



## SISUKORD

Saateks .....	3
Seminar "Statistika eluteadustes" .....	4
E.-M. Tiit. Rahvastiku prognoosist .....	5
A.-M. Parring ja M. Vähi. Korrelatsioonimaatriksi ohtlikud olulisuse tõenäosused .....	24
E. Käärrik. Kordusmõõtmistest .....	33
V. Reeben. Looduslike rütmide lainete kuju kirjeldamisest .....	39
R. Birkenfeldt. Krooniliste mittespetsiifiliste haiguste riskitegurite uurimise metoodikast .....	45
H. Kaarma, L. Saluste, G. Veldre, M. Thetloff. Mitmemõõtmeline statis- tiline analüüs antropomeetriliste probleemide lahendamisel .....	48
K. Põlluste. Biomeetria õpetamine arstiteaduskonnas .....	52
E. Pärna. Kvantitatiivsete tunnuste variatsioon .....	60
A. Kiviste. Metsa kõrguskasvu mudelite koostamise probleeme .....	62
A. Läänelaid. Search for a standardization method to enhance the common signal in tree ring series .....	67
Eesti Statistikaltsi põhikiri .....	81
Konverentsid .....	85

## SAATEKS

Käesolevaga jõuab lugejani Eesti Statistikeseltsi V Teabevihik. Kogumikku on koondatud 1994. aastal toimunud statistikeseltsi seminari ettekannete materjalid, mis käsitlevad statistikaprobleeme valdkonnas, mida kokkuvõtlikult võib nimetada eluteadusteks: meditsiinis, bioloogias, rahvastikuteaduses. Kokku said erinevate teaduste esindajad, kes samadest mõistetest rääkimisel kasutavad tihti erinevaid termineid. Sellisena etendas seminar just seda Eesti statistikuid ühendavat rolli, mis on põhikirjaliselt seltsi peamisi ülesandeid. Samal ajal andis seminar osavõtjatele hea võimaluse enda ja oma valdkonnas tehtava võrdlemiseks naaberaladega. Esinejate hajutatuse mõõda erinevaid asutusi on kindlasti üks põhjus, miks kogumik ilmub alles nüüd. Koostajad vabandavad väljaande hilinemise pärast lootes lugejate heasoovlikkusele.

Kogumiku lõpust leiavad seltsi liikmed uue kinnitatud põhikirja teksti ja lühikese konverentsiinfo.

Eesti Statistikeseltsi juhatus

## SEMINAR "STATISTIKA ELUTEADUSTES"

24. märtsil 1994

Tartu Ülikooli Füüsikahoones Tähe 4, ruum 160

### Programm

- 11<sup>00</sup>-11<sup>15</sup> Seminari avamine  
11<sup>15</sup>-11<sup>35</sup> Tõnu Möls. Biomeetria Eestis.  
11<sup>35</sup>-12<sup>10</sup> Ene-Margit Tiit. Rahvastiku prognoosimisest.  
12<sup>10</sup>-12<sup>30</sup> Anne-Mai Parring ja Mare Vähi. Korrelatsioonimaatriksi ohtlikud olulisustõenäosused.

### *Kohvipaus*

- 12<sup>50</sup>-13<sup>10</sup> Ene Käärik. Kordusmõõtmistest.  
13<sup>10</sup>-13<sup>30</sup> Vello Reeben. Looduslike rütmide lainete kuju adekvaatsest kirjeldamisest.  
13<sup>30</sup>-13<sup>50</sup> Arne Ader. Läbirändelainete otsimisest lindudel.  
13<sup>50</sup>-14<sup>10</sup> Peeter Hõrak. Logit-mudelite kasutamisest.

### *Vaheaeg*

- 15<sup>00</sup>-15<sup>20</sup> Reinhold Birkenfeldt. Krooniliste mittespetsiifiliste haiguste riskitegurite uurimise meetodikast.  
15<sup>20</sup>-15<sup>35</sup> Helje Kaarma. Naise kehaehituse antropoloogilis-statistiline uurimine.  
15<sup>35</sup>-16<sup>00</sup> Kaja Põlluste. Biomeetria õpetamisest arstiteaduskonnas.

### *Kohvipaus*

- 16<sup>15</sup>-16<sup>40</sup> Elli Pärna. Kvantitatiivsete tunnuste variatsioon.  
16<sup>40</sup>-17<sup>00</sup> Andres Kiviste. Statistilised probleemid metsanduses.  
17<sup>00</sup>-17<sup>20</sup> Alar Läänelaid. Aastarõngaseeriade standardiseerimismeetodite otsing.  
17<sup>20</sup>-17<sup>40</sup> Jaak Truu. Populatsiooni parameetrite hindamine mikro-bioloogias.  
17<sup>40</sup>-18<sup>00</sup> Tõnu Möls. Dispersioonanalüüs ökoloogias.

## RAHVASTIKU PROGNOOSIST

E.-M. Tiit

Tartu Ülikool, Matemaatilise Statistika Instituut

**1. Sissejuhatus. Rahvastikuproгноoside ajaloost.** Minge riigi või piirkonna rahvastiku arvukuse ja soolis-vanuselise struktuuri muutuste ettenägemine järgnevateks aastateks või isegi aastakümneteks on suure majandusliku ja poliitilise tähtsusega. Sellepärast kuuluvadki rahvastiku- (demograafilised) ennustused rakeduslikust seisukohast olulisimate statistiliste progноoside hulka. Eesti puhul lisandub rahvastikuproгноosi tavalisele praktilisele väärtusele veel valuline tunnetuslik küsimus – kas me ikka jääme püsima (enamusrahvusena omal põlisel kodumaal)?

On huvitav, et rahva arvukust on püütud progноosida juba ammu, ning võrreldes eelmisel sajandil antud progноose tegelikkusega võib tõdeda, et valdaval enamusel juhtudest on rahvastiku arvukust (selle kasvukiirust) tugevasti alahinnatud, kusjuures ootuspäraselt on viga seda suurem, mida pikema aja peale on tehtud progноos. Viimasel ajal on progноosi täpsus siiski oluliselt paranenud, vt. Näide 1.

**Näide 1.** Rootsi on üks riike, kus on rahvastiku uurimise ja progноosimisega tõsiselt tegeldud juba pikka aega. Tabelis 1 on esitatud Rootsi Statistika Keskbüroo poolt 1960-ndate aastate alguses avaldatud progноos (vt.[7]) ja see kõrvutatud tegelike rahvastikuandmetega (vt.[8]). On näha, et mitmekümne aasta peale ette ennustades on suudetud saavutada selline täpsus, et progноosiviga ei ületa viit protsenti.

Tabel 1. Rootsi rahvastikuprognosis ja tegelik rahvaarv (tuhandetes).

Aasta	1960	1965	1970	1975	1980	1985	1990
Prognoos		7741	8064	8392	8703	8720	8892
Tegelik	7498		8081	8208	8318	8358	8590
Prognoosiviga			-0.2%	2.2%	4.6%	4.3%	3.5%

Järgnevas vaatlemegi rahvastiku prognoosimise meetodeid.

**2. Lineaarne mudel rahvastiku prognoosimiseks.** Kõige lihtsamad prognoosid käsitlevad rahvastiku arvukust (uuritavas piirkonnas) kui aegriidiga  $N(t)$ , kus  $t$  näitab aega (aastates), mis on möödunud vaatluste algushetkest. Suhteliselt lühikese aja jooksul näib iga-aastane juurdekasv ehk iive  $\delta_t = N(t + 1) - N(t)$  olevat ligikaudu konstantne ning siit tulenebki lihtsaim, *lineaarne prognoos*:

$$N(t + k) = N(t) + k\delta, \quad (1)$$

kus  $\delta$  on teatava ajavahemiku põhjal statistiliselt hinnatud aastane iive. Mõnikord annab selline prognoos lühemate ajaperioodide puhul täiesti rahuldavaid tulemusi.

**Näide 2.** Eesti rahvaarv (s.o. Eesti paikse elanikkonna arv) aastail 1970 – 1994 on esitatud tabelis 2 (vt.[1],[3]). See rahvaarv on arvutatud (hinnatud) iga aasta 1. jaanuariks, kasutades selleks rahvaloenduste andmeid ja jooksvat rahvastikusündmuste (sünd, surm, ränne) statistikat.

Kuni 1990-nda aastani tundub iive olevat aastate lõikes üsna ühesugune. See tekitabki mõtte kasutada lineaarset prognoosi. Tavalise *lineaarse regressiooni* teel saame aastate 1970 – 1990 andmeid kasutades võrrandi

$$N(1970 + t) = 1369 + 10.08t,$$

mis kirjeldab vaadeldaval ajavahemikul Eesti rahvastiku kasvu väga hästi, vt. joonis 1. Suurim erinevus teoreetilise ja arvutatud väärtuse vahel on vaid üks protsent. On aga selge, et kasutades sama valemit tulevikuprognosideks võivad tekkida hoopis suuremad vead. Illustreerimaks seda tõsiasja on joonisele kantud ka punktid, mis iseloomustavad rahvaarvu aastail 1991 – 1994.

Tabel 2. Eesti rahvaarv ja selle muutumine (tuhandetes).

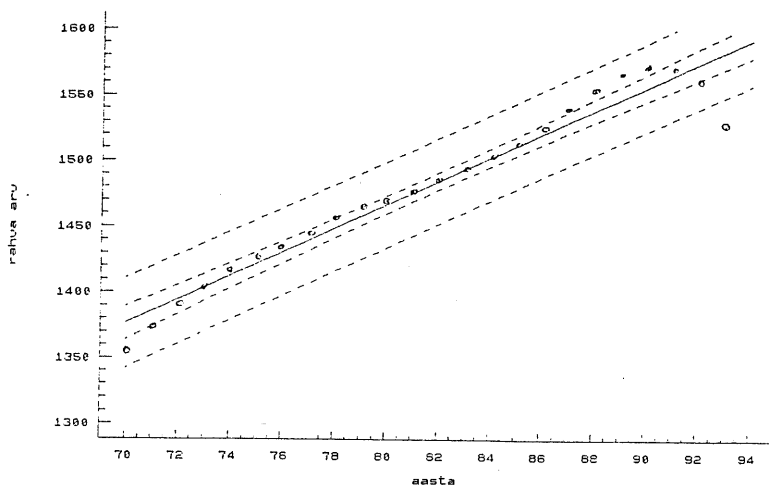
Aasta	Rahvaarv	Iive
1970	1355	
1971	1372	17
1972	1389	17
1973	1403	14
1974	1416	13
1975	1426	10
1976	1435	9
1977	1444	9
1978	1456	11
1979	1464	8
1980	1469	5
1981	1477	8
1982	1485	8
1983	1494	9
1984	1503	9
1985	1513	10
1986	1525	10
1987	1539	14
1988	1554	15
1989	1565	11
1990	1572	7
1991	1570	-2
1992	1562	-8
1993	1526	-36
1994	1507	-19

### 3. Eksponentsiaalne mudel rahvastiku prognoosimiseks.

Eelmises punktis kirjeldatud lineaarse mudeli sobivuse Eesti rahvastiku arvukuse prognoosimiseks tagas tõsiasi, et rahvaarvu muutusi põhjustas neil aastail eeskätt sisseränne, mis püsis ligilähedaselt konstantsena. Kui aga rahvastiku arvukuse muutusi põhjustavad sünni- ja surmaprotsessid, nn. *loomulik muutumine*, siis on otstarbekas rahvastiku prognoosimisel kasutada mõnevõrra keerukamat mudelit, mille aluseks on alljärgnev lihtne arutelu:

Oletame, et vaadeldava rahvastikurühma puhul iga  $b$ -s inimene sünnitab aastas lapse ja iga  $d$ -s sureb aasta jooksul, st, et juhuslikult valitud elaniku puhul on lähema aasta jooksul *sünnitamise*

tõenäosus  $\frac{1}{b}$  ja suremise tõenäosus on  $\frac{1}{d}$ . Olgu lähtehetkel  $t = 0$



Joonis 1.

Eesti rahva arvukus ja selle lineaarne regressioon 1970-1994.

(s.o. algusaasta 1. jaanuaril) rahvaarvu  $N(0)$ . Sel juhul avaldub aastane rahvastiku iive  $\delta(N)$  rahvaarvu  $N(0)$  kaudu järgmise valemi abil

$$\delta(N) = N(0) \left( \frac{1}{b} - \frac{1}{d} \right) = N(0) \frac{d-b}{bd}. \quad (2)$$

Rahvaarvu prognoosimiseks  $t$  aasta pärast omandab valem (2) alljärgneva kuju

$$N(t) = N(0)(1 + \alpha)^t, \quad (3)$$

kus  $\alpha = \left( \frac{1}{b} - \frac{1}{d} \right)$ .

Valemit (2) võime edasi arendada ka pideva ajaskaala jaoks, tähistades sümboliga  $\delta t$  ajavahemiku, mille jooksul rahvastik kasvab suuruse  $\delta N$  võrra. Seos (2) on esitatav diferentsvõrrandina

$$\frac{\delta N}{\delta t} = N(0)\alpha,$$

mis diferentside lähenemisel nullile annab diferentsiaalvõrrandi

$$\frac{dN}{N(0)} = \alpha dt$$

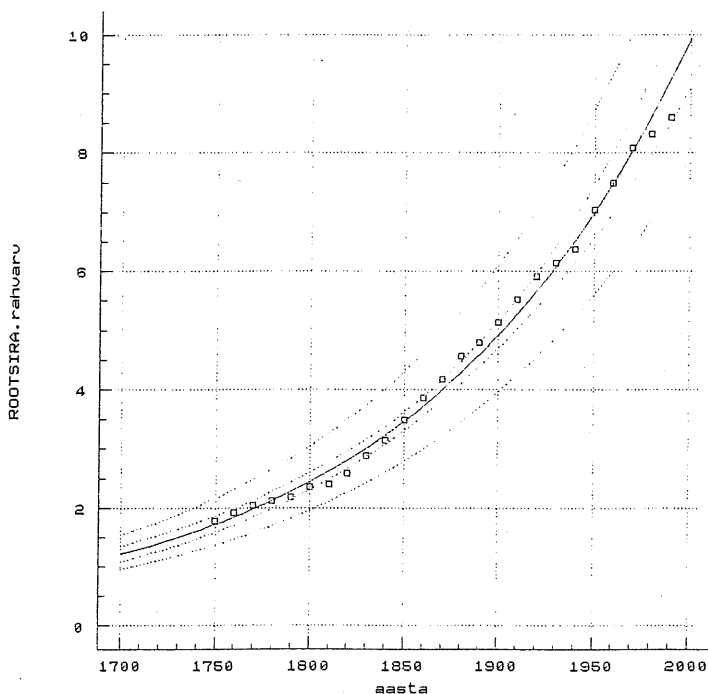


lahendiga

$$N(t) = N(0)e^{t\alpha}. \quad (4)$$

Valem (4) sõltub kahest parameetrist – rahvaarvust  $N(0)$  mingil lähtemomendil, mis enamasti eeldatakse teadaolevaks, ja kasvukiirusest  $\alpha$ . Viimast on võimalik statistiliselt hinnata, kui on olemas küllalt pikk rida vaatlusi.

**Näide 3.** Vaatleme rahvaarvu muutumist Rootsis aastail 1750 – 1950 (vt. joonis 2). Rahvaarvu dünaamikat selle aja jooksul



Joonis 2.

Rootsi rahvastiku dünaamika aastatel 1750 - 1950 ja seda lähendav eksponentfunktsioon  $N(t) = 1700e^{0.007164t}$ .

kirjeldab hästi eksponentsiaalne mudel (3), milles hetkeks 0 on aasta 1750 ja parameetritele saame andmete põhjal järgmised hinnangud

$$N(0) = 1700, \quad \alpha = 0.007164.$$

On näha, et saadud mudel kirjeldab rahvaarvu dünaamikat üsna hästi, ning aastaks 1960 saadud rahvastikuprognosis on üsna lähedane selle suuruse tegelikule väärtusele. Valemiga (4) esitatud mudelit tuntakse *Lotka mudelina* (vt. [6]), ning see leidis kasutamist peaaegselt käesoleva sajandi kolmekümnendatel kuni viiekümnendatel aastatel.

Loomulikult on käesolevat mudelit võimalik modifitseerida, lisades juurde täiendavaid parameetreid (näiteks võtta eksponendi argumentiks mitte lineaarne, vaid ruutfunktsioon jne). Niisugusel täpsustamisel on aga loomulikud piirid, sest prognoosimiseks kasutatava teabe hulk on üpriski piiratud – ainus lähtesuurus, mida kasutatakse, on rahvaarv möödaläinud aastatel. Kindlasti aitaks prognoosi täpsustada detailsema lähteinfo kasutamine.

**4. Rahvaarvu mõjustavad protsessid.** Tähistame, nagu varemgi, tähega  $t$  aasta numbrit (kus  $t = 0$  on vaatluste alghetk). Rahvaarvu muutus ajavahemiku  $t$  jooksul sõltub järgmistest näitajatest:

- 1) Sündide arv  $B(t)$  antud ajavahemiku jooksul;
- 2) Surmade arv  $D(t)$  antud ajavahemiku vältel;
- 3) Sisserännanute arv  $I(t)$  antud ajavahemiku vältel;
- 4) Väljarännanute arv  $E(t)$  antud ajavahemiku vältel.

Rahvaarv  $N(t)$  ajahetkel  $t$  avaldub siis alljärgneva valemiga

$$N(t) = N(0) + B(t) - D(t) + I(t) - E(t). \quad (5)$$

Iive  $N(t) - N(0)$  avaldub loomuliku iibe  $B(t) - D(t)$  ja mehhaanilise iibe ehk rändesaldo  $I(t) - E(t)$  summana. Mõnikord (eriti populaarses ja esseistlikus kirjanduses) mõistetakse iibena ka üksnes loomulikku iivet.

Valemi (5) kasutamiseks rahvastiku prognoosimisel tuleb hinnata suurusi  $B(t)$ ,  $D(t)$ ,  $I(t)$  ja  $E(t)$  vaadeldava aasta jaoks. Pike- ja ajajaliste prognooside puhul on tarvis saada nende suuruste jaoks ka tulevikuhinnanguid.

Sisse- ja väljarännu maht sõltub mitmesugustest majanduslikest ja poliitilistest teguritest ning see on üldiselt küllaltki raskesti prognoositav. Seevastu aga sündide ja surmade arvu on võimalik rohkem või vähem täpselt ette näha, kui on teada elanikkonna soolis-vanuseline koostis. Tõepoolest, sünnitavad ainult suhteliselt noored, nn. *sünnitusealised* naised, ning nende suhtarvust elanikkonnas sõltub olulisel määral oodatav sünnituste arv. Vanurite suhteliselt suur arv aga lubab ennustada kõrget surmajuhtude arvu.

**5. Rahvastikupüramiid.** Edaspidi nimetame uuritavat inimeste hulka (mingi riigi või piirkonna elanike arvu või uurijat huvitavat osa sellest) *populatsiooniks*. Tähistagu  $N(t, x)$   $x$ -aastaste inimeste arvu uuritavas populatsioonis aastal  $t$ . Kui tähistada tähega  $\omega$  *maksimaalne vanus* antud populatsioonis (näiteks 110 aastat), siis saame ilmse valemi rahvaarvu leidmiseks

$$N(t) = \sum_{x=0}^{\omega} N(t, x). \quad (6)$$

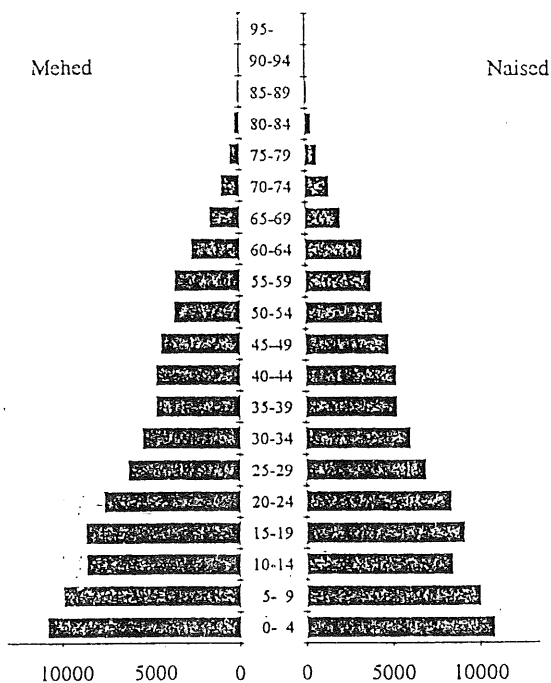
Loomulikult jääb seos (6) kehtima ka siis, kui inimesed on jaotatud näiteks viieaastastesse vanuserühmadesse,  $N(t, x)$  tähistab  $x$ -ndat vanuserühma ning  $\omega$  vanima vanuserühma järjekorranumbrit.

Veelgi detailsema rahvastiku kirjelduse tema *soolis-vanuselise struktuuri mõttes* saame, kui loendame eraldi  $x$ -vanuste meeste arvu (olgu see  $M(t, x)$ ) ja sama vanade naiste arvu  $W(t, x)$  aastal  $t$ . Loomulikult annab nende summa  $x$ -nda vanuserühma,

$$N(t, x) = M(t, x) + W(t, x).$$

Traditsiooniliselt kujutatakse rahvastiku soolis-vanuselist struktuuri *rahvastikupüramiidina*, s.o. graafikuna, milles igale vanuserühmale (vanusele)  $x$  vastab horisontaalne tulp pikkusega  $N(t, x)$ , s.o. nimetatud vanuserühma arvukus (sobivalt valitud skaalal). Graafiku telg on vertikaalne, teljest vasakul on tavaliselt meestele ja paremal naistele vastavad tulbad pikkustega  $M(t, x)$  ja  $W(t, x)$  vastavalt. Alumine tulp vastab noorimale, vanusega 0 ja ülemine – vanimale, vanusega  $\omega$  vanuserühmale. Vanuserühmade suuruseks on kas viis aastat (vt. Joonis 3, lk. 12) või üks aasta (vt. Joonis 4, lk. 13).

Kui rahvastiku areng on kulgenud ühtlaselt, ilma sõdade, näljahädade, kriiside ja epideemiateta, siis on rahvastiku vanuseline koostis üpris korrapärane, kusjuures nooremad põlvkonnad on arvukamad kui vanemad, mille kallal surm on oma hävitustööd pikema aja jooksul teinud. Niisugusele rahvastiku struktuurile vastab korrapärane rahvastikupüramiid, mille astmed alt ülespoole tõustes korrapäraselt vähenevad. Sellist pilti näemegi joonisel 3, millel on kujutatud Eesti elanikkond aastal 1881. Sel ajal oli varasematest hävitavatest sõdadest, katkudest ja näljahädadest möödunud piisavalt aastakümneid, et nende haavad kustutada. Niisugune rahvastikupüramiid on matemaatiliselt üsna lihtsalt kirjeldatav mõne parameetri abil (vt. näit. valemit (4)). Arusaadavalt on sellise rahvastiku koostise puhul demograafiline prognoosimine suhteliselt



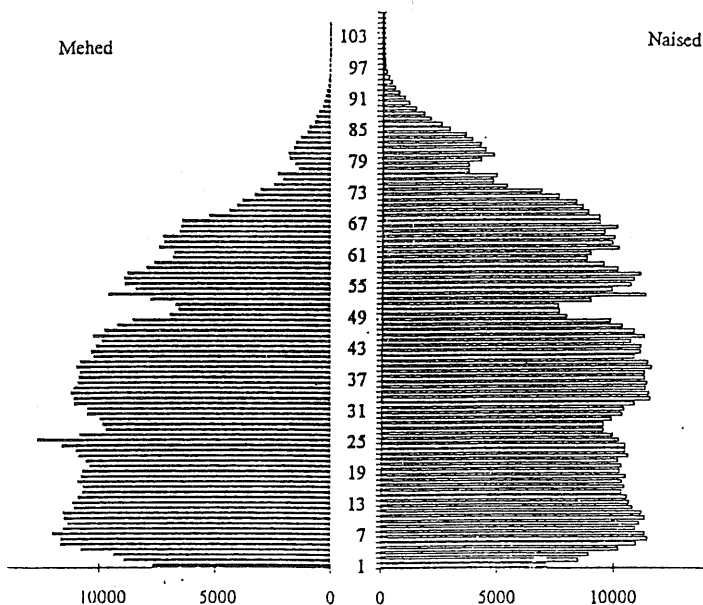
Joonis 3.

