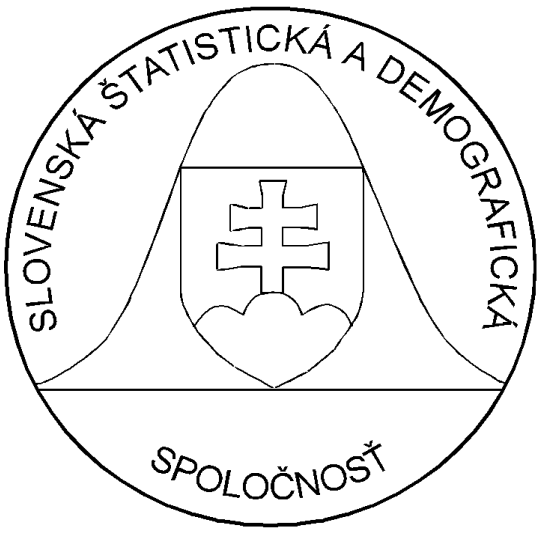
7/2009

FORUM STATISTICUM SLOVACUM



ISSN 1336-7420
9 771336 742001 97

Slovenská štatistická a demografická spoločnosť Miletičova 3, 824 67 Bratislava
www.ssds.sk

Naše najbližšie akcie:

(pozri tiež www.ssds.sk, blok Poriadané akcie)

NITRIANSKE ŠTATISTICKÉ DNI 2010
4. - 5. február 2010, Nitra

POHĽADY NA EKONOMIKU SLOVENSKA 2010
13. 4. 2010, Bratislava, Aula EU

EKOMSTAT 2010, 24. škola štatistiky
tematické zameranie: *Štatistické metódy vo vedecko-výskumnej, odbornej a hospodárskej praxi.*
jún 2010, Trenčianske Teplice

APLIKÁCIE METÓD NA PODPORU ROZHODOVANIA VO VEDECKEJ, TECHNICKEJ A SPOLOČENSKEJ PRAXI
jún 2010, STU Bratislava

15. SLOVENSKÁ ŠTATISTICKÁ KONFERENCIA
2010, SLOVENSKO

19. Medzinárodný seminár VÝPOČTOVÁ ŠTATISTIKA
2. – 3. 12. 2010, Bratislava, Infostat

PREHLIADKA PRÁC MLADÝCH ŠTATISTIKOV A DEMOGRAFOV
2. 12. 2010, Bratislava, Infostat

REGIONÁLNE AKCIE
priebežne

Pokyny pre autorov

Jednotlivé čísla vedeckého časopisu FORUM STATISTICUM SLOVACUM sú prevažne tematicky zamerané zhodne s tematickým zameraním akcií SŠDS. Príspevky v elektronickej podobe prijíma zástupca redakčnej rady na elektronickej adrese uvedenej v pozvánke na konkrétne odborné podujatie Slovenskej štatistickej a demografickej spoločnosti. Názov word-súboru uvádzajte a posielajte v tvare: ***priezvisko_nazovakcie.doc***

Forma: Príspevky písané výlučne len v textovom editore MS WORD, verzia 6 a vyššia do verzie 2003, písmo Times New Roman CE 12, riadkovanie jednoduché (1), formát strany A4, všetky okraje 2,5 cm, strany nečíslovať. Tabuľky a grafy v čierno-bielom prevedení zaradiť priamo do textu článku a označiť podľa šablóny. Bibliografické odkazy uvádzať v súlade s normou STN ISO 690 a v súlade s medzinárodnými štandardami. Citácie s poradovým číslom z bibliografického zoznamu uvádzať priamo v texte.

Rozsah: Maximálny rozsah príspevku je 6 strán.

Príspevky sú recenzované. Redakčná rada zabezpečí posúdenie príspevku členom redakčnej rady alebo externým oponentom.

Štruktúra príspevku: ***(Pri písaní príspevku využite elektronickú šablónu: http://www.ssds.sk/ v časti Vedecký časopis, Pokyny pre autorov.)***

Názov príspevku v slovenskom jazyku (*štýl Názov: Time New Roman 14, Bold, centrovať)*

Názov príspevku v anglickom jazyku *(štýl Názov: Time New Roman 14, Bold, centrovať)*

Vynechať riadok

Meno1 Priezvisko1, Meno2 Priezvisko2 *(štýl normálny: Time New Roman 12, centrovať)*

Vynechať riadok

Abstract: Text abstraktu v anglickom jazyku, max. 10 riadkov (štýl normálny: Time New Roman 12).

Vynechať riadok

Key words: Kľúčové slová v anglickom jazyku, max. 2 riadky (štýl normálny: Time New Roman 12).

Vynechať riadok

Kľúčové slová: Kľúčové slová v jazyku v akom je napísaný príspevok, max. 2 riadky (štýl normálny: Time New Roman 12).

Vynechať riadok

Vlastný text príspevku v členení:

1. **Úvod** *(štýl Nadpis 1: Time New Roman 12, bold, zarovnať vľavo, číslovať)*
2. **Názov časti 1** *(štýl Nadpis 1: Time New Roman 12, bold, zarovnať vľavo, číslovať)*
3. **Názov časti 1. . .**
4. **Záver** *(štýl Nadpis 1: Time New Roman 12, bold, zarovnať vľavo, číslovať)*

Vlastný text jednotlivý častí je písaný štýlom Normal: písmo Time New Roman 12, prvý riadok odseku je odsadený vždy na 1 cm, odsek je zarovnaný s pevným okrajom. Riadky medzi časťami nevynechávajte.

5. **Literatúra** *(štýl Nadpis 1: Time New Roman 12, bold, zarovnať vľavo, číslovať)*

[1] Písať podľa normy STN ISO 690

[2] GRANGER, C.W. – NEWBOLD, P. 1974. Spurious Regression in Econometrics. In: Journal of Econometrics, č. 2, 1974, s. 111 – 120.

Adresa autora (-ov) *(štýl Nadpis 1: Time New Roman 12, bold, zarovnať vľavo, adresy Time New Roman 11, zarovnať vľavo, vpísať do tabuľky bez orámovania s potrebným počtom stĺpcov a s 1 riadkom):*

Meno1 Priezvisko1, tituly1, študenti ročník	Meno2 Priezvisko2 , tituly2, študenti ročník
Katedra, fakulta, škola	Katedra, fakulta, škola
Ulica1, 970 00 Mesto1	Ulica2, 970 00 Mesto2
meno1.priezvisko1@mail.sk	meno2.priezvisko2@mail.sk

FORUM STATISTICUM SLOVACUM

vedecký recenzovaný časopis Slovenskej štatistickej a demografickej spoločnosti

Vydavateľ
Slovenská štatistická a demografická spoločnosť
Miletičova 3
824 67 Bratislava 24
Slovenská republika

Redakcia
Miletičova 3
824 67 Bratislava 24
Slovenská republika

Fax
02/39004009

e-mail
chajdiak@statis.biz
jan.luha@fmed.uniba.sk

Registráciu vykonalo
Ministerstvo kultúry Slovenskej republiky

Registračné číslo
3416/2005

Evidenčné číslo
EV 3287/09

Tematická skupina
B1

Dátum registrácie
22. 7. 2005

Objednávky
Slovenská štatistická a demografická spoločnosť
Miletičova 3, 824 67 Bratislava 24
Slovenská republika
IČO: 178764
DIČ: 2021504276
Číslo účtu: 0011469672/0900

ISSN 1336-7420

Ročník
V.

Číslo
7/2009

Cena výtlačku 20 EUR
Ročné predplatné 80 EUR

ÚVOD

Vážené kolegyne, vážení kolegovia,

siedme číslo piateho ročníka vedeckého časopisu Slovenskej štatistickej a demografickej spoločnosti (SŠDS) Forum Statisticum Slovacum (FSS) je zostavené z príspevkov, ktoré sú obsahovo orientované v súlade s tematikou 18. medzinárodného seminára Výpočtová štatistika a Prehliadkou prác mladých štatistikov a demografov. Tieto akcie sa uskutočnili v dňoch 3. a 4. decembra 2009 na Infostate v Bratislave. Organizátorom seminára bola SŠDS v spolupráci s Fakultou managementu UK v Bratislave, Infostatom a SAS Slovakia, s. r. o.

Akcie, z poverenia Výboru SŠDS, zorganizoval Organizačný a programový výbor v zložení: Ing. Iveta Stankovičová, PhD. – predseda, FM UK v Bratislave, doc. Ing. Jozef Chajdiak, CSc. – podpredseda, vedecký tajomník SŠDS, RNDr. Ján Luha, CSc. – tajomník, LF UK v Bratislave, RNDr. Mária Bohdalová, PhD. – FM UK v Bratislave, RNDr. Jitka Bartošová, PhD. – FM VŠE v Prahe, so sídlom v Jindřichovom Hradci, Doc. RNDr. Bohdan Linda, CSc., - FES Univerzita Pardubice.

Recenziu príspevkov zabezpečili: Ing. Iveta Stankovičová, PhD., RNDr. Ján Luha, CSc., RNDr. Mária Bohdalová, PhD., RNDr. Róbert Bohdal, PhD., RNDr. Jitka Bartošová, PhD., doc. Ing. Jozef Chajdiak, CSc..

Toto číslo vedeckého časopisu FSS 7/2009 zostavili a pre tlač pripravili: Ing. Iveta Stankovičová, PhD. a RNDr. Ján Luha, CSc..

Veľmi nás teší neustály záujem o seminár Výpočtová štatistika. Výbor SŠDS oceňuje aktivitu mladých v rámci Prehliadky prác mladých štatistikov a demografov, čo svedčí tiež o dobrej práci pedagógov a ich študentov. Dúfame, že možnosť prezentácie zvyšuje aj odbornú úroveň mladých štatistikov a demografov.

Výbor SŠDS

Z histórie seminárov Výpočtová štatistika
From the History of Seminars Computational Statistics

Pri príležitosti 18. ročníka semináru Výpočtová štatistika uvádzame stručnú chronológiu predošlých ročníkov.

Prvý seminár sa uskutočnil 9. - 10. 12. 1986 z iniciatívy zamestnancov Katedry štatistiky VŠE v Bratislave a Katedry statistiky VŠE v Prahe, ktorí sa zaoberali problematikou využitia výpočtovej techniky pri riešení štatistických úloh. Príspevky účastníkov boli uverejnené v Informáciách SDŠS č. 3 a č. 4 v roku 1986.

Miestom konania seminárov bola vždy budova Infostat-u a väčšina seminárov sa organizovala v spolupráci so Štatistickým úradom SR (resp. SŠÚ v Bratislave) a Infostat-om Bratislava (resp. VUSEIaR Bratislava).

Druhý seminár prebehol 8. 12. 1987, tretí seminár 11. - 12. 12.1990. Pád socializmu a spoločenské zmeny spôsobili určitú prestávku v organizácii seminárov Výpočtovej štatistiky, a preto sa 4. seminár uskutočnil až 7. - 8. 12. 1994.

Od 5. seminára uskutočneného 5. - 6. 12. 1996 sa už realizuje Výpočtová štatistika opäť každoročne ako medzinárodný seminár.

6. medzinárodný seminár Výpočtová štatistika sa uskutočnil 4.- 5. 12. 1997,

7. medzinárodný seminár Výpočtová štatistika sa uskutočnil 3. - 4. 12. 1998,

8. medzinárodný seminár Výpočtová štatistika sa uskutočnil 2. - 3. 12. 1999,

9. medzinárodný seminár Výpočtová štatistika sa uskutočnil 7. – 8. 12. 2000,

10. medzinárodný seminár Výpočtová štatistika uskutočnil 6. – 7. 12. 2001,

11. medzinárodný seminár Výpočtová štatistika sa uskutočnil 5. - 6. 12. 2002,

12. medzinárodný seminár Výpočtová štatistika sa uskutočnil 4. - 5. 12. 2003,

13. medzinárodný seminár Výpočtová štatistika sa uskutočnil 2. - 3. 12. 2004,

14. medzinárodný seminár Výpočtová štatistika sa uskutočnil 1. - 2. 12. 2005,

15. medzinárodný seminár Výpočtová štatistika sa uskutočnil 7. - 8. 12. 2006,

16. medzinárodný seminár Výpočtová štatistika sa uskutočnil 6. - 7. 12. 2007,

17. medzinárodný seminár Výpočtová štatistika sa uskutočnil 4. - 5. 12. 2008 a

18. medzinárodný seminár Výpočtová štatistika sa uskutočnil 3. - 4. 12. 2009.

Príspevky 2. seminára boli opublikované v Informáciách SDŠS č. 1/1999 a od 3. seminára sa publikujú v samostatnom Zborníku príspevkov príslušného seminára. Od 14. ročníka seminára sú už príspevky publikované vo vedeckom časopise vydávanom Slovenskou štatistickou a demografickou (SŠDS)s názvom FORUM STATISTICUM SLOVACUM.

Zameraním seminára je problematika na rozhraní počítačových vied a štatistiky.

Tematické okruhy posledných seminárov sa nemenia a sú:

- praktické využitie paketov štatistických programov,
- práca s rozsiahlymi súbormi údajov,
- vyučovanie výpočtovej štatistiky a príbuzných predmetov,
- praktické aplikácie výpočtovej štatistiky,
- iné.

V čase konania seminára Výpočtová štatistika sa uskutočňuje aj ***prehliadka prác mladých štatistikov a demografov***. Táto akcia prebieha od 7. seminára. Na 8. medzinárodnom seminári prezentovalo svoje práce 5 mladých štatistikov a demografov, na 9. medzinárodnom seminári už bolo 20 prác mladých štatistikov a demografov, na 10. bolo prihlásených 26 prác a na 11. bolo prihlásených 18 prác, ale vzhľadom na niekoľko prác vypracovaných skupinou autorov bol počet účastníkov vyšší než predošlý rok. Na 12. seminári bolo prihlásených 19 prác, pričom niektoré sú prácou viacerých autorov. Na ďalšom 13. seminári bolo prihlásených 9 prác od 12 autorov. V rámci 14. seminára bolo prihlásených 15 sólových prác mladých autorov. Na 15. seminári bolo prihlásených 20 prác mladých autorov. V rámci 16. seminára bolo prihlásených 17 sólových prác mladých autorov. V rámci 17. seminára bolo prihlásených 15 sólových prác mladých autorov. V aktuálnom ročníku 18. seminára je prihlásených 12 sólových prác mladých autorov.

Prípadní záujemcovia z radov mladých štatistikov a demografov (*Poznámka:* Za mladých štatistikov a demografov považujeme študentov 1. a 2. stupňa vysokolského štúdia pred ukončením vysokej školy. Študenti 3. stupňa vysokolského štúdia (tzv. doktorandi) už do tejto kategórie nepatria.) môžu získať informácie na *www.ssds.sk*, blok akcie a tiež na týchto e-mailových adresách:

chajdiak@statis.biz ; jan.luha@fmed.uniba.sk ; iveta.stankovicova@fm.uniba.sk.

Informácie o najbližšom seminári získate na webovskej stránke SŠDS *http://www.ssds.sk/* resp. *http://www.statistics.sk* v bloku Slovenská štatistická a demografická spoločnosť.

Doc. Ing. Jozef Chajdiak, CSc.
STU v Bratislave
vedecký tajomník SŠDS

RNDr. Ján Luha, CSc.
LF UK v Bratislave
člen sekretariátu Výboru SŠDS

Ing. Iveta Stankovičová, PhD.
FM UK v Bratislave
predsedníčka Programového a organizačného výboru seminára Výpočtová štatistika

Výběrové šetření příjmů domácností v České republice
Sample Survey of Household Incomes in the Czech Republic

Jitka Bartošová

Abstract: The main task of a sample survey of household income is to obtain representative data on income distribution in different types of households. Another task is to survey data on the type of housing, its quality and financial performance, then the equipment long term consumption of household items, etc. This paper deals with the description of the survey of household incomes in the Czech Republic and the problem of generalizing the results to the whole population.

Key words: Equivalent range, Calibration coefficient, Household income, Sample survey

Klíčová slova: ekvivalentní škála, kalibrační koeficient, příjem domácnosti, výběrové šetření

1. Úvod

Výběrové šetření příjmů domácností provádí v ČR Český statistický úřad. Od 50. let minulého století se jednalo o nepravidelné šetření prováděné pod názvem Mikrocensus (v intervalu 2 – 5 let). Po vzniku České republiky byl Mikrocenzus realizován v letech 1989, 1992, 1996 a 2002. Po vstupu ČR do Evropské unie nahradilo nepravidelné šetření Mikrocenzus každoroční zjišťování příjmů a životních podmínek domácností zvané EU – SILC. Poprvé bylo toto šetření provedeno Českým statistickým úřadem v roce 2005 pod názvem Životní podmínky 2005.

Článek vznikl v rámci řešení projektu GAČR 402/09/0515: *Analýza a modelování finančního potenciálu českých (slovenských) domácností*, který ve svých analýzách vychází z datové základny získané výběrovým šetřením příjmů domácností. [1]

2. Mikrocenzus

Účelem tohoto šetření bylo získat reprezentativní údaje o úrovni a struktuře příjmů a základní sociálně demografické charakteristiky domácností a jejich členů.

Mikrocenzní šetření bylo prováděno metodou dvoustupňového stratifikovaného výběru. Tento systém třídění zjišťovaných subjektů se používá tehdy, když základní soubor je příliš velký a prostorově rozptýlený. V tom případě by pouhý prostý náhodný výběr nedokázal zabránit nad či podhodnocování některých skutečností, což by vedlo ke zkreslení získaných výsledků. Postup je tedy takový, že ze základního souboru vybereme nejprve tzv. primární jednotky (obce), a v druhém kole ve vybraných primárních jednotkách vybereme tzv. sekundární jednotky (domácnosti).

Základním členěním subjektů je třídění na tzv. hospodařící domácnosti. Definicí tohoto pojmu se rozumí dobrovolné prohlášení osob bydlících ve vybraném bytě, že společně žijí a hospodaří, tj. hradí výdaje za stravu, ubytování apod.

Poslední Mikrocenzus byl realizován v roce 2002. Výběr domácností probíhal na základě informací z registru sčítacích obvodů, a to pro každý kraj nezávisle, aby bylo dosaženo rovnoměrného rozdělení. Obvody s méně než 24 byty nebyly do výběru zařazeny. Nejprve bylo vybráno metodou znáhodněného systematického výběru s pravděpodobnostmi zahrnutí přímo úměrnými počtu trvale obydlených bytů 50% z počtu plánovaných sčítacích obvodů. Ke každému byl následně vybrán ještě jeden sčítací obvod ze stejné sídelní jednotky,

příp. katastru. V jednotlivých sčítacích obvodech pak byl proveden v druhém kole prostý náhodný výběr dvanácti bytů.

Mikrocenzní šetření obsahuje údaje za tzv. hospodařící domácnosti. Hospodařící domácnost je definována jako skupina lidí, kteří se společně podílejí na hrazení základních výdajů domácnosti, jako je jídlo, náklady na bydlení, služby apod. Údaje o demografických faktorech (vzdělání, rodinný stav apod.), stejně jako údaje o peněžních a naturálních příjmech, byly v Mikrocenzu 2002 zjišťovány podle stavu ke konci roku 2002. Sledované znaky, jako je ekonomická aktivita, druh zaměstnání a odvětví se posuzovaly podle převažujícího stavu. V případě, že osoba ukončila v průběhu roku 2002 školní vzdělání, zaznamenával se stav ve 2. pololetí. V případě, že ve společné domácnosti žila některá osoba jen část sledovaného období (z důvodů stěhování, narození, delší nepřítomnosti dané pobytem v zahraničí, vykonávání základní vojenské služby či výkonu trestu), byla do počtu osob žijících v domácnosti započítávána jen z části.

Důležitou osobou je v mikrocenzním šetření tzv. osoba v čele domácnosti. V úplných rodinách se jako osoba v čele domácnosti bere vždy muž, a to bez ohledu na jeho ekonomickou aktivitu, tedy bez ohledu na to, zda jeho příjmy tvoří opravdu větší (příp. alespoň podstatnější než u partnerky) část rodinných příjmů, či dokonce zda nebyl nezaměstnaný. V neúplných rodinách (tj. pouze jeden rodič s dětmi) a u nerodinných domácností se posuzovala osoba v čele na základě její ekonomické aktivity popř. výše příjmu.

Mikrocenzus 2002 byl proveden jako šetření 11 040 bytů v České republice, což představuje přibližně 0,25% z celkového počtu všech trvale obydlených bytů. Z tohoto počtu se ukázalo 351 bytů (tj. 3,2%) jako neobydlených.

V jednotlivých krajích se lišila úspěšnost vyšetření jednotek zhruba o +/-10%. Nejnižšího vyšetření bylo dosaženo v Praze (61,9%), naopak nejlepšími výsledky se prezentovali tazatelé z Karlovarského kraje (81,3%). Ve výsledcích bylo také oproti původním předpokladům zastoupeno více domácností s členy v důchodovém věku. Byla zjištěna také nižší průměrná velikost domácnosti, než jakou přinesl výsledek sčítání lidu, bytů a domů (SLBD), provedený v roce 2001.

Stejný charakter mělo toto šetření i na Slovensku. Posledním výběrovým šetřením příjmů v SR byl Mikrocenzus 2003, který obsahoval údaje za osoby žijící ve společném bytě k 31. 12. 2002. Jediný rozdíl oproti českému Mikrocenzu 2002 spočíval v tom, že pokud domácnost odmítla odpovědět na zjišťované údaje, tazatel měl možnost při dodržení předepsaných pokynů najít náhradní domácnost. Z celkového počtu 19 569 vyšetřených domácností bylo v tomto šetření osloveno 5365 náhradních domácností, tj. 27,4%.

3. EU – SILC (Životní podmínky)

Účelem tohoto šetření je získat reprezentativní údaje o příjmovém rozdělení jednotlivých typů domácností, dále pak údaje o způsobu, kvalitě a finanční náročnosti bydlení, vybavení domácností předměty dlouhodobého užívání a také o pracovních, hmotných a zdravotních podmínkách dospělých osob žijících v domácnosti.

Hlavním rozdílem oproti dříve prováděným šetřením metodikou Mikrocenzu je především větší detailnost zjišťovaných informací a zpracování výstupů za jednotlivce. Výhodou zůstává velmi podobná metodika výběru domácností, a následná korekce dat přepočítáním dle metod uvedených u Mikrocenzů, včetně odhadů podcenění příjmů a eliminace chybějících příjmů.

V roce 2005 bylo výběrové šetření EU – SILC provedeno na vzorku 7000 bytů, tj. asi 0,16% z celkového počtu všech obydlených bytů. 354 jednotek se ukázalo jako neobydlených, příp. adresa nebyla nalezena nebo nebyla dostupná.

Z výsledků šetření vyplývá prakticky stejná struktura neúspěšných odpovědí v rámci zjišťování domácností, jakou mělo šetření Mikrocenzus. Stejně jako u něho se také opět lišila úspěšnost tazatelů v jednotlivých krajích, která se pohybovala od 51,1% (Praha) do 73,9% (Moravskoslezský kraj). To ukazuje bohužel na trvalý pokles úspěšnosti. Především výsledky z Prahy jsou považlivě nízké, protože se snižujícím se počtem vyšetřených domácností pochopitelně kvalita výběrových dat a tím i následných analýz na těchto datech provedených klesá.

4. Datová základna

Výběrové soubory pocházející z šetření Mikrocenzus a SILC jsou rozděleny do dvou částí. První část obsahuje informace o domácnostech – jejich disponibilních příjmech a různých demografických a sociálně-ekonomických faktorech, jako jsou sociální skupina osoby v čele, její věk, pohlaví a vzdělání, dále pak typ a velikost obce, ve které domácnost žije, počet členů domácnosti, počet nezaopatřených dětí apod. Další část souboru přináší doplňkové informace o osobách žijících v domácnosti, jejich názorech a postojích.

Z datových souborů můžeme získat informace nejen o celkovém disponibilním příjmu domácnosti (tzv. příjem na domácnost) a o příjmu přepočteném na jednoho člena domácnosti (tzv. příjem na hlavu), ale také o ekvivalentním disponibilním příjmu domácnosti (tzv. příjem na spotřební jednotku), který je porovnatelný s dalšími státy.

K transformaci příjmů do ekvivalentní škály používá Český statistický úřad dvojí metodiku:

1. *metodiku OECD*, kde
   - osoba v čele domácnosti je brána s koeficientem 1,0
   - děti ve věku 0 až 13 let s koeficientem 0,5
   - a ostatní děti a osoby s koeficientem 0,7
2. *metodiku EU*, kde
   - děti ve věku 0 až 13 let jsou brány s koeficientem 0,3
   - a ostatní děti a osoby s koeficientem 0,5

Počet spotřebních jednotek je dán součtem vah (koeficientů) přiřazených jednotlivým osobám v domácnosti. Závisí na složení domácnosti a věku dětí. Počet spotřebních jednotek dle definice OECD je v souboru uložen do proměnné sj a je dán vztahem:

$$sj = 1 + 0{,}5 \cdot mladsi\_\det i + 0{,}7 \cdot ostatni\_osoby\,, \qquad (1)$$

Počet spotřebních jednotek dle definice EU je v souboru uložen do proměnné ej a je dán vztahem:

$$ej = 1 + 0{,}3 \cdot mladsi\_\det i + 0{,}5 \cdot ostatni\_osoby\,, \qquad (2)$$

kde $mladsi\_\det i$ jsou děti ve věku 0 – 13 let,

$ostatni\_osoby$ jsou ostatní děti a osoby v domácnosti (kromě osoby v čele).

Pro porovnávání finančního potenciálu domácností v rámci EU je určena ekvivalentní škála s počtem spotřebních jednotek ej. Je zřejmé, že hodnota ej je ve většině případů menší než sj a přináší při mezinárodním srovnávání jiné výsledky. Tato škála je vhodná především k posuzování finanční situace domácností v zemích západní Evropy, kde domácnosti vynakládají mnohem větší částky na společné výdaje (bydlení, voda, energie, paliva apod.) než v postkomunistických zemích východní Evropy. Finanční situaci v zemích východní Evropy lépe vystihuje ekvivalentní škála dle metodiky OECD.

5. Zobecňování výsledků šetření

Výsledky šetření Mikrocenzus a EU – SILC nelze přímo zobecňovat na celou populace, neboť vybrané domácnosti nepředstavují reprezentativní vzorek. K tomuto závěru vede porovnáním s výsledky sčítání lidu, bytů a domů (SLBD). Prosté zobecňování informací získaných z výběrových souborů by proto mohlo vést ke zkreslení. Důvodů vedoucích k porušení reprezentativnosti vzorku je několik. Velkým problémem je stále klesající úspěšnost tazatelů, především v Praze. Na narušení reprezentativnosti se podílí také různá míra úspěšnosti tazatelů v jednotlivých regionech, vyšší zastoupení domácností důchodců a nižší zastoupení domácností, kde osoba v čele je samostatně činná atd. Také průměrná velikost domácnosti ve výběrech se liší od velikosti zjištěné v SLBD.

Z těchto důvodů není možné provést přepočet na celou populaci pomocí koeficientů, které poměřují počet vyšetřených domácností v kraji s jeho celkovým počtem obyvatel. Proto byla data přepočítána pomocí iterační metody kalibrace vah, minimalizující rozdíl mezi odhadnutými a přepočítanými výběrovými charakteristikami, vybranými takto pro každý kraj zvlášť, a to s pomocí těchto charakteristik:

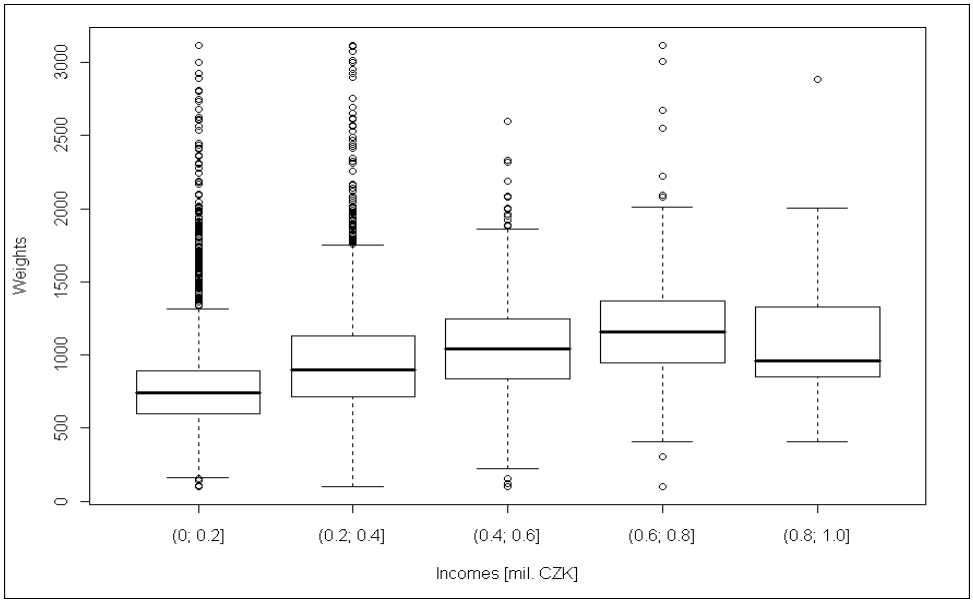
- *počet trvale obydlených bytů* – odhad stanovený na základě výsledků SLDB 2001 a přírůstků resp. úbytků počtu bytů za roky 2001 a 2002
- *počet osob bydlících v bytech* – odvozený ze středního stavu obyvatelstva ke 30. 6. 2002 podle demografické statistiky; protože šetření podléhaly pouze osoby žijící v bytech, byly od údajů z demografie odečteny počty osob žijících v tzv. ústavních domácnostech podle údajů statistiky sociálního zabezpečení za rok 2002
- *počet důchodců (pracujících i nepracujících)* – odvozený z údajů Ministerstva práce a sociálních věcí a České správy sociálního zabezpečení podle stavu ke konci 1. pololetí 2002, přičemž byl odečten počet osob žijících v domovech důchodců apod.,
- *počet nezaměstnaných* – údaje z evidence MPSV za rok 2002 byly povýšeny odhadem neregistrované nezaměstnanosti na základě výsledků VŠPS,
- *počet samostatně činných osob* – odhad stanovený na základě výsledků VŠPS za rok 2002 a výsledků SLBD 2001.

V šetření také dochází k asi 10% podhodnocení příjmů, a to jednak proto, že dotazovaní si na všechny své příjmy nevzpomenou nebo mají snahu udávat nižší příjmy, než odpovídají skutečnosti. Toto zkreslení se velice obtížně kvantifikuje, a proto jsou hodnoty korigovány po porovnání s údaji o průměrných hrubých mzdách. Podobně se postupovalo v případě sociálních dávek, kde uváděné hodnoty naopak překračují skutečnost.

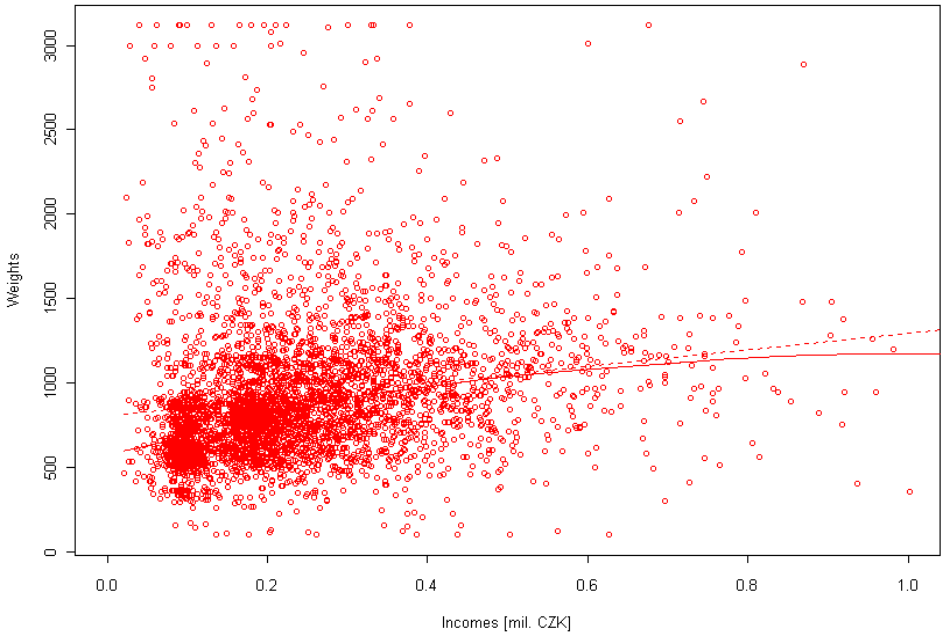
6. Zkoumání vlastností kalibračních vah

Z uvedených skutečností vyplývá, že důležitou součástí výběrových souborů jsou kalibrační váhy domácnosti – koeficienty PKOEF, které slouží k přepočtu výsledků z výběru na celou populaci. Grafy 1 a 2 znázorňují základní charakteristiky rozdělení kalibračních vah domácností v závislosti na výši uváděného příjmu.

Krabičkové diagramy (viz graf 1) zachycují základní kvantilové charakteristiky rozdělení kalibračních koeficientů v závislosti na velikosti ročních příjmů domácností setříděných do intervalů po 200 000,-Kč na (0;0,2) mil.Kč, (0,2;0,4) mil.Kč, (0,4;0,6) mil.Kč, (0,6;0,8) mil.Kč, (0,8;1) mil.Kč. Z diagramů je vidět, že váhy přiřazené jednotlivým domácnostem mají přibližně symetrické či mírně asymetrické rozdělení s množstvím zprava odlehlých hodnot. Kvantilové charakteristiky rozdělení kalibračních koeficientů rostou současně s růstem příjmů. Jejich růst se zastavuje u poslední skupiny domácností, jejíž čisté roční příjmy se pohybují v intervalu (0,8;1) mil.Kč. V této skupině nabývá rozdělení také výrazně asymetrického charakteru.



Graf 1: Box-plot rozdělení kalibračních vah v závislosti na velikosti příjmů domácností



Graf 2: Graf závislosti kalibračních vah na velikosti příjmů domácností

Také z dvourozměrných xy-grafů (viz graf 2) je vidět, že hodnoty kalibračních vah domácností s příjmy mírně rostou, ale tento růst se postupně zpomaluje a zastavuje. Jejich průběh proto bude lépe vystihovat polynomický regresní model (plná čára) než model lineární (přerušovaná čára).

7. Závěr

Výběrová šetření Mikrocensus a EU SILC představují datovou základnu pro získávání informací především o příjmech českých domácností. Z datových souborů můžeme získat informace o celkovém disponibilním příjmu domácnosti, o příjmu na jednoho člena i o ekvivalentním disponibilním příjmu, který je porovnatelný s dalšími státy v rámci EU.

Důležitou součástí výběrových souborů jsou kalibrační váhy domácností, které slouží k přepočtu výsledků z výběru na celou populaci. Kalibrační koeficienty vznikají iteračně, minimalizací rozdílu mezi odhadnutými charakteristikami a výběrovými charakteristikami přepočítanými pro každý kraj zvlášť.

Vliv kalibrace (tj. velikost vah) s růstem příjmů roste, ale tento růst se u vysokých příjmů zastavuje a u těch nejvyšších pak dochází spíše k mírnému poklesu. Z šetření vyplývá, že kalibrační váhy budou mít určitý vliv na souhrnné charakteristiky rozdělení příjmů. Jejich vliv na globální charakteristiky, jako je hustota či distribuční funkce, však bude zanedbatelný a nebude vyžadovat zásadní změny při modelování příjmového rozdělení.

8. Literatura

[1]BARTOŠOVÁ, J. 2009. Analysis and Modelling of Financial Power of Czech Households. Aplimat – Journal of Applied Mathematics, Vol. **2**, Nr. **3**, Slovak Technical University, Bratislava, 2009, s.31-36, ISSN 1337-6365.

[2]BARTOŠOVÁ, J. 2006. Volba a aplikace metod analýzy stavu rozdělení příjmů domácností v České republice po roce 1990. Fakulta informatiky a statistiky VŠE, Praha, 2006, 144s.

[3]Metodické vysvětlivky Mikrocenzus 2002. Dostupné z: <http://www.czso.cz>.

[4]Metodické vysvětlivky EU - SILC 2005, 2006, 2007. Dostupné z: <http://www.czso.cz>.

9. Zdroj dat

EU – SILC 2007

Adresa autora:

Bartošová Jitka, RNDr., Ph.D.
Jarošovská 1117/II
378 21 Jindřichův Hradec
bartosov@fm.vse.cz

Příspěvek byl vytvořen jako součást řešení projektu GAČR 402/09/0515: *Analýza a modelování finančního potenciálu českých (slovenských) domácností.*

Analýza a modelování spotřebního chování českých domácností
Analysis and Modelling of Consumer Behaviour of Czech Households

Jitka Bartošová

Abstract: The aim of the paper is the examination of consumer behaviour of households. Based on the results of the analysis is then developed an economic model describing the influence of various quantitative and qualitative factors in the Czech household spending in 2002 – 2007.

Key words: Consumer behaviour, Economic Model, Household.

Kľúčové slová: domácnost, ekonomický model, spotřební chování.

1. Úvod

Spotřebitelé na světovém trhu se vzájemně liší svým věkem, příjmem, vzděláním, vkusem atd. Nakupují nejrozmanitější druhy zboží a služeb [4]. Jejich volba při rozhodovacím nákupním procesu je ovlivněna především finančním potenciálem, preferencemi a různými demografickými, regionálními a socioekonomickými charakteristikami daného spotřebitele. Nezanedbatelný vliv má na jejich volbu také chování ostatních konzumentů a okolní prostředí. Zkoumání finančního potenciálu a výdajů domácností vzhledem k různým faktorům, které ovlivňují jejich spotřební chování, může být přínosem při rozhodování, jaké výrobky vyrábět a prodávat a komu je nabízet. Známe-li například typy domácností, které utrácejí značnou část svých příjmů za bydlení, můžeme při marketingových aktivitách cílit přímo na tuto skupinu a neplýtvat výdaji při propagaci tohoto zboží v sociálně slabších regionech, jako je například Mostecko. Avšak kromě území zde mohou hrát důležitou roli také další faktory, například věk osoby v čele, její vzdělání a s ním spojený sociální status domácnosti, počet ekonomicky aktivních členů, počet dětí v domácnosti apod.

Z výše uvedených skutečností vyplývá, že finanční potenciál domácností je jedním z klíčových faktorů ovlivňujících trh. Jeho zkoumání tedy přináší důležité informace nejenom pro hodnocení a porovnávání životní úrovně obyvatelstva [6], ale také pro nastavování parametrů výroby a obchodu. Zkoumáním a modelováním finančního potenciálu se zabývá projekt GAČR 402/09/0515: *Analýza a modelování finančního potenciálu českých (slovenských) domácností*, v jehož rámci článek vznikl [1].

Obecně lze očekávat, že hlavním faktorem ovlivňujícím výdaje domácností je jejich příjem. Domácnost s menším příjmem bude většinu svých peněz utrácet za statky nezbytné pro každodenní život, jako jsou potraviny, oblečení a nájem. Naproti tomu domácnosti s vyššími příjmy nebudou mít tendenci nakupovat více potravin, či oblečení, ale budou více utrácet za luxusnější a dražší zboží [2], [5]. V předloženém příspěvku bude provedena analýza celkových průměrných měsíčních výdajů českých domácností v letech 2002 – 2007 a na základě získaných výsledků bude vytvořen regresní model výdajů v závislosti na příjmech a dalších demografických a společensko-ekonomických faktorech.

2. Datová základna

Datovou základnu tvoří údaje ze Statistiky rodinných účtů z let 2002, 2005, 2006 a 2007. Jedná se o pravidelné šetření průměrných měsíčních příjmů a výdajů a dalších charakteristik českých domácností, které je každoročně prováděno Českým statistickým úřadem. Informace zahrnují:

- průměrné měsíční příjmy domácnosti,
- průměrné měsíční výdaje domácností na
    1. potraviny a nealkoholické nápoje
    2. alkoholické nápoje a tabák
    3. odívání a obuv
    4. bydlení, vodu, energii, paliva
    5. bytové vybavení a zařízení domácnosti
    6. zdraví
    7. dopravu
    8. pošty a telekomunikace
    9. rekreaci a kulturu
    10. vzdělávání
    11. stravování a ubytování
    12. ostatní zboží a služby

Data poskytují také širokou škálu dalších doplňkových informací o

- oblasti a velikosti obce, ve které daná domácnost žije,
- počtu osob, dětí a ekonomicky činných osob v domácnosti,
- počtu automobilů,
- vzdělání osoby v čele domácnosti,
- jejím sociálním zařazením
- a další údaje.

3. Model průměrných měsíčních výdajů domácnosti

Pro zjišťování a modelování závislosti spotřeby na různých faktorech byla použita analýza rozptylu a regresní analýza, které nám umožňují vyhledat vlivné faktory (regresory) a vystihnout dynamiku této závislosti [2].

Pracovní hypotéza vychází z předpokladu, že velikost výdajů domácnosti je nejvíce ovlivňována

- příjmy domácnosti,
- počtem spotřebních jednotek v domácnosti.

To znamená, že s růstem příjmů a počtu členů domácnosti budou růst i její výdaje. Nelze však vyloučit ani vliv další faktorů, a to jak kvantitativních, tak i kvalitativních. Příslušný model závislosti průměrných výdajů domácností můžeme tedy popsat vztahem:

$$E = f(I, sj, X), \tag{1}$$

kde E jsou celkové průměrné měsíční výdaje domácnosti,

I jsou čisté průměrné měsíční příjmy domácnosti,

sj je počet spotřebních jednotek v domácnosti podle definice OECD,

X jsou další zvolené kvantitativní a kvalitativní faktory.

Počet spotřebních jednotek sj je součtem vah jednotlivých osob v domácnosti. Výše spotřební jednotky závisí na složení domácnosti a věku dětí. Podle definice OECD je sj dána vztahem:

$$sj = 1 + 0{,}5 \cdot mladsi\_\det i + 0{,}7 \cdot ostatni\_osoby, \tag{2}$$

kde $mladsi\_\det i$ jsou děti ve věku 0 – 13 let,

$ostatni\_osoby$ jsou ostatní děti a osoby v domácnosti (kromě osoby v čele).

Abychom mohli do regresního modelu zabudovat kvalitativní proměnné, musíme je rozepsat do tzv. dummy proměnných a zařadit je do modelu ve formě umělých proměnných. To současně znamená, že musíme vynechat vždy jedenu z variant, aby mezi těmito proměnnými nevznikla lineární závislost (kolinearita) a model byl validní.

Pokud bychom ponechali původní rozdělení kvalitativních proměnných do tříd, které je značně podrobné, vzniklo by v modelu velké množství umělých proměnných. To sebou ovšem přináší nebezpečí porušení podmínek nezávislosti těchto faktorů a často vede k neúnosné míře multikolinearity v modelu. Z tohoto důvodu je potřeba redukovat počet nezávislých proměnných v modelu vhodně volenou agregací (např. u proměnných typ vzdělání osoby v čele domácnosti (vzd_p) nebo velikost obce, v níž domácnost žije (vel) apod.) .

Důvod výběru pouze některých proměnných je dán Parsimony faktorem, který říká, že v matematickém modelu je vhodné použít méně parametrů. Více parametrů by do modelu zanášelo více chyb, což by způsobilo snižování stupňů volnosti a menší vypovídající schopnost modelu. [7]

4. Výběr statisticky významných faktorů a tvorba váženého regresního modelu

Vzhledem k tomu, že rozdělení příjmů i výdajů je zešikmené a obsahuje odlehlé hodnoty, využijeme pro vystižení jejich závislosti logaritmickou transformaci. Z široké nabídky regresních modelů si proto zvolíme logaritmicko-logaritmický model.

K odhadu parametrů tohoto modelu bude vhodné použít váženou metodou nejmenších čtverců, neboť domácnosti nemají v souboru stejnou váhu. Každé domácnosti je váha přiřazena v proměnné $pkoef$, takže v modelu není zajištěna podmínka homoskedasticity. Relativní zastoupení libovolné $i-té$ domácnosti v datovém souboru získáme podílem

$$w_i = \frac{pkoef_i}{\sum_{i=1}^{n} pkoef_i} \tag{3}$$

kde $pkoef_i$ je koeficient přiřazený $i-té$ domácnosti.

Zvolený regresní model má tedy tvar:

$$\sqrt{w_i}\ln E = \sqrt{w_i}a_0 + a_1\sqrt{w_i}\ln I + a_2\sqrt{w_i}sj + a_3\sqrt{w_i}auto + a_4\sqrt{w_i}ea + a_5\sqrt{w_i}vek\_p + ... \tag{4}$$

kde **kvantitativními proměnnými** v modelu jsou:

- E jsou průměrné měsíční výdaje domácnosti celkem
- I jsou čisté průměrné měsíční příjmy domácnosti
- $auto$ je počet osobních automobilů v domácnosti
- sj je počet spotřebních jednotek za jednotlivé osoby v domácnosti
- ea je počet ekonomicky aktivních členů domácnosti
- vek_p je věk (stáří) osoby v čele
- vek_p^2 je věk (stáří) osoby v čele – kvadrát

kvalitativními proměnnými v modelu jsou:

- $pohl\_X$ je pohlaví osoby v čele

pohl _ muz je muž

pohl _ zena je žena

dum _ X je druh domu

dum _1 je rodinný dům stojící samostatně

dum _2 je dvojdomek, řadový domek

dum _3 je bytový dům s méně než 10 byty

dum _4 je bytový dům s 10 a více byty

dum _5 je jiný druh

skup _ X je sociální skupina osoby v čele

skup _1 je nižší zaměstnanec

skup _2 je samostatně činný

skup _3 je vyšší zaměstnanec

skup _4 není nepopsána

skup _6 je důchodce v domácnosti s EA členy

skup _7 je důchodce v domácnosti bez EA členů

skup _8 je nezaměstnaný

skup _9 jsou ostatní

vel _ X je velikost obce, v níž domácnost žije – agregováno do 3 kategorií

vel _1 je 1 000 až 10 000 obyvatel

vel _2 je 10 000 až 100 000 obyvatel

vel _3 je nad 100 000 obyvatel

vzd _ X je vzdělání osoby v čele – do 3 kategorií

vzd _1 je vzdělání nízké

vzd _2 je vzdělání střední

vzd _3 je vzdělání vysoké

Zápis zvoleného modelu si můžeme ještě formálně zjednodušit zavedením nových proměnných:

$$E^* = a_0^* + a_1 \cdot I^* + a_2 \cdot sj^* + a_3 \cdot auto^* + a_4 \cdot ea^* + a_5 \cdot vek\_p^* + a_6 \cdot vek\_p^{*2} + ... \quad (5)$$

kde $E^* = \sqrt{w_i} \ln E$, $I^* = \sqrt{w_i} \ln I$, $sj^* = \sqrt{w_i}\, sj$, $auto^* = \sqrt{w_i}\, auto$, $ea^* = \sqrt{w_i}\, ea$,.....

Po odlogaritmování dostaneme multiplikativní model závislosti průměrných měsíčních výdajů domácností na vybraných faktorech ve tvaru

$$E = e^{a_0} \cdot I^{a_1} \cdot e^{a_2 \cdot sj + a_3 \cdot ea + a_4 \cdot auto + a_5 \cdot vek\_p + a_6 \cdot vek\_p^2 .....} \quad (6)$$

a který můžeme dále upravit na tvar:

$$E = b_0 \cdot I^{a_1} \cdot b_2^{sj} \cdot b_3^{ea} \cdot b_4^{auto} \cdot b_5^{vek\_p} \cdot b_6^{vek\_p^2} \cdot b_7^{pohl\_muz} \cdot b_8^{dum\_1} \cdot b_9^{dum\_3} \cdot ....., \quad (7)$$

kde $b_0 = e^{a_0}$, $b_2 = e^{a_2}$, $b_2 = e^{a_2}$, ...

Následující tabulka 1 osahuje informace o odhadech regresních koeficientů modelu a o jejich statistické významnosti. Výsledky testování významnosti modelu a procento vystižení změn výdajů vybraným ekonomickým modelem jsou uvedeny v tabulce 2.

Tabulka 1: Odhad parametrů modelu celkových výdajů domácností (výstup z programu)

Parameter	**Estimate**	**Standard Error**	**T-Statistic**	**P-Value**
CONSTANT	1,0001	0,0322873	30,975	0,0000
ln_I	0,691336	0,00709285	97,4694	0,0000
auto	9,63602	0,473081	20,3686	0,0000
dum_1	-2,00558	0,584106	-3,43359	0,0006
dum_3	1,26365	1,18148	1,06955	**0,2848**
dum_5	-8,52788	3,03841	-2,80669	0,0050
ea	-2,31134	0,566659	-4,07889	0,0000
pohl_muz	0,349333	0,68554	0,509574	**0,6103**
sj	11,5519	0,475733	24,2823	0,0000
skup_1	-6,74984	1,1778	-5,73088	0,0000
skup_2	6,22497	1,26031	4,93923	0,0000
skup_3	-3,79076	1,17347	-3,2304	0,0012
skup_7	-8,0329	1,37078	-5,86011	0,0000
skup_8	-5,04455	1,89375	-2,66379	0,0077
vek_p	1,52663	0,0909807	16,7797	0,0000
vek_p^2	-0,0164417	0,00095604	-17,1977	0,0000
vel_1	-2,28834	0,543282	-4,21207	0,0000
vel_3	3,97277	0,626675	6,33945	0,0000
vzd_1	-2,12198	0,555796	-3,81791	0,0001
vzd_3	3,49794	0,786682	4,44645	0,0000

Významnost modelu i jeho parametrů byly testovány na pětiprocentní hladině významnosti (tj. $\alpha = 0{,}05$), takže všechny proměnné s hodnotou p-value větší než 0,05 můžeme z modelu vypustit jako nevýznamné. Z tabulky 1 vyplývá, že nejvyšší hodnota p-value je u proměnných $pohl\_muz$ a dum_3. Můžeme je tedy označit za nevýznamné a z regrese vyloučit. Všechny další regresní parametry jsou významné. Výsledky regresní analýzy potvrzují stanovenou pracovní hypotézu, že nejvýznamnějšími faktory, které ovlivňují výdaje domácnosti, jsou příjmy (I) a počet spotřebních jednotek (sj). Dalšími významnými faktory jsou proměnné $auto$ (tj. počet aut v domácnosti) a vek_p (tj. věk osoby v čele domácnosti).

Tabulka 2: Testování významnosti modelu(výstup z programu)

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
Model	2269,57	17	133,504	2676,30	**0,0000**
Residual	609,98	12228	0,0498839		
Total (Corr.)	2879,55	12245			

R-squared = 78,8168 percent

R-squared (adjusted for d.f.) = 78,7874 percent

Standard Error of Est. = 0,223347

Mean absolute error = 0,163889

Durbin-Watson statistic = 1,97265 (P = 0,0651)

Výsledky analýzy rozptylu prokázaly statistickou významnost zvoleného regresního modelu (p-value = 0,0000). Rovněž výstižnost změn výdajů modelem je poměrně dobrá – vybrané regresní proměnné vystihují změny průměrných měsíčních výdajů domácností téměř ze 79% (R-squared (adjusted for d.f.) = 78,7874 percent).

Korelační matice prokázala existenci poměrně vysoké korelace mezi některými regresory. Mezi nezávislými proměnnými bylo detekováno celkem sedm korelačncích koeficientů s absolutními hodnotami většími než 0,5. Přestože tyto hodnoty nepřekročily hranici pro neúnosnou míru autokorelace, mohli bychom uvažovat o jejich vyloučení z modelu. Tím by došlo k dalšímu zjednodušení modelu, které je v případě použití vysokého počtu umělých proměnných žádoucí. Povětšinou se jedná pouze o dummy proměnné jedné kvalitativní proměnné, která určuje sociální skupinu osoby v čele domácnosti ($skup\_X$).

Výsledný ekonomický model závislosti výdajů domácnosti na velikosti jejích příjmů, počtu spotřebních jednotek a dalších vybraných faktorech, má tedy tvar:

$$\begin{aligned} E = {} & e^{0,996099} \cdot I^{0,692397} \cdot e^{11,6155 \cdot sj + 9,70279 \cdot auto + 6,23627 \cdot skup\_2 + 3,95864 \cdot vel\_3} \cdot \\ & \cdot e^{+3,54606 \cdot vzd\_3 + 1,52296 \cdot vek\_p - 0,0163958 \cdot vek\_p^2 - 8,62651 \cdot dum\_5 - 2,08307 \cdot dum\_1} \cdot \\ & \cdot e^{-8,00388 \cdot skup\_7 - 6,74302 \cdot skup\_1 - 4,99488 \cdot skup\_8 - 3,85557 \cdot skup\_3 - 2,23444 \cdot vel\_1} \cdot \\ & \cdot e^{-2,31208 \cdot ea - 2,09599 \cdot vzd\_1} \end{aligned} \qquad (8)$$

5. Závěr

Cílem předloženého příspěvku bylo pomocí regresního modelu najít faktory, které výrazně ovlivňují výdaje domácností. Ukázalo se, že nejvlivnějšími faktory jsou výška příjmů, počet spotřebních jednotek, počet aut v domácnosti a věk osoby v čele. Na základě provedených zjištění v této práci můžeme potvrdit pracovní hypotézu, kterou jsme si na začátku stanovili. Zjistili jsme, že příjmy i počet spotřebních jednotek významně ovlivňují spotřebu jednotlivých domácností. To znamená, že s rostoucími příjmy nebo se zvětšujícím se počtem spotřebních jednotek stoupají výdaje domácností.

Zajímavým zjištěním je informace, že obecně méně významnými faktory jsou velikost obce, ve které domácnost žije, vzdělání osoby v čele a sociální zařazení domácnosti. Lze však konstatovat, že výdaje rostou rovněž s věkem osoby v čele domácnosti. Vyšší výdaje sebou přináší rovněž bydlení ve městech s více než 100 000 obyvateli a také dvě další

charakteristiky osob stojících v čele domácnosti – dosažení vysokoškolského vzdělání a příslušnost ke skupině samostatně činných osob.

6. Literatura

[1]BARTOŠOVÁ, J. 2009. Analysis and Modelling of Financial Power of Czech Households. Aplimat – Journal of Applied Mathematics, Vol. **2**, Nr. **3**, Slovak Technical University, Bratislava, 2009, s.31-36, ISSN 1337-6365.

[2]BARTOŠOVÁ, J. 2007. Mikro a makroekonomické úlohy řešené pomocí programu DERIVE 5. 1. vydání, Praha, Oeconomica 2004. 85 s. ISBN 80-245-0758-7.

[3]HUŠEK R., PELIKÁN J. 2001. Aplikovaná ekonometrie. Praha 2001, ISBN 80-245-0219-4.

[4]TOUFAROVÁ Z. 2007. Identifikace klíčových faktorů ovlivňujících spotřebitelské chování domácností. Diplomová práce. Mendelova zemědělská a lesnická univerzita v Brně, 2007. Dostupné z: < http://is.mendelu.cz/lide/clovek.pl?id=1702;zalozka=13;studium=23676>.

[5]RYTÍŘ. L. 2006. Základní principy ekonometrie. 1. Vydáno dne 19. 08. 2006, Dostupné z: < http://proatom.luksoft.cz/view.php?cisloclanku=2006081901>.

[6]STANKOVIČOVÁ, I., BARTOŠOVÁ, J. 2009. Príspevok k analýze subjektívnej chudoby v SR a ČR. In FORUM STATISTICUM SLOVACUM **3/**2009, Slovenská štatistická a demografická spoločnosť, Bratislava, 2009. (CD ROM) s.1-12. ISSN 1336-7420.

[7]Wikipedia, Dostupné z: <http://en.wikipedia.org/wiki/Parsimony>.

Zdroj dat

SRÚ 2002, 2005, 2006, 2007

Adresa autora

RNDr. Jitka Bartošová, Ph.D.
Vysoká škola ekonomická v Praze
Fakulta managementu
Jarošovská 1117/II
377 01 Jindřichův Hradec
bartosov@fm.vse.cz

Příspěvek byl vytvořen jako součást řešení projektu GAČR 402/09/0515: *Analýza a modelování finančního potenciálu českých (slovenských) domácností.*

Špecifiká záznamu a spracovania dát z Národného registra pacientov s vrodenou vývojovou chybou
Specifics of entering and data processing from National register of patients with congenital anomaly

Daniel Böhmer, Daniela Brašeňová, Ján Luha

Abstract: This article deal with specific tasks entering and data processing multiresponse variables in National register of patients with congenital anomaly. Authors recommend procedures how to solve these problems in environment of SPSS.

Key words: congenital anomalies, frequency, data entering, data processing, multiresponse variables.

Kľúčové slová: vrodené vývojové chyby, frekvencia, záznam dát, spracovanie dát, premenné s viacerými možnosťami odpovede.

1. Úvod

Základné metodologické aspekty zberu a záznamu dát premenných s možnosťou viacerých odpovedí možno nájsť napríklad v práci [7]. V tomto príspevku sa zaoberáme možnosťami a úskaliami štatistického spracovania takýchto otázok na konkrétnej aplikácii z oblasti vrodených vývojových chýb. Pripomenieme si aj špecifiká záznamu premenných s viacerými možnosťami odpovede (multiresponse).

Výskyt vrodených vývojových chýb (VVCh) je evidovaný v hlásení zdravotného stavu Z (MZ SR) 6 - 12 Hlásenie vrodenej chyby (HVCh) v Národnom centre zdravotníckych informácií Bratislava. Bližšie informácie sú na webovej stránke [13] http://www.nczisk.sk/buxus/generate_page.php?page_id=555.

Jednotlivé VVCh sú označované podľa Medzinárodnej klasifikácie chorôb (MKCH-10). V hlásení vrodenej chyby sa sledujú všetky diagnózy kapitoly XVII. Vrodené chyby, deformácie a chromozómové anomálie (diagnózy Q00 až Q99) a vybrané diagnózy z kapitol III. Choroby krvi a krvotvorných orgánov a niektoré poruchy imunitných mechanizmov, IV. Choroby žliaz z vnútorným vylučovaním, výživy a výmeny látok a XI. Choroby tráviacej sústavy. Bližšie o VVCh - viď napr. práce [1] až [5] a [9].

Formulár „Hlásenie vrodenej chyby“ (možno nájsť na stránke http://www.nczisk.sk), okrem nutných demografických charakteristík a charakteristík zdravotného stavu, zachytáva výskyt a opis VVCh. Problém je v tom, že sa u toho istého živonarodeného novorodenca môže vyskytnúť nielen jedna ale aj viac u neho zistených VVCh. Preto formulár umožňuje záznam až do 9 VVCh u jedného novorodenca. Pri zázname možno využiť 3-znakový kód z MKCH-10 alebo rozširujúci 4 znakový kód (na spresnenie podtypu VVCh).

Úskalia pri štatistickom spracovaní závisia od spôsobu záznamu a od štatistického softvéru používaného na spracovanie. Špecifické úskalia sa objavujú pri potrebe zlúčiť kódy – v uvažovanej aplikácii keď potrebujeme spracovať nie 4-znakové kódy chorôb ale redukované na 3-znakové z MKCH-10.

Problémy so záznamom a štatistickým spracovaním premenných s viacerými možnosťami odpovede sú spoločným znakom viacerých oblastí. Napríklad aj iné zdravotnícke registre sa stretávajú s takýmito problémami. V príspevku opísané riešenia aplikované na HVCh sú riešením spracovania takýchto údajov aj pre iné registre.

Štatistický softvér SPSS má najlepšie prepracovaný systém spracovania multiresponse premenných. Poskytuje dva systémy ich spracovania – v procedúrach Multiple response Base

modulu a tiež v module Tables. V module Tables dovoľuje spracovávať znakové (string) aj numericky kódované multirespose premenné a v procedúrach Multiple response Base modulu iba numericky kódované premenné.

2. Záznam VVCh v HVCh

V príspevku sa venujeme iba špecifickým otázkam záznamu a štatistického spracovania dát o hlásených VVCh u živonarodených novorodencov, preto ostatné sledované charakteristiky v HVCh neskúmame.

V nasledujúcej tabuľke uvádzame vybrané príklady záznamu dát o VVCH, ktoré ilustrujú rôzne možnosti ich výskytu u jednotlivých živonarodených novorodencov.
Celý súbor obsahuje v prevažnej miere záznamy novorodencov s jednou vývojovou chybou:

DG_1	DG_2	DG_3	DG_4	DG_5	DG_6	DG_7	DG_8	DG_9
Q210	Q248							
Q66								
Q900	Q158							
Q23	Q24							
Q213								
Q211								
Q700	Q702							
Q210	Q54							
Q900	Q210	Q211						
Q712								
Q909	Q212							
Q620								
Q213								
Q210	Q250							
Q90	Q133	Q211						
Q210	Q211							
Q897	Q048	Q210	Q211	Q360	Q354	Q922		

Ako vidno, pri zázname dát o VVCh v v Hlásení vrodenej chyby máme niekoľko špecifík - dáta sú zaznamenávané:

- alfanumerickým kódom,
- používa sa 3-znakový a 4-znakový kód.

V práci [7] sú uvádzané dva spôsoby záznamu "multiresponse" premenných pri numerickom kódovaní odpovedí. Systém záznamu zvolený v HVCh zodpovedá prvému spôsobu záznamu s tým rozdielom, že bol zvolený záznam alfanumerických kódov podľa MKCH-10. Druhý spôsob uvádzaný v citovanej práci je postup, keď vytvoríme toľko stĺpcov, koľko je možných VVCh a zaznamenávame 1, keď sa daná VVCh vyskytla a 0 aj je to inak. Tento spôsob je ale, vzhľadom na veľký počet možností VVCh, prakticky nevhodný. Teoreticky je v kapitole XVII. možných až do 1 000 kategórií VVCh.

Pri štatistickom spracovaní vyplýva z vlastnosti „multiresponse" a aj zo zvoleného spôsobu kódovania VVCh niekoľko „úskalí", ktoré budeme skúmať v nasledujúcej kapitole.

Prvým je potreba spájania kategórií odpovedí – napríklad pre potreby redukovania 4-znakového kódu na 3-znakový.

Druhé súvisí s vytváraním špeciálnych skupín VVCh – či už priamo definovaných v MKCH-10, alebo pre špecifické účely výskumov.

V nasledujúcej tabuľke uvedieme príklad spájania kódov pri redukcii na 3-znakový kód. Na ilustráciu využijeme prv uvedenú tabuľku, kde bol aplikovaný 4-znakový a tiež aj 3-znakový kód.

D1	D2	D3	D4	DG	DG	DG	DG	DG
Q21	Q24							
Q66								
Q90	Q15							
Q23	Q24							
Q21								
Q21								
Q70	**Q70**							
Q21	Q54							
Q90	**Q21**	**Q21**						
Q71								
Q90	Q21							
Q62								
Q21								
Q21	Q25							
Q90	Q13	Q21						
Q21	**Q21**							
Q89	Q04	**Q21**	**Q21**	Q36	Q35	Q92		

V tabuľke sú hrubo vyznačené prípady, ktoré pri štatistickom spracovaní môžu spôsobiť určité problémy. Napríklad pacient s výskytom VVCh Q700 a Q702 má v novej tabuľke dva rovnaké kódy Q70 a podobne aj v iných prípadoch.

Ak by sme realizovali transformácie, ktoré vytvárajú 0/1 premenné, tak, že hodnota je rovná 1, ak nastane príslušný jav (diagnóza), napr. jav Q70, tak sa pri spracovaní takejto premennej „stratia“ duplicitné, resp. multiciplitné kódy – ako je to napr. v hrubo vyznačených prípadoch v tabuľke a tým dostaneme nepresné výsledky analýz.

3. Vybrané problémy štatistickej analýzy VVCh

Vybrané problémy štatistického spracovania sa týkajú „úskalí“, ktoré sú opísané v predošlej kapitole. Štatistické spracovanie budeme ilustrovať na databáze VVCh za rok 2008 z HVCh v prostredí štatistického softvéru SPSS, v ktorom je problematika spracovania multiresponse premenných najlepšie realizovaná, spomedzi profesionálnych štatistických softvérov. Ak chceme využiť alfanumerické kódovanie VVCH, tak ako je to v HVCh, tak musíme použiť v SPSS modul Tables.

V roku 2008 bolo registrovaných celkom 1309 živonarodených novorodencov s VVCH, pričom celkovo bolo u nich zaregistrovaných 1661 VVCh, teda takmer 127%.

Kvôli úspore miesta (počet rôznych diagnóz a typov VVCh v roku 2008 bol 304) uvádzame iba výrez jednoduchej tabuľky získanej v module Tables. Aby sme mohli využiť na spracovanie procedúry Multiple Response v Base module musíme kódy odpovedí pri premenných DG_1 až DG_9 rekódovať numericky – napr. 1=Q00, 2=Q001, 3= Q002 atď., až 304=E84. Môžeme vytvoriť nové premenné diag1 až diag9 s numerickými kódmi. V module Tables môžeme tieto rekódované premenné samozrejme tiež spracovávať. V ďalšej tabuľke ilustrujeme výstup z procedúry Multiple Response Base modulu. Výsledky sú rovnaké – výstup z modulu Tables môžeme taktiež doplniť o percentá vzhľadom na počet prípadov.

Case Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
$diag(a)	**1309**	100,0%	0	0%	1309	100,0%

Výpočet v module Tables: Počet odpovedí=1661

		Count	Column N %
$DG		**1309**	100
	E03	1	0,076394
	E70	1	0,076394
	E84	1	0,076394
	Q00	1	0,076394
	Q000	3	0,229183
	Q001	2	0,152788
	Q010	2	0,152788
	Q018	2	0,152788
	Q02	6	0,458365
	Q03	9	0,687548
	Q038	2	0,152788

Výpočet pomocou procedúry Multiple Response v Base module:

$diag Frequencies				
		Responses		Percent of Cases
		N	Percent	
$diag(a)	1 Q00	6	0,361228	0,458365
	2 Q01	4	0,240819	0,305577
	3 Q02	6	0,361228	0,458365
	4 Q03	18	1,083685	1,375095
	5 Q04	22	1,324503	1,680672
	6 Q05	19	1,143889	1,45149
	7 Q06	2	0,120409	0,152788
Spolu		1661		

Ďalej ukážeme, ako sa zmení výsledok spracovania po zlúčení kódov zo 4-znakového kódovania na 3-znaky, podľa MKCH-10 pri spracovaní v SPSS module Tables. Vidíme, že vo výstupe je o 50 VVCh menej. Rovnako dopadne spracovanie v module Tables aj po rekódovaní na numerický kód, ako vidno z ďalšej tabuľky.

Výpočet v module Tables s alfanumerickým kódovaním po zlúčení kódov:

Počet odpovedí 1611!!!

		SR	
		Count	Column N %
$d		**1309**	100
	E03	1	0,076394
	E70	1	0,076394
	E84	1	0,076394
	Q00	6	0,458365
	Q01	4	0,305577
	Q02	6	0,458365
	Q03	18	1,375095
	Q04	21	1,604278
	Q05	18	1,375095
	Q06	2	0,152788
	Q07	4	0,305577
	Q10	2	0,152788

Výpočet v module Tables s numerickým kódovaním po zlúčení kódov:

Počet odpovedí 1611!!!

		SR	
		Count	Column N %
$diag	1 Q00	6	0,458365
	2 Q01	4	0,305577
	3 Q02	6	0,458365
	4 Q03	18	1,375095
	5 Q04	21	1,604278
	6 Q05	18	1,375095
	7 Q06	2	0,152788
	8 Q07	4	0,305577
	9 Q10	2	0,152788

V module Tables sa teda „stratia“ duplicitné diagnózy. Keď ale použijeme rekódovanie na numerický kód a spracovanie realizujeme procedúrou Multiple Response, tak je výsledok správny – viď tabuľka:

Výpočet procedúrou Muptiple Response v Base module numerickým kódovaním po zlúčení kódov: **Počet odpovedí 1661!!!**

$diag Frequencies				
		Responses		Percent of Cases
	1309	N	Percent	N
$diag(a)	1 Q00	6	0,361228	0,458365
	2 Q01	4	0,240819	0,305577
	3 Q02	6	0,361228	0,458365
	4 Q03	18	1,083685	1,375095
	5 Q04	22	1,324503	1,680672
	6 Q05	19	1,143889	1,45149
	7 Q06	2	0,120409	0,152788
	8 Q07	4	0,240819	0,305577
	9 Q10	2	0,120409	0,152788

Transformáciou na 0/1 premenné pri pôvodnom kódovaní dostaneme veľa premenných, ktoré nám dovoľujú korektné spracovanie. Treba sa ale vyvarovať transformácii na 0/1 premenné po zlúčení kódov – taktiež sa pri spracovaní „stratia“ duplicitné kódy.

4. Záver

Záznam dát o výskyte VVCh v Hlásení vrodených chýb by bolo vhodné zjednotiť a používať prednostne 4-znakový kód choroby podľa MKCH-10. Tým by sme získali spresnené dáta s možnosťou ich následného zlúčenia.

Pri štatistickom spracovaní multireponse premenných, ktoré sú zaznamenávané alfanumerickým kódom, je možné ich štatistické spracovanie v prostredí SPSS, v module Tables a po rekódovaní na numerický kód aj v procedúrach Multiple Response v Base module.

V prípade potreby zlúčenia kódov nemožno používať na spracovanie modul Tables, ale spracovanie je možné v procedúrach Multiple Response v Base module, s využitím numerického rekódovania.

Riešenie záznamu a štatistického spracovania multiresponse premenných sa týka aj iných registrov - ako napr. Národný register pacientov s vrodenou chybou srdca; Národný register pacientov s cievnym ochorením; Národný register pacientov s chronickým ochorením pľúc a ďalších a prináša spresnenie pre všetky podobné prípady. V príspevku navrhované riešenia prispejú aj k správnemu štatistickému spracovaniu dát z takýchto registrov, ako aj podobných zdravotných záznamov.

5. Literatúra

[1]Böhmer, D., Šticová, A. (1997): Priestorová diferencovanosť územia Východného Slovenska z hľadiska výskytu vrodených vývojových chýb, 6. Demografická konferencia Slovenskej štatistickej a demografickej spoločnosti: Pôrodnosť a vybrané aspekty reprodukcie obyvateľstva, Domaša, 1997, Zborník prednášok in extenso, str. 118-123.

[2]Böhmer, D. (2002): Autoreferát dizertačnej práce, Ústav lekárskej biológie a genetiky, Lekárska fakulta UK v Bratislave, Bratislava, 2002.

[3]Böhmer, D. (2003): Samovoľné reprodukčné straty a vrodené vývojové chyby ako súčasť reprodukčnej biológie človeka, habilitačné práca, Bratislava, 2003, s. 177.

[4]Böhmer, D. (2004): Analýza výskytu vrodených vývojových chýb u živonarodených novorodencov v Slovenskej republike, Gynekológia pre prax, roč. 2, č. 1, 2004, s. 34-36.

[5]Böhmer, D., Luha J., Martináková J. (2009): Distribúcia vrodených vývojových chýb na Slovensku v rokoch 1997 – 2005. FORUM STATISTICUM SLOVACUM 5/2009. SŠDS Bratislava 2009. ISSN 1336-7420.

[6]Luha J.(2003): Matematickoštatistické aspekty spracovania dotazníkových výskumov. Štatistické metódy vo vedecko-výskumnej práci 2003, SŠDS, Bratislava 2003, ISBN 80-88946-32-8.

[7]Luha J.: (2008) Metodologické aspekty zberu a záznamu dát otázok s možnosťou viac odpovedí. FORUM STATISTICUM SLOVACUM 7/2008. SŠDS Bratislava 2008. ISSN 1336-7420.

[8]Luha J. (2009): Korelácia javov. FORUM STATISTICUM SLOVACUM 2/2009. SŠDS Bratislava 2009. ISSN 1336-7420.

[9]Martináková J. (2009): Časová a priestorová diferenciácia vrodených vývojových chýb na Slovensku. Diplomová práca. Prírodovedecká fakulta UK Bratislava 2009.

[10] Manuály SPSS for Windows ver. 15.

[11] Medzinárodná klasifikácia chorôb - MKCH-10.

[12] Zdravotnícka štatistika, Vrodené chyby na Slovensku v roku 2008. Národné centrum zdravotníckych informácií a štatistiky v Bratislave. http://www.nczisk.sk/buxus/generate_page.php?page_id=555

Adresa autora (-ov):

Daniel Böhmer, Doc., MUDr., PhD.
Ústav lekárskej biológie, genetiky a klinickej genetiky LF UK a FNsP Bratislava
daniel.bohmer@fmed.uniba.sk

Daniela Brašeňová, PhDr.
Národné centrum zdravotníckych informácií Bratislava
daniela.brasenova@nczisk.sk

Ján Luha, RNDr., CSc.
Ústav lekárskej biológie, genetiky a klinickej genetiky LF UK a FNsP Bratislava
jan.luha@fmed.uniba.sk

Slovensko a jeho predpoklady pre úspešný prenos poznatkov a technológií z vedy do praxe
Capacity of Slovakia to advance technology transfer and commercialisation of research

Denisa Brighton

Abstract: Knowledge and technology transfer and collaboration with industries is not a new phenomenon at Slovak universities; however the number of registered patents and licences remains low, well below the EU average. There has been limited interest in protecting intellectual property and fully exploiting intellectual property rights. Existing university systems do not motivate or professionally support their research staff to work with businesses and other organisations in order to sell their expertise. Modern universities have been under pressure since the last quarter of the previous century to respond to the needs of the knowledge-based industries, social and economic needs of today societies and at the same time come up with new knowledge and know-how. The Slovak government needs to truly recognise this inevitable trend by formulating a new innovation strategy and action plans fully aligned with the Lisbon agenda.

Key words: Technology transfer, knowledge, transfer, competitiveness, knowledge-based industry, commercialisation, innovation, intellectual property, university collaboration, entrepreneurial university

Kľúčové slová: inovačná spoločnosť, ochrana duševného vlastníctva, prenos poznatkov do praxe, technologický transfer, znalostná ekonomika, inovácie, konkurencieschopnosť

1. Úvod

Väčšina univerzít a výskumných pracovísk v rámci EU15 si založila vlastnú kanceláriu pre prenos poznatkov či technológií do praxe, aby tak rozšírili svoju pôsobnosť a poskytli služby širšej spoločnosti. Kolískou týchto aktivít boli Spojené štáty americké a ich vznik sa datuje skončením druhej svetovej vojny, keď bola nevyhnutná potreba využiť výdobytky vedy na rýchle budovanie novej industriálnej spoločnosti. Americká vláda sa zaviazala dlhodobo financovať základný výskum čím vytvorila silné kapacity pre ďalší rozvoj.

Prenos poznatkov a technológií do praxe alebo inými slovami technologický transfer je proces prenášania poznatkov vedy z jednej organizácie do druhej tak, aby ich sa ich využitie zrealizovalo v praxi. Toto si často vyžaduje veľké investície, ktorými ani univerzity ani verejné výskumné pracoviská nedisponujú, a preto spolupracujú s investormi, podnikateľmi alebo aj inými štátnymi či verejnými organizáciami, ktoré umožnia ďalší rozvoj produktov a technológií, prípadne ich priamo uvedú na trh. Úspešnosť technologického transferu a realizácie inovácií na univerzitách za posledné roky mimoriadne stúpla, a tak i o profesionálne služby kancelárií pre technologický transfer, ktoré podporujú vedcov a výskumníkov, je zvýšený záujem. Ich úsilie prispieva k zvyšovaniu konkurencieschopnosti zainteresovaných podnikov a k rozvoju celej miestnej ekonomiky.

2. Slovensko v meradle európskych inovačných ukazovateľov

Slovensko v tejto oblasti zatiaľ zaostáva, pretože verejné výskumné inštitúcie a univerzity zatiaľ neposkytujú systematickú odbornú pomoc na ochranu duševného vlastníctva a jeho následného využitia v priemyselnej a spoločenskej praxi. Niektoré slovenské inštitúcie

však už začínajú pracovať na tvorbe balíčkov služieb a ich aktivity sú väčšinou čiastkovo financované prostredníctvom štrukturálnych fondov z Operačného programu Výskum a vývoj. Zo skúseností na zahraničných univerzitách je známe, že etablovanie centier pre podporu technologického transferu je dlhodobou záležitosťou, ktorá si vyžiada tri až päť ročné úsilie a značné počiatočné finančné investície.
Prieskum úspešnosti centier pre technologický transfer v rámci Európy prebieha viac ako tri roky a je založený na zbere a vyhodnocovaní siedmich vybraných indikátorov:

1. počet vynálezov
2. počet patentových žiadostí
3. počet udelených patentov
4. počet novozaložených podnikov
5. počet licenčných zmlúv
6. príjem z predaných licencií
7. počet zmlúv na realizáciu výskumu

Údaje však nie sú celkom porovnateľné, existujú odchýlky medzi definíciami niektorých indikátorov a zber dát v jednotlivých krajinách sa neuskutočňuje v rovnakom čase. Mnohé krajiny, ktoré sú lídrami vo výskume ako napr. Švédsko, Nemecko, Rakúsko a Fínsko, nerealizujú vlastný národný prieskum, údaje zbierajú európske asociácie ako ProtTon a ASTP. Informácie o technologickom transfere v dvanástich nových členských štátoch EÚ sú mimoriadne nevýrazné a žiadna slovenská inštitúcia sa do prieskumu zatiaľ nezapojila[1]. V súčasnosti nie je teda možné porovnať Slovensko s ostatnými krajinami na základe týchto prieskumov.

Inovačnú výkonnosť Slovenska je možné posúdiť podľa systému inovačných indikátorov uvedených v *European Innovation Scoreboard 2008*[2] (Európska hodnotiaca tabuľka inovácií) a štatistickej príručky *Science, Technology and Innovation in Europe 2009*[3] (Veda, technológie a inovácie 2009). Tabuľka č.1. zobrazuje zoznam inovačných indikátorov, ktoré sú rozdelené do kategórií. Každá kategória uvádza zdroj primárnych dát a rok ich zberu. Na základe týchto údajov je vyčíslený súhrnný inovačný index (SII) pre každú krajinu EÚ a tento predstavuje inovačnú výkonnosť členskej krajiny.

Hlavná štatistická analýza vypracovávaná zo súhrnného inovačného indexu už v predchádzajúcich piatich rokoch zadeľuje členské štáty do štyroch kategórií:

1. "Inovační lídri" (mimoriadna inovačná výkonnosť) – Fínsko, Nemecko, Švédsko, Švajčiarsko a Veľká Británia
2. „Nasledovníci" (nad priemerom EU27) – Rakúsko, Belgicko, Francúzsko, Írsko, Luxembursko a Holandsko
3. „Mierni inovátori" (inovačná výkonnosť pod priemerom EU27) – Cyprus, Česká republika, Estónsko, Grécko, Island, Taliansko, Nórsko a Portugalsko
4. „Dobiehajúce krajiny" (inovačná výkonnosť hlboko pod priemerom EU27 avšak táto sa postupne približuje k priemeru EU27 s výnimkou Litvy a Chorvátska) – Bulharsko, Maďarsko, Lotyšsko, Malta, Poľsko, Slovensko, Rumunsko, Chorvátsko a Litva

[1] Metrics for Knowledge Transfer from Public Research Organisations in Europe, European Commission 2009
[2] European innovation scoreboard 2008 - Comparative analysis of innovation performance, European Commission 2009
[3] Science, Technology and Innovation in Europe 2009, Eurostat Statistical Books, European Commission 2009

Tabuľka 1: Indikátory Európskej hodnotiacej tabuľky inovácií - EIS 2008-2010

EIS položka	**Zdroj dát, rok zberu**
VSTUPY	
Ľudské zdroje	
1.1.1 Počet absolventov vysokých škôl na 1000 obyvateľov vo veku 20 - 29r.	Eurostat (2006)
1.1.2 Počet absolventov doktorandského štúdia na 1000 obyvateľov vo veku 25 - 64r.	Eurostat (2006)
1.1.3 Počet absolventov vysokých škôl na 100 obyvateľov vo veku 25 - 64r.	Eurostat (2007)
1.1.4 Účasť na celoživotnom vzdelávaní na 100 obyvateľov vo veku 25 - 64r.	Eurostat (2007)
1.1.5 Dosiahnutá úroveň vzdel. mládeže-počet obyv. vo veku 20 - 24r. s ukončeným stredošk. vzdel.	Eurostat (2007)
Financovanie a podpora	
1.2.1 Verejné výdavky na výskum a vývoj (% z HDP)	Eurostat (2007)
1.2.2 Rizikový kapitál investovaný do výskumu a vývoja (% z HDP)	EVCA/Eurostat (2007)
1.2.3 Podnikateľské pôžičky (vo vzťahu k HDP)	IMF (2007)
1.2.4 Prístup podnikov k širokopásmovému internetu (% podnikov)	Eurostat (2007)
PODNIKATEĽSKÉ AKTIVITY a INOVÁCIE	
Podnikateľské investície	
2.1.1 Náklady podnikov na výskum a vývoj (% z HDP)	Eurostat (2007)
2.1.2 Náklady na IKT (% z HDP)	EITO/Eurostat (2006)
2.1.3 Náklady na inovácie netýkajúce sa výskumu a vývoja (% z celkového obratu)	Eurostat (2006)
Spolupráca a podnikanie	
2.2.1 Vnútropodnikové inovácie malých a stredných podnikov (MSP) (% MSP)	Eurostat (2006)
2.2.2 Inovatívne MSP kooperujúce s inými (% MSP)	Eurostat (2006)
2.2.3 Zmeny v počte MSP (počet vznikov a zánikov MSP) (% MSP)	Eurostat (2005)
2.2.4 Publikácie vydané v spolupráci medzi súkromným a verejným sektorom na 1 mil. obyv.	Thomson Reuters/ CWTS (2006)
Výkony	
2.3.1 Európske (EPO) patenty na 1 milión obyvateľov	Eurostat (2005)
2.3.2 Obchodné značky na úrovni EÚ na 1 milión obyvateľov	OHIM/ Eurostat (2007)
2.3.3 Nové dizajny na úrovni EÚ na 1 milión obyvateľov	OHIM/ Eurostat (2007)
2.3.4 Bilancia platob. tokov týkajúcich sa nových technológií (platby a tržby za licencie) (% z HDP)	World Bank (2006)
VÝSTUPY	
Inovačné podniky	
3.1.1 MSP zavádzajúce nové produkty alebo technológie (% MSP)	Eurostat (2006)
3.1.2 MSP zavádzajúce marketingové alebo organizačné inovácie (% MSP)	Eurostat (2006)
3.1.3 Inovátori z oblasti efektívneho využívania zdrojov, nevážený priemer z:	
• Podielu inovatív. podnikov, kt. významne znížili náklady na zamestnancov (% podnikov)	Eurostat (2006)
• Podielu inovatív. podnikov, kt. významne znížili náklady na použ. materiálov a energií (% podnikov)	Eurostat (2006)
Dopad na ekonomiku	
3.2.1 Zamestnanosť v stredne a vysoko technickej *high-tech* výrobe (% zamestnancov)	Eurostat (2007)
3.2.2 Zamestnanosť v službách vyžadujúcich si vysoké znalosti (% zamestnancov)	Eurostat (2007)
3.2.3 Export *high-tech* produktov (% z celkového objemu exportu)	Eurostat (2006)
3.2.4 Export služieb vyžadujúcich si vysoké znalosti (% z celkového objemu exportu služieb)	Eurostat (2006)
3.2.5 Predaj produktov, ktoré sú na trhu novinkami (% z celkového obratu)	Eurostat (2006)
3.2.6 Predaj nových produktov, ktoré podniky predtým nepredávali (% z celkového obratu)	Eurostat (2006)

Tabuľka 2: Hodnoty EIS indikátorov pre SR, EU 27, min a max.

EIS položka	Min.	SR	EU27	Max.	Rovnomerné normovanie
VSTUPY					
Ľudské zdroje					
1.1.1 Počet absolventov VŠ na 1000 obyv.	12,60	24,40	40,30	62,10	0,24
1.1.2 Počet absolventov dokt. štúdia na 1000 obyv.	0,03	0,89	1,11	2,75	0,32
1.1.3 Počet absolventov VŠ na 100 obyv.	9,70	14,40	23,50	36,40	0,18
1.1.4 Účasť na celoživ. vzdelávaní na 100 obyv.	1,30	3,90	9,70	32,00	0,08
1.1.5 Dosiahnutá úroveň vzdelania mládeže	46,40	91,30	78,10	94,60	0,93
Financovanie a podpora					
1.2.1 Verejné výdavky na V a V (% z HDP)	0,21	0,27	0,65	1,26	0,06
1.2.2 Rizikový kapitál investovaný do V a V (% z HDP)	0,01	0,01	0,11	0,48	0,00
1.2.3 Podnikateľské pôžičky (vo vzťahu k HDP)	0,26	0,42	1,31	2,47	0,07
1.2.4 Prístup podnikov k širokopásm. internetu (%)	37,00	76,00	77,00	95,00	0,67
PODNIKATEĽSKÉ AKTIVITY a INOVÁCIE					
Podnikateľské investície					
2.1.1 Náklady podnikov na výskum a vývoj (% z HDP)	0,10	0,18	1,17	2,64	0,03
2.1.2 Náklady na IKT (% z HDP)	1,20	2,50	2,70	3,80	0,50
2.1.3 Náklady na inovácie netýkajúce sa V a V (% z obratu)	0,16	1,51	1,03	3,36	0,42
Spolupráca a podnikanie					
2.2.1 Vnútropodnikové inovácie MSP (% MSP)	15,10	17,90	30,00	46,30	0,09
2.2.2 Inovatívne MSP kooperujúce s inými (% MSP)	2,90	7,20	9,50	27,50	0,17
2.2.3 Zmeny v počte MSP(% MSP)	0,70	4,80	5,10	10,30	0,43
2.2.4 Publikácie vydané v spolupráci na 1 milión obyv.	0,0	4,50	31,40	193,1	0,02
Výkony					
2.3.1 Európske (EPO) patenty na 1 milión obyv.	0,70	5,80	105,70	411,10	0,01
2.3.2 Obchodné značky na úrovni EÚ na 1 milión obyv.	1,90	20,60	124,60	1220,00	0,02
2.3.3 Nové dizajny na úrovni EÚ na 1 milión obyv.	2,60	18,00	121,80	1018,60	0,02
2.3.4 Technologická platobná bilancia (% z HDP)	0,08	0,43	1,07	9,92	0,04
VÝSTUPY					
Inovačné podniky					
3.1.1 MSP zavádzajúce nové produkty alebo technológie (%)	14,40	21,40	33,70	52,90	0,18
3.1.2 MSP zavádzajúce marketing. či organizač. inovácie (%)	15,70	21,50	40,00	68,10	0,11
3.1.3 Inovátori z oblasti efektívneho využívania zdrojov:			–		
• Podiel podnikov, kt. znížili náklady na zamestnancov (%)	6,20	8,00	18,00	34,90	0,06
• Podiel podnikov, kt. znížili náklady na mat. a energie(%)	4,30	10,80	9,60	20,70	0,40
Dopad na ekonomiku					
3.2.1 Zamestnanosť v s *high-tech* výrobe (% zamestnancov)	0,90	9,89	6,69	10,85	0,90
3.2.2 Zamestnanosť v službách náročných na znalosti (%)	5,26	9,86	14,51	23,94	0,25
3.2.3 Export *high-tech* produktov (% z exportu)	11,40	57,20	48,10	74,50	0,73
3.2.4 Export služieb vyžadujúcich si vysoké znalosti (%)	8,90	20,80	48,70	82,40	0,16
3.2.5 Predaj produktov, kt. sú na trhu novinkami (% z obratu)	1,61	7,79	8,60	24,79	0,27
3.2.6 Predaj predtým nepredávaných produktov (% z obratu)	1,25	8,95	6,28	13,69	0,62

Tabuľka 3 - Hodnoty EIS indikátorov usporiadané podľa dosiahnutej hodnoty normovanej premennej

EIS položka	Min.	SR	EU27	Max.	Rovnom. Norm.	Krajina EU (Min.)	Krajina EU (Max.)
1.1.5 Dosiahnutá úroveň vzdelania mládeže	46.40	91.30	78.10	94.60	0.93	Turecko	Chorvátsko
3.2.1 Zamestnanosť v s *high-tech* výrobe (% zamestnancov)	0.90	9.89	6.69	10.85	0.90	Cyprus	Dánsko
3.2.3 Export *high-tech* produktov (% z exportu)	11.40	57.20	48.10	74.50	0.73	Nórsko	Malta
1.2.4 Prístup podnikov k širokopásm. internetu (%)	37.00	76.00	77.00	95.00	0.67	Rumunsko	Island
3.2.6 Predaj nových predtým nepred. produktov(% z obratu)	1.25	8.95	6.28	13.69	0.62	Rumunsko	Litva
2.1.2 Náklady na IKT (% z HDP)	1.20	2.50	2.70	3.80	0.50	Grécko	Švédsko
2.2.3 Zmeny v počte MSP (% MSP)	0.70	4.80	5.10	10.30	0.43	Fínsko	Veľká Brit.
2.1.3 Náklady na inovácie netýkajúce sa V a V (% z obratu)	0.16	1.51	1.03	3.36	0.42	Turecko	Estónsko
• Podiel podnikov, kt. znížili náklady na mat. a energie(%)	4.30	10.80	9.60	20.70	0.40	Nórsko	Grécko
1.1.2 Počet absolventov dokt. štúdia na 1000 obyv.	0.03	0.89	1.11	2.75	0.32	Malta	Portugalsko
3.2.5 Predaj produktov, kt. Sú novinkami (% z obratu)	1.61	7.79	8.60	24.79	0.27	Nórsko	Malta
3.2.2 Zamestnanosť v službách náročných na znalosti (%)	5.26	9.86	14.51	23.94	0.25	Rumunsko	Luxemb.
1.1.1 Počet absolventov VŠ na 1000 obyv.	12.60	24.40	40.30	62.10	0.24	Turecko	Írsko
3.1.1 MSP zavádzajúce nové produkty al. technológie (%)	14.40	21.40	33.70	52.90	0.18	Litva	Švajčiarsko
1.1.3 Počet absolventov VŠ na 100 obyv.	9.70	14.40	23.50	36.40	0.18	Turecko	Fínsko
2.2.2 Inovatívne MSP kooperujúce s inými (% MSP)	2.90	7.20	9.50	27.50	0.17	Rumunsko	Fínsko
3.2.4 Export služieb vyžadujúcich si vysoké znalosti (%)	8.90	20.80	48.70	82.40	0.16	Veľká Brit.	Luxemb.
3.1.2 MSP zavádzajúce market. alebo organiz. inovácie (%)	15.70	21.50	40.00	68.10	0.11	Bulharsko	Nemecko
2.2.1 Vnútropodnikové inovácie MSP (% MSP)	15.10	17.90	30.00	46.30	0.09	Bulharsko	Nemecko
1.1.4 Účasť na celoživ. vzdelávaní na 100 obyv.	1.30	3.90	9.70	32.00	0.08	Rumunsko	Švédsko
1.2.3 Podnikateľské pôžičky (vo vzťahu k HDP)	0.26	0.42	1.31	2.47	0.07	Rumunsko	Írsko
• Podiel podnikov, kt. znížili náklady na zamestnancov (%)	6.20	8.00	18.00	34.90	0.06	Litva	Francúzsko
1.2.1 Verejné výdavky na V a V (% z HDP)	0.21	0.27	0.65	1.26	0.06	Malta	Island
2.3.4 Technologická platobná bilancia (% z HDP)	0.08	0.43	1.07	9.92	0.04	Litva	Írsko
2.1.1 Náklady podnikov na výskum a vývoj (% z HDP)	0.10	0.18	1.17	2.64	0.03	Cyprus	Švédsko
2.2.4 Publikácie vydané v spolupráci na 1 milión obyv.	0.00	4.50	31.40	193.10	0.02	Litva	Švajčiarsko
2.3.1 Európske (EPO) patenty na 1 milión obyv.	0.70	5.80	105.70	275.00	0.02	Rumunsko	Nemecko
2.3.2 Obchodné značky na úrovni EÚ na 1 milión obyv.	1.90	20.60	124.60	1220.00	0.02	Turecko	Luxemb.
2.3.3 Nové dizajny na úrovni EÚ na 1 milión obyv.	2.60	18.00	121.80	1018.60	0.02	Litva	Luxemb.
1.2.2 Rizikový kapitál investovaný do V a V (% z HDP)	0.01	0.01	0.11	0.48	0.00	Slovensko	Veľká Brit.

Výrazné slabiny Slovenska vidieť aj v dostupnosti rizikového kapitálu a podnikateľských pôžičiek. V rozvinutých krajinách na nich stojí komercializácia poznatkov a technológií, ktorá si mnohokrát vyžaduje relatívne vysoké počiatočné vklady, takže ak sa investície v týchto oblastiach nezvýšia, aktivity budú jednoznačne stagnovať alebo aj upadať. Vnútropodnikové inovácie sú takmer na minime, výkonnosťou sa tu Slovensko približuje k najslabšiemu Bulharsku. Väčšina nákladov v podnikateľskom sektore bola nasmerovaná rozvoj nových služieb a nie do výroby, z čoho vyplýva, že výskum a vývoj väčších spoločností pôsobiacich na Slovensku prebieha v iných krajinách.

Štatistiky tiež poukazujú na veľké rozdiely v inovačnej výkonnosti v rámci jednotlivých krajín. Takmer polovica výdajov na výskum a vývoj bola investovaná v Bratislavskom kraji, ktorý zamestnáva takmer polovicu všetkých slovenských vedcov.

Povzbudivé sú náklady na informačné a komunikačné technológie a náklady na inovácie nesúvisiace s výskumom a vývojom. Dobré výstupy sú zaznamenané aj v počte *high-tech* exportov a zamestnanosti v *high-tech* službách. Je však nutné poznamenať, že väčšina týchto štatistík je z roku 2006, a teda nereflektuje dôsledky globálnej hospodárskej krízy, ktorá začala sa v roku 2008.

Počet EPO patentov a ostatných foriem ochrany duševného vlastníctva je na Slovensku minimálny, priemer EU27 v EPO patentoch je viac ako 18-krát vyšší, a Švajčiarsko ako najvýkonnejšia krajina, malo v roku 2005 až 70-krát viac patentov. Slabý záujem slovenských vedcov o efektívnu ochranu duševného majetku medziiným súvisí aj s nedostatočným chápaním jeho dôležitosti a tiež vysokými nákladmi, ktoré sú s celým procesom spojené. Možnosti jeho využitia môžu byť takisto širšie ako by sa na prvý pohľad zdalo a ich zhodnotenie si vyžaduje posudok niekedy aj viacerých znalcov. Malé a stredné podniky sa spolupráci nebránia a aj z praxe je zrejmé, že uvítajú spoluprácu s inými

organizáciami a profesionálne riadenými centrami pre technologický transfer ukotvených na vedecko-výskumných inštitúciách.

3. Záver

Analýza ukazovateľov inovačnej výkonnosti v rámci krajín EU27 potvrdzuje, že Slovensko má v súčasnosti nedostatočne rozvinuté inovačné kapacity, vláda neinvestuje potrebné prostriedky do vývoja, výskumu a rozvoja infraštruktúry a trpia predovšetkým vysoké školy. Súkromný sektor je v investíciách tiež na chvoste a jedným z východísk na zlepšenie spolupráce na realizáciu aplikovaného výskumu je vybudovanie potrebného *know-how* na univerzitách a štátnych výskumných pracoviskách. Na to bude nutná systematická podpora počas troch až piatich rokov a počiatočné nenávratné investície, na ktoré sa nebude viazať bremeno neprimeranej administratívy a obmedzení spojených so štrukturálnymi fondmi. Musí existovať efektívna spolupráca v rámci celej štátnej a verejnej sféry a fondy EU by mali mať doplnkový charakter.

Podporu si budú vyžadovať aj služby pre podnikateľov, ktoré by mali byť navzájom prepojené - napr. technologické inkubátory pre *start-up* firmy a spoločnosti *spin-off*, vedecko-technické parky, priemyselné *klastre* a škála nástrojov pre finančnú podporu a stimuláciu sieťovania. Zlepšenie inovačného profilu Slovenska významne prispeje k rozvoju celej znalostnej ekonomiky a tým aj k napĺňaniu Lisabonskej stratégie.

Literatúra:

[1]METRICS FOR KNOWLEDGE TRANSFER FROM PUBLIC RESEARCH ORGANISATIONS IN EUROPE, EUROPEAN COMMISSION 2009

[2]EUROPEAN INNOVATION SCOREBOARD 2008 - COMPARATIVE ANALYSIS OF INNOVATION PERFORMANCE, EUROPEAN COMMISSION 2009

[3]SCIENCE, TECHNOLOGY AND INNOVATION IN EUROPE 2009, EUROSTAT STATISTICAL BOOKS, EUROPEAN COMMISSION 2009

[4]CHAJDIAK, J.: ŠTATISTIKA V EXCELI 2007. BRATISLAVA, STATIS 2009

[5]www.astp.net, WEB STRÁNKA ASSOCIATION OF EUROPEAN SCIENCE AND TECHNOLOGY TRANSFER PROFESSIONALS

[6]www.autm.net, WEB STRÁNKA ASSOCIATION OF UNIVERSITY TECHNOLOGY MANAGERS V USA

Adresa autora:

Denisa Brighton, Ing.
Know-How Centrum STU
Vazovova 5, 812 43 Bratislava
e-mail: denisa.brighton@stuba.sk

Metóda párového porovnávania a brainstormingu pri manažérskom rozhodovaní vo výrobných procesoch
Pairwise comparison and brainstorming in the management decision - making in manufacturing processes

Jaroslav Dvorský, Marián Majerík, Martin Ešše

Abstract: This paper deals with pairwise comparison and brainstorming. It explains what is pairwise comparison and brainstorming and how do they work. In this article is shown how these mentioned methods can be used. The example, used in this article, presents pairwise comparison and brainstorming, when you have to make a decision which process is the most critical.

Key words: pairwise comparison, brainstorming, process, decision - making

Kľúčové slová: párové porovnávanie, brainstorming, proces, rozhodovanie

1. Úvod

Metódou manažérstva kvality sa rozumie cieľavedomý, premyslený a sústavný postup k riešeniu problémov spojených so zabezpečením a zlepšovaním kvality. Medzi tvorivé metódy patria napríklad: brainstorming a brainwriting, delfská metóda, medzi porovnávacie metódy patrí napríklad metóda párového porovnávania podobnosti. Rôznorodosť procesov, súvisiacich s kvalitou procesov a výrobkov, vzdelanostná úroveň a kvalifikácia jednotlivcov a kolektívov a vyšších štruktúr riadenia kvality vo firmách, ale aj kultúrno-psychologicko-historické aspekty firmy a štátu, veľkosť firmy a zložitosť vyrábaného výrobku vyžadujú od vrcholového vedenia firmy hľadať vhodný systém nástrojov a metód pre zabezpečenie a zlepšovanie kvality. Preto je veľmi dôležité vedieť vhodne použiť danú metódu pre daný proces.

2. Párové porovnávanie

Párové porovnávanie má veľkú poznávaciu silu, ktorá umožňuje postupovať pri zlepšovaní kvality produkcie takým smerom, ktorý uznajú za vhodný všetky zainteresované osoby.

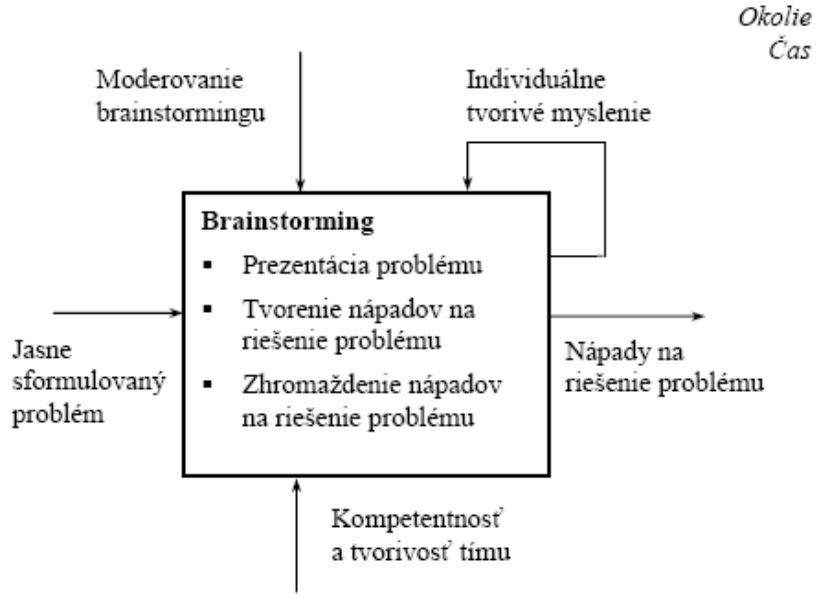
Obrázok 1: Znázornenie párového porovnávania

Ako možno vidieť z obrázku 1 výstupom z metódy párového porovnávania je zistenie závažnosti prvkov a názorový rozptyl pri skúmanej problematike. Vstupmi sú situácie a rôznymi charakteristikami a zainteresované strany. Pri tejto metóde musíme neustále dbať na účel porovnávania. Porovnávanie vykonáva tím odborníkov. Priebežne výsledky a poznatky sú použité pri určovaní závažnosti prvkov.

V prvom kroku postupu sa vykoná dekompozícia situácie na prvky, ktoré logicky súvisia s účelom porovnávania. V druhom kroku sa porovnáva každý prvok z každým, pričom sa ohodnotí nejakou škálou (napríklad od 1 do 10). V treťom kroku sa porovnávajú výsledky tak aby sme dokázali zistiť ich závažnosť. Výsledky môžu byť podporené grafmi alebo ďalšími štatistickými charakteristikami.

3. Brainstorming

Brainstorming je metóda hľadania nápadov na riešenie problémov, má široké použitie a v praxi sa realizuje rôznymi, často dosť odlišnými spôsobmi.

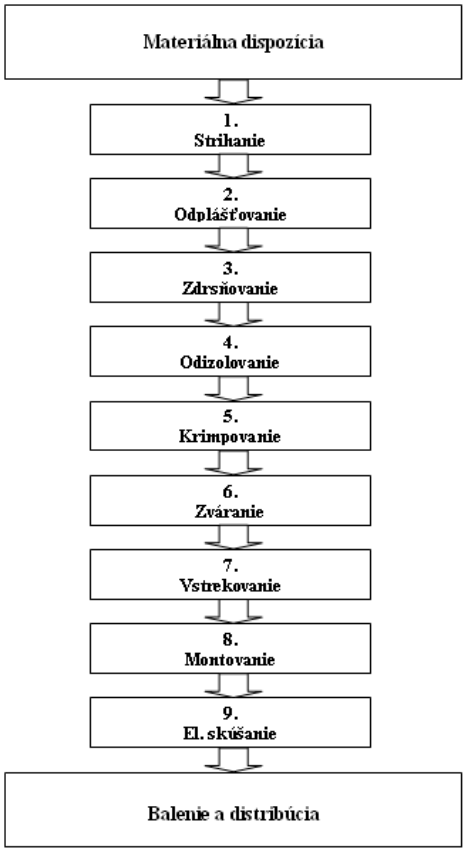


Obrázok 2: Znázornenie procesu brainstormingu

Ako možno vidieť z obrázku 2 výstupom z brainstormingu sú nápady pre riešenie problému. Než sa dostaneme k tomuto výsledku musíme jasne sformulovať problém. Riadiacim orgánom je pri tejto metóde moderátor, obmedzujúcim faktorom je kompetentnosť a tvorivosť tímu, individuálne tvorivé myslenie možno chápať ako priebežné výsledky, ktoré bude použité v ďalšom postupe.

4. Rozhodnutie o výbere najkritickejšieho procesu z pohľadu pretrhnutia žiliek vo vodiči

Pri analýze možnosti skúšania vodičov s rozpoznávaním prerušenia jednotlivých žiliek vo vnútri vodiča nás zaujímalo, kde vo výrobe by mohla byť najväčšia pravdepodobnosť prerušenia žiliek vo vnútri vodiča. Proces výroby káblových zväzkov sme rozdelili do 9 subprocesov (Obr. 3). Okrem poukázania na kritické procesy výroby potrebujeme vedieť tieto miesta aj kvôli zaradeniu jednotlivých metód skúšania, keďže nie všetky metódy sa používajú až na konci výrobného procesu.



Obrázok 3: Dekompozícia procesu výroba káblových zväzkov

Použili sme k takémuto ohodnoteniu metódu párového porovnávania, čo je veľmi jednoduchú a však podľa mnohých autorov vykonateľnú iba na základe potrebných skúseností a dostatočnej objektivity.

Tabuľku 1 sme vytvorili po dekompovaní procesu výroby káblových zväzkov (obrázok 3). Porovnávali sme každý proces s každých, pričom sme určovali ten významnejší z pohľadu možnosti pretrhnutia žilky vo vnútri vodiča.

Postupne sme porovnávali proces 1 so všetkými ostatnými, ak je pravdepodobnosť pretrhnutia žilky vo vodiči počas procesu 1 väčšia ako pri ostatnom porovnávanom procese priradili sme hodnotu 1.

Ak je pravdepodobnosť pretrhnutia žilky vo vodiči počas procesu 1 menšia ako pri ostatnom porovnávanom procese priradili sme hodnotu 0.

Ak je pravdepodobnosť pretrhnutia žilky vo vodiči pri procesoch rovnaká priradili sme hodnotu 0.5.

Ten istý postup sme opakovali pre všetky nasledovné procesy.

Takéto párové porovnávanie bolo vykonané na základe skúseností s výrobným procesom káblových zväzkov. Postupne sme zbierali informácie od operátorov z výroby. Pýtali sme sa kde vzniká prerušenie žilky vo vodiči. S týmito údajmi sme šli na stretnutie s manažérmi kvality. Pri tvorbe tabuľky sme používali zásady brainstormingu t.j.:

- Zákaz kritiky – nepoužívať kritické myslenie s tvorivým, aby nedochádzalo k tlmeniu aktivity účastníkov
- Inšpirácia – inšpirácia z predostretých námetov pre ďalších účastníkov rokovania
- Kvantita – využitie poznatku, že kvantita je predpokladom kvality
- Pohoda – vytvoriť priaznivú atmosféru počas rokovania

Tabuľka 2: Rozhodovacia matica výberu najkritickejšieho procesu z pohľadu možnosti pretrhnutia žiliek

Procesy	1	2	3	4	5	6	7	8	9	Suma	Poradie	Váha
1 Strihanie	-	0,5	1	0	0	1	0	1	1	4,5	**4**	0,125
2 Odplášťovanie	0,5	-	1	0	0	1	0	1	1	4,5	**4**	0,125
3 Zdrsňovanie	0	0	-	0	0	0	0	1	1	2	**7**	0,056
4 Odizolovanie	1	1	1	-	0	1	0	1	1	6	**3**	0,167
5 Krimpovanie	1	1	1	1	-	1	0	1	1	7	**2**	0,194
6 Zváranie	0	0	1	0	0	-	0	1	1	3	**6**	0,083
7 Vstrekovanie	1	1	1	1	1	1	-	1	1	8	**1**	0,222
8 Montovanie	0	0	0	0	0	0	0	-	1	1	**8**	0,028
9 El. skúšanie	0	0	0	0	0	0	0	0	-	0	**9**	0
Spolu:										36		1,000

Z uvedenej tabuľky vyplýva: najväčšia pravdepodobnosť pretrhnutia žiliek na základe skúseností je v procese 7 vstrekovanie, ďalej nasleduje proces 5 krimpovanie. Najmenšia pravdepodobnosť prerušenia žiliek je v procesoch 9 elektrické skúšanie, 8 montovanie a v procese 3 odizolovanie.

5. Záver

Vývoj metód pri manažérskom rozhodovaní sa stále vyvíja. Medzi najjednoduchšie a najprehľadnejšie patria metóda párového porovnávania a metóda brainstormingu. Obidve sú ľahko použiteľné pri rozhodovaní, kde potrebujeme zosúladiť viaceré názory a zároveň sa dopracovať k merateľnému výsledku. V tomto príspevku bolo demonštrované ako možno jednoducho použiť spomínané metódy pri výbere najkritickejšieho procesu z pohľadu prerušenia žiliek vo vodiči pri výrobe káblových zväzkov.

6. Literatúra

[1]ZGODAVOVÁ, K., LINCZÉNYI, A., NOVÁKOVÁ, R., SLIMÁK, I.: Profesionál kvality, Technická univerzita v Košiciach, 2002

[2]KMEŤ, S.: Manažment kvality, Žilina, Edis, 2004

[3]MARSH, J.: Nástroje kvality A – Z, Zvyšovanie kvality metódami Total Quality Management, AF, s.r.o., SR, 1996, ISBN 80-967022-2-X.

Adresa autora (-ov):

Jaroslav Dvorský, Ing.
Študentská 2
911 50 Trenčín
dvorsky@tnuni.sk

Martin Ešše, Ing.
Študentská 2
911 50 Trenčín
esse@tnuni.sk

Marián Majerík, Ing.
Študentská 2
911 50 Trenčín
marian.majerik@tnuni.sk

Kansei Engineering vo výučbe projektovania mechatronických systémov
Kansei Engineering in teaching of mechatronic systems planning

Martin Ešše, Marián Majerík, Jaroslav Dvorský

Abstract: This paper presents linking emotional values into product design through the methodology of Kansei Engineering. The last years has seen a growing interest in emotional design. Today products need to evoke the right emotions within the user to differentiate from other products. Emotional design is an area integrating design, engineering, mathematics, psychology and art. The first part of the paper will describe the Kansei Engineering methodology. In the second part are methods from Kansei Engineering applied to the robot LEGO Mindstorms NXT.

Key words: Kansei Engineerig, Kansei words, Robot Lego Mindstorms NXT

Kľúčové slová: Kansei Engineering, Kansei slová, Robot Lego Mindstorms NXT

1. Úvod

Dosahovanie konkurencieschopnosti mechatronických produktov na globálnom trhu si vyžaduje dôsledne uplatňovať efektívny prístup k ich navrhovaniu, realizovaniu, poskytovaniu zákazníkom, užívaniu aj recyklácii. Pružnosť podniku začína u zákazníka a schopnosti marketingu zistiť jeho požiadavky. Pružnosť sa prenáša do návrhu a vývoja mechatronických produktov, ktorá musí byť schopná rýchlo reagovať na tieto impulzy.

2. Kansei Engineering

Neexistuje presný výraz na vysvetlenie slova *"Kansei"*. Termín Kansei je úzko spojený s japonskou kultúrou, takže je zložité ho preložiť. Kansei je ľahšie poznateľný ako definovateľný. Aj pohľad na obraz evokuje v niektorých ľuďoch „príjemný pocit", ktorý je ťažko opísateľný. Toto je to o čom je Kansei Engineering. Kansei je individuálny dojem z určitého predmetu, prostredia alebo situácie, využívajúci všetky zmysli ako zrak, sluch, cit, čuch, chuť, pochopenie a ich rovnováha. Kansei vyjadruje význam slov [3],[4]: citlivosť – vnímavosť – estetickosť – pocity – emócie – náklonnosť – intuícia.

Jeden pocit v človeku môže následne vyvolať ďalší. Najčastejší spôsob merania Kansei je cez slová. Slová sú vonkajším prejavom ľudskej mysle. Prvky Kansei nemusia byť detailne obsiahnuté v slovách, pretože slová nedokážu popísať všetky naše emócie. Väčšina štúdii Kansei Engineeringu, ktorá bola publikovaná, využíva na meranie Kansei slová (slová vyjadrujúce dojmy z vlastností produktu). Kansei Engineering je teda metódou pre preklad pocitov a dojmov do vlastnosti a funkcií produktu (jej tvorcom je japonský profesor Mitsuo Nagamachi (1970)). Dokáže "merať" dojmy a ukazovať vzťahy k určitým produktovým vlastnostiam (y_{Kansei}**= f (vlastnosť)**). Vo všeobecnosti exituje niekoľko typov Kansei Engineeringu [4]:

- ***Kansei Engineering Typu I - Triedenie do skupín***

  Je to najrýchlejšia a najjednoduchšia metóda. Identifikuje sa produktová stratégia a segment trhu. Vytvorí sa stromová štruktúra identifikovaných zákazníkových emócií. Produktovévlastnosti sa potom spájajú s požiadavkami zákazníka.

- ***Kansei Engineering Typu II - Systém Kansei Engineering***

Matematické a štatistické nástroje sú použité na spojenie Kansei a požiadaviek zákazníka. Toto je počítačom riadený systém využívajúci databázu Kansei slov. Tento systém pozostáva z databázy obsahujúcej Kansei slová, obrázky, dizajnu a farby a znalostí o ich vzájomnom pôsobení.

- ***Kansei Engineering Typu III - Hybridný systém***

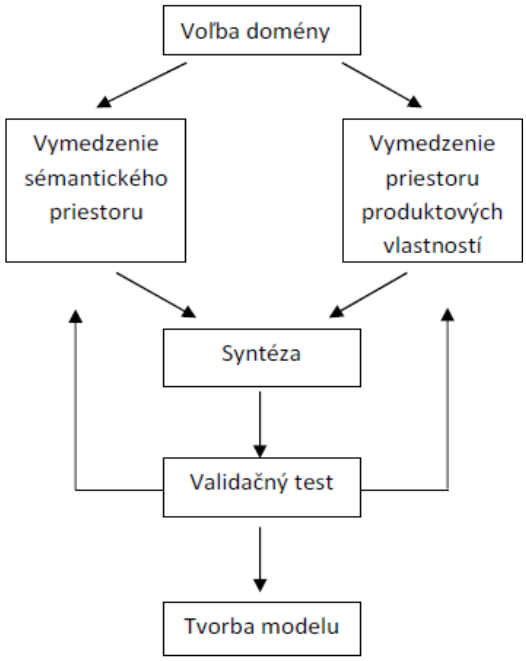
Podobá sa na predchádzajúci typ Kansei Engineering, ale tento môže predvídať. Návrhár môže pridať svoje myšlienky do databázy už prítomných Kansei slov.

- ***Kansei Engineering Typu IV - Kansei modelovanie***

Tento typ je založený na matematickom predvídaní, dokonca dokáže ohodnotiť ľudské pocity zo série slov.

- ***Kansei Engineering Typu V - Virtuálny Kansei***

Tento typ Kansei Engineeringu nahrádza reálny produkt s VR reprezentáciou integrovanej virtuálnej reality so štandardným systémom zbierania dát.



Obr.1: Model Kansei Engineeringu [4]

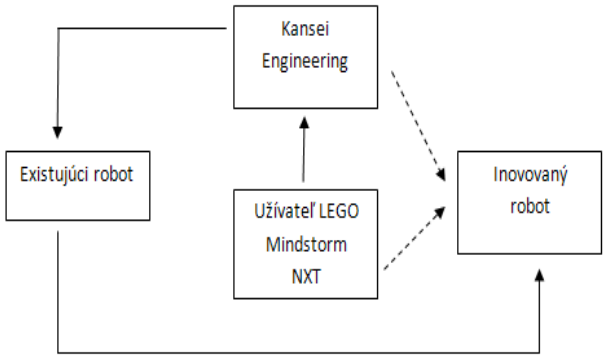
3. Využitie metódy Kansei Engineering pri didaktickej inovácii robota

Pokiaľ chceme získať inovovaný produkt, v našom prípade inovovaného robota, je potrebné, aby sa okrem vývojárov a ďalších zainteresovaných osôb zapojil aj „hlas“ zákazníka. Zákazník je podstatným generátorom tvorby myšlienok súvisiacich so všetkými aspektmi vlastností a funkcií robota, ktoré si mnohokrát samotní vývojoví pracovníci a konštruktéri neuvedomia pri vývoji a konštrukcii jednotlivých častí robota. Pri inovovaní robota LEGO Mindstorms NXT je potrebné poznať jeho účelový funkciu, jeho prvky (komponenty) a väzby medzi nimi. V našom prípade je záujem o jeho využitie vo výučbe, preto je potrebné v prvom kroku získať skupinu „Kansei slov“, ktoré budeme ďalej spracovávať. Identifikácia a postupnosť spracovania Kansei slov je znázornená na Obr. 3 a 5.

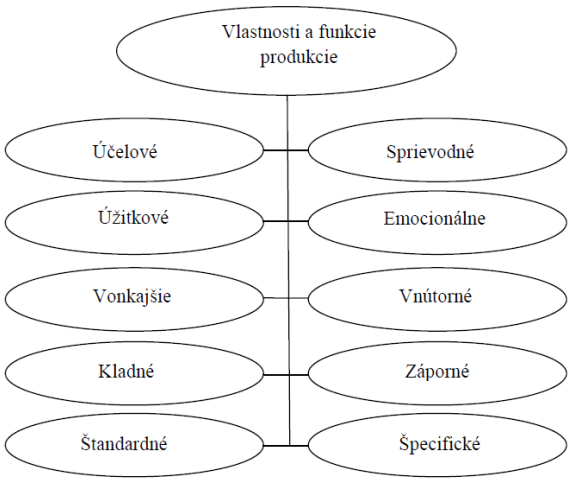
Z obrázku je jasne vidieť, že užívateľ si definuje skupinu Kansei slov vzťahujúcich sa na produkt, tie sa ďalej spracujú prostredníctvom metódy Kansei Engineering. Výsledkom je získanie inovovaného produktu podľa predstáv a požiadaviek zákazníka, v našom prípade robota pre účely výučby. Kansei slová je možné vyšpecifikovať z dotazníka, ktorý si ďalej rozpracujeme. Dotazník je uvedený na Obr.2, prostredníctvom ktorého sa nám podarilo vyšpecifikovať skupinu Kansei slov, ktoré sa stali vstupnými údajmi pre poznanie požadovaných vlastností inovovaného robota. V našom prípade je vývojovou organizáciou študent, ktorý pozná metodiku, ktorú skúša aplikovať na základe poznania požiadaviek zákazníka (– zadanie od vyučujúceho).

Dotazník	
1. Aké rozmery by mal mať robot?	
2. Akú hmotnosť by mal mať robot?	
3. Vyžadujete schopnosť robota manipulovať s predmetmi?	
4. Akým spôsobom by mal robot reagovať a vnímať prostredie?	
5. Akú dlhú dobu vyžadujete výdrž batérie?	
6. Vyžadujete možnosť preprogramovania robota?	
7. Vyžadujete bezdrôtové ovládanie robota? (napr. technológiou bluetooth alebo infraportom?)	
8. Vyžadujete, aby robot reagoval na slovné príkazy?	
9. Akej farby by mal byť robot?	
10. Aký materiál uprednostňujete použiť?	
11. V tejto časti môžete definovať Vaše ďalšie požiadavky...	

Obr.2: Dotazník pre kolekciu Kansei slov



Obr.3: Inovácia legorobota prostredníctvom poznania Kansei slov



Obr.4: Základné členenie vlastností a príznačných funkcií pre vytváranie charakteristík produkcie [2]

Pôvodná skupina slov	Skupina 1	Skupina 2	Kansei slová
Krehký Pevný Ľahký Organický Plastový Drsný Hladký Mäkký	Krehký Ľahký Plastový Pekný	Ľahký Pevný	ĽAHKÝ
Bezpečný Ostrý Spôsobilý Spoľahlivý	Bezpečný Ostrý Spoľahlivý	Bezpečný Spoľahlivý	BEZPEČNÝ
Výkonný	>>>	>>>	VÝKONNÝ
Premyslený Ekologický ……	Ekologický	Ekologický	EKOLOGICKÝ

Obr.5: Identifikácia Kansei slov

Dekompozícia stavebnice LEGO© Mindstorms NXT z hľadiska mechatroniky

Mechatronika sa zaoberá špecificky integrovaným prístupom k novonavrhovaným alebo konštruovaným výrobkom a integrovaným systémom. Na mechatroniku ako technickú disciplínu sa môžeme pozerať z dvoch strán [1]:

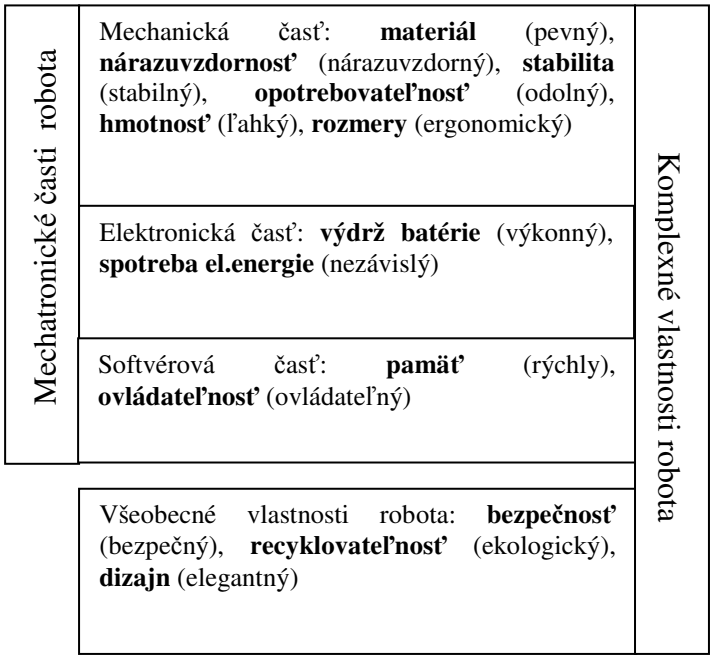
- Z pohľadu výroby – ako využívať kontinuitu výrobného procesu na základe poznatkov z mechatroniky pri návrhu a vlastnej výrobe - návrh mechatronického produktu,
- Z hľadiska užívateľského prístupu k výrobku - životný cyklus produktu.

Základné členenie vlastností a funkcií mechatronických systémov, z ktorých časť je využiteľná v metóde Kansei Engineering je znázornené na Obr.4. Samotnú stavebnicu LEGO Mindstorms NXT je možné dekomponovať z hľadiska prvkov a konštrukcie podľa mechatronického diagramu. Príznačnými funkciami mechatronických systémov sú tie jej väzby s daným okolím v určitom čase, ktorých existencia vyžaduje mechaniku, elektrotechniku a počítačové technológie. Tieto skutočnosti uvádzame a zdôrazňujeme preto, lebo tvoria základ nami uplatňovanej, na systémovom prístupe založenej paradigme kvality [2].

Finálnou skupinou Kansei slov sú:

• Ľahký	• Ovládateľný	• Odolný
• Bezpečný	• El.nezávislý	• Primeraný
• Výkonný	• Lacný	• Ergonomický
• Ekologický	• Kompatibilný	• Elegantný
• Spoľahlivý	• Nárazuvzdorný	
• Rýchly	• Stabilný	

Párovým porovnávaním jednotlivých Kansei slov sme stanovili ich závažnosť a priradili im jednotlivé vlastnosti pre plánovanie robota, ktoré môžu predstavovať vstupné informácie do ďalších metód (QFD, FMEA, ...) [5].



Obr.6: Vlastnosti potrebné pre posúdenie pre inováciu legorobota

4. Záver

Ľubovoľný produkt je kvalitný vtedy, pokiaľ ho za taký považuje aj jeho užívateľ. Tento fakt samozrejme platí aj v prípade mechatronických systémov, preto zavedenie metód a postupov pre vytváranie emocionálnych hodnôt z pohľadu užívateľa by malo mať svoje pevné miesto aj vo výučbe projektovania mechatronických zariadení. Tento článok by mal byť demonštratívnou ukážkou ako postupovať pri takýchto úlohách, samozrejme v oveľa širšom a konkrétnejšom riešení.

5. Literatúra

[1] MAIXNER, L.: Mechatronika. Computer Press, Brno, 2006. ISBN 80-251-1299-3

[2]ZGODAVOVÁ, K., LINCZÉNYI, A., NOVÁKOVÁ, R., SLIMÁK, I.: Profesionál kvality. TU Košice 2001. ISBN 80-7099-669-2

[3]MAJERÍK, M., EŠŠE. M.: The study of Kansei Engineering applied to the VHCQC Web Portal,Október 22-25,2008,Trnava *In: Annals of DAAAM for 2008 & Proceedings of the 19th International DAAAM Symposium "Intelligent Manufacturing & Automation: Focus on next Generation of Intelligent Systems and Solutions"*. - ISSN 1726-9679. - Vienna: DAAM International, 2008. - ISBN 978-3-901509-68-1. - p.0777-0778.

[4]GRIMSAETH, K. (2005): Kansei Engineering – Linking emotions and product features [online] (cit. 15.06.2009)

[5] BREZNICKÁ, A.: FMEA - účinný nástroj na kvantifikáciu rizika. CD ROM, máj 21-23, 2008, Trenčianske Jastrabie *In: Kondor 2008* [elektronický zdroj] : Proceedings of 2nd Conference of PhD. students. - Trenčianske Jastrabie, máj 21-23, 2008. - 1 elektronický optický disk, 220 s. - ISBN 978-80-214-3663-3. - 1 elektronický optický disk, s. 1-4.

[6] <http://mindstorms.lego.com/eng/Egypt_dest/Default.aspx> (cit. 15.06.2009)

Adresa autorov:

Jaroslav Dvorský, Ing.
Študentská 2
911 50 Trenčín
dvorsky@tnuni.sk

Martin Ešše, Ing.
Študentská 2
911 50 Trenčín
esse@tnuni.sk

Marián Majerík, Ing.
Študentská 2
911 50 Trenčín
marian.majerik@tnuni.sk

Klasifikace signálů AE pomocí fuzzy metod a ϕ-divergencí

Classification of AE signals based on fuzzy methods and ϕ-divergences

Farová Zuzana, Kůs Václav

Abstract: We present a new classification method applied to laboratory measured acoustic emission signals. This method combined fuzzy classification approach with parametric versions of ϕ-divergences between two normalized spectrums of the individual acoustic signals. This nonmetric measure of distance and dissimilarities between the original signals is then used as an input numerical parameter into the classification method. Overall comparison of complete acoustic signals is the main advantage of this procedure. Moreover, ϕ-divergences can be adapted through its parameters to achieve optimal statistical properties of classification.

