

Štatistický úrad Slovenskej republiky
The Statistical Office of the Slovak Republic

SLOVENSKÁ ŠTATISTIKA a DEMOGRAFIA

SLOVAK STATISTICS
and DEMOGRAPHY

vedecký časopis/scientific journal

2/2017
ročník 27



ŠTATISTICKÝ
ÚRAD
SLOVENSKEJ
REPUBLIKY

ISSN 1339-6854 (online)
ISSN 1210-1095 (tlačené vydanie)

INFORMÁCIE PRE PRISPIEVATEĽOV

Príspevky prijímame v slovenskom, v českom a v anglickom jazyku. Musia rešpektovať odborné zameranie časopisu a jeho vedecký charakter. Zaslaný príspevok nesmie byť v recenznom konaní v inom časopise, ani uverejnený v odbornej a inej tlači.

Príspevky zasielajte v elektronickej forme vo formáte MS Word alebo Open Office, typ písma Arial, veľkosť 12, riadkovanie 1. Nad titulkom treba uviesť meno autora a jeho pracovisko.

Súčasťou príspevku je abstrakt (základný popis cieľa a spôsobu spracovania faktov v rozsahu do 100 slov), kľúčové slová (maximálne 5), resumé (stručné zhrnutie obsahu článku s dôrazom na jeho prínos a najvýznamnejšie závery v rozsahu do 500 slov), profesijný životopis (v rozsahu do 120 slov) a kontakt (e-mailová adresa autora). Názov článku, abstrakt, kľúčové slová a resumé poskytne autor aj v anglickom jazyku. Zoznam použitej literatúry v abecednom poradí s úplnými bibliografickými údajmi sa uvádza na konci článku. Odkazy na literatúru sa uvádzajú v texte číslami v hranatých zátvorkách. Poznámky s poradovým číslom sú umiestnené pod čiarou na príslušnej strane textu, ku ktorému sa vzťahujú. Podrobnejšie pokyny nájdete u autori na www.statistics.sk.

Maximálny rozsah vedeckých článkov je 15 normostrán, informatívnych článkov 6 normostrán, recenzie, rozhovory a informácie publikujeme v rozsahu maximálne 3 normostrany. Tabuľky, mapy, grafy a obrázky musia mať názov a uvedení zdroj údajov; odporúčame, aby kopírovali šírku textu. Skratky sa používajú len minimálne, pri prvom použití je potrebné skratku v zátvorke rozpísať. Redakcia zabezpečuje jazykovú úpravu textu.

Príspevky sú recenzované. Oponentské konanie je obojstranne anonymné. Konečné rozhodnutie o publikovaní článku vydáva redakčná rada.

Redakcia si vyhradzuje právo zverejniť články schválené redakčnou radou v tlačenej podobe a s odstupom troch mesiacov aj v elektronickej forme na internetovej stránke Štatistického úradu SR.

INFORMATION FOR AUTHORS

Articles are accepted in Slovak, Czech and English languages and must comply with the journal's professional specialisation and scientific nature as well. The submitted articles should not be peer-reviewed by another journal and should not have already been published in any specialised or other press.

Please submit your articles in electronic form, in MS Word or Open Office format, Arial font, size 12 and typed in single spacing. The author's name and workplace should be indicated above the heading.

Articles should contain an abstract (general description of the objective and the processing methods used up to 100 words), key words (max. 5), resume (brief summary of the article's content emphasizing its contribution and the most important conclusions up to 500 words), curriculum vitae of the author (no more than 120 words) and the author's contact (e-mail address). The author should submit the article's title, abstract, key words and resume in English language. List of the literature used with full bibliographic data should be given in alphabetical order at the end of an article. Bibliographic citations should be given in square brackets. References are indicated by numbers in a text in square brackets. Footnotes should be numbered in the order of the corresponding page of a text. Authors can find more details at the website www.statistics.sk.

Maximum scope of a scientific article is up to 15 standard pages, informative articles should be up to 6 standard pages in length, reviews, discussions and information not more than 3 standard pages. Tables, maps, graphs and pictures should have a title and the data source indicated, it is also advised to copy the width of a text. Abbreviations should be used only rarely and should be appropriately explained in parentheses when first used. Language text revisions are provided by the editorial office.

Articles are reviewed. The opponent procedure is mutually anonymous. The final decision on the article's publication is made by the editorial board.

The editorial office reserves the right to publish articles approved by the editorial board in printed form at intervals of at least three months also in electronic form at the website of the Statistical Office of the SR.

OBSAH/CONTENTS

I. VEDECKÉ ČLÁNKY/SCIENTIFIC ARTICLES

Alena KAŠČÁKOVÁ, Gabriela NEDELOVÁ 3
VÝSKUM ŠTRUKTÚRY A HODNOTY NEPLATENEJ PRÁCE V DOMÁCNOSTIACH
NA SLOVENSKU
RESEARCH INTO THE STRUCTURE AND THE VALUE OF UNPAID WORK IN
THE SLOVAK HOUSEHOLDS

Branislav ŠPROCHA 16
RIZIKÁ HODNOTENIA ZDRAVOTNÉHO STAVU OBYVATEĽSTVA SLOVENSKA
PROSTREDNÍCTVOM VÝBEROVÝCH ZISŤOVANÍ EHS A EU SILC
THE RISKS OF RATING THE HEALTH STATUS OF THE SLOVAK POPULATION
BY MEANS OF EHS AND EU SILC SAMPLE SURVEYS

Bianka PARMOVÁ, Mária VOJTKOVÁ 36
SEGMENTÁCIA ČITATEĽOV S VYUŽITÍM TEXT MININGU
SEGMENTATION OF READERS USING TEXT MINING

II. INFORMATÍVNE ČLÁNKY, NÁZORY, RECENZIE, ROZHOVORY, INFORMÁCIE/ INFORMATIVE ARTICLES, OPINIONS, REVIEWS, INTERVIEWS, INFORMATION

Roman PAVELKA 52
ANALÝZY ČASOVÝCH ŘAD POMOCÍ PROGRAMOVACÍHO JAZYKA R
ANALYSES OF TIME SERIES WITH THE R PROGRAMMING LANGUAGE
Informatívny článok/Informative article

Alexander BALLEK 69
I. SVETOVÉ FÓRUM O DÁTACH PRINIESLO VEĽA INŠPIRÁCIE AJ PRE ŠTÁTNU
ŠTATISTIKU
THE FIRST WORLD DATA FORUM HAS BROUGHT MANY INSPIRATION ALSO
FOR THE OFFICIAL STATISTICS
Informácia/Information

Ľudmila IVANČÍKOVÁ/Zuzana ŠTUKOVSKÁ 72
SČÍTANIE SI ZASLÚŽI BYŤ NÁRODNÝM PROJEKTOM
CENSUS DESERVES TO BE A NATIONAL PROJECT
Rozhovor/Interview

Mikuláš CÁR 75
BIELA KNIHA O BÝVANÍ V BRITSKOM PARLAMENTE
WHITE PAPER ON HOUSING IN THE BRITISH PARLIAMENT
Informácia/Information

Viera LABUDOVIČOVÁ 78
Milan Terek: INTERPRETÁCIA ŠTATISTIKY A DÁT, 4. doplnené vydanie
Milan Terek: STATISTICS AND DATA INTERPRETATION, 4th Extended Edition
Recenzia publikácie/Review of Publication

Elena BENKOVÁ **80**
ŠTATISTICKÝ ÚRAD SR ZÍSKAL UZNANIE ZA PARTICIPÁCIU NA
MEDZINÁRODNOM PROJEKTE V KAZACHSTANE
THE STATISTICAL OFFICE OF THE SR GAINED CREDIT FOR THE
PARTICIPATION IN THE INTERNATIONAL PROJECT IN KAZAKHSTAN
Informácia/Information

III. PRIPRAVUJEME/COMING SOON **81**

Alena KAŠČÁKOVÁ, Gabriela NEDELOVÁ
Ekonomická fakulta Univerzity Mateja Bela v Banskej Bystrici

VÝSKUM ŠTRUKTÚRY A HODNOTY NEPLATENEJ PRÁCE V DOMÁCNOSTIACH NA SLOVENSKU

RESEARCH INTO THE STRUCTURE AND THE VALUE OF UNPAID WORK IN THE SLOVAK HOUSEHOLDS

ABSTRAKT

Príspevok sa venuje problematike neplatenj práce v domácnostiach na Slovensku. Primárny prieskum viedol tím z Ekonomickej fakulty Univerzity Mateja Bela, pričom vychádzajúc z metodológie Eurostatu, obsahuje údaje o jednotlivých činnostiach neplatenj práce. V prvej časti príspevku čitateľ nájde opis dátového zdroja, definíciu neplatenj práce a opis činností, ktoré sa považujú za neplatenú prácu. V ďalšej kapitole autorky prezentujú výsledky analýzy neplatenj práce z pohľadu rodových rozdielov a mimoekonomické a geografické aspekty neplatenj práce. Súčasťou článku je aj opis použitého postupu výpočtu hodnoty neplatenj práce a jej podiel na objeme hrubého domáceho produktu (HDP) v roku 2012.

ABSTRACT

The contribution deals with the issue of unpaid work in Slovak households. The primary survey was led by the research team of the Faculty of Economics, Matej Bel University, according to the Eurostat methodology, containing data on individual unpaid work activities. In the first part of the paper, readers may find description of the data sources, definition of unpaid work and the description of activities, that could be considered as unpaid work. In the next part, the authors present the analysis results of unpaid work from the perspective of gender differences, non-economic and geographic aspects of unpaid work. The paper also contains the description of the applied method of calculating the value of unpaid work as well as its share in GDP in 2012.

KLÚČOVÉ SLOVÁ

neplatená práca, domácnosti, výskum využitia času, generalistický a špecialistický prístup

KEY WORDS

unpaid work, households, TUS Survey, generalist and specialist approach

1. ÚVOD

Neplatená práca je už niekoľko desaťročí v centre pozornosti mnohých zahraničných odborníkov, inštitúcií a výskumných tímov (Anderson, D., Anger, S., Baumann, A., Campbell, I., Blythe, J., Boos, L., Davies, S., Denton, M., Kelliher, C., Pannenberg, M., Shi, H. H., Wang, Y., Zeytinoglu, I. U. a ďalší). Na Slovensku, ale aj v Česku sa doteraz neplatenj práci venuje len okrajová pozornosť a skúma sa najmä v súvislosti s otázkami súladu rodinného a pracovného života, rovnosti či nerovnosti žien a mužov a ich deľbou práce v domácnosti, respektíve s dobrovoľníctvom. Sporadicky sa neplatená práca skúma aj vo väzbe na trh práce (napr. Páleník, V. a kol., Přívarová, M., Rievajová, E. a kol., Šimek, M. a kol.,

Uramová, M., Vincúr, P. a kol.) a jeho špecifiká, pričom sa pozornosť sústreďuje viac na otázky politiky zamestnanosti a rovnováhy na trhu práce, nezaplatenej či nedostatočne ohodnotenej práce a na problémy primárneho a sekundárneho trhu práce (Uramová, 2012).

Skúmanie neplatennej práce je opodstatnené z viacerých uhlov pohľadu. Primárne ekonomické hľadisko ju vymedzuje ako hodnotu, ktorá nie je započítaná do výkonu ekonomiky a ktorej podiel sa líši nielen v jednotlivých krajinách, ale aj pri jej viacdimezióne štruktúrovaní. Samotná problematika neplatennej práce je však až prekvapujúco širokospektrálna. Jej skutočný stav a perspektívy vývoja sa dajú identifikovať iba na základe zohľadnenia množstva faktorov, ktoré oscilujú od historických, kultúrnych, demografických, rodových, sociálnych, psychologických, etických, ekonomických až po politické determinanty jej pretrvávania v súčasnosti. Inými slovami, existencia jednotlivých foriem neplatennej práce v domácnostiach má svoje ekonomické, ako aj mimoekonomické dimenzie. To, že obe dimenzie neplatennej práce sa navzájom ovplyvňujú, je evidentné. Možno dokonca konštatovať, a realizované výskumy to potvrdzujú, že samostatne skúmať ekonomické a mimoekonomické determinanty neplatennej práce v domácnostiach ani nie je možné. Aj v prípade, že sa zameriame na čo najrealistickejší výpočet objemu jednotlivých druhov neplatennej práce v konkrétnom časovom intervale (deň, týždeň, sezóna) na domácnosť a využijeme na to adekvátne ekonometrické techniky a metódy, nevyhne sa potrebe zohľadniť pritom typ domácnosti, počet členov domácnosti, vekovú štruktúru jej príslušníkov, generačné vzťahy vnútri rodiny, status aktivity jej jednotlivých príslušníkov, príjmovú kategorizáciu dospelých členov, počet vyživovaných detí až po umiestnenie domácnosti v mestskom či vidieckom prostredí, prípadne regióne.

Okrem týchto faktorov je nevyhnutné zobrať do úvahy aj na prvý pohľad menej dôležité kritériá, ktoré však paradoxne objem neplatennej práce, rozsah výskytu a tendenciu jej pretrvávania v rodinách ovplyvňujú významnejšie, ako by sme predpokladali. Patria k nim motívy a dôvody vedúce k neplatennej práci, postoje k jednotlivým druhom neplatennej práce a rovnako podiel jednotlivých členov domácnosti na ich vykonávaní. Nemenej dôležité sú aj ďalšie determinanty, ako rolové správanie a očakávania rodinných príslušníkov, zapájanie detí do domácich prác či medzigeneračné odovzdávanie skúseností. Všetky sa významným spôsobom podieľajú na zachovávaní tradícií a zvykov vnútri rodinnej komunity, kreovaní a udržiavaní istej úrovne rodinnej intimity, dôvernosti, solidarity a vzájomnej pomoci. Práve vďaka ich vzájomnému pôsobeniu dochádza k vytváraniu sociálnych vzorcov správania a noriem, ich rešpektovaniu a následnému dodržiavaniu aj v širšom spoločenskom prostredí mimo rodiny (susedia, miestna komunita, škola, iné verejné, štátne a súkromné inštitúcie, pracovný kolektív). Všetky uvedené hľadiská a súvislosti pomáhajú poodhaliť a pochopiť každodenné rodinné rituály a postupy formovania, udržiavania a rozvíjania rodinného života a vzťahov medzi členmi domácnosti, čím zároveň dovoľujú spoznať hĺbku a silu ukotvenia jednotlivých druhov neplatennej práce a predikovať možné perspektívy vývoja jej rozsahu a objemu v budúcnosti. Môžeme preto konštatovať, že neplatená práca má svoju významnú ekonomickú, ale aj nezanedbateľnú sociálnu hodnotu. Tá síce nemá kvantifikovateľne vyjadriteľnú podobu, na druhej strane nám však dovoľuje detailnejšie spoznať kvalitatívne charakteristiky života rodín a zabezpečovania chodu domácností [18].

2. ÚDAJOVÉ ZDROJE

Na identifikovanie štruktúry a objemu neplatenj práce jednotlivcov v domácnostiach existuje niekoľko zdrojov údajov. Z úrovne štátnej štatistiky sa v niektorých krajinách vykonáva zisťovanie Time Use Survey (ďalej TUS), ktoré mapuje rozdelenie času, ktorý jednotlivci venujú rôznym činnostiam počas dňa. Toto zisťovanie sa uskutočňuje v 29 európskych i mimoeurópskych krajinách, na Slovensku sa však nerealizuje. Na kvantifikáciu veľkosti neplatenj práce sa v niektorých krajinách využívajú i dáta získané z populačného cenzu, prípadne zo sociologických výskumov. Na odhad veľkosti času stráveného na vykonávanie vybraných činností neplatenj práce v domácnostiach sa v európskom kontexte môžu využiť vybrané otázky z dotazníkového výskumu International Social Survey Programme (ISSP) a z výskumu European Quality of Life Survey.

Veľmi dôležitým zdrojom údajov o objeme domácich prác v domácnostiach na Slovensku sú dva výskumy zamerané na rodovú problematiku. Prvý z nich realizovala Aliancia žien Slovenska prostredníctvom agentúry Focus v polovici 90. rokov minulého storočia, druhý uskutočnil Inštitút pre verejné otázky v auguste 2006. Výsledky oboch výskumov sú obsahom publikácií autorských kolektívov pod vedením Zory Bútorovej [1], [2].

Predchádzajúce už spomenuté zdroje údajov neposkytujú relevantný komplexný zdroj údajov na určenie veľkosti a štruktúry neplatenj práce a vyjadrenie jej hodnoty. Preto sa v roku 2012 na Slovensku prvýkrát uskutočnil primárny prieskum v slovenských domácnostiach. Riadil ho riešiteľský tím projektu VEGA 1/1141/11 s názvom Trh práce v kontexte špecifik neplatenj práce, meranie jej rozsahu a dopadu na domácnosti, podnikateľskú sféru a ekonomiku, ktorého výsledky boli publikované na viacerých fórach [4], [5], [7], [8], [10], [11]. Cieľom výskumu bolo zistiť objem a štruktúru neplatenj práce vykonávanej v domácnostiach na Slovensku, prepočítať jej hodnotu, vyčíslit' podiel neplatenj práce z objemu HDP vytvoreného na Slovensku v roku 2011 a venovať sa poznaniu ekonomických i mimoekonomických faktorov, ktoré objem a štruktúru neplatenj práce na Slovensku ovplyvňujú. Na výsledky nadviazal projekt VEGA 1/0935/13 Neplatená práca ako potenciálny zdroj sociálno-ekonomického rozvoja spoločnosti a determinant individuálneho blahobytu. V rámci tohto projektu poskytol riešiteľský tím viacdimenzionálny pohľad na objem a štruktúru neplatenj práce na Slovensku a jej medzinárodnú komparáciu.

Na zisťovanie údajov o neplatenj práci využili výskumníci z Ekonomickej fakulty Univerzity Mateja Bela dotazníkový prieskum. Realizovali ho vždy na jar v období rokov 2012 – 2016 a odpovede respondentov na položené otázky reflektovali realitu predchádzajúceho roka.

Tabuľka č. 1: Chronologický prehľad prieskumov, ich ciele, náplň a veľkosť výberovej vzorky

Rok zberu údajov	Počet oslovených		Ciele a náplň prieskumu
	domácností	respondentov	
2012	1 564	4 435	Základné poznanie objemu, štruktúry neplatenj práce jednotlivcov, ako aj domácností ako celku, motívov, postojov a očakávaní jej vývoja

Rok zberu údajov	Počet oslovených		Ciele a náplň prieskumu
	domácností	respondentov	
2013	861	2 247	Objem a štruktúra neplatenj práce v domácnostiach, zistenie postojov domácností k možnosti substituovať neplatenú prácu v domácnosti trhovými službami
2014	1 143	-	Využívanie, resp. nevyužívanie platených trhových služieb, frekvencia využívania a dôvody, ktoré k tomuto postoju domácnosti vedú; samostatný modul bol venovaný potenciálnemu dopytu so špecifikáciou rozsahu a finančnej dotácie, ktorú by boli domácnosti ochotné konkrétnej službe venovať
2015	1 854	5 120	Štruktúra a rozsah neplatenj práce, motívy jej vykonávania, postoj k existujúcemu stavu, materiálne vybavenie domácností a individuálny blahobyt jednotlivcov
2016	1 743	4 818	Štruktúra a rozsah neplatenj práce, využívanie služieb, potenciálny dopyt po službách, postoj a motívy domácností k odmietaniu substituovať neplatenú prácu v domácnosti trhovými službami

Poznámka: V roku 2014 sa otázky v dotazníku týkali len domácnosti ako celku.

Zdroj údajov: vlastné spracovanie

3. VYMEDZENIE NEPLATENEJ PRÁCE V DOMÁCNOSTI

Neplatená práca v najširšom zmysle zahŕňa široký okruh činností, ktoré vykonávajú členovia domácnosti pre svoju potrebu alebo pre potrebu osôb mimo domácnosti a za ktoré členovia domácnosti nedostávajú finančnú odmenu [14]. Výnimku môžu tvoriť činnosti súvisiace s chovom hospodárskych zvierat, pestovaním úžitkových rastlín a výstavbou a opravou domu, ktorých výsledky sa za istých okolností môžu stať predmetom predaja na trhu. Zoznam všetkých činností, ktoré zahŕňa pojem neplatená práca v domácnosti, vychádza z klasifikácie činností v rámci prieskumu TUS (Eurostat 2009) a je uvedený v tabuľke č. 2.

Tabuľka č. 2: Druh a stručný opis činností týkajúcich sa domácnosti a starostlivosti o rodinu kvalifikovaných ako neplatená práca

	Druh činnosti	Stručný opis činností
1.	Príprava jedál	Príprava jedál a občerstvenia, umývanie riadu, prestieranie a ďalšia činnosť súvisiaca s prípravou jedál
2.	Domáce práce – upratovanie	Údržba domácnosti, upratovanie, čistenie, kúrenie, nosenie vody, úpravy a údržba pozemku okolo domu, ďalšia údržba domácnosti
3.	Výroba a údržba textílií	Pranie, žehlenie, starostlivosť o odev, topánky, výroba textilu, ručné práce a ďalšia výroba a starostlivosť o textílie a obuv
4.	Záhradkárstvo a starostlivosť o zvieratá	Pestovanie úžitkových a okrasných rastlín, chov hospodárskych zvierat a starostlivosť o domácich miláčikov a pod.
5.	Výstavba a rekonštrukcie	Výstavba a oprava domu, oprava zariadenia, nábytku, oprava a údržba automobilu, výroba náradia do domácnosti, ďalšia výstavba a služby
6.	Nakupovanie a služby	Nákupy tovarov a kapitálu, komerčné, administratívne, zdravotnícke, veterinárne, ďalšie osobné služby a ďalšie nákupy
7.	Starostlivosť o deti	Fyzická starostlivosť, dozor (dohľad), čítanie, hranie, rozprávanie sa s deťmi, učenie a sprevádzanie dieťaťa, navštevovanie školy, jaslí, ďalšie činnosti starostlivosti o dieťa

	Druh činnosti	Stručný opis činnosti
8.	Starostlivosť o dospelých	Fyzická starostlivosť a pomoc závislému (invalidnému) dospelému členovi domácnosti, iná pomoc takémuto človeku
9	Dobrovoľníctvo	Dobrovoľníctvo je slobodne zvolená činnosť v prospech iných vykonávaná bez nároku na odmenu

Zdroj údajov: Eurostat, 2009

Zo zoznamu je zjavné, že ide predovšetkým o tie aktivity, ktoré ľudia vykonávajú každý deň a venujú im veľa svojho času. Napriek tomu, že výsledky niektorých činností môžu byť finančne ohodnotené na trhu, patria medzi aktivity vykonávané bezodplatne v domácnostiach. Budeme ich považovať za súčasť celkového objemu domácich prác vykonávaných v slovenských domácnostiach. Pre ďalšiu analýzu sme uvažovali detailnejšiu štruktúru 4. a 5. skupiny činností, v ktorej sme podrobne skúmali nasledujúce činnosti: záhradkárstvo – pestovanie okrasných rastlín, záhradkárstvo – pestovanie úžitkových rastlín, starostlivosť o domácich miláčikov, chov hospodárskych zvierat, oprava a údržba (bežné opravy v domácnosti) a výstavba a rekonštrukcia (veľké opravy a rekonštrukčné práce).

4. VÝSLEDKY PRIESKUMOV

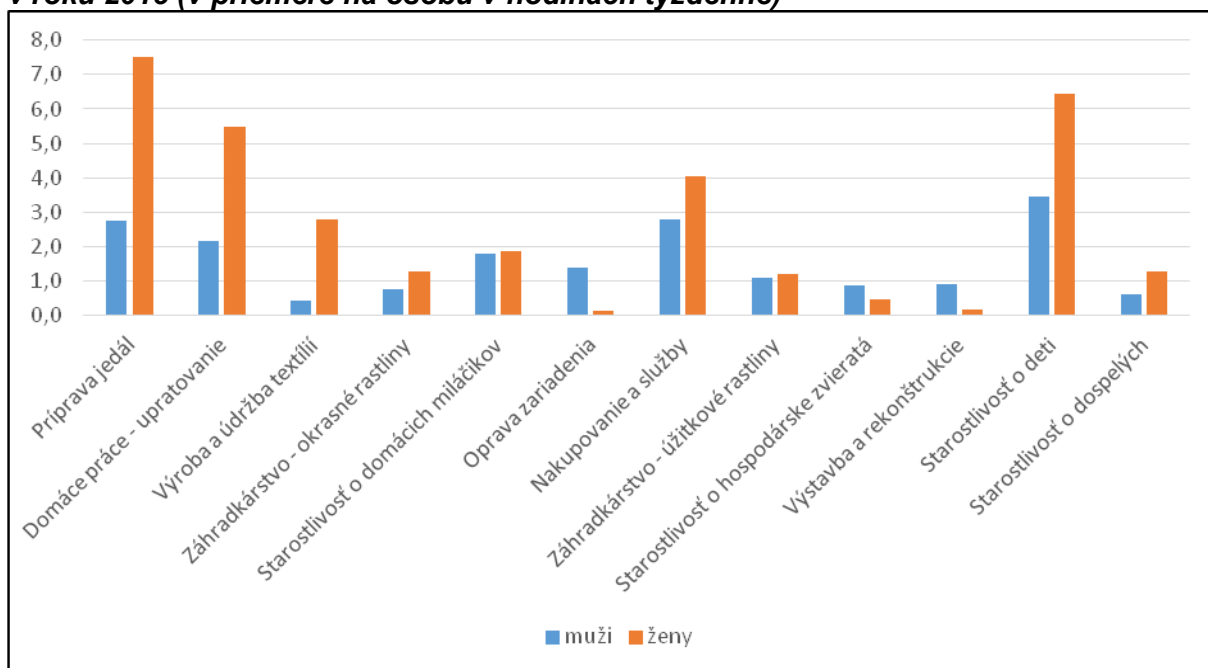
4.1. Domáce práce z pohľadu rodových rozdielov

Spracovanie výsledkov výskumu VEGA 1/1141/11 Trh práce v kontexte špecifik neplatenj práce, meranie jej rozsahu a dopadu na domácnosti, podnikateľskú sféru a ekonomiku poukázalo na výrazné rozdiely v objeme neplatenj práce vykonávanej mužmi a ženami v podmienkach slovenských domácností. Kým ženy oslovené v prieskume v roku 2012 strávili vykonávaním neplatenj práce v domácnosti v priemere 36,37 hodiny týždenne, muži sa za rovnaké obdobie neplatenj práci venovali 21,92 hodiny. Pri vykonávaní väčšiny činností ženy strávili výrazne viac času prípravou jedál, upratovaním, údržbou textílií, starostlivosťou o okrasné rastliny, starostlivosťou o deti aj dospelých. Muži investovali viac času do opráv a údržby, ako aj do výstavby a rekonštrukcií. Štatisticky nevýznamné boli rozdiely v čase strávenom starostlivosťou o domácich miláčikov, pestovaním úžitkových rastlín a dobrovoľníckou prácou. Na testovanie významnosti rozdielov sa použili t-test, Wilcoxonov test a hladina významnosti 0,05.

Pri neplatenj práce v domácnostiach na Slovensku strávia ženy najviac času prípravou jedál ($8,07 \pm 6,59$ hodiny¹), starostlivosťou o deti ($6,37 \pm 19,01$ hodiny) a upratovaním ($5,77 \pm 4,99$ hodiny), muži prípravou jedál ($3,06 \pm 3,74$ hodiny), starostlivosťou o deti ($2,90 \pm 9,58$ hodiny) a upratovaním ($2,84 \pm 3,83$ hodiny).

Na základe údajov zistených v prieskumoch v jednotlivých rokoch možno konštatovať, že časové zaťaženie mužov a žien vykonávaním jednotlivých druhov činností domácich prác bolo veľmi podobné a nepodliehalo výrazným zmenám v čase. Rozdelenie týždenného času venovaného špecifickým činnostiam neplatenj práce v domácnosti z prieskumu v roku 2015 znázorňuje graf 1.

¹ Údaje v zátvorkách informujú o hodnote priemeru a smerodajnej odchýlky.

Graf č. 1: Rozdelenie času na jednotlivé činnosti neplatenj práce podľa pohlavia v roku 2015 (v priemere na osobu v hodinách týždenne)

Zdroj údajov: vlastné spracovanie údajov za rok 2015 zistených pri riešení projektu VEGA 1/0935/13.

V roku 2012 sa na Slovensku v rámci výskumu ISSP Slovensko 2012 modul Rodina uskutočnil prieskum, ktorý obsahoval otázky o neplatenj práci v domácnosti. Výsledky tohto prieskumu boli podobné ako výsledky prieskumu VEGA 1/1141/11. Muži oslovení v tomto prieskume uviedli, že neplatenú prácu v domácnosti vykonávali priemerne 19,98 hodiny týždenne, ženy 35,74 hodiny. Vzhľadom na odlišnú formuláciu otázok nebolo možné detailnejšie opísať časovú dotáciu jednotlivých činností, ale tento výskum obsahoval otázky o participácii mužov a žien na vykonávaní domácich prác a tiež subjektívne vyjadrenie postoja k ich vykonávaniu. Z výsledkov výskumu vyplýva, že rozdelenie domácich prác nezodpovedá ideálnym predstavám mužov a žien. Ženy požadujú, aby sa na domácich prácach podieľali obaja partneri. Až 35 % mužov uznáva, že ich podiel na domácich prácach je nižší, ako by mal byť, a 47 % žien pociťuje, že sa na domácich prácach podieľajú viac ako muži.

Vyššie zaťaženie žien neplatenou prácou otvára otázku ich participácie na platenej práci. Podľa výsledkov ISSP bol priemerný týždenný objem času v platenej práci u mužov 43,08 hodiny a u žien 41,64 hodiny. Podľa prieskumu projektu VEGA muži strávili v platenej práci týždenne priemerne 44,12 hodiny a ženy 39,65 hodiny. Spolu s neplatenou prácou v domácnosti muži za týždeň odpracujú v priemere 67,03 a ženy 75,89 hodiny.

Potvrdili sa očakávané rozdiely v rozsahu práce mužov aj žien. Celkové pracovné zaťaženie žien je zjavne vyššie – v sledovaných domácnostiach týždenne v priemere strávili spolu prácou v zamestnaní a v domácnosti podľa výsledkov ISSP (resp. VEGA) o vyše 14 hodín (resp. o 9 hodín) viac ako muži. Problém tzv. dvojitej záťaže žien je v podmienkach Slovenska špecifický a jeho interpretácia vyžaduje zohľadňovať kultúrno-historické hľadisko, ktoré odkazuje na skutočnosť, že participácia slovenských (československých) žien na trhu práce sa vo zvýšenej miere

začala prejavovať ešte v období budovania socializmu. Nárastom zamestnanosti žien sa však neznižil objem neplatenj práce v domácnostiach. Súvisí to najmä s tým, že vtedajšie domácnosti na Slovensku na rozdiel od domácností v západnej Európe nevyužívali moderné technológie. Ženám chýbali elektrické spotrebiče (práčka, sušička, umývačka riadu, mikrovlnná rúra a pod.) na uľahčenie predovšetkým domácich prác, ktoré sú svojím objemom i významom výrazne zastúpené v jednotlivých kategóriách neplatenj práce. Uvedený stav pretrvávajúcej záťaže žien mimo plateného zamestnania ovplyvnila aj nerozvinutá terciárna sféra v podobe nedostupných (neexistujúcich) trhových substitútov v oblasti služieb (donáška jedla, e-shopy, upratovacie služby, au pair služby, zariadenia pre seniorov, značkový servis automobilov, služby záhradných architektov) [13].

4.2. Mimoekonomické aspekty neplatenj práce

Iný uhol pohľadu na neplatenú prácu v domácnostiach otvára priestor na hlbšiu analýzu kvalitatívnej stránky neplatenj práce v domácnostiach, a to skúmanie:

- motívov vedúcich k neplatenj práci,
- postojov k neplatenj práci,
- pomoci pri neplatenj práci,
- perspektívy vývoja neplatenj práce.

Analýza odpovedí priniesla viacero zaujímavých zistení. Pri motivácii neplatenj práce prevažujú determinanty, ktoré majú citeľne sociálno-vzťahový rozmer, čo dovoľuje predpokladať, že aj postoje, ktoré sa vytvárajú a formujú pri vykonávaní neplatenj práce v atmosfére vzájomných rodinných väzieb, budú pozitívne a relatívne trvalé. Respondenti svojimi odpoveďami potvrdili, že neplatenú prácu vykonávajú v prevažnej miere radi a dobrovoľne. Na domácich prácach, ktoré tvoria podstatnú časť neplatenj práce v rodinách, sa podieľajú významným spôsobom práve rodinní príslušníci. Spoliehanie sa na svojich najbližších, dôverovanie vlastným a blízkym osobám odhaľuje dlhodobu živenú a generačne podporovanú predstavu o istej cyklickosti a akoby uzavretosti recipročného charakteru vzťahov a väzieb medzi najbližšími členmi rodiny, prípadne pokrvného príbuzenstva (gastronomická tradícia zabezpečovania svadobných hostín, rodinná výpomoc pri výstavbe domov, spoločné obrábanie pôdy a zber poľnohospodárskych plodín a iné). Preto neprekvapuje zistenie, že vnímanie perspektívy a vývoja rozsahu neplatenj práce v budúcnosti koreluje s očakávaniami jej zachovania až nárastu. Neplatená práca v podmienkach slovenských domácností je ešte stále prirodzenou a samozrejmovú súčasťou ich každodenného koloritu [9].

Záveru vyplývajúce z uvedených preferencií potvrdzujú tendenciu prevahy konzervatívnych, tradicionalistických postojov k neplatenj práci v podmienkach Slovenska. V tejto súvislosti je preto podstatné zistenie, že ekonomická interpretácia neplatenj práce (vedená snahou o zapracovanie jej kvantifikovaného objemu do HDP a úsilím o znižovanie jej rozsahu vytváraním ponuky trhových substitútov) naráža na mimoekonomické determinanty, ktoré výrazným spôsobom ovplyvňujú pretrvávanie a vykonávanie neplatenj práce v domácnostiach. Sú to práve mimoekonomické dimenzie neplatenj práce, ktoré výrazným spôsobom ovplyvňujú správanie a rozhodovanie rodiny a jej príslušníkov aj o využívaní trhových substitútov adekvátnych jednotlivým formám neplatenj práce vykonávaným v domácnostiach. Problematickosť a jednoznačnosť takejto spotrebiteľskej orientácie však výrazne komplikuje predovšetkým motivácia takéhoto správania, teda dôvody, ktoré by mali

viest' rodiny k vytváraní dopytu po trhových alternatívach. Ak oslovené domácnosti uvažujú o potenciálnom využívaní trhových substitútov, tak tie sa týkajú v rámci jednotlivých kategórií neplatenej práce viac služieb pre domácnosť (príprava jedál, upratovanie či opravy a údržba) ako osobných služieb (starostlivosť o deti a dospelých). Toto zistenie koreluje s výraznou preferenciou vzťahovo-emočných motívov výskytu a pretrvávania neplatenej práce v slovenských rodinách. Preto neprekvapuje, že redukovanie starostlivosti o najbližších (deti, rodičia) je ešte stále vo väčšine prípadov nezlučiteľné s predstavou „zakúpenia“ starostlivosti na trhu ako produktu (služby), ktorý by mal nahrádzať osobnú rodičovskú lásku či splnenie si záväzkov doopatrovania rodičov vlastnými silami [12].

4.3. Geografické aspekty neplatenej práce

Objem neplatenej práce v slovenských domácnostiach vykazuje rozdiely podľa príslušnosti k regiónu. Dá sa povedať, že sa zvyšuje od juhozápadu na severovýchod. Domácnosti v Bratislavskom kraji vykazujú týždenný objem neplatenej práce 63,1 hodiny, v Nitrianskom kraji 66,8 hodiny, v Trnavskom kraji 75,7 hodiny a v Trenčianskom kraji 78,9 hodiny. Stred a východ Slovenska zaznamenáva mierne vyššie hodnoty neplatenej práce pripadajúcej na domácnosť. Domácnosti Banskobystrického kraja mali priemerný týždenný objem neplatenej práce 73,1 hodiny, Žilinského kraja 77,2, Košického kraja 77,7 a Prešovského kraja 80,6 hodiny.

V dotazníkovom prieskume sme tiež zisťovali príslušnosť domácnosti k typu obce, pričom sme podľa veľkosti rozlišovali kategórie: krajské mesto, mesto a iná obec (t. j. vidiecka obec). Priemerná veľkosť objemu neplatenej práce klesala s rastom veľkosti obce. V krajských mestách bola zistená v rozsahu 49,7 hodiny, v mestách 71,4 hodiny a v iných obciach 76,2 hodiny.

Súčasťou transformačného procesu na Slovensku je aj zastavenie populačného rastu miest a pokles ich obyvateľov. Po roku 2000 sa na Slovensku začala zintenzívňovať bytová výstavba zameraná prevažne na výstavbu rodinných domov v suburbánných sídlach a vo vidieckom zázemí miest. Rodinné domy poskytujú obyvateľom širšiu paletu povinností (možností) vykonávania neplatenej práce, čo nepochybne súvisí s nepomerne väčším obytným priestorom, doplnenými nebytovými priestormi a pozemkom, vyžadujúcim opateru a starostlivosť. Tieto skutočnosti sa prejavujú vo vyššej časovej dotácii, ktorú domácnosti v rodinných domoch venujú neplatenej práci (tabuľka č. 3).