Rahvastiku vanuspüramiid 1881.a.

lihtne ja suure tõenäosusega ka üsna täpne – senikaua, kui ei toimu mingit sotsiaalmajanduslikku protsessi, mis rahvastiku senist arengut radikaalselt muudab.

Kahjuks ei ole aga reaalne elu enamasti nii korrapärane, seda mõjustavad sõjad, haigused, ebasoodsad ilmastikuolud ja poliitilised kataklüsmid. Sellepärast ongi kaasajale hoopis iseloomulikumad märksa ebakorrapärasema kujuga rahvastikupüramiidid.

Näiteks selle kohta on joonis 4, kus on kujutatud Eesti rahvastik aastal 1993. Eriti silmapaistev on viimasel püramiidil nooremate põlvkondade (püramiidi aluse) suhteline väiksus keskmiste põlvkondadega võrreldes. Selgesti paistab sellelt püramiidilt silma ka II Maailmasõja mõju, mistõttu 49 – 53-aastaste vanuserühmad on naaberrühmadest märksa väiksemad. II Maailmasõja järelmõjuga, nn. *demograafilise lainega* saab kirjeldada 27 – 28-aastaste väiksemat arvukust. 75-aastaste kohal on täheldatav I Maailmasõja mõju, ning hästi paistab silma ka eakate meeste rühmade suhteline väiksus naistega võrreldes. Umbes 25-aastaste meeste eriti arvukas rühm on



Joonis 4.  
Eesti rahvastik 1993.a.

seletatav rahvastiku arvestuse alase segadusega Nõukogude armee lahkumisel. On võimalik, et tegemist on peale demobiliseerimist Eestisse jäänud noormeestega.

**6. Sündimus, suremus ja loomulik iive.** Sündide arv  $B(t)$  mingis populatsioonis aastal  $t$  sõltub oluliselt rahvaarvust  $N(t)$  selles populatsioonis vaadeldaval aastal. Et võrrelda sündide arvukusi erinevates populatsioonides, võetakse kasutusele mõiste *sündimus*, s.o. sündide arv aasta jooksul 1000 elaniku kohta. Tähistame sündimuse aastal  $t$  tähega  $b(t)$ :

$$b(t) = \frac{1000B(t)}{N(t)}.$$

Samal viisil määratletakse ka *suremus*, s.o. surmajuhtumite arv aastal  $t$  tuhande elaniku kohta  $d(t)$ :

$$d(t) = \frac{1000D(t)}{N(t)}.$$

Sündimuse ja suremuse vahe moodustab loomuliku iibe kordaja  $c(t)$  (aastal  $t$ ),

$$c(t) = b(t) - d(t).$$

See kordaja näitab rahvaarvu loomulikku muutumist 1000 inimese kohta vaadeldaval aastal  $t$ . Loomuliku iibe kordaja märk näitab seda, kas vaadeldaval aastal rahva arv suurenes või vähenes loomuliku iibe tagajärjel.

Erijuhul, kui sündimus  $b(t)$  ja suremus  $d(t)$  jäävad konstantseks,  $b(t) = \beta$ ,  $d(t) = \theta$ , saame rahvastikku prognoosida valemi (2) abil, kus  $b = \frac{1}{\beta}$  ja  $d = \frac{1}{\theta}$ , seega

$$\alpha = \beta - \theta$$

ja  $t$  näitab alghetkest 0 möödunud aastate arvu.

Et saada täpsemat ennustust, tuleb arvestada, et niihästi sündimus kui ka suremus sõltuvad ühelt poolt erinevate vanuserühmade osakaalust elanikkonnas, teiselt poolt aga ka sünitamise ja suremise *sagedusest* (intensiivsusest) erinevates vanuserühmades. Need näitajad muutuvad sõltuvalt mitmesugustest ühiskonna arengut mõjuvatest teguritest.

**7. Elukestusfunktsioon.** Vaatleme inimese eluiga juhusliku suuruse  $X$ . Selle juhusliku suuruse jaotusfunktsioon  $F(x)$  väljendab tõenäosust selleks, et (juhuslikult valitud) vastsündinu eluiga ei ole suurem kui  $x$ , st ta sureb hiljemalt vanuses  $x$ ,

$$P(X \leq x) = F(x).$$

Paneme tähele, et rahvastikustatistikas kasutatakse (erinevalt tõenäosusteooria tavast) mitteranget võrratust jaotusfunktsiooni definitsioonis.

Inimese seisukohalt hoopiski optimistlikum on elukestusfunktsioon  $S(x) = 1 - F(x)$ , mis väljendab tõenäosust selleks, et vastsündinu elab üle vanuse  $x$ ,

$$S(x) = P(X > x).$$

Tõenäosust, et inimese elu katkeb mingis konkreetsetes vanuses  $x$ , saab leida niihästi eluea jaotusfunktsiooni kui ka elukestusfunktsiooni kaudu:

$$P(x < X \leq x + 1) = F(x + 1) - F(x) = S(x) - S(x + 1).$$

Kõik ülalkirjeldatud tõenäosused on arvutatud vastsündinu kohta. Kui aga on teada, et inimene on juba teatava vanuseni elanud, saab edasise elukestuse tõenäosuse avaldamiseks kasutada *tinglikku tõenäosust*. Nii on  $x$ -aastaseks elanud inimese puhul tõenäosus  $q_x$  surra lähima aasta jooksul, st hiljemalt oma  $x + 1$ -sel sünnipäeval, alljärgnev:

$$q_x = P(x < X \leq x + 1 \mid X > x) = \frac{S(x) - S(x + 1)}{S(x)} = 1 - \frac{S(x + 1)}{S(x)}.$$

Saadud valemist ilmneb ühtlasi ka see, et tõenäosus  $p_x$ , et  $x$ -aastane inimene elab üle oma  $x + 1$  sünnipäeva, nn *üleelamistõenäosus*

$$p_x = 1 - q_x \quad (7)$$

avaldub kujul

$$p_x = \frac{S(x + 1)}{S(x)}.$$

Selleks, et saadud tulemuste põhjal rakenduslikke järeldusi teha, on tarvis leida eeskiri valemities esinevate suuruste hindamiseks empiirilise andmestiku põhjal.

Tähistame sümboliga  $D(t, x)$   $x$ -aastaste inimeste surmajuhude arvu aastal  $t$ . Siis suhe

$$\tilde{q}_x = \frac{D(t, x)}{N(t, x)} \quad (8)$$

ongi hinnanguks suremistõenäosusele  $q_x$  aasta  $t$  andmete põhjal. Loomulikult on siis lihtne leida valemist (7) hinnangut ellujäämise tõenäosusele  $p_x$ .

Sündmuse  $[X > x]$ , st et inimene elab üle oma  $x$ -nda sünnipäeva tõenäosuse  $S(x)$  võime avaldada järgmise korrutisena

$$S(x) = S(1) \frac{S(2)}{S(1)} \times \dots \times \frac{S(x)}{S(x-1)} = \prod_{i=1}^x p_i. \quad (9)$$

Kõik üleelamistõenäosused  $p_i$  on aga empiiriliste andmete põhjal hinnatavad valemite (7) ja (8) abil, seega saab nende põhjal valemis (9) leida funktsiooni  $S(x)$  kõik huvipakkuvad väärtused,  $x = 1, \dots, \omega$ .

**8. Keskmise eluiga.** Kui eeldada, et valemis (9) kasutatud üleelamistõenäosused jäävad pikaks ajaks (inimese eluajaks)

konstantseks, st valemiga (8) esitatav suhe on aastast  $t$  praktiliselt sõltumatu, siis saab nende andmete põhjal arvutada välja ka inimese keskmise (oodatava) eluea:

$$\bar{x} = \int_0^{\omega} x f(x) dx, \quad (10)$$

kus  $f(x)$  on elukestuse tihedusfunktsioon. Kasutades ositi integreerimist ( $u = x$ ,  $dv = -f(x)dx = s(x)dx$ ) ja arvestades, et  $S(\omega) = 0$ , saame oodatava eluea avaldada elukestusfunktsiooni kaudu:

$$\bar{x} = \int_0^{\omega} S(x) dx.$$

Funktsiooni  $S(x)$  diskretiseerimisel, st eeldades, et funktsioon  $S(x)$  on konstantne igas intervallis  $(x, x + 1]$ , saame integraali asendada summaga:

$$\bar{x} = \sum_0^{\omega} S(x). \quad (10')$$

Saadud valem on rakendatav praktilistes arvutustes, kui kasutada valemite (7), (8) ja (9) abil saadud elukestusfunktsiooni hinnanguid.

Kui meil on tegemist inimesega, kes on juba  $y$ -aastane, siis on mõtet kõnelda temale *veel elada jäänud keskmisest aastate arvust*, mille tähistame sümboliga  $\bar{x}_y$ . Selle suuruse leidmiseks paigutame valemisse (10) tihedusfunktsiooni  $f(x)$  asemele tingliku tihedusfunktsiooni  $f_y(x) = \frac{f(x)}{S(y)}$ , mis iseloomustab eluea kestust tingimusel, et inimene on juba  $y$ -aastaseks elanud, ning muudame vastavalt ka integreerimisrajasid. Peale ositi integreerimist ja funktsiooni diskretiseerimist saame tulemuseks valemi (10') analoogi

$$\bar{x}_y = \frac{1}{S(y)} \sum_{x=y}^{\omega} S(x), \quad (11)$$

mis näitab  $y$ -aastase inimese oodatavat elada jäänud aastate arvu.

Kuigi eeldus elukestuse näitajate konstantsuse kohta on täiesti ebareaalne, kasutatakse valemiga (10') arvutatud *keskmist eluiga* ehk *oodatavat elukestust* kui väga olulist rahvastikukarakteristikut.

Elukestus on väga oluline näitaja ka rahvastiku prognooside tegemisel. Näiteks kui mingi aja  $t$  jooksul (sündimuse ja migratsiooni konstantsuse korral) keskmine elukestus suureneb  $T$  aastalt



$T_1$  aastani, suureneb teatava aja möödudes samas proportsioonis ka elanikkonna arvukus,

$$N(t) = N(0) \cdot (T_1/T).$$

Tabelis 3 (vt. [4], [5]) esitame mõnede Euroopa riikide elanike keskmised elukestused.

Tabel 3. Mõnede Euroopa riikide elanike keskmised elukestused käesoleva sajandi 80-ndail aastail.

Riik	Mehed	Naised
Austria	69.2	76.6
Eesti	65.8	75.0
Hispaania	71.1	78.6
Iirimaa	69.5	75.0
Inglismaa	70.4	76.6
Jugoslaavia	67.7	73.2
Leedu	66.9	76.3
Luxemburg	70.0	76.7
Läti	65.3	75.2
Moldova	65.5	72.3
Norra	72.5	79.2
Poola	67.1	75.7
Portugal	65.1	72.9
Prantsusmaa	70.4	78.5
Rootsi	73.6	79.0
Saksamaa	69.0	74.8
Soome	70.1	78.5
Šveits	72.7	79.6
Taani	71.3	77.4
Tšehhoslovakkia	66.8	74.0
Ukraina	66.1	74.0
Ungari	66.6	73.6
Valgevene	66.8	76.4
Venemaa	64.2	74.5

Üldiselt on ajaloo vältel elukestused suurenenud, olles arenenud maades märksa kõrgemad kui arengumaades.

**9. Suremustabel.** Elukestusfunktsiooni ja sellega seotud suuruste käepäraseks esitusvormiks on nn. *suremustabel*, mis sisaldab

funktsioonide  $S(x)$ ,  $q_x$  ja  $p_x$  väärtuste empiirilisi hinnanguid kõigi vanuserühmade jaoks,  $x = 1, \dots, \omega$ . Tõenäosuste mugavamaks esitamiseks on nad korrutatud nn. *kohordi alusega*, milleks on tavaliselt arv 100 000, ning seega võib suurust  $l(x) = 100000S(x)$  interpreteerida kui *nende inimeste arvu, kes on veel ellu jäänud 100 000-st  $x$  aasta eest sündinud lapsest*. Arv  $D_x(l) = 100000q_x$  näitab aga nende inimeste arvu *antud kohordist*, kes on *jooksva aasta kestel surnud*.

Arusaadavalt on üleelamise tõenäosused  $p_x$  suured (ühelähe-  
dased) noorte vanuserühmade puhul ja hakkavad kahanema vanema-  
maealiste puhul. Suhteliselt väiksemad on üleelamistõenäosused  
ka alla-aastaste laste puhul. Kuivõrd meeste ja naiste puhul on  
üleelamistõenäosused üldiselt erinevad (naistel enamasti suuremad),  
siis koostatakse ka suremustabelid meeste ja naiste jaoks tavaliselt  
eraldi. Tavaliselt lisatakse suremustabelisse ka elada jäänud aastate  
arv  $\bar{x}_y$ , mille kohta kasutatakse ka tähist  $e_x$ , kus  $x$  tähistab vanust.

Tabelis 4 esitame fragmendi Eesti elanikkonna suremustabelist  
aastal 1990, mis pärineb matemaatikaüliõpilase Anne Pirni bakalau-  
reusetööst (vt. [2]).

Tabel 4. Eesti elanikkonna suremustabel aasta 1990 andmetel  
(kogu elanikkond vanuses 0 – 7 aastat).

$x$	$q_x$	$l_x$	$D_x(l)$	$p_x$	$e_x$
0	0.0131	100 000	1312	0.9869	70.52
1	0.0027	99 731	268	0.9973	69.71
2	0.0016	99 575	156	0.9984	68.82
3	0.0007	99 508	66	0.9993	67.86
4	0.0007	99 438	71	0.9993	66.91
5	0.0005	99 388	50	0.9995	65.94
6	0.0006	99 325	63	0.9994	64.98
7	0.0007	99 201	73	0.9993	64.03

Kui rahvastiku vanusejaotus aastal  $t$  on teada, ning on alust oletada,  
et lähitulevikus ei toimu olulisi muutusi üleelamistõenäosustes, siis  
on suremustabeli abil lihtne prognoosida surmajuhtude arvu  $D(t)$   
aastal  $t$ :

$$D(t) = \sum_{x=1}^s N(t, x)q_x, \quad (12)$$

kus  $N(t, x)$  tähistab  $x$ -nda vanuserühma arvukust aastal  $t$  ja  $q_x$  selle  
vanuserühma esindaja suremistõenäosust.

10. **Viljakus.** Vastsündinute arv  $B(t)$  aastal  $t$  ja samuti sündinute arv 1000 elaniku kohta e. sündimus  $b(t)$  sõltub ühelt poolt sündimisealiste (tavaliselt arvestuslikult 15 – 45-aastaste) naiste osast elanikkonnas, teiselt poolt aga ka *sündimise intensiivsusest ehk viljakusest*.

Olgu  $b_x$  tõenäosus, et  $x$ -aastane naine sünnitab jooksva aasta vältel lapse, see on nn *vanusest sõltuv viljakuskordaja*. Oletades, et see tõenäosus ei muutu aastate lõikes oluliselt, saame olemasoleva soo-vanuse jaotuse abil prognoosida aastal  $t$  sündivate laste arvu. Kui  $W(t, x)$  tähistab  $x$ -vanuste naiste arvu aastal  $t$ , siis saame sündinutavate laste arvu prognoosiks

$$B(t) = \sum_{x=15}^{45} W(t, x) b_x. \quad (13)$$

Valemit (13) on võimalik täpsustada, arvestades, et mingi (õnneks küll kaunis tühine) osa naisi sureb aasta jooksul, seega tuleks summeeritavad liikmed korrutada vastavalt üleelamistõenäosusega  $p_x$ ,

$$B^*(t) = \sum_{x=15}^{45} W(t, x) p_x b_x. \quad (14)$$

Kui oletada, et sündimise tõenäosus  $b_x$  aja jooksul ei muutu, siis saaksime keskmise naise poolt sündinutavate laste arvu leida lihtsalt, summeerides tõenäosused

$$\bar{b} = \sum_{x=15}^{45} b_x,$$

kus summeerimine toimub üle kõigi aastate, mil naine loetakse sündinutusealiseks. Ka seda valemit saab täpsustada, arvestades naise üleelamistõenäosusi. Teine valem, mida tihti valemi (14) asemel kasutatakse, on *naise poolt sündinutavate tütarde keskmine arv*  $\bar{b}_f$ . Selle jaoks võib saada hinnangu lihtsalt, kasutades *tütarlaste suhtarvu* vastsündinute seas  $f$ . Tavaliselt võetakse  $f = 0.488$  ja siis *poiste suhtarv*  $g = 0.512$  vastavalt vanusele. Keskmine tütarde arv on siis

$$\bar{b}_f = f\bar{b}.$$

Kuigi ka *keskmine sündinutavate laste arv* on leitud ebareaalsele eeldusele tuginedes (sündimise aktiivsus ei muutu mitme aastakümne jooksul) ja on seetõttu teatud mõttes fiktiivne, kuulub ta

olulisemate rahvastikukarakteristikute hulka, ning teda nimetatakse *viljakuse kordajaks*. Tabelis 5 (vt.[4],[5]) on toodud nimetatud näitaja väärtused mõnede Euroopa riikide jaoks.

Tabel 5. Viljakuskordaja, sündimuskordaja ja loomuliku iibe kordaja mõnedes Euroopa riikides käesoleva sajandi 90-ndate aastate algul.

Riik	Viljakus	Sündimus	Iive
Austria	1.5	12.1	1.5
Belgia	1.6	12.5	2.1
Bulgaaria	1.4	10.5	-2.2
Eesti	1.8	12.3	-0.3
Hispaania	1.4	10.0	1.3
Holland	1.6	13.0	4.4
Iirimaa	2.0	14.6	5.8
Inglismaa	1.98	13.5	2.5
Island	2.2	17.6	11.1
Itaalia	1.2	9.9	0.3
Jugoslaavia	1.6	19.0	3.6
Kreeka	1.4	10.1	0.6
Leedu	1.9	14.2	3.2
Luxemburg	1.6	13.0	2.8
Läti	1.7	12.0	-1.5
Moldova	2.3	16.0	5.8
Norra	1.9	14.0	3.6
Poola	1.9	13.4	3.1
Portugal	1.5	11.7	1.4
Prantsusmaa	1.7	13.0	3.9
Rootsi	2.1	14.2	3.2
Rumeenia	1.5	11.4	-0.2
Saksamaa	1.3	10.0	-1.0
Soome	1.9	13.2	3.4
Šveits	1.6	12.6	3.6
Taani	1.8	13.1	1.3
Tšehhia, Slovakkia	2.1	14.1	4.0
Ukraina	1.7	11.5	-1.9
Ungari	1.8	11.8	-2.6
Valgevene	2	13.4	2.2

**11. Rahvastiku prognoosimine lineaarse kasvu mudeli (Leslie mudeli) abil.** Oletame, et meil on teada rahvastiku soo-

vanuse jaotus  $M(t_0, x), W(t_0, x)$  ( $x = 0, \dots, s$ ) mingil lähteastal  $t_0$  ja samuti on teada surevustabel, st üleelamistöenäosused  $p_t$  (nii meeste kui naiste jaoks). Eeldame, et on teada ka vanusest sõltuvad viljakuskordajad  $b_x$ , mis (eelduse kohaselt) ei sõltu aastaarvust  $t$ .

Selleks, et leida valemi (5) järgi rahvastiku loomulik muutus mingil aastal, tuleb leida sellel aastal sündinud laste arv – see on leitav valemiga (13) järgi ning sellest lahutada surmajuhtude arv, mis on leitav valemiga (12). Enamasti toimitakse selliselt, et alustatakse lähteastast, arvutatakse selle põhjal sündide ja surmajuhtude arv järgmiseks aastaks. Selle põhjal hinnatakse järgmiseks aastavahtetuseks ka rahvastiku soo-vanusejaotus, st määratakse suurused

$$M(t_0 + 1, x), W(t_0 + 1, x), x = 0, 1, \dots, s,$$

kasutades selleks enesestmõistetavat tõesiasja, et kõik need  $x$ -aastased inimesed, kes antud aastal ei sure, on järgmisel aastal  $x + 1$ -aastased,

$$N(t_0 + 1, x + 1) = N(t_0, x)(1 - q_x).$$

Omaette probleemiks on 0-aastaste arvu määramine. See saadakse vastasündinute arvust, vt. valem (13), mida parandatakse alla aasta vanuste surmatöenäosuse  $q_0$  arvel. Poiss- ja tütarlaste arv määratakse vastavalt tütarlaste ja poiste suhtarvudele vastasündinute seas  $f$  ja  $g$ , mis samuti loetakse konstantseks. Seega

$$M(t, 0) = B(t)g, \quad W(t, 0) = B(t)f.$$

Eeldades suremus- ja viljakusnäitajate konstantsust saab samal viisil prognoosimist jätkata suvalise aastate arvu jaoks.

## 12. Rahvastiku prognoosimine Leslie maatriksi abil.

Nimetatud prognoosieeskiri on lihtsalt esitatav, kasutades üleminekumaatriksit, mille idee pärineb Markovi ahelate teooriast.

Võtame kasutusele alljärgnevad tähistused. *Rahvaarvu vektoriks*  $N_t$  hetkel  $t$  nimetame alljärgnevat veeruvektorit

$$N_t = (M(t, 0), M(t, 1), \dots, M(t, \omega), W(t, 0), W(t, 1), \dots, W(t, \omega))', \quad (15)$$

mille dimensioon on  $k = 2\omega + 2$ . Seega vektor  $N_t$  sisaldab komponentidena kõigi meeste vanuserühmade arvukusi aastal  $t$  ja selle all kõigi naiste vanuserühmade arvukusi samal aastal.

Selleks, et arvutada rahvaarvu vektor  $N_t$  vektorist  $N_{t-1}$  kasutame valemit

$$N_t = AN_{t-1}, \quad (16)$$

kus  $A$  on nn. *üleminekumaatriks*. Valemis (16) nimetame vektorit  $N_{t-1}$  alg- ja vektorit  $N_t$  lõppvektoriks.

Üleminekumaatriksis  $A$  on  $k$  rida ja veergu, kusjuures iga rida vastab lõppvektori  $N_t$  ühele vanusrühmale ja iga veerg algvektori  $N_{t-1}$  ühele vanusrühmale. Näiteks maatriksi  $A$  element  $a_{ij}$  ( $i, j = 0, \dots, \omega$ ) näitab tõenäosust, et mees vanusega  $j$  on aasta pärast samuti mees vanusega  $i$ . Arusaadavalt on selles maatriksis enamus elemente võrdsed nulliga.

Esitame alljärgnevalt maatriksi  $A$  struktuuri.

$$A = \begin{pmatrix} A_m & C_m \\ 0 & A_f \end{pmatrix}, \quad (17)$$

kus  $A_m$  ja  $A_f$  tähistavad vastavalt meestele ja naistele vastavaid üleminekublokke,  $C_m$  iseloomustab meeste sünde ja 0 on null-maatriks. Vaatleme alljärgnevalt blokke detailsemalt.

$$A_m = \begin{pmatrix} 0 & 0 & 0 & \dots & 0 \\ p_0 & 0 & 0 & \dots & 0 \\ 0 & p_1 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & p_{\omega-1} \end{pmatrix}.$$

Siin iseloomustavad üleelamistõenäosused  $p_x$  mehi, kuid lihtsuse mõttes on lisaindeksist loobutud.

Maatriksis  $C_m$  leidub nullist erinevaid elemente ainult esimeses reas, need on veergudes, mille indekseid vastavad viljakas eas naiste vanustele  $j_1 = \omega + 17, \dots, j_{31} = \omega + 47$ . Maatriks  $C$  on seega defineeritud eeskirjaga

$$c_{1,j_i} = \begin{cases} gb_{14+i}, & i = 1, \dots, 31 \\ 0, & \text{muidu.} \end{cases}$$

Maatriks  $A_f$  sisaldab esimese reana maatriksi  $C_m$  esimese reaga sarnast rida, kus vaid kordaja  $g$  asemel on kordaja  $f$ . Maatriksi  $A_f$  ülejäänud osa on sarnane maatriksiga  $A_m$ , kusjuures loomulikult üleelamistõenäosused  $p_x$  iseloomustavad naisi.

Kasutades toodud maatriksimudelit on võimalik leida prognoos  $t$  aasta peale valemist

$$N_t = A^t N_0.$$

Loomulikult mõjutab ka selle valemi kasutamisel prognoositäpsust oluliselt eeldus, et suremuse ja sündimuse karakteristikud jäävad

prognoosiperioodi vältel konstantseks. Siiski annab valem enamasti lühemate prognoosiperioodide jaoks rahuldavaid tulemusi, mis aga järsult halvenevad, kui rahvastikustatistika näitajad (sündimus, suremus) mingi sotsiaalmajandusliku teguri – näiteks kriisi või sõja tõttu radikaalselt muutuvad. Seetõttu on ka praegu raske anda usaldusväärseid prognoose Eesti rahvastiku jaoks, sest ei sündimus ega ka suremus ei ole peale iseseisvumist stabiliseerunud.

