise mõju suurust, kusjuures mõju suund on fikseeritud tunnuste jaotamisega funktsioon- ja argumenttunnusteks.

Struktuurivõrandite mudeli teoreetilise käsitlusega võib lugeja tutvuda näiteks K. A. Bolleni monograafia [1] alusel.

Mis puutub algandmete kohta tehtavatesse kitsendustes, siis eeldatakse struktuurivõrandite mudelite korral enamasti seda, et tunnused oleksid järjestatavate väärtustega, sest aluseks võetakse tunnustevaheline korrelatsioonimaatriks (kovariatsioonimaatriks).

Sõltuvalt konkreetsest statistilisest protseduurist, mille alusel mudel koostatakse, püstitatakse vahel ka veel rangemaid eeldusi, kuid struktuurivõrandite mudelit on võimalik kasutada siiski küllalt laia ülesannete klassi korral, kus algandmed avalduvad järjestusskaaladel.

Tabelis 3 on analoogiliselt tabeliga 2 loetletud struktuurivõrandite koostamise võimalusi meil tuntumate statistikapaketide korral.

Üheks levinumaks lineaarsete struktuurivõrandite lahendamise vahendiks on K. Jöreskogi lähenemisviisi alusel loodud programmpakett LISREL [2].

Tabel 3

PAKETT	MOODUL, KÄSK
BMDP	eraldiseisev pakett EQS
SPSS	vastav meetoodika puudub
SAS	CALIS
SYSTAT	eraldiseisev moodul EzPath
STATGRAPHICS	vastav meetoodika puudub

Nii *log*-lineaarse analüüsi kui ka lineaarsete struktuurivõrandite koostamine kujutavad endast tunnustevahelise seose detailse analüüsi protseduure, mis võimaldavad uuritava nähtuse sügavat ja põhjalikku analüüsi. On aga selge, et mida detailsem on formaliseeritud käsitlus, seda enam peab olema käepärast sisulist lähteinformatsiooni. Kui struktuurivõrandite mudel on püstitatud aineteoreetilise põhjendusega või isegi sellega vastuolus olevana, siis on ka analüüsi tulemused reeglina ekslikud, kuigi see tarvitse tulemustest formaalselt välja paista. Seetõttu tuleb Lisrel-mudeli kasutamisel olla eriti korrektne ülesande püstitamisel, et vältida selle matemaatiliselt korrektse ja rikka meetoodika väärkasutamist.

Kirjandus

1. Bollen, K. A. Structural Equations with Latent Variables. John Wiley & Sons, 1989.
2. Jöreskog, K. G., Sörbom, D. LISREL7 User's Reference Guide. 1st Ed., Sc. Software, Inc., 1989.
3. Santner, Th. J., Duffy, D. E. The Statistical Analysis of Discrete Data. Springer-Verlag, 1989.
4. The Statistical Software Guide '92/93. Computational Statistics & Data Analysis, vol. 15, N. 2, 1993, pp. 241–267.

METSANDUSANDMETE TÖÖTLEMISE KOGEMUSTEST

Andres Kiviste

Eesti Põllumajandusülikool, Metsakorralduse õppetool

Ligikaudu 40% Eestist on kaetud metsaga. Metsast saadav puit on praegu üks Eesti tähtsamaid eksporditartikleid ja metsa kui taastuva loodusrikkuse oskusliku kasutamine on üks meie reaalsemaid tulevikulootusi. Selleks, et metsi arukalt majandada, on vaja hulgalist teavet nii metsade seisundi kui ka nende edasise kasvu kohta.

Eestis kuuluvad kõik metsad 10-aastase ajaintervalli järel hindamisele (takseerimisele), mille käigus koostatakse iga metsaeraldise (puistu) kohta silmamõduline takseerikirjeldus. Takseerikirjeldus sisaldab andmed kasvukohatüübi, puistu päritolu, liigilise koosseisu, liikide kaupa takseertunnuste (vanus, kõrgus, diameeter, tagavara), majandusliku tegevuse jms. kohta. Eesti puistute andmebaas säilitatakse Eesti Metsakorralduskeskuses ja see andmestik on aluseks nii metsandusettevõtete majanduskavade koostamisele kui ka riikliku metsapoliitika kavandamisele.