Tabuľka č. 3: Priemerný týždenný objem neplatenej práce v domácnostiach podľa druhu bývania (v hodinách)

Druh bývania	Priemerný počet hodín	Počet	Smerodajná odchýlka
Rodinný dom	90,14	764	65,89
Byt v obytnom dome	60,62	764	54,65
Iný druh ubytovania	39,11	31	32,84
Spolu	74,66	1 559	62,04

Zdroj údajov: [6]

Vo väčších sídlach je zrejmy nižší objem vykonávanej neplatenej práce v domácnostiach, čo môže byť spôsobené vyššou zamestnanosťou, ale tiež vyšším podielom obyvateľstva bývajúceho v obytných domoch. Menšie sídla sa vyznačujú

typickým vysokým zastúpením obyvateľstva bývajúceho v rodinných domoch, ktoré si vynucujú potrebu realizácie aktivít neplatenj práce. V intenciách už spomenutých zmien možno očakávať tlak na posilnenie objemu neplatenj práce idúci u zamestnaných osôb ruka v ruke s potrebou jej suplovania z externých zdrojov z dôvodu pracovného vyťaženia i časového zaťaženia pracovnou migráciou či dochádzaním do zamestnania.

5. HODNOTA NEPLATENEJ PRÁCE

Na výpočet hodnoty neplatenj práce možno využiť inputovú (market replacement cost approach) alebo outputovú metódu (opportunity cost approach). V prípade, že sa zvolí inputový prístup, je potrebné prijať predpoklad, že hodnota tovarov a služieb vyprodukovaná v domácnosti sa rovná nákladom na jej produkciu. Napriek tomu, že produkcia domácností, ktorú trh neponúka, vyžaduje aj ďalšie vstupy, nielen prácu, používa sa pri nej len práca, ktorá sa následne ohodnotí [15].

Z inputových metód sa najčastejšie používa metóda trhovu-nákladového ocenenia, ktorá vychádza z ocenenia výkonu neplatenj práce porovnaním s cenou adekvátnej (substitučnej) neplatenj práce, t. j. mzdou za túto prácu. Pri generalistickom prístupe sa vychádza zo mzdy osoby profesionálne sa starajúcej o domácnosť, kým pri špecializovanej metóde sa vychádza zo mzdy osôb špecializovaných na určitý druh činnosti (kuchár, upratovačka, opatrovatelka a pod). Veľkosť neplatenj práce je potom možné vypočítať nasledujúcim spôsobom:

$$\sum_{i=1}^N \sum_{j=1}^M X_i \cdot H_{ij} \cdot W_j \cdot 52$$

kde

N – veľkosť výberového súboru,

M – počet skupín činností neplatenj práce,

X_i – extrapolačný faktor (váhy, pomocou ktorých sa zovšeobecňuje hodnota z výberu),

H_{ij} – počet hodín strávených v neplatenj práci z *j*-tej skupiny aktivít za jednu časovú periódu zodpovedajúcej *i*-temu jednotlivcovi,

W_j – priemerná mzda zo špecializovaných pozícií v *j*-tej skupine.

Vzhľadom na disponibilnú štruktúru údajov sme zvolili inputovú metódu a dva prístupy: špecialistický (oceňuje výstup každej činnosti neplatenj práce mzdou príslušného špecialistu) a generalistický (všetky činnosti oceňuje mzdou jednej osoby, ktorá by tieto činnosti potenciálne vykonávala) [16]. Mzdy na ocenenie hodnoty jednotlivých činností sme prevzali z publikácie o štruktúre miezd v SR v roku 2012. Výsledky ocenenia neplatenj práce v domácnosti použitím špecialistického a generalistického prístupu a vyčíslenie hodnoty neplatenj práce a jej podielu z objemu HDP v roku 2012 sú obsahom tabuľky č. 4.

Tabuľka č. 4: Hodnota neplatenej práce vytvorenej v domácnostiach na Slovensku v roku 2012

Metóda	Hodnota	
	mil. EUR	% z HDP
Generalistický prístup		
so zahrnutím trhových činností	14 639	20,28
s vylúčením trhových činností	13 159	18,23
Špecialistický prístup		
so zahrnutím trhových činností	18 148	25,14
s vylúčením trhových činností	15 773	21,85

Zdroj údajov: [16]

Z výsledkov je zrejmé, že hodnota práce vytvorenej v domácnostiach na Slovensku tvorí až štvrtinu hodnoty vytvorenej v národnom hospodárstve, pričom sa do tejto veličiny nezapočítava. Ako už bolo spomenuté, proces jej tvorby má celý rad sociálno-ekonomických súvislostí a vplyvov. Slovensko sa z hľadiska celkového objemu neplatenej práce nachádza mierne nad priemerom štátov OECD. Podobné výsledky sa zaznamenali napríklad v Holandsku a na Novom Zélande. Zo stredoeurópskeho regiónu dosahuje veľmi podobné hodnoty Poľsko. V Poľsku je objem neplatenej práce za deň v priemere vyšší o 7 minút [5].

6. ZÁVER

Z výsledkov primárneho prieskumu objemu a štruktúry neplatenej práce možno vyvodit' mnohé ekonomické i mimoekonomické súvislosti. Časové zaťaženie členov domácností pri vykonávaní neplatenej práce je rodovo nerovnomerne rozdelené. Príprava jedál, upratovanie, údržba textílií, nakupovanie a služby, starostlivosť o deti a dospelých sú jednoznačnou dominantou žien, činnosti týkajúce sa opravy a údržby, výstavby a rekonštrukčných prác zase mužov. Rovnaký alebo približne rovnaký podiel domácich prác vykonávajú muži a ženy len v prípade starostlivosti o domácich miláčikov, chovu hospodárskych zvierat, pestovania úžitkových rastlín a dobrovoľníckej činnosti. Líšia sa aj názory na uvedené rozdelenie. Väčšina mužov a žien sa síce prikláňa k názoru, že vykonávajú taký podiel domácich prác, ako im prislúcha, avšak zvyšná časť žien si prevažne myslí, že vykonáva viac alebo oveľa viac, zatiaľ čo muži priznávajú svoj nižší podiel na domácich prácach [6].

Pri sledovaní motívov, ktoré vedú k vykonávaniu neplatenej práce, prevládajú nad ekonomicko-finančnými motívy so sociálno-vzťahovým efektom. Slovensko sa tak radí ku krajinám s tradicionalistickým a konzervatívnym prístupom domácností k neplatenej práci. Členovia domácností prevažne využívajú pri vykonávaní jednotlivých činností neplatenej práce v domácnosti pomoc ostatných členov a vyzdvihujú jej výchovný vplyv, stmelujúci a podporný efekt na súdržnosť rodiny. Perspektívne sa očakáva rast objemu neplatenej práce.

Z geografického hľadiska možno pozorovať vyšší objem vykonávaných činností neplatenej práce v krajoch stredného a severovýchodného Slovenska, viac neplatenej práce je na vidieku a v malých mestách a prevažne v domácnostiach bývajúcich v rodinných domoch.

V súvislosti so zmenenou demografickou situáciou, pri znižujúcej sa pôrodnosti a náraste poproduktívnej zložky populácie, treba počítať so zvýšeným dopytom po komplexnej inštitucionálnej starostlivosti o starších členov rodiny. Paradox „dvojitej pasce“ starnutia spočíva v tom, že nastupujúca „bezdetná“ produktívna populácia bude tieto služby vzhľadom na absenciu vlastného potomstva zároveň sama potrebovať. Preto možno predpokladať, že sa zvýšia nároky na kvalitu týchto trhových a verejných služieb.

Príspevok vznikol s podporou grantového projektu VEGA 1/0935/13 Neplatená práca ako potenciálny zdroj sociálno-ekonomického rozvoja spoločnosti a determinant individuálneho blahobytu a Agentúrou na podporu výskumu a vývoja na základe zmluvy č. APVV-0309-11.

LITERATÚRA

- [1] BÚTOROVÁ, Z. a kol.: Ona a on na Slovensku: Ženský údel očami verejnej mienky. Bratislava: Focus, 1996. 180 s.
- [2] BÚTOROVÁ, Z. a kol.: Ona a on na Slovensku: Zaostrené na rod a vek. Bratislava: Inštitút pre verejné otázky, 2008. 365 s. ISBN 978-80-89345-10-6.
- [3] EUROSTAT. http://www.timeuse.org/sites/ctur/files/public/ctur_report/4583/ks-ra-08-014-en.pdf. 2009. Prístup k 4. 6. 2016.
- [4] HRONEC, M. – KOLLÁR, J. – KUBIŠOVÁ, Ľ.: Neplatená práca a jej význam pri meraní makroekonomickej aktivity. In: Unpaid work = Neplatená práca = Nieodplátna praca: international scientific conference, Gierlož, 29-30 May 2014, I, economic section. Olsztyn: University of Warmia and Mazury, 2014, s. 95 – 113.
- [5] HUŇADY, J. – ORVISKÁ, M.: Faktory motivácie k neplatenj práci a niektoré politické implikácie. In: Globalizácia a jej sociálno-ekonomické dôsledky. Zborník z medzinárodnej vedeckej konferencie, Rajecké Teplice 9. – 10. október 2013. Žilina: Žilinská univerzita, Fakulta prevádzky a ekonomiky dopravy a spojov, 2013, s. 1 – 12.
- [6] KAŠČÁKOVÁ, A.: Neplatená práca v domácnostiach na Slovensku. In: Rodina na Slovensku v teórii a vo výskume. 1. vyd. Bratislava: STIMUL, 2015. ISBN 978-80-8127-151-9.
- [7] KAŠČÁKOVÁ, A. – TUSCHLOVÁ, M. – URAMOVÁ, M.: Neplatená práca ako potenciálny zdroj ekonomického rastu. In: Rozpočet EÚ, inovácie a ekonomický rast. Zborník príspevkov z medzinárodnej vedeckej konferencie. Banská Bystrica: Regionálne európske informačné centrum, 2013.
- [8] KAŠČÁKOVÁ, A. – NEDELOVÁ, G. – POVAŽANOVÁ, M.: Determinants of the unpaid work in Slovakia. In: Statistika: statistics and economy journal. Praha: Czech Statistical Office, 2013, č. 1.
- [9] KIKA, M. – MARTINKOVIČOVÁ, M.: Neekonomické súvislosti neplatenj práce. In: Trh práce v kontexte špecifík neplatenj práce, meranie jej rozsahu a dopadu na domácnosti, podnikateľskú sféru a ekonomiku. Zborník vedeckých štúdií VEGA 1/1141/11. Banská Bystrica: Univerzita Mateja Bela, Ekonomická fakulta, 2012. ISBN 978-80-557-0437-1.
- [10] KIKA, M. – MARTINKOVIČOVÁ, M.: Neplatená práca v slovenských domácnostiach – výskum, výsledky a súvislosti. In: Sociológia, 2015, č. 5, s. 474 – 503.
- [11] KIKA, M. – VALLUŠOVÁ, A.: Unpaid work in households – selected implications for bussiness management. In: SGEM conference on political sciences, law,

- finance, economics and tourism: conference proceedings, international multidisciplinary scientific conferences on social sciences and arts, Bulgaria 3-9 September 2014. Sofia: STEF92 Technology, 2014, p. 859-865.
- [12] MARTINKOVIČOVÁ, M. – POVAŽANOVÁ, M. – KAŠČÁKOVÁ, A.: Unpaid work as a potential support for employment in the labour market. In: Olsztyn economic journal. Olsztyn: Wydawnictwo Uniwersytetu Warmińsko-Mazurskiego w Olsztynie, 2014, č. 1, s. 1 – 14.
- [13] MARTINKOVIČOVÁ, M. – KAŠČÁKOVÁ, A. – KIKA, M. – KUBIŠOVÁ, Ľ.: Unpaid work in selected households in Slovakia. In: Zborník z medzinárodnej vedeckej konferencie Mobbing in economic, legal and social terms. Ogonki: University of Warmia and Mazury in Olsztyn, 2015, p. 288-300.
- [14] MIRANDA, V.: Cooking, Caring and Volunteering: Unpaid Work Around the World: OECD Social, Employment and Migration Working Papers, č. 116. OECD Publishing, 2011. Dostupné na: www.oecd.org/els/workingpapers.
- [15] PAĽA, J. – POVAŽANOVÁ, M.: Metódy merania monetárnej hodnoty neplatenj práce. In: Nové trendy – nové nápady. Znojmo: Súkromná vysoká škola ekonomická Znojmo, september 2011. ISBN 978-80-87314-20-3.
- [16] POVAŽANOVÁ, M. – VALLUŠOVÁ, A. – URAMOVÁ, M. – KAŠČÁKOVÁ, A.: Assigning monetary values to unpaid work in Slovakia. In: Procedia Economics and Finance: 3rd Economics and Finance Conference, Rome, Italy, April 14-17 2015 and 4th Economics & Finance Conference, London, UK, August 25-28 2015. London: Elsevier, 2015.
- [17] Štruktúra miezd v SR 2012. Dostupné na: www.statistics.sk. Prístup 16. 6. 2015.
- [18] URAMOVÁ, M.: Interdisciplinárny výskum neplatenj práce v podmienkach SR – ciele, metódy a postupy riešenia. In: Trh práce v kontexte špecifík neplatenj práce, meranie jej rozsahu a dopadu na domácnosti, podnikateľskú sféru a ekonomiku. Banská Bystrica: EF UMB Banská Bystrica, 2012. 10 s. ISBN 978-80-557-0437-1.

RESUME

There are various data sources that can be used to estimate the amount and value of unpaid work. However, in order to obtain relevant and complex data source it was necessary to carry out a primary survey and collect these data. For the first time it was performed in 2011, led by a team of researchers at the Faculty of Economics UMB through support from the VEGA grant agency. The data concerning the amount and structure of the unpaid work were used to calculate its total value and share in GDP generated in Slovakia, emphasizing both the economic and non-economic factors that may affect the performance of unpaid work in Slovakia.

At first, the research team needed to specify the types of activities directly concerning unpaid work within a household. These activities were taken over from the TUS survey used by Eurostat since 2009. Most of these activities are carried out by household members on a daily basis without remuneration, despite entitlement to a reward. The data analysis confirmed the gender differences in both the scope and the type of the unpaid activities, within the Slovak households. These patterns resulted also from the survey ISSP Slovakia 2012 Family module. The results concerning the increased amount of unpaid work done by women were examined in correlation with the amount of time women spent doing paid work. The results show that despite the smaller amount of time women spent in paid work, they spend considerably more time working than men. The so-called double burden of women has its cultural and historical background.

Our survey included questions regarding motivation to do unpaid work, the help of household members with unpaid work and their future expectations with unpaid work. The analysis of the responses suggests that unpaid work is largely perceived as a socio-relational issue and a natural and inseparable part of a common household functioning.

Another factor determining the amount of unpaid work is the spatial distribution within Slovakia by a municipality size. The general pattern is that the amount of unpaid work increases eastwards and decreases as the municipality size grows.

For calculating the value of unpaid work, output method and two approaches were used – the specialist and generalist approach. The results in both cases show that the value of unpaid work done within a household accounts for almost a quarter of the domestic output although it is not included in the official statistics on GDP.

PROFESIJNÝ ŽIVOTOPIS

Ing. Alena Kaščáková, PhD., pôsobí ako odborná asistentka na Katedre kvantitatívnych metód a informačných systémov Univerzity Mateja Bela v Banskej Bystrici. V roku 2006 na tomto pracovisku obhájila dizertačnú prácu. Venuje sa ekonomickej demografii, problematike aktívneho starnutia a využitiu štatistických metód v analýze sociálno-ekonomických javov a procesov. Je spoluautorkou viacerých publikácií, zúčastňuje sa na riešení viacerých vedeckovýskumných projektov.

RNDr. Gabriela Nedelová, PhD., je absolventkou Matematicko-fyzikálnej fakulty Univerzity Komenského v Bratislave v odbore pravdepodobnosť a matematická štatistika. V súčasnosti je odbornou asistentkou na Katedre kvantitatívnych metód a informačných systémov na Ekonomickej fakulte Univerzity Mateja Bela. Venuje sa aplikáciám štatistických metód v ekonómii.

KONTAKT

alena.kascakova@umb.sk
gabriela.nedelova@umb.sk

Branislav ŠPROCHA
INFOSTAT – Výskumné demografické centrum
Centrum spoločenských a psychologických vied SAV

RIZIKÁ HODNOTENIA ZDRAVOTNÉHO STAVU OBYVATEĽSTVA SLOVENSKA PROSTREDNÍCTVOM VÝBEROVÝCH ZISŤOVANÍ EHS A EU SILC¹

THE RISKS OF RATING THE HEALTH STATUS OF THE SLOVAK POPULATION BY MEANS OF EHS AND EU SILC SAMPLE SURVEYS

ABSTRAKT

Hlavným cieľom príspevku je poukázať na riziká použitia výberových zisťovaní EHS a EU SILC pri hodnotení zdravotného stavu obyvateľstva na príklade Slovenska. Analyzujeme postavenie Slovenska v európskom priestore z pohľadu zdravotného stavu, a to v spojitosti s úrovňou úmrtnosti. Rozdiely v hodnotách zdravých rokov života sa snažíme dekomponovať na faktor úmrtnosti a faktor zdravotného stavu.

ABSTRACT

The main objective of this paper is to indicate the risks of using the EU SILC and the EHS sample surveys in assessing of the health status on the example of Slovakia. We analyze the position of Slovakia in the European area in terms of health status in conjunction with the level of mortality. The differences in the values of the healthy life years are decomposed by the factor of mortality and health status.

KLÚČOVÉ SLOVÁ

zdravotný stav, európsky minimálny modul zdravia, EU SILC 2014, EHS 2014, úmrtnosť, Slovensko

KEY WORDS

health status, the Minimum European Health Module, EU SILC 2014, EHS 2014, mortality, Slovakia

1. ÚVOD

Predĺžovanie života vyvoláva otázky, či pridané roky sú aj rokmi, ktoré osoby prežijú v dobrom zdravotnom stave, alebo kvalita života za týmto vývojom zaostáva. Z teoretického hľadiska sa vyprofilovali tri hlavné skupiny názorov na túto problematiku [12]. Zástancovia expanzie morbidity [napr. 8, 11, 14] sa prikláňajú k názoru, že zníženie úmrtnosti prostredníctvom zlepšenia liečebných schopností a postupov nevedlo v rovnakom rozsahu k poklesu incidencie ochorení. Predĺženie života tak bolo najmä výsledkom zníženia pravdepodobnosti úmrtia a nie skutočného vyliečenia. Roky navyše, ktoré sú výsledkom týchto zmien, sú tak prežité v horšom alebo zlom zdravotnom stave. Opačný názor prezentuje teória kompresie morbidity

¹ **EHS (European Health Interview Survey)** predstavuje harmonizované výberové zisťovanie o zdraví a o ďalších oblastiach súvisiacich so zdravím. Uskutočňuje sa na reprezentatívnej vzorke populácie všetkých členských krajín EÚ. Na Slovensku sa realizovali zatiaľ dve vlny v gescii Štatistického úradu SR.

EU SILC (Statistics on Income and Living Conditions) je harmonizované výberové zisťovanie, ktorého cieľom je získať informácie o rozdelení príjmov, o úrovni a štruktúre chudoby a o sociálnom vylúčení v jednotlivých krajinách EÚ. Na Slovensku ho každoročne realizuje Štatistický úrad SR.

[4, 5]. Vychádza z predpokladu, že pokles úmrtnosti prebieha súčasne so zlepšovaním zdravotného stavu, a to aj smerom do vyšších vekových skupín. Posledná, tretia skupina je akýmsi prienikom oboch krajných teórií a predpokladá existenciu akejsi dynamickej rovnováhy. Predlžovanie života sa dá čiastočne vysvetliť spomalením miery nárastu chronických ochorení. Znamená to, že osoby síce budú v priemere dlhšie vystavené zhoršeniu zdravotného stavu, ale nepôjde o výrazné zníženie kvality života. Kľúčovou otázkou v spojitosti s touto problematikou je spôsob, ako vlastne merať zdravotný stav obyvateľstva. V posledných desaťročiach sme svedkami viacerých snáh vytvoriť metodický koncept na kvantifikáciu kvality zdravotného stavu obyvateľstva prostredníctvom vhodne zvolených syntetických indikátorov, vzájomne porovnávať tieto údaje medzi jednotlivými krajinami či posudzovať ich vývoj v čase. Okrem toho cieľom je tiež odhaľovať nerovnosti v zdraví, v prístupe k zdravotnej starostlivosti. Získané údaje a z nich konštruované indikátory následne majú slúžiť ako dôležité oporné mechanizmy na formulovanie a zavádzanie zdravotníckych politík a sú aj nástrojmi na ich hodnotenie. Postupne sa stali súčasťou viacerých dôležitých politických opatrení a iniciatív (napr. lisabonská stratégia, Európa 2020, stratégia trvalo udržateľného rozvoja, európske kľúčové ukazovatele o zdraví a pod.).

Primárnym problémom, s ktorým všetky koncepty prichádzajú do kolízie, je určenie bodu, kedy je osoba ešte zdravá a kedy ju už môžeme označiť za chorú, resp. kedy sa signifikantne znížila kvalita jej zdravia [9]. Exaktne určiť takéto hranice je v podstate nemožné, keďže zdravie predstavuje momentálny stav, ktorý neustále ovplyvňuje celý komplex endogénnych a exogénnych faktorov. Tie sa podieľajú na postupných, ale aj náhlych zmenách zdravotného stavu. Okrem toho treba tiež upozorniť, že tieto prechody nemusia byť (a často ani nie sú) jednosmerné. Značná dynamika a multistavovosť spolu s možnosťou viacnásobnej reverzibility tvoria nestabilný systém, ktorého objektívne empirické hodnotenie je veľmi problematické. Jednotlivé aspekty dokážeme v podstate len odhadovať na základe určitých vyjadrení (pocitov) danej osoby a vedomostí o jej zdravotnom stave.

Vo všeobecnosti sa vychádza z troch pohľadov na zdravie [6]. Osobe je diagnostikovaná choroba alebo osoba sama deklaruje, že má alebo nemá určité ochorenie. V druhom prípade sa zisťuje prítomnosť alebo neprítomnosť nejakého zdravotného obmedzenia a v poslednom, treťom koncepte sa analyzujú niektoré ďalšie charakteristiky zdravia, ako napríklad osobou vnímaná kvalita jej zdravia a miera spokojnosti s ním. Celý tento proces je však s výnimkou prevalencie a incidencie ochorenia značne subjektívny. Ak odhliadneme od pravidelných zdravotníckych štatistík, asi najčastejšie využívaným nástrojom sú rôzne výberové zisťovania [2, 3]. Ich cieľom je získať ďalšie informácie o zdravotnom stave, ktoré sa v bežne (a pravidelne) zisťovaných zdravotníckych štatistikách nedajú získať. Vo všeobecnosti môžeme podľa prístupu k danej problematike hovoriť o dvoch základných typoch zisťovaní [2]. Prvým je Health Interview Survey (HIS), v ktorom respondent sám odpovedá opytovateľovi na otázky o svojom vlastnom zdravotnom stave bez možnosti nejakej verifikácie či konfrontácie s lekársym vyšetrením. Aplikuje sa pritom celá škála spôsobov (PAPI, CAPI, CASI², telefonický rozhovor a pod.), ako získať takéto do určitej miery subjektívne druhy údajov. Health

² Ide o techniky získavania odpovedí v dotazníkových zisťovaniach. Podľa spôsobu zberu ide napr. o PAPI (Paper and Pencil Interviewing), CAPI (Computer-Assisted Personal Interviewing), CASI (Computer-Assisted Self Interviewing).

Examination Survey (HES) predstavuje kvalitatívne iný typ zisťovania. Je oveľa náročnejšie, avšak svojou koncepciou prináša objektívnejšie informácie, aj preto, že najčastejšie sa spája s lekársnym vyšetrením (vrátane biochemických odberov) a prípadne sa dopĺňa ešte ďalšími údajmi získanými zo zdravotnej dokumentácie.

Do kategórie HIS zisťovaní možno zaradiť aj časť otázok používaných v EU SILC a EHIS. Konkrétne ide o tri otázky pokrývajúce tzv. minimálny európsky modul zdravia (MEHM) snažiaci sa sledovať kvalitatívny aspekt vlastného zdravia respondenta, prítomnosť chronických a dlhotrvajúcich ochorení a limitáciu bežných denných aktivít. Jeho konštrukcia je výsledkom snáh disponovať nástrojom na zbieranie harmonizovaných údajov o niektorých aspektoch zdravotného stavu členských štátov EÚ28 a niektorých ďalších krajín (Nórsko, Švajčiarsko a Island), v ktorých sa zisťovania pravidelne vykonávajú. Prostredníctvom vhodných metodických nástrojov (predovšetkým tzv. Sullivanovej metódy), kombináciou štruktúry odpovedí na položené otázky a vybraných funkcií prierezových tabuliek úmrtnosti sa následne konštruujú syntetické indikátory umožňujúce analyzovať a porovnávať určité aspekty zdravotného stavu. Veľký význam tohto postupu bol potvrdený prijatím zdravých rokov života (healthy life years) medzi základné štruktúrne indikátory EÚ28. Otázkou však zostáva, či tento prístup skutočne spĺňa očakávania do neho vložené a či ho možno aplikovať bez rizika akýchkoľvek interpretačných a informačných kolízií s realitou.

Cieľom článku je snaha poukázať na niektoré problematické závery, ktoré prináša použitie modulu MEHM v spojitosti so Slovenskom a jeho postavením v európskom priestore. Zdravotný stav hodnotený práve predmetnou koncepciou je konfrontovaný so strednou dĺžkou života ako objektívnym indikátorom odzrkadľujúcim epidemiologickú situáciu v danej populácii. Predpokladáme, že priaznivé úmrtnostné pomery by sa mali odzrkadliť aj na kvalite zdravotného stavu, a teda v krajinách s najdlhším životom by aj jednotlivé koncepty zdravia mali byť najlepšie alebo aspoň patriť k najlepším v Európe. Slovensko sa dlhodobo z pohľadu hodnotenia zdravotného stavu nachádza na pomyselnom konci európskeho rebríčka. Cieľom práce je aj snaha objasniť príčiny tohto stavu. V neposlednom rade sa tiež zameriame na paradoxnú situáciu, ktorá vzniká použitím toho istého modulu v zisťovaniach EU SILC a EHIS. Keďže obe pracujú s tými istými medzinárodne harmonizovanými otázkami, logicky sa dá predpokladať, že aj získané výsledky by mali byť rovnaké alebo aspoň podobné. Ako však naznačil vo svojej práci Mészáros [10], v roku 2009 bola situácia úplne odlišná. My sme sa snažili toto zistenie verifikovať alebo zamietnuť v spojitosti s druhou vlnou EHIS z roku 2014 a v tom istom roku vykonanom zisťovaní EU SILC.

2. DÁTA A METÓDY

Pozornosť venovaná zdraviu, zdravotnému stavu jeho vývoju v kontexte rozvoja spoločností v rámci EÚ28 prispela k snahám vytvoriť jednotný harmonizovaný nástroj, prostredníctvom ktorého by bolo možné analyzovať a porovnávať stav, sledovať jeho zmeny v širšom medzinárodnom kontexte. Celkovo môžeme hovoriť o dvoch zisťovaniach, ktoré spĺňajú požiadavky metodiky na meranie zdravotného stavu európskych populácií vypracovanej v rámci programu the European Health Expectancy Monitoring Unit (EHEMU). Ide o zisťovanie EU SILC a EHIS. Pri medzinárodnom porovnávaní (v rámci EÚ28) pracujeme len s údajmi, ktoré poskytlo zisťovanie EU SILC, pretože zisťovanie EHIS sa vo všetkých členských štátoch

uskutočnilo až v roku 2014 a v čase písania tohto článku neboli ešte primárne údaje voľne dostupné na vedecké účely.

V spojitosti s hlavnými cieľmi je pozornosť venovaná trom otázkam:

- 1) Ako by ste celkovo (vo všeobecnosti) zhodnotili vaše zdravie?
- 2) Máte nejaké dlhotrvajúce ochorenia alebo dlhotrvajúci zdravotný problém?
- 3) Do akej miery ste boli v priebehu minimálne posledných 6 mesiacov obmedzovaný vo vašich bežných činnostiach kvôli zdravotnému problému?

Vstupné údaje sme získali z databáz EUROSTAT-u, ako aj zo špecializovanej databázy JA:EHLEIS (the European Health and Life Expectancy Information System). Okrem týchto zdrojov sme za rok 2014 využili aj primárne údaje zo zisťovaní EU SILC a EHIS, ktoré nám poskytol Štatistický úrad SR.

Na výpočet strednej dĺžky života v dobrom zdraví bez chronických ochorení a obmedzení v bežných denných aktivitách využívame Sullivanovu metódu [7, 13]. Ide o jednoduchú, dátovo nenáročnú, ale na druhej strane medzinárodne uznávanú a na tieto účely široko používanú metódu. Vstupnými údajmi sú tabuľkové počty osôb (l_x) dožívajúcich sa presného veku (x) z prierezových úmrtnostných tabuliek, ďalej tabuľkový počet žijúcich osôb (${}_nL_x$), resp. počet človekorokov, ktoré osoba prežije medzi dvomi presnými vekmi (x , $x + n$), a prevalencia sledovaného zdravotného stavu (s_x) vo veku (x) z výberového zisťovania. Napríklad pre indikátor zdravé roky života (HLY) možno Sullivanov prístup zapísať v tvare:

$$HLY_x = \frac{\sum_x \omega^{-1} (L_x \times s_x)}{l_x}$$

kde s_x predstavuje podiel osôb vo veku x bez obmedzenia bežných denných aktivít.

Na účel presnej identifikácie príspevkov jednotlivých vekových skupín, efektov úmrtnostných pomerov a zisteného zdravotného stavu k rozdielom počtu zdravých rokov života medzi Slovenskom a Švédskom využívame dekompozičnú techniku navrhnutú kolektívom Andreev, Shkolnikov a Begun [1]. Švédsko v tomto prípade reprezentuje európske populácie s najlepšimi úmrtnostnými pomermi a súčasne aj najvyššími hodnotami zdravých rokov života.

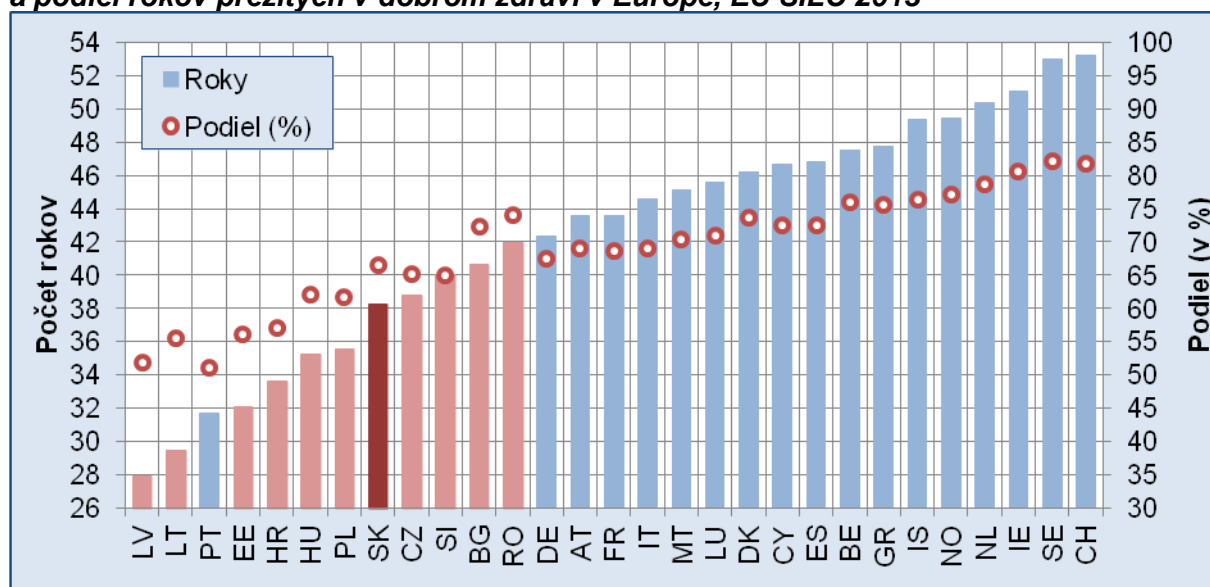
3. POZÍCIA SLOVENSKA V EURÓPE Z POHĽADU ÚMRTNOSTI A ZDRAVOTNÉHO STAVU

Slovensko dlhodobo od začiatku 90. rokov aj napriek pozitívnemu vývoju úmrtnostných pomerov pomerne výrazne zaostáva za demograficky najvyspelejšími krajinami Európy. Napríklad vo Švajčiarsku stredná dĺžka života pri narodení mužov podľa údajov Eurostat-u prekročila v roku 2014 už hranicu 81 rokov, kým na Slovensku dosahovala len niečo viac ako 73 rokov. Naopak, najhoršie na tom boli v rámci EÚ28 muži v Lotyšsku a Litve (69 rokov), nasledovaní mužmi v Bulharsku a Rumunsku (71 rokov). V skupine žien by mali pred sebou najdlhší život pri nezmenených úmrtnostných pomeroch ženy v Španielsku. Stredná dĺžka života tam v roku 2014 presiahla hranicu 86 rokov. Na Slovensku v rovnakom období dosahovala približne o 5,5 roka menej a v Bulharsku bola stredná dĺžka života žien v porovnaní so Španielskom kratšia dokonca až o viac ako 8 rokov.

Stredná dĺžka života v dobrom zdravotnom stave nás informuje o tom, koľko rokov zostáva osobe potenciálne prežiť v dobrom zdraví pri nezmenenej úmrtnosti a nezmenených pomeroch v charaktere zdravotného stavu danej populácie. Kombinuje tak údaje z úmrtnostných tabuliek (a teda intenzitu úmrtnosti) a štruktúru odpovedí z využívaného zisťovania o zdraví (a teda zistený zdravotný stav populácie). Okrem samotného absolútneho číselného vyjadrenia počtu rokov je veľmi dôležitou informáciou aj štruktúra zostávajúcich rokov podľa kvality zdravotného stavu. Hovorí o tom, akú časť z potenciálnych rokov života môže prežiť muž alebo žena v dobrom zdraví.

Najvyššiu hodnotu strednej dĺžky života v dobrom zdraví v presnom veku 16 rokov dosahujú niektoré krajiny severnej a západnej Európy. Ide predovšetkým o Švajčiarsko, Švédsko, Írsko, Holandsko, Nórsko (pozri graf 1 a 2). Muži v týchto krajinách majú šancu prežiť okolo 50 rokov v dobrom zdraví. U žien prvé tri spomínané krajiny dokonca prekračujú hranicu 52 rokov. Rovnako populácie týchto krajín sa vyznačujú najvyšším podielom rokov (70 – 80 %), ktoré môžu prežiť vo veľmi dobrom a dobrom zdraví z celkového počtu potenciálnych rokov.

Graf č. 1: Stredná dĺžka života mužov v presnom veku 16 rokov v dobrom zdraví a podiel rokov prežitých v dobrom zdraví v Európe, EU SILC 2013



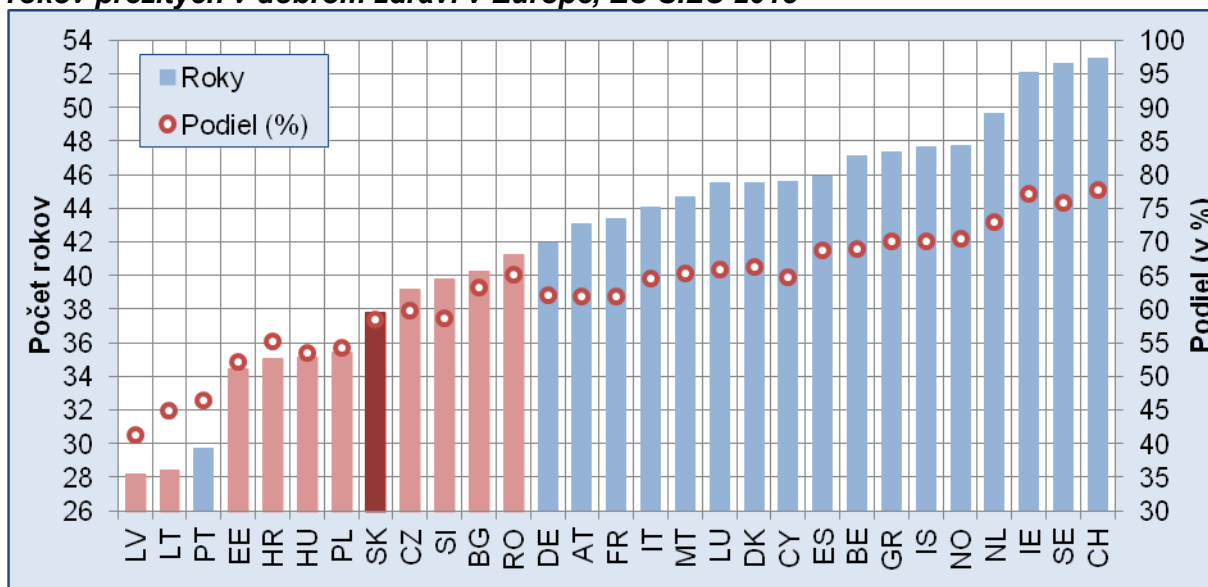
Poznámka: BE – Belgicko, BG – Bulharsko, CZ – Česko, DK – Dánsko, DE – Nemecko, EE – Estónsko, FI – Fínsko, IE – Írsko, GB – Veľká Británia, GR – Grécko, ES – Španielsko, FR – Francúzsko, CH – Švajčiarsko, HR – Chorvátsko, IS – Island, IT – Taliansko, CY – Cyprus, LV – Lotyšsko, LT – Litva, LU – Luxembursko, HU – Maďarsko, MT – Malta, NL – Holandsko, NO – Nórsko, RO – Rumunsko, AT – Rakúsko, PL – Poľsko, PT – Portugalsko, SI – Slovinsko, SK – Slovensko, SE – Švédsko.

