

NEKAJ O NASLOVNICI

STAGE – Statistika & Geografija – je odprtokodna interaktivna spletna kartografska aplikacija, ki uporabnikom omogoča vizualizacijo, interpretacijo in prenos geoprostorskih statističnih podatkov. V STAGE-u:

- spremljamo časovni razvoj različnih pojavov v kombinaciji več kot 300 spremenljivk, 20 statističnih področij in 10 teritorialnih ravni,
- združujemo prostorske enote v nova zaključena območja ter ustvarjamo nove statistike po svoji meri,
- izdelane kartografske poglede delimo z drugimi uporabniki ali jih integriramo v svoje spletne strani.

STAGE 2.0 ima posodobljen uporabniški vmesnik zasnovan v odzivnem načinu in je prilagojen tudi uporabi na malih zaslonih. Napredni uporabniki bodo deležni izboljšane proizvodnje po prostorskih podatkih z možnostjo prikazovanja le-teh tudi v obliki grafikona. STAGE bo dopolnjen tudi vsebinsko, saj prihajajo novi kazalniki s področja zdravstvenih in ekonomskih statistik.

Avtor prispevka je Igor Kuzma.

KAZALO

Uvodnik	5
Priznanja Statističnega društva Slovenije 2017	
Podelitev priznanj	7
Prejemnica Blejčevega priznanja	9
Priznanje odličnosti statističnega poročanja v medijih	10
Častni član Statističnega društva Slovenije	11
Priznanja mladih statistikov	12
<i>Uporaba psevdo vrednosti v analizi relativnega preživetja</i>	13
<i>Generiranje omrežij z upoštevanjem različnih tipov triad</i>	15
<i>Intervali zaupanja za test Mann-Whitney</i>	17
Ob 70-letnici prof. dr. Anuške Ferligoj	19
Domači in mednarodni dogodki	
Konferenca Uporabna statistika 2017	21
<i>Uvajanje e-poročanja na kratkoročnih poslovnih statistikah</i>	21
Statistični dan 2018 – Moč podatkov	22
Strokovni seminar – Kombinirani načini zbiranja podatkov	25
22. srečanje mladih statistikov – Kaj raziskujejo doktorski študenti statistike?	27
Delavnica o simbolni analizi podatkov SDA 2017	30
<i>Razvrščanje simbolnih podatkov z relacijsko omejitvijo: Demografske strukture po spolu in starosti po Evropi</i>	32
Konferenca ICOS2017 v Sarajevu	42
Letno srečanje skupine V7	44
Aktualno	
Dvajset let arhiva družboslovnih podatkov	46
Konzorcij arhivov družboslovnih podatkov (CESSDA)	51
Statistika za boljši svet	53
Evropske statistične igre	57
Dogodki Ekonomske fakultete v Ljubljani	60

Izobraževanje

Predavanja na IBMI – <i>Vzorci v glasbi</i>	61
Povzetki doktorskih disertacij	65

Raziskujemo

Uporaba mikro podatkov za razvoj mikrosimulacijskih modelov	78
Vpliv volitev in drugih pomembnejših dogodkov na zaupanje potrošnikov v Sloveniji	81

Zanimivo

Nemoč anketnih raziskav	85
-------------------------------	----

Skupščina

Zapisnik 26. redne skupščine Statističnega društva Slovenije	87
--	----

Urednica:	Jerneja Čuk
Odgovorni urednik:	dr. Matevž Bren, predsednik Statističnega društva Slovenije
Uredništvo:	člani Izvršnega odbora: dr. Andrej Blejec, dr. Anuška Ferligoj, Boro Nikić, Ema Mišič, mag. Irena Vipavc Brvar
Urednica spletne strani:	mag. Irena Vipavc Brvar
Oblikovanje:	Anica Žinger
Oblikovanje naslovnice:	mag. Matjaž Erker
Tisk:	Birografika Bori, d. o. o.
Slika na naslovnici:	mag. Matjaž Erker
Naklada:	150 izvodov

ISSN 1408-3272

Naslov društva	Statistično društvo Slovenije Litostrojska cesta 54, 1000 Ljubljana
Spletna stran	www.stat-d.si
Elektronski naslov:	info@stat-d.si
Telefon:	01 2416 400

AVTORJEM PRISPEVKOV:

Bilten izhaja enkrat letno, vendar le, če je dovolj prispevkov. Zato vse prijazno vabimo, da strokovne in aktualne prispevke, zapise o udeležbah na različnih strokovnih dogodkih, napovedi ali o drugih zanimivostih, ki se vam zdijo vredni, da jih objavimo in z njimi seznanimo statistično javnost, sproti pošiljate uredništvu. Le tako jih bomo ohranili v spominu in tudi tako pomagali izpolnjevati poslanstvo društva.

Prispevke pošiljajte urednici, predsedniku ali članu izvršnega odbora, ki vam je najbližji. Ne pozabite pošiljati sporočil tudi za našo spletno stran.

Naši elektronski naslovi so:

Jerneja.Cuk@gov.si, Matevz.Bren@fvv.uni-mb.si, Andrej.Blejec@nib.si, Anuska.Ferligoj@fdv.uni-lj.si, Boro.Nikic@gov.si, Ema.Misic@gov.si, Irena.Vipavc@fdv.uni-lj.si

UVODNIK

Spoštovani prijatelji, kolegi, člani društva!

Dejavnost Statističnega društva Slovenije je pestra, kar dokazuje tudi ta številka Biltena pred vami. Kazalo nam kaže bogato, raznoliko vsebino od dejavnosti društva, praznovanja 40-letnice delovanja, nagrajencev za leto 2017 in nagrajencev razpisa za mlade statistike 'Inovativni statistični pristopi', do domačih in mednarodnih dogodkov ter raziskovanja in izobraževanja. Vsa zahvala za tako bogat Bilten gre, poleg avtorjem, v veliki meri urednici Jerneji Čuk in je rezultat njenega prizadevanja poiskati zanimive vsebine in ustrezno nagovoriti avtorje.

V uvodu želim okvirno izpostaviti naše dejavnosti v preteklem letu. V lanskem uvodniku Biltena sem izpostavil, da smo delo društva še posebej usmerili v

- terminologijo, saj lahko prav tu članom in tudi ostalim uporabnikom statistike ponudimo neposredno pomoč ter sledimo novostim stroke in znanosti;
- statistično pismenost z združevanjem prizadevanj vseh, ki na srednje in višje/visokošolskem izobraževanju poučujemo statistiko;
- urejanje spletne strani, dodajanje vsebin, ki bi bile na več ravneh in bi bil samo članom društva omogočen dostop do vseh; predvsem naj bi bil članom omogočen vstop v terminološki kotichek z možnostjo postavljanja vprašanj in predlogov poimenovanja.

Tako smo se na Inštitutu za biomedicinsko informatiko (IBMI) sestali člani terminološke sekcije in se ukvarjali s slovenskimi poimenovanji angleških izrazov, ki so se nabrali kolegom na IBMI. Na tem mestu izrekam zahvalo predsedniku komisije Jaru Lajovicu za temeljito pripravo virov (kako so to poimenovali Nemci, kako Francozi ...) in pojasnila o latinskem ali starogrškem korenu. S tem nam je omogočil celovit vpogled in v veliki meri olajšal naše delo; pa tudi dogovorjena nova poimenovanja je že vključil v glosar www.rosigma.si/glosar Pohvala pa gre tudi ideji, da magistrski študenti statistike v svojih nalogah predložijo tudi spisek novih izrazov, ki še nimajo ustreznega poimenovanja. Naslednje srečanje terminološke sekcije bo na SURS-u, kjer nam že pripravljajo spisek izrazov brez ustreznega slovenskega poimenovanja.

V sodelovanju s Statističnim uradom (uspešno so se prijavi na razpis Eurostata) izvajamo tekmovanje srednješolcev 'Evropske statistične igre'. Za tekmovanje se je prijavilo 142 ekip, ki jih sestavlja 397 dijakov in 36 mentorjev iz 34 šol. Prof. Blejec je v decembru izvedel štiri lepo obiskane in pohvaljene seminarje za mentorje, v žiriji za izbor pa sodelujeta mag. Irena Vipavc Brvar in dr. Aleš Toman. Glavni cilji 'Statističnih iger' so spodbujanje radovednosti in zanimanja za statistiko med dijaki ter spodbujanje učiteljev, da za poučevanje statistike uporabijo realne podatke ter iščejo priložnosti za uporabo pridobljenega statističnega znanja. Še kaj več o statističnih igrah si lahko preberete na <http://www.stat.si/statweb/News/Index/6992>.

Urejanje spletne strani poteka bolj počasi, pa vseeno nam je uspelo prenesti ves arhiv Metodoloških zvezkov na drug strežnik (hvala sodelavcem IBMI) in s tem sprostiti prostor

na zakupljenem strežniku. Iskanje novega ponudnika in ustreznega orodja za oblikovanje in posodabljanje spletne strani pa še poteka.

Iz vsebine želim izpostaviti še jubilej prof. Anuške Ferligoj. Prepričan sem, da delim mnenje vseh članov društva, da je njena vloga pri uveljavljanju statistike v Sloveniji tako na raziskovalnem kot pedagoškem področju neprecenljiva, enako tudi njen prispevek pri delovanju društva. Na kratko 'brez njenega dela nič ne bi bilo tako, kot je' in se ji za opravljeno delo v imenu članov društva zahvaljujem.

Zahvaliti se želim tudi vsem udeležencem in žirantom razpisa za mlade statistike 'Inovativni statistični pristopi', ki smo ga izvedli v okviru praznovanja 40-letnice delovanja društva (dr. Mojca Bavdaž, Boro Nikić, dr. Maja Pohar Perme in dr. Aleš Žiberna). Prispevki poster sekcije in podelitev nagrad so dodali ustrezno vsebino praznovanju. Zahvaljujem se tudi vsem, ki so sodelovali pri pripravi in izvedbi tega dogodka in prijetnega druženja, pa tudi vsem, ki so v društvu delovali in nam s svojim delom omogočili, da tudi sedaj v okviru društva tvorno sodelujemo in živimo skupne cilje.

Predsednik društva
prof. dr. Matevž Bren

PRIZNANJA STATISTIČNEGA DRUŠTVA SLOVENIJE 2017

PODELITEV PRIZNANJ

Statistično društvo Slovenije je konec leta 2017 praznovalo 40. obletnico delovanja. Ta jubilej smo zaznamovali s slavnostno skupščino, ki smo jo pripravili 17. oktobra 2017 na Fakulteti za družbene vede. Dogodka se je udeležilo več kot šestdeset oseb iz različnih ustanov, ki aktivno sodelujejo z nami že vrsto let.

Na slavnostni skupščini smo podelili priznanja Statističnega društva Slovenije, priznanja natečaja za najboljše raziskovalne dosežke mladih statistikov z naslovom '*Novi statistični pristopi in izvirne uporabe statistike*' ter zahvale vsem dozrajšjim predsednikom društva.

Ob tem dogodku smo izdali tudi posebno številko *Biltena*.

Navzoče na skupščini je nagovoril predsednik društva Matevž Bren in na kratko opisal delovanje in aktivnosti društva v vseh letih. Sledil je govor Iva Lavrača, ustanovnega člana društva, ki nas je v svojem govoru popeljal prav na začetek, do nastanka društva in ustanovne skupščine, ki je bila 14. 12. 1977. Nato smo podelili zahvale vsem nekdanjim predsednikom društva, tajniku in urednicama *Biltena*. Hkrati s podelitvami zahval po posameznih letih smo prikazali zanimivo časovnico zadnjih štiridesetih let, ki jo lahko vidite tudi na naslovni strani omenjene posebne izdaje *Biltena*.



Zadaj (od leve proti desni): dr. Ivo Lavrač, dr. Anuška Ferligoj, dr. Anton Kramberger, dr. Andrej Blejec, Bogdan Grmek
Spredaj (od leve proti desni): dr. Janez Malačič, dr. Lea Bregar, dr. Katarina Košmelj, dr. Janez Stare, mag. Katja Rutar, dr. Matevž Bren

Na kratko nas je nagovoril tudi Andrej Blejec, ki je bil v vlogi predsednika društva dvanajst let. Sledil je slavnostni govor direktorice Statističnega urada RS Genovefe Ružić, nato je sledila slavnostna podelitev priznanj društva.

Razpis za zbiranje predlogov prejemnikov priznanj smo objavili v mesecu juniju 2017, prav tako tudi natečaj za najboljše raziskovalne dosežke mladih statistikov. Čas za prijave je bil do 4. septembra 2017. Do takrat smo na sedež društva prejeli dva predloga za Blejčevo priznanje, dva predloga za priznanje za odličnost poročanja v medijih in en predlog za častnega člana društva.

Društvena priznanja, Blejčevo priznanje 2017 (imenovano po prof. dr. Marijanu Blejcu, začetniku moderne statistike v Sloveniji) za vrhunske dosežke v statistiki smo podelili Genovefi Ružić. Priznanje za odličnost statističnega poročanja v medijih za leto 2017 je prejel dr. Jure Stojan, častni član društva za leto 2017 pa je postal dr. Tamas Rudas. Utemeljitev predlogov objavljamo v nadaljevanju.

Po podelitvi priznanj društva je sledila predstavitev razpisa *'Novi statistični pristopi in izvirne uporabe statistike'*. Plakati so bili razstavljeni na hodniku, zato smo vse navzoče na skupščini povabili, naj si jih skupaj z nami ogledajo.

Na natečaj za najboljše raziskovalne dosežke mladih statistikov se je prijavilo pet mladih raziskovalcev. Štiričlanska strokovna komisija je izbrala tri najboljše. Slike teh plakatov s pripadajočimi besedili in z obrazložitvami izbire objavljamo v nadaljevanju.

Sledila je prijetna razprava med udeleženci. Poskrbeli smo tudi za pogostitev in prijetno vzdušje.

Dogodek je bil tudi medijsko odmeven. Krajši članek, ki ga je napisala dr. Mojca Vizjak Pavšič, dobitnica priznanja za odličnost statističnega poročanja v medijih za leto 2016, smo lahko prebrali v časniku Delo. Dogodek bil objavljen tudi na spletni strani Tiskovne agencije STA, na SURS-ovem profilu na Twitterju ter na spletni strani Fakultete za družbene vede Univerze v Ljubljani.

PREJEMNICA BLEJČEVEGA PRIZNANJA ZA LETO 2017: GENOVEFA RUŽIČ

Tatjana Novak
Statistični urad RS

Utemeljitev predloga, da se Genovefi Ružič podeli Blejčevo priznanje Statističnega društva Slovenije.

Genovefa Ružič je generalna direktorica Statističnega urada RS od leta 2013. Pred tem je bila deset let namestnica dotedanje predstojnice SURS-a. V zadnjem obdobju je aktivna v evropskem in mednarodnem prostoru. Sodelovala je pri pripravi vizije evropskega statističnega sistema do leta 2020. Predseduje usmerjevalni skupini (PG - partnerski skupini) Evropskega statističnega sistema (2016-2017), ki se ukvarja s ključnimi strateškimi vprašanji delovanja in nadaljnega razvoja evropskega statističnega sistema. V času njenega vodenja PG-ja so bili razviti osnovni principi za izmenjavo zaupnih podatkov podjetij za statistični namen ter priporočila za modernizacijo statistike blagovne menjave med državami članicami EU. Trenutno PG pripravlja predlog ESS prioritet po letu 2020. Je članica skupine na visoki ravni za posodabljanje (modernizacijo) uradne statistike, ki deluje v okviru Statistične komisije združenih narodov za Evropo (HLG-MOS) in biroja odbora OECD za Statistiko in Politiko statistike (2016-2017). Na nacionalni ravni sodeluje pri pripravi Strategije razvoja Slovenije 2030. Ocenjujemo, da s svojim delom pomembno prispeva tako pri razvoju državne statistike kot prepoznavnosti slovenske statistike v mednarodnem okolju zato jo predlagamo za Blejčevo priznanje.



Prejemnica Blejčevega priznanja Genovefa Ružič in predsednik društva Matevž Bren

PRIZNANJE ODLIČNOSTI STATISTIČNEGA POROČANJA V MEDIJIH ZA LETO 2017 SMO PODELILI DR. JURETU STOJANU

Andraž Zorko
Podjetje Valicon d.o.o.

Dr. Jure Stojan je s svojimi prispevki na SIOL.NET pripeljal statistiko v množične medije skozi velika vrata. Čeprav je do danes objavil »le« osem prispevkov na tem spletnem mestu, pa si ti prispevki zaslužijo pozornost in, po mojem prepričanju, tudi priznanje odličnosti statističnega poročanja v medijih. Za razliko od običajne rabe statistike v medijih, ki še vedno le redkokdaj presega suhoparno navajanje odstotkov, kot presežek pa lahko razumemo že bolj privlačne prikaze le teh s t.i. infografiko, Jure Stojan statistiko uporabi za utemeljevanje – ali zavračanje - trditev, ki se navadno pojavijo že v naslovu. Njegovi prispevki so poglobljeni, uporabljeni viri mnogoteri, prikazi in predstavitev dejstev in ugotovitev pa hkrati dovolj enostavni in predvsem prepričljivi, da so razumljivi in zanimivi tudi širšemu občinstvu. In to je namen tega priznanja – spodbujanje in nagrajevanje avtorjev, ki znajo s svojimi prispevki v medijih statistiko približati širšemu občinstvu, izven kroga ljubiteljev in statističnih zanesenjakov, poglobljeno, nazorno, pa vendar jasno in predvsem razumljivo. Kot primer navajam prispevek »Pod socializmom se ni živelo bolje« (31. 10. 2016, dostop: <http://siol.net/posel-danes/novice/pod-socializmom-se-ni-zivelo-bolje-428796>).



Dr. Matevž Bren in prejemalec priznanja dr. Jure Stojan

V dnevih pred objavo prispevka se je, iz kdo bi še vedel katerega razloga že, na družbenih omrežjih razplamtela debata o tem, ali smo nekoč, v prejšnji državi, živeli bolje kot danes. Verjetno ni potrebno posebej poudariti, da se je diskurz kmalu spustil na precej nizko raven, kjer so se udeleženci razdelili na znana tabora »levih« in »desnih«, »naših« in »vaših«, pri tem pa seveda povsem pozabili na kakršnakoli oprijemljiva dejstva ali argumente, s katerimi bi svoja, goreče zagovarjanja stališča, podkrepili. Jure Stojan je nato objavil prispevek, v katerem je z uporabo različnih statističnih kazalcev dokazal, da se v prejšnji državi pač ni

živelo bolje, ne glede na to, kako nostalgичno lepe spomine imamo morda na to obdobje. V obsežnem, pa vendar zelo jasnem zapisu, je v več kot 10.000 znakih, s 17 grafi in 11 citiranimi viri uporabljenih statistik, jasno zavrnil slej ko prej zgolj nostalgичno in zelo pavšalno trditev »nekoč se je živelo bolje«. Prispevek je bil med njegovimi do sedaj najbolj odmeven, če sodimo po številu komentarjev (242) in deljenj z drugimi (815).

V času neznosne lahкости opletanja s pol- ali celo neresnicami, ki skozi ojačevalec družbenih omrežij dobijo nesorazmerno pozornost, in na žalost tudi težo v posledicah, ki jih nosijo s seboj, so prispevki nominiranca najboljša ilustracija, kako enostavno je pravzaprav, z uporabo statistike, vsa ta »alternativna dejstva« ovreči, ne da bi se pri tem ideološko ali kako drugače opredeljevali. Zgled, ki mu moramo dati priznanje in s tem potrebno težo, upajoč, da novinarstvu, javnemu poročanju in argumentiranemu diskurzu vendarle še ni povsem odzvonilo. Ker je v takšnih in v tem primeru še posebej, ravno statistika ključnega pomena, predlagam, da se priznanje odličnosti statističnega poročanja v medijih podeli dr. Juretu Stojanu.

Vsi prispevki nominiranca, objavljeni na spletni strani SIOL.NET, so objavljeni tule: <http://siol.net/avtorji/dr-jure-stojan-22877/>.

ČASTNI ČLAN STATISTIČNEGA DRUŠTVA SLOVENIJE 2017: PROF. DR. TAMAS RUDAS

Prof. dr. Anuška Ferligoj
Fakulteta za družbene vede

Predlagam, da Statistično društvo Slovenije v letu 2017 imenuje za častnega člana prof. dr. Tamasa Rudasa, rednega profesorja na Univerzi Eotvos Lorand (ELTE) v Budimpešti in generalnega direktorja Centra za družbene vede na Madžarski akademiji znanosti, za njegov izjemni doprinos k slovenski statistiki. Prof. dr. Tamas Rudas je zelo ugleden znanstvenik na področju statistike, raziskovalno se ukvarja predvsem z analizo kategorialnih podatkov. Predaval je na Univerzi Eotvos Lorand, Srednje evropski univerzi v Budimpešti in Varšavi, na univerzah v Pensilvaniji, Toledu, Washingtonu, Princetonu, Grazu, Mannheimu, Erfurtu in seveda v Ljubljani. Njegova bibliografija je izjemno bogata, med drugim je objavil več knjig s področja statistike in verjetnosti pri založbi Sage in članke v najuglednejših revijah s področja statistike. Zelo je aktiven v Evropskem združenju za metodologijo, v letih 2012-2014 je bil njegov predsednik. Prejel je več nagrad in priznanj na Univerzi Eotvos Lorand in je izvoljeni član Evropske akademije za sociologijo.

Prof.dr. Tamas Rudas že več desetletij sodeluje s slovenskimi statistiki. Že od samega začetka se je udeleževal metodoloških srečanj v Preddvoru in Ljubljani, bil je tudi vabljeni predavatelj na letni konferenci Uporabna statistika v Ribnem. Od leta 1998 soorganizira letna srečanja mladih statistikov, sedaj petih držav Avstrije, Hrvaške, Italije, Madžarske in Slovenije. Že od samega začetka je član uredniškega odbora revije Metodološki zvezki, kjer je pomagal predvsem z recenzijami člankov. Pomagal je pri vzpostavitvi doktorskega programa Statistika

na Univerzi v Ljubljani in od leta 2003 naprej izvaja predmet Analiza kategorialnih podatkov. Sedaj redno predava v okviru predmeta Sodobni statistični pristopi. Njegova pomoč pri izvajanju magistrskega programa Uporabna statistika in doktorskega programa Statistika v obliki podpore, priporočil in predavanj je bila in je še vedno izjemnega pomena.

Na podlagi navedenih dejstev menim, da je prof.dr. Tamas Rudas izjemno prispeval k razvoju in uveljavitvi statistične znanosti v slovenskem prostoru in zato predlagam, da se prof. dr. Tamasu Rudasu podeli v letu 2017 naziv častni član Statističnega društva Slovenije.



Od leve proti desni: Anuška Ferligoj, častni član Tamas Rudas, Matevž Bren

PRIZNANJA MLADIH STATISTIKOV

'NOVI STATISTIČNI PRISTOPI IN IZVIRNE UPORABE STATISTIKE'

V nadaljevanju poročamo o treh najboljših raziskovalnih dosežkih mladih statistikov v obliki plakatov (posterjev) z naslovom »Novi statistični pristopi in izvirne uporabe statistike«.

Ob 40. obletnici delovanja statističnega društva smo namreč objavili razpis za izdelavo plakata (posterja), katerega vsebina naj bi bil izviren prikaz znanstvenih dosežkov in dosežkov v uporabi statistike, njegov namen pa je bila promocija statistične stroke v najširšem smislu. Razpis je bil namenjen magistrskim in doktorskim študentom programa Statistika Univerze v Ljubljani, študentom vseh slovenskih univerz ter mladimi do 28. leta starosti.

Priporočljive mere plakatov so bile 90 krat 120 centimetrov. Poleg posterja so udeleženci pripravili tudi krajše povzetke, dolge do 300 znakov. Na razpis se je odzvalo sedem študentov z mentorji. Štiričlanska komisija je izmed prispelih plakatov s povzetki izbrala tri, za katere je presodila, da so v največji meri zadostili njenim merilom (izvirnost, prepričljivost, estetski vtis). Slike zmagovalnih plakatov z utemeljitvami komisije tudi objavljamo.

Priznanje za najboljšo predstavitev z vidika izvirnosti in vizualnosti je prejel Klemen Pavlič.

Utemeljitev komisije:

Prispevek razumljivo prikaže raziskovalni problem in nazorno predstavi izvirnost, to je nov pristop k ocenjevanju preživetja s pomočjo psevdo vrednosti, ki temelji na manj predpostavkah od obstoječih, je uporaben na zveznih in diskretnih podatkih ter omogoča nesporno interpretacijo.

UPORABA PSEVDO VREDNOSTI V ANALIZI RELATIVNEGA PREŽIVETJA

Klemen Pavlič

Inštitut za biostatistiko in medicinsko informatiko

Analiza relativnega preživetja se ukvarja z analizo podatkov o rakavih bolnikih, ki lahko umrejo bodisi zaradi raka bodisi zaradi drugih vzrokov. Za te podatke je značilno, da razlog smrti pogosto ni znan oziroma je nezanesljiv. Cilj tega področja je oceniti mero, ki bi bila neodvisna od umrljivosti zaradi drugih vzrokov in bi zato neposredno odražala tveganje za smrt zaradi raka. Velik napredek se je zgodil v zadnjem desetletju, ko je bila predstavljena nepristranska cenilka (PP) čistega preživetja (tj. mera, ki ni odvisna od tveganja za smrt iz drugih vzrokov). Za to cenilko je značilno, da je bila skonstruirana za zvezno merjen čas in zato daje pristranske rezultate v primeru diskretno merjenega časa. Druga, večja pomanjkljivost, pa izhaja iz same mere, saj njena interpretacija sloni na nepreverljivih predpostavkah.

Reševanja teh problemov smo se lotili s pomočjo psevdo vrednosti, ki predstavljajo orodje, ki omogoča ocenjevanje individualnega preživetja za vse paciente (tudi krnjene) ob vseh časih (tudi po krnjenju). Z njihovo pomočjo smo direktno iz definicije mere skonstruirali novo cenilko, ki je preprosta za izračun in temelji na manj predpostavkah kot cenilka PP. Teoretično in s simulacijami smo pokazali, da se v primeru izpolnjenih predpostavk cenilki praktično ujemata. Poleg tega smo razvili dve formuli za ocenjevanje njene variance. Stranski produkt tega je, da smo dobili metodo za izračun variance same psevdo vrednosti, kar predstavlja teoretično novost. Nova cenilka se enostavno razširi tudi na podatke z diskretno merjenim časom, njena definicija pa naravno ponuja novo interpretacijo, ki ni odvisna od spornih predpostavk. Rezultate bomo ilustrirali na primeru podatkov slovenskih bolnic raka in na simuliranih podatkih.

Uporaba psevido vrednosti v analizi relativnega preživetja

Klemen Pavlič

Inštitut za biostatistiko in medicinsko informatiko, Medicinska fakulteta

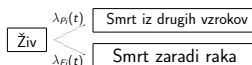


Predstavitel problema in podatkov

Opis podatkov (rakavi bolniki)

Za i -tega pacienta poznamo:

- ▶ čas od diagnoze do smrti
- ▶ status (smrt ali krnjenje)
- ▶ vektor demografskih spremenljivk D_i (tipično spol, starost in leto diagnoze)
- ▶ vektor drugih spremenljivk X_i , ki nas zanimajo
- ▶ $Z_i = (D_i, X_i)$



- ▶ **Skupno tveganje:**

$$\lambda_0(t|Z) = \lambda_E(t|Z) + \lambda_D(t|Z) \quad (1)$$

- ▶ **Pomankljivost podatkov:** razlog smrti neznan/nezanesljiv
- ▶ **Predpostavka:** to pomankljivost odpravimo z vpeljavo informacije o $\lambda_D(t|Z)$ iz populacijskih tabel umrljivosti

Cilj področja

- ▶ Mera, ki povzame λ_E in omogoča primerjavo med različnimi populacijami (državami, časovnimi obdobji)
- ▶ Ne sme biti odvisna od λ_D
- ▶ **Poskus rešitve:** čisto preživetje
- ▶ **Individualno čisto preživetje:**

$$S_E(t|Z) = \exp\left(-\int_0^t \lambda_E(u|Z) du\right)$$

- ▶ **Čisto preživetje skupine:**

$$S_E(t) = \int S_E(t|z) dH(z)$$

- ▶ **Pomankljivost čistega preživetja:** interpretacija - preživetje v hipotetičnem svetu, kjer pacienti lahko umrejo le zaradi raka

Trenutno stanje

- ▶ Cenilka PP [2] za čisto preživetje
- ▶ dosledna*
- ▶ zahteva numerično integracijo
- ▶ izpeljava temelji na nepreverljivih predpostavkah
- ▶ izpeljana je za zvezno merjen čas
- ▶ pristranska v primeru diskretno merjenega časa

Literatura

- [1] P. K. Andersen, J. P. Klein, and S. Rosthøj. Generalized linear models for correlated pseudo-observations, with applications to multi-state models. *Biometrika*, 90:15–27, 2003.
- [2] M. Pohar Perme, J. Stare, and J. Estève. On estimation in relative survival.

Nov predlog rešitve problema

Izhajamo iz enačbe (1), vendar ne zahtevamo, da je količina $S_E(t)$ preživetje:

Cenilka:

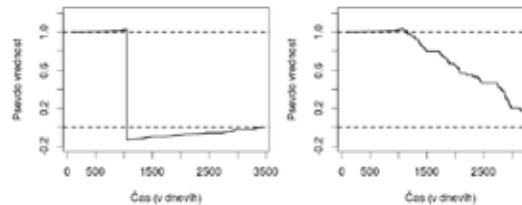
$$\widehat{S}_E(t) = \frac{1}{n} \sum_{i=1}^n \frac{\widehat{S}_{0i}(t)}{\widehat{S}_{Pi}(t)}$$

- ▶ $\widehat{S}_{Pi}(t)$ preberemo iz populacijskih tabel umrljivosti
- ▶ $\widehat{S}_{0i}(t) = i$ -ta psevido vrednost

i -ta psevido vrednost [1]

$$\widehat{S}_{0i}(t) = n \cdot \widehat{S}_0(t) - (n-1) \cdot \widehat{S}_0^{(-i)}(t)$$

- ▶ $n =$ velikost vzorca
- ▶ \widehat{S}_0 in $\widehat{S}_0^{(-i)}$: cenilki skupnega preživetja na celem vzorcu in na vzorcu brez i -tega pacienta (npr. Kaplan-Meier cenilka)
- ▶ **prednost:** i -ta psevido vrednost je definirana za vse čase (ne glede na krnjenje)
- ▶ prispevek i -tega pacienta k skupnemu preživetju [3]



Lastnosti novega predloga ocenjevanja

- ▶ dve cenilki za ocenjevanje variance cenilke (2)
- ▶ stranski rezultat: variabilnost ene psevido vrednosti - **teoretična novost**
- ▶ enostavna implementacija
- ▶ razširitev na podatke z diskretno merjenim časom
- ▶ **nesporna interpretacija:** povprečje kvocientov $\frac{\widehat{S}_{0i}(t)}{\widehat{S}_{Pi}(t)}$
- ▶ **naravna razširitev:** $\frac{\widehat{S}_{0i}(t)}{\widehat{S}_{Pi}(t)}$ lahko uporabimo kot odvisno spremenljivko v GEE (*ang. generalized estimating equations*) modelih

Diskretno merjen čas - pristranskost

- ▶ simulirani podatki
- ▶ $n = 1000$
- ▶ 1000 ponovitev
- ▶ čas merjen na leto natančno
- cenilka PP ... 0.0171
- nova cenilka ... 0.0009
- Manjša pristranskost nove cenilke**

Ilustracija uporabe cenilk



- ▶ slovenske bolnice z rakom debelega črevesa, $n = 1538$
- ▶ diagnosticirane med januarjem 1994 in decembrom 2000
- ▶ stare med 26 in 85 let
- ▶ zvezno merjen čas

Priznanji za svoji predstavitvi sta prejela tudi Marjan Cugmas (njegova mentorja sta bila dr. Anuška Ferligoj in dr. Aleš Žiberna) in Damjan Manevski (mentorica dr. Maja Pohar Perme).