### Viitekirjandus

1. Eesti rahvastik 1993. ESA, Tallinn, 1995.
2. Pirn, A. Surevustabelite koostamisest Eesti andmetel. Bakalau-reusetöö. Tartu, 1995.
3. Rahvastiku vanuskoostis 1994. ESA, Tallinn, 1994.
4. Demographic Statistics 1994.
5. Demographic Yearbook. United Nations, N.York, 1993.
6. Dublin, L.J., Lotka, A.J. On the true rate of natural increase. J. Amer. Stat. Ass., 1925, no. 151.
7. Hyrenius, H. Demometri. Den formella befolkningslärans grunder. Göteborg, 1966.
8. Statistisk Årsbok' 92, 1992.

## KORRELATSIOONIMAATRIKSI OHTLIKUD OLULISUSE TÕENÄOSUSED

A.-M. Parring, M. Vähi

Tartu Ülikool, Matemaatilise Statistika Instituut

Kindlasti ei ole olulisuse tõenäosus iseenesest ohtlik. Ohtlik on aga see, et sageli kiputakse korrelatsioonimaatriksi analüüsimisel olulisuse tõenäosusi kasutama valesti. Püüamegi järgnevalt selgitada, milles seisneb eksimus ja kuidas seda vältida.

Tuletame meelde, kuidas määratakse olulisuse tõenäosus ühe korrelatsioonikordaja analüüsimisel. Esmalt fikseeritakse kontrollitavad hüpoteesid. Kõige sagedamini on need esitatavad kujul, kus sisukas hüpotees on kahepoolne,

$$H_0 : \rho = 0$$

$$H_1 : \rho \neq 0.$$

Normaaljaotusega tunnuste korral kasutatakse otsustamiseks teststatistikut

$$T = r \sqrt{\frac{n-2}{1-r^2}},$$

kus sümboliga  $n$  on tähistatud valimi maht, sümboliga  $r$  aga selle valimi põhjal arvutatud korrelatsioonikordaja. Kui nullhüpotees kehtib, on selle statistiku jaotuseks  $t$ -jaotus parameetriga  $n-1$ .

Seejärel valitakse kriitiline piirkond. Kuna sisuka hüpoteesi kehtivuse korral on võimalikud nii negatiivsed kui ka positiivsed korrelatsioonikordaja väärtused, sobib kasutada kaheosalist kriitilist piirkonda. Selle raja määratakse nõnda, et sõltumatute tunnuste korral oleks teststatistiku kriitilisse piirkonda sattumise tõenäosus parajasti võrdne uurija poolt valitud olulisuse nivooga  $\alpha$ .

Lõpuks arvutatakse otsuse langetamiseks vajaminev statistiku väärtus. Kui see kuulub kriitilisse piirkonda, loetakse sisukas hüpotees tõestatuks, vastasel juhul tuleb jääda nullhüpoteesi juurde. Otsuse langetamiseks võib kasutada kahte võtet:

- 1) võib võrrelda statistiku väärtust valitud olulisuse nivoole vastava kriitilise piirkonna rajaga;



2) võib arvutada olulisuse tõenäosuse – olulisuse nivoo, millele vastava kriitilise piirkonna rajaks oleks leitud statistiku väärtus.

Mõlemad võtted annavad ühe ja sama tulemuse. Kuna teine võte on üldiselt informatiivsem ja arvutis realiseerimiseks mugavam, väljastavad statistikapaketid reeglina statistiku väärtuse ja sellele vastava olulisuse tõenäosuse. Olulisuse tõenäosusel on seega ilus sisuline tähendus – ta on vähim olulisuse nivoo, mille korral saaks antud valimi põhjal vastu võtta sisuka hüpoteesi.

Üheainsa korrelatsioonikordaja analüüsimisel töötab olulisuse tõenäosus tõrgeteta. Eksimise risk seose oluliseks tunnistamisel on uurija enda määrata – kui valida  $\alpha = 0.05$ , tunnistatakse testi 100-kordsel rakendamisel erinevatele valimitele keskmiselt 5 sõltumatut tunnuste paari oluliselt seotuks; kui valida  $\alpha = 0.01$ , tunnistatakse testi 100-kordsel rakendamisel erinevatele valimitele keskmiselt 1 sõltumatute tunnuste paar oluliselt seotuks jne.

Mis juhtub aga, kui ühe korrelatsioonikordaja asemel on neid terve maatriksitäis ja seda suurt hulka korrelatsioonikordajaid analüüsitakse täpselt samuti nagu ühteainsat – vaadatakse iga üksiku kordaja olulisuse tõenäosust ja otsustatakse selle põhjal, kas vastav seos on oluline või mitte.

Intuiivselt peaks olema selge, et tõenäosus kriitilise piirkonna tabamiseks mõne korrelatsioonikordajaga on nüüd oluliselt suurenenud. Ühe "viske" asemel püütakse seda tabada  $p(p-1)/2$  "viskega" – just nii palju on ühes korrelatsioonimaatriksis erinevaid korrelatsioonikordajaid. Muutuse tunnetamiseks sobiks järgmine kujutlus-pilt. On sügis ja te seisate küpseid õunu täis õunapuu all. Kuigi aeg-ajalt potsatab üksik õun alla, tunnete te end suhteliselt julgelt – õunu kukub harva ja tõenäosus, et mõni teid tabab, on tühine. Siis aga haarab keegi õunapuu tüvest ja raputab. Silmapilk olete te õunapuu alt kadunud. Miks? Ühe juhusliku katse asemel vallandus terve seeria katseid. Tõenäosus, et mitte ükski õun ei taba, muutus väga väikeseks. Te püüdsite kaugemale tõmbuda piirkonnast, kuhu õunad langevad. Sama muutus juhtub ka ühe korrelatsioonikordaja analüüsimiseks ettenähtud kriitilise piirkonnaga, kui selle abil kontrollida suure hulga korrelatsioonikordajate olulisust. Katsete seeria korral muutub selle piirkonna tabamise tõenäosus liiga suureks. Piirkonda tuleb "tagasi tõmmata" alalt, kuhu statistiku väärtused nullhüpoteesi kehtides võivad langeda liiga sageli.

Vale meetodika kasutamise tagajärgi on võimalik illustreerida modelleerimise teel. Vaatleme normaaljaotusega tunnusvektorit, millel on 10 sõltumatut komponenti. Iga tunnuste paari jaoks on siis

õige nullhüpotees – tõeline korrelatsioonikordaja on võrdne nulliga. Genereerime soovitava mahuga juhusliku valimi selle tunnusevektori väärtustest ja arvutame valimi põhjal korrelatsioonimaatriksi. Selles maatriksis on 10 rida ja 10 veergu, erinevaid korrelatsioonikordajaid on 45.

Kontrollime saadud korrelatsioonikordajate olulisust nii, nagu ei tohi – loeme seose oluliseks, kui korrelatsioonikordaja olulisuse tõenäosus ei ületa meie poolt valitud olulisuse nivood  $\alpha$ . Tagajärgede kirjeldamiseks loeme ära maatriksid, millede leitakse vähemalt üks oluline korrelatsioonikordaja ning arvutame, millise protsendi moodustavad need maatriksid kõikidest modelleeritud maatriksitest. See on vale meetodika tagajärjel tehtud valede otsustuste protsent.

Modelleerimine on viidud läbi valimi mahtude 12, 60 ja 120 korral. Olulisuse nivool 0.05 saadud tulemused (iga valimi mahuga on tehtud 200 katset) on toodud tabelis 1.

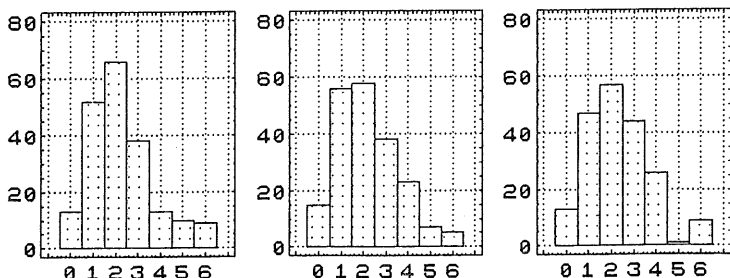
Tabel 1

Tunnused on sõltumatud,  $\alpha = 0.05$

Valimi maht	12	60	300
Pole ühtegi olulist seost	6.5 %	6.5%	7.5%
Vähemalt üks oluline seos	93.5%	93.5%	92.5%

Nagu näha, jätkub heausksele statistikakasutajale üsna palju leidmisrõõmu: enam-vähem vaid iga kümnes korrelatsioonimaatriks on selline, kus ühtegi olulist seost ei avastata. Paraku on rõõm aga põhjendamatu – tegelikult olid tunnused sõltumatud, tõelises korrelatsioonimaatriksis pole ühtegi olulist seost. Tasub tähele panna, et valede otsustuste protsent on stabiilne – see ei muutu oluliselt valimi mahu kasvades.

Oluliseks tunnustatud seoste arv ühes maatriksis on 1 ja 6 vahel, kusjuures vastava arvu suhtelised sagedused on nii suurte kui ka väikeste valimite korral peaaegu ühesugused. Tulemust illustreerivad histogrammid joonisel 1.



valimi maht 12

valimi maht 60

valimi maht 300

Joonis 1. Oluliste seoste arvu jaotus

Tabelis 2 on toodud samade juhuslike valimite korral olulisuse nivool 0.01 saadud tulemused.

Tabel 2

Tunnused on sõltumatud,  $\alpha = 0.01$

Valimi maht	12	60	300
Pole ühtegi olulist seost	64.5 %	59.5%	61.5%
Vähemalt üks oluline seos	35.5%	41.0%	38.5%

Tulemused on veidi paremad. Kuid ikkagi leitakse umbes igas kolmandas korrelatsioonimaatriksis mõned olematud olulised seosed.

Olematu seose oluliseks tunnistamine on esimest liiki viga – selle tegemise tõenäosust peab tõkestama olulisuse nivoo  $\alpha$ . Vale meetoodika rakendamine suurendas esimest liiki vea tegemise riski mitmekümne kordselt. See on vastuolus matemaatilise statistika põhimõtetega – risk esimest liiki vea tegemiseks peab olema range kontrolli all.

Riski lubamatu suurenemise vältimiseks on kaks teed.

Kontrollitavad hüpoteesid võib sõnastada terve korrelatsioonimaatriksi kohta. Nullhüpoteesiks on siis oletus, et kõik tunnused on sõltumatud. Sel juhul on kõik korrelatsioonimaatriksi elemendid, mis ei asu peadiagonaalil, nullid, peadiagonaalil aga asuvad ühed – korrelatsioonimaatriksiks on ühikmaatriks  $I$ . Sisukaks hüpoteesiks tuleks valida eelmise väite eitus. Tähistades tõelise korrelatsioonimaatriksi sümboliga  $\mathfrak{R}$  saame need väited esitada järgmiselt:

$$H_0 : \mathfrak{R} = I$$

$$H_1 : \mathfrak{R} \neq I.$$

Niisuguste hüpoteeside kontrollimiseks kasutatakse Bartlett' statistikut  $B$ ,

$$B = -\left(n - \frac{1}{6}(2p + 1)\right) \ln |R|,$$

kus  $p$  on analüüsitava tunnuste arv,  $n$  kasutatud valimi maht,  $|R|$  aga valimi korrelatsioonimaatriksi determinant. Kui Bartlett' statistiku põhjal võetakse vastu nullhüpotees, pole korrelatsioonimaatriksit üldse mõtet edasi analüüsida.

Nullhüpoteesi kehtivuse korral läheneb Bartlett' statistiku jaotus valimi mahu kasvades  $\chi^2$ -jaotusele parameetriga  $p(p - 1)/2$ . Eelpool analüüsitud valimite korral annab Bartlett' statistik tulemused, mis on toodud tabelis 3.

Tabel 3

Tunnused on sõltumatud, otsus langetatakse  
Bartlett' statistiku põhjal,  $\alpha = 0.05$

Valimi maht	12	60	300
Pole ühtegi olulist seost	78.5 %	95.5%	95.0%
Vähemalt üks oluline seos	21.5%	4.5%	5.0%

Nagu näha, ei tööta Bartlett' statistik 12-elementilise valimi korral veel kuigi hästi. Suurte valimite korral on tulemused ootuspärased – olulisuse nivoo oli 0.05 ja keskmiselt 5 maatriksi korral 100-st analüüsitud langetatakse vale otsus, tunnistades oluliseks mõned olematud seosed.

Bartlett' statistikut arvutavad paketid SYSTAT ja SAS, pakett STATGRAPHICS aga seda statistikut ei arvuta.

Teine tee esimest liiki vea riski suurendamisest hoidumiseks on Bonferroni võrratuse kasutamine. Olgu korrelatsioonimaatriksi elemendid mingil viisil nummerdatud. Tähistame sündmuse, et  $i$ -s korrelatsioonimaatriksi element omandab väärtuse väljaspool kriitilist piirkonda, sümboliga  $A_i$ . Siis sündmus  $A$  – mitte ükski korrelatsioonimaatriksi element ei satu kriitilisse piirkonda – on sündmuste  $A_i$  korrutis

$$A = \cap_i A_i.$$

Sündmus, et mõni korrelatsioonimaatriksi element satub kriitilisse piirkonda, on selle sündmuse vastandsündmuseks,

$$A^c = (\cap_i A_i)^c = \cup_i A_i^c.$$

Kuna tõenäosusteoorias tuntud võrratuse põhjal

$$P(\cup_i A_i^c) \leq \sum_i P(A_i^c),$$

siis saame siit tõkke tõenäosusele, et ükski korrelatsioonimaatriksi element ei satu kriitilisse piirkonda

$$P(A) \geq 1 - \sum_i P(A_i^c).$$

Saadud võrratus kannabki Bonferroni võrratuse nime.

Kui tahame, et sõltumatute tunnuste korral sündmuse  $A$  tõenäosus poleks väiksem kui  $1 - \alpha$ , peame iga üksiku korrelatsioonikordaja kriitilisse piirkonda sattumise tõenäosuseks valima  $\alpha/k$ , kus  $k$  on analüüsitava korrelatsioonimaatriksi erinevate elementide arv, s.t.  $k = p(p - 1)/2$ . Seda tõenäosust nimetataksegi Bonferroni tõenäosuseks. See on tõke, millest väiksemad olulisuse tõenäosused tunnistavad seose oluliseks.

Kui näiteks uuritavaid tunnuseid on 10 ja me tahame kasutada olulisuse nivood 0.05, saame Bonferroni tõenäosuseks suuruse  $0.05/45 \approx 0.0011$ . Oluliseks tuleb  $10 \times 10$  korrelatsioonimaatriksis tunnistada vaid need korrelatsioonikordajad, mille olulisuse tõenäosus on väiksem kui 0.0011. Tulemused, mis saadakse eelpool modelleeritud valimite korral Bonferroni tõenäosust kasutades, on esitatud tabelis 4.

Tabel 4

Tunnused on sõltumatud, otsus langetatakse  
Bonferroni tõenäosuste põhjal,  $\alpha = 0.05$

Valimi maht	12	60	300
Pole ühtegi olulist seost	93.5 %	95.5%	95.5%
Vähemalt üks oluline seos	6.5%	4.5%	4.5%

Näeme, et nende maatriksite protsent, mille korral leitakse olematuid olulisi seoseid, on lähedane kasutatud olulisuse nivoole. Siin on tulemused juba väikeste valimite korral ootuspärased.

Seega on meil olemas kaks üsna mõistlikku strateegiat korrelatsioonimaatriksi analüüsimiseks juhul, kui uuritavad tunnused on tegelikult sõltumatud. Neis on seose oluliseks tunnistamine muudetud raskemaks kui ühe korrelatsioonikordaja analüüsimisel. Sõltumatute tunnuste korral väldib see suure hulga tegelikkuses olematute seoste leidmist.

Nüüd aga võib tekkida küsimus, mis juhtub siis, kui tegelikult tunnused pole sõltumatud? Kas ei jää siis avastamata suur hulk tegelikkuses eksisteerivaid seoseid?

Püüame ka sellist olukorda illustreerida modelleerimise teel. Kuna sõltuvate tunnuste korral võib korrelatsioonimaatriksi struk-

tuur olla äärmiselt mitmekesine, ei saa küsimusele anda ammen-  
davast vastust. Modelleeritud on situatsiooni, kus tegemist on ho-  
mogeense korrelatsioonimaatriksiga – kõik väljaspool peadiagonaali  
asuvad elemendid on võrdsed. Nii omab näiteks  $4 \times 4$  homogeenne  
korrelatsioonimaatriks kuju

$$\begin{pmatrix} 1 & \rho & \rho & \rho \\ \rho & 1 & \rho & \rho \\ \rho & \rho & 1 & \rho \\ \rho & \rho & \rho & 1 \end{pmatrix}.$$

Esimeses modelleerimiste seerias on 10-komponendilise normaaljao-  
tusega juhusliku vektori korral valitud  $\rho = 0.2$ . Piltlikult öeldes  
on meil tegemist olukorraga, kus tunnuste vahel on palju suhteliselt  
nõrku seoseid. Tabelis 5 on toodud nende korrelatsioonimaatriksite  
protsent, mille korral vähemalt üks seos tunnistati oluliseks. An-  
tud olukorras muutub selle arvu tähendus – see on nende maatrik-  
site protsent, mille korral tehti vähemalt üks õige otsus. Rakendati  
kõiki siiani vaadatud meetodikaid: "ohtlikke olulisuse tõenäosusi"  
(OOT), Bartlett' statistikut, Bonferroni tõenäosust. Igale metoo-  
dikale vastab tabeli erinev rida. Nagu varem, on modelleerimisel  
kasutatud kolme erinevat valimi mahtu: 12, 60 ja 300.

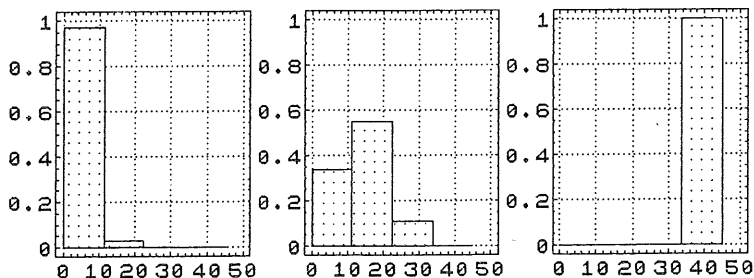
Tabel 5

Tunnused on nõrgalt seotud ( $\rho = 0.2$ ), $\alpha = 0.05$			
Valimi maht	12	60	300
OOT	96.5 %	99.5%	99.5%
BARTLETT'	38.0%	97.5%	100.0 %
BONFERRONI	10.5%	75.0%	100.0 %

Valitud korrelatsioonimaatriksi korral on tegelikult kõikvõima-  
like tunnuste paaride vahel seos, kuigi üsna nõrk seos. Loomulikult  
töötavad niisuguses olukorras rangemad testid halvemini. Väikeste  
valimite korral on korrelatsioonimaatriksite protsent, kus vähemalt  
üks oluline seos avastatakse, suhteliselt väike. See protsent aga suu-  
reneb kiiresti valimi mahu kasvades. Seda tasub meeles pidada –  
nõrkade seoste avastamiseks ei tohi valimi maht olla väike.

Nõrga seose avastamata jätmine on teist liiki viga. Selle vea  
tegemise risk on tõkestatud olulisuse nivoo valikuga, tõkkeks on suu-  
rus  $1 - \alpha$ . See, et väikeste valimite puhul on nõrga seose avastamata  
jätmise risk suur, on statistika omapära. Seda riski on võimalik  
vähendada kahel viisil – suurendades esimest liiki vea tegemise riski  
või suurendades katsete arvu.

Tulemuste täpsemaks kirjeldamiseks on joonistel 2 ja 3 toodud oluliseks tunnistatud seoste arvu jaotushistogrammid 12, 60 ja 300 katselises seerias ohtlike olulisuse tõenäosuste ja Bonferroni tõenäosuse kasutamisel.

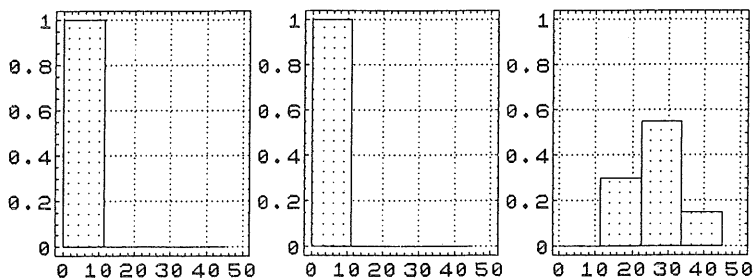


valimi maht 12

valimi maht 60

valimi maht 300

Joonis 2. Oluliste seoste arvu jaotus OOT kasutamise korral



valimi maht 12

valimi maht 60

valimi maht 300

Joonis 3. Oluliste seoste arvu jaotus Bonferroni tõenäosuste korral

Näeme, et avastatud oluliste seoste arv kasvab valimi mahu kasvades üsna kiiresti. Bonferroni tõenäosuste kasutamisel 300-elementiliste valimite korral on avastatud seoste arv rohkem kui pooltel juhtudel 20 ja 30 vahel.

Teises modelleerimiste seerias on 10-komponendilise normaaljaotusega juhusliku vektori korral valitud  $\rho = 0.5$ . Piltlikult öeldes on sel juhul tegemist olukorraga, kus tunnuste vahel on palju keskmise tugevusega seoseid. Modelleerimise tulemused on esitatud tabelis 6.

Tabel 6

Tunnused on keskmise tugevusega seotud ( $\rho = 0.5$ ),  $\alpha = 0.05$

Valimi maht	12	60	300
OOT	99.5 %	100.0%	100.0%
BARTLETT'	85.5%	100.0%	100.0 %
BONFERRONI	64.5%	100.0%	100.0 %

Nagu näha, on keskmise tugevusega seoste korral korrelatsioonimaatriksite protsent, mille korral vähemalt üks seos tunnistatakse oluliseks, tunduvalt kasvanud ka väikeste valimimahtude korral. Suurte valimite korral töötavad ka rangemad kriteeriumid väga hästi.

Kolmandas modelleerimiste seerias on 10-komponendilise normaajaotusega juhusliku vektori korral valitud  $\rho = 0.8$ . Piltlikult öeldes on sel juhul tegemist olukorraga, kus tunnuste vahel on palju tugevaid seoseid. Niisuguses olukorras töötasid kõik kriteeriumid väga hästi – 200-katselises seerias olid ainult sellised korrelatsioonimaatriksid, milles mõni seos oli oluline.

**Kokkuvõte.** Korrelatsioonimaatriksi analüüsimisel ei tohi olulisuse tõenäosust kasutada samal viisil kui ühe korrelatsioonikordaja analüüsimisel. Vältimaks esimest liiki vigade kuhjumist, tuleb seose oluliseks tunnistamise kriteeriumi muuta rangemaks. Rangeimate kriteeriumide ja väikeste valimite korral suureneb risk mõne olulise seose avastamata jätmiseks. See on statistiliste otsustuste omapära, mille vastu saab võidelda ainult valimi mahtu suurendades. Seda omapära arvestab ka nullhüpoteesi vastu võtmise vormel – "antud katse korral ei õnnestunud nullhüpoteesi kummutada". Vastu võetud nullhüpoteesi ei loeta tõestatuks.

Programm valimite modelleerimiseks, korrelatsioonimaatriksi arvutamiseks ja tulemuste analüüsimiseks on kirjutatud SAS-süsteemi IML-protseduuri maatrikskeeles.

### Viitekirjandus

1. SAS Institute Inc. SAS/IML User's Guide, Release 6.03 Edition. Cary, NC: SAS Institute Inc. 1988
2. Tiit E., Viil M. Andmeanalüüs personaalarvutil programmi-paki STATGRAPHIC abil. Tartu, 1992



## KORDUSMÕÕTMISTEST

E. Käärrik

Tartu Ülikool, Matemaatilise Statistika Instituut

Kordusmõõtmistega (ingl.k. *repeated measurements*) andmed esinevad meil siis, kui igal objektil on mingit tunnust (näitajat) mõõdetud rohkem kui üks kord. Tavaliselt on tegemist ajas üksteisele järgnevate mõõtmistega. Teatavas mõttes toimub sellisel juhul objektide ökonoomsem kasutamine, igalt objektilt saame ühe tunnuse kohta terve rea väärtusi, samas aga tuleb arvesse võtta, et ühelt objektilt saadud tulemused on omavahel seotud.