Zdroj údajov: JA:EHLEIS, výpočty autora

Naopak, najnižšiu strednú dĺžku života v dobrom zdraví dlhodobo dosahuje skupina pobaltských štátov (najmä Lotyšsko, Litva) a Portugalsko. Hlboko pod hranicou 40 rokov sa nachádzajú aj ďalšie krajiny bývalého východného bloku. Znamená to tiež, že ich populácie majú aj najnižší podiel rokov, ktoré môžu prežiť v dobrom zdraví. Špecifická situácia je však v Bulharsku a Rumunsku, kde vďaka vysokým podielom respondentov, ktorí označujú svoje zdravie v pozitívnom svetle, identifikujeme aj najvyššiu strednú dĺžku života v dobrom zdraví spomedzi všetkých

sledovaných krajín postsocialistického bloku. Slovensko so svojimi približne 38 rokmi patrí v európskom priestore ku krajinám s nižšou hodnotou strednej dĺžky života v zdraví, čo do určitej miery korešponduje aj s jeho postavením z pohľadu úmrtnostných pomerov.

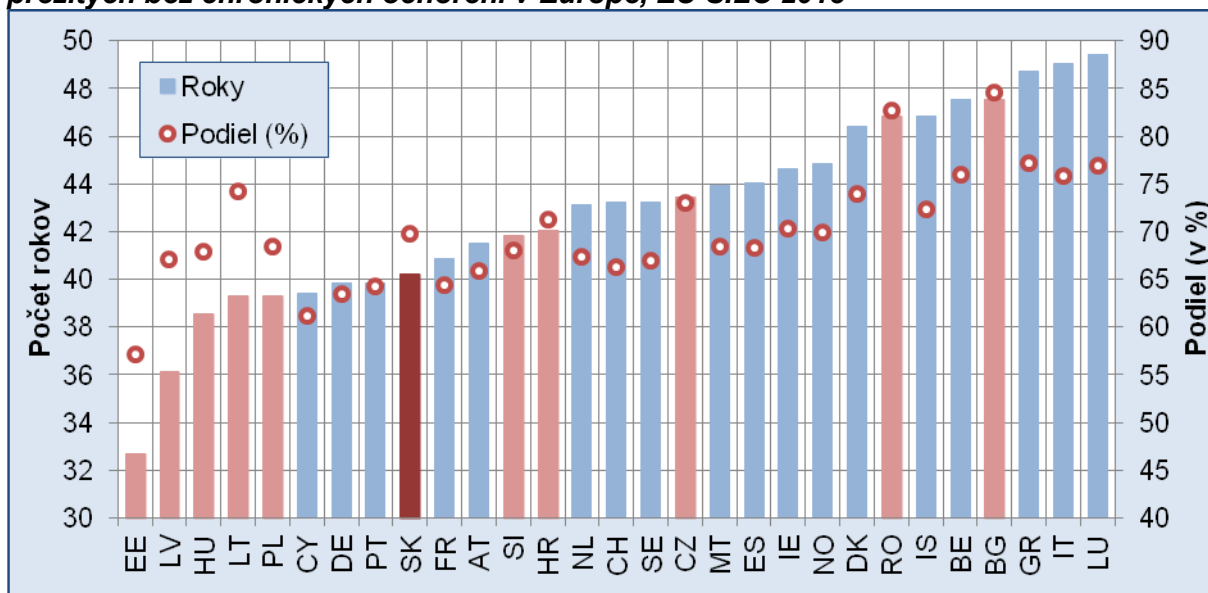
Graf č. 2: Stredná dĺžka života žien v presnom veku 16 rokov v dobrom zdraví a podiel rokov prežitých v dobrom zdraví v Európe, EU SILC 2013



Poznámka: Vysvetlivky skratiek štátov pozri v poznámke pod grafom č. 1.

Zdroj údajov: JA:EHLEIS, výpočty autora

Graf č. 3: Stredná dĺžka života mužov v presnom veku 16 rokov a podiel rokov prežitých bez chronických ochorení v Európe, EU SILC 2013



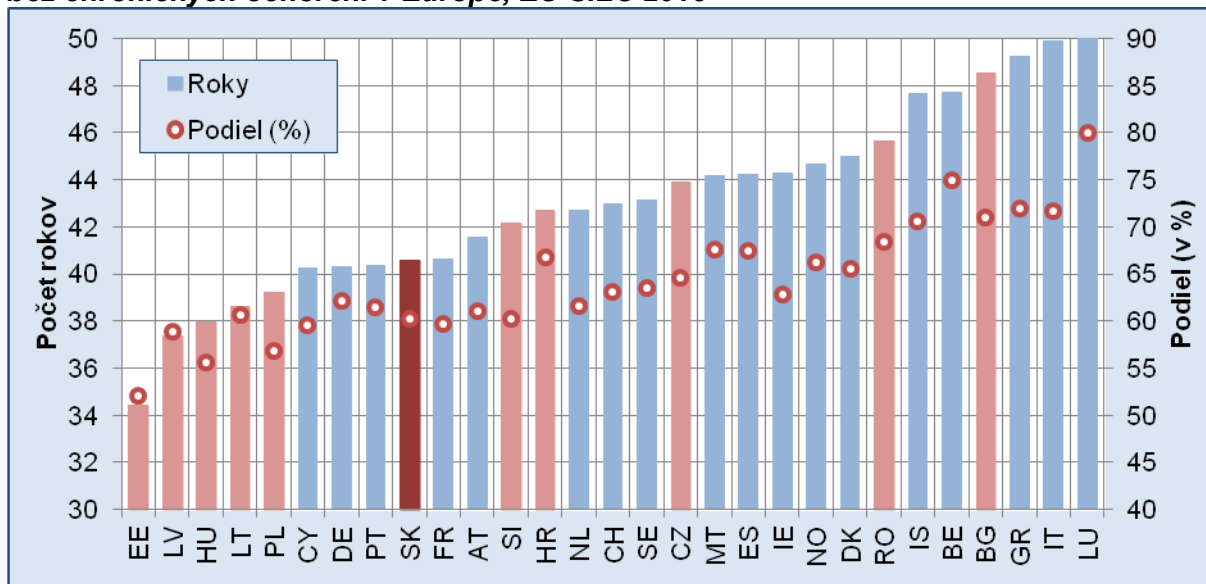
Poznámka: Vysvetlivky skratiek štátov pozri v poznámke pod grafom č. 1.

Zdroj údajov: JA:EHLEIS, výpočty autora

Jednoznačne najnižšiu strednú dĺžku života bez dlhodobých ochorení a zdravotných problémov majú podľa výsledkov EU SILC 2013 a úmrtnostných pomerov v danom roku populácie všetkých troch pobaltských štátov, ďalej Maďarska a Poľska (pozri graf 3 a 4). Paradoxne horšie ako na Slovensku sú na tom aj muži

a ženy v niektorých krajinách (Nemecko, Portugalsko a Cyprus), ktoré majú dlhodobo lepšie úmrtnostné pomery. Podobná situácia ako na Slovensku je aj v Rakúsku a Francúzsku. Na druhej strane Bulharsko a Rumunsko aj napriek svojej pomerne vysokej úmrtnosti (pozri predtým) patria na popredné priečky. Je to dôsledok vysokého zastúpenia odpovedí reflektujúcich absenciu dlhodobých ochorení a zdravotných problémov.

Graf č. 4: Stredná dĺžka života žien v presnom veku 16 rokov a podiel rokov prežitých bez chronických ochorení v Európe, EU SILC 2013



Vysvetlivky skratiek štátov pozri v poznámke pod grafom č. 1.

Zdroj údajov: JA:EHLEIS, výpočty autora

Prítomnosť zdravotného problému alebo dlhodobého ochorenia nemusí byť pre osobu automaticky limitujúcim prvkom pri vykonávaní bežných denných aktivít. Aj preto tretím aspektom MEHM modelu je prítomnosť, resp. miera obmedzenia bežných aktivít v posledných šiestich mesiacoch v dôsledku prítomnosti nejakého zdravotného problému. Respondenti mali možnosť deklarovat' úplnú samostatnosť bez obmedzenia alebo dva stupne obmedzenia: veľmi obmedzený a obmedzený, ale nie veľmi.

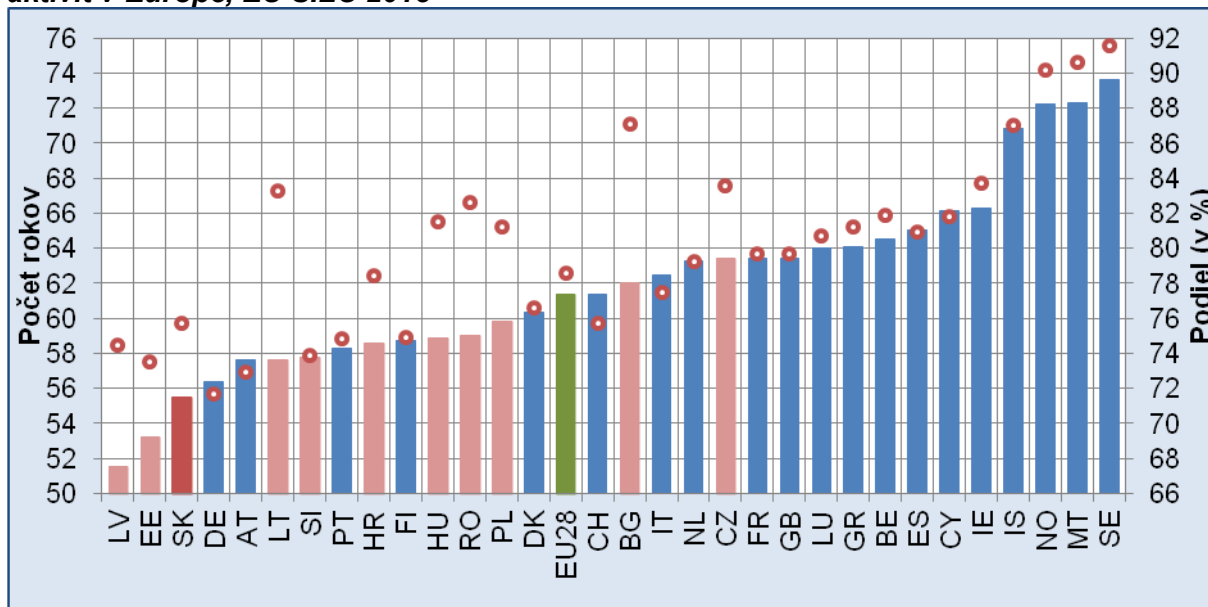
Najlepšiu kvalitu zdravia z pohľadu limitácie bežných denných aktivít vykazujú niektoré severské štáty (Nórsko, Švédsko, Island), Malta, Veľká Británia, Cyprus, ale aj Bulharsko. U mužov v týchto štátoch podiel osôb bez obmedzenia bežných denných aktivít vo veku 16 a viac rokov prekračuje hranicu 80 % a u žien je nad úrovňou 75 %. Slovensko sa dlhodobo zaraďuje medzi najhoršie krajiny EÚ28. Podľa údajov z roku 2013 len 73 % mužov a 64 % žien vo veku 16 a viac rokov uviedlo, že nie sú nijako obmedzovaní pri vykonávaní bežných denných aktivít. Najhoršia situácia bola v Lotyšsku, Estónsku, Nemecku a u žien aj v Holandsku. Výrazné zaostávanie Slovenska sa pritom formuje už v mladšom veku, pričom v seniorskom veku už majú prevahu skupiny, ktoré priznávajú nejakú formu obmedzenia bežných denných aktivít. V EÚ28 to nastáva u mužov až vo veku 75 – 79 rokov a u žien vo vekovej skupine 70 – 74 rokov.

Kombináciou pomerne nízkej strednej dĺžky života a vysokého podielu osôb deklarujúcich nejaké obmedzenie bežných denných aktivít z dôvodu zdravotných

problémov sa Slovensko zaraďuje v EÚ28 medzi krajiny s najhoršími hodnotami indikátora zdravé roky života. U mužov sa dokonca nachádzame na poslednom mieste (graf 5). Len o niečo lepšia bola situácia u žien (graf 6). Podľa posledných dostupných údajov sa tak na Slovensku počet rokov, ktoré môže pri zachovaní úmrtnostných pomerov a štruktúry odpovedí na danú otázku osoba prežiť bez limitácie, pohyboval na úrovni 54,6 roka u mužov a 55,6 roka u žien. Znamená to, že len asi tri štvrtiny z celkovej strednej dĺžky života pri narodení u mužov a 68 % u žien prežijú osoby bez limitácie denných aktivít. Priemer za EÚ28 sa pritom pohybuje na úrovni 61,4 roka (takmer 79 %) v mužskej a 61,8 roka (takmer 74 %) v ženskej časti populácie. S výnimkou Česka a u žien prekvapivo aj Bulharska je v ostatných krajinách bývalého východného bloku hodnota indikátora zdravé roky života podpriemerná.

Okrem postsocialistických krajín je však situácia problematická aj v niektorých demograficky vyspelých štátoch bývalého západného bloku. Ide predovšetkým o Nemecko, Rakúsko, Portugalsko, Fínsko (graf 5) a u žien dokonca aj o Švajčiarsko a Holandsko (graf 6), kde počet zdravých rokov neprekračuje hranicu 60 rokov. Keďže vo väčšine z nich stredná dĺžka života dosahuje vysoké hodnoty, tieto krajiny patria medzi populácie vyznačujúce sa najnižším podielom rokov, ktoré prežijú bez obmedzenia bežných denných aktivít. Ak sa na poradie krajín EÚ28 pozrieme práve optikou podielu rokov bez limitácie, potom Slovensko už nepatrí na samý koniec rebríčka. Ešte nižšiu úroveň tohto ukazovateľa dosahujú u mužov práve niektoré populácie s výrazne lepšími úmrtnostnými pomermi (napr. Nemecko, Rakúsko, Fínsko a pod.). Podobne u žien zisťujeme, že Slovensko nie je jedinou krajinou, ktorá dosahuje taký nízky podiel zdravých rokov života z celkovej hodnoty strednej dĺžky života pri narodení (pozri graf 6).

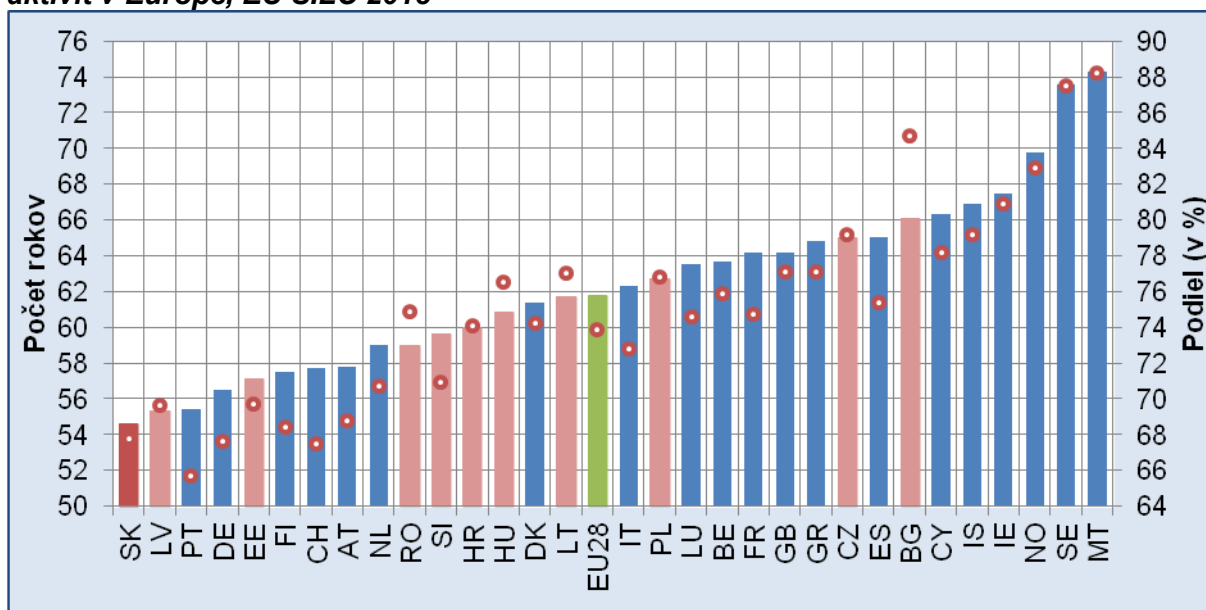
Graf č. 5: Zdravé roky života mužov a podiel rokov bez obmedzenia bežných denných aktivít v Európe, EU SILC 2013



Vysvetlivky skratiek štátov pozri v poznámke pod grafom č. 1.

Zdroj údajov: Eurostat, výpočty autora

Graf č. 6: Zdravé roky života žien a podiel rokov bez obmedzenia bežných denných aktivít v Európe, EU SILC 2013



Vysvetlivky skratiek štátov pozri v poznámke pod grafom č. 1.

Zdroj údajov: Eurostat, výpočty autora

4. PREČO JE POČET ZDRAVÝCH ROKOV ŽIVOTA NA SLOVENSKU TAKÝ NÍZKY?

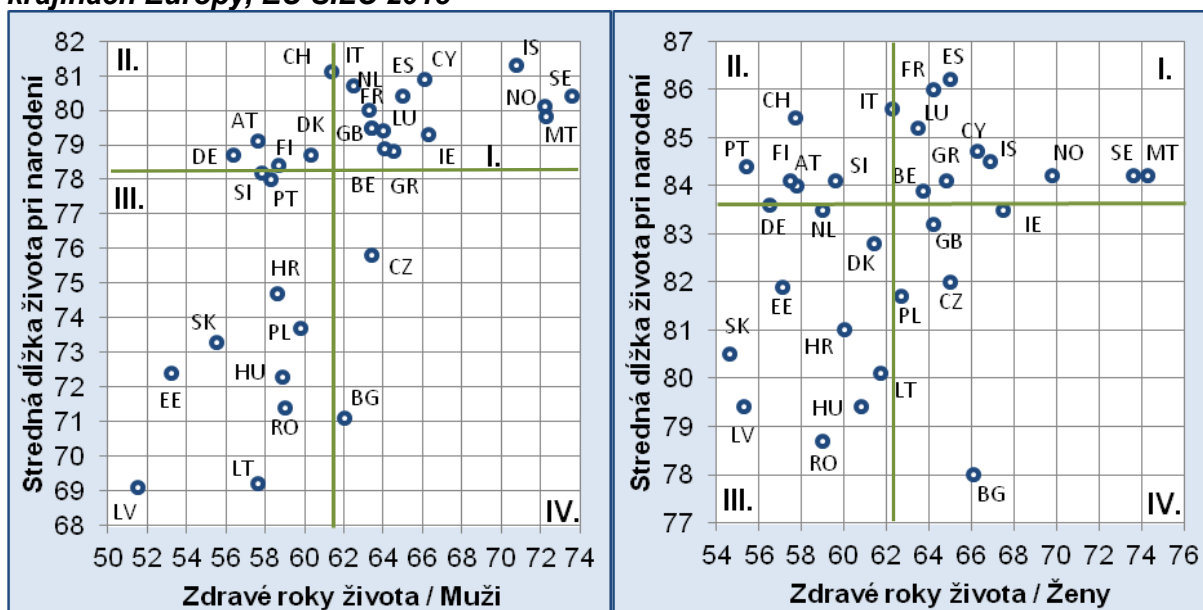
Postavenie Slovenska z pohľadu počtu zdravých rokov života podľa výsledkov zisťovania EU SILC je dlhodobo veľmi nepriaznivé. Slovensko sa už tradične nachádza na posledných priečkach EÚ28. Keďže uvedený indikátor patrí medzi európske štrukturálne ukazovatele a jeho význam zdôrazňuje aj lisabonská stratégia, namiesto je otázka, prečo je pozícia Slovenska taká nepriaznivá, a to aj v porovnaní s krajinami, ktoré sa dlhodobo vyznačujú ešte horšími úmrtnosťnými pomermi ako Slovensko? Pozornosť tomuto paradoxu treba venovať aj v spojitosti s účelom daného indikátora, ktorým je nielen monitorovať zdravie a zdravotný stav populácie, ale aj hodnotiť kvalitu života a prežitých rokov, prispievať k empirickému zhodnoteniu zamestnateľnosti starších osôb a tiež monitorovať pokrok dosiahnutý pri zvyšovaní prístupnosti, kvality a udržateľnosti zdravotnej starostlivosti.

Predchádzajúce výsledky však prinášajú do určitej miery rozporuplné zistenia. Ak budeme predpokladať, že pokles úmrtnosti prebiehal a prebieha súčasne so zlepšovaním zdravotného stavu, potom by sa viaceré krajiny s najvyššou strednou dĺžkou života pri narodení mali nachádzať aj na popredných priečkach z pohľadu zdravých rokov života. Naopak, očakávali by sme, že počet rokov, ktoré prežijú osoby bez obmedzenia v bežných denných aktivitách, v krajinách s vyššou úmrtnosťou bude dosahovať výrazne podpriemernú úroveň. Situácia je však oveľa komplikovanejšia.

Ak budeme sledovať vzťah medzi strednou dĺžkou života pri narodení a zdravými rokmi života, môžeme hovoriť o štyroch skupinách krajín podľa ich pozície vzhľadom na priemerné hodnoty za celú EÚ28 (pozri graf 7 a 8). V prvom kvadrante sa nachádzajú populácie s nadpriemernou strednou dĺžkou života pri narodení, ako aj nadpriemerným počtom zdravých rokov života. Ide o väčšinu krajín bývalého západného bloku, no nie je tu ani jedna krajina zo socialistického bloku. Opakom je

tretí kvadrant s podpriemernou hodnotou strednej dĺžky života a zdravých rokov života. Práve tu leží Slovensko, ako aj väčšina postsocialistických krajín, čo sa svojím spôsobom nevymyká už spomínanému logickému prepojeniu kratšieho života a horšej kvality zdravia. Špecifickým je však štvrtý a druhý kvadrant. Kým v poslednom je počet zdravých rokov nadpriemerný, ale stredná dĺžka života podpriemerná, v druhom je situácia opačná. K vysokej strednej dĺžke života pri narodení sa priraduje podpriemerná úroveň počtu rokov prežitých v zdraví. Patria sem napríklad nemecky hovoriace krajiny (Nemecko, Rakúsko, Švajčiarsko), ale aj Portugalsko, Fínsko, Dánsko či Slovinsko.

Grafy č. 7 a 8: Vzťah strednej dĺžky života a zdravých rokov života vo vybraných krajinách Európy, EU SILC 2013



Poznámka: Zelené čiary prezentujú priemerné hodnoty za EÚ28; vysvetlivky skratiek štátov pozri v poznámke pod grafom č. 1.

Zdroj údajov: Eurostat

Kľúč na pochopenie nízkeho počtu zdravých rokov života na Slovensku, ale aj v niektorých demograficky najvyspelejších krajinách Európy sa skrýva v samotnej metodike výpočtu tohto indikátora, resp. vo vstupných údajoch. Na konštrukciu sa využíva tzv. Sullivanova metóda [7, 13], ktorá pracuje s funkciami prierezových tabuliek úmrtnosti prezentujúcich úmrtnostné pomery danej populácie a s vekovo-špecifickými podielmi osôb, ktoré uviedli, že nemajú žiadne obmedzenie bežných denných aktivít pochádzajúce z výsledkov zisťovania EU SILC. Samotný rozdiel v počte zdravých rokov života je potom výsledkom odlišnej strednej dĺžky života a charakteru odpovedí na túto otázku v zisťovaní.

Prvé priblíženie vplyvu úmrtnostných pomerov a charakteru odpovedí možno urobiť na základe jednoduchšej konštrukcie. Podľa našich údajov konštruovaných z databázy JA:EHLEIS Slovensko zaostáva za Švédskom v počte zdravých rokov života u mužov o 11,7 roka a u žien dokonca o viac ako 12 rokov (pozri tab. 1). Ak sa však pozrieme na hodnoty strednej dĺžky života, rozdiely sú výrazne menšie. Z dôvodu horších úmrtnostných pomerov sa muži na Slovensku v presnom veku 16 rokov môžu pri nezmenenej úmrtnosti dožiť o približne 7 rokov a ženy o 3,4 roka menej ako rovnako staré osoby vo Švédsku. V prípade neexistencie žiadnych

rozdielov v úmrtnostných pomeroch medzi Slovenskom a Švédskom sa diferencia v hodnote zdravých rokov života zníži u mužov len o necelé 2 roky a u žien dokonca len o 0,5 roka. Úplne odlišná situácia vznikne, ak by muži i ženy na Slovensku odpovedali na otázku o limitácii bežných denných aktivít v rovnakej štruktúre ako vo Švédsku. Pri naplnení tohto modelu by počet zdravých rokov života vzrástol u mužov o 8 a u žien o viac ako 10 rokov (tab. 1).

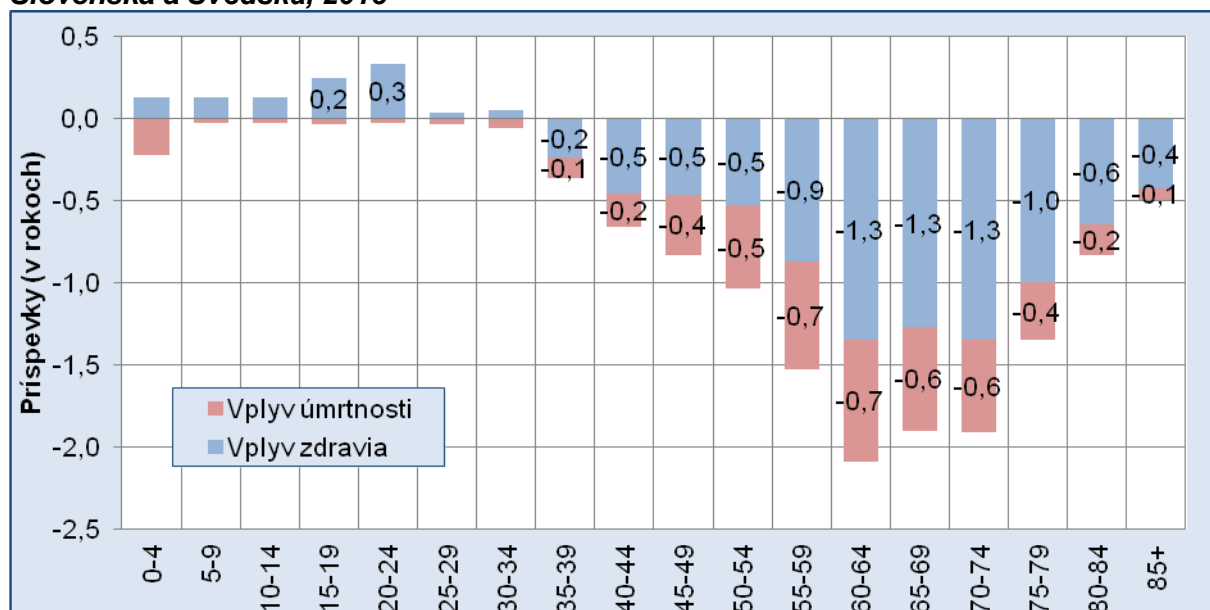
Tabuľka č. 1: Modelové situácie vývoja hodnôt zdravých rokov života na Slovensku podľa úrovne úmrtnostných pomerov a zdravotného stavu vo Švédsku, EU SILC 2013

Model	Zdravé roky života HLY ³ (v presnom veku 16 rokov)	Rozdiel HLY v porovnaní so Švédskom (v rokoch)	Predĺženie HLY (v rokoch)
<i>Muži</i>			
Úmrtnosť Slovensko, zdravie Slovensko	39,5	-11,7	x
Úmrtnosť Švédsko, zdravie Slovensko	41,4	-9,8	1,9
Úmrtnosť Slovensko, zdravie Švédsko	47,5	-3,7	8,0
Úmrtnosť Švédsko, zdravie Švédsko	51,2	0,0	11,7
<i>Ženy</i>			
Úmrtnosť Slovensko, zdravie Slovensko	39,3	-12,2	x
Úmrtnosť Švédsko, zdravie Slovensko	39,8	-11,7	0,5
Úmrtnosť Slovensko, zdravie Švédsko	49,6	-1,9	10,3
Úmrtnosť Švédsko, zdravie Švédsko	51,5	0,0	12,2

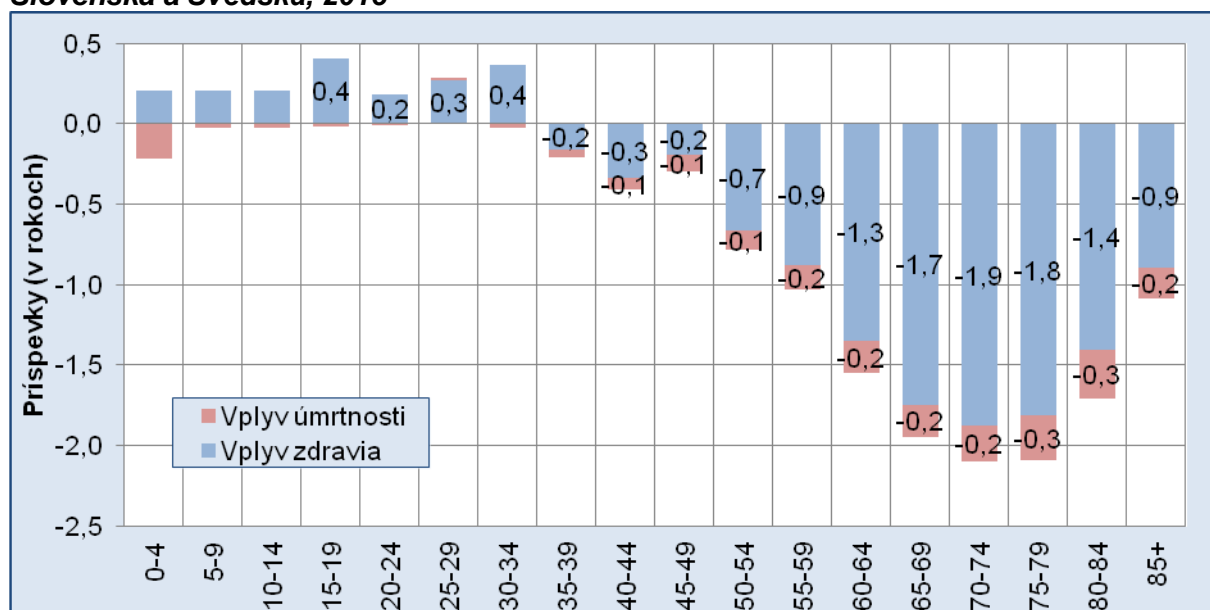
Zdroj údajov: JA:EHLEIS, výpočty autora

Už tieto výsledky pomerne jednoznačne poukázali na kľúčový význam štruktúry odpovedí týkajúcich sa limitácie bežných denných aktivít na hodnotu indikátora zdravé roky života. Definitívne tento záver potvrdila viacrozmerná dekompozícia rozdielov zdravých rokov života podľa veku a pohlavia. Pracovali sme s metodickým konceptom navrhnutým kolektívom autorov Andreev, Shkolnikov a Begun [1]. Ten rozkladá existujúci rozdiel medzi jednotlivé vekové skupiny, pričom rozlišuje efekt (vplyv) diferencií v úmrtnostných pomeroch a efekt zdravia, teda rozdielov v štruktúre odpovedí na otázky merajúce kvalitu zdravotného stavu v použitých zisťovaniach. Výsledky znázorňujú grafy 9 a 10. Je z nich zrejmé, že hlavným faktorom výrazného zaostávania Slovenska za Švédskom v hodnote zdravých rokov života je predovšetkým charakter odpovedí respondentov na danú otázku. U mužov sa podieľajú približne 60 % (čiže 7,6 roka) a u žien dokonca viac ako 80 % (9,5 roka). Zvyšných 40 %, resp. necelých 20 % pripadá na diferencie spôsobené zaostávaním Slovenska v úmrtnostných pomeroch.

³ The Healthy Life Years (HLY).

Graf č. 9: Dekompozícia rozdielu v počte zdravých rokov života medzi mužmi Slovenska a Švédska, 2013

Zdroj údajov: JA:EHLEIS, výpočty autora

Graf č. 10: Dekompozícia rozdielu v počte zdravých rokov života medzi ženami Slovenska a Švédska, 2013

Zdroj údajov: JA:EHLEIS, výpočty autora

Dôležitým je tiež rozloženie príspevkov jednotlivých vekových skupín. Odzrkadľuje jednak horšie úmrtnostné pomery mužov a žien v seniorskom veku na Slovensku, ako aj dynamicky sa zvyšujúci podiel osôb deklarujúcich istú formu obmedzenia. Preto u mužov vo veku 60 a viac rokov nachádzame takmer dve tretiny a u žien takmer 90 % z celkového rozdielu zdravých rokov života.

5. DVE ZISŤOVANIA, DVA RÔZNE VÝSLEDKY?

Na Slovensku v rokoch 2009 a 2014 bolo možné na výpočet zdravých rokov života, ako aj ďalších indikátorov zdravotného stavu použiť dva nezávislé zdroje údajov. Išlo o spomínaný EU SILC a tiež druhú vlnu EHIS. Keďže v oboch

zist'ovaniach sa použil rovnaký súbor otázok v module MEHM a rovnaký bol aj spôsob zist'ovania odpovedí, dalo by sa predpokladať, že rovnaké alebo aspoň podobné z hľadiska určitej miery spoľahlivosti údajov by mali byť aj získané výsledky. Ako však poukázal Mészáros [10], medzi výsledkami EU SILC 2009 a EHIS 2009 existovali pomerne veľké rozdiely, ktoré sa následne odzrkadlili aj na hodnotách vypočítaných indikátorov zdravia. Týkalo sa to predovšetkým prítomnosti chronických ochorení, keď v zist'ovaní EHIS 2009 bol výrazne nižší podiel osôb bez chronického ochorenia. Podobne aj v otázkach zameriavajúcich sa na subjektívne hodnotenie vlastného zdravotného stavu, resp. mieru obmedzenia bežných denných aktivít možno identifikovať v zist'ovaní EHIS 2009 určitú náklonnosť respondentov voliť kategórie odpovedí znamenajúce nižšiu kvalitu zdravia [10].

V roku 2014 sa uskutočnila druhá vlna zist'ovania EHIS, do ktorej sa už zapojili všetky členské štáty EÚ, a rovnako aj zist'ovanie EU SILC, pričom otázky použité v oboch zist'ovaniach v rámci modulu MEHM boli plne harmonizované. Ako ukazujú štruktúry odpovedí na jednotlivé otázky pre osoby vo veku 16 a viac rokov, aj v tomto prípade môžeme identifikovať určité rozdiely. Testovacia charakteristika (chí-kvadrát test, 5 % hladina významnosti) potvrdila, že u oboch pohlaví a vo všetkých troch konceptoch boli rozdiely v štruktúre odpovedí medzi zist'ovaniami štatisticky významné. Najmä v prípade prítomnosti chronických ochorení a miery obmedzenia bežných denných aktivít sú tieto disproporcie zrejme už zo samotného porovnania frekvenčných tabuliek odpovedí (tab. 2 a tab. 3).