Utemeljitev komisije posterja Marjana Cugmasa:

Prispevek z uvedbo novega algoritma generiranja omrežij, ki upošteva vse ali le določene tipe triad, dokaže, da lahko določena globalna struktura omrežja nastane pod vplivom določenih mehanizmov, ki so neodvisni od lastnosti enot. Problem in rešitev sta podana pregledno, temeljito in jasno in tako vsakomur ponujata poučen vpogled v globalne strukture omrežij.

GENERIRANJE OMREŽIJ Z UPOŠTEVANJEM RAZLIČNIH TIPOV TRIAD

Marjan Cugmas
Fakulteta za družbene vede

Prispevek naslavlja vprašanje učinkovitosti generiranja omrežij z določeno globalno strukturo z upoštevanjem različnih tipov triad. Globalna struktura omrežja je definirana z različnimi tipi bločnih modelov, kjer je bločni model omrežje, v katerem enote predstavljajo skupine enakovrednih enot (glede na povezave z ostalimi enotami) preučevanega omrežja. Za generiranje omrežij je predstavljen in uporabljen algoritem prestavljanja povezav. To je iteracijski algoritem, kjer kot začetno omrežje predpostavimo slučajno omrežje z dano gostoto. Nato v vsaki iteraciji prestavimo eno povezavo tako, da se porazdelitev števila izbranih tipov triad čim bolj približa porazdelitvi izbranih tipov triad v omrežju z željeno globalno strukturo (idealnem omrežju brez napak).

Rezultati kažejo, da lahko, z upoštevanjem vseh tipov triad, generiramo omrežja z željeno bločno strukturo. Še več, omrežja z izbrano bločno strukturo je mogoče generirati tudi z upoštevanjem zgolj določenih tipov triad. Navedeno ne drži za primer hierarhičnega bločnega modela brez polnih blokov na diagonalni, kjer je potrebno upoštevati nekatere dodatne lokalne strukture, kot so poti dolžine tri. To privede do pomembnega spoznanja, da lahko določena globalna struktura omrežja nastane pod vplivom določenih mehanizmov, ki so neodvisni od lastnosti enot.

GENERIRANJE OMREŽIJ Z UPOŠTEVANJEM RAZLIČNIH TIPOV TRIAD

Marjan Cugmas (marjan.cugmas@fdv.uni-lj.si), Anuška Ferligoj (anuska.ferligoj@fdv.uni-lj.si), Aleš Žiberna (ales.ziberna@fdv.uni-lj.si)
 Fakulteta za družbene vede, Univerza v Ljubljani

GLOBALNE STRUKTURE OMREŽIJ

Globalne strukture omrežij so opisane z bločnimi modeli. Bločni model je omrežje, kjer enote predstavljajo skupine enakovrednih enot (glede na povezave z ostalimi enotami) preučevanega omrežja. Na slikah spodaj levo so matricno predstavljeni idejni tipi preučevanih bločnih modelov.

Kohезiven tip bločnega modela sestoji iz nekaj skupin enot. Enote znotraj skupin so povezane, med enotami iz različnih skupin pa ni povezav.

Simetričen center-periferija bločni model navadno sestoji iz dveh skupin. Ena skupina enot se imenuje center (to so enote, ki so med sabo vzajemno povezane), druga pa periferija (enote iz periferije med sabo niso povezane). Centralne enote so povezane z enotami iz periferije. Tudi enote iz periferije so povezane z enotami iz centra.

Asimetričen center-periferija bločni model je podoben simetričnemu center-periferija bločnemu modelu, le da v primeru asimetričnega obstajajo povezave le iz periferije k centralnim enotam (ali obratno).

Hierarhičen bločni model s praznimi bloki na diagonalni (oziroma hierarhično neopoznamiti skupin enot). Enote iz prve skupine so povezane z enotami iz druge skupine, enote iz druge skupine so povezane z enotami iz tretje skupine in tako naprej.

Hierarhičen bločni model s polnimi bloki na diagonalni je podoben hierarhičnemu bločnemu modelu s praznimi bloki na diagonalni, le da so enote znotraj skupin, v primeru prvega, med sabo močno povezane.

Tranzitivni bločni model s praznimi bloki na diagonalni je podoben hierarhičnemu modelu s praznimi bloki na diagonalni, le da v primeru tranzitivnega obstajajo povezave iz skupine iz višjega nivoja z enotami, ki pripadajo skupnam na vseh nižjih nivojih.

Tranzitivni tip bločni model s polnimi bloki na diagonalni je podoben tranzitivnemu bločnemu modelu s praznimi bloki na diagonalni, le da so enote znotraj skupin, v primeru prvega, med sabo močno povezane.

LOKALNE STRUKTURE OMREŽIJ

Lokalne strukture omrežja so opredeljene z različnimi tipi triad. Dokazano je, da je obstoje različnih tipov triad potreben in zadosten pogoj za obstoj nekaterih globalnih struktur omrežij, kjer globalne strukture niso opredeljene skozi različne tipe bločnih modelov, kot velja za pričakovanja.

Na sliki desno so predstavljeni vsi možni tipi triad.

TRIADE V BLOČNIH MODELIH

dovoljeni tipi triad za različne tipe bločnih modelov

tip triade	111	112	113	121	122	123	131	132	133	211	212	213	221	222	223	231	232	233	311	312	313	321	322	323	331	332	333	
kolheziven	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
asimetričen center-periferija	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
simetričen center-periferija	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
hierarhičen s praznimi bloki na diagonalni	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
hierarhičen s polnimi bloki na diagonalni	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
tranzitivni s praznimi bloki na diagonalni	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
tranzitivni s polnimi bloki na diagonalni	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*

Dovoljeni tipi triad za določen tip bločnega modela so tisti tipi triad, katerih število v omrežju z določenim tipom bločnega modela je večje od nič.

Prepovedani tipi triad za določen tip bločnega modela so tisti tipi triad, katerih število v omrežju z določenim tipom bločnega modela je enako nič.

RL ALGORITEM

Za generiranje omrežij z določeno globalno strukturo je uporabljen algoritem prestavljanja povezav (RL algoritem), ki je:

- determinističen,
- računsko zahteven,
- podvržen morebitnemu lokalnemu optimumu.

Prepostavka: Enote tedaj k vzpostavljanju takih povezav, ki rezultirajo v taki porazdelitvi izbranih tipov triad, kot bi jo pričakovali v primeru omrežja s predpostavljanim tipom bločnega modela brez napak (vedenje erozijsko je namensko).

Algoritem: Slučajno izbrana povezava je prestavljena: če to izboljša porazdelitev izbranih tipov triad, potem povezava ostane prestavljena, sicer se vrne na prvotno mesto. Algoritem je iterativen.

Desno je shematičen opis RL algoritma.

Z UPOŠTEVANJEM RAZLIČNIH TIPOV TRIAD JE MOGOČE GENERIRATI OMREŽJA Z VEČINO ANALIZIRANIH TIPOV BLOČNIH MODELOV

POSTOPEK PONOVNO 50-KRAT ZA VSAK TIP BLOČNEGA MODELJA IN ZA VSAKO MNOŽICO IZBRANIH TIPOV TRIAD

1. GENERIRAMO SLUČAJNO OMREŽJE

2. NA SLUČAJNEM OMREŽJU UPORABIMO RL ALGORITEM

3. NA SLUČAJNEM OMREŽJU IN NA IZBRANEM MNOŽICI IZBRANIH TIPOV TRIAD UPORABIMO NALOŽNO REKURZIVNO FUNKCIJO

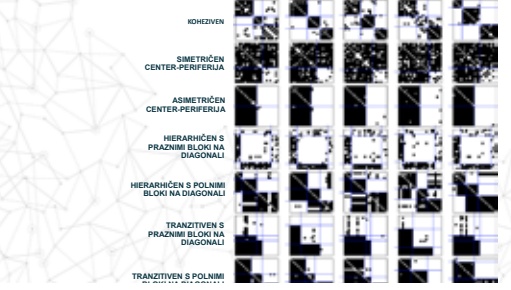
4. IZRAČUNAMO VREDNOSTI KRITERIJSKE FUNKCIJE

5. PRIMERJAMO GENERIRANA OMREŽJA

PORAZDELITVE VREDNOSTI IZBOLJŠANJA ZA RAZLIČNE TIPE BLOČNIH MODELOV IN RAZLIČNE MNOŽICE TIPOV TRIAD



NEKAJ PRIMEROV GENERIRANIH OMREŽIJ (UPOŠTEVANI SO VSI TIPI TRIAD)



GENERIRANJE HIERARHIČNEGA TIPIA BLOČNEGA MODELJA S PRAZNI MI BLOKI NA DIAGONALI

Vsa generirana omrežja imajo želeno bločno strukturo, ki je je mogoče enostavno vizualno prepoznati. Izbrana je omrežja, za katerega smo pričakovali hierarhičen tip bločnega modela s praznimi bloki na diagonalni.

Dobljena struktura se izboljša, če poleg izbranih dovoljenih tipov triad upoštevamo še pododbojne ti.

GLAVNE UGOTOVITVE

- Vsa generirana omrežja imajo željeno globalno strukturo, ki jo je mogoče enostavno vizualno prepoznati.
- Načeloma so globalne strukture najbližje pričakovanim, ko upoštevamo vse možne tipe triad. Upoštevanje zgolj prepovedanih tipov triad rezultira v generiranih omrežjih s pričakovanimi globalnimi strukturami z nekaj več napakami, kot v primeru upoštevanja vseh možnih tipov triad ali vseh dovoljenih tipov triad.
- Dobljeno globalno strukturo je mogoče izboljšati z upoštevanjem nekaterih dodatnih lokalnih struktur omrežja (npr. poti dolžine tri v primeru generiranja hierarhičnega tipa bločnega modela s

Utemeljitev komisije posterja Damjana Manevskega:

Prispevek na jasen in razumljiv način prikazuje pomembno izboljšavo pri računanju intervalov zaupanja za statistiko, ki se uporablja pri Mann-Whitney testu. Komisija je še posebej vesela, da lahko nagrado podeli mlademu statistiku, ki še obiskuje magistrski program Uporabna statistika.

INTERVALI ZAUPANJA ZA TEST MANN-WHITNEY

Damjan Manevski

Študent magistrskega programa Uporabna statistika

Test Mann-Whitney preučuje problem, ko je potrebno oceniti koliko se razlikujeta dve neodvisni slučajni spremenljivki X in Y , oziroma dva neodvisna vzorca. Ker test ne predpostavi porazdelitev obravnavanih vzorcev, se je Mann-Whitney od začetka izpostavil kot neparametrična alternativa testa t . Kot stopnja prekrivanja se pogosto uporabi $\theta = P(X < Y)$, ki je povezana s statistiko U in s krivuljo ROC.

V naši raziskavi smo se osredotočili na izračun intervalov zaupanja za θ . Naredili smo povzetek najbolj pogostih metod, ki se uporabljajo za izračun intervalov zaupanja. Zanimalo nas je kako se različne metode obnašajo, ko v simulacijah privzamemo bolj kompleksne porazdelitve, saj so se večina dozdajšnjih avtorjev nanašali predvsem na porazdelitve iz družine eksponentnih porazdelitev. V tej obsežni raziskavi smo našli primere, ko posamezne metode ne delujejo tako kot bi želeli, zato smo se poglobili v definicijah teh metod in poiskali razloge za njihovo neučinkovitost.

Ko smo enkrat bolje razumeli posamezne metode, smo predlagali dva koraka, ki sta izboljšala metodo DeLonga in s katerimi smo dobili metodo, ki nam je v testiranih primerih dala zadovoljive rezultate. Zato vidimo ta novodobljen pristop kot dobro alternativo (in v določenih primerih izboljšavo) pete Newcombeve metode, ki je sicer pogosto privzeta kot najboljšo metodo za izračun intervalov zaupanja za θ .

Intervali zaupanja za test Mann-Whitney

Damjan Manevski, izr. prof. Maja Pohar Perme

Univerza v Ljubljani, IBMI

Uvod in motivacija

Test Mann-Whitney predstavlja neparametrična alternativa testu t. Test ne predpostavlja porazdelitve populaciji in je njegova moč zelo podobna kot pri testu t. Kljub njegovim pogostim uporabi, se v praksi redko poroka pripadajoči interval zaupanja, poleg tega pa še zmeraj se ni izpostavil enoten način izračuna intervalov zaupanja. Zato smo se v naši raziskavi osredotočili na izračun teh intervalov.

Skozi leta so se pojavili različni pristopi računanja, vendar kot se je izkazalo v naših simulacijah, za vsako izmed pogosto uporabljanih metod lahko najdemo primere, ko metoda spodleti. Zato smo se pogljobili v vsaki metodi in poskusili razumeti razloge zakaj posamezne metode v določenih primerih ne delujejo toliko dobro. Na podlagi tega smo predlagali nekaj izboljšav, že obstoječih pristopov izračuna IZ, s katerimi lahko odpravimo te težave.

θ kot stopnja prekrivanja vzorcev

Imamo dva neodvisna vzorca:

- $X_i, i = 1, \dots, m$, kjer so $X_i \sim F$
- $Y_j, j = 1, \dots, n$, kjer so $Y_j \sim G$

Test Mann-Whitney preveri $H_0: F_X = F_Y$. V naši raziskavi smo kot meto obstojanja vzorcev uporabili $\theta = P(X < Y)$, saj obstaja bijektivna preslikava med θ in testno statistiko za Mann-Whitney. Poleg tega pa θ predstavlja AUROC (ang. *area under the ROC curve*), kar dodatno poveča njeno uporabo. Lahko vidimo, da ko H_0 drži, je $\theta = 0,5$. Zaradi bijektivnosti jo lahko uporabimo tudi v drugi smeri - če $\theta \neq 0,5$, je posredično H_0 kršena.

V praksi običajno je poznamo eksaktno vrednost θ . Zato uporabimo cenilko:

$$\hat{\theta} = \frac{1}{mm} \sum_{i=1}^m \sum_{j=1}^n I(X_i < Y_j) = \frac{1}{m} \sum_{i=1}^m V_{ij}$$

, kjer je $V_{ij} = I(X_i < Y_j)$, $V_{ij} \sim \text{Bern}(\theta)$.

Intervali zaupanja za θ

Najprej nas zanima eksaktno varianco $\hat{\theta}$, ki ima tako obliko.

$$\text{Var}(\hat{\theta}) = \frac{\theta(1-\theta)}{mm} + \frac{(n-1)}{mm} \text{Var} S_Y(X) + \frac{(m-1)}{mm} \text{Var} F_X(Y)$$

Obstaja več načinov kako oceniti to varianco [1],[2]. Tukaj se bomo osredotočili le na pristop, ki ga je predlagal DeLong [1]:

$$\widehat{\text{Var}}(\hat{\theta}) = \frac{1}{m} \text{Var} \hat{S}_Y(X) + \frac{1}{n} \text{Var} \hat{F}_X(Y)$$

Več o cenilkah \hat{S}_Y , \hat{F}_X lahko najdemo v [1].

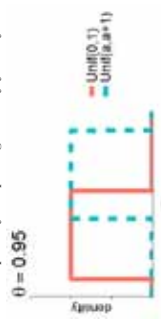
Cenilka $\widehat{\text{Var}}(\hat{\theta})$ se je v naših simulacijah izkazala kot najbolj zanesljiva za različne kombinacije porazdelitev vzorcev in različne velikosti vzorcev. Kot zadnji korak DeLong [1] naredi IZ na naslednji način:

$$\hat{\theta} \pm z_{1-\frac{\alpha}{2}} \cdot \sqrt{\widehat{\text{Var}}(\hat{\theta})} \quad (1)$$

Simulacije in nadaljnje izboljšave

Naredili smo simulacije za obsežen nabor možnih porazdelitev vzorcev. Pri tem je bil naš cilj poglobiti kako se obnaša predlagana metoda v primerjavi z ostalimi obstoječimi - npr. s peto Newcombova metoda [3], ki je pogosto predstavljena kot najbolj zanesljiva metoda.

V naši raziskavi smo vključili razl. kombinaciji porazdelitev, velikosti vzorcev in vrednosti θ , vendar tukaj bomo le predstavlili primere, kjer posamezne metode spodletijo ter predlagali nadaljnje izboljšave.



Slika 1: $X \sim \text{Urn}[0, 1]$, $Y \sim \text{Urn}[0, \alpha + 1]$, $m = n = 20$ in je α določen v obsežnosti od $\theta = 0, 95$.

1. izboljšava

Na Sliki 1 smo prikazali primer povzedelev, ko IZ (1) ne deluje, kot bi želeli, saj je verjetnost pokritja na podlagi simulacij **88%** (v naši simulaciji smo vedno generirali 95% IZ, oz. $\alpha = 0, 05$).

Kot izboljšavo uporabimo idejo, ki jo je predlagala Pepe [4] - najprej interval (1) ustrezno izračunamo na log skali in dobljene meje intervala zopet transformiramo na prvotni skali. Posledično se je verjetnost pokritja na tem primeru povečala na **96%**. S to izboljšavo se je verjetnost pokritja stabilizirala v večini problematičnih primerov.

2. izboljšava

Vendar ta nov pristop še zmeraj ni bil dovolj zanesljiv za vse testirane primere v simulacijah. Izkazalo se je, da metoda ne deluje toliko dobro za majhne velikosti vzorcev. Na Sliki 2 smo prikazali primer, za kateri je verjetnost pokritja **88%**.



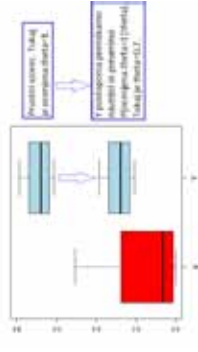
Slika 2: $X \sim N(0, 1)$, $Y \sim N(\mu, 1)$, $m = n = 5$, kjer je μ določen v obsežnosti od $\theta = 0, 8$.

Izkaže se, da je to posledica tega, da se dlogodrek $(\hat{\theta} = 1)$ zgodi prepuščeno. Namreč, če pogledamo (1), vidimo, da pri $(\hat{\theta} = 1)$ doblimo IZ oblike $[1, 1]$, torej meje IZ ne zajamejo pravo vrednost θ . To se tudi zgodi, če generiramo IZ na log skali (s pomočjo prve izboljšave). In ker je verjetnost tega dogodka dokaj visoka ($P(\hat{\theta} = 1 | \theta = 0, 8) = 0, 11$), metoda ima manjšo verjetnost pokritja za $m = n = 5$.

Zato moramo narediti popravek, ko je $\hat{\theta} = 1$. Začnimo z idejo, da mora 95% IZ vsebovati vse vrednosti na $[0, 1]$, kjer je $P(\hat{\theta} = 1 | \theta) > 0, 05$. Da bi poiskali meje tega IZ, fiksiramo vrednosti prvega vzorca, medtem ko vrednosti drugega vzorca postopoma zmanjšujemo, s čimer se tudi $P(\hat{\theta} = 1 | \theta)$ zmanjšuje (verjetnost ocenimo z *bootstrap* vzorčenjem).

2. izboljšava - nadaljevanje

Ko enkrat postane ta ocenjena verjetnost manjša kot 0,05, prognoziramo zadnjo vrednost θ za spodnjo mejo IZ (zgornja meja pa privzamemo, da je enaka 1). Na Sliki 3 smo grafično pokazali ta postopek.



Slika 3: Prikaz postopka za 2. izboljšavo.

Ta pristop se je izkazal za najbolj zanesljiv za vse primere, ki smo jih obravnavali. Če se pa spet vrnemo na primer z Slike 2 cite, s predlaganim popravkom doblimo verjetnost pokritja **95,6%**.

Zaključek

V naši raziskavi smo poiskali primere, ko že obstoječe metode izračuna IZ za θ ne dajo zadostna podkritja. S pomočjo predlaganih popravkov doblimo metodo, ki je v večini primerov zelo stabilna.

Poleg tega, našli smo primere, ko peta, Newcombova metoda [3] (ki je najbolj pogosto uporabljena) ne deluje toliko dobro, a žal tokrat jih nisimo predstavili. Na podlagi rezultatov, lahko vidimo naš predlagani pristop kot primerno alternativo, ali izboljšavo pete Newcombove metode.

References

- [1] DeLong, E., DeLong, D., Clarke-Pearson, D. (1988). Comparing the Areas under Two or More Correlated Receiver Operating Characteristic Curves: A Nonparametric Approach. *Biometrics*, 44(3), 837-851.
- [2] Hsieh, J. A. and McKay, B. J. (1982). The meaning and use of the area under the receiver operating characteristic (ROC) curve. *Biometrics*, 143:29-38.
- [3] Newcombe, R. G. (2006). Confidence Intervals for an effect size measure based on the Mann-Whitney statistics. Part 2: asymptotic methods and evaluation. *Statistics in Medicine* 25:559-573.
- [4] Pepe, M. S. (2003). *The statistical evaluation of medical tests for classification and prediction*. Oxford University Press, Oxford.

OB 70-LETNICI PROF. DR. ANUŠKE FERLIGOJ

Dr. Janez Stare
Inštitut za biostatistiko in medicinsko informatiko

Saj ni res! Pa je!

Ko me je Nuša lani septembra povabila na kosilo v Piran, se pravega razloga za vabilo nisem zavedal. In ko mi je Andrej povedal, za kaj gre, je bila moja reakcija nekako v stilu zgornjega prvega stavka, Andrejeva pa potem v stilu drugega.

Njene točne starosti nisem pravzaprav nikoli natančno poznal. Vedel sem, da je nekaj starejša od mene in to mi je zadoščalo. Človek jo gleda in ona je vedno ista in si pač misliš: Dokler sem mlajši od nje, sem kar v redu.

Sedemdesetletnica tega ne spreminja. No, morda pa le toliko, ker je pač Nušina, da se mi zdaj ta številka ne zdi več prav velika. A vendar, spodobi se, da Nuši ob njej posvetimo nekaj besed. Ker komu pa bomo, če ne njej.

Ne zdi se mi potrebno, da bi tukaj navajal njene biografske podatke, tudi ne njenih funkcij, nagrad in priznanj, publikacij, bivanj v tujini in kar je še takšnega. O tem morda ob njeni 80-letnici, bolj sigurno pa ob 90-letnici, ko bo tudi med statistiki že kakšen premlad, da bi jo poznal. Za tokrat naj vse to naštevanje povzamem v stavku, ki sem ga nekoč že zapisal: Nuša Ferligoj je pionir raziskovalne statistike v Sloveniji. In naj dodam, da ves čas tudi njena najbolj ključna figura.

V nadaljevanju pa o Anuški Ferligoj kot jo vidimo njeni statistični/metodološki kolegi. Predvsem bi izpostavil štiri značilnosti:

Prvič: *radovednost*. VSE jo zanima. Interes za vse, kar se tako ali drugače dotika našega raziskovalnega področja, je po mojem temeljni pogoj za uspešnost v raziskovanju. Nuši, se mi zdi, je bilo to naravno dano, drugi pa se lahko od nje učimo.

Drugič: *pronikljivost*. Vedno nekako vidi bistvo problema. Tu ne mislim samo na stroko, ampak recimo na vprašanja/probleme, o katerih razpravljamo na raznih sejah. Razpravljamo in razpravljamo, ta misli tako, oni drugače, vsi pa nekako čakamo, kaj bo rekla Nuša. In če začne stavek z Ja, kaj pa jaz vem ..., vemo, da se z dosedaj povedanim ne strinja in potem čakamo na stavek Jaz bi rekla ..., da zremo kako in kaj naj naredimo.

Tretjič: *dobronamernost*. Nuša je ena tistih oseb, pri kateri se meni ljuba Konfucijeva misel Ne dovoli zla v svojih mislih! nekako naravno udejanja. Človek seveda ne more imeti vseh rad, tudi strinjati se ni mogoče vedno in povsod z vsemi, in tu in tam je treba koga tudi kritizirati, a pri Nuši tudi vedno čutimo, da je kritika dobronamerna.

Četrtič: *povezovanost*. O tem sem pisal v predlogu za podelitev Blejčevega priznanja in ker sem tako lepo povedal, naj tukaj kar ponovim: brez nje bi bili slovenski statistiki, se bojim, samo naokrog razmetani strokovnjaki, eni boljši, eni slabši, ki bi komaj vedeli drug za drugega in bi samozadovoljno solili pamet strokovnjakom z različnih področij znanosti, ki, na srečo, težko sodijo o naši kompetentnosti. Z njo pa smo ekipa pedagogov in raziskovalcev, ki je povsem primerljiva s statistiki v svetu in ki počasi postavlja slovensko statistiko tudi doma na mesto, ki ji v raziskovalnem svetu gre.

In za konec: zadnje čase nekaj grozi z upokojitvijo. Saj po svoje bi ji jo privoščil, ker vem, da ji tudi v pokoju ne bo nikoli dolgčas, ampak zaenkrat bi ji vseeno odsvetoval. Slovenska statistika jo še vedno potrebuje.



Prof. dr. Anuška Ferligoj

DOMAČI IN MEDNARODNI DOGODKI

KONFERENCA

UPORABNA STATISTIKA – APPLIED STATISTICS 2017

Vsako leto ponosno poročamo o uspešni izvedbi konference *Uporabna statistika / Applied statistics* in tudi tokrat ni nič drugače. Štirinajsta konferenca AS je potekala od 24. do 27. septembra 2017 v hotelu Brdo. Tudi tokrat je ponudila bogat program in tudi tokrat smo se trudili, da smo tematsko zajeli čim več področij statistične stroke.

V nadaljevanju objavljamo enega od zanimivih prispevkov z naslovom *Uvajanje e-poročanja v kratkoročne poslovne statistike v Sloveniji*, ki ga je udeležencem konference predstavila avtorica Nina Češek Vozel s Statističnega urada RS.

Vsi drugi povzetki predstavitev v angleškem jeziku so na voljo tudi na elektronski povezavi <http://conferences.nib.si/AS2017/AS2017-Abstracts.pdf>.

Naj vas hkrati že zdaj povabimo, da se nam pridružite tudi na konferenci AS 2018. Ta bo predvidoma od 23. do 26. septembra 2018. Več pa na povezavi <http://conferences.nib.si/AS2018/>.

UVAJANJE E-POROČANJA NA KRATKOROČNIH POSLOVNIH STATISTIKAH V SLOVENIJI

Nina Češek Vozel
Statistični urad RS

Zaradi učinkovite rabe virov je treba v čim večji meri posodobiti in racionalizirati statistične procese. Modernizacija zbiranja podatkov je eden ključnih elementov. V prispevku je prikazano praktično izvajanje in izid uvajanja spletnih vprašalnikov v kratkoročnih poslovnih statistikah v Sloveniji. Slednje so namreč zelo pomembne, saj omogočajo zgodnje odkrivanje sprememb v gospodarskem razvoju. Pravočasnost in kakovost podatkov sta pomembna.

SURS se je v letu 2013 srečal z izzivom uvajanja spletnih vprašalnikov. Začetna pomanjkljivost je postala izziv in z leti se je tehnologija izboljšala. Postal je bolj prijazen uporabnikom, pri vsaki prenovi vprašalnika oziroma raziskovanja za e-poročanje pa še vedno ostajajo izzivi, saj ima vsako raziskovanje kakšno posebnost.

Veliko začetnih prizadevanj za zamenjavo tiskanih vprašalnikov z elektronskimi se je odrazilo v rezultatih, kot so zagotavljanje hitrega odziva pri opominjanju podjetij, manjše število napak zaradi dodatnih kontrol podatkov ter nižji stroški tako za SURS kot tudi za poročevalske enote.

Po drugi strani smo se soočili z drugimi vrstami napak. V kontekstu uvajanja e-poročanja smo skušali postaviti bolj standardiziran statistični proces za podobna raziskovanja.

Čeprav večina podjetij pošilja svoje podatke v elektronski obliki, majhen odstotek še vedno uporablja tiskane vprašalnike. Njihovo zmanjšanje ostaja naš cilj v prihodnosti. Da bi zagotovili dobro reprezentativnost objavljenih podatkov, moramo namreč upoštevati vsa podjetja v vzorcu, ker je slovensko gospodarstvo relativno majhno. Ker slovenska vlada že dlje časa spodbuja elektronsko komunikacijo s podjetji, pričakujemo, da bo več podjetij pripravljeno za prehod na elektronsko sporočanje podatkov.

STATISTIČNI DAN 2018

MOČ PODATKOV

Irena Svetin
Statistični urad RS

Podatki nas obdajajo vsepovsod, vsi jih tudi uporabljamo. Pravilno uporabljeni so lahko učinkovito orodje pri sprejemanju odločitev, ki lahko vplivajo na naša življenja. Pomembno vlogo pri prenosu informacij, podkrepjenih s podatki, imajo mediji; čedalje pomembnejši so družbeni mediji. Še več, s podatki lahko oblikujemo javno mnenje. Skratka, moč podatkov je velika. V poplavi podatkov se večkrat vprašamo, ali so zgodbe, ki jih gledamo, o katerih beremo ali poslušamo, resnične ali lažne. Čedalje več je razprav o verodostojnosti in dojetanju podatkov, o pravilni interpretaciji, uporabni vrednosti, avtorstvu, etičnosti uporabe, o tem, kako ločiti dejstva od fikcije, kakšen je sploh potencial uporabe podatkov. To so bila izhodišča za 27. statistično posvetovanje z naslovom Moč podatkov, ki se je odvijalo 23. januarja 2018 v Kongresnem centru Brdo pri Kranju.

Vse to je v uvodnem govoru omenjala tudi generalna direktorica SURS-a Genovefa Ružič. Poleg tega je opozorila tudi na veliko količino podatkov, ki se nenehno povečuje in tudi nenehno uporablja. Ampak kako naj vemo, katerim podatkom lahko verjamemo, se je vprašala. In dalje: Ali so podatki SURS-a bolj verodostojni od drugih? Kako drugi gledajo na SURS-ove podatke? Kakšna je vloga medijev pri objavljanju SURS-ovih podatkov?

Ali je anketna metodologija sploh še potrebna, se je v uvodnem nagovoru spraševal Matevž Bren, predsednik Statističnega društva.



Foto: STA

O vlogi uradne statistike je govoril tudi Martin Bajželj, na SURS-u zadolžen za stike z uporabniki. Poudaril pa je naslednje: podatki nimajo nobene vrednosti, če niso uporabljeni, pa naj bodo še tako kakovostni. Kako torej ustvariti zanimanje za uradno statistiko? Kako postati viden, kako privabiti pozornost uporabnika in kako jo tudi ohraniti? Podatki uradne statistike temeljijo na preverjenih metodah, vsi procesi so popisani, metodologija vseh raziskovanj je javno objavljena.

Nekaj odgovorov na ta vprašanja je ponudil Slavko Jerič, novinar in kolumnist MMC RTV Slovenija. Povedal je, kako na številčne podatke gledajo novinarji, in opozoril na vprašanje glede pravilnosti podatkov. Z omembo fake news, lažnih novic, je dal iztočnico Mihi Artnaku, umetniku, ki je zaslovel z lažno novico o prodaji slike. Ta je opisal svoja projekta, s katerima je nastavil ogledalo družbi, ki nekritično sprejema informacije.

Odgovore na vprašanja, zastavljena v prvem delu posvetovanja, so skupaj z vsemi navzočimi iskali udeleženci okrogle mize v drugem delu srečanja: Gal Kočar, specialist za digitalni marketing, Denis Oštir, urednik 24ur na POP TV, Igor E. Bergant, voditelj Odmevov na RTV Slovenija, mag. Karmen Hren, namestnica generalne direktorice slovenskega statističnega urada, in dr. Valentin Bucik, psiholog in profesor.



Foto: Arhiv Andreje Jernejčič

Vsi govorniki so se strinjali, da je moč podatkov zelo velika, še večja pa da je lahko moč napačnih podatkov. Strinjali so se tudi, da so popravki napačnih podatkov v večini primerov smiselni in nujni, da pa skoraj nikoli ne dosežejo tako velikega števila uporabnikov kot napačni podatki. Denis Oštir je spomnil na znani rek Winstona Churchila: Laž je že na pol poti okrog sveta, še preden si resnica natakne hlače. Dejstvo je, da imamo za procesiranje na voljo čedalje manj časa, informacij, ki jih je treba sprocesirati, pa čedalje več. Dr. Bucik je na to problematiko pogledal s psihološkega zornega kota: odločamo se na podlagi najhitrejše bližnjice, in to zato, da bi porabili čim manj miselne energije in časa. Socialna resničnost (dejstva, stališča, mnenja, predsodki, stereotipi) vpliva na naše odločitve; odločamo se na podlagi nepopolnih informacij, zato lahko hitro zapademo v stereotipe. Tudi zato ni pomembna samo pravilnost podatkov, pomembno vlogo ima tudi pravilna interpretacija podatkov. To je izziv tako za medije kot tudi za Statistični urad. Ta mora zato – to se razume – še naprej skrbeti za verodostojnost podatkov. Pa ne samo to, prizadevati si mora tudi za čim višjo raven znanja uporabnikov. Poseben izziv pri vsem tem so mladi, saj informacij ne pridobivajo po tradicionalnih kanalih (iz časopisa, televizije, radia). Zavedati se je treba, da mlade lahko dosežemo na njim bližji način. Torej je treba za širjenje informacij, če naj te dosežejo tudi čim več mladih, uporabiti različne informacijske poti (kanale).