Tüüpiliselt esineb kordusmõõtmisi meditsiinis: ühte näitajat (tunnust) mõõdetakse teatud ajavahemike järel, uuritakse mingi ravi mõju vms. Näiteks antakse ravimit ja mõõdetakse kehatemperatuuri, pulssi ja vererõhku iga päev nädala jooksul. Tihti on tegemist erinevate ravimeetodite võrdlemisega.

Sellist situatsiooni võime vaadelda kui mitmemõõtmelise analüüsi ülesannet.  $T$  erinevat mõõtmist kujutavad endast  $T$ -dimensionaalset juhuslikku vektorit  $\vec{y}_{ij} = (y_{ij}^{(1)}, \dots, y_{ij}^{(T)})$  keskväertusega  $\vec{\mu}_i = (\mu_i^{(1)}, \dots, \mu_i^{(T)})$ , kus  $i = 1, \dots, R$  (grupp, rühm, ravi);  $j = 1, \dots, N_i$  (objekt,  $N = \sum_{i=1}^R N_i$ ) ja kovariatsioonimaatriksiga  $\Sigma$ . Selline lähenemine eeldab, et keskväertus  $\mu_i$  sõltub ainult rühmast (ravist)  $i$  ja kovariatsioonimaatriks on sama iga ravi ja objekti korral.

Aegridade teooria kohaselt kirjeldab keskväertuse vektor  $\mu_i$  trendi ja kovariatsioonimaatriks  $\Sigma$  autokovariatsioonifunktsiooni. Kui tavalises aegridade teoorias pakub tihti huvi eelkõige autokovariatsioonanalüüs, siis korduvate mõõtmiste korral on peaarõhk trendil. Siinkohal me aegridade teorial ei peatu.

Meil on seega andmetabel:

Rühm (ravi)	Objekt	Mõõtmine (test, aeg)			
		1	2	...	T
1	1	$y_{11}^{(1)}$	$y_{11}^{(2)}$	...	$y_{11}^{(T)}$
...	...	...	...	...	...
1	$N_I$	$y_{1N_I}^{(1)}$	$y_{1N_I}^{(2)}$	...	$y_{1N_I}^{(T)}$
...	...	...	...	...	...
...	...	...	...	...	...
R	1	$y_{R1}^{(1)}$	$y_{R1}^{(2)}$	...	$y_{R1}^{(T)}$
...	...	...	...	...	...
R	$N_R$	$y_{RN_R}^{(1)}$	$y_{RN_R}^{(2)}$	...	$y_{RN_R}^{(T)}$

Vaatame MANOVA (Multivariate ANalysis Of VAriance mudelit. MANOVA mudel on ANOVA mudeli üldistus, kus ühe funktsioonitunnuse asemel on mitu tunnust. Need tunnused võivad olla kas täiesti erinevad (näiteks kasv, kaal, IQ jne.) või nad võivad olla ühe tunnuse erinevad mõõtmised. Viimasel juhul ongi meil tegemist funktsioonitunnuse kordusmõõtmistega.

Maatrikskujul on see  $Y = X \cdot \beta + \varepsilon$ , kus  $Y$  on  $N \times T$  kordusmõõtmiste maatriks,  $X$  on  $N \times R$  plaanimaatriks, mis koosneb ühtedest ja nullidest ning näitab, kuhu rühma objekt kuulub,  $\beta$  on  $R \times T$  tundmatute parameetrite maatriks,  $\varepsilon$  on  $N \times T$  juhuslike vigade maatriks, kusjuures  $\varepsilon \sim N(0, I_N \otimes \Sigma)$ .

Korduvate mõõtmistega andmete korral on tegemist faktoritega (rühma faktor ja kordusfaktor), mis on objektiga erinevalt seotud. Objekt on allutatud rühma faktorile. Kordusfaktor aga on objektiga ristseoses, igal korduse tasemel on mõõdetud kõik objektid.

Üldise lineaarse hüpoteesi kuhu rühma ja korduse koosmõju kontrollimiseks on  $H_0 : L\beta M = 0$ , kus  $(R - 1) \times R$  maatriks  $L$  sisaldab rühma (ravi) kontrastid (nulliga võrduvad lineaarkombinatsioonid) ja  $T \times (T - 1)$  maatriks  $M$  sisaldab aja (kordusmõõtmiste) kontrastid.

$$L = \begin{pmatrix} 1 & -1 & 0 & \dots & 0 \\ 0 & 1 & -1 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & 0 & \dots & -1 \end{pmatrix}$$

$$M = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ -1 & 1 & 0 & \dots & 0 \\ 0 & -1 & 1 & \dots & 0 \\ 0 & 0 & -1 & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & 0 & \dots & -1 \end{pmatrix}$$

Seega võime kirjutada

$$(L\beta M)_{ik} = (\beta_{ik} - \beta_{ik+1}) - (\beta_{i+1k} - \beta_{i+1k+1}) = 0,$$

kus  $i = 1, \dots, R$  määrab rühma ja  $k = 1, \dots, T$  määrab korduse.

Hüpoteeside kontrolli juures saame tõmmata paralleele enam-tuntud ANOVA mudeliga:

ANOVA	MANOVA
SS Sum of Squares	Sum of Products matrix H
SSE Error sum of Squares	Error Sum of Products matrix E
SS/SSE	$E^{-1}H$ või $(E + H)^{-1}H$

Arvutatakse neli teststatistikut [2]:

1. Wilks'i lambda  $\Lambda = |E| / |H + E|$
2. Pillai' jälg  $V = \text{tr}(H/(H + E))$
3. Hotelling-Lawley jälg  $U = \text{tr}(E^{-1}H)$
4. Roy' suurim omaväärtus  $\lambda_{max}$ , kus  $\lambda_1 < \lambda_2 < \dots < \lambda_{max}$  on  $E^{-1}H$  omaväärtused.

Leitud statistikute kasutamiseks avaldatakse nende abil ligikaudne  $F$ -statistik. Selleks arvutatakse teatud abisuurused:

$$p = \text{rank}(H + E); \quad q = \text{rank}(L); \quad V = N - R;$$

$$s = \min(p, q); \quad m = (|p - q| - 1)/2; \quad n = (V - p - 1)/2$$

ja  $s = 1$  korral saadakse näiteks Wilks'i lambda abil täpne  $F$ -statistik

$$F = (n + 1)(1 - \Lambda)/(m + 1)\Lambda \sim F(2m + 2, 2n + 2).$$

Märkus: SAS-programm väljastab vaikimisi nii  $s, n, m$ , neli teststatistikut kui ka neile vastavad ligikaudsed  $F$ -statistikud.

Näitena on lahendatud järgmine ülesanne. Meil on tegemist 15 sportlasega (suusatajad), neist 7 meest ja 8 naist ( $R = 2, N_1 = 7$ ,

$N_2 = 8, N = 15$ ). Sportlased on teinud nn. koormustesti, s.t. nad on testitud velogomeetrial, kus on iga kolme minuti tagant tõstetud koormust 25W võrra, koormust on tõstetud 6 korda, algkoormuseks on 50W. Pulssi on mõõdetud enne testi ja igal koormusastmel (võetakse iga koormusastme keskmine pulss), seega  $T = 8$ . Korduvalt mõõdetud tunnuseks on pulss ning faktoriks, mis määrab mõõtmise korduvuse, on koormus (koormuse igal tasemel mõõdetakse pulss). Algtabel on seega järgmine:

Sugu	Indiv nr.	1 0W	2 50W	3 75W	4 100W	5 125W	6 150W	7 175W	8 200W
1	1	62	91	94	107	118	129	137	148
1	2	83	92	97	104	114	111	127	138
1	3	75	92	96	106	117	132	141	158
1	4	74	83	91	98	104	114	120	126
1	5	51	96	93	104	112	121	137	147
1	6	70	91	87	103	110	122	128	132
1	7	66	83	90	96	106	125	145	148
2	1	55	76	88	96	118	140	159	166
2	2	44	75	86	99	110	119	140	151
2	3	90	106	115	126	136	151	160	173
2	4	101	102	110	124	134	152	160	172
2	5	58	94	94	104	115	128	144	156
2	6	85	113	125	133	146	156	166	177
2	7	97	111	116	126	135	141	153	164
2	8	59	89	107	116	125	135	149	156

Andmeid illustreerib järgmine graafik (vt. Joonis 1 lk. 37).

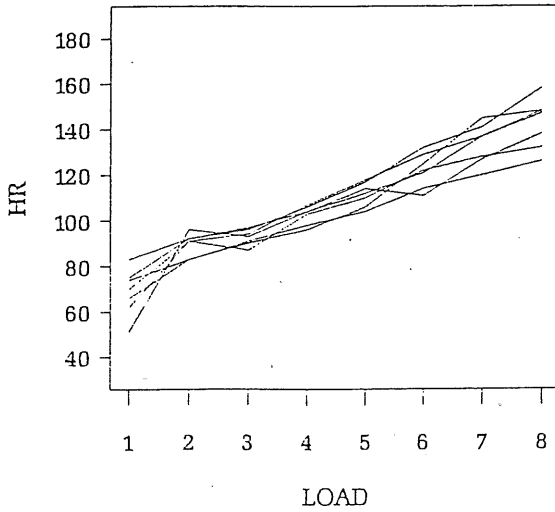
Probleem on selles, kuidas koormus mõjub pulsile ja kas ta mõju on meestele ja naistele erinev. Ülesande lahendamiseks on koostatud järgmine SASi programm:

```

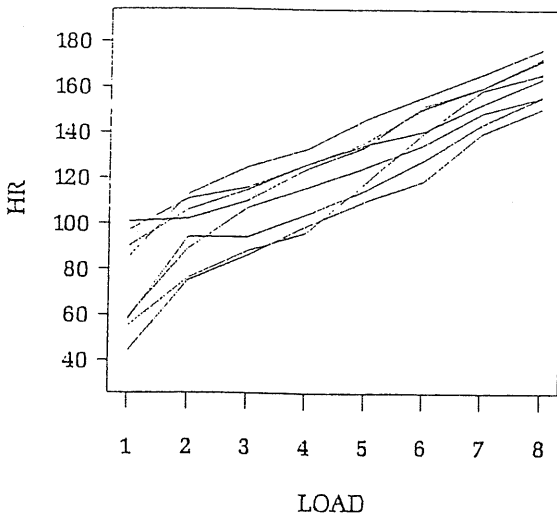
data a1;
infile 'd:\ enesas\ suus.dat'
input sex nr hr1 hr2 hr3 hr4 hr5 hr6 hr7 hr8;
run;
proc glm;
classes sex;
model hr1-hr8=sex;
manova;
repeated koormus contrast (1)/printe printh printm printrv sum-
mary;

```

Step - test  
SEX = 1



SEX = 2



Joonis 1. Pulsisagedused (HR) erinevatel koormustel (LOAD) meestel (SEX=1) ja naistel (SEX=2).

Korduvaid mõõtmisi määrav(ad) faktor(id) tuleb ette anda REPEATED-lausega. Faktori nimi on programmi koostaja poolt vabalt valitav; kui faktoreid on mitu, siis antakse iga nime järel tasemete arv. Korduvate faktorite arv ei ole piiratud. Antud juhul on faktori nimeks 'koormus'. CONTRAST(1) määrab maatriksi  $M$  kuju selliselt, et esimene tase oli korduvalt mõõdetud tunnuse jaoks teatav algtase, mille suhtes teisi analüüsitakse. Kui meid huvitavad erinevused korduvalt mõõdetud tunnuse erinevate tasemete vahel, siis tuleb anda valik PROFILE. Valikud PRINTE, PRINTH, PRINTM, PRINTRV annavad vastavalt maatriksite  $E$ ,  $H$  ja  $M$  ning maatriksi  $E^{-1}H$  omaväärtuste ning omavektorite väljatrüki. PRINTE annab ka kovariatsioonimaatriksi. SUMMARY annab korduva faktori ANOVA tabelid, mis, arvestades valikut CONTRAST(1), näitavad, millised korduva faktori tasemed erinevad oluliselt algtasemest. Väljastatakse ka iga korduva mõõtmise ANOVA tabelid, nende ärakeelamiseks saab kasutada valikut NOUNI MODEL-lauses.

Antud ülesande lahendus näitab, et pulsile avaldavad mõju nii sugu kui koormus (peamõjud). Tõestada ei saanud aga soo ja koormuse koosmõju, s.t. koormuse mõju liitub soo mõjuga, täiendav koosmõju puudub.

Toodud näide peaks veenvalt illustreerima kordusmõõtmistega andmete analüüsi seni võib-olla vähemtuntud MANOVA meetodiga ja näitama, et taoline meetodika võimaldab analüüsida probleemi sügavamalt, määrates ka kordust põhjustava faktori mõju.

### Kirjandus

1. Hand D.J., Taylor C.C. (1991) Multivariate Analysis of Variance and Repeated Measures.
2. Lindsey J. (1993) Models for Repeated Measurements.
3. SAS/STAT User's Guide. Release 6.03 Editor. SAS institute 1988.
4. Skovgaard L.T. (1992) Repeated Measurements for Continuous Responses. Abstract 14. Nordic Conference in Math. Statistics. Roros, June 1992.

# LOODUSLIKE RÜTMIDE LAINETE KUJU KIRJELDAMISEST

V. Reeben

Tartu Ülikool, Füsioloogia Instituudi biofüüsika uurimisgrupp

Teatavasti on statistika üheks haruks looduslike nähtuste aegridade spektrite uurimine. Klassikalisteks näideteks statistikaraamatutes on olnud päikeseplekkide aegrida (andmed alates 1749 a.) ja Lääne-Euroopa nisu hindade aegrida (andmed alates 1260 a.).

Itaallaste M. Pagani jt. hüpotees, et inimese südame rütmi ja vere rõhu lainete spektrite uurimine võimaldab diferentseerivalt diagnoosida autonoomse närvi-süsteemi seisundit on tänapäeval tekitanud buumi jõukamate maade meditsiinis, kus vaatamata aparatuuri kõrgele hinnale loodetakse sellel teel leida uut efektiivset informatsiooniallikat. Ka füüsi- sikud näevad biomeditsiiniliste rütmide uurimises uut ja huvitavat tegevusvälja nendele tuttava Fourier' kiirteisenduse jms. meetodite kaudu.

Tartu Ülikooli Biofüüsika Laboratooriumis (praegu TÜ Füsiol. Inst. biofüüs. uurimisgrupp) on pikaajalised kogemused inimese vere rõhu, südame rütmi jt. füsioloogiliste parameetrite varieeruvuse uurimisel (alates 1963. aastast) ja ka suur sellealane infobank magnetlintidel (alates 1978 a.). Viimastel aastatel on siia lisandunud võimalused nende aegridade otseseks ja üle magnetlindi sisestamiseks arvutile ja registreeritud mitmekanalilise informatsiooni spektraalseks töötlemiseks.

Samal ajal on biofüüsika uurimisgrupis tegeldud ka looduslike rütmide üldisemate ("globaalsete") probleemidega, mis tuginevad Ramanujani tegurirohketele arvudele (sellised täisarvud  $N = 2^{\alpha} \cdot 3^{\beta} \cdot 5^{\gamma} \dots$ , millel on maksimaalselt võimalik algtegurite arv  $dN = (\alpha+1)(\beta+1)(\gamma+1)\dots$ ), (Ramanujan 1915) ja nende kohasusele olla teoreetiliseks aluseks looduses levinud objektide ja nendega seotud looduslike rütmide kirjeldamisel.

Leitud rütmide kirjeldamise seos on osutunud hämmastavalt lihtsaks. Põhiseoses

$$T_x = T_0 / N_x$$

on ainult üks universaalkonstant  $T_0$  ja üks arvuteoreetiliselt määratud muutuja  $N_x$ . Viimane kujutab endast sellist tegurirohket arvu, mis leitakse kahe kriteeriumi alusel:

- $dN_x = \max$ , tegurirohkus, (informatsiooni maksimum),
- $N_x \approx a^2$ , ruutarvu lähedus, (kommunikatiivsuse maksimum).

Konstanti  $T_0$  võib interpreteerida kui galaktilise aasta pikkust.

Urvides kõikvõimalikke looduslikke rütme suures diapasoonis, torkab selgesti silma võimalus eristada kahte erinevat tüüpi perioodilisi protsesse olenevalt sellest, kas tegurirohket arvude grupi  $N(p_i)$  suurim algtegur  $p_{\max}$  on  $4k-1$  (Im) või  $4k+1$  (Re) tüüpi arv. Suurimate algtegurite 3, 7, 11, 19, 23, 31,... puhul on meil tegu suhteliselt regulaarsete rütmidega, suurimate algtegurite 5, 13, 17, 29, 37,... puhul aga nähtustega, mis on rohkem müra tüüpi, sporaadilisemad ja järsemate frontidega, millele statistilisel keskmistamisel aga ikkagi selgub seose  $T_x = T_0 / N_x$  kirjeldusvõime.

Vaatleksime edasi kahte konkreetsemat näidet. Esiteks Päikese aktiivsuse rütmika, millega on seostatud ka inimeste sotsiaalset aktiivsust rääkimata väga mitmesugustest geofüüsikalistest ja bioloogilistest protsessidest. See rütm on seotud kahe unikaalse ruutarvule eriti lähedase arvuga, mis mõlemad on Re-tüüpi:

- 1)  $N_{50}(13) = 2^7 * 3^3 * 5 * 7 * 11 * 13$ , kõige viimane  $a^2 - 1$  tüüpi arv kogu Ramanujani arvude hulgas ( $N = 4159^2 - 1$ ,  $\sqrt{N} = n, 99988$ ,  $T_x = 11,86$  a).
- 2)  $N_{\cdot}(17) = 2^3 * 3^3 * 5 * 7 * 11 * 13 * 17$ ,  $a^2 - 3^2$  tüüpi arv,  $\sqrt{N} = n, 99895$ ,  $N/N_{\text{opt}} = 1,06$ ,  $T_x = 11,16$  a.

Teisele rütmile vastav  $N_x$  on küll ruutarvust kaugemal ja ka 6% Ramanujani optimumist väljas, kuid arvatavasti on tugevate allrütmide (harmooniliste) 2,79 a ja 255 p tõttu tema ilming siiski märgatav. Nende rütmide omavahelisel tuiklemisel tekib 189,7 a rütm.

Vaadates päikeseplekkide Wolfi arvude 11 a. rütmiga lainete tegelikke kujusid tekib paratamatult kahtlus, kas sin-cos funktsioonid on ikka kohased sellise perioodilise, kuid täiesti omanäolise loodusliku nähtuse kirjeldamiseks? Kõrgema amplituudiga aktiivsuslainetel on selgesti laine esimene front järsem. Seda ei saa väheste sin-cos tüüpi harmooniliste funktsioonide abil kirjeldada.

Veelgi rohkem süvendab seda kahtlust järgmine näide, mis pärineb meie labori põhitegevusest, vererõhu nn. kolmandat järku lainete ( $T \approx 11s$ ) uuringutest. Ramanujani arvudele tuginev klassifikatsioon kirjeldab seda rütmi eelkõige unikaalse tegurirohke arvu



$N_{(29)} = 2^6 \cdot 3^5 \cdot 5^2 \cdot 7^2 \cdot 11 \cdot 13 \cdot 17 \cdot 19 \cdot 23 \cdot 29$  ( $a^2 - 630^2$  tüüp,  $\sqrt{N} = n,9918$ , mis on optimumist väljas ainult 2%) kaudu, millele võib lisanduda ka tegurirohke arvu  $N_{123}(31) = 2^7 \cdot 3^3 \cdot 5^2 \cdot 7 \cdot 11 \cdot 13 \cdot 17 \cdot 19 \cdot 23 \cdot 29 \cdot 31$ ,  $\sqrt{N} = n,8796$  teatud mõju.

Vererõhu kõikumised kujutavad endast integraalset makroskoopilist nähtust, mis tuleneb mikrotsirkulatsioonis kapillaaride verevoolu kinni-lahti tüüpi, mürataolisest, kuid teatud tingimustes ilmselt  $T \approx 11s$  perioodi sisaldavast käitumisest.

Sellel rütmil on järgmised iseäralikud tunnused:

- 1) Keskmise sageduse suhteliselt suur püsivus ja karakterne väärtus inimpopulatsioonis (tavaliselt 9-13 s piirides).
- 2) Närvisüsteemi madalal aktiivsusel: rütmi sporaadilisus, ilmumine sporaadiliste tõusvate-kahanevate, ligikaudselt siinuseliste lainete gruppidega.
- 3) Närvisüsteemi kõrgema aktiivsuse korral (eksam, pingeline vestlus jms.): lainete amplituudi kasv ja pidevam esinemine.
- 4) Une ajal üksikud ebaregulaarsed selgesti järsema tõusuga ja aeglasema langusega lained.
- 5) Need unes esinevad harvemad asümmeetrilised vererõhu lained on sageli peaaegu täpselt vastasfaasis nahaverevoolu vasomotoorsete lainetega.

Kapillaaride nivoo rütmika on aga hoopiski sporaadilisem, mürarohkem, äkiliste kinni-lahti frontidega ja seejuures ka keskmiselt pikemate avatud-suletud perioodidega.

Püüe otsida nende keeruliste nähtuste vahelisi sõltuvusi on mind viinud arvamuseni, et just **Re-tüüpi rütmide jaoks** (teoreetiliselt on siin tegemist  $N(29)$  Re-tüüpi rütmidega) ei ole **sin-cos funktsioonid kohased ja on põhjust otsida sobivamat elementaarfunktsiooni.**

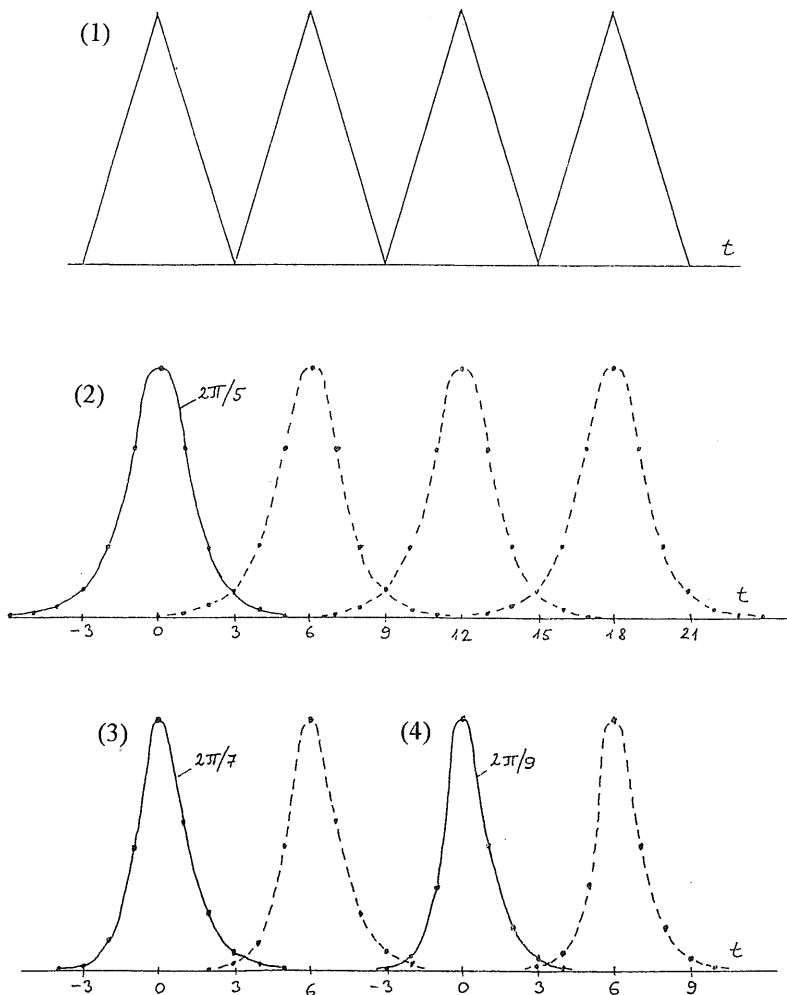
Need otsingud on viinud järelduseni, et sobivam on kasutada alljärgnevat Re-tüüpi ruut- ja kuupvõrranditest pärinevat funktsioone (joon. 1):

1) reaalsete ruutvõrrandite alusel:

$$y = 2/(x_1^t + x_2^t), \text{ kus } x_i = 2\cos(k\pi/5), \\ y = 1/\operatorname{ch}(0,962t);$$

2) reaalsete kuupvõrrandite alusel:

$$y = 3/(x_1^t + x_2^t + x_3^t), \text{ kus } x_i = 2\cos(k\pi/7) \text{ või } 2\cos(k\pi/9).$$



**Joonis 1.** Looduslike RE-rütmide lainete kuju lähendusi. (1) Lineaarne kolmnurkkõver, mis on levinumaks lähenduskõveraks ajakirjas "Cycles". (2)  $2\pi/5$  -Re-tüüpi ruutvõrrandi kõver ( $x_{1,2} = 2\cos k\pi/5$ ) põhikorduvusega (kuid mitte obligatoorse korduvusega) kuus sammu. (3)  $2\pi/7$  -Re-tüüpi kuupvõrrandi kõver ( $x_{1,2,3} = 2\cos k\pi/7$ ) põhikorduvusega 6 sammu. (4)  $2\pi/9$  -Re-tüüpi kuupvõrrandi kõver ( $x_{1,2,3} = 2\cos k\pi/9$ ) põhikorduvusega 6 sammu.