Tabuľka č. 2: Štruktúra odpovedí na otázku „Máte nejaké dlhodobé ochorenie alebo dlhodobý zdravotný problém?“ v zist'ovaniach EHIS a EU SILC na Slovensku v roku 2014 (%)

Zist'ovanie	Pohlavie	Áno	Nie
EHIS 2014	muži	48,3	51,7
	ženy	59,5	40,5
	spolu	54,1	45,9
EU SILC 2014	muži	26,2	73,8
	ženy	34,0	66,0
	spolu	30,3	69,7

Zdroj údajov: ŠÚ SR (primárne údaje EHIS 2014, EU SILC 2014), triedenie a výpočty autora

Tabuľka č. 3: Štruktúra odpovedí na otázku „Do akej miery ste boli v priebehu minimálne posledných 6 mesiacov obmedzovaný kvôli zdravotnému problému v činnostiach, ktoré ľudia bežne vykonávajú?“ v zist'ovaniach EHIS a EU SILC na Slovensku v roku 2014 (%)

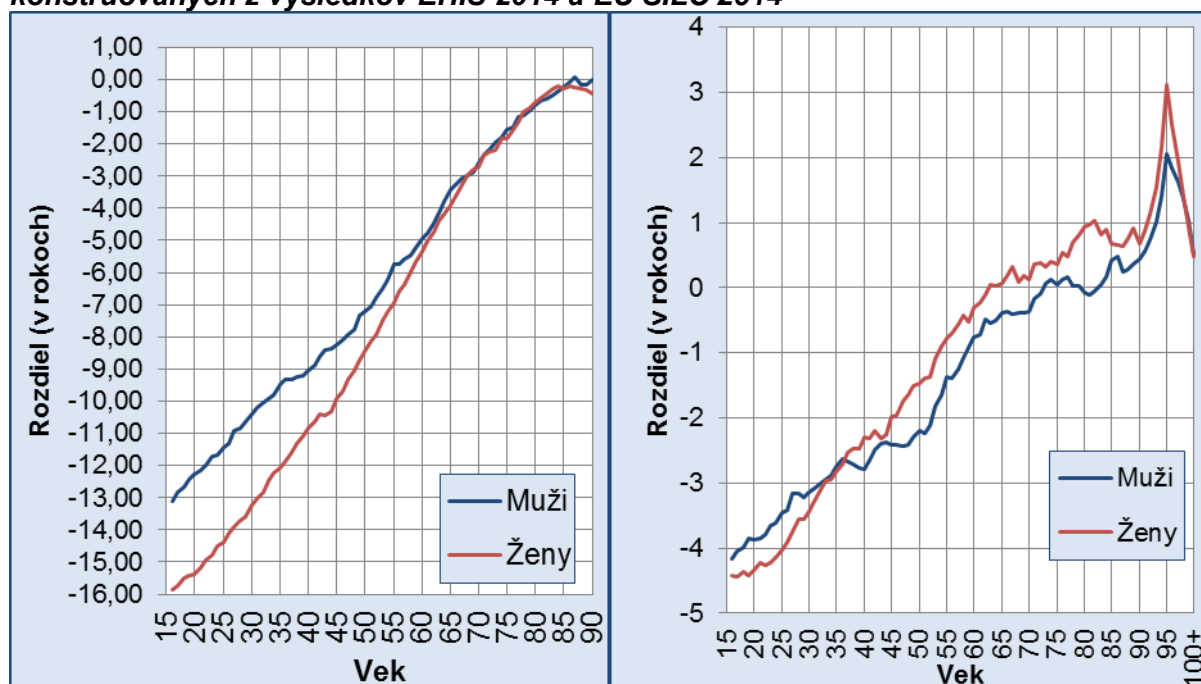
Zist'ovanie	Pohlavie	S obmedzením	Mierne obmedzenie	Bez obmedzenia
EHIS 2014	muži	10,2	24,9	64,9
	ženy	12,4	30,5	57,1
	spolu	11,4	27,8	60,9
EU SILC 2014	muži	8,2	19,9	71,8
	ženy	11,3	24,2	64,5
	spolu	9,8	22,2	68,0

Zdroj údajov: ŠÚ SR (primárne údaje EHIS 2014, EU SILC 2014), triedenie a výpočty autora

Do výpočtu stredných dĺžok života bez chronických ochorení, ako aj bez obmedzení (zdravé roky života) vstupujú úmrtnostné pomery prezentované prierezovou úmrtnostnou tabuľkou a jej vybranými funkciami a štruktúra odpovedí na dané otázky. Keďže pri výpočte boli použité rovnaké úmrtnostné tabuľky, jediným možným zdrojom odchýlok v stredných dĺžkach sú rozdiely medzi odpoveďami respondentov v zisťovaní EU SILC a EHIS 2014. Ako je zrejmé z nasledujúceho grafu 11, rozdiel medzi strednou dĺžkou života v presnom veku 16 rokov vypočítaný z údajov získaných zo zisťovaní EU SILC 2014 a EHIS 2014 bez chronických ochorení je u mužov viac ako 13 rokov a u žien dokonca takmer 16 rokov.

Keďže platí, že vo všetkých vekových skupinách respondenti v zisťovaní EHIS deklarujú prítomnosť dlhotrvajúceho ochorenia alebo zdravotného problému častejšie, najväčšie rozdiely medzi strednými dĺžkami života podľa prítomnosti chronického ochorenia vzhľadom na metodiku výpočtu identifikujeme v najmladšom veku. So zvyšujúcim sa vekom sa tieto diferencie znižujú, no aj napriek tomu zostávajú významné (graf 11). Aj preto platí, že u mužov a predovšetkým žien pri použití údajov EHIS 2014 už od najmladších vekových skupín majú prevahu roky prežité s nejakým chronickým ochorením. Ak použijeme na výpočet štruktúru odpovedí zo zisťovania EU SILC 2014, prevaha rokov prežitých s chronickým ochorením u mužov vzniká až vo veku 49 rokov a u žien vo veku 38 rokov.

Grafy č. 11 a 12: Rozdiely hodnôt strednej dĺžky života mužov a žien bez chronického ochorenia (vľavo) a zdravých rokov života (vpravo) podľa veku a pohlavia konštruovaných z výsledkov EHIS 2014 a EU SILC 2014



Poznámka: Ide o rozdiel vypočítaných hodnôt stredných dĺžok života z údajov EHIS 2014 a EU SILC 2014.

Zdroj údajov: ŠÚ SR (primárne údaje EHIS 2014, EU SILC 2014), triedenie a výpočty autora

Odlíšnosti v štruktúre odpovedí respondentov na otázku o miere obmedzenia bežných denných aktivít v zisťovaniach EHIS 2014 a EU SILC 2014 sa podobne ako v predchádzajúcom prípade premietajú do hodnôt strednej dĺžky života. Kým pri

použití odpovědí v zisťovaní EU SILC stredná dĺžka života bez obmedzenia (HLY) v presnom veku 16 rokov dosahuje u oboch pohlaví takmer 40 rokov, aplikáciou výsledkov zisťovania EHIS 2014 dostávame len hodnotu niečo viac ako 35 rokov. Rovnako logickým je aj vývoj rozdielov zdravých rokov života s vekom. Keďže vo väčšine vekových skupín (s výnimkou najstarších) platí, že podiel osôb bez obmedzenia v EHIS bol nižší, najväčšie rozdiely vznikali v najmladšom veku. Postupne sa však u oboch pohlaví zmenšovali a vďaka štruktúre odpovedí v EU SILC v seniorskom veku dokonca nastáva inverzia. Preto v najstarších vekoch mierne vyšší počet rokov bez obmedzenia bežných aktivít dosahujú muži i ženy pri použití výsledkov EHIS (pozri graf 12).

6. NÁČRT MOŽNÝCH PRÍČIN ROZDIELNYCH VÝSLEDKOV HODNOTENIA ZDRAVIA V EHIS A EU SILC 2014

Veľké rozdiely v štruktúre odpovedí na otázky modulu MEHM medzi zisťovaniami EU SILC a EHIS z roku 2014 predstavujú veľmi zaujímavý fenomén, ktorý vytvára ďalšie značné riziko samotného využitia tohto spôsobu zisťovania informácií o zdravotnom stave obyvateľstva a na neho nadväzujúcich indikátorov. Je potrebné si uvedomiť, že tieto ukazovatele patria dokonca k najdôležitejším pri hodnotení krajiny nielen podľa zdravia a zdravotného stavu jej populácie, ale aj hľadiska kvality zdravotníckej starostlivosti a jej dostupnosti. Možno ich teda považovať aj za pomocný nástroj na formovanie konkrétnych opatrení pri tvorbe zdravotníckej politiky. Ako sa však postaviť ku skutočnosti, že z dvoch zisťovaní, ktoré sa pýtajú na to isté rovnakým spôsobom, dostávame v podstate výrazne odlišné výsledky? Základnou otázkou zostáva, ako a prečo vôbec sú také výrazné rozdiely v štruktúre odpovedí, čo sa následne odráža aj na značných diferenciáciách v hodnotách sledovaných indikátorov zdravia. Pochopenie ich vzniku nám umožňuje nielen vysvetliť podstatu problému, ale prispieva tiež ku korektnejšiemu vnímaniu sledovaného konceptu zdravia a poukazuje na jeho obmedzené využitie a potrebu citlivého prístupu k získaným informáciám.

V prípravnej fáze k prvej vlne EHIS (roky 2008 a 2009) došlo postupne k medzinárodnému zosúladieniu znenia všetkých troch otázok v module MEHM. Preto približne od roku 2009 sú plne harmonizované nielen v rámci členských krajín, ale aj medzi zisťovaniami EU SILC a EHIS. Preto faktor formulácie otázky by nemal zohrávať žiadnu podstatnejšiu úlohu pri vysvetľovaní sledovaných rozdielov. Ďalším z možných determinantov, ktorý sa spája s odlišnou štruktúrou odpovedí, je obsah pokynov na ich vyplnenie. Vzhľadom na to, že nám nie je známe, či sa reálne poskytujú respondentom pred vyplňaním dotazníka alebo až počas vyplňania a ako sú v tejto oblasti inštruovaní samotní opytovatelia, je ťažké odhadnúť, ako výrazne môže tento faktor ovplyvniť obsah odpovedí na položené otázky. Z nášho pohľadu sa však nezdá, že by mohli mať (okrem zisťovania prítomnosti chronických ochorení) zásadnejší význam. Uvádzame len tie najdôležitejšie. V pokynoch k prvej otázke modulu MEHM je v prípade EHIS-u striktné uvedené, že „*pýtame sa na zdravie vo všeobecnosti a nie na súčasný zdravotný stav*“, kým v pokynoch k zisťovaniu EU SILC sa uvádza, že „*pýtame sa na súčasný zdravotný stav všeobecne*“. Ďalej EHIS definuje koncept dočasných zdravotných problémov a uvádza aj ich konkrétne príklady, ktoré nemajú byť zohľadňované pri hodnotení vlastného zdravia respondentom. Podobne aj v pokynoch pre otázku sledujúcu výskyt chronických ochorení môžeme nájsť niektoré drobné odchýlky. Predovšetkým v EHIS-e je po definícii chronického stavu explicitne uvedených niekoľko konkrétnych príkladov

(napr. kardiovaskulárne ochorenia, rakovina, chronické respiračné ochorenia, cukrovka, duševné choroby a tiež alergie). Na druhej strane pokyny EU SILC sú doplnené o upozornenie týkajúce sa respondentov s problémami, ktoré nepovažujú za veľmi závažné (napr. opakovaná senná nádcha), aby kód 1 (trpí chronickým ochorením) uvádzali len v tom prípade, ak sú v dôsledku zdravotného problému obmedzovaní vo svojich bežných aktivitách. Pokyny k tretiemu konceptu zdravia sú vo všeobecnosti obširnejšie a presnejšie definované v EHIS-e. Určitou výnimkou je prípad, keď v EU SILC sú problémy bližšie rozdelené na fyzické alebo mentálne a navyše ešte doplnené o kategóriu choroba a invalidita.

Dôležitým vysvetľujúcim faktorom by mohla byť tiež odlišná koncepcia použitých dotazníkov, ich samotný obsah, štruktúra a rozloženie otázok, ako aj nadväznosť na ďalšie obsahové moduly zisťovaní. Navyše je potrebné si uvedomiť, že kým EHIS je zisťovanie priamo zamerané na zdravie, zdravotný stav, EU SILC sa primárne zaoberá pracovnou aktivitou, trhom práce, zamestnaním, príjmami a pod., otázky MEHM sú len doplnkovým prvkom, na ktorý nenadväzuje a ani mu nepredchádza žiadny ďalší modul o zdraví respondenta. Okrem odlišného psychologického efektu a nastavenia respondenta v EHIS-e na dané 3 otázky priamo nadväzujú bloky otázok zisťujúcich prítomnosť konkrétnych chronických ochorení a určitých obmedzení denných aktivít. Môžeme sa preto domnievať, že odpovede respondenta by mohli byť ovplyvnené touto skutočnosťou. S podobným vysvetlením výrazných rozdielov medzi EHIS 2009 a EU SILC 2009 prichádza v Českej republike aj Daňková [3]. Empiricky potvrdiť alebo vylúčiť tento predpoklad by však umožnilo až testovanie dotazníkov na tých istých respondentoch v ten istý deň, pričom najprv by odpovedali na otázky zisťovania EU SILC a v druhom kole na otázky položené v zisťovaní EHIS.

Prepojenie odpovedí na prítomnosť chronických ochorení a mieru obmedzenia bežných aktivít je do určitej miery možné testovať aj nepriamo. Ak predpokladáme platnosť tohto vzťahu, potom osoba uvádzajúca v dotazníku, že trpí na nejaké chronické ochorenie alebo je nejakým obmedzovaná vo svojich bežných aktivitách, by túto skutočnosť mala vo zvýšenej miere reflektovať aj v odpovediach na otázky sledujúce konkrétne ochorenia, zdravotné problémy alebo ťažkosti s vykonávaním osobných činností, resp. činností pri starostlivosti o svoju domácnosť. Uvedený predpoklad sa potvrdil predovšetkým v prípade chronických ochorení. Ak muž/žena deklarovali, že majú chronické ochorenie, tak v 93 % prípadov aj v nasledujúcom bloku nejaké konkrétne uviedli (pozri tab. 4).

Tabuľka č. 4: Matica závislostí medzi odpoveďami na otázku o prítomnosti chronického ochorenia a deklaráciou konkrétnych chronických ochorení v zisťovaní EHIS 2014 na Slovensku

Prítomnosť chronického ochorenia	Uviedli aspoň jedno chronické ochorenie	Neuviedli ani jedno chronické ochorenie
<i>Muži (v %)</i>		
Má chronické ochorenie	92,9	7,1
Nemá chronické ochorenie	x	100,0
<i>Ženy (v %)</i>		
Má chronické ochorenie	93,3	6,7
Nemá chronické ochorenie	x	100,0

Poznámka: Údaje sa čítajú po riadkoch.

Zdroj údajov: ŠÚ SR (primárne údaje EHIS 2014), triedenie a výpočty autora

Tabuľka č. 4 tiež poukazuje na veľmi úzky vzťah medzi zápornými odpoveďami: ak osoba uviedla, že nemá dlhodobé ochorenie, v 100 % prípadov žiadne neuviedla. Menej zrejma je súvislosť medzi mierou obmedzenia bežných denných aktivít a uvedením aspoň jedného konkrétneho obmedzenia, resp. ťažkosti s vykonávaním činností osobnej starostlivosti alebo starostlivosti o domácnosť. Len približne tretina mužov a necelých 42 % žien, ak deklarovali určitú mieru obmedzenia, v ďalších otázkach v zisťovaní EHIS 2014 uviedli aspoň jedno konkrétne obmedzenie.

Tabuľka č. 5: Matica závislostí medzi deklaráciou konkrétnych obmedzení a odpoveďami na otázku o miere obmedzenia bežných činností v zisťovaní EHIS 2014 na Slovensku

Miera obmedzenia	Počet konkrétnych obmedzení	
	uviedli aspoň jedno obmedzenie	neuviedli žiadne obmedzenie
<i>Muži (v %)</i>		
Obmedzovaní (veľmi alebo nie veľmi)	32,7	67,3
Neobmedzovaní	7,8	92,2
<i>Ženy (v %)</i>		
Obmedzované (veľmi alebo nie veľmi)	41,9	58,1
Neobmedzované	9,0	91,0

Poznámka: Údaje sa čítajú po riadkoch.

Zdroj údajov: ŠÚ SR (primárne údaje EHIS 2014), triedenie a výpočty autora

7. ZÁVER

Slovensko z pohľadu indikátorov zdravotného stavu produkovaných z výsledkov zisťovaní EU SILC v rámci modulu MEHM patrí medzi najhoršie krajiny v EÚ28 a v prípade ukazovateľa zdravé roky života sa dokonca zaradilo na poslednú priečku rebríčka krajín. Pri detailnejšom pohľade však zisťujeme, že nepriaznivé postavenie majú aj viaceré demograficky najvyspelejšie krajiny Európy (s najvyššou strednou dĺžkou života) a, naopak, niektoré krajiny s jednoznačne horšími úmrtnostnými pomermi, ako sú na Slovensku, sa nachádzajú nad priemerom EÚ28. Paradoxy daných ukazovateľov vo vzťahu k strednej dĺžke života sa ešte viac zvyraznia, ak sa pozrieme na štruktúru rokov podľa kvality zdravia. Je zrejme, že vzťah vyššia stredná dĺžka života a lepšie zdravie merané indikátormi konštruovanými z výsledkov zisťovania EU SILC platí len čiastočne, resp. v niektorých krajinách.

Viacrozmerná dekompozícia rozdielov zdravých rokov života označila ako primárny faktor výrazného zaostávania horšie vnímanie kvality vlastného zdravia u respondentov na Slovensku. V najväčšej miere sa na tomto stave pritom podieľajú osoby v poreprodukčnom a poproduktívnom veku.

Porovnanie EU SILC 2014 a EHIS 2014 opätovne potvrdzuje vznik paradoxu pomerne veľkých rozdielov v strednej dĺžke života bez prítomnosti chronických ochorení a čiastočne aj zdravých rokov života počítaných zvlášť z výsledkov oboch zisťovaní. Problém je o to významnejší, ak si uvedomíme, že v oboch zisťovaniach sa používajú rovnaké harmonizované otázky zisťujúce dané koncepty zdravotného stavu. Ako ukázala naša analýza, primárne je to dôsledok výrazne odlišných štruktúr odpovedí respondentov na položené otázky.

V zisťovaní EHIS majú tendenciu muži i ženy deklarovať nižšiu kvalitu svojho zdravia, častejší výskyt chronických ochorení, ako aj obmedzení bežných denných činností. Vzhľadom na získané poznatky sa môžeme domnievať, že formovanie rozdielov v štruktúre odpovedí na otázky modulu MEHM v zisťovaniach EHIS a EU SILC 2014 je výsledkom pôsobenia niekoľkých vzájomne sa podmieňujúcich faktorov. Mieru ich vplyvu síce nie je možné empiricky jednoznačne stanoviť, ale z nášho pohľadu sa domnievame, že kľúčovým je predovšetkým prítomnosť ďalších doplňujúcich otázok v zisťovaní EHIS, ktoré majú konkrétne za cieľ určiť prítomnosť niektorého z vymenovaných chronických ochorení alebo obmedzení vo vykonávaní bežných denných aktivít. Tým predstavujú pre respondenta vlastne akúsi pomôcku, ktorú však v zisťovaní EU SILC nemá k dispozícii. Je zrejmé, že použitie viacerých druhov zisťovaní, resp. aplikácia modulu MEHM v rôznych zisťovaniach môže (a nielen v prípade Slovenska) viesť a ukazuje sa, že aj vedie, k získaniu často až diametrálne odlišných výsledkov. To výrazne obmedzuje výpovednú hodnotu už takto problematického konceptu merania zdravia a zdravotného stavu a znižuje jeho transparentnosť a dôveryhodnosť nielen v odborných kruhoch.

Hodnotenie zdravotného stavu predstavuje značne komplexnú a v mnohých smeroch aj komplikovanú problematiku, ktorú vyčerpávajúco nemôže podchytiť taký jednoduchý indikátor, akým sú zdravé roky života (resp. ostatné indikátory zdravia konštruované v rámci modelu MEHM). Preto je veľmi dôležité pracovať so získanými výsledkami a interpretovať ich opatrne, nepreceňovať ich význam a upozorňovať širšiu vedeckú i laickú verejnosť na všetky aspekty práce s nimi. Domnievame sa, že v takejto podobe nepredstavujú vhodný ukazovateľ na meranie kvality zdravotníctva, zdravotníckych služieb a len v obmedzenej miere ich možno aplikovať pri sledovaní samotného zdravotného stavu.

V súvislosti so získanými výsledkami si preto dovoľujeme odporučiť používať predmetné indikátory len ako jeden z možných prístupov, ktorý však hodnotí určitú špecifickú dimenziu zdravotného stavu. Významným kvalitatívnym posunom by bolo doplnenie o zisťovanie skutočného zdravotného stavu respondenta prostredníctvom lekárskeho vyšetrenia, odbornej analýzy zdravotných záznamov a pod. Prínosom by tiež mohla byť konštrukcia sofistikovanejšieho indikátora zdravotného stavu obyvateľstva, ktorý by sa opieral nielen o subjektívne hodnotenie zdravia, ale aj o niektoré empirické zdravotnícke štatistiky. V neposlednom rade by bolo vhodné pri komunikácii uvedených zistení najmä pre širšiu laickú verejnosť uvádzať aj ďalšie informácie spojené s týmto konceptom. Mohlo by ísť napríklad o hodnotu reálnej strednej dĺžky života v kombinácii so štruktúrou rokov zostávajúcich osobe prežiť v určitej kvalite zdravia.

LITERATÚRA

- [1] ANDREEV, E. M. – SHKOLNIKOV, V. M. – BEGUN, A. Z.: Algorithm for decomposition of differences between aggregate demographic measures and its application to life expectancies, healthy life expectancies, parity-progression ratios and total fertility rates. In: Demographic Research, 2002, 7, p. 499-522.
- [2] DAŇKOVÁ, Š.: Výběrová šetření o zdravotním stavu v ČR a v Evropě. In: Demografický informační portál. 2006 [cit. 2016-12-05]. Dostupné na: http://www.demografie.info/?cz_detail_clanku&artclID=357.

- [3] DAŇKOVÁ, Š.: Srovnání ukazatelů zdravotního stavu v šetření EHIS a SILC. In: Demografický informační portál. 2010 [cit. 2016-12-06]. Dostupné na: http://www.demografie.info/?cz_detail_clanku&artclID=703.
- [4] FRIES, J. F.: The compression of morbidity. In: *Milbank Memorial Fund Quarterly*, 1983, 61, p. 397-419.
- [5] FRIES, J. F.: The compression of morbidity: near or far? In: *Milbank Memorial Fund Quarterly*, 1989, 67, p. 208-232.
- [6] HRKAL, J. – DAŇKOVÁ, Š.: Zdravá délka života u obyvatel EU. In: Demografický informační portál [cit. 2016-05-12]. Dostupné na: http://www.demografie.info/?cz_detail_clanku&artclID=107.
- [7] JAGGER, C. – VAN OYEN, H. – ROBINE, J. M.: Health Expectancy Calculation by the Sullivan Method: A Practical Guide. Newcastle University. Institute for Ageing, 2014 [cit. 2016-05-12]. Dostupné na: https://reves.site.ined.fr/fichier/s_rubrique/20184/rp408.en.pdf.
- [8] KRAMER, M.: The rising pandemic of mental disorders and associated chronic diseases and disabilities. In: *Acta Psychiatrica Scandinavica*, 1980, 62 (Suppl. 285), p. 282-297.
- [9] MÉSZÁROS, J.: Ako dlho žije populácia Slovenskej republiky v zdraví? In: *Slovenská štatistika a demografia*, 2007, č. 1 – 2, s. 133 – 140.
- [10] MÉSZÁROS, J.: Stredná dĺžka života v zdraví podľa EHIS 2009. Bratislava: INFOSSTAT, 2010.
- [11] OLSHANSKY, S. J. – RUDBERG, M. A. – CARNES, B. A. – CASSEL, C. K. – BRODY, J.: Trading off longer for worsening health: the expansion of morbidity hypothesis. In: *Journal of Aging and Health*, 1991, 2, p. 194-216.
- [12] RYCHTAŘÍKOVÁ, J.: Zdravá délka života v současné české populaci. In: *Demografie*, 2006, č. 3, s. 166 – 178.
- [13] SULLIVAN, D. F.: A single index of mortality and morbidity. In: *Health Services Mental Health Administration Health Reports*, 1971, 86, p. 347-354.
- [14] VERBRUGGE, L. M.: Longer life but worsening health? In: *Trends in health and mortality in middle-age and older persons. Memorial Fund Quarterly*, 1984, 62, p. 475-519.

RESUME

The position of Slovakia in terms of health indicators constructed from MEHM module in the EU SILC survey has been negative in the European area. This is particularly true for the healthy life years (HLY), which belong among the lowest value within the EU28. Paradoxically, there is an unfavourable situation also in some countries with the highest life expectancy at birth. On the contrary, some countries (e.g. Romania, Bulgaria) with bad mortality conditions are ranked among the above average. It is obvious that the relationship between mortality and the health status evaluation through the MEHM module is not always clear. The main cause of the shortfall in the HLY value was caused by frequent declaration of the restriction of daily activities by respondents in Slovakia. Our results also confirmed the existence of considerable differences in the answers to the same questions about the health (MEHM module) in the sample survey EU SILC 2014 and EHIS 2014. The formation of these differences is due to several mutually interdependent factors. The key factor is particularly the presence of other supplementary questions in EHIS survey, tracking the presence of any chronic diseases or difficulties in personal or household care.

PROFESIJNÝ ŽIVOTOPIS

RNDr. Branislav Šprocha, PhD., absolvoval magisterské štúdium na Prírodovedeckej fakulte Univerzity Karlovej v Prahe v odbore demografia a demogeografia (2006). V roku 2011 ukončil doktorandské štúdium v programe demografia. Od roku 2007 je vedeckovýskumným pracovníkom Výskumného demografického centra Inštitútu informatiky a štatistiky v Bratislave a od roku 2009 vedeckým pracovníkom Prognostického ústavu Slovenskej akadémie vied. V oblasti demografie sa špecializuje na problematiku sobášnosti a rozvodovosti, populačného vývoja a jeho vplyvu na spoločnosť. Okrem toho sa venuje analýze vybraných populačných štruktúr, reprodukčného správania rómskeho obyvateľstva na Slovensku a problematike populačného prognózovania.

KONTAKT

branislav.sprocha@gmail.com

Bianka PARMOVÁ, Mária VOJTKOVÁ
Katedra štatistiky Fakulty hospodárskej informatiky Ekonomickej univerzity
v Bratislave

SEGMENTÁCIA ČITATEĽOV S VYUŽITÍM TEXT MININGU

SEGMENTATION OF READERS USING TEXT MINING

ABSTRAKT

Analýza správania čitateľov na internetovej stránke poskytuje tvorcom webu i tvorcom webového obsahu cenné informácie, ktorých využitie môže zvýšiť zisk prevádzkovateľovi stránky. Odhalenie najpopulárnejších tém preferovaných čitateľmi umožňuje získať lepší prehľad o tom, o ktorý typ obsahu je najmenší záujem a, naopak, ktorá téma je pre čitateľov najzaujímavejšia. Cieľom tohto príspevku je segmentácia online čitateľov na základe prečítaných článkov podľa tém s využitím hĺbkovej analýzy textu. Metódy hĺbkovej analýzy textu využívame na extrahovanie tém článkov spravodajského webu. Prostredníctvom výsledkov tejto analýzy v spojení s údajmi o návštevnosti stránky vytvoríme profily čitateľov na základe tém, ktoré preferujú.

ABSTRACT

A behavioral analysis of readers at the website, provides valuable information for the creators of the site and content, the utilization of which may help to increase revenue of the website operator. Revealing the most popular topics preferred by readers enables the web creators to gain deeper insights into which topics are the least interesting for users and vice versa, which topics attract readers the most. The aim of this paper is to create segments of online readers using text mining techniques, based on the articles read. We are using text mining techniques to extract topics from the articles of the news website. We can create user profiles based on the topic preferences of readers through the analysis results in connection with web traffic data.

KLÚČOVÉ SLOVÁ

hĺbková analýza textu, extrahovanie tém, model latentnej Dirichletovej alokácie, segmentácia čitateľov

KEY WORDS

text mining, topic modeling, model of Latent Dirichlet allocation, segmentation of readers

1. ÚVOD

V súčasnom trende rozvoja informačných technológií a internetu sa podnikanie v mnohých oblastiach presúva do online prostredia. Výnimkou nie sú ani spravodajské médiá, ktoré ponúkajú svoj obsah na webových stránkach. Výhodou online prostredia je možnosť získania obrovského množstva údajov o zákazníkoch, ktorými sú v prípade online médií samotní návštevníci (čitatelia) stránky. Analýzou dát o návštevnosti stránky v spojení s analýzou textových údajov priamo zo stránky médií môžeme dospieť k cenným informáciám o návštevníkoch stránky, ktoré na základe tradičných nástrojov webovej analytiky nemožno získať.

Hlavným cieľom tohto článku je vytvorenie segmentov čitateľov spravodajského webu s rovnakými preferenciami tém článkov. Tento cieľ pozostáva z dvoch základných parciálnych cieľov. Prvým z nich je charakteristika jednotlivých článkov prostredníctvom tém využitím metód hĺbkovej analýzy textu. Po naplnení prvého základného parciálneho cieľa získame charakteristiku analyzovaných článkov prostredníctvom tém, ktorých sa týkajú. Tieto údaje budú vstupom do druhej časti analýzy a podmienkou naplnenia druhého parciálneho cieľa, ktorým je vytvorenie segmentov čitateľov s rovnakými preferenciami pomocou zhlukovej analýzy.

Výsledok hlavného cieľa práce spočíva v poskytnutí nového, netradičného pohľadu na štruktúru čitateľov spravodajského webu vzhľadom na preferované témy. Kombináciou techniky hĺbkovej analýzy textu a zhlukovej analýzy poskytuje tento článok nový pohľad na segmentáciu v oblasti webovej analytiky, ktorú možno využiť v mnohých ďalších oblastiach okrem aplikovanej oblasti spravodajských médií.

Samotná analýza je založená na dvoch zdrojoch dát. Prvým zdrojom je html kód analyzovanej stránky www.sme.sk, z ktorého sme pomocou kódu v jazyku *R* extrahovali URL adresy článkov a ich kľúčové slová, ktoré vytvárajú autori článkov pri publikovaní. Druhým zdrojom dát sú logy servera analyzovanej stránky, kde sa zaznamenávajú všetky interakcie návštevníkov so stránkou. Tieto dáta sme mali k dispozícii od firmy Piano Media, s. r. o., ktorá prevádzkuje na stránke www.sme.sk služby spoplatnenia online obsahu¹. Prostredníctvom dát o návštevníkoch tohto spravodajského webu budeme analyzovať ich správanie z pohľadu obsahu prečítaných článkov za analyzovaný mesiac, ktorým bol júl 2015.

2. HĽBKOVÁ ANALÝZA TEXTU A JEJ VÝZNAM V SÚČASNOSTI

Definícia hĺbkovej analýzy textu nie je jednoznačná, pretože sa opiera o poznatky z mnohých ďalších oblastí, ako napríklad vyhľadávanie informácií, hĺbková analýza údajov alebo objavovanie znalostí v dokumentoch. V tejto časti článku by sme chceli objasniť vývoj tohto pojmu.

S pojmom procesu hĺbkovej analýzy textu (Text Mining), často nazývaného aj procesom objavovania znalostí v textových dokumentoch (Knowledge Discovery in Texts – KDT), sa v literatúre stretávame prvýkrát v roku 1999 [4], i keď s výskumom tejto problematiky sa možno stretnúť aj skôr. Autorka tu definuje hĺbkovú analýzu textu pomocou už v tom čase známych disciplín, ako napr. vyhľadávanie informácií (Information Retrieval). Kľúčovou vlastnosťou je hľadanie rôznych typov vzorov v textových súboroch, čo považovala za analogické klasickému procesu hĺbkovej analýzy údajov [7].

Publikácia [6] z roku 2003 rozdeľuje hĺbkovú analýzu textu do dvoch kategórií – klasické a inteligentné dolovanie. „Pod klasickým dolovaním v textoch sa chápe napr. kategorizácia textov, zhlukovanie textov, extrakcia lexikálnych a syntaktických znakov, hľadanie asociácií medzi termami, extrakcia liniek medzi entitami v rámci textu a medzi rôznymi textami, ale napr. aj sumarizácia textov. Inteligentným dolovaním v textoch sa potom nazýva interakcia výskumníka a počítačového nástroja, ako aj použitie metód umelej inteligencie s cieľom vytvárať znalosti o okolitom svete na základe odvodených lingvistických znakov a ďalších typov

¹ Menovaná firma ako vlastník dát súhlasí s publikovaním výsledkov v predloženej podobe.

vzorov. Napr. odráža novoidentifikovaná téma v prúde dokumentov realitu? Aké stratégie skúmania implikujú nájdené prepojenia, sú naozaj relevantné? Aké obchodné rozhodnutia možno na ich základe učiniť?“ [7, s. 10].

Publikácia [3] z roku 2007, ktorá sa venuje hĺbkovej analýze textu veľmi podrobne, ju definuje ako „znalostne intenzívny proces, v ktorom používateľ priebežne interaguje s kolekciami dokumentov za pomoci analytických nástrojov“. Autori kladú dôraz na analýzu prepojení medzi informáciami v kolekcii dokumentov, pričom kategorizáciu, zhľukovanie a extrakciu informácií chápu ako súčasť predspracovania textových údajov [7, s. 11].

Náš názor na definíciu hĺbkovej analýzy textu sa vo veľkej miere zhoduje s názorom kolektívu autorov z Technickej univerzity v Košiciach, ktorý vydal publikáciu z oblasti hĺbkovej analýzy textu v slovenskom jazyku *Dolovanie znalostí z textov* [7]. Autori sa prikláňajú k definícii hĺbkovej analýzy textu podľa základnej definície hĺbkovej analýzy údajov, ktorú prezentujú ako „interaktívny a iteratívny proces získavania platných, pre danú aplikáciu užitočných a doposiaľ neznámych znalostí“. Za dôležitú súčasť tohto procesu považujú taktiež interakciu používateľa s analytickým systémom. S týmto názorom sa stotožňujeme, pretože považujeme zohľadňovanie subjektívnych názorov a vedomostí analytika z analyzovanej oblasti za veľmi dôležitú súčasť analýzy textových údajov.

Po definícii hĺbkovej analýzy textu sa zameriame bližšie na jej základné techniky a ich aplikácie v praxi. Medzi základné techniky hĺbkovej analýzy textu môžeme zaradiť výskumnú analýzu (Exploratory Analysis) a kategorizáciu (Categorization) [5, s. 2].

Výskumná analýza zahŕňa techniky, ako extrahovanie tém (Topic Extraction) či zhľukovú analýzu (Clustering). Hlavnou myšlienkou extrahovania tém je vytvorenie skupín dokumentov, ktoré sa týkajú rovnakých tém. Cieľom zhľukovej analýzy v oblasti hĺbkovej analýzy textu je priradiť jednotlivé dokumenty zo súboru k jednému zhľuku, pričom dokumenty v zhľukoch sú si navzájom podobné a jednotlivé zhľuky dokumentov sú navzájom odlišné. Rozdiel medzi zhľukovaním a extrahovaním tém zo súboru dokumentov spočíva v tom, že v prípade extrahovania tém sa môže jeden dokument zaradiť súčasne do viacerých skupín v závislosti od toho, koľko tém je v ňom obsiahnutých [5, s. 111].

Kategorizácia súvisí s oblasťou manažmentu obsahu (Content Management) zameranou na organizáciu veľkého množstva dokumentov získaných z rôznych zdrojov na základe ich obsahu. Hlavným rozdielom medzi kategorizáciou obsahu a extrahovaním tém je, že v prípade kategorizácie ide o učenie s učiteľom, teda je potrebné určiť kategórie obsahu, kam majú byť jednotlivé dokumenty zaradené. V prípade extrahovania tém sa tieto kategórie vytvárajú automaticky na základe štatistických metód (napr. latentná sémantická analýza) [5, s. 159].

Technika kategorizácie textu sa využíva v mnohých oblastiach podnikania. Niektoré online médiá využívajú túto techniku na automatické zaraďovanie nových článkov do sekcií na stránke (napr. šport, politika, veda, ekonomika atď.) alebo tiež na vytváranie personalizovaných odporúčaní na obsah súvisiaci s témami v okruhu záujmov čitateľa.

3. EXTRAHOVANIE TÉM SPRAVODAJSKÝCH ČLÁNKOV

Zdrojom dát našej analýzy bola spomínaná internetová stránka spravodajského portálu www.sme.sk, z ktorej sme prostredníctvom html kódu získali kľúčové slová článkov dostupných na stránke v júli 2015. Vstupné dáta obsahujúce tieto údaje pozostávali z 8 203 unikátnych URL adries jednotlivých článkov. Každý z článkov mal vo svojej URL adrese zahrnutý unikátny identifikátor v tvare /c/“sedem číslic“/, pričom niektoré identifikátory článkov sa v dátovom súbore vyskytovali viac než jedenkrát, z čoho vyplýva, že niektoré články boli zahrnuté viackrát s rôznym tvarom URL adresy, čo je pre našu analýzu nežiaduce. Preto sme najprv odstránili duplicitné pozorovania (články).

Po tejto modifikácii tvorili naše dáta 4 043 unikátnych článkov, ktorých obsah reprezentovali kľúčové slová priradené ku každému článku. V ďalšej časti sme sa zamerali na analýzu kľúčových slov. Vstupom do analýzy textu boli teda samotné kľúčové slová, pričom naša databáza obsahovala 7 408 unikátnych kľúčových slov.

Pred samotnou analýzou bolo však potrebné textové údaje predspracovať. Vo všeobecnosti sa v procese predspracovania textových údajov dodržiavajú nasledujúce kroky:

1. Konverzia na čistý text – znamená úpravu dát vzhľadom na to, že kľúčové slová z formálneho hľadiska nie sú jednotne upravené, niektoré slová sa začínajú veľkými písmenami, iné malými napriek tomu, že ich význam je totožný; jednotlivé slová sú oddelené čiarkami, ale s rôznym počtom medzier za nimi atď. Konverzia na čistý text bola aj v našom prípade úvodným krokom analýzy kľúčových slov.
2. Tokenizácia a segmentácia – rozdelenie textu na elementárne textové jednotky. V našom prípade išlo o rozdelenie zoznamu kľúčových slov pri jednotlivých článkoch pomocou jednotného separátora – medzery.
3. Lematizácia, morfológická analýza – jednotlivé slová sa v textových dokumentoch môžu vyskytovať v rôznych morfológických tvaroch, teda v rôznych pádoch, osobách, číslach atď. Preto je nevyhnutné transformovať ich do základného tvaru, tzv. lemu. V našom prípade sme na vstupné údaje aplikovali všetky úpravy okrem lematizácie, pretože sme pracovali už priamo s kľúčovými slovami článkov, ktoré sa uvádzali v prevažne zjednotenom morfológickom tvare.
4. Eliminácia neplnovýznamových slov – znamená vylúčenie takých slov, pri ktorých sa predpokladá malý prínos ku charakteristike obsahu dokumentov. V našom prípade sme spomedzi kľúčových slov vylučovali číslice, znaky a chybné výrazy pozostávajúce z jedného písmena.
5. Váhovanie termov – označuje sa ako úprava frekvencie slov každého dokumentu v celom korpuse. V našom prípade hral tento krok predspracovania údajov významnú úlohu – až 63 % všetkých kľúčových slov (4 637 z celkového počtu 7 408) bolo priradených k článku iba jedenkrát, 15 % len v dvoch prípadoch. Hranicu na elimináciu nízko frekventovaných slov sme stanovili na úroveň 5 priradení v rámci všetkých článkov.