Skratka, #PodatkiSoZakon. Tak je bil tudi slogan posvetovanja. Vendar bo to veljalo le, če se bomo zavedali, kolikšno moč imajo podatki. Da jo imajo, smo se po koncu posvetovanja verjetno vsi strinjali.

Zanimiv prispevek o tej temi kot odmev na posvetovanje smo lahko prebrali v Delovi prilogi Svet kapitala 26. januarja 2018. Poleg tega so na SURSovi spletni strani (www.stat.si) objavljene fotografije in video posnetki s posvetovanja.

STROKOVNI SEMINAR

KOMBINIRANI NAČINI ZBIRANJA PODATKOV

Nataša Bučar in Mateja Zgonc
Statistični urad RS

SURS je 11. oktobra 2017 v sodelovanju s Fakulteto za družbene vede (FDV) in Nacionalnim inštitutom za javno zdravje (NIJZ) pripravil že tradicionalni strokovni seminar z naslovom Kombinirani načini zbiranja podatkov. Tokrat smo s strokovnjaki iz akademskega okolja in s predstavnikom drugih pooblaščenih izvajalcev dejavnosti državne statistike izmenjavali izkušnje in ideje v zvezi s kombiniranimi načini zbiranja podatkov pri osebah in gospodinjskih v okviru statističnih raziskovanj.

Seminar je bil sestavljen iz dveh vsebinskih sklopov. V prvem smo obravnavali vplive in učinke kombiniranega načina zbiranja podatkov, v drugem, vsebinsko širšem, pa druge metodološke izzive, s katerimi se spopadamo v praksi. Vsak sklop je bil sestavljen iz treh predstavitev. Vsaka udeležena ustanova, FDV, NIJZ, SURS, je v vsakem sklopu prispevala po eno predstavitev. Vsaki predstavitvi je sledila razprava.

Zbrane je najprej pozdravila Tatjana Novak, vodja sektorja za procese in komuniciranje na SURS.

Moderatorica prvega sklopa z naslovom Vplivi in učinki kombiniranega načina anketiranja je bila Andreja Smukavec, vodja oddelka za splošno metodologijo na SURS.

Prvi govorec, Tomaž Burnik s FDV, je predstavil učinke načina anketiranja v kombiniranih anketah na primeru SURS-ovega raziskovanja Mnenje potrošnikov. Predstavil je ocene neto učinkov med metodo pomožnih spremenljivk (angl. backdoor method) in metodo primerjalnega vzorca. Ena glavnih ugotovitev je, da je pri vseh spremenljivkah neto učinek med načinoma pozitiven. V praksi to pomeni, da spletni respondenti zaradi načina zbiranja podatkov podajajo bolj pesimistične odgovore, kot bi jih, če bi odgovarjali po telefonu. S tem so se dodatno potrdila tudi SURS-ova dognanja, da so odgovori spletnih respondentov pri tem raziskovanju bolj negativni. V razpravi je bilo opozorjeno, da metodi za izračun vrednosti uporabljata povprečja, medtem ko se v uradni statistiki uporabljajo ravnotežja. Ti dve metodi sta razviti za povprečja, saj se v tem primeru ocene porazdeljujejo normalno.

Sledila je predstavitev Tine Zupanič z NIJZ. Predstavila je vpliv kombiniranega načina anketiranja – primerjava CINDI 2012 in 2016 (raziskava Z zdravjem povezan življenjski slog). V razpravi je bilo predlagano, da bi bilo dobro prihranek, ki ga dobimo s spletnim zbiranjem podatkov, porabiti za nagrajevanje respondentov. Predlagano je bilo tudi, da bi bilo glede na rezultate smiselno razmisliti tudi o tem, da bi se poštni vprašalniki ohranili le za del ciljne populacije (npr. za starejše od 50 let). Iz rezultatov te raziskave je razvidno, da respondenti odgovarjajo iskreneje prek spleta (več jih npr. prizna depresijo). Predvidevamo, da zato, ker se počutijo neidentificirane. Izkušnje kažejo, da manj izobraženi in tisti v nižjih dohodkovnih razredih na splošno odgovarjajo na vprašalnike v manjšem številu. Pri kombinaciji poštnega in spletnega načina zbiranja podatkov se pojavljajo tudi dvojniki; ista oseba odgovori na spletni in tudi na poštni vprašalnik. Na NIJZ nameravajo (glede na trenutne rezultate) tudi v prihodnje ohraniti kombinacijo poštnega in spletnega načina zbiranja podatkov.

Eva Belak s SURS je predstavila kombinirani način zbiranja podatkov v Anketi o izobraževanju odraslih (AIO); podatki so se zbirali prek spleta, po telefonu in na terenu. Predstavila je analizo odgovorov in stroške zbiranja glede na način zbiranja podatkov. V razpravi je bilo poudarjeno, da smo na SURS pri izvajanju terenskih raziskovanj stroškovno zelo učinkoviti. Za to so v veliki meri zaslužni stalni, že dobro izobraženi anketarji, ki sodelujejo tudi v drugih raziskovanjih. Kontinuiteto dela anketarjev je treba ohranjati tudi v prihodnje.

Moderatorica drugega sklopa z naslovom Metodološki izzivi kombiniranega načina anketiranja je bila Tina Steenvoorden, vodja oddelka za standarde na SURS.

V drugem delu seminarja je prva nastopila Gaja Zager Kocjan z NIJZ. Predstavila je kognitivno testiranje Ankete o tobaku, alkoholu in drugih drogah za različne načine anketiranja. Pri omenjeni raziskavi so izvedli kognitivno testiranje vprašalnika v dveh korakih. Kognitivnemu testiranju v obliki osebnega anketiranja je sledilo še spletno kognitivno testiranje. Na NIJZ se zavedajo pomanjkljivosti obeh oblik testiranj. Iz izsledkov kognitivnih testiranj je razvidno, da s spletnim kognitivnim testiranjem pridobimo manj respondentovih odgovorov na metavprašanja. Pri osebnem kognitivnem testiranju družbeno občutljivih vprašanj pa se dopušča možnost, da bodo testni respondenti podajali socialno zaželene odgovore. Vprašanja udeležencev strokovnega seminarja so bila vezana predvsem na specifične izvedbe kognitivnega testiranja (število metavprašanj, tipi metavprašanj, literatura, po kateri so se zgledovali pri izvedbi testiranja) in na nagrajevanje pri testiranju sodelujočih respondentov (npr. kaj je bilo darilo). Nekatere ugotovitve iz tega testiranja bodo lahko prenesli tudi na druga raziskovanja s podobno občutljivo tematiko.

Katja Rutar s SURS je predstavila analizo izključitev članov skupinskih gospodinjstev na primeru raziskovanja Aktivno in neaktivno prebivalstvo. Analizirala je podatke o članih skupinskih gospodinjstev. V razpravi je bil podan predlog za analizo primerjave aktivnosti, v starostni skupini od 20 do 24 let, med prebivalci skupinskih in zasebnih gospodinjstev. V tej starostni skupini je namreč delež oseb, ki živijo v skupinskih gospodinjstvih, visok. Gre za osebe, ki bivajo v študentskih domovih. Analiza je smiselna, saj se v SURS-ovih raziskovanjih uteži izračunavajo le glede na prebivalce zasebnih gospodinjstev.

V zadnji predstavitvi sta Nejc Berzelak in May Doušak predstavila spletni panel CRONOS za Evropsko družboslovno raziskavo. Predstavila sta namen spletnega panela in analizo preliminarnih podatkov za Evropsko družboslovno raziskavo. V razpravi so se pojavili pomisleki glede plačevanja respondentov. Vsi so se strinjali, da je razmislek o tem glede na prakse drugih držav smiseln.

Tatjana Novak je z vprašanji o prihodnosti na področju anketiranja in formatu prihodnjega srečanja povabila k panelni razpravi sopobudnika strokovnih seminarjev Vasjo Vehovarja s FDV in Metko Zaletel z NIJZ. Strinjali so se, da je metodološka delavnica s temeljito razpravo pomembna in koristna. Smiselno bi jo bilo razširiti tudi na druga metodološka vprašanja.

Dosedanje sodelovanje se je izkazalo za uspešno, zato nameravamo take seminarje pripravljati tudi v prihodnosti. To je zelo dobra priložnost, da predstavimo delo in izzive na SURS ter izmenjamo izkušnje z drugimi institucijami. Ugotovili smo, da bodo naši izzivi in izzivi drugih institucij tudi v prihodnosti podobni. Naslednji posvet bo v letu 2018.

22. SREČANJE MLADIH STATISTIKOV

KAJ RAZISKUJEJO DOKTORSKI ŠTUDENTI STATISTIKE?

Ddr. Joca Zurc

V letu 2017 je potekalo že 22. srečanje mladih statistikov, ki so se ga tradicionalno udeležili podiplomski študenti ter mladi doktorji znanosti statistike s svojimi mentorji iz Slovenije, Hrvaške, Avstrije, Italije in Madžarske. Srečanje je potekalo od petka, 13. do nedelje, 15. oktobra 2017, v odlični organizaciji Hrvaškega statističnega društva, Medicinske fakultete in »Andrija Štampar« Fakultete za javno zdravje Univerze v Zagrebu ter s podporo Statističnega urada Hrvaške in Hrvaškega združenja za biometriko. S svojimi predstavitvami je sodelovalo 15 izbranih mladih statistikov, po trije iz vsake države udeleženke. Študentske predstavnike vsake delegacije sta spremljala najmanj po dva eksperta. Prvi dan so se udeleženci zbrali na lokaciji konference v čudovitem ambientu Škole narodnog zdravlja »Andrija Štampar«, ki je vodilna znanstveno-izobraževalna ustanova na področju javnega zdravja na Hrvaškem, ustanovljena že leta 1927 in poimenovana po njenem ustanovitelju dr. Andriju Štampar, priznanem hrvaškem zdravniku socialne medicine, rektorju Univerze v Zagrebu, predsedniku Jugoslovanske akademije znanosti in umetnosti ter ekspertu Svetovne zdravstvene organizacije pri Društvu narodov, predhodnice današnje OZN. V čudovitih prostorih navedene ustanove je potekal vsebinski program srečanja kot tudi bivanje udeležencev. Hišna restavracija pa je poskrbela za odlično kulinariko s hrvaškimi specialitetami ter obveznimi domačimi štrudli.

Program srečanja se je pričel v petek, 13. oktobra 2017, z registracijo udeležencev in sestankom mentorjev. Slavnostna otvoritev pa je potekala v soboto, 14. oktobra 2017, s pričetkom ob 9. uri, ko se je pričel prvi dan študentskih predstavitev. Udeležence sta nagovorila glavna organizatorica dogodka, red. prof. dr. Ksenija Dumičić iz Oddelka za statistiko, Ekonomske fakultete Univerze v Zagrebu, ter prof. dr. Mirjana Kujundžić Tiljak, direktorica Škole narodnog zdravlja »Andrija Štampar«. Po slavnostnih nagovorih so se pričele predstavitve mladih statistikov, ki so bile razdeljene v štiri sekcije v soboto in eno sekcijo v nedeljo. V vsaki sekciji so bile po tri predstavitve, praviloma zastopane s predstavniki treh različnih držav, pri čemer je bilo za vsako predstavitev skupaj z razpravo namenjenih 30 minut.

Vsebinski program je odprla Avstrijka Eva Benková s predstavitvijo razvoja inovativnega modela za določitev optimalnega raziskovalnega načrta eksperimentalnih raziskav, pri katerem jo je zlasti zanimalo zagotavljanje ustrezne porazdelitve meritev skozi časovni okvir izvedbe raziskave. Mladi statistik Bálint Nagy iz Madžarske je nadaljeval s predstavitvijo robustnega ocenjevanja povprečnega učinka zdravljenj v opazovalnih raziskavah z uporabo analize občutljivosti. Za njim pa je sledil prvi slovenski nastop. Ana Slavec, ki je ena izmed prvih doktorandk prenovljenega bolonjskega Interdisciplinarnega doktorskega študijskega programa Statistika (pričel z izvedbo v študijskem letu 2009/2010), je predstavila rezultate svoje doktorske disertacije na temo odkrivanja neznanih besed v anketnih vprašalnikih s pomočjo predtestiranja vprašalnikov in jezikovnih virov. Na osnovi izvedene raziskave na treh študijah primerov, je avtorica razvila postopek predtestiranja vprašalnikov, s katerim je možno razkriti neznane besede pri anketirancih in predlagati izboljšave v besedilih anketnih vprašanj. Ana Slavec je doktorsko disertacijo z naslovom Izboljševanje ubeseditve anketnih vprašanj z jezikovnimi viri pripravila pod mentorstvom rednega profesorja dr. Vasja Vehovarja in jo zagovarjala 17. junija 2016 na Fakulteti za družbene vede Univerze v Ljubljani.

Drugo sekcijo je pričela hrvaška predstavica Ivana Lobić s področja ekonomske statistike, in sicer jo je na primeru velike recesije zanimalo ocenjevanje zaznavanja in pričakovanj inflacije v anketnih raziskavah s potrošniki. Simona Jokubauskaite iz Avstrije je v nadaljevanju predstavila uporabnost skupnega ocenjevanja diskretne izbire in sistema opredelitev časovnih odhodkov na primeru prihranka časa potovanja, kar daje pomembne usmeritve za prevozno infrastrukturo. Njena rojakinja, Sarah Karasek, pa se je posvetila modelu razvrščanja v skupine za namen ocenjevanja kakovosti slik v proizvodni polprevodnikov - električnih ali fotonjskih vezij. V tretji sekciji smo najprej spremljali prvo italijansko predstavico, Tereso Barrella, ki je predstavila GeCUB modele za subjektivno evalvacijo funkcionalne septoplastike po operaciji nosnih dihalnih poti. Sledila je druga slovenska predstavitev, ki jo je zastopala gostujoča raziskovalka Daria Maltseva. Perspektivna mlada raziskovalka iz Ruske Federacije je zaposlena na National Research University Higher School of Economics v Moskvi, kjer deluje kot namestnica vodje v Mednarodnem laboratoriju za aplikativne raziskave socialnih omrežij. Od 1. marca do 29. maja 2017 je gostovala na Centru za metodologijo in informatiko na Fakulteti za družbene vede Univerze v Ljubljani, kjer je v sodelovanju z zaslužnim profesorjem dr. Vladimirjem Batageljem raziskovala strukturo socialnih omrežij raziskovalcev na primeru objavljenih člankov v ruskih znanstvenih revijah, vključenih v elektronsko bazo eLibrary.ru, ki vsebuje več kot 3.800 znanstvenih periodičnih publikacij v ruskem jeziku. Predstavljeni rezultati raziskave so bili na srečanju mladih statistikov deležni velike pozornosti in plodovite

razprave. Tretjo predstavitev v tej sekciji pa je izvedel György Márk Kis iz Madžarske, ki se je posvetil analizi socialnih omrežij v igri ameriškega nogometa z namenom iskanja ključnih igralcev igre na primeru obrambe ameriškega moštva Minnesota Vikings. V zadnji, četrti sekciji prvega dne predstavitev, je najprej nastopila Iva Budimir iz Splita na Hrvaškem, ki se je predstavila s temo učinki postopkov normalizacije na statistično sklepanje v genomskih, metabolnih in sorodnih raziskavah. Sledili sta predstavitvi italijanskih predstavnic. Daniele Bottigliengo se v svojem statističnem raziskovanju posveča učinkovitosti t.i. hadronoterapije, ki predstavlja nov pristop v radioloških obravnavah pri zdravljenju raka. S pomočjo Bayesovega pristopa razvija novo metodo ocenjevanja relativne biološke učinkovitosti na podlagi in-vitro ter kliničnih podatkov pacientov z rakom prostate. Veronica Sciannameo pa je predstavila uporabo Bayesovega pristopa pri reševanju statističnih problemov na področju raziskav v delovnem okolju. Te raziskave se srečujejo s problemom multiplih primerjav, ki povečujejo napačne pozitivne rezultate ter s problemom pomanjkanja podatkov zaradi majhnega števila predstavnikov različnih poklicnih profesij.

Zadnji dan statističnega srečanja je potekala peta sekcija, ki jo je pričela slovenska predstavnica Joca Zurc, s predstavitvijo ugotovitev svoje doktorske disertacije, v kateri je razvijala temeljne kriterije ocenjevanja veljavnosti v raziskavah z integracijo kvantitativne in kvalitativne metodologije [angl. mixed methods]. Joca Zurc je bila najbolj »sveža« doktorica statističnih znanosti na srečanju, saj je disertacijo na Interdisciplinarnem doktorskem študijskem programu Statistika pod mentorstvom redne profesorice dr. Anuške Ferligoj uspešno zagovarjala tri dni pred srečanjem, 10. oktobra 2017. V nadaljevanju sekcije je Madžar Laszlo Németh predstavil statistično teoretsko raziskovanje porazdelitve odvisnosti cenilk repnega indeksa. Vsebinski program 22. srečanja mladih statistikov pa je sklenil domači predstavnik, Ivan Papić, z raziskavo o koreliranih kontinuiranih časih naključnih korakov in Pearsonovih difuzijah.

22. srečanje mladih statistikov - YSM 22 se je zaključilo z razpravo med udeleženci v sproščenem vzdušju ob kavi in domačih štrudlih, kjer so se organizatorji še enkrat izkazali in potrdili sloves odličnega gostitelja. Vse čestitke in zahvala za brežhibno organizacijo, navdušujočo koordinacijo in izstopajočo kakovost izvedbe letošnjega srečanja gredo profesorici dr. Kseniji Dumičić, glavni gonilni sili YSM 22, ki je v sodelovanju s sodelavci Škole narodnog zdravlja »Andrija Štampar« omogočila vsem prisotnim nepozabno srečanje in čudovito bivanje v Zagrebu. YSM 22 je potrdil svoje poslanstvo in pomen za znanstveno-raziskovalni razvoj mladih statistikov ter pokazal na izjemno širino in inovativnost aktualnih raziskav na področju statistične znanosti v državah območja Alpe Adria.

Naslednje, 23. srečanje mladih statistikov, bo potekalo v organizaciji madžarskih predstavnikov, v mesecu oktobru 2018, v Budimpešti.



Udeleženci 22. srečanja mladih statistikov – YSM 22 pred Šolo narodnega zdravja »Andrija Štampar« v Zagrebu

DELAVNICA O SIMBOLNI ANALIZI PODATKOV SDA 2017

Dr. Simona Korenjak-Černe
Ekonomska fakulteta Univerze v Ljubljani

V času od 12. do 14. junija 2017 je Ekonomska fakulteta v soorganizaciji s Statističnim društvom Slovenije in Inštitutom za matematiko, fiziko in mehaniko gostila mednarodno delavnico SDA 2017 (<http://vldowiki.fmf.uni-lj.si/doku.php?id=sda:meet:lj17>).

Delavnica je namenjena izmenjavi raziskovalnih rezultatov in idej na področju simbolne analize podatkov, ki je del širšega področja analize podatkov in med drugim ponuja tudi orodja za obdelavo velikih podatkovij (big data). Ukvarja se z razvojem metod analize podatkov in njihovo uporabo, ki temeljijo na kompleksnejših predstavitevah podatkov. Poseben primer pri tem predstavljajo npr. agregirani podatki, ki jih namesto s povprečno vrednostjo predstavimo s porazdelitvijo. S tem ohranimo več informacije. Za analizo tako dobljenih zgoščenih podatkov pa so potrebne nove metode. Področje simbolne analize združuje raziskovalce številnih »klasičnih« disciplin, npr. matematike, statistike, računalništva, ekonomije, strojništva in industrije ter drugih, kjer se srečujejo s sodobnimi izzivi analiz podatkov.

Delavnica v Ljubljani je bila že šesta po vrsti. Pretekle delavnice so bile v

1. Avstriji (SDA'2009, October 18-20, 2009, Wienerwaldhof, Austria),
2. Belgiji (SDA'2011, June 7-9, 2011, Namur, Belgium <https://www.info.fundp.ac.be/SDA11/>),
3. Španiji (SDA'2012, November 7-9, 2012, Madrid, Spain <http://eprints.ucm.es/17547/1/SDA2012.pdf>),
4. na Tajvanu (SDA'2014, June 14-16, 2014, Academia Sinica, Taipei, Taiwan <http://www3.stat.sinica.edu.tw/sda2014/>) in
5. v Franciji (SDA'2015, November 17-19, 2015, Orléans, France <http://www.univ-orleans.fr/mapmo/colloques/sda2015/>).

Na vseh smo s predstavitvami svojega dela na tem področju sodelovali tudi slovenski predstavniki.

Na uradnem delu otvoritve je udeležence pozdravila prodekanja za raziskovalno delo na Ekonomski fakulteti prof. Vesna Žabkar, nato se je v času delavnice odvijalo osem sekcij s predstavitvami sodobnih rezultatov. Tokratne delavnice se je udeležilo 29 raziskovalcev iz 11 držav (Avstralije, Belgije, Francije, Italije, Japonske, Kanade, Kostarike, Portugalske, Slovenije, Španije in Združenih držav Amerike). Med udeleženci sta bila tudi pobudnik tega pristopa analize podatkov prof. Edwin Diday iz Francije (Emeritus profesor, Dauphine Université Paris) in soavtorica temeljne knjige s tega področja Symbolic Data Analysis: Conceptual Statistics and Data Mining prof. Lynne Billard (University of Georgia, ZDA), ki je znana poleg svojih raziskovalnih dosežkov tudi po aktivnostih na področju znanosti nasploh, vloge žensk ter je bila nenazadnje tudi predsednica dveh največjih statističnih organizacij International Biometric Society (1994 - 1995) in American Statistical Association (1995 - 1996), ter tudi prof. Paula Brito s Portugalske (Faculty of Economics, University of Porto, v mandatu 2011-2013 predsednica IASC - International Association for Statistical Computing) in prof. Scott Sisson iz Avstralije (UNSW - University of New South Wales, Sydney, tudi predsednik Statistical Society of Australia).

Slovenski raziskovalci smo na delavnici sodelovali s petimi prispevki:

- Scientific collaboration: Comparing self evaluation and administrative data source (avtorja Luka Kronegger in Anuška Ferligoj),
- Panel data estimation in regressions for symbolic data: An application to the clustering of cultural entrepreneurial regimes (avtorja Andrej Srakar in Marilene Vecco),
- Clustering European countries regarding causes of death (avtorji Aleša Lotrič Dolinar, Jože Sambt in Simona Korenjak-Černe),
- Clustering of symbolic data with relational constraint: Demographic sex-age structures in US and Europe (avtorji Nataša Kejžar, Simona Korenjak-Černe in Vladimir Batagelj) in
- Symbolic network analysis of bike sharing systems (avtor Vladimir Batagelj).

Poleg raziskovalcev sta se delavnice udeležila tudi Manca Golmajer in Črt Grahonja s Statističnega urada Republike Slovenije.

Tako organizacijsko kot finančno je bila delavnica uspešno izvedena, kar so ob odhodu potrdili udeleženci s pohvalami in pozitivno finančno stanje po izvedbi.

RAZVRŠČANJE SIMBOLNIH PODATKOV Z RELACIJSKO OMEJITVIJO: DEMOGRAFSKE STRUKTURE PO SPOLU IN STAROSTI PO EVROPI

CLUSTERING OF SYMBOLIC DATA WITH RELATIONAL CONSTRAINT: DEMOGRAPHIC SEX-AGE STRUCTURES IN US AND EUROPE

Nataša Kejžar¹, Simona Korenjak-Černe², Vladimir Batagelj³

1 Uvod

Razvrščanje je opisna statistična metoda, ki omogoča odkrivanje skupin v podatkih. Običajno pri razvrščanju iščemo take skupine, da so si enote znotraj skupin bolj podobne kot pa enote iz različnih skupin. Podobnost para enot v tem prispevku definiramo s pomočjo različnosti med njima in uporabimo hierarhično razvrščanje z združevanjem (Ward 1963, angl. agglomerative hierarchical clustering), ki sta ga Ferligoj in Batagelj prilagodila upoštevajoč relacijske omejitve (Ferligoj in Batagelj, 1982, 1983).

V tem delu pokažemo na primeru evropskih teritorialnih enot za statistiko (NUTS), da hierarhično (in z omejitvami) lahko razvrščamo tudi bolj kompleksne enote (npr. enote, ki so opisane s simbolnimi opisi, v našem primeru s populacijskimi piramidami). Zanima nas, katera povezana področja v Evropi bi lahko na podlagi podatkov o demografskih strukturi po starosti in spolu tvorila dokaj homogene regije.

V prispevku najprej predstavimo opis populacijskih piramid kot simbolnih objektov. Nato pokažemo, kako je potrebno v našem primeru spremeniti osnovno metodo hierarhičnega razvrščanja, da upoštevamo relacijsko omejitev in komentiramo uporabljene različnosti (med enotami in med skupinami). Opišemo rezultate na enotah NUTS in povzamemo glavne točke analize.

2 Simbolni opisi enot

Razvrstiti želimo evropske teritorialne enote NUTS, in sicer glede na njihove demografske strukture po spolu in starosti v letu 2013. Tako strukturo pogosto grafično prikažemo s populacijsko piramido, in sicer kot dva histograma (za vsak spol) ekvidistančnih starostnih razredov.

Na teritorialno enoto lahko gledamo kot na enoto drugega reda (angl. second level unit), ki je agregirana iz posameznikov (enot prvega reda). Vrednosti dveh spremenljivk (v tem primeru spola in starosti) sta pri vsakem posamezniku enostavni. V agregirani enoti ju lahko zapišemo bodisi spet z enostavno vrednostjo (npr. s srednjo vrednostjo starosti in deležem spola) ali pa

¹ Univerza v Ljubljani of Ljubljana, Medicinska fakulteta, IBMI; natasa.kejzar@mf.uni-lj.si

² Univerza v Ljubljani, Ekonomska fakulteta

³ IMFM Ljubljana in Univerza na Primorskem, IAM, Koper

tako, da ohranimo več informacije. Lahko je to za vsako enoto porazdelitev po starosti glede na spol ali pa obratno. Take (kompleksno predstavljene) enote so predmet področja analize simbolnih podatkov (Billard in Diday, 2006) in jih imenujemo simbolni objekti (SO).

V našem primeru je torej vsak SO $X = (V_m, V_z)$ opisan s spremenljivkama V_m in V_z (za moške in ženske) in vsaka od njiju je predstavljena z vektorjem vrednosti $f_{x_i} = [f_{x_i1}, f_{x_i2}, \dots, f_{x_i k_i}]$ kjer je V_i spremenljivka s k_i vrednostmi k_j , $i \in \{m, z\}$, in $f_{x_{ij}}$ frekvenca njene kategorije j . Naj velja $n_{x_i} = \sum_j f_{x_{ij}}$. Potem lahko dobimo pripadajočo opaženo porazdelitveno funkcijo kot

$$p_{x_i} = \frac{1}{n_{x_i}} f_{x_i}.$$

Na podlagi opaženih porazdelitvenih funkcij zgradimo matriko različnosti za vsak par SO-jev. Različnost, ki jo uporabljamo v tem prispevku, je predstavljena v članku Batagelja in sod. (2015), in sicer kot posplošitev kvadratne evklidske razdalje. Za izračun uporabimo opaženo porazdelitveno funkcijo p_X in velikost n_X (število osnovnih enot):

$$d(u, v) = \frac{n_u \cdot n_v}{n_u + n_v} \cdot (p_u - p_v)^2,$$

kjer sta p_u in p_v komponenti spremenljivke i enot X_u in X_v . Da dobimo dokončno različnost med enotama $d(X_u, X_v)$, moramo vse različnosti med komponentami sešteti. Če velikosti n_X ne upoštevamo (torej $n_X = 1$), dobimo navadno kvadratno evklidsko razdaljo. V tem primeru torej izračunavamo relativne različnosti med populacijskimi piramidami. Pri razvrščanju to pomeni, da bomo gledali, kako podobne/različne so enote glede na oblike populacijskih piramid (tj. prehajanje med demografskimi modeli - demographic transition model - DTM, Kirk (1996)). Z upoštevanjem uteži pa bodo bolj obljudene enote bolj vplivale na skupni opis skupine, v katero so dodeljene. Izbira n_X za vrednost uteži omogoča, da je naravni skupni predstavnik skupine kar populacijska piramida celotnega področja, ki mu enote znotraj skupine pripadajo.

V splošnem bi za razvrščanje lahko uporabili katero koli smiselno različnost, ki jo izračunamo za vsak par enot. Matrika različnosti je namreč edini vhodni podatek za metode, ki temeljijo na Lance-Williamsovem obrazcu. Mednje sodi tudi Wardov postopek razvrščanja.

3 Hierarhično razvrščanje z relacijsko omejitvijo z združevanjem

Osnovna ideja razvrščanja z relacijsko omejitvijo je, da omejimo nabor vseh možnih razvrstitev \mathcal{C} , ki v primeru hierarhičnega razvrščanja tvorijo polno drevesno razvrstitev. V našem primeru je omejitev simetrična in refleksivna relacija sosednosti R ($R(X_u, X_v) \equiv$ teritorialni enoti X_u in X_v imata skupno mejo). Relacijo R lahko predstavimo z neusmerjenim grafom. Vsaka

skupina iz dopustne razvrstitve določa v grafu relacije povezan podgraf. V našem primeru jo torej sestavljajo geografsko povezane "regije" teritorialnih enot. Skupini lahko torej združimo le, če v grafu relacije obstaja povezava, ki ima krajišči v obeh skupinah. Relacijo R lahko tako razširimo na skupine:

$$R(C_p, C_q) \Leftrightarrow \exists X_u \in C_p \wedge \exists X_v \in C_q : R(X_u, X_v).$$

Algoritem hierarhičnega razvrščanja z relacijsko omejitvijo z združevanjem je podan v Tabeli 1. Dodatna omejitev je označena rdeče. C_k označuje razbitje (angl. partition) na k -tem koraku, vse enote pa so zbrane v množici \mathbf{U} . V opisu algoritma uporabljamo še množico sosedov $R(U) = \{V : R(U, V)\}$.

Tabela 1: Algoritem hierarhičnega razvrščanja z relacijsko omejitvijo z združevanjem

```

 $C_0 = \{\{X\}; X \in \mathbf{U}\}; h_0 = 0$ 
for  $k:=1$  to  $|\mathbf{U}|$  do
  if  $R = \emptyset$  exit
   $r = (u, v) := \operatorname{argmin}_{(p,q) : C_p, C_q \in C_{k-1} \wedge p \neq q \wedge R(C_p, C_q)} D(C_p, C_q)$ 
   $C_r := C_u \cup C_v$ 
   $C_k := C_{k-1} \setminus \{C_u, C_v\} \cup \{C_r\}$ 
   $h_k := D(C_u, C_v)$ 
   $R(C_r) := R(C_r) \cup R(C_u) \cup R(C_v) \setminus \{C_u, C_v\}$ 
  for  $s : s \neq r \wedge C_s \in C_k \wedge R(C_r, C_s)$  do
     $R(C_s) := \{C_r\} \cup R(C_s) \setminus \{C_u, C_v\}$ 
    update  $D(C_r, C_s)$ 
  endfor
endfor

```

Algoritem je vključen v program Pajek za analizo velikih omrežij (Batagelj in Mrvar, 2017). Ukaz

Network > Create Hierarchy > Clustering with Relational Constraint > Run

kot vhodni podatek potrebuje omrežje sosednosti, ki je uteženo z različnostmi med enotami. Če omrežje sestavlja več delov (komponent), algoritem združi enote v vsakem delu posebej.