Põhjendused on järgmised:

Re-võrrandites ja Re-protsessides ei ole otsest sulutrajektoorseid või negatiivse tagasisidestuse põhjal tekkivat perioodsust, küll aga esineb Re-võrrandite rekurrentsetes arvujadades väikeste algtegurite perioodilise korduvuse nähtus, mis põhimõtteliselt saab olla aluseks ka nendes teatud tüüpi (varieeruvama, sporaadilisema, statistilise) perioodsuse tekkel. Ruutvõrrandi  $x^2 - x - 1 = 0$  puhul, s.o., Fibonacci/Lucas' arvude puhul saab perioodiline korduvus tekkida sammudel ( $n = t$ ) 6, 12, 18, 24, ..., s.o. Fibonacci arvuderea erandlike, väikeste algteguritega, ilma uute algarvudeta liikmete  $a_6 = 8$ ,  $a_{12} = 144$  jne. juures.

Arvan, et see rekurrentsetes ruutjadades esinev eriline väikeste algarvude kordumise fenomen ja üldse Re-protsesside statistilise korduvuse nähtus võib olla aluseks bioloogias laialt levinud siseseostele, millele on ammu otsitud seletust ja mida R.Sheldrake on nimetanud morfogeneetiliseks või morfiliseks resonantsiks.

On huvitav märkida, et ruutvõrrandi  $x^2 + px + q = 0$  selles p-q - kvadrantis, kus tema Bernoulli-trajektoolid on pidevad, annavad mõlemad kõige tähtsamad Im ja Re tüüpi võrrandid ( $x^2 - x + 1 = 0$  ja  $x^2 - 3x + 1 = 0$ ), mis on omavahel duaalsed, minimaalseks elementaarperioodiks sama väärtuse - **kuus samm**.

Selline lähenemine saab aga pakkuda meile sissejuhatavat arusaamist Im-Re duaalsusest ka protsesside perioodilise korduvuse tasandil, mis seni aegridade kirjeldamises puudub. Dualism täisarvuliste koefitsientidega täisperioodiliste Im ja Re ruutvõrrandite vahel eksisteerib aga veelgi üldisemal kujul:

$$x^2 + (2\cos(2\pi/n))x + 1 = 0$$

$$x^2 + (2\operatorname{ch}(nA/12))x + 1 = 0$$

$$x^2 + (2\operatorname{sh}(nA/12))x + 1 = 0.$$

Im-konstandi  $2\pi = 6,283..$  analoogideks on siin Re-konstandid

$$A = \ln((\sqrt{5} + 1)/2) = 5,774.. \text{ ja}$$

$$A = \ln(\sqrt{2} + 1) = 10,576.$$

Täisarvulise  $n$  puhul on Re-perioodid  $12/n$ , milledest tähtsamad on 12, 6, 4, 3, 2. Vastavate Re-ruutvõrrandite lineaarliikmete kordajad on Lucas' arvud  $b_n$  või Fibonacci arvud  $a_n\sqrt{5}$ . Paaris- ja paaritute arvudega kirjeldatud kõverate omavaheline ortogonaalsus loob siin aga teoreetilise aluse, et asuda Fourier' jt. analüüsidele vastavaid duaalseid Re-meetodeid otsima.

Kirjeldatud lähenemisel ilmneb võimalus seletada järsema fronti esinemist perioodilistes protsessides mikromaailma suunas liikudes. Asi on selles, et ruutarvule lähedased tegurirohked arvud võimaldavad sageli seda ruutarvlähedust veelgi suurendada, kuigi paratamatult

Ramanujani optimumist kaugenemise arvel. Vasomotoorika rütm ~11 s lubab seda  $630^2 = 396900$  korda. (Rütmini 50,7 päeva, ruutarvläheduseni  $n,999987$ , kus kaugenemine optimumist on 40 korda). See on küll erandlikult suur kordsus, kuid väiksemates piirides on nähtus sage. Selline nähtus võib olla seotud jõuliste, kuid lühiaegsete protsesside tekkega, mis on võimelised kommuteerima bistabiilseid süsteeme (klassiarvuga 2 või enam Re-süsteeme) ühest stabiilsest seisundist teise. Arvan, et loomse vereringe kapillaarvõrgustikus ongi see teoreetiline võimalus ilmnenu.

Avaneb ka võimalus integraalsete makroskoopiliste rütmide sporaadilisuse ja esimeste frontide järskuse seletamiseks, mida vererõhus väga selgesti täheldatakse une ajal. (Nii allrütmide eksistentsi tõttu, kuna neil on omadus ilmuda põhirütmiga harjast  $1/4$ ,  $1/9$ ,  $1/16$ .. perioodi võrra varem, kui ka kuupvõrrandite Re-kõverate kaju kaudu, vt. kõveraid 3 ja 4 joonisel 1).

Lõpuks tuleks veel märkida, et kuna ruutvõrrand kirjeldab väga universaalselt looduslikke protsesse, ei maksa sugugi üllatuda, kui sama tüüpi matemaatilised kõverad leiavad rakendust ka statistiliste jaotuskõveratena, nagu seda on välja pakutud ELUŠi Teoreetilise Bioloogia XI Kevadkoolis (Reeben 1985).

### Viidatud kirjandus

1. Ramanujan, S., (1915). Highly composite numbers. Proc. London Math. Soc., 2, XIV, 347-419.
- Reeben, V., (1985). Tihedusfunktsioonide uue perekonna genereerimisest algebraaliste võrrandite lahendite baasil. Kogumikus: Stohhastilised mudelid bioloogias. Schola biotheoretica XI, Tartu, 150-158.

## KROONILISTE MITTESPETSIIFILISTE HAIGUSTE RISKITEGURITE UURIMISE METOODIKAST

R. Birkenfeldt  
Tartu Ülikooli Sisekliinik

Krooniliste mittespetsiifiliste haiguste riskitegurite uurimiseks korraldatakse elanike epidemioloogilisi uuringuid. Võrreldakse teatud tunnuste sagedust tervetel ja haigetel. Igat epidemioloogilist tegurit ei tohi aga samastada riskiteguriga. Riskitegurite eristamisele ja konkretiseerimisele peab eelnema pikaajaline perspektiivne epidemioloogiline uuring. See aga on küllalt suuremahuline, meetoodiliselt komplitseeritud ja väga kulukas.

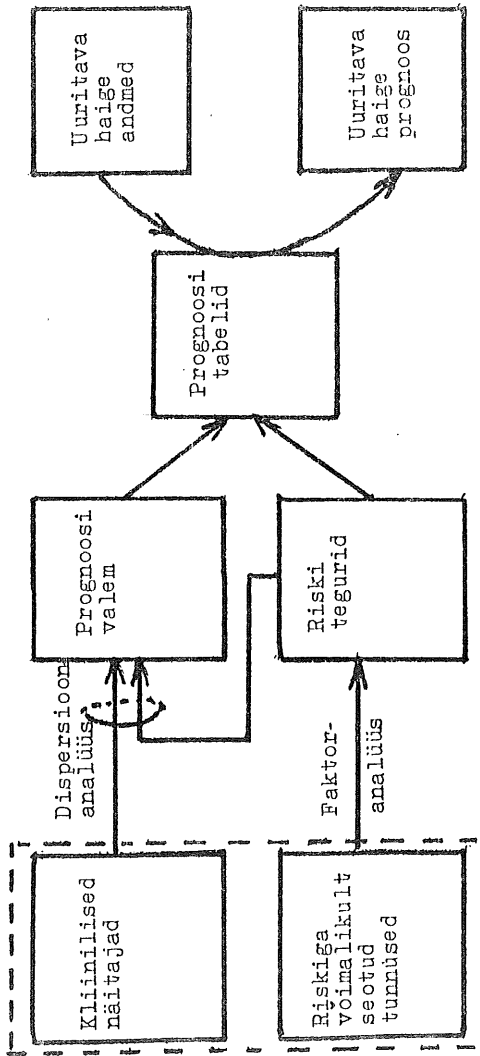
Järgnevalt tutvustame meie poolt reumatismi riskitegurite uurimiseks kasutatud meetoodikat, mille puhul algandmed saadi ainult haigete pikaajalise prospektiivse jälgimisega. Nüüdisaja andmetel on reumatism sidekoe süsteemne põletikuline haigus protsessi eelislokatsiooniga südamevereringesüsteemis (südame klapiirikete teke), haaratud võivad olla ka liigesed ja teised elundid. Haigus võib retsidiveeruda.

Jälgisime pikka aega (15-25 aastat) 1310 reumahaiget Eesti saartel. Matemaatilisele töötlusele võeti 165 riskiga võimalikult seotud epidemioloogilist tunnust ja kliinilist näitajat. Uurimise käik on toodud Joonisel 1. Riskiga võimalikult seotud tunnuste vahel oli hulgaliselt positiivseid ja negatiivseid korrelatsioone. Faktoranalüüs tõi välja viis reuma riskitegurit ja neid moodustavad komponendid. Need olid:

1. krooniline koldenakkus (nakatus A-grupi beeta-hemolüütilise streptokokiga, krooniline tonsilliit, selle hilinevad saneerimine);
2. olmetegur (ebasoodsad olme- ja tööttingimused, A-beeta-hemolüütilise streptokoki kandlus ja krooniline tonsilliit perekonnas);
3. soolis-ealine tegur (lapse- ja noorukiiga, eriti naissugu);
4. päriulik eelsoodumus (perekondlik levik, geneetilised tegurid);
5. puudulik teisene profülaktika.

Riskitegurite ja kliiniliste näitajate dispersioonanalüüs näitas, et esmane reumasse haigestumine oli seotud 1., 2., 3. ja 4. riskiteguriga, haiguse retsidiivi teke 1., 2., 3., 4., ja 5. riskiteguriga, südame klapiirikete

Prospektiivse  
uuringu  
andmed



Joonis 1.

Kroonilise mittespesiifiliste haiguste  
riskitegurite uurimine

formeerumine 1., 2. ja 3. riskiteguriga ning kroonilise südame-vereringe puudulikkuse teke ja haige invaliidistumine 1., 2. ja 5. riskiteguriga. Ühtlasi saime reumasse haigestumise riski, haiguse kulu ja lõppe prognoosi valemid, millede alusel koostati prognoosi tabelid.

Prognoosi tabelite alusel saab määrata iga uuritava isiku esmast riski haigestuda reumasse (riskitegurid 1, 2 ja 3), juba reumasse haigestunud retsidiivide tekke riski (riskitegurid 1, 2, 3 ja 5), võimaliku retsidiivi tekke aega (4. riskitegur), südame klapiरिके tekke riski (riskitegurid 1, 2 ja 4) ning riski kroonilise vereringe puudulikkuse arenguks (riskitegurid 1, 2 ja 5). Seega on prognoosi tabelite abil võimalik määrata reumasse haigestumise riski, haigestunudel aga prognoosida võimalikku haiguse kulgu ja lõpet. Selle alusel on võimalik ühelt poolt välja selgitada reumast ohustatud, teiselt poolt võtta diferentseeritult kestvale jälgimisele erinevaid reumahaigete kontingente. Seega avaneb võimalus kasutada riskitegureid haiguse esmasel ja teisesel profülaktikas.

Lõpetuseks annaksime omalt poolt riskiteguri uue laiendatud määratluse. Haiguse riskitegur on bioloogiline, ökoloogiline või sotsiaalne tunnus, mille suurema väärtuse puhul võib haigus tekkida suurema tõenäosusega või kulgeda raskemalt. Riskitegurite ebasoodne mõju haiguse kulule ja lõppele on eriti tähelepanev krooniliste, perioodiliselt ägenevate haiguste puhul.

## MITMEMÕÕTMELINE STATISTILINE ANALÜÜS ANTROPOMEETRIILISTE PROBLEEMIDE LAHENDAMISEL

H. Kaarma, L. Saluste, G. Veldre, M. Thetloff  
Tartu Ülikool, Füüsilise Antropoloogia Keskus

Käesolevas artiklis analüüsitakse koostöö tulemusi Tartu Ülikooli antropoloogide ja matemaatilise statistika spetsialistide vahel. Koostöö on kestnud alates 1974 aastast ja on puudutanud peaaegu kõiki antropoloogia valdkondi, millega Tartu Ülikoolis tegeldakse. Allpool kirjeldame üksikasjalikumalt üht nendest uurimustest.

### Materjal ja meetodika

Üheks põhiliseks uurimismaterjaliks on olnud noorte eestlannade (18-22a.) kontingent (uuritud 670 naist). Mõõdeti 37 kehamõõtu ja 10 nahavolti. Mõõtmised testiti klassikalisel R. Martini meetodil (Martin, 1928) ja analüüsiti mitmemõõtmelise statistilise analüüsi programme kasutades (Tiit, 1975). Konsultandiks oli prof. E.-M. Tiit.

### Töö tulemused

Uurides kehaehitusstruktuuri leiti, et noore naise kehaehitus kujutab endast lineaarselt hästikorreleeruvat tunnuste süsteemi, kus juhtivad tunnused on pikkus ja kaal. Faktoranalüüsi puhul määravad need tunnused ülejäänud tunnuste varieeruvusest keskmiselt 50%. Ülejäänud 50% jääb tunnuste individuaalse varieeruvuse arvele. Jääkkorrelatsioonide kasutades oli võimalik tõestada, et kogu kehaehituse terviksüsteem laguneb, kui elimineeritakse pikkuse ja kaalu mõju üksiktunnustelt (Kelder, 1977). Kui kaal ja pikkus on juhtivad tunnused, siis peaks ülejäänud tunnuseid uurima mitte isoleeritult, vaid vastavates pikkus-kaalu klassides. Selleks moodustati pikkus-kaalu klassifikatsioon ( $5 \times 5$  klassi, kumbki tunnus varieerub vahemikus  $-2,5\sigma \dots +2,5\sigma$ ) ning arvatati kõigi ülejäänud tunnuste aritmeetilised keskmised kõigi klasside jaoks. Nendes klassides arvatatud tunnuste keskvaartuste erinevust uuriti S-testiga (Kelder, 1978). Leiti, et kõigis pikkusklassides



toimub kaalu suurenemisel statistiliselt oluline laius-sügavus- ja ümbermõõtude suurenemine. Jälgides muutusi viies kaaluklassis, ilmnesid pikkusmõõtude suurenemisel oluline laius-sügavus- ja ümbermõõtude vähenemine. Niisugune süsteemsus lubas kasutada ka keha proportsioonide uurimiseks sama pikkus-kaalu klassifikatsiooni. Moodustati 124 kehaproportsiooni indeksit ja uuriti nende muutumist analoogselt alg tunnustega. Sellises neljamõõtmelises süsteemis ilmnes sama süsteemsus, kusjuures indeksid, mis olid moodustatud laius-sügavus- ja ümbermõõtudest ning pikkusmõõtudest, muutusid palju rohkem kui indeksid, mis olid moodustatud üht ja sama liiki mõõtudest.

On huvitav märkida, et pikkus-kaalu klassides ilmnesid ka keha koostise süsteemsed muutused. Nii oli võimalik demonstreerida, et nii keha tiheduse kui ka nahaaluse rasvkoe hulga muutused on olulises osas määratud uuritavate pikkus-kaalu vahekordadega.

Keha struktuuri terviklikkusest andsid ülevaate ka uuringud, mis tehti selgitamaks iga üksiku kehatunnuse prognoosimise võimalusi ülejäänud tunnuste abil. Valiti 3 argumenttunnuste kombinatsiooni: 1) pikkus ja kaal, 2) pikkus ja kaal ning veel üks uuritavate tunnustega väga tugevasti seotud tunnus, 3) juhuslikult valitud 10 kehatunnust ilma pikkuse ja kaaluta. Kõigil juhtudel oli prognoos statistiliselt oluline. Uuritavate tunnuste varieeruvust oli võimalik kirjeldada 50-83% ulatuses.

Pärast üldkontingendi uurimist selgitati põhilised kehastruktuuri iseärasused ka pükniliste ja leptosoomsete naiste gruppides (Kretschmer, 1961). Sel eesmärgil analüüsiti 32 püknilise ja 39 leptosoomse naise antropomeetrilisi andmeid. Need naised pärinesid üldkontingendist, kuid olid somatoskoopiliselt tüpiseeritud enne antropomeetrilise mõõtmise teostamist.

Nende 2 grupi jaoks arvutati kõigi üksiktunnuste ja 124 indeksi keskmised väärtused ning määrati erinevuste statistiline olulisus. Tulemusena oli võimalik väita, et muutused, mis ilmnevad püknikute kehaehituse struktuuris, meenutavad tendentsi, mis ilmneb, kui üldkontingendis uuritavate kehakaal järkjärgult suureneb ühe ja sama pikkusklassi piires. Ja vastupidi, leptosoomsete naiste iseärasused meenutavad muutusi üldkontingendil, kui kehapiikkus järkjärgult suureneb ühe ja sama kaaluklassi piires.

Analoogsed muutused kehaehitusproportsioonides üldkontingendil, püknikute ja leptosoomsete grupis andsid põhjust järeldada, et püknikute ja leptosoomsete naiste kehaehitusstruktuur baseerub samuti erinevatel pikkus-kaalu vahekordadel. Seda hüpoteesi kontrolliti 2-mõõtmelise statistilise analüüsi abil, mis viidi läbi järgmiselt: valiti 80 indeksit ning uuriti nende seost ühelt poolt Rohreri indeksiga ja

teiselt poolt kehatüübi näitajatega. Selgus, et ka siin valitseb sama seaduspärasus. Järelikult püknikute ja leptosoomsete naiste kehaehitusstruktuur ei kujuta endast mingit iseäralikku süsteemi.

Järgmiseks uuringu teemaks oli selgitada, kas pükniliste ja leptosoomsete naiste kehalisi iseärasusi on võimalik kasutada üldkontingendi tüpiseerimiseks. Selleks valiti 13 indeksi, mille väärtused pükniliste ja leptosoomsete naiste grupis kõige rohkem erinesid. Diskriminantfunktsiooni kasutades oli võimalik jaotada üldkontingendi naised 3 rühma: püknomorfsed, keskmised ja leptosoomsed. Äärmistes rühmades esinesid samad iseloomulikud erinevused, mis ilmsesid somatokoopilisel määratlendud pükniliste ja leptosoomsete naiste juures.

Leitud tulemustest võib järeldada, et inimitüüpide kogu varieeruvus ei koosne üksikutest diskreetsetest somatotüüpidest, vaid tunnuste vahel esinev pidev mittevastavuse põhimõte viib nn. puhaste tüüpide (püknikute ja leptosoomsete) eristamisele varieeruvuse perifeerias. Juhivateks tunnusteks on siin pikkus ja kaal.

## Diskussioon

Teoreetilise meditsiini ja teoreetilise bioloogia üheks lõppeemärgiks on haige ja terve inimese konstitutsionaalsete iseärasuste tundmaõppimine (Frankenberg, 1980, Robinow, 1982, Drobný, 1986, Wienker, 1984). Et selgitada, missugune on keha terviklik mudel, mille sees oleks võimalik individuaalseid morfoloogilisi ja funktsionaalseid iseärasusi süstematiseerida ja võrrelda, selleks on vaja antropoloogilisi statistilisi uuringuid (Kaarma, 1981, 1994, 1995).

Tänu koostööle TÜ Arvutuskeskusega analüüsi eesti noorte naiste kehaehitusstruktuuri olemus, kasutades sealjuures mitmemõõtmelise statistilise analüüsi meetodeid. Uurimus baseerus tuntud antropoloogilistel põhimõtetel kehatunnuste omavahelisest korreleeruvusest (Akinshchikova, 1969, Bashkirov, 1962) ja Kretschmeri somatotüpiseerimismeetodil (Kretschmer, 1961).

Kõige lihtsam pikkus-kaalu klassifikatsioon on kasutatav bioloogias ja mitmetes meditsiini distsipliinides selleks, et analüüsida kehaehituse seost mitmete meditsiiniliste näitajatega.

Tulemustest selgus, et kehaehitusstruktuuris pole olulisi erinevusi üldkontingendi ja pükniliste ning leptosoomsete naiste vahel. See annab aluse järeldada, et somatotüpiseerimiseks on võimalik kasutada erinevaid meetodeid (Heat, Carter, jt 1966). Selleks, et võrrelda omavahel kehastruktuuri erinevate meetoditega tüpiseeritud materjalidel, on siiski vaja läbi viia võrdlevad mitmemõõtmelise statistilise analüüsi uuringud.

## Viitekirjandus

1. Akinshchikova, G., 1969, Inimese kehaehitus ja organismi reaktiivsus. Leningrad. (vene k.)
2. Bashkurov, P., 1962, Õpetus inimese füüsilisest arengust. Moskva. (vene k.)
3. Drobny, L., 1986, Beitrag der Anthropologie im Rahmen des medizinischen Skrining. Anthropologie (CSSR), 24, 2-3, 139-145
4. Frankenberg, R., 1980, Medical anthropology and development: A theoretical perspective. Soc. Sci. and Med, 14, 4, 197-207
5. Heath, B., Carter, J., 1966, A comparison of somatotype methods. Amer. J. of Phys. Anthropology, 24, 87-99
6. Kaarma, H., 1981, Naiste antropomeetriliste tunnuste süsteemi mitmemõõtmeline statistiline analüüs. Tallinn. (vene k.)
7. Kaarma, H., Saluste, L., Veldre, G., Thetloff, M., Loolaid, V., Kuivjõgi, K., 1994, Systematitition of anthropometrical data for comparative statistical estimation of women's individual nutritional state. Intern. J. of Anthropology, 9, 3, 208
8. Kaarma, H., 1995, The role of pelvic measurements in women's whole body structure. Papers on Anthropology VI. Tartu, 99-106
9. Kelder, T., 1977, Vähimruutude meetodi mitmemõõtmeline variant arvutile "Minsk-32". Arvutuskeskuse toimetised, TRÜ, 40, 65-77. (vene k.)
10. Kelder, T., 1978, Scheffe S-meetod keskmiste võrdlemiseks. Programme kõigile, Tartu, 14, 16-24
11. Kretschmer, E., 1961, Körperbau und Character. Berlin-Göttingen-Heidelberg.
12. Martin, R., 1928, Lehrbuch der Antropologie, Jena
13. Robinow, M., 1982, Clinical applications of physical anthropology. Yearb. Phys. Anthropology, 25, New York, 169-179
14. Tiit, E., 1975, Statistilise andmetöötluse üldprintsipiibid. Programme kõigile, Tartu, 9, 4-27
15. Wienker, C., 1984, The emergence of biomedical anthropology and its implications for the future. Amer. J. Phys. Anthropology, 64, 2, 141-146

## BIOMEETRIA ÕPETAMINE ARSTITEADUSKONNAS

K. Põlluste

Tartu Ülikool, Tervishoiu Instituut

Tartu Ülikoolis on õpetatud biomeetriat arstiteaduse üliõpilastele juba enam kui 100 aastat. Selles valdkonnas on kõige tuntum professor B. Körber kui statistilise suuna esindaja tervishoius. Ta kirjutas kaks uurimust Tartu linna hügieeniliste ja demograafiliste olude kohta ning ülevaate Tartumaa mitme kihelkonna demograafiast. 1880ndail aastail valmis tema juhendamisel väitekirju, mis olid pühendatud demograafiliste protsesside uurimisele Liivi- ja Eestimaa kubermangus [1]. Alates 1895. aastast, kui asutati ennetava meditsiini erialade õpetamiseks Tervishoiu Instituut, kuni tänaseni\*, on meditsiinistatistika õpetamine kuulunud nimetatud instituudi kohustuste hulka. Aastatel 1945-1990 toimus meditsiinistatistika õpetamine üleliiduliste programmide järgi sotsiaalhügieeni ja tervishoiuorganisatsiooni kursuse raames. Õppetöö toimus raviosakonnas VIII ja IX semestril ning lõppes eksamiga. Nimeetatud õppeaine kuulus kuni 1993. aastani ka arstiteaduskonna lõpu-eksamite hulka.