Všetky doteraz opísané kroky spracovania textových dát boli nevyhnutné na transformáciu textových dokumentov do formy vektorovej reprezentácie.

Po úpravách sme mali k dispozícii vstupné dáta pozostávajúce zo 4 043 článkov opísaných 922 kľúčovými slovami. Aplikácia techniky hĺbkovej analýzy textu (modelu latentnej Dirichletovej alokácie – LDA)) nám umožnila nahradiť kľúčové slová charakterizujúce obsah článkov témami generovanými spomínaným modelom.

Latentná Dirichletova alokácia je súčasťou oblasti pravdepodobnostného modelovania (*probabilistic modeling*). Generatívne pravdepodobnostné modelovanie je postavené na myšlienke, že analyzované dáta považujeme za výsledok generatívneho pravdepodobnostného procesu, ktorý obsahuje skryté (latentné) premenné (*hidden variables*). Tento generatívny proces je definovaný združeným rozdelením pravdepodobnosti pozorovaných a skrytých premenných. Prostredníctvom tohto združeného rozdelenia pravdepodobnosti vieme vypočítať podmienenú pravdepodobnosť skrytých premenných za predpokladu pozorovaných premenných. V prípade LDA sú pozorovanými premennými slová dokumentov a skrytými premennými rozumieme neznámu štruktúru tém v kolekcii dokumentov, ktorú chceme odhaliť. Výpočtovým problémom odvodenia skrytej štruktúry tém je problém výpočtu aposteriórneho rozdelenia pravdepodobnosti, teda podmienenej pravdepodobnosti skrytých tém za predpokladu pozorovaných premenných [1, s. 77 – 84]. Práve z dôvodu využitia štatistickej inferencie založenej na výpočte podmienenej pravdepodobnosti neznámych premenných za predpokladu známych pozorovaní (slov) sa tento model označuje za bayesiánsky [2, s. 71 – 93].

Tabuľka č. 1: Výsledok procesu extrahovania tém

Téma	Poradie	1	2	3	4	5	Názov témy
1	Kľúčové slovo	eko	zvuk	vzi	video	foto	Ekonomika
	Významnosť	0,1099	0,0674	0,0475	0,0275	0,0258	
2	Kľúčové slovo	maďarsko	migranti	čr	migrácia	počasie	Migranti
	Významnosť	0,0752	0,0624	0,0554	0,0267	0,0248	
3	Kľúčové slovo	iráň	dohoda	Usa	vzi	horúčavy	Zahraničné správy
	Významnosť	0,0590	0,0501	0,0457	0,0354	0,0354	
4	Kľúčové slovo	polícia	vzi	bax	nehoda	bbx	Policajné správy
	Významnosť	0,1024	0,0300	0,0300	0,0300	0,0248	
5	Kľúčové slovo	hokej	nhl	khl	usa	slovan	Hokej
	Významnosť	0,2299	0,0729	0,0540	0,0440	0,0364	
6	Kľúčové slovo	is	sýria	turecko	usa	útok	Blízky východ
	Významnosť	0,0665	0,0525	0,0402	0,0359	0,0350	
7	Kľúčové slovo	futbal	el	prestup	anglicko	usa	Futbal
	Významnosť	0,2553	0,0245	0,0238	0,0224	0,0175	
8	Kľúčové slovo	zákon	nrsr	novela	návrh	prezident	Politika SR
	Významnosť	0,0472	0,0460	0,0331	0,0224	0,0213	
9	Kľúčové slovo	re	ba	bax	doprava	tt	Dopravné správy
	Významnosť	0,1485	0,0748	0,0408	0,0283	0,0261	
10	Kľúčové slovo	cyklistika	tourdefrance	svet	motorizmus	tdf	Cyklistika
	Významnosť	0,1252	0,0429	0,0340	0,0322	0,0322	

Téma	Poradie	1	2	3	4	5	Názov témy
11	Kľúčové slovo	školstvo	vláda	školy	súdy	bbx	Školstvo
	Významnosť	0,0443	0,0246	0,0172	0,0148	0,0123	
12	Kľúčové slovo	tenis	wimbledon	dvojhra	výsledok	výsledky	Tenis
	Významnosť	0,1271	0,0535	0,0524	0,0468	0,0401	
13	Kľúčové slovo	rusko	ukrajina	usa	francúzsko	eko	Rusko, Ukrajina
	Významnosť	0,0867	0,0766	0,0615	0,0242	0,0232	
14	Kľúčové slovo	usa	obete	útok	británia	čina	Krimi zahranicie
	Významnosť	0,0457	0,0431	0,0326	0,0300	0,0222	
15	Kľúčové slovo	eko	grécko	eú	nemecko	eurozóna	EÚ
	Významnosť	0,1351	0,1210	0,0507	0,0244	0,0192	

Poznámka: *eko* – ekonomika, *vzi* – prevzaté články, *čr* – Česká republika, *bax* – okres Bratislava, *bbx* – okres Banská Bystrica, *nhl* – Americká hokejová liga, *khl* – Kontinentálna hokejová liga, *is* – Islamský štát, *el* – Európska futbalová liga, *nrsr* – Národná rada Slovenskej republiky, *re* – regionálne spravodajstvo, *ba* – Bratislava, *tt* – Trnava, *tdf* – Tour de France, *eú* – Európska únia.

Zdroj: *vlastné spracovanie v jazyku R*

Výstupom modelu LDA je okrem iného matica početnosti priradení jednotlivých slov k vytvoreným témam, na základe ktorej dokážeme získať obsah matice ϕ s hodnotami pravdepodobnosti, s akou jednotlivé slová súvisia s vytvorenými témami. Použitím pomocnej funkcie v jazyku R získame pre každú tému päť najvýznamnejších kľúčových slov a ich pravdepodobnosť, s akou súvisia s danou témou. Súčet pravdepodobností všetkých slov sa pre každú vytvorenú tému rovná jednej. Tento výstup upravený do tabuľky 1 nám umožňuje interpretovať a pomenovať jednotlivé témy. V prvom stĺpci tabuľky 1 je uvedené číslo témy, každá z tém je charakterizovaná kľúčovým slovom (v riadku *Kľúčové slovo*) spolu s pravdepodobnosťou súvislosti tohto slova s danou témou (riadok *Významnosť*). Na základe najvýznamnejších slov subjektívne určíme názvy tém, ktoré budeme používať v ďalších častiach článku. V poznámke pod tabuľkou uvádzame podrobnejší opis kľúčových slov, ktoré majú formu skratiek.

V prvej téme je najvýznamnejšie slovo *eko* indikujúce správy z ekonomiky. Ďalšie kľúčové slová sa týkajú prevzatých článkov (*vzi*) či videí a fotiek. Ide o viac-menej všeobecnú tému, nazveme ju *ekonomika*.

Druhá téma sa týka vyslovene problematiky migrácie, medzi najvýznamnejšie kľúčové slová patria slová ako *maďarsko*, *migranti* či *migrácia*. Objavuje sa tu aj *počasie*, no tému napriek tomu nazveme *migranti*.

Tretiu tému tvoria zahraničné správy a najvýznamnejšie kľúčové slová sú *iráň*, *dohoda* a *usa*. Aj v tomto prípade sa medzi prvými piatimi najvýznamnejšími kľúčovými slovami vyskytuje termín týkajúci sa počasia (*horúčavy*), čo je pravdepodobne spôsobené obdobím zberu dát, ktorým bol mesiac júl, keď sa o počasie zaujíma viacero čitateľov.

Medzi pätnástimi témami sa vyskytujú štyri témy týkajúce sa vyslovene športových disciplín: *hokej*, *futbal*, *cyklistika* a *tenis*. V prípade týchto tém sa medzi najvýznamnejšími kľúčovými slovami nachádzajú výlučne slová charakteristické pre dané športy. Napríklad téma *tenis* je charakterizovaná kľúčovými slovami, ako *tenis*,

wimbledon, dvojhra, výsledok, výsledky. V prípade témy *futbal* ide o slová *futbal, el* (európska liga), *prestup, anglicko, usa*.

Ďalšími zaujímavými témami s kľúčovými slovami vzťahujúcimi sa vyslovene na danú tému sú napríklad *politika SR* s kľúčovými slovami, ako *zákon, nrsr, novela, návrh, prezident*, alebo téma *EÚ* charakterizovaná najvýznamnejšími termínmi, ako *eko, grécko, eú, nemecko, eurozóna*.

Doposiaľ sme analyzovali vzniknuté témy len prostredníctvom ich najvýznamnejších kľúčových slov. Na získanie lepších interpretačných výsledkov použijeme vizualizáciu výsledkov extrahovania tém, ktorá nám umožní porovnať jednotlivé témy a tiež zistiť, ktorá z nich má najväčšiu prevahu.

Na vizualizáciu potrebujeme zostrojiť dve matice. Prvou je *theta*, matica opisujúca relevanciu jednotlivých článkov k vytvoreným témam (tabuľka č. 2). Obsahom matice sú koeficienty *theta*, ktoré sme získali vydelením počtu priradení slov dokumentov k jednotlivým témam v rámci iteračného procesu (celkovým počtom priradení slov dokumentu ku všetkým témam). Téma s najvyššou hodnotou tejto pravdepodobnosti vyjadrená koeficientom *theta* bude daný článok charakterizovať najväčšou mierou. Počet iterácií hovorí o počte opakovaní procesu priradovania jednotlivých kľúčových slov k témam a jednotlivých tém k článkom, pričom v našom prípade sme nastavili jeho úroveň na 1000 opakovaní.

Tabuľka č. 2: Ukážka matice *theta* určujúcej závislosti medzi článkami a témami

Kľúčové slová	sr nr sieť poriadok exekučný novela beblavý	usa futbal turné chelsea výhra barcelona sumár	srbsko kosovo albánsko rama návšteva neohlásená	eko rusko meny cudzie nákup pozastavený	sr elektro odborníci priemysel elektro- technický nedostatok zep stretnutie kiska	usa new york černochoch policajt uškrtienie dohoda
Ekonomika	0,0024	0,0016	0,0024	0,0047	0,4698	0,0024
Migranti	0,0024	0,0016	0,4843	0,0047	0,0047	0,0024
Zahraničné správy	0,0024	0,0016	0,0024	0,0047	0,0047	0,4843
Policajné správy	0,0024	0,0016	0,0024	0,0047	0,0047	0,4843
Hokej	0,0024	0,0016	0,0024	0,0047	0,0047	0,0024
Blízky východ	0,0024	0,0016	0,0024	0,0047	0,0047	0,0024
Futbal	0,0024	0,9772	0,0024	0,0047	0,0047	0,0024
Politika SR	0,9663	0,0016	0,0024	0,0047	0,4698	0,0024
Dopravné správy	0,0024	0,0016	0,0024	0,0047	0,0047	0,0024
Cyklistika	0,0024	0,0016	0,0024	0,0047	0,0047	0,0024
Školstvo	0,0024	0,0016	0,0024	0,0047	0,0047	0,0024
Tenis	0,0024	0,0016	0,0024	0,0047	0,0047	0,0024
Rusko, Ukrajina	0,0024	0,0016	0,4843	0,4698	0,0047	0,0024

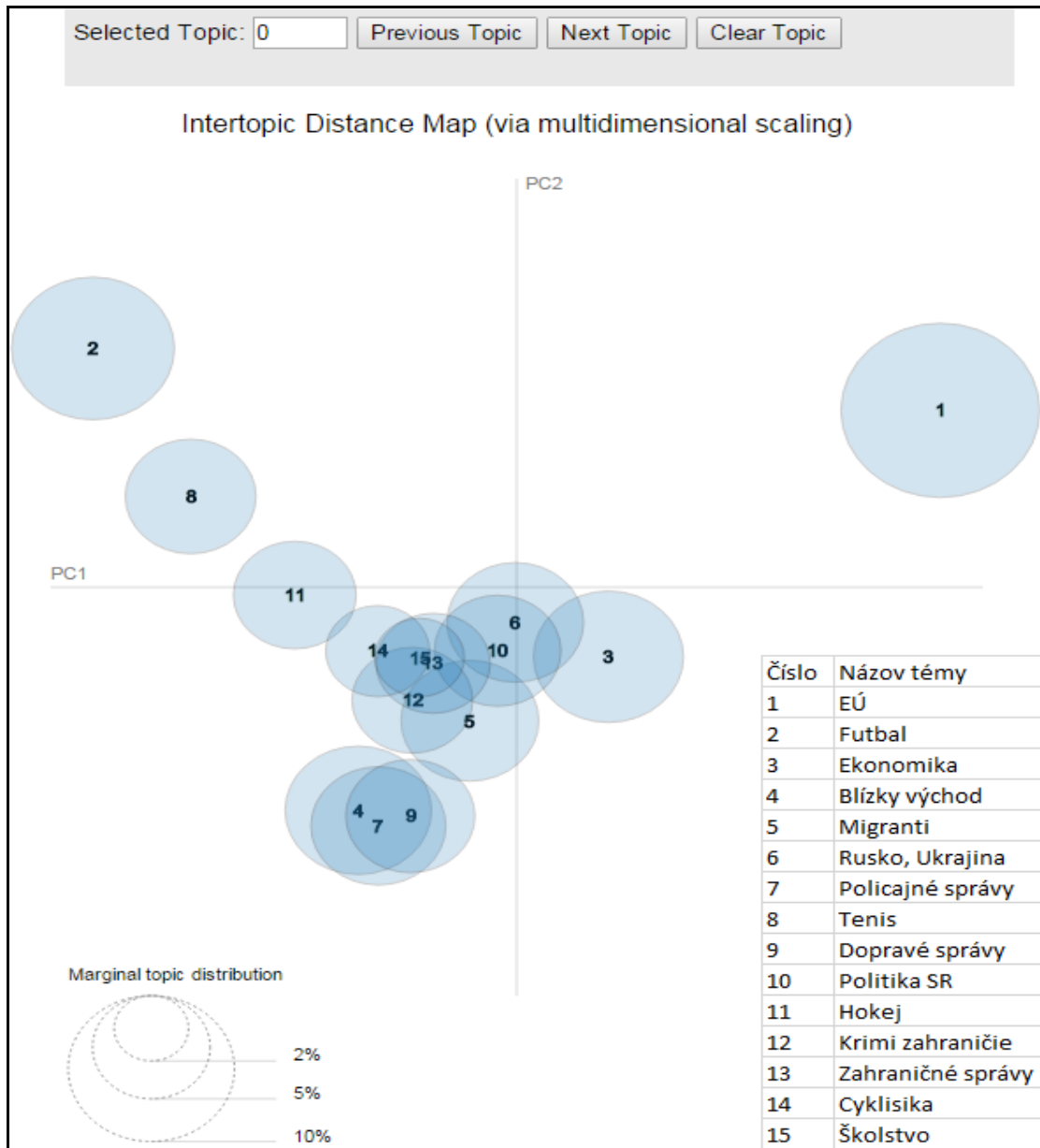
Kľúčové slová	sr nr sieť poriadok exekučný novela beblavý	usa futbal turné chelsea výhra barcelona sumár	srbsko kosovo albánsko rama návšteva neohlásená	eko rusko meny cudzie nákup pozastavený	sr elektro odborníci priemysel elektro- technický nedostatok zep stretnutie kiska	usa new york černochoch policajt uškrtienie dohoda
Krimi zahraničie	0,0024	0,0016	0,0024	0,0047	0,0047	0,0024
EÚ	0,0024	0,0016	0,0024	0,4698	0,0047	0,0024

Zdroj: vlastné spracovanie v jazyku R

Prvý článok s kľúčovými slovami *SR, NR, Sieť, poriadok* atď. sa s pravdepodobnosťou 96,63 % týka témy *politika SR*. S ešte vyššou pravdepodobnosťou je v poradí druhý článok priradený k téme *futbal*. Pravdepodobnosť súvislosti tretieho článku s kľúčovými slovami ako *Srbsko, Kosovo, Albánsko* je rozdelená medzi dve témy – *migranti* a *Rusko, Ukrajina*. Príkladom takéhoto rozdelenia pravdepodobnosti je aj nasledujúci článok, ktorý sa týka témy *ekonomika a politika SR*. Posledný článok zobrazený v našej tabuľke je príkladom článku z oblasti zahraničných policajných správ s kľúčovými slovami, ako *USA, New York, černochoch, policajt*.

Takto upravená matica obsahujúca hodnoty pravdepodobnosti, s akou jednotlivé články súvisia s vytvorenými témami, bude vstupom do ďalšej časti našej analýzy. Charakteristiky článkov prostredníctvom pätnástich premenných využijeme na segmentáciu čitateľov týchto článkov.

Graf č. 1: Interaktívna vizualizácia tém



Zdroj: vlastné spracovanie v jazyku R²

Druhou maticou potrebnou na vizualizáciu tém je matica ϕ charakterizujúca vzťah tém a kľúčových slov, na základe ktorej sme v predchádzajúcej časti určili názvy tém. Hodnoty koeficientov ϕ sme určili ako podiel počtu priradení slov k jednotlivým témam a celkového počtu priradení slov ku všetkým témam spolu.

Vizualizácia vytvorených tém je možná prostredníctvom interaktívnej mapy poskytujúcej všeobecný pohľad na model ako celok. Na základe nej vieme analyzovať dominanciu jednotlivých tém v rámci modelu či vzájomnú podobnosť tém.

Jednotlivé témy znázorňujú kružnice s proporcionálnym obsahom k pomeru³, ktorý

² Číslovanie tém nie je zhodné s označením vo výstupoch modelu, názvy tém uvádzame v legende.

³ LDAvis Vignette, dostupné na internete, dátum posledného prístupu k zdroju: november 2016: <<https://cran.r-project.org/web/packages/LDAvis/vignettes/details.pdf>>

vyjadruje odhadovaný počet slov vybraných na základe danej témy v generatívnom procese. Na základe toho platí, že čím väčšia je celková frekvencia výskytu slov charakteristických pre danú tému, tým väčší je obsah kružnice, ktorou sa označuje daná téma.

V našom prípade (graf č. 1) je najviac dominantná téma *EÚ*, nasledovaná témami *futbal* a *ekonomika*. Zaujímavým spôsobom sa od ostatných tém odlišujú športové témy. Najvýznamnejšou spomedzi športových tém je *futbal* ležiaci najviac vychýlene od ostatných.

Najväčší zhluk tvorí skupina tém, v ktorej sa nachádzajú *politika SR*, *migranti* a *ekonomika*. Najmenej významnou v tejto skupine je *školstvo*. Nižšie umiestnená je skupina, ktorú tvoria témy *Blízky východ*, *dopravné správy* a *policiálne správy*.

4. SEGMENTÁCIA ČITATEĽOV SPRAVODAJSKÝCH ČLÁNKOV

Druhou časťou analýzy bola segmentácia čitateľov stránky www.sme.sk s využitím výsledkov získaných v predchádzajúcej časti práce. Analyzovali sme 4 043 článkov, ktoré sme opísali prostredníctvom pätnástich premenných – tém, ktoré sme získali na základe analýzy ich kľúčových slov modelom LDA.

Údaje o článkoch a ich pravdepodobnosti súvisu s vytvorenými témami sme spojili s údajmi o správaní čitateľov na stránke, a tak sme vytvorili profil čitateľov podľa charakteru prečítaného obsahu na webe.

Na získanie informácií o správaní čitateľov na stránke sme spracovali náš druhý zdroj dát – databázu logov servera stránky. Táto databáza obsahuje informácie o každom pozretí stránky v priebehu jedného mesiaca ako čas pozretia, URL adresa pozretej stránky, unikátny identifikátor návštevníka, zariadenie a prehliadač, ktorý použil, či URL adresa stránky, z ktorej návštevník prišiel. Pre našu analýzu sme vybrali z tejto databázy potrebné premenné a upravili ich.

Z dôvodu nepresnej identifikácie návštevníkov používajúcich mobilné zariadenia sme sa zamerali len na analýzu správania čitateľov používajúcich na prezeranie stránky osobné počítače. Podiel pozretí stránky z mobilných zariadení sa pohyboval na úrovni 37,5 % z celkových 64 miliónov pozretí v júli 2015. Vzhľadom na cieľ, ktorým bola segmentácia čitateľov na základe preferovaného obsahu prečítaných článkov, sme do analýzy zahrnuli iba čitateľov, ktorí v priebehu analyzovaného mesiaca prečítali aspoň 5 článkov. Do kategórie tzv. lojálnych čitateľov patrilo 244 457 čitateľov, čo predstavovalo približne 12 % všetkých návštevníkov stránky v danom mesiaci.

V nasledujúcom kroku sme spojili dva zdroje údajov – tabuľku s údajmi o čitateľoch a prečítaných článkoch s tabuľkou vytvorenou procedúrou LDA charakterizujúcou články pomocou pätnástich tém, a to na základe unikátneho identifikátora článkov. Výstupom týchto krokov bola teda matica s 244 457 riadkami (počet čitateľov) a 16 stĺpcami, obsahujúca unikátny identifikátor čitateľov spolu s pätnástimi koeficientmi charakterizujúcimi relevanciu prečítaných článkov k vytvoreným témam. Súčet pravdepodobností sa pre každé pozorovanie (čitateľa) rovná jednej. Ukážku matice v transponovanej podobe zobrazuje tabuľka č. 3.

V riadkoch ukážky matice vstupujúcej do zhlukovej analýzy je 15 tém, v stĺpcoch zobrazujeme 6 vybraných čitateľov. Čitateľ 1 sa najviac zaujíma o témy *EÚ* a *cyklistika*, no čitateľ 2 sa zaujíma o tému *EÚ* dvakrát viac. U tohto čitateľa výrazne prevládajú ďalšie dve témy (*politika SR* a *zahraničné správy*), o zvyšné sa zaujíma menej v porovnaní s inými čitateľmi. Preferencie čitateľa 3 súvisia najmä s problematikou zahraničnej politiky, v jeho prípade prevažujú témy *migranti* a *Rusko a Ukrajina*. Články prečítané čitateľom 4 sa týkali všetkých tém s približne rovnakou pravdepodobnosťou. Naopak, v prípade čitateľa 5 vidíme, že v jeho zozname článkov prevládali športové články, ktoré súviseli najmä s témami *cyklistika*, *hokej* či *futbal*.

Tabuľka č. 3: Ukážka vstupných dát zhlukovej analýzy

Téma	Čitateľ 1	Čitateľ 2	Čitateľ 3	Čitateľ 4	Čitateľ 5
Ekonomika	0,0558	0,0308	0,0878	0,0487	0,0296
Migranti	0,0683	0,0308	0,2504	0,0993	0,0144
Zahraničné správy	0,0548	0,1695	0,0878	0,0466	0,0144
Policajné správy	0,0509	0,0308	0,0084	0,0726	0,1136
Hokej	0,0573	0,0308	0,0084	0,0698	0,2108
Blízky východ	0,0358	0,0308	0,0878	0,0633	0,0307
Futbal	0,0770	0,0308	0,0084	0,0693	0,1659
Politika SR	0,0360	0,1550	0,0084	0,0429	0,0307
Dopravné správy	0,0138	0,0308	0,0084	0,0436	0,0296
Cyklistika	0,1573	0,0308	0,0888	0,1232	0,1975
Školstvo	0,0393	0,0308	0,0084	0,0358	0,0612
Tenis	0,0645	0,0308	0,0084	0,0328	0,0270
Rusko, Ukrajina	0,0633	0,0308	0,1697	0,0876	0,0459
Krimi zahraničie	0,0599	0,0308	0,0878	0,0480	0,0144
EÚ	0,1659	0,3062	0,0809	0,1164	0,0144

Zdroj: vlastné spracovanie v jazyku R

4.1. Metóda *k*-priemerov

S cieľom získať segmenty čitateľov sme sa rozhodli použiť nehierarchickú metódu *k*-priemerov (*k*-means) [9, s. 140]. Jedným z dôvodov bola vyššia efektívnosť spracovania pomerne veľkého počtu 244 457 pozorovaní. Ako mieru podobnosti jednotlivých čitateľov sme použili euklidovskú vzdialenosť, pričom sme overili nezávislosť vstupných premenných ako prvú podmienku na použitie uvedenej miery vzdialenosti. Keďže články boli reprezentované ako reálne vektory v normovanom tvare, nebolo potrebné uvažovať o Mahalanobisovej vzdialenosti [bližšie 8, s. 51 – 80].

Ďalším predpokladom použitia metódy *k*-priemerov je vopred určený počet zhlukov, ktoré chceme dostať. Heuristicky by sme tento počet určili na úrovni štyroch až ôsmich zhlukov, no naše predpoklady sme najskôr overili pomocou grafickej analýzy, ktorá hodnotila výsledky procedúry *k*-priemerov pre počet zhlukov od 2 do 15. Krivka znázorňujúca vnútrozhlukovú variabilitu výraznejšie poklesla pri počte 5 zhlukov, ktorý sme v konečnom dôsledku zvolili za optimálny.

Výsledkom procesu k -priemerov je päť homogénnych zhlukov podobnej veľkosti (tabuľka č. 4). Výnimkou je prvý segment, do ktorého patrí 40 % zo všetkých 244 457 čitateľov. Ide o segment čitateľov, ktorí čítajú články zo všetkých tém približne rovnakým podielom.

Tabuľka č. 4: Početnosť zhlukov čitateľov

Segment	Absolútna početnosť	Relatívna početnosť (%)
1	98 276	40,4
2	32 917	13,5
3	36 458	15,0
4	41 835	17,2
5	33 792	13,9

Zdroj: vlastné spracovanie v jazyku R

Tabuľka č. 5: Charakteristika piatich segmentov čitateľov

Téma	Segment 1	Segment 2	Segment 3	Segment 4	Segment 5
Ekonomika	0,0604	0,0679	0,0335	0,0505	0,0354
Migranti	0,1124	0,0645	0,0367	0,0570	0,0504
Zahraničné správy	0,0547	0,0463	0,0251	0,0361	0,0321
Policajné správy	0,0612	0,2481	0,0378	0,0516	0,0523
Hokej	0,0528	0,0418	0,1638	0,0408	0,0660
Blízky východ	0,0789	0,0390	0,0246	0,0392	0,0288
Futbal	0,0456	0,0357	0,3030	0,0347	0,0679
Politika SR	0,0568	0,0516	0,0273	0,0394	0,0290
Dopravné správy	0,0510	0,0879	0,0296	0,0341	0,0376
Cyklistika	0,0802	0,0748	0,1050	0,0710	0,3808
Školstvo	0,0310	0,0332	0,0157	0,0213	0,0179
Tenis	0,0329	0,0260	0,0742	0,0236	0,0473
Rusko, Ukrajina	0,1086	0,0632	0,0364	0,0616	0,0452
Krimi zahraničie	0,0575	0,0422	0,0253	0,0351	0,0273
EÚ	0,1159	0,0780	0,0621	0,4037	0,0821

Zdroj: vlastné spracovanie v jazyku R

Tabuľka č. 5 charakterizuje jednotlivé zhluky prostredníctvom všetkých 15 premenných, teda tém, kde sa pre každú tému uvádza hodnota koeficienta θ vypovedajúca o pravdepodobnosti súvislosti medzi danou témou a článkami prečítanými čitateľmi v jednotlivých segmentoch. Súčet hodnôt koeficientov θ pre každý segment sa rovná jednej. Čím je hodnota koeficienta vyššia, tým obľúbenejšia je téma v danom segmente. Najpopulárnejšie témy sú zobrazené zelenou farbou a, naopak, najmenej populárne témy v rámci segmentov znázorňujeme červenou farbou.

Najobľúbenejšou zo všetkých segmentov je u čitateľov téma *EÚ*. Až v štyroch segmentoch sa objavila medzi tromi najpopulárnejšími témami. Už sme spomenuli, že téma *EÚ* dominuje všetkým témam, pretože je charakterizovaná kľúčovými slovami s častým výskytom. Kľúčové slová ako *eko*, *Grécko*, *EÚ*, *Nemecko* či *eurozóna* sa v korpuse dát z júla 2015 objavovali často zrejme najmä preto, že sa v tomto období viedli rokovania o poskytnutí pomoci Európskej únii Grécku.

Z podobného dôvodu sa medzi najpopulárnejšími témami segmentov vyskytuje téma *cyklistika*. O tejto téme sa v období zberu dát veľa písalo, pretože v júli 2015 sa konali preteky Tour de France, ktoré boli na Slovensku mimoriadne sledované.

Medzi najmenej populárne témy segmentov patrí *školsťvo* definované kľúčovými slovami ako *školsťvo*, *vláda*, *školy* či *súdy*. Tieto slová sa v sledovanom období vyskytovali v korpuse s malou frekvenciou. O školsťve sa publikovalo v júli 2015 menej článkov v porovnaní s ostatnými témami.

Na základe výsledkov zhrnutých v tabuľke č. 5 môžeme tvrdiť, že obdobie zberu dát ovplyvňuje výsledky extrahovania tém a tým aj charakteristiku segmentov čitateľov. Napríklad segment 5, pre ktorý je charakteristická najmä téma *cyklistika*, by sme mohli považovať za dočasný; v analýze dát z iného obdobia by sa pravdepodobne neobjavil. Tento fakt však neplatí pre zvyšné segmenty, ktoré môžeme považovať za stabilné segmenty čitateľov webu s nasledujúcimi charakteristikami:

Prvý segment je najpočetnejší spomedzi všetkých piatich, tvorí ho 40 % analyzovaných čitateľov, ktorí čítajú správy zo všetkých tém s približne vyrovnaným podielom. Najviac prevažujú témy *EÚ*, *migranti* a problematika *Ruska a Ukrajiny*. Priemerná hodnota koeficienta *theta* pre tieto témy sa pohybuje na úrovni 0,1, čo je v porovnaní s najdominantnejšími témami iných segmentov nízka hodnota.

Témami s najväčšou prevahou sú v prípade *druhého segmentu* témy z oblasti policajných a dopravných správ, kde medzi najdominantnejšie kľúčové slová patrili slová, ako *polícia*, *nehoda*, *vzi* (indikuje prevzaté články), *tt*, *ba*, resp. *bax* (indikujúce správy z hlavného mesta). Medzi najmenej populárne témy tohto segmentu patria okrem témy školsťva a športových tém najmä témy týkajúce sa zahraničia (*krimi zahraničie*, *Blízky východ*, *zahraničné správy*). Pomocou týchto indikátorov by sme mohli túto skupinu čitateľov označiť ako čitateľov aktualít z lokálnej oblasti.

Tretí segment tvoria čitateľa športových správ. Výrazne dominantnou témou je *futbal*, kde priemerná hodnota pravdepodobnosti relevancie danej témy priradená k článkom prečítaným čitateľmi tohto segmentu dosiahla hodnotu 0,30. Druhou najpopulárnejšou témou tohto segmentu je *hokej* takisto s relatívne vysokou hodnotou koeficientu *theta* 0,16, nasledovaný v tomto období populárnou *cyklistikou*. Téma *tenis* dosahuje v tomto segmente hodnotu 0,07, čo je najvyššia hodnota koeficienta *theta* spomedzi všetkých segmentov pre túto tému.

Vo *štvrtom segmente* výrazne dominuje téma *EÚ*, ktorej priemerná hodnota koeficienta *theta* pre tento segment dosahuje hodnotu až 0,4. Druhé poradie v oblúbenosti tém v tejto skupine čitateľov obsadzuje *cyklistika*, treťou najoblúbenejšou témou je *Rusko a Ukrajina*, no hodnoty koeficientov *theta* pre tieto témy sú výrazne nižšie, dosahujú úroveň 0,07, resp. 0,06. Najmenej oblúbenými témami tohto segmentu čitateľov sú *tenis* a *školsťvo*, ktoré patria medzi najmenej populárne témy aj v ostatných segmentoch. Keďže vo štvrtom segmente extrémne dominuje téma *EÚ*, ktorú charakterizujú kľúčové slová, ako *eko*, *Grécko*, *Nemecko*, *EÚ*, *eurozóna*, *financie*, *banky*, usudzujeme, že ide o segment čitateľov zaujímajúcich sa najmä o správy z oblasti ekonomiky.

Ako sme už spomenuli, *piaty segment*, ktorý tvorí približne 13,9 % všetkých analyzovaných čitateľov, považujeme za dočasný, keďže jeho vznik ovplyvnilo obdobie zberu dát. V tomto zhľuku dominuje v danom období populárna téma *cyklistika*, priemerná hodnota koeficienta *theta* dosahuje pre túto tému úroveň 0,38.

5. ZÁVER

Výsledkami segmentácie spravodajského denníka je päť homogénnych zhľukov čitateľov, ktoré sa navzájom odlišujú preferenciami obsahu článkov. Identifikovali sme segment fanúšikov športových správ, kde dominovali všetky štyri športové témy – *futbal*, *hokej*, *tenis* a *cyklistika*. Taktiež sme zistili, že segment s najväčším počtom čitateľov sa zaujíma o široké spektrum tém. Na základe toho môžeme predpokladať, že čitatelia tohto segmentu čítajú väčšie množstvo článkov z rôznych oblastí a patria tak medzi najlojálnejšiu časť publika. Medzi piatimi segmentmi čitateľov sa objavil jeden dočasný, v ktorom výrazne prevažovala téma cyklistiky, ktorá bola populárna najmä v období zberu dát.

V súčasnosti sa stáva trendom v oblasti spravodajských médií spoplatňovanie obsahu (Paid Content), teda prijímanie platieb od čitateľov za sprístupnenie obsahu vybraných (prípadne všetkých) článkov na stránke. S cieľom zvyšovania zisku prostredníctvom spoplatnenia obsahu sa prikladá dôraz na ďalší cieľ – zvyšovanie spokojnosti a lojality zákazníka. Zaplatenie za sprístupnenie obsahu je viac pravdepodobné u lojálneho čitateľa ako u náhodného návštevníka stránky.

Práve otázka týkajúca sa obsahu spravodajských článkov bola pre nás v tomto príspevku najpodstatnejšia. Analyzovať obsah článkov je možné napríklad na základe sekcie, do ktorej sú priradené, na základe kľúčových slov priradených ku každému článku alebo na základe samotného textu článkov. Výsledkom analýzy tohto druhu môžu byť napríklad segmenty čitateľov, ktorí sú zoskupovaní podľa obsahu článkov, ktoré za určité obdobie prečítali. Výsledky segmentácie čitateľov podľa prečítaného obsahu poskytujú pre tvorcov webu cenné informácie o svojom publiku, ktoré na základe tradičných nástrojov webovej analytiky nemožno získať.

Odhalenie najpopulárnejších tém preferovaných čitateľmi umožňuje tvorcovi webu získať lepší prehľad o tom, o ktorý typ obsahu je najmenší záujem a, naopak, ktorá téma je pre čitateľov najzaujímavejšia. Ak sa napríklad ukáže, že medzi najpopulárnejšie témy patrí taká, na ktorú sa tvorcovia nezameriavajú takou mierou ako na ostatné, tak je ideálne pozmeniť štruktúru obsahu článkov a zvýšiť počet článkov tvorených na dopytovanú tému a zvýšiť tým lojalitu návštevníkov.

Môže tiež nastať prípad, keď tvorcovia obsahu pomocou takejto analýzy zistia, že články zamerané na témy, z ktorých publikujú väčší počet článkov v porovnaní s ostatnými témami, oslovujú iba malú časť publika. V tomto prípade ponuka prevyšuje dopyt a je pre tvorcov optimálne znížiť počet publikovaných článkov na danú tému.

Informácie z analýz takéhoto druhu sa dajú využiť napríklad aj v oblasti marketingu, kde možno zostavovať emaily odberu noviniek na základe obľúbených tém čitateľa a zvyšovať tak frekvenciu návštev a lojalitu odberateľov noviniek. V prípade spravodajských webov spoplatňujúcich obsah je možné na základe segmentácie čitateľov vytvárať ciele marketingové kampane so zľavami na časť

obsahu podľa záujmov čitateľov alebo vytvárať špeciálne balíčky predplatného zamerané napr. len na čitateľov ekonomiky, regionálnych správ alebo športu.

Uvedený článok poskytuje komplexný pohľad na štruktúru čitateľov spravodajského webu vzhľadom na preferované témy. Spojením výsledkov analýzy textových dát charakterizujúcich obsah článkov a údajov o návštevnosti stránky sme opísali charakter čitateľov z nového, netradičného pohľadu. Doplňujúcou úlohou tohto článku bola ukážka aplikácie hĺbkovej analýzy textu v oblasti webovej analytiky ako možnosť zlepšenia tradičných analýz o návštevnosti stránky.