$h_k = D(C_u, C_v)$ določa višino vozlišča v drevesu združevanja – dendrogramu. Zaželeno je, da v drevesu združevanja ni obratov. Ti nastanejo, če je različnost združene skupine manjša od različnosti kake od skupin, ki ju združujemo. Če ima različnost lastnost (Bruynooghe, 1977)

$$D(C_p, C_q) \leq D(C_p, C) \wedge D(C_p, C_q) \leq D(C_q, C) \Rightarrow D(C_p, C_q) \leq D(C_p \cup C_q, C),$$

potem obratov ni. Zlahka se prepričamo, da ta lastnost velja za minimalno, maksimalno in povprečno različnost med skupinami; torej za osnovne tri različnosti, ki jih uporabljamo pri hierarhičnem razvrščanju z združevanjem. V tej analizi smo izbrali maksimalno različnost (različnost med novo nastalo skupino $C_p \cup C_q$ in ostalimi skupinami C $D(C_p \cup C_q, C)$) je izračunana kot $\max(D(C_p, C), D(C_q, C))$. Začnemo s skupinami velikosti 1 – torej z različnostmi med posameznimi enotami, $D(X, Y) = d(X, Y)$.

4 Geografska sosednost

Graf sosednosti teritorialnih enot je bil ustvarjen iz datoteke Shape (© EuroGeographics for the administrative boundaries) in je povezan šele, ko ročno dodamo čezmorske povezave. Omrežje iz tega grafa, kjer je debelina povezav sorazmerna z različnostjo med enotami, je prikazano na Sliki 1 in predstavlja vhodno informacijo za razvrščanje v Pajku.

Slika 1: Omrežje sosednosti evropskih teritorialnih enot. Čezmorske povezave so prikazane v rdečem, različnosti brez uteži (brez velikosti populacije) označuje debelina povezave.

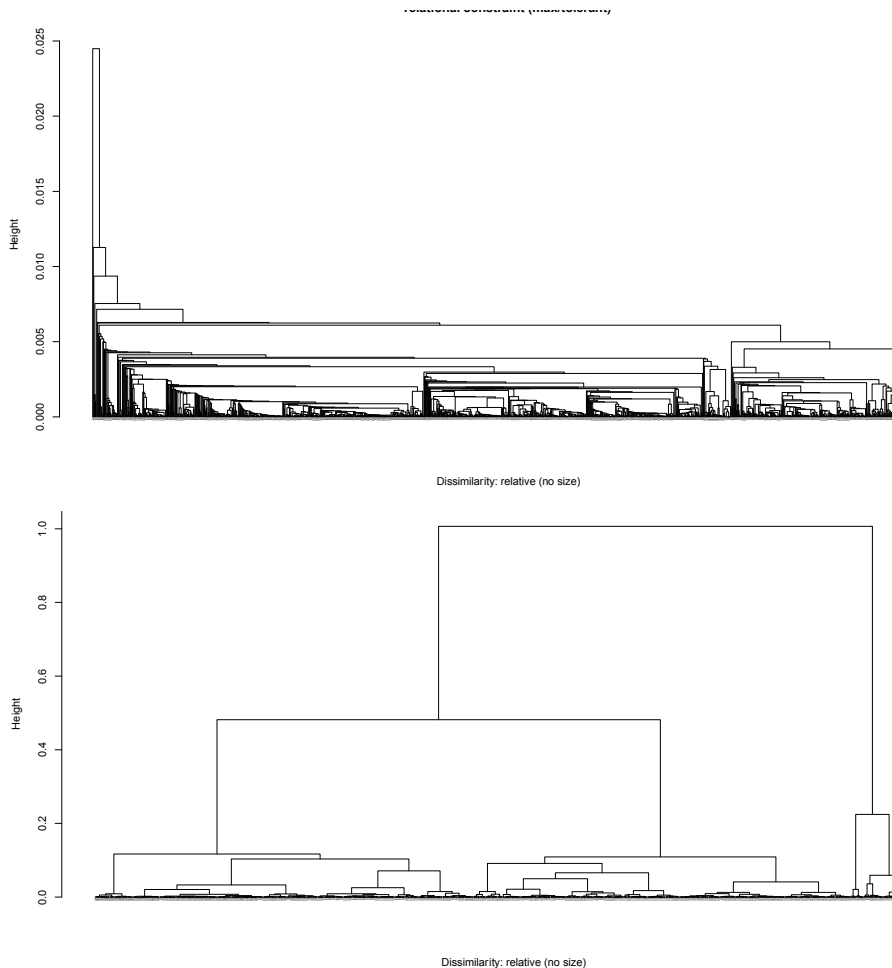


5 Rezultati

Razvrščanje enot NUTS 2013 je bilo narejeno na več načinov. Matrika različnosti je bila izračunana (1) brez upoštevanja velikosti (relativno) in (2) z upoštevanjem velikosti skupin (uteži). Iz vsake izmed matrik različnosti smo dobili svojo različico razvrstitve z omejitvami, ki smo jo primerjali še z razvrstitvijo brez omejitev. V nadaljevanju je bolj obširno predstavljena le različica brez upoštevanja uteži, v zaključkih pa sta omenjeni obe.

Za razvrščanje z omejitvami je bila uporabljena strategija max-tolerant (Batagelj in Ferligoj, 1982, 1983). Iz dreves te razvrstitve (RzO) in razvrstitve brez omejitev (RbO) (Slika 2) vidimo, da pri omejitvah dobimo pogostejše veriženje enot (dodajanje obrobnih enot), kar je pričakovano.

Slika 2: Drevesi razvrstitev RzO z omejitvijo (zgoraj) in brez omejitev RbO (spodaj). Višina v drevesu je sorazmerna z različnostjo med skupinami.



Razvrstitve prikažemo tudi na zemljevidu Evrope. Razvrstitev v 4 skupine pri RbO smo izbrali zaradi večje raznovrstnosti, čeprav bi bila po kriteriju Calinskija in Harabasza (1974) v razvrstitvi brez omejitev najboljša razvrstitev v le dve skupini. Pri RzO smo določili 7 skupin z vsaj 5 enotami, ostale pa smo zbrali v skupini nerazvrščenih. Na Sliki 3 in 4 sta prikazani obe razvrstitvi, prav tako pa tudi oblike populacijskih piramid za vsako skupino. Oblika populacijskih piramid predstavlja povprečno obliko strukture prebivalstva po spolu in starosti vseh teritorialnih enot v skupini. Vsaka teritorialna enota torej prispeva enako k celotni skupini.

Vidimo lahko, da z obema metodama odkrijemo zelo izrazito skupino - skupino 4 v RbO in skupino 6 v RzO - na daljnem vzhodu Turčije z zelo trikotno obliko piramide (tj. veliko rodnostjo in veliko smrtnostjo prebivalstva). Ker RzO dovoli združevanje samo sosednjih enot, lahko identificiramo nekaj ostalih podobnih področij: skupino 4 (NJR Makedonija in del Bolgarije), skupino 5 (Islandija in celoten irski otok z velikim deležem mladih ljudi) in skupino 7 (Škotska in severni del Anglije (Yorkshire)) z zelo staro starostno strukturo glede na DTM (majhno rodnostjo in majhno smrtnostjo). Zanimivo je Skandinavija povezana prek baltskih držav z Vzhodno Evropo in Turčijo (skupina 3). Oblika populacijske piramide te skupine je mlajša (po DTM) kot tista iz skupine 1 (Južna Evropa) ali iz skupine 2 (Zahodna Evropa).

Če za razvrščanje uporabimo metodo z omejitvijo in za izračun različnosti upoštevamo tudi uteži, je končni rezultat podoben RzO, veliko razliko pa opazimo v Skandinaviji, baltskih državah in severnih vzhodnoevropskih državah (do Madžarske), ki se v tej novi razvrstitvi v 7 skupin z vsaj 5 enotami priključijo skupini zahodnoevropskih držav. Ostale vzhodnoevropske države z vključno zahodnim delom Turčije, se priključijo skupini Južne Evrope. To vidimo tudi iz kontingenčne tabele (Tabela 2), kjer so skupine od 1 do 4 v razvrstitvi RzO precej premešane glede na razvrstitev, kjer uporabimo uteži.

Tabela 2: Kontingenčna tabela razvrstitev z omejitvijo. Vrstice predstavljajo rezultat z relativno različnostjo (RzO), stolpci pa razvrstitev v 7 skupin z vsaj 5 enotami z različnostjo med enotami z utežmi.

	Ostali	1	2	3	4	5	6	7
Ostali	18	13	19	5	0	0	0	0
1	0	443	51	0	0	0	0	0
2	5	98	429	0	0	0	0	0
3	0	99	177	41	0	2	0	0
4	0	19	0	0	0	0	0	0
5	0	0	0	0	15	0	0	0
6	2	0	0	0	0	12	0	0
7	0	0	0	0	0	0	26	6

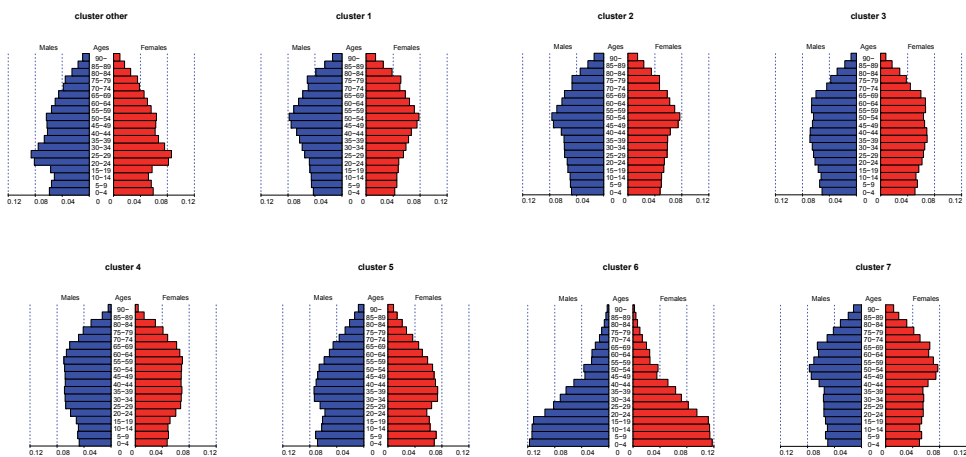
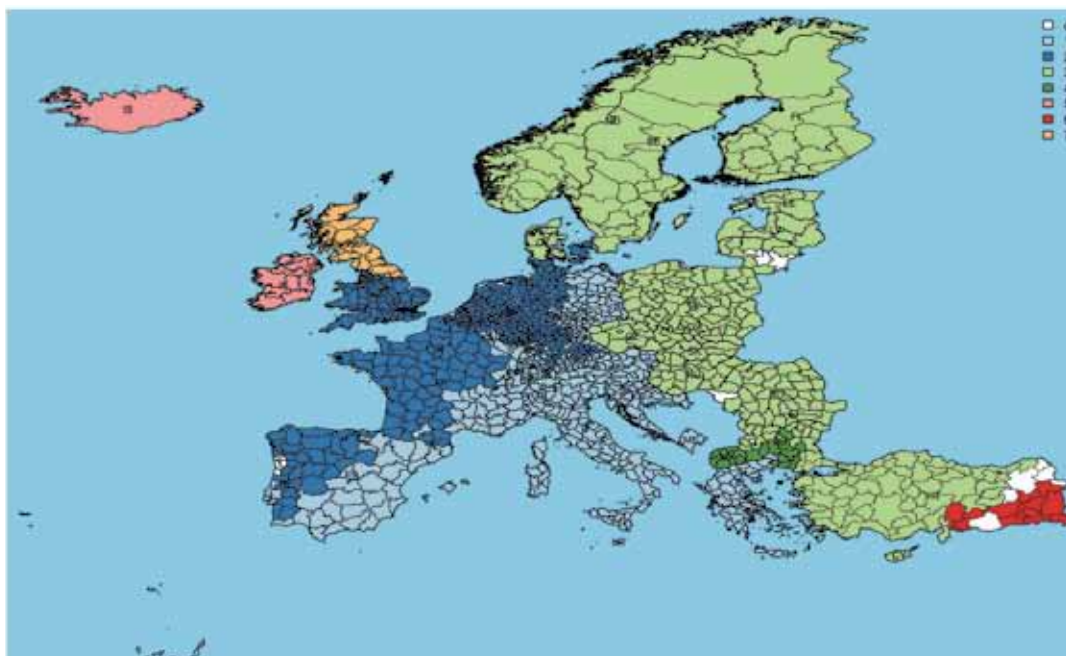
Zanimivo je pogledati populacijsko piramido nerazvrščenih teritorialnih enot (RzO). Tu se nahajajo območja, ki so preveč različna sosedom, da bi bila vključena v skupino. Opazimo lahko, da gre za območja z velikim deležem ljudi med 20. in 35. letom starosti, kar sovпада z mlajšim delom delovno aktivnega prebivalstva. Ti se radi priseljujejo v območja z veliko delovnimi mesti, v mestna in primestna središča. Ob podrobnejšem pregledu lahko vidimo, da so to teritorialne enote na področju velikih mest (npr. London, Berlin, Sofija ipd.).

6 Razprava

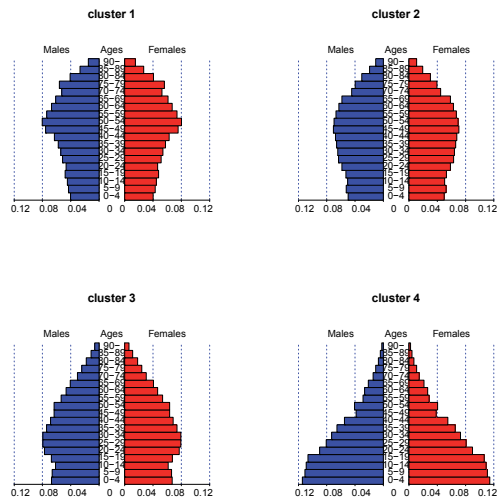
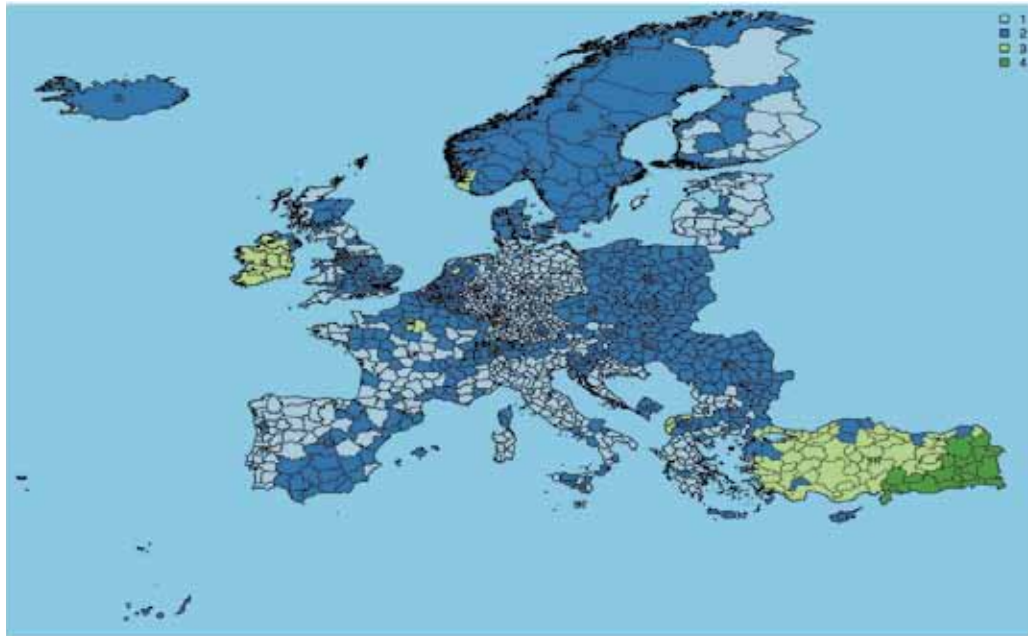
V prispevku smo prikazali uporabo hierarhične metode z združevanjem (z omejitvijo) na simbolnih objektih. Uporabljena metoda razvrščanja kot vhodni podatek potrebuje le matriko različnosti, zato lahko razvrščamo tudi agregirane enote, kjer vrednosti spremenljivk niso nujno skalarji. V primeru NUTS različnosti med enotami izračunamo s pomočjo posplošene kvadratne evklidske razdalje. Če razvrščamo simbolne objekte brez omejitev, so predstavniki skupin namreč intuitivno razložljivi. Gledamo podobnost oblik (brez uteži) ali pa dejanskih populacijskih piramid (z utežmi), Korenjak-Černe in sod., 2015. Hierarhično razvrščanje z omejitvijo je v Pajku narejeno za osnovne različnosti med skupinami, ki v našem primeru niso usklajene z izbranimi različnostmi med enotami. To ni optimalno, zagotavlja pa drevesa razvrstitev brez obratov. Kaj prinese uporaba usklajenih različnosti je zanimivo vprašanje za nadaljnje raziskave.

Razvrstitev z relacijsko omejitvijo brez upoštevanja velikosti populacije na primeru NUTS pokaže tri velika področja s podobno starostno strukturo po spolu: (1) Zahodna Evropa, (2) Jugovzhodna Evropa in (3) Vzhodna Evropa s Skandinavijo. V rezultatih razvrstitve z upoštevanjem uteži sta dobro vidni le dve veliki področji: (1) Zahodna in (2) Južna z Vzhodno Evropo. To kaže, da je vpliv uteži na razvrstitev lahko precejšen in bi ga bilo smiselno raziskati in ovrednotiti v primeru optimalnega razvrščanja; tj. ob uporabi usklajene različnosti med enotami in skupinami.

Slika 3: Rezultati razvrstitve z omejitvijo v 7 skupin po vsaj 5 enot. Enote, ki ne spadajo v nobeno skupino, so označene z belo.



Slika 4: Rezultati razvrstitve brez omejitev, razdelitev v 4 skupine.



Literatura

Batagelj V., Kejžar N., Korenjak-Černe S. (2015): Clustering of Modal Valued Symbolic Data. *arXiv:1507.06683* [stat.ML].

Billard L., Diday E. (2006): *Symbolic Data Analysis: Conceptual Statistics and Data Mining*, JohnWiley, Chichester.

Bruynooghe M. (1977): *Méthodes nouvelles en classification automatique des données taxinomiques nombreuses*, *Statistique et Analyse des Données* 3:24–42.

Calinski T., Harabasz J. (1974): A dendrite method for cluster analysis. *Communications in Statistics - Theory and Methods* 3(1):1-27.

© EuroGeographics for the administrative boundaries:

<http://ec.europa.eu/eurostat/web/gisco/geodata/reference-data/administrative-units-statistical-units>.

Ferligoj A., Batagelj V. (1982): Clustering with relational constraint, *Psychometrika*, 47(4):413-426.

Ferligoj A., Batagelj V. (1983): Some Types of Clustering with Relational Constraints, *Psychometrika*, 48(4):541-552.

Kirk, D. (1996): Demographic transition theory, *Population Studies* 50(3): 361–387.

Korenjak-Černe S., Kejžar N., Batagelj V. (2015): A weighted clustering of population pyramids for the world's countries, 1996, 2001, 2006, *Population Studies* 9(1).

Mrvar A., Batagelj V. (2017): Pajek - Program for Large Network Analysis 5.01. <http://mrvar.fdv.uni-lj.si/pajek/>.

Ward J. (1963): Hierarchical grouping to optimize an objective function, *Journal of the American statistical association*, 58(301):236–244.

KONFERENCA ICOS2017 V SARAJEVU

Ema Mišić
Statistični urad RS

Ekonomška fakulteta (EF) in Federalni statistični zavod BIH sta skupaj organizirala 1. mednarodno konferenco o uradni statistiki ICOS2017: Challenges, Opportunities and Future Directions in Official Statistics, ki je potekala 30. in 31. marca 2017 v Sarajevu. Glavni namen konference je bil povezati strokovnjake iz akademskih krogov in uradne statistike iz Jugovzhodne Evrope in sveta preko predstavitev rezultatov statističnih analiz, izmenjave dobrih praks in izkušenj ter razprav, da bi pripomogli k večji prepoznavnosti in uporabi uradnih statističnih podatkov za različne namene.



Po uvodnih nagovorih organizatorjev, dekana EF in direktorja federalnega zavoda za statistiko BIH, sta udeležence konference pozdravila tudi ambasadorica ZDA v BIH M. Cormack in profesor na londonski fakulteti za ekonomijo dr. N. Bingham.

Konferenca je v dopoldanskih sekcijah postregla z zelo zanimivimi, aktualnimi vabljenimi prispevki. V okviru prve teme **»Podatki kot zaveznik pri pripravi politik in razvojnemu načrtovanju«** so bili predstavljeni trije prispevki, in sicer je bila v prvi predstavitvi predstavljena organiziranost uradne statistike v BIH preko treh statističnih inštitucij in zakonskih podlag za njihovo delovanje ter centralne banke, ki prav tako pokriva pripravo nekaterih statistik ter na drugi strani s tem povezan sistem izobraževanja za pridobitev ustrezne izobrazbe in znanj za delovanje na statističnem področju. V drugi predstavitvi je bilo govora o povečevanju statistične pismenosti kot univerzalnem jeziku in načinu za boljše delovanje in prepoznavanje uporabnosti podatkov. V tretji predstavitvi pa so bili predstavljeni kazalniki trajnostnega razvoja (SDG) in, kakšno vlogo ima/bo imel pri tem statistični sistem.

V okviru druge teme **»Uradna statistika v dobi velikih podatkov«** so bile prav tako predstavljene tri zadeve, v okviru katerih je bilo govora o splošnih značilnostih te nove vrste podatkov, kako je z zaupanjem v njihovo uporabo, zakonske omejitve. Predstavniki italijanskega statističnega urada (ISTAT) je tako predstavil tudi prve poskuse uporabe »Big data« v praksi, in sicer konkretno pri statističnem raziskovanju IKT za podjetja.

Popoldanski del prvega dne konference je potem potekal vzporedno v dveh sekcija, od katerih je bilo v eni govora o različnih poteh/načinih analiziranja podatkov, v drugi pa o modernizaciji uradne statistike. V okviru slednje je imel SURS kar aktivno vlogo, predstavil je modernizacijo poslovnih statistik⁴ in »selektivni editing«⁵ kot nov način urejanja podatkov na našem uradu. Oba prispevka sta bila zelo dobro sprejeta in sta naletela tudi na precej zanimanja. V okviru prve sekcije pa sta predstavnici Ekonomske fakultete Univerze v Ljubljani predstavili prednosti pristopa simbolične analize podatkov v uradni statistiki⁶.

Tudi drugi dan je bil jutranji del namenjen vabljenim predavateljem in razdeljen v dva vsebinska sklopa. V okviru prve teme **»Odprta vprašanja ekonomskih in finančnih statistik«** je naprej predstavnica Eurostata predstavila postopke spremljanja makroekonomskih neravnovesij po državah, kateri mehanizmi se za to uporabljajo, zakaj je to pomembno ter, kje so vse te informacije in podatki na voljo. V drugi predstavitvi je predstavnik Svetovne banke govoril o uporabi podatkov na način, da se z njimi pove čim bolj prepričljivo zgodbo in se jim tako da dodano vrednost. Prav slednje pa naj bi bila tudi ena od nalog uradne statistike. Na Svetovni banki so za te potrebe vzpostavili tudi posebno bazo podatkov in razvili orodje za razmeroma hiter in enostaven prikaz zelenih podatkov, ki pa jim je na koncu potrebno dodati še zanimivo zgodbo. V tretji predstavitvi pa je predstavnik IMF predstavil razvoj finančnih statistik v državah zahodnega Balkana preko predstavitve dela v okviru večletnega projekta.

Pod drugo temo **»Odprta vprašanja pri administrativnih in socialnih statistikah«** pa so bile novosti in aktivnosti podane s strani dveh mednarodnih organizaciji. Predstavnik IMPAQ je govoril o uporabi administrativnih podatkov ter o stroškovni in kratkoročni oceni ter pripravi osnovnih evidenc na primeru BIH, s strani predstavnika Svetovne banke pa so bile podane nove usmeritve in rešitve pri merjenju blaginje v državah zahodnega Balkana.

Popoldanski del konference je prav tako potekal v dveh vzporednih sekcijah. V eni od njih je imel ponovno aktivno vlogo tudi SURS, in sicer so bili predstavljeni ekonomski računi za kmetijstvo⁷.

Konferenca je postregla z veliko novih, uporabnih informacij o razvoju, novostih in tudi prihodnjih načrtih na različnih statističnih področjih kot tudi v okviru različnih mednarodnih organizacij. Zagotovo pa takšni dogodki zelo veliko pripomorejo k večji prepoznavnosti uradne statistike in uporabnosti statistični podatkov.

⁴ **Emma Mišić**, Statistical Office of the Republic of Slovenia, Modernization of Business Statistics: Challenges and Outcomes of the Slovenian Statistical Office.

⁵ **Rudi Seljak, Nejc Jevšnik, Jerneja Pikelj**, Statistical Office of the Republic of Slovenia, Selective Editing – New Approaches at the Statistical Office of the Republic of Slovenia.

⁶ **Aleša Lotrič – Dolinar, Jože Sambt, Simona Korenjak – Černe**, Faculty of Economics, University of Ljubljana, Advantages of Symbolic Data Analysis Approach in Official Statistics.

⁷ **Irena Žaucer**, Statistical Office of the Republic of Slovenia, Economic Accounts as a Tool of Agricultural Policy.



Foto: Vanja Čerimagić

LETNO SREČANJE SKUPINE V7

Dr. Matevž Bren
Predsednik SdS

V petek, 13. oktobra smo se v Bratislavi sestali predstavniki skupine V7 – statističnih združenj držav srednje Evrope, ki jo sestavljamo združenja Avstrije, Češke, Madžarske, Poljske, Romunije, Slovaške in Slovenije. Srečanje je tradicionalno in se iz leta v leto seli po državah članicah. Tokrat nas je gostilo Slovaško statistično in demografsko združenje v prostorih Statističnega urada v Bratislavi.

Dnevni red srečanja

1. Otvoritev in sprejem dnevnega reda
2. Novi pristopi v popisu prebivalstva in gospodinjstev 2021 na Slovaškem
3. Letna poročila o delu sodelujočih društev
4. Obravnava predlogov skupnega delovanja
5. Naslednje srečanje
6. Zaključek

Po sprejetju dnevnega reda smo začeli z delom. Bil sem prijetno presenečen nad vsebino novih e-pristopov v popisu 2021 na Slovaškem in uporabno vrednostjo predstavljenih poročil posameznih društev. Ugotovitev je, da imamo veliko skupnih dejavnosti in težav in je taka izmenjava izkušenj ter predstavitev možnih rešitev zelo dobrodošla. Dogovorili smo izmenjavo informacij o znanstvenih srečanjih in revijah indeksiranih v mednarodnih bibliografskih bazah, ki jih društva organizirajo oz. izdajajo. Kot prvo smo že dobili povabilo Avstrijskega statističnega združenja, da se jim leta 2020 pridružimo na njihovem letnem srečanju 'Avstrijski statistični dnevi'.

V zaključku smo izdelali osnutek in kasneje v e-dogovoru sprejeli skupno izjavo, ki vsebuje (1) skupina V7 se zavezuje k razvoju statistike kot znanstvene discipline za pridobivanje zanesljivih in objektivnih statističnih podatkov, ki bi omogočali dvig statistične kulture v skladu z EUROSTAT-ovo usmeritvijo; (2) v zvezi z volitvami novega vodstva FENStatS izrekamo zahvalo predsedniku Mauriciju Vichi in preteklemu vodstvu za opravljeno delo, novemu vodstvu s predsednikom Walterjem Redemacherjem pa želimo uspešno vodenje združenja; (3) izražamo zaskrbljenost zaradi sodnega preganjanja bivšega predsednika grškega statističnega združenja Andreasa Georgiouisa in bomo sopodpisniki protestnega pisma sestavljenega v sklopu združenja FENStatS; (4) naslednje srečanje skupine V7 bo leta 2018 v Varšavi.



Predstavniki skupine V7 v Bratislavi

AKTUALNO

DVAJSET LET ARHIVA DRUŽBOSLOVNIH PODATKOV

Dr. Mojca Vizjak Pavšič

Pogovor z dr. Sonjo Bezjak in mag. Ireno Vipavc Brvar je bil v krajši obliki objavljen 11. januarja 2018 v prilogi Znanost časopisa Delo.

Kako preveriti kakovost in verodostojnost raziskav

Arhiv družboslovnih podatkov, ki je pred nedavnim obeležil 20 let delovanja v slovenskem in mednarodnem raziskovalnem okolju, je bil ustanovljen leta 1997 na pobudo prof. dr. Nika Toša kot organizacijska enota Inštituta za družbene vede na Fakulteti za družbene vede Univerze v Ljubljani. Arhiv, ki ga vse od ustanovitve vodi doc. dr. Janez Štebe, hrani v svojih zbirkah več kot 650 raziskav, med katerimi so številne del večjih mednarodnih raziskav.

Z opravljanjem infrastrukturne znanstveno-raziskovalne dejavnosti povezuje arhiv ustvarjalce podatkov z njihovimi uporabniki in jim nudi podporo pri predaji gradiva oziroma pri njegovi uporabi, pojasnjujeta **dr. Sonja Bezjak**, raziskovalka v Arhivu družboslovnih podatkov (ADP) in **mag. Irena Vipavc Brvar**, vodja področja v ADP in vodja usposabljanja pri Konzorciju evropskih arhivov družboslovnih podatkov (CESSDA).

Medtem ko se do začetka 21. stoletja znanstveno komuniciranje ni veliko spremenilo, so v zadnjih letih sodobne digitalne tehnologije spodbudile hiter razvoj možnosti znanstvenega objavljanja, ki čedalje bolj vključuje tudi raziskovalne podatke. Kaj je temeljno poslanstvo Arhiva družboslovnih podatkov?

Bezjak: Arhiv družboslovnih podatkov po svojem osnovnem poslanstvu skrbi za izbor kakovostnih raziskovalnih podatkov s področja družboslovja s ciljem, da bi ti bili dolgotrajno shranjeni in na voljo za ponovno uporabo, morda v povsem druge namene, kot so bili sprva ustvarjeni. V tem smislu predstavlja ADP podatkovno zakladnico za družboslovje. Ker smo izvorno vezani na slovenski prostor, nas seveda v prvi vrsti zanimajo podatki, ki so nastali v Sloveniji in so pomembni za proučevanje slovenske družbe ali pa so kako drugače povezani s slovenskim prostorom.

Za uresničevanje svojega poslanstva izvajamo številne aktivnosti. Med pomembnejšimi je ovrednotenje ponujenih raziskovalnih podatkov in izbor za prevzem. Prednost imajo podatki iz znanstveno pomembnih raziskav, ki so teoretsko in metodološko dobro zasnovane, še zlasti longitudinalni in mednarodno primerljivi podatki, ki vključujejo Slovenijo. Ko opravimo t.i. evalvacijo, začnemo s pripravo paketa za dolgotrajno digitalno hrambo ter paketa za distribucijo oz. za dostop za končne uporabnike.

Poleg tega osnovnega poslanstva, ki je nekoliko bolj »tehnične« narave, pa skrbimo tudi za usposabljanje uporabnikov. Raziskovalcem, avtorjem izvornih podatkov nudimo usposabljanja in svetovanje za načrtovanje ravnanja s podatki, ki kasneje omogoča kakovostno arhiviranje podatkov. Na drugi strani pa izvajamo tematske delavnice in pripravljamo priročnike za uporabnike podatkov, kar vključuje osveščeno in kakovostno rabo podatkov.

Evropska komisija, ki je leta 2004 začela izvajati prve aktivnosti v smeri posodobitve znanstvenega komuniciranja, je v okvirnem programu za raziskave in inovacije Obzorje 2020 določila obvezen odprti dostop do vseh recenziranih objav, ki nastajajo v okviru projektov, financiranih iz javnih sredstev. Kako se Arhiv družboslovnih podatkov vključuje v evropsko politiko odprte znanosti?