Key words: Acoustic emission, Fuzzy method, ϕ-divergence, Classification of signal

Klíčová slova: akustická emise, fuzzy metoda, ϕ-divergence, Klasifikace signálů

1. Úvod a teoretický popis klasifikace

Pracujeme na metodě, která by s co nějvětší úspěšností klasifikovala signály akustické emise do skupin, které odpovídají fyzikalní podstatě zdrojů. Hlavním cílem klasifikace akustické emise je odlišit emisi způsobenou vznikajícím defektem ve zkoumaném materiálu od běžných emisí způsobených šumem. K určení typu zdroje akustické emise jsou důležité dva základní body

1. zvolit vhodné charakteristické parametry signálů, podle nichž se naměřené signály co nejlépe klasifikují, tzn. obsahující maximum komprimované informace o odlišnosti těchto srovnávaných signálů,
2. zvolit vhodnou metodu pro klasifikaci.

Klasifikaci je možné provádět pomocí metod shlukové analýzy. V našem případě jsme použili metodu fuzzy klasifikace, která je založena na minimalizaci tzv. objektivní funkce. Parametr pro klasifikaci jsme počítali pomocí ϕ-divergencí, které jsme aplikovali na normovaná spektra signálů akustické emise. Výhodou použití tohoto parametru je zohlednění celého signálu. I při použití pouze tohoto jednoho parametru, jsme dosáhli velmi dobré úspěšnosti při klasifikaci.

2. Fuzzy metody klasifikace

Fuzzy metoda klasifikace patří mezi metody založené na optimalizaci tzv. objektivní funkci. Máme-li N vzorků $(x_k)_1^N$ v R^n a předpokládáme, že chceme vytvořit c shluků, pak objektivní funkci uvažujeme ve tvaru "vážené" sumy kvadratických odchylek mezi jednotlivými vzorky x_k a souborem prototypů shluků $v_1, v_2, \ldots, v_c$ (tj. reprezentativních vzorků). Objektivní funkce má tvar $Q = \sum_{i=1}^{c}\sum_{k=1}^{N} u_{ik} d^2(x_k, v_i)$,

kde $d(x_k, v_i)$ označuje vzdálenost mezi vzorkem x_k a v_i. Důležitou součástí sumy je matice rozdělení $U = [u_{ik}]$, $i = 1,2,\ldots,c$, $k = 1,2,\ldots,N$, jejíž úlohou je přidělit vzorky do shluků. Obecně u metod založených na objektivní funkci jsou prvky matice U binární. Vzorek s

indexem k přísluší ke shluku i, je-li $u_{ik} = 1$, tentýž vzorek není zahrnutý do shluku, pokud $u_{ik} = 0$. V případě klasifikační fuzzy metody představují složky matice U stupně členství vzorků ve shlucích a splňují následující požadavky:

• shluk je netriviální, tzn. není prázdný a neobsahuje všechny vzorky

$$0 < \sum_{k=1}^{N} u_{ik} < N, \quad i = 1,2,\ldots,c \tag{1}$$

• pro součet platí $$\sum_{i=1}^{c} u_{ik} = 1, \quad k = 1,2,\ldots,N. \tag{2}$$

Tyto podmínky platí i pro matici rozdělení u dalších obecných metod založených na objektivní funkci. V případě fuzzy metody je navíc stupeň členství u_{ik} v příslušné objektivní funkci umocněn na tzv. fuzzyfikační faktor $m > 1$, který pomáhá kontrolovat tvary shluků a vytváří rovnováhu mezi stupni členství blízko u 1 nebo 0 a těmi stupni členství s průměrnými hodnotami. Objektivní funkce má tvar

$$Q = \sum_{i=1}^{c}\sum_{k=1}^{N} u_{ik}^{m} d^2(x_k, v_i).$$

Omezení, která jsou kladena na matici rozdělení, začleníme do optimalizace pomocí Lagrangových multiplikátorů, kdy budeme minimalizovat po složkách. Po několika úpravách s využitím omezující podmínky (2) dostaneme

$$u_{sk} = \frac{1}{\sum_{j=1}^{c}\left(\frac{d_{sk}}{d_{jk}}\right)^{2/(m-1)}}.$$

Výpočet prototypů je přímý, neboť na ně nejsou kladena žádná omezení. Extrém Q se počítá ze soustavy $\nabla_{v_s} Q = 0.$ Detailní řešení závisí na volbě vzdálenostní funkce, v případě Eukleidovské vzdálenosti dostaneme výraz $v_s = \frac{\sum_{k=1}^{N} u_{sk}^{m} x_k}{\sum_{k=1}^{N} u_{sk}^{m}}$.

Na algoritmus fuzzy metody klasifikace můžeme pohlížet jako na iterační proces zahrnující postupné počítání prototypů shluků a matic rozdělení. Počáteční vstupní hodnoty parametrů se vkládají předem. Skládají se z následujících položek: počet shluků (c), vzdálenostní funkce (d), fuzzyfikační faktor (m) a omezující kritérium pro iterační metodu (ε), dále se též inicializuje matice rozdělení U tak, aby její složky splňovaly požadavky (1) a (2). Pro klasifikaci jsme použili hodnotu fuzzyfikačního faktoru $m = 2$ a jako vzdálenostní funkci Eukleidovskou vzdálenost.

3. ϕ-divergence

Jako jeden z parametrů pro klasifikaci signálů je možné použít tzv. ϕ-divergenční míru mezi spektrálními hustotami signálů AE, která má výhodu značné flexibility jak již v samotné definici divergence (funkce ϕ), tak v metrických a robustních vlastnostech této míry vzdálenosti na prostorech pravděpodobnostních měr indukovaných odpovídajícími hustotami pravděpodobnosti. ϕ-divergence je definována následovně.

Definice 1 Nechť $\mathsf{P}(\mathsf{Y},\mathsf{A})$ je soubor pravděpodobnostních měr se σ-algebrou A nad prostorem Y, $P,Q \in P(\mathsf{Y},\mathsf{A})$. Označme $p = \mathrm{d}P/\mathrm{d}\mu$ a $q = \mathrm{d}Q/\mathrm{d}\mu$ Radon-Nikodymovy derivace P a Q vzhledem k σ-konečné míře μ. Pro danou generující divergenční funkci $\phi : (0,\infty) \to \mathsf{R}$ konvexní na $(0,\infty)$, definujeme ϕ-divergenci měr *P* a *Q* vztahem

$$D_\phi(P,Q) = \int_{\mathsf{Y}} q\phi\left(\frac{p}{q}\right)\mathrm{d}\mu, \quad P,Q \in \mathsf{P}(\mathsf{Y},\mathsf{A}).$$

Jednoduché parametrické rodiny divergencí lze nalézt pomocí speciální procedury rozšíření. Tyto rodiny zahrnují páry známých a v dnešní době často používaných divergencí. Proceduru rozšíření a podrobnosti k rozšiřování parametrických rodin divergencí lze nalézt v [2]. Naše rozšiřovací procedura se nazývá *konvexní standardizace*. Tato procedura rozšiřuje konvexní nebo konkávní funkce $\psi : (0,\infty) \to \mathsf{R}$, které jsou dvakrát spojitě diferencovatelné v okolí $t = 1$, přičemž pro druhou derivaci platí $\psi''(1) \neq 0$.

Definice 2 Konvexní standard funkce ψ je definován výrazem

$$\phi(t) = \frac{\psi(t) - \psi(1) - \psi'(1)(t-1)}{\psi''(1)}, \qquad t > 0. \tag{3}$$

Za funkci ψ můžeme do vzorce (3) použít libovolnou $\psi : (0,\infty) \to \mathsf{R}$ dvakrát spojitě diferencovatelnou na intervalu $(0,\infty)$ pro kterou platí, že $\psi''(t) \neq 0 \qquad pro\ t > 0.$

Rozšířené power divergence: Konvexní standardizaci budeme aplikovat na třídu funkcí

$$\psi_{\alpha,\beta}(t) = \frac{1}{\beta t^\alpha + 1 - \beta}, \qquad t > 0,$$

pro vhodné parametry (α,β) takové, že odpovídající $\psi_{\alpha,\beta}$ splňují podmínky standardizace.

Konvexní standardizací dostáváme třídu divergenčních funkcí

$$\phi_{\alpha,\beta}(t) = \frac{1}{\alpha(2\alpha\beta - \alpha + 1)}\left[\frac{1 - t^\alpha}{\beta t^\alpha + 1 - \beta} + \alpha(t-1)\right], \qquad t > 0.$$

Rozšířené absolutní divergence: Rozšířenou třídu absolutních divergencí získáme pomocí konvexní standardizace (3), použitím funkce $\psi_{\alpha,\beta}(t) = |t + \beta - 1|^\alpha, \quad t > 0,$ pro vhodné parametry α, β. Z (3) dostaneme třídu divergenčních funkcí

$$\phi_{\alpha,\beta}(t) = \frac{|t + \beta - 1|^\alpha - |\beta|^\alpha - \alpha\, sign(\beta)|\beta|^{\alpha-1}(t-1)}{\alpha(\alpha - 1)}.$$

4. Aplikace na laboratorně naměřená data

Všechny laboratorně měřené signály byly získány pomocí piezoelektrických snímačů na tenké plechové desce o rozměrech $0{,}7m \times 0{,}8m \times 5mm$. Signály akustické emise byly zaznamenávány do počítače pomocí měřícího přístroje DAKEL-XEDO v rozlišení 12 bitů a vzorkování 8 MHz. Rozlišení 12 bitů znamená, že snímací aparatura zaznamenává hodnoty napětí v intervalu [-2048mV, 2048mV]. Akustickou emisi jsme budili čtyřmi způsoby vždy ve středu desky, snímače byly umístěny ve všech čtyřech rozích desky, vždy ve vzdálenosti 10 cmod nejbližších hran. Fotografii desky se snímači můžeme vidět na obr. 1. Způsoby, jakými byla buzena akustická emise jsou: • lámání tuhy o průměru 0,5 mm - ozn. **Tuha 5HB**

- lámání tuhy o průměru 0,7 mm - ozn. **Tuha 7HB**
- pouštění zrnek rýže na desku z výšky 25 cm - ozn. **Rýže**

• údery štětkou na desku - ozn. **Štětka**.

Obrázek 1: Fotografie desky se snímači.

5. Divergence jako parametr

Do metod shlukové analýzy jsme použili parametr spočítaný z normovaného spektra. Spektrum signálu X_f je počítáno pomocí *diskrétní Fourierovy transformace*. Spektrum signálu normujeme na jedničku, čímž získáváme odhad spektrální hustoty posloupnosti $\{X_t\}_{t=0}^{T-1}$, kde T značí délku záznamu emisní události (v našem případě T=6144). Normované spektrum vypadá následovně

$$\tilde{S}(f) = \frac{|X_f|}{\sum_{f=0}^{T-1} |X_f|}.$$

Jako jeden z parametrů pro porovnávání normovaných spekter signálů jsme použili různé typy divergencí mezi spektry signálů AE. Nejprve jsme si vytvořili tzv. *referenční spektrum*, což je vyprůměrované spektrum jednoho typu signálů (Rýže, Štětka, Tuha 5HB, Tuha 7HB) akustické emise, a je dáno vztahem

$$\tilde{S}^{refer}(f) = \frac{1}{m}\sum_{i=1}^{m} |\tilde{S}^i(f)|, \quad pro\ každé\ f = 0,\ldots,T-1,$$

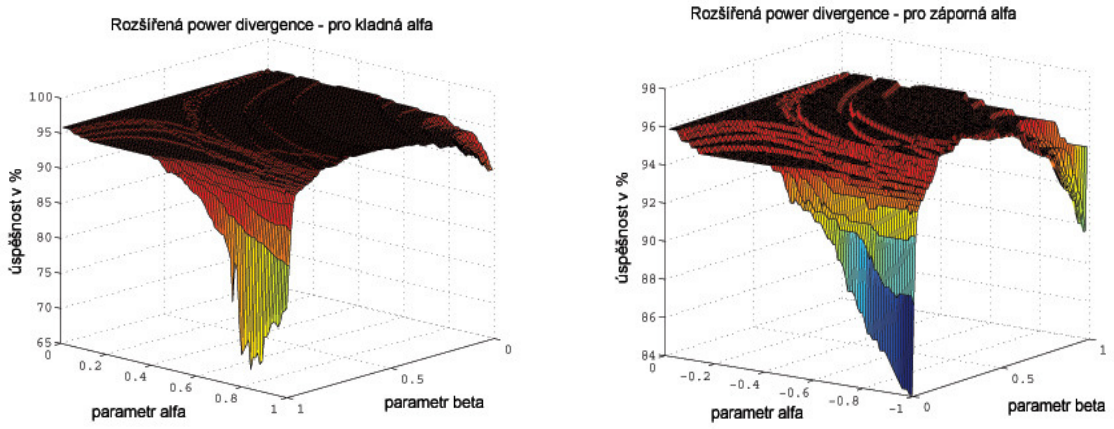
kde m je počet vzorků daného typu signálu, $\tilde{S}^i(f)$ je hodnota $\tilde{S}(f)$ i-tého spektra daného typu. Během výpočtů se ukázalo, že výsledná klasifikace nezávisí na tom, z jakého typu signálu spočítáme referenční spektrum, vůči kterému pak následně počítáme divergenci. Pro výpočet divergence použijeme například referenční spektrum z typu Rýže. Divergenci mezi spektry počítáme v diskrétní podobě, obecně dané vztahem

$$D_\phi(\tilde{S}^{refer}, \tilde{S}) = \sum_{f=0}^{T-1} \tilde{S}(f)\phi\left(\frac{\tilde{S}^{refer}(f)}{\tilde{S}(f)}\right), \tag{4}$$

kde za $\tilde{S}$ postupně dosazujeme normovaná spektra signálů. Za divergenci D_ϕ poté volíme námi vybrané typy divergencí pro zvolené divergenční generující funkce ϕ. Hodnotu divergence mezi referenčním spektrem a spektrem signálu lze použít jako parametr do klasifikační metody.

6. Výsledky fuzzy metody při použití ϕ-divergencí

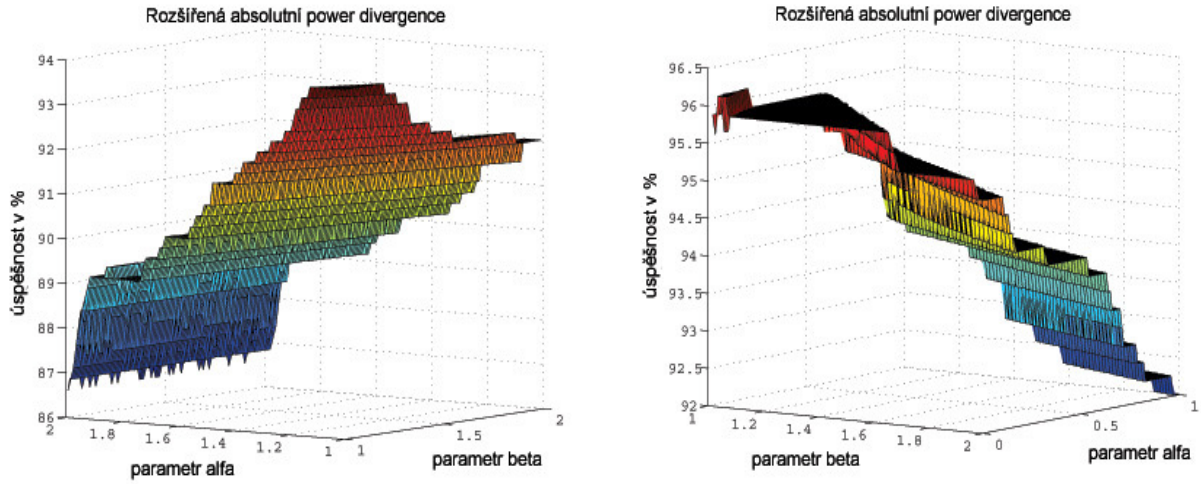
Za divergence jsme volili reprezentanty z rodiny rozšířených power divergencí a absolutních power divergencí uvedených výše. Divergence jsme počítali pro různé kombinace parametrů α, β, pro které je daná rozšířená divergence definována pomocí procesu rozšíření.



Obrázek 2: Úspěšnost fuzzy metody při použití rozšířené power divergence pro hodnoty parametrů $(\alpha,\beta)\in[0,1)\times[0,1]$ ***a*** $(\alpha,\beta)\in[-1,0]\times[0,1)$.

Na obrázku 2 a 3 vidíme, jaké úspěšnosti jsme dosáhli při použití fuzzy metody na rozšířenou power divergenci pro hodnoty parametrů $(\alpha,\beta)\in[0,1)\times[0,1]$ a pro hodnoty $(\alpha,\beta)\in[-1,0]\times[0,1)$. Z těchto grafů je vidět, že fuzzy metoda dosahuje pro většinu kombinací parametrů úpěšnosti více než 95%. Úspěšnost začíná klesat, pokud se začneme s hodnotami parametrů (α,β) blížit k hraničním rohovým bodům oblasti, pro kterou je rozšířená power divergence definována.

- Pro $(\alpha,\beta)\to(1,1)$ klesá úspěšnost k hodnotám pod 70%.
- Pro $(\alpha,\beta)\to(1,0)$ klesá úspěšnost k hodnotě 89%.
- Pro $(\alpha,\beta)\to(-1,1)$ klesá úspěšnost k hodnotě 89%.
- Pro $(\alpha,\beta)\to(-1,0)$ klesá úspěšnost k hodnotě 84%.



Obrázek 3: Úspěšnost fuzzy metody při použití rozšířené absolutní power divergence pro hodnoty parametrů $(\alpha,\beta)\in[1,2]\times[1,2]$ ***a*** $(\alpha,\beta)\in[0,1]\times[1,2]$.

Na obrázku 4 je znázorněna úspěšnost fuzzy medoty při použití rozšířené absolutní power divergence pro hodnoty parametrů $(\alpha,\beta)\in[1,2]\times[1,2]$. V tomto případě se dosahuje největší úspěšnosti v okolí hodnoty parametrů $(\alpha,\beta)=(1,1)$ a to úspěšnosti téměř 94%. Čím více se hodnoty parametrů vzdalují od hodnoty $(1,1)$, tím nižší je úspěšnost fuzzy metody. Pro $(\alpha,\beta)\to(2,2)$ klesá úspěšnost k hodnotám pod 87% a to tak, že tento pokles je pásový a způsobený převážně vzrůstající hodnotou parametru $\beta\to 2$. Obdobný charakter má průběh grafu úspěšnosti pro $(\alpha,\beta)\in[0,1]\times[1,2]$, ovšem v tomto případě jsme dosáhli maximální úspěšnosti přes 96% v okolí bodu $(0,1)$.

7. Závěr

Na závěr uvádíme pro porovnání výsledky fuzzy klasifikce při použití parametrů ze spektra W, $Q_{0.33}$, S, P, jejichž detailnější popis lze nalézt v [4] a [1]. Data byla rozdělena z experimentálních důvodů v těchto případech do třech podsouborů.

Fuzzy metoda			
Data	soubor1	soubor2	soubor3
W, $Q_{0.33}$	86%	89%	86%
W, $Q_{0.33}$, P	86%	89%	86%
W, $Q_{0.33}$, S	96%	90%	94%

Pro úplnost jsme aplikovali fuzzy metodu při použití parametrů W, $Q_{0.33}$, S na všechna data a dosáhli jsme úspěšnosti 91%. Při použití rošířených power divergencí na všechna data jsme dosahovali úspěšnosti přes 95%, v porovnání s předchozí tabulkou je to nejlepší výsledek.

8. Literatura

[1] Farová, Z. – Kůs, V. 2009. Pokročilejší metody klasifikace signálů akustické emise. Výzkumná zpráva, Katedra matematiky FJFI ČVUT Praha, 2009.

[2] Kůs, V. – Morales, D. – Vajda, I. 2008. Extension of the Parametric Families of Divergences Used in Statistical Inference. In: Kybernetika vol.44/1, 2008, s. 95 - 112.

[3] Pedrycz, W. 2005. Knoledge-Based Clustering, From Data to Information Granules. Wiley-Interscience, New Jersey, 2005.

[4] Tláskal, J. 2008. Statistické metody klasifikace signálů. Diplomová práce KM FJFI ČVUT, Praha, 2008.

Adresa autorov:

Farová Zuzana
Katedra matematiky FJFI ČVUT
Trojanova 13, 120 00 Praha 2
zuzana.farova@email.cz

Kůs Václav, Ing. PhD.
Katedra matematiky FJFI ČVUT
Trojanova 13, 120 00 Praha 2
vaclav.kus@fjfi.cvut.cz

Projekce obyvatelstva ČR a jejích krajů
Population Projection of the Czech Republic and of its Regions

Tomáš Fiala, Jitka Langhamrová

Abstract: Latest demographical data of the Czech Republic indicate stagnation of fertility and rapid decrease of net migration. New population projection of the Czech Republic and its regions has been computed in two variants. The variant CSÚ is based on the medium variant of the latest projection computed by the Czech Statistical Office. The variant NL supposes higher increase of fertility and life expectancy and higher net migration. In the first variant the population size of the Czech Republic will be stabilized below 11 millions, the second variant expects continual growth of the population to more then 12 millions of people. The highest increase of population is expected in surroundings of Prague and Plzeň, on the other hand Moravian regions will probably have the lowest increase, or even a slow decrease of population.

Key words: population projection, fertility, mortality, migration, regions of the Czech Republic.

Klíčová slova: populační projekce, plodnost, úmrtnost, migrace, kraje České republiky.

1. Úvod

V posledních letech dochází v ČR i v řadě jiných zemí k častějším změnám demografického chování než v minulosti. V letech 2007 a 2008 jsme byli svědky až nečekaně vysokého nárůstu plodnosti a především velmi výrazného nárůstu migračního přírůstku. Data za první pololetí roku 2009 však naznačují, že úhrnná plodnost v tomto roce zřejmě již dále neporoste, bude zhruba stejná jako v roce 2008. Migrační přírůstek se v roce 2009 očekává výrazně nižší, zhruba na úrovni poloviny až třetiny hodnot roku 2008. Proto byla na naší katedře vypočtena aktualizovaná projekce obyvatelstva ČR i jejích krajů. Populační projekce byla vypočtena ve dvou variantách.

První varianta vychází ze střední varianty projekce Českého statistického úřadu z roku 2009. (Tuto variantu budeme nazývat varianta ČSÚ.)

Druhá varianta vychází z předpokladu, že demografické chování české populace bude – s jistým zpožděním – kopírovat demografické chování populace Nizozemska. Nizozemsko bylo vybráno proto, že se jedná o populaci, kde již byla dokončena transformace plodnosti žen do vyššího věku, plodnost je zde poměrně stabilní a rovněž úmrtnost v Nizozemsku se zdá být poměrně stabilní. Navíc se jedná o populaci geograficky nepříliš vzdálenou a co do velikosti v jistém smyslu srovnatelnou s obyvatelstvem České republiky. (Tuto variantu budeme nazývat varianta NL.)

Vstupní data pro výpočty týkající se obyvatelstva ČR byla získána z internetových stránek ČSÚ, zdrojem dat o obyvatelstvu Nizozemska byl Eurostat.

2. Scénáře projekce pro celou ČR

Výchozí demografickou strukturou pro obě varianty bylo složení obyvatelstva České republiky podle pohlaví a jednotek věku k 1. 1. 2009.

VARIANTA ČSÚ

PLODNOST

Předpokládá se, že po krátké stagnaci na přelomu tohoto desetiletí vzroste úhrnná plodnost do roku 2020 plynule na 1,6, v dalších 30 letech již poroste pomaleji, v roce 2050

dosáhne hodnoty 1,72. Nejvyšší specifické plodnosti budou mít ženy ve věku 28–30 let, to znamená, že struktura plodnosti se již prakticky nebude měnit.

Tabulka 1: Scénář plodnosti – varianta ČSÚ

Rok	2008	2009	2010	2020	2030	2040	2050
Úhrnná plodnost	1,50	1,50	1,50	1,60	1,66	1,69	1,72

ÚMRTNOST

Předpokládáme, že střední délka života se bude i nadále prodlužovat po celé období projekce. Do roku 2030 bude roční nárůst zhruba stejný jako v současné době, po roce 2030 dojde ke zpomalení růstu střední délky života. Jako výchozí byla použita průměrná struktura pravděpodobností úmrtí za léta 2006–2008, předpokládá se, že bude po celou dobu projekce stejná.

TABULKA 2: SCÉNÁŘ ÚMRTNOSTI – VARIANTA ČSÚ

Rok	2008	2009	2010	2020	2030	2040	2050
Střední délka života mužů	73,96	74,23	74,50	77,00	79,50	81,50	83,50
Střední délka života žen	80,14	80,37	80,60	82,80	85,10	86,80	88,40

MIGRACE

Migrační saldo se po celé období projekce předpokládá konstantní ve výši 25 000 osob ročně. Co se týče demografické struktury migrantů, byl přijat předpoklad, že v roce 2009 bude odpovídat váženému průměru za období 2004–2007. V dalších letech do roku 2030 se předpokládalo postupné přibližování této struktury struktuře migračního salda EU (pro kterou je charakteristický především vyrovnaný poměr mužů a žen), po roce 2030 se předpokládala trvale struktura EU.

VARIANTA NL

PLODNOST

Odhad vývoje plodnosti byl v této variantě proveden na základě odhadu vývoje plodnosti jednotlivých „pseudokohort“ (tj. vzájemně se překrývajících kohort žen vždy dvou sousedních ročníků narození), přičemž se předpokládá, že plodnost kohort českých žen bude s určitým zpožděním kopírovat plodnost žen Nizozemska, kde již byl ukončen přesun plodnosti do vyššího věku a kohortní plodnost se zde zdá být poměrně stabilní.

České kohorty žen narozených v letech 1965–1977 již dosáhly vrcholu své plodnosti. Pro každou z těchto kohort byla na základě hodnot posledních známých specifických měr plodnosti a grafu trendu vývoje plodnosti v posledních letech nalezena „podobná“ kohorta nizozemská. Při odhadu neznámých specifických měr plodnosti těchto českých kohort ve vyšším věku se tedy předpokládá, že jejich specifické míry plodnosti budou odpovídat specifickým mírám plodnosti „podobných“ kohort nizozemských – buď budou přímo rovny mírám nějaké kohorty, nebo průměrné plodnosti dvou či více sousedních kohort.

Pro české kohorty 1978 a mladší se předpokládá, že jejich specifické míry plodnosti ve věku 30 a více let budou (podobně jako předchozí kohorty) „kopírovat“ míry plodnosti starších kohort nizozemských. Teprve české kohorty roku 1991 a mladší budou mít stejné tempo vývoje jako Nizozemsko. Pokud pro mladší nizozemské kohorty již nebyla známa plodnost pro nejvyšší jednotky věku, předpokládá se, že tato plodnost bude rovna poslední známé plodnosti v daném věku, tj. že plodnost v Nizozemsku v nejvyšším věku již dále neporoste. O plodnosti výše uvedených českých kohort ve věku do 30 let se předpokládá, že bude postupně klesat na úroveň poslední známé plodnosti nizozemských kohort v daném věku.

Věk nejvyšší plodnosti českých kohort roste. Zatímco kohorty narozené na přelomu 50. a 60. let dosahovaly nejvyšší plodnosti již v 21 letech, plodnost kohorty 1977 vrcholila až ve 30 letech a pro mladší kohorty předpokládáme vrchol plodnosti až ve 31 letech věku.

Na základě odhadnuté plodnosti českých kohort byla zpětně určena průřezová plodnost pro výpočet demografické projekce.

TABULKA 3: SCÉNÁŘ PLODNOSTI – VARIANTA NL

Rok	2008	2009	2010	2020	2030	2040	2050
Úhrnná plodnost	1,50	1,50	1,51	1,70	1,80	1,85	1,90

ÚMRTNOST

V České republice byl od roku 2001 průměrný roční nárůst střední délky života mužů o 0,33 roku, u žen pak o 0,28 roku. Projekce v této variantě předpokládá, že střední délka života mužů i žen poroste stejným tempem i v dalších letech. Vývoj střední délky života zachycuje následující tabulka.

TABULKA 4: SCÉNÁŘ ÚMRTNOSTI – VARIANTA NL

Rok	2008	2009	2010	2020	2030	2040	2050
Střední délka života mužů	73,96	74,33	74,66	77,96	81,27	84,58	87,88
Střední délka života žen	80,14	80,46	80,74	83,53	86,33	89,12	91,92

MIGRACE

Ze složek populačního vývoj je migrace tou nejobtížněji prognózovatelnou. Závisí totiž nejen na případných změnách v legislativě přijímající země, ale, a to snad především, na politické, hospodářské a osobní situaci v zemi, z níž se osoba vystěhovává.

Navzdory hospodářské krizi lze i nadále předpokládat, že Česká republika zůstane zemí imigrační, roční migrační saldo však předpokládáme výrazně nižší než v předchozích letech 2007 a 2008. Věková struktura imigrantů se i nadále bude lišit od věkové struktury osob s občanstvím České republiky (a zároveň s trvalým pobytem na jejím území). Česká republika bude i v budoucnu cílem spíše pracovní migrace, a tak mezi imigranty bude výrazně zastoupena produktivní generace.

V letech 2009–2016 se i nadále počítá zejména s pracovní migraci mužů a s velmi pozvolným nárůstem migračního salda z 30 000 na 35 000 tisíc ročně. Současně předpokládáme mírné změny věkové a pohlavní struktury migrantů vzhledem k migraci žen a dětí, které budou postupně následovat své manžele (partnery) resp. otce.

V letech 2017–2021 se předpokládá stejné věkové a pohlavní složení migračního salda jako v předchozím období a každým rokem tohoto pětiletého období se saldo migrace zvýší o 1 000 osob až ke 40 000.

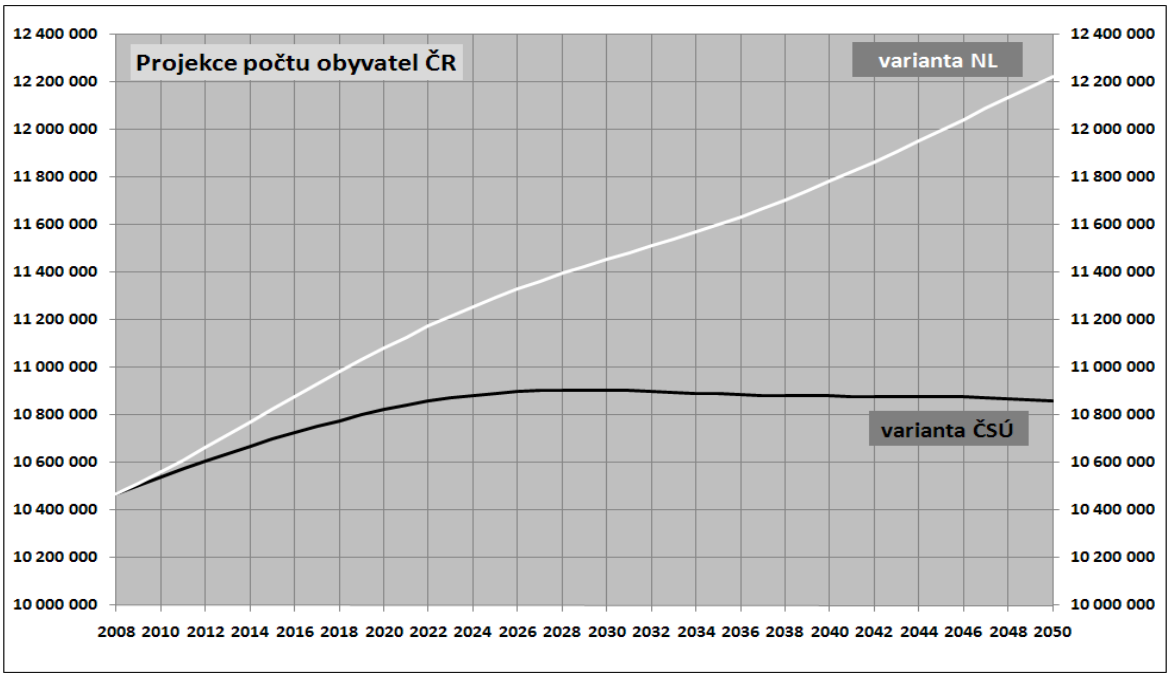
Počínaje rokem 2022 by mohli migranti, kteří přišli do České republiky na počátku období projekce, mít nárok na získání českého občanství. Tato skutečnost by mohla posílit migraci za účelem slučování rodin. Tedy za muži, kteří odešli ze země původu z ekonomických důvodů, by postupně přicházely jejich ženy a děti. To by mělo vliv nejen na věkovou a pohlavní strukturu migračního salda, ale i na jeho výši, a tak se od roku 2022 předpokládá migrační saldo ve výši 40 000 osob ročně.

TABULKA 5: SCÉNÁŘ MIGRAČNÍHO SALDA – VARIANTA NL

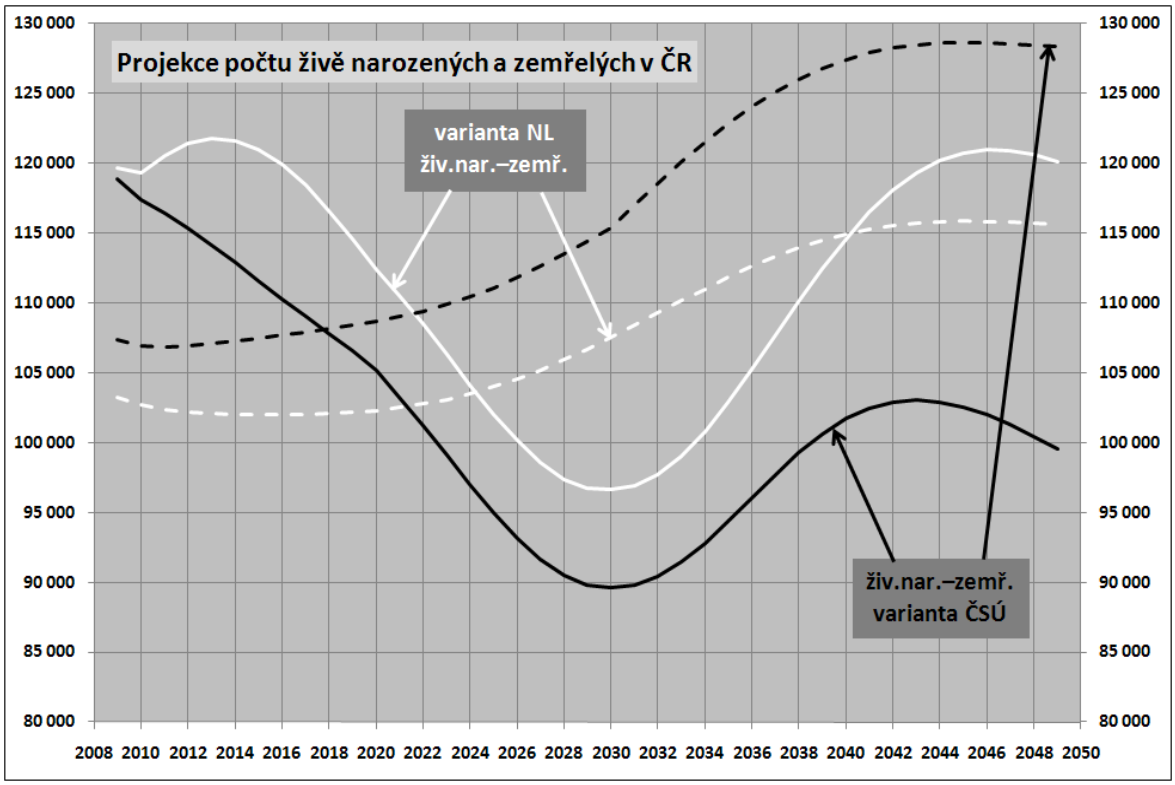
Rok	2008	2009	2010	2015	2020	2022–50
Roční saldo migrace	71 790	30 000	31 000	35 000	38 000	40 000

3. Základní výsledky projekce

Na první pohled se jedná o dvě diametrálně odlišné varianty vývoje. Upravená střední varianta projekce ČSÚ předpokládá, že počet obyvatel ČR se během 20 let přiblíží k 11 milionům, ale poté začne stagnovat či dokonce mírně klesat (viz Obr. 1). Naproti tomu podle varianty NL, předpokládající vyšší plodnost, rychlejší růst střední délky života i vyšší migrační přírůstek, počet obyvatel ČR trvale poroste a překročí hranici 12 milionů.



Obr. 1: Předpokládaný vývoj počtu obyvatel



Obr. 2: Předpokládaný vývoj počtu živě narozených a zemřelých

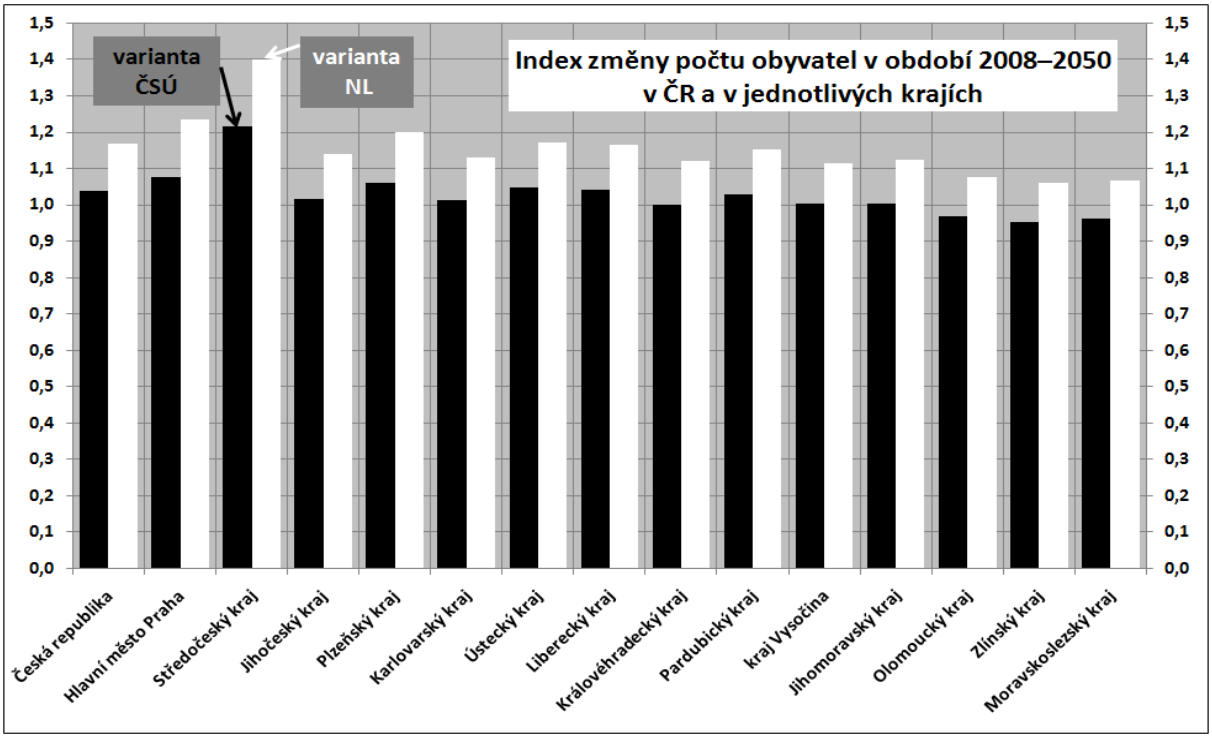
Zcela rozdílný charakter má i vývoj počtu živě narozených a zemřelých (Obr. 2). Podle varianty ČSÚ nastane zhruba za 10 let okamžik, kdy počet živě narozených bude opět nižší než počet zemřelých a tento stav zůstane zachován trvale, přičemž rozdíl se bude stále zvyšovat. Podle varianty NL nastane tento okamžik zhruba až o 5 let později, rozdíly nebudou příliš výrazné a vzhledem k vyšší plodnosti a migraci lze koncem první poloviny tohoto století opět očekávat kladný přirozený přírůstek obyvatelstva.

4. Projekce obyvatelstva v jednotlivých krajích ČR

Demografická projekce vývoje obyvatelstva jednotlivých krajů byla vypočtena rovněž ve dvou variantách a scénáře vývoje plodnosti, úmrtnosti i migrace vycházely z výše uvedených scénářů pro Českou republiku. Výchozí demografickou strukturou bylo pro obě varianty složení obyvatelstva příslušného kraje podle pohlaví a jednotek věku k 1. 1. 2009.

Předpokládalo se, že během prognózovaného období bude struktura plodnosti i úmrtnosti ve všech krajích stejná jako v České republice, úhrnné plodnosti a střední délky života se ale budou v jednotlivých krajích lišit. Na počátku budou indexy hodnot těchto charakteristik vzhledem k hodnotě za Českou republiku na úrovni průměrných indexů za období 2001–2008 a během prognózovaného období se budou rozdíly mezi kraji postupně snižovat tak, že v roce 2050 bude plodnost i úmrtnost ve všech krajích stejná jako v České republice (tj. hodnoty všech indexů budou rovny jedné).

O migraci jsme předpokládali, že během prognózovaného období bude struktura migračního salda ve všech krajích stejná jako v České republice. Na počátku bude podíl imigrantů (z celkového migračního přírůstku České republiky) do každého kraje roven průměrnému podílu migračního salda příslušného kraje (včetně vnitřní migrace v rámci České republiky) k migračnímu saldu celé České republiky za období 2002–2008. Během prognózovaného období se rozdělení migračního salda do jednotlivých krajů bude postupně měnit tak, že v roce 2050 bude podíl imigrantů do každého kraje zhruba úměrný počtu jeho obyvatel.



Obr. 3: Předpokládaná změna počtu obyvatel v ČR a v jednotlivých krajích

Podrobné výsledky projekce pro jednotlivé kraje není možno vzhledem k omezenému rozsahu tohoto článku prezentovat. Uvádíme alespoň přehled předpokládaných změn počtu obyvatel v jednotlivých krajích během období projekce (viz Obr. 3). Vzhledem ke zvoleným scénářům je ve všech krajích nárůst počtu obyvatel při variantě NL vyšší než při variantě ČSÚ. Nejvyšší nárůst obyvatelstva lze předpokládat v Praze, Středočeském kraji a v Plzeňském kraji, tedy v českých krajích v okolí velkých měst. Nejnižší nárůst počtu obyvatel bude naopak v moravských krajích, při variantě ČSÚ se dokonce v těchto krajích (s výjimkou Jihomoravského) očekává mírný pokles obyvatelstva.

5. Závěr

Je těžké říci, která z uvedených variant vývoje je více pravděpodobná. V minulé době jsme byli svědky několika krátkodobých změn trendů demografického vývoje. Málokdo očekával tak velký pokles plodnosti v 90. letech, opětovný růst na počátku 21. století byl také o něco vyšší, než se předpokládalo. Je otázkou, zda letošní stagnace plodnosti je pouze krátkodobým přerušením dalšího růstu či zda se jedná opět o déledobou změnu předchozího trendu. Rovněž migrační přírůstek v předchozích dvou letech byl nečekaně vysoký, jeho další vývoj asi do značné míry závisí na dalším vývoji ekonomické situace v ČR i ve světě.

V každém případě je však zřejmé, že za předpokladu dalšího růstu plodnosti, růstu střední délky života a kladného migračního salda nebude v České republice v nejbližších desetiletích docházet k úbytku obyvatelstva.

6. Literatura

[1]FIALA, T. 2008. Analýza růstu střední délky života v ČR metodou klouzavé lineární regrese. **Forum Statisticum Slovacum [CD-ROM]**, 2008, roč. 6, č. 6, s. 31–35. ISSN 1336-7420.

[2]FIALA, T. 2006. Dva přístupy modelování vývoje úmrtnosti v populační projekci a jejich aplikace na populaci ČR. Bratislava 05.10.2006 – 06.10.2006. In: **Forum Statisticum Slovacum 4/2006.** Bratislava : Slovenská štatistická a demografická spoločnosť, s. 44–55. ISSN 1336-7420.

[3]FIALA, T. – LANGHAMROVÁ, J. 2009. Některé aspekty budoucího demografického vývoje České republiky. **Forum Statisticum Slovacum**, 2009, roč. 5, č. 5, s. 32–36. ISSN 1336-7420.

[4]KOSCHIN, F. 2007. Prognóza lidského kapitálu obyvatelstva České republiky do roku 2050. Praha : Oeconomica. 105 s. (Další autoři: FIALA, T. – FISCHER, J. – HLAVÍNOVÁ, H. – HULÍK, V. – KAČEROVÁ, E. – LANGHAMROVÁ, J. – MAZOUCH, P. – PIKÁLKOVÁ, S. – ŠŤASTNOVÁ, P. – FOŘTLOVÁ, S.).

[5]Readings in Population Research Methodology. 1993. Vol. 5. Population Models, Projections and Estimates. BOGUE, D. J., ARRIAGA, E. E. and ANDERTON, D., L. (eds.). United Nations Population Fund, Social Development Center, Chicago, Illinois.

Adresa autora (-ov):

Tomáš Fiala, RNDr., Csc.
katedra demografie
fakulta informatiky a statistiky
nám. W. Churchilla 4
130 67 Praha 3
fiala@vse.cz

Jitka, Langhamrová, doc., Ing., Csc.
katedra demografie
fakulta informatiky a statistiky
nám. W. Churchilla 4
130 67 Praha 3
fiala@vse.cz

Asymptotické vlastnosti odhadů s minimální vzdáleností
Asymptotic properties of minimum distance density estimators

Hanousková Jitka, Kůs Václav

Abstract: This paper focuses on the minimum distance density estimate of probability density f on the real line. The rate of consistency of Kolmogorov estimate for $n\to\infty$ is known if the degree of variation is finite. The rate of consistency is studied under more general conditions. If our partial degree of variation is finite and some additional assumptions are fulfilled then the Kolmogorov, Lévy and discrepancy distance estimates are consistent of the order $n^{-1/2}$ in L_1-norm and also in the expected L_1-norm. Numerical simulations of these estimates are performed and graphs of consistency are presented.

Key words: Minimum distance estimators, Degree of variation, Consistency, PC simulation

Abstrakt: Příspěvek se zaměřuje na odhady hustot s minimální vzdáleností pro hustoty pravděpodobnosti f na reálné ose, pro které je řád konsistence Kolmogorovského odhadu známý za předpokladu konečnosti tzv. stupně variance rodiny hustot. Za obecnější podmínky a dodatečných předpokladů ukazujeme, že pokud tzv. částečný stupeň variance je konečný, pak odhady s minimální Kolmogorovskou, Lévyho a diskrepanční vzdáleností jsou konsistentní řádu $n^{-1/2}$ v L_1-normě a také ve střední hodnotě L_1-normy. Dále prezentujeme numerické simulace těchto odhadů a příslušné grafy konsistence.

Klíčová slova: odhady s minimální vzdáleností, stupeň variace, konzistence, pc simulace

1. Introduction

We introduce notation used in the following text. Let λ be a $\sigma-$ finite measure on $(\mathsf{R},\mathcal{B})$, with borel $\sigma-$field $\mathcal{B}$ on R. Let $\mathcal{F}_\lambda$ be the set of distributions on $(\mathsf{R},\mathcal{B})$ which are absolutely continuous with respect to a measure λ. Denote by $\mathcal{D}_\lambda$ the set in Banach space $L_1(\mathsf{R},\mathrm{d}\lambda)$ containing densities corresponding to distribution functions in $\mathcal{F}_\lambda$, by $\mathcal{D}$ its arbitrary nonempty subset and by $\mathcal{F}$ a subset of $\mathcal{F}_\lambda$.

Througout the text $\tilde{F}_n$ denotes the distribution function corresponding to the estimator of density $\tilde{f}_n$. Further, let $X_1,..,X_n$ be i.i.d. random sample and F_n be used for the empirical distribution function based on $(X_1,..,X_n)$ with the empirical measure $\nu_n(A)$

$$F_n(x)=\frac{1}{n}\sum_{j=1}^{n} I_{\{X_j\le x\}}, \qquad \nu_n(A)=\frac{1}{n}\sum_{j=1}^{n} I_{\{X_j\in A\}}, A\subset\mathsf{R},$$

where $I_{\{X_j\le x\}}$ means indicator of the event $\{X_j\in(-\infty,x]\}$ and $I_{\{X_j\in A\}}$ is indicator of $\{X_j\in A\}$.

Definition 1. We say that an estimate $\tilde{f}_n$ of a density $f\in\mathcal{D}$ is minimum D distance estimate iff the corresponding distribution function $\tilde{F}_n\in\mathcal{F}$ satisfies the condition:

$$D(\tilde{F}_n, F_n) = \inf_{F \in \mathrm{F}} D(F, F_n) \qquad \text{a.s.}$$

Every metric distance D on $\mathcal{F}$ (R) defines pseudometric ρ_D on $\mathcal{D}_\lambda$: $\rho_D(f,g) = D(F,G)$, where $F, G \in \mathcal{F}_\lambda$ are the distribution functions corresponding to densities $f, g \in \mathcal{D}_\lambda$.

Definition 2. We say that an estimate $\tilde{f}_n$ of a density $f \in \mathcal{D}$ is consistent in given ρ_D distance, respective in expected ρ_D distance, iff $\rho_D(\tilde{f}_n, f) \to 0$ a.s., respective iff $E\rho_D(\tilde{f}_n, f) \to 0$. We say that estimate $\tilde{f}_n$ is consistent of the order $r_n \to 0$ in the distance ρ_D, respective in expected ρ_D distance, iff $\rho_D(\tilde{f}_n, f) = O_p(r_n)$, respective iff $E\rho_D(\tilde{f}_n, f) = O(r_n)$.

We deal with Kolmogorov (ρ_K), Lévy (ρ_L), Cramer-von Mises (ρ_{C-M}), discrepancy (ρ_D) distances and also with the total variation (ρ_V), all defined by expressions:

$$\rho_K(f,g) = K(F,G) = \sup_{x \in \mathrm{R}} |F(x) - G(x)|, \qquad \rho_{C-M}(f,g) = \int_{\mathrm{R}} \big(F(X) - G(x)\big)^2 f(x)\mathrm{d}x,$$

$$\rho_L(f,g) = \inf\{\varepsilon > 0 : G(x-\varepsilon) - \varepsilon \le F(x) \le G(x+\varepsilon) + \varepsilon, \forall x \in \mathrm{R}\}, \quad \rho_V(f,g) = \int_{\mathrm{R}} |f - g| \mathrm{d}\lambda,$$

$\rho_D(f,g) = \sup_{B \in \mathbf{B}} |P(B) - Q(B)|$, where **B** is the set of all closed balls and P,Q are probability measures with distribution functions F,G corresponding to densities *f,g*.

2. Consistence in L_1-norm

Definition 3. We say that ρ_K dominates ρ_V on $\mathcal{D}$ (denoted by $\rho_K \succ \rho_V$) iff for all sequences $f, f_1, f_2, \ldots \in \mathcal{D}$ the convergence $f_n \to f$ in ρ_K for $n \to \infty$ implies the convergence $f_n \to f$ in ρ_V. Further, ρ_K uniformly dominates ρ_V locally with respect to ρ_K on $\mathcal{D}$ (denoted by $\rho_K \succ^u \rho_V/\rho_K$) iff for every density $g \in \mathcal{D}$ there exists $c > 0$ and Kolmogorov neighborhood of g, $B_K(g) \subset \mathcal{D}$, such that $\rho_K(f,g) \ge c\rho_V(f,g)$ for all $f \in B_K(g)$.

Theorem 1. *Let $\rho_K \succ \rho_V$ on $\mathcal{D}$, then arbitrary Kolmogorov estimate of a density from $\mathcal{D}$ is consistent in L_1 - norm. If $\rho_K \succ^u \rho_V/\rho_K$ on $\mathcal{D}$, then Kolmogorov estimate of a density from $\mathcal{D}$ is consistent of the order $n^{-1/2}$ in $L_1 -$ norm and also in the expected $L_1 -$ norm.*

It is known that every pair of densities $f, g \in \mathcal{D}_\lambda$ defines the finite measure ν with density $\frac{\mathrm{d}\nu}{\mathrm{d}\lambda} = f - g$. This measure is the difference of two finite measures on $(\mathrm{R}, \mathcal{B})$, the upper variation ν^+ and lower variation ν^- with the densities $\frac{\mathrm{d}\nu^+}{\mathrm{d}\lambda} = (f-g)^+ = \max\{0, f-g\}$ and ν^- with $\frac{\mathrm{d}\nu^-}{\mathrm{d}\lambda} = (f-g)^- = \max\{0, g-f\}$.

Definition 4. We say that $A \in \mathcal{B}$ separates ν^+ and ν^-, iff it holds either $\nu^+(A) = \nu^+(\mathsf{R})$ and $\nu^-(\mathsf{R} - A) = \nu^-(\mathsf{R})$ or $\nu^+(\mathsf{R} - A) = \nu^+(\mathsf{R})$ and $\nu^-(A) = \nu^-(\mathsf{R})$.

Definition 5. Let $f, g \in \mathcal{D}_\lambda$. Then the degree of variation $DV(f,g) \in [0,+\infty]$ is defined by: $DV(f,g) = 0$, if A separates ν^+ and ν^- with $\lambda(A) = 0$, otherwise

$$DV(f,g) = \inf\left\{ m \in N:\ A = \bigcup_{j=1}^{m} \mathcal{J}_j,\ A \text{ separates } \nu^+,\ \nu^- \right\},$$

where $\mathcal{J}_1, \ldots \mathcal{J}_m$ are nonvoid intervals in R. If the minimized set is empty, i.e. if there does not exist any m with desired properties, we set $DV(f,g) = +\infty$.

Definition 6. For a fixed $f \in \mathcal{D} \subset \mathcal{D}_\lambda$ and $\delta > 0$ we define local degree of variation $LDV_\delta(f)$ of the density f with respect to the Kolmogorov distance in $\mathcal{D}$ by means:

$$LDV_\delta(f) = \sup\{DV(f,g):\ g \in B_{K,\delta}(f) \cap \mathcal{D}\},$$

where $B_{K,\delta}(f)$ is the Kolmogorov ball in $\mathcal{D}$ with diameter δ centered in f. Further, we put the *degree of variation* $DV(\mathcal{D})$ of a family $\mathcal{D} \subset \mathcal{D}_\lambda$ to be:

$$DV(\mathcal{D}) = \sup\{DV(f,g): f, g \in \mathcal{D}\}.$$

Theorem 2. *Let $\mathcal{D} \subset \mathcal{D}_\lambda$ and for all densities $f \in \mathcal{D}$ there exists $\delta > 0$ such that $LDV_\delta(f) < +\infty$. Then ρ_K uniformly dominates ρ_V locally w.r.t. ρ_K ($\rho_K \succ^u \rho_V/\rho_K$) on $\mathcal{D}$.*

3. Consistence in L_1-norm under more general assumptions

Now we generalize the theory of Section 2 and we prove that Kolmogorov estimates are consistent in L_1-norm and in the expected L_1-norm also inside the families of densities while $LDV_\delta(f)$ need not be finite. But the family is forced to satisfy some additional assumptions. For this sake we introduce new variants of domination relations.

Definition 7. We say that ρ_K asymptotically dominates ρ_V of the order a_n locally with respect to ρ_K on $\mathcal{D}$ (denoted by $\rho_K \succeq \rho_V/\rho_K(a_n \to 0)$) iff $(\forall f \in \mathcal{D})$ $(\exists B_K(f))$ such that $(\forall (f_n)_1^\infty \in B_K(f),\ f_n \to f \text{ in } \rho_K)$ $(\exists c > 0)$ in such a way that $\rho_K(f_n, f) \geq c\rho_V(f_n, f) - a_n$, where $B_K(f)$ means Kolmogorov neighborhood of density f and a_n is nonnegative sequence with $\lim_{n \to +\infty} a_n = 0$.

The domination just defined is actually generalization of the original one form Definition 3 (see [7]). Now we state the main theorem the proof of which can be found in [7].

Theorem 3. *Let $\rho_K \succeq \rho_V / \rho_K$ $(a_n \to 0)$ on $\mathcal{D}$, then arbitrary Kolmogorov estimate of a density from $\mathcal{D}$ is consistent in L_1-norm and expected L_1-norm. Moreover if $a_n = o(n^{-1/2})$, then Kolmogorov estimate of a density from $\mathcal{D}$ is consistent of the order $n^{-1/2}$ in L_1-norm and in expected L_1-norm.*

Definition 8. Let $f, g \in \mathcal{D}_\lambda$ and $a \in [0,+\infty]$. Then partial degree of variation $DV_a(f,g) \in [0,+\infty]$ is defined by:

$$DV(f,g) = \inf\left\{ m \in N \cup \{0\} : \; A = \bigcup_{j=1}^{m} \mathcal{J}_j + I, \; A \text{ separates } \nu^+, \; \nu^- \right\},$$

where $\mathcal{J}_1, \ldots \mathcal{J}_m$ are nonvoid disjoint intervals in R and $I \subset \mathsf{R}$ is the set satisfying $\nu^+(I) \le a$ and $\nu^-(I) \le a$ while $\bigcup_{j=1}^{m} \mathcal{J}_j \cap I = \varnothing$. If the minimized set is empty, i.e. if there does not exist any m with desired properties, we set $DV_a(f,g) = +\infty$.

If $a > b > 0$, then it holds $DV_a \le DV_b \le DV_0 = DV$. This partial degree of variation is always less then or equal to previously mentioned degree of variations DV referring us about the number of sign changes in $f - g$. It is not the case of partial degree of variations, but nevertheless we know that finite value $DV_a(f,g) < +\infty$ implies that all the sign changes in $f - g$ are located inside the set I except for a finite number of exceptions.

Theorem 4. *Let $\mathcal{D} \subset \mathcal{D}_\lambda$ and for every density $f \in \mathcal{D}$ there exists Kolmogorov neighborhood $B_K(f)$, constant $K \in [0,\infty)$ and nonnegative sequence a_n with $\lim_{n\to\infty} a_n = 0$ such that $(\forall (f_n)_1^\infty \in B_K(f), f_n \to fv\rho_K)$ it holds that $DV_{a_n}(f_n, f) < K$ for all $\forall n \in \mathbb{N}$. Then ρ_K asymptotically uniformly dominates ρ_V of the order a_n locally with respect to ρ_K on $\mathcal{D}$.*

Now we explore the consistence and the order of consistence of our minimum distance estimates in case of distances differing from Kolmogorov distance. If for a given distance D it holds both inequalities $D \le h_1(\rho_K), \rho_K \le h_2(D)$, where h_1, h_2 are two real functions continuous in zero point with zero value in zero argument and if there exist positive constants K_i such that $h_i(x) \le K_i$ in a neighborhood of zero and moreover the asymptotic domination is fulfilled for family $\mathcal{D} \subset \mathcal{D}_\lambda$, then the minimum D distance estimates of densities from $\mathcal{D}$ are consistent of the order $n^{-1/2}$ in L_1-norm and the expected L_1-norm. (See [7] in details, for specific inequalities derived in spaces of probability densities, see [3].) Above stated conditions are satisfied for example for Lévy and discrepancy distances.

4. Numerical simulation

Unfortunately, it cannot be found from the proofs of above stated theorems how big is the constant of proportionality at the order $n^{-1/2}$ of the asymptotic consistency. To find it and see the range of random sample which the estimates are reliable and reasonably accurate for, we produced numerical simulation. We considered minimum distance estimates with Kolmogorov, Lévy and discrepancy distances which are accompanied by the theoretical results but we also computed minimum Cramer-von Mises distance estimates where we failed to show the desired inequality leading to $n^{-1/2}$ consistency. Six well-known distributions were simulated, the only Normal and Cauchy models are presented here (see [4] and [7] in details).

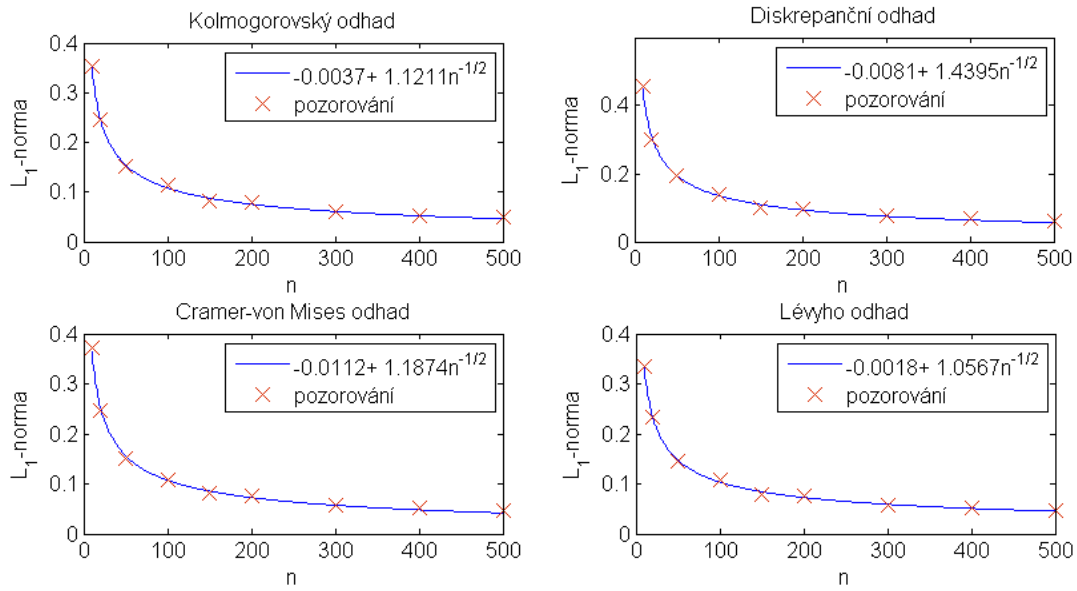


Figure 1: Estimates for Normal distribution with parameters μ=1, σ^2=1, repetitions 500.

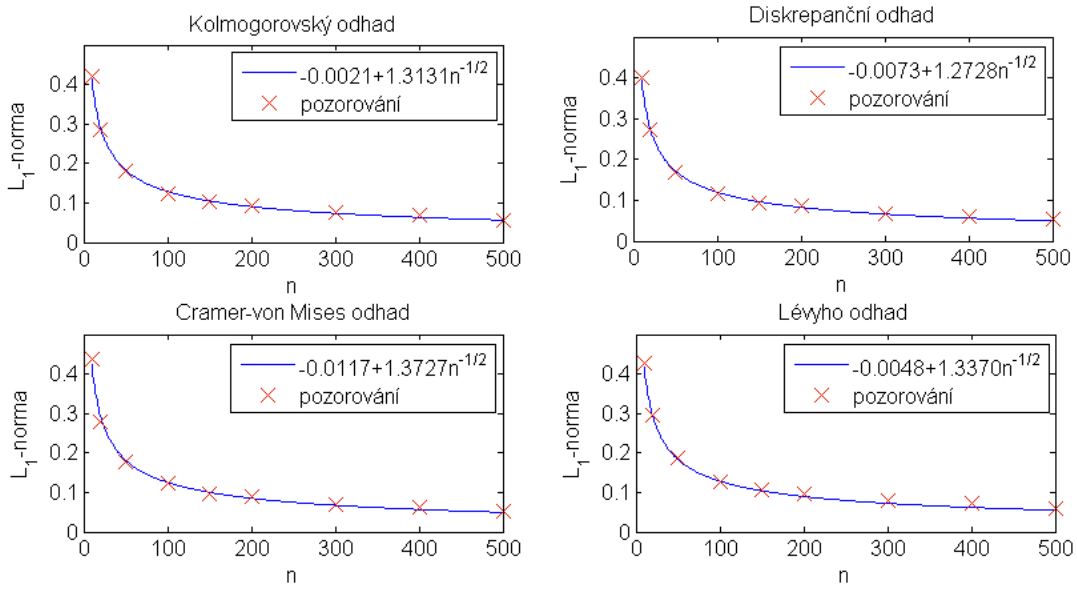


Figure 2: Estimates for Cauchy distribution with parameters a=1,λ=1, repetitions 500.

The functions $g(x) = a_0 + a_1 n^{-1/2}$ were fitted to computed estimates by means of the least squares technique. We can see from Figures 1-2, that data are in very good coincidence with theoretical results already for small sample ranges. The coefficient a_0 is almost zero everywhere as it was expected. The main proportionality constant a_1 varies inside the interval (1, 1.5). Furthermore, we obtained the similarly good results also for the case of minimum Cramer-von Mises distance estimates, which lead us to the idea that these estimates could possess also the same order of consistency $n^{-1/2}$. To verify this, a comprehensive simulation study would be beneficial. In the Tables 1-2 below we present parameter estimates with its variances, distances in L_1-norm and its variances and finally also the Euklidean distances of estimated parameters from the true values of parameters. The best case of parameter estimate in the sense of L_1-norm is emphasized.

Table 1: Normal distribution

n	D	$\overline{\mu}$	$\overline{\sigma^2}$	$\sqrt{S(\mu)}$	$\sqrt{S(\sigma^2)}$	$V(f_0)$	$\sigma_V(f_0)$	$\lvert\mu-\mu_0\rvert$	$\lvert\sigma^2-\sigma_0^2\rvert$
10	K	0.982	0.949	0.308	0.574	0.353	0.202	0.248	0.452
	D	0.968	0.848	0.437	0.578	0.456	0.247	0.359	0.481
	$C-M$	0.832	0.971	0.297	0.581	0.372	0.222	0.280	0.469
	L	0.982	0.976	0.302	0.558	**0.335**	0.191	0.244	0.433
20	K	0.997	0.997	0.235	0.403	0.245	0.130	0.188	0.326
	D	0.991	0.949	0.322	0.405	0.301	0.158	0.257	0.336
	$C-M$	0.922	1.021	0.231	0.404	0.246	0.138	0.195	0.328
	L	0.997	1.015	0.233	0.384	**0.232**	0.118	0.188	0.310
50	K	1.010	0.969	0.150	0.246	0.153	0.079	0.122	0.203
	D	1.019	0.944	0.210	0.258	0.193	0.097	0.172	0.216
	$C-M$	0.978	0.977	0.148	0.241	0.150	0.076	0.121	0.200
	L	1.008	0.982	0.147	0.235	**0.145**	0.075	0.119	0.190
100	K	1.017	0.975	0.117	0.179	0.113	0.061	0.096	0.144
	D	1.019	0.971	0.159	0.194	0.140	0.074	0.131	0.158
	$C-M$	1.001	0.984	0.113	0.177	0.108	0.060	0.092	0.140
	L	1.015	0.977	0.115	0.163	**0.107**	0.056	0.093	0.131
150	K	0.995	0.995	0.085	0.140	0.081	0.047	0.067	0.109
	D	0.991	0.987	0.113	0.149	0.101	0.054	0.091	0.120
	$C-M$	0.985	0.995	0.084	0.138	0.081	0.046	0.067	0.108
	L	0.995	0.991	0.084	0.133	**0.079**	0.046	0.065	0.104
200	K	1.007	0.984	0.080	0.127	0.078	0.042	0.065	0.102
	D	1.005	0.982	0.112	0.135	0.097	0.053	0.090	0.109
	$C-M$	1.000	0.990	0.078	0.125	**0.075**	0.041	0.062	0.099
	L	1.008	0.984	0.078	0.122	**0.075**	0.041	0.063	0.097
300	K	0.998	0.998	0.064	0.097	0.059	0.033	0.050	0.078
	D	0.995	0.997	0.087	0.105	0.075	0.040	0.070	0.085
	$C-M$	0.993	1.001	0.062	0.096	**0.058**	0.032	0.049	0.076
	L	0.996	0.995	0.063	0.092	**0.058**	0.032	0.049	0.071
400	K	1.000	0.990	0.058	0.083	0.053	0.029	0.046	0.067
	D	0.998	0.989	0.080	0.089	0.067	0.035	0.065	0.072
	$C-M$	0.997	0.993	0.056	0.082	**0.051**	0.029	0.044	0.066
	L	0.999	0.988	0.057	0.081	0.052	0.028	0.044	0.066
500	K	1.008	0.994	0.050	0.079	0.049	0.025	0.040	0.064
	D	1.007	0.993	0.069	0.083	0.060	0.031	0.055	0.067
	$C-M$	1.001	0.997	0.049	0.075	**0.047**	0.024	0.039	0.061
	L	1.002	0.992	0.049	0.077	0.048	0.023	0.039	0.061

Table 2: Cauchy distribution

n	D	$\overline{a}$	$\overline{\lambda}$	$\sqrt{S(a)}$	$\sqrt{S(\lambda)}$	$V(f_0)$	$\sigma_V(f_0)$	$\lvert a-a_0\rvert$	$\lvert\lambda-\lambda_0\rvert$
10	K	0.975	1.063	0.685	0.596	0.419	0.214	0.502	0.432
	D	1.006	0.967	0.550	0.536	**0.400**	0.210	0.406	0.406
	$C-M$	0.745	1.026	0.895	0.568	0.436	0.221	0.534	0.418
	L	0.969	1.088	0.714	0.610	0.427	0.218	0.526	0.446
20	K	1.006	1.074	0.413	0.401	0.284	0.138	0.319	0.304
	D	1.012	1.027	0.368	0.350	**0.273**	0.130	0.292	0.270
	$C-M$	0.890	1.057	0.408	0.370	0.279	0.142	0.326	0.280
	L	1.000	1.081	0.429	0.415	0.295	0.142	0.334	0.316
50	K	0.997	1.016	0.253	0.232	0.182	0.095	0.202	0.178
	D	1.001	1.000	0.216	0.216	**0.169**	0.085	0.173	0.168
	$C-M$	0.955	1.011	0.244	0.216	0.176	0.091	0.199	0.167
	L	0.992	1.019	0.268	0.241	0.189	0.101	0.213	0.182
100	K	0.992	1.021	0.173	0.149	0.124	0.065	0.139	0.120
	D	1.003	1.009	0.151	0.149	**0.116**	0.061	0.122	0.118
	$C-M$	0.971	1.016	0.167	0.146	0.122	0.064	0.137	0.116
	L	0.986	1.022	0.180	0.152	0.128	0.069	0.146	0.121
150	K	1.006	0.991	0.138	0.120	0.102	0.053	0.112	0.098
	D	1.007	0.988	0.119	0.110	**0.092**	0.048	0.094	0.091
	$C-M$	0.992	0.989	0.132	0.107	0.096	0.048	0.108	0.088
	L	1.002	0.989	0.147	0.124	0.108	0.055	0.119	0.102
200	K	1.003	1.005	0.128	0.113	0.093	0.048	0.102	0.089
	D	1.003	1.004	0.110	0.110	**0.085**	0.047	0.088	0.087
	$C-M$	0.992	1.005	0.122	0.103	0.088	0.046	0.099	0.082
	L	1.000	1.001	0.132	0.114	0.097	0.049	0.107	0.091
300	K	1.003	0.997	0.099	0.090	0.075	0.038	0.079	0.071
	D	1.008	0.995	0.086	0.085	**0.068**	0.035	0.070	0.068
	$C-M$	0.996	0.995	0.095	0.080	0.070	0.035	0.077	0.064
	L	0.997	0.994	0.105	0.092	0.079	0.041	0.084	0.071
400	K	1.001	1.005	0.090	0.083	0.068	0.035	0.073	0.066
	D	1.001	1.001	0.076	0.079	**0.061**	0.032	0.061	0.063
	$C-M$	0.996	1.002	0.085	0.076	0.064	0.033	0.069	0.061
	L	0.994	1.002	0.096	0.083	0.071	0.036	0.077	0.067
500	K	0.998	1.001	0.074	0.073	0.057	0.032	0.058	0.058
	D	1.001	0.998	0.064	0.069	**0.052**	0.028	0.050	0.056
	$C-M$	0.994	1.000	0.071	0.066	0.054	0.029	0.056	0.054
	L	0.991	0.999	0.079	0.074	0.060	0.034	0.063	0.059

5. References

[1] KŮS, V. 2004. Nonparametric density estimates consistent of the order of $n^{-1/2}$ in the L_1-norm. In: Metrika, 2004, s. 1-14.

[2] DEVROYE, L. – GYÖRFI, L. – LUGOSI, G. 1996. A Probabilistic Theory of Pattern Recognition, New York: SPRINGER, 1996.

[3] GIBBS, A. L. – SU, F. E., 2002. On choosing and bounding probability metrics. In: International Statistical Review,70, 2002, s.419-435.

[4] FRÝDLOVÁ, I. 2004. Odhady pravděpodobnostních hustot s minimální Kolmogorovskou vzdáleností, Diplomová práce, FJFI ČVUT Praha, 2004.

[5] DEVROYE, L. - GYÖRFI, L. 1985. Nonparametric density Estimate, the L_1 – view, New York: WILEY, 1985.

[6] GYÖRFI, L. - VAJDA I. - VAN DER MEULEN, E. 1996. Minimum Kolmogorov Distance Estimates of Parameters and Parametrized Distributions. In: Metrika, 1996, s.237-255.

[7] HANOUSKOVÁ, J. 2009. Asymptotické vlastnosti odhadů s minimální vzdáleností. Diplomová práce, FJFI ČVUT Praha, 2009.

Adresa autorov:

Hanousková Jitka, Ing.
Katedra matematiky FJFI ČVUT
Trojanova 13, 120 00 Praha 2
jitka.meier@centrum.cz

Kůs Václav, Ing. PhD.
Katedra matematiky FJFI ČVUT
Trojanova 13, 120 00 Praha 2
vaclav.kus@fjfi.cvut.cz

Analýza vzájomného vzťahu typu dysfágie a aspirácie a pneumónie
Analysis of interrelation between a type of dysphagia, aspiration and pneumonia

Jozef Chajdiak, Barbora Bunová, Ján Lietava, Daniel Bartko

Abstract: The paper analyses intensity of dysphagia and its relation to predictors of aspiration and aspiration on a sample of 113 patients. The patients, hospitalized at the Neurological Ward of ÚVN-FN in Ružomberok from February until the end of September 2009, were diagnosed with dyspahgia and aspiration by means of a specific method called the videofluoroscopy. Pneumonia was confirmed by the chest x-ray method. The Excel (2007 version) was used for analytical calculations.

Key words: dysphagia, aspiration, pneumonia, frequency, pivot table.

Kľúčové slová: dysfágia, aspirácia, pneumónia, triedenie, kontingenčná tabuľka.

1. Úvod

Prehĺtanie je vysoko komplexný proces, ktorý si vyžaduje presnú súhru asi 50 párov svalov, 5 hlavových a troch cervikálnych nervov. Porucha prehĺtania je dysfágia. Komplikáciou dysfágie okrem iného je aj aspirácia a pneumónia. Aspirácia je preniknutie tekutiny, jedla do dýchacích ciest. Pneumónia je zápal pľúc, jednou z jej príčin môže byť aj aspirácia.

Dysfágiu a aspiráciu sme diagnostikovali prostredníctvom špeciálnej vyšetrovacej metódy - videofluoroskopie. Je to roentgenologické vyšetrenie prehĺtania, ktoré sa zaznamenáva analogicky na videozáznam a následne analyzuje.

V tejto súvislosti sme analyzovali súbor 113 vlastných pacientov s neurologickými ochoreniami v priebehu mesiacov február až september 2009.

2. Hodnoty analyzovaných premenných

Analyzované premenné boli nominálneho typu a mohli nadobudnúť nasledujúce hodnoty:

TYP DYSFÁGIE: 0 – žiadna 1 – ľahká 2 – stredná 3 – ťažká	**PREDIKTORY ASPIRÁCIE** 0 – žiaden 1 – kašeľ po prehltnutí 2 – dysfónia 3 – zmena kvality hlasu po prehltnutí 4 – dyzartia 5 – abnormálny vôľový kašeľ 6 – abnormálny dávivý reflex
ASPIRÁCIA 0 – žiadna 1 – penetrácia 2 – predeglutinačná 3 – intradeglutinačná 4 – postdeglutinačná 5 – tichá	**PNEUMÓNIA** 0 – nie 1 – áno

3. Dysfágia a prediktory aspirácie

V tabuľke 3.1 sú výsledky dvojstupňového triedenia medzi typom dysfágie a prediktormi aspirácie. Tabuľka obsahuje absolútne triedne početnosti a celkové percentuálne triedne početnosti. Z výsledkov v tabuľke môžeme konštatovať, že:

- 10% respondentov nemá dysfágiu ani žiaden prediktor aspirácie,
- najintenzívnejším preditktom na dysfágiu je dyzartia,
- kontingenčný koeficient 0,658 svedčí o vyššej intenzite asociácie.

Tabuľka 3.1 Výsledky dvojstupňového triedenia podľa typu dysfágie a prediktorov aspirácie (absolútne a percentuálne celkové početnosti)

Prediktory aspirácie	Údaje	Typ dysfágie				
		žiadna	ľahka	stredna	Ťažká	Spolu
žiaden	Počet	19	27	2		48
	% z celku	10%	14%	1%		26%
kašeľ po prehltnutí	Počet		6	12	8	26
	% z celku		3%	6%	4%	14%
dysfónia	Počet		6	16	8	30
	% z celku		3%	9%	4%	16%
zmena kvality hlasu po prehltnutí	Počet		1	12	5	18
	% z celku		1%	6%	3%	10%
dyzartria	Počet	1	25	15	4	45
	% z celku	1%	13%	8%	2%	24%
abnormálny vôľový kašeľ	Počet			1	2	3
	% z celku			1%	1%	2%
abnormálny dávivý reflex	Počet		4	9	4	17
	% z celku		2%	5%	2%	9%
Spolu		20	69	67	31	187
% z celku		11%	37%	36%	17%	100%

Dysfágia a aspirácia

V tabuľke 4.1 sú výsledky dvojstupňového triedenia medzi typom dysfágie a aspirácie. Tabuľka obsahuje absolútne triedne početnosti a celkové percentuálne triedne početnosti. Z výsledkov v tabuľke môžeme konštatovať, že:

- šestina respondentov nemá dysfágiu ani žiadnu aspiráciu,
- pri aspirácii sa znak žiaden vyskytuje najčastejšie (59 %),
- najintenzívnejší vplyv na pokročilú dysfágiu má tichá aspirácia,
- kontingenčný koeficient 0,766 svedčí o vysokej intenzite asociácie medzi typom dysfágie a aspirácie.

Tabuľka 4.1 Výsledky dvojstupňového triedenia podľa typu dysfágie a aspirácie (absolútne a percentuálne celkové početnosti)

		Typ dysfágie				
Aspirácia	Údaje	žiadna	ľahka	stredna	ťažká	Spolu
žiaden	Počet	20	48	2		70
	% z celku	17%	40%	2%		59%
	Počet		7	9	1	17
penetrácia	% z celku		6%	8%	1%	14%
	Počet			4	2	6
predeglutinačná	% z celku			3%	2%	5%
	Počet			3	7	10
intradeglutinačná	% z celku			3%	6%	8%
	Počet			3		3
postdeglutinačná	% z celku			3%		3%
	Počet			7	6	13
tichá	% z celku			6%	5%	11%
Spolu		20	55	28	16	119
% z celku		17%	46%	24%	13%	100%

4. Dysfágia a pneumónia

V tabuľkách 5.1 až 5.4 sú výsledky dvojstupňového triedenia medzi typom dysfágie a pneumóniou. Prvá tabuľka obsahuje absolútne triedne početnosti a v ďalších troch tabuľkách je k nim priradená verzia celkových percentuálnych relatívnych početností, riadkových percentuálnych relatívnych početností a stĺpcových percentuálnych relatívnych početností. Z výsledkov v tabuľkách môžeme konštatovať, že:

- pneumónia sa vyskytuje len u strednej a ťažkej dysfágie,
- 20 zo 113 osôb nemá dysfágiu ani pneumóniu,
- 55 zo 113 (prakticky polovica) má ľahkú dysfágiu bez pneumónie.

Tabuľka 5.1 Výsledky dvojstupňového triedenia podľa typu dysfágie a pneumónie (absolútne početnosti)

	Typ dysfágie				
Pneumónia	žiadna	ľahká	stredná	ťažká	Spolu
Nie	20	55	16	1	92
Áno			11	10	21
Spolu	20	55	27	11	113

Tabuľka 5.2 Výsledky dvojstupňového triedenia podľa typu dysfágie a pneumónie (absolútne a percentuálne celkové početnosti)

		Typ dysfágie				
Pneumónia	Údaje	žiadna	ľahka	stredna	Ťažká	Spolu
Nie	Počet	20	55	16	1	92
	%	18%	49%	14%	1%	81%
Áno	Počet			11	10	21
	%			10%	9%	19%
Spolu		20	55	27	11	113
Spolu %		18%	49%	24%	10%	100%

Tabuľka 5.3 Výsledky dvojstupňového triedenia podľa typu dysfágie a pneumónie (absolútne a percentuálne riadkové početnosti)

		Typ dysfágie				
Pneumónia	Údaje	žiadna	ľahka	Stredna	Ťažká	Spolu
Nie	Počet	20	55	16	1	92
	%	22%	60%	17%	1%	100%
Áno	Počet			11	10	21
	%			52%	48%	100%
Spolu		20	55	27	11	113
Spolu %		18%	49%	24%	10%	100%

Tabuľka 5.4 Výsledky dvojstupňového triedenia podľa typu dysfágie a pneumónie (absolútne a percentuálne stĺpcové početnosti)

		Typ dysfágie				
Pneumónia	Údaje	žiadna	ľahka	stredna	ťažká	Spolu
Nie	Počet	20	55	16	1	92
	%	100%	100%	59%	9%	81%
Áno	Počet			11	10	21
	%			41%	91%	19%
Spolu		20	55	27	11	113
Spolu %		100%	100%	100%	100%	100%

6. Záver

Videofluroskopickým vyšetrením sme zistili, že zo 113 pacientov s neurologickými chorobami má dysfágiu 93 pacientov, z toho 55 ľahkú, 27 strednú a 11 ťažkú. Ľahkú dysfágiu ovplyvňuje najviac dyzatria, penetrácia. Strednú dysfágiu ovplyvňuje najviac dysfónia a opäť penetrácia. Ťažkú dysfágiu ovplyvňuje najviac kašel po prehlnutí, dysfónia, intradeglutinačná aspirácia a pneumónia.

7. Literatúra

[1]BARTOLOME, G., SCHROETTER-MORASCH, H.(Hrsg.) 2006. Schluckstoerungen. Diagnostik und Rehabilitation. Urban/Fischer, Muenschen. ISBN-10: 3- 437-47160-0

[2]Bartolome, G. 2003. Funktionelle Dysphagie– Therapie. Kurz pre klinických logopédov. Viedeň 2003

[3]BASSOTI,U., PAGLIARICCI, S., et al. 1998. Esophageal manometric abnormalities in Parkinson s disease. Dysphagia , volume 13, p. 28-31

[4]BIGENZAHN,W., DENK, M.D. 1999. Aetiologie, Klinik, Diagnostik und Therapie von Schluckstoerungen. Georg Thieme Verlag, Stuttgart. ISBN 3-13-115991-X

[5]BUNOVÁ, B., BARTKO,D.2006.Deglutinačné poruchy v starobe. Geriatria č. 1, p.30-36

[6]BUNOVÁ,B.,Tedla,M. 2006. Terapia dysfágie u pacientov po operáciách v oblasti hlavy a krku. Choroby hlavy a krku č.2., p. 9-13. ISSN 1210-0447

[7]BUNOVÁ, B. TEDLA, M. 2009. Špecializované vyšetrenia hltacieho aktu. In: Tedla, M. a kol.: Poruchy polykání.Tobiáš. ISBN 978-80-7311-105-2

[8]DANIELS, S.K., BRAILEY, K., PRIESTLY, D. H. et al. 2000: Clinical predictors of dysphagia and aspiration risk: outcome measures in acute stroke patients. Archives of Physical and Medical Rehabilitation, volume 81,p. 1030- 1033

[9]CHAJDIAK, J. 2003. Štatistika jednoducho. Bratislava: Statis 2003. 194 s. ISBN 80-85659-28-X.

[10] CHAJDIAK, J. 2003. Štatistické úlohy a ich riešenie v Exceli. Bratislava: Statis 2005. 262 s. , ISBN 80-85659-39-5.

Adresa autorov:

Jozef Chajdiak, Doc., Ing., CSc.
Ústav manažmentu STU
Vazovova 5
Bratislava
chajdiak@statis.biz

Barbora Bunová, Dr.
ÚVN-SNP-FN
Neurologická klinika
Ružomberok
bunova@stonline.sk

Ján Lietava, Doc., Mudr., PhD.
II. Interná klinika Lekárskej fakulty UK
Mickiewiczova 13
Bratislava
jan.lietava@yahoo.com

Daniel Bartko, Prof., Mudr., DrSc.
Riaditeľ ústavu medicínskych vied,
neurovied a vojen. zdravotníctva
ÚVN-SNP-FN, Ružomberok
bartkod@uvn.sk

Jak se jmenovali, jak se jmenujeme aneb Křestní jména kdysi a dnes
Frequency of the Christian Names in the Chrudim Area in the Middle of the 17th Century

Eva Kačerová

Abstract: One of the most valuable sources for Czech historical demography is the List of Serfs according to Faith of 1651, giving information not only on the age and family structure of the population but also on the names of the mentioned persons. The presented paper deals with the frequency of the christian names of the cerfs included in the List in the Chrudim area in the middle of the seventeenth century. This article came into being within the framework research project IGA VŠE 15/08.

Key words: Christian names, 17th century, 21st century, List of Serfs, Chrudim area

Klíčová slova: Křestní jména, 17. století, 21. století, Soupis poddaných podle víry, Chrudimsko

1. Úvod

Pramenem pro tento článek byl Soupis poddaných podle víry z roku 1651, analyzovanou oblastí je Chrudimsko. Chrudimsko v polovině 17. století představovalo 40 panství (Bělá, Bítovany, Bylany, Březovice, Čankovice, Hrady Nové, Choceň, Choltice, Chrast, Chroustovice, Chrudim, Bystrá a Čankovice, Košumberk, Lanškroun, Lhota, Líbanice, Litomyšl, Městec Heřmanův a Stolany, Morašice, Vysoké Mýto, Nasavrky, Orel, Pardubice, Polička, Přestavlky, Rosice, Rychmburk, Seč, Skuteč, Slatiňany, Svojano, Trojovice, Třibřichy, Hrochův Týnec, Uhersko, Vlčnov, Zaječice, Zdechovice, Žumberk, České Heřmanice). Byla mezi nimi jak panství komorní, která patřila králi, tak i panství patřící šlechticům a rytířům, stejně tak i obce, které spadaly do vlastnictví měst a i rychty svobodníků.

Jednotlivá panství se lišila svou velikostí. Ve Vlčnově bylo evidováno 9 osob, v Pardubicích bylo zapsáno 8 452 osob, přičemž děti mladší 12 let byly evidovány výjimečně. Celkem na Chrudimsku bylo zapsáno 46 626 osob, přičemž u 1 958 na panství Bystrá a Čankovice, které patřilo Jiřímu Adamovi Bořitovi hraběti z Martinic nebyl věk zaznamenán vůbec.

Kromě jména a v omezené míře i příjmení byly Soupisem zjišťovány následující údaje: stav (panský, rytířský, poddanský), věk (zcela jistě se projevuje zaokrouhlování uváděného věku na násobky 5 či 10), zaměstnání a vyznání. Zjištění údaje o náboženském vyznání, zda-li daná osoba je či není katolík, a pokud není katolík, zda-li u ní je či není naděje na obrácení na víru katolickou, bylo hlavním důvodem pořizování Soupisu.

Křestní jména ze 17. století jsou srovnávána se jmény, která se vyskytují v současné populaci žijící na území celé České republiky.

2. Křestní jména v 17. století

Křestní jméno, které každý člověk dostává záhy po svém narození, bylo až do zavedení stabilních příjmení v roce 1771 vlastně jediným trvalým oficiálním označením jedince.

Během jednotlivých staletí se užívalo 100 až 500 rozličných křestních jmen, jejichž skladba se měnila a často byla pro určitou dobu typická. Z tohoto souboru se 30–50 % vyskytovalo vzácně. Jména označovaná jako oblíbená jsou ta, která přesahují 10% podíl,

jména, jejichž výskyt se v dané populaci pohybuje mezi 3–10 %, jsou označovaná za běžná a jména, jež dostává 1–2 % narozených, považujeme za obvyklá [2].

Při posuzování četnosti výskytu jednotlivých jmen byly brány jak české tak německé ekvivalenty a jejich případné podoby dohromady (zvláště jim bude věnována pozornost později). Bohužel ne všem záznamům v kolonce jméno se podařilo přiřadit nějaké konkrétní jméno. K takovým patří třeba mužské jméno Faltys či Faltin a jeho obměny, které se na eggenbergském panství vyskytlo 43krát na panství Litomyšl a Lanškroun. Celkem se takto nepodařilo zjistit 243 jmen, což je 0,5 %. Více nejasných záznamů se týká mužských jmen – zde se podařilo určit 98,8 %, zatímco u ženských jmen zůstalo neurčeno pouze 0,04 % jmen. Ještě je třeba vzít v potaz, že u některých osob nebylo křestní jméno uvedeno vůbec. Takových bylo celkem 133, z nich 107 mezi ženami.

Je zřejmé, že rozličná křestní jména byla uváděna v různých variantách, tak jak je písař zaznamenal. Pouze na panství Bělá byl Soupis zpracován v němčině, na ostatních panstvích písaři použili češtinu. Nejvíce různých variant téhož jména můžeme zaznamenat u jména Václav. Na 2 501 nositelů tohoto jména připadá 14 rozličných podob zápisu, přičemž 2 446 jich je zapsáno jako Václav. Vykytovaly se i zápisy v domácké podobě Vašek a i různých z němčiny vycházejících formách jako např. Wentzel. Tyto německé formy byly zaznamenány nejenom na panství Bělá, kde byl Soupis veden německy, ale i na jiných panstvích, na nichž byl Soupis pořízen v češtině. Lze se tedy domnívat, že takto zapsaní Václavové používali němčinu jako běžný jazyk.

Tabulka 3: Absolutní a relativní četnost křestních jmen žen, Chrudimsko 1651

Anna	4 643	18,68	Regina	126	0,51	Františka	3	0,01
Kateřina	3 965	15,95	Alena	101	0,41	Kunhuta	3	0,01
Dorota	3 917	15,76	Judita	74	0,30	Polyxena	3	0,01
Magdalena	2 187	8,80	Veronika	53	0,21	Anastázie	2	0,01
Mariana	1 477	5,94	Sibyla	52	0,21	Libuše	2	0,01
Ludmila	1 426	5,74	Johana	37	0,15	Meluzína	2	0,01
Alžběta	1 011	4,07	Klára	28	0,11	Sára	2	0,01
Salomena	751	3,02	Apolena	23	0,09	Agáta	1	0,00
Marie	745	3,00	Ester	22	0,09	Eleonora	1	0,00
Zuzana	671	2,70	Lucie	21	0,08	Eufemie	1	0,00
Eva	660	2,65	Helena	18	0,07	Isabela	1	0,00
Markéta	446	1,79	Sabina	16	0,06	Jana	1	0,00
Barbora	442	1,78	Eliška	14	0,06	Konstancie	1	0,00
Gertruda	342	1,38	Hedvika	13	0,05	Lukrécie	1	0,00
Voršila	324	1,30	Kordula	11	0,04	Rebeka	1	0,00
Rozina	231	0,93	Žofie	11	0,04	Sidonie	1	0,00
Justýna	221	0,89	Juliana	8	0,03	Vanda	1	0,00
Walpurga	161	0,65	Felicia	6	0,02	nezjištěno	10	0,04
Anežka	157	0,63	Emilie	4	0,02	neuvedeno	107	0,43
Kristýna	153	0,62	Benigna	3	0,01	**Celkem**	**24 860**	**100,00**
Marta	142	0,57	Bohumila	3	0,01			

Nejčetnějším ženským jménem byla Anna (4 543, tj. 18,7 %), druhým nejčetnějším jménem byla Kateřina (3 965, tj. 15,95 %), posledním oblíbeným jménem byla Dorota (3 917, 15,76 %). Magdalena, Mariana (nelze odlišit, zdali to bylo samostatné jméno, nebo složenina ze jmen Marie a Anna), Ludmila, Alžběta, Salomena, Marie byly jmény běžnými s četností

výskytu 3–9 %. Zuzana, Eva, Markéta, Barbora, Gertruda a Voršila byly jmény obvyklými. Ostatních 34 ženských jmen se objevuje u méně než 1 % nositelek. 10 jmen se vyskytlo pouze jedenkrát (tab.1).

Tabulka 2: Absolutní a relativní četnost křestních jmen mužů, Chrudimsko 1651

Jan	4 190	19,25
Jiří	2 613	12,00
Václav	2 501	11,49
Matěj	1 383	6,35
Jakub	1 292	5,94
Martin	1 231	5,66
Pavel	1 049	4,82
Mikuláš	626	2,88
Tomáš	584	2,68
Matouš	560	2,57
Ondřej	475	2,18
Adam	408	1,87
Petr	387	1,78
Vavřinec	352	1,62
Bartoloměj	280	1,29
Lukáš	277	1,27
Šimon	276	1,27
Urban	229	1,05
Řehoř	223	1,02
Vít	194	0,89
Michal	166	0,76
Daniel	152	0,70
Marek	151	0,69
Havel	142	0,65
Kryštof	139	0,64
Tobiáš	112	0,51
Kašpar	106	0,49
Jindřich	105	0,48
Samuel	99	0,45
Štěpán	97	0,45
Valentin	86	0,40
Benedikt	81	0,37
Mat…	62	0,28
Filip	43	0,20
Stanislav	43	0,20
Jiljí	40	0,18
David	37	0,17
Fridrich	34	0,16
Matyáš	33	0,15
Vojtěch	32	0,15
Blažej	31	0,14
Nikodém	30	0,14
Eliáš	28	0,13
Burian	27	0,12
Baltazar	25	0,11
Karel	25	0,11
Viktorin	22	0,10
Melichar	21	0,10
Antonín	20	0,09
Prokop	20	0,09
Felix	19	0,09
Bohuslav	18	0,08
František	18	0,08
Jeroným	18	0,08
Vilém	18	0,08
Zachariáš	17	0,08
Diviš	16	0,07
Zikmund	15	0,07
Šebestián	14	0,06
Rudolf	13	0,06
Ambrož	12	0,06
Kryšpín	12	0,06
Jeremiáš	10	0,05
Wolfgang	10	0,05
Augustin	9	0,04
Gabriel	9	0,04
Hanuš	9	0,04
Jáchym	9	0,04
Josef	9	0,04
Kristián	9	0,04
Albrecht	7	0,03
Duchoslav	7	0,03
Ferdinand	6	0,03
Hynek	6	0,03
Ludvík	6	0,03
Severin	6	0,03
Abraham	4	0,02
Bernard	4	0,02
Jonáš	4	0,02
Šalamoun	4	0,02
Alexandr	3	0,01
Dominik	3	0,01
Jaroslav	3	0,01
Klement	3	0,01
Benjamín	2	0,01
Ezechiel	2	0,01
Florián	2	0,01
Heřman	2	0,01
Job	2	0,01
Zdeněk	2	0,01
Absolon	1	0,00
Adrian	1	0,00
Aleš	1	0,00
Ámos	1	0,00
Budislav	1	0,00
Cyprián	1	0,00
Dětřich	1	0,00
Eustach	1	0,00
Ferdinad	1	0,00
Ignác	1	0,00
Izaiáš	1	0,00
Konrád	1	0,00
Leonard	1	0,00
Leopold	1	0,00
Linhart	1	0,00
Maxmilián	1	0,00
Neftalín	1	0,00
Noe	1	0,00
Oldřich	1	0,00
Otmar	1	0,00
Quirin	1	0,00
Richard	1	0,00
Samson	1	0,00
Teofil	1	0,00
Timotej	1	0,00
Zacheus	1	0,00
nezjištěno	233	1,07
neuvedeno	26	0,12
Celkem	**21 766**	**100,00**

Variabilita mužských jmen byla v polovině 17. století ve srovnání s ženskými jmény vyšší – na Chrudimsku je zapsáno 97 mužských jmen a jen 59 ženských jmen. Nejčetnějším mužským jménem v té době bylo jméno Jan (4 190, tj. 19,25 %), druhým nejčastějším byl Jiří (2 613, tj. 12,00 %), třetím oblíbeným mužským jménem byl Václav (2 501, 11,49 %) Jména, Matěj, Jakub, Martin, Pavel byla jmény běžnými s výskytem mezi 4–7 %. 11 mužských jmen (Mikuláš, Tomáš, Matouš, Ondřej, Adam, Petr, Vavřinec, Bartoloměj, Lukáš, Šimon, Urban a Řehoř) patřilo mezi jména obvyklá a zbývajících 79 jmen se vyskytlo méně než 217krát, tedy v méně než 1 % případů, přičemž 26 z nich se vyskytlo pouze jedenkrát (tab. 2).

Je možné sledovat případné generační změny v oblibě jednotlivých jmen. Při tomto přístupu si budeme všímat jmen, která se nevyskytovala příliš často, protože právě u nich lze předpokládat, že teprve „přicházejí do módy, nebo naopak“. Nejjednodušším možným kritériem pro postižení generačních změn v oblibě nějakého jména je, pokud vycházíme ze Soupisu pořizovaného k určitému okamžiku, porovnání průměrného věku nositelek/-ů konkrétního jména s průměrným věkem všech žen/mužů.

Tabulka 3: Průměrný věk podle křestního jména muže, Chrudimsko 1651

Abraham	33,67	Ezechiel	29,00	Klement	26,00	Richard	45,00
Absolon	31,00	Felix	35,74	Konrád	14,00	Rudolf	18,92
Adam	33,07	Ferdinad	66,00	Kristián	31,78	Řehoř	32,64
Adrian	30,00	Ferdinand	18,20	Kryšpín	22,58	Samson	11,00
Albrecht	24,71	Filip	29,73	Kryštof	32,42	Samuel	27,21
Aleš	34,00	Florián	30,50	Leonard	44,00	Severin	28,00
Alexandr	25,33	František	15,67	Leopold	15,00	Stanislav	28,26
Ambrož	38,30	Fridrich	28,07	Linhart	30,00	Šalamoun	13,33
Ámos	21,00	Gabriel	23,63	Ludvík	20,50	Šebestián	38,21
Antonín	26,60	Hanuš	35,67	Lukáš	28,09	Šimon	34,27
Augustin	22,63	Havel	33,95	Marek	31,63	Štěpán	36,08
Baltazar	33,74	Heřman	23,00	Martin	30,93	Teofil	20,00
Bartoloměj	32,61	Hynek	48,00	Mat	32,42	Timotej	33,00
Benedikt	32,45	Ignác	18,00	Matěj	29,34	Tobiáš	32,64
Benjamín	17,50	Izaiáš	40,00	Matouš	31,25	Tomáš	31,16
Bernard	19,00	Jáchym	27,89	Matyáš	24,28	Urban	33,08
Blažej	38,52	Jakub	29,83	Maxmilián	7,00	Václav	30,19
Bohuslav	31,53	Jan	30,28	Melichar	28,00	Valentin	32,41
Budislav	21,00	Jaroslav	32,67	Michal	31,65	Vavřinec	31,19
Burian	34,07	Jeremiáš	30,30	Mikuláš	31,65	Viktorin	45,81
Cyprián	38,00	Jeroným	37,06	Nikodém	31,30	Vilém	27,11
Daniel	30,77	Jiljí	29,69	Noe	26,00	Vít	32,25
David	31,54	Jindřich	24,02	Oldřich	38,00	Vojtěch	28,75
Dětřich	15,00	Jiří	30,71	Ondřej	31,81	Wolfgang	24,40
Diviš	39,94	Job	19,50	Otmar	60,00	Zachariáš	29,44
Dominik	44,67	Jonáš	30,75	Pavel	29,44	Zacheus	35,00
Duchoslav	38,00	Josef	33,44	Petr	31,21	Zdeněk	49,50
Eliáš	33,76	Karel	22,04	Prokop	40,27	Zikmund	25,67
Eustach	20,00	Kašpar	29,69	Quirin	14,00		

Průměrný věk žen zaznamenaných na Chrudimsku byl 28,7 roku. Nejčastěji vyskytující se Anny měly průměrný věk právě shodný s průměrným věkem všech žen. Jméno Justýna se vyskytuje celkem 211krát na 13 panstvích. Průměrný věk nositelek tohoto jména je 23,7 roku, můžeme tedy říci, že je to jméno, které „bylo moderní“, neboť se s ním setkáme spíše u mladších dívek. Nejstarší byla 55letá vdova z Litomyšle. Obliba tohoto jména byla nepochybně regionální. V některých oblastech bychom Justýnu hledali marně, v Chrudimi, kde celkem žilo 1 001 žen, bychom našli pouze čtyři nositelky tohoto jména, Na lanškrounském panství představovaly nositelky tohoto jména 2,4 % žen. Ani jednou bychom mezi Justýnami nenašli příslušnice dvou generací z téže rodiny. Naopak se zdá, že např. jméno Marta „vycházelo z módy“. Na Chrudimsku jich v době pořízení Soupisu žilo 133, jejich průměrný věk byl 35,1 roku.

Průměrný věk mužů na Chrudimsku byl 31,2 roku. A tak jméno Karel (24 výskytů) bychom mohli považovat za jméno moderní, neboť průměrný věk jeho nositelů byl pouze 22 let. Nejstarším byl 60letý pekař z Košumberka, většinou však nositelé tohoto jména byli ještě v synovském postavení.

Tabulka 4: Průměrný věk podle křestního jména ženy, Chrudimsko 1651

Alena	27,49	Eufemie	48,00	Klára	26,18	Meluzína	24,00
Alžběta	25,82	Eva	28,40	Konstancie	40,00	Polyxena	20,33
Anastázie	16,50	Felicia	27,83	Kordula	42,27	Regina	30,80
Anežka	28,62	Františka	19,67	Kristýna	28,79	Rozina	24,05
Anna	28,69	Gertruda	31,59	Kunhuta	26,50	Sabina	22,63
Apolena	27,10	Hedvika	37,54	Libuše	30,50	Salomena	28,12
Barbora	31,43	Helena	38,36	Lucie	26,15	Sára	32,50
Benigna	21,67	Isabela	14,00	Ludmila	28,09	Sibyla	27,73
Bohumila	39,33	Jana	46,00	Lukrécie	18,00	Sidonie	20,00
Dorota	27,94	Johana	29,57	Magdalena	28,12	Vanda	31,00
Eleonora	16,00	Judita	24,68	Mariana	27,04	Veronika	24,27
Eliška	33,33	Juliana	28,43	Marie	27,95	Voršila	31,79
Emilie	25,75	Justýna	23,74	Markéta	29,84	Walpurga	29,28
Ester	31,63	Kateřina	28,07	Marta	35,10	Zuzana	26,73
						Žofie	29,00

3. Křestní jména v současnosti

A k jaké změně ve volbě jmen došlo během staletí? Z údajů Ministerstva vnitra ČR [3] vyplývá, že nejčastějším mužským jménem v České republice je, možná poněkud překvapivě vzhledem k vžité představě, jméno Jiří a s mírným odstupem následuje Jan a překvapivě na třetím místě Petr (tab. 5). Se jmény Řehoř a Vít se mezi současníky setkáme spíše výjimečně a Matouš, ač nepatří mezi 50 nejčastěji se vyskytujících jmen, se zase začíná objevovat u nejmladších chlapců. U současné populace poměrně oblíbené jméno Josef, pravda spíše mezi příslušníky starších generací, jak o tom svědčí pokles četnosti výskytu tohoto jména mezi lety 2002 a 2009, bychom na Chrudimsku našli pouze u 9 mužů. Mezi ženami je nejčastější jméno Marie (tab. 6), které si spojujeme spíše se staršími ženami, a které se na počátku tohoto století zase začíná mezi malými dívkami objevovat. Hanu ani Věru bychom na Chrudimsku v polovině 17. století nepotkali. Určitým překvapením u současné populace četnost výskytu jména Ludmila – nejoblíbenějším bylo toto jméno v 2. polovině 50. let, kde jej ročně dostávaly téměř 3 tisíce narozených dívek.

Tabulka 5: Křestní jména mužů, ČR, k 22.4.2002 a 1.5.2009

Pořadí	Jméno	Četnost	Pořadí	Jméno	Četnost
1	JIŘÍ	328 884	1	JIŘÍ	315 369
2	JOSEF	294 405	2	JAN	295 627
3	JAN	293 121	3	PETR	272 837
4	PETR	267 601	4	JOSEF	249 922
5	JAROSLAV	223 442	5	PAVEL	207 100
6	PAVEL	207 300	6	JAROSLAV	198 664
7	MIROSLAV	176 337	7	MARTIN	181 764
8	FRANTIŠEK	176 215	8	TOMÁŠ	168 136
9	MARTIN	164 534	9	MIROSLAV	164 626
10	ZDENĚK	152 742	10	FRANTIŠEK	147 163
11	VÁCLAV	152 582	11	ZDENĚK	140 601
12	TOMÁŠ	146 668	12	VÁCLAV	137 247
13	KAREL	135 363	13	MICHAL	118 915
14	MILAN	119 532	14	KAREL	118 844
15	MICHAL	108 286	15	MILAN	116 251

Tabulka 6: Křestní jména žen, ČR, k 22.4.2002 a 1.5.2009

Pořadí	Jméno	Četnost	Pořadí	Jméno	Četnost
1	MARIE	385 985	1	MARIE	316 559
2	JANA	276 123	2	JANA	274 303
3	ANNA	164 362	3	EVA	160 317
4	EVA	163 238	4	HANA	150 573
5	HANA	152 215	5	ANNA	148 688
6	VĚRA	140 109	6	VĚRA	124 871
7	LENKA	116 662	7	LENKA	119 366
8	ALENA	111 924	8	KATEŘINA	110 936
9	JAROSLAVA	101 920	9	ALENA	110 046
10	LUDMILA	99 583	10	LUCIE	103 702
11	PETRA	98 716	11	PETRA	102 321
12	KATEŘINA	95 623	12	JAROSLAVA	94 475
13	HELENA	90 548	13	LUDMILA	85 939
14	LUCIE	90 275	14	HELENA	82 985
15	ZDEŇKA	85 415	15	MARTINA	81 543

4. Závěr

Četnost výskytu křestních jmen se během staletí změnila. Volba křestního jména závisela na jménu rodičů, kmotrů nebo podle světce, který měl svátek blízko dne narození či křtu dítěte. Tyto závislosti jsou bohužel zpracovávaným pramenem nepostižitelné, neboť vyžadují individuální údaje o narozených, které by bylo možné získat pouze z křestních matrik. Děti nezřídka dostávaly při křtu více jmen, avšak v rodině pak byly označovány jen jedním z nich.

5. Literatura

[1] Soupis poddaných podle víry – Chrudimsko. Archivní pramen
[2]Heřmánková, M.: Demografický vývoj únětické farnosti v 18. století. In: Historická demografie 24/2000 s. 94.
[3]http://www.mvcr.cz/statistiky/jmena/

Adresa autora:

Eva Kačerová, RNDr.
nám. W. Churchilla 4
130 67 Praha
kacerova@vse.cz

Ekonomická aktivita cizinců v České republice
Economic activity of foreigners in the Czech Republic

Eva Kačerová

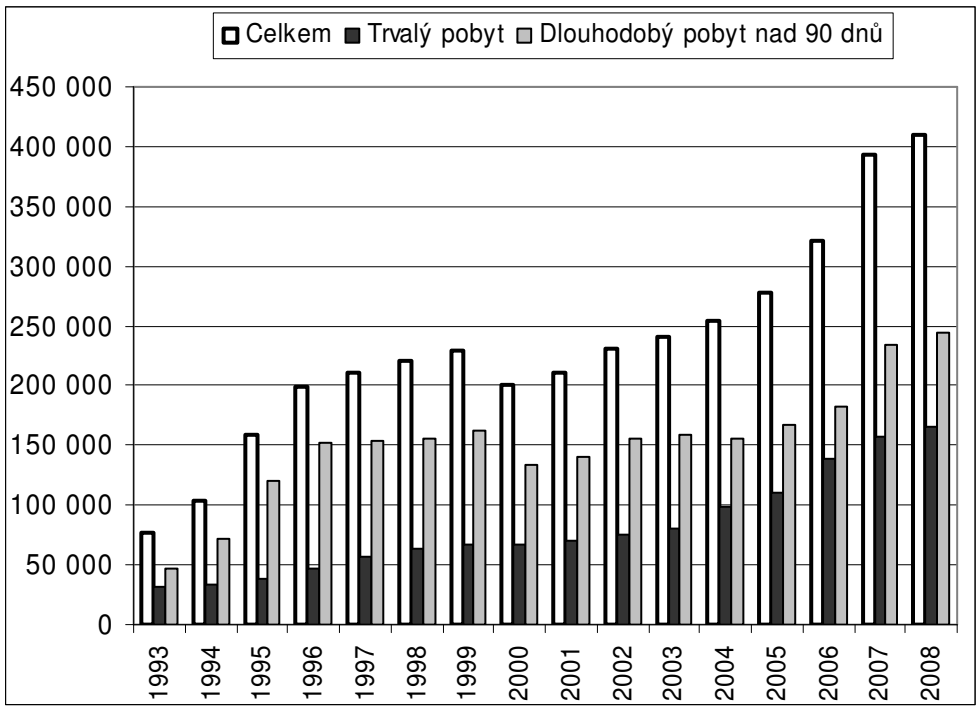
Abstract: The number of long-term or permanently residing foreigners in the Czech Republic exceeded in the 2008 the amount of 410 000. More then 50 % of them are in the age of economic activity. This paper is focused on economic activity of foreigners in the Czech Republic. This article came into being within the framework of the long-term research project 2D06026, "Reproduction of Human Capital", financed by the Ministry of Education, Youth and Sport within the framework of National Research Program II.

Key words: Economic activity, foreigners, Czech Republic

Klíčová slova: ekonomická aktivita, cizinci, Česká republika

1. Úvod

Od roku 1991 je Česká republika zemí migračně ziskovou, s výjimkou roku 2000, avšak toto bylo způsobeno změnou metodiky. Na počátku 90. let převažovala návratová migrace imigrantů převážně z konce 70. let. V období kolem rozdělní federace jsme mohli pozorovat kompenzační migraci. Ve druhé polovině 90. let se počet imigrantů-cizinců do ČR ustálil ročně na zhruba 10 tisících osob ročně. Česká republika se postupně ze země tranzitní měnila v zemi cílovou. Věková struktura cizinců žijících v České republice déle než jeden rok se podstatně liší od věkové struktury populace ČR. Největší podíl cizinců představují občané Ukrajiny, Slovenska, Vietnamu, Polska a Ruské federace.



Obr. 1: Počet cizinců s trvalým a dlouhodobým pobytem nad 90 dnů v ČR v letech 1993-2007 k 31.12. a v roce 2008 k 31.5

Zdroj: MPSV

2. Cizinci v České republice

Základní normou upravující postavení, povinnosti, práva a omezení cizinců při jejich legálním pobytu na území ČR je zákon č.326/1999 Sb. o pobytu cizinců. Aktuální norma pochází z roku 1999 a byla mnohokrát novelizována. Poslední novelizace z konce roku 2007 reaguje na fakt vstupu ČR do schengenského prostoru, upravuje podmínky pro získání povolení k trvalému pobytu (povinnost složit zkoušku z českého jazyka) a zpřísňuje možnost získání trvalého pobytu pro cizince/manžele/ky českých občanů. Na zákon o pobytu cizinců jsou navázány v některých svých částech věnujících se cizincům i další normy, jako např. zákony o zaměstnanosti, o veřejném zdravotním pojištění, o nemocenském pojištění, o hmotné nouzi atd.

Tabulka 1: Vývoj počtu cizinců s povolením k pobytu v ČR - 2001-2007 (stav k 31.12.)

	2001	2002	2003	2004	2005	2006	2007
Cizinci celkem	212 069	232 932	241 934	255 917	280 111	323 343	394 345
Obyvatelstvo ČR	10 206 436	10 203 269	10 211 455	10 220 577	10 251 079	10 287 189	10 381 130
Cizinci zahrnutí do obyvatelstva ČR celkem*	163 805	179 154	195 394	193 480	258 360	296 236	347 649
Trvalé pobyty	69 816	75 239	80 844	99 467	110 598	139 185	157 512
Azyly	1 275	1 334	1 513	1 623	1 799	1 887	2 030
Přechodný EU/ Dlouhodobý pobyt	-	-	-	92 390	145 963	155 164	188 107
% z obyvatelstva	1,60	1,76	1,91	1,90	2,52	2,88	3,35
Obyvatelstvo ČR a zbývající cizinci	10 254 700	10 257 047	10 257 995	10 283 014	10 272 830	10 314 296	10 427 826
Víza nad 90 dní	140 978	156 359	159 577	62 437	21 751	27 107	46 696
Trvalý pobyt (v %)	32,9	32,3	33,4	38,9	39,5	43,0	39,9
Platný azyl (v %)	0,6	0,6	0,6	0,6	0,6	0,6	0,5
Povolení k pobytu (kromě trvalého; v %)	-	-	-	36,1	52,1	48,0	47,7
Pobyt na víza nad 90 dnů (v %)	66,5	67,1	66,0	24,4	7,8	8,4	11,8

Zdroj: ČSÚ

V České republice v devadesátých letech počet legálně usazených cizinců postupně vzrůstal. Mezi lety 1994 a 1999 se více než zdvojnásobil ze zhruba 100 tisíc na počty kolem 200 tisíc pobývajících cizinců. V roce 2000 počet cizinců v ČR poklesl o 30.000 osob, přičemž tento vývoj se všeobecně přičítá změnám legislativy. 1.1.2000 vstoupil v platnost zákon č. 326/1999 Sb., o pobytu cizinců na území ČR, v původní podobě, který podstatně zpřísnil vstupní a pobytový režim většiny cizinců v ČR. Některá ustanovení tohoto zákona

byla zmírněna až novelou platnou od 1. července 2001, která měla za následek opětovný mírný nárůst počtu usazených cizinců. Ten pokračoval až do roku 2007, kdy v ČR bylo Ředitelstvím služby Cizinecké a pohraniční policie evidováno přibližně 394 345 cizinců, z nichž 40 % představovali cizinci s trvalým pobytem.

K 31. 7. 2009 Ředitelství služby cizinecké policie MV ČR v České republice evidovalo 439 762 cizinců, z toho 176.508 cizinců s trvalým pobytem, 263.254 cizinců s některým z typů dlouhodobých pobytů nad 90 dnů (tj. přechodné pobyty občanů EU a jejich rodinných příslušníků, dále víza nad 90 dnů a povolení k dlouhodobému pobytu občanů zemí mimo-EU).

V souvislosti se vstupem ČR do EU došlo k rozšíření kategorií pobytu, kromě pobytů trvalých a víz nad 90 dnů jsou rozlišovány také pobyty dlouhodobé (pobyty navazující na víza nad 90 dnů) a pobyty přechodné pro občany EU a jejich rodinné příslušníky.

K 31. 7. 2009 byli v ČR nejčastěji zastoupeni občané Ukrajiny (133 773 osob, 30 % z celkového počtu cizinců) a Slovenska (76 956 osob, 17 %). Dále následovala státní občanství Vietnamu (60 998 osob, 14 %), Ruska (29 144 osob, 7 %) a Polska (20 700 osob, 5 %).

3. Ekonomická aktivita cizinců na území České republiky

Účelem pobytu většiny cizinců v České republice je zaměstnání či podnikání na živnostenský list (tab. 2)[4]. Celková zaměstnanost[5] cizinců k 31. 12. 2008 činila 361 709 osob, přičemž 79 % těchto cizinců bylo evidováno úřady práce. Mezi tyto cizince jsou řazeni cizinci v postavení zaměstnanců a nově podle zákona č. 435/2004 Sb., o zaměstnanosti rovněž cizinci, kteří pracovali jako společníci obchodních společností, členové družstev či jako členové statutárních orgánů obchodních společností a družstev, a přitom se kromě účasti na řízení společnosti věnovali plnění tzv. běžných úkolů. Zbylých 77 158 cizinců mělo platné živnostenské oprávnění.

Ze zemí EU je u nás zaměstnáno 156 959 osob, mezi nimi je 30,1 % žen. 15 931 osob podniká na základě živnostenského oprávnění (z toho 19,3 % žen). Úřady práce bylo k 31.12.2008 evidováno 141 028 osob, z toho 31,8 % žen. Téměř 110 tis. jich je ze Slovenska. Slováků je tak pochopitelně nejvíce jak mezi živnostníky, tak i mezi ekonomicky aktivními v pozici zaměstnanců. Mezi ekonomicky aktivními Poláky, kterých zde žije 22 tisíc výrazně převažují ti, kteří mají živnostenský list.

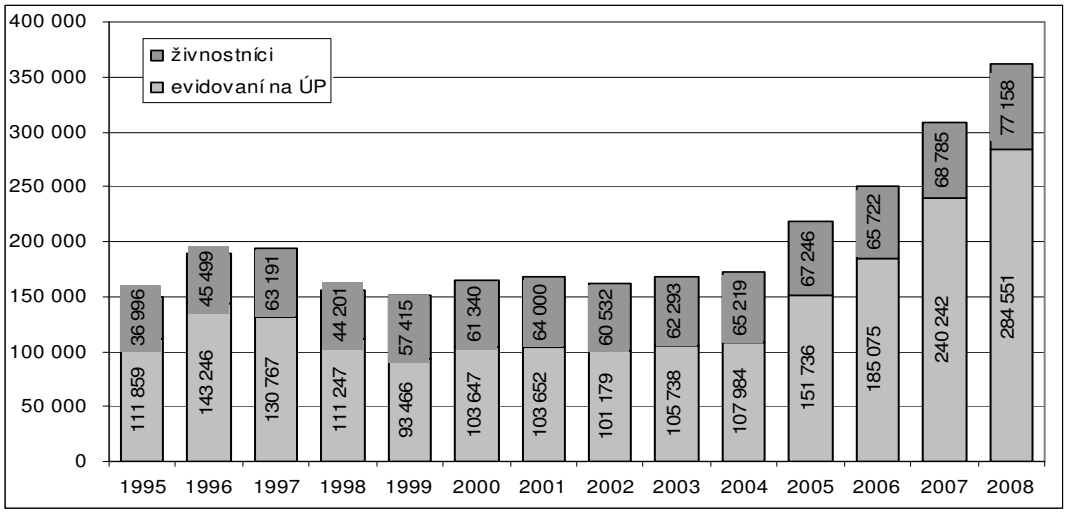
Ze zemí mimo EU bylo v ČR ke konci loňského roku ekonomicky aktivních 204 750 osob, z toho 33,7 % žen. Nejčastěji pocházeli z Ukrajiny, Vietnamu a Mongolska. Téměř 30 % z celkového počtu ekonomicky aktivních ze zemí mimo EU podnikalo na základě živnostenského oprávnění. Nejvíce mezi nimi byli zastoupeni občané Vietnamu, Ukrajiny a s velkým odstupem občané Ruska. Mezi občany Mongolska, kteří představují pátou nejpočetnější skupinu cizinců, bylo pouze 157 osob podnikajících na základě živnostenského

[4] Údaje z hlediska účelu pobytu jsou bohužel zatím k dispozici pouze za rok 2007. Počet cizinců s platným živnostenským oprávněním se nemusí shodovat (a ani neshoduje, neboť je ve všech letech vyšší) s počtem cizinců, u nichž je živnostenské oprávnění účelem pobytu. Někteří cizinci mají jiný účel pobytu a přitom podnikají na základě živnostenského oprávnění.

[5] Údaje za celkovou zaměstnanost cizinců (tj. zaměstnané cizince, celkem) vycházejí ze dvou oddělených evidencí. Z Ministerstva práce a sociálních věcí ČR, resp. Správy služeb zaměstnanosti, která shromažďuje údaje vycházející z evidencí úřadů práce, a to údaje o vydaných platných povoleních k zaměstnání cizinců, údaje o počtu informací o nástupu občanů EU/EHP a Švýcarska k výkonu práce a z informací o nástupu k výkonu práce cizinců s trvalým pobytem z ostatních zemí Druhým zdrojem jsou informace z Ministerstva průmyslu a obchodu ČR, poskytujícího data o cizince s platným živnostenským oprávněním.

oprávnění. Téměř všichni zde žijící Mongolové ve věku ekonomické aktivity jsou v postavení zaměstnanců.

Počet cizinců evidovaných úřady práce (do října 2004 se jednalo výhradně o osoby v postavení zaměstnanců) v druhé polovině devadesátých let prudce klesal až na 93,5 tis. v roce 1999. Po přechodném vzestupu 2000–2001 se jejich počet v roce 2002 opět snížil na 101 tis. osob, přičemž tento pokles byl způsoben úbytkem počtu pracujících občanů Slovenska (obr. 2). Od roku 2003 je však zřejmý mírný nárůst počtu zaměstnanců. V roce 2005 počet cizinců prudce vzrostl a na konci roku 2006 zde již bylo v postavení zaměstnanců-cizinců 185 tis. osob. V roce 2007 přírůstek pracujících cizinců se meziročně zvýšil od 55 tisíc, tj. téměř o 30 %. Na konci roku 2008 zde již bylo 285 tisíc cizinců-zaměstnanců.



Obr. 2 Celková zaměstnanost cizinců podle postavení v zaměstnání

Tabulka 2: Cizinci podle kraje a účelu pobytu (mezinárodní klasifikace) k 31. 12. 2007

Účel pobytu / Kraj	Celkem	Studium a praxe	Podnikání na ŽL	Zaměstnání	Ost. ek. akt.	Volné právo usídlení (krajané)	Povolení k trvalému pobytu)	Rodinní příslušníci a sloučení rodiny	Azyl	Humanitární statut; dočasná ochrana	Ostatní
Pha	129 417	4 281	24 205	47 825	127	498	20 619	29 470	565	81	1 746
StČ	50 379	559	8 089	19 827	17	234	8 253	12 947	137	8	308
JiČ	15 185	679	1 610	4 539	7	48	2 916	4 338	20	3	1 025
Plz	21 048	230	3 917	6 305	2	140	4 067	5 250	72	11	1 054
KaV	19 442	107	3 130	1 926	1	112	4 598	8 271	27	4	1 266
Úst	33 360	128	6 536	6 407	18	188	5 355	10 457	337	17	3 917
Lib	15 442	207	2 129	4 740	5	98	2 775	5 119	200	5	164
KrH	15 594	199	3 001	5 301	2	64	2 162	4 525	125	-	215
Par	10 588	79	1 640	4 818	5	61	1 174	2 704	29	3	75
Vys	8 752	22	1 970	3 402	2	26	1 147	2 107	24	1	51
JiM	32 835	634	5 062	12 264	8	157	4 702	9 240	332	21	415
Olo	10 354	623	1 739	2 480	7	27	1 670	3 676	38	3	91
Zlí	7 652	200	590	2 285	3	35	1 143	3 310	17	2	67
MoS	23 049	550	2 127	7 286	17	56	3 061	9 620	116	45	171
Nezn.	1 248	21	265	329	-	1	36	96	468	14	18
Celk.	394 345	8 519	66 010	129 734	221	1 745	63 678	111 130	2 507	218	10 583

Počet cizinců v postavení zaměstnanců je přímo ovlivňován situací na trhu práce. Oblasti, v nichž je nízké procento nezaměstnanosti, zpravidla vykazují větší počet cizinců, kteří obdrželi povolení k zaměstnání nebo jsou evidováni na úřadech práce (Praha a některé okresy Středočeského kraje).

Tabulka 3: Zaměstnanost cizinců: nejčetnější státní občanství; 2000-2008 (31. 12.)

Státní občanství	2000	2001	2002	2003	2004	2005	2006	2007	2008
Slovensko	70 237	70 606	63 733	66 176	68 575	84 016	99 637	109 915	109 478
Ukrajina	37 155	39 063	39 005	41 241	41 885	61 195	67 480	83 519	102 285
Vietnam	19 382	20 466	20 231	21 201	22 229	22 876	23 602	29 862	48 393
Polsko	8 712	7 712	8 419	8 529	10 133	13 929	18 387	24 931	22 044
Mongolsko	891	1 204	1 375	1 594	1 789	2 013	2 973	7 057	13 157
Moldavsko	1 852	1 793	1 794	1 908	1 933	3 314	4 093	6 433	9 748
Bulharsko	2 697	2 986	2 989	2 884	2 764	2 823	2 859	6 319	6 066
Rusko	2 970	2 795	2 640	2 545	2 691	3 929	3 659	3 716	4 576
Německo	2 289	2 158	2 255	2 417	2 406	2 907	3 583	4 108	4 135
Rumunsko	1 090	942	910	877	798	1 146	1 453	4 538	3 876
Spojené království	1 408	1 343	1 399	1 455	1 263	1 699	2 163	2 446	2 835
Spojené státy	1 911	1 868	2 024	2 029	1 806	1 824	1 698	1 819	2 290
Čína	408	590	440	523	576	1 145	1 157	1 371	1 808
Bělorusko	1 599	1 474	1 574	1 329	1 198	1 362	1 387	1 568	1 771
Francie	731	745	839	870	669	849	1 243	1 446	1 727
Ostatní	9 972	10 075	10 351	10 778	10 923	12 464	14 043	18 528	26 553
Celkem	**164 987**	**167 652**	**161 711**	**168 031**	**173 203**	**218 982**	**250 797**	**309 027**	**361 709**

Počet cizinců s živnostenským oprávněním poprvé kulminoval na konci roku 1997 (necelých 63 tis. osob), když se proti konci roku 1994 zvýšil téměř 3,5krát. V roce 1998 došlo k poklesu, a to až o třetinu proti roku předcházejícímu. Od roku 2000 má na vývoj počtu těchto pracujících cizinců značný vliv novela živnostenského zákona. S tím související zpřísnění podmínek pro získání dlouhodobého víza za účelem podnikání se projevilo v relativně vysokém poklesu počtu těchto osob v roce 2002. Od následujícího roku však počet živnostníků rostl a na konci roku 2005 již dosáhl více než 67 tis. osob. V roce 2006 počet cizinců podnikatelů sice mírně klesl, ale v minulém roce se dostal na historicky nejvyšší úroveň, když dosáhl téměř 69 tis.