LITERATÚRA

- [1] BLEI, D.: Probabilistic Topic Models. In: Communications of the ACM, 2012, Vol. 55, No. 4, p. 77-84.
- [2] BLEI, D. – LAFFERTY, J. D.: Topic models. In: Text Mining: Classification, Clustering and Applications, 2009, p. 71-93.
- [3] FELDMAN, R. – SANGER, J.: The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data. Cambridge University Press, 2007. ISBN 978-0-521-83657-9.
- [4] HEARST, M. A.: Untangling text data mining. In: Proceedings of ACL '99: the 37th annual meeting of the Association for Computational Linguistic. University of Maryland, 1999, p. 3-10.
- [5] CHAKRABORTY, G. – MURALI, P. – SATISH, G.: Text Mining and Analysis: Practical Methods, Examples and Case Studies Using SAS. Cary, NC:SAS Institute Inc., 2013. ISBN 978-1-61290-551-8.
- [6] KROEZE, J. H. – MATTHEE, M. C. – BOTHMA, T. J. D.: Differentiating Data- and Text-Mining Terminology. In: Proceeding of SAICSIT, 2003, p. 93-101. ISBN:1-58113-774-5.
- [7] PARALIČ, J. a kol.: Dolovanie znalostí z textov. Košice: Equilibria, 2010. ISBN 978-80-89284-62-7.
- [8] ŘEZÁNKOVÁ, H. – HÚSEK, D. – SNÁŠEL, V.: Shluková analýza dat. Příbram: Professional Publishing, 2007. ISBN 978-80-86946-26-9.
- [9] STANKOVIČOVÁ, I. – VOJTKOVÁ, M.: Viacrozmerné štatistické metódy s aplikáciami. Bratislava: Iura Edition, 2007. ISBN 978-80-8078-152-1.

RESUME

The outcome of the analysis described in this paper, is to identify the segments of online news readers based on topic preferences. The traditional methods of web traffic are completed with text analyses, in order to describe the readers from a new perspective.

Based on the results of cluster analysis, we concluded that the period of data collection affects the results of topic extraction and thus the readers' segment description. One of the five segments were composed of readers preferring the topic of cycling, which was specific for the analyzed period when the Tour de France was held. The other four segments of readers are considered stable, and expected to appear in a similar form also in other month's analysis. The most important segment was the most numerous one, consisting of approximately forty percent of the analyzed readers. Readers in this segment, prefer reading articles on all topics with approximately the same proportions. We also identified a segment consisting of sports fans, where particularly the sport topics were dominant.

PROFESIJNÝ ŽIVOTOPIS

Ing. Bianka Parmová v máji 2016 ukončila druhý stupeň štúdia na Fakulte hospodárskej informatiky Ekonomickej univerzity v Bratislave v študijnom odbore štatistické metódy v ekonómii. V súčasnosti pracuje ako dátová analytička (Data Scientist) vo firme Piano Media, s. r. o. Článok vychádza z výsledkov jej diplomovej práce, ktorú vypracovala pod vedením doc. Ing. Márie Vojtkovej, PhD.

Doc. Ing. Mária Vojtková, PhD., pôsobí vo funkcii docentky na Katedre štatistiky Fakulty hospodárskej informatiky Ekonomickej univerzity v Bratislave. Vo svojej vedeckovýskumnej a pedagogickej činnosti sa venuje viacrozmerným štatistickým metódam, ktoré sú zamerané na aplikáciu viackriteriálneho hodnotenia v rôznych oblastiach sociálno-ekonomického života, hľadanie skrytých vzťahov pomocou metód zníženia dimenzie, segmentáciu, čiže zhlukovanie podobných objektov charakterizovaných určitými vlastnosťami, a určenie diskriminačnej funkcie ako spôsobu rozlíšenia medzi vytvorenými skupinami a klasifikáciu nových objektov. Je spoluautorkou vedeckých monografií, niekoľkých učebníc, skrípt a mnohých vedeckých článkov publikovaných doma i v zahraničí.

KONTAKT

bianka.parmova@gmail.com

maria.vojtkova@euba.sk

Roman PAVELKA
Štatistický úrad SR

ANALÝZY ČASOVÝCH ŘAD POMOCÍ PROGRAMOVACÍHO JAZYKA R

ANALYSES OF TIME SERIES WITH THE R PROGRAMMING LANGUAGE

1. ÚVOD

Posloupnost hodnot sledovaného ekonomického, fyzikálního či jiného ukazatele, které jsou uspořádány v čase, se nazývá časová řada [6]. S fenoménem časové řady se dříve nebo později setká každý, kdo je postaven před problém analyzovat určitý ekonomický jev, ať už jde o makroekonomické ukazatele – např. vývoj agregátů tvorby a užití hrubého domácího produktu, inflace, nezaměstnanosti apod. – nebo o některé dílčí ukazatele, jako je vývoj kurzů cizích měn, peněžní zásoby, cen akcií na kapitálovém trhu či průmyslové nebo stavební produkce.

Analýza časových řad je jednou z nejvýznamnějších aplikací kvantitativních metod v ekonomii. Kvantitativní metody a techniky používané v analýzách časových řad zpravidla patří k velmi komplexním a výpočetně velmi náročným. Z tohoto důvodu si příslušná statistická analýza vyžaduje zapojení výkonné výpočetní techniky včetně nezbytného specializovaného statistického softwaru. Ke statistickým analýzám časových řad existuje množství statistických programů. Většina specializovaných statistických programů (např. Stata, EViews, SPSS, apod.) nabízí srovnatelné analytické metody a techniky, které umožňují požadovanou analýzu a modelování časových řad. Jejich společnou nevýhodou je však skutečnost, že jejich pořízení a aktualizace je ekonomicky velmi nákladné.

Jedním z volně dostupných (a tedy je jeho použití zcela zdarma) specializovaných statistických programů je programovací jazyk R. Uvedený programovací jazyk se vyvinul v komplexní systém na manipulaci s daty, jejich statistickou analýzu a zpracování i grafickou reprezentaci. Programovací jazyk R vychází z jazyka S, který byl vyvinut v Bell Laboratories (předtím AT&T, nyní Lucent Technologies) jako komerční softwarový nástroj pro analýzy a vyhodnocování dat. Programovací jazyk R však představuje volně dostupný softwarový nástroj, který je šířen a používán v rámci tzv. open-source¹ projektů nadace *Free Software Foundation*². Snadná dostupnost a volné používání programovacího jazyka R umožňuje jeho neustálý rozvoj, který je realizován zejména v podobě mnoha doplňkových programových balíčků (*add-on packages*) s knihovnamí a funkcemi na různé typy specializovaných analýz. Proto programovací jazyk R získává stále větší význam nejen v komerční a akademické sféře, resp. ve vědeckovýzkumných pracovištích, ale nabývá na důležitosti i při

¹ Otevřený software (anglicky *open-source software* nebo *open software*, zkratka *OSS*) je počítačový software s otevřeným zdrojovým kódem. Otevřenost zde znamená jak technickou dostupnost kódu, tak legální dostupnost – licenci software, která umožňuje, při dodržení jistých podmínek, uživatelům zdrojový kód využívat, například prohlížet a upravovat.

² Nadace *Free Software Foundation*, česky *Nadace pro svobodný software*, byla založena v roce 1985 s cílem podporovat práva uživatelů počítačů používat, studovat, kopírovat, modifikovat a redistribuovat počítačové programy.

produkci výstupů státní statistiky v národních statistických úřadech. V současné době se programovací jazyk R používá v produkčních procesech národních statistických úřadů Rakouska, Itálie, Holandska, Rumunska a částečně i Slovenské republiky a jiných států EU i v samotném Eurostatu.

2. POUŽITÉ ČASOVÉ ŘADY

Pro účely tohoto příspěvku byly vybrány ekonomické časové řady, které popisují nejdůležitější ekonomické ukazatele produkované státní statistikou. Jedná se o časovou řadu hrubého domácího produktu měřeného v běžných cenách v mld. EUR, která byla zjišťována se čtvrtletní periodicitou od 1. čtvrtletí 1995 do 2. čtvrtletí roku 2016 [8]. Další sledovanou řadou je časová řada meziměsíční míry inflace (v jednotkách %), jejíž hodnoty byly pozorovány s periodicitou měsíční, a to od ledna roku 2002 do srpna roku 2016 [9]. Jako poslední časová řada použitá v rámci tohoto příspěvku byla vybrána časová řada indexů spotřebitelských cen podle Klasifikace individuální spotřeby podle účelu (COICOP) zjišťovaná měsíčně ve stejném časovém období jako předchozí časová řada [10]. Vybrané časové řady jsou volně dostupné na WWW stránkách Statistického úřadu SR v databázi SlovStat³. Výběr časových řad byl proveden tak, aby časové řady byly dostatečně dlouhé a také významné z ekonomického pohledu. Délka jednotlivých časových řad je určena dostupností údajů v databáze SlovStat.

3. ZÁKLADNÍ METODY ANALÝZY ČASOVÝCH ŘAD POMOCÍ R

3.1. Import časové řady do programovacího jazyku R a její vizualizace

Aby bylo možné časové řady analyzovat v programovacím jazyku R, je nutné data zkoumaných časových řad do tohoto systému importovat. Import dat do R může být proveden různými způsoby – například R nabízí programové prostředky pro import ASCII a binárních dat, dat z jiných aplikací nebo import dat přímo z databází pomocí ODBC připojení. Způsobem přípravy dat, který lze považovat jako optimální, je vytvoření textového souboru, který obsahuje výlučně samotná data časových řad strukturovaně uložených do pojmenovaných sloupců. Data z každého sloupce jsou oddělena středníkem, desetinná místa odděluje čárka. Každý řádek souboru reprezentuje jedno pozorování časové řady v příslušném časovém období. Jakmile importovaná data dosáhnou tuto podobu, celý takto vytvořený soubor se uloží ve formě textového souboru s rozšířením CSV⁴.

Po konverzi dat do textového formátu CSV data časových řad mohou být vloženy do programovacího jazyka R. Příkazem

³ Základní informace o databázi: databáze SLOVSTAT obsahuje časové řady ukazatelů hospodářského a sociálně-ekonomického vývoje za Slovenskou republiku. Údaje z různých statistických okruhů se prezentují ve formě tabulek v měsíčních, čtvrtletních nebo ročních časových řadách.

⁴ CSV (Comma-separated values, hodnoty oddělené čárkami) je jednoduchý souborový formát určený pro výměnu tabulkových dat. Soubor ve formátu CSV sestává z řádků, ve kterých jsou jednotlivé položky odděleny znakem čárka (.). Hodnoty položek mohou být uzavřeny do uvozovek ("), což umožňuje, aby text položky obsahoval čárku. Pokud text položky obsahuje uvozovky, jsou tyto zdvojeny. Jelikož se v některých jazycích včetně češtiny čárka používá v číslech jako oddělovač desetinných míst, existují varianty, které používají jiný znak pro oddělování položek než čárku, nejčastěji středník, případně tabulátor (taková varianta se pak někdy označuje jako TSV, Tab-separated values). Variantu se středníkem (ale stále pod názvem CSV) používá např. Microsoft Excel v české verzi Microsoft Windows (řídí se oddělovačem zadaným v místním a jazykovém nastavení). Díky jednoduchosti, nenáročnosti a čitelnosti i bez specializovaného software se tento formát používá pro výměnu informací mezi různými systémy.

```
HDP <- read.csv("D:/pom/HDP/dataTab_col.csv", header=T, dec="," , sep=";")
```

se časová řada obsahující informace o čtvrtletních hodnotách HDP načte do programovacího jazyka R. Použitý příkaz jazyka R obsahuje mimo potřebné cesty k importovanému souboru ("D:/pom/HDP/dataTab_col.csv") také informaci o tom, jestli importovaný soubor obsahuje i názvy datových sloupců a čím jsou odděleny jednotlivé údaje na 1 řádku. Proměnná nazývaná HDP obsahuje ukazatel na datové pole (v programovacím jazyku R se nazývá *dataframe*). Časové řady se ukládají jako datové matice, k nimž je možné přistupovat pomocí hranatých závorek ([]). Řádky a sloupce datových matic jsou dostupné pomocí příkazů *název_dat_matice[řádek, sloupec]*.

Programové prostředí jazyka R obsahuje z pohledu časových řad struktury pro reprezentaci a analýzu dat časových řad. Fundamentální strukturami [7] pro časové řady jsou programové třídy nazývané *ts* a *zoo*. Objekty třídy *ts* jsou buď numerické vektory v případě jednorozměrných řad, nebo numerická matice pro vícerozměrné řady. Tyto objekty slouží především pro uchovávání rovnoměrně rozložené časové řady (pomocí číselných časových razítek). Proto se velmi dobře hodí pro roční, měsíční, čtvrtletní data, atd. Atribut třídy *ts* nazývaný *tsp* (*time series properties*) reprezentuje základní vlastnosti časové řady. Jedná se o vektor o délce 3 prvků, který obsahuje počátku časové řady, ukončení časové řady (v časových jednotkách) a její periodicitu (frekvenci). Zobecněním objektů typu *ts* jsou objekty typu *zoo*, které umožňují pracovat i s časovými řadami neperiodickými, resp. umožňují práci i s časovými řadami s chybějícími pozorováními.

Příkazem programovacího jazyka R

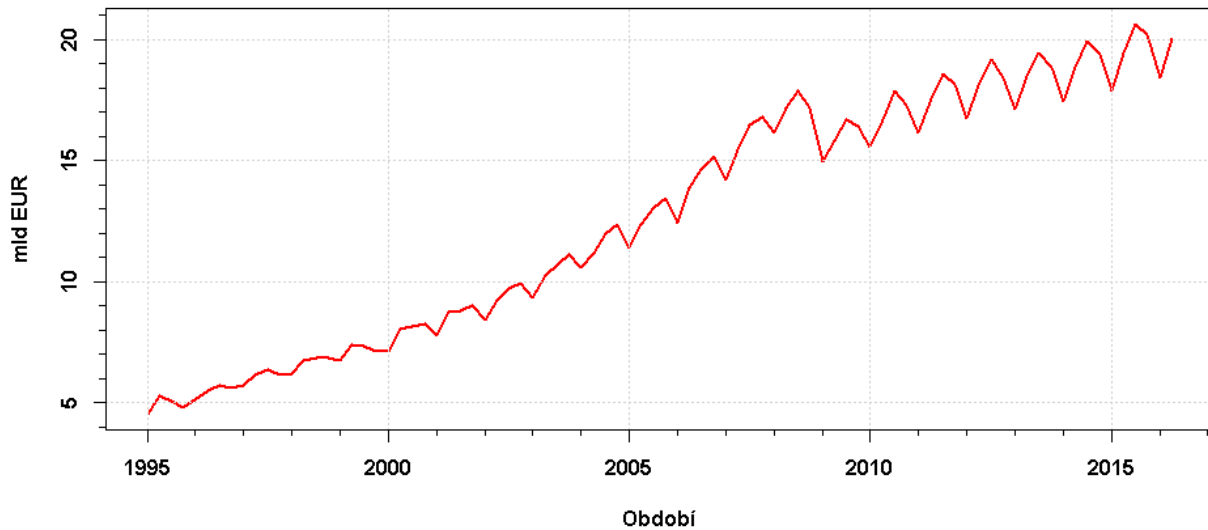
```
HDPTs<-ts(HDP[,2], frequency = 4, start=c(1995,1))
```

se vytvoří v prostředí programu R datová struktura jednorozměrné časové řady, která obsahuje pozorované hodnoty HDP v běžných cenách v miliardách EUR, která začíná v 1. čtvrtletí roku 1995 se čtvrtletní frekvencí pozorování.

Prvotní informace pro analýzu časových řad získáme ze spojnicových grafů. Jejich princip spočívá v zakreslení jednotlivých hodnot časové řady do souřadných os, na kterých jsou vyznačeny příslušné stupnice. Například zakreslení průběhu časové řady HDP v běžných cenách miliard EUR bylo realizováno pomocí příkazu

```
plot(HDPTs, type="l", lwd=2, col="red", xlab="Čtvrtletí", ylab="mld EUR", main="HDP v b. c. v mld EUR")
```

Příkaz k vykreslení grafu *plot()* dovoluje mnoho volitelných parametrů, jeden z parametrů je *type="l"*, který nastavuje typ grafu na spojnicový. Výsledkem uvedeného grafického příkazu je graf průběhu časové řady „HDP v b. c. v mld. EUR“, který je uveden na obrázku 1.

Obrázek č. 1: Vizuální průběh časové řady „HDP v b. c. v mld. EUR“

Zdroj: vlastní konstrukce

Podobným způsobem se zobrazí vizuální průběh časové řady roční míry inflace a časové řada indexů spotřebitelských cen (CPI) podle Klasifikace individuální spotřeby podle účelu (COICOP) zjišťovaných měsíčně.

Do programového prostředí jazyka R lze importovat časové řady nejen textového formátu (s koncovkou .CSV), ale také formátu SAS, strukturovaných formátů, jako je DBF, resp. DB, apod.

3.2. Analýza časové řady pomocí rozkladu na její složky

Klíčovým konceptem v tradiční analýze časových řad je rozklad časové řady Y_t na složku trendu T_t , sezónní složku S_t a složku reziduální e_t .

Nejběžnější metodou pro získání trendu časové řady je použití lineární filtrace [2] na dané časové řadě. Trend T_t časové řady Y_t lze vyjádřit ve tvaru

$$T_t = \sum_{i=-\infty}^{\infty} \lambda_i y_{t+i}, \quad (1)$$

kde λ_i jsou koeficienty filtrace a y_t jsou pozorované hodnoty časové řady v čase t .

K nejpoužívanější třídě lineárních filtrů náležejí tzv. klouzavé průměry se stejnými váhami (koeficienty):

$$T_t = \frac{1}{2a+1} \sum_{i=-a}^a y_{t+i}. \quad (2)$$

Platí-li pro trend časové řady vztah (2), jsou filtrované hodnoty časových řad v periodě φ reprezentované klouzavými průměry $x_{\varphi-a}, \dots, x_{\varphi}, \dots, x_{\varphi+a}$. Potom koeficienty lineární filtrace nabývají hodnot $\left\{ \frac{1}{2a+1}, \dots, \frac{1}{2a+1} \right\}$.

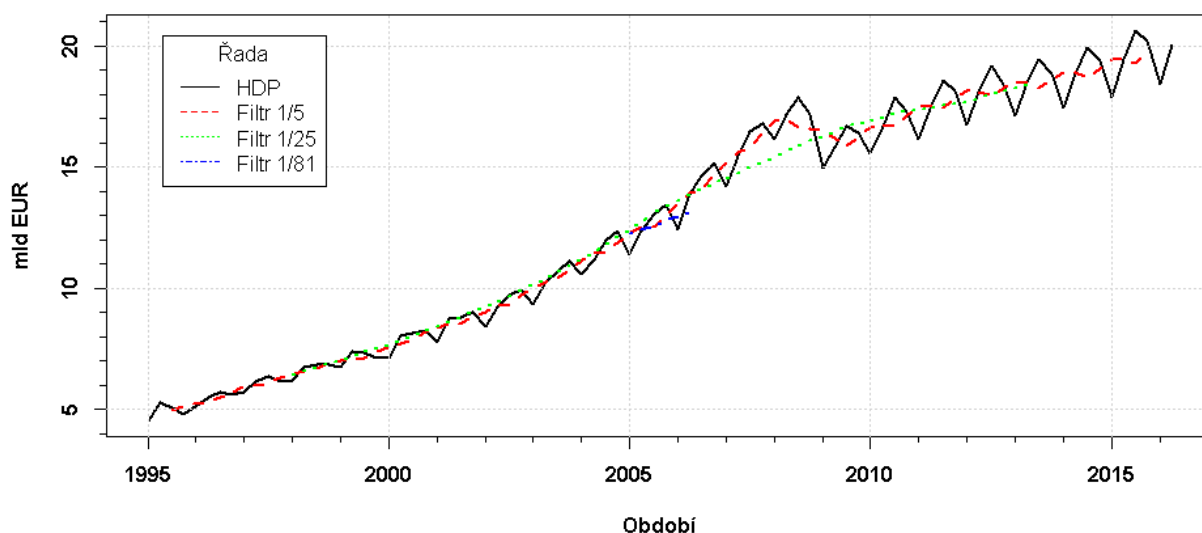
Použije-li se metoda klouzavých průměrů s hodnotami koeficientu $a = 2, 12$ a 40 na hledání trendu časové řady HDP v běžných cenách miliard EUR, vzniknou následující lineární filtry:

$$a = 2 : \lambda_i = \left\{ \frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5} \right\}. \quad (3)$$

$$a = 12 : \lambda_i = \underbrace{\left\{ \frac{1}{25}, \dots, \frac{1}{25} \right\}}_{25 \times} \quad (4)$$

$$a = 40 : \lambda_i = \underbrace{\left\{ \frac{1}{81}, \dots, \frac{1}{81} \right\}}_{81 \times} \quad (5)$$

Obrázek č. 2: Časová řada „HDP v b. c. v mld. EUR“ a odhad trendu pro vybrané filtry



Zdroj: vlastní konstrukce

Filtrování časových řad v programovacím jazyku R je zajištěno pomocí příkazu `filter()`. Sekvence příkazů programovacího jazyka R, které realizují lineární filtraci časové řady HDP v běžných cenách v mld. EUR podle vztahů (3 - 5), je následující:

```
plot(HDPts,type="l", xlab="", ylab="", main="")
HDP.T1 <- filter(HDPts,rep(1/5,5))
HDP.T2 <- filter(HDPts,rep(1/25,25))
HDP.T3 <- filter(HDPts,rep(1/81,81))
lines(HDP.T1, col="red", lty=2, lwd=2)
lines(HDP.T2, col="green", lty=3, lwd=2)
lines(HDP.T3, col="blue", lty=4, lwd=2)
```

Výsledek lineární filtrace časové řady HDP v běžných cenách v mld. EUR je znázorněn na obrázku 2.

Další možností pro odhad trendu časové řady Y_t je dekompozice pomocí funkce `decompose` časové řady Y_t na složku trendu T_t , sezónní složku S_t . Z diferencí $Y_t - T_t$ je odhadnuta složka reziduální e_t . Funkce nejprve determinuje trendovou složku pomocí klouzavých průměrů, kterou vyděluje z časové řady. Potom je počítána sezónní složka průměrováním pozorovaných hodnot pro každou časovou jednotku během celého sledovaného období. Nakonec je odstraněním složky trendové a sezónní determinována reziduální komponenta časové řady. Funkce `decompose` poskytuje multiplikatívní (6) nebo aditivní (7) model dekompozice časové řady, tj.

$$Y_t = T_t * S_t * e_t, \quad (6)$$

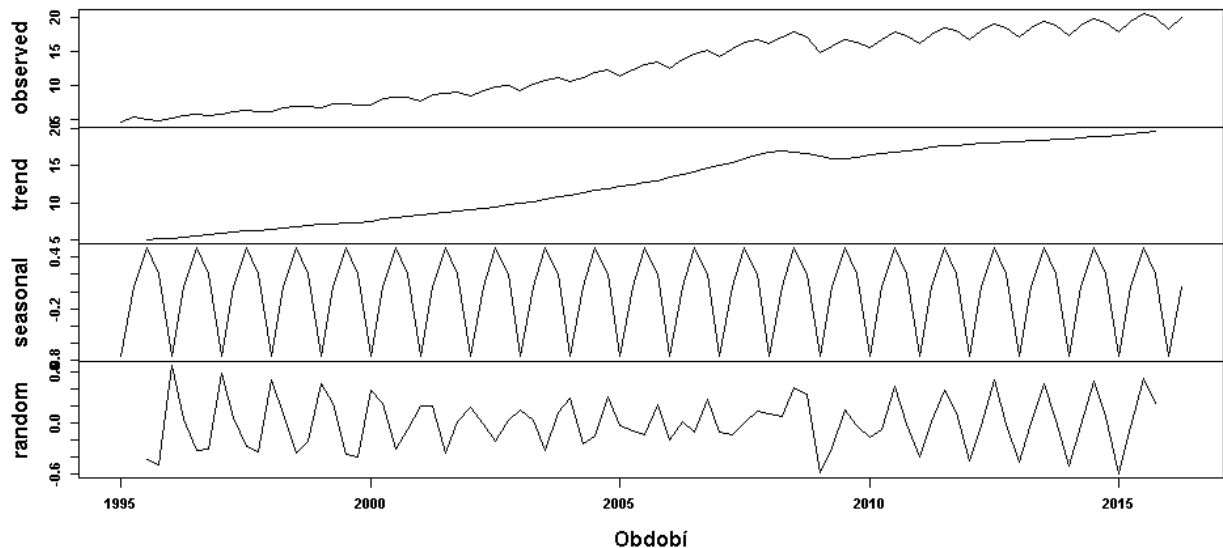
$$Y_t = T_t + S_t + e_t, \quad (7)$$

Symbol S_t představuje označení pro sezónní komponentu časové řady Y_t , e_t reprezentuje reziduální složku řady Y_t . Trendová složka T_t časové řady je označena jako T_t . Funkce *decompose* se používá ve tvaru

```
compYAdd <- decompose(HDPts , type = "additive")
plot(compYAdd)
```

Vytvořený datový objekt *compYAdd* je maticová struktura jednorozměrné časové řady Y_t , která obsahuje pozorované hodnoty HDP v běžných cenách v miliardách EUR, řadu hodnot reprezentující trend T_t časové řady Y_t , hodnoty její sezónní složky S_t a řadu hodnot reziduální složky e_t . Aditivní model podle (7) časové řady HDP v běžných cenách v mld. EUR je znázorněn na obrázku 3.

Obrázek č. 3: Časová řada „HDP v b. c. v mld. EUR“ a její aditivní dekompozice

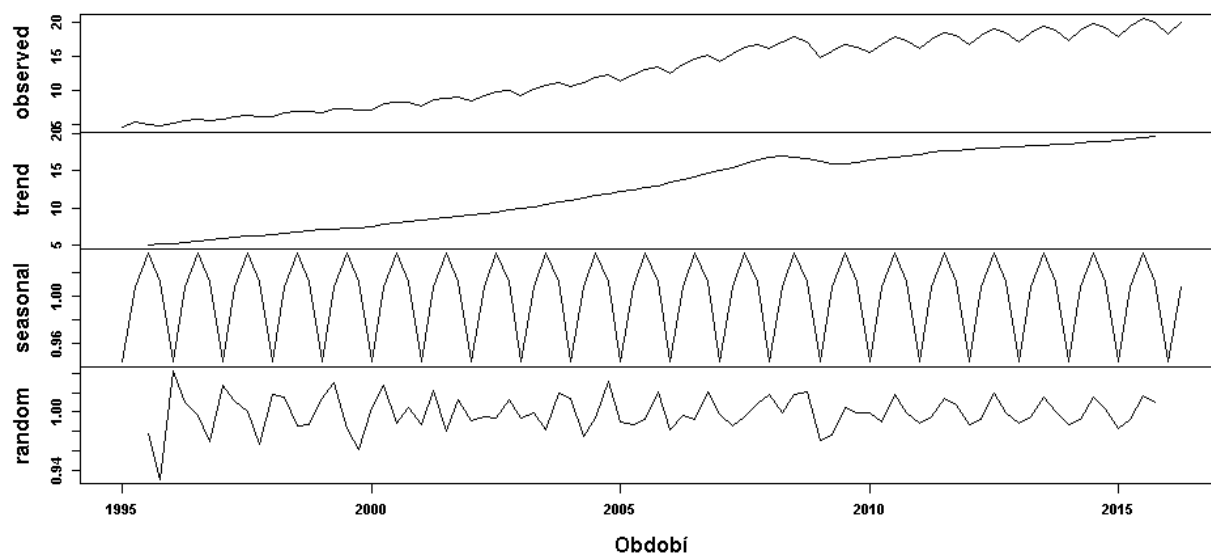


Zdroj: vlastní konstrukce

Stejnou funkcí lze provést také i multiplikativní dekompozici datové řady, tj.

```
compYMul <- decompose(HDPts , type = "multiplicative")
plot(compYMul)
```

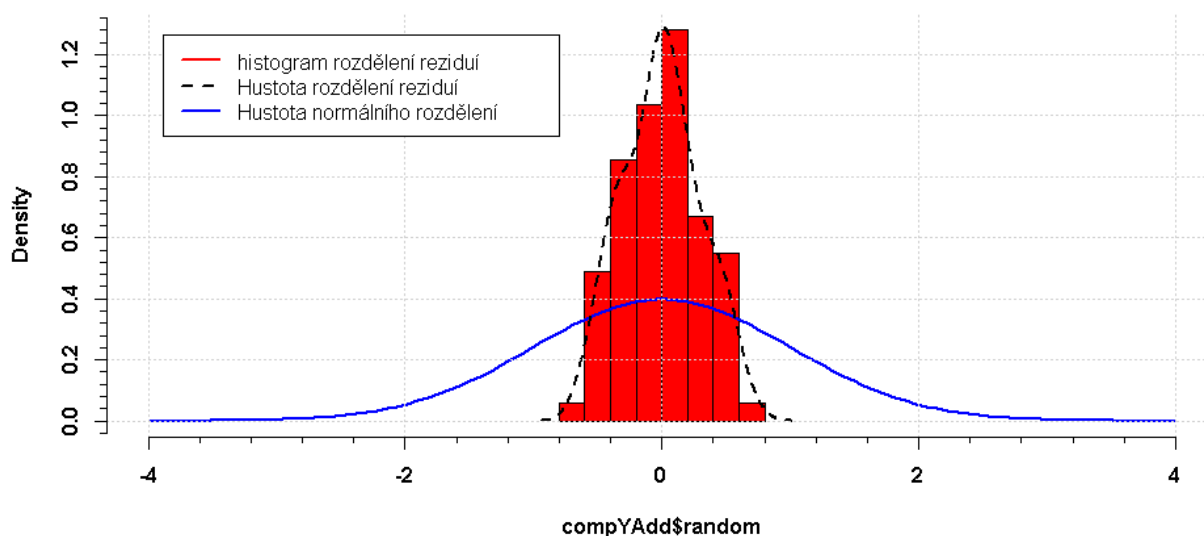
Multiplikativní model podle (6) časové řady HDP v běžných cenách v mld. EUR je znázorněn na obrázku 4.

Obrázek č. 4: Časová řada „HDP v b. c. v mld. EUR“ a její multiplikativní dekompozice

Zdroj: vlastní konstrukce

Často je užitečné posoudit vlastnosti časových řad z hlediska jejich pravděpodobnostního rozdělení. Tento přístup umožňuje na jedné straně porovnávat charakter časových řad, hledat jejich obdobné a rozdílné rysy a činit předběžné závěry o jejich závislosti, na druhé straně vytváří podmínky pro jejich modelování. V této souvislosti je třeba zdůraznit, že při modelování ekonomických časových řad se často vychází z dekompozičního principu, tj. jejich rozkladu na jednotlivé složky – trendovou, cyklickou, sezónní a reziduální. Kritériem kvalitně provedeného rozkladu, tedy oddělení systematické části od části nesystematické, je posouzení vlastností odhadu reziduální složky, tato složka by měla mít nesystematický charakter [2].

Pravděpodobnostní vlastnosti časové řady mohou být posuzovány například pomocí histogramu reziduální složky ve srovnání s hustotou normálního rozdělení.

Obrázek č. 5: Rozdělení reziduální složky časové řady „HDP v b. c. v mld. EUR“

Zdroj: vlastní konstrukce

Ke zjištění pravděpodobnostních vlastností reziduální složky e_t časové řady HDP v běžných cenách v miliardách EUR po aditivní dekompozici (podle (7)) byl zkonstruován histogram a jádrová hustota rozdělení reziduální složky ve srovnání s hustotou normálního rozdělení, což zachycuje obrázek 5. K vytvoření tohoto grafu byly použity následující příkazy:

```
hist(compYAdd$random, xlim=c(-4,4), prob=T, col="red")
mu<-mean(compYAdd$random)
sigma<-sd(compYAdd$random)
x<-seq(-4,4,length=100)
y<-dnorm(x, mu, sigma)
lines(x,y,lwd=2,col="blue")
lines(density(compYAdd$random), lwd=2, col="black", lty=2)
```

Normalitu rozdělení reziduální složky e_t časové řady Y_t lze posuzovat i podle jiných grafů – například pomocí pravděpodobnostních nebo kvantilových grafů, případně je možné reziduální složku testovat pomocí různých statistických testů.

3.3. Regresní analýza časové řady

Pro dekompozici časové řady Y_t lze využít i neparametrickou metodu lokálně váženou polynomičnou regresi. Jedná se o váženou metodu nejmenších čtverců, kde váhová funkce dává největší váhu bodům datovým bodům nejbližším bodu odhadu a nejnižší váhu datovým bodům nejvzdálenějším. Příkladem váhové funkce může být:

$$w(t) = \left(1 - |t|^3\right)^3 \cdot I[|t| < 1], \quad (8)$$

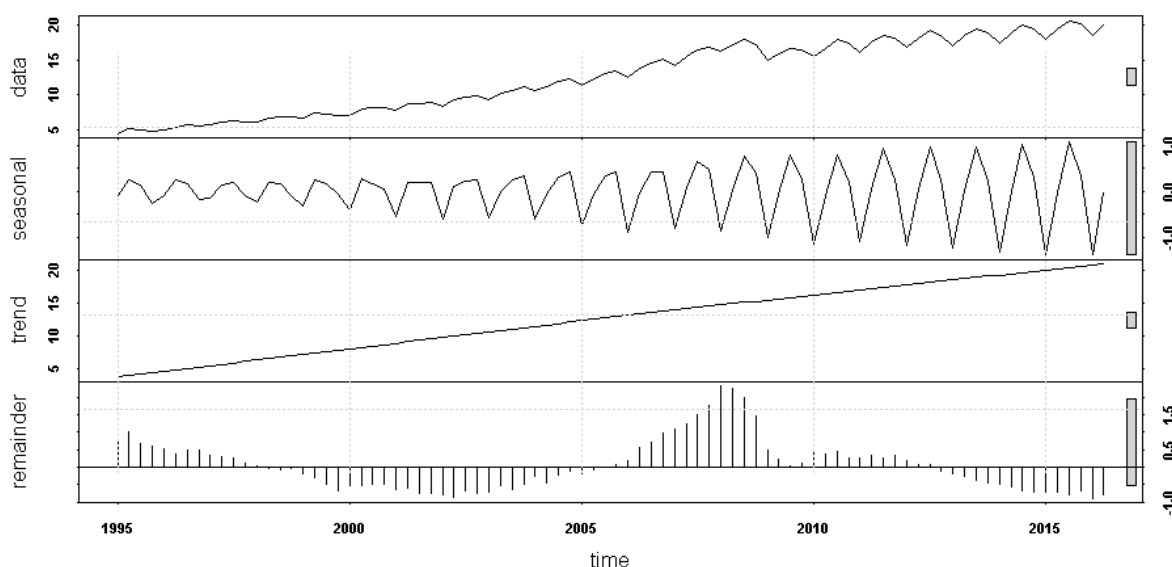
kde $w(t)$ je označením pro váhovou funkci.

V programovacím systému R lze provést dekompozici pomocí neparametrické regrese prostřednictvím příkazu *stl*, v jehož rámci lze zadat odpovídající frekvenci pozorování i délku analyzované časové řady:

```
decompY <- stl(HDPts, s.window = frequency(HDPts), t.window = length(HDPts))
```

Po provedené dekompozici je časová řada i její složky zobrazeny na obrázku 6.

Obrázek č. 6: Modelování časové řady „HDP v b. c. v mld. EUR“

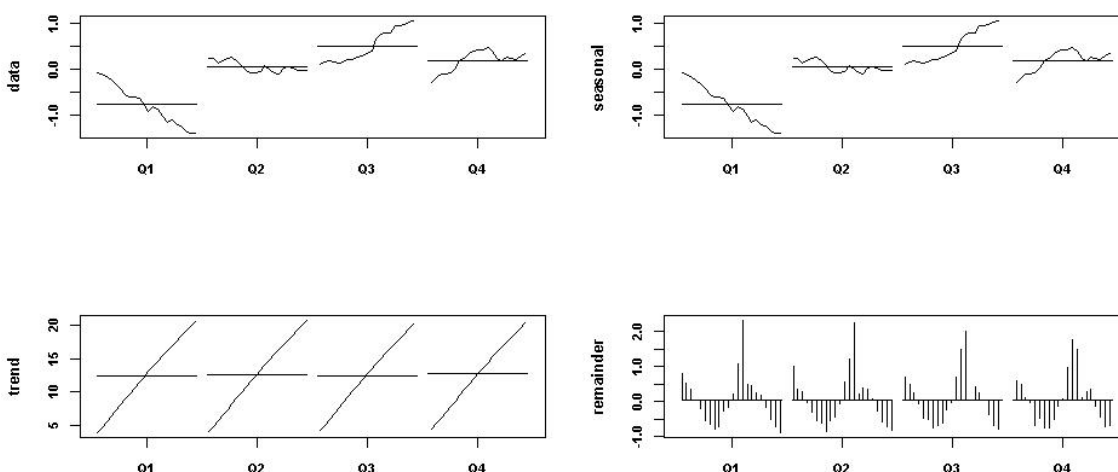


Zdroj: vlastní konstrukce

Příkazem `summary(decompY)` lze získat všechny souhrnné hodnoty odhadnutých složek časové řady Y_t . K jednotlivým složkám časové řady lze přistupovat pomocí vytvořené datové struktury `decompY`, v níž jsou tyto složky uloženy.

Programovací jazyk R umožňuje také analýzu časové řady po částech. Například lze analyzovat časovou řadu včetně dekompozice v rámci jednotlivých čtvrtletí. Toto přibližuje obrázek 7, kde je proveden rozbor chování časové řady pro jednotlivá čtvrtletí.

Obrázek č. 7: Modelování časové řady „HDP v b. c. v mld. EUR“ v jednotlivých čtvrtletích



Zdroj: vlastní konstrukce

Číselné charakteristiky a jednotlivé složky časové řady pro každé čtvrtletí, které jsou graficky reprezentovány na obrázku 7, lze získat pomocí vytvořeného datového objektu *decompY*.

