

Eesti Statistikeselts 20

# Uued suundumused statistikas

Eesti Statistikeseltsi 24. konverents

27.-28. september 2012

Programm ja ettekanded

Tartu 2012

Koostanud Imbi Traat, Kalev Pärna, Kelli Sander  
Autorite tekst muutmata

Tartu Ülikooli Kirjastus  
[www.tyk.ut.ee](http://www.tyk.ut.ee)  
Tellimus nr. ...

## Saateks

Kätte on jõudnud Eesti StatistikaSeltsi juubelisünnipäev, millele on pühendatud seltsi käesolev 24. konverents. Eesti StatistikaSelts asutati 20 aastat tagasi 30. septembril 1992. Täna on seltsil ligi 100 liiget ja 3 auliiget. Liikmeid on üle Eesti ja välismaaltki, uusi liikmeid lisandub igal aastal. Algusaastaist peale on selts olnud aktiivne, mille tunnistuseks on igaaastased konverentsid ja neid kajastavad teabevihikud. Seltsi konverentsid on alati laiemat huvi pakkunud, läheb ju statistikat vaja paljudes elu valdkondades. Veelgi erilisema tähelepanu alla tõuseb statistika aastal 2013, mil üle maailma tähistatakse rahvusvahelist statistika-aastat.

Seekordne konverents keskendub kahele peateemale – rahvaloendusele ja suurte andmete probleemile. Aasta 2011 oli rahvaloenduse aastaks paljudes Euroopa riikides, sealhulgas Eestis. Suurürituse läbiviimisest, kogutud andmetest ja nende analüüsist on palju rääkida. Suurte ja ülisuurte andmete olemasolu ja vajadus nende analüüsiks on suhteliselt uus probleem. Tekivad raskused nii statistikameetodite rakendamisel kui ka algoritmide realiseerimisel. Käesolev konverents heidab veidi valgust sellele valdkonnale, millest ka nimetus "Uued suundumused statistikas".

Juubelikonverentsile omaselt on meil külalisi ka väljastpoolt Eestit, mistõttu on konverents kakskeelne. Õnnitlema on tulnud statistikaSeltside esindajad Poolast, Rootsist, Soomest, Lätist ja teised koostööpartnerid. Pidulikult anname diplomid kätte kahele uuele auliikmele, prof. Hannu Niemile Helsingi Ülikoolist ja prof. Leo Vöhandule Tallinna Tehnikaülikoolist.

Täna kõiki konverentsi organiseerimisele kaasaaitajaid. Eriti suur tänu ettekannete tegijaile. Soovin kuulajatele meeldejäädavaid konverentsielamusi.

Imbi Traat  
Eesti StatistikaSeltsi president

## Preface

The 24th Conference of the Estonian Statistical Society is devoted to its anniversary. The Estonian Statistical Society was founded 20 years ago on September 30, 1992. Today, the Society has nearly 100 members and 3 honorary members. There are members from Estonia as well outside of it, their number increases every year. The Society has been active from the very beginning. Important outcomes of its activity have been annual conferences accompanied by the published Bulletins. The conferences of the Society have always obeyed wide interest, evidently caused by the need of statistics in various fields. Statistics receives even more attention in 2013 when the International Year of Statistics is celebrated worldwide.

The present conference focuses on two main topics – the population census and the Big Data. In 2011 census was carried out in many European countries, including Estonia. There is much to tell about organization, data collection, results and future perspectives of this tremendous event. Nowadays, new problems arise from availability of big, or even super big data sets. Their analysis requires different, often new statistical methods and new, more powerful algorithms. Our conference enlightens this novel topic.

Due to the jubilee, we have guests also from outside Estonia. Therefore the conference languages are Estonian and English. Representatives from statistical societies of Finland, Latvia, Poland and Sweden have come with congratulations. During our festive event we announce two new Honorary Members – Professor Hannu Niemi from the University of Helsinki and Professor Leo Võhandu from the Tallinn Technical University.

Hereby, I thank all people who helped to organize this conference. Special thanks go to the participants with presentations. I wish memorable conference to all participants.

Imbi Traat

President of the Estonian Statistical Society

# Programm

## 27. september

- 10–11 Registreerimine TÜ peahoones aula ees  
Kohv senati saalis aula kõrval
- 11.00 **Avaistung TÜ aulas:** juh. Kalev Pärna  
Eesti Statistikaselts 20 – ESS president **Imbi Traat**  
Tervitused – Riigikogu esimees **Ene Ergma**, külalised  
Auliikmete diplomite üleandmine (Leo Võhandu, Hannu Niemi)
- 11.30 **L. Võhandu** (TTÜ): A long path of creating order out of chaos
- 12.00 **E. Gołata** (Poznań): Poland – demographic perspectives in the view  
of Population Census 2011
- 12.20 **E.-M. Tiit** (TÜ, ES): Census and population size of Estonia
- 12.40–14.00 **Lõuna**

### TÜ raamatukogu saal

#### Rahvastikustatistika: juh. Ene-Margit Tiit

- 14.00 **D. Matteus** (ES): Is register-based census possible in Estonia?
- 14.20 **V. Halapuu, T. Paas** (TÜ): Determinants of people's attitudes to-  
wards immigrants in Europe
- 14.40 **E. Ulicāns** (Rēzekne): Gross domestic product and labour force de-  
velopment tendencies in the Baltic States, 2000–2011
- 15.00 **M. Servinski** (ES): Rahvastikuareng ja Eesti meedia
- 15.20 **Kohvipaus**

#### Suured andmemahud: juh. Tõnu Kollo

- 15.50 **D. von Rosen** (Uppsala): From multivariate to high-dimensional data
- 16.10 **E. Saar** (Tartu Obs.): Statistics for large cosmological surveys
- 16.30 **J. Vilo** (TÜ): Data mining and bioinformatics
- 16.50 **A. Leontjeva** (TÜ): Building predictive models from user behavior  
data in very large social networks
- 17.10 **R. Sirel** (TÜ): Text-mining the Estonian Electronic Health Record

19.00–21.00 **Bankett TÜ Ajaloomuuseumi valges saalis (Toomel)**

## 28. september

### TÜ raamatukogu saal

#### Statistiliste andmete analüüs: juh. Meelis Käärrik

- 9.00 **O. Bogdanov** (SAS Inst.): Mis on BIG DATA ja kuidas seda töödelda
- 9.20 **K.-R. Kont**(TTÜ), **M. Jõgi** (RR): Raamatukogustatistika kogumisest ja analüüsist Eestis
- 9.40 **J. Pelt** (Tartu Obs.): Ernst Öpik ja matemaatiline statistika
- 10.00 **K. Pärna, R. Kangro** (TÜ): Tehisõpe *versus* tavastatistika – näide kindlustuse vallast
  
- 10.20 **Kohvipaus**

#### Statistika ja geneetika: juh. Märt Möls

- 11.00 **M. Remm** (TÜ): Suuremahulised andmed DNA sekveneerimisest
- 11.20 **K. Fischer** (TÜ): Riskidest ja prognoosimisest TÜ Eesti Geenivaramu kohordis
- 11.40 **R. Mägi** (TÜ): Ülegenoomsed assotsiatsioonianalüüsid ja nende meta-analüüs
- 12.00 **T. Kaart** (EMÜ): Korrelatsioonimaatriksi illustreerimisest – näiteid eluteadustest
  
- 12.20–13.30 **Lõuna**

#### Statistika ja eluteadused: juh. Imbi Traat

- 13.30 **M. Haldna, J. Haberman** (EMÜ): Pikaajaliste hüdrobioloogiliste andmete analüüs Võrtsjärve zooplanktoni näitel
- 13.50 **K. Kuljus, B. Ranney** (Umeå): Optimaalne disain Rootsi põdra- ja metskitsepopulatsiooni suuruse hindamiseks ekskrementide loendamise meetodi abil
- 14.10 **A. Sims, A. Kiviste, A. Nilson** (EMÜ): Eesti metsaregistri andmestik ja selle analüüsi probleemidest
- 14.30 **Ü. Kristjuhan** (TTÜ): Eluea ja tervena elatud aastate pikendamise võimalused Eestis lähematel aastakümnetel
  