4. Závěr

Nejčastějším účelem pobytu cizinců je zaměstnání, patrné je výrazné zastoupení tohoto účelu pobytu zejména u mužů (pobyt za účelem zaměstnání mělo přibližně 40 % z nich), dalším významným účelem pobytu je sloučení s rodinou, které je častější u žen (tento typ pobytu mělo 40 % žen). Dále zde cizinci pobývají za účelem podnikání na živnostenský list či za účelem usídlení a to na základě povolení k trvalému pobytu.

5. Literatura

[1] Cizinci v České republice, ČSÚ 2004–2008
[2] www.czso.cz
[3] www.migraceonline.cz
[4] www.antropologie.zcu.cz
[5] www.domavcr.cz
[6] www.policie.cz

Adresa autora
Eva Kačerová, RNDr.
Nám. W. Churchilla 4
130 67 Praha 3
kacerova@vse.cz

Závisí miera nezamestnanosti od veľkosti okresu na Slovensku? Does unemployment rate depend on district area in Slovakia?

Samuel Korónyi

Abstract: The paper deals with formal relationship between unemployment rates and areas of Slovak districts during 2001 - 2008. It really seems that unemployment rate does depend on district area ($p < 0,001$) at first sight. But if we divide all Slovak districts into two groups of nine "town districts" (Bratislava I to V and Košice I to IV) and of seventy other districts then the degree of dependence is questionable.

Keywords: regional disparities, correlation analysis, regression analysis

Kľúčové slová: regionálne disparity, korelačná analýza, regresná analýza

1. Úvod

V nedávno uverejnenom príspevku (Korónyi 2009) sme štatisticky dokázali, že územie Slovenska bolo politickým rozhodnutím rozdelené do piatich zreteľne rozdielnych zhlukov vzhľadom na hustotu obyvateľstva. Jedným zhlukom je skupina deviatich „mestských" okresov Bratislava I až V a Košice I až IV. Ostatné štyri zhluky obsahujú rôzne okresy, ktoré nemajú podľa dostupných informácií rozumnú interpretáciu. Toto rozdelenie má za následok aj multimodalitu rozdelení plochy okresov a počtu ich obyvateľov.

Cieľom príspevku je zistiť, či veľkosť súčasných okresov nesúvisí s aktuálnou mierou evidovanej nezamestnanosti za posledných osem rokov.

2. Dáta

Pre analýzu sme použili voľne dostupné dáta z databázy Regdat na webovej stránke Štatistického úradu SR. Išlo o vybrané údaje na lokálnej úrovni LAU 1 (okresy) o plošnej veľkosti okresov a ich miere evidovanej nezamestnanosti (obidva ukazovatele sú za roky 2001 až 2008). Plošná veľkosť okresov sa v sledovanom období rokov 2001 - 2008 reálne nemení (v r. 2001 - 2005 sa udávala na celé km^2), len sa spresňuje až na úroveň stotiny km^2 (2006 - 2008). Pri výpočtoch je rozdiel medzi plochami okresov po rokoch zanedbateľný.

3. Použité metódy

Pre splnenie cieľa príspevku sme použili korelačnú a regresnú analýzu implementovanú v systémoch NCSS verzia 2001 a SPSS verzia 13. Pre grafické zobrazenie hodnôt sme urobili štandardné rozptylové grafy s regresnými priamkami v štatistickom systéme NCSS.

4. Jednorozmerná deskriptívna analýza ukazovateľov

V tabuľke 1 sú základné štatistické parametre analyzovaných ukazovateľov. Z nej je zrejmé, že rozsah plochy okresov je veľký a siaha od 9,59 km^2 (okres Bratislava I) až po 1551,14 km^2 (okres Levice). Deväť najmenších okresov patrí medzi „mestské" okresy - Bratislava I až V, Košice I až IV (Korec 2005). Ďalší v poradí desiaty je okres Kysucké Nové Mesto (173,68 km^2). Smerodajná odchýlka (367) je relatívne veľká v porovnaní s aritmetickým priemerom (621), tomu odpovedá relatívne veľká hodnota variačného koeficientu 0,591. Všetko toto poukazuje na príliš roztiahnutú distribúciu plochy okresov.

Tabuľka 1: Štatistické parametre okresov - plocha (v km^2) v roku 2008 (A2008 a miera evidovanej nezamestnanosti v r. 2001 - 2008 (N2001 - N2008)

Ukazovateľ	Minimum	Maximum	Priemer	Smer. odch.
A2008	9.59	1551.14	620.72	366.60
N2001	3.79	35.45	19.55	7.68
N2002	3.16	37.22	18.40	8.24
N2003	2.85	30.64	16.29	7.32
N2004	2.29	28.66	13.99	6.64
N2005	1.77	29.24	12.35	6.58
N2006	1.69	28.34	10.35	6.21
N2007	1.56	27.05	8.95	5.82
N2008	1.46	26.83	9.45	5.97

Priemerná nezamestnanosť monotónne klesá o 1 až 3 percentá v danom období s výnimkou posledného roka 2008, kedy naopak stúpla v porovnaní s rokom 2007 následkom krízy o 0,5 percenta.

5. Dvojrozmerná exploračná analýza ukazovateľov

Z ekonomického hľadiska je podstatné, či samotná veľkosť plochy okresov neovplyvňuje niektoré ekonomické ukazovatele. Nezamestnanosť je najčastejšie používaný ukazovateľ pri zisťovaní stavu rozvinutosti na úrovni okresov (Korec 2005). Samotná miera závisí od množstva faktorov a možno ju považovať za syntetický výstupný ukazovateľ. Preto sa natíska otázka, či miera evidovanej nezamestnanosti nesúvisí aj s plochou okresov.

Pre zistenie závislosti sme použili Pearsonove a Spearmanove korelačné koeficienty medzi plochou okresu a mierou evidovanej nezamestnanosti. V tabuľke 2 sú hodnoty Pearsonových koeficientov (Spearmanove sú podobné) po rokoch. Všetky sú významné na hladine menšej ako 0,001. Sú kladné, to znamená, že väčšie okresy majú významne väčšiu nezamestnanosť.

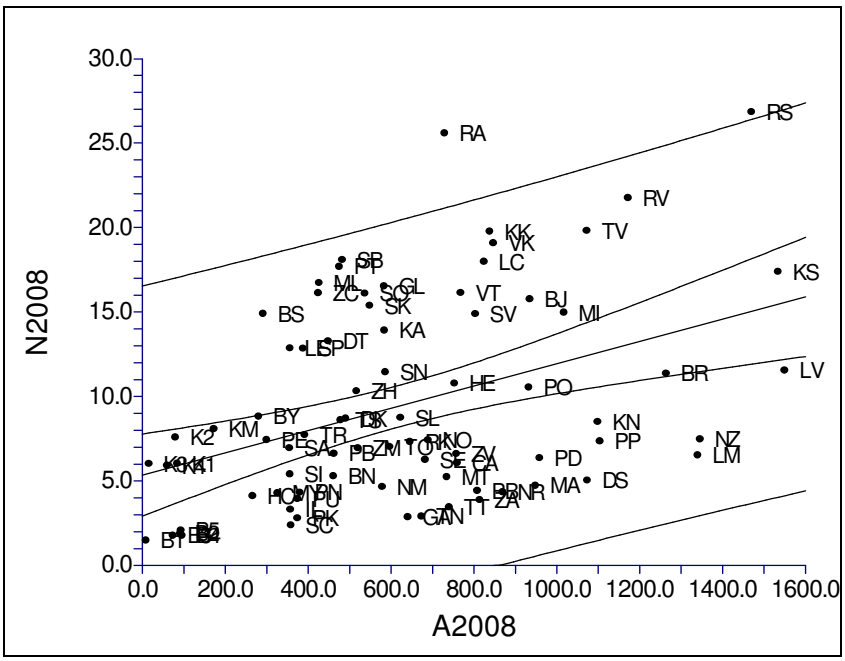
Tabuľka 2: Pearsonove korelačné koeficienty medzi plochou okresu a nezamestnanosťou v rokoch 2001 až 2008 (n = 79)

Rok	2001	2002	2003	2004	2005	2006	2007	2008
R	0,506	0,488	0,472	0,439	0,412	0,420	0,399	0,405

Pre grafické zobrazenie empirického vzťahu medzi nezamestnanosťou a veľkosťou okresu je na grafe 1 stav v poslednom roku 2008. Na grafe sú okrem regresnej priamky aj 95 percentné predikčné intervaly. Z celkového počtu 79 okresov sú dva okresy mimo predikčných intervalov (Rimavská Sobota a Revúca). Pre ostatné roky v období 2001 - 2007 je vzhľad závislosti veľmi podobný s miernymi rozdielmi vo vzájomnej polohe okresov.

Ukazuje sa teda, že za obdobie ôsmich po sebe idúcich rokov 2001 až 2008 závisí nezamestnanosť od veľkosti plochy okresu. To je však len formálny záver.

Na grafoch za všetky sledované roky je nápadná poloha piatich navzájom blízkych bratislavských a štyroch košických okresov, ktoré sú v ľavom dolnom rohu a vzniká možné ovplyvnenie regresnej priamky leverage efektom. V tomto prípade by išlo o zvýšenie sklonu regresnej priamky.



Graf 1: Závislosť nezamestnanosti (N2008) v roku 2008 od plochy okresu(A2008)

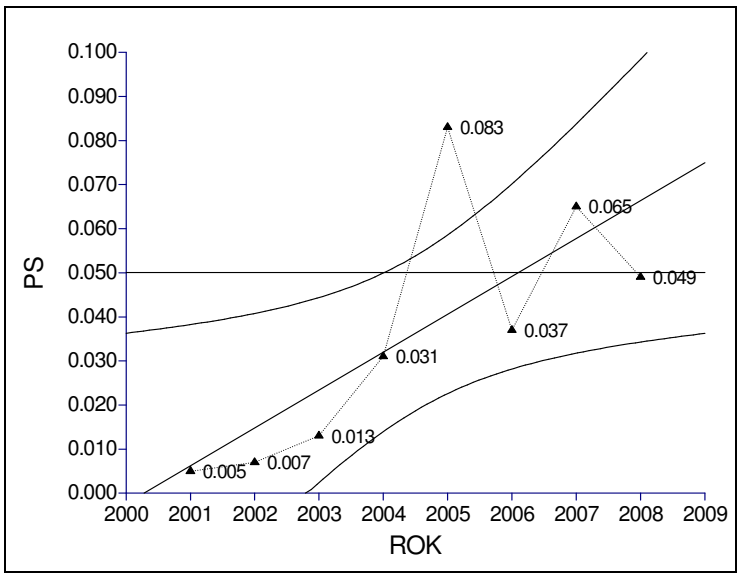
Pre overenie vplyvu bratislavských a košických okresov na sklon regresnej priamky sme uvažovali všetky ostatné okresy bez deviatich mestských okresov. V tabuľke 3 sú Pearsonove a Spearmanove korelačné koeficienty pre ostatné okresy (n = 70), pričom v poslednom riadku je presná *p* hodnota Spearmanovho koeficienta pre všeobecnejšie testovanie monotónnosti vzťahu. V porovnaní s tabuľkou 2 sa znížila hodnota koeficientov minimálne o 0,1.

Tabuľka 3: Pearsonove a Spearmanove korelačné koeficienty medzi plochou okresu a nezamestnanosťou v rokoch 2001 až 2008 bez mestských okresov(n = 70)

Koeficient	N2001	N2002	N2003	N2004	N2005	N2006	N2007	N2008
Pearson	0,371	0,362	0,336	0,292	0,273	0,308	0,285	0,282
Spearman	0,329	0,317	0,297	0,257	0,209	0,250	0,222	0,236
p_S	0,005	0,007	0,013	0,031	0,083	0,037	0,065	0,049

Z *p* hodnôt Spearmanovho korelačného koeficientu je vidieť, že významnosť vzájomného monotónneho vzťahu už nie je tak veľká ako predtým a skôr ukazuje len na marginálnu závislosť medzi mierou evidovanej nezamestnanosti a plošnou veľkosťou okresu.

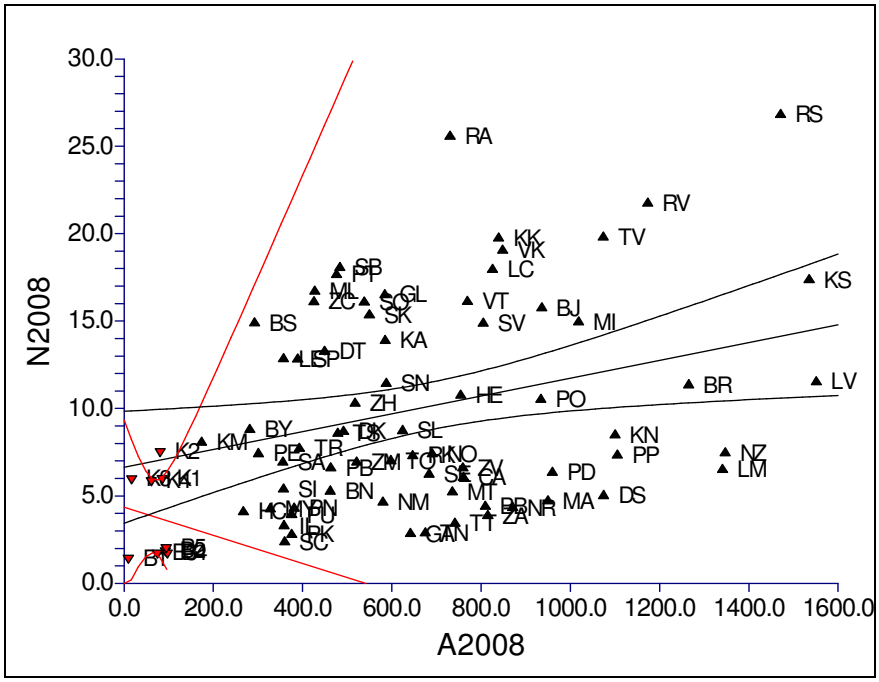
Je otázka, či samotné *p* hodnoty Spearmanovho korelačného koeficientu závislosti medzi veľkosťou okresov a ich nezamestnanosťou majú významný trend. Na jej zodpovedanie sme opäť použili korelačné koeficienty. Obidva sú kladné a významné - Pearsonov korelačný koeficient má $p = 0{,}034$, Spearmanov koeficient má $p = 0{,}010$. Preto môžeme povedať, že *p* hodnoty závislosti medzi nezamestnanosťou a plochou okresu významne rastú. Rast *p* hodnôt je na grafe 2, kde okrem regresnej priamky a jej konfidenčného intervalu je aj „kritická“ hranica 0,05.



Graf 2: Vývoj p hodnôt Spearmanovho koeficientu závislosti nezamestnanosti od plochy okresu pre okresy mimo Bratislavy a Košíc

Vyzerá to, akoby sa závislosť medzi veľkosťou nie mestských okresov a ich nezamestnanosťou vstupom do EÚ stala nevýznamná, čo by tiež svedčilo o určitej konsolidácii ekonomiky.

Na grafe 3 je zobrazená skúmaná empirická závislosť s konfidenčnými intervalmi regresnej priamky pre rok 2008 po rozdelení okresov na mestské a ostatné okresy. Je na nej zreteľné, že pre deväť mestských okresov je sklon priamky nevýznamný a pre ostatných 70 okresov sa jej sklon zmenšil.



Graf 3: Závislosť nezamestnanosti (N2008) v roku 2008 od plochy okresu(A2008) po rozdelení na mestské a ostatné okresy

6. Záver

Na základe údajov databázy RegDat Štatistického úradu SR sme urobili štatistickú analýzu závislosti plošnej veľkosti súčasných okresov a miery evidovanej nezamestnanosti za roky 2001 až 2008. Z jej výsledkov vyplýva, že pri analýze slovenských okresov je potrebné dávať pozor na rozdiely medzi mestskými okresmi (Bratislava I až V, Košice I až IV) a ostatnými okresmi. Ukazuje sa, že niekedy je vhodné posudzovať ich oddelene pre ich principiálne rozdielnu podstatu.

7. Literatúra

[1]KOREC, P. et al. 2005. Regionálny rozvoj Slovenska v rokoch 1989-2004 : Identifikácia menej rozvinutých regiónov Slovenska. 1. vyd. Bratislava : Geo-grafika, 2005. ISBN 80-969338-0-9

[2]KORÓNY, S.: Exploračná analýza počtu obyvateľov a plochy okresov na Slovensku. In: Forum Statisticum Slovacum. Roč.5, č. 6, 2009, s. . s. 69-76. ISSN 1336-7420

Adresa autora:

RNDr. Samuel Koróny, PhD.
Ústav vedy a výskumu UMB
Cesta na amfiteáter 1
974 01 Banská Bystrica
Email: samuel.korony@umb.sk

Kriminalita mladistvých na Slovensku
Juvenile delinquency in Slovakia

Viera Labudová

Abstract: This contribution deals with the specifics of juvenile delinquency in Slovakia. In the paper we analyze the spatial dimensions of juvenile delinquency. We use the principal components method and cluster analysis.

Key words: criminality, juvenile delinquency, principal components method, cluster analysis

Kľúčové slová: kriminalita, trestný čin, hlavné komponenty, zhluková analýza

1. Úvod

„Každý štát, ktorý chce byť zodpovedný za budúci vývoj kriminality v krajine, musí venovať trvalú a sústredenú pozornosť formám a prejavom sociálnopatologických javov u detí a mladistvých a súčasne musí hájiť ich práva a oprávnené záujmy v najširšom slova zmysle. Táto pozornosť nesmie mať len represívny alebo reštriktívny charakter, ale musí mať aj podobu chápajúcej pedagogicko-psychologickej analýzy, ktorá vyúsťuje do vhodných preventívnych či resocializačných zásahov.“[6]

2. Kriminalita mladistvých

Trestná činnosť mladej generácie predstavuje dlhodobo celospoločenský problém nielen u nás, ale aj v krajinách Európskej únie. Ako vyplýva z oficiálnych štatistík, ľudia sa dopúšťajú trestnej činnosti predovšetkým v mladosti a rannej dospelosti.

Trestnoprávna ochrana mládeže je integrálnou súčasťou úloh a funkcií trestného práva v našej spoločnosti. Otázky základov trestnej zodpovednosti mladistvých[7], druhy trestov a ochranných opatrení, skutkové podstaty týkajúce sa mladistvých, sú na Slovensku upravené priamo v platnom Trestnom zákone č. 300/2005 Z. z. a následne v Trestnom priadku č. 301/2005 Z. z. Súčasná právna úprava umožňuje trestné stíhanie mladistvých, ktorí v čase spáchania trestného činu dovŕšili hranicu štrnásť rokov.

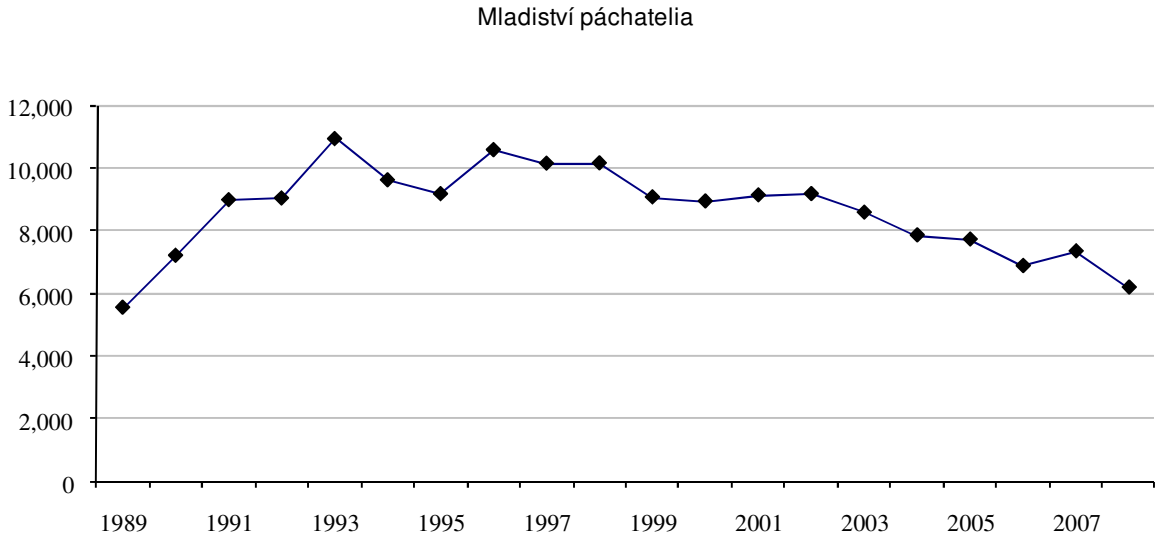
Trend vývoja počtu mladistvých, ktorí sa dopustili v období rokov 1989 až 2008 trestného činu znázorňuje obrázok 1.

Kriminalita je výrazne spätá so zmenami v spoločensko-hospodárskom vývoji. V dôsledku prudkých sociálnych zmien dochádza k oslabeniu sociálnej sebakontroly jedincov, k zmenám hodnotovej orientácie, k poruchám v konaní a prekročení jestvujúcich noriem. Obdobie po roku 1989 sa na Slovensku vyznačuje výrazným nárastom celkovej zločinnosti a samozrejme aj v náraste kriminality mladistvých, ktorá kulminovala v roku 1993 (10 944 mladistvých páchateľov). Od roku 1996 počet mladistvých páchateľov klesá a tento trend pokračuje doteraz.

---

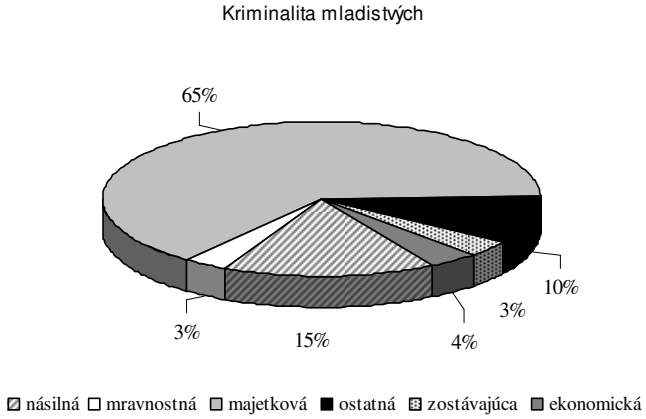
[6] Novotný, O.-Zapletal, J. a kol.: Kriminologie. Praha: ASPI Publishing, 2004, str. 377. ISBN 80-7357-02b-2.

[7] Osoba, ktorá v čase spáchania trestného činu dovŕšila štrnásty rok a neprekročila osemnásty rok svojho veku, sa považuje za mladistvú. Zdroj: ZÁKON z 20. mája 2005 TRESTNÝ ZÁKON (v znení zákona č. 650/2005 Z.z.).

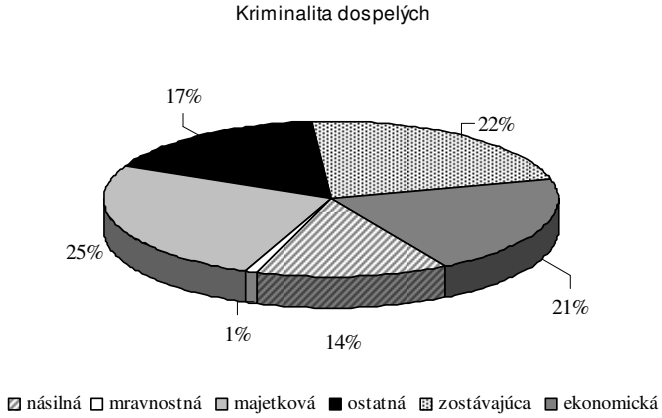


Obrázok 2: Vývoj počtu mladistvých páchateľov v rokoch 1989 až 2008

Zaujímavé je sledovať štruktúru trestných činov mladistvých, ktorej špecifiká sa najlepšie prejavia pri porovnaní so štruktúrou kriminality dospelých páchateľov. Najvýraznejšie sa tento rozdiel prejavuje v podiele majetkovej kriminality, kam patria rôzne druhy krádeží. Tento druh kriminality predstavoval v roku 2008 u mladistvých až 65 % celkovej kriminality, pričom u dospelých to bolo len 25%. V poradí druhou najpočetnejšou skupinou bola násilná kriminalita, kam patria vraždy, lúpeže, úmyselné ublíženie na zdraví (15% kriminality mladistvých). Najmenej trestných činov u mladistvých bolo z oblasti zostávajúcej kriminality (3%) a mravnostnej kriminality (3%). U dospelých zostávajúca kriminalita (dopravné cestné nehody, ohrozenie pod vplyvom návykových látok, vojenské trestné činy a trestné činy proti republike) tvorila až 22 % celkovej kriminality.



Obrázok 2: Štruktúra kriminality mladistvých v roku 2008



Obrázok 3: Štruktúra kriminality dospelých v roku 2008

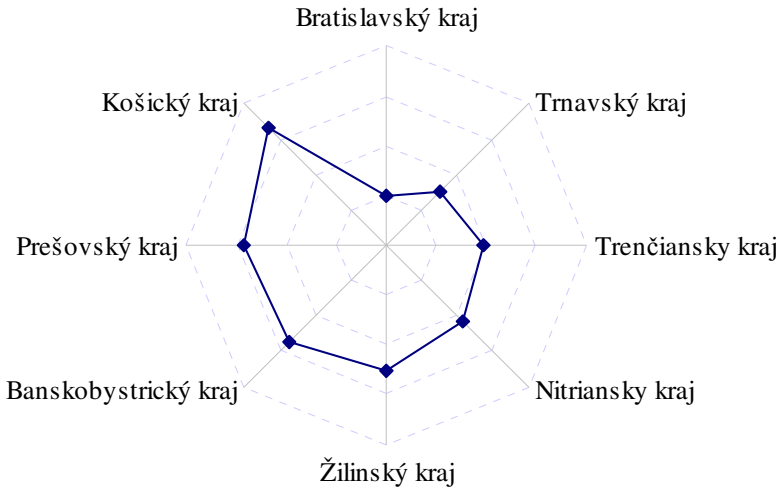
3. Priestorová dimenzia kriminality

Cieľom tohto príspevku nebolo skúmanie determinantov vzniku a vývoja trestnej činnosti detí a mládeže na Slovensku. Pokúsili sme sa o jej priestorovú diferenciáciu.

Analýzu sme založili na regionálnych údajoch o kriminalite detí a mládeže, ktoré zverejnil Ústav informácií a prognóz školstva SR. Údaje boli zbierané na úrovni okresov, resp. krajov. Zoznam premenných je v tabuľke č.1.

Niektoré ukazovatele boli vyčíslené nielen pre skupinu mladistvých, ale aj pre deti, ktoré nepresiahli vekovú hranicu štrnásť rokov a pre mladých vo veku od 19 do 21 rokov.

Jednoduchou metódou poradí, aplikovanou na všetkých 47 premenných, sme určili poradie krajov na základe kriminality mladistvých a mladých páchateľov (obr.4).



Obrázok 4: Úroveň kriminality mladistvých a mladých páchateľov v krajoch Slovenska v roku 2008

Najnepriaznivejší stav kriminality mladistvých a mladých ľudí je v Košickom kraji, ďalej nasleduje Prešovský a Banskobystrický kraj. Najlepšia situácia je v Bratislavskom a Trnavskom kraji.

Pre každý okres boli hodnoty všetkých premenných prepočítané na celkový počet páchateľov daného okresu. Na zredukovaniu počtu 45 vstupných premenných sme použili metódu hlavných komponentov. Množinu vstupných premenných sme nahradili siedmimi hlavnými komponentami. Každý z nich vyhovuje Kaiserovmu pravidlu - jeho vlastné číslo je väčšie ako 1. Komponenty, ktorých vlastné čísla a podiel vysvetlenej variability uvádza tabuľka 2 vysvetľujú spolu 72,81 % variability.

Na základe vytvorených hlavných komponentov bola uskutočnená zhluková analýza okresov Slovenska. Vytvorených osem zhlukov s príslušným okresmi je v tabuľke 3.

Na základe relatívnych početností páchateľov a stíhaných osôb pre jednotlivé druhy kriminality je najvážnejšia situácia v okrese Stará Ľubovňa - uvedený okres predstavuje samostatný zhluk. Zlá situácia je aj v okresoch šiesteho zhluku. Naopak, okresy patriace do druhého zhluku (okresy Bratislavského kraja) majú najnižší podiel páchateľov a stíhaných osôb.

Tabuľka 4: Vstupné premenné použité pri analýze

Označenie premennej	**Názov premennej**
X_1, X_2, X_3	Celková kriminalita_počet páchateľov (do 14 rokov, 14-18 rokov, 19-21 rokov)
X_4, X_5, X_6	Celková kriminalita_počet stíhaných (do 14 rokov, 14-18 rokov, 19-21 rokov)
X_7, X_8, X_9	Násilná kriminalita_počet stíhaných (do 14 rokov, 14-18 rokov, 19-21 rokov)
X_{10}, X_{11}, X_{12}	Násilná kriminalita_počet páchateľov (do 14 rokov, 14-18 rokov, 19-21 rokov)
X_{13}, X_{14}, X_{15}	Mravnostná kriminalita_počet stíhaných (do 14 rokov, 14-18 rokov, 19-21 rokov)
X_{16}, X_{17}, X_{18}	Mravnostná kriminalita_počet páchateľov (do 14 rokov, 14-18 rokov, 19-21 rokov)
X_{19}, X_{20}, X_{21}	Majetková kriminalita_počet stíhaných (do 14 rokov, 14-18 rokov, 19-21 rokov)
X_{22}, X_{23}, X_{24}	Majetková kriminalita_počet páchateľov (do 14 rokov, 14-18 rokov, 19-21 rokov)
X_{25}, X_{26}, X_{27}	Ostatná kriminalita_počet stíhaných (do 14 rokov, 14-18 rokov, 19-21 rokov)
X_{28}, X_{29}, X_{30}	Ostatná kriminalita_počet páchateľov (do 14 rokov, 14-18 rokov, 19-21 rokov)
X_{31}, X_{32}, X_{33}	Zostávajúca kriminalita_počet stíhaných (do 14 rokov, 14-18 rokov, 19-21 rokov)
X_{34}, X_{35}, X_{36}	Zostávajúca kriminalita_počet páchateľov (do 14 rokov, 14-18 rokov, 19-21 rokov)
X_{37}	Počet vrážd
X_{38}	Počet lúpeží
X_{39}	Počet úmyselných ublížení na zdraví
X_{40}	Počet týraní zverenej osoby
X_{41}	Počet sexuálnych násilí
X_{42}	Počet krádeží
X_{43}	Počet prípadov užívania omamných látok
X_{44}	Počet prípadov užitia drog pre vlastnú potrebu
X_{45}	Počet ekonomických podvodov

4. Záver

Po roku 1989 došlo na Slovensku k nárastu sociálnopatologických javov. Zvýšila sa kriminalita, nezamestnanosť, toxikománia, bezdomovectvo, rastie organizovaný zločin. Zvýšila sa delikvencia mladistvých a maloletých detí.

V príspevku sme naznačili priestorovú dimenziu kriminality mladistvých, pričom sme použili jednoduchú metódu poradí a zhlukovú analýzu založenú na vytvorených hlavných komponentoch.

Tabuľka 2: Tabuľka vlastných čísel komponentnej analýzy okresov SR

Eigenvalues of the Correlation Matrix				
	Eigenvalue	**Difference**	**Proportion**	**Cumulative**
1	10.6760948	5.4960867	0.2669	0.2669
2	5.1800081	0.8401919	0.1295	0.3964
3	4.3398161	1.2441056	0.1085	0.5049
4	3.0957106	0.6471935	0.0774	0.5823
5	2.4485171	0.5051742	0.0612	0.6435
6	1.9433429	0.5022679	0.0486	0.6921
7	1.4410751		0.0360	0.7281

Tabuľka 3: Výsledky zhlukovej analýzy okresov SR

Zhluk 1
Bánovce nad Bebravou, Trenčín, Ilava, Myjava, Nové Mesto nad Váhom, Nové Zámky, Šaľa, Dunajská Streda, Levice, Prievidza, Partizánske, Komárno
Zhluk 2
Bratislava I, Bratislava II, Bratislava III, Bratislava IV, Bratislava V, BA okolie, Galanta, Trnava, Hlohovec, Piešťany
Zhluk 3
Čadca, Kysucké Nové Mesto, Michalovce, Sobrance, Žilina, Bytča, Košice, Považská Bystrica, Púchov, Martin, Turčianske Teplice, Brezno, Zvolen, Detva, Krupina, Nitra, Zlaté Moravce, Senica, Rožňava, Skalica, Topoľčany, Liptovský Mikuláš
Zhluk 4
Košice - okolie, Trebišov, Bardejov, Vranov nad Topľou, Žiar nad Hronom, Banská Štiavnica, Žarnovica
Zhluk 5
Prešov, Sabinov, Spišská Nová Ves, Gelnica, Lučenec, Poltár, V. Krtíš, Poprad, Kežmarok, Levoča, R. Sobota, Revúca
Zhluk 6
Dolný Kubín, Námestovo, Tvrdošín, Ružomberok, Humenné, Snina, Medzilaborce, Svidník, Stropkov
Zhluk 7
Stará Ľubovňa
Zhluk 8
Banská Bystrica

5. Literatúra

[1]NOVOTNÝ, O– ZAPLETAL, J. a kol. 2004. Kriminologie. Praha: ASPI Publishing. 2004. 452 s. ISBN 80-7357-02b-2.

[2]MATOUŠEK, O. – KROFTOVÁ, A. 2003. Mládež a delikvence. Praha: Portál, 2003. 344 s. ISBN 80-7178-771-X.

[3] Trestný zákon č. 300/2005 Z. z. . [online]. [cit. 2009-28-10]. Dostupné z: http://www.zbierka.sk/zz/predpisy

[4] Ročenka o deťoch a mládeži. [online]. [cit. 2009-28-10]. Dostupné z: http://www.uips.sk

[5] Štatistika kriminality v Slovenskej republike. [online]. [cit. 2009-28-10]. Dostupné z: http://www.minv.sk/?statistika-kriminality-kopia

Adresa autora:

Viera Labudová, RNDr., PhD.
Dolnozemská cesta 1/b,
853 25 Bratislava
viera.labudova@euba.sk

*Článok vznikol v rámci riešenia interného projektu **IGP č. 23/2009 „**Sociálnopatologické javy v živote súčasnej rodiny“ a grantovej úlohy **VEGA 1/4586/07** „Modelovanie sociálnej situácie obyvateľstva a domácností v Slovenskej republike a jej regionálne a medzinárodné porovnania“.*

Testovanie linearity pre Markov-switching modely
Testing linearity for Markov-switching models

Lenčuchová Jana

Abstract: Many econometricians have got interested in Markov-switching models because of their great modeling abilities and become very popular among them in last two decades, but there are still some open problems. This article discusses another, less time consuming approach to evaluate test of linearity against Markov-switching type of non-linearity by using Hamilton's dynamic specification test for validity of Markov assumptions [3]. We apply the same idea in testing remaining non-linearity to compare 2-regime with 3-regime models. We use selected Slovak macro-economic indicators time series to estimate parameters of Markov-switching models and to test linearity.

Key words: Markov-switching model, Markov assumptions, score function, dynamic specification test

Abstrakt: Markov-switching modely zaujali mnoho ekonómov a stali sa veľmi populárne v posledných dvoch desaťročiach najmä kvôli ich výborným schopnostiam modelovať časový rad. V tejto oblasti sú však ešte stále nevyriešené otázky. Tento článok pojednáva o inom, časovo menej náročnom, spôsobe testovania linearity oproti nelinearite typu Markov-switching, pričom využíva Hamiltonov špecifický test na platnosť Markových predpokladov [3]. Rovnakú myšlienku aplikujeme aj v testovaní zostávajúcej nelinearity na porovnanie 2-režimového modelu s 3-režimovým modelom. Na odhadovanie parametrov Markov-switching modelov a testovanie linearity používame vybrané makroekonomické ukazovatele Slovenska.

Kľúčové slová: Markov-switching model, Markovove predpoklady, skórova funkcia, špecifický test

1. Introduction

We notice some dramatic changes in our lives, which modify existing character suddenly and radically. The economists have remarked such changes in many economic and financial time series with sufficient time length. Typical for such changes is that they happen occasionally and suddenly. Such evident changes can result from events like financial crisis, wars, political development, natural disasters and can cause changes in parameters of observable time series for example in expected value, in variance or in coefficients. Markov-switching models are excellent tool for modeling such time series that we describe above.

In case of Markov-switching models we suppose that the regime or state, which occurs at time t is unobserved and it is determined by random variable s_t. If there are N possible regimes, then random variable s_t can acquire only an integer value from $\{1,2,...,N\}$. We assume, stochastic process $\{s_t\}$ is specified as first-order Markov process, it means, that regime at time t depends only on the regime at time t-1:

$$P(s_t = j \mid s_{t-1} = i, s_{t-2} = k,...) = P(s_t = j \mid s_{t-1} = i) = p_{ij}. \quad (1)$$

We call such process as *N-state Markov chain* with transition probabilities $\{p_{ij}\}_{i,j=1,2,...,N}$. The transition probability p_{ij} represents the probability that regime i will be followed by regime j. Note that

$$p_{i1} + p_{i2} + ... + p_{iN} = 1, \qquad i = 1,2,...,N. \quad (2)$$

Suppose that conditional observable variable $y_t|s_t$ is drawn from normal distribution $N(\mu_{s_t}, \sigma_{s_t}^2)$ and model has form

$$y_t = \phi_{0,s_t} + \phi_{1,s_t} y_{t-1} + ... + \phi_{q,s_t} y_{t-q} + \varepsilon_t, \tag{3}$$

where $s_t = 1,2,..., N$ and $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$. Details about the derivation and the estimation of the Markov-switching model are in ([1],[2]).

2. Classical approach to testing linearity

When we describe time series by Markov-switching model we should check if considered time series has non-linear character at all. So we will test linearity against the alternative of Markov-switching type of non-linearity. A natural approach to assessing the suitability of modeling by Markov-switching model is the likelihood ratio test, where the null hypothesis is linear model and the alternative is 2-regime MSW model - that is $H_0 : \boldsymbol{\varphi_1} = \boldsymbol{\varphi_2}$ - which is tested against the alternative hypothesis $H_1 : \phi_{i,1} \neq \phi_{i,2}$ for at least one $i \in \{1,2,...,q\}$ (where $\boldsymbol{\varphi_i}$ represents AR coefficients for i=1,2). The test statistic has the form:

$$LR = L_{MSW} - L_{AR}, \tag{4}$$

where L_{MSW} and L_{AR} are logarithms of likelihood functions for the appropriate Markov-switching model and the AR model. Hansen proved in [5], that the test statistic (4) has non-standard probabilistic distribution, which cannot be characterized analytically. Critical values to determine the significance of the test-statistic therefore have to be determined by means of simulation. The basic structure of such a simulation experiment is that one generates a large number (minimum is 5000) of artificial time series y_t^* according to the model that holds under the null hypothesis. Next, one estimates parameters for AR and MSW models for each artificial time series and then calculates the corresponding LR-statistic according to (4). From that estimated distribution, we calculate critical values. It is necessary to simulate these values for each time series distinctly (also for each order *p* distinctly).

It is apparent that this classical approach is time consuming. Our suggestion is to use Newey-Tauchen-White test ([6],[7],[8]), score function and specification test for validity of Markov assumptions as in Hamilton [3] to substitute this time consuming classical test.

3. Score function and Newey-Tauchen-White test

Score function $\mathbf{h_t}(\boldsymbol{\theta})$ is defined as derivation of logarithm of conditional probability by parameter vector $\boldsymbol{\theta}$:

$$\mathbf{h_t}(\boldsymbol{\theta}) \equiv \frac{\partial \log f(y_t \mid \Omega_{t-1}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}, \tag{5}$$

where $\boldsymbol{\theta}$ represents $(k \times 1)$ parameter vector, Ω_{t-1} represents observation history and $f(y_t \mid \Omega_{t-1}, \theta)$ is the density of y_t conditional on the Ω_{t-1}. So we have *T* (time series length) score functions $\mathbf{h_t}(\boldsymbol{\theta})$ corresponding to sample $(y_T, y_{T-1},..., y_1)$.

The basic Markov-switching model is described by density of y_t conditional on the random variable s_t taking on the values *j* (regime *j*) and on the history of observations:

$$f(y_t \mid s_t = j, \Omega_{t-1}; \boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi}\sigma_\varepsilon} \exp\left\{ \frac{-(y_t - \boldsymbol{\varphi}_\mathbf{j}' \mathbf{x_t})^2}{2\sigma_\varepsilon^2} \right\}, \tag{6}$$

where $\boldsymbol{\varphi}_\mathbf{j} = (\phi_{0,j}, \phi_{1,j}, ..., \phi_{q,j})'$ is vector of AR coefficients for regime *j*, $\mathbf{x_t} = (1, y_{t-1}, ..., y_{t-q})'$, $\Omega_{t-1} = (y_{t-1}, y_{t-2}, ...)$ is history of observations and σ_ε^2 is residual dispersion for which stands $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$.

Let us denote by $\alpha' = (\boldsymbol{\varphi}_1', \boldsymbol{\varphi}_2', ..., \boldsymbol{\varphi}_N', \sigma_\varepsilon^2)$ the parameter vector which describing conditional density for all regimes and the vector of transition probabilities **p** with omitting redundant parameters, because we know express them by the others:

$$p_{iN} = 1 - p_{i1} - p_{i2} - .... - p_{i,N-1}, \qquad i = 1, ..., N. \tag{7}$$

The parameter vector includes α' and **p**:

$\boldsymbol{\theta}' = (\boldsymbol{\alpha}', \mathbf{p}')$,

where $\mathbf{p}' = (p_{11}, p_{12}, ..., p_{1,N-1}, p_{21}, p_{22}, ..., p_{N,N-1})$.

Hamilton [3] derived score function and it equals:

$$\frac{\partial \ln f(\mathbf{Y}_1 \mid \Omega_0; \boldsymbol{\theta})}{\partial \boldsymbol{\alpha}} = \sum_{s_1=1}^{N} \frac{\partial \ln f(y_1 \mid \mathbf{x_1}, s_1; \boldsymbol{\theta})}{\partial \boldsymbol{\alpha}} P(s_1 \mid \Omega_1) \tag{8}$$

for $t = 1$ and

$$\begin{aligned} \frac{\partial \ln f(y_t \mid \Omega_{t-1}; \boldsymbol{\theta})}{\partial \boldsymbol{\alpha}} &= \sum_{s_t=1}^{N} \frac{\partial \ln f(y_t \mid \mathbf{x_t}, s_t; \boldsymbol{\theta})}{\partial \boldsymbol{\alpha}} P(s_t \mid \Omega_t) + \\ &+ \sum_{\tau=1}^{t-1} \sum_{s_\tau=1}^{N} \frac{\partial \ln f(y_\tau \mid \mathbf{x_\tau}, s_\tau; \boldsymbol{\theta})}{\partial \boldsymbol{\alpha}} \{P(s_\tau \mid \Omega_t) - P(s_\tau \mid \Omega_{t-1})\} \end{aligned} \tag{9}$$

for $t = 2,3,...,T$.

$$\begin{aligned} \frac{\partial \ln f(y_t \mid \Omega_{t-1}; \boldsymbol{\theta})}{\partial p_{ij}} &= p_{ij}^{-1}.P(s_t = j, s_{t-1} = i \mid \Omega_t) - p_{iN}^{-1}.P(s_t = N, s_{t-1} = i \mid \Omega_t) \\ &+ p_{ij}^{-1} \left\{ \sum_{\tau=2}^{t-1} [P(s_\tau = j, s_{\tau-1} = i \mid \Omega_t) - P(s_\tau = j, s_{\tau-1} = i \mid \Omega_{t-1})] \right\} \\ &- p_{iN}^{-1} \left\{ \sum_{\tau=2}^{t-1} [P(s_\tau = N, s_{\tau-1} = i \mid \Omega_t) - P(s_\tau = N, s_{\tau-1} = i \mid \Omega_{t-1})] \right\} \\ &+ \sum_{s_1=1}^{N} \frac{\partial \ln P(s_1; \mathbf{p})}{\partial p_{ij}} [P(s_1 \mid \Omega_t) - P(s_1 \mid \Omega_{t-1})], \end{aligned} \tag{10}$$

for $i = 1,2,..., N$, $j = 1,2,..., N$, $t = 2,..., T$ and

$$\frac{\partial \ln f(y_1 \mid \Omega_0; \boldsymbol{\theta})}{\partial p_{ij}} = \sum_{s_1=1}^{N} \frac{\partial \ln P(s_1; \mathbf{p})}{\partial p_{ij}} P(s_1 \mid \Omega_1) \tag{11}$$

for *t*=1. For calculating (9) and (10) we need to know, how the statistical inference about process in regime s_τ is changing after adding observation y_t to history of observations:

$$P(s_\tau \mid \Omega_t) - P(s_\tau \mid \Omega_{t-1}). \tag{12}$$

Actually this value represents weights of derived logarithm of function (6). The computations of (12) are a by-product of calculating smooth probabilities $P(s_\tau \mid \Omega_T)$. We can get these values also from iterative algorithm introduced by Hamilton in [3].

The unconditional probability of occurring regime s_1 represents vector of ergodic probabilities $\boldsymbol{\pi}$ [4], which we can compute as

$$\boldsymbol{\pi} = (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{e}_{N+1}, \tag{13}$$

where $\mathbf{A}_{(N\times 1)\times N} = \begin{bmatrix} \mathbf{I}_N - \mathbf{P} \\ \mathbf{1}' \end{bmatrix}$, $\mathbf{1}_{Nx1} = (1,\ldots,1)'$. and $\mathbf{e}_{N+1}$ represents $(N+1)$ column of identical matrix $\mathbf{I}_{N+1}$.

White in paper [8] has suggested autocorrelation tests of score functions, which use conditional moment tests from Newey [6] and Tauchen [7]. Hamilton [3] derived it further for Markov-switching models.

We need $(l\times 1)$ vector $\mathbf{c_t}(\hat{\boldsymbol{\theta}})$ for constructing this test. Vector $\mathbf{c_t}(\hat{\boldsymbol{\theta}})$ contains the examined elements from $[\mathbf{h_t}(\boldsymbol{\theta})]\cdot[\mathbf{h_{t-1}}(\boldsymbol{\theta})]'$. The null hypothesis tests independence $\mathbf{c_t}(\boldsymbol{\theta})$ on $\mathbf{c_{t-1}}(\boldsymbol{\theta})$. It means we cannot predict values of vector at time *t* from values gaining at time *t*-1.
Since we want to test validity of Markov assumptions, it means we want to verify these assumptions:

$$P(s_t = i \mid s_{t-1} = i) = P(s_t = i \mid s_{t-1} = i, y_{t-1}) \qquad i = 1,\ldots,N. \tag{14}$$

$$P(s_t = i \mid s_{t-1} = i) = P(s_t = i \mid s_{t-1} = i, s_{t-2} = j) \qquad i, j = 1,\ldots,N. \tag{15}$$

So vector the $\mathbf{c_t}(\boldsymbol{\theta})$ contains $2N$ elements in this form:

$$\frac{\partial \ln f(y_t \mid \Omega_{t-1};\boldsymbol{\theta})}{\partial p_{ii}} \cdot \frac{\partial \ln f(y_{t-1} \mid \Omega_{t-2};\boldsymbol{\theta})}{\partial \phi_i(1)} \qquad i = 1,\ldots,N, \tag{16}$$

$$\frac{\partial \ln f(y_t \mid \Omega_{t-1};\boldsymbol{\theta})}{\partial p_{ii}} \cdot \frac{\partial \ln f(y_{t-1} \mid \Omega_{t-2};\boldsymbol{\theta})}{\partial p_{ii}} \qquad i = 1,\ldots,N. \tag{17}$$

The test statistic has form:

$$\left[T^{-1/2}\sum_{t=1}^{T}\mathbf{c_t}(\hat{\boldsymbol{\theta}})\right]'\left[T^{-1}\sum_{t=1}^{T}\mathbf{c_t}(\hat{\boldsymbol{\theta}})\cdot[\mathbf{c_t}(\hat{\boldsymbol{\theta}})]'\right]^{-1}\left[T^{-1/2}\sum_{t=1}^{T}\mathbf{c_t}(\hat{\boldsymbol{\theta}})\right] \to \chi^2(l) \tag{18}$$

and has asymptotic $\chi^2(l)$ distribution, where *l* is number of $\mathbf{c_t}(\boldsymbol{\theta})$ elements . More details are available in [3].

4. New approach to testing linearity and remaining nonlinearity

For testing linearity, the null hypothesis represents validity of Markov assumptions. It means that if model doesn't confirm assumptions (14) and (15) (we reject null hypothesis), we cannot consider corresponding Markov-switching model as appropriate for examined time series. We calculate test statistic (18) for 2-regime model and find *p*-value from $\chi^2(4)$ distribution. Compare this *p*-value with given significance level α (the most common 0,05 or 0,01). If *p*-value $< \alpha$, we will reject the null hypothesis, in other words, the examined process doesn't confirm first-order Markov chain assumptions and we cannot use Markov-switching model for this time series.

If the previous testing of linearity for 2-regime model doesn't reject the null hypothesis, we can test the remaining non-linearity by the similar manner and use it to compare 2-regime and 3-regime models. We test the null hypothesis about appropriateness of a 2-regime model against alternative hypothesis about 3-regime model. After calculating the test statistic (18) for 3-regime model (we have this test statistic for a 2-regime model already) and corresponding *p*-values (test statistic is from $\chi^2(6)$ distribution for 3-regime model), we compare these *p*-values with the given significance level.

The following alternatives can occur:

- Non-rejecting the null hypothesis for a 2-regime model, rejecting null hypothesis for a 3-regime model. It means that the 2-regime model is appropriate.
- Non-rejecting the null hypothesis for a 2-regime model, non-rejecting the null hypothesis for a 3-regime model. The model with greater *p*-value from testing validity of Markov properties is the appropriate model. In this case we can check this results by other criterions like BIC (Bayesian Information Criterion) values for both types of models (better model has lower BIC value), residual dispersion, forecasting error values, results for testing autocorrelation and so on.

5. Testing for selected Slovak macroeconomic indicators

We applied this testing of linearity against Markov-switching type of non-linearity in modeling selected time series representing some Slovak macroeconomic indicators, specifically gross domestic product (GDP), inflation, unemployment, consumption and real wages.

Testing linearity and remaining Markov-switching type of non-linearity results are in Table 1 (2R means 2-regime model, 3R means 3-regime model and *q* is order of Markov-switching model). We can see that at the significance level α=0.05 there are models with bold *p*-value which pass through testing linearity, so these 2-regime models confirmed Markov assumptions. Further we test remaining non-linearity with these models as follows by calculating *p*-value of test statistic (18) for 3-regime models and then comparing corresponding *p*-values (both for 2 and 3-regime models) and choosing greater the one of them. These *p*-values are highlighted with grey color. It means that the model with highlighted color is more suitable.

Table 1: p-values of test statistic (18) for both 2-regime and 3-regime models

	p-value									
Order	q=1		q=2		q=3		q=4		q=5	
Regimes	2R	3R	2R	3R	2R	3R	2R	3R	2R	3R
GDP	**0.339**	0.006	**0.491**	0.037	0.044	0.007	**0.447**	<0.001	**0.405**	0.010
Consumption	**0.098**	0.373	**0.053**	0.297	**0.237**	<0.001	**0.095**	0.101	**0.307**	0.321
Real wages	**0.375**	<0.001	0.022	0.017	0.003	0.006	**0.201**	0.011	**0.471**	0.033
Inflation	**0.136**	<0.001	0.023	0.070	0.031	0.275	**0.106**	0.006	0.024	0.010
Unemployment	**0.407**	0.049	**0.370**	0.019	**0.145**	0.017	**0.375**	0.132	**0.114**	0.085

6. Conclusion

Markov-switching models are important tool in modeling time series, specifically these ones, which show non-linear character. These time series are mainly sensitive to outside influences and we can observe periods of expansion and recession there.

Diagnostic tests were the centre of the theoretic part in this contribution. We use the score test for validity of Markov assumptions to substitute the classical test linearity against Markov-switching type of non-linearity. This contribution solves problem of time consuming test linearity because of using simulations.

7. References

[1]Hamilton, J.D. 1989. A new approach to the economic analysis of nonstationary time series and the business cycle. In: Econometrica, No. 57, 1989, p. 357-384.

[2]Hamilton, J.D. 1990. Analysis of time series subject to changes in regime. In: Journal of Econometrics, No. 45, 1990, p. 39-70.

[3]Hamilton, J.D. 1996. Specification testing in Markov-switching time series models. In: Journal of Econometrics, No. 70, 1996, p. 127-157.

[4]Hamilton, J.D.1997. Time series analysis. Princeton University Press, 1994. p. 820 ISBN 978-0-691-04289-3.

[5]Hansen, B. E. 1992. The likelihood ratio test under nonstandard assumptions: testing the Markov Switching model of GNP. In: Journal of Applied Econometrics, No. 7, 1992, p. 61- 82.

[6]Newey, W. K.1985. Maximum likelihood specification testing and conditional moment tests. In: Econometrica, No. 53, 1985, p. 1047-1070.

[7]Tauchen, G. 1985. Diagnostic testing and evaluation of maximum likelihood models. In: Journal of Econometrics, No. 30, 1985, p. 415-443.

[8]White, H.1987. Specification testing in dynamic models. In Truman F. Bewley, editors, Advances in econometrics. Fifth world congress, Vol. 2, Cambridge: Cambridge University Press, 1987.

Author's address:

Jana Lenčuchová, Mgr.
Department of Mathematics
Faculty of Civil Engineering
Slovak University of Technology Bratislava
Radlinskeho 11
813 68 Bratislava
lencuchova@math.sk

Acknowledgement: *The support of the grant APVV No. LPP-0111-09 is kindly announced.*

Claim Reserving Calculation
Výpočet pojistných rezerv

Bohdan Linda, Jana Kubanová

Abstract: The aim of this paper is to show how to use non-parametric bootstrap at claim reserving calculation.

Abstrakt: Obsahem tohoto příspěvku je ukázat využití neparametrického bootstrapu při výpočtu pojistných rezerv

Key words: Claims reserves, insurance benefit, claim settlement, Chain Ladder Method, bootstrap method

Klíčová slova: Pojistná rezerva, pojistné plnění, likvidace škody, Chain Ladder metoda, bootstrapová metoda

1. Introduction

The methods of the claim reserving calculation were based on the strictly deterministic methods not so long time ago. The insurance benefits started to be regarded as a random process or a time series as late as the last 30 years. However, we know to count the parameter estimates of the random processes analytically only in the cases, when certain assumptions about probability distribution of the measured factors are fulfilled. These assumptions are in practice often verifiable with difficulties. Therefore, estimates often turn to various asymptotical approximations. Regarding to the small number of available data, the asymptotic theory needn´t be always valid. One of the methods that do not request large files of raw data is bootstrap method. The aim of this paper is to show how these methods can help to solve claim reserving problem.

2. Theoretical basis

In the non-life insurance are common such types of insurance benefits, which are not totally covered in the year of accident. Examples of such insurance are accident insurance, car insurance, property insurance, travel insurance and many others.

To be able to cover these claims, the insurance company has to estimate the future payments and create certain claim reserves. We consider two basic types of claim reserves for past exposures:

a) IBNR (Incurred but not reported). The reserve for insurance benefit from the claims that have occurred but haven't been reported yet corresponds with it.

b) IBNS (Reported but not settled). This means not settled insured accident corresponding with the reserve for the insurance benefit from the claims that have been reported but have not been settled. The payment is expected in the future.

There are many mathematical, statistical and numerical methods that can help us to calculate these reserves. The summarization of the claim reserving problem can be following: On the base of available information about the past are estimated future payments. This information has to concern similar situations, similar amount of payments and so on.

Methods for claim reserves calculation

The classification of these methods is not goal of this paper, so that only the best known and the most often used methods are in the following list.

a) Chain-ladder method, inflation adjusted Chain ladder method

b) The separation method

c) Bornhuetter-Ferguson method

d) Poison model for claim counts

e) Bayesian models – Benktander-Hovinen method

- Cape Cod model

f) Generalized linear model

g) Bootstrap method

3. Chain Ladder method

The mostly used method is Chain ladder method. The values $C_{i,j}$ represent the cumulative claim amounts occurred from the accident year i and development year j. These data are structured in the table 1.The assumptions of application of this method are:

- Maximally n years are needed to the total claim settlement.

- Development of the sums of paid insurance benefits are characterized by the "development coefficient of the insurance benefit" λ_j, it means that

$$C_{i,j} = \lambda_j C_{i,j-1} \tag{1}$$

- Policy portfolio is homogenous.

- Inflation is stable during development years.

Table1 Claim amounts

accident year i	development year j						
	0	**1**	**2**	**...**	***n*-2**	***n*-1**	***n***
0	$C_{0,0}$	$C_{0,1}$	$C_{0,2}$	**...**	$C_{0,n-2}$	$C_{0,n-1}$	$C_{0,n}$
1	$C_{1,0}$	$C_{1,1}$	$C_{1,2}$	**...**	$C_{1,n-2}$	$C_{1,n-1}$	
2	$C_{2,0}$	$C_{2,1}$	$C_{2,2}$	**...**	$C_{2,n-2}$		
...	**...**	**...**	**...**	**...**			
***n* - 1**	$C_{n-1,1}$	$C_{n-1,2}$					
n	$C_{n,1}$						

The point of interest is to estimate the missing values in the rows, which express the amounts that are necessary to cover the claims occurred in the year i. The missed values are calculated according to the model (1). The simplest way of coefficients ´ λ_j estimate is quotient

$$\hat{\lambda}_j = \frac{C_{0j} + C_{1j} + ... + C_{n-j\,j}}{C_{0\,j-1} + C_{1\,j-1} + ... + C_{n-j\,j-1}} \quad j = 1,2,...n \tag{2}$$

Example: The data published by Taylor and Ashe (1983) are used in the example. They express the cumulative values of insurance benefits and these data are stated in the "white" part of the table 2 (the upper triangle matrix).

The cumulative values of insurance benefits are calculated in the "grey" part of the table 2. These values were calculated according to the model (1), in which the coefficients λ_j were calculated according to the formula (2).

Table 2 Cumulative values of insurance benefits

		development year									
i		**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
	0	357848	1124788	1735330	2218270	2745596	3319994	3466336	3606286	3833515	3901463
	1	352118	1236139	2170033	3353322	3799067	4120063	4647867	4914039	5339085	5433719
	2	290507	1292306	2218525	3235179	3985995	4132918	4628910	4909315	5285148	5378826
	3	310608	1418858	2195047	3757447	4029929	4381982	4588268	4835458	5205637	5297906
	4	443160	1136350	2128333	2897821	3402672	3873311	4207459	4434133	4773589	4858200
	5	396132	1333217	2180715	2985752	3691712	4074999	4426546	4665023	5022155	5111171
	6	440832	1288463	2419861	3483130	4088678	4513179	4902528	5166649	5562182	5660771
	7	359480	1421128	2864498	4174756	4900545	5409337	5875997	6192562	6666635	6784799
	8	376686	1363294	2382128	3471744	4075313	4498426	4886502	5149760	5544000	5642266
	9	344014	1200818	2098228	3057984	3589620	3962307	4304132	4536015	4883270	4969825

The main diagonal in the scheme shows the paid insurance benefits for the claims from the accident year $i = 0, 1, 2, ..., n$ till the end of the development year.

The stated model (1) is deterministic, it means, that it assumes $C_{i,j}$ to be constant values. In reality, we have to consider these variables to be random variables.

There are many models that express a dependency between variables $C_{i,j}$. To demonstrate the bootstrap method we selected model

$$C_{i,j+1} = \lambda_j \cdot C_{i,j} + \sigma_j \sqrt{C_{i,j}}\, \varepsilon_{i,j+1}. \qquad (3)$$

that can be considered as an autoregressive process.

Coefficients λ_j, σ_j can be estimated by many methods, for example, least square methods and others.

When the parameters λ_j, σ_j are estimated, we can proceed to estimate benefits $C_{i,j}$ for $n < i + j < 2n$. These benefits´ estimates are marked $\hat{C}_{i,j}$.

$\hat{C}_{n,1} = \hat{\lambda}_1 \cdot C_{n,0} + \sigma_j \sqrt{C_{n,0}}$, $\hat{C}_{n,2} = \hat{\lambda}_2 \cdot C_{n,1} + \sigma_j \sqrt{C_{n,1}}$, $\hat{C}_{n,3} = \hat{\lambda}_2 \cdot C_{n,2} + \sigma_2 \sqrt{C_{n,2}}$, ...

This process shows us how to complete step by step the whole triangle below the diagonal.

4. Bootstrap approach

If we assume that the type of distribution of random variables $C_{i,j}$ is unknown, we have to use non-parametric bootstrap. The base of non-parametric bootstrap is creation of the bootstrap samples from the empirical distribution function of the measured values. By the help of these values are calculated bootstrap replications of the estimate of the wanted parameter, in our case λ_j. The estimates of residuals $\hat{\varepsilon}_{i,j}$ are used in autoregressive models, for generation of the bootstrap samples. Residuals are calculated according to the formula [Wutrich, Merz]: $\hat{\varepsilon} = \dfrac{F_{i,j} - \hat{f}_{j-1}}{\sigma_{j-1}} \sqrt{C_{i,j-1}}$

Where $F_{i,j} = \dfrac{C_{i,j}}{C_{i,j-1}}$ and $\hat{\lambda}_j = \dfrac{\sum_{i=1}^{I-j} C_{i,j+1}}{\sum_{i=1}^{I-j} C_{i,j}}$.

It is recommended to work with so-called adjusted residuals instead of residuals $\hat{\varepsilon}_{i,j}$, which can be obtained according to the following formula:

$$Z_{i,j} = \frac{\hat{\varepsilon}_{i,j}}{\sqrt{1 - \frac{C_{i,j-1}}{\sum_{i=1}^{I-j+1} C_{i,j-1}}}}.$$

The bootstrap replications λ_j^* are earned by the following procedure. The bootstrap sample, element of which are indicated $Z_{i,j}^*$, is obtained by resampling of the data $Z_{i,j}$.

By the help of these adjusted bootstrap residuals the bootstrap sample of measured values is gained, elements of which are $C_{i,j}^*$. This bootstrap sample is used for calculation of the bootstrap replications $\hat{\lambda}_j^*$ by the following way.

We put $C_{i,1}^* = C_{i,1}$

The other values $C_{i,1}^*$ for $i + j < J$ are calculated according to the process stated below.

$$C_{i,j}^* = \hat{\lambda}_{j-1} C_{i,j-1} {}^* + \hat{\sigma}_{j-1} \sqrt{C_{i,j-1}}\, Z_{i,j}^*$$

$$\hat{\lambda}_j^* = \sum_{i=1}^{I-j} \frac{C_{i,j+1}^*}{\sum_{i=1}^{I-j} C_{i,j}^*} \frac{C_{i,j+1}^*}{C_{i,j}^*}$$

Table 3 Cumulative values of the insurance benefit calculated by bootstrap method

		development year									
i		**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
	0	357848	1124788	1735330	2218270	2745596	3319994	3466336	3606286	3833515	3901463
	1	352118	1236139	2170033	3353322	3799067	4120063	4647867	4914039	5339085	5430383
	2	290507	1292306	2218525	3235179	3985995	4132918	4628910	4909315	5154290	5242428
	3	310608	1418858	2195047	3757447	4029929	4381982	4588268	4800246	5039778	5125958
	4	443160	1136350	2128333	2897821	3402672	3873311	4231205	4426687	4647578	4727052
	5	396132	1333217	2180715	2985752	3691712	4060514	4435706	4640635	4872203	4955517
	6	440832	1288463	2419861	3483130	4117060	4528354	4946774	5175315	5433563	5526477
	7	359480	1421128	2864498	4153809	4909802	5400291	5899278	6171824	6479798	6590603
	8	376686	1363294	2383038	3455643	4084570	4492619	4907737	5134474	5390685	5482865
	9	344014	1199783	2097221	3041180	3594675	3953783	4319113	4518656	4744137	4825261

The missing bootstrap values $\hat{C}_{i,j}^*$ for $i + j > J$ are kept count of according to the formula $C_{i,j}^* = \prod_{k=1}^{J-i} \hat{\lambda}_{J-k}^* \hat{C}_{i,i}$.

The above stated process is repeated *R*–times and for each pair of indices *i,j* $i + j > J$. We gain by this way *R* bootstrap replications. The unknown values $C_{i,j}$ are then estimated by average from the appropriate bootstrap replications.

The calculated cumulative values of the insurance benefit are presented in the grey part of the table 3. These values were calculated by bootstrap method, based on the model (3).

5. Conclusion

Both stated methods can be confronted, if we compare the calculated values in the tables 2 and 3. Differences between results calculated by classical Chain ladder method and bootstrap method are stated in the table 4.

Table 4 Cumulative values of the insurance benefit calculated by bootstrap method

		development year									
i		**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
	0	0	0	0	0	0	0	0	0	0	0
	1	0	0	0	0	0	0	0	0	0	3335,461
	2	0	0	0	0	0	0	0	0	130858,7	136398,1
	3	0	0	0	0	0	0	0	35212	165859,1	171947,4
	4	0	0	0	0	0	0	-23745,9	7446,62	126010,8	131147,8
	5	0	0	0	0	0	14484,6	-9159,41	24388,3	149952,3	155654
	6	0	0	0	0	-28381,6	-15175	-44245,6	-8666	128619,5	134293,7
	7	0	0	0	20947,6	-9257,06	9045,61	-23281,3	20737,6	186836,3	194196
	8	0	0	-909,805	16100,8	-9257,63	5807,15	-21234,5	15285,2	153315,7	159400,9
	9	0	1034,29	1006,57	16803,6	-5055,46	8523,49	-14980,4	17359	139133,4	144563,3

6. Literature

[1]ARLT, J. *Moderní metody modelování ekonomických časových řad*, Praha, Grada Publishing, 1999. ISBN 80-7169-593-4.

[2]Box, G. E. P., Jenkins, G. M., Reinsel, G. C. *Time Series Analysis. Wilwy 2008.* ISBN 978-0-470-27284-8.

[3]EFRON, B, TIBSHIRANY, R. J. *An Introduction to the Bootstrap*. Chapman&Hall/CRC, 1993. ISBN 0-412-04231-2.

[4]LINDA, B., KUBANOVÁ, J. *Přesnost odhadu v autoregresních modelech*. In: Forum Statisticum Slovacum 4/2007. SŠDS Bratislava 2007. s.81-88. ISSN 1336-7420.

[5]PACÁKOVÁ, V. *Poistná štatistika. Iura Edition.* Bratislava 2004. ISBN 80-8078-00-8.

[6]PINHEIRO, J.R., ANDRADE e SILVA, J.M., CENTENO, M.,L. *Bootstrap Methodology in Claim Reserving*. Working papers. Centro de Mathemática Aplicada á Previsao e Decisao Económica. Lisboa, 2000.

[7]RUBLÍKOVÁ, E. *Analýza časových radov*. Iura Edition. Bratislava 2007. ISBN 978-80-8078-139-2.

[8]STANKOVIČOVÁ I., VOJTKOVÁ M. *Viacrozmerné štatistické metódy s aplikáciami*. Bratislava. **IURA EDITION** 2007. ISBN 978-80-8078-152-1

[9] VERALL, R.,J. *Statistical Methodsfor the Chain Ladder Technique*. *www.casact.org/pubs/forum/94spforum/94spf393.pdf*

[10] WEI, W. W. S. *Time series analysis*. *Univariate and Multivariate Methods*. Addison-Wesley Publishing Company. 1994

[11] WŰTHRICH, M.V. *Stochastic Claims Reserving Methods in Insurance*. John Wiley 2008.

Addresses of authors:

Bohdan Linda, doc., RNDr., CSc.
Studentská 95, 532 10 Pardubice
bohdan.linda@upce.cz

Jana Kubanová, doc., PaedDr, CSc.
Studentská 95, 532 10 Pardubice
jana.kubanova@upce.cz

Remark: *The article was elaborated with the support of the grant GAČR 402/09/1866, Modelling, simulations and management of insurance risk.*

Srovnání zemí EU z hlediska demografických ukazatelů a shluky podobncýh zemí

Comparison of EU countries from the viewpoint of Demographic data and clusters of similar countries

Tomáš Löster, Jana Langhamrová

Abstract: The ageing of the population is a contemporary problem that many economists and demographers are attempting to tackle. The composition of the age structure and its expected development will have fundamental consequences not only for the actual reproduction of the population, but especially for the functioning of the economy as a whole. With regard to the present trend in population ageing it will be necessary to devote considerable attention to various pillars of the economy such as insurance. The aim of this paper is to evaluate the demographic development of various demographic indicators of the 27 EU member-states. Among these indicators one may include the proportion of persons of the child generation, the parent generation and the grandparent generation, the age index and the index of the dependency of seniors.

Key words: Ageing of the population, demographic development, proportion of persons of the child generation, proportion of persons of the grandparent generation, dendrogram.

Kľúčové slová: Stárnutí obyvatelstva, demografický vývoj, podíl osob dětské generace, podíl osob prarodičovské generace, dendrogram.

1. Úvod

Evropská unie má k počátku roku 2009 celkem 27 členských států, jednou z nich je Česká republika. Problém stárnutí obyvatelstva trápí nejednoho demografa, ale také ekonoma z každého ze států Evropské unie. Je proto vhodné udělat si představu, jak jsou na tom jednotlivé členské země z hlediska různých demografických ukazatelů, mezi které patří podíl osob dětské generace, podíl osob rodičovské generace, podíl osob prarodičovské generace, index stáří a index závislosti seniorů. U těchto ukazatelů se soustředíme na jejich vývoj v závislosti na dostupnosti údajů v letech 1950-2006 a také si stanovíme základní míry dynamiky vývoje časových řad v letech 1993-2006, tj. od doby vzniku Evropské unie.

2. Vývoj demografických ukazatelů v zemích EU

Z tabulky č. 1, která obsahuje průměrné absolutní meziroční přírůstky demografických ukazatelů v letech 1993-2006 vyplývá, že v letech 1993 – 2006 kromě Dánska a Lucemburska docházelo k průměrnému meziročnímu poklesu podílu dětské generace ve všech zemích. K největšímu poklesu dětské generace docházelo v Polsku, kde průměrný meziroční pokles činil 0,605 %. Vývoj podílu rodičovské generace v uvedených letech není u zemí Evropské unie tak jednoznačný jako vývoj podílu dětské generace. V některých členských zemích docházelo k průměrnému meziročnímu nárůstu podílu této věkové skupiny (např. u Bulharska, Estonka, Lotyšska). Situace je zcela jednoznačná u vývoje podílu osob prarodičovské generace. Ve sledovaném období ve všech členských zemích Evropské unie dochází k nárůstu podílu této generace na celkovém počtu obyvatel. Největší meziroční nárůst je ve Finsku, kde činí 0,613 %. Další sloupec tabulky 1 představuje vývoj indexu stáří, který představuje počet osob starších 50ti let připadajících na osobu ve věku 0–14 dokončených let. Z vývoje vyplývá, že v průměru meziročně tento podíl ve všech zemích narůstal. Z vývoje

indexu závislosti seniorů vyplývá, že kromě Irska, Malty, Slovenska a Švédska docházelo k průměrnému meziročnímu nárůstu tohoto ukazatele.

Tabulka 1: Průměrné meziroční přírůstky demografických ukazatelů v letech 1993-2006

Země	podíl osob dětské generace	podíl osob rodičovské generace	podíl osob prarodičovské generace	index stáří	index závislosti seniorů
Belgie	-0,0878	-0,1993	0,2870	0,0260	0,2121
Bulharsko	-0,3969	0,0191	0,3778	0,0801	0,2160
Česká republika	-0,3897	-0,1571	0,5469	0,0778	0,0077
Dánsko	0,1159	-0,4179	0,3021	0,0048	0,0146
Estonsko	-0,4887	0,1687	0,3200	0,0691	0,4873
Finsko	-0,1568	-0,4559	0,6127	0,0501	0,3341
Francie	-0,1097	-0,2835	0,3932	0,0304	0,2108
Irsko	-0,3742	0,1776	0,1966	0,0272	-0,2403
Itálie	-0,0670	-0,2406	0,3076	0,0328	0,5069
Kypr	-0,5593	0,1934	0,3659	0,0506	0,0133
Litva	-0,4826	0,2061	0,2764	0,0565	0,4203
Lotyšsko	-0,5499	0,3164	0,2335	0,0759	0,4330
Lucembursko	0,0144	-0,0689	0,0545	0,0017	0,0579
Maďarsko	-0,2652	-0,1439	0,4091	0,0560	0,1396
Malta	-0,2276	-0,1158	0,3434	0,0382	-0,0762
Německo	-0,1893	-0,0960	0,2854	0,0492	0,6698
Nizozemsko	-0,0205	-0,4282	0,4488	0,0265	0,2040
Polsko	-0,6051	0,0419	0,5633	0,0761	0,1923
Portugalsko	-0,2251	-0,0945	0,3196	0,0453	0,2613
Rakousko	-0,1706	-0,0762	0,2469	0,0350	0,2386
Rumunsko	-0,4785	0,1742	0,3043	0,0611	0,2869
Řecko	-0,2823	0,0217	0,2606	0,0547	0,4245
Slovensko	-0,5689	0,1118	0,4571	0,0647	-0,0277
Slovinsko	-0,3916	-0,1267	0,5183	0,0793	0,4408
Spojené království	-0,1462	-0,0482	0,1945	0,0244	-0,0125
Španělsko	-0,2302	0,0205	0,2098	0,0422	0,0990
Švédsko	-0,1306	-0,1639	0,2945	0,0310	-0,0707

Zajímavý je i pohled na země EU z pohledu vývoje biologických a ekonomických generací. Ve všech zemích je viditelný podobný trend týkající se stárnutí populací sledovaných zemí.

Podíl osob dětské generace (0-14)

Z vývoje podílu osob dětské generace, tj. osob ve věku 0-14 le,t je patrné, že od roku 1950 dochází k neustálému poklesu tohoto podílu. Jak zřejmé, že Česká republika se během sledovaného období nachází přibližně na průměrné úrovni. V ČR v roce 1960 byl podíl této generace přibližně 25 % a v roce 2006 již necelých 15 %, což vyjadřuje znatelný pokles. Za 46 let došlo k poklesu o více než 10 procentních bodů. V roce 1960 byl nejvyšší podíl dětské generace v Polsku, kde činil více než 33 %. Během následujících 46 let byl pokles dětské generace v Polsku více než o 17 % a tak v roce 2006 byl podíl této generace necelých 15 %, podobně jako v České republice.

Podíl osob rodičovské generace (15-49)

Z vývoje je zřejmé, že podíl osob rodičovské generace se v dlouhém období pohybuje přibližně okolo 50 %. Nejnižší hodnoty bylo dosaženo ve Francii v roce 1960. Podíl rodičovské generace zde byl pouze 44 %. Naopak nejvyššího podílu bylo dosaženo na Slovensku v roce 2002, kdy tento podíl činil téměř 55 %. V České republice vývoj tohoto podílu vykazuje určité vlny, přičemž ve sledovaném období dochází k nárůstu podílu těchto osob a to z hodnoty 46 % v roce 1960 na 50 % v roce 2006. Nejvyššího podílu rodičovské generace v České republice bylo dosaženo v roce 1995, kdy podíl osob činil téměř 53 %.

Podíl osob prarodičovské generace (50+)

Z vývoje prarodičovské generace je zcela zřejmá rostoucí tendence podílu těchto osob ve všech zemích EU. Znatelný nárůst prarodičovské generace vykazuje Itálie. Z 25 % v roce 1960 hodnota tohoto podílu vzrostla až na téměř 39 % v roce 2006. Česká republika během celého sledovaného období zastává jednu z předních pozic a podíl této generace je ve srovnání s ostatními zeměmi vysoký. V roce 1960 byl podíl prarodičovské generace necelých 30 % a v roce 2006 byl více než 35 %.

Index stáří

Index stáří vyjadřuje podíl prarodičovské generace k dětské generaci. Z vývoje vyplývá zcela zřejmý rostoucí trend tohoto indexu což znamená, že dochází ve většině zemí ke stárnutí obyvatelstva. V České republice v roce 1960 byl index stáři přibližně 1,1 což vyjadřuje, že obě generace byly přibližně stejně zastoupeny. V roce 2006 byla hodnota indexu stáří v České republice téměř 2,5 což znamená, že prarodičovská generace významně převyšuje podíl dětské generace a dochází k velmi výraznému stárnutí obyvatel.

Z vývoje biologických generací vyplývá, že téměř ve všech zemích dochází k úbytku dětské generace a dochází k nárůstu prarodičovské generace což znamená, že dochází ke stárnutí obyvatelstva a to navíc se zrychlující se úrovní. Rodičovská generace je přibližně na úrovni 50 %.

Podíl osob v produktivním věku (20-64)

Vývoj podílu osob v produktivním věku vykazuje opět patrný rostoucí trend. Česká republika se z hodnoty 57 % v roce 1960 dostává na přední místo a v roce 2006 dosahuje hodnota tohoto ukazatele úrovně 65 %.

Index závislosti seniorů

Index závislosti seniorů je definován jako poměr osob v poproduktivním věku ku osobám ve věku produktivním. I tento index během sledovaného období vykazoval rostoucí tendenci, pouze v období mezi roky 1980-1990 hodnota tohoto indexu klesá, případně stagnuje, téměř ve všech zemích EU. Česká republika hodnotou svého indexu závislosti seniorů reprezentuje průměrnou resp. lehce podprůměrnou úroveň ve srovnání s ostatními zeměmi EU. Hodnota tohoto indexu v roce 2006 činí 22 %.

Úhrnná plodnost

Vývoj úhrnné plodnosti vykazuje v letech 1993-2007 patrný klesající směr vývoje. Největší pokles nastal na Slovensku, kdy z hodnoty 1,9 došlo k poklesu na 1,2. Česká republika v letech 1993-1999 vykazovala velmi dramatický pokles, v roce 1993 došlo k poklesu na nejnižší úroveň z celé EU v celém období, a to na hodnotu 1,13, což je velice hluboce pod záchovnou hranici prosté reprodukce.

Střední délka života žen při narození

Z vývoje střední délky života žen při narození vyplývá, že dochází k jejímu neustálému prodlužování se. Znatelný nárůst této střední délky života při narození vykazuje Česká

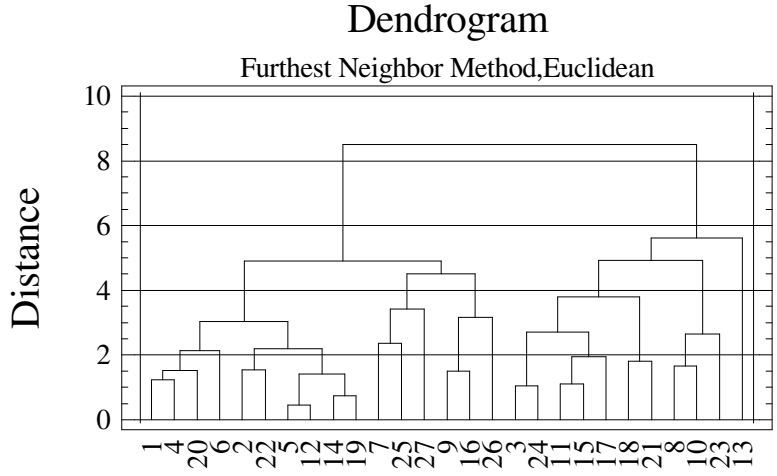
republika, která z hodnoty 76,4 v roce 1993 vzrostla na hodnotu 79,9 let. Nejvyšší střední délku života při narození žen v letech 1993-2007 vykazuje Francie, kde z úrovně 81,45 vzrostla na 84,5. Naopak nejnižší střední délku života žen při narození v celém sledovaném období vykazuje Rumunsko, kde z hodnoty 73,32 let v roce 1993 došlo k nárůstu na 76,14 let v roce 2007.

Střední délka života mužů při narození

Střední délka života mužů při narození vykazuje velmi podobnou tendenci vývoje jako střední délka života žen při narození. Během sledovaného období, tj. v letech 1993-2007 tento ukazatel vykazuje rostoucí trend. Česká republika zaznamenala nárůst z hodnoty 69,2 na 73,67 let. Na Slovensku dochází taktéž k růstu střední délky života, ovšem není tak rychlý, jako v České republice. Z 68,3 na 70,5 let. Nejvyšší střední délku života při narození mužů vykazuje Švédsko, kde z úrovně 75,5 vzrostla na 78,9 let. Naopak nejnižší střední délku života mužů při narození téměř v celém sledovaném období vykazuje Estonsko, kde z hodnoty 62,2 let v roce 1993 došlo k nárůstu na 67,36 let v roce 2007.

3. Shluky podobných zemí EU podle demografických ukazatelů

Pomocí shlukové analýzy, s využitím výše popisovaných proměnných, které představují demografické údaje z let 1993 – 2006 a na základě informacích o HDP na hlavu, počtu křesel v Evropském parlamentu byl sestaven dendrogram, který popisuje vznik jednotlivých shluků. Při měření vzdáleností mezi jednotlivými objekty (dílčí země) bylo použito metody nejvzdálenějšího souseda a při měření vzdáleností mezi shluky byla použita Eukleidovská vzdálenost. Postupné shlukování jednotlivých zemí podle podobnosti obsahuje graf 1.



Graf 1: Dendrogram – shluky podobných zemí podle demografických údajů

Z dendrogramu vyplývá, že existuje 7 významných shluků jednotlivých zemí, které se liší jednotlivými demografickými ukazateli. Čísla na ose X označují indexy zemí. Shluk č. 1 tvoří Belgie, Dánsko, Rakousko a Finsko. Skupina těchto států je charakteristická tím, že podíl prarodičovské generace je o něco vyšší, než je průměr v celé Evropské unii. To se samozřejmě projevuje i nadprůměrnou hodnotou indexu závislosti seniorů, ve srovnání s ostatními státy. Státy v tomto shluku vykazují nadprůměrnou úroveň HDP na obyvatele. Druhý shluk je tvořen Bulharskem, Řeckem Estonskem, Lotyšskem, Maďarskem a Portugalskem. I v tomto shluku je podíl nejstarší (prarodičovské) generace vyšší, než je průměr v EU, což se také projevuje nadprůměrnou úrovní indexu závislosti seniorů. Narozdíl od předchozího shluku mají tyto státy podprůměrnou úroveň HDP na obyvatele. Třetí shluk je

tvořen Francií, Spojeným královstvím a Švédskem, kde je podíl nejstarší generace na průměrné úrovni EU. Tato skupina států je reprezentována nadprůměrnou úrovní HDP. Čtvrtý shluk tvoří Itálie, Německo a Španělsko. Podíl prarodičovské generace je silně nadprůměrný ve srovnání s průměrem EU, stejně tak jako úroveň HDP v EU. Šestý shluk je reprezentován Českou republikou, Slovinskem, Litvou, Maltou a Nizozemskem. Státy v této skupině jsou z hlediska skladby věkové struktury a podílu prarodičovské generace na průměrné úrovni Evropské unie. Předposlední shluk reprezentuje Polsko a Rumunsko, ve kterých je podíl poslední generace spíše pod průměrem Evropské unie. Poslední shluk obsahuje Irsko, Kypr a Slovensko. V těchto státech je podíl nejstarší generace silně pod průměrem Evropské unie, což se projevuje i na hodnotě indexu závislosti seniorů, který je také pod průměrem EU. Posledním nezařazeným státem je Lucembursko, které se z hlediska svých demografických ukazatelů nehodí do žádného popsaného shluku.

4. Závěr

Z hlediska všech 27 států EU je u procesu stárnutí obyvatel nejhorší situace v Polsku, které vykazuje nevětší průměrný meziroční pokles podílu osob dětské generace (0,605 %) a vysoký nárůst podílu osob prarodičovské generace o 0,563 % meziročně. Situace v České republice je také vážná, neboť průměrný meziroční nárůst podílu osob prarodičovské generace (o 0,547 %) není kompenzován nárůstem podílu osob dětské generace, neboť ta v průměru meziročně klesá o 0,389 %.

Vzhledem k jednoznačnému stárnutí obyvatelstva je nezbytně nutné, aby jednotlivé členské státy EU věnovaly vysokou pozornost této problematice a jednotlivé země posilovaly význam různých pilířů například důchodových systémů.

5. Literatúra

[1] HEBÁK, P., HUSTOPECKÝ, J., ŘEZANKOVÁ, H., a kol.: Vícerozměrné statistické metody (1), Informatorium, Praha 2004.

[2] HEBÁK, P., HUSTOPECKÝ, J., ŘEZANKOVÁ, H., a kol.: Vícerozměrné statistické metody (3), Informatorium, Praha 2005.

[3] LANGHAMROVÁ, JITKA, FIALA, TOMÁŠ, LANGHAMROVÁ, JANA.: Ageing of the Population of the Czech Republic and its Economic Consequences in the Sphere of Pension Security and Financing of Health Care. Marrakech 27. 09. 2009 – 02. 10. 2009.

[4] STANKOVIČOVÁ, I., VOJTKOVÁ, M.: Viacrozmerné štatistické metódy s aplikáciami, Ekonómia, Bratislava 2007.

Adresa autorov:

Löster, Tomáš, Ing.
Vysoká škola ekonomická v Praze
Katedra statistiky a pravděpodobnosti
nám. W. Churchilla 4, 130 67 Praha 3
losterto@vse.cz

Langhamrová, Jana
Vysoká škola ekonomická v Praze
Katedra statistiky a pravděpodobnosti
nám. W. Churchilla 4, 130 67 Praha 3
JanaLanghamrova@seznam.cz

Korelácia javov a konfiguračno frekvenčná analýza
Correlation of events and Configural frequency analysis

Ján Luha, Zdenka Ruiselová

Abstract: This article is devoted to the application of specific statistical analysis named configural frequency analysis to some psychological questions and some specific analysis of the correlations of events.

Key words: configural frequency analysis, correlations of events.

Kľúčové slová: konfiguračno frekvenčná analýza, korelácia javov .

1. Úvod

V príspevku sa venujeme korelácii javov a konfiguračno frekvenčnej analýze (configural frequency analysis (**CFA**)) premenných s dvojprvkovými množinami hodnôt pri skúmaní javov v psychológii.

Konfiguračno frekvenčná analýza je jednou z metód analýzy viacrozmerných kontingenčných tabuliek. Ide o metódu (Lienert, 1971, Krauth, 1993, von Eye 1990, von Eye 2002), ktorá je u nás pomerne málo známa a využívaná. Preto sme sa rozhodli na niektoré možnosti jej využitia v tomto príspevku upozorniť. Podstatou tejto metódy je určovanie typov (antitypov), ako určitých konfigurácií viacerých premenných, signifikantne vyššie (nižšie) frekventovaných oproti ich očakávaným frekvenciám.

Aplikáciu v psychológii ilustrujeme na výskume, ktorý realizovala Z. Ruiselová so spolupracovníkmi primárne na súbore 470 respondentov zdravotných sestier (v súbore bolo aj 2,6% mužov zamestnaných ako zdravotná sestra) vo veku 22 – 57 rokov.

Ide o výskum kontrafaktového myslenia a jeho súvislosti s vybranými charakteristikami osobnosti. Na ilustráciu použitia spomínaných štatistických metód vyberáme z tohto výskumu dotazník – súbor 10 otázok, zostavených na báze štandardizovaného rozhovoru a týkajúcich sa kvantitatívnych i kvalitatívnych aspektov procesu kontrafaktového myslenia. Kontrafaktové myslenie je vlastne premýšľaním o neuskutočnených alternatívach riešenia bežných problémov. Stretáva sa s ním každý z nás v každodennom živote. Je to myslenie o tom, čo by mohlo byť, keby nastala iná alternatíva antecedentu (oproti tej, ktorá nastala) a o tom či by situácia dopadla ináč, alebo tak isto, ako v skutočnosti dopadla (rôzny alebo rovnaký konzekvent). Jednou vetou to možno stručne charakterizovať asi takto: Ako by to dopadlo, keby bola východzia situácia iná – a tu možno uvažovať o mnohých alternatívach.

Na priblíženie tohto typu myslenia sa uvádzajú mnohé príklady – uvedieme jeden z nich: Peter išiel včera domov z práce inou cestou ako inokedy a zastavil sa v kníhkupectve. Po ceste domov mal potom autonehodu. Kontrafaktové myslenie tu možno ilustrovať napríklad takto: Keby bol išiel tou istou cestou ako inokedy, nemuselo sa nič stať. Alebo: Keby sa nezdržal v kníhkupectve, nemuselo dôjsť k nehode – a podobne.

Niektorí autori chápu kontrafaktové myslenie ako regulátor správania, ktorého primárna funkcia je centrovaná na jeho koordináciu. Kontrafaktové myšlienky sú silno prepojené na ciele a podieľajú sa na regulácii správania najmä v sociálnych interakciách (Epstude, Roese, 2008, Segura, Morris, 2005). Podľa N. Roesea (2005) môže byť kontrafaktové myslenie funkčné – ak vedie k vhľadu do vhodnejšieho správania, prípadne k riešeniu (náprave) jednotlivcovho problému a môže byť korektívne – ak nasleduje po neúspešnej sociálnej skúsenosti. Vyskytuje sa viac po neúspechu, ako po úspechu. Negatívne skúsenosti evokujú tento typ myslenia omnoho frekventovanejšie ako pozitívne.

Tento typ myslenia silno súvisí s emóciami a so zvládaním náročných situácií. Viaceré výskumy ukázali, že porovnávanie s lepšími alternatívami riešenia (tzv. kontrafakty nahor) navodzuje viac negatívne emócie, kým porovnávanie s horšími alternatívami (tzv. kontrafakty nadol) navodzuje skôr emócie pozitívne. Podstatnou charakteristikou pri výskume kontrafaktového myslenia je aj kontrolovateľnosť antecedentu, teda možnosť ovplyvnenia, resp. vlastného manažmentu situácie.

Výskumný súbor bol vybratý aj vzhľadom na skutočnosť, že profesia zdravotnej sestry je na riešenie problémov a na výskyt záťažových situácií skutočne bohatá.

Toľko aspoň stručne na opis skúmaného problému, na ktorom v ďalšom texte ilustrujeme použitie hore uvedených štatistických metód. Pre lepšiu ilustráciu uvedieme aj dotazník – súbor 10 otázok, týkajúcich sa kontrafaktového myslenia.

PREMÝŠĽANIE O NEUSKUTOČNENÝCH ALTERNATÍVACH RIEŠENIA

(zostavila Z. Ruiselová)

Pohlavie: 1- muž; 2 - žena Vek: (v rokoch)

Každý z nás občas premýšľa o tom, ako by dopadli mnohé situácie (a riešenia problémov) v jeho živote, keby ...

▲▲▲

Zaujímame sa o toto premýšľanie. Odpovedzte, prosím na nasledujúce otázky. Nič sa nehodnotí, nie je tu ani dobré, ani zlé riešenie. Ide o to, pochopiť do akej miery sa ľudia zaoberajú tým, čo by sa bolo stalo (aká alternatíva by nastala), keby boli počiatočné podmienky problému iné (a aké?).

o1) Zaoberáte sa takýmito myšlienkami?

1 - nikdy; 2- zriedka; 3 – často; 4 – veľmi často

o2) Ak ste riešili nejaký problém a nedopadlo to dobre, pôsobí to na vás emočne?

1 – len málo; 2 – silno; 3 – veľmi silno

o3) Pôsobí na vás emočne aj úspech pri riešení problému?

1 – len málo; 2 – silno; 3 – veľmi silno

o4) Uvažujete o tom, ako to mohlo dopadnúť, keby ... – rozosmutňuje vás to?

1 - nikdy; 2- zriedka; 3 – často; 4 – veľmi často

o5) Uvažujete o alternatívach lepších ako tá, ktorá nastala v skutočnosti?

1 - nikdy; 2- zriedka; 3 – často; 4 – veľmi často

o6) Uvažujete o alternatívach horších ako tá, ktorá nastala v skutočnosti?

1 - nikdy; 2- zriedka; 3 – často; 4 – veľmi často

o7) Uvažujete častejšie o alternatívach

1 - lepších; 2 – horších

(ako to mohlo lepšie dopadnúť, keby ...) (ako to mohlo horšie dopadnúť, keby ...)

o8) Pomáha vám tento typ uvažovania pri budúcich problémoch?

1- skôr nie; 2 – skôr áno

o9) Brzdí vás tento typ uvažovania pri riešení problémov?

1- skôr nie; 2 – skôr áno

Otvorená otázka na konci dotazníka nebola do tejto etapy spracovania zaradená a tak ju neuvádzame.

Pre účely ďalšieho spracovania sme odpovede rekódovali, tak aby sme získali dvojprvkové množiny hodnôt a to nasledovne: o1: 1=1+2; 2=3+4; o2: 1=1; 2= 2+3; o3: 1=1; 2= 2+3; o4: 1=1+2; 2=3+4; o5: 1=1+2; 2=3+4 a o6: 1=1+2; 2=3+4. o7, o8 a o9 majú dvojprvkové množiny hodnôt, pričom o7 je logicky rekódovaná: 1=horších, 2=lepších.

Na analýzu sme použili aj premennú socex, reprezentujúcu zmysel pre koherenciu (SOC-Antonovsky) – osobnostnú charakteristiku, ktorá opisuje globálnu efektívnosť zvládania náročných situácií. Vybrali sme zo súboru zdravotných sestier extrémne skupiny podľa kritéria priemer +/- sigma, pričom 1= skupina s nízkou úrovňou SOC a 2= skupina s vysokou úrovňou SOC.

Výhodou otázok s dvojprvkovou množinou hodnôt je možnosť priameho počítania Pearsonovho korelačného koeficienta. Možno ich, prípadne rekódovať na indikátorové premenné, napr. 1=1; 2=0. Keďže sa hodnota korelačného koeficienta takouto manipuláciou nezmení a pre účely CFA je výhodnejšie pôvodné kódovanie hodnôt znakov, tak ho zachováme.

2. Základné štatistické charakteristiky skúmaného súboru

Skúmaný súbor zdravotných sestier sa skladal prevažne zo žien - mužov bolo iba 2,6%. Vekové rozpätie bolo od 22 do 57 rokov. Išlo o vysokoškolské študentky (a študentov) SZU.

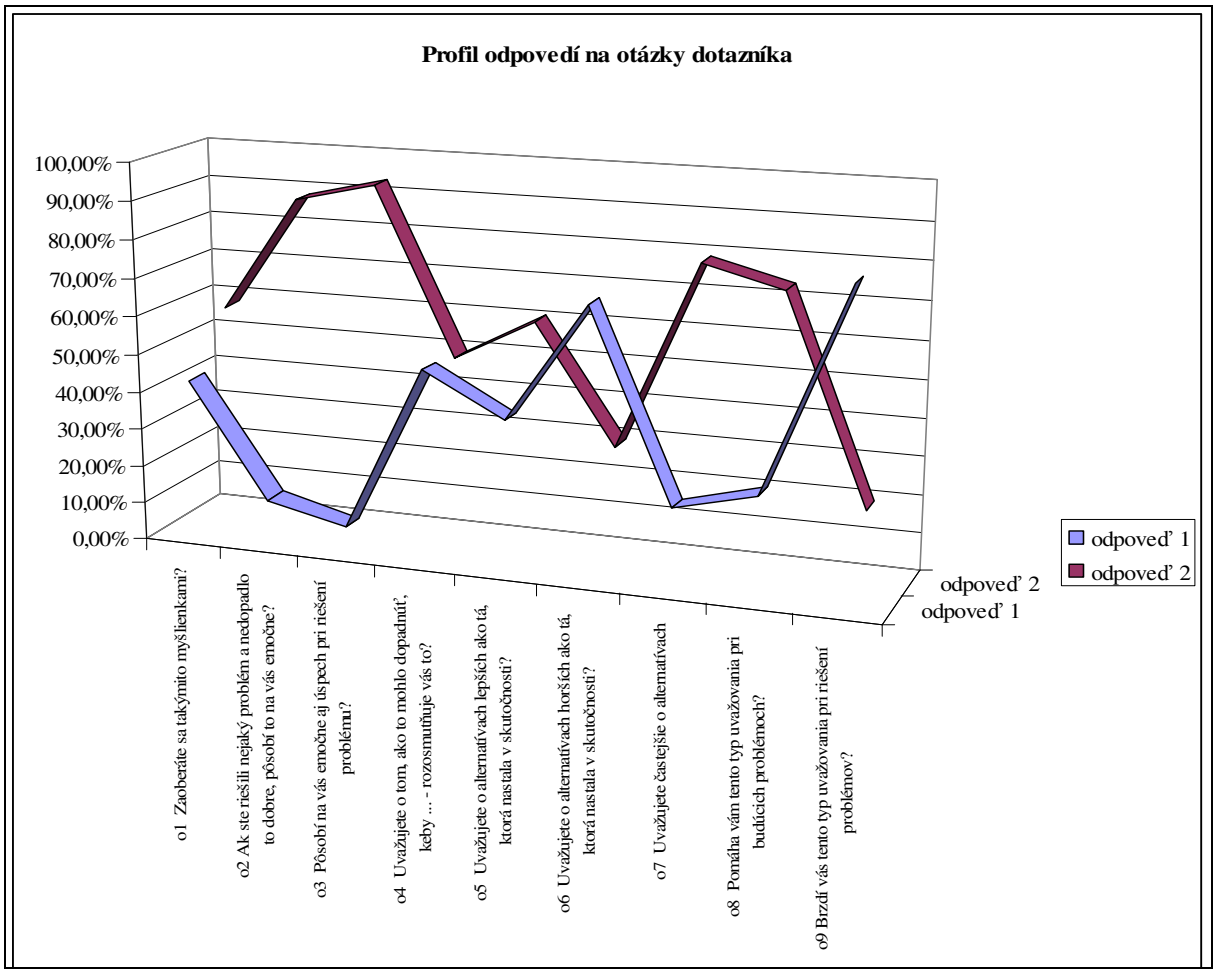
V tabuľke uvádzame prehľad percent odpovedí pre rekódované otázky. Najprv je percento na rekódovanú odpoveď č.1 a potom na č.2.

o1 Zaoberáte sa takýmito myšlienkami?	o2 Ak ste riešili nejaký problém a nedopadlo to dobre, pôsobí to na vás emočne?	o3 Pôsobí na vás emočne aj úspech pri riešení problému?	o4 Uvažujete o tom, ako to mohlo dopadnúť, keby ... – rozosmutňuje vás to?	o5 Uvažujete o alternatívach lepších ako tá, ktorá nastala v skutočnosti?	o6 Uvažujete o alternatívach horších ako tá, ktorá nastala v skutočnosti?	o7 Uvažujete častejšie o alternatívach -horších -lepších?	o8 Pomáha vám tento typ uvažovania pri budúcich problémoch?	o9 Brzdí vás tento typ uvažovania pri riešení problémov?
42,5%	12,2%	7,1%	50,9%	39,7%	71,0%	21,3%	26,6%	80,9%
57,5%	87,8%	92,9%	49,1%	60,3%	29,0%	78,7%	73,4%	19,1%

Frekvencia zriedkavého a častého kontrafaktového myslenia v tomto súbore sa takmer nelíši od našich predchádzajúcich výskumov, uskutočnených so študentkami a iným súborom zdravotných sestier. U detských lekáriek sa zistil nižší výskyt frekventovaného kontrafaktového myslenia (36%), avšak išlo len o malý súbor (n=25), ktorý sa postupne snažíme rozšíriť. Výskumy sa týkajú súvislostí kontrafaktového myslenia s osobnostnými charakteristikami, napr. s úzkosťou, Big Five charakteristikami, koherenciou osobnosti a ďalšími, ktorých priemernými hodnotami v súboroch by na základe už zistených, ako aj ďalších v literatúre dokumentovaných poznatkov, bolo možné frekvenciu kontrafaktového myslenia zdôvodniť (čo však v tejto štúdii bližšie nerozoberáme).

Odpovede dokumentujú silné emočné pôsobenie neúspechu a prevahu kontrafaktového uvažovania o lepších alternatívach, ktoré sú v súlade s poznatkami mnohých autorov, ako aj pomerne vysoké percento pomoci a nízke percento brzdenia tohto typu uvažovania pri riešení budúcich problémov , čo je tiež v súlade s výsledkami našich predchádzajúcich výskumov. Povšimnutia hodné je vysoké percento emočného pôsobenia úspechu – dá sa podľa nášho názoru spojiť s vysokou pracovnou i študijnou motiváciou v tomto súbore.

Grafické zobrazenie profilu odpovedí je v nasledujúcom grafe.



Vzhľadom na psychologickú podstatu skúmaného problému budeme na konfiguračno frekvenčnú analýzu používať aj premennú socex. Uvádzame jej frekvenčnú tabuľku:

socex		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	72	15,3	49,3	49,3
	2	74	15,7	50,7	100,0
	Total	146	31,1	100,0	
Missing	System	324	68,9		
Total		470	100,0		

Vidno, že báza na ďalší výskum sa redukuje maximálne na 146 prípadov vzhľadom na vybraté extrémne skupiny podľa SOC. Profil odpovedí na skúmané otázky na báze súboru 146 respondentov sa mierne zmení – viď tabuľka:

o1 Zaoberáte sa takýmito myšlienkami?	o2 Ak ste riešili nejaký problém a nedopadlo to dobre, pôsobí to na vás emočne?	o3 Pôsobí na vás emočne aj úspech pri riešení problému?	o4 Uvažujete o tom, ako to mohlo dopadnúť, keby ... - rozosmutňuje vás to?	o5 Uvažujete o alternatívach lepších ako tá, ktorá nastala v skutočnosti?	o6 Uvažujete o alternatívach horších ako tá, ktorá nastala v skutočnosti?	o7 Uvažujete častejšie o alternatívach -horších -lepších?	o8 Pomáha vám tento typ uvažovania pri budúcich problémoch?	o9 Brzdí vás tento typ uvažovania pri riešení problémov?
41,80%	17,10%	8,20%	51,70%	40,30%	69,20%	26,60%	30,10%	79,20%
58,20%	82,90%	91,80%	48,30%	59,70%	30,80%	73,40%	69,90%	20,80%

3. Korelácia javov

V práci Luha J. (2009): Korelácia javov. FORUM STATISTICUM SLOVACUM 2/2009. SŠDS Bratislava 2009. ISSN 1336-7420 sú skúmané niektoré špecifické modely korelácie javov. Keď skúmame premenné s dvojprvkovou množinou hodnôt môžeme na skúmanie závislostí použiť Pearsonov korelačný koeficent, čo využijeme aj v tomto príspevku.

Keďže skúmame aj premennú socex, redukuje sa počet respondentov na 146 prípadov, preto pri skúmaní súboru premenných s uvedenou premennou pracujeme s redukovanou množinou respondentov.

Korelačná matica skúmaných 9 rekódovaných otázok a premennej socex je v tabuľke:

Pearson Correlations	**o1**	**o2**	**o3**	**o4**	**o5**	**o6**	**o7**	**o8**	**o9**	**socex**
o1 Zaoberáte sa takýmito myšlienkami?	1	0,17	0,05	0,25	0,37	0,23	0,07	0,08	0,09	-0,14
o2 Ak ste riešili nejaký problém a nedopadlo to dobre, pôsobí to na vás emočne?	0,17	1	0,39	0,37	0,15	0,15	-0,02	0,06	0,18	-0,27
o3 Pôsobí na vás emočne aj úspech pri riešení problému?	0,05	0,39	1	0,14	-0,04	-0,02	0,16	0,08	0,15	-0,10
o4 Uvažujete o tom, ako to mohlo dopadnúť, keby ... - rozosmutňuje vás to?	0,25	0,37	0,14	1	0,30	0,25	-0,02	-0,05	0,22	-0,37
o5 Uvažujete o alternatívach lepších ako tá, ktorá nastala v skutočnosti?	0,37	0,15	-0,04	0,30	1	0,28	0,13	0,10	0,03	-0,08
o6 Uvažujete o alternatívach horších ako tá, ktorá nastala v skutočnosti?	0,23	0,15	-0,02	0,25	0,28	1	-0,28	0,05	0,22	-0,08
o7 Uvažujete častejšie o alternatívach ...	0,07	-0,02	0,16	-0,02	0,13	-0,28	1	0,05	-0,14	0,10
o8 Pomáha vám tento typ uvažovania pri budúcich problémoch?	0,08	0,06	0,08	-0,05	0,10	0,05	0,05	1	-0,45	0,31
o9 Brzdí vás tento typ uvažovania pri riešení problémov?	0,09	0,18	0,15	0,22	0,03	0,22	-0,14	-0,45	1	-0,35
socex	-0,14	-0,27	-0,10	-0,37	-0,08	-0,08	0,10	0,31	-0,3	1

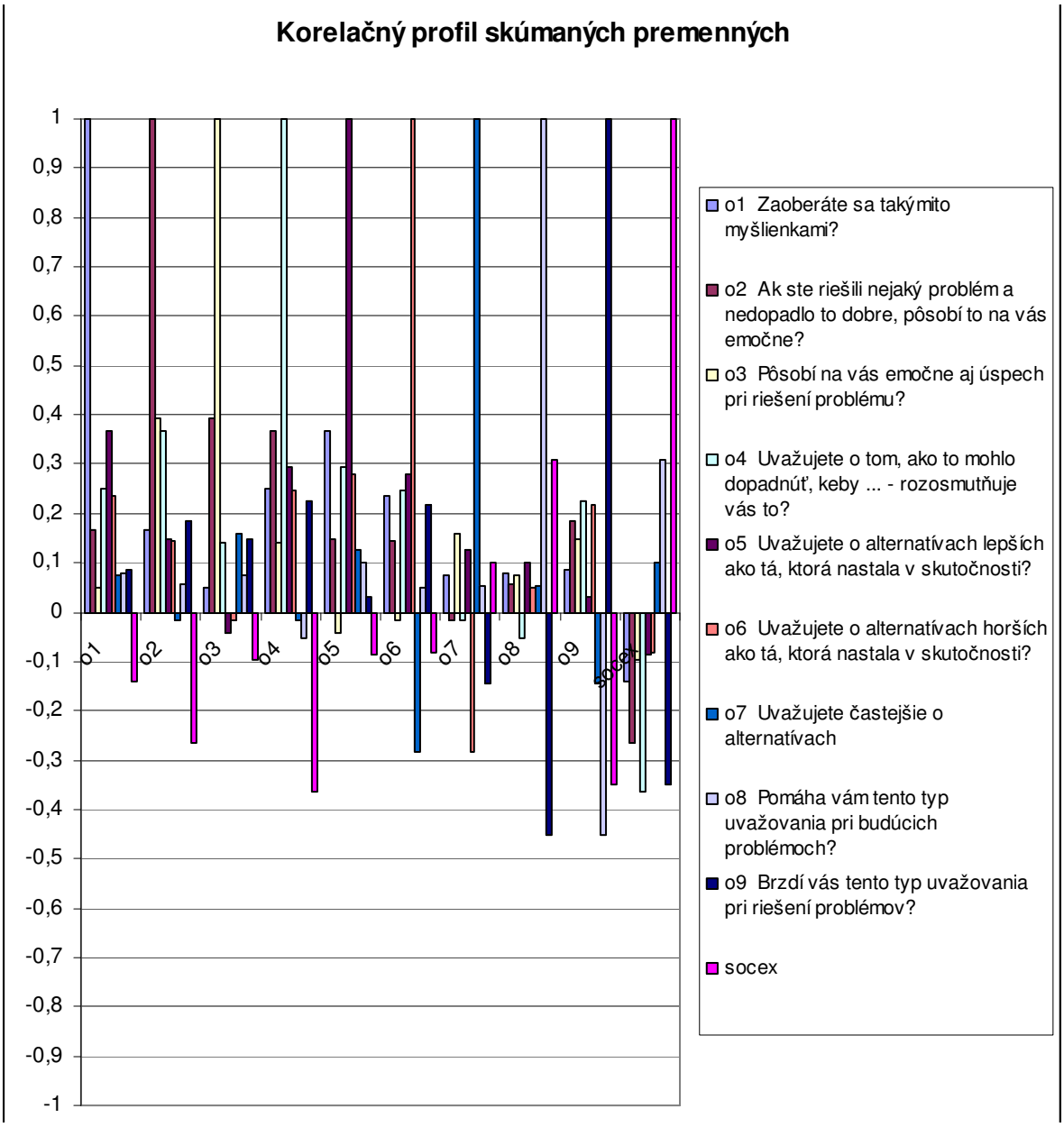
Skúmaním korelačnej matice môžeme zistiť najdôležitejšie párové vzťahy a na ich základe vybrať do CFA menší počet znakov.

Korelačnú tabuľku môžeme graficky prezentovať pomocou korelačného profilu, ktorý prehľadne identifikuje napríklad záporné korelácie a najväčšie korelácie. V tabuľke a aj v grafe sme nechali aj koreláciu znaku „so sebou“, čo je zrejme 1.

Na základe analýzy korelačnej matice a logických vzťahov otázok a odpovedí možno vybrať niekoľko skupín premenných na ďalšiu analýzu pomocou CFA. Napríklad:

1. skupina: o1, o4, o5, o6; (počet konfigurácií 16);
2. skupina: o2, o3, o4, socex; (počet konfigurácií 16);
3. skupina: o1, o2, o4, o5, o6; (počet konfigurácií 32);
4. skupina: o1, o4, o8, o9, socex; (počet konfigurácií 32);
5. skupina: o2, o4, o8, o9, socex; (počet konfigurácií 32);
6. skupina: o8, o9; (počet konfigurácií 4).

Po zvážení rôznych teoretických psychologických predpokladov vyberáme ako relevantnú pre ilustráciu na ďalšiu analýzu skupinu č. 4.



4. CFA pre znaky s dvojprvkovými množinami hodnôt

Konfiguračno frekvenčná analýza skúma viacrozmerné kontingenčné tabuľky. Vytvára všetky možné kombinácie odpovedí skúmaných znakov, ktoré sa nazývajú ***konfigurácie***.

Uvažujme na ilustráciu tri otázky s dvomi možnými odpoveďami, napríklad vybrané tri otázky skúmaného dotazníka. Trojrozmernú kontingenčnú tabuľku zapíšeme v tvare konfigurácií a ich početností. Keďže uvažujeme 3 otázky, pričom každá má dve možnosti odpovede je počet konfigurácii rovný súčinu 2*2*2=8. Z nášho výskumu si na ilustráciu zápisu konfigurácií vyberáme o1, o8 a o9. Príklad je za celý skúmaný súbor. V zápise konfigurácií kontingenčnej tabuľky je 459 prípadov, pretože niektoré konfigurácie mali chýbajúce hodnoty.

Tabuľka konfigurácií

o1	o8	o9	n
1	1	1	45
1	1	2	23
1	2	1	119
1	2	2	8
2	1	1	25
2	1	2	27
2	2	1	182
2	2	2	30

Na analýzu mnohorozmernej kontingenčnej tabuľky je dôležité navrhnúť vhodný model teoretického rozdelenia početností. V tomto príspevku sa sústredíme na objasnenie modelov 0-tého a 1-vého rádu. Modely vyššieho rádu súvisia napríklad s log-lineárnymi modelmi.

Model 0-tého rádu vychádza z predpokladu rovnomerného rozdelenia pravdepodobnosti jednotlivých konfigurácií.

Model 1-vého rádu využíva predpoklad nezávislosti uvažovaných premenných.

Konfiguračno frekvenčná analýza hľadá typy a antitypy testovaním hypotéz na základe diferencie empirických (n_i) a teoretických (e_i) početností. Môžeme pritom vychádzať z Chí-kvadrát modelu $(n_i - e_i)*2/(n_i-e_i)$ (táto veličina má asymptoticky Chí-kvadrát rozdelenie s 1 stupňom voľnosti), alebo z-transformácie, čo je odmocnina Chí-kvadrátu – (táto veličina má asymptoticky normované normálne rozdelenie N(0,1)), prípadne exaktných modelov, ako napr. binomický model.

CFA identifikuje TYPy a ANTITYPy nasledovne:

- konfigurácia i je TYP, ak zvolený test ukáže signifikantný vzťah: $n_i-e_i > 0$;
- konfigurácia i je ANTITYP, ak je signifikantný vzťah $n_i-e_i < 0$.

Na stanovenie simultánnej úrovne významnosti využíva CFA Bonferroniho princíp (von Eye A. (1990), von Eye A. (2002)).

CFA poskytuje jednoduchý nástroj na analýzu mnohorozmerných kontingenčných tabuliek, skúma však aj zložitejšie modely, ktoré sú blízke log-lineárnym modelom.

Ak by sme chceli zostaviť konfigurácie pre 9 otázok s dvomi možnými odpoveďami, tak máme 2*2*2*2*2*2*2*2*2=512 všetkých konfigurácií, čo je v našom prípade viac ako počet skúmaných osôb (470) a tak je nutné hľadať najdôležitejšie vzťahy na zníženie dimenzie problému.

Prehľad o počte konfigurácií pre skúmané rekódované otázky dotazníka je v tabuľke:

počet otázok	1	2	3	4	5	6	7	8	9
počet odpovedí	2	2	2	2	2	2	2	2	2
počet konfigurácií	**2**	**4**	**8**	**16**	**32**	**64**	**128**	**256**	**512**

V predošlej kapitole sme definovali 6 skupín otázok na skúmanie pomocou konfiguračno frekvenčnej analýzy. Vzhľadom na limitovaný rozsah príspevku uvedieme výsledky CFA pre skupinu č.4. Využili sme program autora A. von Eye, ktorý je dostupný na internete.

Configural Frequency Analysis

author of program: Alexander von Eye, 2000

Marginal Frequencies

Variable	Frequencies	
1	59.	84.
2	74.	69.
3	43.	100.
4	113.	30.
5	71.	72.

sample size N = 143

Pearsons chi2 test was used
Bonferroni-adjusted alpha = .0015625
a CFA of order 1 was performed

Table of results

Configuration	fo	fe	statistic	p	
11111	5.	3.602	.543	.46136551	
11112	4.	3.653	.033	.85582209	
11121	1.	.956	.002	.96434458	
11122	3.	.970	4.250	.03923983	
11211	4.	8.377	2.287	.13047784	
11212	20.	8.495	15.583	.00007898	Type
11221	2.	2.224	.023	.88064404	
11222	0.	2.255	2.255	.13316262	
12111	4.	3.359	.122	.72636333	
12112	1.	3.406	1.700	.19234819	
12121	3.	.892	4.985	.02556692	
12122	0.	.904	.904	.34164993	
12211	5.	7.811	1.011	.31454914	
12212	6.	7.921	.466	.49493203	
12221	1.	2.074	.556	.45591879	
12222	0.	2.103	2.103	.14702380	
21111	2.	5.128	1.908	.16715536	
21112	3.	5.201	.931	.33457418	
21121	3.	1.361	1.972	.16024753	
21122	0.	1.381	1.381	.23998756	
21211	7.	11.926	2.035	.15373149	
21212	20.	12.094	5.168	.02300836	
21221	0.	3.166	3.166	.07517442	
21222	0.	3.211	3.211	.07315126	
22111	3.	4.782	.664	.47823	
22112	0.	4.849	4.849	.02766011	
22121	10.	1.269	60.041	.00000000	Type
22122	1.	1.287	.064	.80005142	
22211	16.	11.120	2.141	.14339555	
22212	13.	11.277	.263	.60790181	
22221	5.	2.952	1.420	.23336534	
22222	1.	2.994	1.328	.24917623	

chi2 for CFA model = 127.3649
df = 26 p = .00000000
LR-chi2 for CFA model = 102.4835
df = 26 p = .00000000

Ukazuje sa, že prvý pozorovaný typ (11212) zdôrazňuje kognitívny aspekt kontrafaktového myslenia. Ak by sme toto myslenie chápali ako stratégiu zvládania, potom ide viac o problémovo orientované zvládanie (ako ho definuje Lazarus). To je v súlade aj s Antonovského (autor metódy SOC) poznatkami, že osoby s vysokou úrovňou SOC majú aktívny prístup k riešeniu každodenných aj špecifických problémov a v ich myslení dominuje kognitívne spracovanie všetkých aspektov záťažovej situácie. Druhý pozorovaný typ (22121) reprezentuje viac emočný aspekt kontrafaktového myslenia KM (frekventované KM, rozosmutňujúce KM, nepomáhajúce KM, brzdiace KM), ktoré sa spája s nízkou efektívnosťou zvládania záťažových situácií.

5. Záver

CFA možno využívať ako exploračnú aj konfirmačnú analýzu. V exploračnej fáze nám CFA umožňuje zmapovať možné súvislosti skúmaných premenných. Pri konfirmačnom použití overujeme vopred stanovené hypotézy (v našom prípade napríklad o typoch daných určitou konfiguráciou premenných o1, o4, o8, o9 a socex), vyplývajúce z predpokladov psychologickej teórie.

V rámci ukážky z nášho výskumu sa ukazuje CFA ako vhodná na zisťovanie relevantných vzťahov skúmaných premenných (týkajúcich sa tohto typu myslenia i osobnostných charakteristík).

Tu využitá CFA prvého rádu upozorňuje na relevantnosť spomínaných vzťahov a v ich podrobnejšej analýze možno pokračovať využitím CFA vyšších rádov v kombinácii s loglineárnymi modelmi.

Komplexnejšie využitie CFA dovoľuje i porovnávanie konfigurácií vo viacerých súboroch a uplatnenie interakčnej štruktúrnej analýzy (von Eye, 1990, Krauth , 1993) na podrobnejšiu analýzu vzájomných závislostí premenných, vytvárajúcich typy.

6. Literatúra

[1]Antonovsky, A.(1987): Unraveling the mystery of health. Jossey Bass, San Francisco.

[2]Epstude K., Roese N.J. (2008): The functional theory of counterfactual thinking. Personal and Social Psychology Review, 12, 168-192.

[3]Krauth, J. (1993): Einführung in die Konfigurations-frequenzanalyse (KFA). Beltz – Psychologie Verlags Union, Weinheim, Basel.

[4]Lienert, G.A. (1971): Die Konfigurationsfrequenzanalyse I. Ein neuer Weg zu Typen und Syndromen. Zeitschrift für klinische Psychlogie und Psychotherapie, 19,207-220.

[5]Luha J. a kol. (1985): Metódy štatistickej analýzy kvalitatívnych znakov II. ÚVT VŠ, Bratislava.

[6]Luha J. (1985): Testovanie štatistických hypotéz pri analýze súborov charakterizovaných kvalitatívnymi znakmi. Vydal Odbor Výskumu programov ČST a divákov v SR. Bratislava 1985.

[7]Luha J. (2006): Štatistické metódy analýzy kvalitatívnych znakov. FORUM STATISTICUM SLOVACUM 2/2006. SŠDS Bratislava 2006. . ISSN 1336-7420.

[8]Luha J. (2009): Korelácia javov. FORUM STATISTICUM SLOVACUM 2/2009. SŠDS Bratislava 2009. ISSN 1336-7420.

[9]Roese N.J. (2005): If only. Broadway Books, New York.

[10] Ruiselová Z., Prokopčáková, A., Kresánek, J. (2009): Counterfactual thinking as a coping strategy – cognitive and emotional aspects. Studia psychologica, 51, 2-3, 237-249.

[11] Ruiselová Z., Prokopčáková A., Kresánek J. (2007): Counterfactual thinking in relation to the personality of women – doctors and nurses. Studia psychologica, 49, 4, 333-339.

[12] Řehák J., Řeháková B. (1986): Analýza kategorizovaných dat v sociologii. Academia, Praha.
[13] Segura S., Morris M.W. (2005): Scenario simulations in learning: Forms and functions at the organizational level. In: D.R.Mandel, D.J.Hilton, P.Catellani (Eds.): The psychology of counterfactual thinking (p.94-109). Routledge, New York.
[14] von Eye A. (1990): Introduction to configural frequency analysis. The search for types and antitypes in cross-classification. Cambridge University Press, Cambridge 1990.
[15] von Eye A. (2002): Configural frequency analysis. Methods, Models and Applications. Mahwah, N.J.: Lawrence Erlbaum Associates. Inc. 2002.

Adresa autorov:

Ján Luha, RNDr., CSc.
Ústav lekárskej biológie, genetiky a klinickej genetiky LF UK a FNsP
Sasinkova 4, Bratislava
jan.luha@fmed.uniba.sk

Zdenka Ruiselová, RNDr., CSc.
Ústav experimentálnej psychológie SAV
Dúbravská cesta 9
Bratislava
expszru@savba.sk

Uvedený výskum bol podporovaný grantovou agentúrou VEGA (grant No 2/7034/27) a Centrom excelentnosti SAV – CEVKOG.

Možnosti modelování úmrtnosti
Possibilities of mortality modeling

Petr Mazouch

Abstract: Modeling of mortality is very actual issue now, especially because of necessity of mortality forecasting into the future. There are many of methods, which are engaged in this theme, and based on different bases. This paper tries to build model based on mortality ratio between individual age groups and development of this indicator in the time period. For the comparison were chosen several countries, where the mortality is stabile and on good level in long time period.

Key words: Mortality, modeling, mortality rate.

Kľúčové slová: Úmrtnost, modelování, míra úmrtnosti

1. Úvod

Modelování úmrtnosti je v současné době jedna z hlavních činností demografů a pojistných matematiků. Celá řada institucí, které jsou závislé právě na vývoji tohoto ukazatele, investují nemalé prostředky do vývoje různých modelů, které by zajistily alespoň trochu přesný odhad budoucího vývoje. Mezi takové organizace nepatří pouze pojišťovny nebo penzijní fondy, které jsou na těchto odhadech existenčně závislé, ale také například stát, který vzhledem k nemalému podílu penzí na mandatorních výdajích musí nutně uvažovat o vývoji úmrtnosti v budoucnu.

Předmětem těchto modelů není samozřejmě vývoj celkové míry úmrtnosti, ale jejích jednotlivých specifických měr rozdělených zejména podle věku a také pohlaví. Hodnoty těchto specifických měr v čase klesají pro všechny věkové skupiny. U některých se můžeme setkat se situací, kdy již téměř dosahuje hranice, kterou nelze překonat, tedy limitně se blíží nule. Při pouhém modelování jednotlivých měr se můžeme vystavovat nebezpečí, kdy úmrtnost ve vyšším věku dosáhne hodnot nižších, než ve věku nižším. Tato situace (pro věky 30+) však nastat nemůže, resp. nesmí, více viz [1]. Proto je nutné zvolit jiný model, který bude zajišťovat vyloučení této situace.

Model, který by zajišťoval situaci popsanou výše, může být založen na predikci vývoje vztahů mezi měrami úmrtnosti v jednotlivých věkových skupinách, tedy ne na predikci jednotlivých měr. Tento může být konstruován jako absolutní rozdíl mezi měrami, pak by nesmělo dojít v predikci k situaci, kdy by byl rozdíl menší než nula, nebo založený na relativním vyjádření, pak by se muselo jednat o zamezení situace, kdy by byl poměr menší než jedna.

V předkládaném článku bude zvolena varianta s relativním vyjádřením a předmětem zkoumání bude vývoj tohoto ukazatele v čase a jeho možné využití pro další predikci vývoje úmrtnosti. Také byly zvoleny pouze věkové skupiny 30 – 65 let. Ve věku pod 30 let je možnost poklesu míry úmrtnosti pro vyšší věkovou skupinu proti skupině mladší a pro skupiny nad 65 by bylo nutné využít některou z metod vyrovnávání, což by mohlo ovlivnit jednotlivé vztahy mezi měrami úmrtnosti.

2. Metodika a data

Jednotlivé míry úmrtnosti jsou vypočteny jako podíl počtu zemřelých a středního stavu obyvatelstva příslušné věkové skupiny. Tyto míry úmrtnosti byly dále vyrovnány dle metodiky užívané ČSÚ následujícím způsobem:

$$m^{*}_{t,x} = \frac{1}{315} \cdot \left(105 \cdot m_{t,x} + 90 \cdot \left(m_{t,x+1} + m_{t,x-1}\right) + 45 \cdot \left(m_{t,x+2} + m_{t,x-2}\right) - 30 \cdot \left(m_{t,x+3} + m_{t,x-3}\right)\right) \quad , \quad (1)$$

Jak již bylo uvedeno v předchozí kapitole, v předkládaném článku bude tedy využito následujícího vztahu:

$$\frac{m^{*}_{t,x+1}}{m^{*}_{t,x}} \quad , \quad (2)$$

Takto byly spočteny poměry měr úmrtnosti pro věky 30 – 65 let v letech 1965 – 2006 za Českou republiku (CR), Rakousko (AUT), Franci (FRA), Nizozemsko (NLD), Velká Británie (UK), Švédsko (SWE) pro každé pohlaví zvlášť. Jako zdroj dat byla použita [2].

Protože nebylo možné využít průběhu jednotlivých časových řad pro jejich relativně vysokou variabilitu, byly spočteny základní ukazatele, které budou prezentovány v rámci následující kapitoly. Srovnáván bude zejména průměr za jednotlivé věkové skupiny (resp. průměr poměrů jednotlivých měr úmrtností v jednotlivých věkových skupinách za celé sledované období) a vyznačeny budou také nejméně a nejvíce variabilní hodnoty. Použití těchto jednoduchých charakteristik je možné zejména proto, že ve vývoji žádného poměru v čase není patrný žádný jiný, než konstantní trend.