1990/1991 õppeaastal hakati arstiteaduskonnas järk-järgult üle minema uutele õppeplaanidele, mille kohaselt suurendati ennetava meditsiiniga seotud õppeainete mahtu. Sellega seoses vaadati läbi ka sotsiaalhügieeni ja tervishoiuorganisatsiooni programm ning kursus lahutati kaheks iseseisvaks osaks: biomeetriaks ja tervishoiukorralduseks.

Praegu õpetatakse arstiteaduskonnas biomeetriat raviosakonnale V semestril, aine maht on 64 tundi (32 tundi loenguid, 32 tundi seminare) ning kursus lõpeb hindelise arvestusega. Stomatoloogiaosakonnale õpetatakse biomeetriat III semestril ja vanemmeditsiinidõdede osakonnale II semestril. Viimatinimetatud osakondades on aine maht 32 tundi (16 tundi loenguid ja 16 tundi seminare) ning kursus lõpeb arvestusega. Alates 1994. aastast lõpeb studium arstiteaduskonnas

---

\* vt. Toim. märkus lk. 58

ühise kompleksse lõpueksamiga, milles osa küsimusi on biomeetria valdkonnast.

TÜ Tervishoiu Instituudi koosseisu kuulub neli õppetooli:

- hügieeni ja keskkonnatoksikoloogia õppetool,
- töötervishoiu õppetool,
- tervishoiukorralduse õppetool,
- biomeetria ja epidemoloogia õppetool.

Viimatinimetatud õppetool tegelebki biomeetria õpetamisega. Õppetooli juhib Tartu Ülikooli ning Eksperimentaalse ja Kliinilise Meditsiini Instituudi ühisprofessor Mati Rahu.

Biomeetria õppeprogrammi koostamisel on järgitud Maailma Tervishoiuorganisatsiooni (MTO) juhiseid. 1978. aasta märtsis toimus MTO ja Rahvusvahelise Epidemioloogia Assotsiatsiooni (REA) algatusel Pakistanis regioonidevaheline konverents statistika õpetamisest arstiteaduse üliõpilastele. Konverentsil käsitletud probleemide kohta avaldati otsus "*The Successful Teaching of Statistics to Every Medical Student*" [2], milles esitatud seisukohad on võetud aluseks Tartu Ülikooli arstiteaduskonna biomeetria õppeprogrammi koostamisel. Samuti on eeskujuks võetud mitme Euroopa maa ülikoolide õppeprogramme.

Iga õppeprogrammi koostamisel tuleb lähtuda tegelikest vajadustest. Eriti oluline on teadvustada üliõpilastele just selliste õppeainete tähtsust, mis ei ole ainult arstiteaduslikud.

Põhjused, miks biomeetriat tuleb õpetada igale tulevasele arstile, on järgmised:

1. Statistikaadmised on vajalikud, et mõista loogilisi põhiprintsiipe, millel põhinevad diagnoosimine, prognoos ja haige raviga seotud arstlikud otsused.
2. Arst peab oskama interpreteerida haigevoodi ääres tehtud tähelepanekuid ja laboratoorsete uuringute tulemusi tulenevalt nende mitmekesisusest.
3. Arst peab teadma statistilisi ja epidemioloogilisi andmeid ravitava haiguse kohta ning nendest aru saama, et anda oma patsiendile nõu, kuidas kahjulikke tegureid vältida või nende osa oma elus vähendada.
4. Praktiseerivad arstid, kes on tervise- ja tervishoiu statistika algandmete genereerijad, peavad teadma, kuidas neid andmeid oma igapäevatoos kasutada ning kuidas neid kasutatakse kogu riigi tervishoiu korraldamisel.
5. Arst peab teadma, kuidas interpreteerida ja teha järeldusi statistilistest andmetest, mis iseloomustavad elanikkonna tervise-

probleeme ning nägema võimalusi, kuidas neid probleeme lahendada.

6. Statistika aitab üliõpilastel arendada kriitilist ja deduktiivset mõtlemist, mida nad saavad kasutada nii õpingutes kui ka tulevases praksises.
7. Statistika on oluline teadustöö osa. Ilma statistikast tundmata ei suuda arst aru saada teaduslikes ajakirjades esitatud töödest, samuti ei ole võimalik publitseerida statistiliselt töötlemata andmeid.

Vastavalt eeltoodud vajadustele on MTO ja REA poolt välja pakutud biomeetria õpetamise põhiprogramm, mis koosneb kolmest osast [2]:

1. statistika põhialused ja meetodid;
2. tervise statistika;
3. statistika kliinilises meditsiinis.

TÜ Arstiteaduskonnas on põhitähelepanu pööratud kahele esimesele osale. Põhjus on selles, et biomeetria kursus on ette nähtud studiumi esimeses pooles ning seostub ennekõike profülaktilise meditsiiniga. Ka ei ole üliõpilastel siis veel küllaldaselt teadmisi kliinilistes erialades.

Edaspidi on kavas täiendada biomeetria kursust ka kliinilise meditsiini statistika osas. Seda valdkonda on kohane õpetada viimasel kursusel. Kuna lõpukursusel on õppetöö maht piisavalt suur, tuleks kõne alla sellesisuline valikkursus.

### **Statistika põhialused ja meetodid**

Arstiteaduse üliõpilane ei ole matemaatik ning temal ei ole vaja põhjalikult õppida matemaatilist statistikat. Ka ei ole vajalik valemite päheõppimine. Oluline on osata statistikat rakendada meditsiinis, teada, millal ühte või teist valemit kasutatakse ning kust käsiraamatust vastava valemi võib leida. Mistahes statistika põhimõtteid õpetades peavad need olema seostatud arstiteadusega.

Statistika põhialuseid õpetatakse arstiteaduse üliõpilastele järgmise kava alusel:

1. Sissejuhatus: biomeetria mõiste, ajalugu, statistika osa inimese bioloogias ja meditsiinis.
2. Andmete kogumine ja mõõtmise ulatuse määramine.
3. Andmete esitamine tabelite ja arvjoonistena.
4. Andmete mitmekesisus ja tunnuste jaotumine.
5. Keskmiste mõõtmine: peamised mahu- ja asendikeskmised (aritmeetiline keskmine, mood, mediaan), nende kasutamine ja interpreteerimine.

6. Varieerumise mõõtmine: amplituud, dispersioon ja standardhälve, nende kasutamine ja interpreteerimine.
7. Tõenäosuse mõiste.
8. Väljavõteline uuring: valimi mõiste, selle saamise võimalused, randomiseerimine, valimi esinduslikkuse (representatiivsuse) tagamine, esindusviga.
9. Statistilised järeldused: olulisustestid, nende kasutamine ja tulemuste interpreteerimine.
10. Seose ja põhjuslikkuse mõiste: regressioon ja korrelatsioon, seose mõõtmine, regressiooni- ja korrelatsioonikordajate kasutamine ja interpreteerimine.

Nimetatud küsimusi käsitletakse raviosakonnas 28 tunni ulatuses, stomatoloogia- ja vanem-meditšiiniõdede osakonnas 16 tunni ulatuses. Statistika põhialuste õpetamiseks on kasutatud peamiselt U. Mereste "*Statistika üldteooriat*" [3] ja Saksamaal välja antud raamatut "*Medizinische Biometrie*" [4], aga ka mitmeid meil kättesaadavaid USA ülikoolide statistikaõpikuid mittematemaatikutele [5,6]. Kavas on välja anda ka eestikeelne biomeetria praktikumi juhend.

### Moodne epidemioloogia

Alates 1991. aastast on biomeetria kursusesse uus osana sisse võetud moodne epidemioloogia uue osana, mille maht on raviosakonnas 12 tundi (6 tundi loenguid + 6 tundi seminare), stomatoloogia- ja vanem-meditšiiniõdede osakonnas 2 tundi (loeng). Kursuse omandamisel on abiks A. Ahlbomi ja S. Norelli "*Sissejuhatus moodsasse epidemioloogiasse*" [7]. Moodsa epidemioloogia kursuses käsitletakse järgmisi teemasid:

1. Moodsa epidemioloogia mõiste ja ajalugu.
2. Moodsa epidemioloogia alajaotused (kirjeldav, analüütiline, eksperimentaalne).
3. Kirjeldav epidemioloogia: haigestumuse, suremuse, levimuse ja elulemuse mõõtmine.
4. Analüütiline epidemioloogia: haiguse põhjuse ja riskitegurite uurimine, tegurlähtene e. kohortuuring, suhtelise riski mõõtmine, juht-kontrolluuring.
5. Eksperimentaalne epidemioloogia: uuritava teguri ettekatsetatud muutmise mõju haigestumisele, epidemioloogilise eksperimendi planeerimine ja läbiviimine, pimemetodi kasutamine.

Lisaks eeltoodule on asjahuvilistel alates 1994/95 õppeaastast võimalik kuulata professor Mati Rahu valikkursust "Moodne epidemioloogia edasijõudnutele", kus käsitletakse moodsa epidemioloogia põhimõisteid ning ökoloogilisi, mikroepidemioloogilisi ja kohortuuringuid.

## Tervise statistika

MTO ja REA ekspertide soovitusel võib tervise statistikat õpetada kas iseseisva biomeetria kursuse osana või epidemioloogia ja hügieeni kursuse osana [2]. Tartu Ülikooli arstiteaduskonnas kuulub tervise statistika biomeetria kursuse juurde ning selle maht on raviosakonnas 24, stomatoloogia- ja vanem-mediitsiiniõdede osakonnas 14 tundi. Tervise statistika ja sotsiaalmeditsiini küsimusi käsitletakse järgmise valiku alusel:

1. Arst kui haigestumise ja suremuse statistika andmete tootja ja kasutaja. Haigestumis- ja suremusnäitajate kasutamine tervishoiukorralduses.
2. Populatsioonidünaamika: rahvaarv, struktuur, rahvastiku liikumise näitajad (mehaaniline liikumine - migratsioon, loomulik liikumine - elanikkonna juurdekasv), sündimus-, suremus- ja fertiilsusnäitajad.
3. Rahvaloendus, selle olemus ja tähtsus, läbiviimise kord. Sündide, abielude ja lahutuste registreerimise kord.
4. Suremuse statistika: surmade registreerimise kord, suremuse vanuseline ja sooline struktuur, standardiseeritud suremuse näitajad. Rahvusvaheline haiguste ja surmapõhjuste klassifikatsioon.
5. Haigestumise statistika: haigestumise ja puuete registreerimine, haigestumise mõõtmine (esmane haigestumus e. *incidence*, üldhaigestumus e. *prevalence*, levimus- ja elulemuskordajad). Registrid tervishoius (haigusteregistrid, abordiregister). Peamisteks surmapõhjusteks olevate haiguste (südame-veresoonkonna haiguste, pahaloomuliste kasvajate, vägivaldse surma), alkoholismi ja tuberkuloosi sotsiaalne ja meditsiiniline tähendus (nende osa haigestumise, invaliidistumise ja ajutise töövõimetuse põhjustena). Ajutise töövõimetuse uurimine.
6. Tervishoiuasutuste statistika. Haiglaväliste ja statsionaarsete raviasutuste töö mahu ja kvaliteedi hindamine, andmete kasutamine tervishoiuvõrgu planeerimisel (seda osa käsitletakse X semestril tervishoiukorralduse kursuse raames).



## Statistika kliinilises meditsiinis

Nagu eelpool öeldud, nimetatud biomeetria valdkonda iseseisvalt Tartu Ülikoolis ei käsitleta, kuna V semestril ei ole üliõpilased veel piisavalt õppinud kliinilisi erialasid. Seda valdkonda peetakse kõige raskemaks, kuid samas ka üliõpilastele kõige huvitavamaks.\*

Selles valdkonnas soovivad MTO ja REA eksperdid käsitleda järgmisi probleeme [2]:

1. Mõõtmistulemuste varieerumine meditsiinis, normi ja patoloogiliste väärtuste piiri hindamine (vaadeldakse tunnuste varieerumisest lähtudes).
2. Diagnoosimise, ravi ja prognoosi statistilised alused.
3. Diagnoosimine: kliiniliste mõõtmistulemuste (sümptoomide ja laboratoorsete analüüside tulemuste) võrdlemine nimetatud tunnuste jaotumisega kogu populatsioonis.
4. Prognoos: haigestumuse ja suremuse analüüs ning saadud näitajate interpreteerimine antud haiguse kontekstis.
5. Ravi: ravi- ja profülaktikaplaani koostamine, loodetava efekti hindamine, järelduste interpreteerimine.
6. Statistiliste ja epidemioloogiliste uuringute eetilised probleemid.

Pärast kohustusliku biomeetriakursuse omandamist on arstiteaduskonna doktorantidel võimalik täiendada oma statistikateadmisi TÜ Matemaatilise Statistika Instituudi poolt korraldatud statistilise andmeanalüüsi loengukursusel (32 tundi), millega kaasnevad ka praktilised tööd arvutiklassis 32 tunni ulatuses.

## Kirjandus

1. Tartu Ülikooli ajalugu, II kd. Tallinn, 1982, lk. 249.
2. The Successful Teaching of Statistics to Every Medical Student. The Report of the Interregional Conference on Teaching Statistics to Medical Undergraduates. Karachi, 1978.
3. Mereste, U. Statistika üldteooria. Tallinn, 1975.
4. Heinecke, A., Hultsch, E., Repges, R. Medizinische Biometrie. Springer-Verlag Berlin Heidelberg, 1992.
5. Freedman, D., Pisani, R., Purves, R. Statistics. 1980.
6. Anderson, T. W., Sclove, S. L. Introductory Statistical Analysis. 1974.
7. Ahlbom, A., Norell, S. Sissejuhatus moodsasse epidemioloogiasse. Tallinn, 1993.

---

\* Alates 1995/96 õppeaastast loeb nimetatud kursust valikkursusena MSI teadur Krista Lapp, vt. toimetuse märkus lk. 58-59

## Toimetuse märkus.

### **Meditisiinistatistika õpetamisest TÜ Arstiteaduskonnas Matemaatilise Statistika Instituudi poolt**

Lisaks eelmises artiklis nimetatud statistilise andmeanalüüsi loengukursusele doktorantidele on MSI 1992/93 õppeaastal õpetanud meditsiinistatistikat arstiteaduskonna VI kursuse üliõpilastele 64 tunni ulatuses (32 tundi loenguid ja 32 tundi praktilisi töid arvutiklassides). Alates 1994/95 õppeaastast on MSI õpetanud sama ainet arstiteaduskonna soovil teise nimetuse all (meditsiiniline infoteadus) ja lühendatud kujul (32 tunni ulatuses - 16 tundi loenguid ja 16 tundi praktilisi töid arvutiklassides) kõigile arstiteaduskonna I kursuse üliõpilastele.

Viimatinimetatud kursusesse kuuluvad järgmised teemad:

1. Andmeanalüüsi ülesanne. Ülesannete põhitüübid. Üldkogum, valim.
2. Andmestik. Tunnus, tunnuse tüübid. Andmete kodeerimine. Objekt-tunnus-maatrics. Puuduv väärtus. Andmestiku kontrollimine. Vead. Andmete teisendamine.
3. Üksiktunnuse kirjeldav analüüs. Tunnuse (empiirilise) jaotus, selle esitus tabelina ja illustreerimine graafikutega. Tunnuse empiirilised arvarakteristikud. Variatsioonirida, selle põhjal leitavad karakteristikud.
4. Teoreetilised jaotused: binoomjaotus, Poissoni jaotus, normaaljaotus ja sellest tuletatud jaotused.
5. Valimi kasutamine järelduste tegemiseks üldkogumi kohta. Usalduspiiride mõiste. Usaldustõenäosus ja usaldusnivoo. Keskväärtuse usalduspiirid.
6. Statistiliste hüpoteeside kontrollimine. Sisukas ja nullhüpotees. I ja II liiki viga. Olulisuse nivoo. Hüpoteesid keskmise kohta. Kahe keskmise võrdlemine t-testi abil. Protsentide võrdlemine.
7. Tunnuspaari analüüs. Sagedus- ja jaotustabel, marginaaljaotus, tinglik jaotus. Tunnuspaari seose kirjeldamine. Seose tüübid ja seose karakteristikud.
8. Ühefaktoriline dispersioonanalüüs, lihtne regressioonanalüüs. Korrelatsioonanalüüs. Lisaks toimub iga teema kohta praktiline õppus arvutiklassis, kasutusel on statistikapaketid STATGRAPHICS, SPSS, SYSTAT ja statistilise andmetöötluse osa paketi EXCEL.

Alates 1995/96 õppeaastast on MSI planeerinud valikkursuse "Kliiniliste eksperimentide statistika" (Aine maht: 16 loengutundi, lektor:

Krista Lapp). Selle kursuse jooksul tulevad käsitlemisele järgmised teemad:

1. Statistika kasutusvaldkondadest meditsiinis, kliiniliste eksperimentide vajalikkusest (olukorrad, tingimused).
2. Eksperimentide põhiligid. Patsientide valik.
3. Esimest liiki eksperiment, selleks vajalik patsientide arv ja muud eeldused. Tulemuste analüüs.
4. Teist liiki eksperiment. Võimalikud protseduurid patsientide arvu optimeerimiseks. Tulemuste analüüs.
5. Kolmandat liiki eksperiment. Patsientide arvu kindlaksmääramine.
6. Patsientide rühmitamine juhusliku valiku alusel. Muud meetmed nihete ärahoidmiseks.
7. Eksperimendi võimalikud lõpp-punktid, nende plussid ja miinused.
8. Ülevaade elukestvusandmete analüüsi meetoditest. Kirjeldavad statistikud ja graafilised meetodid.
9. Kahte valimit võrdlevad testid, regressioonimudelid (elukestvusandmete korral).
10. Teisi andmeanalüüsi meetodeid (klassikaline ja logistiline regressioon, mitteparameetrilised meetodid). Erinevate meetodite ja testide võimsusest.
11. Alagruppide analüüsimine eraldi, analüüsi kohandamine erinevatele (segavatele) faktoritele.
12. Vahetulemused (kui tihti analüüsida?)
13. Praktilisi aspekte: katse protokollide koostamine, läbiviimiseks vajalik personal, finantsprobleemid, tulemuste esitamine jne.
14. Meta-analüüs.
15. Eetilised probleemid kliiniliste katsete puhul.
16. Näiteid erinevate eksperimentide kohta, sagedased vead ja probleemid.

## KVANTITATIIVSETE TUNNUSTE VARIATSIION

E. Pärna

Eesti Loomakasvatuse ja Veterinaaria Instituut

Põllumajandusloomade produktsioonitunnused on kombineeritud väljundiks looma paljudele erinevatele füsioloogilistele süsteemidele, mis toetavad tunnuse väärtust kas otseselt, aditiivselt või interaktsiooni kaudu teiste füsioloogiliste mehhanismidega. Produktsioonitunnuseid võime nimetada ka "makrofunktsioonideks", mis koosnevad ensümaatiliste, hormonaalsete jt. biokeemiliste reaktsioonide ahelatest, mida reguleeritakse geeniproduktide poolt. Seega on need tunnused polügeensed, sest nende väärtused on mõjutatud väga suure arvu geenide poolt. Selliseid tunnuseid nimetatakse ka kvantitatiivseteks tunnusteks. Kuna nende avaldumist modifitseerivad keskkonna mõjud, ei saa loomade fenotüübi järgi eristada erinevaid genotüüpe.

Kvantitatiivse tunnuse variatsiooni modelleerimiseks on kaks võimalust. Nendest reaktsiooninormi mudel on kasulik kvantitatiivsete tunnuste mõningate põhikontseptsioonide selgitamisel, matemaatiline lineaarne mudel aga annab teoreetilise aluse praktilises aretustöös kasutatavale metodoloogiale.

Reaktsiooninormi mudel on bioloogiliselt tundlik ja illustratiivne, kuid ta ei anna praktilisi töövahendeid kvantitatiivsete tunnuste analüüsimiseks. Sel eesmärgil vajatakse matemaatilist mudelit, mis kirjeldab fenotüübilisi väärtusi genotüübi ja keskkonna funktsioonina. Kuna geneetilised ja keskkonna mõjud pole otseselt mõõdetavad, pole võimalik geneetilist ja keskkonna dispersioonikomponenti tavalise dispersioonivalemi alusel arvutada. Võib küll arvutada fenotüübilise dispersiooni (ja kovariatsiooni), kuid ainult juhul, kui on tegemist mittesuguluses olevate indiviidide juhusliku valimiga vaadeldavast populatsioonist. Enamikel praktilistel juhtudel pole see nõue täidetud, kuna loomad on omavahel suguluses, mis vähendab nendevahelist variatsiooni ning selektsioon ja mitmesugused süstemaatilised faktorid mõjutavad variatsiooni.

Põhimõtteliselt on ainult üks protseduur, millega uurida bioloogilise variatsiooni täiust põllumajandusloomadel farmi tingimustes. Selleks on dispersioonikomponentide tuletamine sugulaste fenotüübilistest väärtustest. Bioloogiliste dispersioonikomponentide hindamise alternatiivseid meetodeid, mis kasutavad sugulaste sarnasust, on suur hulk. Hindamise protseduuril on kolm astet.

I. Andmete analüüs lineaarsete mudelite abil, mille käigus hinnatakse regressioonikoefitsiente või dispersiooni, mida võib kasutada sugulaste fenotüübilise sarnasuse mõõduna.

II. Tuletatakse bioloogiliste dispersioonikomponentide oodatav seos lineaarsest mudelist hinnatud komponentidega.

III. Viimane võrdsustatakse vastavate hinnangutega I astmest, et formuleerida võrrandite võrk, millest leida bioloogiliste komponentide hinnangud.

## METSA KÖRGUSKASVU MUDELITE KOOSTAMISE PROBLEEME

### A. Kiviste

#### Eesti Põllumajandusülikool, metsakorralduse instituut

Metsandusliku uurimistöö üheks olulisemaks ülesandeks on puistu kasvu mudelite loomine. Metsateadlased üle kogu maailma on selle probleemiga tegelejad juba enam kui saja aasta jooksul ning selle tulemusena on loodud tuhandeid puu ja puistu kasvu mudeleid. Praegusel ajahetkel ei saa rääkida üldtunnustatud metsa kasvu mudelist ega isegi üldtunnustatud metoodikast metsa kasvu mudelite loomiseks. Sellise situatsiooni peamiseks põhjusteks võib tuua järgmist:

1. Mets on keeruline bioloogiline kooslus, mille kasv sõltub väga paljudest faktoritest. Metsa kasvu võib vastavalt püstitatud eesmärgile modelleerida erinevatel tasemetel, millele vastavalt rakendatakse erinevat tüüpi matemaatilisi mudeleid.

2. Metsa liigiline koosseis, struktuur ja kasv on erinevates geograafilistes piirkondades väga erinev, mistõttu isegi naaberriikides loodud mudelid vajavad enne rakendamist üsna hoolikat kontrollimist.

3. Mets võib kasvada mitu inimpõlve, mis tundub raskendab andmete kogumist. Vahelduvad nii uurijad kui ka kasutatavad mõõtmismeetodid. Samuti on metsameeste-praktikute traditsioonid eri maades erinevad, mistõttu erinevates maades tehtud mõõtmistulemused, seega ka neist tulenevad kasvumudelid ei ole võrreldavad.

4. Kõik seni loodud metsa kasvumudelid eeldavad kasvutingimuste konstantsust ajas. Tänu inimtegevusele on täheldatud Maa kliimas arvestatavaid muutusi (näiteks süsihappegaasi kontsentratsiooni suurenemine, üldine temperatuurimuutus, ulatuslikud saastesademed jne.), mis mõjuvad otseselt metsa kasvule ja mille tõttu muutuvad varasemate pikaajaliste kordusmõõtmiste analüüsimisel saadud tulevikuprognosisid kahtlaseks.