Metsakorralduse takseerikirjelduste andmebaas võimaldab arvutil "läbi mängida" paljusid metsa kasvatamise ja metsa kasutamise strateegiaid. Selle tegemiseks on tarvilikud metsa kasvu mudelid, mis kirjeldaksid puistu kõrguse, diameetri, tagavara ja teiste takseerimäitajate muutumist sõltuvalt puistu kasvukohatingimustest, vanusest ja puistu seisundist. Puistu kasvukäigu modelleerimisega on tegeldud juba aastakümneid. Esimesed nn. puistu "kasvuseadused" pärinevad käesoleva sajandi algusest. Metsaadlaste suurest huvist puistu kasvu modelleerimise vastu annab tunnistust fakt, et käesolevaks ajaks on publitseeritud ligikaudu sada erinevat metsa kasvufunktsiooni, millest juba ainuüksi ülevaate saamine on keeruline ülesanne.

Puistu kasvumudelite koostamiseks kasutatakse põhiliselt kaht tüüpi andmestikke:

1. metsakorralduse takseerikirjelduste andmed;
2. metsa proovitükkide mootmisandmed.

Metsakorralduse takseerikirjelduste andmete peamiseks eeliseks on nende rohkus. Andmete puuduseks on taksaatorite süstemaatilised vead metsa silmamõdulisel hindamisel. Proovitükkide andmed on küll piisava täpsusega, kuid neid on napilt. Proovitükkide andmete teiseks puuduseks on asjaolu, et tavaliselt uurija ei paiguta proovitükki metsa mitte juhuslikult, vaid nn. tüüpilisse või parimasse metsatükki. Siit tuleneb metsandusandmete representatiivsuse probleem. Paljudes maades (s.h. Soome ja Rootsi) tehakse riigi metsaresursi hindamiseks igal aastal tuhandeid proovitükke, kusjuures proovitükid paigutatakse "täpselt" arvuti poolt kavandatud "juhuslikesse" punktidesse.

Erilise väärtusega on nn. püsiproovitükid, mis on mõõdetud mitmekordselt 5–10 aastase ajaintervalli järel. Püsiproovitükke kasutatakse peamiselt kasvumudelite kontrolliks.

Metsa kasvu mudelid võib jaotada järgmistest klassidesse:

1. Tervikpuistu takseetunnuste keskmist kasvukäiku prognoosivad mudelid. Need on enamasti mittelineaarsed regressioonivõrrandid, mis on koostatud metsakorralduse takseerandmete või suure hulga proovitükkide põhjal. Seda liiki mudeleid on koostatud kõige arvukamalt, kuid nad neid ei saa kasutada paljude oluliste metsandusülesannete lahendamiseks.

2. "Suured" biofüüsikalised puistu kasvu mudelid. Seda tüüpi mudelid kirjeldavad puistut bioloogilise tunnetuse seisukohalt. Need mudelid on väga keerulised ja nende rakendamine metsamajanduse päevaprobleemide lahendamiseks on komplitseeritud.

3. Üksikpuu kasvu mudelile tuginevad puistu kasvu mudelid. See mudelitüüp on kompromiss lihtsate regressioonimudelite ja "suurte" mudelite vahel. Seda tüüpi mudelite näideteks on rootslaste HUGIN ja soomlaste MELA, mis on osutunud üsna efektiivseteks paljude metsanduslike prognoosülesannete lahendamiseks.

Eestis tegeldakse puistu kasvukäigu modelleerimisega EPMÜ metsakorralduse õppetoolis. Meie kasutada on metsakorralduse takseerikirjelduste andmestik ja suur osa proovitükkide andmetest. Nende andmete põhjal on koostatud regressioonivõrrandite süsteem, mis kirjeldab Eesti männikute keskmiste takseetunnuste muutumist sõltuvalt kasvukohatingimustest. Koostamisel on analoogilised mudelid ka teiste puuliikide jaoks. "Suurte" mudelitega on Eestis tegelenud Tõnu Oja (1984) ning Kalevi ja Olevi Kull (1989). Nn. kompromissmudeleid veel Eestis tehtud ei ole.