3. Výsledky a diskuse

Z následujících výsledků lze pozorovat, že v prezentovaných zemích existuje relativně stabilní hodnota průměru napříč pozorovanými věky. Tyto hodnoty se však liší jak mezi jednotlivými zeměmi, tak zejména lze pozorovat rozdíly mezi pohlavími pro některé země.

Jednotlivé hodnoty charakterizují koeficient růstu míry úmrtnosti mezi po sobě jdoucími věky. Pro českou republiku je tato hodnota přibližně 1,1, což znamená, že s růstem věku o jednotku (o jeden rok) dochází k nárůstu míry úmrtnosti o jednu desetinu, proti věku o jednotku nižšímu. U mužů lze pozorovat, že hodnoty pro Rakousko, Francii a Švédsko lze považovat za přibližně shodné, jejich hodnoty se pohybují mezi 9 – 10 %. Velká Británie a Nizozemsko dosahuje hodnot více, než 11 % a Česká republika se nachází někde mezi těmito skupinami, tedy její hodnoty jsou mezi 10 – 11 %.

Hodnoty, které jsou výrazně nižší, patří nižším věkům, které jsou zatíženy „chybou“ náhodného výskytu úmrtí, tedy je zde ještě stále malá skupina osob zemřelých (a početná skupina osob žijících), což zapříčiňuje značnou variabilitu. Proto jsou téměř všechny hodnoty tučně označené (poměry s nejvyšší variabilitou) zvýrazněné do věku čtyřicet let. Pro predikci je však mnohem důležitější dobře odhadnout vývoj měr úmrtnosti ve věcích vyšších, tedy v těch, kde dochází k největšímu úbytku obyvatelstva, tedy věcích vyšších než 50 let. Pro tyto věky naopak zjišťujeme, že jsou hodnoty označeny kurzívou, což značí, že variabilita je zde nejnižší, což odpovídá tomu, že se jedná o robustní hodnoty neovlivněné některými náhodnými úmrtími v počtu jednotek osob.

Toto zjištění může výrazně podpořit další možnosti pro predikci, které z daných výsledků vyplývají. Je zřejmé, že po oddělení některých věků, které jsou obtížně predikovatelné (nízké věky) pro svou vysokou volatilitu, bude možné predikovat míry úmrtnosti pro skupiny, které by měly být hlavním předmětem našeho zkoumání.

Tabulka 5: Průměrná hodnota poměru míry úmrtnosti věku x ku míře úmrtnosti ve věku x+1 za období 1965 – 2006, muži

Muži						
Věkové skupiny	**AUT**	**CR**	**FRA**	**NLD**	**SWE**	**UK**
31/30	**1,047**	**1,056**	1,033	**1,041**	**1,031**	1,045
32/31	1,050	1,056	**1,043**	**1,045**	**1,042**	1,043
33/32	1,060	**1,046**	1,056	**1,056**	**1,061**	1,053
34/33	**1,054**	**1,067**	*1,068*	1,067	**1,074**	**1,060**
35/34	**1,059**	**1,078**	1,075	**1,072**	1,058	1,065
36/35	**1,073**	**1,104**	**1,077**	1,074	1,074	**1,072**
37/36	**1,087**	**1,103**	**1,076**	**1,087**	**1,062**	1,078
38/37	1,083	1,105	**1,086**	**1,089**	**1,078**	**1,088**
39/38	**1,099**	1,099	*1,087*	1,092	1,079	**1,099**
40/39	1,098	1,104	1,094	1,102	1,090	**1,105**
41/40	1,095	1,109	1,096	1,111	1,085	1,109
42/41	1,094	1,115	1,102	1,115	1,093	**1,102**
43/42	1,091	1,105	**1,099**	1,122	1,082	1,109
44/43	1,091	1,110	**1,099**	1,112	1,087	1,113
45/44	1,095	1,114	**1,092**	1,111	1,097	1,117
46/45	1,096	1,104	1,096	1,112	1,093	1,112
47/46	1,099	1,109	1,090	1,114	*1,103*	*1,115*
48/47	*1,096*	1,111	1,092	1,114	1,095	1,115
49/48	1,096	1,110	*1,090*	1,115	1,096	*1,114*
50/49	1,095	*1,104*	*1,087*	1,113	1,099	1,112
51/50	1,089	1,107	*1,081*	*1,117*	1,109	*1,115*
52/51	*1,102*	*1,099*	*1,083*	1,115	*1,104*	1,114
53/52	*1,104*	*1,103*	1,082	*1,117*	*1,105*	1,113
54/53	*1,097*	*1,101*	*1,083*	*1,109*	*1,101*	*1,112*
55/54	*1,099*	*1,099*	*1,083*	*1,113*	*1,104*	*1,112*
56/55	*1,097*	*1,098*	*1,081*	*1,117*	*1,098*	*1,111*

Pozn.: Tučně jsou vyznačeny hodnoty s nejvyšší variabilitou, kurzívou jsou vyznačeny hodnoty s nejnižší variabilitou.

Pro prognózu pojišťoven, penzijních fondů, ale i státu jsou totiž mnohem podstatnější hodnoty úmrtnosti u osob starších, protože to jsou osoby, které budou v budoucnu čerpat příspěvky ať již ze státních či soukromých fondů. Úmrtnost osob v nižších věkových skupinách je tak nízká, že počet přeživších se téměř nemění.

U žen je situace velmi podobná, jako u mužů. Pouze zde lze pozorovat utvoření dvou skupin. První je složená z Rakouska a Francie, druhá ze zbývajících států. Také nástup hodnot, které jsou „podobné“ těm mužským je přibližně o pět let posunutá do vyššího věku. Situace s variabilitou je stejná, jako u mužů. Nejvíce variabilní jsou zde nižší věky (opět vliv nízkých hodnot počtu zemřelých v těchto věkových kategoriích) a nejméně variabilní vysoké věky.

Tabulka 2: Průměrná hodnota poměru míry úmrtnosti věku x ku míře úmrtnosti ve věku x+1 za období 1965 – 2006, ženy

Ženy						
Věkové skupiny	**AUT**	**CR**	**FRA**	**NLD**	**SWE**	**UK**
31/30	**1,084**	**1,080**	**1,072**	**1,083**	**1,077**	**1,078**
32/31	**1,113**	**1,099**	**1,069**	**1,092**	**1,066**	**1,087**
33/32	**1,093**	**1,103**	**1,081**	**1,087**	**1,079**	**1,092**
34/33	**1,080**	**1,099**	**1,088**	**1,089**	**1,096**	**1,097**
35/34	**1,103**	**1,106**	**1,083**	**1,079**	**1,095**	**1,097**
36/35	1,085	**1,097**	1,090	**1,079**	**1,085**	1,099
37/36	**1,108**	**1,100**	**1,085**	**1,091**	1,091	**1,100**
38/37	**1,105**	1,097	1,082	1,109	1,087	1,103
39/38	1,087	1,107	1,091	1,104	**1,087**	1,102
40/39	1,088	1,121	**1,084**	1,115	**1,094**	1,107
41/40	1,100	1,129	1,092	1,110	1,101	1,113
42/41	1,100	1,112	1,088	1,107	1,112	1,113
43/42	1,112	1,109	1,089	1,104	1,108	1,112
44/43	1,107	1,112	1,087	1,112	1,102	1,108
45/44	1,098	1,106	*1,091*	*1,107*	1,096	1,107
46/45	1,095	1,115	1,084	1,109	1,107	1,108
47/46	1,087	1,108	1,088	1,107	*1,102*	*1,105*
48/47	1,093	1,098	1,085	1,088	1,108	*1,110*
49/48	1,090	*1,103*	1,078	1,081	1,098	*1,101*
50/49	1,086	1,097	1,078	*1,097*	*1,090*	*1,101*
51/50	1,087	1,097	1,076	*1,088*	1,086	1,098
52/51	*1,080*	1,100	1,073	1,090	1,084	*1,094*
53/52	*1,080*	*1,097*	*1,072*	*1,082*	*1,086*	*1,091*
54/53	*1,087*	*1,095*	*1,069*	*1,078*	*1,090*	*1,094*
55/54	*1,079*	*1,106*	*1,070*	1,084	*1,091*	1,095
56/55	*1,093*	*1,101*	*1,072*	*1,093*	*1,099*	*1,098*

Pozn.: Tučně jsou vyznačeny hodnoty s nejvyšší variabilitou, kurzívou jsou vyznačeny hodnoty s nejnižší variabilitou.

4. Záver

Příspěvek se pokusil navrhnout možné postupy vhodné pro využití při predikci měr úmrtnosti v budoucnosti. Jednoznačným přínosem je nalezení poměrně jasných nárůstů měr úmrtnosti podle věku o 9 – 11 %, v závislosti na zemi a pohlaví. Toto může vést k zjednodušení projekce, kde by bylo vhodné najít věk, jehož úmrtnost by byla nejvhodnější pro predikci a tento použít jako referenční úroveň, od které by se úrovně úmrtnosti ostatních věkových skupin daly odvodit. Pro vhodnost tohoto přístupu je možné pokusit se o odhad měr úmrtnosti v období, za které již míry úmrtnosti známe.

Z výše uvedeného lze tedy vyvodit, že výsledky prezentované v tomto článku jsou jen dílčími výsledky, které budou dále rozpracovávány s jasným cílem pokusit se o nalezení komplexního modelu pro odhad budoucího vývoje úmrtností založeným a využívajícím vztahy mezi jednotlivými měrami úmrtnosti mezi jednotlivými věkovými skupinami.

Literatúra

[5]MAZOUCH, P. 2009. Modelling of mortality in Czech Republic. In: SBORNÍK PŘÍSPĚVKŮ MEZINÁRODNÍ KONFERENCE AMSE 2009. 2009. Praha: Fakulta informatiky a statistiky

[1]THE HUMAN MORTALITY DATABASE. WEB: WWW.MORTALITY.ORG/

Adresa autora:

Petr Mazouch, Ing.
Vysoká škola ekonomická v Praze
Nám. W. Churchilla 4
130 67 Praha 3
mazouchp@vse.cz

Příspěvek vznikl za podpory a projektu Národního programu výzkumu II MŠMT ČR č. 2D06026 „Reprodukce lidského kapitálu“.

Conjoint analýza v zákaznícky orientovanom marketingu
Conjoint Analysis in Customer-oriented Marketing

Branislav Pacák

Abstract: Conjoint analysis refers to a group of multivariate statistical methods for analyzing dependencies that are used to determine the preferences that consumers attach to different properties of a product or service and their combinations. We assume that these properties are under the control of the producer (or seller). During the analysis based on consumer evaluations of possible combinations of the goods properties we can found importance of individual properties and assess their most-preferred combination. This article contains example of conjoint analysis using module Market Research Application of the statistical software package SAS 9.1.

Key words: conjoint analysis, preferences, factors, factor levels, stimulus, part-worths

Kľúčové slová: conjoint analýza, preferencie, faktory, úrovne faktora, profily, čiastkové užitočnosti

1. Úvod

V marketingovom prostredí posledných rokov dochádza k radikálnym zmenám v dôsledku viacerých významných celospoločenských zmien, ako sú prehlbujúce sa konkurenčné trhové prostredie, technologický pokrok, globalizácia a deregulácia. Z toho vyplývajú aj neustále sa vyvíjajúce požiadavky na marketingový výskum a nutnosť jeho inovácie. Zákazník sa stáva alfou a omegou celého marketingu. Tradičný marketing, ktorého cieľom bol predaj produktov namiesto snahy o pochopenie a uspokojenie skutočných potrieb zákazníka, je nutné inovovať na zákaznícky orientovaný marketing.

Viaceré viacrozmerné štatistické metódy našli v posledných desaťročiach napriek svojej náročnosti širokú aplikáciu v marketingových výskumoch v zahraničí, ako to vyplýva zo zahraničnej odbornej literatúry. Podľa počtu aplikácií v najrôznejších marketingových výskumoch patrí v zahraničí k najpoužívanejším štatistickým metódam conjoint analýza. Táto metóda je zo všetkých viacrozmerných metód najviac spojená práve s výskumom trhu a v posledných dvadsiatich rokoch zaznamenala značný metodologický aj aplikačný pokrok. V našich marketingových výskumoch sa táto metóda aplikuje len sporadicky.

2. Princíp conjoint analýzy

Názov conjoint analýza označuje skupinu viacrozmerných štatistických metód analýzy závislostí, ktoré sa používajú k určeniu preferencií, ktoré spotrebitelia prikladajú rôznym vlastnostiam určitého výrobku alebo služby a ich kombináciám. Pritom predpokladáme, že sa jedná o vlastnosti, ktoré sú pod kontrolou výrobcu (alebo predajcu). V priebehu analýzy sa na základe spotrebiteľského hodnotenia možných kombinácií vlastností tovaru zistí význam jednotlivých vlastností a odhadne sa ich najpreferovanejšia kombinácia. Na rozdiel od bežne používaných spôsobov spotrebiteľského hodnotenia vlastností produktov, keď respondent posudzuje každú vlastnosť izolovane, pri conjoint analýze potenciálni spotrebitelia hodnotia vopred dané kombinácie vlastností výrobkov.

Preferencie sa uvažujú ako vysvetľované premenné. Môžu byť vyjadrené napr. poradím (najmenšie číslo znamená najväčšiu preferenciu), alebo bodovým ohodnotením na nejakej škále (najmenšie číslo znamená najmenšiu preferenciu). Vysvetľujúcimi premennými sú vlastnosti (atribúty), charakterizujúce výrobky alebo služby. Vzhľadom na využitie analýzy

rozptylu sa nazývajú aj faktory. Jednotlivé kategórie, charakterizujúce faktory, sa nazývajú úrovne faktora alebo vlastnosti. Kombinácie rôznych úrovní všetkých faktorov sa nazývajú profily. Preferencie respondentov sa vzťahujú práve na posudzované profily.

Respondenti pri conjoint analýze zoraďujú jednotlivé profily, resp. varianty produktov podľa svojich preferencií, prípadne pravdepodobností nákupu. Z rôznych poradí variantov produktov sa odhaduje, akú dôležitosť respondent prikladá jednotlivým vlastnostiam produktu a počítajú sa ich čiastkové užitočnosti. Pomocou variačného rozpätia užitočností je možné určiť aj relatívny význam jednotlivých vlastností a na základe toho posúdiť pravdepodobnosť voľby určitej (aj neposudzovanej) kombinácie vlastností produktu.

Dôležitou vlastnosťou conjoint analýzy je teda schopnosť transformovať kvalitatívne údaje o preferenciách na údaje kvantitatívne a tým umožniť ich ďalšiu analýzu.Výhodou conjoint analýzy je to, že umožňuje zistiť význam aj takých vlastností produktov, ktoré je prakticky nemožné od seba oddeliť a skúmať izolovane. Experimentálne skúmanie približuje realite, v ktorej sa spotrebiteľ rozhoduje.

Vzhľadom k tomu, že nesprávny postup experimentu conjoint analýzy nie je možné dodatočne opraviť, je nutné vždy splniť kľúčové podmienky pre získanie objektívnej informácie o skutočných preferenciách zákazníkov vzhľadom k atribútom konkrétneho produktu.

3. Ukážka aplikácie v štatistickom programovom balíku SAS 9. 1

Obchodná spoločnosť zvažovala návrh nového priemyselného čistiaceho prostriedku pre možné použitie v mnohých odvetviach priemyslu. Pri vyvíjaní konceptu tohto výrobku chcel manažment lepšie poznať potreby a preferencie svojich priemyselných zákazníkov. Preto realizoval conjoint experiment na vzorke 86 respondentov - priemyselných spotrebiteľov. Výskum cieľovej skupiny potvrdil, že týchto päť atribútov predstavovalo hlavné určujúce prvky hodnotenia priemyselných čistiacich prostriedkov.

Názov atribútu	Úroveň		
Forma výrobku	Premix	Koncentrovaná tekutina	Prášok
Počet aplikácií na balenie	50	100	200
Pridanie dezinfekčnej zložky	Áno	Nie	
Šetrný k životnému prostrediu	Nie	Áno	
Cena typickej aplikácie	7 korún	10 korún	16 korún

Dátový súbor je prevzatý z prílohy k publikácii [1], dostupnej na internetovej stránke http://mvstats.com/Downloads/Fifth_edition/HATCO_data.ZIP.

Pretože sa kládol dôraz na dôkladné pochopenie štruktúry preferencií a predpokladal sa vysoký záujem respondentov pri hodnotení, bola vybratá tradičná conjoint metóda, ktorá sa javila ako vyhovujúca tak z hľadiska zaťaženia respondentov, ako aj z hľadiska hĺbky získaných informácií.

Aby sa zabezpečili reálne podmienky a bolo možné uprednostniť pre vyjadrenie preferencií bodovanie pred poradím, obchodná spoločnosť sa rozhodla použiť metódu plného profilu pre získanie neagregovaných aditívnych výsledkov s najjednoduchšou metódou odhadov. Výberom aditívneho pravidla vznikla možnosť použiť ortogonálne pole, aby sme nemuseli hodnotiť všetkých 108 možných kombinácií (3x3x2x2x3). V tomto prípade ortogonálne pole obsahuje skupinu 18 plných profilov. Keďže ako miera preferencie bolo použité metrické bodovanie, mohli sme na odhad modelu použiť regresnú analýzu, pretože ortogonálne pole poskytovalo dostatok profilov na odhad neagregovaných modelov.

Odhad čiastkových užitočností pre všetky faktory sa vypočítal najprv jednotlivo pre každého respondenta a výsledky boli potom agregované, aby sme získali kolektívny výsledok. Najskôr sa vypočítali odhady diskrétnych čiastkových užitočností pre každú úroveň. Potom sa skúmali jednotlivé odhady s cieľom odhaliť možné vzťahy medzi čiastkovými užitočnosťami jednotlivých faktorov (napr. či použiť lineárny alebo kvadratický vzťah). Tabuľka 1 ukazuje agregované výsledky pre celú skupinu a tabuľka 2 individuálne výsledky prvých dvoch respondentov.

Tabuľka 6: Odhady čiastkových užitočností pre celkovú vzorku

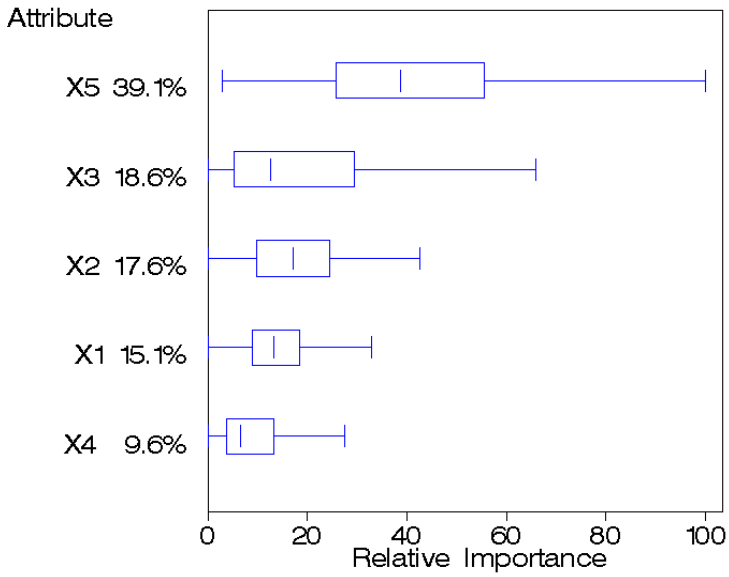
Preference	Attribute	Relative Importance	Attribute Value	Utility
Rank_AGREG	X1	8.1231	Koncentrát	0.16667
		_	Premix	-0.21705
		_	Prášok	0.05039
	X2	14.1128	50	-0.34496
		_	100	0.02326
		_	200	0.32171
	X3	21.6000	nie	-0.51017
		_	áno	0.51017
	X4	6.5231	nie	-0.15407
		_	áno	0.15407
	X5	49.6410	7	1.13178
		_	10	0.08140
		_	16	-1.21318

Tabuľka 2: Odhady čiastkových užitočností pre prvých dvoch respondentov

Preference	Attribute	Relative Importance	Attribute Value	Utility
Rank_1	X1	14.286	Koncentrát	0.61111
		_	Prášok	-0.55556
		_	Premix	-0.05556
	X2	20.408	50	0.44444
		_	100	0.61111
		_	200	-1.05556
	X3	5.102	nie	0.20833
		_	áno	-0.20833
	X4	13.265	nie	0.54167
		_	áno	-0.54167
	X5	46.939	7	1.44444
		_	10	0.94444
		_	16	-2.38889
Rank_2	X1	20.690	Koncentrát	-0.55556
		_	Prášok	0.11111
		_	Premix	0.44444
	X2	17.241	50	-0.05556
		_	100	-0.38889
		_	200	0.44444
	X3	6.897	nie	-0.16667
		_	áno	0.16667
	X4	24.138	nie	-0.58333
		_	áno	0.58333
	X5	31.034	7	0.61111
		_	10	-0.88889
		_	16	0.27778

Dôležitým krokom skúmania štruktúry preferencií čiastkových užitočností je výpočet dôležitosti atribútov. Tieto hodnoty odzrkadľujú relatívny dopad každého atribútu na výpočet celkovej užitočnosti (t.j. *skóre celkovej užitočnosti*). Počítajú sa z hodnôt čiastkových užitočností každého respondenta a poskytujú základ pre ďalší výstižný spôsob porovnávania štruktúr preferencií jednotlivých respondentov.

Systém SAS na základe odpovedí všetkých respondentov poskytuje škatuľkovitý graf (box plot) relatívnych dôležitostí každého faktora, pričom pre každý faktor (atribút) je uvedená jeho užitočnosť v percentách (podiel na užitočnosti všetkých atribútov).



Obrázok 3: Relatívne užitočnosti atribútov pre celú vzorku respondentov

4. Ukážka manažérskej aplikácie výsledkov conjoint analýzy

Simulátorom výberu sme v systéme SAS urobili odhady preferencií výrobkov pre každého respondenta. Odhady očakávaných podielov na trhu boli počítané dvoma modelmi: modelom *maximálnej užitočnosti* a *logitovým* modelom. Model maximálnej užitočnosti využíva informáciu o tom, koľkokrát mal každý z troch výrobkov najvyššiu užitočnosť pre všetkých respondentov. Druhý prístup odhadov podielov na trhu je v systéme SAS možný pomocou logit modelu. Tento model odhaduje relatívnu mieru preferencie každého produktu a odhaduje relatívnu početnosť kúpy produktu pre každého respondenta.

Tabuľka 3: Výsledky simulácie výberu pre tri špecifikované produkty

Simulation Model: Maximum utility

Status	Market Share	_Id_	X1	X2	X3	X4	X5
Active	71.9%	Vyrobok 3	Prasok	200	ano	ano	7
Active	21.3%	Vyrobok 2	Koncentrat	200	ano	ano	10
Active	6.9%	Vyrobok 1	Premix	50	ano	ano	16

Simulation Model: Logit

Status	Market Share	_Id_	X1	X2	X3	X4	X5
Active	63.1%	Vyrobok 3	Prasok	200	ano	ano	7
Active	29.1%	Vyrobok 2	Koncentrat	200	ano	ano	10
Active	7.9%	Vyrobok 1	Premix	50	ano	ano	16

V tabuľke 3 výsledkov simulácie obidvoma metódami na výstupe zo systému SAS vidíme, že výrobok 1 bol preferovaný (mal najvyššiu odhadnutú celkovú užitočnosť) iba u 6,9 % (resp. 7,9 %) všetkých respondentov. Výrobok 2 bol druhý v poradí, preferovalo ho 21,3 % (resp. 29,1 %) respondentov a najviac preferovaný bol výrobok 3, ktorý by uprednostnilo až 71,9 % (resp. 63,1 %) respondentov.

5. Záver

Prehlbujúce sa konkurenčné trhové prostredie vyžaduje nové prístupy k marketingu a vnímaniu zákazníkov. Základom takýchto analýz sú štatistické metódy. Využitie kvalitných softvérových produktov pri aplikácii štatistických metód pri výskume trhu umožnilo ich masovú aplikáciu v zahraničí a je jedinou možnou cestou ich širšieho využívania aj na Slovensku. Pochopiť a odvodiť detailne ich algoritmus je schopný len veľmi zdatný matematik, čo zrejme nie je možné vyžadovať od všetkých marketingových manažérov a analytikov. Pri informačnej explózii v súčasnej dobe ani nie je reálne detailné ovládanie každej metódy jej užívateľmi.

Ďalším limitujúcim činiteľom pri aplikácii všetkých viacrozmerných štatistických metód je ich výpočtová prácnosť, z ktorej vyplýva, že ich aplikácia už len z tohto dôvodu nie je možná bez využitia počítača a počítačového programu. V reálnych marketingových aplikáciách je nutné využitie procedúry pre conjoint analýzu v niektorom štatistickom programovom balíku. Článok obsahuje ukážkou využitia ponuky Conjoint Analysis modulu Market Research Application štatistického programového balíka SAS 9.1.

6. Literatúra

[1]HAIR, J.F. – BLACK, W. C. – BABIN, B. J. – ANDERSON, R. E. – TATHAM, R. L.: Multivariate Data Analysis. Sixth Edition. New Jersey: Pearson Education, Inc., 2007.

[2]HEBÁK, P. – HUSTOPECKÝ, J. – PECÁKOVÁ, I. – PRŮŠA, M. – ŘEZANKOVÁ, H. – SVOBODOVÁ, A. – VLACH, P.: Vícerozměrné statistické metody (3). Praha: Informatorium, 2005. ISBN 0-13-7246659-5.

[3]JOHNSON, R. A. – WICHERN, D. W.: Applied Multivariate Statistical Analysis. Sixth Edition. New Jersey: Pearson Prentice Hall, 2007.

[4]PACÁKOVÁ, V. – STRELCOVÁ, M. – SUŠIENKOVÁ, K.: Conjoint analýza v marketingovom výskume. Nová ekonomika, vedecký časopis OF EU v Bratislave, december 2004, ročník 3, číslo 4 (9), str. 66-73.

[5]ŘEZANKOVÁ, H.: Analýza dat z dotazníkových šetření. Praha: Professional Publishing, 2007. ISBN 978-80-86946-49-8.

[6]STANKOVIČOVÁ, I. – VOJTKOVÁ, M.: Viacrozmerné štatistické metódy s aplikáciami, Bratislava: IURA Edition, 2007. ISBN 978-80-8078-152-1.

[7]SAS Institute Inc.: Marketing Research: Practical Applications Using the SAS system Course Notes, Cary, NC, 1996.

[8]SAS Institute Inc.: Getting Started with the Market Research Application. Cary, NC, 1997.

[9]http://www.conjointpage.com/

Adresa autora:

Branislav Pacák, Ing. PhD.
Jantzen Development, s.r.o.
Hviezdoslavovo námestie 25
811 02 Bratislava,
bp@jantzendevelopment.com

Príspevok je spracovaný v rámci projektu VEGA č. 1/4586/07.

Využitie Paretovho rozdelenia v neproporcionálnom zaistení
Using of Pareto Distribution in Non-proportional Reinsurance

Viera Pacáková

Abstract: The worldwide property-liability insurance industry has been rocked by the increasing catastrophes in recent years and increased demand for catastrophe cover (e.g., per occurrence excess of loss reinsurance), leading to a capacity shortage in property catastrophe reinsurance. Catastrophe events in last years are associated with increases in premiums for some lines of business. Modelling of the tail of the loss distributions in non-life insurance is one of the problem areas, where obtaining a good fit to the extreme tails is of major importance. Thus is of particular relevance in non-proportional reinsurance if we are required to choose or price a high-excess layer. Pareto distribution plays a central role in this matter and an important role in quotation in non-proportional reinsurance.

Key words: non-proportional reinsurance, excess of loss reinsurance, priority, layer, loss premium, Pareto distribution, expected frequency, expected loss

Kľúčové slová: neproporcionálne zaistenie, zaistenie škodovej nadmiery, priorita, vrstva, zaistné, Paretovo rozdelenie, priemerná frekvencia, priemerná škoda

1. Úvod

Zaistenie je široko využívaný nástroj súčasného poisťovníctva. Jeho význam v dnešnej dobe stále viac narastá. Dôvodom sú klimatické, sociálne a technologické zmeny, ktoré majú pre súčasnú spoločnosť okrem pozitívnych aj množstvo negatívnych dôsledkov. Ide napríklad o nárast počtu, ale aj rozsahu prírodných katastrof, ako aj katastrof spôsobených človekom.

Zaistenie je prevod časti rizika, ktoré prevzal poisťovateľ od poistených formou priameho poistenia, na iného nositeľa rizika, označovaného ako zaisťovateľ. Predstavuje vzťah medzi poisťovateľom a zaisťovateľom, v ktorom zaisťovateľ nemá s poisteným žiadny zmluvný vzťah.

Typy zaistenia udávajú spôsoby podielu zaisťovateľa na krytí rizika. Rozlišujeme proporcionálny a neproporcionálny typ zaistenia. Vo všetkých typoch proporcionálneho zaistenia sa poistná suma, poistné a poistné plnenie delí medzi poisťovateľa a zaisťovateľa v rovnakom, zmluvne dohodnutom pomere, pri rešpektovaní limitu zaisťovateľa.

Pri neproporcionálneho zaistení účasť zaisťovateľa začína až od vopred dohodnutej úrovne skutočne vzniknutých škôd. Niekedy sa preto označuje ako škodové zaistenie. Základný podiel poisťovateľa na krytí škôd je vyjadrený v škodovom objeme. Poisťovateľ nesie škody do určitej výšky, v neproporcionálnom zaistení sa táto výška označuje ako *priorita*. Škody presahujúce prioritu hradí zaisťovateľ. Zaistné sa určuje obyčajne nezávisle od poistného, a to na základe pravdepodobnosti, že skutočná výška škôd presiahne vlastný vrub poisťovateľa (prioritu).

2. Neproporcionálne zaistenie WXL/R

Neproporcionálne zaistenie *škodovej nadmiery jednotlivých rizík* (WXL/R – working excess of loss cover per event) má za úlohu chrániť poisťovateľa pred dopadom jednotlivých veľkých škôd. V prípade, že z poistnej zmluvy zaisteného portfólia vyplývajú poistné nároky, ktoré prevyšujú prioritu poisťovateľa, potom vzniknutú nadmieru hradí zaisťovateľ, no len do

výšky jeho limitu *L*. Ak označíme prioritu poisťovateľa *a*, potom z poistného plnenia *X* v rámci určitej poistnej zmluvy poistné plnenie zaisťovateľa X_Z vyjadruje vzťah

$$X_Z = \begin{cases} 0 & if \quad X \le a \\ X - a & if \quad a < X \le L \\ L & if \quad X > L \end{cases}$$

3. Paretovo rozdelenie škôd prekračujúcich prioritu

Základom na určenie zaistného je stanovenie netto zaistného, označované ako tarifovanie. Používajú sa pritom rôzne postupy. Jedným z nich je postup, založený na Paretovom rozdelení, ktorý vychádza z minulej škodovej situácie v kombinácii s poistno-matematickým modelom pre stanovenie výšky zaistených škôd.

Paretovo rozdelenie škôd X_a, ktoré sú vyššie ako priorita *a*, je vyjadrené distribučnou funkciou

$$F_a(x) = 1 - \left(\frac{a}{x}\right)^b, \qquad x \ge a$$

a hustotou pravdepodobnosti v tvare

$$f_a(x) = \frac{b \cdot a^b}{x^{b+1}}, \qquad x \ge a$$

Parameter *b* je potrebné odhadnúť na základe známych výšok škôd počas určitého časového intervalu, obyčajne kalendárneho roku. Škôd, prekračujúcich dosť vysokú prioritu *a*, je obyčajne malý počet, preto zvolíme hodnotu *OP* (observation point), podstatne nižšiu ako je priorita *a*, tak, aby škôd, prekračujúcich *OP*, bolo dostatočne veľa. Tieto škody označíme ako

$$X_{OP,1}, X_{OP,2}, ..., X_{OP,n}$$

a ich rozdelenie je Paretovo rozdelenie s distribučnou funkciou

$$F_{OP}(x) = 1 - \left(\frac{OP}{x}\right)^b, \qquad x \ge OP$$

Maximálne vierohodný odhad parametra *b* je daný vzťahom

$$\hat{b} = \frac{n}{\sum_{i=1}^{n} \ln\left(\frac{X_{OP,i}}{OP}\right)}$$

4. Stanovenie netto zaistného pomocou Paretovho modelu

Postup, založený na Paretovom rozdelení, je jednou z metód stanovenia netto zaistného v neproporcionálnom WXL/R zaistení. Netto zaistné dostaneme ako súčin priemerného počtu škôd, prekračujúcich prioritu *a* a priemernej výšky škôd X_a, prekračujúcich prioritu a, teda poistných plnení zaisťovateľa.

Z minulých údajov vieme odhadnúť iba priemerný počet (aktualizovaných) škôd *LF*(*OP*) nad hodnotou *OP*, pre oveľa vyššiu prioritu *a* musíme použiť odhad

$$LF(a) = LF(OP) \cdot P(X_{OP} \rangle\, a) = LF(OP) \cdot (1 - F_{OP}(a)) = LF(OP) \cdot \left(\frac{OP}{a}\right)^b =$$

$$= \begin{cases} LF(OP) \cdot OP^b \cdot \dfrac{a^{1-b}}{1-b} \cdot (RL^{1-b} - 1) & ak \quad b \neq 1 \\ LF(OP) \cdot OP \cdot \ln RL & ak \quad b = 1 \end{cases}$$

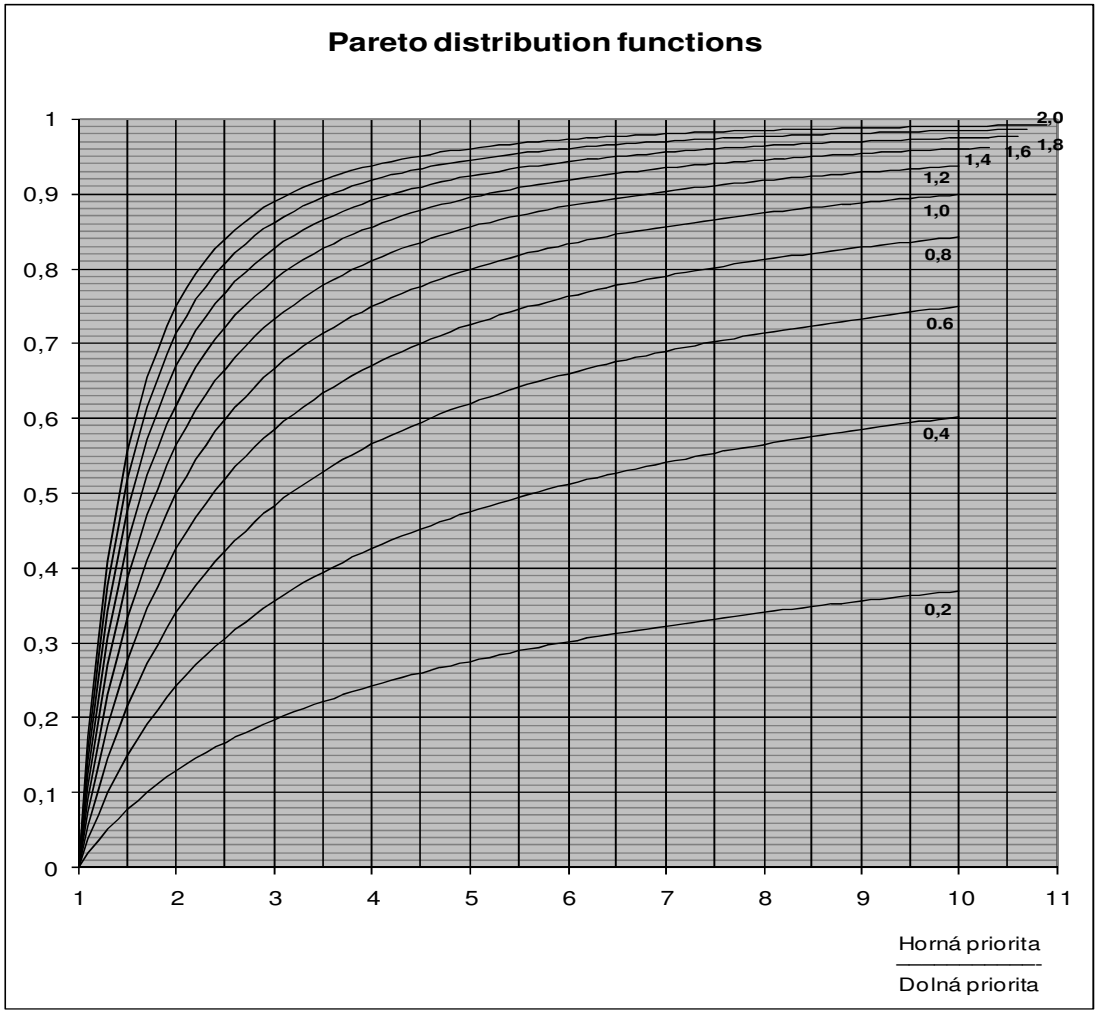
Strednú hodnotu poistných plnení zaisťovateľa, ktorý v danom roku platí škody nad prioritou a do výšky limitu L, vypočítame podľa vzťahu

$$EXL = \int_a^{a+L} (x-a) \cdot f_a(x)dx + \int_{a+L}^{+\infty} L \cdot f_a(x)dx =$$

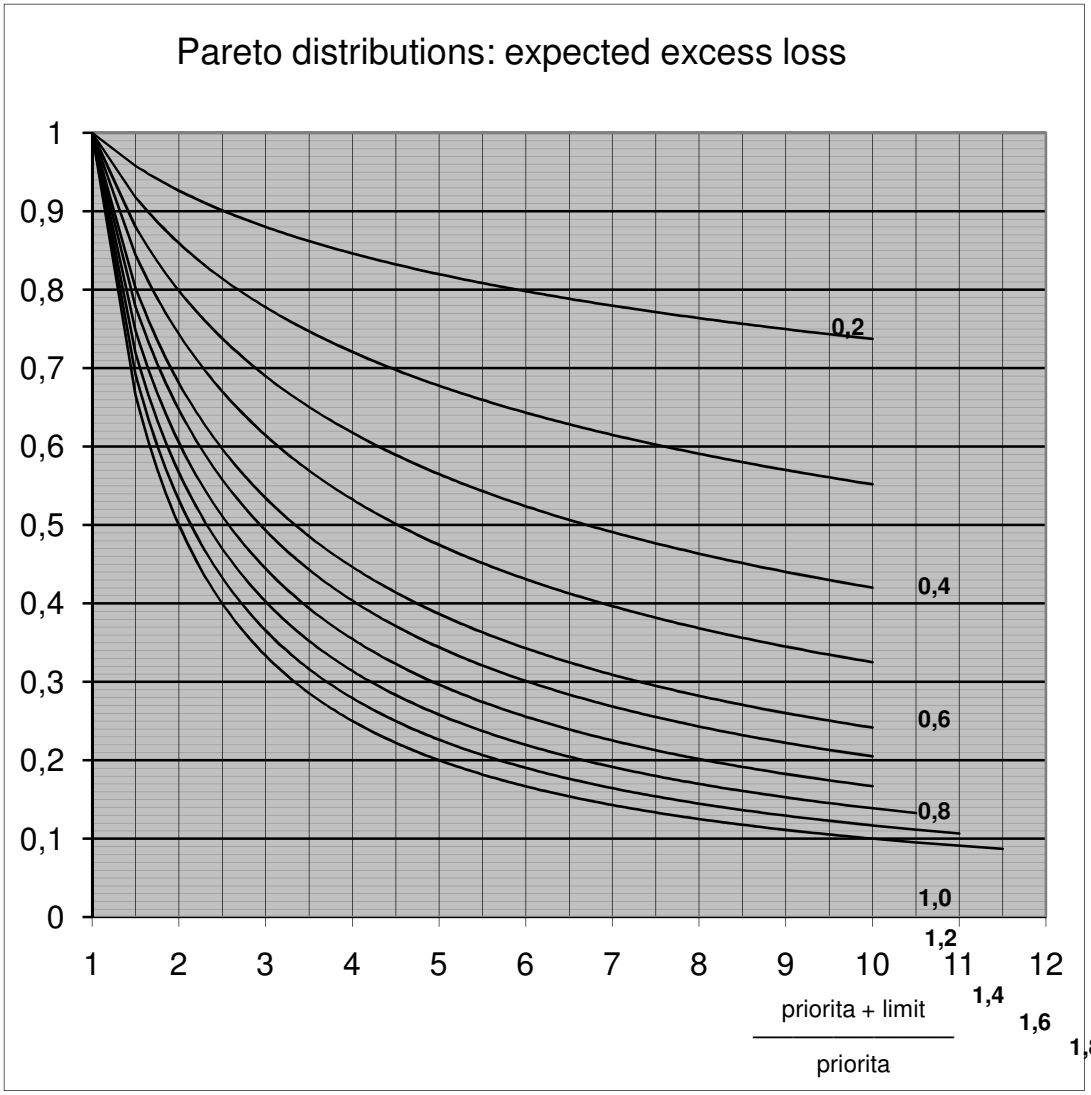
$$= \begin{cases} \dfrac{a}{1-b}(RL^{1-b} - 1) & if \quad b \neq 1 \\ a \cdot \ln RL & if \quad b = 1 \end{cases}$$

kde $RL = \dfrac{a+L}{a}$ je relatívna dĺžka vrstvy (relative layer). Potom pre výpočet netto zaistného dostávame vzťah

$$RP = LF(a) \cdot EXL = \begin{cases} LF(OP) \cdot OP^b \cdot \dfrac{a^{1-b}}{1-b} \cdot (RL^{1-b} - 1) & if \quad b \neq 1 \\ LF(OP) \cdot OP \cdot \ln RL & if \quad b = 1 \end{cases}$$



Obrázok 1: Distribučné funkcie Paretovho rozdelenia pre rôzne hodnoty parametra b



Obrázok 2: Očakávaná škodová nadmiera pre rôzne parametre b

5. Grafický postup stanovenia netto zaistného

V praxi môžeme pre stanovenie netto zaistného využiť krivky na obr. 1 a obr. 2. Stačí poznať parameter *b* Paretovho rozdelenia, prioritu *a* a limit zaisťovateľa *L*. Potom môžeme určiť priemernú škodovú frekvenciu a priemernú škodu nad prioritou *a* v nasledujúcich krokoch.

Parameter b Paretovho rozdelenia sme odhadli ako b = 1,6 a poznáme priemernú škodovú frekvenciu škôd, prekračujúcich hodnotu OP=100 000, ktorá je $LF(100\,000) = 4{,}5$. Našim cieľom je stanoviť netto zaistné, ak $a = 500\,000$ and $L = 500\,000$.

Krok 1: Určíme škodovú frekvenciu pre škody, presahujúce prioritu a= 500 000 ako $LF(500\,000) = P(X_a > 500\,000) \cdot 4{,}5 = (1 - 0{,}92) \cdot 4{,}5 = 0{,}36$, keď sme pomocou kriviek na obr. 1 pre pomer $\frac{a}{OP} = \frac{500\,000}{100\,000} = 5$ určili $P(X_a \le 500000) = 0{,}92.$

Krok 2: Určíme strednú hodnotu škôd, presahujúcich prioritu a. Pomocou kriviek na obr. 2 pre hodnotu relatívnej vrstvy

$$\frac{a+L}{a} = \frac{500\,000 + 500\,000}{500\,000} = 2$$

Nájdeme hodnotu 0,57. Potom $EXL = 0{,}57 \cdot L = 285\,000$.

Krok 3: Netto zaistné dostaneme ako súčin

$$NZ = LF(500\,000) \cdot EXL = 0.36 \cdot 285\,000 = 102\,600.$$

6. Záver

Príspevok je ukážkou využitia Paretovho rozdelenia pre odhad zaistného v prípade neproporcionálneho zaistenia škodovej nadmiery s vysokou prioritou *a*. V takomto prípade máme malý počet škôd, prekračujúcich túto prioritu. Výhodné vlastnosti Paretovho rozdelenia umožňujú odhadnúť priemerný počet aj priemernú výšku škôd prekračujúcich prioritu *a*, ak poznáme priemerný počet škôd presahujúcich nižšiu, vhodne stanovenú hodnotu *OP*. Netto poistné potom dostaneme ako súčin frekvencie a strednej hodnoty škôd nad prioritou *a*. Článok obsahuje aj ukážku grafického riešenia problému stanovenia netto zaistného pomocou kriviek distribučnej funkcie a škodovej nadmiery, ktoré sme pre rôzne hodnoty parametra b zostrojili v tabuľkovom procesore Excel.

7. Literatúra

[1]CIPRA T.: Zajištění a přenos rizik v pojišťovnictví. Praha: Grada Publishing, 2004. ISBN 80-247-0838-8.

[2]RYTGAARD M.: Estimation in the Pareto Distribution, ASTIN Bulletin Volume 20, No. 2.

[3]SCHMITTER H.: Estimating property excess of loss risk premiums by means of Pareto model, Swiss Re, Zürich 1997

[4]SCHMUTZ M. - DOERR R.: The Pareto model in property reinsurance, Swiss Re, Zürich 1998.

[5]SIPKOVÁ, Ľ.- SODOMOVÁ, E.: Modelovanie kvantilovými funkciami. Bratislava: Vydavateľstvo EKONÓM. ISBN 978-80-225-2346-2.

Príspevok je spracovaný v rámci projektov GA ČR č. 402/09/1866 *Modelování, simulace a řízení pojistných rizik* a VEGA č. 1/0724/08 *Riadenie rizík neživotného poistenia podľa direktívy Európskej komisie SOLVENCY II* .

Adresa autora:

Viera Pacáková, prof. RNDr. PhD.
Univerzita Pardubice
Fakulta ekonomicko-správní
Studentská 95
532 10 Pardubice
vpacakova@gmail.com

Mozaikové grafy v analýze kategoriálních dat
Mosaics Plots in Categorical Data Analysis

Iva Pecáková

Abstract: A mosaics plot is a type of statistical graph that helps to examine the relationship between two or more categorical variables. The paper sums up the potential of the mosaics plot and refers to software to use it.

Key Words: categorical data analysis, mosaics plots

Klíčová slova: analýza kategoriálních dat, mozaikové grafy

1. Úvod

Výsledkem každého praktického měření či zjišťování je vlastně diskrétní hodnota veličiny. Dovedeno do důsledků by se tak za kategoriální dala považovat jakákoliv data. Používáme-li však toto označení, máme na mysli situaci, kdy počet hodnot veličiny je ve srovnání s rozsahem statistického souboru relativně velmi malý. Binární proměnné mají jen dvě kategorie, polytomické více nespořádaných (nominální proměnné) nebo uspořádaných kategorií (ordinální proměnné). Kategorie přitom mohou být a často také jsou vyjadřovány slovně, aniž by měly jednoznačný kvantitativní význam. Obvyklé příklady takových proměnných jsou pohlaví (muž – žena), rodinný stav (svobodný, ženatý-vdaná, rozvedený, ovdovělý), politické preference (levice, střed, pravice), ale také věk (18-30, 31-45, 46-60, nad 60) nebo počet dětí (0, 1, 2, 3, více).

Při prezentaci výsledků jednorozměrného třídění kategoriálních dat se běžně používají grafické metody. Frekventované jsou zejména různé typy sloupkových či výsečových diagramů. Použití grafických nástrojů při roztřídění většího počtu kategoriálních veličin je zatím mnohem méně časté, i když myšlenka vyjádření četností pravoúhlými plochami s konkrétním významem jejich rozměrů (délky a výšky) je již dosti stará (E. Halley, J. Graunt). Její renesance od 80. let minulého století souvisí především s rozvojem log-lineárního modelování a výpočetní techniky a je přičítána J. Bertinovi [1] a především J. Hartiganovi a B. Kleinerovi [3]. V posledním období je asi nejvíce v této oblasti vidět práce M. Friendlyho [2].

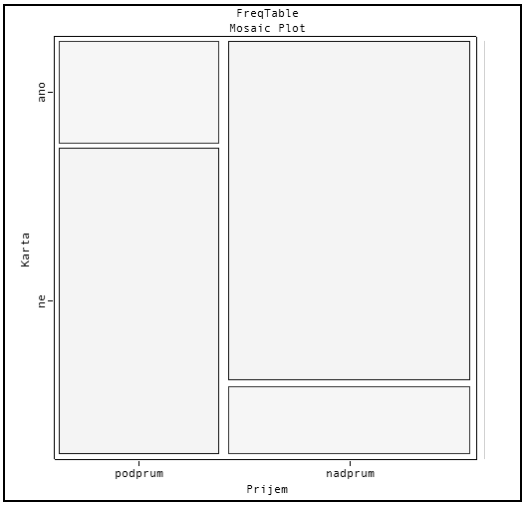
2. Mozaikové grafy

Mozaikový graf (mosaic display) je grafický nástroj pro vizualizaci a analýzu kategoriálních dat ve dvou i vícerozměrném třídění. Základní pravidla konstrukce lze (podle [2]) shrnout takto:

- Jednotková plocha čtverce či obdélníku je rozdělena na sloupky o šířce dané relativními marginálními četnostmi jedné veličiny, $p_{i+} = n_{i+}/n$.
- Tyto sloupky jsou rozděleny na základě podmíněných relativních četností pro druhou veličinu, $p_{j/i} = n_{ij}/n_{i+}$, na „dlaždice"; tyto podmíněné relativní četnosti určují tedy jejich výšku. Zda jednotlivé dlaždice bezprostředně sousedí či jsou odděleny úzkými mezerami, nehraje přitom žádnou roli (pro mozaiku bez mezer se vyskytuje také označení Mondrianovy[8] diagramy). Použití mezer však je výhodnější, jelikož jejich různá šířka umožňuje lepší vnímání různých úrovní třídění, je-li v grafu zachyceno více proměnných.

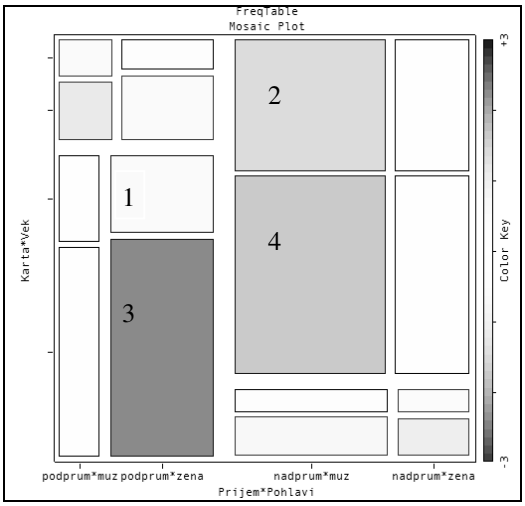
[8] P.C.Mondrian, holandský kubistický malíř začátku 20. století.

– Plocha „dlaždice“ reprezentuje výskyt příslušné kombinace hodnot veličin, podle nichž bylo provedeno dvojné třídění. Pokud odpovídá součinu marginálních četností $n_{i+}\, n_{+j}$, tedy v případě nezávislosti veličin, jsou „dlaždice“ v grafu zarovnány. Příklad mozaikového grafu v obrázku 1 tak naznačuje závislost mezi používáním platební karty (veličina *Karta* s hodnotami *ano*, *ne*) a velikostí příjmu (veličin *Příjem* s hodnotami *podprůměrný* a *nadprůměrný*).



Obrázek 1. Mozaikový graf – dvě veličiny

Popsaný proces konstrukce grafu může pokračovat pro další proměnnou – další proměnné – postupným členěním obrazce na menší a menší pravoúhlé díly. V takovém případě však je prostřednictvím jedné či obou os znázorňováno více veličin, které jsou tak obtížněji identifikovatelné. Obrázek 2 ilustruje podrobnější rozčlenění mozaikového grafu z obr. 1 dalším tříděním podle pohlaví a podle věku (*do 35 let* a *nad 35 let*).



Obrázek 2. Mozaikový graf – vícerozměrné třídění

Pokročilejší formy mozaikového grafu (Friendly [1]) využívají navíc barvy a stínování pro zobrazení velikosti a znamének standardizovaných reziduí modelu, s nímž jsou data konfrontována. Použití dvou základních barev umožňuje graficky rozlišovat kladná rezidua (obvykle modrá barva) a záporná rezidua (obvykle červená barva). Jejich velikost je pak vyjádřena světlejším (malá rezidua) či tmavším (větší rezidua) odstínem použité barvy.

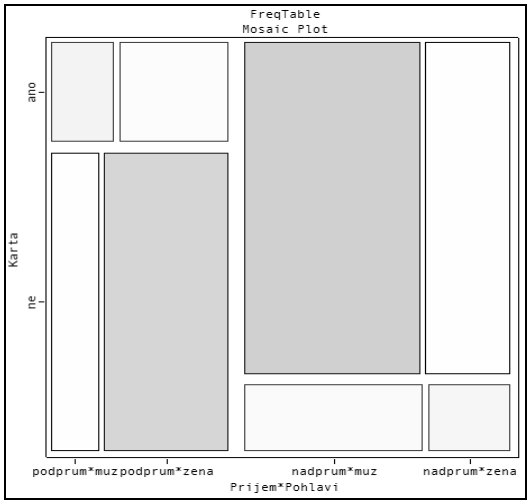
V obrázku 2 jsou očíslovaná pole modrá, ostatní pole, mají-li vůbec nějaký odstín, pak červený.

Mozaikové grafy pro dvojice veličin lze při větším počtu proměnných uspořádat (podobně jako např. bodové diagramy pro číselné proměnné) do matice mozaikových grafů – viz obrázek 3. Z jednotlivých grafů lze zde vyčíst závislost mezi používáním platebních karet a příjmem i pohlavím.

Obrázek 3. Matice mozaikových grafů

Podvojné závislosti zobrazené v matici mozaikových grafů však mohou být pouze zdánlivé. Zobrazení detailnějšího členění v mozaikovém grafu může napomoci takovou zdánlivou souvislost odhalit (podobně jako například loglineární model). Tak například v grafu znázorňujícím trojrozměrné třídění podle všech zúčastněných veličin zjišťujeme (viz obrázek 4), že v rámci obou příjmových skupin používání platební karty na pohlaví nezávisí (viz přibližné zarovnání „dlaždic“ v levé a pravé části grafu). Závislost používání karty na

pohlaví je pouze zdánlivá a je vyvolaná závislostí velikosti příjmu na pohlaví (patrnou z grafu v obrázku 3).



Obrázek 4. Mozaikový graf pro tři proměnné

Mozaikové grafy jsou implementovány do některých komerčních statistických systémů, obvykle však pouze pro dvourozměrné třídění. Pro vícerozměrné třídění lze použít některé volně distribuované programy. Pro ukázky v tomto textu byl použit systém ViSta (lze volně stáhnout na http://forrest.psych.unc.edu/research/index.html). Systém mj. umožňuje použití interaktivních mozaikových grafů pro zobrazení výsledků vícerozměrného třídění a konfrontace různých loglioneárních modelů. Na http://euclid.psych.yorku.ca/cgi/mosaics lze používat aplet s pěknými ukázkami mozaikových grafů a na http://www.math.yorku.ca/SCS/vcd pak jsou k dispozici makra pro mozaikové grafy v SASu.

3. Závěr

Mozaikové grafy, zejména interaktivní, pomohou nejen zobrazit strukturu kategoriálních dat, ale urychlí a zpříjemní také použití loglineárního modelování. Nalezení „dobré shody“ s nějakým modelem znamená dosažení malých reziduí, a tedy v mozaikovém grafu se projevuje jeho „čištěním“. Pokud je dále pořadí kategorií veličin v grafu přizpůsobováno tak, aby kategorie s podobnými rezidui sousedily, může mozaikový graf napomoci k porozumění vztahů mezi proměnnými, k rozhodování mezi různými typy loglineárních modelů a jejich interpretaci.

4. Literatura

[1]BERTIN,J: *Graphics and Graphic Information-processing*. de Gruyter, New York, 1981.

[2]FRIENDLY, M.: Visualizing Categorical Data, SAS Institute Inc., 2004

[3]HARTIGAN, J. A. – KLEINER, B.: Mosaics for contingency tables. In *Computer Science and Statistics: Proceedings of the 13th Symposium on the Interface*, Springer-Verlag, New York, 1981

[4]YOUNG, F.W.: ViSta, Visual Statistics System;
staženo z http://forrest.psych.unc.edu/research/index.html, 11/2009

Adresa autora:
Iva Pecáková, doc. Ing. CSc.
VŠE Praha
pecakova@vse.cz

Modely časových radov s premenlivými režimami
Regime-Switching models

Anna Petričková

Abstract: In this paper we will discuss mainly the class of regime-switching models with regimes determined by unobservable variables – Markov-switching models. These models have a nonlinear character. The goal of this work is to provide an overview of the problem and to describe this class of models. The focus of the work is applications of the obtained results. We estimate all the examined real time series with more types of switching models and compare them using certain goodness-of-fit criteria.

Key words: time series, MSW model, diagnostic test, estimation of the parameters

Abstrakt: V článku sa budeme zaoberať predovšetkým triedou viacrežimových modelov s režimami určenými nepozorovateľnými veličinami – Markovovými modelmi (MSW). Tieto modely majú nelineárny charakter. Cieľom práce je poskytnúť ucelenejší pohľad na problematiku a popísať túto triedu modelov. Ťažiskom práce je aplikácia dosiahnutých výsledkov na reálnych dátach z ekonómie a finančníctva. Pre každý zo skúmaných 20 reálnych časových radov odhadneme optimálne hodnoty parametrov pre viaceré typy modelov s premenlivými režimami, a na základe niekoľkých kritérií ich budeme navzájom porovnávať.

Kľúčové slová: časové rady, MSW model, diagnostický test, odhad parametrov

1. Introduction

One of the features of many real time series is their variable variance. In process of time series the phases with higher and lower dispersion alternate, these periods are short, sometimes longer. Practical analysis of financial time series shows that the autocorrelation in time series of yield in stock prices depends on their volatility [7]. Autocorrelation tends to rise during periods of lower volatility and tends to decline during periods of higher volatility. This character of the relationship is logical, since higher volatility indicates the presence of relatively stronger non-systematic component (and relatively weaker systematic component). The change of volatility, respectively autocorrelation of time series can be understand as the change of regime in the behaviour of time series. This change may be due to the several factors. Some of them are well identified, others hardly or even never.

In recent years there were proposed many time series models, which formalize the idea of the existence of different regimes of behavior in time series. Here belong, for example, models that can be used for modelling of time series financial yields, hydrological and geodetic time series, and so on ([1], [5], [9]). These models have nonlinear character. These include, for example, regime-switching models with regimes determined by observable variables and regime-switching models with regimes determined by unobservable variables.

This article will deal primarily with the class of regime-switching models with regimes determined by unobservable variables - MSW models. In the introduction, however, we make the overview from the simpler models to more complex ones.

2. Overview to time series models

Generally, we can divide stochastic models into 2 classes:

a) *linear stochastic models* – these include, for example ARMA model (see, e.g. [1], [3]):

The classical linear ARMA model for the predictable part X_t is based on the assumption that it is linear combination of p values X_{t-1}, ..., X_{t-p} and q values of i.i.d. process Z_t (consisting of independent equally distributed random variables with zero mean value and dispersion σ_z^2) shifted in time:

$$X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} - \ldots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2} + \ldots + \theta_q Z_{t-q} \quad (1)$$

where ϕ_1, ..., ϕ_p (autoregressive AR) and θ_1, ..., θ_q (moving averages MA) coefficients are unknown parameters. In special case if $p = 0$, we have MA process, if q = 0 - AR process. For details of linear ARMA models, see e. g. [3] or [4].

b) *nonlinear models* –

In the real life we often meet time series that exhibit strong non-linear features, because linear models are not in general always suitable for use. Therefore lately attention increases to nonlinear models, such as bilinear models, neural networks, regime-switching models, etc.. In this paper we focus on the class of regime-switching models, which are well to interpret and are also very suitable for modeling a lot of real data. The basic feature of these models is their „control" with one or more variables.

- regime-switching models with regimes determined by observable variables

Typical models belonging to this class are TAR models („**T**hreshold **A**uto**R**egressive"). They form the basis of regime-switching models with regimes determined by observable variables and were designed by Tong ([11], [12]). These models assume that any regime in time t can be given by any observed variable q_t (*indicator variable*). Values of q_t are compared with *threshold value c*. In case of 2-regimes model the first regime applies if $q_t \le c$, the second if $q_t > c$. A special case arises when the variable q_t is taken to be a lagged value of the time series itself, that is $q_t = X_{t-d}$ for a certain integer $d > 0$. The resulting model is called a **S**elf-**E**xciting **T**hreshold **A**uto**R**egressive (SETAR) model ([4]). For example 2-regime model SETAR with AR(p) in both regimes has form

$$X_t = \left(\phi_{0,1} + \phi_{1,1} X_{t-1} + \cdots + \phi_{p,1} X_{t-p}\right)\left[1 - I\left(X_{t-d} > c\right)\right] + \left(\phi_{0,2} + \phi_{1,2} X_{t-1} + \cdots + \phi_{p,2} X_{t-p}\right) I\left(X_{t-d} > c\right) + a_t$$

where $\{a_t\}$ is the strict white noise process with $E[a_t] = 0$, $D[a_t] = \sigma_a^2$ for all $t \in T$, and I[A] is the *indicator function* with values I[A] = 1 if the event A occurs and I [A] = 0 otherwise. If we replace the indicator function $I[X_{t-d} > c]$ by a continuous function $F(q_t, \delta, c)$ (*transition function*), which changes smoothly from 0 to 1 as X_{t-d} increases, the resultant model is called a **S**mooth **T**ransition **A**uto**R**egressive (STAR) model ([1], [4]). For example 2-regime model STAR with AR(p) in both regimes and indicator variable $q_t = X_{t-d}$ has form

$$X_t = \left(\phi_{0,1} + \phi_{1,1} X_{t-1} + \cdots + \phi_{p,1} X_{t-p}\right)\left[1 - F\left(X_{t-d}, \delta, c\right)\right] + \left(\phi_{0,2} + \phi_{1,2} X_{t-1} + \cdots + \phi_{p,2} X_{t-p}\right) F\left(X_{t-d}, \delta, c\right) + a_t$$

If

$$F\left(q_t; \delta, c\right) = \frac{1}{1 + \exp\left(-\delta\left(q_t - c\right)\right)}, \quad \delta > 0 \qquad \text{model LSTAR} \quad (2)$$

$$F\left(q_t; \delta, c\right) = \left(1 - \exp\left(-\delta\left(q_t - c\right)^2\right)\right), \quad \delta > 0 \qquad \text{model ESTAR} \quad (3)$$

- regime-switching models with regimes determined by unobservable variables

This class of models goes out from assumption that a regime is determined by any unobservable stochastic process, which can be labeled as $\{S_t\}$. It follows that separate regimes can not be identified exactly, but only with some probability. In case of 2 regimes process $\{S_t\}$ can take only values 1 and 2. If we work on a model AR(1), the regime-switching model with regimes determined by unobservable variables is

$$X_t = \phi_{0,S_t} + \phi_{1,S_t} X_{t-1} + a_t \qquad \text{for} \quad S_t = 1,2 \quad (4)$$

Markov switching models

This class of models is based on the assumption that a regime is determined by the discrete ergodic first order Markov process, hence important is only the actual and the previous state:

$$P(q_t = S_j \mid q_{t-1} = S_i, q_{t-2} = S_k, \ldots) = P(q_t = S_j \mid q_{t-1} = S_i) = p_{ij} \qquad 1 \le i, j \le N \qquad (5)$$

where $p_{ij} \ge 0$, $\sum_{j=1}^{N} p_{ij} = 1$ and p_{ij} is a state transition probability (transition from the state S_i at time *t-1* to the state S_j at time *t* in Markov chain), *t* = 1, 2, ..., n are time instants associated with state changes and q_t is actual state at time *t*. For details on MSW models see, e. g. [4], [5], [6].

Empirical specification procedure for nonlinear models – They are recommended the following steps (see, e.g. [2], [4])

1. specify an appropriate linear AR(p) model for the time series under investigation
2. test the null hypothesis of linearity against the alternative regime-switching nonlinearity
3. estimate the parameters in the selected nonlinear model
4. evaluate the model using diagnostic tests
5. modify the model if necessary
6. use the model for descriptive or forecasting purposes.

When we test the null hypothesis of linearity against the alternative of SETAR and MSW – type nonlinearity, we need to know estimates of parameters of the nonlinear model. Therefore, in this case we start a specification procedure for the model with estimation of parameters.

3. Application

We focused at the application of previously mentioned modeling procedures to real data, in our case 20 time series. Modelling of the time series can be (generally) used for any real data. In this paper we focused on macroeconomic indicators, exchange rates and other economic time series. We worked particularly with the slovak data, but also within V4 countries, we also used U.S. data and data of the European Union. Some data were seasonally adjusted, some were not, and we also used monthly as well as quarterly data. All data used in this paper can be downloaded from the pages of National Bank of Slovakia (www.nbs.sk) and European Central Bank (www.ecb.int), which takes some data directly from Eurostat (www.ec.europa.eu / Eurostat).

The first section summarizes the results of estimates of optimal parameter values of regime-switching models - concretely SETAR, LSTAR, ESTAR (regime-switching models with regimes determined by observable variables) and MSW (regime-switching models with regimes determined by unobservable variables). For each model, we consider 2, as well as 3 regimes. We tested linearity of the model to detect, if a linear or a nonlinear model is better and then, if hypothesis of linearity was rejected, we tested the remaining nonlinearity to detect, which of two model types is more suitable for our data: 2-regimes or 3-regimes time series model. We tested also the autocorrelation of residuals. We illustrate the graphs to see "how well" our models describe our real data.

In the second part we engaged to the prediction of values. The analysis of time series we do only on the part of data. The last 5 data we used to compare with the predictions given with models. This way we found out (with prediction errors RMSE and MAE) prediction properties of individual models.

For modelling of time series, we used the computing system *Mathematica*, version 6.
Note: Complete results of testing for each model we can find in [13], in this paper we summarized only selected models that best describe data.

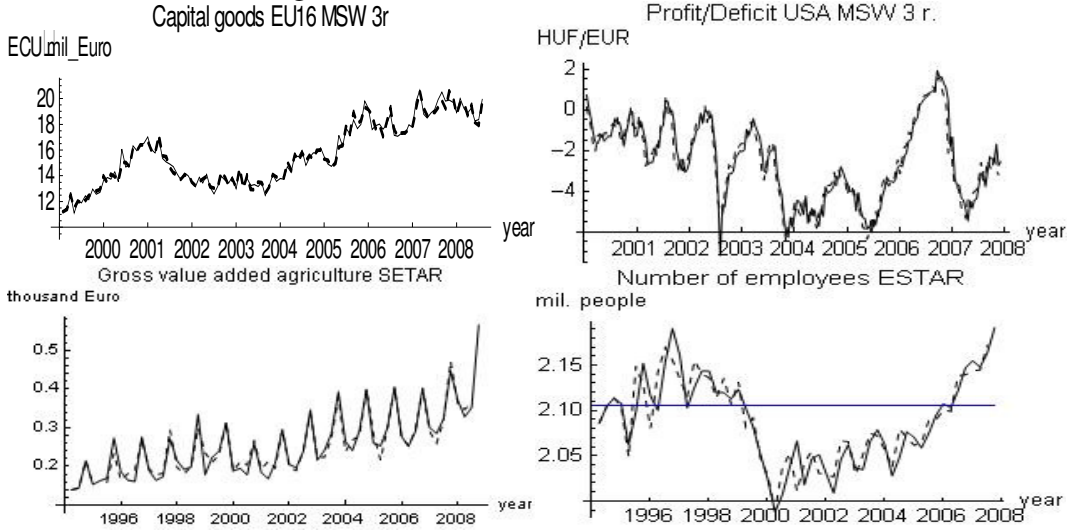
a. The best of the models

For each time series we choose the best model for descriptive purposes. The main criteria are standard deviation of residuals σ_{rez} and information criterion BIC. In the following table (Table 1) is statistics for each class of models.

Table 7: Selected the best models

model	number of selected models			Number of the best models from them
	2 regimes	3 regimes	linear	
SETAR	2	14	4	3
LSTAR	10	6	4	0
ESTAR	7	9	4	2
MSW	9	11	0	15

At the following picture there are selected graphs of models that best describe our time series (continuous line – original time series, dashed line – estimated model).



Picture 4: The best estimated model (4 selected graphs)

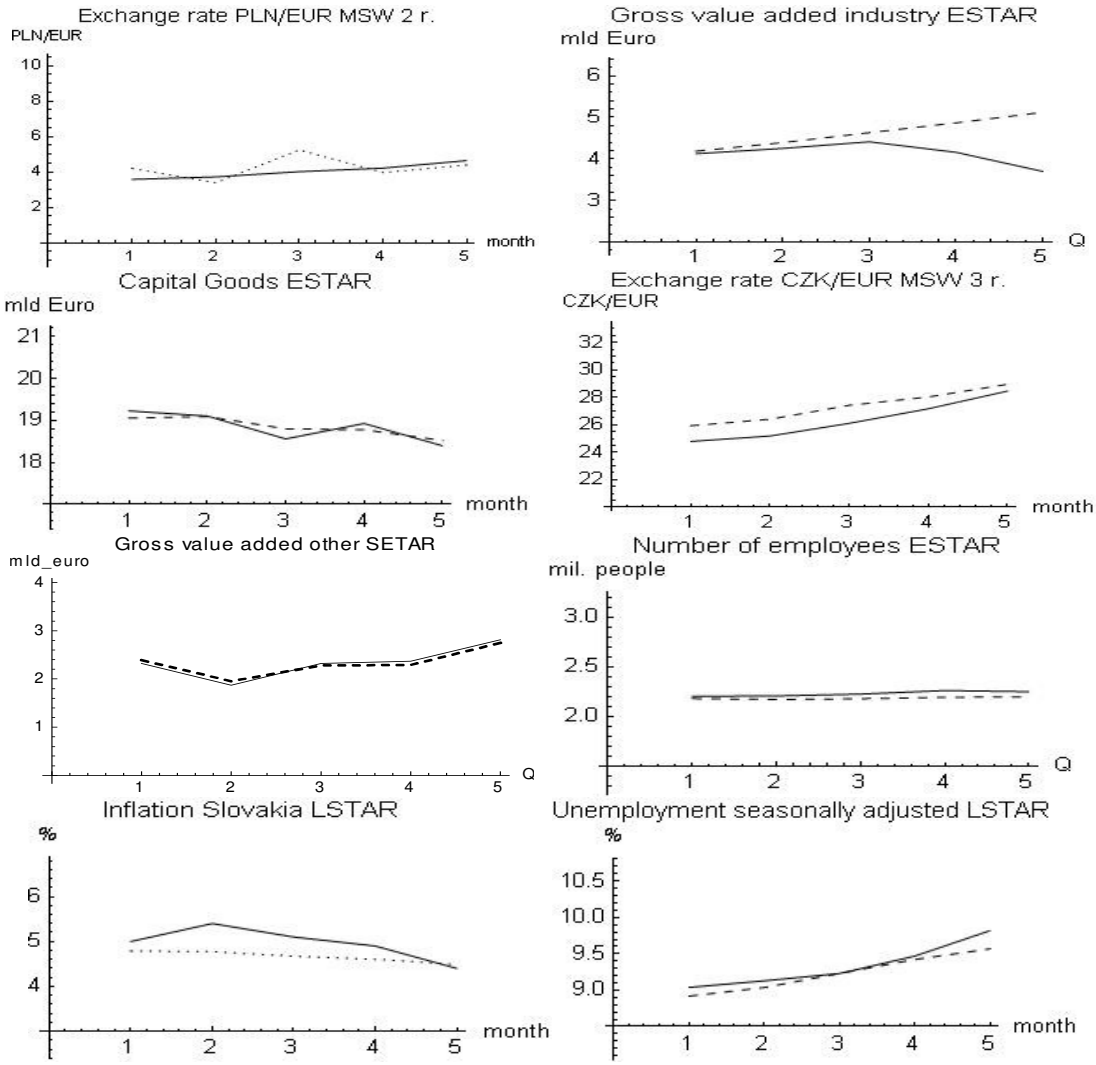
b. Predictions

In the previous section we have chosen the best models for description of real time series from SETAR, LSTAR, ESTAR and MSW class of models. For each time series we choose here the best models for prediction. The main criteria are prediction errors RMSE and MAE. In the following table (Table 2) the statistics for each class of models are presented.

Table 2: Selected the best model for forecasting purpose

Model class	Number of the best model for predictions
SETAR	6
LSTAR	2
ESTAR	8
MSW	4

At the following picture there are selected graphs of models with the best prediction of our time series (continuous line – original time series, dashed line – estimated model).



Picture 5: The best predicted models (8 selected graphs)

It's needed to note that we could get high-quality results only when external conditions didn't change, but these assumptions weren't fulfilled (due to the financial crisis).

4. Summary

In this work we discussed mainly the class of regime-switching models with regimes determined by unobservable variables – Markov-switching models, but in practical section we

compared estimations of all the examined real time series with models with regimes determined by observable variables (SETAR, LSTAR a ESTAR).

15 from 20 considered time series are the best described by models from the class MSW models (9 two-regimes, 6 three-regimes). Model from the class SETAR are the best for 3 and ESTAR for 2 time series.

The best predictive properties for considered time series have models from the class ESTAR (8 from 20 time series). The subsequent order (based on predictive properties) is: models from class SETAR (6), MSW (4) a LSTAR (2).

5. References

[1] AKINTUG, B. - RASMUSSEN, P. F. 2005. A Markov switching model for annual hydrologic time series. Water Resour, Res. 41, 2005.
[2] ARLT JOZEF - ARLTOVÁ MARKÉTA 2003. Finanční časové řady. Grada Publishing a.s., 2003.
[3] BOX, G. E. P. - JENKINS, G. M. 1970. Time Series Analysis: Forecasting and Control. Holden-Day, San Francisco, 1970.
[4] FRANSES, P. H. - D. VAN DIJK 2000. Non-linear time series models in empirical finance. Cambridge University Press, Cambridge, 2000.
[5] HAMILTON, J. D. 1989. A new approach to the economic analysis of nonstationary time series subject to change in regime. In: Econometrica, No. 57, 1989, 357 – 384.
[6] HAMILTON, J. D. 1994. Time Series Analysis. Princeton University Press. 1994.
[7] LEBARON, B. 1992. Some relationships between volatility and serial correlation in stock market returns. Journal of Business 65, 1992, 199 – 219.
[8] PSARADAKIS Z. - SOLA M. 1997. Finite-sample properties of the maximum likelihood estimator in autoregressive models with Markov switching. In: Journal of Econometrics, No. 86, 1997, 369 – 386.
[9] SEAN D. CAMPBELL 2002. Specification Testing and Semiparametric Estimation of Regime Switching Models: An Examination of the US Short Term Interest Rate. Department of Economics, Brown University, 2002.
[10] TIMMERMANNM, A. 2000. Moments of Markov Switching models. In: Journal of Econometrics, No. 96, 2000, 75 – 111.
[11] TONG, H. 1978. On a threshold model. In: C. H. Chen (ed.), Pattern recognition and Signal Processing, Amsterdam, 1978, 101 – 141.
[12] TONG, H. 1990. Non-linear time series: A dynamical systems approach. Oxford University Press, Oxford. 1990.
[13] PETRIČKOVÁ, A. 2009. Modely časových radov s premenlivými režimami. Diplomová práca, Univerzita Komenského Bratislava, 2009.

Address:

Anna Petričková, Mgr.
Department of Mathematics
Faculty of Civil Engineering
Slovak University of Technology Bratislava
Radlinského 11, 813 68 Bratislava
petrickova@math.sk

Acknowledgement: *This work was supported by Slovak Research and Development Agency under contract No. LPP-0111-09.*

Porovnání indexů kvality normálně rozložených skóre s různým rozptylem
Comparison of quality indexes of normally distributed scores with different variance

Martin Řezáč

Abstract: If one wants to assess quality of scoring models, he needs to use some of quantitative indexes like K-S, Gini and so forth. It is possible to use formulae for empirical computation. Nevertheless, formulae based on assumption of normality of scores can be used too. Assumption of equality of variances of scores is a crucial point now. This paper deals with comparison of appropriate results obtained on real data.

Key words: Scoring function, quality indexes, normal distribution, equality of variances.

Klíčová slova: Skóringová funkce, indexy kvality, normální rozdělení, shoda rozptylů.

1. Úvod

Při tvorbě skóringových modelů je zcela přirozená otázka, jak jsou tyto modely kvalitní. Tedy jak dobře daný model rozpozná subjekty, kteří přísluší do jedné či druhé skupiny. Máme-li např. posoudit jak bude klient finanční instituce splácet své závazky, bude kvalita skóringového modelu reprezentovat jeho schopnost rozlišit mezi klienty, kteří své finanční závazky bez problému plní, a klienty, kteří jsou defaultní.

Předpokládejme, že máme k dispozici skóre S představující výstup nějaké skóringové funkce. Toto skóre je typicky odhadem pravděpodobnosti defaultu klienta. Dále máme ke každému klientovi dáno, zda u něj nastal nebo nenastal default. Zavedeme následující označení:

- $F_{BAD}(s)$ - distribuční funkce skóre špatných (defaultních) klientů
- $F_{GOOD}(s)$ - distribuční funkce skóre dobrých (nedefaultních) klientů
- $F_{ALL}(s)$ - distribuční funkce skóre všech klientů

$f_{BAD}(s)$, $f_{GOOD}(s)$ a $f_{ALL}(s)$ jsou příslušné pravděpodobnostní hustoty, μ_b, μ_g a μ jsou příslušné střední hodnoty a σ_b^2, σ_g^2, σ^2 jsou příslušné rozptyly.

Mezi základní charakteristiky popisující kvalitu prediktivního modelu, tj. např. skóringové funkce, patří střední diference (Mahalanobisova vzdálenost) [2], [7], Kolmogorovova-Smirnovova statistika [1], [2], [7], Giniho koeficient [1], [6], [9], Lift [1] a informační statistika [7], [8]. Obecně jsou tyto indexy dány vztahy:

Střední diference:

$$D = \frac{\mu_g - \mu_b}{\sigma^*}, \text{ pro } \sigma_g = \sigma_b = \sigma^* \tag{1}$$

$$D = \sqrt{2}\, D^*, \; D^* = \frac{\mu_g - \mu_b}{\sqrt{\sigma_g^2 + \sigma_b^2}}, \text{ pro } \sigma_g \neq \sigma_b \tag{2}$$

Kolmogorovova-Smirnovova statistika:

$$KS = \max_{s\in[L,H]} \left| F_{BAD}(s) - F_{GOOD}(s) \right| \tag{3}$$

Giniho koeficient:

$$G = 1 - 2\int_0^1 F_{GOOD}\left(F_{BAD}^{-1}(s)\right) ds \tag{4}$$

Lift:

$$Lift(s) = \frac{F_{BAD}(s)}{F_{ALL}(s)} \tag{5}$$

Informační statistika:

$$I_{val} = \int_{-\infty}^{\infty} \left(f_{GOOD}(s) - f_{BAD}(s)\right) \ln\left(\frac{f_{GOOD}(s)}{f_{BAD}(s)}\right) ds\,. \tag{6}$$

Postup při výpočtu empirických hodnot uvedených statistik je buď zřejmý nebo jej lze nalézt v [5].

2. Indexy kvality normálně rozdělených skóre

Předpokládejme nyní, že skóre špatných a dobrých klientů jsou normálně rozdělené se středními hodnotami μ_b, μ_g a s rozptyly σ_b^2, σ_g^2. Pak lze snadno ukázat, že za předpokladu rovnosti rozptylů platí:

$$KS = \Phi\left(\frac{D}{2}\right) - \Phi\left(\frac{-D}{2}\right) = 2\cdot\Phi\left(\frac{D}{2}\right) - 1, \tag{7}$$

kde $\Phi(\cdot)$ je distribuční funkce standardizovaného normálního rozdělení.

Pro Giniho koeficient platí

$$G = 2\cdot\Phi\left(\frac{D}{\sqrt{2}}\right) - 1\,. \tag{8}$$

Lift je dán vztahem

$$Lift_q = \frac{1}{q}\Phi\left(\frac{\sigma_{ALL}}{\sigma}\cdot\Phi^{-1}(q) + p_G\cdot D\right) \tag{9}$$

Konečně, snadno se ukáže, že informační statistiku lze vyjádřit jako

$$I_{val} = D^2\,. \tag{10}$$

Uvažujme nyní situaci normálně rozložených skóre bez předpokladu rovnosti rozptylů. Výpočetní vztahy pro KS, Lift a informační statistiku se výrazně zkomplikují. Vztah pro výpočet Giniho koeficientu je naproti tomu srovnatelně složitý za situace shodných i obecně různých rozptylů. Předchozí vztahy (7)-(10) i následující vztahy (12)-(14) lze nalézt v [4] a [5]. Výjimku tvoří vztah (12) pro KS statistiku, kdy v [4] je uveden nesprávný výraz. Pro KS tedy platí

$$KS = \Phi\left(\frac{a}{b}\sigma_b\cdot D^* - \frac{1}{b}\sigma_g\sqrt{a^2 D^{*2} + 2b\cdot c}\right) - \Phi\left(\frac{a}{b}\sigma_g\cdot D^* - \frac{1}{b}\sigma_b\sqrt{a^2 D^{*2} + 2b\cdot c}\right), \tag{11}$$

kde $a = \sqrt{\sigma_b^2 + \sigma_g^2}$, $b = \sigma_b^2 - \sigma_g^2$, $c = \ln\left(\frac{\sigma_g}{\sigma_b}\right)$. Giniho koeficient je dán vztahem

$$G = 2 \cdot \Phi\left(D^*\right) - 1 \tag{12}$$

Lift se dá vyjádřit pomocí vztahu

$$Lift_q = \frac{1}{q}\Phi_{\mu_b,\sigma_b^2}\left(\mu_{ALL} + \sigma_{ALL} \cdot \Phi^{-1}(q)\right) = \frac{1}{q}\Phi\left(\frac{\sigma_{ALL} \cdot \Phi^{-1}(q) + \mu_{ALL} - \mu_b}{\sigma_b}\right). \tag{13}$$

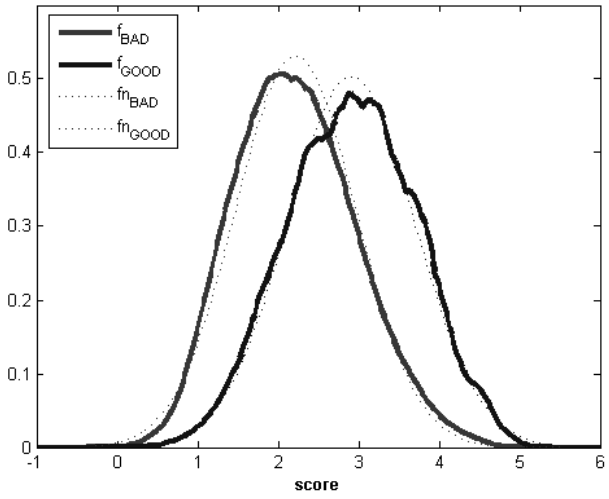
Konečně, informační statistika je dána vztahem

$$I_{val} = (A+1)D^{*2} + A - 1, \tag{14}$$

kde $A = \frac{1}{2}\left(\frac{\sigma_G^2}{\sigma_B^2} + \frac{\sigma_B^2}{\sigma_G^2}\right)$.

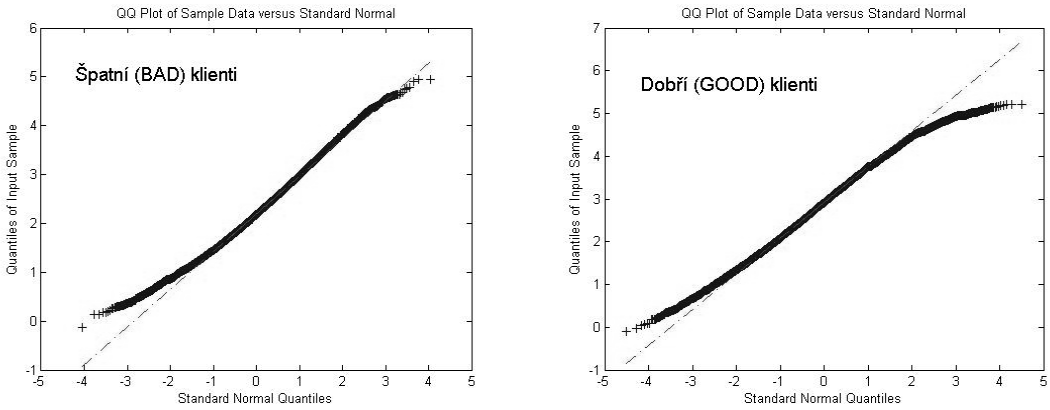
3. Aplikace

Následující výpočty jsou založeny na reálných datech jisté finanční instituce poskytující finanční produkty ve střední a východní Evropě. K dispozici byl vzorek čítající cca 175 000 hodnot skóre a ukazatele dobrého klienta. Skóre představuje transformovaný odhad pravděpodobnosti splácení finančních závazků klienta. Na obrázku 1 jsou zobrazeny pravděpodobnostní hustoty skóre dobrých a špatných klientů. Plné čáry představují jejich jádrové odhady, čárkované čáry pak aproximace pomocí hustoty normálně rozdělených náhodných veličin s empiricky odhadnutými parametry střední hodnoty a rozptylu.



Obrázek 1: Hustoty skóre dobrých a špatných klientů

Je zřejmé, že skóre dobrých i špatných klientů je rozloženo normálně jen velmi přibližně. Testování hypotézy o normalitě, viz [3], těchto skóre dopadne negativně. V obou případech je p-hodnota Lillieforsova testu <0,001. Nicméně vzhledem k rozsahu datového souboru jde o očekávaný výsledek. Následující obrázek 2 obsahuje Q-Q grafy pro obě skóre a je zřejmé, že ani z těchto grafů neplyne perfektní shoda s normální rozdělením.



Obrázek 2: Q-Q graf skóre dobrých a špatných klientů

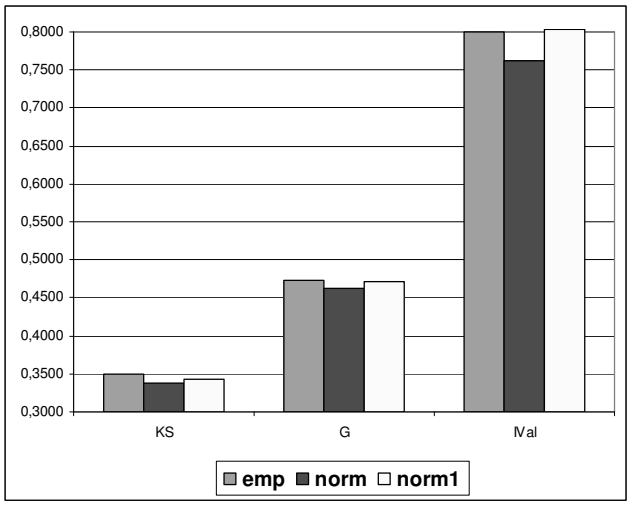
Nicméně podoba Q-Q grafů nebrání posouzení přesnosti odhadů míry kvality skóringové funkce počítaných pomocí empirických vzorců s výsledky, které obdržíme při použití vztahů (7) až (10) a vztahů (11) až (14).

Následující tabulka 1 udává hodnoty Kolmogorovova-Smirnovova koeficientu, Giniho koeficientu a informační statistiky vypočtené empiricky, pomocí vztahů pro normálně rozložená skóre při shodnosti rozptylů, tj. s využitím vztahů (7), (8) a (10), a konečně pomocí vztahů pro normálně rozložená skóre s obecnými rozptyly, tj. pomocí (11), (12) a (14).

Tabulka 1:KS, Giniho koeficient, informační statistika

	KS	G	I_{Val}
emp	0,3492	0,4723	0,7998
norm	0,3374	0,4629	0,7617
norm1	0,3424	0,4714	0,8027

Vzhledem k rozsahu datového vzorku lze považovat empiricky spočtené hodnoty za velmi přesné odhady hodnot uvažovaných koeficientů. Jak je z tabulky 1 vidět, hodnoty získané pomocí vztahů (7), (8) a (10) významně podhodnocují hodnoty všech tří zkoumaných koeficientů. Naproti tomu výsledky získané pomocí vztahů (11), (12) a (14) vykazují mnohem větší shodu s hodnotami získanými empirickým výpočtem. Graficky je vše zachyceno na obrázku 3.



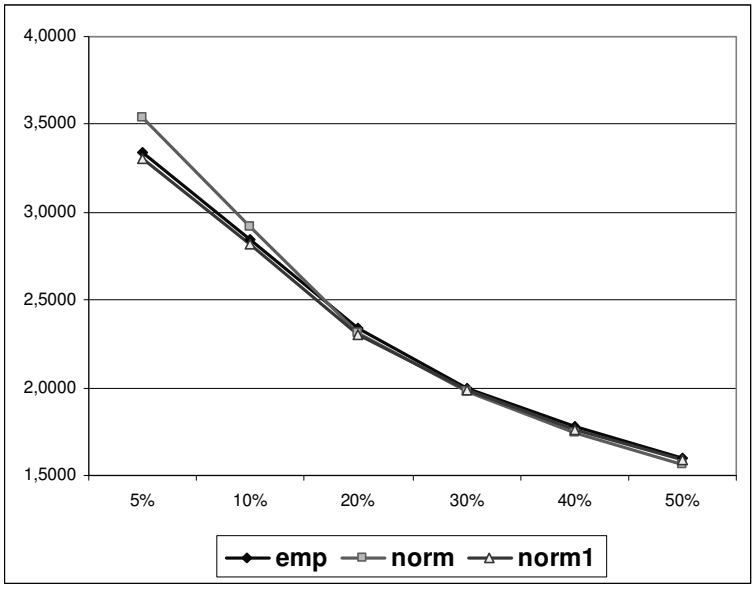
Obrázek 3: KS, Giniho koeficient, informační statistika

Hodnoty Liftu pro *q = 5 , 10, 20,…,100 %* jsou uvedeny v tabulce 2.

Tabulka 2:Lift

	$Lift_{0,05}$	$Lift_{0,1}$	$Lift_{0,2}$	$Lift_{0,3}$	$Lift_{0,4}$	$Lift_{0,5}$	$Lift_{0,6}$	$Lift_{0,7}$	$Lift_{0,8}$	$Lift_{0,9}$	$Lift_{1}$
emp	**3,3397**	**2,8426**	**2,3366**	**1,9978**	**1,7757**	**1,5963**	**1,4385**	**1,3055**	**1,1908**	**1,0901**	**1,0000**
norm	**3,5438**	**2,9126**	**2,3130**	**1,9761**	**1,7424**	**1,5635**	**1,4183**	**1,2955**	**1,1882**	**1,0916**	**1,0000**
norm1	**3,3069**	**2,8185**	**2,3037**	**1,9913**	**1,7649**	**1,5863**	**1,4380**	**1,3106**	**1,1981**	**1,0961**	**1,0000**

Také v tomto případě je vidět, že hodnoty vypočtené pomocí vztahu (13), tj. s obecně různými rozptyly, jsou v lepší shodě s hodnotami empirickými. Nejvýraznější je rozdíl pro úroveň *q = 5 %* a *q =10 %*, což jsou nejčastěji užívané hodnoty. Především proto, že okolo úrovně 10 % je obvykle nastavena úroveň zamítání pomocí skóringové funkce a je tudíž důležité mít silný skóringový model právě v těchto místech. Situaci ilustruje i následující obrázek 4.



Obrázek 4: Porovnání hodnot Liftu pro q=5, 10, …,50 %

4. Závěr

Situace kdy je třeba posoudit kvalitu skóringového modelu je v praxi velmi častá. K dispozici je řada indexů či měr, pomocí kterých lze tuto kvalitu určit. Mezi ty nejpoužívanější patří indexy uvedené v tomto článku. Jejich výpočet je možné provést pomocí empirických vztahů nebo vztahů založených na předpokladu normality skóre dobrých a špatných klientů. V druhém případě je nutné rozlišit situaci, kdy rozptyly obou skóre považujeme za shodné nebo připouštíme jejich obecnou různost.

Z uvedené aplikační studie vyplývá, že předpoklad shodných rozptylů skóre dobrých a špatných klientů nemusí být správnou volbou. Na zkoumaných datech vede tento předpoklad k podhodnocení odhadů indexů KS, Giniho koeficientu i informační statistiky. V případě Liftu došlo naopak k nadhodnocení pro hodnoty q = 5 % a q = 10 %. Ovšem hodnoty všech indexů vypočtených za předpokladu normality skóre a obecně různých jejich rozptylů vykazují velmi dobrou shodu s hodnotami vypočtenými empiricky. Celkem se tedy dá říci, že použití vztahů (11) až (14) je mnohem vhodnější než užívání vztahů (7) až (10).

5. Literatura

[1] ANDERSON, R. 2007. The Credit Scoring Toolkit: Theory and Practice for Retail Credit Risk Management and Decision Automation, Oxford University Press, Oxford. ISBN 9780199226405.

[2] HAND, D.J. and HENLEY, W.E. 1997. Statistical Classification Methods in Consumer Credit Scoring: a review. In: Journal. of the Royal Statistical Society, Series A, 160,No.3:523-541.

[3] LILLIEFORS, H.W. 1967. On the Komogorov-Smirnov test for normality with mean and variance unknown. In: Journal of the American Statistical Association, 62:399-402.

[4] ŘEZÁČ, M. 2009. Indexy kvality normálně rozložených skóre. In: Forum Statisticum Slovacum, 2009, č. 2., 144-148.

[5] ŘEZÁČ, M., ŘEZÁČ, F. 2009. Quality indexes of predictive models in risk and portfolio management. In: IMAEF 2009 Proceedings. Ioannina, Greece.

[6] SIDDIQI, N. 2006. Credit Risk Scorecards: developing and implementing intelligent credit scoring, Wiley, New Jersey.

[7] THOMAS, L.C., EDELMAN, D.B., CROOK, J.N. 2002. Credit Scoring and Its Applications, SIAM Monographs on Mathematical Modeling and Computation, Philadelphia. ISBN 9780898714838.

[8] WILKIE, A.D. 2004. Measures for comparing scoring systems. In: Readings in Credit Scoring 2004, Thomas, L.C., Edelman, D.B., Crook, J.N. (Eds.). Oxford University Press, Oxford, 51-62.

[9] XU, K., 2003. How has the literature on Gini's index evolved in past 80 years? [online]. [cit. 2009-01-15] URL: <http://economics.dal.ca/RePEc/dal/wparch/howgini.pdf>.

Adresa autora:

Martin Řezáč, Mgr. Ph.D.
Ústav statistiky a operačního výzkumu PEF MZLU
Zemědělská 1, 613 00 Brno
rezac@mendelu.cz

Optimalizace čistého zisku marketingové kampaně
Net income optimization of marketing campaign

Martin Řezáč

Abstract: Non-differentiated marketing may be strongly uneconomical, while target marketing is a way how to fight off still growing competition pressure at the present time. Concerning the optimal marketing strategy, it is necessary to know an estimate of probability of response for each client. Knowing this and a few of other parameters it is possible to model and optimize the effectiveness of marketing campaign with respect to net income.

Key words: Target marketing, probability of response, net income.

Klíčová slova: Cílený marketing, pravděpodobnost responze, čistý zisk.

1. Úvod

Vhodným nástrojem optimalizace marketingové kampaně je tzv. cílený marketing. Úkolem je nalézt podskupinu (segment) klientů, kteří s vysokou pravděpodobností kladně odpoví na učiněnou marketingovou nabídku. Nezbytností je v tomto případě konstrukce modelu odhadujícího pravděpodobnost pozitivní responze. Osloveni jsou pak ti klienti, u nichž model predikuje pozitivní responzi vyšší než je jistá mezní hodnota (cut-off).

Metodologií konstrukce zmíněného modelu existuje celá řada, stejně tak jako různých typů těchto modelů. Mezi nejznámější a v praxi nejpoužívanější patří model logistické regrese, viz [3].

2. Finanční kvantifikace modelu

Porovnáváme-li vzájemně kvalitu několika různých modelů, srovnáváme obvykle jejich Giniho index (viz. [3], [4]), případně index navýšení (Lift) na desátém percentilu, viz [1], [2] a [5]. Definici a výpočetní vzorce těchto indexů lze nalézt v uvedených publikacích, nicméně pro úplnost uveďme, že distribuční funkce pozitivně/negativně respondujících klientů jsou dány vztahy

$$F_{n.POS}(a) = \frac{1}{n}\sum_{i=1}^{n} I\left(s_i \le a \wedge Y = 1\right), \tag{1}$$

$$F_{m.NEG}(a) = \frac{1}{m}\sum_{i=1}^{m} I\left(s_i \le a \wedge Y = 0\right),\ a \in [0,1], \tag{2}$$

kde s_i je skóre i-tého klienta, n, resp. m, je počet klientů s pozitivní, resp. negativní, responzí. Dodejme, že distribuční funkci skóre všech klientů

$$F_{N.ALL}(a) = \frac{1}{N}\sum_{i=1}^{N} I\left(s_i \le a\right),\ a \in [0,1], \tag{3}$$

kde $N = n + m$ je počet všech klientů.

Giniho index je při tomto označení možné vyjádřit jako

$$Gini = 1 - \sum_{k=2}^{n+m} \left(F_{m.NEG_k} - F_{m.NEG_{k-1}} \right) \cdot \left(F_{n.POS_k} + F_{n.POS_{k-1}} \right). \quad (4)$$

Lift je dán vztahem

$$Lift(a) = \frac{1 - F_{n.POS}(a)}{1 - F_{N.ALL}(a)}, \; a \in [0,1]. \quad (5)$$

Obecně je velmi těžké stanovit nějakou mez, od které lze model považovat za dobrý, protože odhad responze u odlišných typů marketingových kampaní muže být velmi různě obtížný.

Porovnání kvality dvou modelu je sice důležité, v praxi je však mnohem důležitější vědět, jaký finanční dopad bude aplikace modelu mít. Máme-li k dispozici prediktivní model a známe-li pro něj funkci zisku $y = G(x)$, která pro dané procento oslovených klientů x udává procento pozitivně respondujících klientů $G(x)$, pak k výpočtu jeho finančního přínosu potřebujeme znát hodnotu následujících čtyř parametrů:

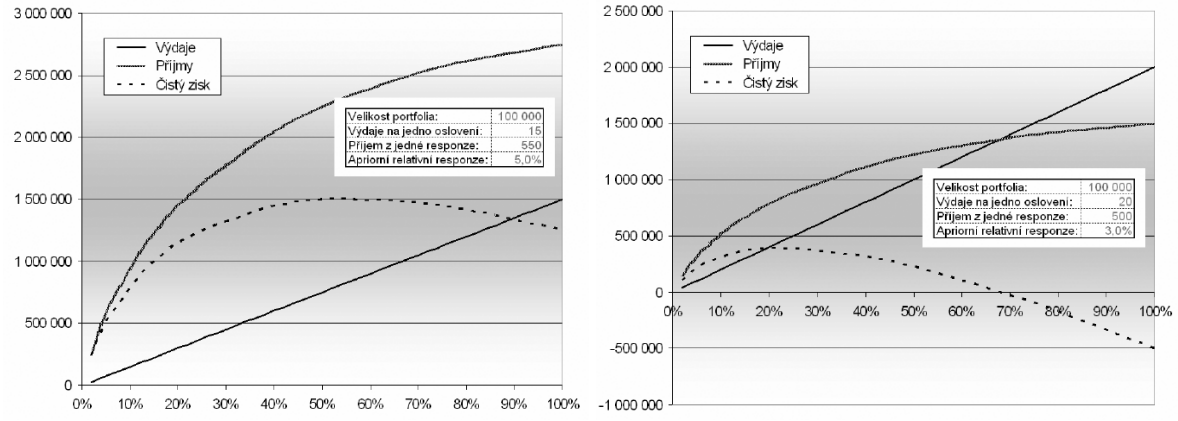
1. celkové náklady C na obeslání jednoho klienta
2. čistý příjem I z jednoho respondenta (po odečtení rizikových nákladů)
3. předpokládanou apriorní responzi R, tj. takovou relativní responzi, jakou bychom dosáhli bez použití modelu
4. počet klientů N, kteří potenciálně mohou být osloveni.

Za předpokladu znalosti těchto parametrů, lze snadno sestrojit křivku zisku (obrázek 1 a 2), která nám přesně řekne, jaké procento klientů bychom měli oslovit, abychom maximalizovali zisk z dané marketingové kampaně. Čistý zisk je totiž dán výrazem

$$\left(G(x) \cdot I \cdot R - x \cdot C \right) \cdot N, \quad (6)$$

kde x značí procento oslovených klientů a $G(x)$ je výše zmíněná funkce zisku. Více viz [2]. Je zřejmé, že pro náhodný výběr platí $G(x) = x$, což znamená, že příjmy i čistý zisk rostou lineárně a zisk je nejvyšší v bode $x = 100$ %.

Čtvrtý údaj N (velikost portfolia) nemá vliv na průběh křivek a tedy ani na optimální procento oslovených klientů, tato informace však slouží k tomu, aby y-ová osa vyjadřovala srozumitelný údaj, a to čistý zisk akce v peněžních jednotkách. Podíváme-li se na obrázek 1, snadno vidíme, že uplatnění prediktivních modelů je tím nevyhnutelnější, čím máme vyšší jednotkové náklady na oslovení, nižší příjem z respondujícího klienta a nižší předpokládanou apriorní responzi. Také je patrné, že poměrně malá změna těchto parametrů má velký vliv na křivku zisku. Jsou-li parametry nastaveny tak, jak ukazuje levá část obrázku 1, pak je čistý zisk maximalizován při oslovení 54 % klientského portfolia. Ovšem pravá část obrázku 1 dává i po malé změně parametrů podstatně pesimističtější výsledky: nejvyššího zisku dosáhneme při oslovení 21 % klientů a pokud jich oslovíme více než 68 %, dostaneme se dokonce do ztráty. Samozřejmě však mohou nastat situace, kdy je zisk maximalizován až při oslovení všech klientů – v tom případě prediktivní model není zapotřebí. Lze si ovšem představit, že nás tlaky konkurenčního prostředí zavedou do takové konstelace parametrů, kdy bez použití responzního modelu k rozumnému zisku nedospějeme.



Obrázek 1: Graf čistého zisku marketingové kampaně 2

3. Závěr

Uvedené indexy, tedy Giniho index a Lift, velmi dobře poslouží při porovnání kvality modelů předikujících pozitivní responzi marketingové kampaně. Máme-li k dispozici nejlepší možný model, vyvstává otázka jak je použít. Zcela zásadní je odhad čistého zisku plynoucího z oslovených klientů v rámci dané kampaně. Jak bylo ukázáno můžou se křivky zisku výrazně lišit i při poměrně malých změnách vstupních parametrů. Nicméně popsaným způsobem je při znalosti všech parametrů možné optimalizovat jednu marketingovou kampaň. Vyšší úrovně optimalizace ovšem docílíme, pokusíme-li se optimalizovat vytěžovací proces jako celek. Pokud máme například možnost oslovit klienty několika různými nabídkami, z nichž každá s sebou nese odlišné náklady na oslovení, jinou výši příjmu z responze i rozdílnou předpokládanou apriorní responzi, stojíme před poměrně náročnou optimalizační úlohou.

4. Literatura:

[1] COPPOCK, D.S. 2002. Why Lift? DM Review Online. [online]. [cit. 2009-01-15]. URL: <http://www.dmreview.com/news/5329-1.html>.

[2] NEPIL, M. 2007. Data mining v praxi. In: Datakon 2007. Brno. s. 25-38. ISBN 978-80-7355-076-9.

[3] THOMAS, L.C., EDELMAN, D.B., CROOK, J.N. 2002. Credit Scoring and Its Applications, SIAM Monographs on Mathematical Modeling and Computation, Philadelphia. ISBN 9780898714838.

[4] ŘEZÁČ, M., ŘEZÁČ, F. 2009. Quality indexes of predictive models in risk and portfolio management. In: IMAEF 2009 Proceedings. Ioannina, Greece.

[5] ŘEZÁČ, M. 2008. Kvantitativní charakteristiky modelu pozitivní response marketingové kampaně. In: Aktuálne marketingové trendy v teórii a praxi. Žilina: EDIS - vydavateľstvo Žilinskej univerzity s. 39--43. ISBN 978-80-8070-964-8.

Adresa autora:

Martin Řezáč, Mgr. Ph.D.
Ústav statistiky a operačního výzkumu PEF MZLU
Zemědělská 1, 613 00 Brno
rezac@mendelu.cz

Rodová analýza na základe EU-SILC
Gender Analysis according to EU-SILC Data

Ľubica Sipkova

Abstract: Article analyzes the conditions between men and women in area of medical care, education and labor, based on official Slovak statistical survey EU-SILC 2007, in comparison with the officially published indicators of multilateral economic institutions and the Slovak authorities. The results have been compared with the similar analysis provided by Trexima, Ltd. Co. The main objective of this article is to analyze how suitable are data for regular survey EU-SILC, which are provided by the Slovak Statistical Office in line with the provision of the European Council for comparison social and living conditions for citizens in the European Union. Based on analysis for EU-SILC 2007, it was not find possible applicability for advanced multivariate statistical methods and models in order to find causes significant wage disparities between men and women in Slovakia.

Key words: Gender_gap, Gender_equality, Mainstreaming, gender inequality, EU SILC, Pivot_table, Pivot_graph, Gender_analysis, Gender_pay_gap

Kľúčové slová: rodová_nerovnosť, rodová_rovnosť, rodový_mainstreaming, EU SILC, kontigenčné tabuľky, rodová_analýza, rodový_mzdový_rozdiel

1. Úvod

Dôvody skúmania rodovej nerovnosti a snahy o odstraňovanie nesúladu možno posudzovať zo sociálneho aj ekonomického hľadiska. Z ekonomického hľadiska možno rozdielne podmienky mužov a žien na trhu práce považovať za nedostatočné využitie ľudského potenciálu, schopností a talentu žien, a tým za nevyužité možnosti rozvoja ekonomiky. Z hľadiska sociálneho je rovnosť pre mužov a žien považovaná za kľúčový predpoklad decentného života celej spoločnosti.

Chápanie pojmu „rovnosť“ sa však často zamieňa s významom „zhodnosť“, alebo „rovnakosť“. Preto je potrebné zdôrazniť, že v článku je diskutovaná rovnosť šancí, rovnaký prístup k finančným zdrojom a k ponúkaným službám v spoločnosti, a tiež rovnosť v odmene za rovnakú prácu. Odmieta sa „feministický“ prístup s jednostranným nevedeckým hodnotením a snahou dosiahnuť celkovú rovnakosť pohlaví v ekonomickej a politickej oblasti.

V príspevku je porovnaná situácia mužov a žien v oblasti zdravia, vzdelania a práce na základe výsledkov štatistického zisťovania EU-SILC 2007 s oficiálne prezentovanou úrovňou svetovými inštitúciami a slovenskými úradmi. Výsledky sú hodnotené aj vzhľadom na závery z podobných analýz organizácie Trexima, s.r.o., ktorá monitoruje rodovú mzdovú diskrimináciu v Čechách a na Slovensku a projekt je podporovaný z prostriedkov Európskeho sociálneho fondu.

Hlavným cieľom v príspevku je posúdiť vhodnosť pravidelného výberového zisťovania EU-SILC, ktoré zabezpečuje Štatistický úrad Slovenskej republiky podľa nariadení Európskej rady pre účely porovnania sociálnych a životných podmienok obyvateľov EÚ. Posúdenie uvedeného súboru údajov je len z hľadiska možností aplikácie pokročilých viacrozmerných štatistických metód a tvorby modelov s cieľom odhaliť dôvody významných mzdových rozdielov mužov a žien na Slovensku.

2. Rodová nerovnosť a hodnotenie jej úrovne

Prístup, ktorý je prezentovaný v článku, je založený na posune od hodnotenia a skúmania postavenia ženy v spoločnosti k exaktnej analýze rodových súvislostí, t. j. gendrových procesov. Kategória rod (anglicky gender s obmenami man a woman) sa zakladá na analýzach z aspektu možností a príležitostí zahrňujúc rovnako mužov ako ženy.

Analýza exaktnými štatisticko-matematickými metódami prináša hodnotenie nerovnosti rovnako medzi kategóriami muž a žena, ako aj nerovnosti v rámci každej kategórie. Pri intersekcionálnom prístupe sa ku kategórii rod pristupuje adekvátne ako napríklad k národnostným, vekovým, regionálnym, náboženským, príjmovým, vzdelanostným, alebo iným členeniam spoločnosti. Rodové analýzy, t. j. gendrové analýzy, t. j. analýzy z rodového pohľadu, majú slúžiť ekonomickému rozvoju smerujúcemu k redukcii chudoby, rastu celkovej prosperity spoločnosti a rastu uspokojovania potrieb.

Rodová rovnosť (gender equality), t. j. rovnosť medzi mužmi a ženami, znamená, že všetky ľudské bytosti, muži aj ženy, sú slobodní v rozvoji svojich osobných schopností a robia rozhodnutia bez obmedzení stereotypmi, prísnymi rodovými rolami a predsudkami. Rodová rovnosť znamená, že „odlišné správanie, ašpirácie a potreby žien a mužov sú zohľadňované, hodnotené a preferované rovnako. Neznamená to, že ženy a muži musia byť takí istí, ale že ich práva, zodpovednosti a príležitosti nebudú závisieť od toho, či sa narodili ako muž alebo žena.“ [3]

Rodová analýza vychádza z tzv. rodovo členených údajov, t. j. rodovej štatistiky, tiež uvádzané v množnom čísle – rodových štatistík. Kategória pohlavie mužské a ženské (anglicky sex s obmenami male a female) je definovaná z biologického aspektu a v texte sa vyskytuje vzhľadom na to, že ide o jeden zo sledovaných štatistických znakov v rodových štatistikách, t. j. v údajovej základni s členením podľa pohlavia.

Rodovú nerovnosť (v literatúre označovanú aj ako gendrová nerovnosť, gendrové rozdiely, rodová priepasť, disparita medzi mužmi a ženami, alebo priamo anglickým výrazom gender gap) možno všeobecne definovať ako „rozdiel medzi ženami a mužmi v miere participácie, možnostiach prístupu, právach, ako aj výsledkoch v ktorejkoľvek oblasti verejnej sféry; nerovnosť medzi ženami a mužmi alebo dievčatami a chlapcami v prístupe k moci, zdrojom, sociálnym výhodám, vzdelaniu, ochrane zdravia a pod.” [5]

V súvislosti s rodovou nerovnosťou a rodovou rovnosťou možno definovať aj pojem rodová spravodlivosť. Znamená stav, keď ženy aj muži rovnako plnia sociálne funkcie a majú rovnaký prístup k zdrojom a výhodám nezávisle od príslušnosti k pohlaviu. Neznamená absolútnu zhodu v postavení mužov a žien. Z pohľadu rodovej spravodlivosti má distribučný princíp v spoločnosti dvojaký aspekt. Prvý, aspekt rovnakej deľby, znamená rovnaké zaobchádzanie v zhodných prípadoch, druhý je aspekt založený na uznaní odlišností medzi pohlaviami a rovnaká úroveň u oboch sa môže dosiahnuť len pri nerovnakom zaobchádzaní.

V exaktných štatisticko-matematických metódach sú rodové nerovnosti a rodové mzdové (príjmové) nerovnosti vyjadrené kvantitatívne pomocou viacerých štatistických mier, t. j. charakteristík, ktoré poskytujú kvantitatívny obraz o postavení mužov a žien v rôznych oblastiach života spoločnosti. Podľa definície rodovej spravodlivosti by sa však nemalo očakávať, že všetky ukazovatele a štatistické miery by pri rodovej rovnosti mali dosahovať rovnaké hodnoty. Hodnotenie úrovne rodovej rovnosti si preto vyžaduje aplikáciu viacrozmerných štatistických metód, ako aj výpočet kompozitných a komparatívnych mier. Komplexnosť rodových vzťahov spoločnosti však nie je možné vystihnúť jednou štatistickou mierou a ani popísať jedným viacrozmerným modelom. Modelovanie v štruktúrach obyvateľstva a v čase pozri napr. [1]. Rôzne prístupy k ohodnoteniu a z rôznych hľadísk sa vzájomne dopĺňajú a mali by poskytovať nezaujatú základňu pre politické rozhodnutia smerujúce k spravodlivej spoločnosti.

Stav rodovej rovnosti na Slovensku v súlade s definíciami EÚ monitoruje Rada vlády SR pre rodovú rovnosť. Zameriava sa na najvýraznejšie rodové rozdiely a indikátory rodových nerovností, bez výraznejšieho hodnotenia, alebo hľadania príčinných súvislostí. „Opierame sa o základné štrukturálne indikátory rodovej rovnosti na európskej úrovni a vychádzame prevažne, aj keď nie výlučne, z vymedzení a konštrukcie indikátorov umožňujúcich celoeurópske porovnanie," konštatuje sa v nej [4]. Hodnotenie úrovne rodovej rovnosti na Slovensku vychádza z údajových základní a štatistických publikácií Štatistického úradu SR, Eurostatu, Slovstatu, regionálnych databáz a špecifických databáz jednotlivých ministerstiev, ktoré aj samostatne poskytujú údaje a uverejňujú rôzne analytické hodnotenia v skúmanej oblasti rodovej nerovnosti aj na svojich webových stránkach.

Každoročne Rada pre rodovú rovnosť zosumarizuje dostupné informácie z medzinárodných hodnotení SR a z všetkých vyššie uvedených zdrojov a publikuje Súhrnnú správu. Všetky v nej aplikované indikátory rodovej nerovnosti sú vymenované spolu so stručným uvedením spôsobu výpočtu na konci tejto hodnotiacej správy.

Spolufinancovaný výskum a monitorovanie rodových nerovností Európskym sociálnym fondom má organizácia Trexima Bratislava, s.r.o. V oblasti rodovej nerovnosti sa zameriava hlavne na monitorovanie rodovej mzdovej nerovnováhy v podmienkach Slovenska a jeho regiónov. Má vypracovaný Jednotný systém monitorovania rodovej nerovnosti na náhodnom výbere podnikov.

Z mesačných hlásení spracováva priebežné štvrťročné hodnotenia o úrovni (z priemerov aj mediánov) rodového mzdového rozdielu v SR a vplyvu veku, dĺžky a úrovne vzdelania, sféry podnikateľskej a nepodnikateľskej, dĺžky odbornej praxe, druhu úväzku aj podľa štruktúry zamestnaní - KZAM na jeho výšku. Sleduje koncentráciu mužov a žien v jednotlivých zamestnaniach a odvetviach podľa veku a vzdelania. Aplikuje viaceré metódy rozkladu mzdových rozdielov, napr. Oaxaca-Blinder dekompozíciu a počíta Duncanov segregačný index. Vybrané analytické výsledky sú poskytované na vyžiadanie, verejne prezentované a publikované aj na webových stránkach, napr. [6].

3. Východiská našich analýz o rodovej nerovnosti a jej meraní v SR

V príspevku prezentované výsledky vychádzajú zo súboru údajov EU-SILC 2007, UDB_08/08/20, poskytnutých Štatistickým úradom SR. Získané boli stratifikovaným náhodným výberom, formou zisťovania PAPI za rok 2007 len pre 2. rotačnú skupinu v 5840 domácnostiach v SR v súlade s platnými nariadeniami Európskej komisie ohľadom výberu, spôsobu zisťovania a definovania premenných. Informácie o 12 573 osobách vo veku 16 rokov a viac sú sústredené do samostatného súboru osobných prierezových údajov, ktorý je označený ako P_súbor. Každá osoba má svoje osobné ID RB041.

Zistené údaje boli kontrolované a upravované pre neodpovede, nekontaktovanie domácnosti, nevyčíslenie a zistené chyby. Imputácie, predovšetkým na základe priemerovania, boli uplatnené len pre príjmové premenné. Váženia by mali byť aplikované na základe v P_súbore uvedených individuálnych prierezových váh PB040. Kalibrácia váh bola na externé počty osôb len podľa veku a pohlavia v krajoch NUTS III tak, aby sa úhrn váh rovnal počtu obyvateľov SR vo veku 16 a viac.

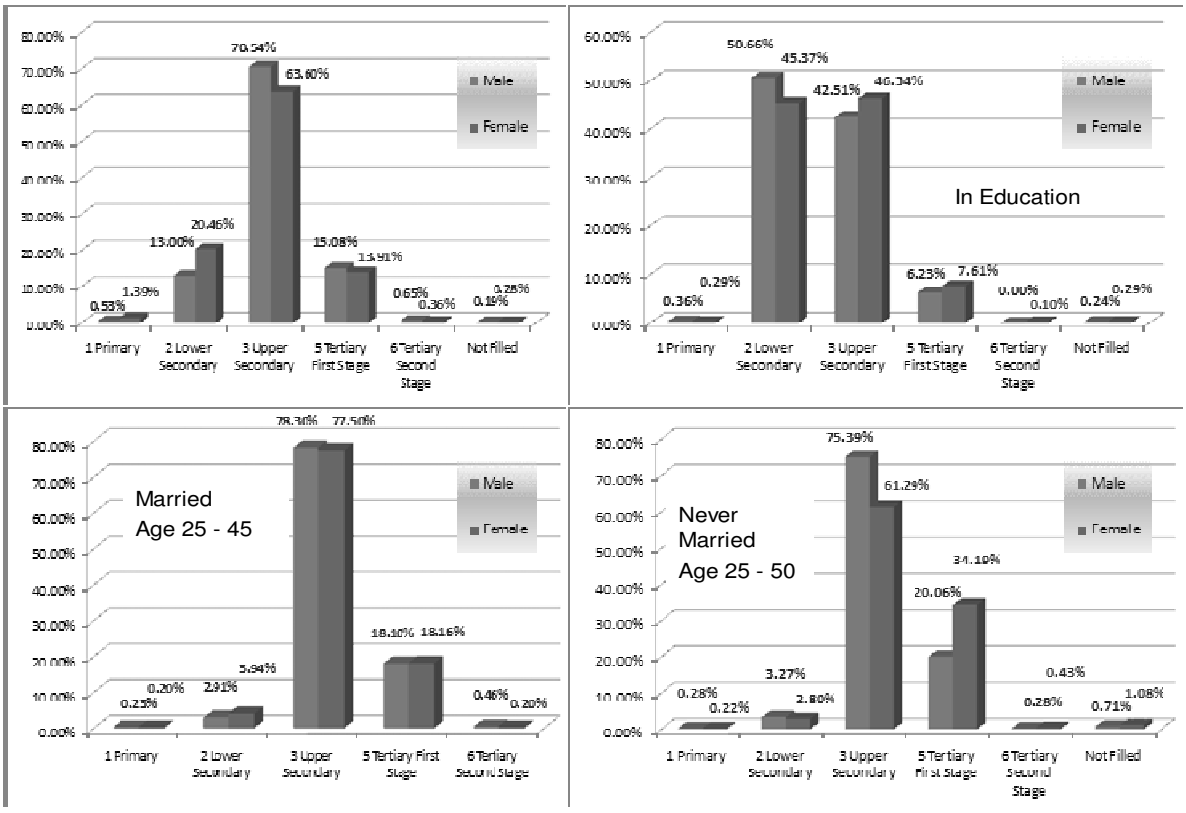
Analyzované premenné, obsiahnuté v P_súbore, sú v príspevku označené v súlade s označovaním v databáze, spolu so stručným názvom podľa nariadení EK a oficiálnych príloh k databáze. Preto ich v článku nedefinujeme, definície cieľových premenných možno nájsť na oficiálnej stránke EK v Nariadení Komisie (ES) č. 1983/2003 zo dňa 7. 11.2003. Primárne premenné sú rozdelené do piatich oblastí. Prvou oblasťou sú základné osobné údaje (22 premenných - ID, váha a demografické údaje, premenná PB150 pre pohlavie). Druhou sú údaje o vzdelaní (4 premenné – súčasná vzdelávacia aktivita a jej úroveň, rok dosiahnutého najvyššieho vzdelania a vzdelanie podľa ISCED 19997). Treťou oblasťou sú premenné o

zdravotnom stave a poskytovanej zdravotnej starostlivosti (7 premenných – všeobecné zdravie, trpí chronickými chorobami a o naplnení/nenaplnení potrieb druhov medicínskeho vyšetrenia a dôvody). V štvrtej oblasti sú informácie o práci, ekonomickej aktivite a postavení (38 premenných – klasifikácia podľa ISCO 1988, dĺžke pracovného času, zamestnanosti, klasifikácia odvetvia činnosti NACE, postavenie type pracovnej zmluvy, dôvodoch zmeny, počte odpracovaných rokov, ...). Poslednou oblasťou sú údaje o osobnom príjme (15 premenných – séria komponentov hrubých príjmov a séria komponentov čistých príjmov, namiesto PY200G Hrubá mesačná mzda zo zamestnania je v databáze uvedená „slovenská premenná“ SPY200G).

Použité pre hodnotenie úrovne rodovej nerovnosti podľa vyššie uvedenej údajovej základne boli najskôr metódy triedenia a grafického zobrazovania v hlbších štruktúrach dát pomocou Pivotných tabuliek (sústavy prierezových viacstupňových kontigenčných tabuliek umožňujúcich aj vyčíslenie základných deskriptívnych charakteristík) a k nim príslušných Pivotných grafov v softvéri EXCEL 2007. K aplikácii metód štatistickej indukcie bol použitý štatistický programový balík STATGRAPHICS CENTURION XV.

4. Niektoré získané výsledky a hodnotenia rodovej nerovnosti v SR

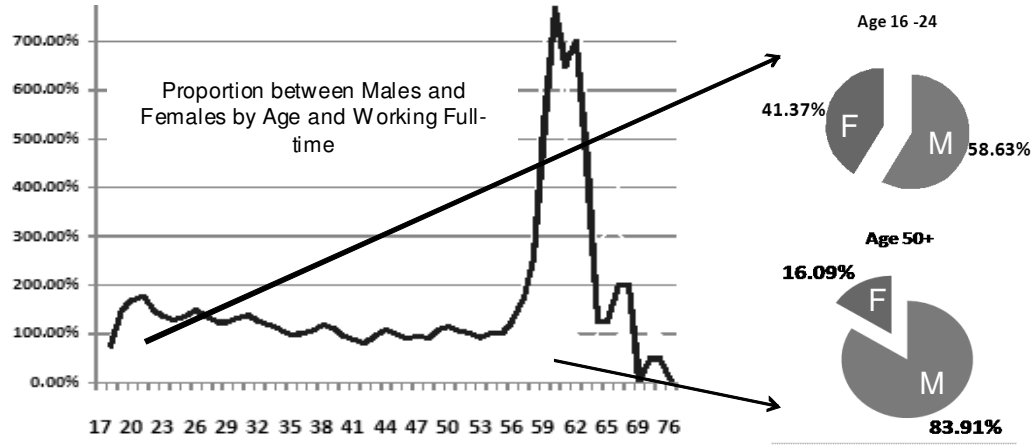
V článku sú prezentované len niektoré výsledky analýz z dôvodu, že hlavným cieľom v príspevku nie je analyzovať úroveň rodovej nerovnosti v SR, ale posúdiť vhodnosť poskytnutej údajovej základne pre komplexnejšie viacrozmerné štatistické analýzy v skúmanej oblasti nerovnosti na Slovensku. Zdrojom nasledujúcich obrázkov a tabuľky sú vlastné grafické zobrazenia a výpočty.



Obrázok 6: Najvyššie dosiahnuté vzdelanie podľa ISCED

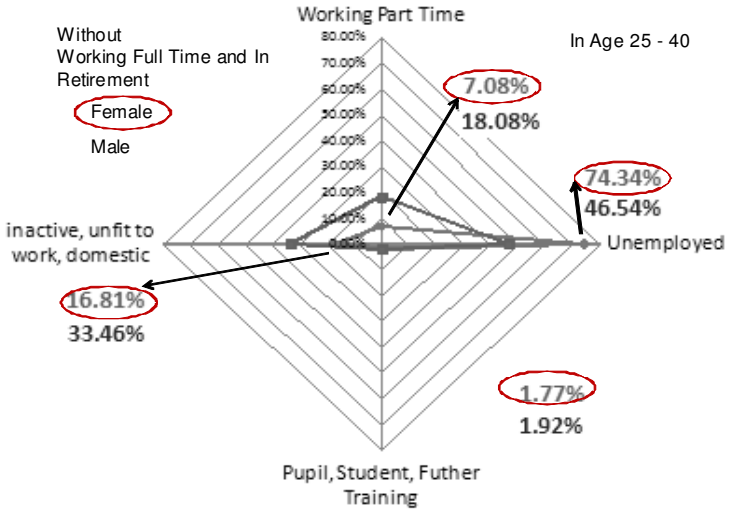
Percentuálny podiel žien s vysokoškolským vzdelaním vo veku 16 rokov a viac, ktoré boli v čase zisťovania vo vzdelávacom procese je vyšší ako u mužov (obrázok 1), t. j. 46,34% žien oproti 42,51% mužov. Vo vedúcom postavení nikdy nevydatých vysokoškolsky vzdelaných žien je 46,75% (oproti 42,17% nikdy neženatých vysokoškolsky vzdelaných

mužov, ktorí sú v riadiacej pozícii) v zamestnaní na plný úväzok. Znamená to, že na Slovensku sa na vedúce postavenie v zamestnaní na plný úväzok dostane ľahšie žena práve s vysokoškolským vzdelaním ako takýto muž, ale aj ak sa pritom ešte ďalej vzdeláva.