- 14.50 **Kohvipaus**

**Statistika ja majandus:** juh. Tanel Kaart

15.20 **M. Täht** (ES): Eesti sisemajanduse koguprodukti prognoosimisest

15.40 **D. Teneng** (TÜ): The normal inverse Gaussian distribution: exposition and applications

16.00 **Ümarlaud: Statistika tulevikusuundumused**,  
vestlust juhib Krista Fischer

17.00–18 **ESS üldkoosolek**





# Eesti statistikaseltsi 24. konverents

## *Uued suundumused statistikas*

**Ene-Margit Tiit**

*Statistikaamet, Tartu Ülikool*

Eesti statistikaselts on jõudnud tudengiikka ja tähistab käesoleval sügisel oma ümmargust sünnipäeva. Seltsi loomise eesmärk 20 aasta eest oli ühendada Eestis erinevates valdkondades ja erinevates paikades töötavaid statistikuid, anda neile võimalus omavaheliseks suhtlemiseks ja koosmõtlemiseks. Selle eesmärgi saavutamiseks on igal aastal korraldatud vähemalt üks konverents, mille teemavalik on käinud läbi väga erinevad statistika rakendus- alad – statistika õpetamisest ning statistilisest kirjaoskusest rahvastiku- statistikani, registritest elukvaliteedi ja kvaliteedistatistikani. Ka paiknemise mõttes on konverentsid käinud läbi suure osa Eestist – konverentsikülalisi on võõrustanud lisaks Tartule ja Tallinnale ka Pärnu, Paide, Viljandi, Otepää ja Rakvere. Kõigist konverentsidest on jäänud jälg raamatukese, Statistikaseltsi teabevihiku näol, mis tänaseks raamaturiulil juba kokku enam kui kahe tuhande leheküljega võrra ruumi võtavad ja avardavad statistikahuviliste ning statistikatarbivate maailmapilti eestikeelsete uurimusartiklitega. Kuna valdavalt on konverentsid olnud eestikeelsed, on neil olnud oluline roll eesti statistikakeele ja siitkaudu ka statistilise mõtlemise kujundamisel.

Statistikaseltsi loomise ajal püstitatud eesmärgid on tänaseni aktuaalsed, võimalik, et veelgi aktuaalsemad kui toona. Internetiajastul muutub inimlik silmast silma suhtlemine aina väärtuslikumaks, erinevate valdkondade teadmiste sünergia omandab hindamatu väärtuse.

Siiski erineb noorukiikka jõudnud statistikaseltsi sünnipäevakonverents varasematest. Esmakordselt toimub konverentsi avaistung Tartu Ülikooli kõige auväärsemas paigas – aulas. Kuna külla on kutsutud ka naaberseltside esindajaid, on konverents seekord kakskeelne – rida ettekandeid esitatakse inglise keeles niihästi külaliste poolt kui ka lugupidamisest külaliste vastu.

Sünnipäevakonverentsil ei räägita peamiselt ühest ainevaldkonnast, nagu seda on tehtud enamasti varem. Otsides teemat, mis ühendaks erinevaid statistikavaldkondi ja oleks tänapäeval eriti oluline, sügav ja paeluv, jõudis seltsi juhatus äratundmisele – see on suured andmestikud! Nimelt suured ja ülisuured andmestikud, mille analüüsimine ja töötleminegi oli mõnekümne aasta eest võimatu ja kümnekond aastat tagasi keerukas, iseloomustavad tänast teadust ja eriti statistikat. Ülisuuri andmehulki kogutakse nii taevast – astronoomiliste ja meteoroloogiliste vaatluste käigus, merest kui ka maa peal. Miljonitesse ulatuvaid andmehulki analüüsivad rahvastikustatistikud ja geneetikud, kusjuures statistika meetodid võivad olla ühised sõltumata

sellest, mis erialal ja mille kohta tehakse järeldusi. Erinevates statistikavaldkondades tekivad uued probleemid, mis on paljuski ühised. Saab selgeks, et klassikalise statistika vorm jääb tihti uute, tänapäeva jaoks tähtsate ülesannete jaoks kitsaks, sageli on tarvis tavapäraseid eeldused asendada uutega, leida uusi lähenemisi. Kõik need on teemad ja suundumused, mida saabuval sünnipäevakonverentsil käsitletakse.

Eesti Statistikaselts on viimaste aastate jooksul nimetanud auliikmeid, need on kas Eesti teenekad statistikud või välismaalased, kellel on silmapaistvaid teeneid statistika arendamisel Eestis. Ka sellel konverentsil saab selts kaks uut auliiget – üks neist on Helsingi ülikooli statistikaproffessor, kes on väga tõhusalt kaasa aidanud kahe naaberülikooli statistikute – nii õppejõudude kui ka tudengite – omavahelise suhtluse arendamisele. Teine on Eesti statistika *grand old man* Leo Võhandu, kelle enam kui pooleaja aasta eest Tartu Ülikoolis peetud faktoranalüüsi loengud olid äratuskellaks, mis pani paljude erialade teadlaste pähe idanema mõtte – statistikameetodid on vajalikud, nendest võib ka minu töös olla kasu. Sünnipäevakonverentsi avaettekande teebki Tallinna Tehnikaülikooli emeriitprofessor Leo Võhandu.

Kuna käesoleval aastal lõppes rahvaloendus, on rahvastiku teema käesoleval statistikakonverentsil aukohal. Sellele on pühendatud esimene istung, kus esineb ka külalisprofessor Poolast. TÜ raamatukogu saalis jätkuval istungil tuleb kõneks ka rahvaloenduste tulevik – kas ja millal asendub klassikaline loendus, mille üheks komponendiks on ükselt uksele käimine, inimeste andmete kokkupanemisega olemasolevatest riiklikest registritest. Euroopas on viimase kümnendi jooksul migratsiooni teema omandanud erilise kõla, see jagab ka muidu rahumeelseid inimesi arvamusrühmadesse ja mõjutab poliitikatki, seda arvestades on eriti huvitav uuring, mis käsitleb eurooplaste hoiakuid rändurite suhtes.

Esimese konverentsipäeva õhtupoolne istung enne banketti Ajaloomuuseumi valges saalis on see, kus suurte ja ülisuurte andmestikega seonduvaid statistika uusi suundumusi teoreetiliselt käsitletakse niihästi matemaatilise statistika kui ka infotehnoloogia vaatevinklist.

Teine konverentsipäev algab statistiliste andmete analüüsile pühendatud istungiga, mille käigus kosmoloogia andmestikest jõutakse märksa maisema ja argisema raamatukogustatistikani.

Teise konverentsipäeva järgmised istungid on pühendatud geneetikale ja eluteadustele, kus tähelepanu all on nii Võrtsjärvest kogutud kui ka Eesti metsaregistris talletatud andmete analüüs, Rootsimaa põtrade arvukuse analüüsi meetodika ja Eesti inimeste eluea pikendamise võimalused. Päeva viimane istung käsitleb mõningaid majandusstatistika olulisi teemasid ja punkt pannakse konverentsile statistika tulevikusuundumusi käsitleval ümarlaulal.

# A long path of creating order out of chaos

Leo Võhandu

*Tallinna Tehnikaülikool*

Seven different high schools in Viluste, Reola, Tartu, Elva and finally Tartu again. Silver medal with algebra a 4 instead of 5. Naturally I had to study mathematics at Tartu University. Additional activities: chess teacher, shorthand writer for 8 years at the University (dissertation defences dialogues spreading from child birth to space) and chess player.

Somehow I finished university with honors and defended my dissertation about iterative methods in Banach spaces in 1955 (there are citations of those results even nowadays).