K regresní analýze časových řad nabízí programovací jazyk R funkce *lsfit* především pro regrese metodou nejmenších čtverců a *lm* pro obecnější metody regrese. Větší možnosti pro regresní analýzu, zejména pokud jde o testování významnosti odhadovaných koeficientů, nabízí funkce *lm*. Proto potřebná regresní analýza pro účely analýzy časových řad by měla být pomocí příkazu *lm*.

3.4. Exponenciální vyrovnání a predikce časové řady

Přirozeným odhadem pro předpovídání další hodnoty dané časové řady Y_t v čase $t = \tau$ je vážená suma minulých pozorování, tj.

$$\hat{y}_{(t=\tau)}(1) = \lambda_0 \cdot y_\tau + \lambda_1 \cdot y_{\tau-1} + \dots, \quad (9)$$

kde $\hat{y}_{(t=\tau)}(1)$ značí hodnotu časové řady Y_t predikovanou v čase $t = \tau$, λ_i představují koeficienty váhy pro každé pozorování, $i = 1, \dots, n$.

Pro predikci pozorování časové řady se zdá rozumné, že nedávná pozorování mají větší vliv než pozorování v minulosti. Z tohoto důvodu je jednou z možností použití geometrických vah, které jsou dány tvarem:

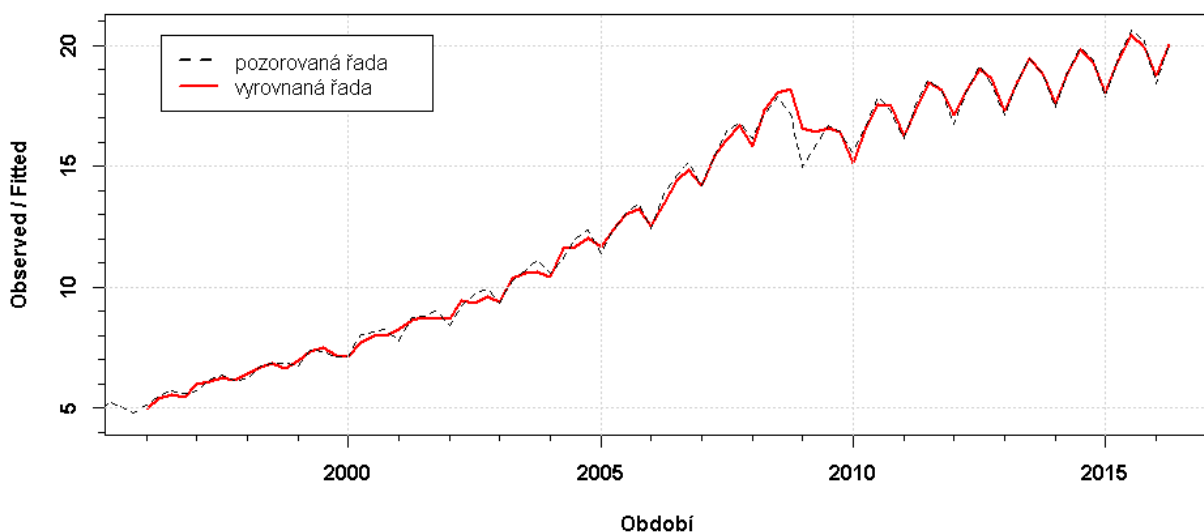
$$\lambda_i = \alpha(1 - \alpha)^i; \text{ pro } 0 < \alpha < 1, i = 1, \dots, n. \quad (10)$$

A po dosazení do vztahu (9) pro predikované pozorování dostaneme výraz

$$\hat{y}_{(t=\tau)}(1) = \alpha \cdot y_\tau + \alpha \cdot (1 - \alpha)^1 y_{\tau-1} + \alpha \cdot (1 - \alpha)^2 y_{\tau-2} + \dots, \quad (11)$$

Exponenciální vyrovnávání ve své základní formě (11) by mělo být použito pouze pro časové řady bez systematického trendu anebo sezónních komponent. Za účelem úspěšného použití exponenciálního vyrovnávání i pro časové řady obsahující jak trendovou složku, tak i složku sezónní, obsahuje programovací jazyk R zobecněnou verzi tohoto vyrovnávání, tzv. Holtovo-Wintersovo vyrovnávání. V tomto případě jsou tři parametry vyhlazování, a to α pro úroveň, β pro zachycení trendu a γ pro sezónní variace časových řad.

Obrázek č. 8: Exponenciální vyrovnávání časové řady „HDP v b. c. v mld. EUR“



Zdroj: vlastní konstrukce

3.5. Modelování časové řady pomocí Boxovy-Jenkinsovy metodologie

Boxova-Jenkinsova metodologie bere v úvahu při konstrukci modelu časové řady reziduální složku, která může být tvořena korelovanými (závislými) náhodnými veličinami. Boxova-Jenkinsova metodologie tedy nejen může zpracovávat časové řady s navzájem závislými pozorováními, ale dokonce těžiště jejich postupů spočívá právě ve zkoumání těchto závislostí neboli tzv. korelační analýze. Kombinují se autoregresivní modely $AR(p)$ (p je řád AR procesu) s modely klouzavých průměrů reziduální složky $MA(q)$ (q je řád MA procesu).

V případě nestacionární časové řady se provádí stacionarizace např. diferencováním a zjišťuje se řád diferencování s parametrem d . Výsledný model se potom označuje jako $ARIMA(p,d,q)$, v případě sezónních vlivů $SARIMA$ modely.

Informaci o síle lineární závislosti mezi veličinami y_t a y_{t-k} poskytuje autokorelační funkce definovaná výrazem

$$\rho(t, t-k) = \frac{E[(y_t - \mu_t)(y_{t-k} - \mu_{t-k})]}{\sqrt{E(y_t - \mu_t)} \cdot \sqrt{E(y_{t-k} - \mu_{t-k})}}, \quad (12)$$

kde y_t je pozorování časové řady v čase t , μ_t je střední hodnota časové řady Y_t .

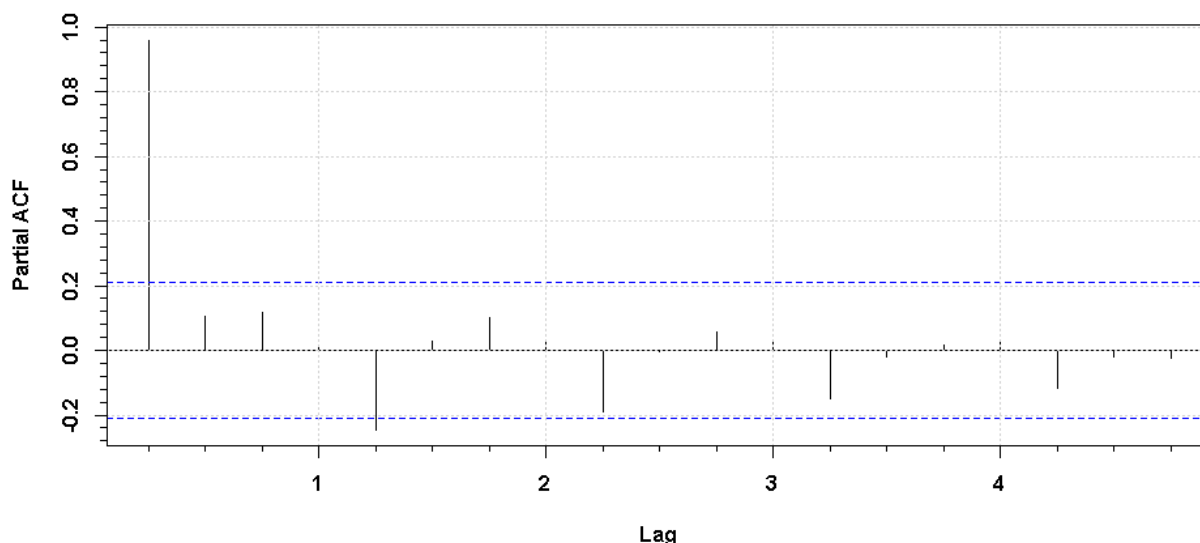
Informaci o síle lineární závislosti mezi veličinami y_t a y_{t-k} , která je očištěna o vliv veličin ležících mezi nimi, poskytuje výběrová parciální autokorelační funkce definovaná rekurentním vztahem [2]

$$f_{kk} = \frac{r_k - \sum_{j=1}^{k-1} f_{k-1,j} r_{k-j}}{1 - \sum_{j=1}^{k-1} f_{k-1,j} r_j}, \quad \text{kde } f_{kj} = f_{k-1,j} - f_{kk} f_{k-1,k-j}, \quad \text{a } j = 1, 2, \dots, k-1. \quad (13)$$

Výběrovou parciální autokorelační funkci časové řady HDP v běžných cenách v miliardách EUR lze v programovacím jazyku R získat pomocí příkazu `pacf(HDPts, main="HDP v b. c. v mld EUR")`

Výsledkem vykonání příkazu je graf výběrové parciální autokorelační funkce časové řady, z níž je patrná sezónní korelace mezi pozorování posunutými o čtyři čtvrtletí. Graf výběrové parciální autokorelační funkce časové řady je uveden na obrázku 9.

Obrázek č. 9: Výběrová parciální autokorelační funkce řady „HDP v b. c. v mld. EUR“



Zdroj: vlastní konstrukce

Identifikace modelu časové řady je provedena na základě výběrové parciální autokorelační funkce. Na základě [1] je vhodné zvolit jako model časové řady založený na Boxově-Jenkinsově metodologii ve tvaru $ARIMA(1, 0, 1)$, tedy řád AR procesu $p = 1$, řád MA procesu $q = 1$ a řád diferencování je $d = 0$. Ve skutečnosti se jedná o model $ARMA(1, 1)$, protože se diferencování neuplatňuje.

Odhady parametrů modelu časové řady v programovacím jazyku R se provádí pomocí příkazu *arima*, kde jako parametry příkazu vstupují modelovaná časová řada a parametry odhadovaného modelu, tj.

```
HDPts_arima<-arima(HDPts,order=c(1,0,1))
```

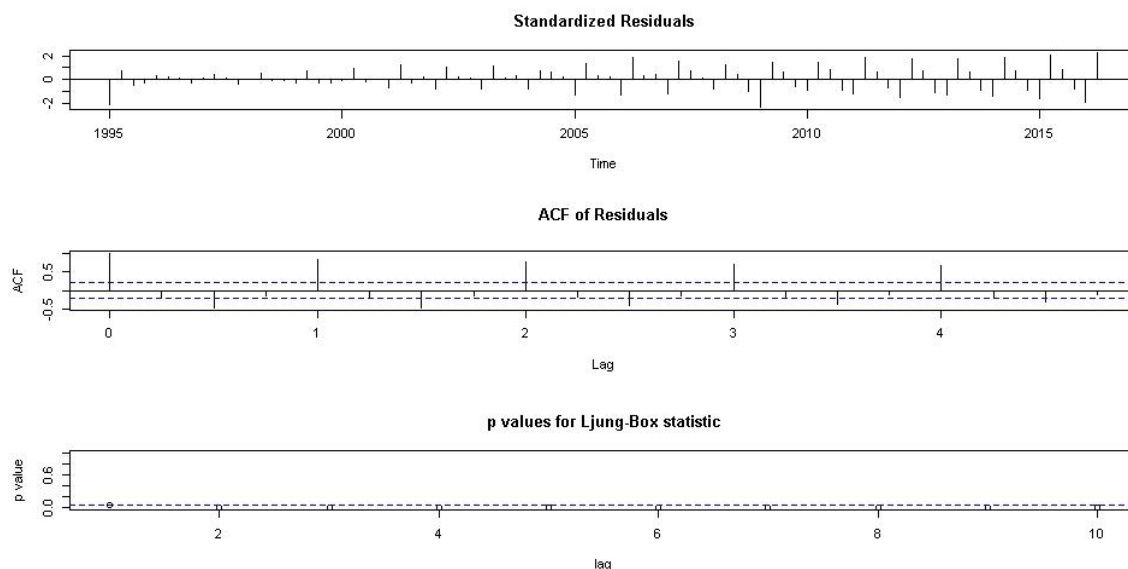
Po provedení odhadu parametrů navrženého modelu $ARMA(1, 1)$ vzniká datová struktura v prostředí programovacího jazyka R s odhadnutými koeficienty, reziduálními hodnotami i hodnotami k diagnostice modelu (například Akaikeho informační kritérium a další).

Následuje diagnostika odhadnutého modelu, která spočívá zejména v analýze reziduí. K tomu má programovací jazyk R funkci *tsdiag*, která zajistí celkovou diagnostiku navrženého modelu, tj.

```
tsdiag(HDPts_arima)
```

Výsledky diagnostiky modelu $ARMA(1, 1)$ navrženého jako modelu pro časovou řadu HDP v běžných cenách v mld. EUR jsou naznačeny na obrázku 10. I model $ARMA(1, 1)$ potvrdil sezónní korelaci mezi pozorováními posunutými o čtyři čtvrtletí.

Obrázek č. 10: Výsledky testování modelu $ARMA(1, 1)$ řady „HDP v b. c. v mld. EUR“



Zdroj: vlastní konstrukce

Jakmile je model identifikován a byly odhadnuty jeho parametry, je dalším cílem předpovědět budoucí hodnoty časové řady. Budoucí hodnoty časové řady (predikce) umožňuje programovací jazyk R modelovat z odhadnutého modelu pomocí funkce

```
HDP.pred<-predict(HDPts_arma,n.ahead=20).
```

Uvedený příkaz dovoluje na základě odhadnutého ARMA(1, 1) modelu časové řady HDP v běžných cenách v mld. EUR predikovat následujících 20 pozorování modelované časové řady.

3.6. Spektrální analýza časové řady

Časová řada (po odstranění trendové složky) může být při spektrální analýze reprezentována jako součet sinusovek a kosinusovek od nižších po vyšší frekvenci o různých amplitudách.

Frekvence ω časové řady je definována jako počet pozorování během 1 cyklu. Časová řada, kdy jsou jednotlivé hodnoty pozorovány měsíčně, je frekvence $\omega = 12$, u čtvrtletních pozorování je frekvence $\omega = 4$. Obecně nemusí být rozložení frekvencí u časové řady známé. Relativní hustotu každé frekvence ve spektru odvozené z analyzované časové řady lze odhadovat periodogramem $I(\omega)$. Periodogram $I(\omega)$ je definován ve tvaru [2]

$$I(\omega) = \frac{1}{2\pi n} \left[\left(\sum_{t=1}^n y_t \cos(\omega t) \right)^2 + \left(\sum_{t=1}^n y_t \sin(\omega t) \right)^2 \right], \quad (14)$$

kde n je počet pozorování časové řady, y_t jsou pozorované hodnoty časové řady v čase t a ω reprezentuje frekvenci. Periodogram umožňuje identifikovat hodnoty frekvencí ω , které jsou v analyzované řadě významné.

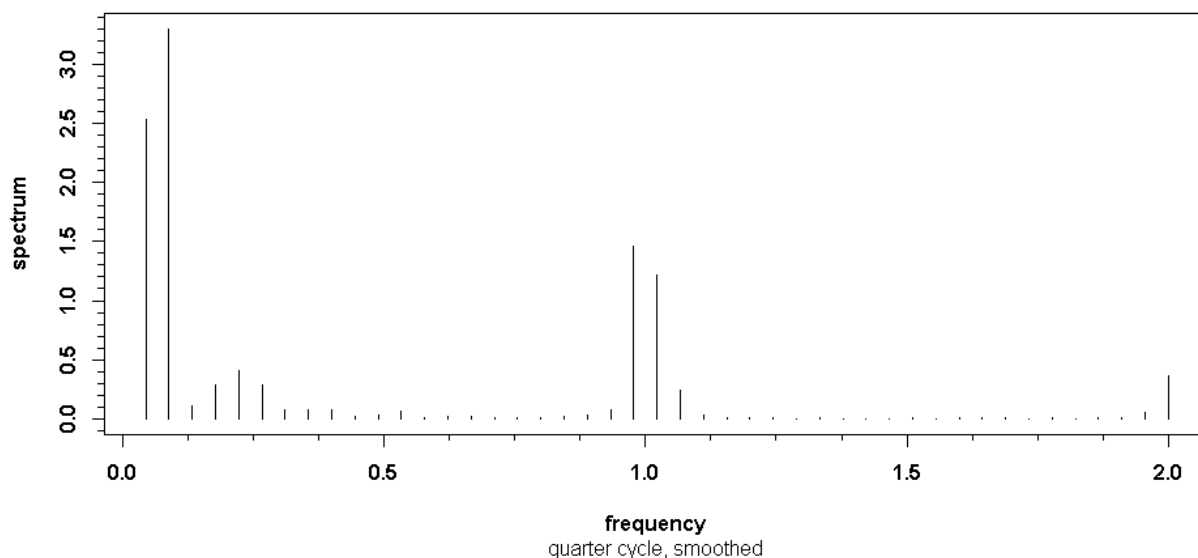
V případě čtvrtletní časové řady HDP v běžných cenách v mld. EUR je možné předpokládat, že frekvence s nejvyšší amplitudou bude 4.

Periodogram analyzované časové řady se v prostředí programovacího jazyka R vytvoří pomocí příkazu

```
spectrum(HDPts, log = "no", main = "Periodogram řady HDP v b. c. v mld EUR", sub = "quarter cycle, smoothed", type = "h").
```

Odhadnutý periodogram vybrané sledované řady je znázorněn na obrázku 11.

Obrázek č. 11: Periodogram časové řady „HDP v b. c. v mld. EUR“



Zdroj: vlastní konstrukce

4. SHRNUÍ A ZÁVĚR

Základní funkcionální programového prostředí R je dodávána s množstvím funkcí užitečných pro časové řady, zejména v doplňkovém programovém balíčku (add-on package) statistiky. Toto je doplněno dalšími doplňkovými programovými balíčky (add-on packages), které jsou vybaveny dalšími funkcionalitami pro oblast analýzy časových řad. Jedná se jak o klasické metody, tak i o velmi pokročilé techniky [5]. Doplňkové programové balíčky pro analytické zkoumání a modelování časových řad neobsahují pouze funkcionality pro samotné časové řady, ale i funkcionality z hlediska postavení časových řad v ekonometrii a financích a do značné míry se svými nástroji pro časové řady překrývají. Z tohoto pohledu lze zhruba strukturovat doplňkové programové balíčky do následujících témat, které se dotýkají problematiky časových řad:

- Základní funkcionality pro manipulaci s napozorovanými údaji ve tvaru časových řad představují zejména doplňkové programové balíčky *Time Series Analysis* [5].
- Pokročilé funkcionality programového prostředí jazyka R pro analytické metody a předvídání časových řad nabízejí doplňkové programové balíčky *Econometrics* [3].
- Metody a techniky analýzy časových řad z oblasti finanční a pojistné matematiky nabízí skupina doplňkových programových balíčků *Empirical Finance* [4].

Článek nabízí – aspoň pro některé doplňkové programové balíčky, které jsou nejdůležitější z pohledu používaných metod a technik na poli časových řad – stručný přehled o možnostech programu R a předložil pár jednoduchých a názorných příkladů praxe v používání programu R. Cílem článku bylo především upozornit na obrovský potenciál a velmi vhodné možnosti používání programu R při modelování časových řad.

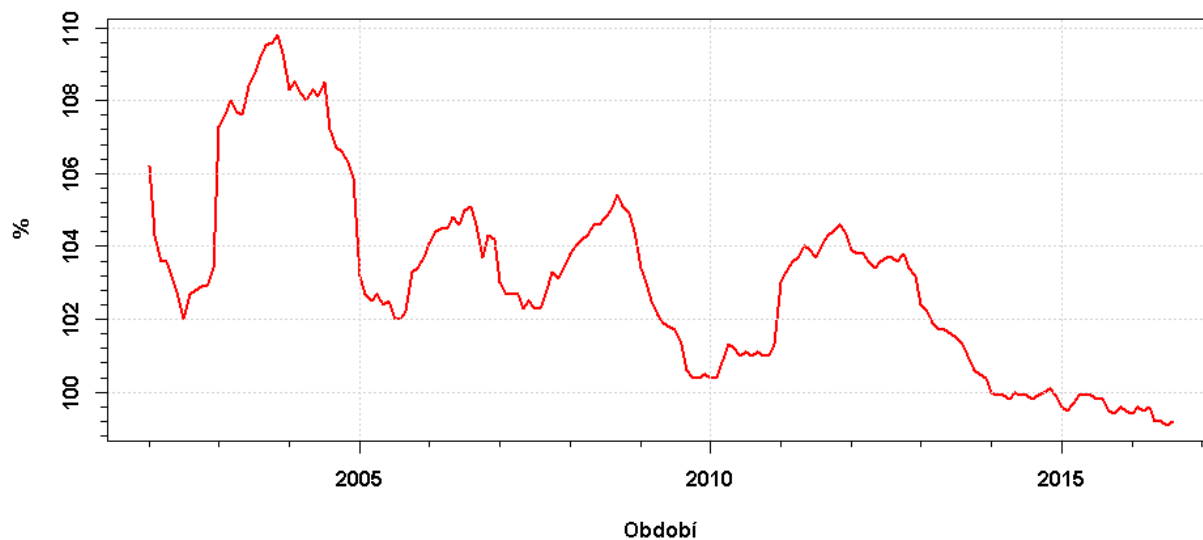
Využití specializovaných programových balíčků v rámci programovacího prostředí R na metody analýzy a prognózování časových řad dovoluje zvládat problémy, které jsou řešitelné ostatními komerčními statistickými programy. Na rozdíl od jiných specializovaných statistických programů lze programovací prostředí R získat a používat zcela zadarmo. Neustálý vývoj nových programových balíčků, na jejichž vzniku se zpravidla podílejí významní odborníci i zkušení praktici v této oblasti, zajišťuje možnost přístupu k moderním metodám a technikám analýzy časových řad a jejich předpovídání celé komunitě uživatelů programu R.

Výše provedený přehled metod a technik, které lze uplatnit na poli modelování časových řad, představuje pouhý výsek možností, které poskytuje programové prostředí programu R. Využitím tohoto programového vybavení, a to zejména doplňkových programových balíčků, lze získat velmi efektivní a sofistikované nástroje (a to zcela zdarma) k řešení i těch nejsložitějších odborných problémů a otázek v oblasti časových řad. Jedná se především o odhad parametrů klasického modelu časové řady, dekompozici časové řady, otázky Boxovy-Jenkinsovy metodologie, nestacionaritu časových řad, regresní metody i strukturální modely a další odborné otázky v této oblasti. Vícerozměrné časové řady, jejich diagnostika, spektrální analýza, modely GARCH a další lze poměrně jednoduchým způsobem řešit pomocí zabudovaných funkcionalit programového prostředí jazyka R.

V příloze jsou uvedeny ukázky analýzy ostatních vybraných ekonomických časových řad z databáze SlovStat.

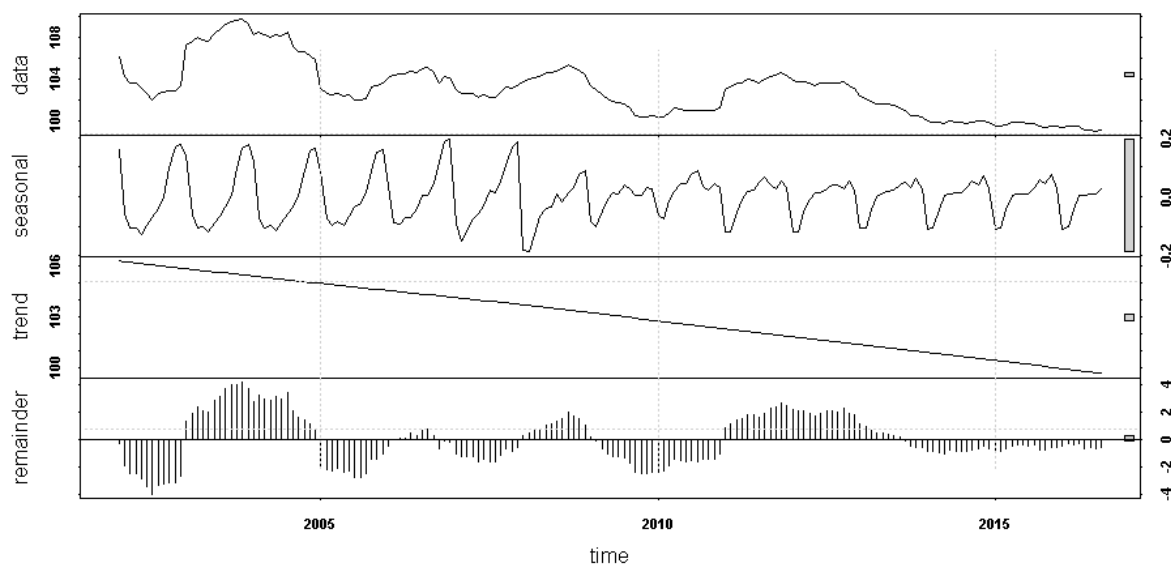
PŘÍLOHA

Obrázek č. 1: Vizuální průběh řady „Indexy spotřebitelských cen podle COICOP“



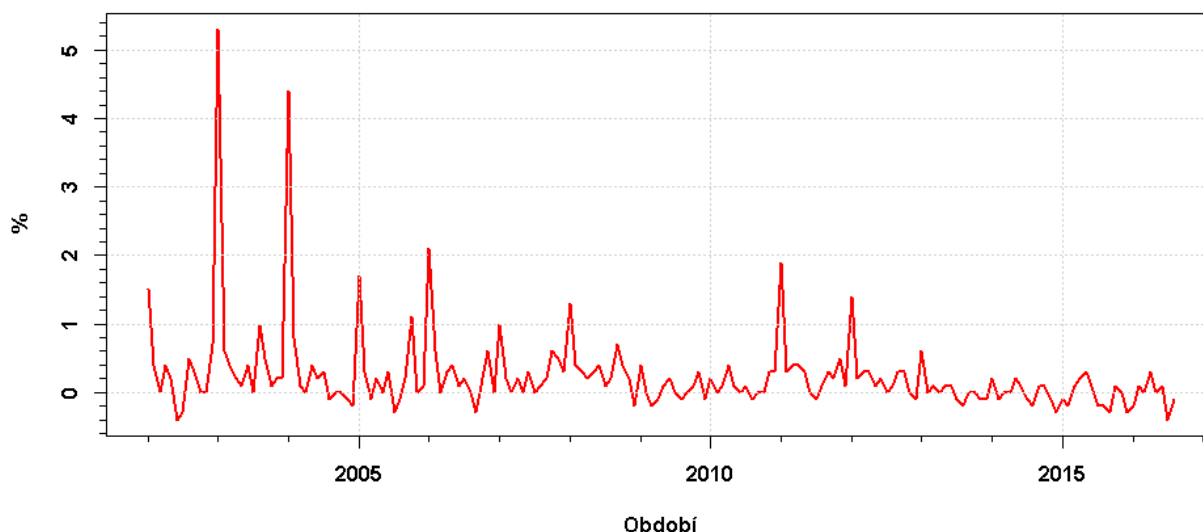
Zdroj: vlastní konstrukce

Obrázek č. 2: Modelování časové řady „Indexy spotřebitelských cen podle COICOP“



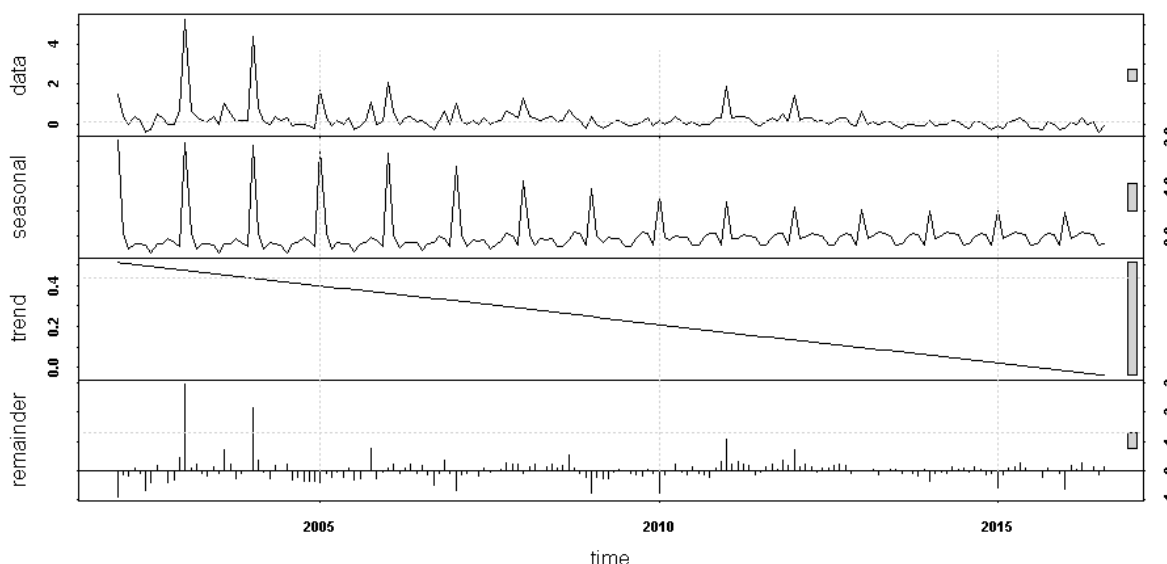
Zdroj: vlastní konstrukce

Obrázek č. 3: Vizuální průběh časové řady „Meziměsíční inflace v %“



Zdroj: vlastní konstrukce

Obrázek č. 4: Modelování časové řady „Meziměsíční inflace v %“



Zdroj: vlastní konstrukce

LITERATURA

- [1] ARLT, J. – ARLTOVÁ, M. – RUBLÍKOVÁ, E.: Analýza ekonomických časových řad s příklady. VŠE Praha, 2004. 148 s. ISBN 80-245-0777-3.
- [2] BROCKWELL, P., J. – DAVIS, R., A.: Introduction to Time Series and Forecasting. Second Edition. New York: Springer-Verlag New York, Inc., 2002. 450 p. ISBN 0-387-95351-5.
- [3] CRAN Task View: Econometrics [on-line]. 23. 04. 2016 [cit. 09. 05. 2016]. WWW: <<https://cran.rstudio.com/web/views/Econometrics.html>>.
- [4] CRAN Task View: Empirical Finance [on-line]. 14. 04. 2016 [cit. 09. 05. 2016]. WWW: <<https://cran.rstudio.com/web/views/Finance.html>>.
- [5] CRAN Task View: Time Series Analysis [online]. 22. 01. 2016 [cit. 09.05.2016]. WWW: <<https://cran.rstudio.com/web/views/TimeSeries.html>>.
- [6] CYHELSKÝ, L. – SOUČEK, E.: Základy statistiky. Praha: VŠFS, o. p. s., 2009. 164 s. ISBN 978-80-7408-013-5.

- [7] KLEIBER, CH. – ZEILEIS, A.: Applied Econometrics with R. New York: Springer Science+BusinessMedia, LLC, 2008. 222 p. ISBN: 978-0-387-77318-6 (eBook).
- [8] SlovSTAT – HDP na obyvateľa v stálych cenách prechádzajúceho roka (s. c. p. r.) (tis. EUR) – metodika ESA 2010, podľa štvrtročných NÚ. WWW: <http://www.statistics.sk/pls/elisw/objekt.send?uic=1854&m_sso=3&m_so=81&ic=188>.
- [9] SlovSTAT – Inflácia meraná CPI (národná metodika) – celková, jadrová, čistá (2002M01 - 2016M03). WWW: <http://www.statistics.sk/pls/elisw/objekt.send?uic=1486&m_sso=3&m_so=16&ic=46>.
- [10] SlovSTAT – Inflácia meraná CPI (národná metodika) – celková, jadrová, čistá (2002M01 - 2016M03). WWW: <http://www.statistics.sk/pls/elisw/objekt.send?uic=1486&m_sso=3&m_so=16&ic=46>.

Ing. Roman Pavelka, PhD., pracuje v odbore metód štatistických zisťovaní Štatistického úradu SR.

Informácia/Information

I. SVETOVÉ FÓRUM O DÁTACH PRINIESLO VEĽA INŠPIRÁCIE AJ PRE ŠTÁTNU ŠTATISTIKU

THE FIRST WORLD DATA FORUM HAS BROUGHT MANY INSPIRATION ALSO FOR THE OFFICIAL STATISTICS

V dňoch 15. – 18. januára 2017 sa v juhoafrickom Kapskom Meste uskutočnilo I. svetové fórum o dátach.¹ Usporiadal ho Štatistický úrad Juhoafrickej republiky pod záštitou štatistickej divízie Hospodárskej a sociálnej rady Organizácie Spojených národov (OSN), Štatistická komisia OSN a Vysoká skupina pre partnerstvo, koordináciu a budovanie kapacít v štatistike pre agendu trvalo udržateľného rozvoja 2030. I. svetové fórum o dátach bolo zamerané na zavedenie systému zberu a analýz údajov potrebných na dosiahnutie strategických cieľov a ukazovateľov na podporu trvalo udržateľného rozvoja. Cieľom podujatia, na ktorom sa zúčastnilo približne 1 300 delegátov z viac ako 100 krajín sveta, bolo zintenzívnenie spolupráce odborných skupín z verejného a súkromného sektora, zástupcov vedeckej a akademickej obce a používateľov údajov.

Na podujatí boli prezentované inovatívne spôsoby využitia údajov a štatistiky na meranie svetového progresu a na kvalifikované rozhodovanie v rámci Agendy 2030. K najdiskutovanejším témam patrili: budovanie štatistických kapacít, štatistická gramotnosť, synergia medzi tradičnou štatistikou a novými zdrojmi údajov vrátane tzv. Big Data, inovatívne technológie produkcie dát, mobilizácia potrebných zdrojov, dôverynosť a ochrana údajov, normy a systémy geopriestorových informácií.

V rámci témy Big Data boli predstavené inovatívne projekty a nástroje, ktoré naznačili možnosti potenciálneho využitia tohto typu údajov. Ide napríklad o analýzu obsahu rozhlasového vysielania v reálnom čase, odhady chudoby založené na údajoch zo satelitov, systém manažmentu údajov na vizualizáciu odpadového hospodárstva v utečeneckých táboroch či zber údajov z rozličných zdrojov (sociálne médiá, občianske hlásenia...) v prípade prírodných katastrof. Prezentácie obsahovali aj reálne príklady humanitárnej a krízovej pomoci, očkovacie kampane a cenzy, ako aj analýzy s použitím schém a agregovaných údajov v časopriestore. Účastníkov fóra zaujali aplikácie, ktoré kombinujú satelitné snímky s vysokým rozlíšením a senzorické údaje s multitematickými zisťovaniami v domácnostiach, respektíve so zisťovaniami fariem. Tieto aplikácie umožňujú zlepšiť meranie rozvoja poľnohospodárstva v rurálnom priestore a dopravy v mestskom prostredí. Poskytujú možnosť vytvárania máp s vysokým rozlíšením, ktoré prinášajú napr. nový spôsob chápania chudoby prostredníctvom jej priestorového rozloženia alebo nový pohľad na úrodnosť pôdy. Z príspevkov témy Big Data ďalej vyplynulo, že kombináciou rôznych zdrojov a použitím nových matematických modelov bude možné zaplniť biele miesta na mape a napomôcť prijatie kvalifikovaných a transparentných rozhodnutí.

¹ Podrobnosti pozri na: <http://undataforum.org/> (prístup k 13. 3. 2017).

Agenda 2030 obsahuje súbor cieľov a ukazovateľov, ktoré majú v horizonte 15 rokov pomôcť jednotlivým štátom monitorovať dosiahnutý pokrok v oblasti trvalo udržateľného rozvoja. Všetky majú vo svojej podstate štatistický a geografický charakter. Riešenie mnohých problémov, ktoré brzdia trvalo udržateľný rozvoj, si vyžaduje integrované zdroje údajov a informácií vhodných na analýzu, modelovanie a vytváranie máp v priestore a čase. I. svetové fórum o dátach poskytlo národným informačným systémom aj praktické ukážky využitia geopriestorových a štatistických informácií ako podkladu na efektívne rozhodovanie v oblasti trvalo udržateľného rozvoja. Nezávislá poradná skupina expertov (IAEG-SDG) zhrnula obsah prednášok a diskusií a navrhla možné spôsoby implementácie ukazovateľov trvalo udržateľného rozvoja na svetovej i národnej úrovni. Fórum bolo tiež príležitosťou získať politickú podporu poskytovaniu údajov k Agende 2030.

Medzi účastníkmi fóra zarezovala myšlienka *Nezabúdajme na nikoho* (Leave No-one Behind – LNB), ktorá obsahuje nasledujúce posolstvá: ciele nie sú splnené, ak nie sú splnené pre všetkých, údaje budú členené minimálne na kvintily (20 % populácie) podľa pohlavia, veku a zdravotného obmedzenia, budú dostupné a kombinované tak, aby priniesli prospech pri prijímaní rozhodnutí. Výstupom akceptácie spomínanej myšlienky (LNB) by mala byť povinnosť členiť údaje v záujme hľadania riešení pre mnohopočetné a navzájom sa prekrývajúce skupiny obyvateľstva a lepšie pochopenie toho, že údaje neslúžia len na meranie, ale sú prostriedkom na dosiahnutie cieľov.

Tému využívania administratívnych zdrojov údajov prezentovali v Kapskom Meste zástupcovia štatistických úradov Nórska, Dánska, Švédska a Fínska, krajín s vysoko rozvinutým využívaním administratívnych zdrojov. Vo svojich vystúpeniach predstavili základné postupy pri práci s registrami v štatistike, venovali sa otázkam kvality a ich prezentácie obsahovali aj praktické ukážky integrovaného štatistického systému.