Eesti metsanduse süsteemis on juurutamisel IBM PC tüüpi personaalarvutid. Praegusel hetkel on personaalarvutid kõikides maakondade metsaametites ja umbes kolmandikul metskondadest. See seab metsandusteaduskonnale tõsise nõude üliõpilaste informaatikaalase ettevalmistuse osas. Tänu Eesti Vabariigi Riiklikule Metsaametile kingiti 1992. aasta lõpul metsandusteaduskonnale 12-kohaline IBM PC-tüüpi arvutiklass, mis võimaldab metsandustudengeile arvuti kasutamise oskust õpetada. Metsandustudengite õppeplaani kavandanud esimeste õppeainetena lisaks informaatika algõpetusele (DOS, Norton Commander) andmebaasisüsteemi (Foxplus) ja tabelarvutuse (Quattro) õpetamist. Matemaatilist statistikat õpetatakse praegu käsikäes programmiga STATGRAPHICS. Geoinfosüsteemide õppeaines õpetatakse programmi IDRISI kasutamist. Kõiki neid programme kasutatakse järgnevatel kursustel paljudes teistes õppeainetes.

KOOSTÖÖ

ON STATISTIKASÜSTEEMI TÄIUSTAMISE ALUS

Euroopa Ühenduse, Euroopa Vabakaubanduse Assotsiatsiooni (EFTA) ning Balti riikide statistikaorganisatsioonide esindajad kõhtusid 15.-17. novembril Tallinnas, et jätkata tšüd 1992. a. alustatud koostööprogrammi raames.

Euroopa Ühenduse / EFTA statistikaalasel 3-aastasel koostööprogrammil on järgmised finantseerimisallikad:

- Euroopa Ühenduse rahalised vahendid PHARE fondide näol
- EFTA rahalised vahendid
- EFTA riikide rahalised vahendid

PHARE fondid on mõeldud Kesk- ja Ida-Euroopa riikide abistamiseks ning jagunevad:

- regionaalne e. piirkondlik fond
- riiklik e. rahvuslik fond

Regionaalset fondi kasutatakse kindla piirkonna, näit. Balti riikide abistamiseks, riikliku fondi vahendite kasutamise otsustab iga konkreetse riigi valitsus, kellele abi osutatakse.

Koostööprogrammi eesmärk on Balti riikide statistikasüsteemi arendamine põhiliselt PHARE regionaalse fondi vahendite arvel ja tehnikaabi PHARE riikliku fondi vahendite arvel. Programm tagab osavõtu rahvusvahelistest nõupidamistest ning statistikute väljaõppe seminaridel ja õppekäikudel. Igal Balti riigil on koordinaator, kes annab konsultatsiooni kohapeal.

Üks tähtsamaid üritusi programmi raames on seminarid statistikaarbijatele. Esimene taoline toimus 3. detsembril 1992. a. Riias ja teine 15. novembril 1993. a. Tallinnas Rahvusraamatukogu auditooriumis. Selle seminari ülesanne oli selgitada arbijatele erinevusi statistiliste ja administratiivsete (ametkondlike) andmete vahel, statistiliste andmete kasutamise korda ning algandmete kaitset reguleerivat seadusandlust. Statistika rollist demokraatlikes turumajanduse riikides, statistika ja tema arbjaja vahelistest seostest rääkis hr. Jüri Kõll Rootsi Statistika Keskbüroost. Hr. Kõll on EC/EFTA ja Baltimaade koostööprogrammi statistikakonsultant Eestis.