I started teaching with 33 hours per week. Numerical mathematics, variational calculus, analysis, graphical geometry. I started also with free-lance lectures about statistics for biology, medicine and veterinary people. Headed the Biophysics laboratory for 7 years. Awful lot of consultations.

I did visit in 1964–65 Chicago University Committee for Mathematical Biology (Rashevsky, Rosen, Landahl). My main result was an idea about so-called Maximal Correlation Path (1959, 1964) as a variational minimal representation of different multivariable systems. In Russian internet there are listed 168 dissertations, which cite this method as main method of data representation and analysis. In nowadays Estonian science systems that counts as a round ZERO. One Moscow big advertising firm sells the method as "method of Estonian mathematician L:K.Võhandu".

In 1966 I did emigrate to Tallinn Technical University, Why? Free flat, Finnish TV. Most importantly I had some time to think how scientific hypothesis are created. As a result one of my doctorates Jossif Mullat did prove a theorem about monotone systems, which has served us very effectively up to now. I did build a method of conformity measures for empirical systems. For statistical proofs we are using Bootstrapping. Our method has been very effective also in solving different NP-hard problems (Cliques, dicliques, TSP, FCA ...). Just now we are studying effective diagonalisation methods for sparse matrices (useful in computer linguistics). Altogether I have had 44 defended candidate and doctoral dissertations under my supervision and consultation. That is the summary of one pretty old Datadigger's life work!

# Poland – demographic perspectives in the view of Population Census 2011

Elżbieta Gołata

*Poznan University of Economics, Poland*

**Keywords:** registered based census, nonsampling errors, residents, people registered for permanent residence and actually living, census coverage survey.

The aim of the paper is to discuss demographic perspectives of Poland in the view of National Census 2011 data. The 2011 Population Census in Poland was the first one using data from administrative sources and gathering information directly from the people in the sample survey.

Despite long tradition as well as a well-developed research methodology, censuses do not provide 'perfect' results. First of all, the census as a comprehensive study may be burdened with the errors of another than random character. Taking into account all the different methods for conducting censuses, including the traditional method, use of data from administrative records with sample surveys and the mixed approach, different sources of error are analyzed.

The study attempts to identify census population in the light of international standards: residents, people registered for permanent residence and actually living (de facto population), as well as the resulting consequences. Considering sources of errors, methods of their identification and elimination are discussed.

Evaluation of population census is 'traditionally' carried out based on the results of post-census survey. As Polish experiences in this field are rather small, particular attention is given to methods based on demographic analysis. Coverage errors are illustrated by examples from previous Polish censuses referring to research conducted by J. Paradysz as well as own analysis of the discrepancies observed in the last census. Discussion includes also harmonization problems, particularly the danger of divergent results and estimates, in terms of several data sources.

# Census and population size of Estonia

## Rahvaloendus ja rahvaarv

Ene-Margit Tiit

*Statistikaamet, Tartu Ülikool*

Rahvaloendused – maailma vanim statistikatöö, tegemist väga suurte andmestikega, eriti arvutieelsel ajastul. Eesmärk – hinnata rahvastiku arvukust.

See on alati olnud hinnang, milles sisaldub põhimõtteliselt 2 viga – alaja ülekaetus. Keskmiselt need tasakaalustuvad vastastikku (kui loendusesse on tõsiselt suhtunud). Pikka aega on lähtunud eeldusest: rahvaarv = loendustulemus.

Kui ülekaetust õnnestub kõrvaldada isikute täpsema identifitseerimise teel, siis osutub loendatud isikute arv tegeliku rahvaarvu jaoks nihkega hinnanguks. Alakaetust süvendavad kaasaajal muudki põhjused (inimeste liikuvus, negatiivsed hoiakud loenduse suhtes, isikud, kes ei kuulu ühessegi riiki).

Kui on tarvis tegelikku rahvaarvu teada, tuleks alakaetuse viga parandada. Võimalused:

- Alakaetuse mahu hindamine ja loendusandmete kaalumise;
- Isikute imputeerimine mingi eeskirja järgi;
- Registrite kasutamine loenduskogumi täiendamiseks.

Eesti REL2011 andmestiku ja riiklike registrite põhjal on töötatud välja alakaetuse parandamise meetodika (1) eksperthinnangute baasil; (2) diskriminantanalüüsi baasil; (3) logistilise analüüsi baasil.

Sellega seostub rida täiendavaid probleeme: kas peaks hindama ka ülekaetust? Mis on primaarne, kas register või ütlus? Kuidas hinnata hinnanguvigu? Kas tulemused on selletõttu väga usaldusväärsed, et aluseks on väga suured andmestikud?

# Is register-based census possible in Estonia? Kas registriandmetel põhinev rahvaloendus on võimalik?

Doris Matteus

*Statistikaamet*

2011. aasta rahva ja eluruumide loenduse andmete avaldamiseks ettevalmistamine alles käib, kuid juba on algust tehtud ettevalmistustega järgmiseks, 2020. aasta loendusvooruks. Eesmärgiks on viia järgmine loendus läbi registripõhisena, kasutades traditsioonilise küsitluse asemel andmeid, mida riik ühel või teisel moel ühte või teise andmekogusse juba kogub.

Eesti ei ole selles osas pioneer. Võrreldes eelmise, 2000. aasta loendusvooruga suurenes nende riikide arv, kes kasutasid lisaks traditsioonilisele küsitlusele loenduse läbiviimisel kas suuremal või vähemal määral andmeid, mida riik ühte või teise andmekogusse juba kogunud on, olulisel määral. Euroopa riikidest viisid 2010. aasta vóorus täielikult registritel põhineva loenduse läbi juba 8 riiki. Lisaks kasutasid mitmed riigid, nende hulgas Eesti, kombineeritud meetodit – traditsioonilist loendust toetasid mingil moel registriandmed.

Registripõhise loenduse ettevalmistamisel on esimeseks suureks tööks analüüsida erinevatesse riiklikesse andmekogudesse kogutava teabe sobivust rahva ja eluruumide loenduse otstarbeks ning hinnata selle teabe kvaliteeti. Võrreldes riikidega, kes juba kasutavad loendusel vaid registriandmeid või registriandmeid kombineerituna muude meetoditega, ei ole Eesti riiklike andmekogude olukord halb, kuid olulisi probleeme, millele tuleb 2020. aastaks lahendus leida, on mitmeid. Probleemidest tõsiseim on asjaolu, et inimeste aadressid rahvastikuregistris ei vasta tihti nende tegelikult elukohale.

# Inimeste hoiakuid immigrantide suhtes kujundavad tegurid Euroopas

Vivika Halapuu, Tiiu Paas

*Tartu Ülikool, majandusteaduskond*

Riikide ja regioonide tulevane majanduskasv ja konkurentsivõime on teiste tegurite kõrval oluliselt seotud ka sellega, kui võrd tolerantsed ollakse erisuste suhtes ning kui võrd suudetakse integreerida erinevast rahvusest inimesi. Siit tulenevalt on uurimistöö sisuks selgitada hoiakuid immigrantide suhtes kujundavaid tegureid Euroopa riikides võttes siinjuures aluseks hoiakute individuaalseid ja kollektiivseid tegureid käsitleva interdistsiplinaarse teoreetilise raamistiku. Töö empiiriline osa tugineb Euroopa Sotsiaaluuringu neljanda vooru küsitluste andmetele. Uurimustulemustest nähtub, et immigrantide suhtes on reeglina tolerantsemad vähemusrahvuste hulka kuuluvad inimesed, kõrgema haridusega ning parema sissetulekuga inimesed ja ka need, kellel on olemas välismaal elamise kogemus. Vastajate hoiakutega immigrantide suhtes on tugevas positiivses seoses ka usaldus riigi institutsioonide ning majapidamiste kindlus oma sotsiaalmajandusliku olukorra stabiilsuse suhtes. Hoiakud immigrantide suhtes erinevad ka tulenevalt mitmetest riigispetsiifilistest teguritest (kultuuriline ja ajalooline taust, immigrantide osakaal, riigi üldine majanduslik olukord jt), mis läbiviidud ökonomeetrilises analüüsis on arvesse võetud riikide fiktiivsete muutujate kaudu. Riikide immigratsioon- ja integratsioonipoliitika meetmete kujundamisel tuleb lisaks riigi rahvastiku koosseisu ning majandusliku ja ajaloolise arengu omapärade arvestamisele rakendada ka meetmeid, mis oluliselt tõstavad inimeste usaldust riigi institutsioonide ning majapidamiste sotsiaalmajandusliku olukorra stabiilsuse suhtes.