Fórum vytvorilo priestor aj na diskusiu o dôsledkoch zmien v dostupných údajoch a náraste algoritmických analýz na meranie, monitorovanie a propagovanie sociálneho rozvoja. Diskusia zahŕňala viaceré okruhy otázok, napríklad meranie zdravia a pokroku v spoločnostiach v nasledujúcich 10 – 15 rokoch a neskôr, zostavovanie a účel starých a nových ukazovateľov (sociálna kohézia a dôvera, odolnosť spoločnosti atď.), použitie inštitucionálnych a právnych ustanovení v kontexte súkromných spoločností, ktoré vlastnia väčšinu osobných údajov, a pod.

Z pohľadu trendov prezentovaných na I. svetovom fóre o dátach sa Štatistický úrad SR uberá správnou cestou. Už pri najbližšom sčítaní obyvateľov, domov a bytov v roku 2021 využije údaje z vybraných registrov a administratívnych zdrojov a pri výpočte harmonizovaného indexu spotrebiteľských cien (HICP) sa pripravuje na využívanie transakčných dát od obchodných reťazcov (Scanner Data). Využívanie dát z administratívnych a ďalších externých zdrojov umožní na jednej strane znižovať záťaž respondentov a na druhej strane prinesie do štatistickej produkcie vyššiu kvalitu, efektívnosť a znižovanie nákladov.

Ďalší rozvoj informačných technológií úradu sa nezaobíde bez budovania kapacít na integráciu Big Data a geopriestorových informácií. Aktuálne vízie naznačujú, že už v blízkej budúcnosti vstúpia do produkcie údajov oficiálnej štatistiky nové rôzne zdroje (telekomunikačný a energetický sektor, satelitné snímky, elektronické médiá,

sociálne siete...), ktoré sa ešte dnes javia ako nekompatibilné a s najväčšou pravdepodobnosťou budú vo vlastníctve nadnárodných spoločností. Úlohou štatistických úradov bude získať prístup k týmto dátam a vytvoriť prostredie na ich bezpečné a efektívne spracovanie. Výsledkom štatistických procesov už nebude len produkcia štandardných štatistických ukazovateľov, ale najmä ponuka multidimenzionálnych produktov a služieb s vysokou pridanou hodnotou. Vzhľadom na vývoj technológií a používateľských preferencií bude tiež potrebné štatistické výstupy prezentovať inovatívnymi spôsobmi a interaktívnou formou. To sú všetky výzvy, na ktoré budú musieť reagovať národné štatistické úrady, ale aj Európsky štatistický systém.

Ing. ALEXANDER BALLEK

Autor je predsedom Štatistického úradu Slovenskej republiky.

SČÍTANIE SI ZASLÚŽI BYŤ NÁRODNÝM PROJEKTOM

CENSUS DESERVES TO BE A NATIONAL PROJECT



PhDr. Ľudmila Ivančíková, PhD.

Štatistický úrad SR zintenzívňuje prípravu sčítania obyvateľov, domov a bytov v roku 2021. Dôvodom je viac. Sčítanie je vyčerpávajúce, na Slovensku obsahovo najrozsiahlejšie, a tým aj najnáročnejšie štatistické zisťovanie. Najbližšie sčítanie bude opäť medzinárodne koordinované Európskou úniou a Organizáciou Spojených národov a navyše na Slovensku ho sprevádzajú viaceré novinky.

Práve o nich sme sa rozprávali s generálnou riaditeľkou sekcie sociálnych štatistík a demografie Štatistického úradu SR Ľudmilou Ivančíkovou.

- **Štatistický úrad SR oznámil, že v roku 2021 uskutoční na Slovensku prvýkrát v histórii integrované sčítanie. Čo konkrétne si majú pod tým naši čitatelia predstaviť? Čím bude integrované sčítanie iné ako doterajšie cenzy?**

Integrované sčítanie je sčítanie založené na kombinácii údajov z registrov, administratívnych zdrojov a údajov získaných priamo od obyvateľov. Integrovaným sčítaním sa na Slovensku skončí takmer 200-ročné obdobie realizácie tradičných cenzov.

Trend prechodu od tradičného cenzu ku kombinovanému sčítaniu alebo k sčítaniu postavenému výlučne na registroch je celosvetovým trendom, ktorý možno pozorovať už v cenzoch okolo roku 2011. Tradičným spôsobom nateraz posledné sčítanie realizovala menej ako polovica z 28 krajín Európskej únie (13). Zvyšné štáty použili buď metódu kombinácie zdrojov (kombinovali údaje z registrov a z vyčerpávajúceho zisťovania, resp. z registrov a existujúceho výberového zisťovania), alebo uskutočnili sčítanie výlučne z údajov získaných z registrov.

- **Prečo sa integrované sčítanie javí ako najlepší cenzový model pre rok 2021?**

Modernizácia preniká aj do takej konzervatívnej oblasti, akou je štatistika. Pribúdajú nové nástroje, nové metódy i nové zdroje, ktorými požadované údaje dokážeme získať napríklad rýchlejšie, kvalitnejšie, a čo je rovnako dôležité, aj bez priamej účasti obyvateľov a pri maximálnom využití už existujúcich zdrojov.

Integrované sčítanie spája pozitíva tradičného sčítania a sčítania založeného výlučne na registroch. Očakávame, že okrem efektívnosti nový spôsob sčítania zníži záťaž obyvateľov a zvýši kvalitu údajov a informácií, najmä vzhľadom na rastúcu mieru neodpovedí obyvateľov pri tradičnom sčítaní.

Zvažovali sme, samozrejme, aj náklady, keďže tradičné sčítania sú stále najdrahšie. Pod rozhodnutie uskutočniť na Slovensku v roku 2021 integrované

sčítanie sa podpísala najmä postupujúca modernizácia služieb, rozvoj nových informačných technológií a digitalizácia našej spoločnosti.

- **V cenze 2021 budú teda obyvatelia síce dôležitým, ale už len jedným z viacerých zdrojov údajov a informácií. Boli vypočítané hlasy tých, ktorí v roku 2011 vyčítali Štatistickému úradu SR, že duplicitne žiada od obyvateľov aj údaje nachádzajúce sa v rôznych administratívnych zdrojoch a registroch, alebo zavážili iné relevantné okolnosti?**

Ako som už spomenula, konceptom integrovaného sčítania sleduje Štatistický úrad SR dva ciele – zníženie záťaže obyvateľov a udržanie kvality údajov. Pod rozhodnutie sa okrem vami spomenutej skutočnosti podpísal aj stav registrov na Slovensku, ktorý je dnes z hľadiska potrieb sčítania vhodnejší ako pred desiatimi rokmi. Informatizácia verejnej správy priebežným „čistením“ registrov zvýšila nielen ich kvalitu, ale aj možnosti integrácie údajov, ktoré sa v registroch nachádzajú.

Treba si uvedomiť, že časť údajov z registrov sa využije ako zdroj, časť na účely štatistického spracovania a časť na účely porovnávania a koherencie, t. j. na hodnotenie kvality údajov. Preto plánujeme využiť všetky v súčasnosti dostupné a zmapované administratívne i sekundárne zdroje, napr. informačné systémy ministerstiev.

Netreba však zabúdať ani na národné požiadavky spojené s cenom. Spomeniem údaje o národnostiach, ktoré sú v našich podmienkach jedinečné z hľadiska časových radov a aktuálneho používania, alebo údaje o náboženskom vyznaní. Neexistuje žiadny iný zdroj obsahujúci tieto premenné. Dajú sa získať len a len od obyvateľov, ktorí v sčítaní sami deklarujú príslušnosť k národnosti a svoje náboženské vyznanie. Aj v tomto prípade je však hľadisko zníženia záťaže obyvateľov prioritou. Zaoberáme sa preto myšlienkou uskutočniť v roku 2021 elektronické multikanálové sčítanie, aby sme obyvateľom poskytovanie údajov čo najviac uľahčili. Na druhej strane očakávame od obyvateľov pochopenie a spoluprácu pri realizácii zisťovania, ktorým sa zabezpečujú úlohy celospoločenského významu.

- **Čo ukázala vaša analýza, sú údaje a informácie vo vytypovaných registroch a administratívnych zdrojoch využiteľné na štatistické účely takpovediac okamžite?**

Kvalita registrov je výrazne lepšia ako v roku 2011. Do roku 2021, ktorý je rokom sčítania, by jednotlivé registre mali byť taktiež plne integrované. To umožní ich využitie na účely sčítania v maximálne možnej miere.

Na druhej strane registre sú budované na špecifické, nie štatistické účely a už vôbec nie na účely sčítania obyvateľov, domov a bytov. Ich využitie si preto okrem analýzy vyžaduje predspracovanie údajov, následne ich spracovanie, vytvorenie integrovanej databázy a až potom integráciu s údajmi z elektronického zberu.

- **Ambíciou Štatistického úradu SR je urobiť zo sčítania 2021 celonárodný projekt. Prostriedkom na to má byť Národný akčný plán Sčítania obyvateľov, domov a bytov 2021 na roky 2017 – 2020. Prečo ste sa vybrali práve touto cestou?**

Nová koncepcia sčítania na Slovensku si vyžaduje sústrediť sa v oveľa väčšej miere ako pri tradičnom cenze na prípravu zisťovania. Aj preto národný akčný plán nepokrýva celý projekt sčítania a jeho komplexnú prípravu, ale sústreďuje sa na nové oblasti a partnerov z externého prostredia. Celkovo vymedzuje osem strategických cieľov zameraných okrem už spomínanej problematiky administratívnych zdrojov a registrov napr. na oblasť harmonizácie údajov o počte obyvateľov, ďalej na územnú prípravu, elektronické sčítanie a špecifické prístupy k sčítaniu obyvateľov, domov a bytov, napr. vo veľkých mestách, alebo k sčítaniu bezdomovcov.

Očakávame, že plynulá diskusia spojená s neustálym informovaním o sčítaní a zainteresovaním všetkých zodpovedných orgánov a organizácií, ale aj verejnosti, obyvateľov respondentov, sa môže pozitívne prejavíť na výsledku komunikácie, a teda aj na výsledku samotného sčítania.

- **Aké poslanstvo chce štatistický úrad komunikovať prostredníctvom národného akčného plánu sčítania a kto predovšetkým by mal byť jeho adresátom?**

Treba si uvedomiť, že sčítanie je celospoločenskou akciou. Aby Štatistický úrad SR splnil svoju úlohu pri jeho príprave a realizácii, je potrebné sústrediť sa najmä na orgány a organizácie, kde predpokladáme najintenzívnejšiu spoluprácu; spomeniem len správcov jednotlivých administratívnych zdrojov, obce a mestá. Bez ich zapojenia a otvorenej spolupráce nebude možné na Slovensku realizovať moderné sčítanie ani teraz, ani v budúcnosti.

- **Ako konkrétne môže úspešné integrované sčítanie ovplyvniť rozvoj demografickej štatistiky a sociálnych štatistík na Slovensku po roku 2021?**

Čoraz viac sa hovorí o tom, že sčítanie by nemalo byť osamotenou akciou a je potrebné vychádzať z jeho napojenia na demografickú štatistiku. Európsky štatistický systém už v súčasnosti pracuje na koncepte tzv. postcenzu 2021.

Hovorí sa o zmenách v oblasti včasnosti poskytovaných údajov (dáta by mali byť k dispozícii do 12 mesiacov od zberu), frekvencie ich poskytovania (skrátene 10-ročnej periodicity) a pokrytia (koncept obvyklého pobytu a poskytovanie časti údajov na detailnejšej úrovni, ako sú obce).

Cenzus má teda svoju budúcnosť, aj keď už nemusí byť sčítaním, ako ho poznáme doteraz. Myslím si, že údaje zo sčítania sú nevyhnutné pre súčasné generácie. Nezabúdajme však, že sú dôležité aj pre ďalšie generácie, ktoré nás budú poznať a hodnotiť aj cez prizmu výsledkov sčítania obyvateľov, domov a bytov. Viete, ako sa napríklad realizoval vo Veľkej Británii televízny projekt s názvom *Who do you think you are?*¹ Práve s použitím výstupov z minulých cenzov.

Za rozhovor ďakuje ZUZANA ŠTUKOVSKÁ

¹ Pozri na <http://yesterday.uktv.co.uk/shows/who-do-you-think-you-are/> (prístup k 23. 2. 2017).

Informácia/Information

BIELA KNIHA O BÝVANÍ V BRITSKOM PARLAMENTE**WHITE PAPER ON HOUSING IN THE BRITISH PARLIAMENT**

Na začiatku februára 2017 predstavila vláda Veľkej Británie v parlamente vyše 100-stránkový dokument ***Fixing our broken housing market***¹, ktorý mapuje aktuálnu situáciu s bývaním v krajine a prichádza s návrhmi na zlepšenie terajšieho stavu.

Vládna správa poukazuje na relatívne zlú situáciu v uspokojovaní potrieb bývania obyvateľov Veľkej Británie. Ide o dôsledok klesajúcej a zároveň najmenej miery prírastku bytov na obyvateľa za posledné tri dekády v rámci krajín západnej Európy. Riešenie tejto nepriaznivej situácie je v okamžitom začatí výstavby väčšieho počtu nehnuteľností určených na bývanie na správnom mieste.

Vláda Veľkej Británie chce vyvinúť tlak na miestne orgány a staviteľov, aby zrýchlili proces výstavby nehnuteľností, napríklad aj prostredníctvom povinných objednávok a skrátením času z troch na dva roky medzi vydaním stavebného povolenia a začatím výstavby.



Britská premiérka Theresa Mayová v úvodnom príhovore vládnej správy pripúšťa, že tak kúpa, ako aj prenájom bývania sú v krajine pre mnohých ľudí nedostupné. Dnes je potrebné na priemerné bývanie vo Veľkej Británii vynaložiť takmer osemročný priemerný zárobok, čo je najviac v doterajšej histórii. Zhoršujúca sa

¹ Materiál je k dispozícii na:

https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/590463/Fixing_our_broken_housing_market_-_accessible_version.pdf.

dostupnosť bývania spôsobila napr. zdvojnásobenie počtu nájomného bývania od roku 2000. Viac ako 2,2 milióna britských domácností s podpriemerným príjmom spotrebujú viac ako tretinu ich disponibilného príjmu na výdavky na bývanie.

Britská vláda je odhodlaná pôsobiť na súčasný trh s bývaním tak, aby sa bývanie stalo cenovo dostupnejšie a ľudia cítili istotu pri plánovaní budúcnosti. Panuje všeobecná zhoda v tom, že terajší stav trhu s bývaním je jednou z najväčších prekážok pokroku v dnešnej Británii. Jediným riešením je komplexným prístupom k súčasnému systému bývania zabezpečiť riešenia všetkých jeho slabých stránok. Vyžiada si to realizáciu viacerých opatrení.

Predovšetkým **je potrebné mať k dispozícii viac pozemkov** tam, kde ľudia chcú žiť. Predpokladá to získať spoľahlivé informácie o existujúcich nehnuteľnostiach v jednotlivých regiónoch a prehľad o požiadavkách potenciálnych záujemcov o bývanie v daných lokalitách. Tieto úlohy by mali spoľahlivo plniť predovšetkým jednotlivé miestne orgány.

Druhou oblasťou je **zabezpečenie rýchlej výstavby nehnuteľností po získaní stavebných povolení**. V tomto smere bude potrebné dobre zainvestovať do vytvorenia otvorenejších a prístupnejších plánovacích systémov, zlepšiť koordináciu verejných investícií do infraštruktúry, podporovať včasné pripojenie na inžinierske siete a operatívne riešiť problémy, ktoré by mohli výstavbu nehnuteľností zbytočne predĺžiť. Tieto nástroje by mali pomôcť miestnym orgánom a staviteľom zrýchliť samotnú výstavbu. Zvyšujúca sa ponuka by mala následne znížiť tlak na rast cien za bývanie i prenájom.

Tretím zámerom je **diverzifikovanie trhu s bývaním** tak, aby sa viac otvoril aj menším staviteľom a tým, ktorí prichádzajú s inovatívnymi a efektívnymi metódami výstavby. Preto by sa mali vo Veľkej Británii výraznejšie podporovať bytové družstvá a miestne orgány, aby viac stavali, a zároveň motivovať noví investori, aby sa oživila bytová výstavba vrátane výstavby bytov a domov určených na prenájom.

V neposlednom rade je potrebné **priebežne pokračovať aj v zabezpečovaní pomoci s bývaním ohrozeným skupinám obyvateľstva** vrátane primeranej ochrany sociálneho nájomného bývania a pomoci bezdomovcom.

Zmena súčasnej nie najlepšej situácie na britskom trhu s bývaním bude vecne aj časovo náročná. Britská vláda sa však rozhodla urobiť všetko pre jeho zásadnú revitalizáciu. Na riešenie súčasného deficitu vo výstavbe bytov a dosiahnutie súladu medzi rastom populácie a potrebami jej bývania bude potrebné stavať zhruba 250-tisíc bytov ročne (od roku 1970 pribudlo každý rok len približne 160-tisíc bytov).

Odhodlanie britskej vlády zlepšiť bývanie vo Veľkej Británii potvrdzuje aj zoznam konkrétnych úloh (s. 18 – 19), ktoré je potrebné splniť na zabezpečenie už spomenutých hlavných zámerov v oblasti bytovej výstavby. V nasledujúcich častiach materiálu sú tieto úlohy detailne rozpracované. Príloha obsahuje ďalšie podrobnosti súvisiace s koncipovaním týchto úloh a konzultáciami v jednotlivých navrhovaných krokoch.

Pri plnení konkrétnych úloh sa počíta s aktívnym prístupom miestnych orgánov, súkromných developerov, bytových družstiev, poskytovateľov infraštruktúry, úverových inštitúcií, ako aj rôznych miestnych komunít. Je potrebné spoločne prejsť, obrazne povedané, od vytvoreného plánu k tehlám a malte.

Súčasný trh s bývaním sa vo Veľkej Británii hodnotí vo všeobecnosti pomerne kriticky. Napriek tomu kroky na dosiahnutie zmeny naznačené v materiáli triezvejšie hodnotenia označujú skôr za evolučné ako revolučné. Materiál sa bude ďalej rozpracúvať s cieľom rozšíriť ho o ďalšie podnety a návrhy. S aktualizáciou sa počíta v priebehu roka.

Tento materiál by mohol byť dobrou inšpiráciou pre rezortné ministerstvo, ktoré je gestom bytovej výstavby v Slovenskej republike. Aj v našich podmienkach je potrebné kriticky hodnotiť stav bytového fondu. S 370 bytmi na tisíc obyvateľov sme na tom podľa posledného sčítania obyvateľov, domov a bytov ešte horšie ako Veľká Británia, kde pripadá 412 bytov na tisíc obyvateľov. V rámci prípravy vyhodnotenia plnenia úloh z Koncepcie štátnej bytovej politiky do roku 2020² by bolo preto vhodné orientačne porovnať napr. aj stav bytového fondu s teoretickou potrebou bývania v podmienkach Slovenska.

Ing. MIKULÁŠ CÁR, PhD.

Autor je expert Národnej banky Slovenska na makroekonomické súvislosti trhu s bývaním.

² *Koncepciu štátnej bytovej politiky do roku 2020 schválila vláda Slovenskej republiky 7. januára 2015. Predstavuje rámcový dokument štátu so zameraním na smerovanie ďalšieho rozvoja bývania.*

Recenzia publikácie/Review of Publication

Milan Terek:
INTERPRETÁCIA ŠTATISTIKY A DÁT, 4. doplnené vydanie

Košice: EQUILIBRIA, 2016. 428 s.

ISBN 978-80-8143-177-7

Milan Terek:
STATISTICS AND DATA INTERPRETATION, 4th Extended Edition

Košice: EQUILIBRIA, 2016. 428 p.

ISBN 978-80-8143-177-7

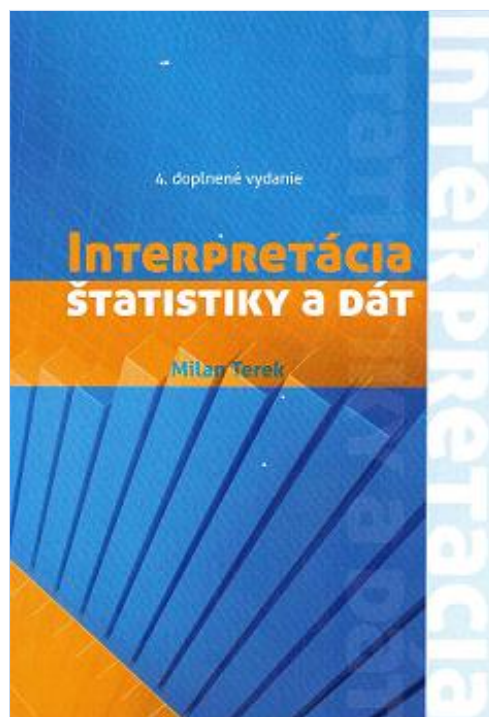
Vysokoškolská učebnica *Interpretácia štatistiky a dát, 4. doplnené vydanie* vyšla vo vydavateľstve EQUILIBRIA v roku 2016. Autorom učebnice je profesor Milan Terek, pôsobiaci na Katedre štatistiky Fakulty hospodárskej informatiky Ekonomickej univerzity v Bratislave.

Učebnica je určená predovšetkým študentom kurzu štatistiky na Vysokej škole manažmentu v Trenčíne/City University of Seattle. Svojím obsahom pokrýva celý rozsah tohto kurzu odporúčaný v roku 2016. Môžu ju, samozrejme, používať aj študenti ostatných vysokých škôl a univerzít a pracovníci z praxe, ktorí analyzujú dáta v najrôznejších oblastiach ekonomiky.

Učebnica má 428 strán, jej obsah je tematicky rozdelený do sedemnástich kapitol. Časť s prílohami obsahuje okrem tabuliek kvantilov a hodnôt distribučnej funkcie pre najviac využívané pravdepodobnostné rozdelenia aj tabuľku náhodných čísel a hodnoty Durbinovej-Watsonovej štatistiky. Užitočnou časťou učebnice je slovník termínov s definíciami, resp. vysvetlením použitých štatistických pojmov. Veľkým pozitívom je prehľadnosť textu, ktorá je výsledkom premysleného narábania s voľným priestorom a grafickým zvýraznením dôležitých informácií a definícií. Trochu rušivo pôsobí nejednotnosť grafickej úpravy obrázkov. Metodicky premyslené je zhrnutie najdôležitejších poznatkov, ktoré spolu s námetmi na ich precvičovanie uzatvára každú kapitolu.

V prvých troch kapitolách (*Dáta a štatistika, Opisná štatistika – tabuľkové a grafické prezentácie dát, Opisné charakteristiky*) sú uvedené základné štatistické pojmy, spôsoby prezentácie dát a základné opisné charakteristiky. Za pridanú hodnotu považujem časť venovanú odľahlým dátam a ich detekcii.

Ďalšie tri kapitoly (*Základy teórie pravdepodobnosti, Rozdelenia pravdepodobnosti, Náhodné vyberanie a výberové rozdelenia*) poskytujú základné vedomosti nevyhnutné na zvládnutie a pochopenie techník induktívnej štatistiky, ako je odhadovanie a testovanie štatistických hypotéz. Prínosom je zaradenie



problematiky náhodného vyberania, ktorej sa v učebniciach podobného zamerania venuje veľmi malá pozornosť.

Induktívna štatistika a jej techniky sú obsahom siedmej až jedenástej kapitoly (*Bodové odhadovanie, Intervalové odhadovanie, Testovanie štatistických hypotéz, Induktívne úsudky o parametroch dvoch základných súborov, Chí-kvadrát testy*).

Problematika lineárnej párovej regresie, viacnásobnej regresie a korelácie a formulácie regresného modelu je obsahom kapitol *Lineárna párová regresia, Viacnásobná regresia a korelácia* a kapitoly *Formulácia regresného modelu*. Hodnotným je autorovo zamyslenie nad kauzálnymi vzťahmi medzi premennými a netradičným je zahrnutie problematiky navrhovania experimentov.

V porovnaní s predchádzajúcimi vydaniaми je učebnica rozšírená o kapitoly *Analýza asociácie medzi kategoriálnymi premennými, Analýza časových radov a prognózovanie* a o kapitolu *Indexy*. Riešenie problematiky merania sily asociácie medzi kategoriálnymi premennými je neštandardne rozšírené o možnosti využitia korigovaných normovaných rezíduí, porovnania podielov a pomerov šancí.

Učebnica prináša nielen množstvo teoretických poznatkov, ale aj možnosť vidieť ich aplikáciu na celom rade praktických príkladov. Riešenia týchto úloh spolu s vysvetľujúcim slovným komentárom ukazujú možnosti ich praktického využitia.

Osobitosťou učebnice je aj snaha poukázať na etické aspekty aplikácie štatistických metód v praxi. Autor sa pri niektorých témach venuje dodržiavaniu etických princípov pri aplikácii štatistických metód v ekonomickej oblasti a pri prezentácii a interpretácii získaných výsledkov (v častiach *Opisná štatistika a etika, Niektoré etické aspekty interpretácie pravdepodobnosti, Intervaly spoľahlivosti a etické problémy*).

Učebnica *Interpretácia štatistiky a dát* má vzhľadom na svoj špecifický charakter nezastupiteľné miesto nielen medzi publikáciami, ktoré svojím obsahom pokrývajú základné kurzy štatistiky, ale hlavne medzi tými, ktorých cieľom je nájsť uplatnenie medzi odborníkmi v praxi. Jej význam a dôležitosť podčiarkuje aj skutočnosť, že sa v priebehu krátkočasového obdobia dočkala svojho v poradí už štvrtého, rozšíreného vydania.

Doc. RNDr. VIERA LABUDOVÁ, PhD.

Autorka je docentkou na Katedre štatistiky Fakulty hospodárskej informatiky Ekonomickej univerzity v Bratislave.

Informácia/Information

ŠTATISTICKÝ ÚRAD SR ZÍSKAL UZNANIE ZA PARTICIPÁCIU NA MEDZINÁRODNOM PROJEKTE V KAZACHSTANE

THE STATISTICAL OFFICE OF THE SR GAINED CREDIT FOR THE PARTICIPATION IN THE INTERNATIONAL PROJECT IN KAZAKHSTAN

Federálny štatistický úrad Nemecka DESTATIS ocenil Štatistický úrad Slovenskej republiky za výbornú spoluprácu, podporu a prístup k riešeniu problémov pri realizácii projektu *Posilnenie národného štatistického systému Kazachstanu* financovaného z pôžičky, ktorú poskytla Svetová banka. Štatistický úrad SR sa na projekte zúčastnil ako člen konzorcia, ktoré pod vedením Federálneho štatistického úradu Nemecka DESTATIS tvorili Slovensko, Fínsko, Rusko, Južná Kórea a Česká republika. Neskôr, z dôvodu náročnosti úloh spojených s realizáciou projektu, sa k odborníkom z týchto krajín pridali aj experti štatistických inštitúcií z Talianska, Dánska, Nórska, Kanady a Francúzska.

Projekt bol zameraný na inštitucionálne budovanie národného štatistického systému Kazachstanu, využívanie informačných technológií a zvyšovanie kvality jednotlivých štatistík. Kládol si za cieľ vytvoriť v Kazachstane podmienky na poskytovanie včasných, relevantných a spoľahlivých štatistických údajov v súlade s medzinárodnou metodikou. Prvé pracovné aktivity na projekte sa začali v roku 2012. Koncom roka 2016 projekt slávnostne ukončili na konferencii v Aстане.

Projekt bol rozdelený do siedmich základných okruhov (komponentov), ktoré sa ďalej členili na štyridsaťjeden subkomponentov. Komponenty zahŕňali už spomínané inštitucionálne budovanie národného štatistického systému vrátane inštitucionálneho rámca, zlepšenie operačného, informačného a komunikačného systému, riadenia ľudských zdrojov, štatistickej infraštruktúry, používaných noriem a metodiky a ďalší rozvoj jednotlivých štatistík.

Odborníci Štatistického úradu SR sa podieľali na štyroch oblastiach. Partnerom z Kazachstanu odovzdali poznatky zo zavádzania manažmentu kvality, z tvorby evidencie odpracovaného času, demografie a riešili s nimi otázky súvisiace s problematikou národných účtov. Absolvovali sedem misií v Kazachstane a kazašskí kolegovia sa zúčastnili na piatich študijných návštevách na Slovensku. Spolupráca sa uskutočňovala formou implementácie praktických odporúčaní konzultantov vrátane analytických aspektov. Kazašskí kolegovia sa pri konzultáciách zaujímali aj o fungovanie slovenského štatistického systému ako celku.

Expertí Štatistického úradu SR využili participáciu na projekte na prezentáciu svojej odbornosti a šírenie dobrého mena slovenskej štatistiky. Účasť na projekte bola pre nich príležitosťou získať širší rozhľad a nadviazať kontakty s kazašskými partnermi. Za prínos možno považovať aj spoluprácu a výmenu skúseností s členmi konzorcia, ktorá im umožnila detailnejšie sa oboznámiť s činnosťou a systémom práce partnerských štatistických inštitúcií.

Ing. ELENA BENKOVÁ

Autorka je riaditeľkou odboru európskych záležitostí a medzinárodnej spolupráce Štatistického úradu SR.

PRIPRAVUJEME

monotematické číslo *Slovenskej štatistiky a demografie* 3/2017 so zameraním na ***progressívne štatistické metódy a postupy v praxi***.

Oslovili sme renomovaných autorov z Fakulty hospodárskej informatiky Ekonomickej univerzity v Bratislave, Fakulty matematiky, fyziky a informatiky Univerzity Komenského v Bratislave a Fakulty informatiky a štatistiky Vysokiej školy ekonomickej v Prahe, aby predstavili využitie *rozhodovacích stromov* v prediktívnom modelovaní, *zhlukovú analýzu* a *bayesovské metódy*. Témou pripravovaných vedeckých článkov bude ďalej proces *navrhovania komplexných štatistických prieskumov a niektoré možnosti analýzy dát získaných z týchto prieskumov*. Potenciál zaujať má aj prezentácia *analýzy rozptylu*, metódy matematickej štatistiky známej pod skratkou ANOVA (Analysis of variance).

Vzhľadom na široké uplatnenie progresívnych štatistických metód a postupov v praxi by monotematické číslo 3/2017 *Slovenskej štatistiky a demografie* mohlo osloviť odbornú čitateľskú verejnosť pôsobiacu v rôznych oblastiach spoločenského života.

ONLINE VERZIA KOMPLETNÉHO ČÍSLA 2/2017 SLOVENSKEJ ŠTATISTIKY A DEMOGRAFIE BUDE VEREJNE DOSTUPNÁ na internetovej stránke Štatistického úradu SR www.statistics.sk **15. JÚLA 2017**.

* * *

COMING SOON

a monothematic issue of the *Slovak Statistics and Demography* No. 3/2017 focused on ***progressive statistical methods and procedures in practice***

We approached renowned authors from the Faculty of Economic Informatics of the University of Economics in Bratislava, Comenius University Faculty of Mathematics, Physics and Informatics and the Faculty of Informatics and Statistics of the University of Economics in Prague about presenting the uses of *decision trees* in predictive modelling, *cluster analysis and the Bayesian methods*. The prepared scientific articles will further focus on the process of *proposing complex statistical surveys and some possibilities of an analysis of data derived from these surveys*. The presentation on the *analyses of variance*, methods of mathematical statistics known by the acronym ANOVA (Analysis of variance) has also a potential to attract.

Due to the widely applied progressive statistical methods and procedures in practice, the monothematic issue No. 3/2017 of the *Slovak Statistics and Demography* should address the professional reading public from different areas of social life.

THE FULL ONLINE VERSION OF THE JOURNAL SLOVAK STATISTICS AND DEMOGRAPHY No 2 (2017) WILL PUBLICLY BE AVAILABLE AT THE WEBSITE OF THE STATISTICAL OFFICE OF THE SR www.statistics.sk on JULY 15, 2017.

SLOVENSKÁ ŠTATISTIKA A DEMOGRAFIA

je jediný recenzovaný vedecký časopis so zameraním na prezentáciu moderných štatistických a demografických metód a postupov. Propagujeme miesto a význam slovenskej štatistiky v Európskom štatistickom systéme, spoluprácu Eurostatu a národných štatistických úradov pri harmonizácii zisťovaní a multidimenzionálny rozmer štatistiky. Podporujeme rozvoj štatistickej teórie a jej prepojenie s praxou. Naším cieľom je prispievať k využiteľnosti štatistických výstupov v rôznych oblastiach a k zvyšovaniu ich kvality a efektivity.

Publikujeme analytické články, prognózy, názory, diskusné príspevky, recenzie, rozhovory, informácie a oznamy z rôznych oblastí štatistiky (národné účty, produkčné štatistiky, sociálne štatistiky, štatistika životného prostredia a pod.) a demografie (demografická štatistika, teoreticko-metodologické východiská demografie, historická demografia a pod.), vrátane sčítania obyvateľov, domov a bytov ako neodmysliteľnej súčasti demografickej štatistiky.

Vydáva:

Štatistický úrad SR

Identifikačné číslo vydavateľa:

IČO 00166197

Vychádza:

Štyrikrát ročne

Dátum vydania:

15. apríl 2017

Tlač:

Reprografické stredisko
Štatistického úradu SR

Predplatné:

20 eur (na rok)
5 eur (za jeden výťažok)

Objednávky prijíma:

Informačný servis
Štatistického úradu SR
Tel.: +4212/502 36 339
+4212/502 36 335
E-mail: info@statistics.sk

SLOVAK STATISTICS AND DEMOGRAPHY

is the only scientific peer-reviewed journal focusing on the presentation of modern statistical and demographic methods and procedures. Our aim is to promote the position and importance of Slovak statistics in the European statistical system, cooperation between the Eurostat and the national statistical offices in the field of survey harmonisation and the multidimensional character of statistics as well. We support the development of statistical theory and its connection with practice. We aim to contribute to the utility of statistical outputs in various fields and to the improvement of quality and efficiency.

We publish analytic articles, prognoses, views, discussion contributions, reviews, discussions, information and announcements from various statistical fields (national accounts, production statistics, social statistics, environmental statistics etc.) and demography (demographic statistics, theoretical and methodological bases of demography, historical demography etc.) including the population and housing census as an essential part of demographic statistics.

Issued by:

Statistical Office of the SR

Company registration number:

00166197

Published:

Four times a year

Date of issue:

15th April 2017

Press:

Reprographic centre of the
Statistical Office of the SR

Subscription:

20 Eur (per year)
5 Eur (for one copy)

Orders are to be addressed to:

Information Service of the
Statistical Office of the SR
Tel.: +4212/502 36 336
+4212/502 36 335
E-mail: info@statistics.sk



SLOVENSKÁ ŠTATISTIKA A DEMOGRAFIA

Recenzovaný vedecký časopis založený v roku 1991. Od roku 2014 sú jednotlivé čísla dostupné čitateľskej verejnosti s trojmesačným odstupom aj v elektronickej forme na www.statistics.sk. Názory autorov článkov sa nemusia zhodovať s názormi vydavateľa.

Zahraniční poradcovia/Foreign Consultants

Gabriela Czanner

University of Liverpool
Veľká Británia/United Kingdom

Jitka Langhamrová

Vysoká škola ekonomická v Praze
University of Economics in Prague
Česká republika/Czech Republic

Estefanía Mourelle Espasandín

Universidade da Coruña
Španielsko/Spain

Michaela Potančoková

Vienna Institute of Demography Austrian
Academy of Sciences, Wittgenstein Centre
for Demography and Global Human Capital
Rakúsko/Austria

Hana Řezanková

Vysoká škola ekonomická v Praze
University of Economics in Prague
Česká republika/Czech Republic

Milan Stehlík

Universidad Técnica Federico Santa María,
Valparaíso, Čile/Chile
Johannes Kepler University, Linz
Rakúsko/Austria

Výkonná redaktorka/Executive Editor

Zuzana Štukovská

Jazykové redaktorky/Language Editors

Slovenský jazyk/Slovak Language

Silvia Duchková

Anglický jazyk/English Language

Andrea Okenková

Adresa redakcie/Address of Editorial Office

Slovenská štatistika a demografia
Štatistický úrad SR
Miletičova 3, 824 67 Bratislava
Slovenská republika

SLOVAK STATISTICS AND DEMOGRAPHY

The scientific peer-reviewed journal founded in 1991. From 2014 individual copies of the journal will be available at intervals of three-months also in electronic form at the website www.statistics.sk. The opinions of the authors do not necessarily correlate with the opinions of the publisher.

Redakčná rada/Editorial Board

Ľudmila Ivančíková

(predsedníčka/chairwoman)
Štatistický úrad SR/Statistical Office of the SR

Mikuláš Cár

Národná banka Slovenska
National Bank of Slovakia

Ján Haluška

INFOSTAT Bratislava/INFOSTAT Bratislava

Ivan Janiga

Slovenská technická univerzita v Bratislave
Slovak University of Technology in Bratislava

Iveta Stankovičová

Univerzita Komenského v Bratislave
Comenius University in Bratislava

Erik Šoltés

Ekonomická univerzita v Bratislave
University of Economics in Bratislava

Pavol Tišliar

Univerzita Komenského v Bratislave
Comenius University in Bratislava

Boris Vaňo

INFOSTAT - Výskumné demografické centrum,
Bratislava
INFOSTAT - Demographic Research Centre,
Bratislava

Obálka/Cover

Klára Smutná

E-mailová adresa/E-mail address

SSaD@statistics.sk

www.statistics.sk