Bezjak: V ADP zelo dobro razumemo pomen odprte znanosti, seveda še posebej pomen dostopa do raziskovalnih podatkov, ki so podlaga znanstvenim objavam. Dostop do podatkov je pomemben zaradi transparentnosti znanstvenega dela, pa tudi zato ker postane tako znanje lažje dostopno drugim uporabnikom izven znanstvenih krogov. S tem smo se v ADP podrobno seznanili že v okviru projekta Ministrstva za izobraževanje, znanost in šport (MIZŠ) Odprti podatki, v katerem smo izvedli vrsto intervjujev z akterji v slovenskem raziskovalnem prostoru. Sodelovali smo tudi v delovnih skupinah za pripravo strategije ter Akcijskega načrta izvedbe nacionalne strategije odprtega dostopa do znanstvenih objav in raziskovalnih podatkov v Sloveniji 2015-2020. Pa tudi sicer smo v okviru Konzorcija evropskih arhivov družboslovnih podatkov (CESSDA) vključeni v številne aktivnosti za promocijo odprte znanosti.

Seveda pri tem ne gre spregledati kompleksnosti varnega ravnanja s podatki, ki so lahko občutljivi, zaupne narave in jih je morda težko razumeti.

Včasih prihaja do nesporazumov, ko govorimo o odprtih podatkih. Osnovna zamisel je, da naj bi bili podatki čim bolj odprti in čim bolj dostopni širokemu krogu potencialnih uporabnikov. Ampak ker v družboslovju preučujemo pojave, ki lahko vključujejo občutljive podatke o ljudeh ali podjetjih in drugih organizacijah, smo dolžni le-te skladno z zakonodajo in etiko tudi skrbno varovati. Občutljivi raziskovalni podatki ne morejo biti prosto dostopni, vsekakor pa je potrebno zagotoviti, da je raziskava vpisana v katalog ADP, da je torej mogoče najti informacije o njej ter pogoje, pod katerimi je eventualno vendarle lahko dostopna, npr. recenzentom.

Vipavc Brvar: Naj dodam, da tako doma kot v okviru CESSDA izvajamo usposabljanje za načrtovanje ravnanja z raziskovalnimi podatki, ki postaja pomembna sestavina raziskovalnih projektov. 15. decembra 2017 smo kot CESSDA izdali obsežen spletni učbenik na to temo, ki je brezplačno dostopen na spletu. Spletni učbenik je namenjen samostojnemu učenju, hkrati pa bo v pomoč tudi mentorjem pri usposabljanju raziskovalcev za ustrezno ravnanje z raziskovalnimi podatki. ADP je pri tem aktivno sodeloval, zadalženi smo bili za pripravo enega od šestih poglavij.

Katere vrste podatkov sprejemate v arhiv?

Vipavc Brvar: Naj poudarim, da se velikokrat z raziskovalci pogovarjamo o tem, kaj sploh so podatki. Vprašanje se zdi preprosto, ampak včasih potrebujemo kar nekaj časa, da ugotovimo, kaj je tisto, kar bi želeli in morali shraniti v ADP. Včasih se zgodi, da raziskovalec reče, saj sem že vse objavil v tabelah v članku. Nas seveda ne zanimajo samo tabele, pač pa predvsem t.i. surovi podatki oz. mikropodatki, ki jih raziskovalci zberejo na terenu.

Skladno z našim poslanstvom sprejemamo družboslovne podatke, izjemoma tudi z drugih področij, če ni možnosti, da bi jih raziskovalci objavili drugje. Če pogledamo tematsko, so v katalogu ADP najpogosteje zastopani podatki z naslednjih področij: množične politične navade, stališča in mnenja, družbene navade in stališča, delo in zaposlovanje, informiranje in komuniciranje ter mednarodna politika in organizacije.

Glede vrst podatkov velja omeniti, da sicer prevladujejo kvantitativni podatki iz anketnih merenj, vse več pa je tudi kvalitativnih raziskav, ki so nastale z metodo intervjujev, fokusnih skupin in tudi nove vrste podatkov, kot so npr. podatki z družbenih omrežij.

Kdo ima dostop do podatkov, ki jih hranite?

Vipavc Brvar: Metapodatki o arhiviranih raziskavah so dostopni v Katalogu ADP, kjer jim je dodeljen stalni identifikator. S tem zagotavljamo možnost, da je raziskavo mogoče najti in jo povezati z objavami, ki so nastale na podlagi teh podatkov. Metapodatki in povezana gradiva, kot so vprašalniki, so brez registracije dostopni vsem obiskovalcem spletne strani ADP.

Dostop do večine mikropodatkov, na katerih lahko delamo analize, je možen ob standardni registraciji. Skrbimo, da so podatki objavljeni pod Creative Commons licencami in tako dostopni čim širšemu krogu. V praksi to pomeni, da je večina podatkov dostopnih vsem ob registraciji in obvezi, da uporabnik sprejema pogoje uporabe. Določene raziskave so omejene na raziskovalce z raziskovalno šifro, ki morajo jasno opravičiti potrebo po dostopu do občutljivih podatkov. Najbolj občutljivi podatki pa so dostopni samo v nadzorovani uporabi znotraj t.i. varne sobe. Med običajne uporabnike raziskovalnih podatkov v ADP štejemo raziskovalce, pedagoge, študente in ostalo javnost.

Omogoča vpogled v arhivirane podatke tudi preverjanje kakovosti opravljenih raziskav in verodostojnosti zbranih podatkov?

Bezjak: V prvi vrsti želim poudariti, da v ADP veliko truda vložimo v evalvacijo raziskave. V fazi prevzema raziskave skrbno preverimo vse gradivo, ki nam ga preda raziskovalec. Pri tem nas zanima tako metodološki kot konceptualni vidik ter celovitost dokumentacije, denimo, ali vsebuje informirano soglasje udeležencev, vprašalnik, kodirnik ipd., kar vse omogoča razumevanje raziskave ter preverjanje, ali so zagotovljeni zakonski in etični vidiki (npr. varstvo osebnih podatkov in avtorske pravice). Morebitne pomanjkljivosti in nejasnosti razrešimo z avtorjem in tako poskrbimo, da so podatki razumljeni ter da so tveganja za zlorabe zmanjšana.

Ko so podatki objavljeni v katalogu ADP, so dostopni uporabnikom in kot takšni ponujeni v ponovno presojo. Nema lokrat so podatki in z njimi povezano gradivo pomembni za razvoj

samih metod, za interdisciplinarno sodelovanje ter nadgradnjo obstoječih znanj. Znanost se seveda ne sme bati odkritja morebitnih napak, saj je le tako mogoče napredovati. V tem smislu je zagotavljanje dostopa vsekakor v interesu raziskovalcev in znanosti same.

Kako zagotavljate trajno digitalno obstojnost in dostopnost arhiviranih podatkov?

Vipavc Brvar: V ADP pri oblikovanju svoje informacijske tehnologije upoštevamo priporočila in zahteve konzorcija CESSDA. Redno spremljamo aktualne trende razvoja strojne in programske opreme, ki jo tudi stalno posodobljamo. Ohranjanje zbirk podatkov, distribucija podatkov in celotna infrastruktura arhiva temelji na prilagojeni IT infrastrukturi, pri čemer imajo dostop do sistema le zaposleni in registrirani uporabniki.

ADP uporablja za svoje delovanje tudi omrežje Akademске in raziskovalne mreže Slovenije ARNES, ki skrbi za stabilno, varno in učinkovito delovanje informacijsko-komunikacijske infrastrukture. Pri zagotavljanju storitev dolgotrajne hrambe pa redno sodelujemo tudi z Narodno in univerzitetno knjižnico (NUK).

Na kakšen način urejate etična vprašanja glede varovanja zasebnosti, saj družboslovne raziskave pogosto temeljijo na številnih podatkih zaupne narave?

Bezjak: Temu področju namenjamo veliko pozornosti. V družboslovju namreč pogosto ravnamo z osebnimi podatki, čeprav nas najpogosteje ne zanimajo posamezniki kot takšni, pač pa značilnosti skupin. Ob prevzemu podatkov preverimo, če jih spremlja soglasje vključenih v raziskavo in če so ustrezno anonimizirani. V preteklem letu smo se v to tematiko posebej poglobili, saj bo v letu 2018 stopila v veljavo nova evropska uredba o varstvu osebnih podatkov. Naj omenim, da uredba raziskovalcem določa pravice pri rabi osebnih podatkov za raziskovalne namene. Te, na nek način privilegije, je mogoče opravičiti s posebnim poslanstvom raziskovanja, ki cilja k novim spoznanjem in izboljšanju kvalitete življenja, zato je dostop do podatkov nujen. Hkrati pa seveda nalaga skrb in odgovornost. V ADP smo oblikovali in vodili delovno skupino za spremljanje zakonodaje s področja varstva osebnih podatkov ter izoblikovali pripombe k ZVOP-2. Pri tem smo seveda na podlagi praktičnih izkušenj preučili tiste člene, ki so pomembni za uresničevanje raziskovalnega poslanstva in poslanstva dolgotrajne hrambe podatkov za raziskovalne namene.

Sicer pa se tudi v svoji vsakodnevni praksi aktivno ukvarjamo s tem vprašanjem. Raziskovalce usposabljam glede t.i. načrtovanja ravnanja z raziskovalnimi podatki. Tak načrt je potrebno pripraviti, preden gremo na teren zbirati podatke. Pomembno je npr. opredeliti, kako bomo pridobili soglasje udeležencev v raziskavi. Pojavijo se lahko tudi različna etična vprašanja, ki jim je potrebno nameniti določen čas v raziskovanju.

Druga točka v procesu je, ko raziskovalec želi podatke predati v ADP, takrat z njim preverimo, ali je odstranil neposredne identifikatorje, o čemer se tudi izjasni v t.i. Izjavi o izročitvi. V primerih pa, ko bi odstranitev občutljivejših osebnih podatkov pomenila osiromašenje raziskave, če bi denimo morali iz intervjuja izbrisati preveč informacij, izbere raziskovalec reguliran dostop do podatkov, npr. dostopnost samo pod omejenimi pogoji v varni sobi.

In kako urejate probleme, ki se pojavljajo glede avtorskih pravic?

Vipavc Brvar: Avtorske pravice so posebno poglavje. Glede avtorstva raziskave in podatkov znotraj nje se o tem seveda dogovorijo raziskovalci, sodelavci. Nemalokrat pa gre tudi za vprašanje avtorstva izpeljanih podatkov, npr. kadar podatke snamemo s spleta ali kadar kombiniramo podatke iz zares mnogih virov in podobno.



Mag. Irena Vipavc Brvar (levo) in dr. Sonja Bezjak
Foto: Mojca Vizjak Pavšič

KONZORCIJ ARHIVOV DRUŽBOSLOVNIH PODATKOV (CESSDA)

Dr. Sonja Bezjak in mag. Irena Vipavc Brvar
Arhiv družboslovnih podatkov

Arhiv družboslovnih podatkov (ADP) je pri svojem delu tesno vpet v razvojne aktivnosti Konzorcija arhivov družboslovnih podatkov (CESSDA). Od leta 2017 tudi vodi področje t.i. usposabljanja (an. »training«), ki velja za enega od štirih podpornih stebrov te organizacije.

Usposabljanje vključuje aktivnosti za izvajalce podatkovnih storitev pri CESSDA, kar obsega oblikovanje navodil, gradiv za delavnice, učna gradiva, spletne seminarje, podporo pri uvajanju storitev in politik in drugo. Poleg tega potekajo tudi aktivnosti za usposabljanje strokovnjakov in raziskovalcev, ki niso direktno povezani s CESSDA. To obsega vsebine o načrtovanju ravnanja z raziskovalnimi podatki, delavnice s področja digitalne hrambe in iskanja podatkov, OAIS (Open Archival Information System).

V preteklem letu smo evropski arhivi družboslovnih podatkov, združeni v konzorcij CESSDA, velik del pozornosti namenili vsebinam v povezavi s pripravo načrta ravnanja z raziskovalnimi podatki (v nadaljevanju NRRP, an. Research Data Management Plan). Aktualnost teme je povezana z zahtevami financerjev raziskovanja o odprtih podatkih (npr. okvirni program EU za raziskave in inovacije Obzorje 2020, Akcijski načrt izvedbe nacionalne strategije odprtega dostopa do znanstvenih objav in raziskovalnih podatkov v Sloveniji 2015-2020) ter znanstvenih revij o dostopu do podatkov, na katerih temeljijo članki. NRRP pa lahko bistveno pripomore tudi k lažjemu arhiviranju in objavi raziskovalnih podatkov v repozitorijih, kakršen je tudi ADP. Po izkušnjah arhivov NRRP raziskovalcem še zmeraj predstavlja precejšen izziv. Iz te potrebe smo raziskovalcem v pomoč pripravili učbenik, neke vrste vodič po življenjskem krogu podatkov.

Spletni učbenik Expert tour guide on Data Management (<https://www.cessda.eu/Research-Infrastructure/Training/Expert-tour-guide-on-Data-Management>) je namenjen samostojnemu učenju raziskovalcev s področja družboslovja, ki želijo pridobiti več znanja o ravnanju z raziskovalnimi podatki (research data management). Paket znanj vključuje teoretične podlage, praktične primere in koristne napotke, ki raziskovalca vodijo skozi življenjski krog podatkov. Gre za obsežen nabor vsebin, cca. 15 ur samostojnega učenja, ki jih je mogoče osvajati po vrsti, kakor si sledijo poglavja. Lahko pa se k vračamo po potrebi, po modelu turističnih ogledov »hop-on, hop-off«.

V jedro učbenika je postavljen NRRP, ki raziskovalcu pomaga razumeti življenjski krog podatkov. Raziskovalcu skuša pokazati, kako načrtovati korake, opisovati in beležiti ravnanje s podatki in nasploh komunicirati o raziskovalnih postopkih, vse s ciljem, da bi zagotovili čim višjo kakovost in varnost podatkov, maksimizirali potencial podatkov za ponovno oz. drugo rabo ter podprli dolgotrajno ohranjanje podatkov.

Vodič je razdeljen na šest poglavij, ki sledijo življenjskemu krogu podatkov:

1. Skrbno načrtovanje ravnanja s podatki
2. Dobra organizacija in dokumentiranje ravnanja s podatki
3. Ustrezna priprava podatkov za analizo in deljenje
4. Varna hramba podatkov
5. Premišljena zaščita podatkov
6. Ahiviranje in objava podatkov

V poglavjih pa se ponavljajo tudi teme, ki raziskovalca napotijo k praktičnim vidikom ravnanja s podatki:

- Dopolni svoj Načrt ravnanja z raziskovalnimi podatki (ang. Adapt your DMP)
- Evropska raznolikost (European diversity)
- Strokovni namigi (Expert tips)
- Kvalitativni in kvantitativni vodič (Tour operators)

V ADP smo sodelovali pri zasnovi in izdelavi celotnega učbenika, sicer pa smo bili odgovorni za pripravo poglavja z naslovom »Arhiviranje in objava podatkov«. Gre za zadnji korak v življenjskem krogu podatkov, s katerim raziskovalec svoje podatke dolgotrajno shrani na najprimernejšem mestu, npr. v zaupanja vrednem področnem repozitoriju, in jih z objavo ponudi v rabo drugim. Posebno pozornost v tem poglavju smo namenili pojmu podatkovna objava, s čimer smo želeli opozoriti na pomembne razlike, prednosti in slabosti med načini in mesti objave podatkov na spletu (objava v podatkovni reviji, samo-arhiviranje, objava v različnih tipih repozitorijev).

Posvetili smo se tudi izboru ustrezne licenc, s čimer raziskovalec opredeli možnosti uporabe njegovih raziskovalnih podatkov in spremljajočih gradiv. Z objavo podatkov na ustreznem mestu zagotovi citatne pogoje svojega samostojnega raziskovalnega rezultata (naslov podatkovne datoteke, avtor(ji), leto in kraj izdelave, mesto dostopa) in omogoči drugim, da ga ustrezno navajajo pri rabi.

Objava v javno dostopnem katalogu raziskave približa potencialnim uporabnikom. Je lahko prvi korak novega raziskovalnega procesa, v katerem raziskovalec med objavljenimi podatki poišče take, ki mu bodo v pomoč pri reševanju njegovega raziskovalnega problema.

Spletni učbenik Expert tour guide on Data Management bomo evropski arhivi družboslovnih podatkov v letu 2018 skušali približati raziskovalcem tudi s pomočjo delavnic in webinarjev. Prva delavnica bo potekala 11. aprila 2018 v Ljubljani.

Povezava do spletnega učbenika: <https://www.cessda.eu/Research-Infrastructure/Training/Expert-tour-guide-on-Data-Management>.

STATISTIKA ZA BOLJŠI SVET

Dr. Mojca Vizjak Pavšič

Mag. Irena Križman, dobitnica priznanja Mednarodnega statističnega inštituta: Uradni statistični podatki so temeljnega pomena v boju proti skrajni revščini, podnebnim spremembam in upadanju biotske raznovrstnosti.



Podelitev priznanja mag. Ireni Križman za prizadevno delo v ISI v Marakešu

Podatkovna revolucija z eksponentnim večanjem obsega in virov podatkov, močno spreminja vlogo uradne statistike, pravi mag. Irena Križman, ki je pred nedavnim na svetovnem statističnem kongresu v Marakešu v Maroku prejela priznanje Mednarodnega statističnega inštituta (International Statistical Institute - ISI) s sedežem v Haagu, ki sodi med najstarejša znanstvena združenja v svetu.

Mag. Irena Križman, ki sopedreduje Odboru ISI za statistično izobraževanje in usposabljanje, zadnja štiri leta pa je bila v ožjem vodstvu ISI, je bila pred upokojitvijo dva mandata (2003-2013) generalna direktorica Statističnega urada RS (SURS), pred tem pa šestnajst let namestnica generalnega direktorja SURS. V letih 2009-2011 je vodila Mednarodno združenje za uradno statistiko. Med drugim je aktivno sodelovala v Evropskem statističnem sistemu, v Konferenci evropskih statistikov pri Združenih narodih (UNECE) ter v več strokovnih skupinah v okviru Svetovne banke in drugih mednarodnih organizacij.

»Statistika je mednarodni jezik, saj je eno njenih najpomembnejših načel mednarodna primerljivost,« poudarja mag. Irena Križman, ki je s svojo vizijo o sodobni nacionalni statistiki pomembno vplivala na razvoj slovenske uradne statistike in njeno umestitev v mednarodno statistično skupnost. V času njenega vodenja je SURS uspešno sodeloval pri vstopu Slovenije v Evropsko unijo, Evrsko območje in OECD, eden od takratnih mednarodno najodmevnejših projektov SURS pa je bila izvedba registrskega popisa prebivalstva v letu 2011, ki je nadomestil drago in zamudno zbiranje podatkov na terenu po vsej Sloveniji, za kar je bilo potrebno vzpostaviti sodelovanje s številnimi zunanjimi strokovnjaki in institucijami ter razviti nove metodološke pristope.

Ob nedavnem Dnevu evropske statistike so po vsej Evropi potekale številne prireditve pod geslom Uradna statistika kot javno dobro. Kaj je temeljno sporočilo tega mednarodnega gibanja?

Podobno kot Svetovni statistični dan, katerega geslo je Boljši podatki, boljše življenje, je tudi namen dogodkov, ki so potekali v Evropi, povečati zavedanje o tem, zakaj so podatki, ki jih zbirajo nacionalni statistični uradi, pomembni, zakaj jih je potrebno ustrezno analizirati in posredovati na uporabniku prijazen način, zakaj je potrebno tiste, ki naj bi jih uporabili, prepričati, da je za njih dobro, da jih uporabijo in jih hkrati tudi izobraževati, da bodo znali te podatke pravilno uporabiti. V zvezi s tem so še posebej pomembni projekti, katerih cilj je povečati statistično pismenost in tudi podatkovno pismenost. Razumevanje podatkov ni nujno le za politične, gospodarske in poslovne namene, ampak tudi za vsakodnevne odločitve.



Dobitnica priznanja mag. Irena Križman
Foto: Mojca Vizjak Pavšič

S katerimi izzivi se v digitalni dobi sooča statistika?

Zaradi hitrega razvoja informacijsko-komunikacijskih tehnologij imamo danes na eni strani ogromne zbirke masivnih podatkov, na drugi strani pa veliko število različnih deležnikov, ki se s temi podatki ukvarjajo, ne zgolj tiste, ki podatke o nas uporabnikih zbirajo in analizirajo, kot sta Google in Facebook, temveč tudi druga podjetja, ki želijo iz teh podatkov nekaj narediti.

Vprašanje varnosti podatkov je danes izredno pomembno, saj se v teh sistemih zbirajo tako občutljivi osebni podatki, kot tudi poslovne in druge informacije. Hitro lahko pride do večjih ali manjših zlorab. Problem, s katerim se soočamo, je pomanjkanje strokovnjakov, ki bi znali tovrstna vprašanja učinkovito reševati. Skozi to prizmo postajajo združenja, kot je Mednarodni statistični inštitut vse bolj pomembna, saj povezujejo številne organizacije, podjetja in posameznike, ki se v svetu ukvarjajo s podatki.

Torej v Mednarodnem statističnem inštitutu namenjate veliko pozornosti etičnim, strokovnim in tehničnim standardom?

Zelo si prizadevamo za uveljavitev teh standardov na lokalni, nacionalni in globalni ravni, vendar je inštitut zgolj prostovoljna organizacija, ki lahko promovira dobre prakse in opozarja na določene izzive. Predvsem na področju uradne statistike obstaja veliko število standardov, ki jih na najvišji ravni podpira Statistična komisija Združenih narodov, vendar so ti standardi še v precejšnjem delu sveta zelo slabo uveljavljeni.

Izjemno pomembna so etična načela uradne statistike, kot jih opredeljujejo Združeni narodi v dokumentu Fundamental Principles of Official Statistics. Pri pripravi tega dokumenta je aktivno sodeloval tudi ISI, ki ima status opazovalca v OZN. Na osnovi teh načel je ISI izdelal svojo Deklaracijo poklicne statistične etike, ki jo lahko uporabimo tudi, če želimo zaščititi svoje člane.

Ob tem naj omenim primer prejšnjega direktorja grškega statističnega urada Andreasa Georgiouisa, zoper katerega že več let poteka sodni pregon, ker je vztrajal pri standardu za sistem nacionalnih računov. Njegov urad je pripravil pravilne podatke za dolg in primanjkljaj, vendar so Grki te podatke zavrnili, ker so jih že pred tem prirejali, saj, kot vemo, so bili podatki, ko se je Grčija vključila v EU, napačni. ISI ta sodni pregon pozorno spremlja. Na svoji spletni strani redno objavlja stališča do tega dolgotrajnega procesa, ki je zelo drag in smo v ta namen za kolega že tudi zbirali finančno pomoč. ISI namreč izrecno poudarja, da je »objektivna in visoko kakovostna statistika nujna za demokratično delovanje družbe in blagostanje državljanov«.

Tudi Globalno partnerstvo za statistične podatke o razvoju (Global Partnership for Statistical Development Data) opozarja, da je temeljna ovira v boju proti neenakosti v svetu, problem nezanesljivih ali neobstojećih podatkov ter pomanjkanje veščin ali pripravljenosti, da bi jih uporabili.

Tako je. Tudi ISI se je pridružil temu globalnemu gibanju, ki je bilo ustanovljeno pred dvema letoma z namenom, da bi nove tehnologije in izjemno zmogljive računalniške ekosisteme uporabili za doseganje ciljev trajnostnega razvoja, kot jih opredeljuje razvojna agenda OZN do leta 2030. Ob tem poudarjajo, da so prav uradni statistični podatki temeljnega pomena za

razumevanje skrajne revščine v nerazvitem svetu, podnebnih sprememb, upadanja biotske raznovrstnosti in drugih problemov ter za iskanje novih pristopov, kako naj bi jih reševali. Globalno partnerstvo namerava to vrzel odpraviti na nacionalni ravni s pridobivanjem kakovostnih podatkov in z najbolj naprednimi analitičnimi veščinami. Poleg drugih pomembnih aktivnosti razvija tudi t.i. digitalno tržnico, ki spodbuja izvajanje strategije odprtih podatkov.

V Globalno partnerstvo je včlanjenih že več kot 250 ustanov, med njimi ameriška in britanska vlada, številni državni statistični uradi, nekatere mednarodne organizacije, organizacije civilne družbe in zasebna podjetja, med drugimi tudi IBM, Facebook in Google ter multinacionalno podjetje Esri, ki razvija geografske informacijske sisteme.

»Pojava, ki ga ne moreš meriti, tudi upravljati ne moreš,« je dejal Peter Drucker. Obstaja vizija, kako naj bi merili uresničevanje ciljev trajnostnega razvoja?

Zahteve po podatkih, ki bi verodostojno spremljali uresničevanje ciljev trajnostnega razvoja so ogromne in predstavljajo velik izziv za vse države. Trditev utemeljitelja sodobnega managementa Petra Druckerja, ki je bil tudi eden od odličnih statistikov, drži. Stvari, ki jih ne moremo izmeriti, tudi ne moremo ovrednotiti in oceniti, v katero smer se razvijajo. Ne moremo jih izboljševati, ker sploh ne vemo, kaj se je zgodilo. Kakovost in statistika – to je vse povezano. Izbrati moramo prave kazalnike, znati izmeriti njihove vrednosti in oceniti velikost njihove spremembe.

Vloga nacionalnih statističnih uradov in mednarodnih statističnih organizacij je izjemno pomembna za usklajevanje, informiranje in sodelovanje vseh, ki lahko prispevajo k spremljanju ciljev trajnostnega razvoja. Statistične organizacije so se v okviru Statistične komisije Združenih narodov dogovorile za 232 kazalnikov, s katerimi naj bi na globalni ravni merili uresničevanje ciljev Agende 2030 in jih je julija letos potrdila tudi Generalna skupščina OZN.

So bila to tudi osrednja vprašanja, ki ste jih obravnavali na prvem Svetovnem forumu Združenih narodov o podatkih?

Da, cilj foruma, ki bo odslej organiziran vsaki dve leti, je s povezovanjem vseh deležnikov sodelovati pri vzpostavitvi sistemov za zbiranje in analiziranje podatkov, potrebnih za spremljanje ciljev trajnostnega razvoja po vsem svetu. Zaradi vseh teh novih zahtev je dala namreč skupina neodvisnih strokovnjakov na poziv generalnega sekretarja OZN pobudo za organizacijo prvega svetovnega foruma o podatkih, ki je potekal januarja letos v Cape Townu v Južni Afriki. Na forumu, kjer sem sodelovala kot članica programskega odbora, se je zbralo okoli 1.500 strokovnjakov iz več kot sto držav in številne skupine, ki so predstavljale oblikovalce politik, javni in zasebni sektor, izobraževalno in raziskovalno področje ter organizacije civilne družbe.

Razprave v Cape Townu so pokazale, da se nacionalni statistični uradi vse bolj srečujejo z drugimi podatkovnimi skupnostmi, kar prinaša številne izzive, kot so potreba po vzpostavljanju novih partnerstev, dostop do sekundarnih podatkovnih virov ter kakovost, transparentnost in etičnost objavljanja podatkov, ki jih zbirajo deležniki izven uradnih statističnih sistemov. Ob

tem so udeleženci foruma opozarjali tudi na potrebe po spremembi statistične zakonodaje, uvajanju etičnih in procesnih standardov ter novih poslovnih modelov. Izjemno pomemben dosežek foruma pa je sprejem globalnega akcijskega načrta, ki poziva nacionalne vlade in mednarodno skupnost k sodelovanju pri izvajanju ključnih ukrepov za uresničevanje ciljev Agende 2030.

Pogovor dr. Mojco Vizjak Pavšič z mag. Ireno Križman je bil objavljen 9. novembra 2017 v prilogi Znanost časopisa Delo.

Več o odboru ISI: <https://isi-web.org/index.php/news-from-isi/5245-statistical-capacity-building-2>.

EVROPSKE STATISTIČNE IGRE

PRILOŽNOST ZA ŠIRJENJE IN KREPITEV STATISTIČNE PISMENOSTI MED ŠOLAJOČO SE MLADINO

Simona Klasinc
Statistični urad RS

Kakšne igre?

Evropske statistične igre so tekmovanje, ki ga je pripravil in ga izvaja Statistični urad RS v sodelovanju z Evropsko komisijo in Statističnim društvom Slovenije. Srednješolci preverjajo pod vodstvom mentorjev svoje znanje iz poznavanja osnov statistike in iz poznavanja in pravilnega razumevanja ter uporabe statističnih podatkov. To tekmovanje je drugačno od običajnih slovenskih, saj spodbuja ekipno delo, iznajdljivost in kreativnost dijakov ter hkrati preverja raven njihove statistične pismenosti.

Kako potekajo?

Na tekmovanje ESI se je prijavilo približno 140 ekip (sestavljenih iz enega, dveh ali največ treh članov), v katerih sodeluje skoraj 400 dijakov in 36 mentorjev iz 35 srednjih šol. Razdeljeni so v dve tekmovalni kategoriji: ena vključuje dijake prvih in drugih letnikov, druga pa dijake tretjih in četrtih letnikov. Tekmovanje se je začelo oktobra 2017 in bo trajalo do junija 2018 in obsega različne aktivnosti, in sicer na nacionalni in evropski ravni.

Prvi krog nacionalnega tekmovanja je za nami. V njem so sodelujoče ekipe odgovarjale prek spleta na vprašanja zaprtega tipa. Pomerile so se:

- v osnovnem znanju statistike,
- v uporabi uradnih statističnih virov (podatkovna vira: spletni strani SURS in Eurostata),
- v razlaganju in razumevanju (interpretiranju) statističnih objav (podatkovni vir: interaktivna publikacija Življenje žensk in moških v Evropi).

Zgledi vprašanj:

Par pelje svojega šestmesečnega otroka k pediatru. Ta ga stehta in izmeri njegovo višino. Razloži jima, da predstavlja višina njunega otroka 70. percentil. To pomeni:

- A. Da je med stotimi otroki njegove starosti visok 70 cm.
- B. Da je otrok visok 70 cm.
- C. Da je 70 % otrok njegove starosti nižjih od njega.
- D. Da je otrok nižji od 70 % otrok njegove starosti.

Katera slovenska občina je bila najbolj priljubljeni cilj tujih turistov v letu 2016?

- A. Bled.
- B. Piran.
- C. Kranjska Gora.
- D. Ljubljana.

V Evropski uniji je bilo v 2016 v vseh starostnih skupinah več žensk kot moških, razen v skupini:

- A. Do 18 let.
- B. Od 18 do 44 let.
- C. Od 44 do 65 let.
- D. Več kot 65 let.

V drugem krogu nacionalnega tekmovanja (ta je potekal prav v času pisanja tega članka) se je 70 ekip iz obeh kategorij (polovica vseh) pomerilo v raziskovalnem delu. Uporabiti so morale podatkovni niz v obliki Excel, ga obdelati, analizirati in rezultate predstaviti v obliki razumljive vizualizacije podatkov. Rezultate raziskovalne naloge so morale predstaviti v programu Power Point in shraniti v zapisu pdf.

Najboljše tri ekipe iz vsake kategorije bodo napredovale na evropsko tekmovanje. Tam se bodo s tekmovalci iz drugih sodelujočih držav (Hrvaške, Cipra, Portugalske, Španije, Italije, Francije, Poljske, Bolgarije, Grčije) pomerile v izdelavi videoposnetka s temo Zakaj je uradna statistika pomembna v naši družbi? Glavna merila, ki bodo usmerjala delo ekip, bodo: kreativnost videa, prepričljivost sporočila glede na postavljeno vprašanje ter učinkovitost videa za posredovanje sporočila. Zmagovalna ekipa evropskega finala bo nagrajena s prenosnimi računalniki ter potovanjem v Krakov na Q konferenco 2018 (kakovost v statistiki).

Motivi za sodelovanje dijakov in mentorjev

Mentorji, ki so se odločili za sodelovanje in k temu spodbudili dijake, so statistične igre prepoznali kot priložnost zase in za dijake. Nekateri so morda s tem popestrili redne učne ure in dijakom približali statistiko ob praktičnih primerih uporabe podatkov, drugi so temu posvetili nekaj časa med urami, namenjenimi raziskovalnemu in podobnemu delu. Tisti, ki so se udeležili seminarjev, ki so bili sestavni del priprav na to tekmovanje, so svoje znanje lahko tudi izpopolnili in poglobili.