## Determinants of people's attitudes towards immigrants in Europe

The key elements of global competition and economic success are no longer trade of goods and services and flows of capital, but the competition for people. Economic success also depends on how tolerant is a society to newcomers and how supportive are public policies and institutions in integrating immigrants. Therefore our paper focuses on examining the attitudes of European people to immigration using micro-data of the European Social Survey fourth round database and relying on interdisciplinary theoretical framework of individual and collective determinants of people's

attitudes towards immigrants. The study is intending to provide empirical evidence-based grounds for the development of policy measures to integrate ethnically diverse societies, taking into account the composition of the country's population as well as other country's peculiarities. The outcomes of the empirical analysis show that European peoples' attitudes towards immigrants vary depending on 1) personal characteristics of the respondents; 2) country's characteristics; 3) peoples' attitudes towards countries' institutions and socio-economic security. Surprisingly, respondents' labour market status (employed, unemployed) does not have a statistically significant relationship with their attitudes towards immigrants. Thus, in order to support the integration of ethnically diverse societies, the implementation of policy measures that support the improvement of people's attitudes towards a country's institutions and socio-economic situation are inevitably necessary.



# Gross domestic product and labour force development tendencies in the Baltic States, 2000–2011

Einārs Ulnicāns

*Rezekne Higher Education Institution, Latvia*

**Keywords:** Gross domestic product, employed persons, unemployed persons, labour productivity, loss of unemployment, Baltic States.

The presentation with calculations analyses development tendencies of gross domestic product, employment, unemployment, labour productivity and loss of unemployment in the Baltic States during 2000–2011. The results of the calculations are explained in the description of these trends and their obvious and possible causes. A brief concept of the theoretical background and the main formula for the calculation of labour productivity is provided as well. Conclusions are drawn about the overall character of development trends. Overall trends of economic development are similar in all three countries. Gross domestic product, employment and labour productivity is growing until 2007. From 2008 to 2010 they are reducing as a result of economic crisis, but in 2011 all the indicators are rising again. GDP growth pace in Latvia, Lithuania and Estonia from 2000 to 2007 was higher than the average in the European Union. This made it possible to reduce the backlog comparing with more developed European Union countries. Major, including the negative changes in the differences more frequently are observed in Lithuania and Estonia. Differences between countries appear in nuances, especially in Lithuania. For example, Lithuania for the longest time has managed to maintain GDP growth since it has decreased against the previous year only in 2009. On the other hand, the fastest growth of Estonian GDP is taking place in 2010 – final stage of the crisis. Labour productivity in Lithuania in 2008, with emergence of the first indications of crisis, still continued to increase, but the number of employed persons had already started to decrease. In Latvia and Estonia there was an opposite development trend. The most significant statistical difference between the Baltic States is that labour productivity is always the highest in Estonia, Lithuania is taking the second place, and Latvia is consistently having the lowest indicator.

# Rahvastikuareng ja Eesti meedia

Mihkel Servinski

*Statistikaamet*

- "Tõenäoliselt langeb 2050. aastaks Eesti elanike arv 800 000–900 000 peale, ja see on üsna vana elanikkond."(Joakim Helenius – Postimees, 26. mai 2012 (Arter))
- "Mina olen jõudnud sügavale veendumusele, et käesolev ei ole mitte kinnisvara-, rahandus- ega majanduskriis. Tegemist on sügava väärtuste kriisiga."(Andres Arrak – Postimees, 8. veebruar 2012)
- "Kas põhiseaduse kirjapanijad mõistsid tõesti, et säilitamine on meie ambitsioonide lagi? Et ratsionaalselt võttes oleme siiski hääbuv kultuur, millega saja aasta pärast saavad etnograafiahuvilise tutvuda Raadil asuvas Euroopa suurimas rahvamuuseumis."(Armin Kõomägi – Äripäev, 23. veebruar 2012)
- "Meie suurim probleem ongi püsiva töö puudumine. Seda ollakse sunnitud otsima väljaspool Eestit. See on traagika perekonnale ja lastele."(Tiit Vähi – Maaleht, 21. juuni 2012)
- "Rahvaarvu trendi arvestades on Eesti viimane koht, kuhu investeerida."(Raivo Sormunen – Äripäev, 1. november 2011)
- "Me vajame immigratsiooni, et maksta oma pensione. Kust võiks immigrandid tulla? Võib-olla osa tuleb Ukrainast või Venemaalt, aga tegelikult on noorte inimeste reserv muslimi maailmas."(Timothy Garton Ash – Postimees, 10. juuli 2012)
- "Peame endale teadvustama, et väike sissetulek on praegu probleem number üks ja see tuleb lahendada."(Igor Rõtov – Äripäev, 20. juuni 2012)
- "... väikeste perede probleem ei ole raha vähesus ja kui probleem pole rahas, ei saa seda ka rahaga lahendada."(Arved Breidaks – Lääne Elu, 29. juuni 2012)
- "Kokkuvõttes: eesmärk ei saa olla rahvaarv, vaid inimeste õnn ja heaolu."(Kaarel Tarand – Eesti Päevaleht, 21. juuli 2010)
- "Seitse või üheks miljardit nimest ei ole jube. Jube oleks see, kui me ei suudaks kohandada oma elu ja suhteid siin maakeral ümber uue arenguperioodi vajaduste järgi."(Erik Terk – Postimees, 22. oktoober 2011)
- "Olen veendunud, et Eesti rahval on jõudu elada üle ka praegune ebakindel aeg."(Andrus Ansip – Postimees, 25. veebruar 2012)

# From multivariate to high-dimensional data

Dietrich von Rosen

*Department of Engineering and Technology,  
Swedish University of Agricultural Sciences,  
Department of Mathematics, Linköping University*

High-dimensional statistical analysis is, with today's huge amount of available data, of the utmost interest. Various different high-dimensional approaches are natural extensions of classical multivariate methods. A general characterization of high-dimensional analysis is that there are more dependent variables than independent observations available. It is driven by theoretical challenges as well as numerous applications such as applications within signal processing, finance, bioinformatics, environmetrics, chemometrics, etc. The area comprises, but is not limited to, random matrices, Gaussian and Wishart matrices with sizes which turn to infinity, free probability, the R-transform, free convolution, analysis of large data sets, various types of  $p, n$ -asymptotics including the Kolmogorov asymptotic approach, functional data analysis, smoothing methods (splines); regularization methods (Ridge regression, partial least squares (PLS), principal components regression (PCR), variable selection, blocking); and estimation and testing with more variables than observations.

If one considers the asymptotics with  $p$  indicating the number of dependent variables and  $n$  the number of independent observations, there are a number of different cases:  $p/n \rightarrow c$ , where  $c$  is a known constant, and both  $p$  and  $n$  go to infinity without any relationship between  $p$  and  $n$ . The latter case, however, has to be treated very carefully in order to obtain interpretable results. For example, one has to distinguish if first  $p$  and then  $n$  goes to infinity or vice versa, or  $\min(p, n) \rightarrow \infty$ . When studying proofs of different situations in the literature, it is not obvious which situation is considered and many results can only be viewed as approximations and not as strict asymptotic results, at least on the basis of the presented proofs.