Obrázok 2: Proporcia medzi mužmi a ženami pracujúcimi na plný úväzok podľa ich veku

Z kompozície grafov (obrázok 2) je zrejmý rozdiel v percentuálnych bodoch u mužov a žien hlavne vo veku 16 až 24 rokov, ale hlavne päťdesiat a viac ročných, kedy na jednu ženu pripadne viac ako 7 mužov v zamestnaní na plný úväzok.



Obrázok 3: Rodová nerovnosť u nepracujúcich na plný úväzok

V radiálnom grafickom zobrazení (obrázok 3) mužov a žien, ktorí nepracujú na plný úväzok a nie sú na dôchodku, vidieť rodovú disparitu. Muži udávajú skôr že sú nezamestnaní a ženy, že sú osobami v domácnosti, alebo inak neaktívne. Ženy skôr pracujú na čiastočný úväzok ako muži. Posunutie štvoruholníka u žien je smerom k čiastočnému úväzku a k osobe v domácnosti, kým muži ne-dôchodkári, ktorí nepracujú na plný úväzok sú radšej v kategórii nezamestnaný.

Zaujímavým je samohodnotenie úrovne všeobecného zdravia dôchodcami a dôchodkyňami. Ide o subjektívne hodnotenie a podľa porovnaní percentuálnych bodov u mužov a žien možno postrehnúť väčšiu nespokojnosť so svojím zdravím u žien dôchodkýň, ako u mužov dôchodcov (tabuľka 1), ale len pokiaľ ide o „priemerné“ a „skôr zlé“ zdravie.

Tabuľka 8: Samo-ohodnotenie svojho zdravia dôchodcami v % rodových úhrnov

Health PH010	Very Good	Good	Fair	Bad	Very Bad
Male	1.85%	9.45%	39.57%	34.01%	15.11%
Female	1.69%	8.57%	40.91%	34.33%	14.50%
Gender Gap	**0.16%**	**0.88%**	**-1.33%**	**-0.32%**	**0.61%**

Podobných hodnotení a deskriptívnych analýz pomocou Pivotných tabuliek v prepojení s pivotnými grafmi je možné robiť v rôznych štruktúrach a členeniach pri výpočte rôznych charakteristík.

Pri skúmaní rodových mzdových rozdielov sa výsledky výrazne nelíšili od Treximov prezentovaných záverov. Zistený mediánový mzdový rozdiel (vyjadrenie disparity mediánových hrubých hodinových miezd mužov a žien v percentách mediánovej úrovne miezd mužov) je 18,75% a priemerný 54,20%. Potvrdilo sa s 99 % spoľahlivosťou, že hrubá hodinová mzda u mužov je významne vyššia ako u žien, pričom aj rozptyl miezd je významne väčší. Nepodarilo sa potvrdiť významné rozdiely v podieloch mužov a v podieloch žien v rôznorodých štruktúrach podľa volených premenných zisťovaných v EU SILC v oblasti vzdelania, zdravia a pracovných podmienok.

5. Záver

Databáza EU-SILC predstavuje primeraný zdroj údajov pre meranie mzdových rozdielov, ale je obmedzená čo sa týka rozsahu hlbších štrukturálnych analýz. A to v dôsledku nedostatku informácií z oblasti zdravia, vzdelania a o štruktúre zamestnanosti z hľadiska povolaní, tiež pre nedostatok informácií o hrubých mesačných a hodinových mzdách.

Vzhľadom k skutočnosti, že zisťované znaky neboli volené pre rodovú analýzu, nie sú volené vhodne a nie je pomocou nich možné definovať príčiny rodových disparít na Slovensku. Časové rady na základe EU-SILC sú veľmi krátke. Váženie získaných hodnôt váhami, ktoré zohľadňujú len externé počty osôb podľa pohlavia a veku v krajoch, v rodových analýzach zdravia, vzdelania, pracovných podmienok nie je vhodné. Stratifikovaný náhodný výber v rozsahu uskutočnenom v SR za rok 2007 nepovažujeme za primeraný na reprezentatívnu viacrozmernú štatistickú analýzu rodových nerovností a na modelovanie v štruktúrach obyvateľstva.

6. Literatúra

[1] BARTOŠOVÁ, J. – BÍNA, V. Modeling of income distribution of Czech households in years 1996 – 2005, In. Acta Oeconomica Pragensia 3/2009, Oeconomica, Praha, 2009, s. 3-18, ISSN 0572-3043.

[2] Glosár rodovej terminológie, dostupné na http://glosar.aspekt.sk, 6.11.2009

[3] KVAPILOVÁ, E. – PORUBANOVÁ S. Rodová rovnosť: Prečo ju potrebujeme?, Stredisko pre štúdium práce a rodiny, Bratislava, 2003. ISBN 80-89048-10-2.

[4] Súhrnná správa o stave rodovej rovnosti na Slovensku a o činnosti Rady vlády SR pre rodovú rovnosť za rok 2008. Rada vlády pre rodovú rovnosť, 2008. Dostupné na *www.gender.gov.sk/index.php?id=627*

[5] VRANOVÁ, Z. – FILADELFIOVÁ J., http://slovnik.aspekt.sk , 6.11.2009

[6] Základné informácie – zenymuzi.sk, dostupné na http://www.zenymuzi.sk/, 6.11.2009

Adresa autora:

Ľubica Sipková, Ing., PhD., Javorinská 9A, 811 03 Bratislava, SR, lubica.sipkova@euba.sk

Príspevok je súčasťou riešenia výskumných projektov IGP č. 21/2008-2009, VEGA 1/4586/2007-2009 a GAČR 402/09/0515

Analýza monetárnej chudoby v domácnostiach Českej republiky
Analysis of monetary poverty in households in Czech Republic

Iveta Stankovičová

Abstract: The aim of this paper is to present logistic regression model of risk of monetary poverty in Czech households. We used data from statistical survey EU SILC 2007. The goal of this survey is to obtain information on the distribution of income, the level and structure of poverty and the social exclusion. It is a harmonized survey, which has been carried out from year 2005 in all 25 (27) EU member countries.

Key words: risk-of-poverty rate, statistical survey EU SILC 2007, Czech Republic, system SAS

Kľúčové slová: miera rizika chudoby, štatistické zisťovanie EU SILC 2007, Česká republika, systém SAS

1. Úvod

Chudoba a proces jej prehlbovania v čase hospodárskej krízy, je považovaná za vážny spoločenský problém, ktorý je potrebné sledovať a riešiť. Je to problém, ktorý sa nedotýka len obyvateľov tretieho sveta, ale aj obyvateľov rozvinutých krajín, nevynímajúc ani Európu. V štátoch EÚ 25 bolo v roku 2006 ohrozených tzv. monetárnou chudobou až 16% obyvateľstva.

V príspevku sa pokúsime o viacrozmerné modelovanie rizika monetárnej chudoby v Českej republike v roku 2007 s využitím logistickej regresie. Použili sme údaje EU SILC 2007 ČR, kde reprezentatívnu vzorku tvorilo 9675 českých domácností. Výpočty sme uskutočnili v prostredí softvéru SAS® Enterprise Guide 4.1 a SAS® Enterprise Miner 5.3.

Príspevok vznikol v procese analýzy súborov údajov výberového zisťovania o príjmoch a životných podmienkach domácností EU SILC 2007. Tieto súbory dát za Slovenskú a Českú republiku tvoria údajovú základňu pre riešenie dvoch výskumných projektov:

- projektu VEGA 1/4586/07 s názvom *Modelovanie sociálnej situácie obyvateľstva a domácností v Slovenskej republike a jej regionálne a medzinárodné porovnania*, ktorý je riešený na Katedre štatistiky FHI EU v Bratislave pod vedením prof. V. Pacákovej,
- projektu GAČR 402/09/0515 s názvom *Analýza a modelování finančního potenciálu českých (slovenských) domácností*, ktorý je riešený na Katedre managementu informácií FM VŠE v Jindřichovom Hradci pod vedením Dr. J. Bartošovej.

2. Miera rizika chudoby

Hranica rizika monetárnej chudoby je stanovená ako 60% mediánu národného ekvivalentného príjmu, v prepočte na paritu kúpnej sily a na Euro. Inými slovami, za monetárne chudobnú domácnosť sa považuje tá domácnosť, v ktorej je ekvivalentný disponibilný príjem pod hranicou rizika chudoby. V SR to bola v roku 2006 čiastka asi 7 tis. Sk na mesiac (7394 Sk) a v ČR to bolo zhruba 8 tis. Kč mesačne (7797 Kč). Údaje boli publikované národnými Štatistickými úradmi, ŠÚSR a ČSÚ, na základe výpočtov z dát EU SILC 2007. Tieto hraničné hodnoty použijeme aj v našej analýze.

Miera rizika chudoby (risk-of-poverty rate) predstavuje podiel osôb s ekvivalentným disponibilným príjmom pod hranicou 60% národného mediánu ekvivalentného príjmu. Podľa výsledkov EU SILC 2007 SR bolo v roku 2006 ohrozených rizikom peňažnej (monetárnej) chudoby 10,7 % obyvateľov Slovenska. Oproti roku 2004 sa miera rizika chudoby znížila o

2,6 percentuálneho bodu a v porovnaní s rokom 2005 poklesla o 0,9 percentuálneho bodu. V ČR peňažnou chudobou bolo ohrozených vyše 995 tis. osôb (t.j. 9,76 % zo všetkých osôb) a v podstate sa tento podiel oproti roku 2005 nezmenil. Na základe údajov EU SILC 2007 bolo v ČR vyše 4 mil. domácností a z nich až 9,77% bolo ohrozených chudobou (Tabuľka 1). V ďalších tabuľkách môžeme sledovať podiely ohrozených domácností chudobou podľa niektorých významných kategoriálnych premenných (Tabuľka 2 až 5).

Tabuľka 1: Domácnosti ČR podľa miery rizika chudoby (RISK60)

RISK60	Počet	v %
0	3648178	90.23%
1	**395162**	**9.77%**
Spolu	4043340	100.00%

Tabuľka 2: Výskyt rizika chudoby podľa pohlavia prednostu domácnosti

POHL_P	Podiel domácností	RISK60		Súčet
		0	1	
1	**77.47%**	93.44%	6.56%	**100%**
2	**22.53%**	79.17%	20.83%	**100%**
Súčet	**100.00%**	**90.23%**	**9.77%**	**100%**

Tabuľka 3: Výskyt rizika chudoby podľa vzdelania

DOM_VZD	Podiel domácností	RISK60		Súčet
		0	1	
1	**9.03%**	71.85%	28.15%	**100%**
2	**75.63%**	90.84%	9.16%	**100%**
3	**15.34%**	98.04%	1.96%	**100%**
Súčet	**100.00%**	**90.23%**	**9.77%**	**100%**

Tabuľka 4: Výskyt rizika chudoby podľa sociálnej skupiny

SKUP	Podiel domác.	RISK60		Súčet
		0	1	
1	**24.80%**	93.49%	6.51%	**100%**
2	**12.50%**	93.62%	6.38%	**100%**
3	**24.44%**	98.10%	1.90%	**100%**
6	**4.44%**	99.23%	0.77%	**100%**
7	**28.13%**	89.17%	10.83%	**100%**
8	**4.67%**	31.94%	68.06%	**100%**
9	**1.03%**	37.98%	62.02%	**100%**
Súčet	**100.00%**	**90.23%**	**9.77%**	**100%**

Tabuľka 5: Výskyt rizika chudoby podľa pracovnej aktivity

DOM_EA	Podiel domác.	RISK60		Súčet
		0	1	
1	**67.48%**	94.90%	5.10%	**100%**
2	**4.55%**	28.07%	71.93%	**100%**
3	**27.21%**	90.73%	9.27%	**100%**
4	**0.75%**	28.95%	71.05%	**100%**
Súčet	**100.00%**	**90.23%**	**9.77%**	**100%**

3. Príprava premenných pre modelovanie

Ekvivalentný disponibilný príjem (ročný) je ročný disponibilný príjem domácnosti vydelený ekvivalentnou veľkosťou domácnosti. Tento príjem je potom priradený každému členovi domácnosti (definícia Eurostatu).

Ekvivalentná škála je škála koeficientov použitá na výpočet indikátorov chudoby v súlade s metodikou Eurostatu (tzv. modifikovaná škála OECD), kde koeficient 1 sa použije pre prvého dospelého člena domácnosti, 0,5 pre druhého a každého dospelého člena domácnosti, 0,5 pre 14-ročných a starších a 0,3 pre každé dieťa mladšie ako 14 rokov. Používa sa na výpočet ekvivalentnej veľkosti domácnosti.

V dátach EU SILC 2007 ČR sa nachádza premenná CP_PRIJ (čistý peněžní příjem domácnosti v Kč za rok) a tiež premenná EJ (počet spotřebních jednotek - def. EU). Aby sme získali *ekvivalentný disponibilný príjem* podľa definície Eurostatu, tak sme si vytvorili premennú EU_PRIJ podľa vzťahu: EU_PRIJ = CP_PRIJ / EJ. V ďalšom kroku sme odpočítali od EU_PRIJ *hranicu rizika monetárnej chudoby* (t.j. 60% mediánu národného ekvivalentného príjmu). Hranica rizika chudoby z dát EU SILC 2007 bola stanovená ČSÚ na 93560 Kč na rok (t.j. 7797 Kč na mesiac), čiže zistili sme premennú ROZDIEL=EU_PRIJ - 93560 pre

každú domácnosť v súbore dát. Potom sme vytvorili novú binárnu premennú RISK60. Hodnotu 1 v tejto premennej majú domácnosti pod hranicou chudoby (ROZDIEL ≤ 0) a hodnotu 0 domácnosti, ktoré nemusia čeliť monetárnej chudobe podľa vyššie uvedenej definície (ROZDIEL > 0). Pri výpočtoch sme používali frekvenčnú premennú PKOEF (přepočítací koeficient pro přepočet výsledků z výběru na celý soubor v populaci).

4. Interpretácia výsledkov modelu logistickej regresie

Modelom logistickej regresie modelujeme podmienenú pravdepodobnosť, že domácnosť je pod hranicou chudoby v závislosti od rôznych činiteľov. V našom prípade bola závislá (modelovaná, target) binárna premenná RISK60, konkrétne jej hodnota 1, t.j. domácnosť je pod hranicou chudoby.

Zoznam významných vysvetľujúcich premenných v našom modeli je uvedený v tabuľke (Tabuľka 6). Získali sme ich použitím selekčnej metódy *stepwise* na hladine významnosti 0,05 pre zaradenie aj vyradenie z modelu. Celkový počet významných premenných v modeli je 11, z toho sú 4 kvantitatívne a 7 kategoriálnych.

Ako kvantitatívne vysvetľujúce premenné boli vybrané nasledovné premenné: počet členov v domácnosti (OSOB), počet nezaopatrených detí (DETI), počet ekonomicky aktívnych členov (EA) a tiež druh domácnosti podľa pracovnej aktivity DOM_PI, kde PI je koeficient pracovnej intenzity členov domácnosti vo veku 16-64 rokov (je to podiel počtu mesiacov počas ktorých boli osoby ekonomicky aktívne (EA) na celom roku). Ide síce o kategoriálnu premennú, ale jej 4 kategórie sú definované nasledovne: 1 - zcela nepracující (PI=0), 2 - spíše nepracující (0<PI<0.5), 3 - spíše pracující (0.5<PI<1) a 4 - plně pracující (PI=1). Takúto poradovú kategoriálnu premennú je možné použiť aj ako kvantitatívnu premennú v modeli.

Ostatné premenné v modeli sú kategoriálne. Ako významné sa ukázali sociálna skupina (SKUP), druh vzdelania (DOM_VZD) a druh ekonomickej aktivity (DOM_EA), ale aj typ obce (TYP), veľkosť obce (VEL) a stupeň urbanizácie (OBLAST). Významné je aj pohlavie osoby v čele domácnosti (POHL_P).

Tabuľka 6: Zoznam významných premenných

Type 3 Analysis of Effects				
Effect	**DF**	**Wald Chi-Square**	**Pr > ChiSq**	**Popis významných premenných**
SKUP	6	75305.94	<.0001	Soc. skupina osoby v čele: 1-nižší zamestnanec, 2-samostatne činný, 3-vyšší zamestnanec, 6-dôchodca s EA členmi, 7-dôchodca bez EA členov, 8-nezamestnaný, 9-ostatní (7 skupín)
DOM_EA	3	57166.97	<.0001	Druh domác. podľa prac. aktivity: 1-pracujúci, 2-nepracujúci, 3-dôchodca, 4-nepracujúci ostatní
EA	1	49340.16	<.0001	Počet ekonomicky aktívnych (pracujúcich) členov
DOM_VZD	2	35693.55	<.0001	Druh domácnosti podľa vzdelania: 1-nízka, 2-stredná, 3-vysoká úroveň
POHL_P	1	28267.15	<.0001	Pohlavie osoby v čele (1-muž, 2-žena)
DETI	1	20617.58	<.0001	Počet nezaopatrených detí (najviac do 25 rokov vrátane)
TYP	3	10985.81	<.0001	Typ obce: 1-Praha, 2-krajské mesto, 3-mestská obec, 4-dedinská obec
DOM_PI	1	6953.91	<.0001	Druh domác. podľa prac. intenzity (PI): 1 až 4-plne pracujú
VEL	8	6297.25	<.0001	Veľkosť obce: 9 skupín (1-do 199, …, 9-nad 100 tis. obyv.
OSOB	1	1523.30	<.0001	Počet členov domácnosti
OBLAST	2	1008.25	<.0001	Oblasť (stupeň urbanizácie): 1-husto, 2-stredne, 3-riedko

Kvalita nami prezentovaného modelu je dobrá (viď Tabuľka 7). Pomer zhodných párov dosiahol až 89,4% a plocha pod ROC krivkou je 0,896 (*c* štatistika, tzv. ROC index). Svedčí to o vysokej zhode napozorovaných a predikovaných hodnôt závislej premennej prezentovaným modelom. Výskyt extrémnych hodnôt meraných štatistikou C je znázornený na obrázkoch (Obrázok 1 a 2).

Na interpretáciu logistického modelu použijeme pomery šancí (ods ratio), ktoré uvádzame v tabuľke (Tabuľka 8). Pomery šancí z intervalu (0, 1) znamenajú, že so zvyšovaním hodnoty vysvetľujúcej premennej o jednotku, sa znižuje šanca domácnosti stať sa chudobnou. Pomery šancí z intervalu (1, ∞) znamenajú, že so zvyšovaním hodnoty vysvetľujúcej premennej o jednotku, sa zvyšuje šanca domácnosti dostať sa pod hranicu chudoby (za predpokladu, že ostatné vysvetľujúce premenné sa nemenia).

Z výsledkov v tabuľke 8 je zrejmé, že so zvyšovaním počtu členov domácnosti (OSOB) a tiež so zvyšovaním počtu ekonomicky aktívnych členov (EA) sa šanca domácnosti prepadnúť pod hranicu chudoby znižuje. Platí to aj pre pracovnú aktivitu domácnosti (DOM_PI). Naopak, rastúci počet závislých detí v domácnosti (DETI) zvyšuje túto šancu, v priemere 2,27:1 na každé ďalšie dieťa.

Pri interpretácii pomerov šancí pre kategoriálne premenné sú porovnávané šance jednotlivých kategórií s poslednou kategóriou (Tabuľka 8). Najvýznamnejšou premenou sa ukázala sociálna skupina osoby v čele domácnosti (SKUP). Pomery šancí ku kategórii 9-ostatní ukazujú, že najnižšiu šancu dostať sa pod hranicu chudoby majú domácnosti 3-vyšších zamestnancov (0,14:1) a 6-dôchodcov s EA členmi (0,08:1). Naopak najvyššiu šancu stať sa chudobnými majú domácnosti 8-nezamestnaných (4,2:1).

Ďalšou významnou premennou je druh domácnosti podľa ekonomickej aktivity (DOM_EA). Pri porovnávaní s kategóriou 4-nepracujúci (nikto nie je EA ani nezamestnaný a ani dôchodca), paradoxne vychádza, že šance pre domácnosti 1-pracujúcich (aspoň 1 člen EA) dostať sa pod hranicu chudoby sú vyššie (4,43:1). Môže to byť spôsobené tým, že táto kategória je v korelácii s počtom EA členov v domácnosti alebo tiež preto, že táto kategória (4-nepracujúci) je veľmi široko a nepresne definovaná. V procese ďalšieho precizovania modelu bude nutné toto nejako vyriešiť. Bude asi vhodnejšie namiesto tejto premennej vybrať premennú DOM_OECD (druh domácnosti: 1-plně zaměstnaná (jeden dospělý je EA anebo 2+ dospělých, z nichž aspoň 2 jsou EA), 2-nezaměstnaná (nikdo z dospělých není EA), 3-částečně zaměstnaná, 4-důchodecká (jen osoby ve věku 65+, nejsou EA).

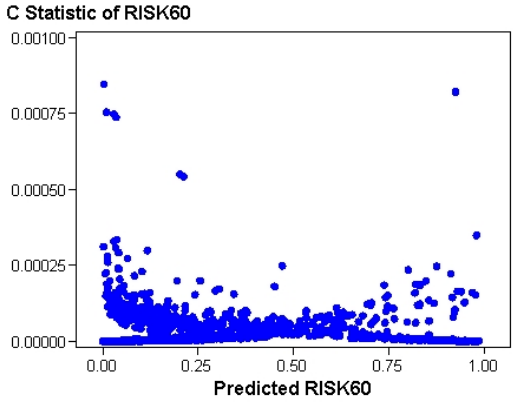
Pri členení domácností podľa vzdelania (DOM_VZD), model porovnáva šance s domácnosťami 3-vysoká úroveň vzdelania, kde má aspoň jeden z partnerov vysokoškolské vzdelanie. Je logické, že šance rodín s nižším vzdelaním dostať sa pod hranicu chudoby sú vyššie. U domácností 1-nízka úroveň vzdelania (obaja partneri ZŠ alebo SŠ) je táto šanca až 8,39:1 a v domácnostiach 2-stredná úroveň vzdelania (aspoň jeden partner SŠ) je to 3,98:1.

Čo sa týka pohlavia prednostu domácnosti (POHL_P), tak šanca domácnosti dostať sa do chudoby, kde na čele je 1-muž, je nižšia, ako keď tam je 2-žena (0,38:1). Ak si vypočítame prevrátenú hodnotu tejto šance (1/0,38=3,18) tak môžeme povedať, že šanca domácnosti na čele so ženou je 3-krát vyššia ako keď na čele domácnosti je muž. Z tabuľky 2 je to tiež zrejmé. Podiel domácností v súbore ak je na čele žena pod hranicou chudoby je až 20,83%.

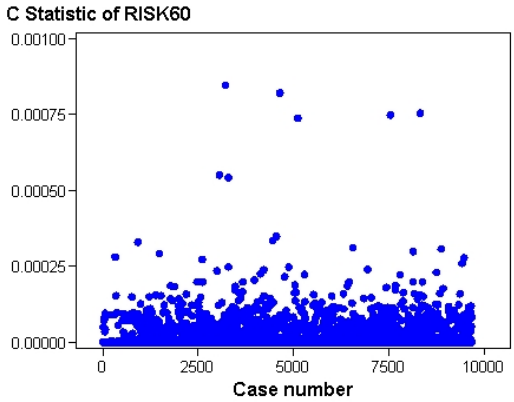
Pri premennej typ obce (TYP) je najnižšia šanca dostať sa do chudoby v 1-Prahe hl. meste v porovnaní s 4-dedinskou obcou (0,58:1). Čo sa týka veľkosti obce (VEL), tak nižšie šance na chudobu majú domácnosti z obcí pod 1000 obyvateľov (kategórie 1 až 3) a vyššie šance zas domácnosti z obcí kde je tisíc až 100 tis. obyvateľov (kategórie 4 až 8). Čo sa týka stupňa urbanizácie (OBLAST), tak oproti kategórii 3-riedke osídlenie, v ostaných kategóriách urbanizácie je nižšia šanca dostať sa do chudoby.

Tabuľka 7: Štatistiky kvality modelu

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	89.4	Somers' D	0.792
Percent Discordant	10.2	Gamma	0.796
Percent Tied	0.4	Tau-a	0.140
Pairs	1.44E+12	c	0.896



Obrázok 1: Extrémne hodnoty podľa predikovanej pravdepodobnosti



Obrázok 2: Extrémne hodnoty podľa poradia

Tabuľka 8: Odhady pomerov šancí

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
OSOB	0.84	0.83	0.85
DETI	2.27	2.25	2.30
EA	0.14	0.14	0.14
DOM_PI	0.67	0.67	0.68
TYP 1 vs 4	0.58	0.55	0.60
TYP 2 vs 4	1.60	1.55	1.66
TYP 3 vs 4	1.01	0.98	1.03
POHL_P 1 vs 2	0.38	0.37	0.38
OBLAST 1 vs 3	0.88	0.86	0.90
OBLAST 2 vs 3	0.84	0.83	0.85
VEL 1 vs 9	0.92	0.87	0.96
VEL 2 vs 9	0.83	0.79	0.87
VEL 3 vs 9	0.96	0.92	1.00
VEL 4 vs 9	1.35	1.30	1.41
VEL 5 vs 9	1.70	1.64	1.77
VEL 6 vs 9	1.24	1.20	1.28
VEL 7 vs 9	1.34	1.30	1.38
VEL 8 vs 9	1.05	1.02	1.07
SKUP 1 vs 9	0.44	0.41	0.46
SKUP 2 vs 9	0.56	0.53	0.59
SKUP 3 vs 9	0.14	0.13	0.15
SKUP 6 vs 9	0.08	0.07	0.08
SKUP 7 vs 9	1.97	1.85	2.09
SKUP 8 vs 9	4.20	3.96	4.45
DOM_EA 1 vs 4	4.43	4.12	4.76
DOM_EA 2 vs 4	0.62	0.58	0.67
DOM_EA 3 vs 4	0.04	0.04	0.04
DOM_VZD 1 vs 3	8.39	8.20	8.59
DOM_VZD 2 vs 3	3.98	3.90	4.06

5. Záver

Pri analýze faktorov monetárnej chudoby v domácnostiach Českej republiky sme využili model logistickej regresie. Odhalili sme 11 významných faktorov, ktoré pôsobia na riziko chudoby (premenná RISK60). Ich vplyv sme kvantifikovali pomocou pomerov šancí. Ukázalo sa, že model je síce štatisticky významný, ale má ešte niektoré problematické miesta v interpretačnej oblasti. V ďalšej etape sa budeme preto snažiť model zlepšiť využitím interakčných a polynomických premenných a tiež úpravou niektorých premenných (napr. kategorizácia spojitých premenných, spájanie kategórií u vybratých kategoriálnych premenných). Môžeme však už teraz konštatovať, že model odhalil a kvantifikoval zaujímavé činitele, ktoré vplývajú na riziko chudoby v Českej republike.

Konečným naším cieľom bude odhadnúť parametre modelov logistickej regresie z dát EU SILC ČR aj SR a porovnať faktory pôsobiace na riziko chudoby v českých a slovenských

domácnostiach. Tento bod porovnávania pokladáme za problematický, lebo štruktúra dát EU SILC 2007 nie je úplne rovnaká v SR a ČR.

6. Literatúra

[1] BARTOŠOVÁ, J.: Analysis and Modelling of Financial Power of Czech Households. In Aplimat – Journal of Applied Mathematics, Vol. 2, Nr. 3, Slovak Technical University, Bratislava, 2009, s.31-36, ISSN 1337-6365.

[2] BARTOŠOVÁ, J., STANKOVIČOVÁ, I.: Diferenciace příjmu a chudoba českých a slovenských domácností. In Sborník příspěvku MSED 2009. VŠE Praha 2009. ISBN: 978-80-86175-66-9 (CD).

[3] LABUDOVÁ, V.: Analýza monetárnej chudoby na Slovensku. SAS Forum 2008. Bratislava, október 2008.

[4] STANKOVIČOVÁ, I., BARTOŠOVÁ, J.: Príspevok k analýze subjektívnej chudoby v SR a ČR. In Forum Statisticum Slovacum 3/2009. SŠDS, Bratislava 2009. ISSN: 1336-7420 (CD).

[5] STANKOVIČOVÁ, I., VOJTKOVÁ, M.: Viacrozmerné štatistické metódy s aplikáciami. Bratislava : Iura Edition, 2007. 261 s. ISBN 978-80-8078-152-1.

[6] ŽELINSKÝ, T., HUDEC, O.: Odhad subjektívnej chudoby na Slovensku založený na distribučnej funkcii rozdelenia príjmov. In Forum Statisticum Slovacum 7/2008. SŠDS, Bratislava, s. 152-157. ISSN 1336-7420.

[7] Materiály Štatistického úradu SR, web-stránka na : <http://portal.statistics.sk>

[8] Materiály Českého štatistického úradu, web-stránka na: <http://www.czso.cz>

[9] Životní podmínky (EU-SILC). Dostupné na (9. 5. 2009): <http://www.czso.cz/csu/redakce.nsf/i/zivotni_podminky_(eu_silc)>

Adresa autora:

Ing. Iveta Stankovičová, Ph.D.
Katedra informačných systémov, Fakulta managementu UK v Bratislave
Odbojárov 10, P. O. Box 95, 820 05 Bratislava 25
iveta.stankovicova@fm.uniba.sk

Príspevok bol vytvorený ako súčasť riešenia projektu VEGA 1/4586/07: *Modelovanie sociálnej situácie obyvateľstva a domácností v Slovenskej republike a jej regionálne a medzinárodné porovnania* a projektu GAČR 402/09/0515: *Analýza a modelování finančního potenciálu českých (slovenských) domácností.*

Problematika nízké síly Jarque-Bera testu a jeho modifikací proti rozdělením s velmi krátkými chvosty
The poor power of the Jarque-Bera test and its modifications against very short tailed distributions

Luboš Střelec

Abstract: This paper deals with problems of the poor power of the classical Jarque-Bera test and its modifications (the Jarque-Bera-Urzúa test and the robust Jarque-Bera test) to detect selected very short tailed alternatives (bimodal, uniform and beta distributions), especially in small sample sizes. The aim of this paper is to compare the power of the classical Jarque-Bera test in respect to the Jarque-Bera-Urzúa test, the robust Jarque-Bera test, the Anderson-Darling test, the Cramér-von Mises test, the Lilliefors (Kolmogorov-Smirnov) test, the Shapiro-Wilk test, the D`Agostino test, the Pearson test, the directed SJ test, the Geary test, the Uthoff test, the skewness test and the kurtosis test. For this purpose we utilize the Monte Carlo simulations.

Key words: Tests of normality, Monte Carlo simulations, Power comparison, Short tailed distributions, Bimodal distributions.

Klíčová slova: testy normality, Monte Carlo simulace, komparace síly testu, rozdělení s krátkými chvosty, bimodální rozdělení.

1. Úvod

Jarque-Bera test (Jarque a Bera, 1980) patří společně se Shapiro-Wilkovým testem (Shapiro a Wilk, 1965) a Lillieforsovým (Kolmogorov-Smirnovovým) testem (Lilliefors, 1967) mezi nejběžněji používané omnibus testy normality. Jarque-Bera test je pak nejvíce využívaným omnibus testem především v oblasti ekonomie a financí. Jarque-Bera test vykazuje velmi dobrou sílu proti většině typů alternativních rozdělení. Přesto však je možné najít případy, kdy Jarque-Bera test vykazuje velmi nízkou sílu (např. uniformní rozdělení) či dokonce nulovou sílu (např. bimodální rozdělení) a to především pro malé rozsahy souborů, na což upozornili již Thadewald a Bünning (2007). Mimo klasického Jarque-Bera testu byly odvozeny i jeho modifikace. Za první modifikaci lze považovat Urzúovu modifikaci Jarque-Bera testu, která využívá standardizace šikmosti a špičatosti využité v klasické Jarque-Bera testové statistice (blíže viz Urzúa, 1996). Za jiný typ modifikace Jarque-Bera testu lze považovat robustní Jarque-Bera test zkonstruovaný Gelovou a Gastwirthem (2008), který v čitateli šikmosti i špičatosti využívá průměrné absolutní odchylky od výběrového mediánu. Cílem této práce tak je provést na základě výsledků Monte Carlo simulací komparaci síly klasického Jarque-Bera testu normality a jeho modifikací s dalšími vybranými testy normality.

2. Materiál a metodika

Pro potřeby komparační analýzy bude sledována síla klasického Jarque-Bera testu (JB), Jarque-Bera-Urzúa testu (JBU), robustního Jarque-Bera testu (RJB) a vybraných běžně používaných testů normality – konkrétně Shapiro-Wilkova testu (SW), Lilliefors (Kolmogorov-Smirnovova) testu (LT), Anderson-Darlingova testu (AD), Cramér-von Misesova testu (CM), D`Agostinova testu (DT), Pearsonova testu (PT), directed SJ testu (SJ – blíže viz Gelová, Miao a Gastwirth, 2007), Gearyho testu (GT), Uthoffova testu (UT), testu výběrové šikmosti (SKT) a testu výběrové špičatosti (KT) – blíže uvedené testy viz např.

Thode (2002). Pro všechny sledované testy bude využito exaktních kritických hodnot získaných na základě Monte Carlo simulací. Jako alternativy jsou zvoleny vybraná rozdělení s krátkými a velmi krátkými chvosty – konkrétně uniformní rozdělení, beta rozdělení a rovněž vybraná bimodální rozdělení jako směsi dvou normálních rozdělení dané následující funkcí F

$$F = 0{,}5N(0,1) + 0{,}5N(\mu_2,1), \text{ kde } \mu_2 \in \{3,4,5\}. \quad (1)$$

3. Komparace síly sledovaných testů normality

V Tabulce 1 jsou uvedeny hodnoty síly sledovaných testů normality proti vybraným rozdělením s lehkými a velmi lehkými chvosty zahrnující uniformní rozdělení, beta(2,2) rozdělení, beta(4,4) rozdělení a výše vymezená bimodální rozdělení pro různé rozsahy souborů a to vše pro hladinu významnosti $\alpha = 0{,}05$. Všechny sledované alternativy jsou symetrické se špičatostí nižší než je špičatost normálního rozdělení (viz Tabulka 2).

Z uvedených výsledků je patrné, že nejsilnějšími testy proti uvedeným symetrickým lehkochvostým alternativám jsou především test výběrové špičatosti a pro bimodální rozdělení pak Uthoffův test. Naopak klasický Jarque-Bera test, Jarque-Bera-Urzúa test a robustní Jarque-Bera test vykazují v porovnání s ostatními testy velmi slabou sílu testu a to pro všechny sledované rozsahy souborů. Např. klasický Jarque-Bera test má téměř nulovou sílu pro uniformní rozdělení pro velmi malé soubory ($n \le 40$) a v případě středně velkých a velkých souborů ($60 \le n \le 100$) vykazuje sílu významně nižší než nejsilnější testy vůči této alternativě, kam lze zařadit test výběrové špičatosti, Uthoffův test, Gearyho test, Shapiro-Wilkův test či Anderson-Darlingův test. Ještě nižší sílu než klasický Jarque-Bera test vykazuje pro uniformní rozdělení Jarque-Bera-Urzúa test, robustní Jarque-Bera test, directed SJ test a test výběrové šikmosti. Poslední tři zmiňované testy vykazují prakticky nulovou sílu i pro velký rozsah souboru $n = 100$. Důvodem v případě sledovaných verzí Jarque-Bera testu je symetrie daného rozdělení a špičatost nevýznamně odlišná od špičatosti normálního rozdělení (hodnoty šikmosti a špičatosti jsou uvedeny v Tabulce 2). Podobné výsledky lze vyčíst i v případě alternativ beta(2,2) rozdělení a beta(4,4) rozdělení s tím, že téměř nulovou sílu pro velký rozsah souboru $n = 100$ vykazuje mimo robustního Jarque-Bera testu, directed SJ testu a testu výběrové šikmosti i klasický Jarque-Bera test a Jarque-Bera-Urzúa test.

Tabulka 9: Síla sledovaných testů normality pro hladinu významnosti $\alpha = 0{,}05$ a různé rozsahy souborů

	uniformní					beta(2,2)				
	20	**40**	**60**	**80**	**100**	**20**	**40**	**60**	**80**	**100**
AD	0,171	0,439	0,692	0,866	0,953	0,058	0,103	0,165	0,240	0,325
CM	0,141	0,335	0,535	0,718	0,846	0,057	0,090	0,133	0,191	0,252
DT	0,129	0,598	0,885	0,979	0,997	0,029	0,139	0,297	0,477	0,635
GT_{st}	0,238	0,529	0,744	0,874	0,944	0,097	0,192	0,286	0,395	0,491
JB	0,002	0,001	0,042	0,379	0,748	0,003	0,002	0,001	0,008	0,047
JBU	0,001	0,000	0,004	0,167	0,563	0,003	0,001	0,000	0,002	0,013
KT_{st}	***0,336***	***0,749***	***0,940***	***0,989***	***0,999***	***0,116***	***0,264***	***0,460***	***0,624***	***0,764***
LT	0,102	0,202	0,325	0,471	0,598	0,052	0,075	0,099	0,127	0,151
PT	0,101	0,147	0,230	0,347	0,447	0,059	0,067	0,075	0,094	0,107
RJB	0,002	0,000	0,000	0,000	0,008	0,004	0,001	0,000	0,000	0,000
SJ_{dir}	0,002	0,000	0,000	0,000	0,000	0,006	0,001	0,000	0,000	0,000
SKT_{st}	0,007	0,004	0,002	0,001	0,001	0,005	0,004	0,002	0,002	0,002
SW	0,199	0,576	0,865	0,972	0,996	0,050	0,109	0,199	0,324	0,463

UT_{st}	0,219	0,504	0,722	0,859	0,936	0,096	0,184	0,280	0,389	0,486
	beta(4,4)					bimodální (μ_2 = 3)				
	20	40	60	80	100	20	40	60	80	100
AD	0,043	0,053	0,059	0,075	0,085	0,119	0,308	0,507	0,681	0,803
CM	0,047	0,052	0,057	0,069	0,079	0,116	0,308	0,498	0,672	0,798
DT	0,021	0,039	0,064	0,104	0,133	0,120	0,376	0,580	0,746	0,846
GT_{st}	0,054	0,084	0,101	0,133	0,157	0,346	0,605	0,781	0,901	0,947
JB	0,012	0,005	0,002	0,002	0,004	0,001	0,000	0,007	0,069	0,212
JBU	0,011	0,003	0,001	0,001	0,001	0,001	0,000	0,001	0,021	0,101
KT_{st}	0,053	***0,087***	***0,128***	***0,176***	***0,225***	0,321	0,549	0,717	0,842	0,908
LT	0,046	0,051	0,055	0,061	0,061	0,083	0,217	0,351	0,498	0,611
PT	***0,056***	0,050	0,045	0,056	0,057	0,100	0,180	0,256	0,374	0,470
RJB	0,012	0,005	0,002	0,001	0,001	0,000	0,000	0,000	0,000	0,002
SJ_{dir}	0,015	0,009	0,005	0,003	0,003	0,000	0,000	0,000	0,000	0,000
SKT_{st}	0,016	0,011	0,006	0,005	0,006	0,001	0,000	0,000	0,000	0,000
SW	0,038	0,043	0,055	0,068	0,084	0,099	0,256	0,432	0,605	0,741
UT_{st}	0,053	0,080	0,102	0,133	0,158	***0,364***	***0,617***	***0,786***	***0,900***	***0,951***
	bimodální (μ_2 = 4)					bimodální (μ_2 = 5)				
	20	40	60	80	100	20	40	60	80	100
AD	0,396	0,839	0,977	0,997	0,999	0,770	0,994	***1,000***	***1,000***	***1,000***
CM	0,397	0,846	0,976	0,997	***1,000***	0,781	0,995	***1,000***	***1,000***	***1,000***
DT	0,424	0,862	0,973	0,996	0,999	0,778	0,994	***1,000***	***1,000***	***1,000***
GT_{st}	0,770	0,977	0,998	***1,000***	***1,000***	0,974	***1,000***	***1,000***	***1,000***	***1,000***
JB	0,000	0,002	0,192	0,683	0,920	0,000	0,051	0,774	0,991	**1,000**
JBU	0,000	0,000	0,038	0,458	0,828	0,000	0,000	0,429	0,963	0,999
KT_{st}	0,698	0,936	0,989	0,999	***1,000***	0,936	0,998	***1,000***	***1,000***	***1,000***
LT	0,277	0,674	0,893	0,975	0,993	0,574	0,958	0,999	***1,000***	***1,000***
PT	0,254	0,555	0,790	0,927	0,974	0,503	0,909	0,992	0,999	***1,000***
RJB	0,000	0,000	0,000	0,000	0,305	0,000	0,000	0,000	0,076	0,973
SJ_{dir}	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
SKT_{st}	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
SW	0,342	0,768	0,953	0,992	0,999	0,705	0,987	***1,000***	***1,000***	***1,000***
UT_{st}	***0,806***	***0,979***	***0,999***	***1,000***	***1,000***	***0,982***	***1,000***	***1,000***	***1,000***	***1,000***

Pozn.1: Počet Monte Carlo simulací je 10000.
Pozn.2: Zvýrazněny jsou testy s nejvyšší sílou pro danou alternativu a rozsah souboru.

V případě sledovaných bimodálních rozdělení vykazují velmi dobré výsledky síly Uthoffův test, Gearyho test a test výběrové špičatosti. Pro středně velké a velké rozsahy souborů ($60 \le n \le 100$) pak mimo uvedených testů vykazují dobré výsledky síly rovněž Anderson-Darlingův test, Cramer-von Misesův test, D`Agostinův test, Lilleforsův test, Pearsonův test a Shapiro-Wilkův test. Naopak klasický Jarque-Bera test a jeho modifikace (Jarque-Bera-Urzúa test a robustní Jarque-Bera test) vykazují v porovnání s ostatními testy nízkou sílu – pouze v případě bimodálního rozdělení pro $\mu_2 \in \{4,5\}$ a velký rozsah souboru $n = 100$ vykazuje klasický Jarque-Bera test a Jarque-Bera-Urzúa test sílu srovnatelnou

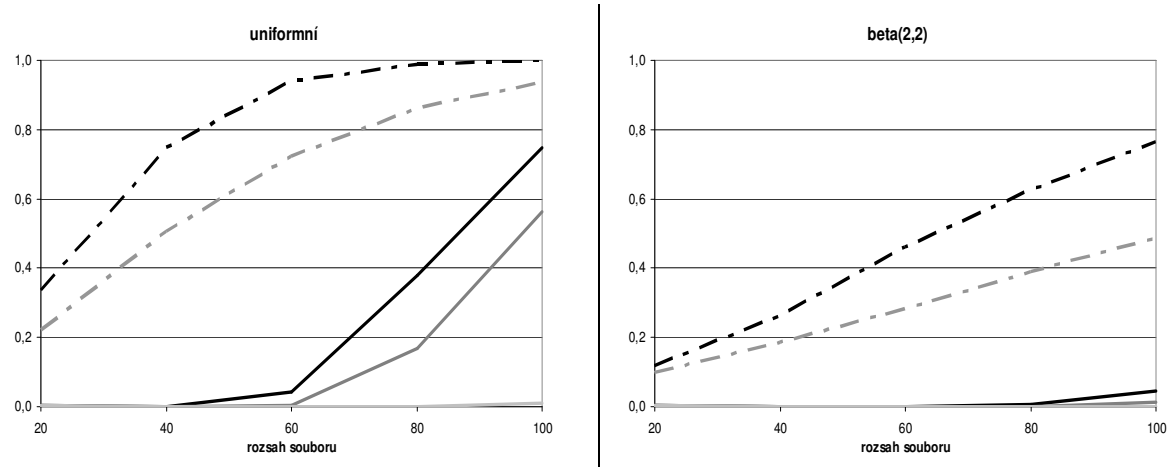
s nejsilnějšími testy pro danou alternativu. Důvodem je, že zatímco pro malé rozsahy souborů není rozdíl mezi hodnotami špičatosti sledovaných bimodálních rozdělení a hodnotou špičatosti normálního rozdělení významný rozdíl, tak pro středně velké a velké rozsahy souborů je tento rozdíl již významný, což vede k zamítnutí hypotézy o normalitě rozdělení.

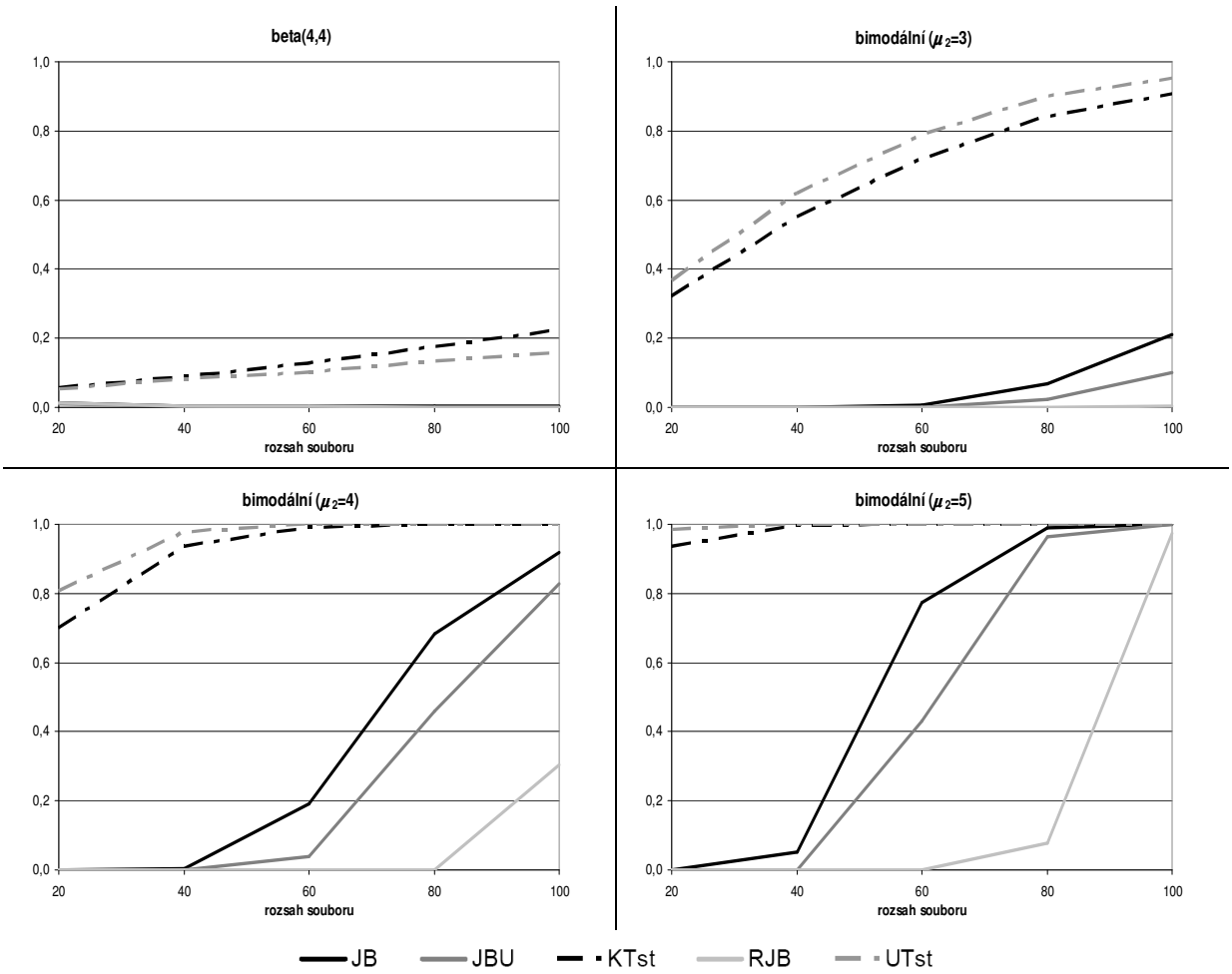
Tabulka 10: Charakteristiky šikmosti β_1 a špičatosti $\sqrt{\beta_2}$ pro sledovaná lehkochvostá rozdělení

	uniformní	beta(2,2)	beta(4,4)	bimodální		
				$\mu_2 = 3$	$\mu_2 = 4$	$\mu_2 = 5$
β_1	0,00	0,00	0,00	0,00	0,00	0,00
$\sqrt{\beta_2}$	1,80	2,14	2,46	2,04	1,72	1,51

Z výsledků provedených Monte Carlo simulací lze rovněž vypozorovat, že zatímco v případě Jarque-Bera testu, Jarque-Bera-Urzúa testu a robustního Jarque-Bera testu je patrný s růstem rozsahu souboru nárůst i jejich síly proti sledovaným bimodálním alternativám, tak test výběrové šikmosti a directed SJ test vykazují pro sledovaná bimodální rozdělení a všechny sledované rozsahy souborů nulovou sílu. Důvodem v případě testu výběrové šikmosti je symetrie sledovaných bimodálních rozdělení a v případě directed SJ testu fakt, že tento test je zkonstruován se zaměřením především na těžkochvostá rozdělení, tudíž logicky v případě lehkochvostých rozdělení vykazuje velmi nízkou sílu testu.

Ze srovnání klasického Jarque-Bera testu, Jarque-Bera-Urzúa testu a robustního Jarque-Bera testu (viz Obrázek 1) je patrné, že nejsilnější z nich je klasický Jarque-Bera test, následovaný Jarque-Bera-Urzúa testem. Na Obrázku 1 je rovněž pro srovnání zachycena síla nejsilnějších testů pro sledované alternativy, kterými jsou Uthoffův test a test výběrové špičatosti – viz předchozí výsledky.





Obrázek 7: Komparace síly Jarque-Bera testu, Jarque-Bera-Urzúa testu, robustního Jarque-Bera testu a nejsilnějších testů pro sledované alternativy – testu výběrové špičatosti a Uthoffova testu

Jak je z Obrázku 1 patrné, tak tyto testy ve všech sledovaných případech s výjimkou bimodálního rozdělení pro $\mu_2 = 5$ a $n \geq 80$ vykazují významně vyšší sílu než klasický Jarque-Bera test i jeho modifikace.

4. Závěr

Z provedených Monte Carlo simulací je patrné, že klasický Jarque-Bera test, Jarque-Bera-Urzúa test a robustní Jarque-Bera test vykazují v případě sledovaných lehkochvostých alternativ (uniformní rozdělení, beta(2,2) rozdělení, beta(4,4) rozdělení a tři bimodální rozdělení vzniklá jako směsi dvou normálních rozdělení dané funkcí $F = 0{,}5N(0{,}1) + 0{,}5N(\mu_2, 1)$, kde $\mu_2 \in \{3, 4, 5\}$) především pro malé rozsahy souborů ($n \leq 40$) velmi nízkou sílu (prakticky nulovou sílu), což je způsobeno především kombinací symetrie sledovaného rozdělení a pro malé rozsahy souborů nevýznamného rozdílu mezi špičatostí daného rozdělení a špičatostí normálního rozdělení – tyto testy založené na kombinaci výběrové šikmosti a výběrové špičatosti tudíž nejsou schopny odchylky od normality rozpoznat. Ze srovnání uvedených tří variant Jarque-Bera testu je patrné, že nejvyšší sílu pro všechny sledované lehkochvosté alternativy a všechny sledované rozsahy souborů vykazuje klasický Jarque-Bera test, naopak nejslabším je robustní Jarque-Bera test. Důvodem je zaměření robustního Jarque-Bera testu především na těžkochvosté alternativy, kde vykazuje vyšší sílu testu než klasický Jarque-Bera test – viz Gelová a Gastwirth (2009).

Ještě horší výsledky pro sledované lehkochvosté alternativy vykazuje test výběrové šikmosti (důvodem je symetrie všech sledovaných lehkochvostých alternativ) a rovněž directed SJ test (test je zaměřen především na těžkochvosté alternativy, kde v porovnání s ostatními testy vykazuje velmi vysokou sílu – viz Gelová, Miao a Gastwirth, 2007), které pro všechny sledované lehkochvosté alternativy a všechny sledované rozsahy souborů vykazují prakticky nulovou sílu testu.

Naopak velmi dobré výsledky síly testu vůči sledovaným lehkochvostým alternativám vykazují test výběrové špičatosti, Uthoffův test, Gearyho test, Shapiro-Wilkův test, Anderson-Darlingův test a Cramer-von Misesův test. Zatímco nejsilnějším z nich pro alternativy uniformního, beta(2,2) a beta(4,4) rozdělení je test výběrové špičatosti, tak pro alternativy bimodálního rozdělení je nejsilnějším Uthoffův test. V obou případech se jedná o testy zaměřené na špičatost rozdělení a dokáží tak odhalit i nepatrně nižší špičatost než je špičatost normálního rozdělení. Tyto testy by tak měly být preferovány v případech, kdy lze očekávat rozdělení s nižší špičatostí a krátkými a velmi krátkými chvosty či v případech, kdy lze očekávat bimodální rozdělení.

5. Literatura

[1]GEL, Y. R., GASTWIRTH, J. L. 2008. A robust modification of the Jarque-Bera test of normality. In: Economics Letters, 99(1), 2008, s. 30 – 32.

[2]GEL, Y.R., MIAO, W., GASTWIRTH, J.L. 2007. Robust directed tests of normality against heavy-tailed alternatives. In: Computational Statistics & Data Analysis, 51(5), 2007, s. 2734 – 2746.

[3]JARQUE, C. M., BERA, A. K. 1980. Efficient tests for normality, homoscedasticity and serial independence of regression residuals. In: Economics Letters, 6(3), 1980, s. 255 – 259.

[4]LILLIEFORS, H. 1967. On the Kolmogorov-Smirnov test for normality with mean and variance unknown. In: Journal of the American Statistical Association, 62(318), 1967, s. 399 – 402.

[5]SHAPIRO, S. S., WILK, M. B. 1965. An analysis of variance test for normality (complete samples). In: Biometrika, 52 (3,4), 1965,. s. 591 – 611.

[6]THADEWALD, T., BÜNING, H. 2007. Jarque-Bera test and its competitors for testing normality – a power comparison. In: Journal of Applied Statistics, Taylor and Francis Journals, 34(1), 2007, s. 87 – 105.

[7]THODE, H. C. 2002. Testing for normality. New York: Marcel Dekker, 2002. 368 s. ISBN 978-0824796136.

[8]URZÚA, C.M. 1996. On the correct use of omnibus tests for normality. In: Economics Letters, 53(3), 1996, s. 247 – 251.

Příspěvek vznikl za podpory programu AKTION Česká republika – Rakousko, spolupráce ve vědě a vzdělání, projektu č. 54p21 s názvem „Robust testing for the normality and its applications for the economy, sports and Basel II“.

Adresa autora:

Luboš Střelec, Ing.
ÚSO PEF MZLU v Brně
Zemědělská 1, 613 00 Brno
lubos.strelec@mendelu.cz

Monte Carlo simulace konstant C_1 a C_2 pro RT třídu Jarque-Bera testů normality
Monte Carlo simulations of the constants C_1 and C_2 for the RT class of the Jarque-Bera normality tests

Luboš Střelec

Abstract: The aim of this paper is to estimate two constants for the RT class of the Jarque-Bera normality tests introduced by Střelec and Stehlík (2009b). The RT class contains a vast amount of tests, which can be obtained for different settings of the parameters used in the general form of the RT test statistic. Choosing of the appropriate constants C_1 and C_2 is the hardest aspect of the variants of the RT tests. Among others, we can obtain these constants from the Monte Carlo simulations. Finally, this paper presents the results of the Monte Carlo simulations of the constants C_1 and C_2.

Key words: Monte Carlo simulations, RT class of normality tests, variance, skewness, kurtosis, Jarque-Bera test.

Klíčová slova: Monte Carlo simulace, RT třída testů normality, variance, šikmost, špičatost, Jarque-Bera test.

1. Úvod

Problematika testů normality nabývá v současné době stále více na významu a to především v oblasti analýzy jejich síly v souvislosti s rozvojem výpočetní techniky. Mimo nejvyužívanějších klasických testů, mezi něž lze zařadit především Shapiro-Wilkův test normality (Shapiro a Wilk, 1965) a Jarque-Bera test normality (Jarque a Bera, 1980), je v posledních letech zaměřena pozornost především na robustní testy normality. Mezi robustní testy normality lze zařadit i obecnou třídu RT testů normality (viz Střelec a Stehlík, 2009b), což je skupina testů zobecňující klasický Jarque-Bera test využitím výběrového mediánu a trimmovaného aritmetického průměru na místo výběrového aritmetického průměru ve výběrových charakteristikách šikmosti a špičatosti. Třída RT testů v sobě tak zahrnuje velké množství testů lišící se navzájem právě využitím aritmetického průměru, výběrového mediánu a trimmovaného aritmetického průměru ve výběrových charakteristikách šikmosti a špičatosti v obecné testové statistice třídy RT testů normality. Jedním z problémů je různá variabilita jednotlivých variant charakteristik šikmosti a špičatosti a z toho vyplývající problematika volby konstant C_1 a C_2, které zajišťují, že obecná testová statistika RT má asymptoticky chí-kvadrát rozdělení se dvěma stupni volnosti.

Volba odpovídajících konstant C_1 a C_2 je nejobtížnějším výpočetním aspektem třídy RT testů. Pro získání těchto konstant lze úspěšně využít Monte Carlo simulací. Cílem této práce tak je provést Monte Carlo simulace za účelem získání konstant C_1 a C_2 pro vybrané skupiny testů pocházejících z třídy RT testů normality.

2. Materiál a metodika

Třída RT testů normality je založena na obecné testové statistice RT (viz Střelec a Stehlík, 2009a,b)

$$RT = \frac{k_1(n)}{C_1}\left(\frac{M_{i_1,j_1}^{\alpha_1}}{M_{i_2,j_2}^{\alpha_1}} - K_1\right)^2 + \frac{k_2(n)}{C_2}\left(\frac{M_{i_3,j_3}^{\alpha_3}}{M_{i_4,j_4}^{\alpha_4}} - K_2\right)^2, \qquad (1)$$

kde $M_{i,j}$ je všeobecnou empirickou variantou centrálního momentu náhodné veličiny ve tvaru

$$M_{i,j} = \frac{1}{n}\sum_{m=1}^{n} \varphi_j\left(X_m - M_{(i)}\right), \tag{2}$$

kde $i \in \{0,1,2\}$ označuje použití výběrového aritmetického průměru $M_{(0)} = \overline{X}$, použití výběrového mediánu $M_{(1)} = M_n$ nebo použití trimmovaného aritmetického průměru $M_{(2)} = \frac{1}{n-2s}\sum_{i=s+1}^{n-s} X_{i:n}$ pro pořadovou statistiku $X_{1:n} < X_{2:n} < \ldots < X_{n:n}$, přičemž φ_j pro $j \in \{0,1,2,3,4\}$ označuje typ proměnlivé spojité funkce následovně: $\varphi_0(x) = |x|$, $\varphi_1(x) = x$, $\varphi_2(x) = x^2$, $\varphi_3(x) = x^3$ a $\varphi_4(x) = x^4$.

Třída RT testů pak obsahuje širokou škálu testů závislou především na volbě jednotlivých parametrů *i*, *j*, α. V rámci třídy RT testů normality tak lze najít i vybrané všeobecně známé testy normality pro konkrétní volbu parametrů – např. Jarque-Bera test, Gearyho test, Uthoffův test, test výběrové šikmosti, test výběrové špičatosti apod. (blíže viz Střelec a Stehlík, 2009b,c). Konkrétně např. Jarque-Bera test normality je jeden z testů RT třídy pro následující volbu parametrů: $k_1(n) = n$, $k_2(n) = n$, $C_1 = 6$, $C_2 = 18$, $\alpha_1 = 1$, $\alpha_2 = 3/2$, $\alpha_3 = 1$, $\alpha_4 = 2$, $i_1 = i_2 = i_3 = i_4 = 0$, $j_1 = 3$, $j_2 = 2$, $j_3 = 4$, $j_4 = 2$, $K_1 = 0$ a $K_2 = 3$. Testovou statistiku klasického Jarque-Bera testu pak lze zapsat ve tvaru

$$JB = \frac{n}{6}\left(\frac{M_{0,3}}{M_{0,2}^{3/2}}\right)^2 + \frac{n}{18}\left(\frac{M_{0,4}}{M_{0,2}^{2}} - 3\right)^2. \tag{3}$$

V rámci obecné třídy RT testů pak lze vymezit i skupinu RT_{JB} testů normality, což je skupina testů založená na obecné testové statistice RT_{JB}

$$RT_{JB} = \frac{n}{C_1}\left(\frac{M_{i_1,3}}{M_{i_2,2}^{3/2}}\right)^2 + \frac{n}{C_2}\left(\frac{M_{i_3,4}}{M_{i_4,2}^{2}} - 3\right)^2, \tag{4}$$

tj. jedná se o skupinu testů normality pocházející z obecné RT třídy testů normality pro následující volbu parametrů obecné RT testové statistiky: $k_1(n) = n$, $k_2(n) = n$, $\alpha_1 = 1$, $\alpha_2 = 3/2$, $\alpha_3 = 1$, $\alpha_4 = 2$, $j_1 = 3$, $j_2 = 2$, $j_3 = 4$, $j_4 = 2$, $K_1 = 0$ a $K_2 = 3$, přičemž jednotlivé testy v rámci této třídy se liší nastavením parametrů i_1, i_2, i_3, i_4 pro $i \in \{0,1,2\}$.

Jak již bylo uvedeno výše, tak volba odpovídajících konstant C_1 a C_2 je nejobtížnějším výpočetním aspektem třídy RT testů. Pro získání těchto konstant lze úspěšně využít Monte Carlo simulací. Jelikož třída RT testů zahrnuje velké množství různých testů, tak pozornost bude zaměřena pouze na skupinu testů třídy RT_{JB} testů, tj. cílem je na základě Monte Carlo simulací o 100000 replikací získat konstanty C_1 a C_2 pro výběrové charakteristiky šikmosti a špičatosti uvedené v Tab. 1.

3. Výsledky simulace konstant C_1 a C_2

Konstanty C_1 a C_2 vycházejí z rozdělení sledovaných charakteristik šikmosti a špičatosti. V Tab. 2 až 5 tak jsou pro sledované charakteristiky šikmosti a špičatosti uvedeny empirické i asymptotické střední hodnoty a variance. Následně jsou uvedeny výsledky simulací konstant C_1 a C_2 (viz Tab. 6 a 7).

Jak je patrné z Tab. 2 a 3, tak střední hodnoty všech sledovaných charakteristik šikmosti konvergují k nulové střední hodnotě a střední hodnoty všech sledovaných charakteristik

špičatosti konvergují k hodnotě 3, přičemž je patrná rychlejší konvergence střední hodnoty charakteristik šikmosti, kde již pro rozsah souboru $n = 1000$ je empirická střední hodnota jednotlivých variant šikmostí rovna asymptotické hodnotě.

Tabulka 11: Označení výběrových charakteristik šikmosti a špičatosti

$SK = M_{i_1,3} \big/ M_{i_2,2}^{3/2}$	$K = M_{i_3,4} \big/ M_{i_4,2}^{2}$
$SK_1 = M_{0,3} \big/ M_{0,2}^{3/2}$	$K_1 = M_{0,4} \big/ M_{0,2}^{2}$
$SK_2 = M_{0,3} \big/ M_{1,2}^{3/2}$	$K_2 = M_{0,4} \big/ M_{1,2}^{2}$
$SK_3 = M_{1,3} \big/ M_{0,2}^{3/2}$	$K_3 = M_{1,4} \big/ M_{0,2}^{2}$
$SK_4 = M_{1,3} \big/ M_{1,2}^{3/2}$	$K_4 = M_{1,4} \big/ M_{1,2}^{2}$
$SK_5 = M_{0,3} \big/ M_{2,2}^{3/2}(s=1)$	$K_5 = M_{0,4} \big/ M_{2,2}^{2}(s=1)$
$SK_6 = M_{2,3} \big/ M_{0,2}^{3/2}(s=1)$	$K_6 = M_{2,4} \big/ M_{0,2}^{2}(s=1)$
$SK_7 = M_{2,3} \big/ M_{2,2}^{3/2}(s=1)$	$K_7 = M_{2,4} \big/ M_{2,2}^{2}(s=1)$
$SK_8 = M_{0,3} \big/ M_{2,2}^{3/2}(s=5)$	$K_8 = M_{0,4} \big/ M_{2,2}^{2}(s=5)$
$SK_9 = M_{2,3} \big/ M_{0,2}^{3/2}(s=5)$	$K_9 = M_{2,4} \big/ M_{0,2}^{2}(s=5)$
$SK_{10} = M_{2,3} \big/ M_{2,2}^{3/2}(s=5)$	$K_{10} = M_{2,4} \big/ M_{2,2}^{2}(s=5)$

Pozn.: parametr *s* je parametr trimmovaného aritmetického průměru (viz výše).

Tabulka 12: Simulace střední hodnoty pro různé varianty šikmosti

SK	n							
	25	50	100	200	500	1000	5000	$n \to \infty$
SK_1	-0,001	-0,002	0,001	0,000	0,000	0,000	0,000	0
SK_2	-0,001	-0,002	0,001	0,000	0,000	0,000	0,000	0
SK_3	-0,003	-0,004	0,001	0,001	-0,001	0,000	0,000	0
SK_4	-0,003	-0,003	0,001	0,001	0,000	0,000	0,000	0
SK_5	-0,001	-0,002	0,001	0,000	0,000	0,000	0,000	0
SK_6	-0,002	-0,003	0,001	0,000	0,000	0,000	0,000	0
SK_7	-0,002	-0,003	0,001	0,000	0,000	0,000	0,000	0
SK_8	-0,001	-0,002	0,001	0,000	0,000	0,000	0,000	0
SK_9	-0,002	-0,003	0,001	0,000	0,000	0,000	0,000	0
SK_{10}	-0,002	-0,003	0,001	0,000	0,000	0,000	0,000	0

Pozn.: počet Monte Carlo simulací 100000. **Zdroj:** vlastní simulace

Podobně lze analyzovat vývoj konvergence variance sledovaných charakteristik šikmosti a špičatosti (viz Tab. 4 a 5). Charakteristikami šikmosti s největší variabilitou jsou šikmosti *SK*3 a *SK*4, které mají asymptotické rozdělení s nulovou střední hodnotou a rozptylem 18/*n*. Ostatní sledované charakteristiky šikmosti mají asymptotickou hodnotu rozptylu 6/*n*, přičemž nejpomaleji konvergují hodnoty pro *SK*9 a *SK*10.

Rychlost konvergence variability charakteristik špičatosti je všech sledovaných případech přibližně stejná, přičemž všechny mají asymptotické rozdělení se střední hodnotou 3 a rozptylem 24/*n*.

Tabulka 13: Simulace střední hodnoty pro různé varianty špičatosti

K	n							
	25	50	100	200	500	1000	5000	$n \to \infty$
K_1	2,768	2,883	2,939	2,971	2,988	2,994	2,999	3
K_2	2,656	2,824	2,907	2,955	2,982	2,990	2,998	3
K_3	3,054	3,024	3,011	3,008	3,003	3,001	3,000	3
K_4	2,911	2,958	2,978	2,991	2,996	2,997	2,999	3
K_5	2,763	2,882	2,939	2,971	2,988	2,994	2,999	3
K_6	2,815	2,898	2,944	2,973	2,989	2,994	2,999	3
K_7	2,810	2,897	2,943	2,973	2,989	2,994	2,999	3
K_8	2,734	2,875	2,937	2,971	2,988	2,993	2,999	3
K_9	2,918	2,932	2,955	2,976	2,989	2,994	2,999	3
K_{10}	2,880	2,925	2,953	2,976	2,989	2,994	2,999	3

Pozn.: počet Monte Carlo simulací 100000. **Zdroj:** vlastní simulace

Tabulka 14: Simulace variance pro různé varianty šikmosti

SK	n							
	25	50	100	200	500	1000	5000	$n \to \infty$
SK_1	0,188	0,107	0,057	0,029	0,012	0,006	0,001	$6/n$
SK_2	0,172	0,102	0,055	0,029	0,012	0,006	0,001	$6/n$
SK_3	0,629	0,321	0,166	0,085	0,034	0,017	0,003	$18/n$
SK_4	0,543	0,298	0,160	0,083	0,034	0,017	0,003	$18/n$
SK_5	0,187	0,106	0,057	0,029	0,012	0,006	0,001	$6/n$
SK_6	0,260	0,129	0,064	0,031	0,012	0,006	0,001	$6/n$
SK_7	0,258	0,129	0,064	0,031	0,012	0,006	0,001	$6/n$
SK_8	0,181	0,106	0,057	0,029	0,012	0,006	0,001	$6/n$
SK_9	0,415	0,181	0,081	0,037	0,013	0,006	0,001	$6/n$
SK_{10}	0,397	0,179	0,081	0,037	0,013	0,006	0,001	$6/n$

Pozn.: počet Monte Carlo simulací 100000. **Zdroj:** vlastní simulace

S variabilitou jednotlivých charakteristik šikmosti a špičatosti pak přímo souvisejí hodnoty konstant C_1 a C_2 (viz Tab. 6 a 7). Nejpomaleji konvergují hodnoty konstanty C_1 pro charakteristiky šikmosti *SK*3 a *SK*4, což je způsobeno využitím výběrového mediánu v empirickém centrálním momentu v čitateli charakteristiky šikmosti. Pro většinu charakteristik šikmosti a špičatosti je patrná shoda empirických a teoretických hodnot konstant C_1 a C_2 pro rozsahy souborů $n \geq 500$.

4. Závěr

Z výsledků provedených Monte Carlo simulací je patrné, že asymptotické hodnoty konstanty C_2 jsou pro všechny sledované varianty výběrové charakteristiky špičatosti shodné, tj. všechny sledované varianty charakteristiky špičatosti mají asymptotickou varianci $24/n$ a konstanta C_2 je tedy rovna hodnotě 24. V případě sledovaných variant výběrových charakteristik šikmosti již výsledky tak jednotné nejsou. Většina výběrových charakteristik šikmosti má asymptotickou varianci $6/n$, tj. konstanta C_1 je pro tyto varianty rovna hodnotě 6. Výjimku tvoří výběrové charakteristiky šikmosti *SK*3 a *SK*4, které mají asymptotickou varianci $18/n$, tj. konstanta C_1 je pro tyto varianty rovna hodnotě 18.

Tabulka 15: Simulace variance pro různé varianty špičatosti

K	n							
	25	50	100	200	500	1000	5000	$n \to \infty$
K_1	0,530	0,355	0,204	0,112	0,047	0,024	0,005	$24/n$
K_2	0,523	0,351	0,203	0,111	0,047	0,024	0,005	$24/n$
K_3	0,781	0,429	0,225	0,117	0,048	0,024	0,005	$24/n$
K_4	0,624	0,388	0,214	0,114	0,047	0,024	0,005	$24/n$
K_5	0,520	0,353	0,204	0,112	0,047	0,024	0,005	$24/n$
K_6	0,611	0,373	0,208	0,112	0,047	0,024	0,005	$24/n$
K_7	0,598	0,372	0,208	0,112	0,047	0,024	0,005	$24/n$
K_8	0,505	0,349	0,204	0,112	0,047	0,024	0,005	$24/n$
K_9	0,688	0,394	0,213	0,113	0,047	0,024	0,005	$24/n$
K_{10}	0,633	0,386	0,212	0,113	0,047	0,024	0,005	$24/n$

Pozn.: počet Monte Carlo simulací 100000. **Zdroj:** vlastní simulace

Tabulka 16: Simulace konstanty C_1 pro různé varianty šikmosti

SK	n							
	25	50	100	200	500	1000	5000	$n \to \infty$
SK_1	4,71	5,33	5,66	5,84	5,89	5,94	5,97	6
SK_2	4,31	5,08	5,52	5,76	5,86	5,92	5,97	6
SK_3	15,72	16,07	16,63	16,90	16,95	17,05	17,06	18
SK_4	13,56	14,92	15,98	16,56	16,81	16,98	17,05	18
SK_5	4,68	5,32	5,66	5,84	5,89	5,94	5,97	6
SK_6	6,50	6,47	6,37	6,26	6,10	6,05	6,00	6
SK_7	6,45	6,46	6,37	6,26	6,10	6,05	6,00	6
SK_8	4,52	5,28	5,65	5,83	5,89	5,94	5,97	6
SK_9	10,37	9,05	8,07	7,34	6,66	6,39	6,10	6
SK_{10}	9,92	8,96	8,05	7,34	6,66	6,39	6,10	6

Pozn.: počet Monte Carlo simulací 100000. **Zdroj:** vlastní simulace

Tabulka 17: Simulace konstanty C_2 pro různé varianty špičatosti

K	n							
	25	50	100	200	500	1000	5000	$n \to \infty$
K_1	13,26	17,75	20,45	22,33	23,38	23,71	23,80	24
K_2	13,08	17,54	20,33	22,26	23,38	23,70	23,79	24
K_3	19,51	21,46	22,51	23,42	23,77	23,92	23,83	24
K_4	15,60	19,39	21,42	22,86	23,57	23,81	23,81	24
K_5	13,00	17,67	20,43	22,33	23,38	23,71	23,80	24
K_6	15,28	18,67	20,81	22,47	23,42	23,73	23,80	24
K_7	14,96	18,58	20,79	22,46	23,41	23,72	23,80	24
K_8	12,64	17,47	20,36	22,31	23,38	23,71	23,80	24
K_9	17,20	19,71	21,30	22,67	23,47	23,75	23,80	24
K_{10}	15,83	19,32	21,19	22,65	23,47	23,74	23,80	24

Pozn.: počet Monte Carlo simulací 100000. **Zdroj:** vlastní simulace

5. Literatura

[1]JARQUE, C. M., BERA, A. K. 1980. Efficient tests for normality, homoscedasticity and serial independence of regression residuals. In: Economics Letters, 6(3), 1980, s. 255 – 259.

[2]SHAPIRO, S. S., WILK, M. B. 1965. An analysis of variance test for normality (complete samples). In: Biometrika, 52 (3,4), 1965, s. 591 – 611.

[3]STŘELEC, L. STEHLÍK. M. 2009a. On robust testing for normality. In: Proceedings of the 6th St. Petersburg Workshop on Simulation. St. Petersburg: VVM com. Ltd., Ed. by S.M.Ermakov, V.B. Melas and A.N.Pepelyshev. 2009, s. 755 – 760.

[4]STŘELEC, L. STEHLÍK. M. 2009b. Some properties of robust tests for normality and their applications. IFAS Research report, 2009.

[5]STŘELEC, L. STEHLÍK. M. 2009c. On robust testing for normality. 11.8.2009 submitted to Computational Statistics and Data Analysis.

Příspěvek vznikl za podpory programu AKTION Česká republika – Rakousko, spolupráce ve vědě a vzdělání, projektu č. 54p21 s názvem „Robust testing for the normality and its applications for the economy, sports and Basel II“.

Adresa autora:

Luboš Střelec, Ing.
ÚSO PEF MZLU v Brně
Zemědělská 1, 613 00 Brno
lubos.strelec@mendelu.cz

Diskrétna transformácia 3D dát
Discrete transformation of 3D data

Imrich Szabó

Abstract: The goal of this article is to introduce the world of transformation of 3D data. Its main part is dedicated to four-point transformation, especially to its 3D version. A definition of this transformation will be made here, along with some of its properties (interaction with power functions and polynomials) and plans of its future usage. Also the 2D version will be discussed in this article, due to its status of being the entry point of construction of the 3D four-point transformation. There will be some information about the inverse transformation and IZA representation of polynomials mentioned marginally in the article.

Keywords: function transformation, power functions, polynomial degree reduction, piecewise approximation

Kľúčové slová: transformácia funkcií, mocninové funkcie, zníženie stupňa polynómu, úseková aproximácia

1. Úvod

Štúdiom problematiky transformácie 3D dát sa zistilo, že v danej oblasti podnikol významné kroky ruský bádateľ N. D. Dikoussar, ktorý zadefinoval *diskrétnu projektívnu transformáciu (DPT)*, používajúcu určitý počet referenčných bodov. Konkrétne ich počet bol stanovený na tri.

Podobnou cestou sa vydali aj autori práce [2], ktorí ukázali, že podobná transformácia sa dá vykonať aj s dvomi referenčnými bodmi. Túto svoju myšlienku pretavili do zaujímavých výsledkov, nezávisle od toho, či hovoríme o rovine (krivky/funkcie jednej premennej), alebo priestore (plochy/funkcie dvoch premenných).

Na základe týchto výsledkov sa dospelo k záveru, že by sa tieto transformácie mohli dať nejakým spôsobom zovšeobecniť a mohli by tak vzniknúť transformácie pracujúce s ľubovoľným počtom referenčných bodov. No neostalo to len pri úvahách. Konkrétne výsledky v súvislosti s touto témou produkoval Cs. Török, ktorý vo svojej, zatiaľ ešte nepublikovanej, práci [4] uvádza tzv. *r-bodovú transformáciu*, pre $r \geq 2$, platnú pre jednorozmerné útvary v rovine, presnejšie krivky.

Obsah ďalších kapitol bude venovaný konceptu štvorbodovej transformácie. Druhá kapitola nám priblíži štvorbodovú transformáciu v rovine, jej definíciu, princíp činnosti a ďalšie teoretické úvahy s ňou spojené. V podobnom duchu sa nesie aj najdôležitejšia kapitola 3, obsahom ktorej je štvorbodová transformácia v priestore, používajúca pri svojej činnosti dovedna 16 referenčných bodov. Článok uzavrie štvrtá kapitola o zámeroch využitia transformácie v budúcnosti. V tejto súvislosti sa spomenie potreba vytvoriť inverznú podobu transformácie, z ktorej by mala byť odvodená forma reprezentácie polynómov (IZA) vhodná pre následné vytvorenie aproximačného modelu.

2. Štvorbodová transformácia v rovine

Pôvod výsledkov uvedených v tejto kapitole je potrebné hľadať v práci [2]. Dôvodom ich uvedenia na tomto mieste je skutočnosť, že nasledujúca tretia kapitola sa o spomínané výsledky opiera.

Definícia 1

Dopredná štvorbodová transformácia ľubovoľnej spojitej funkcie jednej premennej *f(x)*, založená na štyroch referenčných bodoch $\mathcal{R} \equiv \mathcal{R}_v \equiv \mathcal{R}_4(v) = \{[v_i, f(v_i)]; i = 0,\ldots,3\}$, je definovaná nasledujúcim vzťahom

$$T_{R_v} f(x) = H_{R_v,0}(x) f(x) + \sum_{i=1}^{3} H_{R_v,i}(x) f(v_i),$$

kde $H_{R_v,i}(x)$, pre $i = 0,\ldots,3$, sú definované nasledovne:

$$H_{R_v,0}(x) = \frac{(v_0 - v_1)(v_0 - v_2)(v_0 - v_3)}{(x - v_1)(x - v_2)(x - v_3)},$$

$$H_{R_v,1}(x) = \frac{(v_0 - x)(v_0 - v_2)(v_0 - v_3)}{(v_1 - x)(v_1 - v_2)(v_1 - v_3)},$$

$$H_{R_v,2}(x) = \frac{(v_0 - x)(v_0 - v_1)(v_0 - v_3)}{(v_2 - x)(v_2 - v_1)(v_2 - v_3)},$$

$$H_{R_v,3}(x) = \frac{(v_0 - x)(v_0 - v_1)(v_0 - v_2)}{(v_3 - x)(v_3 - v_1)(v_3 - v_2)}.$$

Poznámka 1

Namiesto $T_{R_v} f(x)$ budeme často písať $T_{4,v} f(x)$.

Lemma 1 (Štvorbodová 2D transformácia mocninovej funkcie)

Nech je daná mocninová funkcia jednej premennej $f(x) = x^p$ pre $p \in Z^{0+}$. Potom štvorbodová 2D transformácia danej funkcie bude

a) konštanta pre $p < 4$ a platí:

$$T_{4,v} x^p = v_0^p,$$

b) polynóm stupňa $p-3$ pre $p \geq 4$ a platí:

$$T_{4,v} x^p = v_0^p + z_v(x)\, S^v_{p-4,4},$$

kde

$$z_v(x) = (x - v_0) \prod_{i=1}^{3} (v_0 - v_i), \tag{1}$$

a $S^v_{p-4,4}$ sa vypočíta na základe rekurzívneho vzťahu

$$S^v_{j,k} = S^v_{j,k-1} + v_k S^v_{j-1,k}, \tag{2}$$

pre $1 \le j$; $1 \le k \le 4$, pričom $S^{v}_{0,k} = 1$, $k \ge 1$, $S^{v}_{j,0} = v_0^j$, $j \ge 1$ a $v_4 \equiv x$.

Príklad 1
Nech $p = 6$. Potom výsledok štvorbodovej transformácie funkcie $f(x) = x^p$ bude:

$$v_0^6 + (x - v_0)(v_0 - v_1)(v_0 - v_2)(v_0 - v_3)\,(v_0^2 + v_1(v_0 + v_1) + v_2(v_0 + v_1 + v_2)$$
$$+ v_3(v_0 + v_1 + v_2 + v_3) + x(v_0 + v_1 + v_2 + v_3 + x))$$

Veta 1 (Štvorbodová 2D transformácia polynómu)
Nech $p \in Z^{0+}$. Potom štvorbodová transformácia polynómu $P_p(x) = a_0 + a_1x + a_2x^2 + \ldots + a_px^p$ bude

a) konštanta pre $p < 4$ a platí:

$$T_4P_p(x) = P_p(v_0),$$

b) polynóm stupňa $p-3$ pre $p \ge 4$ a platí:

$$T_4P_p(x) = P_p(v_0) + z_v(x)\,A_{p-4,4},$$

kde $z_v(x)$ je definované vo vzťahu (1) a

$$A_{p-4,4} = \sum_{i=4}^{p} S^{v}_{i-4,4}\,a_i.$$

3. Štvorbodová transformácia v priestore

Táto kapitola je zovšeobecnením výsledkov uvedených v práci [2] na 4x4 referenčné body a opiera sa tiež o informácie z predchádzajúcej kapitoly.

Definícia 2
Doprednú 4x4-bodovú 3D transformáciu ľubovoľnej spojitej funkcie dvoch premenných F(x,y), založenú na šestnástich referenčných bodoch $\mathcal{R}_{\mu,\nu} = \{[\mu_i, \nu_j, F(\mu_i, \nu_j)], i = 0,\ldots,3, j = 0,\ldots,3\}$ definujeme vzťahom

$$T_{R_{\mu,\nu}}F(x, y) = H_{00}F(x, y) + \sum_{i=1}^{3} H_{i,0}F(\mu_i, y) + \sum_{j=1}^{3} H_{0,j}F(x, \nu_j) + \sum_{i=1}^{3}\sum_{j=1}^{3} H_{i,j}F(\mu_i, \nu_j),$$

kde $H_{i,j} = H_{R_{\mu,i}}(x)\; H_{R_{\nu,j}}(y)$ (hodnoty $H_{R_{\mu,i}}(x)$, resp. $H_{R_{\nu,j}}(y)$ získame spôsobom uvedeným v definícii 1) a $x \ne \mu_i$, $y \ne \nu_j$, $\mu_i \ne \mu_j$, $\nu_i \ne \nu_j$ pre $i, j = 0,...,3$.
Namiesto $T_{R_{\mu,\nu}}F(x, y)$ budeme často písať $T_{4,4,\mu,\nu}F(x, y)$.

Poznámka 2

Dá sa ukázať, že $\sum_{i=0}^{3}\sum_{j=0}^{3} H_{i,j} = 1.$

Nasleduje lemma, ktorá hovorí o dôležitej vlastnosti štvorbodovej transformácie.

Lemma 2
Dopredná 4x4-bodová 3D transformácia ľubovoľnej spojitej funkcie F(x,y) disponuje vlastnosťou linearity. Teda pre ľubovoľné konštanty c_1 a c_2 platí:

$$T_{4,4,\mu,\nu}(c_1\ F(x,y)+c_2)=c_1 T_{4,4,\mu,\nu}F(x,y)+c_2$$

Dôkaz
Z definície 2 dostávame:

$$T_{4,4,\mu,\nu}(c_1F(x,y)+c_2)=H_{00}(c_1F(x,y)+c_2)+\sum_{i=1}^{3}H_{i0}(c_1F(\mu_i,y)+c_2)+\sum_{j=1}^{3}H_{0j}(c_1F(x,\nu_j)+c_2)$$

$$+\sum_{i=1}^{3}\sum_{j=1}^{3}H_{ij}(c_1F(\mu_i,\nu_j)+c_2)=c_1[H_{00}F(x,y)+\sum_{i=1}^{3}H_{i0}F(\mu_i,y)+\sum_{j=1}^{3}H_{0j}F(x,\nu_j)$$

$$+\sum_{i=1}^{3}\sum_{j=1}^{3}H_{ij}F(\mu_i,\nu_j)\ ]+c_2\sum_{i=0}^{3}\sum_{j=0}^{3}H_{ij}.$$

Po aplikácii rovnosti z poznámky 2 je náš dôkaz kompletný.

■

Predpokladajme, že $F(x,y) = f(x)\ g(y)$. Potom z rovnosti $H_{i,j} = H_{R_{\mu,i}}(x)\ H_{R_{\nu,j}}(y)$
a z definície 2 nám vyplýva nasledujúca lemma

Lemma 3
Nech $F(x,y) = f(x)\ g(y)$. Potom platí:

$$T_{4,4,\mu,\nu}F(x,y)=T_{4,\mu}f(x)\ T_{4,\nu}g(y).$$

Lemma 4 (4x4-bodová 3D transformácia mocninovej funkcie)
Nech je daná mocninová funkcia dvoch premenných $F(x,y) = x^p y^q,\ p,q \in Z^{0+}$. Potom 4x4-bodová 3D transformácia danej funkcie bude

a) konštanta pre $p,q < 4$ a platí:

$$T_{4,4,\mu,\nu}x^p y^q=\mu_0^p\nu_0^q,$$

b) pre $p \geq 4,\ q < 4$:

$$T_{4,4,\mu,\nu}x^p y^q=\mu_0^p\nu_0^q+z_\mu(x)S^{\mu}_{p\text{-}4,4}\nu_0^q,$$

c) pre $p < 4,\ q \geq 4$:

$$T_{4,4,\mu,\nu}x^p y^q=\mu_0^p\nu_0^q+z_\nu(y)S^{\nu}_{q\text{-}4,4}\mu_0^p,$$

d) pre $p \geq 4,\ q \geq 4$:

$$T_{4,4,\mu,\nu}x^p y^q=\mu_0^p\nu_0^q+z_\nu(y)S^{\nu}_{q\text{-}4,4}\mu_0^p+z_\mu(x)S^{\mu}_{p\text{-}4,4}\nu_0^q+z_\mu(x)\,z_\nu(y)\,S^{\mu}_{p\text{-}4,4}\,S^{\nu}_{q\text{-}4,4},$$

kde výpočet $z_\mu(x)$ a $z_\nu(y)$ je uvedený vo vzťahu (1) a $S^{\mu}_{p\text{-}4,4}$, $S^{\nu}_{q\text{-}4,4}$ sa počítajú pomocou vzťahu (2) .

Dôkaz
Podľa lemmy 1 vytvoríme vzťahy pre x^p, resp y^q. Dostávame teda:

- pre $p < 4$: $\mathrm{T}_{4,\mu}\mathrm{x}^{\mathrm{p}} = \mu_0^{\mathrm{p}}$,
- pre $p \geq 4$: $\mathrm{T}_{4,\mu}\mathrm{x}^{\mathrm{p}} = \mu_0^{\mathrm{p}} + \mathrm{z}_{\mu}(\mathrm{x})\ \mathrm{S}^{\mu}_{\mathrm{p\text{-}4,4}}$,
- pre $q < 4$: $\mathrm{T}_{4,\nu}\mathrm{y}^{\mathrm{q}} = \nu_0^{\mathrm{q}}$,
- pre $q \geq 4$: $\mathrm{T}_{4,\nu}\mathrm{y}^{\mathrm{q}} = \nu_0^{\mathrm{q}} + \mathrm{z}_{\nu}(\mathrm{y})\ \mathrm{S}^{\nu}_{\mathrm{q\text{-}4,4}}$.

Lemma 3 pre tento prípad hovorí, že $\mathrm{T}_{4,4,\mu,\nu}\mathrm{x}^{\mathrm{p}}\mathrm{y}^{\mathrm{q}} = \mathrm{T}_{4,\mu}\mathrm{x}^{\mathrm{p}}\ \mathrm{T}_{4,\nu}\mathrm{y}^{\mathrm{q}}$. Pri jej aplikácii máme tieto možnosti:

1. *p, q* < 4. V tom prípade $\mathrm{T}_{4,4,\mu,\nu}\mathrm{x}^{\mathrm{p}}\mathrm{y}^{\mathrm{q}} = \mu_0^{\mathrm{p}}\ \nu_0^{\mathrm{q}}$, čím sme pokryli prípad **a)** lemmy 4.
2. $p \geq 4$, $q < 4$. Potom
   $\mathrm{T}_{4,4,\mu,\nu}\mathrm{x}^{\mathrm{p}}\mathrm{y}^{\mathrm{q}} = (\mu_0^{\mathrm{p}} + \mathrm{z}_{\mu}(\mathrm{x})\ \mathrm{S}^{\mu}_{\mathrm{p\text{-}4,4}})\ \nu_0^{\mathrm{q}} = \mu_0^{\mathrm{p}}\nu_0^{\mathrm{q}} + \mathrm{z}_{\mu}(\mathrm{x})\ \mathrm{S}^{\mu}_{\mathrm{p\text{-}4,4}}\nu_0^{\mathrm{q}}$. Pokrytý je prípad **b)**.
3. $p < 4$, $q \geq 4$. Potom
   $\mathrm{T}_{4,4,\mu,\nu}\mathrm{x}^{\mathrm{p}}\mathrm{y}^{\mathrm{q}} = \mu_0^{\mathrm{p}}(\nu_0^{\mathrm{q}} + \ \mathrm{z}_{\nu}(\mathrm{y})\ \mathrm{S}^{\nu}_{\mathrm{q\text{-}4,4}})\ = \mu_0^{\mathrm{p}}\nu_0^{\mathrm{q}} + \mathrm{z}_{\nu}(\mathrm{y})\ \mathrm{S}^{\nu}_{\mathrm{q\text{-}4,4}}\mu_0^{\mathrm{p}}$. Pokrytý je prípad **c)**.
4. *p, q* ≥ 4. Potom $\mathrm{T}_{4,4,\mu,\nu}\mathrm{x}^{\mathrm{p}}\mathrm{y}^{\mathrm{q}} = (\mu_0^{\mathrm{p}} + \mathrm{z}_{\mu}(\mathrm{x})\ \mathrm{S}^{\mu}_{\mathrm{p\text{-}4,4}})\ (\nu_0^{\mathrm{q}} + \mathrm{z}_{\nu}(\mathrm{y})\ \mathrm{S}^{\nu}_{\mathrm{q\text{-}4,4}})$.
   Roznásobením zátvoriek dostávame vzťah pokrývajúci prípad **d)** a teda náš dôkaz je kompletný.

Veta 2 (4x4-bodová 3D transformácia polynómu)

Nech $p, q \in Z^{0+}$. Potom 4x4-bodová 3D transformácia polynómu $P_{pq}(x, y) = \sum_{i=0}^{p}\sum_{j=0}^{q} a_{ij}\, x^i\, y^j$ bude

a) konštanta pre *p, q* < 4 a platí:
$$\mathrm{T}_{4,4,\mu,\nu}\mathrm{P}_{\mathrm{pq}}(\mathrm{x},\mathrm{y}) = P_{pq}(\mu_0, \nu_0),$$

b) pre $p \geq 4$, $q < 4$:
$$\mathrm{T}_{4,4,\mu,\nu}\mathrm{P}_{\mathrm{pq}}(\mathrm{x},\mathrm{y}) = P_{pq}(\mu_0, \nu_0) + \mathrm{z}_{\mu}(\mathrm{x})\sum_{\mathrm{i=4}}^{\mathrm{p}}\sum_{\mathrm{j=0}}^{\mathrm{q}} a_{ij}\nu_0^j S^{\mu}_{i-4,4},$$

c) pre $p < 4$, $q \geq 4$:
$$\mathrm{T}_{4,4,\mu,\nu}\mathrm{P}_{\mathrm{pq}}(\mathrm{x},\mathrm{y}) = P_{pq}(\mu_0, \nu_0) + \mathrm{z}_{\nu}(\mathrm{y})\sum_{\mathrm{i=0}}^{\mathrm{p}}\sum_{\mathrm{j=4}}^{\mathrm{q}} a_{ij}\mu_0^i S^{\nu}_{j-4,4},$$

d) pre $p \geq 4$, $q \geq 4$:

$$\mathrm{T}_{4,4,\mu,\nu}\mathrm{P}_{\mathrm{pq}}(\mathrm{x},\mathrm{y}) = P_{pq}(\mu_0,\nu_0) + \mathrm{z}_\mu(\mathrm{x})\sum_{\mathrm{i}=4}^{\mathrm{p}}\sum_{\mathrm{j}=0}^{\mathrm{q}} a_{ij}\nu_0^j S_{i-4,4}^{\mu} + \mathrm{z}_\nu(\mathrm{y})\sum_{\mathrm{i}=0}^{\mathrm{p}}\sum_{\mathrm{j}=4}^{\mathrm{q}} a_{ij}\mu_0^i S_{j-4,4}^{\nu}$$

$$+ \mathrm{z}_\mu(\mathrm{x})\,\mathrm{z}_\nu(\mathrm{y})\sum_{\mathrm{i}=4}^{\mathrm{p}}\sum_{\mathrm{j}=4}^{\mathrm{q}} a_{ij}\mathrm{S}_{\mathrm{p\text{-}4,4}}^{\mu}\,\mathrm{S}_{\mathrm{q\text{-}4,4}}^{\nu},$$

kde výpočet $\mathrm{z}_\mu(\mathrm{x})$ a $\mathrm{z}_\nu(\mathrm{y})$ je uvedený vo vzťahu (1) a $\mathrm{S}_{\mathrm{p\text{-}4,4}}^{\mu}$, $\mathrm{S}_{\mathrm{q\text{-}4,4}}^{\nu}$ sa počítajú pomocou vzťahu (2) .

Dôkaz
Majme polynóm v tvare uvedenom v predpoklade vety. Potom na základe liem 2, 4 dostávame

$$\mathrm{T}_{4,4,\mu,\nu}\mathrm{P}_{\mathrm{pq}}(\mathrm{x},\mathrm{y}) = \mathrm{T}_{4,4,\mu,\nu}\sum_{i=0}^{p}\sum_{j=0}^{q} a_{ij}x^i y^j =^{L2} \sum_{i=0}^{p}\sum_{j=0}^{q} a_{ij}\mathrm{T}_{4,4,\mu,\nu}x^i y^j$$

$$=^{L4} \sum_{i=0}^{3}\sum_{j=0}^{3} a_{ij}\mu_0^i\nu_0^j + \sum_{i=4}^{p}\sum_{j=0}^{q} a_{ij}(\mu_0^i\nu_0^j + z_\mu(x)S_{i-4,4}^{\mu}\nu_0^j) + \sum_{i=0}^{p}\sum_{j=4}^{q} a_{ij}(\mu_0^i\nu_0^j + z_\nu(y)S_{j-4,4}^{\nu}\mu_0^i)$$

$$+ \sum_{i=4}^{p}\sum_{j=4}^{q} a_{ij}(\mu_0^i\nu_0^j + z_\mu(x)z_\nu(y)S_{i-4,4}^{\mu}S_{j-4,4}^{\nu})$$

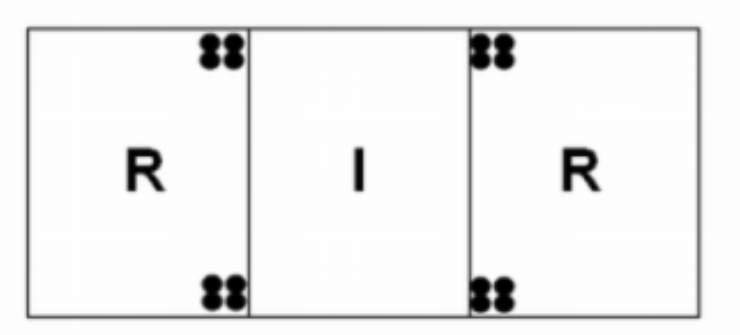
$$= P_{pq}(\mu_0,\nu_0) + \mathrm{z}_\mu(\mathrm{x})\sum_{\mathrm{i}=4}^{\mathrm{p}}\sum_{\mathrm{j}=0}^{\mathrm{q}} a_{ij}\nu_0^j S_{i-4,4}^{\mu} + \mathrm{z}_\nu(\mathrm{y})\sum_{\mathrm{i}=0}^{\mathrm{p}}\sum_{\mathrm{j}=4}^{\mathrm{q}} a_{ij}\mu_0^i S_{j-4,4}^{\nu}$$

$$+ \mathrm{z}_\mu(\mathrm{x})\,\mathrm{z}_\nu(\mathrm{y})\sum_{\mathrm{i}=4}^{\mathrm{p}}\sum_{\mathrm{j}=4}^{\mathrm{q}} a_{ij}\mathrm{S}_{\mathrm{p\text{-}4,4}}^{\mu}\,\mathrm{S}_{\mathrm{q\text{-}4,4}}^{\nu}.$$

4. Aktivity do budúcnosti, záver

V tejto časti článku sa zmienime o možnostiach využitia 4x4 bodovej transformácie pre vyhladzovanie - aproximáciu. No pred tým, než by sme pristúpili k detailom, venujme trochu pozornosti myšlienke prínosu tohto článku, ktorým je opis prechodu z roviny do priestoru. Kým autori zdrojov, z ktorých sa pri tvorbe čerpalo, sa väčšinou venujú transformáciám funkcií jednej premennej používajúcim dva, respektíve tri referenčné body, výsledkom mojej práce je transformácia funkcií dvoch premenných používajúca 4 referenčné body v oboch smeroch. A prízvuk je predovšetkým na posledných slovách, pretože, ako sa to už v úvode spomína, autor zatiaľ ešte nepublikovaného článku [4] uvádza transformáciu pomocou *r* referenčných bodov, ale iba v rovine. Mojou doménou by mal byť predovšetkým trojrozmerný priestor.

Nadväzujúc na prvú vetu tejto kapitoly o vyhladzovaní – aproximácii poukážeme teraz na známu skutočnosť, že klasická regresia (MNŠ) nie je dostatočná na vyhladzovanie zašumených dát s komplexnou štruktúrou a je potrebné použiť efektívnejšie spôsoby, medzi ktoré patrí *úseková aproximácia*. Podstatou tejto techniky, z nášho uhla pohľadu (viď práce [2] a [4]), je rozdelenie



Obr. 1: Schéma z troch častí R-I-R a 16 referenčných bodov

aproximovanej oblasti na segmenty (menšie časti znižujúce zložitosť štruktúry) dvoch základných typov - **R** a **I**. Rozdelenie na segmenty musí byť striedavé, tak ako to ilustruje aj obrázok 1. V tomto konkrétnom prípade sa na krajné **R**-segmenty nezávisle aplikuje klasická regresia (metóda najmenších štvorcov) a stredný **I**-segment treba aproximovať tak, aby sa zabezpečil hladký prechod medzi susednými segmentmi. Je dôležité pritom vytvoriť inverznú podobu vyššie spomenutej transformácie, z ktorej bude následne odvodená IZA reprezentácia polynómov. Pomocou tejto reprezentácie a tiež vhodného výberu (lokalizácie) referenčných bodov (čierne body na obrázku 1) sa navrhne model na aproximáciu stredného segmentu. Pod vhodným výberom referenčných bodov sa rozumie ich umiestnenie do tesnej blízkosti hraníc **R**-segmentov s **I**-segmentom.

Iná schéma (schéma z dvoch častí) je použitá v *sekvenčnom* modeli, v ktorom sa na začiatku nachádza jeden **R**-segment a všetky nasledujúce sú **I**-segmenty. Tieto je potrebné aproximovať pomocou IZA reprezentácie. Ako už z názvu vyplýva, udeje sa to postupne, kedy sa prvý **I**-segment vypočíta pomocou počiatočného **R**-segmentu, druhý **I**-segment podľa prvého **I**-segmentu a tak ďalej.

Rozdelenie aproximovanej oblasti na segmenty dvoch typov **R** a **I** umožňuje okrem už spomenutých schém vytvárať rôzne ďalšie schémy. Predmetom ďalšej práce budú schémy šachovnicového typu na aproximáciu 3D dát zložitej štruktúry pomocou polynómov.

5. Literatúra

[1] Dikoussar N. D., Function Parametrization by Using 4-Point Transforms,Comp. Phys. Commun., 1997

[2] Matejčíková A., Török Cs.,doc. RNDr. CSc., Two-point Transformation and Polynomials, Košice, 2007

[3] Révayová M., Török Cs.,doc. RNDr. CSc., Piecewise approximation and neural networks, Kybernetika 43, 2007

[4] Török Cs.,doc. RNDr. CSc., Reference Points Based Transformation and Aproximation, UPJŠ Košice, 2008 (to be published)

Adresa autora:

Imrich Szabó, Mgr.
Prírodovedecká fakulta UPJŠ v Košiciach, Ústav informatiky
Jesenná 5, 040 01 Košice
imrich.szabo@gmail.com

Analýza ekonomických softvérov aktuálne ponúkaných softvérovými spoločnosťami na slovenskom trhu
Analysis of the Economic Software currently offered by software cooperations in the Slovak market

Mária Szabóová

Abstract: The contribution describes the economic software, as a comprehensive system of corporate agendas, from each supplier in the Slovak market. Current overview of larger and smaller producers and suppliers of the software existing in the Slovak market is given in first part. Second part includes more details about individual products offered by software companies and the risk of financial fraud.

Key words: economic software, modules, versions

Kľúčové slová: ekonomický softvér, moduly, verzie

1. Úvod

Používanie informačného systému predstavuje nielen veľmi moderný, ale aj efektívny spôsob využívania špičkových technológií, ktorý má za následok zvýšenie výnosov, zvýšenie produktivity, zníženie nákladov na administratívu, logistiku, lepšie spracovanie zákaziek a iné. V súčasnosti je ekonomický softvér ponúkaný viacerými softvérovými spoločnosťami, a preto výber vhodného softvéru za prijateľnú cenu predstavuje často zložitý proces. Na druhej strane počítačové spracovanie informácií znižuje možnosť výskytu podvodov. V dnešnej dobe sa podvody často vyskytujú v podnikateľskej sfére hlavne z dôvodu stále pretrvávajúcej finančnej krízy.

2. Výrobcovia a dodávatelia software existujúci na slovenskom trhu

Produkty softvérových spoločností sú v dnešnej dobe určené pre riadenie podnikových procesov malých, stredných, ale aj veľkých firiem. Sú prispôsobené potrebám jednotlivých používateľov pre zabezpečenie komplexného spracovania firemnej agendy u príslušných podnikateľských subjektov s cieľom poskytnúť užívateľom štandardné výstupy.

Tabuľka 1: Cieľové zameranie vybraných slovenských softvérových spoločností

Vybrané slovenské softvérové spoločnosti a produkty	malé firmy	stredné firmy	veľké firmy
A3Soft*	x		x
ABRA Software*	x		x
Aurus	x	x	x
Cígler software*	x	x	x
Datalan	x		
Exalogic	x	x	
KROS	x	x	x
MADO	x	x	x
RERO Software	x	x	x
Stormware*	x	x	
SUNSOFT	x	x	
Tangram a.s.		x	x

Zdroj: vlastné vypracovanie

* spoločnosť ponúkajúca produkty na trhu v Slovenskej republike aj Českej republike

3. Charakteristika jednotlivých produktov ponúkaných softvérovými spoločnosťami a riziko finančných podvodov

A3 Soft dodáva informačné systémy pre maloobchodné ako aj veľkoobchodné firmy. Prednosťou ponúkaného ekonomického softvéru je široká variabilita použitia. Avšak pri jeho kúpe sa musí počítať s vyššou vstupnou investíciou.

Produkty:

a) A3 ACCOUNT - účtovný software A3 Account zabezpečuje údaje a analyzuje podklady pre potreby operatívneho a strategického riadenia firiem.

Moduly A3 Account:

-homebanking, banka, evidencia bankových príkazov, pokladňa, devízový účet, valutová pokladňa, odberateľské faktúry, dodávateľské faktúry, účtovné doklady, daňové účtovné doklady, zálohové faktúry, vzájomné zápočty, elektronické bankovníctvo pre zahraničné banky, doklady v cudzích menách podľa kurzového lístku, tuzemské doklady v cudzej mene, penalizácia, upomienky.

b) A3 POS je zameraný na zabezpečenie jednoduchého spôsobu predaja na pokladniach v malých firmách.

c) A3 STORE predstavuje skladový systém zameraný na malé a veľké podnikateľské subjekty.

Softvérová spoločnosť ***ABRA Software*** sa zameriava na poskytovanie systémov na riadenie firemných procesov a ekonomický softvér na trhu.

Produkty:

a) ABRA G1 - toto softvérové riešenie je vhodné pre živnostníkov, ktorý účtujú v sústave jednoduchého účtovníctva a malé podnikateľské subjekty.

b) ABRA G2 - ekonomický softvér ABRA G2 je vhodný pre podnikateľské subjekty účtujúce v sústave podvojného účtovníctva.

c) ABRA G3

d) ABRA G4

ABRA G3 a ABRA G4 sú určené pre veľké podniky, pretože sa vyznačuje dobrými vlastnosťami pri používaní veľkým počtom užívateľov.

Moduly ABRA Software:

- predaj, nákup, skladové hospodárstvo, banka a homebanking, pokladňa, majetok, podvojné účtovníctvo, jednoduché účtovníctvo, call-centrum, mzdy a personalistika, maloobchodný predaj, splátkový predaj, evidencia pošty, polohované sklady, kompletizácia, projektová dokumentácia, výroba, business Intelligence, SCM, skriptovanie.

Výhodou produktov ABRA G1 až G4 je predovšetkým stabilita, bezpečnosť a otvorenosť. Vďaka svojej pružnosti a prispôsobivosti dokážu ponúknuť riešenia všetkých požiadaviek kladených na moderné informačné systémy. Všetky moduly majú jednoduché a veľmi ľahké ovládanie.

Produkt spoločnosti ***AURUS*** sa špecializuje predovšetkým na oblasť pôdohospodárstva a potravinárstva so zameraním na malé, stredné, veľké podniky, rozpočtové a príspevkové organizácie.

Produkt:

EKOPACKET obsahuje stabilné a rýchle ekonomické programy s jednotným dizajnom jednotlivých modulov a veľkou údajovou základňou a vysokou prispôsobivosťou špecifickým požiadavkám jednotlivých užívateľov. Charakteristickým znakom tohto softvérového riešenia je vzájomná uzávierková nezávislosť príslušných podsystémov.

Moduly EKOPACKET:

- správca, účtovníctvo, fakturácia, pokladňa, sklad, dlhodobý majetok, drobný majetok, utility, dane, banka, výroba, odbyt, mzdy, personalistika, doplnkové dôchodkové poistenie, e-commerce.

Prednosťou uvedeného produktu je mohutná údajové základňa, ktorá sa užívateľovi prispôsobí tak, že ho pri zadávaní vstupov a čítaní výstupov neobťažujú "nadbytočné" údaje, ale keď sa stanú potrebnými, sú rýchlo dostupné. Jednotlivé moduly majú jednotný dizajn**,** jednotné ergonomické ovládanie a previazanosť údajov tak, že každý sa zadáva maximálne raz. Široká a prepracovaná konfigurovateľnosť tohto produktu zabezpečuje vysokú nezávislosť jednotlivých podsystémov na zmenách legislatívy a vysokú prispôsobiteľnosť špecifickým požiadavkám užívateľov.

Ekonomický software ponúkaný spoločnosťou ***Cíger software*** je zameraný na vedenie účtovníctva a riadenia malých a stredných firiem.

Produkty:

a) Money S3 - softvérové riešenie obsahuje široký rozsah modulov s množstvom nadštandardných funkcií a jednoduchou obsluhou. Patrí medzi najrozšírenejšie systémy pre malé a stredné firmy. Umožňuje jednoduchú a účelnú prácu s dátami, s ich vyhľadávaním, triedením, spracovaním alebo tlačou.

b) Money S5 - predstavuje najpoužívanejší informačný systém na našom trhu.

Moduly Money:

- adresár, účtovníctvo, majetok, fakturácia, sklady, katalógy, cenníky, objednávky.

Výhodou produktov je, že sa vyznačujú najmä jednoduchosťou a prehľadnosťou ovládania pre ekonómov, obchodníkov, ako aj manažérov. Ponúkajú veľké množstvo obvyklých aj nadštandardných funkcií, ktoré na každom kroku uľahčujú prácu užívateľa.

Spoločnosť ***Datalan*** ponúka množstvo služieb v oblasti softvérových aplikácii a inovatívnych podnikových riešených firiem. Softvér spracováva firemné agendy podnikateľských subjektov a malých podnikov.

Produkt:

MEMPHIS je určený pre začínajúce firmy povinné viesť podvojné účtovníctvo s možnosťou kedykoľvek produkty modifikovať podľa aktuálnych požiadaviek. Poskytuje prehľad o majetku, jeho zaradení, zaúčtovaní, stave odpisovania a upozorňuje na vyradenie odpísaného majetku z evidencie. Je základom kompletného firemného informačného systému zostaveného z modulov, ktoré je možné jednoducho rozšíriť a navzájom kombinovať. Hlavnými výhodami produktu sú možnosť kombinácie s inými ERP systémami, prehľadné reporty o stave firmy a známe užívateľské prostredie MS Office.

Moduly MEMPHIS:

- financie, fakturácia, obchod, logistika, majetok, účtovníctvo, výroba, personalistika, správa dokumentov, CRM zabezpečujúci prehľad o všetkých činnostiach súvisiacich s obchodnou činnosťou firmy.

Spoločnosť ***Exalogic*** sa zaoberá predovšetkým tvorbou účtovného softvéru OBERON pre malé a stredné podniky z rôznych oblastí podnikania, pričom dôraz kladie na jednoduchosť a logickosť. Produkt OBERON je určený pre malé a stredne veľké firmy. Prednosťou tohto produktu je stabilita v jednoužívateľskej aj sieťovej prevádzke, sledovanie legislatívnych zmien, bezplatný hot-line a technická podpora, možnosť požiadať o vzdialenú pomoc prostredníctvom internetu, nové verzie na stiahnutie na internete.

KROS patrí medzi 3 najvýznamnejšie spoločnosti na slovenskom trhu s ekonomickým softvérom. Hlavnou prednosťou tejto spoločnosti je, že patrí medzi najúspešnejšie slovenské

softvérové firmy zaoberajúce sa zaoberá vývojom a predajom ekonomického softvéru. Ako popredný slovenský výrobca ekonomického softvéru získal ocenenie „Jednotka na trhu“. Spoločnosť dosiahla ku koncu prvého štvrťroka 2009 počet 54 tisíc zákazníkov. Prezídium Nadácie Slovak Gold po dôslednom zhodnotení funkcií a vlastností produktov rozhodlo o udelení prestížneho ocenenia kvality – medaily Slovak Gold a ocenenie Grand Prix Slovak Gold. Toto ocenenie získala ako prvý softvér na Slovensku.

Produkty:

a) ALFA – je určený pre vedenie ekonomickej agendy jednoduchého účtovníctva firmy. Predstavuje profesionálny účtovný softvér s jednoduchým a praktickým ovládaním. Moduly: účtovníctvo, sklad, fakturácia, kniha jázd, pokladnice, kalendár úloh, manažérske funkcie, homebanking, evidencia majetku a pošty, daňové priznanie FO a iné.

b) OLYMP - je profesionálny nástroj určený na výpočet miezd. Vďaka praktickému ovládaniu je práca s ním jednoduchá a pohodlná. Silnou stránkou programu je veľký počet výstupných zostáv. Samozrejmosťou sú výkazy odovzdávané do zdravotných poisťovní, Sociálnej poisťovne ako aj na Daňový úrad, pričom tieto výkazy je možné tlačiť aj do originálnych tlačív alebo zasielať elektronicky. Moduly: personálny modul, mzdový modul, tlačový a exportný modul, dochádzkový modul.

c) OMEGA - predstavuje komplexné riešenie v oblasti vedenia podvojného účtovníctva, skladového hospodárstva, fakturácie, dlhodobého a krátkodobého majetku, cestovných príkazov, evidencie jázd a ostatných podporných evidencií. Dôraz je kladený na maximálne zjednodušenie práce užívateľa a samozrejme možnosť akýchkoľvek spätných opráv v jednotlivých moduloch programu. Moduly: podvojné účtovníctvo, DPH, daňové priznanie PO + účtovné poznámky, CRM (riadenie vzťahov so zákazníkmi), sklady a skladové karty, výrobné čísla a šarže, makrokarty, fakturácia, objednávky a rezervácie, majetok, kniha jázd a iné.

MADO sa zameriava na vývoj ekonomického softvéru, vedenie účtovníctva a ekonomické poradenstvo. Svoje produkty poskytuje nielen podnikateľským subjektom, ale aj organizáciám nezriadeným za účelom podnikania, t. j. pre neziskové organizácie. Výhodou produktov je, že pozostávajú z viacerých podsystémov čo vytvára pre organizácie možnosť zostaviť si ekonomický softvér podľa vlastných potrieb, pričom každý z uvedených podsystémov je možné využívať aj samostatne.

Produkty ponúkané spoločnosťou MADO:

a) MADO PROFIT - určený pre podnikateľské subjekty.

Moduly: podvojné účtovníctvo, fakturácie, jednoduché účtovníctvo, skladové hospodárstvo, evidencia majetku.

b) MADO NONPROFIT – určený pre neziskové organizácie, rozpočtové a príspevkové organizácie.

Moduly: podvojné účtovníctvo, fakturácie, jednoduché účtovníctvo, skladové hospodárstvo, evidencia majetku.

RERO software s.r.o. poskytuje služby v oblasti projektovania, vývoja, implementácie a údržby podnikových informačných systémov. Produkt poskytovaný spoločnosťou RERO je REPO ekonomika vhodný pre výrobné podniky, spoločnosti zaoberajúce sa obchodom, účtovnícke domy, logistické spoločnosti a pod. Výhodou softvéru je prispôsobiteľnosť tak malým organizáciám, ako aj organizáciám so zložitou organizačnou a geografickou štruktúrou.

Moduly: podvojné účtovníctvo a saldokonto, jednoduché účtovníctvo, majetok, kalkulácie, pokladňa, sklady materiálu, fakturácia a sklady tovaru, voľná fakturácia, výroba, personalistika a mzdy, a iné.

Softvérová spoločnosť ***Stormware*** vyrába a dodáva informačné systémy pre platformu MS Windows. Ponúka ekonomický software POHODA ponúka pre firmy jednoduché aj podvojné účtovníctvo a je vhodná pre platiteľov aj neplatiteľov DPH. Jeho výhodou je, že patrí medzi obľúbený a často využívaný nástroj na vedenie účtovníctva a riadenie ekonomiky malých, stredne veľkých a väčších firiem. Obsahuje širokú škálu nadštandardných funkcií, poskytuje profesionálny vzhľad firemných dokumentov a podrobný prehľad o vlastnom hospodárení. Vďaka prepracovanému užívateľskému rozhraniu, interaktívnym sprievodcom a rozsiahlemu systému kontextového pomocníka uľahčuje svojmu užívateľovi každodennú prácu. Nevýhodou je, že okrem nákladov na nasadenie systému, je potrebné vynaložiť dodatočné náklady na zaškolenie zamestnancov.

Moduly: adresár, jednoduché účtovníctvo, fakturácie, sklady, mzdy, majetok, tlač, homebanking, e-shop, eform, XML.

SUNSOFT sa zaoberá vývojom ekonomického softvéru a predajom výpočtovej a kancelárskej techniky a služieb. Produkt EcoSun je plne modulárny ekonomický software určený pre lokálne aj sieťové spracovanie s podporou výmeny dát medzi vzdialenými pracoviskami.

Moduly: podvojné účtovníctvo PU, obchodná a skladový informačný systém, personalistika a mzdy, informačný systém pre manažérov, riadenie vzťahov so zákazníkmi, pokladničný systém a jednoduché účtovníctvo.

Hlavné výhody systému EcoSun sú dokladovo riadené základné funkcie, minimalizácia duplicitných vstupov, okamžitá dostupnosť všetkých potrebných informácií na jednom mieste, užívateľská prívetivosť a jednoduchosť obsluhy a automatika na požiadanie.

Spoločnosť ***Tangram*** ponúka ekonomický softvér Tangram určený pre stredné a väčšie firmy aj pre firmy s viacerými navzájom prepojenými pobočkami. Vyznačuje sa vysokým komfortom obsluhy, jednoduchým ovládaním a nízkymi nárokmi na zaškolenie obsluhy. Moduly sú navzájom prepojené do jedného celku, účtovanie dokladov je automatizované pomocou šablón a predkontačných tabuliek.

Moduly : podvojné účtovníctvo, financie, vydané faktúry, prijaté faktúry, saldokonto, DPH, majetok, evidencia obchodných partnerov.

V dnešnej dobe sa v podnikateľskej sfére často vyskytujú podvody hlavne z dôvodu stále pretrvávajúcej finančnej krízy. Kabát, L., Stehlíková, B., Tirpáková, A. v článku Identifikácia zamestnaneckých finančných podvodov [1] uvádzajú, že k vážnym problémom v oblasti finančných podvodov patria opakované a vysoko frekventované pokusy o podvodné konanie zo strany zamestnancov firiem a iných organizačných štruktúr v snahe získať osobný prospech na úkor vlastnej organizácie, alebo na úkor veriteľov, ktorý tejto organizácii zverili do spravovania svoje osobné zdroje. Cieľom finančného podvodu v podnikateľskom prostredí je poskytovať zavádzajúce a mylné informácie odberateľom resp. ostatným účastníkom trhu o výsledkoch, ktoré podnik dosahuje. Ide teda o úmyselné klamanie dôsledkom ktorého sú finančné straty a iné negatívne následky týchto nepravdivých informácií. Jedným zo spôsobov zníženia výskytu podvodných konaní v rámci podnikateľského prostredia je počítačové spracovanie informácií, a teda zavádzanie efektívnych softvérových produktov do riadenia firemných procesov.

4. Záver

Cieľom implementácie ekonomického softvéru je predovšetkým zníženie nákladov a vytvorenie lepšej podpory pre rozhodovanie. Z hľadiska požiadaviek trhu do popredia vystupujú najmä efektivita, racionalizácia a optimalizácia tých firemných procesov, ktoré sú nevyhnutné pre existenciu daného podnikateľského subjektu na trhu. Najväčším problémom pri výbere vhodného ekonomického softvéru je nedostatočný prehľad podnikateľov o situácii na trhu s ekonomickým softvérom. Najvýznamnejšiu úlohu v procese výberu zohráva šikovnosť predajcu a poskytovanie informácii podnikateľských subjektom. Okrem samotného prevedenia ekonomického softvéru a konzultácií týkajúcich sa jeho výberu sú rozhodujúce servisné služby, bezplatné zaškolenia používateľov a telefonická podpora. Zohľadnenie služieb uľahčujúcich prechod na nový ekonomický softvér zohráva tiež dôležitú úlohu pri jeho výbere. Vo všeobecnosti využívanie ekonomických softvérov výrazne uľahčuje svojim užívateľom prácu a umožňuje rýchlejšie, kvalitnejšie a bezchybné spracovanie údajov. Sú vytvárané s dôrazom kladeným nielen na maximálne zjednodušenie práce používateľa, ale predovšetkým aj na možnosť akýchkoľvek spätných opráv v jednotlivých moduloch programu. Uvedené kritériá v plnej miere spĺňa predovšetkým spoločnosť KROS, ako jedna z najvýznamnejších spoločností na slovenskom trhu, ktorej produkty sú považované za najpredávanejšie na Slovensku už od roku 2004. Keďže produkty ponúkané touto spoločnosťou preukázali svoju nadštandardnú kvalitu, počet zákazníkov spoločnosti má neustále rastúcu tendenciu. Z vyššie uvedenej komparácie produktov vybraných softvérových spoločností vyplýva, že produkty spoločnosti KROS patria nielen medzi najpredávanejšie na slovenskom trhu, ale zároveň ide o produkty vysokej kvality so zameraním na širokú škálu zákazníkov.

5. Literatúra

[1] Kabát, L., Stehlíková, B., Tirpáková, A.: Identifikácia zamestnaneckých finančných podvodov. FORUM STATISTICUM SLOVACUM 2/2009. SŠDS Bratislava 2009.
[2] http://www.ekonomickysoftware.sk/vyrobcovia [cit. 2009-10-12]
[3] http://www.a3soft.sk/sk/stranka/software-4 [cit. 2009-10-12]
[4] http://www.abra.sk/html/produkty/vstup-pro-uzivatele-abra-gx.php [cit. 2009-10-09]
[5] http://www.aurus.sk/prod_eko.html [cit. 2009-10-09]
[6] http://www.kros.sk/ [cit. 2009-09-14]
[7] http://www.madosoft.sk/produkty.php [cit. 2009-09-14]
[8] http://www.rerosoftware.sk/REROEkonomika [cit. 2009-09-24]
[9] http://www.softveroveriesenia.sk/ext_files/file_ext_document_27.pdf [cit. 2009-10-09]
[10] http://www.stormware.sk/pohoda/ [cit. 2009-10-12]
[11] http://www.sunsoft.sk/ekonomicky_software_ecosun.asp [cit. 2009-09-14]
[12] http://www.tangram.sk/ [cit. 2009-10-09]

Adresa autora (-ov):

Maria Szabóová, Ing.
externá doktorandka na Katedre financií
Slovenská poľnohospodárska univerzita
Fakulta ekonomiky a manažmentu
Trieda Andreja Hlinku 2, 949 76 Nitra
szaboova.m@gmail.com

Aktuárske výpočty v prostredí programu R
Actuarial calculation in programme R

Alena Tartaľová

Abstract: This paper presents calculation of aggregate claim amount distribution using package *actuar* in statistical system R. We have showed how to use function *aggregateDist* to compute or approximate the *cdf* of the aggregate claim amount distribution for various methods.

Key words: R, software, package actuar, aggregate distribution

Kľúčové slová: R, softvér, balík actuar , rozdelenie celkových škôd

1. Úvod

Programovací jazyk a prostredie programu **R** je vhodné na štatistické výpočty a tvorbu grafických výstupov. Ide o *Open Source* implementáciu jazyka S, ktorý používajú profesionálne štatistické programy. Softvér je voľne šíriteľný v zmysle Všeobecnej verejnej licencie GNU a je to program s otvoreným zdrojovým kódom. Program je podporovaný takmer všetkými bežne používanými operačnými systémami (UNIX, Linux, Windows, MacOS). Okrem bežne používaných štatistických metód , vďaka širokej užívateľskej základni je k dispozícii množstvo prídavných balíkov a možnosť naprogramovania si vlastných procedúr, ktoré jeho schopnosti ešte posilňujú. Preto sa tento program používa vo viacerých oblastiach štatistiky, napr. aj v meraní chudoby [7].

Cieľom príspevku je popísať možnosti programu **R** v aktuárstve. Balík pre aktuárske výpočty je balík **`actuar`**, ktorý bol prvýkrát spustený v roku 2005 a stále sa pracuje na jeho zlepšovaní. Umožňuje použitie programu v **R** v troch hlavných oblastiach poistnej matematiky: Rozdelenie individuálnej výšky škôd (Loss Distribution), Teória rizika (Risk theory) a Teória kredibility (Credibility theory). V príspevku sa venujeme použitiu na určenie rozdelenia celkových škôd.

2. Modely kolektívneho rizika

Základom riešenia rozhodujúcich otázok pre poisťovateľa, súvisiacich so stanovením poistného, so spoluúčasťou, zaistením, pravdepodobnosťou krachu a pod., je poznať rozdelenie náhodnej premennej opisujúcej celkovú škodu, ako aj základné charakteristiky rozdelenia celkového (agregátneho) poistného plnenia. Model, ktorý pomáha nastolené aktuálne otázky riešiť sa nazýva kolektívny model rizika.

Definujme náhodnú premennú S ako celkovú škodu vyplývajúcu z konkrétneho portfólia nezávislých poistných zmlúv, ktoré vzniknú v období jedného roka. Nech náhodná premenná N opisuje počet škôd, ktoré v sledovanom období nastanú a náhodná premenná X_i opisuje výšku i-tej škody. Potom celkovú škodu vyjadríme ako súčet všetkých individuálnych škôd čo zapíšeme vzťahom:

$$S = X_1 + X_2 + \cdots + X_N,$$

pričom $S=0$, ak $N=0$. Ďalej musia byť splnené tieto predpoklady:

- $\{X_i\}_{i=1}^{\infty}$ je postupnosť nezávislých, identicky rozdelených náhodných premenných,
- Náhodná premenná N je nezávislá od $\{X_i\}_{i=1}^{\infty}$.

Náhodná premenná S má zložené rozdelenie, čo označujeme

$$S \approx Co\left(p_N(n), F_X(x)\right),$$

kde $p_N(n)$ - zákon rozdelenia počtu škôd, $F_X(x)$ - zákon rozdelenia resp. distribučná funkcia individuálnych výšok škôd.

a. ROZDELENIE CELKOVEJ ŠKODY S A ZÁKLADNÉ CHARAKTERISTIKY

Distribučnú funkciu celkového poistného plnenia $F_S(x)$ vyjadríme ako:

$$F_S(x) = P(S \le x) = P\left(\bigcup_{n=0}^{\infty}(S \le x \wedge N = n)\right) = \sum_{n=0}^{\infty} P(S \le x \cap N = n) = \sum_{n=0}^{\infty} P(S \le x / N = n).P(N = n).$$

Keďže máme fixný počet poistných plnení $\{X_i\}_{i=1}^{n}$, tak

$$P(S \le x / N = n) = P(X_1 + X_2 + \cdots + X_n \le n) = F^{*n}(x),$$

kde $F^{*n}(x)$ je n - násobná konvolúcia distribučných funkcií. Úpravou dostávame:

$$F_S(x) = \sum_{n=0}^{\infty} P(N = n).F^{*n}(x).$$

Výpočet momentov náhodnej premennej S je založené na podmienenej pravdepodobnosti dvoch náhodných premenných, pre ktoré momenty existujú.