Andmekaitse reegleid ja algandmete kaitset reguleerivat seadusandlust tutvustasid hr. Tapio Leppo EFTA Statistikaametist, hr. Erik Aurbakken Norra Statistika keskusest ja hr. Erkki Pentimäki Soome Statistika keskusest. Viimane valgustas ka Rahvusvahelise Statistika Instituudi kutse-etiika deklaratsiooni.

Seminaril osalesid ministerruumide ja ametite esindajad, kuid vaba juurdepääs oli tagatud kõigile. Kahjuks viibis auditooriumis statistika arbijaid vähe. Ilmselt

oli põhjuseks informatsiooni levitamise puudulikkus ja ürituse toimumine põhiliselt inglise keeles.

Kõlama jäi põhimõte: Statistikaamet peab olema sõltumatu, et toota obje...tiivset ja erapooletut statistikat, mis vastaks kõigi arbijate nõuetele. Algandmed peavad olema täielikult kaitstud ja kindlustatud vastava seadusandlusega.

Koostööprogrammi koordinaatsioonikomitee nõupidamine toimus 16. novembril Statistikaameti saalis. Nõupidamist juhtis hr. Erich Bader, Austria Statistikaameti president. Kohal viibisid hr. Klaus Löning Eurostatist ja hr. Tapio Leppo, statistikakonsultandid Rootslst, Soomest ja Norrast ning Balti riikide statistikaorganisatsioonide peadirektorid koos esindajatega.

Nõupidamise eesmärk oli teha kokkuvõtte 1993. aasta tšüdist ja koostada 1994. aasta koostöökava. Järgmise aasta tegevuskava koostab seminaride, konsultatsioonide ja õppereiside ajagraafikust, milles on määratletud ürituse (projekti) eest vastutav EFTA riigi statistikaorganisatsioon ja koht, kus väljaõpe toimub.

Tegevuskava haarab praktiliselt kõik statistikaarvad – 15 prioriteetset valdkonda. Tähtsamatena võib välja tuua rahvamajanduse arvepidamissüsteemi ja hinnastatistika edasiarendamise, valikvaatluse metoodika kasutamise ettevõtlusstatistikas ja rahvastikustatistika arendamise, kuid alahinnata ei tohi ühtki alustatud projekti.

Kahjuks ei saanud koostööprogrammi lülitada inglise keele kursusi, nende tarvis tuleb leida teised võimalused.

Statistikaameti teenistuste juhatajatel oli hea võimalus koordinaatsioonikomitee nõupidamisel osaleda. Isiklike kontaktide loomine aitab kaasa koostöö organiseerimisele.

Teenistuste juhatajatega toimusid kohtumised ja vestlused ka 17. novembril. Huvitavaid kujunesid raamatukogu külastamine ja hr. Aleksander Kaarna poolt hästi ettevalmistatud tutvumise informaatika teenistuse tšüga. Leedu Statistikaameti esindajad viibisid Peaarvutuskeskuse registre ja klassifikatorite info osakonnas ja tutvusid Ettevõttereistriga. Nähtuga jäid kolleegid väga rahule.

Kohtumise raames viibisid külalised Tallinna vanalinnas ja Toompea lossis, kuulasid ühel õhtul ansambli "Hortus Musicus" meeldivat esinemist.

Järgmine koordinaatsioonikomitee nõupidamine toimub Vilniuses 1994. a. kevadel.

Helina Vigla
peadirektori aseäidja

KILDE STATISTIKA AJALOOST

Praktilise statistika algus

Elajaloolisel ajal, mil tekkidid esimesed riigid, sai alguse ka statistiliste andmete kogumine. Andmeid elanikkonnas, ühikute valduses oleva maa ja vara suuruse kohta koguti juba mitu tuhat aastat tagasi. Nendel kaugtel aegadel täitsid statistiku funktsioone preestrid, kõrged võimukandjad, riigiametnikud. Ilma hädapäraste informatsioonita ei olnud võimalik ka siis juhtida riiki, kaitsa tema piire välivaenlaste kallaletungide eest, korraldada majanduselu.