One of the main problems in multivariate statistical analysis as well as high-dimensional analysis occurs when the inverse dispersion matrix,  $\Sigma^{-1}$  has to be estimated. If  $\Sigma$  is known, it often follows from univariate analysis that the statistic of interest is a function of  $\Sigma^{-1}$ . Then one tries to replace  $\Sigma^{-1}$  with an estimator. If  $\mathbf{S}$  is an estimator of  $\Sigma$ , the problem is that  $\mathbf{S}^{-1}$  may not exist or may perform poorly due to multicollinearity, for example. If  $\mathbf{S}$  is singular, then  $\mathbf{S}^+$  has been used. Moreover, "ridge type" estimators of the form  $(\mathbf{S} + \lambda \mathbf{I})^{-1}$  are in use (Tikhonov regularization). Sometimes a shrinking takes place through a reduction of the eigenspace by removing the

part which corresponds to small eigenvalues. A different idea is to use the Cayley-Hamilton theorem and utilize the fact that

$$\Sigma^{-1} = \sum_{i=1}^p c_i \Sigma^{i-1},$$

where  $\Sigma$  is of the size  $p \times p$  and since  $\Sigma$  is unknown the constants  $c_i$  are also unknown. Then an approximation of  $\Sigma^{-1}$  is given by

$$\Sigma^{-1} \approx \sum_{i=1}^a c_i \Sigma^{i-1}, \quad a \leq p,$$

and an estimator is found via  $\widehat{\Sigma}^{-1} \approx \sum_{i=1}^a \widehat{c}_i \mathbf{S}^{i-1}$ . When determining  $c_i$  a Krylov space method, partial least squares (PLS), is used.

# Statistics for large cosmological surveys

**Enn Saar**

*Tartu Observatory*

The amount of data obtained by surveys in cosmology is growing year by year. I list shortly the present and planned large surveys, their goals, typical features and data access methods (large databases, virtual observatories).

I describe the specifics of the statistical methods used presently to extract information from these data. In conclusion, I give several examples of real-life applications, both from my own practice and those developed by other cosmologists.

# Building predictive models from user behavior data in very large social networks

Anna Leontjeva

*The Software Technology and Applications Competence Centre/  
Tarkvara Tehnoloogia Arenduskeskus*

The Software Technology and Applications Competence Center (STACC) is a research-oriented organization that works in the field of software technology and aims to bridge the gap between scientific and economic innovations. One of the STACC projects deals with the social network analysis, which has been a hot topic in recent years due to the enormous data available for analysis from social networks.

The aim of my presentation is to introduce the social network analysis that we perform in STACC and give some insights what are the main topics of research in our team.

Firstly, the general information about the data and the complications of this field of research will be discussed. Then, the comparison of two giant social networks as Skype and Facebook will be shown. The aim of this comparison is to understand if it is possible to apply the information from one social network to another.

I will also show some results on bursty egocentric network evolution, e.g. how users add new contacts in an online social network and if there are any groups of people with similar patterns.

Next, the problem of service adoption and fraud detection in the social network will be discussed. I will demonstrate how the knowledge about the neighbours in a graph can be used to predict the future behaviour of a user.

# Text-mining the Estonian Electronic Health Record

Raul Sirel

*The Software Technology and Applications Competence Centre/  
Tarkvara Tehnoloogia Arenduskeskus*

The Electronic Health Record (EHR) is a nationwide system that integrates data from Estonia's different healthcare providers to create a common record for each patient. It is a powerful tool for practitioners, for all the concerning doctors can access and complement the electronic patient record with little effort. Furthermore, EHR provides data for statisticians and researchers interested in national health, pharmacological and epidemiological research, etc.

However, in many cases the valuable information is locked away as narrative texts, which need to be linguistically processed in order to access the semantics carried by the texts. These texts may include anamneses, discharge letters, radiological reports, etc. To make things worse (or more interesting), the texts in electronic health records are often grammatically and vocabularily extremely diverse and thus difficult to process.

The problem is actual worldwide and usually adressed via Natural Language Processing (NLP), which is used to transform facts encoded in natural language into machine-understandable format.

It is the aim of the presentation to give a generic overview of challenges in processing and mining textual clinical data in EHR.

# Tehisõpe *versus* tavastatistika – näide kindlustuse vallast

Kalev Pärna ja Raul Kangro

*Tartu Ülikool, matemaatilise statistika instituut*

Tehisõpe (statistical learning, machine learning) nagu matemaatiline statistikagi tegeleb klassifitseerimis- ja prognoosimeetodite väljatöötamisega. Tehisõppe peamised arendajad on arvutiteadlased, mistõttu selle valdkonna terminoloogia on mõnevõrra erinev traditsioonilises statistikas kasutatavast (nt valim = treeningandmestik, hindamine = treenimine). Tehisõppe meetodite puhul tehakse tunduvalt vähem eeldusi andmete kohta ning nad on sageli suhteliselt arvutusmahukad.

Me näitame, kuidas tehisõppe meetodeid kasutada ühe kindlustusest pärit probleemi lahendamisel: milline on kindlustuspoliisi õige hind? Poliisi puhaspreemia on summa, mis on võrdne analoogsete poliiside keskmise kahjusummaga, seega puhaspreemia katab täpselt keskmised kahjud. Tegemist on tüüpilise regressiooniülesandega, kus tuleb leida keskmine kahju  $E(Y|x)$  poliisi  $x$  korral. Poliisivektori  $x$  komponentideks on siin nii kindlustusvõtja enda kui ka kindlustatava objekti parameetrid (sugu, vanus, auto mark, hind jmt). Antud regressiooniülesannet võib üritada lahendada traditsioonilise statistika meetoditega, kasutades näiteks üldistatud lineaarseid mudeleid (GLM). Alternatiivseks lähenemiseks on tehisõppe meetodid, millede hulgas tundusid olevat sobivamad kandidaadid lähinaabrite meetod ja regressioonipuud. Lähinaabrite meetodi puhul leitakse meid huvitavale poliisile  $x$  teatud mõttes kõige lähedased poliisid ehk lähinaabrid, misjärel poliisi  $x$  puhaspreemia leitakse kui tema lähinaabrite keskmine kahju. Viimane on olemasoleva kahjuandmestiku põhjal otseselt arvutatav. Konkreetsemal empiirilisel andmestikul läbi viidud testimise tulemusena selgus, et lähinaabrite meetod ületab prognoosi täpsuse osas nii GLM mudeleid kui ka regressioonipuud. Samas tuleb märkida, et lähinaabrite meetod ei anna meile valmis prognoosivalemit vaid üksnes algoritmi, mille töö tulemuseks on prognoos. Lähinaabrite meetod on suhteliselt arvutusmahukas, kuid kaasaja arvutustehnika võimaldab selle potentsiaali edukalt ära kasutada.



# Mis on BIG DATA ja kuidas seda töödelda

Oleg Bogdanov

*SAS Institute*

Iga aasta toob kaasa andmekandjate hindade langemise ja samas ka kogutava ja salvestatava andmemassi kasvu. Viimase kümne aastaga on gigabaidi andmete säilitamise hind langenud kümneid kordi. Samas võimaldab infotehnoloogia ja uued kommunikatsioonivõimalused koguda tohutul hulgal tähtsat ja ka vähemtähtsat informatsiooni. Tekkinud on uus paradigma – BIG DATA. Ehk andmed kogunevad sellise kiirusega ja sellisel hulgal, et ei jõua neid enam operatiivselt töödelda ja sõklaid teradest eraldada.

Kõik see muudab analüütika ja anlüütiku rolli järjest tähtsamaks. Vaja on uusi lähenemisi ja teistsuguseid tehnoloogiaid, et eraldada vähemtähtsatest andmetest tähtsad andmed, teha seda kiiresti, et siis analüüsi tulemusena võtta vastu elutähtsaid otsuseid.