Ne nazadnje so bile morda spodbuda za dijake tudi praktične nagrade. Finalisti slovenskega finala bodo nagrajeni s tabličnimi računalniki, vsi udeleženci pa s praktičnimi darili, ki so hkrati

tudi promocijsko gradivo SURS-a (majčke, vrečke, svinčniki, čokolade). Če bo slovenska ekipa zmagala v evropskem finalu, se bo udeležila podelitve nagrade v Krakovu na Poljskem (nagrada bodo prenosni računalniki).

Uspešno povezovanje in sodelovanje

Statistično društvo in Statistični urad pri izvedbi tekmovanja zgledno sodelujeta na več področjih. Dr. Andrej Blejec je v imenu društva izvedel štiri seminarje o poznavanju osnov statistike, statističnih podatkov, o uporabni vrednosti statistike, o pravilnem razumevanju in pravilni uporabi statistike za srednješolske učitelje (v Ljubljani, Mariboru, Postojni, Novem mestu). Od petintridesetih v tekmovanju sodelujočih učiteljev se jih je enega od seminarjev udeležilo kar dvajset, kar pomeni, da so bili seminarji tisto najpomembnejše, kar je pritegnilo sodelujoče, da so se prijavili na tekmovanje.

V žiriji, ki bo ocenjevala najboljše raziskovalne naloge ter določila finaliste in zmagovalca iz vsake od kategorij, sodelujejo tudi dva predstavnika SURS-a in tudi dva predstavnika Statističnega društva.

Drugo leto?

Počakali bomo na odzive, vendar prvi že kažejo na to, da Evropske statistične igre ne bodo enkratni dogodek. Verjamemo, da z izobraževalnimi in promocijskimi aktivnostmi, ki jih izvajamo med mladimi, svojo energijo usmerjamo učinkovito in da spodbujamo uporabo in pravilno rabo statističnih podatkov pri svojih prihodnjih uporabnikih. In tega smo zelo veseli.

UNIVERZA V LJUBLJANI

DOGODKI EKONOMSKE FAKULTETE

Ekonomski fakulteta Univerze v Ljubljani v sodelovanju z Eurostatom organizira več dogodkov za mrežo EMOS (European Master in Official Statistics). Vodja projekta je **izr. prof. dr. Mojca Bavdaž**.

Med aktivnostmi izpostavljam serijo 12 webinarjev, ki jih Ekonomski fakulteta gosti med januarjem in junijem 2018 in je primarno namenjena magistrskim programom EMOS in institucijam uradne statistike po Evropi (statistični uradi, centralne banke itd.). Serijo je 10. januarja 2018 odprla Mariana Kotzeva, nova generalna direktorica Eurostata. Program že izvedenih in prihajajočih webinarjev na teme s področja uradne statistike je dostopen na <http://konference.ef.uni-lj.si/emos/webinars/>.

6. marca 2018 je v Ljubljani potekala delavnica EMOS z naslovom "*Taking EMOS to the next level*", na kateri so se srečali predstavniki že akreditiranih magistrskih programov in tistih, ki bi to radi postali, predstavniki institucij evropskega statističnega sistema in evropskega sistema centralnih bank, kakor tudi drugi deležniki. Program dogodka je dostopen na <http://konference.ef.uni-lj.si/emos/workshop/programme>.

Prof. dr. Irena Ograjenšek z Ekonomske fakultete je v sodelovanju z dr. Iddom Galom z Univerze v Haifi v reviji *Journal of Official Statistics* februarja 2017 objavila znanstveni članek z naslovom *Official Statistics and Statistics Education: Bridging the Gap*. Zainteresiranim bralcem je članek prosto dostopen na naslednji spletni povezavi: <https://doi.org/10.1515/jos-2017-0005>.

V članku avtorja uvedeta pojem uradne statistične pismenosti (angl. official statistics literacy), opredelita njene ključne elemente ter nakažeta možne pristope k njenemu izboljšanju v populaciji odraslih uporabnikov storitev institucionalnih ponudnikov uradne statistike.

Senat Univerze v Ljubljani je **prof. dr. Ireno Ograjenšek** z Ekonomske fakultete na svoji 3. seji v mandatnem obdobju 2017-2019 dne, 19. decembra 2017 izvolil v naziv redne profesorice in znanstvene svetnice za področje statistike. V utemeljitvi ob izvolitvi je med drugim zapisano:

"V slovenski in mednarodni znanstveni javnosti velja za priznano raziskovalko na področju statistike. V svojih najpomembnejših oziroma najodmevnejših delih sistematično preverja, nadgrajuje in integrira parcialne pristope k merjenju zadovoljstva zaposlenih in uporabnikov ter kakovosti storitev, merjenju kakovosti življenja in oblikovanju alternativnih kazalcev blaginje ter k identifikaciji in kvantifikaciji dejavnikov konkurenčnosti ter gospodarske rasti. Posebno področje njenega znanstvenega udejstvovanja predstavljajo izzivi statističnega izobraževanja."

Na Ekonomski fakulteti v aprilu 2018 pričenjamo s projektom ARROW (Improvement of Research and Innovation Skills in Mongolian Universities) v okviru sheme Erasmus +. V njem bo konzorcij partnerskih univerz iz Evropske unije pomagal partnerskim univerzam iz Mongolije razviti raziskovalno in inovacijsko platformo. Naloga raziskovalcev z Ekonomske fakultete Univerze v Ljubljani pod vodstvom **prof. dr. Irene Ograjenšek** je z razvojem kurikulumov za predmete metodološko-statistične vertikalne prispevati k dvigu kakovosti pouka statistike in posledično statistične pismenosti mongolskih študentov in bodočih raziskovalcev.

IZOBRAŽEVANJE

PREDAVANJA NA IBMI

Na Inštitutu za biostatistiko in medicinsko informatiko (IBMI) je približno enkrat na mesec predavanje, ki je namenjeno vsem, ki jih zanima statistika. Zapis enega izmed takih predavanj z naslovom *Vzorci v glasbi*, ki ga je imela 7. 12. 2017 mag. Tinka Majaron, sledi v nadaljevanju.

VZORCI V GLASBI

Mag. Tinka Majaron
Konservatorij za glasbo in balet Ljubljana

V tem povzetku si bomo ogledali, kako lahko s programskim jezikom R poiščemo vzorce v melodiji neke skladbe. Predstavitev je delni povzetek mojega magistrskega dela na Fakulteti za matematiko in fiziko Univerze v Ljubljani.

Vnos podatkov

S statistično analizo melodije želimo poiskati vzorce, ki se največkrat ponovijo. V ta namen moramo notni zapis spremeniti v smiselne številske vrednosti. Za pisanje not je na tržišču kar nekaj programov. V svojem magistrskem delu sem uporabila ABC Plus, saj je prosto dostopen in kompatibilen z LaTeXom. V programskem jeziku C sem nato za računanje iskanih številskih vrednosti preprogramirala sintaktični analizador (parser) za ABC Plus avtorja Rema Dentata.

Poiskali bomo vzorce v prvem stavku Partite za flavto solo (BWV 1013) Johanna Sebastiana Bacha. Na Sliki 1 sta prva dva takta te skladbe, v Tabeli 1 pa izpis številskih vrednosti za prvih 7 not, ki ga ustvari že omenjeni program.

Slika 1



Tabela 1

Centi	Trajanje	Čas v skladbi	Čas v taktu
1.600,00	0,062500	0,062500	0,062500
2.100,00	0,062500	0,125000	0,125000
2.005,00	0,062500	0,187500	0,187500
2.100,00	0,062500	0,250000	0,250000
2.400,00	0,062500	0,312500	0,312500
2.100,00	0,062500	0,375000	0,375000
1.600,00	0,062500	0,437500	0,437500

Centi predstavljajo višino note, polton meri 100 centov. Privzeto je, da je enočrtni C na višini 0 centov, vsi ostali toni imajo vrednost centov preračunano iz tega izhodišča: -100 centov za vsak polton navzdol in +100 centov za vsak polton navzgor. Dvočrtni C ima tako višino 1.200 centov. Poleg tega je za višaj prištetih 5 centov, za nižaj pa odštetih 5 centov, da po potrebi lahko ločimo zvišane in znižane note. Dvočrtni Cis ima tako višino 1.305 centov, Des pa 1.295 centov.

Za trajanje note je privzeto, da celinka traja 1 enoto, vse ostale notne vrednosti so ustrezno preračunane. Četrtnika tako traja 0,25 enote, osminka 0,125 enote, šestnajstinka 0,0625 enote in tako naprej.

Čas v skladbi je merjen v enakih enotah kot trajanje not, teče od začetka skladbe naprej. Čas v taktu je merjen podobno, le da se vsak takt začne šteti znova (z 0).

Analiza podatkov

Notam smo priredili številске vrednosti, zato za analizo melodije lahko uporabimo poljuben program za statistično analizo. V svojem delu sem uporabila prosto dostopen programski jezik R.

Del programa, ki poišče vzorce:

```
patterns <- function(deg, dur, n, fil) {
  write(paste("length=",n), fil, append=TRUE, ncolumns=1)
  k <- length(deg)-n; j <- 1
  if (k > 2) {stringDD <- c(1:k); for (j in c(1:k)) {stringDD[j] <- ""}
    j <- 1; m <- 1
    for (j in c(1:k)) {for (m in c(1:(2*n))) {if (m <= n) {stringDD[j] <- paste(stringDD[j],
  "", deg[j+m-1], sep="")} else {stringDD[j] <- paste(stringDD[j], "R", sprintf("%.4f", dur[j+m-1]),
  sep="")}}}
  stringTab <- as.data.frame(table(stringDD),stringsAsFactors=FALSE)
  re <- sort.int(stringTab[,2], decreasing=TRUE, index.return=TRUE)
  write(paste("number of repetitions=", re$x[1]), fil, append=TRUE,
ncolumns=1)
```



```

if (re$x[1] > 2) {write(stringTab[re$ix[1],1], fil, append=TRUE, ncolumns=1)
  j <- 2
  while (re$x[j] == re$x[1]){write(stringTab[re$ix[j],1], fil, append=TRUE,
ncolumns=1); j <- j+1}
  }
}

i <- 1
for (i in c(2:30)) {patterns(intStr, durInt, i, filNam)}

```

Program izpiše najpogostejše intervalne vzorce dolžine od treh do enaintridesetih tonov (od motiva do teme). Pri tem upošteva tudi ritem - dva vzorca sta enaka, če se ujemata v intervalih in v ritmičnih razmerjih. Običajen interval je razlika med višinama sosednjih tonov, ritmično razmerje pa je količnik dolžin (trajanj) teh dveh tonov.

Rezultati

Program najdene vzorce pretvori nazaj v note. Oglejmo si izpis vzorcev glede na intervale med zaporednimi toni. Izpis je brez taktovskega načina, en takt predstavlja en vzorec - število tonov v taktu predstavlja dolžino vzorca. Na začetku takta je izpisano, kolikokrat se v celotni skladbi vzorec ponovi. Izpišejo se le vzorci, ki se ponovijo vsaj trikrat. Če obstaja več vzorcev, ki se ponovijo enako mnogokrat, izpiše vse. Opazujemo le intervale, ni pomembno, na kateri višini se pojavijo. Izpis vzorcev skladb, ki so napisane v molu, se vedno začne na enočrtnem A-ju, skladb, ki so napisane v duru, pa na dvočrtnem C-ju. Dolžina prve note je četrtninka, dolžine ostalih not so izračunane glede na ritmična razmerja. Tukaj zaradi preglednosti zapišimo le najkrajše (Slika 2) in najdaljše vzorce (Slika 3).

Slika 2



Slika 3

The image shows a musical score for Slika 3, consisting of 11 staves of music. The music is written in treble clef with a key signature of one sharp (F#). The score features various rhythmic patterns and melodic lines, with some measures marked with a '4' below the staff, likely indicating a four-measure phrase or a specific rhythmic grouping. The notation includes eighth and sixteenth notes, rests, and accidentals.

Opazujemo lahko tudi vzorce glede na stopnje, vzorce glede na začetek v taktu in še kaj.

Zaključek

Pri iskanju teh vzorcev smo v resnici uporabili zelo malo statistike, zato za konec naštejmo še nekaj statističnih parametrov, ki veliko povejo o skladbah, glede na njihove vrednosti pa skladbe lahko tudi razvrščamo v skupine.

Najosnovnejša informacija o melodiji je njen obseg ali ambitus, to je interval med najvišjim in najnižjim tonom. Obseg se uporablja kot merilo za razvitost melodije. V melodiji so toni bolj ali manj razpršeni okrog njihove srednje vrednosti. Standardni odklon tonov v centih je merilo za razgibanost melodije. Avtokorelacija v melodiji išče morebitne periodičnosti. Izračunamo vse navzkrižne avtokorelacije (z razmakom od 0 do števila tonov v melodiji) in si ogledamo lokalne maksimume, ki nakazujejo periode. Entropija je pričakovana informacija o pojavljanju tonov pri naključnem izbiranju tona,

$$H = - \sum_{i=1}^k (p_i \log_2 p_i),$$

pri čemer je p_i verjetnost pojavitve posameznega tona, k pa število tonov. Entropijo lahko privzamemo kot merilo za razvitost melodijske zgradbe.

Raziskovanja so pokazala, da je za ocenjevanje glasbenega sloga najpomembnejša kurtozis. Vrednosti kurtozis so skozi obdobja umetnostne zgodovine enakomerno naraščale in so zanesljivo merilo stopnje razvitosti glasbenega sloga. Zanimivo je, da ima dvanajsttonska glasba ali dodekafonija nizko vrednost kurtozis (v nasprotju z moderno) in največjo vrednost entropije. Dodekafonija ima torej najbolj razvito zgradbo, ocene njenega glasbenega sloga pa se spustijo nazaj na vrednosti iz renesanse in baroka.

Literatura

- T. Majaron: Vzorci v glasbi, magistrsko delo, Fakulteta za matematiko in fiziko Ljubljana, 2016.
 G. J. McLachlan: Discriminant Analysis and Statistical Pattern Recognition, John Wiley & Sons, 2004.
 B. Ravnikar: Osnove glasbene akustike in informatike, Državna založba Slovenije, 2001.
 D. Temperley: The Cognition of Basic Musical Structures, Massachusetts Institute of Technology, 2001.

POVZETKI DOKTORSKIH DISERTACIJ

V letu 2017 so na interdisciplinarnem doktorskem študijskem programu Statistika uspešno zagovarjali doktorske disertacije štirje študenti. Podrobneje smo o študijskem programu Statistika 2017/2018 že poročali v posebni publikaciji, izdani ob 40. obletnici delovanja društva.

V nadaljevanju objavljamo povzetke doktorskih disertacij Vilme Sem in Joce Zurc ter Tanje Istenič. Prvi dve sta uspešno doktorirali na doktorskem študijskem programu Statistika, tretja pa na doktorskem študijskem programu Ekonomske in poslovne vede.

UVRŠČANJE NEURAVNOTEŽENIH VISOKO RAZSEŽNIH PODATKOV V VEČ RAZREDOV

Dr. Vilma Sem
Mentorica izr. prof. dr. Lara Lusa
Medicinska fakulteta, Univerza v Ljubljani

Uvod

Cilj uvrščanja je zgraditi model, ki na podlagi merjenih spremenljivk neznani enoti določi pripadnost razredu. Model smo zgradili na osnovi znanih enot, za katere poleg vrednosti merjenih spremenljivk poznamo tudi dejansko pripadnost razredu. Začetni namen doktorskega dela je bil poiskati model, ki bi na podlagi podatkov bližnje infrardeče spektroskopije (NIRS, Near Infrared Spectroscopy) napovedal vrsto plastik (polimerov), ki so pogost sestavni del predmetov v muzejskih zbirkah. Namen je bil poiskati učinkovit in praktično uporaben model uvrščanja, ki pa je postal kompleksen zaradi lastnosti podatkov (neuravnoteženost, visoka razsežnost, koreliranost in veliko število različnih polimerov – razredov).

Ker se metodo bližnje infrardeče spektroskopije v zadnjem času zaradi cenovne ugodnosti, preproste uporabe in nedestruktivnosti vse pogosteje uporablja v raziskavah na različnih področjih (varnost hrane, medicina, kemija, biologija ...) smo želeli pripraviti tudi praktične napotke za uvrščanje podatkov z opisanimi lastnostmi, za kar je bilo potrebno metode uvrščanja raziskati s simulacijami. Ker se obstoječi pristopi generiranja podatkov niso dovolj približali realnemu stanju podatkov, je bilo potrebno razmišljati o novih načinih generiranja podatkov.

Uporabljene metode

V vzorčni zbirki zgodovinskih polimerov je bilo 534 predmetov (enote), ki so pripadali 45 različnim vrstam polimerov (razredi). Pravilo za uvrščanje smo zgradili na podlagi odboja svetlobe v NIR območju pri 191 valovnih dolžinah (spremenljivke). Za raziskovanje metod uvrščanja smo uporabili računalniške simulacije. Za namen simulacij smo generirali podatke na štiri različne načine, pri čemer sta bila dva načina preprostejša in že večkrat uporabljena v literaturi, dva načina pa sta bila predstavljena na novo. Na simuliranih in realnih podatkih smo za obdelavo uporabili štiri metode predobdelave podatkov (brez predobdelave, standardna normalna vektorska transformacija, kvantilna normalizacija in prvi odvod izračunan z metodo Savitsky-Golay), štiri metode za zmanjšanje dimenzije podatkov (brez zmanjšanja dimenzije, izbor spremenljivk z največjo varianco, izbor spremenljivk z največjo F-statistiko in metoda glavnih komponent), štiri metode uvrščanja (linearna diskriminantna analiza, metoda k-najbližjih sosedov, odločitvena drevesa in metoda podpornih vektorjev), trije pristopi za zmanjšanje neravnotežja (vsak-proti-vsakem, večkratno zmanjšanje večjega razreda in prilagoditev praga za uvrščanje). V postopku vrednotenja uvrščanja smo uporabili dva pristopa (100-krat ponovljeno prečno preverjanje s pregibanjem in 500-krat ponovljena stratificirana razdelitev na učno in testno množico) ter 16 različnih mer za vrednotenje uvrščanja. Rezultate 16 mer za vrednotenje uvrščanja smo dodatno analizirali z analizo usklajenosti (analiza konkordance).

Rezultati

Zgradili in ovrednotili smo 300 modelov uvrščanja 45 vrst polimerov. Izbrani model (metoda predobdelave: standardna normalna vektorska transformacija, metoda zmanjšanja dimenzije: brez, metoda zmanjšanja vpliva neravnotežja: brez, metoda uvrščanja: linearna diskriminantna analiza) napoveduje neznane enote z dovolj visoko natančnostjo, da je praktično uporaben za identificiranje polimernih materialov v muzejskih zbirkah (povprečna napovedna točnost razreda: 0,88). Napake v napovedih so večinoma posledica zelo podobne kemijske zgradbe polimerov v nekaterih skupinah in za uporabnika v muzeju niso zelo moteče.

Analiza konkordance, ki je bila narejena za 16 mer za vrednotenje uvrščanja je pokazala visoko stopnjo skladnosti (Kendallov koeficient konkordance $W=0,95$). Študija, ki bi obravnavala toliko različnih mer za vrednotenje uvrščanja, do sedaj še ni bila narejena. V posameznih študijah so bile primerjane le nekatere od obravnavanih mer. V doktorskem delu predstavljeni rezultati se ujemajo z večino teh študij, ne pa z vsemi.

Predstavljeni rezultati kažejo, da lahko napačne metode generiranja podatkov pripeljejo do zavajajočih zaključkov. Za namen simulacij smo zato predlagali dva nova pristopa generiranja podatkov, ki sta se izkazala za primernejša od preprostejših pristopov, ki se sicer pogosto uporabljajo v simulacijskih študijah (Slika 1).

Uporaba predobdelave pri uvrščanju podatkov bližnje infrardeče spektroskopije se je izkazala za nujno, medtem ko so metode za zmanjšanje dimenzije podatkov rezultate uvrščanja večinoma poslabšale. Ta ugotovitev potrjuje smiselnost rutinske uporabe metod predobdelave pri raziskavah s področja bližnje infrardeče spektroskopije, pod vprašaj pa postavlja uporabo prav tako zelo pogosto uporabljene metode glavnih komponent za zmanjšanje dimenzije podatkov. Za zmanjšanje vpliva neravnotežja, sta se prilagoditev praga za uvrščanje in metoda vsak-proti-vsakem izkazali za obetajoči. Od obravnavanih metod uvrščanja je, kljub kršeni predpostavki o neodvisnosti spremenljivk, linearna diskriminantna analiza izkazala najboljše rezultate, na kar je bilo v literaturi že večkrat opozorjeno. Najbolj občutljiva na neravnotežje pa je bila metoda klasifikacijskih dreves.

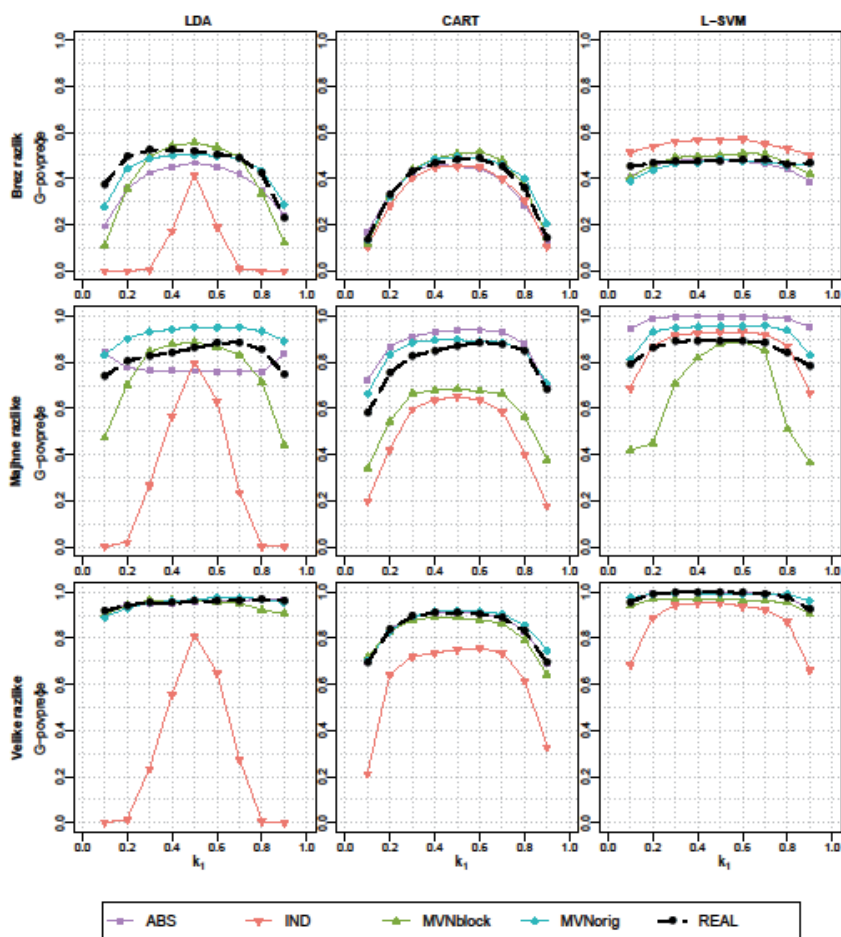
Zaključek

V doktorskem delu smo predstavili rešitev problema uvrščanja zgodovinskih plastik v 45 razredov, ki je prva za uporabnika sprejemljiva rešitev uvrščanja zgodovinskih plastik na podlagi bližnje infrardeče spektroskopije. Rešitev je bil model uvrščanja, izbran izmed 300 modelov uvrščanja, ki so bili zgrajeni z različnimi kombinacijami metod za uvrščanje, predobdelavo podatkov, zmanjšanje dimenzije podatkov in zmanjšanje vpliva neravnotežja. Modeli so bili ovrednoteni s 16 merami za vrednotenje uvrščanja.

V nalogi smo predlagali dva nova pristopa umetnega generiranja podatkov bližnje infrardeče spektroskopije, ki sta uporabna za raziskovanje lastnosti različnih statističnih metod, ki so kategorizirani v skupine.

Na podlagi simulacij smo za linearno diskriminantno analizo, odločitvena drevesa in metodo podpornih vektorjev naredili sistematično študijo, ki proučuje lastnosti uvrščanja podatkov bližnje infrardeče spektroskopije v dva, tri in, v omejenem obsegu, tudi 45 razredov, ob različnih stopnjah neravnotežja in ob uporabi metod za predobdelavo podatkov, zmanjšanje dimenzije podatkov in zmanjšanje vpliva neravnotežja.

Slika 1 Povprečne vrednosti mere G pri uvrščanju v dva razreda z metodami LDA, CART in L-SVM brez uporabe predobdelave in brez uporabe metod za zmanjšanje dimenzije podatkov. Različne barve predstavljajo različne metode generiranja podatkov: ABS – nov pristop generiranja podatkov na podlagi teoretičnih absorpcij funkcionalnih skupin v NIR območju, IND – spremenljivke so generirane neodvisno iz normalne porazdelitve, MVNblock – spremenljivke so generirane iz multivariatne normalne porazdelitve z bločno kovariančno matriko, MVNorig – nov pristop, kjer so spremenljivke generirane iz multivariatne normalne porazdelitve s parametri ocenjenimi iz realnih podatkov, REAL – realni podatki.



KRITERIJI OCENJEVANJA VELJAVNOSTI V RAZISKAVAH Z INTEGRACIJO KVANTITATIVNE IN KVALITATIVNE METODOLOGIJE

Ddr. Joca Zurc

Mentorica prof. dr. Anuška Ferligoj

Fakulteta za družbene vede, Univerza v Ljubljani

Doktorska disertacija obravnava področje ocenjevanja kakovosti raziskovanja na primeru raziskav, ki hkrati vključujejo statistično oz. kvantitativno metodologijo ter kvalitativno raziskovanje [angl. mixed methods research]. Raziskovanje z integracijo metod se je izredno razvilo v zadnjih 30 letih, saj ta metodologija izhaja iz samega raziskovalnega fenomena oz. paradigme pragmatizma in omogoča holističen pogled na predmet raziskovanja (Teddle in Tashakkori 2009). Danes je integracija metod poznana in priznana kot tretja metodološka skupina v znanosti (Johnson in Onwuegbuzie 2004). Dosedanja raziskovanja so se ukvarjala z različnimi odprtimi vprašanji te metodologije, kot so njena konceptualizacija, razvoj različnih tipologij raziskovalnih načrtov ter njihova uspešnost pri raziskovanju kompleksnih raziskovalnih vprašanj v različnih znanstvenih disciplinah. Metodološko področje, ki je izredno aktualno in še neraziskano, je področje veljavnosti raziskav z integracijo metod. Veljavnost je ob zanesljivosti ključna komponenta kakovosti raziskav (Ferligoj in drugi 1995; Kogovšek 1998; Silverman 2005; Neuman 2011) in pomeni ugotavljanje resničnosti, verodostojnosti in prepričljivosti raziskave ter odgovor, ali so dobljene ugotovitve dovolj avtentične glede na resnično realnost, da jim lahko zaupamo pri posploševanju (Guba in Lincoln 2005; Neuman 2011).

Med obstoječimi teoretskimi modeli ocenjevanja veljavnosti raziskav z integracijo metod se med avtorji kažejo razlikovanja v poudarkih različnim vidikom veljavnosti, neenotni terminologiji in majhni pozornosti samim kriterijem ocenjevanja veljavnosti (Sale in Brazil 2004; Bryman 2008; Bryman in drugi 2008; O'Cathain 2010). Eksperti opozarjajo, da obstaja potreba po naboru in soglasju temeljnih kriterijev veljavnosti, ki bodo dali smernice za izvedbo kakovostne raziskave z integracijo metod, vzpostavili skupni metodološki jezik ter bodo v pomoč številnim interesnim skupinam, kot so raziskovalci, financerji, učitelji, tvorci politik in drugi zainteresirani uporabniki.

Naša doktorska raziskava je bila namenjena iskanju optimalnih kriterijev oz. standardov za ocenjevanje veljavnosti v raziskavah z integracijo kvantitativne in kvalitativne metodologije. Zanimalo nas je, kateri so ključni kriteriji veljavnosti raziskave z integracijo metod, ki jih mora upoštevati vsaka raziskava ne glede na temo ali znanstveno področje proučevanja. Cilj pa je bil tudi razviti merski instrument za samooceno veljavnosti, ki bo vključeval temeljne kriterije veljavnosti in ga bo možno uporabiti pri načrtovanju, izvajanju ali evalviranju vsake raziskave z integracijo metod ter ga bodo lahko uporabljali sami raziskovalci kot tudi druge interesne skupine ne glede na znanstveno vedo in metodološko predznanje.

Doktorska disertacija je bila izvedena v obliki treh zaporednih empiričnih raziskovalnih faz in z uporabo metodološkega modela IDCIV – Instrument Development and Construct Validation (Onwuegbuzie in drugi 2010), ki preko 10 interaktivnih korakov usmerja razvoj novega merskega instrumenta. V prvi fazi je s kvalitativnim pristopom metasinteze, s katerim smo iskali skupne tematske kategorije, potekal sistematični pregled literature z vidika obstoječih kriterijev za ocenjevanje veljavnosti v raziskavah z integracijo metod, pri čemer smo se osredotočili na analizo kvantitativnega dela, kvalitativnega dela ter postopka integracije obeh delov v isti raziskavi. Dobljeni rezultati so predstavljali temelj za razvoj druge raziskovalne faze, v kateri smo s kvalitativnimi polstrukturiranimi poglobljenimi intervjuji proučevali mnenja devetih uveljavljenih pionirjev metodologije integracije metod. Druga raziskovalna faza je potekala na mednarodni ustanovni konferenci združenja Mixed Methods International Research Association (2014) v Bostonu, ZDA. Za analizo transkriptov intervjujev je bila uporabljena kvalitativna vsebinska analiza, ki je potekala s pomočjo programskega orodja NVivo, verzija 11.0. V tretji raziskovalni fazi pa smo pozornost posvetili razvoju in testiranju novega merskega instrumenta za ocenjevanje veljavnosti. Pilotno različico merskega instrumenta smo trikrat testirali na vzorcu 56 dodiplomskih in podiplomskih študentov na Turku University of Applied Sciences na Finskem. Na osnovi ugotovitev pilotnih testiranj je bila izdelana izpopolnjena verzija merskega instrumenta. Revidirani merski instrument smo testirali na vzorcu treh izkušenih recenzentov in 10 objavljenih raziskav z integracijo metod iz različnih področij znanosti. Testiranje revidiranega merskega instrumenta je vključevalo statistično analizo kot tudi kvalitativna mnenja recenzentov, ki so jih izrazili na treh usklajevalnih sestankih, namenjenih iskanju soglasja o temeljnih kriterijih ocenjevanja veljavnosti v raziskavah z integracijo metod. Soglasnost recenzentov je bila analizirana z indeksom soglasnosti in analizo rangov. Raziskovalne faze so bile izvedene v zaporedju, vsaka predhodna je bila pogoj za razvoj naslednje.

Rezultati prve raziskovalne faze na vzorcu 18 raziskav so pokazali tri osrednje značilnosti ocenjevanja veljavnosti: 1) spreminjanje koncepta veljavnosti in terminologije skozi čas, 2) različni pristopi in vrste k ocenjevanju veljavnosti v kvantitativnih raziskavah, kvalitativnih raziskavah in raziskavah z integracijo metod ter 3) potreba po razvoju novih specialnih kriterijev ocenjevanja veljavnosti v raziskavah z integracijo metod. Rezultati druge raziskovalne faze so izpostavili mnenje ekspertov o izrednem pomenu veljavnosti kot temeljnem konceptu kakovosti vsake raziskave ne glede na metodološko orientacijo. Analiza je izločila deset temeljnih kriterijev veljavnosti, ki bi jih morala upoštevati vsaka raziskava z integracijo metod. Eksperti so se strinjali, da je razvoj merskega instrumenta za ocenjevanje veljavnosti v raziskavah z integracijo metod možen, pri čemer je ključno presoditi o številu in vsebini vključenih kriterijev, njihovi prioritetenosti ter sami obliki instrumenta, da bo širše dosegljiv za različne ocenjevalce in namene.