Stredná hodnota. Vzhľadom na to, že celková škoda S je závislá od počtu škôd, ktorý je opísaný náhodnou premennou N, platí

$$E(S) = E(E(S / N)),$$

pričom na základe vlastností strednej hodnoty môžeme písať

$E(S / N = n) = E\left(\sum_{i=1}^{n} X_i\right) = E(X_1) + E(X_2) + \ldots + E(X_n) = n.m_1$, z čoho dostávame strednú hodnotu náhodnej premennej S v tvare:

$$E(S) = E(E(S / N)) = E(N.m_1) = E(N).E(X).$$

Teda očakávaná celková škoda sa rovná súčinu očakávaného počtu škôd za jednu časovú sledovanú jednotku a očakávanej individuálnej škody.

Disperzia. Obdobnými úvahami odvodíme vzťah pre disperziu D(S), za predpokladu, že postupnosť náhodných premenných $\{X_i\}_{i=1}^{\infty}$ je postupnosť nezávislých náhodných premenných.

$D(S / N = n) = D\left(\sum_{i=1}^{\infty} X_i\right) = D(X_1) + D(X_2) + \ldots + D(X_n) = n.\left(m_2 - m_1^{\,2}\right)$. Využitím vlastností disperzie dostávame disperziu celkovej škody

$$D(S) = E(N).D(X) + E^2(X).D(N).$$

b. VÝPOČET DISTRIBUČNEJ FUNKCIE CELKOVÉHO POISTNÉHO PLNENIA

V súlade s predchádzajúcou časťou budeme označenie $F_S(x)$ používať pre distribučnú funkciu celkového poistného plnenia S, teda

$$F_S(x) = P(S \le x).$$

Ďalej budeme predpokladať, že poznáme rozdelenie počtu poistných plnení N, aj identické rozdelenie výšky individuálnych poistných plnení X_i, pre $i = 0,1,\ldots,N$, pričom toto spoločné rozdelenie individuálnych poistných plnení je ***diskrétne rozdelenie***, definované pre kladné celé čísla. To znamená, že možné hodnoty individuálnych plnení sú 1,2,3,… a možné hodnoty celkového poistného plnenia S sú 0,1,2,3….

Budeme používať takéto označenie pre spoločnú pravdepodobnostnú funkciu identicky rozdelených individuálnych poistných plnení X_i a funkciu pravdepodobnosti celkového poistného plnenia S:

$$f_k = P(X_i = k) \text{ pre } k = 1,2,3,\ldots$$

$$g_k = P(S = k) \text{ pre } k = 0,1,2,\ldots$$

Predpoklad diskrétnych rozdelení X_i a S sa na prvý pohľad zdá značne obmedzujúci a vyvoláva zdanie, že výsledky budú ťažko aplikovateľné v praxi vzhľadom na spojitosť rozdelení náhodných premenných X_i a S. Tento problém sa dá úspešne vyriešiť tzv. *diskretizáciou*, resp. *preškálovaním* spojitej náhodnej premennej X_i. Postup diskretizácie je podrobne popísaný napr.[5].

Na výpočet distribučnej funkcie celkového poistného plnenia S môžeme použiť niektorú z nasledujúcich metód.

- **Panjerov rekurentný vzorec na výpočet $F_S(x)$**

Rekurentný vzorec, ktorý odvodil H.H. Panjer v roku 1981, môžeme použiť vtedy, ak je splnená podmienka týkajúca sa rozdelenia počtu poistných plnení. Označme:

$$p_N(k) = P(N = k).$$

Predpokladajme, že pravdepodobnostná funkcia $p_N(k)$ môže byť vyjadrená pomocou známeho Panjerovho vzťahu:

$$p_N(k) = \left(a + \frac{b}{k}\right) p_N(k-1), k = 1,2,3,\ldots$$

kde a a b sú konštanty, pričom začiatočná hodnota pre rekurentný výpočet hodnôt náhodnej premennej N opisujúcej počet škôd je hodnota $p_N(0) > 0$.

Potom ak sú individuálne poistné plnenia X_i diskrétne, s hodnotami 1,2,3,... a $f_k = P(X_i = k)$, pravdepodobnosti $g_k = P(S = k)$ môžeme vypočítať podľa Panjerovho rekurentného vzorca:

$$g_0 = p_N(0)$$

$$g_k = \sum_{j=1}^{k} \left(a + \frac{b.j}{k}\right) f_j . g_{k-j}$$

Odvodenie tohto vzorca možno nájsť napr. v[1],[2].

- **Presný výpočet $F_S(x)$ konvolúciou distribučných funkcií**

V predchádzajúcej časti sme odvodili vzťah pre distribučnú funkciu celkového poistného plnenia v tvare:

$$F_S(x) = \sum_{n=0}^{\infty} P(N = n).F^{*n}(x),$$

kde $F^{*n}(x)$ je n - násobná konvolúcia distribučných funkcií individuálnych výšok škôd. Ak individuálne poistné plnenia majú diskrétne rozdelenie, tak n-násobnú konvolúciu definujeme vzťahom:

$$F^{*n}(x) = \begin{cases} I\{x \geq 0\}, k = 0 \\ F(x), \; k = 1 \\ \sum_{y=0}^{x} F^{*(k-1)}(x-y) f(y), \; k = 2,3,\ldots. \end{cases}$$

Vypočítať konvolúciu distribučných funkcií je pre niektoré typy rozdelení veľmi zložité (napr. [6]), a tak bez použitia vhodného počítačového programu je ich aplikácia prakticky nemožná.

- **Aproximácia rozdelenia celkových škôd normálnym rozdelením**

Predpokladajme, že vieme spoľahlivo odhadnúť základné charakteristiky kolektívneho rizika – strednú hodnotu $E(S)=\mu$ a rozptyl $D(S)=\sigma^2$. Pretože $S=\sum_{i=1}^{N} X_i$ je súčtom nezávislých a identicky rozdelených náhodných premenných, podľa centrálnej limitnej vety sa ponúka normálna aproximácia rozdelenia *S*. Teda pre ľubovoľné *s* platí

$$F_S(x)=P(S\le s)=P\left(\frac{S-\mu}{\sigma}\le\frac{s-\mu}{\sigma}\right)\approx\Phi\left(\frac{s-\mu}{\sigma}\right)$$

Čím väčší je počet *N* poistných plnení, tým je aproximácia $F_S(x)$ pomocou distribučnej funkcie normovaného normálneho rozdelenia lepšia.

Menej známou a aplikovanou je aproximácia rozdelenia transformovaným normálnym rozdelením

$$F_S(x)\approx\Phi\left(-\frac{3}{\gamma}+\sqrt{\frac{9}{\gamma^2}+1+\frac{6}{\gamma}\frac{x-\mu}{\sigma}}\right)$$

kde $\gamma=E\left[(S-\mu)^3\right]/\sigma^{3/2}$. Aproximácia sa dá použiť v prípade, ak $x>\mu$ a dáva uspokojujúce výsledky, ak $\gamma<1$ (podrobnejšie pozri v[3]).

Vážnym nedostatkom normálnej aproximácie je to, že hustota normálneho rozdelenia je symetrická a konverguje veľmi rýchlo k nule. *Aproximácia posunutým gama rozdelením* čiastočne odstráni, resp. zníži podhodnotenie pravdepodobnosti vysokých poistných plnení v porovnaní s normálnou aproximáciou (pozri[5]).

- **Určenie rozdelenia celkových škôd využitím simulácie**

Predošlé prístupy sú založené na použití zložitých analytických postupov a značných matematických znalostí. Simulácia hodnôt distribučnej funkcie pomocou metódy Monte Carlo vychádza z priameho napodobňovania reálneho systému (pozri v [1]).

3. Určenie rozdelenia celkových škôd v programe R

Funkcia **`aggregateDist`** z balíka **`actuar`** vypočíta distribučnú funkciu celkových škôd využitím piatich metód. Používame ju v tvare

```
aggregateDist(method = c("recursive", "convolution", "normal",
                         "npower", "simulation"),
              model.freq = NULL, model.sev = NULL, p0 = NULL,
              x.scale = 1, moments, nb.simul, ...,
              tol = 1e-06, maxit = 500, echo = FALSE)
```

Pred použitím tejto metódy je nutné preškálovať rozdelenie individuálnych škôd pomocou funkcie **`discretize`**. Táto funkcia na preškálovanie používa štyri metódy.

```
discretize(cdf, from, to, step = 1,
           method = c("upper", "lower", "rounding", "unbiased"),
           lev, by = step, xlim = NULL)
```

Použitie funkcie **`aggregateDist`** si ilustrujeme na niekoľkých praktických úlohách, pričom cieľom nie je zistiť, ktorý z daných modelov je najlepší, iba ukázať postup výpočtu v programe **R**.

Príklad 1: Predpokladajme, že počet škôd N sa riadi Poissonovým rozdelením s parametrom λ=10, a rozdelenie individuálnych škôd X_i je Gamma(2,1). Našou úlohou je nájsť rozdelenie celkových škôd *S* pomocou Panjerovho rekurentného vzťahu. Najprv preškálujeme gama rozdelenie na intervale (0,22) a krokom 0,5 pomocou funkcie `discretize`:

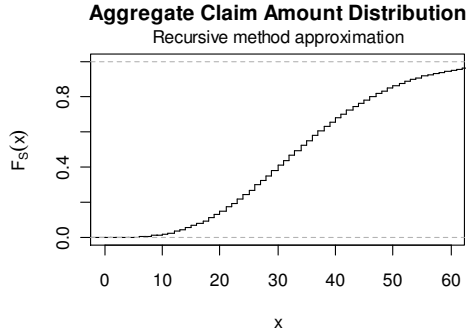
```
>fx <- discretize(pgamma(x,2,1), from=0, to=22, step=0.5, method="upper")
##Panjerov rekurentný vzťah
> Fs <- aggregateDist("recursive", model.freq = "poisson", model.sev = fx,
lambda = 10, x.scale = 1)
# vo funkcii zvolíme metódu výpočtu „recursive“, model.freq = rozdelenie
počtu škôd, model.sev= rozdelenie individuálnych škôd
> summary(Fs)             #výpočet základných charakteristík celkovej škody
Aggregate Claim Amount Empirical CDF:
     Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
  0.00000  24.00000  33.00000  35.00331  43.00000 130.00000
> quantile(Fs)            #kvantily rozdelenia celkovej škody
  25%   50%   75%   90%   95%  97.5%   99%  99.5%
 12.0  16.5  21.5  26.5  30.0  32.5   36.5  39.0
> quantile(Fs, 0.999      # 99.9% kvantil rozdelenia napr. na výpočet VaR
99.9%
 44.5
```

Graf distribučnej funkcie Fs je na Obrázku1. (získame ho funkciou `>plot(Fs)`)

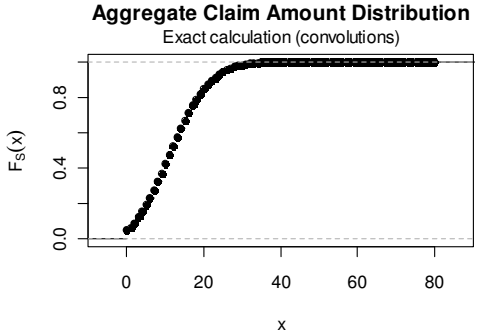
Príklad 2: Určenie rozdelenia celkových škôd pomocou konvolúcie ukážeme na príklade, ktorý možno nájsť v [4], str.522, Example 6.6.

```
##Konvolučná metóda
>fx <- c(0, 0.15, 0.2, 0.25, 0.125, 0.075,0.05, 0.05, 0.05, 0.025, 0.025)
#vektor hodnôt hustoty rozdelenia individuálnych škôd
>pn <- c(0.05, 0.1, 0.15, 0.2, 0.25, 0.15, 0.06, 0.03, 0.01)
#vektor pravdepodobnosti počtu škôd, prvá hodnota musí zodpovedať prípadu
P(N=0)
>Fs <- aggregateDist("convolution", model.freq = pn, model.sev = fx,
x.scale = 1)
> summary(Fs)
Aggregate Claim Amount Empirical CDF:
   Min.  1st Qu.  Median   Mean   3rd Qu.    Max.
   0.00    6.00   11.00   12.58   16.00   80.00
```

Graf distribučnej funkcie Fs je na Obrázku 2.



Obrázok1:Rozdelenie celkových škôd z pr.1



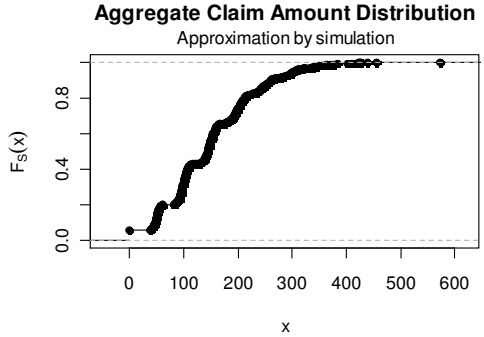
Obrázok2:Rozdelenie celkových škôd z pr.2

Príklad 3: Ďalšou z možností na nájdenie distribučnej funkcie celkových škôd je použitie metódy simulácie.

```
## Simulačná metóda
>model.freq <- expression(data = rpois(3))     #rozdelenie počtu škôd
>model.sev <- expression(data = rgamma(100, 2))#rozdelenie výšky škôd
>Fs <- aggregateDist("simulation", nb.simul = 1000, model.freq, model.sev)
```

```
> summary(Fs)
Aggregate Claim Amount Empirical CDF:
     Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
  0.0   96.66383 148.68182 152.88710 203.73621 460.16133
```

Graf distribučnej funkcie Fs je na Obrázku3.



Obrázok3:Rozdelenie celkových škôd z pr.3

4. Záver

Cieľom príspevku bolo ilustrovať použitie balíka **`actuar`** na určenie rozdelenia celkových škôd. Na konkrétnych príkladoch sme ukázali použitie funkcie **`aggregateDist`** pre rôzne metódy odhadu. Použitím tejto funkcie sa náročné matematické vyjadrenia celkovej výšky škôd dajú jednoduchšie vypočítať.

Literatúra

[1] HORÁKOVÁ, G. 2006. Teória rizika v poistení, Vydavateľstvo EKONÓM. ISBN 80-225-2141-8

[2] DAYKIN, C.D., PENTIKÄINEN, T. and PESONEN, M. 1994. Practical Risk Theory for Actuaries, Chapman & Hall. ISBN 0-412-42850-4

[3] GOULET, V. 2008 An R Package for Actuarial Science, *Journal of Statistical Software*, Volume 25, Issue 7. http://www.jstatsoft.org/v25/i07/paper

[4] KLUGMAN, S. A., PANJER, H. H. and WILLMOT, G. E. 2004, Loss Models, From Data to Decisions, Second Edition, Wiley. ISBN 0471238848

[5] PACÁKOVÁ,V. 2004. Aplikovaná poistná štatistika, Edícia Ekonómia, 2004, 261 s. ISBN 80-8087-004-8

[6] TARTAĽOVÁ, A.2008.Problém stability rozdelenia pri modelovaní celkovej výšky škôd, In: *Forum Statisticum Slovacum*, roč.4, č.7 (2008). ISSN 1336-7420.

[7] ŽELINSKÝ, T. 2009. Odhad vybraných ukazovateľov chudoby a ich štandardných chýb na regionálnej úrovni SR v prostredí R. In: *Forum Statisticum Slovacum*. roč. 5, č. 7 (2009). ISSN 1336-7420.

Tento príspevok je podporovaný grantom VEGA č. 1/0724/08, „Riadenie rizík neživotného poistenia podľa direktívy Európskej komisie SOLVENCY II“.

Adresa autora:

Alena Tartaľová, Mgr.
Technická univerzita v Košiciach, Ekonomická fakulta
Katedra aplikovanej matematiky a hospodárskej informatiky
Nemcovej 32, 040 01 Košice
alena.tartalova@tuke.sk

Úsekové vyhladzovanie pomocou spoločných parametrov
Piecewise smoothing using shared parameters

Csaba Török

Abstract: Thanks to the modification of one part of a recently proposed two-part approximation scheme and leveraging the models' mutual parameters we provide a new approach to piecewise parametric smoothing. Both parts of the scheme are modelled by the IZA representation of polynomials. While the original two-part scheme uses a sequential computation of the models' parameters the new approach computes them at the same time. Consequently the transition between the gained approximants is smoother.

Key words: polynomial model, LS, IZA representation, splines.

Kľúčové slová: polynomiálny model, MNŠ, IZA reprezentácia, splajny.

1. Úvod

Práca sa zaoberá úsekovým vyhladzovaním zašumených dát komplexnej štruktúry, na modelovanie ktorých je polynomiálny model nepostačujúci. Uvažujeme iba dva úseky, ako lokálny model, ale prístup k hladkému prechodu medzi úsekmi, založený na spoločných parametroch, je možné zovšeobecniť na globálnu aproximáciu.

Navrhujeme schémy z dvoch častí na vyhladzovanie zašumených dát pomocou polynómov s hladkým prechodom v ich spoločnom bode zadaných na susedných intervaloch a modelovaných na základe IZA reprezentácie (pozri vetu) polynómov.

Predovšetkým sformulujeme úlohu. Uvažujme nad intervalmi $[u_0, u_1]$, $[u_1, u_2]$ polynómy

$$P(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3, \quad Q(x) = b_0 + b_1 x + b_2 x^2 + b_3 x^3,$$

medzi ktorými predpokladáme C^1 spojitosť

$$P(u_1) = Q(u_1), \qquad P'(u_1) = Q'(u_1)$$

kvôli hladkému prechodu medzi $P(x)$, $Q(x)$ v ich spoločnom uzle u_1 a zašumené dáta

$$\{[x_{i,M}, \mathfrak{y}^*_{i,M}],\ \mathfrak{y}^*_{i,M} \equiv \tilde{P}(x_{i,M}) = P(x_{i,M}) + \varepsilon^*_{i,M},\quad i = \overline{1,M}\},$$

$$\{[x_{i,N}, \mathfrak{y}_{i,N}],\ \mathfrak{y}_{i,N} \equiv \tilde{Q}(x_{i,N}) = Q(x_{i,N}) + \varepsilon_{i,N},\quad i = \overline{1,N}\},$$

kde $\varepsilon^*_{i,M}$ a $\varepsilon_{i,N}$ sú (vzájomne) nekorelované náhodné veličiny so strednou hodnotou 0 a disperziou σ^2. Našim cieľom je odhadnúť polynómy $P(x)$ a $Q(x)$. Poznamenáme, že napr. Hermitove polynómy je možné používať na riešenie danej úlohy, ale štandardne iba sekvenčne a aj vtedy jeho pravý/druhý polynóm je o niečo zložitejší ako náš (pozri schému 1, 2).

Uvedieme dve verzie schémy z dvoch častí *so spoločnými parametrami* v modeloch. V dôsledku reparametrizácie a podobnej reprezentácie polynómov súhrnný počet parametrov v týchto dvoch verziách je rovnaký a menší ako súhrnný počet koeficientov dvoch uvažovaných polynómov. Prvá verzia schémy z dvoch častí vyjadruje dva polynomiálne modely pomocou spoločných funkčných hodnôt ako parametrov. V druhej verzii pravý (druhý) model používa všetky koeficienty - parametre ľavého (prvého) modelu. Napriek tomu, že pravý model druhej verzie je vyjadrený pomocou zvýšeného počtu parametrov, súhrnný počet parametrov dvoch modelov sa rovná súhrnnému počtu parametrov modelov

prvej verzie. Dôležité je, že toto číslo je menšie ako súčet počtu koeficientov dvoch polynómov. Kým prvá verzia zabezpečuje kvázi-hladký prechod medzi dvomi aproximantmi, pozri vzťah (1), druhá verzia garantuje hladký prechod.

Nasledujúca časť uvádza všeobecný vzorec IZA reprezentácie polynómov, ako základ našich schém. Časť 3 prezentuje schémy na vyhladzovanie a nasledujúca ilustruje nový prístup na dátach.

2. IZA reprezentácia polynómov

Pred uvedením schém z dvoch častí na vyhladzovanie sformulujeme *r*-bodovú IZA reprezentáciu polynómov stupňa *p*, $p \geq r \geq 2$, pozri [3] (skratka IZA je z vyjadrenia polynómu v tvare $P = I + ZA$, kde I je *Incomplete interpolation*, Z je *Zeroing polynomial* a polynóm A sa používa pre *Aproximáciu*).

Veta. Nech $p \geq r \geq 2$. Potom ľubovoľný polynóm $P_p(x) = a_0 + a_1 x + ... + a_p x^p$ môže byť vyjadrený pomocou jeho *r* rozličných (referenčných) bodov

$$R \equiv R_v = \{[v_i, y_i],\ y_i = P_p(v_i),\ i = \overline{0, r-1}\},$$

ako

$$\begin{aligned} P_p(x) &= I_R(x) + Z_R(x)\, A_{p-r,r}(x) \\ &= I_R(x) + w_R(x)^T \alpha \end{aligned},$$

kde $I_R(x) = \sum_{i=0}^{r-1} \Pi_i(x)\, y_i$ je neúplný interpolačný polynóm s

$$\Pi_i(x) = \prod_{v \in V_i} \frac{x - v}{v_i - v},\ \ V_i = V \setminus \{v_i\},\ \ V = \{v_0, v_1, \dots, v_{r-1}\},\ i = \overline{0, r-1},$$

a

$$Z_R(x) = \prod_{i=0}^{r-1} (x - v_i),\ \ A_{p-r,\,r}(x) = S^T \cdot \alpha,\qquad \alpha = (a_r, a_{r+1}, \dots, a_p)^T,$$

$$S = (S_{0,r}, S_{1,r}, \dots, S_{p-r,r})^T,\ S_{0,r} = 1,\ r \geq 0,\ \ S_{j,0} = x_0^j,\ j \geq 0,$$

$$S_{j,k} = S_{j,k-1} + v_k\, S_{j-1,k},\ \ j \geq 1,\ k \geq 1,\ \ v_r \equiv x,$$

$$w_R(x) = (w_{0,r}(x),\ \dots,\ w_{p-r,r}(x))^T,\qquad w_{j,r}(x) = Z_R(x) S_{j,r}.$$

Napríklad, kubický polynóm $P(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3$ pomocou dvoch referenčných bodov $R = \{[v_0, P_0], [v_1, P_1]\}$ ležiacich na $P(x)$ môžeme vyjadriť v súlade s reprezentačnou vetou nasledovne

$$P(x) = \frac{(x - v_1)}{(v_0 - v_1)} P(v_0) + \frac{(x - v_0)}{(v_1 - v_0)} P(v_1) + (x - v_0)(x - v_1)(a_2 + (v_0 + v_1 + x) a_3).$$

3. Schémy z dvoch častí

Našim zámerom je porovnať stručne popísané nové schémy v časti 1 s nedávno navrhnutými dvomi schémami. Preto, aby čitateľ nebol odkázaný na ďalšie publikácie, uvedieme tu štyri schémy. Prvá a tretia schéma bola získaná z IZA reprezentácie polynómov [2, 4] a používa **referenčné body**. Druhá a štvrtá schéma boli odvodené z prvej a tretej

schémy. O nový prístup, založený na spoločných parametroch, sa opierajú schéma tri a štyri. Čo sa týka prvej a druhej schémy, odvolávame sa na práce [2], [4].

V prvej a tretej schéme predpokladáme, že

$$P(v_0) = Q(v_0) \quad \text{a} \quad P(v_1) = Q(v_1),$$

kde $v_0 = u_1$, $v_1 = u_1 - \tau$, τ je malé kladné reálne číslo. Tento predpoklad zaručuje kvázi C^1 spojitosť

$$P(u_1) = Q(u_1), \quad P'(u_1) = Q'(u_1) + \tau, \tag{1}$$

tzn. $|P'(u_1) - Q'(u_1)| \leq cO(\tau)$.

1. schéma používa mimo klasického polynomiálneho modelu aj IZA model, vychádzajúci z IZA reprezentácie (pripomíname, že tu $P(v_0) = Q(v_0)$, $P(v_1) = Q(v_1)$)

$$Q(x) = \frac{(x - v_1)}{(v_0 - v_1)} P(v_0) + \frac{(x - v_0)}{(v_1 - v_0)} P(v_1) + (x - v_0)(x - v_1)(b_2 + (v_0 + v_1 + x)b_3). \tag{2}$$

Prvá schéma z dvoch častí je určená modelmi

$$\tilde{Y}^* = X^* a + \varepsilon^*,$$
$$\tilde{Y} = I_R + W\beta + \varepsilon,$$

kde $a = (a_0, a_1, a_2, a_3)^T$, $R = \{[v_i, P(v_i)],\ v_i = u_1 - i\tau,\ i = 0,1\}$, $\beta = (b_2, b_3)^T$ a matica W je skonštruovaná pomocou bázových funkcií $(x - v_0)(x - v_1))$ a $(x - v_0)(x - v_1)(v_0 + v_1 + x))$. Vektor β je vypočítaný z $\tilde{Y} - I_{\hat{R}} = W\beta$ po odhadnutí *a*, kde

$$\hat{R} = \{[v_i, \hat{P}(v_i)],\ v_i = u_1 - i\tau,\ i = 0,1\}.$$

2. schéma používa nasledujúce vyjadrenie Q(x)

$$\mathrm{Q(x)} = a_0 + a_1\mathrm{x} + (-v_0^2 + 2\mathrm{x}v_0)a_2 + (3\mathrm{x}v_0^2 - 2v_0^3)a_3 + (-v_0 + \mathrm{x})^2(b_2 + (\mathrm{x} + 2v_0)b_3),$$

ktoré je získané z (2) vďaka $\tau \to 0$ a zaručuje v spoločnom bode $u_1 \equiv v_0$ polynómov C^1 hladkosť: $P(u_1) = Q(u_1)$, $P'(u_1) = Q'(u_1)$.

Druhá schéma z dvoch častí je zadaná pomocou

$$\tilde{Y}^* = X^* a + \varepsilon^*,$$
$$\tilde{Y} = X_2 a + W\beta + \varepsilon,$$

kde matice X_2 a W_2 sú skonštruované na základe bázových funkcií

$$g_1 = 1,\ g_2 = x,\ g_3 = -v_0^2 + 2\mathrm{x}v_0,\ g_4 = 3\mathrm{x}v_0^2 - 2v_0^3 \ \text{ a } \ g_5 = (x - v_0)^2,\ g_6 = (x - v_0)^2(x + 2v_0). \tag{3}$$

Po odhadnutí a je vektor β vypočítaný zo vzťahu $\tilde{Y} - X_2\hat{a} = W\beta$.

Parametre modelov v prvej a druhej schéme sú vypočítané sekvenčne. Novinkou tretej a štvrtej schémy je to, že oni počítajú parametre naraz, pozri rovnice (4), (5).

3. schéma využíva IZA reprezentáciu ako pre $P(x)$ tak aj pre $Q(x)$

$$P(x)=\frac{(x-v_1)}{(v_0-v_1)}P(v_0)+\frac{(x-v_0)}{(v_1-v_0)}P(v_1)+(x-v_0)(x-v_1)(a_2+(v_0+v_1+x)a_3),$$
$$Q(x)=\frac{(x-v_1)}{(v_0-v_1)}P(v_0)+\frac{(x-v_0)}{(v_1-v_0)}P(v_1)+(x-v_0)(x-v_1)(b_2+(v_0+v_1+x)b_3).$$

Tretia schéma, ako *schéma z dvoch častí so spoločnými parametrami*, je daná modelmi

$$\tilde{Y}^* = I_R + W\alpha + \varepsilon^*,$$
$$\tilde{Y} = I_R + W\beta + \varepsilon,$$

kde $R=\{[v_i,P_i],\ v_i=u_1-i\tau,\ P_i=P(v_i),\ i=0,1\}$, $\alpha=(a_2,a_3)^T$, $\beta=(b_2,b_3)^T$. Parametre P_0, P_1 a prvky vektorov α, β sú odhadnuté pomocou MNŠ naraz z rovnice

$$\tilde{Y}_3 = X_3\ (a_2,a_3,P_0,\ P_1,b_2,b_3)^T, \tag{4}$$

kde $\tilde{Y}_3=\begin{pmatrix}\tilde{Y}^*\\ \tilde{Y}\end{pmatrix}$,

$$X_3=\begin{pmatrix}
f_1(x_{1,M}) & f_2(x_{1,M}) & f_3(x_{1,M}) & f_4(x_{1,M}) & 0 & 0\\
\dots & \dots & \dots & \dots & \dots & \dots\\
f_1(x_{M,M}) & f_2(x_{M,M}) & f_3(x_{M,M}) & f_4(x_{M,M}) & 0 & 0\\
0 & 0 & f_3(x_{1,N}) & f_4(x_{1,N}) & f_5(x_{1,N}) & f_6(x_{1,N})\\
\dots & \dots & \dots & \dots & \dots & \dots\\
0 & 0 & f_3(x_{N,N}) & f_4(x_{N,N}) & f_5(x_{N,N}) & f_6(x_{N,N})
\end{pmatrix}$$

a bázové funkcie f_i, $i=\overline{1,6}$ sú definované vzťahmi

$$f_1(x)\equiv f_5(x)=(x-v_0)(x-v_1),$$
$$f_2(x)\equiv f_6(x)=(x-v_0)(x-v_1)(v_0+v_1+x),$$
$$f_3(x)=\frac{(x-v_1)}{(v_0-v_1)},\ f_4(x)=\frac{(x-v_0)}{(v_1-v_0)}.$$

4. schéma Štvrtá schéma, ako *schéma z dvoch častí so spoločnými parametrami*, používa tie isté modely ako druhá sekvenčná schéma

$$\tilde{Y}^* = X^* a + \varepsilon^*,$$
$$\tilde{Y} = X_2(a^T,b_2,b_3) + \varepsilon,$$

ale a^T,b_2,b_3 sú odhadnuté pomocou MNŠ na základe rovnice

$$\tilde{Y}_4 = X_4\ (a_0,a_1,a_2,a_3,b_2,b_3)^T, \tag{5}$$

tu $\tilde{Y}_4\equiv\tilde{Y}_3$ a bázové funkcie *f.*(x) a *g.* (x) pre X_4 sú zadané pomocou $f_i(x)=x^{i-1}$, $i=\overline{1,4}$ a (3), kde

$$X_4 = \begin{pmatrix} f_1(x_{1,M}) & f_2(x_{1,M}) & f_3(x_{1,M}) & f_4(x_{1,M}) & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ f_1(x_{M,M}) & f_2(x_{M,M}) & f_3(x_{M,M}) & f_4(x_{M,M}) & 0 & 0 \\ g_1(x_{1,N}) & g_2(x_{1,N}) & g_3(x_{1,N}) & g_4(x_{1,N}) & g_5(x_{1,N}) & g_6(x_{1,N}) \\ \dots & \dots & \dots & \dots & \dots & \dots \\ g_1(x_{N,N}) & g_2(x_{N,N}) & g_3(x_{N,N}) & g_4(x_{N,N}) & g_5(x_{N,N}) & g_6(x_{N,N}) \end{pmatrix}$$

Spoločné v daných štyroch schémach je to, že každá z nich

- využíva IZA reprezentáciu polynómov,
- obsahuje rovnaký súhrnný počet parametrov (6), ktorý je menší ako súhrnný počet koeficientov polynómov (8).

Vyjadrenie modelov	**Hladkosť**	Sekvenčne	Naraz	**Výpočet parametrov**	
Poly+ Iza	Kvázi-C[1], τ	a_i + F(v_i), b_i **(sch.1)**	a_i,F(vi) + F(v_i), b_i **(sch.3)**	F(v_i)	**Spoločné parametre**
Iza + Iza	C[1]	a_i + a_i, b_i **(sch.2)**	a_i + a_i, b_i **(sch.4)**	a_i	

Rozdiely je možné sformulovať na základe toho (pozri tabuľku 2),

- ako je ľavý a pravý model vyjadrený
- ako sú vypočítané ich parametre
- či modely používajú τ
- v akej miere je prechod medzi polynómami hladký (kvázi-hladkosť či hladkosť)

a v prípade verzií 3 a 4 aké sú spoločné parametre .

4. Príklad

Uvažujme nad intervalmi [0; 1], [1; 1.9] polynómy (u_1=1)

$P(x) = 0.03 + 0.432x - 1.2x^2 + 0.8x^3$,

$Q(x) = 0.03 + 0.432x - 1.2x^2 + 0.8x^3 - 3(x - 1)^3 = 3.03 - 8.568x + 7.80x^2 - 2.2x^3$

a ich pomocou generované dáta (SNR=1):

x	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
$\tilde{P}(x)$	0.02	0.057	0.079	0.066	0.071	0.058	0.078	0.037	0.001	0.046	0.079

x	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9
$\tilde{Q}(x)$	0.141	0.273	0.479	0.207	0.246	0.144	0.424	0.108	-0.41

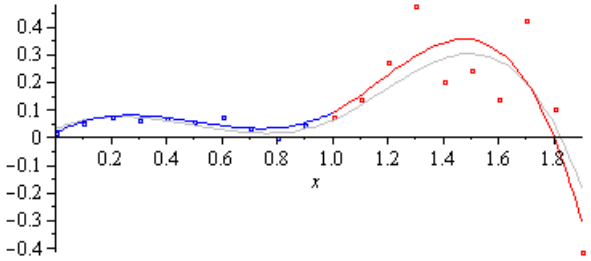
Poznamenávame, že práca [4] obsahuje o jeden bod [1; 0.16] pre $\tilde{Q}(x)$ viac. Z tretej schémy na základe (3) obdržíme odhady (zaokrúhlené na tri desatinné miesta)

$$\hat{P}(x) = 0.502x - 0.411 + (x - 1)(x - 0.999)(0.428 + 0.896x),$$

$$\hat{Q}(x) = 0.502x - 0.411 + (x - 1)(x - 0.999)(4.159 - 2.737).$$

Obrázok 2 znázorňuje body, polynómy a ich odhady $\hat{P}(x)$, $\hat{Q}(x)$. Vidíme, že v $x = u_1 = 1$ prechod medzi $\hat{P}(x)$, $\hat{Q}(x)$ je hladký. Grafy výsledkov na základe (4) a (5) zo schém 3, 4 sú prakticky

rovnaké. Ale život prináša aj také dáta, na ktorých schémy 1, 2 (na rozdiel od schém 3, 4) dajú prijateľnú aproximáciu iba vhodným výberom spoločného uzla u_1.



Obrázok 8: Polynómy, body a ich odhady

5. Záver

Podarilo sa nám odhadnúť susedné polynómy s hladkým prechodom v ich spoločnom bode pomocou MNŠ naraz vďaka navrhnutým modelom so *spoločnými parametrami* bez riešenia dodatočných splajn rovníc. Nami navrhnutý prístup na vyriešenie danej úlohy potrebuje riešiť sústavu iba šiestich rovníc. Známe parametrické metódy potrebujú k tomu zvyčajne viac rovníc.

Na základe IZA reprezentácie je možné odvodiť a navrhnúť schémy z dvoch častí pre polynómy vyššieho stupňa, $p \geq 2$, ktoré majú C^k hladkosť, $k \geq p-1$. Daný postup je možné zovšeobecniť aj na globálne parametrické vyhladzovanie, ktoré poskytuje parametrický *IZA splajn* s menším počtom segmentov ako dobre známy neparametrický *vyhladzujúci splajn* (smoothing spline) [1].

6. Literatúra

[1]Eubank R.L., Nonparametric regression and spline smoothing, Marcel Dekker, Inc., 1999

[2]Matejčiková A., Török Cs., Two-point Transformation and Polynomials, Kybernetika, sent for publication

[3]Török Cs., The Fusion of Interpolation and Approximation, MD CEF TU Košice, to appear

[4]Török Cs., On-line splines and approximation models, Forum Statisticum Slovacum, 7/2008, Bratislava, str.136-140

Poďakovanie

Táto publikácia vznikla vďaka podpore grantového projektu VEGA 1/4003/07 a v rámci OP Výskum a vývoj pre projekt: CaKS - Centrum excelentnosti informatických vied a znalostných systémov (ITMS: 26220120007), spolufinancovaný zo zdrojov Európskeho fondu regionálneho rozvoja

Adresa autora:

Csaba Török, doc. RNDr. CSc.
Ústav informatiky, PF UPJŠ
Jesenná 5, 04001 Košice
csaba.torok@upjs.sk

The impact of human capital on total factor productivity in industries in the Slovak Republic
Vliv lidského kapitálu na souhrnnou produktivitu faktorů ve slovenských odvětvích

Kristýna Vltavská, Jakub Fischer

Abstract: The aim of this paper is to show the impact of human capital (measured as the level of education) on productivity in all industries in Slovakia. It concentrates on development in the period between the years 2004 and 2007. The article implicitly examines the necessity of investments in human capital for increasing both national and international competitiveness of industries. It is important to take into account the changing composition of labour force. Employing decomposition of the contribution of labour services into the contribution of hours worked and the contribution of labour composition one can find out the impact of labour skills on the productivity.

Abstrakt: Cílem příspěvku je ukázat vliv lidského kapitálu na souhrnnou produktivitu faktorů. Lidský kapitál v příspěvku odhadujeme pomocí nejvyššího dosaženého stupně vzdělání a vliv analyzujeme v jednotlivých slovenských odvětvích v období 2004 – 2007. Oproti běžně používaným postupům je brána v úvahu změna ve složení pracovní síly, což umožňuje provést odhad vlivu služeb práce (labour services) jako součet vlivu počtu odpracovaných hodin a vlivu složení pracovní síly.

Key words: labour composition effect, labour services, total factor productivity

Klíčová slova: služby práce, souhrnná produktivita faktorů, vliv složení pracovní síly

1. Introduction

Total factor productivity is one of the most appropriate indicators to evaluate economic performance. In measuring TFP the labour input reflects the effort, skills and work time of the workforce. While the data of hours worked capture the time dimension, they do not reflect the skill size. This is achieved by using the labour services as the labour input. By employing the decomposition of the labour services into the contribution of labour composition and contribution of hours worked one can find out the influence of labour skills on total factor productivity.

The aim of this paper is to estimate the impact of human capital (measured as the level of education) on TFP in all Slovak industries in the period between 2004 and 2007.

2. Computation of total factor productivity

Index of productivity of two factors by the gross value added is computed by indices of product (Y), capital (K) and labour (L):

$$\frac{Y_1}{Y_0} = \frac{A_1}{A_0}\left(\frac{K_1}{K_0}\right)^{1-\alpha}\left(\frac{L_1}{L_0}\right)^{\alpha}, \qquad (1)$$

where Y_1/Y_0 is the index of value added,

K_1/K_0 is the index of fixed assets,

L_1/L_0 is the index of hours worked,

α is the average share of compensation of employees on value added.

The analysis uses data from the Statistical Office of the Slovak Republic which include value added, hours worked and net fixed assets divided by the industries in the period between 2004 and 2007. The ESA 95 recommends the hours worked being used as the input of labour. The net fixed assets are used as the capital input. The data are divided into 14 industries (industry A is put together with industry B because this industry is too small for the individual analysis). Because of using the multiplicative computation the results do not work out.

Table 1: Calculation of total factor productivity index, using hours worked and net fixed assets as inputs, total growth from 2004 to 2007 (%)

	VA	H	FA	TFP
Total	29.76	5.11	3.88	18.84
A + B Agriculture, forestry, fishing	18.57	-2.24	-38.88	98.43
C Mining and quarrying	4.52	9.88	31.60	-27.73
D Manufacturing	79.25	5.44	13.24	50.11
E Electricity, gas and water supply	-27.03	-0.12	27.22	-42.57
F Construction	40.75	6.89	18.69	10.94
G Wholesale and retail trade; repairs	74.23	7.38	3.28	57.10
H Hotels and restaurants	19.15	12.60	-4.06	10.29
I Transport, storage and communication	-9.78	7.39	18.84	-29.31
J Financial intermediation	-33.13	0.97	-2.41	-32.14
K Real estate, renting and business activities	8.84	5.86	3.50	-0.66
L Public administration and defence	40.28	7.05	-5.81	39.13
M Education	-5.13	0.55	-5.68	0.04
N Health and social work	0.17	2.44	30.40	-25.01
O Other community, social and personal service activities	54.57	-1.56	-4.34	64.14

Note: VA – change in value added in %, H – contribution of change in hours worked to change in VA, FA – contribution of change in net fixed assets to change in VA, TFP – total factor productivity growth in %
Source: Statistical Office of Slovak Republic, computations of authors

Gross value added of the whole economy grew 29.76% in the period between 2004 and 2007. The main proportion of this was constituted by TFP (18.84%). Hours worked and fixed assets represent a lower influence with 5.11% and 3.88% respectively. The highest growth of gross value added was achieved in the "G" industry – Wholesale and retail trade with 74.23% in the period between 2004 and 2007. The main proportion of this was constituted by TFP (57.10%). However, one needs to be careful with the conclusions because values of inputs are preliminary.

3. Alternative computation of total factor productivity

The productivity of various types of labour input is different. This paper distinguishes four levels of education – primary, secondary without A levels, secondary with A levels, tertiary. Standard measures of labour input do not take these differences into account. But it is necessary to measure labour input that takes the diversity of labour force into account when analysing productivity and the contribution of labour to output growth. These measures are

called labour services [2], as they allow for differences in the amount of services delivered per unit of labour in the growth accounting approach. It is assumed that the flow of labour services for each labour type is proportional to hours worked, and workers are paid according to their marginal productivities[9]. Then the index of labour services input LS is given by:

$$\Delta \ln LS_t = \sum_l \bar{v}_{l,t} \Delta \ln H_{l,t}, \tag{2}$$

where $\Delta \ln H_{l,t}$ indicates the growth of hours worked by labour type l and weights are given by the average shares of each type in the values of labour compensation. Thus, aggregation takes into account the changing composition of the labour force. A shift in the share of hours worked by low-skilled workers to high-skilled workers will lead to a growth of labour services which is larger than the growth in total hours worked. This difference is called the labour composition effect.

Because of using the labour services it is necessary to modify (1) into:

$$\frac{Y_1}{Y_0} = \left(\frac{K_1}{K_0}\right)^{1-\alpha} \left(\frac{LS_1}{LS_0}\right)^{\alpha} \left(\frac{A_1^*}{A_0^*}\right), \tag{3}$$

where LS_1/LS_0 is the index of labour services.

The index of labour services is divided into the index of hours worked and the index of labour composition:

$$\frac{LS_1}{LS_0} = \left(\frac{H_1}{H_0}\right)\left(\frac{LC_1}{LC_0}\right). \tag{4}$$

At first, the labour services are computed then the development of labour composition as proportion of the development of labour services and the development of hours worked.

The data from the Statistical Office of the Slovak Republic and Trexima[10] are used for the first part of computation. These data sets include the numbers of employed divided into 14 industries by the level of education, average hours worked and average wages for each industry in the period between 2004 and 2007. The analysis needs the hours worked for each industry and the level of education. In this case the average hours worked and the number of employed are multiplied for each industry.

[9] In the analysis average wages are used instead of marginal productivities because one can assume that the average wages are indicators of marginal productivities.

[10] From the ISPV= Information system on the average wages.

Table 2: Calculation of total factor productivity index, using labour services and net fixed assets as inputs, total growth from 2004 to 2007 (%), total growth from 2004 to 2007 (%)

	VA	LS	H	LC	FA	TFP*
Total	29.76	7.98	5.11	2.74	3.88	15.67
A + B Agriculture, forestry, fishing	18.57	-0.61	-2.24	1.66	-38.88	95.18
C Mining and quarrying	4.52	14.67	9.88	4.35	31.60	-30.74
D Manufacturing	79.25	9.95	5.44	4.28	13.24	43.95
E Electricity, gas and water supply	-27.03	2.51	-0.12	2.64	27.22	-44.05
F Construction	40.75	9.56	6.89	2.49	18.69	8.24
G Wholesale and retail trade; repairs	74.23	12.29	7.38	4.57	3.28	50.23
H Hotels and restaurants	19.15	21.56	12.60	7.95	-4.06	2.17
I Transport, storage and communication	-9.78	14.35	7.39	6.48	18.84	-33.61
J Financial intermediation	-33.13	1.59	0.97	0.61	-2.41	-32.56
K Real estate, renting and business activities	8.84	6.17	5.86	0.29	3.50	-0.95
L Public administration and defence	40.28	10.66	7.05	3.37	-5.81	34.59
M Education	-5.13	-3.09	0.55	-3.62	-5.68	3.80
N Health and social work	0.17	9.61	2.44	6.99	30.40	-29.92
O Other community, social and personal service activities	54.57	2.04	-1.56	3.65	-4.34	58.36

Note: VA – change in value added in %, LS – contribution of change in labour services to change in VA, H – contribution of change in hours worked to change in VA, LC – contribution of change in labour composition to change in VA, FA – contribution of change in net fixed assets to change in VA, TFP* – total factor productivity growth in %, adjusted of labour composition effect
Source: Statistical Office of Slovak Republic, computations of authors

Using the decomposition of contribution of labour services into contribution of hours worked and contribution of labour composition, one can find out how a shift in the portion of hours worked by low-skilled workers to high-skilled workers leads to a growth of labour services that is larger than the growth of hours worked, mainly in industries H, N, I, G, C and D. The labour composition effect caused movement of more educated workforce from less productive industries to more productive ones.

The value of contribution of labour composition shows that during the period between 2002 and 2006 the level of education grows in following industries: H – Hotels and restaurants (7.95%), N – Health and social work (6.99%), I – Transport, storage and communication (6.48%), G – Wholesale and retail trade (4.57%) and C – Mining and quarrying (4.35%). One can see that the qualification level of labour force of the sector in question increased. The main cause for the increase was the growth in number of workforce with secondary education with A levels and people with college.

4. Conclusion

There is an obvious impact of human capital change (measured as the change in level of education) on productivity growth in the Slovak industries between the years 2004 and 2007. The decomposition of contribution of labour services on TFP into contribution of hours worked and contribution of labour composition showed that high-skilled workforce shifted to

the more productive industries (e.g. Hotels and restaurants, Health and social work, etc.) in the examined period of time.

A comparison of contribution of labour composition in the Czech Republic and in Slovakia should prove very interesting. However, this will be part of further studies.

5. References

[1]JÍLEK, J. – MORAVOVÁ, J. 2007. Ekonomické a sociální indikátory. Praha: Futura, 2007. 248 s. ISBN 978-80-86844-29-9.

[2]O'MAHONY, M. – TIMMER, P.M. – VAN ARK, B. EU KLEMS Growth and Productivity Accounts: Overview November 2007 Release, www.euklems.net.

[3]Statistical Office of the Slovak Republic: National Accounts 2009.

[4]Statistical Office of the Slovak Republic: Labour Force Survey, 2004 – 2007.

[5]Trexima: ISPV survey 2004 -2007.

[6]VLTAVSKÁ, K. 2009. The impact of human capital on productivity in all industries in the Czech Republic, AMSE 2009.

Contacts:

Kristýna Vltavská, Ing.
KEST FIS VŠE v Praze
Nám. W.Churchilla 4
130 67 Praha 3
kristyna.vltavska@vse.cz

Jakub Fischer, doc. Ing. Ph.D.
KEST FIS VŠE v Praze
Nám. W. Churchilla 4
130 67 Praha 3
fischerj@vse.cz

This paper has been prepared under the support of the project of the Czech Science Foundation No. 402/07/0387 „Capital Services in National Accounts and its Impact on GDP of the Czech Republic".

Niektoré vlastnosti odhadu regresných parametrov v lineárnom zmiešanom modeli
Some Properties of the Estimator of Regression Parameters in Linear Mixed Model

Gejza Wimmer

Abstract: It is a well known fact that in small-sample size cases a conventional approach to estimating the precision and making inference about regression parameters in linear mixed model (i.e. an approach based on asymptotic properties of some estimator of the parameters) is inappropriate. Many authors have studied this problem and proposed some modifications of these classical methods to remove their deficiencies. Based on a simulation study, this contribution discusses some properties of such an estimator of regression parameters and compares it with a conventional estimator of regression parameters for longitudinal data analysis with AR(1) errors in small sample sizes.

Key words: Confidence region, Linear mixed model, Regression parameters.

Kľúčové slová: konfidenčné oblasti, lineárny zmiešaný model, regresné parametre.

1. Úvod

V praktických úlohách je často potrebné analyzovať údaje získané opakovanými meraniami na viacerých subjektoch v rôznych časových okamihoch. V takomto experimente je potom prirodzeným predpokladom, že jednotlivé vektory pozorovaní na rôznych subjektoch sú navzájom nezávislé, pričom pozorovania získané na danom subjekte môžu vykazovať nejakú formu korelácie. Informácie dosiahnuté z takto navrhnutého experimentu sa nazývajú longitudinálne dáta.

Jedným z primárnych cieľov analýzy longitudinálnych dát je potom popísať spoločnú črtu všetkých pozorovaných subjektov. Pri ich modelovaní je však vhodné uvažovať okrem týchto spoločných efektov navyše aj individuálne vplyvy jednotlivých subjektov na svoje opakované merania. Tu sa ukazuje využitie lineárneho zmiešaného modelu (LZM) ako veľmi vhodný prístup k ich ďalšiemu spracovaniu, pretože už v samotnom zápise zohľadňuje možnosť oboch spomenutých efektov a navyše ponúka istú formu závislosti opakovaných meraní na danom subjekte pomocou náhodných chýb.

Nech teda opakované pozorovania na *i* –tom $(i = 1, 2, \ldots, N)$ subjekte spĺňajú LZM

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\varepsilon}_i\,, \tag{1}$$

kde $\mathbf{X}_i$ sú známe $(n_i \times p)$ –rozmerné matice plánu, $\boldsymbol{\beta}$ je neznámy p –rozmerný vektor (pevných) regresných parametrov, spoločný pre všetky subjekty. $\mathbf{Z}_i$ sú taktiež známe $(n_i \times r)$ –rozmerné matice plánu pre r –rozmerné individuálne náhodné efekty $\mathbf{b}_i$, ktoré sú, podobne ako $\boldsymbol{\beta}$, neznáme. Navyše u nich predpokladáme, že sú navzájom nezávislé s rozdelením $N(\mathbf{0}, \mathbf{D})$. $\boldsymbol{\varepsilon}_i$ sú n_i-rozmerné, navzájom nezávislé vektory náhodných chýb jednotlivých subjektov, pričom pre jednotlivé $\boldsymbol{\varepsilon}_i$ platí, že sú nezávislé od $\mathbf{b}_i$ a $N(\mathbf{0}, \mathbf{R}_i)$ rozdelené.

Kovariančná matica *i* –teho subjektu spĺňajúceho popísaný model (1) je potom

$$Var(\mathbf{y}_i) = \mathbf{V}_i = \mathbf{Z}_i\mathbf{D}\mathbf{Z}_i' + \mathbf{R}_i. \tag{2}$$

Predpokladajme, že matica $\mathbf{D}$ je funkciou nejakého parametra $\boldsymbol{\theta}_D$ a matice $\mathbf{R}_i$ (pre všetky *i*) sú funkciou iného parametra $\boldsymbol{\theta}_R$. Označme $\boldsymbol{\theta} = (\boldsymbol{\theta}_D, \boldsymbol{\theta}_R)'$ ako vektor všetkých kovariančných

komponentov daného modelu. Teraz je zrejmé, že matica (2) je funkciou vektora $\boldsymbol{\theta}$ a teda $Var(\mathbf{y}_i) = \mathbf{V}_i = \mathbf{V}_i(\boldsymbol{\theta})$.

2. Odhad neznámych regresných parametrov a jeho charakteristiky

V prípade, že kovariančné parametre $\boldsymbol{\theta}$ modelu (1) sú známe (a teda poznáme i hodnoty matíc $\mathbf{V}_i$), odhad neznámych regresných parametrov $\boldsymbol{\beta}$ získame použitím váženej metódy najmenších štvorcov, pričom vieme, že najlepší lineárny nevychýlený odhad

$$\widetilde{\boldsymbol{\beta}} = (\textstyle\sum_{i=1}^{N} \mathbf{X}_i' \mathbf{V}_i^{-1} \mathbf{X}_i)^{-1} (\sum_{i=1}^{N} \mathbf{X}_i' \mathbf{V}_i^{-1} \mathbf{y}_i) \tag{3}$$

je asymptoticky normálne rozdelený so strednou hodnotou

$$E[\widetilde{\boldsymbol{\beta}}] = \boldsymbol{\beta} \tag{4}$$

a kovariančnou maticou

$$Var[\widetilde{\boldsymbol{\beta}}] = (\textstyle\sum_{i=1}^{N} \mathbf{X}_i' \mathbf{V}_i^{-1} \mathbf{X}_i)^{-1} \equiv \boldsymbol{\Phi}. \tag{5}$$

Napriek tomu, že z hľadiska analýzy longitudinálnych dát sú v našom prvoradom záujme najmä odhady neznámych regresných parametrov $\boldsymbol{\beta}$ spomenutý postup často nie je možné v praxi uplatniť, pretože kovariančné parametre $\boldsymbol{\theta}$ nie sú známe. Zaužívaným spôsobom k ich dosiahnutiu je tzv. REML funkcia vierohodnosti ktorej maximalizáciou dostaneme požadované REML odhady kovariančných parametrov $\widehat{\boldsymbol{\theta}}$. REML odhady neznámych kovariančných matíc modelu (1) sú potom $\mathbf{V}_i(\widehat{\boldsymbol{\theta}}) = \widehat{\mathbf{V}}_i$.

V prípade, že kovariančné parametre $\boldsymbol{\theta}$ modelu (1) sú neznáme, avšak máme k dispozícii ich REML odhady, prevažná časť autorov základnej literatúry o longitudinálnych dátach odporúča na získanie odhadu neznámych regresných parametrov $\boldsymbol{\beta}$ a jeho kovariančnej matice využiť vzťah (3) a (5), pričom v uvedených vzorcoch sa nahradia kovariančné matice $\mathbf{V}_i$ ich REML odhadmi $\widehat{\mathbf{V}}_i$. REML odhad regresných parametrov je potom

$$\widehat{\boldsymbol{\beta}} = \left(\textstyle\sum_{i=1}^{N} \mathbf{X}_i' \mathbf{V}_i^{-1}(\widehat{\boldsymbol{\theta}}) \mathbf{X}_i\right)^{-1} \left(\sum_{i=1}^{N} \mathbf{X}_i' \mathbf{V}_i^{-1}(\widehat{\boldsymbol{\theta}}) \mathbf{y}_i\right). \tag{6}$$

Takto získaný odhad je taktiež asymptoticky normálne rozdelený so strednou hodnotou $E[\widehat{\boldsymbol{\beta}}] = \boldsymbol{\beta}$, avšak jeho kovariančná matica $Var[\widehat{\boldsymbol{\beta}}] \neq (\sum_{i=1}^{N} \mathbf{X}_i' \widehat{\mathbf{V}_i}^{-1} \mathbf{X}_i)^{-1} \equiv \widehat{\boldsymbol{\Phi}}$. Na túto nerovnosť poukazuje už samotný fakt, že $\widehat{\boldsymbol{\Phi}}$ je vychýleným odhadom $\boldsymbol{\Phi}$ a taktiež skutočnosť, že takto definovaná kovariančná matica odhadu regresných parametrov $\widehat{\boldsymbol{\beta}}$ by nezohľadňovala vo svojom zápise neistotu, ktorú so sebou prináša odhadovanie neznámych kovariančných parametrov modelu (1). Obe tieto nepresnosti sa vo veľkej miere prejavujú najmä v prípade malých rozsahov výberu. Na základe týchto zistení odvodili autori [1] aproximáciu odhadu kovariančnej matice pre REML odhad regresných parametrov $\widehat{\boldsymbol{\beta}}$, ktorá môže byť zapísaná ako súčet odhadu asymptotickej kovariančnej matice $\widehat{\boldsymbol{\Phi}}$, odhadu jeho výchylky oproti skutočnej hodnote $\widehat{\boldsymbol{\Lambda}}$ a odhadu neistoty spôsobenej odhadovaním hodnôt $\boldsymbol{\theta}$, ktorý je $(\widehat{\boldsymbol{\Lambda}} + \widehat{\mathbf{A}})$. Navrhnutý odhad $Var[\widehat{\boldsymbol{\beta}}]$ je potom

$$\widehat{Var(\widehat{\boldsymbol{\beta}})} = \widehat{\boldsymbol{\Phi}} + 2\widehat{\boldsymbol{\Lambda}} + \widehat{\mathbf{A}} \equiv \widehat{\boldsymbol{\Phi}}_{mod}. \tag{7}$$

3. Konfidenčné oblasti

Uvažujme teraz o štatistických inferenciách ohľadne l lineárnych kombinácií prvkov $\boldsymbol{\beta}$, konkrétne zostrojením konfidenčnej oblasti pre $\mathbf{L}\boldsymbol{\beta}$, kde $\mathbf{L}$ je známa $(l \times p)$–rozmerná matica.

Opäť, v prípade, že poznáme kovariančné parametre modelu $\boldsymbol{\theta}$, je z predchádzajúcej časti zrejmé, že

$$(\mathbf{L}\widetilde{\boldsymbol{\beta}} - \mathbf{L}\boldsymbol{\beta})' (\mathbf{L}\boldsymbol{\Phi}\mathbf{L}')^{-1} (\mathbf{L}\widetilde{\boldsymbol{\beta}} - \mathbf{L}\boldsymbol{\beta}) \tag{8}$$

má asymtoticky χ^2 rozdelenie s l stupňami voľnosti a teda vieme skonštruovať asymptoticky presné konfidenčné oblasti pre požadovanú lineárnu kombináciu regresných parametrov $\boldsymbol{\beta}$.

V prípade, že nepoznáme kovariančné parametre modelu $\boldsymbol{\theta}$, avšak máme k dispozícii ich REML odhady, autori v [1] ukázali, že je vhodné uvažovať, že štatistika

$$F^* = \lambda(1/l)(\mathbf{L}\widehat{\boldsymbol{\beta}} - \mathbf{L}\boldsymbol{\beta})'(\mathbf{L}\widehat{\boldsymbol{\Phi}}_{mod}\mathbf{L}')^{-1}(\mathbf{L}\widehat{\boldsymbol{\beta}} - \mathbf{L}\boldsymbol{\beta}), \qquad (9)$$

má $F_{l,m}$, teda Fisherovo–Snedecorovo rozdelenie s l a m stupňami voľnosti. Tento vzťah zohľadňuje okrem náhodnosti vektora $\widehat{\boldsymbol{\beta}}$ aj náhodnú štruktúru matice $\mathbf{L}\widehat{\boldsymbol{\Phi}}_{mod}\mathbf{L}'$. V tom istom článku autori navrhli i odhad škálovacieho parametra λ, ako aj odhad stupňov voľnosti m. S využitím týchto výsledkov potom vieme i v tomto prípade skonštruovať požadované (približné) konfidenčné oblasti pre $\mathbf{L}\boldsymbol{\beta}$.

V ponúknutej simulačnej štúdii porovnávame simulované 95% pravdepodobnosti pokrytia skutočnej hodnoty regresného parametra $\boldsymbol{\beta}$ pre konfidenčné oblasti, vyplývajúce z oboch uvedených prístupov.

4. Simulačná štúdia

Pre rôzne kombinácie počtu subjektov N a počtu meraní na jednotlivých subjektoch n_i sme pomocou 10 000 simulácií spočítali empirické pravdepodobnosti pokrytia skutočnej hodnoty $\boldsymbol{\beta}$ pre konfidenčnú oblasť získanú pomocou (8) nahradením neznámeho parametra $\boldsymbol{\theta}$ jeho odhadom, ako aj pre konfidenčnú oblasť navrhnutú v [1], opierajúcu sa o modifikovaný odhad kovariančnej matice $Var[\widehat{\boldsymbol{\beta}}]$ a štatistiku F^*. V prezentovaných simuláciách sa uvažuje 2 –rozmerný regresný parameter $\boldsymbol{\beta} = (0,1)'$, 2 –rozmerný vektor náhodných efektov $\boldsymbol{b}_i = (b_{i,1}, b_{i,2})'$ s kovariančnou maticou $\mathbf{D} = \mathbf{I}$ (jednotková). V záujme zavedenia určitej formy korelácie jednotlivých subjektov na svoje opakované merania sme pre rozdelenie chýb uvažovali AR(1) proces s rozptylom $\sigma^2 = 1$ a autokorelačným koeficientom $\rho = 0.8$

Tabuľka 18a: Empirické pravdepodobnosti pokrytia β pre rôzny počet subjektov s rovnakým počtom 5 pozorovaní na každom subjekte založené na štatistike (8) pre konfidenciu 0.95

počet subjektov	počet opakovaných meraní na subjekte	empirická pravdep. pokrytia
5	5	0,8354
10	5	0,9018
30	5	0,9353

Tabuľka 1b: Empirické pravdepodobnosti pokrytia β pre rôzny počet subjektov s rovnakým počtom 5 pozorovaní na každom subjekte založené na výsledkoch [1] pre konfidenciu 0.95

počet subjektov	počet opakovaných meraní na subjekte	empirická pravdep. pokrytia
5	5	0,8576
10	5	0,9152
30	5	0,9410

5. Záver

Z uvedenej simulačnej štúdie možno na základe tabuliek 1a a 1b usudzovať, že s rastúcim počtom subjektov, aj napriek nízkemu počtu pozorovaní na jednotlivých subjektoch, sa obe vyšetrené konfidenčné oblasti približujú požadovanej teoretickej hodnote. Rovnaký výsledok ponúka i vyhodnotenie tabuliek 2a a 2b, kde však obe spomenuté

konfidenčné oblasti s rastúcim počtom opakovaných pozorovaní na malom počte subjektov dosahujú svoju požadovanú teoretickú hodnotu oveľa pomalšie.

Zo všetkých uvedených situácií je taktiež zrejmé, že použitie modifikovaného odhadu kovariančnej matice odhadu regresných parametrov $\widehat{\boldsymbol{\Phi}}_{mod}$ a následná aproximácia testovacej štatistiky F^* pomocou $F_{l,m}$ rozdelenia uvedená v [1] síce zlepšuje výsledky oproti „klasickému" prístupu, bohužiaľ len nepatrne. Tento nedostatok by mohol byť odstránený zahrnutím odchýlky odhadu kovariančných parametrov modelu (1) do záverov uvedených v [1], keďže v spomínanom článku autori pri odvodzovaní hodnôt $\widehat{\boldsymbol{\Phi}}_{mod}$ využívajú nevychýlenosť (resp. len malú odchýlku) odhadov $\widehat{\boldsymbol{\theta}}$ od svojej skutočnej hodnoty $\boldsymbol{\theta}$, čo vo všeobecnosti pre REML odhady neplatí. V [2] sa autori zamerali práve na túto skutočnosť a navrhujú nový odhad $Var[\widehat{\boldsymbol{\beta}}]$, ktorý v sebe zahŕňa aj nepresnosť v odhadovaní kovariančných parametrov modelu. Porovnanie postupu z [2] s nami uvažovanými postupmi je predmetom ďalšieho výskumu.

Tabuľka 2a: Empirické pravdepodobnosti pokrytia β pre rovnaký počet subjektov (N = 5) s rôznym počtom pozorovaní na každom subjekte založené na štatistike (8) pre konfidenciu 0.95

počet subjektov	počet opakovaných meraní na subjekte	empirická pravdep. pokrytia
5	5	0,8354
5	10	0,8535
5	30	0,8838
5	50	0,9012

Tabuľka 2b: Empirické pravdepodobnosti pokrytia β pre rovnaký počet subjektov (N = 5) s rôznym počtom pozorovaní na každom subjekte založené na výsledkoch [1] pre konfidenciu 0.95

počet subjektov	počet opakovaných meraní na subjekte	empirická pravdep. pokrytia
5	5	0,8354
5	10	0,8776
5	30	0,9152
5	50	0,9285

6. Literatúra

[1] KENWARD, M.G. – ROGER, J.H. 1997. Small Sample Inference for Fixed Effects from Restricted Maximum Likelihood. In: Biometrics, č. 53, 1974, s. 983 – 997.

[2] KENWARD, M.G. – ROGER, J.H. 2009. An improved approximation to the precision of fixed effects from restricted maximum likelihood. In: Computational Statistics and Data Analysis, doi:10.1016/j.csda.2008.12.013

Adresa autora:

Gejza Wimmer, Mgr.
Matematický ústav SAV, Štefánikova 49, 814 73 BRATISLAVA
gejza.wimmer.ml@mat.savba.sk

Výskum bol riešený v rámci projektov VEGA 1/0077/09, APVV-SK-AT-0003-09 a APVV-RPEU-0008-06.

Konsolidace energetických odvětví při sestavování tabulek dodávek a užití
Consolidation of energy branch in supply use tables

Jaroslav Zbranek, Petr Musil

Abstract: The aim of this article is to present consolidation in the annual national accounts in the Czech Republic. It is necessary to take into account changes in real economy. Consolidation means an adjustment of output and intermediate consumption. It is an important adjustment in order to provide reliable information about output, distribution and consumption of electricity and gas at current and previous year's prices.

Key words: consolidation, tables of supply and consume, deflation, energetic

Klíčová slova: konsolidace, tabulky dodávek a užití, deflace, energetika

1. Úvod

Systém národního účetnictví představuje v současné době asi nejpropracovanější model národního hospodářství (NH). Vzhledem k tomu, že jeho metodologický rámec je zakotven ve standardech národního účetnictví SNA 1993 potažmo ESA 1995, lze konstatovat, že hodnoty ukazatelů NH, které soustava národních účtů poskytuje, jsou mezinárodně srovnatelné.

Cílem příspěvku však není výklad či popis zmíněných mezinárodních standardů, potažmo systému národního účetnictví. Smyslem je pouze přiblížit metodologické změny při sestavování tabulek dodávek a užití, které bylo nutné implementovat, aby národní účty nadále přinášely kvalitní a nezkreslené informace o NH České republiky. Změny se týkaly pouze odvětví elektrické energie, plynárenského průmyslu a odvětví rafinérských zpracovatelů ropy, které v posledních letech prošly rozsáhlou restrukturalizací. Bylo tedy nanejvýš vhodné tyto faktické změny promítnout i do národních účtů, které vlastně představují ekonomický model reálného hospodářství. Energetická odvětví představují z globálního hlediska jeden z nejvýznamnějších segmentů hospodářství každé země. Důvod pro toto tvrzení tkví zejména v tom, že v podstatě žádné další odvětví se již v dnešní době při své produktivní činnosti neobejde bez energie. Změny v metodice podporují také reálná čísla, ukazující nemalý reálný význam těchto odvětví, což je vidět na příkladu roku 2006, kdy uvedená odvětví představovala v úhrnu podíl 3 % na hrubé přidané hodnotě (HPH), což není zajisté zanedbatelný vliv.

Jak se v následujících odstavcích ukáže, provedené metodické změny, které se v praxi Českého statistického úřadu (ČSÚ) označují jako „konsolidace" zde uvedených odvětví, které jsou rovněž v souladu se standardy ESA 1995, příznivě působí na vyjádření HPH těchto odvětví ve stálých cenách tedy na deflaci.

2. Současná situace na trzích elektrické energie, plynu a rafinérských zpracovatelů ropy

Trh s elektrickou energií se v ČR začal uvolňovat v roce 2002. V roce 2005 odstartoval v ČR proces označovaný jako „unbundling", tedy oddělení aktivit provozovatele přenosových sítí od ostatních aktivit (výroba, obchod). Tento proces tedy souvisí s liberalizací trhu s elektřinou [7]. K liberalizaci neboli uvolnění zmíněných trhů se ČR zavázala svým přistoupením k EU v roce 2004. Povinnost liberalizace energetických trhů ve všech zemích EU je dána směrnicí Evropského parlamentu a rady 2003/54/ES s názvem „O společných pravidlech pro vnitřní trh s elektrickou energií" ze dne 26. 6. 2003, která nahradila původní

směrnici 96/92/ES [6]. Úplné otevření trhu s elektřinou bylo touto směrnicí stanoveno na 1. 7. 2007. V ČR se ovšem podařilo plně tento trh uvolnit již k 1. 1. 2006. Nejprve došlo k oddělenému vedení účetních záznamů pro jednotlivé činnosti. Od roku 2005 pak bylo nutné oddělení manažerské a právní. Nyní energetický zákon ukládá provozovatelům distribučních soustav, kteří dodávají elektrickou energii více než 90 tisícům odběratelů, povinnost oddělit činnosti distribuce od ostatních licencovaných činností. V praxi toto uvolnění znamená, že poslední skupina koncových spotřebitelů elektrické energie získala možnost výběru dodavatele elektřiny. Regulaci nyní podléhají pouze činnosti, které se vyznačují monopolním charakterem. Jako takové lze označit dopravu elektřiny od výrobního zdroje prostřednictvím přenosového a distribučního systému konečnému spotřebiteli, a dále činnosti spojené se zajištěním stability energetického systému z technického i obchodního hlediska.

Na trhu s plynem je situace podobná. Možnost obchodovat s plynem v ČR pro všechny subjekty, které získají potřebná oprávnění a také možnost konečných spotřebitelů vybrat si svého dodavatele byla příčinou rozsáhlé transformace českého plynárenského trhu, která proběhla v letech 2006 a 2007. Také na tomto trhu došlo k oddělení výrobních a obchodních aktivit od činnosti provozovatele rozvodné sítě, jehož činnost nadále podléhá regulaci (unbundling) [11]. To znamená, že od počátku roku 2006 nemá provozovatel sítě právo zároveň obchodovat s plynem. Ve stejné době se na českém plynárenském trhu objevuje nový pojem, kterým je „provozovatel přepravní soustavy“. Do této doby nebyl trh plně liberalizován a funkce provozovatele přepravní soustavy (tzn. dopravce plynu) nebyla striktně oddělena od obchodu s plynem. Obchodník s plynem tedy nově může obhospodařovat pouze činnosti obchodní a uskladňování plynu. Český plynárenský trh je dále tvořen z regionálních distribučních společností a menších obchodníků s plynem, kteří mají spíše lokální význam. Dalším výsledkem transformace v tomto odvětví je rozšíření trhu o nové dovozce plynu. V souvislosti s dovozem plynu lze konstatovat, že ČR je především tranzitní zemí, protože přibližně pouze čtvrtina celkového množství plynu, který je do ČR dopraven, zde také zůstává. Zbývající množství je přes ČR dopravováno do dalších zemí Evropy. Vlastní zdroje zemního plynu v ČR, kterých je méně než 1 % domácí spotřeby, se těží především na jižní Moravě. Soustava zemního plynu v ČR je tedy téměř zcela závislá na dovozu ze zahraničí. Přibližně dvě třetiny dovozu plynu pocházejí z Ruska, menší podíl dovozu plynu má svůj původ v Norsku.

Poněkud odlišná od výše popsaných trhů je situace na trhu s rafinérskými produkty, nicméně obdobně došlo k rozsáhlé transformaci vlivem vlastnických a organizačních změn v českém rafinérském průmyslu. Rafinerie přešly na systém tzv. přepracovacích rafinérií, které neprodukují výrobky, ale provádí službu spojenou s výrobou rafinérských produktů pro další subjekty (převážně podniky obchodující s pohonnými hmotami). Tím došlo k výrazné změně dosavadního všeobecně přijímaného pojetí odvětví zpracování ropy a odvětví obchodu. Odvětvová struktura agregátů národního účetnictví by měla být tvořena souhrnem činnostních jednotek, což je v praxi, kdy jsou dostupná data pouze za institucionální jednotky, velmi obtížné splnit. Odvětvové účty založené na institucionálních jednotkách v tomto případě poskytují zkreslené informace, protože zmizela výroba rafinérských produktů.

3. Konsolidace energetických odvětví

V národních účtech jsou zachyceny převážně nekonsolidované transakce. Ke konsolidaci se přistupuje především tehdy, je-li riziko, že transformační procesy v podnikové sféře povedou ke komplikacím při zachycení a analýze ukazatelů národních účtů. Konsolidací se rozumí vyloučení transakcí mezi subjekty v rámci některých odvětví, které spolu věcně přímo souvisejí. Absence konsolidace a spoléhání se pouze na podnikové účetnictví subjektů

z odvětví, kde proběhly transformační změny, by představovalo mnohdy výrazné zkreslení celkové produkce, ale také mezispotřeby.

Zde uvedené změny v metodologii jsou zahrnuty již v definitivní sestavě národních účtů za rok 2005, v semidefinitivní a definitivní za rok 2006, v předběžné a semidefinitivní sestavě za rok 2007. V předběžné sestavě národních účtů za rok 2008 je nově zohledněna tzv. ekologická daň. Případné další změny ve sledovaných odvětvích v následujících obdobích budou sledovány a pružně zohledňovány v budoucích sestavách.

Proces konsolidace, který je zde popsán, by však byl nemyslitelný bez předchozího dostatečného poznání hospodářských vazeb na příslušných trzích a bez dostupných dodatkových informací, relevantních pro dosažení racionálních závěrů celé konsolidace. V tomto ohledu je situace jednodušší, protože při deflaci produktů elektrické energie, plynu a rafinérských výrobků mohou být výsledné agregáty konfrontovány s naturálními ukazateli z energetické bilance. Jinými slovy, například v odvětví výroby a rozvodu elektřiny by měl index fyzického objemu produkce elektrické energie odpovídat meziročnímu indexu skutečně vyrobené elektrické energie (z bilance elektřiny v Gwh).

Konsolidace v odvětví výroby a distribuce elektřiny (OKEČ 401, CZ-NACE 351)

Pro národní účetnictví resp. korektní odhad HPH představuje značnou komplikaci nejen samotné faktické oddělení výroby a distribuce elektřiny, ale také stávající převažující praxe v podnikovém účetnictví firem na energetickém trhu ČR. Nepřesnosti pocházejí zejména od distribučních společností na trhu s elektřinou, které účtují o nakoupené elektřině jako o nakoupené elektrické energii v rámci provozních nákladů, místo správného zaúčtování jako o nákladech na zakoupené zboží. A také při prodeji této energie dalšímu článku na trhu nebo konečnému spotřebiteli o této energii účtují v provozních výnosech jako výnosech za vlastní produkci, místo aby věcně správně účtovaly jako o výnosech za prodej zboží. Dochází tak k nadhodnocování produkce a mezispotřeby v odvětví výroby elektřiny. Úhrn zkreslení představuje částka, o které mělo být účtováno jako o operacích se zbožím.

Následně je sporná také interpretace matice mezispotřeby energetických odvětví z hlediska input-output analýzy, kdy hlavní vstup při výrobě elektrické energie představuje spotřebovávaná elektrická energie. Zde je zjevně zahrnut vnitřní obrat, což představuje nesoulad s metodikou národního účetnictví, která předpokládá vyloučení vnitřního obratu.

Konsolidace se provádí tak, aby úroveň HPH v běžných cenách zůstala zachována. Protože k úpravě v podobě snížení dochází jednak v případě produkce, ale i mezispotřeby, a to ve stejné výši.

Základním předpokladem konsolidace je zahrnutí výroby elektrické energie standardním způsobem do produkce, zatímco všechny ostatní navazující články energetického řetězce jsou považovány pouze za obchodníky. Jejich produkce je dána obchodní marží, což je rozdíl mezi prodanou a nakoupenou elektřinou.

Konsolidace v odvětví výroby a rozvodu plynných paliv prostřednictvím sítí (OKEČ 402, CZ-NACE 352)

V roce 2006 došlo k vyčlenění potrubní přepravy plynu, proto bylo cílem při konsolidaci v plynárenských odvětvích zachovat stávající způsob zachycení výroby a distribuce plynu v národních účtech. Distribuce plynu je v ročních národních účtech zachycena na principu jednostupňového procesu „výroby“, ač o výrobu v pravém slova smyslu nejde. Plyn je do ČR převážně dovážen a zdejšími plynárenskými společnostmi je pouze aromatizován a následně distribuován spotřebitelům. Vstupní surovinou je dovážený plyn (SKP 111, CZ-CPA 06), která se zahrnuje do mezispotřeby plynárenských společností. Jejich produkcí je pak zemní plyn pro spotřebitele, zahrnutý v SKP 402 (CZ-CPA 352), což

není úplně správné. K těmto účelům je pak speciálně konstruován použitý cenový index pro deflaci takto „vyrobeného“ plynu.

Všechny další mezičlánky distribučního řetězce jsou pak považovány jako na trhu s elektřinou za obchodníky, jejichž produkce je dána pouze výší obchodní marže a z jejich mezispotřeby je proto nakoupený plyn, coby zboží vyloučen. Smyslem tohoto procesu „výroby“ je získat jediný produkt, který je užit uživateli, ačkoli podle klasifikace produktů (klasifikace SKP, CZ-CPA) neexistuje. Klasifikace SKP totiž striktně rozlišuje plyn (SKP 111) a služby rozvodu (SKP 402). Zjišťování spotřeby samotného plynu a služeb rozvodu není racionální, neboť by neúměrně zatěžovalo respondenty.

Konsolidace v odvětví rafinérského zpracování ropy (OKEČ 232, CZ-NACE 192)

Jak již bylo výše zmíněno, v odvětví rafinérského zpracování ropy tkví problém v nesouladu mezi institucionálními jednotkami, u nichž jsou data zjišťována a místními činnostními jednotkami, které předpokládá systém národních účtů. Problém je tedy ten, že jednotky zatříděné dle převažující činností do odvětví obchodu (OKEČ 50 a 51, resp. CZ-NACE 45 a 46) se staly výrobci rafinérských produktů (SKP 232 resp. CZ-CPA 192). Tyto jednotky („procesoři“) nakupují ropu a na zakázku si ji nechávají zpracovat rafinériemi. Rafinérie pak účtuje pouze hodnotu svých služeb. Produkce a mezispotřeba procesorů jsou pak podhodnoceny o hodnotu spotřebované ropy a vyrobených rafinérských produktů.

Konsolidace zde představuje korekci produkce i mezispotřeby odvětví ve stejné výši a přičtení této částky do odvětví rafinérského průmyslu. Jinými slovy lze napsat, že jde o přesun tržeb i nákladů souvisejících s výrobou rafinérských produktů od firem obchodujících s pohonnými hmotami do odvětví rafinérského průmyslu. HPH v běžných cenách je tím pádem po konsolidaci opět beze změny. K eliminaci vlivu organizačních změn se zdá jako ideální objemová metoda, kdy celková produkce je odvozena od cenového a objemového vývoje naturálních ukazatelů (tuny rafinérských produktů).

4. Deflace

Prvním krokem při odhadu reálné změny makroagregátů je přepočet hodnoty makroagregátů v běžných cenách do stálých cen předchozího roku. Na kvalitě přepočtu do stálých cen (deflace) přímo závisí kvalita odhadu indexu fyzického objemu (reálné změny). Každý makroagregát vyžaduje specifický způsob deflace, který závisí na typu ocenění.

Hrubá přidaná hodnota

Hrubá přidaná hodnota je jedním z nejdůležitějších ukazatelů každého odvětví NH. Podle výrobní metody odhadu HDP je součet HPH ve všech odvětvích a čistých daní na výrobky roven HDP. Deflace HPH se neprovádí přímo, ale pomocí tzv. dvojité deflace, kdy je nejprve provedena deflace produkce, následně deflace mezispotřeby do stálých cen předchozího roku. HPH ve stálých cenách předchozího roku je spočtena jako jejich rozdíl.

Produkce

Obecně je nejvhodnějším cenovým indexem pro deflaci produkce index cen průmyslových cen výrobců (PPI), viz například [3].

U výroby elektrické energie je ještě nutné vzít v úvahu typ zákazníka, kterému je elektrická energie dodávána, neboť cenové podmínky jsou pro různé typy zákazníků velmi odlišné. Z tohoto důvodu byla v českých národních účtech rozdělena komodita SKP 401 na dvě komodity. Komodita SKP 401A00 zahrnuje elektrickou energii, která je prodávána domácnostem a je deflována indexem spotřebitelských cen. Komodita 401000 obsahuje elektrickou energii pro ostatní uživatele a je přepočtena indexem cen průmyslových výrobců.

Tento způsob deflace umožňuje reagovat na rozdílné trendy ve vývoji cen pro domácnosti a pro ostatní zákazníky.

U plynu je nutné upozornit i na jiné metodické vymezení SKP 402 v národních účtech od vymezení ve Standardní klasifikaci produkce. Podle definice národních účtů obsahuje toto SKP služby rozvodu i samotný plyn, zatímco podle Standardní klasifikace produkce jsou v SKP 402 zahrnuty pouze služby rozvodu. Tuto změnu v metodickém vymezení je nutné vzít v úvahu při deflaci, a proto se sestavují tři různé indexy cen průmyslových výrobců:

1. index zahrnující pouze samotný plyn
2. index zahrnující pouze služby rozvodu
3. vážený index obsahují plyn i jeho distribuci

Komodita SKP 111A (surový plyn) je deflována indexem číslo ad 1. Pro účely deflace je komodita SKP 402 rozdělena na SKP 402001 zahrnující plyn pro domácnosti a SKP 402000 obsahující plyn pro ostatní uživatele. Princip a důvody jsou stejné jako u elektrické energie, tj. reagovat na odlišný vývoj cen v obou segmentech trhu. Plyn pro domácnosti (SKP 402A00) je deflován indexem spotřebitelských cen, plyn pro ostatní uživatele je deflován váženým indexem (index č.3).

Mezispotřeba a tvorba hrubého kapitálu

Deflace těchto makroagregátů je náročnější proces, protože jsou v tabulkách dodávek a užití zachyceny v kupních cenách a je nutné nejprve odhadnout odpočtové matice v běžných cenách (matice DPH, obchodních a dopravních přirážek, daní na produkty a dotací na produkty). Mezispotřeba a tvorba hrubého kapitálu v základních cenách, tj. po odečtení všech matic, jsou deflovány obdobně jako produkce s tím, že je nutné rozlišit, jak velká část pochází z dovozu. Kvalita deflace je závislá na kvalitě odpočtových matic. Princip jejich deflace je popsán na příkladu DPH, u ostatních matic je postup obdobný. DPH ve stálých cenách předchozího roku se vypočte jako součin sazby daně z předchozího roku a základu daně přeceněného do cen předchozího roku. Deflátor daně je následně vypočten jako implicitní: DPH v běžných cenách k DPH ve stálých cenách.

Výdaje na konečnou spotřebu domácností

Výdaje na konečnou spotřebu domácností (KSD) jsou zachyceny v klasifikaci COICOP a na rozdíl od ostatních makroagregátů jsou deflovány v kupních cenách indexy spotřebitelských cen. Do klasifikace SKP jsou údaje v běžných i stálých cenách převedeny pomocí převodníku.

5. Závěr

Konsolidace energetických odvětví je jednou z nejvýznamnějších změn, ke které došlo při sestavování tabulek dodávek a užití v poslední době. Tato změna má zásadní význam při interpretaci input-output tabulek, neboť údaje po konsolidaci mají mnohem vyšší vypovídací schopnost o reálné ekonomice. Konsolidace nemá vliv na hrubou přidanou hodnotu v běžných cenách, ovšem HPH ve stálých cenách předchozího roku je ovlivněna z důvodu použití různých cenových indexů pro vstupy a výstupy.

6. Literatura

[1]Evropský systém účtů (ESA 1995). ČSÚ, Praha 2000.
[2]FISCHER, Jan, FISCHER, Jakub. Měříme správně hrubý domácí produkt? In: Statistika. 2005, roč. 42, č. 3, s. 177–187. ISSN 0322-788X.
[3]Handbook on price and volume measures in national accounts. Eurostat. 2001.

[4]HRONOVÁ, S., HINDLS, R.: Národní účetnictví - koncept a analýzy. C. H. Beck. Praha. 2000.
[5]JEDLIČKOVÁ, E., KONVIČKA, S., MUSIL, P., SIXTA, J. Nové metody deflace HDP. In: Statistika. 2009, roč. 89, č. 3, s. 193–206. ISSN 0322-788X.
[6]CHEMIŠINEC, I., RODRYČ, P. a KOL. Změna dodavatele elektrické energie v České republice z pohledu Operátora trhu s elektřinou. [online]. Dostupný z WWW: <www.pro-energy.cz/clanky1/2.pdf>
[7]Národní zpráva České republiky o elektroenergetice a plynárenství za rok 2006.
[8]Postup přepočtu tabulek dodávek a užití do stálých cen. ČSÚ. Praha. 2008.
[9]ONDRUŠ, V. Výpočet hrubého domácího produktu a jeho revize. In: Statistika. 2003, roč. 83, č. 3, s. 24-43. ISSN 0322-788x.
[10] SIXTA, J. Odhady spotřeby fixního kapitálu. In: Statistika. 2007, roč. 87, č. 2, s. 156–163. ISSN 0322-788X.
[11] www.bilancnicentrum.cz
[12] www.eru.cz

Adresa autorů:

Jaroslav Zbranek, Ing.
Vysoká škola ekonomická v Praze
Fakulta informatiky a statistiky
Katedra ekonomické statistiky
nám. W. Churchilla 4, 130 67 Praha 3
zbranek.jaroslav@gmail.com

Musil Petr, Ing.
Vysoká škola ekonomická v Praze
Fakulta informatiky a statistiky
Katedra ekonomické statistiky
nám. W. Churchilla 4, 130 67 Praha 3
petr-musil@post.cz

Odhad vybraných ukazovateľov chudoby a ich štandardných chýb na regionálnej úrovni SR v prostredí R

Estimation of Selected Poverty Indicators and their Standard Errors on Regional Level in Slovakia with R

Tomáš Želinský

Abstract: The aim of paper is to perform poverty analysis on regional level of Slovakia with **R**. **R** is an open source programming environment for statistical computations. Basic poverty measures of Foster-Greer-Thorbecke group of poverty measures together with estimation of standard errors and coefficients of variation are estimated. Weights of households are considered in the calculations. Simple programs to perform calculations in **R** are alleged. Results for the NUTS 3 level of Slovak regions according to the EU SILC 2007 data are indicated.

Key words: poverty, **R**, EU SILC, bootstrapping.

Kľúčové slová: chudoba, **R**, EU SILC, bootstrapping.

1. Úvod

R [1] je jazyk a prostredie na uskutočňovanie štatistických výpočtov a grafické prezentovanie výsledkov. **R** poskytuje širokú škálu štatistických a grafických nástrojov, jeho moduly sa pravidelne dopĺňajú a odstraňujú prípadné nedokonalosti.

Za významnú výhodu tohto jazyka možno považovať skutočnosť, že ide o voľne šíriteľný softvér (*open source*) šírený v zmysle Všeobecnej verejnej licencie GNU. Program je podporovaný takmer všetkými bežne používanými operačnými systémami (UNIX, Linux, Windows, MacOS).

Okrem bežne používaných štatistických metód a neustále sa dopĺňajúcich nástrojov v podporných balíkoch má **R** veľmi dobré možnosti na programovanie vlastných nástrojov a metód [2].

Cieľom príspevku je ilustrovať využitie softvéru **R** na analýzu chudoby, konkrétne na odhad vybraných mier chudoby (zo skupiny Fosterových-Greerových-Thorbeckeových mier chudoby) vrátane odhadu štandardných chýb a variačných koeficientov odhadnutých ukazovateľov.

2. Základné charakteristiky chudoby

Z veľkého množstva mier chudoby v príspevku popíšeme skupinu Fosterových-Greerových-Thorbeckeových mier chudoby [3] založených na predpoklade známeho rozdelenia zvoleného indikátora chudoby. V rámci tejto skupiny mier za najznámejšie považujeme *mieru rizika chudoby* (podiel chudobného obyvateľstva), *index priepasti chudoby* a *Fosterovu-Greerovu-Thorbeckeovu mieru vážnosti („drsnosti“) chudoby*.

2.1 Miera rizika chudoby (*H*)

Miera rizika chudoby (head-count index) udáva podiel obyvateľstva ohrozeného rizikom chudoby a možno ho zapísať nasledovne:

$$H = \frac{q}{n} \tag{1}$$

kde q je počet jednotlivcov, u ktorých platí $y \leq z$, t. j. hodnota ukazovateľa blahobytu dosahuje nanajvýš hodnotu zvolenej hranice chudoby,
n je počet obyvateľov krajiny (komunity a pod.).

Ide o ľahko pochopiteľnú mieru chudoby. Má dobré použitie pri porovnávaní (napr. úspechu v znižovaní chudoby po uplatnení určitej politiky; medzi krajinami a pod.). No má aj svoje nevýhody – napríklad pri sledovaní účinkov konkrétnej politiky. Ak sa vybraná chudobná osoba stane ešte chudobnejšou (čiže hodnota jej spotreby, resp. iného ukazovateľa sa zníži), v indexe H sa to neprejaví. Znamená to, že nie je úplne citlivý na zmenu v hĺbke (priepasti) chudoby. (Tento nedostatok odstraňuje napríklad index PG).

2.2 Index priepasti chudoby (PG)

Hĺbkou (priepasťou) chudoby sa myslí vzdialenosť úrovne blahobytu y_i jednotlivca i od určenej hranice chudoby z, čo možno zapísať $(z - y_i)$. Relatívnou vzdialenosťou potom rozumieme $\frac{z - y_i}{z}$, resp. $1 - \frac{y_i}{z}$, čo je základom pre určenie ďalšej miery chudoby. Relatívna priepasť chudoby jednotlivca nadobúda hodnoty z intervalu [0; 1]. Napríklad hodnota 0,7 znamená, že osoba dosahuje len 30% minimálneho stanoveného príjmu, spotreby, príp. iného indikátora.

Index priepasti chudoby je mierou hĺbky chudoby, ktorá je založená na agregovanom deficite chudoby chudobných v pomere k hranici chudoby. Podáva dobrý obraz o hĺbke chudoby, ktorá závisí na vzdialenosti hodnoty indikátora od hranice chudoby.

Index priepasti chudoby PG potom môžeme vypočítať aj ako:

$$PG = \frac{1}{n} \sum_{i=1}^{q} \frac{z - y_i}{z} \tag{2}$$

Zo zápisu (2) je zrejmé, že ukazovateľ PG vyjadruje priemernú relatívnu priepasť chudoby obyvateľstva.

2.2 Fosterova-Greerova-Thorbeckeova miera vážnosti (drsnosti) chudoby (P_2)

Miera vypovedá o chudobe najchudobnejších, teda o „drsnosti“, resp. vážnosti chudoby a nepriamo vyjadruje nerovnosť medzi chudobnými. Analogicky ako PG môžeme aj P_2 zapísať ako:

$$P_2 = \frac{1}{n} \sum_{i=1}^{q} \left(\frac{z - y_i}{z} \right)^2 \tag{3}$$

Miera má využitie napríklad pri porovnávaní politík, ktorých cieľom je zasiahnuť skupinu najchudobnejších, ale sa ťažko interpretuje a v súčasnosti sa v praxi málo používa.

Pre lepšie pochopenie skutočností, aké odrážajú jednotlivé miery, ilustrujeme výpočet mier na jednoduchom príklade. Majme rozdelenia spotreby obyvateľov v troch lokalitách: A(20, 40, 60, 80), B(40, 40, 40, 90) a C(50, 50, 50, 100). Definujeme hranicu chudoby vo všetkých lokalitách ako $z = 60$.

Ak porovnáme úroveň chudoby v lokalitách na základe ukazovateľa podielu chudobných (miery rizika chudoby) H, dospeli by sme k záveru, že úroveň chudoby je vo všetkých lokalitách rovnaká: $H_A = H_B = H_C = 0{,}75$.

Ďalšou možnosťou je zhodnotenie situácie na základe ukazovateľa indexu priepasti (hĺbky) chudoby PG. Dostávame: $PG_A = PG_B = 0{,}25$ a $PG_C = 0{,}125$. Priemerná relatívna priepasť chudoby je v prvých dvoch lokalitách vyššia ako v lokalite C. Na základe výpočtu môžeme tvrdiť, že v A a B je úroveň chudoby rovnaká a je vyššia ako v C.

Porovnaním spotreby najchudobnejšej osoby v každej z týchto lokalít zisťujeme, že najchudobnejšia osoba v *A* dosahuje len 50% spotreby najchudobnejšej osoby v *B* a len 40% úrovne spotreby najchudobnejšej osoby v *C*. Môžeme tak predpokladať, že chudoba je „najdrsnejšia“ práve v *A*. Na overenie tohto predpokladu vypočítame hodnotu indexu P_2 pre každú z lokalít: $P_{2A} \approx 0{,}139$; $P_{2B} \approx 0{,}083$ a $P_{2C} \approx 0{,}021$. Skutočne môžeme tvrdiť, že chudoba je najvážnejšia (najdrsnejšia) v lokalite *A*.

Jednoduchým príkladom sme chceli poukázať na vážnosť voľby konkrétnej miery chudoby, nakoľko každá zo spomínaných mier pokrýva iný rozmer chudoby.

3. Metódy

Vstupnými údajmi boli mikroúdaje zisťovania EU SILC 2007 [4]. Za hranicu chudoby považujeme národnú hranicu chudoby (t. j. 60% mediánu ekvivalentného disponibilného príjmu za celú SR).

Za indikátor blahobytu budeme v modeli pokladať *ekvivalentný disponibilný príjem domácnosti*, t. j. podiel celkového ročného disponibilného príjmu domácnosti (premenná HY020) a ekvivalentnej veľkosti domácnosti (premenná HX050). Kvôli korigovaniu neodpovedí jednotlivcov v rámci domácnosti bola premenná HY020 vynásobená inflačným faktorom (premenná HY025).

Domácnosťou rozumieme hospodáriacu domácnosť, t. j. súkromnú domácnosť tvorenú osobami v byte, ktoré spoločne žijú a spoločne hospodária, vrátane spoločného zabezpečovania životných potrieb. Za znak spoločného hospodárenia sa považuje spoločná úhrada základných výdavkov domácnosti (strava, úhrada nákladov za bývanie, elektrina, plyn a pod.).

Pri výpočte bola zohľadnená prierezová váha domácností (premenná DB090) [5], a preto pôvodné vzťahy (1), (2) a (3) bolo potrebné upraviť a nadefinovať v prostredí **R**.

Na odhad štandardných chýb a variačných koeficientov odhadov bola použitá metóda *bootstrapping* (pozri napr. [6]), pričom bolo použitých 10 000 replikácií, celá procedúra bola uskutočnená v prostredí softvéru **R** [1].

Miera rizika chudoby (H):

$$H = \frac{\sum_{i=1}^{q} w_i}{\sum_{i=1}^{n} w_i} \tag{4}$$

kde w_i je prierezová váha *i*-tej domácnosti,
q je počet domácností, pre ktoré platí: $y_i \leq z$,
n je počet sledovaných domácností.

Predpokladajme, že v externom súbore „sr2007.csv“ máme uložené vstupné údaje o príjme a prierezových váhach domácností zo zisťovania EU SILC 2007. V prostredí **R** sme následne naprogramovali celý postup výpočtu:

```
> sr <- read.csv2("sr2007.csv")
> y <- sr$prijem          # ekv. disp. prijem
> w <- sr$vaha            # prierezova vaha domacnosti
> z <- 88722              # definovanie hranice chudoby
> y1 <- sr$prijem         # pomocny vektor prijmov
> y1[y1 <= z] <- 1        # vektor s bin. premennou (chudobne domacnosti)
> y1[y1 > z] <- 0         # vektor s bin. premennou (nechudobne domacnosti)
> sucin <- y1*w           # zohladnenie vah chudobnych domacnosti
> hc <- function(x) sum(x)/sum(w)  # definovanie ukazovatela chudoby H
> resamples.h <- lapply(1:10000, function(i) sample(sucin, replace = T))
> r.hc <- sapply(resamples.h, hc)  # bootstrapping odhadu
> mean(r.hc)                       # odhad strednej hodnoty ukazovatela
> sqrt(var(r.hc))                  # odhad S. E. odhadu
> sqrt(var(r.hc))/mean(r.hc)*100   # odhad koeficientu variacie
```

Index priepasti chudoby (PG):

$$PG = \frac{1}{\sum_{j=1}^{n} w_j} \sum_{i=1}^{q} \frac{z - y_i}{z} w_i \tag{5}$$

kde w_i je prierezová váha *i*-tej domácnosti,
q je počet domácností, pre ktoré platí: $y_i \leq z$,
n je počet sledovaných domácností,
z je hranica chudoby,
y_i je ekvivalentný disponibilný príjem *i*-tej domácnosti.

V prostredí **R** sme naprogramovali celý postup výpočtu:

```
> sr <- read.csv2("sr2007.csv")
> y <- sr$prijem          # ekv. disp. prijem
> w <- sr$vaha            # prierezova vaha domacnosti
> z <- 88722              # definovanie hranice chudoby
> y1 <- sr$prijem         # pomocny vektor prijmov
> y1[y1 <= z] <- 1        # vektor s bin. premennou (chudobne domacnosti)
> y1[y1 > z] <- 0         # vektor s bin. premennou (nechudobne domacnosti)
> sucin <- y1*w           # zohladnenie vah chudobnych domacnosti
> rozdiel <- (z - y)      # citatel zlomku v algebrickom sucte vztahu (5)
> sucin2 <- rozdiel*sucin # zohladnenie odpovedi chudobnych domacnosti
> pg <- function(x) sum(x)/(sum(w)*z)  # definovanie ukazovatela PG
> resamples.h <- lapply(1:10000, function(i) sample(sucin2, replace = T))
> r.pg <- sapply(resamples.h, pg)  # bootstrapping odhadu
> mean(r.pg)                       # odhad strednej hodnoty ukazovatela
> sqrt(var(r.pg))                  # odhad S. E. odhadu
> sqrt(var(r.pg))/mean(r.pg)*100   # odhad koeficientu variacie
```

Miera drsnosti chudoby (P_2):

$$P_2 = \frac{1}{\sum_{j=1}^{n} w_j} \sum_{i=1}^{q} \left(\frac{z - y_i}{z} \right)^2 w_i \tag{6}$$

kde w_i je prierezová váha *i*-tej domácnosti,
q je počet domácností, pre ktoré platí: $y_i \leq z$,
n je počet sledovaných domácností,
z je hranica chudoby,
y_i je ekvivalentný disponibilný príjem *i*-tej domácnosti.

V prostredí **R** sme naprogramovali celý postup výpočtu:

```
> sr <- read.csv2("sr2007.csv")
> y <- sr$prijem          # ekv. disp. prijem
> w <- sr$vaha            # prierezova vaha domacnosti
> z <- 88722              # definovanie hranice chudoby
> y1 <- sr$prijem         # pomocny vektor prijmov
> y1[y1 <= z] <- 1        # vektor s bin. premennou (chudobne domacnosti)
> y1[y1 > z] <- 0         # vektor s bin. premennou (nechudobne domacnosti)
> sucin <- y1*w           # zohladnenie vah chudobnych domacnosti
> rozdiel <- (z - y)^2    # citatel zlomku v algebrickom sucte vztahu (6)
> sucin2 <- rozdiel*sucin # zohladnenie odpovedi chudobnych domacnosti
> p2 <- function(x) sum(x)/(sum(w)*z^2) # definovanie ukazovatela P2
> resamples.h <- lapply(1:10000, function(i) sample(sucin2, replace = T))
> r.p2 <- sapply(resamples.h, p2)  # bootstrapping odhadu
> mean(r.p2)                       # odhad strednej hodnoty ukazovatela
> sqrt(var(r.p2))                  # odhad S. E. odhadu
> sqrt(var(r.p2))/mean(r.p2)*100   # odhad koeficientu variacie
```

4. Výsledky

S využitím uvedených postupov dostávame nasledovné odhady:

Tabuľka 1: Odhady charakteristík chudoby pre kraje SR

Kraj	*Hodnoty ukazovateľov pre EU SILC 2007 [%]*								
	H	**S. E.**	**CV**	***PG***	**S. E.**	**CV**	P_2	**S. E.**	**CV**
SR	10,65	0,51	4,91	2,69	0,18	6,78	1,12	0,10	9,26
BA	6,02	1,26	20,92	1,96	0,54	27,72	1,04	0,38	36,37
BB	12,53	1,50	11,97	2,89	0,46	16,06	1,13	0,28	24,36
KE	12,05	1,52	12,72	3,12	0,56	18,37	1,33	0,32	24,28
NR	11,91	1,42	11,91	3,36	0,48	15,95	1,34	0,28	20,77
PO	15,52	1,76	11,34	4,08	0,60	14,62	1,64	0,31	17,73
TN	6,96	1,13	16,87	1,36	0,34	25,14	0,50	0,18	35,55
TT	8,04	1,31	16,32	1,84	0,44	24,08	0,77	0,30	39,05
ZA	8,46	1,22	14,35	2,02	0,41	20,30	0,95	0,27	28,08

Zdroj: vlastné výpočty

Z výsledkov vyplýva, že situácia je najlepšia na západe Slovenska a najhoršia na východnom Slovensku.

5. Záver

Cieľom článku bolo ukázať možnosti softvéru **R** pri odhade vybraných mier chudoby vrátane odhadov ich chýb. Po charakteristike mier chudoby zo skupiny Fosterových-Greerových-Thorbeckeových mier chudoby sme uviedli modifikované vzťahy pre ich výpočet po zohľadnení prierezových váh domácností. Následne sme uviedli jednoduché programy na uskutočnenie odhadu v prostredí **R**. Výsledky podľa uvedených vzťahov boli vypočítané pre regióny SR na úrovni NUTS3 podľa údajov EU SILC 2007.

Literatúra

[1] R DEVELOPMENT CORE TEAM: *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing, 2008. ISBN 3-900051-07-0, URL http://www.R-project.org.

[2] TARTAĽOVÁ, A.: Aktuárske výpočty v prostredí programu R. In: *Forum Statisticum Slovacum*. roč. 5, č. 7 (2009). ISSN 1336-7420.

[3] FOSTER, J., GREER, J., THORBECKE, E.: A Class of Decomposable Poverty Measures. In: *Econometrica*. roč. 52. č. 3 (máj, 1984). s 761 – 766. ISSN 0012-9682.

[4] ŠTATISTICKÝ ÚRAD SR: *EU SILC 2007: UDB verzia 20/08/2008*. Bratislava: Štatistický úrad SR, 2008.

[5] ŠTATISTICKÝ ÚRAD SR: *EU SILC 2006: Zisťovanie o príjmoch a životných podmienkach domácností v SR*. Bratislava: Štatistický úrad SR, 2007.

[6] EFRON, B., TIBSHIRANI, R. J.: *An introduction to the bootstrap*. New York: Chapman & Hall, 1993. ISBN 0-412-04231-2.

Adresa autora:

Tomáš Želinský, Ing. PhD.
Ekonomická fakulta, TU Košice
Němcovej 32, 040 01 Košice
tomas.zelinsky@tuke.sk

Príspevok bol napísaný s podporou Vedeckej grantovej agentúry MŠ SR a SAV v rámci riešenia vedecko-výskumného projektu VEGA 1/0370/08 Regionálny prístup k meraniu sociálneho kapitálu a chudoby.

Vývoj a cielenosť sociálnych dávok pre domácnosti s deťmi
The development and mainstreaming of the state social benefit for houshold with dependent children

Vladimíra Želonková

Abstract: The contribution contains the analysis of the development of state social benefits for households with dependent children in Household Budget Survey. The alignment chart shows the development trend in the period 2005 – 2008. Definitions of state social benefits want to point out their importance in social policy in the Slovak republic.

Key words: households by number of dependent children, dependent child, state social benefits, children´s allowance, parental allowance

Kľúčové slová: domácnosti podľa počtu nezaopatrených detí, nezaopatrené dieťa, štátne sociálne dávky, prídavok na dieťa, rodičovský príspevok

1. Úvod

Cieľom príspevku je opísať význam sociálnych dávok v rámci sociálnej politiky pre rodiny s nezaopatrenými deťmi. Na základe tabuliek a grafov z údajov štatistiky rodinných účtov v priebehu rokov 2005 až 2008 je vidieť nárast týchto príjmov v závislosti od pravidelného zvyšovania peňažných dávok, najmä prídavku na dieťa a rodičovského príspevku. V štatistike rodinných účtov je tiež možné pozorovať každoročný nárast podielu peňažnej pomoci pre rodiny s deťmi na čistých peňažných príjmoch.

2. Štátne sociálne dávky a ich vývoj

Štátne sociálne dávky zabezpečujú poskytovanie dávok štátnej sociálnej podpory, ktoré sú najrozsiahlejšou formou finančnej podpory štátu orientovanej najmä na rodiny s nezaopatrenými deťmi. Štátne sociálne dávky a príspevky sú financované zo štátneho rozpočtu. Prostredníctvom nich sa štát podieľa na riešení rôznych životných situácií rodiny, ktoré tvoria prirodzenú súčasť rodinného životného cyklu, napr. príchod dieťaťa do rodiny, výchova a osobná starostlivosť o dieťa v rodine, jeho príprava na povolanie alebo svojou prítomnosťou špecificky menia spôsob života jednotlivca či rodiny, napr. prítomnosť dieťaťa s ťažkým postihnutím v rodine, prechodná alebo dlhodobá neprítomnosť jedného z rodičov v rodine.

Štátne sociálne dávky prešli svojim vývojom. Od 1.1.1995 boli vyčlenené z nemocenského poistenia. V oblasti sociálnej starostlivosti nastali zmeny v dôsledku transformácie sociálneho zabezpečenia prijatím zákona č.195/1998 Z.z. o sociálnej pomoci. Došlo k systémovej zmene v oblasti poskytovania dávok sociálne odkázaným občanom. Dovtedy poskytované dávky sociálnej starostlivosti, ktoré boli adresované rôznym skupinám obyvateľstva, boli pretransformované do jedinej dávky sociálnej pomoci.[11]

Dňom 1.1.2009 nadobudli účinnosť nové zákony, resp. novely zákonov aj pre rodiny s nezaopatrenými deťmi, patrí sem novela zákona o príspevku pri narodení dieťaťa a príspevok na starostlivosť o dieťa.

[11] ŠÚSR - Databáza časových radov SLOVSTAT

3. Štátne sociálne dávky pre domácnosti s deťmi

V rámci štátnej sociálnej podpory sa v súčasnosti na Slovensku poskytujú tieto štátne sociálne dávky pre rodiny s deťmi:

1. príspevok pri narodení dieťaťa
2. príplatok k príspevku pri narodení dieťaťa
3. príspevok rodičom, ktorým sa súčasne narodili tri alebo viac detí alebo ktorým sa v priebehu dvoch rokov opakovane narodili dvojčatá
4. rodičovský príspevok
5. príspevok na starostlivosť o dieťa
6. prídavok na dieťa
7. príplatok k prídavku na dieťa
8. príspevky na podporu náhradnej starostlivosti o dieťa
9. príspevok na pohreb

1. Príspevok pri narodení dieťaťa zákon č. 235/1998 Z.z.

Príspevok pri narodení dieťaťa je štátna sociálna dávka, ktorou štát prispieva na pokrytie výdavkov spojených so zabezpečením nevyhnutných potrieb novorodenca.

- výška dávky: **151,37 €** / 4 560 Sk
- typ dávky: štátna sociálna dávka jednorazová
- poskytovateľ: ÚPSVaR príslušný podľa miesta trvalého pobytu oprávnenej osoby

2. Príplatok k príspevku pri narodení dieťaťa zákon č. 235/1998

Príplatok k príspevku je štátna sociálna dávka, ktorou štát pripláca na zvýšené výdavky spojené so zabezpečením nevyhnutných potrieb dieťaťa, ktoré sa matke narodilo ako prvé dieťa, druhé dieťa alebo tretie dieťa, ktoré sa dožilo aspoň 28 dní.[12]

- výška dávky: **678,49 €** / 20 440 Sk
- typ dávky: štátna sociálna dávka jednorazová od 1.1.2009 na každé dieťa
- poskytovateľ: ÚPSVaR

3. Príspevok rodičom , ktorým sa súčasne narodili tri alebo viac detí alebo ktorým sa v priebehu dvoch rokov opakovane narodili dvojčatá zákon č.235/1998 Z.z.

Príspevok rodičom je štátna sociálna dávka, ktorou štát prispieva raz za rok rodičom alebo osobe, ktorá prevzala deti do starostlivosti nahrádzajúcej starostlivosť rodičov na základe právoplatného rozhodnutia súdu, na zvýšené výdavky, ktoré vznikajú v súvislosti so starostlivosťou o súčasne narodené tri deti alebo viac detí alebo v priebehu dvoch rokov opakovane narodené dvojčatá alebo viac detí súčasne.

- výška dávky do 6 rokov **81, 99 € /** 2 470 Sk
  od 6 do 15 rokov **101,25 € /** 3 050 Sk
  vo veku 15 rokov **107,55 € /** 3 240 Sk
- typ dávky štátna sociálna dávka jednorazová
- poskytovateľ ÚPSVAR

4. Rodičovský príspevok zákon č.280/2002 Z.z.

Rodičovský príspevok je štátna sociálna dávka, ktorou štát prispieva rodičovi na zabezpečovanie riadnej starostlivosti o dieťa do troch rokov veku dieťaťa, s dlhodobo

[12] MPSVR – dávkový informačný servis

nepriaznivým zdravotným stavom do šiestich rokov veku dieťaťa alebo do šiestich rokov veku dieťaťa, ktoré je zverené do starostlivosti nahrádzajúcej starostlivosť rodičov, najdlhšie tri roky od právoplatnosti rozhodnutia súdu o zverení dieťaťa do tejto starostlivosti.

- výška dávky **164,22 €** / 4 947 Sk od 1.9.2009 (opatrenie MPSVR SR č.326/2009)
- typ dávky štátna sociálna dávka opakovaná
- poskytovateľ ÚPSVAR

5. Príspevok na starostlivosť o dieťa zákon č. 561/2008 Z.z.

Poskytovaním príspevku štát prispieva rodičovi alebo fyzickej osobe, ktorej je dieťa zverené do starostlivosti nahrádzajúcej starostlivosť rodičov (ďalej len „rodič"), na úhradu výdavkov vynaložených na starostlivosť o dieťa.

- výška dávky v sume preukázaných výdavkov, najviac v sume rodičovského príspevku
- typ dávky štátna sociálna dávka opakovaná
- poskytovateľ ÚPSVAR

6. Prídavok na dieťa zákon č. 600/2003 Z.z.

Prídavok na dieťa (ďalej "prídavok") je štátna sociálna dávka, ktorou štát prispieva oprávnenej osobe na výchovu a výživu nezaopatreného dieťaťa.

- výška dávky **21,25 €** / 640 Sk
- typ dávky štátna sociálna dávka opakovaná
- poskytovateľ ÚPSVAR

7. Príplatok k prídavku na dieťa zákon č. 600/2003 Z.z.

Príplatok k prídavku na dieťa (ďalej len "príplatok k prídavku") je štátna sociálna dávka, ktorou štát pripláca oprávnenej osobe k prídavku na dieťa na výchovu a výživu nezaopatreného dieťaťa, na ktoré nemožno uplatniť daňový bonus podľa osobitného predpisu.

- výška dávky **9,96 €/** 300 Sk
- typ dávky štátna sociálna dávka opakovaná
- poskytovateľ ÚPSVAR

8. Príspevky na podporu náhradnej starostlivosti o dieťa zákon č.627/2005 Z.z.

Zákon o príspevkoch na podporu náhradnej starostlivosti o dieťa upravuje poskytovanie finančných príspevkov ako sociálnych dávok, ktorými štát podporuje náhradnú starostlivosť o dieťa, ak náhradnú starostlivosť o dieťa vykonáva osobne na základe rozhodnutia súdu alebo na základe rozhodnutia príslušného orgánu iná fyzická osoba ako rodič

9. Príspevok na pohreb zákon č. 238/1998 Z.z.

Príspevok na pohreb je štátna sociálna dávka, ktorou štát prispieva na úhradu výdavkov spojených so zabezpečením pohrebu zomretého.

- výška dávky **79,68 €/** 2 400 Sk
- typ dávky štátna sociálna dávka jednorazová
- poskytovateľ ÚPSVAR

4. Štátne sociálne dávky a ich cielenosť

Základným cieľom systému štátnej sociálnej podpory je formou dávok cielene a diferencovane poskytovať podporu rodinám nachádzajúcim sa v určitých sociálnych situáciách (tam, kde štát má záujem stimulovať, podporiť a chrániť) a tým prispievať k úhrade

nákladov spôsobených určitými sociálnymi udalosťami, ale nezabezpečovať ich úplnú náhradu.[13]

V rámci štátnej podpory pre rodiny s deťmi je **prídavok na dieťa** najväčšou štátnou sociálnou dávkou. V roku **2005** predstavoval 55,1 % z objemu štátnych sociálnych dávok. Na dávku bolo vyplatených 8 676 mil. Sk.

Na druhú objemovo najväčšiu dávku **rodičovský príspevok** bolo v roku **2005** vyplatených 6 531 mil. Sk. Výdavky na túto dávku vzrástli o 2,8 % oproti roku 2004, kedy bolo vyplatených 5 790 mil. Sk. V tabuľke č.1 môžeme vidieť tento trend v priebehu období rokov 2005-2008.

Tabuľka č.1: Vývoj finančných prostriedkov na prídavok na dieťa a rodičovský príspevok

Štátna sociálna dávka	*2005*	*2006*	*2007*	*2008*
Prídavok na dieťa	8 676 mil. Sk	8 462 mil. Sk.	8 254 mil. Sk	8 065 mil. Sk
Rodičovský príspevok	6 531 mil. Sk.	7 059 mil. Sk	7 370 mil. Sk	7 558 mil. Sk

Tabuľka č.2: Prídavok na dieťa a rodičovský príspevok v období rokov 2004 – 2009

Štátna sociálna dávka	*2004*	*2005*	*2006*	*2007*	*2008*	*2009*
Prídavok na dieťa	500 Sk	540 Sk	540 Sk	540 Sk	540 Sk	640 Sk
Rodičovský príspevok	4 110 Sk	4 230 Sk	4 440 Sk	4 560 Sk	4 780 Sk	4 947 Sk

5. Spôsob zisťovania štátnych sociálnych dávok v štatistike rodinných účtov

Štátne sociálne dávky sa sledujú v štatistike rodinných účtov ako sociálne príjmy v súkromných domácnostiach na minimálnej vzorke, približne 4 700 jednotiek za rok. V sociálnych príjmoch sú zahrnuté starobné dôchodky, iné dôchodky, dávky v chorobe, podpora v nezamestnanosti, dávky sociálnej pomoci, peňažná pomoc rodinám s deťmi, iné sociálne príjmy.

Informácie o sociálnych príjmoch sú zisťované v špeciálnych formulároch, kde si spravodajská jednotka zaznamenáva bežné výdavky a príjmy. Údaje o príjmoch a výdavkoch sa vyhodnocujú na úrovni celej domácnosti spolu za obdobie dvoch mesiacov.

Formulár – mesačný rozpočet domácnosti RÚ 6-12

V module 1340 sa zisťujú u domácností s deťmi za sledovaný mesiac peňažné príjmy domácností, patria sem sociálne príjmy všetky pravidelné a nepravidelné príjmy, ktoré dostal ktorýkoľvek člen domácnosti od Sociálnej poisťovne, obecného mestského úradu, úradov práce a sociálnych vecí a rodiny (prídavok na dieťa, daňový bonus, materské).

Formulár – denník spravodajskej domácnosti RÚ 4-12

V module 1308 pre domácností s deťmi sa zaraďujú nárokovateľné sociálne príjmy členov ako materské a rodičovský príspevok. V module 1309 sa sledujú sociálne peňažné príjmy domácmostí s deťmi za domácnosť, sem patrí peňažná pomoc rodinám s deťmi, t.j. prídavok na dieťa, príspevok pri narodení dieťaťa, vyrovnávacia dávka v materstve a tehotenstve, vr. príplatkov k príspevku, príspevok pestúna, daňový bonus a iné peňažné dávky podporujúce uhradiť špecifické náklady súvisiace s potrebami rodín pri výchove detí a iné sociálne príjmy, vr. príspevku na pohreb.

[13] Vojtech Stanek a kolektív Sociálna politika Bratislava, r. 2002

Podkladom na analýzu vývoja štátnych sociálnych dávok v domácnostiach s deťmi boli publikácie Príjmy, výdavky a spotreba súkromných domácností SR za roky 2005 až 2008. Každý rok sa v nich publikujú údaje o príjmoch a výdavkoch súkromných (vrátane sociálnych príjmov).Do náhodného výberu sa dostávajú súkromné domácnosti aj podľa počtu nezaopatrených detí .[14]

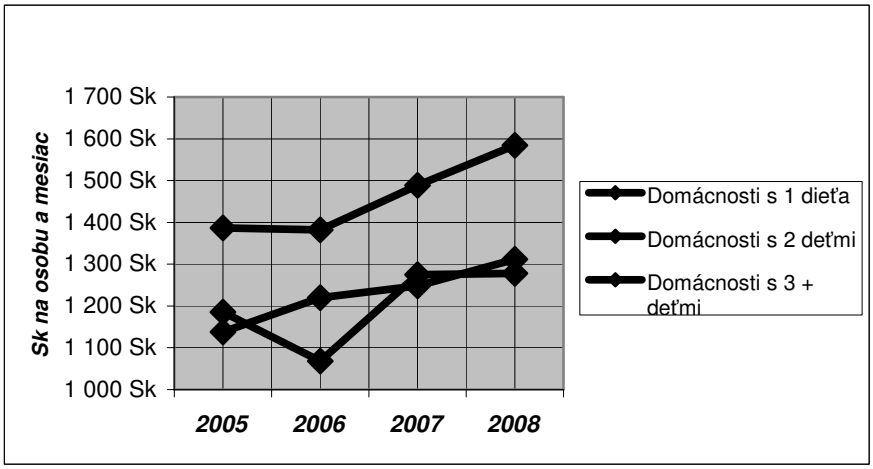
Z grafu na obrázku č.1 je možné vyčítať, že viacdetné domácnosti s vyšším počtom detí napriek tomu, že dostávajú vyššiu sumu finančných prostriedkov v rámci ŠSD ako domácnosti s 1 dieťaťom, pri prepočte na 1 osobu v domácnosti dostávajú menej ako rodiny 1 dieťaťom.

Tabuľka č.3: Peňažná pomoc rodinám podľa počtu nezaopatrených detí 2005 – 2008

Štátna sociálna dávka	*2005*	*2006*	*2007*	*2008*
domácnosť 1 dieťa	5 474 Sk	5 825 Sk	6 774 Sk	7 405 Sk
domácnosť 2 deti	6 876 Sk	7 616 Sk	8 369 Sk	8 699 Sk
domácnosť 3 + detí	7 958 Sk	8 427 Sk	9 903 Sk	10 221 Sk

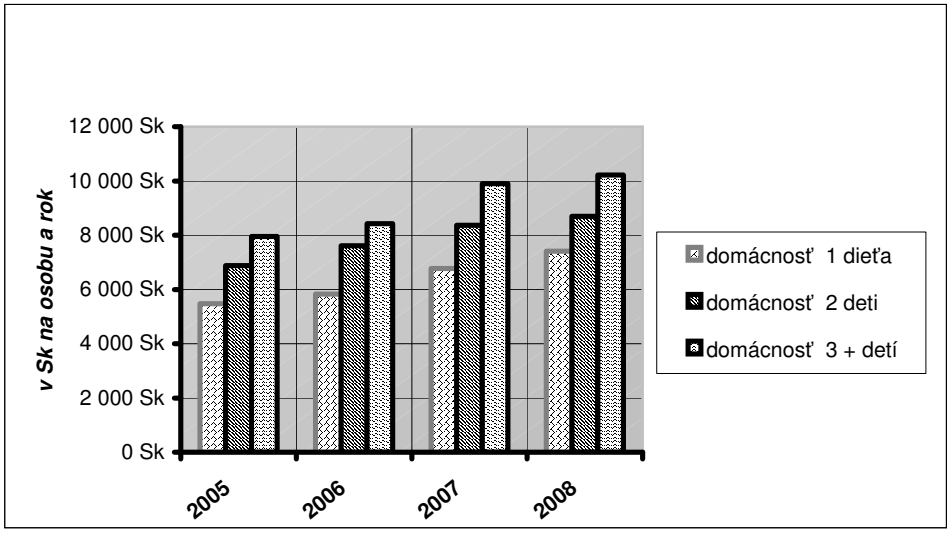
Tabuľka č.4 : Peňažná pomoc rodinám s deťmi podľa počtu nezaopatrených detí (roky 2005 – 2008)

Rok zisťovania	*2005*	*2006*	*2007*	*2008*
Veľkosť vzorky domácnosti	4 710	4 701	4 698	4 718
Čisté peňažné príjmy spolu	102 038 Sk	118 304 Sk	118 215 Sk	129 542 Sk
Peňažná pomoc rodinám s deťmi	3 082 Sk	4 578 Sk	5 053 Sk	5 476 SK
Podiel k čistým príjmom	***3,11 %***	***4,43 %***	***4,36%***	***4,30%***



Obrázok 1: Peňažná pomoc rodinám s deťmi podľa počtu nezaopatrených detí na osobu a mesiac rok 2005 až 2008.

[14] ŠÚSR Príjmy, výdavky a spotreba súkromných domácností SR, Bratislava 2009



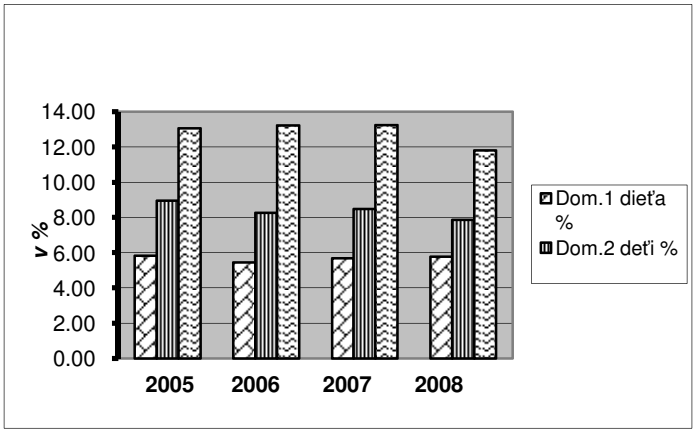
Obrázok 2: Peňažná pomoc rodinám s deťmi podľa počtu nezaopatrených detí na rok 2005 až 2008.

Graf na obrázku č.2 ukazuje percentuálny podiel pomoci rodinám s deťmi k čistým peňažným príjmom, ktorý vykazuje stúpajúcu tendenciu. Podiel peňažnej pomoci rodinám s deťmi na celkových príjmoch domácnosti s deťmi sa pohybuje na úrovni približne 3-4 %. V roku 2006 oproti predchádzajúcemu roku tento podiel vykázal výrazný nárast takmer až na 4,5 %, pričom v súčasnosti sa pohybuje na úrovni 4,3 %. Tento podiel je však stále príliš nízky na to, aby príjmy rodín s deťmi v podobe ŠSD dokázali výraznejšie ovplyvniť kvalitu ich života.

Tabuľka č.5: Peňažná pomoc rodinám s deťmi, percentuálny podiel z čistých soc .príjmov

Rok zisťovania	**2005**	**2006**	**2007**	**2008**
Čisté príjmy spolu 1	94 119 Sk	107 087 Sk	119 596 Sk	128 465 Sk
Čisté príjmy spolu 2	76 929 Sk	92 293 Sk	98 930 Sk	110 754 Sk
Čisté príjmy spolu 3	60 893 Sk	63 776 Sk	74 837 Sk	86 577 Sk
Domácnosti s 1 dieťaťom	5 474 Sk	5 825 Sk	6 774 Sk	7 405 Sk
Domácnosti s 2 deťmi	6 876 Sk	7 616 Sk	8 369 Sk	8 699 Sk
Domácnosti s 3 + deťmi	7 958 Sk	8 427 Sk	9 903 Sk	10 221 Sk
Dom. 1 dieťa %	**5,82**	**5,44**	**5,66**	**5,76**
Dom. 2 deti %	**8,93**	**8,25**	**8,46**	**7,85**
Dom. 3 + deti %	**13,06**	**13,21**	**13,23**	**11,81**

Na grafe obrázku č.3 je zaujímavé, že u domácnosti s 3 a viacerými deťmi tvorí peňažná pomoc rodinám s deťmi vyšší podiel z čistých príjmov domácnosti (približne 13 %) než u domácností s 1, prípadne 2 deťmi. To znamená, že peňažná pomoc rodinám s deťmi u viacdetných rodín predstavuje relatívne významný zdroj príjmov.



Obrázok 3: Peňažná pomoc rodinám s deťmi podľa počtu detí, percentuálny podiel z čistých sociálnych príjmov.

6. Záver

Narastajúci trend výšky a poberania sociálnych dávok pre domácností s deťmi je možné pozorovať aj v zisťovaní štatistiky rodinných účtov. Obdobie rokov 2005 – 2008 naznačuje rast objemu peňažných príjmov vyplatených vo forme ŠSD najmä prídavku na dieťa a rodičovského príspevku. Ak chceme podporiť mladých ľudí v rozhodovaní sa v prospech založenia rodiny, mali by sme vziať na vedomie skutočnosť, že v prípade, ak sa jeden z rodičov rozhodne pre celodennú starostlivosť o dieťa, dochádza k výpadku jedného pracovného príjmu a tým k obmedzeniu finančných prostriedkov na výživu a výchovu detí. Starostlivosť o dieťa, najmä v raných fázach vývoja, predstavuje investíciu do budúcnosti. Tu vidíme veľkú príležitosť prostredníctvom opatrení sociálnej politiky ovplyvniť životnú úroveň rodín s deťmi a predpokladáme, že zvýšením podpory sa vytvoria optimálne podmienky aj pre nárast pôrodnosti. Na druhej strane treba vytvoriť aj vhodné mechanizmy na zabránenie zneužívania týchto opatrení.