Räägime antud ettekandes veidi lähemalt kolmest suunast suurte andmehulkade töötlemisel.

1. Hajus-andmetöötlus, ehk kuidas jaotada andmete töötlus laiali paralleelselt paljude arvutite vahel.
2. Analüütika viimine andmebaasidesse sisse ehk sinna, kus andmed "asuvad".
3. Analüütika teostamine otse mälus, et kiirendada analüüsi protsessi ja hõlvata mudelite ehitamisse rohkem andmeid saamaks täpsemaid tulemusi.

Kasutades ja kombineerides paindlikult neid kolme tehnoloogiat on võimalik tavaline analüütika viia täiesti uuele tasemele.

# Ernst Öpik ja matemaatiline statistika

Jaan Pelt

*Tartu Observatoorium*

Ernst Julius Öpik (1893–1985) on ilma kahtluseta maailmas kõige tuntum eesti päritoluga astronoom. Tema huvide ring oli nii lai, et kõige selle kajastamiseks oleks vaja korraldada eraldi konverents. Tuleval aastal möödub 120 aastat kuulsa mehe sünnist ja on seega ka paslik aeg sellele mõelda. Ettekandes käsitlen põhjusi, miks ka statistikud võiksid Ernst Öpiku teaduslikust pärandist huvitatud olla. Ma ei räägi siin ainult tema matemaatilise statistika loengutest, mida ta luges Harvardis, vaid ka väga paljudest konkreetsetest uurimustest. Nii klassikalisi kui ka enda poolt välja töötatud matemaatilise statistika meetodeid rakendas ta erinevatele objektidele – päikesesüsteemi väikekehadest alates kuni tähesüsteemideni välja. Pole see- ga ime, et tema nimi on antud nii asteroidile kui ka Marsi kaaslaste Phobose ühele kraatrile.

# Raamatukogustatistika kogumisest ja analüüsist Eestis

Kate-Riin Kont<sup>1</sup> ja Margit Jõgi<sup>2</sup>

<sup>1</sup>TTÜ Raamatukogu, <sup>2</sup>Rahvusraamatukogu

Selleks, et raamatukogud saaksid toetada teadus- ja õppetegevust, elanikkonna haritust, tagada riigi ja ühiskonna arenguks vajaliku informatsiooni kättesaadavuse, peavad raamatukogu juhtkond ja rahastajad olema piisavalt informeeritud vajalike otsuste vastuvõtmiseks ja nende elluviimiseks.

Eesti Rahvusraamatukogu on raamatukogunduse vabariiklik arenduskeskus Eestis, mis koordineerib raamatukogunduse standardimist, raamatukogustatistika kogumist ning riiklike vaatluste korraldamist. Lähtudes riikliku statistika seadusest kogutakse vastavalt rahvusvahelisele raamatukogustatistika standardile (EVS-ISO 2789:2007. Informatsioon ja dokumentatsioon. Rahvusvaheline raamatukogustatistika) aastast-aastasse andmeid raamatukogude tegevuse ja ressursside kohta. Samas toimub ka andmete kontrollimine, töötlemine ja analüüsimine ning andmete kogumise aluste väljatöötamine ja kaasajastamine kooskõlas kõikide riiklike õigusaktidega ja juhitud ISO standarditest. Raamatukogud kasutavad statistikat peamiselt strateegiliseks planeerimiseks, tõendusmaterjalina raha taotlemiseks ja otsuste tegemiseks.

Raamatukogustatistika mõõtmisel on sisendiks tulud ja kulud, kogud ja lisandumised kogudesse, personali koosseis, arvutitöökohad, kataloogimise andmed jmt. Väljundiks loetakse laenutusi (kohalkasutust, kojulaenutusi, avariikulikasutust), teatmeteendindust, RVL-i teenust, kasutaja koolitust.

Tulemuseks on pakutavate teenuste kvaliteedi ja sisu vastavus raamatukogu kasutaja ootustele; suhted rahastajatega; teenindaja pädevus, oskused ja teadmised; kasutaja infokirjaoskus. Väljundi ja mõju hindamist vajavad raamatukogud, et tõdeda, milline oli saadud efekt, mis järgnes muudatustele (nt. kokkuhoiust saadud tulu ja selle otstarbekas kasutus, teenuste uuendamise ja kvaliteedi parendamisega kaasnenud efekt, struktuuride ja tööprotsesside muutustega kaasnenud parem tulemus jm.). Oluline on, et tulemus avalduks nii raamatukogu arengutendentsides kui ökonoomikas.

# Suuremahulised andmed DNA sekveneerimisest

**Maido Remm**

*Tartu Ülikool, Molekulaar- ja Rakubioloogia Instituut*

Ettekandes tutvustan järjestuse andmeid, mida saadakse niinimetatud 2. põlvkonna sekveneerimise tehnoloogiate abil. Tänapäeva bioloogias uuritakse DNA sekveneerimise abil nii inimeste kui bakterite DNA varieeruvust; tulemusi rakendatakse väga erinevates tervise- ja keskkonnauuringutes. Teen väikese ülevaate sellest mida bioloogid nende järjestuste abil tahavad küsida ja kuidas saaks statistiline maailmavaade selle juures abiks olla. Juttu tuleb ka saadavatest järjestuse andmete formaatidest ja andmeanalüüsi probleemidest.

# Riskidest ja prognoosimisest TÜ Eesti Geenivaramu kohordis

**Krista Fischer**

*TÜ Eesti Geenivaramu*

TÜ Eesti Geenivaramu ees seisab ülesanne anda geenidoonoritele tagasisidet nende geneetiliste riskide kohta. Selleks on võimalik kasutada suurte ülegenoomsete assotsiatsiooniuuringute põhjal saadud teadmisi haigusega seotud geenimarkerite kohta ja neile vastavate parameetrite hinnanguid. Samas tuleb arvestada ka sellega, et enamlevinud krooniliste haiguste korral (nt südame-veresoonkonnahaigused) võib geenimarkerite abil kirjeldada vaid väikest osa riskist. Märksa suurem mõju on erinevatel käitumuslikel ja keskkonnaga seotud teguritel, millest enamik on samuti geenidoonori küsimustiku abil registreeritud. Lisaks rahvusvaheliste uuringute põhjal saadud teadmistele ja hinnangutele on riskitegurite mõju võimalik hinnata TÜ Eesti Geenivaramu kohordis, kuhu kuulub enam kui 50000 geenidoonorit. Neist enam kui 15000-l on praeguseks hetkeks olemas ka info erinevate geenimarkerite kohta. Arvestada tuleb veel sellega, et erinevate riskitegurite prognoosivõime uurimiseks ei piisa vaid statistilise olulisuse hindamisest – hoopis olulisem on teada, kui palju reaalselt erineb uuritava haiguse levimus kõrge ja madala riskihinnangu saanud indiviididel ning kui paljude inimeste riskihinnang muutub korrektsemaks pärast geenimarkerite arvessevõtmist. Selleks on võimalik kasutada näiteks ROC-analüüsi kui ka reklassifitseerimise indeksit – nii ühel kui teisel on oma eelised ja puudused.

Ettekandes käsitleme neid teemasid TÜ Eesti Geenivaramu kohordi ja kahe levinud kroonilise haiguse – hüpertensiooni ja 2. tüüpi diabeedi – näitel.

# Ülegenoomsed assotsiatsioonianalüüsid ja nende meta-analüüs

Reedik Mägi

*TÜ Eesti Geenivaramu*

Ülegenoomne assotsiatsioonianalüüs on osutunud edukaks meetodiks komplekstunnuseid ja haiguseid mõjutavate uute geneetiliste lookuste leidmisel. Paraku on aga selgunud, et selliste tunnuste geneetika on varem arvatust palju keerulisem – komplekstunnuseid mõjutavad nii geneetilised, keskkonna-, kui ka elustiilist tulenevad faktorid, milledest paljud on seni veel tuvastamata. Neid tunnuseid ja haigusi mõjutavad kümned ja sajad geenilookused milledest igaihel on väga väike efekt. Selleks, et neid efekte tuvastada, on vaja kasutada väga paljude geenidonorite andmeid koos ja seda on võimalik saavutada rahvusvaheliste konsortsiumite raames kogutud andmeid meta-analüüsid.