On üldiselt teada, et esimesi statistilisi ülestähendusid Eesti ala kohta leiab näiteks Taani Riigiarhiivis säilitatavat käsikirjalisest kõigest *Liber Censu Danica*, mida meil tuntakse "Taani hindamisraamatu" nime all. Sellest leiab andmeid Põhja-Eesti põllumajandusliku maaomanduse, läänimeeste ning neile kuuluvate valduste kohta. Baltiaks tuntud ajaloolase prof. Paul Johanseni uurimuste järgi on need nimistud koostanud 1219—1220 Põhja-Eestis inimese ristitud preestrid. Tallinna piiskopi Thorikilli poolt 1241. a. lõplikult redigeeritud "Taani hindamisraamatu" andmeid oli Eestis tol ajal 21 000—22 000 adramaad, rahvaarv XIII sajandi algul oli 140 000—160 000.

Tänapäeval on sõnal statistika (sks. *Statistik*, it. *stato* 'riik') vähemal kolm eri tähendust. Statistika, kui refereerida ENE-see koostatud informatsiooni — see on massnähtuste omadusi (arvukust, suurus, levikut, struktuuri jms.) ning massnähtuste vallast toimivaid seaduspärasusi käsitlev teadus, ühiskonna- ja loodusteaduste piirteadus.

Statistika kui teadus tekkis XVII sajandil Sakaamaal nn. oliikolostatistikana ja Inglismaal poliitilise aritmeetikana. Esialgu arenes statistika eriti rahvastikuprotsesside uurimisel, kus tihedasti olulisi seaduspärasusi; koostati esimesed suremüstabelid (E. Halley, 1659—1742), loodi suremüsteenosuse mõiste (Chr. Huygens, 1629—95) ja pandi seega alus elukindlurustatistikale.

Kui statistika all mõista rahvastikku ja kogu rahvajagundust hõlmava informatsiooni hankimise, töötlemise, säilitamise ja edastamise süsteemi, üht peamist arvestuse liiki, siis on tema funktsioonid hoopis konkreetsemad. Statistikaga tegelevad nii sellealased riigiasutused (riiklik statistika) kui ka sel eesmärgil omavalitsusasutuste, ametühingute ning ühiskondlike ja teaduslike organisatsioonide juurde loodud bürood, keskbürood jms. (ametkondlik statistika). Eestis juhib statistikaalast tegevust Statistikaamet, statistikateenistused on mitmel ametil, organisatsioonidel, ettevõtetel.

Argiteadvuses mõistetakse statistika all ka mis tahes objektide kohta hängitud arvandmete hulka.

Kui võtta aluseks praktiline statistika, siis võib "Taani hindamisraamatu" järjekes pida meie kirjuraamatuid. Nendest leiab Eesti ala kohta statistilisi andmeid alates 1670-ndatest aastatest. Rahvaarvu muutuste kohta saab teavet hingederevisjonide lehtedest. Eestimaa kubermangus toimus seitse, Liivimaa kokku kaheksa revisjoni, mille kohta on arhiivmaterjalid säilinud. Teatavasti moodustati Eestimaa ja Liivimaa kubermang Vene imperiumi koosseisus pärast Põhjasõda, täpsemalt selle sõja ajal 1713. a.

Millal loodi Eestimaa kubermangus statistikakomitee?

Et tegemist on Eestis pidevalt statistikaga tegeleva riikliku organisatsiooni stundantümiga, siis huvitab see küsimus mitte üksnes statistikuid, vaid ka laia üldsust. Seda enam, et eespool mainitud küsimuses on esitatud lahkevaid arvamus, aastaarve, kuupäevi, isegi fantaseeritud. Sellest johlvalt tuli pöörduda ürikute poole, mida säilitatakse Tartus Eesti Riigiarhiivis.