Literatúra

[1]Vojtech Stanek a kolektív Sociálna politika Bratislava, r. 2002

[2]Príjmy, výdavky a spotreba súkromných domácností SR,r.2005,r.2006,r.2007,r.2008 Bratislava

[3]ŠÚSR - Databáza časových radov SLOVSTAT

[4]MPSVR – dávkový informačný servis

[5]Vybrané ukazovatele zo sociálneho zabezpečenia v SR v roku 2005,2006,2007

[6]Správa o sociálnej situácii obyvateľstva SR za rok 2007,2008 MPSVR

Adresa autora:

PhDr. Vladimira Želonková
ŠÚSR, Miletičova 3, 824 67 Bratislava
Vladimira.Zelonkova@statistics.sk

PREHLIADKA PRÁC MLADÝCH ŠTATISTIKOV A DEMOGRAFOV

Analýza finančného portfólia
Analysis of financial portfolio

Peter Bialoň

Abstract: This work is aimed to bring insight into optimal portfolio selection, whereby set of fictitious market portfolios was created, in order to clarify this topic. The main contribution of this paper lies in familiarization with portfolio creation and choice process; however, a great deal of attention is also paid to parameters such as risk and return. These parameters are firstly evaluated and flowingly integrated in Markowitz's Modern Portfolio Theory. The work also concentrates on utility theory, which is applied as a tool for optimal portfolio selection. This theoretical knowledge is then used to perform the analysis of fictional portfolios, where the expected return and risk are discussed closely. In searching for effectiveness, the efficiency set is determined, which leads to optimal portfolios' findings by means of optimization and indifference curves.

Key words: financial portfolios, portfolio management, return, risk, Modern portfolio theory, attainable set, efficiency frontier, utility, indifference curves.

Kľúčové slová: finančné portfóliá, portfólio manažment, výnos, riziko, Moderná teória finančných portfólií, prípustná množina, efektívna množina, užitočnosť, indiferentné krivky.

1. Úvod

Predmetom tejto práce sú finančné portfóliá ako investičný nástroj, ktorý je svojmu držiteľovi resp. investorovi schopný prinášať úžitok v prípade efektívnej správy portfólia. Súčasná nestabilita a vysoká miera rizika sú príznačnými prívlastkami kapitálových trhov, čo si vyžaduje väčší dôraz na efektívne rozhodnutia v oblasti riadenia investícií. Prijímaním efektívnych rozhodnutí v rámci riadenia finančných portfólií sa rozumie správa jednotlivých zložiek, pričom tento proces začína ešte pred výberom konkrétnych aktív. Turbulentné prostredie na finančných a kapitálových trhoch prináša rôzne prekážky, ktorým musí investor čeliť v podobe rizika. Za účelom vyhýbania sa riziku resp. jeho minimalizácie vytvárajú racionálni investori a správcovia investícií finančné portfóliá. Správa a rozhodnutia týkajúce sa vedenia portfólia však nebývajú činnosťou jednoduchou a jednoznačnou. Primárnou snahou tejto práce je využitie teoretických prístupov v praxi a poukázanie na skutočnú účinnosť Markowitzovho (MPT) finančného modelu a teórie užitočnosti. Východiskovým bodom tejto analýzy je použitie historických dát, ktoré poskytuje portál Yahoo! s jeho aplikáciami ako vstup pre finančnú analýzu. Cieľom tejto práce je teda poskytnúť názornú ukážku voľby fiktívneho finančného portfólia pre konkrétneho investora so zreteľom na hlavné faktory, ktoré ovplyvňujú výkon a existenciu portfólia.

2. Popis analýzy

Skúmanými prvkami budú náhodne vybraté aktíva, s ktorými sa obchoduje na svetových trhoch. Vychádzajúc z finančnej aplikácie, ktorú poskytuje portál Yahoo!, bude analyzovaný vývin historických cien akcií odlišných spoločností a jednej komodity. Tieto prvky tak vytvoria fiktívne portfólio, ktorého výkon sa budem snažiť analyzovať a následne optimalizovať. Akcie spoločností boli vybraté ľubovoľne a majú napomôcť aplikovať model „Modernej Teórie Finančných Portfólií". Skupina aktív vytvára reprezentatívnu zložku, kde medzi aktívami sú zahrnuté spoločnosti zaoberajúce sa informačnými technológiami, finančnými a bankovými službami, výrobou automobilov, pričom v portfóliu sa nachádza aj

zástupca z oblasti drahých kovov. Vybrané spoločnosti boli nasledovné: International Business Machine Corp. (IBM), Goldman Sachs Group Inc. (GS), Royal Bank of Scotland Group Plc. (RBS.L) a Volkswagen AG (VW.DE). Uvedené spoločnosti sú kótované na svetových finančných trhoch a s ich akciami sa obchoduje na burzách v New Yorku (NYSE), Frankfurte (XETRA) a Londýne (LSE). V analýze je zahrnutá aj jedna komodita – zlato.

Analyzované sú historické údaje z obdobia piatich kvartálov a ako nástroj hodnotenia dát bol použitý program Microsoft Excel. Dáta použité v nasledovnej analýze sú získané z internetovej domény Yahoo! Finance a ide o súhrn denných cien cenných papierov a zlata od 1. januára 2008 do 30. apríla.2009, resp. ide o 345 denných pozorovaní vývoja cien aktív na finančných trhoch (n=345).

Primárnou podmienkou analýzy portfólia ako celku je nutnosť preskúmať a otestovať jeho jednotlivé zložky. Každé aktívum prináša do portfólia istú mieru jednak výnosu a jednak rizika, ktoré v konečnom dôsledku ovplyvňujú očakávaný celkový výnos a riziko portfólia. Z tohto dôvodu je nutné prvotne vypočítať výnosy jednotlivých aktív v skúmanom horizonte. Keďže historický vývoj cien cenných papierov jednotlivých spoločností a zlata nebol jednotvárny, je potrebné vypočítať konkrétne výnosy medzi jednotlivými časovými periódami, ktorými boli v tomto prípade dni. Podmienkou nasledovnej analýzy je však jednotnosť dát k určitému časovému okamihu. Historické výnosy boli rôznorodé a keďže išlo o aktíva, s ktorými sa obchoduje na rozličných svetových trhoch, predpokladom pre ďalšie napredovanie bola jednotnosť časového horizontu. Preto bolo potrebné dáta najskôr vytriediť, aby bol ku každému dňu priradený práve jeden historický výnos z každého aktíva. V takomto prípade sa vytvorila nová reprezentatívna vzorka (n=306), z ktorej sa ďalej vychádzalo. Z výsledkov jednotlivých peňažných výnosov boli následne vypočítané očakávané výnosy pre každú korporátnu akciu a komoditu. Obdobný postup bol použitý aj pri kalkulácii rizika aktív. Keďže riziko je definované ako disperzia jednotlivých hodnôt od očakávanej hodnoty, na výpočet miery rizika boli použité vzorce rozptylu a štandardnej odchýlky. Výsledky opísaných výpočtov sú uvedené v nasledovnej tabuľke.

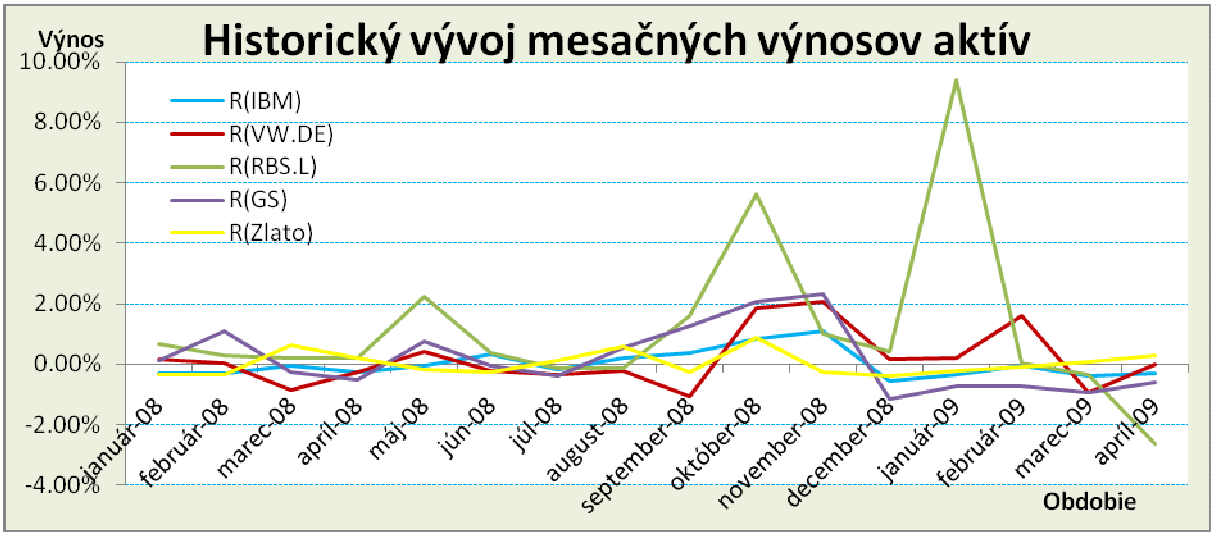
Tabuľka č. 1: Výnos a riziko jednotlivých aktív (triedená vzorka)

	Výnos a riziko jednotlivých aktív(vzorka)				
	IBM	**VW.DE**	**RBS.L**	**GS**	**Zlato**
μ (denný)	-0,0006%	0,1355%	0,7680%	0,1619%	0,0174%
μ (mesačný)	-0,0169%	4,1469%	25,7988%	4,9743%	0,5221%
μ (ročný)	-0,2031%	62,8390%	1470,7900%	79,0595%	6,4485%
σ(denné)	2,3631%	7,8498%	7,7695%	5,0864%	1,9787%
σ(mesačné)	12,9435%	42,9952%	42,5554%	27,8594%	10,8379%
σ(ročné)	44,8375%	148,9399%	147,4163%	96,5079%	37,5437%

Tabuľka zobrazuje očakávanú mieru výnosu a rizika v rôznych časových horizontoch. Napriek tomu, že analyzované historické dáta boli skúmané v dennom horizonte, vo finančnom kontexte sa väčšinou používa mesačné alebo ročné vyjadrenie. Prvé riadky tabuľky zobrazujú očakávania výnosov konkrétneho aktíva, pričom vidíme, že akcie spoločnosti IBM majú zápornú očakávanú mieru výnosu. Racionálny investor by s takýmto aktívom nemal do budúcnosti kalkulovať. Z tohto dôvodu sa môžu akcie tejto spoločnosti z nasledovnej analýzy vylúčiť, a teda ďalej figurovať nebudú.

Historický vývoj priemerných mesačných výnosov je zobrazený v grafe č. 1. Už z grafickej ukážky je možné postrehnúť istú závislosť medzi jednotlivými skupinami aktív. Existuje tu teda pravdepodobnosť, že zmena ceny jedného aktíva bude podnecovať istý pohyb

ceny druhého aktíva. Závislosť jednotlivých aktív môže byť negatívna alebo pozitívna. V prípade finančných portfólií je žiaducou závislosťou aktív práve závislosť negatívna. Negatívna preto, lebo zmena ceny jedného aktíva vyvoláva zmenu ceny iného aktíva (resp. skupiny aktív) opačného charakteru. Tento druh korelácie aktív umožňuje diverzifikačný proces finančného portfólia, a tak aj minimalizáciu rizikovosti celkového portfólia.



Graf č. 1: Historický vývoj mesačných výnosov aktív

Ďalšie skúmanie, ktoré si vyžaduje pozornosť, bolo nutné vykonať v oblasti korelácie jednotlivých aktív medzi sebou. Cieľom tohto skúmania bolo poprieť resp. potvrdiť existenciu vzájomných závislostí medzi zložkami investícií, ktorá má patričný dopad na voľbu a výkon portfólia samotného. Vychádzajúc z triedenej vzorky dát, ktorá má reprezentovať pôvodné historické dáta, môžeme ďalej analyzovať existenciu, charakter a silu závislosti jednotlivých zložiek. Už z grafu č. 1 je možné pozorovať istú mieru závislosti, ktorá vystupuje medzi aktívami. Pre hlbšie skúmanie je však nutné konkretizovať túto mieru resp. ju kvantifikovať. Na tento účel bola použitá kovariancia (Cov) a korelácia, konkrétne koeficient korelácie (Corr). Výsledkom boli dve štvorcové matice, a to korelačná a kovariančná matica, ktoré vyjadrujú existenciu, charakter a silu vzájomnej závislosti. Obe tieto štvorcové matice sú uvedené v tabuľke č. 2.

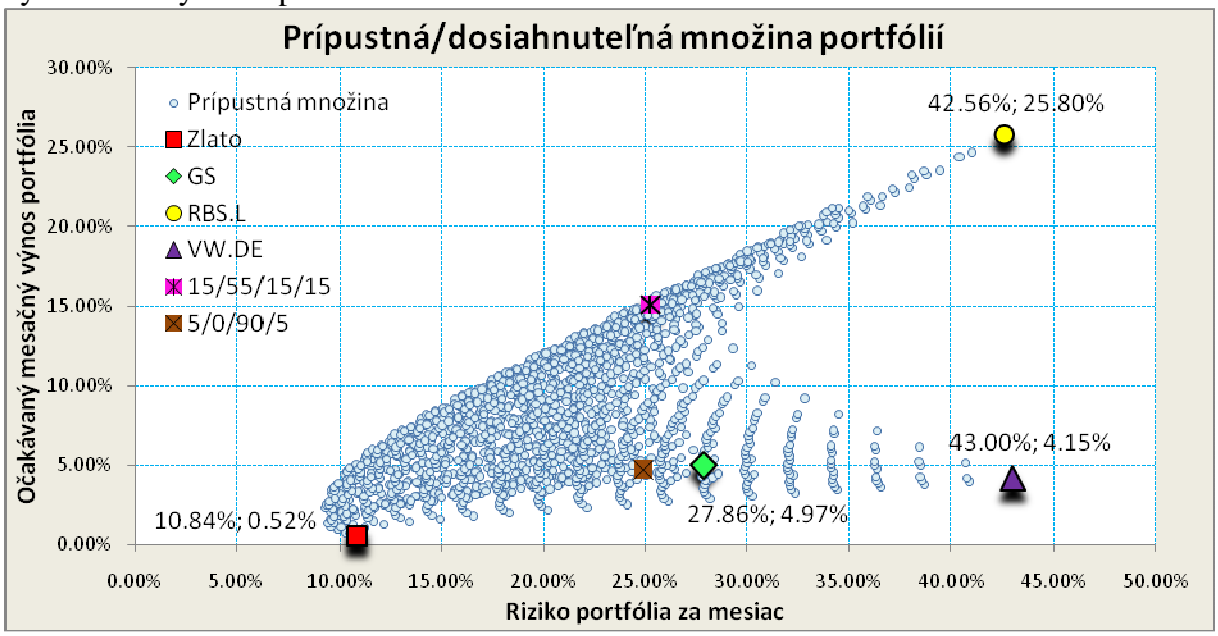
Tabuľka č. 2: Korelačná a kovariančná matica

$Corr_{ij}$ / Cov_{ij}	**(VW.DE)**	**(RBS.L)**	**(GS)**	**(Zlato)**
(VW.DE)	1,000000 0,006162	-0,000668 -0,000668	-0,106216 -0,000424	-0,007558 -0,000424
(RBS.L)	-0,109507 -0,000668	1,000000 0,006037	0,386080 0,001526	-0,093553 -0,000144
(GS)	-0,106216 -0,000424	0,386080 0,001526	1,000000 0,002587	-0,085185 -0,000086
(Zlato)	-0,007558 -0,000012	-0,093553 -0,000144	-0,085185 -0,000086	1,000000 0,000387

Výsledky tohto skúmania sa môžu javiť pozoruhodnými. Vo väčšine prípadov sme svedkami negatívnej korelácie, ktorá nastáva medzi cennými papiermi a zlatom. Tento druh závislosti je pre celkové portfólio žiaduci, pretože je schopný priniesť efektívnu možnosť optimalizácie, a teda aj diverzifikácie portfólia. Na druhej strane sa však dá tvrdiť, že

korelačné koeficienty nadobúdajú prevažne nepatrné hodnoty, čo znižuje možnosť kompenzácie nepriaznivého vývoja výnosu aktív. Jediným aktívom s pozitívnou koreláciou voči inému aktívu sú akcie spoločností Goldman Sachs a Royal Bank of Scotland. Korelácia tu je pozitívna, avšak nenadobúda veľmi vysoké hodnoty. Napriek tomu môžeme tvrdiť, že prípadná zmena ceny jedného cenného papiera bude mať istý obdobný dopad na zmenu ceny cenného papiera druhej spoločnosti.

Keďže disponujeme všeobecnými veličinami, ako sú výnos a riziko, a závislosti medzi jednotlivými skupinami aktív sú nám tiež známe, môžeme uvažovať o konkrétnom portfóliu. Vychádzajúc z našich dát máme momentálne k dispozícii tri rôzne cenné papiere a jedno aktívum komoditného charakteru, do ktorých môžeme alokovať kapitál. Hovoríme teda o tvorbe portfólia, ktoré môže v našom praktickom príklade pozostávať zo štyroch, troch, dvoch alebo jednej zložky. Je teda len na investorovi, aký podiel svojho kapitálu vloží do možností ponúkaných trhom. Naše okliešťené podmienky v podobe štyroch možností investovania slúžia na redukciu komplexnosti, no napriek tomu poskytujú dobrý základ na objasnenie finančného modelu. Na účely dodatočného zjednodušenia modelu budeme uvažovať, že investor môže rozdeliť svoj kapitál v násobkoch čísla 5 resp. 5%. Toto zjednodušenie vedie k redukcii dosiahnuteľnej množiny, ktorá sa napriek tomu bude vyznačovať vysokou početnosťou.

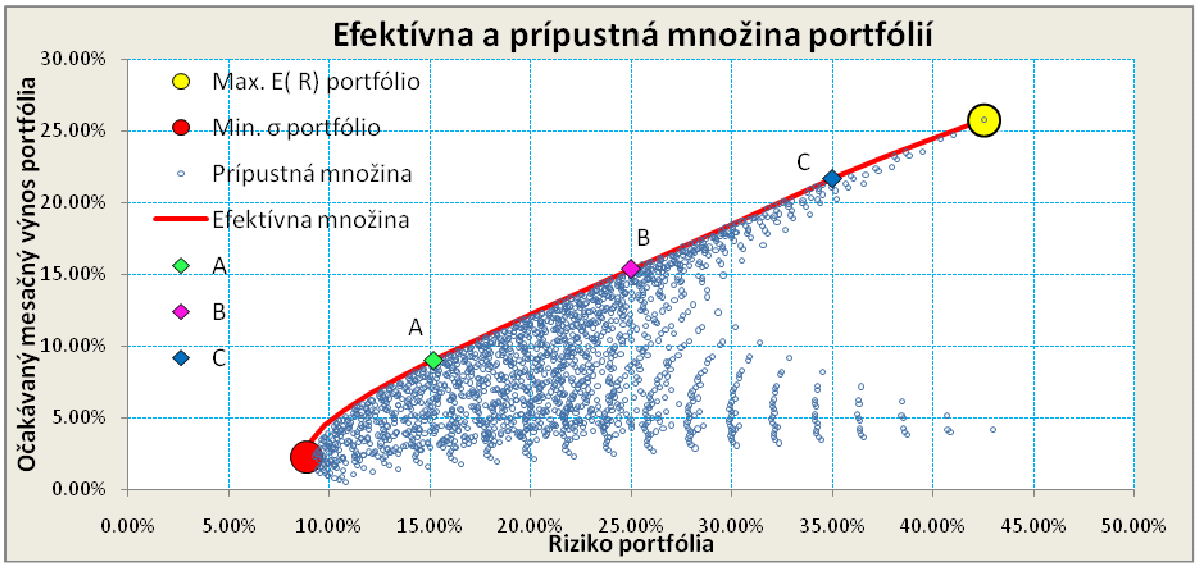


Graf č. 2: Prípustná množina portfólií

Najprv bude potrebné zvážiť, aké sú možnosti, ktoré má investor k dispozícii. Ide teda o definovanie všetkých možností váh priradených jednotlivým aktívam, ktoré tak vytvoria portfólio. Vychádzajúc z toho, že uvažujeme pridelenie váh na základe násobkov 5%, pričom máme možnosť výberu zo štyroch portfólií, existuje 1730 možností alokácie kapitálu. Na základe pridelenia váh jednotlivým aktívam vznikajú portfólia, pre ktoré boli kalkulované veličiny: očakávaný výnos a riziko. Takýmto spôsobom bol vytvorený súhrn portfólií, ktoré sú jedinečné a jednoznačne definované týmito dvomi charakteristikami. Tieto portfólia v našom prípade predstavujú prípustnú alebo dosiahnuteľnú množinu, z ktorej si investor na základe svojich preferencií môže vybrať portfólio, ktoré najlepšie vyhovuje jeho osobitným cieľom a zámerom. Prípustná množina všetkých portfólií vytvorená na základe našich historických dát je zobrazená v grafe č. 2. Každý bod grafu je ekvivalentom jedného portfólia, ktoré má prinášať určitý mesačný výnos a vykazuje príslušnú mieru rizika. Farebne zvýraznené sú 4 jednoprvkové portfóliá, ktoré predstavujú samostatnú držbu konkrétnych

aktív. Dodatočne sú vyznačené dve náhodné portfóliá s patričnými váhami. Tieto body majú poukázať na to, že nie všetky prvky množiny sú efektívnymi portfóliami, pretože pre porovnateľnú mieru rizika je úplne odlišný očakávaný výnos. Racionálny investor musí preto prípustnú množinu podrobiť ďalším skúmaniam a testom.

Keďže dosiahnuteľná množina definuje skupinu portfólií, ktoré investor môže vytvoriť, ale nezohľadňuje racionalitu, túto skupinu dát je potrebné ďalej očistiť. Dve a viaceré odlišné portfóliá prípustnej množiny môžu mať rovnakú mieru očakávaného výnosu pri odlišnej rizikovosti alebo vykazovať rovnaký očakávaný výnos pri odlišnej miere rizika, a preto je nutné tieto skutočnosti zohľadniť. Investor bude samozrejme preferovať vyšší výnos a nižšie riziko, čo vedie k nutnosti screeningu prípustnej množiny. Vytvorí sa tak podmnožina dosiahnuteľnej množiny, ktorá vystupuje pod názvom efektívna množina. Na efektívnej množine ležia všetky portfóliá, ku ktorým nie je možné nájsť paralelné portfólio s väčšou mierou očakávaného výnosu pri danom riziku alebo portfólio s nižšou mierou rizika pri danom očakávanom výnose. Táto množina predstavuje skupinu efektívnych portfólií, ktoré sa odlišujú jednak výnosom a jednak mierou rizika. Efektívna množina pre našu skupinu historických dát je zobrazená v grafe č. 3. Efektívna množina ako skupina efektívnych portfólií je v grafe vyznačená neprerušovanou čiarou, pričom začína v bode portfólia s minimálnou mierou rizika a končí sa portfóliom s najväčším očakávaným výnosom.



Graf č. 3: Efektívna množina portfólií

Okrem portfólií s minimálnym rizikom a maximálnym očakávaným výnosom ležia na efektívnej množine aj náhodne vybraté body A, B, C, ktoré predstavujú portfóliá s individuálnymi charakteristikami bližšie popísané v nasledovnej tabuľke.

Tabuľka č. 3: Vybrané portfóliá ležiace na efektívnej množine

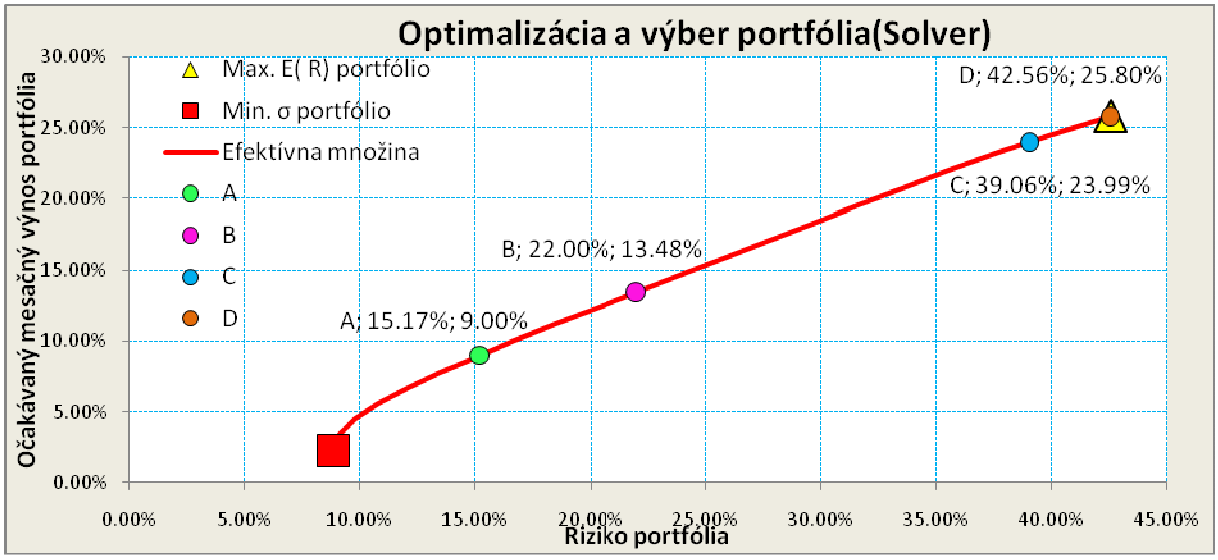
	Váhy/podiel v portfóliu				Mesačne	
Portfólio:	**w(VW.DE)**	**w(RBS.L)**	**w(GS)**	**w(Zlato)**	μ_p	σ_p
Max. E(R)	0,00%	100,00%	0,00%	0,00%	25,80%	42,56%
Min. σ	8,17%	4,13%	11,51%	76,19%	2,26%	8,85%
A	12,21%	33,71%	1,95%	52,13%	4,66%	15,00%
B	17,67%	72,83%	0,00%	9,50%	15,37%	25,00%
C	7,69%	92,31%	0,00%	0,00%	21,66%	35,00%

Napriek tomu, že efektívna množina je jasným medzníkom, ktorý oddeľuje efektívne portfóliá od neefektívnych, stále sa nedá jednoznačne určiť, ktoré portfólio je najvhodnejšie. Efektívna množina totiž vymedzuje podmnožinu portfólií, ktorá však obdobne ako prípustná množina vykazuje značnú početnosť. V prípade, že sa investor chce rozhodnúť pre konkrétne portfólio, musí byť zohľadnená subjektívna zložka, a teda jeho preferencie a averzia voči riziku. Je však už isté, že optimálne portfólio leží na efektívnej množine. Pre jeho konkrétnu špecifikáciu je nutné aplikovať funkciu užitočnosti v spojitosti s indiferentnými krivkami. Práve tieto dva nástroje analýzy finančných portfólií nás privedú k optimálnemu portfóliu.

3. Optimalizácia a výber efektívneho portfólia

Ako je uvedené vyššie, želané portfólio leží na efektívnej množine, ostatné dosiahnuteľné finančné portfóliá ležiace mimo tejto množiny a môžu byť preto v nasledovnej analýze vynechané. Bude sa teda kalkulovať iba s efektívnou množinou. V takomto prípade ide vlastne o optimalizačnú úlohu. Optimalizačnou úlohou je taktiež samotná efektívna množina, pretože ide o portfóliá, ktoré pri danej miere rizika poskytujú najväčší výnos. Či už presné definovanie efektívnej množiny alebo voľba efektívneho portfólia, obe tieto úlohy sa dajú riešiť optimalizačnou aplikáciou Solver v programe Microsoft Excel. Hľadá sa teda portfólio alebo skupina portfólií s minimálnym rizikom pre istý očakávaný výnos, pričom počítanými optimalizovanými premennými sú váhy aktív. Na základe optimalizácie proporcionality aktíva sa portfóliu priradia parametre portfólia, riziko a výnos.

V tejto časti sa už nebude aplikovať prvotný predpoklad, že váhy aktív môžu byť iba násobky 5%, prípadne 0. Tento predpoklad mal redukovať veľkosť prípustnej množiny, avšak v reálnom investovaní takéto obmedzenia neexistujú, a teda investor môže voliť váhy aktív ľubovoľne. Výber želaného portfólia je jednoduchou optimalizačnou úlohou, kde ide o snahu minimalizovať riziko pri istom očakávanom výnose, maximalizovať očakávaný výnos pri danej hodnote rizika alebo dosiahnuť určitú hodnotu očakávaného výnosu. Výber z takýchto optimalizačných úloh je znázornený v nasledovnom grafe.



Graf č. 4: Optimalizácia a výber efektívneho portfólia (Solver)

V nasledovnej tabuľke je uvedený výber z optimálnych portfólií efektívnej množiny pri riešení danej optimalizačnej úlohy. Uvedené sú teda optimalizované portfóliá, ktoré majú prinášať želaný výnos pri minimalizovanej miere rizika, alebo portfóliá maximalizujúce výnos investície pri určitej miere rizika. V prípade, že výnos želaný investorom prevyšuje

skutočné možnosti portfólia, výsledkom optimalizačnej úlohy bude hraničná hodnota, čiže hodnota portfólia s maximálnou možnou mierou očakávaného výnosu (portfólio D).

Tabuľka č. 4: Efektívne portfóliá (Solver)

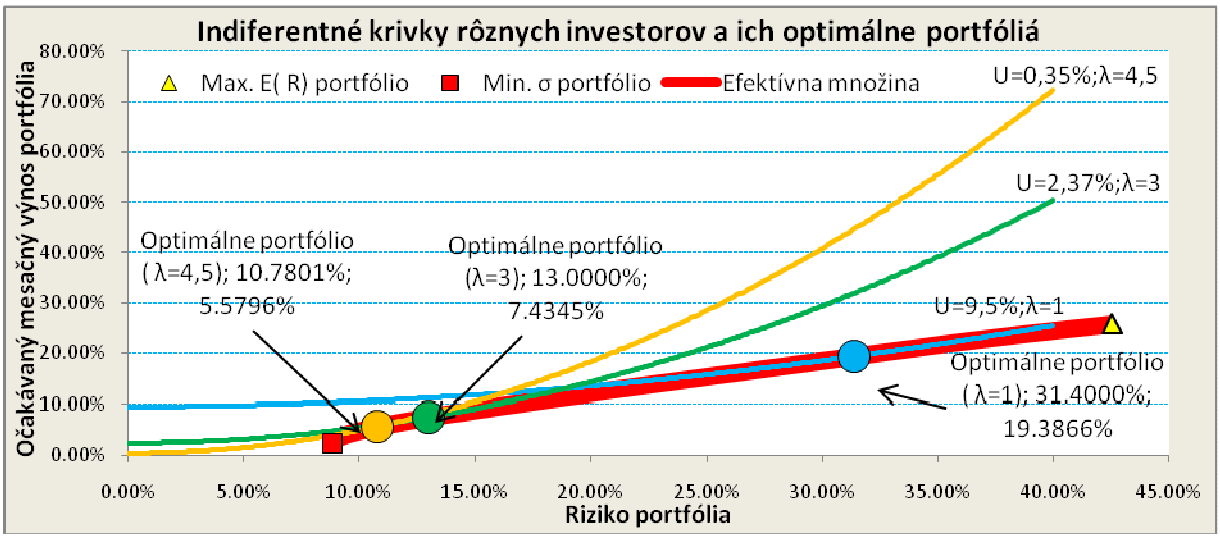
		Váhy/podiel v portfóliu				Mesačne	
Portfólio	**Optimalizácia**	**w(VW.DE)**	**w(RBS.L)**	**w(GS)**	**w(Zlato)**	μ_p	σ_p
A	**E(R)=9%; min. σ**	12,21%	33,71%	1,95%	52,13%	9,00%	15,17%
B	**max. E(R); σ=22%**	17,67%	72,83%	0,00%	9,50%	13,48%	22,00%
C	**E(R)>=23,99%;min. σ**	7,69%	92,31%	0,00%	0,00%	23,99%	39,06%
D	**E(R)=30%; min. σ**	0,00%	100,00%	0,00%	0,00%	25,80%	42,56%

Efektívna množina ako výsledok optimalizácie váh aktív stavia investora pred skupinu riešení, pri ktorých sa nedá jednoznačne povedať, ktoré je nadradené alebo podradené. Týmito riešeniami sú finančné portfóliá, ktoré sa odlišujú mierou očakávaného výnosu a rizika. Ďalší krok, ktorým je výber konkrétneho portfólia, je už úlohou investora, pretože iba ten vie racionálne zvážiť, aký dodatočný podiel rizika je schopný a ochotný znášať za zvýšenú mieru očakávaného výnosu. Nasledovné rozhodnutie bude teda ovplyvnené subjektívnym vnímaním skutočnosti investora resp. jeho preferenciami a averziou voči riziku. V takomto prípade sa vychádza z funkcie užitočnosti a indiferentných kriviek. Keďže každý investor má vlastnú predstavu o tom, aké riziko je ochotný vymeniť za dodatočné navýšenie očakávaného výnosu, funkcia užitočnosti nadobúda u rôznych investorov rôzne podoby. Zjednodušene možno investorov klasifikovať na základe ich averzie voči riziku λ, kde vyššie hodnoty λ indikujú väčšiu averziu voči riskantným investíciám. Vyššia hodnota λ spôsobuje nižšiu mieru úžitku pre investora. Naopak nízky koeficient λ znamená vyššiu rizikovosť investícií, s ktorou sa viaže vyšší očakávaný výnos, a teda aj vyšší úžitok. Na základe koeficientu λ a z neho sa odvíjajúcej funkcie užitočnosti nasledovná tabuľka zjednodušene klasifikuje investorov a priraďuje im optimálne portfóliá s konkrétnou mierou rizika a očakávaného výnosu.

Tabuľka č. 5: Efektívne portfóliá a funkcie užitočností

Investor	Váhy				Mesačne		**Užitočnosť (U)**
Averzia voči riziku (λ)	**w(VW.DE)**	**w(RBS.L)**	**w(GS)**	**w(Zlato)**	μ_p	σ_p	
λ=1	17,81%	73,81%	0,00%	8,39%	19,39%	31,40%	9,53%
λ=3	11,29%	27,00%	4,12%	57,59%	7,43%	13,00%	2,36%
λ=4,5	10,19%	18,92%	6,71%	64,17%	5,58%	10,78%	0,35%

Portfóliá v tabuľke boli investorom priradené na základe indiferentných kriviek, ktoré boli dotyčnicami efektívnej množiny. Každá takáto krivka má jednu hodnotu užitočnosti a jasne vymedzuje efektívne portfólio, ktoré vyhovuje konkrétnemu investorovi. Indiferentné krivky ako dotyčnice efektívnej množiny sú zobrazené v nasledujúcom grafe, ktorý zobrazuje optimálne portfóliá 3 spomínaných investorov.



Graf č. 5: Indiferentné krivky rôznych investorov a ich optimálne portfóliá

4. Záver

Cieľom tejto práce bolo ozrejmiť a priblížiť proces výberu finančného portfólia, ktorý bol aplikovaný na konkrétnom príklade finančného trhu. Analýze výnosu a rizika portfólia v spojení s mierou užitočnosti doviedla fiktívneho investora k prijatiu rozhodnutia v podobe voľby optimálneho portfólia. Markowitzovo pravidlo očakávaného výnosu a rozptylu bolo aplikované na vzorke 1730 fiktívnych portfólií, ktoré boli vytvorené zo štyroch aktív. Keďže výnosom sa rozumela pozitívna zložka investičnej činnosti a na druhej strane stálo riziko alebo rozptyl výnosov ako nežiaduci protipól, pravidlo výnosu a rozptylu prispelo k jednoznačnej klasifikácii a vyčleneniu efektívnych portfólií. Takýmto spôsobom vznikla efektívna množina portfólií, ktorá optimalizovala očakávaný výnos portfólia pri danom riziku resp. mieru rizika pri danej miere očakávaného výnosu. Efektívna množina bola ohraničená portfóliom s minimálnou mierou rizika, čo bolo v tomto prípade σ_p=8,85%, a konečným bodom efektívnej množiny resp. portfóliom s maximálnym očakávaným výnosom, ktorým bolo portfólio s μ_p=25,80%. Efektívna množina bola výsledkom optimalizačnej úlohy a bola súhrnom efektívnych portfólií. Keďže každé takéto portfólio bolo odlišné mierou výnosu a rizika, nedalo sa jednoznačne stanoviť, ktoré je nadradené. Boli preto využité dve možnosti voľby optimálneho portfólia, ktoré vychádzali zo subjektívnych preferencií investora. Ako prvá bola použitá optimalizačná aplikácia Solver. Vychádzalo sa z toho, že investor určí požadovaný výnos alebo akceptovateľnú mieru rizika, prípadne iné doplnkové podmienky a obmedzenia trhu, a pri danej hodnote bude minimalizovaná rizikovosť resp. maximalizovaný výnos portfólia. Druhou cestou k dosiahnutiu optimálneho portfólia bolo použitie funkcie užitočnosti a indiferentných kriviek investora. Bola vytvorená zjednodušená klasifikácia investorov na základe ich vnímania rizika. Tak vznikli traja fiktívni investori definovaní koeficientom λ. Prvý investor (λ=1) preferoval rizikové investície, na rozdiel od druhého investora (λ=4,5), ktorý sa jednoznačne riziku vyhýbal. Akýmsi medzistupňom bol tretí investor (λ=3), ktorý investoval svoj kapitál do portfólia s preferovanou výškou výnosu a opodstatnenou mierou rizika. Na základe indiferentných kriviek investorov, ktoré vychádzali z individuálnej miery užitočnosti a konkrétnej miery averzie investora voči riziku, bolo špecifikované optimálne portfólio pre každého investora. S vyššou mierou averzie investora voči riziku bola spojená nižšia hodnota užitočnosti a naopak. Optimálne portfólio investora, ktorý sa jednoznačne stránil rizikových investícií, má tak jasne nižšiu mieru očakávaného výnosu, ako aj užitočnosť investície. Na druhej strane investor ochotný vystaviť

svoj kapitál vyššej miere rizika si zvolil optimálne portfólio s nadštandardným očakávaným výnosom, ktoré vykazuje vyššiu užitočnosť investície.

5. Literatúra

[1]BENNINGA, S.: *Financial Modeling Modeling.* 2. vydanie. Cambridge : MIT Press, 2000. ISBN 0-262-02482-9

[2]BENNINGA, S.: Principles of Finance with Excel. In: *Portfolio Analysis and the Capital Asset Pricing Model*, New York : Oxford University Press, 2006. ISBN 0-19-530150-1

[3]MARKOWITZ, H. M.: *Portfolio Selection: Efficient diversification of investments.* 2. vydanie. Malden : Wiley-Blackwell, 1991. ISBN 1-55786-108-0

[4]MARKOWITZ, H. M.: Portfolio Selection. In: *Journal of Finance,* Vol. 7, No.1, marec 1952, s. 77-91

[5]Yahoo.com: Yahoo! Finance – Business Finance, Stock Market, Quotes, News. [online]. [citované 30.04.2009] Dostupné z <http://finance.yahoo.com/q?s=>

Adresa autora

Peter Bialoň, 1.ročník magisterského štúdia
Fakulta managementu
Univerzita Komenského v Bratislave
Odbojárov 10, 820 05 Bratislava
e-mail: Peto.Bialon@gmail.com

Demografický vývoj v okresoch Komárno a Liptovský Mikuláš v rokoch 2001 až 2008
Demographic trends in the districts of Komárno and Liptovský Mikuláš between 2001 and 2008

Maroš Čavojský

Abstract: The aim of this paper is presentation of the demographic population of two districts of the Slovak Republic: Komárno and Liptovsky Mikuláš. Author is comparing the contribution of the state population, natural and overall population growth, ethnic composition, life expectancy and age categories in both districts, and also compares these variables between districts.

Key words: state population, natural growth, overall growth, ethnic composition, age structure, average life expectancy.

Kľúčové slová: stav obyvateľov, prirodzený prírastok, celkový prírastok, národnostné zloženie, vekové kategórie, stredná dĺžka života.

1. Úvod

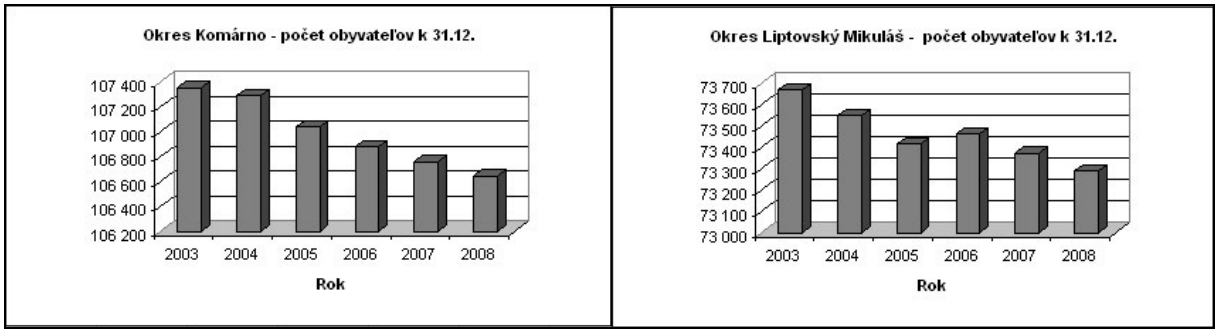
Tento príspevok porovnáva demografické údaje Štatistického úradu SR z rokov 2001 až 2008 za okresy Komárno a Liptovský Mikuláš. Poukazuje na vývoj osídlenia oboch oblastí, vývoj pôrodnosti a mortality, národnostné zloženia, vekové kategórie a strednú dĺžku života v oboch okresoch.

2. Stav obyvateľov

Počet obyvateľov v porovnávanom období zaznamenal pokles v oboch okresoch pod jeden percentuálny bod (Komárno -0,8 % a Liptovský Mikuláš -0,5 %). Snáď treba podotknúť, že podiel počtu obyvateľov, či už v roku 2003 alebo 2008, zostal prakticky rovnaký 45 %. Je zrejmé, že rozdiely v počtoch obyvateľov sú dôsledkom historicky a geograficky rôznymi podmienkami urbanizácie oblastí.

Tabuľka 1: Stav obyvateľov

Počet obyvateľov	2003	2004	2005	2006	2007	2008
Okres Komárno	107 355	107 290	107 037	106 876	106 761	106 645
Okres Liptovský Mikuláš	73 668	73 549	73 418	73 464	73 373	73 289



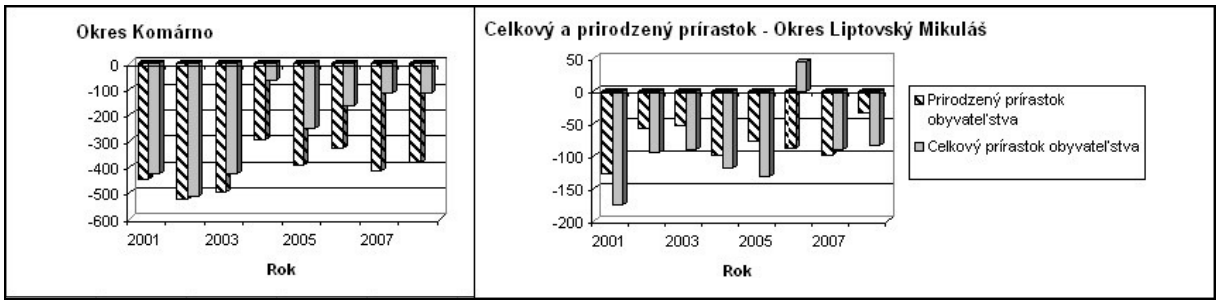
Obrázok 1: Stav obyvateľov

3. Prirodzený a celkový prírastok obyvateľov

O prirodzenom a celkovom prírastku v oboch porovnávaných okresoch vlastne nemôže byť reč, na koľko štatistické údaje vykazujú záporné čísla – teda úbytok obyvateľstva. Celkovo tieto údaje sú potešiteľné, nakoľko vyjadrujú pozitívny vývoj v pôrodnosti ako aj pokles nárastu mortality. Údaje hovoria o prakticky nulovom náraste zomrelých. Tu treba pripomenúť, že na tomto vývoji participuje zvýšené prisťahovateľstvo (okres Komárno +44% okres Liptovský Mikuláš +58%) oproti vysťahovanému obyvateľstvu (okres Komárno +11% okres Liptovský Mikuláš +9%) .

Tabuľka 2: Prirodzený a celkový prírastok obyvateľov

	2001	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**
Okres Komárno								
Prirodzený prírastok	-449	-525	-494	-292	-393	-327	-413	-380
Celkový prírastok	-423	-511	-428	-65	-253	-161	-115	-116
Okres Liptovský Mikuláš								
Prirodzený prírastok	-128	-58	-54	-99	-78	-89	-99	-34
Celkový prírastok	-175	-94	-90	-119	-131	46	-91	-84



Obrázok 2:Prirodzený a celkový prírastok

4. Stredná dĺžka života – muži, ženy, spolu

Údaje jednoznačne preukazujú, že ženy sa dožívajú dlhšieho veku ako muži. Tento fakt je poplatný pre oba okresy. Avšak je badať, že v sledovanom období vzrástla stredná dĺžka života u mužov o 1,5% (v oboch okresoch), kým stredná dĺžka života žien nedosahovala takúto mieru progresu (okres Komárno +0,7% , okres Liptovský Mikuláš +0,3%).

Tabuľka 3: Stredná dĺžka života

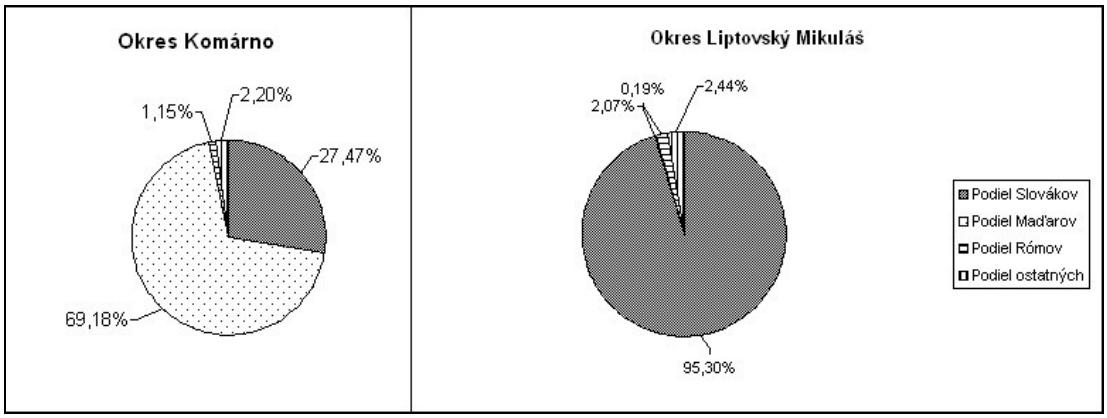
Stredná dĺžka života	***2001***	***2002***	***2003***	***2004***	***2005***	***2006***	***2007***	***2008***
Komárno								
Muži	68,25	68,34	68,24	68,71	68,54	68,87	68,87	69,27
Ženy	76,3	76,33	76,58	76,67	76,47	76,47	76,59	76,8
Spolu	72,275	72,335	72,41	72,69	72,505	72,67	72,73	73,035
Liptovský Mikuláš								
Muži	69,75	70	70,27	70,56	70,76	70,81	70,75	70,89
Ženy	79,29	79,25	79,16	79,06	79,39	79,35	79,46	79,51
Spolu	74,52	74,625	74,715	74,81	75,075	75,08	75,105	75,2

5. Národnostné zloženie obyvateľstva

Pokiaľ údaje v predchádzajúcich parametroch vykazovali minimálne štatistické rozdiely, v tomto ukazovateli vidíme diametrálne rozdielne hodnoty. Je to prirodzené, veď okresy Komárno a Liptovský Mikuláš boli v minulosti osídľované obyvateľstvom s rozdielnym národnostným zložením. Kým v oblasti Komárna to boli v minulosti hlavne Maďari, potom Liptov je možné považovať za výsostne slovenské územie. Tento rozdiel sa generačne preniesol do súčasnosti, čo potvrdzujú aj údaje v tabuľke číslo 4.

Tabuľka 4: Národnostné zloženie

Podiel	2001	2002	2003	2004	2005	2006	2007
Okres Komárno							
Slovákov	27,67%	27,60%	27,51%	27,47%	27,45%	27,36%	27,25%
Maďarov	69,09%	69,13%	69,19%	69,21%	69,17%	69,23%	69,23%
Rómov	1,12%	1,12%	1,14%	1,14%	1,16%	1,17%	1,18%
Ostatných	2,12%	2,15%	2,17%	2,18%	2,22%	2,24%	2,35%
Okres Liptovský Mikuláš							
Slovákov	65,06%	65,27%	65,44%	65,35%	65,36%	65,44%	65,39%
Maďarov	0,13%	0,13%	0,13%	0,13%	0,13%	0,13%	0,13%
Rómov	1,40%	1,41%	1,42%	1,42%	1,42%	1,43%	1,43%
Ostatných	1,61%	1,63%	1,64%	1,65%	1,68%	1,74%	1,78%



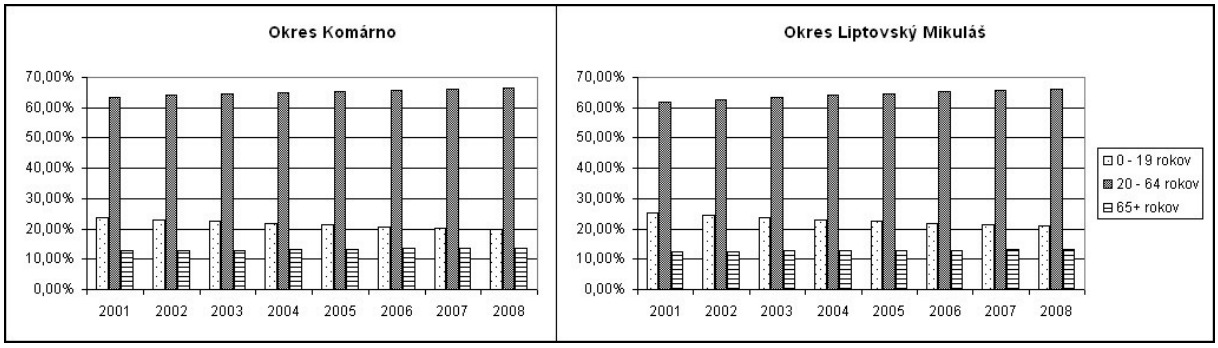
Obrázok 3: Národnostné zloženie

6. Vekové kategórie obyvateľstva

Ak sa pozrieme na údaje v tabuľke, zistíme, že napriek rôznym percentuálnym podielom v jednotlivých kategóriách medzi okresmi sú až zarážajúco podobné odchýlky medzi rokmi 2001 a 2008 (v kategórii do 19 rokov je to pokles o 4 percentuálne body, v kategórii do 64 rokov nárast o 4 percentuálne body a v kategórii nad 64 rokov nárast o 1 percentuálny bod). Znamená to, že bez ohľadu, na ktorý okres sa pozeráme, obyvateľstvo starne. V tejto súvislosti je vhodné pripomenúť riziká spadajúce do sociálnej oblasti, zamestnanosti a dôchodkového zabezpečenia.

Tabuľka 5: Vekové kategórie

	2001	2002	2003	2004	2005	2006	2007	2008
Okres Komárno								
Kategória 0 - 19 rokov								
Podiel kategórie	23,71%	23,02%	22,40%	21,76%	21,20%	20,68%	20,21%	19,74%
Kategória 20 - 64 rokov								
Podiel kategórie	63,46%	64,13%	64,68%	65,11%	65,50%	65,84%	66,19%	66,56%
Kategória 65+ rokov								
Podiel kategórie	12,83%	12,85%	12,92%	13,13%	13,29%	13,48%	13,60%	13,70%
Okres Liptovský Mikuláš								
Kategória 0 - 19 rokov								
Podiel kategórie	25,43%	24,63%	23,83%	23,13%	22,40%	21,81%	21,32%	20,81%
Kategória 20 - 64 rokov								
Podiel kategórie	61,98%	62,78%	63,45%	64,07%	64,70%	65,23%	65,59%	65,96%
Kategória 65+ rokov								
Podiel kategórie	12,60%	12,59%	12,72%	12,80%	12,89%	12,95%	13,10%	13,24%



Obrázok 4: Vekové kategórie

7. Záver

Cieľom tohto príspevku bola prezentácia demografického vývoja obyvateľstva dvoch okresov Slovenskej republiky: okresy Komárno a Liptovský Mikuláš. Autor v príspevku zohľadnil stav obyvateľov, prirodzený a celkový prírastok obyvateľstva, národnostné zloženie, strednú dĺžku života a vekové kategórie v oboch okresoch a taktiež porovnanie ukazovateľov medzi týmito okresmi.

8. Literatúra

[1]ŠTATISTICKÝ ÚRAD SLOVENSKEJ REPUBLIKY, Regionálna databáza, http://px-web.statistics.sk/PXWebSlovak/

Adresa autora (-ov):

Maroš Čavojský
1. ročník Bc. štúdia, STU, FEI
Iľkovičova 3, 812 19 Bratislava 1
xcavojskym@stuba.sk

Analýza zamestnanosti v regiónoch EÚ
Analysis of employment in EU regions

Branislav Gajarský

Abstract: The main purpose of the article is to provide reader an information on situation in EU regions (NUTS 2) according to data of Eurostat from 2007. By using of proper statistical method an indicator was created for every EU region to show distance from goals of Lisbon Strategy in field of total employment, employment of women and employment of older people. The article interprets results for particular regions with impact on possible causes and anomalies.

Key words: Lisbon strategy, NUTS 2, rate of employment

Kľúčové slová: Lisabonská stratégia, NUTS 2, miera zamestnanosti

1. Úvod

V novembri 2003 predložila pracovná skupina vedená vtedajším belgickým premiérom Wimom Kokom správu „Jobs, Jobs - Creating more employment in Europe“. V tejto správe poukázala skupina na neplnenie cieľov Lisabonskej stratégie z roku 2000. O rok neskôr predložila táto skupina správu „Facing the challenge – The Lisbon Strategy for growth and employment in Europe“. V nej zdôrazňovala vysokú dôležitosť Lisabonskej stratégie a poukázala na nedostatočné plnenie jej počiatočných cieľov. Tie boli stanovené ešte v marci 2000, keď sa vtedajších 15 lídrov krajín EÚ na sneme v Lisabone dohodlo na spoločnom postupe pri zvyšovaní miery rastu, zamestnanosti, sociálnej súdržnosti a udržateľnosti životného prostredia.

Na odmeranie rozvoja politiky Lisabonskej stratégie bolo v decembri 2003 určených niekoľko štruktúrnych indikátorov. Štruktúrnymi sa nazývajú, pretože popisujú štruktúru a kľúčové aspekty v každej oblasti. Štruktúry sú charakteristiky, ktoré sa nemenia príliš rýchlo a preto popisujú vývoj spoločnosti z dlhodobého hľadiska. Od prvého zverejnenia v roku 2001 sa množstvo indikátorov strojnásobilo. Bolo potrebné, aby vznikol užší zoznam štruktúrnych indikátorov, ktorý by jednoduchšie a prehľadnejšie popisoval strategické ciele. Užší zoznam z roku 2003 obsahoval 15 kľúčových indikátorov. Článok sa zameriava len na tri z nich:

1. mieru zamestnanosti,
2. mieru zamestnanosti žien,
3. mieru zamestnanosti starších obyvateľov.

V oblasti zamestnanosti boli Lisabonskou stratégiou prijaté tieto strategické ciele s víziou pre rok 2010:

- miera zamestnanosti na úrovni 70 %,
- miera zamestnanosti žien na úrovni 60 %,
- miera zamestnanosti starších pracovníkov medzi 55 – 64 rokov na úrovni 50 % (cieľ bol prijatý v roku 2005).

V roku 2001 sa uskutočnilo v Štokholme zasadanie Európskej rady, kde sa vytýčili aj strednodobé ciele v oblasti zamestnanosti. Do roku 2005 mala dosiahnuť celková zamestnanosť 67 % a zamestnanosť žien 57 %. Tento článok je zameraný na vytvorenie poradia regiónov pomocou metódy vzdialenosti od fiktívneho bodu. V tomto prípade sú fiktívnymi bodmi ciele Lisabonskej stratégie pre roky 2005 a 2010.

2. Súbor dát

Analyzovaný súbor tvorili tieto ukazovatele tri ukazovatele:

- Celková miera zamestnanosti mužov a žien vo veku 15-64 rokov.
- Miera zamestnanosti žien vo veku 15-64 rokov.
- Celková miera zamestnanosti starších mužov a žien vo veku 55-64 rokov.

Zdrojom ukazovateľov bola stránka Eurostatu (www.*ec.europa.eu/eurostat*). Analyzovaný bol rok 2007, pretože v čase písania článku boli najaktuálnejšie dáta práve z tohto roku. Súbor bol tvorený regiónmi EÚ-27, okrem regiónov Slovinska a Dánska. Spomínané dve krajiny nezverejnili údaje o mierach zamestnanosti pre regióny. Celkovo tvorí súbor 264 regiónov podľa kategorizácie NUTS 2.

3. Popis metódy a interpretácia výsledkov

V štatistickom súbore bolo na základe metódy fiktívneho bodu určené poradie regiónov v EÚ-27.

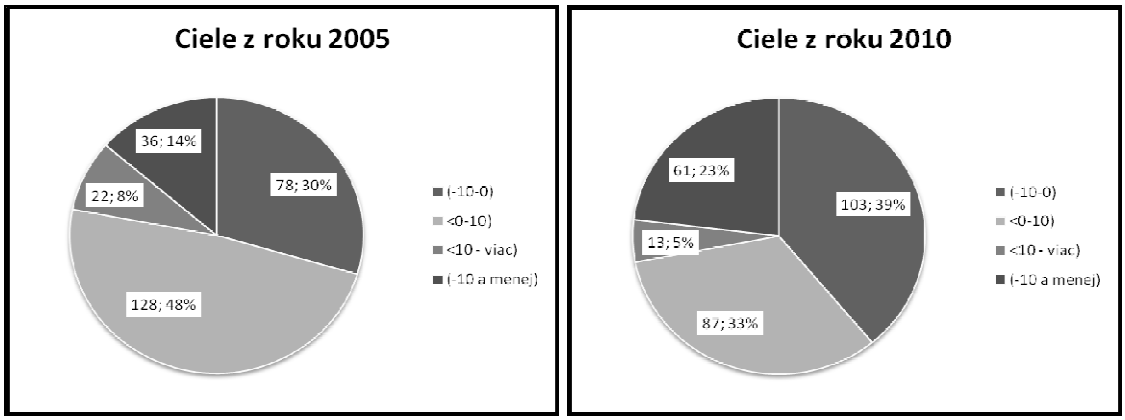
Tabuľka 2: Názvy analyzovaných ukazovateľov a ciele pre LS

Názov ukazovateľa	Označenie v analýze	Cieľ 2005	Cieľ 2010
Miera zamestnanosti žien	ženy 15-64	57%	60%
Celková miera zamestnanosti	total 15-64	67%	70%
Miera zamestnanosti starších	total 55-64	-	50%

Pre regióny sa vypočítajú vzdialenosti ukazovateľov od jednotlivých cieľov, z týchto vzdialeností sa urobí priemer. Priemernú vzdialenosť vypočítame podľa vzťahu:

$$d_i = \frac{1}{p}\sum_{j=1}^{p}(x_{ij} - x_{oj}) \text{ pre } \quad i=1,2...n \quad \text{a} \quad j=1,2...p \tag{1}$$

V tomto prípade i je počet štatistických jednotiek (n =264) a j je počet ukazovateľov (p_{2005} = 2 a p_{2007} =3). Poradie regiónov sa určí na základe priemerov, ktoré budú nadobúdať kladné a záporné hodnoty pre jednotlivé regióny. Kladné znamenajú presiahnutie vytýčených cieľov a záporné nedosiahnutie cieľov. Z takto vzniknutých skóre bol vytvorený priemerný výsledok pre krajinu z dôvodu rozsahovej náročnosti. Krajiny boli podľa nich zoradené. V grafe pod textom je percentuálny podiel regiónov podľa dosiahnutého bodového hodnotenia (počet regiónov, percento).



Graf 1: Podiel regiónov podľa dosiahnutého skóre v rokoch 2010 a 2005

Z 264 regiónov, ktoré boli v porovnávaní sa urobil priemer dosiahnutého skóre pre jednotlivé krajiny. Najlepšie skóre dosiahlo Švédsko (SE), kde boli ciele presiahnuté vo všetkých regiónoch. Tieto regióny sa aj nachádzajú v prvej desiatke z celkového počtu regiónov. V druhom Estónsku (EE), ktoré bolo reprezentované jedným regiónom, len celková miera zamestnanosti zaostávala približne 1 % za cieľom. Spojené kráľovstvo (UK) skončilo na 3. mieste, pričom v priemere ciele stratégie naplnilo. Nasledovalo Holandsko (NL), ktoré dosahovalo nadpriemerné hodnoty. Iba v oblasti zamestnávania starších cieľovú hodnotu nedosiahli regióny Groningen, Zeeland a Limburg. Fínsko (FI) dosiahlo rovnaké priemerné skóre ako Holandsko (4,8). Jediným regiónom so slabším skóre bol Itä-Suomi (-3,8). Naopak regiónom, ktorý dosiahlo z celkového rebríčka najvyššie bodové hodnotenie bolo autonómne súostrovie Ålandy, ktoré dosahovalo priemernú mieru zamestnanosti takmer 80 %.

Kladné priemerné skóre dosiahli ešte krajiny Lotyšsko (LV), Cyprus (CY), Nemecko (DE), Írsko (IE) a Litva (LT). Lotyšsko, Litva a Cyprus boli reprezentované jedným regiónom. Z týchto krajín, len Litva nedosiahla cieľ 70 % v zamestnanosti. V Nemecku (skóre: 1,8) dosiahli tradične nízku mieru zamestnanosti a nízke priemerné skóre regióny východného Nemecka, ktoré tvorili bývalú NDR. Rozdiely medzi „najhorším“ a „najlepším“ regiónom však neboli až také markantné (Freiburg – 8,7 a Arnsberg - -3,8).

Nasledujú regióny s nižším skóre ako 0. Tesne pod hranicou ležia Portugalsko (PT) a Rakúsko (AT). Čo sa týka Portugalska vysoké skóre dosahujú Centro (6,4), kde sa združuje priemysel a Algavre (2,0), ktoré je známou turistickou destináciou. Najhoršie je na tom južná časť Alentejo (-2,2), Azorské ostrovy (-7,0) a Madeira (-3,9). V Rakúsku výrazne znižuje skóre zamestnanosť starších obyvateľov (-10,6 bodu). Španielsko (ES), ktoré dopadlo tak zle najmä vďaka autonómnym mestám Ceuta (-23,8) a Melilla (-18,4) na pobreží Afriky. Výrazne nízke skóre dosiahli južné regióny Andalúzia, Extramadura a Casiilla La Mancha. Regióny s kladným skóre boli Katalánsko (1,4) a okolie Madridu (0,8), kde je najväčšia koncentrácia priemyslu. Katalánsko je tradične známe svojimi ambíciami oddeliť sa od Španielska.

Na 14. mieste sa umiestnilo Bulharsko (BG), ktorému výrazne zvyšuje priemer región Yugozapaden (1,0) s hlavným mestom Sofia. Ostatných 5 regiónov dosiahlo v priemere skóre -9,1. V Českej republike (CZ) výrazne najvyššie skóre dosiahla Praha (5,0) a Stredné Čechy (-1,4). Najnižšie naopak regióny Severozápad (-8,0) a Moravskoslezsko (-8,6). Je viditeľné, že najvyššie skóre si udržiavajú oblasti v blízkosti hlavných miest a hlavné mestá sa výrazne odlišujú od ostatných regiónov. Po Českej Republike nasledovalo Francúzsko, ktorému znižujú priemer skóre najmä malé ostrovné štáty (Guadaloupe, Martinik a Réunion), Guayana a Korzika. Bez nich by malo skóre -6,3. Treba však konštatovať, že ani v jednom regióne neboli dosiahnuté ciele celkovej zamestnanosti a zamestnanosti starších.

Na 17. priečke sa umiestnilo Slovensko (SK). Takisto ako pri Českej republike, najvyššie hodnotenie dosiahlo hlavné mesto, kde už boli dosiahnuté všetky ciele stratégie. V tabuľke je vidieť výrazné regionálne rozdiely medzi Bratislavským krajom a ostatnými regiónmi.

Tabuľka 3: Skóre dosiahnuté v regiónoch SR

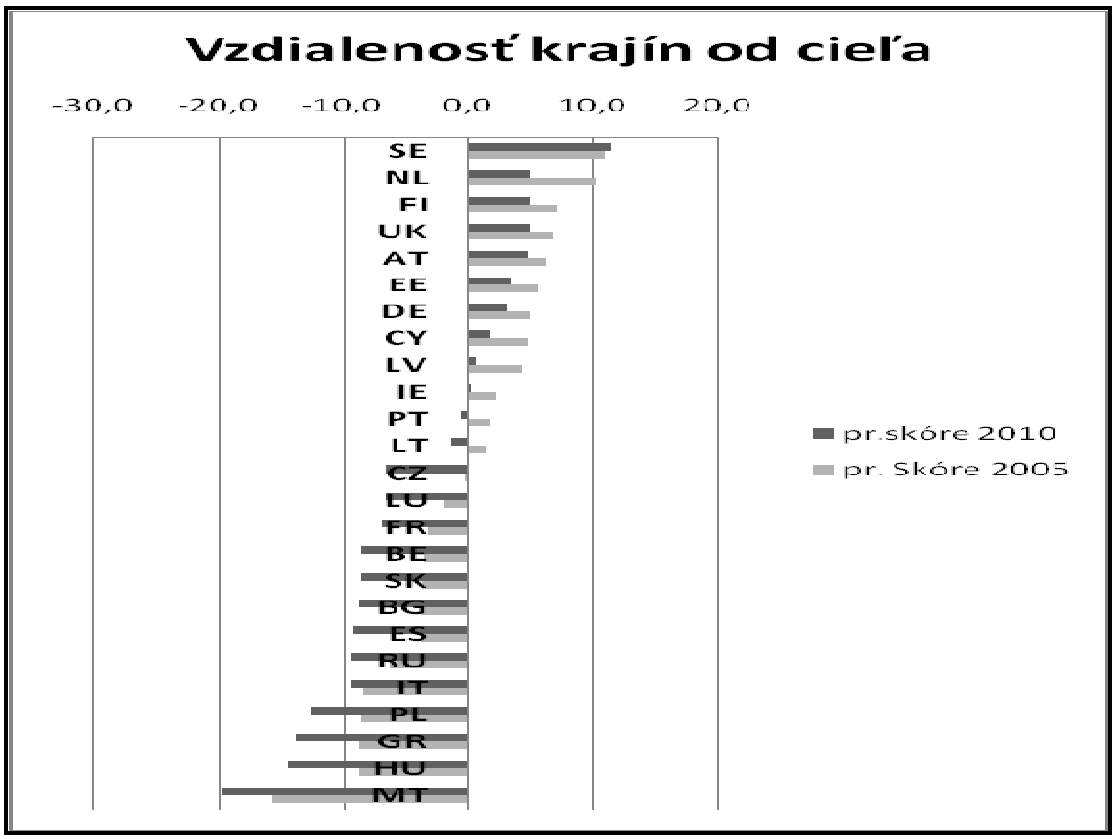
	Ukazovatele			Vzdialenosť od cieľa z roku 2010			
Kraj	total 55-64	ženy 15-64	total 15-64	total 55-64	ženy 15-64	total 15-64	priemer
Západ	36,9	55,8	63,6	-13,1	-4,2	-6,4	-7,9
Stred	30,6	48,9	57,7	-19,4	-11,1	-12,3	-14,3
Východ	29,8	47,9	55,5	-20,2	-12,1	-14,5	-15,6
Bratislava	53,9	65,7	71,0	3,9	5,7	1,0	3,5

So skóre -8,6 ďalej nasleduje Belgicko (BG). Dôvodom je najmä slabá zamestnanosť starších obyvateľov. Nižšiu zamestnanosť dosahovali v priemere regióny Valónska oproti Flámskym (58,6 % oproti 66,1 %). Luxembursko (LU) dosiahlo podobné skóre ako Belgicko, ale bolo reprezentované len jedným regiónom. Rumunsko (RO) a Grécko (GR) sú poslednými krajinami, čo dosiahli skóre nad -10. V Rumunsku dosiahol najbližšie skóre k 0 región Nord-Est (Severovýchod), naopak najhoršie hodnotenie malo centrálne a juhovýchodné Rumunsko. Grécko dosiahlo najlepšie skóre na Peloponézskom polostrove, kde je väčšia koncentrácia veľkých miest. Opačne to bolo v regiónoch susediacich s Macedónskom.

Na posledných priečkach sa umiestnili Taliansko (IT), Maďarsko (HU), Poľsko (PL) a Malta (MT). Tieto krajiny dosiahli skóre menej ako -10. V Taliansku sa podieľajú na nízkom priemernom hodnotení regióny južného Talianska (Campania, Puglia, Basilicata, Calabria, Sicília a Sardínia) a regióny stredného Talianska (Molise a Abrudzo), kde je priemerné skóre -20,6. Ostatné regióny dosahujú priemerné skóre 7,9.

Poľsko a Maďarsko sú krajiny, kde skóre v regiónoch dosahuje extrémne nízke hodnoty. V Maďarsku sú na tom najhoršie južné a severovýchodné regióny (-18 bodu v priemere). O niečo lepšiu hodnotu dosahuje Budapešť a okolie a severozápad Panónie hraničiaci s Rakúskom. V Poľsku paradoxne dosiahli najlepšie hodnotenia regióny na východe republiky. Hodnotenia sa pohybovali v intervale <-19,1; -8,9> bodu za región. Poslednou hodnotenou krajinou bola Malta, ktorá dosiahla -19,7 bodu.

Skóre za jednotlivé krajiny bolo vypočítané ako priemerné hodnotenie regiónov, preto sa určite líši od poradia krajín na základe celkovej miery zamestnanosti. Z analýzy vyplýva, že stále existujú veľmi veľké rozdiely medzi jednotlivými regiónmi EÚ a je potrebné ich odstraňovať. V grafe 2 je znázornené poradie štátov podľa vypočítaného priemerného skóre.



Graf 2:Poradie krajín podľa metódy vzdialenosti od fiktívneho bodu

4. Záver

Požitie metódy vzdialenosti od fiktívneho bodu odhalilo celkové poradie regiónov na základe ich vzdialenosti od cieľov vytýčených Lisabonskou stratégiou. Toto poradie umožňuje identifikovať regionálne rozdiely v rámci krajiny a poskytuje tak hlbší pohľad na skutočnú situáciu v krajine. Ak krajina aj dosiahne celkové hodnotenie, ktoré je uspokojivé, nemusí to vypovedať o skutočnosti v celej krajine. Môžu v nej totižto existovať regióny, ktoré dosahujú nadpriemerne vysoké hodnoty a regióny s hodnotami opačnými. Analýza regiónov pomohla odhaliť, ktoré regióny sú tie najslabšie a ktorým je nutné pomáhať vo vyrovnaní sa s tými „lepšími“, napríklad aj väčšími dotáciami z fondov EÚ. Treba však zdôrazniť, že kvôli kríze môžu byť súčasné štatistiky cieľom z Lisabonu veľmi vzdialené.

5. Literatúra

[1] STANKOVIČOVÁ I. – VOJTKOVÁ, M. 2007. Viacrozmerné štatistické metódy. Bratislava: Iura Edition, 2007. ISBN 978-80-8078-152-1

[2] EUROPEAN COMMISSION. Facing the challenge: Report from High Level Group chaired by Wim Kok. [online] 2.11.2009. [cit. 3.11. 2009] Dostupné z <*ec.europa.eu/growthandjobs/pdf/kok_report_en.pdf*>

[3] EUROPEAN COMMISSION. Jobs, jobs, jobs creating more employment in Europe: Report of the Employment Taskforce chaired by Wim Kok. [on-line] 2.11.2009. [cit. 3.11 2009] Dostupné z <http://ec.europa.eu/employment_social/publications/2004/ke5703265_en.pdf>

[4] STANKOVIČOVÁ I. – VOJTKOVÁ, M.: Viackriteriálne hodnotenie členských krajín EÚ na základe vybraných ukazovateľov Lisabonskej stratégie. [on-line] 2.11.2009. [cit. 4.11. 2009] Dostupné z <http://www.fses.uniba.sk/fileadmin/user_upload/editors/Studium/AE/MSD/Stankovicova_clanok_LS_cely.pdf>

[5] GAJARSKÝ, B: Analýza zamestnanosti v regiónoch EÚ [Bakalárska práca] – Univerzita Komenského v Bratislave. Fakulta managementu; Katedra informačných systémov. – Vedúci bakalárskej práce: Ing. Iveta Stankovičová PhD.: 2009, 48 s.

Adresa autora:

Branislav Gajarský, Bc.,
1. ročník magisterského štúdia
Fakulta managementu UK v Bratislave
branislavus@gmail.com

Analýza vývoja zahraničného obchodu SR
Analysis of the foreign trade development of Slovak Republic

Juriga Ján

Abstract: In this paper I compare both territorial and commodity structures of Slovak foreign trade in years 2002 and 2008. I also analyse total imports and exports development since 1997 as well as the trade balance development since 1997.

Key words: foreign trade, development, Slovakia, imports, exports, trade balance, structure

Kľúčové slová: zahraničný obchod, vývoj, Slovensko, import, export, obchodná bilancia, štruktúra

1. Úvod

Slovenská republika zaznamenala v uplynulých rokoch nebývalý ekonomický rast. Tento rast bol spojený s rýchlym poklesom nezamestnanosti a zvyšovaním životnej úrovne obyvateľov. Menej často spomínaným, ale o to dôležitejším faktom je výrazný nárast objemu zahraničného obchodu. V tejto práci analyzujem zmeny v teritoriálnej ako aj komoditnej štruktúre zahraničného obchodu Slovenska, sledujem vývoj celkového exportu, importu a obchodnej bilancie SR od roku 1997. Ako zdroj informácií mi poslúžili voľne dostupné databázy. Štatistický úrad Slovenskej republiky na svojich internetových stránkach pravidelne zverejňuje publikáciu „Zahraničný obchod Slovenskej republiky“. V nej sú zaznamenané dáta o zahraničnom obchode Slovenskej republiky za každý mesiac roku, jeho teritoriálna aj komoditná štruktúra. Analyzované štatistiky zahŕňajú medzinárodný obchod tovarov v hodnotách typu FOB t.j. zahŕňajú transakčnú hodnotu tovaru a hodnotu služieb (napr. doprava, poistenie, prekládka, skladovanie tovaru a pod.) vykonaných na dodanie tovaru na hranicu vyvážajúcej krajiny

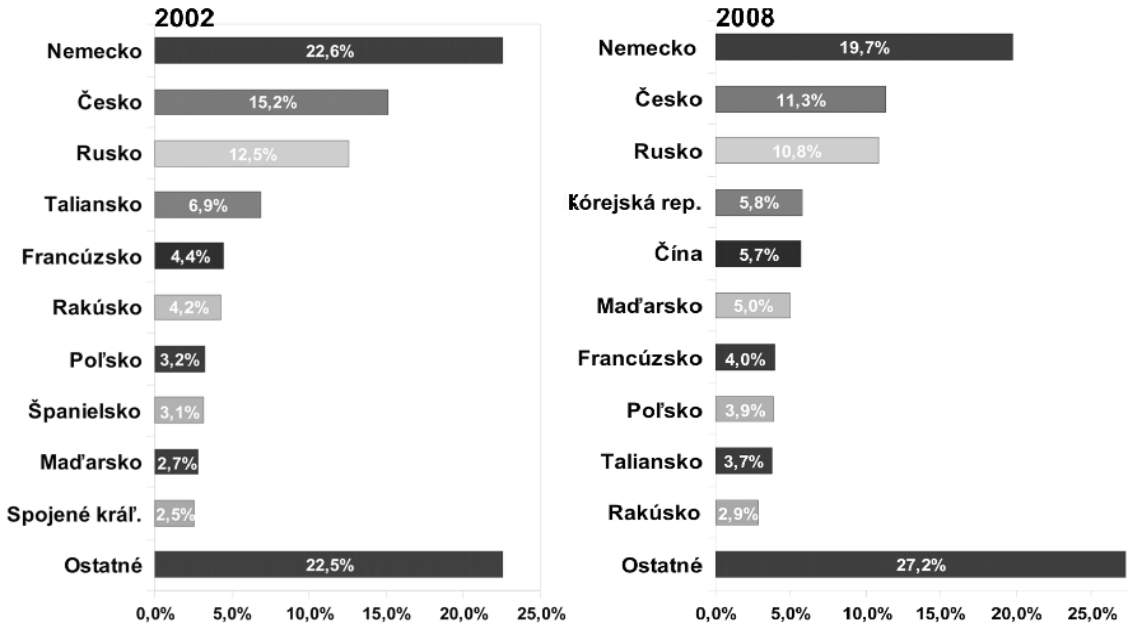
2. Porovnanie teritoriálnej štruktúry importu do SR v rokoch 2002 a 2008

V roku 2002 sa na Slovensko doviezol tovar v celkovej hodnote 747 975,47 miliónov SKK (16 628,8 mil. USD). V roku 2008 sa doviezol tovar v celkovej hodnote 1 514 056 milión. SKK (71 264 milión. USD, 48 421 milión. EUR). Keď porovnáme celkový objem importu v rokoch 2002 a 2008, vidíme, že sa v bežných cenách v SKK takmer zdvojnásobil. V dolárovom vyjadrení vďaka posilneniu slovenskej meny voči USD stúpol až 4,3-násobne. Zvýšila sa aj miera otvorenosti ekonomiky na strane importu: zo 67,4% na 74,6%. To znamená, že v roku 2008 sa na Slovensko doviezol tovar v hodnote takmer až ¾ HDP Slovenska. Podiel hlavných importérov (Nemecka, Českej republiky a Ruska) na celkovom importe sa znížil, zatiaľ čo 2,2-násobne vzrástol podiel ázijských krajín (Kórejská republika, Čína), viď Obrázok 1. Podiel krajín EÚ sa zvýšil o jednu tretinu oproti roku 2002, čo môžeme pripísať najmä vstupu nových členských krajín do EÚ.

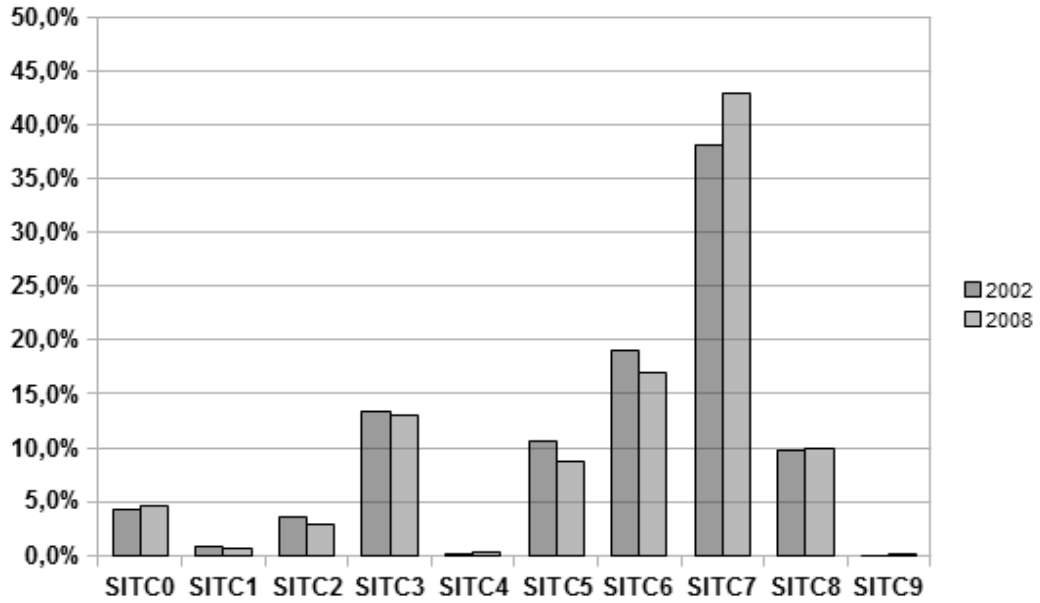
3. Porovnanie komoditnej štruktúry importu do SR v rokoch 2002 a 2008

Z hľadiska tovarovej štruktúry importu podľa medzinárodnej klasifikácie SITC nedošlo v rokoch 2002-2008 k významným zmenám. Svoj podiel ešte zväčšila trieda SITC7 – Stroje a zariadenia a to 1,062-násobne na 42,9% z celkového importu, čo znamená v nominálnom peňažnom vyjadrení hodnotu 649,17 miliardy SKK. 1,47-násobne vzrástol podiel importu olejov a tukov (SITC4) stále však tvorí len 0,3% z celkového importu. Významnejšie klesol podiel nápojov a tabaku (SITC1) a surových materiálov (SITC2) na 74% resp. 81% z podielu

v roku 2002 (Obrázok 2). Podľa jednotlivých druhov tovarov sa na Slovensko v roku 2008 doviezlo najviac televíznych a rozhlasových prijímačov - 142 574,42 mil. SKK, 9,4% z celkového dovozu. Pritom v roku 2002 bol ich podiel na celkovom importe len 1,4%. Ropy a zemného plynu sa v roku 2008 doviezlo za 140 726 milión. SKK, čo je 9,3% z importu. V roku 2002 to bolo až 10,5%. Častí, súčastí a príslušenstva pre motorové vozidlá a ich motory sa v roku 2008 doviezlo za 112 765 milión. SKK, (7,45% oproti 6,5% v 2002) a motorových vozidiel (99 529 milión. SKK, 6,7% oproti 6,5% v roku 2002).



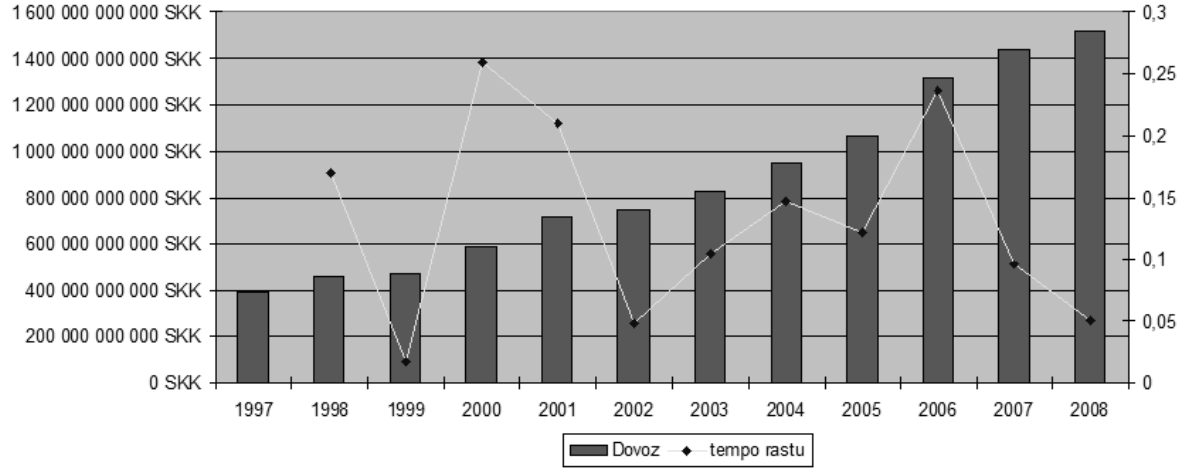
Obrázok 2: Porovnanie teritoriálnej štruktúry importu do SR v rokoch 2002 a 2008



Obrázok 3: Porovnanie komoditnej štruktúry importu SR v rokoch 2002 a 2008

4. Analýza časového radu importu tovarov do SR v rokoch 1997-2008

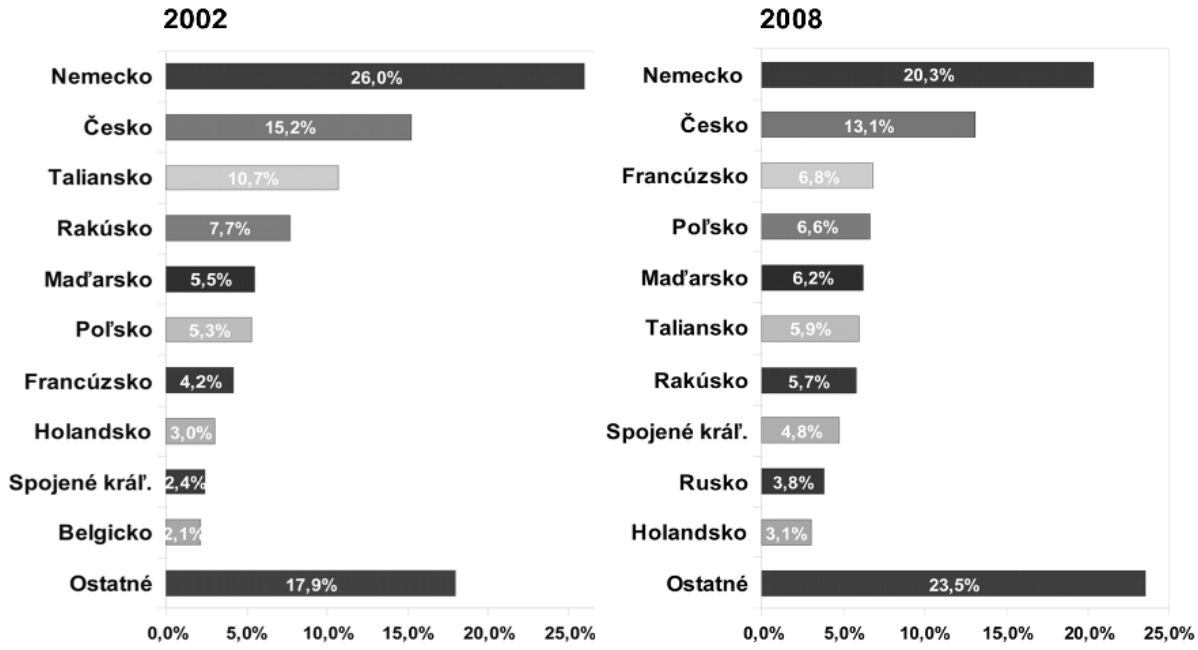
Pri analýze znakov „celkový import v bežných cenách“ a „koeficient rastu importu“ štatistickej jednotky „Slovenská republika“ som dospel k nasledovným zisteniam. V sledovanom období rokov 1997-2008 stúpol celkový import tovarov v bežných cenách na Slovensko 3,84-krát. Zatiaľ čo v roku 1997 predstavoval sumu 393 973 milión. SKK, v roku 2008 to bolo už 1 514 056 milión. SKK. Priemerný koeficient rastu vypočítaný ako geometrický priemer z koeficientov rastu za každý rok sa rovná 10,5%. Pri koeficientoch rastu za každý rok však pozorujeme významné odchýlky. Po tom ako v roku 1998 celkový import vzrástol iba o 1,8%, nasledovali 2 roky s viac ako 20% rastom ročne. Po menšom útlme rastu v roku 2002 import v ďalších rokoch rástol viac ako 10%-ným tempom čo vyvrcholilo rastom o 23,6% v roku 2006 oproti predchádzajúcemu roku. Tento rast ťahali najmä zahraničné investície, ale i rastúca spotreba na Slovensku. V posledných dvoch rokoch sme boli svedkami opätovného poklesu rastu až na 5% v roku 2008 (Obrázok 3). Štatistický úrad SR nezverejňuje dáta zahraničného obchodu v stálych cenách, tie preto neboli predmetom analýzy.



Obrázok 4: Vývoj celkového dovozu SR (1997-2008) v bežných cenách

5. Porovnanie teritoriálnej štruktúry exportu zo SR v rokoch 2002 a 2008

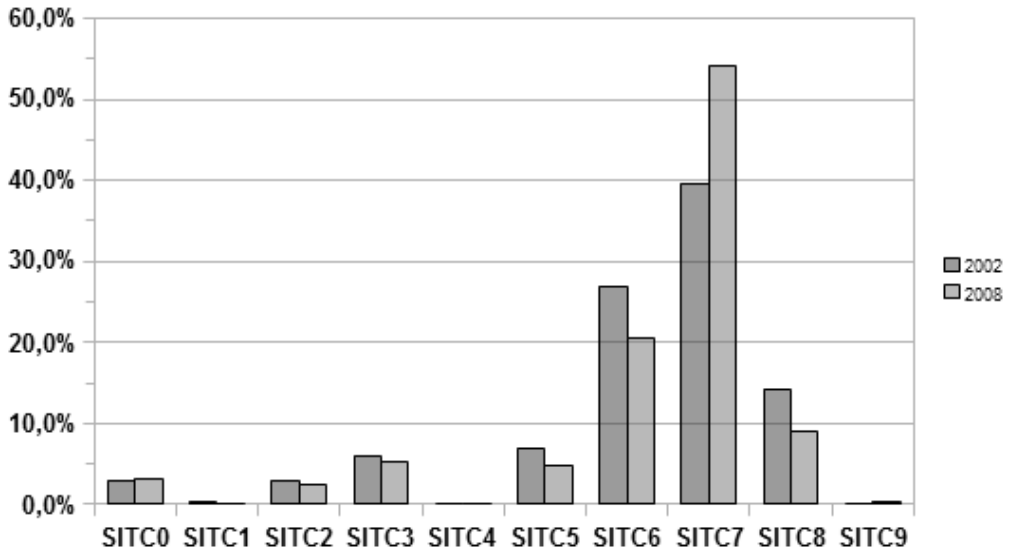
V roku 2002 sa zo Slovenskej republiky vyviezol tovar v celkovej hodnote 652 017,85 milión. SKK (14 477 milión. USD) v bežných cenách. V roku 2008 sa zo Slovenskej republiky vyviezol tovar v celkovej hodnote 1 492 550 milión. SKK (70 273 milión. USD, 47 720 milión. EUR). V roku 2008 bol slovenský export v bežných cenách v SKK až 2,3krát vyšší ako v roku 2002. V dolárovom vyjadrení stúpol až 4,85-násobne. Miera otvorenosti ekonomiky na strane exportu sa zvýšila z 59% na 73,6%. To znamená, že až 73,6% z hodnoty HDP sa vyviezlo za hranice. Pozitívne je možné hodnotiť fakt, že v teritoriálnej štruktúre exportu došlo k diverzifikácii, pretože klesol podiel najväčších obchodných partnerov a stúpol podiel menej významných obchodných partnerov (označených ako „Ostatné“).



Obrázok 5: Porovnanie teritoriálnej štruktúry exportu zo SR v rokoch 2002 a 2008

6. Porovnanie komoditnej štruktúry exportu SR v rokoch 2002 a 2008

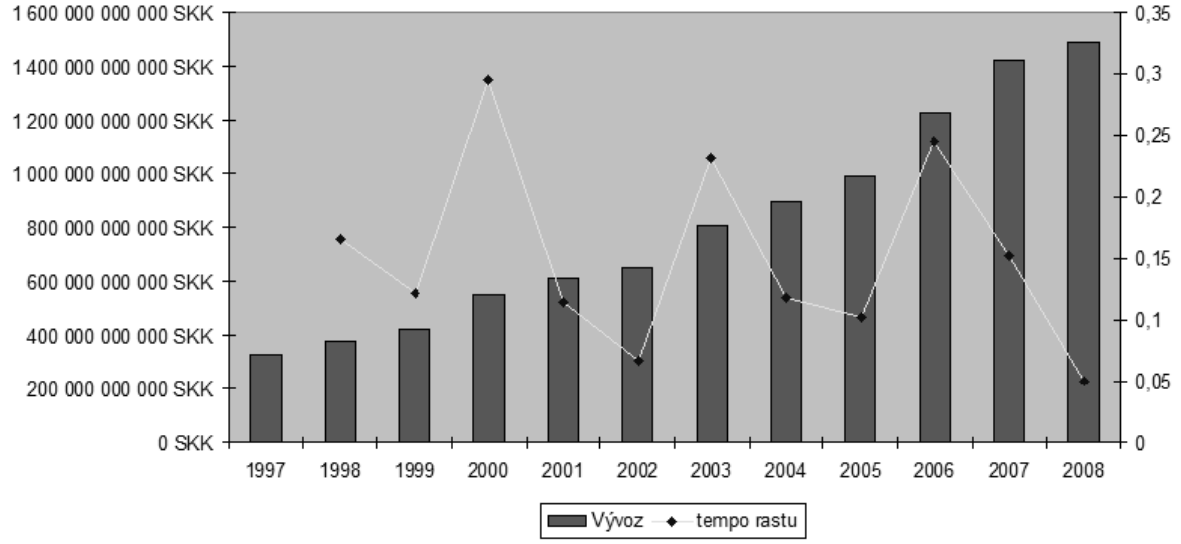
Z hľadiska tovarovej štruktúry podľa klasifikácie SITC došlo v priebehu rokov 2002 – 2008 k významnej zmene v exporte Slovenskej republiky. Ako vidno z obrázku 5, Export strojov a zariadení (SITC7) stúpol o 37% a teraz tvorí až 54,1% percenta z celkového exportu SR. Slovensko sa tak ešte vo väčšej miere stalo závislým napr. od vývozu automobilov (16,4% z vývozu SR) a televíznych a rozhlasových prijímačov a zariadení na záznam zvuku a videa (14,5% z vývozu SR). Za SITC7 nasleduje trieda SITC6 – Trhové výrobky, v ktorej majú veľký podiel na Slovensku už tradične železo, oceľ a ferozliatiny (6,5% z celkového exportu).



Obrázok 6: Porovnanie komoditnej štruktúry exportu SR v rokoch 2002 a 2008

7. Analýza časového radu exportu tovarov zo SR v rokoch 1997-2008

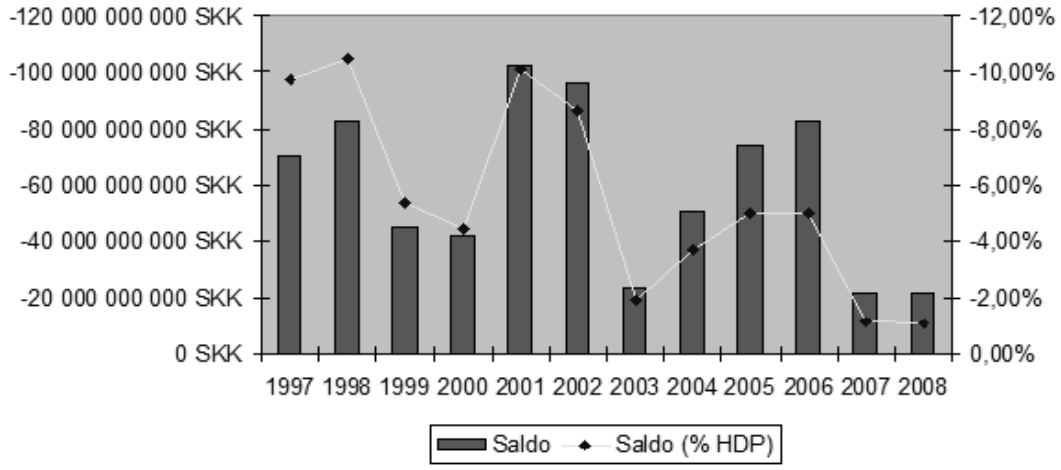
Pri analýze znakov „celkový export v bežných cenách“ a „koeficient rastu exportu“ štatistickej jednotky „Slovenská republika“ som dospel k nasledovným zisteniam. V sledovanom období rokov 1997-2008 stúpol celkový export tovarov za rok v bežných cenách zo Slovenska 4,61-krát. V roku 1997 export predstavoval sumu 324 017 milión. SKK, v roku 2008 to už bolo 1 492 550 milión. SKK. Priemerný koeficient rastu vypočítaný ako geometrický priemer z koeficientov rastu za každý rok má hodnotu 13,3%. Rovnako ako pri koeficiente rastu importu, aj tu pozorujeme významné odchýlky. Po tom ako bol v roku 1998 zaznamenaný nárast exportu o 16,6% oproti roku 1997, nasledoval pokles rastu na 12,1% v roku 1999. Nasledoval rekordný nárast o 29,5% v roku 2000, nasledovaný dvoma rokmi poklesu rastu až na úroveň 6,7% v roku 2002. Počas rokov 2003-2007 sme zaznamenali obdobie prudkého rastu exportu (každoročne od 10,2% do 24,5%). V roku 2008 však rast klesol až na hodnotu 5,1%, čo je minimum počas sledovaného obdobia (Obrázok 6).



Obrázok 7: Vývoj celkového vývozu SR (1997-2008) v bežných cenách

8. Bilancia zahraničného obchodu SR

V sledovanom období rokov 1997-2008 Slovensko vždy vykázalo pasívnu obchodnú bilanciu. Pomerne k HDP to bolo najviac v roku 1998: až 10,5% HDP, najmenej v roku 2008: 1,06% HDP. V absolútnych číslach bolo najvyššie pasívne saldo zahraničného obchodu v roku 2001 a to 102 745 milión. SKK, najnižšie v roku 2007 – 21 384 milión. SKK. V priebehu 6 rokov sa tak saldo zahraničného obchodu znížilo o viac ako 81 miliárd SKK. Túto skutočnosť pripisujeme najmä tomu, že podniky, do ktorých predtým zo zahraničia plynuli investície, rozbehli svoju produkciu naplno v posledných dvoch rokoch. K poklesu salda zahraničného obchodu teda prispel aj podstatný pokles prílevu priamych zahraničných investícií – z úrovne 185,6 miliárd SKK v roku 2002 na len 27,4 miliardy SKK v roku 2007. To, že Slovensko v rokoch 2007 a 2008 nezaznamenalo aktívnu obchodnú bilanciu, pripisujeme najmä rastúcej spotrebe zahraničných tovarov na Slovensku. Napríklad import televíznych a rozhlasových prijímačov a prístrojov na reprodukciu a záznam zvuku sa v roku 2008 zvýšil oproti roku 2003 až takmer 13-násobne na úroveň 142,6 miliardy SKK. Rovnako majú na importe už tradične stále vysoký podiel palivá, najmä ropa a zemný plyn (viď Obrázok 7).



Obrázok 8: Vývoj salda zahraničného obchodu SR v rokoch 1997-2008

9. Záver

Ukazovatele zahraničného obchodu Slovenskej republiky sa v poslednej dekáde značne zlepšili. Od roku 1998 do roku 2008 rástol import vyjadrený v bežných cenách v priemere 10,5%-ným tempom ročne a export až 13,3%-ným tempom. Zvýšila sa miera otvorenosti našej ekonomiky na jednu z najvyšších v Európskej únii. To na jednej strane umožňuje Slovensku čerpať všetky z výhod medzinárodného obchodu, na druhej strane ho to robí viac závislým od ekonomiky iných krajín - jeho obchodných partnerov. Pozitívny je fakt, že došlo k diverzifikácii teritoriálnej štruktúry slovenského importu, ako aj exportu. Slovensko je ale stále veľmi závislé od obchodu s členskými krajinami EÚ a obchod s mimoeurópskymi krajinami je napriek miernemu nárastu stále veľmi malý. Opačný trend ako teritoriálna štruktúra vykazuje komoditná štruktúra slovenského exportu. Tu došlo v priebehu predchádzajúcich rokov (2002-2008) k značnej špecializácii sa na výrobu automobilov, televíznych a rozhlasových prijímačov. Takáto orientácia výroby na tovary s nízkou pridanou hodnotou dokáže síce krátkodobo dostať krajinu z ekonomických problémov, nezaručuje však napredovanie ekonomiky aj do budúcnosti. Naopak treba rátať s poklesom tempa rastu slovenského exportu, ktorého dôvodom je nielen pokles dopytu po našich výrobkoch v zahraničí, ale aj rapídny pokles prílevu priamych zahraničných investícií.

Literatúra

[1] CONDÍKOVÁ, J. *Zahraničný obchod Slovenskej republiky* [online]. 12/2008. Bratislava : Štatistický úrad SR,Mar. 2009 [cit. 2009-11-01]. Dostupné na internete: <http://portal.statistics.sk/showdoc.do?docid=5727>.

[2] CHAJDIAK, J. *Základy hospodárskej štatistiky*. Bratislava : Statis, 2004. 192 s. ISBN 80-85659-38-7.

[3] JURIGA, J. *Analýza zahraničného obchodu krajín sveta.* (Bakalárska práca) Bratislava : Univerzita Komenského, 2009, 48s.

[4] PACÁKOVÁ, V. *Štatistika pre ekonómov*. Bratislava : Iura Edition, 2003. 358 s. ISBN 80-89047-74-2.

[5] ŠTATISTICKÝ ÚRAD SR. *Intrastat-SK* [online]. [cit. 2009-11-01]. Dostupné na internete: <http://portal.statistics.sk/showdoc.do?docid=3752>.

Adresa autora:

Bc. Ján Juriga, Fakulta managementu, Univerzita Komenského v Bratislave
Ul. Odbojárov 10, 820 05 Bratislava, e-mail: janjuriga@aol.com

Zmena reprodukčného správania obyvateľov Bratislavy za posledných 20 rokov a perspektíva budúceho vývoja

The Change of Reproductive Behaviour of Bratislava Population in Last 20 Years and Perspective of Future Development

Michal Katuša

Abstract: This article is focused on changes in reproductive behaviour of Bratislava population, in last 20 years, which were caused by second demographic transition. As Bratislava population is in advance before other population of Slovakia in most of demographic characteristics, it is necessary to aim the study to future development of reproductive behaviour of Bratislava population. This work tries to forecast future reproductive behaviour of young people in Bratislava using a survey study made in 2009 on sample of 1011 respondents.

Key words: Bratislava, second demographic transition, reproductive behaviour, survey study

Kľúčové slová: Bratislava, druhý demografický prechod, dotazníkový prieskum

1. Úvod

Populácia Bratislavy ako aj celého Slovenska zaznamenala v posledných 20 rokoch výrazné zmeny v reprodukčnom správaní. Tieto zmeny súvisia s druhým demografickým prechodom, ktorý mení demografické správanie obyvateľov, pričom obyvateľstvo Bratislavy v mnohých demografických ukazovateľoch, ovplyvnených druhým demografickým prechodom, predbieha vo vývoji ostatné obyvateľstvo Slovenska (Katuša M., 2007, 2008). Preto je na mieste zamerať cieľ výskumu na budúci vývoj reprodukčného správania obyvateľov Bratislavy.

V súvislosti s budúcim vývojom sa v tomto článku snažím porovnať základné charakteristiky reprodukčného správania obyvateľov Bratislavy za posledných 20 rokov s výsledkami dotazníkového prieskumu medzi mladými ľuďmi Bratislavy, ktorý bol uskutočnený v septembri a októbri 2009. Tento dotazník bol zameraný na reprodukčné a rodinné správanie mladých ľudí Bratislavy a zachytáva ich predstavy a plány na život. Bolo zozbieraných 1011 dotazníkov od obyvateľov Bratislavy vo veku 20-30 rokov. Vzorku tvorilo 616 žien a 395 mužov. Z týchto údajov sa môžeme pokúsiť odhadnúť budúci vývoj reprodukčného a rodinného správania obyvateľov Bratislavy, ako aj celého Slovenska.

2. Plodnosť žien podľa veku

Za posledných 20 rokov zaznamenala plodnosť najvýraznejšie zmeny hlavne v intenzite a časovaní plodnosti. Časovanie plodnosti malo stály priebeh, pričom priemerný vek pri pôrode sa kontinuálne zvyšoval počas celého obdobia. Tento nárast je spôsobený na jednej strane odkladaním sobášov a následne aj pôrodov do vyššieho veku, v čase nestabilnej ekonomickej situácie po roku 1989 a na druhej strane zmenami v správaní a hodnotovom rebríčku obyvateľov. Odkladanie a zmeny v hodnotovom rebríčku obyvateľov spôsobila socio-ekonomická transformácia po roku 1989. Tieto zmeny mali však rôznu charakteristiku

u žien v rôznych obdobiach ich reprodukčného obdobia (Potančoková, Vaňo, Pilinská, Jurčová, 2008).

Zmeny v hodnotovom rebríčku sú považované za jeden z prejavov druhého demografického prechodu, pričom sa menia priority obyvateľov. Na prvých miestach už nefiguruje dieťa, rodina a partner, ale prevláda individualizmus a teda na prvom mieste je kariéra, čo sa potvrdilo aj v dotazníkovom prieskume (tabuľka č.1), kde až 39,24% respondentov zaradilo na prvé miesto ako prioritný cieľ v dohľadnej dobe profesionálne uplatnenie, a na druhej strane sa založenie rodiny a túžba po vlastných deťoch umiestnili na posledných miestach.

Tabuľka č.1 : Ciele ktoré by v dohľadnej dobe chceli respondenti dosiahnuť v dohľadnej budúcnosti

poradie dôležitosti	podiel respondentov v %						
	profesionálne uplatnenie	**dosiahnuť vedúce postavenie**	**spokojné užívanie života**	**mať vlastné deti**	**mať veľa peňazí**	**cestovať a poznávať cudzie krajiny**	**založiť rodinu(uzavrieť manželstvo)**
1.	39,24	4,28	28,19	1,79	4,88	6,77	14,84
2.	22,12	13,41	22,02	11,31	7,81	9,41	13,91
3.	13,93	11,22	17,84	13,03	17,43	14,43	12,12
4.	11,14	12,45	13,55	14,76	20,18	12,65	15,26
5.	6,33	17,49	8,24	19,10	16,38	17,99	14,47
6.	4,63	16,60	6,14	18,71	18,11	14,69	21,13
7.	2,92	24,55	3,92	21,43	15,09	24,14	7,95

V kombinácii s rastúcou vzdelanosťou a emancipáciou žien sa posúva pôrodný vek žien do vyšších vekov. Pokým maximum vekovo špecifických mier plodnosti žien bolo v roku 1992 okolo veku 23 rokov v roku 2007 je to až vo veku 30 rokov (graf č.1).

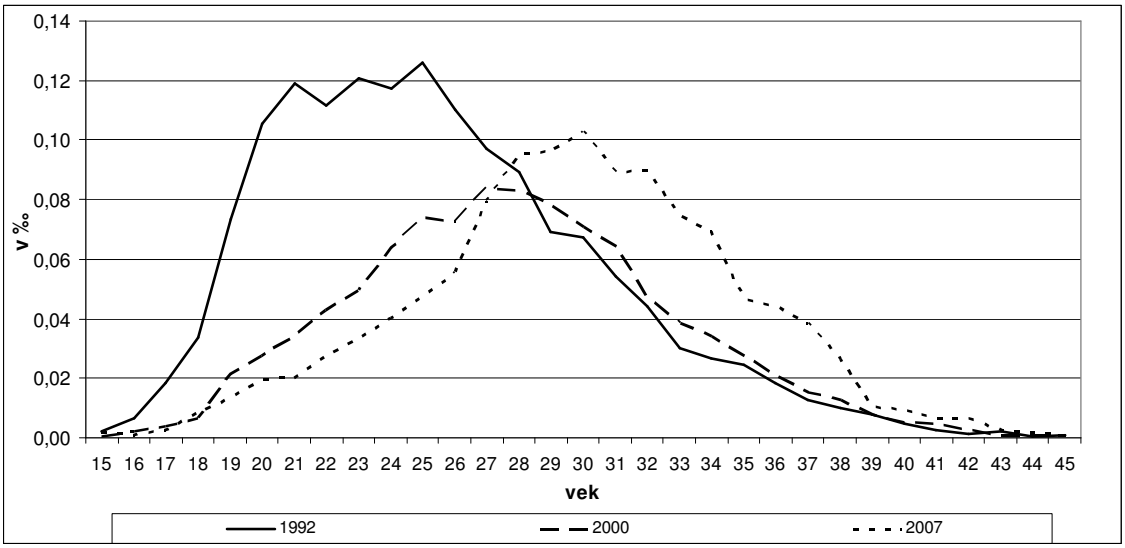
Dochádza aj k istej koncentrácii plodnosti, oproti roku 1992, kedy bola plodnosť žien rozdelená rovnomerne do viacerých vekov (20-26), na druhej strane v roku 2007 sa vytvára isté úzke vekové rozhranie v živote žien, kedy je najvyššia koncentrácia pôrodov a to vo veku 28-30 rokov. toto obdobie sa ukazuje ako ideálne pre mnohé ženy, keďže už aj ženy s vysokoškolským vzdelaním majú v tej dobe istý čas odpracovaný a isté profesionálne uplatnenie už dosiahli. Potvrdzuje to aj dotazníkový prieskum kde až 58 % respondentov označilo za vhodný čas mať dieťa možnosť „až po istom čase v zamestnaní“, 22 % označilo možnosť „po využití možnosti spoznať svet“ a viac ako 16 % označilo možnosť „po dosiahnutí vedúceho postavenia“.

Zmena intenzity plodnosti žien má v priebehu sledovaného obdobia 2 fázy. V prvej fáze dochádzalo k výraznému poklesu intenzity plodnosti a to najmä v nižších vekoch (graf č.1), na jednej strane bol tento pokles spôsobený odkladaním sobášov a pôrodov do vyšších vekov, ale najmä zmenou v správaní a prechodom k modelu 1. detnej rodiny ako ideálu rodiny, a teda sa neuskutočňovali mnohé pôrody vyššieho poradia. Táto prvá fáza trvala do konca 90. rokov. V druhej fáze od roku 2000 zaznamenávame nárast intenzity plodnosti a to najmä vo vyšších vekoch (graf č.1), keďže dochádza k rekuperácii odložených pôrodov.

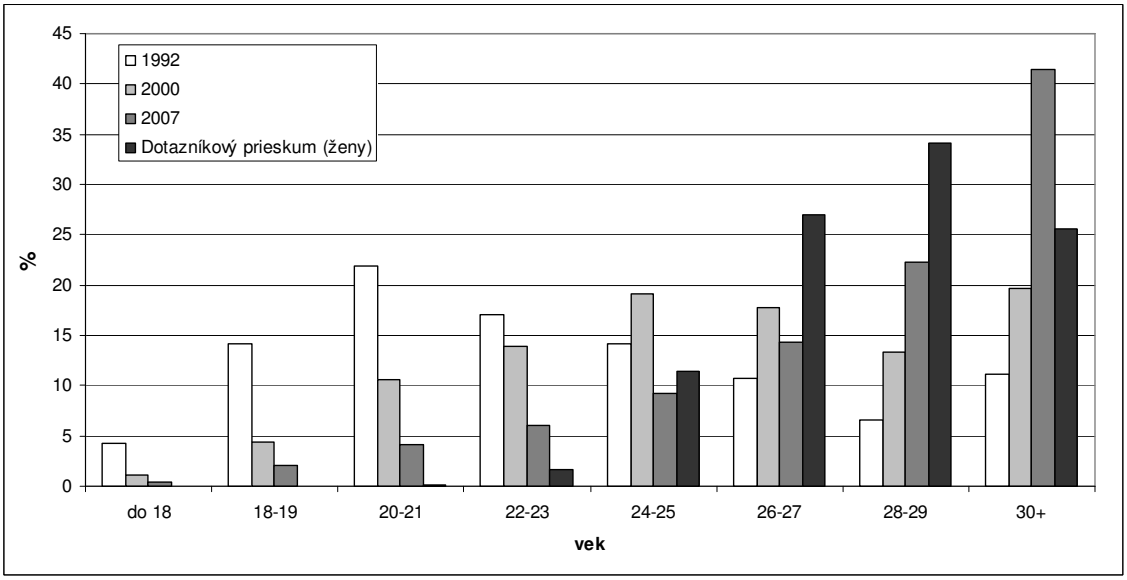
Zmeny v časovaní a intenzite pôrodov možno lepšie pozorovať na podiely narodených detí 1.poradia, kde nevstupujú pôrody vyšších poradí. Z grafu č.2 je zjavný posun medzi rokmi 1992 a 2007. V roku 1992 bol najvyšší podiel pôrodov 1. poradia vo veku 20-21

a podiel pôrodov je rozložený rovnomernejšie medzi viac vekov. Na druhej strane v roku 2007 bolo až viac ako 40% pôrodov prvého poradia po 30. veku matky.

Dotazníkové zisťovanie ukázalo, že viac ako 80 % bezdetných respondentiek chce mať prvé dieťa až po 26 roku, pričom maximum je vo veku 28-29 rokov.



Graf č.1: Vekovo špecifické miery plodnosti žien Bratislavy



Graf č.2: Podiel narodených detí 1. poradia v jednotlivých vekových intervaloch v rokoch 1992, 2000 a 2007 a plánovaný vek počatia 1. dieťaťa ženám bez detí z dotazníkového prieskumu

3. Plodnosť žien podľa poradia

Zmeny v hodnotovom rebríčku ľudí v rámci druhého demografického prechodu a do istej miery aj zmeny socio-ekonomickej situácie sa odrážajú aj na zmenách podielu narodených detí podľa poradia. Dominantným modelom sa stáva 1 detná rodina, čo má za následok nárast podielu pôrodov 1. poradia. tento podiel medzi rokmi 1992 a 2007 vzrástol o 10 % z 48,6 % na 58,6 %. Pokles zaznamenávame v podieloch 2., 3. a 4 a vyššieho poradia.

Najvyšší pokles zaznamenali pôrody druhého poradia a v roku 2007 tvorili 31,9 % všetkých pôrodov. pôrody 4 a vyššieho poradia poklesli o polovicu z 4,4 % na 2,2 %.

Zaujímavým zistením z dotazníkového prieskumu je želaný počet detí opýtaných respondentov, ktorí väčšinou chcú mať dve deti (65,59% ženy a 65,57% muži). Výrazne vysoké podiely zaznamenali aj troj detné modely a na druhej strane výrazne nízky v porovnaní z reálnymi údajmi z roku 2007 je podiel jedno detného modelu ako u mužov tak aj u žien. Do istej miery je tento stav zapríčinený výchovou, alebo skôr ideálnou predstavou o rodine, ktorá sa stále prejavuje tým, že väčšina respondentov chce v budúcnosti 2 deti. Avšak jedná sa o chcený počet detí, istý ideál, ktorý býva vo väčšine prípadov vyšší, ako počet detí, ktoré nakoniec reálne budú ľudia mať. Reálny počet detí, je súhrnom mnohých faktorov a v mnohých prípadoch sú hlavným faktorom tie ekonomické, ktoré výrazne prispievajú k znižovaniu pôrodnosti. Keďže väčšina respondentov prioritne chce dosiahnuť profesionálne uplatnenie, spokojné užívanie života a mať veľa peňazí (tabuľka č. 1) a až potom mať deti je možné, že ten stav dosiahnu až vo vysokom veku alebo sa ich predstava po skúsenostiach v živote zmení. Nezanedbateľným je aj fakt, že mať dieťa je ekonomicky veľmi náročné a v súčasnosti ani štát nevytvára ideálne podmienky, ktoré by motivovali mladých ľudí mať deti. Negatívne v mnohých prípadoch vplýva aj prvé dieťa na rozhodnutie či mať aj druhé, aj kvôli už spomínaným ekonomickým faktorom.

Tabuľka č.2: Podiel narodených detí podľa poradia a podiel počtu želaných detí

Rok	Podiel narodených detí podľa poradia v %			
	1.	2.	3.	4+
1992	48,61	36,95	10,07	4,36
2000	56,42	30,91	8,83	3,84
2007	58,61	31,87	7,27	2,24
	Podiel želaného počtu detí v %			
Dotazník (ženy)	9,54	65,59	21,98	2,90
Dotazník (muži)	6,08	65,57	17,72	3,04

4. Záver

Pôrodnosť zaznamenala v posledných 20 rokoch výrazné zmeny, ktoré ovplyvnili aj reprodukciu obyvateľstva Bratislavy. Tieto zmeny v pôrodnosti môžeme rozdeliť do dvoch fáz. V prvej fáze trvajúcej do konca 90. rokov zaznamenáva populácia Bratislavy výrazný pokles intenzity plodnosti a zároveň aj posun v časovaní plodnosti. Pokles intenzity je do istej miery spôsobený odkladom sobášov a teda zmenami v časovaní pôrodov, ale taktiež zmenami v rodinnom správaní súvisiacimi s druhým demografickým prechodom, konkrétne v zmenách v hodnotovom rebríčku. rozmáha sa individualizmu, emancipácia žien, vzdelanosť žien sa zvyšuje aj ich kariérne a osobné ambície, čo posúva rodinu a deti na nižšie miesta v hodnotovom rebríčku, čo dokazuje aj dotazníkový prieskum uskutočnený medzi mladými ľuďmi Bratislavy.

Druhá fáza začína počiatkom 21. storočia a je charakteristická pokračujúcimi zmenami v časovaní avšak už miernym nárastom intenzity plodnosti, z dôvodu rekuperácie odložených pôrodov. V súčasnosti prevláda model jednodetnej rodiny ako ideálneho konceptu, čo sa prejavuje aj v dominancii podielu pôrodov 1. poradia, aj keď želaný počet detí je u mladých ľudí vyšší.

V súčasnosti sa oproti roku 1992 výrazne skoncentrovala plodnosť žien do niekoľkých vekov. Najvyššie hodnoty vekovo špecifických mier plodnosti dosahujú ženy vo veku 28-32, pričom v roku 1992 bola plodnosť rovnomerne rozdelená do rokov 19-28. Tak isto aj

dotazníkové zisťovanie potvrdzuje, že toto obdobie života ženy (28-32 rokov) sa stáva môžeme povedať ideálnym vekom pre pôrod. Akýmsi kompromisom medzi individualistickým zmýšľaním žien, kedy dávajú ženy prednosť vzdelaniu, kariérnemu postupu a vlastným záujmom, na jednej strane a altruistyckým zmýšľaním, kedy ženy začnú myslieť na rodinu a deti, na druhej strane.

5. Použitá literatúra:

[1] KATUŠA M., 2007, Demografická analýza Bratislavy - Diplomová práca. Prírodovedecká fakulta Univerzity Komenského 2007.

[2] KATUŠA M., 2008, Zmeny v rodinnom a reprodukčnom správaní obyvateľov Bratislavy po roku 1989. Forum Statisticum Slovacum, 7/2008.

[3] POTANČOKOVÁ M., VAŇO B., PILINSKÁ V., JURČOVÁ D., 2008, Slovakia: Fertility between tradition and modernity. Demographic research, vol. 19, july 2008, 973-1018.

6. Zdroje údajov:

[1] HLÁSENIA O NARODENÍ 1992-2007, Štatistický úrad Slovenskej republiky, údaje na požiadanie.

[2] DOTAZNÍKOVÝ PRIESKUM 2009, Michal Katuša

Adresa autora:

Michal Katuša Bc.
Prírodovedecká fakulta UK v Bratislave
michal.katusa@gmail.com, aso@chello.sk,

Potreba chrániť dáta vo výpočtovej štatistike
Data Security in the Computational Statistics

Róbert Krchnavý

Abstract: Computers are everywhere and are used mostly for everything. One very useful domain of them is to simplify computation and calculations of the statistics also. There are many different statistics and they use sensitive data very often. It should be a basic thing to think about their security. Unfortunately it is not so obvious for everyone. And lots of data could be a serious problem with damage like financial loss, loss of reputation and many other impacts. In this article we focus on the techniques and solutions used for securing and protection data, way to implement them and possible consequences in case we do not give the necessary attention to the security threats.

Key Words: access, authentication, data, database, encryption, key, password, protection, security, statistics

Kľúčové slová: autentifikácia, bezpečnosť, databáza, dáta, heslo, kľúč, ochrana, prístup, šifrovanie, štatistika

1. Úvod

Počítače máme všade a už radu rokov sa využívajú pre štatistiku a výpočtové spracovávanie štatistických údajov. Výsledky rôznych prieskumov, štatistických analýz a podobne nás obklopujú z každej strany a stretávame sa s nimi denne. Existuje však aj nepreberné množstvo štatistických analýz v rámci firemného prostredia, mnohokrát narábajúce s citlivými údajmi. Často sa analýzy nechávajú spracovávať externými firmami a aj tu vzniká otázka, ako tieto organizácie chránia naše dáta. Pri spracovávaní údajov by sme mali totiž vždy pamätať na ich bezpečnosť a to, že veľká časť z nich môže byť zneužitá ak sa dostane do nesprávnych rúk. Riziko zneužitia je naozaj vysoké. Technické prostriedky všade dostupné a znalosti mnohých, v tejto oblasti, veľmi nízke. Naopak, ak má niekto záujem a vie, že by mohol zo získaných údajov prosperovať, je pre neho veľakrát až príliš jednoduché sa k nim dostať. Na nasledujúcich riadkoch si prejdeme základné princípy ochrany dát, ako aj to prečo ich aplikovať a aké následky môže mať ich zanedbanie.

Na oblasť bezpečnosti dát sa pozrieme z dvoch pohľadov. Prvým bude zabezpečenie dát kontrolou prístupu, t.j. že len legitímny užívateľ sa dostane k údajom, ku ktorým má mať prístup. Druhým už bude samotná ochrana dát na pracovných staniciach, t.j. šifrovanie.

2. Kontrola prístupu a spôsoby autentifikácie

Väčšina firiem, úradov ale aj jednotlivcov sa dnes spolieha na počítačové siete, a tak im z hľadiska bezpečnosti väčšinou venujú i dostatok potrebnej pozornosti. Menej z nich však už pozerá na zabezpečenie uložených dát a to predovšetkým na koncových staniciach. Štandardné zabezpečenie na desktopoch a notebookoch je dnes ľahko prekonateľné. Na prekonanie hesla do Windows dnes stačí niekoľko minút a nie je k tomu potrebné nič viac než internet. Všeobecne platí, že bezpečnosť má šancu len vtedy, keď bude užívateľsky nenáročná. Kto si má však pamätať desiatky hesiel a neustále ich niekam zadávať? I keď sa už nejakú dobu používajú aj Smart Cards alebo rôzne USB tokeny, ktoré zvyšujú bezpečnosť

autentizácie, stále nútia užívateľov nosiť so sebou (a nezabúdať) ďalšie z „nevyhnutných“ denných pomôcok. Nehovoriac o tom, že s ťažkosťou dosiahneme, aby pracovníci nenechávali „múdre“ karty v notebookoch a čítačkách.

Na autentizáciu a identifikáciu užívateľov máme viacero možností. Poďme sa pozrieť, ktoré to sú, aké sú ich výhody a nevýhody.

MENÁ A HESLÁ

Počítačové heslá paria medzi najstarší a najčastejší spôsob autentizácie užívateľa, stali sa bežnou súčasťou celého radu aplikácií. Ide aj o najlacnejšie riešenie, ktoré pri dobre nastavenej bezpečnostnej politike stále dobre slúži. Sú však najviac náchylné k zneužitiu. Každý z nás používa niekoľko hesiel a povedzme si úprimne, väčšinou veľmi primitívnych odvodených od mien príbuzných a pod. Heslo takejto „kvality“ je možné zlomiť behom pár sekúnd a netreba na to žiadne extra vedomosti, len jednoduchý generátor stiahnutý z internetu. Ďalší nepríjemný scenár vzniká, ak sú heslá uložené v otvorenej forme, t.j. nešifrované, a teda stačí len vedieť, kde ich hľadať. Tu nám nepomôže ani heslo silné. Silným heslom hovoríme takému, ktoré nie je odvodené od zmysluplného slova/textu, obsahuje číselné údaje a pokiaľ je možné i špeciálne znaky. Veľkú úlohu tu teda hrá disciplína užívateľov a uvedomovanie si dôležitosti pravidelnej zmeny hesiel. Heslá majú totiž obrovskú výhodu – sú pomerne jednoduché na správu, užívatelia si na ne zvykli a náklady spojené s ich administráciou patria medzi najzanedbateľnejšie položky.

Ukladanie hesiel

Z dôvodu bezpečnosti nemôžme ani len uvažovať nad ukladaním hesiel v otvorenej forme, tým strácajú na význame. Slušné zabezpečenie prináša ukladanie hesiel v podobe zodpovedajúceho odtlačku. Pre získanie odtlačku z hesla, sa využívajú rozličné hash funkcie, teda jednosmerné funkcie, u ktorých je možné zo vstupu ľahko vypočítať výstup, avšak z výstupu je výpočtovo zložité získať pôvodný vstup. Klasické prirovnanie s objektami reálneho sveta môže byť na pohári. Hash funkcia funguje ako rozbitie pohára – jej rozbitie je veľmi jednoduché, avšak presne zložiť späť z črepín celkom nemožné. Hash funkcia sa navyše pri malej zmene na svojom vstupe prejaví veľkou zmenou na výstupe.

Tento spôsob však nerieši otázku bezpečnosti úplne. V prípade ukladania len odtlačkov hesla (a porovnávaní pri snahe prihlásiť sa) útočník môže uspieť pri jednoduchom odpočúvaním na sieti, kde by získal login a príslušnú hash hodnotu hesla. Tá by síce sama o sebe nič nehovorila, ale v kombinácii s prihlasovacím menom sa ponúka možnosť využiť útok, keď sa útočník zaslaním zodpovedajúcej zachytenej dvojice (login, odtlačok) sám mohol pokúsiť o prihlásenie k cudziemu účtu. V praxi býva tento problém riešený rôznymi časovými známkami, keď sú prihlasovacie údaje doplnené ďalším parametrom, ktorý im dáva platnosť iba v rámci stanoveného časového okna – legitímny užívateľ vie svoje heslo, zadá ho, vytvorí sa hash hodnota + časová známka a to sa overí s hodnotou uloženou v systéme. V prípade odchytenia takejto dvojice útočníkom mu to nepomôže – iná časová známka, iný hash rovnakého hesla sú teda vďaka pridaniu rôznych solí, výsledné odtlačky rôzne (odlišné).