Ühe metsandusliku statistilise probleemi näitena vaatleme ühevanuseliste puhtmännikute ülakõrguse kasvumodeli koostamist Rootsi

Põllumajandusülikooli 156 kordusmõõtmisega püsiproovitüki andmeil. Uuritud männipuistud olid valitud üle kogu Rootsi kõigist kasvukoha-tüüpidest. Igal proovitükil oli tehtud 2-6 kordusmõõtmist, kokku 609 kirjet. Iga kirje sisaldas proovitüki numbrit, puistu vanuse mõõtmishet- kel ja ülakõrguse (100 kõige suurema puu keskmine kõrgus hektaril) hinnangu.

Puistu kõrguse ( $H$ ) sõltuvust puistu vanusest ( $A$ ) kirjeldatakse tavaliselt üheparameetrilise s-kujuliste kõverate parvena, mille üldkuju on:

$$H = f(A, HB), \quad (1)$$

kus  $HB$  on puistu kõrgus 50 aasta vanuses või mingis teises baasvanu- ses.

Parameetrit  $HB$  nimetatakse puistu kasvukohaindeksiks ehk bonitee- diks. Kõrguse kasvufunktsiooni (1) nimetatakse ka puistu boniteerimis- mudeliks, sest see võimaldab hinnata puistu keskmist kasvukiirust ehk "kasvukoha headust" vastavalt konkreetse puistu vanusele ja kõrgusele.

Võtmeküsimuseks puistu kasvukäigu modelleerimisel on sobiva kasvufunktsiooni valik. Puistu kasvu modelleerimiseks on kasutatud umbes sadakonda erineva kujuga kasvufunktsiooni. Puistu kõrguse kas- vu modelleerimisel eeldatakse:

1. funktsioon peab läbima nullpunkti;
2. funktsioon peab olema kasvav;
3. vanuse suurenemisel peab funktsioon lähenema rõhtasümptoodile;
4. funktsioonil peab olema üks käänupunkt.

Erinevate funktsioonide sobivust puistu kasvufunktsiooniks on uuritud paljude uurijate poolt. Paraku universaalset ja üldtunnustatud kasvufunktsiooni ei ole. Küll on aga välja selgitatud mõned mitte kuigi keerulised funktsioonid, mis rahuldavad ülalloeletud nõudeid kasvu- funktsioonile ja on piisavalt paindlikud mõõtmisandmete lähendamise- seks. Antud andmete lähendamisel kasutati 14 erinevat kasvufunkt- siooni, millest kõige sobivamaks osutus 1822 a. Saksa metsateadlase Hossfeldi poolt soovitatud kolmeparameetriline kasvufunktsioon:

$$H = b_0 / (1 + b_1 / A^{b_2}). \quad (2)$$

Veelgi enam kui sobiva kasvufunktsiooni valik mõjutab mudelit tema parameetrite hindamise meetod. Vastavalt kogutud andmete struktuurile jaotatakse puistu kõrguse kasvumudelite koostamise mee- todid kolme rühma (vt. [2]):

1. giidkõvera meetod;

2. parameetrite prognoosimise meetod;
3. diferentsvõrrandi meetod.

Kõige enam on kasutatud boniteerimismudelite konstrueerimiseks nn. giidkõvera meetodit. Selle meetodi puhul lähendatakse kõik vanuskõrgus andmepaarid valitud kasvufunktsiooniga. Tulemusena saadakse parameetrite hinnangud, millega kirjeldatakse kogu punktisarve läbiv keskmine kõver ehk nn. giidkõver. Kõverate parv konstrueeritakse eeldusest, et kõverad on omavahel proportsionaalsed. Võttes baasvanuseks 50 aastat, saaksime Rootsi männikute kõrguse kasvumudeli kujul:

$$H = 1.74H_{50}/(1 + 469/A^{1.65}). \quad (3)$$

Giidkõvera meetodi peamiseks puuduseks on asjaolu, et saame mudeli, mis kirjeldab puistute kõrguse tingitud keskmise sõltuvust vanusest, kuid see ei pruugi kokku minna puistute tegeliku kasvuga. Meetodi teiseks puuduseks on asjaolu, et kasvukõverate proportsionaalsuse eeldus ei ole tegelikkuses täidetud.

Parameetrite prognoosimise meetodi kasutamiseks on vajalik empiiriline materjal eelnevalt rühmitada ja moodustada teatud arv kõrguse vanuseridu. Iga rida lähendatakse kasvufunktsiooniga, mille tulemusena saadakse iga rea jaoks funktsiooni parameetrite hinnangute komplekt. Meetodi teise sammuna leitakse parameetrite seosed boniteediga HB. Andmete rühmitamiseks ja ridade moodustamiseks kasutatakse mitmeid mooduseid:

1. andmed jagatakse mullastiku ja alustaimestiku kirjelduste järgi metsakasvukohatüüpidesse;
2. vanuseread moodustatakse empiiriliste tinglike kvantiilide ridadena;
3. read koostatakse puutüve analüüside põhjal jne.

Piisavalt pikaajaliste vaatlusseeriade olemasolu korral võiks kõrguse vanuseridadeks olla ka vahetult proovitüki kordusmõõtmiste rida. Et antud püsiproovitükkidel olid tehtud vaid mõned kordusmõõtmised, siis antud materjalil osutus otstarbekaks konstrueerida kõrguse vanuseread Tveite meetodil (vt. [4]). Tveite meetodil saadud ridade lähendamise tulemusena osutus täpseimaks mudel:

$$H = H_{50}((1 + b_1/50^{b_2}((1 + b_1/A^{b_2}))^{b_3}), \quad (4)$$

kus

$$\begin{aligned} b_1 &= 0.1377 \exp(0.4316H_{50}), \\ b_2 &= 0.0988H_{50} - 0.192, \\ b_3 &= 22.5/(H_{50} - 5.5) - 0.55. \end{aligned}$$



Saadud kõrguse kasvumudel on praktilisteks arvutusteks tülikas. Näiteks olgu vanuses  $A1$  puistu kõrgus  $H1$ . Selleks, et prognoosida selle puistu kõrgust mingis vanuses  $A2$ , tuleks esmalt lahendada võrrandisüsteem (4) parameetri  $H50$  suhtes, kus  $H = H1$  ja  $A = A1$ . Seejärel saab võrrandisüsteemi (4) järgi prognoosida vaadeldava puistu kõrgust suvalises vanuses. Samuti on meetodi puuduseks materjali eelneva rühmitamise nõue.

Eelpool mainitud puudustest on vaba Cieszewski, Bella poolt väljatöötatud diferentsmeetodi erijuhtum Hossfeldi funktsiooni (2) jaoks (vt.[1]). Hossfeldi funktsiooni (2) võib esitada kujul:

$$H = H25(1 + b_1/25^{b_2}) / (1 + b_1/A^{b_2}),$$

kus  $H25$  on puistu kõrgus 25 aastases baasvanuses.

Baasvanuseks valiti 25 aastat seetõttu, et sellisel juhul on parameetrid  $H25$  ja  $b_1$  omavahel seotud suhteliselt lihtsa sõltuvusega:

$$b_1 = \beta/H25.$$

Seega omandab kasvufunktsioon kuju:

$$H = (H25 + \beta/25^{b_2}) / (1 + (\beta/H25)/A^{b_2}). \quad (5)$$

Diferentsmeetod on sobiv kasutamiseks kordusmõõtmistega andmestiku jaoks, mille korral saab vaatlusandmete põhjal konstrueerida vaatluspaaride hulga  $\{(A1, H1), (A2, H2)\}$ , kus  $H1$  on puistu kõrgus vanuses  $A1$  ja  $H2$  on puistu kõrgus vanuses  $A2$ . Kasvufunktsioonist (5) võib parameetri  $H25$  avaldada tingimusest, et  $H1 = f(A1)$ . Selle tulemusena puistu kõrgus  $H2$  vanuses  $A2$  avaldub:

$$H2 = (H1 + d + r) / (2 + 4\beta A2^{-b_2} / (H1 - d + r)), \quad (6)$$

kus

$$d = \beta/25^{b_2};$$

$$r = \text{sqrt}((H1 - d)^2 + 4\beta H1/A1^{b_2}).$$

Rootsi männikute kordusmõõtmistega andmestiku põhjal osutus võimalikuks moodustada 453 vanus-kõrgus paari, mille lähendamisel ülal- toodud funktsiooniga saadi parameetrite hinnanguteks  $\beta=6900$  ja  $b_2=1.76$  (vt.[3]).

Diferentsmeetod osutus üsna paindlikuks Rootsi männikute kõrguskasvu mudeli koostamiseks kordusmõõtmistega andmete põhjal. Saadud mudelit on suhteliselt lihtne kasutada puistu ülakõrguse prognoosimiseks, kui on teada puistu kõrgus antud hetkel.

Käesolevas töös kasutati mudeli loomisel Rootsi andmeid. Analoogilist mudelit Eesti metsa jaoks ei ole võimalik veel koostada, sest Eestis on püsiproovitükkide kordusmõõtmisi liiga vähe.

#### Viitekirjandus

1. Cieszewski, C.J., Bella, I.E. Polymorphic height and site index curves for lodgepole pine in Alberta. *Can. J. For. Res.* Vol.19, 1989, 1151-1160.
2. Clutter, J.L., Fortson, J.C., Pienaar, L.V., Brister, G.H., Bailey, R.L. *Timber management: a Quantitative Approach.* New York, 1983.
3. Elfving, B., Kiviste, A. Development site index curves for *Pinus silvestris* on basis of permanent plot data in Sweden. Umeå, 1992. 30 pp. (Manuscript at Institute of Silviculture, Swedish University of Agr. Sci.)
4. Tveite B. A method for construction of site index curves. *Meddelelser fra det Norske Skogforsøksvesen.* Nr.94, 1969, 134-159.

# SEARCH FOR A STANDARDIZATION METHOD TO ENHANCE THE COMMON SIGNAL IN TREE RING SERIES

A. Läänelaid

Institute of Botany and Ecology, Tartu University

## Introduction

A need for working method of dendrochronological dating has forced to search for a suitable standardization technique of the tree ring series for that task.

In the beginning of this century, A.E. Douglass dated the early buildings of pueblo indians using the so-called skeleton-plots of tree rings of the construction timber (Douglass, 1929). Compared with studies in America, dendrochronological dating practice in Europe was long delayed due to technical difficulties (Gräslund, 1984). In 1964 a short report about using a digital computer method for dating short tree ring series marked the beginning of a new stage in this field of tree ring studies (Howland *et al.*, 1964). According to the method referred, the correlation coefficient was computed between the unknown and known tree ring vectors for each possible date. The method presumes that the correlation between the matched ring-width series is the highest in the case of the actual date, thus indicating the right shift in years of the synchronous series. In fact, this assumption may be invalid in some cases, e.g. if the highest correlation is caused by coinciding age curves and not by the actual synchronous position of the two series. This difficulty can be overcome by standardization of the tree ring series (Fritts, 1976). Standardization also helps to amplify synchronous fluctuations of ring widths, thus increasing the probability of finding the right matching position of the series.

There are various models describing tree ring series (e.g. Cherkashin, Kuzmichov, 1977; Maslakov, 1978; Abrosova, Mazepa, 1984; Cook *et al.*, 1990; Dupouey *et al.*, 1992, *etc.*). In this study the formula proposed by Graybill (1982) will be used:

$$R(t) = C + B + D1 + D2 + E$$

where  $R(t)$  is the width of the tree ring in year  $t$ ,  $C$  is the macroclimatic term common to trees at a site,  $B$  is the biological curve as a function of tree age,  $D1$  is the disturbance term possible unique to a single tree due to random events that affected its growth,  $D2$  is the signal common to most or all specimens at the site due to fire, insect damage, or other disturbance.  $E$  is the random growth signal unique to each specimen.

Among these signals, only  $C$  and  $D2$  describe factors common to the trees of a site. These components probably increase the cross-correlation between tree ring series and contribute to the finding out the actual (right) shift of the series. The signals  $B$  and  $E$  do not need coincide in time and may be regarded as a noise in dendrochronological dating. Thus, standardization of ring width series of trees has to serve as a tool to enhance the signal-to-noise ratio (Fritts, 1982).

The problem of the present paper is to find out the most suitable standardization method for ring width series to reveal the common signals  $C$  and  $D2$  in relation to other signals. In tree ring studies, various standardization methods have been practised, e.g. smoothing the sequence by a negative exponential function and a polynomial function (Fritts, 1976), by simple moving average (Rudakov, 1951; Bitvinskas, 1974, *etc*), by maximal possible increment curve (Shiyatov, 1970, 1976; Cook *et al.*, 1990), *etc*. An earlier review of standardization methods in dendrochronology used in the Soviet Union is given by Tarankov (1973); a more systematic introduction to modern methods can be read in Cook *et al.*, (1990). Wide use has been made of growth indices calculated as ratios of ring-widths to the function (or moving average) used in standardization. For dating purposes, semilogarithmic curves are often used (Kolchin, 1963; Kolchin, Chernykh, 1977; Aluve, 1978, 1980; Aniol, 1983). It must be mentioned that logarithmization suppresses but does not eliminate the age curve of a tree ring series.

### Material and Methods

To find a suitable standardization method for dating purposes an experiment was carried out using tree ring series with known sampling year. Borer samples were taken from 17 trees of Scots pines (*Pinus sylvestris* L.), growing in a cowberry pine forest in R apina forest district, south-eastern Estonia. In the sample area, depth of moraine ranges from 0.85 to more than 2 metres. From each sample tree one core was extracted at the height of 1.3 m. The average length of the cores (tree age) was 97 years varying from 90 to 102 years. The ring widths were measured to 0.05 mm exactness. Biological age curve could be followed in all the tree ring graphs.

All the ring series were pairwise correlated, moving the series in relation to each other by one year. Thus 136 pairs of series were studied. Series shifting and calculation of cross-correlations was continued until the overlapping period had shortened to 30 years. For each pair of series, the maximal positive correlation coefficient was revealed and the corresponding shift of the series was recorded. If the matched samples were taken at the same year (both before or both after the growing season), the highest correlation coefficient between their tree ring sequences was expected to occur at a shift of nought years (i.e. the outermost ring widths of the series were not shifted in relation to each other). At that position tree rings of the same calendar years were matched. It was regarded as the right position. If the maximal positive correlation occurred at any other position, the shift was accounted as a wrong one.

To investigate the effect of standardization, nine techniques were compared. In each case the series were moved and correlated in pairs as mentioned above. Percentage of the right position estimations and the average correlation coefficient at the right shift were calculated for each method. Altogether more than 80,000 correlation coefficients were calculated. The following standardization techniques were applied:

$I_{21}$  - growth indices as percentages of ring widths from symmetrical 21-year running mean period;

$\log I_{21}$  - logarithms of the previous indices;

$I_{11}$  - growth indices as percentages of ring widths from symmetrical 11-year running mean period;

$I_{20p}$  - growth indices as percentages of ring widths from 20-year running mean period preceding the index year;

$\log I_{20p}$  - logarithms of the previous indices;

$I_{10p}$  - growth indices as percentages of ring widths from 10-year running mean period preceding the index year;

$I_{4p}$  - growth indices as percentages of ring widths from 4-year running mean period preceding the index year;

$FH$  - digital high-pass filter that removes most of the variation at frequencies lower than  $1/8$  cycles per year (Fritts, 1976);

$FB$  - digital 11-year band-pass filter retaining variation with periods of 6 to 15 years (LaMarche, Fritts, 1972).

First results forced to correct the computer program: the minimal allowed overlapping period for calculating cross-correlation between series was changed from 30 to 40 years and the  $FB$ -standardization was omitted. Four techniques were added:

$\log I_{11}$  - logarithms of the indices  $I_{11}$ ;

$\log I_{10p}$  - logarithms of the indices  $I_{10p}$ ;

$\log I_{4p}$  - logarithms of the indices  $I_{4p}$ ;

$H$  - Hollstein's method: logarithms of ratios of ring widths to the width of the preceding tree ring (Fritts, 1976, p. 281).

### Results

Table 1 contains summary of the results when using the first version of the program where the series were shifted and matched up to the minimal overlapping period of 30 years. The standardization methods are ordered by the percentage of the correctly estimated shifts.

**Table 1.** Percentage of right shifts and the corresponding average correlation coefficient between series by different standardization methods. The tree ring series overlap at least 30 years.

Standardization Method	Percentage of correct datings	Average correlation at the right shift
$\log I_{20p}$	58.8	0.48
$I_{10p}$	55.1	0.48
$I_{11}$	53.7	0.42
$I_{4p}$	50.7	0.46
$FH$	50.0	0.46
$I_{20p}$	48.5	0.51
$I_{21}$	45.6	0.48
$\log I_{21}$	45.6	0.47
$FB$	14.7	0.82

After modifying the computer program (extension of the minimal overlapping period from 30 years to 40 years) and including some alternative standardization methods, the probability of right shift changed.

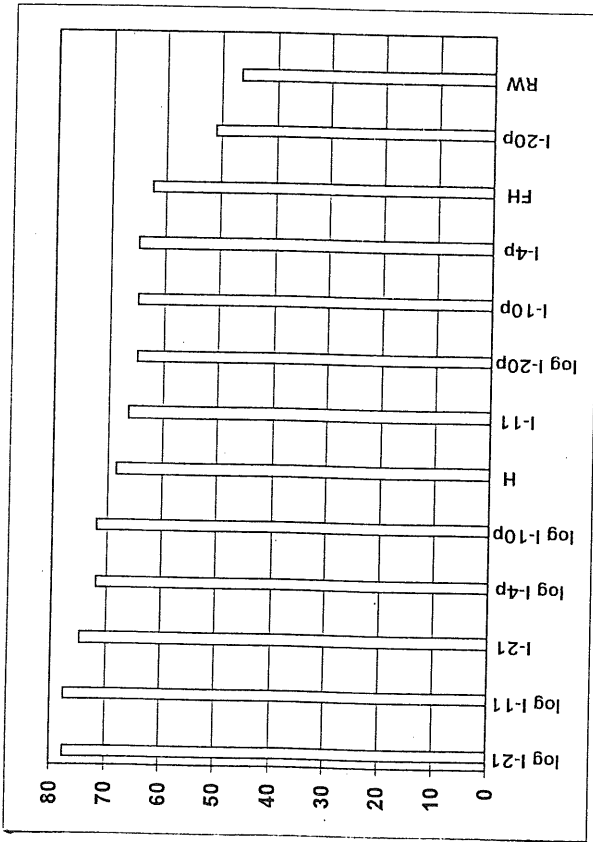
Table 2 represents pairwise correlations of the ring width series transformed by  $\log I_{21}$  method.

**Table 2.** Maximal correlations between  $\log I_{27}$ -standardized ring-width series. Above the main diagonal - shift in years yielding the highest correlation; below the diagonal - the ratios 'Correlation coefficient/critical value of the coefficient at  $\alpha < 0.001$ '. Wrong shifts and the corresponding correlation ratios are given in brackets.

Core	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	-	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	(0)
2	2.67	-	1	1	1	(3)	(3)	1	1	1	1	(2)	1	1	1	2	(3)
3	2.49	2.69	-	0	0	(5)	0	0	0	0	0	(5)	(2)	(5)	(2)	(6)	(6)
4	2.27	2.38	2.49	-	0	(14)	0	0	0	0	0	0	0	0	0	(0)	(10)
5	2.30	2.19	2.25	2.07	-	0	(2)	0	0	0	(-2)	0	0	0	0	(-1)	1
6	2.05	(1.86)	(1.95)	(1.81)	2.07	-	0	0	0	0	0	0	0	0	0	1	1
7	2.56	(2.60)	2.51	2.32	(2.18)	1.95	-	(32)	0	0	(-2)	0	(-1)	0	0	1	1
8	1.53	1.6	1.89	1.77	2.13	1.51	(1.48)	-	0	0	(-32)	(6)	0	0	0	(-46)	(6)
9	2.39	2.36	2.19	2.10	1.99	1.90	2.26	1.78	-	0	0	0	0	0	0	1	1
10	2.54	2.55	2.56	2.38	2.28	1.91	2.57	1.72	2.39	-	0	0	(-1)	0	0	1	(0)
11	2.48	2.41	2.43	2.11	(2.09)	1.89	(2.49)	(1.61)	2.31	2.55	-	0	0	(6)	0	1	1
12	2.61	(2.59)	(2.54)	2.22	2.18	2.23	2.59	(1.54)	2.47	2.54	2.43	-	0	0	0	1	1
13	2.42	2.47	(2.37)	2.15	2.29	2.24	(2.25)	1.72	2.36	(2.38)	2.45	2.43	-	(6)	0	1	1
14	1.83	1.94	(2.06)	1.92	1.95	2.11	1.83	2.21	1.72	2.00	(1.65)	2.38	(1.82)	-	0	1	(0)
15	2.56	2.55	(2.35)	2.32	2.29	2.21	2.27	1.71	2.28	2.39	2.38	2.51	2.49	2.05	-	1	(0)
16	2.45	2.52	(2.22)	(2.23)	(2.11)	2.08	2.50	(1.46)	2.28	2.50	2.44	2.66	2.28	1.98	2.41	-	0
17	(2.44)	(2.57)	(2.46)	(2.17)	2.06	2.32	2.49	(1.70)	2.32	(2.50)	2.47	2.83	2.53	(2.30)	(2.53)	2.61	-

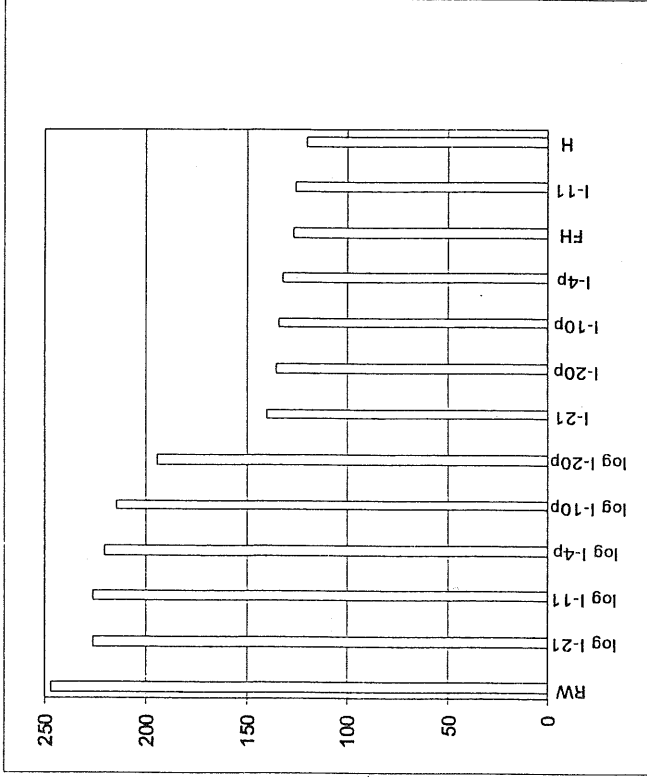
Fig. 1 presents the ordered standardization methods with respect to the percentage of right positions. Fig. 2 graphs the average relative correlation coefficients in the case of right position using various standardizations.

For dating purposes it is important to consider both the percentage of right datings (right positions), as well as the value of the corresponding correlation coefficient (Fig. 3).

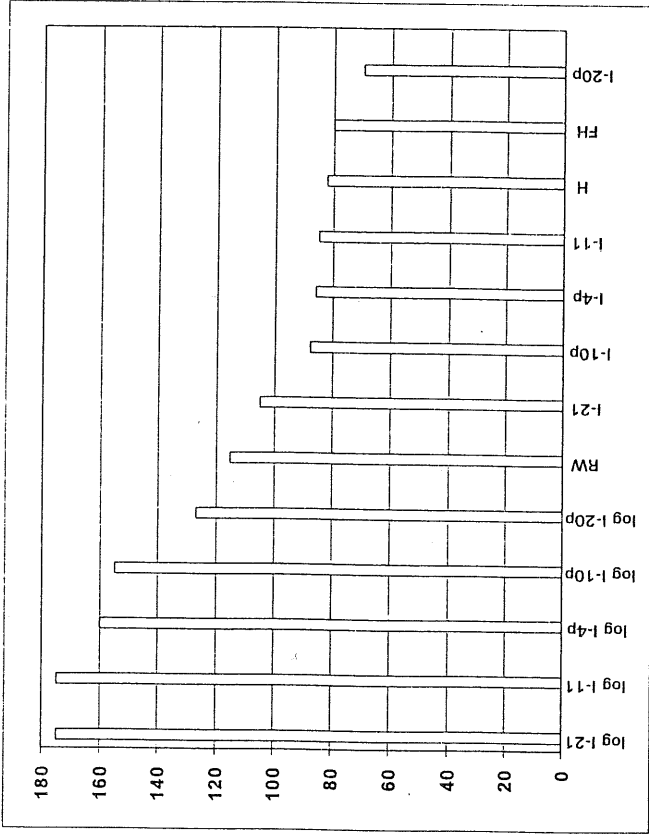


**Figure 1.** Ranking of standardization methods according to the percentage of right shifts as derived from cross-correlating 17 standardized tree ring series up to minimally 40-year overlap. RW - untreated ring-widths.