# Korrelatsioonimaatriksite illustreerimisest – näiteid eluteadustest

**Tanel Kaart**

*Eesti Maaülikool*

Kahe tunnuse vaheline seos on kirjeldatav ühe korrelatsioonikordajaga, kolme tunnuse omavaheliste paarikaupa seoste tarvis on vaja juba kolme, nelja tunnuse puhul kuut, kümne tunnuse puhul 45 ja 100 tunnuse puhul 4950 korrelatsioonikordajat. Kümneid korrelatsioonikordajaid (ja infot nende statistilise olulisuse kohta) sisaldav korrelatsioonimaatriks võib olla atraktiivne küll mõne spetsiifilise seose vastu huvi tundvale erialateadlasele, aga üldisema pildid saamiseks, võimalike mustrite tuvastamiseks või erinevais gruppides ilmnevate seoste võrdlemiseks paljalt tihedalt esitatud arvudest ei piisa. Kuigi enamike kirjeldatud ülesannete lahendamiseks on olemas omad mitmemõõtmelise analüüsi meetodid, on vahel hea osata graafiliselt esitada ka kõige tavalisemat korrelatsiooni- (või mistahes erinevuse- või sarnasusemõõtude) maatriksit – selle alusel tehtavad järeldused on intuiivselt mõistetavad ka suhteliselt statistikakaugetele lugejatele-kuulajatele.

Käesolevas ettekandes esitataksegi mitmeid viimase aasta jooksul erinevatel erialadel (loomakasvatus, ökoloogia, psühholoogia) teadusartiklites ja aruannetes publitseeritud ettekande autori poolt konstrueeritud jooniseid ning diskuteeritakse nende sobivuse ja täiendamisvõimaluste üle.

# Pikaajaliste hüdrobioloogiliste andmete analüüs Võrtsjärve zooplanktoni näitel

Marina Haldna ja Juta Haberman

*Eesti Maaülikooli Limnoloogiakeskus*

Võrtsjärve zooplanktonit on igakuiselt ühe ja sama isiku poolt uuritud alates 1964. aastast käesoleva ajani, see tähendab 48 aastat. Andmebaasis on esitatud järve avavees esinenud 30 zooplanktoni liiki, millede kohta on määratud nende arvukused (hulk ühes kuupmeetris vees). Iga liigi pikkuse (mõõdetud mikroskoobi all) alusel on arvutatud vastavad biomassid. Antud andmebaas võimaldab teha järeldusi mitte ainult zooplanktoni, vaid zooplanktoni baasil ka kogu järve troofilise taseme ja ökosüsteemi kohta. Matemaatiliste meetodite kasutamine taoliste andmete töötlemisel on vältimatu. Töö põhieesmärk oli välja selgitada zooplanktoni indikatiivne väärtus Võrtsjärve seisundi hindamisel. Antud töö raames analüüsiti zooplanktonit liikide, perekondade, rühmade ja kogu zooplanktoni arvukuste põhjal. Iga nimetatud näitaja omab järve troofsuse hindamisel oma spetsiifilise indikatiivse väärtuse. Kasutades mitmemõõtmelise analüüsi meetodeid (statistikapaketi R koosluseökoloogia kirjeldamise programmi *vegan* erinevaid protseduure) tehti kindlaks olulised muutused zooplanktoni liigilises koostises. Algaastatel oli domineerivaid liike rohkem kui vahepealsetel, intensiivsema põllumajanduse perioodidel, puhtaveeliste järvede zooplanktoni liigid on Võrtsjärvest kadunud, domineerivad tugeva eutroofsuse indikaatorliigid. Zooplanktoni kolme rühma (verikirbulised, aerjalgsed, keriloomad) ja nende perekondade indikatiivne väärtus seisneb selles, et nende osakaalude (%) muutumine kogu zooplanktonis iseloomustab adekvaatselt veekogus toimunud pikaajalisi muutusi, mis langesid kokku varemavaldatud tulemustega Võrtsjärve hüdrokeemiliste ja -bioloogiliste muutuste kohta. Võrtsjärve troofsus tõusis 1970ndatel, on praeguseks stabiliseerumas ning mõne näitaja osas paranemas.



# Optimaalne disain Rootsi põdra- ja metskitsepopulatsiooni suuruse hindamiseks ekskrementide loendamise meetodi abil

Kristi Kuljus ja Bo Ranneby

*Rootsi Põllumajandusülikool, Umeå/  
Swedish University of Agricultural Sciences, Umeå*

Põtrade ja metskitsede inventeerimist ekskrementide loendamise meetodi abil viiakse Rootsis läbi igal aastal. Inventeerimise põhieesmärgiks on jälgida muutusi põdra- ja kitsepopulatsioonis. Ekskrementide loendamine toimub alaliste ruudukujuliste proovialade külgedele süstemaatiliselt paigutatud ringikujulistes proovitükkides. Optimaalse disaini leidmisel tuleb arvesse võtta nii statistilisi kui inventeerimise tasuvust puudutavaid aspekte. Seetõttu on ruudukujuline prooviala tavaliselt suurusega, mis vastab loendajate ühepäevasele töökoormusele. Peamiseks statistiliseks kriteeriumiks selles uurimuses on ühel maa-alaühikul leiduvate ekskrementide gruppide arvu hinnangu dispersiooni minimeerimine. Töös on vaadeldud erinevaid disaine – varieeritud on ruudukujulise prooviala külje pikkust ning ringikujuliste proovitükkide arvu, suurust ja paigutust – ja võrreldud neile vastavaid dispersioone.

## Optimum pellet-group count design for estimating the size of the Swedish elk and roe deer populations

**Keywords:** pellet group counting, elk, roe deer, optimum design.

Pellet-group count inventories for studying the size of the elk and roe deer populations in Sweden is carried out annually. The main purpose is to obtain information about changes in the elk and roe deer populations. Pellet group counting is performed in permanent landscape tracts, where a number of circular study plots are allocated along the sides of each square-shaped tract. To find an optimum design, both statistical and cost-benefit aspects have to be considered. This leads usually to a tract size corresponding to one working day for the field crew. The main statistical design criterion used in this study has been to minimize the variance of the estimated number of pellet groups per area unit for the studied region. To achieve the criterion, the length of the tract side, and the number, size and layout of the study plots along the tract sides have been varied and the corresponding variances compared.

# Eesti metsaregistri andmestik ja selle analüüsi probleemidest

Allan Sims, Andres Kiviste, Artur Nilson

*Eesti Maaülikool*

Metsaressursi arvestuse riiklik register ehk metsaregister on metsa inventeerimise andmete riiklik register, kuhu kantakse metsamajanduskavade jaoks tehtavate metsade inventeerimise andmed. Üldiselt ei ole inventeerimise andmed vanemad kui 10 aastat ning andmeid uuendatakse pidevalt. Viimase 10 aasta jooksul siiski kõiki Eesti metsi inventeeritud ei ole ning seetõttu on metsaregistris andmed umbes 75% Eesti metsade kohta.

Metsaregistrisse kantud andmete hulgas on nii mõõdetud kui ka mudelitega arvutatud tunnuseid. Kuna mudeleid täiustatakse ning samade tunnuste arvutamiseks on erinevatel aegadel kasutatud erinevaid mudeleid, siis sellest tulenevalt muutuvad ka tulemused, kuid metsaregistris hoitakse andmeid sellisel kujul nagu need mõõtmisjärgselt esitati. Sellest tulenevalt on andmestikus olevad arvutatud tunnused erinevate väärtustega, kuna need on arvutatud erinevatel aegadel.