Rezultati vseh treh testiranj pilotnega merskega instrumenta so pokazali, da so testiranci najvišje povprečne ocene dodelili primernosti ocenjevanja veljavnosti v kvantitativnem delu raziskave. Rezultati testiranja revidiranega instrumenta pa se izkazujejo z naborom 10 soglasno sprejetih temeljnih kriterijev ocenjevanja veljavnosti, in sicer: 1) utemeljitev raziskovalnega vprašanja, 2) utemeljitev uporabe integracije metod, 3) raziskovalni načrt, 4) integracija, 5) utemeljitev metodologije, 6) pristop treh stebrov k ocenjevanju veljavnosti, 7) utemeljitev ugotovitev raziskave, 8) družbena občutljivost, 9) upoštevanje razlik, 10)

prepričljivost. Testirani merski instrument za ocenjevanje veljavnosti raziskave z integracijo metod vključuje kriterije z opisi in tristopenjsko ocenjevalno lestvico: 1 – kriterij ni prisoten, 2 – kriterij je delno prisoten, 3 – kriterij je prisoten.

Potrebne so nadaljnje raziskave za izračun kvantitativne mere oz. indeksa veljavnosti raziskave z integracijo metod, testiranjem na reprezentativnih vzorcih in analizo doprinosa posameznih kriterijev. Dobljene ugotovitve ponujajo raziskovalcem praktično metodološko orodje za zasnovo kakovostne raziskave. Visokošolskim institucijam pa dajejo izhodišča za razvoj metodoloških študijskih programov in predmetov ter posledično dvig vsesplošne kakovosti raziskovanja posameznih znanstvenih disciplin.

Literatura

- Bryman, Alan. 2008. Why do Researchers Integrate/Combine/Mesh/Blend/Mix/Merge/Fuse Quantitative and Qualitative Research? V *Advances in Mixed Methods Research: Theories and Applications*, ur. Manfred Max Bergman, 87–100. London, Velika Britanija: Sage Publications Ltd.
- Bryman, Alan, Saul Becker in Joe Sempik. 2008. Quality criteria for quantitative, qualitative and mixed methods research: a view from social policy. *International Journal of Social Research Methodology* 11 (4): 261–276.
- Ferligoj, Anuška, Karmen Leskošek in Tina Kogovšek. 1995. *Zanesljivost in veljavnost merjenja*. Ljubljana: Fakulteta za družbene vede, Metodološki zvezki 11.
- Guba, Egon G. in Yvonna S. Lincoln. 2005. Paradigmatic controversies, contradictions, and emerging confluences. V *The Sage Handbook of Qualitative Research (Third Edition)*, ur. Norman K. Denzin in Yvonna S. Lincoln, 191–215. London, Velika Britanija: Sage Publications Ltd.
- Johnson, Burke R. in Anthony J. Onwuegbuzie. 2004. Mixed methods research: a research paradigm whose time has come. *Educational Researcher* 33 (7): 14–26.
- Kogovšek, Tina. 1998. *Kvaliteta podatkov v kvalitativnem raziskovanju*. Magistrsko delo, Ljubljana: Univerza v Ljubljani, Fakulteta za družbene vede.
- Mixed Methods International Research Association (MMIRA). 2014. *Mixed Methods International Research Association Inaugural Conference: Challenges, Advances, and Diversity*, 27–29 junij 2014. Boston, Massachusetts: Boston College. Dostopno: <https://mmira.wildapricot.org/page-1859599> (25. 8. 2017).
- Neuman, Lawrence W. 2011. *Social research methods: qualitative and quantitative approaches (Sixth Edition)*. Boston, Massachusetts: Pearson Education, Inc.
- O'Cathain, Alicia. 2010. Assessing the quality of mixed methods research: toward a comprehensive framework. V *Handbook of Mixed Methods Research in Social & Behavioral Research (Second Edition)*, ur. Abbas Tashakkori in Charles Teddlie, 531–555. Thousand Oaks, California, ZDA: Sage Publications, Inc.
- Onwuegbuzie, Anthony J., Rebecca M. Bustamante in Judith A. Nelson. 2010. Mixed research as a toll for developing quantitative instruments. *Journal of Mixed Methods Research* 4 (1): 56–78.

- Sale, Joanna E.M. in Kevin Brazil. 2004. A strategy to identify critical appraisal criteria for primary mixed-method studies. *Quality & Quantity* 38 (4): 351–365.
- Silverman, David. 2005. *Doing Qualitative Research: A Practical Handbook (Second Edition)*. Thousand Oaks, California, ZDA: Sage Publications, Inc.
- Teddle, Charles in Abbas Tashakkori. 2009. *Foundations of Mixed Methods Research: Integrating Quantitative and Qualitative Approaches in the Social and Behavioral Sciences*. Thousand Oaks, California, ZDA: Sage Publications, Inc.

RAČUNI NACIONALNIH TRANSFERJEV ZA ANALIZO EKONOMSKE ODVISNOSTI SKOZI ŽIVLJENJSKI CIKEL

Dr. Tanja Istenič
Mentor izr. prof. dr. Jože Sambt
Univerza v Ljubljani, Ekonomska fakulteta

Uvod

V svojem življenju se posamezniki soočajo z dvema obdobjema ekonomske odvisnosti, ko njihova potrošnja presega njihov delovni dohodek – to je v času otroštva in v času starosti. Ekonomska odvisnost mladih in starih je financirana s strani delovno aktivne populacije, ki s svojim delom proizvaja več kot troši (Lee & Mason, 2011). V ekonomski analizi je obdobje ekonomske odvisnosti običajno definirano z uporabo fiksnih starostnih mej, neodvisnih od države, leta analize, spola itd. Poleg tega se običajno predpostavlja enaka velikost ekonomske odvisnosti, ne glede na to, koliko so odvisni posamezniki stari. Metodološki okvir računov nacionalnih transferjev (angl. National Transfer Accounts – NTA), ki ga uporabljam čez celotno disertacijo, nam s pomočjo koncepta primanjkljaja življenjskega cikla omogoča natančno merjenje obdobja in velikosti ekonomske odvisnosti. Z vpeljavo dimenzije starosti v Sistem nacionalnih računov (SNA) nam NTA omogoča ocenjevanje starostnospecifičnih vrednosti ekonomskih kategorij, kot so delovni dohodek, dohodek iz premoženja, transferji, potrošnja in varčevanje. Z uporabo starostnospecifičnih vrednosti ekonomskih kategorij nam NTA omogoča tudi celovito ocenjevanje virov, s katerimi je ekonomska odvisnost financirana.

Namen doktorske disertacije je celostno ocenjevanje in analiziranje posameznikove ekonomske odvisnosti skozi življenjski cikel ter medgeneracijskih tokov, s katerimi je ekonomska odvisnost financirana. Za še bolj poglobljen uvid v posameznikovo ekonomsko odvisnost sta v običajno analizo NTA dodani časovna dimenzija ter dimenzija spola. V prvem poglavju doktorske disertacije prvič oblikujem retrospektivne NTA za Slovenijo v zadnjih 30-ih letih in s tem pokažem, kako se delovni dohodek, potrošnja in ekonomska odvisnost spreminjajo v času tranzicije iz socializma v tržno gospodarstvo. Za zadnjih 20 let pa oblikujem še celovite rezultate NTA in s tem prikažem evolucijo medgeneracijskih transferjev skozi čas ter spreminjajoči se pomen posameznih virov financiranja ekonomske odvisnosti. V drugem

poglavju definiram in prvič predstavim razlike v ekonomski odvisnosti med spoloma skozi čas. V tretjem poglavju doktorske disertacije pripravim tudi primerljive rezultate NTA po spolu za deset držav Evropske unije v letu 2010. Z namenom odprave spolne pristranskosti običajnih NTA v analizo dodam še vrednosti neplačanega gospodinjanskega dela in na ta način običajne NTA združim z računi nacionalnih transferjev časa (angl. National Time Transfer Accounts – NTTA). Na podlagi primerljivih in celovitih rezultatov javnih in zasebnih medgeneracijskih tokov predstavim vrednosti petih kazalnikov, ki zajemajo glavne značilnosti režimov blaginje. Na ta način prispevam tudi k obstoječemu znanju povezave med režimi blaginje in režimi starostne prerazdelitve.

Podatki in raziskovalne metode

Metodologija NTA temelji na proračunski omejitvi posameznika, kjer se morajo prilivi izenačiti z odlivi. Med prilive štejemo delovni dohodek (τ^+), dohodek iz premoženja (Y^A) in prilive transferjev (τ^+); med odlive pa potrošnjo (C), odlive transferjev (τ^-) in varčevanje (S). S preureditvijo členov dobimo naslednjo računovodsko identiteto (Lee & Mason, 2011):

$$\underbrace{C(x) - Y^l(x)}_{\text{Primanjkljaj življenjskega cikla}} = \underbrace{\tau^+(x) - \tau^-(x)}_{\text{Neto transferji}} + \underbrace{Y^A(x) - S(x)}_{\text{Prerazdelitev iz naslova sredstev}} \quad (1)$$

Otroci in starejši se soočajo s pozitivnim primanjkljajem življenjskega cikla (definiran kot razlika med javno in zasebno potrošnjo ter delovnim dohodkom). Primanjkljaj življenjskega cikla določa njihovo ekonomsko odvisnost. Na drugi strani, v času delovne aktivnosti, delovni dohodek presega potrošnjo, kar vodi v negativen primanjkljaj življenjskega cikla, tj. presežek življenjskega cikla. Presežek življenjskega cikla omogoča zaposlenim financiranje primanjkljaja življenjskega cikla odvisnega prebivalstva. Ekonomska odvisnost se financira z različnimi oblikami medgeneracijskih tokov: neto transferji (prilivi minus odlivi transferjev) in prerazdelitvijo iz naslova sredstev (dohodek iz premoženja minus varčevanje). Računovodska identiteta velja za vsakega posameznika določene starosti (x) in tudi za nacionalno gospodarstvo kot celoto.

Metodologija NTA celovito analizira javne in zasebne transferje ter javno in zasebno prerazdelitev iz naslova sredstev. Javni prilivi transferjev so na primer javno financirano izobraževanje, javno zdravstvo in pokojnine, medtem ko odlivi predstavljajo davke in socialne prispevke. Zasebni transferji vključujejo transferje med gospodinjstvi in transferje znotraj gospodinjstva. Javna in zasebna prerazdelitev iz naslova sredstev izhaja iz udeležbe na kapitalskih in finančnih trgih (vključuje obresti na kredit, donos na kapital idr.). Eden glavnih prispevkov metodologije NTA je možnost oceniti velikost zasebnih transferjev (Lee & Mason, 2011). Ker je preučevana enota posameznik, predstavljajo javne in zasebne institucije zgolj posrednike med posamezniki.

Z namenom analize ekonomske odvisnosti in ekonomskih tokov, s katerimi je odvisnost financirana, moram oceniti številne starostne profile. Starostni profili so starostni vzorci (povprečja) nadalje razčlenjenih spremenljivk, vključenih v računovodsko identiteto. V prvem

koraku s pomočjo SNA in drugih podobnih podatkovnih virov izračunam agregatne vrednosti posameznih spremenljivk. V drugem koraku izračunam relativno porazdelitev določene spremenljivke po starostnih skupinah s pomočjo anketnih in/ali administrativnih podatkov. Starostne profile iz drugega koraka pomnožim z ustreznim faktorjem prilagajanja (angl. adjustment factor), tako da se starostni profili, pomnoženi s starostno strukturo prebivalstva, ujemajo z agregatnimi vrednostmi iz prvega koraka (United Nations, 2013). Rezultati NTA so torej presečni rezultati, pridobljeni s pomočjo sekundarnih virov podatkov. Za ocenjevanje retrospektivnega NTA za Slovenijo, predstavljenega v prvem poglavju doktorske disertacije, so uporabljeni mikropodatki iz Ankete o porabi gospodinjestev za leta 1983, 1988, 1993, 1998, 2000, 2003, 2005, 2008, 2010 in 2012.

NTA po spolu so ocenjeni podobno kot običajni NTA. Kadar so starostni profili izračunani na podlagi anketnih podatkov, je edina razlika, da moramo poleg povprečij po starostnih skupinah izračunati tudi povprečja po spolu. Starostni profili so prilagojeni na način, da se zagotovi skladnost z makroekonomskimi agregati in tudi s starostnimi profili običajnega NTA (Donehower, 2014). Z namenom popolne primerljivosti podatkov, potrebnih za ocenitev NTA po spolu v Sloveniji, je časovno obdobje skrajšano na obdobje med letoma 2000 in 2012. Poleg grafične predstavitev razlik v potrošnji, delovnem dohodku in primanjkljaju življenjskega cikla drugo poglavje doktorske disertacije prikazuje tudi rezultate dveh kazalnikov, ki kažeta razlike v ekonomski odvisnosti med spoloma in sta bila razvita v okviru te doktorske disertacije.

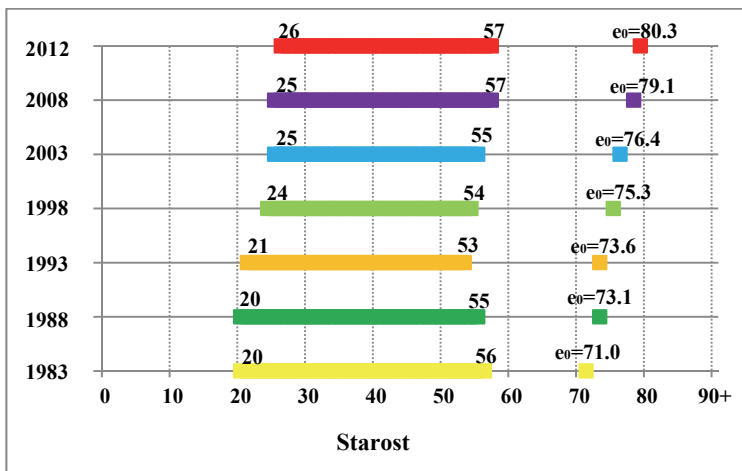
Za ocenjevanje starostnih profilov dohodka v državah EU iz tretjega poglavja doktorske disertacije so uporabljeni mikropodatki »Statistike EU o dohodkih in življenjskih pogojih« (angl. EU Statistics on Income and Living Conditions – EU-SILC). Usklajeni anketni podatki za oblikovanje starostnih profilov zasebne potrošnje na ravni EU v času pisanja tretjega poglavja doktorske disertacije še niso bili na voljo. Zato sem starostne profile zasebne potrošnje ocenila na podlagi najnovejših razpoložljivih relativnih starostnih profilov nacionalnih skupin NTA, prilagojenih na dejansko agregatno vrednost iz leta 2010. Uporabljen način bi moral biti sprejemljiv, saj se vzorec zasebne potrošnje skozi čas bistveno ne spreminja (npr. Hammer, 2014). Rezultati NTTA so izračunani na podlagi ankete o porabi časa, ki nam omogoča določiti čas, namenjen za proizvodnjo gospodinjskih aktivnosti po starosti in spolu. Rezultati NTTA iz tretjega poglavja doktorske disertacije so pridobljeni iz podatkovne baze projekta AGENTA (Vargha et al., 2016).

V tretjem poglavju doktorske disertacije je najprej opisno predstavljena povezava med režimi blaginje in režimi starostne prerazdelitve za prototipe oz. predstavnike posameznih režimov blaginje. Poleg tega poglavje prikazuje vrednosti petih različnih kazalnikov NTA NTTA za vseh deset analiziranih držav EU, ki določajo glavne značilnosti posameznih režimov blaginje. Na podlagi vrednosti kazalnikov so države razporejene v režime starostne prerazdelitve in primerjane s skupinami držav, ki pripadajo določenemu režimu blaginje.

Rezultati

Analiza NTA za Slovenijo med letoma 1983 in 2012 kaže, da ima delovni dohodek v vseh letih tipično obliko obrnjene U-krivulje, medtem ko je potrošnja, z izjemo vrha v času izobraževanja mladih, razmeroma konstantna pri vseh letih starosti. Rezultati kažejo, da mladi skozi čas ostajajo daljše obdobje odvisni od javnih in zasebnih transferjev. Na drugi strani so se starejši sposobni sami financirati dlje časa, vendar je zadnje manj očitno. V zadnjih treh desetletjih se je tako obdobje ekonomske neodvisnosti posameznikov skrajšalo iz 36–37 let v času socializma na 31–33 let v času tržnega gospodarstva. Kljub temu da se je v obdobju 1983–2012 življenjsko pričakovanje ob rojstvu podaljšalo za 9,3 leta, se je doba ekonomske neodvisnosti skrajšala za 5 let (glej Sliko 1).

Slika 1: Razpon starosti, v kateri se v Sloveniji sami vzdržujemo, in pričakovana življenjska doba ob rojstvu (e_0) v izbranih letih 1983–2012

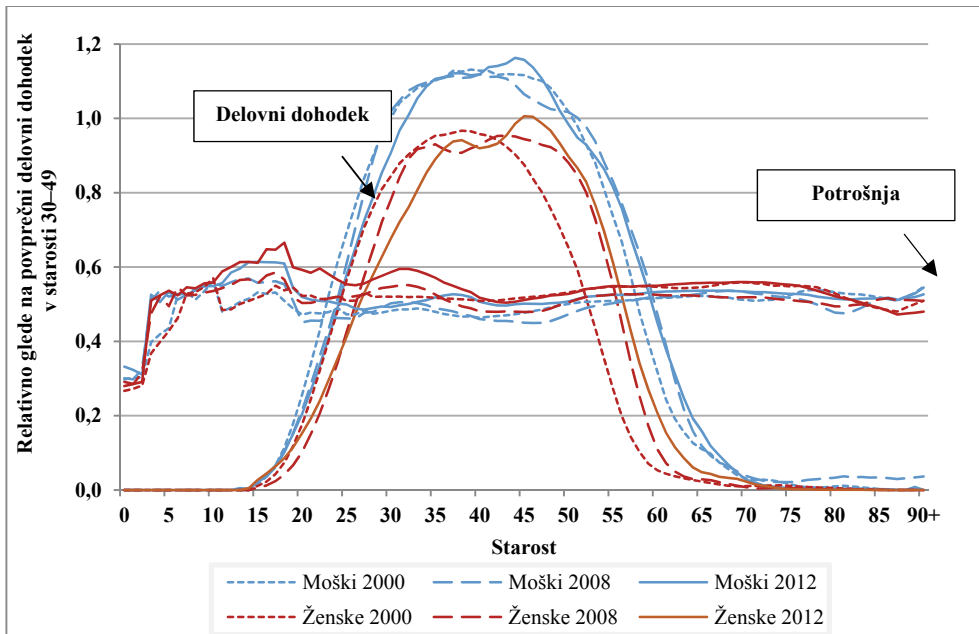


Vir: UMAR (2017). Strategija dolgožive družbe (po Istenič in Sambt, 2017).

Rezultati kažejo tudi, da je ekonomska odvisnost mladih pretežno financirana z zasebnimi transferji (v višini 60 % celotne odvisnosti). Vendar pa se skozi čas pomen zasebnih transferjev zmanjšuje, medtem ko se pomen javnih transferjev za financiranje ekonomske odvisnosti mladih povečuje. Medtem ko se mladi še vedno pretežno financirajo z zasebnimi transferji, je ekonomska odvisnost starejših pretežno financirana z javnimi transferji (v višini 80–90 % celotne odvisnosti). Skozi čas je tudi odvisnost starejših v višji meri financirana z javnimi transferji, kar skupaj s hitrim staranjem prebivalstva še dodatno ogroža vzdržnost javnofinančnega sistema.

V drugem poglavju predstavim rezultate retrospektivnih NTA po spolu v Sloveniji v obdobju 2000–2012. Rezultati kažejo, da je razlika med spoloma v delovnem dohodku prisotna pri vseh starostih, ko ženske zaslužijo manj kot moški, medtem ko je razlika v potrošnji očitna predvsem v času materinstva, ko ženske porabijo več kot moški (glej Sliko 2).

Slika 2: Delovni dohodek in potrošnja moških in žensk, Slovenija, 2000–2012



Vir: Istenič, Ograjenšek, & Sambt, 2018.

Zaradi nižjega delovnega dohodka, a hkrati višje potrošnje žensk v primerjavi z moškimi, so ženske ekonomsko neodvisne krajše obdobje življenjskega cikla kakor moški. V letu 2012 so bile ženske sposobne financirati svojo potrošnjo za obdobje 29,0 let, medtem ko moški 35,9 let. Vendar pa se v času razlika v dolžini ekonomske neodvisnosti skrajšuje, iz 7,9 let v letu 2000 na 6,9 let v letu 2012. Prav tako se skozi čas zmanjšuje tudi razlika med spoloma v velikosti ekonomske odvisnosti. Rezultati kažejo tudi, da je ekonomska kriza prekinila približevanje obdobja ekonomske neodvisnosti in velikosti ekonomske odvisnosti žensk k vrednostim za moške, a so se razlike med spoloma tudi v celotnem obdobju krize zmanjšale.

Na podlagi rezultatov NTA in NTTA po spolu je v tretjem poglavju doktorske disertacije definiranih pet kazalnikov, ki zajemajo glavne značilnosti režimov blaginje in omogočajo preverbo povezave med režimi blaginje in režimi starostne prerazdelitve v desetih državah EU za leto 2010. Prvi trije kazalniki kažejo pomen javnih transferjev v primerjavi z ostalimi viri financiranja odvisnosti mladih in starih. Rezultati kažejo, da so mladi v vseh desetih državah EU odvisni predvsem od zasebnih transferjev. Zasebni transferji so še posebej pomembni v mediteranskih državah, medtem ko je njihov pomen v socialdemokratskih državah manjši. Ekonomska odvisnost starih je financirana pretežno z javnimi transferji, z izjemo Nemčije in Velike Britanije, kjer je odvisnost starejših v večji meri financirana s prerazdelitvijo iz naslova sredstev. Četrti in peti kazalnik merita razlike med spoloma v velikosti ekonomske neodvisnosti. Medtem ko četrti kazalnik meri razlike med proizvodnjo in potrošnjo iz naslova plačanega dela, peti kazalnik vključuje vrednosti plačanega in neplačanega dela. V tretjem poglavju

doktorske disertacije trdim, da je vključitev dimenzije spola ključnega pomena, ko iščemo povezavo med režimi blaginje in režimi starostne prerazdelitve. Z vključitvijo rezultatov NTA in NTTA po spolu se namreč mesto določenih držav precej spremeni. Na koncu oblikujem tri skupine držav: socialdemokratske (Švedska, Madžarska in Slovenija), konservativne (Francija, Finska, Avstrija, Italija in Španija) in liberalne (Velika Britanija in Nemčija). Kot glavno izjemo prepoznam Nemčijo, ki na podlagi kazalnikov NTA-NTTA spada v liberalni in ne konservativni režim.

Viri in literatura

1. Donehower, G. (2014). Incorporating gender and time use into NTA: National Time Transfer Accounts methodology (verzija 4, maj 2014). *Interno gradivo projekta NTA*.
2. Hammer, B. (2014). *The Economic Life Course: An Examination Using National Transfer Accounts*. Doktorska disertacija, Vienna University of Technology, Dunaj, Avstrija.
3. Istenič, T., Ograjenšek, I., & Sambt, J. (2018). The gender gap in economic dependency over the life cycle: Some theoretical and practical considerations. *Economic Research–Ekonomška Istraživanja*, 31(1), 188–205.
4. Lee, R., & Mason, A. (2011). *Population aging and the generational economy: A global perspective*. Cheltenham, UK; Northampton, MA: Edward Elgar.
5. UMAR. (2017). *Strategija dolgožive družbe*. Ljubljana: UMAR.
6. United Nations. (2013). *National Transfer Accounts Manual: Measuring and Analysing the Generational Economy*. New York: United Nations.
7. Vargha, L., Šeme, A., Gál, R. I., Hammer, B., & Sambt, J. (2016). *Manual of the NTTA methodology and guidelines to the AGENTA NTTA data explorer. AGENTA project report*. Najdeno 15. septembra 2018 na spletni strani <http://www.agenta-project.eu/Jacomo/upload/publications/d-2.3-submitted.pdf>.

RAZISKUJEMO

UPORABA MIKRO PODATKOV ZA RAZVOJ MIKROSIMULACIJSKIH MODELOV

Dr. Boris Majcen in dr. Nataša Kump
Inštitut za ekonomska raziskovanja

Inštitut za ekonomska raziskovanja ima že več kot petnajstletno zgodovino v razvoju mikrosimulacijskih modelov. S pojmom mikrosimulacija razumemo tehniko, ki jo uporabljamo za modeliranje kompleksnih dogodkov iz resničnega življenja s simuliranjem dejanj posameznikov in/ali učinkov določenih sprememb na posameznike, ki sestavljajo sistem, kjer do dogodka pride. Mikrosimulacijski model vedno začne z reprezentativno začetno populacijo v določenem trenutku (presečno populacijo), ki jo nato spreminja v času skladno s pričakovanim prihodnjim razvojem in politikami.

Mikrosimulacijske modele lahko razdelimo na statične in dinamične mikrosimulacijske modele. Statični mikrosimulacijski modeli ne vključujejo časovne komponente in vedenjskih odzivov ter so uporabno orodje za oceno neposrednih učinkov reform, za določanje neposrednih »zmagovalcev in poražencev« določenega reformnega ukrepa (običajno na področju davkov in socialnih transferjev) ter za ocenjevanje učinka ukrepov na proračunske prihodke oziroma odhodke. Statični mikrosimulacijski modeli so dejansko aritmetični kalkulatorji, kjer se znotraj modela simulirajo spremembe v razpoložljivem dohodku posameznikov ali gospodinjstev zaradi sprememb v plačilu davkov ali prejemanja socialnih prejemkov ob predpostavki, da se vedenje posameznikov ne spremeni. Dinamično mikrosimulacijo v družbenih znanostih na drugi strani lahko dojemamo kot laboratorij, ki omogoča pripravo eksperimentov z virtualno družbo tisočih/milijonov posameznikov, ki so ustvarjeni in katerih celoten potek življenja je razvit v računalniku. Posamezniki se odločajo o izobrazbi in karieri, se poročajo, rojevajo, se zaposlijo, plačujejo davke in pridobivajo socialne transferje, se ločujejo, migrirajo, se upokojijo, dobijo pokojnino in umrejo. Mikrosimulacija nam tako lahko pomaga razumeti učinke posameznih procesov in spremembe v procesih na končni rezultat. Večje ko je število medsebojno povezanih procesov, težje je razumeti vpliv posameznega dejavnika na končni rezultat.

Na Inštitutu za ekonomska raziskovanja smo od leta 2001 naprej v razvili več mikrosimulacijskih modelov:

- v sodelovanju z Ekonomsko fakulteto v Ljubljani je bila v obdobju med 2001 in 2004 razvita prva različica statičnega mikrosimulacijskega modela za Slovenijo, ki je temeljila APG podatkih,
- statični mikrosimulacijski model davkov in socialnih prejemkov (dve verziji modela, prva končana 2006 in druga 2010),
- EUROMOD evropski statični mikrosimulacijski model davkov in socialnih prejemkov (sistemi od leta 2006 do leta 2017),

- dinamični pokojninski mikrosimulacijski model s statičnim staranjem in
- dinamični pokojninski mikrosimulacijski model DYPENSI.

Predpogoj za možno izdelavo mikrosimulacijskega modela so primerni mikro podatki. Prvo izgradnjo je izgradnjo mikrosimulacijskega modela je omogočila baza mikro podatkov, ki jo je za leto 2004 pripravil SURS in je obsegala:

- demografske podatke o posameznikih: spol, starost, občina prebivališča,
- podatke o prejemkih, ki jih izplačuje ZPIZ,
- podatke, ki omogočajo identificiranje študentov, ter podrobnejše podatke o štipendistih (letnik študija, leto prejemanja štipendije, občina šole, višina in vrsta štipendije, itd.),
- podatke iz Statističnega registra delovno aktivnega prebivalstva in
- podatke Zavoda Republike Slovenije za zaposlovanje.

Ta baza podatkov je služila za prvo kompleksnejšo verzijo statičnega mikrosimulacijskega modela davkov in socialnih prejemkov. S tem modelom, v povezavi z modelom splošnega ravnovesja, smo na primer ocenjevali učinke uvedbe enotne davčne stopnje, ki je bila aktualna v letu 2006.

Podobno bazo podatkov je SURS pripravil tudi za leto 2007. Ta baza je vključevala zgoraj našete podatke za 112.839 oseb v 40.000 gospodinjstvih in je služila kot podatkovna osnova za drugo verzijo statičnega mikrosimulacijskega modela davkov in socialnih prejemkov, ki v svoji osnovni različici simulira sistem socialnih prejemkov, dohodnino in prispevke za socialno varnost v letu 2010. Z modelom smo ocenjevali učinke posameznih idejnih sprememb zakonodaje, ki ureja socialne prejemke, ravno v času, ko sta se pripravljala Zakon o uveljavljanju pravic iz javnih sredstev in Zakon o socialno varstvenih prejemkih.

Mikrosimulacijski model smo uporabili za iskanje odgovorov na vprašanja kot so na primer:

- Kdo bo na boljšem in kdo na slabšem ob predvideni spremembi?
- Kako bo predvidena sprememba vplivala na dohodkovno porazdelitev v družbi? Kako na revščino?
- Kako bo predvidena sprememba vplivala na agregatna potrebna sredstva, ki jih bo morala država zagotoviti za posamezne socialne transferje?

Ker se pravico do posameznih socialnih prejemkov uveljavlja v točno določenem vrstnem redu, se v dohodek, ki se upošteva za presojanje pravice do posameznega socialnega prejemka, upoštevajo prejeti socialni prejemki, ki se v vrstnem redu nahajajo pred tem. To pomeni, da sprememba pri posameznem socialnem prejemku ne spreminja le višine tega socialnega prejemka, ampak vpliva tudi na upravičenost oz. višino ostalih socialnih prejemkov, ki v vrstnem redu sledijo. Učinke spremenjene zakonodaje v takšnem sistemu je mogoče oceniti le z mikrosimulacijskim modelom.

Trenutni nacionalni statični mikrosimulacijski modela davkov in socialnih prejemkov je zastarel, saj do leta 2018 ni bilo nobene posodobitve. Kot statični model davkov in socialnih prejemkov pa se za Slovenijo lahko uporablja model EUROMOD, kjer je modeliranih nekaj manj socialnih prejemkov in subvencij kot v nacionalni različici. Inštitut za ekonomska raziskovanja je od leta 2004 naprej vključen v projekt »EUROMOD«, ki ga vodi University of Essex in je financiran z

evropskimi sredstvi. EUROMOD omogoča simulacijo davčnih sistemov in sistemov socialnih prejemkov v daljšem časovnem obdobju za države EU. Za Slovenijo je modeliran sistem davkov in socialnih prejemkov za obdobje 2005-2017.

Razvoj dinamičnih mikrosimulacijskih modelov je bila naslednja stopnja v razvoju tega področja, saj smo hoteli odgovoriti na bolj dolgoročna vprašanja, kot so na primer:

- Kakšne bodo pokojnine za prihodnje generacije in kako vzdržen bo slovenski pokojninski sistem?
- Kakšno bo razmerje med aktivno in neaktivno populacijo v prihodnje?
- Kako lahko obdržimo solidarnost glede na naraščajočo neenakost med generacijam in znotraj njih?

Najprej smo razvili dinamični pokojninski mikrosimulacijski model s statičnim staranjem, nato pa smo razvili prvo različico pravega dinamičnega mikrosimulacijskega pokojninskega modela DYPENSI. Za oba modela uporabljamo že omenjeno bazo mikro podatkov iz leta 2007. Oba modela sta razvita z namenom ocenjevanja pokojninskega sistema, vendar so med njima nekatere pomembne razlike. Pri statičnem staranju s spreminjanjem uteži sedanjo populacijo »spremenimo« v bodočo. Starajočo družbo na ta način simuliramo tako, da se uteži mlajših oseb postopoma zmanjšujejo, uteži starejših oseb pa se postopoma povečujejo. Zaradi svojega (večinoma) statičnega značaja tak model ne omogoča celovite analize prihodnjih pokojninskih reform. Zato je bil razvoj prvega polnega dinamičnega mikrosimulacijskega modela v Sloveniji (DYPENSI) naravna poteza in prehod na bolj zapletena modelska orodja. V pravem dinamičnem mikrosimulacijskem modelu se posamezniki soočajo z določenimi verjetnostmi za spremembo posameznih lastnosti ali stanj in tako do teh sprememb prihaja v samem procesu simulacije. Modelirani so demografski procesi (rojstva, smrtnost, migracije, partnerstva), procesi na trgu dela (vstop na trg dela, prehod iz zaposlenosti v brezposelnost in obratno, prehod iz aktivnosti v neaktivnost in obratno) ter na koncu tudi upokojevanje (invalidske in družinske ter vdovske oziroma delne vdovske pokojnine, ki se izračunavajo na podlagi delovne in plačne zgodovine posameznikov). Na ta način se izgradijo sintetične zgodovine za vsakega posameznika v modelu. Modela sta bila uporabljena za pripravo poročila o pokojninah Slovenia Country Fiche (Pension Expenditures and Revenues: Projections and Models), ki se redno pripravlja za potrebe AWG skupine pri Evropski komisiji. V letu 2016 sta bila uporabljena tudi za izračune številnih scenarijev izdelanih za Belo knjigo o pokojninah.