Ideálne heslo

Na túto otázku sa dá pozerať z rôznych pohľadov. Samozrejmé je, že iné heslo bude vyžadovať užívateľ, ktorý chce chrániť pár svojich „citlivých“ dokumentov, a iné užívateľ pre prístup k vysoko tajným súborom či aplikáciám. Za všeobecné platné pravidlá môžeme vymenovať nasledovné:

- heslo by malo obsahovať malé aj veľké písmená z abecedy, číslice aj špeciálne znaky,
- heslo by malo byť dostatočne dlhé; 10 a viac znakov,

- heslo by nemalo nijako súvisieť s už skôr (predtým) používanými heslami, nemalo by obsahovať meno alebo login užívateľa a taktiež by nemalo byť bežným slovom či menom.

Ani to však nezabezpečí úplnú ochranu, ak je zle implementovaná hash funkcia, alebo iná časť reťazca pracujúca s heslom a jeho administratívou. Z nášho pohľadu bežných užívateľov je to ale maximum čo môžeme spraviť ohľadom hesla.

Autentizácia pomocou grafických prvkov

O svoje miesto pod slnkom sa v ostatnom čase bijú aj rôzne grafické varianty hesiel. O čo ide? Pri použití grafických hesiel užívateľ nezadáva klasický textový vstup, ale rôznymi spôsobmi vyberá dopredu definované grafické symboly. Pre jednoduchšiu orientáciu a jednoduchšie zapamätanie je pozornosť sústredená predovšetkým na ikony bežných vecí (lopta, dom, more a pod.), pričom sa rôzne algoritmy líšia hlavne v postupe volenia správnej kombinácie. Vo fáze prihlasovania sa potom tieto obrázky zobrazia náhodne rozmiestnené v kope ďalších grafických motívov – napríklad tri obrázky tvoriace heslo sa na prvý pohľad ľahko stratia v matici ďalších, o rozmeroch napríklad 10 x 10. Akonáhle užívateľ identifikuje svoju prihlasovaciu trojicu, musí tlačítkom myši klepnúť do pomyselného trojuholníka, ktorého vrcholy tvoria. Pre väčšiu bezpečnosť je samozrejme vhodné, aby tieto ohraničené oblasti boli čo najmenšie a náhodne rozmiestnenie s identifikáciou sa i niekoľkokrát opakovalo.

SMARD CARDS

Múdre karty si vyžadujú k svojej prevádzke čítačku. Sú už pomerne rozšírené, podporované systémy a nie príliš nákladné. Je možné ich ľahko využiť pre pokročilejšiu autentizáciu do siete a súčasne ako rýchly identifikačný prvok pre kontrolu prístupu v konkrétnych priestoroch. Užívatelia ich môžu zabudnúť alebo stratiť, čo komplikuje ich správu.

USB TOKENY

Sú veľmi podobné múdrym kartám, ale nevyžadujú čítačky, je možné ich pripojiť priamo k počítaču. Na druhej strane ich nevyužijete pre kontrolu prístupu v budovách a nezmestí sa na ne fotografia užívateľa pre jeho rýchlu identifikáciu. USB tokeny sú lacnejšie ako múdre karty a poskytujú takmer porovnateľné bezpečnostné funkcie.

ČÍSELNÉ TOKENY

Zariadenia, ktoré generujú jednorázové heslá, ktoré sú synchronizované s autentizačným systémom. Využívajú sa najčastejšie pre vzdialený prístup. Cena je porovnateľná s USB tokenmi, ale je nutné počítať s nákladmi na serverovú časť.

BIOMETRIKA

Existuje veľa spôsobov overovania na základe biometrických údajov, ktoré sa líšia cenou a mierou zabezpečenia. Technológia zatiaľ nie je natoľko vyspelá ako múdre karty a tokeny, ale má zďaleka najväčší potenciál. Je bezpečnejšia a nie je toľko závislá na disciplíne užívateľov. Avšak je tu riziko zneužitia a počet náhradných prstov je obmedzený.

Snímače odtlačkov prstov sú dnes dostupné v dvoch variantoch – buď je snímaný odtlačok ako reliéf prstu, podobne ako napr. v policajnej kartotéke, alebo vo variante, keď snímač sníma obraz povrchového napätia prstu. V tomto prípade nefungujú útoky možné na prvý druh snímača, t.j. podstrčenie „mŕtveho“ prsta (kópie), lebo prst oddelený od ruky stráca približne po desiatich minútach svoju štandardnú povrchovú vodivosť a snímač teda nemôže autentizáciu uskutočniť. Veľká časť snímačov bežne dostupných na trhu umožňuje iba zjednodušenie prihlasovania do systému, nezvyšuje v podstate bezpečnosť systému.

3. Šifrovanie dát a bezpečnostná politika

I po správnom aplikovaní autentizácie užívateľov je stále nutnosťou dáta chrániť i šifrovaním. Šifrovanie dát je problém, ktorý nie je potrebné príliš vysvetľovať. Každý z nás chápe, k čomu je také šifrovanie dobré a prečo ho používať. A možno nie každý. Inak by sme sa pravdepodobne denne nedočítavali o strate dát z notebookov zabudnutých v lietadle alebo aute. O strate dát za miliardy dolárov, informáciách o stovkách tisíc zamestnancov alebo zákazníkov renomovaných nadnárodných spoločností. Každá firma, a nielen tá, každý jednotlivec, úrad, organizácia by si mala uvedomiť cenu dát, ktoré máme v počítačoch uložené, a cenu, ktorú v prípade straty zaplatí. Nehovoríme tu len o obnove dát, ale hlavne o tom, že sa môžu dostať do rúk konkurencie alebo byť dokonca zverejnené. Vynaložené prostriedky na bezpečnostné riešenia sú potom typicky len veľmi malým zlomkom v porovnaní s možnými následkami úniku dát z menej zabezpečených systémov pri krádeži či neoprávnenom použití. Napríklad výskum CSI/FBI z roku 2003 zameraný na počítačovú kriminalitu a bezpečnosť dát ukázal, že pre 80% respondentov je zásadným problémom zneužitie vlastnej siete zvnútra. Z prieskumu zároveň vyplynulo, že za krádež citlivých informácií sú často zodpovední nespokojní zamestnanci, ktorí chceli spoločnosti uškodiť. Ľudský faktor je teda obvykle kľúčovým prvkom ovplyvňujúcim bezpečnosť a bezpečnostné prvky.

Jedným z pilierov bezpečnosti je správne nastavená bezpečnostná politika firmy a s tým spojené správanie sa užívateľov. Návrh a implementácia technického riešenia bezpečnosti je v praxi záležitosťou niekoľkých dní. Samozrejme správne nastaviť bezpečnostnú politiku je vec, ktorá zaberie mnoho týždňov, a do tohto procesu musia byť zapojené všetky oddelenia firmy a predovšetkým najvyšší manažment.

Ochrana dát

Poďme sa ale vrátiť od rizík a k ich riešeniu. A jediným, môžeme povedať skoro stopercentným riešením, ochrany citlivých informácií je ich správne šifrovanie. Všetky iné prekážky, ako sú firewally, obmedzenie prístupu alebo politiky Windows, iba útočníkovi sťažujú prácu a predlžujú dobu potrebnú k získaniu prístupu.

Útok na nezabezpečené informácie na harddisku či inom médiu pritom nevyžaduje žiadne špeciálne znalosti ani prostriedky útočníka. Harddisk je možné pripojiť do iného počítača a s dátami pracovať, prípadne po nabootovaní zo špeciálnych CD, ktoré sa dajú ľahko získať z internetu, získať alebo resetovať heslo užívateľa a bez obmedzenia sa prihlásiť jeho menom. Tieto základné metódy dnes zvláda žiak 5. triedy základnej školy a i neznalý užívateľ nájde správny návod na internete maximálne za 5 minút.

Naopak, pokiaľ získa útočník notebook, na ktorom sú všetky citlivé súbory zašifrované, potom získal iba hardware s nič nevypovedajúcim obsahom pevného disku. Predpokladom je samozrejme správna voľba zabezpečenia šifrovacích kľúčov. Tie by mali byť zabezpečené tak, aby ich nebolo možné ľahko získať spolu so získaním prístupu k užívateľskému účtu. Cieľom je vždy čo najskôr zašifrovať užitočné dáta. Tie sú uložené v podobe súborov a adresárov na pevných a vymeniteľných diskoch, prípadne na sieti, a preto musíme obvykle šifrovať aspoň celé adresáre. Užívateľské aplikácie napríklad veľmi rozšíreného balíka Microsoft Office, totiž neupravujú priamo daný zašifrovaný súbor, ale vytvoria si jeho kópiu v rovnakom adresári, ktorá je nechránená. Ak budeme šifrovať na úrovni adresárov, musíme si ďalej uvedomiť aj to, že aplikácie používajú rôzne ďalšie odkladacie a dočasné adresáre, či súbory (tempy) a pod., mali by sme teda šifrovať i tie.

Rovnako je na zváženie, či nešifrovať dáta a konfigurácie systému. Rovnako dôležité je chrániť zavádzacie údaje disku, ako je MBR (Master Boot Record) alebo „boot sector“. To ale súborové šifrovanie z princípu neponúka, k tomu potrebujete šifrovať celý diskový oddiel – samozrejme ten systémový. Tu si treba uvedomiť, že aby sa systém vôbec začal načítavať,

musíme už pri zapnutí počítača zadať prístupové heslo. Šifrovaním systémového oddielu sa teda pridáva i akési riadenie prístupu k počítaču ako takému.

Úrovne šifrovania

Pozrime sa na šifrovanie na úrovni súborov na *tzv. vysokej hladine*. Ako už bolo povedané, ponúka sa možnosť šifrovania jednotlivých súborov a adresárov. Nastavenie je tým pádom veľmi flexibilné. Môžeme rôzne súbory šifrovať rôznymi kľúčmi, a oddeliť tak vzájomne od seba určitých užívateľov. Nie je možné ale obvykle chrániť systémové súbory, stránkovacie súbory ani registre. Dôvodom je fakt, že šifrovacie moduly sa zavádzajú do jadra operačného systému až v priebehu jeho štartovania. Užívateľ tak nemá principiálne možnosť zadať heslo, ktoré by dešifrovalo systém ešte pred jeho štartom. To je klasický problém. Riešením je šifrovanie celého diskového oddielu. Jedná sa o ochranu na *tzv. nízkej hladine*. Operačný systém totiž vidí celé oddiely, ich písmená, súbory a adresáre až po tom, čo sú dešifrované. Máme tak chránené nielen dáta, ale i systém samotný. Tento prístup má aj svoje nevýhody. Užívateľ má prístup buď k celému oddielu alebo k ničomu. Nie je to teda veľmi vhodné na oddeľovanie rôznych skupín užívateľov – teda pokiaľ pre ne nevytvoríme vlastné diskové oddiely. Rovnako tak, pokiaľ sa pristupuje k dátam vzdialene, budú dešifrované už na servery a prenášané cez sieť otvorene.

V tomto ohľade môžeme využiť šifrovanie *na strednej hladine*. Teda šifrovanie virtuálnych súborových oddielov. Program si vytvorí na disku obyčajný súbor. Jeho obsah bude šifrovať. Tento súbor sa začne tváriť ako (virtuálny) disk. Všetko, čo potom uložíte na takýto „disk“ bude v skutočnosti uložené v onom šifrovanom súbore. Nemôžeme tak chrániť systém, ale keď sa pripojíme zo siete, bude sa súbor dešifrovať až u nás na stanici, t.j. prenos bude chránený.

Šifrovacie algoritmy

K zašifrovaniu dát na diskoch používame symetrickú kryptografiu. Tá funguje na princípe jedného kľúča, ktorý je použitý pre šifrovanie a dešifrovanie. (Na rozdiel od asymetrickej kryptografie, kde je k šifrovaniu použitý kľúčový pár (jeden kľúč k zašifrovaniu a druhý k dešifrovaniu). Kľúče môžu byť odvodené buď z užívateľom zadávaného hesla (názvy ako password, passphrase, user key, preshared key) alebo vypočítané za pomoci kalkulátorov, či čipovej karty (tzv. kryptografický predmet). Ako sme už spomínali vyššie i v tomto prípade sú posledné dve možnosti o triedu bezpečnejšie ako jednoduché heslo, pretože ich je nutné mať fyzicky pri sebe.

Výhodou symetrických šifier je ich rýchlosť a vysoká bezpečnosť aj pri relatívne krátkom použitom kľúči. Bežne používaným štandardom vo firemnej a bankovej sfére je dnes 128 bitové šifrovanie niektorou z metód AES (Advanced Encryption Standard). V asymetrickej kryptografii sa naopak používajú kľúče s dĺžkou v tisícoch bitov, najpoužívanejší je v dnešnej dobe je kľúč dlhý 1 024 bitov. Častým omylom potom býva predstava, že asymetrická kryptografia musí byť zákonite bezpečnejšia. Dĺžku kľúča u týchto dvoch metód nemôžeme porovnávať iba jednoduchou logikou – čím dlhší, tým lepší. Bohužiaľ algoritmy sú natoľko odlišné, že algoritmus asymetrický (typický RSA) je i pri použití 10 x dlhšieho kľúča menej odolný voči rozlúšteniu. Navyše vďaka obrovskému rozdielu v rýchlosti je pre šifrovanie dát v rádoch kB alebo MB nepoužiteľný.

K porovnaniu dĺžky kľúčov algoritmov z hľadiska kryptografickej odolnosti voči rozlúšteniu poslúži priložená tabuľka.

	Ekvivalenty v odolnosti šifrovacích algoritmov	
	Symetrický algoritmus AES	**Asymetrický algoritmus RSA**
Dĺžka kľúča v bitoch	56	512
	64	768
	80	1 024
	112	2 048
	128	3 072
	192	7 680
	256	15 360

4. Záver

Prešli sme základné princípy ochrany dát v dnešnej počítačovej dobe a dôvody prečo je nevyhnutné dáta chrániť. Ako vidíme, dostať sa k cudzím dátam nemusí byť až také zložité ako sa zdá a len heslá nás v žiadnom prípade neochránia. Čo všetko sa dá, s takto nadobudnutými dátami v konkurenčnom boji získať, si vie každý z nás veľmi dobre predstaviť. Preto si dávajme dobrý pozor na naše dáta a neverme bezhlavo počítačom a ľuďom okolo nás. Nikdy nevieme kto nás môže sledovať...

5. Literatúra

[1]BITTO, O. : Bezpečná hesla?. In: Connect!, č. 02, 2007, s. 63

[2]BITTO, O.: Hesla, jejich solení a další autentizační koření. In: Connect!, č. 02, 2007, s. 61 – 62.

[3]JIROVSKÝ, V.: Kybernetická kriminalita. 1.vyd. 2007. ISBN 978-80-247-1561-2

[4]ONDRÁČEK, M.: Buďte tajemní, vyplatí se to! In: Connect!, č. 12, 2006, s. 60 – 61.

[5]PUŽMANOVÁ, R.: Pro vaše oči bych vraždil. In: Connect!, č. 03, 2006, s. 18 – 19.

[6]RODRYČOVÁ, D., STAŠA, P.: Bezpečnost informací (jako podmínka prosperity firmy). 1.vyd. 2000. ISBN 80-7169-144-5

[7]ŠEVEČEK, O.: Šifrujte data třikrát jinak. In: Connect!, č. 12, 2007, s. 64 – 67.

Adresa autora (–ov)

Róbert Krchnavý, Bc.
FEI STU, AI, 5.ročník
Ilkovičova 3, 812 19 Bratislava 1
robko.djafro@gmail.com

Vplyv konfesie na demografické správanie obyvateľstva v mikromierke vybraných obcí Slovenska

The influence of religion on demographic behavior of population in microscale of chosen communities in Slovakia

Ivana Madžová

Abstract: All events in the life of people which affect their personal status are recorded in the registry. In the registry book are registered facts about birth, marriage or death of the people, i.e. events that can help us to identify person in our state. Thanks to historical research, which consists of excerption form the registry, we can analyse and compare the basic population processes, like natality, mortality or nupciality in chosen communities. This article shows us differences between demographic behaviour of Lutherans and Catholics in the congregation Veľký Grob – Čataj. The purpose of this work is to provide a reader the sight about population development of chosen religious structure in 19th and 20th century and approximate him whole demographic characteristic of communities in this time.

Key words: demographic behavior, congregation Veľký Grob – Čataj, Lutherans, Roman Catholics, religion, historical research

Kľúčové slová: demografické správanie, farnosť Veľký Grob – Čataj, evanjelici a.v., rímskokatolíci, náboženstvo, historický výskum

1. Úvod

Historická demografia je súčasťou demografickej vedy už oddávna, a preto samotná demografia má výrazný historický aspekt. Analýza populačných zákonitostí je totiž nezmyselná bez poznania populačných trendov z hľadiska dlhodobého vývoja a dôkladnej znalosti meniacej sa sociálnej situácie.

Matričná činnosť si ako jedna z mála zachovala svoju vážnosť a výpovednú hodnotu do dnešných čias. V súčasnosti sa historickému bádaniu na základe excerpcie z matičných spisov venuje málo pozornosti, napriek tomu, že záznamy v matrikách sú jediní svedkovia doby, ktorí nám dovoľujú pozorovať populačné procesy od konca 16. storočia na mikroúrovni (v obciach).

Intenzita religiozity hrala najmä v minulosti dôležitú úlohu v predmanželskom, ako i manželskom správaní. Tento tradičný vplyv rodinného správania sa v súčasnosti oslabuje v dôsledku prebiehajúcej sekularizácie. Dochádza k vytláčaniu náboženského vplyvu zo všetkej ľudskej činnosti. Celé krajiny a národy, v ktorých kedysi prekvitalo náboženstvo a kresťanský život dnes prechádzajú ťažkými skúškami a sú výrazne poznamenané šíriacou sa sekularizáciou. (Bridová, 2004)

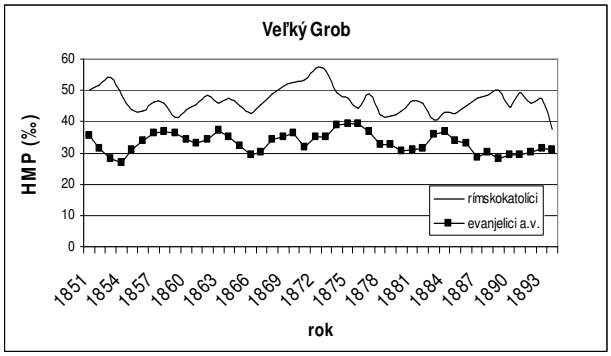
Práca sa sústreďuje na analýzu a komparáciu procesov vybraných náboženských štruktúr v dvoch konkrétnych obciach. Výber obcí Čataj a Veľký Grob bol podmienený konfesionálnym zložením obyvateľstva, pozostávajúceho zo silného zastúpenia evanjelikov a. v. Z hľadiska náboženskej štruktúry tvorili v minulosti obce Veľký Grob a Čataj akúsi enklávu spomedzi okolitých obcí regiónu. V súčasnosti sa pomery v obciach zmenili, preto sme si dovolili nahliadnuť do historických záznamov a prostredníctvom analýzy jednoduchých demografických ukazovateľov sme sa pokúsili nájsť rozdiely v demografickom správaní evanjelikov a.v. a rímskokatolíkov. V našich výskumoch sme vychádzali zo všeobecnej hypotézy mnohodetných rímskokatolíckych rodín a menejdetných (jednodetných)

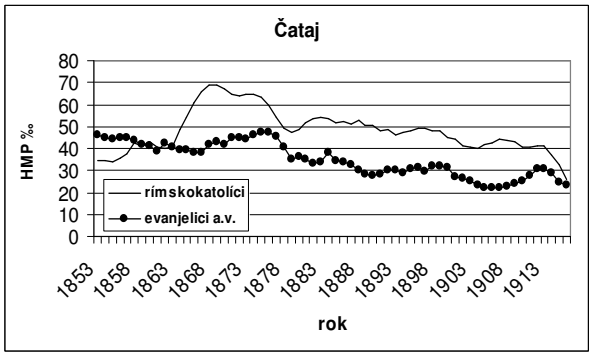
evanjelických a.v. rodín. Naše poznatky o dogmatizme a sile jednotlivých náboženstiev sme sa snažili aplikovať v tomto lokálnom spoločenstve.

2. Pôrodnosť

Na základe údajov získaných z historických matrík možno tvrdiť, že počas celého sledovaného obdobia sa hodnoty hrubej miery pôrodnosti (ďalej HMP) držali na stabilnej úrovni. Z dlhodobého hľadiska však sledujeme mierny pokles pôrodnosti, ktorý sa považuje za jednu zo zákonitostí demografického vývoja každej populácie. Dôkazy možno hľadať najmä v minulosti, kedy hospodárske a kultúrne pomery na Slovensku neboli jednoduché. Podľa E. Stodolu (1912) mali dobré roky priaznivý vplyv na počet pôrodov a naopak.

Vo Veľkom Grobe sa od 50. rokov 19. storočia až do konca sledovaného obdobia držala HMP evanjelického a.v. obyvateľstva na úrovni 30 až 40 ‰. Relatívna stabilita tohto ukazovateľa bola však sprevádzaná veľkými výkyvmi hodnôt z roka na rok. Maximálny počet narodených evanjelikov a.v. v obci bol 35 a minimálny 18. V Čataji možno pozorovať podobný vývoj HMP, ktorý bol v druhej polovici sledovaného obdobia sprevádzaný rýchlejším poklesom počtu pôrodov. Napriek tomu, že Čataj bol počtom obyvateľov menší, sociálne i ekonomicky zaostalejší ako Veľký Grob, dosahoval vyššie hodnoty HMP po celé obdobie výskumu. Maximálny počet narodených evanjelikov a.v. v obci Čataji bol 26 a minimálny 10. Príčinu rozdielneho vývoja ukazovateľa pôrodnosti možno hľadať i v odlišnom národnostnom zložení obyvateľov týchto dvoch obcí. Zatiaľ čo vo Veľkom Grobe žilo v tomto období viac ako 60% Maďarov, v Čataji tvorili obyvatelia tejto národnosti nepatrný podiel. Sám E. Stodola (1912, s. 65) vo svojej štúdii tvrdí, že *„v plodnosti prevyšujú Slováci Maďarov i iné národnosti v Uhorsku.“*





Graf č. 1 a 2: Vývoj hrubej miery pôrodnosti vo Veľkom Grobe (1851 – 1893) a Čataji (1853 -1893)

Zaujímavé je však porovnať hodnoty ukazovateľa HMP na základe konfesií. V oboch sledovaných obciach bola pôrodnosť katolíkov výrazne vyššia. Vo Veľkom Grobe sa pohybovala HMP v intervale od 40 po 50 ‰. Maximálny počet narodených katolíkov bol v sledovanom období 22 a minimálny 10. V Čataji pozorujeme zo začiatku obdobia podobný vývoj ukazovateľa HMP u oboch konfesií. Zmena nastáva v 60. rokoch 19. storočia, kedy hodnoty HMP rímskokatolíkov vystupujú do extrémnych polôh. V 80. rokoch 19. storočia hodnoty HMP klesli na 50 ‰ a v závere sledovaného obdobia ešte výraznejšie, a to na 30 ‰. Treba však spomenúť, že v jednej i druhej obci predstavovali rímskokatolíci len jednu tretinu všetkého obyvateľstva. Vyššiu pôrodnosť katolíkov možno spájať s dogmatizmom tejto viery, ktorá najmä v minulosti prísne zakazovala požívanie antikoncepčných prostriedkov i samotné potraty. (Botíková, 1997) Na základe týchto i ďalších poznatkov možno tvrdiť, že vo Veľkom Grobe i Čataji bol prítomný model mnohodetných katolíckych rodín, zatiaľ čo bohatší evanjelici a.v. v snahe nerozdrobovať majetok, preferovali menejdetné rodiny.

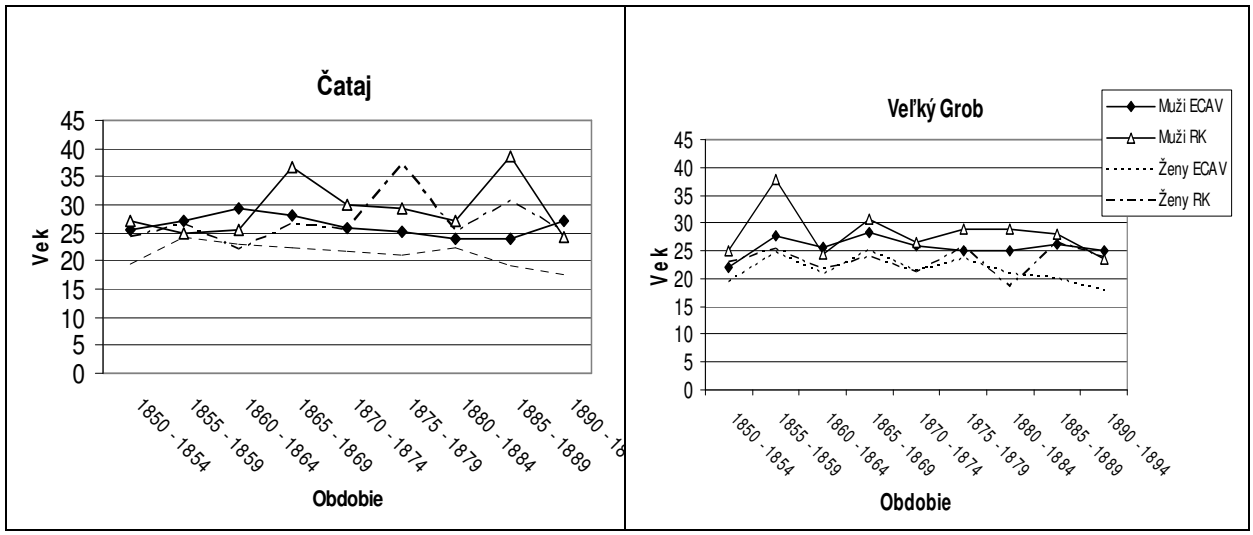
3. Sobášnosť

Pri sledovaní zmien v reprodukcii obyvateľstva sa dôraz kladie na štúdium plodnosti a úmrtnosti. Sobášnosť je skôr odrazom pomerov v socioekonomickej sfére. Avšak demografický význam získala vzhľadom na kultúrne tradície ľudstva, pretože najmä v minulosti sa plodnosť spájala s pojmom rodina. Vyššia úroveň mimomanželskej plodnosti bola výnimkou. Pri retrospektívnom pohľade na vývoj hrubej miery sobášnosti (ďalej HMS) je často ťažké určiť trend vývoja tohto ukazovateľa. Hoci je sobášnosť považovaná za veľmi stály ukazovateľ, pri výskume v mikromierke podlieha sezónnym výkyvom. V prípade menšej komunity totiž závisí od veľkosti vydaja schopného obyvateľstva i od celkovej sociálnej a ekonomickej klímy. (Fialová, 1985)

Vplyv spoločenských a kultúrnych tradícií, či hospodárskych pomerov sa bezprostredne prejavuje na veku, v ktorom muži i ženy vstupujú prvýkrát do manželstva. Často krát sa totiž nekryje s vekom, v ktorom snúbenci pohlavne dospievajú. V období výskumu boli ešte stále samozrejmosťou dohodnuté sobáše rodičmi.(Bagiová et al., 1994) Počas bádania sme došli k záverom, že manželstvá vo farnosti Veľký Grob – Čataj boli silno endogamné. V prípade rímskokatolíkov sme sa nestretli ani s jedným zmiešaným manželstvom, v prípade evanjelikov a.v. bol počet zmiešaných manželstiev minimálny. Všeobecne možno teda tvrdiť, že dominovala sociálna homogamia, ktorá bola znásobená geografickou homogamiou.

Priemerný vek, v ktorom vstupovali evanjelické a.v. ženy do manželstva sa pohyboval v prípade oboch obcí v intervale 20 až 25 rokov. Len v závere sledovaného obdobia ešte klesol pod hranicu 20 rokov, čo mohlo byť dôsledkom vydaja niekoľkých neplnoletých dievčat. Muži – evanjelici a.v. vstupovali do manželstva o čosi neskôr ako ženy, a to v priemere medzi 25 a 30 rokom života.

V prípade rímskokatolíkov nenastali výrazné odlišnosti. Priemerný vek čatajských žien pri sobáši bol okolo 25 rokov, čiže vyšší ako u evanjeličiek a.v. Túto hodnotu však mohli naopak výrazne ovplyvniť opakované sobáše ovdovelých žien. I priemerný vek mužov vstupujúcich do manželstva bol značne vyšší, a to 25 až 35 rokov. Vo Veľkom Grobe vstupovali ženy do manželstva medzi 20 a 25 rokom života a muži medzi 25 a 30 rokom.



Graf č. 3 a 4: Vplyv konfesie na priemerný vek snúbencov vo Veľkom Grobe a Čataji (1850 – 1894)

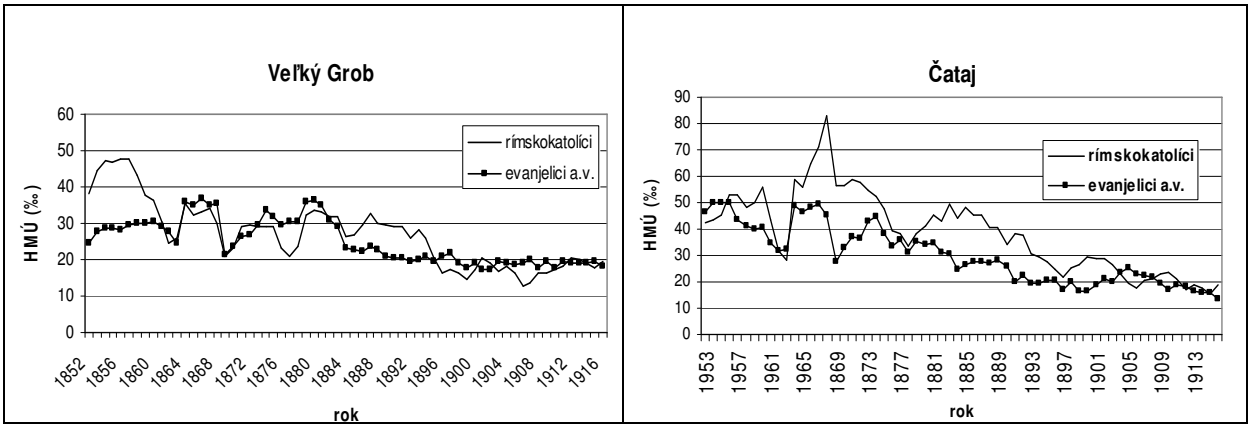
4. Úmrtnosť

Úmrtnosť nepatrí medzi stabilné ukazovatele, nakoľko odráža všetky nepriaznivé udalosti v obci. Pri spätnom vysvetľovaní intenzity úmrtnosti treba do úvahy brať aj intenzitu pôrodnosti, keďže ide o historické obdobie s charakteristickou vysokou detskou úmrtnosťou.

E. Stodola (1912) píše vo svojej štúdii *„tam, kde sa mnoho detí rodí, ktoré ako známo často mrú, tam je i bez iných činiteľov väčšia úmrtnosť.“ (s. 69)* K dojčenskej úmrtnosti často pristupujú aj ďalšie negatívne udalosti, ako chorobnosť populácie, epidémie, nevhodné sociálne a bytové podmienky, neúroda a pod., ktoré dotvárajú našu predstavu o životných podmienkach, ktoré panovali v obci.

Napriek tomu, že časové rady boli vyrovnané pomocou 5 – ročných kĺzavých priemerov, vidíme v jednej i druhej obci výraznú fluktuáciu hodnôt hrubej miery úmrtnosti (ďalej HMÚ). Presun hodnôt na stabilnú úroveň, a to 20 ‰ vidíme až v závere sledovaného obdobia. Vo Veľkom Grobe sa HMÚ evanjelikov a.v. udržiavala do polovice 80. rokov 19. storočia na úrovni 25 až 35 ‰, čo bolo spôsobené vysokou dojčenskou úmrtnosťou i možným prenosom cholerovej epidémie zo susednej obce Čataj. Až začiatkom 20. storočia sa hodnoty HMÚ ustálili vďaka poklesu dojčenskej úmrtnosti na úrovni 20 ‰ a pod túto úroveň neklesli ani počas prebiehajúcej prvej svetovej vojny. Maximálny počet úmrtí v obci bol v sledovanom období 63 a minimálny 7. Rímskokatolíci mali na začiatku sledovaného obdobia vyššiu úmrtnosť (40 až 50 ‰) ako evanjelici a.v., neskôr však so zlepšujúcimi sa zdravotnými podmienkami klesla a zhruba o jednu dekádu neskôr ako v prípade evanjelikov a.v., sa hodnoty HMÚ katolíkov ustálili tiež na hranici 20 ‰. Maximálny počet úmrtí katolíkov v priebehu jedného roka bol 27 a minimálny 3.

V dôsledku vyššej intenzity pôrodnosti v Čataji, tu viac ľudí aj umieralo. Maximálny počet úmrtí v priebehu jedného kalendárneho roka dosahoval v prípade evanjelikov hodnotu 57 a u katolíkov 28. Počiatočné obdobie je sprevádzané výstupom hodnôt HMÚ evanjelikov a.v. do extrémnych polôh, 40 až 50 ‰. V prípade rímskokatolíkov aj nad hranicu 60 ‰. Tento extrém bol následkom vypuknutia cholerovej epidémie v obci, bol ojedinelý a v histórii obce nemá obdobu. Od konca 70. rokov 19. storočia sledujeme intenzívnejší pokles úmrtnosti evanjelikov a.v., zatiaľ čo úmrtnosť katolíkov sa udržuje stále na vysokej úrovni, a to 40 ‰. V závere obdobia sa však hodnoty HMÚ evanjelikov a.v. i katolíkov ustálili na úrovni 20 ‰. Treba poznamenať, že z ekonomického i spoločenského hľadiska zastávali katolíci vo Veľkom Grobe i Čataji nižšie pozície, žili v chudobnejších podmienkach a na ich vysokej úmrtnosti sa často podpísali i niekoľkoročné neúrody. (Beňušková, 2004)



Graf č. 5 a 6: Vývoj hrubej miery úmrtnosti vo Veľkom Grobe (1852 – 1916) a Čataji (1853 - 1917)

5. Záver

Religiozita je popri etnicite azda najvýraznejší determinant formovania kultúry lokálneho spoločenstva, ba celých kultúrnych areálov. Viac či menej sa v rámci určitej geografickej zóny podieľa na spôsobe života lokálnych komunít. V sekularizovanom svete 20. storočia

náboženský život už síce nie je hybnou silou dejín ako to bolo v predchádzajúcich storočiach, avšak niektoré sféry spoločenských a populačných procesov i medziľudských vzťahov sú náboženstvom hlboko poznačené a ich pochopenie nie je možné mimo jeho rámca.

Cieľom tejto práce bolo interpretovať získané poznatky z historických matrík, priblížiť čitateľovi význam i metódy historicko – demografického výskumu, bez ktorého by bola analýza súčasných populačných štruktúr nemysliteľná.

6. Použitá literatúra

[1] BAGIOVÁ, M. et al. 1994, Čataj – vlastivedná monografia, Kežmarok, 205 s.

[2] BEŇUŠKOVÁ, Z. (2004): Religiozita a medzikonfesionálne vzťahy v lokálnom spoločenstve, Ústav etnológie SAV, Bratislava, s. 38 – 53

[3] BOTÍKOVÁ, M. et al. 1997, Tradície slovenskej rodiny, Veda, Bratislava, s. 148 – 160

[4] BRIDOVÁ, V. 2004, Religiózna štruktúra obyvateľstva Slovenska, Diplom.práca, 82 s.

[5] FIALOVÁ, L. 1985, Příspěvek k možnostem studia sňatečnosti v českých zemích za demografické revoluce, In: Historická demografie 9, Ústav československých a světových dějin ČSAV, Praha, s. 89-123

[6] MADŽOVÁ, I. 2008, Demografické správanie evanjelikov a.v. v mikromierke vybraných obcí Slovenska, Bakalárska práca, 64 s.

[7] MAUR, E. 1983, Základy historické demografie, Univerzita Karlova, Praha, 194 s.

[8] MLÁDEK, J. et al. (2006): Demografická analýza Slovenska, Univerzita Komenského v Bratislave, 221 s.

[9] SEBŐK, L. (zost.) 2005. Az 1869. évi népszámlálás vallási adatai. Budapest, TLA Teleki L. I.

[10] STODOLA, E. (1912): Štatistika Slovenska, Turčiansky Sv. Martin, 158 s.

7. Zdroje

[1] A magyar korona országaiban az 1881. év elején végrehajtott népszámlálás főbb eredménye megyék és községek szerint részletezve 2. zv. Budapest: Pesti Könyvnyomda – Részvény – Társaság, 1882. 326 s.

[2] A magyar korona országainak 1900. évi népszámlálása. 1. zv. Budapest: Pesti Könyvnyomda – Részvény – Társaság, 1902. 609 s.

[3] A magyar szent korona országainak 1910. évi népszámlálása. 42. zv. Budapest: Athenaeum, 1912. 880 s.

[4] A magyar szent korona országainak 1903, 1904, és 1905 népmozgalma. Budapest, Pesti Könyvnyomda-részvénytársaság. Magyar statistikai közlemények, 22 zväzok

[5] A magyar szent korona országainak 1906, 1907, és 1908 népmozgalma. Budapest, Pesti Könyvnyomda-részvénytársaság. Magyar statistikai közlemények, 32 zväzok

[6] A magyar szent korona országainak 1909, 1910, 1911 és 1912 népmozgalma. Budapest, Pesti Könyvnyomda-részvénytársaság. Magyar statistikai közlemények, 50 zväzok.

[7] Matrika narodených evanjelického a.v. farského úradu Veľký Grob – Čataj, Štátny archív v Bratislave, 1837 – 1888 a 1889 – 1895

[8] Matrika narodených rímskokatolíckeho zboru Veľký Grob – Čataj, Štátny archív v Bratislave, 1850 – 1895

[9] Matrika sobášených rímskokatolíckeho zboru Veľký Grob – Čataj, Štátny archív v Bratislave, 1850 – 1895

[10] Matrika sobášených evanjelického a.v. farského úradu Veľký Grob – Čataj, Štátny archív v Bratislave, 1787 – 1896

[11] Matrika zomretých rímskokatolíckeho zboru Veľký Grob – Čataj, Štátny archív v Bratislave, 1850 – 1895

[12] Matrika zomretých evanjelického a.v. farského úradu Veľký Grob – Čataj, Štátny archív v Bratislave, 1786 – 1895

[13] Matriky úmrtí, Obecný úrad vo Veľkom Grobe, 1895 – 1902, 1903 – 1906 a 1907 – 1922

[14] Matriky sobášov, Obecný úrad vo Veľkom Grobe, 1895 – 1902, 1903 – 1906 a 1907 – 1922

[15] Matriky narodených, Obecný úrad vo Veľkom Grobe, 1895 – 1902, 1903 – 1906 a 1907 – 1922

Adresa autora (–ov)

Ivana Madžová, Bc.
Prírodovedecká fakulta UK v Bratislave
ivicka12@gmail.com, ivanamadzova@yahoo.com

Porovnanie životnej úrovne obyvateľov Slovenskej a Českej republiky
Comparison of living standard of Slovak and the Czech citizens

Lucia Majerníková

Abstract: The aim of my thesis was a comparison of living standards of Slovak and Czech regions by using the methods of multidimensional comparison. This thesis is divided into two chapters. The first chapter describes the methods of multidimensional comparison. The second chapter compares the living standards of Slovak and the Czech regions on the basis of a set of indicators.

Key words: standard of living, Slovak republic, The Czech republic, regions, districts, multidimensional comparison

Kľúčové slová: životná úroveň, Slovenská republika, Česká republika, kraje, regióny, viacrozmerné porovnávanie

1. Úvod

Cieľom práce je komparácia životnej úrovne obyvateľov jednotlivých krajov Českej a Slovenskej republiky. Porovnávať tieto dve krajiny je zaujímavé z viacerých dôvodov. Česko a Slovensko tvorili dlhé obdobie spoločný štát, zdieľajú spoločnú minulosť a v oboch krajinách sa uskutočnila transformácia ekonomiky. Napriek rozdeleniu spoločného Československého štátu, tieto dve krajiny navzájom spolupracujú aj dnes a mnoho vecí ich spája (napr. spoločné úsilie o vstup do Európskej únie, ktoré sa uskutočnilo v roku 2004, Operačný program Slovenská republika – Česká republika 2007 – 2013, budovanie spoločnej česko-slovenskej bojovej skupiny EÚ).

Na porovnanie regiónov sme použili metódy viacrozmerného porovnávania. Výpočty sme realizovali prostredníctvom počítačových aplikácií Microsoft Office Excel 2007, Statgraphics Plus 5.1 a SAS Enterprise Guide. V prvej časti sa venujeme výberu ukazovateľov charakterizujúcich životnú úroveň, ich definovaniu a redukcii ich počtu.

V druhej časti sme na komparáciu životnej úrovne krajov Českej a Slovenskej republiky použili štyri metódy viacrozmerného porovnávania: metódu poradí, bodovaciu metódu, metódu normovanej premennej a metódu vzdialenosti od fiktívneho objektu.

2. Porovnanie životnej úrovne krajov SR a ČR

Výber premenných

Pri výbere ukazovateľov sme zohľadňovali ich vplyv na životnú úroveň a možnosť získať ich hodnoty pre sledovaný súbor krajov. Cieľom bolo uskutočniť porovnanie krajov na čo možno najaktuálnejších údajoch. V čase zisťovania boli najdostupnejšie údaje za rok 2006. Pre SR sme čerpali údaje z Regionálnej databázy Štatistického úradu SR a pre ČR z Verejnej databázy Českého štatistického úradu. Hodnoty niektorých ukazovateľov sú z roku 2005, z roku 2004 alebo staršieho, vzhľadom na nedostupnosť aktuálnejších údajov v čase ich získavania (november 2008).

Redukcia počtu premenných

Pôvodný súbor šestnástich premenných sme zredukovali najskôr o tie premenné, ktorých hodnota variačného koeficienta bola veľmi nízka ($V_j \leq 10\%$). Ďalej sme postupovali podľa Hellwigovej parametrickej metódy a vylúčili sme tie premenné, ktoré boli silne skorelované s ostatnými premennými, t.j. spĺňali nerovnosť: $r_{is} \geq 0{,}6$.

Určovanie typu premenných z hľadiska ich vplyvu na sledovaný jav

Pred aplikáciou jednotlivých metód sme určili charakter premenných, ktoré ostali po redukcii pôvodného súboru, t.j. či ide o stimulujúce alebo destimulujúce premenné.
Medzi stimulujúce sme zaradili:

– miera zamestnanosti (%),
– dokončené byty (počet),
– počet miest v domove dôchodcov na 1000 obyvateľov regiónu vo veku 65+ (počet),
– dĺžka ciest pripadajúca na rozlohu 100 km^2 (km),
– počet osobných vozidiel na 1000 obyvateľov regiónu.

Medzi destimulujúce sme zaradili tieto premenné:

– emisie oxidu uhoľnatého (t/km^2).

Aplikácia jednoduchých metód viacrozmerného porovnávania na vybraný súbor premenných

Na súbor vybraných premenných sme aplikovali jednotlivé metódy a kraje usporiadali podľa poradia. Aplikovali sme nasledujúce metódy: metódu poradí, bodovaciu metódu, metódu normovanej premennej, metódu vzdialenosti od fiktívneho objektu. Na všetky výpočty bola použitá počítačová aplikácia MS EXCEL 2007.

2.1.1 Metóda poradí

Kraje sme usporiadali podľa každého ukazovateľa a to tak, že objektu s najlepšou hodnotou ukazovateľa sme priradili poradie, rovnajúce sa počtu objektov v súbore, v našom prípade $m = 22$. Nasledujúcemu kraju sme priradili poradie $m - 1$ až po posledný kraj, ktorému sme priradili poradie 1. Ak malo viac objektov rovnakú hodnotu jednej premennej, priradili sme im rovnaké poradie, rovnajúce sa priemeru poradí.

Konečné poradie krajov sme určili na základe syntetickej premennej d_i, ktorú sme vypočítali ako aritmetický priemer všetkých poradí pre daný objekt. Kraj s najväčšou hodnotu syntetickej premennej sa umiestnil na prvom mieste.

Použitím metódy poradí sa na prvom mieste umiestnil Stredočeský kraj, ktorého hodnota syntetickej premennej bola najvyššia ($d_i = 17,5$). Na jeho umiestnenie mala vplyv hlavne premenná *dokončené byty* (5961 bytov), *dĺžka ciest pripadajúca na rozlohu 100 km^2* (87,13 km ciest), *počet osobných vozidiel na 1000 obyvateľov regiónu* (437 osobných vozidiel) a *miera zamestnanosti* (57,5%) . Na druhom mieste sa umiestnil Juhočeský kraj ($d_i = 17,0$) hlavne vplyvom hodnoty premennej *emisie oxidu uhoľnatého* (1,0 t/km^2), *počet miest v domove dôchodcov na 1000 obyvateľov regiónu vo veku 65+* (3,1 miesta) a *počet osobných vozidiel na 1000 obyvateľov regiónu* (428 osobných vozidiel). Tretie miesto obsadil kraj Vysočina ($d_i = 15,0$).

Z krajov SR sa umiestnili v prvej desiatke dva kraje, Trnavský kraj ($d_i = 14,176$)a Bratislavský kraj ($d_i = 13,176$).

Na posledných priečkach sa umiestnilo šesť slovenských krajov, z toho na dvadsiatom prvom až dvadsiatom druhom Banskobystrický kraj ($d_i = 4,167$) a Košický kraj ($d_i = 4,167$). Pozícia Košického kraja bola ovplyvnená vysokými hodnotami *emisie oxidu uhoľnatého* (15,45 t/km^2) a nízkymi hodnotami počtu *dokončených bytov* (799 bytov). U Banskobystrického kraja sa dá takéto umiestnenie pripísať hlavne nízkej hodnote *miery nezamestnanosti* (47,5 %).

2.1.2 Bodovacia metóda

Pri tejto metóde sme hodnoty ukazovateľov substituovali príslušnými bodmi. Pre každý ukazovateľ sme našli objekt s najlepšou hodnotou tohto ukazovateľa s prihliadnutím na charakter premennej a tomuto objektu sme priradili pre túto premennú 100 bodov. Hodnoty

premenných ostatných objektov sme dopočítali ako ich percentuálny podiel na najlepšej hodnote danej premennej. Integrálny ukazovateľ sme vypočítali znovu ako aritmetický priemer všetkých bodov pre daný objekt.

Prvenstvo si udržal Stredočeský kraj (d_i = 75,62) a to vďaka premenným *dokončené byty* (5961 bytov) a *dĺžka ciest pripadajúca na rozlohu 100 km*^{2} (87,13 km ciest), pri ktorých získal najväčší možný počet bodov. Zaujímavý je posun Juhomoravského kraja (d_i = 71,58) na druhé miesto oproti siedmemu miestu dosiahnutému pri predchádzajúcej metóde. O sedem miest sa zo štvrtej priečky na jedenástu oproti metóde poradia prepadol Plzenský kraj (d_i = 61,215). Z krajov SR si polepšil len Bratislavský kraj (d_i = 65,58), ktorý postúpil z deviateho na piate miesto vďaka najvyššej *miere zamestnanosti* (88,29%).

Posledné miesta, tak ako pri metóde poradí, znovu obsadilo šesť krajov SR .

2.1.3 Metóda normovanej premennej

Pri aplikácii tejto metódy sme najskôr hodnoty ukazovateľov previedli na normovaný tvar. Syntetickú premennú sme vypočítali ako aritmetický priemer normovaných hodnôt ukazovateľov.

Prvé miesto si znovu udržal Stredočeský kraj (d_i = 1,034). Do prvej trojky, na druhé miesto, sa dostal aj jeden kraj SR a to Bratislavský kraj (d_i = 0,710). Na takto „dobré“ umiestnenie mal najväčší vplyv ukazovateľ *miera zamestnanosti* (88,29%), v ktorom Bratislavský kraj dosahuje najvyššiu hodnotu a ukazovateľ *dokončené byty* (4307 bytov). Tretie miesto obsadil Ústecký kraj (d_i = 0,491) a to práve vďaka premennej *počet miest v domove dôchodcov na 1000 obyvateľov regiónu vo veku 65+* (3,98 miest).

O šesť miest oproti predchádzajúcej metóde sa prepadol Juhomoravský kraj (d_i = 0,344), ktorý klesol z druhého na ôsme miesto. Trnavský kraj (d_i = 0,013) sa prepadol o päť miest v porovnaní s výsledkami bodovacej metódy.

Na jedno z posledných miest sa zaradil aj jeden z krajov ČR a to Moravskosliezsky kraj (d_i = -0,557), ktorý obsadil devätnáste miesto. Tento kraj má najvyššiu hodnotu premennej *emisie oxidu uhoľnatého* (24,4 t/km^2). To zapríčinilo, že i napriek tomu, že má lepšie hodnoty ostatných piatych premenných ako Nitriansky kraj (d_i = -0,529), tak sa ocitol v poradí až za ním.

Tak ako pri predchádzajúcich metódach, na posledných priečkach sa umiestnilo znovu šesť krajov SR. Na posledné dvadsiate druhé miesto sa dostal Košický kraj (d_i = -1,143).

2.1.4 Metóda vzdialenosti od fiktívneho objektu

Pri použití metódy vzdialenosti od fiktívneho objektu sme pracovali s normovanými hodnotami premenných z predchádzajúcej metódy (metódy normovanej premennej). Vytvorili sme fiktívny objekt, ktorý má najlepšie možné hodnoty v každej premennej. Jednotlivé kraje sme potom porovnávali s týmto objektom.

Syntetická premenná vyjadruje euklidovskú vzdialenosť kraja od fiktívneho objektu. Kraj, ktorý bol k fiktívnemu objektu „najbližšie“, t.j. mal najmenšiu hodnotu syntetickej premennej, bol prvý. A naopak, kraj, ktorý bol mal najväčšiu hodnotu syntetickej premennej, bol „najďalej“ od fiktívneho objektu a preto bol posledný.

Prvé a druhé miesto zostalo nezmenené oproti predchádzajúcej metóde, t.j. Stredočeský kraj (d_i = 1,753) na prvom a Bratislavský kraj (d_i = 1,946) na druhom mieste.

Juhomoravský kraj (d_i = 2,229) si oproti metóde normovanej premennej polepšil a postúpil z ôsmeho na štvrté miesto a to práve vďaka premennej *dokončené byty* (3965 bytov). Naopak, Ústecký kraj (d_i = 2,368) sa v porovnaní s predošlou metódou prepadol z tretieho na ôsme miesto. Najväčší podiel má na tomto prepade premenná *dokončené byty* (1119 bytov) a *miera zamestnanosti* (51,4%).

Na posledných siedmych miestach sa znovu umiestnilo šesť krajov SR a jeden kraj ČR, tak ako pri metóde normovanej premennej. Prvý krát sa na posledné miesto nezaradil Košický kraj (d_i = 3,363), ale Prešovský kraj (d_i = 3,464).

Poradie krajov SR a ČR využitím všetkých štyroch metód

V nasledujúcej tabuľke uvádzame získané výsledky prostredníctvom štyroch metód viacrozmerného porovnávania.

Tabuľka 4: Poradie krajov SR a ČR využitím všetkých štyroch metód

Kraje	***Metóda poradí***	***Bodovacia metóda***	***Metóda normovanej premennej***	***Metóda vzdialenosti od fiktívneho objektu***
	Poradie	*Poradie*	*Poradie*	*Poradie*
Bratislavský	9	5	2	2
Trnavský	7 a 8	9	14	10
Trenčiansky	18	18	16	16
Nitriansky	17	17	18	18
Žilinský	19	20	17	17
Banskobystrický	21 a 22	21	20	20
Prešovský	20	19	21	22
Košický	21 a 22	22	22	21
Praha	12 a 13	15	9	12
Stredočeský	1	1	1	1
Juhočeský	2	3	4	3
Plzenský	4	11	11	13
Karlovarský	16	14	15	15
Ústecký	12 a 13	7	3	8
Liberecký	14	13	13	14
Královohradecký	6	10	5	6
Pardubický	5	8	7	7
Vysočina	3	4	6	5
Juhomoravský	7 a 8	2	8	4
Olomoucký	10	6	10	9
Zlínsky	11	12	12	11
Moravskosliezsky	15	16	19	19

Z výsledkov v tabuľke môžeme vidieť, že z celkového počtu dvadsaťdva porovnávaných krajov má len jediný kraj (Stredočeský kraj) rovnaké poradie podľa všetkých štyroch metód. U štyroch krajov sa líši poradie o jednu poradovú hodnotu (Nitriansky kraj, Banskobystrický kraj, Košický kraj, Liberecký kraj) a u troch krajov (Trenčiansky kraj, Juhočeský kraj, Karlovarský kraj) tvorí rozdiel dve poradia.

Dôvodom, prečo všetky štyri metódy nedávajú rovnaké výsledky, je spôsob, akým sa u jednotlivých metód, posudzuje variabilita ukazovateľov. Napr. metóda poradí vôbec nezohľadňuje variabilitu medzi hodnotami premennej. U bodovacej metódy sa posudzuje variabilita absolútne, avšak u metódy normovanej premennej sa posudzuje relatívne. Metóda vzdialenosti od fiktívneho objektu pracuje so štvorcami odchýlok, čo spôsobuje väčšiu citlivosť na zmeny hodnôt premenných.

3. Záver

Cieľom práce bolo porovnanie životnej úrovne krajov Českej a Slovenskej republiky. Keďže hovoríme o kategórii, na charakteristiku ktorej sa používa celý rad premenných, využili sme na riešenie tejto problematiky metódy viacrozmerného porovnávania.

Životnú úroveň sme charakterizovali prostredníctvom súboru šestnástich ukazovateľov. Výber premenných bol ovplyvnený dostupnosťou údajov pre jednotlivé kraje. Snahou bolo získať, čo najaktuálnejšie údaje pre čo najväčší počet krajov. Tými boli v čase zisťovania (november 2008) údaje za rok 2006, niektoré údaje sú z predchádzajúceho obdobia. Pôvodný súbor sme zredukovali na základe variability a korelácie na súbor šiestich premenných. Všetky vybrané premenné, ktoré boli použité v analýze, obsahovali údaje z roku 2006. Na zistené údaje sme aplikovali štyri metódy viacrozmerného porovnávania – metódu poradia, bodovaciu metódu, metódu normovanej premennej a metódu vzdialenosti od fiktívneho objektu.

Z porovnania výsledkov získaných jednotlivými metódami vyplýva, že jednoznačne najvyššiu životnú úroveň dosahujú obyvatelia Stredočeského kraja. Tento kraj sa umiestnil prvý pri každej zo štyroch metód. Z krajov SR sa najlepšie umiestnil Bratislavský kraj na treťom až štvrtom mieste a Trnavský kraj na jedenástom mieste. Ostatných šesť slovenských krajov sa umiestnilo vždy na posledných priečkach. Košický kraj sa pri troch metódach zaradil na posledné miesto. To potvrdzuje tvrdenie o bohatšom Západe a chudobnejšom Východe. Z krajov ČR sa najhoršie umiestnil Moravskosliezsky kraj a to v priemere na sedemnástom mieste. Zaujímavé je umiestnenie Prahy, ktorá bola v priemere na trinástom mieste. Tento kraj (spolu so Stredočeským krajom) sa považuje za kraj s najvyššou životnou úrovňou v ČR. Z nášho porovnávanie však vyplýva, že desať krajov Česka má vyššiu životnú úroveň ako Praha. Možno teda povedať, že výsledok analýzy životnej úrovne tesne súvisí s výberom ukazovateľov. Na základe iných ukazovateľov by sme pravdepodobne dostali aj iné výsledky.

4. Literatúra

[1] ČECHOVÁ, M. 2006. Česká republika a Slovensko – 12 let poté. Masarykova univerzita v Brně,

[2] FERŤALOVÁ, M. 2008. Čechom pomohol lepší štart, teraz ich dobiehame. In: Hospodárske noviny, 24.10. 2008.

[3] GABRIELOVÁ, H. – KLAS, A. - MIKELKA, E. - OKÁLI, I – SCHMÖGNEROVÁ, B. 1992. Ekonomika Slovenska na začiatku transformačného procesu. Bratislava: Ústav ekonomickej teórie Slovenskej akadémie vied, 1992.

[4] JANČÍK, Ľ. 2005. Životnú úroveň Čechov dostihneme o desať rokov. In: Hospodárske noviny, 8.4. 2005.

[5] KIŠŠ, Š. – ŠIŠKOVIČ, M. 2004. Porovnanie životnej úrovne na Slovensku v rokoch 1989 – 2005. Inštitút finančnej politiky Ministerstvo financí SR, 2004.

[6] KŘIVÁNKOVÁ, P. 2008. Výzkum pro řešení regionálních disparit. Ministerstvo pro místní rozvoj, 2008.

[7] PAŽITNÁ, M. – LABUDOVÁ, V. 2007. Metódy štatistického porovnávania. Bratislava: Ekonóm, 2007. s. 148 - 173.

[8] STANEK, V. 2004. Sociálna politika. EKONÓM, Bratislava, 2004.

[9] STANKOVIČOVÁ, I. – VOJTKOVÁ, M. 2007. Viacrozmerné štatistické metódy s aplikáciami. Bratislava: Iura Edition, 2007. s. 15 – 43.

[10] http://px-web.statistics.sk/PXWebSlovak/index.htm

[11] http://portal.statistics.sk

[12] http://www.czso.cz/

[13] http://www.euractiv.sk/regionalny-rozvoj/zoznam_liniek/regionalna-politika-v-sr

Adresa autora (-ov):

Lucia Majerníková, Bc.
1.ročník 2.stupňa štúdia, Fakulta hospodárskej informatiky
Ekonomická univerzita v Bratislave
luciamajernikova@gmail.com

Marketingový výskum produktu Kofola
Marketing research on Kofola

Michaela Nazadová

Abstract: This contribution utilizes the use of statistical methods in marketing research. The research was conducted by the means of a standardized questionnaire. The key point of the investigation was to examine customers' attitude towards the product Kofola; whether the customer prefers this product over other like products; how does the customer realize the price and quality ratio of the product and the customers' ability to recognize the slogan and the advertisement of the product. This contribution consists of three parts. The first section is dedicated to the methodology of the research conducted, the technique used to collect and analyze data and the size of the sample. The next section examines the results of the investigation and evaluates the hypotheses. In the last section, concrete suggestions for the company Kofola a.s. based on the outcome of the investigation are recommended.

Key words: marketing research, inquiry, standardized questionnaire, Kofola.

Kľúčové slová: marketingový výskum, dopytovanie, štandardizovaný dotazník, Kofola.

1. Úvod

Využitie štatistických metód vo výskumnej praxi rôznych odvetví a oblastí života je v dnešnej dobe samozrejmosťou. Nie je tomu inak ani v oblasti marketingového výskumu.

Produkt Kofola je na slovenskom trhu veľmi obľúbený medzi všetkými vekovými kategóriami obyvateľov. Dôvtipné reklamné kampane, prispôsobené ročnému obdobiu, prehľadná web stránka, kvalitná marketingová komunikácia a dlhoročná tradícia sú atribúty, ktoré na spotrebiteľa výrazne vplývajú. Potreby a želania spotrebiteľa sú základom pri tvorbe efektívnej marketingovej stratégie. Hlavnou úlohou marketingového výskumu je takéto informácie poskytnúť a pomôcť tým manažmentu poznať bližšie svojich zákazníkov a prispôsobovať vlastnosti produktu jednotlivým trhovým segmentom.

Cieľom uskutočneného výskumu je predložiť poznatky o postojoch zákazníka k produktu, či ho preferuje pred ostatnými, vnímaní ceny a kvality a jeho schopnosti rozpoznať slogan a konkrétnu reklamu. Zozbierané údaje budú analyzované a vyhodnotené pomocou štatistických metód a výsledky analýz budú formulované do záverov a odporúčaní.

2. Metodika výskumu

Predkladaný výskum má skúmať postoje, vnímanie a motívy spotrebiteľa, zvolená bola preto metóda osobného dopytovania s použitím štandardizovaného dotazníka, ktorý je formulovaný tak, aby pôsobil dynamicky. Prevažuje využitie dichotomických otázok a otázok s viacerými alternatívami odpovede, v prípade skúmania vnímania ceny, chute a reklamy bola na hodnotenie využitá päťstupňová škála. Dotazník sa skladá z 11 otázok s uzavretými odpoveďami, ktorých výhodou je správna interpretácia respondentmi a jednoduchosť spracovania a vyhodnocovania takto zozbieraných údajov.

Výskum bol uskutočnený na vzorke 80 respondentov v období 17. – 24. marca 2009 na území mesta Bratislava. Pomer zastúpenia pohlaví je muži 64% a ženy 36%. Najväčšmi zastúpená je veková kategória od 21 do 30 rokov so 77,5%. Samotnému výskumu predchádzal predprieskum na menšej vzorke opýtaných, aby sa overila správnosť porozumenia jednotlivých otázok zo strany respondentov. Zvolenú formu, osobné dopytovanie,

považujeme za veľmi vhodné, pretože dáva priestor na usmernenie respondenta a prípadné vysvetlenie vzniknutých nejasností. Zozbierané údaje boli spracované v programoch MS Excell a SAS Enterprise Guide 4. Po spracovaní, boli dáta pomocou štatistických metód analyzované, vyhodnotené a transformované do výstupov vo forme záverov a odporúčaní.

3. Zhrnutie výsledkov výskumu

Proces marketingového výskumu chápeme ako chronologickú činnosť pozostávajúcu z prípravy, realizácie a vyhodnotenia výsledkov výskumu. Aby sme vedeli vyhodnotiť výsledky, je potrebné stanoviť si ciele, ktoré súčasne definujú potrebu konkrétnych informácií. Stanovenie hypotéz na začiatku výskumu umožní sústrediť sa na problém, formulovať dotazník a podľa výsledkov ich vyhodnotiť. Stanovili sme si nasledovné tri hypotézy:

H1: Viac ako polovica respondentov by produkt Kofola uprednostnila pred ostatnými.

H2: 80% žien privítalo rozšírenie produktového radu Kofola o Kofolu light.

H3: Viac ako 30% respondentov považuje reklamu na nápoj Kofola za výbornú.

Úvodná otázka predkladaného dotazníka je filtračná a má za úlohu preveriť či respondent produkt pozná, slúži teda na selekciu vhodných a nevhodných respondentov. Z osemdesiatich opýtaných na túto otázku odpovedali všetci kladne, môžeme teda konštatovať, že produkt Kofola ľudia poznajú.

Tabuľka 5: Znalosť produktu Kofola

Poznáte produkt Kofola?		
Odpoveď	**Počet odpovedí**	**Počet Percent**
Áno	80	100.00

Druhá otázka sa zameriava na pozíciu nápoja Kofola medzi ostatnými najčastejšie ponúkanými kolovými nápojmi. 45% respondentov uviedlo, že dáva prednosť Kofole pred ostatnými kolovými nápojmi. Za ňou nasleduje Coca cola, ktorú uviedlo 33.75%, 6.25% uviedlo nápoj Pepsi a 15% dáva prednosť inému kolovému nápoju ako predchádzajúcim uvedeným.

Tabuľka 6: Preferencia druhu kolového nápoja

Druh kolového nápoja		
Odpoveď	**Počet odpovedí**	**Počet Percent**
Coca cola	27	33.75
Pepsi	5	6.25
Kofola	36	45.00
Iný	12	15.00

Z uvedeného vyplýva, že väčšina respondentov preferuje nápoj Kofola pred ostatnými, skúmali sme teda ďalej ako vplýva pohlavie na preferenciu Kofoly. Pre zjednodušenie sme transformovali tabuľku, v ktorej sme rozdelili nápoje na Kofolu a iné, pričom pod iné sme zaradili všetky odpovede Coca cola, Pepsi a iný kolový nápoj.

Tabuľka 7: Preferencia Kofoly podľa pohlavia

Preferencia Kofoly podľa pohlavia			
Druh nápoja	**Pohlavie**		**Spolu**
	Žena	**Muž**	
Iné	17	27	44
Kofola	12	24	36
Spolu	29	51	80

Pomocou chí-kvadrát testu sme skúmali závislosť preferencie kofoly od pohlavia. Keďže P hodnota je rovná 0,6235 a je väčšia ako 0,05, prijímame teda nulovú hypotézu a môžeme potvrdiť, že preferencia Kofoly nezávisí od pohlavia respondenta.

V nasledujúcej otázke sme skúmali vnímanie nápoja z hľadiska *ceny, chute a reklamy*, pričom respondentom bola ponúknutá jednoduchá päťstupňová hodnotiaca škála, na ktorej mohli tieto atribúty nezávisle od seba ohodnotiť.

Cena bola najčastejšie hodnotená známkou 3, urobilo tak 42,5% opýtaných, 35% ju hodnotilo známkou 2 a 17,5% známkou 1. Iba 5% respondentov uviedlo známku horšiu ako 3, môžeme teda konštatovať, že cenu Kofoly vnímajú respondenti ako dobrú až veľmi dobrú.

Tabuľka 8: Hodnotenie ceny

Hodnotenie - cena		
Známka	**Počet odpovedí**	**Počet Percent**
1	14	17.50
2	28	35.00
3	34	42.50
4	3	3.75
5	1	1.25

Chuť nápoja bola hodnotená najčastejšie ako výborná, teda známkou 1, a urobilo tak 43,75% opýtaných, ako veľmi dobrú ju hodnotilo 30% a dobrú 21,25 opýtaných. Iba 5% respondentov uviedlo známku horšiu ako 3, môžeme teda konštatovať, že chuť nápoja Kofola vnímajú respondenti ako výbornú alebo veľmi dobrú.

Tabuľka 9: Hodnotenie chuti

Hodnotenie – chuť		
Známka	**Počet odpovedí**	**Počet Percent**
1	35	43.75
2	24	30.00
3	17	21.25
4	3	3.75
5	1	1.25

Reklama bola takmer polovicou respondentov hodnotená ako výborná, túto možnosť uviedlo 47,5%, ako veľmi dobrú ju hodnotilo 26,25%, známku 3 by jej udelilo 17,5%. Iba 8,75% respondentov by udelilo reklame známku horšiu ako 3, môžeme teda konštatovať, že aj reklama nápoja Kofola je vnímaná ako výborná, čí veľmi dobrá.

Tabuľka 10: Hodnotenie reklamy

Hodnotenie – reklama		
Známka	**Počet odpovedí**	**Počet Percent**
1	38	47.50
2	21	26.25
3	14	17.50
4	4	5.00
5	3	3.75

Na testovanie hodnotenia premenných v škálových otázkach sme použili Wilcoxonov test. Škála hodnotenia bola postavená na princípe známkovania od 1do 5, tak ako v škole. Respondenti boli vyzvaní, aby hodnotili tri atribúty: cenu, chuť a reklamu.

V prvom prípade sme testovali hypotézu, že cena a chuť boli hodnotené rovnako. Pomocou Wilcoxonovho testu sme zistili, že P hodnota je 0,003 a to je menej ako 0,05, preto zamietame nulovú hypotézu a prijímame alternatívnu hypotézu, že cena a chuť neboli hodnotené rovnako a súčasne vieme povedať, že chuť bola hodnotená lepšie.

V druhom prípade sme testovali kategórie chuť a reklama a porovnávali ich hodnotenie. Ako Wilcoxonov test dokázal, chuť a reklama boli na testovanej hladine významnosti hodnotené rovnako, keďže vypočítaná P hodnota je väčšia ako 0,05.

Z počtu 80 opýtaných respondentov je 51 mužov a 29 žien. Pomocou krížovej tabuľky máme možnosť skúmať aký druh Kofoly by uprednostnili muži a aký ženy. Z 29 žien by si Kofolu Orignal vybralo 21, čo je takmer 73%, 17,24% by si vybralo Kofolu Citrus a 10,34% nerozlišuje, ktorý uprednostňuje. Zaujímavé je, že ani jedna zo žien neodpovedala, že by preferovala Kofolu Light. Muži takisto preferujú Kofolu Original – 82,35% pred Kofolou Citrus – 9,8%, medzi jednotlivými druhmi Kofoly nerozlišuje 5,88% a jeden muž uprednostňuje Kofolu Light.

Tabuľka 11: Preferencia druhu Kofoly v závislosti od pohlavia

Pohlavie vz. Druh Kofoly					
Pohlavie	**Druh Kofoly**				**Spolu**
	Original	**Citrus**	**Light**	**Nerozlišujem**	
Žena	21	5	0	3	**29**
	72.41	17.24	0.00	10.34	
Muž	42	5	1	3	**51**
	82.35	9.80	1.96	5.88	
Spolu	**63**	**10**	**1**	**6**	**80**

4. Vyhodnotenie hypotéz

H1: Viac ako polovica respondentov by produkt Kofola uprednostnila pred ostatnými.

Väčšina opýtaných dáva prednosť nápoju Kofola pred ostatnými kolovými nápojmi, preferuje ju 45% opýtaných, pričom 33,75% by uprednostnilo Coca colu.

H2: 80% žien privítalo rozšírenie produktového radu Kofola o Kofolu light.

Táto hypotéza bola postavená na teórii, že ženy uprednostňujú light výrobky. Absolútne sa však nepotvrdila, keďže žiadna žena v dotazníku neuviedla, že spomedzi produktového radu Kofola preferuje Kofolu Light.

H3: Viac ako 30% respondentov považuje reklamu na nápoj Kofola za výbornú.

Súčasťou dotazníka boli aj hodnotiace škály, kde sa respondenti mohli vyjadriť ako vnímajú reklamu. Pri známkovaní od 1 do 5, tak ako v škole, 47,5% hodnotilo reklamu Kofoly ako výbornú, môžeme teda konštatovať, že táto hypotéza sa potvrdila.

5. Záver

Medzi oslovenými tvorili najväčšiu skupinu mladí ľudia vo veku od 21 do 30 až 77,5%. 50% tejto skupiny respondentov by uprednostnilo Kofolu pred inými kolovými nápojmi a súčasne táto kategória hodnotila cenu produktu lepšie ako respondenti iných vekových kategórií. Z výskumu ďalej vyplýva, že spotrebitelia hodnotia chuť Kofoly na výbornú až veľmi dobrú, zhodlo sa na tom 74% a obľubujú chuť najmä klasickej Kofoly Original, uviedlo tak 79% opýtaných.

V oblasti pôsobenia na jednotlivé pohlavia alebo vekové skupiny, z hľadiska frekvencie konzumácie a preferencie produktu, neboli výskumom dokázané žiadne závislosti, preto v tejto oblasti nevidíme dôvod robiť cielenú kampaň a zamerať sa na určitý segment. Priestor na zmenu vidíme vo vnímaní ceny, ktorá je u oboch pohlaví závislá od veku. Zatiaľ čo mladí spotrebitelia vo veku do 30 rokov hodnotili cenu v priemere ako veľmi dobrú, staršia generácia ju vidí ako dobrú. Toto vnímanie ceny skresľujú rôzne faktory, ktoré sme v našom výskume nemali možnosť detailne skúmať, a však čo vnímanie ceny výrazne ovplyvňuje sú cenové akcie.

6. Literatúra

[1] MESÁROŠOVÁ, M., MESÁROŠ, F. Marketingový výskum. Bratislava: EKONÓM, 2002, s. 248. ISBN 80-225-1606-6

[2] RICHTEROVÁ, K. a kol. Kapitoly z marketingového výskumu. Bratislava: EKONÓM, 2000, s.244. ISBN 80-225-1312-1

Adresa autora

Michaela Nazadová, Bc.,
2. ročník magisterského štúdia
Fakulta managementu UK v Bratislave
misel.nazadova@gmail.com

Modelovanie rizika monetárnej chudoby obyvateľstva SR
Risk modelling of monetary poverty of inhabitants in SR

Lukáš Pastorek

Abstract: By using the method of logistic regression, the author is modelling the risk of monetary poverty within the population of Slovak republic. The households are scored according to the level of poverty threshold, in our case according to the level - 60% of the equivalent disposable income. The paper is analyzing several selected variables and their influence on the target variable, valuating suitability of the model from the point of various statistical test criteria.

Key words: multidimensional methods, logistic regression, poverty indicators, poverty threshold, EU SILC, monetary deprivation, equivalent disposable income, risk of poverty rate

Kľúčové slová: viacrozmerné metódy, logistická regresia, indikátory chudoby, hranica chudoby, EU SILC, monetárna deprivácia, ekvivalentný disponibilný príjem, miera rizika chudoby

1. Úvod

So vstupom Slovenska do Európskej únie, sme sa užšie začlenili do skupiny krajín intenzívne bojujúcich s hmotnou i finančnou núdzou domáceho obyvateľstva a jej následkami. Implementujúc taktiku a smernice Európskej komisie o meraní indikátorov materiálnej a monetárnej deprivácie je možné dosiahnuť medzinárodné porovnanie týchto veličín na vyššej kvalitatívnej úrovni.

Pod taktovkou nariadení Eurostatu štatistické úrady každoročne realizujú národné moduly spoločného celoeurópskeho výberového zisťovania - European Union - Statistics on Income and Living Conditions (EU SILC). EU SILC je výberové zisťovanie získavané a spracovávané na ročnej báze, ktorého účelom je zozbierať informácie o stave a prerozdelení monetárnych zdrojov jednotlivých typov domácností, finančnej náročnosti a kvalite bývania, pracovných a zdravotných podmienkach, o charaktere a štruktúre chudoby, ako aj o indikátoroch sociálnej exklúzie v danej lokalite.

V tomto príspevku sa sústredíme na viacrozmerné modelovanie rizika chudoby v Slovenskej republike v roku 2007 s využitím logistickej regresie. Použili sme údaje EU SILC 2007 SR, kde reprezentatívnu vzorku predstavuje 4941 slovenských domácností. Výpočty sme uskutočnili v prostredí softvéru SAS® Enterprise Guide 4.1 a SAS® Enterprise Miner 5.3.

2. Miera rizika chudoby

Najdôležitejším ukazovateľom, ktorý sa sleduje a prezentuje verejnosti, je podiel chudobou ohrozeného obyvateľstva v sledovanej krajine - *miera rizika chudoby* (at-risk-of-poverty rate). Táto veličina udáva podiel osôb v krajine s ekvivalentným disponibilným príjmom nižším ako je 60% mediánu ekvivalentného disponibilného príjmu v národnom hospodárstve. Inými slovami, za monetárne chudobnú domácnosť sa považuje tá domácnosť, v ktorej je ekvivalentný disponibilný príjem pod hranicou rizika chudoby.

V Slovenskej republike bol v roku 2007 podiel osôb pod hranicou rizika monetárnej chudoby na úrovni 11% z celkového počtu obyvateľstva (10% muži, 11% ženy). Hranica monetárnej chudoby (t.j. 60% mediánu) bola stanovená na 88 727 Sk na rok, resp. 7 394 Sk

na mesiac na osobu domácnosti. Tieto čísla boli vypočítané na základe údajov EU SILC 2007 a publikované Štatistickým úradom SR.

3. Výpočet ekvivalentného disponibilného príjmu obyvateľstva

Postup výpočtu hranice chudoby z dát EU SILC 2007 SR bol nasledovný:

- Výpočet ekvivalentného disponibilného príjmu na osobu, ktorý sa priradí každému členovi domácnosti a z ktorého sa potom počíta medián je nasledovný:

Ekvivalentný disponibilný príjem *(HX100)* **=** (***Celkový disponibilný príjem domácnosti*** *(HY020)* * ***Inflačný faktor v rámci neodpovedajúcej domácnosti*** (*HY025))* **/** ***Ekvivalentná veľkosť domácnosti*** *(HX050), čiže: (HX100) = (HY020 * HY025) / (HX050)*

- Premenná *Ekvivalentná veľkosť domácnosti (HX050)* vyjadruje veľkosť domácnosti na základe modifikovanej OECD škály (váha 1 pre prvého dospelého, váha 0.5 pre každého ďalšieho dospelého člena a váha 0.3 pre každé dieťa mladšie ako 14 rokov).

- *Medián Ekvivalentného disponibilného príjmu* domácností z výberového zisťovania za rok *2007* sa vypočíta ako medián (vážený) z ekvivalentného disponibilného príjmu osoby (*HX100)*, ktorá sa nachádza v strede celého súboru osôb preváženého jednotlivými osobnými prierezovými váhami *(RB050)*. Jeho hodnota je 147 879 Sk.

- Na základe porovnania hodnoty ekvivalentného disponibilného príjmu jednotlivej domácnosti *(HX100)* s hranicou chudoby, t.j. z 60% mediánového ekvivalentného príjmu za rok 2007 (88727 Sk), bola osobám, resp. aj domácnostiam, v novovytvorenej binárnej premennej *RISK* priradená hodnota 1 (*je chudobná)* a ostatným hodnota 0 (*nie je chudobná)*.

4. Logistická regresia

Zo súborov EU SILC 2007 SR (H, D, R, P) sme vyselektovali len tie premenné, ktoré vykazovali vecnú a štatisticky významnú koreláciu na modelovanú (target) premennú RISK. Väčšinu premenných sme pre potreby logistického modelu museli upraviť (preškálovať, zoskupiť alebo prekódovať) s využitím rôznych metód v systéme SAS Enterprise Miner (uzlov *Transform variables* a *Interactiv binning*) a označili sme ich ako „*UPR*“. V tabuľke 1. sú premenné rozdelené do 2 skupín. Prvá skupina predstavuje premenné, ktoré sa vzťahujú k domácnostiam a druhá k osobám ich prednostov. V tabuľke sa vyskytuje 1 frekvenčná premenná (váhy), 2 kvantitatívne a 16 kategoriálnych premenných.

Tabuľka 12: Vysvetľujúce premenné

Premenná	Popis
RB050	Osobná prierezová váha (frekvenčná premenná)
HH020_UPR	Vlastnícky status obydlia - upravená
HS040	Schopnosť dovoliť si zaplatenie raz ročne jedného týždňa dovolenky mimo domu
HS050	Schopnosť dovoliť si mäsité jedlo (vegetariánsky ekvivalent) každý druhý deň
HS060	Schopnosť čeliť neočakávaným finančným výdavkom
HS120_UPR	Schopnosť vystačiť s peniazmi - upravená
HS140	Celkové náklady na bývanie ako finančná záťaž pre domácnosť
HX050	Ekvivalentná veľkosť domácnosti
HX070	Počet členov domácnosti
HX090_UPR	Počet závislých detí - upravená
HX020_UPR	Typ domácnosti - upravená
DB100	Stupeň urbanizácie
PB150	Pohlavie
PB190_UPR	Rodinný stav - upravená
PE040_UPR	Najvyššia dosiahnutá úroveň vzdelania - upravená
PH010_UPR	Všeobecné zdravie - upravená
PH020	Trpí rôznymi chronickými chorobami alebo stavmi
PL030_UPR	Aktuálny samo-definovaný status ekonomickej aktivity - upravená
PX010_UPR	Vek na konci príjmového referenčného obdobia - upravená

V modeli logistickej regresie *modelujeme podmienenú pravdepodobnosť, že domácnosť je chudobná* (t.j. jej disponibilný príjem je pod hranicou chudoby), čiže premenná RISK nadobúda hodnotu 1.

Pomocou selekčnej metódy *stepwise* bolo do modelu zaradených všetkých 18 nami zvolených premenných, t.j. všetky sú významné na hladine významnosti α=0,05. Hodnota Waldej chí-kvadrát štatistiky v tabuľke 2 nám odkryla aj pôsobenie najvýznamnejších premenných, ktoré predikujú výskyt chudobných domácností. Ako najsilnejší faktor sa ukázala premenná **PL030_UPR -** *Aktuálny samo-definovaný status ekonomickej aktivity*. Potom nasledujú dve premenné: **PB190_UPR -** *Rodinný stav* a **PE040_UPR -** *Najvyššia dosiahnutá úroveň vzdelania***.** Medzi najsilnejšie faktory sa zaradila aj premenná **HX020_UPR -** *Typ domácnosti***.**

Na základe štatistických testov môžeme konštatovať, že logistický model je ako celok významný (nízke *p*-hodnoty pre *Likelihood Ratio, Score* a *Waldov* test). Kvalitu modelu vyjadrujú miery asociácie. Percento zhodných párov je vysoké, dosahuje hodnotu **86%** (Percent Concordant)**.** Hodnota ROC indexu je **0,863.**

Tabuľka 13: Premenné vybrané do modelu logistickej regresie metódou stepwise

Analysis of Effects			
Effect	**DF**	**Wald Chi-Square**	**Pr > ChiSq**
PL030_UPR	5	**60469.3**	<.0001
PB190_UPR	4	**19463.9**	<.0001
PE040_UPR	2	**19159.0**	<.0001
HX020_UPR	4	**14703.5**	<.0001
HX090_UPR	3	9508.2	<.0001
HS140	2	8858.3	<.0001
PX010_UPR	3	8573.4	<.0001
DB100	2	7644.0	<.0001
HS060	1	6565.1	<.0001
HS050	1	6393.6	<.0001
HS120_UPR	1	3739.5	<.0001
PH010_UPR	2	3076.7	<.0001
HS040	1	2681.5	<.0001
HX050	1	2091.3	<.0001
HX070	1	2005.8	<.0001
HH020_UPR	3	1257.7	<.0001
PB150	1	73.0	<.0001
PH020	1	65.0	<.0001

5. Interpretácia výsledkov modelu logistickej regresie

Na interpretáciu logistického modelu použijeme pomery šancí (ods ratio), ktoré uvádzame v tabuľke 3.

Najvyššie šance prepadu pod hranicu chudoby sa objavujú v premennej PL030_UPR – *Aktuálny samo-definovaný status ekonomickej aktivity.* Všetky parametre sú porovnávané s kategóriou *„práca na úväzok" (6)*. Najvyššie šance domácnosti stať sa chudobnou (10,12:1), sa objavujú pri porovnaní prednostu, ktorý je *„nezamestnaný" (2)*, nasledujúc *„osoby v domácnosti a iné neaktívne osoby" (1)* (4,38:1). Šance parametrov *„žiak, študent" (3)*, *„dôchodca" (4)*, *„trvalá invalidita alebo nespôsobilosť" (5)* sa pohybujú medzi 2:1 až 3:1.

Rodinný stav - PB190 ukázal vysokú rozdielnosť v šanciach prednostov a ich domácností stať sa chudobnou. Najpriepastnejší rozdiel sa ukazuje medzi porovnaním *„slobodný" (1)* k *„ženatý / vydatá" (5)*. V prípade, že je prednosta domácnosti stále slobodný, jeho šanca stať sa chudobným vzrastie 3,19-násobne v porovnaní s manželským párom. Rovnako vyššiu šancu vykazuje aj domácnosť, na ktorej čele stojí *„rozvedený" (2)* alebo *„žijúci oddelene" (3)*. Na prvý pohľad trošku prekvapujúco pôsobí fakt, že v prípade *„vdova / vdovec" (4)* je táto šanca nižšia (0,9:1), nie však nejako radikálne. Ale keďže veľkú časť týchto obyvateľov predstavujú práve dôchodci, keďže priemerný vek vdov/vdovcov v našom

zisťovaní sa pohybuje v intervale 57-77 rokov a ktorí podľa viacerých správ NBS nepredstavujú najrizikovejšie skupiny, na rozdiel od mladých rodín s deťmi, je tento fakt čiastočne pochopiteľný.

PE040_UPR **-** *Najvyššia dosiahnutá úroveň vzdelania [primárne (1), sekundárne (2), terciárne (3)]* sa ukazuje, že bez ohľadu, či máte vyštudovanú strednú školu alebo len základnú školu, šanca prepadu do chudobou ohrozeného obyvateľstva je približne rovnaká v porovnaní s vysokoškolským vzdelaním, a to 3 krát vyššia.

Pre kategoriálnu premennú HX020_UPR – *Typ domácnosti* boli potvrdené predpokladané vzťahy medzi domácnosťami. *„Domácnosti bez závislých detí“ (1)* sa ukázali byť najmenej ohrozené (0,29:1) v protiklade s premennou *„domácnosť s jedným rodičom a závislými deťmi“ (2),* ktoré sú ohrozené najviac (1,28:1). Všetky parametre boli porovnávané s kategóriou *„jednočlenná domácnosť“(5). „Domácnosť 2 dospelých so závislými deťmi“ (3)* a *„Ostatné domácnosti so závislými deťmi“ (4)* sa prejavili ako menej rizikové vzhľadom na jednočlennú domácnosť (0,85:1; 0,45:1).

Rast *počtu závislých detí* - HX090_UPR tiež ovplyvňuje riziko monetárnej deprivácie. Šanca rodín s „*nula alebo jedno dieťa“ (1)* v porovnaní s rodinami „*štyri a viac detí“ (4)* je mizivá (0,09:1). Táto šanca sa však s pribúdajúcim počtom detí „*dve deti“ (2),* „*tri deti“ (3)* zvyšuje.

Premenná HS140 - *Celkové náklady na bývanie ako finančná záťaž pre domácnosť* nadobúdala hodnoty „*veľmi zaťažuje“ (1)*, „*trochu zaťažuje“ (2)* a „*vôbec nezaťažuje“ (3).* Model paradoxne ukazuje, že ľudia, ktorý subjektívne ohodnotili náklady na bývanie ako zaťažujúce majú väčšiu pravdepodobnosť sa v reálnom živote vyhnúť prepadnutiu do chudobou ohrozeného obyvateľstva (0,67:1; 0,41:1).

PB150 - *pohlavie* [*muž (1)*, *žena (2)*] ukazuje, že šanca je z rodového hľadiska približne rovnaká, s trochu vyšším rizikom u mužov (1;06:1).

Tabuľka 3: Hodnoty odhadov pomerov šancí a ich intervaly spoľahlivosti

Odds Ratio Estimates									
Effect		Point Estimate	95% Wald Confidence Limits		Effect		Point Estimate	95% Wald Confidence Limits	
HX050		**3.76**	3.56	3.98	**PB190_UPR**	4 vs **5**	**0.9**	0.88	0.92
HX070		**0.53**	0.51	0.54	**PE040_UPR**	1 vs **3**	**3.08**	2.97	3.19
HH020_UPR	1 vs **4**	**0.45**	0.43	0.48		2 vs **3**	**2.93**	2.88	2.98
	2 vs **4**	**1.08**	1.07	1.10	**PH010_UPR**	1 vs **3**	**0.83**	0.81	0.84
	3 vs **4**	**1.48**	1.38	1.57		2 vs **3**	**0.65**	0.64	0.66
HS040	1 vs **2**	**0.65**	0.64	0.66	**HX090_UPR**	1 vs **4**	**0.09**	0.09	0.09
HS050	1 vs **2**	**0.61**	0.60	0.61		2 vs **4**	**0.23**	0.22	0.24
HS060	1 vs **2**	**0.56**	0.56	0.57		3 vs **4**	**0.36**	0.35	0.38
HS120_UPR	0 vs **1**	**0.45**	0.44	0.46	**PL030_UPR**	1 vs **6**	**4.38**	4.27	4.50
HS140	1 vs **3**	**0.67**	0.65	0.69		2 vs **6**	**10.12**	9.93	10.31
	2 vs **3**	**0.41**	0.40	0.42		3 vs **6**	**2.92**	2.73	3.14
HX020_UPR	1 vs **5**	**0.29**	0.28	0.30		4 vs **6**	**2.54**	2.48	2.61
	2 vs **5**	**1.28**	1.24	1.32		5 vs **6**	**2.1**	2.03	2.18
	3 vs **5**	**0.85**	0.83	0.88	**PH020**	1 vs **2**	**1.06**	1.05	1.08
	4 vs **5**	**0.45**	0.43	0.47	**PX010_UPR**	1 vs **4**	**2.37**	2.29	2.46
PB150	1 vs **2**	**1.06**	1.05	1.08		2 vs **4**	**1.7**	1.64	1.75
PB190_UPR	1 vs **5**	**3.19**	3.11	3.27		3 vs **4**	**0.64**	0.62	0.65
	2 vs **5**	**2.64**	2.59	2.70	**DB100**	1 vs **3**	**0.47**	0.46	0.48
	3 vs **5**	**2.64**	2.50	2.78		2 vs **3**	**0.76**	0.75	0.77

V premennej HH020 sú všetky parametre: „*ubytovanie je poskytované zadarmo“ (1)*, „*nájomník alebo podnájomník platiaci bežné nájomné alebo trhovú cenu“ (2), „ubytovanie prenajímané za zníženú cenu“ (3*) porovnávané s hodnotou „*majiteľ / vlastník“ (4).* Ako vidíme v tabuľke 3, najvyššia šanca, a teda najviac riziková skupina, bola priradená parametru *(3)* so šancou cca. 1,5 vyššou ako je pri *majiteľovi*. Naopak najmenej rizikový sú tí, ktorí bývajú bezplatne (0,45:1).

V binárnych premenných HS040, HS050, HS060 bola hodnota „*áno“ (1)* porovnávaná s „*nie“ (2).* HS120_UPR - *Schopnosť vystačiť s peniazmi* bola podobne porovnávaná hodnota „*bez ťažkostí“ (0)* s hodnotou „*s ťažkosťami“ (1)* (0,45:1).

Z pohľadu vplyvu zdravia na chudobu PH010_UPR **-** *Všeobecné zdravie*, tí respondenti, ktorí označili svoj zdravotný stav ako „*dobrý“ (1),* čelia menšiemu riziku chudoby (0,83:1) vzhľadom na tých, ktorí označili svoj zdravotný stav ako „*zlý“ (3)*. Relatívne paradoxne vyznieva fakt, že osoby s ohodnotením „*priemerný zdravotný stav“ (2)* sú pred chudobou chránené viac, ako prednostovia s dobrým fyzickým stavom (0,65:1).

V ďalšej kategoriálnej premennej venujúcej sa zdravotnému stavu PH020 - *Trpí rôznymi chronickými chorobami alebo stavmi,* sa porovnáva hodnota kladného „*áno“ (1)* s hodnotou „*nie“ (2)* (1,06:1).

PX010 - *Vek na konci príjmového referenčného obdobia,* poukazuje práve na paradox verejnej mienky o tom, ktoré segmenty obyvateľstva a domácností sú najviac ohrozené chudobou. Výsledky ukazujú na fakt, že rodiny s mladými prednostami „*osoby do 37 rokov“ (1)*, „*osoby 37-57 rokov“ (2)* čelia ďaleko vyššiemu riziku (2,37:1; 1,37:1) ako osoby starších generácií „*osoby medzi 57- 77“ (3)*, „*osoby medzi 77- 97“ (4).* Fakt je spôsobený počtom závislých detí, medzi ktoré musia domácnosti svoj príjem deliť, na rozdiel od dôchodcov. Šanca mladej domácnosti je viac ako 2-násobne vyššia ako rodiny, kde má prednosta medzi 77-97 rokov, kde je vysoký predpoklad, že domácnosť tvorí 1 alebo max. 2 osoby.

DB100 - *Stupeň urbanizácie* dáva do pomeru „*územie s hustým osídlením“ (1)*, „*územie s priemerne hustým osídlením“* (2) s parametrom „*územie s riedkym osídlením“ (3).*

6. Záver

S využitím logistickej regresie sa nám podarilo odhaliť významné činitele, ktoré pôsobia na riziko prepadu domácností pod hranicu monetárnej chudoby a tiež vyčísliť vplyv týchto faktorov na cieľovú premennú RISK. Získali sme tak predikované pravdepodobnosti výskytu monetárne chudobných domácností pri určitej kombinácií týchto činiteľov. Prezentovaný model nepovažujeme za konečný. V ďalšej práci sa budeme snažiť model ešte optimalizovať.

7. Literatúra

[1] BARTOŠOVÁ, J.: Analysis and Modelling of Financial Power of Czech Households. In Aplimat – Journal of Applied Mathematics, Vol. 2, Nr. 3, Slovak Technical University, Bratislava, 2009, s.31-36, ISSN 1337-6365.

[2] LABUDOVÁ, V.: Analýza monetárnej chudoby na Slovensku. SAS Forum 2008. Bratislava, október 2008.

[3] STANKOVIČOVÁ, I., VOJTKOVÁ, M.: Viacrozmerné štatistické metódy s aplikáciami. Bratislava : Iura Edition, 2007. 261 s. ISBN 978-80-8078-152-1.

[4] MATERIÁLY ZO ŠTATISTICKÉHO ÚRADU SR, web stránka dostupná na: <http://portal.statistics.sk>

Adresa autora:

Lukáš Pastorek, Bc.,
2. roč. magister. štúdia
Fakulta managementu
UK v Bratislave
lukas.pastorek@st.fm.uniba.sk

Poďakovanie: *Príspevok bol vytvorený ako súčasť riešenia projektu VEGA 1/4586/07: Modelovanie sociálnej situácie obyvateľstva a domácností v Slovenskej republike a jej regionálne a medzinárodné porovnania a projektu GAČR 402/09/0515: Analýza a modelování finančního potenciálu českých (slovenských) domácností.*

Analýza vybraného historického zdroja pomocou metódy joint count statistics
Analysis of historical data by the means of joint count statistics

Ivica Paulovičová

Abstract: Historical data usually are not as detailed as nowadays data. Therefore the analysis of historical data is much more difficult. The main idea of this article is to show joint count statistics as a new method of analysing historical data. In this article, there is a comparison between spatial autocorrelation in the year 1773 and 2001 based on joint count statistics.

Key words: joint count statistics, neighbourhood matrix, table of joints, historical data.

Kľúčové slová: joint count statistics, matica susedstiev, tabuľka spojov, historické dáta.

1. Úvod

V príspevku sa zameriam na možnosť použitia metódy joint count pri analýze vybraného historického zdroja. Použitie metódy budem prezentovať na jazykovej štruktúre osád v Hornom Ostrovnom okrese Bratislavskej stolice v roku 1773 a 2001.

2. Metóda joint count statistics

Joint count statistics ako najjednoduchšia metóda identifikácie priestorovej autokorelácie porovnáva namerané a očakávané počty „spojov" (susedstiev). Primárne je určená na štúdium dvojfarebných máp, resp. dichotomických premenných lokalizovaných v polygónoch. Po úprave sa metóda môže aplikovať na k-farebné mapy, teda polymické premenné. [6]

Na začiatku analýzy sa vytvorí binárna matica susedstiev (napr. ***Obrázok 10***), farby polygónov predstavujú ich typy, jednotka indikuje existenciu susedstva. Na určovanie susedstiev existujú tri pravidlá: „queen-based", „rook-based" a „bishop-based". [1]

Tabuľka spojov sa vyplní skutočne nameranými počtami z matice susedstiev a vypočítanými očakávanými počtami. V prípade rovnakého typu regiónov sa očakávaný počet ($E(J_{AA})$) vypočíta podľa vzorca (1), v prípade rôznych typov ($E(J_{AB})$) podľa (2). [6]

$$E(J_{AA}) = J * \frac{n_A * (n_A - 1)}{n * (n-1)} \quad (1)$$

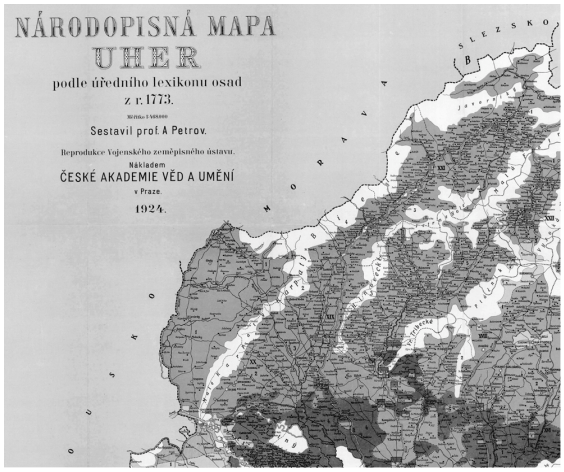
$$E(J_{AB}) = 2 * J * \frac{n_A * n_B}{n * (n-1)} \quad (2)$$

Vo vzorcoch (1) a (2) predstavuje J počet všetkých susedstiev, n počet všetkých priestorových jednotiek, n_A počet priestorových jednotiek typu A a n_B počet priestorových jednotiek typu B.

Porovnaním jednotlivých počtov dostaneme informáciu o charaktere priestorovej autokorelácie. Ak sú namerané a teoreticky očakávané počty veľmi podobné, ide o priestorovú náhodnosť. V prípade počtov susedstiev rovnakých typov priestorových jednotiek nižších ako teoreticky očakávaných ide o negatívnu priestorovú autokoreláciu. V opačnom prípade je priestorová autokorelácia pozitívna. **[6]**

3. Vybraný historický zdroj

Historickým zdrojom v príspevku je Národopisná mapa Uher podľa Úradného lexikónu osád z roku 1773. [3] Ako vidieť na obrázkoch, obsahuje mapovú (viď. *Obrázok 9a)* a podrobnú textovú časť (viď. *Obrázok 9b).* [4]



a)

JAZYKY.

1. Bulharský.

Vinga, m. (k., grnu. — 2) XLV. 1.

2. Český.

Kak (h.) XXXII. 1.

3. Chrvatský.

Agaro, IX. 5.
Alek, Sent ~ (k. — 1) VII. 2.
Arač IX. 5.
Atad, Nadj ~ (k. — 1) IX. 5.
Ata XI. 1.

Babolča, ms. (k. — 1) IX. 5.
Badličan VIII. 2.
Bagola IX. 1.

Cogerštof (k. — 1) II. 1.
Crkovljani VIII. 2.

Čajta (1) VII. 1.
Čakovec — Csáktornya, ms. VIII. 2.
Čatar, D. ~ VII. 2.
Čatar, G. ~ (1) VII. 2.
Čehovec VIII. 2.
Čenča VII. 2.

b)

Obrázok 9: a) Mapová časť Národopisnej mapy Uher
b) Ukážka z textovej časti [4]

Údaje zisťované pri sčítaní a zároveň zachytené v spomínanom diele v časti „Zoznam osád“ (viď. *Obrázok 9b*) sa nevzťahujú na obyvateľov, ale na osady ako celok, takže ide o polytomickú premennú osady (počty obyvateľov neboli zisťované). Z diela možno vyčítať administratívne členenie, jazykovú štruktúru, rozmiestnenie kňazov rôznych vierovyznaní a počty národných učiteľov v jednotlivých osadách. Pre analýzu som si vybrala jazykovú štruktúru osád. V tej dobe sa jazyková a národnostná štruktúra stotožňovali. [4]

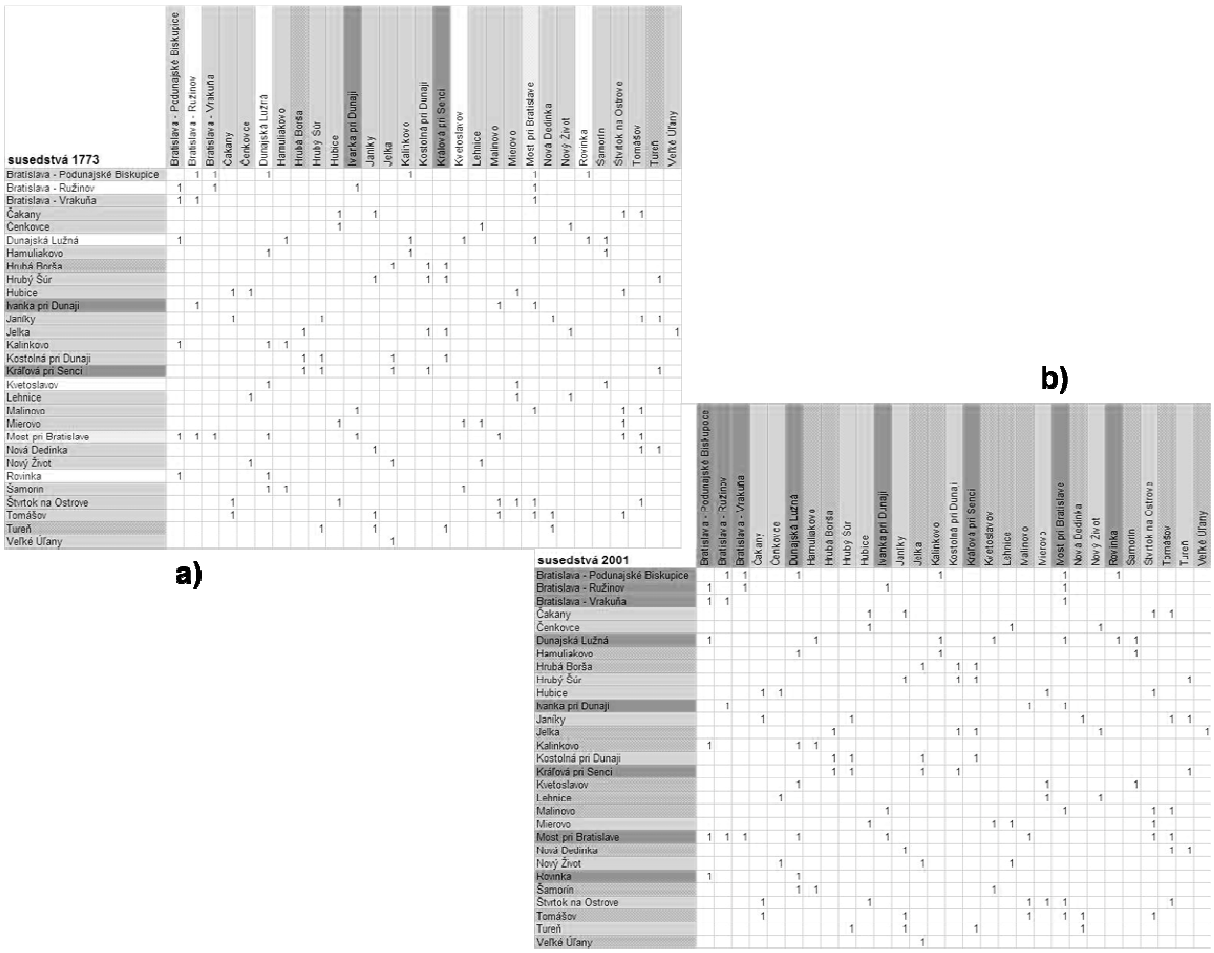
4. Analýza historického zdroja

Opísaná metóda sa využíva prevažne na porovnávanie údajov v dvoch časových obdobiach. [1]**,** [2] Rozhodla som sa porovnať údaje z roku 1773 so sčítaním z roku 2001. Rok 2001 som sa vybrala z dôvodu, že údaje z Národopisnej mapy Uher sú rekonštruované do územnosprávneho členenia z roku 2001. [3]

Údaje o národnostnej štruktúre obyvateľov obcí zo sčítania z roku 2001 boli upravené na národnostné typy, aby bolo možné údaje porovnávať. Národnostné typy boli vytvorené podľa opisu vytvárania jazykových typov pri sčítaní v roku 1773. [3]

V Hornom Ostrovom okrese sa nachádza 24 obcí, medzi ktorými je 58 susedstiev. Keďže dáta za okres sú polytomické, pri analýze som použila upravenú metódu joint count statistics, pre analýzu k-farebných máp. Na určovanie susedstiev som zvolila metódu rook-based.

Jednotlivé susedstvá za roky 1773 a 2001 sú zobrazené v binárnej matici susedstiev (***Obrázok 10***, *a)* rok 1773, *b)* rok 2001). Z matíc vyplýva, že počty susedstiev ani ich lokalizácia sa medzi rokmi nemenia. Menia sa jazykové typy jednotlivých obcí a typy susedstiev. V roku 1773 mali obce rozmanitejšiu jazykovú štruktúru (5 typov) ako v roku 2001 (3 typy).



Obrázok 10: Matica susedstiev Horného Ostrovného okresu: a) rok 1773, b) rok 2001

V roku 2001, ako možno pozorovať v tabuľke spojov (*Tabuľka 14*), sú skutočné počty väčšinou nižšie ako očakávané, preto sa nejedná o náhodné usporiadanie, teda existuje tu istá autokorelácia. Zároveň rozloženie jazykových typov v roku 2001 vykazuje charakteristiky priestorovej náhodnosti.

Jedným z nedostatkov metódy joint count statistics je, okrem prílišnej jednoduchosti, previazanosť na polygóny. Keďže pri analýze okresu bolo použité súčasné územnosprávne členenie, prišlo k istým stratám tým, že nie za všetky súčasné obce boli dostupné informácie za rok 1773.

5. Záver

Ako bolo ukázané v príspevku, historické zdroje dát sa napriek menšej podrobnosti dajú skúmať niektorými jednoduchšími metódami. V tom ale spočíva aj ich nevýhoda.

6. Literatúra

[1] ARTHUR, J. – LEMBO, Jr. 2007. Spatial modeling and analysis: Spatial Autocorrelation: Join count Analysis. In: course Spatial modeling and analysis. College of Agriculture and Life Science, Cornell University. *dostupné na: http://www.css.cornell.edu/courses/ 620/css620.html#resource dňa: 08.11.2009*

[2] LEE, J. – WONG, D. 2000. Statistical Analysis with ArcView GIS. New York: John Wiley & Sons. *dostupné na: http://gis.esri.com/library/userconf/proc00/professional/ papers/PAP392/P392.HTM dňa: 18.10.2008*

[3] PAULOVIČOVÁ, I. 2008. Rekonštrukcia národnostnej štruktúry z Národopisnej mapy Uher. *Bakalárska práca*. Bratislava: Prírodovedecká fakulta Univerzity Komenského v Bratislave. s. 45

[4] PETROV, A. L. 1924. Národopisná mapa Uher. Praha: Nakladateľstvo Českej akadémie vied. 132 s.

[5] ROGERS, A. 2008. Demographic Modeling of the Geography of Migration and Population: A Multiregional Perspective. In: Geographical Analysis, č. 3. Wiley-Blackwell, Oxford. s. 276 – 296. *dostupné na: http://www3.interscience.wiley.com/cgi-bin/fulltext/ 120779702/PDFSTART. dňa: 20.10.2008*

[6] SADAHIRO, Y. 2009. Advanced Urban Analysis – Spatial Analysis and GIS: Spatial Analysis: Spatial autocorrelation. In: course Advanced Urban Analysis – Spatial analysis and GIS. Faculty of Engineering, University of Tokio. *dostupné na: http://ua.t.u-tokyo.ac.jp/ okabelab/sada/docs/pdf_class/Ch06_4c.pdf dňa: 08.11.2009*

Tabuľka 14: Počty spojov v Hornom Ostrovnom okrese v roku 1773 a 2001

	rok 1773		rok 2001	
	skutočný počet	očakávaný počet	skutočný počet	očakávaný počet
všetky spoje	58	58,00	58	58,00
slovensko-slovenské	0	0,21	12	11,43
slovensko-maďarské	4	8,41	4	4,00
maďarsko-maďarské	27	39,93	12	6,43
maďarskoslovensko-slovenské	2	0,84	10	12,57
maďarskoslovensko-maďarské	5	8,41	13	15,71
maďarskoslovensko-maďarskoslovenské	0	0,21	7	7,86
nemecko-nemecké	2	1,26	0	0,00
slovensko-nemecké	1	1,68	0	0,00
maďarsko-nemecké	9	16,81	0	0,00
nemeckomaďarsko-slovenské	1	0,42	0	0,00
nemeckomaďarsko-maďarské	5	4,20	0	0,00
nemeckomaďarsko-nemecké	2	0,84	0	0,00
nemecko-maďarskoslovenské	0	1,68	0	0,00
nemeckomaďarsko-maďarskoslovenské	0	0,42	0	0,00
nemeckomaďarsko-nemeckomaďarské	0	0,00	0	0,00
všetky priestorové jednotky	29		29	
slovenské	2		8	
maďarské	20		10	
maďarskoslovenské	2		11	
nemecké	4		0	
nemeckomaďarské	1		0	

Adresa autora:
Ivica Paulovičová, Bc.
Prírodovedecká fakulta UK v Bratislave
ivica.paulovicova@gmail.com

Demografická budúcnosť krajín V4
Demographic future of countries of V4

Jana Tichá

Abstract: The pattern of recent population development in all countries of V4 has been characterized by several changes which have significant impact on future development. Therefore is very important to make estimates of the future development of population and its age structures. This article is focused on ageing of population which is currently perceived as a global phenomenon. The steady increase in the oldest age groups adversely affects the social, economic and cultural aspects of human life. Notwithstanding the increasing intensity of fertility, mortality and migration, all countries of the Visegrad group is waiting on a short term reduction in population growth, which gradually turns to the decline in population and intensification of the aging process.

Keywords: age structures, population ageing, counties of V4, fertility, mortality, migration, population growth

Kľúčové slová: veková štruktúra, populačné starnutie, krajiny V4, pôrodnosť, úmrtnosť, migrácia, populačný rast

1. Úvod

Súčasné obdobie je obdobím výrazných zmien v demografickom správaní obyvateľstva, ktoré sú výsledkom predovšetkým aktuálnej spoločensko–ekonomickej situácie. Neustále znižovanie pôrodnosti a plodnosti, predlžovanie strednej dĺžky života, zmenšovanie prirodzeného prírastku, zvyšovanie priemerného veku pri sobáši ako aj pri pôrode, starnutie obyvateľstva, a tým aj zvyšujúce sa ekonomické zaťaženie obyvateľstva. To všetko sú fakty, ktoré sú ovplyvňované demografickým správaním obyvateľstva. Otázkou ostáva, či ide o dlhodobý trend vo vývoji alebo či možno v budúcnosti očakávať zmenu k lepšiemu. Práve v tom tkvie dôležitosť a potreba tvorby populačných prognóz. Výsledkom prognózy býva veková štruktúra a veľkosť budúcej populácie zistená na základe vstupných parametrov, použitej metódy a vybraného scenára prognózy. Dôležitým, no nie veľmi priaznivým výsledkom populačných prognóz je v poslednom období starnutie populácie, ktoré patrí medzi základné rysy súčasného populačného vývoja. Odhad budúceho vývoja počtu ako aj vekovej štruktúry obyvateľov je stále viac žiadaná informácia. Podľa Káčerovej a Blehu (2007) je intenzita a dôležitosť tohto procesu výrazná v globálnom meradle, hlavne v posledných sto rokoch. Populačné starnutie je teda kauzálne spojené s demografickým prechodom a jeho ukončením vo viac vyspelej časti sveta.

V nasledujúcom texte budú priblížené výsledky národných prognóz pre jednotlivé štáty Višegrádskej štvorky. Zamerali sme sa predovšetkým na prognózy Eurostatu (Europop 2004) a OSN (2006) kvôli jednotnosti údajov. Pokiaľ nebude uvedené inak, bude sa text zaoberať stredným (najpravdepodobnejším) variantom.

2. Očakávaný vývoj celkového počtu obyvateľov

Základným výsledkom prognóz je zníženie počtu obyvateľov všetkých štátov Višegrádskej štvorky a jeho demografické starnutie. Príčinou poklesu bude neustále sa prepadajúci prírastok prirodzenou menou, ktorý bude len z časti vyrovnaný pozitívnym migračným saldom (bilanciou).

Tabuľka 1: Predpokladaný vývoj celkového počtu obyv. krajín V4 v rokoch 2005-2025

Prirodzený prírastok/úbytok					
Roky	**2005**	**2010**	**2015**	**2020**	**2025**
Česká republika	-6 016	-3 413	-13 505	-23 512	-36 509
Maďarsko	-36 382	-31 329	-34 907	-38 762	-43 227
Poľsko	-29 320	3 454	-17 411	-60 746	-110 518
Slovensko	706	795	-1 996	-7 752	-15 129
Migračný prírastok/úbytok					
Roky	**2005**	**2010**	**2015**	**2020**	**2025**
Česká republika	15 790	25 857	27 692	24 739	21 260
Maďarsko	14 745	19 086	22 064	22 407	18 016
Poľsko	-27 837	-15 291	8 484	13 983	4 855
Slovensko	2 020	3 210	4 955	5 001	4 029
Celkový prírastok/úbytok					
Roky	**2005**	**2010**	**2015**	**2020**	**2025**
Česká republika	9 774	22 444	14 187	1 227	-15 249
Maďarsko	-21 637	-12 243	-12 843	-16 355	-25 211
Poľsko	-57 157	-11 837	-8 927	-46 763	-105 663
Slovensko	2 726	4 005	2 959	-2 751	-11 100

Zdroj:Europop 2004

Tabuľka 2: Očakávaný vývoj počtu obyvateľov v krajinách V4 (všetky varianty)

Rok	**Česko**			**Slovensko**			**Maďarsko**			**Poľsko**		
	nízky	**stredný**	**vysoký**	**nízky**	**stredný**	**vysoký**	**nízky**	**stredný**	**vysoký**	**nízky**	**stredný**	**vysoký**
2005	10 201	10 236	10 273	5 375	5 384	5 392	10 086	10 086	10 086	38 196	38 196	38 196
2010	10 141	10 283	10 432	5 370	5 401	5 439	9 850	9 980	10 029	37 533	38 272	38 272
2015	10 041	10 283	10 553	5 351	5 416	5 503	9 560	9 783	10 004	36 645	38 512	38 512
2020	9 875	10 284	10 700	5 307	5 417	5 568	9 244	9 621	9 994	35 511	38 629	38 629
2025	9 386	10 102	10 823	5 233	5 396	5 627	8 929	9 449	9 963	34 229	38 412	38 412

Zdroj: OSN 2006

Početná veľkosť populácie jednotlivých štátov sa v horizonte prognózy na základe jednotlivých variantov výrazne odlišuje. Najmarkantnejší pokles počtu obyvateľov zaznamená Maďarsko, ktoré bude mať pri nízkom aj strednom variante už na budúci rok menej ako 10 miliónov obyvateľov. Pri vysokom variante k tomu dôjde až v horizonte prognózy. Pri ostatných štátoch vysoký variant predpokladá nárast početnosti na celej dĺžke prognózovaného obdobia, zatiaľ čo pri strednom sa očakáva pomalý plynulý rast len do roku 2020. Potom nastane mierny pokles. Pesimistický nízky variant predpokladá znižovanie počtu obyvateľov na celej dĺžke sledovaného obdobia. Príčinou celkového úbytku obyvateľov Maďarska bude prirodzený úbytok, ktorý nedokáže vyrovnať ani kladná migračná bilancia. Obdobne ako pri Maďarsku, tak aj pri Česku očakávame až do horizontu prognózy vysoké úbytky obyvateľov prirodzenou menou. Tie budú vyrovnané pozitívnym migračným saldom okrem roku 2025, kedy je predpokladaný celkový úbytok obyvateľov. Na celej dĺžke sledovaného obdobia budeme pri Poľsku zaznamenávať celkové úbytky obyvateľov. Do roku 2015 budú zapríčinené negatívnou migráciou, zatiaľ čo od tohto roku do horizontu predovšetkým prirodzeným úbytkom. Na Slovensku zaznamenáme v poslednej tretine prognózovaného obdobia, t.j. po roku 2020 celkový úbytok vyvolaný záporným prirodzeným prírastkom.

3. Očakávané starnutie obyvateľstva

Závažnejší ako samotný úbytok obyvateľov je jeho starnutie charakterizované nárastom počtu aj relatívneho zastúpenia starších osôb v populácii na úkor mladého a produktívneho obyvateľstva. Hlavnou príčinou je pokles pôrodnosti a neustále predlžovanie strednej dĺžky života.

Tabuľka 3: Očakávaná veková štruktúra obyvateľov krajín V4

	Rok	0-14		15-64		65+		80+	
		abs	%	abs	%	abs	%	abs	%
PL	2005	590 087	15,48	27 083 387	71,1	5 131 376	13,46	1 140 239	2,99
	2010	572 583	15,03	27 212 635	71,4	5 153 481	13,53	1 314 176	3,45
	2015	577 495	15,19	26 311 643	69,2	5 929 459	15,60	1 487 799	3,91
	2020	589 930	15,59	24 976 993	66,0	6 953 590	18,38	1 566 446	4,14
	2025	560 570	14,97	23 988 339	64,1	7 844 056	20,95	1 537 400	4,11
ČR	2005	1 518 945	15,17	7246615	72,39	1 245 325	12,44	307 406	3,07
	2010	1 406 600	13,75	7253316	70,89	1 571 275	15,36	358 645	3,51
	2015	1 376 140	13,74	6811752	68,04	1 824 123	18,22	384 716	3,84
	2020	1 364 114	13,78	6478731	65,43	2 059 003	20,79	392 595	3,96
	2025	1 325 622	13,51	6286602	64,07	2 199 453	22,42	482 308	4,92
SR	2005	893 725	16,57	3 851 111	71,64	631 052	11,74	130 997	2,44
	2010	816 668	15,06	3 887 428	72,50	658 033	12,27	145 716	2,72
	2015	804 670	14,79	3 818 299	71,34	728 924	13,62	157 366	2,94
	2020	801 356	14,73	3 658 250	68,75	861 409	16,19	163 575	3,07
	2025	760 053	14,06	3 515 556	66,80	986 928	18,75	185 473	3,52
HU	2005	1 552 638	15,41	6 925 347	68,74	1 596 076	15,84	352 067	3,49
	2010	1 461 333	14,64	6 852 211	68,65	1 668 378	16,71	389 407	3,90
	2015	1 420 282	14,44	6 642 368	67,54	1 771 600	18,01	427 057	4,34
	2020	1 396 954	14,41	6 324 732	65,25	1 971 596	20,34	453 375	4,68
	2025	1 371 757	14,31	6 111 103	63,73	2 105 514	21,96	513 839	5,36

Zdroj: OSN 2006

Typickým znakom vekovej štruktúry je jej nepravidelnosť, charakterizovaná zárezmi spôsobenými hlavne druhou svetovou vojnou a hospodárskou krízou z 30-tych rokov minulého storočia. Prechod silných a slabých vekových generácií jednotlivými vekovými kategóriami zapríčiní zmeny relácií medzi hlavnými vekovými skupinami. Do budúcnosti treba počítať s pokračujúcim znižovaním početnosti a relatívneho zastúpenia detskej aj produktívnej zložky obyvateľov na úkor starého obyvateľstva. Predpokladá sa pokles predproduktívneho obyvateľstva na celej dĺžke prognózovaného obdobia, a to o 2-4 percentuálne body. Výraznejšie zmeny nastanú v kategórii 15-64 ročných, kde sa očakávajú poklesy v rozpätí 5-10%. Oslabenie po roku 2010 nastane v dôsledku odchodu populačne silných ročníkov narodených na konci druhej svetovej vojny a po jej skončení a príchodu početne slabých ročníkov narodených v 90-tych rokoch. Ďalšia vlny poklesu nastane v horizonte prognózy, kedy hranicu 64 rokov začnú pomaly opúšťať generácie narodené v 70-tych rokoch. K najmarkantnejším zmenám dôjde vo vekovej kategórii 65 a viac ročných, kde je očakávaný nárast 6-10%. V Poľsku to znamená zvýšenie početnosti tejto skupiny do roku 2025 o 2,5 mil. obyvateľov, v Čechách o 900 a v Maďarsku o 600 tisíc obyvateľov, na Slovensku to predstavuje približne 350 tisíc obyvateľov. Štruktúrne zmeny nastanú aj vo vnútri tejto kategórie. Najrýchlejší nárast sa očakáva v kategórii 80 a viac ročných, a to o približne 1-3 percentuálne body. Prvé výrazné zvýšenie nastane v roku 2010, kedy budú hranicu 80-tich rokov prekračovať generácie z 30-tych rokov. Druhá vlna je očakávaná v závere prognózy pri prechode ročníkov narodených na konci druhej svetovej vojny. Zatiaľ

čo doteraz sme mohli hovoriť o starnutí populácie zdola vekovej pyramídy, ktoré je charakterizované poklesom podielu detskej zložky obyvateľstva a zároveň pomalým nárastom starého obyvateľstva. V súčasnosti už môžeme v dôsledku prudkého nárastu početnosti aj podielu starého obyvateľstva a neustále sa zmenšujúceho zastúpenia detskej zložky obyvateľstva hovoriť o starnutí zhora vekovej pyramídy. Starnutie populácie vyvolá v blízkej budúcnosti vyššie zaťaženie ekonomicky aktívneho obyvateľstva, a to aj napriek postupnému zvyšovaniu vekovej hranice odchodu do dôchodku. Na prahu prognózy pripadala približne jedna závislá osoba na dve nezávislé, zatiaľ čo v horizonte prognózy sa toto zaťaženie zvýši o približne 20%, t.j. na 100 nezávislých osôb bude pripadať okolo 70 závislých osôb. Bude sa meniť aj štruktúra závislých osôb, pričom oveľa vyššie zastúpenie bude mať obyvateľstvo poproduktívneho veku.

Tabuľka 4: Očakávaný index ekonomického zaťaženia v krajinách V4

Rok	2005	2010	2015	2020	2025
Česko	53,22	58,54	66,06	69,72	72,68
Slovensko	48,72	49,61	55,59	61,84	65,83
Poľsko	51,05	51,11	57,71	66,62	71,53
Maďarsko	58,72	59,63	65,55	70,69	71,81

Zdroj: Europop 2004

4. Záver

Jednou z najviditeľnejších spoločenských a ekonomických zmien v súčasnosti je starnutie populácie. Proces starnutia sa bude v najbližších desaťročiach zrýchľovať, čo je výsledkom znižovania pôrodnosti a predlžovania strednej dĺžky života. Ide o nezvratný proces, ktorý nie je možné zastaviť ani spomaliť. Prirodzeným dôsledkom starnutia je nárast počtu osôb s nárokom na starobný dôchodok. Aj keď sa chorobnosť obyvateľstva bude pravdepodobne znižovať, tak vyšší podiel starého a najstaršieho obyvateľstva spôsobí obrovský nárast výdavkov na lekársku starostlivosť. Analogický je vplyv populačného starnutia aj na viaceré ďalšie oblasti spoločenského života. Na druhej strane sa nájdu aj názory (Loužek, M., 2008), podľa ktorých by sme sa starnutia populácie rozhodne nemali obávať. Rodí sa síce čím ďalej tým menej detí a dožívame sa neustále vyšších vekov, ale oba tieto fakty svedčia skôr o blahobyte ako o chudobe či hroziacej katastrofe.

5. Literatúra

[1] BLEHA, B. – KÁČEROVÁ, M. 2007. Teoretické východiská populačného starnutia a retrospektívny pohľad na starnutie Európy. In: Slovenská štatistika a demografia, č.3, 2007, ISSN 1210-1095, s. 43-61.

[2] BLEHA, B. – VAŇO, B. 2007. Prognóza vývoja obyvateľstva SR do roku 2025. Bratislava: INFOSTAT, 2007, 62 s.

[3] ČESKÝ STATISTKY ÚŘAD. 2003. Projekce obyvatelstva České republiky do roku 2050. Praha: 23 s.

[4]Europop 2004:
http://epp.eurostat.ec.europa.eu/portal/page?_pageid=1996,45323734&_dad=portal&_schema=PORTAL&screen=welcomeref&open=/popula/proj&language=en&product=EU_MASTER_population&root=EU_MASTER_population&scrollto=0

[5] LOUŽEK, M. 2008. Je stárnutí populace tragédií?. In: Ekonomický časopis, č.6, 2008, s. 565 - 581.

[6] THE WORLD POPULATION PROSPECTS. The 2006 Revision. 2007. New York: United Nations, 2007.

Adresa autora:

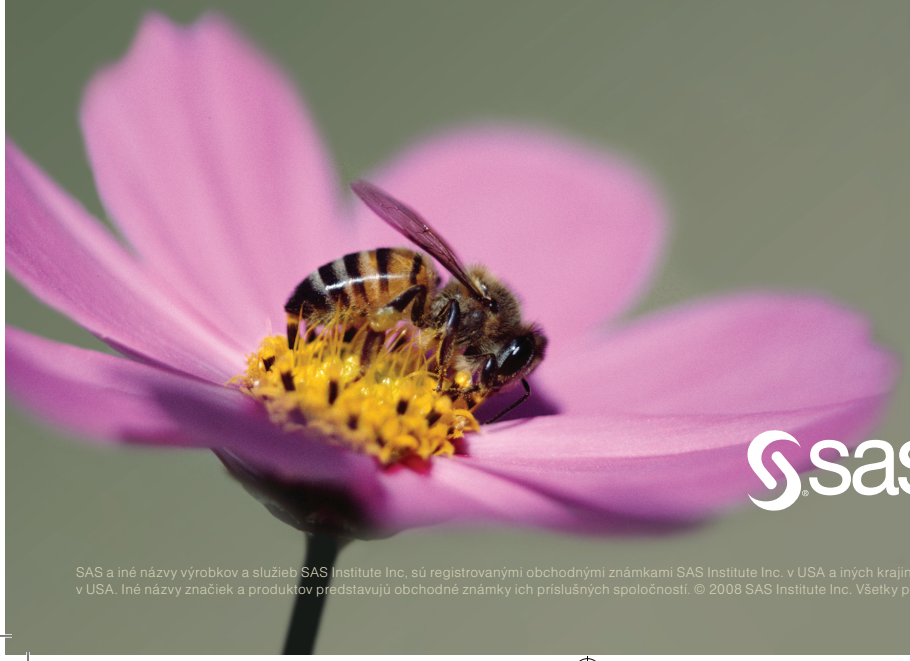
Jana Tichá, Bc., 2. ročník Mgr. štúdia,
Prírodovedecká fakulta UK v Bratislave, e-mail: janka.ticha@gmail.com

OBSAH / CONTENT

Jedna včela vyrobí za svoj život jednu dvanástinu lyžičky medu

Nevidí ako sa jej tvrdá práca pretaví na niečo veľké a aký má skutočne význam.
Vy to však vidieť môžete. S osvedčeným softvérom pre vzdelávanie od spoločnosti SAS.

www.sas.com/slovakia/uni
www.sas.com/bees



sas | THE POWER TO KNOW.