Sealt saadud dokumentide koopiatele tuginedes võime öelda, et Vene riigis, mille koosseisu Eesti tol ajal kuulus, asutati kubermangude statistikakomiteed vastavalt Valitseva Senati seadlusele 20. detsembrist 1834. Selles kinnitati Siseministeeriumi juures asuva Statistikaajakonna ja kubermangude statistikakomiteede reeglid mitmesuguste statistiliste andmete kogumiseks ja töötlemiseks. Eesmärgiks oli võimalikult üksikasjalik ja täpne informatsioon kõikides Siseministeeriumi alluvuses olevates üksustes (§ 2), isegi uute linnade plaanide eelnev lõbivaatamine, ettepanekud linnade tulude ja kulude-määrade kohta jpm. Statistikaajakonnale olid delegeeritud suured õigused, antud laialdased volitused informatsiooni saamiseks ja säilitamiseks.

Tegelikult ei jõutunud see 1834. a. seadlus mitte kõikides kubermangudes, sealhulgas ka Eestimaal. Põhjuseks oli balti aadlike vastuseis Vene keskvalitsuse korraldustele, viited Balti erikorrale. Seetõttu ei saa seda aastat pidada meie statistika algusajaks.

1853. a. lõpul tuli Vene keskvalitsus statistikakomiteede küsimuse juurde tagasi. Siseministeeriumi Statistikaakomitee tsirkulaaris nr. 190 30. detsembrist 1853 "Kubermangukomitee tegevemise vajalikkust suunast" anti suunised, mille kohaselt tuli korraldada statistikakomiteede tegevust Balti kubermangudes. Kindraladjutant Bibikov nõudis tsaaari nimel kätu vastuvaidlematu tähtsust kõikjal. Balti kindralkubermeri tsirkulaaris nr. 115 18. jaanuarist 1854 tehti Eestimaa, Liivimaa ja Kurlamaa kubernerile kohutuseks käivitada kiiresti statistikakomiteede tegevus.

Pidades silmas, et dokumendid hakkasid laekuma statistikaga seotud fondi alates 1854. aastast ja Balti kindralkubermeri tsirkulaari, loevad arhiivspetsialistid 18. jaanuari 1854 statistikaakomitee sünnitähiseks.

18. jaanuaril k. a. võisime tähistada Eesti statistika 140. aastapäeva. Nimetatud datum on kõige arvestatavam meie ametliku riikliku statistika tekkeloos.

(Järg infolehes nr. 6)

Leidur Rannamets

Tolmetaja: Raivo Rohla
Kujundaja: Kaja Põder

Tolmetuse tel. 45 04 29



AMETLIKU STATISTIKA PÕHIPRINTSIIBID EUROOPA MAJANDUSKOMISJONI PIIRKONNAS

Vastu võetud ÜRO Euroopa Majanduskomisjoni
47. istungil 15. aprillil 1992 Genfis, Rahvastepalees

Euroopa Majanduskomisjon,

pidades meeles, et ametlikule statistilisele informatsioonile tugineb olulisel määral majanduse, demograafilise ja sotsiaalsfääri ning keskkonna areng, samuti teabe ja kauba vahetus piirkonna riikide ja rahvaste vahel,

pidades meeles, et üldsus usaldab ametlikku statistilist informatsiooni niivõrd, kui võrd ta austab neid põhiväärtusi ja põhimõtteid, millele toetub iga demokraatlik ühiskond, mis püüdleb enesemõistmisele ja oma liikmete õiguste austamisele,

pidades meeles, et ametliku statistika kvaliteet ja seega valitsusele, majandusele ja üldsusele kättesaadava informatsiooni kvaliteet sõltub suuresti nende kodanike, ettevõtete jt. koostööst, kes annavad statistilisi algandmeid,

tuletades meelde sellekohaseid üldsäiteid ja normatiive, mis sisalduvad Euroopa inimõigustealases konventsioonis, Euroopa Nõukogu konventsioonis 28. jaanuarist 1991 "Üksikisikute kaitseks seoses isiklike andmete automaatotlusega", Euroopa julgeoleku ja koostöö nõupidamise Helsingi lõppaktis, Euroopa majanduskoostöö Bonni nõupidamise lõppdeklaratsioonis ning Uue Euroopa Pariisi hartas,