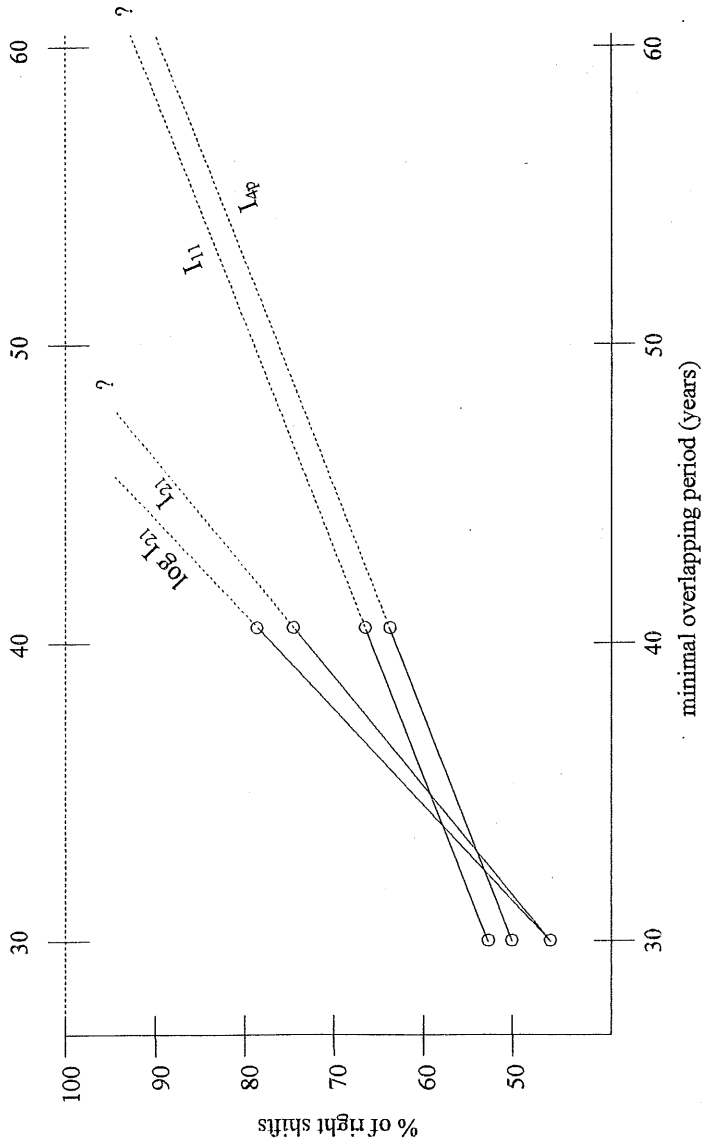




**Figure 2.** Ranking of standardization methods according to the average relative cross-correlation between correctly shifted series. On Y-axis the average ratio 'correlation coefficient / critical correlation coefficient corresponding to  $\alpha=0.001$ ' is given. RW - untreated ring-widths.



**Figure 3.** Ranking of standardization methods according to the product of the percentage of right positions (shifts) and the average relative correlation coefficients at the right positions.



**Figure 4.** Relation between the probability of right shifts and the minimal overlapping period of the matched pair of tree ring sequences

## Discussion

Using the first version of the program, the method  $\log I_{20p}$  has produced right positions in 59% of the cases, which appeared to be the highest value among the tested methods. The lowest percentage of right positions using the 11-year band-pass filter is obviously due to asynchronous phytocoenotic cycles of the same periodicity. Often the position of the maximum correlation deviated from the right one only 1 or 2 years. The fact of interest is, that correlation of series of raw (unstandardized) ring widths (RW) comprising age curve, has yielded right positions only in 31.6% of all cases. In spite of that, the average correlation coefficient at right positions was high: 0.84.

Unexpected by low percentage of right positions forced to search for means to raise it by 1) increasing the minimal overlapping period of the correlated series and 2) including additional standardization techniques. At the same time, criterion of recorded shift was altered: instead of the maximal significant positive coefficient of correlation, the relative values of coefficients of correlation were regarded as basis in search for right positions.

Alteration of the computer program resulted in remarkable increase in the percentage of right positions (cf. Table 1 and Fig. 1). Ranking of the standardization techniques also changed. Ordering of methods in respect of the relative values of the correlation coefficients has moved the popular method  $I_{21}$  backward (Fig. 1). It is remarkable, that the relative value of the correlation coefficients of logarithms of growth indices exceeds the corresponding unlogarithmed indices. This fact evidently refers to the greater concentration of the common synchronous growth fluctuations in narrower tree rings. The high relative value of the correlation coefficients (2.47) in the case of raw ring widths can be easily explained by the coinciding biological curves (signal B in the Graybill's formula) of the even-aged sample trees. If matching trees of different age, the unstandardized ring widths should have no advantage over the standardized ring widths for dating purposes.

To study the effect of the amount of right positions and the corresponding relative values of correlation simultaneously, these values were multiplied. It came out that the arrangement of the standardization methods in regard to the products of the two parameters did not alter the first half of the sequence (Fig. 3):  $\log I_{21}$  and  $\log I_{11}$  retained their leading positions.

For dating relatively short chronologies (about one hundred years) it is essential to know the minimal allowed overlapping period of the correlated sequences. On the other hand, probability of getting wrong positions

increases with decrease of the overlapping segments of the chronologies. Hence the minimal overlapping period for correlating two sequences should be increased to a desirable probability of right datings (positions). Using the most appropriate techniques, e.g.  $\log I_{21}$ , the expected optimal length of the minimal overlapping period would be derived graphically (Fig. 4). It appeared that the recommended minimal overlapping period for tree ring sequences standardized by  $\log I_{21}$  would be about 50 years. Comparing that number with some dendrochronological studies we can see that a minimal overlapping period of 40 years has been used in Italy (Pignatelli, Preto, 1984). But 33 years is evidently too short overlapping for getting significant results (Pilcher *et al.*, 1984). Nevertheless, there are examples by F. Renner of possible wrong datings even in the case of much longer than 40 years overlappings (Schweingruber, 1983, p. 87). In general, overlapping of about one hundred years is considered to be sufficient to obtain right datings (Baillie *et al.*, 1985; Fritts, 1985; McGrail, 1978). Although B. Huber has written that search for the correct position "... is not - as it may appear to many - subjective by intuition and by "God's glance"; it is more a case of an acquired technique with objective result" (Huber, 1978) as a rule, seldom more than 2/3 of samples were datable in European and American laboratories (Schweingruber, 1983).

### Conclusions

1. Among the 12 standardization methods under the study the highest percentage, 77, of right datings was obtained using logarithms of growth indices calculated from symmetrical 21-year running mean period ( $\log I_{21}$ ) as well as using logarithms of indices calculated from symmetrical 11-year running mean period ( $\log I_{11}$ ). Calculations were carried out, shifting the sequences in relation of each other to minimal overlapping of 40 years. These techniques,  $\log I_{21}$  and  $\log I_{11}$ , may appear to have perspective in further dendrochronological dating.

2. The same detrending techniques,  $\log I_{21}$  and  $\log I_{11}$ , appeared to give the highest relative value of correlation - 2.26 times over the 0.999 significance level in average. These methods, revealing best the common growth fluctuations among trees, may come of use for dendroclimatological studies.

3. For the case sample plot of recent Scots pines from R pina, South-eastern Estonia, the desired minimal overlapping of sequences of 50 years would theoretically yield approximately 100% of right datings, if the best standardization techniques are used.

## Acknowledgements

I would like to thank late Dr. Erich Lõhmus, senior researcher of the Estonian Forest Research Institute, for kind donation of his tree ring measurements of pines of Rāpina for this study. I am also indebted to Tõnu Sepp, former computer manager at Tartu University, for appropriate computer programs.

## References

1. Abrosov, N. S., Mazepa, V. S. 1984. Sovremennye podhody pri izuchenii dendrohronologicheskoy informacii i analiz vremennykh ryadov. Preprint 38-B. Inst. fiziki im. L. V. Kirenskogo SO AN SSSR. Krasnoyarsk, 51 p. (in Russian)
2. Aluve, K. 1978. Eesti NSV lääneosa ehitusmälestiste dendrokronoloogilisest dateerimisest. - Ehitus ja Arhitektuur, 2, lk. 18-23.
3. Aluve, K. 1980. O dendrohronologicheskoy datirovke pamyatnikov arhitektury Zapadnoy Estonii. - Izv. AN Estonskoy SSR, 29. Obshestvennye nauki, 4, p. 342-354. (in Russian)
4. Aniol, R. W. 1983. Procedures and statistics in dendrochronology. - In: Eckstein, D., Wrobel, S., Aniol, R. W. (eds.), Dendrochronology and archaeology in Europe. Proceedings of a Workshop of the European Science Foundation (ESF), held in Hamburg April 28-30, 1982. Mitteilungen der Bundesforschungsanstalt für Forst- und Holzwirtschaft, Nr. 141, Oktober, 1983. Hamburg, p. 21-31.
5. Baillie, M. G. L., Hillam, J., Briffa, K. R., Brown, D. M. 1985. Re-dating the English art-historical tree ring chronologies. - Nature, Vol. 315, No. 6017, p. 317-319.
6. Bitvinskas, T. T. 1974. Dendroklimaticheskiye issledovaniya. Gidrometeoizdat, Leningrad, 172 p. (in Russian)
7. Cherkashin, V. P., Kuzmichov, V. V. 1977. Statisticheskii analiz ryadov shiriny godichnykh kolec derevyev. Krasnoyarsk, 50 p. (in Russian)
8. Cook, E., Briffa, K., Shiyatov, S., Mazepa, V. 1990. Tree Ring Standardization and Growth-Trend Estimation. - In: Cook, E. R., Kairiukstis, L. A. (eds.), Methods of Dendrochronology. Applications in the Environmental Sciences. Kluwer Acad. Publ., Dordrecht, p. 104-123.
9. Douglass, A. E. 1929. The secret of the Southwest solved by talkative tree rings. - Nat. Geogr. Mag., Vol. 56, No. 6, p. 736-770.

10. Dupouey, J. L., Denis, J. B., Becker, M. 1992. A new method of standardization for examining long-term trends in tree ring chronologies. - In: Eggertsson, O. (ed.), *Tree Rings and Environment. Proceedings of the International Dendrochronological Symposium*, Ystad, South Sweden, 3-9 September 1990. LUNDQUA Report 34. Lund, p. 85-88.
11. Fritts, H. C. 1976. *Tree Rings and Climate*. Acad. Press, London, 567 p.
12. Fritts, H. C. 1982. An overview of dendroclimatic techniques, procedures and prospects. - In: Hughes, M. K. *et al.* (eds.), *Climate from tree rings*. Cambridge Univ. Press, Cambridge, p. 191-197.
13. Fritts, H. C. 1985. Tree-ring analysis (dendroclimatology). - In: Oliver, J. E. (ed.), *The Encyclopedia of Climatology*. Fairbridge, Rh. W. (ed.), *The Encyclopedia of Earth Sciences*. Hutchinson-Ross, Stroudsburg, PA, 21 p.
14. Gräslund, M. 1984. The history of Dendrochronology in the Nordic Countries. - *Dendrochronologia*, 2, p. 31-62.
15. Graybill, D. A. 1982. Chronology development and analysis. - In: Hughes, M. K. *et al.* (eds.), *Climate from tree rings*. Cambridge Univ. Press, Cambridge, p. 21-31.
16. Howland, H. C., Sharrock, G., Raskin, J. 1964. A Digital Computer Method for Dating Short Tree Ring Series. - *Nature*, Vol. 201, No. 4917, p. 412-413.
17. Huber, B. 1978. Dendrochronology. - In: Fletcher, J. (ed.), *Dendrochronology in Europe*. Greenwich, p. 15-26.
18. Kolchin, B. A. 1963. Dendrokronologiya Novgoroda. - *Materialy k issledovaniyu po arkheologii SSSR*, No. 117, t. 3. *Novye metody v arkheologii*, p. 5-103. (in Russian)
19. Kolchin, B. A., Chernykh, N. B. 1977. Dendrokronologiya Vostochnoy Evropy (Absolyutnye dendrokronologicheskiye shkaly s 788 po 1970 g.). *Nauka*, Moskva, 128 p. (in Russian)
20. LaMarche, V. C. Jr., Fritts, H. C. 1972. Tree rings and sunspot numbers. - *Tree Ring Bull.*, Vol. 32, p. 19-33.
21. Maslakov, Ye. L. 1978. O vliyaniy kolebaniy klimata na prirost derevyev v molodnyakakh sosny. - In: Bokhovkin, I. M. (ed.), *Tezisy dokladov k III Vsesoyuznoy konferencii po dendroklimatologii "Dendroklimaticheskiye issledovaniya v SSSR" (4- & yulya 1978 g.)*. Arkhangel'sk, p. 57-58. (in Russian)
22. McGrail, S. 1978. Dating ancient wooden boats. - In: Fletcher, J. (ed.), *Dendrochronology in Europe*. Greenwich, p. 239-258.

23. Pignatelli, O., Preto, P. 1984. Analisi dendrocronologica su alcune strutture lignee dell'eta del Bronzo rinvenute a Canale Anfora (Udine). - *Dendrochronologia*, 2, p. 99-101.
24. Pilcher, J. R., Baillie, M. G. L., Schmidt, B., Becker, B. 1984. A 7,272-year tree ring chronology for western Europe. - *Nature*, Vol. 312, No. 5990, p. 150-152.
25. Renner, F. 1982. Beiträge zur Gletschergeschichte des Gotthardgebietes und dendroklimatologische Analysen an Fossilen Hölzern. - *Physische Geographie. Univ. Zürich*, Vol. 8, 182 S.
26. Schweingruber, F. H. 1983. Der Jahrring. Standort, Methodik, Zeit und Klima in der Dendrochronologie. Verl. Paul Haupt, Bern, 234 S.
27. Shiyatov, S. G. 1970. K metodike rascheta indeksov prirosta derevyev. - *Ekologiya*, No. 3, p. 85-87. (in Russian)
28. Shiyatov, S. G. 1986. Dendrohronologiya verkhney granicy lesa na Urale. Moskva, 137 p. (in Russian)
29. Tarankov, V. I. 1973. Vvedeniye v dendroklimatologiyu Dalnego Vostoka. - *Tr. Biol.-pochv. inst. Dalnevostochnogo nauchnogo centra AN SSSR*, t. 12, No. 115, p. 7-23. (in Russian)



22. märtsil 1995. a. toimus Eesti Statistikaltsi üldkoosolek, kus arutati parandusi ja täiendusi seltsi põhikirja muutmiseks seoses muutustega seltside tööd reguleerivas seadusandluses. Täiendatud põhikiri registreeriti 19. aprillil 1995. a. Tartu Maavanema korralduse nr. 229 alusel 13. aprillist 1995. a.

Asutatud Tartus,  
30. sept. 1992. a.

## **Eesti Statistikaltsi PÕHIKIRI**

### **1. Seltsi eesmärgid ja ülesanded**

1.1. Eesti Statistikalts on vabatahtlik isikuid ühendav mittetulunduslik ühing, mis soodustab statistika arengut, levikut ja rakendamist Eestis.

1.2. Seltsi ülesanneteks on

- soodustada statistikaalast uurimistööd;
- integreerida eri valdkondades töötavate statistikute erialast tegevust;
- aidata kaasa statistilise informatsiooni kogumise ja töötlemise meetoodika täiustamisele ja efektiivsuse tõstmisele;
- tõsta oma liikmete erialast kvalifikatsiooni ja aidata kaasa statistikute kaadri ettevalmistamisele;
- aidata kaasa statistika-alaste teadmiste levitamisele Eestis;
- hoolitseda statistika õpetamise taseme tõstmise eest Eesti haridussüsteemis;
- stimuleerida statistika-alase informatsiooni ulatuslikku kasutamist Eesti majandus- ja ühiskonnaelus;
- teha rahvusvahelist koostööd teiste statistikaseltsidega;
- arendada sidemeid lähedastel teadusaladel tegutsevate Eesti seltsidega.

1.3. Nende ülesannete täitmiseks selts

- organiseerib teoreetilisi ja praktilisi nõupidamisi, seminare ja konverentse;
- korraldab kvalifikatsiooni tõstmise kursusi ja seminare;
- korraldab teaduslikke ja meetoodilisi ettekandekoosolekuid ja juubeli- ning memoriaalistungeid;

- määrab ja taotleb oma liikmetele stipendiume ja komanderinguid;
- korraldab konkursse statistikateaduste alal, seab sisse preemiaid silmapaistvate tööde eest;
- osaleb statistika-alase teadusliku, õppemethodilise ja aimekirjanduse väljaandmisel ning vastavate plaanide aruteludes;
- propageerib statistikat ja statistika rakendamist ajakirjanduse, televisiooni, raadio, loengulise töö jms. kaudu;
- teeb Eesti Vabariigi valitsusele ettepanekuid statistikaalase tegevuse parandamiseks ja korraldamiseks;
- loob vajaduse korral erialaseksioone ning komisjone ja töörühmi seltsi ees seisvate ülesannete lahendamiseks.

## **2. Seltsi koosseis, liikmete õigused ja kohustused.**

### **2.1. Seltsi koosneb tegev- ja auliikmetest:**

- tegevliikmeks võib olla isik, kes on tuntud seltsile oma tegevusega statistika uuringute, nende rakenduste või statistika-alase hariduse valdkonnas;
- auliikmeks võib valida isiku, kellel on erilisi teeneid statistika arendamisel või seltsi tegevuses.

2.2. Seltsi asutajaliikmed on seltsi asutamiskoosolekust osavõtjad, kes peale registreerimiskaardi täitmist ja sisseastumismaksu tasumist saavad seltsi tegevliikmeteks.

2.3. Tegevliikmeks astumisel tuleb esitada juhatusele avaldus koos kahe soovitusena tegevliikmetelt. Juhatuse otsustab liikmeks vastuvõtu lahtisel hääletamisel.

2.4. Tegevliikmed saavad liikmepileti. Auliikmetele antakse auliikme diplom.

2.5. Seltsi sisseastumismaksu ja liikmemaksu suurus määratakse seltsi üldkoosolekul, auliige on seltsi maksudest vabastatud.

2.6. Seltsi liikmetel on otsustav hääleõigus seltsi üldkoosolekutel.

2.7. Seltsi liikmetel on õigus osaleda seltsi üritustel.

2.8. Seltsi tegevliige on kohustatud oma tegevusega kaasa aitama seltsi eesmärkide ja ülesannete elluviimisele ja täitma põhikirja nõudeid.

2.9. Kustutamine seltsi liikmete nimekirjast toimub juhatuse koosolekul. Kustutamise põhjuseks võib olla liikme surm, omal soovil lahkumine või põhikirja sätete rikkumine, sh. liikmemaksu tasumata jätmine enam kui kahe aasta jooksul. Juhatuse otsuse kinnitab üldkoosolek lahtisel hääletamisel.

### 3. Seltsi organid

3.1. Seltsi organiteks on üldkoosolek, juhatus eesotsas presidendiga ja revisjonikomisjon.

3.2. Üldkoosolek on seltsi kõrgeim organ ja toimub vähemalt üks kord aastas. Üldkoosoleku päevakorra ja toimumisaja teatab seltsi juhatus igale liikmele vähemalt kaks nädalat ette. Juhatus on kohustatud kokku kutsuma erakorralise üldkoosoleku vähemalt ühe viiendiku seltsi liikmete allkirjastatud nõudmise alusel.

3.3. Üldkoosolek

- kuulab ära, arutab ja kinnitab juhatuse ja revisjonikomisjoni aruanded;
- valib seltsi presidendi, juhatuse ja revisjonikomisjoni kolmeks aastaks kinnise hääletamise teel kooskõlas "Mittetulundusühingute ja nende liitude seadusega" 23. märtsist 1994. a.;
- võtab vastu otsuseid ja soovitusi seltsi tegevuse kohta.

Otsused neis küsimustes võetakse vastu kohalolevate liikmete liithääaltenamusega.

3.4. Otsuseid seltsi põhikirjas muudatuste või täienduste tegemise, uue põhikirja vastuvõtmise ning seltsi reorganiseerimise või likvideerimise kohta teeb üldkoosolek kvalifitseeritud hääaltenamuse (2/3 kohalolijaist) alusel.

3.5. Seltsi juhatus koosneb (lisaks seltsi presidendile) 7-12 isikust ja valib oma koosseisust asepresidendi(d), sekretäri ja laekuri.

Juhatus asukoht: Tartu, J. Liivi tn. 2.

3.6. Juhatus

- juhhib seltsi tegevust üldkoosolekute vahelisel ajal;
- planeerib ja organiseerib seltsi ülesannete täitmist;
- käsutab seltsi varasid;
- kutsub kokku seltsi üldkoosolekuid;
- viib ellu üldkoosolekute otsuseid ja soovitusi.

Juhatus peab oma koosolekuid vähemalt kaks korda aastas, tehes otsuseid kohalolijate hääaltenamuse alusel lahtise hääletamise teel.

3.7. Revisjonikomisjon koosneb kolmest isikust, valib oma koosseisust esimehe ning viib vähemalt kord aastas läbi seltsi tegevuse revisjoni. Teeb otsuseid lahtise hääletamise teel liithääaltenamuse alusel.

3.8. Seltsi juhatus ja revisjonikomisjon annavad oma tegevusest aru üldkoosolekule.

#### **4. Seltsi juriidilised õigused ja vahendid.**

4.1. Selts on kasumit mittetaotlev juriidiline isik, kellel on oma pitsat ja pangaarve.

4.2. Seltsi rahalised vahendid moodustuvad

- liikmete sisseastumis- ja aastamaksudest;
- ühekordestest sisseasketest (sponsorlus, annetused, konverentside ja kokkutulekute osavõtumaksud, jms.);
- tulust, mida saadakse seltsi väljaannete kirjastamisest ja levitamisest või kultuurhariduslike ning muude seltsi põhikirjalistest eesmärkidest tulenevate tasuliste ürituste korraldamisest;
- muust tulust, mis saadakse seltsi põhikirjalise tegevuse arendamisel.

4.3. Seltsi omandis võib olla igasugune vara, mis on vajalik seltsi põhikirjaliste eesmärkide saavutamiseks ja mille omandamine ei ole vastuolus seadusega.

4.4. Seltsi tegevusest laekunud tulu kasutatakse põhikirjajärgsete eesmärkide täitmiseks.

4.5. Seltsi rahaliste vahendite kohta peetakse raamatupidamise arvestust ja makstakse makse riigi- ja kohalikku eelarvesse seadusandluses ettenähtud korras.

#### **5. Tegevuse likvideerimise kord.**

5.1. Selts lõpetab oma tegevuse

- seltsi üldkoosoleku otsuse alusel;
- põhikirja kinnitanud organite otsuse alusel kooskõlas "Mittetulundusühingute ja nende liitude seadusega" 23. märtsist 1994. a.

5.2. Seltsi likvideerimine toimub vastavalt 23. 03. 94 välja antud "Mittetulundusühingute ja nende liitude seaduses" kehtestatud korrale.

## KONVERENTSID

Käesoleva aasta augustis toimuvad kaks kaalukat sündmust rahvusvahelises statistikaelus, millest nende ürituste organisaatorid on palunud informeerida Eesti statistikuid. Neist laiaulatuslikum,

### **4-th World Congress of the Bernoulli Society**

toimub Viinis 26.-31. aug. 1996. a.

Meile mitmes mõttes lähedasem

### **16-th Nordic Conference on Mathematical Statistics**

leiab aset nädal varem Lahtis 18.-21. aug. 1996. a. Täpsema informatsiooni saamiseks palun pöörduda TÜ matemaatilise statistika instituudi poole, Tartu J. Liivi 2 (e-mail [kollo@cs.ut.ee](mailto:kollo@cs.ut.ee)).

Kõigile meeldetuletuseks: meie oma rahvusvaheline konverents Tartus arvutusstatistika ja statistika õpetamise alal toimub 2.-8. juunini k.a. Kõik on teretulnud.

Seltsi sekretär T. Kollo

Tartu Ülikooli Kirjastuse trükikoda  
Tiigi 78, EE2400 Tartu  
Tellimus nr. 188.