Tako kot fiziki potrebujejo raziskovalno infrastrukturo CERN da bi lahko razumeli fiziko delcev, raziskovalci na področju družboslovja potrebujemo raziskovalno infrastrukturo v obliki programskih orodij in urejenih baz podatkov, ki predstavljajo, morda za mnoge čuden/specifičen, a vendarle še kako uporaben raziskovalni laboratorij. Ta laboratorij pa je seveda treba redno vzdrževati, posodabljati in razvijati. V okviru raziskovalnega dela na področju proučevanja in razvoja ustreznih metodologij in orodij bomo nadaljevali z delom na vzdrževanju, razvoju kot tudi povezovanju posameznih modelskih orodij. Hkrati pa načrtujemo širjenje sedanje družine mikrosimulacijskih modelov na področja dolgotrajne oskrbe, zdravstva in izobraževanja, podjetij in trga dela; pomemben korak bo predstavljalo tudi nadaljevanje uvajanja dinamike v sicer statične modele in njihovo povezovanje z makro modelom splošnega ravnovesja slovenskega gospodarstva.

VPLIV VOLITEV IN DRUGIH POMEMBNEJŠIH DOGODKOV NA ZAUPANJE POTROŠNIKOV V SLOVENIJI

THE EFFECT OF ELECTIONS AND OTHER MAJOR EVENTS ON CONSUMER CONFIDENCE IN SLOVENIA

Kratka vizualna analiza

Luka Zupanc
Statistični urad RS

Povzetek

Z vizualno analizo kazalnika zaupanja potrošnikov smo skušali ugotoviti, ali pomembnejši politični dogodki v Sloveniji vplivajo na gibanje omenjenega kazalnika. Pregledali smo časovno vrsto tega kazalnika od januarja 2005 do oktobra 2017. Pri analizi je bil zaznan kratkotrajen pozitiven vpliv parlamentarnih volitev na razpoloženje potrošnikov, saj je v mesecu volitev, v manjši meri pa tudi v predhodnem mesecu kazalnik zaupanja potrošnikov naraščal. V mesecu po volitvah se je njegova vrednost spet znižala. Podobno se je zaupanje potrošnikov okrepilo tudi v obdobju večjih protestov proti politični korupciji konec leta 2012 in na začetku leta 2013. Med migrantsko krizo konec 2015 in na začetku 2016 pa se je razpoloženje poslabšalo. Omenjeni dogodki vplivajo le na razpoloženje potrošnikov, ne pa tudi na njihovo dejansko potrošnjo.

Sledi predstavitev v angleškem jeziku.

Based on the special topic in the publication European Business Cycle Indicators for the 3rd quarter of 2017 and the study done by the INSEE in March 2017, a brief visual analysis of the consumer confidence time series was done to see if elections and some other major events have an impact on the consumer confidence in Slovenia.

After a visual inspection of the time series of the consumer confidence indicator (CCI) and its component series it is evident that there is an effect of parliamentary elections and some other major events. A seasonally adjusted time series of the CCI and its components from 2005 to October 2017 were taken into consideration. During this period there were three parliamentary elections in Slovenia. Additionally, there are also local and presidential elections in Slovenia, but these do not have such a big impact on the consumer confidence.

Chart 1: General economic situation in Slovenia over the next 12 months

Source: SURS

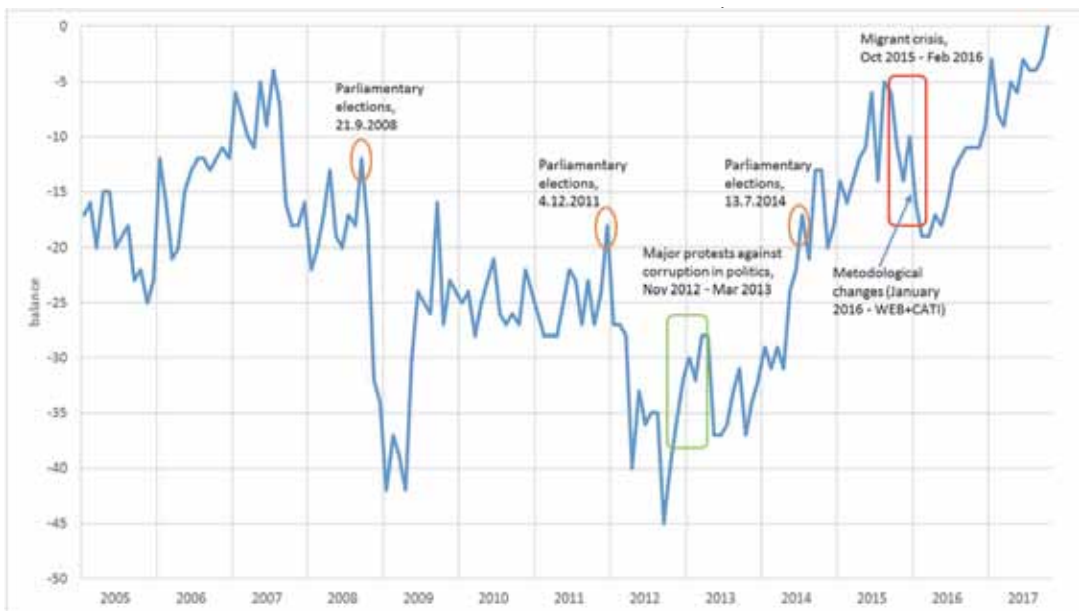
It is evident from Chart 1 that parliamentary elections have a positive and very short-lived effect on consumers' expectations about the country's economy. The positive effect is present in the election month and to a smaller extent also in the previous month. This is the period in which the election campaign dominates the news. In the month after the elections the confidence deteriorates. However, there are a few considerations to keep in mind. The first two elections (2008 and 2011) took place at the beginning of the two waves of the last economic recession in which the consumer opinion along with all other economic indicators dropped to historic low points. The 2014 election, however, was held in the period of economic recovery after the recession and the consumer confidence grew very rapidly.

Not only elections influence the consumer opinion in Slovenia. The effect of two other major events is also evident on the chart. At the end of 2012 and the beginning of 2013 there were major protests against corruption in politics throughout Slovenia. It all started in Maribor, the second largest town in Slovenia, with the protest against the mayor. At the beginning of 2013 the Commission for the Prevention of Corruption released the report accusing the prime minister of corruption. Also the leader of the main opposition party was accused of corruption. This led to mass protests mainly in Ljubljana, the capital of Slovenia, but also in other places across the country. This movement reached its peak in February 2013. The consumer confidence grew considerably in those months and after the end of this period dropped again almost to the same level as before the protests.

On the other hand, consumer expectations dropped at the end of 2015 and the beginning of 2016 when the migrant crisis was at its peak in Slovenia. This coincides, however, with the methodological changes in the area of sampling and data collection, when a combined data collection method (WEB+CATI) was implemented in the Consumer Survey in Slovenia in January 2016. There is a between-mode effect (different mode effect for CATI and WEB) or, in other words, the data collection method does affect the opinion. The opinions expressed on the WEB questionnaire are generally more pessimistic.

The elections and other major events described above also have a similar effect on the CCI as a whole (Chart 2) and on its other two components; expectations about households' finances and unemployment level. On the other hand, elections and other major events have a much weaker effect on the expectations about savings, which is also a CCI component. Although this is only a brief visual analysis, it shows that there is an effect of elections and other major events on consumer confidence in Slovenia, but of course a more detailed analysis would be necessary to get a clearer picture of these effects.

Chart 2: Consumer confidence indicator, Slovenia

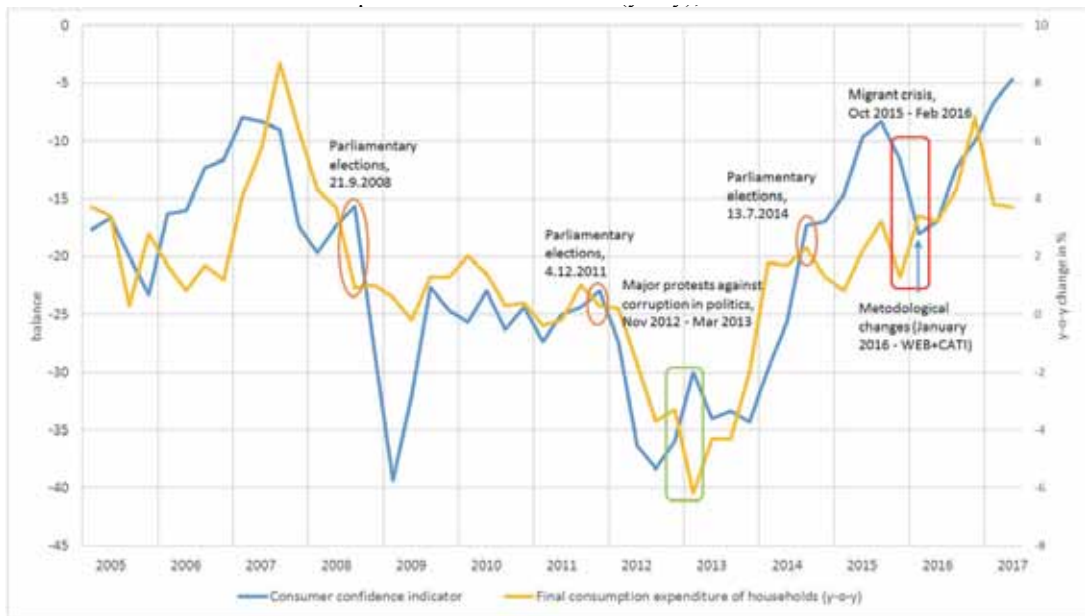


Source: SURS

The improved consumer opinion during the election months and major protests does not translate into increased household consumption. Chart 3 presents the CCI (converted to a quarterly basis) and final consumption expenditure of households in Slovenia. We can see that the peaks in consumer confidence during the election months are not coupled by the increased household consumption. During the time of the major protest against corruption in politics, household consumption even decreased considerably. In the first quarter of 2016,

during the migrant crisis, the household consumption increased, while the CCI fell quite considerably. However, at that time the methodological changes in the Consumer Survey were made and this can at least partially explain the discrepancy between the CCI and household consumption in the first quarter of 2016.

Chart 3: Consumer confidence indicator (converted to a quarterly basis) and final consumption expenditure of households (y-o-y), Slovenia



Source: SURS

References

- French National Institute of Statistics and Economic Studies (2017). 'Electoral periods have a positive albeit short-lived effect on household confidence'. French National Institute of Statistics and Economic Studies, Conjoncture in France, March.
- Reuter, A. The effect of elections on consumer confidence in Europe. European Business Cycle Indicators – 3rd Quarter 2017, 18-26.

ZANIMIVO

NEMOČ ANKETNIH RAZISKAV

Mag. Katja Rutar
Statistični urad RS

Jeseni 2017 je Martin Thöring, anketar s petindvajsetletnimi izkušnjami in nekdanji lastnik dveh manjših podjetij za zbiranje podatkov, obiskal nemško državno tožilstvo in sam sebe prijavil kot ponarejevalca anket. Policija ga je potem večkrat zaslišala. Dokumenti o njem že nekaj mesecev nedotaknjeni čakajo na okrožnem državnem tožilstvu v Münstru.

Verjetno je bil to eden izmed razlogov, da so se novinarji nemškega tednika Der Spiegel poglobili v raziskovanje kakovosti podatkov, pridobljenih z anketnimi raziskavami in svoje ugotovitve objavili pod naslovom Die Akte Marktforschung na: www.spiegel.de/thema/die_akte_marktforschung. Organizirali so tudi televizijsko soočenje na temo anketnih raziskav. Informirajo nas, da problem ni nov. Leta 1994 je nek drugi anketar, Heiner Dorroch, objavil knjigo o svojih izkušnjah s ponarejanjem anket z naslovom Meinungsmacher-Report. Wie Umfrageergebnisse entstehen (Poročilo oblikovalca mnenj. Kako nastajajo anketni podatki), ki je takrat pretresla anketne raziskovalce (Der Spiegel 26/1994, str. 41). Ti so se večinoma strinjali, da je večina anketnih raziskav slabe kakovosti. A število le-teh in prihodki od njih so vztrajno naraščali. Ob tem so se izboljševale možnost organizacij, ki zbirajo podatke, za nadzor nad delom anketarjev. Iz večinoma osebnega anketiranja na papirju po terenu se zbiranje podatkov seli na računalnike in telefone. Hkrati pa postajajo vprašalniki vse daljši, raziskovalne teme vse bolj občutljive ali pa kar prosto nametane skupaj, ciljne osebe vse bolj specifične (npr. begunci), zneski, ki so jih naročniki pripravljene plačati za zbrane podatke, stalno nižji in stopnje odgovorov določene v pogodbah. Če sestavimo vse to skupaj, se ne smemo čuditi, da se pri zbiranju podatkov s pomočjo anket načrtno manipulira, povzamejo novinarji. Pri tem ne sodelujejo le anketarji, ampak tudi vodje projektov in nadzorniki. Večina poznavalcev anketnega raziskovanja to sluti, mnogo jih je informiranih o tem, vendar problema na uspejo rešiti. Kljub temu, da je vse več govora o kakovosti podatkov in da se ESOMAR-jev kodeks, ki to področje ureja, stalno posodablja.

Problem je resnejši, ko se z anketami zbira podatke za zdravstvene ali farmacevtske raziskave. Ali ko se meri odnos ljudi do kakih strateških političnih vprašanj in so na podlagi odgovorov oblikovani ukrepi. Velike finančne posledice za podjetja, stranke in zaposlene imajo lahko tudi podatki iz običajnih tržnih raziskav.

Med vpletenimi nemškimi tržnoraziskovalnimi podjetji, na katere se njihovi interni viri nanašajo, omenjajo ugledna velika podjetja, ki imajo podružnice tudi v Sloveniji. Pri njih so podatke naročale tudi že slovenske državne in raziskovalne ustanove ter Evropska unija.

Reakcije nemške javnosti na februarja 2018 objavljene zgodbe so bile različne. Predstavniki gospodarstva so ponarejanje anketnih podatkov označili kot kriminalno dejanje. Nekateri politiki so v parlamentu zahtevali, da če gre za kazniva dejanja, morejo ta biti kaznovana. S pravosodnega ministrstva so sporočili, da je zaskrbljujoče, da cenovni pritisk vpliva na kakovost podatkov in da naj potrošniki bolj sprašujejo, od kje podatki prihajajo. Predstavniki zveze potrošnikov je vodjem podjetij in politikom svetoval, naj sami sprejemajo odločitve in za njimi tudi stojijo, namesto da se skrivajo za številkami iz anketnih raziskav, saj da se neljube odločitve prepogosto neupravičeno pripiše rezultatom raziskav. Predstavniki tržnoraziskovalne organizacije Forsa pa je naročnikom svetoval, naj pri naročanju podatkov ne upoštevajo samo ponujene cene, ampak dajo pomen tudi kakovosti, ki je običajno premo sorazmerna s ceno.

Štiri dni po objavi prvih Spiegel-ovih zgodb so se srečali predstavniki štirih nemških državnih združenj tržnih in družboslovnih raziskovalcev (ADM, BVM, DGOF in ASI) in sestavili t.i. Weinheimsko izjavo o kakovosti in transparentnosti v tržnem in javnomnenjskem raziskovanju, ki se je dotaknila tudi nove evropske uredbe o varovanju osebnih podatkov. Četrta točka izjave poudarja osrednjo vlogo anketarjev v anketnem raziskovanju. Anketarji morajo imeti socialne in jezikovne kompetence ter dobro poznati vsebino raziskovanj. Organizacije, ki jih zaposlujejo, pa jih morajo ustrezno izbrati, izobraziti, informirati in plačati. Ostalih osem točk govori o strokovnosti, metodologiji, naročnikih raziskav, potrebnem času, varovanju podatkov, interpretaciji rezultatov, transparentnosti postopkov.

SKUPŠČINA

STATISTIČNO DRUŠTVO SLOVENIJE
Litostrojska cesta 54, LJUBLJANA

Z A P I S N I K

26. redne skupščine Statističnega društva Slovenije,

ki je potekala v ponedeljek, 27. marca 2017, od 14.00 do 15.30 ure
v sejni sobi Statističnega urada Republike Slovenije na Litostrojski cesti v Ljubljani.

Navzoči člani: Andrej Blejec, Matevž Bren, Simona Korenjak-Černe, Jerneja Čuk, Matjaž Erker, Ida Repovž Grabnar, Bogdan Grmek, Valentina Hlebec, Milena Jankovič, Nadja Svetlin Kastelic, Laura Šuštar Kožuh, Jaro Lajovic, Ema Mišič, Mojca Noč Razinger, Katja Rutar, Genovefa Ružič, Pavle Sicherl, Janez Stare, Tina Šijanec, Mojca Škrlec, Jana Vajda, Aleš Žiberna

1. Odprtje skupščine in izvolitev delovnega predsedstva, zapisnikarja in dveh overovateljev zapisnika
2. Predavanje: Big data na SURS-u (Črt Grahonja)
3. Poročilo predsednika in nadzornega odbora
4. Obravnava finančnega poročila za leto 2016 (Bogdan Grmek)
5. Poročilo o aktivnostih: Uporabna statistika 2016 (Andrej Blejec), Statistični dan 2016 (Genovefa Ružič), Mladi statistiki (Andrej Blejec), Metodološki zvezki (Valentina Hlebec), magistrski in doktorski študij (Janez Stare)
6. Razprava o poročilih in sprejetje sklepov skupščine
7. Razno.

K 1

Skupščino Statističnega društva Slovenije je s pozdravnim nagovorom odprl predsednik društva dr. Matevž Bren ter predstavil dnevni red 26. redne skupščine.

Za kandidate delovnega organa skupščine je predlagal:

- dr. Andreja Blejca za predsednika skupščine,
- Jernejo Čuk za zapisnikarico,
- Mojco Škrlec in Jano Vajda za overiteljici zapisnika.

Sklep: Navzoči so soglasno potrdili delovne organe skupščine.

K 2

V nadaljevanju je predsednik skupščine Andrej Blejec za govorniški oder povabil Črta Grahonjo z Oddelka za podporo izhodnih procesov in internetne tehnologije na Statističnem uradu RS (v nadaljevanju SURS). Črt nam je predstavil delo na področju t.i. Big Data na SURS-u. Najprej je izpostavil je, da še ni uradno določen slovenski termin za te vrste podatkov, zato se pri komunikaciji uporabljata tudi termina masivni podatki oziroma velepodatki. SURS pri pripravi statistik uporablja različne vire podatkov, poleg podatkov raziskovanj, še v čim večji meri podatke administrativnih in drugih obstoječih virov. Ker SURS nenehno strmi k bolj kakovostnim in hitreje objavljenim statistikam, kakor tudi čim večji razbremenitvi poročevalskih enot, sedaj preverja še pridobitev in uporabo različnih Big Data virov. Te podatke zaznamuje njihova velikost, nestrukturiranost in (ali) velika hitrost njihovega nastajanja. Tipični primeri teh podatkov so npr. podatki mobilne telefonije, skenirani podatki trgovskih verig, podatki pametnih števec in strgani podatki s spleta. Ker se ti podatki generirajo z veliko hitrostjo in so (potencialno) na voljo v referenčnem obdobju, za katerega nas zanimajo določene statistike, nam seveda omogočajo tudi zgodnejše oziroma pogostejše objave statistik. Tako se SURS v zadnjem obdobju ukvarja tudi z novimi metodami, ki nam omogočajo hitrejšo objavo statistik (tako imenovani nowcasting). Predstavljene so bili testni rezultati objave zgodnjih ekonomskih kazalnikov in drugih statistik na podlagi podatkov števec prometa in strganih podatkov spletnih portalov, ki oglašujejo prosta delovna mesta. Prav tako so bili predstavljeni rezultati EU granta, kjer smo testirali uporabnost podatkov mobilne telefonije.

SURS je zelo dejaven pri delu na mednarodnih projektih, ki testirajo uporabo masivnih podatkov, prav tako pa je član konzorcija, financiranega s strani Eurostata, ki pripravlja smernice za etično uporabo teh virov, smernica za izobraževanje statistikov za delo s temi podatki ter pregled pravih predpisov, ki nam (ne) omogočajo dostop do teh podatkov. Prav tako SURS pri raziskovanju masivnih podatkov in izobraževanju sodeluje tudi s slovenskimi znanstvenimi organizacijami, kot sta Inštitut Jožef Štefan in Fakulteta za računalništvo in informatiko.

Kot eden največjih izzivov je bil izpostavljen dostop do teh podatkov in etični vidik njihove uporabe, saj so nekateri masivni podatki zelo občutljivi z vidika poslovnih in osebnih podatkov, obstajajo pa tudi nekatere neusklajenosti nekaterih pravnih predpisov glede dostopa do njih. Tako SURS posebno pozornost namenja vzpostavitvi partnerstva s skrbniki podatkov, ki nam omogoča lažji dostop do teh podatkov. V zadnjih nekaj letih je bil kljub omenjenim izzivom narejen velik napredek pri raziskovanju teh virov, pripravi testnih statistik ter pridobivanju novih metodoloških in IT znanj, ki nam omogočajo obdelavo teh podatkov, zato smo lahko zelo optimistični pri uvajanju teh virov v reden statistični proces.

K 3

Predsednik društva je kronološko na kratko predstavil dogodke v času od zadnje skupščine. V maju 2016 je bila volilna skupščina. Za obdobje od 2016 do 2020 so izvolili nov sedemčlanski izvršni odbor s predsednikom društva na čelu, tričlanski nadzorni odbor in petčlansko častno razsodišče. Marca 2016 je izšla 59. številka Biltene statističnega društva. V mesecu septembru smo soorganizirali konference Uporabna statistika, Srečanje mladih statistikov in Statistični

dan ter izdali 13. številko Metodoloških zvezkov in društvenega informativnega publikacije. Žal se 10. rednega letnega srečanja predstavnikov evropskih in regionalnih statističnih društev v Bukarešti (V7) lani nismo udeležili.

V januarju 2017 smo na Brdu pri Kranju sodelovali v organizaciji tradicionalnega 'Statističnega dne' in podelili priznanja Statističnega društva Slovenije 2016. Blejčevo priznanje je prejel prof. dr. Andrej Blejec, priznanje za odličnost statističnega poročanja v medijih je prejela dr. Mojca Vizjak Pavšič. Častna članica Statističnega društva Slovenije 2016 je postala profesorica doktorica Blaženka Košmelj, ki je na začetku januarja 2017 umrla – tako smo ji priznanje podelili posthumno.

Bogdan Grmek je v imenu nadzornega odbora povedal, da so pregledali finančno poslovanje društva in ugotovili, da je bil v letu 2016 presežek prihodkov nad odhodki.

Sklep: Navzoči so soglasno sprejeli poročilo predsednika o delu društva v preteklem letu, poročilo nadzornega odbora in finančno poročilo za leto 2016.

K 4

Bogdan Grmek je pojasnil številke iz finančnega poročila. Iz naslova pridobitne in nepridobitne dejavnosti ter od članarin je društvo dobilo približno 12.000 evrov. Stroški v preteklem letu so znašali nekaj več kot 11.900 evrov in so bili povezani v glavnem z organizacijo strokovnih srečanj. Za kritje stroškov konferenc mora društvo samo poskrbeti, saj nič ne kaže, da bi Ministrstvo za visoko šolstvo, znanost in tehnologijo v prihodnje sofinanciralo strokovna srečanja. Na dan 31. 12. 2016 je imelo društvo na transakcijskem računu 7.224,71 evra.

Del prihodkov dobi društvo iz letnih članarin. Društvo enkrat letno pošlje članom plačilni nalog za letno članarino, ki znaša 10 evrov, za upokojence 6 evrov. Pri nekaterih članih že dlje časa opazamo, da ne plačujejo članarine, čeprav jih društvo redno opominja. Članarina se v letu 2017 ne bo zvišala.

Sklep: Člane, ki kljub opominjanju ne poravnajo svojih obveznosti do društva tri leta, jih bo društvo črtalo iz seznama. Člane bomo zaprosili za elektronske naslove, da jih bomo v prihodnje opominjali za plačilo članarine po elektronski pošti. S tem bomo posodobili spisek članov in optimizirali nastale stroške.

K 5

Dr. Andrej Blejec nam je na kratko podal poročilo o konferenci Uporabna statistika 2016, ki je potekala od 18. do 21. septembra 2016 v Ribnem pri Bledu. Zbralo se je več kot sedemdeset statistikov, analitikov, uporabnikov in drugih, ki jih statistika spremlja pri delu ali pa jih preprosto zanima. Med njimi je bilo malo manj kot polovica tujih udeležencev. Nastopali so ugledni vabljeni predavatelji.

Srečanje mladih statistikov, že enaindvajseto po vrsti, je bilo tokrat po štirih letih spet v Piranu. Srečevali smo se od 4. do 6. novembra 2016 v Morski biološki postaji Nacionalnega inštituta

za biologijo. Udeležilo se ga je 60 mladih statistikov, od tega približno pol iz Slovenije. Na srečanju se je zvrstilo 15 predstavitev, iz vsake države po tri. Letošnje srečanje bo v oktobru gostila Hrvaška v Zagrebu.

Generalna direktorica SURS-a ga. Genovefa Ružič nam je povedala, da je 24. januarja 2017 na Brdu pri Kranju potekal 26. posvet statistikov – Statistični dan, tokrat z naslovom Lačni podatkov? Statistika o hrani. Priznani slovenski strokovnjaki so razpravljali o hrani, o skrbi za prehransko varnost ter o podatkovnih osnovah, ki omogočajo vpogled v stanje na tem področju. Prva okrogla miza je osvetlila širino obravnavane teme, predvsem pa je nakazala, kaj še lahko storimo za večjo učinkovitost kmetijske proizvodnje in gospodarno porabo hrane. Hkrati se je že dotaknila teme naslednje okrogle mize: podatkovnih osnov za spremljanje obravnavanega področja. Udeležencev posveta je bilo približno 300 iz Slovenije in tujine. Statistični dan se je končal s slovesno podelitvijo društvenih priznanj Statističnega društva Slovenije. SURS je v letu 2016 sodeloval pri projektu Evropska noč raziskovalcev, ki je bil spet odlična priložnost za stike z uporabniki. To je del dveletnega evropskega projekta, imenovanega Znanost za življenje, katerega glavni namen je popularizirati poklic raziskovalca znanstvenika in ga približati širši javnosti, zlasti mladim. SURS je za Evropsko noč raziskovalcev izpeljal dve aktivnosti: Dan odprtih vrat na SURS in stojnico Statomanija v ljubljanskem Cityparku.

Sourednica dr. Valentina Hlebec iz Fakultete za družbene vede nam je povedala, da sta v letu 2016 izšli dve številki revije Metodološki zvezki z devetimi prispevki (dva več kot pred enim letom). Urednici sta pri izboru prispevkov za trinajsti letnik revije Metodološki zvezki upoštevali ustaljen recenzentski postopek, po katerem se zahtevata vsaj dve pozitivni anonimni recenziji za vsak sprejeti prispevek. Skladno s tem merilom je bila tudi letos zavrnjena polovica prispelih prispevkov. V recenzentskem postopku je sodelovalo 19 recenzentov iz Slovenije in sedem recenzentov iz tujine. Žal se je znižalo število slovenskih prispevkov in so večinoma objavili kakovostne prispevke iz tujine. Letos je bil med avtorji prispevkov samo eden iz Slovenije, vse druge so prispevali tuji avtorji (Velika Britanija, Španija, Avstralija, Francija, Portugalska, Tajska, Srbija, Iran, Indija). Revija Metodološki zvezki je indeksirana v naslednjih mednarodnih bibliografskih bazah: SCOPUS, STMA-Z, EBSCO, ECONIS in ProQuest. Revija se je prebila med ¼ revij v bazi SCOPUS in ima tudi A' za leto 2015. Je tudi edina revija s področja statistike v Sloveniji. V prihodnje želijo povečati število objav in citiranost revije. Trudijo se, da bi pridobili več člankov od slovenskih raziskovalcev na področju statistike in metodologije. Načrt je povečati število prispevkov, da bi se Metodološki zvezki uvrstili v bazo TomReuters. Težava so nizka sredstva in premalo člankov. Sredstva, ki se trenutno prejemajo, so iz naslova ARRS. Vsa sredstva se porabijo za izdajo revije. Društvo bo pozvalo člane, da se včlanijo v Mednarodni statistični inštitut.

Prof. dr. Janez Stare je povedal, da je bil magistrski program Uporabna statistika v študijskem letu 2016/17 razpisan četrtič. V prvi letnik se je vpisalo 15 študentov, kar je največje število doslej in kaže na to, da je program čedalje boljše prepoznaven. Skromen vpis v drugi letnik (5 študentov) je po drugi strani posledica slabega vpisa lansko leto, ko se je vpisalo le 8 študentov. Modul Uradna statistika je pridobil mednarodno akreditacijo z oznako EMOS

(European Master in Official Statistics), ki deluje pod okriljem Eurostata, največje uradne statistične ustanove v Evropi. Program je tako eden od 12 magistrskih programov v Evropi, ki so kot prvi dobili to oznako. Študenti na študiju EMOS so bili konec oktobra 2016 povabljeni v Luksemburg na European Statistical Week - Study visit to Eurostat. Z našega študija se ga je udeležila študentka 2. letnika Maja Pleško. V letu 2016 smo dobili štiri nove magistrice uporabne statistike. V naslednjem študijskem letu bo program ponovno razpisan, vpisnih mest je 25. Študente moramo spodbujati k včlanitvi v Statistično društvo in k pisanju prispevkov za Metodološke zvezke.

V letu 2016 je bila ustanovljena delovna skupina za področje terminologije, ki jo vodi dr. med. Jaro Lajovic, oblikovana je bila terminološka komisija in pripravljen bil je načrt za delovanje terminološke skupine. Sestava komisije je takšna, da njeni člani pokrivajo osnovna področja stroke. Kratkoročne dejavnosti obsegajo dostop do terminoloških virov, komuniciranje z uporabniki in tekoče terminološko delo. Med dolgoročne dejavnosti je uvrščeno oblikovanje področnih skupin, terminografija, navzočnost v javnosti (tudi prek medijev) in zagotavljanje usklajenosti izrazja.

Sklep: Navzoči so soglasno sprejeli poročilo poročevalcev o delu društva v preteklem letu.

Za izdajanje Metodoloških zvezkov je treba priskrbeti več sredstev. Na IO društva bo podana pobuda za povečanje sredstev, namenjenih Metodološkimi zvezkom, ter podpora nadaljnjim prizadevanjem za priključitev k TomReutersu.

Dolgoročno se bodo oblikovale področne skupine za pripravo fondov izrazja, ki še ni vključeno v obstoječi glosar, ter za morebitno izvedbo delovnih srečanj članov terminološke skupine za razreševanje odprtih vprašanj. Obnovil se bo terminološki slovar, okreplili pa bomo njegovo vidnost na spletu.

K 6

Udeleženci skupščine so razpravljali o dogodkih in dejavnostih v letu 2016 in tudi o tistih, ki jih bo društvo organiziralo ali soorganiziralo v letu 2017. Poudarili so, da je treba doseči večjo prepoznavnost društva in okrepiti terminološko delo. Govorili so tudi o vzdrževanju spletne strani, izdaji Biltene ter arhivskem gradivu. Novih predlogov na skupščini ni bilo. Udeleženci skupščine so soglasno sprejeli sklepe skupščine.

Predsednik se je navzočim zahvalil za udeležbo in sodelovanje na skupščini.

Zapisa: Jerneja Čuk

Overiteljici zapisnika:
Mojca Škrlec
Jana Vajda

Predsednik delovnega predsedstva:
dr. Andrej Blejec

STAGE