tuletades meelde statistikas tegevate valitsus- ja mittevalitsusorganisatsioonide jõupingutusi riikidevahelisi võrdlusi võimaldavate normatiivide ja kontseptsioonide sisseseadmiseks, tuletades meelde ka Rahvusvahelise Statistika Instituudi deklaratsiooni kutse-eesikast, olles teadlik, et Euroopa statistikutel konverents leidis vajaliku olevat kujundada printsiibid, millest oma tegevuses juhenduksid kõnealuse piirkonna ning iga üksiku liikmesriigi ametlikud statistikaorganid,

võtab vastu käesoleva resolutsiooni

I. Ametlik statistika on demokraatliku ühiskonna infosüsteemi vajalik osa, mis varustab valitsust, majandust ja üldsust andmetega majandusliku, demograafilise, sotsiaalse ja keskkonna olukorra kohta. Et oleks tagatud kodanike õigus saada avalikku informatsiooni, peab ametlik statistika vastama praktilisele kasutatavusele ja erapooletuse nõudele.

II. Et säilitada usaldust ametliku statistika vastu, tuleb statistikaorganitel professionaalselt, teaduslikult ja kutse-eesikast arvestavalt välja töötada statistiliste algandmete kogumise, töötlemise, säilitamise ja esitamise meetodika ja menetlused.

III. Et aidata andmeid korrektselt tõlgendada, tuleb statistikaorganitel esitada teaduslike normatiivide kohast informatsiooni statistika allikate, meetodika ja menetluste kohta.

IV. Statistikaorganitel on õigus kommenteerida statistika eksslikku tõlgendust ja väärkasutust.

V. Statistilistel eesmärkidel võib algandmeid ammutada kõikidest allikatüüpidest, olgu need siis statistilised vaatlused või administratiivaruanded. Statistikaorganitel tuleb allikas valida, pidades silmas kvaliteeti, ajakulu, maksumust ja andmeandjate koormust.

VI. Statistikaorganite poolt koondite jaoks kogutud isiklikud andmed, olgu need siis seotud era- või juriidiliste isikutega, peavad olema rangelt konfidentsiaalsed ja neid võib kasutada ainult statistilistel eesmärkidel.

VII. Seadused, määrused ja meetmed, mille alusel statistikasüsteemid toimivad, tuleb avalikustada.

VIII. Oluline on iga riigi siseselt statistikaorganite tööd koordineerida, et saavutada statistikasüsteemi kooskõlastatus ja efektiivsus.

IX. Rahvusvaheliste mõistete, klassifikatsioonide ja meetodikate kasutamine iga riigi statistikaorganite töös aitab jõuda kõikide ametlike tasandite statistikasüsteemide töö kooskõlastatuse ja efektiivsuseni.

X. Kahe- ja mitmepoolne statistikaalne koostöö annab panuse ametlike statistikasüsteemide täiustamisse kõikides riikides.

ÜLEVAADE 1993. A. TEHTUD TÖÖST JA 1994. A. TEGEVUSKAVAST

1993. a. oli Riigi Statistikaametile reorganiseerimise aasta nii sisuliselt kui ka organisatsiooniliselt. 1992. a. sügisel käivitunud ja 1993. a. täie hoo saanud EFTA/EC ja Balti riikide statistikaalane koostööprogramm kujunes metoodika ümberkorraldamise ja sellega kaasneva kaadriümberõppe aluseks. Et Eestis ei ole statistikaalast erikoolitust, sai sellest programmist, mis 1993. a. haaras 15 statistikaalast, ESA töötajate koolituse peamine kanal. 1993. a. oktoobris PHARE-programmi toetusel saadud tehnikaabi (64 personaalarvutit, 3 kohtvõrku ja keskarvuti, eksploatatsioonimaterjalid üheks aastaks) lõi võimaluse EFTA-programmi raames omandatud metoodiliste teadmiste rakendamiseks.

Milliseid 1993. a. tulemusi tahaks eraldi rõhutada?