Metsaregistri andmeid kasutatakse metsandusliku statistika tegemiseks, mille üheks väljundiks on Eesti metsanduse aastaraamat, mis annab ülevaate metsaressursist. Metsad on inventeeritud erinevatel aegadel, aga kuna metsade seisund pidevalt muutub ning arvutusteks on kasutatud erinevaid mudeleid, siis andmestikku korrastamine adekvaatse ülevaate saamiseks on esmane probleem, mis vajab lahendamist.

Pika aja jooksul kogutud mahukas andmestikus esineb nii süstemaatilisi kui ka juhuslikke vigu. Andmesisestuse käigus kasutatavad vigade kontrolli programmi pidevalt täiendatakse ning uued sisestatud andmed sisaldavad järjest vähem vigu. Paraku ei saa veatuvastus programme rakendada varem sisestatud andmetele, kuna antud andmestik on koopia metsaomaniku käes olevatest andmetest. Vigu saab küll kõrvaldada uurimistööks kasutatavas andmebaasi koopias, kuid metsaregistrisse jäävad vead alles. Antud andmestikus on vigade kõrvaldamise protseduur uurimistöös igakordne pärast uue koopia tegemist.

Eesti metsaregistri andmestik on olnud ja on ka praegu hea andmestik puistu kasvukäigu ja paljude muude metsa mudelite loomiseks ja arendamiseks. Viimaste aastate uuringutest on selgunud, et seoses keskkonnaningimuste muutumisega on metsaregistri andmetel moodustatud aegread erinevad võrreldes puistute tegeliku kasvuga, mis on saadud püsiproovitükkide ja puuanalüüsi andmetest. See on omaette tõsine probleemistik metsa kasvu alases uurimistöös.

# Eluea ja tervena elatud aastate pikendamise võimalused Eestis lähematel aastakümnetel

Ülo Kristjuhan

*Tallinna Tehnikaülikooli emeriitdotsent*

Inimene on alati unistanud pikast elust, seejuures tervena, ilusana, vabana haigustest. Keskmine oodatav eluiga riigis on oluline elukvaliteedi näitaja, mis sõltub paljudest tegureist: füüsilistest, keemilistest ja psühholoogilistest keskkonnategureist, tervisekäitumisest, meditsiinilisest teenedusest, tehtavast tööst, kasutatud arstimeist ning toidust. Aastal 2005 oli keskmine eluiga Eestis 72,82. Aastaks 2010 oli keskmise eluea pikkus Eestis tõusnud 75,84 aastani, sealhulgas meestel 70,62 ja naistel 80,52 aastani. Vabariigi valitsuse kava kohaselt peaks oodatav eluiga aastaks 2015 tõusma meestel 72,5 ja naistel 82,2 aastani. Mis saab järgnevatel aastakümnetel? Reservid on suured. Sellele osutavad statistilised andmed nende maade ja piirkondade kohta, kus eluiga on oluliselt kõrgem. Juba aastal 2001 oli Londoni Kensingdon ja Chelsea piirkonnas meeste eluiga 84,3 aastat ja naiste eluiga 88,9 aastat. Viimastel andmetel on seal ühe väiksema piirkonna, Queen's Gate, meeste keskmine eluiga 88.3 ehk ligi 18 aastat pikem kui Eestis. Vastavalt *World Factbook* hinnangu andmetele on aastal 2012 Monaco inimese keskmine eluiga juba 89.68 aastat ehk ligikaudu 14 aastat pikem kui Eestis.

Uuringud näitavad, et inimese eluiga on suurtes piirides muudetav. Longitudinaalsed uuringud on näidanud, et eluviisi mõne teguri kompleksi kaudu on võimalik lisada kuni 14 eluaastat. Eksperimentaalsed uuringud on näidanud, et kasutades kõiki teadaolevaid võimalusi on inimesel võimalik edasi lükata vananemisprotsesse kuni 20 aastat. Seega on lähematel aastakümnetel võimalik keskmine eluiga Eestis 90 aastat. Eestis oleks kasulik koostada "Elanikkonna noorusaastate pikendamise ja vananemise edasilükkamise strateegia".

# Eesti sisemajanduse koguprodukti prognoosimisest

Mihkel Täht

*Statistikaamet*

Riigi heaolu on võimalik hinnata teatud koondnäitajate ja nende käitumise analüüsi kaudu. Need näitajad võivad olla näiteks tarbija- ja tootjahinnaindeks, sisemajanduse koguprodukt (SKP), monetaarsed ja teised indikaatorid. Vastava analüüsi tulemused on abiks teadlastele, poliitikutele, poliitika kujundajatele, analüütikutele, panganduse juhtidele ja teistele spetsialistidele.

Väga oluline koondnäitaja on SKP. On teada, et vastava kvartali SKP-d avaldatakse teatud hilinemisega. Statistikaamet avaldab SKP esimese hinnangu (kiirhinnangu) mitte hiljem kui 42. ja järgmise 70. päeval pärast vastava kvartali lõppu. Nende hinnangute saamiseks kasutatakse erinevaid andmeallikaid, seetõttu on nende väärtused erinevad. Kui me tahaksime kasutada saadud SKP hinnangut lühiajaliste arengustsenaariumide jaoks, siis võivad tekkida probleemid just andmete hilinenud avaldamise tõttu. Teisiti öeldes, me tahame saada andmeid varem. Sellisel juhul tuleks SKP hinnangute arvutamiseks lähtuda nende prognoosidest.

Autor on uurinud suhteliselt lihtsate meetodite rakendamise võimalusi SKP lühiajaliseks prognoosimiseks. Ettekandes lähtutakse kahest prognoosimise võimalusest:

1. SKP aegrea prognoosimine sesoonse *ARIMA* mudeli abil.
2. SKP aegrea jaotamine nähtamatuteks komponentideks (trend, sesoonne ja irregulaarne komponent) aditiivse või multiplikatiivse mudeli alusel ja nende komponentide hilisem prognoosimine. Sellisel juhul modelleeritakse kõik komponendid eraldi ja seejärel prognoositakse nende väärtused. Trendi modelleerimisel ja prognoosimisel on kasutatud polünoomiaalseid mudeleid. Polünoomi järku saab kindlaks määrata Akaike informatsiooni kriteeriumi (AIC) abil. Konkreetse kvartali sesoonse komponendi modelleerimisel saab rakendada sama meetodikat. Eeldatakse, et irregulaarne komponent kujutab endast "valget müra".

Esitav meetodika on rakendatav vaid lühiajaliste (ühe või kahe kvartali jaoks) prognooside korral. Ettekandes antakse lühiülevaade nende meetodite, prognooside kvaliteedi hindamise, saadud tulemuste ja praktiliste soovitude kohta.

# The normal inverse Gaussian distribution: Exposition and applications

Dean Teneng

*University of Tartu*

We expose the unique properties of the normal inverse Gaussian distribution (NIG) useful for modeling asset prices, indexes and foreign exchange (FX) closing prices through the generalized hyperbolic class of distributions. We consider short, medium and long term exchange rate closing prices and demonstrate that traditional beliefs in exchange rates not being independently identically distributed random variables are fundamentally flawed.

## Indeks

Bogdanov, 25

Fischer, 29

Gołata, 12

Haberman, 32

Halapuu, 15

Haldna, 32

Jögi, 27

Kaart, 31

Kangro, 24

Kiviste, 34

Kont, 27

Kristjuhan, 35

Kuljus, 33

Leontjeva, 22

Matteus, 14

Mägi, 30

Nilson, 34

Paas, 15

Pelt, 26

Pärna, 24

Ranneby, 33

Remm, 28

von Rosen, 19

Saar, 21

Servinski, 18

Sims, 34

Sirel, 23

Teneng, 37

Tiit, 9, 13

Traat, 3, 4

Täht, 36

Ulnicāns, 17

Võhandu, 11