Kõigepealt muidugi keskasutuse reorganiseerimine, mille põhieesmärgiks oli organisatsioonilise struktuuri vastavusse viimine täidetavate funktsioonidega. Ma arvan, et kõigi ametikohtade komplekteerimine konkursi alusel oli õige otsus ja me saime sellega päris edukalt hakkama. Reorganiseerimise üheks osaks oli tööde ümberjagamine sektorite ja teenistuste vahel vastavalt uuele tööjaotusele ja struktuurile.

Eelmisel aastal alustasime ettevalmistustega maakondlike statistikabüroode reorganiseerimiseks piirkondlikeks statistikabüroodeks. Sellel aastal tuleb kavandatu teoks teha.

Eelmisel aastal algas üleminek valikvaatlustele, kusjuures ettevõtteregistri baasil moodustati statistiline profiil. Vaatamata raskustele teooria omandamisel ja vigadele koostöös Ettevõtteregistriga jõudsimeni ka siin päris heade tulemusteni, igatahes saime suure hulga kogemusi.

Väga tähtsaks tuleb pidada informaatikaalase töö ületoomist Peaarvutuskeskusest Statistikaameti teenistustesse. Suure töö tegid ära meie informaatikud PHARE-programmi toetusel saadud arvutite ja litsentseeritud tarkvara evitamisel.

Tubli oli meie statistika levi teenistus statistikaalaste väljaannete ajakohastamisel. Üleminek tähtsamate väljaannetes paralleelselt eesti ja inglise keele kasutamisele tegi info kättesaadavaks paljudele uutele tarbijatele.

Kindlasti tuleb ära märkida raamatukogu avamine ja selle lausa üllatavalt edukas tegutsemine, samuti info levitamise uute võimaluste (infotelefon) otsimine ja leidmine.

Peale ülalootetu oli 1993. a. veel terve rida häid tulemusi ja toredaid kordaminekuid, kuid paljud protsessid alles algasid ja paljud probleemid jäid lahenduseta, õigemini tuleb lahendada sel aastal.

Millised on Statistikaameti 1994. a. tegevuskava põhipunktid?

Tähtsamaks tööks kujuneb Eesti statistikasteemi pikaajalise arengukava väljatöötamine. Seejuures tuleb määrata reaalsusel põhinevad prioriteetsed suunad lühiajalises arenguprogrammis.

Tõsine töö tuleb ära teha statistilises aruandluses kasutusel oleva majandusterminoloogia ja definitsioonide kooskõlla viimisel teistes riikides kasutusel olevate mõistetega, samuti rahvusvaheliste klassifikatsioonisteede rakendamisel vastavalt nende tõlgete valmimisele ja kooskõlastamisele teiste ametkondadega.

Ees ootab statistikaalase riigitellimuse rakendamine ja kooskõlastamine ministereeriumide ja ametkondadega.

Alanud aastal tuleb tagada täielik üleminek valikvaatlusmetoodikale, mille üheks eelduseks on statistilise profiili koostamise tehnoloogia täiustamine.

- Tähtsamate tööde hulka kuuluvad kindlasti ka
- statistikaseaduse ja ESA põhimääruse uue versiooni koostamine;
 - statistikanõukogu ja ESA teenistuste ekspertgruppide moodustamine ja nende töö käivitamine;
 - piirkondlike statistikabüroode tööerakendamine;
 - PHARE-programmi 1993. ja 1994. a. eelarvest finantseeritava abi korras saabuva tehnika ja tarkvara rakendamine keskasutuses ja piirkondlikes büroodes.

Ülaltoodu ei ole loomulikult täielik loetelu lahendust nõudvatest probleemidest. Igal teenistusel ja sektoril on terve rida spetsiifilisi ülesandeid. Arvan, et kui me suudame arendada koostöötahtet nii maja sees kui majast välja, saame kõigea hakkama. Edu teile!

Rein Veetõusme
ESA peadirektor

Tartu Ülikooli Kirjastuse trükikoda
Tiigi 78, EE2400 Tartu
Tellimus nr. 220. Trükiarv 200.